Environmental Modelling and Software 107 (2018) 175–185

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Long-term forecasting of nitrogen dioxide ambient levels in metropolitan areas using the discrete-time Markov model

Asaf Nebenzal[a], Barak Fishbain[b,*]

[a] Department of Mathematics, Technion – Israel Institute of Technology, Haifa, 320003, Israel
[b] Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa, 320003, Israel

ARTICLE INFO

Keywords:
Air pollution modeling
Discrete-time Markov model
Long-term forecasting
Modeling
Risk assessment
Nitrogen dioxide (NO_2)

ABSTRACT

Air pollution management and control are key factors in maintaining sustainable societies. Air quality forecasting may significantly advance these tasks. While short-term forecasting, a few days into the future, is a well-established research domain, there is no method for long-term forecasting (e.g., the pollution level distribution in the upcoming months or years). This paper introduces and defines *long-term* air pollution forecasting, where *long-term* refers to estimating pollution levels in the next few months or years. A Discrete-Time-Markov-based model for forecasting ambient nitrogen oxides patterns is presented. The model accurately forecasts overall pollution level distributions, and the expectancy for tomorrow's pollution level given today's level, based on longitudinal historical data. It thus characterizes the temporal behavior of pollution. The model was applied to five distinctive regions in Israel and Australia and was compared against several forecasting methods and was shown to provide better results with a relatively lower total error rate.

1. Introduction

Air pollution is a significant risk factor for multiple health issues and impacts negatively on the environment (Venkatadri and Rao, 2014; Heroux et al., 2015). The estimation of future air pollution levels is one of the major tools for air pollution management and control (Ott, 1978). However, predicting future pollution levels may fail to be sufficient. Recent findings (Di et al., 2017; Zhang, 2017) suggest that pollution levels in general not only damage the human body and its physiological processes but daily differences in pollution are also key factors. This underscores the urgent need for forecasting not only pollution levels but also its daily transitions. Although a wide variety of techniques have been developed for forecasting air quality, none have addressed long-term (i.e., 12 months) or the daily gradients.

Broadly speaking, forecasting methods can be categorized into empirical, statistical and deterministic approaches (Zhang et al., 2012). The empirical approach is based on observations of several past data points. It covers the persistence (Garner and Thompson, 2012), climatology (Broday et al., 2012; Dye et al., 2003), moving average (Gardner, 1985), exponential smoothing (Gardner, 1985) and double exponential smoothing (Holt-Winters method) (Gardner, 1985) methods.

The statistical approaches consist of Classification and Regression Trees, CART, (Athanasiadis et al., 2006), Regression Analysis, RA, (Agirre-Basurko et al., 2006; Tranmer and Elliot, 2008; Donnelly et al., 2015), Artificial Neural Networks, ANN, (Ibarra-Berastegi et al., 2008), and a combination of these methods (Gong and Ordieres-Meré, 2016). These methods are based on finding correlations in the historical data and exploiting them to estimate future pollution levels. This is typically done on short historical windows.

Deterministic models forecast air pollution through mathematical formulations and simulations of emissions, transport, diffusion, transformation, and removal of air pollution such as meteorological and emission models (Finardi et al., 2008; Manders et al., 2009; Corani and Scanagatta, 2016).

While this extensive body of work addresses air pollution forecasting, all of them only provide short-term prediction, typically 24–72 h. Clearly, however, any planning and implementation of air pollution strategies can benefit from long-term forecasting. There is no consensual definition of long-term air-pollution forecasting in the literature. Long-term forecasting makes little sense if the purpose is to predict actual values (e.g., what will the pollution levels be on a specific date in the distant future, say in a year from now). For this reason, we define long-*term* forecasting in terms of air quality behavior; i.e., *distribution* and *transitions*, in the next few months or years.

The Discrete-Time Markov Chain (DTMC) is a well-known probabilistic model used to describe and analyze stochastic processes (Ye, 2000). DTMC makes it possible to predict and investigate the behavior of a given process and has been proved to be effective and useful in a

* Corresponding author.
E-mail addresses: asaf.n@technion.ac.il (A. Nebenzal), fishbain@technion.ac.il (B. Fishbain).

https://doi.org/10.1016/j.envsoft.2018.06.001
Received 14 January 2018; Accepted 1 June 2018
Available online 04 June 2018
1364-8152/

Software availability

Name of software Long-Term Air-Pollution Forecasting
Developer A. Nebenzal, Dept. of Applied Math, Technion – Israel Institute of Technology, Email: asaf.n@technion.ac.il
Year first available 2018
Software required Matlab 2015 (and up)
Program language Matlab scripting language
Program size 100 KBytes
Availability Source code available at: http://fishbain.net.technion.ac.il

wide range of areas such as chemistry (Vereecken et al., 1997), medicine (Zipkin et al., 2010) and computer science (Almasizadeh and Azgomi, 2008). In the field of air pollution, DTMC has been used for estimating the probability that a daily maximum ozone level would exceed a predefined threshold in a metropolitan area of Mexico City (Rodrigues and Achcar, 2013) and for developing a transportation emission model (Crisostomi and Kirkland, 2011).

One form of noxious air pollution comes from nitrogen-dioxide gas or NO_2, which is known to be harmful to human health. Short-term exposure may cause respiratory problems such as wheezing, coughing, bronchitis and to aggravate respiratory diseases such as asthma (Koenig, 1999; Clark and Demers, 2010). Long-term exposure may also increase the risk of asthma (United States EPA, 2017). NO_2 originates primarily from combustion sources in vehicles and industrial plants. It is also the precursor of several harmful secondary air pollutants, such as ground-level ozone, and nitrates which contribute to increased respirable particle levels in the atmosphere and plays a role in the formation of acid rain. Thus, analyzing and understanding NO_2 pollution level behavior can be used as an indicator for general pollution conditions and as an exposure assessment tool in a region (Clark et al., 2014). NO_2 has highly spatiotemporal variability, especially in urban areas (Mayer, 1999). NO_2 has a short lifetime (hours), and its temporal (i.e., diurnal, weekly and seasonal) cycles are affected by many factors including traffic and industry emissions, metrological conditions, and chemical processes (Liu et al., 2018; O'Leary and Lemke, 2014).

Here we implemented an Air Pollution Discrete-Time Markov Chain (AP-DTMC) model for long-term forecasting of NO_2 levels. The AP-DTMC is harnessed here to provide a new means of assessing and more importantly estimating future air-pollution characteristics and pollution characteristics; namely: (a) the number of days within a time period, e.g. a year, pollution levels will exceed a predefined threshold; (b) the pollution distribution, i.e., the fraction of time the pollution will have a specific level within this time period, which was defined here as the stationary distribution; (c) the probability of observing any degree of pollution in the proceeding time period, given the current level, or brief historical data. This set of probabilities forms the transition matrix. Both the stationary distribution and the transition matrix constitute novel tools for analyzing air quality and more importantly its behavior. This model was compared to well-known long-term forecasting methods such as multiple linear regression (Zeger et al., 2000; Schwartz et al., 2002; Sousa et al., 2007), moving average (Schwartz et al., 1996), exponential and double smoothing (Schwartz, 1991, 1994; Kandya et al., 2009) and the persistence method (Kang et al., 2005) and was shown to be superior in terms of estimation error.

2. Method

2.1. Study area

This study is based on NO_2 pollution levels in five different regions. Four were situated along the Israeli coastline and included the Haifa metropolitan area, the urban coastline between Haifa and Tel-Aviv (Sharon region), the Tel-Aviv metropolitan area and the Israeli Southern Coastal Plain (SCP), see Fig. 1a; and the fifth in Queensland, Australia. These regions were selected to present both geographically close and remote locations to better assess the ability of the AP-DTMC to predict future air pollution levels based on previous measurements regardless of the chosen location. The sensor locations in all are presented in Fig. 1 b–f.

The Haifa metropolitan area is about 250 km^2 and has a population of roughly 500,000. It is located along the northern part of the Israeli Mediterranean coastline. The area includes the city of Haifa and its satellite towns. The Haifa Bay area in the center of the region is one of the major industrial complexes in Israel and houses petrochemical industries and refineries. The Haifa seaport is also part of this area. The Haifa bay area is the main transportation hub between central and northern Israel. Overall these features generate a complex pollution pattern composed of heavy industry, a high transportation volume, and a dense population.

