Measuring infrastructure and community recovery rate using Bayesian methods: a case study of power systems resilience

H. Baroud & S. Murlidar
Vanderbilt University, Nashville, USA

ABSTRACT: With the increasing frequency and severity of disasters resulting especially from natural hazards and impacting both infrastructure systems and communities, thus challenging their timely recovery, there is a strong need to prepare for more effective response and recovery. Communities have especially struggled to understand the aspects of recovery patterns for different systems and prepare accordingly. Therefore, it is essential to develop models that are able to measure and estimate the recovery trajectory for a certain community or infrastructure network given system characteristics and event information. The objective of the study is to deploy the Poisson Bayesian kernel model developed and tested in earlier work in risk analysis to measure the recovery rate of a system. In this paper, the model is implemented and tested on a resilience modeling case study of power systems. The model is validated using a comparison to other count data models such as Poisson generalized linear model and the negative binomial generalized linear model.

1 INTRODUCTION

Recent disasters severely impacting both infrastructure systems and communities emphasize the need to prepare for more effective response and recovery. Communities have especially struggled in understanding the aspects of recovery patterns for different systems. Therefore, there is a strong need to develop models that are able to measure and estimate what are the recovery prospects for a certain community or infrastructure network given system characteristics and event information. In addition, the models need to account for uncertainty underlying the information that has been or being gathered before, during, and after the disruption.

Prior work on recovery rate modeling of infrastructure systems focuses on the time to recovery from power outages as a function of event attributes and impact of the disaster (Mackenzie & Barker 2013, Barker & Baroud 2014, Barabadi & Ayele 2018). In this research, the goal is to incorporate the uncertainty in estimating the resilience of systems after disruption. More specifically, the objective of the study is to analyze the recovery rate of a system or a community that has been impacted by a disaster. The response variable considered in this work is the average recovery rate computed based on the impact of the event and the total time to network recovery as well as other variables.

In order to integrate information from experts with data on the disruptive event and recovery process, this work proposes the use of a Poisson Bayesian kernel model which accommodates count data while accounting for prior information and uncertainty in the estimates. The model has been developed and tested using sample data in earlier work (Floyd et al. 2014) and has been applied to a risk analysis case study to predict the frequency of disruptive events in inland waterway (Baroud et al. 2013). However, the method has never been implemented in post-disaster scenarios, more specifically to model recovery rate. In this paper, the model is implemented and tested on a resilience modeling case study of power systems. More specifically, the recovery rate of a community from power outages is represented by a parameter following a Gamma distribution. This prior distribution is updated using historical data of disruptive events as well as a set of attributes that are represented by the kernel function, a measure of similarity between the new data point and the training set. The model performance is evaluated in comparison to other count data models such as the Poisson generalized linear model and the negative binomial generalized linear model.

Section 2 provides background literature on community resilience modeling and count data methods with an outline of the paper's contributions. Section 3 briefly describes the Poisson Bayesian kernel method and provides a structure to the model comparison and performance measures. Section 4 describes the case study with an overview of the data and a summary of the results of the models used in this work. Finally, concluding remarks are provided in section 5.

2 BACKGROUND AND CONTRIBUTION

2.1 *Community Resilience Modeling*

The ultimate goal of recovery measures after a disaster is to insure the society is able to bounce back from the losses incurred and reach normalcy as fast as possible, in recent studies this has been termed as "community resilience." One common definition for community resilience refers to the ability for a social system to respond and recover from a disaster. While vulnerability was previously used as an indicator, researchers and government policy have realized the advantages of utilizing resilience as an indicator to measure the ability of a community to not only recover during the post-disaster phase, but also advance beyond the pre-disaster state and adapt or transform to improve preparedness to future events. Furthermore, resilient communities are also less vulnerable to hazards than an equivalent less resilient community. Initially, community resilience modeling research focused on qualitative approaches founded in a set of metrics and indicators that describe the resilience of a community (Johansen et al. 2016). The concept of resilience can be useful when quantified and used as a decision-making tool, however, this can be challenging due to the uncertainty in many factors impacting resilience as well as the lack of data in recovery measures. As such, a number of research initiatives have focused on quantifying resilience ranging from stochastic modeling to simulation and data-driven approaches, among others.

