7-31-2015

The Informative g-Prior vs. Common Reference Priors for Binomial Regression With an Application to Hurricane Electrical Utility Asset Damage Prediction

Nathan R. Lally
University of Connecticut - Storrs, nathan.lally@uconn.edu

The Informative g-Prior vs. Common Reference Priors for Binomial Regression

With an Application to Hurricane Electrical Utility Asset Damage Prediction

Nathan Ross Lally

B.A. University of Connecticut, 2007, 2014

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science at the University of Connecticut 2015

APPROVAL PAGE

Masters of Science Thesis

The Informative g-Prior vs. Common Reference Priors for Binomial Regression

With an Application to Hurricane Electrical Utility Asset Damage Prediction

Presented by Nathan Ross Lally, B.A. Mathematics/Statistics

Advisor____

James Bridgeman

Associate Advisor____

Brian Hartman

Associate Advisor____

Yung S Choi

University of Connecticut 2015

The Informative g-Prior vs. Common Reference Priors for Binomial Regression

Abstract:

Eliciting appropriate prior information from experts for a statistical model is no easy task. Expressing this information in terms of hyperparameters of prior distributions on abstract model parameters can be nearly impossible, especially after data augmentation or transformation. In previous work on logistic and binomial regression models, Hanson et al. (2014) assert that ``experts are confident only in their assessment of the population as a whole'' and propose a version of the g-prior which effectively places a standard beta distributed prior on the overall population probability of success. We explore the efficacy of using the informed g-prior in a real prediction problem involving electrical utility asset damages due to hurricanes in Connecticut. Prior information is elicited from a group of engineers at the electrical utility and several methods are used to select hyper-parameters for the g-prior. The out-of-sample predictive accuracy of these informed models is compared to the performance of models constructed under common reference priors (Jeffreys's, Gelman et al. (2008), and a noninformative specification of the g-prior) using IS-LOO (Vehtari et al., 2014; Vehtari and Gelman, 2014), root mean squared error (RMSE), and other statistics. In this application, with carefully selected hyper-parameters, binomial regression models using the informed g-prior match the predictive accuracy of common reference priors and offer no distinct advantage. Careless selection of hyper-parameters can however, lead to substantial reduction in predictive accuracy. Surprisingly, the noninformative specification of the g-prior performed marginally better than all other models tested in this paper; contradicting one of the findings in Hanson et

al. (2014) In addition, we show the predictive accuracy gained by modeling spatial correlation in the residuals and prove that such models substantially outperform some statistical learning models growing in popularity in this field.

The Informative g-Prior vs. Common Reference Priors for Binomial Regression

Table of Contents:

Bayesian Analysis , Number , pp.

The Informative g-Prior vs. Common Reference Priors for Binomial Regression in an Application to Hurricane Electrical Utility Asset Damage Prediction

Nathan Lally

1 Introduction

Eliciting appropriate prior information from experts for a statistical model is a difficult prospect. Garthwaite et al. (2005) suggest the elicitation process involves two main parties, the facilitator (perhaps a statistician) and the expert (familiar with the process or phenomena being modeled). They have the goal of formulating the expert's knowledge in probabilistic form. If the problem at hand, like our own as we will demonstrate, requires the construction of a binomial regression model with a potentially large model space (the set of possible model parameterizations) this expert knowledge can be difficult to translate to regression parameters in the form of probability distributions. The task is made even more difficult with transformations to the data space such as standardization of covariates, log transformations, inclusion of polynomial terms, orthogonal rotation or others.

Until recently, research on informative priors for binomial and logistic regression models has focused principally on model selection (see Chen et al. (1999, 2003) for examples) which is relevant for problems where inference is the most important outcome. If at least some historical data is available, empirical Bayes methodology can be used to derive prior distributions, or a power prior approach (Ibrahim and Chen, 2000) may be feasible.

In our particular application we seek to build a model to predict future electrical utility (Eversource Energy) asset damages due to hurricanes in Connecticut with only

data from the recent storms Irene and Sandy. Having no historical data prior to these storms, the power prior method is not ideal because we prefer to use one storm's data as a training set and the others a test set to validate the model's out-of-sample predictive accuracy. Further, because our set of potential predictor variables is large, heavily correlated, and we are mostly interested in predictive accuracy over inference, we perform an orthogonal rotation of the data matrix. This serves to reduce the dimension of the problem and subsequently the need for probabilistic variable selection.

The appeal of using a subjective Bayes approach in our hurricane damage modeling application is clear. We lack historical data in a usable form for predictive modeling but we have access to several experienced engineers at the utility company with knowledge of asset damages in previous storms. However, as previously stated, selecting reasonable prior distributions on regression model parameters associated with transformed variables is intractable. Fortunately, Hanson et al. (2014) proposes a version of the g-prior (Zellner, 1983) which allows the statistician or practitioner to effectively place a standard beta distributed prior on the overall population probability of success in a binomial regression model, i.e. $\theta \sim \mathcal{B}e(a_\pi, b_\pi)$. Using this informative g-prior we construct three predictive models where the hyperparameters were elicited from subject matter experts using three different methodologies. These models with informed priors are compared by their respective predictive accuracy to models utilizing common reference priors such as the Jeffreys's prior, a weakly informative Cauchy prior with data augmentation suggested by Gelman et al. (2008), and a noninformative specification of the g-prior (where $\theta \sim \mathcal{B}e(a_\pi = 1/2, b_\pi = 1/2)$).

We conclude that models using informed g-priors perform comparably, yet without specific advantage, to models with common reference priors provided the hyperparameters selected are reasonable. Models without well considered hyperparameters perform significantly worse than their noninformative counterparts and caution is advised in the elicitation process. Surprisingly, in our application the prior specification that produced the best predictive performance was the noninformative default specification of the g-

prior. This finding contradicts the results of a simulation study in Hanson et al. (2014) showing the Gelman prior outperformed the default specification of the g-prior.

Additionally, to make our research comparable to literature on hurricane damage modeling in the field of reliability engineering and risk analysis, we explore the benefit of modeling spatial correlation (a subject of some debate) in the residuals of our best predicting binomial regression model. We also compare our models to two types of tree based models growing in popularity in this discipline. We conclude that properly accounting for random spatial effects significantly improves predictive accuracy and suggest our methodology may produce more accurate predictions than tree based machine learning methods in this setting.

Our paper is organized as follows, first in section 2 we introduce the three types of prior distributions used in our binomial regression models. In 3 we then discuss our methodology for estimating spatial correlation in the residuals of our best fitting models. Next in 4 we provide a brief overview of the criteria used to evaluate the predictive performance of our models. Section 5 describes the hurricane damage data used in this study, potential spatial correlation in the response variable, and our data preparation process. Finally section 6 details our prior elicitation process for g-prior hyperparameters, section 7 summarizes our results, and section 8 concludes the paper.

