Low-cost air quality sensors and their use for urban-scale modelling

Philipp Schneider

with contributions from
Hai-Ying Liu, Nuria Castell, Paul Hamer, Matthias Vogt, Franck Dauge, Alena Bartonova

NILU
Norsk institutt for luftforskning
Norwegian Institute for Air Research

Application areas

Two obvious application areas for low-cost AQ sensors with regards to modelling:

1. Comparison of sensor network data with model results
2. Combination of sensor network data with model output, i.e. through data fusion or data assimilation

NILU

Part 1

On the feasibility of using low-cost sensors for model validation

NILU

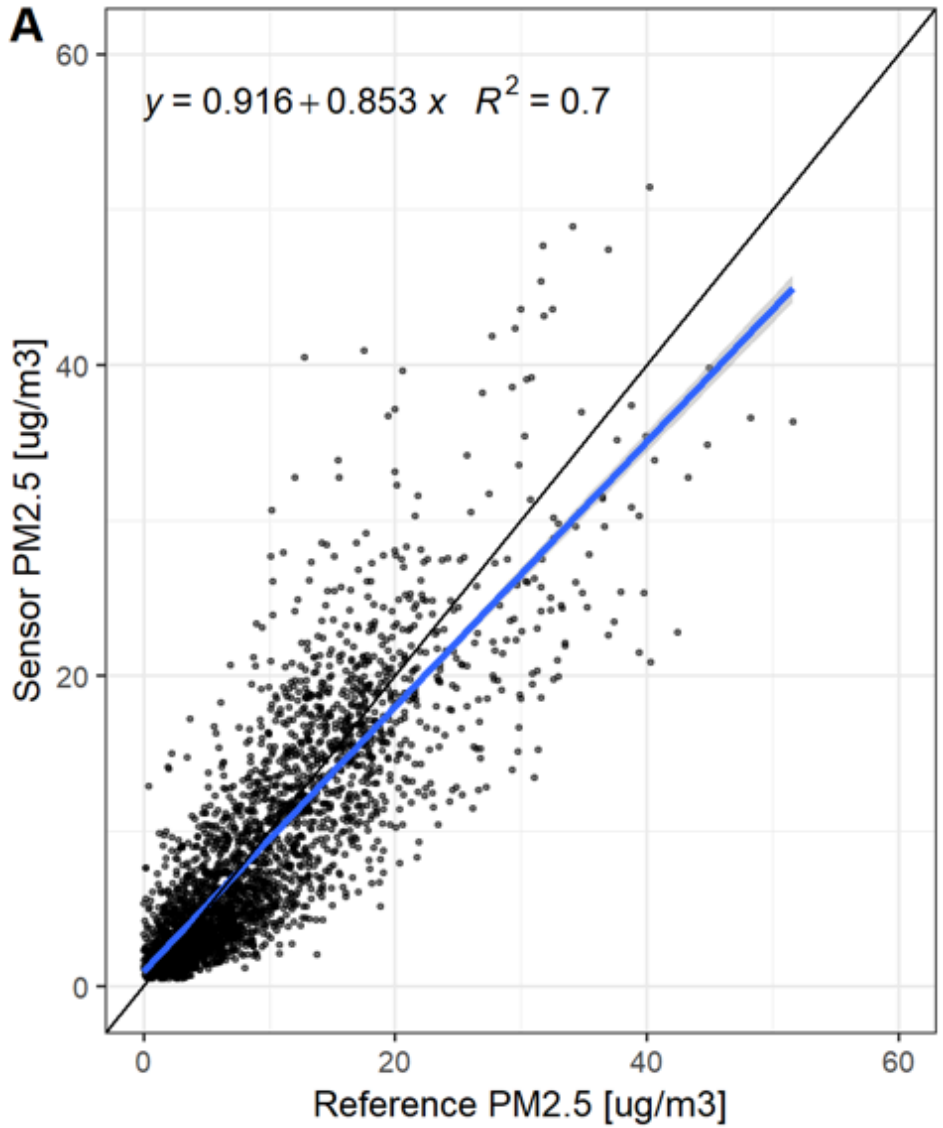
How usable are low-cost sensors these days?

- Using sensor observations for model validation requires mostly one thing: **accuracy**
- In previous years there was often very **questionable performance** of low-cost sensors
- There continues to be **high variability in accuracy** between sensor systems and pollutants
- However, more recently, the **accuracy has improved significantly for $PM_{2.5}$** from low-cost particle sensors (nephelometers)

NILU

Example 1

- SDS011 sensor
- Very cheap (ca.30-50 EUR)
- Widely used
- Consistently good out-of-box performance for $PM_{2.5}$
- Relative effects sensor accuracy for RH > 80%
- PM_{10} less useful at this point due to physical design principles of the sensor

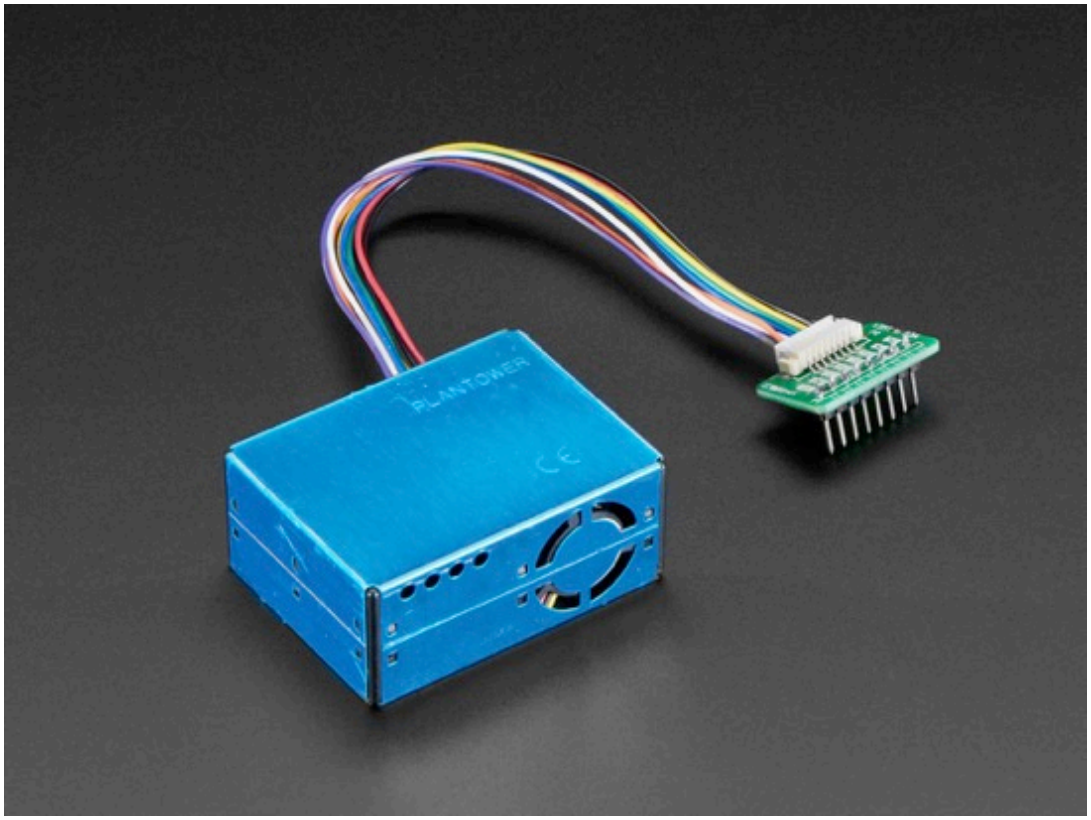


Comparison of hourly PM2.5 from SDS011 against an AQ monitoring station for a 4-month period in Oslo, Norway.

