Application of ARIMA Models for the Analysis of Utilization Process of Military Technical Objects

Anna Borucka
Military University of Technology, Poland

The newest solutions in Polish Armed Forces are implemented gradually and focus mainly on soldiers' combat readiness. Many concurrent processes occur, for which proper analysis and interpretation could constitute command process and task realization support; however poor and standing (paper) record seems to be an obstacle in their modelling. Therefore the author of the article tried to depict the process of military technical objects exploitation based on archived data according to present methods of documents preparation, circuit and record, applicable in Polish Armed Forces. Based on that, the method of research the readiness of aircraft ships from military air base, powered by ARIMA model, was proposed. Using empirical data of two years of exploitation, the identification of researched time series, and then a few models estimation was made. Finally, the best model was chosen and verified.

Keywords: aircraft, forecasting, ARIMA models, military data analysis, data quality issue.

1. INTRODUCTION

Nowadays the development of enterprises is extremely dynamic because of the strong market expansion, large competition and pressure on time and prices. This development is supported by modern technologies allowing receiving, processing and gathering huge amount of information which is recorded in real time (online).The analysis and conclusions based on such database require appropriate knowledge and creativity, but also advanced algorithms and equipment of high computational power. Obtained results may significantly adjust the enterprises functionality, however the selection process itself is a huge challenge, and the problem so called "big data" is being discussed both by scientific and business communities. In this context Polish Armed Forces seem to function in old-fashion way. The goals and tasks of both military and civil logistics are generally similar and relate to the processes of satisfying the demand and organization functioning preservation by providing the appropriate supply and maintaining the stock level with its adequate disposal. However, the main goal of logistics in companies is to make the financial profit, and all of the actions (including logistics) aim to optimize the outlay and achieved effects. The above mentioned implementation of the newest technological and technical achievements – especially in the IT sector – is an efficient tool to acquire a client. The most important factor deciding about the market dominance is the information, so it must be precise, accessible, high quality and should not generate high costs. Therefore, the systems that collect, process and present data develop at rapid pace, integrating not only the various spheres of business functioning, but also the links throughout the supply chain. Transport is also an area of high exploitation. The tools that use IT, telecommunication and automatic control technologies, adapted and integrated with the attended physical systems, support the flow of goods proper management, but also ensure full identification, location and monitoring of means of transport. The vehicles transceivers allow to gather the information about current location, driving direction with its parameters (time, velocity, fuel combustion), performed activities status (work, service, rest), but also information about vehicle exploitation, determining the parameters of engine operation, fuel consumption, tires etc. The possibility to acquire such multivariable and precise data ensures efficient and effective fleet management. It allows increasing the transport safety, preventing drivers misuse and their discipline improvement; it also gives a chance for a quick reaction in case of irregularities, but primarily simplifies scheduling, analysis and proper decisions and diagnosis making based on forecast of high probability, which causes the increase of savings and consequently the company profit.

The implementation of modern technologies in army is not that simple, mainly because of the vehicles condition and age. The newest technologies are gradually implemented, and only in the new units. The implementation delay of some useful solutions results from many factors, however the differences in the objectives of functioning between the Polish Armed Forces and civilian enterprises should be emphasized. Their task is to provide widely understood safety and peace, which have to be achieved in the training and preparation processes for conducting combat operations. The costs saving is not army's priority, in fact its preparation must require adequate input. Admittedly, past years show that costs optimization is possible and advisable, which can be successively accomplished with tools and methods applied in civilian enterprises logistics, but at the moment, while the civilian market is struggling with the excess of digitized resources, in the army a paper documents circuit occurs. Existing print designs and the method of recording, unaltered from many years, contain only key elements for realized task, lacking of detailed information which may be useful for analysis. The significant disadvantage records gathered this way is their high exposure to errors resulting from so-called "human factor", which undermines credibility to a considerable extent. The implementation of modern technologies in army is slow and gradual process, not only because of its specification, but also because of its size. Nonetheless, the information possible to acquire presently, although lacking in precision and the volume of mentioned "big data", may be the basis of the study, also in the technical objects exploitation area. Modelling of objects realized processes, predicting potential trends and changes allows to control the willingness not only of technical objects, but the whole military unit. Therefore, the author attempts to model the process of military units exploitation based on the archived data according with current methods of preparing, circulation and documents record, compulsory in Polish Armed Forces.

