Evaluation of the transition probabilities for daily precipitation time series using a Markov chain model

Liana Cazacioc and Elena Corina Cipu

Abstract

The Markov models are frequently proposed to quickly obtain forecasts of the weather "states" at some future time using information given by the current state. One of the applications of the Markov chain models is the daily precipitation occurrence forecast. There is tested a Markov chain model with two states for the daily precipitation in summer and winter seasons of 1961-1990 at several stations in Romania. The states of the Markov chain are precipitation occurrence and precipitation non-occurrence, that is wet and respectively dry days. There are computed the sets of conditional (or transition) probabilities for first-order, second-order and third-order Markov chain. To find the most appropriate model order among the different orders of the Markov chains for the daily precipitation series, the Bayesian information criterion (BIC) was used.

Mathematics Subject Classification: 76A05, 76A10, 76M25.
Key words: Markov chnanel model, transition probabilities, daily precipitation.

1 Introduction

There are three methods used for precipitation forecasting: the subjective prognosis based on the experience of the forecasters, the deterministic prognosis obtained from the models of numerical weather prediction and the statistical prognosis. The two last are part of the objective techniques.

One of the statistical techniques is the Markov chain used to predict precipitation on short term, at meteorological stations. The Markov chain models have two advantages: (1) the forecasts are available immediately after the observations are done because the use as predictors only the local information on the weather and (2) they need minimal computation after the climatological data have been processed.

The idea of applying the stochastic processes theory in meteorology belongs to Gabriel and Neuman (1962) who have developed a Markov chain model for the daily precipitation in Tel Aviv. Using a set of data containing the daily mean values of

Proceedings of The 3-rd International Colloquium "Mathematics in Engineering and Numerical Physics" October 7-9 , 2004, Bucharest, Romania, pp. 82-92.

departmental precipitation over 11 years of observations, Ladoy (1974) have calculated the transition probabilities of the daily spatial distributions in France. The Australians researchers (Hess et al. 1989) have demonstrated that this technique is good enough to build an objective base of the precipitation occurrence probability in short-term prognosis. It can be also reminded the study on evaluating the monthly precipitation amounts anomalies of the Romania's territory (Mares 1976), based on a Markov chain model of 1^{st} and 2^{nd} order. A 5-order Markov chain was applied to analyse the producing of extreme events in the monthly precipitation field at meteorological stations in Romania (Mares and Mares 1993).

In this paper there are presented the results of testing the Markov chain of first, second and third-order from the daily precipitation series in summer and winter seasons at some representative meteorological stations in Romania.

2 First-Order Markov Chain

A first-order Markov chain is one in which knowing one variable (like cloudiness, precipitation amount, temperature, fog, frost, wind) at time t is sufficient to forecast it at some later time.

The simplest kind of discrete variable is that which has binary values (yes/no), corresponding to the two states in which it can exist. For the daily precipitation, those are their occurrence or non-occurrence. A sequence of daily observations of "precipitation" and "no precipitation" from a meteorological station constitute a time series of that discrete variable.

For the first-order Markov chain, the transition probabilities controlling the future state of the studied variable depend only on its current state (Wilks 1995). Knowing that at time t the variable X is either in state 0 (no precipitation occurs and $X_t = 0$), or in state 1 (precipitation occurs and $X_t = 1$), conditional transition probabilities are computed at time t+1, where the time step is one day. That is,

$$
\begin{aligned}
p_{00} &= Pr(X_{t+1} = 0 \mid X_t = 0) \\
p_{01} &= Pr(X_{t+1} = 1 \mid X_t = 0) \\
p_{10} &= Pr(X_{t+1} = 0 \mid X_t = 1) \\
p_{11} &= Pr(X_{t+1} = 1 \mid X_t = 1)
\end{aligned} \tag{2.1}
$$

The transition probabilities estimation procedure consists of computing the conditional relative frequencies, as follows:

$$
\hat{p}_{01} = \frac{n_{01}}{n_{0\bullet}}\ , \quad \hat{p}_{11} = \frac{n_{11}}{n_{1\bullet}} \tag{2.2}
$$

Notations used were:n_{01}-number of transitions from state 0 to state 1, n_{11}-number of pairs of time steps in which the system keeps in state 1,$n_{0\bullet}$ -number of states 0 in the series followed by another data point, and $n_{1\bullet}$ -number of states 1 in the series followed by another data point. That is, $n_{1\bullet} = n_{10} + n_{11}$ and $n_{0\bullet} = n_{00} + n_{01}$.

