

HIDDEN MARKOV MODELS WITH COVARIATES FOR ANALYSIS OF DEFECTIVE INDUSTRIAL MACHINE PARTS

Pornpit Sirima and Premysl Pokorny

Department of Manufacturing Systems and Processes,
Faculty of Mechanical Engineering, Technical University of Liberec, Czech Republic

Received 2014-04-23; Revised 2014-06-09; Accepted 2014-07-26

ABSTRACT

Monthly counts of industrial machine part errors are modeled using a two-state Hidden Markov Model (HMM) in order to describe the effect of machine part error correction and the amount of time spent on the error correction on the likelihood of the machine part to be in a “defective” or “non-defective” state. The number of machine parts errors were collected from a thermo plastic injection molding machine in a car bumper auto parts manufacturer in Liberec city, Czech Republic from January 2012 to November 2012. A Bayesian method is used for parameter estimation. The results of this study indicate that the machine part error correction and the amount of time spent on the error correction do not improve the machine part status of the individual part, but there is a very strong month-to-month dependence of the machine part states. Using the Mean Absolute Error (MAE) criterion, the performance of the proposed model (MAE = 1.62) and the HMM including machine part error correction only (MAE = 1.68), from our previous study, is not significantly different. However, the proposed model has more advantage in the fact that the machine part state can be explained by both the machine part error correction and the amount of time spent on the error correction.

Keywords: Hidden Markov Model (HMM), Machine Part Errors, Defective and Non-Defective States, Bayesian Method

1. INTRODUCTION

Industrial machine part data, including a number of machine part errors and amounts of time spent on error correction for each part over an entire time period can be collected. These data are useful for the study of effectiveness of machine part error correction and the amount of time for error correction. A model for these data can be constructed and used to estimate the effect of the machine error correction and the amount of time spent on error correction.

In the study of industrial machine parts, it is legitimate to hypothesize an unobserved machine part state that governs individual errors. The normal error rate corresponds to a “non-defective” state and an excess error rate corresponds to “defective” state. The

probability of being in the “defective” or “non-defective” state for a particular part in a given month will differ. It depends on its past state in which that part was in and other possible covariates including specifically error correction and an amount of time spent on error correction. The terms “defective” and “non-defective” are used throughout this study as labels for the two different states. It is important to point out that the two states reflect periods of high and low machine errors which are the surrogates for the concepts of “defective” and “non-defective” respectively. As such there may be periods of frequent machine errors corresponding to a “defective” state being predicted by the models that do not represent a “defective” period in the machine part and vice versa for infrequent machine errors and the “non-defective” state.

Corresponding Author: Pornpit Sirima, Department of Manufacturing Systems and Processes,
Faculty of Mechanical Engineering, Technical University of Liberec, Czech Republic

In this study we propose a model for “defective” and “non-defective” unobserved machine states as a hidden Markov chain including some covariates. It is the extension of our previous study of defective industrial machine parts in which a Hidden Markov Model (HMM) with machine part error correction dummy variables (Sirima and Pokorny, 2014) was proposed. The major research question is whether the machine part error correction and the amount of time spent on error correction reduce the probability of subsequently entering the “defective” state as measured by the machine part errors. We consider fitting the two-state hidden Markov model to the number of machine part errors per month. A Bayesian method is used for parameter estimation. Within the Bayesian method, the MCMC technique is used to sample from the posterior distribution of the parameters (Albert and Chib, 1993). Moreover, Gibbs sampling MCMC is implemented in the existing OpenBUGS software which is adopted in this study. For the structure of this study, the methodology and application, including HMM, Bayesian method, Gibbs sampling, MCMC convergence and an application, are given in section 2. The results are presented in section 3. Section 4 and 5 give some discussion and conclusion, respectively.

