A Bayesian Approach for Modelling the Post-Earthquake Recovery of Electric Power Systems in Japan

Yuki Handa
Former Graduate Student, Dept. of Civil, Environmental and Geomatic Engineering, University College London, London, United Kingdom

Eyitayo A. Opabola
Senior Research Fellow, Dept. of Civil, Environmental and Geomatic Engineering, University College London, London, United Kingdom

Carmine Galasso
Professor, Dept. of Civil, Environmental and Geomatic Engineering, University College London, London, United Kingdom

ABSTRACT: Post-disaster recovery modelling of engineering systems has become an important facet of disaster risk management. The post-disaster recovery trajectory of a civil infrastructure system can be quantified using (a) the initial post-disaster functionality level, Q_o (i.e., the ratio of the number of serviced customers/end users post-disaster to that pre-disaster); (b) rapidity h (i.e., the rate of functionality restoration); and (c) recovery time (R_t) (i.e., the total time to restore full functionality to the entire community). This study uses a Bayesian estimation approach to develop probabilistic models for characterising the relationships between seismic intensity, exposed population (PEX), Q_o, h, and R_t of electric power networks (EPNs) using post-earthquake recovery data of large earthquakes occurring in Japan. Firstly, a data collection exercise was carried out to aggregate publicly available data on the aforementioned parameters for different seismic events in Japan. Based on the quality of available information, 16 earthquake events between 2003 and 2022 were selected. Next, a set of probabilistic models to estimate Q_o, R_t, and h were developed using Bayesian parameter estimation to capture uncertainties. The data analysis suggests that the initial post-disaster functionality level depends on the seismic intensity and exposed population. The post-disaster recovery time depends on the initial post-disaster functionality level, event magnitude, and year of occurrence. The rapidity of recovery depends on the initial post-disaster functionality level. Apart from being an efficient stand-alone tool, the proposed data-driven models can be a useful benchmarking tool for simulation-based models.

1. INTRODUCTION

Field-based evidence has shown that disaster-induced damages to critical infrastructure systems can cause significant direct and indirect socioeconomic losses, including casualties. One of the most affected critical infrastructure systems is electric power networks (EPN). Apart from economic losses associated with business disruptions, power outages can result in human losses, increased waste from perishable food items, failure of security systems, increased disease spread, significant direct repair and restoration costs, and several other forms of a nuisance to the general public (e.g., Chang et al. 2007; Dugan et al. 2023).

Accurately estimating power outages and recovery can help EPN management authorities and other private/public organisations reliant on electric power define effective short-term and long-term recovery strategies to improve community-level resilience. For example, business owners and homeowners can set up appropriate measures (e.g., backup power sources and power disruption insurance) if potential post-

disaster recovery trajectories of EPNs at different hazard intensities are known.

Figure 1 presents a typical post-disaster recovery trajectory for a component/system (or even a community) with a pre-disaster functionality level Q_{pre}, reduced to Q_o (referred to as the initial post-disaster functionality level) after an event at a time t_o. Repairing or replacing damaged system components will result in the gradual restoration of functionality. The time required to restore the system's functionality to a desired level Q_R (which can be similar, close to, or better than the pre-disaster functionality level) is referred to as the recovery time (R_t). The recovery path of the damaged system (i.e., from $\{t_o, Q_o\}$ to $\{t_R, Q_R\}$) can be quantified by a recovery function $Q(t)$. The recovery rapidity h quantifies the recovery rate. An ideal recovery modelling approach should be able to estimate Q_o, R_t, h, and $Q(t)$. Furthermore, any sources of uncertainties in these metrics need to be accounted for.
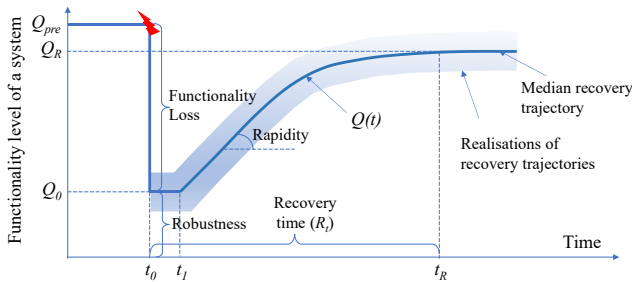

Figure 1: Post-disaster recovery trajectory of an engineering system or community