The intention was to check if they are sufficient, predicting based on them achievable, and the results obtained are reliable and possible to implement in command process, because the commanders decisions should be made taking into account proper mathematical methods and models. In a situation where analyzed process is random, the theory of stochastic processes application is possible, where the object in successive moments can be in different exploitation states. In each state it is possible to mastermind the analyzed process, which leads to the technical objects and the system as a whole readiness. The model described in the article allows to military vehicles readiness evaluation, and ultimately its increase by proper areas analysis. The desired and significantly important issue is valid system construction, whose task is the efficient and on-time action ensuring the safety of human health and life, especially in units that realize precise schedules and/or sudden, requiring immediate intervention commands.

Detailed knowledge about processes realized by technical objects and possibility of their reliable forecasting is very important for the commanders, but also for logistic component, responsible for supporting and securing of realized combat, training and economic operations. Thanks to that the constant and systematic supply of high quality stock is ensured.

2. RESEARCH OBJECT CHARACTERISTICS

The fundamental problem of technical objects research in the army is a way of information gathering and storing. The above mentioned lack of implementation of IT solutions and automation of these processes results in poor and difficult source material to analyze. None the less it should be mentioned that the IT systems, although they have a lot of advantages, there exist a risk that enemy can capture some information through them. So the implementation of IT systems in military requires involvement of cryptology and cyber security specialists, which results with other difficulties and extra time consumption.

However, before implementation of such systems, the analysis of available databases and possibility of effective forecasting based on them were discussed in this article.

The research was carried out based on the measurements from 2 consecutive years (2014 and

2015). The measurements concern activities as part of the military technical objects (aircraft) exploitation process in the selected military unit. Collected evidence, in great majority in a form of paper documents, required electronic data base creation. It was a long process, with a high risk of making mistakes, which caused the necessity of numerous verifications and corrections. The quickness of making decisions in command process is a very important factor, so such long-term process of information preparation is unfavourable.

The first analysis stage consists of objects tasks specification. For this purpose five activities were specified, obtaining disconnected exploitation states of them, shown in the Table 1.

Table 1. Exploitation states, differentiated for technical objects under research.

Exploitation state	Function realized
State 1 – task realization	State in which the object realizes its destined task. For analyzed aircraft vehicles this state is the flight. Verification and control flights were also taken into account.
Stage 2 – control and repair processes realization	State in which control and repair processes on the aircraft are realized. They are executed on a technically efficient object. They check and prepare the aircraft for flight. Among others these actions include: preparing for action by technician, and trials before flight; on ground trial, overhang trial and refuelling.
Stage 3 – readiness to perform the task	Stage in which the aircraft is checked and ready in technical point of view, with a pilot inside waiting for the first or next task realization.
Stage 4 – readiness state without pilot	State in which aircraft is checked and ready for flight, however without possibility of task realization because of lack of a pilot.
Stage 5 – garaging	The garaging process is performed. To realize tasks again the control actions performing is required.

This article is focused only at the Stage 1, because the most important factor for the unit base exploitation continuity is the task realization. The forecasting of Stage 1 will allow obtaining information about intensity of flights, and thanks to that proper aircraft and staff arrangement. The research itself also had to show that improving the exploitation of military systems is possible and reasonable by using time series methodology.

3. INTRODUCTION TO ARIMA MODEL

We understand the term 'time series' as a realization of stochastic process, in which the field of interest is time [1; 5; 6]. Therefore it is an information series in chronological order. The time series analysis is done primarily to establish the nature of the researched phenomenon, expressed by the set of collected observation, and to make prognosis (anticipation of the next ones) of the incoming time series variables [1; 2; 3; 4; 5].

Models based on the time series analysis assume existing and constant in time predictor connections (otherwise explanatory, egzogenic or external variable) with explanatory variables (endogenic, forecasted or internal variable) all over the set of possible variables.

This property enables to describe existing dependence with one analytical formula. It should be mentioned here that properly prepared prognosis is not the final decision, or an unequivocal answer on how to form the future actions. Their role has to be only advisory, and can only be a hint referring to making decisions, defining strategies, goals and tasks both for military units commanders and civilian enterprises managers.

