Quality Assurance of PEMS Emissions Data aimed for the Development of Real-world Vehicle Emission Factors

Petros KATSIS*, Robin SMIT**, Leonidas NTZIACHRISTOS***, Ting-Shek LO**** & Carol WONG****

** Emisia S.A., Antoni Tritsi 21, P.O. Box 8138, Thessaloniki, GR-57001, Greece*
- email: petros.k@emisia.com

*** School of Civil Engineering, University of Queensland, Brisbane, 4072, Australia*

**** Laboratory of Applied Thermodynamics, Department of Mechanical Engineering, Aristotle University, Thessaloniki, GR-54124, Greece*

***** Environmental Protection Department, the Hong Kong SAR Government*

Abstract

Portable Emission Measurement Systems (PEMS) have recently become widespread in collecting real-world emission information from vehicles. The Environmental Protection Department of Hong Kong (HKEPD) has been carrying out vehicle emission measurements by means of PEMS. The collected data are used to generate real-world emission factors for the particular environment and vehicle types. For this reason, emission information needs to be verified in a systematic manner by using extensive Quality Assurance / Quality Control (QA/QC) methods.

In this study, the authors attempt to devise proper criteria to verify PEMS data measurements and subsequently develop corresponding automated algorithms aiming to highlight a variety of potential abnormalities such as single or clustered outliers, analyser measurement saturation and zero drift.

Results show that the proposed set of detection and verification algorithms can be used for PEMS data to create a robust, efficient and cost-effective screening protocol that will guarantee a quality level, adequate enough to check the data prior to using them for emission factor development. Visual time-series analysis is still a necessary tool to rule out any prediction errors.

Key-words*: (5 words) PEMS, vehicle measurements, quality assurance, quality control, real-world.*

Introduction

Portable Emission Measurement Systems (PEMS) have been under the spotlight as of late, in view of the developments and resulting need for accurate emission measurements in realistic driving conditions. PEMS systems are versatile emission and exhaust flow measurement equipment with the capacity to provide accurate, second-by-second emission information for gaseous and particulate pollutants. PEMS are installed on board the vehicle and are powered by an independent source (on-board batteries). With PEMS on, each vehicle can be operated in its usual driving conditions, while its emissions are being measured. PEMS therefore provide realistic emission information of high resolution, which is difficult, if not impossible, to obtain otherwise.

As a result of the high resolution, PEMS generate large datasets with pollutant emission rates as well as other secondary information. From the perspective of accurate vehicle emission prediction, a large database of empirical data is essential to adequately reflect the large variability in real-world emission profiles from different vehicles and different engine and driving conditions. However, it is also vital to use verified emissions data in emission factor development to prevent prediction errors. Given the large amount of data, efficient and effective screening methods are needed to filter out erroneous measurements. Even data from high quality test facilities such as certification-grade laboratories require scrutiny and verification as there are several steps in data collation that can lead to errors, e.g. manual data entry typos and erroneous data can still be overlooked and included in the database.

The Environmental Protection Department of Hong Kong (EPD) has collected and continues to collect a large number of vehicle emission measurements utilizing PEMS. The large number of emission data produced even on a single vehicle are useful to check the emission performance of the particular vehicle, but also, more generally, to understand emission phenomena. Combined with emission measurements from other vehicles and properly processed, these data may generate real-world emission factors for the particular environment and vehicle types. For this reason, the collected emission information needs to be verified and organized systematically; the sheer quantity of available PEMS data requires effective quality control and assurance (QA/QC) methods that use specific verification statistics and automated computation and visualisation techniques to ensure invalid data are flagged, corrected or removed, while ensuring that valid outliers are maintained.

The Hong Kong landscape combined with the driving regulations and typical traffic conditions constitute a measurement environment with certain characteristics; it is mainly an urban area with a hilly terrain. The speed is controlled by traffic lights and is usually maintained below 50km/h in urban areas and between 80 and 110 km/h in highways, while the speed limit for heavy duty vehicles is set to 70 km/hr.

The PEMS measurements conducted by EPD are mainly focused on NO_x and PM emissions. Moreover, On-Board Diagnostics (OBD) are not readily provided for most vehicles. Because of Hong Kong landscape particularities, i.e. high buildings in dense urban environment, GPS recordings may contain gaps due to satellite signal loss. Due to that, speed measured by a speedometer on the wheel of the tested vehicle with dead reckoning is used alongside GPS. Finally, the entire procedure for PEMS data measurement, collection and processing is consistent with the requirements of CFR1065, ISO16183 and Regulation (EU) No 582/2011 with corresponding UNECE Regulation No 49.

EPD has already in place an established method to collate and organize data from different PEMS measurement campaigns. With regard to measurements, five main components are being recorded:

- Gaseous pollutants, including CO, CO_2, NO, NO_2, and total hydrocarbons (THC), utilizing a number of different PEMS systems of different manufacturers and generations.
- Integrated particulate matter (PM) and real-time PM, again utilizing different PEMS units.
- Pollutants speciation and non-regulated pollutants, like NH_3, N_2O using Fourier-Transform Infrared Spectroscopy (FTIR).
- Global Positioning System (GPS) data, including position, altitude and speed.
- Vehicle relevant information, including speed measured on the wheel of the tested vehicle and OBD data if available.

Vehicles equipped with PEMS devices and peripherals will follow a target vehicle with the similar characteristics, which will then operate on its regular driving route (car chasing). Apart from this method, PEMS-equipped vehicles can also follow a fixed course. Both of these approaches are used for all vehicles except buses.

The current measurement sample exceeds 250 vehicles in several categories by using a variety of PEMS and other monitoring and recording devices over a range of routes and operational conditions.

Various data summary statistics can be used in a quality assurance and screening process to identify unrealistic and suspicious data. They range from basic statistics such as ‘maximum value’ and ‘number of missing values’ to the results produced by specific verification algorithms. A challenge with the use of summary statistics is the determination of accurate and robust pass/fail criteria. These criteria are developed through inter-comparison of PEMS data with vehicle-specific data, and, in some cases, for several vehicles within a particular vehicle class.

In this paper, we present different QA/QC procedures in order to verify the PEMS data being collected by EPD. The analysis focused on second-by-second data of gaseous and particulate pollutants. The proposed screening protocol applies algorithms to detect issues in the overall test statistics and includes data pre-processing, speed smoothing, driving behaviour verification and basic statistical analysis of emission rates. A second step examination focuses on specific pollutant traces and includes specific routines for clipping detection, statistical analysis indices, baseline drift detection and outlier detection. The criteria behind these procedures aim to pinpoint a variety of abnormalities such as single or burst outliers, analyser measurement saturation and zero drift.

1. Methodology

The proposed methodology applies verification procedures in two consecutive phases, each applied at a different level of detail:

- verification of overall test validity and integrity, and
- verification of individual pollutant trace integrity.

The concept of this approach is first to subject the data to a general screening, so as to remove any obviously compromised measurements and ensure that abnormalities are detected and flagged or corrected; this would often mean that entire tests would be discarded or modified as a whole. Then, the second phase deals with more specific issues, focusing on individual pollutant traces to identify occurrence of errors either at a test level or at a time stamp level (single errors or bursts of errors).

In order to apply the verification protocol, first a number of statistics have to be calculated using the original dataset. The proposed screening protocol is designed to handle issues on gaseous and particulate real-time (second-by-second) PEMS data measurements in conjunction with real-time speed data measurements. The two distinct levels of verification steps will hereinafter be denoted as Phase 1 (overall test level) and Phase 2 (individual trace level) respectively, while the preparatory step will be referred to as Phase 0. A summary of the approach is presented in Figure 1. Details for each procedure are provided in the following paragraphs.

