Time Series in the Study of Seismic Regime of Vrancea (Romania) Seismic Zone

R.Z. Burtiev*

Moldavian Academy of Sciences, Institute of Geology and Seismology, 3 Academiei Str, Chisinau, Moldova

Abstract: In the series of monthly number of earthquakes is present long-term systematic component. The assumption about stationary of the average value and the variance is rejected. Statistically significant autocorrelation coefficients mean that the time series is not random, and there is some connection between successive levels. In order to predict a series of exponential smoothing method was used. Best model is simple seasonal for the logarithm of the levels, and Winters additive for the square root level of the original series. The research of time series confirms that the *ARIMA* (0, 1, 2) x (1, 0, 1) $_s$ model can be used to the analysis and prediction of uniform non-stationary time series with a nonlinear trend, such as a polynomial of low degree. Prediction for 2012 is computed using simple Simple seasonal, Winters and ARIMA models.

Keywords: Simple seasonal, winters, ARIMA models, autocorrelation, periodogramm.

1. INTRODUCTION

The theory of time series (TS) may be useful in the study the phenomena like period of increased seismic activity and variations of seismicity in time, etc. Observations of active seismic zones show that the temporal characteristics of seismicity are not stationary. The periods of variations of earthquakes number (monthly, yearly) are scattered randomly on the time axis. Under the schedule of the monthly number of Vrancea focus earthquakes, we cannot definitely judge about the presence of the regularities in the duration of the periods of variations of the number of earthquakes and in the alternation of periods of seismic quiescence and high seismic activity. The stimulus for this study was the desire to analyze the structure of some formal methods for finding regularities in the statistical variations of seismicity parameters in time.

Seismicity is characterized by: 1) frequency of earthquakes; 2) statistical distribution of thrust force (magnitude); 3) spatial distribution of lesions; 4) macroseismic observations of strong seismic events.

For the investigation of seismicity can be applied spatial models and time series models. Spatial model describes a set of parameters of seismicity at a given time.

Time series represents a series of regular observations over a certain parameter of seismicity at successive times or time intervals. In this paper, a time series model is used to investigate the probabilistic structure of the flow of seismic events. Generally, the purpose of the study is to identify the time series patterns in the changing of series levels and the elaboration of the model in order to predict and study the relationships between phenomena.

The theory of time series is used to solve the following main tasks: 1) determination of the nature the series; 2) determining the main parameters of the series; 3) prediction of future values (monthly number of earthquakes) time series according to available data. Time series can include the following components:

1) Trend;

2) Cyclical component;

3) Seasonal component;

4) Random component.

Time series consists of deterministic and random parts. Deterministic part (trend, cyclical and seasonal components) is used for prediction of future values of the series. For a correct reflection of the real process by time series data sample must meet the following conditions: comparability; homogeneity; sustainability; a sufficient volume of data.

2. DATA ANALYSIS

Analysis is subject to a set of monthly number of earthquakes in the Vrancea earthquake zone during the period from 1978 to 2011. Levels of the series are monthly number of earthquakes. The sample observation period must be sufficiently long to identify changes in seismic events. In practice, a reference

*Address correspondence to this author at the Moldavian Academy of Sciences, Institute of Geology and Seismology, 3 Academiei Str, Chisinau, Moldova; Tel: +(373) 78097790; Fax: 37322 739729;
E-mail: burtiev_rashid@mail.ru

E-ISSN: 2410-3624/14

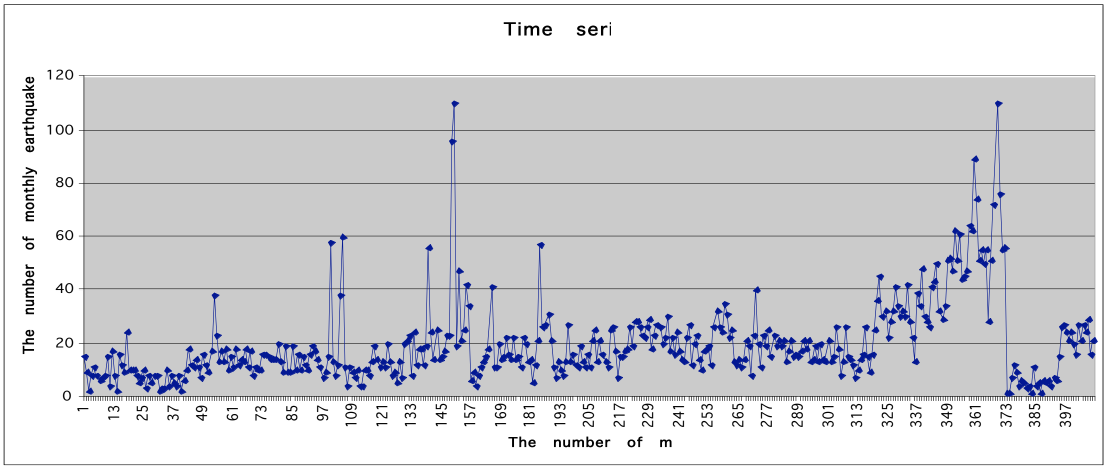


Figure 1: The series schedule.

period including more than 50 observations is considered sufficient [1]. The series of monthly number of earthquakes in Vrancea seismic source is analyzed.

