IDŐJÁRÁS

Markov chain analysis of the probability of days in a heat wave period

Árpád Fekete

University of Public Service
Faculty of Water Sciences, Department of Hydraulic Engineering
Bajcsy-Zs. utca 12-14, 6500 Baja, Hungary

*Corresponding Author e-mail: fekete.arpad@uni-nke.hu

Abstract—The impact of global climate change is also felt in Hungary. An undesirable effect of climate change (associated with rising temperatures) is mainly the increase in the frequency, intensity, and length of summer heat waves. In this respect, the region of the Southern Great Plain and the Danube-Tisza Region are particularly endangered, but the number of heat wave days has increased throughout the country in recent decades.

This study takes the climatic data sets of Baja into account, and on this basis, it gives the probability that a particular day falls into a heat wave period. Based on this information we draw a general conclusion about the long-term change of this climatic characteristic of the Southern Great Plain. The mathematical model used in the research applies the theory of Markov chains, which is relatively new in statistical analysis.

Key-words: heat wave, Markov chain, the matrix of transition probabilities, equilibrium distribution, return time of heat wave periods

1. Introduction

In the last century, the climate has warmed in Hungary, too. Compared to the last 100 years, the temperature in Hungary has increased by 1 °C, and a further 2.6 °C is expected by 2050 (*Láng et al.*, 2007). The increase in temperature is most intense in summer (*Bartholy et al.*, 2007). Rising temperatures are associated with heat waves. Heat waves as a hazard often have a negative effect causing heat stress in the man. Heat and drought events are of great importance in most climate regions. Heat waves also have a very detrimental effect on the development of plants, for example, maize does not tolerate heat stress above 35 °C, its development slows down or even stops (*Somfalvi-Tóth*, 2018). It is worth looking

at the yield of maize for the period 2010–2019 (*KSH*, 2019). Weaker yields in 2012, 2015, 2017 may correlate with the number of 31, 32, 33 heat wave days occurring in these years.

The annual number of heat waves was 0.5 in the period from 1961 to 1990 in Hungary (*Bartholy et al.*, 2013). The study in the previous reference predicts an average of 4-5 heat waves per year for the end of the century, i.e., the period from 2071 to 2100. This work also supports the previous estimate. We calculated an average of 3.3 heat waves per year at Baja in the period from 2010 to 2019. We will also show an increasing trend.

Serbian researchers (*Pecelj et al.*, 2019) have also found, among other things, that in addition to the increase in the frequency of heat waves, a growth trend is also detected in their duration.

In this study we take the climatic data sets of Baja into account, and on this basis, our goal is to give the probability that a particular day falls into a heat wave period. In a previous article we examined the annual precipitation sums of the last 30 years in the Baja region and the precipitation relations for six months belonging to the vegetation period (based on the SPI6 precipitation index) using Markov chains, as well (*Fekete* and *Keve*, 2020).

2. The definition of heat wave

It is difficult to determine the universal definition of a heat wave that could be applied in all climate zones. There have been numerous responses to the definition of a heat wave in recent years. However, there is no consistent, globally accepted definition; neither the World Meteorological Organization (WMO) nor the World Health Organization (WHO) (*Robinson*, 2001) have developed a uniform, globally acceptable definition of heat wave.

The national meteorological services have all dealt with the issue used to their own territory and climate, and accordingly, the meaning of the term heat wave is not consistent in the published studies either. Basically, a heat wave is an "unusually" warm period, but it means different things by continent, country, and even region. Therefore, it is difficult to find a threshold that suits all climates. The threshold can be the air temperature or an index that takes the interaction between the human body and its environment into account (*Robinson*, 2001). One way to create a threshold is to define a fixed value that shows the lower limit of the physiological heat wave. The conditions above the fixed limit established in this way lead to a decrease in comfort and increase in health risk. The disadvantage of this method is that it cannot be extended to a large area.

