Applied Ecology and Environmental Sciences, 2022, Vol. 10, No. 8, 498-502
Available online at http://pubs.sciepub.com/aees/10/8/1
Published by Science and Education Publishing
DOI:10.12691/aees-10-8-1

Markov Chain Model for Observing Changing Behaviours of Air Quality Index

Sumithra P*, Loganathan A, Deneshkumar V

Department of Statistics, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, India
*Corresponding author: 350455@msuniv.ac.in

Received June 20, 2022; Revised July 25, 2022; Accepted August 01, 2022

Abstract The rising industrial revolution, urbanization and increasing human activities are the some of the primary reasons for increasing air pollution. Increasing air pollution adversely affects human health, agricultural production, wild animals. It is also an underlying cause for climate changes and these ultimately culminate in economic losses. Analysis of air quality index (AQI) can help to make recommendations to restrict outdoor air exposure and plan for demand on the healthcare system during periods of higher pollution. This study examines the stochastic behaviour of the AQI in Chennai city. Probabilities have been computed transitions among various states of AQI. Furthermore, steady state probability and mean first passage time have been evaluated.

Keywords: *Markov Chain, mean first passage time, transition probability matrix, AQI*

Cite This Article: Sumithra P, Loganathan A, and Deneshkumar V, "Markov Chain Model for Observing Changing Behaviours of Air Quality Index." *Applied Ecology and Environmental Sciences*, vol. 10, no. 8 (2022): 498-502. doi: 10.12691/aees-10-8-1.

1. Introduction

The presence of air pollutants in the atmosphere has been caused by increased human activities, urbanization, and industrialization. Increasing air pollution seeks much attention from researchers because planning and regulating air pollution requires a detailed analysis of the AQI as mentioned in Alyousifi *et. al.*, [1]. Air pollution is one of the significant issues affecting agricultural crop yield, human health and climate change. Zakaria *et. al.*, [2] stated that analysis of air pollution and identifying air pollution behaviour can help to provide information about the air quality, enabling individuals to take precautionary measures to avoid being exposed to unhealthy levels of air pollution.

Analysis of categorical data, and discrete time Markov chain (MC) are extensively used in many disciplines. Craig and Sendi [3] used a discrete time MC to model the transition behaviour of diseases and to monitor the HIV patient's health status by categorizing CD4 cell counts into three levels.

Several researchers have considered MC as a powerful model for analysing the probabilistic behaviour and underlying patterns in air pollution data. Sheskin [4] designed a new technique to compute the mean first passage time in MC. This technique used reduction of transition probability matrix into submatrix, and this procedure is useful in solving many numerical problems in a four-state MC. Oettl *et al.* [5] introduced a dispersion model for air pollution caused due to road traffic. Their model has been mostly based on gaussian equations. The results from the Markov Chain-Monte Carlo model reveal that this dispersion approach is useful and lead to reasonable consideration levels near motorways compared to observations. Alyousifi *et al.* [7] used discrete time MC model to study the stochastic behaviour of AQI in Klang. The mean first passage time and the mean return time have been evaluated to identify the risk of occurrence of unhealthy events. Wang *et al.* [8] applied fuzzy time series model for predicting urban air quality. They have developed deterministic prediction module, uncertainty analysis module, and assessment module. Zakaria *et al.* *[2]* developed a MC model to predict the AQI of Sarawak and evaluated the distribution of pollution in various levels. The initial state vector and transition probability matrix have been used for forecasting AQI. Chen and Jim Wu [9] proposed a discrete time MC for forecasting AQI. Ozone, nitrogen dioxide and particulate matter have been identified using the Markov model as prime pollutants in Taiwan. The prediction performance gives better results for forecasting short intervals of time than the long interval of time. Holmes and Hassini [10] predicted the air quality health index for Ontario using the discrete-time MC. The AQI has been divided into three states based on the risk factor. They found out that life of common public in Ontario is frequently high-risk state. Masseran and Safari [11] proposed a MC model to study the particulate matter (PM). The stochastic behaviour of PM10 data was studied using the first and higher-order MC. Their results proved that the Markov model is the best-fitted model to obtain information about the PM10 pollution index. Alyousifi *et al.* [12] applied the Fuzzy Markov chain model for AQI estimation. They also analysed the uncertainty in the occurrence of air pollution. These models can enable to manage the environmental risk. Alyousifi *et al.* [13] have performed a trend analysis on the air pollution index for 37 air monitoring stations in peninsular Malaysia based on a nonparametric test, the Mann-Kendall test. Moreover, the change point detection of the mean and maximum of AQI was studied using the Pettitt test. This study helped to analyse the air quality trends and their atmospheric aspects and was used to find a solution for some air pollution problems in Malaysia.