Preliminary stage of statistical analysis included checking the uniformity of the sample. The statistical test used to measure uniformity is the coefficient of variation [2-4]:

$$V = \frac{\sigma}{a} \quad (1)$$

where σ is the standard deviation, and $\bar{a}$ is the arithmetic mean of the sample. The data are considered to be uniform if the coefficient of variation does not exceed 33%. The coefficient of variation Vq = 77% shows that the sample of monthly number of earthquakes is not uniform. According to Kolmogorov-Smirnov test the hypothesis according to a Poisson distribution of the monthly number of earthquakes rejected.

3. RESEARCH OF THE SERIES STRUCTURE

The first idea of the nature of time series can be composed according to its schedule. On the schedule of TS (Figure **1**) we can see a slow steady increase in levels of the series, thus constituting a time series components enter into it multiplicatively. Multiplicative models differ from additive models in that the additive model seasonal fluctuations do not depend on a number of levels, and in the multiplicative model amplitude of seasonal fluctuations varies depending on the values of the series. Often the number of aftershocks in the spatial and temporal vicinity of strong earthquakes increases, forming a swarm of earthquakes, including the previous (foreshocks) and subsequent (aftershocks) earthquake. Multiplicative model can be explained by a swarm of earthquakes, and the lack of seismic data availability smooth increase in the number of monthly earthquakes (trend). The graph shows that the series contains a cyclical component, and the nature of earthquakes repeated monthly. Outliers indicate periods of increased seismic activity. However, for a complete analysis of a number of structures to be built correlogram.

The independence of time series parameters can specify a stationary character from the starting time. A series is divided into two intervals, and the hypothesis of equality of means and variances of intervals is checked for revealing of this circumstance.

Table 1: Group Statistics

Group	N	Average	Standard Deviation	St. Error of Average
1	204	15,32	12,925	0,905
2	204	24,39	16,437	1,151

The data series for the periods 1978-1994 and 1995-2011 years was made up for this purpose. The criterion of Livin dispersions equality is applied to test the null hypothesis H_0, about the equality of both groups' dispersions (Table **2**). If the hypothesis is rejected, i.e. observed significance level is less than 0.05, it is necessary to use *t*-criterion with separate dispersions for comparison of average values. In this case the *F*-statistics for series is equal to F =15.085 the corresponding probability of significance α = 0.0 is less than the critical probability of 0.05, i.e. the hypothesis about equality of dispersions is improbable. *F*-statistics for the Livin criterion is obtained by conducting the one-

Table 2: Criteria for Independent Samples

	Criteria of Livin Dispersions Equality		Degrees of Freedom	T-Criterion of Comparison of Averages	
Dispersions	F	Significance		t	significance
equal	15.085	0.0	406	-6.918	0.0
not equal			384, 606	-6.918	0.0

factorial dispersive analysis for module deviation of each observation from the group average. The statistics *t* =-6.918 is obtained by using *t*-test with separate dispersions (equality of dispersions isn't supposed). This value exceeds the tabulated value corresponding to a significance level α = 0.01 i.e., about equality of average values of two groups with different dispersions is improbable. Thus, the assumption of stationary character of series on the average value and on the dispersion is rejected.

In *t*-tests of paired samples of the null hypothesis H_0: $\bar{u}$ =0 assumes equality to zero of the average of observations' differences in the two groups $u_i = x_i - y_i$ i = 1,..., n. In order to test the H_0 hypothesis statistics it is used the *t*-criterion:

$$t = \frac{\bar{u}}{\frac{\sigma}{\sqrt{n}}},\ \bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i^2,\ \sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n} u_i \quad (2)$$

It is compared with the tabulated value corresponding to a significance level α = 0.05 with (n-1) degrees of freedom (Table **2**). This criterion is applied when it is necessary to compare the conditions in which the observations are hold and the values of the time series at different time intervals. By sampling data, hypothesis about the equality to zero of average value of the series level differences for the periods 1978-1994 and 1995-2011 years is disproved - the observed probability of significance is less than α=0.05.

