Anna van Buerenplein 1
2595 DA Den Haag
P.O. Box 96800
2509 JE The Hague
The Netherlands

www.tno.nl

T +31 88 866 00 00

TNO report

TNO 2017 R11015

Review of RDE legislation: legislation text, evaluation methods and boundary conditions on the basis of RDE test data

Date	25 August 2017
Author(s)	Norbert E. Ligterink, Bernd J. Rietberg, Willem Hekman, Pim van Mensch, Rob F.A. Cuelenaere, Sam van Goethem
Copy no	2017-STL-RAP-0100307878
Number of pages	111 (incl. appendices)
Number of appendices	2
Sponsor	DG-GROW
Project name	RDE evaluation
Project number	060.27375

Summary

The European Commission has the mandate to keep under review the procedures, tests and requirements, as well as the test cycles used to measure vehicle emissions as part of the Euro-5/6 legislation. The development of the new Real Driving Emissions (RDE) legislation is considered to be the way forward to secure low on-road emissions in the years to come.

The Commission has contracted TNO to review the evaluation methods from current legislation and a few alternative options, on the basis of the test data and experience from the stakeholders. In addition, boundary conditions were assessed as well. For this review RDE data from different stakeholders were used, as collected by JRC for this purpose.

Also the opportunity was given to stakeholders to provide information on the appropriateness of the RDE boundary conditions. Only information on the normal driving behaviour of N1 vehicles, i.e., vans, was supplied for this purpose. JRC has added this new data to the overview on N1 driving behaviour.

The main objective of this report was to provide a comprehensive overview of the analysis of the current RDE data-evaluation methods, alternative proposals for evaluation, and RDE boundary conditions, on the basis of the experiences of the stakeholders with RDE testing, reflected in the RDE test data. The data collected were received from a wide variety of stakeholders, including OEM's, technical services, type-approval authorities, national authorities, independent institutes, consumer organizations, and NGO's. The results reflect the experiences with RDE testing in the last years, based mainly on the RDE1 and RDE2 legislative packages.

The evaluation methods under review are: (based on RDE1, RDE2, and RDE3)
- Raw emissions, no boundary conditions
- Raw emissions with test validity on the basis of RDE boundary conditions (annex IIIa and appendix 7a and 7b, reference method)
- Moving Average Windows (MAW, appendix 5 RDE)
- Power Binning (PB, appendix 6 RDE including the adaptions of RDE3)
- Looped MAW, variant of MAW with windows looping from the end to the beginning.
- NOx/CO2 (appendix 7c, intended for hybrid vehicles)
- NOx/CO2 with ICE distance fraction (ACEA proposal with test validity on MAW boundary conditions)
- Raw emissions with validity with MAW boundary conditions (T&E proposal)

The review of the evaluation methods contained multiple aspects. It was examined if elements of the RDE legislation are fit-for-purpose. Secondly, the practicality of the different elements was examined. Thirdly, the technological neutrality and applicability for all emission reduction technologies has been considered. Fourthly, it was examined whether elements are prone to abuse or exploitation. Finally, the transparency, ease of use, and unambiguous understanding of the methods were examined. These aspects were interpreted in the context of the study, i.e., what does the evaluation method do with the data, as robustness, sensitivity, effectiveness and bias.

The three overarching conclusions of the review are:

- The fraction of invalid trips is large. The evaluation methods exclude many more tests on top of the exclusions based on trip composition and trip dynamics boundary conditions.
- The results of the evaluation methods do not seem to be consistent. Large corrections of the raw, or measured, results occur in both directions, up and down.
- However, these corrections are not fully random, but seem to be systematic to some extent, varying with different vehicles. Systematic corrections are prone to exploitation. The extent of systematic corrections could not be established on the basis of the data provided.

The effectiveness of the methods in correlating the evaluation results in a systematic manner to the raw results is regarded as an important evaluation aspect. All of the methods cause scatter in the results, unexplained by the underlying aspects.

Below, the most important conclusions from this study are summarized. The report supplies more details on each of the elements below.

