Extending Urban Air Quality Maps Beyond the Coverage of a Mobile Sensor Network: Data Sources, Methods, and Performance Evaluation

Ali Marjovi, Adrian Arfire and Alcherio Martinoli
Distributed Intelligent Systems and Algorithms Laboratory
School of Architecture, Civil and Environmental Engineering
École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
firstname.lastname@epfl.ch

Abstract

Targeting the problem of generating high-resolution air quality maps for cities, we leverage four different sources of data: (i) in-situ air quality measurements produced by our mobile sensor network deployed on public transportation vehicles, (ii) explanatory air-quality and meteorological variables obtained from two static monitoring stations, (iii) land-use data of the city, and (iv) traffic statistics. We propose two novel approaches for estimating the targeted pollutant level at desired time-location pairs, extending also to areas of the city that are beyond the coverage of our mobile sensor network. The first is a log-linear regression model which is built over a virtual dependency graph based on land-use data. The second is a deep learning framework that automatically captures the dependencies of the data based on autoencoders. We have evaluated the two proposed approaches against three canonical modeling techniques considering metrics of coefficient of determination (R^2), root mean square error (RMSE), and the fraction of predictions within a factor of two of observations (FAC2). Using more than 45 million real measurements in the models, the results show consistently superior performance in respect to the canonical techniques.

1 Introduction

Assessing the relationship between human health and long-term exposure to urban air pollutants is a major goal of many medical studies. This is a momentous problem considering the fact that statistically more than 7 million premature human deaths are annually linked to air pollution [39]. In the framework of numerous medical projects (e.g., CoLaus [32]), various health related parameters (e.g., blood pressure, renal salt excretion and physical activities) and location of a large number of participants are frequently collected. For physicians, one of the main missing pieces of the puzzle to accurately link specific medical disorders with air pollution is the availability of spatial high-resolution air quality maps of the urban areas in which participants reside and work. One way of generating such maps is to densely sense the very same air that people breathe. Unfortunately, this is not yet achievable with the current technology for a large number of participants.

In most countries, monitoring air quality is carried out through the use of air quality stations operated by national environmental protection agencies. These reference stations provide highly accurate measurements from a limited number of specially selected sites. In most cases, the distance between two stations is in the order of tens of hundreds of kilometers resulting in very low spatial resolution of the generated air quality maps. Although this low spatial resolution is restrictive, using data of such few static stations and through pilot experiments, the physicians have shown that short-term exposure to higher levels of air pollution (e.g., particulate matters) is associated with many health disorders such as higher night-time systolic blood pressure, reduced nocturnal blood pressure dipping, and a reduced ability of the kidney to excrete sodium during the day [36]. Further understanding of these assessments clearly requires highly-resolved air quality maps.

1.1 Mobile Sensing

As opposed to traditional air quality monitoring stations, the use of networks of low-cost sensors is quickly emerging, aiming at providing air quality data with unprecedented temporal and spatial resolution. Mobile sensors can dynamically move throughout the environment and provide high spatial resolution data. Innovative sensing strategies such as integrating air quality sensing nodes into wearable devices [21] and smart-phones [15] are proposed. These crowd-sourcing strategies may open exciting new opportunities for the study of urban air quality and its impact on health, if deployed in large numbers. However, most of such systems currently suffer from noisy measurements produced by low-cost sensors. Moreover, in many cases they are operated by inexpert users and the provided data is not fully trustworthy due to unknown sampling conditions and high uncertainty in the calibration and maintenance procedures.

Aiming at generating high-quality and high-resolution air quality maps, we use a heterogeneous system to gather data

from four sources: (i) a mobile network using moderate-cost sensors which is deployed and maintained in the framework of our project, OpenSense II[1], in Lausanne, Switzerland, (ii) two traditional static air quality and meteorological stations maintained by NABEL (the Swiss National Air Pollution Monitoring Network) and MeteoSwiss (the Swiss Federal Office of Meteorology and Climatology), respectively, (iii) land-use data (as the correlated measures for air quality) gathered and reported by the Swiss Federal Office of Topography (Swisstopo) and the Swiss Federal Office of Statistics (GeoStat), and (iv) traffic counts of the streets of the city provided by the Transitec company. For the first data category, we have anchored our sensing nodes on top of ten public buses in Lausanne [27, 4]. This innovative deployment has been later followed by some other researchers (e.g., Gao et al. [13]). This approach which adds mobility to monitoring platforms brings significant benefits in comparison to more traditional static Wireless Sensor Networks (WSNs): finer spatial resolution, coverage of a wider area (including urban, suburban, and even rural) with fewer nodes, cheaper maintenance, etc. However, not much literature exists on field estimation using non-stationary sensor networks. The movements of the nodes are not under our control and not even predictable since the buses are assigned to different lines every few hours depending on real-time needs of the public transportation company. The spatial coverage of the mobile sensor network in respect to the urban area of interest is in any case partial: the bus network is limited to a subset of the streets of the city, meaning that we never have any measurement for the streets/areas that are not covered by bus lines. Moreover, the coverage changes dynamically over time because of the dynamic assignment of buses mentioned above. Therefore at any given period of time there are many parts of the city that do not have any measurements. Given these conditions, generating consistent and as complete as possible maps with high resolution is a tough challenge. In this paper, we use the three other mentioned sources of information (i.e. measurements of static air-quality/meteorological stations, land-use, and traffic data) to generate extended maps through data-driven models.

1.2 Modeling and Estimation of Air Quality

To generate extended maps out of sparse measurements, we need to develop air quality models. 'Deterministic simulation' and 'statistical modeling' are the two main categories of works on air quality estimations [24]. Deterministic simulations use physico-chemical models of airborne gas dispersion and estimate the concentrations having the sources of emissions as input. GRAL [31] is an advanced example of this category which mathematically models the motion of pollution plume particles in the atmosphere using a Lagrangian dispersion model. Berchet et al. [7] have developed a GRAMM/GRAL simulation framework for Lausanne, Switzerland and achieved ten years of NO_x concentration maps with a spatial resolution of 5 m. One drawback of this category is that they require accurate data of emission inventories (e.g., type and number of vehicles in the streets), structural and geographical details of the environment (e.g., building dimensions), and meteorological data (e.g., temperature and wind speed), which are not always available in high temporal resolutions. Moreover, these simulations cannot adapt without manual intervention to real-world changes respective to their set-up parameters (e.g., topography or traffic flow modifications) and cannot capture particular pollution events (e.g., construction sites, fires, sudden changes in the weather conditions, etc.)

Statistical models treat air quality as a random distribution and try to derive an analytical description for it. These methods can be divided into two classes. The first class is represented by the purely field-driven models which aim at finding all the dependencies and variables from the measurements of the targeted pollutant. Spatial interpolation methodologies (e.g., inverse distance weighing interpolation [38], and K-Nearest Neighbor (KNN) [11]) are the most common approaches in this class. The performance of such methods drops drastically if the pollution distribution is dynamic and multi-variant [27] (which is usually the case for urban environments under short term observational conditions). The second class treat other modalities correlated to air quality (AKA explanatory variables, e.g., temperature) as random variables to derive a model for the target distribution. In other words, this class consists of statistical models which work not only based on the field measurements but also take one or more explanatory variables into account. These methods usually show higher performance compared to the purely field-driven models.

As concrete examples of the second class, Niska et al. [30] designed an architecture of a Multi-Layer Perceptron (MLP) Artificial Neural Network (ANN) model for forecasting concentrations of nitrogen dioxide at an urban station. They proposed a parallel genetic algorithm for selecting the input variables. Yi et al. [40] proposed another ANN structure for modeling ozone based on meteorological variables in an urban environment through a pattern recognition approach. Hussein et al. [23] developed a linear regression model on the data of a stationary monitoring node to predict aerosol particles. Mølgaard et al. [28] proposed a Bayesian regression model to predict Ultra-Fine Particles (UFPs) concentrations of an urban monitoring station using meteorology and traffic data as inputs. To obtain better prediction performance, Clifford et al. [9] proposed a generalized additive model using meteorological data, time, solar radiation and rainfall as explanatory variables. Reggente et al. [33] employed a Gaussian process regression to estimate UFPs in an urban air pollution monitoring network based on local and remote concentrations of NO_x, O_3, CO, and UFPs. Zheng et al. [41] inferred the real-time and fine-grained air quality information based on the historical and real-time data reported by static monitoring stations and a variety of explanatory data sources such as meteorology, traffic, human mobility, structure of road networks, and point of interests. They presented a semi-supervised learning approach on a framework that consists of two stacked models; an ANN-based model for capturing spatial, and a linear-chain conditional random field method for capturing temporal dependencies. None of the mentioned works have considered mobile sensor networks as source of in-situ measured air quality data.

