TOWARDS LARGE-SCALE AIR QUALITY MONITORING

Petteri Nurmi
petteri.nurmi@helsinki.fi

MOTIVATION

- Over 4.2 million deaths per year linked with air pollution (11.6% of all deaths)
- 2-5% of GDP spent on air quality related diseases globally
- Only 1 in 10 people breathe air that is safe

STATE-OF-THE-ART

- Professional-grade measurement towers
  - Provide highly accurate environmental data
  - But very expensive and laborious (to maintain), typical cost over a ($/€/£)
  - Large and bulky ➔ restricted to fixed locations
- Industrial sensors
  - Cost in tens of thousands
  - Partially mobile
  - Lower measurement accuracy
- Low-cost sensors
  - Cost < 1000 $/€/£
  - Highly inaccurate

Low-cost

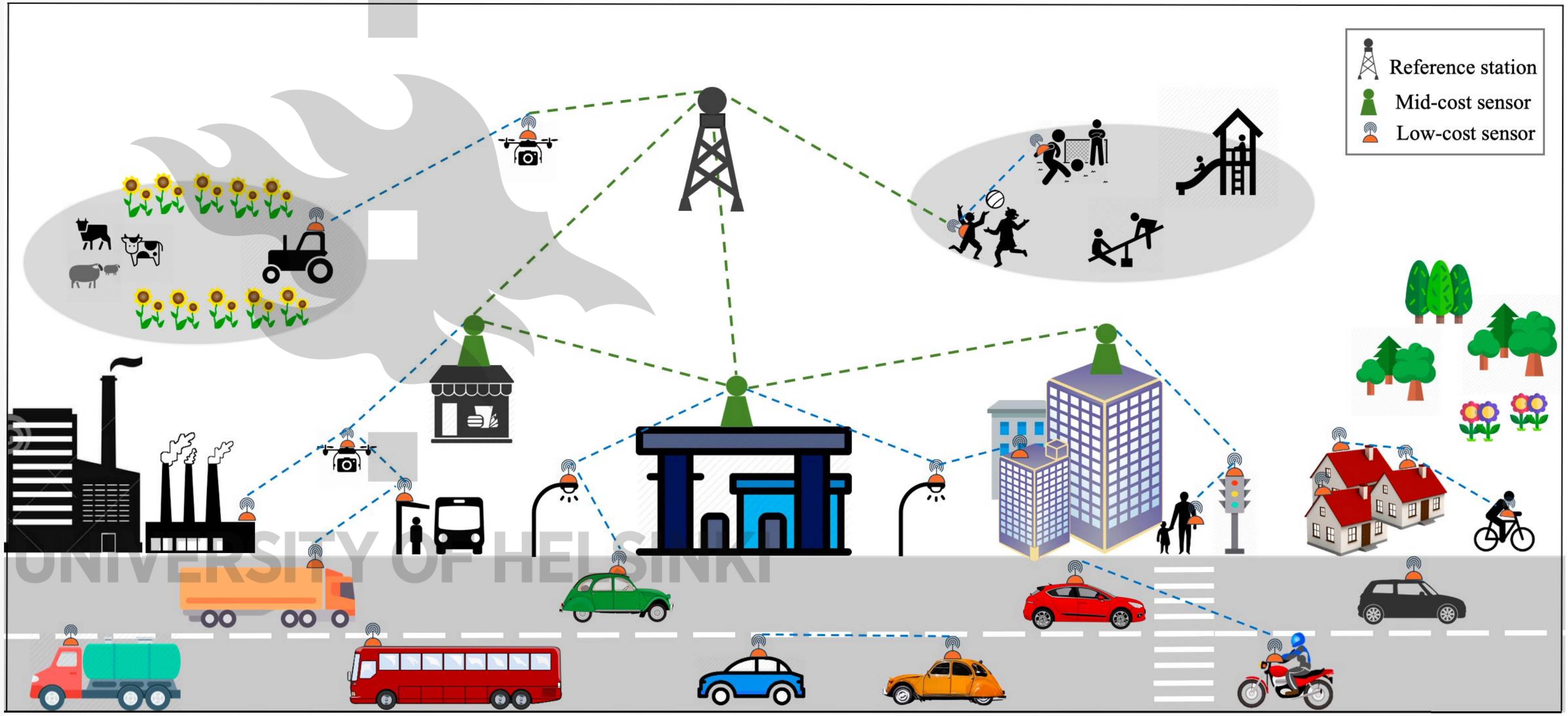
Professional

Industrial

Source: N. H. Motlagh, E. Lagerspetz, P. Nurmi, X, Li, S. Varjonen, J. Mineraud, M. Siekkinen, A. Rebeiro-Hargrave, T. Hussein, T. Petäjä, M. Kulmala, S. Tarkoma, "Toward massive scale air quality monitoring", IEEE Communications Magazine, 58(2), pp. 54-59, IEEE, 2020

VISION FOR AIR QUALITY MONITORING

Dense observation networks that combine different types of sensors



Source: N. H. Motlagh, E. Lagerspetz, P. Nurmi, X, Li, S. Varjonen, J. Mineraud, M. Siekkinen, A. Rebeiro-Hargrave, T. Hussein, T. Petäjä, M. Kulmala, S. Tarkoma, "Toward massive scale air quality monitoring", IEEE Communications Magazine, 58(2), pp. 54-59, IEEE, 2020

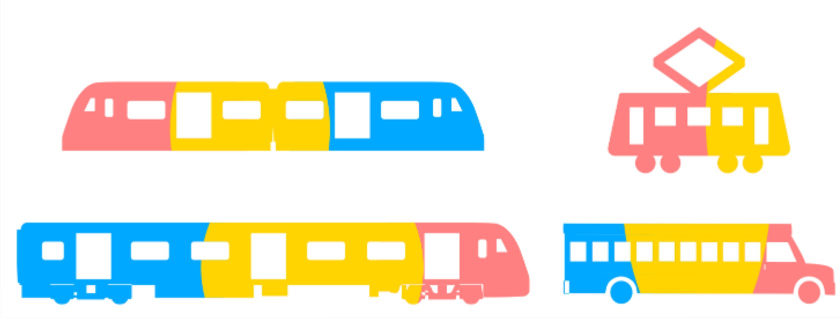
WHY NEW VISION?

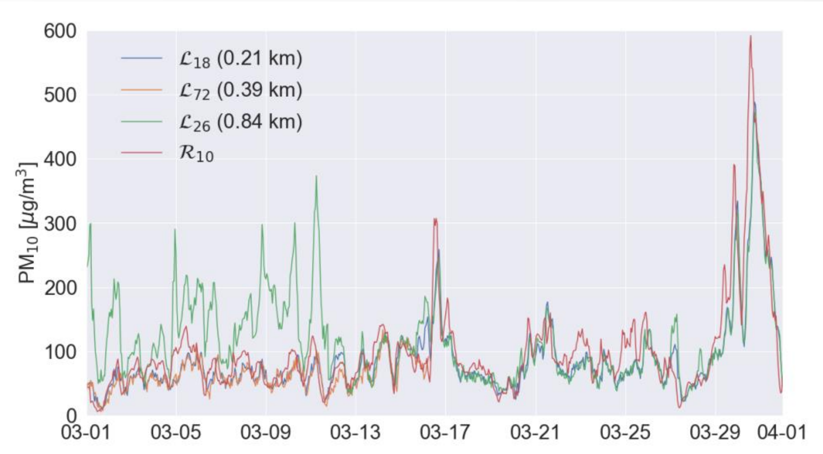
- Professional grade (and industrial) sensors have limited spatial and temporal coverage due to their high cost
  - Analysis necessarily limited to aggregated information instead of providing details of localized pollutant distributions
- But pollutant concentrations can vary drastically even within 30 meter distance ➔ demand for high spatial and temporal resolution
  - Need around 1000 sensors / square mile or tens of thousands of sensors / city district
- Achieving accurate yet dense information only possible by combining different types of sensors!
  - How to ensure sufficient accuracy?
  - How to maintain large-scale deployments?
  - How to deploy and design sensors?

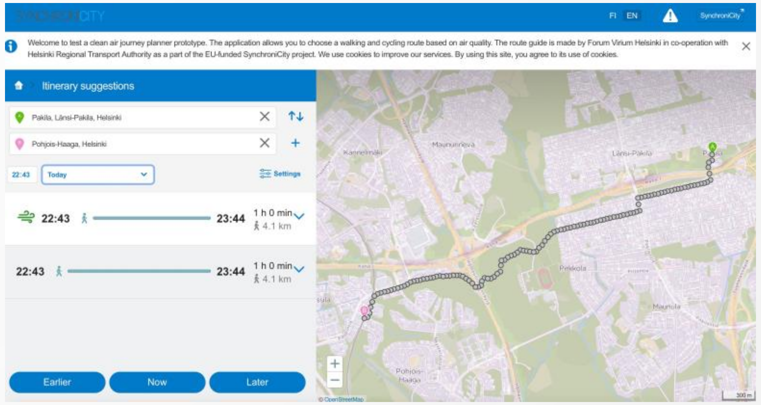
Source: N. H. Motlagh, E. Lagerspetz, P. Nurmi, X, Li, S. Varjonen, J. Mineraud, M. Siekkinen, A. Rebeiro-Hargrave, T. Hussein, T. Petäjä, M. Kulmala, S. Tarkoma, "Toward massive scale air quality monitoring", IEEE Communications Magazine, 58(2), pp. 54-59, IEEE, 2020

APPLICATION AREAS

- High resolution air quality is highly important for emerging applications
  - Localized monitoring of pollutants to identify emission hotspots or other areas of variation
  - Green routing to suggest routes that avoid heavy pollution intake
  - Detection and analysis of pollutants inside public transportation vehicles
- As well as analysing overall impacts of pollutants
  - For example, development of alternative / new air quality indexes







PORTABLE SENSORS AND COVERAGE

- Powerful way for increasing coverage is to equip inhabitants with sensors (or to integrate sensors on their personal devices)
    - Allows estimating personal impact better
    - But data also biased toward personal routines and retention an issue (i.e., people stop using devices)
- Examples on the right sensors developed as part of the MegaSense programme at University of Helsinki
    - Sensors collaboration between industry and academia
    - Collect particulate matter (PM), gaseous pollutants, and diverse environmental variables
    - Can be attached to a bag or other equipment with a simple clip

Source: Motlagh, N. H., Zaidan, M. A., Fung, P. L., Lagerspetz, E., Aula, K., Varjonen, S., Siekkinen, M., Rebeiro-Hargrave, A., Petäjä, T., Matsumi, Y., Kulmala, M., Hussein, T., Nurmi, P., & Tarkoma, S. (2021). Transit pollution exposure monitoring using low-cost wearable sensors. Transportation Research Part D: Transport and Environment, 98, 102981.

