Mathematics and Statistics 8(2A): 28-35, 2020 http://www.hrpub.org
DOI: 10.13189/ms.2020.081305

Markov Chain: First Step towards Heat Wave Analysis in Malaysia

Nur Hanim Mohd Salleh[1,*], Husna Hasan[1], Fariza Yunus[2]

[1]School of Mathematical Sciences, Universiti Sains Malaysia, Malaysia
[2]Malaysian Meteorological Department (Jabatan Meteorologi Malaysia), Malaysia

Received August 8, 2019; Revised October 8, 2019; Accepted February 20, 2020

Abstract Extreme temperature has been carried out around the world to provide awareness and proper opportunity for the societies to prepare necessary arrangements. In this present paper, the first order Markov chain model was applied to estimate the probability of extreme temperature based on the heat wave scales provided by the Malaysian Meteorological Department. In this study, the 24-year period (1994-2017) daily maximum temperature data for 17 meteorological stations in Malaysia was assigned to the four heat wave scales which are monitoring, alert level, heat wave and emergency. The analysis result indicated that most of the stations had three categories of heat wave scales. Only Chuping station had four categories while Bayan Lepas, Kuala Terengganu, Kota Bharu and Kota Kinabalu stations had two categories. The limiting probabilities obtained at each station showed a similar trend which the highest proportion of daily maximum temperature occurred in the scale of monitoring and followed by the alert level. This trend is apparent when the daily maximum temperature data revealed that Malaysia is experiencing two consecutive days of temperature below 35˚C.

Keywords Markov Chain, Maximum Temperature, Heat Wave, Transition Probability Matrix, Limiting Probabilities

1. Introduction

One of the most important climate parameters which affect natural and social phenomena is the temperature. The temperature that has exceeded the average temperature for a given area is considered as a heat wave. Recently, on February 2019, the Malaysian Meteorological Department (MMD) issued a level 1 alert for ten areas in the country amidst the ongoing nationwide heat wave. The ten areas are located in the western region of Peninsular Malaysia including Perlis, Kedah, Perak, Kuala Lumpur and Johor [1]. The level 1 alert was issued as the daily maximum temperature exceeding 35˚C for three consecutive days.

2. Background of Study

Stochastic processes are progressively being used for modeling and predicting climate data. The arising need of using this approach in climate and weather models is caused by the inability to resolve all necessary processes and scales in comprehensive numerical weather and climate prediction models [2]. One of the famous stochastic models applied to evaluate climate data is Markov chain. The Markov chain model is a commonly used tool for simulating time series of discrete random variables [3]. This model is considered as a memoryless process due to the probability of transitioning from one state to another depends only on the current state and not on the past. The detailed theory and structure of Markov chain can be referred to in some textbooks stochastic processes [4, 5]. Many researchers around the world have applied Markov chain model to analyze the climate data. In Malaysia, several similar analyses also have been attempted by the researchers to model the daily observation of rainfall [6], wind speed [7] and maximum temperature [8]. In the year 2015, Hasan [8] proposes the application of Markov chain onto daily maximum temperature data. The study covers only the northern region of Peninsula Malaysia. A scale for the state of transition was determined by using the physiological equivalent temperature (PET). The same scale was applied by Hassan and Hasan [9] in the year 2017 to determine the steady-state probability for the daily maximum temperature across Peninsular Malaysia. This study is the extension of these two works. The main purpose of the current work is to describe the heat wave condition in Malaysia by applying Markov chain model onto the daily maximum temperature data.

Figure 1. Location of 17 selected stations in Malaysia

3. Data Characteristics

The data used in this study are daily maximum temperature measured in degree Celsius (˚C) from 17 meteorological stations across Malaysia. These data are provided by Malaysian Meteorological Department (MMD) with less than 2% missing data. Each station has at least 24 years of daily data except KLIA station (19 years). The location of each station is illustrated in Figure 1. Fourteen stations are located in the west part of Malaysia (Peninsular Malaysia) while remaining three stations are located in the eastern region of Malaysia (Sabah and Sarawak).

