Discrete-Time Markov Chain Modelling of the Ontario Air Quality Health Index

Jason Holmes · **Sonia Hassini**

Abstract The Air Quality Health Index (AQHI) is an aggregate indicator of air pollution used to communicate to Canadians the health impact of short-term exposure to current air pollutant levels. Understanding the stochastic behaviour of the AQHI can aid public health officials in predicting air pollution levels, determining the likelihood and duration of air quality advisories, and planning for increased strain on the health care system during periods of higher air pollution. Previous research has applied discrete-time Markov chains to investigate stochastic behaviour of air pollution indices but only in a handful of regions and none with the same climatic characteristics as Canadian regions. In this study, we investigated the stochastic behaviour of AQHI risk categories in Ontario (34 air monitoring stations) for 5 years from 2015 to 2019. We employed discrete-time Markov chains using three of the AQHI risk categories (Low Risk, Moderate Risk, High Risk) as states to determine (1) the transition probabilities between these states, (2) the long-run proportion of time spent in each state, and (3) the mean persistence time of each state. These results were then used to assess spatial trends in the stochastic behaviour of AQHI risk categories and the likelihood and duration of air quality advisories. Overall, the air quality (as characterised by the AQHI) in Ontario tends to decrease as population density increases. Urban areas spent a greater proportion of time in higher risk categories, and tended to remain in the higher risk categories for longer before transitioning.

1 Introduction

Air pollution is a global concern due to its detrimental impact on human health. Numerous studies have identified links between air pollution and increases in mortality and hospital admissions due to respiratory and cardiovascular disease (Brunekreef & Holgate, 2002). Predicting air pollution levels can help public health officials make recommendations to limit outdoor air exposure and plan for increased strain on the health care system during periods of higher air pollution.

The Air Quality Health Index (AQHI) was developed by Health Canada and Environment Canada to communicate the health impact of short-term exposure to current air pollutant levels to Canadians (Environment and Climate Change Canada, 2019). The AQHI is an aggregate indicator of the average trailing 3-h concentrations of three pollutants: ozone (O3), nitrogen dioxide (NO2), and fine particulate matter (PM$_{2.5}$). The AQHI is reported using 11 values ranging from 1 to 10+. A value of 1 represents the lowest health risk, while a value of 10+

J. Holmes · S. Hassini (✉)
Department of Civil Engineering, McMaster University,
Hamilton, ON L8S 4L7, Canada
e-mail: hassins@mcmaster.ca

J. Holmes
e-mail: holmej28@mcmaster.ca

represents a very high health risk. The AQHI values are further categorised according to their respective health risk; Low Risk (1–3), Moderate Risk (4–6), High Risk (7–10), and Very High Risk (10+). Each risk category has a recommended action for both the general population and the at-risk population (Environment and Climate Change Canada, 2015). For example, the Low Risk category recommends continuing all outdoor activities, while the Very High Risk category recommends avoiding all strenuous outdoor activities.

Ontario uses a modified form of the AQHI, where hourly concentrations of additional pollutants are also considered when assigning risk categories. In Ontario, ozone, nitrogen dioxide, sulphur dioxide, carbon monoxide, and total reduced sulphur compounds are the pollutants that may change the AQHI categories. For instance, if hourly air concentrations of one or more of the contaminants mentioned above exceed Ontario's Ambient Air Quality Criteria (AAQC), a desirable level of a contaminant in air, and the AQHI value is currently considered Low Risk or Moderate Risk, then the AQHI risk category is adjusted to a High or Very High Risk category (Ontario Ministry of the Environment, Conservation and Parks, 2019a).

The discrete-time Markov chain is a probabilistic model used to analyse stochastic processes. It has been employed in a wide range of applications such as modelling precipitation (da Silva et al., 2019; Schoof & Pryor, 2008), infrastructure deterioration (Baik et al., 2006), and wind speed (Sahin & Sen, 2001). Discrete-time Markov chains have also been employed to model air pollution (e.g., Asadollahfardi et al., 2016; Caraka et al., 2019; Mohamad et al., 2018; Romanof, 1982). Nebenzal and Fishbain (2018) found that, for forecasting long-term NO2 pollution, Markov chain models reduce the total error compared to other forecasting methods (e.g., multiple linear regression, moving average, exponential smoothing, Holt, and persistence methods).

Most air pollution studies that used Markov models focused on modelling specific contaminant concentrations such as particulate matter (Asadollahfardi et al., 2016; Caraka et al., 2019; Mohamad et al., 2018), nitrogen dioxide (Nebenzal & Fishbain, 2018), ozone (Rodrigues et al., 2019), and sulphur dioxide (Romanof, 1982), while few studies have focused on directly modelling an air pollution index similar to the AQHI (Alyousifi et al., 2018; Alyousifi et al., 2019; Zakaria et al., 2019). From a public health standpoint, directly

modelling the AQHI is more informative since the AQHI categories are directly related to health risks and outdoor air quality advisories.

This study aims to provide further insight into the stochastic behaviour of AQHI risk categories measured in Ontario, which has different climate conditions than those in the other studies (i.e., Alyousifi et al., 2018; Alyousifi et al., 2019; Zakaria et al., 2019). A discrete-time Markov chain was employed using three of the AQHI risk categories (Low Risk, Moderate Risk, High Risk) as states to determine (1) the transition probabilities between these states, (2) the long-run proportion of time spent in each state, and (3) the mean persistence time of each state. These results were then used to analyse spatial trends in the AQHI risk categories that could help public health officials understand the characteristics of air quality advisories, including their likelihood and duration.

2 Methodology

2.1 Study Area

Ontario is a province located in Central Canada. It is Canada's most populous province and second-largest by total land area. There are thirty-nine air monitoring stations in Ontario operated by the Ontario Ministry of the Environment, Conservation and Parks (MECP). Most of the thirty-nine air monitoring stations are located in populated areas in Southern Ontario, with the highest number in the Golden Horseshoe area around Toronto. Five of the thirty-nine monitoring stations were excluded from the study due to insufficient data. The locations of the selected air stations for this study are shown in Fig. 1. The characteristics of these stations are illustrated in Table 1.

2.2 Data

Historical Ontario AQHI data is available in hourly increments for a period of 24 h from the Ontario Ministry of the Environment, Conservation & Parks (MECP) website (Ontario Ministry of the Environment, Conservation and Parks, 2019b). Most air monitoring stations had AQHI data available for the period from January 1, 2015 to the current date, despite Ontario replacing their former Air Quality Index with the AQHI on June 24, 2015.

