U.P.B. Sci. Bull., Series C, Vol. 83, Iss. 1, 2021 ISSN 2286-3540

IMPROVED ALGORITHM FOR SEISMIC EXTREME EVENTS PREDICTION

Theodora DUMITRESCU[1], Radu DOBRESCU[2]

The paper proposes an improved variant of the classic M8 extreme events prediction algorithm, opting for a modeling framework with additional parameters which allow better adaptation to the local characteristics of the seismic zone of interest, in this case the Vrancea region. These characteristics include: the assessment of seismic cycles, the distribution of earthquakes by magnitude, the spatio-temporal clustering of earthquakes and a seismicity model that can foreshadow large magnitude earthquakes. The efficiency of the proposed algorithm is compared with those of the classical algorithm by simulations based on numerous scenarios combining the intervals of selection of historical and prediction data, respectively.

Keywords: extreme events, earthquakes occurrence, prediction algorithms, time series

1. Introduction

In recent decades, biological threats and ecological disturbances have become more and more numerous, mainly due to economic and political globalization, accelerated technology development, global warming, etc. These real disasters can be considered as Extreme Events (EE) Not only because of their complexity, but especially because of their rarity, EE are difficult to study and even more difficult to predict.

There are numerous studies that present a lot of methods for describing, understanding and predicting EE that can occur in a number of natural phenomena. Evolved software tools have been developed to extract the features of these events from real data, as well as mathematical models of the chaotic processes that govern the dynamics of the system. On the other hand, due to the lack of concrete knowledge and limited computing power, the application of these models is sometimes imperfect and can lead to inaccurate results and as such unacceptable. In this paper we wanted to develop improved computing algorithms, which reduce the calculation time by extracting from the historical

[1] PhD eng., Automatic Control and Industrial Informatics Department, University POLITEHNICA of Bucharest, Romania, e-mail: theodora_dumitrescu@yahoo.com
[2] Prof., Automatic Control and Industrial Informatics Department, University POLITEHNICA of Bucharest, Romania, e-mail: rd_dobrescu@yahoo.com

time series a small number of essential samples, without reducing the calculation accuracy and as such the quality of the forecast.

Research on EE occurring in highly complex nonlinear dynamic systems is largely based on the development of a mathematical framework for the construction of appropriate nonlinear models that aim to predict both observed and hidden EE, using observed time series (sometimes with uncertain data). Unlike classical regression models based on ad-hoc data, these nonlinear models deal also with the physical constraints that condition them.

A recent paper written by Chen and Majda [1] summarizes observations on results obtained in the last decade by numerous EE specialists (the paper has 136 references), and also presents a new mathematical framework that allows the incorporation of various strategies in the development of appropriate models for EE prediction based on both observed and hidden data. As examples, we mention the development of a new dynamic statistical model to predict geophysical EE and climatic anomalies [2]; construction of a parametric model for EE prediction in complex geophysical flows [3] as well as forecasting statistical responses associated with uncertainty quantification ([4], [5]); efficient schemes for solving problems related to rare events [6] and control algorithms that can suppress instability at an early stage [7].

2. Prediction of seismic activity in Vrancea region – M8 algorithm

It is obvious that when discussing Extreme Events in Romania, the first topic is that of major earthquakes that can occur in the Vrancea region. In the literature, seismic activity in the area was first reported in [8], a paper that also contains a prediction study based on the M8 algorithm proposed by Kossobokov since 1993. This study was returned with various tests included in applications in real time both globally and regionally ([9], [10], [11]), but the consecration is resumed in a complex high-resonance study developed by a large team of specialists led by M. Ghil [12]. The purpose of the M8 algorithm is to decide based on the known seismic activity before the moment t, if an earthquake of magnitude $M \geq M_0$ will take place or not in the period of time (t, $t+1$). In particular, for the Vrancea region the magnitude M_0 was considered 6 on the Richter scale.

The Vrancea seismic region (red area in Fig. 1) [17] has experimented some of the most severe earthquakes in Central Europe. During the documented history, the most important earthquakes in Vrancea were on 26.10.1802 (magnitude 7.7), 10.11.1940 (magnitude 7.7, hypocenter at 140-150 km depth); 4.03. 1977 (magnitude 7.7; hypocenter at 109 km depth). That is why the seismic data of this area were selected in various research projects to test different methodologies for analysis and prediction of seismic activity.

