

Pushing the Spatio-Temporal Resolution Limit of Urban Air Pollution Maps

David Hasenfratz, Olga Saukh, Christoph Walser, Christoph Hueglin*, Martin Fierz†, and Lothar Thiele
Computer Engineering and Networks Laboratory, ETH Zurich, Switzerland

*Laboratory for Air Pollution and Environmental Technology, EMPA, Switzerland

†Institute for Aerosol and Sensor Technology, University of Applied Sciences Northwestern Switzerland
{hasenfratz, saukh, walser, thiele}@tik.ee.ethz.ch christoph.hueglin@empa.ch martin.fierz@fhnw.ch

Abstract—Up-to-date information on urban air pollution is of great importance for health protection agencies to assess air quality and provide advice to the general public in a timely manner. In particular, ultrafine particles (UFPs) are widely spread in urban environments and may have a severe impact on human health. However, the lack of knowledge about the spatio-temporal distribution of UFPs hampers profound evaluation of these effects. In this paper, we analyze one of the largest spatially resolved UFP data set publicly available today containing over 25 million measurements. We collected the measurements throughout more than a year using mobile sensor nodes installed on top of public transport vehicles in the city of Zurich, Switzerland. Based on these data, we develop land-use regression models to create pollution maps with a high spatial resolution of $100\text{ m} \times 100\text{ m}$. We compare the accuracy of the derived models across various time scales and observe a rapid drop in accuracy for maps with sub-weekly temporal resolution. To address this problem, we propose a novel modeling approach that incorporates past measurements annotated with metadata into the modeling process. In this way, we achieve a 26% reduction in the root-mean-square error—a standard metric to evaluate the accuracy of air quality models—of pollution maps with semi-daily temporal resolution. We believe that our findings can help epidemiologists to better understand the adverse health effects related to UFPs and serve as a stepping stone towards detailed real-time pollution assessment.

I. INTRODUCTION

Air pollution is a major concern in many cities worldwide. Atmospheric pollutants considerably affect human health; they are responsible for a variety of respiratory and cardiovascular illnesses and are known to cause cancer if humans are exposed to them for extended periods of time [1]. Additionally, air pollution is responsible for environmental problems, such as eutrophication and acidification of ecosystems.

Most countries have mass emission limits for particulate matter PM_{10} and $\text{PM}_{2.5}$ (*i.e.*, particles with a diameter of less than $10\text{ }\mu\text{m}$ and $2.5\text{ }\mu\text{m}$, respectively), but have no restrictions on ultrafine particles (UFPs). UFPs are particles with a diameter of less than 100 nm . In ambient air, UFPs are mainly man-made as byproducts of specific high temperature processes, such as combustion reactions in car engines. The adverse health effects of UFPs are most probably underestimated when they are traditionally monitored by mass as part of PM_{10} and $\text{PM}_{2.5}$ [2]. This is because UFPs make a dominant contribution to the total number of urban particle concentrations, but their contribution to the total particle mass is small [3]. Therefore, UFPs were not considered particularly hazardous in the past. There are strong indications, however, that adverse health

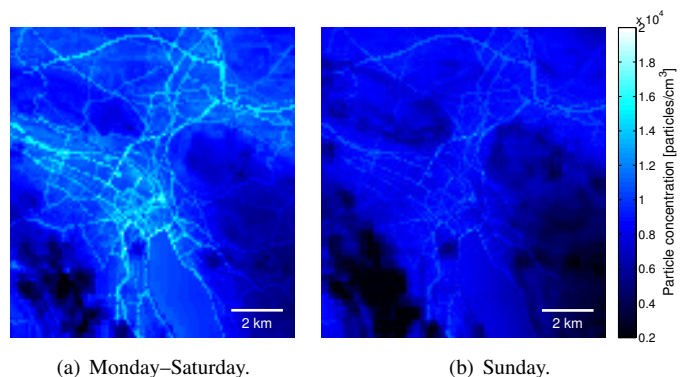


Figure 1. Novel ultrafine particle concentration maps for Zurich (Switzerland). The particle concentrations are higher during the week (Monday–Saturday) than on weekends (Sunday) due to higher traffic volumes.

effects are more related to particle concentration rather than to particle mass [2]. To better understand the adverse health effects of UFPs, it is essential to have spatially resolved UFP concentration measurements at hand [4].

Nowadays, air pollution is monitored by networks of static measurement stations operated by official authorities. These stations are highly reliable and able to accurately measure a wide range of air pollutants. However, their high acquisition and maintenance costs severely limit the number of installations. As a result, very little is known about the spatial distribution of air pollutants in urban environments and there is a lack of accurate intraurban air pollution maps. However, for air pollutants with high spatial variability, such as UFPs, the public availability of reliable pollution maps is essential. They raise the citizens’ awareness about air pollution and empower environmental scientists to craft and evaluate new policies.

Contributions and road-map. To tackle the challenges above, we propose to use a mobile measurement system. Node mobility trades off temporal resolution against spatial resolution, enabling a high spatial resolution across large areas without the need for thousands of fixed sensors. However, due to the lower temporal resolution of any covered location, it is a formidable challenge to derive pollution maps with a high temporal resolution at daily or hourly time scales. In this paper, we demonstrate that a mobile measurement system can effectively be used to derive accurate UFP pollution maps with high spatio-temporal resolution.

Our mobile measurement system consists of ten sensor nodes installed on top of public transport vehicles, which cover a large urban area on a regular schedule. The sensor nodes are equipped with a novel measurement device (MiniDiSCs [5]) to monitor UFP particle concentrations. Throughout more than a year, we collected over 25 million UFP measurements. Based on these data, we develop land-use regression (LUR) models to produce accurate pollution maps with high spatio-temporal resolution, such as those depicted in Fig. 1. LUR models use a set of explanatory variables (land-use and traffic data) to model pollution concentrations at locations not covered by the mobile sensor nodes. In a first step, we evaluate the dependencies between the explanatory variables and the measurements. Then, we exploit these relationships to predict the pollution levels for all locations without measurements but with available land-use information. Using this method and our mobile measurement system, we derive accurate and fine-grained pollution maps, which are valuable to environmental scientists, epidemiologists, and the general public.

In summary, this paper makes the following contributions:

- We introduce in Sec. II our mobile measurement system, which is deployed in the city of Zurich (Switzerland) collecting a highly spatially resolved data set of UFP measurements. As of today, we collected more than 25 million measurements.
- Assessing the quality of the measurements is difficult due to very sparse ground truth data. We post-process the measurements (calibration and filtering) and propose in Sec. III a three-fold validation approach to evaluate the quality of the processed data. Our analysis indicates a high data quality.
- We use the validated measurements in Sec. IV to derive LUR models for UFP pollution maps with a high spatial resolution of $100\text{ m} \times 100\text{ m}$. In Sec. V we apply standard metrics to analyze the quality of the models from yearly up to semi-daily temporal resolutions. We find a good quality of pollution maps with yearly to weekly time scales, while models with sub-weekly temporal resolutions perform less well.
- To tackle this problem, we propose in Sec. VI a novel modeling approach that incorporates past measurements (annotated with metadata, such as meteorological conditions) into the modeling process. In this way, we increase the quality of pollution maps with a high temporal resolution. For example, we decrease the root-mean-square error—a standard metric to evaluate the accuracy of air quality models—of semi-daily pollution maps by 26%.