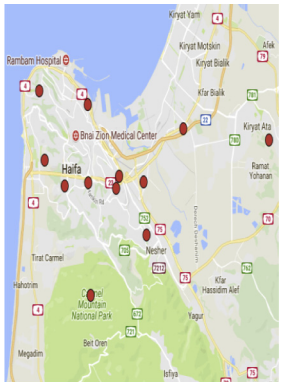
The Sharon plain lies between the Mediterranean Sea to the west and the Samarian hills to the east and extends from Mount Carmel in the north to the Yarkon river in the south. It covers about 650 km^2 and has a population of about 800,000 residing primarily in 4 mid-sized cities (Netanya, Herzliya, Hadera, and Ra'anana). The largest combustion power plant in Israel is located on the Hadera coast and represents about 20% of the Israel Electric Corporation's total generation capacity. This region is relatively sparsely populated with medium-low inter-city traffic activity.

The most densely populated and traffic-heavy region in Israel is the Tel-Aviv metropolitan area. It covers about 150 km^2 and has a population of roughly 2,000,000. It includes Tel-Aviv, the most highly populated city in Israel, and other large municipalities such as Ramat-Gan, Petah-Tikva, Rishon LeTsiyon, and Holon. Its borders, as defined here were the Mediterranean Sea to the west, Petah-Tikva to the east, Ramat-Hasharon to the north and Rishon LeTsiyon to the south.

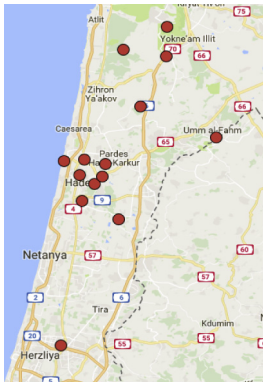
The southern coastal plain begins at the city of Yavne in the north and extends all the way to the Gaza Strip in the south, and from the Mediterranean Sea to the west and the Shfela hills to the east. It covers about 820 km^2. The main cities in this region are Ashdod, Ashkelon, and Rehovot, and has a population of about 400,000. It incorporates the Ashdod seaport, which is the main maritime port of entry in Israel and, at its southwest border, the Rutenberg Power Station, the second largest power station in Israel.

Brisbane is the capital and most populous city of the Australian state of Queensland, and the third most populous city in Australia with a population of about 2,200,000. The current study focused on center-east Brisbane, which covers about 260 km^2 from St Lucia to the west and Raby Bay to the east and from Brisbane Airport to the north and Salisbury to the south. Most of the area is relatively densely populated and includes residential and industrial zones including the Brisbane air and sea ports.

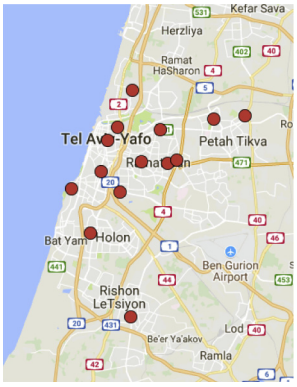
(a) Coastline of Israel (scale: 1:64,000,000)



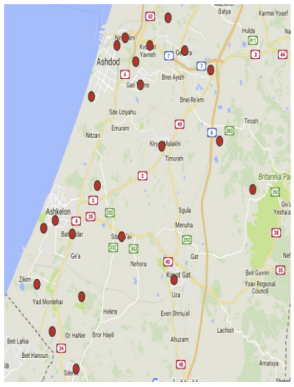
(b) Haifa Metropolitan Area (scale 1:500,000)



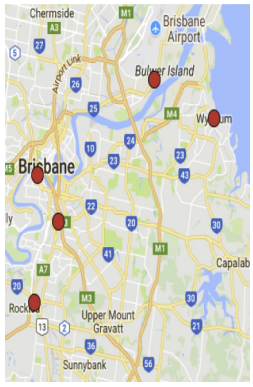
(c) Sharon Region (scale 1:1,000,000)



(d) Tel-Aviv Metropolitan Area (scale 1:500,000)



(e) SCP (scale 1:750,000)



(f) Central-East Brisbane (scale 1:500,000)

Fig. 1. Focal Areas. Red circles show Air Quality Monitoring Stations North is at the top of all maps. (source: Google maps).

2.2. Data

NO_2 data collected between 2010 and 2015 were used. This lengthy temporal window insured that both seasonality and long-term trends were present in the data and thus accounted for. In Israel, the data were collected by the Israeli Ministry of Environmental Protection (I-MEP) and Non-Government Organizations (NGOs). The raw data acquired from I-MEP went through validation and data correction processes (Yuval and Broday, 2006). In all the Air Quality Monitoring (AQM) stations, the data are saved as half-hourly mean concentrations, in ppb units. In Australia, the data were gathered by the Department of Environment and Heritage Protection, Queensland Government (Environment and Heritage Protection, Queensland, 2015). These data are saved as one-hour averages in ppm units. In both Israel and Australia, the NO_2 levels were measured by chemiluminescence-based measuring equipment (Vacher et al., 2018), which was calibrated and maintained according to ISO-17025 laboratory equipment standards (ISO, 2005).

The collected data were divided into historical (i.e., training) and validation sets. The former consisted of NO_2 measurements acquired from 2010 to 2014, and the latter comprised NO_2 data collected throughout 2015. Over the years there have been some changes in the number of AQMs: in some cases, stations have stopped operating, and in other cases, new stations have been set up. We only chose stations that were active during the entire period to make an accurate comparison between the training and test sets. The number of stations used in each region is as follows: Haifa - 14 (out of 16 active AQMs overall in both periods); Sharon 13 (18); Tel-Aviv - 14 (15); SCP 20 (22); and center-east Brisbane - 5 (7).

In this study, the pollution levels at 6:00 p.m. were used as the daily *characteristic measurement*. Six p.m. was chosen since in most locations the level of pollution is at its peak or close to that at this time. This is because one of the most significant sources of NO_2 is transportation, and at 6:00 p.m., the highest quantity accumulates before it reacts with other substances or disperses. There are at least three advantages for using this time of day rather than any other time or an average. The first is that a single point of time serves as a better *daily characteristic measurement*, due to the high diurnal variability of NO_2 (Mayer, 1999). Second, daily-max values and percentiles are features of interest in regulations and air quality control policies (EPA, 2018), and third, this is consistent with several studies designed to analyze pollution temporal patterns; e.g., (Broday et al., 2012). Note that this model was tested at different times of day (e.g., AM peak and night time) as the characteristic measurement as well as hourly and daily averages. In all these experiments the same promising results were obtained.

The representative pollution level, for each region was thus the average NO_2 level calculated over all measurements acquired by the AQM stations in that region at 6:00 p.m. AQM stations are categorized by purpose, such as monitoring transportation or industrial pollutants. The specific purpose of an AQM determines which pollutants it monitors and its physical configuration. Since we were interested in the representative level at 6:00 p.m., all AQMs were included in the calculations. Although there may be considerable differences between a busy street and a station located in a park nearby, the characteristic levels of pollution of the entire area were taken into consideration.

2.3. Air pollution discrete-time Markov model

Let $a_t^r \in A^r$ be region r's daily *NO_2 characteristic measurement* at day t, where $t \in T = \{1,2,3...\}$ is the day index and A^r is the entire region's time series consisting of the time series $\{a_t^r\}$. Let S be a set, $\{s_i\}$, of pollution severity labels, e.g., {"low"; "high"} or {"low"; "medium-low"; "medium-high"; "high"}. The number of labels in S is denoted by K; i.e., $|S| = K$. The label, $s_t^r \in S$ is the severity tag associated with a_t^r. The label s_t^r can be determined in several ways. It can be set according to a standard, such as the Air Quality Index standard (AQI) (Cheng et al., 2007; Bishoi et al., 2009), or with respect to other characteristic measurements obtained in the same region; i.e., with respect to A^r, or with respect to all measurements acquired on the national level. Here we use a regional labeling system and label each region's measurements independently; i.e., w.r.t A^r for each r. Similar to the definition of A^r, S^r is the labels' time series of region r. It is assumed that S^r fulfills the *Markov property*; i.e., the conditional probability distribution of the next label, s_{t+1}^r, depends solely on the present label s_t^r:

$$\boldsymbol{P}\{\boldsymbol{s}_{t+1}^r|\boldsymbol{s}_t^r,\ \boldsymbol{s}_{t-1}^r,\ldots,\ \boldsymbol{s}_1^r,\ \boldsymbol{s}_0^r\} = \boldsymbol{P}\{\boldsymbol{s}_{t+1}^r|\boldsymbol{s}_t^r\} \tag{1}$$

This assumption stems from the typical gradual behavior of air pollution and meteorology. Both phenomena tend to show stability over a cycle of a few days; e.g., diurnal and weekly patterns (Wilby and Tomlinson, 2000; Broday et al., 2012). Due to this slow change, the assumption is that the previous day is sufficient to indicate the next day's pollution level and thus air pollution fulfills the Markov property. Sections 3.3 and 3.4 show that this assumption indeed holds for all the regions studied here.

