

Forecasting of Forest Fires in Portugal Using Parallel Calculations and Machine Learning

Yaroslav Vyklyuk¹, Milan M. Radovanović², Volodymyr Pasichnyk³,
Nataliia Kunanets³, Sydor Petro¹

¹Bukovinian University, Chernivtsi, Ukraine

²Geographical Institute “Jovan Cvijić”, Serbian Academy of Sciences and Arts,
Belgrade, Serbia,

³Information Systems and Networks Department, Lviv Polytechnic National University,
Lviv, Ukraine

Abstract. Forest fires that occurred in Portugal on June 18, 2017 caused several dozens of human casualties. The cause of their emergence, as well as many others that occurred in Western Europe at the same time, remained unknown. The heliocentric hypothesis has indirectly been tested, according to which charged particles are a possible cause of forest fires. We must point out that it was not possible to verify whether in this specific case the particles by reaching the ground and burning the plant mass create initial phase of the formation of flame. Therefore, we have tried to determine whether during the critical period, i.e. on June 15–19, there was a certain statistical connection between certain parameters of the solar wind (SW) and meteorological elements. Based on the hourly values of the charged particles flow, a correlation analysis was performed with hourly values of individual meteorological elements including time lag at Monte Real station. The application of the adaptive neuro-fuzzy inference systems has shown that there is a high degree of connection between the flow of protons and the analyzed meteorological elements in Portugal. However, further verification of this hypothesis requires further laboratory testing.

Keywords: forest fires, heliocentric hypothesis, ANFIS models, Portugal

1 Introduction

Forest fires that occurred on June 18, 2017 in the central part of Portugal are among the most endangered in the history of this country. About 60 forest fires were reported during the night of June 17/18, 2017. The number of victims was 64, including six fire-fighters. Many died in their cars or near the vehicles when they tried to escape from fire. About 200 people were injured. On June 20, a plane with two crew members who participated in the extinguishing of the fire collapsed. On that day, the fires were localized. More than 45,000 hectares of forest burnt. The question arose whether it was possible to predict these natural disasters, and therefore take measures to eliminate them and minimize human and material casualties.

Severe weather conditions were mentioned as a possible reason. It is possible that the fire was caused by a thunderstorm, as the investigators found a tree that was hit by a “thunderstorm without rain,” the Portuguese media reported, referring to police sources. In addition, some officials expressed the view that fires were deliberately set.

Gomes and Radovanović [2] presented a number of critical views on the so-called generally accepted attitudes regarding explanation of the initial phase of the flame. In short, there was a suspicion of anthropogenic activity being possibly responsible in 95% of cases, as it was accepted not only in the media, but also in many scientific circles. And first of all, because it is hard to believe that, for example, over 30,000 fires have been reported over the years, and the area of the burnt vegetation is over 350,000 ha.

Milenković et al. [5] found the connection between the number of forest fires in Portugal and the Atlantic multidecadal oscillation. Considering absolute values, there some other so-called potential candidates look suspicious, such as lightning or high air temperature. Thunderstorms are followed by rain almost as a rule, and when it comes to lightning, they strike without rain. As for air temperature, it has been determined that a minimum of 300 °C is required to show the initial phase of the flame [12]. This temperature has never been measured on the ground by far (including desert areas), not to mention the air temperature.

As far as we know, Gomes and Radovanović [1] argued for the first time that the processes on the Sun could represent a potential explanation for the occurrence of forest fires. In the meantime, more research was published on this subject: Radovanović et al. [10], Radovanović, Vyklyuk et al. [8], Radovanović, Pavlović et al. [9], Radovanović, Vyklyuk et al. [7], etc. In these works, the effectiveness of adaptive neuro-fuzzy inference systems (ANFIS) and artificial neural networks (ANN) was proved. In this paper, therefore, we tried to re-examine whether there is justification for this kind of approach. According to the mentioned hypothesis, it is necessary to have a coronary hole and/or energy region in the geoeffective position on the Sun before the occurrence of a fire. On June 15, coronary holes CH807 and CH808, as well as energy region 12663 were in the geoeffective position. The characteristic of this region was beta-gamma. The maximum speed of the SW at La Grange Point was 594 km/s¹. On that day, on June 16, according to the same source, the Ap index was around 3–48.