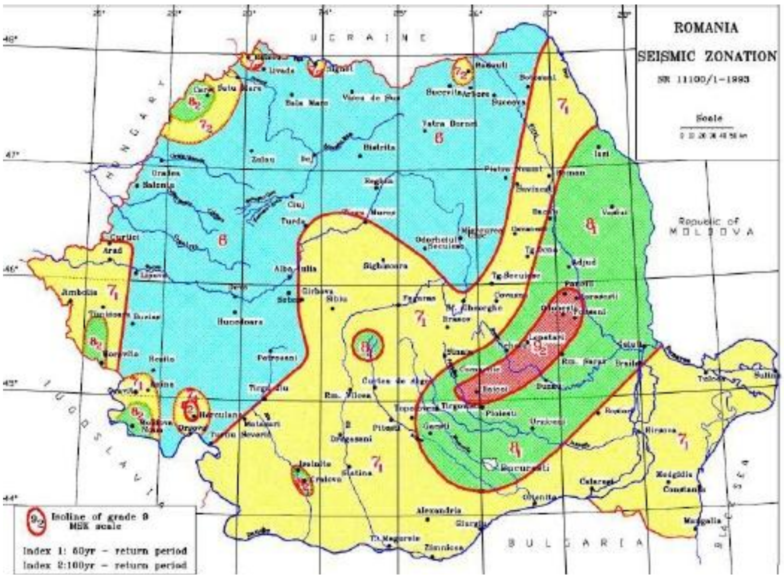


Fig.1. The seismic map of Romania

The data set on seismic activity in the Vrancea region was compiled from the seismic catalogue available at the National Institute of Earth Physics (https://web.infp.ro/#/romplus). Fig. 2 presents a selection of the results obtained by applying the M8 algorithm, using the half-yearly updates of the RomPlus catalog [16] from January 2007 to July 2010 (inclusive).

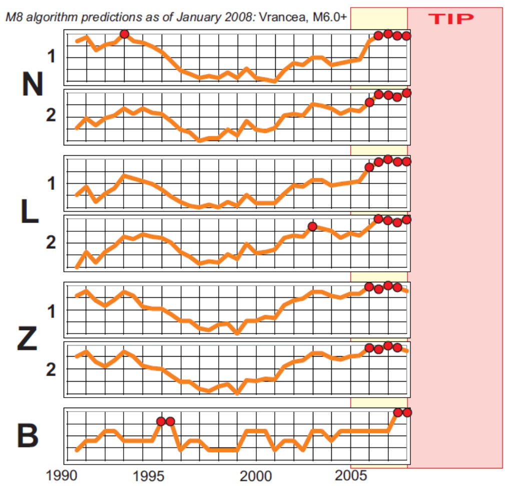


Fig. 2. Simulation of seismic activity using M8 algorithm (taken from [12])

The diagrams predict that the region will enter in a Time of Increased Probability (TIP) range for 6.5+ magnitude earthquakes in mid-2006. This TIP was due to expire by June 2011.

The diagrams are normalized plots having the following meaning: seismic speed (rate of occurrence) *N* (two measures, *N*1 and *N*2), differential rate *L* (two measures, *L*1 and *L*2), dimensionless concentration *Z* (both measures *Z*1 and *Z*2), as well as a measure *B* that characterizes the degree of clustering (concentration in a small area of the geographical area) of earthquakes. The values of these quantities which exceed an imposed threshold are marked by red circles. The M8 algorithm triggers a TIP alarm when the state vector (*N*1; *N*2; *L*1; *L*2; *Z*1; *Z*2; *B*) enters a predefined subset of this phase space. In order to verify the veracity of the graphs in Fig. 2, Table I [16] shows the real data regarding the earthquakes with magnitude $M \geq 4.7$ between January 2007 and August 2010.