Using our measurement system and modeling approach, we create UFP pollution maps with an accuracy that is comparable to state-of-the-art air pollution maps [6], while achieving unprecedented spatio-temporal resolution. We survey related work in Sec. VII, and conclude in Sec. VIII.

II. MOBILE AIR POLLUTION MONITORING SYSTEM

Starting in 2012, we gradually equipped ten streetcars of the public transport network in Zurich, Switzerland, with air quality measurement stations as part of the OpenSense project [7]. The sensor nodes cover on a regular schedule a

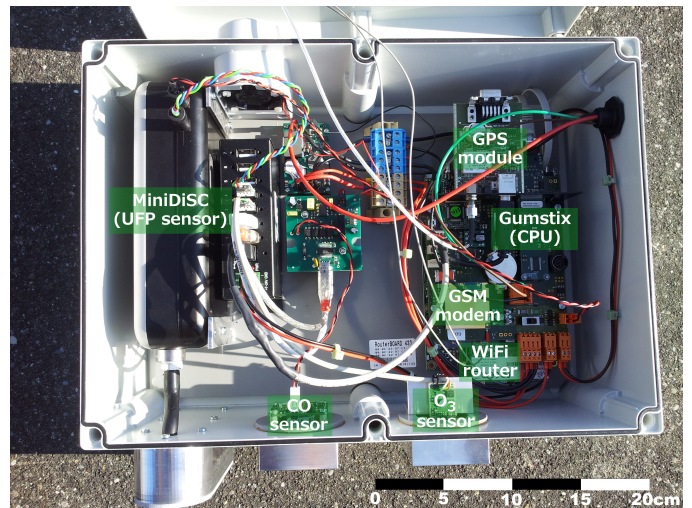


Figure 2. The air quality sensor node is equipped with UFP, O_3 , and CO sensors. Geotagged and time-stamped measurements are transmitted over GSM (cellular network) to the back-end server for further processing.

large urban area of 100 km^2 and monitor a wide range of air pollutants and environmental parameters. The measurements are locally stored in a database and transmitted in real-time over GSM (cellular network) to the back-end server running Global Sensor Network (GSN) [8], a software middleware that facilitates data collection in sensor networks. The sensor readings are removed from the local database once their reception is acknowledged by the back-end server. The web-based caching and visualization tool Vizzly [9] displays the measured particle concentrations on top of Google Maps. All measurements are publicly available.¹

A. Air Quality Sensor Node

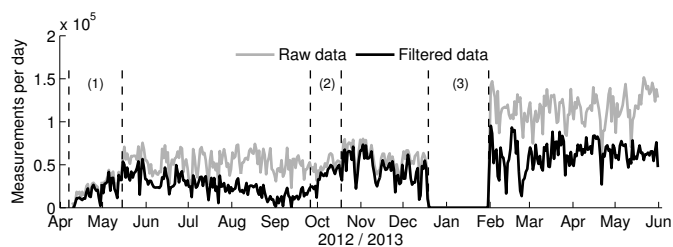
The core of the sensor node, depicted in Fig. 2, is a Gumstix (embedded computer) with a 600 MHz CPU running an embedded Linux operating system. A GPS receiver supplies the station with precise geospatial information. The station supports bidirectional communication over GSM and WiFi. The former is used under normal conditions while the latter is for reliable access during debugging and maintenance phases. While the streetcars are in operation they supply the nodes with power. During the night, typically from 1:00 AM to 5:00 AM, the streetcars are in their depots and the nodes are turned off.

All air intakes are equipped with protection covers against water and dust. To monitor air pollution, the sensor nodes are equipped with low-cost ozone (O_3) and carbon monoxide (CO) sensors and a novel compact device to measure UFP concentrations. Additionally, the nodes monitor environmental parameters, including temperature and humidity. In this work, we focus on analyzing the UFP measurements.

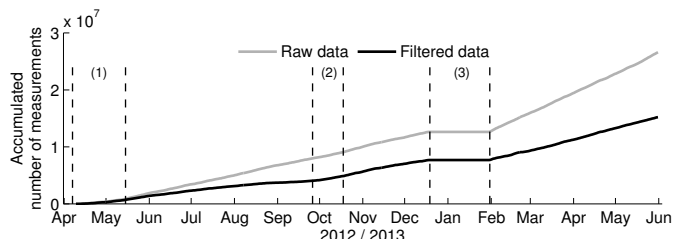
B. Data Set of Ultrafine Particle Concentrations

The sensor nodes are equipped with Miniature Diffusion Size Classifiers (MiniDiSCs) [5] (see Fig. 2), a novel tool for UFP monitoring. Traditionally, UFPs are monitored by mass, hence their adverse health effects are very

¹<http://www.opensense.ethz.ch>



(a) Number of measurements per day across all installed devices.



(b) Accumulated number of measurements across all installed devices.

Figure 3. Number of UFP measurements per day and accumulated number of measurements over the course of 14 months with three service phases (1)–(3).

likely underestimated [2]. Whereas the MiniDiSC, a compact hand-held device (4 x 9 x 18 cm), is able to monitor particle number concentrations. It can detect concentrations between 10^3 – 10^6 particles/cm³, so it embraces the average daily range in urban environments of 10^4 – 10^5 particles/cm³.

The functional principle of the MiniDiSC is based on unipolar charging of aerosol particles, followed by detection in two electrometer stages. In short, the particles are first charged in a standard positive unipolar diffusion charger, which imparts an average charge on the particles that is approximately proportional to the particle diameter. The charged particles then flow through a diffusion stage and generate a current, which is dependent on the average particle size and is used to calculate the number of particles per cm³.

We deployed the first five sensor nodes with integrated MiniDiSCs in April and May 2012 and an additional five devices in January 2013, as shown in Fig. 3. The MiniDiSCs sample UFP every 50 ms. The measurements are aggregated to one sample per 5 s, to reduce the amount of transmitted data. Each sensor node transmits around 10,000 measurements per day to the back-end infrastructure, as depicted in Fig. 3(a) (raw data). In total, we collected over 25 million aggregated measurements, as shown in Fig. 3(b). Further, we depict in Fig. 3 the three main service phases since the start of the deployment: (1) Initial installation of the first five sensor nodes, (2) cleaning of the MiniDiSCs on top of the streetcars (no deinstallation required) to increase data quality, and (3) deinstallation of the MiniDiSCs for cleaning and re-calibration, and installation of five additional devices.