Fig. 1 Locations of air monitoring stations in Ontario, Canada
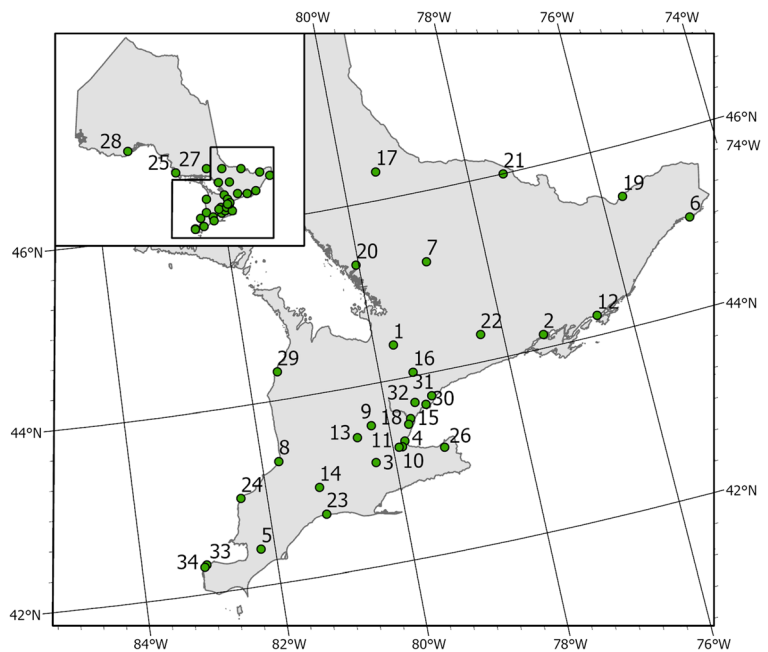
The AQHI is an aggregate indicator of the average trailing 3-h concentrations of three pollutants: ozone (O_3), nitrogen dioxide (NO_2), and fine particulate matter ($PM_{2.5}$). The AQHI is calculated according to the following formula: $AQHI = \frac{1000}{10.4} \times (e^{0.000537 \times O_3} - 1 + e^{0.000871 \times NO_2} - 1 + e^{0.000487 \times PM_{2.5}} - 1)$ (Szyszkowicz, 2019). Ozone and nitrogen dioxide concentrations are inputted as parts per billion, and fine particulate matter as $\mu g/m^3$. The AQHI is then rounded to the nearest integer. Additionally, if hourly air concentrations of ozone, nitrogen dioxide, sulphur dioxide, carbon monoxide, or total reduced sulphur compounds exceed Ontario's Ambient Air Quality Criteria (AAQC), and the AQHI value is currently considered Low Risk or Moderate Risk, then the AQHI value is adjusted to a High or Very High Risk category (Ontario Ministry of the Environment, Conservation and Parks, 2019a).

A Python script was developed to compile the data from the MECP website by automating the process of downloading the hourly data for each 24 h. This study used the available data for each air monitoring station from January 1, 2015 to December 31, 2019. Five of the thirty-nine air monitoring stations were excluded from the study due to insufficient data (missing more than a year's worth of data). The total number of observations in the historical dataset for the air monitoring stations during the study period was 1,463,652 h. This amount

represents an overall data coverage of 98.3% (i.e. 1.7% of the data is missing). The data coverage for each air monitoring station is included in Table 1.

2.3 Discrete-time Markov Chain Model

Markov chains are probabilistic models used to analyse stochastic processes. The choice of discrete-time or continuous-time Markov chain depends on the analysed time series. In this application, the discrete-time Markov chain is used since the time series consists of discrete hourly increments. Discrete-time Markov chains are characterised by the number of states in the state space and the number of previous states the transition probability is dependent on.

Consider a stochastic process $\{X_n, \ n = 1, 2, 3, \ldots\}$ that takes on a finite or countable number of possible values defined as the state space, S. If $X_n = i$, then the process is said to be in state i at time n. The stochastic process is a Markov chain if the conditional distribution of any future state is independent of the past states and depends only on the present state. This condition is known as the Markov or memoryless property and can be expressed as $P\{X_{n+1} = j \mid X_n = i, \ X_{n-1} = i_{n-1}, \ \ldots, X_2 = i_2, X_1 = i_1\} = P\{X_{n+1} = j \mid X_n = i,\} = p_{ij}$ for all states $i_1, \ i_2, \ \ldots, i_{n-1}, \ i, \ j$ and all $n \geq 0$. The value p_{ij} is the one-step transition probability and represents the probability of transitioning from state i to state j. p_{ij}^n is used to represent the elements of the nth power of the transition

Table 1 Ontario AQHI air monitoring station details

Number	Name	Latitude	Longitude	Station type	Elevation ASL (m)	Height of air intake (m)	No. of observations	Data coverage
1	Barrie	44.382361	− 79.702306	Urban	226	5	42967	98.2%
2	Belleville	44.150528	− 77.395500	Urban	75	10	42600	97.3%
3	Brantford	43.138611	− 80.292639	Urban	205	5	43332	99.0%
4	Burlington	43.315111	− 79.802639	Urban	78	5	43387	99.1%
5	Chatham	42.403694	− 82.208306	Urban	179	15	42778	97.7%
6	Cornwall	45.017972	-74.735222	Urban	55	4	43340	99.0%
7	Dorset	45.224278	− 78.932944	Rural	318	3	43022	98.3%
8	Grand Bend	43.333083	− 81.742889	Rural	185	5	42804	97.8%
9	Guelph	43.551611	− 80.264167	Urban	330	4	43109	98.5%
10	Hamilton Downtown	43.257778	− 79.861667	Urban	90	4	43155	98.6%
11	Hamilton West	43.257444	− 79.907750	Urban	96	3	42919	98.0%
12	Kingston	44.219722	− 76.521111	Urban	84	5	43267	98.8%
13	Kitchener	43.443833	− 80.503806	Urban	325	5	42768	97.7%
14	London	42.974460	− 81.200858	Urban	244	5	43309	98.9%
15	Mississauga	43.546970	− 79.658690	Urban	105	5	43182	98.6%
16	Newmarket	44.044306	− 79.483250	Urban	268	5	43507	99.4%
17	North Bay	46.322500	− 79.449444	Urban	212	4	43244	98.8%
18	Oakville	43.486917	− 79.702278	Urban	165	12	42844	97.9%
19	Ottawa Downtown	45.434333	− 75.676000	Urban	68	4	43054	98.4%
20	Parry Sound	45.338261	− 80.039269	Urban	176	5	42988	98.2%
21	Petawawa	45.996722	− 77.441194	Rural	174	6	43229	98.8%
22	Peterborough	44.301917	− 78.346222	Urban	226	10	43155	98.6%
23	Port Stanley	42.672083	− 81.162889	Rural	212	5	42443	97.0%
24	Sarnia	42.990263	− 82.395341	Urban	182	5	43203	98.7%
25	Sault Ste Marie	46.533194	− 84.309917	Urban	244	8	42794	97.8%
26	St Catherines	43.160056	− 79.234750	Urban	105	4	42964	98.1%
27	Sudbury	46.491940	-81.003105	Urban	271	5	43032	98.3%
28	Thunder Bay	48.379389	− 89.290167	Urban	192	15	42952	98.1%
29	Tiverton	44.314472	− 81.549722	Rural	226	4	42818	97.8%
30	Toronto Downtown	43.662972	-79.388111	Urban	105	10	42624	97.4%
31	Toronto East	43.747917	− 79.274056	Urban	168	4	43289	98.9%
32	Toronto West	43.709444	− 79.543500	Urban	141	8	43165	98.6%
33	Windsor Downtown	42.315778	− 83.043667	Urban	176	4	43166	98.6%
34	Windsor West	42.292889	− 83.073139	Urban	180	4	43242	98.8%

