Application of ANFIS Models for Prediction of Forest Fires in the U.S.A. – Preliminary Results

ARTICLE · JANUARY 2015

Yaroslav Vyklyuk
Bukovinian University

Milan Radovanovic
Serbian Academy of Sciences and Arts

Application of ANFIS Models for Prediction of Forest Fires in the U.S.A. - Preliminary Results

YAROSLAV VYKLYUK[a], MILAN M. RADOVANOVIĆ,[b] MILAN MILENKOVIĆ[b] ANA JOVANOVIĆ[b], DARKO VUKOVIĆ[b], MILAN STEVANČEVIĆ[c], NATALIYA MATSIUK[d]

[a] Bukovynian University, Chernivtsi, Ukraine
[b] Geographical Institute "Jovan Cvijić", Serbian Academy of Sciences and Arts - SASA, Serbia
[c] Ex Federal Ministry of Telecommunication of Yugoslavia, Serbia
[d] Bukovynian State Finance and Economics University, Chernivtsi, Ukraine
vyklyuk@ukr.net   http://www.bukuniver.edu.ua

Abstract: - In this research we search for a functional dependence between the occurrence of forest fires in the U.S.A. and the factors which characterize the solar activity. For this purpose we used several methods (R/S analysis, Hurst index) for establishing potential links between the influx of some parameters from the sun and the occurrence of forest fires with lag of several days. We find evidence for a connection and developed a prognostic scenario based on the ANFIS technique. This scenario allows the predicting of up to 93% of forest fires in some cases.

Key-Words: - forest fires, heliocentric hypothesis, ANFIS models, U.S.A.

1 Introduction

Forest fires are an important ecological problem, particularly because adequate prevention measures do not exist. In essence, the ability to prevent the spread of the fire is based on reactions to the occurrence of fire. Indeed, there is no consensus on the origin of many forest fires. Analysing the FAO data, Radovanović & Gomes [24] concluded that in Europe for the period 1999-2001 there were 42.7% of the cases for which the causes were not established.

According to Nikolov [21], in average, 58.8% of the total number of forest fires in the countries of the Balkan Peninsula for the period 1988-2004 have human origin, 3.3% have natural origin and 37.9 % have unknown origin. The largest percentage of forest fires with human origin was recorded in Croatia (75.3%) and the smallest percentage in Bulgaria (30.4%). On the other hand, Bulgaria has the largest percentage of unknown causes (67.9%).

The total number of forest fires was 25 221 in Portugal in 2011, whereof 40% of the cases was treated as fires with unknown cause. In Germany, out of 888 fires in the same year, the causes of 48% of the fires were unknown. On the other hand, in Hungary 95% of fires is human-induced. Overall, the investigations against forest fire crimes in 2011 carried out by the Territorial Garrison of Italian Forest Corps, resulted in the reporting of 455 people to the Court of Justice, including 9 taken under arrest or under custody measures for fire arson. The total number of forest fires in 2011 was 8 181 in Italy [30].

The sources from which the data were downloaded for this study (the number of fires in the U.S.) indicate that all fires occurred either by human activity (85.5%) or a lightning strike (14.5%). Citing the results of the National Interagency Fire Centre[28], pointed out the opposite data: "for 1997 for North America three quarters of the land burned – 76% was due to lightning." It is known that lightning can also be an important factor for the origin of the initial phase of the flame (http://www.predictiveservices.nifc.gov/intelligence/archive.htm). However, in this domain also there are strongly opposing opinions. It is obvious that the precipitation quantity in such situations defines whether the fire would spread or be extinguished, since lightning is mostly followed by precipitation [15]. It seems that the lack of more detailed studies on this theme does not offer the strong enough support to understand the question to what extent electric discharges participate in the initial phase of the fire phenomenon. As Hall points out "From 1990 to 1998, over 17 000 naturally ignited wildfires were observed in Arizona and New Mexico on US federal land during the fire season of April through October. Lightning strikes associated with these fires accounted for less than 0.35% of all recorded cloud-to-ground lightning strikes that occurred during the fire season during that time".

On the other hand, according to Cumming [6], in the period 1961-1993 in mixed forests of Alberta (Canada) 67.6% of fires were caused by thunder strike. Research related to relatively recent data have shown that the share of lightning as a fire cause is almost ½ of analyzed cases. "Across Canada, from 1991 to 2000, ~8000 forest fires occurred per year, 48% of which were ignited by lightning" [38]. Sannikov *et al*. [29] have suggested that in western Siberia almost all fires are caused by thunder strike. Beside different temporal intervals of data processing, the ranges of impacts of lightening on forest burning are therefore at least contradictory.

There is a suggested link between the relatively high air temperatures and the occurrence of fire, but it also does not also give an adequate explanation. It is well known that a minimum of 300°C is necessary for the mentioned initial phase [35]. It is not necessary to point out that such a high air temperature has never been measured on the Earth by standard meteorological measures, even when it is soil temperature about.

Based on these results, we could see that there are confusing information about the extent to which a man can be the cause of the fire. We have seen that these percentages range from 95% (in the case of Hungary) to ~ 43% of the cases for which the cause has not been determined. Data on lightning strikes as the cause are also contradictory. The percentages range from 0.35% in the case of Arizona and New Mexico to almost 100% in western Siberia. The suspicion that in the usual way we can explain most of the fires is also presented by Guyette *et al*. [11]: "Furthermore, we have found that, at relatively coarse spatial and temporal scales (that is, regions and centuries), fire frequency variability caused by local factors such as vegetation type, topography, grazing, and human ignitions becomes less important".