Models of community resilience often include a variety of social factors. In one study, community resilience was modeled as categorical variables based on four primary sets of adaptive capacities-Economic Development, Social Capital, Information and Communication, and Community competence (Norris et al. 2008). It is proposed in this work that advancements within each category will aim to create a community that is more resilient to disasters as a whole. More specifically, one example of the hypothesis proposed in Norris et al. (2008) is the ability to measure infrastructure and economic resilience in terms of power restoration time which can therefore be used as a proxy to understand community resilience.

A more robust model for community resilience uses a composite index of social and geographical factors, the Baseline Resilience Indicators for Communities (BRIC) (Cutter et al. 2014). This relative value measure of resilience can point to counties and tracts within a specific geographic location that are particularly vulnerable to disasters and require more attention and more time to fully recover. This measure was found to have significant negative correlation with the previously established Social Vulnerability Index (SoVI).

Analysis has been performed to identify recovery rate specifically following a disaster. However, two relevant primary issues are dealing with missing data as well as homogeneity and heterogeneity across the data set and the fact that some models are so specific that they need to be adapted for different situations. In addition, most studies have aimed to provide restoration curves that give information on the number of customers with service over time. A lack of literature exists to model recovery rate specifically. One study focuses on the need to not only develop recovery rate plots but to be able to select the appropriate models based on the characteristics of a specific data set (Barabadi & Ayele 2018).

2.2 *Methods for Modeling Count Data*

Modeling the recovery rate requires methods that can accommodate count data as the response variables in this case constitutes the number of recovered subjects per unit of time.

Generalized Linear Models (GLM) are widely used within regression models when count data is present. Within this class of models, the Poisson density function is often used with a log-link function, if the variance of the counts is higher than the mean of the counts, it is common to also use a negative binomial GLM. In certain special cases, extensions of these models can accommodate specific situations. For example, zero-truncated models and zero-inflated models can be used when there are excess zero counts (Shankar & Mannering, 1997), and both use an underlying Poisson distribution.

However, both Poisson and negative binomial lack the flexibility to handle data that is, for example, both underdispersed and overdispersed. As such, other models have been developed. One example is the Conway-Maxwell Poisson (COM) distribution GLM (Guikema & Goffelt, 2008). The model functions by having underdispersed data yield a Bernoulli distribution, overdispersed data yield a geometric distribution, and a Poisson distribution when the variance is equal to the mean.

Using a Bayesian framework to account for the uncertainty in the regression parameters, it is possible to improve on their accurate estimation by updating the parameter distributions with new data. Other approaches of analyzing count data using a Bayesian framework are conjugate priors. These methods are quite attractive as they offer the benefit of uncertainty modeling using Bayesian techniques without adding any computational cost. Given a specific prior distribution and a specific likelihood function, the posterior distribution will have the same form as the prior distribution but with updated posterior parameters. There are different forms of conjugate priors, one of which is the Gamma conjugate prior used to model count data in the model presented in this paper. The method assumes that the rate of occurrence follows a Gamma prior and updates the distribution using information represented by a Poisson likelihood. The Gamma conjugate prior is the foundation of the Poisson Bayesian kernel model used in this paper and will be further discussed in the following section. This method allows the user to model and understand the uncertainty around each variable and estimate them by considering their probability distributions as opposed to point estimate

2.3 *Contributions*

This paper presents new analysis for data-driven community resilience modeling. A Bayesian approach developed and tested in prior work is implemented and tested in a case study of community recovery from power outages. The work presented here constitutes a first step in advancing data-driven methods for applications in infrastructure and community resilience.

