SmartAQnet 2020: A New Open Urban Air Quality Dataset from Heterogeneous PM Sensors

Chaofan Li[1*], Matthias Budde[1,2], Paul Tremper[1], Klaus Schäfer[3], Johannes Riesterer[1], Johanna Redelstein[4], Erik Petersen[4], Mohamed Khedr[5], Xiangsheng Liu[5], Marcel Köpke[1], Sajjad Hussain[1], Felix Ernst[1], Michal Kowalski[6], Markus Pesch[7], Johannes Werhahn[8], Markus Hank[7], Andreas Philipp[4], Josef Cyrys[6], Jürgen Schnelle-Kreis[5], Hans Grimm[3], Volker Ziegler[7], Annete Peters[6], Stefan Emeis[8], Till Riedel[1], Michael Beigl[1]

[1]*Karlsruhe Institute of Technology (KIT), TECO, Karlsruhe, Germany*
[2]*Disy Informationssysteme GmbH, Karlsruhe, Germany*
[3]*Aerosol Akademie e.V., Ainring, Germany*
[4]*Uni Augsburg, Physical Geography and Quantitative Methods, Augsburg, Germany*
[5]*Helmholtz Zentrum München, Comprehensive Molecular Analytics, Munich, Germany*
[6]*Helmholtz Zentrum München, Institute of Epidemiology II, Munich, Germany*
[7]*GRIMM Aerosol Technik Ainring GmbH & Co. KG, Ainring, Germany*
[8]*Karlsruhe Institute of Technology (KIT), IMK-IFU, Karlsruhe, Germany*
[*]*li@teco.edu*

Abstract

The increasing attention paid to urban air quality modeling places higher requirements on urban air quality datasets. This article introduces a new urban air quality dataset—the SmartAQnet2020 dataset—which has a large span and high resolution in both time and space dimensions. The dataset contains 248,572,003 observations recorded by over 180 individual measurement devices, including ceilometers, Radio Acoustic Sounding System (RASS), mid- and low-cost stationary measuring equipment equipped with meteorological sensors and particle counters, and low-weight portable measuring equipment mounted on different platforms such as trolley, bike, and UAV.

Keywords *Air Quality; Dataset; Aerosol; Particulate Matter; Open Data*

ISSN: 2283-5954
Selection and peer-review under responsibility of DUST 2021 Scientific Committee

1. Introduction

Air pollution causes severe damage to human health. The WHO Air Quality Guidelines (Hoffmann et al., 2021) state that adverse health effects of air pollution can be observed not only in high exposures but also at very low concentration levels. Due to the large concentration of human activities in urban areas, information on urban air pollution is particularly interesting. However, fine-grained monitoring and forecasting of urban air pollution remain a major challenge. The Smart Air Quality Network (SmartAQnet) (Budde et al., 2017a) concentrates on recording urban meteorology and aerosol measurement data in fine granularity using heterogeneous measurement technology. Since 2017, this project has collected over 300 million observations in the model region of the City of Augsburg, Germany, and this number is still rapidly increasing as time goes on.

In this paper, we share the data collected by SmartAQnet between 2017 and 2020 as the open SmartAQnet2020 dataset. The dataset contains 248,572,003 observations recorded by over 180 individual measurement devices, including ceilometers, Radio Acoustic Sounding System (RASS), mid- and low-cost stationary measuring equipment using meteorological sensors and particle counters, and low-weight portable measuring equipment mounted on different platforms such as trolley, bike, and UAV.

2. Background

Current urban air pollution models can be classified into two main types: physical and statistical models. The traditional process of the physical model is first to estimate the possible emission sources. Then take emission sources and meteorological data as inputs, input them into a series of physics equations, which simulate the transfer, diffusion, and chemical reactions of pollutants (Slørdal et al., 2003; Karl et al., 2019).