For two labels $s_i, s_j \in S$, let P_{ij} denote the probability of shifting to label s_j, given the current label s_i:

$$\boldsymbol{P_{ij}} = \boldsymbol{P}\{\boldsymbol{s}_{t+1}^r = \boldsymbol{s_j}|\boldsymbol{s}_t^r = \boldsymbol{s_i}\},\quad \boldsymbol{s_i}, \boldsymbol{s_j} \in \boldsymbol{S} \tag{2}$$

For K labels in S, the AP-DTMC transition probability matrix, $\boldsymbol{P}$, describes all probabilities to shift from any label s_i to any label s_j, and is denoted by:

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P_{11}} & \boldsymbol{P_{12}} & \ldots & \boldsymbol{P_{1K}} \\ \boldsymbol{P_{21}} & \boldsymbol{P_{22}} & \ldots & \boldsymbol{P_{2K}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{P_{K1}} & \boldsymbol{P_{K2}} & \ldots & \boldsymbol{P_{KK}} \end{bmatrix} \tag{3}$$

The probabilities populating matrix $\boldsymbol{P}$ are derived from the historical data, simply by counting how many times, out of the total days, s_j followed s_i . Utilizing the entire historical database, which incorporates both major pollution gradients as well as moderate changes; i.e., seasonality, allows the method to account, to some extent, for these changes in the analysis. Each row in the matrix is a probability vector; hence $\sum_{i=1}^{K} P_{ij} = 1$ for each j. The *n-day* transition probability matrix, P_{ij}^n, describes the probability of having label s_j at time $t + n$, given the label s_i at time t:

$$\boldsymbol{P_{ij}^n} = \{\boldsymbol{s}_{t+n}^r = \boldsymbol{s_j}|\boldsymbol{s}_t^r = \boldsymbol{s_i}\} = \boldsymbol{P^n} \tag{4}$$

Based on these definitions, a label time series, S^r, is an *ergodic* Markov chain. Hence, it is irreducible, positive recurrent and aperiodic (Pakes, 1969). This also means that the rows of $\boldsymbol{P^n}$, for $\boldsymbol{n}\gg$, are the same and represent the stationary distribution, π_j, of the AP-DTMC:

$$\boldsymbol{\pi_j} = \lim_{\boldsymbol{n}\to\infty} \boldsymbol{P_{ij}^n} \tag{5}$$

Recapitulating the notion above, the transition matrix, Eq. (3), expresses the probability of having label s_j, at the next time step $t+1$, provided that the current time step label is s_j. This can be extended to represent more complicated connections, accounting for a longer history, say past M *days*:

$$\boldsymbol{P}\{\boldsymbol{s}_{t+1}^r|\boldsymbol{s}_t^r,\ \boldsymbol{s}_{t-1}^r,\ \boldsymbol{s}_{t-2}^r,\cdots,\boldsymbol{s}_{t-M+1}^r\} s.\ t.\ \{\boldsymbol{s}_t^r,\ \boldsymbol{s}_{t-1}^r,\ \boldsymbol{s}_{t-2}^r,\cdots,\boldsymbol{s}_{t-M+1}^r\} \in \boldsymbol{S} \tag{6}$$

2.4. Applications

2.4.1. Threshold analysis

In many air pollution mitigation plans, the percentage of days where NO_2 levels exceeds a given threshold is key. The threshold can be set to any value, such as the level considered by the authorities as hazardous, or a specific air-control policy goal (Rodrigues and Achcar, 2013; Gong and Ordieres-Meré, 2016). In this case, two labels are considered in the Markov model: i.e., $S = \{$*"low"; "high"*$\}$, which correspond to below and above the threshold. Thus, the transition probability matrix (Eq. (3)) is of the form:

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P_{L,L}} & \boldsymbol{P_{L,H}} \\ \boldsymbol{P_{H,L}} & \boldsymbol{P_{H,H}} \end{bmatrix} \tag{7}$$

The analysis here would result in obtaining the stationary distribution π_j, Eq. (5), which represents the future, estimated, distribution of time between "low" and "high" pollution levels. This information can be used for example to apply a policy whose goal is to achieve a better distribution than expected by the model (thus, a higher portion of "low" days) in the upcoming year.

2.4.2. Stationary distribution

The stationary distribution, π_j, facilitates the investigation of pollution signal behavior in the long run. In particular, π_j can serve to identify the fraction of time AP-DTMC assumes each label. This information is extremely helpful in designing and applying air pollution mitigation policies where decision makers can use the expected distribution to set lower goals; i.e., a lower proportion of days with high pollution levels.

The stationary distribution also facilitates the Air Quality Index (AQI) measure (Kyrkilis et al., 2007), which is commonly used for assessing air quality in general and personal exposure assessment in particular. This is because personal exposure is affected by many factors; hence, the variance of the dose response function is typically high and dominates the attributed relative risks/hazard ratio results regardless of sensor accuracy (Lebret, 1990; Zeger et al., 2000; Jerrett et al., 2005). Therefore, a common practice for estimating individual exposure is to use a coarse scale such as the AQI, rather than the sensors' actual readouts. The stationary distribution enables these types of analyses in present and future times.

2.4.3. Transition analysis

By applying matrix $\boldsymbol{P}$, the AP-DTMC provides the probabilities for tomorrow's pollution's label given today's label. Essentially, this information constitutes short-term forecasting. Hence, AP-DTMC has a predictive value both in short and long-term forecasting. In short-term forecasting, AP-DTMC is considered a good benchmark for tomorrow's forecast (Garner and Thompson, 2012). The AP-DTMC model provides two distinct observations: *daily stationary distribution levels* (see Section 2.4.2), and the *transition probabilities*. Whereas the former describes the distribution of pollution levels in the upcoming year; i.e., the fraction of time pollution presents a specific level within a time period, the latter provides an indication of the pollution level tomorrow, given today's pollution level; i.e., the transition behavior. Since the daily differences also present a health risk (Di et al., 2017; Zhang, 2017), this information is highly important and is, to date, only provided by the AP-DTMC model.

2.4.4. Higher order Markov chain

Matrix $\boldsymbol{P}$ can describe multiple day forecasts by constructing a higher dimensional transition matrix $\boldsymbol{P}$. This is done by formulating Eq. (6) so that the pollution label of Day $t+1$ is predicted based on the past M days $\{t,\ t-1,\cdots,t-M+1\}$. In the case presented here, the model describes the pollution behavior in a three-day sequence. Thus, $\boldsymbol{P}$, in this case, presents the probabilities for all possible pollution labels,

$P(s^r_{t+1})$ on day $t+1$, given the pollution level at day t, and day t-1. For four pollution levels, $\boldsymbol{P}$ is a 16 × 4 matrix.

3. Results and discussion

The AP-DTMC model was applied on the NO_2 datasets described in Section 2.2. The data were divided into a training set, consisted of the data acquired 2010–2014, and a validation test set, which consisted of the 2015 data. The performance of the AP-DTMC algorithm in the applications (Section 2.4) was compared against the results obtained by multiple linear regression (Zeger et al., 2000; Schwartz et al., 2002; Sousa et al., 2007), moving average (Schwartz et al., 1996), exponential smoothing (with $\alpha = 0.80$) (Schwartz, 1991, 1994), double exponential smoothing, Holt's method (with $\alpha = 0.80$, $\gamma = 0.2$) (Kandya et al., 2009), and the persistence method (Kang et al., 2005). The α and γ parameters for the exponential smoothing and the double exponential smoothing methods were selected so that the best prediction results would be obtained.

3.1. Threshold analysis

In threshold analysis, the goal is to estimate the number of days pollution will exceed a predefined threshold. For purposes of comparison, the Tel-Aviv metropolitan area, with its 14 AQM stations was chosen (see Section 2.1). The results of all other study areas are detailed in Appendix A and are consistent with the results of Tel-Aviv region. For the performance analysis, both the transition matrix (Eq. (3)) and the stationary distribution (Eq. (5)) were evaluated. All training data acquired throughout 2010–2014 were used for this analysis. The threshold was arbitrarily set so that the highest 20% of the samples in the years 2010–2014 would be considered above the threshold. In the Tel-Aviv area, the actual value was 35 ppb.