The easiest way to classify prognosis is by the time horizon point of view, dividing them into short-term, mid-term and long-term (prospective). Short-term prognosis are made for the maximum one year period, mid-term describe the interval from 2 to 5 years, and long-term prognosis – over 5 years. This division is however only conventional, and in the literature there exist also other ones, referred to e.g. characteristics of the forecasted phenomenon change. According to this criteria for the short-term prognosis, only the quantitative changes is taken into account, so only the change of the forecasted variable conditioned by current regularity (e.g. trend) is observed. In mid-term prognosis not only the quantitative changes exist, but also minor quality changes (at least minimal extrapolation is observed), while in long-term prognosis both quantitative and quality changes are important, which makes them crucial factors of researched phenomena [2;5].

Auto-regression and moving average models, which constitute the components of ARIMA model (Autoregressive Integrated Moving Average Model), belong to short-term prognosis. They find their application in modelling of stationary time series (that means for those whom oscillations around the mean value is only random), or non-stationary, brought to stationary.

The basis of creation of such models is an autocorrelation phenomena, which is the correlation of forecasted variables with the values of the same variables delayed in time. There are three basic types [5; 7]:

- auto-regressive models (AR)
- moving average models (MA)
- combined auto-regression and moving average models (ARMA).

In autoregressive model (AR) the values of the variables are explained by their delayed values. The differential level is indentified by p parameter of this process that is auto-regression order.

AR model form is the following (1):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t \tag{1}$$

where:
$y_t, y_{t-1}, y_{t-2}, y_{t-p}$-values of forecasted variable at the moment or period $t, t-1, t-2, \ldots, t-p$;
$\alpha_0, \alpha_1, \alpha_2, \alpha_p$- model parameters;
ε_t - error (rest) of the model at the moment or period t;
p - delay value.

In the moving average model (MA) the values of the explained variable are expressed as a function of delayed values of stationary, random component. Parameter q of this process, that is MA process order, shows the level of assumed delays of the model.

MA model form is the following (2):

$$y_t = \beta_0 + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 y_{t-2} - \ldots - \beta_q \varepsilon_{t-q} \tag{2}$$

where:
$y_t, y_{t-1}, y_{t-2}, y_{t-q}$- values of forecasted variable at the moment or period $t, t-1, t-q;$
$\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-q}$- error (rest) of the model at the moment or period t$t, t-1, t-2, \ldots, t-q$;
$\beta_0, \beta_1, \beta_2, \beta_q$- model parameters;
q- delay value.

By combining the mentioned above models we obtain the ARMA model. It enables to achieve higher efficiency and elasticity in model adjustment to the time series. ARMA model assumes, that the value of the forecasted variable in the time t depends on its past values and difference between the past real values of forecasted variable, and its values obtained by the model – the errors of the prognosis. The ARMA model form is the following (3):

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \\ &+ \varepsilon_t \beta_0 + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 y_{t-2} - \ldots - \beta_q \varepsilon_{t-q} \end{aligned} \tag{3}$$

Classical ARMA models find their application only for the stationary series. If the series is non-stationary, but brought to the stationary form, the ARIMA model may be implemented [2; 5; 7]. The '*I*' letter in the name means that the researched time series is being differentiated, which enables to obtain the stationary series. The *d* parameter informs about the level of integration of series (the number of differentiation operations that have been done).

In the 70s of the last century Box and Jenkins developed a selection of appropriate model procedure of three main stages. These are: the identification, estimation and tests, whose positive results allow to final model implementation.

4. RESEARCH METHOD

According to the above mentioned methodology, the first stage consists of the analysis and estimation of researched time series. Because of discontinuity of gathered daily data, it was decided to apply the weekly data set, and data of such arrangement was further verified. Researched time series constituted 98 observations, gathered in the years 2014 and 2015. For research purposes 92 of them were used; the 6 remaining were left to forecast verification. It is a sufficient number to make use of the proposed econometric model.

The ARIMA models – as mentioned – find their application in a case of stationary series or non-stationary, brought to stationary series, so at the beginning it is essential to examine this feature [7]. According to the established procedure the first step is to analyze the chart of the researched time series (Fig.1), and then the charts of the ACF function correlation and PACF partial autocorrelation (Fig.2). They indicate the stationary character of the process, which is also confirmed by the additionally performed Dickey-Fuller (Gretl) test, obtaining a negligible value of p=1.26E-8, for a 0th hypothesis assuming the occurrence of a unitary element.

