Original Article

THE FITTING OF SARIMA MODEL ON PEADS PATIENTS COMING AT OUTPATIENTS MEDICAL LABORATORY (OPML), MAYO HOSPITAL, LAHORE

Muhammad Ibrahim[*], Muhammad Jamil[**], Safoora Akbar[***], Abdul Majeed Akhtar[****], Zaheer-ul-Hassan Mir[*****], [5] Abu Bakar Imran[******], Azka Jamil Akhtar[*******]

[*] Department of Statistics, Govt. MAO College, Lahore;

[**]OPML, OPD, Mayo Hospital, Lahore;

[***]Safoora Akbar;

[****]Program Manager Provincial TB Program, Punjab;

[*****]Additional Medical Superintendent (OPD), Mayo Hospital Lahore;

[******]FMH College of Medicine and Dentistry Lahore. Lahore;

[*******]Bank Alfalah Ltd. Islamic Banking

ABSTRACT

BACKGROUND & OBJECTIVE: Forecasting medical phenomena like the patients volume, emergency overcrowding, stay length in hospital on surgical procedure and bed occupancy in hospital wards, patients' influx or patient arrival drawn considerable attention since last two decades. This study was carried out to fit a best uni-variate ARIMA model (Box-Jenkin Methodology) to forecast the Pediatrics patient's incoming at Outpatients Medical Laboratory (OPML), Outpatients Department (OPD), Mayo Hospital, Lahore to determine any seasonal impact on patient's incoming.

METHOD: Time series data of Peads patients coming/reporting in OPML, OPD Mayo Hospital Lahore, from September 2007 to December 2013 were used for fitting the best model.

RESULT & CONCLUSION: The appropriate model for Peads data as ARIMA (1, 1, 1) and SARIMA (1, 0, 1) at seasonal points=4 after residuals diagnostic checks. The estimated number of Peads patients in the month of January 2014 was 119 using ARIMA model and 92 using SARIMA model as compared to actual which is 102 patients. SARIMA model has ability to forecast the number of incoming patients accurately. It is now concluded that the fitted SARIMA model showed a seasonal impact on incoming patients and can be used to forecast the patients' incoming to OPML, OPD for future planning and management to ensure quality diagnostic healthcare service to the patients/clinician.

KEYWORDS: Forecast, Medical Time Series Data, SARIMA Model, Box-Jenkin Methodology

INTRODUCTION:

Patients incoming with increasing trends in Outpatients Department (OPD) and Outpatients Medical Laboratory (OPML) has become serious problem throughout Pakistan as case around the world, which consequents in increasing care costs, causing stress in clinician/surgeons, Medical Laboratory Professionals (MLP's), patients and their attendants and also, affecting adversely outcomes of medical & clinical systems. One side of problem is the difficulty of anticipating the timing and magnitude of overcrowded condition. Modeling and forecasting the patients volume using some most appropriate statistical tools, may provide useful information for hospital administration and OPML management, which may be useful in

Corresponding Author:
Muhammad Ibrahim
Department of Statistics,
Govt. MAO College, Lahore
Email: ibrahim.ap12@gmail.com

planning, expansion, allocating resources and, recruitment and selection of Medical Laboratory Professionals.

Box-Jenkin Methodology[1] is very popular tool since a long time, in developing time series model for forecasting the variables which are related to time. The non-seasonal ARIMA was introduced in 1976. Seasonal ARIMA (known as SARIMA) is used in time series data analysis, when data exhibits seasonal variations or data contains some seasonal behavior.

