Progress in IS

Volker Wohlgemuth
Frank Fuchs-Kittowski
Jochen Wittmann *Editors*

Advances and New Trends in Environmental Informatics

Stability, Continuity, Innovation

Springer

Editors
Volker Wohlgemuth
Environmental Informatics
Hochschule für Technik und Wirtschaft Berlin (HTW Berlin), University of Applied Sciences
Berlin
Germany

Jochen Wittmann
Environmental Informatics
Hochschule für Technik und Wirtschaft Berlin (HTW Berlin), University of Applied Sciences
Berlin
Germany

Frank Fuchs-Kittowski
Environmental Informatics
Hochschule für Technik und Wirtschaft Berlin (HTW Berlin), University of Applied Sciences
Berlin
Germany

ISSN 2196-8705 ISSN 2196-8713 (electronic)
Progress in IS
ISBN 978-3-319-44710-0 ISBN 978-3-319-44711-7 (eBook)
DOI 10.1007/978-3-319-44711-7

Library of Congress Control Number: 2016947762

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Optimal Noise Filtering of Sensory Array Gaseous Air Pollution Measurements

Barak Fishbain, Shai Moshenberg and Uri Lerner

Abstract One of the fundamental components in assessing air quality is continuous monitoring. However, all measuring devices are bound to sensing noise. Commonly the noise is assumed to have zero mean and, thus, is removed by averaging data over temporal windows. Generally speaking, the larger the window, the better the noise removal. This operation, however, which corresponds to low pass filtering, might result in loss of real abrupt changes in the signal. Therefore, the need arises to set the window size so it optimally removes noise with minimum corruption of real data. This article presents a mathematical model for finding the optimal averaging window size. The suggested method is based on the assumption that while real measured physical phenomenon affects the measurements of all collocated sensors, sensing noise manifests itself independently in each of the sensors. Hence, the smallest window size which presents the highest correlation between the collocated sensors, is deemed as optimal. The results presented here show the great potential of the method in air quality measurements.

Keywords Air pollution measurements · Noise filtering · Micro sensing units

1 Introduction

Air quality has a tremendous effect on public health and the environment (Künzli et al. 2000). Many studies have associated various adverse effects to general air pollution and its specific components such as nitrogen dioxide (NO_2), ozone (O_3) carbon monoxide (CO) and particular matter (PM), to name a few (Kampa and Castanas

B. Fishbain (✉) · S. Moshenberg · U. Lerner
Faculty of Civil & Environmental Engineering,
Technion - Israel Institute of Technology, 32000 Haifa, Israel
e-mail: fishbain@technion.ac.il

S. Moshenberg
e-mail: shaisho@tx.technion.ac.il

U. Lerner
e-mail: uriler@technion.ac.il

V. Wohlgemuth et al. (eds.), *Advances and New Trends in Environmental Informatics*, Progress in IS, DOI 10.1007/978-3-319-44711-7_22

2008). These pollutants, for example, affect the respiratory system, the cardiovascular system and other systems in the human body (Laumbach and Kipen 2012). Some of the pollution is due to natural phenomena and some due to anthropogenic activity (Robinson and Robbins 1970; Cullis and Hirschler 1980). Regardless of its sources, air pollution undergoes a set of chemical processes in the atmosphere, depending on initial concentration and ambient conditions. The large number of sources and the intricateness of the chemical processes, lead to the creation of complex scenarios, displaying highly variable spatial and temporal pollution patterns rendering the analysis of air-pollution and its effects as a challenging task (Nazaroff and Alvarez-Cohen 2001; Levy et al. 2014).

One of the primary tools for assessing air-pollution patterns is continuous monitoring of pollutants' ambient levels. To accomplish that, numerous chemical-physical methods have been developed and standardized Air Quality Monitoring (AQM) station networks have been spread around the world. However, as any other sensor, these AQM stations are bound to measurement errors due to sensors' and circuitry noise. This noise limits AQM's capability to accurately capture ambient pollution levels and thus, hinders the study of air-pollution (Duyzer et al. 2015). With the growing usage of Micro Sensing Units (MSUs) for measuring ambient pollutants' levels (Künzli et al. 2000; Kampa and Castanas 2008; Mead et al. 2013; Williams et al. 2013; Moltchanov et al. 2015; Lerner et al. 2015), this problem increases as MSUs are more error prone than the standard measuring equipment (Tchepel and Borrego 2010; Mead et al. 2013; Williams et al. 2013; Moltchanov et al. 2015; Lerner et al. 2015). Thus, in order to better utilize the sensing equipment, noise must be effectively filtered out.