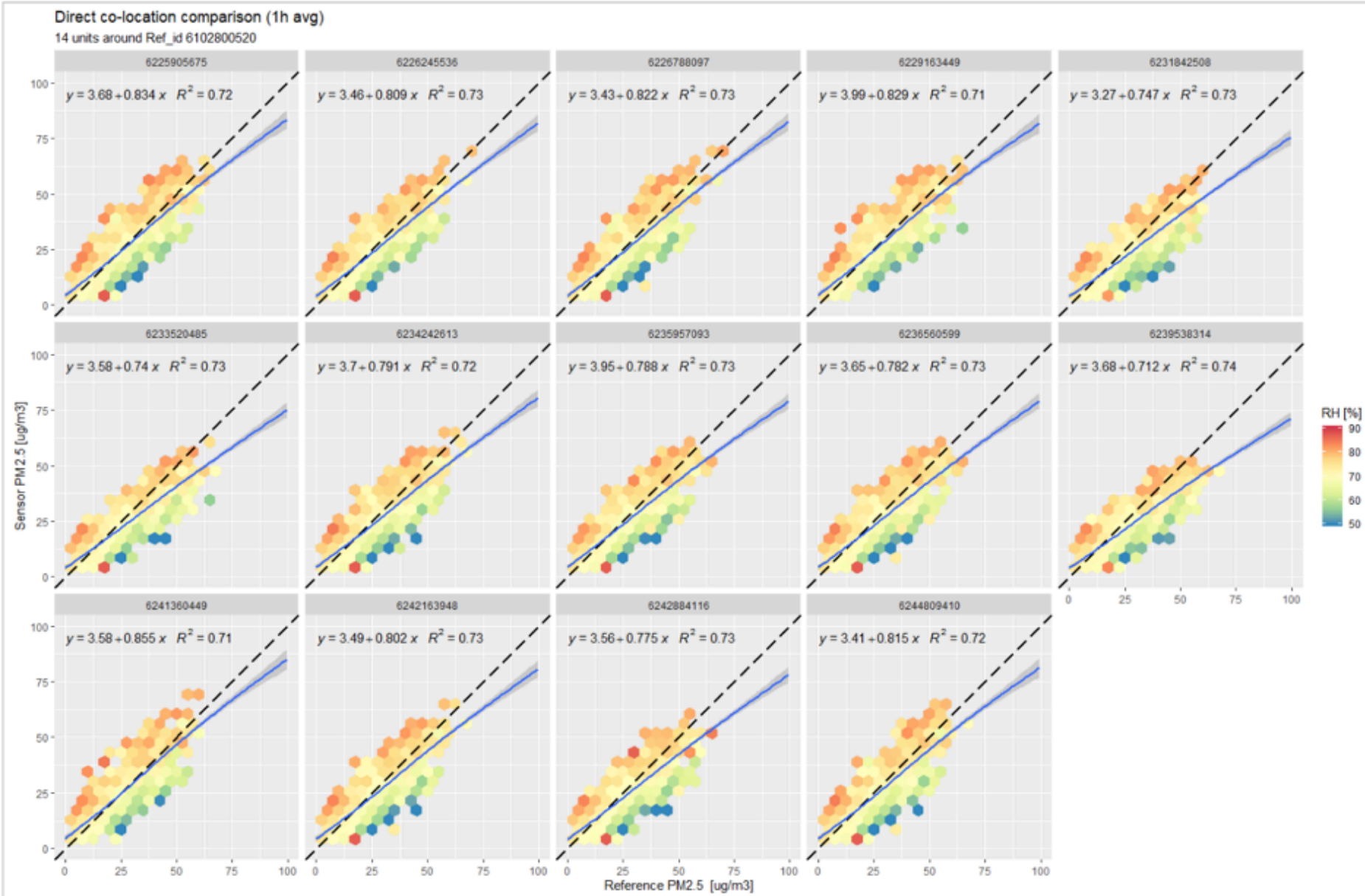
NILU

Example 2

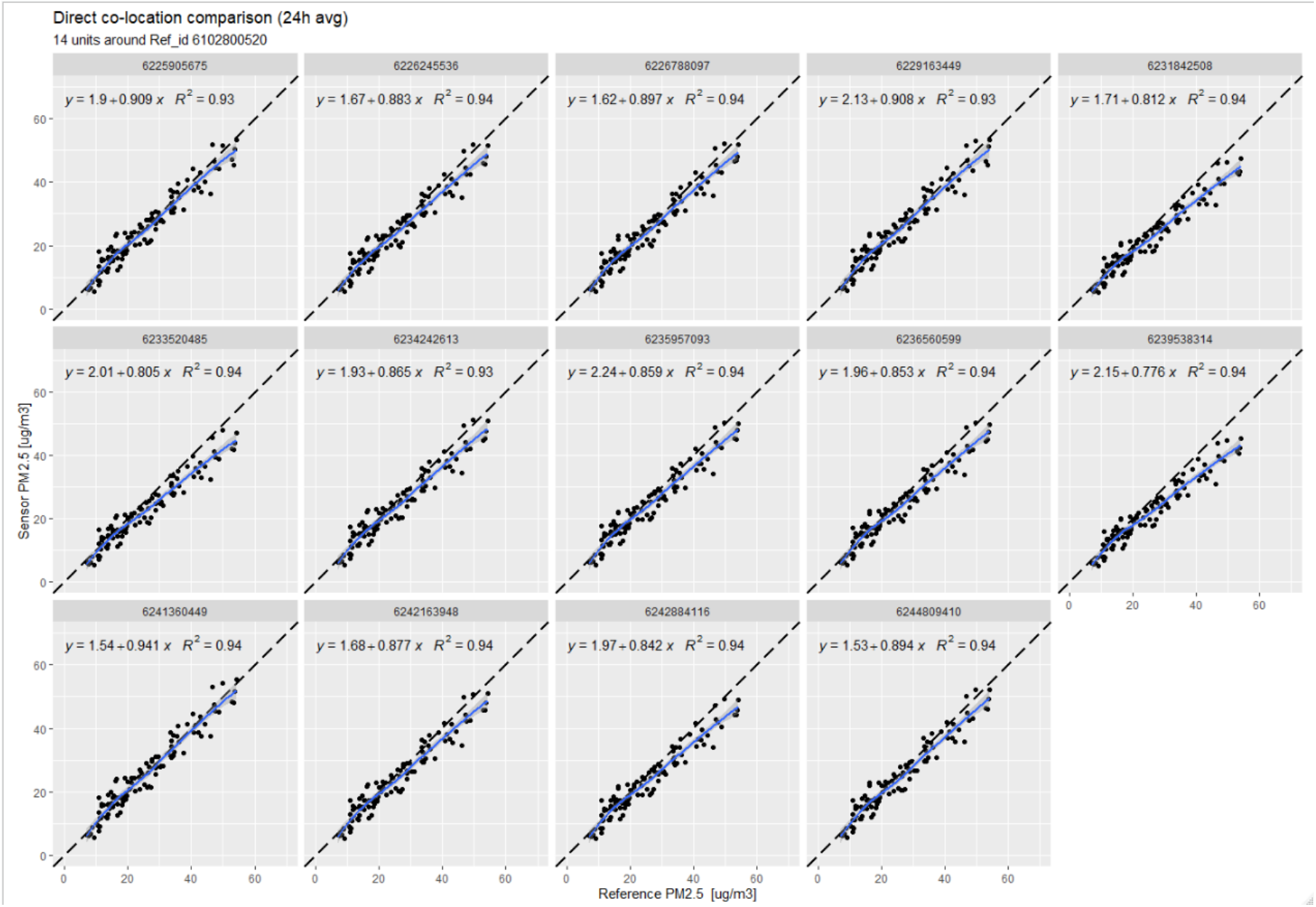
- Plantower PMS5003 sensor
- Price Ca. 100 EUR
- Widely used
- Consistent out-of-the-box performance for $PM_{2.5}$
- Some dependence on relative humidity
- PM_{10} less useful at this point due to physical design principles of the sensor