4. Methodology

4.1. Heat Wave Scales

The daily maximum temperature model based on a Markov chain describes the temperature changes by following the heat wave scales. Four levels of heat wave scale provided by the Malaysian Meteorological Department are used in this study. The daily temperature data in this study is assigned to its heat wave scales. The descriptive terms for each scale together with the range of daily maximum temperature are listed in Table 1.

Table 1. The details of heat wave scales

Scale	Descriptive term	Range of daily maximum temperature (˚C)
1	Monitoring	below 35
2	Alert level	[35,37]
3	Heat wave	(37, 40]
4	Emergency	above 40

4.2. A Markov Chain Model for Daily Maximum Temperature

There are two properties need to be distinguished when applying Markov chain model to the climate data. The first property is the "state", defined as the number of different values that the variable can have and the second property is the "order", described as the number of previous values used to determine the state-to-state transition probabilities [3]. In this study, the Markov chain model applied on the daily maximum temperature is the first order model with four states which can be described as follows. Let X_t be defined as the state of temperature at t^{th} day where $X_t = i$, $i = 1,2,3,4$

1. if day t is at state of monitoring,
2. if day t is at state of alert level,
3. if day t is at state of heat wave,
4. if day t is at state of emergency.

Assuming that the state of temperature is dependent on the condition of the previous day, then X_t follows the first-order Markov chain. As an example, the below equations express the conditional probabilities of the temperature at state j on day t depending on the temperature at state i on day $t-1$. The transition probabilities estimated from the historic measurements, signify the probabilities of temperature from state i to state j. In general, the transition probability from state i into state j can be written as

$$p_{ij} = P[X_t = j \mid X_{t-1} = i], \quad i, j = 1,2,3,4$$

4.3. Development of Transition Probability Matrix

To obtain the transition probability matrix $\mathbf{P}$, the data with the previous state i and current state j are first counted. Then, the counted values, m_{ij} will be sorted and placed into a transition count matrix $\mathbf{M}$. Since there are four states of transition used in this study, the transition of daily maximum temperature can be divided into the 16 cases which can constitute the following count matrix $\mathbf{M}$. Define

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix}$$

which is then transformed into the transition probability matrix, $\mathbf{P}$, that is

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

where $\mathbf{P}_{ij} = \dfrac{m_{ij}}{\sum_j m_{ij}}$ and $\sum_{j=1}^{4} \mathbf{P}_{ij} = 1$

4.4. Development of Limiting State Probabilities

The n-step transition probabilities can be defined as follows:

$$\mathbf{P}^n = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}^n$$

As n increases, the n-step transition probabilities converge to certain probabilities:

$$\lim_{n\to\infty} \mathbf{P}^n = \lim_{n\to\infty} \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}^n = (\pi_1, \pi_2, \pi_3, \pi_4)$$

$$\sum_{1}^{4} \pi_j = 1$$

The probabilities that $\pi_1, \pi_2, \pi_3, \pi_4$ represent the mean occurrence probabilities of daily maximum temperature will be in each state, namely monitoring, alert level, heat wave and emergency state respectively after a sufficiently long time. These probabilities are also called the limiting or stationary state probabilities. The limiting state probability can be obtained by first, multiplying the transition probability matrix by itself. Then the resulting matrix is iterated until all the transition probability matrices reach the equilibrium [8].

4.5. Analyzing Markov Chain Using *R* Software

R is a free software environment used for statistical computing and graphics. *R* comes with a standard set of packages and other packages that can be downloaded and installed. These packages contain *R* functions, data and compiled code in a well-defined format. In this study, the package named 'markovchain' has been used to manage discrete time Markov chain more easily. It was developed and maintained by Giorgio Alfredo Spedicato and published on 21st January 2019.

An individual transition count matrix for each station can be created using *a createsequenceMatrix* function while the transition probability matrices can be computed using *markovchainFit* function. Using a *plot* function, a better understanding of the Markov Chain model can be gained through a transition diagram. This transition diagram is a graphical representation of a Markov chain which is equivalent to its transition probability matrix [10]. Furthermore, a steady-state distribution or a limiting state probability can also be computed using *steadyStates* function.