Currently, two main approaches for infrastructure recovery modelling are widely adopted – empirical and simulation-based. Empirical models for EPN assessments (e.g., Guikema and Quiring 2012; Liu et al. 2005; Nojima and Sugito 2003) are based on correlations from data analysis of historical spatial outages and recovery times. Given the data resolution used in developing these models, most empirical models do not explicitly account for the network topology and fragility of individual components and subcomponents in the EPNs. However, empirical models are considered helpful in forecasting region-level disaster-induced outages. It is noted that the majority of existing empirical models were developed for hurricane and storm hazards. Furthermore, available empirical models on post-earthquake recovery of EPNs do not adequately link earthquake features (e.g., magnitude, intensity, etc.), exposed population, post-earthquake functionality level, and recovery time. Hence, it is important to develop more efficient empirical models for post-earthquake recovery of EPNs.

Simulation-based methods (Çağnan et al. 2006; Guidotti et al. 2016; Ouyang and Dueñas-Osorio 2014) use network analysis to simulate the post-disaster functionality level and recovery trajectory of EPNs, while explicitly accounting for component and system fragilities. One of the fundamental challenges of simulation-based models is that, due to security reasons, EPN topologies are not publicly available data. Also, reliable information on component and system fragilities and repair/replacement times and sequences are needed to develop reliable estimates of post-disaster recovery trajectories. The adequacy of simulation-based methods can be improved through appropriate validation exercises using real-life events. Empirical models may be helpful benchmarking tools whenever such validation exercises are not feasible.

This study's main objective is to enhance disaster risk management of EPNs by developing empirical models that adequately link earthquake intensity, exposed population, post-earthquake functionality level, rapidity, and recovery time. To this end, this study (a) collects and aggregates data on the performance and restoration of EPNs from past earthquake events in Japan; and (b) develops simple probabilistic models for post-disaster recovery trajectory estimation using relatively easy-to-obtain information such as seismic intensity, exposure data (in terms of number of households exposed to each local seismic intensity level), number of serviced households, and EPN restoration trajectory data.

2. DATA COLLECTION AND AGGREGATION

The first step is establishing a database of EPN performances/damages and restorations from past earthquakes in Japan. The data collection exercise for this study focused on major earthquakes that caused significant damage to EPNs. Another criterion was the need to have sufficient information on the initial post-disaster functionality level, recovery trajectory and recovery time. Based on these criteria, 16 events between 2003 and 2022 were considered. Information collected includes the earthquake magnitude, time of occurrence, exposed population data, number of serviced households, and recovery time for the EPNs. The collated database is available in Handa (2022). The subsequent subsections provide details on the data collection process.

2.1. Earthquake magnitude

This study adopts the Japan Meteorological Agency (JMA) magnitude (M_{JMA}) to quantify the magnitude of each event. More details on the JMA magnitude scale can be found in JMA (2014). The M_{JMA} for each event was extracted from Japan Real-time Information for earthQuake (J-RISQ) reports (J-RISQ 2015).

2.2. Population data

The number of households and population in the affected regions are taken from publicly available census data (SBJ 2020). For the number of households for the entire region, if the municipality where the affected households are located is known, the total number of households within that area is used; if not, the total number of households within the jurisdiction of the electricity company's sales office or prefecture is used. In some instances, information on the total number of households was unavailable. However, it was inferred from events with information on both the number of households and the population, that the average ratio of population to the number of households equals 2.8 (i.e., 2.8 persons per household). Hence, if the number of households was unavailable, it was inferred by dividing the reported population by 2.8.

2.3. Number of serviced customers (households)

This is the number of households with continued access to power immediately after the event. The number of serviced customers and total customers were extracted from published literature. Priority was given to data published by municipal electricity companies. In cases where incomplete information is available from electricity companies, data from newspapers and journal articles were used. Q_o is defined as the ratio of the number of households with continued access to power immediately after the disaster to the total number of households in the affected region.

2.4. PEX (population exposed to the earthquake) data

Data on the exposed population at each JMA (local) seismic intensity scale level were extracted from J-RISQ reports (J-RISQ 2015). JMA intensity scale intensities range from 0 to 7. More information on the JMA intensity scales can be found in JMA (2015).

