Avicenna J Environ Health Eng. 2017 June; 4(1):e58031. doi: 10.5812/ajehe.58031.
Published online 2017 June 25. Research Article

Application of Poisson Hidden Markov Model to Predict Number of PM2.5 Exceedance Days in Tehran During 2016-2017

Fatemeh Sarvi,[1] Azam Nadali,[2] Mahmoud Khodadost,[3,4] Melika Kharghani Moghaddam,[5] and Majid Sadeghifar[6,*]

[1]Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, IR Iran
[2]Department of Environmental Health Engineering, School of Public Health, Hamadan University of Medical Sciences, Hamadan, IR Iran
[3]Department of Epidemiology, School of Public Health, Shahid Behesht University of Medical Sciences, Tehran, IR Iran
[4]Department of Epidemiology and Biostatistics, Iran University of Medical Sciences, Tehran, IR Iran
[5]Department of Public Health, School of Public Health, Hamadan University of Medical Sciences, Hamadan, IR Iran
[6] Department of Statistics, Faculty of Basic Sciences, Bu-Ali Sina University, Hamadan, IR Iran

[*]*Corresponding author*: Majid Sadeghifar (PhD), Department of Statistics, Faculty of Basic Science, Bu-Ali Sina University, Hamadan, IR Iran. Tel: +98-9188115417, Fax: +98-8138271541, E-mail: sadeghifar@basu.ac.ir

Received 2017 April 01; **Accepted** 2017 May 16.

Abstract

PM2.5 is an important indicator of air pollution. This pollutant can result in lung and respiratory problems in people. The aim of the present study was to predict number of PM2.5 exceedance days using Hidden Markov Model considering Poisson distribution as an indicator for people susceptible to that particular level of air quality. In this study, evaluations were made for number of PM2.5 exceedance days in Tehran, Iran, from Oct. 2010 to Dec. 2015. The Poisson hidden Markov model was applied considering various hidden states to make a two-year forecast for number of PM2.5 exceedance days. We estimated the Poisson Hidden Markov's parameters (transition matrix, probability, and lambda) by using maximum likelihood method. By applying the Akaike Information Criteria, the hidden Markov model with three states was used to make the prediction. The results of forecasting mean, median, mode, and interval for the three states of Poisson hidden Markov model are reported. The results showed that the number of exceedance days in a month for the next two years using the third state of the model would be 5 to 16 days. The predicted mode and mean for the third months afterward at the third state were 11 and 11. These predictions showed that number of exceedance days (predicted mean of 6.87 to 11.39 days) is relatively high for sensitive individuals according to the PM2.5 Air Quality Index. Thus, it is essential to monitor levels of suspended particulate air pollution in Tehran.

Keywords: PM2.5 Pollution, Poisson Hidden Markov Model (HMM), Predicting, Tehran, Air Pollution

1. Background

World health organization (WHO) reported in 2012 that many people die of the adverse effects of exposure to fine particles such as cardiovascular and respiratory diseases and cancers (1). Numerous studies have been conducted around the world on the effects of short-term and long-term exposure to air pollution. Among the investigated air pollutants, $PM_{2.5}$ is widely studied due to its negative effects on people's health (2-5). It should be noted that this type of pollutants is mostly emitted to the environment from industrial sources (6). Based on the previous reports, there is a strong correlation between concentration of $PM_{2.5}$ and rate of hospitalization due to cardiovascular diseases such as stroke and ischemic and respiratory diseases such as acute respiratory diseases, chronic obstructive pulmonary disease (COPD), and cancer (2, 7, 8).

The issue of air pollution is considered as the Tehran's most acute problem for those living in this metropolitan (9). Although air quality of Tehran has improved slightly due to removing lead from fuels and decreased concentrations of sulfur and carbon monoxide, annual reports show that air quality is classified as unhealthy for one in three days. This is mostly due to the accumulative effects of population growth, use of old and heavy vehicles, high number of commercial institutions, and geographical and meteorological factors (9, 10). It should be noted that poor air quality in Tehran makes people stay at home or use a mask in public places (9). Recently, EPA has released a revised standard as the index for $PM_{2.5}$ air pollution (6). It should also be mentioned that retaining pollutants under the standard values set for each pollutant could not be considered as completely safe for children, pregnant women, the elderly, and those with critical diseases.