2 Three Types of Prior Distribution for Binomial Regression Models: From Noninformative to Strongly Informative

Throughout this article we assume the following definitions. The binomial regression model has a conditional density defined by,

$$p(y_i|\boldsymbol{x}_i, N_i, \beta) = \binom{N_i}{y_i} [F(\boldsymbol{x}_i'\beta)]^{y_i} [1 - F(\boldsymbol{x}_i'\beta)]^{N_i - y_i} \text{ for } i = 1, 2, ..., n \tag{1}$$

where $0 \leq y_i \leq N_i$ is a realization of a random binomial response variable, N_i is the total possible number of successes for a given observation i. The vector of covariates for each observation is given by $\boldsymbol{x}_i' \,\epsilon\, \mathbb{R}^p$ (rows of the design matrix $\boldsymbol{X} \,\epsilon\, \mathbb{R}^{n\text{x}p}$) with the first element $x_{i,0} = 1$ associated with the intercept term. The vector of regression coefficients is given by $\beta \,\epsilon\, \mathbb{R}^p$. In our application the function $F(\cdot)$ may be any continuous twice differentiable cumulative density function (cdf) with associated probability density function (pdf) $f(z) = {}^{dF(\cdot)}/_{dz}$. The link function is therefore $F^{-1}(\cdot)$. In this article we work exclusively with the logit link where,

$$F(\boldsymbol{x}_i'\beta) = \frac{1}{1 + e^{-\boldsymbol{x}_i'\beta}} \tag{2}$$

$$F^{-1}(\boldsymbol{x}_i'\beta) = \frac{F(\boldsymbol{x}_i'\beta)}{1 - F(\boldsymbol{x}_i'\beta)} = \boldsymbol{x}_i'\beta \tag{3}$$

and additionally, the likelihood for β is defined by,

$$L(\beta|\boldsymbol{X}, \boldsymbol{y}) = \prod_{i=1}^{n} \binom{N_i}{y_i} [F(\boldsymbol{x}_i'\beta)]^{y_i} [1 - F(\boldsymbol{x}_i'\beta)]^{N_i - y_i} \text{ for } i = 1, 2, ..., n \tag{4}$$

2.1 Noninformative: Jeffreys's Prior & the Informed g-Prior as a Reference Prior

Jeffreys's Prior

The Jeffreys's prior is one of the most widely used noninformative priors for generalized linear models (GLM). Defined as the square root of the determinant of the Fisher information matrix (see (Jeffreys, 1946) for details), for binomial regression models it takes the following form,

$$\pi(\beta|\boldsymbol{X}) \propto |\boldsymbol{X'W}(\beta)\boldsymbol{X}|^{1/2} \tag{5}$$

where, $\boldsymbol{W}(\beta) = \text{diag}(w_1(\beta), w_2(\beta), ..., w_n(\beta))$, and

$$w_i(\beta) = \frac{N_i\,[f(x_i'\beta)]^2}{F(x_i'\beta)\,[1 - F(x_i'\beta)]} \text{ for } i = 1, 2, .., n \tag{6}$$

One of the main advantages of the Jeffreys's prior for GLMs over common alternatives such as the improper flat uniform prior is that it guarantees a proper posterior distribution in many if not most cases (Ibrahim and Laud, 1991). Additionally, for binomial regression models where the design matrix $\boldsymbol{X}$ is of full rank and the logistic link function is used, the Jeffreys's prior has many appealing properties. It can be shown the prior is proper, a moment generating function (mgf) exists, symmetric about $\mathbf{0}$, unimodal, has well understood tail behavior, and is invariant under one-to-one linear transformations in the covariates (Chen et al., 2008).

In light of these characteristics, the Jeffreys's prior serves as a default reference prior for the binomial regression models considered in this paper. All other model's measures of fit and predictive performance will be compared to the models evaluated under the Jeffreys's prior.

2.2 Weakly Informative: Cauchy Priors

Though appealing due to some of their intrinsic properties, completely or near completely noninformative priors are not without their limitations. For example, using uniform priors or similarly a model evaluated under maximum likelihood estimation (MLE), can be computationally unstable and fail when separation occurs (Zorn, 2005). Since various penalized likelihood corrections and Bayesian estimation techniques can mitigate the separation problem, perhaps more importantly, noninformative priors such as the Jeffreys's are difficult to interpret as prior information in regression models and

may not adequately constrain parameter estimates (Gelman et al., 2008). To address these issues Gelman et al. (2008) suggest placing a minimally informative prior on the regression coefficients which assumes that there is little probability a change in the value of a covariate will correspond to a change as large as 5 on the log scale. The prior is meant to be an application invariant default prior (as opposed to fully informative or noninformative) and has the form,

$$\beta_k \overset{\text{i.i.d.}}{\sim} \mathcal{C}(0, 2.5) \text{ for } k = 1, ..., p \tag{7}$$

or independent and identically distributed (i.i.d.) Cauchy priors with location 0 and scale 2.5.

The Gelman prior also requires a corresponding covariate modification scheme; binary covariates are shifted to have a mean 0 and to differ by 1 in their upper and lower conditions while continuous covariates are centered on 0 and scaled to have a standard deviation of 0.5 to match the binary covariates.

One of the potential criticisms of the Gelman prior is the inability to accommodate correlation between covariates. However, in our application, an orthogonal transformation of the design matrix (see 5.2) makes this potential problem irrelevant.

2.3 Informative: Informative g-Prior

Recently, Hanson et al. (2014) established the following construction of the g-prior (see Zellner (1983) for background on g-priors / reference informative priors (RIPs))for β in binomial regression models,

$$\beta \sim \mathcal{N}_p(b \cdot \boldsymbol{e}, gn(\boldsymbol{X'X})^{-1}) \tag{8}$$

where the vector $\boldsymbol{e} \in \mathbb{R}^p$ and the element $e_1 = 1$ with all others $(e_2, ..., e_p = 0)$ and the

parameters b and g are fixed and are functions of the hyperparameters of a $\mathcal{B}e(a_\pi, b_\pi)$ density. The g-prior is location-scale invariant, accommodates correlation between covariates, and mitigates the problem of quasi or complete separation in logistic regression (Hanson et al., 2014).

As its name suggests, the informative g-prior can be used to construct informed prior distributions for binomial regression parameters. Since the g-prior is a conditional means prior (Hanson et al., 2014; Bedrick et al., 1996) we can view our prior as being for the mean response given a particular set of covariates. In this application we consider a prior placed on a parameter θ representing the overall probability of success in a population given a binomial likelihood. In 6 we detail several methods for eliciting prior information and selecting values for the g-prior hyperparameters. Models constructed with these hyperparameters are compared to each other and to the other models with less informative priors.

The authors suggest setting the hyperparameters $a_\pi = b_\pi = 0.5$ (conveniently the Jeffreys's prior for a binomial random variable) for problems where prior information is lacking. This special case is also employed as a noninformative reference prior in our models. We compare the performance of the g-prior as a reference prior to our other noninformative, and weakly informative prior specifications. Also, we compare the aforementioned noninformative form of this prior to the case where a_π and b_π are estimated from prior information elicited from expert opinion and historical data in a predictive modeling application.

3 Estimating Spatial Correlation in the Residuals

Several articles have investigated the potential for improved predictive performance through spatial effects when modeling hurricane electrical utility asset damage or outages. Notably, Liu et al. (2008) makes the case that spatial correlation should be considered when predicting outages due to hurricanes and uses a negative binomial generalized

linear mixed model (GLMM) with spatial random effects to do so. The authors realize little to no benefit from incorporating spatial random effects. We believe this conclusion may be influenced by model misspecification (see 5.1 for details) and the computational disadvantages of classical inference procedures rather than a lack of significant spatial correlation.