Filtering the noise out requires full characterization of either the noise or the signal. The statistical properties of the sensing noise may be known from the certification of the monitoring system. However, in many applications these data are unavailable. Further, it was shown that MSUs' accuracy, i.e. sensory noise level, varies over time, which makes any characterization futile, as it is valid for only a limited time period (Künzli et al. 2000; Gupta et al. 2011; US Environmental Protection Agency 2012; Mead et al. 2013; Moltchanov et al. 2015).

Sensing noise is often characterized as Additive White Gaussian Noise (AWGN) (Schwartz and Marcus 1990; Rao and Zurbenko 1994; Varotsos et al. 2005). Thus, for x_i, the true pollutant's ambient level and ε_i, the noise at time step i, the measurement y_i is given by: $y_i = x_i + \varepsilon_i$, where ε_i is a normally distributed random variable with zero mean and unknown variance (Wu and Huang 2009).

Realistically, changes in the composition of the atmosphere happen over relatively long period of time when compared to the sampling rate, i.e. order of tens of minutes with respect to the sampling rates of tens of seconds (Rao and Zurbenko 1994; Wang et al. 2003; Peng et al. 2006). Even when considering photochemistry in hot regions, a global change in air-pollution composition takes a much longer time than the sampling rate (Leighton 2012; Weinstein et al. 2016). Combined with the assumption of AWGN, noise may be filtered out by averaging the signal over a temporal sliding window, i.e., replacing each measurement, y_i, with the computed average of samples within a temporal window centered at i (Schwartz and Marcus 1990). This procedure is also known as Kolmogorov-Zurbenko (KZ) (Zurbenko 1986) or Sinc filtering (Yaroslavsky 2014), and for a window size of $2K + 1$ is given by:

$$y_i = \frac{1}{2K+1} \sum_{j=-K}^{K} y_{i+j} \tag{1}$$

The KZ operator essentially suppresses abrupt changes in the signal. The larger the temporal window is, the smoother the output signal (Yaroslavsky 2014). This is equivalent to removing higher frequencies of the signal, thus, low-pass filtering (Zurbenko 1986; Schwartz and Marcus 1990; Yaroslavsky 2014).

To this end, low-pass filtering in its simplest form means zeroing all signal's frequency coefficients above a given frequency, called the cut-off frequency. The larger the window size, the lower the cut-off frequency.

Previous analyses suggested to set the cutoff frequency so it maximizes the coefficient of determination, R^2, of a regression model associating mortality (Peng et al. 2006) or temperature (Rao and Zurbenko 1994) with air-pollution measurements. In both cases, the temporal window size found was considerably large (order of days), heavily smoothing the signals. This outcome is expected as signal's temporal local variations, whether originated from genuine signal's fluctuations or from noise, degrade R^2 value. Thus, removing these perturbations improves the regression model, but deteriorates the signal's high frequencies.

Therefore, using such a filter for noise filtering calls for a method to determine the ideal cutoff frequency or the size of the temporal window, so it eliminates as much noise as possible, while preserving real data. Here we present a mathematical model to optimally set the window's size.

2 Materials and Methods

2.1 Optimal Filtering Window Size

Typically, as the level of noise increases, the correlation between the real signal and the measured signal decreases (Fishbain et al. 2008). Assuming that the noise affects each sensor independently, if a pollution signal is measured by two separate collocated devices, the correlation between them is expected to decrease as the noise level grows. This is illustrated in Fig. 1, where Fig. 1a depicts real-life NO_2 time series, A_N, acquired between January 1st and December 31st, 2010 (16,949 samples) by a standard AQM station located at the heart of the Haifa industrial/commercial area (LAT/LON: 32.78919/35.04038)—see (Moltchanov et al. 2015) for more details on the study area. A_N's maximum measurement was 48 [ppb], its average was 4.67 [ppb] and its standard deviation was 5.98. From this signal a synthetic noisy signal, S_k, is generated by adding random AWGN, ε_σ, with zero mean and standard deviation, so the signal to noise ratio (SNR) is 5. The process is then repeated with