For a Markov chain describing the daily occurrence or non-occurrence of precipitation, the stationary probability for precipitation (state 1), π_1, corresponds to the (unconditional) climatological probability of precipitation. Using the transition probabilities p_{01} and p_{11},

$$\pi_1 = \frac{p_{01}}{1 + p_{01} - p_{11}}. \tag{2.3}$$

The transition probabilities also imply a serial correlation degree or persistence for the time series. The lag-1 autocorrelation or persistence parameter of the time series writes in terms of the transition probabilities as

$$r_1 = p_{11} - p_{01}. \tag{2.4}$$

The mean length of the persistence sequence is computed (Yao 1982) with the formula:

$$L_E = \frac{1}{1 - p_E}, \quad E = 0, 1. \tag{2.5}$$

where p_E is the occurrence probability of event E which may take value 0 or 1, respectively.

3 Higher-Order Markov Chains

Consider for instance a second-order Markov chain. Second-order time dependency means that the transition probabilities depend on the states at lags of both one and two time periods. Then, the transition probabilities for a second-order Markov chain require three subscripts: the first refers to the state at time t-1, the second to the state at time t, and the third specifies the state at time t+1. The notation for the transition probabilities of a second-order Markov chain can be defined as

$$p_{hij} = Pr(X_{t+1} = j \mid X_t = i, X_{t-1} = h), \quad h, i, j = 0, 1. \tag{3.6}$$

The transition probabilities for a second-order Markov chain yield with formula

$$\hat{p}_{hij} = \frac{n_{hij}}{n_{hi\bullet}}. \tag{3.7}$$

That is, knowing that the value of the time series at time t-1 was $X_{t-1} = h$ and at time t, $X_t = i$, the probability that the future state of the time series $X_{t+1} = j$ is p_{hij}.

Similarly, for a third-order Markov chain, the notation requires four subscripts on the transition counts and transition probabilities: the first refers to the state at time t-2, the second to the state at time t-1, the third to the state at time t, and the fourth specifying the state at time t+1. The notation for the transition probabilities of a third-order Markov chain can be defined as

$$p_{ghij} = Pr(X_{t+1} = j \mid X_t = i, X_{t-1} = h, X_{t-2} = g), \; g, h, i, j = 0, 1. \tag{3.8}$$

The transition probabilities for a third-order Markov chain are obtained from the conditional relative frequencies of the transition counts $X_{t-2} = g$, the value of the time series at time t-1 was $X_{t-1} = h$ and the value of the time series at time t, $X_t = i$, the probability that the future state of the time series $X_{t+1} = j$ is p_{ghij}.

4 Determining the Order of the Markov Chain

Two criteria are use to decide among different orders of the Markov chain models: the Akaike criterion (AIC) and Bayesian criterion (BIC). Both are based on the log-likelihood functions for the transition probabilities of the Markov chains constructed on a certain data series. These log-likelihoods depend on the transition counts and the estimated transition probabilities. The log-likelihoods for Markov chains of order 0, 1, 2 and 3 are

$$
\begin{aligned}
L_0 &= \sum_{j=0}^{s-1} n_j ln(\hat{p}_j) \\
L_1 &= \sum_{i=0}^{s-1}\sum_{j=0}^{s-1} n_{ij} ln(\hat{p}_{ij}) \\
L_2 &= \sum_{h=0}^{s-1}\sum_{i=0}^{s-1}\sum_{j=0}^{s-1} n_{hij} ln(\hat{p}_{hij}) \\
L_3 &= \sum_{g=0}^{s-1}\sum_{h=0}^{s-1}\sum_{i=0}^{s-1}\sum_{j=0}^{s-1} n_{ghij} ln(\hat{p}_{ghij})
\end{aligned}
\tag{4.9}
$$

Here the summations are performed over all the states s of the Markov chain. For two states there will be included just two terms corresponding to the binary time series. Statistics of the two criteria differ slightly as can be seen below

$$
\begin{aligned}
AIC(m) &= -2L_m + 2s^m\ (s-1), \\
BIC(m) &= -2L_m + s^m\ (ln(m)).
\end{aligned}
\tag{4.10}
$$

The order m is chosen as appropriate that minimizes either equation $(3.6)_1$ or equation $(3.6)_2$. BIC criterion tends to be more conservative, indicating lower orders than AIC. It is preferable to use BIC statistic for sufficiently long time series, i.e. for those containing a number of data from around n=100 to over n=1000.