NILU

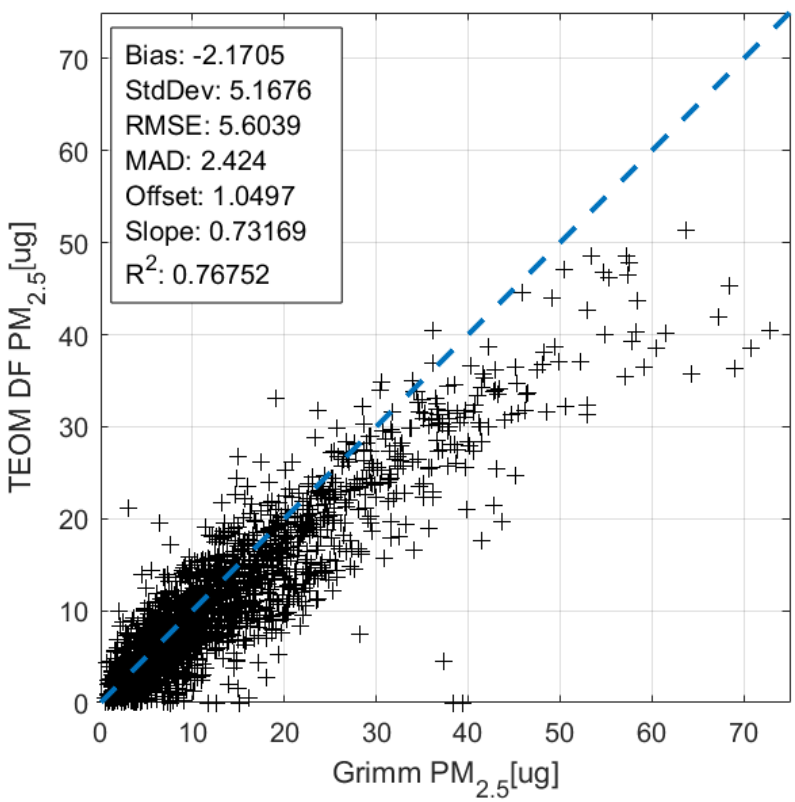


Comparison of hourly $PM_{2.5}$ provided by 14 PMS5003 units against data from an AQ monitoring station with reference equipment (co-location). Tested over a 4 month period.

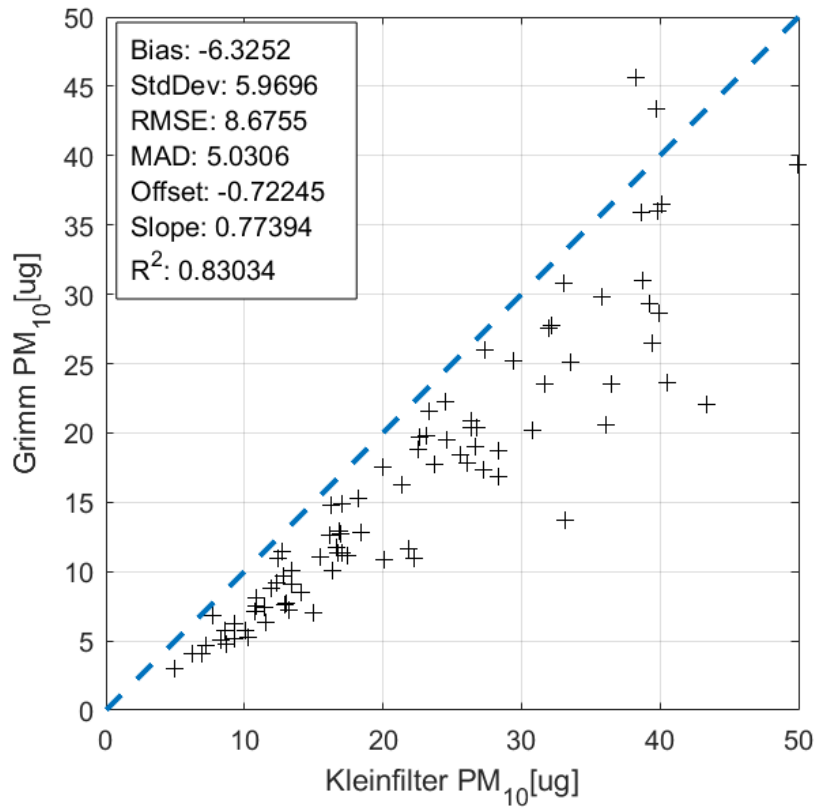


Comparison of daily average $PM_{2.5}$ provided by 14 PMS5003 units against data from an AQ monitoring station with reference equipment (co-location). Tested over a 4 month period.

Putting things in perspective: Official PM monitors for comparison…



Hourly $PM_{2.5}$ from **two widely used reference-equivalent instruments** compared to each other over several months



Daily average PM_{10} from a **reference-equivalent instrument** compared to the **true gravimetric reference** (Kleinfiltergerät) over several months

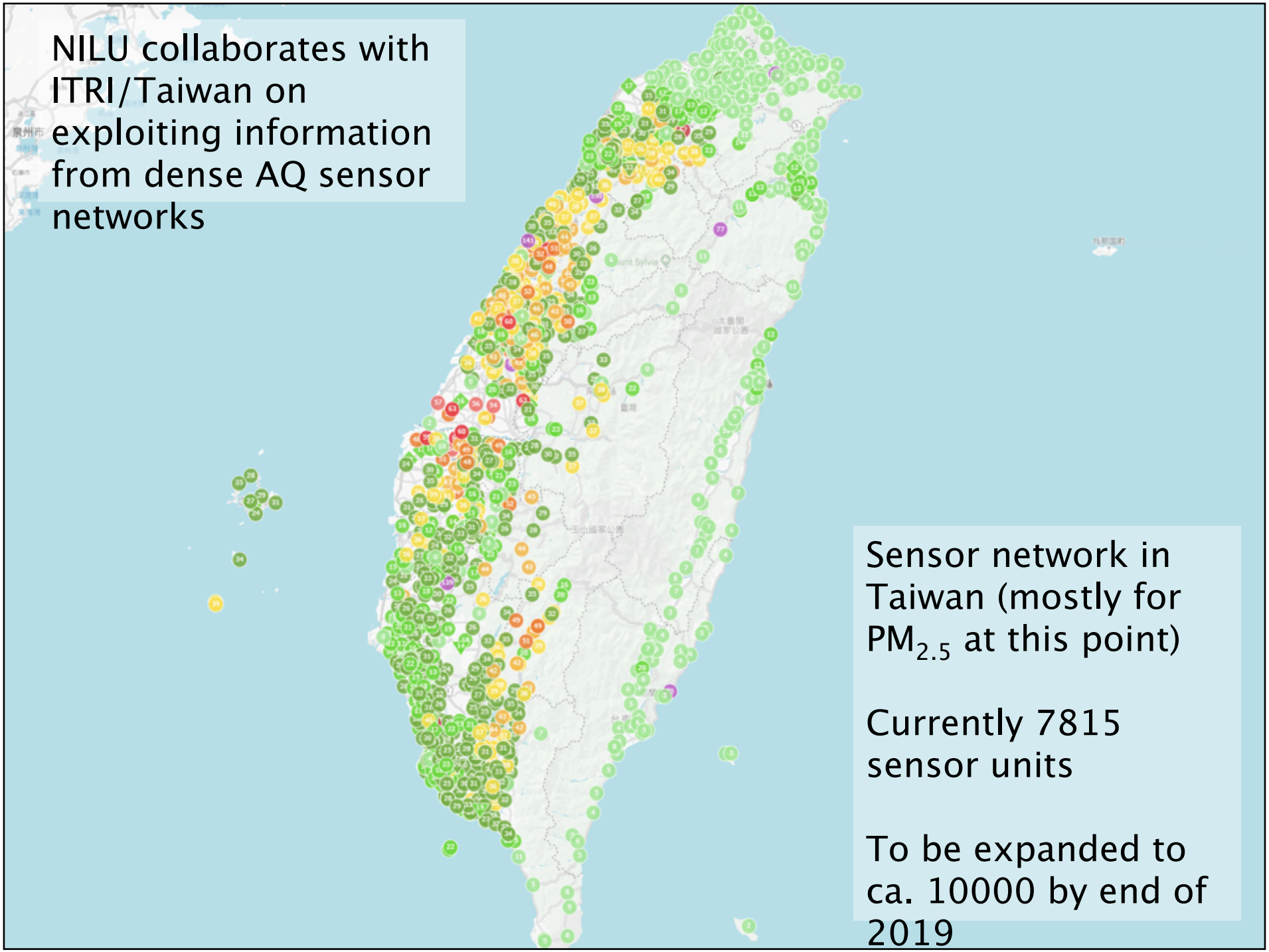
→ Measuring PM is very challenging and even AQM stations typically have substantial errors. Reference instruments and $PM_{2.5}$ sensor systems are not worlds apart anymore.