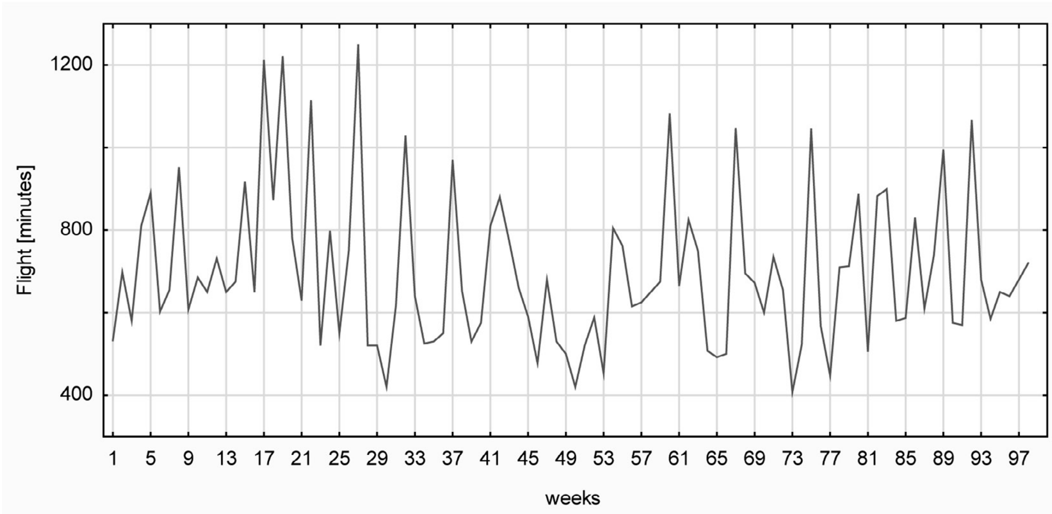

Fig. 1. Chart of researched time series – aircraft flight in minutes.

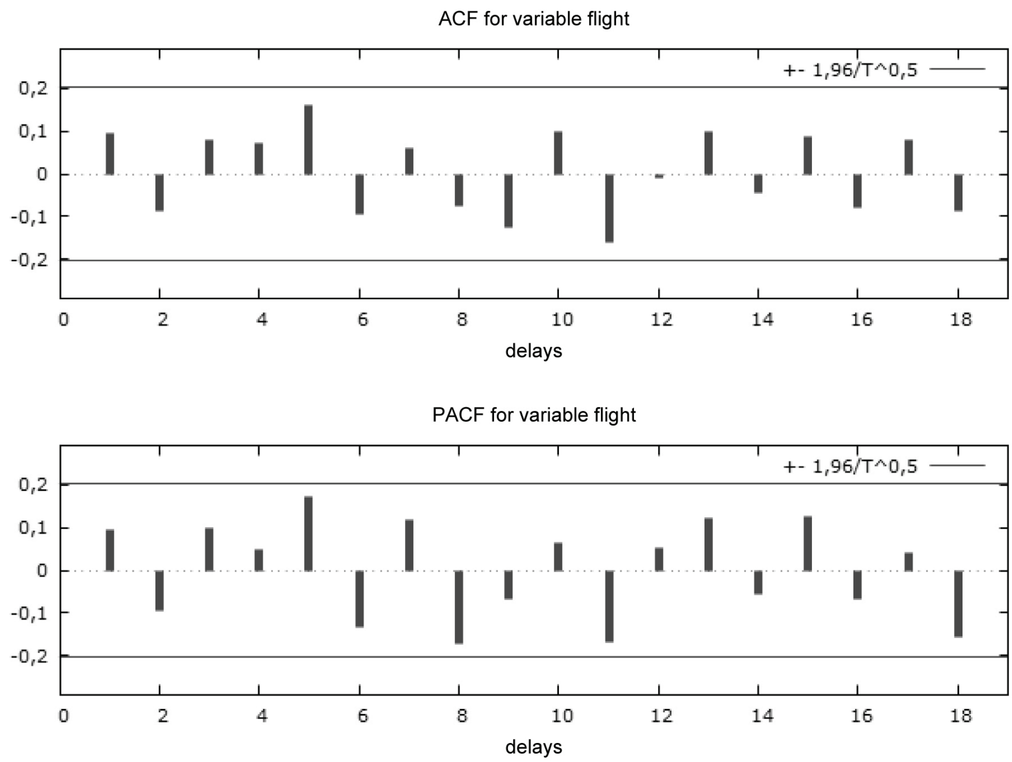

Fig. 2. Charts of ACF and PACF autocorrelation of researched series.

