

Pre-Thesis

In fulfillment of the requirements for the PhD candidacy examination

Optimal deployment of static sensors for industrial leak detection using a Multiobjective Evolutionary Algorithm (MOEA)

בעברית:

**Submitted by:**

Idit Belachsen

Faculty of Civil and Environmental Engineering Technion-Israel Institute of Technology

31.07.2019

**Advisors:** Associate Professor Barak Fishbain, Associate Professor Shai Kendler

**Abstract**

TBD.

**Introduction**

Degraded air quality is a consequence of heightened emissions from a wide range of pollution sources, both anthropogenic, such as transportation or industry and natural, such as dust plumes, wildfires or vegetation. Dense urban centers and industrialized regions serve as “hotspots” in a continuum of concentrations, dictated by the rate by which all emitted substances undergo transport, diffusion, chemical transformation and deposition to the ground (Atmospheric chemistry and physics). Fossil fuel combustion processes that govern todays’ industrial and transportation activities are major emitters of gaseous pollutants to the troposphere, mainly sulfur dioxide (SO2), nitrogen oxides (NOx, nitrogen monoxide (NO) and dioxide (NO2)), carbon monoxide (CO) and dioxide (CO2), volatile organic compounds (VOCs) and others. Particulate pollutants, known as particulate matter (PM), are emitted as well in combustion processes and can be divided to primary particles (such as black carbon (BC)) and secondary particles (like sulfates and nitrates) which are formed in the atmosphere by oxidation of primary gaseous pollutants. (coarse/fine/ultrafine particles?) Another secondary pollutant is the ozone (O3). Like other secondary pollutants, it is produced naturally in the troposphere by photochemical oxidation of primary pollutants by the hydroxyl radical (OH). Many other anthropogenic pollutants are emitted from various industrial processes, among them are X that is emitted from Y, X1 from Y1.

The main incentive to reduce and control emission rates from anthropogenic sources is of course insuring (assuring?) population health. World health organization (WHO) estimates that 4.2 million premature deaths globally are linked to ambient air pollution, mainly from heart disease, stroke, chronic obstructive pulmonary disease, lung cancer, and acute respiratory infections in children. Pollutants with the strongest evidence for public health concern include particulate matter (PM), ozone (O3), nitrogen dioxide (NO2) and sulfur dioxide (SO2) (WHO,2019).

Epidemiologic studies (others as well?) try to evaluate (assess?) past and present population exposure to air pollution and correlate the level of exposure to observed health effects in the population. The major challenge in this type of work is in producing accurate pollution concentration maps of high spatial and temporal resolution that can enable finding such correlations at a personal level (e.g. estimating concentrations in the exact place of residence of a subject). These studies rely on data of ambient pollution concentrations usually obtained by two methods; i) routine measurements reported by standard air quality monitoring (AQM) stations, ii) short-term measurement campaigns which usually utilize large number of sensors. Evaluations using the first method are considered very accurate, since AQM stations are equipped with expensive, pollutant-designated measuring devices, that are maintained and calibrated on a regular basis by regulatory authorities. However, these tend to suffer from a few apparent flaws; i) their span is usually sparse, and cannot represent well the spatial and temporal variability of a typical pollutant (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233947/), ii) samples of air are taken a few meters above street level and hence cannot represent well the extent of exposure of a passerby, and iii) they are costly to maintain. Short-term measurement campaigns on the other hand can provide a higher spatial resolution of the desired region of research, but for a limited time span only, that is rarely adequate for a comprehensive epidemiological study.

WDESN… + CITIZEN SCIENCE

In the past couple of years, citizen science projects had become more common, and large amounts of air quality data are being collected today by individuals, usually using low-cost sensors of various types. Using such data however requires an exhaustive preprocessing work as reliability of some of the measurements might be questionable.