3 METHODOLOGY

3.1 *Poisson Bayesian Kernel Model*

Poisson Bayesian kernel methods estimate the rate of occurrence of the event rather than estimating a deterministic value for the number of times the event is estimated to occur. A common distribution to model count data within a Bayesian framework is the Gamma-Poisson conjugate prior. The development of the Poisson Bayesian kernel method discussed can be found in Baroud et al. (2013) and Floyd et al. (2014). The approach uses the Gamma conjugate prior as the basis of the model.

It is assumed that the parameter to be estimated is the rate of occurrence, $\lambda > 0$, which follows a Gamma prior distribution with parameters $\alpha > 0$ and $\beta > 0$, as shown in Eq. (1).

$$P(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} \mathrm{e}^{(-\beta\lambda)} \quad (1)$$

For the likelihood function, the product of the Poisson density function, shown in Eq. (2), is used, since this is a Gamma-Poisson conjugate prior approach.

$$L = \prod_{i=1}^{m} P(y_i) = \prod_{i=1}^{m} \frac{\left(\lambda_i^{y_i} e^{-\lambda_i}\right)}{y_i!} = \frac{\lambda_i^{\sum_{i=1}^{m} y_i} e^{-m\lambda_i}}{\prod_{i=1}^{m} y_i!} \quad (2)$$

Thus, the posterior distribution is the product of Eqs. (1) and (2). Rearranging the product of the likelihood function and the prior distribution function results in a Gamma posterior distribution where $\alpha^* = \sum_{i=1}^{m} x_i + \alpha$ and $\beta^* = m + \beta$.

$$\begin{aligned} P(\lambda|x) &= \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}\right) \left(\lambda^{\sum_{i=1}^{m} y_i} e^{-m\lambda}\right) \\ &= \frac{\lambda^{\left(\sum_{i=1}^{m} y_i + \alpha - 1\right)}\, e^{-\lambda(m+\beta)} (n+\beta)^{\sum_{i=1}^{m} y_i + \alpha}}{\Gamma\left(\sum_{i=1}^{m} y_i + \alpha\right)} \\ &= \text{Gamma}\left(\alpha^*,\, \beta^*\right) \end{aligned} \quad (3)$$

This result is the basic Gamma conjugate prior approach used in Bayesian analysis. This approach assumes the notion of exchangeability meaning that for different sets of training and testing data, the resulting posterior parameters will be similar since they are a function of the prior parameter, the size of the dataset, and the summation of all the data points. The characteristics of each outcome are not taken into consideration in this case, but rather the overall property of the dataset (MacKenzie et al., 2014).

The Poisson Bayesian kernel approach extends the notion of the conjugate prior such that the posterior parameters computation not only depends on the prior parameters and the historical data but also on the attributes through the kernel matrix. The parameters for the Bayesian kernel model for counts are expressed in Eqs. (4) and (5). **K** is the $m \times m$ kernel matrix, **Y** is an $m \times 1$ vector containing the output data associated with the m observations of **X**, and **V** is an $m \times 1$ vector containing ones. Each entry in the kernel matrix represents the similarity measure between the attributes of the testing set and the training set, respectively. As such, the new data point is compared with the training set and according to the similarities of the attributes, new values for the parameter of the posterior distribution are computed. Note that in this case, the training and testing sets are assumed to have the same size, m. However, when the model is deployed, the sets can be of different sizes, and in some cases, the testing set could include only one data point such as in a leave-one-out analysis.

