

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)

**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.

**Literature review**

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

**Research objectives**

**Research contribution**

**Research plan**

**Methods**

1. **The Gaussian plume model**

Atmospheric transport and dispersion (ATD) modeling refers to the mathematical description of contaminant transport in the atmosphere. The term dispersion is used to describe the combination 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 the mentioned 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 contaminants 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 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 is the release rate from the stack [kg/s], H is the effective height (i.e., the sum of the actual stack height h and the plume rise Δh) [m], x, y and z are the downwind, crosswind and vertical distances [m], is the mean wind speed at the height h of the release [m/s]. and [m] are key parameters in the model and represent the standard deviation of the gaussian concentration distribution in the crosswind and vertical direction.

The solution steps will not be shown, however, the simplifying assumptions that had to be made in order to reach the above stated solution (X) should be stated:

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. This assumption is reasonable if the wind velocity and all other parameters are independent of time and the time scale of interest is long enough.
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.

parameters are used instead of the diffusion/turbulence coefficient K [m2/s], due to the fact that they are much easier to determine experimentally. ﻿A variety of functional forms have been proposed, with one of the most common being a simple power law σ2(x)= axb [33]. Experimental measurements have been used to estimate the coefficients a and b under a variety of atmospheric conditions, and typical values are shown in Table 3.1. This type of dependence of σ on downwind distance can be justified by noting that as one moves further from a source, the plume becomes broader and hence σ must increase

Z=0 we take, so solution looks like:

For multiple it is just the sum

scale of models

no solution in 100 meters?

A close up of a map

Description automatically generated

1. **Meteorology**
2. ?). סוגי יציבות, מה ההגדרה של אטמוספירה יציבה לעומת לא יציבה, מפל אדיאבטי...
3. **Evolutionary algorithms and Borg MOEA**

אלגוריתם אבולוציוניים

כלי האופטימיזציה בורג

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

מה זה PED, הפרדה בין מקורות

1. **Data**

**מטאורולוגיה**

נתונים מטאורולוגיים - מהירות וכיוון רוח ויציבות אטמוספרית (חישבתי לפי… מעידה גם על קרינה סולארית וטמפרטורה בעקיפין(

התייחסות לסוג המזהם

1. **Simulation set**

גובה הארובה צריך להיות בגובה נתוני הרוח שאני לוקחת! למרות שנראה לי שיש הנחה שהרוח קבועה בכל גובה. 5 מטר זה בכל מקרה נמוך מידי ובגלל זה הריכוזים שלי לא הגיעו רחוק.

**בעתיד:**

**להשתמש במודל דיספרסיה מתוחכם יותר**

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

**אולי לוותר על המודל דיספרסיה ולעשות ML ?**

**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).

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**