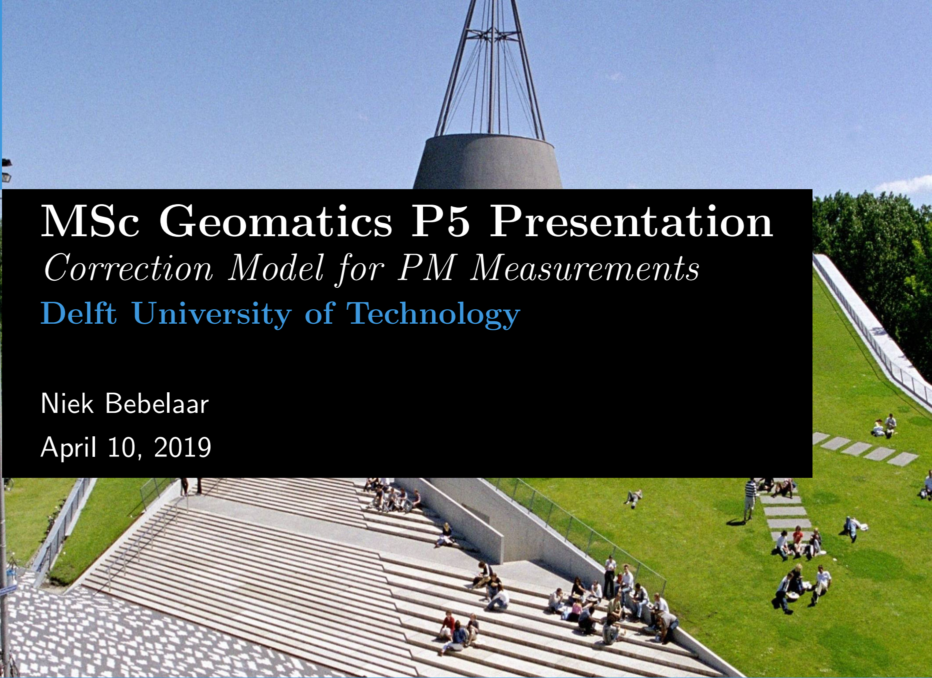


MSc Geomatics P5 Presentation

Correction Model for PM Measurements

Delft University of Technology

Niek Bebelaar

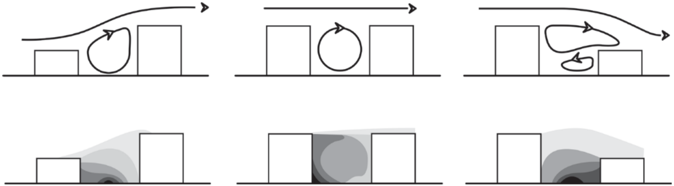
April 10, 2019

Outline

Outline

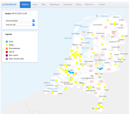
TU Delft

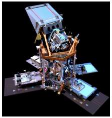
- Motivation
    - Effect of air pollution on human health [Pope & Dockery, 2006; WHO, 2006; Bentayeb et al, 2015]
    - Pollutants mainly present in urban areas [Pijpers-Van Esch, 2015]



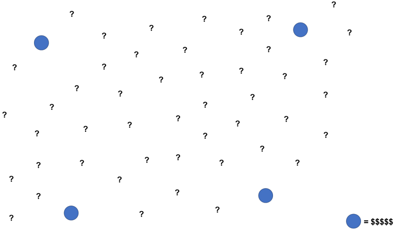
Images: Pijpers-Van Esch, 2015

- Motivation
    - Various initiatives to monitor the air quality [Van Alphen & Pot, 2014; Apte et al, 2017; De Vries et al, 2016]





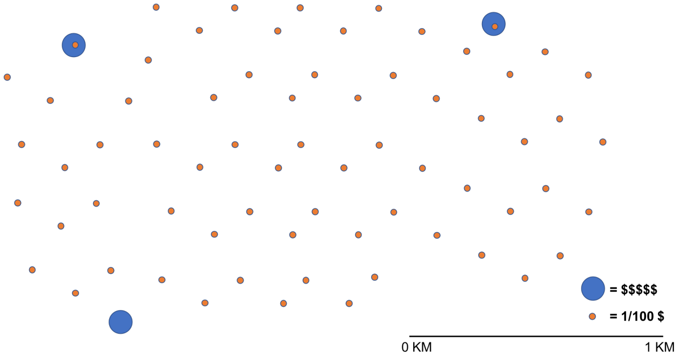
Use of low-cost sensors in Smart Cities [Salim, 2012]