Instead of using classical inference procedures to model spatial effects, we use a fully Bayesian conditional auto-regressive (CAR) approach best known as the BYM model Besag et al. (1991). Under the BYM model two additional components are added to our binomial regression model. We define parameter τ_i as the geographically unstructured component of heterogeneity in damage risk and ϕ_i as the spatial component of between area variation in damage risk (Best et al., 2005) such that,

$$\tau_i \sim \mathcal{N}(0, \sigma_\tau^2) \tag{9}$$

$$\phi_i | \phi_{j \neq i} \sim \mathcal{N}\left(\bar{\phi}_i, \frac{\sigma_\phi^2}{m_i}\right) \tag{10}$$

where $\bar{\phi}_i = {}^1/_{m_i} \sum_{j \epsilon \delta_i} \phi_j$, δ_i is the set of neighbors of region i, and m_i is number of neighboring regions. Though various neighborhood and weighting schemes can be employed for the prior on $\phi_i | \phi_{j \neq i}$ we use the simple queen's case neighborhood structure and weight all neighbors equally since spatial correlation is not the focus of this paper.

In our application, we evaluate the spatial properties of the residuals of our best fitting model under the various prior specifications. In this manner we consider the benefit to predictive accuracy of a random spatial component but we do not complicate the likelihood, nor computation, of our original models.

For example, suppose $\hat{\boldsymbol{\mu}} = \boldsymbol{X}\hat{\beta}$ such that $\hat{\mu}_i = \boldsymbol{x}_i'\hat{\beta}$ is estimated under the Jeffreys's, Gelman, or g-prior. To estimate the spatial correlation of a given model's residuals our likelihood becomes simply,

$$L(S|\hat{\boldsymbol{\mu}}, \boldsymbol{y}) = \prod_{i=1}^{n} \binom{N_i}{y_i} [F(\hat{\mu}_i + S_i)]^{y_i} [1 - F(\hat{\mu}_i + S_i)]^{N_i - y_i} \text{ for } i = 1, 2, ..., n \tag{11}$$

where each $S_i \,\epsilon\, \boldsymbol{S} = \tau_i + \phi_i | \phi_{j \neq i}$. Allowing us to make inference about predictive distribution of the spatial models comparable to the models without the spatial component. We place weakly informative conjugate inverse-gamma priors on the scale parameters of the heterogeneity and spatial components such that $\sigma^2_{\tau,\phi} \sim \mathcal{IG}(\alpha = 2, \beta = 2000)$.

4 Model Evaluation Criteria

To determine which prior distribution specification resulted in the best predictive model or if modeling spatial correlation is beneficial we use several criteria. Models are evaluated using an approximation of leave-one-out cross validation (LOO), several predictive accuracy statistics obtained from out-of-sample (test set) predictions, and we also compare our binomial regression models to several tree based machine learning methods popular in reliability engineering and risk analysis.

4.1 Approximate Leave-One-Out Cross Validation (IS-LOO)

The primary criteria we use is an approximation (due to computational convenience) of Bayesian leave-one-out cross validation (LOO). The LOO statistic serves as a measure of expected predictive performance given new data generated from the same random process. For future observations and covariates ($\tilde{y}$ and $\tilde{x}$ respectively) the predictive distribution is given by,

$$\pi(\tilde{y}|\tilde{x}, \mathcal{D}) = \int \pi(\tilde{y}|\tilde{x}, \mathcal{D}, \beta)\pi(\beta|\tilde{x}, \mathcal{D})d\beta \tag{12}$$

where $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$ is our observed data (training set). Then the following is given by Vehtari et al. (2014). The expected predictive performance using the log score and unknown true distribution of the future observation $\pi_t(\tilde{y})$ is

$$\int \pi_t(\tilde{y}|\tilde{x}) \log(\pi(\tilde{y}|\tilde{x}, \mathcal{D})) d\tilde{x} d\tilde{y} \tag{13}$$

Which can be approximated by re-using the observations and computing leave-one-out Bayesian cross-validation estimate

$$\text{LOO} = \frac{1}{n} \sum_{i=1}^{n} \log(\pi(y_i|x_i, \mathcal{D}_{i-1})) \tag{14}$$

where $\mathcal{D}_{i-1}$ contains all observations x and y except (x_i, y_i). Due to a large sample size $n =$ we do not perform a complete leave-one-out cross validation but use an approximation obtained by importance sampling (IS-LOO) described in Vehtari and Gelman (2014).

4.2 Watanabe-Akaike Information Criterion (WAIC)

In addition to IS-LOO we also consider the Watanabe-Akaike Information Criterion (Watanabe, 2010) a statistic which is asymptotically equivalent to Bayes cross-validation loss. WAIC is defined as the log pointwise predictive density of the observed data minus the estimated effective number of parameters. WAIC seeks to approximate the same value as IS-LOO but has come under some criticism as of late due to concerns over accuracy. It is therefore considered secondary to IS-LOO in this study.

4.3 Measures of Out-Of-Sample Predictive Accuracy

In addition to IS-LOO and WAIC, we calculate several other simple statistics indicative of predictive accuracy. Since our data is from two separate hurricanes (Irene and Sandy) occurring over the same geographic space, we can naturally use one hurricane as a model

training set and the other as a test set. Alternating each hurricane's data as a training set and then test set, we compute the following statistics on the out-of-sample data (test set) in each instance.

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{\sum_{i-1}^{n}(\hat{y}_{i,\text{test}} - y_{i,\text{test}})^2}{n_{\text{test}}}} \tag{15}$$

$$\text{Coverage\%} = \frac{1}{n_{\text{test}}}\sum_{i=1}^{n}\mathbb{I}_{\left\{\overset{L}{\hat{y}}_{i,\text{test}} \leq y_{i,\text{test}} \leq \overset{U}{\hat{y}}_{i,\text{test}}\right\}} \tag{16}$$

$$\text{Difference: (Predicted vs. Actual Trouble Spots)} = \sum_{i-1}^{n}\hat{y}_{i,\text{test}} - \sum_{i-1}^{n} y_{i,\text{test}} \tag{17}$$

Where $\hat{y}_{i,\text{test}}$ and $y_{i,\text{test}}$ are the out-of-sample (test set) predicted median and observed response respectively, $\overset{L}{\hat{y}}_{i,\text{test}}$ and $\overset{U}{\hat{y}}_{i,\text{test}}$ are the lower and upper values a $100(1-\alpha)\%$ highest posterior density (HPD) interval for $\hat{y}_{i,\text{test}}$, n_{test} is the test set sample size (in our application $n_{\text{test}} = n_{\text{training}}$), and $\mathbb{I}_{\{\cdot\}}$ is the indicator function taking a value 1 if the condition is true and 0 otherwise.

4.4 Comparison to Tree Based Predictive Models for Hurricane Damages

Our goal for this paper is twofold; to compare the predictive performance of various popular prior distributions for binomial regression models, and also to show that well specified binomial regression models can outperform other common methods used to predict electrical utility asset damages due to hurricanes. To that end, we compare the predictive performance of our models to a Bayesian additive regression trees (BART) (Chipman et al., 2010) model currently used in the University of Connecticut / Eversource Energy Storm Modeling Group's operational system for trouble spot (damages) prediction (see He et al. (2015)). In addition we compare our models to the non-Bayesian Ensemble Forest method of Wanik et al. (2015) also used in the same operational system.

5 Connecticut Electrical Utility Asset Damage Data from Hurricanes Irene and Sandy

The data set we use to test the predictive performance of the Jeffreys's, Gelman, and g-prior consists of electrical utility asset damages, weather simulations, distribution infrastructure, land cover and socioeconomic details for the Connecticut service territory of a major regional electrical provider. We focus specifically on modeling damages produced by two major hurricanes, Irene and Sandy with most attention being given to training a model on the Irene data to predict Sandy given the natural chronology.