Unlike physical models that pay more attention to physical and chemical rules, statistical models mainly focus on using methods such as machine learning to summarize the statistical characteristics of historical observation records (Singh et al., 2012), (Cheng et al., 2018). With the rapid development of machine learning technology, more and more researchers have paid attention to statistical models in recent years. Training for good statistical models is becoming more and more data-hungry.

Traditionally, urban air quality data is usually collected by a few stationary, high-precision professional measuring stations (Budde et al., 2014). They are accurate, well maintained, but expensive and need experienced personnel. Only relatively few organizations with sufficient technical and financial resources can establish such measurement networks. Although such stations can provide high-quality data, it is challenging to base fine-grained models on their data because of their scarcity in numbers.

However, in urban areas, factors related to air quality, such as human activities, meteorology, and land use, are highly complex and may change rapidly. In order to meet the current needs, the paradigm of air quality monitoring has started to shift towards monitoring urban air quality by deploying a large number of low-cost measuring sensors (Budde et al. 2013) to achieve higher spatial and temporal resolutions (Li et al., 2012; Snyder et al., 2013). Our measurement network, the Smart Air Quality Network (SmartAQnet), incorporates both data from high-quality measurement stations, as well as lower-cost and -fidelity measurement equipment.

3. The SmartAQnet2020 Dataset

The SmartAQnet2020 dataset includes all the data collected by the SmartAQnet project during the time interval from January 1, 2017, to December 31, 2020. The dataset contains 248,572,003 observations collected in the model region of the City of Augsburg, Germany. Various aerosol and meteorological features are measured and collectively referred to as observed properties. Table 1 below shows the number of observations included in each observed property. There are sometimes multiple observed properties in the Table corresponding to the same measured object because the SmartAQnet 2020 dataset comes from a large number of heterogeneous sensors.

As mentioned above, the SmartAQnet project contains a large number of low-cost sensors. These sensors are cheap and simple for massive deployment, but on the other hand, their working conditions are not as stable as those expensive sensors. Therefore, the actual deployment scale of our project and the number of observations fluctuate over time. Figure 1(a) below shows how the number of observations included in each month in the dataset changes. Figure 1(b) shows the average number of daily available devices for each month.

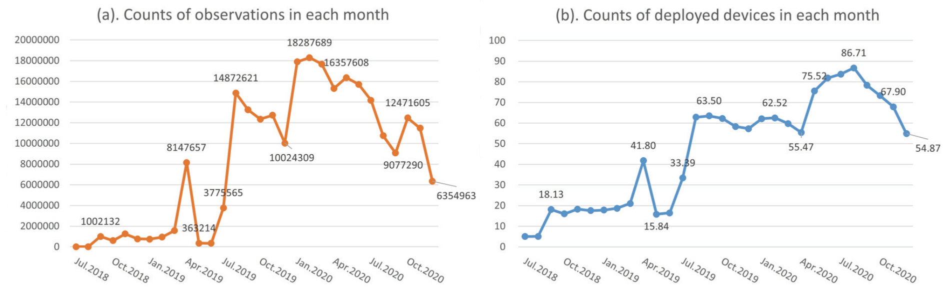

Fig. 1. (a). Counts of observations in each month. (b). Average counts of deployed devices in each month.