Images: luchtmeetnet.nl; apte.caee.utexas.edu/google-air-mapping/; De Vries et al, 2016

Outline

- Problem statement
  - Quality of the low-cost sensors is doubtful [Gerboles et al, 2017]
  - Assess the data quality [Lewis & Edwards, 2016; Morawska et al, 2018]
  - Thus: assess and improve the data quality from low-cost PM sensors



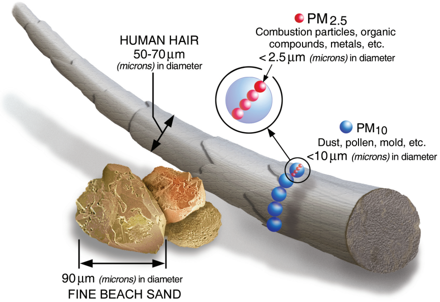
Outline

- Research question
  - How can **accuracy and precision** of **Particulate Matter** measurement results from a **low-cost** outdoor sensor network be improved by using a **correction model**, using data from **reference sensors** and additional sensors measuring **inferencing phenomena**?
- Two parts:
  - Data quality assessment
  - Data quality improvement

Outline

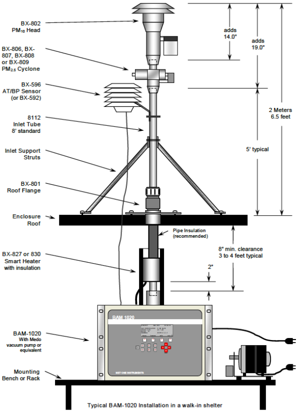
TU Delft

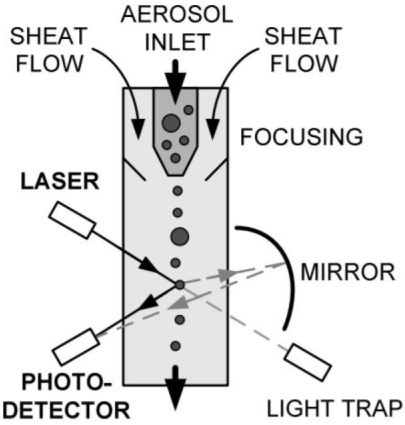
- Particulate Matter
  - What is Particulate Matter (PM) and where does it originate from?
    - PM1
    - PM2.5
    - PM10



Source image: Environmental Protection Agency, 2016

- How is Particulate Matter monitored?





Source images: Met One Instruments, 2016; Carminati et al, 2011

Outline

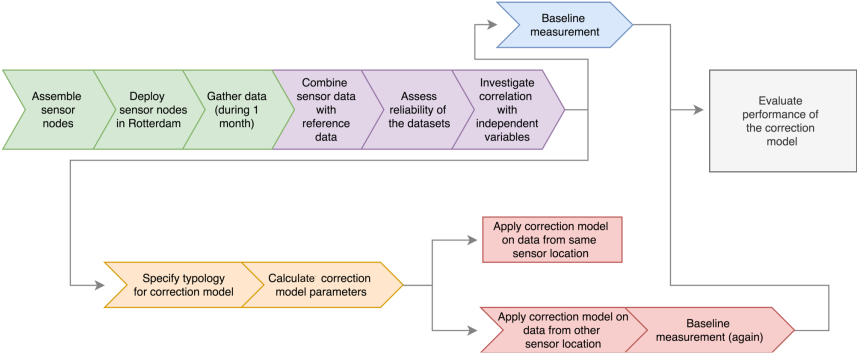
TU Delft

- Air quality monitoring with low-cost PM sensors
  - Humidity and other meteorological variables can have an inferencing effect on PM from low-cost sensors [Cross et al, 2017; Mukherjee et al, 2017]
  - Performance varies from unit to unit: examine data quality of each sensor before use [Castell et al, 2017]
  - (Continuous) calibration and each new sensor needs to have data quality assessment [Li & Biswas, 2017]

