ORIGINAL ARTICLE

Prediction model of seasonality in the construction industry based on the accidentality phenomenon

Bożena Hoła[1] · Mariusz Topolski[2] · Iwona Szer[3] · Jacek Szer[3] · Ewa Blazik-Borowa[4]

Abstract

The construction industry is an economic sector that is characterized by seasonality. Seasonal factors affect the volume of production, which in turn affects the accident rate. The aim of the research presented in the article was to develop a model for predicting the number of people injured in occupational accidents in the construction industry. Based on the analysis of statistical data and previous studies, the occurrence of certain regularities of the accidentality phenomenon was found, namely the long-term trend over many years, as well as seasonality and cyclicality over the course of a year. The found regularities were the basis for the assumptions that were made for the construction of the model. A mathematical model was built in the non-linear regression dimension. The model was validated by comparing the results of prediction errors generated by the developed model with the results of prediction errors generated by other known models, such as ARIMA, SARIMA, linear and polynomial models, which take into account the seasonality of the phenomenon. The constructed model enables the number of people injured in accidents in the construction industry in selected months of future years to be predicted with high accuracy. The obtained results can be the basis for making appropriate decisions regarding preventive and prophylactic measures in the construction industry. Commonly known mathematical tools available in the STATISTICA package were used to solve the given task.

Keywords Construction industry · Seasonality in the construction industry · Nonlinear regression · Occupational safety · Accidentality prediction · Long-term trend

1 Introduction

The construction industry is one of the most important sectors of the economy of every country. Due to the high degree of diversity and complexity of construction works, it is characterized by a high level of hazard to the life and health of employees. The consequence of this is a high accident rate for this sector of the economy, which is measured by the number of people injured in accidents each year. This statement applies to most countries in the world [1–3].

The accidentality phenomenon has a negative impact on the economy and causes a serious global public health problem. In many countries of the world, to improve the existing accident situation at workplaces, actions that aim to achieve a zero number of accidents have been initiated [4]. This goal can be achieved by legislative changes in labor law, systematic employee education, shaping the appropriate safety climate and safety culture in enterprises, and with appropriate management. To properly shape the policy of changes that are introduced, and the directions of accident prophylaxis

✉ Bożena Hoła
bozena.hola@pwr.edu.pl

Mariusz Topolski
mariusz.topolski@pwr.edu.pl

1 Faculty of Civil Engineering, Wroclaw University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

2 Faculty of Information and Communication Technology, Wroclaw University of Science and Technology, Wybrzeze Wyspianskiego 27, 50-370 Wrocław, Poland

3 Faculty of Civil Engineering, Architecture and Environmental Engineering, Lodz University of Technology, Politechniki 6, 90-924 Lodz, Poland

4 Faculty of Civil Engineering and Architecture, Lublin University of Technology, Nadbystrzycka 40, 20-618 Lublin, Poland

and prevention, it is necessary to observe trends in occupational accidents and to analyze forecasts in this area.

The accident rate in the construction industry depends on many factors of a technical, human, managerial, legislative, environmental and economic nature. The accidentality phenomenon in the construction industry is also influenced by unknown factors that have not yet been investigated [5–8]. Research conducted by scientists in many countries around the world has shown that the number of accidents in the construction industry varies from year to year, as well as from month to month.

The aim of the research was to develop an accident prediction model in the construction industry. Statistical data published by the Central Statistical Office in Poland was analyzed. The occurrence of certain regularities in the accidentality phenomenon was found, namely, a decreasing long-term trend, as well as seasonality and cyclicality of this phenomenon. The found regularities were the basis for the assumptions that were made to build the model. These assumptions included: non-linear regression, as well as the seasonality and cyclicality of the phenomenon. The STATISTICA package was used for the calculations.

Building the model will allow the number of people injured in accidents in selected months of the coming years to be predicted. The obtained results can be the basis for making appropriate decisions regarding preventive and prophylactic actions. Knowledge about hazards in a workplace, and also the adopted methods of preventing accidents, which were provided systematically to employees, may change their attitudes in terms of the perception of dangerous situations at work and the use of safe methods of work. It can also be the basis for forecasting the future costs of occupational accidents, which in turn may have an impact on prices in the construction industry, as well as on financial and insurance policies.

2 Literature survey

The construction industry is considered to be one of the most dangerous sectors of the economy. Statistical data published in various countries of the world concerning occupational accidents indicate that the problem of accidents in the construction industry is global. The risk of a fatal accident in the construction industry is three times higher than in other industries [9, 10]. Each year, the construction industry around the world suffers more than 60,000 deaths. The phenomenon of accidentality in the construction industry should, therefore, be given special attention, as it has a negative impact on the economy, employers, employees, and their families.

When analyzing the accident rate in the construction industry on a monthly basis, it can be seen that the number of accidents in individual months of the year is different [11]. This may indicate the seasonality of this phenomenon. This thesis is confirmed by research conducted by scientists in many countries. In Singapore, for example, the highest number of deaths in accidents in the construction industry occurs in October before the rainy season [1]. In Taiwan, in the period from 2000 to 2009, most accidents occurred in the summer [12]. In turn, in South Australia, the number of accidents in the period from 2002 to 2011 was the highest in Autumn [2]. Based on a study of occupational injuries and diseases in the United States in the years 2003–2010, it was found that most of the injuries occurred during the summer months [13]. Accidents in Korea occurred less frequently in winter than in other seasons [14].

The above-mentioned examples show that there are seasonal factors that influence the occurrence of accidents in the construction industry. Research conducted by Traczyk and Trzebski [15], which concerned work-related physical effort, showed that a too high or low temperature, high humidity, high and low atmospheric pressure, and also changes in wind speed may cause changes in the body of an employee—even in acclimatized people. Such changes could become a cause of an accident. It was also noted that the number of accidents depends on the climatic zone. In Spain, for example, in the period 2003–2008, the majority of accidents occurred in the Mediterranean and continental zones and accounted for 48% and 35% of all accidents, respectively [3]. The influence of climatic zones and the period of the year was also analyzed in China [16]. It was observed that in the years 2010–2016, August was the month with the highest average number of accidents, and the incidence rate was the highest in regions with a mild climate. Therefore, a constant pattern of the accidentality phenomenon in the construction industry can be observed. The fact that it repeats each year confirms the seasonality of the phenomenon.

The seasonal nature of the accidentality phenomenon in the construction industry is influenced by many factors. W. J. Granger defined a seasonal factor as any pattern that repeats itself every year in the same month, regardless of the cause. He classified the factors influencing seasonality in the construction industry into four categories, namely [17]:

- climate-related factors, including weather changes, which have a direct impact on the efficiency of
- construction production;
- calendar-related factors, including the number of actual working days in a month, which affect monthly production;
- date-related factors such as the number of public holidays during which production is stopped;
- expected factors that influence seasonal behavior related to, e.g., the date of VAT settlements, the payment of other taxes, the end of the year, and also the need to

comply with both the production result and the annual budget.