WHY COVERAGE MATTERS?

Source: Motlagh, N. H., Zaidan, M. A., Fung, P. L., Lagerspetz, E., Aula, K., Varjonen, S., Siekkinen, M., Rebeiro-Hargrave, A., Petäjä, T., Matsumi, Y., Kulmala, M., Hussein, T., Nurmi, P., & Tarkoma, S. (2021). Transit pollution exposure monitoring using low-cost wearable sensors. Transportation Research Part D: Transport and Environment, 98, 102981.

- Example from monitoring **personal exposure** to pollutants using portable sensors
- Pollutant concentrations vary depending on
  - Mode of transport
  - Route / environment
  - Location inside the vehicle
- Location within vehicle can result in up to 25% increase in daily exposure
  - This compared to dedicated sensors in each vehicle
  - Using scientific measurement stations would result in even coarser estimates

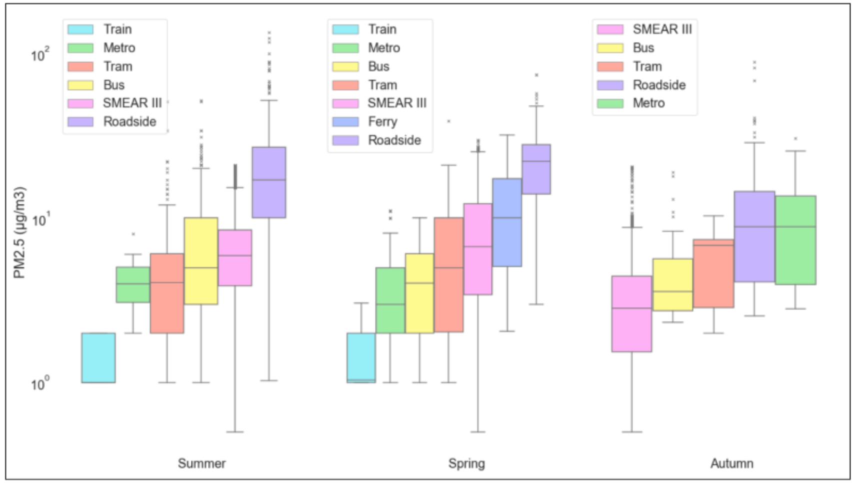
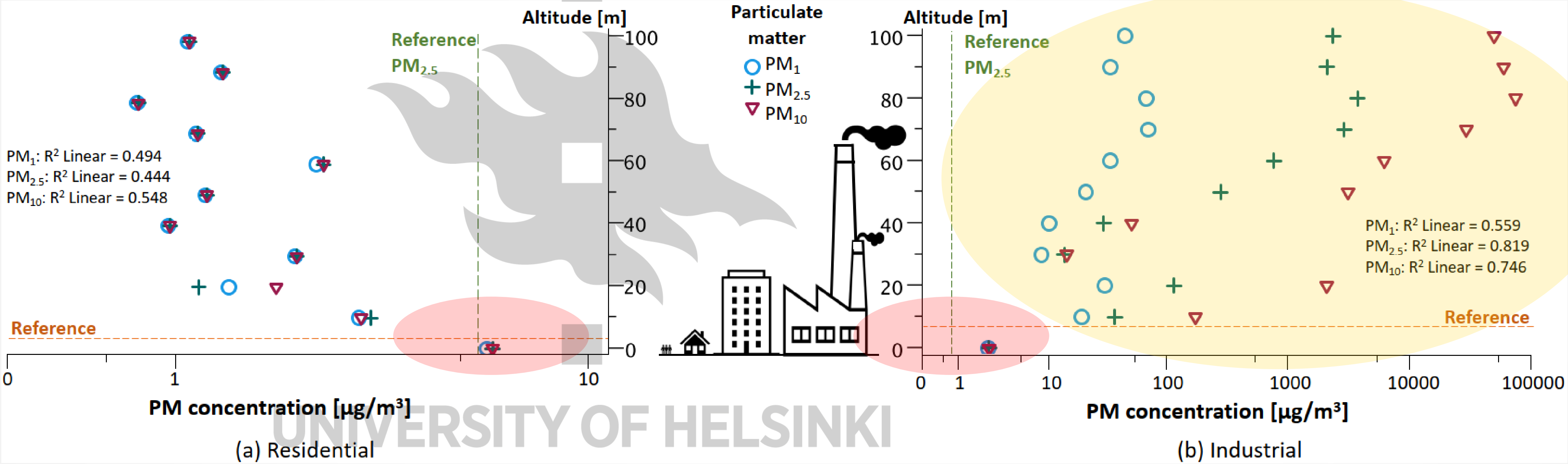


Fig. 4. The $PM_{2.5}$ level in different transportation systems during our experiments.

Transport System	Duration (*min*)	Front (Device 1) $PM_{2.5}$ $\mu g/m^3$	Front (Device 1) DD (μg)
Bus	60	3.43	2.66
Metro	41	8.61	3.88
Tram	44	6.24	3.02
Roadside	82	8.81	7.94

WHY COVERAGE MATTERS?

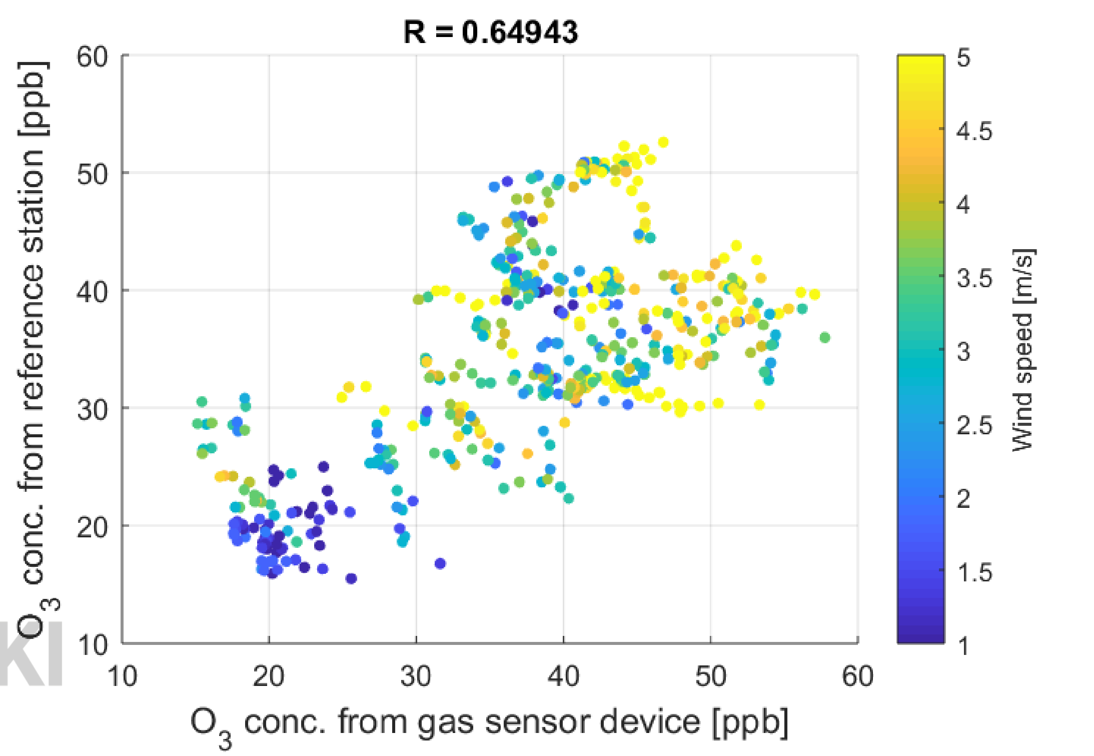


- Localized variations in air quality ➔ need fine-grained coverage to capture these
- Industrial sites result in dispersion of pollutants ➔ need to know their pattern to estimate their impact to other areas

Source: Motlagh, N. H., Irjala, M., Zuniga, A., Lagerspetz, E., Rantala, V., Flores, H., Nurmi, P. & Tarkoma, S. (2022). Toward Blue Skies: City-Scale Air Pollution Monitoring using UAVs. *IEEE Consumer Electronics Magazine*.