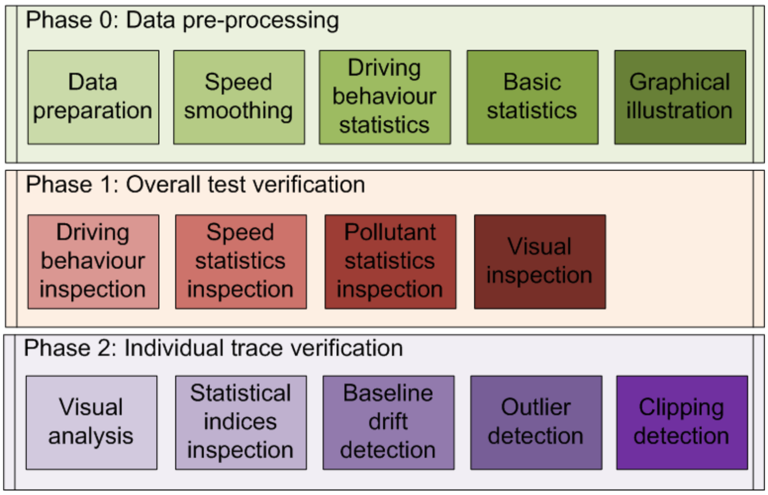

Figure 1. Screening protocol process

Phase 0: Data pre-processing

This first step addresses time gaps developed when intermediate time-stamps are missing from the PEMS data set e.g. due to transcription errors. This step enables the application of the follow-up verification phases, which require complete time series at least split on a sub-trip level. Sub-trips are defined as continuous and uninterrupted periods of travel time for a unique vehicle. Sub-trips are identified by computing the time difference (Δt) between consecutive time stamps, and flagging $\Delta t > 1$ s.

The addition of this metadata yields continuous second-by-second values within each sub-trip; this is necessary to allow speed smoothing to take place and more importantly to calculate acceleration values as part of the Phase 1 verification.

Speed smoothing may be optionally applied to the speed-time column to account for measurement noise and to prevent unrealistic computations of acceleration and engine power, in particular at higher speeds. A T4253H smoothing can be used in this approach (Velleman 1975, Velleman 1980); T4253H is a nonlinear data smoothing filter which can provide a practical method of finding smooth traces for data confounded with possibly long-tailed or occasionally spiky noise. The smoothing algorithm is resistant to the effects of extreme observations that are not part of the local pattern and capable of responding rapidly to well-supported patterns. Speed-time smoothing is also conducted at the sub-trip level within each vehicle data test to prevent unrealistic jumps at sub-trip end points.

In the next part, statistics pertaining to second-by-second speed and individual pollutant traces will be extracted. It is useful to examine derived variables as they can readily reveal unrealistic recordings, which can then be traced back to particular issues with the PEMS measurements. The following speed trace summary statistics are computed for each unique vehicle data test:

- idle time (s),
- proportion of idle time (-),
- number of missing values in the speed trace (s),
- proportion of missing values in the speed trace (-),
- total length of all trips (s),
- average absolute deviation (AAD) in between raw and smoothed speeds (km/h),
- maximum absolute deviation (MAD) in between raw and smoothed speeds (km/h),
- minimum/maximum speed (km/h),
- minimum/ maximum acceleration (m/s^2), with acceleration being computed using smoothed speed.

A number of basic statistics are computed for the pollutant emission rates (expressed in g/s) at test level:

- minimum, mean, maximum (MIN, MEAN, MAX),
- proportion of zero values (PZERO),
- proportion of missing values (PNA),
- proportion of negative values (PNEG),
- standard deviation (SDEV),
- coefficient of variation (COV),
- peak-to-mean ratio (PTM).

Figure 2 illustrates an example of the basic statistics table for PM sec-by-sec data, containing the metrics described above.

trip	min	mean	max	pzero	pna	pneg	sdev	cov	ptm
PM_1	0	0.0014	0.0418	0.0	0.0	0.0	0.004	2.96	30.5
PM_2	0	0.0023	0.1000	0.0	0.0	0.0	0.007	2.97	44.3
PM_3	0	0.0026	0.0602	0.0	0.0	0.0	0.005	2.10	23.0
PM_4	0	0.0012	0.0725	0.0	0.0	0.0	0.005	3.89	62.2
PM_5	0	0.0023	0.1024	0.0	0.0	0.0	0.009	3.81	45.4
PM_6	0	0.0018	0.0942	0.0	0.0	0.0	0.007	3.89	51.8
PM_7	0	0.0017	0.0697	0.0	0.0	0.0	0.006	3.41	41.3

Figure 2. Example of summary statistics table (PM)

Finally, before proceeding to the actual verification algorithms, visualisation of empirical data is an effective way of summarizing various key aspects of PEMS data, and facilitating further detailed analysis of potential issues. Time-series and other summary plots need to be generated for each vehicle data set for different variables, and saved for possible use in the screening process.

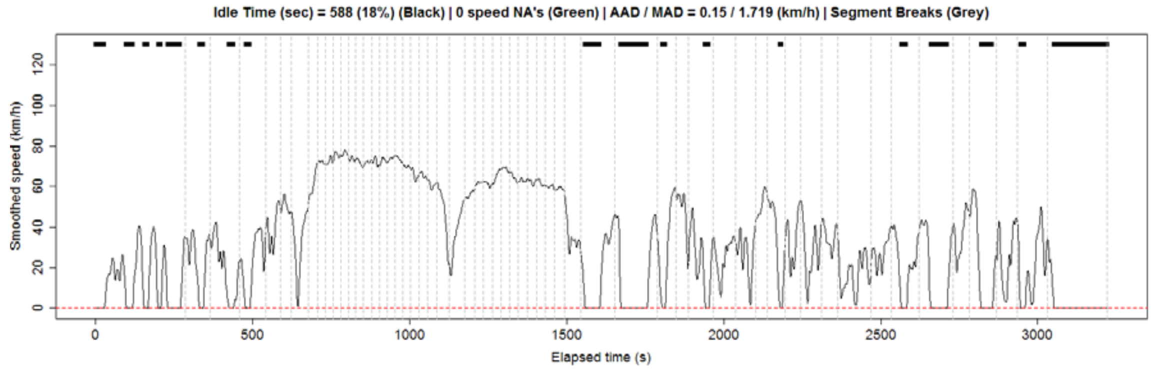

Figure 3. v-t-power plot example

Figure 3 shows an example of a 'v-t-power' plot which depicts the vehicle speed versus the elapsed test time derived from a PEMS data test. Coloured bars are used in these plots to provide additional time-series information, such as 'idling' (black bars), and 'missing/not available values' (green bars, not visible in this plot). The vertical grey dotted lines show 500 m drive segments. On the top of this figure, additional average statistics, as noted previously, are illustrated: Idling time and its proportion vs. the total test time, the average absolute deviation between raw and smoothed speeds and the maximum absolute deviation in between raw and smoothed speeds.

Phase 1: Overall test validity and integrity

The goal of this verification phase is to study the pre-processed data provided after the completion of Phase 0 in order to investigate whether basic trip metrics such as speed, acceleration and pollutant basic statistics at an aggregated test level fall within plausible ranges. The approach is valid only for real-time, sec-by-sec measurement data.

The overall test integrity procedures are expected to target the following issues:

- Instrument errors, including calibration and zero/span issues, data gaps in the recordings, misalignment in synchronization and time scales, measurements below detection limits, over the range (instrument saturation) values, etc.
- Errors in the transcription of the different files, such as value separation issues, digits changes (often confusion with point or comma decimal separator), unknown or variable units scales, lack or erroneous transfer of manual information stored in log files, etc.
- Vehicle operation specific abnormalities, e.g. cold start operation, regeneration modes, in particular for DPF equipped vehicles, and specific operation windows.
- Protocol gaps, including lack of correct monitoring of environmental conditions (e.g. on-road water load during rainy conditions), lack of monitoring of vehicle and instrument preconditioning patterns, etc.
- Other reasons which may be random or systematic, but may not be possible to identify (i.e. specific instrument errors in the recordings, PEMS system faults, etc.)

A number of procedures were developed to address these issues. The procedures fall in the following four categories:

1. Operation feasibility verification
2. Speed trace statistics verification
3. Pollutant trace average statistics verification
4. Visual time-series inspection

The first three procedures provide methods to verify statistics related to the speed trace or pollutant traces at an aggregated level (overall test). The fourth procedure can be used to verify the validity of the potential issues raised by procedures 2 and 3. These procedures are analysed below.