Seasonal factors affect the amount of construction production, which in turn affects the accidentality rate. Mach et al. [18] identified cyclical fluctuations in relation to the construction industry in 16 European countries. Moreover, the authors of article [19] showed the dependence between the number of accident and production cycles in the construction industry in South Korea. The authors of paper [20] analyzed the influence of seasonal factors on the costs and duration of the implementation of a project. The relationship between the accident rate and the value of production was proved by, among others, Fabiano et al. Research concerning the Italian industry showed that there is a cyclical trend in the number of fatal accidents and that the number of accidents depends to a large extent on the value of production [21]. Similar conclusions were formulated by Dong et al. [22]. Their research, concerning falls from roofs, confirmed the relationship between the number of fatalities and the economic cycle. Research conducted by Hoła and Szostak [23], which referred to the construction industry in the European Union countries, also indicated the existence of a long-term cyclical trend in the accidentality phenomenon.

Analysis of the accidentality phenomenon in the construction industry can also be found in [24–27]. The research conducted by Hoła [28] regarding the Polish construction industry indicated that the accident rate in the construction industry changes every year, and that there is a clear downward trend. Studies of trends in the accidentality phenomenon and the prognostic models that are built on their basis can be used to predict future events. Moreover, based on this knowledge, decisions in the area of labor law, education and prophylactic and preventive actions can be undertaken.

For the modeling of the accidentality phenomenon in the construction industry, among others, linear mathematical models, descriptive and graphical models, and IT models (e.g., neural networks) can be used. For example, the research included in [28] involved qualitative and quantitative analyses that resulted in the creation of mathematical models for the development of the trend of the accidentality phenomenon in the Polish construction industry. In turn, Indian scientists proposed the use of artificial neural networks (ANNs) to predict the behavior of construction workers in a risk situation [29]. In American studies [30], an approach based on the Bayesian network (BN) was proposed to diagnose the risk of falls from a height. The tests performed on the model enabled the probability of an occurrence of an accident related to various safety hazards to be determined.

Thanks to IT technologies, it is possible to virtually model the construction environment and its hazards. Paper [31] proposes the use of a construction site model developed in a virtual environment for training and educational purposes. In turn, studies [32] propose the use of a virtual reality environment to identify hazards, to determine the rules of movement of employees on a construction site, and to specify the impact of the building environment on its surroundings. In research [33], the authors proposed a model that enables the tracking of construction workers in real time using Real-time locating systems (RTLS). The developed tool can be used to control the exposure of construction workers to hazards that change over time.

The authors of study [34] proposed an approach based on machine learning (ML) to assess occupational safety. According to the authors, all accidents follow certain patterns, which in turn allows them to be predicted. The ML technique was used to predict slip-trip-fall (STF) accidents.

Based on the literature review, it can be concluded that the accidentality phenomenon in the construction industry is seasonal, and that the number of occupational accidents changes cyclically over a period of time covering one calendar year. When conducting studies concerning the accidentality phenomenon during a multi-year period, it is also possible to identify a general trend in the accident rate in the construction industry. The constructed prediction model should, therefore, take into account both of these phenomena simultaneously.

To predict data, various methods are used, such as linear, logistic and Bayesian regression [35]. To analyze the time series in the situations of many explanatory variables, various selection and extraction methods can be used, which will allow a set of features that best discriminate the explained variable to be obtained. Such methods include Principal Component Analysis [35, 36].

The review of the methods used to build prediction models shows a certain research gap, namely that all regression methods encounter difficulties in a situation when, on the basis of numerical data, the simultaneous occurrence of a trend and seasonality is found. Although there are methods that take into account the seasonality and cyclicality of phenomena, such as: ARIMA and SARIMA [37, 38], their application is limited in many cases. This is due to the fact that these methods can be used to model stationary series, i.e., series in which there are only random fluctuations around the mean value, or non-stationary series that are reducible to the stationary form [39]. In addition, the ability to predict events that may only occur in the near future can be seen to be a limitation.

Taking into account the above observations, the authors attempted to build a model for predicting the number of people injured in accidents in the construction industry, which takes into account both the long-term trend and the seasonal nature of the phenomenon with regards to non-linear regression. The information obtained on the basis of the model on the expected number of accidents in particular months of

subsequent years may be the basis for making decisions in the field of legislative, educational, prophylactic and preventive actions, as well as for estimating the costs incurred due to accidents. This in turn may affect prices in the construction industry, as well as the financial and insurance policy of enterprises.

3 Materials and methods

3.1 Statistical data

To develop the prediction model, statistical data published by the Central Statistical Office in Poland were used. The data concerned the number of people injured in occupational accidents in the construction industry in particular months of the year during the period covering the years from 2007 to 2019 [40]. Figure 1 shows, in the form of a three-dimensional graph, how the total number of people injured in occupational accidents in the construction industry ($x_{i,j}$) changed in the individual months of the analyzed years.

This figure shows that the number of people injured in accidents in the construction industry was high at the beginning of the analyzed period, while from 2012 onwards it shows a clear downward trend. When analyzing the distribution of the number of people injured in accidents in the following months of the studied years, a certain regularity can also be noticed: the accident rate in January is low, then increases, reaches its maximum in July and August, and decreases again in the following months. The above
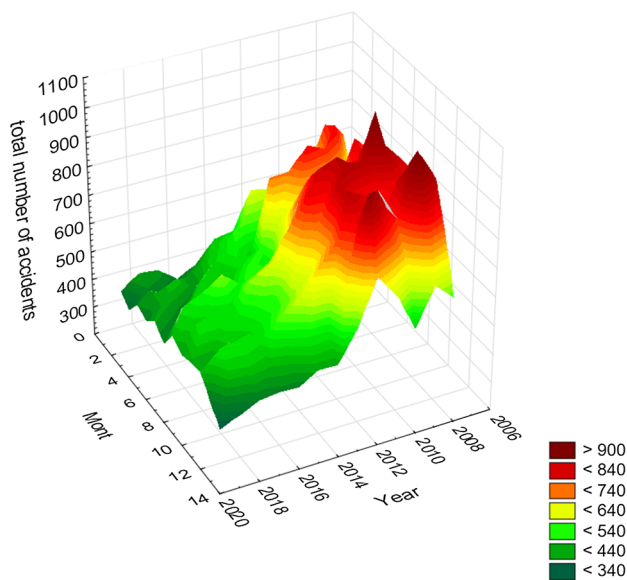

Fig. 1 Graph of the variability of the total number of people injured in the construction industry in individual months of the analyzed years

observations suggested examining the correlation between the following years of the analyzed period of time and the number of people injured in the construction industry, and also the correlation between individual months of the year and the number of people injured in those months. The obtained values of the correlation coefficients will indicate whether there is correlative dependency, and what the actual strength of the relationship between these variables is.