SENSOR ACCURACY

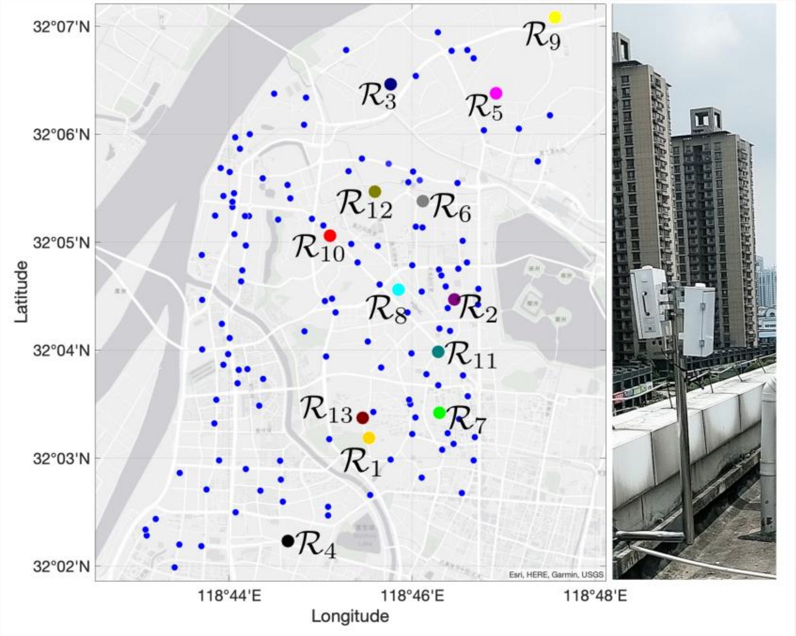
- Coverage can only be increased with sensors that are cheaper and easier to use
- These tend to have lower accuracy than what scientific measurement devices provide
- Many sources of error can influence the measurements
  - Air quality: cross-pollutant sensitivity, weather conditions, drift, characteristics of the location
- Need ways to reduce errors and to understand what kind of errors happen



Source: Lagerspetz, E., Motlagh, N. H., Zaidan, M. A., Fung, P. L., Mineraud, J., Varjonen, S., Siekkinen, M., Nurmi, P., Matsumi, Y., Tarkoma, S. & Hussein, T. (2019, July). Megasense: Feasibility of low-cost sensors for pollution hot-spot detection. In *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)* (Vol. 1, pp. 1083-1090). IEEE.

ACCURACY IN LARGE-SCALE DEPLOYMENTS

- The previous slide characterized accuracy of *individual* sensors compared to reference stations
- What happens when tens or hundreds of sensors are deployed?
- Our recent work explores this question in **dense** deployments of low-cost sensors
  - Deployments in Nanjing, China, cover roughly an area of 55km^2 with 1.1 million inhabitants
  - 126 low-cost sensors
  - 13 reference sensors
  - 6 different types of areas within this region (e.g., roadside, construction, monitoring sites)

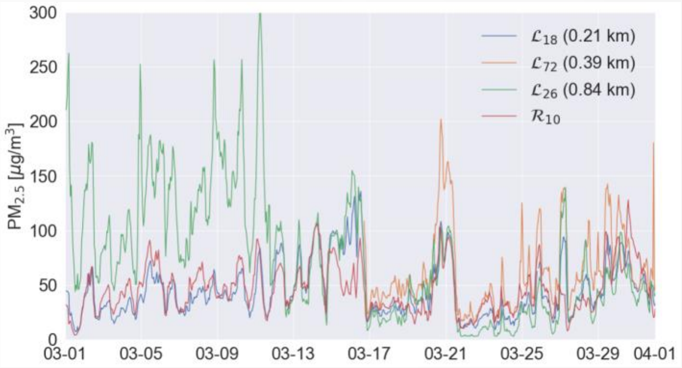


Source: M. A. Zaidan, Y, Xie, N. H. Motlagh, B. Wang, W. Nei, P. Nurmi, S. Tarkoma, T. Petäjä, A. Ding, M. Kulmala, "Dense Air Quality Sensor Networks: Validation, Analysis and Benefits", IEEE Sensors, 2022.

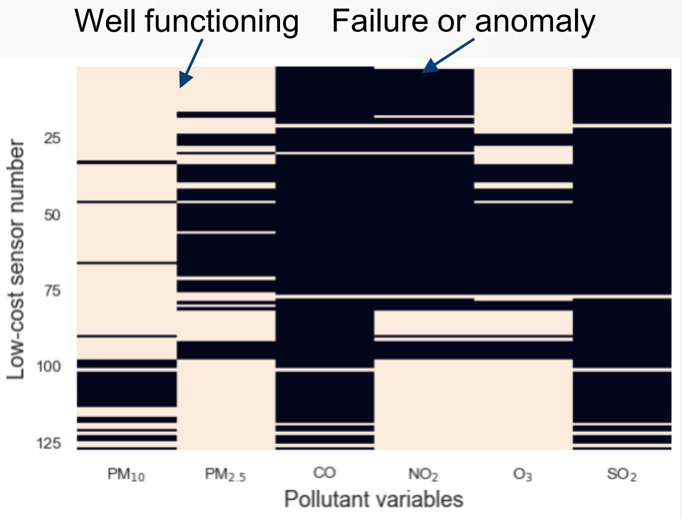
VALIDATION METHODS

Three LCSs validation methods:

- ***Reliability investigation*** to evaluate **all** LCSs observe if they provide reliable measurements as a whole
- ***Accuracy tests*** on **few** of LCSs nearest to the reference stations
  - $PM_{2.5}$ measurements are similar with the PM concentrations measured at R_{10}
- ***Failure and anomaly detection*** on **individual** LCSs to evaluate if they generate reliable air quality
  - Almost all sensors for CO and SO2 are in anomaly or in failure modes ➔ filtered out



Measurements of $PM_{2.5}$ from the 3 nearest LCSs



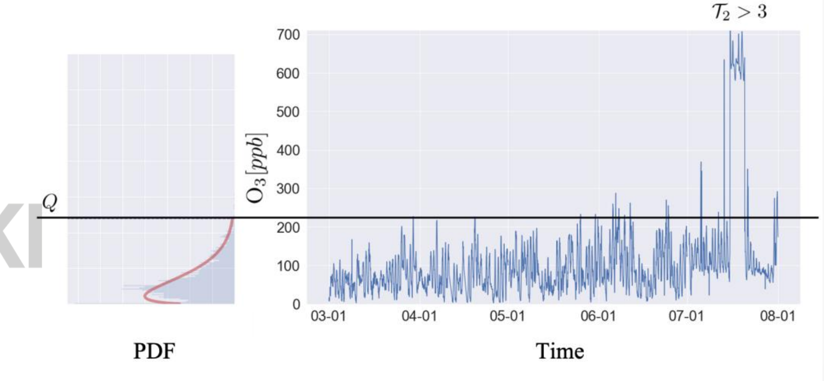
Individual analysis: sensors failure and anomaly

ACCURACY IN LARGE-SCALE DEPLOYMENTS
SELECTED RESULTS

- Median values largely similar for reference sensors (R) and low-cost sensors (L)
- Also variation (MAD) similar
- Mean and variance contain significantly more variation for low-cost sensors ➔ measurements contain outliers
- Anomalies typically continuous periods where something unexpected happens
  - Detected by modelling the distribution of pollutant values at nearest reference station and determining a probability threshold
  - Values exceeding threshold for a long period extremely unlikely and considered outliers
  - Weibull distribution used for sensor measurements, chosen using AIC weights

TABLE IV
STATISTICAL PROPERTIES OF POLLUTANT VARIABLES.

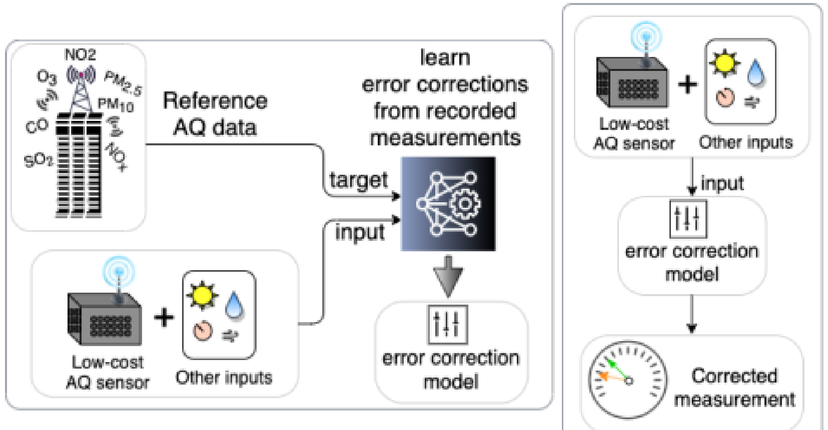
Pollutant	$\mathcal{R}$					$\mathcal{L}$				
Variables	Median	Mean ± CI	Std	MAD	Skewness	Median	Mean ± CI	Std	MAD	Skewness
AQI	51	54.04 ± 4.46	33.70	16.96	4.46	48.89	52.60 ± 19.04	33.54	20.00	4.20
PM_{10} [$\mu g/m^3$]	48	33.03 ± 5.23	48.53	19.97	4.51	46	55.18 ± 21.11	79.87	21.17	72.83
$PM_{2.5}$ [$\mu g/m^3$]	25	19.19 ± 9.95	112.52	11.18	85.30	21.65	26.65 ± 19.10	83.55	11.15	76.90
CO [ppb]	0.81	0.85 ± 0.30	0.38	0.19	9.87	0.4	1.41 ± 18.64	5.88	1.00	7.45
NO_2 [ppb]	24	29.26 ± 9.23	19.48	12.26	1.57	20	22.55 ± 8.85	13.82	9.25	1.5
O_3 [ppb]	56	68.09 ± 13.82	49.16	32.08	1.09	65.21	78.00 ± 30.42	54.37	35.00	2.98
SO_2 [ppb]	8	8.56 ± 1.54	6.79	1.56	43.66	4	4.92 ± 1.48	6.25	1.91	20.98