Data calibration and filtering. We ensure a high data quality by calibrating and filtering the timestamped and geo-tagged measurements. For one minute in every hour the devices go into a self-calibrating phase to measure their null-offsets, which we use offline to adjust the offset of all measured particle concentrations. Then, we use a two-stage filtering process to remove faulty and unreliable measurements.

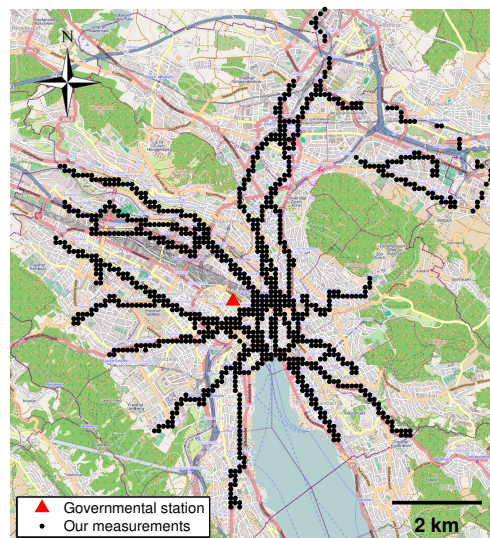


Figure 4. Ten mobile sensor nodes deployed on top of public transport vehicles achieve a good coverage of the city of Zurich (Switzerland). The dots denote locations with at least 50 measurements over the course of 14 months.

First, a GPS-based filter eliminates measurements with horizontal dilution of precision (HDOP) values above 3. The HDOP value specifies the GPS location’s precision based on the geometric positioning of the GPS satellites. Values below 3 denote a good to excellent positioning within a few meters. We consider only these measurements (99 % of the total).

The second filter examines the internal status variables of the MiniDiSCs, which are transmitted to the back-end server in addition to the particle concentration numbers. These variables allow to draw conclusions about the proper functioning of the devices. We discard measurements if a status variable indicates a malfunctioning, such as a too low air flow. Additionally, the MiniDiSCs need a warm-up phase; therefore, all measurements within one hour after start-up are discarded. The two-stage filtering process invalidates around 40 % of the measurements (mainly the second filter), as shown in Fig. 3, but is crucial to achieve high data quality, as we will detail in the next section.

Spatial coverage. A good spatial coverage of the measurements is essential to precisely assess the concentration distribution in urban environments. The spatial coverage of the filtered UFP measurements is shown in Fig. 4. The dots denote locations (100 m x 100 m) with at least 50 measurements over the course of 14 months. The ten mobile sensor nodes achieve a good coverage, in particular compared to the single governmental station monitoring UFPs in Zurich, denoted with a triangle in Fig. 4. Our measurements cover a large set of diverse location characteristics. For example, the data set includes measurement locations at terrain elevations from 400–610 m and at diverse traffic densities ranging from vehicle-free zones to areas with over 90,000 vehicles per day.

III. DATA VALIDATION

Good data quality is a must for the development of reliable pollution maps. Evaluating the quality of measurements spread over a large urban area is a challenging task, especially if no (or only sparse) ground truth is available. The MiniDiSC has been thoroughly tested in laboratory environment and in the

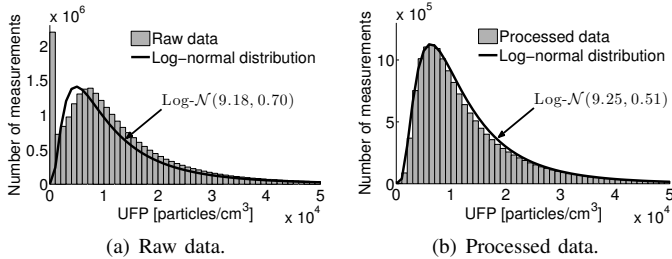


Figure 5. The log-normal distribution (black) with mean and standard deviation of the UFP data (gray). After calibration and filtering (processed data), the log-normal distribution accurately fits the distribution of the measurements.

field [5]. However, despite these tests it is unclear whether the harsh deployment setting on top of streetcars has an impact on the quality of the measurements. Among others, the devices have to endure mobility, constant vibrations, high temperature and humidity variations, and long, unattended operating times.

Thus, we propose a three-fold validation approach to assess the quality of our measurements. We analyze (i) the statistical distribution of the monitored particle concentrations, (ii) the baseline signal of each device, and (iii) compare our measurements to data from two high-quality stations collected during the same time period but at different locations in Switzerland.

A. Statistical Distribution

The statistical distribution of data from many different scientific disciplines, including the concentration of ambient air pollutants, closely follows a log-normal distribution, *i.e.*, the logarithm of the measurements is normally distributed [10]. Previous work indicates that also UFPs are approximately log-normally distributed [11]. Here, we confirm that UFPs are log-normally distributed in urban environments. We show in Fig. 5 the distribution of the raw and processed (*i.e.*, applied calibration and filtering) data and the log-normal distributions with mean and standard deviation of the raw and processed data, respectively. Raw data are not log-normally distributed, as apparent from Fig. 5(a). However, as we show in Fig. 5(b), the distribution of the processed data nicely fits the log-normal distribution $\text{Log-}\mathcal{N}(\mu, \sigma^2)$ with $\mu = 9.25$ and $\sigma^2 = 0.51$.

B. Baseline Signal

We examine the correct offset of the MiniDiSCs, by looking at their baseline signals, *i.e.*, low-pass filtered measurements. We expect a similar baseline signal across all devices, because they all take measurements in the same region (streetcars are not bounded to specific lines). We construct the baseline signal of each device with a simple low-pass filter. We take for every 90 min time interval, the average time for a streetcar to cross the city twice, the 20th percentile of the measured concentrations. Our analysis confirms that over time the baseline signals are similar across all ten devices.

C. Comparison to High-Quality Data Sets

We compare our data set to UFP concentrations measured by static stations of the Swiss National Air Pollution Monitoring Network (NABEL). Five out of 16 NABEL stations are equipped with accurate UFP counters located in urban,

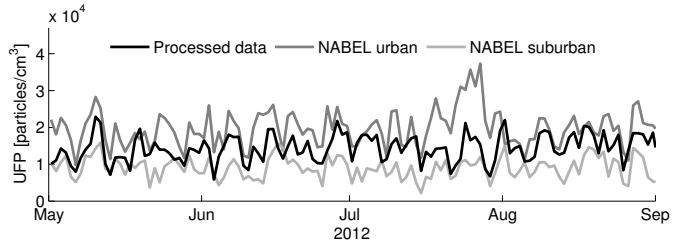


Figure 6. Daily average UFP concentration measured by the mobile sensor nodes in Zurich corresponds well to measurements of two static NABEL stations with urban (Pearson $r=0.49$) and suburban ($r=0.55$) locations.

suburban, and rural environments. For the comparison we choose data from the stations with urban heavy traffic (Bern-Bollwerk) and suburban (Basel-Binningen) backgrounds, as these locations reflect best the topographical extremes of our measurement region in Zurich.