5 Results

There are presented the results of the application of the Markov chains of 1, 2nd and 3rd order to the daily precipitation series in the summer and winter seasons at 8 representative stations (Table 1) from Romania.

5.1 First-Order Markov Chain

The stationary (climatological) probability for the summer precipitation (see Figure 1b), calculated with equation (2.3), has the highest values in the northwest of the country and in the mountain area. The western atmospheric circulations carrying wet air masses mostly influence these areas and determine here a bigger number of wet days than in the rest of the territory. A similar distribution is pointed out in winter (see Figure 1a) when the stationary probability is bigger than in summer, but is not so homogeneous in the mountain area.

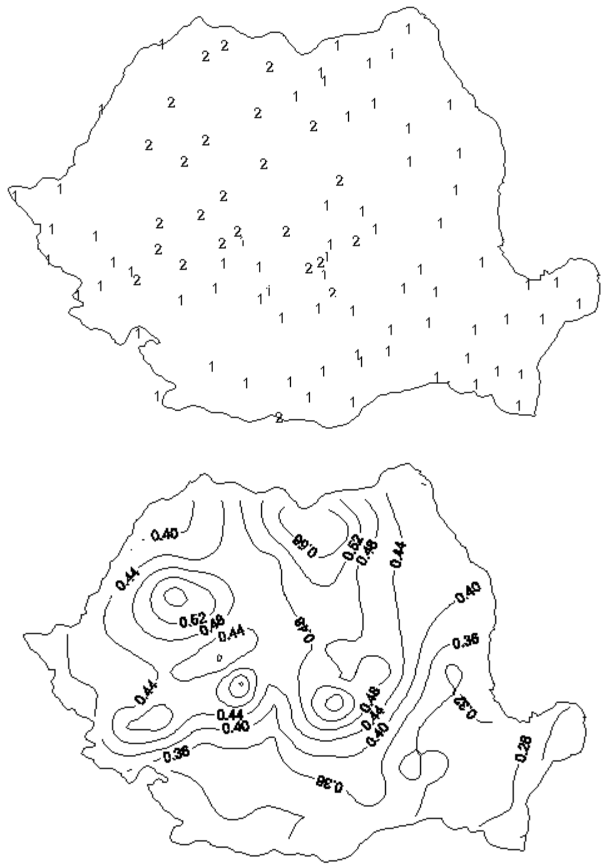

Figure 1: Climatological probability of daily precipitation occurance over Romania: a) in winter and b) in summer (1961-1990)

The persistence parameter given by (2.4), shows in both seasons a substantial serial correlation in the time series in the northwestern part of Romania (see Figure 2a and Figure 2b), where are placed the highest values. The persistence parameter field has a much higher gradient in winter than in summer, and is better highlighted in the cold season than in the warm one, accordingly to the nature of the precipitation.

The sequential analysis of the precipitation occurrence during the summer shows that the greatest mean lengths of the persistence sequences, calculated with the relation (2.5), are reached in the mountain area: 2.46 days in the northern part of the Eastern Carpathians, 2.4 days at Vladeasa Mount and 2.42 days in Bucegi Mounts (see Figure 3b). In winter, the mean length of the persistence sequence is longer