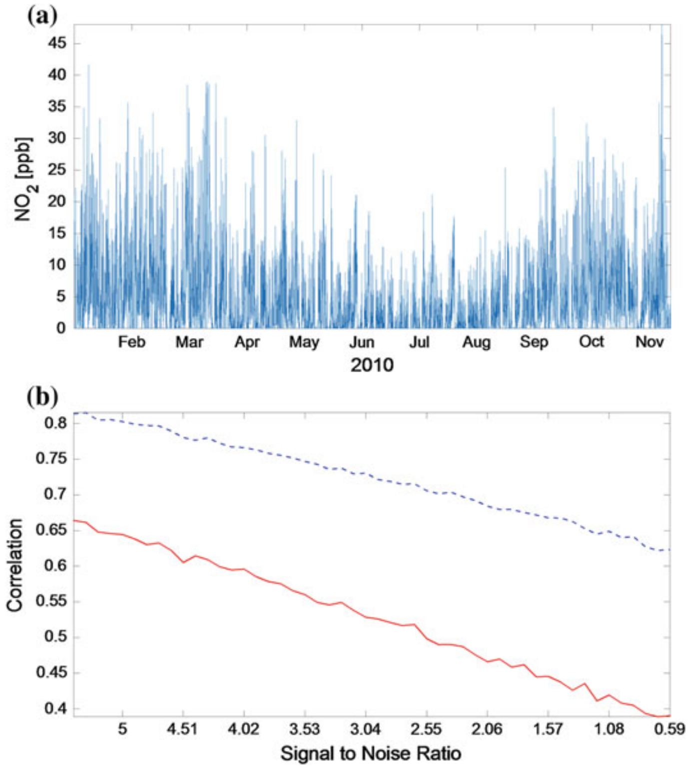

Fig. 1 Correlation coefficient as a function of noise standard deviation. **a** Real-life NO_2 time series acquired between January 1st and December 31st, 2010 (16,949 samples) by a standard AQM station located at the heart of the Haifa industrial/commercial area (LAT/LON: 32.78919/35.04038). **b** Correlation between two sets of synthetic noisy signals as a function of added noise characteristics (*solid-red*) and between the original signal and one of the synthetic signals (*dashed-blue*)

SNR=$\{4.9, 4.8, \ldots, 0.1\}$. Hence, a set of fifty signals with noise increasing standard deviations is created. Two such sets are used here.

Figure 1b shows the correlation between the two sets of synthetic noisy signals as a function of the added AWGN's standard deviation (solid-red) and between the original signal, A_N, and one of the synthetic signals (dashed-blue). It is evident that indeed the correlation drops as the noise level increases.

Signal's energy is a characteristic used in signal processing for quantifying the amount of data within a signal. For a continuous signal, $p(t)$, the energy is given by:

$$E = \int_{-\infty}^{\infty} |p(t)|^2 dt \tag{2}$$

Following the Parseval's theorem (Boas 1966), the energy of a signal is equal to the energy of its frequency transform, $P(\omega)$:

$$E = \int_{-\infty}^{\infty} |P(t)|^2 d\omega \tag{3}$$

Hence, the function $|P(\omega)|^2$ represents the energy distribution in the frequency domain. Applying discrete sampling, (3) becomes:

$$E = \sum_{\omega=1}^{N} |P(\omega)|^2 \tag{4}$$

Given (4) and the notion presented in Fig. 1, the optimization goal is to find the highest cut-off frequency such that as little information, i.e., energy, in the higher frequencies is removed, while the correlation between two collocated sensors reaches its maximum.

This is the essence of the suggested filtering scheme. For removing AWGN, a temporal window is suggested. For finding the optimal window's size, one should balance between the window size which presents the highest correlation between two sensors measuring the same physical phenomenon, and by evaluating the signal's spectrum in search of a cut-off frequency, which removes as little as possible of a signal's energy, i.e., information.