Part 2

Mapping urban air quality by assimilating sensor observations into a model

NILU

NILU collaborates with ITRI/Taiwan on exploiting information from dense AQ sensor networks



Sensor network in Taiwan (mostly for $PM_{2.5}$ at this point)

Currently 7815 sensor units

To be expanded to ca. 10000 by end of 2019

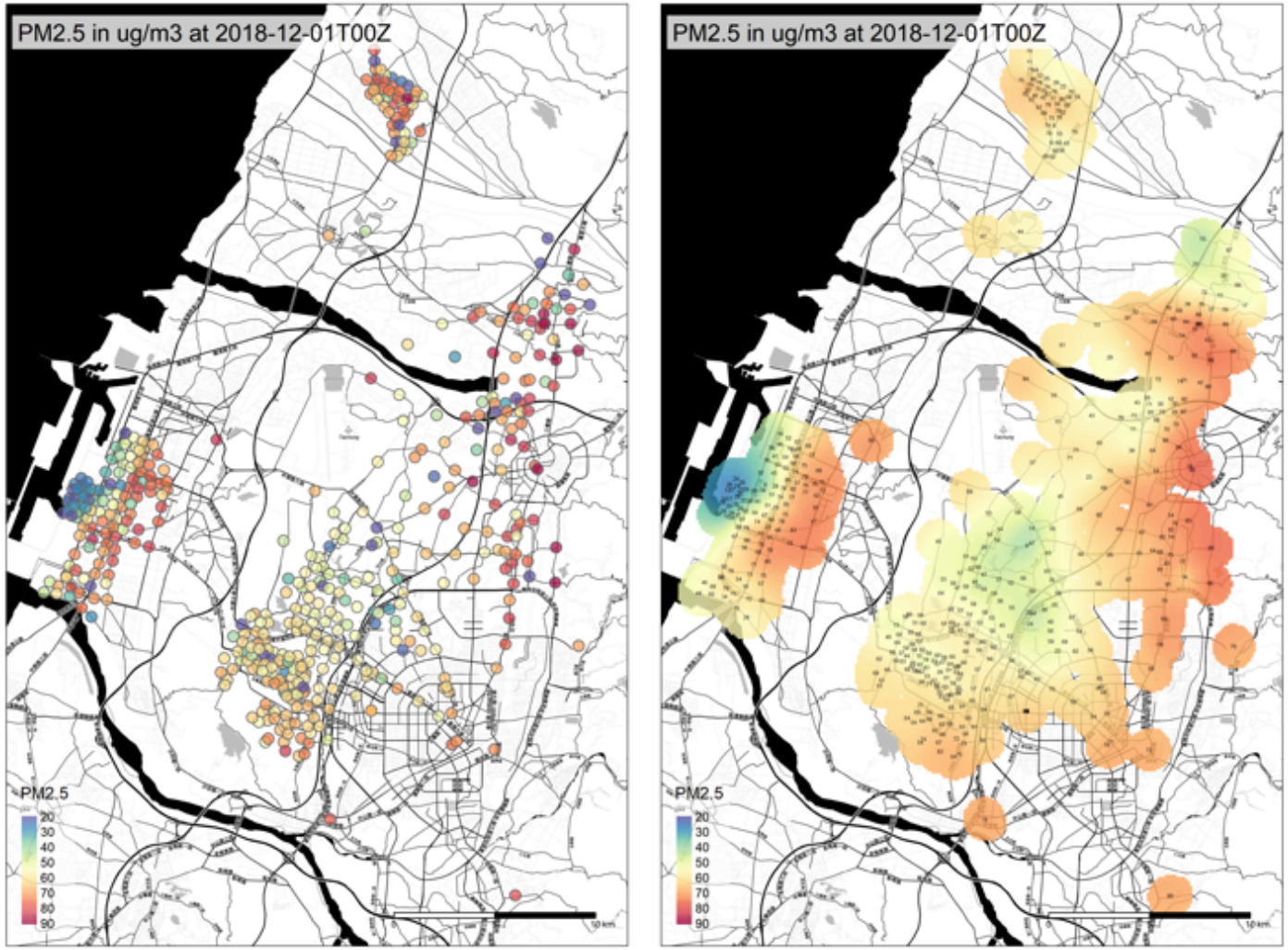
Purely observation-based mapping

Using only observations for urban-scale AQ mapping is very **challenging** due to the **high spatial variability of air pollution**

Typically not feasible or meaningful unless one has a **very dense sensor network**

Solution: **Combine sensor network with model output**

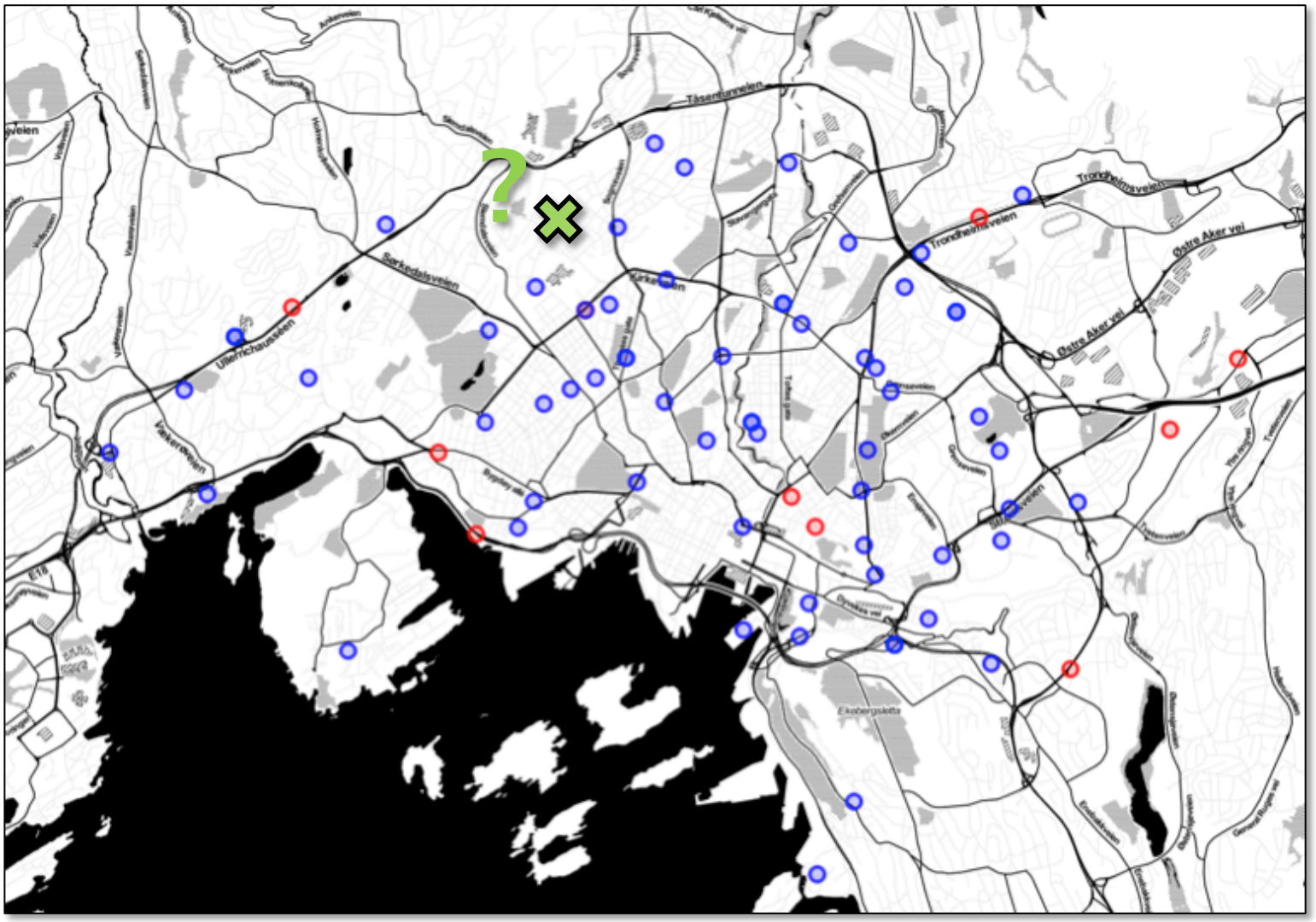
(use the model as "*a priori*" information in areas without observations)