5. Result and Discussion

5.1. Descriptive Statistics

The descriptive statistics consisting of mean, maximum and minimum values (˚C) of the daily maximum temperature from the selected metrological stations are presented in Table 2. The mean values of the observed data range from 31.30˚C to 32.94˚C. The highest value of maximum temperature is recorded at Chuping (40.10˚C) followed by Alor Setar (39.10˚C) station. It is interesting to note that both stations are located in the northern region of Peninsular Malaysia.

Table 2. Descriptive statistics of daily maximum temperature

Station name	Mean	Maximum	Minimum
Chuping	32.87	40.10	23.70
Alor Setar	32.70	39.10	24.10
Bayan Lepas	31.69	36.00	25.10
Sitiawan	32.28	36.40	23.90
Subang	32.94	37.90	24.60
KLIA	32.11	37.40	24.20
Seremban	32.24	38.30	23.30
Melaka	32.13	38.00	24.40
Mersing	31.30	38.20	23.60
Senai	31.96	37.20	23.40
Muadzam Shah	32.50	37.70	23.30
Kuantan	31.99	37.80	23.20
Kuala Terengganu	31.39	35.80	23.80
Kota Bharu	31.30	36.40	23.80
Kuching	31.75	37.30	23.60
Labuan	31.31	36.60	25.00
Kota Kinabalu	32.17	36.50	24.30

5.2. Count and Transition Probability Matrices

The number of categories indicates the dimension of their respective count and probability matrices. To obtain the count matrix, the transition between two consecutive days is tallied and totaled. We notice that a large number of days are accumulated in cell m_{11} compared to the other cells. This common trend explains that all of the stations tend to experience two consecutive days at the state of monitoring. The second highest counted values are observed at cell m_{12} (the transition of temperature from state monitoring to alert level) for most of the station and at cell m_{22} (the temperature stays at alert level state for two consecutive days) for Chuping and Alor Setar stations. As explained in the methodology section, the transition probability matrix, presented in Figure 2, can be computed from the count matrix. The result evaluated from the transition probability matrix shows that the daily maximum temperature has the highest probability ($p > 0.900$) to remain at the state of monitoring for two consecutive days.

Furthermore, to obtain a better understanding of transition probability for each of the station, the transition diagram, as shown in Figure 3, is plotted. Having obtained both transition probability and transition diagram, the behavior of the daily maximum temperature can be assessed and the existence of heat wave at the studied location can be determined. Referring to the transition diagram, all states communicate with each other. Thus, this Markov chain model is irreducible. The transition of temperature from the monitoring or alert level to heat wave state is probably happening at Chuping, Alor Setar, Subang, KLIA, Seremban, Melaka, Mersing, Senai, Muadzam Shah, Kuantan and Kuching stations as their transition probability value, $p_{13} > 0.000$ or $p_{23} > 0.000$.

To be considered as heat wave, the temperatures have to be outside the historical averages for a given area for at least two or more days [11]. As mentioned above, the daily maximum temperature in Malaysia falls into heat wave category (Scale = 3) when the temperature is above 37˚C. Based on the transition probability and transition diagram, several stations located in the western region of Peninsular Malaysia experience two consecutive days at the state of heat wave with their transition probability value, $p_{33} > 0.000$. The stations are Chuping, Alor Setar, Seremban, Melaka and Mersing. This result means that the western part of Peninsular Malaysia is likely to experience heat wave compared to the other regions in Malaysia.

$$\begin{array}{cc} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} 0.960 & 0.040 & 0.001 & 0.000 \\ 0.333 & 0.619 & 0.048 & 0.000 \\ 0.032 & 0.272 & 0.690 & 0.006 \\ 0.000 & 0.000 & 1.000 & 0.000 \end{bmatrix} \end{array}$$

Chuping

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.966 & 0.332 & <0.001 \\ 0.333 & 0.616 & 0.051 \\ 0.034 & 0.449 & 0.517 \end{bmatrix} \end{array}$$