Source image: http://mappingforchange.org.uk/projects/citizen-science-used-to-map-community-air-quality

Outline

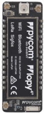
- Overview methodology

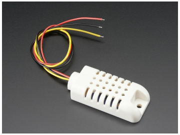


Outline

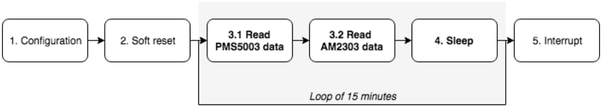
TU Delft

- Sensor nodes design and engineering
  - Two sensor nodes: two different locations in the study area
  - Two PM sensors per sensor node
  - Hardware: assembled, configured and tested at Science Centre Delft
    - Plantower PMS5003
    - PyCom LoPy
    - AM2302





  - Software

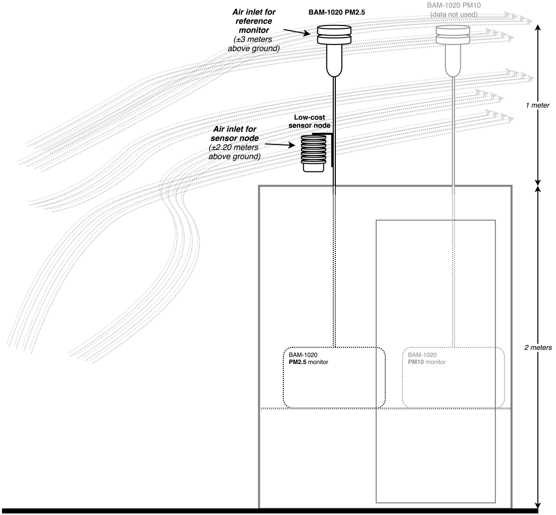


Source images: aqicn.org/sensor/pms5003-7003/; pycom.io/product/lopy4/; adafruit.com/product/393

Outline

- Study area and data collection
  - Study area in Rotterdam: Pleinweg and Zwartewaalstraat

Study area and data collection



- Study area and data collection
    - Data from Sensor Node
        - Particulate matter
        - Relative humidity
        - Temperature
    - Data from environmental agency (via API)
        - Particulate matter (ground truth)
    - Data from meteorological agency (via API)
        - Barometric air pressure
        - Wind speed
        - Wind direction
        - Relative humidity (extra)
        - Temperature (extra)
    - Data collected during 1 month
        - 15 May to 16 June 2018

Outline

TUDelft

- Combining and preprocessing
- Relationships between the variables
  - Relationships between independent variables
  - Relationships between independent and dependent variables

TU Delft

Outline

- Baseline measurement
  - Evaluation metric: Root Mean Square Error (RMSE)
    - Requirements when using RMSE [Chai & Draxler, 2014]
    - Formula

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(z_i^M - z_i^R)^2}{n}} \quad (1)$$

- Baseline measurement
  - RMSE of normalized datasets before and after removing systematic error

	RMSE normalized data	**Standard Deviation normalized data**	**R**	**RMSE after SE removal**	**Standard Deviation after SE removal**
PW sensor 1	0.1813	0.1702	0.86	**0.1249**	0.1756
PW sensor 2	0.1327	0.1403	0.87	**0.0918**	0.1448
ZW sensor 1	0.2045	0.1620	0.86	**0.1706**	0.1714
ZW sensor 2	0.1527	0.1316	0.83	**0.1256**	0.1392

