Sensor collocation data analysis in Mehan Garden, Manila

ASIA BLUE SKIES PROGRAM IN THE CITY OF MANILA

Date of report preparation: October 12, 2023
Report prepared by: Clean Air Asia (Everlyn Tamayo, Air Quality and Climate Change Science Lead)

This report aims to summarize the correction of sensor data in Mehan Garden, which has been collocated with the Department of Environment and Natural Resources - Environmental Management Bureau (DENR-EMB) reference station in the same location in Manila, Philippines from August 2020 to May 2022 (excluding periods when the reference station was under maintenance).

This activity is part of the implementation of the **Asia Blue Skies Program** supported by 3M and implemented by Clean Air Asia in partnership with the City Government of Manila and the DENR-EMB National Capital Region (NCR). The Asia Blue Skies Program aims to deliver science-based air quality solutions for a healthier and more livable Manila City. For more information about this program and other Clean Air Asia activities, please visit https://www.cleanairasia.org/. For questions about this report, please email aqccmanila@cleanairasia.org or everlyn.tamayo@cleanairasia.org.

Instrument information

Air quality sensor

The non-reference sensor used in the study is the Clarity Node-S, which is a solar-powered IoT air quality monitoring system which transmits data through global cellular communications. It measures particulate matter (PM) and nitrogen dioxide (NO_2), through laser light scattering and an electrochemical cell, respectively. In the collocation experiment, only PM performance was studied because the reference station only measures PM. The range for PM measurements is 0-1000 μg/m^3, with a 1 μg/m^3 resolution. The technical specifications of the sensor can be accessed through (https://www.clarity.io/products/clarity-node-s).

Preliminary analysis shows less agreement of PM_{10} measurements. This analysis is therefore focused on the performance assessment of the sensors based on $PM_{2.5}$ readings.

Reference air quality station

The reference station by the DENR-EMB NCR provides real-time monitoring data of PM mass concentrations through the Teledyne T640 analyzer which uses scattered light spectrometry for measurement. In particular, it employs broadband spectroscopy using 90° white-light scattering with Polychromatic LED. The particle size resolution covers 256 sizes over 0.18 – 20μm range, combined to 64 channels for mass calculation. The mass resolution measurement range is 0.1 - 10,000 μg/m^3 while the measurement and display resolution is 0.1 μg/m^3. More details on the instrument can be found in the Teledyne website (https://www.teledyne-api.com/products/particulate-instruments/t640).

Instrument deployment and collocation period

The Clarity Node-S unit studied in the collocation, with unit code AFXWSXVD, was first deployed in Mendiola, Manila on February 4, 2020 before being transferred to Mehan Garden on August 28 of the same year for the collocation experiment. The reference station on the other hand, was installed on the first week of August and was operational before August 27, 2020 after the standard calibration.

The sensor was installed in the railings of the reference station housing, around 1 meter from the inlet of the Teledyne instrument. The **collocation period was officially started on August 29, 2020. For this collocation analysis, the coverage of the data is from August 29, 2022 to May 23, 2022 (total of 633 days)**. Periods with no data from the reference instrument due to maintenance and other issues were identified to be from October 24-30, 2020; November 12-23, 2020; June 18-July 2, 2021; September 14-15, 2021; October 12-13, 2021; and December 13, 2021 (total of 39 missing days). Sensor data was complete for the whole period from August 29, 2020 to May 23 ,2022.

(Once reference data is shared from May 2022 onwards, the analysis can be further expanded.)

Collocation results

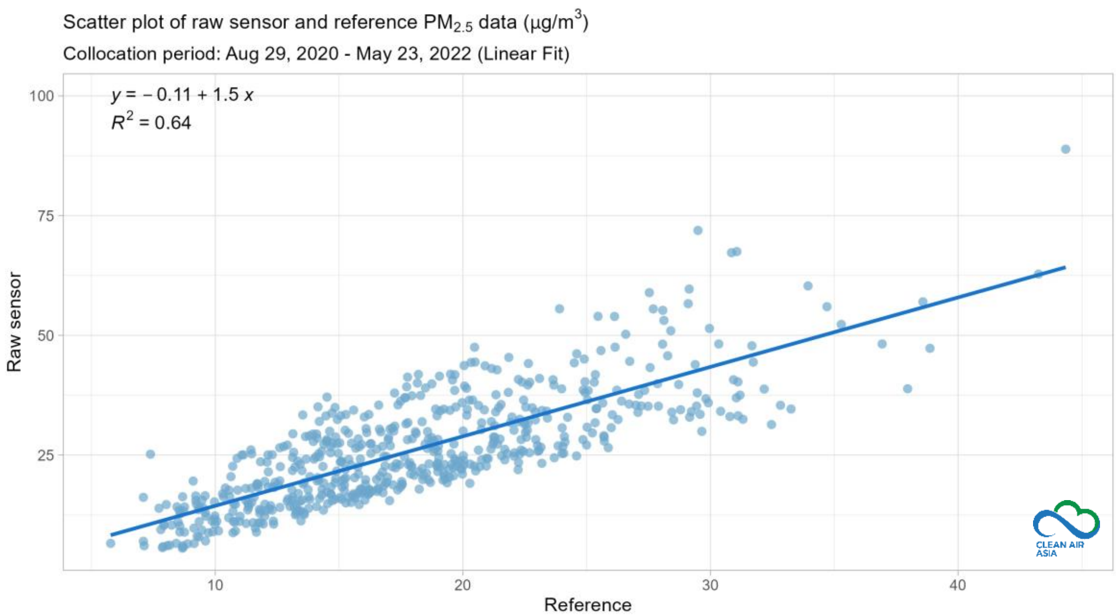
Data import

Monitoring data from the reference and sensor instruments were in hourly resolution and were averaged daily before importing in matching time stamps.

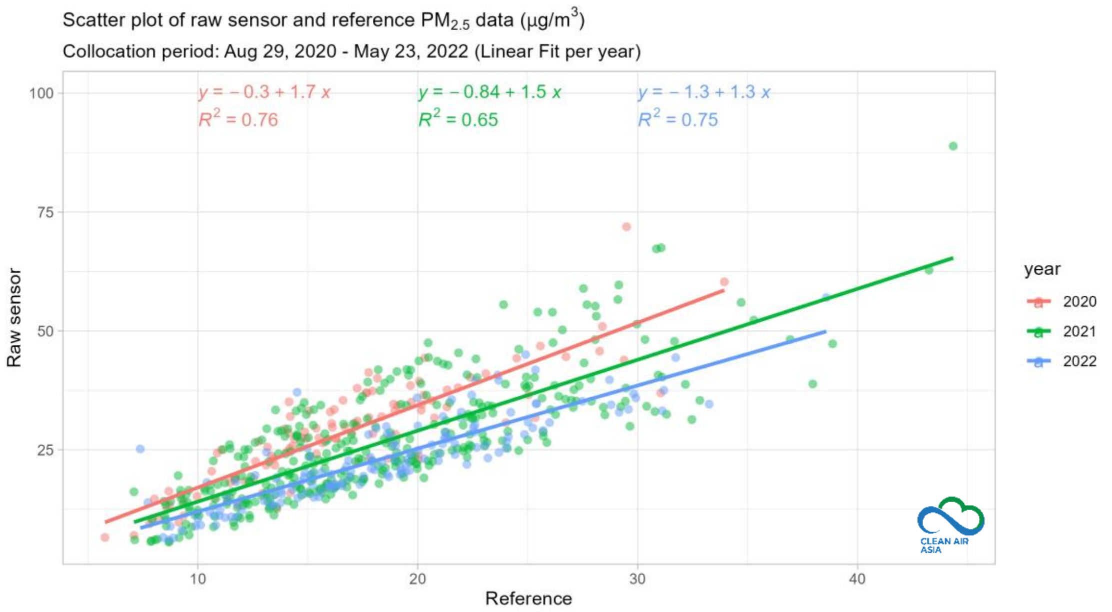
Linear regression

The scatter plot of the data shows a positive relationship between the $PM_{2.5}$ levels measured by the reference station and the sensor. From the plot it can be inferred that as the concentration goes higher, the spread of the data, thus the standard error (SE), increases.