Various researchers used SARIMA model in various fields when time series contains seasonal component in data. Suhartono[2] used SARIMA model to forecast tourist arrival in Indonesia, Etuk[3] used SARIMA model in forecasting of monthly Naira-Euro exchange rate, Etuk & Igbudu[4] also used same model for Nigerian Naira-British Pound rates, Okereke & Bernard[5] applied SARIMA to forecast GDP in Nigeria, Tankersley *et al*[6] employed this model to forecast ground level fluctuation, Pozza *et al*[7] used SARIMA to analysis of PM 2.5 and PM10-2.5 mass concentration in the city of Sao Carlos Brazil, Durdu [8] used this model for forecasting of boron in Western Turkey, Abraham *et al*[9] applied SARIMA model for short-term forecasting of emergency inpatient flow, Tseng *et al*[10] used this model for forecasting the production value of the mechanical industry in Taiwan, Ediger and Akar[11] applied SARIMA for forecasting production of fossil fuel sources in Turkey and; Scortti *et al*[12] employed this model for forecasting canine rabies in Argentina, Bolivia and Paraguay

Time series data analysis has been observed in medical literature for forecasting the patients' volume, emergency overcrowding, stay length in hospital on surgical procedure and bed occupancy in hospital wards, patients influx or patient arrival, moreover to estimate the cost of hospital stay or any medical or surgical procedure.[13,14,15]

Various researchers used time series data analysis tools like AR (Autoregressive) MA (Moving Average) ARIMA (Autoregressive Integrated Moving Average) Model in forecasting the patient's arrival behavior and trend in hospitals. Various researches are available where the authors used SARIMA (seasonal ARIMA) where seasonal impact on data is observed Time series Models and techniques provide better results than traditional approaches such as descriptive statistics, multiple regression and ANOVA.[16,17,18]

The health/patients related data, which is collected over a sequence of time at regular interval normally shows a non-stationary behavior and clinical researcher are often interested to forecast the future occurrence of such related phenomena. The ARIMA model has the ability to handle the non-stationary time series data to forecast. The objective of this study is to forecast the incoming of patients at OPML, OPD Mayo Hospital Lahore and to determine any seasonal impact on patient's incoming.

1. SARIMA Model

The most popular model for forecasting of uni-variate time series data is ARIMA model, which was introduced by Box-Jenkin in 1976.[1,2] As many researchers used this model for forecasting of different medical phenomena, in paper we has used ARIMA model for forecasting of future incoming of patients at OPML Mayo Hospital Lahore. The ARIMA model is combination of two time series processes Autoregressive (AR) and Moving Average (MA). The ARIMA model is defined as:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots\dots\dots\dots + \Phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots\dots\dots\dots\dots + Z_q \theta_{t-q} \quad \dots\dots\dots\dots\dots 1$$

The ARIMA model is used for integrated stationary process and is best when time series is non-stationary, the non-seasonal ARIMA is classified as ARIMA (p,d,q)[13,16], the process X_t is said to be ARIMA if

$$W_t = \nabla^d X_t = (1-B)^d X_t$$

$$W_t = \alpha_1 W_{t-1} + \alpha_2 W_{t-2} + \dots\dots\dots\dots + \alpha_p W_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots\dots\dots\dots + \theta_q Z_{q-t} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots 2$$

We can say the process is stationary after differencing the process "d" times, normally in practice the "d" is taken one. The generalized form of ARIMA which contained seasonal component is multiplicative SARIMA model (Box 1994, 2) and is written as

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)Z_t \quad ....3$$

Where B is backshift operator, $\phi_p, \Phi_P, \theta_q, \Theta_Q$ are polynomials of order p, P, q, Q and Z_t is pure random process. W_t is shown as below

$$W_t = \nabla^d \nabla_s^D X_t$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - .... - \phi_p B^p$$

$$\Phi_P(B^s) = 1 - \Phi_1 B - \Phi_2 B^{2S} - .... - \Phi_P B^{Ps}$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - .... - \theta_q B^q$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B - \Theta_2 B^{2s} - .... - \Theta_Q B^{Qs}$$

The variable W_t is derived from original variable X_t and not only simple differences but also by seasonal differencing to remove the seasonality. In above model "d" and "D" is non-seasonal and seasonal order of differences. The "s" is value of seasonal period; it may be "12" or "4" as whatever the situation is. The model is expressed as SARIMA (p,d,q) * $(P,D,Q)_s$. When there is no seasonal impact, then SARIMA reduces to normal ARIMA.

