Analysis and Forecast of Mining Accidents in Pakistan

Kausar Sultan Shah1*, Izhar Mithal Jiskani2, Niaz Muhammad Shahani2, Hafeezur Rehman3, Naseer Muhammad Khan3 and Sajjad Hussain1

1. Department of Mining Engineering, University of Engineering and Technology Peshawar, Peshawar, Pakistan
2. School of Mines, China University of Mining and Technology, Xuzhou, China
3. Department of Mining Engineering, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan

Abstract

In the mining sector, the barrier to obtain an efficient safety management system is the unavailability of future information regarding the accidents. This paper aims to use the auto-regressive integrated moving average (ARIMA) model, for the first time, to evaluate the underlying causes that affect the safety management system corresponding to the number of accidents and fatalities in the surface and underground mining in Pakistan. The original application of the ARIMA model provides that how the number of accidents and fatalities is influenced by the implementation of various approaches to promote an effective safety management system. The ARIMA model requires the data series of the predicted elements with a random pattern over time and produce an equation. After the model identification, it may forecast the future pattern of the events based on its existing and future values. In this research work, the accident data for the period of 2006-2019 is collected from Inspectorate of Mines and Minerals (Pakistan), Mine Workers Federation, and newspapers in order to evaluate the long-term forecast. The results obtained reveal that ARIMA (2, 1, 0) is a suitable model for both the mining accidents and the workers’ fatalities. The number of accidents and fatalities are forecasted from 2020 to 2025. The results obtained suggest that the policy-makers should take a systematic consideration by evaluating the possible risks associated with an increased number of accidents and fatalities, and develop a safe and effective working platform.

1. Introduction

Pakistan is endowed with a vast mineral potential comprising coal, metallic and industrial minerals, rock salt, dimension stone, and precious stones [1-3]. Unfortunately, the acceleration of the turn-over of the mining industry in Pakistan does not comply with its potential [4-6]. The country has initiated an open-pit coal mine intended to achieve 7.8 million tons per year from block-I. Currently, small-scale coal mining is a major contributor to coal production in Pakistan [7]. Yet this sector is neglected regarding safety, leading towards loss of workers’ lives. Therefore, it is mandatory to carry out effective research works for the improvement of safety conditions and reducing the probability of mining accidents [8]. While supervising mining safety, the officer should acquire the precise forecasted data to evaluate future safety in order to prevent common workplace accidents and cause a safe work. Accurate and useful forecasting can not only diminish worker casualties and financial damage but also effectively improve the mine safety management system.

The most frequently referred factors responsible for mining accidents are electrocution, gas and dust explosions, mine collapses, haulage system failure, gas suffocation, mine machinery, falling from rocks, transportation, and explosives [7-10]. All of these factors responsible for mining accidents can be categorized into three levels: proximal causes, workplace causes, and systematic factors. The
...proximal causes recognize the human errors and workplace factor address violation or errors producing conditions leading to mining accidents, while the systematic factor evaluates the way in which management contributes to error or violation [11]. Due to the dynamic of safety issues in the mining industry, it is not possible to describe it with a simple data model.

Various statistical methods have been used to predict the upcoming situation [12-15]. The time series forecast has extensively used methods for predicting air pollution, variation and forecasting of precipitation and temperature, economic growth, jet engine remaining useful life, and civil and mining engineering accidents. In 1976, Box and Jenkins presented an ARIMA model to forecast the time series [16]. The ARIMA model uses time series of the forecasted data extracted with respect to the original time series as a random series. After choosing the accurate model, it may build a forecast of the future values according to the actual values of the time series. Many researchers have focused on predicting the mining and road accidents using the ARIMA model. Kher and Yerpude [17] have applied the time series forecasting models to forecast future accidents in the Indian coal mines. Al-Zyood [18] has examined the best ARIMA model to predict the future car road accidents in Saudi Arabia. Li [19] has built the ARIMA model based on the Mann-Kendall trend and mutation analysis approach to forecast the number of civil aviation accidents and causalities in the world. Ghédira, Kammoun [20] have assisted the policy-making process related to road safety in Tunisia using the ARIMA model for a time series analysis. Rajaprasad [21] has forecasted the number of factory accidents in India using the ARIMA model, revealing that the number of accidents could be increased in the future; therefore, the policy-makers should pay attention towards the uncertain situations. Wu et al. [22] have developed a grey auto-regressive moving average model based on the empirical mode decomposition (EMD-GM-ARMA) in order to predict the mining safety situation, and have compared the results obtained with those obtained for the ARMA model, particle swarm optimization-support vector machine (PSO-SVM), EMD-GM-ARMA, and neural network. The results obtained revealed that the EMD-GM-ARMA model exhibited a high accuracy prediction compared to the other models. Shao et al. [23] have used a combination of various predictive models with ARIMA to predict the traffic accidents. They established a combination of ARIMA with BP-neural network and ARIMA with support vector machine (SVM). The results obtained revealed that a combination of the models provided more accuracy than a single model. Yixuan et al. [24] have developed a hybrid prediction model based on ARIMA and the support vector regression (SVR) model to predict the traffic accidents. The results obtained revealed that the hybrid prediction model provided a higher accuracy than the single ARIMA model. The above researchers focused on forecasting the future accidents or evaluating the accidents to provide the respective enhanced weight to factors responsible for the accidents. They found that the forecasting values of accidents increased or decreased but the results obtained also showed that the time series analysis provided a better overview to the safety officers to improve the safety management system.