Importantly, the geographic region of the utility's service territory is divided into 2x2 km grid cells. Each observation i therefore represents the aggregate data from one of these grid cells and spatial correlation will also be assessed at this level.

5.1 Data Description & Justification for Binomial Regression

The response for our models (denoted $\boldsymbol{y} = \{y_1, ...y_n\}$) is the count of electrical utility asset damages (hereby referred to as trouble spots or TS) per grid cell. In addition to the count of trouble spots, our data also contains the number of total assets (or isolating devices) per grid cell which we denote as $\boldsymbol{N} = \{N_1, ...N_n\}$. Several papers (see (Liu et al., 2008; Han et al., 2009b) for examples) using parametric models to perform hurricane damage prediction to electrical assets, treat this or similar variables as a linear predictor in their models. We believe this is theoretically inconsistent and can lead to spurious conclusions (as in Liu et al. (2008)) that the number of assets in a region is somehow a strong predictor of damage to an asset rather than the environment or the forces acting on it. We would certainly expect close to zero trouble spots per grid cell in the absence of a significant meteorological event regardless of the number of assets present in the area. Also, because of the geographical nature of this variable, it potentially masks any legitimate spatial properties inherent to the damage generating process which could

otherwise be modeled explicitly. With these considerations in mind, we treat these values ($\boldsymbol{N}$) not as a covariate but as a fixed parameter in a binomial regression model which represents the number of opportunities for trouble spot occurrence per grid cell.

Table 1 provides brief summary statistics for the response and assets in each hurricane Irene and Sandy.

	Irene	Sandy
Sample Size: n	2,851	2,851
Total Assets: $\sum_{i=1}^{n} N_i$	354,586	354,586
Total Trouble Spots (TS): $\sum_{i=1}^{n} y_i$	15,033	15,249
Average TS per Grid Cell	5.2729	5.3487
Median TS per Grid Cell	3.0000	2.0000
Variance of TS per Grid Cell	48.4245	83.8138

Table 1: Hurricane Response Variable and Electrical Utility Asset Summary

In addition to the response we have data including 38 covariates containing the results of Weather Research and Forecasting Model (WRF) (Skamarock et al., 2008) weather simulations for the hurricanes, information on power distribution infrastructure, land cover and socioeconomic details. A detailed description of the the variables included in our data are provided in Wanik et al. (2015) and He et al. (2015).

5.2 Orthogonal Rotation & Other Data Preparation

Orthogonal Rotation

Many of the meteorological, infrastructure, and socioeconomic variables in our data are highly correlated as is common with data of this nature (Han et al., 2009b). To avoid problems with multicollinearity we perform an orthogonal rotation of our training and test data matrices (often referred to as principal component analysis (PCA) (Pearson, 1901; Hotelling, 1933)) using the following procedure.

Given the training and test data matrices $\boldsymbol{X}_{tr}^{*} \, \epsilon \, \mathbb{R}^{n\mathrm{x}(p-1)}$ and $\boldsymbol{X}_{te}^{*} \, \epsilon \, \mathbb{R}^{n\mathrm{x}(p-1)}$ (the design matrix without the leading column of ones), we perform mean centering of each column obtaining,

$$\boldsymbol{X}_{\bar{tr}}^{*} = [x_{tr,1} - \bar{x}_{tr,1}|...|x_{tr,p} - \bar{x}_{tr,p}] \text{ and,} \tag{18}$$

$$\boldsymbol{X}_{\bar{te}}^{*} = [x_{te,1} - \bar{x}_{te,1}|...|x_{te,p} - \bar{x}_{te,p}] \tag{19}$$

by singular value decomposition of the empirical covariance matrix associated with the data,

$$\boldsymbol{X}_{\bar{tr}}^{*}{}^{T}\boldsymbol{X}_{\bar{tr}}^{*} = \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{U}^{T}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{W}^{T} = \boldsymbol{W}\boldsymbol{\Sigma}^{2}\boldsymbol{W}^{T} \tag{20}$$

where $\boldsymbol{W}$ is a matrix containing the right singular vectors of $\boldsymbol{X}_{tr}^{*}$, $\boldsymbol{U}$ is a matrix containing the left singular vectors of $\boldsymbol{X}_{tr}^{*}$, and $\boldsymbol{\Sigma}$ is diagonal matrix of singular values $\sigma_{(k)}$. We then obtain the orthogonally transformed training data matrix,

$$\boldsymbol{T}_{tr}^{*} = \boldsymbol{X}_{tr}^{*}\boldsymbol{W}_{s} \tag{21}$$

and under the same linear transformation $\boldsymbol{W}_{s}$ a transformed test data matrix. Where orthogonality is not guaranteed but correlation between column vectors is assumed to be minimal

$$\boldsymbol{T}_{te}^{*} = \boldsymbol{X}_{tr}^{*}\boldsymbol{W}_{s} \tag{22}$$

where $\boldsymbol{W}_{s} = \boldsymbol{W} * [e_1|...|e_s]$ and each $e_{1,...,s}$ is a vector in the standard basis of $\mathbb{R}^{n\mathrm{x}(p-1)}$ allowing us to select the desired number of principal components (column vectors of $\boldsymbol{T}_{tr}^{*}$ or $\boldsymbol{T}_{te}^{*}$ ranked in descending order by their associated singular values) to include in our model. Finally, by adding a leading column of ones to each $\boldsymbol{T}_{tr}^{*}$ and $\boldsymbol{T}_{te}^{*}$ we obtain $\boldsymbol{T}_{tr}$ which will be used as the design matrix to train our regression models, and $\boldsymbol{T}_{te}$ which is our transformed test data.

$$\tag{23}$$

For the models in this study we select a value $s = 6$ such that $\boldsymbol{W}_6 \,\epsilon\, \mathbb{R}^{p=38\text{x}6}$ which provided a good balance of dimension reduction and predictive strength. This procedure has the added benefit of ensuring that our design matrix is of full rank; alleviating any computational issues that may arise with our various prior specifications. Also, since the column vectors of the resultant design matrix are linearly independent (LIN) we also ensure that the i.i.d. assumption of the Gelman prior holds.

Scaling per Gelman et al. (2008)

In addition to the orthogonal rotation, we also scale (already with mean zero) the column vectors of the transformed design matrices to each have a standard deviation of 0.5 per the requirements of the Gelman prior. While unnecessary for our other prior specifications, we use this scaling throughout so our estimated parameters will be comparable between models.

5.3 Spatial Properties

Several other previously mentioned papers in the fields of reliability engineering and risk analysis investigate the potential benefit of modeling the spatial correlation in hurricane damages and power outages. Before fitting spatial models, we briefly explore the spatial properties of our data.

Figure 1 below plots the aggregate trouble spots (on a log scale for visual convenience) for each 2x2 km grid cell in the Eversource Energy CT service territory.