$$\alpha^* = \mathbf{KY} + \alpha \quad (4)$$

$$\beta^* = \mathbf{KV} + \beta \quad (5)$$

As with other statistical and mathematical models, there are a few assumptions underlying the deployment of such modeling approach. Even though the form of the prior distribution is known from the conjugate prior, the model user would still need to identify the values of the prior parameters. While there are formal ways to determine the prior parameters (Kass & Wasserman, 1996), the selection of such parameters might not always be considered (Montesano & Lopes, 2009; Mason & Lopes, 2011). Oftentimes, the priors are either assumed to be known or are assigned such that the prior distribution is non informative. In other cases, these parameters are estimated using data and prior knowledge by matching the sample mean and variance to those of the prior distribution (MacKenzie et al. 2014; Carlin & Louis, 2008). Another assumption to consider is the choice of the kernel function which depends on the application and the model user. This research uses the most popular kernel function, the radial basis function (RBF) in Eq. (6), where $k(\mathbf{x}_i, \mathbf{x}_j)$ is one entry in the matrix $\mathbf{K}$ representing the kernel function between the attributes of the i^{th} and j^{th} data points.

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (6)$$

In addition to being commonly used in kernel methods, RBF has nice properties. The function has only one parameter, σ, to be tuned to an optimal value. This reduces computation efforts significantly in comparison to other kernel functions with two or more parameters requiring a grid search to estimate them. Also, the structure of the function is based on the Euclidean distance, whereby similar data points are closer to each other in the feature space. Finally, the kernel matrix of the RBF has full rank and the entries fall between zero and one resulting in kernel functions of the data points acting as weights in the computation of the posterior parameters (Schölkopf & Smola, 2002). More discussion on the impact of the RBF parameter, σ, on the performance of the model will follow in the case study presented in section 4.

The estimated rate for the new data point follows then a Gamma distribution with parameters α^* and β^*. As a point estimate for this parameter, the expected value of the posterior distribution is considered, shown in Eq. (7) as the ratio of the Gamma distribution parameters α^* and β^*.

$$\hat{\lambda} = \frac{\alpha^*}{\beta^*} \quad (7)$$

Note that a different point estimate for the rate can be used such as the median, the mode, or the variance, depending on the type of problem and the model users.

3.2 *Predictive Accuracy Measures*

The ultimate objective of developing and identifying predictive models is their application in risk and resilience analysis problems, such as predicting the frequency of disruptions in a particular network system or the recovery rate of infrastructure and communities. While the goodness of fit is important to assess whether the model is capturing the pattern and variability in the data, is it equally important to analyze the prediction power of a statistical model if it is going to be used for forecasting purposes. Prediction accuracy is assessed by the out-of-sample error, which accounts for the discrepancy between the estimated parameter and the actual observation of data points that were not in the set used to train the model. In order to validate the prediction power of the models, several metrics are evaluated to assess the out-of-sample error, and they are summarized in Table 1.

Table I: Prediction Accuracy Metrics Formulae

Prediction accuracy metrics	Formula
Root Mean Square Error (RMSE)	$\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}_i\right)^2}$
Normalized Root Mean Square Error (NRMSEM & NRMSED)	$\frac{\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}_i\right)^2}}{sd(Y_i)}$ $\frac{\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}_i\right)^2}}{Y_{maximum} - Y_{minimum}}$
Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - \hat{\lambda}_i\right\|$

While RMSE and MAE are the most commonly used measurements of error, the normalized RMSE is also considered to account for the variability across different samples of training sets generated by the multi-iteration validation process. NRMSE can either be normalized based on the standard deviation of the observed values, $sd(Y_i)$, or the range of values in the testing set, $Y_{maximum} - Y_{minimum}$, and both cases are considered in this paper.

3.3 *Comparative Analysis*

In order to assess the performance of the models, the predictive accuracy measures are used to evaluate the models. More specifically, Poisson Bayesian kernel model is compared to a Poisson generalized linear model and a negative binomial generalized linear model (Cameron & Trivedi, 1986, 2013).