We compare the daily measured average particle concentrations from the mobile sensor nodes to data of these two static stations. Thereof we depict in Fig. 6 an excerpt of 4 months. It shows that the three data sets have a similar trend, further reflected by the good Pearson correlations of $r=0.49$ and $r=0.55$ between our processed data and the urban and suburban located stations, respectively. The two NABEL data sets have a slightly lower correlation of $r=0.46$. Furthermore, as we would expect, the daily average UFP concentration in Zurich ranges between the daily averages measured at the urban location with heavy traffic and at the suburban location.

The above analyses indicate a good quality of our data. In the following, we use the processed data set to derive land-use regression models for accurate high-resolution pollution maps.

IV. DEVELOPING LAND-USE REGRESSION MODELS TO CREATE HIGH-RESOLUTION POLLUTION MAPS

Land-use regression (LUR) models are widely used to assess spatial variation of air pollutants, typically at an intraurban scale [6], [12]. LUR models use land-use and traffic characteristics (explanatory variables) to predict pollution levels for locations not covered by measurement devices. The general concept is based on two steps:

- 1) At all measurement locations the dependencies between explanatory variables (*e.g.*, population density, traffic volume, and terrain elevation) and monitored pollution levels are evaluated using linear regression.
- 2) The found relationships between monitored concentrations and explanatory variables are used to predict concentration levels at locations without measurements but with available land-use data.

There are different approaches to construct LUR models. We use Generalized Additive Models (GAMs) [13], because they support non-linear relationships between monitored concentration levels and explanatory variables. Furthermore, GAMs have been used recently to analyze and model the spatio-temporal variability of particulate matter [14], [15].

A. Generalized Additive Models

We divide our data into different time scales and build for each time scale a separate model. This yields 989 mod-

Table I. 12 EXPLANATORY VARIABLES ARE EXAMINED TO BUILD THE AIR QUALITY MODELS FOR UFP POLLUTION MAPS.

Variable [unit]	Variable [unit]
Population [inhabitants/ha]	Industry [industry buildings/ha]
Building height [floor levels/ha]	Heating [oil and gas heatings/ha]
Terrain elevation [average m/ha]	Road type [busiest road type/ha]*
Distance to next road [m]	Distance to next large road [m] [†]
Terrain slope [average degree/ha]	Terrain aspect [average degree/ha]
Traffic volume [vehicles per day/ha]	Distance to next traffic signal [m]

*Five road types: residential, tertiary, secondary, primary, and freeway.

[†]Road types classified as large: secondary, primary, and freeway.

els with yearly, seasonal, monthly, biweekly, weekly, daily, and semi-daily (midnight–noon and noon–midnight) temporal resolutions. For all models we use the following relationship between pollution concentration and explanatory variables:

$$\ln(c_{\text{num}}) = a + s_1(A_1) + s_2(A_2) + \dots + s_n(A_n) + \epsilon, \quad (1)$$

where c_{num} denotes the UFP concentration, a the intercept, $s_1(A_1) \dots s_n(A_n)$ the smooth functions $s_1 \dots s_n$ with explanatory variables $A_1 \dots A_n$, and ϵ the error term. The non-parametric functions $s_1 \dots s_n$ are smooth regression splines with an upper limit of 3 on the degrees of freedom. Through empirical evaluations we found that the logarithmic link function, $\ln(c_{\text{num}})$ in (1), yields the best model residuals, which are the differences between measured and predicted values.

B. Selecting Explanatory Variables

We collected a set of 12 explanatory variables, as listed in Table I. Data to calculate population density, industry density, building heights, heating type, and elevation, slope, and aspect of the terrain are from the Swiss Federal Statistical Office. Road types and distances to next road and traffic signal are extracted from OpenStreetMap data. Average daily traffic volumes are obtained from the Department of Waste, Water, Energy, and Air of the Canton of Zurich. Some of the data sets are based on measurements from 2007–2010. We assume that these explanatory variables did not change considerably compared to 2012/2013, which is the monitored and modeled time period. Most explanatory variables have a 100 m x 100 m (1 ha) resolution, which we use in general for all variables, as indicated in Table I. This automatically sets the spatial resolution of the generated pollution maps.

It is crucial to examine the correlation between each pair of explanatory variables. Removing variables that have high correlation with each other helps to better distinguish individual contributions of different variables to the modeled particle concentration and prevents redundancy in the model [16]. We do not use population density and number of gas and oil heating households as both have a strong linear relationship with each other and with the number of floor levels (R^2 is larger than 0.6 [14]). Using the distance to the next traffic signal did not improve any of the tested models, and is, therefore, also removed from the modeling process.

The terms on the right side in (1) have a multiplicative relationship with the particle concentration c_{num} , *i.e.*, $c_{\text{num}} = e^a \cdot e^{s_1(A_1)} \cdot e^{s_2(A_2)} \dots$. Thus, we can specify the influence of each explanatory variable A_n as a multiplicative influence factor $e^{s_n(A_n)}$. Fig. 7 depicts typical influence factors observed in our models. Building height and daily traffic volume have an amplifying while terrain elevation has a reducing

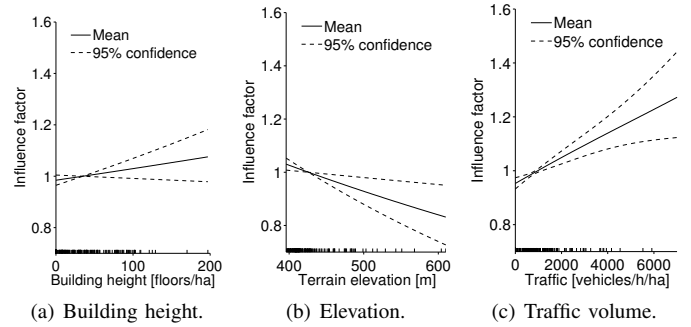


Figure 7. Typical influence factors. Building height and traffic volume have a positive while terrain elevation has a negative impact on UFP concentrations.

impact on the particle concentration. The rugs on the x-axis show that a reduced number of measurements is available at the extremes resulting in decreased confidence bands.

C. UFP Data Aggregation

According to the resolution of the explanatory variables, the measured UFP concentrations of one year (April 2012 to March 2013) are projected on a grid with 13,200 cells, each of size 100 m x 100 m, covering the complete region of interest depicted in Fig. 4. The used subset of the data depends on the desired temporal resolution of the model, *i.e.*, for daily maps we consider measurements from a single day while for seasonal maps those from the entire season. For each grid cell i , we fetch the n_i measurements located in the cell to calculate the mean pollution concentration c_i^m . The applied GPS filter ensures that the localization is precise enough to correctly assign the measurements to their cells with high probability.

