



Jun 28th, 9:00 AM - 10:20 AM

Long-Term Forecasting of Nitrogen Dioxide Ambient Levels in Metropolitan Areas Using the Discrete-Time Markov Model

Asaf Nebenzal
asaf.n@technion.ac.il

Barak Fishbain
Technion Israel Institute of technology, fishbain@technion.ac.il

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

Nebenzal, Asaf and Fishbain, Barak, "Long-Term Forecasting of Nitrogen Dioxide Ambient Levels in Metropolitan Areas Using the Discrete-Time Markov Model" (2018). *International Congress on Environmental Modelling and Software*. 49.
<https://scholarsarchive.byu.edu/iemssconference/2018/Stream-C/49>

This Oral Presentation (in session) is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Long-Term Forecasting of Nitrogen Dioxide Ambient Levels in Metropolitan Areas Using the Discrete-Time Markov Model

Asaf Nebenzal^a and Barak Fishbain^b

^a Department of Mathematics, Technion – Israel Institute of Technology, Haifa 320003, Israel,
(asaf.n@technion.ac.il)

^b Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 320003, Israel, (fishbain@technion.ac.il)

Abstract: Air pollution management and control are key factors in maintaining sustainable societies. Air quality forecasting is a key factor in these tasks. While short-term forecasting, few days into the future, is a well-established research domain, no method exists for long-term forecasting (e.g., what will be the pollution levels distribution in the next year), which is essential for environmental planning and management as well as epidemiological and exposure studies. This research defines *long-term* air pollution forecasting and presents a Discrete-Time-Markov-based model for forecasting ambient nitrogen oxides patterns. Specifically, the model provides an estimate for the chance for tomorrow's pollution level, given today's level. Thus, it provides an understanding of pollution's temporal behaviour. The model was applied on two distinctive regions in Israel and Australia, giving a solid base for evaluation. The model did manage to forecast accurately the future transition probabilities.

Keywords: Air pollution; Discrete-Time Markov model; Long-term forecasting; Modelling; Scenario analysis;

1. INTRODUCTION

Air pollution is a significant risk factor for multiple health issues and impacts negatively on the environment (Venkatadri and S.Rao, 2014; Heroux *et al.*, 2015). 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 not only the pollution **levels** impair the human body and its physiological processes but the daily differences in pollution are also a key factor. Thus, arises the need for forecasting not only pollution levels but the daily transitions.

A wide variety of techniques has been developed for forecasting air quality. Generally speaking, these methods can be categorized to empirical, statistical and deterministic approaches (Zhang *et al.*, 2012), all of them only provide short-term prediction, typically 24-72 hours. 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.

Nitrogen-dioxide, NO₂, is a gas with known harmful effects on human health. Short-term exposure may cause respiratory problems like wheezing, coughing, and bronchitis and worsen respiratory diseases like asthma (Koenig, 1999; Clark *et al.*, 2010). Long-term exposure might also increase the risk of onset asthma as well as incidence and exacerbation of asthma attacks (United States EPA 2017). NO₂ is mainly originated from combustion sources in vehicles and power plants. It also precursors to several harmful secondary air pollutants, such as ground-level ozone production, and nitrates which contribute

to increased respirable particle levels in the atmosphere and play a role in the formation of acid rain. Hence, NO₂ levels can be used as an indicator for the general pollution condition of a region (Clark, Millet and Marshall, 2014).

This research introduces long-term forecasting of NO₂ levels through Air Pollution Discrete-Time Markov Chain (AP-DTMC) model. The AP-DTMC provides new means for assessing and more importantly estimating future air-pollution temporal characteristics and behavior. Specifically, behavior here refers to the probability to obtain any degree of pollution in the proceeding time period, given the current level, or a short historical data.

