

Hierarchical Bayesian reliability assessment of energy networks

Tomas Iešmantas^{1,2}, Robertas Alzbutas^{1,2}

¹*Faculty of Mathematics and Natural Sciences, Kaunas University of Technology*
K. Donelaicio 73, LT-44029 Kaunas

²*Laboratory of Nuclear Installations Safety, Lithuanian Energy Institute*
Breslaujos 3, LT-44403 Kaunas
E-mail: tomas.iesmantas@ktu.lt. robertas.alzbutas@ktu.lt

Abstract. It is often the case that data used for the systems reliability assessment comes from more than one information source. Whether they are power plants at different geographical locations, gas transmission pipelines operating in different environment or power transmission networks deployed within various areas. Therefore, different operating conditions, varying maintenance programs and efficiencies have its share in influencing the vulnerability and variability of reliability data. However, in practice it is usually the case that this heterogeneity is neglected leading to the underestimation of underlying uncertainty of the data. Bayesian models are capable of dealing with this kind of uncertainty as opposed to the frequentists statistical methods. Hierarchical Bayesian modelling technique provides means to quantify not only within-source, but also between-source uncertainties. Even in the case of small data samples it performs well, unlike for example the classical likelihood method which may provide degenerate estimates. In this paper authors investigate the possibility to incorporate this kind of uncertainty into the systems reliability and vulnerability assessment through the Bayesian framework in several cases: gas transmission networks and power transmission grids.

Keywords: Bayesian reliability, networks, heterogeneity, hierarchical Bayesian methods.

Introduction

There is an issue with systems reliability data that deserves a wider attention from the scientific community. Consider a case when data is collected from components or systems operating in various areas. For example, overhead electricity lines deployed through entire area of the country, pipelines operating in varying soil and environmental conditions. Or maybe some class of rare components, for which the data is collected from nuclear power plants operating in different countries or continents. These are examples of cases when in order to assess reliability and vulnerability, the usual choice is to pool the failure data.

In reliability engineering applying statistical data analysis it is a quite common practice to pool statistical information from different sources. At first sight it seems quite natural to make such decision: if systems are similar and perform the same function, then samples also should be treated as similar. However, such aggregation causes a loss of information about specificity of those systems and impact of their environment. However, similar components can have different ageing behaviour when

operating in different environmental and maintenance conditions. In addition, there is another issue, leading to the decision to pool the data – highly reliable systems, especially if they are age-dependent, do not supply sufficient statistical information for long-term reliability investigation, so that aggregation would strengthen statistical inferences. The drawback of such heterogeneous and rare data pooling is that one may get estimates with smaller uncertainty bounds, than it would be when between-source (separate samples) variability is considered. Those too optimistic uncertainty bounds will lead to less strict safety margins what itself causes higher risk of safety limits exceedance [2]. In this paper we are concerned with modelling reliability data heterogeneity, since it has a significant influence on the overall system reliability and its prediction considering data uncertainty. There are two approaches that could be used for the statistical analysis of heterogeneous data: mixed-effect models [5] and hierarchical Bayesian models [1]. The second class was chosen and together with Bayesian methods will be employed through this entire paper without turning to the frequentist concepts. We investigate the possibility to incorporate heterogeneity into the systems reliability and vulnerability assessment through the Bayesian framework in several cases: gas transmission networks and power transmission grids.

