Hough-Transform-Based Interpolation Scheme for Generating Accurate Dense Spatial Maps of Air Pollutants from Sparse Sensing

Asaf Nebenzal, Barak Fishbain

Hough-Transform-Based Interpolation Scheme for Generating Accurate Dense Spatial Maps of Air Pollutants from Sparse Sensing

Asaf Nebenzal[1] and Barak Fishbain[2*]

[1] Department of Mathematics, Technion – Israel Institute of Technology, Haifa, Israel.
[2] Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa, Israel.

{asaf.n, fishbain}@technion.ac.il

Abstract. Air pollution is a significant health risk factor and causes many negative effects on the environment. Thus, arises the need for studying and assessing air-quality. Today, air-pollution assessment is mostly based on data acquired from Air Quality Monitoring (AQM) stations. These AQM stations provide continuous measurements and considered to be accurate; however, they are expensive to build and operate, thus scattered sparingly. To cope with this limitation, typically, the information obtained from those measurements is generalized with interpolation methods such as IDW or Kriging. Yet, the mathematical basis of those schemes defines that pollution extremum values are obtained at the measuring points. In addition, they are not considering the location of the pollution source or any physicochemical characteristics of pollutant hence do not reveal the real spatial air-pollution patterns. This research introduces a new interpolation scheme which breaks the interpolation process into two stages. At the first stage, the source of pollution and its estimated emission rate are inferred through a detection procedure which is based on the Hough Transform. At the second stage, based on the detected source location and emission, spatial dense pollution maps are created. The method requires, for its computation, to assume a dispersion model. To this end, *any* model can be used as sophisticated as it may be. Spatial maps created with simplified dispersion models in a computational simulation, show that the suggested interpolation scheme manages to create more accurate and more physically reasonable maps than the state-of-the-art.

Keywords: Air-quality modeling; Spatial maps; Interpolation; Hough transform; Source detection

1 Introduction

Air pollution is a significant risk factor for multiple health situations including eye irritation, breathing difficulties, lung cancer, heart diseases and respiratory infections [1].

In addition, air-pollution causes many negative effects on the environment like decreased visibility, acid rain, global warming, climate change, water quality deterioration and ecosystems destruction [2]. Thus, arises the need for studying and assessing air-quality's characteristics, dispersion patterns and behavior.

Today, numerous air-pollution studies are based on data acquired from Air Quality Monitoring (AQM) Stations [3]. However, AQM are typically scattered sparingly, mainly near main roads, industrial factories, or near highly populated areas [4]. Thus, the AQM network has a limited ability to account for spatial variability of pollution levels in heterogeneous regions, such as urban areas, which in return, renders exposure assessment as a difficult task [5]. To cope with the measurements sparsity, the information obtained from those measurements is often generalized with mathematical methods to improve the spatio-temporal coverage. To this end, interpolation schemes are sought.

Interpolation is a mathematical method of constructing a continuous function that obtains the measured values (or close values) at the measuring point. Environmental interpolation is based on the assumption that data attributes are continuous over space and spatially dependent [6]. Grossly speaking, interpolation methods can be divided into deterministic and geostatistical methods. The first include Inverse Distance Weighted (IDW), Nearest Neighbor (NN) and radial basis functions [6], while the latter involve, for example, various types of Kriging methods [7]. Next, we focus on IDW and ordinary Kriging, owing to their frequent use in spatial maps creation.

There are many studies in the field of air pollution modeling that used IDW or Kriging for creating dense spatial map of air pollution. IDW, for example, was applied for examining the ratio between low birth weight and air pollution exposure during pregnancy [8]. In that research, the IDW interpolation was utilized for estimating PM10 levels at future mothers' home address. Clark et. al [9] examined the effect of early life exposure to air pollution on development of childhood asthma. For estimating the average exposure level of an area, IDW interpolation was applied. Ventura et. al [10] introduced multi-objective pollutant AQMs optimization. In their research, they applied a Kriging interpolation scheme for creating dense spatial maps. Sarigiannis and Saisana [11] used Kriging interpolation method to create pollution maps of CO and O3 as an additional input to their multi–objective optimization scheme, which was based on remote sensing satellites.