probability matrix P^n. The transition probability matrix, P, for a Markov chain with k states is composed of $k \times k$ one-step transition probabilities, p_{ij}, where $0 \leq p_{ij} \leq 1$ and the sum of the probabilities in each row is equal to 1 ($\sum_{j=1}^{k} p_{ij} = 1$).

Each element, p_{ij}, of the transition probability matrix was calculated for a given state i, where n_{ij} is the observed frequency of one-step transitions in the historical data from state i to state j, using:

$$p_{ij} = \frac{n_{ij}}{\sum_{j=1}^{k} n_{ij}} \qquad (1)$$

The validity of fitting a Markov chain k states to observed data can be investigated using the Chi-square (χ^2) test statistic

$$\chi^2_{\text{calc}} = \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{\left(n_{ij} - e_{ij}\right)^2}{e_{ij}} \tag{2}$$

where n_{ij} represents the observed transition frequency, and e_{ij} represents the expected transition frequency (Wilks, 2011). The null hypothesis is that the observed data is serially independent. The alternative hypothesis is that the observed data was generated by a Markov chain. Under the null hypothesis of independence, e_{ij} is calculated as follows:

$$e_{ij} = \frac{\left(\sum_{i=1}^{k} n_{ij}\right)\left(\sum_{j=1}^{k} n_{ij}\right)}{\left(\sum_{i=1}^{k} \sum_{j=1}^{k} n_{ij}\right)} \tag{3}$$

For a Markov chain with k states, the test statistic follows the χ^2 distribution with $(k-1)^2$ degrees of freedom.

2.3.1 Stationary Distribution

The stationary distribution shows the long-term proportion of time each air monitoring station spends in a specific AQHI risk category. The estimated time can be used to evaluate the air quality at each air monitoring station by comparing the time spent in the different AQHI risk categories. For example, a longer portion of time spent in the Low Risk category would indicate better air quality.

The stationary distribution of a Markov chain refers to the long-run probability distribution that remains unchanged as time progresses. A finite-state Markov chain that is ergodic will have each row of the limiting distribution ($\lim_{n \to \infty} \boldsymbol{P}^n$) converge to the stationary distribution. Let π_j represent the long-run proportion of time spent in a state j. If the finite-state Markov chain is ergodic, the stationary distribution is unique, and π_j can be calculated with the following equations (Ross, 2014):

$$\pi_j = \sum_{i=1}^{k} \pi_i p_{ij} \tag{4}$$

$$\sum_{j=1}^{k} \pi_j = 1 \tag{5}$$

A finite-state Markov chain is said to be ergodic if it is aperiodic and irreducible. A finite-state Markov chain is irreducible if all states communicate with each other. State i and state j are said to communicate if they are accessible from each other. State j is said to be accessible from state i if $p_{ij}^n > 0$ for some n time-steps. A Markov chain is said to be aperiodic if all of its states are aperiodic. A state is aperiodic if it is not periodic. A state is said to be periodic if the chain can return to the state only at multiples of some specific integer larger than 1.

2.3.2 Mean Persistence Time

The expected amount of time that, after the Markov chain enters an AQHI risk category, it remains in the same AQHI risk category before exiting can be referred to as the mean persistence time. Developing a model to estimate the mean persistence time of a specific AQHI risk category can be useful for public health officials to predict the length of air quality advisories. When public health officials communicate the health message associated with each AQHI risk category, additional information about the expected duration could be included as well. This information would allow the concerned population to better plan their activities to comply with the health message. For example, if, after entering the AQHI High Risk category, the mean persistence time is 8 h, then an air quality advisory to reduce or reschedule strenuous outdoor activities for at least 8 h can be issued before requiring a reassessment.

The transition probability matrix can be used to calculate the expected time it takes to enter any set of absorbing states from a transient state. A transient state is a non-recurrent state. It means that there is a non-zero probability that a Markovian process starting in a transient state will never return to that state. An absorbing state is defined as a state that cannot be transitioned out of after it is entered. A state i can be transformed into an absorbing state by setting $p_{ii} = 1$ and $p_{ij} = 0$ for all $j \neq i$. For an ergodic Markov chain with k states, after transforming b states into absorbing states, a new matrix with $t = k - b$ states can be defined as

$$\boldsymbol{P}_T = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1t} \\ p_{21} & p_{22} & \cdots & p_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ p_{t1} & p_{t2} & \cdots & p_{tt} \end{bmatrix} \tag{6}$$

where the elements of $\boldsymbol{P}_T$ are the reordered and renumbered one-step transition probabilities p_{ij} for all

non-absorbing states (Ross, 2014). Let s_{ij} be the expected amount of time before absorption that a Markov chain spends in state j, given it started in state i. A $t \times t$ matrix S composed of elements s_{ij} can be calculated using $S = (I - P_T)^{-1}$ where I is the $t \times t$ identity matrix (Ross, 2014).

By making all states, except for state i, an absorbing state, it is possible to calculate the mean persistence time of state i (the expected amount of time that the Markov chain will remain in state i before transitioning out of state i). For this case, s_{ii} can be calculated using:

$$s_{ii} = \frac{1}{1 - p_{ii}} \tag{7}$$

3 Results and Discussion

3.1 Discrete-Time Markov Chain Model

We used a discrete-time Markov chain to model the occurrence of AQHI risk categories at each air monitoring station as a stochastic process. We defined the state space as $S = \{1, 2, 3\}$, where states 1, 2, and 3 represent the Low Risk, Moderate Risk, and High Risk AQHI categories, respectively. The Very High Risk AQHI category was excluded from the study since only one occurrence was observed during the study period across all air monitoring stations (i.e. 1 out of 1,463,652 h). Additionally, the High Risk AQHI state was removed from the state space of the Markov chain of specific air monitoring stations if there were no observed occurrences during the study period.