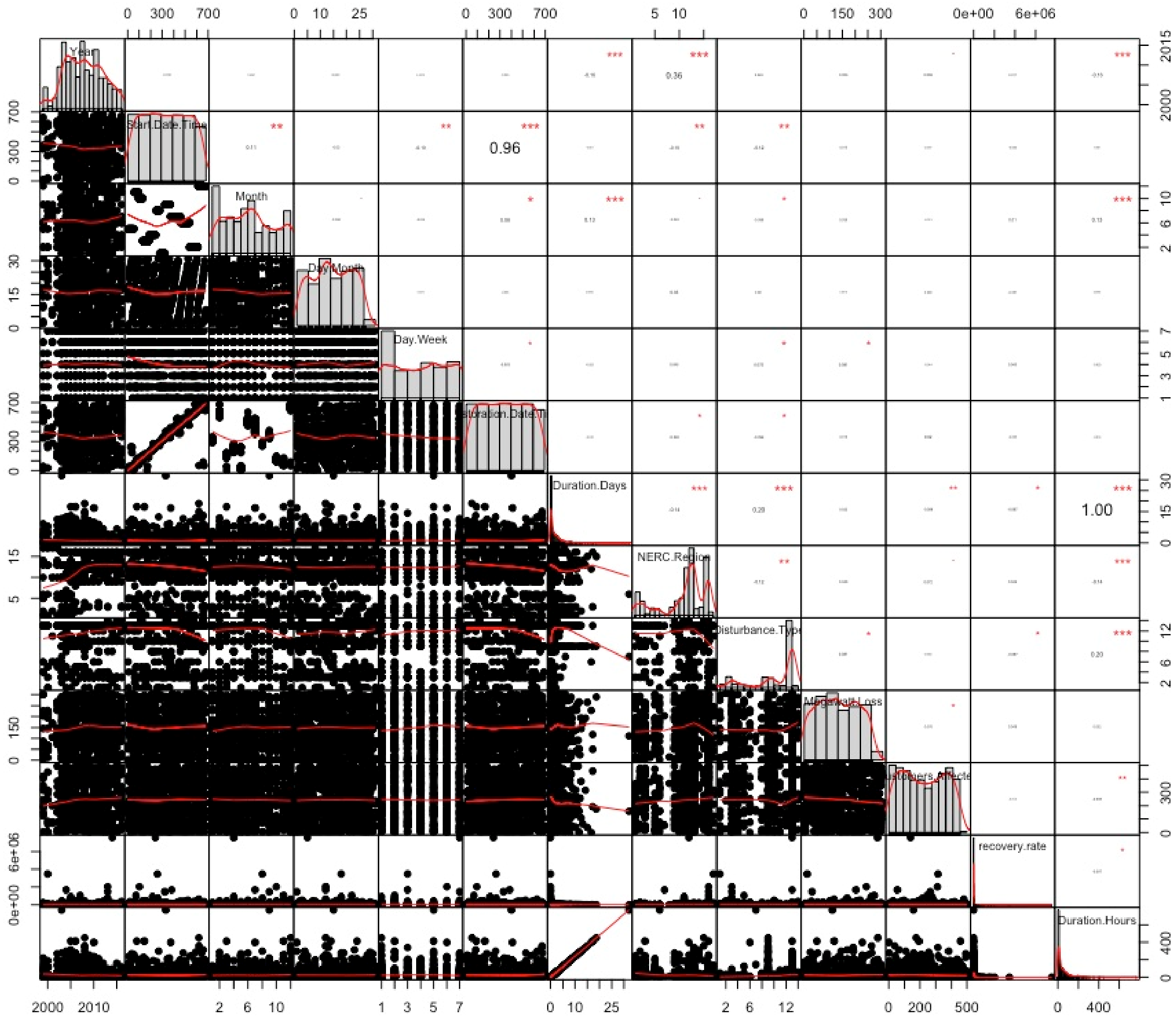
Figure 1. Pairwise Scatter Plots for all the Variables in the data

The Poisson GLM assumes that the rate to be estimated has an exponential relationship with a set of covariates representing coefficients for the different attributes, $\hat{\lambda}_{PGLM} = e^{\beta_i X}$, while the predicted rate for the PBK is equal to the expected value of the posterior probability distribution, $\hat{\lambda}_{PBK} = \frac{\mathbf{KY}+\alpha}{\mathbf{KV}+\beta}$.