[1] http://opensense.epfl.ch

One thrust of research has focused on modeling the air quality based on land-use data. Land-use features (in the context of environmental engineering) are measures of population density, building heights, heating type, terrain elevation, terrain slope, types of roads, and so on, and even sometimes average traffic volumes. Li et al. [26] proposed a Gaussian process regression (AKA Kriging) model using land-use characteristics to estimate urban UFP levels from measurements collected from trams within different grid-cells. The main problem with land-use data is that usually they are not available in high temporal resolutions, implying that, they are usually considered only as long-term representative for trends of air quality. One way to overcome this issue is to generate different models for every target time period. Hasenfratz et al. [17] and Li et al. [26] (in two separate works) built up two sets of models, each targeting one time period (e.g., one model per day) for one city. These models cannot be used for time periods other than those they have been trained for. In the method of Hasenfratz et al. [17], measurements gathered in a previous period are also used in the model to increase the accuracy of high resolution maps. In particular, they annotate the UFP measurements obtained during one year with the corresponding meteorological (e.g., temperature) and time data (e.g., weekday). Then based on the current meteorological conditions and time, they fetch the most relevant historic UFP measurements and use them to augment the current dataset represented by the real-time UFP measurements. This method significantly increases the accuracy of the maps, although the real-time meteorological data are not directly used in the model itself. On the other hand, Li et al. [26] did not consider meteorological parameters.

In our previous work [27], based on an analysis of the Pearson correlation of the targeted pollutant in various regions of the city, we built a graphical network that connects the regions (i.e., street segments) with highest statistical correlations. Using this network, we proposed a Probabilistic Graphical Model (PGM) to estimate the concentration of the target pollution for the nodes using Bayesian inference on the historical probability distribution of the nodes. Although promising results were reported, this method had a few drawbacks. First, PGMs require a huge amount of data to generate the probability distributions accurately (for instance, this is why the reported method could not build the model for a monthly time resolution). Second, marginalization of the probability distributions is a very heavy process. Third, that method is limited to the area covered by the sensors meaning that it cannot estimate the air quality outside the coverage of the mobile WSN. Finally, that work did not incorporate any land-use data in the model due to the fact that the land-use data usually does not have a temporal property and building probability distributions out of them is not trivial.

1.3 Our Contribution

To address the stated problem, this paper presents three contributions:

1. We propose to leverage four sources of data: (i) targeted air quality data provided by a mobile WSN, (ii) explanatory air quality and meteorology variables, (iii) land-use, and (iv) traffic data. These modalities together will enable us to generate air quality maps of the city that go beyond the actual sampling locations. There are only a few previous works which have integrated these types of data (e.g., [41]) to estimate the air quality. However, to the best of our knowledge none of them has ever dealt with the peculiarities of a mobile sensor network. The details of our mobile sensor network and the three other types of data are presented in Section 2. Similar to [25, 27] we discretize the area topologically based on the street segments in the city.

2. A network-based log-linear regression model is proposed that takes into account the measurements gathered by the mobile sensor network values, nine meteorological (e.g., temperature and precipitations) and gaseous (e.g., NO_2 and CO) explanatory variables obtained from the two static stations (explained in Section 2.2), eleven land-use modalities, and six traffic measures. This model is similar to the second model proposed in [27] with the following two improvements: (i) we generate the dependency network based on the similarities of street segments in terms of land-use and traffic data (while in [27] the network is built exclusively based on correlation of air quality data); (ii) we integrate the land-use data as another useful source of data in the log-linear regression model (while in [27], we did not use land-use data at all). Obviously log-linear models cannot capture the non-linearities of the data, however, we present how this vastly used class can be upgraded to predict the air quality even outside the coverage area of the mobile sensor network. Although these types of models are vastly studied [23, 27, 28, 33], the use of land-use data in building a virtual network that naturally captures the dependencies between street segments, the number of explanatory variables, the scale of the data, and the time resolutions which we consider in this paper are beyond the framework of many previous works in this area. Section 3.1 provides details of this contribution.

3. A deep learning framework is proposed in which an ANN is built while meaningful features of data are automatically captured. We found this framework very suitable for the stated problem since its predictions are not limited to the spatio-temporal domain used for the training process. By presenting this model, we show how an advanced machine learning technique can further improve the performance of air quality maps. To the best of our knowledge, none of the previous works in this area have designed a deep learning ANN to capture automatically the features of the four types of data. This powerful tool is explained in Section 3.2. Compared to the work of Zheng et al. [41] which proposes two stacked models; one for capturing spatial dependencies, and the other for temporal ones using a static sensor network, our model simultaneously learns the temporal and spatial dependencies using data gathered by a mobile sensor network.

2 System Overview and Data Sources

Our proposed heterogeneous system consists of four sources of information. This section presents the details of each one of them.

2.1 Mobile Sensor Network

Among the airborne pollutants (CO, SO_x, NO_x, NH_3, O_3, etc.), a growing attention is being placed towards studying Particulate Matter (PM) due to their significant adverse impact on human health. Most of the previous studies (e.g., [8]) have focused on PM_{10} or $PM_{2.5}$ which describe the fraction of particles smaller than 10 μm or 2.5 μm in a given volume, respectively. The most commonly used metrics for particulate matter are particle mass and particle count. However, the mass or number of particles are not necessarily the best measures for all impacts to human health. In fact, the surface area of the particles also matters. It is well-known that finer particles are potentially more harmful than coarse particles [34]. Studies have shown that measuring the surface of nano-particles, rather than their mass or number, is more meaningful for quantifying their health impact [5, 29]. UFPs are able to travel deeper into the lungs and, due to their large surface-to-volume ratio, have higher reactivity which can result in higher toxicity. Therefore, following our previous work presented in [27], we are interested in measuring and estimating the Lung-Deposited Surface Area (***LDSA***) which is a measure that describes the deposited surface of particles per volume of air inhaled.

2.1.1 Mobile sensing platforms

We have designed and developed dedicated sensing nodes anchored to ten public buses which measure multiple air quality parameters including LDSA. The measurements are geo- and time- stamped locally by the sampling node. The localization of the mobile nodes is achieved through fusion of GNSS and vehicle odometry data, and accurate timestamps are also obtained from the GNSS module. The measurements are sent through GPRS to a database server. Along with these, there are several meta-data information that are sent to the server to indicate the health state of the measurement network. The final deployment of our mobile sensor network started in October 2013, and is ongoing. The LDSA sensors are Naneos Partector [2], devices with negligible aging and drift which are calibrated by the manufacturer and therefore ready to use without further calibration efforts. The response time for this active sampling device is in the order of fractions of a second; multiple sub-second measurements are aggregated to result in a still relatively fast and reliable measurement output at 1 Hz frequency. The extension of the techniques presented in this paper to *gas-phase* pollutants (e.g., CO, NO_x) is beyond the scope of this work. However, this would be feasible once the issues of maintaining chemical sensor calibration [16] and mitigating mobility induced measurement distortions [3] due to the slow response of the chemical sensors are addressed. Figure 1 shows one of the sensing nodes used in this project.