Table I

EARTHQUAKES WITH MAGNITUDE ≥ 4.7 (OCCURRRED 01.2007 – 08.2010)

Date	Hour	Long.	Lat.	Hypocenter (km)	Magn.
17.01.2007	13:17:21	45,60N	26,39E	128,6	M=4,9
14.02.2007	06:56:37	45,49N	26,26E	149,8	**M=5,0**
15.02.2007	02:32:53	45,82N	26,76E	101,9	M=4,7
09.03.2007	21:44:16	45,56N	26,31E	159,6	M=4,8
30.01.2008	13:09:30	45,58N	26,48E	146,1	M=4,9
11.02.2008	03:31:30	45,62N	26,60E	117,0	M=4,8
25.05.2008	09:31:33	45,53N	26,28E	158,4	M=4,7
31.05.2008	04:32:43	45,66N	26,60E	147,8	M=4,7
03.07.2008	00:55:11	45,58N	26,33E	159,5	M=4,5
04.07.2008	23:19:26	45,63N	26,53E	151,0	M=4,9
02.10.2008	14:04:48	45,63N	26,47E	148,4	M=4,8
01.12.2008	21:23:03	45,66N	26,51E	151,0	M=4,7
25.04.2009	17:18:48	45,68N	26,61E	109,6	**M=5,5**
12.05.2009	01:15:12	45,55N	26,39E	134,5	M=4,8
27.05.2009	03:12:50	45,69N	26,49E	151,9	M=4,8
24.07.2009	20:27:09	45,70N	26,61E	140,2	**M=5,1**
26.12.2009	23:04:39	45,75N	26,72E	107,3	M=4,5
08.06.2010	15:16:00	45,60N	26,41E	113,0	M=4,7

The original M8 algorithm must be run in 5 stages [8]:

1. Determining the target threshold magnitude.

The algorithm was designed to forecast earthquakes with magnitudes greater than M_0 (with a maximum value of 8 for M_0, hence the name of the algorithm). The range of values of the target events is between M_0 and $M_0 + \Delta M$, with ΔM less than 1, the recommended value being $\Delta M = 0.5$.

2. Determining the area of investigation.

The surface of the studied region is completely covered by overlapping Investigation Circles (IC), with a fixed radius $R(M_0)$ which is chosen depending on the size of the area where the target earthquakes occur, being usually three to five times smaller than this.

3. Preparation of the initial data set (also called catalog).

In the original version, only the main shock sequence is considered in each IC[R], and for this it is necessary to eliminate the replicas (aftershocks).

4. Calculation of the main functions M8.

For each IC[R], in a time window (t_s, t) and with the magnitude range [M, M_0], such that $M \leq m<M_0$, several mean values are calculated to quantify the characteristics of the seismic sequence, as it follows:

(i) $N(t)$: this function represents the number of earthquakes of magnitude m or greater in the time window (t-s, t).

(ii) $L(t)$: Deviation of seismic activity from the trend of long-term evolution in a period (t_0, t).

(iii) $Z(t)$: The ratio of the mean value of the diameter of the sources at the average value of the distances between the sources, representing the concentration of main shock sources.

$$Z(t) = Z(t \mid m, M_0 - g, s, \alpha, \beta), \tag{1}$$

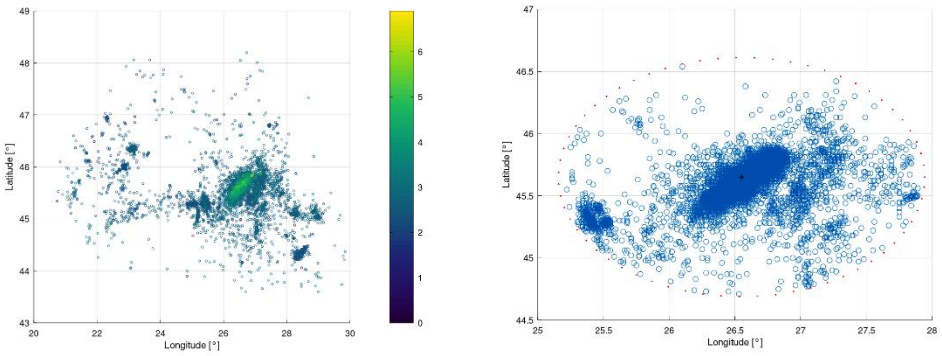
In this expression g is a non-negative weighting coefficient, with a maximum value of 1, which depends on the distance between the locations of the sources, and α and β are parameters used to determine the average diameter of earthquake sources in the chosen range of magnitudes.

(iv) $B(t)$: the maximum number of replicates, $b_i(e)$, between the events having the magnitude in the range [M_0-p, M_0-q] in the time interval (t-s, t).