1 Hierarchical Bayesian modelling of reliability data

Due to the widespread use of the Bayesian methods we assume prior knowledge of basic probabilistic language in constructing the Bayesian models: likelihood function, prior distribution, and posterior distribution. Consider N samples of failure events (reliability data) observations, i.e. we observe N systems and register the failures of some particular group of components. For the sake of simplicity we assume that the number of data points in each sample is the same and is equal to M . One option is to pool all samples to obtain only one, as already discussed in the introduction section, but this would most likely lead to the incorrect variance estimates. Another option is to build a hierarchy and to introduce a partial pooling. Suppose we have a reason to believe that each source generates data from similar but not exactly the same probabilistic model. Denote the model of the i -th source by $f(x|\theta_i)$. This means that we do not assume that data from different sources are identically distributed. These models form the first level of the hierarchy. We have now sets of unknown parameters that needs to be estimated. Estimating these parameters as though they are independent from each other would lead to large variances especially if the samples are small. However, we assume that these unknown parameters are related to each other by enforcing a stochastic model onto them. We treat parameters θ_i as unobserved data (or unobservables) and assume a model $\pi(\theta_i|\zeta)$, where ζ is another unknown parameter that has to be estimated. This model is called the second level of the hierarchy. It seems like we made our lives more difficult by introducing even more unknown parameters. However, this second level model enables the sharing of information between those N samples, so that they are partially related (partially pooled). Now we put all the hierarchies into a mathematical expressions:

$$\begin{aligned} X_{i,j}|\theta &\sim f(x|\theta_i), \\ \theta_i|\zeta &\sim \pi(\theta_i|\zeta), \\ \zeta &\sim \pi_0, \end{aligned}$$

where $X_{i,j}$ is the j -th observation for i -th source, π_0 is a prior distribution for the parameter ζ . The so called posterior distribution $\pi(\theta_i|\zeta)$ is proportional to $L(X, \theta) \cdot \pi_0$, i.e. the product of likelihood and prior distribution. All the inferences are then made from this posterior distribution. The calculations are another thing most likely posterior distribution will not have an analytic form and approximating algorithms must be employed. For this the most common choice is Markov Chain Monte Carlo algorithms [4].

2 Demonstration of hierarchical reliability modelling

2.1 Power transmission grid

13 overhead lines of 330 kV rating were observed over the period of 42 years. To be more specific, not all of the lines were observed over exactly 42 years, some were put in operation several years later comparing to others. Therefore we have 13 data sources with registered outages. More about the application of hierarchical Bayesian models in power grid context see the recent work of authors [3].

Variability of data clearly indicates that one cannot assume the same stochastic model for all overhead lines – frequencies are simply too different. Therefore we consider building a hierarchical Bayesian model. If X_i denotes the number of outages over the entire observation period Δ_i for the i -th overhead line of length L_i and if we assume a Poisson distribution for the count data, the stochastic model is as follows:

$$\begin{aligned} X_i|\theta &\sim \text{Poisson}(\exp(\lambda_i \Delta_i L_i)), \\ \lambda_i|\zeta &\sim N(\mu, \sigma^2), \\ (\mu, \sigma^2) &\propto 1, \end{aligned}$$

where $\exp(\lambda_i)$ denotes the failure rate for i -th line when length and period are assumed to be equal to 1. Exponential function was chosen due to the simplifications, because now λ_i are not restricted to the fixed positive values and normal distribution with parameters μ, σ^2 can be imposed over these parameters. The likelihood function can be expressed as follows:

$$L(X|\lambda) \propto \sigma^{-N} \exp \left[\sum_{i=1}^N (-e^{\lambda_i} + X_i \lambda_i) - \frac{1}{2\sigma^2} \sum_{i=1}^N (\lambda_i - \mu)^2 \right].$$

Since the prior distribution for the parameters μ, σ^2 is chosen to be uniform, the posterior is proportional to the likelihood function $L(X|\lambda)$. The estimates of the outage rates are presented in Fig. 1.

If one would treat the outage rates as coming from the same probabilistic model, one would get posterior outage estimate equal to 2.07, obviously inappropriate value for the variation level present in the data sample.