3.2 Investigation of the correlative dependencies between variables

Due to the fact that one of the variables assumes ordinal values, and that the other one assumes quantitative values, the Spearman's rank correlation was used to determine the strength of the relationship between the studied variables [41, 42]. The study of the correlative dependence was carried out for two sets of numerical data, namely:

- ordinal variable $r_i; i = 1, \ldots, 13$ represents the subsequent year in the period from 2007 to 2019, and quantitative variable $x_i; i = 1, \ldots, 13$ is the number of people injured in occupational accidents in the construction industry in year r_i.
- ordinal variable $m_j; j = 1, \ldots, 12$ is the subsequent i-th month of the year, and quantitative variable $x_j; j = 1, \ldots, 12$ is the sum of the number of people injured in occupational accidents in the construction industry in all months $m_j; j = 1, \ldots, 12$ within the time interval from 2007 to 2019

The following scale of evaluating the strength of the relationship between variables r_i and x_i and m_j and x_j, which was expressed by the Spearman Rho coefficient [43], was adopted:

- Rho = 0—the variables are not correlated,
- 0 < Rho < 0.1—very weak correlation,
- 0.1 ≤ Rho < 0.3—weak correlation,
- 0.3 ≤ Rho < 0.5—average correlation,
- 0.5 ≤ Rho < 0.7—high correlation,
- 0.7 ≤ Rho < 0.9—very high correlation,
- 0.9 ≤ Rho < 1—nearly full correlation.

The calculations were conducted using the STATISTICA package. Table 1 presents descriptive statistics of variable x_i—the total number of people injured in occupational accidents in the construction industry in the i-th calendar year.

The calculated values of the Rho and p statistics (rho = − 0.847; $p < 0.01$) indicate a very high negative correlation. It can be concluded that the total number of accidents decreases over the period from 2007 to 2019. In turn, Table 2 contains descriptive statistics for variable x_j—which

Table 1 Descriptive statistics of variable x_t—the total number of people injured in occupational accidents in the construction industry in the years from 2007 to 2019

i	Year r_i	Number of injured people x_i	Monthly average M_i	Standard deviation SD_i	Minimal value min$[x_i]$	Maximal value max$[x_i]$
1	2007	8895	741.25	96.97	540	880
2	2008	9934	827.83	111.15	601	989
3	2009	8684	723.67	106.72	470	872
4	2010	9098	758.17	115.58	555	892
5	2011	9222	768.50	99.56	598	934
6	2012	8125	677.08	86.76	533	805
7	2013	6712	559.33	92.85	423	708
8	2014	6264	522.00	52.12	426	609
9	2015	5776	481.33	55.68	387	567
10	2016	5468	455.67	57.56	364	568
11	2017	5390	449.17	49.08	371	520
12	2018	5247	437.25	46.83	348	500
13	2019	4743	395.25	39.43	320	448

Table 2 Descriptive statistics of the total number of people injured in occupational accidents in the construction industry in month m_j; $j = 1, \dots, 12$ during the period from 2007 to 2019

j	Month m_j	The sum of people injured in the years 2007–2019 in month j x_j	Long-term average M_j	Standard deviation SD_j	Minimal value min$[x_j]$	Maximal value max$[x_j]$
1	January	6828	525.23	135.10	364	774
2	February	6920	532.31	139.59	345	806
3	March	7625	586.54	158.74	401	803
4	April	7287	560.54	138.43	384	828
5	May	7794	599.54	160.42	387	808
6	June	8580	660.00	181.74	380	987
7	July	8883	683.31	176.28	448	907
8	August	8461	650.85	159.27	418	849
9	September	8363	643.31	175.48	425	897
10	October	8862	681.69	181.53	442	989
11	November	7843	603.31	179.48	375	934
12	December	6112	470.15	113.69	320	689

is the sum of people injured in occupational accidents in the construction industry in months m_j; $j = 1, \dots, 12$ during the time interval from 2007 to 2019.

The calculated values of the Rho and p statistics (rho $= 0.094$; $p = 0.245$) indicate a very weak correlation between the variables, which proves their non-linear dependence and the seasonality of the phenomenon.

To illustrate the seasonality of the accidentality phenomenon in the construction industry, graphs of the variability of the total number of people injured in accidents in the construction industry in particular months m_j; $j = 1, \dots, 12$ of consecutive years were developed. Figure 2 shows the correlation curves and the distribution charts of the monthly number of people injured in occupational accidents in the construction industry in individual years of the analyzed time period.

Figure 2 provides very important information. Each year, there is a very repeatable scenario in which the total number of accidents over a period of 1–12 months can be approximated by a quadratic function. The maximum value of this function falls on the 7-8th month of the year. The distribution of the parabola arms is also comparable in individual years, and it can, therefore, be concluded that we are dealing with the phenomenon of an increase in accidents in the summer, and a decrease in the winter.

As a measure of adjustment of the correlation curve (in the form of a quadratic function) to the empirical values, the coefficient of determination R_i^2 was adopted. The value of this coefficient is presented in Table 3.

When analyzing the results, it can be noticed that the value of the determination coefficients varies from $R_{12}^2 = 0.305$ (which means an average correlation) to
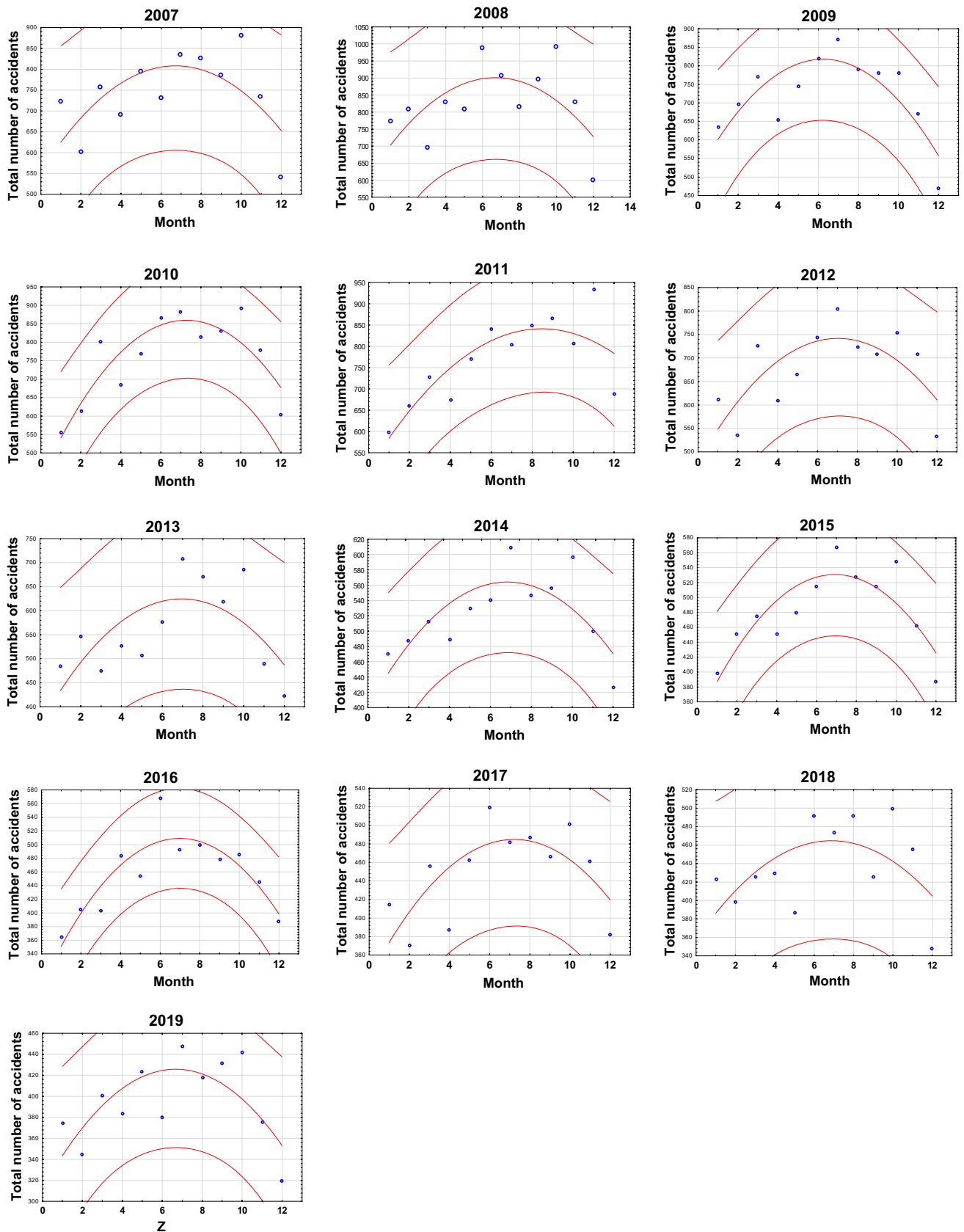
Fig. 2 Scatterplots of the total number of people injured in accidents—broken down into years and months