The measurements are (unevenly) distributed among 300–1300 different cells depending on the analyzed temporal subset of the data. For example, every black dot in Fig. 4 denotes a cell with at least 50 measurements when projecting the full 14-month data set onto the grid. As model input we use the 200 cells with the highest number of measurements, which are mainly cells containing a streetcar stop. This ensures that the calculated means c_i^m are reliable and provides a good trade-off between spatial input distribution and model performance, as we will show in Sec. V. Introducing prior weights on the cells (*e.g.*, using number of measurements per cell) does not improve the models.

D. Model Output: High-Resolution Pollution Maps

We use the described method above to develop models for pollution maps with different temporal resolutions. For instance, the output of the models with seasonal resolution (winter, spring, summer, and fall) is shown in Fig. 8. In general, terrain elevation, building heights, and traffic density have the highest influence on the predicted pollution levels. Further, we see that the pollution levels are higher in winter and fall than in spring and summer. This is due to frequent high-inversion fog in Zurich from October to March preventing the pollutants to be lifted from the surface and therewith increasing pollution concentration near the ground [17]. The topographical structure around Zurich makes the city prone to inversion fog as it is closely surrounded by hills in the north-east and south-west.

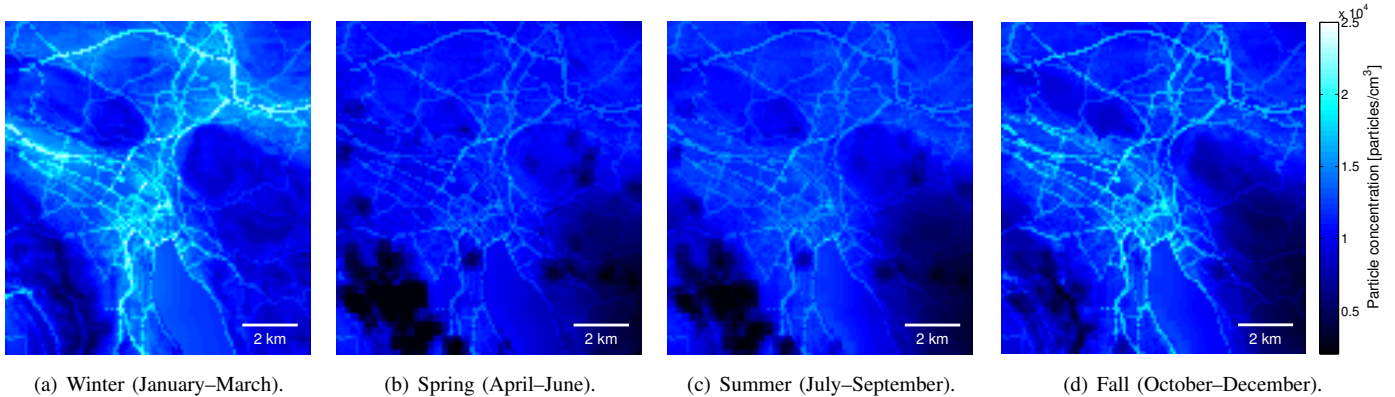


Figure 8. Seasonal UFP pollution maps with a spatial resolution of $100\text{ m} \times 100\text{ m}$ for Zurich (Switzerland), based on data collected by mobile sensor nodes throughout a year. Terrain elevation, building heights, and traffic density have the highest impact on the modeled particle concentrations. Frequent high-inversion fog in winter and fall lead to higher particle concentrations than in spring and summer.

V. REVEALING THE TEMPORAL RESOLUTION LIMIT

This section evaluates the performance of the 989 generated air quality models with yearly to semi-daily temporal resolutions based on our measurements from April 2012 to March 2013. Our evaluation reveals the following findings:

- Models for pollution maps with yearly to weekly time scales have an accuracy that is comparable to recently published state-of-the-art air quality models.
- Pollution maps with daily and semi-daily temporal resolutions are less accurate, because a lower number of measurements are available to build the models.

A. Metrics and Evaluation Methodology

We use three standard metrics to evaluate our models [18].

Factor of 2 measure (FAC2): Quantitatively analyzes scatter plots (predicted concentrations plotted against measured concentrations) by evaluating the fraction of data points that lie inside the factor two area, *i.e.*, fraction of data that satisfy:

$$0.5 \leq \frac{c_i^p}{c_i^m} \leq 2.0, \quad (2)$$

where c_i^p is the model predicted and c_i^m is the average measured concentration of grid cell i . This measure is based on the belief that an accurate model for pollution maps should have a relative scatter less than a factor of two [18].

Adjusted coefficient of determination (R^2): Indicates from 0 to 1 how well the predicted concentrations fit the measurements ($R^2=1$ denotes a perfect fit). R^2 reflects the linear relationship between predicted and measured values, hence, it is insensitive to additive and multiplicative errors.

Root-mean-square error (RMSE): Quantifies the difference between the predicted and measured particle concentrations:

$$\text{RMSE} = \left(\frac{\sum_{i=1}^N (c_i^p - c_i^m)^2}{N} \right)^{0.5}, \quad (3)$$

where N denotes the number of cells used in the evaluation.

For all evaluations we perform a 10-fold cross-validation. That is, we randomly remove 10% of the measurements and

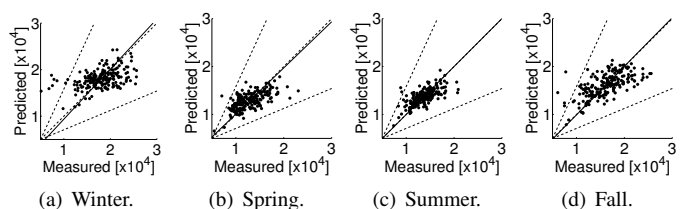


Figure 9. Scatter plots. Model predicted versus measured UFP concentrations (particles/cm³) for seasonal pollution maps. The dashed lines depict the factor of two area (slopes of 0.5 and 2) and the 1-to-1 relation (slope of 1). All linear regression lines (solid) are very close to the 1-to-1 lines.

create (calibrate) the model using the remaining 90% of the data. We use the predictions at the locations of the removed 10% to evaluate the metrics above. We repeat this procedure 40 times to have a good coverage of the complete data set.

B. Model Performance

We first ensure that the models have no systematic bias and then use the metrics above to evaluate their accuracy.

Trustworthy models should not have any systematic bias in their predictions. In general, the predictions did not show any bias. We show this exemplary in Fig. 9 on scatter plots of seasonal pollution maps (see Fig. 8). The dashed lines denote the factor of two area, *i.e.*, slopes of 0.5 and 2, and the optimal 1-to-1 relation, *i.e.*, slope of 1. The linear regression lines (solid) are very close to the 1-to-1 lines for all depicted seasons, indicating the absence of systematic model errors.