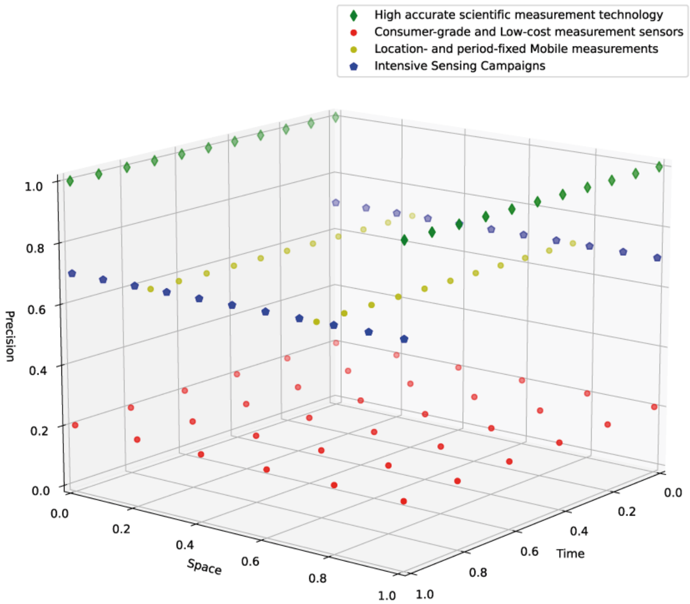

Fig. 2. Characteristics of different parts in terms of time resolution, spatial resolution, and accuracy

Table 1. The number of observations included in each observed property

Abbreviation in dataset	Counts	Description
saqn:op:absp	665,663	Attenuated Backscatter Profile
saqn:op:bc	27,048	Black Carbon
saqn:op:blh	1,262,184	Boundary Layer Height
saqn:op:ca	665,669	Cloud Amount
saqn:op:dp	4,470,887	Dew Point
saqn:op:globalrad	283,541	Global Radiation
saqn:op:hur	48,313,049	Relative Humidity
saqn:op:irbcc	30,696	Infrared Particulate Matter (IRPM)
saqn:op:mcpm	47,088	PM Total Mass Concentration
saqn:op:mcpm1	14,760,604	PM1 Mass Concentration
saqn:op:mcpm10	53,903,439	PM10 Mass Concentration
saqn:op:mcpm2p5	53,759,429	PM2.5 Mass Concentration
saqn:op:mcpm4	1,871,276	PM4 Mass Concentration
saqn:op:mcpmtotal	120,517	PM Total Mass Concentration
saqn:op:mcresp	120,519	PM4 Mass Concentration
saqn:op:ncpm	27,048	PM Total Particle Number Concentration
saqn:op:ncpm1	54,096	PM1 Particle Number Concentration
saqn:op:ncpm10	9,465,732	PM10 Particle Number Concentration
saqn:op:ncpm2p5	81,144	PM2.5 Particle Number Concentration
saqn:op:plev	5,556,495	Air Pressure
saqn:op:pnc0p2-1	177,075	Particle number concentration in size range 0.02 - 1 μm
saqn:op:precip	310,545	Precipitation
saqn:op:sigmaw	16,206	Sigma of the Vertical Wind
saqn:op:ta	48,526,008	Air Temperature
saqn:op:td	1,331,371	Temperature in Device
saqn:op:theta_a	16,229	Acoustic Potential Temperature
saqn:op:total	2,273	PM Total Particle Number Concentration
saqn:op:uvbcc	30,696	Ultraviolet Particulate Matter (UVPM)
saqn:op:wchill	15,104	Wind Chill
saqn:op:wdir	404,999	Wind Direction
saqn:op:wspeed	766,799	Wind Speed
saqn:op:zcb	665,663	Cloud Base Altitude
saqn:op:zcl	665,676	Cloud Layer Altitude

So far, the measuring equipment involved in the SmartAQnet project could be categorized into four parts:

- High accurate scientific measurement technology
- Consumer-grade and Low-cost measurement sensors
- Location- and period-fixed Mobile measurements
- Intensive Sensing Campaigns

Each of these four parts has its own characteristics of time resolution, spatial resolution, and accuracy. As shown in Figure 2, they cooperated in providing a detailed and comprehensive observation system.

3.1 High accurate scientific measurement technology

SmartAQnet is integrated into the existing measurement network in the Augsburg area. In other words, it incorporates the publicly available data of high-precision measurement equipment provided by local authorities.

First of all, there are four state air quality monitoring stations provided by the LfU Bayern (Bavarian State Office for the Environment), which (among other parameters) collects PM10 readings.

The second existing measurement network is a ground-based remote sensing network consisting of a VAISALA CL51 ceilometer and a METEK Radio-Acoustic Sounding System (RASS) (see Emeis et al. (2009); Emeis et al. (2012)), both located on the campus of the University of Augsburg. This measurement network serves to observe cloud, wind, and temperature data by height profiles.

