



## ARIMA MODEL USED TO ANALYZE THE DEMAND FOR SWIMMING POOL SERVICES

Mieczysław Pawlisiak

### Abstract

Many factors influence customer preferences among those who choose active leisure. A wide range of market products makes for many opportunities, and sports facilities are required to be fully prepared to provide services. It is helpful to create forecasts that enable to determine the predictable number of clients. An example prediction made with respect to swimming pools is presented in this article. For this purpose, the ARIMA model was used, based on the assumption that the value of the endogenous variable is affected by the value of this variable lagged in time.

**Keywords:** ARIMA model, lifeguarding and water rescue, readiness, swimming pool, swimming baths, safety

**JEL classification:** C2, C22.

### Introduction

The knowledge of demand formation is an extremely useful element in the functioning of any enterprise. It not only allows for a better adjustment to the needs reported by the market, but is also an excellent tool for gaining a competitive advantage. Forecasts also allow shaping an important element of each company, which is its readiness to perform specific activities. This applies not only to civilian-run enterprises operating within a market economy, but also to state institutions, in which readiness to perform tasks is one of the most important parameters (Borucka, 2018). The concept of readiness is usually identified with the exploitation of technical objects (Waśniewski, Borucka, 2011; Borucka, 2018) and expressed as a probability that the object will be ready to fulfill its tasks in a given moment or in a given period of time. The literature on the subject examines the readiness of individual elements of machines and devices (Świdorski, Borucka, 2018), vehicles (Borucka, 2013) or entire systems. Particularly frequently considered are those that require quick and appropriate actions such as emergency

medical services or the fire brigade, but also those that operate on the basis of fixed timetables, such as passenger transport (Borucka, 2018) or delivery of goods (Bielińska, 2007, Borucka, 2018). As regards the flow of means of transport, the influence of factors that may hinder it is also analyzed, e.g. road accidents (Czyżycki, Hundert, Klóska, 2007; Świdorski, Borucka, Jacyna-Golda, Szczepański, 2019; Borucka, 2014) or congestion in urban traffic (Borucka, 2018). Readiness applies to the facilities, systems as well as the staff operating them. The degree of competence and preparation of employees affects the reliability of the entire system (Dittmann, Szabebela-Pasierbińska, Dittmann, Szpulak, 2011). In facilities such as swimming pools, the level of readiness determines the speed of response to a threat by lifeguards on continuous watch, i.e. persons with knowledge and skills in water rescue and swimming techniques as well as other qualifications useful in this line of duty, (Świdorski, Skoczyński, Borucka, 2018).

## 1. Introduction to the ARIMA model

A sequence of information ordered in time is a time series. Autoregressive models are some of the possible forecasting models, based on time series analysis, (1), comprising a group for which it is assumed that there is a relationship between the values of a time series at a given moment and the prior values of this series, distant from each other by a certain time interval (Sokołowski, 2016).

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, \varepsilon_t) \quad (1)$$

Therefore, the value of the forecast variable is estimated on the basis of its own components, which are distant in time. They are a function, i.e. they depend on the value of the variable  $y$  in previous periods and on the random component. Such models have restrictions in their applicability. They can only be used for stationary series or series converted into stationary form by means of specific transformations (Borucka, 2014; Żurek, Ziółkowski, Borucka, 2017). For series characterized by a variable average value, variance or slope, the operation enabling the reduction to stationary form is the procedure of  $d$ -fold differentiation, i.e.  $d$ -fold calculation of differences in adjacent terms of the series. These series are referred to as integrated series and can be represented by autoregressive models which include:

- autoregressive models of the AR( $p$ ) type,
- moving average model MA( $q$ ),
- ARMA (Autoregressive moving average) models,
- ARIMA (Autoregressive integrated moving average) models.

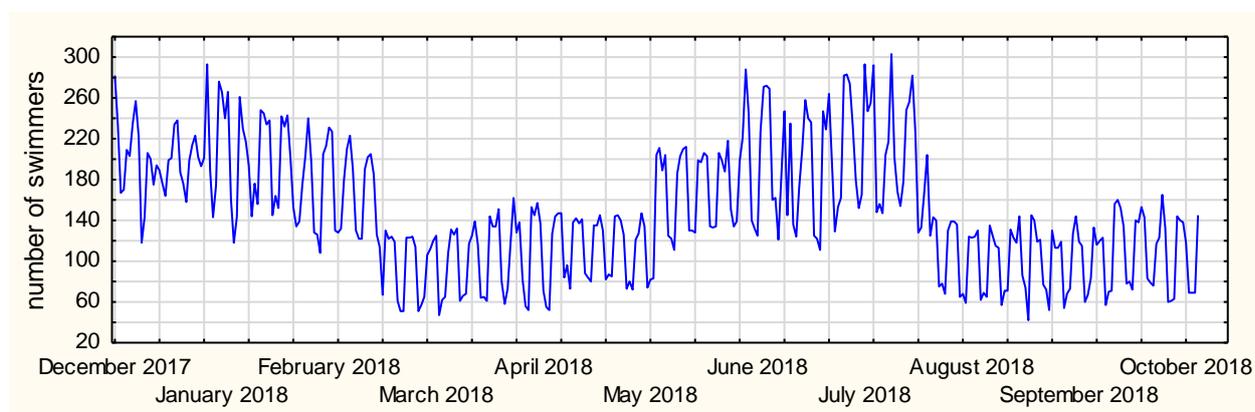
In the AR autoregressive model, a relation occurs between the values of the forecast variable and its values lagged in time. Its estimation consists in determining the parameter  $p$  (order of autoregression), which informs how far one should reach into the past when selecting exogenous variables in the model. If the value of the examined series is correlated with its previous value, then we are dealing with a first order autoregressive model - AR(1). In the MA moving average process, the values of the endogenous variable are expressed as a function of the lagged values of the stationary random component. The parameter  $q$  of this process, i.e. the order of the MA process, indicates the level of lags adopted for the model (Borucka, 2018).