JMA (2015) identifies seismic intensities of '5 Lower' or more to influence electric power disruptions. Hence PEX data collection focused on the population exposed to seismic intensities '5 Lower' (P_{5L}), '5 Upper' (P_{5U}), '6 Lower' (P_{6L}), '6 Upper' (P_{6U}), and 7 (P_7). The PEX data for each event was normalised by the total population in the high-intensity zones (as defined in section 2.2) (i.e., $p_{5u} = P_{5U}/(P_{5L}+P_{5U}+P_{6L}+P_{6U}+P_7)$).

2.5. Recovery time

The recovery time is defined as the time taken to restore power to all affected households in the affected region. The recovery time was extracted from publicly available information. Priority was given to information published by electricity companies, local authorities and relevant ministries and agencies. In certain cases, newspaper data (typically in the form of government announcements) were used.
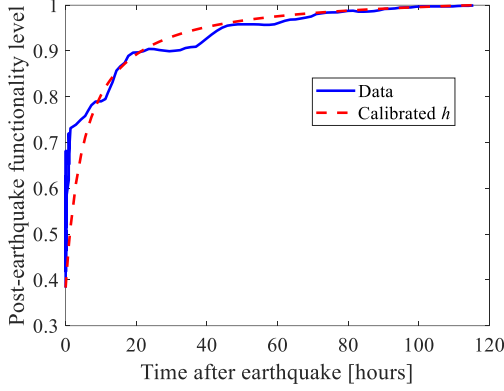
Figure 2: Post-recovery trajectory following the 2016 Kumamoto earthquake showing calibrated rapidity coefficient h (Data from METI (2016)*)*

2.6. Rapidity

As mentioned earlier, rapidity characterises the recovery rate of the EPN. Opabola and Galasso (2023) proposed a recovery trajectory function $Q(t)$ that is dependent on Q_o, mobilisation time t_1 (i.e., the time after which restoration work starts), R_t, and rapidity coefficients g and h (See Equation (1)).

$$Q(t) = Q_o + \frac{g\left(\frac{t - t_1}{R_t - t_1}\right)}{1 + gh\left(\frac{t - t_1}{R_t - t_1}\right)}, \qquad (1)$$

The values of g and h express the level of preparedness, resource availability, technical know-how and societal conditions of a community. For known values of h and $Q(R_t)$, g can be computed from Equation (1) as:

$$g = \frac{Q(R_t) - Q_o}{Q_o h - h + 1}, \qquad (2)$$

A concave recovery curve corresponds to $h \ll 0$ and represents the recovery trajectory of a community with a poor level of preparedness, a high level of resource unavailability, and unfavourable societal conditions. In contrast, $h \gg 0$ represents a community with good preparedness and resource availability and favourable societal conditions.

The rapidity coefficient h from each event was estimated by fitting Equation (1) to the field data (see Figure 2 for an example) so that the Sum of Squares Error (SSE) between the observed and predicted functionality trajectory during the recovery phase is minimised.

3. DATA ANALYSIS AND PROPOSED MODELS

3.1. Correlation analysis

A correlation analysis for the aggregated database was first conducted. The correlation matrix is presented in Figure 3. As shown in the figure, Q_0 is negatively correlated with the PEX (more significant correlation with p_7 and p_{6u}). This trend captures that the higher the proportion of a population exposed to a more significant seismic intensity, the lower the initial post-disaster functionality level.

R_t is shown to be highly correlated with Q_o ($r = -0.84$), year of occurrence and earthquake magnitude. Low Q_o corresponds to increased damage to EPN components, resulting in a longer time to recover. The significant correlation of PEX with R_t may be attributed to the fact that PEX and Q_o are also highly correlated.

Furthermore, the rapidity coefficient h is shown to be correlated with Q_o and R_t with r equals 0.49 and -0.42, respectively. As earlier described, a higher level of h depicts a system with good preparedness, good resource availability, and favourable societal conditions. Hence, the positive correlation between h and Q_o can be attributed to the fact that a higher Q_o is also associated with good preparedness and favourable societal conditions.

3.2. Bayesian linear regression approach

Unlike frequentist approaches, a Bayesian parameter estimation yields a probability distribution of model parameters instead of a single value. Various literature (e.g., Gelman et al. 1995; Reich and Ghosh 2019) provides a detailed introduction to Bayesian methods. This paper provides a brief summary of Bayesian parameter estimation.