Predicting concentrations of PM is an important issue in controlling and decreasing concentrations of air pollutants (6). So far, various methods have been applied to predict the concentration of PM. The two most commonly applied methods based on mathematical models are briefly explained as below (6, 11):

1.1. Deterministic Models Known also as Chemical Transmission Models (CTMs)

1.2. Statistical Models

The first model focuses on the source and transmission of various chemical elements.

Statistical methods can be used to determine the correlation between data on air pollution and factors related to meteorology. A wide range of statistical methods have been proposed in previous studies in relation to air quality; however, the data required for predicting PM using such methods are scarce, particularly in Iran (12-14).

Among the proposed statistical methods, the hidden Markov model (HMM) has a stronger mathematical structure and it can be used to determine mathematical functions between hidden and observed patterns. HHMs have been used recently to predict the high concentrations of $PM_{2.5}$ in California, USA; based on reports extracted from that study, this method is capable of predicting the number of $PM_{2.5}$ exceedance days with high precision (11).

One method for overcoming the problem of overdispersion due to heterogeneity is to use a mixture model with consideration of dependency between observations. This method uses the dependency and Markov property (each observation is dependent on only one previous observation). Poisson–hidden Markov model is obtained by allowing this assumption (15).

The aim of this study was to forecast the future status of $PM_{2.5}$ air pollution in Tehran. This issue is important in health and medical areas. This study was carried out using Poisson hidden Markov model that respects dependency principle in observations. The manuscript is organized in the following sections: the first section presents data collection and discusses the used method and then, results are shown in understandable figures; the second section presents a discussion on the results using statistical methods.

2. Methods

In the present study, Tehran was selected because it is, as the capital of the country, an important and crowded city. It is also considered as one of the most polluted cities in Iran due to its dense traffic. The required data were acquired from the air quality control center (AQCC) from Oct. 2010 to Dec. 2015. Figure 1 shows the distribution of the stations.

According to the guidelines for air pollution suggested by EPA, the 24-hour standard for $PM_{2.5}$ has set at 50 $\mu g.m^{-3}$ and levels exceeding 35 $\mu g.m^{-3}$ are considered unhealthy for susceptible people (16).

The aim was to determine number of exceedance days for sensitive people (children, the elderly and pregnant women) and predict $PM_{2.5}$ exceedance days in Tehran for 2016 and 2017 (a 24-month period).

Figure 2 shows the number of $PM_{2.5}$ exceedance days at various months from Sep. 2010 to Dec. 2015.

Based on studies on air pollution prediction at different cities all around the world (17-19), the HMM was used for prediction considering Poisson distribution of number of exceedance days for various months in the years 2016 and 2017.

The Markov chain is a randomized process that displaces among modes based on studied mechanism (20). This chain is based on the assumption that a current mode depends on its previous mode and the method is used to model future events.

The hidden Markov model (HMM) is a mixed model that constitutes a randomized two-part process. One part is not directly observable and includes hidden modes (Ct) and the other part includes a series of randomized variables that are dependent on the hidden state of the first part (X_t).

X_t is not dependent on previous Ct state; it is only dependent on the current state.

In the present study, evaluating the number of exceedance days at each month was considered as a randomized variable, distributions from randomized variables had hidden status, and randomized variables were dependent only on distribution.

$$P\left(C_t|C^{t-1}\right) = P\left(C_t|C_{t-1}\right) t = 2, 3, \ldots \quad (1)$$

Where:

$C^{(t-1)} = C_1, C_2, \ldots, C_{t-1}$

$$P\left(X_t|X^{(t-1)}, C^{(t)}\right) = P\left(X_t|C_t\right) t \epsilon N \quad (2)$$

$C^{(t)} = (C_1, C_2, \ldots, C_t)$

$X^{(t-1)} = (X_1, X_2, \ldots, X_{t-1})$

$$P_i\left(x\right) = P\left(X_t = x|C_t = i\right) \quad (3)$$

P_i indicates probability mass function of X_t at time t lie in i-th state.