Alor Setar

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.998 & 0.002 \\ 0.938 & 0.063 \end{bmatrix} \end{array}$$

Bayan Lepas

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.993 & 0.007 \\ 0.703 & 0.297 \end{bmatrix} \end{array}$$

Sitiawan

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.990 & 0.010 & 0.000 \\ 0.604 & 0.389 & 0.007 \\ 0.000 & 1.000 & 0.000 \end{bmatrix} \end{array}$$

Subang

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.986 & 0.014 & <0.001 \\ 0.660 & 0.333 & 0.007 \\ 0.667 & 0.333 & 0.000 \end{bmatrix} \end{array}$$

KLIA

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.978 & 0.022 & <0.001 \\ 0.517 & 0.466 & 0.017 \\ 0.200 & 0.500 & 0.300 \end{bmatrix} \end{array}$$

Seremban

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.989 & 0.104 & <0.001 \\ 0.447 & 0.518 & 0.035 \\ 0.091 & 0.636 & 0.273 \end{bmatrix} \end{array}$$

Melaka

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.992 & 0.008 & 0.000 \\ 0.574 & 0.365 & 0.061 \\ 0.250 & 0.625 & 0.125 \end{bmatrix} \end{array}$$

Mersing

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.990 & 0.010 & 0.000 \\ 0.604 & 0.389 & 0.007 \\ 0.000 & 1.000 & 0.000 \end{bmatrix} \end{array}$$

Senai

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.949 & 0.051 & 0.000 \\ 0.602 & 0.388 & 0.010 \\ 0.286 & 0.714 & 0.000 \end{bmatrix} \end{array}$$

Muadzam shah

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.984 & 0.016 & 0.000 \\ 0.630 & 0.361 & 0.009 \\ 0.500 & 0.500 & 0.000 \end{bmatrix} \end{array}$$

Kuantan

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.998 & 0.002 \\ 0.400 & 0.600 \end{bmatrix} \end{array}$$

Kuala Terengganu

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.996 & 0.004 \\ 0.842 & 0.158 \end{bmatrix} \end{array}$$

Kota Bharu

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.988 & 0.012 & 0.000 \\ 0.667 & 0.327 & 0.007 \\ 0.000 & 1.000 & 0.000 \end{bmatrix} \end{array}$$

Kuching

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.998 & 0.002 \\ 0.655 & 0.345 \end{bmatrix} \end{array}$$

Labuan

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.993 & 0.007 \\ 0.616 & 0.384 \end{bmatrix} \end{array}$$