The transition matrix analysis is detailed in Table 1, which depicts the actual (ground truth) and predicted values of the 2015 transition matrix $\boldsymbol{P}$; i.e., {*low* → *low*; *low* → *high*; *high* → *low*; *high* → *high*}. The table shows the results derived by the different methods and summarizes the different methods' total error, which is the sum of the absolute error of each label. As can be seen from Table 1, all methods predicted the probabilities for the transition from the 'low' to 'low' state and from 'low' to 'high' fairly accurately but performed less well when predicting the other two transitions. The results show that the AP-DTMC achieved the lowest prediction error, at 8.9%.

The two-state stationary distribution (Eq. (5)) and the total error are presented in Table 2. As shown, the AP-DTMC method outperformed linear regression, exponential smoothing, Holt's method and the Persistence method and was comparable to moving average, which was slightly better.

Combining the results of Tables 1 and 2 shows that the AP-DTMC method provides better predictions than all the other methods. Further, the transition matrix, $\boldsymbol{P}$, and the behavior of the model at infinity; i.e., π_j, are unique observations provided solely by the Markov model. These features can be extracted by using all the other methods as shown here,

Table 1
Forecasting the transition matrix for threshold analysis; i.e., two states - 'L' and 'H', for the selected long-term forecasting methods vs. the ground truth and the total error of each method.

Method	LL	LH	HL	HH	% total error
Ground Truth	0.91	0.09	0.428	0.571	
Markov	0.901	0.098	0.465	0.534	8.94%
Multi Linear Regression	0.906	0.093	0.515	0.484	18.10%
Moving Average	0.902	0.097	0.470	0.529	10.02%
Exponential Smoothing	0.894	0.105	0.459	0.540	9.40%
Holt's method	0.916	0.083	0.474	0.525	10.40%
Persistence	0.911	0.088	0.473	0.526	9.38%

Table 2
The stationary distribution of threshold analysis and the total error of each method.

Method	L	H	% total error
Ground Truth	0.826	0.173	
Markov	0.825	0.174	0.12%
Multi Linear Regression	0.845	0.154	3.88%
Moving Average	0.826	0.173	0.07%
Exponential Smoothing	0.812	0.187	2.80%
Holt's method	0.849	0.150	4.64%
Persistence	0.843	0.157	3.32%

but these insights are inherent to the DTMC method. Based on the findings in Di et al. (2017) that day-to-day *changes* in PM2.5 and ozone ambient concentrations were significantly associated with a higher risk of all causes of mortality at levels well below the current standards, the availability of $\boldsymbol{P}$ and π_j are crucial. The availability of future air pollution transients can allow authorities to set goals and determine policy with respect to these features, and enables individuals at increased risk to reduce or mitigate their exposure (Zhang, 2017).

The AP-DTMC has additional benefits in that it requires less training data than the other methods. For the multiple linear regression, at least 5 years' worth of data (2010–2014) is needed for a simple two-state model, whereas AP-DTMC only requires 3 year data for its forecasting.

3.2. Stationary distribution

To assess the stationary distribution performance of the different methods, $\boldsymbol{P}$ was calculated based on the 2012–2014 historical data, as three years provided sufficient information to obtain adequate results, and π_j was then deduced from $\boldsymbol{P}$. Although the model is applicable for any given number of labels, K (see 2.3), the analysis was conducted for a quantized scale of four pollution labels: 'low', 'medium-low', 'medium-high' and 'high', {L, ML, MH, H}. The mechanism that assigns a label to each measurement can be a threshold defined by an air quality standard or can be set according to a specific official policy. Here, arbitrarily, for each of the study regions the measurements were assigned a label according to their quartile, such that the measurements within the first, second, third and fourth quartiles were labeled 'L', 'ML', 'MH', and 'H' respectively. Hence, the value range that correspond to each of the labels was region dependent. The average values of the measurements assigned to each label, for all study areas, are detailed in Table 3.

The results are detailed in Fig. 2 where the observed values of 2015 are on the left (represented by red circles) and the estimated values are on the right (represented by yellow circles). The gap between the observed and the model's expected values ranged from 0.20% to 8.39%. The average difference was 3.45% with a standard deviation of 2.19%. It is evident that the suggested model predicts the pollution distribution for the different labels with high accuracy in the five distinctive regions suggesting that it can serve as a tool for assessing the general behavior of air pollution by allowing for hotspots (Cole et al., 2005; Kandlikar, 2007) and exposure (Hystad et al., 2011; Adgate et al., 2014) analyses, and thus facilitating the implementation of pollution mitigation policies.

Table 3
NO_2 label averages in ppb.

Score	L	ML	MH	H
Haifa	3	7	10	20
Sharon	3	5	9	19
Tel-Aviv	7	13	20	40
SCP	3	5	8	20
Center-East Brisbane	5	10	14	23

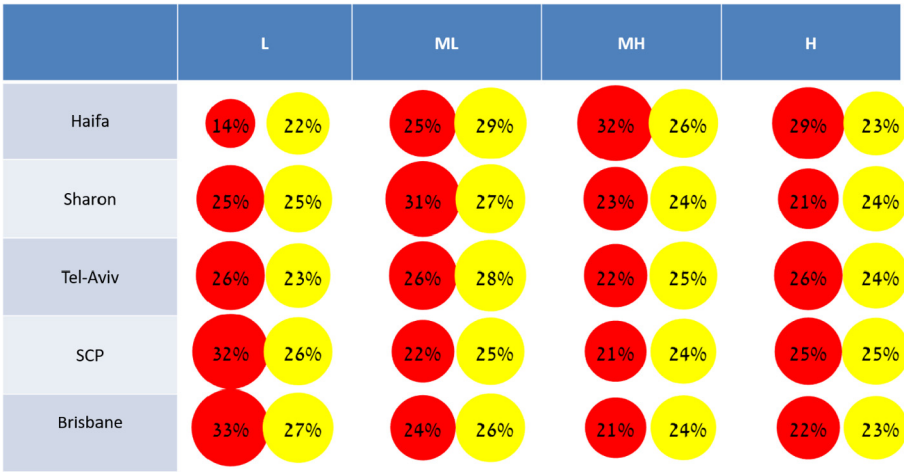


Fig. 2. Pollution stationary distribution. Observed in red versus expected in yellow.

3.3. Transition analysis

For the transition analysis, we estimated the 2015 transition matrix, $\boldsymbol{P}$, describing the probability of the pollution level at day $t+1$, s_{t+1}^r, from the preceding day's pollution level, s_t^r. The 2015 estimation was based on 2012–2014 historical data. Fig. 3 presents the true values (red) and the estimated values (yellow) of the 2015 transition matrix for the Sharon region; the matrices of other regions are depicted in Appendix B and exhibit similar behavior. Table 4 summarizes the mean absolute estimation error (MAE) of $\boldsymbol{P}$ at all the selected regions. It is worthwhile noting that the model also accounts for significant transitions; i.e., from low to high (e.g., the shift from a weekend to a weekday) and high to low (e.g., from a weekday to the weekend). To illustrate, in Fig. 3, the transition matrix of the Sharon region, there are 1%/2% (observed/expected) transitions between 'L' to 'H' and 5%/7% from 'H' to 'L'. This coincides with the fact that there are 52 weekends a year; i.e., ~2% and thus, these transitions may be attributed to weekend vs. weekday pollution behavior.

As can be seen in both Fig. 3 and in Table 4, the model predicts the behavior of pollution distribution with high accuracy. The results also show the tendency of air pollution systems to maintain stability, in that the diagonal values are mostly higher than the off-diagonal ones. Therefore, the highest probability for a specific pollution level on day $t+1$ is to maintain the current pollution level. Thus, if the level was $s_{t+1}^r = 'L'$, on day t, there is a greater likelihood that the level will be 'L' on the following day as well. Although the likelihood of shifting from a specific pollution level to a neighboring level (e.g., from 'ML' on day n to 'L' or 'MH' on day $t+1$) is lower than maintaining the same level, it does present a higher likelihood than shifting two levels, for instance from 'ML' to 'H'. Shifting three levels, which can occur solely in the transition from 'L' to 'H' and vice-versa, is very rare.

The likelihood of maintaining level 'H' on day $t+1$ after having 'H' on day t are more than double from its stationary distribution (Fig. 2); namely, 57% (observed) as compared to 21%, and 57% (expected) as compared to 24%. Similar behavior was observed for both 'ML' and 'MH' levels but at lower statistical significance.