2 Data and Methods

Bearing in mind that it was not possible to directly record the possible propagation of particles to the ground, as a potential reason that causes initial phase of the flame in the burning plant mass, we decided to test the heliocentric hypothesis indirectly. The following hourly averaged real-time data were used as input parameters: differential electron (energy ranges 38–53 and 175–315 keV)—*E1*, *E2*, proton flux (energy ranges 47–68, 115–195, 310–580, 795–1193 and 1060–1900 keV) —*P1–P4*, and integral proton Flux—*I1*, *I2*; hourly averaged real-time bulk parameters of the solar wind plasma proton density (p/cc) —*W1*, bulk speed (km/s)—*W2*, and ion temperature (degrees K)

¹ <http://www.solen.info/solar/indices.html>

² ftp://ftp.swpc.noaa.gov/pub/lists/ace2/201706_ace_epam_1h.txt

$W3^3$. ACE Satellite—Solar Wind Electron Proton Alpha Monitor is located at La Grange point so that it measures the data in real time that come from the Sun to our planet. Hourly meteorological data relating to Monte Real Station have been used as an output (Latitude: 39° 49' 52" N, Longitude: 8° 53' 14" W). This station is located in the military air-base near Leiria. The data include air temperature (°C), humidity (%), and air pressure (hPa). All the data used in the paper refer to June 15–19, 2017. This station was selected because it is located near the fire-affected area and the data are available on the Internet (Table 1).

The goal of calculations was to investigate the functional dependencies between the characteristics of the SW and air temperature T , humidity H , and pressure P . The measurement step was 1 hour.

The solution of this problem consists of several stages.

Table 1. Tested input fields and output fields and correlation between them

Input fields		Correlation (R)		
		T	H	P
Differential Flux particles/cm2-s-ster-MeV, electrons				
$E1$	38–53	–0.11	0.14	0.16
$E2$	175–315	0.06	–0.02	–0.17
Differential Flux particles/cm2-s-ster-MeV, protons				
$P1$	47–68	0.01	–0.03	0.11
$P2$	115–195	0.14	–0.12	–0.07
$P3$	310–580	0.15	–0.13	–0.17
$P4$	795–1193	0.11	–0.08	–0.30
$P5$	1060–1900	–0.10	0.06	0.22
Integral Proton Flux				
$I1$	> 10MeV	–0.57	0.50	0.85
$I2$	> 30 MeV	–0.55	0.48	0.85
Solar Wind				
$W1$	Proton Density (p/cc)	–0.11	0.16	0.41
$W2$	Bulk Speed (km/s)	0.34	–0.34	–0.47
$W3$	Ion Temperature (degrees K)	0.14	–0.12	–0.06

2.1 Preliminary Analysis

The feature of this dataset is the presence of missed data (gaps) with a maximum duration of 3 hours. The spline interpolation using not-a-knot end conditions was used to fill in these gaps. The interpolated value at a query point is based on a cubic interpolation of the values at neighboring grid points in each respective dimension [3].

Correlation analysis was performed to establish the presence of a linear connection between input and output fields. Calculations show that Pearson correlation coefficients (R , Table 1) are sufficiently small in all cases except $I1$ and $I2$. It means that any linear dependencies of these data are not observed. High values of R for $I1$ and $I2$ indicate the presence of strongly expressed nonlinear relationships. The presence of lagging

³ ftp://ftp.swpc.noaa.gov/pub/lists/ace2/201706_ace_swepam_1h.txt

(time) between input and output fields may be another reason for small correlation coefficients.

To establish the lag dependence, the transformation of the dataset was conducted. Output fields were fixed, after that the time series of each input field was shifted vertically downward by the number of rows equal to the lag studied. After that, the correlation coefficient between input and output fields was calculated. We investigated the lag from 0 to 5 hours.