The statistical moments calculated for different intervals of the series are independent random variables. A parametric method for testing stationary of a series is used when probabilistic structure of the series is known. Nonparametric methods are used, when there is no information on the frequency nature of a series. Methods such as inversion method can find trends in time series and identify the statistical independence of samples [6-8]. Many statistical tests suggest that the order in which data were collected has no importance. If yes, then the sample is not random and these observations are not independent. The criterion of series is applied for hypothesis checking about a random of sample. The randomnicity in the case of time series means the absence of a trend. To use the runs test, each level is compared with the value M (median, mode, or the average of time series) [9, 10].

If the level of series- x_t is more than *M*, this value is assigned to (+); if it is less than (-), and if it is equal to *M* the sign isn't given. Series is a sequence of consecutive symbols (+) or (-). The sample is considered random if the following two inequalities are carried out [11]:

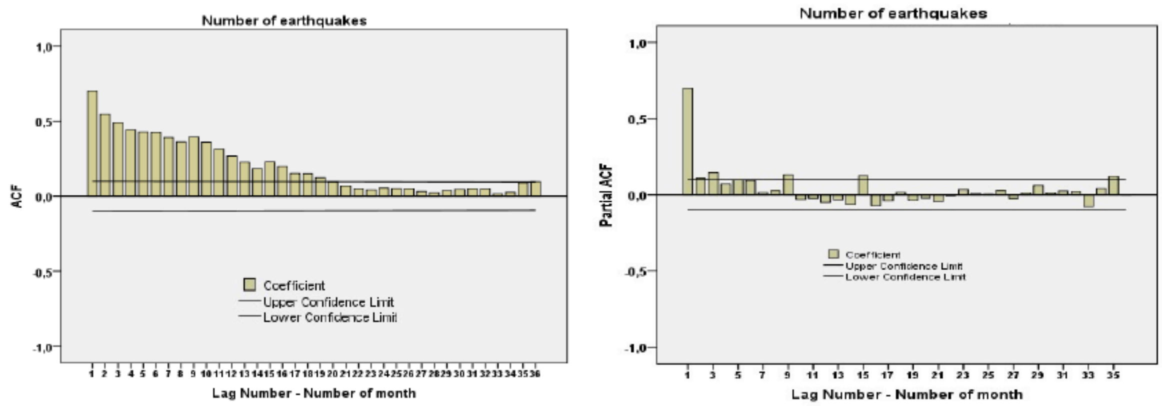
$$L_{\max} < 3.3\lg(n+1), w\ \ v_t < 0,5(n+1-u_{1-\frac{\alpha}{2}}\sqrt{n-1}) \quad (3)$$

The hypothesis about the random nature of the series is rejected, because the corresponding probability value is negligible, observed significance level is less than α=0.05.

U-test by the method of Mann and Whitney is a test for nonparametric comparison of two independent samples. It is based on the use of a common sequence of values of both samples. One variation series is made from two series. The null hypothesis assumes that the value of one sample will be distributed in regular intervals among the values of other samples. The alternative hypothesis would be consistent if the situation when the value of one of the samples will dominate on one of ends of the combined series [12, 13]. In this case, the intersection of two lines will be minimal. By the method of Mann and Whitney, hypothesis about randomness for is rejected.

Carrying out the Mozes test, the first of two series is regarded as a control. The values of both series are arranged in the order statistics and they are assigned the appropriate rank location. The scope between the rank positions is counted up in the control group. By Mozes test, the hypothesis that both series are components of one series, twice the length, also is rejected.

The basis of the Kolmogorov-Smirnov test is the calculation of the maximum difference between the



(**a**) ACF (**b**) PACF

Figure 2: (**a**)**:** Autocorrelation function. (**b**) particular autocorrelation function of the series of the series.

cumulative incidences of both series, on which the calculation of the probability of significance is based. According to the criterion of Kolmogorov-Smirnov test, the null hypothesis is rejected [14]. If the assumption of a random nature of the oscillation observations in the test series really, there should be no relation between the levels of the series. An alternative assumption allows for dependence between successive levels, that is, in this case, the time series is not random. To assess the relationship between the successive levels can be used autocorrelation function time series. However, in this case, the autocorrelation coefficient can be estimated for non stationary series, but in this case its probabilistic interpretation is lost [10].

4. AUTOCORRELATION FUNCTION OF SERIES

The aim of time series analysis is to identify a number of models. Analysis of the autocorrelation function and correlogram (4) allows revealing the structure of the series, that is, to determine in a series the presence of one or another, not priori known frequency periodic components.

$$ACF(l) = \frac{\sum_{k=l+1}^{n}(y_k - \bar{y})(y_{k-l} - \bar{y})}{\sum_{k=1}^{n}(y_k - \bar{y})^2}, \text{ k=1,…, n-1} \qquad (4)$$

Although the autocorrelation function is defined only for stationary processes, it can be calculated for any series and used to analyze the structure of the series. So, if the highest is first-order autocorrelation, which means that, the analyzed series contains only a tendency. If the time series is close to white noise, the correlogram oscillates close to the horizontal axis, and its value is close to 0 [3]. For high values of τ estimate ACF (4) of the autocorrelation coefficient has errors. This is due to the partial summation as well as observation of summing discarded. Therefore correlograms for large values of l does not reflect the true structure of the series. For fixed number of ACF (4) decreases rapidly with increasing l. If there is a trend autocorrelation function takes the form of slowly falling curve. In the case of seasonal periodicity in the graph there are peaks of ACF for lags of multiple periods of seasonality. But these peaks can be hidden by the presence of a trend or a large variance of the random component [15].

