On the need for explanations, visualisations and measurements in data-driven air quality monitoring and forecasting

Sigmund Akselsen[1, *], Pontus Edvard Aurdal[6], Kerstin Bach[2, *], João Paulo Costeira[5], Ilias Kalamaras[3], Andreas Jacobsen Lepperød[2], Pedro Lima[5], Ieva Martinkenaite[1], Ole Jakob Mengshoel[2], Arne Munch-Ellingsen[1], Hai Thanh Nguyen[1], Dimitrios Tzovaras[3], Tiago Veiga[5], Konstantinos Votis[3], Leendert Wienhofen[4], Weiqing Zhang[1], Pinar Øzturk[2]

[1] Telenor, Snarøyveien 30, 1360 Fornebu, Norway
[2] Norwegian University of Science and Technology, 7491 Trondheim, Norway
[3] Center For Research and Technology / Information Technologies Institute, GR 57001 Thermi, Thessaloniki, Greece
[4] Trondheim Municipality, 7004 Trondheim, Norway
[5] ISTR, Instituto Superior Técnico, 1049-001 Lisbon, Portugal
[6] UiT - The Arctic University of Norway, 9037 Tromsø, Norway

* Corresponding authors: sigmund.akselsen@telenor.com, kerstin.bach@ntnu.no; Authors are listed in alphabetical order.

Keywords: Air Quality Monitoring and Prediction, Explainable AI, Physical AI, IoT, Time Series Data, Human-Centered AI Systems

Motivation

Air pollution is a growing problem due to industrialisation and urbanisation. Emissions and concentrations have increased worldwide. According to the Air Quality in Europe - 2018 report [1]: Air pollution continues to have significant impact on the health of Europeans, particularly in urban areas. It also has considerable economic impacts, cutting lives short, increasing medical costs and reducing productivity through working days lost across the economy. Europe's most serious pollutants in terms of harm to human healthcare are PM, NO_2 and ground-level O_3. Air pollution also damages vegetation and ecosystems. It has environmental impacts, which affect vegetation and fauna directly, as well as the quality of water and soil, and the ecosystems they support.

Monitoring and forecasting of air quality is challenging because of limited reliability of data and complexity of the models. The main approaches to forecasting are classical dispersion models and data-driven models. Data-driven models are increasing in importance as more data is becoming available, in real-time, but with varying data quality. Several air quality forecasting services are appearing, both from official entities like the Norwegian Environment Agency[1], the Norwegian Institute for Air Research[2] crowdsourced initiatives like HackAir[3] and

[1] https://luftkvalitet.miljostatus.no/
[2] http://luftkvalitet.info/home.aspx
[3] https://www.hackair.eu/

Luftdaten[4]. The services provide information in various formats. One example is the Real-time Air Quality Index Visual Map for Europe[5] provided by the World Air Quality Index project[6]. Data-driven models for air quality forecasting have shown good results as compared to forecasts based on classical dispersion models [2, 3]. They have a potential to improve current forecasts but are limited in explanatory power for upcoming situations.

Data-Driven Air Quality Monitoring and Forecasting: The AI4IoT Pilot

Air quality monitoring and forecasting is one of eight pilots of the AI4EU project[7], the European Union's landmark Artificial Intelligence project, which seeks to develop a European AI ecosystem, bringing together the knowledge, algorithms, tools and resources available, and making that a compelling solution for users. The objectives of the AI4IoT pilot are to develop [4]: i) a better understanding of AI capabilities for IoT, and ii) PoCs (Proof of Concepts) for more precise, real-time measurements, estimations and predictions of air pollution and data-driven decision tools.

Our AI4IoT pilot aims to show that applying AI methods can utilize the high variety and velocity of data within the complex processes of forecasting air quality. Moreover, such methods can incorporate real-time data made available from different sensors, ranging from high-end sensors run by scientists to low-cost sensor devices run by citizens. These methods may also involve large-scale data analysis to provide accurate air quality forecasting, due to the variety of different real-time data sources. Data veracity is another important issue. IoT devices that monitor air quality may provide misleading data caused by failures or malfunctioning, thus standardized data preprocessing is needed to improve data quality. This mash-up of data provides new opportunities for the development of tailored, personalised services.