This paper aims to use the ARIMA model, for the first time, to evaluate the underlying causes that affect the safety management system corresponding to the number of accidents and fatalities in the surface and underground mining in Pakistan. This original application of the ARIMA model to the problem under discussion provides an insight into how the number of accidents and fatalities is influenced by the implementation of various approaches to promote an effective safety management system. In the view of the reviewed literature, it could be stated that a majority of the forecasted models focus on the number of accidents, while, in this work, the ARIMA model has been developed based on the statistical analysis for both the number of accidents and the number of fatalities.

2. Materials and Methods
2.1. Data Sources

This research work was carried out according to the collected statistical records of both the surface and the underground mining accidents and causalities in Pakistan from 2006 to 2019. The data also involved the factors responsible for mining accidents and fatalities, which included electrocution, gas and dust explosions, mine collapses, haulage system failure, suffocation, mine machinery, falling of rocks, transportation, and explosives. The statistical data related to mining accidents was collected from Inspectorate of Mines and Minerals and Pakistan Mine Workers Federation (http://www.pmwf.info/), and newspapers. The data was collected from five highly distributed newspapers in Pakistan, The Express Tribune, DAWN, Frontier Post, Pakistan...
today, and The INTERNATIONAL News. The newspapers were searched on the internet for mining accident records, and reviewed independently for information on the fatalities and accidents. The accident information was recorded manually, and the findings were verified with the data obtained from the Inspectorate of Mines and Minerals, and the Pakistan Mine Workers Federation.

2.2. Research methods
The time-series forecasting models assist in expressing the sort of relationship between the parameter past and future values [25]. The performance of the time series model is more significant in the absence of the explanatory model, which relates the forecasted variable with explanatory variables. In this research work, a time series model, the ARIMA model, was used to model the accident data in Pakistan.

2.2.1. ARIMA model
The ARIMA model contains significant statistical properties that make it suitable to forecast the linear data. It is assumed that for the stationary time series, the ARIMA model is a linear (i.e. regression type) function of lags of the past observations and lags of forecasted errors of the variables. Therefore, the basic approach of generating the time series equation can be represented through Equation 1.

\[ y_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \varepsilon - \beta_1 \varepsilon_{t-1} - \ldots - \beta_q \varepsilon_{t-q} \]  

where \( y_t \) represents the actual values, while \( \varepsilon_t \) denotes the random error at the time period \( t \); accordingly, \( \alpha_i \) (i = 1, 2, 3,……….., p) and \( \beta_i \) (i = 1, 2, 3,……….., p) are the model parameters, and p and q are the integers that identify the order of the model.

Moreover, \( \varepsilon_t \) is the uniformly distributed over the zero mean and constant variance of \( \sigma^2 \). As the order parameters, p and q define the type of model, the model considered as the auto-regressive (AR) model when p = 0 corresponding to the moving average (MA) type model if q = 0. Similarly, it may be inferred that the ARIMA model Equation 1 depicts the optimistic order of the model parameters p and q. Box, Jenkins [16] have presented a methodology to perform the time series analysis using the ARIMA model including identification, estimation, and diagnostic checking.