2.1.2 LDSA data

More than 45 million geo- and time- stamped LDSA measurements gathered by our urban mobile sensor network during the year 2014 are used in this paper. The buses circulate only from 5 A.M. to mid-night approximately and the allocation of the buses to the transport lines is out of our control and driven by the transportation company needs. Therefore the distribution of LDSA measurements varies both

[2] http://www.naneos.ch

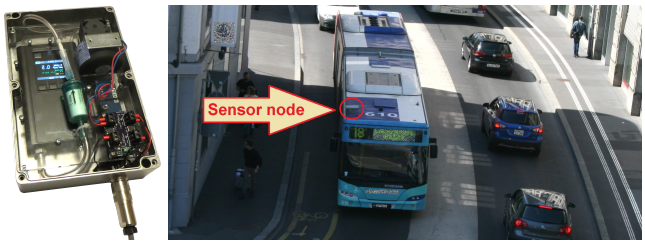

Figure 1. One of the LDSA sensing modules (left). One sensor node anchored on top of a public bus (right).

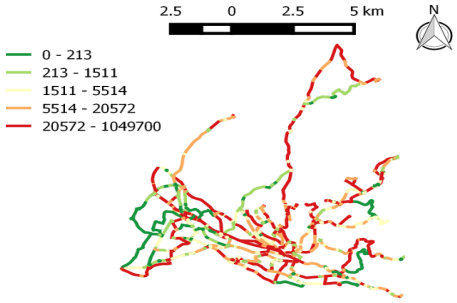

Figure 2. Number of LDSA measurements per street segment during one year.

in time and space. Streets of the city center and bus stops have a higher number of measurements as can be seen in Figure 2. Note that these are the total number of measurements for the whole year so, in practice, many of the streets will not have enough measurements to generate daily air quality maps. It is important to note that about 5% of the LDSA data provided by the mobile network were filtered out because of sensor malfunctioning using the meta-information that the sensors send alongside the measurements. The challenge is to produce a model of air quality over the whole city using these data that are still relatively sparse both in time and space.

2.1.3 Street segmentation

Most of the previous works (e.g., [17, 22]) divide the city into uniform grid cells, assuming that the measurements inside a cell have the same conditions in terms of weather, wind, traffic, and homogeneity of distribution of pollutants. Depending on the cell-size, one cell can cover several streets which have different environments and traffic conditions. To overcome this issue, Jutzeler et al. [25] proposed to use regions of homologous emissions in order to divide the city into partitions with similar daily traffic estimations. We showed that compared to grid-based, a partitioning based on the streets segments produces better predictions for pollutants distributions [27]. In this paper, we follow the same concept for space tessellation using the street segments of the city (acquired from the online OpenStreetMap [14] database). The use of this space discretization method will inherently result in higher resolutions in the downtown areas where street segments are shorter and the heterogeneity of the measured field is expected to also be higher than in suburban areas. In total we extracted 1669 street segments for the city under consideration.

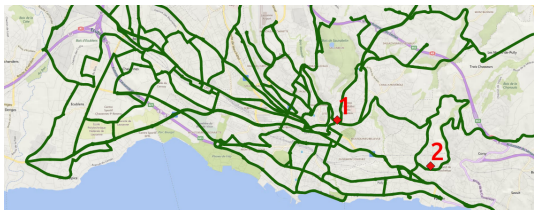

Figure 3. The red diamonds represent the location of the two stationary stations: air-quality measuring station (1) and the meteorological station (2). The green lines show partially the coverage of the bus network in Lausanne.

Table 1. The modalities measured by two static stations (left). The parameters of land-use data gathered from various agencies (right).

Parameter	Unit
CO	mg/m^3
NO	mg/m^3
O3	mg/m^3
NO_2	mg/m^3
Rain	mm
Radiation	W/m^2
Temperature	K
Relative Humidity	%
Wind speed	m/s

Parameter	Unit
Altitude	m
Slope	angle in °
Population	#/ha
Households	#/ha
# of buildings	#/ha
# of buildings using gas heating	#/ha
# of buildings using oil heating	#/ha
Industry	#/ha
Primary Industries	#/ha
Secondary Industries	#/ha
Tertiary Industries	#/ha

Table 2. The traffic data modalities.

Parameter	Symbol
Daily mean charge	TJM
Daily traffic volume mean during working days	TJOM
Hourly charge during morning rush hours	HPM
Hourly charge during night rush hours	HPS
Estimated effect of highways	HT

2.1.4 Back-end database

In order to manage the massive amount of data sent in real-time by the buses, we have adopted GSN (Global Sensor Networks) [2] for our back-end server, a middleware solution designed to handle the life-cycle of sensors. This includes sensor data acquisition, metadata management, storage, real-time processing, and publishing of measurement data.

2.2 Stationary Nodes

In addition to the LDSA measurements collected by the buses, we consider two static monitoring stations in our system (their location is shown in Figure 3). One is an air-quality measuring station which monitors many parameters (e.g., CO, NO, NO_2). The other is a monitoring station which provides meteorological parameters (e.g., precipitations, radiation and humidity). These two stations report their measured values (listed in in Table 1) every ten minutes.

2.3 Land-use Data

As mentioned already land-use features are measures of types of vegetation, population density, building heights, heating type, terrain elevation, etc. Since these are relevant variables to air quality, we have selected a number of modalities of interest for the city under study. It is important to note that we have extracted these data not only for the street segments that are covered by the mobile sensor network but also for the other parts of the city. This will later enable us to estimate the air quality for the segments that are beyond our sensor network coverage.

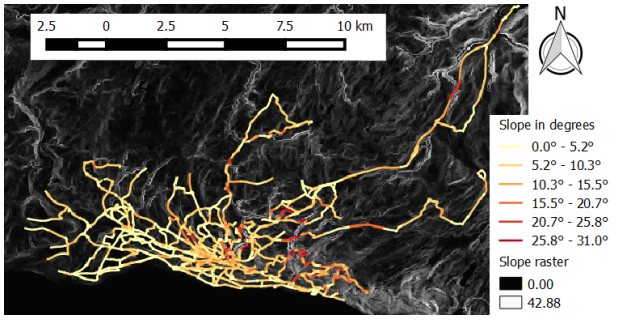

Figure 4. Slope profile of street segments of Lausanne.

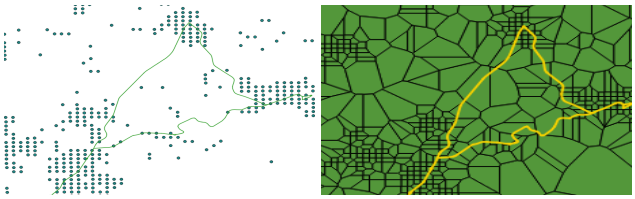

Figure 5. Data points of land-use data in hectare grid format (left) and its corresponding Thiessen polygon partitioning (right). Each point (in the left picture) represents one land-use data point for one hectare. Since the land-use grid data is sparse and non-uniformly distributed, the Thiessen polygon partitioning is used to assign the mean data to the streets segments that traverse a partition.

• **Altitude** and **slope:** These are potentially two important land-use parameters. We obtained a digital height library of the city that represents the 3D form of the earth's surface with spatial resolution of 25 m and average error of 1.5 m (provided by Swisstopo). Applying a digital elevation model (DEM) tool in a GIS software we extracted the mean altitude and the mean slope of the street segments of the city. Figure 4 is an example that shows the slope raster data (in the background) and the extracted corresponding data of the street segments.

• **Density of population households, buildings, industries:** GeoStat has provided a large variety of land-use and population statistics averaged in $100 \times 100\ m^2$ (i.e. hectare) grids. Number of habitants, buildings, buildings which use natural gas for heating, industry facilities, etc. are the main modalities of this dataset. Table 1 (right) gives a complete list of these modalities.

Translating the land-use statistics data (which is in hectare grid format) to our street segments spatial model was not a trivial task particularly due to the fact that the land-use data is sparse and not every point in the hectare grid has a value (see Figure 5, left). Moreover, street segments are in different lengths, so they traverse different number of grids and each would be corresponding to various land-use statistics values. To produce an adequate predictor based on these data, we firstly partitioned the whole area of the city using Thiessen polygons. This divided the area in polygons having the value of the closest point with a land-use statistical measurement

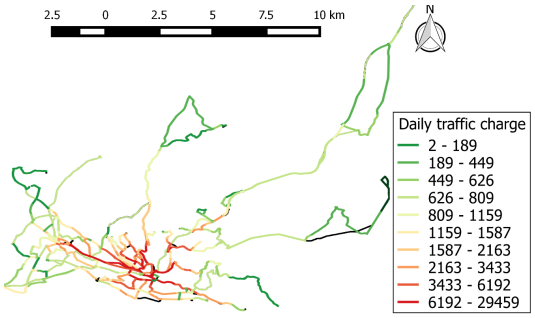

Figure 6. Average daily traffic charge (TJM) of street segments.