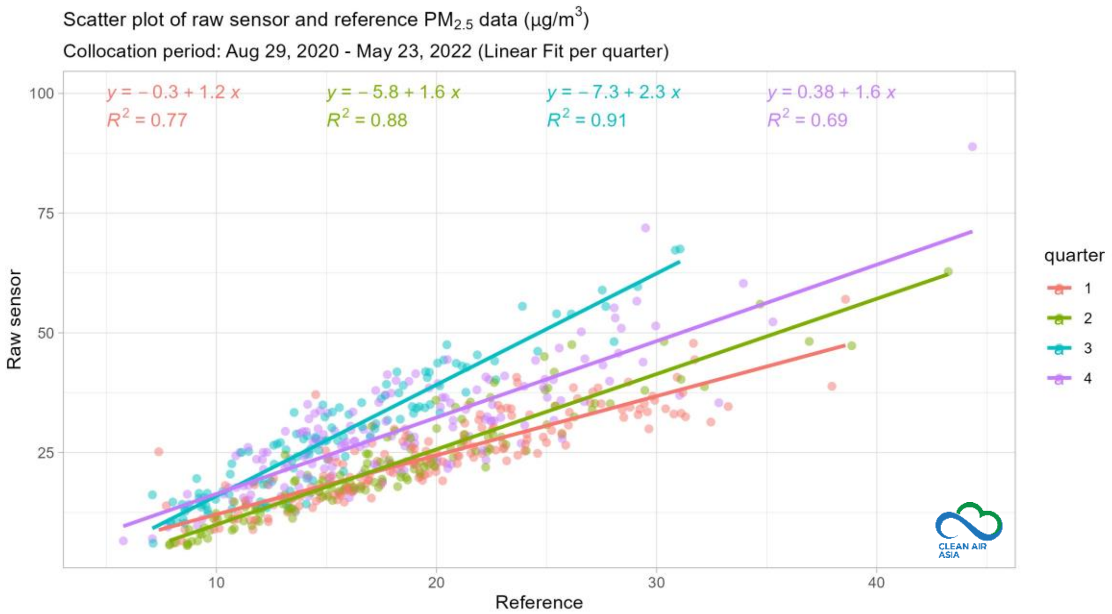
The scatter plot of the data shows varying degree of the correlation strength depending on the period of assessment. The coefficient of determination (R^2) is the value which shows the proportion of variance in the dependent variable (raw sensor data) can be predicted or explained by an independent variable (reference data). If the whole collocation period of Aug 2020 to Oct 2022 is considered, the R^2 is 0.6383813 while the equation of the line is y = 1.4506x - 0.1147.



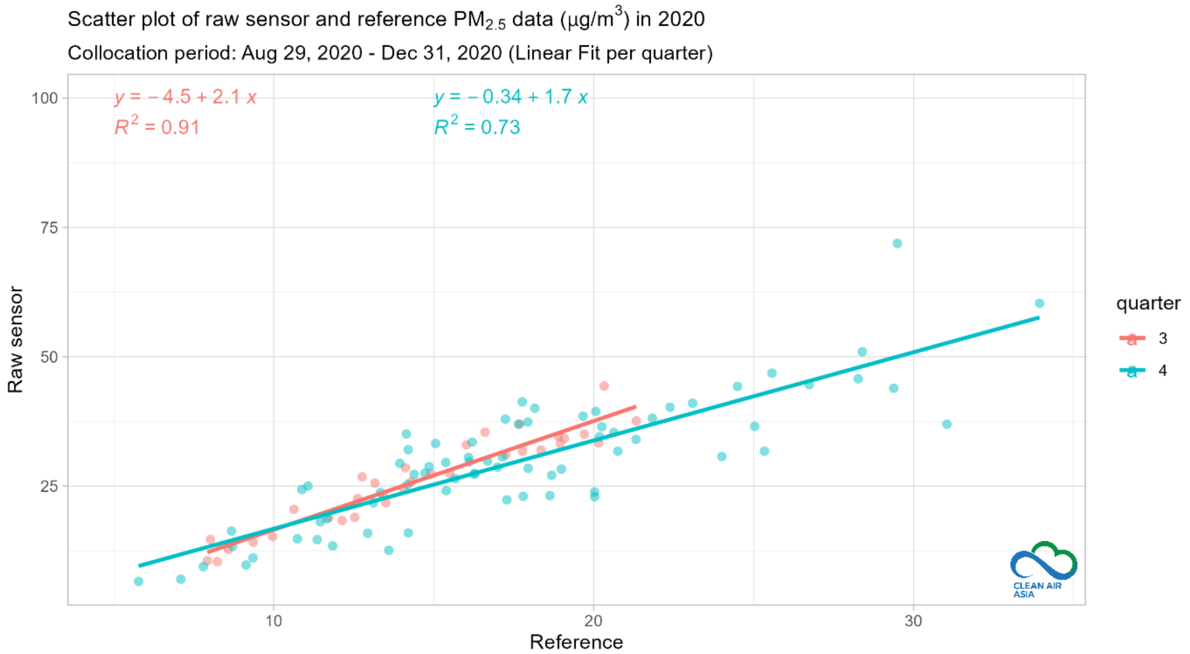
The general expectation is that the performance of sensors declines through time. Since the R^2 is one of the direct metrics of sensor agreement with the reference station, comparing it across the years can be done. The plot shows best R^2 for the dataset in 2020 (0.76), as shown below. Although the value decreased in 2021 (0.65), it increased again in 2022 (0.75).