The second option is to examine the deviation from the local normal (expected value). Normal means daily averages or all observations (data so far). The degree of exceedance is a given value or a fixed standard deviation, or a fixed percentage. The use of this method is already advantageous in a larger area, as it takes the local conditions into account.

In Hungary, the heat wave is defined as a period of at least 3 consecutive days in which the average daily temperature exceeds 25 °C (National Institute of Environmental Health). This definition is based on an analysis of meteorological and mortality data between 1970 and 2000. (The choice of threshold is based on the expected value of the daily mortality data. The average daily temperature above 25 °C causes a 15% increase in mortality.) The definition created in this way is a fixed threshold definition, which is practical and easy to use in a small area (*Páldy* and *Bobvos*, 2008). The Hungarian Meteorological Service (OMSZ) defines the heat wave day similarly, i.e., when the average daily temperature exceeds 25 °C.

3. Theoretical background of Markov chains

We consider a discrete-time, discrete-space stochastic process, which we write as $X(t) = X_t$, for $t = 0, 1, \dots$. The state space S is discrete, i.e., finite or countable, so we can let it be a set of integers, as in $S = \{1, 2, \dots, k\}$ or $S = \{1, 2, \dots\}$.

The process $X(t) = X_0, X_1, X_2, \dots$ is a discrete-time Markov chain if it satisfies the *Markov property*:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = x_{n-1}, \dots X_1 = x_1, X_0 = x_0) = P(X_{n+1} = j | X_n = i).$$

Markov property means that the past and future are independent when the present is known. This means that if one knows the current state of the process, then no additional information of its past states is required to make the best possible prediction of its future.

The quantities $P(X_{n+1} = j | X_n = i)$ are called the *transition probabilities*. In general, the transition probabilities are functions of the initial state, end state, and time *(i, j, n)*. It is convenient to write them as

$$P_{ij}^{n,n+1} := P(X_{n+1} = j | X_n = i).$$

The Markov chain $X(t)$ is *time-homogeneous* if $P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i)$, i.e., the transition probabilities do not depend on time *n* (*Karlin* and *Taylor*, 1985). If this is the case, we write $P_{ij} = P(X_1 = j | X_0 = i)$ for the probability to go from *i* to *j* in one step, and $\mathbf{P} = (P_{ij})$ for the *transition probability matrix*. In detail

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & \dots & P_{0k} \\ P_{10} & P_{11} & \dots & P_{1k} \\ & \vdots & & \vdots \\ & \vdots & & \vdots \\ P_{k0} & P_{k1} & \dots & P_{kk} \end{bmatrix}.$$

The elements P_{ij} are non-negative numbers, and the sum of the elements on each row yields 1, so $\mathbf{P}$ is a stochastic matrix.

The *n-step transition probabilities* $P_{ij}^{(n)}$ are defined by

$$P_{ij}^{(n)} = P(X_n = j | X_0 = i).$$

Using the law of total probability and the Markov property, it is obvious that

$$P_{ij}^{(n)} = \mathbf{P}^n.$$

The transition matrix $\mathbf{P}$ of a chain is called *regular* if some power of $\mathbf{P}$ has no entry equal to zero. We then call the Markov chain itself *regular.*

We assume that at the start (at time 0) there is some initial distribution, that is, the distribution of X_0 is given. Let us denote it by φ_0. It is important that it is a row vector that gives the probabilities of being in each state. The Markov chain is determined completely by the transition matrix $\mathbf{P}$ and the initial distribution φ_0. The Kolmogorov equation gives that the distribution of X_n is $\varphi_n = \varphi_0 \cdot \mathbf{P}^n$.