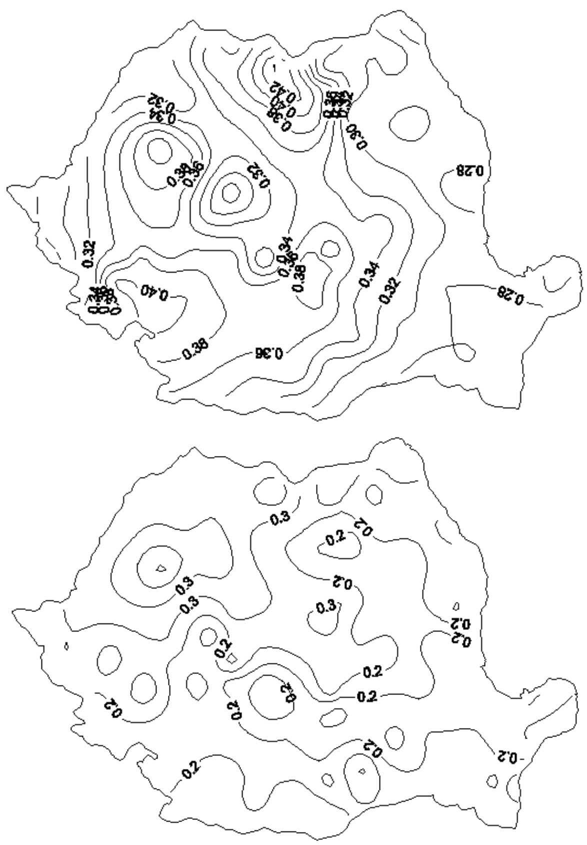

Figure 2: Persistence parameter for daily precipitation over Romania: a) in winter and b) in summer (1961-990)

than in summer (see Figure 3a) reaching 2.64 days in Bucegi Mounts, 2.54 days at Vladeasa Mount and 2.5 days in Eastern Moldova. The longest mean persistence sequence of the precipitation non-occurrence that results from (2.5), is being recorded in the southwest and southeast parts of the country (see Figure 4a and Figure 4b), with higher values in summer than in winter.

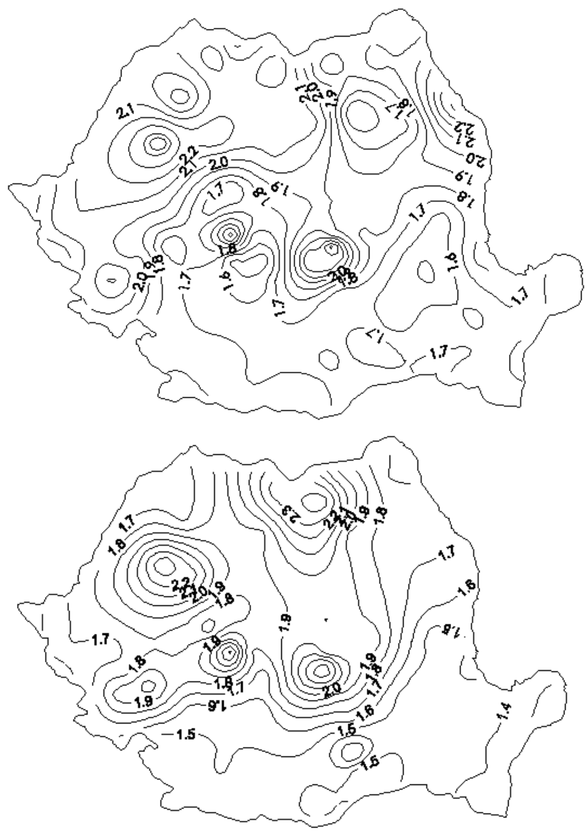

Figure 3: Mean length persistence sequence of the precipitation occurrence (days) over Romania: a) in winter and b) in summer (1961-1990)

The transition probabilities of the dry sequence, p_{00} are higher than those of the wet sequence, p_{11} in both seasons at all the stations (Table 1). The highest probability that a dry day will be followed by a dry day too is being registered in wintertime in southwest (Piclisa) Romania and in summertime in southeast (Constanta) Romania. A wet day will be followed with the highest probability by another wet day in the western part of territory during the winter (Baia Mare) and in the central part during the summer (Tg. Mures).