The same physical phenomenon can be identically measured when the sensors are collocated (Mead et al. 2013; Moltchanov et al. 2015; Williams et al. 2013). This mode of operation is applicable mainly when MSUs are in use. Due to MSUs' inherent limitations, collocating is currently the common practice (Fishbain and Moreno-Centeno 2016; Lerner et al. 2015; Mead et al. 2013; Moltchanov et al. 2015; Williams et al. 2013). When the sensors are not collocated, measuring the same phenomenon can be achieved when it is uniform in all measuring points (Moltchanov et al. 2015).

2.2 Frequency Representation

In this study the transformation of the pollutants' time-series to the frequency domain is executed through the 2nd Discrete Cosine Transform (DCT). The DCT is well documented to have high energy-compaction, i.e., most of the signal's energy, in the frequency domain, lays with a small number of low-frequencies coefficients (Zurbenko 1986). Using DCT increases the amount of information in the lower frequencies, limiting true signal's information in the higher frequencies. For a pollutant time series, A_N, that is composited of N data points—a_k, the frequency coefficient, α_r is given by:

$$\alpha_r = \frac{2}{\sqrt{2N}} \sum_{k=0}^{N-1} \left(\frac{a_k \cos\left[\pi\left(k + \frac{1}{2}\right)\right]}{N} r \right) \tag{5}$$

2.3 Data

For demonstrating the suggested filtering scheme, two Air-Quality MSU pods (AQMesh 2015) were placed near an AQM station in Haifa, Israel (Lat:32.78741, Lon: 35.02119, height above ground level: 12 [m], height above sea level: 208 [m]). Each AQMesh unit was equipped with five environmental sensors: NO, NO_2, O_3, atmospheric pressure (AP), and relative humidity (RH). Additionally, the AQMesh measured the unit's internal temperature (Temp). Each pod has its own battery and communication device, wirelessly-transmitting the measurements to a central server every 15 min.

In order to compare the AQM and the MSU measurements, the time resolution of both should be the same. If that is not the case, the time series with the fine temporal resolution is aggregated so it fits the coarser resolution. The MSU measurements were acquired at a 15 min resolution, while the AQM time-series had a 30 min resolution. Hence, MSU measurements were averaged (without overlapping) to produce a time-series that corresponds to the AQM temporal resolution.

3 Results

For simulating true sensors' data post-processing, the measured signals of the two MSUs were low-pass filtered by averaging, with no overlapping windows and decreasing filter size, i.e., lowering the cut-off frequency at each iteration. For each window's size the correlation between the two averaged sequences was calculated. As seen in Fig. 2 for O_3, there is a peak at around 500 min. Also evident is that the variance of the correlation increases with the window size. This is attributed to the smaller number of window's positions, which decreases with the window's size.

The DCT transformation of the ozone time series is plotted in Fig. 3. Setting the cut-off frequency so 90 % of signal's energy is preserved, the cut-off frequency was found to be 53 [1/min]. This is equivalent to averaging the signal over 672 min. Evaluating this result with respect to Fig. 2, this value is sufficiently close to the highest correlation (found around 500 min) and thus noise can be filtered out without compromising on the correlation between the two signals. The 11 h average that was found by the suggested method agrees with the National Ambient Air Quality Standards (NAAQS) of the United States Environmental Protection Agency (US-EPA), which suggests an 8 h average (US Environmental Protection Agency 2012) for monitoring ozone.

Figure 4 illustrates the filtered signal (in red) versus the original noisy signal, in blue. It is noticeable that the filtered signal manages to describe the measurements truthfully, while giving a smoother behavior, without peaking at extreme high or low values.