The distribution φ_n approaches a limit as the number of steps approaches infinity, i.e., $\lim_{n \to \infty} \varphi_n = \pi$. This limit π is the *limiting (or invariant or stationary) distribution* of the Markov chain and it is a row vector: $\pi = (P_0, P_1, \dots, P_k)$, with $P_j > 0$ for all j. It is important that $\sum_{j=0}^{k} P_j = 1$, so that π is a probability distribution. If π is a limiting distribution, then it satisfies that $\lim_{n \to \infty} P_{ij}^{(n)} = P_j$ for all j=0,1,…,k. In other words

$$\mathbf{P}^* := \lim_{n \to \infty} \mathbf{P}^n = \begin{bmatrix} P_0 & P_1 \dots & P_k \\ P_0 & P_1 \dots & P_k \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ P_0 & P_1 \dots & P_k \end{bmatrix}$$

is the *limiting matrix.*

There are two methods to get the limiting distribution: $\mathbf{P}$ is exponentiated until the rows of the matrix are identical or applying the idempotence of $\mathbf{P}^* \mathbf{P} = \mathbf{P}^*$, it leads to a linear equation system:

$$\left. \begin{array}{c} P_j = \sum_{v=0}^{k} P_v \cdot P_{vj} \quad (j = 0, 1, \dots, k) \\ \sum_{j=0}^{k} P_j = 1 \end{array} \right\}.$$

The solution of the equation system gives the limiting distribution.

The Markov model is also applicable to predict the frequency of a heat wave of n days in a fixed period, in our case for the summer period. To realize this, we use the following formula (*Freidooni et al.*, 2015):

$$H_n = 1 + \frac{(N-n) \cdot P_{10} \cdot P_{01} \cdot P_{11}^{n-1}}{P_{10} + P_{01}}, \tag{1}$$

where H_n is the frequency of the heat wave of n days, N is the total number of days in the whole period, n is the number of heat wave days, and $\begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix}$ is the transition probability matrix. This equation will be applied in Section 5.

4. Presentation of heat wave periods

From the early 1980s, the OMSZ carried out a kind of categorization of heat waves to illustrate the occurrences of heat wave periods and their strength. The evaluation was made on the basis of national daily average temperatures using data from 1981–2016 (*OMSZ*, 2017). *Fig. 1* shows the following heat wave characteristics:

- duration of periods in which the daily average temperature exceeds 25 °C for at least 3 days;
- the highest daily average temperature during a heat wave;
- intensity (sum of temperatures above 25 °C), illustrated by the size of the circles.
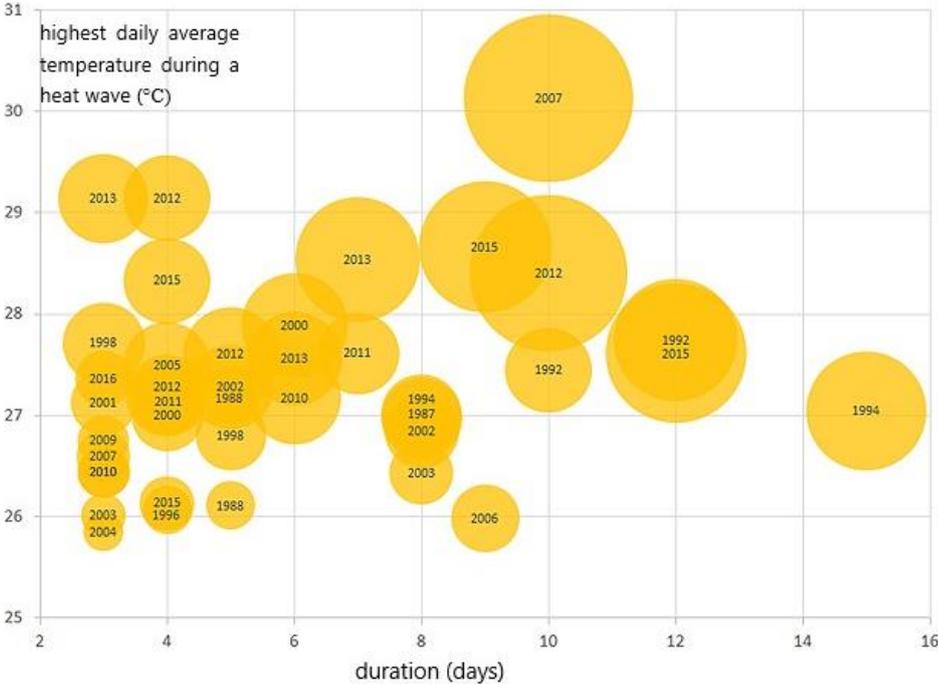
Fig. 1. Heat waves in the period 1981–2016 (*OMSZ*, 2017).