Calculations show that the smallest R is observed for electron and proton flux (E and P). It means that these input fields do not impact output fields for all lags. The largest R is observed for $I1$ and $I2$ relating to air pressure. R grows weakly with increasing lag. It means that there are nonlinear inertial dependencies between these fields and output fields. These lags mean that it is possible to predict output fields for few hours forward. Similar situation was observed for fields $W1$, $W2$, $W3$.

For further research, an autocorrelation analysis should be carried out to reconcile the interconnection between the input fields. Calculations show strong linear relationship between $I1$ and $I2$ fields. It means that only one of them should be used in calculations.

2.2 Search of Best Models

As can be seen from lag correlation and autocorrelation analysis, the best models for all output fields must be dependent on integral proton flux and solar wind with lag 5:

$$T(H, P) = F((I1 \text{ or } I2)_5, W1_5, W2_5, W3_5), \quad (1)$$

where subscribe index 5 means lag 5.

We must know which of them ($I1$ or $I2$) is better. Therefore we tested models $T(H, P) = F(I1_5, W1_5, W2_5, W3_5)$ and $T(H, P) = F(I2_5, W1_5, W2_5, W3_5)$.

For checking this decision, models with all possible combinations of lags from 0 to 5 were tested ($6^4 = 1,296$ models). Models with $I1$ or $I2$ input fields were tested separately. Theoretical investigations [6] showed that electrons must have nonlinear impact on output fields. Therefore, similar calculations for models containing one of the fields $E1$ or $E2$ were carried out ($6^5 = 7,776$ models). In addition, models that take into account only differential flux of electrons and protons were tested ($6^7 = 279,936$ models). Linear regression analysis and ANFIS were used as models in this investigation. Two Gauss membership functions were created for each input field in ANFIS.

Since ANFIS is a Sugeno-type system, the output membership function type was selected as constant. Each ANFIS system was trained during 100 epochs, initial step size was 0.01, step size decrease rate was 0.9, step size increase rate was 1.1. Hybrid method was selected as optimization method for membership function parameter training. For learning process, the dataset was split into training and test sets in proportion 90/10. This method is a combination of least-squares estimation and back-propagation.

Taking into account that 894,240 models have been investigated, and all of them are independent of each other, the parallel calculation was used to solve this problem. It decreased time of calculation about 3.5 times. The longest calculation lasted about 60

hours. The total time consisted of about 200 hours (~8 days). Results of these calculations are presented in Table 2.

Table 2. Comparison of correlation coefficients of the best models with models with lag 5 input fields

Models	Linear		ANFIS		Number of models	Position of (1) model (linear/ANFIS)
	Lag 5	Best	Lag 5	Best		
Temperature						
$F(I1, W1, W2, W3)$	0.8199	0.8251	0.8529	0.8621	1296	19 16
$F(I2, W1, W2, W3)$	0.8113	0.8178	0.8483	0.8540	1296	11 11
$F(I1, W1, W2, W3, E1)$	0.8337	0.8431	0.8698	0.8839	7776	54 41
$F(I1, W1, W2, W3, E2)$	0.8239	0.8287	0.8843	0.9051	7776	65 359
$F(E1, E2, P1, P2, P3, P4, P5)$	0.3372	0.4241	0.3567	0.4733	279936	95,927 196,615
Humidity						
$F(I1, W1, W2, W3)$	0.7559	0.7680	0.8026	0.8247	1296	20 24
$F(I2, W1, W2, W3)$	0.7437	0.7597	0.7953	0.8115	1296	26 24
$F(I1, W1, W2, W3, E1)$	0.7705	0.7861	0.8216	0.8455	7776	46 123
$F(I1, W1, W2, W3, E2)$	0.7672	0.7831	0.8538	0.8910	7776	50 733
$F(E1, E2, P1, P2, P3, P4, P5)$	0.3553	0.4363	0.3689	0.4610	279936	56,329 130,930
Pressure						
$F(I1, W1, W2, W3)$	0.8985	0.9037	0.9483	0.9637	1296	178 247
$F(I2, W1, W2, W3)$	0.8767	0.8988	0.9338	0.9506	1296	932 512
$F(I1, W1, W2, W3, E1)$	0.8994	0.9055	0.9521	0.9679	7776	917 929
$F(I1, W1, W2, W3, E2)$	0.8991	0.9061	0.9576	0.9673	7776	1280 309
$F(E1, E2, P1, P2, P3, P4, P5)$	0.4090	0.5879	0.4338	0.6153	279936	256,184 276,420