With this in mind the results of Gomes & Radovanovic [9] put forward the "heliocentric hypothesis" that those forest fires without established causes are caused by a burning plant mass under the action of charged particles that come to us from the sun. The authors suggested that the occurrence of the fire should be preceded by, and correlated with, a sudden influx of the mentioned particles toward our planet. Without going into the feasibility of the proposed model of the propagation of particles to the ground, in this study we tested the heliocentric hypothesis using the R/S method, and the Hurst index, to see if any relevant correlations could be found to support or disprove this hypothesis. We find evidence for correlations between the sudden influx of charged particles and the occurrence of large forest fires with a delay of one to four days. Based on these results, we propose a prognostic ANFIS model which may allow a more precise prediction of forest fires

2 Problem Formulation

The decision to test the hypothesis especially in the case of the U.S.A. was made due to the availability of data on fires in a relatively large area and on a daily basis. The study comprised the period from May to October in each of the years 2004 – 2007. Data on forest fires are retrieved from http://www.predictiveservices.nifc.gov/intelligence/archive.htm. Data on the number of new small fires were used (F^{small}), as well as the new large fires (F^{large}). According to this source large (significant) fires are those that exceed 300 acres in grass and brush fuels and 100 acres in timber fuels. Data on the flow of protons, electrons and solar flux are retrieved from: http://www.swpc.noaa.gov/ftpmenu/warehouse.html . Data on solar wind speed (km/s) are retrieved from: http://umtof.umd.edu/pm/crn/, wherein the maximum values were used on a daily basis.

The number of large fires (F^{large}) and small fires (F^{small}) are taken to be the output variables for this research. The input parameters (as the indicators of the conditionally said solar activity) were selected as follows: the flow of >1 MeV protons (X_1), the flow of >10 MeV protons (X_2), the flow of >100 MeV protons (X_3), the flow of >0.6 MeV electrons (X_4), the flow of >2 MeV electrons (X_5), the 10.7 cm solar flux (X_6), and the solar wind speed (X_7). Emission from the Sun at centimetric (radio) wavelength is due primarily to coronal plasma trapped in the magnetic fields overlying active regions. It is an excellent indicator of overall solar activity levels (Tapping 1987). It is important to note that the data related to the solar activity are downloaded from the ACE satellite, which is always located between the Earth and the Sun. Previous research have indicated that in certain situations there is some causality between the abrupt influx of protons and/or electrons and the occurrence of fire on relatively large areas [8, 25, 26]. Bearing in mind that some areas may or may not be under the influence of both charged particles, (X_6) and (X_7) were selected as the general indicators of the solar activity. The studied period refers to the last phase of the solar cycle 23 [7]. Already in April of the following 2008, the solar activity was at a minimum, so that in this way we wanted to look at a situation for which we can say that is characterized by a continuous downward solar activity in the aforementioned cycle.

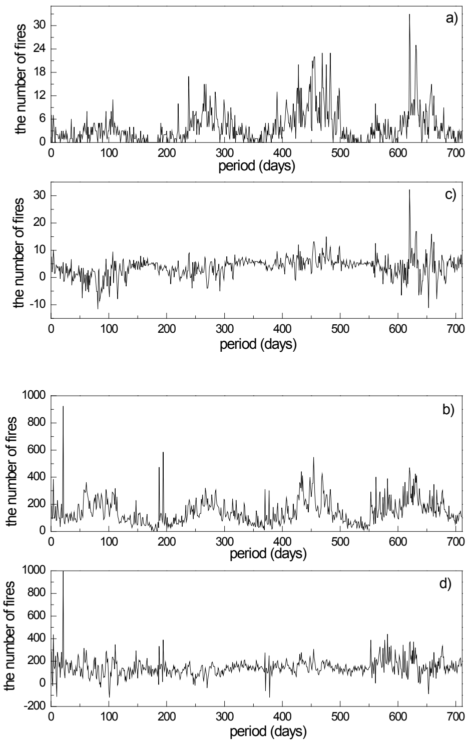

Figure 1 Number of large (a), (c) and small (b), (d) fires. Real data (a), (b), data with the seasonal component removed (c), (d)

As can be seen in Figure 1 (a, b), a cyclical occurrence of fires for F^{small} and F^{large} can be observed. Time series which have been corrected for this seasonal component are shown in parts (c) and (d). Here one can see that sudden outbreaks of fires are observed during the studied period. As can be seen in the same figure the amplitude of number of fires is not time dependent. Therefore a decomposition of the time series F^{small} and F^{large}, applying an additive model should be used. The additive model of the time series in our case is following:

$$F^{small(large)} = T^{small(large)} + S^{small(large)} + \tilde{F}^{small(large)},$$

where $T^{small(large)} = \left\{ t_j^{small(large)} \right\}_{j=\overline{1,n}}$ – trend component, i.e. long-term change course of the number of small (large) fires,

n - quantity of the observations, in our case 710 (days of the period from May to October in each of the years 2004 – 2007),

$S^{small(large)} = \left\{ s_j^{small(large)} \right\}_{j=\overline{1,n}}$ – seasonal component of the number of small (large) fires, which are associated with the temperature increase (decrease) within a year or with influence of holidaymakers on the appearance of forest fires,

$\tilde{F}^{small(large)} = \left\{ \tilde{f}_j^{small(large)} \right\}_{j=\overline{1,n}}$ – irregular component, which is related with some other factors, for example solar activity.

By removing seasonal and trend components from the initial time series we had prepared the time series to the research of solar activity influence on the appearance of small and large forest fires [5]. We used the classical method of seasonality indexes in order to filter out the seasonal component [4]. The technique of removing the seasonal component is following:

Step 1 The smoothing of the time series F^{small} and F^{large} using a simple moving average.

Step 2 The calculation of seasonal component $S^{small(large)}$ through doing two additional steps:

1) The finding the centred moving average. This step is necessary because of shifting of the obtained values of the moving average relatively to the real values of time series.
2) The calculation of the correctional coefficient which provides that the sum of all seasonal indexes equals zero, i.e. the seasonal effects for the entire annual cycle cancel each other for the additive model.