Toward Large-Scale Air Quality Monitoring
Petteri Nurmi
petteri.nurmi@helsinki.fi
http://pds.cs.helsinki.fi/

IMPROVING ACCURACY: MACHINE LEARNING BASED CALIBRATION

- Most sensor-enabled devices integrate multiple sensors ➔ possible to use different sensors to estimate and mitigate errors
- Machine learning based calibration builds on this idea, learns a correspondence function that can be used to "correct" measurements
- Requires co-locating inaccurate sensors close to a "reference" (or gold standard) measurement instrument to learn the mapping



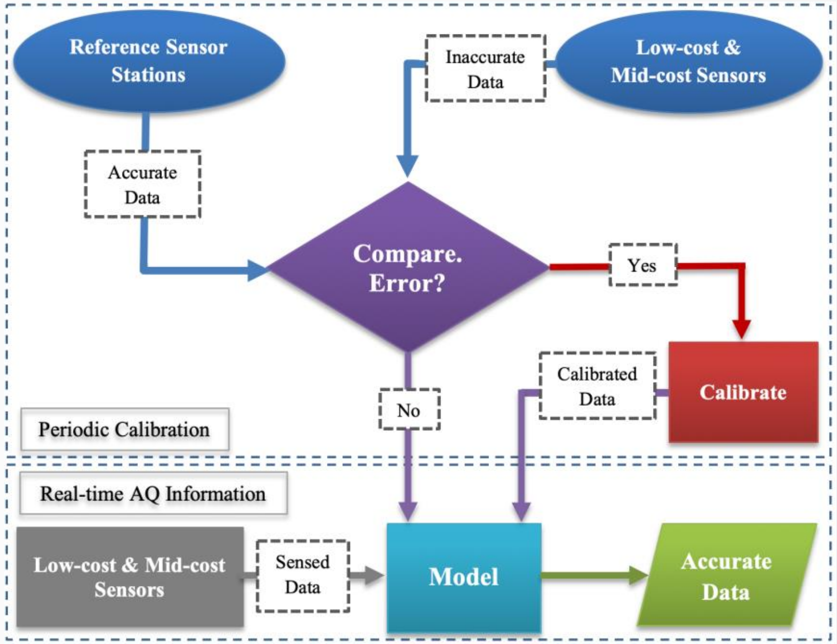
(a) Step 1

(b) Step 2

Source: Aula, K., Lagerspetz, E., Nurmi, P., & Tarkoma, S. (2022). Evaluation of Low-Cost Air Quality Sensor Calibration Models. *ACM Transactions on Sensor Networks (TOSN).*

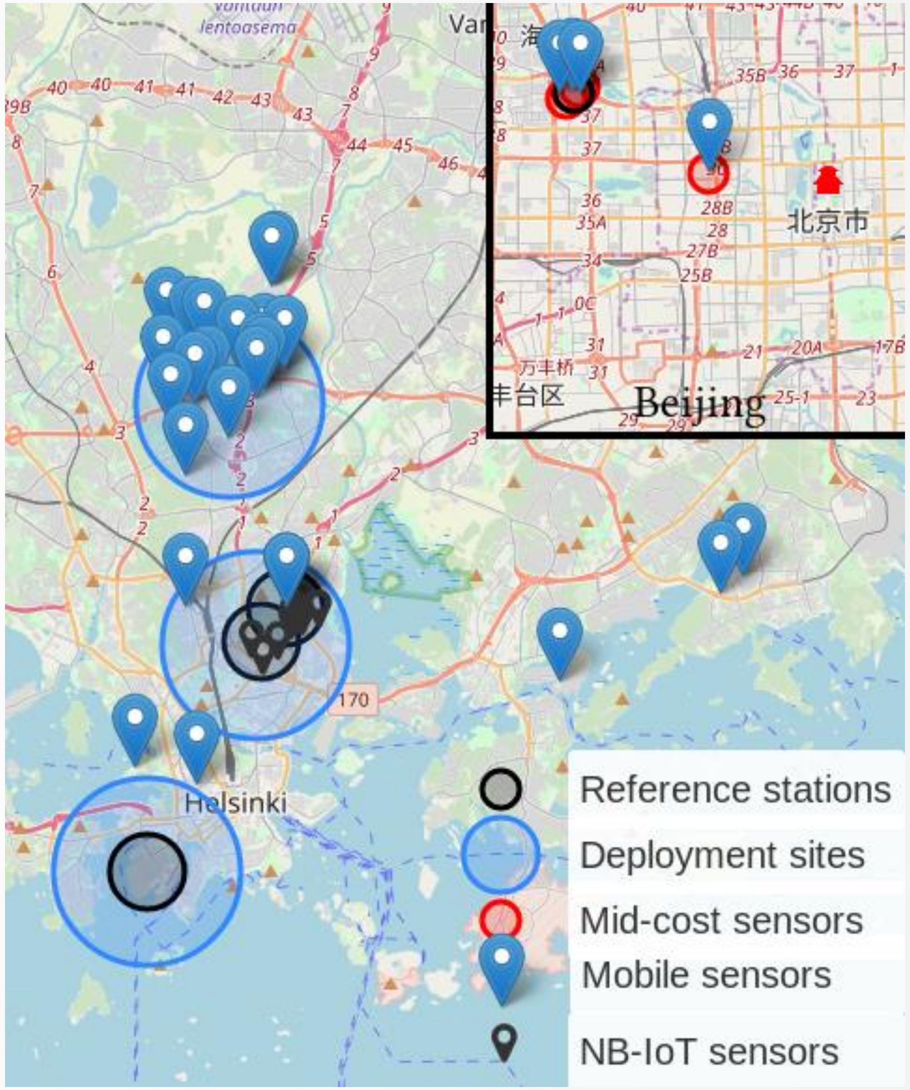
OPPORTUNISTIC SENSOR CALIBRATION

- Periodically co-locating sensors next to each other not feasible in practice ➔ need for alternative methods
  - Preferably allow sensors to operate for long periods of time without manual intervention
  - …co-locating sensors are makes little sense as then could just use the reference station
- ***Opportunistic sensor calibration***
  - collects measurements ***opportunistically*** whenever a device is close to a reference station
  - shares training data from these opportunistic encounters to learn a global calibration model



EXPERIMENTS FOR SENSOR CALIBRATION

- Sensors deployed in several areas of Helsinki and Beijing
- "Triangulation": areas with different characteristics, different spatial and temporal scales for deployments
  1. Shipping district with congested traffic
  2. Residential area away from industry and traffic
  3. Mixed residential and university area close to congested roads
  4. Two deployments at business district areas in Beijing
  - 100+ sensors in total across all areas



OPPORTUNISTIC SENSOR CALIBRATION: RESULTS

1. Machine-learning based calibration using only 2.5 days of data (from a co-located deployment) reduces errors of low-cost sensors by 56%
2. Small amounts of training data sufficient for learning calibration models, quality of measurements more important than quantity
3. Mixing data between industrial and low-cost sensors feasible for calibration, can halve the error

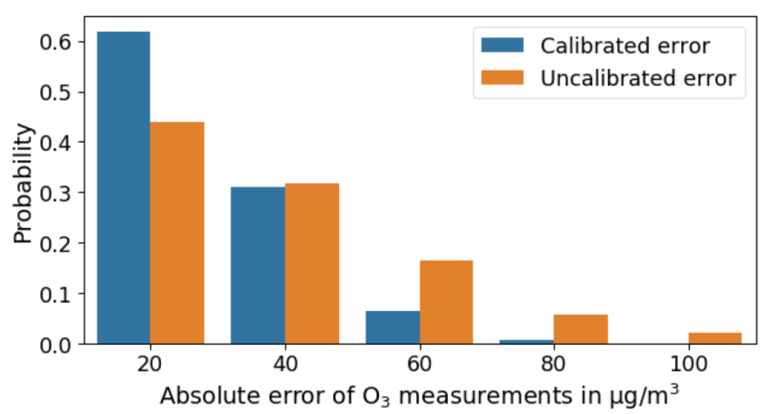
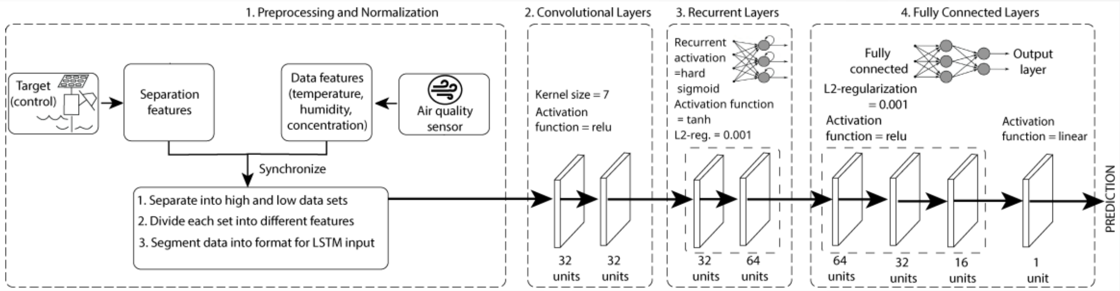


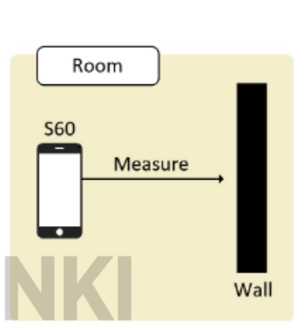
Table I. RESULTS OF FEASIBILITY EVALUATION.