Kota Kinabalu

Figure 2. Transition probability matrix for 17 stations

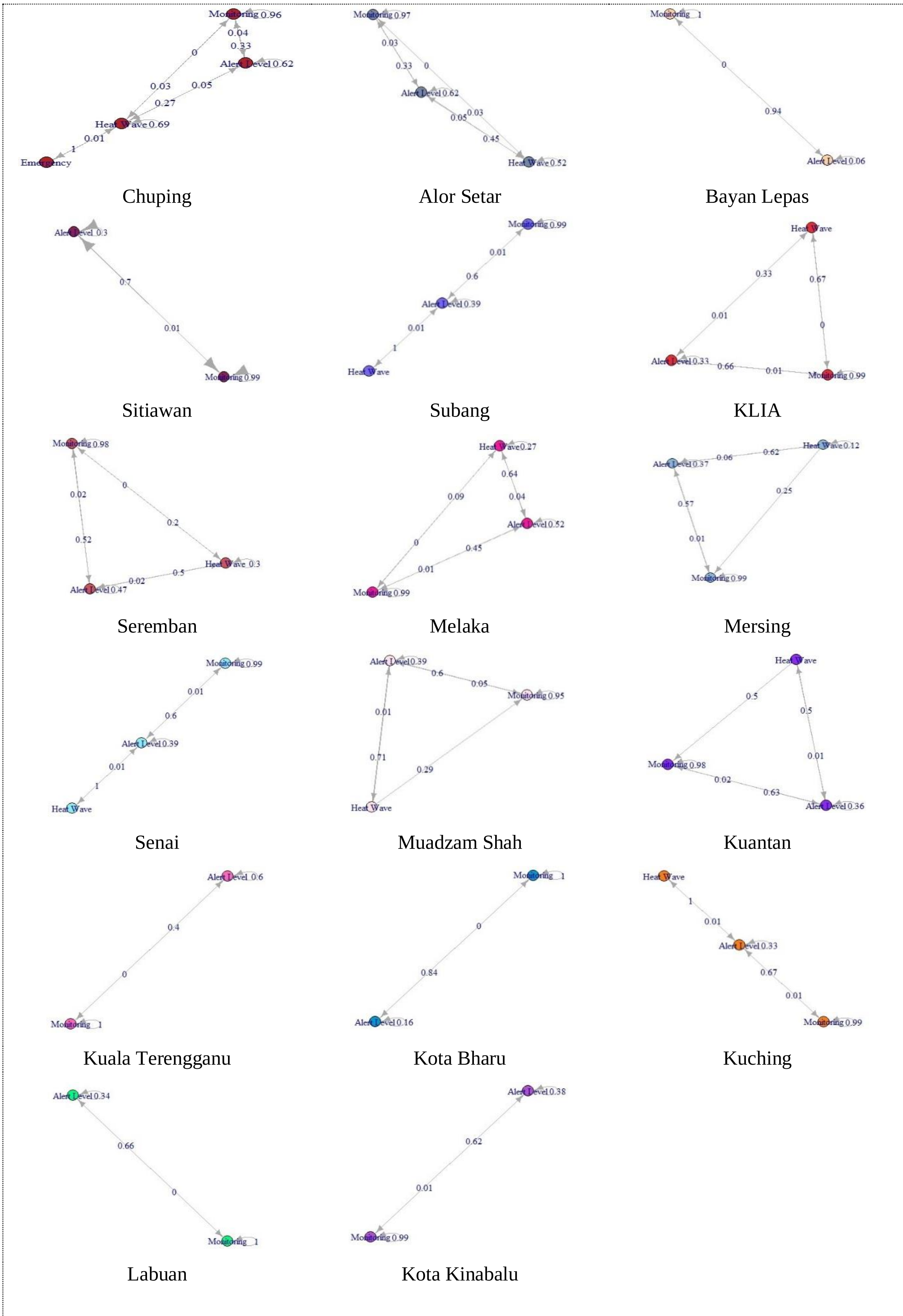

Figure 3. Transition diagram for each station

5.3. Limiting State Probabilities

The limiting state probabilities can be interpreted as the probabilities that remain unchanged in the Markov chain as time progresses. These probabilities are computed by multiplication of the transition probability with itself until the stationary distribution is attained. After a sufficiently long time, the limiting state probabilities are achieved when each row of the transition probability has the same probability value.

Table 3. Summary of stationary distribution.

Station Name	State of transition			
	1	2	3	4
Chuping	0.878	0.104	0.018	<0.001
Alor Setar	0.900	0.090	0.010	-
Bayan Lepas	0.998	0.002	-	-
Sitiawan	0.990	0.010	-	-
Subang	0.983	0.016	<0.001	-
KLIA	0.979	0.021	<0.001	-
Seremban	0.959	0.040	0.001	-
Melaka	0.976	0.023	0.001	-
Mersing	0.986	0.013	0.001	-
Senai	0.984	0.016	<0.001	-
Muadzam Shah	0.922	0.077	0.001	-
Kuantan	0.975	0.025	<0.001	-
Kuala Terengganu	0.995	0.005	-	-
Kota Bharu	0.996	0.004	-	-
Kuching	0.983	0.017	<0.001	-
Labuan	0.997	0.003	-	-
Kota Kinabalu	0.989	0.011	-	-

It can be understood from Table 3 that the highest mean occurrence probabilities of daily maximum temperature for all stations will be at monitoring state after a sufficiently long time. The proportion value of the occurrence at the state of monitoring is quite high ($\pi_1 > 0.800$) compared to the other states. Chuping, Alor Setar, Subang, KLIA, Seremban, Melaka, Mersing, Senai, Muadzam Shah, Kuantan and Kuching stations show that there are chances of daily temperature occurring at the state of heat wave. In contrast to the other stations, Chuping proves that there exists a small probability that the daily maximum temperature will be at the emergency state in the future.

