Modelling of fault in RPM using the GLARMA and INGARCH model

Ji-Yong Kim, Hee-Young Kim[✉], Daihee Park and Yongwha Chung

According to the of time series of faults in railway point machines (RPMs), forecasting approach based on the generalised linear autoregressive moving average (GLARMA) models and the integer-valued generalised autoregressive conditional heteroscedastic (INGARCH) models are presented. The conditional distribution of observed fault counts of given previous faults and weather conditions are assumed to be Poisson or negative binomial distributions. The forecasting future fault counts of RPM are obtained by one-step-ahead forecasts and the performance evaluation shows that the GLARMA method performs better than the traditional autoregressive moving-average (ARMA) model and generalised linear model (GLM).

Introduction: The railway occupies an important position in transportation due to its high speed and transportation capability. As a consequence, the maintenance of the constituent components of railways is a crucial task, and the railway point machine (RPM), which uses an electric motor to move a switch blade from its current position to the opposite position to offer different routes to trains, is the most critical device for railway infrastructure.

To date, most fault diagnosis methods of RPM are classification methods using electric current. In [1], classification methods using a support vector machine (SVM) with discrete wavelet transform were proposed, and in [2], a one-class SVM with the similarity measure of edit distance with the real penalty was reported. In addition, Kim *et al.* [3] proposed to manage the variation in duration of RPM movement using dynamic time warping. Recently, Sa *et al.* [4] focused on the aging effect of the replacement of RPMs using electric current, and Lee *et al.* [5] presented a method for identifying faults using of sound signals.

However, there are few statistical time series modelling approaches for fault detection of RPM. A method for detecting failures was presented in [6], by comparing the expected electric current signals with those actually observed, and the expected shape is modelled as a forecast of a time series model, vector autoregressive moving average (VARMA), and harmonic regression, based on fault-free data. In addition, Pedregal *et al.* [7] used a state space model and a harmonic regression model for the required forecasts.

Despite these preliminary efforts mentioned above, all studies about RPMs were found to be insufficient to address the characteristics of our data, which was collected sequentially over time, daily, with 40 different kinds of faults, so a key feature is that each series is a time series of correlated count data. In [6, 7], VARMA and state space models assume a Gaussian continuous-valued response time series. However, it is well known that the VARMA model is restricted by linearity; although this model is suited to capture short range dependence, it is not well suited to deal with non-linearity, not appropriate to count time series, which result in poor prediction accuracy.

Over the last few years, various modelling approaches for time series of counts have been proposed, taking into account observations that are non-negative integers and capturing suitable dependence among observations. In this Letter, we compare seven models for analysing and forecasting 40 different types of fault count data, measured daily for the period of February 2015 to October 2015 and obtained from RPMs in 15 stations in three regions of Seoul, South Korea.

Models: Let $\{Y_t : t = 1, \ldots, T\}$ be the available response observations. The generalised linear autoregressive moving average (GLARMA) model in [8] is a class of non-Gaussian non-linear state space models, in which the response process depends on regressors and past values of the observed process. There is a r-dimensional covariate vector, say $\boldsymbol{x}_t = (x_{t,1}, \ldots, x_{t,r})'$, for $t = 1, \ldots, T$. Let F_t be the σ-field generated by $\{Y_s : s < t-1,\ \boldsymbol{x}_s : s \le t\}$, i.e. it denotes the past available information on response and the past and current information on regressors. The conditional distribution of Y_t given F_t is assumed to be of the exponential family, with density

$$f(y_t|F_t) = \exp\{y_t W_t - a_t \mathrm{b}(W_t) + c_t\} \tag{1}$$

where W_t is the canonical parameter (state variable), and a_t and c_t are sequences of constants depending on the y_t.

The response distribution in (1) covers popular discrete distributions, such as Poisson, and binomial distributions. The negative binomial (NB) distribution is also available with an additional parameter, α. To extend the Gaussian ARMA time-series model to non-Gaussian time series data, the GLARMA(p, q) model allows that the state variable W_t includes autoregressive moving average terms as follows:

$$W_t = x_t\beta + O_t + Z_t,$$
$$Z_t = \sum_{j=1}^{\infty} r_j e_{t-j}, \tag{2}$$

where O_t is an additional offset term and r_j are given as coefficient in the power series, $\sum_{j=1}^{\infty} r_j \zeta^j = \theta(\zeta)/\phi(\zeta) - 1$, with $\phi(\zeta) = 1 - \phi_1\zeta - \cdots - \phi_p\zeta^p$ and $\theta(\zeta) = 1 + \theta_1\zeta + \cdots + \theta_q\zeta^q$ being autoregressive and moving average polynomials, respectively. If Z_t were not involved in (2), then it would be a generalised linear model (GLM), in which observations Y_t are independent. The GLM framework has been commonly used to detect faults [9]. However, GLM is not tailored for time-dependent response variables, yielding invalid model fits.