We can see that there was no heat wave until 1986, in the first six years of the observed period 1981-2016. The first heat wave was in 1987. The longest heat wave was in 1994, when the heat lasted for 15 days. However, the intensity of this heat period was not as significant as the heat wave in 2007 (July 15-24), although it lasted only 10 days. The national heat record was also detected: 41.9 ° C was measured in Kiskunhalas on 20 July, 2007. (*OMSZ*, 2017).

In recent years, several researchers have used Markov chains to study heat waves (*Yazdanpanah* and *Alizadeh*, 2000, *Freidooni et al.*, 2015). In our research, we examine the average daily temperatures of Baja in summer (June 1–August 31) for the period 2010–2019 and determine the heat wave periods occurring in each year from the data set. This is illustrated in *Table 1* below:

Table 1. The heat wave periods from 2010 to 2019 in Baja

Year	Number of periods	Number of days in the period	Date
2010	3	15	June 10–14, July 12–18, August 13–15
2011	2	13	July 8–15, August 23–27
2012	4	31	June 18–22, June 29–July 11(!), August 2–7, August 20–26
2013	3	17	June 17–22, July 27–29, August 2–9
2014	0	0	
2015	4	32	July 5–8, July 16–25, August 3-16(!), August 28–31
2016	3	10	June 23–26, July 11–13, July 23–25
2017	7	33	June 22–24, July 6–11, July 19-24, July 30–August 7, August 9–11, August 17–19, August 25–27
2018	3	21	July 29–31, August 5–14, August 17–24
2019	4	27	June 10–16, June 25–27, August 10–13, August 19–31(!)

It is worth mentioning, that 199 heat wave days were detected in the period 2010–2019, while 142 such days were numerated in the period 2000–2009.

5. Calculations

With the use of Markov chains we calculate the probability that a particular day falls into a heat wave period for each year (2010–2019) (*Freidooni et al.*, 2015). A matrix that shows the frequency at which the system transitions from each state to another state is called a *frequency matrix*. Let state 0 be the non-heat wave day, while let state 1 be the day in the heat wave period. The detailed calculation will be made for the year 2010. Based on this, we expect the probabilities for the other years, as well.

The transition frequency matrix is:

$$\begin{pmatrix} g_{00} & g_{01} \\ g_{10} & g_{11} \end{pmatrix} = \begin{pmatrix} 73 & 3 \\ 3 & 12 \end{pmatrix}.$$

Interpret the g_{01} element. Since there were 3 heat wave periods, it was three times that we switched from a non-heat wave day to a heat wave day and vice versa with respect to the element g_{10}. Interpret the g_{11} element. We have a total of 15 days in the heat wave period, so there are 12 transitions between the days in this state (the transitions start from the first days of the periods, so since there are 3 heat wave periods, there are 12 transitions). Studying the element g_{00}, we realize that there are 77 non-heat wave days in 4 periods (transitions start from the first days of the periods, so since there are 4 non-heat wave periods, there are 73 transitions). Using the transition frequency matrix we get the transition probability matrix:

$$\mathbf{P} = \begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix} = \begin{pmatrix} \dfrac{g_{00}}{g_{00} + g_{01}} & \dfrac{g_{01}}{g_{00} + g_{01}} \\ \dfrac{g_{10}}{g_{10} + g_{11}} & \dfrac{g_{11}}{g_{10} + g_{11}} \end{pmatrix} = \begin{pmatrix} \dfrac{73}{76} & \dfrac{3}{76} \\ \dfrac{3}{15} & \dfrac{12}{15} \end{pmatrix}.$$