The combination of these processes results in the ARMA process, which allows for greater efficiency and flexibility in adapting the model to the time series. Its construction is based on the assumption that the value of the forecast variable at time  $t$  depends both on its past values and on the differences between the past actual values of the forecast variable and its values obtained from the model (forecast errors). When the analyzed series is non-stationary but has been converted to stationary by determining the differences of the appropriate order, called the process integration order, the ARIMA model is obtained. The use of the letter I means that the series has been subjected to the  $d$ -fold differentiation operation. The presented models are short-

term forecasts and are one of the most important forecasting tools (Żurek, Ziółkowski, Borucka, 2017).

## 2. Characteristics of the tested subject

The tested subject is one of the swimming pools in Warsaw, Poland. The information provided refers to the number of its customers. It was gathered between December 2017 until mid-December 2018. The last 14 observations were not included in the estimation, but were left as test observations in order to verify the model. Missing data, resulting from the national and religious holidays, were replaced by an average value calculated for specific days of the week in individual months, which resulted from the weekly and monthly seasonality visible on the graph (Fig. 1).



**Figure 1:** Daily number of customers at the swimming pool in the period: December 2017 – November 2018.

Source: author's own work.

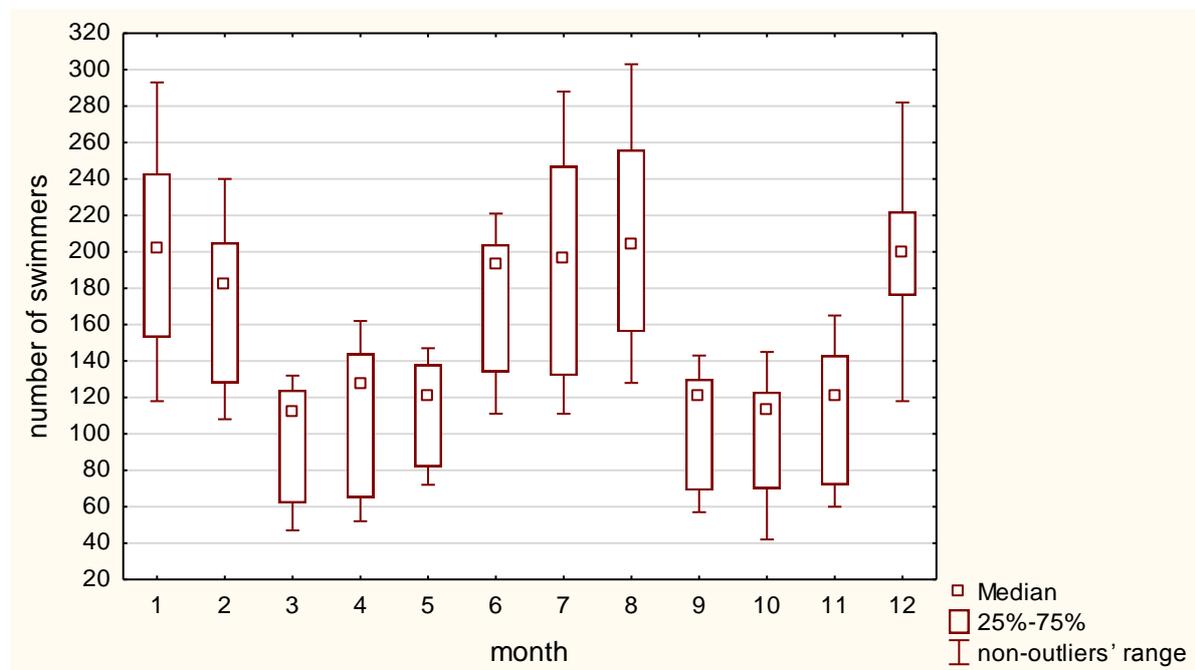
First, a visual analysis of the graph (Fig. 1) was carried out. It seemed that the best solution would be a multiple regression model, but after its preparation and investigation, it turned out that the distribution of residuals was not consistent with the normal distribution and that dependencies unexplained by the model occurred in it (significant correlations on the graph of autocorrelation and partial autocorrelation functions). Therefore, another method of forecasting, using the ARIMA model, is proposed in this article. Since the seasonality of the process is clearly visible in on the graph (Fig. 1), it was decided to divide the year into two periods, determining the high season, during which the number of customers is much higher than the average value and the low season, characterized by a lower daily frequency of attendance. The gathered number of observations allowed for such a procedure, because the ARIMA model requires a minimum of 60 observations. In addition to the basic measures of descriptive statistics contained in the table 1, the frame diagram presented in Fig. 2 showing the monthly seasonality of the process also proved to be helpful.

**Table 1:** Basic measures of descriptive statistics for the entire sample.

Month	Number of observations	Mean [number of persons]	Median [number of persons]	Minimum [number of persons]	Maximum [number of persons]	Standard deviation [number of persons]	Coefficient of variation [%]
The entire sample	365	149	138	42	344	61.31	41.24

Month	Number of observations	Mean [number of persons]	Median [number of persons]	Minimum [number of persons]	Maximum [number of persons]	Standard deviation [number of persons]	Coefficient of variation [%]
January	31	203	202	118	293	50.14	24.76
February	28	171	183	108	240	42.70	25.00
March	31	95	112	47	132	31.81	33.40
April	30	110	128	52	162	39.58	36.08
May	31	110	121	72	147	28.88	26.26
June	30	174	193	111	221	37.14	21.32
July	31	198	196	111	344	65.16	32.91
August	31	209	204	128	303	55.70	26.60
September	30	104	121	57	143	31.45	30.23
October	31	99	113	42	145	32.10	32.35
November	30	111	121	60	165	36.83	33.19
December	31	199	200	118	282	33.35	16.76

Source: author's own work.



