

Pre-Thesis(research proposal?)

In fulfillment of the requirements for the PhD candidacy examination

Optimal deployment in time and space of heterogenous sensor array for better representation of air pollution and its sources

פריסה אופטימלית בזמן ובמרחב של רשת חיישנים הטרוגנית לייצוג טוב יותר של זיהום האוויר ומקורותיו

**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

## Air pollution

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 [1]. Fossil fuel combustion processes that govern todays’ industrial and transportation activities are major emitters of gaseous pollutants to the troposphere, mainly nitrogen oxides, NoX, i.e., 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 which are formed in the atmosphere by oxidation of primary gaseous pollutants. Another secondary pollutant is the ozone (O3), produced naturally in the troposphere by photochemical oxidation of primary pollutants and mostly a summer problem in urban environments [2]. Many other anthropogenic pollutants are emitted from various industrial processes, among them are ammonia, methane and fluorinated gases, emitted for example from the fertilizer industry, agricultural practices or waste decay, and aluminum or semiconductor manufacturing, respectively [3]. Emissions are commonly separated to two types: point source emissions - channeled through a pipe, such as a chimney stack or a vent, and non-point source emissions – caused by direct contact of volatile substances or PM with the environment, where tracing pollution back to a single source is difficult [4].

The main incentive to reduce and control emission rates from anthropogenic sources is of course assuring population health. World health organization (WHO) estimates that 4.2 million premature deaths every year 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 fine and ultrafine particulate matter (less than 2.5 and 0.1 in diameter, respectively), ozone, nitrogen dioxide and sulfur dioxide [5]. Another not less important aspect is the affect air pollution has on climate. Climate change is driven by air pollution, as many pollutants disturb the steady state condition in earth’s energy balance when interacting with solar and terrestrial radiation, leading to changes in earth’s surface temperature and in climate [6], [7].

## Monitoring air pollution

Monitoring air pollution is therefore necessary and may serve for preventing population exposure by detecting pollution peaks, for urban planning and development, for climate research or for public health studies that try to assess past and present population exposure to air pollution and correlate the level of exposure to observed health effects [8]. Regulatory authorities monitor air pollution to enforce plants that temporarily exceed their emission permit, reveal leaks or new unknown sources. For all these purposes, the major challenge is in producing accurate pollution concentration maps of high spatial and temporal resolution. In epidemiological studies for example, it can enable finding correlations at a personal level (e.g. estimating concentrations in the exact place of residence of a subject) or help in finding health impacts triggered by short-term exceedances of pollution concentrations.

Ambient air pollution concentrations are usually obtained by two methods: i) routine measurements reported by standard air quality monitoring (AQM) stations [9] and ii) short-term measurement campaigns which usually utilize large number of sensors [10]–[12]. Data obtained by the first method are considered very accurate, since AQM stations are equipped with pollutant-designated measuring devices of high quality, 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; ii) samples of air are taken a few meters above street level and hence cannot represent well the extent of exposure of a passerby, if required, 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 only, and thus, would fail to account for seasonality or any phenomenon that presents longer cycles than the time of the campaign.

An alternative or complementary solution is to use Wireless Distributed Environmental Sensor Network (WDESN), usually comprised of portable and relatively low-cost Micro Sensing Units (MSUs), that can measure, process and transmit data to some base station. Operational costs of WDESNs are much lower than of AQM stations and they can be deployed in large numbers with far less operational requirements in various sites. These properties enable covering a larger area and obtaining a finer spatial and temporal resolution of measurements than the standard methodology. Nevertheless, the reliability of MSUs’ measurements is still questionable. Recent laboratory and field evaluations of MSUs show that these units are less accurate than standard laboratory equipment or AQM stations. However, it has been shown that they are able to effectively capture the spatial and temporal variability of the pollution [10], [11], [13]–[15].

## Air pollution representation in space

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 [16]. 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 [17]. Nevertheless, atmospheric transport and dispersion (ATD) models are the only models that can quantify the deterministic relationships between sources’ emissions and concentrations in space and in time. 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 [18]. ATD models can forecast the spread of the pollutants, when available results from monitoring data are used for calibration and evaluation of model performance. 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 [19].

## Source term estimation

For an accurate forecast, several variables are needed as input to ATD models, including, among others: meteorological data, the quantities of the pollutants emitted and the emissions’ locations. 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 quantity, time or even location of pollutants’ releases are frequently unknown. ﻿Pollutant release and transfer register (PRTR), an inventory of pollutants released to the environment, requires reporting of emissions from various industrial sectors to the regulator, on a yearly basis. However, their reliability is often questionable, not necessarily because of any bias, but due to inherent difficulties in the application of measurement techniques [20]–[22].

As a consequence, sensor measurements are often being used to determine source parameters in an inverse modelling problem of source-term estimation (STE) [23]. In these methods, source parameters serve as input to an ATD model and are modified until the difference between the calculated and observed measurements is minimal. Usually, an optimization technique is applied in order to search the solution space and reach an accurate evaluation. The estimation of source parameters can also serve as an objective of its own, rather than generating high resolution concentration maps; in the case of a chemical attack or an industrial leak, locating the position and the leak rate of the cause is the primarily step before taking any action.