The autocorrelation function of the series shows a significant peak at lag 1 (Figure **2**). The series contains only a tendency. A statistically significant autocorrelation coefficient indicates that the levels are not random, and between successive observations there is some connection. The comparative smoothness of the series is explained by the positive correlation between consecutive levels.

The linear coefficients of autocorrelation characterize a closeness only of a linear connection between current and previous levels of the series. Therefore, it is possible to judge about the presence or absence of linear dependence by using the coefficients of autocorrelation. None zero values of autocorrelation coefficients indicate an on linear trend. The linear coefficients of autocorrelation for time series, consisting of the logarithms of the initial levels, were calculated for series checking on the presence of several non-linear trends (Figure **3**).

Formal methods of determining the trend are:

- Method of least squares
- Method of a sliding average.

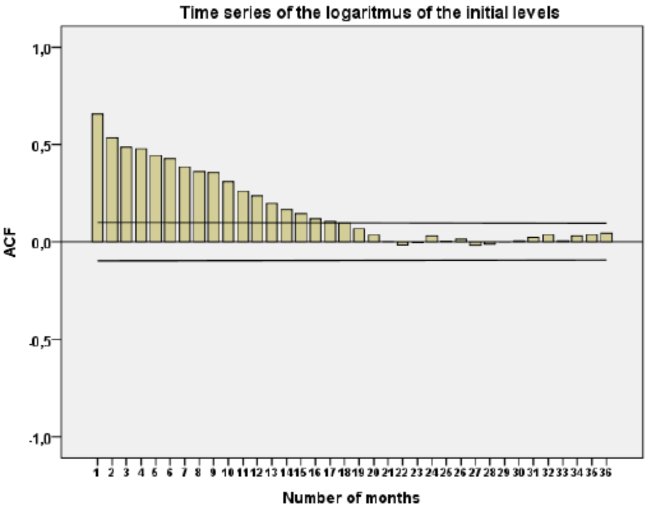


Figure 3: Autocorrelation function of the logarithms of the initial levels.

The analytic function, which characterizes the dependence of levels of a series from time, is constructed for modeling of the trends of the time series. This method is called an analytic alignment of time series.

For trend description the following functions are used:

- Linear
- Hyperbolic
- Exponential
- Polynomial.

For example, for detection of a trend in the form of sedate function, it is possible to use regression of levels of a time series on time (10):

$$x_t = a + \sum_{j=1}^{p} \beta_j t^j + \varepsilon_t \tag{5}$$

$\varepsilon_\tau \sim \iota\iota\delta(0,\sigma^2)$. If regression is not significant, it means that the trend in the time series is absent. Let's consider the linear regression model for time series levels $\xi_\tau=\alpha+\beta_\tau+\varepsilon_\tau$. Thus, the hypothesis about the absence of a trend means the plausibility of hypothesis H_0: $\alpha=0,\ \beta_\varphi=0,\ \varphi=1,\dots,\pi$- about the equality of regression coefficients to zero. For testing the null hypothesis H_0 it is calculated *F*-statistic criterion (6) [1,8,13]:

$$F = \frac{(RSS_H - RSS)/(p-1)}{RSS/(n-p)} \tag{6}$$

$$F > F^{\alpha}_{p-1,n-p} \tag{7}$$

where ν –number of observations, π -number of parameters. If (7) is valid, then the hypothesis H_0 is rejected. It means that the regression is significant and all regressors should be taken into account.

In our case, the hypothesis H_0 is rejected at the 5% significance level, the sample value F-test indicates a statistically significant regression relationship of levels over time on all models of the trend. That is, in a number of monthly numbers of earthquakes there is a tendency of growth of levels over time. The presence of a trend in a number of monthly numbers of earthquakes can be explained by the fact that:

- Improved operation of the system of seismic observations. After earthquake catalogs do not reflect the image of seismicity and do not contain information about all the seismic events that occurred in the earthquake zone. The minimum threshold magnitude earthquake catalogs completeness decreases with time. That is, the observed trend is not due to the nature of seismicity, and is caused by the lack of a representative volume of information on earthquakes.

