CETRA 2010 17–18 May 2010, Opatija, Croatia
First International Conference on Road and Rail Infrastructure

FORCASTING MODELS OF ACCIDENT RISKS ON THE RAILWAY

Boriss Misnevs, Alla Melikyan
Department of Computer Science, Transport and Telecommunication Institute, Riga, Latvia

Abstract

A forecasting model made on the basis of auto regression model combined with the moving average (ARIMA) used for making predictions of the railway accident number on the railroad is described in the article. The results of forecasting railway accident number on the Latvian Railroad and the real number of accidents during the forecasting period are given. The reviled weather factors that influence the incidence of railway accidents are named. In accordance with these results the deviations from the forecasting number of accidents and the ones observed in reality are explained.

Keywords: weather conditions, forecast, security, statistics, railway

1 Introduction

It is important to identify the factors that influence greatly on the number of railway accidents for the safety measures improvement and elaboration. Taking into the consideration the fact that the railroad transportation is the most reliable means of transport which is also resistant to the weather conditions as well as having the restricted traffic, the authors of this article did not have the intention to explain completely the reasons of railway accidents in this transport sphere. Being aware of the fact that the malfunction in railroad transport work appears mostly due to the subjective circumstances that is justified by the statistic data, the authors allowed themselves to monitor and evaluate the influence of objective weather conditions on the number of accidents though not hoping for the full and exhaustive description.

2 Time-series analysis

The aim of the study is to construct a forecasting model in order to predict the future values of railway accidents number basing on analysis of time series of recorded accidents number. The author collected and systematized data on emergencies which occured on the Latvian railways for the period since 2001 till 2008 [1]. As a result time series of railways accidents containing information on monthly number of accidents since 2001 till 2008 which took place during that period was built Fig.1. The kind of time series, their correlation and the Fourier series confirmed that the statistics of railways accidents has definite seasonal component. Thus, harmonic with the period of a year dominates on the periodogram Fig.2. The idea to forecast accidents taking into account the seasonal part of the series appeared.

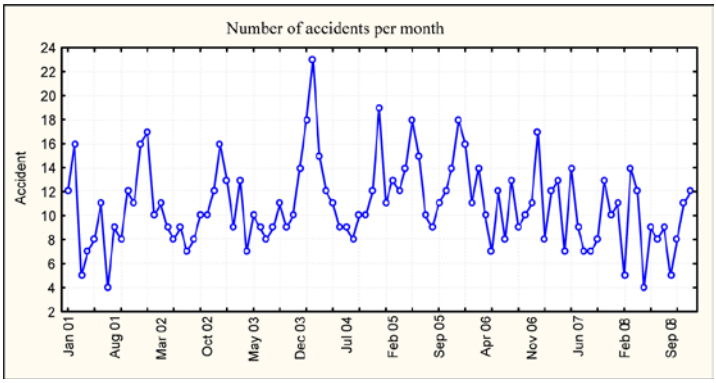

Figure 1 The time series of accidents

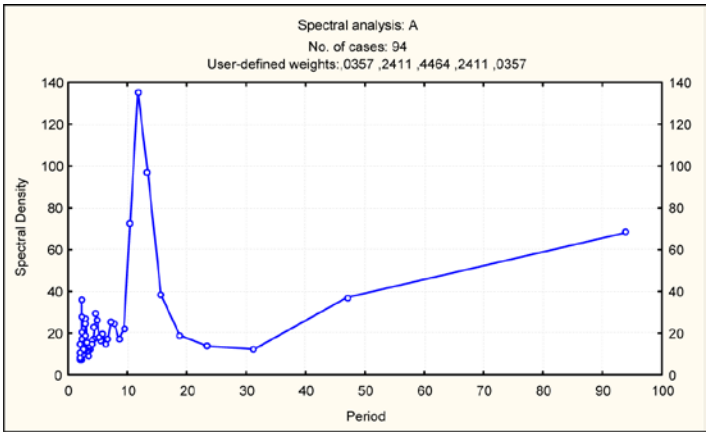

Figure 2 Periodogram of the time series of accidents

3 Method of autoregressive and pre-integrated moving average

ARIMA – the method of "Auto-Regressive Integrated Moving Average" developed by George Box and Gwilym Jenkins in 1976 was chosen as a method of forecasting [2]. The ARIMA model itself is a joint of two classic models - autoregressive model and moving average model. Autoregressive model suggests explaining the current value of x_t observed variable X its values in the preceding moments of time x_{t-k}. More remote in time observations have less weight in the model, set by the parameters of autoregressive φ_k.