4 CASE STUDY

A case study is presented in this paper to demonstrate the use of the Poisson Bayesian kernel model in assessing the resilience of communities. More specifically, the study is focused on major power outage events that happened in the US between 1999 and 2016. The goal is to compare the performance of the model against classical methods and assess its ability to predict, with a high level of accuracy, the recovery rate after these major events.

The ability to accurately measure and predict the recovery rate from power outages allows responders and recovery crews to improve their strategies and resource allocations before, during, and after a disruption.

4.1 *Data*

The data used in the case study is collected from the Energy Information Administration and includes information on the time, date and length of an outage occurred, the magnitude of the power outage (Megawatt Loss & Customers Affected) and the disturbance type (severe weather, equipment failure, among others). The dependent variable to be modeled is recovery rate which is the number of customers affected divided by the duration of outage. To model the rate using a Poisson linear model, an offset of duration was used. Recovery rate is modeled based on 10 regression coefficients that represent information on the cause of the outage, the severity, the location, the duration, and the time of the day and month.

Figure 1 is a scatterplot of all variables in the data set, each square represents a pairwise plot between the corresponding pair of variables on the x-axis and the y-axis, the red line represents a local regression line of the two variables. The numbers shown in the upper side of the scatterplot represent correlations of the pairs of variables which, in this data set, are not significant with the exception of a couple of variables. Examining Figure 1, it is difficult to identify visually any particular relationships beyond the expected linear correlations due to multicollinearity such as start date and time with restoration time. The plot provides histograms for the different variables and it can be seen that there is a large variance for many predictors.

Further examination of the patterns in the data focus on the impact of seasonal variations and types of disturbance on the recovery process from power outages. Rates of recovery are generally slower in the winter than in the summer months (Figure 2).