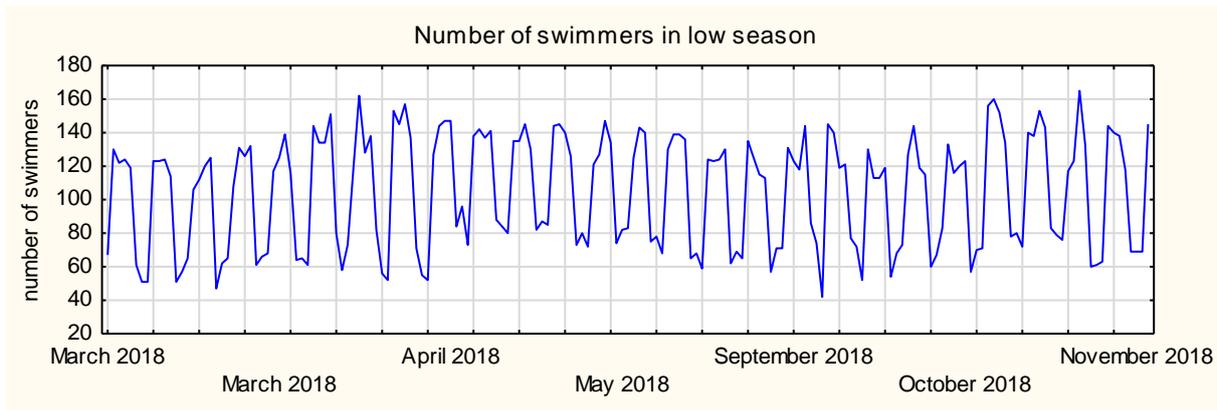
**Figure 2:** Monthly frame diagram for the entire sample.

Source: author's own work.

On the basis of the above analyses, it was decided to divide the studied period into two seasons. The following months were qualified for the low season: March, April, May, September, October and November. The high season included: January, February, June, July, August and December. Each period is described in a separate model.

### 3. Low season assessment

The first stage of the study is a visual analysis of the low season observation graph (Fig. 1 3)



**Figure 3:** Number of customers in low season.

Source: author’s own work.

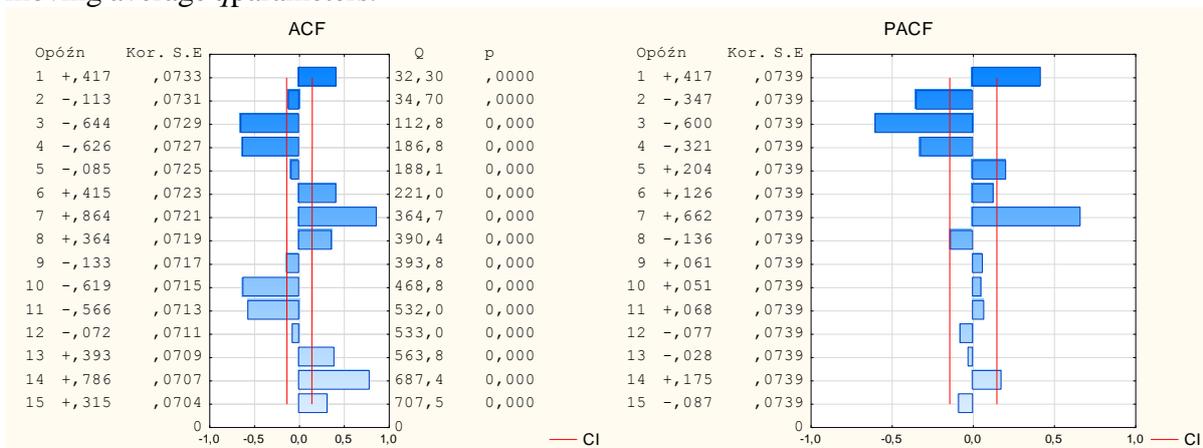
The graph presents weekly fluctuations, but their level seems to be constant over time. In such a defined sample the coefficient of variation and standard deviation decreased in comparison to the results for all observations (table 2).

**Table 2:** Values of selected measures of descriptive statistics for the low season.

Descriptive statistics (low season)						
Number of observations	Mean [number of persons]	Median [number of persons]	Minimum [number of persons]	Maximum [number of persons]	Standard deviation [number of persons]	Coefficient of variation [%]
183	104	11	42	165	33.66	32.12

Source: author’s own work.

A sample prepared in such a way allows estimating the parameters of the ARIMA model. Helpful in determining their type and number are the diagrams of autocorrelation (ACF) and partial autocorrelation (PACF), presented in the Figure 4. The sinusoidal shape of the ACF and the high values of the PACF suggest a positive value of both the autoregressive  $p$  and the moving average  $q$  parameters.



**Figure 4:** ACF and PACF graph for the low season.

Source: author’s own work.

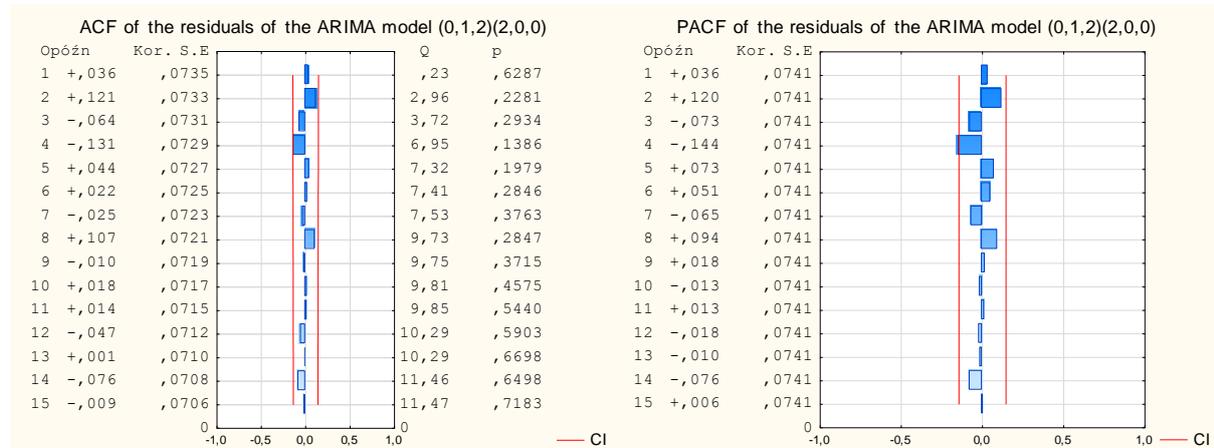
Several models were estimated and on the basis of the evaluation of statistical significance of their parameters and the value of the MS error, the best one was selected, described by two moving average parameters (q) and two seasonal autoregressive parameters (Ps) - ARIMA (0.1.2)(2.0.0). In order to eliminate the diagnosed weekly seasonality, a seasonal differentiation with a lag equal 7 was carried out. The results are presented in the Table 3.