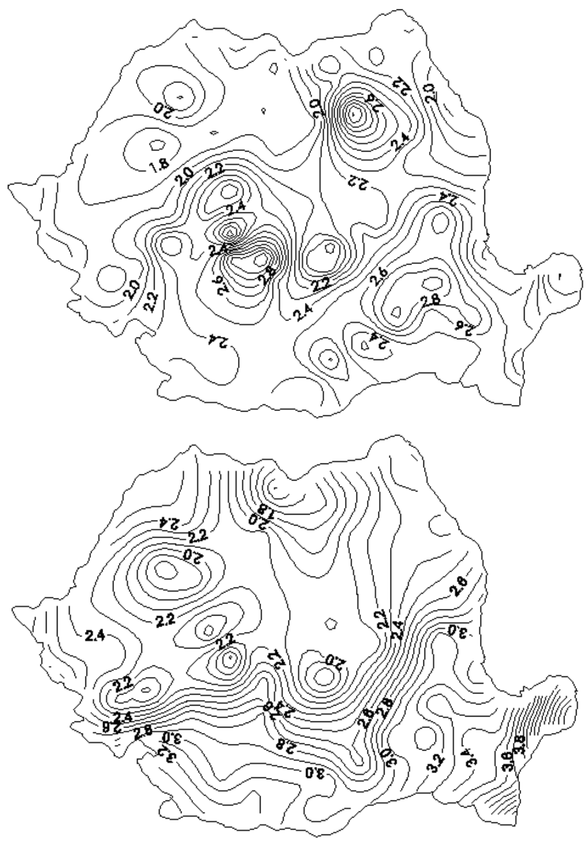

Figure 4: Mean length persistence sequence of the precipitation non-occurrence (days) over Romania: a) in winter and b) in summer (1961-1990)

5.2 Second-Order Markov Chain

For the second-order Markov chain with two states, applied to the time series of the daily precipitation in summer and in winter of 1961-1990 at the meteorological stations in Romania, the transition probabilities have been calculated accordingly to the relation (3.2) and a few results are presented in Table 1. The probability that two dry days will be followed by another dry day, p_{000}, is higher at the southern stations than in the rest of the territory, during the summer season. In winter, the highest probability of lack of precipitation in three consecutive days was found in southwest of the country (Turnu Severin and Piclisa). Two wet days will be followed with the highest probability by another wet day in northwestern Romania (Baia Mare), in both summer and winter. Comparing the two seasons, the probability of precipitation occurrence in three consecutive days, p_{111}, is higher in winter than in summer, at almost all the stations, in accordance with the dominant precipitation nature in each season.

Table 1. Transition probabilities computed for 1st, 2nd and 3rd order Markov chains at several stations from Romania

Season		Avra-meni	Bacau	Baia Mare	Buc. Baneasa	Con-stanta	Turnu Severin	Pi-clisa	Tirgu Mures
Winter	p_{00}	0.669	0.665	0.65	0.699	0.681	0.735	0.753	0.656
Summer		0.69	0.682	0.682	0.709	0.799	0.759	0.659	0.651
Winter	p_{11}	0.597	0.628	0.732	0.643	0.596	0.67	0.58	0.687
Summer		0.535	0.553	0.599	0.561	0.463	0.462	0.575	0.608
Winter	p_{000}	0.677	0.676	0.699	0.725	0.695	0.753	0.776	0.689
Summer		0.698	0.686	0.679	0.711	0.812	0.773	0.673	0.655
Winter	p_{111}	0.595	0.614	0.727	0.638	0.592	0.677	0.555	0.676
Summer		0.527	0.537	0.612	0.537	0.435	0.464	0.591	0.588
Winter	p_{0000}	0.69	0.663	0.711	0.732	0.708	0.764	0.784	0.716
Summer		0.698	0.703	0.69	0.719	0.81	0.771	0.671	0.647
Winter	p_{1111}	0.615	0.653	0.742	0.659	0.557	0.694	0.603	0.694
Summer		0.55	0.557	0.608	0.526	0.47	0.497	0.588	0.593

5.3 Third-Order Markov Chain

For the third-order Markov chain with two states, applied to the time series of the daily precipitation for the summer and winter seasons of 1961-1990, the transition probabilities have been calculated accordingly to the relation (3.4). The results for a few stations in Romania are presented in Table 1. There is a significant difference between the precipitation regime for southeast (Constanta) and northwest (Baia Mare) or west (Piclisa) of the territory. As concerning the other stations, the results obtained for Bucharest are similar to the ones in Bacau, both stations being representative for the extra-Carpathian area of Romania's territory.

The performances of the third-order Markov chain are better in winter than in summer for the wet sequence, p_{1111}, at all stations. The results for the dry sequence, p_{0000}, must be used locally.

5.4 Determining the Order of the Markov Chain

The order of the most appropriate Markov chain model to represent the series of daily precipitation is determined accordingly to the BIC criterion $(3.6)_2$ at each station in both summer and winter. The BIC criterion values have been calculated based on the transition gure 5), a first-order Markov chain is the best representation for the daily precipitation occurrence.