## Sensor network deployment

When resources are unlimited, the challenge of representing the pollution field well or sufficiently estimating source parameters, becomes trivial. All that is required is the deployment of an extensive number of sensors to reach full cover of the area. However, as financial and human resources are usually limited, a smart and rigorous deployment of a network of sensors is needed, one which can provide the best performance for its designated application with the lowest possible cost. The cost of a network is usually determined by the number of sensors and their attributes, such as sensitivity and dynamic range. Placement locations need to be considered as well, as the deployment of sensors close to a leak may be costly due to additional operational costs to cope with the expected harsh environment. On the other hand, it can enable using less-sensitive sensors, which are usually cheaper. It is also noted that in many industrial sites sensors locations may be limited due to economic and practical considerations.

Sensor network redeployment…

The following proposal offers a multi-objective optimization model for the deployment of WDESN in time and space, aimed at finding a protocol for placing MSUs, under a set of fiscal and geographical constraints, so the best representation of the pollution field is obtained. The offered tool can serve stakeholders when either establishing a new network or planning a span of sensors during a routine sampling task for emission increase detection. It considers the time varying meteorological conditions (wind velocity and atmospheric stability) and uses as an objective a quantitative measure of the complexity of the given set of sources/sensors geometries, making it possible to evaluate "scene complexity", i.e. the potential to separate overlapping pollution plumes, once sensors are deployed at specific locations.

# Research objectives

The general objective of the proposed work is to find a set of optimal solutions for the deployment of a WDESN, for the purpose of obtaining early warning in the case of an increase in industrial gas emission or for the purpose of generating pollution dense maps. The set of solutions will be retrieved using multi-objective optimization techniques that balance between detection capabilities of the network of sensors and operational requirements (e.g. cost and physical field constraints). The multi-objective optimization will allow decision makers to discover tradeoffs between performance criteria and to consider alternative modes of action that may not have been apparent prior to the analysis. Later, the proposed scheme will be used to address the problem of a dynamic network, with the goal of finding the optimal redeployment in each time step that induces minimal transfer effort (…???). מינימום חיישנים שזזים לעומת מינימום מרחק להזיז. שני אובייקטים אלו מתנגשים.

אפשר גם מרחק בריבוע כך שיש עונש גדול יותר להזיז רחוק.

The specific objectives are:

* Design a *spatial* optimization model that computes sensors’ deployment, which minimizes the cost of the deployment while maximizing the sensitivity of the network to changes in the source term. The model will capture the distribution of weather conditions that characterizes the regime.
* Design a *temporal* optimization model that computes sensors’ redeployment, which minimizes the change (in what???) from the previous deployment. While maximizing the PED as before???
* Design a spatial-temporal optimization model. In this phase, the findings of the two previous objectives will be integrated, and probability of change in weather conditions will be considered, so it is most probable that future deployments require minimum changes.
* Conduct a set of simulations for each of the above-mentioned models to derive engineering insights for effective deployments of air pollution sensors at various leak rates, sources/sensors configurations and weather conditions.
* Examine the potential improvement achieved by the deployment of a heterogenous network, comprised of different types of sensors with different characteristics (sensitivity, dynamic range, cost), compared to a homogeneous network of sensors.

הטרוגניות של חיישנים (סקירה של אורי) -

Cross-talk – you use one sensor to deduct on other metirials.

המאמר הראשון לא מתאים למצבים משתנים – הוסיפו עוד מפעל...

המאמר השני בא ליצג מצבי חירום שנשמכים עד כמה ימים.

# Research contribution

﻿Technological developments in recent years have made Wireless Distributed Environmental Sensor Network (WDESNs) feasible to deploy, in a relatively low operational cost, using portable Micro Sensing Units (MSUs). These units can easily transmit data directly to cloud-based platforms, making it possible to immediately obtain the latest mapping of the pollution level in the environment. These advancements have produced new challenges in the field of environmental air pollution monitoring and modelling, such as selecting optimal sensors’ locations of placement.

This proposed work will provide engineering tools that facilitate effective sensor deployment schemes which successfully balance between detection and operational requirements. The work will allow decision makers to examine tradeoffs between performance criteria of the multi-objective optimization model and to consider alternative modes of action according to the prevailing circumstances.

This proposed work’s innovativeness stems from the implementation of a multi-objective optimization model, allowing a different perspective on the way operational problems may be approached. In addition, it considers the time-varying meteorological conditions that characterize the region of study, and thus produces more robust and reliable solutions.

MAYBE ADD AN EXAMPLE OF REAL WORLD CASES – planning that is stuck because of places constraints, devises that get broken or changes in the conditions that we calculated by first.