Values of the seasonal component, obtained in such a way, represent the ratio of the number of fires in a given day of the year to the average number of fires per year and thus receive either positive or negative values.

Step 3. The removing of the seasonal component from the original time series. In this way we had the time series of the number of the forest fires without seasonality impacts:

$$\hat{F}^{small(large)} = F^{small(large)} - S^{small(large)} = T^{small(large)} + \tilde{F}^{small(large)}$$

Step 4. The removing of the trend component from the $\hat{F}^{small(large)}$ with the least square method [13].

ISBN: 978-1-61804-276-7

So it was received only the occasional component $\tilde{F}^{small(large)}$, which we had used for the identification of the functional dependence between solar activity and the forest fires appearance.

To test this hypothesis, the correlation analysis was made between the factors X_i and the number of fires taking into account time delay (lag) between the onset of fires and solar activity. The results of this analysis are shown in the Table 1.

Table 1 Pair correlation coefficients between input (X_i, $i=\overline{1,7}$) and output ($\tilde{F}_L^{small(large)}$) variables with time lag $L=\overline{0,5}$

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$\tilde{F}_0^{large}$	-0.02	0.01	0.00	0.04	-0.02	-0.15	0.05
$\tilde{F}_1^{large}$	-0.04	-0.03	-0.01	0.02	-0.04	-0.16	0.04
$\tilde{F}_2^{large}$	-0.04	-0.02	-0.02	0.00	-0.02	-0.17	0.02
$\tilde{F}_3^{large}$	-0.04	-0.03	-0.03	-0.01	-0.02	-0.18	0.02
$\tilde{F}_4^{large}$	-0.05	-0.03	-0.03	-0.01	-0.02	-0.18	0.02
$\tilde{F}_5^{large}$	-0.02	-0.02	-0.02	0.01	-0.04	-0.19	0.02
$\tilde{F}_0^{small}$	-0.02	-0.01	-0.01	0.03	-0.02	0.09	-0.04
$\tilde{F}_1^{small}$	0.01	0.01	-0.01	0.00	-0.02	0.09	-0.03
$\tilde{F}_2^{small}$	-0.02	0.02	0.01	0.00	-0.01	0.07	-0.03
$\tilde{F}_3^{small}$	-0.04	-0.02	0.03	0.01	0.02	0.07	-0.02
$\tilde{F}_4^{small}$	-0.05	-0.04	0.01	0.01	0.04	0.07	-0.07
$\tilde{F}_5^{small}$	-0.03	-0.03	-0.02	0.00	0.03	0.05	-0.07

As it can be seen any correlation coefficient is not higher than 0.2. It means that there are no linear relationships between mentioned factors. Therefore it is necessary to apply methods of nonlinear analysis to test the hypothesis of a functional relationship between the onset of fires and solar activity.

3 R/S analysis

For determination of the degree of randomness for time series of input and output parameters, the R/S analysis was conducted [18, 19, 36]. The R/S analysis makes possible to determine whether the time series are stochastic ones or they have long-terminal correlation (long-terminal memory). To do this, the following equation was solved for each of the factors [23]:

$$R/S = c\cdot n^H, \quad (1)$$

where R/S – normalized magnitude, i.e. scope of partial sums of deviations of time series from its average, scaled by the standard deviation, c – constant, H – the Hurst index.

We had solved this equation for each of the input X_i and output time series $\tilde{F}^{large}$ and $\tilde{F}^{small}$. In this paper the way of solution for $\tilde{F}^{large}$ was shown. For other time series the process was the same.

At first, the initial time series $\tilde{F}^{large}$ with length n was transformed into a sequence $F=\left\{f_j\right\}_{j=\overline{1,n-1}}$, where $f_j=\ln\left(\frac{\tilde{f}_j^{large}}{\tilde{f}_{j-1}^{large}}\right)$. After that investigated time series were divided into A contiguous sub-periods with length l. Each sub-period has been marked as L^a, $a=\overline{1,A}$, and each element of the sub-period – $f_{(a-1)l+k}$, $k=\overline{1,l}$. Then for each sub-period the average meaning $\overline{f^a}=\frac{1}{l}\cdot\sum_{k=1}^{l} f_{(a-1)l+k}$ was determined and the scope of accumulated sums $R^a=\max_a\left(\left\{\sum_{k=1}^{l}\left(f_{(a-1)l+k}-\overline{f^a}\right)\right\}\right)-\min_a\left(\left\{\sum_{k=1}^{l}\left(f_{(a-1)l+k}-\overline{f^a}\right)\right\}\right)$ in terms of each sub-period was calculated.

Standard deviation S^a for each sub-period was defined as:

$$S^a=\sqrt{\frac{1}{l}\cdot\sum_{k=1}^{l}\left(f_{(a-1)l+k}-\overline{f^a}\right)^2}. \quad (2)$$

Each scope of accumulated sums R^a was normalized by dividing its corresponding standard deviation S^a. Then the average value $(R/S)_l$ for length l was defined as:

$$\left(R/S\right)_l=\frac{1}{A}\cdot\sum_{a=1}^{A}\frac{R^a}{S^a}. \quad (3)$$

Increasing the length of sub-periods l to integer $(n-1)/2$ and calculating for all of them $(R/S)_l$, the Hurst (H_l) index was determined by solving the simple least-squares linear regression equation using logarithmic transformation:

$$\log((R/S)_l) = \log(c) + H_l \cdot \log(l) \quad (4)$$

The value of the Hurst index can be interpreted as follows:

- If *H=0.5*, time series are stochastic ("white noise");
- If *0.5<H<1*, time series own property of persistence, i.e. time series has long-memory effect ("black noise"). This means that a decreasing time series, then, it is more probable that the time series will continue to decrease. In theory, the trend at a particular point in time affects the remainder of the time series;
- If *0<H<0.5*, time series own property of antipersistence, i. e. time series changes their trajectory faster than in case of stochastic process ("pink noise"). This means that a decreasing time series, then, it is more probable that the time series will show an increasing trend [23].