We need to calculate the limiting matrix $\mathbf{P}^* = \begin{pmatrix} P_0 & P_1 \\ P_0 & P_1 \end{pmatrix}$. Applying the idempotence of $\mathbf{P}^*\mathbf{P} = \mathbf{P}^*$,

$$\begin{pmatrix} P_0 & P_1 \\ P_0 & P_1 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{73}{76} & \dfrac{3}{76} \\ \dfrac{3}{15} & \dfrac{12}{15} \end{pmatrix} = \begin{pmatrix} P_0 & P_1 \\ P_0 & P_1 \end{pmatrix}.$$

This leads to the solution of the following system of equations:

$$\left.\begin{array}{r}\frac{73}{76}P_0 + \frac{3}{15}P_1 = P_0 \\ \frac{3}{76}P_0 + \frac{12}{15}P_1 = P_0 \\ P_0 + P_1 = 1\end{array}\right\}.$$ (2)

We get that $P_0 = 0.835$ and $P_1 = 0.165$. Based on the data of the year 2010, there is a 16.5% chance of a heat wave day in the summer in the future.

The limiting distribution is calculated in a similar way for the next years (2011–2019). We do not detail the calculations, but the transition probability matrices and the limiting distributions are summarized in *Table 2*.

Table 2. Transition probability matrices and limiting distributions (2010-2019)

Year	P	(P_0, P_1)
2010	$\begin{pmatrix} 0,96 & 0,04 \\ 0,2 & 0,8 \end{pmatrix}$	(0,835, 0,165)
2011	$\begin{pmatrix} 0,97 & 0,03 \\ 0,15 & 0,85 \end{pmatrix}$	(0,83, 0,17)
2012	$\begin{pmatrix} 0,93 & 0,07 \\ 0,13 & 0,87 \end{pmatrix}$	(0,65, 0,35)
2013	$\begin{pmatrix} 0,96 & 0,04 \\ 0,18 & 0,82 \end{pmatrix}$	(0,82, 0,18)
2014	there was no heat wave in the summer	(1, 0)
2015	$\begin{pmatrix} 0,93 & 0,07 \\ 0,125 & 0,875 \end{pmatrix}$	(0,64, 0,36)
2016	$\begin{pmatrix} 0,96 & 0,04 \\ 0,3 & 0,7 \end{pmatrix}$	(0,88, 0,12)
2017	$\begin{pmatrix} 0,88 & 0,12 \\ 0,21 & 0,79 \end{pmatrix}$	(0,64, 0,36)
2018	$\begin{pmatrix} 0,96 & 0,04 \\ 0,14 & 0,86 \end{pmatrix}$	(0,78, 0,22)
2019	$\begin{pmatrix} 0,94 & 0,06 \\ 0,15 & 0,85 \end{pmatrix}$	(0,71, 0,29)

Fig. 2 shows the years with their associated P_1 limiting probabilities.
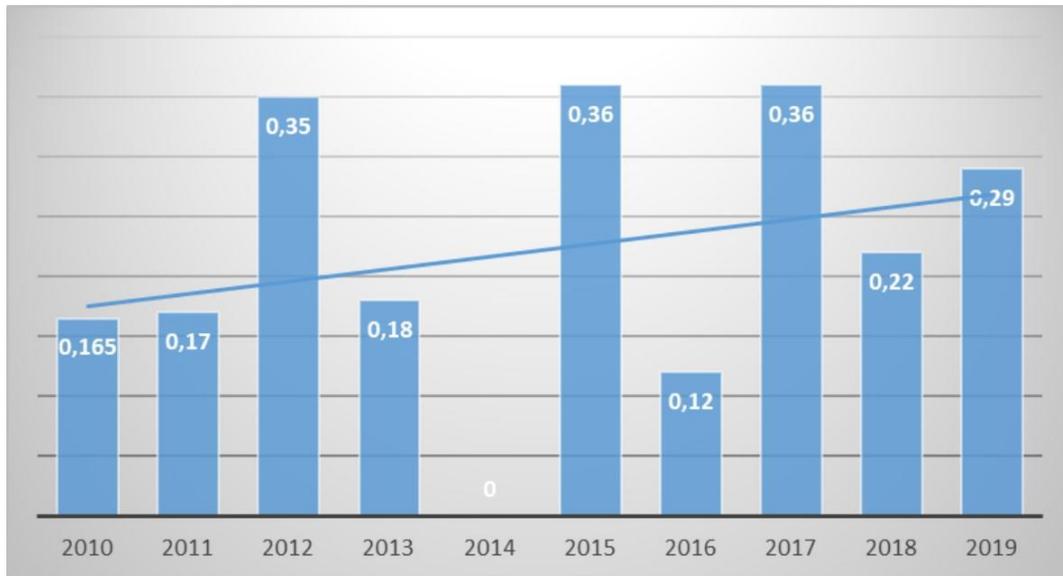
Fig. 2. Limiting probabilities of days in the heat wave period (2010–2019).