2.2 Reliability of natural gas pipeline network

Now consider a problem of evaluating the reliability natural gas transmission pipeline network. It is often the case that for this task international databases are used because

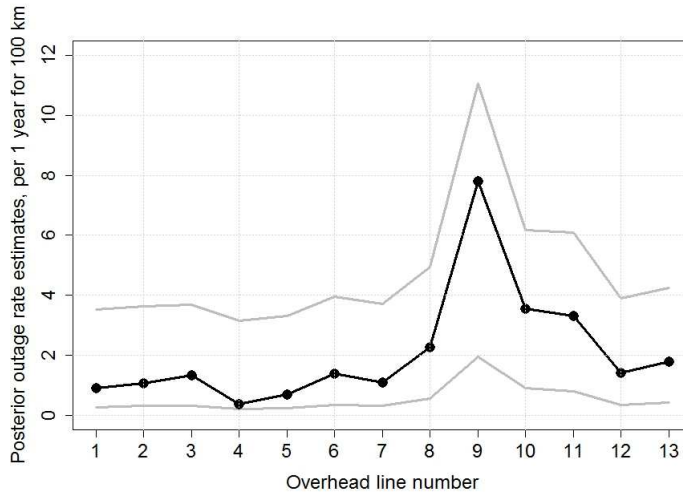


Fig. 1. Posterior outage rates estimates together with 0.95 uncertainty bounds (in grey).

of large data samples contained in there. Most widely known such databases are OPS PHMSA [2], EGIG [2], UKOPA [9], and NEB [10]. However, OPS data sample has two break points at the years when the data collection criterion was changed. In addition, for our case study we include Lithuanian data sample, which is very small and has only 7 data points with one breakpoint. Therefore we have a 5 data samples coming from different sources albeit representing the similar components. How to deal with those data breakpoints and changing collection criteria see [8, 6, 7]. The model constructed to deal with additional breakpoints was called a Criteria-Dependent Poisson model (or CDP model). The hierarchical model in this particular case is quite complex as it includes age-dependency, change in collection criteria and the hierarchical structure:

Level I

$$X_t^1 | \theta^t \sim \begin{cases} \text{Poisson}(E_t^1 \lambda(t; \theta^1)), & t = 1, \dots, 14 \\ \text{Poisson}(E_t^1 \lambda(t; \theta^1)(1 - p^1 - p^2)), & t = 15, \dots, 33 \\ \text{Poisson}(E_t^1 \lambda(t; \theta^1)(1 - p^1)), & t = 34, \dots, 42 \end{cases}$$

for OPS case

$$X_t^1 | \theta^t \sim \begin{cases} \text{Poisson}(E_t^2 \lambda(t; \theta^2) p^1), & t = 19, \dots, 34 \\ \text{Poisson}(E_t^2 \lambda(t; \theta^2)), & t = 35, \dots, 44 \end{cases}$$

for Lithuanian case.

$$X_t^1 | \theta^t \sim \text{Poisson}(E_t^k \lambda(t; \theta^k)), \quad k = 3, \dots, 5$$

for EGIG, UKOPA, NEB cases.

Unknown parameters are assumed to have lognormal distribution. Here X_t^k denotes the k -th sample which is age-dependent, E_t^k is the explosion, p^1 , and p^2 , are probabilities of data falling under certain collection criteria (think about it as an adjustment factor resulting from different data collection criterion), also we made an assumption that $\lambda(t; \theta)$ has two unknown parameters $\theta^k = (\theta_1^k, \theta_2^k)$ of which the first

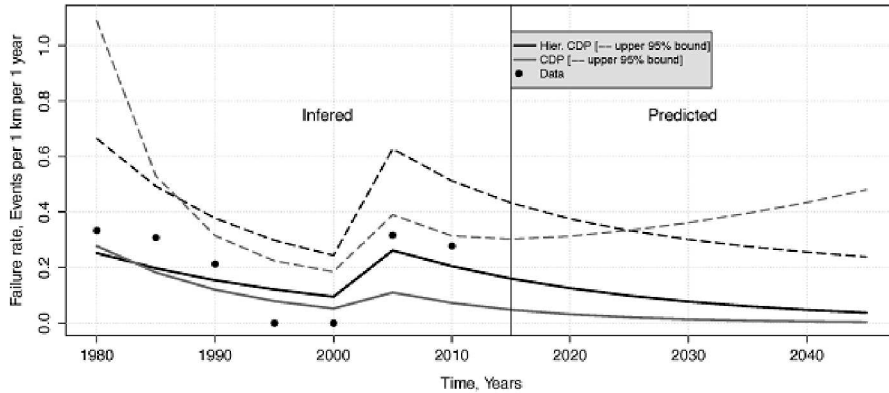


Fig. 2. Inferred and predicted failure frequencies for simple CDP and its hierarchical modification for Lithuanian gas transmission.