The usage of persistence or antipersistence property of the time series allows forecasting of the research process development in a relatively simple way on the base of its history.

On the basis of the Hurst exponent could be calculated another indicator – fractal dimension *D*:

$$D = 2 - H. \quad (5)$$

A fractal dimension shows us how a detail in a pattern (strictly speaking, a fractal pattern) changes with the scale at which it is measured. The results of these calculations are shown in the Table 2. As it can be seen from the table the Hurst index of $X_1 - X_5$ variables is closer to 0.5. It means these variables describe some stochastic processes.

Table 2 Results of *R/S* analysis for time series

Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7	$\tilde{F}^{sm}$	$\tilde{F}^{la}$
Hurst index	0.58	0.56	0.49	0.56	0.55	0.92	0.69	0.72	0.93

On the contrary, the Hurst index that is within 0.69 – 0.72 (for X_7, $\tilde{F}^{small}$) and 0.92 – 0.93 (for X_6, $\tilde{F}^{large}$) shows the dependence of the dynamics of these factors on their values in previous periods. The value of the Hurst index for X_6, X_7, $\tilde{F}^{small}$, $\tilde{F}^{large}$ mean that these processes are fractals and it cannot be used the classical linear statistics to research such time series. The similarity of the fractal dimensions (formula 5) for X_7 – $\tilde{F}^{small}$ and X_6 – $\tilde{F}^{large}$ means existence of the same rules of changing for such time series with scaling. That allows us to conclude that the dynamics of these time series is heavily depended on the same factors or on each other [28].

4 Adaptive neuro fuzzy inference system (ANFIS) models

ANFIS is a kind of neural network that is based on Takagi–Sugeno fuzzy inference system. Since it integrates both neural networks and fuzzy logic principles, it has potential to capture the benefits of both in a single framework. Its inference system corresponds to a set of fuzzy IF–THEN rules that have learning capability on large databases to approximate nonlinear functions [1, 14]. These methods are well examined not only in the natural sciences, but also in some social sciences [37]. Hence, ANFIS can be used to test our hypothesis.

In general form the problem is reduced to finding the dependence in the form: $M^{small(large)} : X_1 \times ... \times X_7 \to \tilde{F}^{small(large)}$. For this task tacking into account lag two training sets in the form of corteges were created:

$$Tr^{small} = \left\{\left\langle \bar{x}_{1,j-L}, ..., \bar{x}_{1,j-L}, \tilde{f}_j^{small} \right\rangle\right\}_{j=\overline{1,n}} \quad (6)$$

$$Tr^{large} = \left\{\left\langle \bar{x}_{1,j-L}, ..., \bar{x}_{1,j-L}, \tilde{f}_j^{large} \right\rangle\right\}_{j=\overline{1,n}} \quad (7)$$

where L – lag, $\bar{x}_{i,j}$ – normalized components of X_i time series, ($\bar{x}_{i,j} = \dfrac{x_{i,j} - \min(X_i)}{\max(X_i) - \min(X_i)}$).

The necessity of normalization of the all input parameters values is caused by significant difference between of the absolute max-min values of the component input vectors that can vary between one to five orders of magnitude (X_1 and X_6 for example). There are observed very large difference between absolute values of different input vectors too. For example $\max(X_4)-\max(X_6)\approx 10^{11}$, $\min(X_4)-\min(X_6)\approx 10^{8}$ (table 3.). Computer calculation without normalization can create big rounding mistakes, which completely neutralizes the objectivity of the model [2, 3, 10, 16, 20, 23, 31-34, 39].

comparative analysis of a coincidences number of small and large forest fires for real data and models (Figure 3). Also it should analyse false peaks and difference in the amplitudes.

Table 3 statistical characteristic of input and output parameters

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	$\widetilde{F}^{small}$	$\widetilde{F}^{large}$
Max	1100000000	74000000	500000	180000000000	9300000000	175	1005	996	32
Min	55000	11000	2100	230000000	650000	65	276	-121	-12
Average	8523106	404424	5487	21438042254	182332930	87	478	144	4
Average of $\overline{X}_i$	0.008	0.005	0.007	0.118	0.020	0.200	0.276	-	-

For determination of the lag between the events 6 ANFIS models for small and 6 ones for large forest fires for $L=\overline{0,5}$ ($\widetilde{f}^{small(large)}_{j\,ANFIS} = M^{small(large)}_{L}\left(\overline{x}_{1,j-L},...,\overline{x}_{7,j-L}\right)$) were created. For this all input parameters were presented as linguistic variables. Since the nonlinear dependence is present each of linguistic variables was identified by nonlinear Gauss terms. As test calculations shown the optimal count was 3 Gauss terms for each X_i (21 terms for each model). For 2 Gauss term was obtaining not objective models. If Gauss terms are bigger than 3, the numbers of empirical parameters exceed the volume of the training dataset. The Sugeno function of zero order was selected as a method of output fuzzy system.

The hybrid method that integrates back-propagation method with the least squares method was used as a method of learning. As a result the productive knowledge bases that contained 6 561 fuzzy rules were obtained.

4 Results

A correlation analysis between the time series $\widetilde{F}^{small(large)}$ and $M^{small(large)}_{L}$ was provided for the determination of the time lag between the onset of forest fires and solar activity (Fig. 2).