For Poisson observation model, we have:

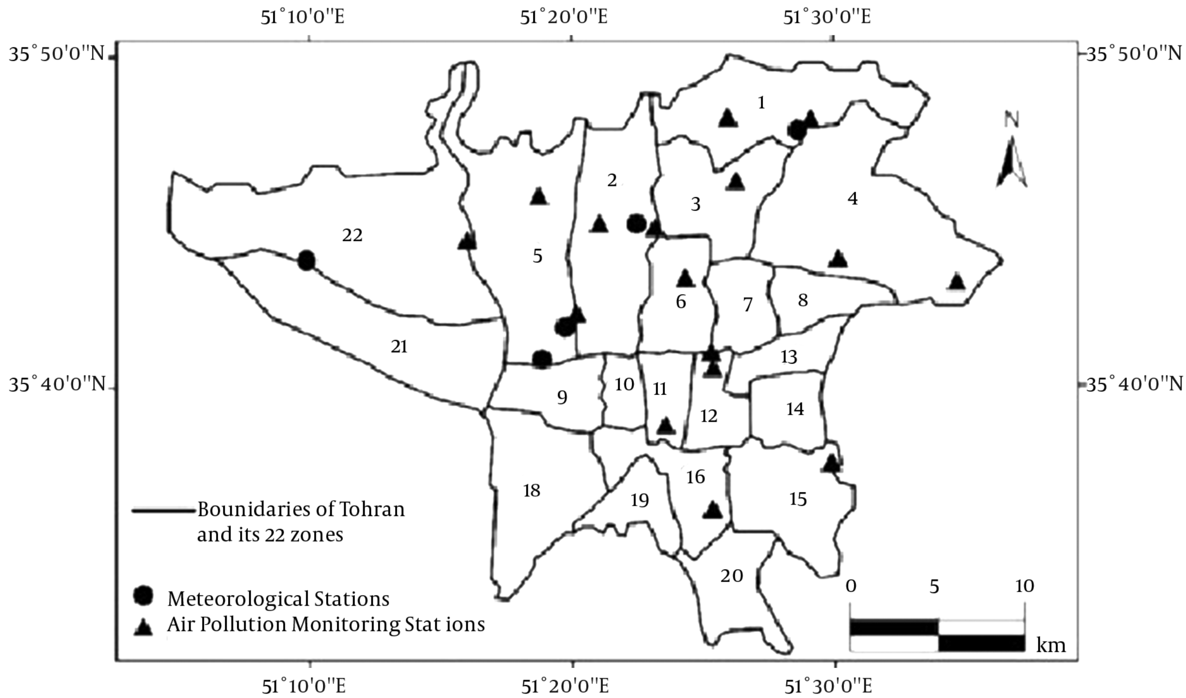

Figure 1. Location of Air Pollution Measurement Stations in Tehran

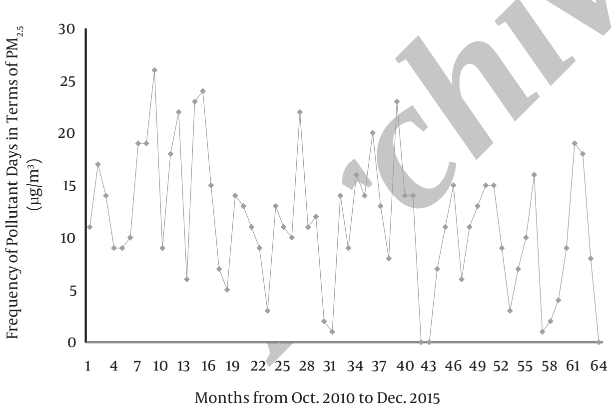

Figure 2. Frequency of Monthly $PM_{2.5}$ Exceedance Days in Tehran from Oct. 2010 to Dec. 2015

$$P_i(x) = P(X_t = x | C_t = i) = \frac{e^{\lambda_i}\lambda_i^x}{x!} \quad (4)$$

The likelihood function for this model is shown by: Equation 5.