The third existing network is a local meteorological station network consisting of 7 stations, collecting various meteorology properties, such as Temperature, Wind, Pressure, Precipitation, etc.

The three aforementioned networks provide observations with very high accuracy, as well as satisfactory temporal resolution. However, the spatial resolution is relatively poor due to the sparse number of stations.

3.2 Consumer-grade and low-cost measurement sensors

The inexpensive sensors involved in the SmartAQnet project could also be briefly classified into two different precision levels: Consumer-grade sensors and low-cost sensors. Consumer-grade sensors can be seen as a compromise between high-precision measuring stations and low-cost sensors. In terms of price and maintenance costs, they are somewhere between the above two. At the same time, the precision of the data they can provide is also high enough to be used as a reference device. We deployed 6 Grimm EDM-164OPC Sensors during the project as consumer-grade sensors, which serve as reference devices of the Grimm sensor network.

Compared with high-precision stations and consumer-grade sensors, low-cost sensors reduce precision for lower price and maintenance costs, enabling them to be massively deployed to provide much higher temporal and spatial resolutions. In the SmartAQnet project, part of the low-cost sensors come from integrating existing local equipment, and a much more considerable amount was directly deployed during the project. We deployed 22 Grimm EDM-80NEPH Sensors and 35 Grimm EDM-80OPC Sensors. Together with the above-mentioned 6 Grimm EDM-164OPC Sensors, these sensors form the so-called Grimm network, in which low-cost sensors could make intelligent signal evaluation through extensive comparison measurements to reference devices. We also have 84 Crowdsensing Nodes composed of a Nova SDS011 Ultrafine Particulate Sensor and a Bosch BME280 Sensor. All the sensors mentioned above can provide PM and basic meteorological observations, and they together constitute the main body of this part. In addition to these newly deployed devices, two existing devices provided by Helmholtz Zentrum München

(HMGU) are integrated into the project, known as HMGU EPI PM Container and HMGU EPI Meteo Container.

In the SmartAQnet project, the locations of the deployed sensors were also carefully designed. We selected a rectangle area of about 4 x 6 km in the centre of Augsburg as the Central Activity Zone (CAZ), which northwest located at 48.39°N, 10.87°E and southeast located at 48.33°N, 10.92°E. We have consciously increased the deployment density of sensors in the CAZ since it contains three of the four high-precision stations so that the sensors deployed in this area can get better references. The following Figure 3 shows the locations of the CAZ and different kinds of stationary sensors.