**Table 3:** Parameters of the ARIMA model (0,1,2)(2,0,0)

Parameter	Model:(0,1,2)(2,0,0) Seasonal lag: 7 MS= 247.30					
	Parameter value	Standard error	t( 178)	p	Lower limit 95% CI	Upper limit 95% CI
q(1)	0.70	0.07	10.51	0.00	0.56	0.83
q(2)	0.29	0.07	4.33	0.00	0.16	0.42
Ps(1)	0.66	0.07	8.82	0.00	0.51	0.81
Ps(2)	0.27	0.08	3.58	0.00	0.12	0.42

Source: author's own work.

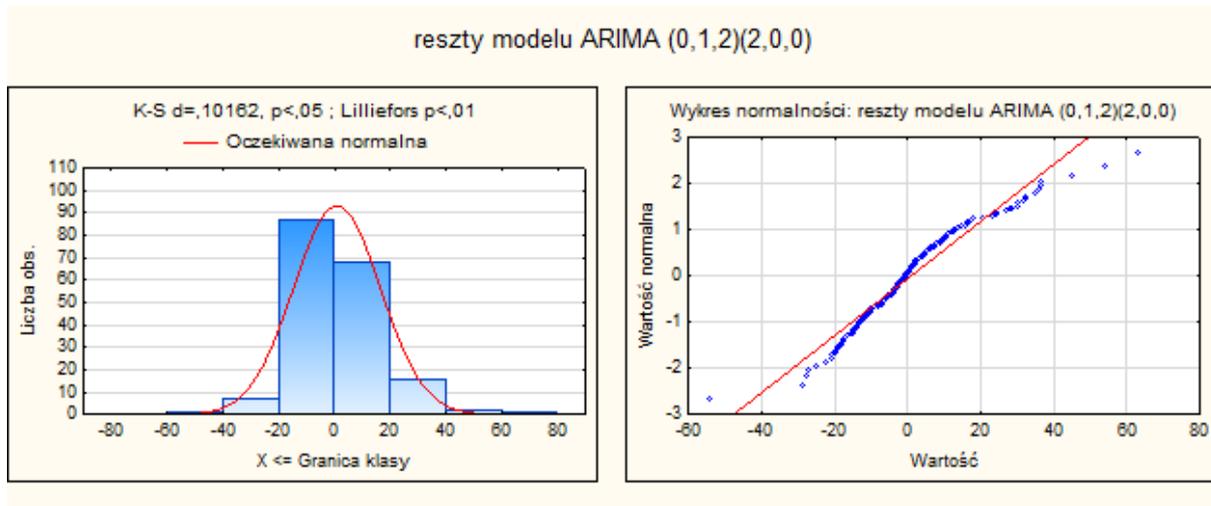
Diagnostics of the developed model consists mainly in checking whether there are any significant correlations between individual lags in the residuals of the model. For this purpose, the autocorrelation and partial autocorrelation function was drawn up again, this time for the residuals of the model (Fig. 5).



**Figure 5:** ACF and PACF graphs for the residuals of the ARIMA model (0,1,2)(2,0,0).

Source: author's own work.

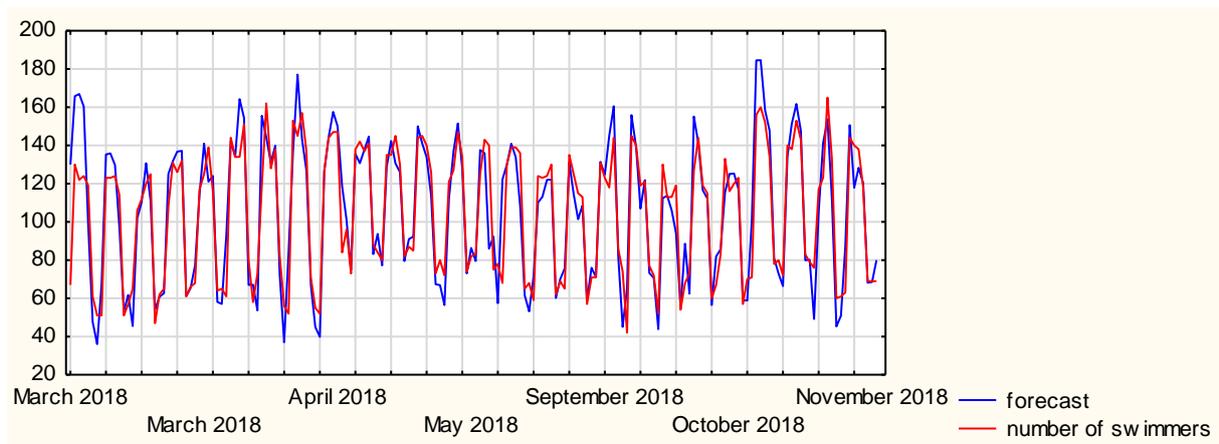
The criterion of normality of residuals was also checked. Unfortunately, the Shapiro–Wilk test did not confirm this feature of the distribution (Fig. 5)



**Figure 6:** Normality test of residuals of the ARIMA model (0,1,2)(2,0,0).

Source: author's own work.

The non-normality of the distribution results from numerous overestimations of the forecast visible in Fig. 6. The forecast works much better on days when the number of swimmers is lower. However, the average forecast error equals 11 people. Such a value makes it possible to consider the model useful in view of the assumed objectives of adjusting the number of lifeguards to the number of swimmers.