Table 3 The coefficients of determination R^2 of the adjustment of data for individual years between 2007 and 2019

i	Year r_i	Coefficients of determination R_i^2	Relevance p_i
1	2007	0.413	0.020
2	2008	0.375	0.021
3	2009	0.679	0.011
4	2010	0.751	0.008
5	2011	0.692	0.010
6	2012	0.510	0.016
7	2013	0.446	0.018
8	2014	0.577	0.014
9	2015	0.706	0.010
10	2016	0.782	0.007
11	2017	0.510	0.016
12	2018	0.305	0.023
13	2019	0.519	0.016

$R_{10}^2 = 0.782$ (which means a very high correlation). The best adjustment of the correlation curve to empirical values occurs for 2016. For further considerations, the correlation curve for this year was adopted as a model of seasonal changes in the accident rate in the construction industry.

The charts presented in Fig. 2, when combined with each other (Fig. 3), show how the total number of people injured in occupational accidents in the construction industry changed in individual months in the period from 2007 to 2019. Figure 3 also shows the decreasing trend of the above-mentioned feature, which was confirmed by a very high negative correlation (rho $= -0.835$, rho$^2 = 0.697$, $p < 0.01$).

3.3 Mathematical model of the accidentality phenomenon in the construction industry

When building the model, the following assumption was made: if the changes in the total number of accidents in the same months of subsequent years are comparable, then it is possible to estimate such a constant numerical value that will allow a trend for subsequent annual cycles to determined.

For this purpose, the following were calculated: the average annual increase/decrease in the total number of accidents, and the average monthly increase/decrease in the total number of accidents.

The average annual increase/decrease in the total number of people injured in accidents Δr can be written as follows:

$$\Delta r_i = \begin{cases} \frac{100}{k} \sum_{j=1}^{k} \left(\frac{1}{n-1} \sum_{i=2}^{n} \frac{-|x_{i,j} - x_{i-1,j}|}{\max [x_{i,j}; x_{i-1,j}]} \right) [\%], & \text{for } x_{i,j} < x_{i-1,j}, \\ \frac{100}{k} \sum_{j=1}^{k} \left(\frac{1}{n-1} \sum_{i=2}^{n} \frac{|x_{i,j} - x_{i-1,j}|}{\max [x_{i,j}; x_{i-1,j}]} \right) [\%], & \text{for } x_{i,j} \geq x_{i-1,j}, \end{cases}$$

$$(1)$$

where n determines the number of years on the basis of which the model is being built. For the analyzed data, $n = 13$ was assumed. k determines the number of months on the basis of which the model is being built. For the analyzed data, $k = 12$ was assumed. $x_{i,j}$ is the total number of accidents in the i-th month of year j. i is the subsequent year, $i = 1,2,…,13$. j is the subsequent month, $j = 1,2,…,12$.

Fig. 3 Scatterplots of the total number of people injured in accidents at work in the construction industry, broken down into years and months
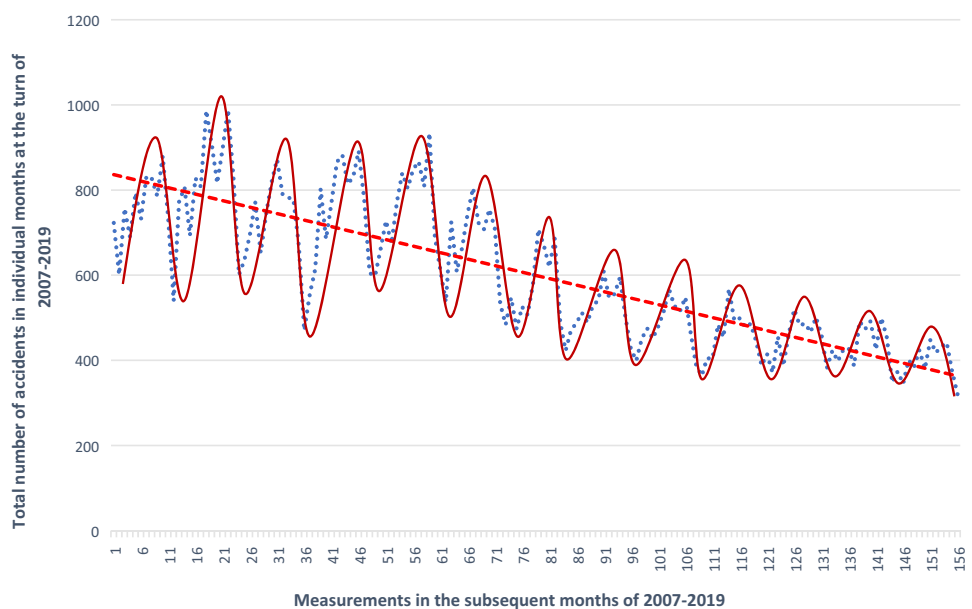
Similarly, the average monthly increase/decrease in the total number of people injured in accidents can be determined as follows:

$$
\Delta m_j =
\begin{cases}
\frac{100\%}{n} \sum_{i=1}^{n} \left(\frac{1}{k-1} \sum_{j=2}^{k} \frac{-|x_{i,j}-x_{i-1,j}|}{\max\,[x_{i,j};x_{i-1,j}]} \right) [\%], & \text{dla } x_{i,j} < x_{i,j-1} \\[2ex]
\frac{100\%}{n} \sum_{i=1}^{n} \left(\frac{1}{k-1} \sum_{j=2}^{k} \frac{|x_{i,j}-x_{i-1,j}|}{\max\,[x_{i,j};x_{i-1,j}]} \right) [\%], & \text{dla } x_{i,j} \geq x_{i,j-1}
\end{cases},
\tag{2}
$$

where n determines the number of years on the basis of which the model is being built. For the analyzed data, $n = 13$ was assumed. k determines the number of months on the basis of which the model is being built. For the analyzed data, $k = 12$ was assumed. $x_{i,j}$ is the total number of accidents in the j-th month of year i. i is the subsequent year, $i = 1,2,\ldots,13$. j is the subsequent month, $j = 1,2,\ldots,12$.