(shown in Figure 5, right). Then, the weighted average (with weights corresponding to length of street segment in each polygon) of the values of the polygons that were intersected by the street segments was calculated and assigned to each street segment to be used as a predictor.

2.4 Traffic Data

Many previous works have shown that traffic has a significant impact on urban air quality [18]. Various parameters of traffic counts for the streets are provided by Transitec in collaboration with the transport office of the city. We have extracted five traffic modalities for our street segments, listed in Table 2. Figure 6 is an example that presents the resulting allocation of one of these modalities (i.e. average daily traffic charge) over the street segments of the city. The first four modalities (in Table 2) have been directly provided by the datasets, while the last parameter was produced by us. Since the traffic volume of the highways are very likely to have an important effect on the LDSA measurements, a variable to model the "estimated effect of highways" was produced as follows:

$$HT_i = \begin{cases} \frac{T_H}{(\frac{d}{d_{min}})^2} & \text{if } d \geq d_{min} \\ T_H & \text{if } d < d_{min} \end{cases} \quad (1)$$

where T_H denotes the mean traffic charge in the closest highway and d is the distance of the street segment i to that highway. This predictor variable estimates the effect of highways by taking the traffic volume of the closest highway to each street segment divided by the normalized square of the distance to that highway. If the distance is less than d_{min} (a constant thresholding parameter, which is set to $d_{min} = 1\ m$ in this paper), we simply give the value of the traffic volume to this predictor.

It is important to mention that the data of all the above modalities have been pre-processed through statistical data cleansing to remove corrupted data and finally a unity-based normalization is performed.

3 Modeling Methods

Among the 1699 street segments of the city, there are many streets that do not have enough LDSA measurements, thus they are not useful for training the models. In this paper we consider 300 street segments, each having more than 20,000 measurements, to train and validate the models. The models then can be used in the other streets of the city for prediction of LDSA. In terms of time, we consider four resolutions: 'quarterly', 'monthly', 'weekly' and 'daily'. The LDSA measurements and the data of the explanatory variables were aggregated in intervals of the considered time resolution. For example, for the 'monthly' time resolution, the LDSA measurements of each month were averaged for each street segment, as well as all the explanatory variables (e.g., CO).

3.1 Log-linear Regression Model based on Land-use and Traffic Network

The goal is to estimate the LDSA values for the location-time pairs that the mobile sensor network has not covered. Many studies have shown that the mathematical links between many environmental parameters are logarithmic [17, 28]. We previously built a network-based log-linear regression model that fulfills this objective with one major limitation [27]: the measured LDSA values of other segments were taken into account for predicting the LDSA values for a given segment in a given time window. Then a virtual dependency network was built in which each segment was one node and a directed edge was drawn between node S_m and S_n if node S_n was considered as a variable in the model of node S_m. The metric used to generate this dependency network was the Pearson correlation between the LDSA measurements of the streets (i.e. two nodes are connected if their LDSA correlation is high). Therefore that model works only for the location-time pairs that have been covered at least once by the bus network, otherwise no correlation could be built. This means that the estimations could not go beyond the area actually covered by the mobile sensor network. Here we tackle this limitation by changing the way we build the dependency network and also by using land-use and traffic data. We refer to this approach as a Land-Use Network-based Log-Linear model ***(LU-NLL)***.

3.1.1 The model

The mathematical formulation of our model is defined by the following equation:

$$\log(L_{S_m}) = \alpha + \sum_{i=1}^{9} \beta_i \,.\, \log(v_i) + \sum_{i=1}^{16} \gamma_i \,.\, \log(U_{i,S_m}) + \sum_{[m-n]\in E} \delta_n \,.\, \log(L_{S_n}) \quad (2)$$

where L_{S_m} denotes the LDSA estimated value in segment S_m, α the intercept, v_i the explanatory variable i provided by the stationary nodes, U_{i,S_m} the land-use and traffic variables i in segment S_m (see Table 1 and Table 2 for the list of variables), E the edge list of the dependency network, and finally β, γ and δ the coefficients of each variable. Hereafter, the operator 'log' actually represents a 'shifted-log' transformation due to many zeros in environmental data (basically a positive number, in this case one, has been added to the data before log-transforming, i.e. $\log(1+v_i)$). The last term in (2) indicates that we have used the LDSA of street segment n as an input to estimate the LDSA of street segment m. At this point, the question is how to define the dependency network.

3.1.2 Virtual graph

Instead of defining the dependency network based on the LDSA correlations of the street segments, we define it based on the land-use and traffic data. This provides a significant advantage since the dependency graph (based on land-use and traffic data) no longer requires LDSA measurements and can be extended to all the streets of the city even beyond the coverage of the mobile sensor network. Of course this extension is limited to the regions that have environmental, geographic and structural properties similar to the coverage

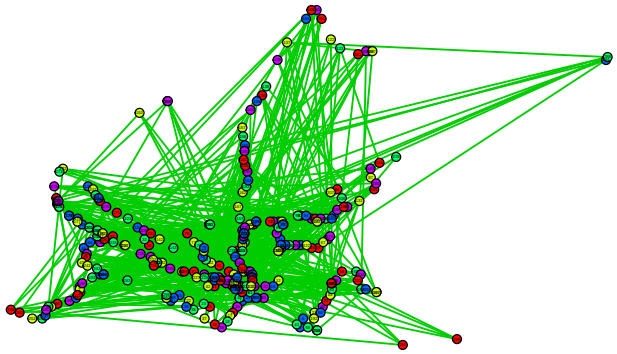

Figure 7. Virtual graph generated for the street segments of the city. Each node represents one street segment and the edges are drawn based on similarity in the land-use and traffic data. For increasing readability, not all the nodes and edges are shown in this figure.

area of the sensors. In the virtual graph, each street segment is one node and we connect the nodes with highest similarity in terms of their land-use and traffic data. Therefore the criteria for connectivity of nodes in the dependency graph is defined as follows:

$$D_L(S_m, S_n) = \sqrt{\sum_{i=1}^{16} \left(\widehat{\log}(U_{i,S_m}) - \widehat{\log}(U_{i,S_n})\right)^2} \quad (3)$$

where $\widehat{\log}$ denotes the normalized shifted-log transformation. $D_L(S_n, S_m)$ is the Euclidian land-use and traffic distance between two arbitrary street segments S_m and S_n. We calculate this distance for all combinations of street segments and then we connect each street to the one with the smallest distance. We keep connecting nodes with smallest land-use and traffic distance until all the graph becomes connected. Figure 7 shows partially such a graph generated for the city under study.

To the best of our knowledge no previous work has ever built a virtual network based on land-use and traffic data and integrated it into such a model.

3.1.3 Training the model

To train and evaluate the model, we randomly divide the available data into two subsets, the "training set" and the "validation set", using 10-fold cross validation. We use the QR decomposition algorithm [35] on the training set to solve the linear least squares problem in order to find the coefficients of the model.