# Literature review

## The problem of sensors’ deployment

The problem of optimal deployment of any type of sensors consists of determining sensors’ positions, while ensuring the coverage and the connectivity of the network [24]. ﻿By optimizing coverage, the deployment strategy would guarantee that the area of the sensing field is optimally covered by sensors, as required by the underlying application. By ensuring that the network is connected, it is also ensured that the sensed information is transmitted to other nodes and possibly to a centralized base-station that can make some valuable decision [25]. ﻿Zhang and Liu [26] added to these challenges prolonging the network lifetime, balancing the load to save energy and improving the accuracy of the transmitted data. However, some argue that when dealing with a WDESN, challenges can be reduced to coverage area and data accuracy only, since most MSUs in use today are independent units which transmit data directly to a centralized computer [27], [28]. Hence, the following review regards works that dealt with optimizing the property of coverage as the main objective.

## The importance of network’s application

An air quality network of sensors is usually deployed for several designated applications. The two main ones include i) the management of a chemical leak due to an industrial accident, a disaster or an attack, that requires the detection of the resulted plume and the mapping the pollutant level in the environment [23] and ii) monitoring concentrations of pollutants emitted from routine activities. The different applications may form various goals that guide the optimization process aimed at finding the optimal sensors’ placement. For example, ﻿Kanaroglou et. al. [29] used population **exposure assessment** as an objective. In their method, they first generate a “demand surface”, which represents the spatial variability in pollution concentrations. Then, their “demand surface” is modified to also consider the density of a population of interest, such as children or elderly. Eventually, the “demand surface” is used as input to an algorithm that solves ﻿a constrained optimization problem ﻿and identifies the optimal locations for a predefined number of air pollution sensors. Their methodology however is limited to already monitored areas, since it relies on existing measurements from monitoring stations. Lerner et. al. [28] developed a method to optimally deploy a network of low-cost sensors of two types for the general purpose of **environmental monitoring** of ozone, NO and NO2 emissions. The optimal locations were found using an optimization process that seeks for the set of locations, constrained by available locations and a given budget, that maximizes the overall utility of the sensor network. The latter is comprised of the suitability of the type of sensor to the location of deployment and the rank of that location, relative to its surrounding. Lerner et. al. focused on the land use of the region of interest for specific atmospheric conditions.Carter and Ragade [30] proposed a probabilistic model which takes into account the detection probabilities of sensors of different types, assuming a decay with distance from a “target” or an “event” (i.e., pollution source). Their optimization procedure utilizes a genetic algorithm to ensure **a certain** **level of detection** of an event is reached by the distributed sensors, while minimizing costs. In air pollution monitoring, unlike in many other fields, pollutants detected by air quality sensors are governed by a time varying wind field. Simply assuming an exponential decay of the detection probability with distance therefore lacks physical base. Boubrima et. al. [31] accounted for the various weather conditions characterizing their region of study, and designed a model for the deployment of sensors with the objective of **detecting threshold crossings** in order to trigger an adequate alert. They used a Gaussian ﻿dispersion model and formulated an integer linear programming problem to minimize operational cost while assuring the detection of a threshold crossing. In their model however, sensors are eventually placed at the highest mean pollution zones, as calculated by assuming constant emission rates from sources.

In another work, Boubrima et. al. [32] simulated the deployment of a network with the goal of achieving the most effective **data assimilation** of air pollution measurements for the correction of physical model simulations. Similarly, Berman et. al. [33] wished to **improve the performance** of an interpolation-based model, using a geostatistical simulation. The researchers used an associated measure of the kriging interpolation method, which considers the uncertainty of the prediction. In the first step, they assessed how many sensors to add to a given network, using a Monte Carlo approach which evaluates how additional monitors change prediction precision through minimized uncertainty. In the second step, they assessed where to place the new monitors, using a similar Monte Carlo scheme which ﻿considers locations that improve prediction uncertainty and provide high prediction accuracy. Their method however requires a robust preliminary deployed network that already captures the spatial variability adequately, in order to be applied. In addition, their implementation did not provide a simultaneous assessment of number of sensors and their placement, due to the great affect it had on computation time.

## Sensors’ redeployment

Another aspect of network deployment dealt with in many works, is the use of a network of sensors mounted on mobile platforms [23], which enables a real-time adaptive deployment of sensors according to the spreading of the pollutant. Kuroki et al. [34] developed an expert system for navigating unmanned aerial vehicles for optimal contamination mapping that enables estimation of source parameters. Belkhiri et. al. [35] proposed several deployment models, that simulates air pollution concentrations as ground truth and estimates concentrations where no sensors are deployed using IDW interpolation. Sensors are redeployed so estimation error (i.e., ﻿the absolute difference between ground truth and estimation) is minimized and execution time is considered.

## Optimization aspect

Another aspect to be considered is the common methods used to solve the problem of network deployment. As the basic problem of sensors’ optimal deployment aims at maximizing the utility of the network, while minimizing its cost, the problem is equivalent to the “0-1 knapsack” problem [36]. In the “0-1 knapsack” problem, a subset of items, out of n items, possessing each some value and some cost , should be selected such that the sum of the values is maximized, while keeping the summed cost within some capacity . The knapsack problem is NP-complete [37], meaning that the time required to solve the problem using any currently known algorithm increases rapidly as the size of the problem grows. (The validity of a solution for a NP-complete problem, however, can be tested in polynomial time). As a result, approximation algorithms, which focuses on finding good solutions (“global optimums”) instead of provably optimal solutions, are required beyond a certain size of problem.