The arithmetic mean of the limiting probabilities in *Fig. 2* is 0.2215. It means that a day of the summer period will be a heat wave day with probability 0.2215 in the future. The return time of a heat wave day in summer is therefore 4.51 days (1/0.2215), which is quite worrying for the future.

Eq.(1). is used to calculate the frequency of heat waves of 3, 4,…12 days in future summers. We take the average transition matrix calculated from the 10-year-long data set based on *Table 2*.

$$\mathbf{P} = \begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix} = \begin{pmatrix} 0,936 & 0,064 \\ 0,222 & 0,778 \end{pmatrix}.$$

Then we substitute the numbers in Eq.(1). The results of our calculations are included in *Table 3*.

Table 3. Frequencies of heat wave periods of n days in the future during the summer season

n	3	4	5	6	7	8	9	10	11	12
frequency	3,67	3,06	2,58	2,21	1,93	1,71	1,55	1,42	1,32	1,25

The table shows that in summer, a 3-day heat wave can occur 3.67 times, a 7-day heat wave roughly twice, but a 12-day heat wave can occur once on average in the future.

We calculate not only the frequency but also the return time of a heat wave of n days with the theory of Markov chains. The following formula is used:

$$R_n = \frac{N}{H_n}, \tag{3}$$

where R_n is the return time of the heat wave of n days. *Table 4* shows the calculated return times.

Table 4. Return times of heat wave periods of n days in the future during the summer season

n	3	4	5	6	7	8	9	10	11	12
return time	25,07	30,07	35,65	41,63	47,67	53,8	59,35	64,79	69,7	73,6

Markov chains is also applicable to determine the persistent nature of a process, i.e., whether it is a trend-enhancing process. Persistence is usually measured within the R/S (rescaled range) analysis by the Hurst coefficient (H). It has a value between 0 and 1 and shows that a process is anti-persistent (H<0.5), persistent (H>0.5) or random (H = 0.5), behavior. Persistence in the heat wave study means that if the number of heat wave days increases in one period, the increase is expected to continue in the next period. Alijani (*Alijani*, 2011) used a simple formula to determine persistence related to the theory of Markov chains. If

$$r = P_{11} - P_{01}$$

is positive, then the process is persistent. If it is negative, then the process shows a lack of persistence. Calculating the value of r based on the transition probability matrices (*Table 2*), we get that $r>0$, so our process is persistent. The strength of persistence is not examined here now.

6. Conclusions

In summary, examining the data set of the last 10 years in Baja, the number of heat waves increased by 40% compared to the previous period 2000–2009. Based on the data of the last ten years, we calculated the arithmetic mean of the limit probabilities of the days in the heat wave period and the frequency and return time of the heat waves of n days (n = 1, 2,…, 12).