The same process was performed on an NO_2 signal and is described in Figs. 5, 6 and 7. The cut-off frequency was obtained at 2,657. The 2,657 [1/ min] cut-off

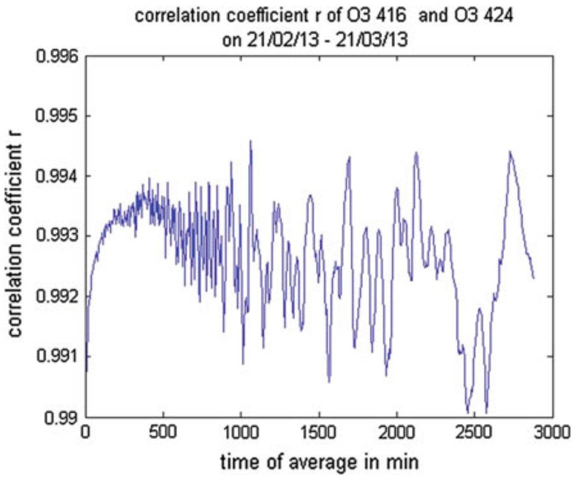

Fig. 2 Correlation between two MSUs as a function of the averaging temporal window size for O_3 measurements

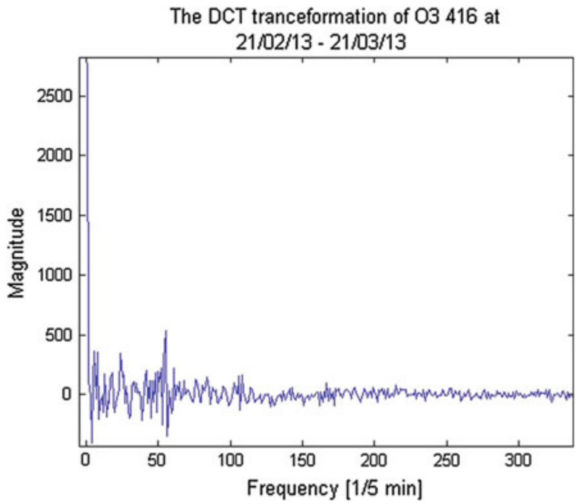

Fig. 3 DCT transformation of the O_3 time series

is equivalent to averaging over a temporal window of 15 min. The correlation, is highest when averaging the signal over a window of 50 min (Fig. 5). The US-EPA NAAQS for NO_2 is one hour (US Environmental Protection Agency 2012), which is close to the window suggested by our method.

In Fig. 7 the original noisy NO_2 signal can be seen in blue, and the filtered signal is in red, and again, it is evident that the filtered signal changes more gradually over time, and presents lower noise level.

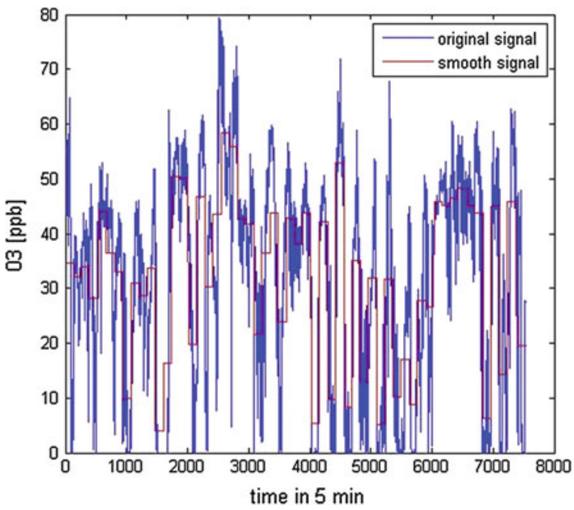

Fig. 4 DCT transformation of the O_3 time series

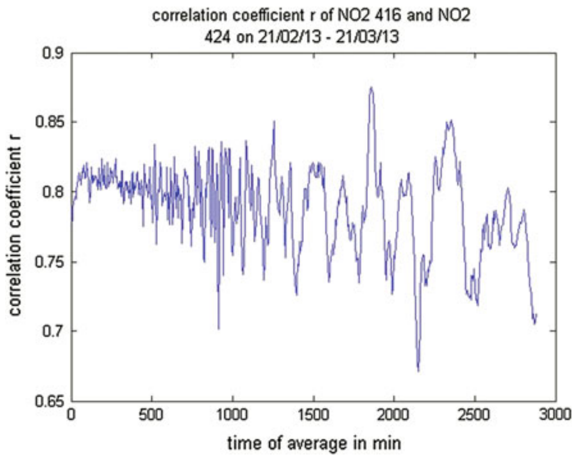

Fig. 5 Correlation between two MSUs as a function of the averaging temporal window size for NO_2 measurements