There is no reliable ground truth data to evaluate the models in such systems. This is due to the fact that ground truth would be only acquirable with a dense deployment of high-end stations, a nonexistent solution given the resulting unaffordable cost. Therefore, similar to many previous works in environmental sensing, we use the validation data set (i.e. 10 % of the measurements that are separated from the training) as ground truth to assess the performance of the models. Working on four time resolutions we developed four models for the whole city (differently from [27] in which one model per street was generated). In fact, land-use and traffic data are used in this model as particular distinguishing features of

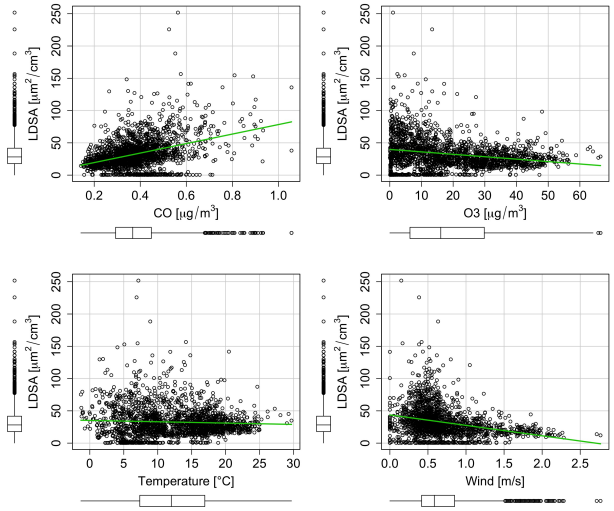

Figure 8. Correlation of four variables from static stations with LDSA concentrations in one street segment. The data aggregated by hour, for one street in the city that has a particularly large amount of bus measures.

the street segments and naturally categorize them into similar classes. The results are reported in Section 4.

3.2 Deep Learning Model (DLM)

The previously explained model inherently assumes that the LDSA distribution is a linear function of input variables. In fact, if the correlations between the input variables and the target variable (i.e. LDSA) are linear, then the regression model can estimate the target pollutant very well. However, studying the correlation between these variables shows that the dependencies are non-linear in most of the cases. Figure 8 shows the correlation between four exemplar variables with LDSA in one particular street segment. As can be seen on the plots, the data are quite scattered and linear correlations cannot always be observed. Some variables (e.g., CO) slightly correlate linearly with LDSA, while the others not at all. Therefore, we need a model that is able to capture the non-linearities and complex relationships between the variables.

ANNs are usually useful to automatically capture the dependencies and interpreting non-linear correlations between the input variables, without needing to know the underlying function. Theoretically, an ANN function is defined as a composition of other functions which can further be defined as a composition of other functions. This can be conveniently represented as a network structure, with arrows depicting the dependencies between functions and variables. The universal approximation theorem [20] states that a feed-forward network with a single hidden layer containing a finite number of neurons (AKA Multi-Layer Perceptron, MLP) can approximate any continuous function under mild assumptions.

Studying the history of multi-layer ANNs, their difficult optimization has mostly prevented obtaining expected benefits of going beyond one or two hidden layers [6], especially considering high-dimensional data. However, this situation has recently changed with the promising approach of Deep

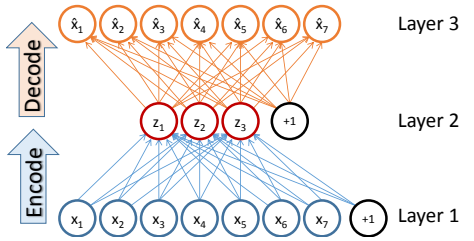

Figure 9. An exemplar autoencoder with three layers. The outputs ($\hat{x}_i$) are learned to predict the input variables (x_i). After training, z (layer 2) will be an encoded representation of x with lower dimension.

Neural Networks (DNNs) and stacked Auto-Encoders (AEs) [6, 19, 37]. Since ANN are effectively trained using a supervised learning approach, it has been repeatedly and experimentally proven that learning the coefficients gets difficult (if not impossible) when the number of input variables, number of layers, and number of neurons gets large. A key component of the DNN success is the procedure of partially pre-training the network using an unsupervised training criterion. The input data itself is used to partially train the neural network. This way, the network can learn intrinsic information about the data without the help of a target vector (i.e. the measured LDSA data in our case). In this process, the network automatically learns high-level features from the input data and it captures the non-linear dependencies between them. The learned information is stored as the initial weights of the network. Moreover, through this unsupervised learning process the dimensions of the input data is reduced while the amount of information in the data is not significantly lost. This is known as feature hierarchy, and it is a hierarchy of increasing complexity and abstraction. It makes deep-learning networks capable of handling very large, high-dimensional data sets (which is the case in our work) with billions of parameters that pass through nonlinear functions. AEs and restricted Boltzmann machines are two equally efficient methods to generate and pre-train DNNs. In this paper, we have used AEs to automatically capture the useful features of the data.

3.2.1 Automatic feature extraction using AEs

AEs can be considered as multi-layer sparse coding networks. The differences between AEs and MLPs are that (i) in an AE the output layer has the same number of nodes as the input layer, and (ii) instead of being trained to predict the target value given inputs, AEs are trained to reconstruct their own inputs. In other words, AEs try to learn an approximation to the identity function, such that the output $\hat{x}$ is similar to the input x. By placing constraints on the network, such as by limiting the number of hidden units (z), AEs can discover interesting structures about the data.

An AE first maps the scaled d-dimension input vector $x \in [-1,1]^d$ to a hidden d'-dimension representation $z \in [-1,1]^{d'}$ (where $d' < d$) through a deterministic mapping function:

$$z = f_\theta(x) = s(Wx+b), \tag{4}$$

parameterized by $\theta = \{W, b\}$, where s is the activation function (e.g., sigmoid, tanh, softmax), W is a $d' \times d$ weight matrix, and b is a bias vector. The resulting latent representation z is then mapped back to a reconstructed vector $y \in [-1,1]^d$ of the input vector, thus:

$$\hat{x} = g_{\theta'}(z) = s(W'z+b') \tag{5}$$

with $\theta' = \{W', b'\}$. The parameters of this model (θ and θ') are optimized to minimize the average reconstruction error:

$$\begin{aligned}(\theta^*, \theta'^*) &= \arg\min_{\theta,\theta'} \frac{1}{d}\sum_{i=1}^{d}(x_i - \hat{x}_i)^2 \\ &= \arg\min_{\theta,\theta'} \frac{1}{d}\sum_{i=1}^{d}\Big(x_i - g_{\theta'}\big(f_\theta(x_i)\big)\Big)^2\end{aligned} \tag{6}$$

Eq. (6) shows that θ^* and θ'^* are functions of only the input vector x, implying that a completely unsupervised back-propagation learning approach can learn optimized weights.

Figure 9 shows an example of an AE network with three layers. In this example, there are seven inputs x_i and there are 3 hidden units (z_i) in the middle layer, so the AE network is forced to learn a compressed (low-dimensional) intermediate representation of the input and then re-construct the 7 input data ($\hat{x}_i$) using z in an unsupervised manner. If there are structures in the data (e.g., if some of the input features are correlated), then the AE will be able to discover some of those correlations.

Once the weights of the AE network was learned (through an optimization method, in our work AdaGrad explained later in Section 3.2.3), the "decode" part of the AE network (colored with orange in Figure 9) will be eliminated and the middle layer (z) will be the input of the next layers of a feed forward DNN which will try to model the target modality. The obtained coefficients of the AE will be used as the initial values of the weights of the larger multi-layer DNN.

3.2.2 The structure of the deep network

As explained in Section 2, we propose to use four sources of information to estimate LDSA in various points of the city. Our proposed network structure uses AEs for each of these

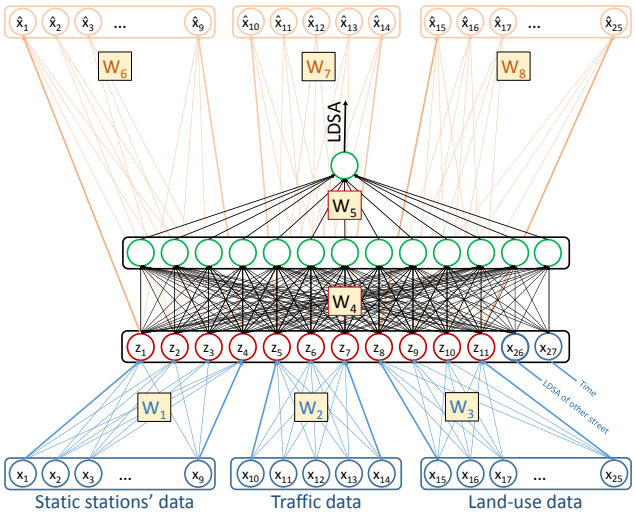

Figure 10. The proposed DNN structure. The blue nodes ($x_1 - x_{27}$) represent the input data, the red nodes ($z_1 - z_{11}$) represent the lower dimension features of the input data and the orange ($\hat{x}_1 - \hat{x}_{25}$) shows the decode part of the AEs.

sources to (i) extract informative features of the data, and (ii) lower the dimensions of the input data.