Mathematically this model is described by the equation

$$x_t = \xi + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \ldots + \varepsilon_t \qquad (1)$$

where

ξ –constant (average value of an observable variable),

φ_k – parameters of autoregressive,

ε_t – random perturbation.

Parameters φ_k are selected so that the sum of squared deviations of calculated values Eq. (3) from observed x_t is a minimum.

The moving average model explains the current value of x_t observed variable x by values of random perturbation ε_{t-k} in the preceding moments of time

$$x_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots + \varepsilon_t \quad (2)$$

where

μ – constant (average value of an observable variable),

θ_k –parameters of moving average,

ε_t – random perturbation.

As in the autoregressive model, early observations have less weight than later, and in the construction of smoothed series $\tilde{x}_t$ as an estimate value of random perturbations ε_t deviations of observed values x_t from the calculation

$$\tilde{x}_t = \mu - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} \quad (3)$$

where $e_t = x_t - \tilde{x}_t$ are used

Parameters θ_k are selected so that the sum of squared of these deviations is minimum.

These models give adequate results during the work with stationary series, i.e. with the series which have constant and are unchanged over time sample variance and autocorrelation. The series where seasonal component is presented as a rule is required pre -differentiate on seasonal lag as well as one or two short lags. The series suitable for using of these models should have the autocorrelogram on all lags in white noise.

Thus, the seasonal model ARIMA has six parameters:

p – the number of autoregressive parameters and respectively the past observations used to explain the current value of the observed values;

q – the number of parameters of moving average, i.e. number of past mistakes, which is accounted in smoothing;

d – the number of lags where the original series is differentiated on to bring it to the stationary series;

ps – the number of seasonal parameters of autoregressive;

qs – the number of seasonal parameters of moving average;

ds – number of seasonal lags which series is differentiated.

Compactly, it is written down as ARIMA (p, d, q) (ps, ds, qs).

The first problem to be solved by using the model is to determine the parameters d and ds, sufficient to obtain stationary series y_t from initial x_t.

In our case, it was enough to pre-differentiate series twice - with the off-season lag of 1 month (d = 1) and the seasonal lag of 12 months (ds = 1). The graph and autocorrelogram of obtained series looked quite satisfactory. The second, more difficult, problem is selection of the remaining parameters of the model. The several forecasting models with different parameters p, q, ps and qs were constructed and compared the accuracy of their forecasting with real data. Relative prediction error was used as a criterion of the adequacy of the model

$$\delta = \frac{\sqrt{\frac{1}{12}\sum_{t=1}^{12}(\hat{x}_t - x_t)^2}}{\bar{x}} \quad (4)$$

where $\hat{x}_t$ - forecasting value of the number of accidents per month t, x_t – real number of accidents recorded per month t,

$\bar{x} = \frac{1}{12}\sum_{t=1}^{12} x_t$ - average number of accidents per month.

The result of calculations presented in table 1. The 2007–2008 forecast and real data for this period were used for the calculation. Figure 3 and 4 presents the result of the comparison of forecasted and real numbers of accidents for two models under consideration during the year 2008 and 2007.

Table 1 The result of calculations

Model for 2008	ARIMA (0,1,1) (0,1,1)	ARIMA (1,1,0) (1,1,0)	ARIMA (1,1,1) (1,1,1)	ARIMA (1,1,0) (0,1,1)	ARIMA (0,1,1) (1,1,0)
δ	0,31	0,36	0,29	0,33	0,27