The pilot is run in Trondheim, Norway, a city that has around 200 000 inhabitants. The air quality in Trondheim is typically good. However, there is high variation, and days of severe pollution can occur, especially in the winter months. The status for the pilot as of July 2019 is that several machine learning models (exploring a multivariate time series approach to modelling and forecasting pollution) have been trained with publicly available data from the Trondheim area. These data include *pollutants* from Norwegian Environment Agency[8] measured by industrial sensors, *weather* data from the Norwegian Meteorological Institute[9], and *traffic* data from the Norwegian Road Authorities[10] counted by inductive loops.

These data-driven models show promising results when it comes to the ability of forecasting severe conditions for a selection of pollutants (PM2.5, PM10, and NO_2) 24 and 48 hours

[4] https://luftdaten.info/en/home-en/
[5] http://aqicn.org/map/europe/
[6] http://aqicn.org/contact/
[7] https://www.ai4eu.eu/
[8] https://api.nilu.no/
[9] https://frost.met.no/
[10] https://www.vegvesen.no/trafikkdata

ahead at their respective location. For the evaluation, we have used data from three air quality measurement stations located within the Trondheim city limits and compared the data-driven models' predictions with the official forecast service provided by the Norwegian Environment Agency[11]. The data-driven models seem to have the potential to complement the classical dispersion models that the official forecast is based upon [5].

Encouraged by these results, we have studied how data quality and services can be improved by combining pollution data captured by low-cost IoT devices (micro-sensors) with other data sources. Initial tests of the quality of data collected from pollutant micro-sensors placed in fixed locations and on mobile vehicles have provided mixed results. In some instances, the scales and trends of these measurements have corresponded closely to data from more expensive industrial sensors, while in other instances this has not been the case. It is still an open question whether the data can be used to enhance the forecasts [6]. An overview of low-cost sensors and systems for air quality monitoring has been provided by the Norwegian Institute for Air Research[12].

AI4IoT for Explainable and Physical AI methods

Our work on the AI4IoT pilot contributes to further development of explainable and physical AI methods. Examples of hybrid (i.e., both data-driven and physics-based) and explainable methods can be found elsewhere in earth science, for example in models to predict soil moisture from data about current soil moisture conditions and rainfall forecast [7]. These improved soil moisture prediction models fulfil three criteria: they are data-driven, accurate and explainable. Here, “explainable” means that the model and its parameters can easily be understood from the earth science perspective, and clearly relate to or directly map to soil or hydrological properties.

Knowledge from physical models such as *relevant patterns* can help estimating or classifying the pollution levels. When incorporated, such methods can generate *advice for decision support systems*. In a «sense-understand-plan-act» view of Physical AI systems that interact with their environment, the AI4IoT pilot aims at closing the loop from collecting air quality data (sense), interpreting the situations (understand), considering various actions (plan) to eventually executing them (act).

We have identified the following research challenges from the AI4IoT pilot for Explainable AI and in Physical AI [8]:

(1) Collection and monitoring of large amounts of data from low-cost IoT networked devices, and ultimately provide recommendations and decision support to organizations is challenging.
(2) **IoT devices** may provide misleading data due to failures or malfunctioning. Explainable AI methods can *offer insights on which features (and hence sensors) are contributing to a decision and explain the rationale of a decision, leading to more robust and manageable IoT systems*.

[11] https://luftkvalitet.miljostatus.no/
[12] https://www.nilu.no/en/research/urban-air-quality/low-cost-sensors-for-monitoring-air-quality/

(3) **AI techniques** can help make existing pollution models more precise, and help citizens take more informed decisions or act to meet global or local requirements.
(4) **Data visualisation** can assist in the presentation of large amounts of available data to the decision maker, but also in the interpretation of trained models and the ways they arrive at their decisions.

