Malaysia Statistics Conference 2016
15 November 2016
Sasana Kijang, Bank Negara Malaysia
Strengthening Statistical Usage for Decisions and Innovation

Universiti Kebangsaan Malaysia
The National University of Malaysia

MARKOV CHAIN MODLING FOR AIR POLLUTION INDEX IN MALAYSIA

Yousif Alyousifi, Nurulkamal Masseran, Kamarulzaman Ibrahim

Outline

- Introduction
- Study area and Dataset
- Objectives of the study
- Methodology
- Results and discussion
- Simulation
- Conclusion
- References

Introduction

- Air pollution may be described as contamination of the atmosphere by gaseous, liquid, solid wastes or by-products that can endanger life, attack materials and reduce visibility.
- Air pollution worldwide is a threat to human health and the natural environment.
- Air pollution can be caused due to the burning of wood, coal, oil, petrol, or by spraying pesticides.
- The air pollution index (API) is highly important for measuring the air quality in the environment.

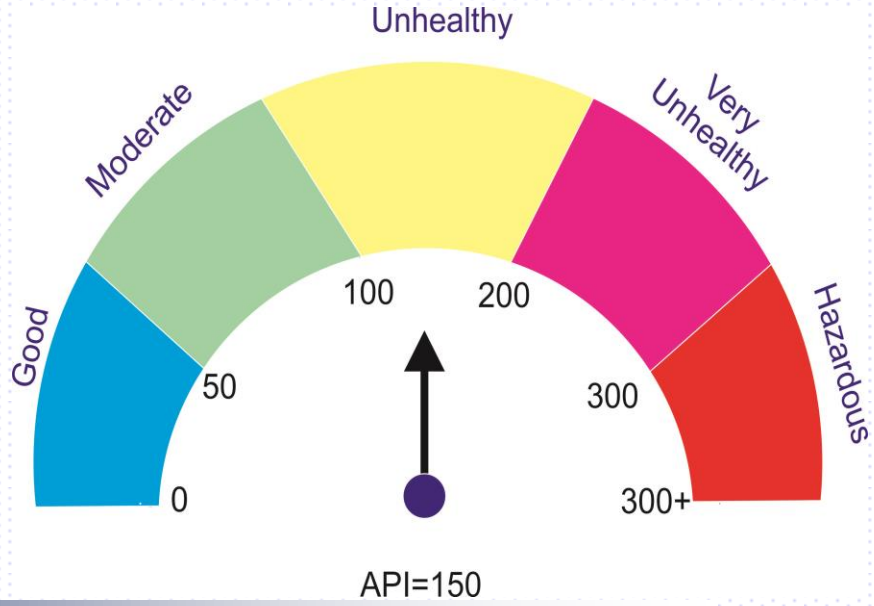
Introduction

- The API in Malaysia has been established by the Department of environment .
- The API is determined based on the highest average value of individual indices for all the variables, which include (SO2, NO2, CO, O3, and PM10) at a particular hour.
- A Markov chain is a random process where all information about the future is contained in the present state.

Study area and dataset

- A case study is conducted based on hourly API values from Klang City, Malaysia for the period of three years (2012 – 2014).

- The observed air pollution data that takes on values in the range from 0 to ∞ are classified into a five-state Markov chain. S={A, B, C, D, E}.



The Objectives of the Study

- This paper accomplishes three main objectives, as follows:

  - Provide the Markov chain model for describing the dependence behavior of API data.
  - Determine the long-run proportion of API data .
  - Determine the mean return time for each status of air pollution data.

- In summary, this research addresses the following research question:

Q1: Can Markov chain model be utilized to describe the stochastic behavior of air pollution?

Research Methodology

- In this study, we discussed DTMC model for describing the probabilistic behavior of air pollution states .

- The behavior of the air pollution can be classified as a stochastic process $\mathbf{X} = \{X_m, m = 0, 1, 2, \ldots, N\}$.

