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

by

Milan M. RADOVANOVIĆ [a*], Yaroslav VYKLYUK [b], Slavica B. MALINOVIĆ-MILIĆEVIĆ [c], Dejana M. JAKOVLJEVIĆ [a], and Milovan R. PECELJ [d]

[a] Geographical Institute "Jovan Cvijić" of the Serbian Academy of Sciences and Arts, Belgrade, Serbia
[b] Bukovinian University, Chernivtsi, Ukraine
[c] University Center for Meteorology and Environmental Modelling, University of Novi Sad, Serbia
[d] Geographical Faculty, University of Belgrade, Serbia

Original scientific paper
DOI: 10.2298/TSCI141103150R

The work examines the potential causative link between the flow of protons, i. e. temperature of the particles that are coming from the Sun and forest fires in the USA. For determination of the degree of randomness for time series of input (temperature of protons) and output parameters (number of forest fires), the R/S analysis is conducted. The analysis of fractal dimension provides us an opportunity to compare self-similar processes in the influx of protons and the time series of forest fires flashes. Therefore we developed and conducted the sensitivity analysis of model based on hybrid neural networks ANFIS. As the calculations showed, only 16% of the real forest flashes cannot be predicted by the model.

Key words: *forest fires, temperature of protons, forecast, ANFIS models, USA*

Introduction

Forest fires, besides emitting the burning products in the atmosphere, considerably affect the relatively limiting heat conditions of the area where they occur. To which extent they represent heat, *i. e.* climate modifier (even in relatively small areas, *i. e.* regions) there have not been reliable estimations so far.

In the recent years there were several papers where the forest fires were brought into the connection with the solar activity [1-6]. In the research, it has been come to a hypothesis that the charged particles of the solar wind (SW) in certain conditions reach the surface of the Earth, where they can cause fires in vegetation. Many of the mentioned papers refer to the individual cases of fires in the periods not longer than a month. Gomes and Radovanović [7] analysed 11 cases of the forest fires on different locations in Europe in the period November 2002 – August 2005. In all these cases the fires were preceded by similar situations on the sun in

* Corresponding author; e-mail: m.radovanovic@gi.sanu.ac.rs

which the energy regions in the geoeffective position emitted strong electromagnetic waves towards the Earth. The velocities of the SW particles were moving in the range of 550 km/s to over 1,000 km/s, while their temperature exceeded 1,000,000 K in some periods.

Due to the need to check the hypothesis on the SW as the cause of fires on as larger number of cases as possible, Radovanović *et al.* [8] used the daily dynamics of the forest fires in the USA in the period 2004-2007 (May-October). As the parameters of the solar activity, the following ones were used: the flow of >1 MeV protons, the flow of >10 MeV protons, the flow of >100 MeV protons, the flow of 0.6 MeV electrons, and the flow of >2 MeV electrons, the 10.7 cm solar flux and the SW speed. The Hurst index was used in the research, and the similar fractal characteristics ($0.72 = H = 0.92$) were obtained for the 10.7 cm solar flux, the SW speed and forest fires data series. In attempt to ascertain more precisely the quantitative parameters of the potential causative-effective explanation of the formation of the initial phase of fire, here in the paper we focused on the link between the dynamics of the flow of protons, *i. e.* their temperature and the forest fires in a longer time period.

In this paper, we did not try to elaborate the proposed model of the propagation of the charged particles to the ground [9], but we only tried to examine the quantitative relationships between the dynamics of the proton temperature and the occurrence of fires in the USA and the deviation of the number of fires from the predictions on the basis of the long-term variations. The obtained results, based on the ANFIS (application of adaptive neuro fuzzy inference system) model, enabled us the development of a prognostic model.

Data and methods

Starting from this, the main goal of the paper was to analyse the relationship between the temperature of protons and the number of forest fires in the time period of 15 years and the deviation from the predictions on the basis of the long-term variations. The decision to test the hypothesis especially in the case of the USA was made due to the availability of data on fires in a relatively large area and on a daily basis. The daily data on forest fires in the USA comprise the period May-October in which the largest number of phenomena is recorded. On the basis of the structure of available data bases, it was decided to take the period 1999-2013 for the research. The number of fires (F) is taken to be the output variables. The dataset for (F) has been downloaded from [10]. These are daily values from Incident Management Situation Report by the National Interagency Coordination Centre, Boise, Id., USA.