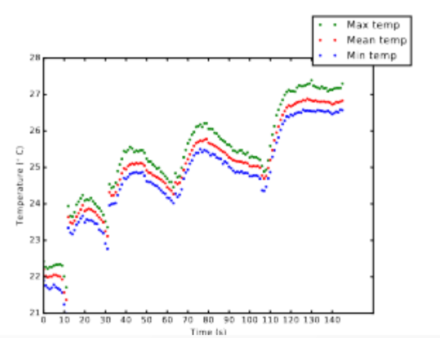
	$PM_{2.5}$		PM_{10}	
train:	high	low	high	low
test:	low	high	low	high
10%	1.38	9.09	2.88	17.84
20%	1.38	8.71	3.23	16.79
30%	1.40	8.67	3.14	15.41
40%	1.38	8.29	3.40	15.04
50%	1.39	8.26	4.79	15.58
60%	1.79	8.48	3.69	15.89
70%	1.79	8.30	4.42	14.84
80%	1.98	7.95	3.86	15.15
90%	3.00	7.64	3.52	15.32
100%	1.38	7.96	4.63	15.88
Mixed	2.33	7.59	6.91	14.42
Orig. err	5.43	10.84	21.34	30.04

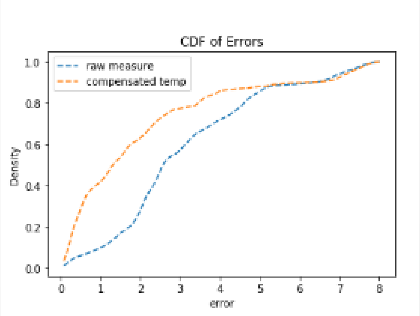
SENSOR CALIBRATION

- Generally best results tend to come with models that combine different structures
- Most urban air quality data manifests linear and non-linear dependencies
- Our work generally uses deep learning models that combine
    - convolutional layers (feature extractors)
    - recurrent layers to capture temporal dependencies
    - fully connected layers to obtain final outputs
- Sensor calibration a **generic problem** with lots of application areas
    - Current work covers sensing for air quality, heart rate, and thermal imaging





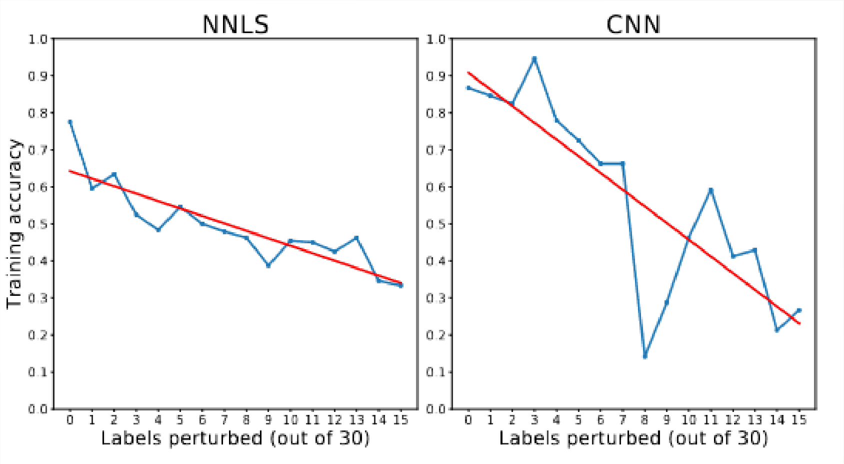




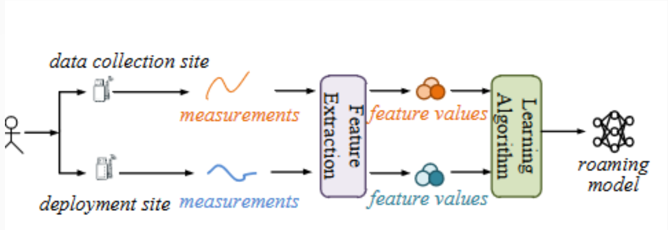
Source: Malmivirta , T , Hamberg , J , Lagerspetz , E , Li , X , Peltonen , E , Flores , H & Nurmi , P 2019 , Hot or Not? Robust and Accurate Continuous Thermal Imaging on FLIR cameras . in 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom) . IEEE , IEEE International Conference on Pervasive Computing and Communications , Kyoto , Japan , 11/03/2019 . https://doi.org/10.1109/PERCOM.2019.8767423

SENSOR CALIBRATION: OVERFITTING

- Model selection non-trivial issue: deep learning models can have excellent performance but tend to overfit on the distribution of the data
  - Figure on the right highlights how changes in data distribution impact deep learning vs. traditional regression methods (using WiFi interference detection as example)
- Transfer learning a potential way to improve performance
  - CrossSense: train separate "expert" models for different environments, select the best matching expert to improve performance

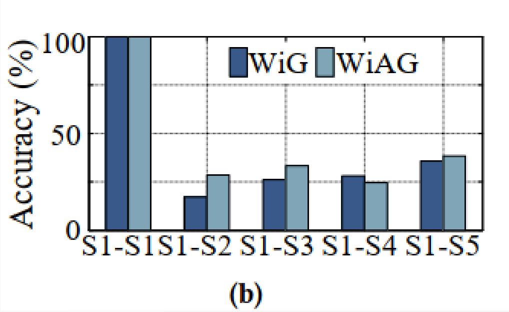


Source: Pulkkinen , T , Nurminen , J K & Nurmi , P 2021 , Understanding WiFi Cross-Technology Interference Detection in the Real World . in 2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS) . IEEE International Conference on Distributed Computing Systems , IEEE, pp. 954-964 https://doi.org/10.1109/ICDCS47774.2020.00061

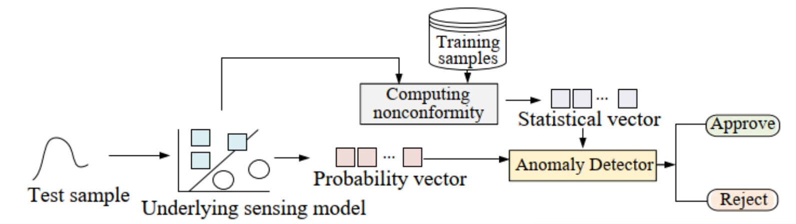


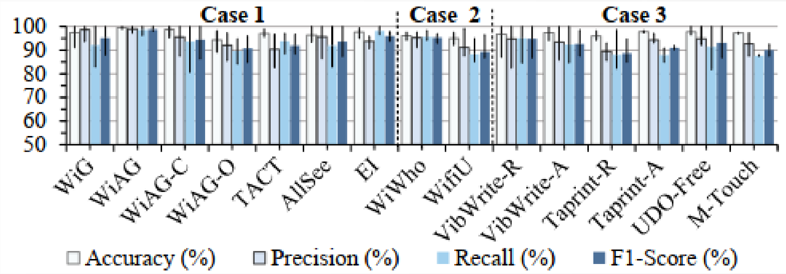
Source: Zhang, J., Tang, Z., Li, M., Fang, D., Nurmi, P., & Wang, Z. (2018, October). CrossSense: Towards cross-site and large-scale WiFi sensing. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking* (pp. 305-320).

SENSOR CALIBRATION: DRIFT



- Machine learning models can lose performance over time as sensors lose accuracy or the environment changes
  - Figure on the right uses WiFi sensing to illustrate how 15cm change in sensor location can result in 75% drop of accuracy
  - Requires re-training ML models and/or feeding new data into the training (e.g., federated learning) but how to detect this?
- **RISE**: system for detecting model drift
  - Examines changes in the output of a ML model (by looking at class probability vector)
  - Compares data to those used during training
  - If either detector rejects a sample ➔ update model
  - 1-2 samples (1.12 on average) needed to retrain model to environmental changes!





Source: Zhai, S., Tang, Z., Nurmi, P., Fang, D., Chen, X., & Wang, Z. (2021, October). RISE: Robust wireless sensing using probabilistic and statistical assessments. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking* (pp. 309-322).

FOR MORE ON LOW-COST AIR QUALITY MONITORING AND CALIBRATION…

…check our survey article in ACM TOSN

Concas , F , Mineraud , J , Lagerspetz , E , Varjonen , S , Liu , X , Puolamäki , K , Nurmi , P, Tarkoma , S, “Low-Cost Outdoor Air Quality Monitoring and Sensor Calibration: A Survey and Critical Analysis” , ACM Transactions on Sensor Networks (TOSN) , vol. 17, no. 2, 20, pp. 1-44 . https://doi.org/10.1145/3446005

Low-Cost Outdoor Air Quality Monitoring and Sensor Calibration: A Survey and Critical Analysis

FRANCESCO CONCAS, University of Helsinki, Finland
JULIEN MINERAUD, University of Helsinki, Finland
EEMIL LAGERSPETZ, University of Helsinki, Finland
SAMU VARJONEN, University of Helsinki, Finland
XIAOLI LIU, University of Helsinki, Finland
KAI PUOLAMÄKI, University of Helsinki, Finland
PETTERI NURMI, University of Helsinki, Finland
SASU TARKOMA, University of Helsinki, Finland

The significance of air pollution and the problems associated with it are fueling deployments of air quality monitoring stations worldwide. The most common approach for air quality monitoring is to rely on environmental monitoring stations, which unfortunately are very expensive both to acquire and to maintain. Hence environmental monitoring stations are typically sparsely deployed, resulting in limited spatial resolution for measurements. Recently, low-cost air quality sensors have emerged as an alternative that can improve the granularity of monitoring. The use of low-cost air quality sensors, however, presents several challenges: they suffer from cross-sensitivities between different ambient pollutants; they can be affected by external factors, such as traffic, weather changes, and human behavior; and their accuracy degrades over time. Periodic *re-calibration* can improve the accuracy of low-cost sensors, particularly with machine-learning-based calibration, which has shown great promise due to its capability to calibrate sensors in-field. In this article, we survey the rapidly growing research landscape of low-cost sensor technologies for air quality monitoring and their calibration using machine learning techniques. We also identify open research challenges and present directions for future research.