The ARIMA model provides a transformation of non-stationary time series into a stationary time series, and presents the dependent variables in the form of the lag value of the residual errors. In practice, the ARIMA model is really basic and requires the dependent variables besides using the additional independent variables. The alternative predictive models such as the neural network and linear regression models require the independent variables for forecasting. However, the mining accidents and fatalities forecasted by this research work consist of a number of independent variables that are inconsistent and unpredictable. Consequently, the ARIMA model has the ability to solve a related problem to forecast the properties of the data. Hence, the ARIMA model meets the requirement of the research work.

3. Results and Discussion
3.1. Mining accident change trend analysis in Pakistan
In this work, we analyzed the variability trend in a number of accidents and fatality of mining accidents in Pakistan. The number of accidents exhibits the increments in a fluctuant way over the years (see Figure 1). From 2006 to 2019, the number of accidents was inconsistent but it was high in 2019, indicating that the accidents increased with a continuous increase in the mining activities. The R-square value for the trend is 0.54, which is statistically highly significant. In this work, the accident reports were analyzed in detail. In Figure 2, the number of accidents is categorized by causes of the accidents. The data obtained revealed that most accidents occurred due to the roof fall followed by suffocation from gases.

3.2. Analysis of mining accident change with time regarding cause of accidents
Figure 2 illustrates the number of accidents with a variety of causes including electrocution, gas and dust explosions, mine collapses, haulage system failure, suffocation from gases, mine machinery, falling of rocks, transportation, and explosives. In the roof fall and suffocation from gases, the number of accidents is 29.3% and 20.4% of the total accidents. The continuous increase in roof fall and suffocation by cases implies that the accidents still prevail but the safety conditions remain unchanged over the years.
3.3. Mining fatality change trend analysis in Pakistan

As shown in Figure 3, there is a significant variation in the trend of the number of accidents and fatalities in the mining sector of Pakistan. From 2006 to 2019, the fatalities are continuously fluctuating, showing a periodic up and downtrend. The R-square value for the trend is 0.42, which is statistically significant. The data obtained revealed that there was an average of 143 fatalities annually. The most frequent cause of a worker’s fatality is the fall of the roof followed by the suffocation from gases (see Figure 4). This is due to the conventional and outdated methods, capitalistic production, precarious working, lack of implementation of mine law and regulations, unskilled workers, and particularly overburden mining inspectorate.

According to Figure 3, the fatalities can be divided into two periods according to change. In the first period (2006 to 2010), the number of fatalities is around 148 and exhibits a continuous decrease. In the same period, there was a rapid grew within 2011 and 2012 and then reduced inconsistently. Similarly, during the second period (2014 to 2019), the estimated fatalities consistently increased and reached 171 in 2018; meanwhile, there was no fluctuant decrease for five years. There are numerous mining accidents; however, each province has its mine management system. Therefore, the year 2019 with the highest number of accidents does not reflect the highest number of fatalities. The statistics of the mining accidents and fatality revealed that the mining sector in Pakistan was in a high safety risk.

3.4. Analysis of mining fatality change with time regarding cause of accidents

Similar to the number of accidents, the fatalities of mining workers are higher in fall of roofs and suffocation from gases. In the year 2015, the number of fatalities was higher by suffocation from gases compared to roof fall; the number was minimum in the electrocution, explosives, and mine machinery. In accordance with the law of time variation, specific cause of accidents from 2006 to 2019, the mining workers’ fatality exhibited a random nature and increased with an inconsistent raise excluding the factors including electrocution, explosives, and haulage.
3.5. Model Selection for time series forecast on key indicators of number of mining accidents

For the time series analysis, the ARIMA model accounts for the type of data (stationary or non-stationary). Usually, the ARIMA model desires the stationary data, although the non-stationary data is pretreated to be stationary. In the first step, the logarithmic transformation is used to eliminate heteroskedasticity in the time series; the second step involves the application of the first-order difference method to transfer time series to the stationary data. The autocorrelation function (ACF) and partial autocorrelation function (PACF) plot (see Figures 5 and 6) are used to obtain the values of ARIMA th parameters p, q, and d.