Regarding the research method and data received:

- The data were collected from different sources by JRC and delivered to TNO. The data sets received varied in (lack of) comprehensiveness and quality, but many of these aspects could not be assessed within the scope of the study. The study focussed on analysing the RDE test data. For example, vehicle precondition, vehicle technology, and cold start, may lead to unknown variations in the results. The variation in emissions have no clear relation with known variation in test execution (i.e. driving behaviour and trip composition) for the same vehicles (except for a few vehicles). The has affecting the evaluation. In particular, for a substantial fraction of the data the fuel type, petrol or diesel, was missing. Some data did cluster at low NO_x emissions below 200 mg/km, which is likely to be mainly petrol. However, technological neutrality also means the fuel type should not make a difference in the evaluation, so the lack of this data did not affect the outcome.
- About 350 test files were received, of which 252 RDE-like trips were identified as appropriate input for the evaluation methods.
- The PB (power binning evaluation) and the MAW (moving average window evaluation) require specific WLTP data to run properly, which was often not available. This data were simulated based on default values. It was shown that PB and MAW are sensitive to differences in these WLTP input values. Input values were available for 50% and 70% of the analysed data for PB and MAW, respectively.
- Certain test conditions and vehicle state and condition were not known, such as ambient temperature, altitude, weather conditions, road conditions, soak time, OBD read-out, etc.. These aspects could not be assessed.

Regarding boundary conditions:

- All boundary conditions together are an important reason that many of the RDE tests were deemed invalid. The general boundary conditions included in the RDE legislation, for instance on trip composition, driving behaviour, reduced the number of valid trips by one third (from 252 to 168 trips).

- Of the trips assessed compliant with the general boundary conditions, close to 60% is considered invalid according to the specific boundary conditions included in the MAW evaluation method and even close to 80% according to PB. Some improvement occurs with variants of the methods and changes in the input data.
- The boundary conditions, part of the evaluation methods, PB and MAW, have a greater impact on the test validity than trip composition and trip dynamics boundary conditions. This is deemed an undesired effect of these evaluation methods.
- Consequently, the outcome of the different evaluation methods are possibly biased by the limited number of valid tests for each method.
- To uncover stringent boundary conditions within these methods, an in-depth analysis of MAW and PB boundary conditions was done. For MAW, motorway share and the urban part of the CO_2 band are important factors for invalidity on MAW test normality. For PB, the power bins P1+2, P3 and P5 are important factors. A more ‘relaxed’ version of PB, with elements from the software version and an alternative interpretation of the legal text, was also implemented for our analysis, showing more than twice as much valid trips (see Appendix B).
- There are conflicts between RDE boundary conditions and evaluation method boundary conditions. Different boundary conditions are complementary to each other. They do not seem to invalidate the tests according based on the same principles of test normality.
- In a substantial number of cases, the outcome of the invalidity check is inconsistent between methods (e.g. some trips were *valid* for MAW, *invalid* for PB). This seems to suggest the methods assess test normality in different ways.
- PB and MAW boundary conditions are sensitive to WLTP input. The span of RDE test conditions is large, while the WLTP is a precise description to achieve reproducibility. Consequently, sensitivity for WLTP input may restrict RDE tests to conditions related to the WLTP in a diffuse manner.
- Some RDE boundary conditions, especially those of legal package RDE3, are relatively often exceeded, such as urban idling and the 60km/h speed limit during urban driving. Trips were not excluded on these grounds, also not on altitude signal quality.

Regarding legislative text and general purpose of boundary conditions:

- Boundary conditions and evaluation methods are two sides of the same coin. If the boundary conditions are restrictive, this leads to average testing, which may contain many aspects of normal driving, but in a distinct combination. In that case, the evaluation method may be unnecessary, as the test variation is limited. On the other hand, if boundary conditions allow too much freedom in testing, evaluation methods will be an essential part to compensate, or correct, for overrepresentation of certain parts of driving in the test.
- The boundary conditions related to acceleration and altitude require high quality of the velocity and altitude signals. This quality is defined in limited terms in the legislative text, leading to difficulty in assessing boundary conditions based on these signals.