Each of the functions N, L and Z is calculated twice for the magnitude limits $\widetilde{N}$ = 10 and $\widetilde{N}$ = 20. Therefore, the seismic sequence is characterized by the average values of the seven functions, $N1$, $N2$, $L1$, $L2$, $Z1$, $Z2$ and B.

5. For each function the threshold values that can trigger an alarm are determinated. The condition is that their number exceeds a percentage Q [%] of the total values encountered. If at least 6 of the 7 functions mentioned above are higher than the selected threshold in a narrow time window (t_u, t) an alarm is triggered and at the same time the increased probability interval (TIP) is initiated for a duration Δ = 5 years.

For achieving a high level of confidence of EE prediction with the M8 algorithm, the seismic area for which the investigation is made must be very correctly mapped, for the coverage with CI investigation circles to be effective. For example, we present in Fig. 3a the graph of the geographical distribution of earthquakes with magnitude greater than or equal to the threshold M = 4.5 and the investigation circle in the Vrancea area with the center at the point with the coordinates 45.68 north latitude and 26.61 east longitude in the period 01.01.1990 - 31.03.2020. In Fig. 3b is represented only the area of interest (from the investigation circle) after the application of an aftershocks filtering algorithm (mandatory in the running of the M8 algorithm).



a b

Fig. 3 Geographical distribution of earthquakes with magnitude ≥ 4.5 in a IC in the Vrancea area

Summarizing the characteristics of the M8 algorithm and the results of different applications published in the literature we can say that:

1. It is necessary to map as accurately as possible the seismic area for which the forecast is desired, so as to cover this area with overlapping investigation circles.

2. The selection of the number of seismic events of interest is performed by the parameterization of the seismicity level in each investigation circle.

3. The determination of an interval with increased probability (TIP) is made on the basis of the numerical values obtained in the previous step.

Returning to the diagrams presented in Fig. 2 and to their correlation with the real data from Table I one can observe that the simulation based on this model predicted the occurrence of 4 earthquakes with $M \geq 6.0$ or larger in the interval 2007 - 2010. In reality, from the concrete data on major earthquakes that occurred between January 2007 and August 2010, only 3 earthquakes with a magnitude greater than or equal to 5 occurred (see Table I). This leads us to the conclusion that in the model there are systematic errors that could be compensated by adjusting the parameters or by adding additional weighting parameters. The next section describes how these adjustments can be made.

4. Improved version of the M8 earthquake prediction algorithm

The main limitation of the performance of the M8 algorithm is related to the way in which the seismic region is covered with ICs. To optimize this coverage, we used a combination of two digitized maps: the location of the ICs centers in the most prone to earthquakes and the seismic clusters (density of anomalies), to take into account the impact of both factors, simultaneous. Therefore, all algorithms were run on the ICs with the center in points with high seismic potential, but at the same time placed in areas with high density of such points. At the same time, we included in the algorithm changes in the radius of the investigation circle and the lower magnitude limit, to take into account the aftershocks with a magnitude exceeding 40% of the magnitude of the main earthquake. To increase the prediction accuracy we added a new term, β, in the formula for calculating the investigation radius. The resulting IC[R] radii for different values of the β parameter are listed in Table II.

Table II

CI RADII CALCULATED FOR DIFFERENT β VALUES

β	1	1.25	1.5	1.75	2
R (km)	108	114	123	148	160

The other change made to the M8 algorithm was related to the change of the limit of the minimum cutting magnitude, $\widetilde{M} = M_{min}(\widetilde{N})$, where $\widetilde{N}$ is the annual average number of events (AANE). Consequently, in simulations we tested various pairs of AANE, such as 5-10, 10-15, 10-20 years.

A notable observation about the changes related to the β parameter is that as the investigation radius increases, the difference between the predictions made with the two algorithms is reduced. In fact, increasing the investigation radius leads to an increase in the seismicity rate of each investigation circle. We can conclude that the alarm zone is not influenced by the smaller cutting quantities.

The mentioned changes allowed a greater freedom in grouping historical data and a better adaptation to the specifics of the local area (Vrancea). However, these changes did not address the main problem reported in relation to predictive performance that discussed in relation to the results in Table I, namely that the predictions of the M8 algorithms for large-magnitude earthquakes were far beyond real events. To temper these predictions and bring them closer to reality, we adapted the “classical” model by adding a pattern-based prediction algorithm, which is more sensitive to trends in the evolution of stochastic parameters.