Spatial interpolation methods (Kriging, inverse distance weighting (IDW), spline) may help overcome sparse representation of concentration and can be used to produce pollution dense maps of certain locations and times. Land use regression (LUR) models can link environmental variables (e.g., road type, traffic volume, topography, land cover) with measurements in monitoring locations and then use these variables as model predictors at unmonitored locations, and possibly as predictors in time as well (?). Nevertheless, atmospheric transport and dispersion (ATD) models are the only models that can quantify the deterministic relationships between sources’ emissions and concentrations in space. ATD models can forecast the spread of the pollutants, when available results from monitoring data are used for calibration and evaluation of model performance. Population exposure estimation is one possible application of ATD models. Oftentimes, these models serve as a useful tool for regulatory authorities to assess baseline ambient concentrations, analyze the relative importance of various emission sources, or test emission reduction strategies (Stein, Ariel F).

For an accurate forecast, several variables are needed as input to ATD models, including, among others: meteorological data, the strength of the release and its location. While meteorological data of reasonable resolution (of a few square kilometers and 15-30 minutes) are usually available­­ from local weather stations or weather models, the **strength**, **time or even location of the release** of the pollutant **are frequently unknown.** As a consequence, sensor measurements are often being used to determine source parameters in an inverse modelling problem of source-term estimation (STE) (review paper). In these methods, source parameters (sent as input to an ATD model) are modified until the difference between the calculated and observed measurements is minimal. A certain optimization technique is usually applied in order to search the solution space and reach an accurate evaluation. Once the source term is calculated, it is used as input to an ATD model that can generate high resolution concentration maps.

As mentioned above, the momentary rate in which pollution is emitted from most sources is usually unknown. In Israel for example, an inventory of pollutants released to the environment by all industrial businesses (production plants?) is produced by the ministry of environmental protection on a yearly resolution, and is based mainly on the plant’s statement. Some plants possess monitoring equipment that enables high resolution evaluation of pollutant emission rates, but the majority (?) rely on occasional samplings which only enable the calculation of a ratio between the emission rate of the pollutant and the production rate during the time of the sampling. Most plants are given a certain emission permit that refer to a certain average emission rate. Certain standard deviations are allowed, with a unique value for each pollutant. Enforcement of these permits relies on the ministry’s inspection.

The motivation behind estimating source parameters in an adequate temporal resolution is now clear. Regulatory authorities may wish to enforce plants that temporarily exceed their emission permits, reveal leaks or new unknown sources. Epidemiological studies may wish to account for high temporal variability of pollution concentrations since health impacts might also be triggered by short-term exceedances of pollution concentrations as well as long term exposure. Such exceedances might be missed in studies that focus on long-term continuous exposure to average daily, monthly or yearly pollution levels.

Nevertheless, estimating source parameters is not an easy task. As the number of sources increases, the complexity of the problem increases and a rigorous deployment of a network of sensors is needed to provide the right representation of the contamination levels in the field. Measurements achieved by this network…. Many approaches exist to what is the best way to span such a network, considering the limitations and constrains (e.g. cost, location, human resources available) that characterize the region. These will be reviewed thoroughly in section X.

The following proposed work offers a modelling tool that can serve stakeholders when either planning the positioning of a stationary network of sensors or planning a span of sensors during a routine sampling task. It is based on meteorological data (wind speed and direction and atmospheric stability) and on a quantitative measure of the complexity of the given set, making it possible to evaluate how hard it may be to separate pollution plumes that overlap. Optimization….

**Research objectives**

TBD.

**Research contribution**

TBD.

**Literature review**

Sensors placement, WDESN, optimization methods, dispersion model types used?, evolutionary algorithms

**Methods and Research plan**

1. **The Gaussian plume model**

Atmospheric transport and dispersion (ATD) modeling refers to the mathematical description of pollutant transport in the atmosphere. The term dispersion is comprised of diffusion (due to turbulent eddy motion) and advection (due to wind) that occurs within the air near the Earth’s surface [1]. Several types of fundamental dispersion models exist: box models, Gaussian plume models, Lagrangian models, Eulerian dispersion models and Dense gas models as well as extensions and combinations of all of the above [2].