Operation feasibility verification

Ideally, second-by-second engine power could be used to identify unrealistic vehicle operation. This could be done by using either measured or computed instantaneous engine power and normalizing these values over rated engine power. Any exceedance of a value of 1 (plus an error margin, e.g. 25%) would indicate unrealistic operation, i.e. an erroneous recording. The issue with this approach is that there is often a lack of accurate road gradient information to be used to predict second by second engine power.

For this reason, an alternative approach was implemented. The operation feasibility verification algorithm aims to flag extreme accelerations for vehicles falling into different power-to-weight ratios (PTW). In order to accomplish this, a model was devised to portray the relationship between acceleration and vehicle class. The maximum possible acceleration is expected to be a function of rated engine power, or rather PTW ratio. For example, heavy duty vehicles (HDVs) with a low PTW ratio will not be able to accelerate as quickly, as light duty vehicles (LDVs) with a high PTW ratio.

First, a range of typical drive cycles per vehicle type and driving mode were designed (Smit, 2006), on the basis of statistical analysis of smoothened speed-trace data of the vehicles in the sample. Different driving profiles were designed according to the vehicle PTW ratio, defined as rated engine power divided by Gross Vehicle Weight (GVW).

Examples of this procedure are shown in Figure 4 for passenger cars and buses, which represent two different PTW classes. This is essentially a scatter plot of all individual speed points and associated accelerations of the drive cycles. It is clear that the maximum possible acceleration is a function of both instantaneous speed and vehicle class.

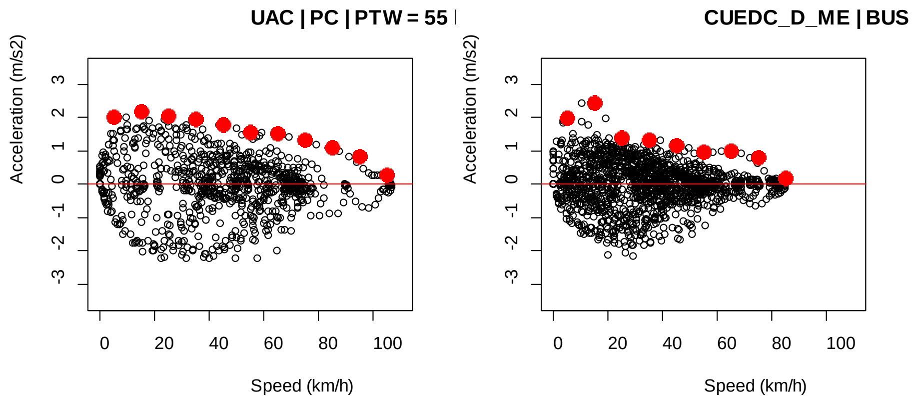

Figure 4. Scatter plots of instantaneous speed versus acceleration (Left: PC; Right: BUS).

To complete the model, first the maximum acceleration was computed for 10 km/h speed bins for each PTW class. These values are depicted as red dots in Figure 4. A non-linear regression model was fitted to these data. It was assumed that a reverse sigmoid function would best describe a maximum feasible acceleration model for each drive cycle. Figure 5 shows the results of the regression for four PTW classes and two vehicle types, i.e. HDVs (PTW of 8 or 15 kW/ton GVW) and LDVs (PTW of 25 or 55 kW/ton GVW). In general, the reverse sigmoid functions appear to fit the data well, in particular for the HDVs, but they may overestimate maximum feasible acceleration at higher speeds (>90 km/h). An error margin of 30% was added to the model and this constitutes the limit above which acceleration appears infeasible.

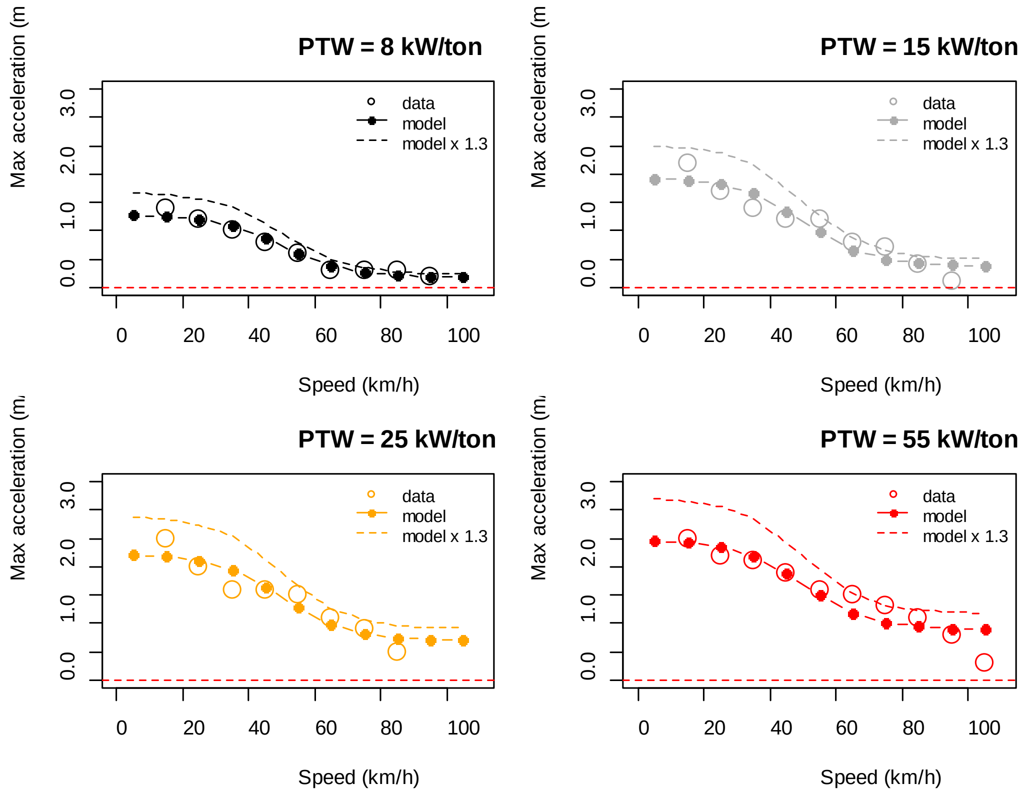

Figure 5. Maximum feasible acceleration models

The final models for LDVs and HDVs extracted via this process are shown by the following equations:

$$\boldsymbol{LDV}: a_{max} = (-0.25 + 0.01\ PTW) + 1.30\left(0.70 + \frac{1.53}{1 + EXP\left(\frac{v - 50}{10}\right)}\right)$$

$$for\ PTW \ \geq 25\ kW/ton\ GVW$$

$$\boldsymbol{HDV}: a_{max} = (0.43 + 0.07\ PTW) \times 1.30\left(0.18 + \frac{1.11}{1 + EXP\left(\frac{v - 50}{10}\right)}\right)$$

$$for\ PTW \ \geq 8\ kW/ton\ GVW$$

Here, a_{max} represents the maximum feasible acceleration (m/s^2), *v* is the instantaneous vehicle speed (km/h) and *PTW* is the ratio of rated engine power to GVW.

With the maximum acceleration – PTW – vehicle class modelling in place, the operation feasibility verification procedure only requires the vehicle information, which is casually available with the PEMS data. With this data, the verification procedure calculates the PTW of the vehicle and classifies the vehicle as either LDV or HDV. It then uses the respective equation above to verify whether the calculated acceleration values of the PEMS data are acceptable or not, flagging such test/time stamps as having a suspicious driving behaviour.

Speed trace statistics verification

This procedure was used to investigate whether other basic speed-time statistics comply with expected values:

- The proportion of idling time may indicate a problem in the test (measurement issue).
- The proportion of missing values in the speed trace.

- Large deviations between the smoothed and original speed traces.

The exact limits for these metrics may need to be customized to the studied data set, but e.g. if the missing speed values percentage within a test is close to 100%, it should probably be discarded.

Pollutant trace average statistics verification

This procedure requires a blanket screening of basic pollutant trace statistics in the overall test level to conduct a high level scan of the integrity of pollutant measurements and identify any issues. This verification procedure uses the minimum/maximum pollutant recording in the dataset and the proportion of zero values, negative values and missing values in the entire test duration.