Regarding emission results:

- The evaluation methods lead to a large variation in corrections, both up and down. However, no significant systematic upwards or downwards shift, of the

complete data set, was found in any of the methods. MAW and “MAW looped” show highly similar emission results.
- MAW, PB and NO_x/CO_2 give some outliers for the correction, indicating the possibility of misuse.
- If data sets of multiple tests with the same vehicle were available, further analysis indicate the existence of systematic corrections that depend on the combination of vehicle and evaluation method. The direction (up or down) of the systematic corrections might vary with evaluation methods.
- The trip validity seems to be the most restrictive part of the different evaluation methods. The magnitude of the corrections in the different methods should not be seen independently from the number of valid trips it applies to. If more than half of the trips are invalid for trip normality of the evaluation method, some bias may be expected in the emission results based on the valid trips only.
- The effectiveness of an evaluation generate systematic yet unbiased effect is regarded as an important quality criterion. It was shown that none of the methods show a high correlation between raw results and evaluation results.. MAW even showed a slight increase in scatter on top of the variation observed in the raw test results with individual vehicles. Some systematic effects per vehicle might exist in PB, some vehicles seem to have systematic upward corrections in the tests while other vehicles systematic downward.
- In contrast to the trip validity check in MAW and PB, that turned out to be very sensitive for WLTP input values. The emission correction results of MAW and PB were not highly sensitive to input WLTP and vehicle values.

Transparency, simplicity and unambiguous understanding:
- PB and MAW contain unclear statements in their respective law text. Choices had to be made on the interpretation of the text to come up with working software implementations.
- PB and MAW were found to show disadvantages on unambiguous understanding, transparency and simplicity. It is difficult to trace back certain effects to the root cause. NO_x/CO_2 and especially raw showed clear advantages on these three points.
- MAW boundary conditions are also used for the validity in the T&E proposal and the ACEA NOx/CO2 proposal, thus show the same complexities, and draawbacks, as MAW evaluation itself. The by ACEA proposed adaption of the boundary conditions does not change this conclusion.
- The invalidity of tests have a convoluted relation with the test execution. It is therefore difficult to supply driver instructions to limit the number of invalid tests.

Results
- The table below summarizes the results, and shows the large range of corrections and the limited bias. It is expected that the average corrections are similar, and limited. If one method gives a different average for the correction, it can be interpreted as bias.

Table 1 The average correction varies with the number of tests included. The results for all tests, the valid test according to the method, and the RDE valid test, which include both the method boundary conditions and the trip and driving dynamics conditions.

Method	range valid tests [%]		range all tests [%]		average correction[%]			number of RDE valid tests
	max	min	max	min	all tests	method valid	RDE valid	
Option 1: Only Trip Composition	0%	0%	0%	0%	0%	0%	0%	217
REFERENCE (RAW with RDE boundary conditions)	0%	0%	0%	0%	0%	0%	0%	168
RAW with MAW boundary conditions (T&E)	0%	0%	0%	0%	0%	0%	0%	75
MAW (EMROAD)	47%	-51%	87%	-51%	0%	2%	1%	75
LOOPED MAW (windows going round to the beginning)	47%	-51%	112%	-51%	-1%	1%	3%	61
PB (CLEAR)	25%	-34%	139%	-51%	-3%	-4%	10%	34
NOx/CO2	58%	-49%	59%	-52%	-7%	-6%	-6%	168
NOx/CO2 * ICE with MAW boundary (ACEA)	23%	-31%	59%	-52%	-14%	-10%	-9%	84

The plug-in hybrid vehicle is excluded from the analysis of the range.

Discussion:

- Certain effects of the evaluation methods rely on the underlying assumption that somehow the emissions are directly linked to driving behaviour and test execution, in the way that the evaluation methods are linked to driving behaviour. However, this may not be the case. The analyses are restricted by the fact that variations in emissions have different causes as well, producing scatter in the results. This scatter obscures the effects of the evaluation methods to some extent.
- Putting too much focus on WLTP input values may be unwarranted, because these values are related to strict protocol testing and conditions which may deviate substantially from the conditions in the RDE test.

PB and MAW show inconsistencies and high numbers of invalid trips on boundary conditions, disadvantages on emission outliers, effectiveness, robustness of the test normality, transparency, simplicity and unambiguous understanding. However, the NO_x/CO_2 and raw emissions methods show some disadvantages as well. NO_x/CO_2 shows a similarly low effectiveness as MAW and PB, and it shows highly similar results to raw emissions. Moreover, evaluation methods bring limited benefit over raw results to justify the risks of deviating results.