Entirely observation-based mapping of $PM_{2.5}$ using a dense sensor network deployed in Taiwan

NILU

More typical deployment density in Europe



Red markers: Locations of Air Quality Monitoring stations for NO_2

Blue markers: Deployment sites of low-cost sensors

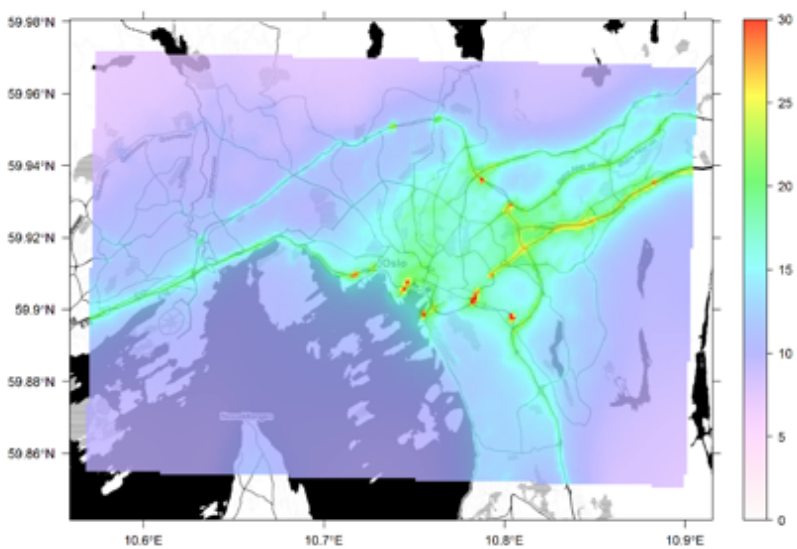
NILU

An example of a previous sensor network for NO_2 deployed in the city of Oslo, Norway (65 units total)

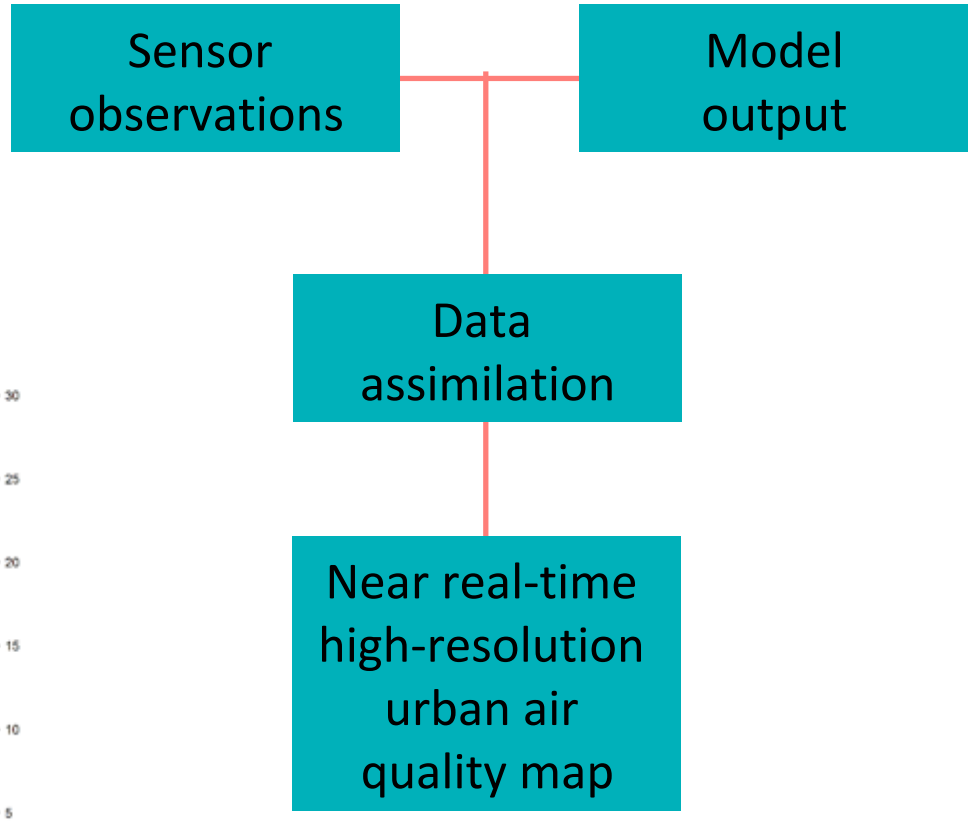
Combination with model output

Combining observations with model output through data fusion or data assimilation **adds value to both** input data sets:

- **Model is constrained** by actual observations
- **Observations are interpolated in space** in a physically meaningful way



Annual average concentration of NO_2 for Oslo as computed by the EPISODE urban air quality model



Data assimilation methodology

- DA has **long heritage** in numerical weather prediction
- Methodologically **similar to geostatistical techniques** (e.g. universal kriging) but easier to directly specify spatial covariance structure etc.
- Specifically takes into account **varying uncertainty of observations**
- Produces **pixel-level uncertainty estimates** of the output map ("analysis error covariance")