The SARIMA model building procedure which is also known as Box-Jenkin method, involves the following four steps:

1. Identification of Model
2. Estimation of Model Parameters
3. Diagnostic Checks (Model Checking)
4. Forecasting

The brief description of each step is as:

1.1 IDENTIFICATION OF MODEL

In this step, the order of respective time series model is determined, the order of differencing "d", the order of AR process which is "p", order of MA process is "q" order of SAR process is "P" and order of SMA is "Q". The model is written as SARIMA (p,d,q)*(P,D,Q). Differencing is the procedure which is used to remove the trend in the time series data, we use differencing to remove the trend in the data. It is critical to difference the time series data, until stationary is getting done. Any seasonal variation can be removed through some appropriate seasonal differencing. The stationarity of time series means that arithmetic mean, variance and autocorrelation remain same over time. It means that if time series data is stationary, then the arithmetic mean of any major subset of data of time series data does not significantly different of any other major subset of the series.[19,20]

PROCEDURE FOR CHECKING THE STATIONARITY

The time plot, correlogram and unit root test are used to test the stationarity of the time series data. The time plot steers us about the trend of data, whereas correlogram guides us regarding the order of SARIMA model and behavior of the series. The unit root test is performed to determine the significance of stationarity in the data. There are several available statistical methods for unit root test of non-seasonal time series data, those tests are Dickey-Fuller (DF), Augmented Dickey-Fuller (ADF, 1979) Zivot (1992), and KwaitKowski (1992). For a given data set, there may be several model with different orders, for example SARIMA (1, 1, 1)*(1, 1, 1), SARIMA (2, 1, 1)*(1, 0, 1) and more, but final model is selected using statistical tools like AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) or Schwarz Criterion. The AIC and BIC are tools to measure the goodness of fit of model. Among number of models, that model is considered best which has lower value of AIC or BIC.[21,27]

1.2 ESTIMATION OF PARAMETERS

Next step after the tentative SARIMA model is the estimation of parameters. Least square or maximum likelihood method can be used to estimate the parameters. The parameters are estimated on the basis of previously available data.

1.3 DIAGNOSTIC CHECK

The tentative suggested SARIMA model is then checked for reasonable fit to the data or we can say that in this stage, we verify the statistical adequacy of the fitted model. Statistically an adequate model must satisfy the four properties:

I. The estimated parameters must be significant.

ii. All Auto Regression (AR) estimated

parameters must be within the "bound of statisonarity". the condition of bound stationarity is as:

SAR order	Condition of bound
SAR(1)	$\lvert\Phi_1\rvert < 1$
SAR(2)	$\lvert\Phi_2\rvert < 1$
	$(\Phi_2 - \Phi_1) < 1$
	$(\Phi_2 + \Phi_1) < 1$

iv. All Moving Average (MA) estimated parameters must lie in the bound of stationarity.

v. The Residuals of model (the unexplained part of data) should be least and R^2 should be maximum, the possible test for minimum residuals, Q-statistic is introduced by Ljung-Box (Ljung and Box 1979, Harry 1993).

1.4 FORECASTING

The eventual target of SARIMA model is to forecast. When the model has undergone all the diagnostic check, the model becomes adequate for forecasting. Various researches showed that SARIMA model performs best in term of forecasting as compared to other models.

2. DATA ACQUISITION AND ANALYSIS STRATEGY

The data of monthly patient's records at OPML, OPD Mayo Hospital Lahore from September 2009 to December 2013 was acquired for modeling. Computer software E-view 5 (Quantitative Micro Software (QMS), USA) and R (Freeware Software) (28) was used to analyze the time series data. The following tools were used for fitting the model.

Step	Statistical Tools	Purpose
Identification	Time Plot Correlogram	To see the trend in data To show the relationship among current and past observations
		To check the stationarity in the data
	Unit Root Test	
Estimation	Maximum Likelihood Estimation (MLE)	To estimate the parameters
Diagnostic Check	R^2 , Goodness of fit AIC, BIC	To find out the most appropriate model
Forecasting and its Accuracy	RMSE(Root Mean Squared Error), MAE(Mean Absolute Error), MAPE(Mean Absolute Percentage Error),	To forecast and check the accuracy in model using forecast error term

3. RESULTS

The original time series data of Peads patients were plotted to observe the behavior or trend in the data, plotting of data is considered a first step to identify the time series model. The line charts of data are shown in Fig 1.