- A Markov chain is a model in which the value of X_m depends only on the previous value of X_{m-1}.

$$p_{ij} = P(X_{m+1} = j | X_m = i, X_{m-1} = i_{m-1}, \ldots, X_0 = i_0)$$
$$= P(X_{m+1} = j | X_m = i)$$

, for all $m, j, i, i_0, i_1, \cdots i_{m-1}$ in S , for all m =0,1,2,…,N.

Research Methodology

TRANSITION MATRIX AND DIAGRAM

Assuming the states are 1, 2,⋯, r, then the state transition matrix of a Markov chain can be written as following

$$\mathrm{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{bmatrix}$$

The quantities p_{ij} need to satisfy the conditions $p_{ij} \geq 0$, and for all i, j,

$$\sum_{k=1}^{r} p_{ik} = \sum_{k=1}^{r} P(X_{m+1} = k | X_m = i) = 1$$

Research Methodology

- **TRANSITION PROBALILITY**

Finding the probability of going from state *i* to state *j* in *n*-steps, i.e,

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i), \text{ for } n = 0,1,2,\cdots,$$

- **CLASSIFICATION OF STATES**

Accessible state ,Communication state, recurrent state ,irreducible state, periodic ,aperiodic and ergodic Markov chain .

- **LIMITING DISTRIBUTION**

$$\pi_j = \sum_{i=1}^{S} \pi_j P_{ij} \ , \qquad \sum_{j \in S} \pi_j = 1.$$

- **MEAN RETURN TIME**

$$r_j = 1/\pi_j \quad , \qquad m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{jk}$$

Results and discussion

- Descriptive statistics of observed API data

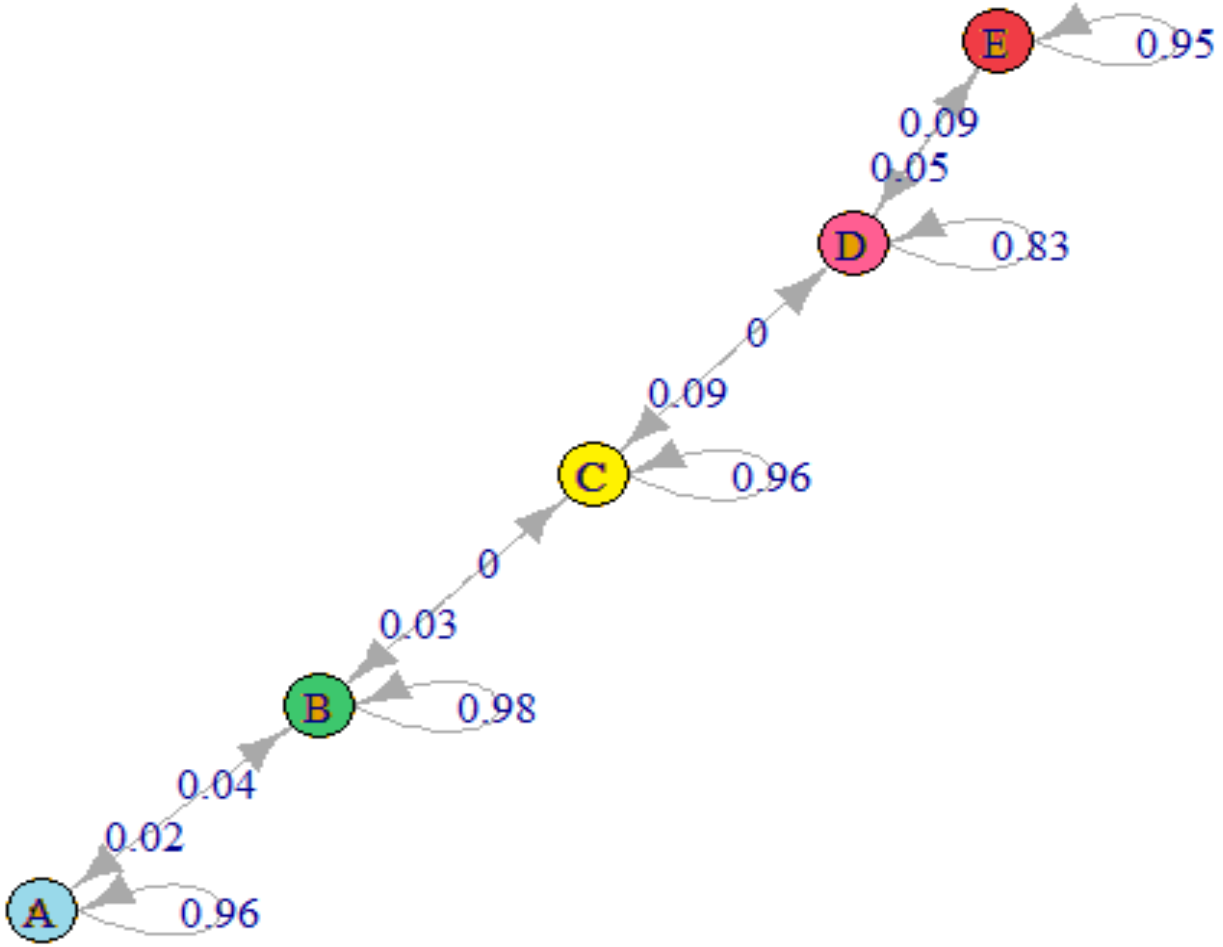
Mean	S.D	Min	Max	API<100	API>100	API>200	API>300
56.9	25.13	7	495	0.965176	0.0339872	0.0037256	0.002395