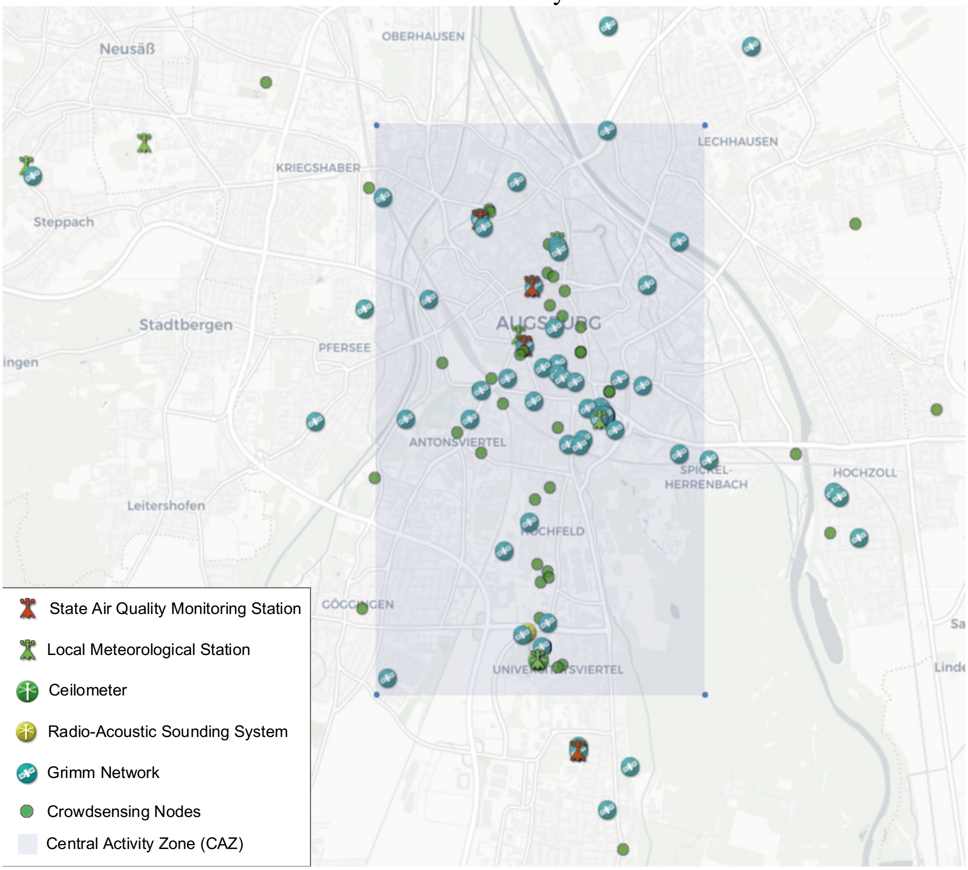

Fig. 3. Locations of the CAZ and different stationary sensors in and around the City of Augsburg, Germany.

3.3 Location- and period-fixed mobile measurements

Mobile measurements were carried out between December 2019 and September 2020 in a fixed period and location using two UAV systems. One is a self-constructed fixed-wing aerial vehicle, and the other is a rotorcraft of type DJI Matrice 600 pro. Both UAVs are equipped with a measuring device that integrates low-weight weather parameter sensors (Sensirion SHT75/SHT85) and particle counters (Alphasense OPC-N2/OPC-N3), special sensor inlets and pumps are installed to reduce the influence of the UAV. Thus, they could detect relative humidity, temperature, and PM concentrations data.

The mobile measurement data provided by the UAV observes the detailed three-dimensional dynamics of the lower atmosphere with fine granularity, which can be used as a powerful supplement to the remote sensing measurement provided by the ceilometer and RASS. However, due to the limitation of UAV load capacity, these measurement data can only be collected with low-weight sensors, so the observation precision cannot match with the highly accurate scientific measurement technology. In addition, since the UAV flight requires the operation of the pilot on the ground, it can only be carried out regularly at a certain frequency, thus cannot fully cover the time dimension.

3.4 Intensive Sensing Campaigns

During the data collection period of SmartAQnet, several intensive sensing campaigns were held. Mobile measuring devices were installed on trolleys and bicycles during the campaigns and were carried through the city by participating personnel.

Between August 2018 and June 2020, several intensive sensing campaigns were carried out using trolleys equipped with portable sensors. The trolleys are equipped with GPS (GPSMAP 64s, Garmin, USA), which records the position with the 1-second resolution, and is equipped with a variety of portable PM sensors, such as DustTrak DRX Aerosolmonitor, P-Trak Ultrafine Particle Counter 8525, Grimm 11e, Aethlabs microAeth MA200, Hand-Held Condensation Particle Counter Model 3007, Testo DISCmini, Aerocet 531S, etc. Figure 4(a) below shows the route of the trolleys' measurements.

Since January 2020, we have introduced another mobile measurement method. Backpacks equipped with GPS, low-weight weather parameter sensors (Sensirion SHT75/SHT85), and particle counters (Alphasense OPC-N2/OPC-N3) were installed on bicycles for mobile measurement. Measurements made with bikes are more frequent compared to measurements performed with trolleys. One could find multiple measurement activities in most months of 2020. Figure 4(b) shows an example route from the bike measurements.