While GLARMA models provide a natural extension of the GLM to include serial dependence terms, another useful approach to handle time series of counts is to imitate the classical GARCH model, the Poisson INGARCH (p,q) model, as in [10]. It is defined as follows:

$$Y_t|\,F_{t-1} \sim \mathrm{Poisson}(\lambda_t),$$
$$\lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j} \tag{3}$$

where $\alpha_0 > 0$, $\alpha_i \ge 0$, $\beta_j \ge 0$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, $p \ge 1$, $q \ge 1$ and F_{t-1} is the σ-field generated by $\{Y_{t-1}, Y_{t-2}, \ldots\}$. Recently, Zhu [11] proposes a negative INGARCH (p, q) model that can deal with both overdispersion and potential extreme observations simultaneously.

Experiment and results: Daily counts of 40 different types of faults are recorded for the period of February 2015 to October 2015, obtained by RPMs in 15 stations in three regions of Seoul, South Korea.

Fig. 1 shows an RPM and monitoring system. Typical point machines widely installed in South Korea are NS-type, NS-AM, and MJ81 electric point machines. The condition monitoring system equipped with an alarm, manufactured by the Sehwa Company and widely used in South Korea, contains alarms for 40 different faults; e.g. AS current, the failure to position movement, the 1st–4th of adherence detection left indication signal input, and indication voltage input and so on.

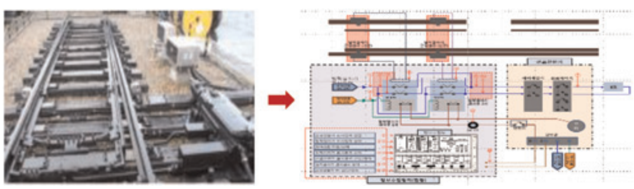

Fig. 1 *RPM and monitoring system*

Among 40-faults series, the fault series of *AS current*, and *indication voltage input* show dependence in the process. Also, the series of the *sum of all faults* shows its dependence. Fig. 2 depicts the original series and Fig. 3 displays the sample autocorrelation function (SACF) and sample partial autocorrelation function (SPACF). Therefore, these provide strong evidence of the serial dependence of each series.

Table 1 gives the Akaike information criterion (AIC) and mean absolute percentage error (MAPE) of seven models for three series: the *sum of all faults* (Y_1), *AS current* (Y_2), and *indication voltage* (Y_3). For GLM and GLARMA, we use the temperature as a regressor x_t in (2). In modelling series Y_1, the logarithm of the number of switching is also included as an offset term O_t in (2), so this provides a model for the fault per unit the number of switching.

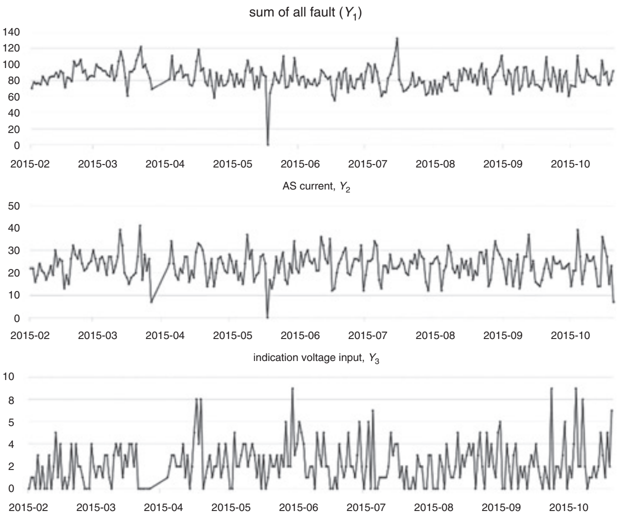

Fig. 2 *Time series plot for sum of all faults (Y_1), fault of AS current (Y_2) and fault of indication voltage input (Y_3) from February to October 2015*