The term global optimization refers to the process of attempting to find the solution out of a set of possible solutions S that has the optimal value for some fitness function f, such that [38]. Two main categories of methods to find exist; the first are **deterministic methods**, which find by an exhaustive search over the set S while making certain assumptions about the fitness function to avoid huge calculations. **Stochastic methods** involve random elements to determine the global optimum point, each one trying to reduce the computational burden of pure random search. ﻿At the outset, a random sample of points in the set S is picked. Then, each method manipulates the sample points in a different manner, using different **heuristics** [39]**. Heuristics** may be thought of as sets of rules for deciding which potential solution out of S, should next be generated and tested (i.e., an intelligent search in space). For some randomized heuristics, such as simulated annealing and certain variants of Evolutionary Algorithms (EAs), convergence proofs exist. The problem in these algorithms however is that they will not identify the suggested solution as being globally optimal, rather as simply the best solution seen so far. **Local search algorithms** (often referred to as **hill climbers**), such as “gradient descent”, work by taking a starting solution x, and then searching the candidate solutions in the ﻿neighboring environment for one x’ that performs better than x. Although they may be quick to identify a good solution, this process will eventually lead to the identification of a local optimum, and no guarantee can be offered for the quality of the solution found, compared to . As a results, local searches are usually incorporated in stochastic methods to yield candidate global optimum, from which the best point is eventually picked.

EAs are Metaheuristic (i.e., problem-independent) optimization algorithms. Inspired by the biological theory, if given a population of individuals (i.e., a set of candidate solutions), the environmental pressure causes natural selection and according to a fitness measure (i.e., an objective function), the better candidates have a higher chance to survive and reproduce (i.e., to stay in the set of candidate solutions and generate new solutions by variation operators such as crossover and mutation). Crossover of two or more selected parents (i.e., selected solutions) may result in one or more offsprings (i.e., new solutions) that based on their fitness will compete with the old candidates for a place in the next generation [38].

The ability of EAs to maintain a diverse set of solutions, by creating new solutions from a non-uniform distribution, not only provides a means of escaping from one local optimum [23] and handling poor initial estimates; it provides a means of coping with large and discontinuous search spaces. As a consequence, EAs were shown to provide near-optimal results in many studies (e.g., [30], [40]).

This proposed work offers new approaches to real-world air pollution problems, by combining environmental aspects with state-of-the-art optimization technique, the Borg Multi-Objective Evolutionary Algorithm (MOEA). Using the proposed methodology would allow decision makers to retrieve sets of optimal solutions, each suited for different use-cases.

# Methods and Research plan

## Background

The proposed methodology requires the use of an atmospheric transport and dispersion (ATD) model in order to connect sources’ emissions and predicted concentrations in all potential locations to place the sensors at. 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 [23]. High resolution modeling tools for complex urban environments are available for most model types, such as the Gaussian plume AERMOD and the Eulerian grid-based CMAQ, both developed by the US Environmental Protection Agency (EPA). Other commonly used tools are the HYSPLIT and GRAL models, which simulate ﻿pollutant emission by releasing a fixed number of Lagrangian particles from the source [19], [41], [42]. For the sake of simplicity, we will first use the Gaussian plume model. Nevertheless, our proposed methodology is invariant to the selected ATD model, and can easily be replaced in the future, as all computations of the various weather conditions and source combinations are computed in advance. Once our method is proved to be effective, a more complex ATD model will be considered.

### The Gaussian plume model

The Gaussian plume model is one of the simplest and widely used models that offers an analytical steady state 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. The Gaussian plume model (Eq. 1) eventually describes the pollutants’ concentration C [kg/m3] in a certain position in space:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

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. Several simplifying assumptions are made in order to reach the Gaussian plume model equation (Eq. 1) [18]:

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 eddy diffusion coefficient K [m2/2] 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 non-reactive in the atmosphere, with a negligible deposition rate and that background pollution is negligible as well. parameters are used instead of the diffusion 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 (cite): and . 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 [43], using Pasquill-Gifford stability class categories [44]. These coefficients are specified in [Table X](http://homepages.see.leeds.ac.uk/~lecrrb/dispersion/index5.html).

### Meteorology

As reviewed above, the meteorological parameters that play a role in any gas dispersion 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 processes such as solar heating, radiative cooling or winds (cold or warm advection). In general, three regimes of atmospheric stability exist: **unstable, neutral and stable**. In an unstable atmosphere, 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 due to **thermal turbulence**. At a **stable atmosphere,** an air parcel will resist an upward vertical motion and will tend to spread out horizontally. In stable conditions, the reduction of vertical exchange reduces pollutant mixing, and may even damp out some of the **mechanical turbulence**, caused by the friction with Earth’s surface. Between the above two conditions is the condition characterized as **neutral**, when temperatures decrease slightly with height, in a rate close to the dry adiabatic rate.