The stationary character of process allowed to move to the second stage, which consists of model parameters establishment. Three parameters are distinguished: autoregressive parameter (*p*), integration order (*d*), moving average parameter (*q*), which is registered in form: ARIMA (*p, q, d*) or with numbers (e.g. 0, 1, 1).

Several different variants are commonly proposed, and then, based on the selected criteria analysis, which may be the significance of model parameters, forecast errors, information criteria (e.g. AIC - Akaike's Information Criterion, BIC - Bayesian Information Criterion or FPE - Final Prediction Error), the best variable is chosen.

Five different variants were estimated. Despite the times series stationary confirmation, other models with series integration, with delay equal 1, were proposed. Obtained models and selected criteria of choice, that is: model parameters significance (statistically significant parameters were underlined), mean squared error (MSE) and information criteria (AIC), are shown in the Table 2.

Table 2. Estimation of selected models results (underlined values are statistically significant).

	Model (1,0,1)	Model (0,0,1)	Model (1,0,0)	Model (0,1,1)	Model (1,1,0)
Constant	640,74	639,12	638,92	0,91336	3,5816
p	-0,7479	x	0,09689	x	-0,4047
q	-0,9603	-0,1252	x	0,93959	x
d	x	x	x	1	1
AIC	1291,205	1299,788	1297,011		1322,379
MSE	66876,63636	72567,435	72748,541	77760,957	112058,75

Analyzing the above mentioned models, firstly the statistical significance of estimated parameters should be checked, because the statistically insignificant parameter informs that variable does not affect on model adjustment and may be omitted in it.

Statistical significance of all model parameters (1,0,1), and the lowest value of MSE and AIC decided of its choice, so in the next stage the diagnosis of this model was performed. It consisted of the analysis of the rests, which in the properly constructed model should be random and symmetrical. Moreover, the stationary character and normality of their decomposition should be examined.

The lack of statistical significant autocorrelations and partial autocorrelations is shown on the chart of ACF and PACF functions (Fig. 3), while normality chart (Fig.4) does not diverge from normal decomposition and allows to accept the rest of the model as white noise process.

Randomness of the decomposition was proved with Wald-Wolfowitz series test, obtaining assumed random value of p=0.85 for a 0th hypothesis, confirming its randomness.