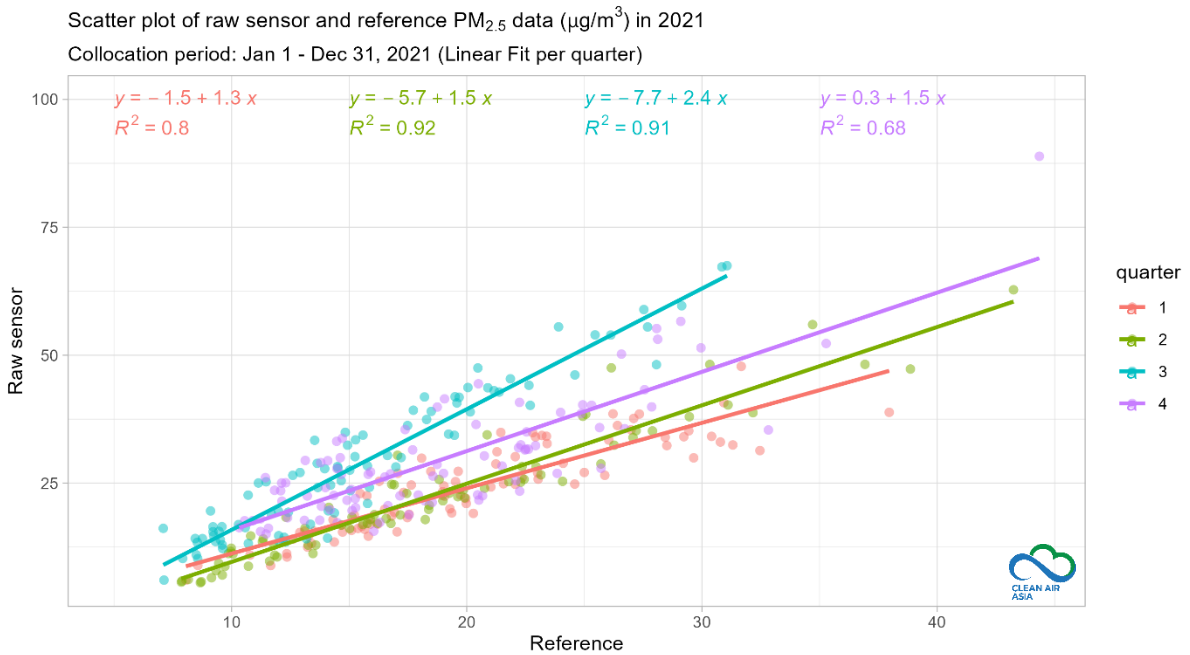


However, quarterly values can also provide more insight as shown in the next plot. From September to December, higher activities and mobility are observed due to the holidays and special events especially in Mehan Garden. Higher $PM_{2.5}$ concentrations have been observed during this time of the year. When comparing the R^2 values, the 4th quarter also shows the lowest R^2 and the highest measurements from both instruments.



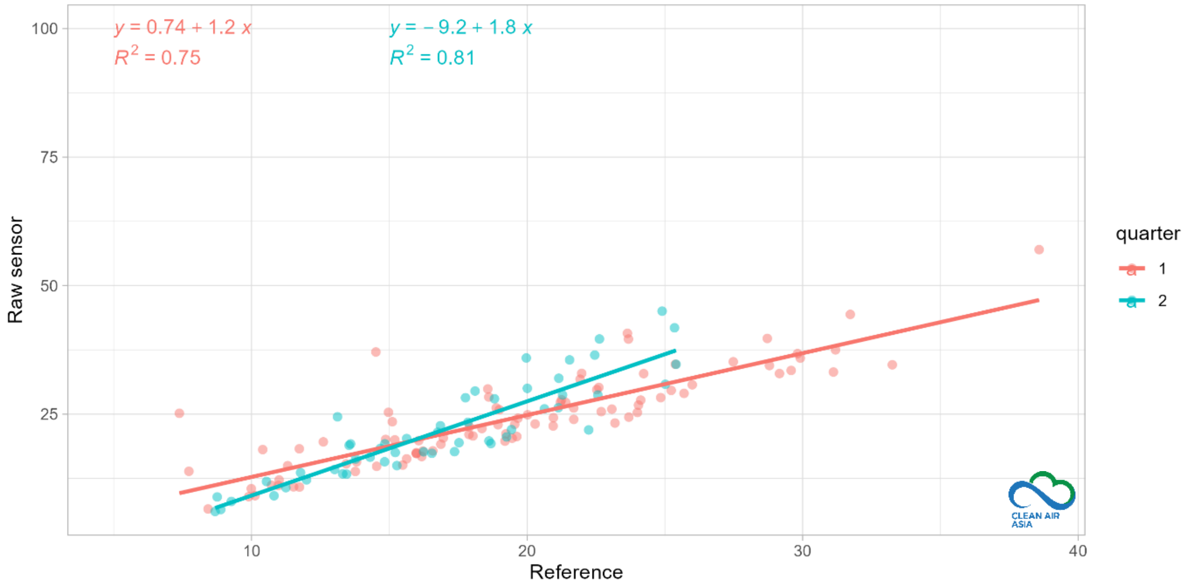
To look for more closely which quarter has characteristic R^2 values from 2020-2022, the quarterly correlation was analyzed per year, as shown below. In 2020, R^2 decreased from 0.91 to 0.73 from the Q3 to Q4. In 2021, Q4 also has the lowest R^2 value at 0.68, compared to the rest that were all 0.8 or above. In 2022 however, R^2 value increased from 0.75 to 0.81 for Q1 and Q2, respectively. This shows no apparent decrease in sensor and reference data agreement through time.





Scatter plot of raw sensor and reference $PM_{2.5}$ data (µg/m³) in 2022

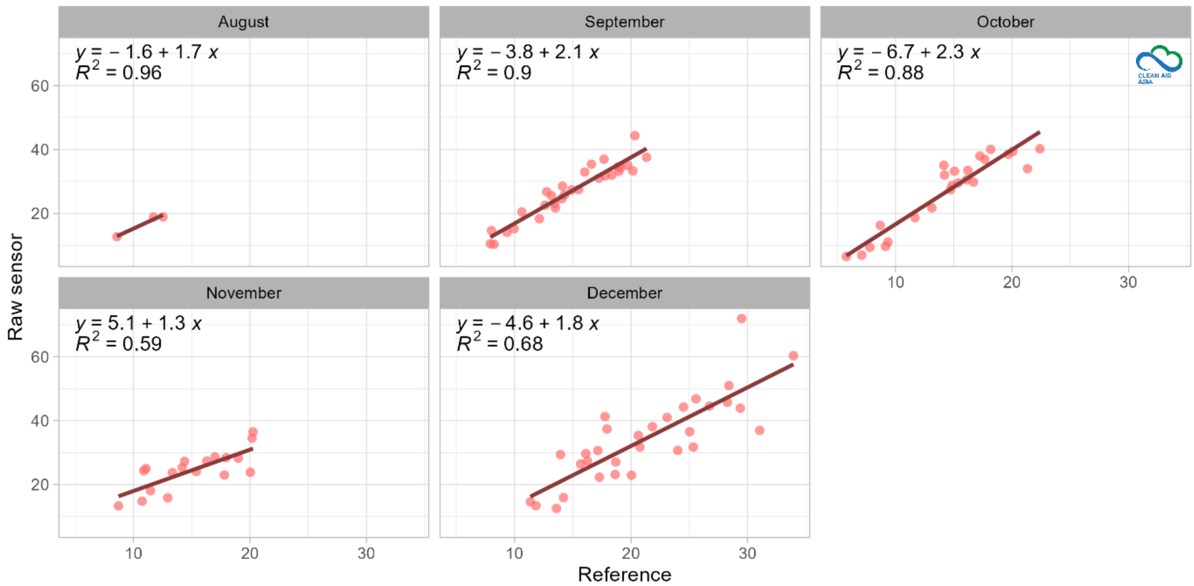
Collocation period: Jan 1 - May 23, 2022 (Linear Fit per quarter)



Air quality data can vary per month, and analysis of the correlation between the ra w sensor and reference datasets can provide more specific information on which specific months lower R^2 can be observed. In 2020, lowest R^2 were observed for months November (0.59) and December (0.68). In 2021, December R^2 was only 0.43, while in 2022 March had the lowest value at 0.66.