The test normality checks of PB or MAW lead to a large number of invalid tests. The large number of test disqualifications are the results of a number of complementary conditions, often on different parts of the test. For example, a valid total trip may be invalidated on one aspect of the urban part. The so-called stacking of boundary conditions seems to leave a narrow margin to achieve a valid tests. Since these conditions are only determined afterwards on the total test, there is little guidance of appropriate test execution. For example, since methods are velocity based, low velocity driving at the end of the motorway part may invalidate a test on the basis of trip composition.

The raw results and NO_x/CO_2 with ICE distance fraction, without additional MAW boundary conditions, do not pose the problems with invalid tests, unlike the other approaches. These two approaches have their own advantages and disadvantages, but not according to the aspects studied here.

Contents

1 Introduction

Background
The development of the new Real Driving Emissions (RDE) legislation is considered to be the best way forward to secure low on-road emissions from light-duty vehicles in the EU in the years to come. From 1 September 2017, RDE legislation will be mandatory for type approval of new emission types of vehicles and will be implemented fully in various steps till 2022.

Among other things, an RDE test needs to comply with the boundary conditions as defined in the legislation. Boundary conditions are implemented to prevent testing under extreme conditions that would not represent the majority of the normal use of the vehicle. The defined boundary conditions are related to ambient and road conditions, trip composition, driving behaviour and vehicle conditions.

RDE tests are processed by using one of two possible evaluation methods in accordance with the RDE legislation. These evaluation methods are meant to compensate the effects of deviating test executions in the results in the emission results. The Commission has contracted TNO to review the evaluation methods from the current legislation and a few alternative options. In addition, the percentage of driving that is not covered by the boundary conditions is to be assessed as well, if new data were available from the stakeholders. For the review of evaluation methods RDE data collected from different stakeholders was to be used. No data were supplied which allowed for the evaluation of boundary conditions against normal vehicle use.

Aim and Approach
The European Commission will have to decide on the final RDE text. The main objective of this report is to provide a comprehensive overview of the analysis of the current RDE data evaluation methods and coverage of RDE boundary conditions. The analyses are subject to the availability of data or other evidence provided by the stakeholders at the request of the European Commission.

The review of the evaluation methods covers multiple aspects. First, it should examine whether they are fit for purpose. Second, it should examine the practicality of the different elements. Third, the technological neutrality and applicability for all technologies should be warranted. Fourth, it should check whether elements are prone to abuse or exploitation, for example, defeat devices and test executions generating systematic deviating results, compared to the raw, measured emission data. Finally, the transparency, ease of use, and unambiguous understanding of the methods should be examined.

In the report, the terms ‘robustness’ and ‘effectiveness’ are used to qualify the evaluation methods, which will be explained further in the relevant chapters. The terms are used in the limited meaning concerning the analyses of data, rather than in the design of experiments.

The definitions used in this report find their basis in system theory. Robustness refers to the clustering of the results, with limited variation and limited outliers. Sensitivity is the magnitude of the variation of the result with the small variation of input data. The term effectiveness reflects the purpose, or desired effect, of the evaluation methods to limit scatter, and show correlation for tests with the same vehicles. For one vehicle, the evaluation input is same, and the tests are compared against the same WLTP reference.

The options for evaluation methods under investigation are listed in the table below. Not all methods affect the results, but they may only lead to additional invalidation of the RDE tests. Hence, in terms of the results, the picture is less complex than with the different boundary conditions used. There are four distinct evaluation methods that affect emission results: raw, Moving Average Windows (MAW), Power Binning (PB), and NOx/CO2, with two minor variations: looped MAW and ACEA NOx/CO2 * Internal Combustion Engine (ICE) fraction.