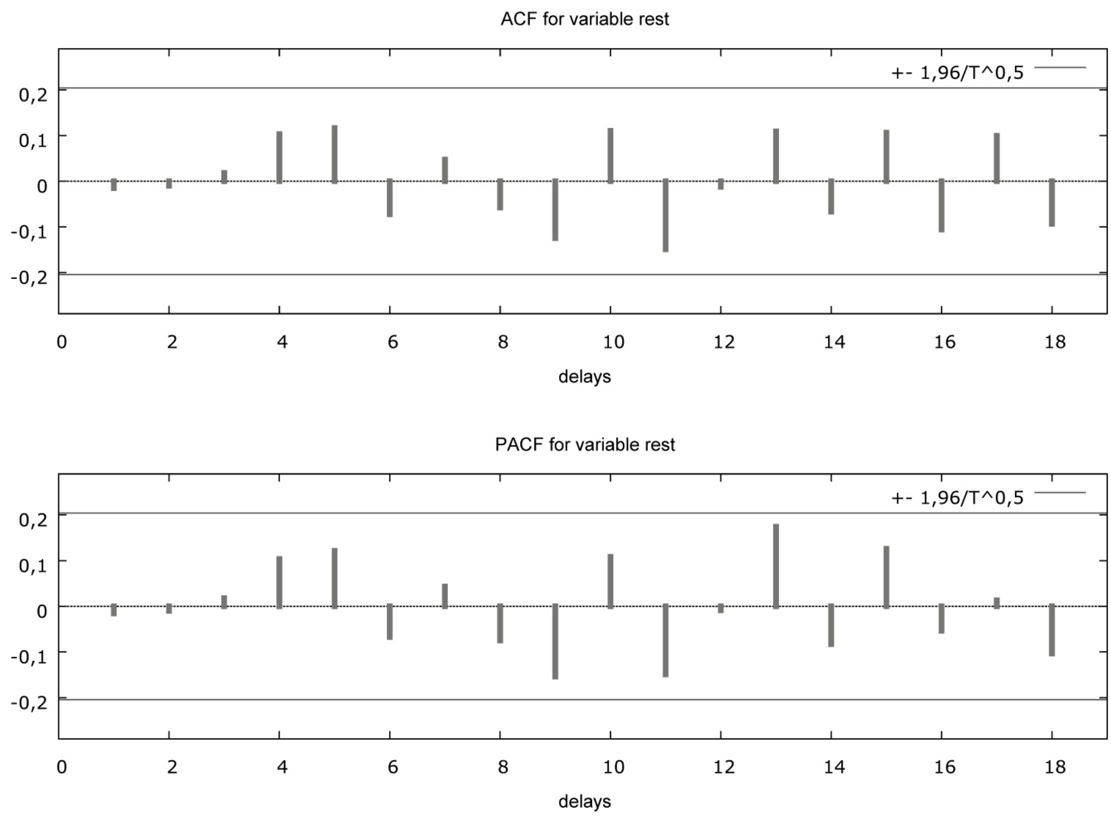

Fig. 3. Charts of ACF autocorrelation and PACF partial autocorrelation functions.

A positive verification allowed to designate the forecast with its value based on known observations, which were stored as tested, shown with the forecast relative error in Table 3.