This inspection allows the identification of traces within each data test that should not be used. The specific limits need to be calibrated on each measured dataset, but some general guidelines are quite straightforward:

- When maximum/minimum values are way off expected limits. This is vehicle- and test-dependent (driving conditions).
- When the proportion of missing values reaches 100% (e.g. no data if all data are missing). Obviously, a lower value should also invalidate the values of this trace in the data test.
- When the proportion of negative values is too high. This also raises the issue of how to proceed, i.e. whether the test will be discarded or if remedial action is needed before further usage (e.g. correction of negative values by zeroing or marking them as missing).
- For tests with too many zero values, similar action could be taken, especially in comparison with the speed behaviour. Some situations are quite intuitive, e.g. a large percentage of zero CO_2 values while not idling is suspicious.

Visual time-series inspection

In Phase 1, the overall test examination focuses on the visual examination of second-by second speed values. While, the previous verification procedures may suggest potential issues, a visual inspection is necessary to corroborate such erroneous behaviour and flag or correct those tests or identify situations where average statistics check may not be sufficient.

Moreover, the computed summary statistics guide the selection of datasets that require visual examination and vice versa to identify untypical errors. For instance, visual examination may point out issues that were actually missed by the operation feasibility verification procedure.

Phase 2: Individual data trace validity and integrity

This phase specifically focused on second-by-second pollutant data trace validity and integrity. The individual pollutant trace verification algorithms included:

- Use of test statistics values such as coefficient of variation and peak-to-mean ratios of recorded data. A challenge with these new statistics is to create accurate and robust pass/fail criteria, as no generic limits can be applied.
- The identification of transient operation conditions, such cold start, fuel enrichment periods, regeneration patterns for DPF devices, etc.
- Methods to detect untypical analyser drift, which may lead to shift in the measured levels or even saturation.

The proposed verification procedures used for the data trace validity and integrity testing were:

1. Visual time-series analysis (pollutant trace level),
2. Clipping detection,
3. COV-PTM analysis,

4. Baseline drift detection,
5. Outlier detection.

Visual time-series analysis

Information-dense time-series plots are created for each individual pollutant trace, and they can be called upon when necessary to examine second-by-second data traces combined with the speed-time behaviour.

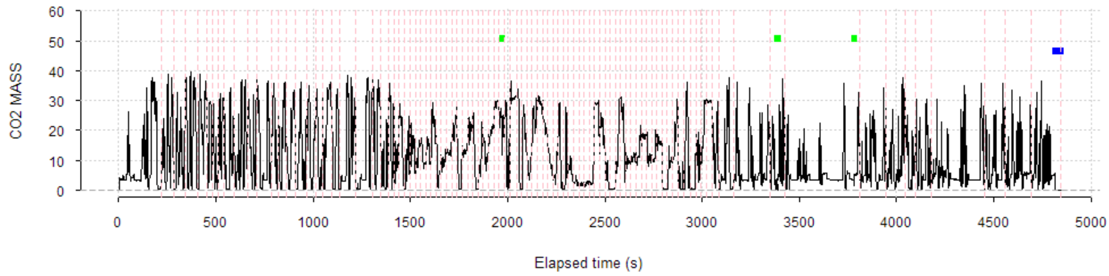

Figure 6. CO_2 rate time-series plot.

In these plots, blue bars indicate zero values, green bars indicate missing values and red bars indicate periods with constant and high values. These situations may be indicative of periods with e.g. clipping and DPF regeneration. This figure also contains additional statistics to enable a combination of visual and statistical analysis.

Clipping detection

Clipping is defined as the replacement of measured concentrations with the analyser's maximum range value, when the concentration exceeds the analyser's range. Clipping results to characteristic plateaus for a few seconds in the measured signal. Clipping may occur e.g. due to improper calibration during the PEMS spanning process or if the measured quantity trace is actually outside the instrument range.

Clipping correction is required when high constant values are monitored for a given period. In this approach "clipping" is quantified with vector C, which is computed as follows:

$$C = e/m$$

where *e* and *m* represent the actual and maximum measured emission rate in g/s for a particular pollutant and particular vehicle. An indicator which illustrates how frequently clipping appears in a test is calculated by dividing the number of flagged values by the total trip duration; this indicator is noted as PCLIP.

COV-PTM analysis

Tests with relatively low PTM and COV values may reflect a significant proportion of clipped emissions data, or have small emission peaks, or experience a combination of both. Tests with relatively high PTM and COV values are indicative of data with large emission peaks ('spiky' data). This is the context around which this verification procedure is built.

COV is computed as the standard deviation divided by the mean. COV is a normalized measure of the dispersion in the data. PTM is a dimensionless indicator reflecting the relative magnitude of spikes in the data. This 'validation by comparison' approach is an effective way to visually identify outliers in large datasets that warrant further analysis. It is noted that the PTM/COV data can vary substantially with pollutant, as well as with engine and emission control technology. The data should therefore be examined for each pollutant individually, and the data need to be categorized using an appropriate vehicle classification.

Baseline drift detection

A measurement drift algorithm was developed to quantify analyser zero drift and at the same time indicate the possibility for other emission events such distinction between cold/hot conditions. Measurement drift represents a shift in measured values under the same conditions at different points in time. Significant drift results in erroneous emission results, so this test parameter is quite important for emissions level accuracy. Despite the fact that measurement drift is usually automatically verified before and after a test, it is still necessary to identify tests with potential issues.

The verification procedure aims to extract emission values for homogeneous engine conditions at different time intervals throughout the test. To quantify the level of measurement drift, idling segments need to be identified first. An idling segment is defined as a chronological time period of a few seconds duration with engine on and zero vehicle speed. The measured emission rates in these idling segments are then extracted. A certain amount of time before and/or after the idling period may be omitted to prevent boundary effects related to transient engine operation and extract stabilized emission rates.

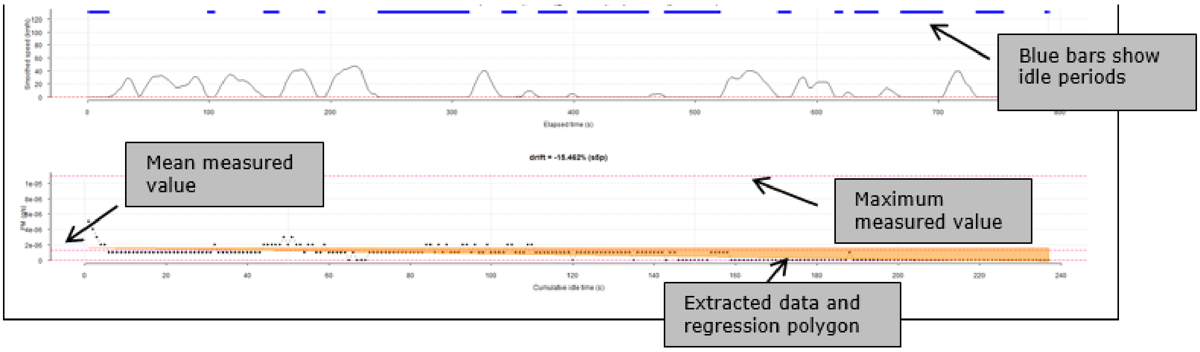

Figure 7. Example of speed and baseline drift time-series plots.

A robust least-squares linear model (RLM) is then fitted to the extracted data, and a drift variable D is computed as follows:

$$D = 100\, n\, b/m$$

where *n* represents the number of extracted idling data points (s), *b* represents the regression coefficient in g/s and *m* represents the maximum measured emission rate in g/s for a particular pollutant and particular test. RLM is not sensitive to outliers, and is expected to better quantify any general trends.

Therefore, *D* quantifies the percentage change in idling emission rates over time relative to the peak emission value. It can produce meaningful results for time-aligned hot running tests with a significant number of idling segments. Cold start tests will create large *D* values due to highly elevated emission levels at the start of the test. Finally, the 95% confidence interval (CI) for *D* can be compared with the intercept to verify if *D* is statistically significant (i.e. D is outside the 95% CI). An example of this procedure is shown in Figure 7.