$$L_T = P\left(X^{(T)} = x^{(T)}\right) \quad (5)$$

$$= \delta P(x_1)\Gamma P(x_2)\ldots\Gamma P(x_T)1'$$

In this function, initial distribution is δ and P(x) is diagonal matrix. The elements of P(x) are $P_i(x)$s. Γ is transition probability matrix. The parameters are obtained by maximum likelihood estimation

Forecasting distribution can be obtained from this equation:

$$P\left(X_{T+h} = x | X^T = x^T\right) = \varnothing_T \Gamma^h P(x) 1' \quad (6)$$

Where:

$\emptyset_T = \alpha_T / \alpha_T 1'$

And:

$\alpha_T = \alpha_{T-1}\Gamma P(X_t)$

To predict the number of exceedance days, the several states of Poisson Hidden Markov Models were fitted (15). To compare the fitting of different states of model, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) were applied.

In this case, number of states was considered in terms of transition matrix states, the proportion of those cases transferred from one state to other states, and those states that did not change were computed as elements of this matrix.

Forecasting mean, mode, and probability of staying on the new state were calculated for the next two years (monthly).

All analyses were made using R (3.1.0) software.

3. Results and Discussion

In the present study, number of $PM_{2.5}$ exceedance days for sensitive individuals was evaluated from Oct. 2010 to Dec. 2015 and number of exceedance days was predicted using HM statistical procedure for 2016 and 2017.

Mean $PM_{2.5}$ concentration for the study years is presented in Table 1 indicating the high concentration of $PM_{2.5}$ for sensitive individuals during most of the study seasons.

Figure 3 shows $PM_{2.5}$ concentrations over recent years, representing a decreasing trend from 2010 to 2015. A higher number of exceedance days were observed in May 2011.

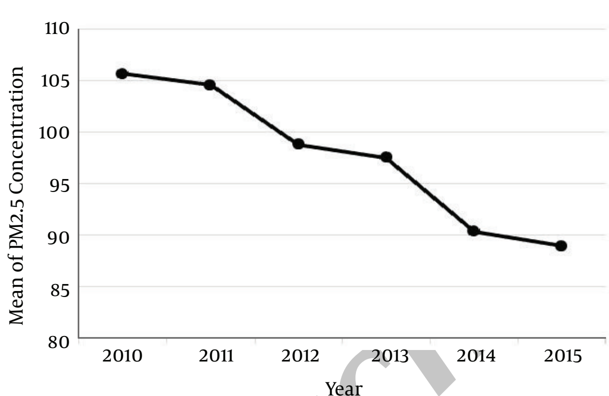

Figure 3. The trend of mean PM2.5 concentration between 2010 and 2015

The number of PM2.5 exceedance days was 63 days from Oct. 2010 to Dec. 2015. Maximum and minimum numbers of exceedance days for the sensitive group were 26 and 0 days, respectively.

Mean and variance of exceedance days were 11.54 and 40.35, respectively, indicating that considering Poisson distribution, the predicted variance was estimated to be less than the real value (since the Poisson model assumes equal mean and variance) and thus, over-dispersion occurs (21). Therefore, the use of Poisson model was not appropriate and instead, the use of Poisson Hidden Markov Model was more favorable.

Fitting hidden Markov model in various modes (from 1 to 3 different modes) and comparison of these modes showed that the model with three modes gave the lower values of AIC and BIC; therefore, the three-mode model was selected to fit our data (Table 2).

Minimum and maximum exceedance days in the first, second, and third clusters were 0.4, 17.26, and 5.16, respectively. Therefore, these clusters were considered as status space in the Markov chain.

In order to estimate parameters associated with Poisson Hidden Markov at the transferring matrix, the possibilities, and their corresponding parameters of the Poisson distribution, maximum likelihood method was used (Table 3).

The results of predictions for the next 24 months (2016 to 2017) showed the possible number of exceedance days predicted by the third mode (with number of exceedance days from 5 to 16 days, with possibility more than 0.5) was higher than that predicted by the other two modes. Mean and mode for this course were 11 and 11, respectively. Evaluations for predicted intervals of exceedance days were as follows: the first month (0.7), the second month (0.10), and the other months (1.11) (Table 4). In addition, Figure 4 shows the trend of falling possibility in various modes. As can be observed in Figure 4, falling possibility in the third mode was more than that of the other two modes. It means that number of exceedance days in a month for sensitive individuals would be between 5 to 16 days in the years 2016 and 2017.

