Analysis and Visualization of Urban Emission Measurements in Smart Cities

Ahlers, Dirk; Kraemer, Frank Alexander; Braten, Anders Eivind; Liu, Xiufeng; Anthonisen, Fredrik; Driscoll, Patrick Arthur; Krogstie, John

Published in:
Proceedings of the 21st International Conference on Extending Database Technology (EDBT)

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Analysis and Visualization of Urban Emission Measurements in Smart Cities

Dirk Ahlers  
NTNU – Norwegian University of Science and Technology  
dirk.ahlers@ntnu.no

Frank Alexander Kraemer  
NTNU – Norwegian University of Science and Technology  
kraemer@ntnu.no

Anders Eivind Braten  
NTNU – Norwegian University of Science and Technology  
anders.e.braten@ntnu.no

Xiufeng Liu  
DTU – Technical University of Denmark  
xiuli@dtu.dk

Fredrik Anthonisen  
AIA Science AS  
Trondheim, Norway  
fredrik.anthonisen94@gmail.com

Patrick Driscoll  
NTNU – Norwegian University of Science and Technology  
patrick.arthur.driscoll@ntnu.no

John Krogstie  
NTNU – Norwegian University of Science and Technology  
john.krogstie@ntnu.no

ABSTRACT
Cities worldwide aim to reduce their greenhouse gas emissions and improve air quality for their citizens. Therefore, there is a need to implement smart city approaches to monitor, model, and understand local emissions to better guide these actions. We present our approach that deploys a number of low-cost sensors through a wireless Internet of Things (IoT) backbone and is thus capable of collecting high-granular data. Based on a flexible architecture, we built an ecosystem of data management and data analytics including processing, integration, analysis, and visualization as well as decision-support systems for cities to better understand their emissions. Our prototype system has so far been tested in two Scandinavian cities. We present this system and demonstrate how to collect, integrate, analyze, and visualize real-time air quality data.

1 INTRODUCTION
Urban emissions contribute over 60% to global greenhouse gas emissions. Cities aim at reducing their emissions through tailored policy and integration to Smart City approaches. Smart City approaches facilitate easier integration of emission sensing into city systems and fulfill city requirements through novel and low-cost approaches [5, 6, 8, 10, 11].

The overall aim of our project is to fulfill the information needs of cities that need specific data for emission reduction actions by providing complementary on-the-ground emission data for improved understanding and decision making [2]. In short, the need based on future challenges faced by cities will be better and more high-granularity measurements to complement existing official measurement stations.

Some Nordic cities have specific challenges in that they have already implemented a range of climate actions, which means that future impact on a certain class of emissions can only be achieved by a more detailed and granular understanding and analysis of emissions, since many broad measures are already in place. The next step then is to get better insight into more difficult to measure components, also to be able to adapt policy in fast feedback loops and at varying scales. This includes impact assessment of measures ranging from small-scale such as closing down certain streets (and being able to observe spillover and evasion effects in surrounding parts of the city) to large-scale such as changes in public transport or denser urban development.

A high spatial granularity of sensor deployments is obviously not possible with the existing expensive high-quality measurement stations that are often provided nationally. Our approach, in contrast, is to use low-cost sensors to cover a city’s spatial footprint with a much higher sensor density. This enables a trade-off of high number and high granularity of low-cost sensors that can compensate for their relatively lower accuracy.

Existing official measurement stations are equipped with high-quality sensors that cost up to $500,000. Our low-cost approach could provide a very dense coverage of a city with 250 additional sensors for the price of one additional station by using sensor units of around $2,000 each. For ease of installation, this requires standalone sensor units that do not need cabling for electricity or connectivity. We achieve that by deploying solar-powered sensor nodes with a wireless data link over the LoRaWAN standard for Smart City IoT applications, which also enables us to quickly scale up the sensor deployment. The approach allows to quickly prototype system components on the hardware and software side for the overall goal of linking the measurement data to the information needs of the cities for emission reduction both for baseline and continuous data collection.

After having built and deployed the general IoT sensor network before [2], we focus here on the integration of data sources and the data analysis infrastructure for Smart City applications.