Outlier detection

Outliers are observations with characteristics that are distinctly different from the other observations. Before PEMS data are used in, e.g. emission factor development, it is vital to identify these outliers and determine if they are valid data points. They can also have a significant impact on emission factors. Outliers can also arise from specific situations or issues, including errors that were produced during testing (e.g. incorrect measurement settings) or during the data transfer and reporting phase (e.g. data entry error). Outliers can be determined in various ways such as via examination of univariate and multivariate distributions and at different scales. The scale of outlier assessment can be at data trace level (single test), vehicle level (all tests) or vehicle class level (all pooled test data for all vehicles with similar characteristics).

The method proposed here focuses on univariate data trace level, as any outliers detected and verified at this fundamental level, will prevent propagation of data issues to higher scales of assessment. Nevertheless, it is still recommended that emission factors based on PEMS data are compared at vehicle and vehicle class level to ensure any suspicious data are investigated further.

It is noted that observations may occur normally in the outer ranges of the distribution, so the analysis attempts to identify those distinctive observations and designate them as outliers. Outliers can be detected using ‘standard z-scores’. A z-score represents the distance between the observation and the mean in terms of the number of standard deviations. The z-score is negative when the observation is below the mean value and positive when it is above.

There is, however, one issue with this approach. Vehicle emissions are typically highly skewed with long tails to the right, reflecting occasional emission spikes. On the other hand, outlier detection using z-scores assumes an approximately normal distribution. As a consequence, PEMS emissions data need to be transformed before outlier detection can be applied. The aim is to achieve a more or less symmetrical and normal distribution of observed emission rates.

Due to the large number of data sets, the Box-Cox procedure has been used to automatically select the best data transformation from a family of power transformations, and compute the transformation variable 'lambda' (λ). Transformed data equals the actual data to the power of lambda where, by definition, $\lambda = 0$ suggests a log-transformation.

The z-scores are computed for pollutant traces for each test as follows:

$$d_i^* = 1 + d_i + |\min(d_i)|$$

$$d_i^{*\prime} = (d_i^*)^{\lambda}$$

$$z_i = (d_i^{*\prime} - d^{*\prime})/s_{d*\prime}$$

where d_i is an actual pollutant data point, d_i* represents a vector of shifted d_i values, d_i*' represents a vector of transformed d_i* values, z_i represents the computed z-score for transformed data point *i*, *d**' and s_d*' are the mean and standard deviation of the vector of d_i*' values, respectively. Using the previous conversion, the minimum of d_i* is achieved when d_i is minimum or:

$$\min(d_i^*) = 1 + \min(d_i) + |\min(d_i)| \geq 1$$

The d_i* term is computed in order to create data values greater than zero; this is required for the Box-Cox procedure, which includes a log transformation of d_i* (thus the logarithm must be higher than zero). Z-scores have a mean of zero and a standard deviation of one. Z-score values are then used to determine if an observation d_i qualifies as an outlier. An example of this transformation is illustrated in Figure 8.

In addition, the impact of removing outliers is quantified to assess the relevance of the outliers with respect to computed emission factors (expressed in g/km). The relevance is computed as the ratio of the test averaged emission factor with outliers removed (expressed in g/km) to the emission factor with all data included. Significant deviation from unity indicates remarkable impacts of the outliers. An outlier relevance impact greater or equal to 0.95 is not considered significant given the uncertainties in the PEMS measurements.

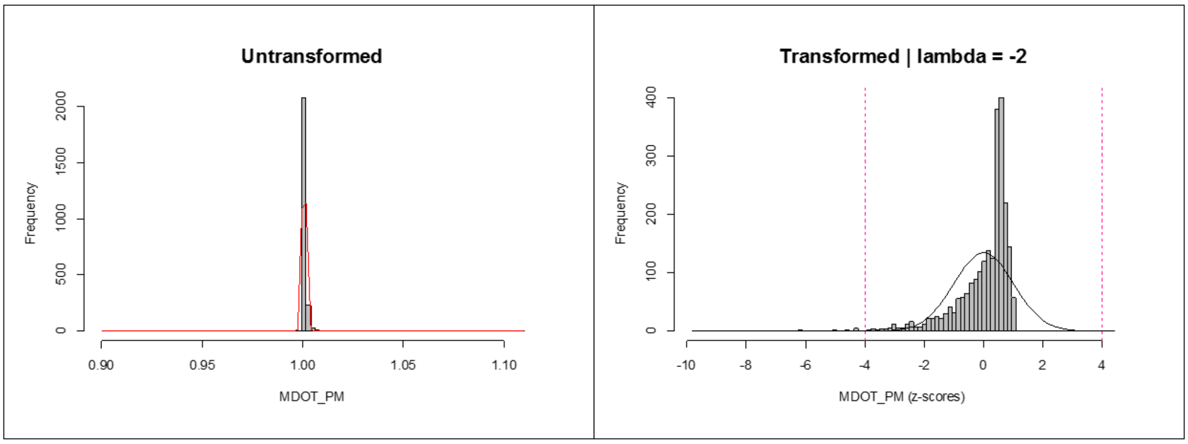

Figure 8. Example of outlier detection in pollutant traces using z-scores

The equivalent real-time plot for the bottom example is shown in Figure 9. The red bubbles indicate the time-stamps where the z-scores method has identified as suspicious if the z-score limit is set to 4.

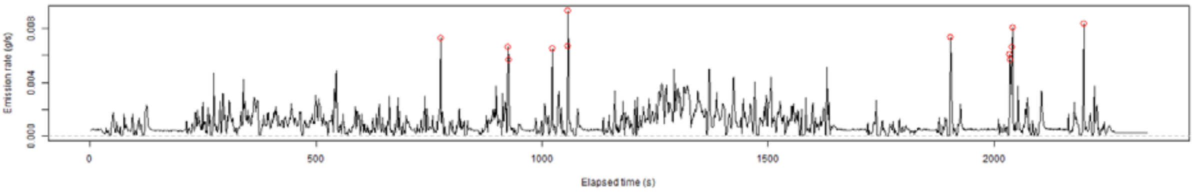

Figure 9. Real-time plot corresponding to the transformed data above (Figure 8)

After outliers have been identified, a decision must be made on the retention or deletion of each outlier. Outliers are not by definition invalid as they may be indicative of specific cases that, although relatively unique, are part of real-world vehicle emissions behaviour (e.g. emission spikes).

Given the large variability in vehicle emissions and especially if high emitters are present in a dataset, outliers should be initially retained, unless there is clear evidence that the data are invalid due to e.g. errors in recording, miscalculation, malfunctioning test equipment or other technical reasons.

2. Results and Discussion

The methodology analysed in the previous paragraphs was developed into verification routines using custom coding. The methodological approach was then tested on the EPD PEMS dataset in order to suggest calibration parameters and demonstrate examples of application for the screening. The pre-processing phase has been carried out prior to proceeding with the verification phases.

Overall test verification

For the first phase, the driving behaviour verification scheme application can outline tests with exceeding maximum acceleration values. Figure 10 shows an example of the summary statistics table extracted from a sample array of second-by-second data tests. The highlighted fields indicate maximum/minimum acceleration exceedance as defined in the developed maximum acceleration model (Operation feasibility verification procedure).

segm	idle_time	idle_prop	na_time	Pna_spd	length_cycle	AAD	MAD	min_speed	max_speed	min_acc	max_acc
PM_9	1	0.00	0	0	459	0.2	1.2	0.0	55.4	-1.2	1.0
PM_10	18	0.06	0	0	301	0.1	1.2	0.0	43.3	-1.2	0.7
PM_1	786	0.29	0	0	2,698	0.1	1.1	0.0	67.8	-1.6	1.0
PM_2	531	0.20	0	0	2,595	0.1	1.3	0.0	81.4	-1.9	17.3
PM_3	215	0.08	0	0	2,629	0.1	1.3	0.0	64.6	-2.0	1.2
PM_4	762	0.23	0	0	3,316	0.1	1.2	0.0	56.4	-1.7	1.0
PM_5	225	0.09	0	0	2,472	0.1	1.3	0.0	84.9	-2.3	1.2
PM_6	1,233	0.32	0	0	3,854	0.1	1.3	0.0	60.9	-1.7	1.1
PM_7	58	0.35	0	0	164	0.1	0.7	0.0	34.7	-1.2	0.8
PM_8	247	0.10	0	0	2,527	0.1	1.7	0.0	74.0	-8.8	1.3
PM_9	649	0.19	0	0	3,467	0.1	1.4	0.0	72.3	-4.1	1.1
PM_10	633	0.21	0	0	3,067	0.1	2.1	0.0	79.7	-2.1	1.2
PM_11	783	0.23	0	0	3,419	0.1	1.5	0.0	67.6	-2.2	1.2
PM_12	797	0.22	0	0	3,625	0.1	1.9	0.0	66.4	-1.8	1.1
PM_13	769	0.23	0	0	3,349	0.1	1.3	0.0	93.1	-2.1	1.0
PM_14	238	0.07	0	0	3,471	0.2	1.7	0.0	71.8	-1.9	1.1
PM_15	34	0.12	0	0	274	0.1	0.6	0.0	56.9	-10.6	0.8
PM_16	1,135	0.17	0	0	6,819	0.1	1.5	0.0	76.9	-2.7	1.2
PM 17	208	0.29	0	0	723	0.1	1.1	0.0	73.1	-1.8	1.1

Figure 10. Example of outlier detection in pollutant traces using z-scores

The obtained statistical properties of the speed-time trace are also presented. A similar table for individual trace (PM) statistics was presented in Figure 2. A visual analysis of the speed trace is necessary to validate possible issues.