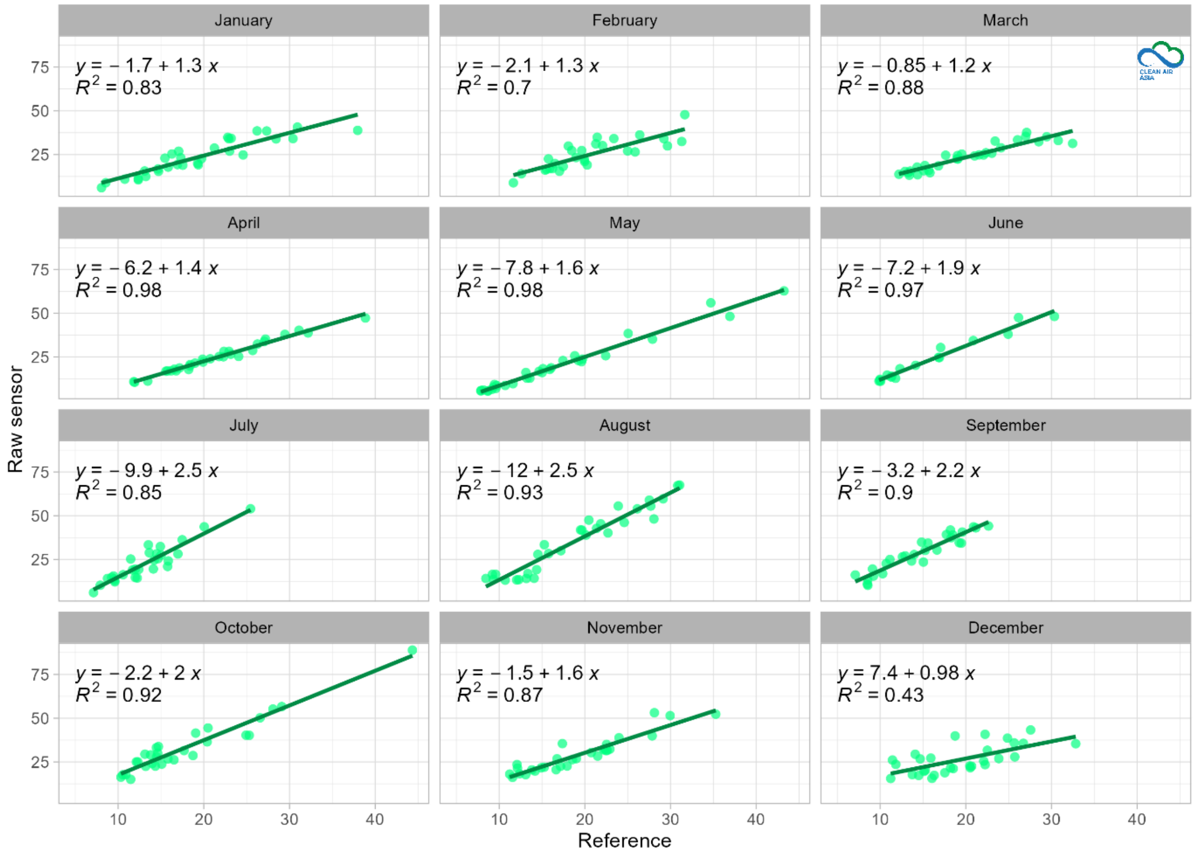
Scatter plot of raw sensor and reference $PM_{2.5}$ data (µg/m³) in 2020

Collocation period: Aug 29, 2020 - Dec 31, 2020 (Linear Fit per month)



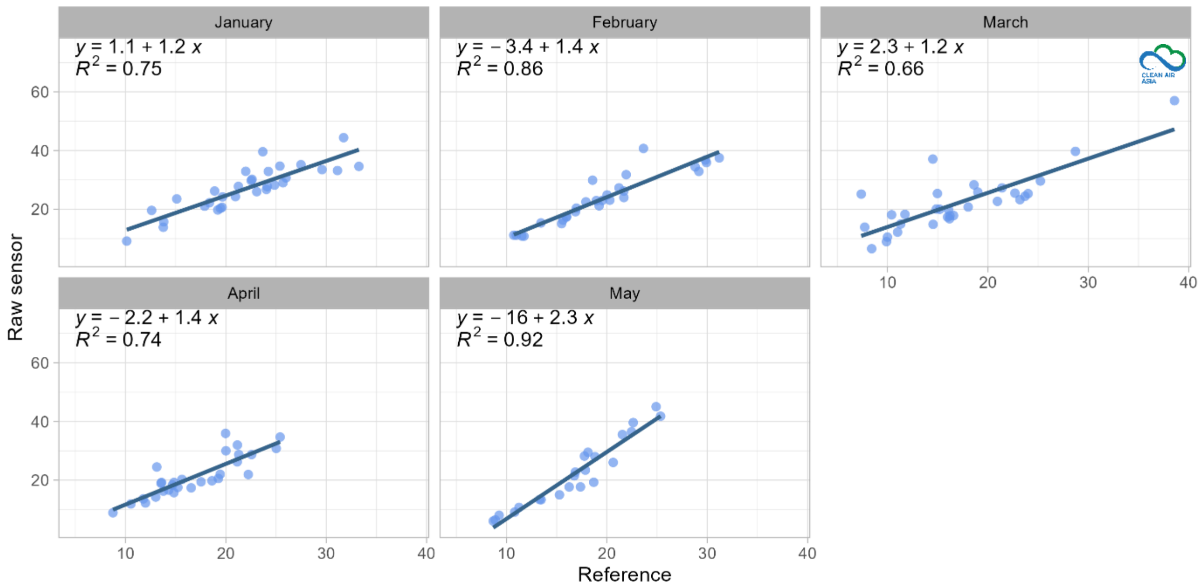
Scatter plot of raw sensor and reference $PM_{2.5}$ data (µg/m^3) in 2021

Collocation period: Jan 1 - Dec 31, 2021 (Linear Fit per month)



Scatter plot of raw sensor and reference $PM_{2.5}$ data (µg/m^3) in 2022

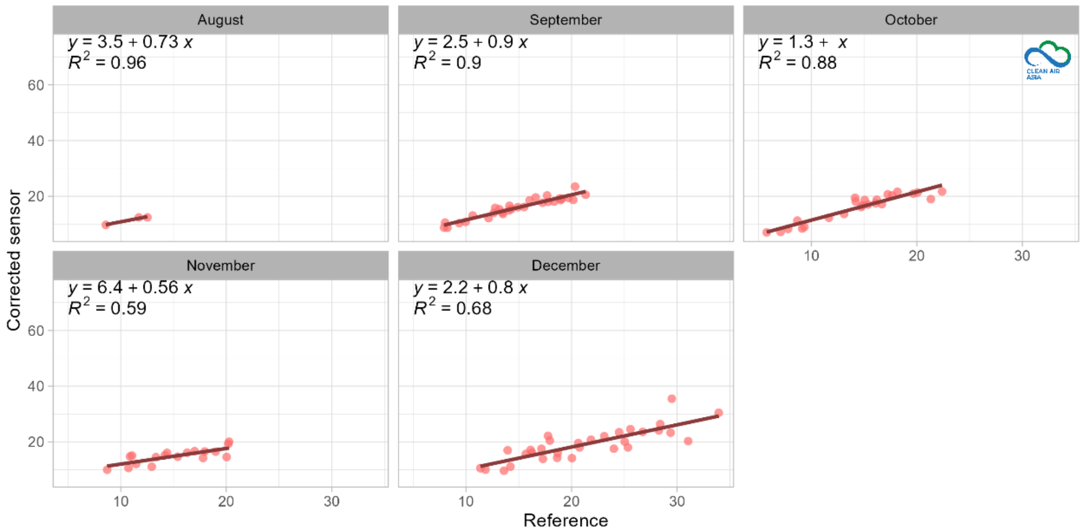
Collocation period: Jan 1 - May 23, 2022 (Linear Fit per month with data)