The study comprised the period from May to October in each of the mentioned years. Data processing in the colder part of the year is not done, because of the relatively small number of fires during the winter in the whole northern hemisphere. The dataset consisted of 2,760 records. The input factors (like indicators of solar activity) have been selected as the proton temperature (Tp). The dataset has been downloaded from [11]. These are daily averages. It is important to note that these data are obtained from the satellites that are always located between the Sun and Earth. In this way, in the real time, we get the information about the arrival of protons of corresponding temperatures, which are directed towards our planet. Unfortunately, the data on the electron temperature are not measured.

The preliminary analysis has shown that the datasets contain the missing data (*holes*). The deleting of records that contain *holes* creates the data time gap which prevents the further inspection into the *lag* (time between the onset of fires and the solar activity) dependence. The number of the *holes* is a relatively small percentage of the Tp and F data (~10%). Also, we do not have the information about the time distributions of data. In this case, the maximum-likelihood estimation (MLE) is used as the basic method for solving these types of issues [12]. The method gives a unified approach to estimation, which is well-defined in the case of the normal distribution and many other problems.

Results

The time dependence of the input and output factors is presented in the fig. 1. Large stochasticity in the whole studied range is present in this figure. As it can be seen, the graph F contains the periodic *waves* associated with the annual season changes and not dependent on the Sun flashes, because there are no visible similar *waves* on Tp graph. Therefore, to further study the impact of the solar activity on forest fires it is necessary to eliminate the seasonality of the factor F.

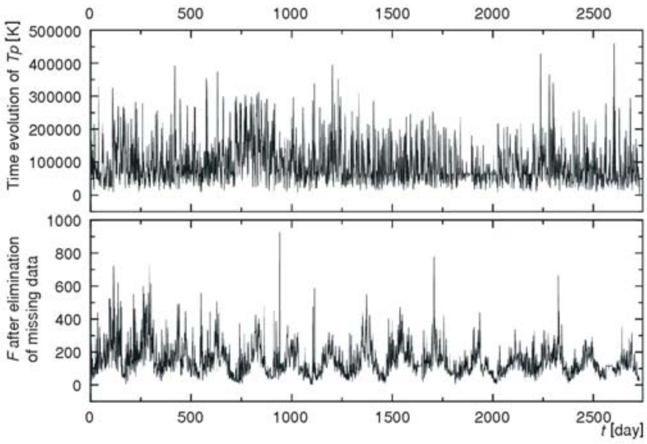

Figure 1. Time series of *Tp* (upper panel) and *F* fields after elimination of missing data (lower panel)

The additive model was used to remove the seasonal component decomposition of the F time series [13]. This model for the F time series is:

$$F = T + S + \widetilde{F} \tag{1}$$

where T is the trend component, *i. e.* long-term change course of the number of fires; S – the seasonal component of the number of fires, and $\widetilde{F}$ – the irregular component, which is related to some other factors, for example solar activity. The classical method of seasonality index was used to remove seasonality. The method of least squares was used to obtain the trend [14]. The obtained linear trend model is:

$$T = -0.024t + 180.9 \tag{2}$$

where t is the time in days.

So the occasional component $\widetilde{F}$ was received, cleared from seasonal and trend components (fig. 2). This component is a deviation from the average values of seasonal and trend components for the whole study period. That is analogous to *white noise*. Such deviations can be both positive and

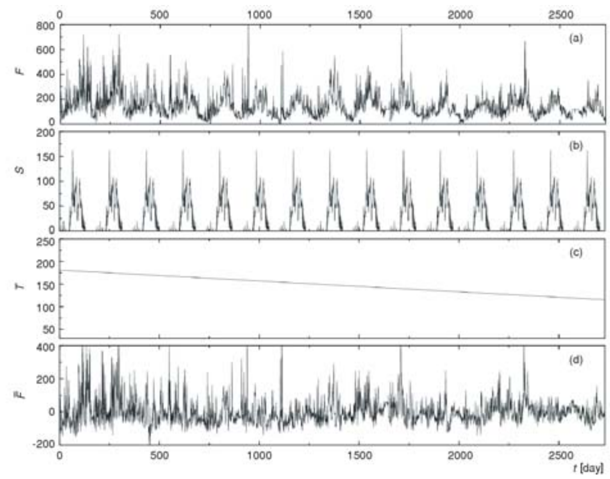

Figure 2. Time series of *F*; (a) initial data, (b) seasonal component, (c) trend, (d) time series, cleared from seasonal and trend components

negative. According to the heliocentric hypothesis, $\tilde{F}$ depends on the characteristics of the solar activity fluctuations. It was used in the next calculations as output field.