2. MATERIALS AND METHODS

2.1. Hidden Markov Models (HMMs)

Let $y = (y_1, \dots, y_T)^T$ be a vector of observed variables, indexed by time $t = 1, \dots, T$. HMMs (Poritz, 1998; Rabiner, 1989) assume that the distribution of each observed data point y_t depends on an unobserved (hidden) variable, denoted as s , that takes on values from 1 to k . The hidden variable $s = (s_1, \dots, s_T)^T$ characterizes the “state” which the generating process is at any time t . HMMs further postulate a Markov Chain for the evolution of the unobserved state variable and, hence, the process for s_t is assumed to depend on the past realizations of y and s only through s_{t-1} Equation 1:

$$P(s_t = j | s_{t-1} = i) = \lambda_{ij} \tag{1}$$

where, λ_{ij} is the generic element of the transition matrix $A = (\lambda_{ij})$, with a vector of stationary probability π satisfying $\pi^T A = \pi^T$ **Fig. 1** illustrates the dependency structure in a HMM, showing that each observation y_t is conditionally independent of all other unobserved and

observed data, given s_t . Applications of HMMs appear in various fields. Dorj *et al.* (2013), as an example, proposed a data-driven approach for anomaly detection in electronic systems based on a Bayesian Hidden Markov model classification technique. Shi and Sun (2012) studied on the HMM model based on system calls anomaly detection in order to improve the detection accuracy. Conesa *et al.* (2011) introduced a framework of models for the early detection of the onset of an influenza epidemic. The process of the observed cases was modelled via a Bayesian Hierarchical Poisson model. Wall and Li (2009; MacDonald and Zucchini, 1997) used Poisson HMM for count data. Nan *et al.* (2008) proposed an application of multiScale hidden Markov modeling wavelet coefficients to Functional Magnetic Resonance Images (fMRI) activation detection Bayesian models. Amutha and Ponnaivaikko (2009) presented a practical target tracking system based on the Hidden Markov Model in a distributed signal processing framework. Hongkong and Pattanasethanon (2013) proposed a HMM for predicting global and diffuse solar radiation on the horizontal plane.

2.2. Bayesian Models

Suppose y is a vector of observations, $y = (y_1, \dots, y_m)$ and θ is a vector of parameters, $\theta = (\theta_1, \dots, \theta_k)$ that are not observable. For Bayesian models (Congdon, 2010), Let $f(y|\theta)$ represent the probability density function of y given θ which is equivalent to the likelihood function and $\pi(\theta)$ is a prior for θ . Then, the posterior probability density function of θ is given by:

$$p(\theta | y) = \frac{f(y|\theta)\pi(\theta)}{\int f(y|\theta)\pi(\theta)d\theta} \tag{2}$$

Since $\int f(y|\theta)\pi(\theta)d\theta$ is a normalizing constant, Equation 2 can be expressed as Equation 3:

$$p(\theta | y) \propto f(y|\theta)\pi(\theta) \tag{3}$$

The goal of Bayesian inference is to get the posterior. In particular, some numerical summaries may be obtained from the posteriors. For example, to keep things simple, a Bayesian point estimator for a univariate θ is often obtained as the posterior mean:

$$E(\theta | y) = \int \theta p(\theta | y) d\theta = \int \theta f(y|\theta)\pi(\theta) d\theta \tag{4}$$

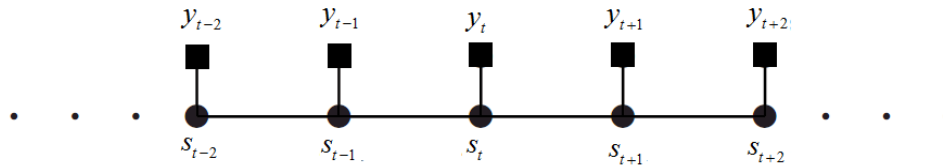


Fig. 1. Dependency structure in a HMM

Markov Chain Monte Carlo (MCMC) methods are proposed to handle the computation in Equation 4. Tongkhaw and Kantanantha (2013) used Bayesian models for time series with covariates, trend, seasonality, autoregression and outliers. Lekdee and Ingsrisawang (2013) used a hierarchical Bayesian method in generalized linear mixed models with spatial random effects for spatio-temporal data.