- Return period of large seismic events is tens, sometimes hundreds of years, while the series of observations cover much shorter periods of time. For example, in this case, we investigate a monthly number of earthquakes that occurred from 1978 to 2011. The age of earthquake sources has millions years, and the length of the series must exceed the return period for the largest seismic event in zone, and the quantization step time should be not only a month, but longer intervals. Under the length of the series refers to the time elapsed from the first to the last observation in the series.

5. THE METHOD OF EXPONENTIAL SMOOTHING AND FORECASTING TIME SERIES

The theory of time-series offers large number of series prediction methods that implement the scheme extrapolation. That is, the number of investigated and it is expected that its properties do not change in the future. One of the simplest methods is simple exponential smoothing method. Despite the simplicity of used mathematical apparatus, the forecast potential of the method is not inferior to methods where applicable deeper mathematical extrapolation methods. Exponential smoothing method refers to the nonparametric method of time series analysis, since its application does not depend on the type of distribution of the random component. Exponential smoothing method makes it possible to obtain an estimate of the parameters of the trend in the non-average number, and the trend followed-up to the last observation.

One of the nonparametric methods of revealing of a trend m_t is the moving average method, in which random deviations are repaid. With this method, the smoothing values of the levels of the series are replaced by average values that characterize the mid-point of the time intervals of moving [2,16]. When we subtract from the values of levels the values of the moving average $\varepsilon_t = x_t - m_t$ we can select a random component of the series, which is used for construction of the autoregressive model to forecast. The method of exponential smoothing is used for forecasting of non-stationary time series (8):

$$S_t = \alpha x_t + \beta S_{t-1} \qquad (8)$$

where S_t-value of the exponential average at the time t; α-smoothing parameter, $0<\alpha<1$; $\beta=1-\alpha$. This formula is applied recursively - each new value is calculated as the weighted average of the observation (which is also forecast) and the smoothed series.

For the series of monthly number of earthquakes have been tried non-seasonal and seasonal

models:

- Non-seasonal models: simple; Holt and Brown (linear trend);
- Seasonal model: simple seasonal; Winters additive; Winters multiplicative.

When the series contains non-linear trend, it makes sense to convert the data. The package SPSS for data conversion used two functions: the natural logarithm and square root. To forecast time series, contains the logarithm of levels monthly number of earthquakes, optimal, based on largest sample test statistic quality control, model is simple seasonal model. Multiplicative seasonal model Winters is best for the series, which consists of the square root level of the original series. Sample values of the probability of significance 0.194 and 0.15 Box-Ljung statistics (9) is less than the significance level α = 0.05.

$$Q_{LB} = T(T+2)\sum_{\tau=1}^{m}\frac{r_\tau^2}{T-\tau} \sim \chi^2(m) \qquad (9)$$

If the series contains a seasonal component, it makes sense to smooth exponentially this component, with some additional parameter δ. Simple seasonal exponential smoothing model differs from the simple exponential smoothing, the fact that the additive model to forecast the seasonal component is added, and the multiplicative model is multiplied. In the additive model, the forecast is based on the following formula:

$$P_t = S_t + I_t \qquad (10)$$

where I_t -seasonal index adjusted at t-p, which is given by the expression [17]:

$$I_t = I_{t-s} + \delta(1-\alpha)e_t \qquad (11)$$

In multiplicative model addition is replaced by multiplying:

$$P_t = S_t I_t \qquad (12)$$

$$I_t = I_{t-s} + \delta(1-\alpha)\frac{e_t}{S_t} \qquad (13)$$

where δ is seasonal smoothing parameter, and takes values between 0 and 1, S_t - value exponentially smoothed level of series at time t, and I_{t-s} denotes the smoothed seasonal factor at time t minus s (s - number of periods in the seasonal cycle), and e_t is observed minus predicted value of the series at time t.

Winters multiplicative model used for series with a linear trend and seasonal variations, varying with the value of the series. In the multiplicative model Winters forecast for k steps ahead given by the formula:

$$Y'_{t+p} = (L_t + pT_t)S_{t-s+p} \qquad (14)$$

where Y '(t + p) levels predicted on the p periods ahead. L_t component describes the smoothed series, T_t is the value of the trend at time t, and S_t is used to assess the seasonality and estimated by the formulas:

- Smoothing the original series:

$$L_t = \frac{\alpha y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + T_{t-1}) \qquad (15)$$

- Smoothing trend

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \qquad (16)$$

- Evaluation of the seasonal component

$$S_t = \gamma\left(\frac{y_t}{L_t} + (1-\gamma)S_{t-s}\right) \qquad (17)$$

Table 3: The Predicted Values for the Series of the Season Exponential Smoothing and Multiplicative Winters Models

Month	1	2	3	4	5	6	7	8	9	10	11	12
2011	24	21	24	20	16	27	21	27	24	29	16	21
Simple seasonal 2012	20	18	23	27	29	26	26	34	31	33	26	31
Winters 2012	21	19	24	28	30	27	27	34	32	34	27	32

α, β, γ smoothing parameters which take values in the interval [0, 1].