Consider a standard linear regression model given as:

$$y_n = \boldsymbol{\beta}^{\mathbf{T}} \boldsymbol{x_n} + \varepsilon_n \qquad (3)$$

where y_n is a scalar response, $\boldsymbol{\beta}$ is a vector of model parameters (or coefficients) for the vector

covariance matrix of Gaussian prior of $\boldsymbol{\beta}$; and $\boldsymbol{\mu}$ is the mean hyperparameter of Gaussian prior on
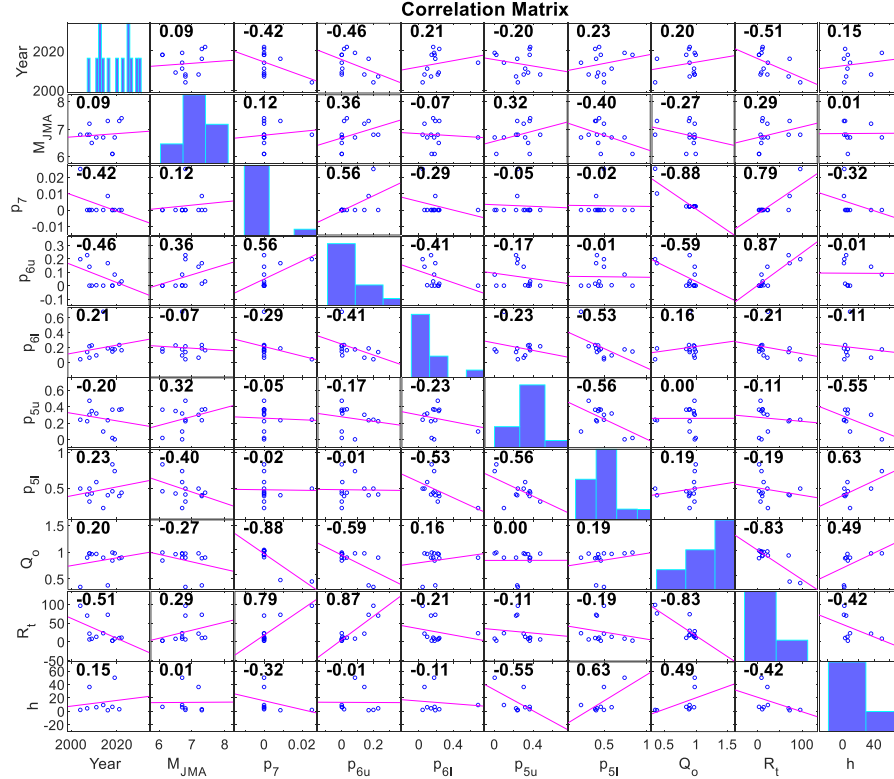
Figure 3: Correlation matrix for extracted data

of regressor $\boldsymbol{x_n}$, and error ε_n is a zero mean Gaussian noise term with a non-zero variance σ^2.

The probabilistic interpretation of Equation (3) can be written as:

$$y_n|\boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}\left(y_n|\boldsymbol{\beta}^{\mathbf{T}}\boldsymbol{x_n}, \sigma^2\right) \qquad (4)$$

In Equation (4), y_n is a random variable with mean of $\boldsymbol{\beta}^{\mathbf{T}}\boldsymbol{x_n}$ and a corresponding non-zero variance of σ^2.

The Bayesian treatment of a linear regression model introduces a prior probability distribution over the model parameters β and σ. In this study, we consider the semiconjugate prior for which the prior and posterior share the same parametric family. The natural semiconjugate prior of parameters β and σ^2 is a normal-inverse-gamma distribution, with the form:

$$p(\boldsymbol{\beta}, \sigma^2) = p(\beta|\sigma^2)p(\sigma^2) \qquad (5)$$

where $(\boldsymbol{\beta}|\sigma^2) \sim \mathcal{N}(\boldsymbol{\mu}, V)$ and $\sigma^2 \sim IG(A, B)$. $\mathcal{N}$ is the normal distribution, and IG is the inverse-gamma distribution. V is the conditional

β. A and B are the shape and scale hyperparameters of the inverse-gamma prior on σ^2, respectively.