Weather conditions have enormous influence on air pollution. Meteorological parameters such as wind speed and direction, temperature, humidity, rain, clouds and solar radiation clearly affect air quality (Zannetti, 1990). Synoptic-scale weather typically exhibits similar characteristics for several days, such that weather changes occur gradually rather than immediately (Dye, 2010; Broday et al., 2012). The characteristics described above for the transition matrix can partially be attributed to this weather behavior. Both Israel and Brisbane are characterized by unstable weather, especially in the transitional seasons (Goldreich, 2003; Ren and Tong, 2006). However, even in times

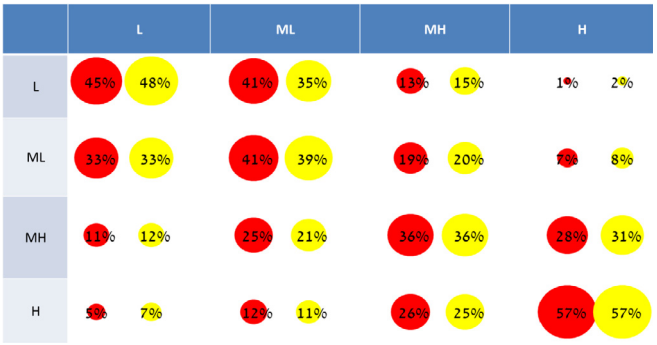


Fig. 3. Transition matrix for the Sharon region. Observed in red versus expected in yellow.

Table 4
Transition matrix Mean Absolute Estimation error.

Region	Haifa	Sharon	Tel-Aviv	SCP	Brisbane
MAE (%)	5.37	1.90	3.66	4.49	2.94

of instability, a synoptic system tends to last a few days; thus, due to the influence of meteorological conditions on air pollution, there is a clear tendency to remain at the same pollution level as shown in the model (Dye, 2010; Broday et al., 2012). Evaluating the results for all study regions (Appendix B) indicates that the model obtained good results in both Israel and Australia. This underscores the ability of the AP-DTMC to predict future air pollution levels based on previous measurements regardless of the chosen location.

Recent findings suggest that mortality as a result of air pollution can be attributed not only to the actual levels but also to transients; i.e., the daily gradients of the pollution signal (Di et al., 2017; Zhang, 2017). The transition analysis, presented here, provides the means to do exactly that.

3.4. Higher order Markov chain

Fig. 4 depicts the probabilities in the transition matrix of the Center-East Brisbane region for a three-day sequence (the results for the other regions appear in Appendix C). Thus, obtaining a specific label s_{t+1}^r, given the labels of the two preceding days, $\{s_t^r,\ s_{t-1}^r\}$.

$$\boldsymbol{P}\{\boldsymbol{s_{t+1}^r}|\boldsymbol{s_t^r}, \boldsymbol{s_{t-1}^r}\} s.\ t.\ \{\boldsymbol{s_{t+1}^r},\ \boldsymbol{s_t^r}, \boldsymbol{s_{t-1}^r}\} \in \boldsymbol{S} \quad (8)$$

Similar to the stationary and transition analyses, the training data

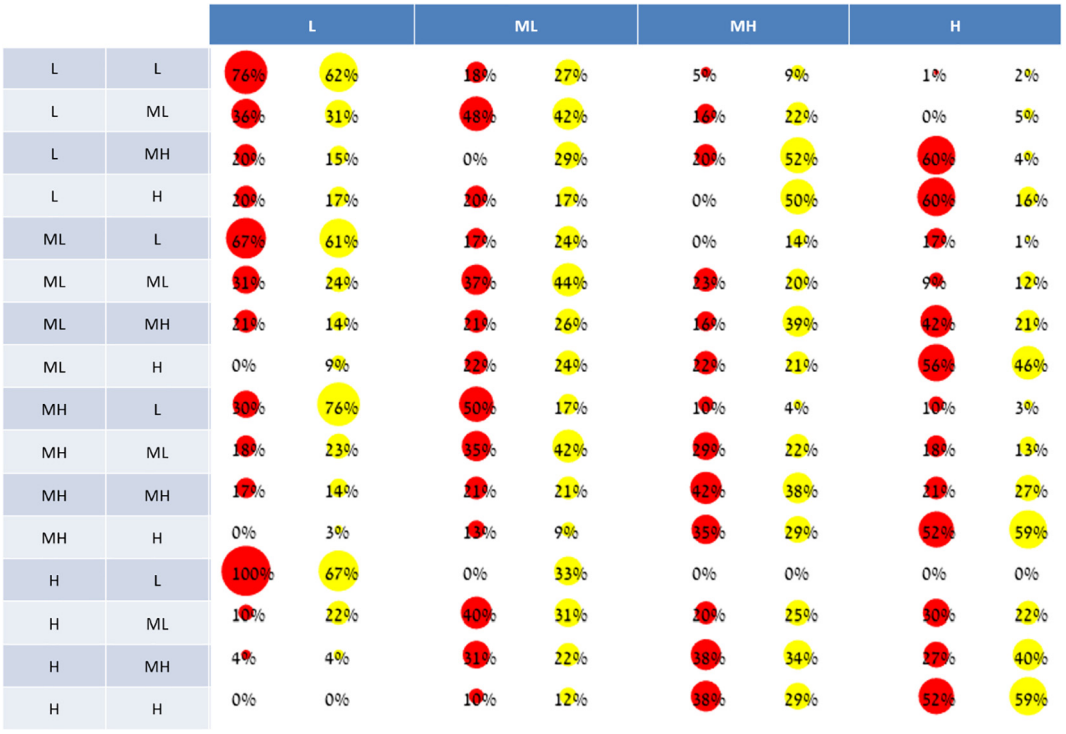


Fig. 4. Three-day sequence transition matrix for Center-East Brisbane. Red circles indicate the observed probability and yellow the expected.

used here were comprised of NO2 measurements acquired throughout 2012–2014. In terms of the three-day transition matrix, sequences showing stability were much more frequent than unstable ones. Tuple sequences of the same level were very frequent. For example {L, L, L} had a 76%/62% (Observed/Expected) likelihood; {H, H, H} 52%/59% (Observed/Expected). Stability at intermediate labels was less frequent: {L, ML, MH} had a 16%/22% (Observed/Expected), {H, MH, ML} had a 31%/21% (Observed/Expected). Unstable sequences were significantly less frequent - {H, H, L} had 0% (both Observed and Expected), {L, L, H} - 1%/2% (Observed/Expected), and {H, L, H} had 0% (both Observed and Expected).

A comparison of the results of section 3.3, transition matrix analysis, and the results of this section suggests that the preceding day has more impact than the day before it This means that the Markov property (Eq. (1)) assumption does hold and that the Markov chain model is suitable for the analysis of long-term air pollution patterns.

4. Conclusion

This paper introduced the concept of long-term air pollution forecasting in terms of pollution distribution and transition probability analyses. These two properties of the pollution signal were derived by modeling the pollution time series as a Markov chain model. The model was applied to NO_2 measurements acquired from 2010 to 2014, in four distinctive regions in Israel and one in Australia. This large number of regions enabled proper evaluation of the model. It was able to accurately predict both the pollution distribution and transition matrices of 2015 based on the historical data. The model was shown to be useful and appropriate for several applications; namely, threshold exceeding estimation, pollution level pattern distribution and transition matrix analysis. As is the case for any other prediction method, the AP-DTMC presented here is vulnerable to dramatic changes in natural or anthropogenic conditions. Future research should address the issue of the rapid detection of changes in transition probabilities, $\boldsymbol{P}$, estimated versus observed and to determine changes in these conditions.

Acknowledgments

This work was partially supported by the Technion Center of Excellence in Exposure Science and Environmental Health (TCEEH), the New-York Metropolitan Research Fund, and the Environmental Health Foundation (EHF). The contribution of Professor Oren Louidor and his course on stochastic models in operations research to this manuscript is gratefully acknowledged. The authors wish to thank Dr. Maya Leventer-Roberts for pointing out the significance to health of transition analyses.

Appendices

Appendix A – Threshold exceeding model results

For the model predicting the percentage of days in a given period exceeding a predefined threshold, the threshold was set as the top 10% highest pollution level for each region. The results are detailed in Fig. 5. The mean differences in all regions was 4.06% (13.38% - Haifa; 1.28% - Sharon; 3.10% - Tel-Aviv; 2.22% - SCP; and 0.28% - Center-East Brisbane) with a standard deviation of 4.52%. The prediction was based on the data acquired throughout 2012–2014 and was validated on data acquired in 2015.