![Figure 5. Auto-correlation plot for the number of accidents.](image)

![Figure 6. Partial auto-correlation plot for the number of accidents.](image)

Furthermore, Akaike information criteria (AIC), Mean absolute error (MAE), R-square, and Root mean square error (RMSE) are used as the evaluating criteria to select the optimum model. A model with a high accuracy exhibits a smaller value for AIC, MAE, and RMSE and a larger value for R-square. By evaluation of the parameters in Table 1, ARIMA (2, 1, 0) is selected optimum with constant, where RMSE = 16.18, MAE = 12.25, AIC = 5.89, and R-square = 0.626. The consequent model can be employed to predict the time series of the number of mining accidents in Pakistan.

| Table 1. Comparison of the parameters for ARIMA model in mining accidents. |
|-----------------|-------|-------|-------|-------|
| Model           | RMSE  | MAE   | AIC   | R-Square |
| ARIMA (2, 1, 0) | 16.81 | 12.25 | 5.89  | 0.626    |
| ARIMA (0, 1, 0) | 16.78 | 12.45 | 5.59  | 0.557    |
| ARIMA (1, 1, 0) | 17.46 | 12.39 | 5.81  | 0.557    |
| ARIMA (2, 1, 0) | 16.70 | 12.48 | 5.87  | 0.631    |
| Natural logarithm transformation | 16.18 | 12.52 | 5.59  | 0.553    |
| Natural logarithm transformation | 17.46 | 12.39 | 5.81  | 0.557    |

The ARIMA (2,1,0) model is employed to forecast the number of mining accidents in Pakistan from 2006 to 2019. The actual, fitted, and forecasting values for the number of accidents are given in Table 2. In this work, the number of accidents was forecasted from 2020 to 2025 using the selected model with respect to the data for the number of accidents from 2006 to 2019. Afterwards, the forecasted values were added to the existing time series to develop a model adjusted to the advanced time series to forecast the number of mining accidents from 2020 to 2025 (see Figure 7).

![Figure 7. Time sequence plot for the observed and forecasted number of a mining accidents.](image)
Table 2. Actual and predicted values for the number of mining accidents.

<table>
<thead>
<tr>
<th>Time (Year)</th>
<th>Actual values</th>
<th>Forecasting values</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>194</td>
<td>194.38</td>
<td>-0.38</td>
</tr>
<tr>
<td>2016</td>
<td>202</td>
<td>193.08</td>
<td>8.92</td>
</tr>
<tr>
<td>2017</td>
<td>205</td>
<td>199.07</td>
<td>5.93</td>
</tr>
<tr>
<td>2018</td>
<td>231</td>
<td>206.76</td>
<td>24.24</td>
</tr>
<tr>
<td>2019</td>
<td>234.66</td>
<td>234.66</td>
<td>-1.66</td>
</tr>
<tr>
<td>2020</td>
<td>227.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td>231.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>242.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2023</td>
<td>244.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2024</td>
<td>248.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2025</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparison of the predicted values and real values of the number of accidents and causalities of the mining sector in Pakistan in 2015 to 2019 identifies a significant relative error, especially in 2019. Generally, the time series for the selected model is optimum because the entire residual error is lower. However, in the view of the uncertainty related to time series, the authenticity of the entire simulation can only be assured.

3.6. Model Selection for time series forecast on key indicators of mining fatalities

A similar procedure was used to model the mining accident fatalities in Pakistan from 2006 to 2019. The ACF and PACF plots for the number of fatalities in mining accidents in Pakistan are given in Figures 8 and 9. From these plots, the values for p, q, and d were selected for the ARIMA model.

Consequently, ARIMA (2, 1, 0) was selected as the optimum model with constant by considering the parameters RMSE = 8.23, MAE = 6.19, AIC = 4.46, and R-square = 0.561 (see Table 3). The optimum model should not only provide a comparatively accurate forecast, but it also should exhibit the residual noise with an irregular pattern.

Therefore, the ARIMA (2, 1, 0) model was used to forecast the number of mining fatalities from 2020 to 2025 with respect to the original time series of fatalities from 2006 to 2019. The actual values and forecasted values using the ARIMA model for the number of mining fatalities in Pakistan are given in Table 4.