Two of the most interesting and relevant issues from the point of view of Physical AI, following a «sense-understand-plan-act» view are:

(5) **Sense.** Obtain reasonably accurate forecasts from a network of mobile sensors, less precise and reliable but larger in number.
(6) **Understand**. Improve data quality by actively placing mobile sensors (or directing them, if they are autonomous) in more informative paths or locations.

We will focus on research challenges tied to missing data and drifting sensors, probabilistic models accounting for sensors' varying precision, different approaches to data visualisations and addressing the needs for explanations. Data visualisation through interactive and reactive visual analytics workflows can assist both in the summarization and understanding of large amounts of input data, and in the interpretation of the trained AI models, towards more informative decision making. We aim to discuss in more detail the topic of AI4IoT visualisation in an upcoming conference paper [9].

The long term objective of the AI4IoT pilot is to improve the state-of-the-art decision making on air quality. Therefore we will benchmark the data collected through IoT devices with deployed installations focusing on precision and granularity of the predictions. Together with our collaborator, Trondheim municipality, we plan to validate our results by developing prototypes and evaluate our results in the form of case studies.

Acknowledgements. This work is part of the AI4EU project (https://www.ai4eu.eu/) which has received funding from the European Union's Horizon 2020 research and innovation programme, under the Grant Agreement No 825619.

References

1. Air quality in Europe — 2018 report, European Environment Agency, EEA Report No 12/2018, ISSN 1977-8449, https://www.eea.europa.eu/publications/air-quality-in-europe-2018
2. Yi, X., Zhang, J., Wang, Z., Li, T., Zheng, Y., 2018. Deep distributed fusion network for air quality prediction. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, pp. 965–973.
3. Cassano, F., Casale, A., Regina, P., Spadafina, L. and Sekulic, P. A Recurrent Neural Network approach to improve the Air Quality Index prediction. In P. Novais et al. (Eds.): ISAmI 2019, AISC 1006, pp. 36–44, 2020. https://doi.org/10.1007/978-3-030-24097-4_5
4. Akselsen, S., Bach, K. and Kalamaras, I. Pilot AI4IoT: Air Quality Monitoring. Presentation given at AI4EU Joint WP5/6/7/9 Workshop AI, Paris, France: May 2-3, 2019, http://ai4eu.aass.oru.se/M4-Workshop/
5. Lepperød, A. J. Air quality prediction with machine learning. Master thesis, NTNU. June 2019.
6. Lidia Morawska, Phong K Thai, Xiaoting Liu, Akwasi Asumadu-Sakyi, Godwin Ayoko, Alena Bartonova, Andrea Bedini, Fahe Chai, Bryce Christensen, Matthew Dunbabin, Jian Gao, Gayle SW Hagler, Rohan Jayaratne, Prashant Kumar, Alexis KH Lau, Peter KK Louie, Mandana Mazaheri, Zhi Ning, Nunzio Motta, Ben Mullins, Md Mahmudur Rahman, Zoran Ristovski, Mahnaz Shafiei, Dian Tjondronegoro, Dane Westerdahl, Ron Williams. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? Environment International, Volume 116, July 2018, Pages 286-299. https://doi.org/10.1016/j.envint.2018.04.018

7. Basak, A. Mengshoel, O. J., Schmidt, K. and Kulkarni, C. Wetting and Drying of Soil: From Data to Understandable Models for Prediction. In Proc. IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA-2018), 2018, pp. 303-312.
8. Alessandro Saffiotti (main editor). AI4EU Deliverable D7.1 Outcomes from the Strategic Orientation Workshop. June 30, 2019.
9. Ilias Kalamaras, Ioannis Xygonakis, Konstantinos Glykos, Sigmund Akselsen, Arne Munch-Ellingsen, Hai Thanh Nguyen, Andreas Jacobsen Lepperød, Kerstin Bach, Konstantinos Votis and Dimitrios Tzovaras. Visual analytics for exploring air quality data in an AI-enhanced IoT environment, 11th International ACM Conference on Management of Digital EcoSystems (MEDES'19), submitted.