Figure 10 presents the proposed structure of the DNN. For increasing the readability we have not shown the usual 'bias' nodes in each layer. The details of this structure are as follows:

1. The nine modalities of the meteorological and air-quality measures obtained from the two static stations (listed in Table 1 (left) are coded into a feature vector with four nodes ($z_1 - z_4$).

2. The five measures related to traffic (listed in Table 2) are mapped into a feature vector with three nodes ($z_5 - z_7$).

3. The eleven measures related to land-use (listed in Table 1, right) are mapped into a feature vector with four nodes ($z_8 - z_{11}$).

4. In order to take into account the land-use network shown in Figure 7, we have inserted node x_{26} into this DNN. This node represents the LDSA measurement of the most similar street segment in terms of land-use data. Since this is only one node, there is no need to propose an AE for it.

5. The last modality taken into account is "time", represented by x_{27} in the DNN. The value of this node is dependent on the time resolution of the model. Since four time resolutions (daily, weekly, monthly and quarterly) have been considered, this nodes has four different implication (i.e. day number, week number, month number, and quarter number).

In each intermediate layer of the DNN, we use the following *nonlinear* neural transfer function to compute the output of the nodes:

$$h_{W,b}(x) = \tanh(\sum_{i=1}^{N} W_i x_i + b) \tag{7}$$

where x_i represents the input nodes, N is the number of inputs, while b and W_i are the bias and the weights to be optimized. The hyperbolic tangent that is used as activation function is suitable for our model due to the fact that all the inputs and the target modalities are scalar numbers that are normalized to $[-1,1]$.

3.2.3 Training and implementation

We use the same exact training set, validation set and land-use network explained in Section 3.1 for this model too. Using the training set, first the AEs of the DNN are trained (unsupervised), and optimized values of their weights ($W_1, W_2, W_3, W_6, W_7, W_8$) are obtained. Then the upper part of the AEs (colored orange) is ignored and the whole DNN is trained (supervised) starting with the partial initial weights' values obtained from the AEs.

For optimizing the weights and the biases, we use a recent adaptive sub-gradient optimization method called AdaGrad [12] which has shown remarkably good results on large scale multi-dimensional learning tasks in a distributed environment [10]. AdaGrad automatically scales the learning rate of each parameter in every iteration and tries to dampen the extreme parameter updates, while increasing the learning rate of parameters that have got few or small updates. The loss function used as the criteria of the optimization process was simply the root square error of the estimated values relative to the measured values.

We have implemented our models in the framework of Tensorflow [1] which is an open-source framework for numerical computations using data flow graphs, recently released by Google. This framework significantly facilitates the process of building and training the networks and provides useful optimization libraries.

4 Performance Evaluation

The proposed models are used to estimate LDSA values for locations/times of interest. In contrast to many previous studies (including our previous work in [27]), the models are able to predict air quality even outside the coverage of the sensor network.

4.1 Canonical Baseline Methods

To assess the performance of the proposed methods, we need to compare them with canonical baseline techniques. Most of the previous works compare their models with purely unsupervised data-driven statistical models (e.g., KNN) which usually show very poor results in real environmental datasets. We instead compare our proposed models with three supervised log-linear regression models which have already delivered acceptable results. In order to optimize their parameters, we have trained these models using the very same training set that is used for training the two new proposed methods. The three baseline methods are as follows.

• ***Basic log-linear regression model (BLL)***: As the first baseline model we use only the data of the static stations to predict the LDSA values. This is similar to many previous works including the first model presented in [27]. The mathematical formulation of this model is defined by:

$$\log(L_{S_m}) = \alpha + \sum_{i=1}^{9} \beta_i \,.\, \log(v_i) \tag{8}$$

Since the static stations do not provide spatial variety, we have to generate one model per street segment. This implies that this model cannot be further extended to street segments outside the coverage area of the mobile sensor network.

• ***Network-based log-linear regression (NLL)***: The second baseline method is similar to the first proposed method (see Section 3.1) but instead of building the virtual dependency network based on the similarity in the land-use data, we build the virtual network based on the Pearson correlation between the LDSA measurements of the streets. This is similar to the second method proposed in our previous work [27].

$$\log(L_{S_m}) = \alpha + \sum_{i=1}^{9} \beta_i \,.\, \log(v_i) + \sum_{[m-n]\in E'} \delta_n \,.\, \log(L_{S_n}) \tag{9}$$

Note that unlike Eq. (2), here E' is defined as the edges of the network generated based on the Pearson correlation of LDSA measurements. As already mentioned the estimations of this model cannot go beyond the area actually covered by the mobile sensor network.

• ***Basic log-linear regression with land-use (BLL-LU)***: As the third baseline model, we have integrated also the land-use and the traffic data into a log-linear regression model. Since the land-use and the traffic data inherently provide spatial variety, we can build one general model for the city and it can predict even for regions outside our sensor network coverage. The mathematical formulation of this model is defined by:

$$\log(L_{S_m}) = \alpha + \sum_{i=1}^{9} \beta_i \,.\, \log(v_i) + \sum_{i=1}^{16} \gamma_i \,.\, \log(U_{i,S_m}) \quad (10)$$

This basic model is similar to our first proposed method explained in Section 3.1 with the significant difference that it does not integrate in the log linear regression a virtual network built on land-use data (compare Eq. (10) and (2)).

4.2 Evaluation Metrics

We have considered three metrics for statistical evaluation of the methods. In the validation datasets, for every estimated value (model's prediction), there is an observed measurement. Denoting M as the set of modeled values and O as the set of corresponding observations, we consider these metrics:

1. R^2: The coefficient of determination shows how well the model fits the observed values. This metric is defined as:

$$R^2 = 1 - \frac{\sum_i (O_i - M_i)^2}{\sum_i (O_i - mean(O))^2} \quad (11)$$

where $mean(O)$ represents the mean of all observations which are considered in the validation sets. $R^2 = 1$ indicates a perfect linear fit between the model and the observations.

2. RMSE: The root mean square error is computed as the following:

$$RMSE = \sqrt{\frac{1}{L}\sum_i (O_i - M_i)^2} \quad (12)$$

where L is the number of estimations provided by the model. Obviously, the lower the RMSE is, the better the model works.

3. FAC2: The factor of two measure, is the percentage of ratios O_i/M_i that fall between 0.5 and 2. i.e.

$$0.5 < \frac{O_i}{M_i} < 2 \quad (13)$$

The more close to 1 this metric is, the better the model has estimated the observations.