A transition probability matrix was determined for each air monitoring station. The observed frequency of transitions is contained in Table 2. The transition probabilities were estimated using Eq. (1). Each element of the transition probability matrices is included in Table 3, and the spatial distribution is shown in Fig. 2. An example of the average transition probability matrix observed at air monitoring stations in Ontario is:

$$P = \begin{bmatrix} 0.990 & 0.010 & 0.000 \\ 0.134 & 0.865 & 0.002 \\ 0.003 & 0.443 & 0.554 \end{bmatrix}$$

The validity of fitting a Markov chain to the observed data for each individual air monitoring station was investigated using Eq. (2). The Chi-square test statistic, χ^2_{calc}, was calculated for each air monitoring station, and compared to the χ^2 distribution with $(k-1)^2$ degrees of freedom. For this analysis, the significance level was chosen to be 0.01. The calculated p value and degrees of freedom for each air monitoring station are contained in Table 2. The p value for the observed data from each air monitoring station was found to be less than 0.001. At the 1% significance level, there is sufficient evidence to reject the null hypothesis that the observed data is serially independent.

Overall, the transition probability matrix for each air monitoring station is similar. The transition probabilities for the transition from the Low Risk AQHI state to the Low Risk AQHI state was the highest occurring probability in the transition probability matrix for each of the air monitoring stations (mean of 0.990, ranging from 0.978 in Toronto West to 0.999 in Petawawa). This implies that the Low Risk AQHI state is generally stable since once the Low Risk AQHI state is entered, it is expected to remain for a while. Transition probabilities for moving directly from the Low Risk AQHI state to a High Risk state were generally non-existent for every air monitoring station. This means that the sudden onset of a High Risk AQHI state from the Low Risk AQHI state is extremely unlikely.

The transition probabilities for the transition out of the Moderate Risk AQHI state were highest for the transition to the Moderate Risk AQHI state (mean of 0.865; ranging from 0.782 in Thunder Bay to 0.890 in Toronto Downtown). These results imply that the Moderate Risk AQHI state is generally stable since once the Moderate Risk AQHI state is entered, it is likely to remain in that state. The next likeliest transition was to the Low Risk AQHI state (mean of 0.134, ranging from 0.110 in Toronto Downtown to 0.209 in Sault Ste Marie). The least probable transition was to the High Risk AQHI state (mean of 0.002, ranging from 0.000 at multiple air monitoring stations to 0.012 in Grand Bend).

Due to the low number of observed occurrences of High Risk AQHI category during the study period, the transition probabilities from the High Risk AQHI state to other states are less consistent across the air monitoring stations. Generally, the most probable transition from the High Risk AQHI state was to the High Risk AQHI state (mean of 0.554, ranging from 0.000 in Cornwall to 0.714 in Kitchener). The next likeliest transition was to the Moderate Risk AQHI state (mean of 0.443, ranging from 0.000 in Thunder Bay to 1.000 at

Table 2 Markov chain transition observation frequency

Air monitoring station		Transition observation frequency									Chi-square statistic		
Number	Name	n_{11}	n_{12}	n_{13}	n_{21}	n_{22}	n_{23}	n_{31}	n_{32}	n_{33}	DF	χ^2	p value
1	Barrie	39917	398	0	394	2258	0	0	0	0	2	30,387.5	< 0.001
2	Belleville	40588	262	0	265	1452	10	0	10	13	4	43,526.0	< 0.001
3	Brantford	40790	325	0	331	1868	6	0	6	6	4	41,549.3	< 0.001
4	Burlington	38082	603	0	608	4078	5	0	5	6	4	44,602.1	< 0.001
5	Chatham	39679	368	0	370	2354	3	0	3	1	4	33,965.6	< 0.001
6	Cornwall	41765	221	0	220	1132	1	0	1	0	4	29,999.7	< 0.001
7	Dorset	42467	93	0	91	371	0	0	0	0	2	27,474.6	< 0.001
8	Grand Bend	40814	237	1	248	1429	20	0	22	33	4	46185.8	< 0.001
9	Guelph	40432	342	0	345	1985	2	0	2	1	4	35,497.6	< 0.001
10	Hamilton Downtown	35324	774	0	784	6246	7	0	7	13	4	50,674.0	< 0.001
11	Hamilton West	37090	638	0	635	4554	1	0	1	0	4	31,783.6	< 0.001
12	Kingston	41506	242	0	246	1265	3	0	3	2	4	36,888.6	< 0.001
13	Kitchener	39990	334	0	336	2090	4	0	4	10	4	52914.3	< 0.001
14	London	40849	307	0	308	1839	3	0	3	0	4	31,269.3	< 0.001
15	Mississauga	39832	426	0	429	2492	1	0	1	1	4	41,472.5	< 0.001
16	Newmarket	40422	362	0	363	2326	10	0	10	14	4	46,597.4	< 0.001
17	North Bay	41917	220	1	221	876	1	0	2	4	4	46,414.0	< 0.001
18	Oakville	38748	457	0	458	3172	4	0	4	1	4	33,572.8	< 0.001
19	Ottawa Downtown	40704	284	0	283	1783	0	0	0	0	2	31,539.5	< 0.001
20	Parry Sound	41873	150	0	150	813	1	0	1	0	4	30,406.1	< 0.001
21	Petawawa	42939	45	0	45	200	0	0	0	0	2	28,733.5	< 0.001
22	Peterborough	40890	284	0	286	1684	5	0	5	1	4	32,278.5	< 0.001
23	Port Stanley	39833	319	0	321	1939	8	0	8	15	4	48,628.0	< 0.001
24	Sarnia	38447	528	0	525	3653	18	0	18	14	4	40,186.2	< 0.001
25	Sault Ste Marie	41695	190	0	190	719	0	0	0	0	2	26,467.8	< 0.001
26	St Catherines	39990	375	0	375	2224	0	0	0	0	2	30,780.8	< 0.001
27	Sudbury	40774	299	0	300	1659	0	0	0	0	2	30,347.8	< 0.001
28	Thunder Bay	41760	211	1	213	766	0	1	0	0	4	26,010.2	< 0.001
29	Tiverton	41246	191	0	192	1164	9	0	9	7	4	39,333.2	< 0.001
30	Toronto Downtown	35299	730	0	723	5863	3	0	3	3	4	42,875.8	< 0.001
31	Toronto East	37438	582	0	584	4642	10	0	10	23	4	53,946.2	< 0.001
32	Toronto West	35620	799	0	800	5913	9	0	9	15	4	48,661.0	< 0.001
33	Windsor Downtown	36102	757	0	765	5511	10	0	10	11	4	43,581.8	< 0.001
34	Windsor West	37305	712	0	709	4468	14	0	14	20	4	45,695.4	< 0.001

multiple air monitoring stations). Finally, the least probable transition was to the Low Risk AQHI state (mean of 0.003; ranging from 0.000 at all air monitoring stations except Thunder Bay to 1.000 in Thunder Bay).