2.3. Gibbs Sampling

The Gibbs sampling (Geman and Geman, 1984), an MCMC method, decomposes the joint posterior distribution into full conditional distributions for each parameter in the model and then sample from them. The sampler can be efficient when the parameters are not highly dependent on each other and the full conditional distributions are easy to sample from. It does not require an instrumental proposal distribution as Metropolis methods do. However, while deriving the conditional distributions can be relatively easy, it is not always possible to find an efficient way to sample from these conditional distributions.

Suppose $\theta = (\theta_1, \dots, \theta_k)^T$ is the parameter vector, $f(y|\theta)$ is the likelihood and $\pi(\theta)$ is the prior distribution. The full posterior conditional distribution of $p(\theta_i|\theta_j, i \neq j, y)$ is proportional to the joint posterior density; that is, $p(\theta_i|\theta_j, i \neq j, y) \propto f(y|\theta)\pi(\theta)$. For instance, the one-dimensional conditional distribution of θ_1 given $\theta_j = \theta_j^*, 2 \leq j \leq k$, is computed as Equation 5:

$$p(\theta_1 | \theta_j = \theta_j^*, 2 \leq j \leq k, y) \propto \int \mu f(y | \theta = (\theta_1, \theta_2^*, \dots, \theta_k^*)^T) \pi(\theta = (\theta_1, \theta_2^*, \dots, \theta_k^*)^T) d\theta_2 \dots d\theta_k \quad (5)$$

The Gibbs sampler works as follows:

Set $t = 0$ and choose an arbitrary initial value of $\theta^0 = (\theta_1^0, \dots, \theta_k^0)$. Generate each component of θ as follows:

- draw $\theta_1^{(t+1)}$ from $p(\theta_1 | \theta_2^{(t)}, \dots, \theta_k^{(t)}, y)$
- draw $\theta_2^{(t+1)}$ from $p(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_k^{(t)}, y)$
- ...
- draw $\theta_k^{(t+1)}$ from $p(\theta_k | \theta_1^{(t+1)}, \theta_3^{(t+1)}, \dots, \theta_{k-1}^{(t+1)}, y)$

Set $t = t + 1$. If $t < T$, the number of desired samples, return to step 2. Otherwise, stop. In the MCMC, there are other related processes, called convergence, which are described in the following topics.

2.4. Assessing MCMC Convergence

Simulation-based Bayesian inference requires using simulated draws to summarize the posterior distribution or calculate any relevant quantities of interest. We have to decide whether the Markov chain has reached its stationary, or the desired posterior distribution and to determine the number of iterations to keep after the Markov chain has reached stationarity. Convergence diagnostics help to resolve these issues (Brooks and Roberts, 1997). The common ones are visual analysis via history plots and Kernel density plots.

2.5. An Application

The data were collected from a thermo plastic injection molding machine in a car bumper auto parts manufacturer in Liberec city, Czech Republic. Altogether 27 machine parts were randomly chosen for this study, during the time period from January 2012 to November 2012. The number of machine parts errors were recorded. A Hidden Markov model for the number of machine part errors y_{it} has the following details:

$$y_{it} | s_{it} \overset{iid}{\sim} \text{Pois}(\theta_{it})$$

$$\log(\theta_{it}) = \lambda_0 + \lambda_1 s_{it} \quad (6)$$

$$s_{it} | s_{i(t-1)} \sim \text{Bin}(p, 1)$$

$$\log it p = \beta_0 + \beta_1 s_{i(t-1)} + \beta_2 \text{Correct}_{it} + \beta_3 \text{Time}_{it}$$