CCS Concepts: • **Applied computing** → **Environmental sciences**; • **Hardware** → *Sensor applications and deployments*; • **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*.

Additional Key Words and Phrases: air quality sensors, calibration, low-cost, machine learning, review, survey

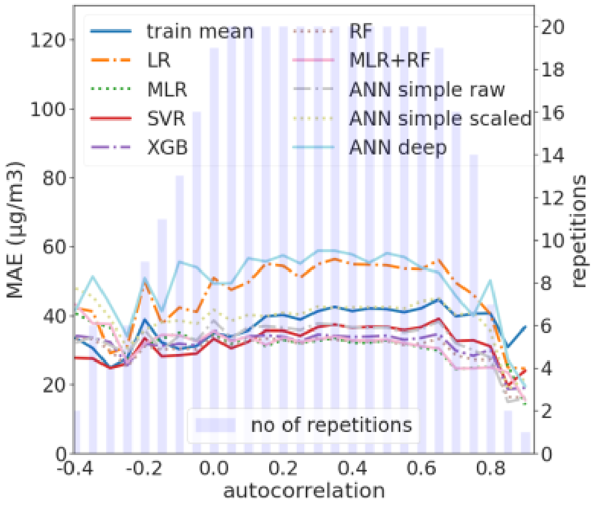
ACM Reference Format:
Francesco Concas, Julien Mineraud, Eemil Lagerspetz, Samu Varjonen, Xiaoli Liu, Kai Puolamäki, Petteri Nurmi, and Sasu Tarkoma. 2021. Low-Cost Outdoor Air Quality Monitoring and Sensor Calibration: A Survey

Authors' addresses: Francesco Concas, francesco.concas@helsinki.fi, University of Helsinki, P.O. 68 (Pietari Kalmin Katu 5), Helsinki, Finland, FI-00014; Julien Mineraud, julien.mineraud@helsinki.fi, University of Helsinki, P.O. 68 (Pietari Kalmin Katu 5), Helsinki, Finland; Eemil Lagerspetz, eemil.lagerspetz@cs.helsinki.fi, University of Helsinki, P.O. 68 (Pietari Kalmin Katu 5), Helsinki, Finland, FI-00014; Samu Varjonen, samu.varjonen@helsinki.fi, University of Helsinki, P.O. 68 (Pietari Kalmin Katu 5), Helsinki, Finland, FI-00014; Xiaoli Liu, xiaoli.liu@helsinki.fi, University of Helsinki, P.O. 68 (Pietari Kalmin Katu 5), Helsinki, Finland, FI-00014; Kai Puolamäki, kai.puolamaki@helsinki.fi, University of Helsinki, P.O. 68 (Pietari Kalmin Katu 5), Helsinki, Finland, FI-00014; Petteri Nurmi, ptnurmi@cs.helsinki.fi, University of Helsinki, P.O. 68 (Pietari Kalmin Katu 5), Helsinki, Finland, FI-00014; Sasu Tarkoma, sasu.tarkoma@helsinki.fi, University of Helsinki, P.O. 68 (Pietari Kalmin Katu 5), Helsinki, Finland, FI-00014.

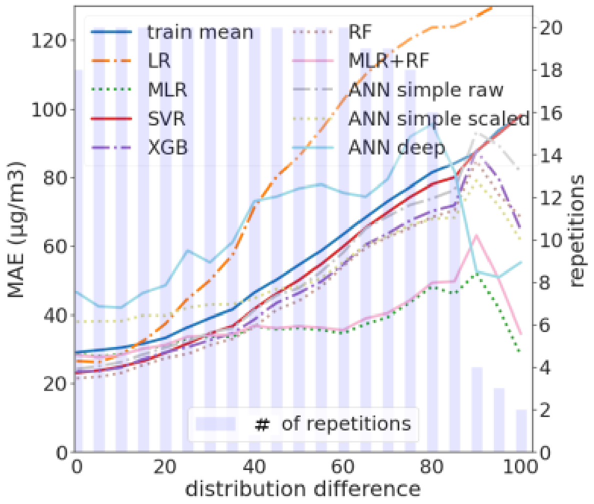
1550-4859/2021/3-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

BEYOND ACCURACY: ROBUSTNESS & CONSISTENCY

- Practical deployments must work on highly different environmental conditions
- Regulations place strict requirements on monitoring that include robustness to variations in environmental conditions
  - E.g., air quality: need to operate robustly against low and high concentrations and in different humidity & temperature
- This needs to be explicitly modelled in the machine learning solutions that operate on data
  - Autocorrelation in data ➔ standard evaluation models incorporate temporal dependencies that give overly optimistic views
  - Distribution of data varies over time ➔ deployment may see data that is not visible during testing / training at all



(a) Segment size=1 week



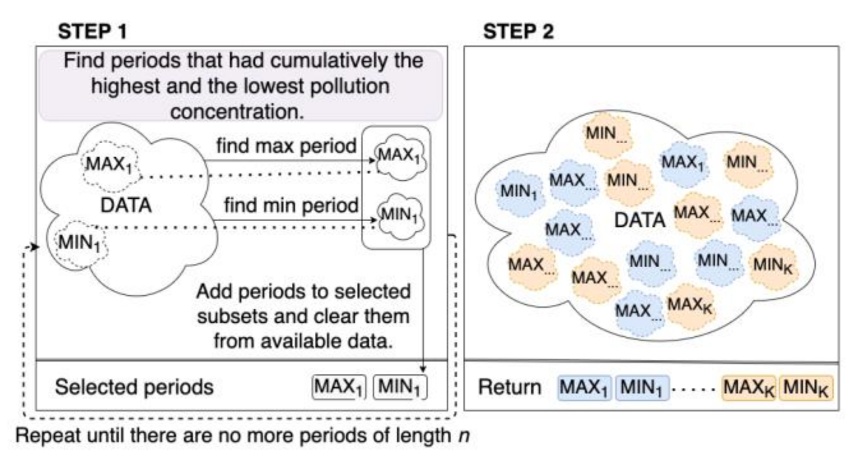
(a) Segment size=1 week

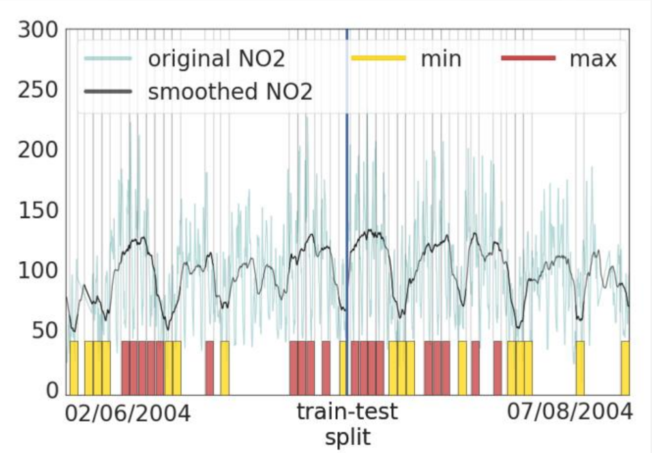
Source: Aula, K., Lagerspetz, E., Nurmi, P., & Tarkoma, S. (2022). Evaluation of Low-Cost Air Quality Sensor Calibration Models. *ACM Transactions on Sensor Networks (TOSN)*.

IMPROVING ROBUSTNESS

Source: Aula, K., Lagerspetz, E., Nurmi, P., & Tarkoma, S. (2022). Evaluation of Low-Cost Air Quality Sensor Calibration Models. *ACM Transactions on Sensor Networks (TOSN)*.

- Robustness of ML models for environmental data can be improved by breaking dependencies
- Diverse data selector a method for creating partitions that help to enforce robustness
  1. Partition data into continuous segments (e.g., a week or a month)
  2. Score segments according to selected criteria (e.g., distributional difference or magnitude of values)
  3. Select partition with highest score and assign it to a pool of measurements
  4. Recompute segment scores and repeat until no more segments available





IMPROVING ROBUSTNESS: SELECTED RESULTS

Source: Aula, K., Lagerspetz, E., Nurmi, P., & Tarkoma, S. (2022). Evaluation of Low-Cost Air Quality Sensor Calibration Models. *ACM Transactions on Sensor Networks (TOSN).*

- Diverse data selector results in data splits that are representative of actual environmental variations
- Standard evaluation methods give overly optimistic views of performance
- Diverse data selector better at assessing performance in practical deployments
- Training with diverse data can significantly improve robustness of machine learning models
- Main effect comes from increasing distributional difference between measurements but also having control over data selection helps

	low		high		diverse		
window	mean	sd	mean	sd	mean	sd	set size
1	78.35	28.53	179.69	49.83	130.27	75.62	720
3	82.87	32.21	172.85	51.87	130.04	70.93	720
5	85.44	34.0	170.86	55.62	131.73	69.22	720
10	90.26	32.81	163.74	58.77	139.92	64.81	720
14	91.91	28.96	161.23	59.12	129.89	62.94	672
15	93.04	31.47	158.96	54.24	128.19	63.25	720
All data	109.63	46.46	109.63	46.46	109.63	46.46	9357