Table 2 Options for evaluation methods under investigation

	evaluation method	description	distinct values in the results
Option 1	no trip dynamics	Raw emissions	REFERENCE
Option 2 = REFERENCE		Full boundaries + no evaluation method	REFERENCE
Option NL	hybrids	Raw emissions/CO2	NOx/CO2
Baseline EMROAD	MAW	RDE 3 Baseline (EMROAD)	MAW
Baseline CLEAR	PB	RDE 3 Baseline (CLEAR)	PB
EMROAD modified	looped MAW	Circular windows calculation	LOOPED
In-between (T&E)	MAW	raw emissions	REFERENCE
ACEA new proposal	wider MAW	MAW boundary with NOx/CO2 * ICE fraction	as NOx/CO2 except for PHEV vehicle

The evaluations fall apart in three categories:

- Raw emission results with a variation of trip and driving boundary conditions. The reference method is raw emission results, without evaluation, but with all other RDE boundary conditions applied.
- Evaluation methods, as described in the RDE text, like MAW, PB, and NOx/CO2.
- Augmented methods, which include additional boundary conditions, like MAW trip normality, as in the T&E proposal, or adaptions thereof as in ACEA and looped MAW

In three cases the same raw results are combined with increasing number of boundary conditions: all raw (Option 1), RDE valid raw (Option 2, reference), and MAW valid raw results (T&E). The results remain the same except for the test validity, and these values are referred to as “raw”. Therefore, in the case of reference, the difference lies only in the validity of the RDE tests, not the results.

An important boundary condition, separate from the effect of the evaluation method, is the Moving Average Window (MAW) boundary condition, also referred to as trip normality in MAW. In three cases the MAW boundary condition is applied: in MAW itself, in raw emissions with valid RDE trips according to MAW, and in an adapted version for the ACEA proposal of NO_x/CO_2 weighted with the fraction of the distance the ICE was operating.

Apart from the evaluation methods, also the boundary conditions are to be evaluated. In the past doubts was raised whether boundary conditions does not exclude normal driving. However, little information is provided by the stakeholders to examine boundary conditions for their appropriateness. Boundary conditions form also an integral part with the evaluation methods. Boundary conditions are both part

of the different evaluation methods, and additional requirements which are in part complementary to those in the boundary conditions.
The measurement data which are collected by different stakeholders is the basis of this study. The different evaluation methods are applied to this data. As a result, interpreting the results is somewhat convoluted as it is based on interpreting the action of the evaluation method on this data, and the results they produce. This may not necessarily lead to a uniform or consistent picture. Possible indirect effects, depending on unknown parameters, may cause certain effects for specific tests. But it is assumed that the large number of parties provided data with a broad coverage and limited bias. Potential bias and errors of the analysis is discussed in-depth.

Structure of the report
This report starts with a short overview of the process of RDE testing, the different types of evaluation methods, and the boundary conditions, in Chapter 2. Next, an overview of the received RDE data forms Chapter 3. In Chapter 4, some remarks and discussion on a possible bias or error in our analysis show the limitation of this study. Chapter 5 is a discussion on the quality of given signals and the dependence of the RDE legislation on this quality. In general results depend only weakly on the quality of the velocity signal. Some exceptions are noted.
From Chapter 6 and on the results of this study is presented. Chapter 6 gives a discussion on the general purpose of boundary conditions, analyse of the data in the light of these boundary conditions, and it remarks on current RDE boundary conditions.
Chapter 7 deals with MAW evaluation method. The boundary conditions of MAW are discussed, as its practicality, transparency, and sensitivity to exploitation. Also, the evaluated emissions are compared with the raw emission result, to check for systematic effects. Lastly, the Looped variant of MAW is discussed.
Chapter 8 is similar to Chapter 7, but for PB. Chapter 9 focuses on the NO_x/CO_2 based evaluation methods: discussing the ACEA proposal and analysing the effect on PHEVs. The NOx/CO2 and the ACEA proposal have special treatment of PHEVs vehicles, for PB and MAW the tests with PHEVs yield mainly invalid results.
Chapter 10 discusses the effectiveness of the evaluation results. There, the emission results will be analysed from a ‘per vehicle’ perspective, giving insight if evaluation methods are effective in systematic corrections with limited scatter.
Chapter 11 compares different methods with each other, to investigate consistency between methods. In Chapter 12, the WLTP values used as input for MAW and PB are varied to study the effect on the results of the evaluation methods.