Similarly, the number of mining fatalities was forecasted from 2020 to 2025 using the selected ARIMA (2, 1, 0) model with respect to the data for the number of accidents from 2006 to 2019. Afterwards, the forecasted values were added to the existing time series to develop a model with advanced time series to forecast the number of mining fatalities from 2020 to 2025 (see Figure 10).
By comparing the results with the literature, it was revealed that the ARIMA model developed in this work held the highest prediction accuracy. The accuracy of the ARIMA model was obtained from the average difference between the actual values and the predicted values. In this work, ARIMA (2, 1, 0) was selected as the best model for both the number of mining accidents and the number of fatalities with an average difference of 3.4 and 6.1. Similarly, Li [19] has found the average difference of 7.2 and 268.8, while, Kher and Yerpude [17] have found 12.2. The increase in the average difference between the actual and forecasted values is due to the increase in the number of accidents; however, overall, the model selected in this work exhibited a higher accuracy than selected in the literature.

Forecasts of the number of mining accidents and fatalities can be taken as a benchmark for developing an efficient safety management system. The accuracy of the model decreases with increase in the length of the forecast period because the forecasted error slowly increases with the extension of the period. From the results of the ARIMA model, it can be deduced that the forecasts of the number of accidents and causalities are quite correct.

The mining accidents analysis is a significant part of the safety research that highlights the key issues and loopholes the related safety management system. In this research work, long-term sequence mining accidents and fatality statistics were used to evaluate the characteristics of mining accidents and forecast the future tendency of mining accidents. The methodology of this research work mainly focused on the statistical analysis and time series analysis, and certain useful conclusion were obtained. Understanding the key factors behind the mining accidents is a complex task because of the frequent categories such as electrocution, gas and dust explosions, mine collapses, haulage system failure, suffocation from gases, mine machinery, falling of rocks, transportation, and explosives. Therefore, it is difficult to develop a model to forecast the mining accidents and fatalities. We considered the time variation for time series analysis and developed a model to predict the future values of mining accidents and fatalities. In this research work, we explored the time regularity of the mining accidents but we could not find how to move towards the end results of the accidents occurring. In the future, we can develop an accident mechanism model to simulate and evaluate the process of mining accidents and fatalities. It is recommended to use other dependent variables to enhance the accuracy of the time series forecasting such as unsafe behavior and unsafe physical condition. Similarly, multiple independent variables such as the age of worker, occupation, and workplace can be used to develop a time series model.

4. Conclusions

In this research work, we presented a statistical analysis and a time series forecasting using the ARIMA model for the number of mining accidents and fatalities in Pakistan from 2006 to 2019, and the following conclusions were drawn:

1. Overall, the number of mining accidents in Pakistan are increasing in a fluctuant way over the years. In 2006 to 2019, there were two periods of changes in the number of accidents. The number decreased by 28 in the first period (2006-2010) and by 25 in the second period (2011-2013) and exhibit increment by 34 in the first period (2010-2012) and by 48 in the second period (2014-2019). A similar trend was exhibited by the number of fatalities from 2006 to 2019. In the first period, there was a decrease by 11 (2006-2010), and by 17 in the second
period (2011-2013); however, overall, the number of fatalities showed a decrement in 2019. The statistical analysis revealed that the mining sector in Pakistan is at a high safety risk; therefore, there is an imperative need to adopt an appropriate and most effective safety management system to reduce the number of accidents and fatalities. The analysis mentioned above indicates that there is a profound professional incompetence related to the prior evaluation of safety risk; therefore, there was an urgency to develop a time series analysis model to evaluate the future safety risk in Pakistan.

2. From ACF and PACF, plots of the best ARIMA (2, 1, 0) model was selected for the number of mining accidents and fatalities, and the forecasted results showed a significant increase in the number of mining accidents and fatalities in Pakistan. The results obtained can provide an insight into how the number of accidents and fatalities is influenced by the implementation of various approaches to promote an effective safety management system. Hence, observing the safety risk by forecast can provide a prompt detection for the concealed danger of mine safety and promote a safe and effective working platform.

Acknowledgment: Kausar Sultan Shah is extremely thankful to the Higher Education Commission (HEC) of Pakistan for HRDI-UESTPs scholarship.

Conflicts of Interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

References


کورس سلطان شاه، ایسهار میتال جیسکانی، نیاز محمد شاهانی، حافظ الرحمن، ناصر محمد خان و سجاد حسین

بررسی و پیش‌بینی حوادث معدن در پاکستان

کلمات کلیدی: روش‌های معادلات حضور رگرسیون، فویت‌ها سیستم مدل‌بردار، پیش‌بینی ایمنی، پیش‌بینی ایمنی معدن.