(b) NO_2 (in µg/m^3)

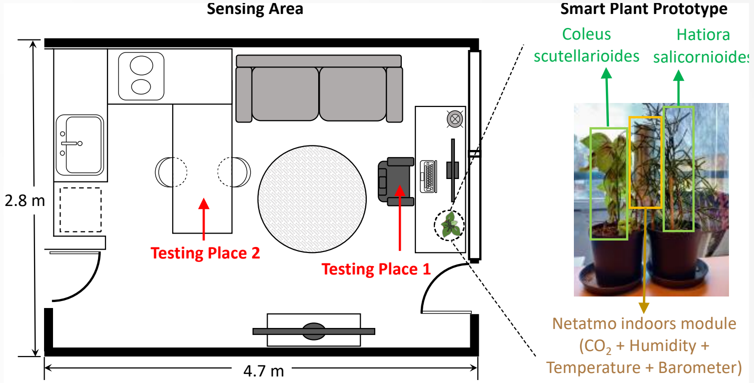
	cont.	low	high	diverse
Training mean	32.77	25.29 (-23%)	88.25 (+169%)	65.0 (+98%)
LR	31.4	27.07 (-14%)	113.8 (+262%)	77.21 (+146%)
MLR	30.05	27.86 (-7%)	**42.77 (+42%)**	43.02 (+43%)
SVR	27.61	25.55 (-7%)	64.97 (+135%)	54.5 (+97%)
XGB	28.6	23.63 (-17%)	73.17 (+156%)	55.21 (+93%)
RF	27.32	**21.74 (-20%)**	66.77 (+144%)	51.83 (+90%)
MLR+RF	**27.31**	22.7 (-17%)	52.61 (+93%)	44.32 (+62%)
ANN simple raw	32.56	28.63 (-12%)	52.92 (+63%)	48.22 (+48%)
ANN simple scaled	30.77	29.01 (-6%)	43.73 (+42%)	44.03 (+43%)
ANN deep	29.35	30.65 (+4%)	43.76 (+49%)	**40.57 (+38%)**
AVG	29.77	26.21 (-12%)	64.28 (+116%)	52.39 (+76%)
4 months (2880 data points w/ 75:25 training-testing split)				

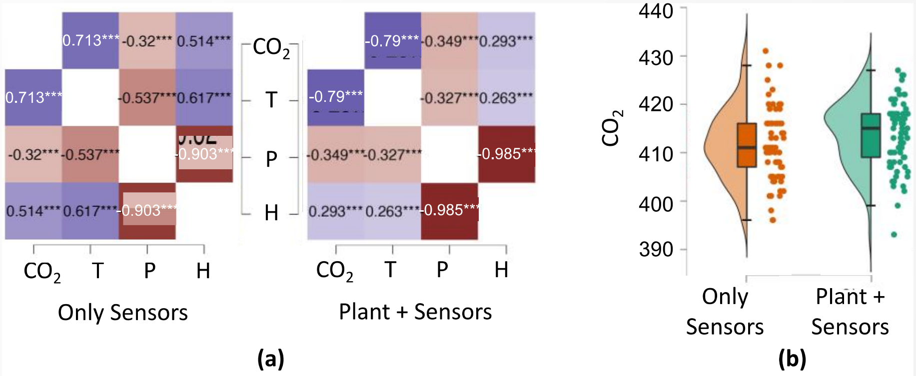
(b) NO_2 (in µg/m^3)

MONITORING INDOOR ENVIRONMENTS USING SMART PLANTS

Source: Zuniga, A., Motlagh, N. H., Flores, H., & Nurmi, P. (2022). Smart Plants: Low-Cost Solution for Monitoring Indoor Environments. IEEE Internet of Things Journal.

- Smart plants offer an easy to deploy and maintain solution for sensing indoors air quality
- Supplement other forms of infrastructure used to monitor, i.e., thermal comfort, workplace productivity
- Besides the sensors for monitoring plants growth, smart plants can integrate CO_2 and temperature sensors to measure environmental conditions
- The sensors in a plant container results in similar values as using a dedicated sensor device in different conditions
- Watering the plants has not a significant effect on the measurements.

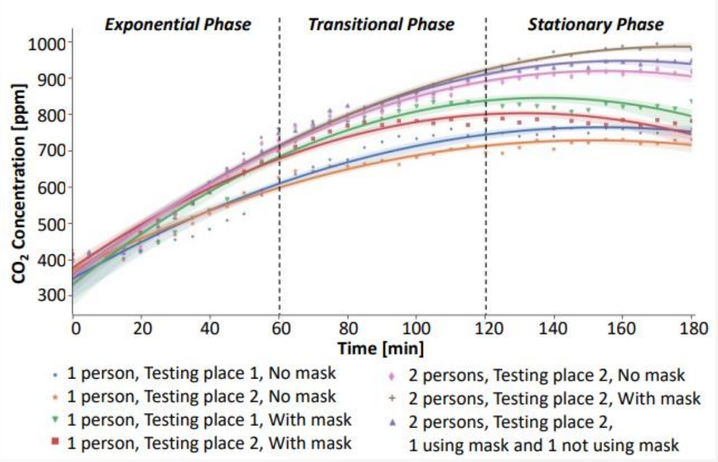




MODELLING FACE MASK USE AND OCCUPATION ESTIMATION

Source: Zuniga, A., Motlagh, N. H., Flores, H., & Nurmi, P. (2022). Smart Plants: Low-Cost Solution for Monitoring Indoor Environments. IEEE Internet of Things Journal.

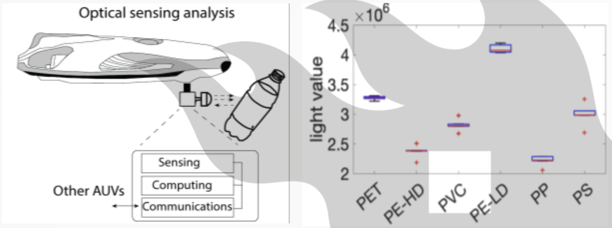
- Smart plants can support coarse-grained classification to distinguish diverse indoor environment conditions
- The analysis focus on the speed of change in the CO_2 levels, as it correlates with the number of people in a space and whether the people use masks or not
- Information about face mask use (or non-use) and occupancy can be obtained using only few minute time windows (5 – 10 min)
- The main source of prediction errors is in the case where mask use is mixed between the occupants
- Using CO_2 levels together with other measurements provided by the smart plant sensors (e.g., temperature) significantly increases the performance of the model



model → predicted	Classifier	Time window size [minutes] 5	10	15	20	25	30
CO_2 → face mask use	RF	65.9	63.8	65.6	63.4	62.9	64.4
	GB	63.7	63.7	63.8	64.3	64.9	65.1
	AB	66.9	67.6	66.5	66.2	64.3	64.1
	Mean	**65.5**	**65.0**	**65.3**	**64.6**	**64.0**	**64.5**
CO_2 → amount of people	RF	81	82.2	82.4	81.3	80.7	80.9
	GB	84.1	84.7	83.1	84.1	84.7	84.3
	AB	89.7	88.8	89.4	88.8	89.4	90.1
	Mean	**84.9**	**85.2**	**85.0**	**84.7**	**84.9**	**85.1**
(CO_2, T) → face mask use	RF	70.9	69.6	70.4	69.7	68.4	67.8
	GB	71.5	72.8	72.2	72.8	70.3	70.3
	AB	69	68.5	67.1	68.1	69.3	69
	Mean	**70.5**	**70.3**	**69.9**	**70.2**	**69.3**	**69.0**
(CO_2, T) → amount of people	RF	85.3	86.9	84.4	85.6	84.1	83.1
	GB	89.4	89.9	89.3	89	88.7	87.8
	AB	92.5	92.4	92.6	92.6	91.9	92.4
	Mean	**89.1**	**89.7**	**88.8**	**89.1**	**88.2**	**87.8**

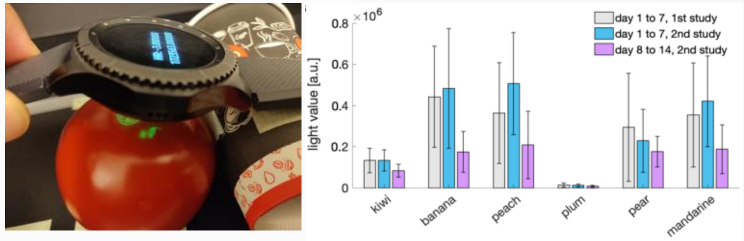
BEYOND AIR QUALITY: SENSING FOR ENVIRONMENTAL SUSTAINABILITY

Marine Plastics



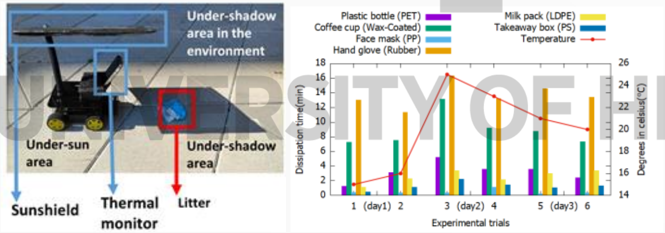
Source: Flores, H., Zuniga, A., Motlagh, N.H., Liyanage, M., Passananti, M., Tarkoma, S., Youssef, M. and Nurmi P., 2020, June. PENGUIN: aquatic plastic pollution sensing using AUVs. In *DroNet@ MobiSys* (pp. 5-1).