2. METHODS

2.1. Study area

This study is based on NO₂ pollution levels measured in two different regions: the Tel-Aviv metropolitan area in Israel and Brisbane, Queensland, Australia. Tel-Aviv metropolitan area is the most populated region in Israel. It is about 150 km² (58 squared miles) with a population of 2,000,000 people and heavy-traffic activity. It includes Tel-Aviv, the biggest city in Israel, and its suburbs cities - Ramat-Gan, Petah-Tikva, Rishon LeTsiyon, and Holon. Its borders, as prescribed in this research are the Mediterranean Sea to the west, Petah-Tikva to the east, Ramat-Hasharon to the north and Rishon Letsion to the south. Brisbane is the capital and most populous city in the Australian state of Queensland, and the third most populous city in Australia with about 2,200,000 people. This research focused on center-east Brisbane, which is about 260 km² (100 squared miles), 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 Brisbane air and seaports.

2.2. Data Collection

NO₂ data, collected between 2012-2015, were used. Such long-time period assures that both seasonality and long-term trends are 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). Data is saved as half-hourly mean concentrations, in ppb units. In Australia, data obtained by the Department of Environment and Heritage Protection, Queensland Government (Environment and Heritage Protection, Queensland 2017). Data are saved as one-hour average in ppm units.

The collected data were divided into historical (i.e., training) and validation sets. Training data consisted of measurements acquired throughout the years 2012-2014 and then the validation was done on 2015 data. Hence, the prediction results were compared against the actual measurements of 2015.

In this study, the pollution levels at 6:00 PM are used as a *daily characteristic measurement* as this was the practice in several studies that aimed at analyzing pollution temporal patterns, e.g., (Broday, Alpert and others, 2012). The time 6:00 PM was chosen as in most locations the level of pollution is at its peak or close to that at this time. A possible explanation is that one of the significant sources of NO₂ is transportation, and at 6:00 PM, the highest quantity accumulates before it reacts with other substances or disperses. The representative pollution level, for each region was thus the average NO₂ level calculated over all measurements acquired by the AQM stations in that region at 6:00 PM.

2.3. Air pollution Discrete-Time Markov model

Discrete-Time Markov Chain (DTMC) is a well-known probabilistic model used to describe and analyze stochastic processes (Ye, 2000). DTMC allows predicting and investigating the behavior of a given process and has been proved to be effective and useful in a wide range of topics such as chemistry (Vereecken, Huyberechts and Peeters, 1997), medicine (Zipkin, Jennelle and Cooch, 2010) and computer science (Almasizadeh and Azgomi, 2008). In the scope of air pollution, DTMC was 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 an emission model arising from transportation (Crisostomi and Kirkland, 2011).

Next we define the notation that constitutes the formulation of the AP-DTMC. Let $a_t^r \in A^r$ be region r 's daily NO_2 characteristic measurement at day t , where $t \in T = \{1, 2, 3, \dots\}$ is the day index; and let A^r be the entire region's time series consists 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, Prakash and Jain, 2009), or with respect to other characteristics 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 . Similarly, 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 only on the present label s_t^r :

$$P\{s_{t+1}^r | s_t^r, s_{t-1}^r, \dots, s_1^r, s_0^r\} = P\{s_{t+1}^r | s_t^r\} \quad (1)$$

This assumption stems from the typical gradual behavior of air pollution and of meteorology. Both phenomena tend to show stability over few days cycles, e.g., diurnal and weekly patterns (Wilby and Tomlinson, 2000; Broday, Alpert and others, 2012). Due to the slow change, the assumption is that the previous day is sufficient to indicate on the next day's pollution level and thus air pollution fulfills the Markov property.

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 is s_i :

$$P_{ij} = P\{s_{t+1}^r = s_j | s_t^r = s_i\}, \quad s_i, s_j \in S \quad (2)$$

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

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1K} \\ P_{21} & P_{22} & \dots & P_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ P_{K1} & P_{K2} & \dots & P_{KK} \end{bmatrix} \quad (3)$$

The probabilities populating the matrix P are derived from the historical data, simply counting how many times, out of the total days, s_j followed s_i . Utilizing the entire historical data, which incorporate both major pollution gradients, as well as moderate changes, i.e., seasonality, allowing 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 . Having the matrix P in hand, the AP-DTMC provides the probabilities for tomorrow's pollution's label, given today's label. This facilitates both short and long-term forecasting. In the short-term forecasting, AP-DTMC is considered as a good benchmark for tomorrow forecast (Garner and Thompson, 2012). In the long-term forecasting, this model shed a light on the general pollution distribution and histogram of the daily differences.