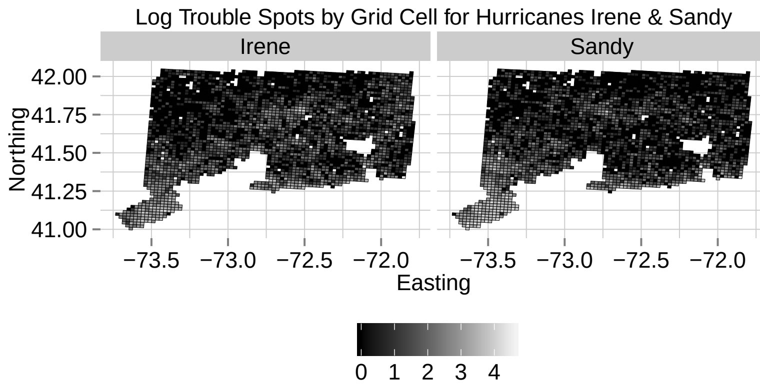

Figure 1: Log Trouble Spots by Grid Cell for Hurricanes Irene & Sandy

Clearly, a similar concentration of trouble spots between the two storms is evident. Both storms have a high concentration of trouble spots in the southwest corner of the state (Fairfield County), the central region of the state (Hartford County), and the shoreline. It does however seem that the damages due to hurricane Irene form stronger clusters in the central and eastern portions of the state than those due to hurricane Sandy. Nonetheless, the spatial distribution of trouble spots appears to be similar enough between the two storms to justify including a spatial component in our predictive models.

In addition to plotting trouble spots over the map, we also calculated a semi-variogram using the modulus estimator outlined in Cressie and Hawkins (1980).

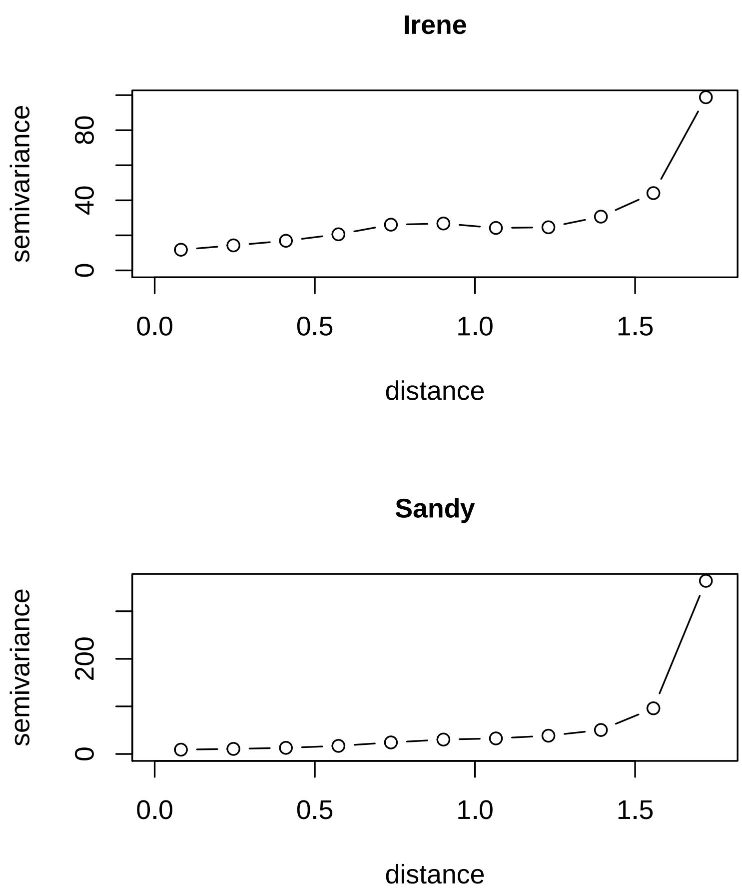

Figure 2: Semivariograms for Hurricanes Irene and Sandy

Both storms have a similar trend with respect to semivariance as distance increases though trouble spots in hurricane Irene appear to have stronger spatial correlation. This finding reinforces the intuition obtained from 1.

6 g-Prior Hyperparameter Elicitation

To select hyperpriors a_π and b_π for the g-prior we solicited information from several knowledgeable (in terms of hurricane damages) employees from the electrical utility company. Because, both hurricane Irene and Sandy occurred several years ago and we have access to no other hurricane data in the same format, we asked the employees to consider their experience and related historical records prior to these hurricanes. The employees then provided informed estimates for the total number of trouble spots they would have expected in a hurricane in this time period. In addition to these estimates, a principal engineer at the same company, provided a single point estimate for the proportion of expected trouble spots (per total assets) in a hurricane and also a measure of potential dispersion around this point. We explore three methods to estimate the hyperpriors a_π and b_π from the aforementioned information.

6.1 Method 1: Using Point Estimates for Expected Trouble Spots from Subject Matter Experts

An appealing approach to estimate g-prior hyperparameters is to treat the elicited trouble spot count estimates as realizations of a sequence of random variables Z_i where each $Z_i \sim \mathcal{B}\text{in}(N, \theta)$ where N is the total number of electrical utility assets in the Eversource Energy CT territories and θ is the true proportion of damaged assets. Assuming a conjugate prior on θ such that $\theta \sim \mathcal{B}e(a_\pi^*, b_\pi^*)$ our posterior distribution is simply,

$$\theta|\boldsymbol{z}, N \sim \mathcal{B}e\left(a_\pi^* + \sum_{i=1}^{k} z_i,\ b_\pi^* + kN - \sum_{i=1}^{k} z_i\right),\ \text{where } \boldsymbol{z} \in \mathbb{R}^k \tag{24}$$

and the hyperparameters selected for the g-prior by this method will be given by,

$$a_{1\pi} = a_\pi^* + \sum_{i=1}^{k} z_i \tag{25}$$

$$b_{1\pi} = b_\pi^* + kN - \sum_{i=1}^{k} z_i \tag{26}$$

We chose the parameters a_π^* and b_π^* of our prior distribution on θ by the following reasoning.

Given that $\hat{\theta}_{\text{MLE}} = \sum_{i=1}^{n} y_i / \sum_{i=1}^{n} N_i$ (where $\sum_{i=1}^{n} y_i$ is the sum of successes per observation and $\sum_{i=1}^{n} N_i$ is the fixed sum of opportunities per observation) is a natural estimator for θ, we can relate the parameters a_π^* and b_π^* of our prior distribution on θ to the MLE estimator for θ since,

$$\mathbb{E}[\theta] = \frac{a_\pi^*}{a_\pi^* + b_\pi^*} \text{ this implies that,} \tag{27}$$

$$a_\pi^* \widehat{=} \sum_{i=1}^{n} y_i \tag{28}$$

$$b_\pi^* \widehat{=} \sum_{i=1}^{n} N_i - y_i \tag{29}$$

and therefore given some reference information we can choose prior parameters for θ.

The largest non-hurricane storm in our historical data (a snow storm) had approximately 25,000 trouble spots. While this storm was not a hurricane, it nonetheless informs what to expect in a major damaging storm in the Eversource Energy territories. Therefore, we set $a_\pi^* = 25,000$ and $b_\pi^* = N - 25,000 = 329,586$.

The vector of trouble spot estimates from the Eversource Energy employees is,

$$\boldsymbol{z} = [2,308 \quad 6,153 \quad 15,000 \quad 22,000 \quad 25,000 \quad 30,500] \tag{30}$$

and therefore our posterior distribution is,

$$\theta | \boldsymbol{z}, N \sim \mathcal{B}e\left(a_{1\pi} = 125,961, b_{1\pi} = 2,356,141\right), \text{ where } \boldsymbol{z} \,\epsilon\, \mathbb{R}^k \tag{31}$$

where,

$$\mathbb{E}[\theta|\boldsymbol{z},N] \approx 0.05075 \tag{32}$$

$$\mathbb{V}\mathrm{ar}[\theta|\boldsymbol{z},N] \approx 1.9408 \cdot 10^{-8} \tag{33}$$

$$\mathbb{SD}[\theta|\boldsymbol{z},N] \approx 0.0001393 \tag{34}$$

and $a_{1\pi}$ and $b_{1\pi}$ are our hyperparameters chosen by this method for the informed g-prior.