The results show that in summer the best model is the first-order Markov chain for all stations in Romania, accordingly to the convective nature of the precipitation and their low persistence. In winter, the daily precipitation occurrence is well simulated by a second-order Markov chain in the almost northwestern Romania and the mountain region. As for the rest of the territory (see Figure 5), a first-order Markov chain is the best representation for the daily precipitation occurrence.

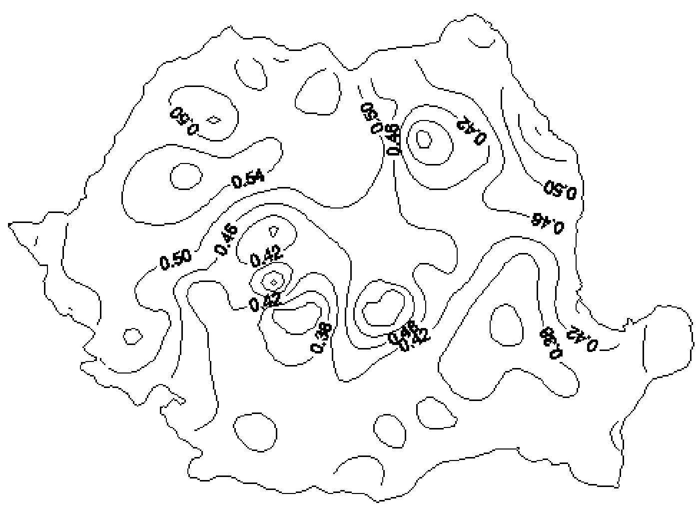

Figure 5: The order of the most appropriate Markov chain model for the daily winter precipitation occurrence (1961-1990)

6 Conclusions

The Markov chain approach for simulating daily precipitation occurrence points out significant differences between the precipitation regime in summer and winter. The results depend on the space domain too.

The occurrence probabilities of the wet sequences (2, 3 or 4 consecutive days with precipitation) are higher in winter than in summer at almost all the stations used in this study.The conditions of the occurrence of precipitation are less favourable in the extra-Carpathian area than in the rest of the territory.

A first-order Markov chain gives the most appropriate representation of the daily precipitation occurrence in the summer season. In winter, the occurrence of the daily precipitation is well simulated by a first-order Markov chain in the extra-Carpathian area and by a second-order one in the rest of Romania's territory. This result points out the influence of the shape and massiveness of the Carpathian on daily precipitation occurrence and their variability.

An interesting application of the Markov chain model for daily precipitation occurrence is in relation to the short-range forecasting of the precipitation probabilities. In this case, it could be used the transition probabilities calculated with the first-order or the second-order Markov chains to find the future distribution of the daily precipitation over Romania

References

[1] Gabriel K.R. and Neuman I., *Markov chain model for daily rainfall occurrence at Tel Aviv*, Q. J. Roy. Met. Soc, 88 (1962), 90-95.

[2] Hess G.D., Leslie L.M., Guymer A.E. and Fraedrich K., *Application of a Markov technique to the operational, short-term forecasting of rainfall*, Australian Meteorological Magazine, 37 (1989), 2, 83-91.

[3] Ladoy Ph., *Etude des probabilité s de transitions pour les types journaliers de distribution spatiale de la pluie.*, Actes du 3è me colloque sur l'analyse des données en géographie, Besançon 1974, 177-202.

[4] Mares Ileana, *A Markov chain for evaluation of Moon rainfalls* (in Romanian), Studii si Cercetari de Meteorologie , Part I/139, 4 (1976), 203-219.

[5] Mares Ileana and Mares C., *On extreme events in the precipitation field. Extreme Hydrological Events: Precipitation, Floods and Droughts*, International Symposium, Yokohama, Japan, Eds. Z.W.Kundzewicz, D.Rosbjerg, S.P.Simonovic and K.Takeuchi , 20-23 July 1993, 307-313.

[6] Wilks D.W., *Statistical methods in the atmospheric sciences*, Eds. Renata Dmowska and J.R.Holton, Academic Press, 1995.

[7] Yao S.C., *A statistical approach to historical record of flood and drought*, J. Appl. Met., 21 (4) (1982), 2-3 , 588-594.

Liana Cazacioc
National Institute of Meteorology and Hydrology,
Bucharest, Romania

Elena Corina Cipu
University Polytehnica of Bucharest,
Splaiul Independentei 313, RO-060042, Bucharest, Romania
e-mail: corinac@math.pub.ro, corinac@math.math.unibuc.ro