3. RESULTS AND DISCUSSION

The AP-DTMC model was applied to the NO_2 datasets described in Sect. 2.2 The data was divided into a training set, consisted of the data acquired 2012-2014, and to a validation, test set, which consists of 2015 data. For the transition analysis, we estimate 2015 transition matrix, 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 . Figure 1(a) and (b) represents Tel-Aviv metropolitan and Brisbane true values (red) and estimated values (yellow) of 2015 transition matrix.

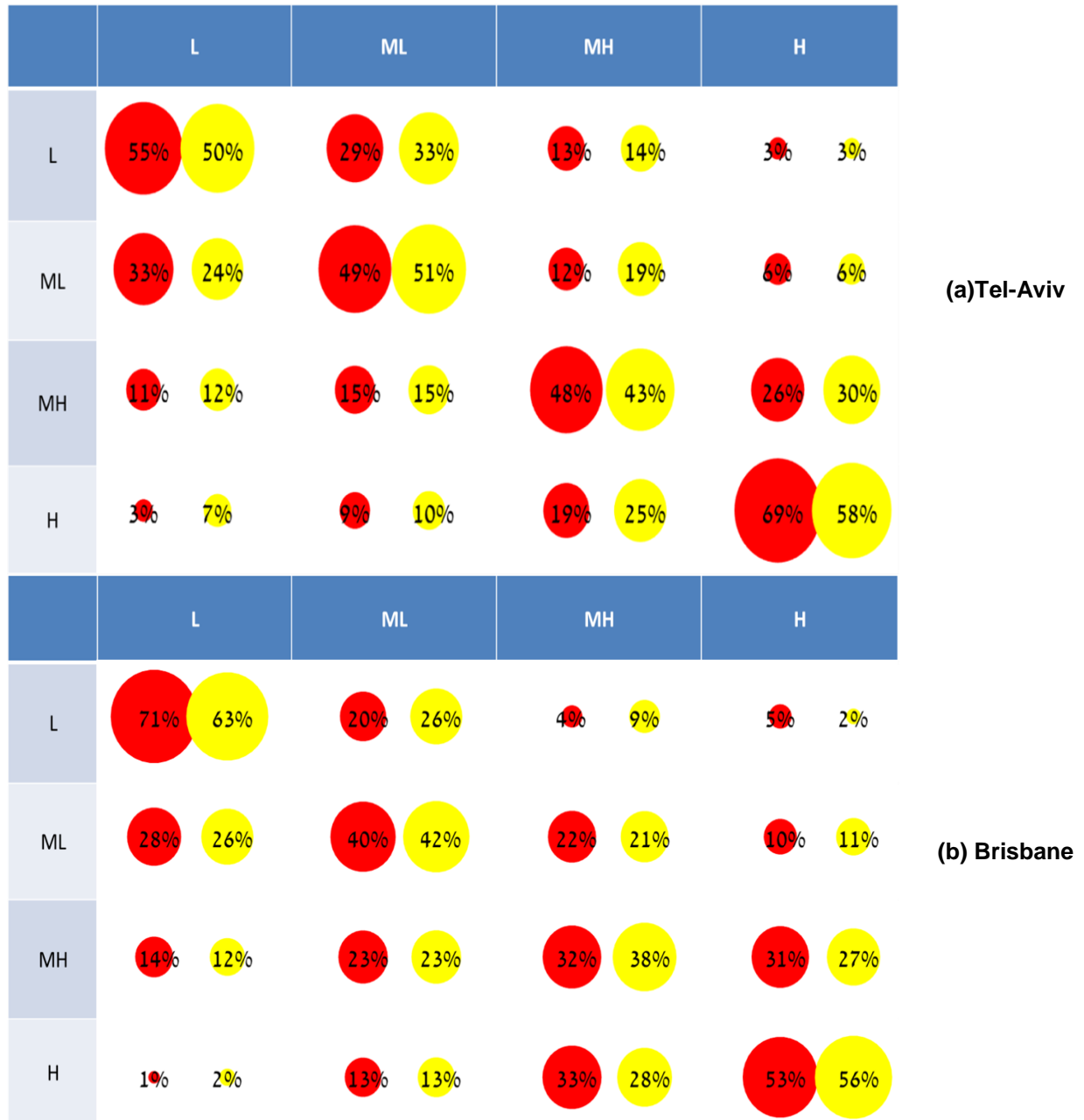


Figure 1. Transition matrices - observed (left hand) in red versus expected (right hand) in yellow. The balloon size represents the probability of each pollution level.

The mean absolute estimation error (MAE) rate for Tel-Aviv and Brisbane is 3.66% and 2.94% respectively. As can be seen in Figure 1, the model does manage, with high accuracy, to predict the behavior of pollution distribution. The results also show the tendency of air pollution system to maintain stability, as 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 $s_{t+1}^r = 'L'$, at day t , chances are that the level would be 'L' in the following day as well. While the chance of shifting from a specific pollution level to a neighboring level (e.g., from 'ML' in day t to 'L' or 'MH' at day $t+1$) is lower than maintaining the same level, it does present higher chance than shifting two levels, like from 'ML' to 'H'. Shifting three levels, which can happen only in the transition from 'L' to 'H' and vice versa, is very rare.

Weather conditions have great influence on air pollution. Meteorological parameters as wind speed and direction, temperature, humidity, rain, clouds and solar radiation clearly affect air quality (Zannetti 2013). Synoptic-scale weather typically exhibits similar characteristics for several days, hence weather

changes occur gradually rather than immediately (Dye et al. 2003; Broday et al. 2012). The characteristics, described above of the transition matrix, are partially attributed to this weather behavior.

4. CONCLUSIONS AND FUTURE WORK

This paper introduces the concept of long-term air pollution forecasting in terms of transition probabilities analyses. These two properties of the pollution signal are derived through modeling the pollution time series as a Markov chain model. The model was applied to NO₂ measurements acquired through 2012-2014, in Israel and Australia. The independence between those two regions allows for an adequate evaluation of the model. The suggested scheme was able to predict accurately the transition matrices, based on the historical data, of 2015. As any other prediction method, the