Outline

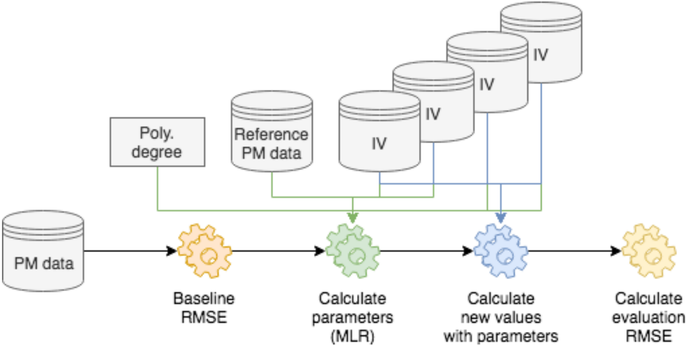
- Creating correction models
  - Calculating the parameters with stepwise Multiple Linear Regression (MLR)

$$Y = a + p_1 X_1 + p_2 X_2 + \ldots + p_k X_k \tag{2}$$

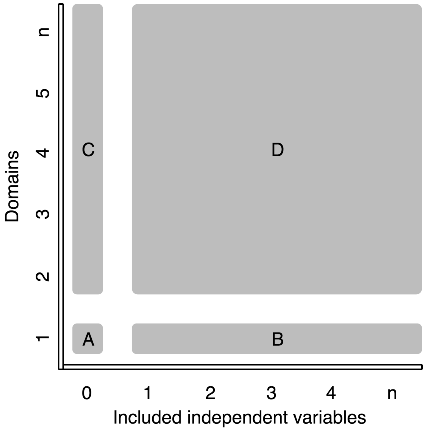
or

$$Y = a + p_1 X_1 + p_2 X_1^2 + \ldots + p_k X_1^k \tag{3}$$

- Creating correction models
  - Algorithm that performs the following steps

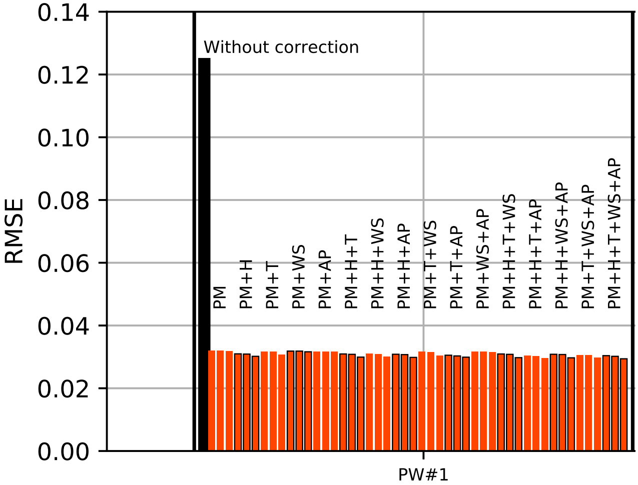


- Creating correction models
  - Proposed typologies for correction models



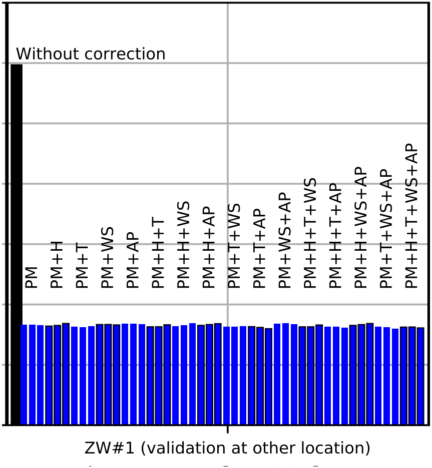
Outline

- Type A and B models: models created with data from Pleinweg sensor 1
  - And applied on Pleinweg sensor 1...