Generally, transitions out of an AQHI state were almost always to an AQHI state with a risk category one level above or below it. For example, only three transitions from the Low Risk AQHI state directly to the High Risk AQHI state were observed out of 1,365,195 transitions from the Low Risk AQHI state, and 1,463,652 transitions in the study period. Only one transition from the High Risk AQHI state directly to the Low Risk AQHI was observed out of the 386 h spent in total in the High Risk AQHI state and 1,463,652 h in the study period. This indicates that the AQHI risk states usually transition gradually, and

Table 3 Markov chain transition probabilities

Air monitoring station		Transition probabilities								
Number	Name	P_{11}	P_{12}	P_{13}	P_{21}	P_{22}	P_{23}	P_{31}	P_{32}	P_{33}
1	Barrie	0.990	0.010	N/A	0.149	0.851	N/A	N/A	N/A	N/A
2	Belleville	0.994	0.006	0.000	0.153	0.841	0.006	0.000	0.435	0.565
3	Brantford	0.992	0.008	0.000	0.150	0.847	0.003	0.000	0.500	0.500
4	Burlington	0.984	0.016	0.000	0.130	0.869	0.001	0.000	0.455	0.545
5	Chatham	0.991	0.009	0.000	0.136	0.863	0.001	0.000	0.750	0.250
6	Cornwall	0.995	0.005	0.000	0.163	0.837	0.001	0.000	1.000	0.000
7	Dorset	0.998	0.002	N/A	0.197	0.803	N/A	N/A	N/A	N/A
8	Grand Bend	0.994	0.006	0.000	0.146	0.842	0.012	0.000	0.400	0.600
9	Guelph	0.992	0.008	0.000	0.148	0.851	0.001	0.000	0.667	0.333
10	Hamilton Downtown	0.979	0.021	0.000	0.111	0.888	0.001	0.000	0.350	0.650
11	Hamilton West	0.983	0.017	0.000	0.122	0.877	0.000	0.000	1.000	0.000
12	Kingston	0.994	0.006	0.000	0.162	0.836	0.002	0.000	0.600	0.400
13	Kitchener	0.992	0.008	0.000	0.138	0.860	0.002	0.000	0.286	0.714
14	London	0.993	0.007	0.000	0.143	0.855	0.001	0.000	1.000	0.000
15	Mississauga	0.989	0.011	0.000	0.147	0.853	0.000	0.000	0.500	0.500
16	Newmarket	0.991	0.009	0.000	0.134	0.862	0.004	0.000	0.417	0.583
17	North Bay	0.995	0.005	0.000	0.201	0.798	0.001	0.000	0.333	0.667
18	Oakville	0.988	0.012	0.000	0.126	0.873	0.001	0.000	0.800	0.200
19	Ottawa Downtown	0.993	0.007	N/A	0.137	0.863	N/A	N/A	N/A	N/A
20	Parry Sound	0.996	0.004	0.000	0.156	0.843	0.001	0.000	1.000	0.000
21	Petawawa	0.999	0.001	N/A	0.184	0.816	N/A	N/A	N/A	N/A
22	Peterborough	0.993	0.007	0.000	0.145	0.853	0.003	0.000	0.833	0.167
23	Port Stanley	0.992	0.008	0.000	0.142	0.855	0.004	0.000	0.348	0.652
24	Sarnia	0.986	0.014	0.000	0.125	0.871	0.004	0.000	0.563	0.438
25	Sault Ste Marie	0.995	0.005	N/A	0.209	0.791	N/A	N/A	N/A	N/A
26	St Catherines	0.991	0.009	N/A	0.144	0.856	N/A	N/A	N/A	N/A
27	Sudbury	0.993	0.007	N/A	0.153	0.847	N/A	N/A	N/A	N/A
28	Thunder Bay	0.995	0.005	0.000	0.218	0.782	0.000	1.000	0.000	0.000
29	Tiverton	0.995	0.005	0.000	0.141	0.853	0.007	0.000	0.563	0.438
30	Toronto Downtown	0.980	0.020	0.000	0.110	0.890	0.000	0.000	0.500	0.500
31	Toronto East	0.985	0.015	0.000	0.112	0.887	0.002	0.000	0.303	0.697
32	Toronto West	0.978	0.022	0.000	0.119	0.880	0.001	0.000	0.375	0.625
33	Windsor Downtown	0.979	0.021	0.000	0.122	0.877	0.002	0.000	0.476	0.524
34	Windsor West	0.981	0.019	0.000	0.137	0.861	0.003	0.000	0.412	0.588
Ontario Average		0.990	0.010	0.000	0.134	0.865	0.002	0.003	0.443	0.554

Note: Transition probabilities that were assigned a value of N/A could not be calculated due to the AQHI risk category not being entered during the period of the study (January 1, 2015–December 31, 2019). The subscripts of the transition probability, 1, 2, and 3 represent the Low Risk, Moderate Risk, and High Risk AQHI category, respectively

sudden, large changes are not expected. This fact means that public health officials will likely not need to issue an air quality advisory for the High Risk AQHI category without an ongoing air quality advisory for the Moderate Risk AQHI category. Additionally, this observation indicates that if any air quality mitigation measures were instituted during the study period, they were not effective enough to

cause an immediate reduction from the High Risk AQHI category to the Low Risk AQHI category.