To test the hypothesis on the presence of the functional dependencies between the component of the solar activity and forest fires flashes, the correlation analysis was made between the factor *T*p and the number of fires $\tilde{F}$, taking into account time delay (lag). As shown in [13], this delay can last up to 0-5 days. The results of the analysis are shown in the tab. 1.

Table 1. Pair correlation coefficients between input (Tp^L) and output ($\tilde{F}$) variables on time lag $L = \overline{0.5}$

Lag	5	4	3	2	1	0
Correlation (R)	–0.018	–0.019	–0.026	–0.041	–0.022	0.001

As it can be seen, any absolute value of the correlation coefficient is not higher than 0.041. It means that there are no linear relationships between the mentioned factors even taking into account lag. Therefore it is necessary to apply methods of the nonlinear analysis to further research.

As we showed in previous work [14, 15], the classical statistical linear methods cannot explain the *heliocentric hypothesis*. As it was shown, the most informative was the nonlinear analysis such as R/S and Data Mining. Therefore they were used in this research.

For determination of the degree of randomness for the time series of the input and output parameters, the R/S analysis was conducted [16-18]. The R/S analysis enables determining the fractal dimensions and whether the time series are the stochastic ones or they have long-terminal correlation (long-terminal memory). The results of these calculations are shown in the tab. 2.

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

	Hurst index	Fractal dimension
Tp	0.84	1.16
$\tilde{F}$	0.74	1.26

As it can be seen from the tab. 2, the Hurst index of the similarity of the fractal dimensions for *Tp* means the 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 dependent on the same factors or on each other [8]. Close to 1, the value of fractal dimension indicates the average level of volatility of the time series. This means that the nonlinear forecasting methods can be used. However, this fact needs further verification.

The problem of finding the hidden dependencies in relatively large databases is related to problems of Data Mining. Therefore, the comparative analysis of models based on ANFIS was investigated and proved in the work. This approach, as demonstrated by our previous calculations, allowed obtaining better results, in contrast to neural networks [19].

Generally, the problem is reduced to finding the dependence in the form:

$$M: Tp^0 \times .. \times Tp^5 \to \tilde{F} \tag{3}$$

The training sets in the form of corteges were created for learning of ANFIS model:

$$Tr = \left\{ \left\langle tp_j^0, tp_j^1, tp_j^2, tp_j^3, tp_j^4, tp_j^5, \tilde{f}_j \right\rangle \right\}_{j=\overline{1,n}} \tag{4a}$$

where

$$Tp^L = \left\{ tp^L \right\}_{j=\overline{1,n}} \tag{4b}$$

and *n* is the size (number of records) of training set (after lag transformation it had 2,654 records).

When forming the training set we should take into account the time delay (lag) between the Tp and the fires caused by it. To do this it is necessary to transform the time series Tp taking into account gaps of time series in the winter months.

For building the ANFIS models, all input parameters were presented as linguistic variables. Since the non-linear dependence is present, each of the linguistic variables is identified by the non-linear Gauss terms. As the test calculations have shown, the optimal count was 3 Gauss terms for Tp and $\widetilde{F}$. The objective models were not obtained for 2 Gauss terms. If Gauss terms are larger than 3, it leads to a sharp increase in the training time. The Sugeno function of zero order [19] was selected as a method of output fuzzy system because of the training time, too. The hybrid method that integrates back-propagation method with the least squares method was used as a method of learning. Built and trained, ANFIS models consist of 729 fuzzy rules and are shown in fig. 3.