Fig. 10 shows the three metrics for all analyzed temporal time scales. Models with yearly to weekly temporal resolutions have very high FAC2s, while daily and semi-daily predictions have a considerable number of models with lower FAC2 values, *i.e.*, many predictions are more than a factor of two off. The RMSEs grow with increasing temporal resolution. Yearly to weekly pollution maps have low RMSEs, which are in range of recently published air quality models [6]. Models with higher temporal resolutions, such as daily and semi-daily pollution maps, have on average 38% higher RMSEs.

The average R^2 of yearly to monthly maps is 0.38, and slightly decreases for shorter time scales. Recently published

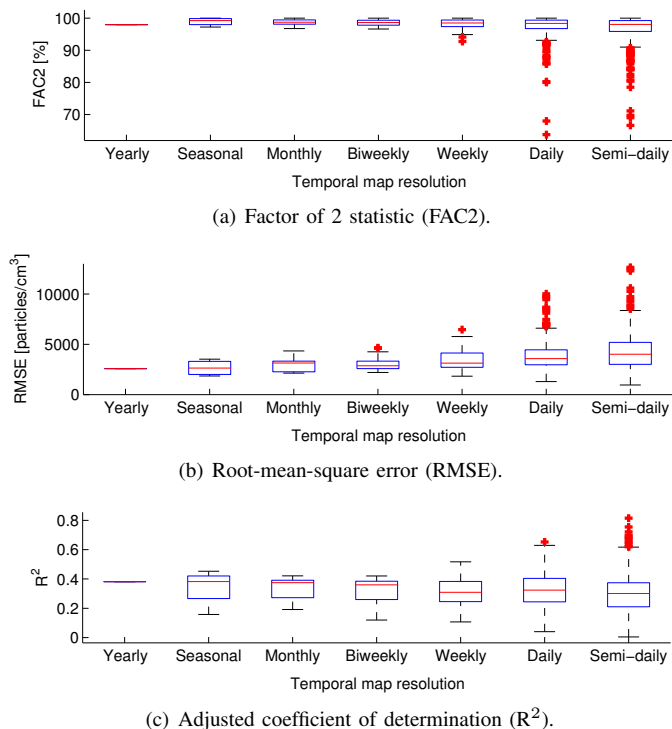


Figure 10. Models with yearly to weekly temporal resolutions have a good quality with high FAC2 values, low RMSEs, and acceptable R^2 coefficients. Models with higher temporal resolutions perform less well.

air quality models for particulate matter have R^2 values in 0.17–0.82 [6]. Most of these models are based on installations that are just in place for a short time (*i.e.*, days or weeks) with measurement devices employed at a carefully selected set of around 40 locations, on average. In our deployment we have a large number of measurement locations, but (perhaps counterintuitively) with many locations it is more difficult to get predictions with high R^2 values [19], [20]. We found that a restriction to 200 grid cells leads to a good trade-off among the considered metrics, as exemplified in Fig. 11 for yearly pollution maps. Models created with a small number of grid cells have good R^2 s but their validity at locations not covered by the input cells is limited, which results in poor FAC2 and RMSE values. Models developed with many grid cells (> 300) have poor R^2 s and RMSEs. Among others, this is because the number of cells with unreliable average cell concentrations increases as the calculated means are more and more based on a limited number of measurements (remember that the cells are favored according to the number of measurements per cell).

C. Challenges of Developing Temporal High-Resolution Maps

We observe on all examined performance metrics that for higher temporal resolutions it is more difficult to derive accurate pollution maps. The main problem is the lower number of measurements available to calculate the mean particle concentration per cell, leading to the following complications:

- The calculated mean of a grid cell is less robust as it is based on a smaller number of measurements.
- Erroneous and inaccurate measurements (*e.g.*, due to outliers, sensor noise, and outdated calibration) have a higher impact on the mean value of a cell.

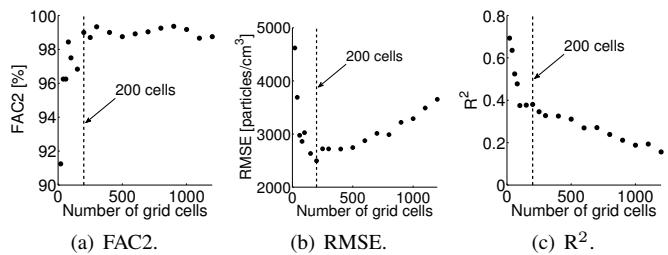


Figure 11. Yearly pollution maps modeled with 20–1200 grid cells. The models achieve a good performance trade-off with 200 cells (dashed line).

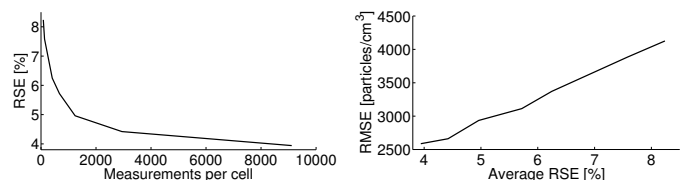


Figure 12. With increased number of measurements the RSE of the cells' mean decreases. Models based on cells with small RSEs have low RMSEs.

To quantitatively support our claims, we examine the relative standard error (RSE) of each cell, which is a good indicator of the reliability of a cell's mean:

$$RSE_i = \frac{\sigma_i^m}{\sqrt{n_i \cdot c_i^m}} \cdot 100, \quad (4)$$

where σ_i^m is the standard deviation of the n_i measurements of cell i . Fig. 12(a) illustrates how the average RSE declines with increasing number of measurements per cell. Models derived from cells with small RSEs have lower RMSEs, as shown in Fig. 12(b). The RSE increase from 4% to 8% results in a 60% rise of the models' RMSEs. Note that we verified that using the median instead of the mean of a cell does not yield improved model performance.

In the following, we propose a new modeling approach to increase the number of measurements per cell and, therewith, improve the accuracy of models with high temporal resolution.