3.2 Stationary Distribution

The stationary distribution of the Markov chain for each air monitoring station was calculated. Each Markov chain was shown to be irreducible and ergodic, which meant that the stationary distribution could be calculated and that it was unique. It is worth mentioning that the primary source of

air pollution in Ontario is population growth. In the long term, this factor may increase the probability of air pollution over time; however, the study period (2015–2019) is relatively short. Therefore, the change in air-pollution probability over time is considered insignificant that means the Markov chain adopted in this study is assumed to be homogeneous in time. The stationary distribution of each Markov chain is contained in Table 4. An example of the average stationary distribution for all of the air monitoring stations is $[\pi_1, \pi_2, \pi_3] = [0.933, 0.067, 0.000]$.
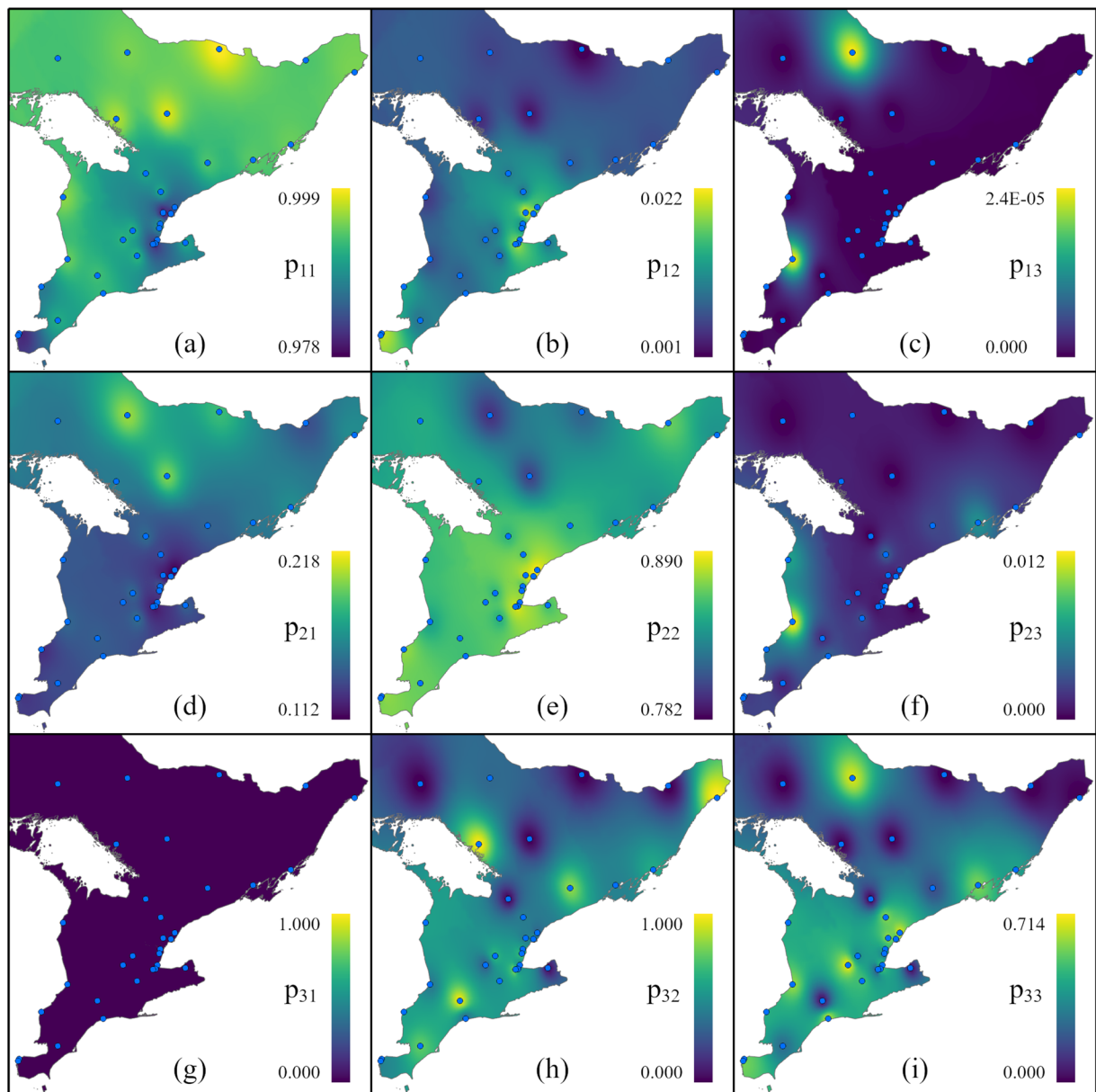

Fig. 2 Transition probabilities. **a** p_{11}. **b** p_{12}. **c** p_{13}. **d** p_{21}. **e** p_{22}. **f** p_{23}. **g** p_{31}. **h** p_{32}. **i** p_{33}

Table 4 Stationary distribution

Air monitoring station		Stationary distribution							
Number	Name	π_1	π_2	π_3	Number	Name	π_1	π_2	π_3
1	Barrie	0.938	0.062	0.000	18	Oakville	0.915	0.085	0.000
2	Belleville	0.959	0.040	0.001	19	Ottawa Downtown	0.952	0.048	0.000
3	Brantford	0.950	0.050	0.000	20	Parry Sound	0.978	0.022	0.000
4	Burlington	0.892	0.107	0.000	21	Petawawa	0.994	0.006	0.000
5	Chatham	0.936	0.063	0.000	22	Peterborough	0.954	0.045	0.000
6	Cornwall	0.969	0.031	0.000	23	Port Stanley	0.946	0.053	0.001
7	Dorset	0.989	0.011	0.000	24	Sarnia	0.902	0.098	0.001
8	Grand Bend	0.961	0.038	0.001	25	Sault Ste Marie	0.979	0.021	0.000
9	Guelph	0.946	0.054	0.000	26	St Catherines	0.940	0.060	0.000
10	Hamilton Downtown	0.838	0.161	0.000	27	Sudbury	0.955	0.045	0.000
11	Hamilton West	0.879	0.121	0.000	28	Thunder Bay	0.977	0.023	0.000
12	Kingston	0.965	0.034	0.000	29	Tiverton	0.968	0.032	0.000
13	Kitchener	0.943	0.056	0.000	30	Toronto Downtown	0.844	0.156	0.000
14	London	0.950	0.049	0.000	31	Toronto East	0.879	0.121	0.001
15	Mississauga	0.933	0.067	0.000	32	Toronto West	0.844	0.156	0.001
16	Newmarket	0.938	0.062	0.001	33	Windsor Downtown	0.855	0.144	0.000
17	North Bay	0.974	0.025	0.000	34	Windsor West	0.879	0.120	0.001
Ontario Average		0.933	0.067	0.000					

The subscripts, 1, 2, and 3 represent the Low Risk, Moderate Risk, and High Risk AQHI category, respectively

The highest proportion of time for each monitoring station was spent in the Low Risk AQHI category (mean of 0.933, ranging from 0.838 in Hamilton Downtown to 0.994 in Petawawa). The second highest proportion of time was spent in the Moderate Risk AQHI category (mean of 0.067, ranging from 0.006 in Petawawa to 0.161 in Hamilton Downtown). Very little time was spent in the High Risk AQHI category (mean of < 0.000, ranging from 0.000 at most air monitoring stations to 0.001 at eight air monitoring stations (Table 4).