5. Experimental results

The results presented in this section are a synthesis of countless simulations of different scenarios of combining three data sets: the real data selection interval (grouped into 5-year sub-intervals), the threshold values for the magnitude of the main shock, respectively for aftershocks filtering, and forecast interval (also achieved by concatenating 5-year subintervals). Various variants were developed for these combinations of basic scenarios, taking into account the

arrangement of the centers of the investigation circles and the radius of these circles, the membership of these centers in clusters of points with intense confirmed seismic activity and the classification of TIP intervals.
In the simulations, two prediction algorithms were used - the classic M8 algorithm (although the term “classic” is not really appropriate, as the algorithm was improved by adding adjustable coefficients - the so-called Levinson - Durbin coefficients [13] - to be as close as possible to the newer variants of the M8 algorithm (*eg*. [14]) and the improved algorithm, described in section 4, named by the acronym M8T (T from Theodora).

In Table III comparative results are passed on the number of earthquakes with magnitude $M \geq 5$ forecasted with the two mentioned algorithms, for 5 combinations of input data selection interval and forecast interval respectively. In all these scenarios, real data from a cluster of 7 points (centers of investigation circles) were used, which had at least one earthquake of magnitude at least 5 in the historical period of the input data.

Table III

PERFORMANCE OF COMPARED PREDICTION ALGORITHMS

Historical data range (actual input data)	*Forecast interval*	*Number of earthquakes with M≥5 (actual from the IFP catalog)*	*Number of predicted earthquakes with classic M8 (M8LD)*	*Number of predicted earthquakes with improved M8 (M8T)*
1995- 2005	2005-2015	5	3	4
1995- 2005	2005-2020	8	5	6
1995-2010	2010-2015	2	1	1
1995-2010	2010-2020	5	3	4
1990-2000	2000-2020	10	5	7

In Fig. 5, Fig. 6, Fig. 7, Fig. 8 and Fig. 9 are presented plots confirming the results in Table III. In these figures the real data are represented in green, those calculated with the M8 algorithm adjusted with coefficients Levinson-Durbin (M8LD) are represented in red and those calculated with the improved algorithm (M8T) are represented in blue. Magnitude values were scaled to avoid overlapping curves and easier identification of points of interest. Only the N2 function which illustrates the earthquake rate of occurrence, expressed by the average annual number of earthquakes with a magnitude equal to or greater than the predetermined magnitude threshold was represented.

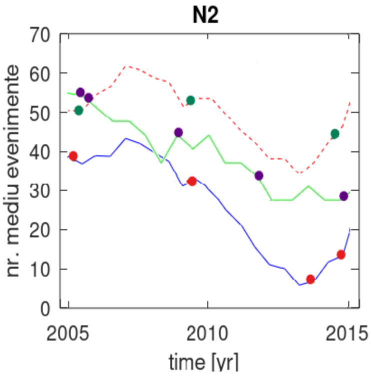


Fig. 5. ExEvs forecasted for interval 2005-2015 on historical data from interval 1995-2005.

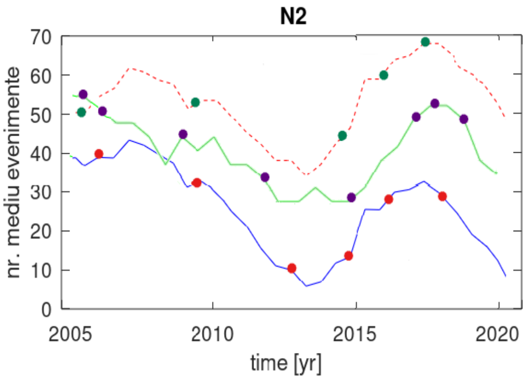


Fig. 6. ExEvs forecasted for interval 2010-2020 on historical data from interval 1995-2005.

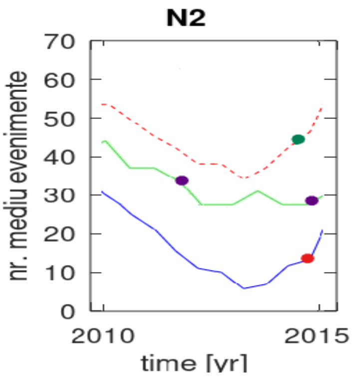


Fig. 7. ExEvs forecasted for interval 2010-2015 on historical data from interval 1995-2010.