The method is based on the use of the exponential weighted average for the season series.

The predicted values for selected models of the monthly number of earthquakes in 2012 are not significantly different (the second and third row (Table **4**.). The first line (Table **4**.) provides information about the number of monthly earthquakes in 2011. Numbers 1, 2, 12 correspond to the name of months: 1st January; February 2nd, and so on.

The quality of the forecast is determined by the maximum coefficient of determination R^2, and the optimum values of numerical characteristics of residues: RMSE; MAPE; MAE; MaxAPE; MaxAE; BIC (Table **4**). If the model is adequate, than residues is "white noise". Charts autocorrelation and partial autocorrelation functions of a number of residues have no significant peaks, and are within the 95% confidence interval for all joists.

The dynamics of the time series may have different forms, and there are many models that can be more or less reliably describe time series. Selection of the optimal model for the purpose of forecasting is difficult and is the main problem of time series theory. The time series consists deterministic and random factors, and study measures the relationship between them can carry out the structure of the series and determine its components.

6. THE SPECTRAL ANALYSIS OF TIME SERIES

The research of the frequency structure series is carried out by the procedure "Spectral analysis" of software package SPSS. Almost any periodic function can be approximated by Fourier series, the sum of sinus and cosines (22) [9, 18, 19, 20]

$$xt = a\cos(\omega t) + b\sin(\omega t), \qquad 0 < \omega \le \pi \tag{18}$$

The Fourier spectrum contains frequencies that are present in the original series. The deviation of rests from the white noise may indicate unaccounted seasonal variations. Considered series cover the period T=408 months. The sinusoidal fluctuations are applied to the entire observation period, when $\omega = 2\pi/408$. With a frequency of 2ω we will obtain two variations on the entire range of observations, because in this case, the length of the period is $2\pi/2\omega = (2\pi)\,408/4\pi = 204$ months. The frequency of $k\omega$, $k = 1, 2, \ldots, 204$ mean that there are k fluctuations, each with a period of 408/k months accordingly. Thus, the time series can be decomposed into a sum of sinusoidal oscillations of different frequencies and amplitudes (19):

$$x_t = \sum_{k=1}^{K} a_k \cos(k\omega t) + b_k \sin(k\omega t), \text{k=1}, \tag{19}$$

where: a_k, b_k are the Fourier coefficients. A graphical representation of the comparative values of the oscillations of different frequencies is called the periodogram (Eq. 24):

$$I(\omega_i) = \frac{2}{K}\left[\left(\sum_{k=1}^{K} a_k \cos(k\omega_i)\right)^2 + \left(\sum_{k=1}^{K} b_k \sin(k\omega_i)\right)^2\right] \tag{20}$$

where $\omega_i = 2\pi i/408$ is frequency. The periodograms are calculated for the detection of periodic oscillations in “white noise” [13]. The periodogram values are estimated on the basis of sample and are connected with random fluctuations. To remove random fluctuations periodogram is smoothed by the method weighted moving average.

The width of the window moving average is taken to be an odd number $m=5$. The spectral density is calculated as the average $m/2$ of the preceding and subsequent periodogram values.

Periodogram (Figure **4**) shows a sequence of random peaks, and it is impossible to talk about the significant periodic cycles of different frequencies. Form of periodogram indicates the trend [9].

In the frequency domain, the trend can be seen as a vibration with an infinitely large period and, accordingly, with a very low frequency. This trend could affect the value of the spectral density function at the left end of the range of frequencies.

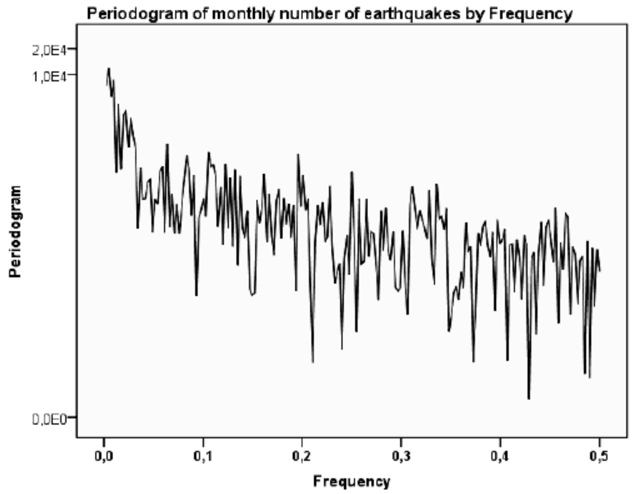


Figure 4: The periodogram of time series.