Based on the analysis of the data series for the period 1971-2019, the number of hot days, tropical nights, and heat wave days increased the most intensively. The length of the heat waves also increased. This also confirms the result of the Prudence project under the 5th Framework Program of the European Union. It states that the temperature in the Carpathian Basin clearly shows a warming trend (*Bartholy et al.*, 2007). The increasing frequency of extreme weather events needs to be given very important attention, because not only the elderly, the chronically ill, and the infants are at risk, but it can also cause a sensitive and then vulnerable condition in anyone.

In our study, we considered data sets from the last 10 years to examine heat waves. Obviously, we will get more accurate results if we work with a longer data set, but our calculations still supported (the increasing trend and the persistence process are proven) the claims of other climate researchers (*Bartholy et al.*, 2013) as we referred to it in the introduction.

Using our calculation method, it is worthwhile to calculate the limit probability of the days in the heat wave period for the future in other areas. Based on this, we could edit a heat wave forecast map.

References

Alijani, B., 2011: Evaluation of the Structure of Continuance of the Two Situations of Raining in South Iran by Means of Markov Hidden Status Model. *Geogr. Develop. J. 25*, 1–20.

Bartholy, J., Pongrácz, R., and *Gelybó, Gy*., 2007: Expected regional climate change in Hungary at the end of the 21st century. *Appl. Ecology Env. Res. 5.,* 1–17. https://doi.org/10.15666/aeer/0501_001017

Pongrácz, R. and *Bartholy, J*. (Eds.), 2013: Alkalmazott és városklimatológia. ELTE TTK, Budapest. (In Hungarian)

Fekete, Á. and *Keve, G*., 2020: csapadékösszegek és az aszályos időszakok vizsgálata Markov-láncokkal. *Hidrológiai Közlöny 100* (4), 60–70. (in Hungarian)

Freidooni F., Ataei, H., and *Shahriars, F*., 2015: Estimating the Occurence Probability of Heat Wave Periods Using the Markov Chain Model. *J. Sustain. Develop. 8* (2), 26–45. https://doi.org/10.5539/jsd.v8n2p26

KSH, 2019: Statistical Mirror. Budapest. http://www.ksh.hu/docs/hun/xftp/stattukor/fobbnoveny/2019/index.html

Karlin S. and *Taylor H.M.*, 1985: Sztochasztikus folyamatok. Gondolat Kiadó, Budapest, Hungary. (In Hungarian)

Láng, I., Csete, L., and *Jolánkai, M*., 2007: A globális klímaváltozás: hazai hatások és válaszok. *Agrokémia és Talajtan. 57.* 199–202.

OMSZ, 2017: Heat waves. https://www.met.hu/ismeret-tar/erdekessegek_tanulmanyok/index.php?id=1969&hir=Hohullamok:_ami_ma_szelsoseges,_az_a_jovoben_valoszinuleg_atlagos_lesz (In Hungarian)

Páldy, A. and *Bobvos, J*., 2008: A 2007. évi magyarországi hőhullámok egészségi hatásainak elemzése – előzmények és tapasztalatok. *„KLÍMA-21”, 52*, 3-15. (In Hungarian)

Robinson, J.P., 2001: On the Definition of a Heat Wave. *J. Appl. Meteorol. 40*, 762–775. https://doi.org/10.1175/1520-0450(2001)040<0762:OTDOAH>2.0.CO;2

Somfalvi-Tóth, K., 2017: Kukorica az éghajlatváltozás tükrében. *Agrofórum Extra* 72, pp. 8-10. https://agroforum.hu/ szakcikkek/novenytermesztes-szakcikkek/a-kukorica-az-eghajlatvaltozas-tukreben/ (In Hungarian)

Pecelj, M., Lukić, Z., Filipović, J., and *Protić, M.,* 2019: Summer variation of the UTCI index and Heat Waves in Serbia. *Nat. Hazard. Earth Syst. Sci. 270*, 1–19. https://doi.org/10.5194/nhess-2019-270

Yazdanpanah, H. and *Alizadeh, T.*, 2000: Estimate the probability of occurrence of heat waves and periods of continuity in the province with the help of Markov chains, *J. Geograph. Res.26(*3), 51–72.