Typical diurnal changes in the stability of the lower atmosphere exist. During the night, especially when winds are light and skies are clear, the radiative cooling of the ground surface often leads to surface air that is colder than the air above it. A stable layer thus exists in the lower hundred or so meter in the atmosphere. Pollutants emitted during the night inside this shallow layer get trapped and can reach relatively high concentrations. As the sun rises, the ground and the air next to it start warming up and the temperature profile corresponding to an unstable atmosphere is established. This change occurs over a period of a few hours in the morning and results in breaking of the inversion usually before noon [1], [43], [45].

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 [46], requires knowledge on cloud cover and cloud ceiling. Alternative empirical methods were developed for situations where these data are not available [47]. For the proposed work, we chose to use a turbulence-based method ( method), because of its simplicity and our currently available data (section X). ﻿The method 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).

### The Borg Multi-Objective Evolutionary algorithm (MOEA) framework

Unlike single objective optimization, which employs a single criterion for identifying the best solution among a set of alternatives, multi-objective optimization employs two or more criteria, i.e., two or more objectives. As multiple objectives can conflict with one another such that improving one objective leads to the deterioration of another - there is no single optimal solution to multi-objective problems. The tradeoffs in a Multi-objective problem are captured by solutions which are superior in some objectives but inferior in others. Such pairs of solutions which are both superior and inferior with respect to certain objectives are called non-dominated. The set of all nondominated solutions in a given generation is referred to as the Pareto frontor tradeoff surface (cite?)*.*

For the proposed work, the self-adaptive Borg MOEA will be used.14 The Borg MOEA is classified as a hyper-heuristic global13 multi-objective search tool as it uses internal feedbacks in search progress to dynamically adapt an ensemble search operators (or heuristic), rewarding those that maximize search progress (i.e., the dominance of new generated solutions). To this end, the Borg utilizes ε-dominance,[[1]](#endnote-1) ε-progress,[[2]](#endnote-2) randomized restart[[3]](#endnote-3),[[4]](#endnote-4) and auto-adaptive multi-operator recombination[[5]](#endnote-5),[[6]](#endnote-6) into a unified framework that has been shown to have a proof of convergence.[[7]](#endnote-7)-[[8]](#endnote-8)[[9]](#endnote-9) The approximate set, , in each iteration, is refined through the following internally competing genetic mating and mutation operators: Simulated Binary (SBX)[[10]](#endnote-10), Patent-Centric (PCX)[[11]](#endnote-11), Unimodal Normal Distribution (UNDX)24, and Simplex (SPX)[[12]](#endnote-12) crossovers, as well as Differential Evolution (DE)[[13]](#endnote-13), and Uniform Mutation (UM)[[14]](#endnote-14). The Borg MOEA auto-adapts the probability of which genetic operators to use, according to the operators’ offspring’s success rate in previous iterations. To avoid a convergence of the algorithm to a local minimum, a stochastic restart mechanism exploiting Uniform Mutation is built into the BORG MOEA to automatically detect and avoid pre-mature convergence to a local optimum and to achieve a diverse set of solutions.14

## Methodology – problem formulation and optimization

### General formulation

Let be the region of interest, i.e., the industrial area and its surroundings, where we wish to place our network of sensors at. 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:

|  |  |
| --- | --- |
|  | (2) |

For multiple sources scenario, each sensor readings are simply the sum of all sources’ contributions, which are the sources’ emissions multiplied by their corresponding transfer function, , i.e.:

|  |  |
| --- | --- |
|  | (3) |

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, such as the Gaussian plume model.

### Pairwise Euclidean distance (PED)

With the objective of finding a possible increase in industrial gas emissions, a criterion that incorporates the sensitivity of the deployed network to changes in the source term needs to be defined. We propose a new criterion for evaluating how well a hypothetical network of sensors, placed in hypothetical locations, would respond to slight changes in the source term. The criterion is the pairwise Euclidean distance (PED) between calculated readings of two sensor network sets with different number of active sources. That PED value of two sets of active sources, as measured with the same WDESN having different {S}' and {S}", is given by:

|  |  |
| --- | --- |
|  | (4) |

A network set that is sensitive to changes in the source term will produce high PED values. On the other hand, low PED values will characterize a non-sensitive network of sensors and might indicate on low ability of such network to resolve the specific source term estimation problem.

### Problem formulation

In our problem, we aim at finding an optimal deployment (number and locations) 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 Eq. 4). Given that the span of a set of 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 generate the highest PED value, which represents the best ability of those sensors to detect changes in industrial gas emissions. The first objective is then to find the minimal number of sensors deployed:

|  |  |
| --- | --- |
|  | (5) |

It is possible to phrase several different objective functions that maximizes the PED value. Our implementation 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 number of 355 possible combinations (NPC) of two sets of number of active sources with different sizes {S}' and {S}". The second objective is::

|  |  |
| --- | --- |
|  | (6) |

The network of sensors deployed should consider the varying meteorological conditions. 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 states exist, out of them only X were found to be physically meaningful according to the data archive used (see section X).