The prior probability distribution of a parameter can then be combined with the likelihood of the observed data to obtain the posterior distribution of the parameter (see Equation (6)).

$$p(\boldsymbol{\beta}, \sigma^2|y) \propto p(y|X, \boldsymbol{\beta}, \sigma^2)p(\boldsymbol{\beta}, \sigma^2) \qquad (6)$$

This study adopts the Markov Chain Monte Carlo (MCMC) algorithm (precisely, Gibbs Sampling) (e.g., Gelfand 2000) for the posterior estimation. The Bayesian regression analyses presented in this study drew 10,000 samples from the posterior distribution generated using a burning period of 10,000 and a thinning level of 10. It is also noted that a stepwise removal process was carried out via a Bayesian estimation method to identify the governing predictors for Q_o, R_t, and h.

3.3. Initial post-disaster functionality level Q_o

The Bayesian approach outlined in section 3.2 was adopted in defining the probabilistic formulation for Q_o. Two data points were removed from the analysis. The 2011 Tohoku event was excluded because the functionality loss and recovery trajectory data are assumed to be significantly influenced by the earthquake and tsunami sequence (rather than earthquake only, as for the other considered events). The 2018 Hokkaido event was excluded because the total blackout was a precautionary measure due to a fire event at one power plant.

The stepwise removal process showed that the main predictors for Q_o are the normalised PEX data. Of the PEX data, p_{5l} was determined to be the least significant in predicting Q_o; hence it was not considered in the Bayesian analysis.

Table 1 presents the posterior summary statistics for the proposed model. The model assumes that the higher the proportion of a population exposed to a more significant seismic intensity, the lower the initial post-disaster functionality level, which is intuitive.

It is noted that the lower- and upper-bound values of Q_o are zero and unity, respectively. Also, it is worthwhile to constrain the model parameters to negative values (i.e., to capture reduction in Q_o). In line with the mean values, another constraint may be to ensure $|\beta(p_7)| > |\beta(p_{6u})| > |\beta(p_{6l})| > |\beta(p_{5u})|$.

Table 1: Posterior summary statistics for Q_o model

Parameter	Mean	Standard deviation	95% credible region
Intercept	1.0	0.13	[0.79, 1.3]
p_7	-24.7	6.6	[-37.7, -11.5]
p_{6u}	-0.4	0.6	[-1.79, 0.6]
p_{6l}	-0.22	0.29	[-0.84, 0.3]
p_{5u}	-0.1	0.29	[-0.78, 0.4]
σ^2	0.02	0.01	[0.006, 0.05]

Figure 4 shows the relationship between the estimated mean and measured Q_o. As shown in the figure, the model provides adequate estimates of Q_o with an R^2 of 0.81 and normalised root mean squared error (NRMSE) of 0.1.
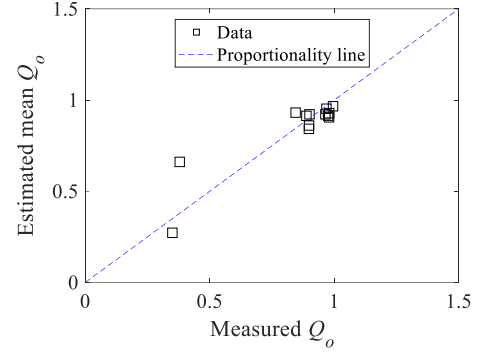

Figure 4: Relationship between measured and estimated mean Q_o

3.4. Recovery time R_t

The stepwise removal process showed that three main predictors for R_t are M_{JMA}, Q_o, and the occurrence year (Y). As discussed earlier, Q_o accounts for PEX. It is presumed that the sensitivity of recovery time to event magnitude is because the earthquake severity on other infrastructure systems and the entire community determines the resources dedicated to repairing damaged EPNs. For example, larger events could have severe impacts on other lifelines which are interdependent with EPNs.

Based on the estimated NRMSE, the predictors relate better with R_t in the natural log space. Hence, the Bayesian analysis was carried out in natural log space (Equation (7)).

$$\ln(y_n) = \boldsymbol{\beta^T} \cdot \ln(\boldsymbol{x_n}) + \varepsilon_n \qquad (7)$$

Table 2 presents the posterior summary statistics for the proposed model. The model captures the increase in recovery time with an increase in magnitude and proportion of customers without service (i.e., $1 - Q_o$). Similarly, the sensitivity of recovery time to occurrence year can be attributed to disaster management agencies learning from past events to reduce the post-earthquake recovery time.