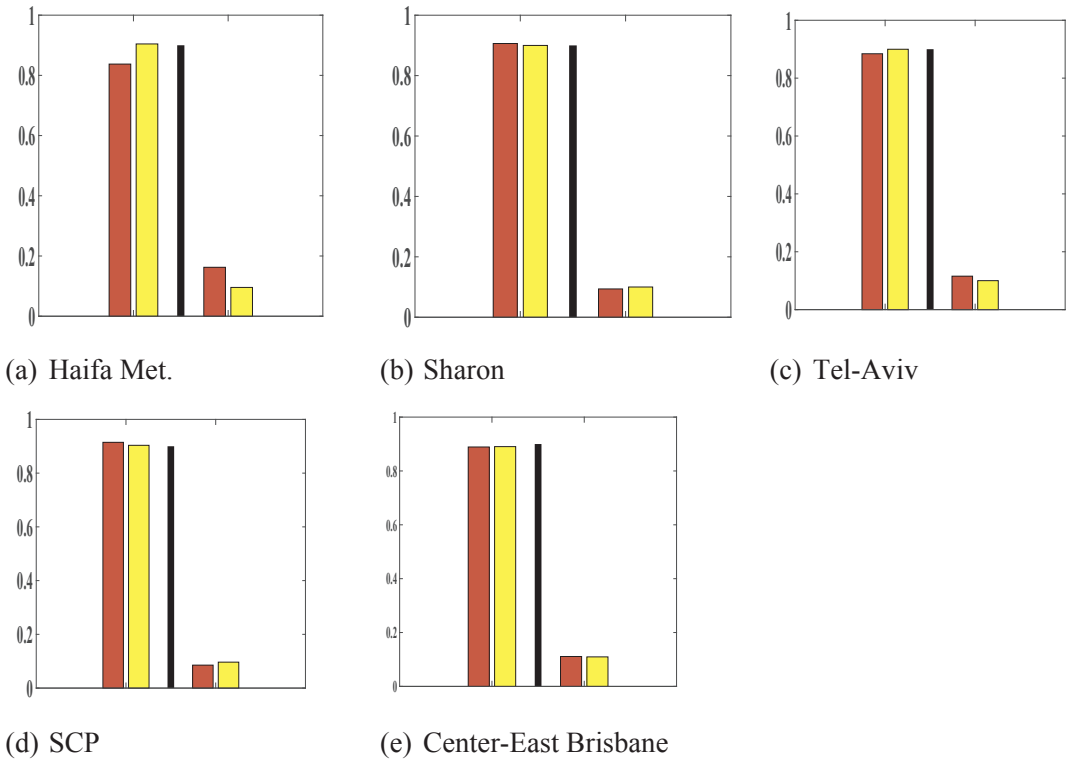


Fig. 5. Threshold exceeding percentages. Observed in red versus expected in yellow. X-axis: the black line divides below from above threshold. Y-axis: the probability of occurrences of these events in a given time period. The threshold was defined as the highest 10% of the total pollution levels.

Appendix B – Transition analysis results

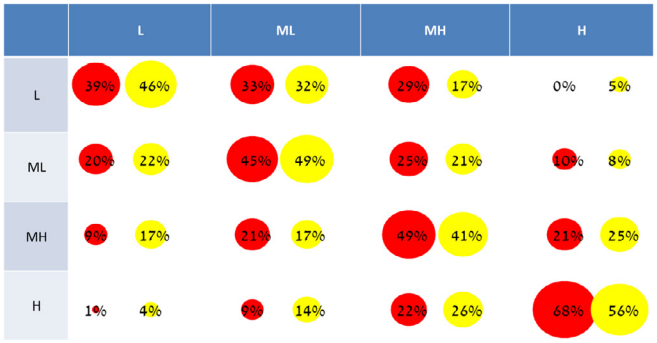


Fig. 6. Daily Transition matrix for the Haifa region.

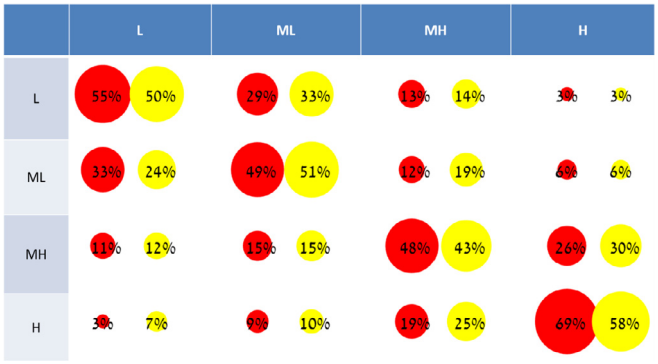


Fig. 7. Daily Transition matrix for the Tel-Aviv metropolitan area.

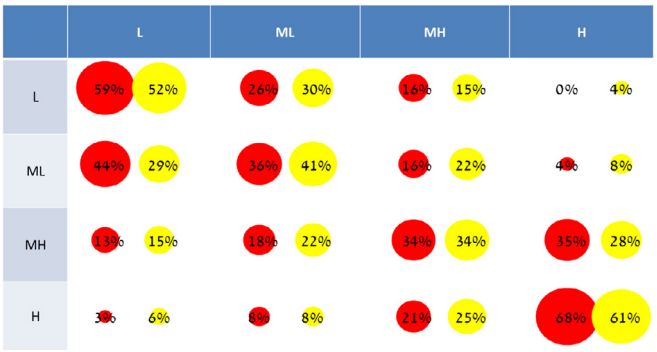


Fig. 8. Daily Transition matrix of the SCP region.

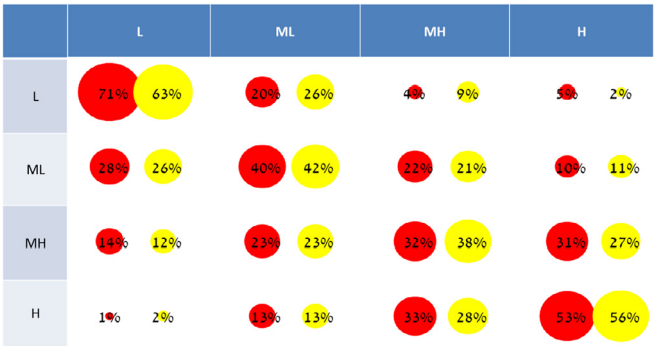


Fig. 9. Daily Transition matrix of the Center-East Brisbane region.

Appendix C – Higher order Markov model results

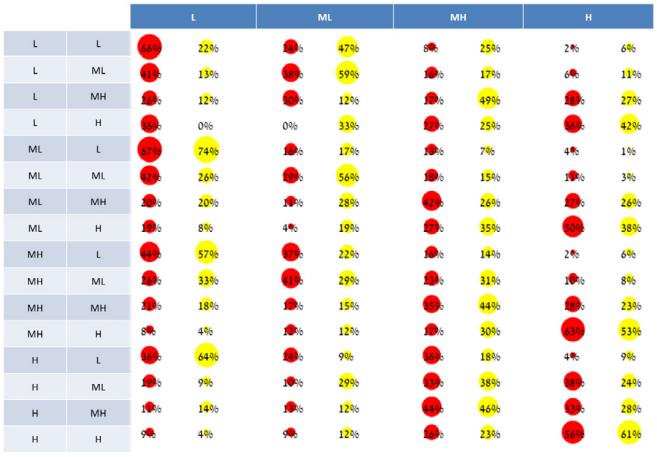


Fig. 10. Three day sequence transition matrix analysis for the Haifa region. Observed measurements from 2015 in red and expected results based on data from 2012 to 2014 in yellow.

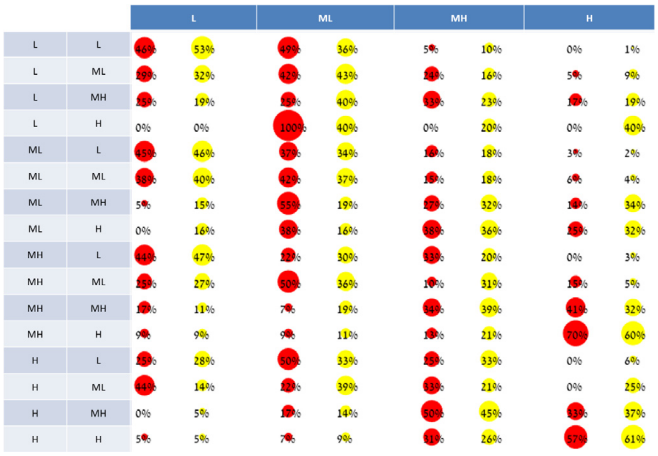


Fig. 11. Three day sequence transition matrix analysis for the Sharon region. Observed measurements from 2015 in red and expected results based on data from 2012 to 2014 in yellow.