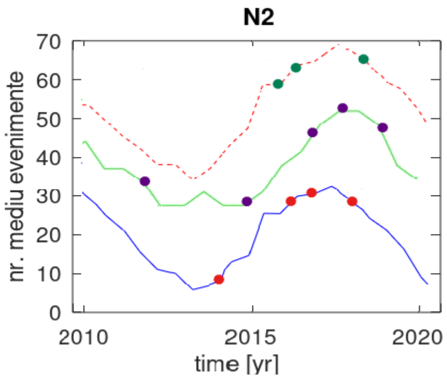


Fig. 8. ExEvs forecasted for interval 2005-2020 on historical data from interval 1995-2010.

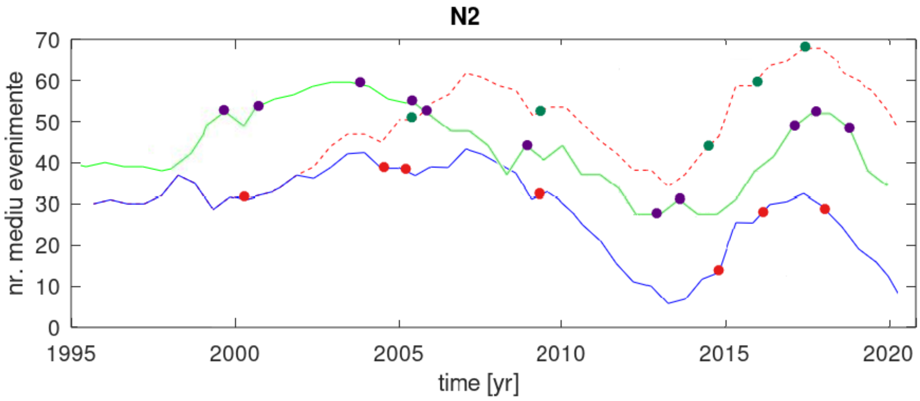


Fig.9. ExEvs forecasted for interval 2000-2020 on historical data from interval 1990-2000.

From the qualitative analysis of the plots from Fig.5, Fig.6, Fig.7, Fig.8 and Fig.9 some objective conclusions can be drawn:

1. The dynamics of the frequency of earthquakes is relatively correct determined by both prediction algorithms, the appearance of the predicted curves being almost similar (number of maxima and minima, increase and decrease gradients, plateau intervals) and quite close to the allure of the real events curve. The only obvious discrepancy is recorded for year 2014, where both prediction algorithms indicate a minimum, but a relative maximum appears in the curve of real events. More annoying is the fact that this relative maximum also corresponds to an earthquake with a magnitude above the threshold, which of course both prediction algorithms missed.

2. The distribution in time of events with magnitude above the predetermined threshold correctly predicted is similar for the two tested algorithms, but they detected fewer events than the real ones in all scenarios. The M8T algorithm has always found at least as many events with magnitude above the threshold as the M8LD (in fact the equality was in a single scenario, over a very narrow forecast period). We can say that at least in terms of error rate, M8T has superior performance.

3. Regarding the tendencies to reach an extreme event, it is found that both algorithms have gradients of increase, respectively decrease, very close in value, but that this value is higher (in the mode) than that of the gradients in the real curve, which it is for this reason much more flattened and presents numerous plateau intervals, almost non-existent in the curves of the predicted events. A strong anomaly is present in both forecasts in the period 2005-2008 (Fig. 6) in which the prediction algorithms present an increasing gradient, while the curve of real events has a decreasing gradient. We did not find a convincing explanation for the occurrence of this anomaly, but it seems to reflect the absence of historical data over at least twice the forecast interval.

4. There is no clear difference between scenarios that use the same interval for historical data, but perform the forecast at different intervals, which shows that the algorithms have a robust behavior. Instead, it is obvious that if different historical data ranges are used for forecasts over the same range, the error rate is lower when the historical data range is longer.

5. On the diagram in Fig. 9, which illustrates a forecast over the longest period (20 years), states (except for the anomaly already commented on for the period 2005-2008) that it would be possible to identify repeatable patterns. Of course, it is too little to discuss characteristics of self-similarity, but it is enough to signal the need for in-depth research, using a much larger volume of historical data and also a much longer forecast period (in our opinion at least 50 years).