Figure 5 shows the relationship between the estimated mean and measured R_t. As shown in the figure, the model provides adequate estimates of R_t with an R^2 of 0.75 and NRMSE of 0.16.

Table 2: Posterior summary statistics for R_t model

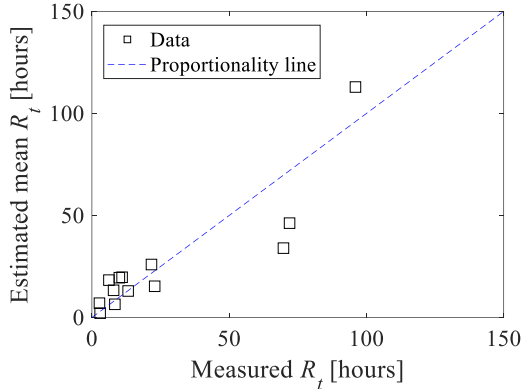| Parameter	Mean	Standard deviation	95% credible region
ln(Intercept)	-7.5	7.4	[-22.2, 7.2]
ln(M_{JMA})	7.2	3.77	[-0.35, 14.7]
ln($1-Q_o$)	0.34	0.17	[0.00, 0.69]
ln($Y-2000$)	-1.0	0.42	[-1.9, -0.22]
σ^2	0.56	0.35	[0.21, 1.45]

Figure 5: Relationship between the measured and estimated recovery time

3.5. Rapidity coefficient h

As shown in Figure 3, *h* is positively correlated with the initial post-earthquake functionality level and negatively correlated with R_t. There are also significant correlations between spatial seismic intensities and *h*. The stepwise removal process suggests that $h = f(1 - Q_o)$. R_t was excluded during the removal process due to a high correlation with Q_o ($r = -0.84$). Furthermore, as shown in Figure 6, there is a nonlinear relationship between shape constant *h* and $1 - Q_o$. Hence, the Bayesian analysis was carried out in natural log space also in this case (see Equation (7)).

Table 3 presents the posterior summary statistics for the proposed model. The adequacy of the mean estimate is represented in Figure 6. As shown in the figure, the proposed h model provides a good estimate of *h* with an R^2 of 0.84 and NRMSE of 0.17.

4. DISCUSSIONS AND CONCLUSIONS

As a result of the unavailability of EPN topology for several earthquake-prone countries, simulation-based recovery modelling studies often adopt synthetic test-beds. The adequacy of simulation-based methods can be improved through appropriate validation exercises using realistic data. Empirical models may also be helpful benchmarking tools whenever such validation exercises are not feasible. Furthermore, such empirical models can serve as simple stand-alone tools for integrating community-level and building-level recovery modeling frameworks.
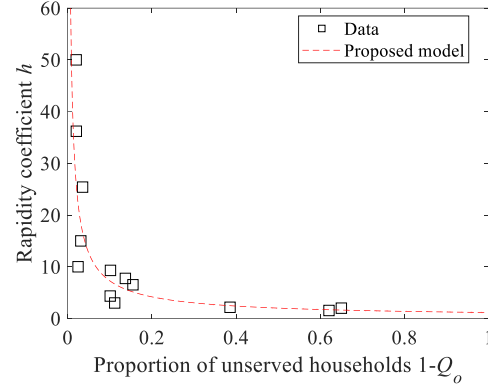

Figure 6: Relationship between proportion of unserved households ($1-Q_o$) and rapidity coefficient h

Table 3: Posterior summary statistics for h model

Parameter	Mean	Standard deviation	95% credible region
ln(Intercept)	0.14	0.28	[-0.41, 0.7]
ln($1-Q_o$)	-0.8	0.11	[-1.0, -0.59]
σ^2	0.26	0.13	[0.1, 0.6]

This study has developed a set of probabilistic models to predict Q_o, R_t, and *h*. using Bayesian parameter estimation to capture uncertainties. The order of applying the proposed formulations for predicting the recovery trajectory of EPNs, using posterior summary statistics presented in Tables 1, 2, and 3, is presented in Figure 7.