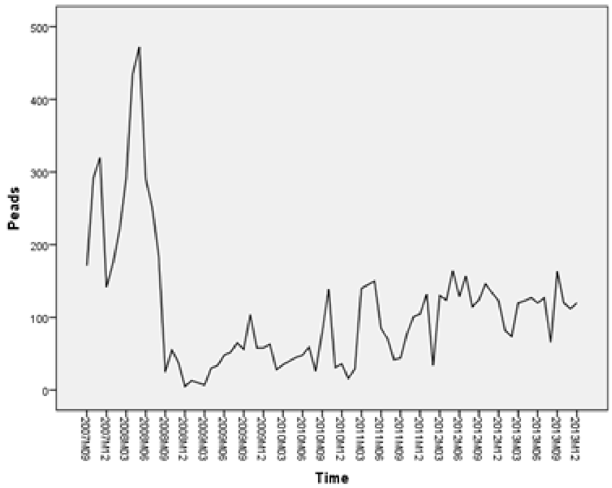
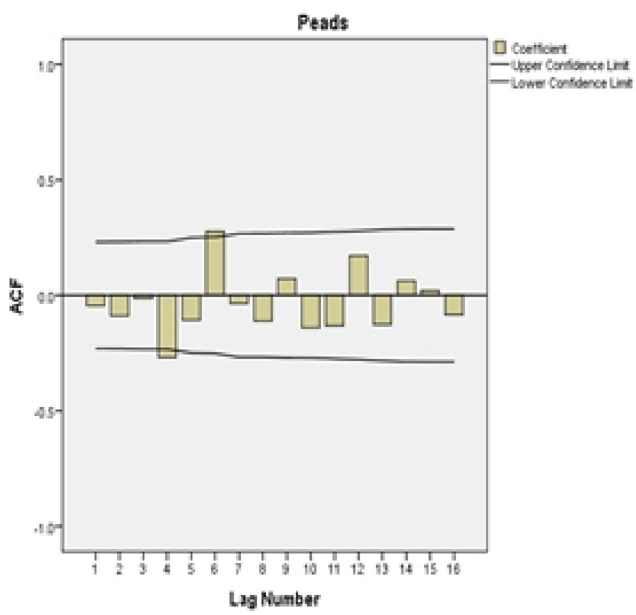


Fig No.1 Time plot of the data

3.1 STATIONARITY CHECKING

Autocorrelation Function (ACF) is a widespread tool to check the stationarity of the time series data. ACF & PACF are also used to propose the order of the tentative model. The data is adjusted for stationarity at d=1 and seasonality at D=0. Apart from the graphic methods, the ADF unit root test, also a very popular method to check the stationairty of the data was also performed to confirm the stationarity of the data shown in table No. 1.



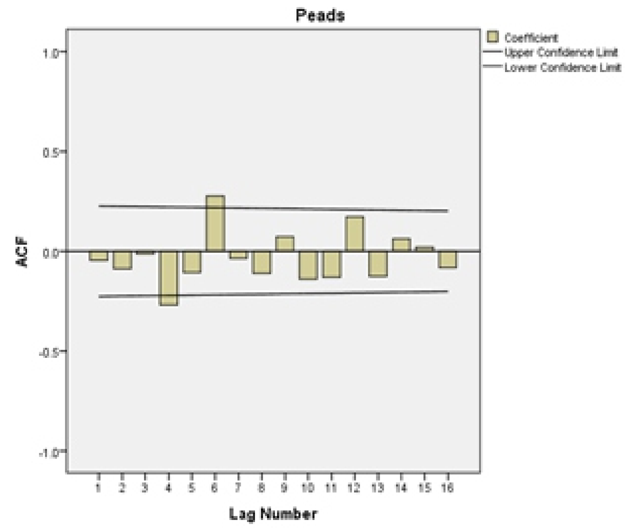


Fig No.3 Correlogram Showing Stationary Behavior of Time Series Data

Table No.1 ADF test confirming stationary behavior of time series data at different level