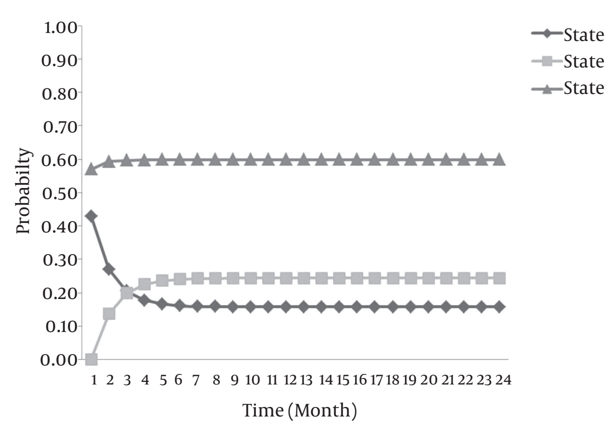

Figure 4. The Trend of Probability of Exccedance Days for Three States

Numerous studies have been done to assess concentrations of air suspended particulates and their relation to various diseases in the world. In addition, some authors have tried to predict particulate concentrations using statistical and mathematical models. However, relatively limited studies have been conducted on prediction using these models in Iran. The HM model could predict number of $PM_{2.5}$ exceedance days with little error. Time

Table 1. Mean, Minimum, and Maximum of PM2.5 Concentration ($\mu g/m^3$) at Various Seasons of 2010 to 2015

Years		Autumn (October-December	winter (January-Mars)	Spring (Aprils-June)	Summer (July-September)	Annual
2010	Min	-	-	24	51	24
	Mean	-	-	105.89	105.53	105.69
	Max	-	-	177	168	177
2011	Min	52	70	26	49	26
	Mean	108.06	107.97	97.77	104.47	104.62
	Max	204	145	146	173	204
2012	Min	57	69	50	42	42
	Mean	99.93	93.59	97.62	104.23	98.81
	Max	192	145	164	180	192
2013	Min	48	69	48	56	48
	Mean	80.02	97.98	101.09	112	97.54
	Max	141	130	170	165	170
2014	Min	40	69	41	45	40
	Mean	72.61	97.82	93.99	97.39	90.35
	Max	125	190	146	152	190
2015	Min	28	58	34	-	28
	Mean	80.03	89.03	92.53	-	88.93
	Max	162	187	162	-	187

Table 2. Comparison of Stationary Poisson Hidden Markov Model with Different States

Number of States	AIC	BIC
1	528.38	530.56
2	432.33	467.13
3	421.89	441.45

Table 3. Parameter Estimation by Maximum Likelihood Method for Three State Poisson Hidden Markov[a]

	State 1	State 2	State 3
Lambda	1.555	19.015	10.876
	0.429	0.000	0.571
Transition matrix	0.000	0.412	0.588
	0.147	0.239	0.613
probability	0.155	0.244	0.601

[a]Lambda: Mean of distribution in each state; the elements of transition matrix: the proportion of those cases transferred from one state to other states and those states that did not change; probability: the probability that the frequency of new exceedance days lines in state 1, 2, or 3.

series techniques have been used to predict future status. Nevertheless, these models for counting data could result in over-dispersion and reduced prediction ability. An HM model using Poisson distribution can mediate the potential for over-dispersion. Results of the present study showed a combined model of three hidden modes that could predict monthly number of exceedance days in the range of 5 - 16 days, which are considered high. In total, 826 days (42.34%) had higher concentrations than the standard level for sensitive individuals, while 1071 days (57.66%) had lower concentrations than the standard level.