AP-DTMC presented here is vulnerable to changes in current trends in the natural or anthropogenic conditions. Future research may address the application of the method on more regions, to see if the assumptions hold and to exploit other properties of the DTMC model for analyzing air pollution, such as the model's stationary distribution of the different labels. In addition, a comparison between the AP-DTMC to a chemistry transport models such as CHIMERE or CMAQ may enable a better assessment of the proposed model.

ACKNOWLEDGMENTS

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

REFERENCES

- Almasizadeh, J. and 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. doi: 10.1109/SecTech.2008.35.
- Bishoi, B., Prakash, A. and 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 Quality Research*, 9(1), pp. 1–17.
- Broday, D. M., Alpert, P. and others (2012) 'Exploring the applicability of future air quality predictions based on synoptic system forecasts', *Environmental pollution*. Elsevier, 166, pp. 65–74.
- Cheng, W.-L., Chen, Y.-S., Zhang, J., Lyons, T. J., Pai, J.-L. and Chang, S.-H. (2007) 'Comparison of the Revised Air Quality Index with the PSI and AQI indices', *Science of The Total Environment*. Elsevier, 382(2–3), pp. 191–198. doi: 10.1016/J.SCITOTENV.2007.04.036.
- Clark, L. P., Millet, D. B. and Marshall, J. D. (2014) 'National patterns in environmental injustice and inequality: outdoor NO₂ air pollution in the United States.', *PloS one*, 9(4), p. e94431. doi: 10.1371/journal.pone.0094431.
- Clark, N. A., Demers, P. a, Karr, C. J., Koehoorn, M., Lencar, C., Tamburic, L. and Brauer, M. (2010) 'Effect of early life exposure to air pollution on development of childhood asthma.', *Environmental health perspectives*, 118(2), pp. 284–90. doi: 10.1289/ehp.0900916.
- Crisostomi, E. and 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. and Dominici, F. (2017) 'Association of Short-term Exposure to Air Pollution With Mortality in Older Adults', *JAMA. American Medical Association*, 318(24), p. 2446. doi: 10.1001/jama.2017.17923.
- Dye, T. S., AIRNow Program (U.S.) and Sonoma Technology Inc (2003) 'Guidelines for Developing an Air Quality (ozone and PM_{2.5}) Forecasting Program'. US Environmental Protection Agency, Office of Air Quality Planning and Standards, Information Transfer and Program Integration Division, AIRNow Program, p. 1. doi: EPA-456/R-03-002.
- Garner, G. G. and Thompson, A. M. (2012) 'The Value of Air Quality Forecasting in the Mid-Atlantic Region', *Weather, Climate, and Society*, 4(1), pp. 69–79. doi: 10.1175/WCAS-D-10-05010.1.
- 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. and Walton,