The results of parameters (1) and (2) are presented in Tables 4 and 5. The results multiplied by 100% determine the percentage increase.

Table 6 presents the results of the analysis of the correlation between the values of the average annual increases (Δr_i) and average monthly increases (Δm_j) in the number of people injured in accidents and the subsequent years and months of the year.

The obtained values of the correlation coefficients indicate: a weak correlation between the average annual increases in the total number of people injured in accidents and the subsequent years, and a weak correlation between the average monthly increases in the total number of people injured in accidents and the subsequent months of the year. The obtained results allow for the conclusion that the increases do not have a trend, and are, therefore, comparable with regards to months and years.

Due to the fact that the trend and seasonality are the strongest for 2016 ($R_{10}^2 = 0.782$; $p = 0.007$) (Table 4), a regression curve for this year was determined by approximating the seasonal fluctuations in the accident rate in the construction industry using a polynomial function in the form:

$$
n_{m,2016} = \left(-3.7234 m^2 + 52.969 m + 306.58 \right), m = 1,\ldots,12
\tag{3}
$$

The above function refers to the adopted pattern of seasonal changes in the accident rate in the subsequent months of 2016. Due to the fact that the long-term trend should also be taken into account in the prediction model, the long-term trend coefficient was introduced into formula 3, and expressed as

$$
\beta = 1 + (2016 - r)\Delta_{r,\text{śr}}, [\%],
\tag{4}
$$

Table 4 Average annual increases/decreases $\Delta r_{i,\text{sr}}$ [%] of the total number of accidents from January to December

Month	2008–2007	2009–2008	2010–2009	2011–2010	2012–2011	2013–2012	2014–2013	2015–2014	2016–2015	2017–2016	2018–2017	2019–2018	$\Delta r_{i,\text{sr}} \rightarrow$
1	0.066	− 0.178	− 0.127	0.072	0.023	− 0.209	− 0.027	− 0.153	− 0.088	0.123	0.019	− 0.116	− 0.050
2	0.254	− 0.136	− 0.118	0.071	− 0.188	0.018	− 0.108	− 0.076	− 0.102	− 0.084	0.068	− 0.133	− 0.044
3	− 0.085	0.102	0.039	− 0.093	− 0.004	− 0.346	0.074	− 0.072	− 0.149	0.114	− 0.068	− 0.056	− 0.045
4	0.168	− 0.211	0.047	− 0.016	− 0.098	− 0.135	− 0.070	− 0.078	0.068	− 0.200	0.098	− 0.105	− 0.044
5	0.017	− 0.079	0.034	0.000	− 0.135	− 0.239	0.042	− 0.095	− 0.052	0.017	− 0.162	0.087	− 0.047
6	0.258	− 0.170	0.054	− 0.030	− 0.114	− 0.224	− 0.062	− 0.050	0.095	− 0.085	− 0.054	− 0.228	− 0.051
7	0.082	− 0.039	0.011	− 0.088	0.001	− 0.120	− 0.140	− 0.069	− 0.131	− 0.022	− 0.019	− 0.053	− 0.049
8	− 0.013	− 0.032	0.029	0.041	− 0.148	− 0.072	− 0.185	− 0.035	− 0.055	− 0.024	0.010	− 0.150	− 0.053
9	0.124	− 0.128	0.060	0.040	− 0.183	− 0.126	− 0.102	− 0.074	− 0.072	− 0.025	− 0.088	0.016	− 0.046
10	0.110	− 0.212	0.127	− 0.094	− 0.066	− 0.091	− 0.131	− 0.081	− 0.113	0.030	− 0.002	− 0.116	− 0.053
11	0.115	− 0.190	0.140	0.165	− 0.241	− 0.309	0.020	− 0.076	− 0.037	0.035	− 0.013	− 0.176	− 0.047
12	0.101	− 0.218	0.223	0.122	− 0.226	− 0.206	0.007	− 0.092	0.003	− 0.015	− 0.089	− 0.080	− 0.039
Average value $\Delta r_{\text{śr}}$													− 0.0473

Table 5 Average monthly increases/decreases $\Delta m_{j,\text{sr}}$ [%] of the total number of accidents in the years 2007–2019

Month	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
2–1	− 0.17	0.04	0.09	0.10	0.10	− 0.12	0.12	0.03	0.12	0.10	− 0.11	− 0.06	− 0.08
3–2	0.21	− 0.14	0.10	0.24	0.09	0.26	− 0.13	0.05	0.05	0.00	0.19	0.06	0.14
4–3	− 0.09	0.16	− 0.15	− 0.15	− 0.07	− 0.16	0.10	− 0.04	− 0.05	0.17	− 0.15	0.01	− 0.04
5–4	0.13	− 0.02	0.12	0.11	0.12	0.09	− 0.04	0.08	0.06	− 0.06	0.16	− 0.10	0.09
6–5	− 0.08	0.18	0.09	0.11	0.08	0.10	0.12	0.02	0.07	0.20	0.11	0.21	− 0.10
7–6	0.12	− 0.08	0.06	0.02	− 0.04	0.08	0.19	0.11	0.09	− 0.13	− 0.07	− 0.04	0.15
8–7	− 0.01	− 0.10	− 0.09	− 0.08	0.05	− 0.10	− 0.05	− 0.10	− 0.07	0.01	0.01	0.04	− 0.07
9–8	− 0.05	0.09	− 0.01	0.02	0.02	− 0.02	− 0.08	0.02	− 0.02	− 0.04	− 0.04	− 0.14	0.03
10–9	0.11	0.09	0.00	0.07	− 0.07	0.06	0.10	0.07	0.06	0.02	0.07	0.15	0.02
11–10	− 0.17	− 0.16	− 0.14	− 0.13	0.13	− 0.06	− 0.29	− 0.16	− 0.16	− 0.08	− 0.08	− 0.09	− 0.15
12–11	− 0.26	− 0.27	− 0.30	− 0.22	− 0.26	− 0.25	− 0.14	− 0.15	− 0.16	− 0.13	− 0.17	− 0.24	− 0.15
$\Delta m_{j,\text{sr}}$ ↓	− 0.023	− 0.020	− 0.022	0.008	0.014	− 0.011	− 0.009	− 0.007	− 0.002	0.004	− 0.008	− 0.017	− 0.013