All possible situations (X weather states with all options of active/non-active sources) are applied in the Gaussian plume model to obtain steady state solutions of a pollution concentration map. Each concentration map is then given a weight according to the frequency distribution derived.

(other objectives).

### Tri-objective problem

The second objective 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 forms a tri-objective optimization problem and requires creating a separated frequency distribution of weather states for day and for night, in the same way described above.

### Optimization procedure

Once all weighted maps of concentrations are obtained, 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 possible PEDvalues of that set (Eq. 9). Then, based on the two or three 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.

## Meteorological data

For our simulation set, we used 10 min wind data obtained from the Israel Meteorological Service (IMS) for the years 2004-2018 from Sde Boker station. This station was chosen due to its proximity to a relatively idealized location: a desert area with little land cover, low topography and very few pollution sources except for Ramat-Hovav industrial zone. That way, we reserve the option to implement our methodology in the future on a real-world problem, possibly in that area.

# Research plan

* **Evaluation and Validation** – using real-world data sets or generate synthetic data of emissions.
* **Improve the modelling tool** – use as a constraint or as an objective, the demand for solving the STE problem at least for at least X percent of the scenarios and of maximum deviations from emission permits.
* **Create more complex scenarios** – allow different types of sensors with different sensitivities and dynamic ranges. possibly consider different pollution sources.
* **Conduct measurements**
* **Chicago database** - Evaluate how well the current placement of sensors is, suggest where to place additional sensors.
* **Mobile sensors and online deployment**

# Initial results (TBD)

### Simulation set

An initial simulation set was tested. We configured a 1000x1000 meter flat area. The simulated sensors were assumed to be situated at ground level (i.e., z=0 in Eq. 1). Five point-sources (stacks) were given average yearly emission rates of 1000, 1500, 600, 1900 and 300 . 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 (+). A total of X optional locations are spread in a grid, 50 m apart. Stack heights (effective heights) were taken at 10 m.

# Work schedule (TBD)

# References

[1] J. H. Seinfeld and S. N. Pandis, *Atmospheric chemistry and physics - from air pollution to climate change*, Second edi. Wiley-Interscience, 2006.

[2] D. J. Jacob and D. A. Winner, “Effect of climate change on air quality - citation,” *Atmos. Environ.*, vol. 43, no. 1, pp. 51–63, 2009.

[3] “Greenhouse Gas Emissions,” *EPA, United States Environmental Protection Agency*, 2019. [Online]. Available: https://www.epa.gov/ghgemissions/overview-greenhouse-gases. [Accessed: 23-May-2019].

[4] “Israel ministry of environmental protection,” *The Impact of Air Pollution from Industry*, 2015. [Online]. Available: http://www.sviva.gov.il/English/env\_topics/Health-and-Environment/Health-Impact-of-Environmental-Nuisances/Pages/The-Impact-of-Air-Pollution-from-Industry.aspx. [Accessed: 23-May-2018].

[5] WHO, “Ambient air pollution: Health impacts,” 2019. [Online]. Available: https://www.who.int/airpollution/ambient/health-impacts/en/. [Accessed: 10-Mar-2019].

[6] IPCC, *Climate Change 2013*, vol. 5. 2014.

[7] A. M. Fiore, V. Naik, and E. M. Leibensperger, “Air quality and climate connections,” *J. Air Waste Manag. Assoc.*, vol. 65, no. 6, pp. 645–685, 2015.

[8] M. Jerrett *et al.*, “A review and evaluation of intraurban air pollution exposure models,” *J. Expo. Anal. Environ. Epidemiol.*, vol. 15, no. 2, pp. 185–204, 2005.

[9] F. Kizel *et al.*, “Node-to-node field calibration of wireless distributed air pollution sensor network,” *Environ. Pollut.*, vol. 233, pp. 900–909, 2018.

[10] 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, pp. 8–16, 2015.

[11] 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, pp. 537–547, 2015.

[12] A. Marjovi, A. Arfire, and A. Martinoli, “High Resolution Air Pollution Maps in Urban Environments Using Mobile Sensor Networks,” *2015 Int. Conf. Distrib. Comput. Sens. Syst.*, pp. 11–20, 2015.

[13] M. I. Mead *et al.*, “The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks,” *Atmos. Environ.*, vol. 70, pp. 186–203, 2013.

[14] R. Piedrahita *et al.*, “The next generation of low-cost personal air quality sensors for quantitative exposure monitoring,” *Atmos. Meas. Tech.*, vol. 7, no. 10, pp. 3325–3336, 2014.

[15] N. Castell *et al.*, “Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?,” *Environ. Int.*, vol. 99, pp. 293–302, 2017.

[16] A. Nebenzal and B. Fishbain, *Hough-Transform-Based Interpolation Scheme for Generating Accurate Dense Spatial Maps of Air Pollutants from Sparse Sensin*, vol. 507. Springer International Publishing, 2017.

[17] P. H. Ryan and G. K. LeMasters, “A Review of Land-use Regression Models for Characterizing Intraurban Air Pollution Exposure,” *Inhal. Toxicol.*, vol. 19, no. sup1, pp. 127–133, Jan. 2007.