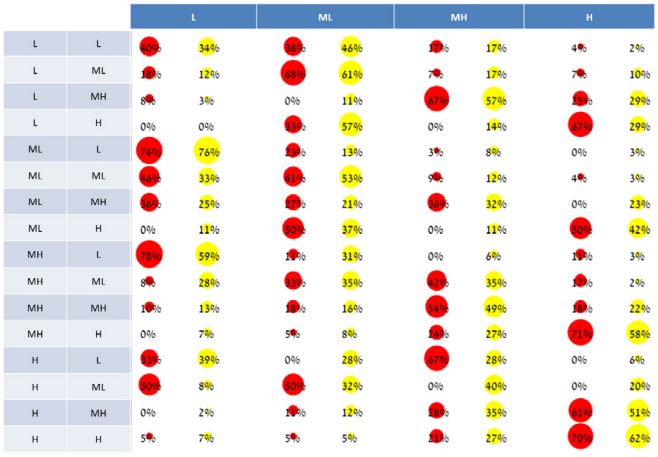


Fig. 12. Three day sequence transition matrix analysis for the Tel-Aviv region. Observed measurements from 2015 in red and expected results based on data from 2012 to 2014 in yellow.

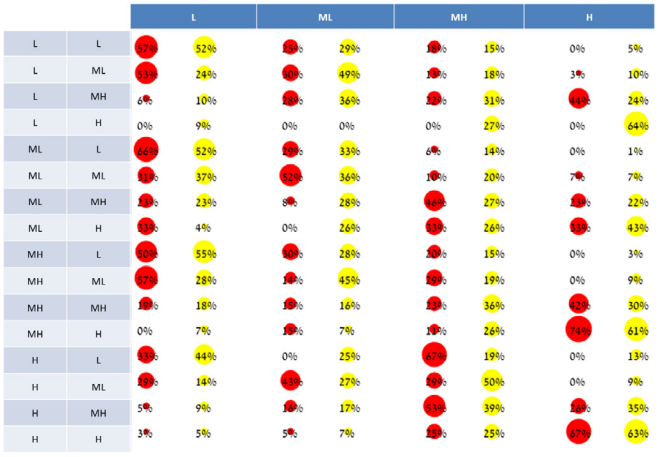


Fig. 13. Three day sequence transition matrix analysis of SCP region. Observed measurements from 2015 in red and expected results based on data from 2012 to 2014 in yellow.

References

Adgate, J.L., Goldstein, B.D., McKenzie, L.M., 2014. Potential public health hazards, exposures and health effects from unconventional natural gas development. Environ. Sci. Technol. 48 (15), 8307–8320. http://dx.doi.org/10.1021/es404621d.

Agirre-Basurko, E., Ibarra-Berastegi, G., Madariaga, I., 2006. Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area. In: Environmental Modelling and Software. Elsevier, pp. 430–446. http://dx.doi.org/10.1016/j.envsoft.2004.07.008.

Almasizadeh, J., Azgomi, M.A., 2008. A new method for modeling and evaluation of the probability of attacker success. In: 2008 International Conference on Security Technology. IEEE, pp. 49–53. http://dx.doi.org/10.1109/SecTech.2008.35.

Athanasiadis, I.N., Athanasiadis, I.N., Karatzas, K.D., Karatzas, K.D., Mitkas, P., Mitkas, P., 2006. Classification techniques for air quality forecasting. In: Fifth ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence, 17th European