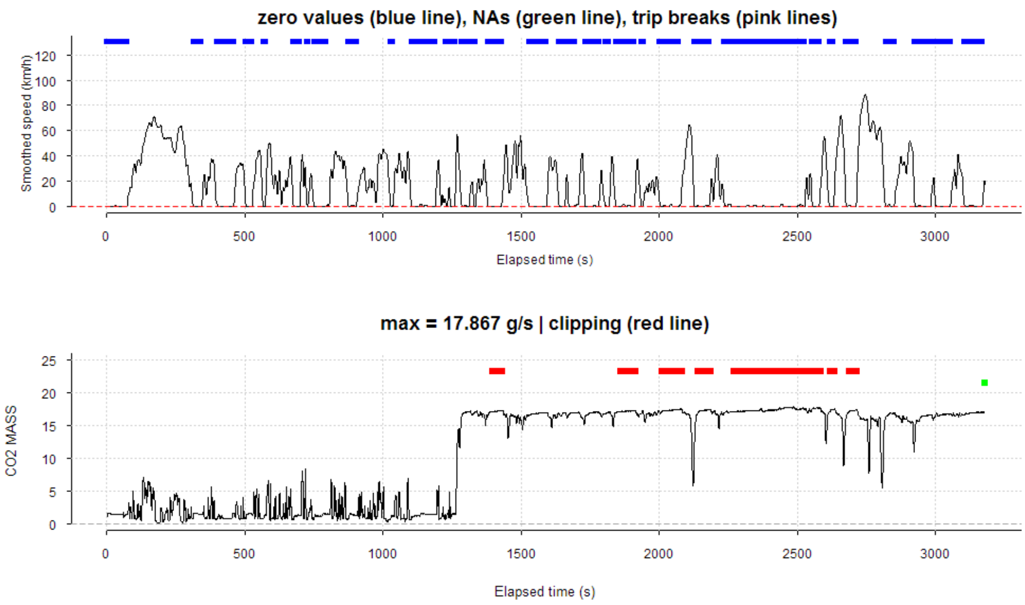

Figure 11. Example of gaseous file with a PCLIP value of 0.11 for CO2.

Individual traces verification

For the practical application of the clipping detection algorithm, the following approach can be used: potential clipping is flagged whenever C > 0.995 and the number of consecutive potentially clipped values is at least 15 s long. In this example, the idling segment is defined as a duration of at least 9 s while the first 5 s and the last 3 s of each idling segment are omitted.

An example of application is shown in Figure 11; it corresponds to a trip travelled by a petrol Euro 4 passenger car with questionably steady CO_2 values for most of the trip duration, despite the fluctuations in vehicle speed. Another sample test which demonstrates potential clipping is shown in Figure 12. The same, relatively low clip value is observed for both NO_x and NO measurements. This trip was performed by a heavy Euro II bus equipped with DPF.

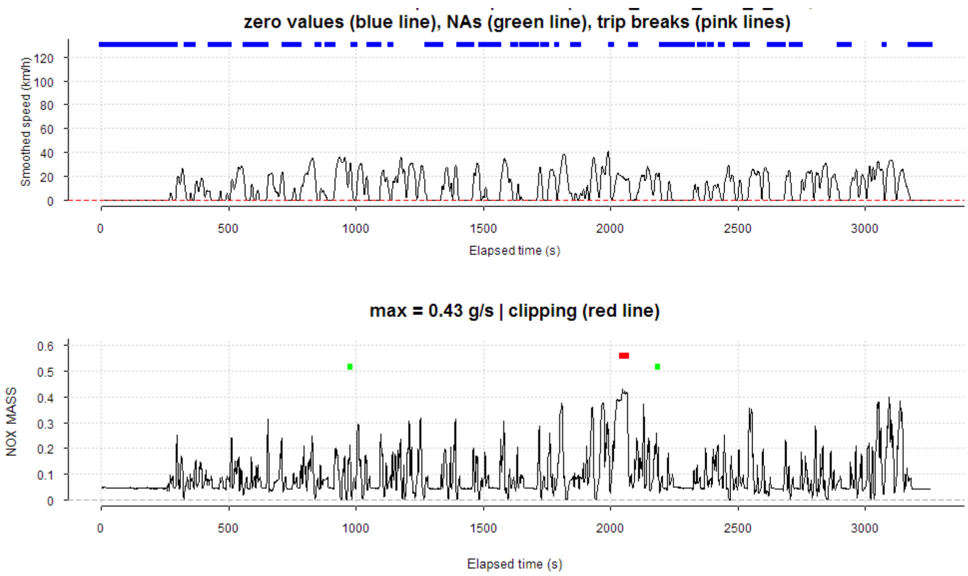

Figure 12. Example of gaseous file with a PCLIP value of 0.006 for NO_x.

In order to apply the COV/PTM analysis procedure, previous lab-based work was used to define typical COV/PTM criteria for modal emission traces (Smit, 2013), as shown in Table 1. Note that the criteria for PM in this table reflect HDV vehicle technology classes up to Euro IV.

Table 1. Emission trace verification criteria for PTM/COV

Variable	CO, HC, NO_x	CO_2, FC	PM
COV	> 1.0	> 1.2	> 1.0
PTM	> 2.5	> 2.5, < 12	> 2.5, < 50

To apply this verification procedure, the comparison may use these limits for modal emission traces from laboratory experiments as a starting point. Differences between PEMS and laboratory testing (sensitivity, stability, etc.) may affect the specific PTM/COV threshold values; thus, a further calibration is required to account for PEMS testing upon careful data inspection.

PEMS tests of the EPD dataset indicate that there is general compliance with the lab-level criteria of Table 1, although it appears that PEMS PM data for newer technology vehicles exceed the PTM threshold of 50, indicating more spiky PM data than were observed in laboratory test conditions for Euro I-III HDVs. This does not mean that the PEMS data are invalid, as lower baseline PM emission rates with relatively high spikes are expected for vehicles with advanced emission control (in particular DPF).

For this particular data set, the following PEMS outlier identification criteria for PM can be adopted instead:

- COV < 0.7
- PTM < 2.5 or PTM > 60 (no DPF technology)
- PTM < 2.5 or PTM > 150 (with DPF technology)

Figure 13 is an example of a vehicle test with very large PTM (304) and COV (10) values. This is a light bus with a DPF and has essentially zero PM emission levels (blue bar) throughout the trip, except for two consecutive peaks in the middle of the trip.

Figure 13. Example of PM traces with large PTM (304) and COV (10) values

For the gaseous emission, a minimum COV limit of 0.7 has been initially used for CO, NO_x, THC and CO_2. Upon reviewing the CO_2 data results visually, new boundaries that denote potential outliers are suggested to account for real-world measurements:

- COV<0.25, for passenger cars
- COV<0.65 for all other vehicles

Figure 14 shows an example of a vehicle test with very low PTM (1.1) and COV (0.11) values for CO_2 data. CO_2 are unrealistically steady and stay so even during idling periods.