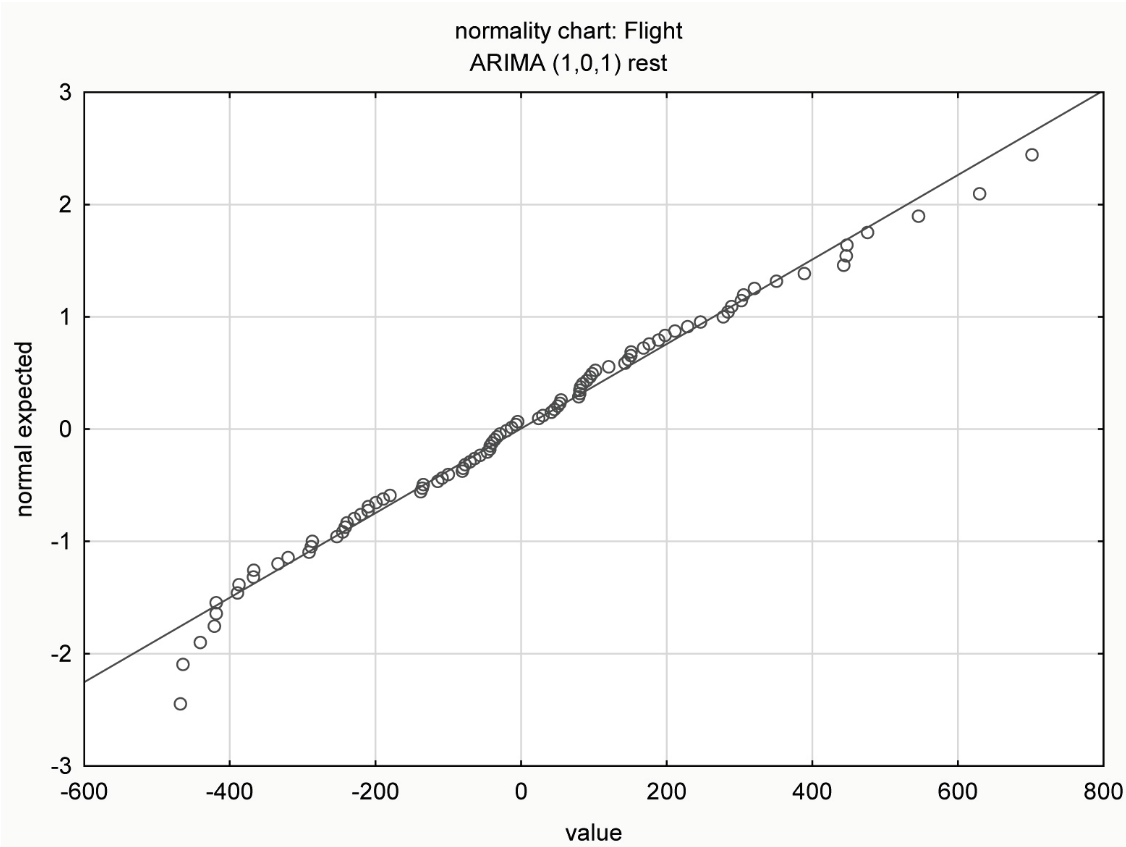

Fig. 4. Normality chart of the rests of estimated model.

Table 3. Obtained forecast verification.

Observation number	Forecast	Empirical data	Forecast relative error Ψ [%]
93	735,2048	712	-3,259101
94	570,081	585	2,550256
95	693,585	650	-6,70538
96	601,2104	640	6,060875
97	670,3018	680	1,426206
98	618,625	720	14,07986
99	657,2766		
100	628,3672		
101	649,9899		
102	633,8173		

Obtained results of errors are satisfactorily low. The chart of researched time series and forecast is shown on Fig. 5.

According to the Fig. 5, the worst forecast is in the case of high intensity of flights, when the mean values reach over 1000 minutes. However for most observations, the forecast relative errors reach the level below 20%, which is a satisfactory and admissible value in logistics. The chart of forecast relative errors is shown on Fig. 6.

The forecast made in this way should be of course monitored and verified. The research shows that the present methods of data obtaining, processing and archiving – used in the army – allow the creation of reliable models, which may find their application in decision-making processes for military operating systems. A combination of commanders' leadership experience with mathematical methods in the process may turn out very effective.

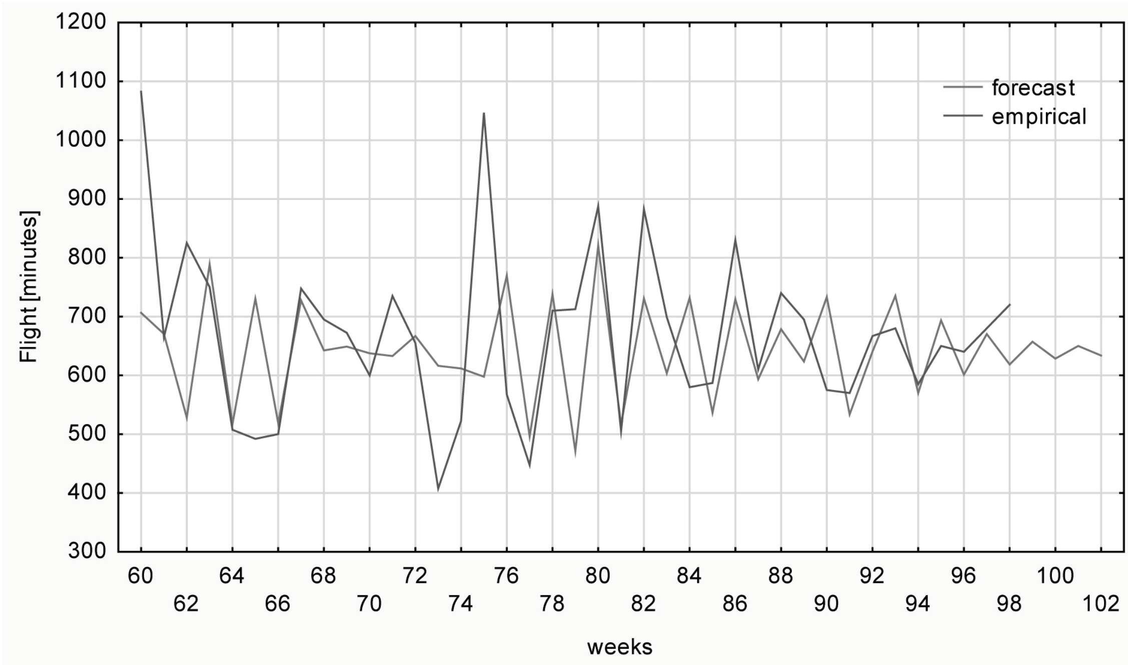

Fig. 5. Chart of researched time series with its forecast.