As it can be seen from the Figure 2, there are peaks for lag 1 and 4 in large fires case. It can means there is nearly 1 or 4 days and nights delay from the solar activity and large forest fires caused by it. Similar situation is observed for small fires. The maximum correlation is observed for lag 0 and 3 days. On the other hand differences between correlation coefficients are not big. Therefore hypothesis about lag dependences should be check by the

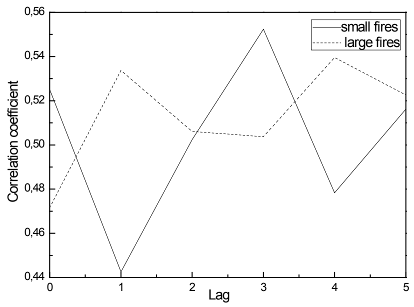

Figure 2 Dependence of correlation coefficient $\widetilde{F}^{small(large)}$ and $M^{small(large)}_{L}$ on lag L

As it can be seen from the previous figures, the models based on hybrid neural-networks give the possibility to explain the main solar activity of either large or small forest fires. Almost every peak on the model graph corresponds to the peak in the graph of real fires. This indicates the adequacy of models.

To check the accuracy the comparative analysis between a number of real fires flashes (peaks in the Fig.3, (a, d)) and flashes predicted by models (peaks on the fig.3. (b, c, e, f)) have been provided. The two cases were examined:

- The flash of the fires, predicted by the model, will occur in the same day;
- The flash of the fires, predicted by the model, will be observed ±1 day and night.

The results of this analysis are shown in the Table 4. As it can be seen from the Table 4, the model is able to predict a relatively good proportion of fires within one day.

ISBN: 978-1-61804-276-7

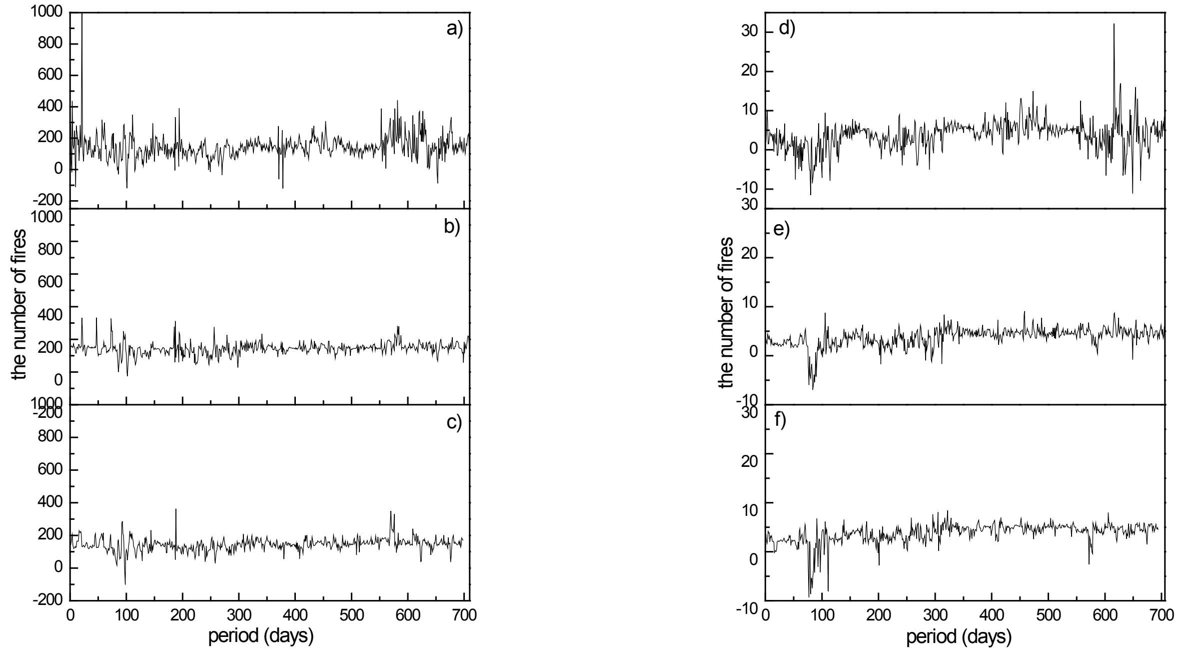

Figure 3 Comparison of modelling results with real data for a number of fires: small fires – a – real data, b – model data (lag=0), c – model data (lag=3); large fires – d – real data, e – model data (lag=1), f – model data (lag=4)

Table 4 Accuracy analysis of fires flashes prediction for ANFIS models

	Real fires flash es	Model fires flashe s	Explaine d by model fires flashes in the same day		Average difference in amplitude	False peaks in the same day		Cannot predict in the same day		Explained by model fires flashes with 3 days accuracy		Average difference in amplitude	False peak swith 3 days accuracy	
1	2	3	4		5	6		7		8		9	10	
Lag	Small fires													
0	207	189	73	35%	-4,6%	116	61%	48	23%	169	82%	-4,4%	20	11%
1	206	187	59	29%	-7,7%	128	68%	53	26%	170	83%	-3,4%	17	9%
2	204	197	78	38%	-5,1%	119	60%	44	22%	178	87%	-3,4%	19	10%
3	202	185	78	39%	-5,1%	107	58%	41	20%	170	84%	-2,1%	15	8%
4	202	180	65	32%	-1,2%	115	64%	42	21%	162	80%	-2,9%	18	10%
5	201	182	76	38%	6,7%	106	58%	38	19%	159	79%	-4,1%	23	13%
Lag	Large fires													
0	229	191	71	31%	11,6%	120	63%	55	24%	186	81%	-6,2%	5	3%
1	229	210	82	36%	2,1%	128	61%	60	26%	210	93%	2,3%	0	0%
2	226	194	75	33%	-1,9%	119	61%	52	23%	194	86%	12,8%	0	0%
3	225	189	66	29%	-2,4%	123	65%	58	26%	188	88%	13,7%	1	1%
4	223	193	69	31%	33,1%	124	64%	56	25%	177	79%	3,2%	16	8%
5	222	197	71	32%	13,3%	126	64%	56	25%	189	85%	22,3%	8	4%

ISBN: 978-1-61804-276-7

The biggest accuracy of the small fires prediction is observed for lag 3 and lag 2 (the same value is for lag 5), and for large fires – lag 1 and 2 (column 4). ANFIS models can predict around 39% of small fires and 36% of large fires with a same day accuracy. But this accuracy is much bigger when a delay of one day is allowed. For example 87% of small fires (for lag 2) and 93% of large fires (for lag 1) can be predicted by these models (column 8). Some peaks on the fig 3 of real fires, that are less than 21% (100%-column 8 - right), aren't explained by designed models.