Generally, the air monitoring stations with the lowest expected proportion of time spent in the Low Risk AQHI category are near densely populated areas in urban environments. A summary of the spatial analysis of the stationary distributions is shown in Fig. 3.
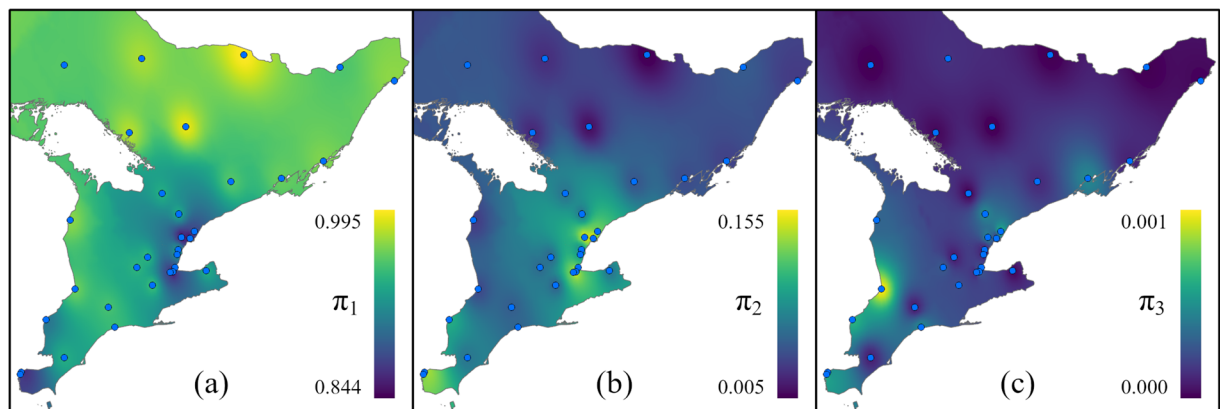

Fig. 3 Long-run proportion of time spent in **a** AQHI Low Risk category, **b** AQHI Moderate Risk category, and **c** AQHI High Risk category

Table 5 Mean persistence time of an AQHI category

Number	Name	s_{11}	s_{22}	s_{33}	Number	Name	s_{11}	s_{22}	s_{33}
Air monitoring station		Mean persistence time in hours							
1	Barrie	101.3	6.7	N/A	18	Oakville	85.8	7.9	1.3
2	Belleville	155.9	6.3	2.3	19	Ottawa Downtown	144.3	7.3	N/A
3	Brantford	126.5	6.5	2.0	20	Parry Sound	280.2	6.4	1.0
4	Burlington	64.2	7.7	2.2	21	Petawawa	955.2	5.4	N/A
5	Chatham	108.8	7.3	1.3	22	Peterborough	145.0	6.8	1.2
6	Cornwall	190.0	6.1	1.0	23	Port Stanley	125.9	6.9	2.9
7	Dorset	457.6	5.1	N/A	24	Sarnia	73.8	7.7	1.8
8	Grand Bend	172.5	6.3	2.5	25	Sault Ste Marie	220.4	4.8	N/A
9	Guelph	119.2	6.7	1.5	26	St Catherines	107.6	6.9	N/A
10	Hamilton Downtown	46.6	8.9	2.9	27	Sudbury	137.4	6.5	N/A
11	Hamilton West	59.1	8.2	1.0	28	Thunder Bay	198.0	4.6	1.0
12	Kingston	172.5	6.1	1.7	29	Tiverton	216.9	6.8	1.8
13	Kitchener	120.7	7.1	3.5	30	Toronto Downtown	49.4	9.1	2.0
14	London	134.1	6.9	1.0	31	Toronto East	65.3	8.8	3.3
15	Mississauga	94.5	6.8	2.0	32	Toronto West	45.6	8.3	2.7
16	Newmarket	112.7	7.2	2.4	33	Windsor Downtown	48.7	8.1	2.1
17	North Bay	190.7	4.9	2.3	34	Windsor West	53.4	7.2	2.4
Ontario average		101.7	6.7	1.5					

The subscripts, 1, 2, and 3 represent the Low Risk, Moderate Risk, and High Risk AQHI category, respectively

3.3 Mean Persistence Time of an AQHI Category

The mean persistence time of each state of the Markov chains was calculated for each air monitoring station; the results are reported in Table 5, and the spatial distribution is shown in Fig. 4. The average mean persistence time in hours for all of the air monitoring stations in Ontario is $[s_{11}, s_{22}, s_{33}] = [101.7, 6.7, 1.5]$.

The mean persistence time was the highest for the Low Risk AQHI category (mean of 101.7 h, ranging from 46.6 h in Hamilton Downtown to 955.2 h in Petawawa). The second highest mean persistence time was for the Moderate Risk AQHI category (mean of 6.7 h, ranging from 4.6 h in Thunder Bay to 9.1 h in Toronto Downtown). The lowest mean persistence time was for the High Risk AQHI category (mean of 1.5 h, ranging
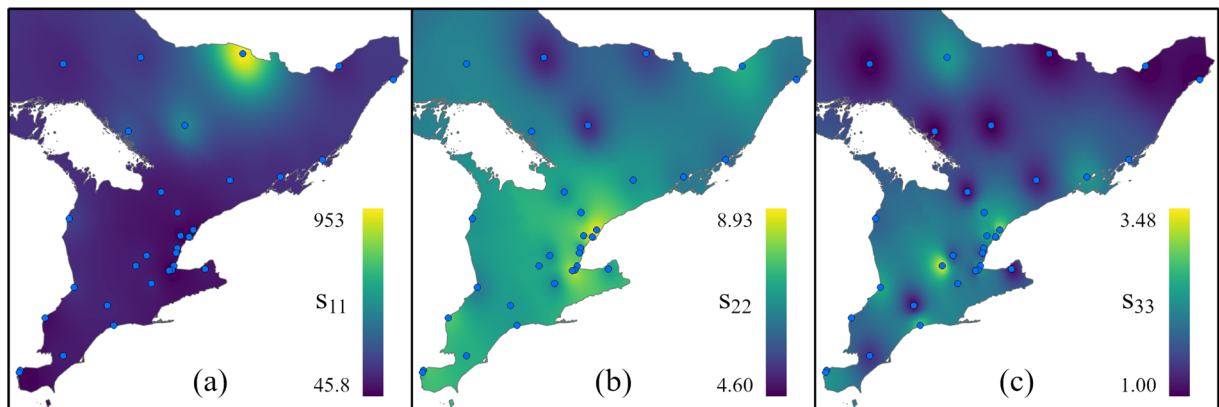

Fig. 4 Mean persistence time of **a** AQHI Low Risk category, **b** AQHI Moderate Risk Category, and **c** AQHI High Risk category

from 1.0 h at five air monitoring stations to 3.5 h in Kitchener).