The Gaussian plume model is one of the simplest and widely used models that offers an analytical solution to the advection-diffusion equation for idealized circumstances, corresponding to a continuous point source that emits pollutants into a unidirectional wind blowing in a domain of infinite extent (see illustration is figure X). The advection-diffusion equation can be derived by the mass conservation equation for the pollutant concentration C [kg/m3]:

﻿where S [kg/m3s] is a source or sink term and the vector function represents the mass flux [kg/m2s] of a pollutant owing to the combined effects of diffusion and advection. ﻿The diffusive flux () is proportional to the concentration gradient through the diffusion/turbulence coefficient K [m2/s]:

﻿The second contribution to the flux is due to simple linear advection by the wind, which can be expressed as:

﻿Substitution X into the equation of conservation of mass (X) yields the three-dimensional advection-diffusion equation:

The Gaussian plume model, which is the solution of the three-dimensional advection-diffusion equation (x), eventually describes the pollutants’ concentration C [kg/m3] in a certain position in space:

Where Q [kg/sec] is the release rate from the stack, H [m] is the effective height (i.e., the sum of the actual stack height h and the plume rise Δh), x, y and z [m] are the downwind, crosswind and vertical distances, respectively, and [m/s] is the mean wind speed at the height h of the release. and [m] are key parameters in the model and represent the standard deviation of the gaussian concentration distribution in the crosswind and vertical direction (figure X).

For simplicity, the steps for the solution of equation X were not specified. However, it is important to mention the simplifying assumptions that had to be made in order to reach the Gaussian plume model equation (X) [1]:

1. The pollutant is emitted in a constant rate Q from a single point source located at height H above ground surface.
2. The wind velocity is constant and aligned with the positive x-axis.
3. ﻿The solution is steady state.
4. ﻿The diffusion/turbulence coefficient K is a function of the downwind distance x only, and diffusion is isotropic so that .
5. ﻿Wind velocity is sufficiently large that diffusion in the x-direction is much smaller than advection.
6. ﻿Variations in topography are negligible so that the ground surface can be taken as the plane z=0.
7. ﻿The pollutant does not penetrate the ground.

To these assumptions we add that the material diffused is a stable gas or aerosol, with a negligible deposition rate and that background pollution is negligible. parameters are used instead of the diffusion/turbulence coefficient , due to the fact that they are much easier to determine experimentally and can usually be described by a simple power law of the form: , . This kind of dependence on downwind distance (x) expresses the widening of the plume when moving further from the source. ﻿Experimental measurements have been used to estimate the coefficients a, b, c, d and f under a variety of atmospheric conditions, and the most used values are those obtained by Turner 1994, using Pasquill-Gifford stability class categories [Pasquill, 1961]. These coefficients are specified in Table X.

In order to apply the suitable parameters to the Gaussian plume model, the Pasquill-Gifford stability category which characterizes the simulated set, should first be determined. The original method to do so, developed by Turner 1964, requires knowledge on cloud cover and cloud ceiling. Alternative methods were developed for situations where these data are not available. They include a radiation-based method which uses measurements of solar radiation during the day and delta-T at night and turbulence-based methods which use wind fluctuation statistics [Meteorological Monitoring Guidance]. For the current simulation, we chose to use a turbulence-based method ( method), because of its simplicity and our currently available data. ﻿The method (sometimes referred to as ) uses the standard deviation of the wind direction () in combination with the scalar mean wind speed (), to determine the Pasquill-Gifford stability category (see Tables X and X). Wind direction standard deviation values of 10-min resolution were obtained from the Israel Meteorological Service (See section data). To minimize the effects of wind meander (﻿long period oscillations associated with light wind speed conditions), it is recommended to calculate the 1-hour value using 10-min or 15-min averages, as specified in equation X (i.e., calculating the root mean square) [Meteorological Monitoring Guidance].