In this study, probabilistic behaviour of air pollution in Manali, an industrial region in the city of Chennai, India, is examined applying MC model. For each state of air pollution, the mean first passage time, stationary probability, and transition probability matrix are calculated. Following the introduction, Section 2 outlines the study area and data, Section 3 explains the methodology, Section 4 describes the results and discussion and finally Section 5 presents the concluding remarks.

2. Study Area and Data

Chennai is one of the India's major urban cities in Tamil Nadu, India. Manali is located in the north part of Chennai. This region has a variety of enterprises like oil refineries, fertilizer plants, chemicals, fabric yarn and steel *etc.*, including Chennai Petroleum Corporation Limited, Madras Fertilizers Limited, Tamil Nadu Petroproducts Ltd, Manali Petro Chemical and other industries are contributing significant pollution in this region [14]. Central Pollution Control Board (CPCB) of India proposed National Ambient Air Quality Standards for pollutant health break point consideration. The air quality sub-indices and health breakpoints have been classified into six categories namely *Good, Satisfactory, Moderate, Poor, Very Poor, and Severe* (https://app.cpcbccr.com/ccr_docs/About_AQI.pdf). Table 1 describes the breakpoint classification of the air quality index.

In this study, consider the Manali daily AQI data for the period from 01-12-2018 up to 23-03-2020 which have been collected from the CPCB. The data consists of 479 daily data.

Table 1. Classification of air quality index

State	AQI range	AQI Status
0	0-50	Good
1	51-100	Satisfactory
2	101-200	Moderate
3	201-300	Poor
4	301-400	Very Poor
5	401-500	Severe

3. Methodology

The sequence of event X_n is a random process, where $\{X_n\}$ is a discrete time MC, which takes on the finite state space and the process of state space are non-negative integers T= (0, 1, 2, …, N). The discrete time MC have been defined as

$$\begin{aligned} &P\left(X_{n+1}=j \mid X_n=i, X_{n-1}=i_{n-1}, \ldots, X_0=i_0\right) \\ &=P(X_{n+1}=j \mid X_n=i) \end{aligned} \quad (1)$$

For all $n, j, i, i_0, i_1, \ldots i_{n-1}$ in T, for all n= 0, 1, 2, …, N. A discrete time MC transition probability states do not depend on time, which can be written as

$$\begin{aligned} &P\left(X_{n+1}=j|X_n=i\right)=P\left(X_1=j|X_0=i\right) \\ &=p_{ij}, p_{ij} \geq 0 \ for \ \forall_{i,j} \ in \ T. \end{aligned} \quad (2)$$

where p_{ij} represent one step transition probability values from state i state j [15]. Markov process has stationary transition probabilities when one-step transition probabilities are independent of the variable. The probabilities that are the components of the transition probability matrix P_{p*p} of the MC, where p is the total number of states, represented by the values of p_{ij}.

$$\sum_{k=1}^{p} p_{ik}=1. \quad (3)$$

The (n + q) step state transition probability of the MC can be written as for any positive integers n and q.

$$p_{ij}^{(n+q)}=P(X_{n+q}=j \mid X_0=i) \sum_{k \in T} p_{ik}^n p_{kj}^q \quad (4)$$

Eq. (5) figures out the steady state distribution of a MC (π_j). This probability shows the percentage of time spent in a particular state *j*.

$$\pi_j=\sum_{i=1}^{T} \pi_j P_{ij} \quad (5)$$

$$\sum_{j \in T} \pi_j=1, \forall_{i,j} \in T \quad (6)$$

Find the limiting distribution, the mean first passage time between one state to another should be identified. If current state was start in *i*, the expected number of steps to return to state *i* for the first time is the mean return time for state *i*. denoted as m_{ii}.