Or:

$$\frac{\log p}{\log(1-p)} = \beta_0 + \beta_1 s_{i(t-1)} + \beta_2 \text{Correct}_{it} + \beta_3 \text{Time}_{it} \quad (7)$$

$$s_{i_0} : Bin(\pi_0, 1)$$

$$\pi_0 \sim Beta(1, 1)$$

Note that Pois, Bin and Beta stand for Poisson, Binomial and Beta distribution, respectively. $\lambda_0, \lambda_1, \beta_0, \beta_1, \beta_2$ and β_3 are regression coefficients. θ_{it} can be viewed as the mean of the Poisson. It is determined by the unobserved a machine part state s_{it} . This unobserved machine parts state follows a Markov chain, with transition probability modeled by a logistic regression with the previous machine part state $s_{i(t-1)}$. The parameter π_0 represents the initial probability of being in the “defective” machine state at the first month t_{i1} . The dummy variables $Correct_{it} = 1$ indicates the status of month t for part i as being after error correction (with before correction as the reference group, $Correct_{it} = 0$). Thus, the estimate for the coefficient of $Correct_{it}$ is of primary interest to see if the probability of being in the “defective” state has significantly decreased after correction. $Time_{it}$ are the covariates refer to the amount of time (minute) spent on error correction for machine part i at time t .

The number of machine part errors y_{it} in a particular month t , is governed by the two state hidden variable s_{it} . More specifically, y_{it} comes from a two state Poisson distribution where the two different means of the Poisson distribution correspond to the two different values of the hidden variable s_{it} which in turn depends on the previous state $s_{i(t-1)}$. To make the unobserved states identifiable, we assume that the lower mean corresponds to $s_{it} = 0$ and the higher mean corresponds to $s_{it} = 1$, which is operationalized by constraining λ_1 to be larger than zero. Thus $s_{it} = 0$ corresponds to the “non-defective” state and $s_{it} = 1$ corresponds to the “defective” state.

2.6. Parameter Estimation

The MCMC Gibbs sampling for parameter estimation are performed in OpenBUGS software. The joint posterior is decomposed into the full conditional posterior distribution with respect to each parameter and the Gibbs sampling is used. Once the chain converges, the empirical joint posterior distribution for all the parameters can be used to obtain the posterior mean and the 2.5 and 97.5% quantiles can be used as the credible interval for all the parameters. The non-informative priors were chosen. Assume $N(0, \tau_{\lambda_0}^2), N(0, \tau_{\beta_0}^2), N(0, \tau_{\beta_1}^2), N(0, \tau_{\beta_2}^2)$ and $N(0, \tau_{\beta_3}^2)$ prior for $\lambda_0, \beta_0, \beta_1, \beta_2$ and β_3 , respectively and $N(0, \tau_{\lambda_1}^2)$ with positive value restriction was used for λ_1 . Similarly, assume Inverse Gamma (IG)

for $\tau_{\lambda_0}^2, \tau_{\lambda_1}^2, \tau_{\beta_0}^2, \tau_{\beta_1}^2, \tau_{\beta_2}^2$ and $\tau_{\beta_3}^2$. The visual analysis, history plots and Kernel density plots are used for the MCMC convergence diagnostics. To evaluate the model performance, the proposed model is compared with our proposed previous model (Sirima and Pokorny, 2014), a HMM with a machine part error correction dummy variables, using Mean Absolute Error (MAE) as a criterion. The HMM without covariates is the same as Equation 6 and 7 except no $\beta_3 Time_{it}$ term.

3. RESULTS

We performed 25,000 MCMC iterations with burn-in of 5,000. The visual analysis is used for MCMC convergence diagnostics. The history plots in **Fig. 2-7** show no trend and snake around the mean and the Kernel density plots in **Fig. 8-13** look more bell-shaped or not multi-modal. These indicate each parameter is converged to a stationary density.

Table 1 shows the results of the HMM fit to the machine part error counts per month. The estimate of β_2 (0.7002) implies that the odds of transitioning to or remaining in the “defective” state in any given month after correction is $\exp(0.7002) = 2.0142$ of what it was before correction. This provides evidence in favor of the machine part error correction not improving the machine part status of individual part. Similarly, although the amount of time spent on error correction increases, the odds of transitioning to or remaining in the “defective” state in any given month is $\exp(0.7432) = 2.1027$ of what it was before correction. This also provides evidence that the increase in the amount of time spent on error correction does not improve the machine part status of individual part.