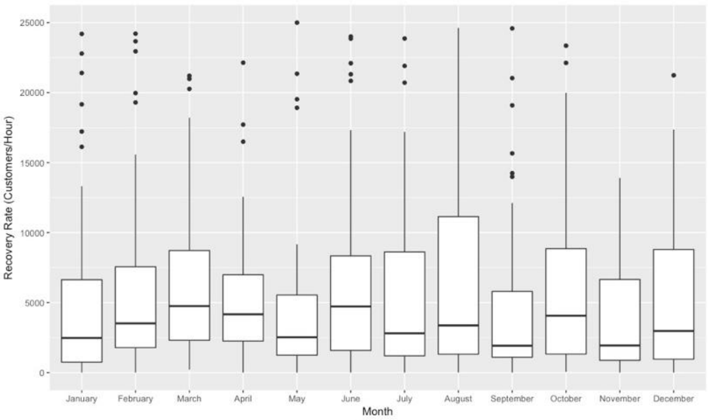

Figure 2. Recovery Rate as a Function of Month

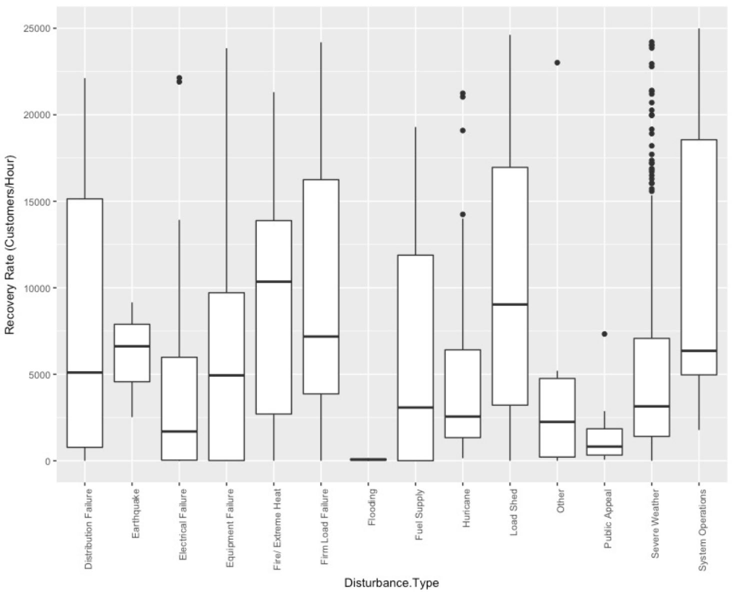

Figure 3. Recovery Rate by Disturbance Type

While wide variations are observed in the recovery rate by the type of disturbance, outages due to load shed and fire/extreme heat experience the highest average recovery rate. Disasters such as flooding and hurricane, however, have much slower recovery rates (Figure 3). Also, Severe Weather events result in the largest number of outliers in the data.

4.2 *Results*

The Poisson Bayesian kernel model referred to as PBK, the Poisson GLM referred to as PGLM, and the negative binomial GLM referred to as NBGLM were used to model the data and predict the recovery rate as a function of the predictors related to the time, location, disruption, and other characteristics. The error measures discussed earlier and presented in Table I were calculated for each model and summarized in Table II.

Across all predictive accuracy measures, PBK performs yields small errors overall. PGLM results in very large errors that could be driven by the extreme values under Severe Events for instance, whereas PBK is able to control for that and provide more stable estimates. For two of the predictive error measures, NRMSED and MAE, the PBK outperforms the NBGLM.

Table II: Prediction Error Values for all the Models

Model	RMSE	NRMSED	NRMSEM	MAE
PBK	2435	2.03	0.25	1258
PGLM	12961	10.88	1.34	5579
NBGLM	1706	2.06	0.17	2039

Overall the performance of PBK and NBGLM is comparable from a predictive accuracy standpoint. However, using PBK would provide an assessment of the uncertainty in the estimates through the prior and posterior distributions of the recovery rate, the outcome is a probability distribution of a comprehensive range of possible values for the recovery rate. As a result, it is possible for a decision maker to identify multiple point estimates based on their risk preference. For example, if the decision maker or infrastructure operator is risk averse, he/she will rely on a more extreme (lower) value than the expected value of the recovery rate posterior distribution since a more conservative mitigation and recovery strategy is preferred. However, if the decision maker is risk taking, the preference would be to save on cost of mitigation and recovery and the upper tail of the distribution will be considered as an optimistic measure of the recovery rate. The choice of the posterior point estimate is not the only way a decision maker is involved in this process. Stakeholders play an important role in identifying multiple initial parameters in the model.

As mentioned earlier, the definition of the prior is an important consideration for any Bayesian approach. In this case, a non-informative prior was assumed. However, another important consideration is the value of the parameter in the kernel function. The results in the table above were obtained based on an arbitrary value of sigma. In order to understand the effect of this parameter on the predictive accuracy, Figure 4 shows the value of the root mean squared error as a function of $1/\sigma$.

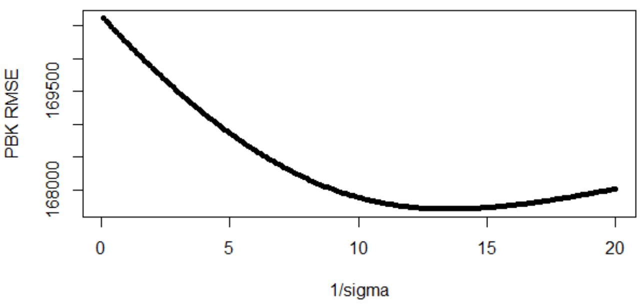

Figure 4. PBK RMSE as a Function of Different Values for the Tuning Parameter of the Kernel Function

There is clearly an optimal value for this parameter valued at approximately 12. It would be ideal if the parameter is tuned to minimize the error during the training process. The drawback of doing so is the additional computation time for tuning which would exponentially increase as more parameters are considered in other forms of the kernel function.