It should be noted that the model with the accuracy of 1 day prediction predicts on average 60-65% cases of false flashes. These false predictions are observed for both large and small fires. More important information is how many real fires flashes are that model failed to predict. To test it we counted the number of cases where on the graph of real fires the peaks were observed and on the modeling graph at the same time value was below the average. As the calculations showed only 19-26% of the real small flashes cannot be predicted by the model. For large fires, this number is similar 23-26%.

However, if the prediction accuracy is 3 days, the number of false peaks is less than 13% for all calculations (column 10). There are also no flashes of real fires that cannot be predicted.

The predicted amplitude of peaks (a number of predicted fires flashes in a particular day in the U.S.) can also be interesting. As shown in the table (column 5), in the case of small fires the amplitude is usually smaller on average by 5% than the actual number of flashes. If the prediction has to be done for 3 days, then the error in the amplitude is reduced to -4 – -2% (column 9). This means the actual number of fires are on 2-5% higher than predicted by the model.

To determine the degree of response of fire to the change of specific factors, a sensitivity analysis was conducted. To do this the values of all input factors were fixed to their averages (Table 3) and the dependence of fire occurrence from sequential changes in each factor has been analyzed. Results of this analysis are presented in the Figure 4.

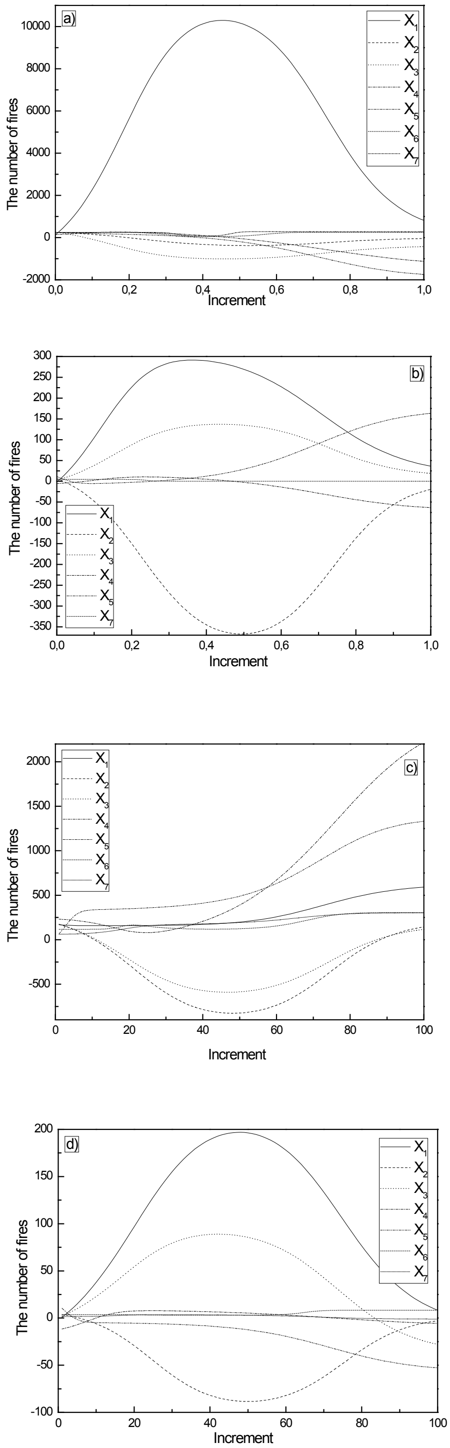

Figure 4 Sensitivity of number of small (a – lag 0, c – lag 3) and large (b – lag 1, d – lag 4) forest fires on X_i factors

As it can be seen from the previous figures the dependence of the fire onset on input factors is nonlinear. In particular, small fires are more sensitive to the X_1 factor (for lag 0 case). The dependence on the last factor has a quadratic form.

When the activity of the X_1 factor is increasing from average value 0.008 to 0.5 a number of fires flashes is quickly increasing. The increase of this factor from 0.5 to 1 leads to the decrease of flashes. It can be explained that never before the so big increase of this factor alone without change of another factors have been observed. The X_2- X_7 factors didn't impact on small fires flashes. Completely different situation is observed for lag 3 (Fig 4. c). The most powerful factor is X_5 for 0 – 0.1 diapason. From 0.1 to 0.5 this factor doesn't affect on fires flashes. After 0.5, the increase of this factor leads to a sharp increase of fires again. However, after 0.6 the most powerful factor becomes X_4.

A different situation is observed for large forest fires. The dependences for lag 1 and lag 4 are similar. As it can be seen from the Figure 4 b, d, the most important impact factors are: the dependence of a number of large fires on X_1 and X_3 is as analogue as small fires on X_1. The dependences on X_5 have exponential form. It means that a number of large fires is quickly increasing when X_5 is bigger than 0.5 (only for lag 1 case).