As it can be seen from Table 2, correlation coefficient between real data and models' data was selected as a criterion of accuracy. First of all, it should be noted that all ANFIS models have higher correlation coefficient than linear ones. It is clearly seen that models based on differential flux of electrons and protons have the smallest R . This means that they are not the main factors of influence on output fields.

Comparing models containing $I1$ and $I2$ factors allows us to conclude that factor $I1$ makes it possible to better describe output fields. It is true for linear and ANFIS models.

As the calculations have shown, taking into account the factor describing differential flux of electrons allowed slight increase of the correlation coefficients for linear and ANFIS models. It should be noted that the influence of factors $E1$ and $E2$ is approximately the same. Therefore, factor $E2$ was chosen for further calculations.

So, in the next stage we investigated the most accurate models:

$$T(H, P) = F(I1_1, W1_1, W2_1, W3_1, E2_1), \quad (2)$$

where subscript 1 stands for lag 1.

As it can be clearly seen from Table 2, there is a large number of models that are more accurate than (1) (columns 2–3 and 4–5). As the last column of the table shows, models (2) occupy by far not the first places among exact models. Therefore, the classical approach to the definition of exact models (1), described in the lag analysis stage, is not suitable for this class of problems. Figure 1 represents the distribution of correlation coefficients for all models (2). As it can be seen from Fig. 1, there are lots of models that can make forecasts of output fields with high level of accuracy. It can be the base for the creation of multimodels expert system for forecasting crisis events.

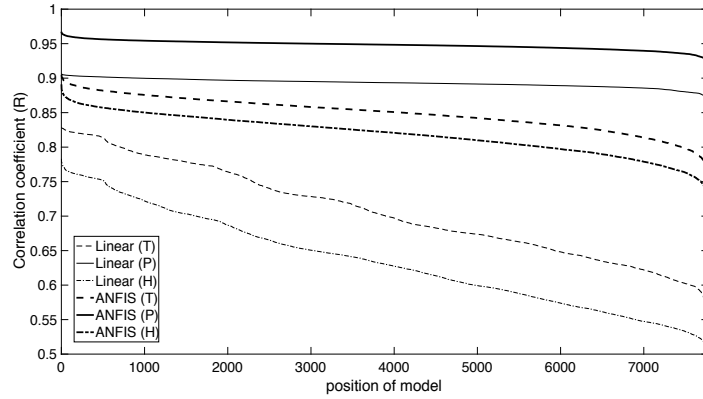


Fig. 1. Distribution of correlation coefficients of all possible models (2)

2.3 Accuracy Analysis

To confirm this conclusion, the adequacy of the three most accurate models was tested:

- linear with higher correlation coefficient;
- ANFIS with higher correlation coefficient;
- model with higher total ($R_{Linear} + R_{ANFIS}$) correlation coefficient.

Information about input fields of these models is presented in Table 3.