VI. PUSHING THE TEMPORAL RESOLUTION LIMIT

We propose a novel modeling approach, which is able to make use of past pollution measurements to increase the accuracy of highly temporally resolved pollution maps. For that, we introduce a history database containing our measurements from the last 12 months annotated with metadata describing the environmental conditions at time of measurement. We exploit that concentrations of air pollutants show a high correlation with various environmental conditions (*e.g.*, temperature) and the current weekday [15], [21], [22]. The data selector depicted in Fig. 13, selects from the history database those measurements that were performed on the same weekday and under similar environmental conditions as the average condition of the modeled time period. These historical measurements (β) are used together with measurements from the modeled time period (α) to calculate temporally resolved pollution maps, as shown in Fig. 13. For example, we want to create a pollution map for a sunny but windy Saturday, hence, we enrich the

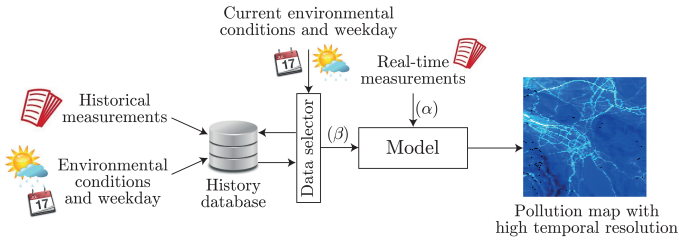


Figure 13. Additional data (β) from a database with historical measurements is used to enhance the original data set (α) to derive pollution maps with high temporal resolutions. The data selector ensures that only those historical measurements are used, which were measured under similar environmental conditions as the average condition of the modeled time period.

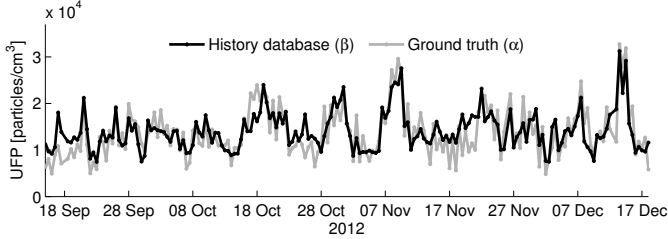


Figure 14. Average semi-daily concentrations based on the history database closely matches ground truth data measured on these days (Pearson $r = 0.74$).

original measurements (α) from that Saturday with historical measurements (β) gathered on a past sunny, windy Saturday.

In the following, we show the feasibility of this novel approach on the case study of semi-daily pollution maps, where the standard model did not deliver a satisfactory accuracy.

A. Data Annotation

We feed the history database with 15 million UFP measurements from a complete year (April 2012 to March 2013). We annotate the measurements with the environmental conditions measured in the city center during this time period by a high-quality static station. The station provides 30 min averages of a diverse set of environmental variables, namely humidity, atmospheric pressure, radiation, precipitation, temperature, wind direction, wind speed, and nitrogen oxide (NO_x) concentration.

B. Data Selector and Quality of Selected Data

The data selector, depicted in Fig. 13, fetches based on the average environmental condition of the modeled half-day those measurements from the history database, which are most likely to be similar to the real measurements from that half-day. To find the best data selector, we evaluate all possible combinations of environmental variables. We introduce a deviation parameter, which controls how closely the metadata of the fetched measurements has to match the given conditions. For example, allowing 20% deviation and a temperature of 15 °C, the data selector returns all data measured at 15 °C \pm 20%, whereas 100% refers to the maximum range in the database.

We evaluate the best data selector by comparing for a complete year the average semi-daily UFP concentration supplied by the two input data sets (β) and (α) depicted in Fig. 13. We observe that the fetched data from the history database (β) are closest to the actual measured concentrations (α) if the selection is based on the following three criteria: temperature, wind

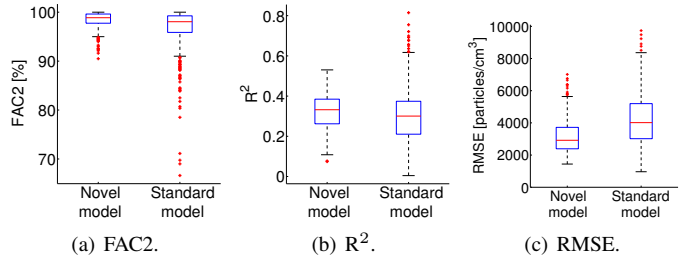


Figure 15. The novel method with a history database helps to increase the quality of semi-daily pollution maps. On average, the RMSE is 26% lower using the novel model compared to the standard model.

direction, and NO_x concentration with an allowed deviation of 10%. The two data sets have a high Pearson correlation of 0.74 and a low average absolute difference of 2,500 particles/ cm^3 , as shown on a three-month extract in Fig. 14. In this way, we attain on average a 14x increase in data volume from 19,000 (α) to 260,000 ($\alpha + \beta$) measurements per half-day.

C. Increased Quality of Semi-Daily Pollution Maps

The extended number of measurements helps to derive pollution maps with a higher accuracy. We compare our modeling approach with the standard approach using the three performing metrics. Fig. 15 shows that using our novel modeling approach increases FAC2s by 2%, R^2 s by 6%, and decreases RMSEs by 26%, on average. The semi-daily pollution maps created with the new models achieve a similar accuracy as the weekly pollution maps created with the standard models.

This new modeling approach advances the generation of accurate pollution maps with high temporal resolutions and simultaneously enables the forecasting of pollution maps for specific environmental conditions, *e.g.*, by using weather forecast data to create pollution maps for the next hours or days.

VII. RELATED WORK

Monitoring airborne pollutants with mobile, low-cost sensors has gained much attention in recent years. Low-cost gas sensors are often embedded in custom-built, personal sensor nodes that are part of participatory sensing networks [23], [24]. Participants can directly monitor the exposure level at locations where they spent their time. However, reaching the critical mass of users to get a coherent picture of the exposure situation in the area of interest is a formidable challenge and may require hundreds to thousands of contributors. Our approach uses a small number of non-personal sensor nodes deployed on top of public transport vehicles to automatically obtain a constant coverage in the area of interest.

There are different kinds of models that can be developed to predict intraurban pollution concentrations, such as proximity-based assessment, statistical interpolation, LUR, and line dispersion models [12]. In this work, we develop LUR models since they have, compared to other models, a relatively low computational overhead, which is beneficial when deriving many hundreds of models. Above all, in the past, LUR models were applied to predict the concentration of a wide range of air pollutants [6]. Most of these models are based on data from static monitoring stations installed at a small set of hand-picked

locations (typically around 40) for some days or couple of weeks [6]. In contrast, our mobile sensor nodes cover a much larger number of locations, *i.e.*, every day around 300 grid cells of size 100 m x 100 m. Therewith, we develop pollution maps with a high spatio-temporal resolution.

A number of studies address the challenge of developing models with high temporal resolutions. Liu et al. [15] derive daily PM_{2.5} maps with a spatial resolution of 4 km by using satellite data in combination with meteorological features. Su et al. [19] extend their NO₂ and NO models with hourly meteorological conditions, such as wind speed and wind direction, to create pollution maps with an hourly resolution. However, it remains unclear how much the meteorological data helps to improve their models. Furthermore, the performance of the hourly resolved pollution maps was not evaluated.

Clifford et al. [21] and Mølgaard et al. [11] develop UFP models with high temporal resolutions by using not only traffic and land-use characteristics data but also meteorological variables as model covariates. Both works solely model temporal variability as their data originates from a single static station.