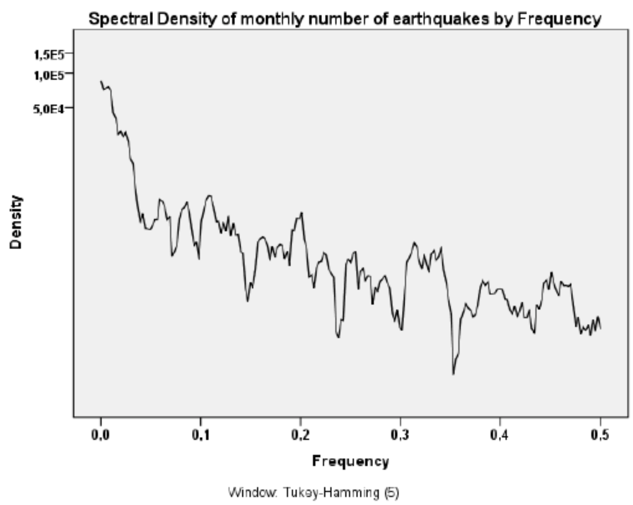


Figure 5: The spectral density function of the series.

Consequently, for the correct application of spectral analysis, a series should be decomposed into components and the trend should be removed. Different approaches to the identification of a trend given its different options, according to different values of levels refined from the trend of series. There is a question, whether these differences affect the values of the spectral density. The particular interest is represented by the low-frequency oscillations, as long-term trends in dynamics of the series are contained in the field of low frequencies. The better the model trend component, the more vibrations it contains, as the approximation is better.

7. LINEAR TIME SERIES MODELS

Above it was shown that the time series is non-stationary. Non-stationary time series can be modeled as an *ARIMA*-model based on a 3-component process: the autoregressive order p; moving average of order q; "random walk" $\xi_t \sim N(0,\sigma^2)$. The procedure “Design Models” package *SPSS* is used to construct the time series models. The mathematical form of the model has the form [21]:

$$\Delta^d x_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j \Delta^d x_{t-j} - \sum_{n=1}^{q} \beta_n \varepsilon_{t-n} + \xi_t \tag{21}$$

If the series is non-stationary, it can be reduced to a weakly stationary series by taking successive differences of same order (22):

$$\begin{aligned} &\Delta x_j = x_j - x_{j-1}, \Delta^2 x_k = \Delta x_k - \Delta x_{k-1}, \\ &\Delta^d x_l = \Delta^{d-1} x_l - \Delta^{d-1} x_{l-1} \end{aligned} \tag{22}$$

The original series are returned at *d*-fold summation of the integrated series. *SPSS* package contains a procedure for integrating the time series and the choice of suitable orders of *ARIMA* -process (*p, d,* and *q*). The diagnostic test of the adequacy of models is based on the analysis of the series of rests. If the model is adequate, then the rests are "white noise”. It means

Table 4: The Values of Numerical Characteristics of Residues

Model	Model Fit Statistics								Ljung-Box Q(18)		
	Stationary R-squared	R-squared	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC	Statistics	DF	Sig.
Simple seasonal	0.624	0.490	11.037	61.05	7.028	5848.347	68.358	4.832	20.610	16	0.194
Winters' Additive	0.623	0,487	11,086	61,762	7,051	6004,718	68,258	4,856	20,598	15	0.150
ARIMA(0,1,2) x*(1,0,1)$_s$*	0.216	0.468	11.309	66.082	7.121	6878.756	71.526	4,896	17,296	15	0.302

Table 5: The Values of Series Predicted by the Models

Month	1	2	3	4	5	6	7	8	9	10	11	12
Observed in Last Years												
2011	24	21	24	20	16	27	21	27	24	29	16	21
Forecasted Values of Monthly Number of Earthquakes in 2012.												
Simple seasonal	20	18	23	27	29	26	26	34	31	33	26	31
Winters' Additive	21	19	24	28	30	27	27	34	32	34	27	32
*ARIMA(0,1,2)*x *(1,0,1)$_s$*	22	21	24	27	29	27	27	32	31	32	28	31

that the correlation coefficients between the levels of a series of rests are equal to zero and the normalized cumulative periodogram deviates little from the straight line connecting the point (0,0) and (0.5; 1). The models (*p,d,q*)*x(P,D,Q)* at different values of traditional and seasonal options were considered in the constructor of the time series models. *ARIMA (0, 1 ,2)*x*(1,0,1)$_s$* model were determined in the constructor of *SPSS* package models for the time series. The plausibility of the null hypothesis H_0 that the autocorrelation coefficients between the levels of a series of rests are equal to zero is tested by the Ljung-Box criterion (9).