$$m_{ii}=1+\sum_{k} p_{ik} m_{ik} \quad (7)$$

The mean first passage time from state *i* to state *j*, referred as m_{ij}, is the expected number of steps taken by a MC to arrive at state *j* after starting from state *I* [6].

$$m_{ij}=1+\sum_{k \neq j} p_{ik} m_{jk} \quad (8)$$

The discrete time MC model was used to explain the stochastic behaviour of the AQI, where Xn denotes the value of the AQI at time n and T = 1, 2, …, p, where p = 6. Moreover, it is possible to index the state space T = (0-50, 51-100, 101-200, 201-300, 301-400, 401-500) which represents all six states. For example, if the process is in state 2 at time n, then X_n = 2, if a process is in state 4 at time n, then X_n = 4. The maximum likelihood estimator was used to fit the transition probabilities of the MC model for p_{ij}. The current study conducted makes use of R software [16]

4. Results and Discussion

Air pollution $\{X_n\}$ behaviour can be described as a stochastic process, X_n denoting the AQI values of the air pollution state at time n with state space T= {0, 1, 2, 3, 4, 5}.

Table 2. Transition Frequency

States	Good	Satisfactory	Moderate	Poor	Very Poor	Severe
Good	9	10	2	0	0	0
Satisfactory	12	161	48	8	3	0
Moderate	0	57	80	13	4	1
Poor	0	4	16	9	7	0
Very Poor	0	1	7	5	3	1
Severe	0	0	1	1	0	0

Table 2 shows that most of the AQI values are between 51 - 100 which are satisfactory and less amounts are in the states for which the AQI values range 301-500. Only two times the AQI values reach 401 - 500 state from 101 - 200 which is moderate and 301 - 400 which is very poor.

Using Eq. (2) to fit the observed data, the transition probability matrix calculated as $P = p_{ij}$. As a result, the MC is ergodic given the entries in matrix P. If p > 0 indicates it is possible to move from state *i* to state *j*, it is reachable from state *i.*

Table 3 shows that in the estimation of the transition probabilities of the air quality index, there was no improvement in AQI values from the severe state. The probability of transition from the state of good to satisfactory and moderate are 48% and 9% and there was no transition to the remaining state. The probability of transition from satisfactory to good, moderate, poor and very poor are 5%, 7 %, 21%, 3% and 1 % respectively, and there was no transition to moderate. The probability of transition from moderate to satisfactory, moderate, poor and very poor are 37%, 52%, 8% and 2% respectively, and there was no transition to good and severe states. The probability of transition from poor to satisfactory, moderate, poor and very poor are 11%, 44%, 25% and 19% there was no transition to a severe state. The probability of transition from very poor to satisfactory, moderate, poor, very poor and severe are 6%, 41%, 29%, 17% and 5% and there was no transition to the good state. The probability of transition from severe to moderate and poor are 50% and 50%. Figure 1 displays the transition probability.

Table 3. Transition Probability Matrix

States	Good	Satisfactory	Moderate	Poor	Very Poor	Severe
Good	0.4286	0.4762	0.0952	0.0000	0.0000	0.000
Satisfactory	0.0517	0.6939	0.2069	0.0345	0.0129	0.000
Moderate	0.0000	0.3677	0.5161	0.0839	0.0258	0.0065
Poor	0.0000	0.1111	0.4444	0.2500	0.1944	0.0000
Very Poor	0.0000	0.0588	0.4118	0.2941	0.1765	0.0588
Severe	0.0000	0.0000	0.5000	0.5000	0.0000	0.0000