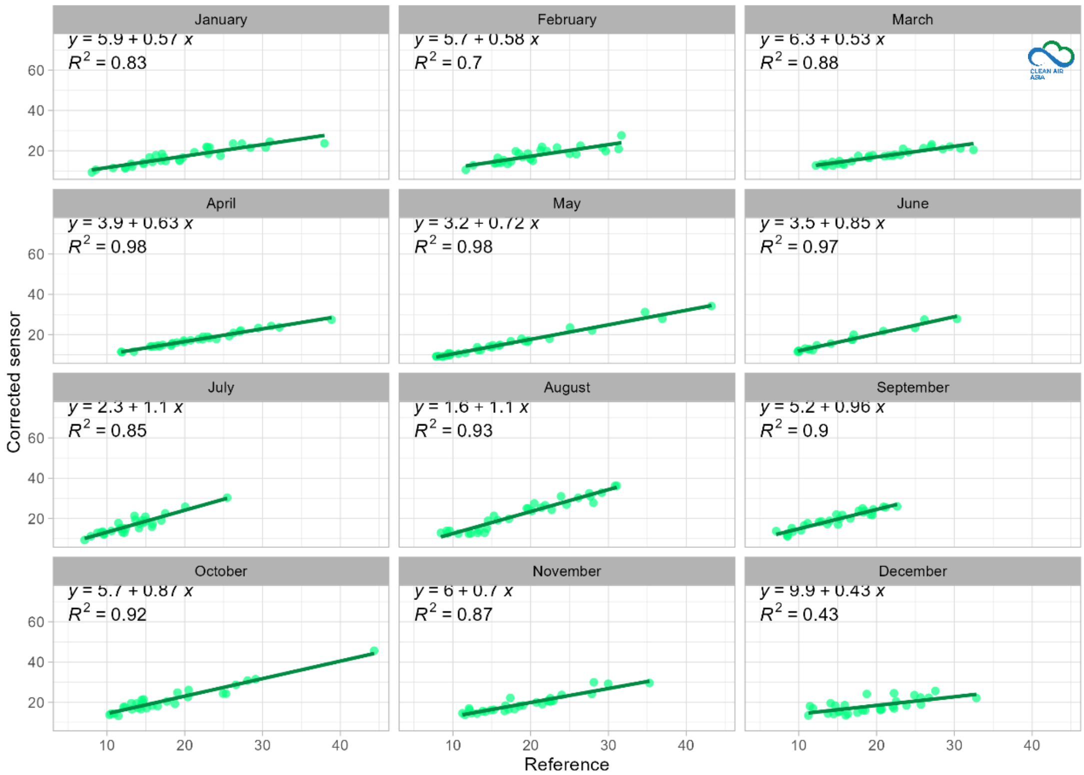
Data correction using linear regression

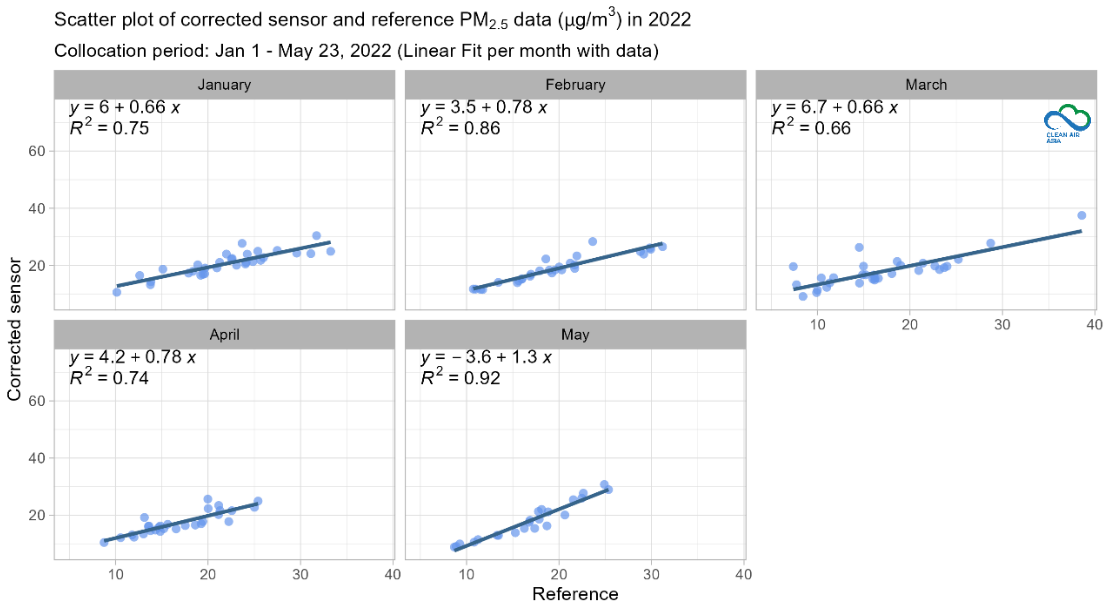
To correct the raw sensor data, the linear regression results per year will be applied for the year of measurement. This means using 2020 linear regression equation and R^2 to correct raw sensor values measured in 2020, and so on. To check correlation of the corrected sensor data and the reference instrument, scatter plots per year are shown below.

Scatter plot of corrected sensor and reference $PM_{2.5}$ data (μg/m^3) in 2020
Collocation period: Aug 29, 2020 - Dec 31, 2020 (Linear Fit per month)



Scatter plot of corrected sensor and reference $PM_{2.5}$ data (μg/m^3) in 2021
Collocation period: Jan 1 - Dec 31, 2021 (Linear Fit per month)





Calculating the root mean square error (RMSE) and mean average error (MAE) for the corrected sensor values show that the linear regression done in this analysis has better outcomes (lower RMSE and MAE) than using the internal correction of Clarity. **This means that, when possible, linear correction must be used instead of the built-in correction of the Clarity system**. (Note that the corrected values were based on per year correction, not overall data)

Correcting sensor data: Difference from reference station

Correction approach	RMSE	MAE
Linear regression	3.532009	2.715967
Clarity built-in correction	4.275194	3.278569

Data correction using multiple linear regression

Sensor performance can also be influenced by temperature and relative humidity (RH). A multiple linear regression analysis was also performed for the collocation data, using the temperature and RH data from the sensor. The reference for the analysis is https://www.ldeo.columbia.edu/~danielmw/code/MLR-Tutorial.html.