4 Conclusion

A methodology for finding the optimal averaging window size for noise removal in air-quality time series is suggested. The window's size is set by balancing between two criteria: maximum correlation between two signals obtained by collocated sensors, and applying a low-pass filter with the highest cut-off frequency. Using this method, the noise affecting the quality of the air pollution signal can be filtered out based on the actual measurement taken (and not by a common rule of thumb), thus giving a better assessment of the monitored signal, improving understanding of the environment.

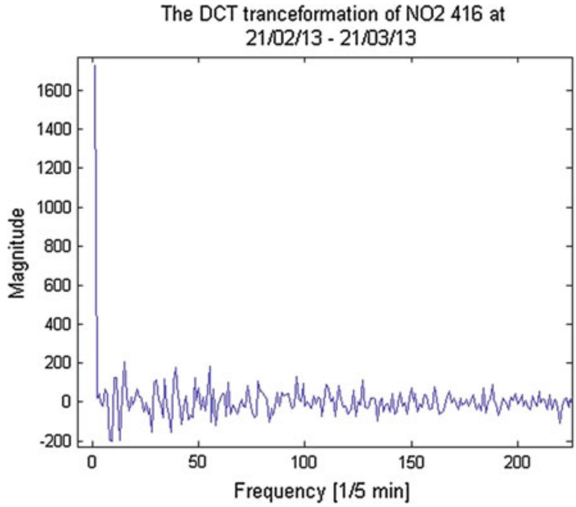

Fig. 6 DCT transformation of the NO_2 time series

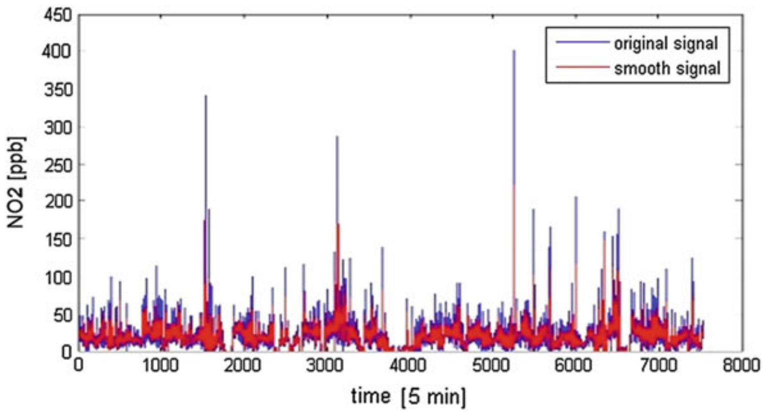

Fig. 7 Original (*blue*) and filtered (*red*) nitrogen-dioxide signal

More research regarding the optimal percent of energy preserved is needed. We assumed that disregarding 10 % of the energy from a long signal would not overly degrade the signal but a guiding methodology is needed. Further studies, which implement the suggested method on different pollutants acquired from different places would also be beneficial in supporting further the argument of the suitability of the method for the general case.

Acknowledgments This work was partially supported by the 7th European Framework Program (FP7) ENV.2012.6.5-1, grant agreement no. 308524 (CITI-SENSE), the Technion Center of Excellence in Exposure Science and Environmental Health (TCEEH) and the Environmental Health Fund (EHF).

References

AQMesh. (2015). Aqmesh website.

Boas, M. L. (1966). *Mathematical methods in the physical sciences* (vol. 2). New York: Wiley.

Cullis, C., Hirschler, M. (1980). Atmospheric sulphur: Natural and man-made sources. *Atmospheric Environment (1967)*, *14*(11), 1263–1278.

Duyzer, J., van den Hout, D., Zandveld, P., & van Ratingen, S. (2015). Representativeness of air quality monitoring networks. *Atmospheric Environment*, *104*, 88–101.

Fishbain, B. & Moreno-Centeno, E. (2016). Self calibrated wireless distributed environmental sensory networks. *Scientific reports*, 6.