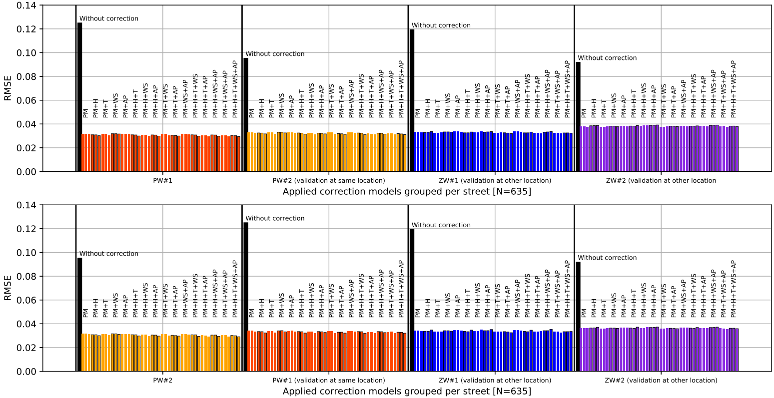


TUDelft

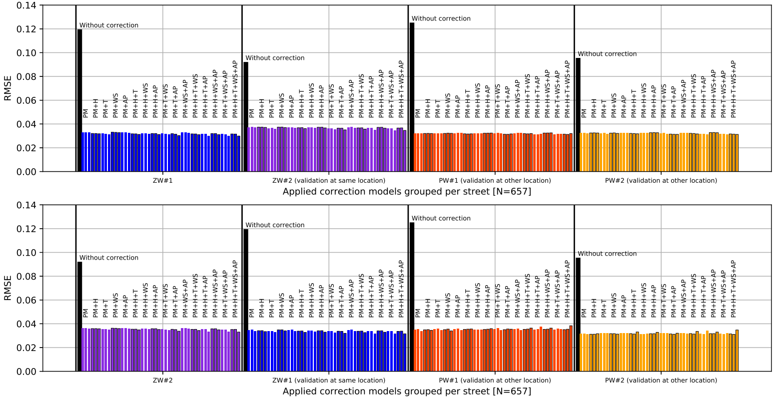
- Type A and B models: models created with data from Pleinweg sensor 1
  - And applied on Zwartewaalstraat sensor 1



- Type A and B models: models created with data from Pleinweg sensor 1 and sensor 2



- Type A and B models: models created with data from Zwartewaalstraat sensor 1 and sensor 2



Outline

- Analysis of "Type A and B" models
  - Best correction model per location
  - Time series for best correction model applied on the data

Best correction model (created with data from Pleinweg: applied on Zwartewaalstraat)

Parameters calculated with data from Pleinweg sensor 1			
Best correction model for data from:	Included variables	RMSE baseline	RMSE corrected
Zwartewaalstraat sensor 1	PM + T + WS + AP	0.11920	0.03192
Zwartewaalstraat sensor 2	PM + T	0.09180	0.03759

Parameters calculated with data from Pleinweg sensor 2			
Best correction model for data from:	Included variables	RMSE baseline	RMSE corrected
Zwartewaalstraat sensor 1	PM+T+WS+AP	0.11920	0.03237
Zwartewaalstraat sensor 2	PM+T+WS+AP	0.09180	0.03552

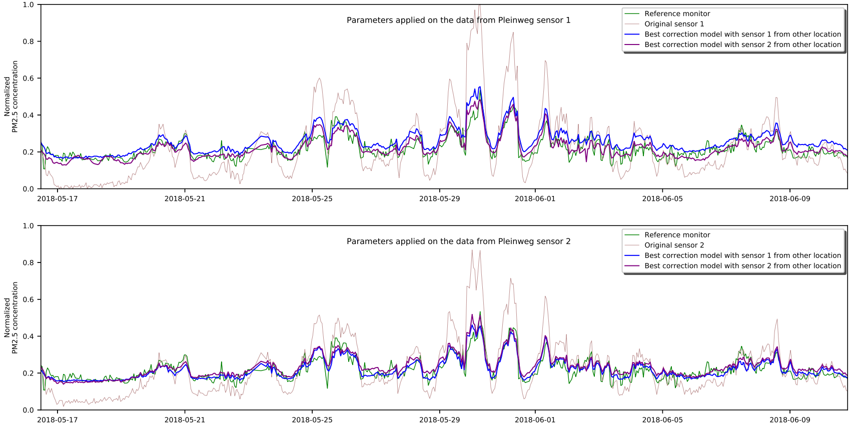
TU Delft

Best correction model (created with data from Zwartewaalstraat: applied on Pleinweg)