Chapter 13 contains a discussion the transparency, simplicity and unambiguous understanding of the evaluation methods. Chapter 14 contains discussion overall topics of the evaluation methods. Chapter 15 are concluding remarks, which summarises the main points of this study. Appendix 1 discusses the core legislative text itself, errors and recommendations for changes, and Appendix 2 offers a large number of extra graphs for the reader to study further.

2 Description of boundary conditions and evaluation methods

In 2016, TNO performed, for the Dutch government, a rather extensive assessment of the strengths and weaknesses of the RDE test procedure[1]. Based on the aforementioned assessment, this chapter describes the most important RDE test and vehicles conditions in brief. As these constitute limitations for RDE testing, the conditions are summarized as RDE boundaries. Moreover, a short description of the evaluation methods is given.

Not all boundary conditions could be applied due to a lack of data. In each subchapter, the used boundary conditions are mentioned, and an overview of the applied boundary conditions, as these were given in RDE1 to RDE3, are given at the end of this chapter.

2.1 Description of the RDE boundary conditions

In this chapter an overview is provided of several boundary conditions. An RDE trip executed within the normal and extended boundaries qualifies as being valid.

<u>Ambient temperature and road conditions</u>
Table 3 provides an overview of RDE boundaries, in terms of ambient and road conditions. This table makes a distinction between ‘normal’ and ‘extended’ boundary conditions. Under these extended conditions, it is more difficult to comply with the emission limits. Hence, the RDE legislation allows for a reduction factor of 1.6 for the emissions measured during driving events under extended conditions.

The ambient temperatures for the obtained trips were not available, therefore these boundary conditions are not applied. Absolute altitude was not available or measured with limited accuracy, thus was not applied. This is elaborated in more detail in chapter 5. All other road conditions, such as road surface, head wind, etc., were also not available.

Table 3 Boundaries for ambient temperature and road conditions

Condition	**Boundary(ies)**	
	Normal	**Extended**
Ambient temperature	0 - 30°C Temporary[2]: 3 - 30°C	-7 - 0°C and 30 - 35°C Temporary: -2 - 3°C and 30 - 35°C
Altitude	Maximum 700 m	700 - 1300 m
Road surface	Paved road only	-

[1] Strengths and weaknesses of the new European RDE test procedure, see: http://publications.tno.nl/publication/34622349/F3ewoI/TNO-2016-R11227.pdf

[2] Temporary boundary conditions apply till 1 September 2019 for new type approvals and 1 September 2020 for all registrations.

Condition	Boundary(ies)	
	Normal	Extended
Road incline	Only indirectly restricted by maximum average cumulative altitude gain over total RDE trip	-
(Head) wind, air pressure and air humidity	No restrictions	-

Trip composition

The RDE legislation contains several requirements for the composition of a valid RDE trip setting boundaries on the duration of the trip, the sequence of urban, rural and motorway driving, the minimum trip length and the number and duration of vehicle stops. Table 4 provides an overview of these boundaries for the trip composition.

The altitude gain could not be thoroughly assessed, and therefore is not applied as a boundary condition in our implementation.

Table 4 Boundary conditions for the trip composition

Condition	Boundary(ies)	Margins
Duration	90-120 minutes	-
Shares Urban (U), Rural (R) and Motorway (M) driving[3]	34%, 33%, 33% of trip distance	29% ≤ U ≤ 44% 23% ≤ R ≤ 43% 23% ≤ M ≤ 43%
Sequence is fixed: Urban driving followed by Rural and Motorway driving	-	-
Length of each section (U/R/M)	At least 16 km	-
Characterisation of urban, rural and motorway driving	Urban: up to 60 km/h Rural: between 60 and 90 km/h Motorway: range between 90 and at least 110 km/h	
Cold or hot start	Maximum of 15 seconds idling after initial engine start and a limitation of 90 seconds for the vehicle stop in the entire cold start period	-
Stops	Several stops ≥10s may be included. Total stoppage time shall be 6-30% of time of urban driving. If a stop lasts over 300s, the test is void	-
Total cumulative positive altitude gain	<1200 m per 100km RDE trip distance, calculated over the full RDE trip. Road incline as such is not regulated	-
Altitude start and end point	Shall not differ by more than 100 m.	-

Driving behaviour

[3] Urban driving is defined as all events with vehicle speed up to 60 km/h included, rural driving by speeds between 60 and 90 km/h and motorway driving by speeds above 90 km/h.