- Transition probability matrix of API values

States:	A	B	C	D	E
A	0.96383809	0.03616191	0.00000000	0.000000000	0.00000000
B	0.02239234	0.97588517	0.00172150	0.000000000	0.00000000
C	0.00000000	0.03391960	0.962311558	0.003768844	0.00000000
D	0.00000000	0.00000000	0.085714286	0.828571429	0.08571429
E	0.00000000	0.00000000	0.000000000	0.047619048	0.95238095

Results and discussion

- Transition Diagram of API Values



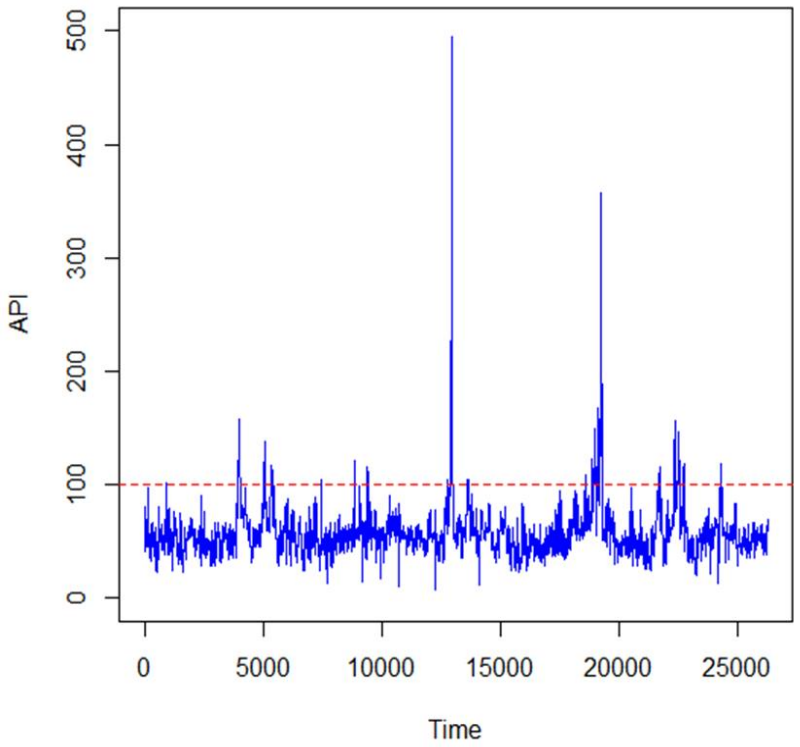
Healthy status

↔

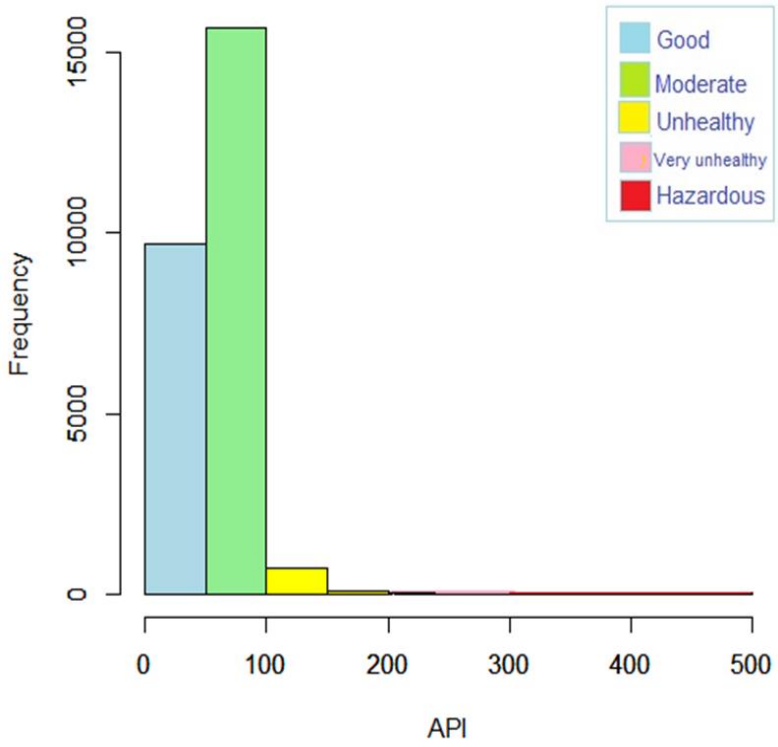
Unhealthy status

Results and discussion

- Time series plot for the data of API values



- Histogram of API values



Results and discussion

- The system of linear equations for **limiting distribution** is

Equations	Number of equation
$\pi_1 = 0.96383809\,\pi_1 + 0.02239234\,\pi_2$	(1)
$\pi_2 = 0.03616191\,\pi_1 + 0.97588517\pi_2 + 0.03391960\,\pi_3$	(2)
$\pi_3 = 0.00172150\,\pi_2 + 0.962311558\,\pi_3 + 0.085714286\,\pi_4$	(3)
$\pi_4 = 0.003768844\,\pi_3 + 0.828571429\,\pi_4 + 0.047619048\,\pi_5$	(4)
$\pi_5 = 0.08571429\,\pi_4 + 0.95238095\,\pi_5$	(5)
$\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1$	(6)

Results and discussion

- By solving the system of linear equations we obtained the steady states or the **limiting Distribution** of API data

$$\pi_j = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{bmatrix} = \begin{bmatrix} 0.36940890 \\ 0.59656680 \\ 0.03029456 \\ 0.00133204 \\ 0.00239768 \end{bmatrix}$$

- **The mean return time** for each states can be calculated by reciprocal the limiting distribution π_j. i.e

$$\boldsymbol{r} = \frac{1}{\pi_j} \text{ for all j} \in \text{S}$$

$$r_j = \begin{bmatrix} 2.707027 \\ 1.676258 \\ 33.00923 \\ 750.7242 \\ 417.0690 \end{bmatrix}$$