In addition to this main finding, the results from the model also characterize a very strong month-to-month dependence of machine part states ($\beta_1 = 3.6220$) where the odds of remaining in the “defective” state in the current month if an machine part was in the “defective” state the last month is estimated to be $\exp(3.6220) = 37.4123$ times the odds of newly transitioning to the “defective” state if an individual part was “non-defective” in the previous month. Using the MAE criterion for model comparison, the performance of the proposed Model (MAE = 1.62) and of our previous model (Sirima and Pokorny, 2014), a HMM with Machine part error correction dummy variables (MAE = 1.68), are not significantly different. They both have good performance.

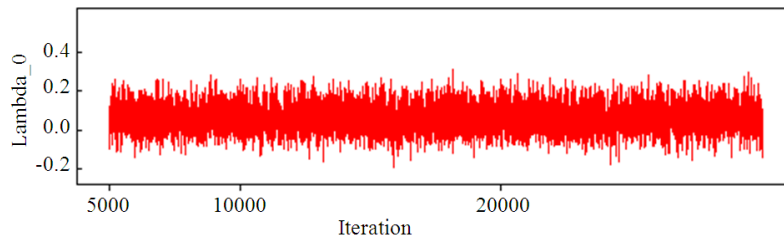


Fig. 2. History plot of λ_0 indicating convergence

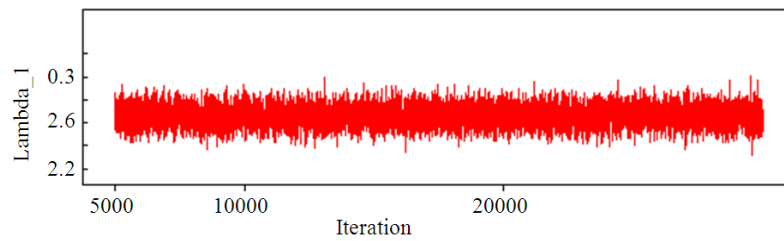


Fig. 3. History plot of λ_1 indicating convergence

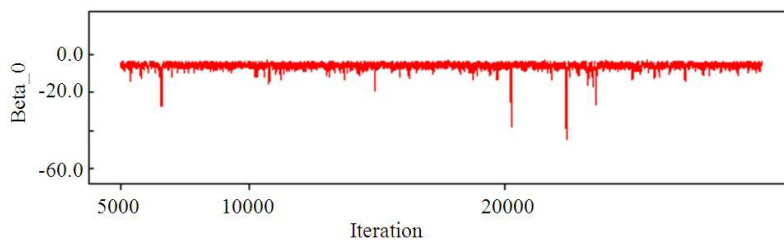


Fig. 4. History plot of β_0 indicating convergence

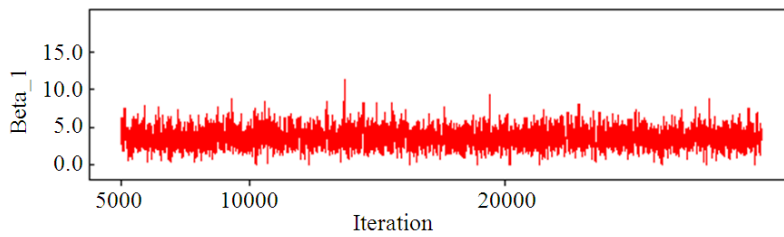


Fig. 5. History plot of β_1 indicating convergence

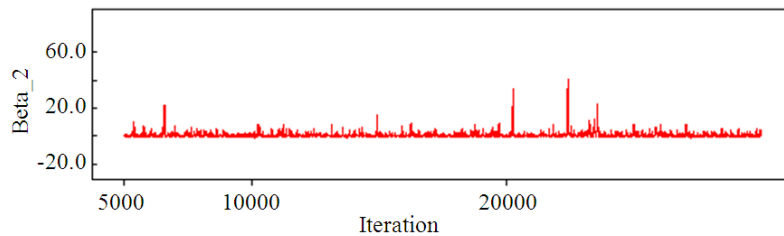


Fig. 6. History plot of β_2 indicating convergence

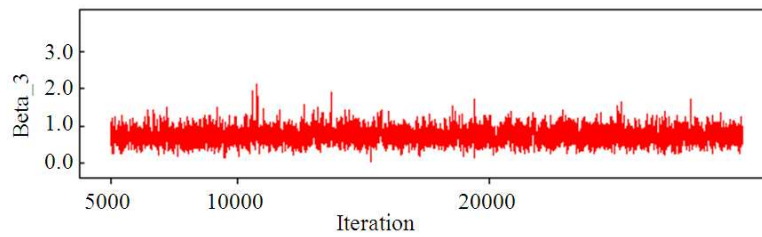


Fig. 7. History plot of β_3 indicating convergence

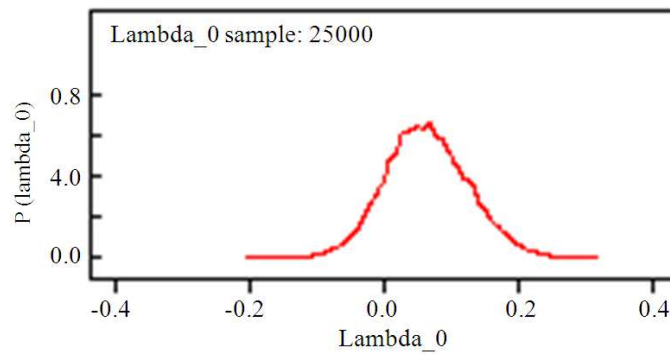


Fig. 8. Kernel density plot of λ_0 indicating convergence

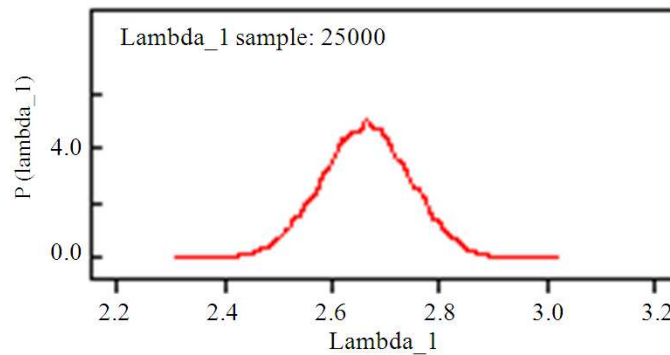