4 Conclusion

We find evidence for the presence of nonlinear relationships between the onset of the forest fires and the solar activity. This gives the possibility to use nonlinear methods of Soft Computing for discovery and analysis of functional dependences between them. It has been observed that between the solar activity and forest fires onsets there are lags that consist of 24 hours and several days for large fires case. This opens the possibility to predict when fires will occur and take steps to prevent them. Based on the developed ANFIS models, a prognosis for small fires with a delay of 2 days has an accuracy of 87%, while the accuracy is 93% for large fires with a delay of one day. This is true for the models forecast with 3 days accuracy. For the prognostication flashes in the same day, accuracy drops to 36-39% and increases the number of false peaks. Despite this, only 22% of fires flashes developed methods cannot predict. In all cases, accuracy in predicting of the fires number amplitude is less then 5%.

In contrast, it should be noted that in the understanding of this subject there are certain weaknesses which are reflected in the results. For example, if the satellite measures the increased inflow of any parameter of the solar activity, this does not prove that the charged particles will come into contact with the plant mass. Even if it could be proven in a laboratory that they could cause the initial phase of the flame, it does not necessarily mean that any sudden influx of particles would hit the U.S. territory. Additionally, when this occurs in conditions of increased humidity and/or cloudiness, the charged particles tend to not reach the ground, because the moisture in the air acts as absorbent [17].

Also, it is necessary to bear in mind that a certain time period is necessary from the moment of the rapid increase of flow of particles in certain energy ranges to the moment of registration of fire. This may explain relatively poor results of the prediction models that refer to the same day (lag 0). Though, in the case of accuracy of the models for fires that occur on the same day when there is a sudden influx of particles, it ranges from 29-39 %. In other words, in certain conditions, it is possible that in one (same) day it comes to increased solar activity and to detection of minor and/or larger fires. But as we note above the dependence on the lag is not very strong and requires further research.

On the other hand, the question is whether all small fires can be registered, especially given the fact that large parts of the U.S. territory are uninhabited. In case of large fires, the quality of the results is certainly burdened by the fact that at this moment we do not have reliable information how many of them emerged by coalescence of small fires and how soon.

As it can be seen from the Figure 2 and the Table 4 the results do not depend strongly on the lag. The difference between the results of the research is within 10% depending on the lag. This can be caused by several factors:

- forest fires flashes across the United States are analyzed in the paper. The difference in climate and atmospheric conditions, due to the large area, and at the same time in the plant world, leads to different inertia in the processes of ignition. This, in its turn, "lubricates" lag dependence;
- fires flashes depend on other factors than solar activity and they were not included in the model;
- in time series noise associated with fluctuations in climate, weather and other stochastic factors may be present.

The results of analysis indicate that the solar activity in specific narrow energy range (i.e. X_1) may lead to an increase of a number of forest fires. Therefore, despite the complexity of the analysis, the registration of all these factors enables to predict occurrence of forest fires just the same or next few days after the solar activity. According to Figure 4 we can conclude that X_1 is the most influential factor for lag 0. The increasing intensity of X_4 and X_5 factors lead to flashes of small fires with 3 days delays.

However, despite the relatively high values of prognostic ANFIS models, in order that the hypothesis is accepted, it is necessary to carry out experimental laboratory research. Prediction of place and time could be followed up in subsequent attempts. So far, the results indicate the possibility of a notice of potential hazards in terms of the timeline of events, while to assess the vulnerability of certain areas it is necessary to involve teams of different professional orientations.

References:

[1] Abraham, A. (2005). Adaptation of Fuzzy Inference System Using Neural Learning. In: *Fuzzy Systems Engineering: Theory and Practice* (eds. Nedjah, N. & de Macedo Mourelle, L.). Springer Verlag, pp. 53–83.

[2] Amini, M., Abbaspour, K.C. & Johnson, C.A. (2010). A comparison of different rule-based statistical models for modeling geogenic groundwater contamination. *Environ. Modell. Softw.*, 25(12), 1650-1657.

[3] Bektas Ekici, B. & Aksoy, U.T. (2011). Prediction of building energy needs in early stage of design by using ANFIS. *Expert Syst. Appl.*, 38(5), 5352-5358.

[4] Bell, W.R., Holan, S.H. & McElroy, T.S. (2012). *Economic Time Series: Modeling and Seasonality*. Chapman and Hall/CRC.

[5] Boxall, M. *et al.* (2009). ESS Guidelines on Seasonal Adjustment. Eurostat. Available at: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-09-006/EN/KS-RA-09-006-EN.PDF. Last accessed 26 December 2013.

[6] Cumming, S.G. (2001). Forest type and wildfire in the Alberta boreal mixedwood: What do fires burn? *Ecol. Appl.*, 11(1), 97-110.

[7] de Jager, C. (2012). Solar Forcing of Climate. Surv Geophys, 33:445–451.

[8] Ducic, V., Milenkovic. M. & Radovanovic, M. (2008). Contemporary Climate Variability and Forest Fires in Deliblatska pescara. *Journal of the Geographical institute Jovan Cvijic SASA*, No 58, p. 59-74.

[9] Gomes, J.F.P. & Radovanovic, M. (2008). Solar activity as a possible cause of large forest fires - a case study: Analysis of the Portuguese forest fires. *Sci. Total Environ.*, 394(1), 197–205.

[10] Güneri, A.F., Ertay, T. & Yücel, A. (2011). An approach based on ANFIS input selection and modeling for supplier selection problem. *Expert Syst. Appl.*, 38(12), 14907-14917.

[11] Guyette, P.R., Stambaugh, C.M., Dey, C.D. & Muzika, R-M. (2012). Predicting Fire Frequency with Chemistry and Climate. *Ecosystems*, 15(2), 322–335.

[12] Hall, L.B. (2007). Precipitation associated with lightning-ignited wildfires in Arizona and New Mexico. *Int. J. Wildland Fire*, 16(2), 242–254.

[13] Hansen, B.E. (2014). Econometrics. University of Wisconsin, Department of Economics. p. 378.

[14] Jang, J.-S.R., Sun, C.-T. & Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Inteligence*. Prentice Hall.