In addition, numerous studies have reported seasonal and daily changes in air pollution worldwide. Farah Halek et al. (2010) reported $PM_{2.5}$ concentrations at four stations in the northwest of Tehran and found higher accumulation of particles during warmer seasons compared to cold seasons and the total mean $PM_{2.5}$ concentration was reported as 210.5 $\mu g.m^{-3}$ while it was 100 $\mu g.m^{-3}$ in the present study. This difference could be due to lower sample size and different study areas that some of them may not be representative of the whole city (22). AditiKulshresta et al. (2009) studied seasonal variations of particle concentration in Agra, India, and reported 104.9 $\mu g.m^{-3}$ $PM_{2.5}$ concentrations in urban areas, which was equal to

Table 4. Results of Forecasting State, Mean, Median, Mode, and Interval by Three States of Poisson Hidden Markov Model

Month		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	Probability	0.430	0.270	0.205	0.178	0.166	0.161	0.159	0.158	0.158	0.158	0.158	0.158
2		0.000	0.137	0.198	0.224	0.236	0.240	0.242	0.243	0.243	0.243	0.244	0.244
3		0.570	0.593	0.596	0.598	0.598	0.599	0.599	0.599	0.599	0.599	0.599	0.599
Forecast mean		6.87	9.47	10.57	11.04	11.24	11.32	11.36	11.37	11.38	11.38	11.38	11.39
Forecast median		7	10	11	11	11	11	11	11	11	11	11	11
Forecast mode		1	1	11	11	11	11	11	11	11	11	11	11
Forecast interval		(0,7)	(0,10)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)
Month		13	14	15	16	17	18	19	20	21	22	23	24
1	Probability	0.158	0.158	0.158	0.158	0.158	0.158	0.158	0.158	0.158	0.158	0.158	0.158
2		0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244
3		0.599	0.599	0.599	0.599	0.599	0.599	0.599	0.599	0.599	0.599	0.599	0.599
Forecast mean		11.39	11.39	11.39	11.39	11.39	11.39	11.39	11.39	11.39	11.39	11.39	11.39
Forecast median		11	11	11	11	11	11	11	11	11	11	11	11
Forecast mode		11	11	11	11	11	11	11	11	11	11	11	11
Forecast interval		(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)

the mean concentration of the pollutant in the present study. They also found the higher concentrations during winter months (November to February) (23).

Data on 1999 - 2005 show variations in PM10 and $PM^{2.5}$ in relation to aerial particles studied in different areas of Spain. In this study, urban areas and areas with heavy traffic had much higher concentrations than the countryside and rural areas; this result confirmed a relationship between air pollution particles and dispersions from anthropogenic sources of PM in urban areas with high-density of population.

Xiujuan Zhao et al. (2008) studied seasonal and daily variations in $PM_{2.5}$ concentration in urban and rural areas of Beijing (Jan. 2005 to Dec 2007) and reported higher $PM_{2.5}$ concentrations in winter (112 $\mu g/m_3$ in 2006 and 98 μg/m3 in 2007). Lower concentrations were recorded in spring and summer especially in years that had no dust storms. Increased $PM_{2.5}$ concentration in winter could be resulted from rising emissions from heating sources combined with low elevation of boundary layer. In addition, increasing concentrations of particles in the cold seasons in Tehran could be due to the city's geography; the city is surrounded by mountains in the North that causes air to become trapped. Increased concentrations of particles due to temperature inversion and reduced mixing height in winter may be another reason (24).

Based on a literature review, there has been little research on the application of HMM for predictions. Nevertheless, considering that the model could predict number of exceedance days, the required action should be planned to deal with the problem of air pollution. Junko Murakami used the Bayes method to estimate parameters of the Hidden Markov Model. Their results showed that for a low sample size and areas with little observational data, the Bayes method could be preferable to the maximum likelihood method for estimations (25).

Zhang et al. (2012) studied ozone levels using HMM by considering gamma distribution and found that according to ozone content, days could be divided into healthy and unhealthy days; they suggested using the model to predict concentrations that would exceed specific levels in the Livermore Valley, San Francisco. Since the purpose of this study was to predict number of days rather than concentrations, HMM was utilized considering Poisson distribution (18).

Peretz and Glamsch (2016) predicted hourly $PM_{2.5}$ concentrations in Santiago de Chile with an emphasis on night episodes. This was a seasonal prediction between August and April using a neural network. Their results showed higher concentrations at night episodes. They also showed that the accuracy of the model was higher than that of the other similar methods due to considering whole the possible models (26).