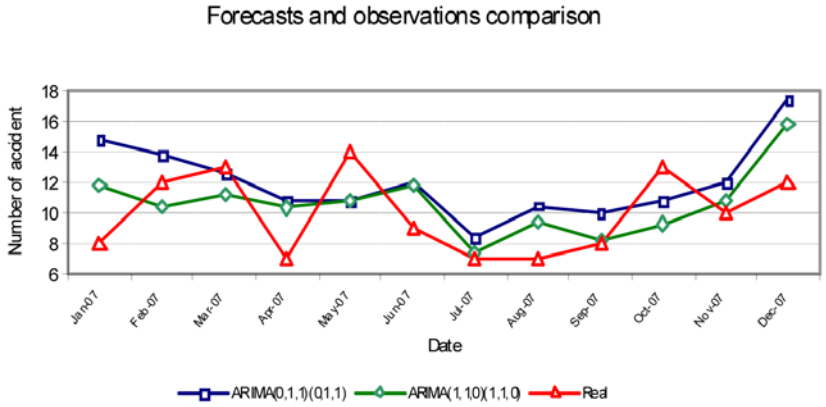

Figure 3 Comparison of 2007 forecast with observations

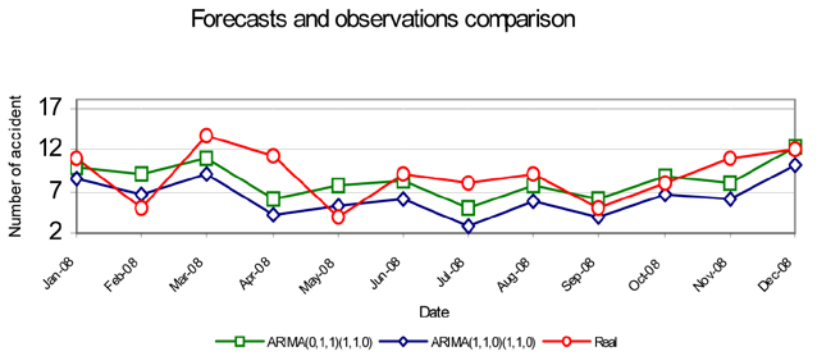

Figure 4 Comparison of 2008 forecast with observations

The lowest forecast error for the year 2008 0,27 where as for the year 2007 0,26 was obtained for the model ARIMA (0,1,1)(1,1,0), so it was decided to use one nonseasonal summands of moving average (q=1) and one seasonal autoregressive summand (ps=1).

As it is seen the number of emergences registered in 2007 is far more less than it was forecasted, and the peak of accidents happened in May. However these deviations from the forecast are explained within the research conducted. While considering the reason of the seasonal emergency fluctuations the information on air temperature that is the evident nature factor tightly connected with the season was collected and dependence between the number of accidents and temperature monthly pattern was discovered. The primary analysis of the data was carried out in 2007 basing on the 2005-2006 statistics and the results of which were published in the article"Statistical research of weather conditions influence on the railway transport accident rate" [3]. As it was found out during the research the average air temperature as well as different abnormalities (seasonal temperature deviations, sharp temperature

jumps, rates of temperature change during the short period of time) influence on emergency situations on the railway. It has turned out that accident risk during the cold time of the year is higher than during the hotter period and it also grows when the sharp temperature jumps occur during the short period of time.

The fast temperature rise has also turned to be the factor that increases the number of emergency situations. Relatively small amount of emergencies that took place in January is explained by the high and quite stable temperature that was set this month as compared with a norm. But at the same time the peak of accidents which occurred in May 2007 coincides with the big amount of the sharp temperature jumps against its fast rise.

The corresponding model ARIMA (0, 1, 1) (1, 1, 0) is described by the equation

$$y_t = \mu - \theta_1 \varepsilon_{t-1} + \varphi_{12} y_{t-12} + \varepsilon_t \qquad (5)$$

After estimating the parameters, the model forecast for 2009 was constructed. Obtained series was pre-integrated on the same lags 1 and 12 to return to original value - accidents number xt. Simulation result is presented in Fig.5. The research of model adequacy can be carried out by analysing the residuals, Fig. 6.

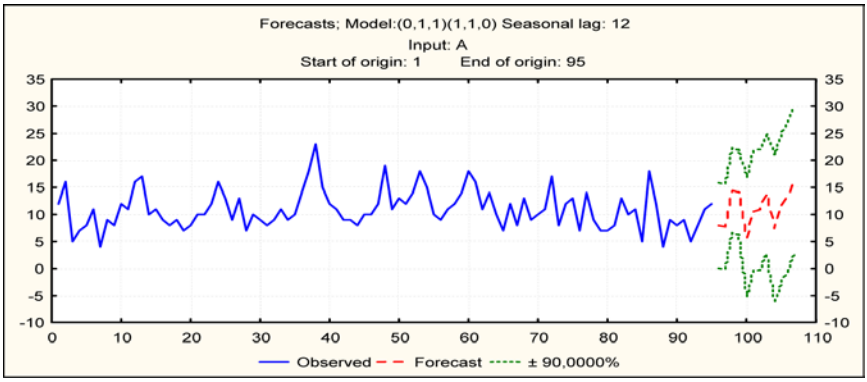

Figure 5 Forecast of accidents

The allocation of residuals is fairly well described by the normal allocation therefore the model adequately depicts the dynamics of the process under investigation.