[15] Kourtz, P.H. & Todd, J.B. (1991). Predicting the daily occurrence of lightning-caused forest fires. Forestry Canada, Petawawa National Forestry Institute, Chalk River, Ontario. Information Report PI-X-112. Available at: http://cfs.nrcan.gc.ca/publications/?id=10706. Last accessed 26 December 2013.

[16] Kurtulus, B. & Flipo, N. (2012). Hydraulic head interpolation using anfis—model selection and sensitivity analysis. *Comput. Geosci.*, 38(1), 43-51.

[17] Labitzke, K. (2003). The global signal of the 11-year sunspot cycle in the atmosphere: When do we need the QBQ? *Meteorolog. Zeitshrift*, 12(4), 209-216.

[18] Lenskiy, A.A. & Seol, S. (2012). The analysis of R/S estimation algorithm with applications to WiMAX network traffic. *International Journal of Multimedia and Ubiquitous Engineering*, 7(3), 27-34.

[19] Mitra, S. K. (2012). Is Hurst Exponent Value Useful in Forecasting Financial Time Series? Asian Social Science Vol. 8, No. 8; July 2012, pp. 111-120.

[20] Mohandes, M., Rehman, S. & Rahman, S.M. (2011). Estimation of wind speed profile

using adaptive neuro-fuzzy inference system (ANFIS). *Appl. Energy*, 88(11), 4024-4032.

[21] Nikolov, N. (2006). Global Forest Resources Assessment 2005 – Report on fires in the Balkan Region. Forestry Department, FAO of the UN, Fire Management Working Papers FM/11/E, Rome. Available at: http://www.fao.org/docrep/009/j7567e/j7567e00.htm. Last accessed 26 December 2013.

[22] Özger, M. (2011). Prediction of ocean wave energy from meteorological variables by fuzzy logic modeling. *Expert Syst. Appl.*, 38(5), 6269-6274.

[23] Peters, E.E. (2003). *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*. John Wiley & Sons.

[24] Radovanovic, M. & Gomes, J.F.P. (2009). *Solar Activity and Forest Fires*. Nova Science Publishers Inc.

[25] Radovanović, M. (2010). Forest fires in Europe from July 22nd to 25th 2009. *Arch. Biol. Sci.*, 62(2), 419-424.

[26] Radovanovic, M. (2011). Solar Activity – Climate Change and Natural Disasters in Mountain Regions. In: *Sustainable Development in Mountain Regions* (ed. Zhelezov G.). Springer Science+Business Media B.V., pp. 9-17.

[27] Radovanović, M., Vyklyuk, Y., Jovanović, A., Vuković, D., Milenković, M., Stevančević, M. *et al.* (2013): Examination of the correlations between forest fires and solar activity using Hurst index. *Journal of the Geographical institute Jovan Cvijic SASA*, 63(3), 23-32.

[28] Rowell, A. & Moore, F.P. (2000). Global Review of Forest Fires. WWF, IUCN, 64. Available at: http://2010vision.org/WWF_Forest_Fires_Report.pdf. Last accessed 26 December 2013.

[29] Sannikov, S.N., Zakharov, A.I., Smol'nikova, L.G. & Sannikova, N.S. (2010). Forest fires caused by lightning as an indicator of connections between atmosphere, lithosphere, and biosphere. *Russ. J. Ecol.+*, 41(1), 1–6.

[30] Schmuck, G., San-Miguel-Ayanz, J., Camia, A., Durrant, T., Boca, R., Whitmore, C. *et al.* (2012). Forest Fires in Europe, Middle East and North Africa 2011. Joint Reseacrh Centre of the European Commission, 109. Available at: http://forest.jrc.ec.europa.eu/media/cms_page_media/9/forest-fires-in-europe-2011.pdf. Last accessed 26 December 2013.

[31] Shiri, J., Kisi, O., Yoon, H., Lee, K.-K. & Nazemi, A.H. (2013). Predicting groundwater level fluctuations with meteorological effect implications—A comparative study among soft computing techniques. *Comput. Geosci.*, 56, 32-44.

[32] Soltani, F., Kerachian, R. & Shirangi, E. (2010). Developing operating rules for reservoirs considering the water quality issues: Application of ANFIS-based surrogate models. *Expert Syst. Appl.*, 37(9), 6639-6645.

[33] Talebizadeh, M. & Moridnejad, A. (2011). Uncertainty analysis for the forecast of lake level fluctuations using ensembles of ANN and ANFIS models. *Expert Syst. Appl.*, 38(4), 4126-4135.

[34] Tan, Z., Quek, C., & Cheng, P.Y.K. (2011). Stock trading with cycles: A financial application of ANFIS and reinforcement learning. *Expert Syst. Appl.*, 38(5), 4741-4755.

[35] Viegas, D.X. (1998). Forest fire propagation. *Phil. Trans. R. Soc. London Ser. A*, 356, 2907–2928.

[36] Velásquez Valle, M. A., García, G. Medina, Cohen, Ignacio Sánchez, Oleschko, L. Klaudia, Corral, J. A. Ruiz, and Korvin Gabor. (2013). Spatial Variability of the Hurst Exponent for the Daily Scale Rainfall Series in the State of Zacatecas, Mexico. Journal of Applied Meteorology and Climatology, December 2013, Vol. 52, No. 12 : pp. 2771-2780

[37] Vyklyuk, Y (2013). Simulation of spatial form of urban systems by diffusion methods. *Journal of the Geographical Institute "Jovan Cvijic" SASA*, 63, #1,89–100, #2, 67–77

[38] Wotton, M.B., Stocks, J.B. & Martell, L.D. (2005). An index for tracking sheltered forest floor moisture within the Canadian Forest Fire Weather Index System. *Int. J. Wildland Fire*, 14(2), 169-182.

[39] Yilmaz, I. & Kaynar, O. (2011). Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. *Expert Syst. Appl.*, 38(5), 5958-5966.