4.3 Results

All five modeling methods have been tested using the same data set, with models being derived for daily, weekly, monthly, and, in the case of models that exploit land-use and traffic data, also quarterly time resolutions. The results of calculating the metrics for all the proposed models are shown in Figure 11. Although not directly comparable, some observations can be made on the difference between the attainable performance for methods that derive models for each street segment (i.e. local models that do not use spatial explanatory variables), and global models derived for the whole city that make use of land-use and traffic data. As can be expected, the BLL and NLL methods are able to deliver locally higher performance than single models for the whole city. This is because they generate one model per street segment, thus each local model is trained and validated on a small (and usually highly homogeneous) fraction of the data taken in one street segment. The variability of this performance is, however, large. Also, as the temporal aggregation is increased, these methods start having a harder time fitting their local models because of the reduced number of aggregated data points within the 12 months considered in each street segment. This starts affecting their performance at the monthly scale, while

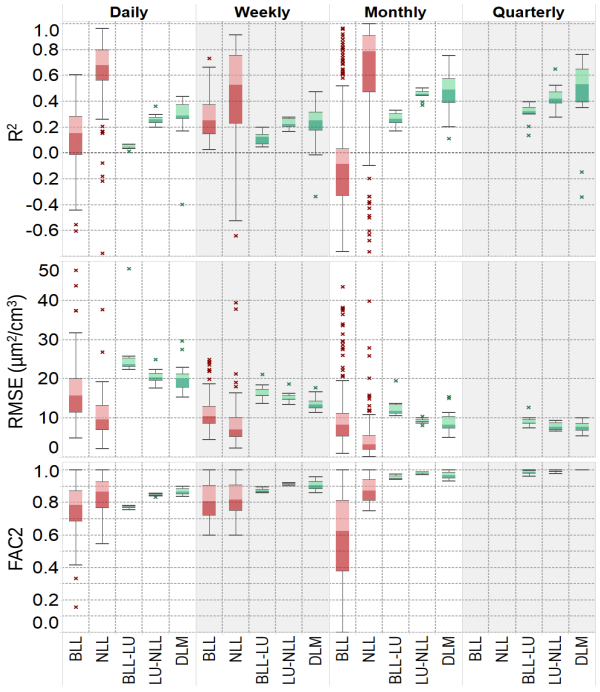

Figure 11. The evaluation results for the proposed modeling methods. The results of the methods that derive one model per street segment (BLL and NLL) are shown with red boxplots, while the results for the other modeling techniques which derive one model for the whole city are depicted with green boxplots.

making it impossible to fit local models at the quarterly levels (there are only four complete year-quarters in the considered interval). Anyhow, as already mentioned, the local models (BLL and NLL) cannot be extended to the areas beyond the coverage area of the sensor network.

A significant advantage of the three models that also exploit land-use and traffic data (BLL-LU, LU-NLL and DLM) is that one single model is generated for the whole city that can capture all the dependencies between all the street segments and also the dependencies between the LDSA values of each segment with the input variables. Basically, only a few models are built for the whole city (one model per time resolution) and they can generate high spatial resolution maps even for locations beyond the coverage area of the mobile sensor network where both land-use and traffic data are available. Figure 12 shows two daily air quality maps of the city generated by the proposed DLM model.

The quantitative performances of the three global methods (BLL-LU, LU-NLL and DLM) is more consistent, with smaller variations compared to the two local models (BLL and NLL) for the three metrics (see Figure 11). Although the NLL shows higher mean R^2 and FAC2 and lower mean RMSE for most of the time resolutions, its standard variation is very large. Considering the fact that NLL is a local modeling approach, this means that there are a considerable number of street segments that are associated with very poor models (and of course there are also many streets with really good models). On the other hand, the global models that use the land-use and traffic data, can always guarantee an

acceptable performance over the whole city. This consistency in the performance results is an important achievement that is required for generating global air quality maps with bounded errors for the whole city. This proves that exploiting the land-use together with the other three sources of data provides a significant advantage in this estimation/map building effort.

The two proposed methods provide very satisfactory results (even compared to the only global baseline method BLL-LU): mean R^2 always above 0.2, mean FAC2 always above 0.85 and mean RMSE always bellow 21 $\mu m^2/cm^3$. However, their performance depends also on the targeted time resolution. For the lower time resolutions (e.g., quarterly) the performance of these models is better (i.e. higher R^2 and FAC, while lower RMSE) than for the high time resolution (e.g., daily). This is due to the fact that lowering time resolution implies aggregating more data together and therefore damping high frequency fluctuations, which makes the signals smoother, thus easier to be modeled.

The results prove the impact of using a virtual network in this system. For both cases when a virtual dependency network is used (i.e. for the NLL and LU-NLL methods) this has a significant positive impact on performance when compared to the respective basic log-linear regression methods (i.e. BLL and BLL-LU). Differently from NLL, the proposed land-use virtual network also allows us to extend the maps beyond the coverage of the mobile WSN .

Finally, the second proposed method (DLM) outperforms the other two global models (BLL-LU and LU-NLL), delivering the best performance for the three metrics. However, this superiority comes with the price of higher complexity and higher computational cost, it takes about 20 hours to train the DLM on the daily resolution, while the LU-NLL takes less than five minutes to be trained on the same computer platform with the same dataset. Therefore, the suitability of the methods depends also on the application. For an online service that should model the air quality every few minutes the first proposed method is appropriate, while for an offline survey service that should accurately process the history of data and generate a more precise air quality map, the DLM method is suitable.

5 Conclusion

In this paper, we showed how to use a heterogeneous system, which consists of a mobile sensor network, two static monitoring stations and two other sources of land-use and traffic data, to generate extended high-resolution air quality maps of a city. We proposed two novel modeling approaches in order to be able to estimate the air quality for time-location pairs with no associated measurements. The first method was a log-linear regression model based on a virtual land-use network of the street segments of the city. The second method was a deep learning approach using autoencoder structures which automatically capture the dependencies of the input variables and extracts their informative features.

The two proposed models are universal and consistent in a sense that one single model for one time resolution can capture all the dependencies between all the street segments and also the dependencies between the LDSA values of each segment with the input variables. The resolution and the area coverage of the maps generated by the proposed models are beyond the ones of our mobile sensor network, and are extended to the resolution and coverage of the land-use and traffic data. The main limitation of the presented methods lies in the fact that the coverage of the output maps cannot be extended beyond the spatial information provided by both land-use and traffic data. While land-use data are typically available for the whole territory, traffic data, especially if gathered with canonical vehicle counting techniques, are more sparse in time and space and represent therefore the main bottleneck for further mapping extension.

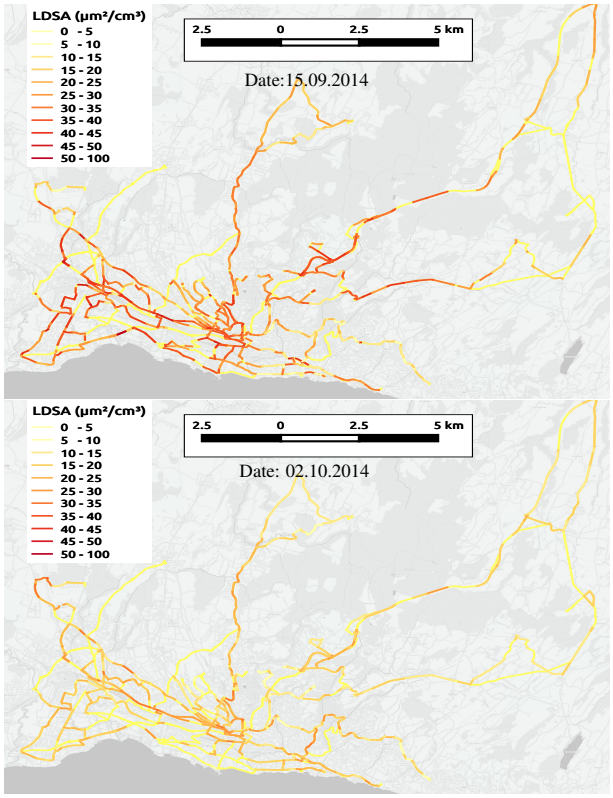

Figure 12. Two examples of LDSA daily maps of Lausanne, generated by the DLM model.

Comparing with three baseline methods, we conclude that both proposed methods show very good results considering three metrics of R^2, RMSE and FAC2. Nevertheless, the second proposed method based on Deep Neural Networks outperforms the other global models, implying that its increased complexity and computational cost is adding significant value in terms of estimation performance.

6 Acknowledgments

This work was funded by NanoTera.ch, a research initiative scientifically evaluated by the Swiss National Science Foundation and financed by the Swiss Confederation, in the framework of OpenSense II, a follow up on the original OpenSense project.

7 References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. Corrado, A. Davis, et al. Tensorflow: Large-scale machine learning

on heterogeneous systems. *White paper, Google Research*, 2015.

[2] K. Aberer, M. Hauswirth, and A. Salehi. A middleware for fast and flexible sensor network deployment. In *Proc. of the 32nd Int. Conf. on Very large data bases*, pages 1199–1202, 2006.