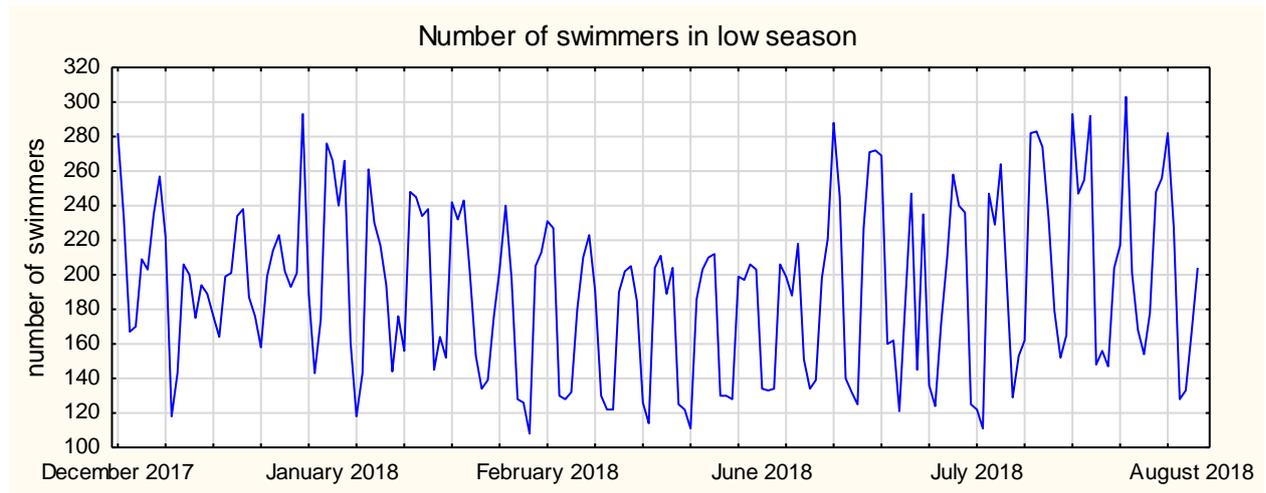


**Figure 7:** Graph of empirical and projected values for the ARIMA model (0,1,2)(2,0,0).

Source: author's own work.

#### 4. High season assessment

Similarly to the low season, the first stage of the study of high season is its visual analysis (Fig. 7).



**Figure 8:** Number of customers in high season.

Source: author's own work.

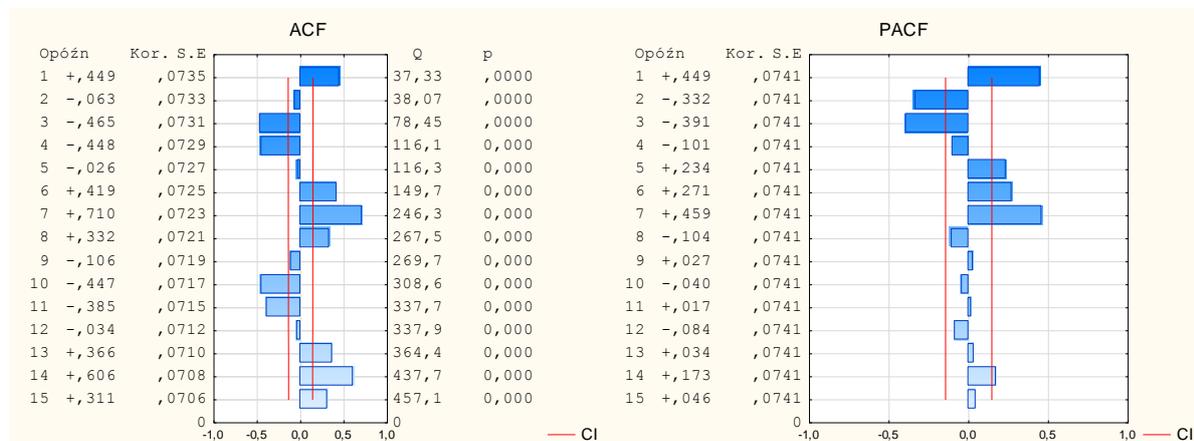
The high season graph also shows weekly fluctuations of a relatively constant level over time. This is the information to be included in the model, performing seasonal differentiation with a lag equal 7. For the second subsample the basic measures of descriptive statistics were determined, presented in Table 3. The value of the coefficient of variation is lower than in the low season, but the standard deviation has increased. The mean value for the high season increased by more than 80%, which confirms the legitimacy of such a division.

**Table 4:** Values of selected measures of descriptive statistics for the high season.

Descriptive statistics (high season)						
Number of observations	Mean [number of persons]	Median [number of persons]	Minimum [number of persons]	Maximum [number of persons]	Standard deviation [number of persons]	Coefficient of variation [%]
182	192	197	108	303	48.78	25.39

Source: author's own work.

In order to support the process of estimation of ARIMA model parameters, the ACF and PACF was determined again, this time for the high season 8.



**Fig. 9.** ACF and PACF graph for the high season.

Source: author’s own work.

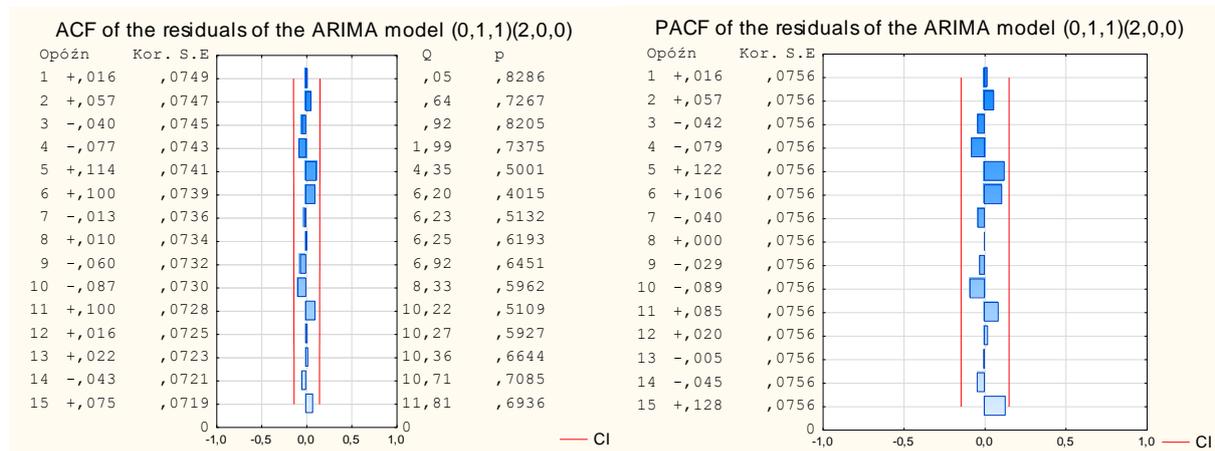
The shape of the ACF is very similar to the low season graph, but the values of individual lags are smaller. The PACF is also similar, the layout of statistically significant lags is slightly different and their value is also lower than in the low season. Therefore, the model will also consist of both moving average parameters  $q$  and autoregressive parameters  $p$ . The best of the models was selected again on the basis of evaluation of statistical significance of their parameters and the value of the MS error. In order to even out the variance, the differentiation was made with a lag equal 7. After this procedure the ARIMA model (0,1,1)(2,0,0) was selected, described by one moving average parameter ( $q$ ) and two seasonal autoregressive parameters ( $P_s$ ). 4.