Table 6 Correlation statistics of annual Δr and monthly Δm increases

	Rho Spearman	p	Min (%)	Max (%)	Average (%)
Δr	0.258	0.394	3.93	5.33	4.75
Δm	− 0.112	0.729	1.40	2.30	0.82

$$n_{r,m} = \left[1 + (2016 - r)\Delta_{r,\text{sr}}\right] \\ \times \left(-3.7234m^2 + 52.969m + 306.58\right), \qquad (5) \\ r = 1, \ldots, 13; \ m = 1, \ldots, 12$$

where for empirical data, $\Delta r_{\text{sr}} = 0.0473\%$ (as shown in Table 6) is the average value of Δr_i. This coefficient shows how many percent the value of the prediction changes for each month of any year in relation to 2016. Finally, a prediction model of the total number of people injured in occupational accidents in the construction industry $n_{r,m}$ was obtained for year r and month m, and is expressed by the following formula:

4 Results

Using the proposed prediction model expressed by Eq. (5), the prediction error was determined as the difference between the real value of the total number of people injured in accidents in the construction industry and the predicted value obtained from the model. The prediction error values for individual years and months are summarized in Tables 7 and 8, and also in Fig. 4. In Table 7, the average value of the prediction error for each year, as well as other statistics, are calculated for 12 months. In turn,

Table 7 Statistics showing the values of prediction errors for individual years

Year r	n	Average value (%)	Median (%)	Minimum (%)	Maximum (%)	Standard deviation (%)	Standard error (%)
2007	12	13.84	14.28	4.37	29.75	7.90	2.28
2008	12	24.33	25.29	6.77	36.56	9.02	2.61
2009	12	18.47	19.07	6.36	25.45	5.99	1.73
2010	12	23.09	23.96	13.76	33.17	6.61	1.91
2011	12	27.20	26.82	15.75	41.87	6.02	1.74
2012	12	20.46	20.30	9.35	30.81	7.64	2.20
2013	12	10.72	9.95	0.33	22.73	7.58	2.19
2014	12	6.72	4.66	0.75	17.28	5.56	1.61
2015	12	5.29	5.46	0.07	11.32	3.56	1.03
2016	12	3.95	3.34	0.39	13.67	3.66	1.06
2017	12	6.92	6.56	1.23	18.33	5.62	1.62
2018	12	9.70	9.54	2.51	23.87	5.93	1.71
2019	12	5.96	4.83	0.32	18.42	5.18	1.50

Table 8 Statistics showing the values of prediction errors for individual months without being divided into years

Month	n	Average value (%)	Median (%)	Minimum (%)	Maximum (%)	Standard deviation (%)	Standard error (%)
1	13	20.71	18.42	2.25	36.56	9.55	2.65
2	13	12.74	10.78	1.17	31.92	9.79	2.71
3	13	14.82	9.77	3.96	30.87	9.97	2.76
4	13	8.16	6.17	0.33	23.52	6.56	1.82
5	13	10.57	10.60	0.99	23.12	7.42	2.06
6	13	13.82	10.18	0.07	31.44	10.81	3.00
7	13	15.05	15.19	0.39	27.89	10.07	2.79
8	13	11.86	15.07	0.93	28.28	9.16	2.54
9	13	12.28	11.09	0.77	31.24	10.66	2.96
10	13	20.06	20.64	4.54	35.27	9.53	2.64
11	13	13.91	12.74	0.32	41.87	13.15	3.65
12	13	9.08	8.09	1.23	27.07	6.42	1.78
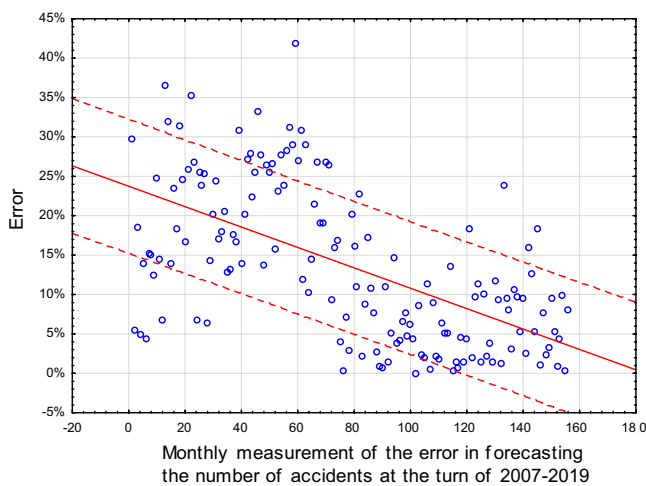
(a) Error scattering plot for subsequent measurements
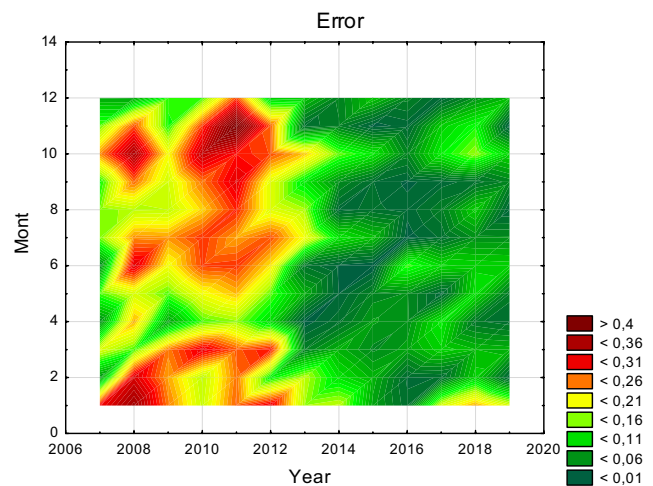
(b) Matrix of the size of the year/month prediction errors

Fig. 4 The trend of forecast errors in successive measurement points and years

in Table 8, the values are calculated for the same month in individual years.

The mean prediction error obtained for all the data is 13.59%, with a standard deviation of 9.95%. The lowest prediction error is 0.07%, and the highest is 41.87%.

Figure 4 shows (a) a scatterplot of the prediction errors for all the subsequent measuring points, and (b) a plane plot showing the distribution of prediction errors broken down into months and years.

From the scatterplot of the prediction error presented in Fig. 4a, it can be concluded that the prediction error decreases with each subsequent measurement, and since 2014 has been at the level of 5–10%. The chart showing the prediction error broken down into years and months, which is presented in Fig. 4b, shows that the prediction error for each month is comparable for individual years. However,

this error decreases in subsequent years. Based on the obtained results, it can be concluded that in the years after 2014 the prediction error is significantly lower than in the years before 2014. The prediction error after 2016 enables the total number of people injured in accidents to be forecasted with high accuracy.