The RDE legislation contains several requirements for the driving behaviour to prevent a valid RDE trip from being driving consistently aggressive or smooth. Table 5 provides an overview of the boundaries of 'normal' driving, in terms of driving behaviour.

The $v*a_{pos}$(95%), the highest 95% percentile of the product of vehicle speed and (positive) acceleration, is commonly used as an indicator for high(er) dynamics of a trip and RPA, the relative positive acceleration, as an indicator for the lack of dynamics in a trip. These indicators are calculated for the urban, rural, and motorway velocity bins of each trip.

Table 5 Boundary conditions for driving behaviour

Parameter	**Boundary(ies)**	**Comment**
$V*a_{pos}$	RDE trip is invalid if (per speed bin) $\bar{v}_k \leq 74.6\mathrm{km/h}$ and $(v \cdot a_{pos})_{k\text{-}}[95] > (0.136 \cdot \bar{v}_k + 14.44)$ or $\bar{v}_k > 74.6 km/h$ and $(v \cdot a_{pos})_{k-}[95] > (0.0742 \cdot \bar{v}_k + 18.966)$	To exclude extremely high dynamics
RPA	RDE trip is invalid if (per speed bin) $\bar{v}_k \leq 94.05\mathrm{km/h}$ and $\mathrm{RPA}_k < (-0.0016 \cdot \bar{v}_k + 0.1755)$ or $\bar{v}_k > 94.05 km/h$ and $\mathrm{RPA}_k < 0.025$	To include sufficient dynamics
Average speed during urban driving	15 km/h ≤ v_{avg_urban} ≤ 40 km/h	
Maximum speed	V_{max} ≤ 145 km/h	For no more than 3% of the duration of motorway driving speeds up to 160 km/h are allowed.
Speed range of motorway driving	Shall properly cover a range between 90 and at least 110 km/h. Speed shall be above 100 km/h for at least 5 minutes.	Vehicles with speed limitations have modified boundaries
Gear selection	No restrictions	

<u>Vehicle conditions</u>

The RDE legislation contains several requirements for the condition of the test vehicle prior to or during the RDE test. Table 6 provides an overview of the vehicle condition requirements.

It must be stated that the detailed information on the compliance of these requirements for the trips in the dataset were not known for this evaluation.

Table 6 Boundaries for vehicle conditions

Parameter	**Condition**

Air conditioning systems and other auxiliary devices	Operation shall correspond to possible use by a consumer at real driving on the road
Fuels, lubricants and reagents	Within specifications issued by the manufacturer for vehicle operation by the customer
Payload	Besides the driver, a witness, test equipment and power supply, artificial payload may be added (up to 90% of the maximum payload)
Preconditioning for cold start	Driven for at least 30 minutes, then engine off for 6 to 56 hours.

2.1.1 Complete overview

A complete overview of all RDE boundary conditions and its application in the implementation in this study is given in Table 7.

Table 7 Overview of available and considered boundary conditions in this study

Boundary condition RDE		**Availability**	**Considered**
Ambient conditions	Ambient temperature	Seldom	NO
	Altitude	Some poor	NO
	Road surface	Unknown	NO
Trip composition	Total duration		YES
	Shares U/R/M		YES
	Sequence of driving	Limited	NO
	Length of each section		YES
	Cold start	Unknown	NO
	Idling periods		YES
	Stops		YES
	Altitude gain	Some poor	NO
Driving behaviour	vapos		YES
	RPA		YES
	Average speed		YES
	Maximum speed		YES
	Speed range		YES
Vehicle conditions	Payload	Few	NO
	Preconditioning	Unknown	NO

2.2 Description of assessed evaluation methods

Evaluation methods are meant to correct for the raw emissions in the results associated with deviations in the RDE test executions. In this report four main evaluation methods can be distinguished: raw, Moving Average Window (MAW), Power Binning (PB), and NO_x/CO_2. The other options for evaluation methods are in essence based on these four with some modifications. Raw emission results, i.e., no correction to the results are central to the raw, the reference and the T&E proposal. In terms of boundary condition, the MAW boundary conditions appear many times: in MAW, the T&E proposal, and a part in the ACEA proposal. The analyses of MAW boundary conditions apply to all these methods. Likewise, the looped MAW is a variant of MAW, and best compared against MAW. The correction of NOx/CO2 and the ACEA proposal on the emissions are variants of the same approach, differing only for PHEVs. Therefore, corrections of the NOx/CO2 and the ACEA method are often analysed and discussed in conjunction.