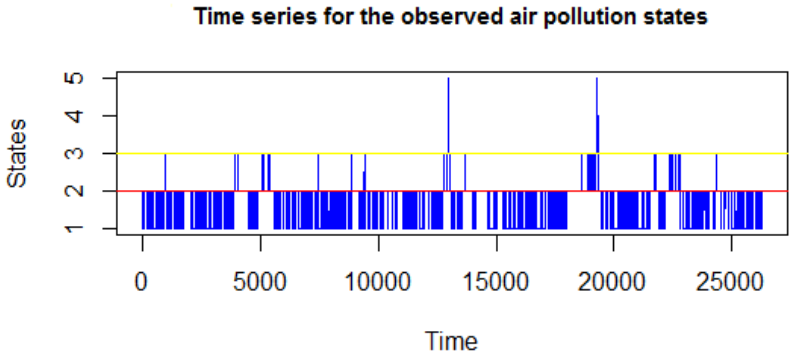
Results and discussion

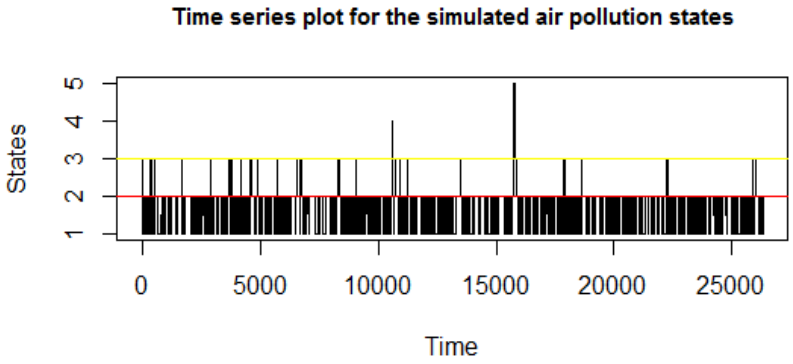
- **The mean return time from any state *i* to state *j*** (from the healthy states to unhealthy states or from unhealthy states to healthy states) is shown in the following matrix

States:	Good	Moderate	Unhealthy	Very unhealthy	Hazardous
Good	2.707027	27.65341	967.70320	9693.48650	18430.932
Moderate	47.20513	1.676258	940.04980	9665.83308	18403.279
Unhealthy	80.31623	67.28813	33.00923	8725.78317	17463.229
Very unhealthy	112.98289	65.77779	32.66665	750.7242	8737.447
Hazardous	133.98289	86.77779	53.66669	21.00041	417.0690

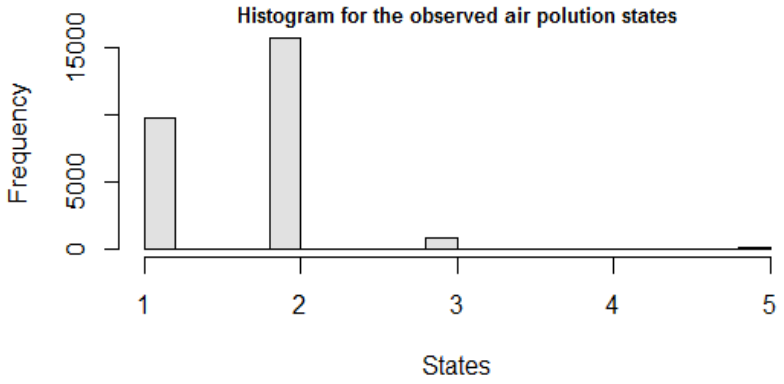
Simulation

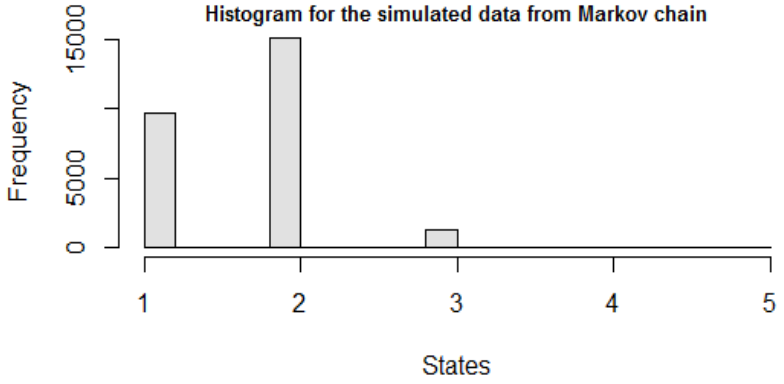
- Time series plot for observed and simulated data





- Histogram plot for observed and simulated data





Simulation

- Comparison of statistical properties between the observed states and simulated states