Analysis field

Analysis, Background, Weights, Obser-vations, Observation operator

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W}[\mathbf{y}_0 - H$$

Weights

LTP of Obs. Op., Background error covariance, Observation error covariance

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T$$

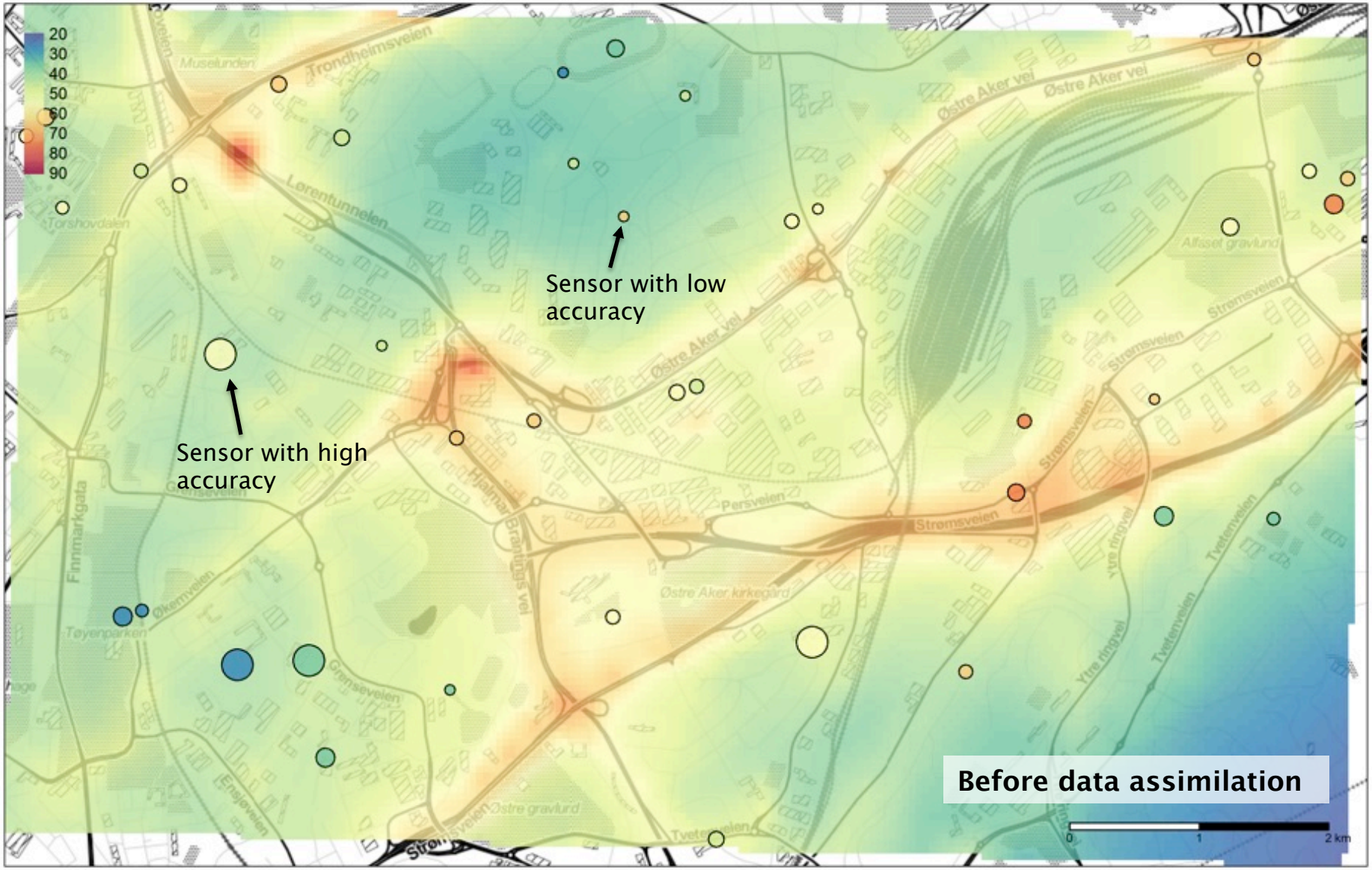
Analysis error covariance

Analysis error, Identity matrix

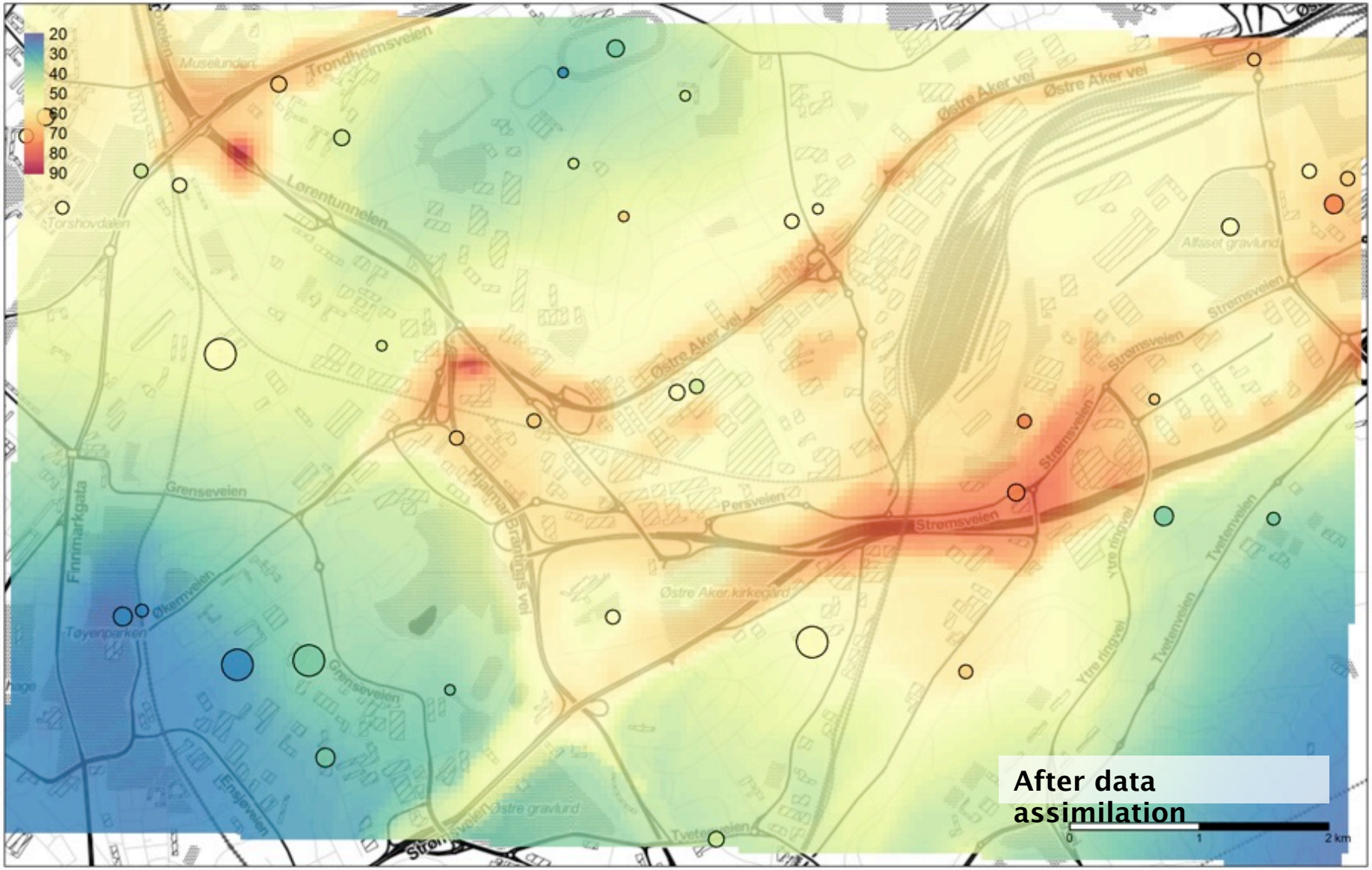
$$\mathbf{P}_a = (\mathbf{I} - \mathbf{W}\mathbf{H})\mathbf{B}$$

- Schneider, P., et al., 2017. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. Environ. Int. 106, 234–247.
- Schneider, P.,et al., 2018. A Network of Low-Cost Air Quality Sensors and Its Use for Mapping Urban Air Quality, in: Bordogna, G., Carrara, P. (Eds.), Mobile Information Systems Leveraging Volunteered Geographic Information for Earth Observation. Springer International Publishing, Cham, pp. 93–110.
- Lahoz, W.A., Schneider, P., 2014. Data assimilation: making sense of Earth Observation. Front. Environ. Sci. 2, 1–28.