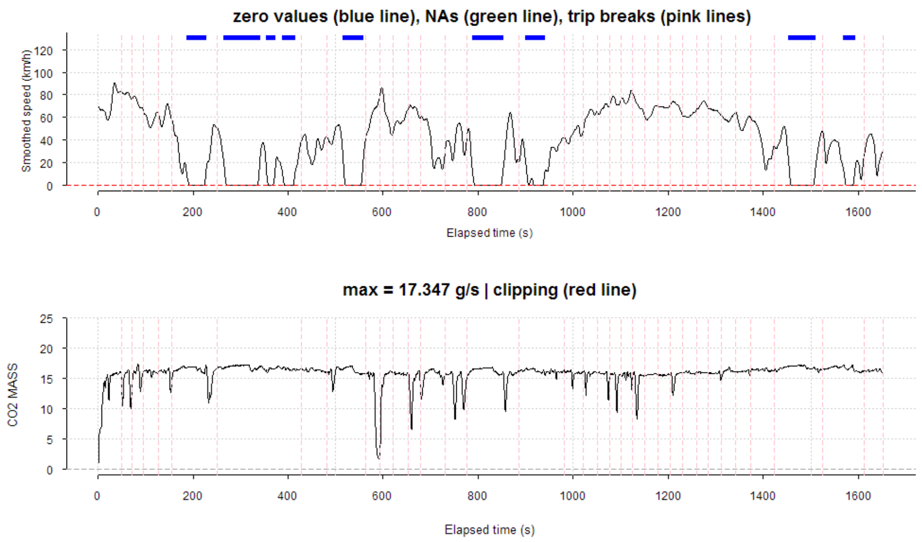

Figure 14. Example of CO_2 trace with very low PTM and COV values.

For the other gaseous pollutants, application of COV and PTM limits individually may flag several normal-looking trips as possible outliers; therefore a combination of COV/PTM boundaries was tested instead. The final thresholds were set

- COV<0.5 and PTM<3 for CO
- COV<0.5 and PTM<2.5 for NO_x
- COV<0.5 and PTM<2.5 for THC

In order to test the baseline analyser drift algorithm, the limit for computed drift variable D (as defined in the methodology) values was set to less than -5% or more than +5%. These values are flagged as showing significant drift. The example illustrated in Figure 15 yields a very high positive D value equal to 113% for CO_2 emissions. It corresponds to the same data test depicted in Figure 11 (clipping detection application). The top part of Figure 15 shows the speed-time visualisation of the observed test, while the bottom part illustrates the cumulative idling time of the same test. The orange-coloured shape is the regression polygon representing the drift. Zero values in the beginning of this test are around 2 g/s, but they exceed 15 g/s later on through the test. The combination of clipping and high drift will result in marking this trip as an outlier.

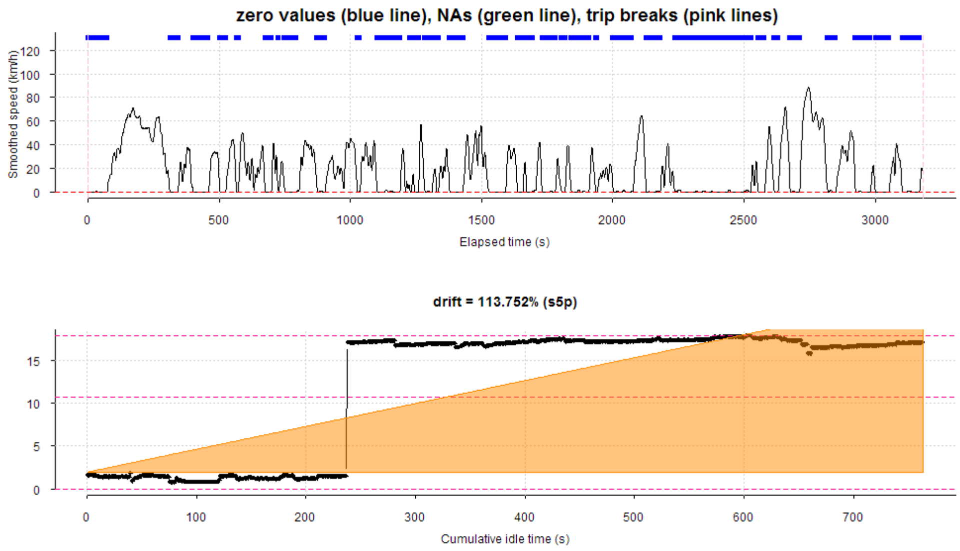

Figure 15. Example of CO_2 trace with large positive baseline drift value (+113%).

Figure 16 shows the THC drift plot of a petrol passenger car with a negative D value of -91%. The plots show a large negative drift of the THC emissions traces over time as shown in the bottom part starting from 0.012 g/s and reaching 0 g/s; this clearly illustrates a cold start and this trip is flagged as such. The corresponding speed-time visualization of this test is shown in Figure 17.

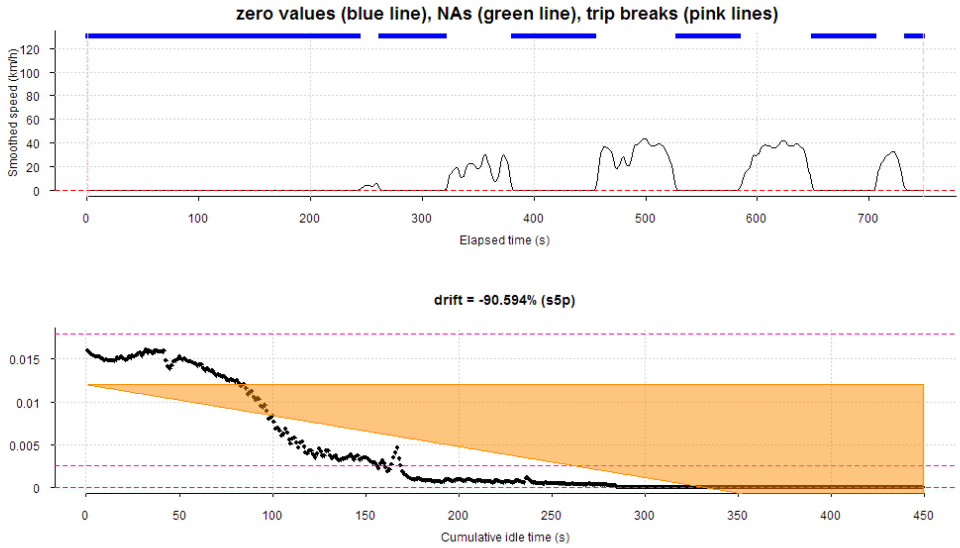

Figure 16. Example of CO trace with large negative baseline drift value (-21%). Drift time-series plot

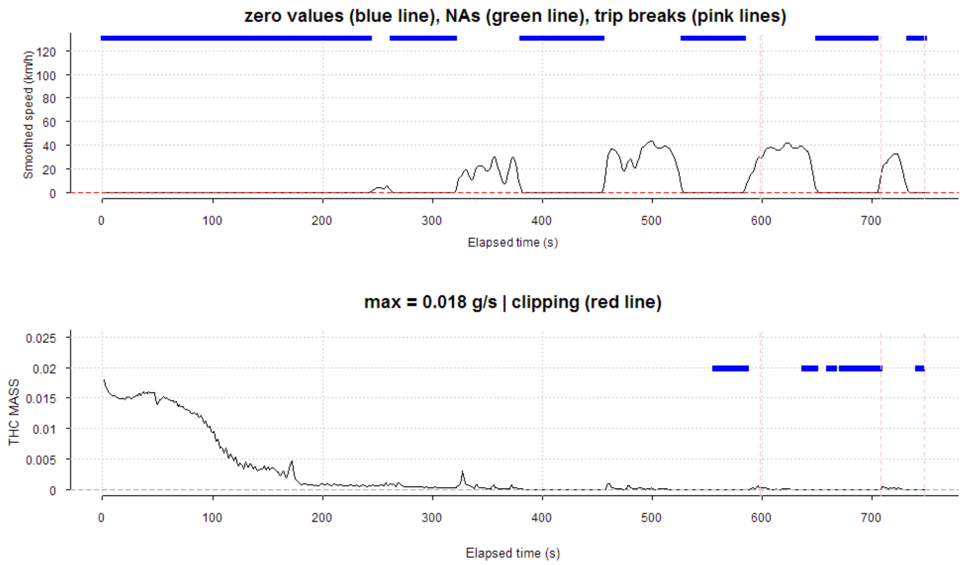

Figure 17. Example of CO trace with large negative baseline drift value (-21%). Speed time-series plot

For the application of the outlier detection scheme, the z-score limits must be defined. The sample test size in this dataset is typically large (n ~ 3000 s on average), so an absolute z-score threshold value larger than 3-4 could typically be used to flag potential outliers for normally distributed data. Should the transformed PEMS data exhibit notable, non-normal behaviour, a more conservative value of 5 can make the method more robust.