[3] A. Arfire, A. Marjovi, and A. Martinoli. Mitigating slow dynamics of low-cost chemical sensors for mobile air quality monitoring sensor networks. In *Int. Conf. on Embedded Wireless Systems and Networks*, pages 159–167, 2016.

[4] A. Arfire, A. Marjovi, and A. Martinoli. Model-based rendezvous calibration of mobile sensor networks for monitoring air quality. In *IEEE SENSORS*, DOI: 10.1109/ICSENS.2015.7370258, (4 pages), 2015.

[5] M. Auffan, J. Rose, J.-Y. Bottero, G. V. Lowry, J.-P. Jolivet, and M. R. Wiesner. Towards a definition of inorganic nanoparticles from an environmental, health and safety perspective. *Nature Nanotechnology*, 4(10):634–641, 2009.

[6] Y. Bengio. Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[7] A. Berchet, K. Zink, D. Brunner, and L. Emmenegger. Assessing spatial and temporal variability of air quality at the city scale using building-resolving dispersion modelling. In *Int. Workshop on Physical Modeling of Flow and Dispersion Phenomena*, Switzerland, 2015.

[8] H.-J. Chu, B. Huang, and C.-Y. Lin. Modeling the spatio-temporal heterogeneity in the PM10-PM2.5 relationship. *Atmospheric Environment*, 102:176–182, 2015.

[9] S. Clifford, S. L. Choy, T. Hussein, K. Mengersen, and L. Morawska. Using the generalised additive model to model the particle number count of ultrafine particles. *Atmospheric Environment*, 45(32):5934–5945, 2011.

[10] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, M. Ranzato, A. Senior, P. Tucker, K. Yang, Q. V. Le, and A. Y. Ng. Large scale distributed deep networks. In *Advances in Neural Information Processing Systems*, pages 1223–1231, 2012.

[11] E. G. Dragomir. Air quality index prediction using k-nearest neighbor technique. *Bulletin of PG University of Ploiesti, Series Mathematics, Informatics, Physics, LXII*, 1:103–108, 2010.

[12] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research*, 12:2121–2159, 2011.

[13] Y. Gao, W. Dong, K. Guo, X. Liu, Y. Chen, X. Liu, J. Bu, and C. Chen. Mosaic: A low-cost mobile sensing system for urban air quality monitoring. In *IEEE INFOCOM Int. Conf. on Computer Communications*, DOI: 10.1109/INFOCOM.2016.7524478, (9 pages), 2014.

[14] M. Haklay and P. Weber. OpenStreetMap: User-generated street maps. *Pervasive Computing,*, 7(4):12–18, 2008.

[15] D. Hasenfratz, T. Arn, I. de Concini, O. Saukh, and L. Thiele. Health-optimal routing in urban areas. In *Int. Conf. on Information Processing in Sensor Networks*, pages 398–399, New York, NY, USA, DOI: 10.1145/2737095.2737135, (2 pages), 2015.

[16] D. Hasenfratz, O. Saukh, and L. Thiele. On-the-fly calibration of low-cost gas sensors. In *European Conference on Wireless Sensor Networks*, pages 228–244, 2012.

[17] D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, T. Arn, J. Beutel, and L. Thiele. Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive and Mobile Computing*, 16:268–285, 2015.

[18] S. B. Henderson, B. Beckerman, M. Jerrett, and M. Brauer. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environmental Science & Technology*, 41(7):2422–2428, 2007.

[19] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[20] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

[21] K. Hu, T. Davison, A. Rahman, and V. Sivaraman. Air pollution exposure estimation and finding association with human activity using wearable sensor network. In *Proceedings of the ACM Workshop on Machine Learning for Sensory Data Analysis*, DOI: 10.1145/2689746.2689749, (48 pages), 2014.

[22] X. Hu, L. A. Waller, M. Z. Al-Hamdan, W. L. Crosson, M. G. Estes, S. M. Estes Jr., D. A. Quattrochi, J. A. Sarnat, and Y. Liu. Estimating ground-level PM 2.5 concentrations in the southeastern US using geographically weighted regression. *Environmental Research*, 121:1–10, 2013.

[23] T. Hussein, A. Karppinen, J. Kukkonen, J. Härkönen, P. P. Aalto, K. Hämeri, V.-M. Kerminen, and M. Kulmala. Meteorological dependence of size-fractionated number concentrations of urban aerosol particles. *Atmospheric Environment*, 40(8):1427–1440, 2006.

[24] A. Jakeman, R. Simpson, and J. Taylor. Modeling distributions of air pollutant concentrations-III. The hybrid deterministic-statistical distribution approach. *Atmospheric Environment*, 22(1):163–174, 1988.

[25] A. Jutzeler, J. J. Li, and B. Faltings. A region-based model for estimating urban air pollution. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 424–430, 2014.

[26] J. J. Li, A. Jutzeler, and B. Faltings. Estimating Urban Ultrafine Particle Distributions with Gaussian Process Models. *In S. Winter and C. Rizos (Eds.): Research@ Locate14*, pages 145–153, 2014.

[27] A. Marjovi, A. Arfire, and A. Martinoli. High resolution air pollution maps in urban environments using mobile sensor networks. In *IEEE Int. Conf. on Distributed Computing in Sensor Systems*, pages 11–20, 2015.

[28] B. Mølgaard, T. Hussein, J. Corander, and K. Hämeri. Forecasting size-fractionated particle number concentrations in the urban atmosphere. *Atmospheric Environment*, 46:155–163, 2012.

[29] A. Nel, T. Xia, L. Mädler, and N. Li. Toxic potential of materials at the nanolevel. *Science*, 311(5761):622–627, 2006.

[30] H. Niska, T. Hiltunen, A. Karppinen, J. Ruuskanen, and M. Kolehmainen. Evolving the neural network model for forecasting air pollution time series. *Engineering Applications of Artificial Intelligence*, 17(2):159–167, 2004.

[31] D. Oettl and U. Uhrner. Documentation of the Lagrangian Particle Model GRAL (Graz Lagrangian Model) Vs. 12.10. *Amt d. Stmk. Landesregierung, FA17C, Technische Umweltkontrolle, No: Lu-03-12*, 2012.

[32] B. Ponte, M. Pruijm, P. Marques-Vidal, P.-Y. Martin, M. Burnier, F. Paccaud, G. Waeber, P. Vollenweider, and M. Bochud. Determinants and burden of chronic kidney disease in the population-based CoLaus study: a cross-sectional analysis. *Nephrology Dialysis Transplantation*, 28(9):2329–2339, 2013.

[33] M. Reggente, J. Peters, J. Theunis, M. Van Poppel, M. Rademaker, P. Kumar, and B. De Baets. Prediction of ultrafine particle number concentrations in urban environments by means of Gaussian process regression based on measurements of oxides of nitrogen. *Environmental Modelling & Software*, 61:135–150, 2014.

[34] J. Schwartz and L. M. Neas. Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren. *Epidemiology*, 11(1):6–10, 2000.

[35] G. A. Seber and A. J. Lee. *Linear regression analysis*, volume 936. John Wiley & Sons, 2012.

[36] D.-H. Tsai, M. Riediker, G. Wuerzner, M. Maillard, P. Marques-Vidal, F. Paccaud, P. Vollenweider, M. Burnier, and M. Bochud. Short-term increase in particulate matter blunts nocturnal blood pressure dipping and daytime urinary sodium excretion. *Hypertension*, 60(4):1061–1069, 2012.

[37] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *The Journal of Machine Learning Research*, 11:3371–3408, 2010.

[38] D. W. Wong, L. Yuan, and S. A. Perlin. Comparison of spatial interpolation methods for the estimation of air quality data. *Journal of Exposure Science and Environmental Epidemiology*, 14(5):404–415, 2004.

[39] World Health Organization (WHO). News release, 25 March 2014, Geneva.

[40] J. Yi and V. R. Prybutok. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environmental Pollution*, 92(3):349–357, 1996.

[41] Y. Zheng, F. Liu, and H.-P. Hsieh. U-air: when urban air quality inference meets big data. In *ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 1436–1444, 2013.