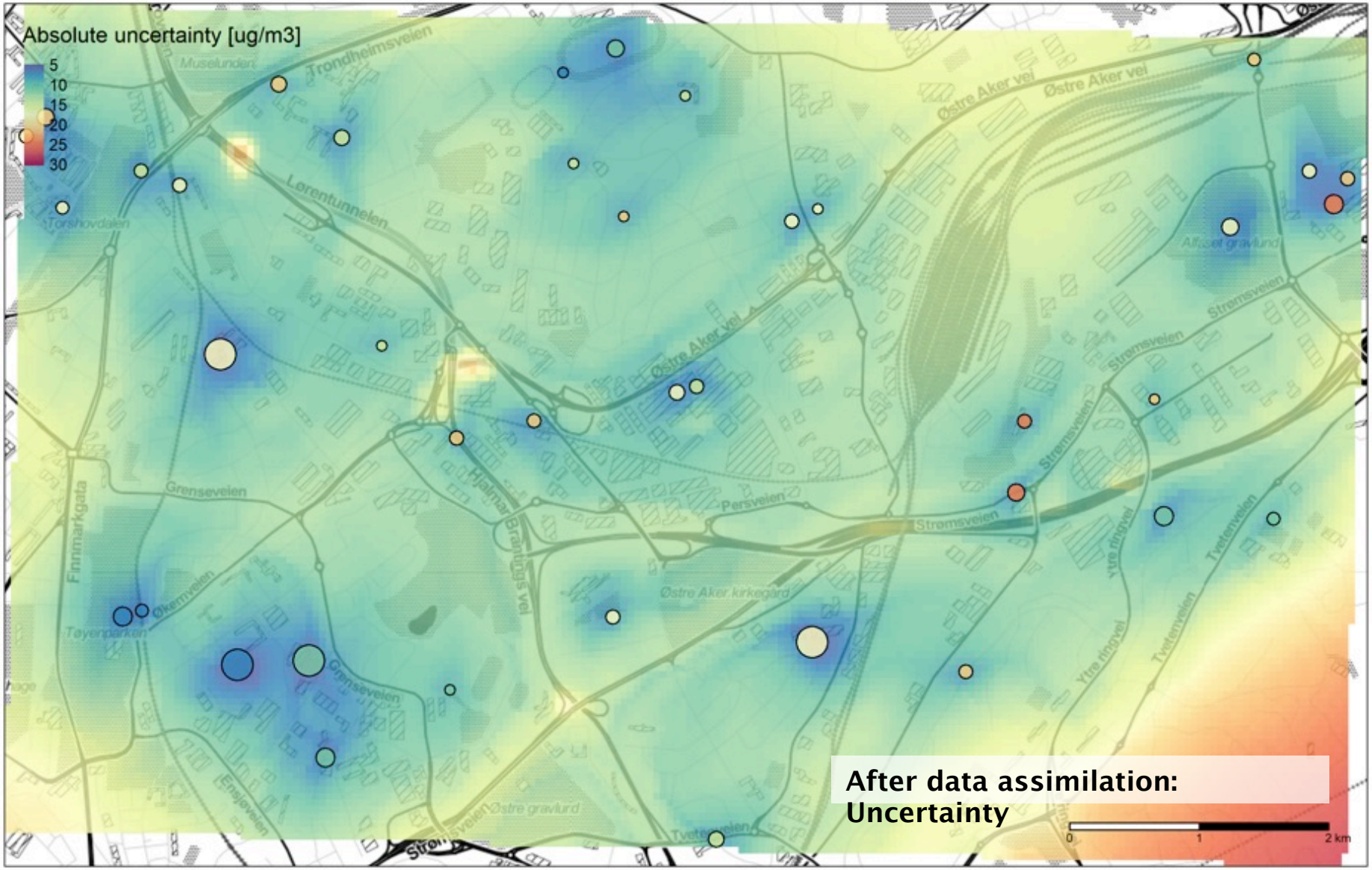
NILU



Urban-scale data assimilation of low-cost sensors in Norway
Model output at 25 m spatial resolution ("*a priori*") and hypothetical observations of NO_2 [in units of $\mu g/m^3$] from AQM stations and a low-cost sensor network of variable accuracy. The size of the marker indicates the accuracy of each observation (inverse of uncertainty).



Urban-scale data assimilation of low-cost sensors in Norway
Data assimilation results ("*analysis*") at 25 m spatial resolution and hypothetical observations of NO_2 [in units of µg/m^3] from AQM stations and a low-cost sensor network of variable accuracy. Marker size indicates the accuracy of each observation (inverse of uncertainty).



Urban-scale data assimilation of low-cost sensors in Norway
Absolute uncertainty of the analysis field and hypothetical observations of NO_2 [in units of μg/m^3] from AQM stations and a low-cost sensor network of variable accuracy. Marker size indicates the accuracy of each observation (inverse of uncertainty).

Conclusions

- The **accuracy** of low-cost sensors is **improving**, opening up possible applications for modelling
- In particular some **sensors for $PM_{2.5}$** consistently **reach R^2 values of 0.7 to 0.9** against reference instruments
- Using a dense network of such sensors systems can contribute to **validation of urban-scale models** (particularly with respect to spatial patterns)
- **Assimilating data** from a dense sensor network into urban-scale models can add value to both datasets and **improve real-time urban-scale AQ mapping**