The proposed probabilistic tool can capture aleatory and epistemic uncertainties in the model. However, additional data collection exercises are needed to update the probability distribution functions of the parameters with new observations. Future studies will also look at collecting data from other countries to improve the applicability range of the proposed formulations.
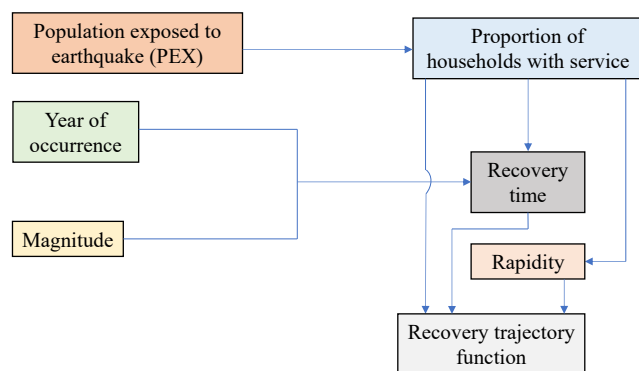
Figure 7: Order of analysis using the proposed models

5. REFERENCES

Çağnan, Z., Davidson, R. A., and Guikema, S. D. (2006). "Post-Earthquake Restoration Planning for Los Angeles Electric Power." *Earthquake Spectra*, 22(3), 589–608.

Chang, S. E., McDaniels, T. L., Mikawoz, J., and Peterson, K. (2007). "Infrastructure failure interdependencies in extreme events: power outage consequences in the 1998 Ice Storm." *Natural Hazards*, Springer, 41, 337–358.

Dugan, J., Byles, D., and Mohagheghi, S. (2023). "Social vulnerability to long-duration power outages." *International Journal of Disaster Risk Reduction*, 85, 103501.

Gelfand, A. E. (2000). "Gibbs Sampling." *Journal of the American Statistical Association*, 95(452), 1300–1304.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Guidotti, R., Chmielewski, H., Unnikrishnan, V., Gardoni, P., McAllister, T., and van de Lindt, J. (2016). "Modeling the resilience of critical infrastructure: the role of network dependencies." *Sustainable and Resilient Infrastructure*, 1(3–4), 153–168.

Guikema, S. D., and Quiring, S. M. (2012). "Hybrid data mining-regression for infrastructure risk assessment based on zero-inflated data." *Reliability Engineering & System Safety*, 99, 178–182.

Handa, Y. (2022). "Characterising post-earthquake recovery times for electric power systems." University College London, UK.

J-RISQ. (2015). "J-RISQ earthquake report." <https://www.j-risq.bosai.go.jp/report/en/R-20230302154801-0010> (Sep. 1, 2022).

JMA. (2014). "User's guide. The Seismological Bulletin of Japan." <https://www.data.jma.go.jp/svd/eqev/data/bulletin/catalog/notes_e.html> (Aug. 4, 2022).

JMA. (2015). "Tables explaining the JMA Seismic Intensity Scale." <https://www.jma.go.jp/jma/en/Activities/inttable.html> (Aug. 4, 2022).

Liu, H., Davidson, R. A., Rosowsky, D. V., and Stedinger, J. R. (2005). "Negative Binomial Regression of Electric Power Outages in Hurricanes." *Journal of Infrastructure Systems*, 11(4), 258–267.

METI. (2016). *Facility damage and power outage restoration response to the 2016 Kumamoto earthquake.*

Nojima, N., and Sugito, M. (2003). "Development of a Probabilistic Assessment Model for Post-Earthquake Residual Capacity of Utility Lifeline Systems." *Advancing Mitigation Technologies and Disaster Response for Lifeline Systems*, American Society of Civil Engineers, Reston, VA, 707–716.

Opabola, E. A., and Galasso, C. (2023). "A Probabilistic Framework for Post-Disaster Recovery Modeling of Buildings and Utility Networks." *Earthquake Spectra (under review).*

Ouyang, M., and Dueñas-Osorio, L. (2014). "Multi-dimensional hurricane resilience assessment of electric power systems." *Structural Safety*, 48, 15–24.

Reich, B. J., and Ghosh, S. K. (2019). *Bayesian Statistical Methods*. Chapman and Hall/CRC, New York.

SBJ. (2020). "Population and Households." *Statistics Bureau of Japan*, <https://www.stat.go.jp/english/data/kokusei/index.html> (Aug. 4, 2022).