6.2 Method 2: Using Point Estimates for Expected Trouble Spot Proportions from Subject Matter Experts

This technique is used in method 1 is appealing due to the use of a conjugate beta prior and its intuitive formulation. However, the assumption that in a given storm all $N = 354,586$ assets are at risk, or at least at equivalent risk, of being damaged is naive. Though the elicited mean may be reasonable, variance is severely underestimated since the model involves what amounts to $k \cdot 354,586$ Bernoulli trials (a very large sample size). Hence, we re-express the problem in terms of proportions. Defining a new sequence of random variables $V_i = {}^{Z_i}/_N$ and assuming then that each $V_i \sim \mathcal{B}e(a_\pi, b_\pi)$ we obtain the vector trouble spot proportion estimates,

$$\boldsymbol{v} = [0.006509 \;\; 0.01735 \;\; 0.04230 \;\; 0.06204 \;\; 0.07050 \;\; 0.08602] \tag{35}$$

and since we have little data suggesting a similarity between trouble spot proportion variance in snow storms vs hurricanes we are not able to construct reasonable informed prior distributions for the g-prior hyperparameters using this method. Therefore, we place noninformative uniform priors on a_π and b_π such that $a_\pi, b_\pi \sim \mathcal{U}(0,\infty)$. We sample from the joint posterior distribution $\pi(a_\pi, b_\pi|\boldsymbol{v})$ with the "No-U-Turn Sampler" (NUTS) sampler available in Stan (Stan Development Team, 2014) and obtain the

estimated posterior median values for a_π and b_π which we denote by $\widehat{a_\pi}_M = 2.29$ and $\widehat{b_\pi}_M = 48.81$. We use these estimates as our g-prior hyperparameters and assume,

$$\theta \sim \mathcal{B}e(\widehat{a_\pi}_M = 2.29, \widehat{b_\pi}_M = 48.81) \text{ which implies} \tag{36}$$

$$\mathbb{E}[\theta] \approx 0.04481 \tag{37}$$

$$\mathbb{V}\text{ar}[\theta] \approx 0.0008216 \tag{38}$$

$$\mathbb{SD}[\theta] \approx 0.02866 \tag{39}$$

Though slightly less convenient mathematically, we feel this latter method produces a much more sensible variance estimate for the proportion of trouble spots due to hurricanes. The values $a_{2\pi} = \widehat{a_\pi}_M = 2.29$ and $b_{2\pi} = \widehat{b_\pi}_M = 48.81$ are ultimately the g-prior hyperparameters we use to construct our predictive models from this method of prior elicitation.

6.3 Method 3: Choosing Hyperparameters by Moment Matching

Prior to our consultation with a larger group of employees, a principal engineer at the electrical utility company supplied us with an estimate of the expected proportion of trouble spots in a typical hurricane as well as an estimate of uncertainty around this proportion. Again, treating θ as the true proportion of damaged electrical utility assets, we were given,

$$\hat{\theta} = 0.05 \pm 0.02 \tag{40}$$

The distribution of the proportion of trouble spots for historical storms prior to Irene and Sandy appears to be unimodal and heavily right skewed. In settings such as this where small deviations from the mean predominate, research suggests that subject

matter experts tend to underestimate variance estimates in the prior elicitation process (Garthwaite et al., 2005). Additionally, the assumption that hurricanes produce highly variable damages is reinforced by author's tendencies to select negative binomial GLMs for similar problems (see (Liu et al., 2008; Han et al., 2009a,b)). With these considerations in mind, rather than treat the interval estimate in 40 as a 95% credible set, we treat 0.02 as the estimated standard deviation (denoted s) and we obtain the estimated variance $s^2 = 0.0004$, and letting $\bar{x} = 0.05$ be the estimated mean proportion of damaged assets, then by equating sample moments to the mean and variance of the standard beta distribution,

$$a_{3\pi} = \bar{x}\left(\frac{\bar{x}(1-\bar{x})}{s^2} - 1\right) \tag{41}$$

$$= \frac{471}{80} = 5.8875 \tag{42}$$

$$b_{3\pi} = (1-\bar{x})\left(\frac{\bar{x}(1-\bar{x})}{s^2} - 1\right) \tag{43}$$

$$= \frac{8949}{80} = 111.8625 \tag{44}$$

and these values $a_{3\pi}$ and $b_{3\pi}$ are the g-prior hyper-parameters chosen by method 3.

7 Results

Table 2 contains the results of models both fit to Irene to predict trouble spots in Sandy, and fit to Sandy to predict trouble spots in Irene. All point estimates for predicted trouble spots are obtained from the posterior median of the predictive distribution and the 95% highest posterior density interval (HPDI) was used to estimate coverage and the upper and lower bounds on total predictions.Of primary importance is the model fit to Irene to predict Sandy. This sequence of storms has the correct chronology and also provides some insight on how to model a future Sandy-like storm; a topic of importance due to Sandy's high damage costs. The first six rows of each sub-table list the results of

the binomial regression model's under the various prior distributions. The model named "CAR" in both sub-tables, contains the results of the conditional autoregressive BYM model fit to the residuals of the g-prior model under the default hyperparameters. The last two rows in each sub-table contain the results from the BART and Forest Ensemble models. Unfortunately, due to the nature of these models, there are no WAIC or IS-LOO statistics available. Additionally, the Forest Ensemble model does not produce measures of uncertainty around its point estimates, so no coverage percentage or lower and upper bounds on total estimated trouble spots are available for this model.

Model Fit to Irene to Predict Sandy									
Model	WAIC	IS-LOO	RMSE	Coverage%	$\sum_{i=1}^{n} \hat{y}^{L}_{i,\text{Sandy}}$	$\sum_{i=1}^{n} \hat{y}_{i,\text{Sandy}}$	$\sum_{i=1}^{n} \hat{y}^{U}_{i,\text{Sandy}}$	$\sum_{i=1}^{n} y_{i,\text{Sandy}}$	Difference
Jeffreys's Prior	15746.99	15747.02	6.3295	0.7938	5633	15404.5	26264		156
Gelman Prior	15747.04	15747.05	6.3194	0.7899	5590	15414	26242		165
$g-$Prior Default	15746.59	15746.53	6.3186	0.7906	5610	15397	26261		148
$g-$Prior Method 1	17386.34	17386.34	7.117	0.7759	6494	17464.5	29165		2216
$g-$Prior Method 2	15747.61	15747.64	6.3284	0.7934	5580	15408	26253	15249	159
$g-$Prior Method 3	15747.32	15747.41	6.3245	0.7906	5578	15412	26242		163
CAR	12124.43	12429.94	5.3085	0.9165	4379	15653.5	29518		405
BART			6.0383	0.9144	14208	14596.9	14939		-652
Forest Ensemble			7.1738			19651.5			4403