IDW and Ordinary Kriging are both well-known and widely used interpolation methods. However, these methods are not appropriate for creating air-pollution spatial dense maps for several reasons: The mathematical basis of those schemes defines that all interpolated values over the study area are essentially a weighted average of the measurements points, thus extremum values cannot be obtained at any other place than the measuring points. In addition, these methods do not consider the location of pollution sources or any physicochemical characteristics of pollution. Hence, the resulted dense pollution maps do fall short in describing accurately the real spatial patterns of pollution. Regarding these in the interpolation process is expected to result in better and more accurate interpolation methods.

This research introduces a Hough Transform-Based Interpolation (HTBI) method, which generates accurate dense pollution maps through finding sources' locations and

the utilization of an air pollution dispersion model. The Hough Transform is a mathematical method, originated in image processing, used for detecting geometric shapes, like lines, circles or ellipses [12]. The main idea is converting from representing the shape in x,y coordinates (Cartesian) system to a parametric space, where the feature of interest is best represented. In this research, a feature space, which will represent best the source location is devised.

The method consists of two phases: at first, based on ambient concentration and assuming a dispersion model, the HTBI detects the sources' emission rates and locations. Then, using this information, the interpolation scheme builds the continuous pollution field. The suggested HTBI scheme applies no constraint on the assumed dispersion model. Hence, any dispersion model found in the literature (e.g., [13]–[15]), as sophisticated as it may be, can be incorporated into the suggested scheme.

2 Methodology

2.1 Notation

The following notation facilitate the description of the method. Let $\{S\}$ be a set of sources of a specific pollutant, with emission rates $\{Q\}$. Let A be the specific pollutant's continues signal generated by $\{S\}$, defined over a geographical area Ω. $\{S\}$ are located at $\{\gamma\}\epsilon\Omega$. Let $\{a\}$ be a finite set of samples of signal A, taken in locations $\{\omega\} \subset \Omega$. Interpolation aims at estimating A over the entire space Ω, based on the set of samples $\{a\}$. This is achieved here by first finding sources' locations, $\{\gamma\}$. It is worthwhile noting that the discussion here is limited to a single pollutant interpolation, i.e., the generation of a dense map of the specific pollutant is based on a set of sparse measurements of the same pollutant.

2.2 Interpolation Scheme

Each sample $a_i \in \{a\}$, represents a measurement in ω_i. W.L.O.G, if we order $\{S\}$, and $\{Q\}$ is sorted so Q_i is the emission rate of source S_i; a_i is a weighted combination of the contributions from all the sources, $\vec{Q}$. Assuming a dispersion model, M, so the kth element of the vector $\overrightarrow{M_i}$ is the decay coefficient of source k, Q_k, in location *i;* sensor *i*'s measurement, a_i, is all sources contributions at *i* and is given by:

$$a_i = \overrightarrow{M_i} \cdot \vec{Q}^T \tag{1}$$

Consequentially, forming the set $\{a\}$ as a vector, all sensors' measurements can be represented by the following matrices multiplication:

$$\vec{a} = M \cdot \vec{Q}^T \tag{2}$$

Given $[M]$, we assume that there exists a matrix E, which satisfies:

$$\vec{Q} = [E] \cdot \vec{a}^T \tag{3}$$

For finding Q and γ, a search on the entire Ω is suggested. To this end, Ω is divided into N disjoint catchments. We assume that each catchment, $C_n \subseteq \Omega$ is small enough so the pollution is uniform all over it. For each of the catchments an estimated emission rate $\hat{Q}^i{}_n$ is calculated, based on accepted measurements from single sample a_i; where e is a single row of E:

$$\hat{Q}_n^i = e \cdot a_i \tag{4}$$

Thus, $\hat{Q}^i{}_n$ introduces the estimated emission rate from the single source S, had it was located at C_n, based on the single measured sample at a_i.
The same process is applied for all C_n for each of the sensors:

$$\vec{\hat{Q}}_n = [E] \cdot \vec{a}^T \tag{5}$$

Applying Equation (5), results in each C_n having its unique set of $\vec{\hat{Q}}_n$, one estimate for each sensor. Using the standard deviation (STD) of the estimates, the catchment with the lowest STD is the approximated location of S. Once the source location, γ, is obtained, the emission rate of S is estimated by the average of the catchment's estimates:

$$\hat{Q}_\gamma = \bar{\vec{\hat{Q}}}_\gamma \tag{6}$$

Having the estimated emission rate Q, of the source S, and its estimated location, γ, with the dispersion model M, we can now estimate the dense pollution map over Ω:

$$C_n = \vec{M} \cdot \hat{Q} \tag{7}$$

The process is illustrated in the simple example of **Fig. 1**, where three sensors are deployed in a region with one source (see **Fig. 1**a). While the catchments can assume any geographical region and shape, for the sake of simplicity, the region, Ω, is divided into squared catchments, forming a squared grid. Sensor 1, which measures a pollution level of 33 (i.e. a_1 = 33), is located at catchment (1,3); Sensor 2, located at (2,4), measures 29; and Sensor 3, at (3,3) measures a level of 30. Keeping in mind the source's location, γ, is unknown, **Fig. 1**b demonstrates the execution of Equation (4), where each catchment is assigned with the estimated source's emission rate if the source was located in this catchment, given Sensor 1's measurement, and an exponential isotropic decay dispersion model, with an extinction coefficient λ. I.e., for r, the Cartesian distance from the source, the pollution level at each location on the map is given by [16]:

$$a_i = Q \cdot e^{-\lambda |r|} \tag{8}$$

If the source was located at (2,2), then the estimated emission rate, $\hat{Q}$, based on Sensor 1's measurement, should have been 38. If the source was located at (1,4), then $\hat{Q}$, according to Sensor 1, would be 47.3. **Fig. 1**c and **Fig. 1**d are the estimation maps, generated in the same fashion as b, for Sensor 2 and Sensor 3 respectively.

Assuming the dense pollution maps are a collection of isolines, the estimated emission rate values of the three sensors should agree in one grid location [17]. To evaluate the agreement, we compute, in each C_n, the three sensors' estimates' standard deviation. The lower the STD, the higher the agreement. This is illustrated in **Fig. 1**e. The smallest STD, indeed is obtained at location (1, 1), where, in this example the source is located.

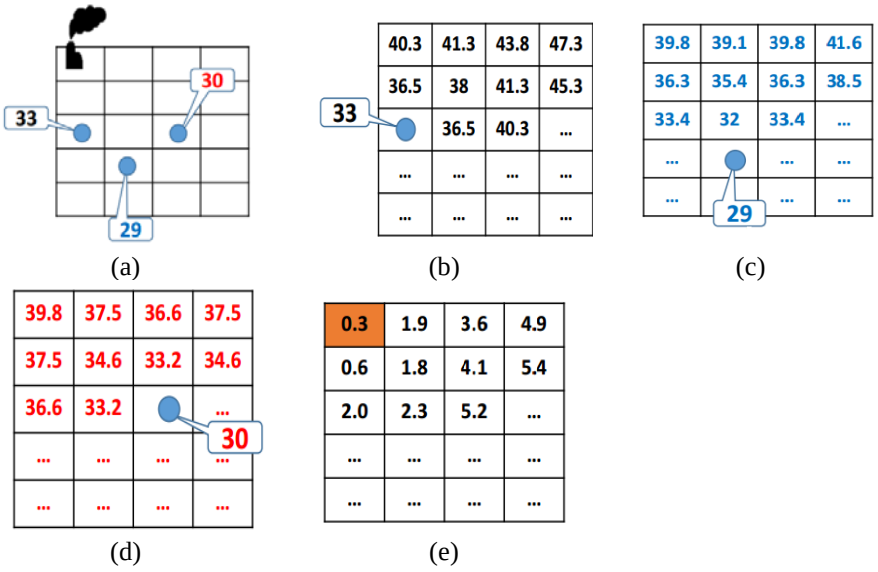


Fig. 1. Source location identification through HTBI assuming simple radial dispersion model.

3 Results and Discussion

3.1 Dispersion Models

As mentioned earlier, M represents the pollution decay function of the dispersion model. The suggested scheme, HTBI, does not apply any constraint on the dispersion model used. It can be any model, as long as it allows to compute the expected pollution on any given location on the map, given the emission rate Q and all other meteorological parameters required by the specific model in use. In this research, two models were used, the above isotropic decay dispersion model [16] (Equation (8)) , and the well-known Gaussian Plume Dispersion (GPD) model [18]:

$$a_i(\mathrm{x},\mathrm{y},\mathrm{z}) = \frac{Q}{2\pi\sigma_y\sigma_z\bar{u}}\exp\left(-\frac{y^2}{2\sigma_y^2}\right) \cdot \left[\exp(-\frac{(z-H)^2}{2\sigma_z^2}) + \exp(-\frac{(z+H)^2}{2\sigma_z^2})\right] \quad (9)$$

Where x is the downwind, y is the crosswind and z is the vertical distances of a_i from the source; $\bar{u}$ is the time-averaged wind speed at the hight of release H; and σ_y and σ_z represent the standard deviations of the crosswind and vertical Gaussian distribution of the pollutant concentration, respectively. The model also assumes full reflection from the ground.