Summary of MLR equation coefficients per year

Year	(Intercept)	pm2.5_sensor.raw	temp_sensor	rh_sensor
2020	60.9460388	0.5247176	-0.8371624	-0.4606217
2021	46.2722370	0.5074841	-0.5580834	-0.3491031
2022	57.7104561	0.6856612	-0.7851730	-0.4698224

Correlation and MAE before and after MLR correction of sensor

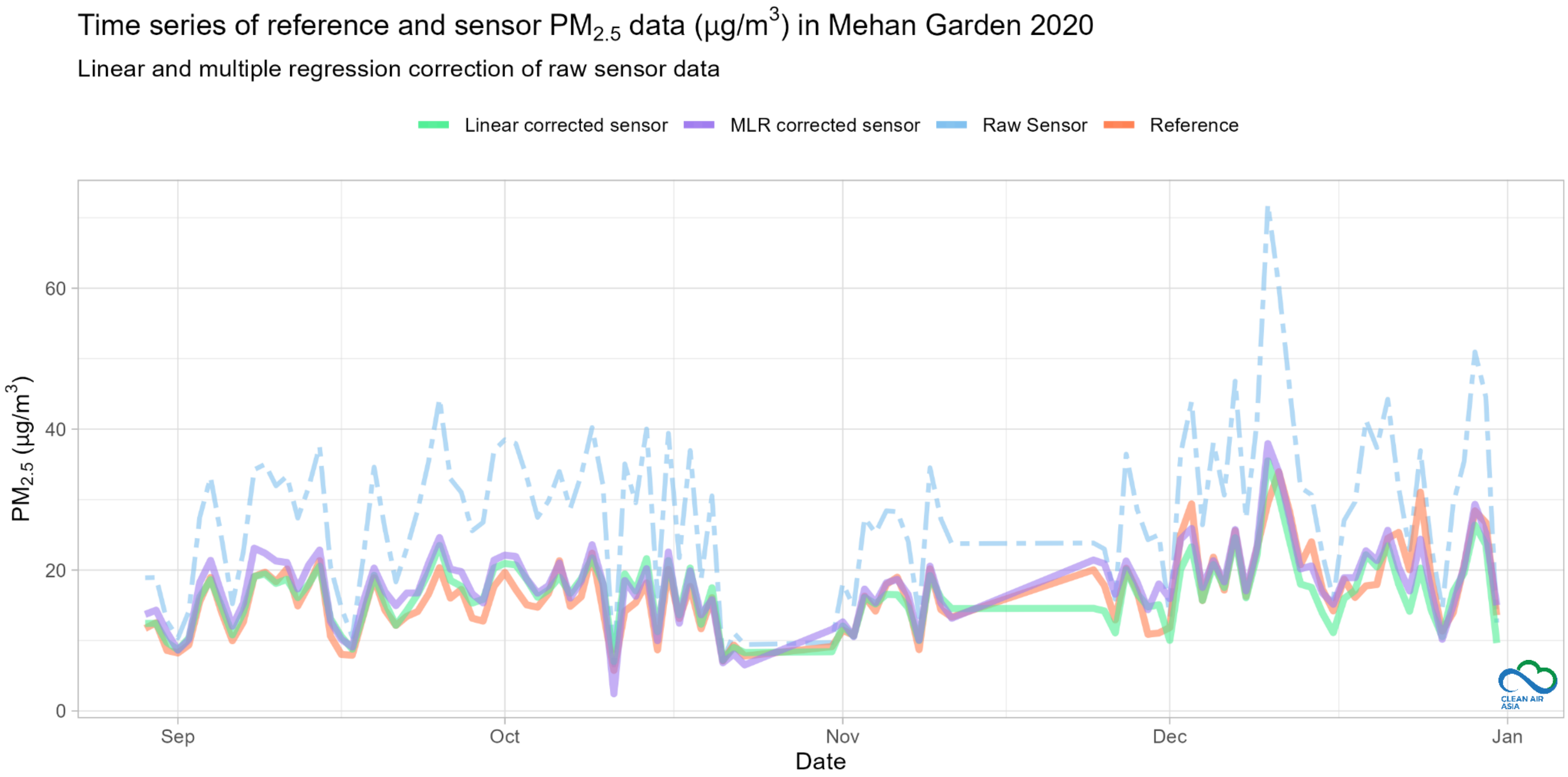
Sensor data	Pearson (R)	Pearson^2 (R^2)	MAE
Before MLR correction	0.7989877	0.6383813	8.199052
After MLR correction	0.9018855	0.8133975	2.148919

Comparison of all corrected values

The comparison of pearson coefficient (R^2), root mean square error (RMSE) and mean average error (MAE) show the **need for a multiple linear regression correction of raw sensor data to ensure higher accuracy of air quality measurements**.

Correction approach	R^2	RMSE	MAE
None (raw)	0.6383813	11.02748	8.199052
Linear regression	0.6911611	3.601986	2.796485
Clarity built-in correction	0.6490282	4.265189	3.283891
MLR	0.8133975	2.74652	2.148919

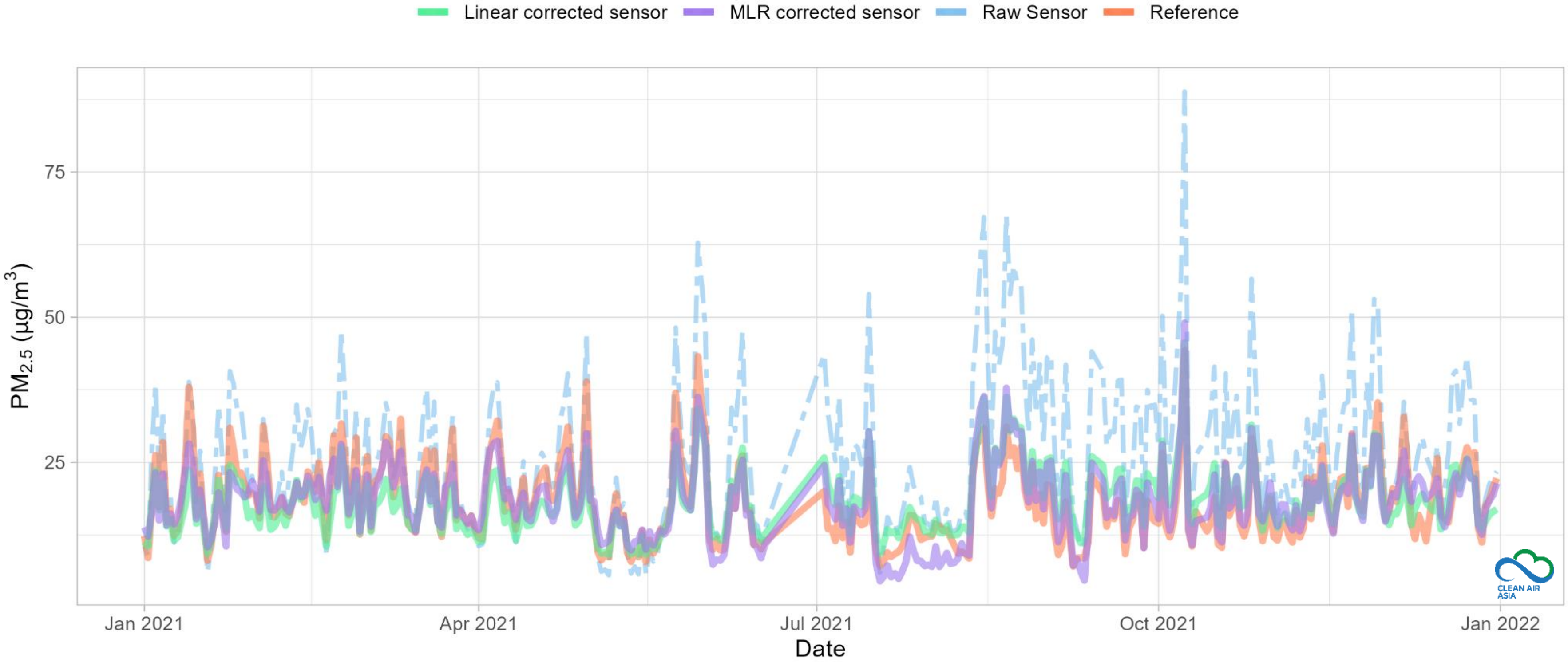
Time series plots show that the correction of data must be done to make sensor readings as close as possible to reference station measurements. Sensor data corrected through MLR is closest to the reference station measurements.