Turbulence is essentially the motions of the wind over the time scales smaller than the averaging time used to determine the mean wind. Turbulence consists of circular whirls or eddies of all possible directions that serve to disperse pollutants by mixing with air having lower pollutant concentrations. It is common to divide the causes of turbulence to mechanical turbulence and buoyant generation of turbulence. Fluctuations in wind direction may represent mechanical turbulence that is formed by the friction with Earth’s surface. Distinguishing between daytime and nighttime takes into consideration buoyant turbulence generated by … all kinds of processes….NOT SURE ABOUT THIS

1. **Meteorology**

Sources: Meteorology today, atmospheric chemistry and physics, workbook

As reviewed above, the meteorological parameters that play a role in the Gaussian plume model are wind speed, wind direction, and the atmospheric stability condition. The stability condition is a function of the vertical temperature profile of the atmosphere, which is govern by winds (cold or warm advection) or processes such as solar heating and radiative cooling. In general, three regimes of atmospheric stability exist: **unstable, neutral and stable**. In an absolutely **unstable** atmosphere, the environmental lapse rate (the rate at which air temperature changes with altitude) is greater than the dry adiabatic lapse rate (100 for every 1000 m), and an air parcel that starts to move upward will continue to rise because it is warmer and less dense than the air around it. Pollutants will be mixed rapidly in extreme vertical motions. In an absolutely **stable** atmosphere, the environmental lapse rate is less than the moist adiabatic rate (60 for every 1000 m) and an air parcel will resist an upward vertical motion and will tend to spread out horizontally. If the temperature of the atmosphere increases with altitude in a certain layer in the atmosphere, then this layer is called an inversion, which is a form of an **absolutely** **stable** atmosphere as well. Such conditions might lead to high pollution concentrations and minimal amount of dispersion. Between the above two extremes is the condition characterized as neutral. In this condition, the environmental lapse rate is characterized by a slight decrease of temperature with height, close to the dry adiabatic rate. These conditions can be caused for example by transitional situations near sunrise or sunset when the atmosphere changes its state from stable to unstable or vice versa. With neutral atmospheric conditions, intermediate level of dispersion takes place. It should be mentioned that meteorologists define a **conditionally unstable** atmosphere to distinguish between the absolutely stable and absolutely unstable conditions (instead of a neutral atmosphere). In this situation, the environmental lapse rate lies between the moist and dry adiabatic rates, and the stability of the atmosphere depends on whether the rising air is saturated or not. When the air parcel is not saturated, then the dry adiabatic lapse rate is the relevant reference state and the atmosphere is stable. For a saturated air parcel inside a cloud, the reference is the moist adiabatic rate and the atmosphere is unstable.

Add somewhere about day and night, also because of the PED separation.

Workbook +

*A picture containing text, newspaper

Description automatically generated*

A picture containing text, newspaper

Description automatically generated

1. **Evolutionary algorithms and Borg MOEA**

אלגוריתם אבולוציוניים

כלי האופטימיזציה בורג

1. **Pairwise Euclidean distance (PED)**

A criterion for evaluating how well a hypothetical network of sensors, placed in hypothetical locations, would be able to distinguish between overlapping plumes, was defined for the optimization problem. The criterion is the pairwise Euclidean distance (PED) between calculated readings of two sensor network sets with different number of active sources. That is, the PED of

two sets of active sources with different sizes {S}' and {S}" is:

In a generated network set that is able to distinguish well between overlapping plumes (if given a source term estimation (STE) problem), a change in the number of active sources is expected to have a dramatic effect on the sensor network readings, resulting in high PED values. On the other hand, low PED values will characterize a network of sensors that will not be able to sufficiently resolve a problem of source separation and estimation. An example of a matrix of PED values is shown in Figure X for different combinations of 1-5 active sources. In our defined problem, we use the PED calculation as an objective function we wish to maximize. SEPARATE DAY AND NIGHT.

1. **Data**

**מטאורולוגי**

For our simulation set, we used 10 min wind data obtained from the Israel Meteorological Service (IMS) for the years X and for one station Nevatim/Negev junction. This station was chosen since in future work, we may wish to implement our methodology on a real-world problem, possibly in an idealized location such as Ramat Hovav industrial zone in southern Israel.