3.2 Computational Simulation

For generating a continuous pollution field, the two types of dispersion models, described above were used. Specifications of the models are: Q =8 ton/hr; wind speed (for the GPD model):4 m/hr; wind direction (GPD model): 285°; effective stack-height: 120m.

The continuous fields were sampled by the set of sensors described in **Fig. 2** . To simulate real conditions, additive white Gaussian noise with Signal to Noise Ratio (SNR) of 10% (10 dB) was added to the readings of the sensors. Each sensor is now reporting the ambient level in its location as derived from the dispersion model with noise added. See **Table 1** for ambient data measured in each sensor, for the radial and the GPD models.

Using *only* the noisy readings obtained from the sensors, $\{a\}$, the source's location is estimated and then the dense pollution maps are created.

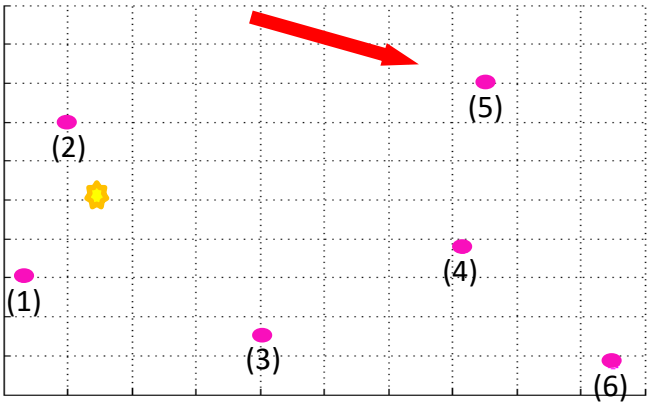
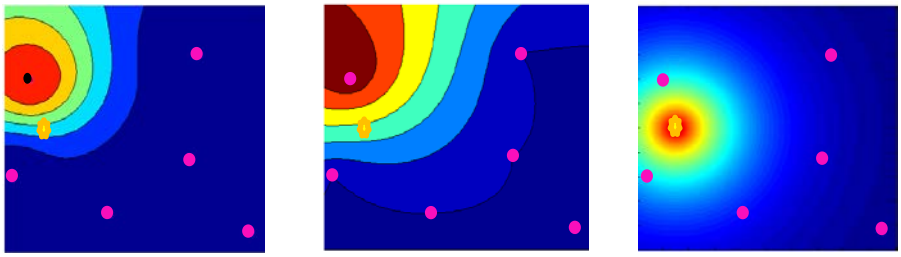


Fig. 2. A 20 km^2 geographical area, Ω. The sensors' locations are marked with pink circles, with a serial number below and the source (which typically its location is unknown) is marked by yellow star. Wind direction represented by red arrow.

Table 1. Ambient data measured by the sensors (units are in μg/m^3) for the radial and the GPD dispersion models

Sensor #	Radial	GPD
(1)	0.180	0
(2)	2.586	0
(3)	3.359*e-10	6.590*e-30
(4)	1.606*e-16	282.5981
(5)	1.659*e-20	6.20*e-17
(6)	1.233*e-30	1.1097

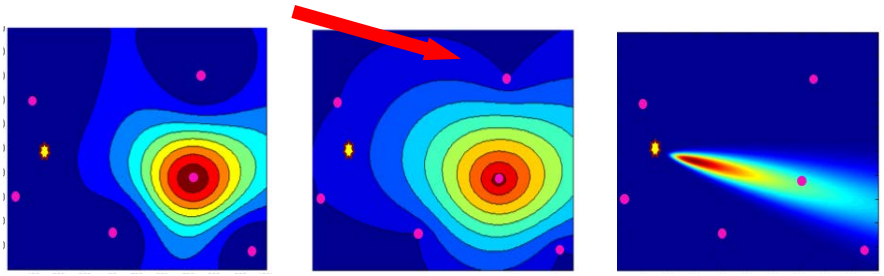
The results obtained for the radial dispersion model (Equation (8)) are displayed in **Fig. 3**. The highest ambient pollution level is located at the source location and exponentially decay as moving away. However, both IDW and Kriging models create a pollution map in which the maximum pollution level is obtained at the closest sensor to the sources' location, and decay as the distance from the source decreases (**Fig. 3**(a) and (b) respectively). HTBI, on the other hand, find the accurate source's location and then computes the accurate dense pollution map (**Fig. 3** (c)).