**Table 5:** Parameters of the ARIMA model (0,1,1)(2,0,0).

Parameter	Model:(0,1,1)(2,0,0) Transformations: D(7) Seasonal lag: 7 MS= 975.03					
	Parameter value	Standard error	t( 178)	p	Lower limit 95% CI	Upper limit 95% CI
q(1)	-0.27	0.07	-3.76	0	-0.41	-0.13
Ps(1)	-0.44	0.07	-5.93	0	-0.59	-0.3
Ps(2)	-0.27	0.08	-3.46	0	-0.43	-0.12

Source: author’s own work.

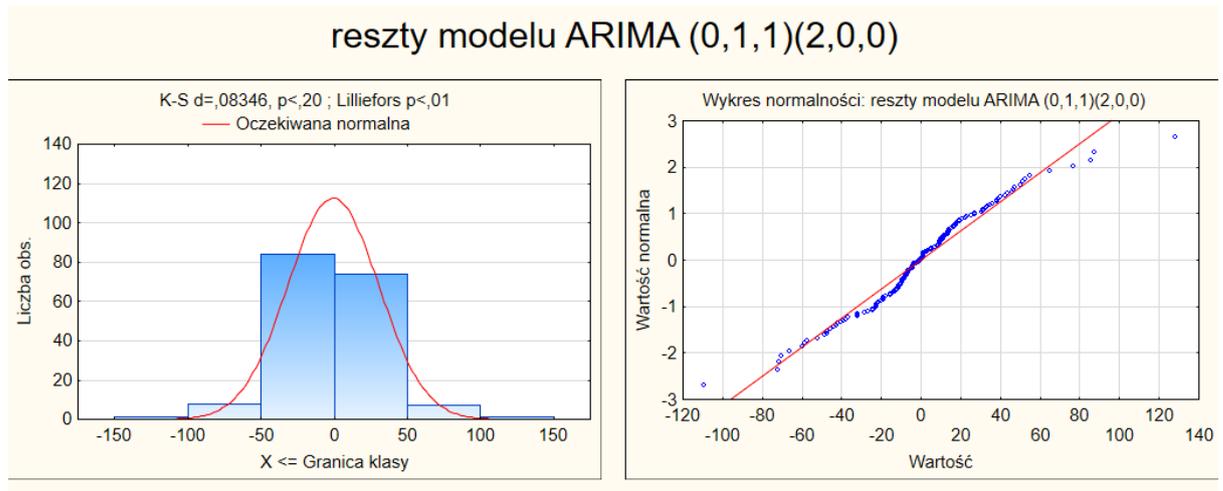
The most important element of diagnostics is, of course, the analysis of the ACF and PACF graphs of the residuals of the constructed model, in order to check the statistical significance of lag values. They could indicate existing dependencies unexplained by the model (Fig. 9).



**Figure 10:** ACF and PACF graphs for the residuals of the ARIMA model (0,1,1)(2,0,0).

Source: author’s own work.

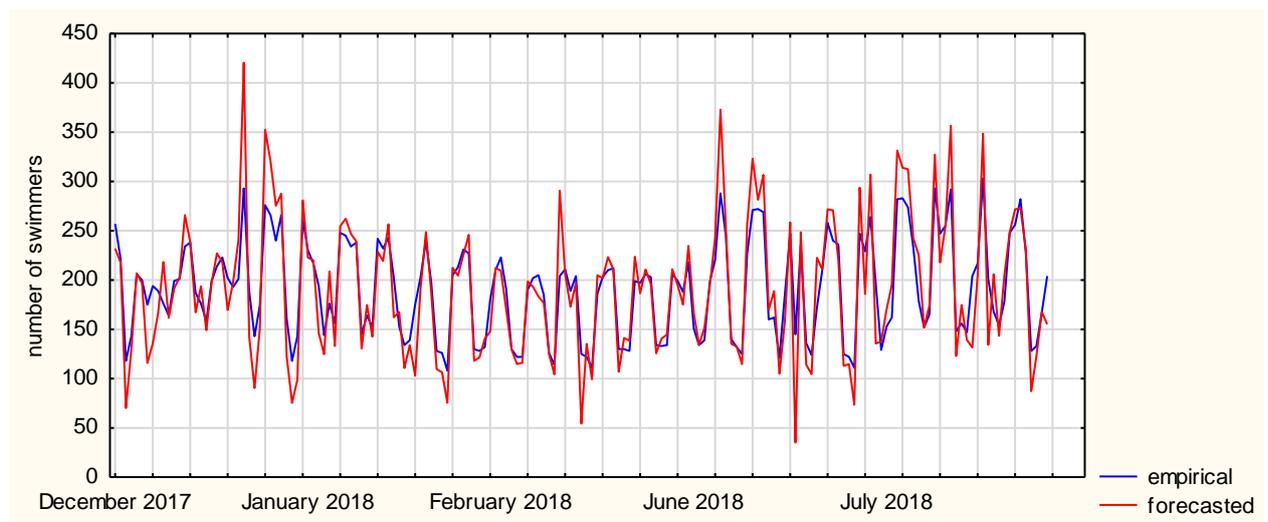
The distribution of residuals is closer to normal distribution ( $p$ -value=0.00029 in the Shapiro-Wilk test), however, it still did not allow for the assumption of the  $H_0$  null hypothesis of normality of distribution at the level of significance  $\alpha=0.05$  (Fig. 10).



**Figure 11:** Normality test of residuals of the ARIMA model (0,1,1)(2,0,0).

Source: author's own work.

The non-normality of distribution was again influenced by high variability of empirical data, resulting in overestimation of the forecast, as shown in Fig. 12. For the high season, the average forecast error was 22 people, which is much higher and indicates a worse fit compared to the previous model for the low season.



**Figure 12:** Graph of empirical and projected values for the ARIMA model (0,1,1)(2,0,0).

Source: author's own work.