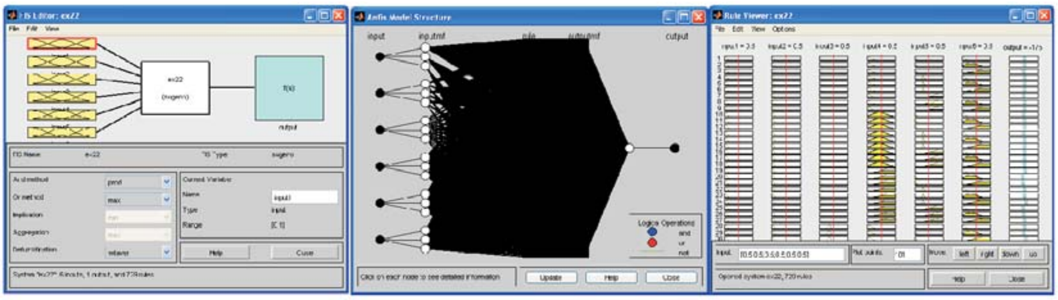

Figure 3. Structure of ANFIS model (general structure, fuzzy rules)

Since the model consists of a lot of fuzzy rules, it is necessary to analyse the adequacy and accuracy of the models using comparative analysis of the real and simulated data. The simulation results are presented in the fig. 4. As it can be seen from this figure, the graphs have many joint peaks. Large number of peaks requires machine analysis.

The model from the previous figure gives the possibility to explain the influence of the solar activity on forest fires represented by proton temperature. Almost every peak on the model graph corresponds to the peak in the graph of real fires by position and amplitude. This indicates the adequacy of the model. To check the accuracy, the comparative analysis between a number of real fires flashes and flashes predicted by model have been provided using created program. To do this, we calculated the number of flashes (peaks) on figures from graph 4, and compared their number, number of coincidences and non-coincidences in the particular days, and mean-square errors in peaks amplitude (number of fires in particular day when peak was observed). As the peak of the fires we assumed the day in which the number of fires was higher than the mean-seasonal and trend indi-

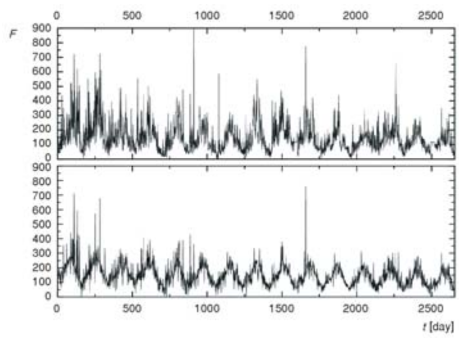

Figure 4. Comparison of simulation results (bottom panel) with real data (upper panel)

cators and was higher than the number of fires in the previous and the next day. The results of computer calculations are shown in the tab. 3.

Table 3. Accuracy analysis of fires flashes prediction for ANFIS model

Correlation coefficient between real data and simulation results	Real fires flashes	Model fires flashes	Explained by model fires flashes		Average difference in amplitude	False peaks		Cannot predict	
1	2	3	4	5	6	7	8	9	10
0.75	770	821	378	49%	1.3%	443	54%	127	16%

As it can be seen from tab. 3, the model is characterized by high accuracy of prediction, as evidenced by the high correlation coefficient between real data and simulation results. As it can be seen from the tab. 1, we were able to largely eliminate the non-linearity and to increase the correlation coefficient from ~ –0.041 to ~ 0.75. This indicator is higher than in our previous work, despite the fact that only one solar wind characteristic – *Tp* has been taken into account. This indicates the adequacy of the approach. In addition, all data presented in the table are better than in the previous works [13]. As it can be seen from the previous table, the model can predict for up to 49% of the flash fires (column 5). It should be noted that the model predicts on average 54% of the cases of the false flashes (column 8). This may indicate the effectiveness of preventive measures on the spread of fire flashes in the USA. It should be noted that the model predicts a little more flash fires than it is (column 2 and 3). More important information is how many real fire flashes the 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 modelling graph at the same time value was below the average (column 10). As the calculations have shown, only 16% of the real forest flashes cannot be predicted by the model.

The predicted amplitude of peaks (a number of predicted fire flashes in a particular day in the USA) can also be interesting. As shown in the previous table (column 6), the amplitude is usually smaller on average by 1.3% than the actual number of flashes.