2 APPROACH
Our approach is to build an ecosystem of relevant tools and methods to better understand city emissions and work with data, such as analytics [9, 12], visualizations [7], and decision support systems [5, 6, 11] around local emission measurements and the integration of external data sources. This is an important aspect of Smart Cities [9], and can also be used as a case study to understand and build similar systems. Our system is piloted in the two cities of Trondheim, Norway, and Vejle, Denmark.

In this paper, we describe key aspects of this ecosystem of data analysis and visualization that strongly relates to challenges and...
requirements of the cities. We further demonstrate the integration and aggregation of data sources for a smart city.

2.1 Architecture

The system architecture and data flow is sketched in Fig. 1, which consists of four components: a city-wide IoT sensor network, cloud-based systems for data collection and storage, integration of external data, and analysis and visualization platforms for stakeholders. The architecture is flexible through an ecosystem approach and accommodates different components for a range of related tasks. Our technology stack follows common concepts for IoT and Smart City systems [10] with project-specific adaptations.

The sensor network is composed of sensor nodes deployed within the city, which measure emissions and air parameters: CO$_2$, NO$_x$, PM$_x$ (particulate matter); temperature, pressure, and humidity. The data is transmitted to the IoT backbone, which forwards collected data to the cloud storage, from where it is available for analysis and visualization, using relevant external data sources. The backbone uses LoRaWAN as a radio-based urban sensor networks through a number of gateways covering the pilot regions [2]. Data forwarding and cloud sensor management was built through the event-driven MQTT communication protocol.

Visualizations and analyses are connected to all stages of the data processing. Examples are network monitoring and early data validation close to the sensors, stream processing on measurement data, up to C&C centers, satellite measurement grounding, integration into GML-based 3D city models, and other forms of mapping and integration that we describe in the following.

2.2 Data Integration

Apart from the direct sensor data, there is a range of municipal and national data sets available as well as other external data sources that need to be included in the data analytics and visualization to support analyses and improve data quality. Table 1 gives an overview of these sources and how they can be utilized. They range from direct measurements of air quality that can be used to validate and calibrate the sensor network to other data sources that help to understand emissions in the context of a city, for example through traffic patterns [12] or integration into city tools and systems.

The sensor network has the usual issues of missing data that is dealt with on a technical monitoring level and being handled by standard methods in the analyses, as well as the aggregation of data from multiple sensor units. More interesting are the challenges posed in the data integration. The sources contain highly heterogeneous data, with different timescales, measurement frequencies, spatial distributions and granularities, measurement technologies, and a complex set of related uncertainties and inaccuracies in the data.

2.3 Network Metadata Analysis and System Status Monitoring

The network, server components, gateways and sensors are subject to transient and permanent failures, which can ultimately result in missing data. Although the later analysis tasks can detect such losses of data, they do not analyze the cause for the error, or prevent further losses. Instead, failures in the system should be detected as quickly as possible, so that data loss is kept at a minimum. We therefore built a monitoring application (the dataport) to monitor the status of all sensors, gateways and the network [3]. It is built with the Akka framework, which facilitates the creation of fault-tolerant applications based on the actor model [4]. Actors are independent, supervised processes that encapsulate data and control logic and communicate via messages. Each device in the real world corresponds to a dedicated actor that acts as its digital twin, which is a virtual model of the sensor or gateway. It keeps track of its state in real-time, monitors all communication and triggers alarms if data is not received as expected. Incoming data contains meta-data that identifies the originating sensor and the gateway from which it was received. In this way, the digital twin for a gateway can detect if a gateway operates as expected.

Faults of a more complex nature, such as decaying sensors, erroneous behavior of sensor nodes, or missing data patterns need specific analysis. For example, a single missing measurement is expected occasionally. Based on the measurement frequency of individual sensors, it takes some cycles to determine a failure with certainty. As sensors nodes can adapt their frequency based on battery levels, a complex model of the sensor node and its status is needed for detection. Actors are organized hierarchically. On higher levels, failures can be grouped so that for example a distinction can be drawn between sensor failures versus a gateway outage that would make a set of sensors invisible.

The dataport also monitors the larger system, such as the The Things Network (TTN) cloud backend and the MQTT connection. If any of the components on the data path from the sensors to the data storage fails, the dataport generates a notification. If the dataport itself fails, it is detected by an external watchdog service, in this case AppBeat. The dataport further drives a visualization of the network itself, shown in Fig. 3, of the structure of digital twins for sensors and gateways, their location, the connections and live data transmission between sensors and gateways. Apart from the practical value of monitoring the network, it is also a useful illustration of the spatial and measurement characteristics.