4. Conclusions

The main challenge we sought to address by developing this article was to modernize and improve the well-known and most applied earthquake seismic prediction algorithm (M8) so as to increase the level of confidence in forecasting extreme events that could occur in Vrancea region. The proposed model has additional parameters that offer adaptation to the geographical data of the area, avoiding sudden gradient changes (which occur in the classic model, but are rarely found in the dynamics of real events) and. especially, reducing the error rate by avoiding false alarms as much as possible (it is known that the original M8 algorithm tends to estimate earthquakes with a magnitude higher than the one actually recorded). We consider that the paper, offers a rich study material for anyone interested in the long-term forecast of the most feared seismic activity in Romania. Moreover, from scientific point of view, the idea that the curve of the dynamics of earthquakes in the area could have self-similar patterns, which would suggest the existence of repetitive cycles, is the most important challenge for future research.

R E F E R E N C E S

[1]. *N. Chen, A. J. Majda,* "Predicting observed and hidden extreme events in complex nonlinear dynamical systems with partial observations and short training time series", Chaos, 30, pp. 033101:1-39, 2020

[2]. *A. J. Majda, M. Moore, D. Qi,* "Statistical dynamical model to predict extreme events and anomalous features in shallow water waves with abrupt depth change," Proc. Natl. Acad. Sci. 116, pp. 3982–3987, 2019.

[3]. *D. Qi, A. J. Majda*, "Predicting extreme events for passive scalar turbulence in two-layer baroclinic flows through reduced-order stochastic models", Commun. Math. Sci., 16, pp. 17–51, 2018.

[4]. *T. Sapsis,* "New perspectives for the prediction and statistical quantification of extreme events in high-dimensional dynamical systems", Philosophical Transactions of the Royal Society A., 376, pp. 20170133:1-18, 2018

[5]. *M. Farazmand, T. Sapsis,* "Extreme Events: Mechanisms and Prediction", ASME Appl. Mech. Rev., 71(5), pp. 050801:1-25, 2018.

[6]. *Z. K. Gao, M. Small, J. Kurths*, "Complex network analysis of time series", Europhys. Lett., 116, p.50001, 2016.

[7]. *B. L. Mediouni*, "Modeling and Analysis of Stochastic Real-Time Systems", PhD Thesis, Université Grenoble Alpes, 2019

[8]. *V. Kossobokov*, "User Manual for M8, in: Algorithms for Earthquake Statistics and Prediction", edited by: Healy, J. H., Keilis-Borok, V. I., and Lee, W. H., IASPEI Software Library, pp.167–222, 1997

[9]. *V. Kossobokov, et al*., "Stabilizing intermediate-term middle-range earthquake predictions", J. Seismol. Earthq. Eng., 8, pp. 11–19, 2002.

[10]. *V. Kossobokov, P. Shebalin,* "Earthquake Prediction, in: Nonlinear Dynamics of the Lithosphere and Earthquake Prediction", edited by: Keilis-Borok, V. I. and Soloviev, A. A., Springer Verlag, pp. 141–207, 2003

[11]. *V. Kossobokov*, "Quantitative Earthquake Prediction on Global and Regional Scales", in: Recent Geodynamics, Georisk and Sustainable Development in the Black Sea to Caspian Sea Region, edited by: Ismail-Zadeh, A., American Institute of Physics Conf. Proceedings, 825, pp. 32–50, 2006.
[12]. *M. Ghil, et al.*, Extreme Events: Dynamics, Statistics and Prediction, Nonlinear Processes in Geophysics, 18, pp. 295–350, 2011
[13]. *V. Teja, et al.*, "Spectral analysis of seismic signals using Burg algorithm", International Journal of Pure and Applied Mathematics, Vol. 114, No. 10, pp. 163-17, 2017
[14]. *M. Mojarab, et al.*, "An application of earthquake prediction algorithm M8 in eastern Anatolia at the approach of the 2011 Van earthquake", Journal of Earth System Science, 2015
[15]. https://web.infp.ro/#/romplus
[16]. http://www.infp.ro/index.php?i=romplus – RomPlus Catalog
[17]. STAS 2923-63 (incd.ro) **-** Seismic intensity zoning map