The ideal distribution for air quality health impacts would have a very high mean persistence time for the Low Risk AQHI category, and very low mean persistence times for the Moderate Risk, and High Risk AQHI categories. Generally, this pattern was observed at air monitoring stations in less populated, rural environments. At the air monitoring stations in more populated, urban environments, this pattern was also observed to a lesser extent. Based on the results of this analysis, public health officials can, on average, expect to issue air quality advisories in Ontario with health messages lasting for 6.7 h for the Moderate Risk AQHI category and 1.5 h for the High Risk AQHI category.

4 Conclusions

In this study, we used the Markov chain model to investigate the pattern of AQHI risk categories in Ontario for a period of 5 years, from 2015 to 2019. We estimated the transition probability matrix for each air monitoring station. We identified a general trend in the transition probability patterns for AQHI risk categories. We found that each AQHI risk category generally tends to be stable across each air monitoring station. And transitions between the AQHI risk categories occur gradually; sudden, large transitions between risk categories two levels away are not expected to occur.

The transition probability matrix facilitates the calculation of the stationary distribution showing the long-term proportion of time that each air monitoring station spends in a specific AQHI risk category. We calculated the mean persistence time for each AQHI risk category and identified the average duration of air quality advisories in Ontario for the Moderate Risk and High Risk AQHI categories. Overall, we found that air monitoring stations in less populated, rural environments had better air quality and spent less time in Moderate Risk and High Risk AQHI categories than air monitoring stations in more populated, urban environments.

The discrete-time Markov chain analysis done in this study can be extended to include other air monitoring stations outside of Ontario to broaden the study area and determine if the trends identified in this study can be generalised to other areas. The analysis could also be split into multiple periods to identify temporal trends in the observed AQHI risk category data. In the future, if

there is enough data, a similar analysis of AQHI data recorded during the period of the COVID-19 pandemic would provide more insights about the impact of reduced human outdoor activities on air quality.

Notation   The following symbols are used in this paper:
I identity matrix;
n_{ij} observed frequency of transitions from state i to state j;
P transition probability matrix;
P_T transition probability matrix consisting of only the transient states;
p_{ij} one-step transition probability from state i to state j;
S expected amount of time before absorption matrix;
s_{ij} expected amount of time before absorption that Markov chain spends in state j given it started in state i;
X_n stochastic process;
π_j long-run proportion of time spent in state j;

References

Alyousifi, Y., Masseran, N., & Ibrahim, K. (2018). Modeling the stochastic dependence of air pollution index data. *Stochastic Environmental Research and Risk Assessment, 32*(6), 1603–1611.

Alyousifi, Y., et al. (2019). Markov chain modeling for air pollution index based on maximum a posteriori method. *Air Quality, Atmosphere & Health, 12*(12), 1521–1531.

Asadollahfardi, G., Zangooei, H., & Aria, S. H. (2016). Predicting PM 2.5 concentrations using artificial neural networks and Markov chain, a case study Karaj City. *Asian Journal of Atmospheric Environment (AJAE), 10*, 2.

Baik, H.-S., Jeong, H. S., & Abraham, D. M. (2006). Estimating transition probabilities in Markov chain-based deterioration models for management of wastewater systems. *Journal of water resources planning and management, 132*(1), 15–24.

Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. *The lancet, 360*(9341), 1233–1242.

Caraka, R. E., et al. (2019). Prediction of Status Particulate Matter 2.5 using state Markov chain stochastic process and HYBRID VAR-NN-PSO. *IEEE Access, 7*, 161654–161665.

Environment and Climate Change Canada. 2015. "Understanding Air Quality Health Index messages." https://www.canada.ca/en/environment-climate-change/services/air-quality-health-index/understanding-messages.html. Accessed 23 Mar 2020.

Environment and Climate Change Canada. 2019. "About the Air Quality Health Index." https://www.canada.ca/en/environment-climate-change/services/air-quality-health-index/about.html. Accessed 23 Mar 2020.

Mohamad, N. S., Deni, S. M., & Ul-Saufie, A. Z. (2018). Application of the first order of Markov chain model in describing the PM10 occurrences in Shah Alam and Jerantut, Malaysia. *Pertanika Journal of Science & Technology, 26*(1), 367–368.

Nebenzal, A., & Fishbain, B. (2018). Long-term forecasting of nitrogen dioxide ambient levels in metropolitan areas using

the discrete-time Markov model. *Environmental modelling & software, 107*, 175–185.

Ontario Ministry of the Environment, Conservation and Parks. (2019a). *Air Quality Health Index (AQHI) Historical Search*. Air Quality Ontario. http://www.airqualityontario.com/aqhi/search.php. Accessed 20 Mar 2020.

Ontario Ministry of the Environment, Conservation and Parks. (2019b). *Frequently asked questions*. Air Quality Ontario. http://www.airqualityontario.com/press/faq.php. Accessed 24 Mar 2020.

Rodrigues, E. R., Tarumoto, M. H., & Tzintzun, G. (2019). Application of a non-homogeneous Markov chain with seasonal transition probabilities to ozone data. *Journal of Applied Statistics, 46*(3), 395–415.

Romanof, N. (1982). A Markov chain model for the mean daily SO2 concentrations. *Atmospheric Environment (1967), 16.8*, 1895–1897.

Ross, S. M. (2014). *Introduction to probability models*. Academic press.

Sahin, A. D., & Sen, Z. (2001). First-order Markov chain approach to wind speed modelling. *Journal of Wind Engineering and Industrial Aerodynamics, 89*(3-4), 263–269.

Schoof, J. T., & Pryor, S. C. (2008). On the proper order of Markov chain model for daily precipitation occurrence in the contiguous United States. *Journal of Applied Meteorology and Climatology, 47*(9), 2477–2486.

da Silva, J. J., et al. (2019). Application of Markov chain on daily rainfall data in Paraíba-Brazil from 1995 to 2015. *Acta Scientiarum. Technology, 41*, e37186–e37186.

Szyszkowicz, M. (2019). The Air Quality Health Index and all emergency department visits. *Environmental Science and Pollution Research, 26*(24), 24357–24361.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

Zakaria, N. N., et al. (2019). Markov chain model development for forecasting air pollution index of Miri, Sarawak. *Sustainability, 11*(19), 5190.