The selected statistic (9) is less than the critical value of χ^2 distribution with *(k-p-q)* degrees of freedom corresponding to the level of significance α=0.05. The dimension of the observed probability value *0.302* for series, exceed the probability of error of the first kind α=0.05.

The best models of a series are defined by the maximum coefficient of determination and by the optimal values of the numerical characteristics of a series of rests: *RMSE; MAPE; MAE; MaxAPE; MaxAE; BIC*. Typically, the forecast shouldn't go beyond the set time of one cycle (Tables **3-5**).

8. CONCLUSION

In the series of monthly number of earthquakes is present long-term systematic component. The assumption about stationary of the average value and the variance is rejected, because the autocorrelation at lag 1 is maximal. Statistically significant autocorrelation coefficients mean that the time series is not random, and between successive levels there is some connection. In order to predict a series of exponential smoothing method was used. Best model was simple seasonal, for the logarithm of a levels, and Winters additive model for the square root level of the original series. The research of time series confirms that the autoregressive methods - *ARIMA(0,1,2)*x*(1,0,1)$_S$* can be used to the analysis and prediction of uniform non-stationary time series with a nonlinear trend, such as a polynomial of low degree.

ACKNOWLEDGEMENTS

The author is grateful to Prof. Vasile Alcaz for useful suggestions for carry out time series analysis of seismic regime.

REFERENCES

[1] Box J, Jenkins G. Analiz vremennyh ryadov. Prognoz i upravlenie (in Russian), Vyp 1 Mir: Moskva 1974.

[2] Christine Bernhardt. Modellierung von Elektrizitätspreisen durch lineare Zeitreihenmodelle und Value-at-Risk Schätzung mittels Methoden aus der Extremwerttheorie (in German), Technische Universität München, Zentrum Mathematik, München 2007.

[3] Rasid Burtiev, Fernanda Greenwell, Vitalii Kolivenko. Time series analysis of wind speed and temperature in Tiraspol, Moldova. Environmental Engineering and Management Journal 2013; 12: 23-33.

[4] Eliseeva II, Juzbashev MM. Obshaya teoriya statistiki. (in Russian). Financy i Statistika: Moskva 2004.

[5] Fisz M. Probability and Mathematical Statistics (in German), VEBDeutscher Verlag der Wissenschaften: Berlin 1966.

[6] Marcel Dettling, Angewandte Zeitreihenanalyse (in German). Zürcher Hochschule für Angewandte Wissenschaften: Zürich 2008.

[7] Kislyak NV. Econometrics (in Russian), Ekaterinburg 2007.

[8] Toutenburg H, Heumann C, Deskriptive Statistik. Eine Einfuhrung in Methoden und Anwendungen mit R und SPSS (in German). Springer-Verlag: Berlin-Heidelberg, 2008.

[9] Kendall M. Vremennye ryady (in Russian). Nauka: Moskva 1981.

[10] Spezialvorlesung. Zeitreihenanalyse mit Beispielen in Mathematica (in German). Institut für Stochastik. Johannes Kepler Universitat Linz: Linz 2006.

[11] Soong TT. Fundamentals of Probability and Statistics for Engineers. State University of New York at Buffalo: New York 2004.

[12] Shanchenko NI. Lekcii po ekonometrike (in Russian). Ulyanovsk 2008.

[13] Teusch A. Einführung in die Spektral- und Zeitreihen analyse mit Beispielen aus der Geodäsie (in German). Verlag der Bayerischen Akademie der Wissenschaften: München 2006.

[14] Kramer G. Matematicheskie metody statistiki (in Russian), Mir, Moskva 1975.

[15] Vaidogas ER, Juocevicius V. A critical estimation of data on extreme winds in Lithuania. Journal of Environmental Engineering and Landscape Management 2012; 19: 178-88. http://dx.doi.org/10.3846/16486897.2011.579452

[16] Leshinsky OL, Ryazantseva VA, Yunkova OO (in Ukrainian). Ekonometriya: Kiev 2003.

[17] Glass GV, Glass GV, Willson VL, Gottman JM. Design and Analysis of Time- Series Experiments. Colorado Associated University Press: USA 2009.

[18] Kolmogorov AN, Fomin SV. Elementy teorii funkcii i funkcional'nogo analiza (in Russian), Nauka: Moskva 1976.

[19] Piskunov NS. Differencial'noe i integral'noe ischislenie (in Russian), Nauka: Moskva 1964.

[20] Rytov SM. Vvedenie v statisticheskuyu radiofiziku (in Russian), Nauka: Moskva, 1976.

[21] SPSS Trends ™ 14.0, Copyright ©2005SPSS Inc. Chicago 2005.

Received on 14-12-2014 Accepted on 24-12-2014 Published on 10-01-2015

DOI: http://dx.doi.org/10.15377/2410-3624.2014.01.02.4