6. Conclusions

In this paper, the behavior of daily maximum temperature data from 17 meteorological stations across Malaysia is assessed through Markov chain modeling. Previous studies have used physiological equivalent temperature (PET) scale to determine the state of transition. For the data analyzed here, the heat wave indications; monitoring, alert level, heat wave and emergency are used to categorize the data. The data classification result shows that the data can be categorized into four states of transition; monitoring, alert level, heat wave and emergency. The majority of the stations have three states of transition; monitoring, alert level and heat wave except for Bayan Lepas, Sitiawan, Kota Bharu, Kuala Terengganu, Labuan and Kota Kinabalu stations which have two states and Chuping possesses all four states.

It is not surprising to note that the highest transition probability value for each station is to remain at the state of monitoring for two consecutive days. Only five stations located in the western region of Peninsular Malaysia experience state of heat wave for two consecutive days. Investigation of the limiting state probabilities shows the decreasing trend from the state of monitoring towards the emergency state. After a sufficiently long time, all the stations have the highest proportion of time to experience monitoring state (temperature below 35˚C). This paper is limited to only investigate daily maximum temperature data. A more complete study of heat wave might consider other variables including daily minimum temperature or measure of heat stress.

Acknowledgements

We would like to thank the Malaysian Meteorological Department for providing the data.

REFERENCES

[1] H. K Kannan, Heat wave: MetMalaysia issues 'Level 1' alert for 10 areas nationwide", NewStraitsTime, Online available from https://www.nst.com.my/news/nation/2019/02/463630/heatwave-metmalaysia-issues-level-1-alert-10-areas-nationwide, last visit: 25.04.2019

[2] C. L. E. Franzke, T. J. O'kane, J Berner, P. D. Williams & V. Lucarini, Stochastic climate theory and modeling. WIREs Climate Change, Vol.6, 63–78, 2014.

[3] J. T. Scoof & S. C. Pryor, On the proper order of Markov chain model for daily precipitation occurrence in the contiguous United States, Journal of Applied Meteorology and Climatology, Vol.47, No.9, 2477–2486, 2008.

[4] D. R. Cox & H. D. Miller, The Theory of Stochastic Processes, Methuen, London, 1965.

[5] S. M. Ross, Introduction to Probability Models, Academic Press, New York, NY, 2010.

[6] S. Mohd Deni, Fitting optimum order of Markov chain models for daily rainfall occurrences in Peninsular Malaysia, Theoretical and Applied Climatology, Vol.97, No.1, 109-121, 2007.

[7] H. Hasan, A. Mohamad & N. H. Mohd Salleh, Application of Markov chain to wind speed in Northern Peninsular

Malaysia, Journal of Applied and Physical Sciences, Vol.3, No.2, 52-57, 2017.

[8] H. Hasan, M. A. Che Nordin & N. H. Mohd Salleh, Modeling Daily Maximum Temperature for Thermal Comfort in Northern Malaysia, Advances in Environmental Biology, Vol.9, No.926, 12-18, 2015.

[9] S. Hassan & H. Hasan, Determining the Steady-State Probability for the Daily Maximum Temperature in Peninsular Malaysia, ESTEEM Academic Journal, Vol.13(Special Issue), 129-138, 2017.

[10] F. Kachapova, Representing Markov Chains with Transition Diagrams, Journal of Mathematics and Statistics, Vol.9, No.3, 149-154, 2013.

[11] NOAA SciJinks, What Is a Heat Wave? NOAA SciJinks, Online available from https://scijinks.gov/heat/, last visit:25.04.2019