Fig. 9. Kernel density plot of λ_1 indicating convergence

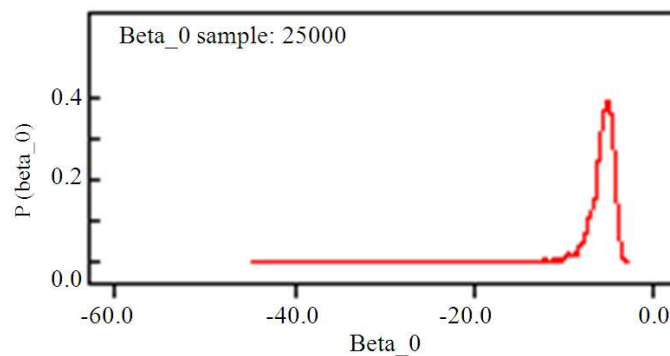


Fig. 10. Kernel density plot of β_0 indicating convergence

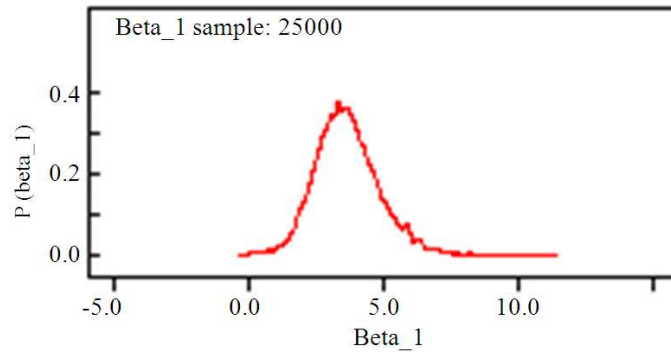


Fig. 11. Kernel density plot of β_1 indicating convergence

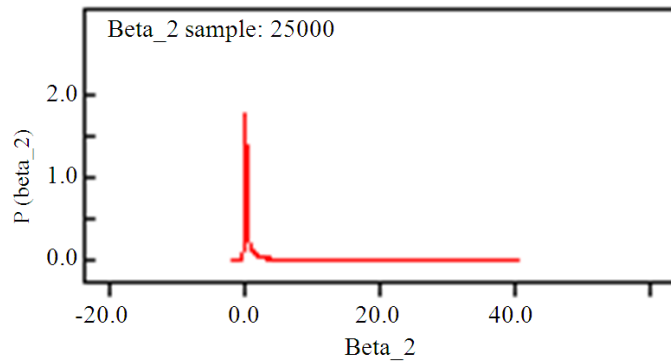


Fig. 12. Kernel density plot of β_2 indicating convergence

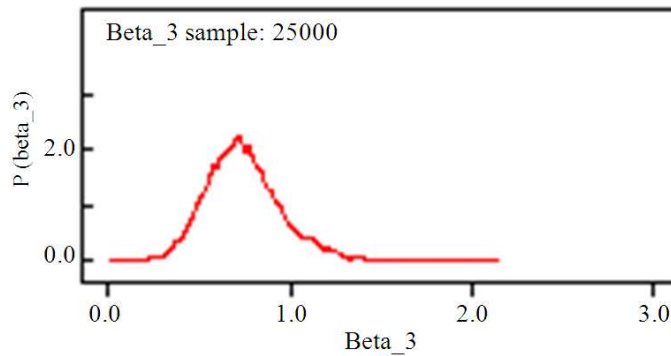


Fig. 13. Kernel density plot of β_3 indicating convergence

Table 1. Parameter estimates from the HMM

Parameter	Mean	SD	95% Credible	Interval
λ_0	0.0627	0.0620	-0.0543	0.1887
λ_1	2.6630	0.0820	2.5020	2.8220
β_0	-5.8570	2.0980	-10.0100	-3.9070
β_1	3.6220	1.2180	1.4660	6.2940
β_2 (Correct)	0.7002	2.0250	-0.2725	4.8950
β_3 (Time)	0.7432	0.2074	0.3868	1.2050

4. DISCUSSION

A HMM with covariates for machine part errors is proposed. We assume there are unobserved machine part states that govern the machinery care utilization of a particular machine part and the machine part state is governed only by the frequency of errors. The main purpose of the HMM is to model changing machine part states over time not necessarily modeling the changing number of machine part errors. In a HMM, the observed machine errors are only a surrogate for “machine parts status”. Measurement errors is allowed between the observed machine errors and the underlying machine state. The HMM with covariates explains the data very well (MAE = 1.62). Even though the performance of the proposed model is not significantly different from our previous model (Sirima and Pokorny, 2014). Nevertheless, the proposed model has more advantage in the fact that the machine part state can be explained by the amount of time spent on error correction.

5. CONCLUSION

This study proposes a two-state hidden Markov model with covariates in order to describe the effect of machine part error correction and the amount of time spent on error correction on the likelihood of the machine parts to be in a “defective” or “non-defective” state. A Bayesian method is used for parameter estimation. The result shows that the machine part error correction and the amount of time spent on error correction do not improve the machine part status of individual part and there is a very strong month-to-month dependence of machine part states. Using the MAE, the performance of the proposed model is not significantly different from of our previous model (Sirima and Pokorny, 2014). Nonetheless, the proposed model is more advantageous in the fact that the machine part state can be explained by the amount of time spent on error correction. For further study, the proposed model can be applied to other similar problems and can be extended to multivariate Poisson count data.

6. ACKNOWLEDGMENT

We gratefully thank the Department of Manufacturing Systems and Processes, the Faculty of Mechanical Engineering, Technical University of Liberec, Czech Republic for technical support.