Zhang et al. (2013) predicted 8-hour concentration of ozone using HMM combined with generalized linear models and utilized data of 8 Livermore Valley areas from 2000 to 2007 to construct the model and used data from 2008 to 2009 to assess the validity of the model. Their results showed that the model worked well for predicting all ozone exceedance days. In the present study, HMM combined with Poisson generalized linear model had a great ability for prediction (27).

In this study, data from 2008 to 2009 were used to assess the model validity and results showed that the model operated well for prediction of total number of exceedance days. In the present study, the HMM combined with Poisson generalized linear model had a great prediction ability.

However, to our knowledge, the present study was the first to predict number of $PM_{2.5}$ exceedance days. The study provided evidence on many days of $PM_{2.5}$ exceedance for sensitive individuals.

A limitation of the present study was that it only focused on data from 63 months and the sample size was not sufficient for prediction of number of days in the next 24 months. Therefore, it is recommended that larger sample sizes and effective meteorological parameters such as temperature, moisture, and wind velocity be considered in further research.

4. Conclusions

In the present study, Tehran was selected for the study due to its high level of pollution. Poisson hidden Markov model was used to forecast numbers of $PM_{2.5}$ exceedance days. Results of the present study showed that 826 days (42.34%) had higher concentrations than the standard level for sensitive individuals for the next two years.

Hidden Markov Model considering poisson distribution had a great ability to predict number of exceedance days and this prediction can play an important role in decreasing and monitoring levels of $PM_{2.5}$ as well as in policy making for air pollution control.

Footnote

Conflict of Interest: There is no conflict of interest to declare.