However, it is best to assess the proposed models in real life conditions, verifying how they managed to determine the potential number of swimming pool customers. For this purpose, the recorded test observations from the first days of December, which were not included in the construction of models, were used. The Table 5 presents the actual observation value and the forecast together with the relative error of the forecast.

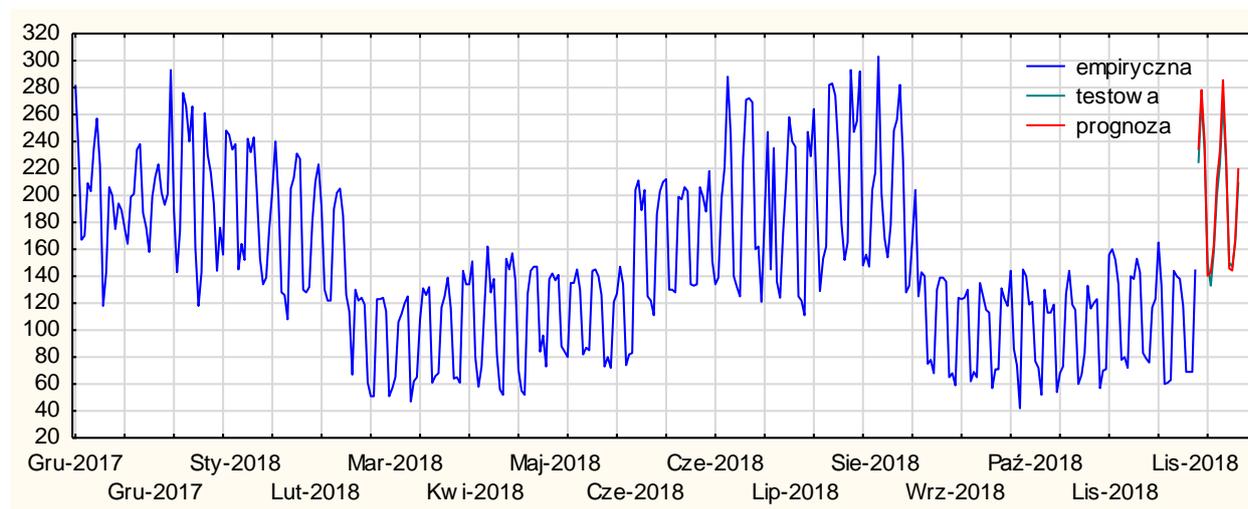
**Table 6:** Comparison of forecast and test observations for December 2018

date	observation no.	forecast	empirical data	relative error of the forecast $\psi$ [%]
------	-----------------	----------	----------------	---

date	observation no.	forecast	empirical data	relative error of the forecast $\psi$ [%]
2018-12-01	366	233.7795	224	4.37
12/2/2018	367	278.1835	271	2.65
12/3/2018	368	240.3599	233	3.16
12/4/2018	369	140.2660	150	6.49
12/5/2018	370	142.8584	133	7.41
12/6/2018	371	163.3989	158	3.42
12/7/2018	372	211.4756	200	5.74
12/8/2018	373	232.9656	223	4.47
12/9/2018	374	285.6192	270	5.78
12/10/2018	375	227.3252	222	2.40
12/11/2018	376	145.7672	150	2.82
12/12/2018	377	144.2306	147	1.88
12/13/2018	378	168.0041	165	1.82
12/14/2018	379	220.1948	210	4.85

Source: author's own work.

The average forecast error is 4%, which is a very good result. However, most forecasts are overestimated, which from the point of view of the swimming pool is a better result than the opposite situation, as it is better to secure the level of safety of users by anticipating their larger number, which will allow a certain reserve of readiness to be maintained. The graph of the analyzed time series including test observations and forecasts is presented in Fig. 11.



**Figure 13:** Graph of the examined series, test observations and forecasts.

Source: author's own work.

The forecasts constructed in such a way allow determining the appropriate number of lifeguards in relation to the users of the swimming pool and not its area. For example, they make it possible to determine the number of permanent employees, which can be supplemented by seasonal employees during periods of increased interest. Moreover, they make it possible to

determine the necessary rescue equipment (lifebuoys, life jackets, medical equipment, medicines and sanitary articles) not only in accordance with the law (Żurek, Ziółkowski, Borucka, 2017), but in relation to the actual number of customers of the swimming pool.

## Conclusions.

The article presents an example of a practical use of the ARIMA model for forecasting the number of swimming pool users. Using the analysis of time series based on the dependence of the examined feature (variable) on time, conclusions were formulated concerning the dynamics of the studied phenomenon in the nearest future. Such forecasts may supplement the existing legislation, which defines only minimum requirements for the safety of swimmers with regard to the area of the pool and not the degree to which they are populated by swimmers. The obtained forecast errors result in it being only able to play an advisory role, but the task of each forecast is only to support management processes, and not to provide ready-made answers. In addition, each of them requires constant monitoring and adaptation to changing circumstances.

## References

- Bielińska E. (2007). *Prognozowanie ciągów czasowych*, Gliwice: Wydawnictwo Politechniki Śląskiej.
- Borucka, A. (2013). Analiza polskiego transportu samochodowego, *Systemy Logistyczne Wojsk*, 39, 13-23.
- Borucka, A. (2018). Analysis of the effectiveness of selected demand forecasting models, *Współczesna Gospodarka*, 2(9), 41-55.
- Borucka, A. (2018). Application of ARIMA Models for the Analysis of Utilization Process of Military Technical Objects, *Logistyka i Transport*, 1(37), 13-22
- Borucka, A. (2018). *Forecasting of fire risk with regard to readiness of rescue and fire-fighting vehicles*, Interdisciplinary Management Research XIV, Croatia, 397-395.
- Borucka, A. (2013), Funkcjonowanie wojskowych oddziałów gospodarczych w nowym systemie logistycznym sił zbrojnych, *Logistyka*, 6, 39 - 48.
- Borucka, A. (2018). Markov models in the analysis of the operation process of transport means, Proceedings of the ICTTE *International Journal For Traffic And Transport Engineering Conference*, Belgrad, 1073-1082.
- Borucka A. (2014). Przewóz osób w Polsce w świetle zmian legislacyjnych, *Logistyka*, 6, 468-475.
- Borucka, A. (2018). Risk Analysis of Accidents in Poland Based on ARIMA Model, *Transport Means*. Proceedings of the 22nd International Scientific Conference part I, Lithuania, 162-166.
- Borucka, A. (2018). *Three-state Markov model of using transport means*, Business Logistics In Modern Management, Proceedings of the 18th International Scientific Conference, Croatia, 3-19.
- Borucka, A., (2018). Model of the operation process of aircraft in the transport system, Proceedings of the ICTTE *International Journal For Traffic And Transport Engineering Conference*, Serbia, 22-30.
- Cieślak M. (2001), *Prognozowanie gospodarcze*, Warszawa: Wydawnictwo PWN.
- Czyżycki R., Hundert M., Klóska R. (2007), *Wybrane zagadnienia z prognozowania*, Szczecin: Economicus.