In the next step it was necessary to examine the influence of parameter lag. To do this, the values of *Tp* for all lags were fixed to its average (92,482 °K) and the dependence of fire occurrence from sequential changes in each input parameter (lag) was analysed. The results of this analysis are presented in the fig. 5.

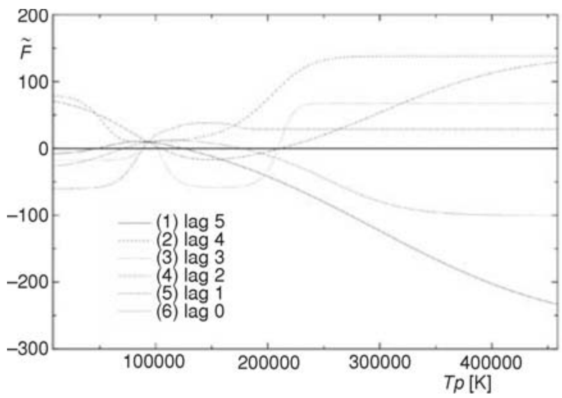

Figure 5. Dependence of $\widetilde{F}$ on proton temperature for different lags

For checking the adequacy of model, it is necessary to analyse the behaviour of calculated plots on the graph. If the temperature of protons for 6 days did not change, the model should show zero value of flash fires. It is clearly seen from the graph (point 92,482 K). It should be expected that with the decreasing intensity of the proton temperature from the mean value, the number of forest fires should decrease, or remain unchanged. On the contrary, increasing the values of the factors, the number of fires should increase, too. But we observed different situation. This means that the situation modelled in sensitivity analyses never occurs in nature. Therefore, this analysis will set only the most influential lags. For example, at small temperature fluctuations, the most influential is the lag 1. When significant *Sun flashes* occur, the most significant lags are 5, 3 and 2. The time difference between the sudden influx of the proton flow and the occurrence of forest fires can occur due to several reasons. If it is hypothetically assumed that the charged particles arriving at the Earth are measured on the satellite at the end of the day (UTC), then it can be expected that the eventual causative-effective connection with the fires may occur the next day earliest. Also, if in such developments, the particles reach the ground and succeed to cause the initial phase of ignition, there are a number of factors that influence the potential spread of fire (relative humidity and air temperature, humidity of combustible materials, the distribution of winds, *etc.*). In remote, *i. e.* unpopulated areas numerous fires remain unregistered. For those fires that are observed, big problem is the determination of the exact time of their occurrence. This means that the period of smouldering can last for several days before the flame covers the wider area, that is, before it is recorded.

Conclusions

We can conclude that the developed model is surprisingly accurate and relatively precise for real data. Flashes of forest fires are the most sensitive to changes in temperature of protons in the previous 2, 4, and 5 days.

Despite the prediction accuracy, both in time and in amplitude, these models do not allow to predict the geographic location of sources of fire. The reason is the lack of geospatial information in the training set. This disadvantage can be eliminated if the information is attached to the database. In addition, the study did not take into account the parameters related to the electrons even despite the fact that there are results that they can emerge as a potential cause of occurrence of fire [20]. Based on the above, we can also assume that better results could be expected if we possessed the data about the fires on a larger surface. Theoretically, it is possible to expect that a sudden influx of the charged particles which come to us from the sun does not necessarily reach the surface in the territory of the USA only.

Acknowledgments

The results are a part of the project III 47007, funded by the Ministry for Education, Science, and Technological Development of the Republic of Serbia.

Special thanks go to Milan Stevančević, without whose help, this work would not exist. We are also grateful to the reviewers on their helpful suggestions.

We thank the ACE SWEPAM instrument team and the ACE Science Center for providing the ACE data.