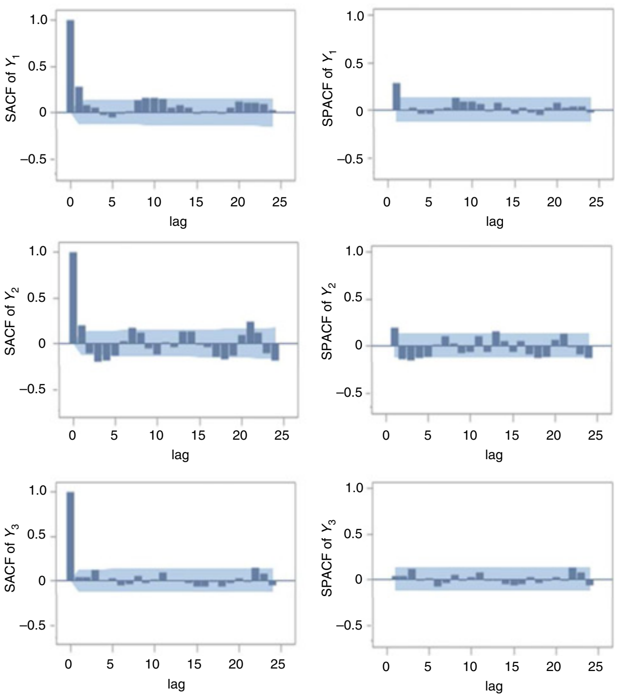

Fig. 3 *SACF and SPACF plots of three series*

Table 1a: Performance comparison

Model	AIC					
	Sum of all faults (Y_1)		AS current (Y_2)		Indication voltage input (Y_3)	
	Poisson	NB	Poisson	NB	Poisson	NB
GLM	2279.5	2138.8	1666.2	1645.6	1019.5	986.1
GLARMA	2208.4	2109.7	1642.1	1628.9	1015.9	983.5
INGARCH	2190.1	2101.8	1652.6	1636.9	1020.5	987.5
ARMA	normal		normal		normal	
	2181.3		1705.1		1081.6	

Table 1b:

Model	MAPE					
	Poisson	NB	Poisson	NB	Poisson	NB
GLM	0.1174	0.1174	0.2165	0.2165	0.3849	0.3853
GLARMA	0.2165	0.1254	0.2006	0.2005	0.3823	0.3851
INGARCH	0.1157	0.1157	0.2081	0.2081	0.3853	0.3853
ARMA	normal		normal		normal	
	0.1251		0.2068		0.3829	

The GLARMA and ARMA modelling procedures consist of three iterative steps: model identification, parameter estimation, and diagnostic checking. These three steps are repeated until an adequate model is identified.

The AIC is a well-known information criterion, based on likelihood

$$\text{AIC} = \frac{2}{T}\ln(\textit{likelihood}) + \frac{2}{T}\left(\textit{number of parameters}\right), \qquad (4)$$

where the likelihood function is evaluated at the maximum likelihood estimates, and T is the sample size. MAPE is commonly used statistic to measure the performance of point forecasts and is calculated using one-step-ahead forecasts. It can be observed in almost all cases that AIC and MAPE obtained with GLARMA with NB distribution are considerably lower than those calculated with other models.

Conclusion: This Letter analyses in-field data of RPMs in 15 stations of three regions in Seoul, South Korea. We forecast the future fault count of RPM, based on GLM, GLARMA, INGARCH, and ARMA models. The GLARMA with NB distribution outperforms the other models.

Acknowledgment: This study was supported by the project: Small & Medium Business Administration under Project S2312692 'Technological Innovation Development Business' for the innovative company in the year 2015.

Submitted: *25 September 2017* E-first: *18 January 2018*
doi: 10.1049/el.2017.3398
One or more of the Figures in this Letter are available in colour online.

Ji-Yong Kim and Hee-Young Kim (*Division of Economics and Statistics, National Statistics, Korea University, Sejong, Korea*)

✉ E-mail: starlike.kim@gmail.com

Daihee Park and Yongwha Chung (*Department of Computer Convergence Software, Korea University, Sejong, Korea*)

References

1 Asada, T., Roberts, C., and Koseki, T.: 'An algorithm for improved performance of railway condition monitoring equipment: alternating-current point machine case study', *Transp. Res. C, Emerg. Technol.*, 2013, **30**, pp. 81–92
2 Vileiniskis, M., Remenyte-Prescott, M., and Rama, D.: 'A fault detection method for railway point systems', *J. Rail Rapid Transit*, 2016, **230**, pp. 852–865
3 Kim, H., Sa, J., Chung, Y., *et al.*: 'Fault diagnosis of railway point machines using dynamic time warping', *Electron. Lett.*, 2016, **52**, (10), pp. 818–819
4 Sa, J., Choi, Y., Chung, Y., *et al.*: 'Replacement condition detection of railway point machines using an electric current sensor', *Sensors*, 2017, **17**, p. 263
5 Lee, J., Choi, H., Park, D., *et al.*: 'Fault detection and diagnosis of railway point machines by sound analysis', *Sensors*, 2016, **16**, p. 549
6 García, F.P., Pedregal, D.J., and Roberts, C.: 'Time series methods applied to failure prediction and detection', *Reliab. Eng. Syst. Safety*, 2010, **95**, pp. 698–703
7 Pedregal, D.J., García, F.P., and Roberts, C.: 'An algorithmic approach for maintenance management based on advanced state space systems and harmonic regression', *Annals Oper. Res.*, 2009, **166**, pp. 109–124
8 Dunsmuir, W.T.M., and Scott, D.J.: 'The Glarma package for observation-driven time series regression of counts', *J. Stat. Softw.*, 2015, **67**, (7), pp. 1–36
9 Wang, J., Wu, Z., Shu, Y., *et al.*: 'An imperfect software debugging model considering log-logistic distribution fault content function', *J. Syst. Softw.*, 2015, **100**, pp. 167–181
10 Ferland, R., Latour, A., and Oraichi, D.: 'Integer-valued GARCH process', *J. Time Ser. Anal.*, 2006, **27**, pp. 923–942
11 Zhu, F.: 'A negative binomial integer-valued GARCH model', *J. Time Ser. Anal.*, 2011, **32**, pp. 54–67