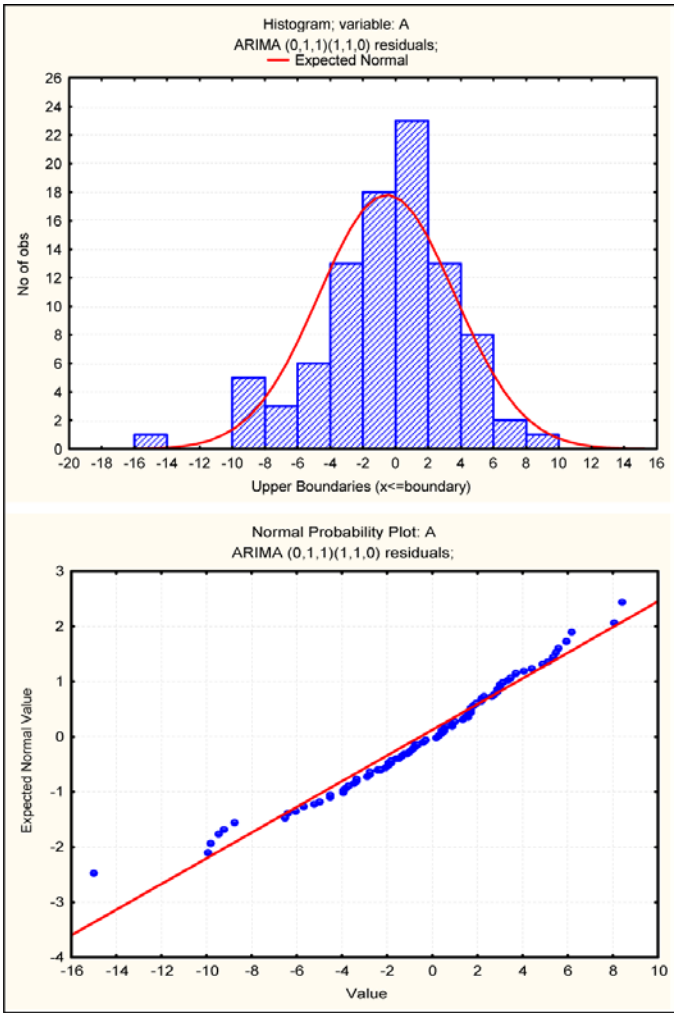

Figure 6 Allocation of residuals

4 Conclusion

For explaining the noticeable deviations from the forecasted value it is suggested to use the model where the weather abnormalities are considered as possible reasons of the increasing number of railroad emergency situations. In order to use the gathered results practically and to improve safety on the railway, we should be armed with detailed and long-term (5-10days) weather forecast. When visible deviation from the norm of factors described in the article is expected, railway workers should increase vigilance and take additional safety measures.

Acknowledgements

The article is written with the financial assistance of European Social Fund. Project Nr. 2009/0159/1DP/1.1.2.1.2/09/IPIA/VIAA/006 (The Support in Realisation of the Doctoral Programme "Telematics and Logistics" of the Transport and Telecommunication Institute).

References

[1] A. Melikyan. Evaluation of transport accident frequency change dynamics on Latvian railway). In: RESEARCH and TECHNOLOGY – STEP info the FUTURE, 2007, Vol.2, No1., Transport and Telecommunication Institute ,Riga ,Latvia, 2007. pp.135.

[2] Time Series Analysis: Forecasting and Control, 4th Edition. George, E. P. Box, Gwilym, M. Jenkins, Gregory, C. Reinsel. ISBN: 978-0-470 -27284-8 2008. 784p.

[3] A. Melikyan, B. Misnevs, R. Seregin Statistical Research of Weather Conditions Influence one the Railway Transport Accident Rate. In: Proceedings of the 7th International Conference RELIABILITY and STATISTICS in TRANSPORTATION and COMMUNICATION (RelStat'07), Transport and Telecommunication Institute, Riga, Latvia, 2007. pp. 120-126.