Data cleaning, analysis and visualisation

When you decide to use low-cost sensors, you should have a clear definition of the purpose and aims of using them. This should inform the analysis you need for your data and how you would like to visualise the results. This may include several levels of objectives. For example, regular management of daily work or task exposures; retrospective analysis over the last month, year, etc; identifying areas for improvement; evaluating control measures; assisting with workers training. Data indicators and analysis may differ based on these objectives.

Before data can be analysed and visualised, data cleaning and correction may be necessary.

Data cleaning and correction

Data cleaning is the process of identifying and fixing errors, inconsistencies, and discrepancies in low-cost sensor data. It's an important step to ensure that data is accurate, reliable, and of high quality, especially when working with large, complex data sets. You should put procedures in place to describe what the data cleaning process will include, and guidance sheet the actions taken to generate the final data set used for analysis.

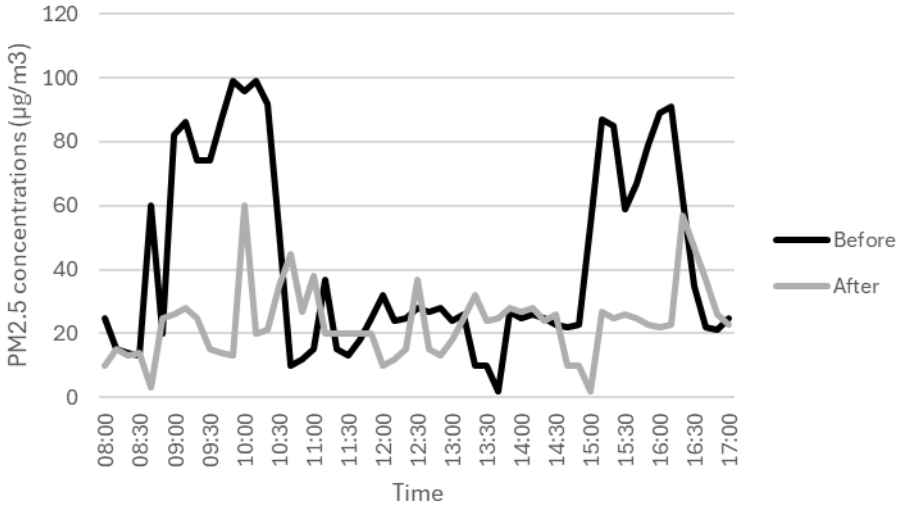
As discussed in the "Quality Assurance and Quality Control" section, it may be necessary to apply correction factors to the data before its analysed so that it is comparable with measurements that would be obtained if using a reference monitoring device. In such instances the correction factors applied to the data should be clearly documented, with details of how these correction factors were generated being described in any accompanying reports.

Data analysis and visualisation

In terms of analysis and visualisation of the (cleaned/corrected) data set, this will depend on the data indicators of interest (e.g., concentrations averaged over a given time-period, exposure peaks, temporal trends over a shift etc), the target audience (e.g., senior management, shop floor workers) and the objective of the communication (e.g., demonstrating a reduction in exposure following adoption of new control measures).

Figure 1 in our guidance sheet "Why use low-cost sensors" shows an example of how low-cost sensors can help identify sources or activities that influence exposure. Figure 2 shows an example of how low-cost sensors can help evaluate the effectiveness of introduced exposure control measures. The figure shows that the PM2.5 peaks have been reduced, and have largely disappeared, following the introduction of local exhaust ventilation over the mixing stations. The use of such visual presentation of the data can also help enhance worker training and reinforce the need for using control measures and the impact these have on reducing worker exposure.

Figure 2: Example of real-time data for PM2.5 collected over two work shifts in a bakery, one before and one after the implementation of local exhaust ventilation over the mixing station.



It is important that the uncertainties surrounding the data, what the data can and cannot be used for, are also clearly identified. The use of the data is very much linked with communication and engagement with the workforce. Further information on this topic is given in our guidance sheet.