5 Discussion

The results of the analysis of the number of accidents in the construction industry in Poland in the years 2007–2019, which are included in this study, and the results of studies conducted in other parts of the world, confirmed that it is a seasonal phenomenon. The published statistical data confirm that the number of accidents in the summer season is

greater than in the winter months [1, 2, 12, 13]. In the graphs presented in Fig. 2, a similar pattern of variability in the number of people injured in occupational accidents in the construction industry is repeated, which proves the seasonal and cyclical nature of this phenomenon. Similar conclusions with regards to seasonality in the construction industry were formulated by Mach et al. in [18]. In long-term studies concerning accident rates in many European and world countries, the formation of long-term trends in accident rates can be noticed. Thanks to the "zero accidents" action carried out in the construction industry around the world, there is a downward long-term trend in many countries. Both these phenomena, namely seasonality and cyclicality, as well as the long-term trend, were included in the developed model of predicting the number of people injured in occupational accidents in the construction industry.

A model based on non-linear regression assumptions was developed. The developed model was verified by comparing the values calculated on the basis of the model with the empirical results. The average prediction error obtained for all the data is 13.59%, with a standard deviation of 9.95%. The lowest prediction error is 0.07%, and the highest is 41.9%. When analyzing graphs 4a and b, it can be noticed that the prediction error decreases when approaching the end of the analyzed time interval. It can be assumed that in the years after 2014 the prediction error is significantly lower than in the years before 2014.

The developed model was verified by comparing the prediction calculated on the basis of the model with the results obtained using the following methods: linear regression, polynomial regression, as well as the ARIMA and SARIMA methods. In the calculations carried out with the above-mentioned methods, the experimental protocol was adopted as a fivefold cross-validation. Within the same data set, fourfold was used for learning, and one was used for testing and validation. Thus, in each method, the learning set accounted for 80% of the samples, with 20% of the samples being included in the testing set. The results of prediction errors obtained using verifying methods, i.e., linear regression, polynomial regression, and ARIMA and SARIMA, are presented in Table 9.

Using the Wilcoxon test, the significance of the differences of the prediction errors between the different base models was assessed at the level of statistical significance ($\alpha = 0.05$). For example, for model (e), the obtained results were a, b and d. This means that the prediction error of 8.9% for the authors' model is statistically significantly ($p < 0.05$) lower than in the case of using: a—linear regression, for which the prediction error amounts to 32%; b—polynomial regression, for which the prediction error amounts to 28%; and d—the SARIMA model, for which the prediction error is equal to 11.7%. This result can be compared with the result obtained using the ARIMA method, in which the prediction error is equal to 10%. For the considered task, the developed model turned out to be better than the models of linear and polynomial regression, and also the SARIMA model. The SARIMA and ARIMA models provided similar results and were better than the methods that are based on the least squares method.

For the forecast covering the period from 2014 to 2019, the prediction error for the developed model was equal to 5.45%. The ARIMA model generated an error of 7.89%. The other methods generated the following prediction error values: SARIMA—8.6%, RL—26%, and RW—23%. The model for predicting the number of people injured in occupational accidents in the construction industry, which was developed and presented in this article, reflects well the attributes of the accidentality phenomenon in the construction industry, such as the long-term trend, seasonality and cyclicality.

When building the model, data concerning the number of accidents in the construction industry in individual months of the year were used. An accident situation is more accurately reflected in the indicator that refers to "the number of accidents in relation to the number of employees". However, in Poland, precise data on the number of employees in the construction industry in particular months of the year are not available. The lack of these data limit our research to only the number of accidents.

6 Conclusions

The aim of the research that was undertaken and described in this article was to develop a model for predicting the number of people injured in occupational accidents in the construction industry based on statistical data published by the

Table 9 Statistics of prediction errors for individual methods—provided with the results of the Wilcoxon test

Type of model	a	b	c	d	e
	RL	RW	ARIMA	SARIMA	MODEL
Prediction error	32% (28–36%)	28% (26–35%)	10% (9–13%)	11.7% (9–14%)	8.9% (7–11%)
Wilcoxon test	–	a	a, b	a, b	a, b, d

RL—linear regression; RW—6th-order polynomial regression (for a larger degree of the polynomial, an excessive adjustment to the data was obtained)

Central Statistical Office in Poland. The model was built on the basis of data that covered the period from 2007 to 2019, and the commonly known mathematical tools available in the STATISTICA software were used to solve the given task.

Based on the analysis of statistical data and previous studies, the occurrence of certain regularities of the accidentality phenomenon was found, such as a decreasing long-term trend over many years, as well as seasonality and cyclicality of the phenomenon over the course of a year. The found regularities were the basis for the assumptions that were made for the purpose of building the model.

Prediction errors were calculated for the data obtained from the model. The obtained results of the prediction errors indicate that the model reflects well the accidentality phenomenon in the construction industry, and is better than the linear and polynomial regression models and the ARIMA and SARIMA models. The average prediction error for the developed model was equal to 8.9%. The information obtained on the basis of the model concerning the expected number of people injured in accidents in particular months of subsequent years may be the basis for making decisions in the field of legislative, educational, prophylactic and preventive actions, as well as for estimating the costs associated with accidents. This in turn may have an impact on shaping prices in the construction industry, as well as on the financial and insurance policy of enterprises. The gained information can also be used to estimate the probability of an occurrence of an accident in the future, and the knowledge obtained as a result of the prediction can be the basis for managing occupational safety in the construction industry.

The model presented in the article was built on the basis of data concerning the number of people injured in occupational accidents in the entire construction branch. Therefore, an analysis of the model's correctness was performed for data covering the entire construction industry. However, the model is universal and can be used to forecast accident rates in individual subsections, e.g., industrial and residential constructions, road infrastructure, bridges, and others.

Declarations

Conflicts of interest   Not applicable.

Ethics approval   Not applicable.

Consent to participate   Not applicable.

Consent for publication   All the authors have read and agreed to the published version of the article.

References

1. Ling FYY, Liu M, Woo YC. Construction fatalities in Singapore. Int J Project Manage. 2009;27(7):717–26. https://doi.org/10.1016/j.ijproman.2008.11.002.

2. Dumrak J, Mostafa S, Kamardeen I, Rameezdeen R. Factors associated with the severity of construction accidents: the case of South Australia. Aust J Constr Econ Build. 2013;13(4):32–49. https://doi.org/10.5130/ajceb.v13i4.3620.

3. López Arquillos A, Rubio Romero JC, Gibb A. Analysis of construction accidents in Spain, 2003–2008. J Saf Res. 2012;43(5–6):381–8. https://doi.org/10.1016/j.jsr.2012.07.005.

4. Elvik R. Can injury prevention efforts go too far? Reflections on some possible implications of Vision Zero for road accident fatalities. Accid Anal Prev. 1999;31:265–86. https://doi.org/10.1016/S0001-4575(98)00079-7.

5. Szer I, Błazik-Borowa E, Szer J. The influence of environmental factors on employee comfort based on an example of location temperature. Arch Civ Eng. 2017;LXII:3. https://doi.org/10.1515/ace-2017-0035.

6. Hola B, Nowobilski T, Szer I, Jacek Szer J. Identification of factors affecting the accident rate in the construction industry. Procedia Eng. 2017;208:35–42.

7. Lipecki T, Jaminska-Gadomska P, Bec J, Blazik-Borowa E. Facade scaffolding behaviour under wind action. Arch Civ Mech Eng. 2020;20:1.