- H. (2015) 'Quantifying the health impacts of ambient air pollutants: recommendations of a WHO/Europe project', *International Journal of Public Health*. Birkhauser Verlag AG, 60(5), pp. 619–627.
- Koenig, J. Q. (1999) 'Air Pollution and Asthma', *Journal of allergy and clinical immunology* 4(104), pp. 717–722. doi: 10.1007/978-94-009-4003-1.
- Ott, W. R. (1978) 'Environmental indices: theory and practice'. Ann Arbor Science Publishers, Inc., Ann Arbor, MI.
- Queensland government Department of Environment and Heritage Protection (2017) Air quality. Available at: <https://www.qld.gov.au/environment/pollution/monitoring/air>.
- Rodrigues, E. R. and Achcar, J. A. (2013) *Applications of Discrete-time Markov Chains and Poisson Processes to Air Pollution Modeling and Studies*. Springer Science & Business Media. doi: 10.1007/978-1-4614-4645-3.
- United States Environmental Protection Agency (2017) Nitrogen Dioxide (NO₂) Pollution.
- Venkatadri, M. and S.Rao, P. (2014) 'A survey on Air Quality forecasting Techniques', *International Journal of Computer Science and Information Technologies*, 5(1), pp. 103–107.
- Vereecken, L., Huyberechts, G. and Peeters, J. (1997) 'Stochastic simulation of chemically activated unimolecular reactions', *The Journal of Chemical Physics*. AIP Publishing, 106(16), p. 6564. doi: 10.1063/1.473656.
- Wilby, R. L. and Tomlinson, O. J. (2000) 'The "Sunday Effect" and weekly cycles of winter weather in the UK', *Weather*. Blackwell Publishing Ltd, 55(7), pp. 214–222. doi: 10.1002/j.1477-8696.2000.tb04064.x.
- Ye, N. (2000) 'A markov chain model of temporal behavior for anomaly detection', *Proceedings of the 2000 IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop*, (4), pp. 171–174.
- Yuval and Broday, D. M. (2006) 'High-resolution spatial patterns of long-term mean concentrations of air pollutants in Haifa Bay area', *Atmospheric Environment*, 40(20), pp. 3653–3664. doi: 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. doi: 10.1017/CBO9781107415324.004.
- Zhang, J. (2017) 'Low-Level Air Pollution Associated With Death', *JAMA*. American Medical Association, 318(24), p. 2431. doi: 10.1001/jama.2017.18948.
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C. and Baklanov, A. (2012) 'Real-time air quality forecasting, part I: History, techniques, and current status', *Atmospheric Environment*. Elsevier Ltd, 60, pp. 632–655. doi: 10.1016/j.atmosenv.2012.06.031.
- Zipkin, E. F., Jennelle, C. S. and Cooch, E. G. (2010) 'A primer on the application of Markov chains to the study of wildlife disease dynamics', *Methods in Ecology and Evolution*, 1(2), pp. 192–198. doi: 10.1111/j.2041-210X.2010.00018.x.