Emission data were obtained from the Ministry of Environmental Protection. These data were used for “evaluation” of our methodology.

1. **Methodology – problem formulation and optimization**

**General problem formulation**

Let be the region of interest, i.e., the industrial area and its surroundings, where we wish to place our network of sensors in. Let be the set of sources, where each source, is located in and for a specific time t, emits [Kg/sec]. Similarly, the set is the set of sensors, where each sensor, , is located in and records a pollution level of . The location of the sources are known, these are the locations of the industrial plants.

Let be the pollution transfer function of the dispersion model, which associates sensor’s r readings, , with the emissions of source s. Thus, the model’s estimated contribution of source to the pollution level in , is given by:

For multiple sources scenario, each sensor readings are simply the sum of all sources’ emissions, i.e.:

As each source owns its unique parameters and location with respect to the sensors, the values of for each source-sensor combination is determined by the dispersion model, in our case, the Gaussian plume model.

**Our problem formulation**

In our problem, we wish to find an optimal deployment of a set of sensors . In general, when the number of sensors () increases, the PED values, calculated between two sets of active sources with different sizes {S}' and {S}", increases as well (see equation X). Given that the span of a set of static sensors is an expensive task due to sensors’ cost and maintenance, we may wish to minimize the number of sensors in use, and place them in a set of locations that together form the highest PED value, which represents the best ability of those sensors to separate the different sources. The first objective is then to find the minimal number of sensors deployed:

It is possible to phrase several different objective functions that maximizes the PED value. Our algorithm seeks to find the optimal set of sensors , so that the percentage of cases where PED values are greater than a certain threshold is maximized, considering all possible combinations of two sets of number of active sources with different sizes {S}' and {S}". For example, in the case of 5 sources, 31 options of 1-5 active sources exist, creating a total of 355 combinations (i.e., 355 PED values) of two sets of number of active sources with different sizes {S}' and {S}". The second objective we define is then phrased as:

The network of sensors deployed should consider the varying meteorological conditions during the year. Hence, our objective takes into account the discrete frequencies of each possible meteorological combination of wind speed, wind direction and Pasquill–Gifford atmospheric stability class. Hourly mean wind directions were binned into 16 categories (22.5o each) and hourly mean wind speeds into 7 categories (<1, 1-2, 2-4, 4-6, 6-8, 8-10, >10 m/s). With six possible Pasquill–Gifford classes for atmospheric stability (A-F), 672 weather situations exist, out of them only X are physically meaningful (i.e., were extracted from the data archive used). In addition, all possible situations were applied in the Gaussian plume model to obtain steady state solutions of a pollution concentration map. Each concentration map was then given a weight according to the frequency distribution derived.

Furthermore, objective 2 can be separated to two different objectives, one which considers PED values obtained during night hours and one which considers PED values obtained during the day:

This required creating a separated frequency distribution of weather states for day and for night, in the same way described above.

**Technical description of the optimization**

Once we have a weighted map of concentrations computed in advance, the optimization process starts. The Borg MOEA algorithm (see section X) searches for the optimal set of sensors by starting with an arbitrary guess. Then the algorithm projects the set on and evaluates all PEDvalues of that set (equation X). Then, based on the two objectives ( and ), the Borg refines the set of decision variables, to locate a set of solutions on the Pareto frontier. Technically, we assume that each location may host a sensor. The decision variables are held in a data array, each value represents a sensor, linked to a certain location . The sensor is “placed” in that location if the decision variable is 1 and is not “placed” if the decision variable is 0.

1. **Simulation set (research plan?)**

We simulated a 1000x1000 meter flat area. The simulated sensors were assumed to be situated at ground level, and so equation X was reduced to:

Five point-sources (stacks) were given average yearly emission rates of 1000, 1500, 600, 1900 and 300 kg/s. All model runs used the same average yearly release rates of the sources, and not some momentary emission rates (), assuming these are usually not available.