2.4 Data Analyses and Visualizations

A range of analyses work on the collected data streams as illustrated in Fig. 1 apart from the more operational network analysis. Examples are ongoing data collection and analysis, understanding of patterns, as well as comparison of sensor measurements to air quality measurement stations to ground the network and calibrate the sensors. There are very few official stations; to
Table 1: Examples of external data integration

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Official air quality measurements</td>
<td>NILU data (Norwegian Air Quality Institute)</td>
<td>Ground truth for certain pollution types, grounding and calibrating measurements to high-quality reference stations</td>
</tr>
<tr>
<td>Remote sensing</td>
<td>NASA OCO-2 satellite CO₂ measurements</td>
<td>Ground truth top-down measurements for certain emission types, large-scale coverage, low spatial resolution, coupling to large-scale modeling and validation</td>
</tr>
<tr>
<td>Traffic data</td>
<td>Traffic density from here.com</td>
<td>Estimate traffic emissions by correlating continuous external traffic density to emission measurements</td>
</tr>
<tr>
<td>3D city models</td>
<td>Municipal traffic counts</td>
<td>Validate traffic estimations, but only available for short periods</td>
</tr>
<tr>
<td>National statistics</td>
<td>Municipal 3D model of Vejle</td>
<td>Integration into existing visualization tools. Use of city geometry in future emission modeling</td>
</tr>
<tr>
<td>Other municipal data and tools</td>
<td>GHG emission estimates from national statistics office</td>
<td>Down-scaled national GHG emission data, often with high uncertainties</td>
</tr>
<tr>
<td></td>
<td>GIS, statistics, decision support, etc.</td>
<td>Understanding emissions in the context of the city</td>
</tr>
</tbody>
</table>

Figure 3: Visualization of sensors, gateways, and links

support the grounding and calibration, we have co-located one of our sensor units to the only station in the pilot area. This allows to compare both absolute and relative accuracy and calibrate the local sensor and, through larger-scale correlated trends, the network, but with lower certainty. In connection with the network monitoring, it also allows the identification of outliers and malfunctioning sensors. Main ongoing work is modeling dependencies of NO₂, PM₁₀, and CO₂, especially from transport emissions, which therefore also looks at linking to traffic patterns [12]. We discuss some analytics around this data in the following.

Battery levels depend on the charging of the autonomous sensor units through their solar panels. Charged occurs during daytime, and is affected by weather conditions. It is important to monitor the battery level to keep the nodes running. Fig. 4 shows the battery level as a function of time (left), and the difference in battery-level from previous sent package versus time of day, and where red indicates whether the nodes could have been charged by sunlight since the previous package (right). This allows to estimate battery depletion.

Dynamics of CO₂ emissions and possible links to traffic in the form of a traffic jam factor (from here.com data) is shown in Fig. 5. According to the plots, we can conclude for this sensor location that traffic is not the only factor that accounts for the dynamics of the CO₂ emission as they exhibit different patterns, and have no apparent correlation. In fact, CO₂ emission dynamic is a more complex issue that may be affected by many factors, including traffic, wind speed, temperature, humidity and other weather conditions, as well as daily and seasonal patterns, which we will further investigate in our future work.

3 DEMONSTRATION

In this first full demonstration of the CTT air quality system, we show the architecture and implementations of IoT and analytics...
Attendees can vary system and analysis properties, and observe data showing different pollution levels. We interact with attendance data, and demonstrate the pattern and its correlation to analyze CO₂, which is crucial for understanding the data flow to evaluate the flexibility of the data stream analysis.

Finally, we demonstrate how to generate dashboards to streamline the whole data flow, including segmentation, chaining, and automation. Further, with more data collected, we will be able to tune models for emission distribution and dispersion to overcome some of the issues and provide improved analysis with better models. Integration into decision support systems is a far goal. Urban emission monitoring needs a range of heterogeneous data and we are continuing to build useful urban systems around it.

ACKNOWLEDGEMENTS

This paper is part of the Carbon Track & Trace (CTT) project. Funding for this project was in part provided by the LoCaL Climate-KIC flagship programme. We thank all our project partners and specifically VCS for the 3D city model demo. Parts of this article are based on project deliverables [1].

REFERENCES