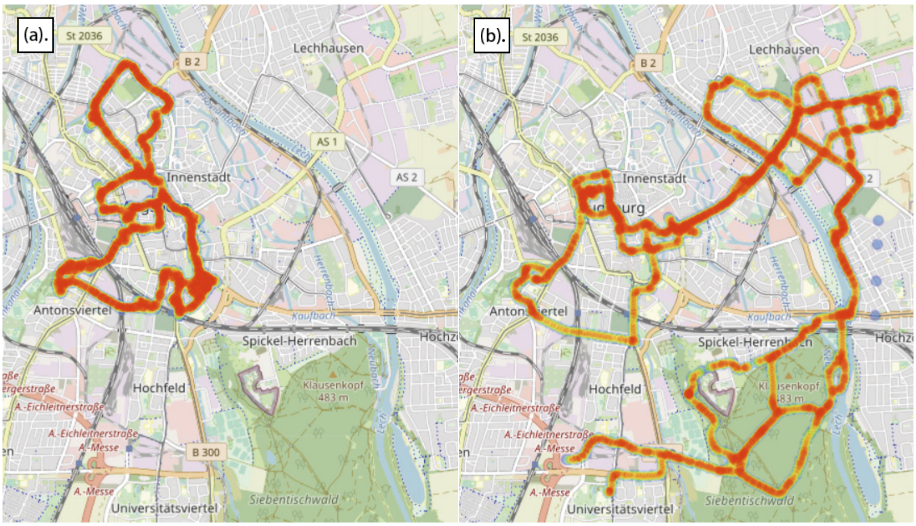

Fig. 4. (a). Route of the trolley measurements. (b). An example route from the bike measurements.

The coverage in the time dimension is more limited due to the effort required to organize activities. However, this part of the data has extremely high coverage in the space dimension.

4. Discussion

In this section, we would like to discuss the further opportunities we believe the SmartAQnet 2020 dataset provides. Work already done based on data from the SmartAQnet project includes the performance evaluation of low-cost PM sensors (Budde et al., 2018), the development of novel calibration approaches (Schlund et al., 2020), spatial modeling (Shen et al., 2019) and interpolation (Tremper et al., 2021) approaches, and higher-level applications such as air-quality-based bike routing (Janßen et al., 2021). Even meta-level discussions have been informed by the experiences in collecting distributed air quality data from heterogeneous sensors, such as work towards sustainable business models for high-resolution air quality assessment (Schäfer et al., 2021).

The first opportunity the SmartAQnet 2020 dataset provides is modeling urban air quality. In our dataset, we have a large number of observations, observing the model area with high resolution. It can meet the requirements of statistical modeling very well. On the other hand, since we have also recorded various meteorological data, it could also be used for physical modeling.

Secondly, the dataset can also be used to evaluate Spatial-temporal interpolation algorithms. The data provided by the high-precision measuring station and the intensive sensing campaigns offer multiple possibilities for evaluation.

Thirdly, the dataset provides good opportunities for understanding the performance of low-cost sensors in the field. SmartAQnet features long-term, high-volume, low-cost sensor usage. Therefore, the dataset holds the possibility to illustrate problems or limitations in data quality (Budde et al., 2017b) when deploying low-cost sensors. It can potentially also be used to benchmark the real-world applicability of different existing (e.g., Budde et al., 2015) or newly proposed data cleaning or distributed calibration methods for low-cost sensors (e.g. (Hasenfratz et al., 2012; Markert et al., 2016; Delaine et al., 2019).

Fourth, the dataset could be used for understanding how the Coronavirus (SARS-CoV-2) – or rather the accompanying changes in urban activity – affected urban air quality. Our dataset covers both the first wave of the Corona period, which means the second half of 2020, and the non-Corona-period. During the Corona period, with different restrictive policies in effect, reduced human activity possibly affected the distribution of particulate matter. By analysing the difference between these two periods, we could gain a deeper understanding of such effects.