Figure X depicts some of the configurations. Sources are marked in red circles (o) and optional locations of sensors are marked in black crosses (+). Optional locations are spread in a grid, 50 m apart. Stack heights (effective heights) were taken at 10 m, following the height of the obtained wind data.

no solution in 100 meters?

**בעתיד(research plan?):**

**ולידציה לתוצאות**

**לקחת מקרה אמיתי (רמת חובב)**

**להשתמש במודל דיספרסיה מתוחכם יותר, אולי אין צורך?**

**להשתמש במקרה בעולם האמיתי. המעבר לעולם האמיתי הוא קשה. הוא דורש טופוגרפיה, כיסוי שטח, וידע על כל המקורות באזור המחקר וריכוזי רקע. וכנראה גם מודל מתוחכם יותר.**

**הדאטה של שיקגו**

**הצעה של שי – לא זוכרת !**

**Initial results**

A simulation using a synthetic set of stacks was performed in order to evaluate the proposed method for sensors’ placement. Usually, several constraints exist when placing static sensors; i) law requires sensors to be employed in populated areas, i.e., inside neighborhoods, ii) placing sensors inside the industrial region could have an effect on the network ability to separate between sources (?), iii) other physical constraints that may characterize the given study area (vegetation, public buildings).

This is especially important in the case of complex industrial site, where access to the leak source might require significant amount of resource such as trained personal, protection gear etc.

Basic idea:  using PED, using winds and atmospheric stability states that characterize the area.

**Simulation 1:** all sensors are the same and have perfect ( detection threshold is zero ?) sensitivity and infinite dynamic range (?).

**Simulation 2:** running a simulation for reasonable sets of detection thresholds and dynamic ranges. See what type of sensor suits.

**Simulation 3:** ~~change the density of the stacks inside a defined area of constant size.~~ Run for average yearly emissions Vs. table of factors emissions (my problem is with stacks that don’t work and then my PED calculation is for 5 and not less stacks). (in a different simulation I take the hourly averages taken from the ministry of environmental protection?).

Assumptions: the optimization process is based on average emission rates of the stacks.So basically we don’t take into consideration activity hours of the plants. These are indirectly considered during the optimization process by PED calculations. Technically we could correlate weather conditions with stacks’ working hours.

Such work may serve regulators and stakeholders.

**What to display:**

1. Pollution dense map for an average wind direction, speed, and average emission rates of stacks for substance X.
2. Distribution of wind speed and direction of the area.
3. Distribution of stability classes.
4. Assuming a sensor in each node of the grid, what is the distribution of my concentrations?
5. Show a map of the weighted PED values (maybe higher resolution than the grid I run).
6. Show a map of solutions for X sensors - an average of 10 runs. Is it clustered in groups?
7. Perform random seed analysis
8. Show the average pareto front of 10 runs.
9. Show pareto fronts of different sensors types.
10. Results for different density of stacks.
11. What is the PED of my set of sensors I just chose, for different emission scenarios?
12. להראות את מפת הריכוזים הממוצעים לעומת מפת הPED. לאו דווקא PED גבוה קורלטיבי עם ריכוזים גבוהים...

**Work schedule**

**References**

**הערות**

זה אולי לא המודל הכי מדויק אבל לצורך הבעיה שלנו הוא יכול להיות מספיק בהחלט בגלל שמסתכלים על ממוצעים שנתיים.

*The pollutant concentrations in the plume are inversely proportional to the wind speed along the x axis, as can be seen from equation X. For example, if we double the wind speed, the concentration in the plume would be half of that with a lighter wind.*

*Incorporating considerations of both…*

*השיטה של מהירות הרוח בעצם לוקחת בחשבון יותר טורבולנציה מכאנית בפני הקרקע. החלוקה ליום ולילה*

**איורים**

A close up of a map

Description automatically generated

A close up of a map

Description automatically generated Pasquill stability class categories: A screenshot of a cell phone

Description automatically generated

The criteria in Tables X and X are for data collected at 10 m and a roughness length of 0.15 m. Wind

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generated