References

[1] Radovanović, M., *et al.*, Endangerment of Undeveloped Areas of Serbia by Forest Fires, *Journal of the Geographical Institute "Jovan Cvijić" SASA, 59* (2009), 2, pp. 17-35
[2] Gomes, J. F. P., *et al.*, Wildfire in Deliblatska Pescara (Serbia) – Case Analysis on July 24th 2007, in: *Forest Fires: Detection, Suppression and Prevention* (Eds. E. Gomez, K. Alvarez), Nova Science Publishers, New York, 2009, pp. 89-140
[3] Radovanović, M., Solar Activity – Climate Change and Natural Disasters in Mountain Regions, in: *Sustainable Development in Mountain Regions* (Ed. G. Zhelezov), Springer Science+Business Media B.V., 2011, pp. 9-17

[4] Radovanović, M., Astrophysical Hypothesis of Forest Fires with Special Regards on Baikal Lake Region, *Proceedings*, International scientific conference "Baikal – the World Treasure", UNESCO Headquarters, Paris, 2012, pp. 192-198
[5] Milenković, M., *et al.*, Fire Protection Problems with Large Forest Fires in Deliblatska Pescara (Serbia), *Journal of the Geographical Institute "Jovan Cvijić" of the Serbian Academy of Sciences and Arts, 63* (2013), 3, pp. 269-278
[6] Radovanović, M., *et al.*, The Influence of Solar Activities an Occurrence of the Forest Fires in South Europe, *Thermal Science*, *19* (2015), 2, pp. 435-446
[7] Gomes, F. P. J., Radovanović, 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, 394* (2008), 1, pp. 197-205
[8] Radovanović, M., *et al.*, Examination of the Correlations between Forest Fires and Solar Activity Using Hurst Index, *Journal of the Geographical Institute "Jovan Cvijić" SASA, 63* (2013), 3, pp. 23-31
[9] Gomes, F. P. J., *et al.*, On the Relationship between Solar Activity and Forest Fires, in: *Handbook on Solar Wind: Effects, Dynamics and Interactions* (Ed. H. E. Johannson), Nova Science Publishers, 2011, pp. 1-79
[10] ***, Incident Management Situation Report (IMSR) Archives, National Interagency Coordination Center, Boise, Id., USA, http://www.predictiveservices.nifc.gov/intelligence/archive.htm
[11] ***, ACE Data Set: SWEPAM Daily Averaged Solar Wind Parameters, NASA, http://www.srl.caltech.edu/cgi-bin/dib/rundibviewswel2/ACE/ASC/DATA/level2/swepam?swepam_data_1day.hdf
[12] Little, R. J. A., Rubin, D. B., *Statistical Analysis with Missing Data*, John Wiley and Sons, Hoboken, N. J., USA, 2002
[13] Radovanović, M., *et al.*, Application of ANFIS Models for Prediction of Forest Fires in the USA – Preliminary Results –, *Thermal Science,* 2015*,* DOI:10.2298/TSCI150210093R
[14] Radovanović, M., *et al*., Conceptual Modelling of Forest Fires Flashes by Data Mining Ontology-based (Part 1) (Konceptual'noe Modelirovanie Vspyshek Lesnych Pogarov na Osnove Ontologicheskogo Podhoda Data Mining, in Ukraine), *System Research and Information Technologies, 4* (2014)*,* pp. 111-120
[15] Radovanović, M., *et al.*, Conceptual Modelling of Forest Fires Flashes by Data Mining Ontology-Based (Part 2) (Konceptual'noe Modelirovanie Vspyshek Lesnych Pogarov na Osnove Ontologicheskogo Podhoda Data Mining, in Ukraine), *System Research and Information Technologies, 1* (2015), (in press)
[16] Lenskiy, A. A., Seol, S., The Analysis of R/S Estimation Algorithm with Applications to WiMAX Network Traffic. *International Journal of Multimedia and Ubiquitous Engineering, 7* (2012), 3, pp. 27-34
[17] Bell, W. R., *et al.*, *Economic Time Series: Modelling and Seasonality*, Chapman and Hall/CRC, 2012
[18] Hansen, B. E., *Econometrics*, University of Wisconsin, Department of Economics, 2014, http://www.ssc.wisc.edu/~bhansen/econometrics/
[19] Zadeh, L. A., Fuzzy Algorithms, *Information and Control, 12* (1968), 2, pp. 94-102
[20] Radovanović, M., Forest Fires in Europe from July 22nd to 25th, 2009, *Archives of Biological Sciences, 62* (2010), 2, pp. 419-424

Paper submitted: November 3, 2014
Paper revised: December 12, 2014
Paper accepted: December 26, 2014