5. Accessing the Dataset

The SmartAQnet 2020 dataset is publicly available with a CC BY 4.0 Attribution license. The latest version of the dataset as of this writing can be accessed with the following DOI: 10.35097/540. All works that make use of the SmartAQnet 2020 data should include a reference using that DOI, as well as giving scholarly credit by citing both the SmartAQnet project (DOI: 10.1117/12.2282698) and this dataset paper itself (DOI: 10.14644/dust2021.001).

Since most of the sensors used in the SmartAQnet project are still in operation, we will release our new datasets at a certain frequency as a supplement to the SmartAQnet 2020 Dataset. Information about the latest release of the dataset can be obtained by visiting the official homepage of the SmartAQnet project (www.smartaq.net).

The SmartAQnet 2020 dataset itself only includes PM- and meteorology-related records. Other datasets of factors closely related to the distribution of urban air pollutants, such as land use, traffic volume, etc., are not released together for distinct reasons. If there is a need for these data, we are willing to assist within our capacity.

6. Conclusion

In this paper, we presented the SmartAQnet 2020 Dataset. It was collected in the model region of Augsburg, Germany between 2017 and 2020 and contains 248,572,003 observations recorded by over 180 individual devices, including ceilometers, Radio Acoustic Sounding System (RASS), mid- and low-cost stationary measuring equipment using meteorological sensors and particle counters, and low-weight portable measuring equipment mounted on different platforms such as trolley, bike, and UAV.

We have used the dataset ourselves for a variety of analyses, including the development and evaluation of modeling, spatial interpolation, and distributed calibration approaches. We provide the dataset to the public under a permissive open data license as an opportunity to develop or benchmark existing or novel approaches based on heterogeneous, real-world air quality data.

7. Acknowledgements

This work was partially funded by the German Federal Ministry of Transport and Digital Infrastructure (BMVI) as part of project "SmartAQnet" (19F2003).

References

Budde M., El Masri R., Riedel T., Beigl M. (2013). Enabling low-cost particulate matter measurement for participatory sensing scenarios. Int. Conference on Mobile and Ubiquitous Multimedia (MUM2013).

Budde M., Zhang L., Beigl M. (2014). Distributed, Low-cost Particulate Matter Sensing: Scenarios, Challenges, Approaches. ProScience 1, pp. 230-236. doi:10.14644/dust.2014.038.

Budde M., Köpke M., Beigl M. (2015). Robust, In-situ Data Reconstruction from Poisson Noise for Low-cost, Mobile, Non-Expert Environmental Sensing. Int. Symposium on Wearable Computing (ISWC'15), 179-182.

Budde M., Riedel T., Beigl M., Schäfer K., Emeis S., Cyrys J., Schnelle -Kreis J., Philipp A., Ziegler V., Grimm H., Gratza T. (2017a). SmartAQnet: Remote and In -Situ Sensing of Urban Air Quality, Proc. SPIE 10424, Remote Sensing of Clouds and the Atmosphere XXII, 104240C.

Budde M., Schankin A., Hoffmann J., Danz M., Riedel T., Beigl M. (2017b). Participatory Sensing or Participatory Nonsense? – Mitigating the Effect of Human Error on Data Quality in Citizen Science. Interactive, Mobile, Wearable and Ubiquitous Technology (IMWUT) 1(3).

Budde M., Schwarz A. D., Müller T., Laquai B., Streibl N., Schindler G., Köpke M., Riedel T., Dittler A., Beigl M. (2018). Potential and limitations of the low-cost SDS011 particle sensor for monitoring urban air quality. ProScience, 5, 6-12.

Cheng W., Shen Y., Zhu Y., Huang L. (2018). A neural attention model for urban air quality inference: Learning the weights of monitoring stations. In Thirty-second AAAI conference on artificial intelligence.

Delaine F., Lebental B., Rivano H. (2019). In Situ Calibration Algorithms for Environmental Sensor Networks: A Review. IEEE Sensors Journal, 2019, 19, 11p. 10.1109/JSEN.2019.2910317.