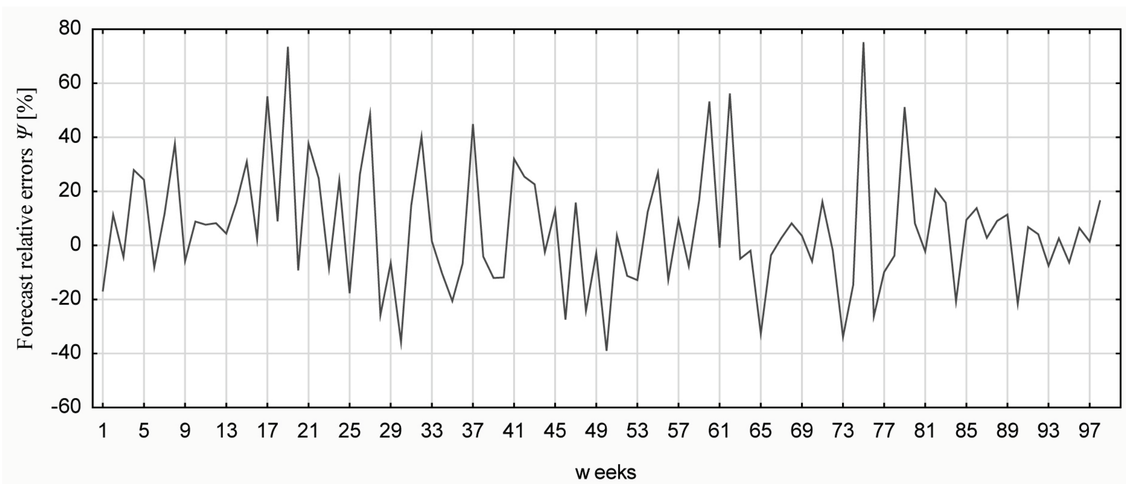

Fig. 6. Forecast relative errors Ψ [%].

5. CONCLUSIONS

The article presents the example of practical usage of ARIMA model in military systems exploitation forecasting. Using the time series analysis, whose basis is the dependence of researched feature (variable) in respect to time, conclusions about the dynamics of the phenomenon in the near future were formulated. Short-term prognoses are significant not only in the enterprise management process, but also in military commanding. Basically, they enable to delegate tasks, ensure necessary forces and resources to realize them, and more accurate decisions making related to preparation of potential orders executing.

This research allowed to predict the flight realization request with high probability. Thanks to that, the commander can plan the tasks for the staff more precisely, also to control the aircrafts and ultimately probe the system readiness with high reliability.

REFERENCES

[1] Cieślak M., 2001. *Prognozowanie gospodarcze*. Warszawa PWN.

[2] Czyżycki R., Hundert M., Klóska R., 2007. *Wybrane zagadnienia z prognozowania.* Economicus. Szczecin.

[3] Czyżycki R., Klósk R., 2011. *Ekonometria i prognozowanie zjawisk ekonomicznych w przykładach i zadaniach*. Economicus. Szczecin.

[4] Dittmann P., 2000. *Metody prognozowania sprzedaży w przedsiębiorstwie*. Wydawnictwo Akademii Ekonomicznej. Wrocław.

[5] Dittmann P., Szabela-Pasierbińska E., Dittmann I. Szpulak A., 2011. *Prognozowanie w zarządzaniu sprzedażą i finansami przedsiębiorstwa*. Wolters Kluwer Polska Sp. Z o.o. Warszawa.

[6] Pawełek B., Wanat S., Zeliaś A., 2013. *Prognozowanie ekonomiczne Teoria przykłady zadania*. PWN Warszawa.

[7] Sokołowski A., 2016. *Prognozowanie i analiza szeregów czasowych. Materiały szkoleniowe*. StatSoft Polska. Kraków.

Date submitted: 2017-11-22
Date accepted for publishing: 2018-02-12

Anna Borucka
Military University of Technology, Poland
anna.borucka@wat.edu.pl