5 CONCLUSION

The work presented in this paper evaluates the use of Poisson Bayesian kernel models to measure and predict the rate of recovery. The ultimate goal of the research is to be able to quantify community resilience in order to inform resource allocation before, during, and after a disruption. The proposed approach to model the rate of recovery was compared to traditional count data models such as Poisson and negative binomial generalized linear models.

The advantage of using Bayesian techniques is their ability to provide probability distribution of the estimates, accounting for the uncertainty in resilience metrics. Another important benefit is the ability to update predictions as new information on the evolvement of the disaster and the corresponding response of the community becomes available.

An initial comparison to other methods shows that PBK provides a higher accuracy than traditional models with the added benefit of accounting for uncertainty and the decision maker's opinion and prior knowledge.

6 REFERENCES

Barker, K., & Baroud, H. (2014). Proportional hazards models of infrastructure system recovery. *Reliability Engineering & System Safety*, *124*, 201-206.

Baroud, H., Barker, K., Lurvey, R., and Mackenzie, C. (2013, January). Bayesian kernel model for disruptive event data. *In Proceedings of IIE Annual Conference*. (p. 1777). Institute of Industrial Engineers-Publisher.

Barabadi, A., & Ayele, Y. Z. (2018). Post-disaster infrastructure recovery: Prediction of recovery rate using historical data. *Reliability Engineering & System Safety*, *169*, 209–223. https://doi.org/10.1016/J.RESS.2017.08.018

Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, *1*(1), 29-53.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression Analysis of Count Data* (Vol. 53). Cambridge university press.

Carlin, B. P., & Louis, T. A. (2008). *Bayesian Methods for Data Analysis*. CRC Press.

Cutter, S. L., Ash, K. D., & Emrich, C. T. (2014). The geographies of community disaster resilience. *Global Environmental Change*, *29*, 65–77. https://doi.org/10.1016/J.GLOENVCHA.2014.08.005

Floyd, M. S., Baroud, H., & Barker, K. (2014). Empirical analysis of Bayesian kernel methods for modeling count data. In *Systems and Information Engineering Design Symposium (SIEDS), 2014* (pp. 328-333). IEEE.

Guikema, S. D., & Goffelt, J. P. (2008). A Flexible Count Data Regression Model for Risk Analysis. *Risk Analysis*, *28*(1), 213–223. https://doi.org/10.1111/j.1539-6924.2008.01014.x

Johansen, C., Horney, J., & Tien, I. (2016). Metrics for Evaluating and Improving Community Resilience. *Journal of Infrastructure Systems*, *23*(2), 04016032.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association, 91*(435), 1343-1370.

MacKenzie, C. & Barker, K. (2013). Empirical Data and Regression Analysis for Estimation of Infrastructure Resilience with Application to Electric Power Outages. *Journal of Infrastructure Systems*, *19*(1), 25–35. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000103

MacKenzie, C. A., Trafalis, T. B., & Barker, K. (2014b). A Bayesian Beta kernel model for binary classification and online learning problems. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *7*(6), 434-449.

Mason, M., & Lopes, M. (2011, March). Robot self-initiative and personalization by learning through repeated interactions. *In Conference on Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International* (pp. 433-440). IEEE.

Montesano, L., & Lopes, M. (2009, June). Learning grasping affordances from local visual descriptors. In *Development and Learning, 2009. ICDL 2009. IEEE 8th International Conference on* (pp. 1-6). IEEE.

Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche, K. F., & Pfefferbaum, R. L. (2008). Community Resilience as a Metaphor, Theory, Set of Capacities, and Strategy for Disaster Readiness. *American Journal of Community Psychology*, *41*(1–2), 127–150. https://doi.org/10.1007/s10464-007-9156-6

Schölkopf, B. & Smola, A.J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press