(a) (b) (c)

Fig. 3. Dense pollution maps based on the *radial dispersion model*: (a) IDW; (b) Ordinary Kriging; (c) HTBI. Sensors' locations are marked in pink and the source's location is marked with a yellow star. Pollution level is represented in a blue (low)- to red (high) color scale.

The interpolation results for the GPD model are presents in **Fig. 4**. As both IDW and Kriging do not consider physicochemical characteristics nor atmospheric conditions, the maximum of the dense pollution maps is found at the closest sensor downwind from the sources' location (**Fig. 4**(a) and (b) respectively). Moreover, the created maps demonstrate a roughly radial dispersion around this point, which is not the true condition, due to the wind. HTBI, as presented in **Fig. 4**(c), does manage to create a dense spatial map which complies with the Gaussian plume behavior. This is attributed to the fact that the HTBI method does incorporates the Gaussian model, as it can incorporate any dispersion model.



(a) (b) (c)

Fig. 4. Spatial maps based on the Gaussian dispersion model: (a) IDW; (b) Ordinary Kriging; (c) HTBI. Pollution level is represented in a blue (low)- to red (high) color scale. The red arrow represents the wind direction.

The suggested algorithm is deterministic in nature, i.e., for the same input, the system will produce the same output. Therefore, the *uncertainty* in the system stems from the uncertainty of the measurements, i.e., measurements noise [19]–[22]. The results of **Fig. 3** and **Fig. 4** were obtained at a noise level of 10%. (SNR, of 10dB). For evaluating the robustness of the algorithm, different noise levels were tested with the system. The radial model (Equation (8)) showed stability even with up to 50% errors (SNR of 3dB). The Gaussian model's (Equation (7)) robustness showed dependency on the catchments size. For larger catchment sizes (e.g., cell size of $40m^2$), our algorithm showed stability up to 10% SNR. However, increasing the spatial resolution to a cell size of $20m^2$, the HTBI showed higher sensitivity to measurement noise and showed the correct source location and interpolation maps for noise levels of up to 5% (13 dB). For lower SNR values the algorithm faced difficulties in locating the source and consequentially generating the dense pollution maps.

4 Conclusions

IDW and Ordinary Kriging are well-known and commonly-used interpolation methods for creating dense spatial maps, however they are not considering the physicochemical properties of the pollution characteristics nor the source location, therefore not accurate for this task. In this research, we introduced the Hough Transform Based Interpolation (HTBI), a two-phase interpolation scheme, which addresses these limitations. At the first phase, the HTBI detects sources' locations and their estimated emission rate. Using this information, at the second phase, a dense pollution spatial map is built. The method incorporates an air-pollution dispersion model into its calculations. This may be *any* dispersion model that can be found in the literature. Comparing between the dense pollution maps created by the HTBI, IDW and Ordinary Kriging shows that the HTBI creates spatial maps, which represents the true pollution maps better and thus, is more

accurate and sensible interpolation scheme. However, this work showed a computational simulation of a simple configuration with only one emission source. Implementing the method to a real-word situation is challenging. Air pollution emitted from many sources including industrial zones and transportation (line source). Hence, HTBI should be adjusted to face with this complex situation of multi sources detection.

Despite the above, HTBI indeed can be used in its current form, a single source detection. We can imagine at least two scenarios in which such configuration applies. The first is indeed when a single source can be identified. For example, when considering SO_2 which is emitted mainly from factories, and the study area contains only single industrial zone. The second is a case of leaks and we would like to identify the leak's source. In these cases, HTBI will be able to produce better and accurate spatial pollution maps than the existing methods.

Current work, carried out these days, is focusing on the implementation of HTBI in exactly such scenarios.