7. REFERENCES

- Albert, J.H. and S. Chib, 1993. Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts. *J. Bus. Econ. Stat.*, 11: 1-15. DOI:10.1080/07350015.1993.10509929
- Amutha, B. and M. Ponnaivaikko, 2009. Energy efficient hidden markov model based target tracking mechanism in wireless sensor networks. *J. Comput. Sci.*, 5: 1082-1090. DOI: 10.3844/jcssp.2009.1082.1090
- Brooks, S.P. and G.O. Roberts, 1997. Assessing convergence of markov chain monte carlo algorithms. *Stat. Comput.*, 8: 319-335. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.2642>
- Conesa, D., M.A. Martinez-Beneito, R. Amorós and A. Lopez-Quilez, 2011. Bayesian hierarchical poisson models with a hidden Markov structure for the detection of influenza epidemic outbreaks. *Stat. Methods Med.*, 25: 1-38, DOI: 10.1177/0962280211414853
- Congdon, P.D., 2010. *Applied Bayesian Hierarchical Methods*. 1st Edn., CRC Press, Boca Raton, ISBN-10: 1584887214, pp: 604.
- Dorj, E., C. Chen and M. Pecht, 2013. A Bayesian hidden Markov model-based approach for anomaly detection in electronic systems. *Proceedings of the IEEE Aerospace Conference*, Mar. 2-9, IEEE Xplore Press, Big Sky, MT, pp: 1-10. DOI: 10.1109/AERO.2013.6497204
- Geman, S. and D. Geman, 1984. Stochastic relaxation Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelli.*, 6: 721-741. DOI: 10.1109/TPAMI.1984.4767596
- Hongkong, S. and S. Pattanasethanon, 2013. Observation modeling for continuous predicting global and diffuse solar radiation on horizontal surface. *Am. J. Environ. Sci.*, 9: 201-209. DOI: 10.3844/ajessp.2013.201.209
- Lekdee, K. and L. Ingsrisawang, 2013. Generalized linear mixed models with spatial random effects for spatio-temporal data: An application to dengue fever mapping. *J. Math. Stat.*, 9: 137-143. DOI: 10.3844/jmssp.2013.137.143
- MacDonald, I.L. and W. Zucchini, 1997. *Hidden Markov and Other Models for Discrete-Valued Time Series*. 1st Edn, CRC Press, ISBN-10: 0412558505, pp: 256.

- Nan, F., Y. Wang and X. Ma, 2008. Application of multiscale hidden Markov modeling wavelet coefficients to FMRI activation detection. *J. Math. Stat.*, 4: 255-263. DOI: 10.3844/jmssp.2008.255.263
- Poritz, A.M., 1988. Hidden Markov models: A guided tour. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Apr. 11-14, IEEE Xplore Press, New York, pp: 7-13. DOI: 10.1109/ICASSP.1988.196495
- Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77: 257-286. DOI: 10.1109/5.18626
- Shi, S. and M. Sun, 2012. Study on HMM based anomaly intrusion detection using system calls. Proceedings of the 2nd International Conference on Electronic and Mechanical Engineering and Information Technology, (EIT' 12), Atlantis Press, Paris, France, pp: 139-144. DOI: 10.2991/emeit.2012.27
- Sirima, P. and P. Pokorny, 2014. Hidden Markov models for analysis of defective industrial machine parts. Proceedings of the International Multi Conference of Engineers and Computer Scientists, Mar. 12-14, IEEE Xplore Press, Hong Kong, pp:1100-1104.
- Tongkhaw, P. and N. Kantanatha, 2013. Bayesian models for time series with covariates, trend, seasonality, autoregression and outliers. *J. Comput. Sci.*, 9: 291-298. DOI: 10.3844/jcssp.2013.291.298
- Wall, M.M. and R. Li, 2009. Multiple indicator hidden Markov model with an application to medical utilization data. *Stat. Med.*, 28: 293-310. DOI: 10.1002/sim.3463