Fishbain, B., Yaroslavsky, L. P., & Ideses, I. (2008). Spatial, temporal, and interchannel image data fusion for long-distance terrestrial observation systems. *Advances in Optical Technologies*.

Gupta, M., Shum, L. V., Bodanese, E., & Hailes, S. (2011). Design and evaluation of an adaptive sampling strategy for a wireless air pollution sensor network. In *2011 IEEE 36th Conference on Local Computer Networks (LCN)* (pp. 1003–1010). IEEE.

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution*, *151*(2), 362–367.

Künzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., et al. (2000). Public-health impact of outdoor and traffic-related air pollution: a european assessment. *The Lancet*, *356*(9232), 795–801.

Laumbach, R. J., & Kipen, H. M. (2012). Respiratory health effects of air pollution: Update on biomass smoke and traffic pollution. *Journal of Allergy and Clinical Immunology*, *129*(1), 3–11.

Leighton, P. (2012). *Photochemistry of air pollution*. Elsevier.

Lerner, U., Yacobi, T., Levy, I., Moltchanov, S. A., Cole-Hunter, T., & Fishbain, B. (2015). The effect of ego-motion on environmental monitoring. *Science of the Total Environment*, *533*, 8–16.

Levy, I., Mihele, C., Lu, G., Narayan, J., & Brook, J. R. (2014). Evaluating multipollutant exposure and urban air quality: Pollutant interrelationships, neighborhood variability, and nitrogen dioxide as a proxy pollutant. *Environmental Health Perspectives*, *122*(1), 65.

Mead, M., Popoola, O., Stewart, G., Landshoff, P., Calleja, M., Hayes, M., et al. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, *70*, 186–203.

Moltchanov, S., Levy, I., Etzion, Y., Lerner, U., Broday, D. M., & Fishbain, B. (2015). On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Science of The Total Environment*, *502*, 537–547.

Nazaroff, W., & Alvarez-Cohen, L. (2001). *Environmental Engineering Science*. John Wiley.

Peng, R. D., Dominici, F., & Louis, T. A. (2006). Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *169*(2), 179–203.

Rao, S. T., & Zurbenko, I. G. (1994). Detecting and tracking changes in ozone air quality. *Air & waste*, *44*(9), 1089–1092.

Robinson, E., & Robbins, R. C. (1970). Gaseous nitrogen compound pollutants from urban and natural sources. *Journal of the Air Pollution Control Association*, *20*(5), 303–306.

Schwartz, J., & Marcus, A. (1990). Mortality and air pollution j london: A time series analysis. *American Journal of Epidemiology*, *131*(1), 185–194.

Tchepel, O., & Borrego, C. (2010). Frequency analysis of air quality time series for traffic related pollutants. *Journal of Environmental Monitoring*, *12*(2), 544–550.

US Environmental Protection Agency. (2012). Epa national ambient air quality standards. Technical report.

Varotsos, C., Ondov, J., & Efstathiou, M. (2005). Scaling properties of air pollution in Athens, Greece and Baltimore, Maryland. *Atmospheric Environment*, *39*(22), 4041–4047.

Wang, T., Poon, C., Kwok, Y., & Li, Y. (2003). Characterizing the temporal variability and emission patterns of pollution plumes in the pearl river delta of China. *Atmospheric Environment*, *37*(25), 3539–3550.

Weinstein, B., Steyn, D., & Jackson, P. (2016). Modelling photochemical air pollutants from industrial emissions in a constrained coastal valley with complex terrain. In *Air pollution modeling and its application XXIV* (pp. 289–294). Springer.

Williams, D. E., Henshaw, G. S., Bart, M., Laing, G., Wagner, J., Naisbitt, S., et al. (2013). Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network. *Measurement Science and Technology*, *24*(6), 065803.

Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, *1*(01), 1–41.

Yaroslavsky, L. (2014). Signal Resotoration by means of linear filtering. In *Digital signal processing in experimental research—Fast transform methods in digital signal processing* (pp. 67–80). Sharjah, U.A.E: Bentham Science Publishers Ltd.

Zurbenko, I. (1986). *The spectral analysis of time series*.North-Holland, Inc.: Elsevier.