[18] J. M. Stockie, “The Mathematics of Atmospheric Dispersion Modeling,” *SIAM Rev.*, vol. 53, no. 2, pp. 349–372, 2011.

[19] A. F. Stein, V. Isakov, J. Godowitch, and R. R. Draxler, “A hybrid modeling approach to resolve pollutant concentrations in an urban area,” *Atmos. Environ.*, vol. 41, no. 40, pp. 9410–9426, 2007.

[20] R. L. Burritt and C. Saka, “Quality of physical environmental management accounting information, Lessons from Pollutant Release and Transfer Registers,” *Sustain. Account. Report.*, pp. 373–407, 2006.

[21] D. Kerret and G. M. Gray, “What do we learn from emissions reporting? Analytical considerations and comparison of pollutant release and transfer registers in the United States, Canada, England, and Australia,” *Risk Anal.*, vol. 27, no. 1, pp. 203–223, 2007.

[22] R. Sullivan and A. Gouldson, “Pollutant release and transfer registers: Examining the value of government-led reporting on corporate environmental performance,” *Corp. Soc. Responsib. Environ. Manag.*, vol. 14, no. 5, pp. 263–273, 2007.

[23] M. Hutchinson, H. Oh, and W. H. Chen, “A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors,” *Inf. Fusion*, vol. 36, pp. 130–148, 2017.

[24] B. Liu, O. Dousse, P. Nain, and D. Towsley, “Dynamic coverage of mobile sensor networks,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 2, pp. 301–311, 2013.

[25] Q. Zhao and A. Swami, “Coverage and Connectivity in Wireless Sensor Networks,” *Adapt. Cross Layer Des. Wirel. Networks*, pp. 301–323, 2010.

[26] H. Zhang and C. Liu, “A Review on Node Deployment of Wireless Sensor Network,” *IJCSI Int. J. Comput. Sci. Issues*, vol. 9, no. 6, pp. 378–383, 2012.

[27] B. Fishbain, U. Lerner, N. Castell, D. M. B. a Tom Cole-Hunter c, d, Olalekan Popoola e, and A. B. b Tania Martinez Iñiguez c, d, Mark Nieuwenhuijsen c, Milena Jovasevic-Stojanovic f, Dusan Topalovic f, g, Roderic L. Jones e, Karen S. Galea h, YaelEtzion a, FadiKizel a, Yaela N. Golumbic a, i, Ayelet Baram-Tsabari i, Tamar Yacobi a, Dana Drahler a, Johan, “An evaluation tool kit of air quality micro-sensing units,” *Sci. Total Environ.*, vol. 575, no. September 2016, pp. 639–648, 2017.

[28] U. Lerner, O. Hirshfeld, and B. Fishbain, “Optimal Deployment of a Heterogeneous Environmental Sensor Network,” *Jorunal Environ. Informatics*, no. X, pp. 1–9, 2018.

[29] P. S. Kanaroglou *et al.*, “Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach,” *Atmos. Environ.*, vol. 39, no. 13, pp. 2399–2409, 2005.

[30] B. Carter and R. Ragade, “A probabilistic model for the deployment of sensors,” *SAS 2009 - IEEE Sensors Appl. Symp. Proc.*, pp. 7–12, 2009.

[31] A. Boubrima, W. Bechkit, and H. Rivano, “Optimal WSN Deployment Models for Air Pollution Monitoring,” *IEEE Trans. Wirel. Commun.*, vol. 16, no. 5, pp. 2723–2735, 2017.

[32] A. Boubrima, W. Bechkit, H. Rivano, and L. Soulhac, “Leveraging the potential of WSN for an efficient correction of air pollution fine-grained simulations,” *Proc. - Int. Conf. Comput. Commun. Networks, ICCCN*, vol. 2018–July, 2018.

[33] J. D. Berman, L. Jin, M. L. Bell, and F. C. Curriero, “Developing a geostatistical simulation method to inform the quantity and placement of new monitors for a follow-up air sampling campaign,” *J. Expo. Sci. Environ. Epidemiol.*, vol. 29, no. 2, pp. 248–257, 2019.

[34] Y. Kuroki, G. S. Young, and S. E. Haupt, “UAV navigation by an expert system for contaminant mapping with a genetic algorithm,” *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4687–4697, 2010.

[35] A. Belkhiri, W. Bechkit, H. Rivano, and M. Koudil, “Context aware MWSN optimal redeployment strategies for air pollution timely monitoring,” *IEEE Int. Conf. Commun.*, vol. 2018–May, 2018.

[36] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*. Springer, 2004.

[37] R. M. Karp, “Reducibility among combinatorial problems,” in *Complexity of computer computations*, Springer, 1972, pp. 85–103.

[38] A. E. Eiben and J. E. Smith, *Introduction to evolutionary computing*, 2nd Editio., vol. 53. Springer, 2007.

[39] G. O. Concepts, *Introduction to Optimum Design, Chapter 16 -Global Optimization Concepts and Methods*, Fourth Edi. Elsevier Inc., 2017.