Variable	ADF test statistic (p value) d=0	ADF test statistic (p value) d=1
Peads	-2.85 (0.06)	-8.97 (0.000)

3.2 MODEL SPECIFICATION

The order of model must be determined before estimation of parameters of the model. The ACF & PACF of Peads data is shown in Fig 5. The correlogram (Graph of ACF & PACF) provides the basis of order determination of the model. Seasonality is easily observed through correlogram at different lags. Spike at regular lags is an indication of seasonality. The tentative models are ARIMA (1, 1, 1) and SARIMA (1, 0, 1) at seasonal point s=4. The estimated (1, 1, 1) * $(1, 0, 1)_4$ SARIMA model is proposed and can be described as

$$W_t = \phi_1 W_{t-1} + \Phi_1 W_{(t-1)4} + Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{(t-1)4}$$

All parameters of the models are estimated using Maximum Likelihood Estimation (MLE) method. The diagnostic checks were also done using Ljung & Box methods. In addition to that, the best model was selected on the following criteria with maximum R^2 and minimum AIC/BIC, RMSE, MAE and MAPE (Table No.2). The final fitted model is as below after diagnostic checks:

$$W_t = 97.61 + 0.7353W_{t-1} + 0.2953W_{(t-1)4} - 0.1036Z_{t-1} - 0.1036Z_{(t-1)4}$$

3.4 FORECASTING

Three months forecasting of Peads patient's is estimated on the basis of best selected model is shown in table No. 3.

4. CONCLUSION

A best model for Peads patient's data is developed for forecasting on the basis of available tools. The empirical analysis indicated that SARIMA (1, 1, 1)*(1, 0, 1)$_4$ is best fitted model for patients data for short run forecasting. The estimated number of patients for next three months is given in Table No. 10 and compare with actual data. The estimated results of model showed that Peads incoming is influenced by seasonal variation of data. The explained 67% variation is also accounted for model. This is good indicator of the model.

The finding of this study may be useful for future planning, clinical researchers, hospital administration, OPML management to ensure quality diagnostic healthcare service, and Government also for the welfare and betterment of patients and healthcare professionals.

1.ACKNOWLEDGMENT

We are highly thankful to the staff of OPML, who helped us in compiling data especially Mr. Muhammad Zubair Qamar.

2.ETHICAL CONSIDERATION

This study was carried out on the patient's data available at OPML records and there was no direct interaction with patient, no patients personal identification was used, so the study has not reviewed by Institutional Review Board (IRB).

Table No. 2: the summery of diagnostic check for peads data

Parameters	Proposed models for peads data	
	SARIMA (1,1,1)*(1,1,1)$_4$	SARIMA (1,1,1)*(1,0,1)$_4$
R^2	0.095	0.67
BIC	11.03	10.98
AIC	10.88	10.80
RMSE	84.48	77.51
MAPE	167.61	152.37
MAE	66.22	61.67

Table No.3: the forecasted and actual number of patients

Month	ARIMA(1,1,1)		SARIMA (1,1,1)*(1,0,1)$_4$	
	Estimated	Actual	Estimated	Actual
Jan 2014	119	102	92	102
Feb 2014	118	98	90	98
March 2014	117	110	110	110

REFERENCES:

1. BOX G.E.R., JENKINS G.M (1976). Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco.
2. Box GEP, Jenkins GM, Reinsel GC: Time Series Analysis: Forecasting and Control 3rd edition. Englewood Cliffs, NJ: Prentice-Hall; 1994
3. Suhartono & Muhammad Hisyam Lee (2011). Forecasting of Tourist Arrivals Using Subset, Multiplicative or Additive Seasonal ARIMA Model. MATEMATIKA, 27(2), 169–182
4. Etuk, E. H. (2012). Forecasting Nigerian Naira-US Dollar Exchange Rates by A Seasonal Arima Model. American Journal of Scientific Research, 59, 71-78
5. Etuk E.H., Igbudu R.C. (2013). A SARIMA fit to Monthly Nigerian Naira- British Pound Exchange Rates. Journal of Computations & Modeling, :3(1) 133-144
6. Okereke O.E., Bernard C.B. (2014). Forecasting Gross Domestic Product In Nigeria Using Box-Jenkins Methodology. Journal of Statistical and Econometric Methods: 3(4) 33-46
7. Tankersley, C. D., Graham, W. D. and

Hatfield, K (1993). Comparison of univariate and transfer-function models of groundwater fluctuations. Water Resources Research. 29: 3517–3533.