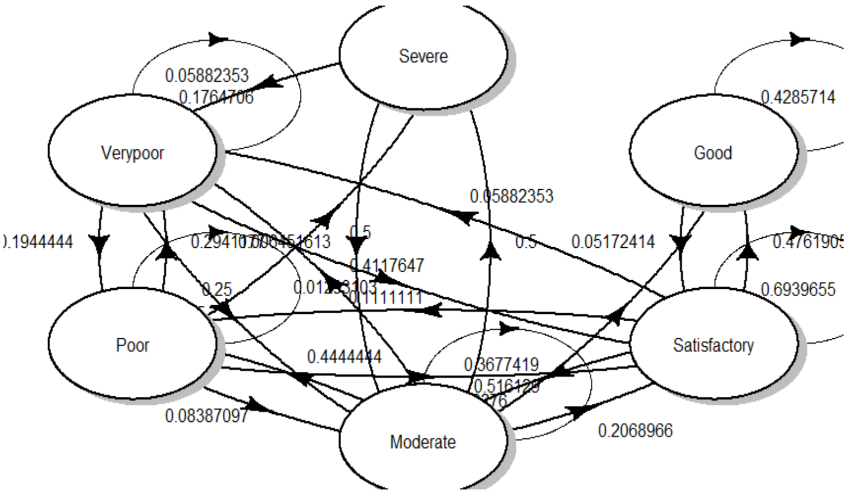

Figure 1. Transitions between states

Table 4. Steady state probability

$$p^{10} = \begin{pmatrix} 0.0464 & 0.5061 & 0.3302 & 0.0769 & 0.0363 & 0.0042 \\ 0.0459 & 0.5049 & 0.3310 & 0.0773 & 0.0365 & 0.0042 \\ 0.0455 & 0.5042 & 0.3317 & 0.0776 & 0.0367 & 0.0043 \\ 0.0450 & 0.5029 & 0.3327 & 0.0781 & 0.0369 & 0.0043 \\ 0.0449 & 0.5027 & 0.3329 & 0.0781 & 0.0370 & 0.0043 \\ 0.0449 & 0.5027 & 0.3328 & 0.0781 & 0.0369 & 0.0043 \end{pmatrix}$$

Table 5. Mean First passage time

States	Good	Satisfactory	Moderate	Poor	Very Poor	Severe
Good	0.0	2.2746	5.6131	20.1302	37.6818	237.5009
Satisfactory	36.0465	0.0	4.6357	18.6567	36.1757	236.2583
Moderate	39.1939	3.1475	0.0	16.9977	34.7122	233.2141
Poor	40.4656	4.4191	2.7159	0.0	27.2629	230.8301
Very Poor	40.7544	4.7079	2.6838	11.7242	0.0	217.1362
Severe	40.8298	4.7833	2.3579	9.4988	31.9876	0.0

Table 4 describes the steady state matrix, which explains that the value of AQI is at the state good (i.e., AQI value 0 -50), remains staying in the same state is 0.04, stays in a satisfactory state (i.e., AQI value 51-100) is 0.50, staying in the moderate state (i.e., AQI value 101 - 200) is 0.33 and staying in the poor state (i.e., AQI value 201 - 300) is 0.078 and staying in the very poor state (i.e., AQI value 301 - 400) is 0.036 and staying in the final severe state (i.e., AQI value 401 - 500) is 0.0043. The steady state was attained on the 10^{th} day of the transition step. From the steady-state matrix, the most prominent state, which accounts for 50.6% of the variability of the observed AQI occurrences is satisfactory. The second most significant percentage of the variability in the data of the AQI is explained by the moderate state, which can be calculated to measure nearly 33.3% of the total variability, and the remaining steady states are good, poor, very poor and severe states presented in smaller amounts.

The above Table 5 shows the mean first passage time of states. The average time taken to reach the severe state (i.e., AQI value is 401 - 500) for the first time from the good state (i.e., AQI value 0 -50) was 237 days. The average time taken to reach the very poor state (i.e., AQI value 301 - 400), poor state (i.e., AQI value 201 - 300), moderate state (i.e., AQI value 101 - 200), and satisfactory state (i.e., AQI value 51 - 100) for the first time from the good state was approximately 37.68, 20.13, 5.61 and 2.27 days respectively. The average time for the remaining states was calculated as before. The average time taken to return to the same state is zero. The average time for entering from the initial state to the final state was short compared to the time taken to reach the final state to the initial state. Additionally, the mean first passage time to passing very poor or severe states from any states was exceptionally long.