References

1. WHO . pollution 2015. Available from: http://www.who.int/mediacentre/factsheets/fs313/en/.
2. Fattore E, Paiano V, Borgini A, Tittarelli A, Bertoldi M, Crosignani P, et al. Human health risk in relation to air quality in two municipalities in an industrialized area of Northern Italy. *Environ Res.* 2011;**111**(8):1321–7. doi: 10.1016/j.envres.2011.06.012. [PubMed: 21764052].
3. Matus K, Nam KM, Selin NE, Lamsal LN, Reilly JM, Paltsev S. Health damages from air pollution in China. *Glob Environ Change.* 2012;**22**(1):55–66. doi: 10.1016/j.gloenvcha.2011.08.006.
4. Chafe ZA, Brauer M, Klimont Z, Van Dingenen R, Mehta S, Rao S, et al. Household cooking with solid fuels contributes to ambient PM2.5 air pollution and the burden of disease. *Environ Health Perspect.* 2014;**122**(12):1314.
5. Zanobetti A, Dominici F, Wang Y, Schwartz JD. A national case-crossover analysis of the short-term effect of PM2.5 on hospitalizations and mortality in subjects with diabetes and neurological disorders. *Environ Health.* 2014;**13**(1):38. doi: 10.1186/1476-069X-13-38. [PubMed: 24886318].
6. Dong M, Yang D, Kuang Y, He D, Erdal S, Kenski D. PM2.5 concentration prediction using hidden semi-Markov model-based times series data mining. *Expert Systems Appl.* 2009;**36**(5):9046–55. doi: 10.1016/j.eswa.2008.12.017.
7. Buczynska AJ, Krata A, Van Grieken R, Brown A, Polezer G, De Wael K, et al. Composition of PM2.5 and PM1 on high and low pollution event days and its relation to indoor air quality in a home for the elderly. *Sci Total Environ.* 2014;**490**:134–43. doi: 10.1016/j.scitotenv.2014.04.102. [PubMed: 24852612].
8. Hoek G, Krishnan RM, Beelen R, Peters A, Ostro B, Brunekreef B, et al. Long-term air pollution exposure and cardio- respiratory mortality: a review. *Environ Health.* 2013;**12**(1):43. doi: 10.1186/1476-069X-12-43. [PubMed: 23714370].
9. Atash F. The deterioration of urban environments in developing countries: Mitigating the air pollution crisis in Tehran, Iran. *Cities.* 2007;**24**(6):399–409. doi: 10.1016/j.cities.2007.04.001.
10. Mawer C. Air pollution in Iran. *BMJ.* 2014;**348**:g1586. doi: 10.1136/bmj.g1586. [PubMed: 24554183].
11. Sun W, Zhang H, Palazoglu A, Singh A, Zhang W, Liu S. Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in Northern California. *Sci Total Environ.* 2013;**443**:93–103. doi: 10.1016/j.scitotenv.2012.10.070.
12. Reich SL, Gomez DR, Dawidowski LE. Artificial neural network for the identification of unknown air pollution sources. *Atmospher Environ.* 1999;**33**(18):3045–52. doi: 10.1016/s1352-2310(98)00418-x.
13. McKendry IG. Evaluation of Artificial Neural Networks for Fine Particulate Pollution (PM10and PM2.5) Forecasting. *J Air Waste Manag Assoc.* 2002;**52**(9):1096–101. doi: 10.1080/10473289.2002.10470836.
14. Domanska D, Wojtylak M. Application of fuzzy time series models for forecasting pollution concentrations. *Expert Systems Appl.* 2012;**39**(9):7673–9. doi: 10.1016/j.eswa.2012.01.023.
15. Zucchini W, MacDonald I. Hidden Markov models for time series: an introduction using R. CRC press; 2009.
16. EPA . NAAQS Table 2015. Available from: https://www.epa.gov/criteria-air-pollutants/naaqs-table.
17. Berhane K, Gauderman WJ, Stram DO, Thomas DC. Statistical Issues in Studies of the Long-Term Effects of Air Pollution: The Southern California Children?s Health Study. *Statistic Sci.* 2004;**19**(3):414–49. doi: 10.1214/088342304000000413.
18. Zhang H, Zhang W, Palazoglu A, Sun W. Prediction of ozone levels using a Hidden Markov Model (HMM) with Gamma distribution. *Atmospher Environ.* 2012;**62**:64–73. doi: 10.1016/j.atmosenv.2012.08.008.
19. Álvarez LJ, Rodrigues ER. Trans-dimensional MCMC algorithm to estimate the order of a Markov chain: an application to ozone peaks in Mexico City. *Int J Pure Appl Math.* 2008;**48**:315–31.
20. Sadeghifar M, Seyed-Tabib M, Haji-Maghsoudi S, Noemani K, Aalipur-Byrgany F. The application of Poisson hidden Markov model to forecasting new cases of congenital hypothyroidism in Khuzestan province. *Journal of Biostatistics and Epidemiology.* 2016;**2**(1):14–9.
21. Agresti A, Kateri M. Categorical data analysis. Springer; 2011.
22. Halek F, Kianpour-Rad M, Kavousirahim A. Seasonal variation in ambient PM mass and number concentrations (case study: Tehran, Iran). *Environ Monitor Assess.* 2010;**169**(1):501–7.
23. Kulshrestha A, Satsangi PG, Masih J, Taneja A. Metal concentration of PM(2.5) and PM(10) particles and seasonal variations in urban and rural environment of Agra, India. *Sci Total Environ.* 2009;**407**(24):6196–204. doi: 10.1016/j.scitotenv.2009.08.050. [PubMed: 19793609].
24. Zhao X, Zhang X, Xu X, Xu J, Meng W, Pu W. Seasonal and diurnal variations of ambient PM2.5 concentration in urban and rural environments in Beijing. *Atmospher Environ.* 2009;**43**(18):2893–900. doi: 10.1016/j.atmosenv.2009.03.009.
25. Murakami J. Bayesian posterior mean estimates for Poisson hidden Markov models. *Comput Statistics Data Analysis.* 2009;**53**(4):941–55. doi: 10.1016/j.csda.2008.11.012.
26. Perez P, Gramsch E. Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes. *Atmospher Environ.* 2016;**124**:22–7. doi: 10.1016/j.atmosenv.2015.11.016.

27. Sun W, Zhang H, Palazoglu A. Prediction of 8 h-average ozone concentration using a supervised hidden Markov model combined with generalized linear models. *Atmospher Environ.* 2013;**81**:199–208. doi: 10.1016/j.atmosenv.2013.09.014.