Time series of reference and sensor PM2.5 data (µg/m3) in Mehan Garden 2020
Linear and multiple regression correction of raw sensor data
Linear corrected sensor
MLR corrected sensor
Raw Sensor
Reference
PM2.5 (µg/m3)
60
40
20
0
Sep
Oct
Nov
Dec
Jan
Date
CLEAN AIR ASIA

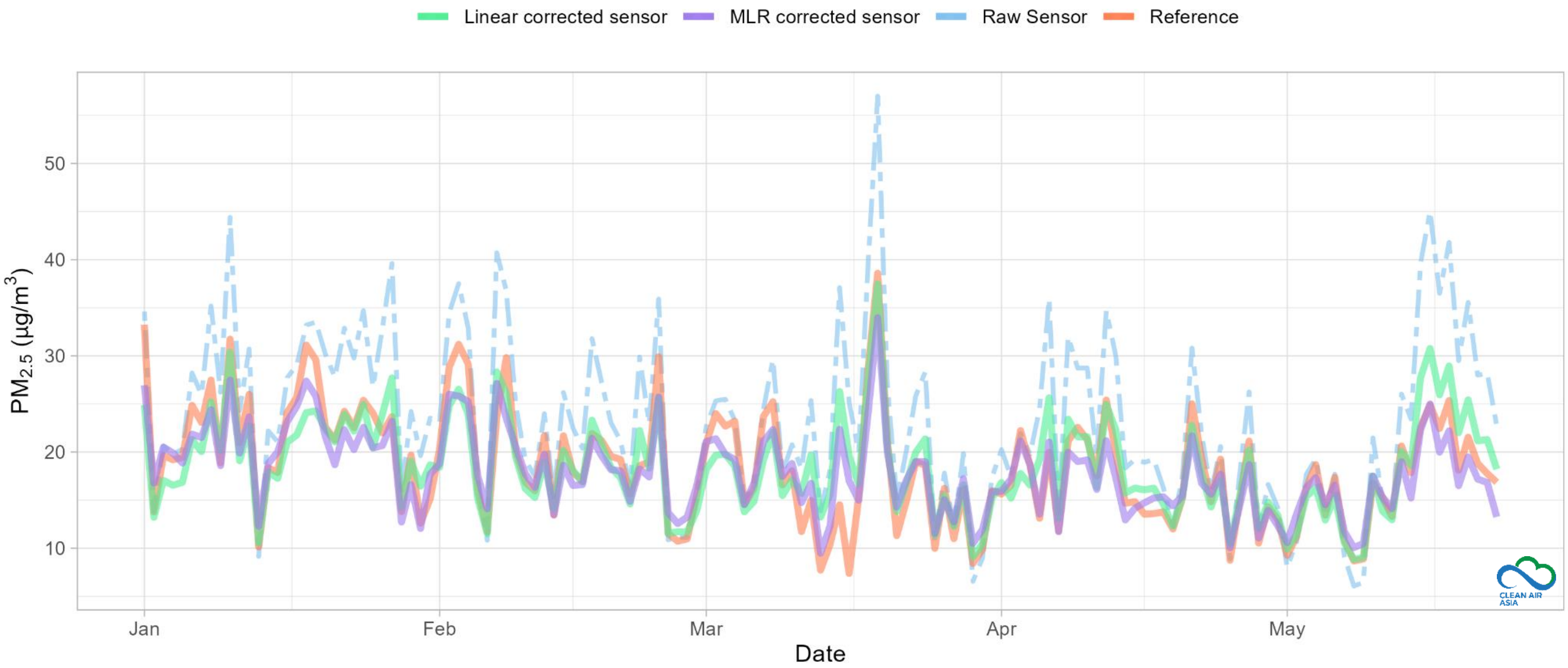
Time series of reference and sensor $PM_{2.5}$ data (µg/m³) in Mehan Garden 2021

Linear and multiple regression correction of raw sensor data



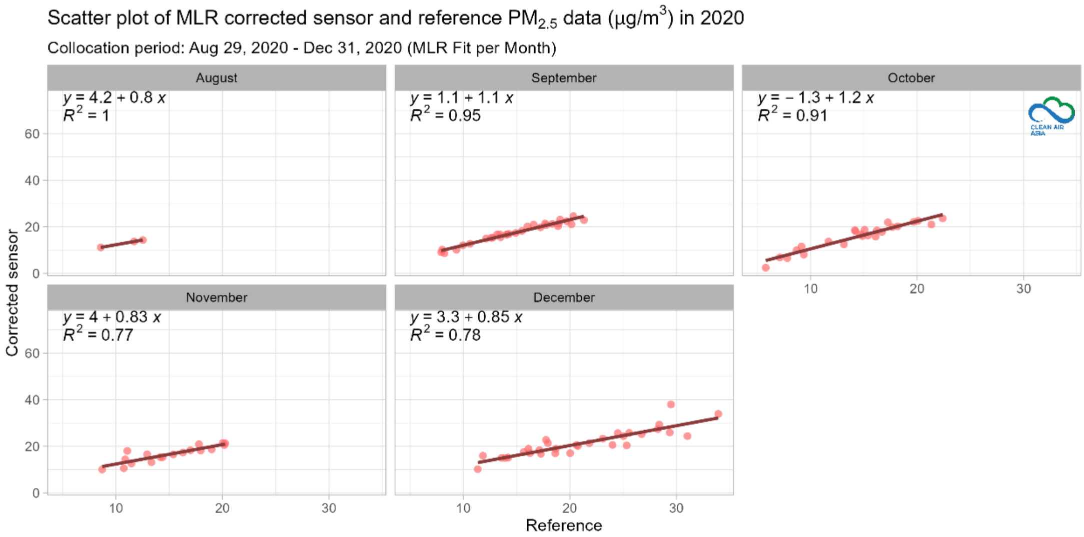
Time series of reference and sensor $PM_{2.5}$ data ($\mu g/m^3$) in Mehan Garden 2022

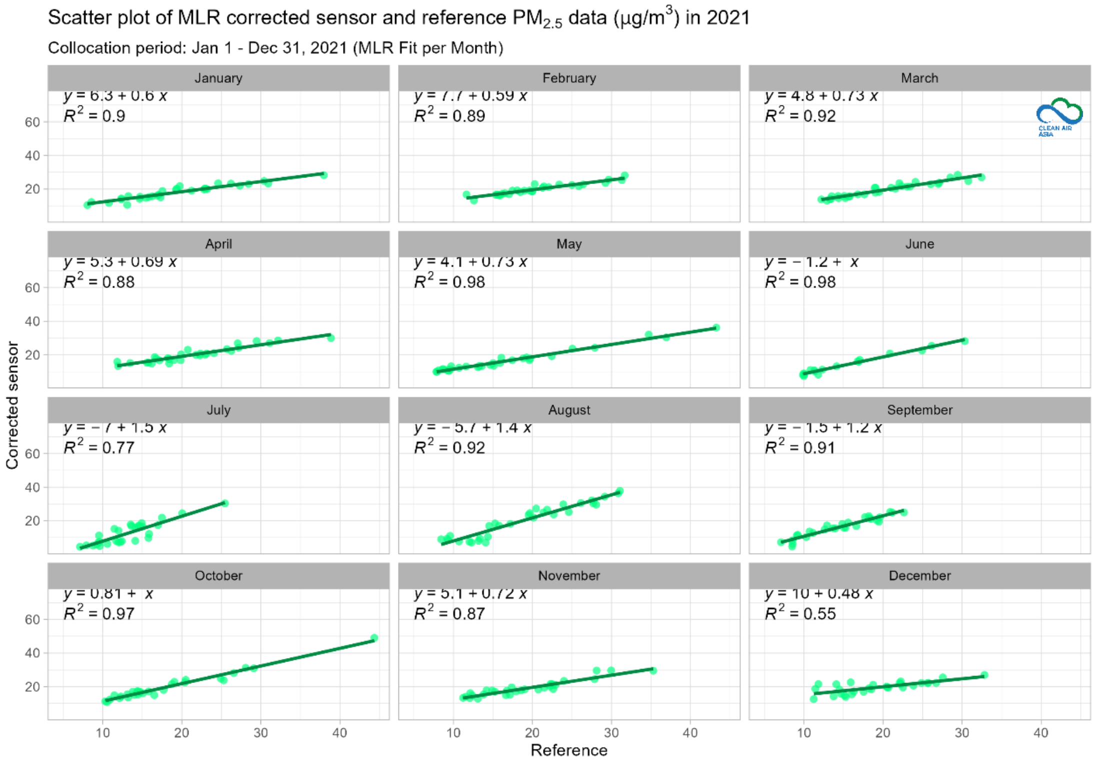
Linear and multiple regression correction of raw sensor data