- Czyżycki R., Klósk R. (2011), *Ekonometria i prognozowanie zjawisk ekonomicznych w przykładach i zadaniach*, Szczecin: Economicus.
- Dittmann P., Szabela-Pasierbińska E., Dittmann I. Szpulak A. (2011), *Prognozowanie w zarządzaniu sprzedażą i finansami przedsiębiorstwa*, Warszawa: Wolters Kluwer Polska.
- Mikosz, B., Borucka, A. (2008), *Organizacja gospodarki odpadami w siłach zbrojnych na tle zmian militarnych i nowych wyzwań stawianych polskiej armii*, Gliwice: Archiwum Gospodarki Odpadami i Ochrony Środowiska, 8, 1-12.
- Mitkow, Sz., Borucka, A. (2018), Mathematical model of travel times related to a transport congestion: an example of the capital city of Poland – Warsaw, *Business Logistics In Modern Management*, Proceedings of the 18th International Scientific Conference, Croatia, 501-526.
- Rozporządzenie Ministra Spraw Wewnętrznych z dnia 23 stycznia 2012 r. w sprawie minimalnych wymagań dotyczących liczby ratowników wodnych zapewniających stałą kontrolę wyznaczonego obszaru wodnego. Dz.U. 2012 poz. 108.
- Rozporządzenie Ministra Spraw Wewnętrznych z dnia 27 lutego 2012 roku – w sprawie wymagań dotyczących wyposażenia wyznaczonych obszarów wodnych w sprzęt ratunkowy i pomocniczy, urządzenia sygnalizacyjne i ostrzegawcze oraz sprzęt medyczny, leki i artykuły sanitarne Dz.U. 2012 poz. 261
- Sokołowski A. (2016), *Prognozowanie i analiza szeregów czasowych*, Materiały szkoleniowe, Kraków: StatSoft Polska.
- StatSoft (2006), *Elektroniczny Podręcznik Statystyki PL*, Krakow, WEB: <http://www.statsoft.pl/textbook/stathome.html>.
- Skoczyński, P., Świdorski, A., Borucka, A. (2018), Characteristics and Assessment of the Road Safety Level in Poland with Multiple Regression Model, *Transport Means*, Proceedings of the 22nd International Scientific Conference part I, Lithuania, 92 – 97.
- Szczepański, E., Jacyna-Gołda, I., Świdorski, A., Borucka, A. (2019) *Wear of brake system components in various operating conditions of vehicle in the transport company*, *Eksploatacja i Niezawodność – Maintenance and Reliability*, 1(21), 1-9. [dx.doi.org/10.17531/ein.2019.1.1](https://doi.org/10.17531/ein.2019.1.1)
- Świdorski, A., Borucka, A. (2018), Mathematical Analysis of Factors Affecting the Road Safety in Selected Polish Region, *Transport Means*, Proceedings of the 22nd International Scientific Conference part II, Lithuania, 651- 654.
- Ustawa z dnia 18 sierpnia 2011 r. o bezpieczeństwie osób przebywających na obszarach wodnych. Dz. U. z 2016 r., poz. 656 z późn. zm.
- Ustawa z dnia 20 lipca 2017 r. - Prawo wodne Dz.U. 2017 poz. 1566.
- Waśniewski, T., Borucka, A. (2011), Sieciowe rozwiązania w łańcuchu dostaw w oparciu o technologię radiowej identyfikacji towarów, *Systemy Logistyczne Wojsk*, 37, 223 – 233.
- Wielgosik, M., Borucka, A. (2016), Istota i znaczenie służby przygotowawczej i szkolenia rezerw, *Systemy Logistyczne Wojsk*, 45, 51-66.
- Żurek, J., Ziółkowski, J., Borucka, A. (2017), *A method for determination of combat vehicles availability by means of statistic and econometric analysis*, *Safety and Reliability. Theory and Applications*, ESREL, 2925-2934.
- Żurek, J., Ziółkowski, J., Borucka, A. (2017), *Application of Markov processes to the method for analysis of combat vehicle operation in the aspect of their availability and readiness*, *Safety and Reliability. Theory and Applications*, ESREL, 2343-2352.
- Żurek, J., Ziółkowski, J., Borucka, A. (2017), *Research of automotive vehicles operation process using the Markov model*, *Safety and Reliability. Theory and Applications*, ESREL, 2353-2362.

## MODEL ARIMA W ZASTOSOWANIU DO ANALIZY POPYTU NA USŁUGI PŁYWALNI

### Streszczenie

Na preferencje klientów wybierających aktywną formę spędzania wolnego czasu wpływa wiele czynników. Bogata oferta rynkowa daje szerokie możliwości, a od obiektów sportowych wymaga się, aby pozostawały w pełnej gotowości do świadczenia usług. Pomocne w takim przygotowaniu jest konstruowanie prognoz umożliwiających określenie przewidywalnej liczby klientów. Przykład takiej predykcji, sporządzonej na podstawie pływalni, przedstawiono w niniejszym artykule. W tym celu wykorzystano model ARIMA bazujący na założeniu, że na wartość zmiennej objaśnianej wpływa wartość tej zmiennej opóźniona w czasie.

**Słowa kluczowe:** model ARIMA, ratownictwo wodne, gotowość, pływalnia, bezpieczeństwo

**Klasyfikacja JEL:** C2, C22.

Mieczysław Pawlisiak  
Military University of Technology  
Ul. Urbanowicza 2, 00-908 Warszawa  
mieczyslaw.pawlisiak@wat.edu.pl