5. Conclusions

In this study, a discrete-time MC model has been constructed to study the probability behaviour of the AQI of Manali, Chennai. If the status of AQI at a time point in Manali is at either *Good or Satisfactory* state, then the status of AQI at immediate succeeding time point will be more likely *Satisfactory*. If the AQI at a time point is in *Moderate* state, then the AQI at immediate succeeding time point will also be more likely at *Moderate state*. If the AQI at a time point is at either *Poor or Very Po*or state, then the AQI at immediate succeeding time point will be more likely at either *Moderate* or at least *Poor* state. Similar transition behaviour is also observed, if the AQI is in *Severe* state. Moreover, the average time taken for the air quality in Manali to transit from *Good* state to *Severe* state is very long, approximately 237 days. The likelihood for experiencing *Severe* state of air quality in Manali is low.

References

[1] Alyousifi, Yousif, Kamarulzaman Ibrahim, Wei Kang, Wan Zawiah Wan Zin, Markov chain modeling for air pollution index based on maximum a posteriori method. Air Quality, Atmosphere & Health 12, no. 12 (2019): 1521-1531.

[2] Zakaria, Nurul Nnadiah, Mahmod Othman, Rajalingam Sokkalingam, Hanita Daud, Lazim Abdullah, and Evizal Abdul Kadir. Markov chain model development for forecasting air pollution index of miri, Sarawak. Sustainability 11, no. 19 (2019): 5190.

[3] Craig, Bruce A., and Peter P. Sendi. Estimation of the transition matrix of a discrete-time Markov chain. Health economics 11, no. 1 (2002): 33-42.

[4] Sheskin, Theodore J. Computing mean first passage times for a markov chain. International Journal of Mathematical Education in Science and Technology 26, no. 5 (1995): 729-735.

[5] Ottl, Dietmar, R. A. Almbauer, Peter-Johann Sturm, and Gerhard Pretterhofer. Dispersion modelling of air pollution caused by road traffic using a Markov Chain-Monte Carlo model. Stochastic Environmental Research and Risk Assessment 17, no. 1 (2003): 58-75.

[6] Alyousifi, Yousif, Nurulkamal Masseran, and Kamarulzaman Ibrahim. Modeling the stochastic dependence of air pollution index data. Stochastic environmental research and risk assessment 32, no. 6 (2018): 1603-1611.

[7] Zakaria, Nurul Nnadiah, Rajalingam Sokkalingam, Hanita Daud, and Mahmod Othman. Forecasting air pollution index in Klang by markov chain model. International Journal of Engineering and Advanced Technology 8, no. 6 Spec (2019): 635-639.

[8] Wang, Jianzhou, Hongmin Li, and Haiyan Lu. Application of a novel early warning system based on fuzzy time series in urban air quality forecasting in China. Applied Soft Computing 71 (2018): 783-799.

[9] Chen, Jeng-Chung, and Yenchun Jim Wu. Discrete-time Markov chain for prediction of air quality index. Journal of Ambient Intelligence and Humanized Computing (2020): 1-10.

[10] Holmes, Jason, and Sonia Hassini. Discrete-Time Markov Chain Modelling of the Ontario Air Quality Health Index. Water, Air, & Soil Pollution 232, no. 4 (2021): 1-13.

[11] Masseran, Nurulkamal, and Muhammad Aslam Mohd Safari. Modeling the transition behaviors of PM10 pollution index. Environmental monitoring and assessment 192, no. 7 (2020): 1-15.

[12] Alyousifi, Yousif, Ersin Kıral, Berna Uzun, and Kamarulzaman Ibrahim. New Application of Fuzzy Markov Chain Modeling for Air Pollution Index Estimation. Water, Air, & Soil Pollution 232, no. 7 (2021): 1-13.

[13] Alyousifi, Y., K. Ibrahim, W. Z. W. Zin, and U. Rathnayake. Trend analysis and change point detection of air pollution index in Malaysia. International Journal of Environmental Science and Technology (2021): 1-22.

[14] Jayanthi, V., and R. Krishnamoorthy. Key airborne pollutants - Impact on human health in Manali, Chennai. Current Science (2006): 405-413..

[15] Karlin, Samuel. A First Course in Stochastic Processes: Second Edition. Acadamic press. 2012.

[16] Spedicato, Giorgio Alfredo, and Mirko Signorelli. The markovchain Package: The R package "markovchain": Easily Handling Discrete Markov Chains in R. Cran,(Nicholson 2013) (2014).