Model Fit to Sandy to Predict Irene									
Model	WAIC	IS-LOO	RMSE	Coverage%	$\sum_{i=1}^{n} \hat{y}^{L}_{i,\text{Irene}}$	$\sum_{i=1}^{n} \hat{y}_{i,\text{Irene}}$	$\sum_{i=1}^{n} \hat{y}^{U}_{i,\text{Irene}}$	$\sum_{i=1}^{n} y_{i,\text{Irene}}$	Difference
Jeffreys's Prior	16756.56	16755.26	5.6261	0.8141	5342	14398.5	24757		-634.5
Gelman Prior	16755.77	16755.86	5.4064	0.8078	5308	14383.5	24738		-649.5
$g-$Prior Default	16754.09	16753.97	5.4089	0.8074	5350	14388.5	24764		-644.5
$g-$Prior Method 1	20584.67	20584.67	5.2734	0.8183	6522	17510.5	29211		2477.5
$g-$Prior Method 2	16755.26	16755.1	5.4085	0.8085	5313	14376.5	24748	15033	-656.5
$g-$Prior Method 3	16754.56	16754.94	5.4107	0.8071	5344	14393.5	24776		-639.5
CAR	11722.52	12074.24	5.0842	0.8846	4464	14781	28176		-252
BART			5.7598	0.8930	14650	15298	15399.75		265
Forest Ensemble			5.4948			14294.2			-738.8

Table 2: Model Evaluation Criteria

When fit to both data sets, the models under the default parameterization of the g-prior had a small performance edge over the other non-spatial binomial regression models when considering most statistics. They outperform other binomial models consistently in terms of IS-LOO and WAIC but are bested in terms of RMSE, coverage, and total estimated trouble spots by other models on occasion when fit to Sandy to predict Irene. Because they have superior IS-LOO and WAIC statistics, and tend to predict well when using Irene to predict Sandy, the residuals from these models were used as the response for the CAR models.

The three other g-prior models with hyperparameters derived through expert opin-

ion performed very similarly to the reference prior models with the exception of the model produced using method 1. Method 1 generated a prior distribution on the mean proportion of trouble spots that severely underestimated variance. As a consequence, these models forced all predictions towards the prior mean which resulted in extreme over-prediction of total trouble spots since the observed distribution of trouble spots is right skewed.

The two statistical learning models (BART and Ensemble Forest) did not generally perform as well as the best binomial regression models with and without spatial random effects. The BART model tends to be too aggressive with its error estimates (credible intervals) and the Forest Ensemble model produced the worst estimates of total trouble spots. These models are currently used in the University of Connecticut / Eversource Energy Storm Modeling Group's operational system for trouble spot prediction and perform well when trained to larger training sets of diverse storms. However, a purpose-made parametric model (similar to the examples in this paper) is likely be a better choice for modeling infrequent and extreme events such as hurricanes.

Finally, The best performing models overall, leading in terms of IS-LOO, WAIC, RMSE, and producing consistently accurate total trouble spot estimates were the CAR models fit to the residuals of the default g-prior models. The spatial random effect serves as a proxy for important (for predicting trouble spots) geo-spatial features missing from, or poorly explained by our data. Standard generalized linear models and statistical learning models can not capture this information. The spatial distribution of observed and the CAR model's predicted trouble spots is illustrated in figure 3.

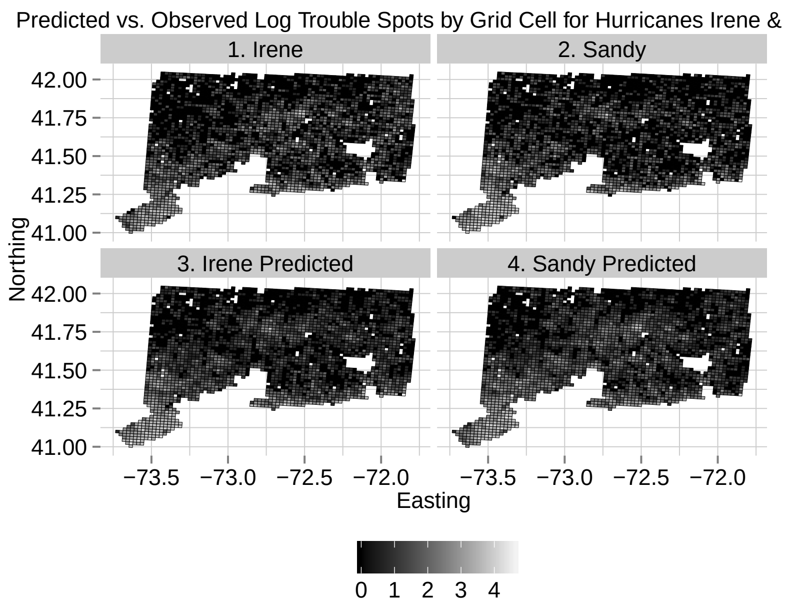

Figure 3: Predicted vs. Observed Log Trouble Spots by Grid Cell for Hurricanes Irene & Sandy

8 Conclusions

In this application, with carefully selected hyper-parameters, binomial regression models using the informed g-prior have similar predictive accuracy as models using common reference priors. The inclusion of prior information from subject matter experts and offers no distinct advantage. However, careless selection of hyper-parameters can lead to substantial reduction in predictive accuracy. Surprisingly, the noninformative specification of the g-prior performed marginally better than all other binomial regression

models tested in this paper; contradicting one of the findings in Hanson et al. (2014). In addition, fitting the BYM CAR model to the residuals of the default g-prior model improved upon predictive accuracy; beyond the performance of some statistical learning models growing in popularity in the field of reliability engineering.

Appendix A: Parameter Estimates

Model Fit to Irene to Predict Sandy																		
Model	Jeffreys's Prior			Gelman Prior			g-Prior Noninformative			g-Prior Method 1			g-Prior Method 2			g-Prior Method 3		
	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$
β_0	-3.1726	-3.1552	-3.1337	-3.1732	-3.1547	-3.1344	-3.1748	-3.1546	-3.1366	-2.9344	-2.9322	-2.9302	-3.1737	-3.1552	-3.1352	-3.1733	-3.1542	-3.1365
β_1	-0.2279	-0.2060	-0.1831	-0.2260	-0.2058	-0.1826	-0.2267	-0.2056	-0.1824	-0.0108	-0.0064	-0.0024	-0.2294	-0.2057	-0.1852	-0.2276	-0.2051	-0.1852
β_2	0.0549	0.0780	0.1050	0.0540	0.0782	0.1045	0.0517	0.0782	0.1043	0.0024	0.0066	0.0106	0.0520	0.0779	0.1043	0.0527	0.0779	0.1035
β_3	-0.1015	-0.0646	-0.0319	-0.0991	-0.0654	-0.0294	-0.1011	-0.0646	-0.0313	-0.0036	0.0006	0.0049	-0.1044	-0.0657	-0.0314	-0.0986	-0.0640	-0.0277
β_4	-0.2380	-0.2042	-0.1711	-0.2350	-0.2029	-0.1680	-0.2358	-0.2035	-0.1697	-0.0073	-0.0031	0.0013	-0.2355	-0.2027	-0.1694	-0.2355	-0.2026	-0.1708
β_5	0.0892	0.1160	0.1418	0.0894	0.1153	0.1414	0.0911	0.1156	0.1428	-0.0009	0.0032	0.0071	0.0870	0.1153	0.1406	0.0902	0.1154	0.1420
β_6	-0.4548	-0.4208	-0.3859	-0.4541	-0.4203	-0.3843	-0.4574	-0.4198	-0.3858	-0.0111	-0.0069	-0.0027	-0.4547	-0.4201	-0.3870	-0.4538	-0.4186	-0.3845
CAR							σ^2_{lower}	$\tilde{\sigma}^2$	σ^2_{upper}									
σ^2_τ							0.1230	0.2058	0.3064									
σ^2_ϕ							0.5969	0.7500	0.9403									