As it can be seen from Table 3, the best linear model for temperature is:

$$T_1 = F(I1_1, W1_5, W2_0, W3_0, E2_5).$$

Table 3. Lags (2) of the best models for forecasting T, H, P

Model	Best for:	II_1	WI_1	$W2_1$	$W3_1$	$E2_1$	R
Temperature							
T_1	Linear	1	5	0	0	5	0.8287
T_2	ANFIS	0	5	5	3	5	0.9051
T_3	Linear+ANFIS	5	5	2	3	5	1.7215
Humidity							
H_1	Linear	1	5	0	0	5	0.7831
H_2	ANFIS	5	4	4	2	5	0.8910
H_3	Linear+ANFIS	4	5	5	3	4	1.6392
Pressure							
P_1	Linear	4	3	3	3	4	0.9061
P_2	ANFIS	5	0	5	3	1	0.9673
P_3	Linear+ANFIS	5	0	5	3	0	1.8684

The best ANFIS model is as follows:

$$T_2 = F(II_0, WI_5, W2_5, W3_3, E2_5).$$

The model with higher total correlation coefficient is as follows:

$$T_3 = F(II_5, WI_5, W2_2, W3_3, E2_5).$$

As it can be seen, lags of input fields of these models are different.

In addition, lags of the input fields of these models are sometimes much smaller than in (1). On the other hand, it confirms that multimodels approach can make forecasting on different time periods from 0–5 hours.

According to the data for the lags from Table 3, nine linear and nine ANFIS models were constructed.

After training each ANFIS model, we obtained a set of membership functions, rules, fuzzification, and defuzzification methods etc. [4]. The attributes of obtained membership functions of input factors are presented in Table 4. As we mentioned above, each input factor consists of 2 membership functions.

As can be seen from Table 4, during training, only parameters of II membership functions were changed. It confirms that this field is most significant in these models.

For testing accuracy of these models, results of model forecasting were compared with real data and correlation coefficient was calculated (Table 5).

As can be seen from Table 5, all models have high R . All ANFIS models have higher correlation coefficient than linear ones. It confirms that ANFIS models are more accurate and take into account nonlinear effects. For visual comparison of results, predicted values obtained by the models in comparison with real data are presented in Fig. 2.

2.4 Adequacy Analysis

As can be seen from Fig. 2, ANFIS models better describe and predict fluctuation in amplitude. It is clearly seen for humidity models. Despite the high coefficient of ANFIS

models, linear models also accurately describe explored output fields. Therefore, an adequacy analysis and sensitivity analysis is required to select the correct type of model.

Table 4. Parameters of membership functions for ANFIS models

Model	[σ c]				
	I_1	WI_1	$W2_1$	$W3_1$	$E2_1$
Temperature					
T_1	[0.05 1.97] [0.03 2.10]	[11.0 1.0] [11.0 27.0]	[111 362] [111624]	[16706130599] [167061423999]	[63953-99999] [6395350600]
T_2	[0.02 1.95] [0.04 2.09]				
T_3	[0.041.95] [0.05 2.09]				
Humidity					
H_1	[0.06 1.98] [0.04 2.11]	[11.0 1.0] [11.0 27.0]	[111 362] [111624]	[16706130599] [167061423999]	[63953-99999] [6395350600]
H_2	[0.04 1.96] [0.05 2.09]				
H_3	[0.04 1.96] [0.053 2.097]				
Pressure					
P_1	[0.03 1.96] [0.03 2.11]	[11.0 1.0] [11.0 27.0]	[111 362] [111624]	[16706130599] [167061423999]	[63953-99999] [6395350600]
P_2	[0.03 1.97] [0.04 2.10]				
P_3	[0.04 1.96] [0.04 2.10]				

Table 5. Pierson correlation coefficients of models from Table 4

Model	Linear	ANFIS
T_1	0.8287	0.8714
T_2	0.7697	0.9051
T_3	0.8204	0.9012
H_1	0.7831	0.8374
H_2	0.7076	0.8910
H_3	0.7629	0.8763
P_1	0.9061	0.9581
P_2	0.9010	0.9673
P_3	0.9032	0.9652

For this purpose, for each model, the following calculations were carried out:

1. For each row of the training set, each value of the input parameter of turning was changed by 10%.
2. A relative change in the output field on change of separate input field was calculated.
3. Data was averaged on all records.

Results of these calculations are presented in Table 6.