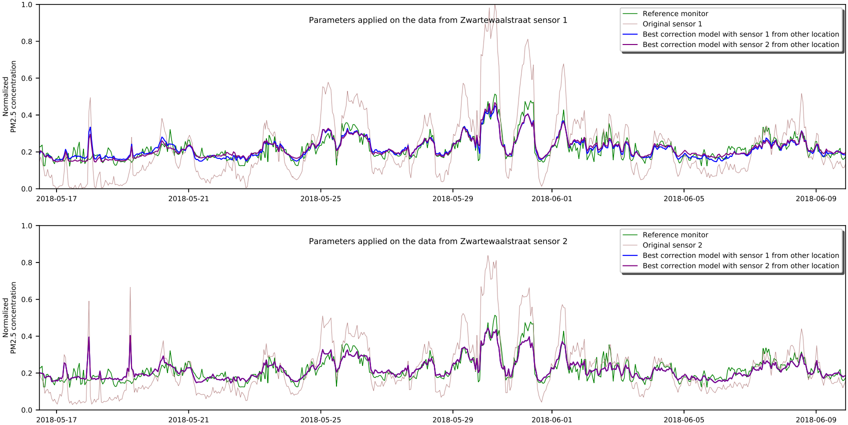
Parameters calculated with data from Zwartewaalstraat sensor 1			
Best correction model for data from:	Included variables	RMSE baseline	RMSE corrected
Pleinweg sensor 1	PM + H + T+ WS + AP	0.12490	0.03125
Pleinweg sensor 2	PM + T + WS + AP	0.09520	0.03116

Parameters calculated with data from Zwartewaalstraat sensor 2			
Best correction model for data from:	Included variables	RMSE baseline	RMSE corrected
Pleinweg sensor 1	PM + WS	0.12490	0.03400
Pleinweg sensor 2	PM + H + T+ WS + AP	0.09520	0.03110

Time series for best correction model applied on the data from Pleinweg



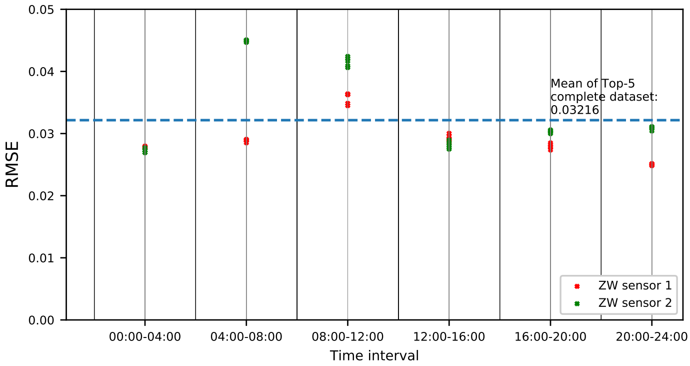
Time series for best correction model applied on the data from Zwartewaalstraat



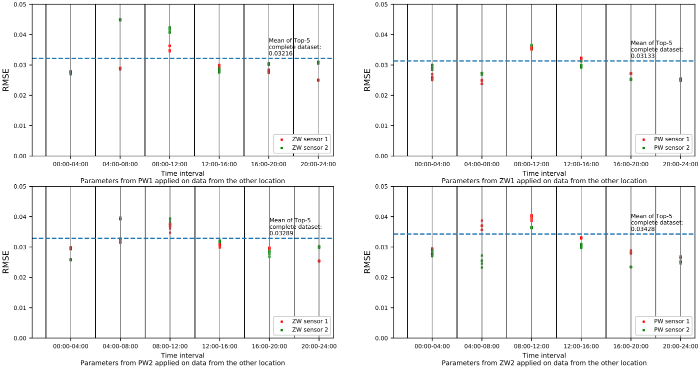
TU Delft

Outline

- Results of "Type C and D" models
  - Grouping based on time interval



- Results of "Type C and D" models
  - Grouping based on time interval



Outline

TU Delft

- Answering research question
    - How can **accuracy and precision** of **Particulate Matter** measurement results from a **low-cost** outdoor sensor network be improved by using a **correction model**, using data from **reference sensors** and additional sensors measuring **inferencing phenomena**?
    - With the proposed method the data quality improves
        - Before correction: RMSE ranging from 0.09180 to 0.12490 on various locations
        - After correction: RMSE ranging from 0.03110 to 0.03759
    - Not necessary to include IVs temperature, humidity and air pressure

Outline

- Recommendations
  - Redo with larger dataset (to get data for Type C & D models)
  - Real time sensor data
  - Instead of stepwise MLR, apply another method
  - Reference monitors also as input variable in the correction model

- Thank you for your attention!
- Questions?