Field Evaluation and Collocation Study: Performance Evaluation Metric used in this study for $PM_{2.5}$ Sensors

Sl No	Parameter	Description	Equation	Purpose	Mode of Evaluation	Acceptable Criteria
1	Bias	Systematic measurement error for sensor values with respect to reference	$Bias = \left(\frac{\dot{x}}{\bar{y}}\right) - 1$	To test and improve sensor's data quality	Evaluated before and after field calibration	Should be close to zero
2	Mean	Arithmetic mean patterns of sensor & reference values	$Mean = \frac{\sum_{i=1}^{n} x}{n}$	To predict subsequent data points and baseline trend	Compared daily average patterns of sensor data with the reference system.	N.A.
3	Mean Absolute Percentage Error (MAPE)	Absolute percentage difference between sensor & reference values	$MAPE = \left(\frac{1}{n}\right) \times \sum_{i=1}^{n} \left(\frac{x-y}{y}\right) \times 100$	To assess the accuracy of the sensor-based monitors	Calculated MAPE before and after field calibration	15-30
4	Correlation	Strength of the linear relationship between sensor & reference values by evaluating R^2	$R^2 = \left[\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[\{n\sum x^2 - (\sum x)^2\}\{n\sum y^2 - (\sum y)^2\}]}}\right]$	To measures the strength and direction of the linear relationship	Calculated the best-fitting regression curve between two sets of data i.e., sensor data and reference data.	$R^2 \geq 0.70$
5	Intra-model variability	Descriptive statistics-mean, median, and standard deviation (SD) between collocated sensors	$SD = \sqrt{\frac{\sum_{i=1}^{n}(x - \overline{x})}{n-1}}$	To determine the similarity.	Calculated mean, median, and standard deviation of the same sensors collocated in the same sensors' batch	SD ≤ 5 µg/m³
6	Data Recovery	Data completeness rate	$Data\ recovery\ (\%) = \left(\frac{n_x}{t_x}\right) \times 100$	To be able to acquire reliable and comprehensive data for analysis.	Assessed the percentage of usable sensor data points	>75%

Note-

1. x & y indicate observation points from the sensor and reference grade station respectively
2. n represents the number of data points
3. n_x and t_x are the number of valid sensor data points during the testing period and the total number of data points for the testing period (from start to end).

Source: CII CABL Analysis (forthcoming 2024) based on primary data and USEPA (2021, 2014) & AQ-SPEC (2017)