The outlier relevance is probably a more important metric to judge if the validity of a possible outlier. Initial visual examination of these pollutant plots may not yield unrealistic behaviour, so the outliers should probably be retained.

Examples of outlier detection results are shown in Figure 18 and in Figure 19; the former one shows THC emissions and is indicative of cold start, while the latter one shows a trip with a single spike in NO_2 values which yields a relevance of 0. It is however, recommended that further checks are performed for some of the extreme cases.

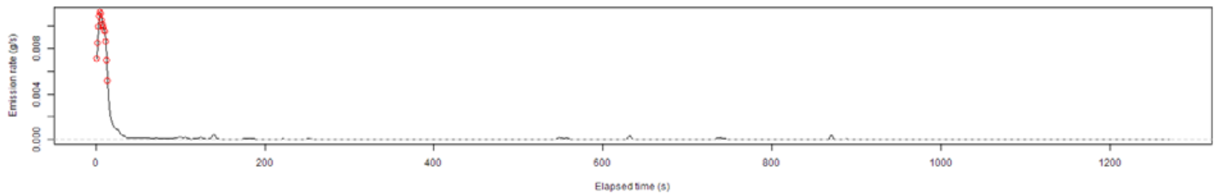

Figure 18. Examples of trip with identified outliers (THC, z-score = -11, relevance = 0.25)

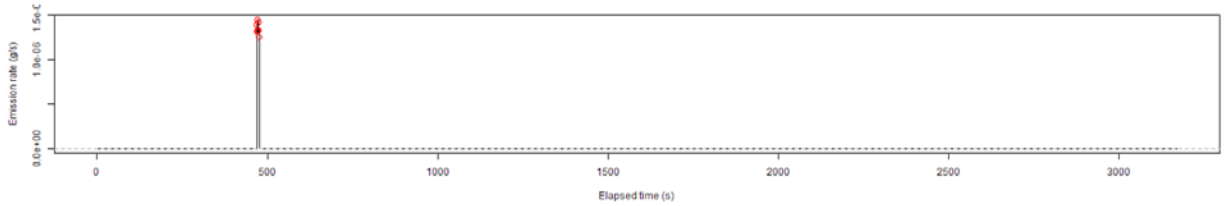

Figure 19. Examples of trip with identified outliers (NO_2, z-score = -23, relevance = 0)

The application of the visual time-series analysis is useful tool to validate issues raised by the other verification procedures and especially for outlier detection.

3. Further work

So far, a comprehensive approach for the verification of a variety of PEMS data measurements has been presented. Nevertheless, there is room for further enhancements. An option for consideration is to use verification algorithms that apply multivariate analysis across a number of pollutants. Univariate analysis does not reveal observations that fall within the ordinary range of values on each of the pollutants, but are unique in their combination of values for the different pollutants.

This option would require an analysis of multivariate distributions and could be achieved e.g. by using a statistical variable called the Mahalanobis Distance, which measures the distance in multi-dimensional space of each observation from the mean centre of the observations.

Also, it is recommended that emission factors based on PEMS data are compared at both vehicle and vehicle class level to ensure any suspicious data are further investigated. As a first step, the data need to be categorized to ensure comparison of emission tests for similar vehicles in similar test conditions. An example is illustrated in Figure 20, where several buses with similar characteristics (vehicle class, emission control technology, fuel, euro standard) are depicted together in terms of PTM vs. COV.

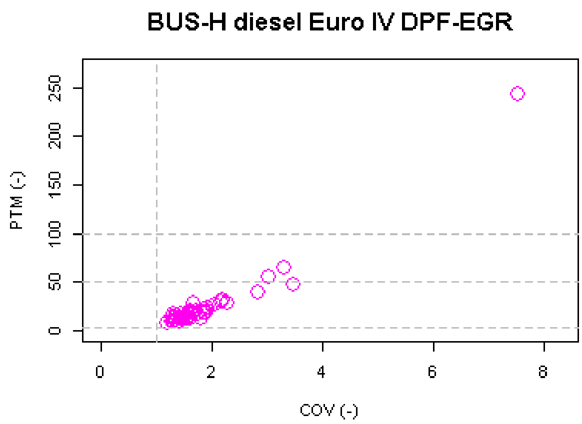

Figure 20. COV-PTM plots for PM emissions for individual vehicle classes.

Therefore, subsets of test data (denoted as groups) can be created, where each subset represented a particular combination of vehicle class, driving conditions and driving mode (hot, cold). Secondly, the data need to be shifted and transformed and z-scores computed for each data point. Thirdly, univariate or multivariate analysis can be used to detect outliers.

To the extent possible, additional protocols should also be devised to deal with missing values, including consideration of possible imputation methods, and outliers (valid, invalid).

Summary and Conclusion

This study aimed to devise a set of automated procedures to extract verification statistics and visualization schemes in order to quality assure PEMS data. The overall goal is to ensure invalid data are flagged, corrected or removed, while at the same time valid outliers are maintained. During the first step of the procedure, a pre-processing of the PEMS data set is carried out to extract useful information for the verification process. Then, in the first verification step, an assessment of overall PEMS test validity is performed. Basic trip metrics are investigated to establish that general abnormalities are detected and flagged. The verification algorithm included a procedure to detect unrealistic accelerations as a function of instantaneous vehicle speed, generic vehicle type and power-to-weight ratio. A non-linear model was fitted to these data by adopting a reverse sigmoid function. The next verification phase focused on individual pollutant traces to identify occurrence of such errors in the dataset. For the gaseous and PM data, apart from visual analysis, the verification algorithms examined clipping, coefficient of variation / peak-to-mean ratio (COV/PTM) analysis, analyser drift and outlier detection. Initial comparisons used verification criteria for modal emission traces from laboratory experiments, but these must modified to account for PEMS testing upon careful data inspection. The baseline analyser drift algorithm was developed to quantify analyser drift and indicate the possibility for other emission events such as cold/hot operation conditions. The outlier detection method used in this analysis focused on univariate data trace level, as any outliers detected and verified at this fundamental level, will prevent propagation of data issues to higher scales of assessment. Standard z-scores were used to implement the outlier detection protocol and the review was based on outlier relevance to measure their impact.

These QA/QC procedures were applied to the Hong Kong dataset collected by EPD to illustrate how they can be fitted to take account of any particularities of the PEMS measurements in general as well as environment-specific ones.

The overall approach provides a variety of tools to tackle different types of data measurements and examples of application indicate that it can successfully detect a variety of issues and suspicious behaviours. This verification methodology can be further expanded to include multivariate analysis across a number of pollutants. Also, emission factors based on PEMS data can be compared at both vehicle and vehicle class level to ensure any suspicious data are investigated further.

Acknowledgments

This work was carried out in the framework of the project "Provision of Services for Quality Assurance of the Portable Emission Measuring Systems (PEMS)", funded by the HKEPD.

References

Smit, R. (2006), An Examination of Congestion in Road Traffic Emission Models and Their Application to Urban Road Networks, Ph.D. Dissertation, Griffith University, Brisbane, Australia.

Smit, R. (2013), A procedure to verify large modal vehicle emissions databases, Proceedings of the CASANZ Conference, Sydney, 7-11 September 2013.

Velleman, P. F. (1976), Robust Nonlinear Data Smoothers; Theory, Definitions, and Applications, PhD thesis, Princeton University.

Velleman, P. F. (1980), Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms, Journal of the American Statistical Association, vol. 75, number 371, pages 609-615.