Statistical properties	Observed states	Simulated states
Mean	1.6700502	1.6655642
Standard deviation	0.5642186	0.5518739
Variance	0.3183426	0.3045648
Kurtosis	2.0327917	2.05871656
Skewness	0.4817832	0.4782365

Simulation

- Proportion and number of hours from the of observed and simulated states

Class of API	Observed states		Simulated states	
	Proportion	**number of hours**	**Proportion**	**number of hours**
API≤100	0.9660	25410	0.9745	25634
API>100	0.0339	894	0.0254	670
API>200	0.0037	98	0.0035	93
API>300	0.0023	63	0.0028	74

- Additionally, the coefficient of determination R^2 is found to be as 0.9567. Almost 95.6% of the data can be described by the Markov chain model.

Conclusion

- The sequence data of the air pollution state is suitable to be fitted with Markov chain model.
- There is a quite small risk of observing API values in Klang, although the risk remains troubling and thus should be considered.
- The value of R^2 is very large, which indicated that most of the observed data can be described by the Markov chain model.
- The results of this study indicate that the Markov chain model is very useful for describing the probabilistic behaviors of air pollution data.

Conclusion

- Markov chain model preserves a similar frequency between the observed data and the simulated data.
- The formulated Markov chain model had shown its flexibility to deal with API hourly data.
- The air quality standard in Klang lies within an acceptable limit and controllable condition.

References

1. Alley E. R., Cleland, W. L. and Stevens, L. B. (1998). Air Quality Control Handbook.New York, McGraw-Hill Professional,.
2. Robert, H. and Robert, K. (1999). Sources and Control of Air Pollution. New Jersey, Prentice-Hall.Inc.
3. Siew, L. Y., Chin, L. Y. & Wee, P. M. J.(2008). Arima and integrated arfima models for forecasting air pollution index in Shah Alam, Selangor. Malaysian Journal of Analytical Sciences 12(1): 257- 263.
4. Wong, T., Tam, W., Yu, I., Wong, A., Lau, A., Ng, S., Yeung, D. & Wong, C.(2012). A Study of the Air Pollution Index Reporting System. Final Report, Tender Ref. AP 07-085.
5. DOE. (2000). A Guide to Air Pollutant Index in Malaysia (API). Department of Environment. Kuala Lumpur, Malaysia: Ministry of Science, Technology and the Environment.
6. Masseran, N., Razali, A. M., Ibrahim, K. & Latif, M. T.(2016) . Modeling air quality in main cities of Peninsular Malaysia by using a generalized pareto model. Environmental monitoring and assessment 188(1): 1-12.
7. Rodrigues, E. R. and Achcar, J. A. (2012). Applications of Discrete-Time Markov Chains and Poisson Processes to Air Pollution Modeling and Studies. New York, Springer Science & Business Media.
8. Hoyos, L., Lara, P., Ortiz, E., Bracho, R. L., and González, J. (2010). Evaluation of air pollution control policies in mexico city using finite markov chain observation model. Revista de Matemática: Teoría y Aplicaciones 16(2), 255-266.
9. Hossien, P-N. (2014). Introduction to Probability, Statistics, and Random Processes. United States, Kappa Research.
10. Privault, N. (2013). Understanding Markov chains: examples and applications, Springer Science & Business Media, New York.
11. Grinstead, C. M., and Snell, J. L. (2012). Introduction to Probability. New York, American Mathematical Society.
12. Larsen, L., Bradley, R. & Honcoop, G. (1990). A new method of characterizing the variability of air quality-related indicators. *Air and Waste Management Association's InternationalSpecialty Conference of Tropospheric Ozone and the Environment. Los Angeles, CA,* hlm.
13. Masseran, N(2015). Markov chain model for the stochastic behaviors of wind-direction data. Energy Conversion and Management 92:266-274.
14. Pishro,H. & Nik (2014). Introduction to Probability, Statistics, and Random Processes .Kappa research ,LLC, United states.

Praise be to Allah

Thank you