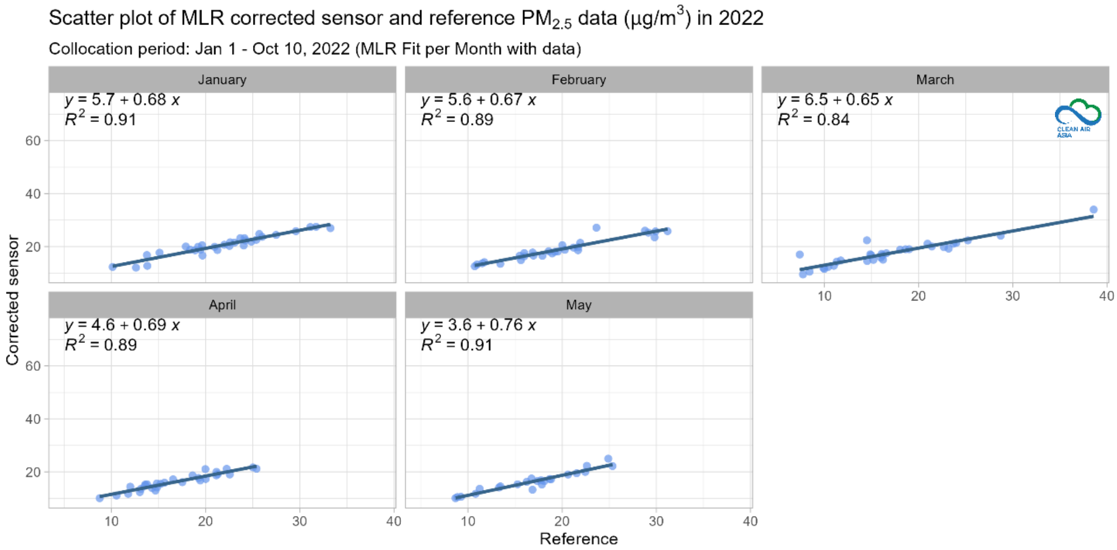


Overall relative accuracy and comparison of averages

The DENR published a Memorandum entitled "Guidelines on the Collocation of Air Quality Monitoring Stations/Equipment for Evaluating Air Quality Sensors or any Non-US EPA Reference/Equivalent Method in Ambient Air Quality Monitoring" on July 3, 2023. One of the metrics is the monthly R^2 that must be ≥ 0.8. In 2020, this criterion was met from Aug - Dec. In 2021, all months met this criteria except for December (0.6). In 2022, this criterion was met for all months from Jan - May.







The average concentration and % Relative Accuracy must also be calculated. In the table below, the values were calculated after the MLR correction. Per year values were provided.

Year	Reference	Experimental (corrected sensor)	Relative Accuracy (%)
2020 (Aug-Dec)	16.6	18	91.6
2021 (Jan-Dec)	18.2	18.1	99.5
2022 (Jan-May)	18.6	18	97.1

Summary, conclusions, and way forward

The collocation analysis presented in this study under the Asia Blue Skies Program is the first and longest-running field collocation experiment performed in order to assess the use of non-reference air sensors to complement the reference-level air quality monitoring network. Upon availability of the reference data in Mehan Garden from July 2022 onwards, the analysis can be further expanded. Results show no apparent decline of sensor performance through time, although lower values for R^2 are observed mostly during months when higher concentrations are measured both by the reference and sensor instruments. It is recommended that performance assessments per concentration range are done in future analyses.

Given the high relative accuracy ranging from 91.6 to 99.5% and an MAE of 2.15 µg/m^3, **it is highly encouraged that these non-reference sensors are used to complement existing monitoring stations, as long as multiple linear regression correction in consideration of temperature and humidity is applied**. This would help in filling data gaps and expanding the spatial (and temporal) coverage of real-time air quality data, information action towards air quality management. For other types of sensors, similar (or more extensive) performance evaluation must be done before use and distribution of data to avoid confusion of the public.

Contact

For questions or more information on this document, please contact **Everlyn Tamayo**, Air Quality and Climate Change Science Lead (everlyn.tamayo@cleanairasia.org) or email aqccmanila@cleanairasia.org and media@cleanairasia.org. *Please do not share this document without proper citation and coordination with Clean Air Asia.*

Clean Air Asia
Unit 3505 Robinsons Equitable Tower
ADB Avenue, Pasig City, 1605, Philippines
Tel +63 2 8631 1042
Fax +63 2 8631 1390
info@cleanairasia.org

Clean Air Asia China Office
11-152, JianGuoMenWai Diplomatic Residence Compound,
No.1 XiuShui Street, ChaoYang District
Beijing 100600 China
Tel/Fax: + 8610 8532 6172
china@cleanairasia.org

Clean Air Asia India Office
Basement C-3, Green Park Extension
New Delhi 110016, India
Tel + 91 11 40395508
india@cleanairasia.org

Country Networks: Indonesia, Malaysia, Nepal, Philippines, Sri Lanka, Vietnam

cleanairasia.org

About Clean Air Asia

Clean Air Asia is an international non-government organization whose vision for the decade is a world with clean air, blue skies, and a stabilized climate for people and the planet. Our mission is to reduce air pollution and greenhouse gas emissions in Asia and contribute to the development of a more sustainable, equitable, and healthier region. We empower governments and cities by inspiring action and focusing efforts on the planning and implementation of solutions at scale, fostering regional cooperation to address issues of mutual concern, and ensuring the integration of air quality management across sectors and institutions.

We work with partners to reduce air pollution and greenhouse gas emissions across Asia by building capacity, advocating for effective and appropriate policies and practice, and informing stakeholders of air pollution and climate change impacts. We aim to reduce air pollution and greenhouse gas emissions in 1000+ cities in Asia through a range of innovative policies and programs covering air quality, transport and industrial emissions, and energy use.

We work with energy, environment, health and transport ministries, cities, the private sector, development agencies, academia and civil society to provide leadership and technical knowledge in Air Quality and Climate Change, and Sustainable Transportation (Low Emissions Urban Development, Clean Fuels and Vehicles, Green Freight and Logistics).

Since 2008, we have been a United Nations-recognized partnership comprised of more than 250 organizations in Asia and internationally, with country networks across the region. Our headquarters are in Manila, Philippines, and we have offices in Beijing, China and New Delhi, India.