Emeis S., Schäfer K., Münkel C. (2009). Observation of the structure of the urban boundary layer with different ceilometers and validation by RASS data. Meteorologische Zeitschrift 18(2), 149.

Emeis S., Schäfer K., Münkel C., Friedl R., Suppan P. (2012). Evaluation of the interpretation of ceilometer data with RASS and radiosonde data. Boundary-Layer Meteorology 143(1), 25-35.

Hasenfratz D., Saukh O., Thiele L. (2012). On-the-fly calibration of low-cost gas sensors. In European Conference on Wireless Sensor Networks (pp. 228-244).

Hoffmann B., Boogaard H., de Nazelle A., Andersen Z.J., Abramson M., Brauer M., Brunekreef B., Forastiere F., Huang W., Kan H., Kaufman J.D., Katsouyanni K., Krzyzanowski M., Kuenzli N., Laden F., Nieuwenhuijsen

M., Mustapha A., Powell P., Rice M., Roca-Barceló A., Roscoe C.J., Soares A., Straif K., Thurston G. (2021). WHO Air Quality Guidelines 2021–Aiming for Healthier Air for all: A Joint Statement by Medical, Public Health, Scientific Societies and Patient Representative Organisations. International Journal of Public Health, 66.

Janßen J., Tremper P., Riedel, T. (2021). Adaptives luftqualitätsgewichtetes Fahrradrouting mittels Land-use Regression auf Basis offener Daten. INFORMATIK 2021.

Karl M., Walker S.E., Solberg S., Ramacher M.O. (2019). The Eulerian urban dispersion model EPISODE–Part 2: Extensions to the source dispersion and photochemistry for EPISODE–CityChem v1.2 and its application to the city of Hamburg. Geoscientific Model Development 12(8), 3357-3399.

Li J.J., Faltings B., Saukh O., Hasenfratz D., Beutel J. (2012). Sensing the air we breathe—the OpenSense Zurich dataset. In Twenty-Sixth AAAI Conference on Artificial Intelligence.

Markert J.F., Budde M., Schindler G., Klug M., Beigl M. (2016). Private rendezvous-based calibration of low-cost sensors for participatory environmental sensing. International Conference on IoT in Urban Space (pp. 82-85).

Schäfer K., Lande K., Grimm H., Jenniskens G., Gijsbers R., Ziegler V., Hank M., Budde M. (2021). High-Resolution Assessment of Air Quality in Urban Areas—A Business Model Perspective. Atmosphere, 12(5).

Schlund R., Riesterer J., Köpke M., Kowalski M., Tremper P., Budde M., Beigl M. (2020). Calibration of Low-Cost Particulate Matter Sensors with Elastic Weight Consolidation (EWC) as an Incremental Deep Learning Method. In International Summit Smart City 360 (pp. 596-614).

Shen Y., Lehmler S., Murshed S.M., Riedel T. (2019). Characterizing Air Quality in Urban Areas with Mobile Measurement and High Resolution Open Spatial Data: Comparison of Different Machine-Learning Approaches Using a Visual Interface. In International Summit Smart City 360 (pp. 115-126).

Singh K.P., Gupta S., Kumar A., Shukla S.P. (2012). Linear and nonlinear modeling approaches for urban air quality prediction. Science of the Total Environment 426, 244-255.

Slørdal L.H., Solberg S., Walker S.E. (2003). The urban air dispersion model EPISODE applied in AirQUIS2003. Technical description. NILU TR.

Snyder E.G., Watkins T.H., Solomon P.A., Thoma E.D., Williams R.W., Hagler G.S., Shelow D., Hindin D.A., Kilaru V.J., Preuss P.W. (2013). The changing paradigm of air pollution monitoring. Environmental Science & Technology 47(20), 11369-11377.

Tremper P., Riedel T., Budde M. (2021). Spatial Interpolation of Air Quality Data with Multidimensional Gaussian Processes. INFORMATIK 2021.