8. Jablonski M, Szer I, Szer J. Probability of occurrence of health and safety risks on scaffolding caused by noise exposure. J Civ Eng Manag. 2018;24(6):437–43.

9. European Statistical Recovery Dashboard https://ec.europa.eu/eurostat/data/database.

10. Occupational Safety and Health Administration. 2020. https://www.osha.gov/. Accessed 4 Nov 2020.

11. Szer I, Szer J, Kaszubska M, Miszczak J, Hoła B, Błazik-Borowa E, Jabłoński M. Influence of the seasons on construction site accidents. Arch Civ Eng. 2021;67:3.

12. Liao CW. Pattern analysis of seasonal variation in occupational accidents in the construction industry. Procedia Eng. 2012;29:3240–4. https://doi.org/10.1016/j.proeng.2012.01.473.

13. Pierce B. The seasonal timing of work-related injuries, pp 2371–2381. 2013. http://www.bls.gov/tus/.

14. Kang K, Ryu H. Predicting types of occupational accidents at construction sites in Korea using random forest model. Saf Sci. 2019;120:226–36. https://doi.org/10.1016/j.ssci.2019.06.034.

15. Traczyk WZ, Trzebski A. Human physiology with elements of applied and clinical physiology (in polish). PZWL, Warszawa. 2001.

16. Meng WL, Shen S, Zhou A. Investigation on fatal accidents in Chinese construction industry between 2004 and 2016. Nat Hazards. 2018;94(2):655–70. https://doi.org/10.1007/s11069-018-3411-z.

17. Granger WJ. Seasonality: causation, interpretation, and implications. 1979. http://www.nber.org/chapters/c3896. Accessed 11 May 2021.

18. Mach L, Zmarzly D, Dabrowski I, Fracz P. Comparison on sub-annual seasonality of building construction in European countries. Eur Res Stud J. 2020;2:4. https://doi.org/10.35808/ersj/1682.

19. Kim DK, Park S. Business cycle and occupational accidents in Korea. Saf Health Work. 2020;11:314–21. https://doi.org/10.1016/j.shaw.2020.05.002.

20. Ruiz-Fernandez JP, Marco JB, Lopez MA, Gascuena NV. Influence of seasonal factors in the earned value of construction. Appl Math Nonlinear Sci. 2019;4(1):21–34. https://doi.org/10.2478/AMNS.2019.1.00003.

21. Fabiano B, Parentini I, Ferraiolo A, Pastorino R. A century of accidents in the Italian industry: relationship with the production cycle. Saf Sci. 1995;21:65–74. https://doi.org/10.1016/0925-7535(95)00043-7.

22. Dong XS, Fujimoto A, Ringen K, Men Y. Fatal falls among Hispanic construction workers. Accid Anal Prev. 2009;41:1047–52. https://doi.org/10.1016/j.aap.2009.06.01.

23. Hoła B, Szóstak M. Analysis of the state of the accident rate in the construction industry in European Union Countries. Arch Civ Eng. 2015. https://doi.org/10.1515/ace-2015-0033.

24. Hola B, Szostak M. Methodology of analysing the accident rate in the construction industry. In: Conference: 12th international conference on modern building materials, structures and techniques (MBMST), Vilnius, Lithuania, 26–27, 2016 Procedia engineering, vol 172; 2017. pp. 355–362.

25. Lee JY, Yoon YG, Oh TK, Park SS. A study on data pre-processing and accident prediction modelling for occupational accident analysis in the construction industry. Appl Sci. 2020;10(21):7949. https://doi.org/10.3390/app10217949.

26. Zhang J, Zhang W, Peixui X, Chen N. Applicability of accident analysis methods to Chinese construction accidents. J Saf Res. 2018;68:4. https://doi.org/10.1016/j.jsr.2018.11.006.

27. Wingea S, Albrechtsena E, Mostueba BA. Causal factors and connections in construction accidents. Saf Sci. 2019;112:130–41.

28. Hoła B. Methodology of estimation of accident situation in building industry. Arch Civ Mech Eng. 2012;9:29–46.

29. Andolfo C, Sadeghpour F. A Probabilistic accident prediction model for construction sites. Proc Procedia Eng. 2015;123:15–23.

30. Nguyen LD, Tran DQ, Chandrawinata MP. Predicting safety risk of working at heights using Bayesian networks. J Constr Eng Manag. 2016;142:04016041.

31. Le Q-T, Park C-S Construction safety education model based on second life. In: Proceedings of the proceedings of IEEE international conference on teaching, assessment, and learning for engineering (TALE) 2012; Hong Kong; 2012. pp. H2C–1–H2C–5.

32. Klempous R, Kluwak K, Idzikowski R, Nowobilski T, Zamojski T. Possibility analysis of danger factors visualization in the construction environment based on Virtual Reality model. In: Proceedings of the 8th IEEE international conference on cognitive infocommunications, CogInfoCom 2017—proceedings; vol. 2018; 2018. pp. 000363–000368.

33. Isaac S, Edrei T. A statistical model for dynamic safety risk control on construction sites. Autom Constr. 2016;63:66–78.

34. Sarkar S, Raj R, Vinay S, Maiti J, Pratihar DK. An optimization-based decision tree approach for predicting slip-trip-fall accidents at work. Saf Sci. 2019;118:57–69.

35. Topolski M. Algorithm of multidimensional analysis of main features of PCA with blurry observation of facility features detection of carcinoma cells multiple myelom. Progress Comput Recogn Syst. 2020;2020:286–94.

36. Topolski M. Application of the stochastic gradient method in the construction of the main components of PCA in the task diagnosis of multiple sclerosis in children computational sciences—ICCS 2020; 35–44. In: Yurekli K, Ability of stochastic modelling to forecast crop reference evapotranspiration. Irrigation and Drainage. 2021. https://doi.org/10.1002/ird.2598.

37. Eddine BI, Guastaldi E, Zirulia A, Bengusmia M, Brancale D. Trend analysis and spatiotemporal prediction of precipitation, temperature, and evapotranspiration values using the ARIMA models: case of the Algerian Highlands 2021. Arab J Geosci. 2021;13:24. https://doi.org/10.1007/s12517-020-06330-6.

38. Brockwell P, Davis R. Introduction to time series and forecasting. Berlin: Springer; 2002.

39. Freedman DA. Statistical models: theory and practice. Cambridge: Cambridge University Press; 2009.

40. Główny Urząd Statystyczny, Wypadki przy pracy. 2020. https://stat.gov.pl/publikacje/publikacje-a-z/szukaj.html?letter=W&page=5.

41. Kendall MG. Rank correlation methods. London: Charles Griffin & Company Limited; 1948.

42. Koronacki J, Mielniczuk J. Statystyka dla studentów kierunków technicznych i przyrodniczych. Warszawa: WNT. 2006. ISBN 83-204-3242-1.

43. Stanisz A. Przystępny kurs statystyki z zastosowaniem statistica PL na przykładach z medycyny. Tom1. Statystyki podstawowe, StatSoft Polska Sp. z o.o., Kraków. 2006.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.