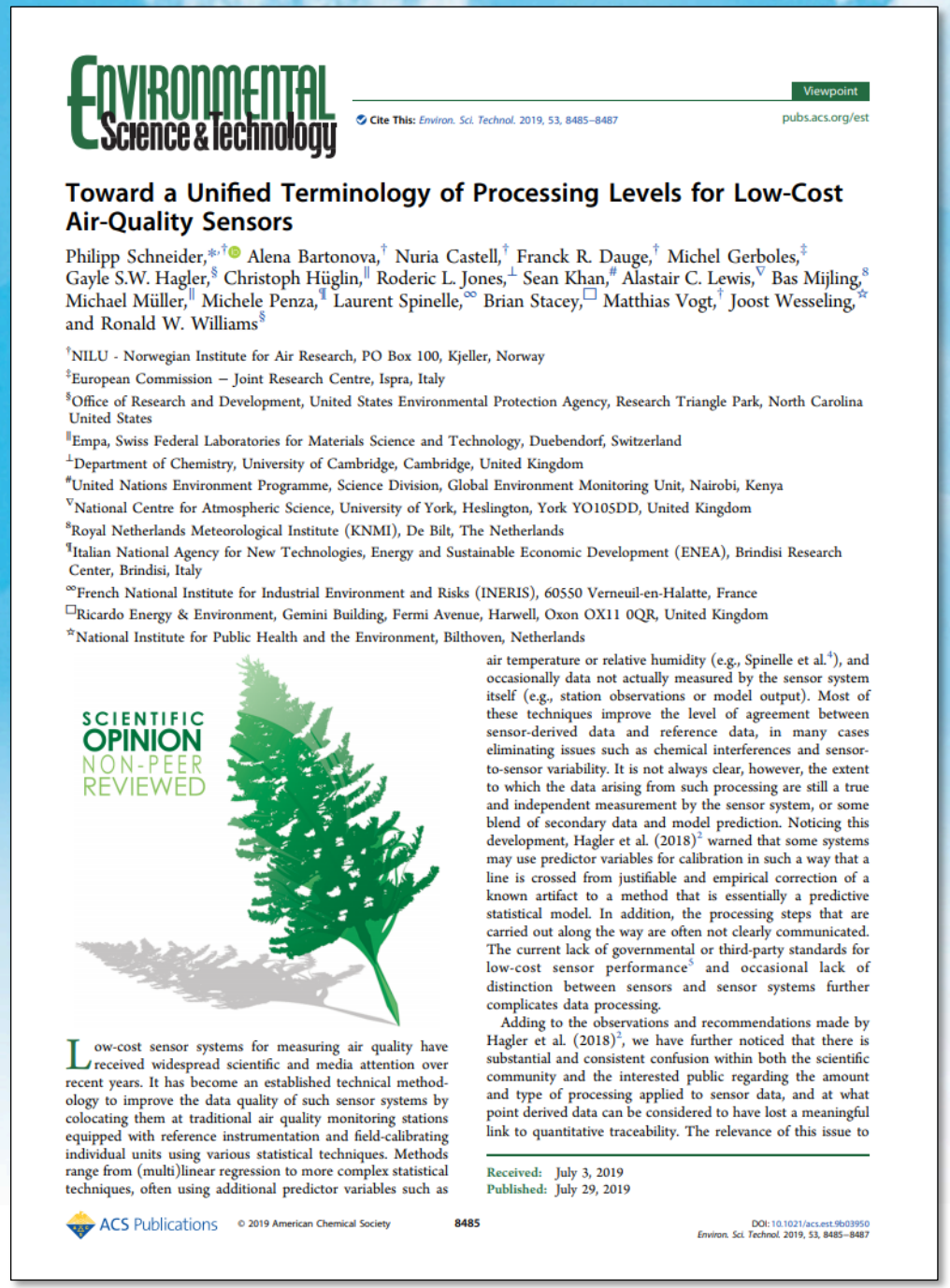
NILU

A short public service announcement…

New paper introducing standardized processing levels for low-cost sensors

Schneider, P., A. Bartonova, N. Castell, F. R. Dauge, M. Gerboles, G. S. W. Hagler, C. Hüglin-, R. L. Jones, S. Khan, A. C. Lewis, B. Mijling, M. Müller, M. Penza, L. Spinelle, B. Stacey, M. Vogt, J. Wesseling, R. W. Williams (2019). ***Toward a Unified Terminology of Processing Levels for Low-Cost Air-Quality Sensors***. Environmental Science & Technology, 2019, 53, 15, 8485-8487.

NILU



Environmental Science & Technology

Viewpoint

Cite This: Environ. Sci. Technol. 2019, 53, 8485−8487

pubs.acs.org/est

Toward a Unified Terminology of Processing Levels for Low-Cost Air-Quality Sensors

Philipp Schneider,[*,†] Alena Bartonova,[†] Nuria Castell,[†] Franck R. Dauge,[†] Michel Gerboles,[‡] Gayle S.W. Hagler,[§] Christoph Hüglin,[∥] Roderic L. Jones,[⊥] Sean Khan,[#] Alastair C. Lewis,[∇] Bas Mijling,[8] Michael Müller,[∥] Michele Penza,[¶] Laurent Spinelle,[∞] Brian Stacey,[□] Matthias Vogt,[†] Joost Wesseling,[☆] and Ronald W. Williams[§]

[†]NILU - Norwegian Institute for Air Research, PO Box 100, Kjeller, Norway

[‡]European Commission – Joint Research Centre, Ispra, Italy

[§]Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, North Carolina United States

[∥]Empa, Swiss Federal Laboratories for Materials Science and Technology, Duebendorf, Switzerland

[⊥]Department of Chemistry, University of Cambridge, Cambridge, United Kingdom

[#]United Nations Environment Programme, Science Division, Global Environment Monitoring Unit, Nairobi, Kenya

[∇]National Centre for Atmospheric Science, University of York, Heslington, York YO105DD, United Kingdom

[8]Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

[¶]Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Brindisi Research Center, Brindisi, Italy

[∞]French National Institute for Industrial Environment and Risks (INERIS), 60550 Verneuil-en-Halatte, France

[□]Ricardo Energy & Environment, Gemini Building, Fermi Avenue, Harwell, Oxon OX11 0QR, United Kingdom

[☆]National Institute for Public Health and the Environment, Bilthoven, Netherlands

SCIENTIFIC OPINION NON-PEER REVIEWED

Low-cost sensor systems for measuring air quality have received widespread scientific and media attention over recent years. It has become an established technical methodology to improve the data quality of such sensor systems by colocating them at traditional air quality monitoring stations equipped with reference instrumentation and field-calibrating individual units using various statistical techniques. Methods range from (multi)linear regression to more complex statistical techniques, often using additional predictor variables such as air temperature or relative humidity (e.g., Spinelle et al.[4]), and occasionally data not actually measured by the sensor system itself (e.g., station observations or model output). Most of these techniques improve the level of agreement between sensor-derived data and reference data, in many cases eliminating issues such as chemical interferences and sensor-to-sensor variability. It is not always clear, however, the extent to which the data arising from such processing are still a true and independent measurement by the sensor system, or some blend of secondary data and model prediction. Noticing this development, Hagler et al. (2018)[2] warned that some systems may use predictor variables for calibration in such a way that a line is crossed from justifiable and empirical correction of a known artifact to a method that is essentially a predictive statistical model. In addition, the processing steps that are carried out along the way are often not clearly communicated. The current lack of governmental or third-party standards for low-cost sensor performance[5] and occasional lack of distinction between sensors and sensor systems further complicates data processing.

Adding to the observations and recommendations made by Hagler et al. (2018)[2], we have further noticed that there is substantial and consistent confusion within both the scientific community and the interested public regarding the amount and type of processing applied to sensor data, and at what point derived data can be considered to have lost a meaningful link to quantitative traceability. The relevance of this issue to

Received: July 3, 2019
Published: July 29, 2019

DOI: 10.1021/acs.est.9b03950
Environ. Sci. Technol. 2019, 53, 8485−8487

Level	Name	Definition	Example: Gas-sensors	Example: Particle-sensors
Level-0	Raw measurements	Original measurand produced by sensor system	Voltage corresponding to measured quantity, such as current for electrochemical and infrared sensors, resistance/ conductance for metal-oxide sensors	Voltage corresponding to current due to light scattered in nephelometers, or to binned counts for optical particle-counters
Level-1	Intermediate geophysical quantities	Estimate derived from corresponding Level-0 data, using basic physical principles or simple calibration equations, and no compensation schemes.	For electrochemical sensors, NO_2 concentration in µg/m^3 or ppb, using only Level-0 data from the NO_2 sensor itself with no additional corrections beyond factory calibration ("raw data in concentration units")	Binned particle-counts or PM mass in µg/m^3 derived from Level-0 data using simple calibration/assumed particle-density
Level-2A	Standard geophysical quantities	Estimate using sensor plus other on-board sensors demonstrated as appropriate for artifact correction and directly related to measurement principle (Hagler et al., 2018)	NO_2 concentration in µg/m^3 or ppb, derived from onboard NO_2/NO/O_3 sensors, corrected for interferences and/or T/RH effects using onboard data	PM concentration in µg/m^3, corrected for T/RH effects with onboard-measured T/RH
Level-2B	Standard geophysical quantities-extended	As Level-2A but using external data demonstrated as appropriate for artifact correction and directly related to measurement principle (Hagler et al., 2018)	As Level-2A but using external data from nearby station related to correcting for interferences based on the measurement principle (e.g. O3, T/RH)	As Level-2A but using external T/RH from nearby station
Measurement/prediction boundary				
Level-3	Advanced geophysical quantities	Estimate using sensor plus internal/ external inputs, not constrained to data proven as causes of measurement bias or related to measurement principle (Hagler et al., 2018)	NO_2 concentration in µg/m^3 or ppb, derived from Level-2A or Level-2B data, further corrected by proxies known to be correlated with NO_2, e.g. emissions or modeled NO_2	PM concentration in in µg/m^3 , derived from Level-2A or Level-2B data, further corrected by proxies known to be correlated with PM, e.g. emissions or modeled PM
Level-4	Spatially continuous geophysical quantities	Spatially continuous maps derived from network of sensor systems	Map of NO_2 concentrations in µg/m^3 or ppb, e.g. by assimilation of network data into a physical model	Map of $PM_{2.5}$ concentrations in µg/m^3, e.g. by assimilation of network data into a physical model

Schneider, P., A. Bartonova, N. Castell, F. R. Dauge, M. Gerboles, G. S. W. Hagler, C. Hüglin-, R. L. Jones, S. Khan, A. C. Lewis, B. Mijling, M. Müller, M. Penza, L. Spinelle, B. Stacey, M. Vogt, J. Wesseling, R. W. Williams (2019). ***Toward a Unified Terminology of Processing Levels for Low-Cost Air-Quality Sensors***. Environmental Science & Technology, 2019, 53, 15, 8485-8487.

Thank you!

For more info please contact:

Philipp Schneider (ps@nilu.no)

NILU