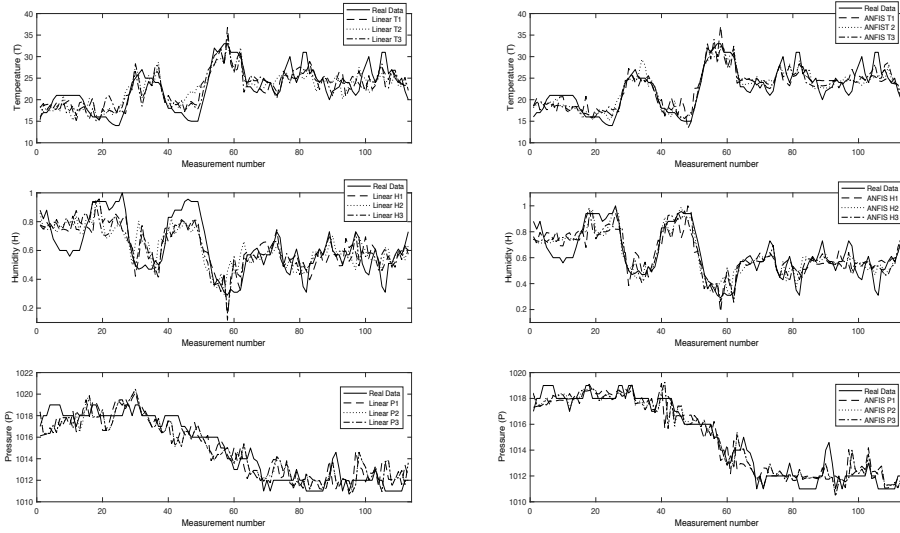


Fig. 2. Comparison of predicted values obtained by the linear and ANFIS models from Table 5 in comparison with the real data

Table 6. Sensitivity analysis of models from Table 2

Model	Best for:	II_1	WI_1	$W2_1$	$W3_1$	$E2_1$
Temperature						
T_1	Linear	-82%	2%	3%	-1%	0%
	ANFIS	-39%	0%	3%	0%	0%
T_2	Linear	-80%	1%	2%	-1%	0%
	ANFIS	-30%	3%	31%	-1%	0%
T_3	Linear	-91%	1%	0%	0%	0%
	ANFIS	-57%	1%	11%	1%	0%
Humidity						
H_1	Linear	96%	-3%	-7%	3%	0%
	ANFIS	72%	-1%	-15%	1%	0%
H_2	Linear	116%	-2%	3%	0%	0%
	ANFIS	91%	1%	-16%	-1%	0%
H_3	Linear	121%	-2%	1%	0%	0%
	ANFIS	58%	-2%	-18%	2%	0%
Pressure						
P_1	Linear	1.29%	0.00%	-0.05%	0.01%	0.00%
	ANFIS	0.36%	0.00%	-0.07%	0.01%	0.00%
P_2	Linear	1.30%	0.00%	-0.04%	0.01%	0.00%
	ANFIS	0.39%	0.00%	-0.05%	0.02%	0.00%
P_3	Linear	1.30%	0.00%	-0.04%	0.01%	0.00%
	ANFIS	0.43%	-0.01%	-0.07%	0.01%	0.00%

As can be seen from Table 6, -82% for linear $T1$ means that on average if factor $I1$ increases by 10% , the temperature will decrease by 82% after 1 hour. Similarly, the same increase in $I1$ will lead to decreasing temperature after 5 hours by 57% according to ANFIS model. As can be seen, the most significant factors are $I1$ and $W2$. These results have also confirmed that electrons do not impact investigated output fields. The results show that increasing $I1$ will lead to decreasing temperature and increasing humidity. In contrast, increasing $W2$ will lead to increasing temperature and decreasing humidity. It is clearly seen that input factors have weak impact on pressure despite the highest correlation coefficient of models.