is positive, hence has lognormal distribution, and the second is normally distributed. Normality and log-normality are assumptions (could be changed by another) and the sensitivity of them has not been investigated in this paper. Now we turn to the inference for Lithuanian pipeline network. Due to the small size of the network and the incident criterion used until 2004, the data is quite scattered and the inferences based on it alone would be questionable. Therefore it is an advantage to be able to support this small sample with the information contained in the data samples from other databases.

When comparing the estimated trend (see Fig. 2, inferred zone), hierarchical model generally provides wider credibility bounds, than obtained from non-hierarchical variant of CDP model (only considering Lithuanian case). This is due to the fact that hierarchical structure of the model allows incorporating and quantifying additional level of uncertainty, i.e. variation throughout different databases is now accounted as well. In addition, two data points that were collected under new incident criterion (since 2003) are less underestimated by hierarchical model.

3 Conclusions and further comments

In this paper we have tackled the task of including additional level of uncertainty into the model of systems reliability. We briefly described the mathematical machinery and illustrated it with three quite different examples. The implications of including heterogeneity of the failure data are wide ranging – uncertainty due to heterogeneity pierces every reliability concept not only failure rate. The easy-to-interpret Bayesian modelling language serves well this purpose and enables to create complex models with various degrees of uncertainty taken into account. We have demonstrated that modelling of uncertainty due to data heterogeneity provides more realistic inferences and thus gives us practical approach to make more informed decisions related to reliability and vulnerability of various systems.

References

- [1] P.D. Congdon. *Applied Bayesian Hierarchical Methods*. CRC Press, 2010.

- [2] EGIG. *Gas Pipeline Incidents*, European Gas Pipeline Incident Group. Technical report, EGIG, 2008.
- [3] E. Golub, J. Greenfeld, R. Dresnack, F.H. Griffis and L.J. Pignataro. *Pipeline accident effects for natural gas transmission pipelines. Final report*. Technical report, New Jersey Institute of Technology, 1996.
- [4] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1):97–109, 1970.
- [5] J. Hox. *Multilevel Analysis: Techniques and Applications*. Erlbaum, Mahwah, NJ, 2002.
- [6] T. Iesmantas and R. Alzbutas. Age-dependent hierarchical Bayesian modelling for reliability assessment under small data sample. In *11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference*, Helsinki, Finland, 2012.
- [7] T. Iesmantas and R. Alzbutas. Methodology for gas transmission network age-dependent reliability assessment considering variation of incident criteria. In *10th Annual International Conference of Young Scientists on Energy Issues (10 CYSENI Anniversary)*, Kaunas, Lithuania, 2013.
- [8] T. Iesmantas and R. Alzbutas. Bayesian assessment of electrical power transmission grid outage risk. *Int. J. Electr. Power Energy Syst.*, **58**:85–90, 2014.
- [9] R. McConell and D.J.V. Haswell. *UKOPA Pipeline Product Loss Incidents (1962–2010)*. Technical report, UKOPA, 2011.
- [10] National Energy Board. *Focus on Safety and Environment: a Comparative Analysis of Pipeline Performance, 2000–2009*. Technical report, National Energy Board, 2011.

REZIUOMÉ

Hierarchinių Bajeso metodų taikymas energetinių tinklų patikimumui vertinti

T. Iešmantas, R. Alzbutas

Straipsnyje nagrinėjama energetikos objektų patikimumo duomenų heterogeniškumas. Dviems atvejams pademonstruotas hierarchinių Bajeso metodų taikymas ir įvertinta, kaip papildomas neapibrėžtumo lygmuo keičia patikimumo charakteristikų įverčius.

Raktiniai žodžiai: Bajeso patikimumo metodai, tinklai, heterogeniškumas, hierarchiniai Bajeso metodai.