8. Pozza, S. A., Lima, E. P., Comin, T. T., Gimenes, M. L. and Coury, J. R.(2010) Time series analysis of PM2.5 and PM10-2.5 mass concentration in the city of Sao Carlos, Brazil. International Journal of Environment and Pollution. 41: 90–108.
9. Durdu, O. F.(2010) Stochastic approaches for time series forecasting of boron: a case study of western turkey. Environmental Monitoring and Assessment. 169: 687–701.
10. Abraham, G., Byrnes, G. B. and Bain, C. A.(2009) Short-term forecasting of emergency inpatient flow. Ieee Transactions on Information Technology in Biomedicine. 13: 380–388.
11. Tseng, F. M., Tzeng, G. H. and Yu, H. C.(1999). Fuzzy seasonal time series for forecasting the production value of the mechanical industry in Taiwan. Technological Forecasting and Social Change 60: 263–273.
12. Ediger, V. S. and Akar, S.(2007). ARIMA forecasting of primary energy demand by fuel in Turkey. Energy Policy. 35: 1701–1708.
13. Scortti, M., Cattan, P. and Canals, M. (1997) Forecast of canine rabies in Argentina, Bolivia and Paraguay, using time series analysis. Archivos de Medicina Veterinaria. 29: 83–89.
14. Imhoff M., Bauer M., Gather U., Löhlein D. (1997). Time series analysis in intensive care medicine. Appl Cardio Patho. (6): 263-281
15. Hwang, U, Richardson, L.D, Sonuyi, T.O, Morrison, R.S. (2006). The effect of emergency department crowding on the management of pain in older adults with hip fracture, Journal of the American Geriatrics Society. 54(2): 270–5.
16. Touloumi G, Pocock SJ, Katsouyanni K, Trichopoulos D. (1994).Short-term effects of air pollution on daily mortality in Athens: a time-series analysis. Int J Epidemiol. 23: 957-67
17. Akhtar S, Rozi S. (2009). An autoregressive integrated moving average model for short-term prediction of hepatitis C virus seropositivity among male volunteer blood donors in Karachi, Pakistan. World J Gastroenterol 15(13): 1607-1612
18. Berman JJ, Meyer J, Coats G. (19984) Effects of program characteristics on treatment outcome: an interrupted time-series analysis. J Stud Alcohol 45: 405-10
19. Lambert, C.R,. Raymenants, E, Pepine C.J.(1995). Time-series analysis of long-term ambulatory myocardial ischemia: effects of beta-adrenergic and calcium channel blockade. Am Heart J 129: 677-84
20. Hoot NR, Zhou C, Jones I, et al.(2007). Measuring and forecasting emergency department crowding in real time. Ann Emerg Med 49: 747-55
21. Akaike H. (1978). On the likelihood of a time series model. Statistician 27: 217-35
22. Schwarz G. Estimating the dimension of a model. Ann Statist 1978; 6: 461-4
23. Dickey, D.A., Fuller, W.A.(1979). Distribution of estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. 74: 427–431
24. Perron, P. (1998). Trends and random walks in macroeconomic time series: further evidence from a newapproach. J. Economic Dynamics and Control12, 297–332.
25. Zivot, E. A., Donald, W.K. (1992). Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis, Journal of Business & Economic Statistics, American statistical Association, 10(3): 251-70
26. Kwiatkowski D., Peter C.B.P., Peter, S., Yangchow, S. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root? Journal of Econometrics 54(1-3): 159-178
27. Ljung, G.M., Box, G.E.P. (1978) on a measure of lack of fit in time series models. Biometrika 65: 297-303

Submitted for publication: 05.06.2015
Accepted for publication: 10.02.2016