Conference on Artificial Intelligence, (August), pp. 1–7.
Bishoi, B., Prakash, A., Jain, V.K., 2009. A comparative study of air quality index based on factor analysis and US-EPA methods for an urban environment. Aerosol and Air Qual. Res. 9 (1), 1–17.
Broday, D.M., Alpert, P., et al., 2012. Exploring the applicability of future air quality predictions based on synoptic system forecasts. Environ. Pollut. 166, 65–74 Elsevier.
Cheng, W.-L., Chen, Y.-S., Zhang, J., Lyons, T.J., Pai, J.-L., Chang, S.-H., 2007. Comparison of the revised air quality index with the PSI and AQI indices. Sci. Total Environ. 382 (2–3), 191–198. http://dx.doi.org/10.1016/J.SCITOTENV.2007.04.036. Elsevier.
Clark, L.P., Millet, D.B., Marshall, J.D., 2014. National patterns in environmental injustice and inequality: outdoor NO2 air pollution in the United States. PLoS One 9 (4), e94431. http://dx.doi.org/10.1371/journal.pone.0094431.
Clark, N.A., Demers, P.a., Karr, C.J., Koehoorn, M., Lencar, C., Tamburic, L., Brauer, M., 2010. Effect of early life exposure to air pollution on development of childhood asthma. Environ. Health Perspect. 118 (2), 284–290. http://dx.doi.org/10.1289/ehp.0900916.
Cole, M.A., Elliott, R.J.R., Shimamoto, K., 2005. Industrial characteristics, environmental regulations and air pollution: an analysis of the UK manufacturing sector. J. Environ. Econ. Manag. 50 (1), 121–143. http://dx.doi.org/10.1016/j.jeem.2004.08.001.
Corani, G., Scanagatta, M., 2016. Air pollution prediction via multi-label classification. Environ. Model. Software 80, 259–264. http://dx.doi.org/10.1016/j.envsoft.2016.02.030. Elsevier.
Crisostomi, E., Kirkland, S., 2011. Markov Chain Based Emissions Models: a Precursor for Green Control. Green IT: Technologies and Applications, (FoE 1999), pp. 381–400.
Di, Q., Dai, L., Wang, Y., Zanobetti, A., Choirat, C., Schwartz, J.D., Dominici, F., 2017. Association of short-term exposure to air pollution with mortality in older adults. JAMA Am. Med. Assoc. 318 (24), 2446. http://dx.doi.org/10.1001/jama.2017.17923.
Donnelly, A., Misstear, B., Broderick, B., 2015. Real time air quality forecasting using integrated parametric and non- parametric regression techniques. Atmos. Environ. 103 (2), 53–65. http://dx.doi.org/10.1016/j.atmosenv.2014.12.011. Elsevier Ltd.
Dye, T.S., 2010. Guidelines for Developing an Air Quality (Ozone and PM2.5) Forecasting Program. US Environmental Protection Agency, Office of Air Quality Planning and Standards, Information Transfer and Program Integration Division, AIRNow Program doi: EPA-456/R-03-002.
EPA, U.S., 2018. Criteria Air Pollutants. Available at: https://www.epa.gov/criteria-air-pollutants.
Finardi, S., De Maria, R., D'Allura, A., Cascone, C., Calori, G., Lollobrigida, F., 2008. A deterministic air quality forecasting system for Torino urban area, Italy. Environ. Model. Software 23 (3), 344–355. http://dx.doi.org/10.1016/J.ENVSOFT.2007.04.001. Elsevier.
Gardner, E.S., 1985. Exponential smoothing: the state of the art. J. Forecast. 4 (1), 1–28.
Garner, G.G., Thompson, A.M., 2012. The value of air quality forecasting in the mid-atlantic region. Weather, Climate, and Society 4 (1), 69–79. http://dx.doi.org/10.1175/WCAS-D-10-05010.1.
Goldreich, Y., 2003. Rainy season (winter and transitional seasons) climate. In: The Climate of Israel. Springer US, Boston, MA, pp. 23–40. http://dx.doi.org/10.1007/978-1-4615-0697-33.
Gong, B., Ordieres-Meré, J., 2016. Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: case study of Hong Kong. Environ. Model. Software 84, 290–303. http://dx.doi.org/10.1016/J.ENVSOFT.2016.06.020. Elsevier.
Heroux, M.E., Anderson, H.R., Atkinson, R., Brunekreef, B., Cohen, A., Forastiere, F., Hurley, F., Katsouyanni, K., Krewski, D., Krzyzanowski, M., Knzli, N., Mills, I., Querol, X., Ostro, B., Walton, H., 2015. Quantifying the health impacts of ambient air pollutants: recommendations of a WHO/Europe project. Int. J. Publ. Health 60 (5), 619–627 Birkhauser Verlag AG.
Hystad, P., Setton, E., Cervantes, A., Poplawski, K., Deschenes, S., Brauer, M., van Donkelaar, A., Lamsa, L., Martin, R., Jerrett, M., Demers, P., 2011. Creating national air pollution models for population exposure assessment in Canada. Environ. Health Perspect. 119 (8), 1123–1129.
Ibarra-Berastegi, G., Elias, A., Barona, A., Saenz, J., Ezcurra, A., Diaz de Argandoña, J., 2008. From diagnosis to prognosis for forecasting air pollution using neural networks: air pollution monitoring in Bilbao. Environ. Model. Software 23 (5), 622–637. http://dx.doi.org/10.1016/j.envsoft.2007.09.003.
ISO, B., 2005. IEC 17025: 2005 General requirements for the competence of testing and calibration laboratories. In: ICS.
Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. J. Expo. Sci. Environ. Epidemiol. 15 (2), 185–204. http://dx.doi.org/10.1038/sj.jea.7500388. Nature Publishing Group.
Kandlikar, M., 2007. Air pollution at a hotspot location in Delhi: detecting trends, seasonal cycles and oscillations. Atmos. Environ. 41 (28), 5934–5947. http://dx.doi.org/10.1016/j.atmosenv.2007.03.044.
Kandya, A., Mohan, M., Khas, H., Delhi, N., 2009. Forecasting the urban air quality using various statistical. In: 7th International Conference on Urban Climate.
Kang, D., Eder, B.K., Stein, A.F., Grell, G.A., Peckham, S.E., McHenry, J., 2005. The new England air quality forecasting pilot program: development of an evaluation protocol and performance benchmark. J. Air & Waste Manag. Assoc. (1995) 55 (12), 1782–1796. http://dx.doi.org/10.1080/10473289.2005.10464775.
Koenig, J.Q., 1999. Air pollution and asthma. J. Allergy Clin. Immunol. 104 (4), 717–722. http://dx.doi.org/10.1007/978-94-009-4003-1.
Kyrkilis, G., Chaloulakou, A., Kassomenos, P.A., 2007. Development of an aggregate Air Quality Index for an urban Mediterranean agglomeration: relation to potential health effects. Environ. Int. 33 (5), 670–676. http://dx.doi.org/10.1016/J.ENVINT.2007.01.010. Pergamon.
Lebret, E., 1990. Errors in exposure measures. Toxicol. Ind. Health 6 (5), 147–156.
Liu, M., Lin, J., Wang, Y., Sun, Y., Zheng, B., Shao, J., Chen, L., Zheng, Y., Chen, J., Fu, M., Yan, Y., Zhang, Q., Wu, Z., 2018. Spatiotemporal variability of NO2 and PM2.5 over Eastern China: observational and model analyses with a novel statistical method. Atmos. Chem. Phys. Discuss. 1–34. http://dx.doi.org/10.5194/acp-2017-1180.
Manders, A.M.M., Schaap, M., Hoogerbrugge, R., 2009. Testing the capability of the chemistry transport model LOTOS-EUROS to forecast PM10 levels in The Netherlands. Atmos. Environ. 43 (26), 4050–4059. http://dx.doi.org/10.1016/j.atmosenv.2009.05.006.
Mayer, H., 1999. Air pollution in cities. Atmos. Environ. 33 (24–25), 4029–4037. http://dx.doi.org/10.1016/S1352-2310(99)00144-2. Pergamon.
O'Leary, B.F., Lemke, L.D., 2014. Modeling spatiotemporal variability of intra-urban air pollutants in Detroit: a pragmatic approach. Atmos. Environ. 94, 417–427. http://dx.doi.org/10.1016/J.ATMOSENV.2014.05.010. Pergamon.
Ott, W.R., 1978. Environmental Indices: Theory and Practice. Ann Arbor Science Publishers, Inc., Ann Arbor, MI.
Pakes, A.G., 1969. Some conditions for ergodicity and recurrence of Markov chains. Oper. Res. 17 (6), 1058–1061. http://dx.doi.org/10.1287/opre.17.6.1058. INFORMS.
Queensland Government Department of Environment and Heritage Protection, 2015. Air Quality. Available at: https://www.qld.gov.au/environment/pollution/monitoring/air.
Ren, C., Tong, S., 2006. Temperature modifies the health effects of particulate matter in Brisbane, Australia. Int. J. Biometeorol. 51 (2), 87–96. http://dx.doi.org/10.1007/s00484-006-0054-7. Springer-Verlag.
Rodrigues, E.R., Achcar, J.A., 2013. Applications of Discrete-time Markov Chains and Poisson Processes to Air Pollution Modeling and Studies. Springer Science & Business Mediahttp://dx.doi.org/10.1007/978-1-4614-4645-3.
Schwartz, J., 1991. Particulate air pollution and daily mortality in Detroit. Environ. Res. 56 (292120158), 204–213.
Schwartz, J., 1994. Nonparametric smoothing in the analysis of air pollution and respiratory illness. Can. J. Stat. 22 (4), 471–487. http://dx.doi.org/10.2307/3315405.
Schwartz, J., Laden, F., Zanobetti, A., 2002. The concentration-response relation between PM2.5 and daily deaths. Environ. Health Perspect. 110 (10), 1025–1029. http://dx.doi.org/10.1289/ehp.021101025.
Schwartz, J., Spix, C., Touloumi, G., Bacharova, L., Barumamdzadeh, T., le Tertre, A., Piekarksi, T., Ponce de Leon, A., Ponka, A., Rossi, G., Saez, M., Schouten, J.P., 1996. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. J. Epidemiol. Community Health 50 (Suppl. 1), S3–S11. http://dx.doi.org/10.1136/jech.50.Suppl_1.S3. BMJ Publishing Group Ltd.
Sousa, S., Martins, F., Alvimferraz, M., Pereira, M., 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Environ. Model. Software 22 (1), 97–103. http://dx.doi.org/10.1016/j.envsoft.2005.12.002. Elsevier.
Tranmer, M., Elliot, M., 2008. Multiple Linear Regression. The Cathie Marsh Centre for Census and Survey Research (CCSR).
United States Environmental Protection Agency, 2017. Nitrogen Dioxide (NO2) Pollution.
Vacher, M., Fdez Galván, I., Ding, B.-W., Schramm, S., Berraud-Pache, R., Naumov, P., Ferré, N., Liu, Y.-J., Navizet, I., Roca-Sanjuán, D., Baader, W.J., Lindh, R., 2018. Chemi- and bioluminescence of cyclic peroxides. Chem. Rev. http://dx.doi.org/10.1021/acs.chemrev.7b00649. American Chemical Society, p. acs.chemrev.7b00649.
Venkatadri, M., Rao, P.S., 2014. A survey on air quality forecasting techniques. Int. J. Comput. Sci. Inf. Technol. 5 (1), 103–107.
Vereecken, L., Huyberechts, G., Peeters, J., 1997. Stochastic simulation of chemically activated unimolecular reactions. J. Chem. Phys. 106 (16), 6564. http://dx.doi.org/10.1063/1.473656. AIP Publishing.
Wilby, R.L., Tomlinson, O.J., 2000. The “Sunday Effect” and weekly cycles of winter weather in the UK. Weather 55 (7), 214–222. http://dx.doi.org/10.1002/j.1477-8696.2000.tb04064.x. Blackwell Publishing Ltd.
Ye, N., 2000. A Markov chain model of temporal behavior for anomaly detection. In: Proceedings of the 2000 IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop, vol. 4. pp. 171–174.
Yuval, Broday, D.M., 2006. High-resolution spatial patterns of long-term mean concentrations of air pollutants in Haifa Bay area. Atmos. Environ. 40 (20), 3653–3664. http://dx.doi.org/10.1016/j.atmosenv.2006.03.037.
Zannetti, P., 1990. Air Pollution Modeling - Theories, Computational Methods and Available Software, Theories Computational Method and Available Software. http://dx.doi.org/10.1017/CBO9781107415324.004.
Zeger, S.L., Thomas, D., Dominici, F., Samet, J.M., Schwartz, J., Dockery, D., Cohen, A., 2000. Exposure measurement error in time-series studies of air pollution: concepts and consequences. Environ. Health Perspect. 108 (5), 419–426. http://dx.doi.org/10.1289/ehp.00108419. National Institute of Environmental Health Science.
Zhang, J., 2017. Low-level air pollution associated with death. JAMA 318 (24), 2431. http://dx.doi.org/10.1001/jama.2017.18948. American Medical Association.
Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A., 2012. Real-time air quality forecasting, part I: history, techniques, and current status. Atmos. Environ. 60, 632–655. http://dx.doi.org/10.1016/j.atmosenv.2012.06.031. Elsevier Ltd.
Zipkin, E.F., Jennelle, C.S., Cooch, E.G., 2010. A primer on the application of Markov chains to the study of wildlife disease dynamics. Meth. Ecol. Evol. 1 (2), 192–198. http://dx.doi.org/10.1111/j.2041-210X.2010.00018.x.