Many of the above works use meteorological data to enhance their models, *i.e.*, [11], [15], [19], [21]. However, their methods of integrating meteorological data into the modeling process only works if (i) the specific meteorological conditions are known at each measurement location and (ii) the meteorological conditions are significantly different among locations for a given instant in time. Most often, both conditions do not apply for spatially resolved intraurban data sets.

VIII. CONCLUSIONS

Today, little is known about the spatial distribution of air pollutants in urban environments. We address this problem by deploying mobile sensor nodes on top of public transport vehicles in the city of Zurich, Switzerland. We take advantage of the transport network's coverage and regular schedule to collect one of the largest spatially resolved data set containing over 25 million ultrafine particle measurements. We develop land-use regression models to create pollution maps with a high spatial resolution and study their temporal resolution limit. We reveal that the accuracy of pollution maps with sub-weekly temporal resolution suffers from the limited number of measurements available to model the pollution concentrations. We tackle this problem by proposing a novel modeling approach, which is able to make use of past measurements to increase the available data volume. Therewith, we develop accurate ultrafine particle pollution maps with a high spatio-temporal resolution. These maps are valuable to the general public as well as to environmental scientists and epidemiologists to shed more light on the adverse health effects of ultrafine particles. We believe that our findings promote the accurate, detailed, and timely assessment of air quality in urban environments.

Acknowledgements. We thank Marco Zimmerling and the anonymous reviewers for their valuable comments, which helped us to improve this paper. This work was funded by NanoTera.ch with Swiss Confederation financing.

REFERENCES

[1] D. Mage, G. Ozolins, P. Peterson, A. Webster, R. Orthofer, V. Vandeweerd, and M. Gwynne, "Urban air pollution in megacities of the world," in *Elsevier Atmospheric Environment*, 1996.

[2] P. Penttinen, K. L. Timonen, P. Tiittanen, A. Mirme, J. Ruuskanen, and J. Pekkanen, "Number concentration and size of particles in urban air: effects on spirometric lung function in adult asthmatic subjects." *Environmental Health Perspectives*, vol. 109, no. 4, 2001.

[3] R. Harrison and J. Yin, "Particulate matter in the atmosphere: which particle properties are important for its effects on health?" *Science of the Total Environment*, vol. 249, no. 1, 2000.

[4] G. Hoek, H. Boogaard, A. Knol, J. de Hartog, P. Slottje, J. Ayres, P. Borm, B. Brunekreef, K. Donaldson, and F. e. a. Forastiere, "Concentration response functions for ultrafine particles and all-cause mortality and hospital admissions: results of a european expert panel elicitation," *Environmental Science and Technology*, vol. 44, no. 1, 2009.

[5] M. Fierz, C. Houle, P. Steigmeier, and H. Burtcher, "Design, calibration, and field performance of a miniature diffusion size classifier," *Aerosol Science and Technology*, vol. 45, no. 1, 2011.

[6] G. Hoek, R. Beelen, K. de Hoogh, D. Vienneau, J. Gulliver, P. Fischer, and D. Briggs, "A review of land-use regression models to assess spatial variation of outdoor air pollution," *Atmospheric Environment*, vol. 42, no. 33, 2008.

[7] J. J. Li, B. Faltings, O. Saukh, D. Hasenfratz, and J. Beutel, "Sensing the air we breathe—the OpenSense Zurich dataset," in *AAAI*, 2012.

[8] K. Aberer, M. Hauswirth, and A. Salehi, "A middleware for fast and flexible sensor network deployment," in *VLDB*, 2006.

[9] M. Keller, J. Beutel, O. Saukh, and L. Thiele, "Visualizing large sensor network data sets in space and time with vizzly," in *IEEE SenseApp*, 2012.

[10] E. Limpert, W. A. Stahel, and M. Abbt, "Log-normal distributions across the sciences: Keys and clues," *BioScience*, vol. 51, no. 5, 2001.

[11] B. Mølgaard, T. Hussein, J. Corander, and K. Hämeri, "Forecasting size-fractionated particle number concentrations in the urban atmosphere," *Atmospheric Environment*, vol. 46, 2012.

[12] M. Jerrett, A. Arain, P. Kanaroglou, B. Beckerman, D. Potoglou, T. Sahuvaroglu, J. Morrison, and C. Giovis, "A review and evaluation of intraurban air pollution exposure models," *Exposure Science and Environmental Epidemiology*, vol. 15, no. 2, 2004.

[13] T. Hastie and R. Tibshirani, "Generalized additive models," *Chapman & Hall*, 1990.

[14] I. Barmpadimos, C. Hueglin, J. Keller, S. Henne, and A. Prévôt, "Influence of meteorology on PM₁₀ trends and variability in Switzerland from 1991 to 2008," *Atmospheric Chemistry and Physics*, vol. 11, no. 4, 2011.

[15] Y. Liu, C. J. Paciorek, and P. Koutrakis, "Estimating regional spatial and temporal variability of PM_{2.5} concentrations using satellite data, meteorology, and land use information," *Environmental Health Perspectives*, vol. 117, no. 6, 2009.

[16] L. Jackson, N. Carslaw, D. Carslaw, and K. Emmerson, "Modelling trends in OH radical concentrations using generalized additive models," *Atmospheric Chemistry and Physics*, vol. 8, no. 4, 2009.

[17] Federal Office for the Environment, "Measurement results from the Swiss national air pollution monitoring network NABEL," 2009.

[18] J. Chang and S. Hanna, "Air quality model performance evaluation," *Meteorology and Atmospheric Physics*, vol. 87, no. 1-3, 2004.

[19] J. G. Su, M. Brauer, B. Ainslie, D. Steyn, T. Larson, and M. Buzzelli, "An innovative land use regression model incorporating meteorology for exposure analysis," *Science of the Total Environment*, vol. 390, no. 2, 2008.

[20] X. Basagaña, M. Rivera, I. Aguilera, D. Agis, L. Bouso, R. Elosua, M. Foraster, A. de Nazelle, M. Nieuwenhuijsen, J. Vila, and N. Kuenzli, "Effect of the number of measurement sites on land use regression models in estimating local air pollution," *Atmospheric Environment*, vol. 54, 2012.

[21] S. Clifford, S. Low C., T. Hussein, K. Mengersen, and L. Morawska, "Using the generalised additive model to model the particle number count of ultrafine particles," *Atmospheric Environment*, vol. 45, no. 32, 2011.

[22] Y. Zheng, F. Liu, and H.-P. Hsie, "U-air: When urban air quality inference meets big data," in *ACM KDD*, 2013.

[23] P. Dutta, P. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodru, "Demo abstract: Common sense - participatory urban sensing using a network of handheld air quality monitors," in *ACM SenSys*, 2009.

[24] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, "Participatory air pollution monitoring using smartphones," in *Mobile Sensing: From Smartphones and Wearables to Big Data*, 2012.