3 Conclusions

Forest fires that occurred on June 18 and 19, 2017 in Portugal are among the most catastrophic ones in the history of the country. As in many other cases, the cause of their emergence has remained unknown. Relying on recent results, we have tried to test the heliocentric hypothesis of the occurrence of forest fires in this case. ACE satellite registered a sudden inflow of temperature, speed, and density of the SW particles a couple of days before the formation of fires. The basic starting point was that if there is any connection between the process on the Sun and forest fires, then during critical days the meteorological parameters would have to “react” to some extent to certain parameters of the SW. In that sense, we have tried to determine whether there is any statistical connection between the flow of protons and electrons in some energy ranges on the one hand and the air temperature, relative humidity, and air pressure in Monte Real, on the other one. Calculations included hourly values, but with a time lag shift from 0 to 5 hours in the period June 15–19, 2017. The largest R is observed for ($I1$) proton flux > 10 MeV and ($I2$) proton flux > 30 MeV (0.89 and 0.86 respectively) relative to air pressure.

Linear regression analysis and ANFIS were used as models in this investigation. Taking into account that 894,240 models have been investigated, and all of them are independent of each other, parallel calculations were used to solve this problem. As it can be seen from Fig. 1, there are lots of models that can make forecasts of output fields with high level of accuracy. According to the data for the lags from Table 3, nine linear and nine ANFIS models were constructed. As it can be seen from Table 5, ANFIS models are more accurate and take into account nonlinear effects.

Obtained results indicate the need of further improvement of the presented methods for the purpose of creation of scientifically based Web-oriented multimodels expert system for making forecasting of crisis events in different time periods of 0–5 hours. Especially if we keep in mind that, depending on the repeatability of certain processes in the Sun, we can expect more or less similar weather and environmental conditions in certain locations on Earth [11].

References

1. Gomes JFP, Radovanovic M. Solar activity as a possible cause of large forest fires a case study: Analysis of the Portuguese forest fires. *Science of the total environment* 2008;394(1):197–205.
2. Gomes JFP, Radovanovic M. *Solar Activity and Forest Fires*. Nova Science Publishers, New York, 2009:1-139, ISBN: 978-1-60741-002-7.
3. Hazewinkel M. *Encyclopaedia of Mathematics: Monge - Ampère Equation - Rings and Algebras*, Springer, 2013;929P, ISBN: 978-0-7923-2976-3.
4. Jang J.-SR ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems Man and Cybernetics*; (1993);23(3):665-685.
5. Milenković M, Yamashkin AA, Ducić V, Babić V, Govedar Z. Forest fires in Portugal - the connection with the atlantic multidecadal oscillation (AMO). *Journal of the Geographical Institute Jovan Cvijic SASA* 2017;67(1):27-35.
6. Radovanović M. Forest fires in Europe from July 22nd to 25th 2009. *Archives of Biological Sciences*, 2010;62(2):419-424.
7. Radovanović M, Vyklyuk Y, Jovanović A, Vuković D, Milenković M, Stevančević M, Matsiuk N. Examination of the correlations between forest fires and solar activity using Hurst index. *Journal of the Geographical Institute Jovan Cvijic SASA* 2013;63(3):23-32.
8. Radovanović MM, Vyklyuk Y, Malinović Miličević BS, Jakovljević MD, Pecelj RM. Modelling of forest fires time evolution in the USA on the basis of long term variations and dynamics of the temperature of the solar wind protons. *Thermal Science*, 2014;19(Suppl. 2):S437-S444.
9. Radovanović MM, Pavlović TM, Stanojević GB, Milanović MM, Pavlović MA, Radivojević AR. The influence of solar activities an occurrence of the forest fires in south Europe. *Thermal Science*, 2014;19(2):435-446.
10. Radovanović MM, Vyklyuk Y, Milenković M, Vuković BD, Matsiuk N. Application of ANFIS models for prediction of forest fires in the USA on the basis of solar activity. *Thermal Science*, 2015;19(5):1649-1661.
11. Todorović N, Vujović D. Effect of solar activity on the repetitiveness of some meteorological phenomena. *Advances in Space Research* 2014;54(11):2430–2440.
12. Viegas DX. A mathematical model for forest fires blowup. *Combust SciTechnol*, 2005;177:27–51.