Model Fit to Sandy to Predict Irene																		
Model	Jeffreys's Prior			Gelman Prior			g-Prior Noninformative			g-Prior Method 1			g-Prior Method 2			g-Prior Method 3		
	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$	$\beta_{p,\text{lower}}$	$\tilde{\beta}_p$	$\beta_{p,\text{upper}}$
β_0	-3.3090	-3.2882	-3.2680	-3.3091	-3.2887	-3.2671	-3.3089	-3.2883	-3.2677	-2.9341	-2.9320	-2.9300	-3.3093	-3.2884	-3.2679	-3.3075	-3.2876	-3.2674
β_1	-0.5913	-0.5716	-0.5517	-0.5921	-0.5718	-0.5512	-0.5914	-0.5715	-0.5518	-0.0254	-0.0209	-0.0166	-0.5908	-0.5711	-0.5512	-0.5895	-0.5704	-0.5488
β_2	-0.2128	-0.1892	-0.1654	-0.2118	-0.1891	-0.1642	-0.2121	-0.1894	-0.1642	-0.0063	-0.0021	0.0021	-0.2128	-0.1891	-0.1649	-0.2124	-0.1890	-0.1649
β_3	0.1202	0.1516	0.1836	0.1217	0.1523	0.1818	0.1214	0.1512	0.1824	0.0021	0.0064	0.0103	0.1191	0.1519	0.1810	0.1183	0.1521	0.1801
β_4	-0.3062	-0.2757	-0.2442	-0.3076	-0.2763	-0.2464	-0.3060	-0.2761	-0.2457	-0.0075	-0.0032	0.0008	-0.3066	-0.2752	-0.2444	-0.3054	-0.2746	-0.2451
β_5	0.0189	0.0450	0.0693	0.0204	0.0453	0.0706	0.0211	0.0449	0.0714	-0.0027	0.0018	0.0056	0.0196	0.0455	0.0716	0.0198	0.0453	0.0715
β_6	-0.2817	-0.2457	-0.2118	-0.2823	-0.2456	-0.2112	-0.2825	-0.2454	-0.2107	-0.0064	-0.0020	0.0023	-0.2820	-0.2453	-0.2094	-0.2817	-0.2446	-0.2094
CAR							σ^2_{lower}	$\tilde{\sigma}^2$	σ^2_{upper}									
σ^2_τ							0.2192	0.2858	0.3660									
σ^2_ϕ							0.6414	0.8081	1.0371									

Table 3: Parameter Estimates: Posterior Median and 95% HPD Intervals

References

Bedrick, E. J., Christensen, R., and Johnson, W. (1996). "A new perspective on priors for generalized linear models." *Journal of the American Statistical Association*, 91(436): 1450–1460. 7

Besag, J., York, J., and Mollié, A. (1991). "Bayesian image restoration, with two applications in spatial statistics." *Annals of the institute of statistical mathematics*, 43(1): 1–20. 8

Best, N., Richardson, S., and Thomson, A. (2005). "A comparison of Bayesian spatial models for disease mapping." *Statistical methods in medical research*, 14(1): 35–59. 8

Chen, M.-H., Ibrahim, J. G., and Kim, S. (2008). "Properties and implementation of Jeffreys's prior in binomial regression models." *Journal of the American Statistical Association*, 103(484): 1659–1664. 5

Chen, M.-H., Ibrahim, J. G., Shao, Q.-M., and Weiss, R. E. (2003). "Prior elicitation for model selection and estimation in generalized linear mixed models." *Journal of Statistical Planning and Inference*, 111(1): 57–76. 1

Chen, M.-H., Ibrahim, J. G., and Yiannoutsos, C. (1999). "Prior elicitation, variable selection and Bayesian computation for logistic regression models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1): 223–242. 1

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). "BART: Bayesian additive regression trees." *The Annals of Applied Statistics*, 266–298. 11

Cressie, N. and Hawkins, D. M. (1980). "Robust estimation of the variogram: I." *Journal of the International Association for Mathematical Geology*, 12(2): 115–125. 16

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). "Statistical methods for eliciting probability distributions." *Journal of the American Statistical Association*, 100(470): 680–701. 1, 22

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). "A weakly informative default prior distribution for logistic and other regression models." *The Annals of Applied Statistics*, 1360–1383. 2, 6, 15

Han, S.-R., Guikema, S. D., and Quiring, S. M. (2009a). "Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models." *Risk analysis*, 29(10): 1443–1453. 22

Han, S.-R., Guikema, S. D., Quiring, S. M., Lee, K.-H., Rosowsky, D., and Davidson, R. A. (2009b). "Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region." *Reliability Engineering & System Safety*, 94(2): 199–210. 12, 13, 22

Hanson, T. E., Branscum, A. J., Johnson, W. O., et al. (2014). "Informative g-Priors for Logistic Regression." *Bayesian Analysis*, 9(3): 597–612. 2, 3, 6, 7, 26

He, J., Wanik, D., Hartman, B., Anagnostou, E., and Astitha, M. (2015). "Nonparametric Tree-based Predictive Modeling of Storm Damage to Power Distribution Network." *Pending Publication*. 11, 13

Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components." *Journal of educational psychology*, 24(6): 417. 13

Ibrahim, J. G. and Chen, M.-H. (2000). "Power prior distributions for regression models." *Statistical Science*, 46–60. 1

Ibrahim, J. G. and Laud, P. W. (1991). "On Bayesian analysis of generalized linear models using Jeffreys's prior." *Journal of the American Statistical Association*, 86(416): 981–986. 5

Jeffreys, H. (1946). "An invariant form for the prior probability in estimation problems." *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007): 453–461. 4

Liu, H., Davidson, R. A., and Apanasovich, T. V. (2008). "Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms." *Reliability Engineering & System Safety*, 93(6): 897–912. 7, 12, 22

Pearson, K. (1901). "LIII. On lines and planes of closest fit to systems of points in space." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11): 559–572. 13

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., and Barker, D. (2008). "A description of the Advanced Research WRF version 3. NCAR Tech." Technical report, Note NCAR/TN-4751STR. 13

Stan Development Team (2014). "Stan: A C++ Library for Probability and Sampling, Version 2.5.0."

URL http://mc-stan.org/ 20

Vehtari, A. and Gelman, A. (2014). "WAIC and cross-validation in Stan." 10

Vehtari, A., Tolvanen, V., Mononen, T., and Winther, O. (2014). "Bayesian leave-one-out cross-validation approximations for Gaussian latent variable models." *arXiv preprint arXiv:1412.7461*. 10

Wanik, D. W., Parent, J. R., Anagnostou, E. N., and HaHartman, B. M. (2015). "Using Vegetation Management, LiDAR and Infrastructure Data in Modeling Weather-Induced Damage to Electric Distribution Networks." *Under Revision*. 11, 13

Watanabe, S. (2010). "Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory." *The Journal of Machine Learning Research*, 11: 3571–3594. 10

Zellner, A. (1983). "Applications of Bayesian analysis in econometrics." *The Statistician*, 23–34. 2, 6

Zorn, C. (2005). "A solution to separation in binary response models." *Political Analysis*, 13(2): 157–170. 5