[40] R. Ramadan, H. El-Rewini, and K. Abdelghany, “Optimal and approximate approaches for deployment of heterogeneous sensing devices,” *Eurasip J. Wirel. Commun. Netw.*, vol. 2007, 2007.

[41] A. Berchet *et al.*, “A cost-effective method for simulating city-wide air flow and pollutant dispersion at building resolving scale,” *Atmos. Environ.*, vol. 158, pp. 181–196, 2017.

[42] A. Berchet, K. Zink, D. Oettl, J. Brunner, L. Emmenegger, and D. Brunner, “simulations over the city of Zürich , Switzerland,” vol. 2, no. 2, pp. 3441–3459, 2017.

[43] D. B. Turner, “Workbook of Atmospheric Dispersion Estimates; 2nd ed;” 1994.

[44] F. Pasquill, “The estimation of the dispersion of windborne material,” *Met. Mag.*, vol. 90, no. 1161, pp. 33–49, 1961.

[45] C. D. Ahrens and R. Henson, *Meteorology Today: An Introduction to Weather, Climate, and the Environment*, 12th ed. Cengage Learning, Inc, 2017.

[46] D. B. Turner, “A diffusion model for an urban area,” *J. Appl. Meteorol.*, vol. 3, no. 1, pp. 83–91, 1964.

[47] USEPA- United States Environmental Protection Agency, “Meteorological Monitoring Guidance for Regulatory Modeling Applications,” *Epa-454/R-99-005*, p. 171, 2000.

1. Reed, Patrick, Joshua B. Kollat, and V. K. Devireddy. "Using interactive archives in evolutionary multi-objective optimization: A case study for long-term groundwater monitoring design." Environmental Modelling & Software 22.5 (2007): 683-692.‏ [↑](#endnote-ref-1)
2. Iwema, Joost, et al. "Land surface model performance using cosmic-ray and point-scale soil moisture measurements for calibration." Hydrology and Earth System Sciences 21.6 (2017): 2843-2861.‏ [↑](#endnote-ref-2)
3. Martí, Rafael, Mauricio GC Resende, and Celso C. Ribeiro. "Multi-start methods for combinatorial optimization." European Journal of Operational Research 226.1 (2013): 1-8.‏ [↑](#endnote-ref-3)
4. White, Jeremy T. "A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions." Environmental Modelling & Software 109 (2018): 191-201.‏ [↑](#endnote-ref-4)
5. Gu, Fangqing, Hai-Lin Liu, and Kay Chen Tan. "A hybrid evolutionary multi-objective optimization algorithm with adaptive multi-fitness assignment." Soft Computing 19.11 (2015): 3249-3259. [↑](#endnote-ref-5)
6. Hadka, David, and Patrick Reed. "Large-scale parallelization of the Borg multi-objective evolutionary algorithm to enhance the management of complex environmental systems." Environmental Modelling & Software 69 (2015): 353-369.‏ [↑](#endnote-ref-6)
7. . Laumanns, M., L. Thiele, K. Deb, and E. Zitzler (2002), Combining convergence and diversity in evolutionary multiobjective optimization, *Evolutionary Computation*, *10*(3), 263-282. [↑](#endnote-ref-7)
8. . Hadka, D., and P. M. Reed (2013), Borg: An Auto-AdaptiveMany-Objective Evolutionary Computing Framework, *Evolutionary Computation*, *21*(2), 231-259. [↑](#endnote-ref-8)
9. . Rudolph, G., and A. Agapie (2000), Convergence properties of some multi-objective evolutionary algorithms, *Congress on Evolutionary Computation (CEC 2000)*, *2*, 1010-1016. [↑](#endnote-ref-9)
10. . Ercan, Mehmet B., and Jonathan L. Goodall. "Design and implementation of a general software library for using NSGA-II with SWAT for multi-objective model calibration." Environmental Modelling & Software 84 (2016): 112-120.‏ [↑](#endnote-ref-10)
11. . Bi, Weiwei, Graeme C. Dandy, and Holger R. Maier. "Use of domain knowledge to increase the convergence rate of evolutionary algorithms for optimizing the cost and resilience of water distribution systems." Journal of Water Resources Planning and Management 142.9 (2016): 04016027.‏ [↑](#endnote-ref-11)
12. . Jain, Ashu, and Sanaga Srinivasulu. "Development of effective and efficient rainfall‐runoff models using integration of deterministic, real‐coded genetic algorithms and artificial neural network techniques." Water Resources Research 40.4 (2004).‏ [↑](#endnote-ref-12)
13. . Andrews, F. T., Barry FW Croke, and Anthony J. Jakeman. "An open software environment for hydrological model assessment and development." Environmental Modelling & Software 26.10 (2011): 1171-1185.‏ [↑](#endnote-ref-13)
14. . Pelletier, Gregory J., Steven C. Chapra, and Hua Tao. "QUAL2Kw–A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration." Environmental Modelling & Software 21.3 (2006): 419-425.‏ [↑](#endnote-ref-14)