Another aspect this work sheds light on is the number of sensors and the way they scattered in the study area. It is obvious that the higher the number of sensors, the easier it will be to locate the source. There is a need for further research in finding the optimal number of sensors in a given area. The parameters that should be considered are the size of the study area, the characters of it (an open area is not the same as crowded urban area.), the coverage capacity and accuracy of the sensors and more.

5 References

1. M. E. Heroux *et al.*, "Quantifying the health impacts of ambient air pollutants: recommendations of a WHO/Europe project," *Int. J. Public Health*, vol. 60, no. 5, pp. 619–627, May 2015.
2. M. Venkatadri and P. S.Rao, "A survey on Air Quality forecasting Techniques," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 1, pp. 103–107, 2014.
3. H. Özkaynak, L. Baxter, and K. Dionisio, "Air pollution exposure prediction approaches used in air pollution epidemiology studies," *J. Expo.* 2013.
4. E. Goswami, T. Larson, and T. Lumley, "Spatial characteristics of fine particulate matter: identifying representative monitoring locations in Seattle, Washington," *J. Air Waste*, 2002.
5. Rao *et al.*, "Environmental Modeling and Methods for Estimation of the Global Health Impacts of Air Pollution," *Environ. Model. Assess.*, vol. 17, no. 6, pp. 613–622, 2012.
6. A. Akkala and V. Devabhaktuni, "Interpolation techniques and associated software for environmental data," *Environ. Prog. Sustain. Energy*, vol. 29, no. 2, pp. 134–141, 2010.
7. J. Li and A. D. Heap, *A review of spatial interpolation methods for environmental scientists* . 2008.
8. X. Xu *et al.*, "PM10 air pollution exposure during pregnancy and term low birth weight in Allegheny County, PA, 1994-2000.," *Int. Arch. Occup. Environ. Health*, vol. 84, no. 3, pp. 251–7, Mar. 2011.
9. N. A. Clark *et al.*, "Effect of early life exposure to air pollution on development of childhood asthma.," *Environ. Health Perspect.*, vol. 118, no. 2, pp. 284–90, Feb. 2010.
10. A. Trujillo-Ventura and J. Hugh Ellis, "Multiobjective air pollution monitoring network design," *Atmos. Environ. Part A. Gen. Top.*, vol. 25, no. 2, pp. 469–479, Jan. 1991.

11. D. A. Sarigiannis and M. Saisana, “Multi-objective optimization of air quality monitoring.,” *Environ. Monit. Assess.*, vol. 136, no. 1–3, pp. 87–99, Jan. 2008.
12. P. Hough, “Method and means for recognizing complex patterns,” *U.S. Pat.*, vol. 3,069,654, 1962.
13. P. Hystad *et al.*, “Creating national air pollution models for population exposure assessment in Canada,” *Environ. Health Perspect.*, vol. 119, no. 8, pp. 1123–1129, Aug. 2011.
14. P. Zannetti, *Air pollution modeling: theories, computational methods and available software* ..
15. Y. Tominaga and T. Stathopoulos, “CFD simulation of near-field pollutant dispersion in the urban environment: A review of current modeling techniques,” *Atmos. Environ.* 2013.
16. M. Buhmann, “Radial basis functions: theory and implementations,” *Cambridge Monogr. Appl. Comput. Math.*, vol. 12, pp. 147–165, 2003.
17. D. H. Ballard, “Generalizing the Hough transform to detect arbitrary shapes,” *Pattern recognition*, vol. 13, no. 2, pp. 111–122, 1981.
18. D. Ermak, “An analytical model for air pollutant transport and deposition from a point source,” *Atmos. Environ. (1967)*, 1977.
19. B. Fishbain *et al.*, “An Evaluation Tool Kit of Air Quality Micro-Sensing Units,” 2015.
20. B. Fishbain and E. Moreno-Centeno, “Self Calibrated Wireless Distributed Environmental Sensory Networks,” *Sci. Rep.*, vol. 6, 2016.
21. U. Lerner, T. Yacobi, I. Levy, S. A. Moltchanov, T. Cole-Hunter, and B. Fishbain, “The effect of ego-motion on environmental monitoring,” *Sci. Total Environ.*, vol. 533, 2015.
22. S. Moltchanov, I. Levy, Y. Etzion, U. Lerner, D. M. Broday, and B. Fishbain, “On the feasibility of measuring urban air pollution by wireless distributed sensor networks,” *Sci. Total Environ.*, vol. 502, 2015.