Food Waste



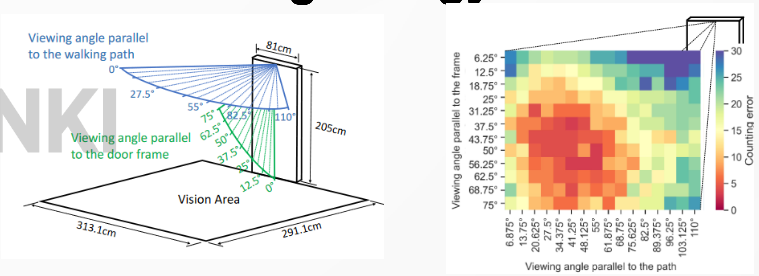
Source: Zuniga, A., Flores, H. and Nurmi, P., 2021. Ripe or Rotten? Low-Cost Produce Quality Estimation Using Reflective Green Light Sensing. IEEE Pervasive Computing, 20(3), pp.60-67.

Waste Recycling



Source: Yin, Z., Olapade, M., Liyanage, M., Dar, F., Zuniga, A., Motlagh, N. H., Su, X., Tarkoma, S., Hui, P., Nurmi, P. & Flores, H. (2022). Toward City-Scale Litter Monitoring Using Autonomous Ground Vehicles. *IEEE Pervasive Computing*.

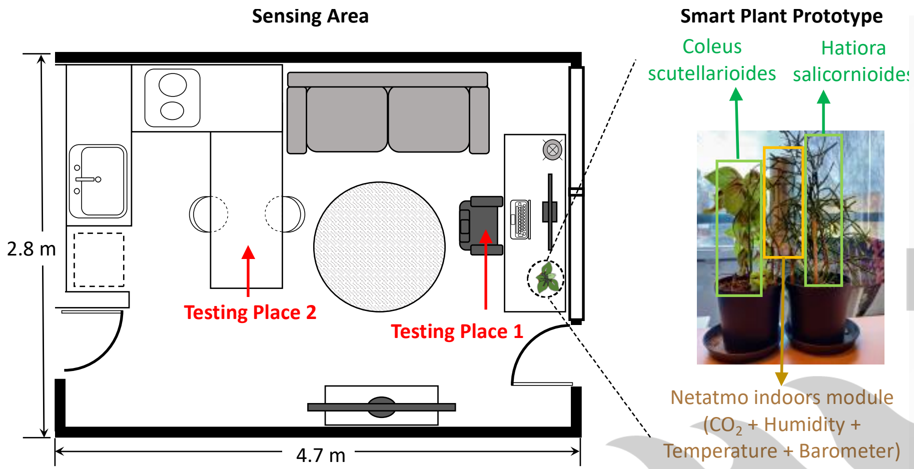
Building Energy Use

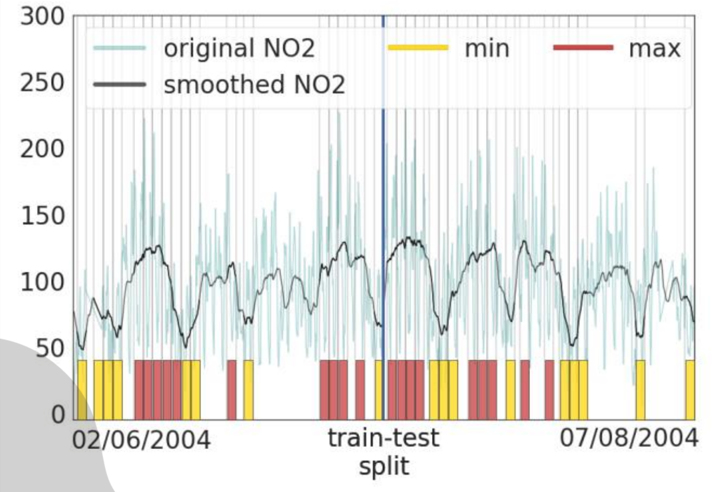


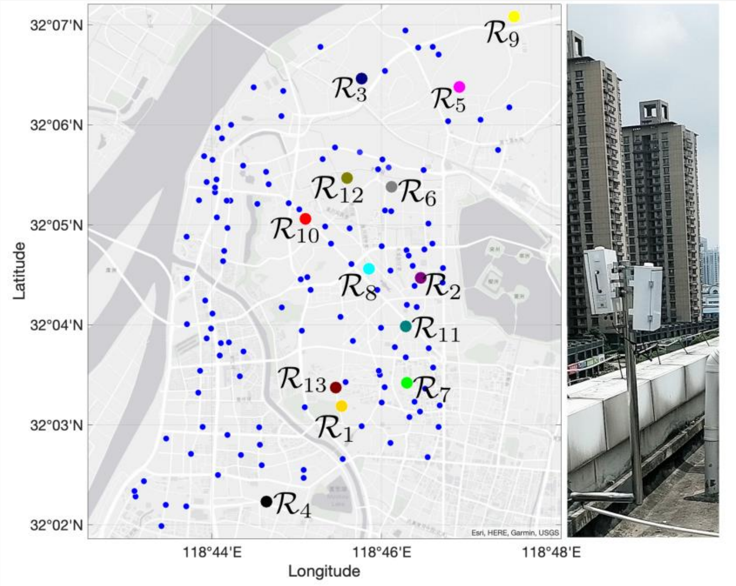
Source: Rinta-Homi, M., Motlagh, N. H., Zuniga, A., Flores, H., & Nurmi, P. (2021). How low can you go? performance trade-offs in low-resolution thermal sensors for occupancy detection: A systematic evaluation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *5*(3), 1-22.

SUMMARISING

- Air quality monitoring a critical challenge for future smart cities as majority of people are subjected to poor breathable air
- Pollution distributions can vary considerably even within small distances ➔ need for dense deployments of sensors
- Dense deployments only possible using inexpensive sensors ➔ need to combine different technologies to ensure high accuracy
- Sensor calibration a potential way to improve accuracy
  - Ensuring model does not overfit critical
  - Transfer learning and drift detection can help
  - Diverse data selector helps to improve model generality and ensure robust performance
- Optimally also monitor indoor air quality, smart plants a potential infrastructure for achieving this

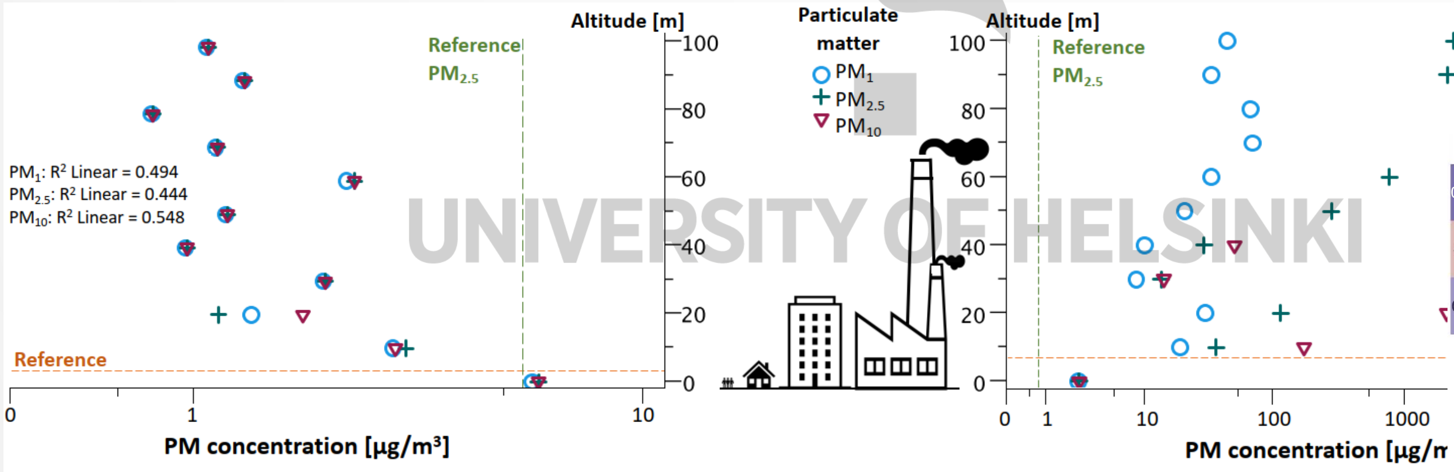






Thank you!

petteri.nurmi@helsinki.fi



(a) Residential

(b) Industrial

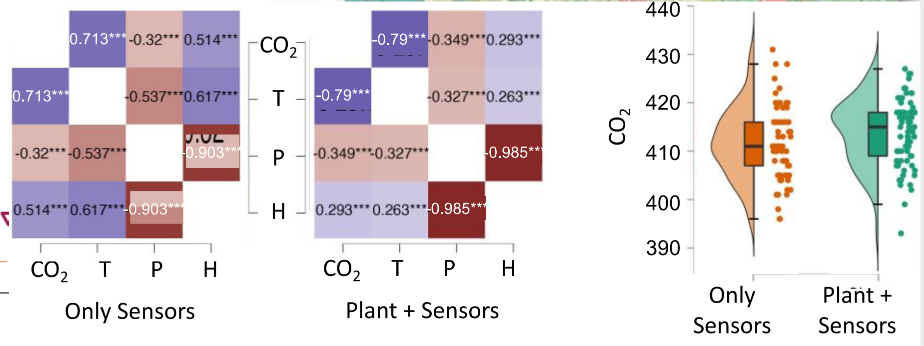


Fig. 2: Comparison between sensing methods. (a) Correlation of air factors, 24-hour sampling, one occupant. (b) CO_2 variation, 6-hour sampling, no occupants.

18.6 – 22.6

MOBISYS 2023 HELSINKI, FINLAND