Regulation 2016/427 (RDE1) describes the MAW and PB methods in Appendix 5 and 6 respectively. Appendix 7c of Regulation 2017/1154 (RDE3) describes the NO_x/CO_2 method. The method were raw measurement results are evaluated is clearly not meant to correct deviations in the results associated with the RDE test executions. This method only determines if the trip is valid or not according to the boundary conditions. The other three evaluation methods can cause differences in the final emission results. In this paragraph these three evaluation methods are explained in brief.

2.2.1 *MAW*

MAW or EMROAD found its basis in heavy-duty legislation. The average test data, i.e., windows, comparable to the length and operation of the WLTP is considered. In this manner idling and high accelerations are combined with other driving to normal operation.

An important step of MAW (EMROAD) is that emissions are averaged over windows. The length of a window is determined by the cumulative emitted mass of CO_2 for each time-window. It has to be equivalent to half the amount of CO_2 emitted during a WLTP test cycle. All consecutive windows move in increments of the sampling period (i.e. 1 second).

In another step the validity of a trip is assessed by comparing the CO_2 emission of a window against the so called CO_2 characteristic curve, which is determined by the the type-approval WLTP test. A trip is valid if at least 50% of the windows is within the CO_2 band, which is 25% below and a maximum of 30% above the vehicle characteristic CO_2 curve. This is one of the MAW boundary conditions. Failure to mee this criteria leads to an invalid test for MAW.

The final emission result of a valid RDE trip is obtained by a weighted average of the individual windows. The weight given to a window depends on the percentual difference with the vehicle characteristic CO_2 curve (wich range from 1 to 0).

A variation of MAW, called 'MAW looped', is also implemented and analyzed. Here, the first and last datapoint are connected, so that windows do not suddenly end at the end of the trip, but are continued through the first part of the trip. This ensures a sufficient coverage of the first and last part of the trip.

2.2.2 *PB*

The PB method bins emission data on the basis of the power demand of the engine. If tests have more or less than normal high power, or low power, moments, the data are reweighted such that emissions at the fractions of different powers are normalized.

In PB (CLEAR), a power curve or Veline is defined on the basis of CO_2 emissions during the WLTP type-approval test. It is a linear relation between power and CO_2 emission rate. Using this, instantaneous CO_2 mass flow during an RDE test can be used to estimate the power at the wheel. To assess the driving behaviour, this power signal will be compared to a power frequency table considered as ‘normal driving’. Power classes are defined on the basis of the vehicle mass and vehicle rated power, and each power class is only allowed a certain amount during a trip. This is called the power frequency table.
To calculate which part of the driving falls in each power class, the instantaneous powers are averaged in 3-second windows. Using these windows a histogram with power classes is made, counting the number of times a window falls in power class 1, 2, etc.

By comparing the power class histogram to the power frequency table, an assessment is done of how much of the driving is deviated from what is considered ‘normal’. The emissions are then accounted for this. Also, an upper and lower limit exists for each power class and acts as a PB boundary condition.

2.2.3 *NO_x/CO_2*

In the case the engine operation is decoupled from the driving behaviour, such as with hybrid vehicles, which can drive electric and charge the battery while driving, the CO2 emission can be a measure of engine operation and associated pollutant emissions. Scaling with CO_2 emissions may compensate for test variations and variation in engine operation.

This method sums all NO_x and CO_2 instantaneous emissions, and divides these. The outcome is multiplied by the CO_2 outcome in [g/km] of the WLTP type-approval test, thus obtaining a g/km outcome.

$$NO_x\ [g/km] = (NO_x[g]/CO_2[g]) * CO_2\_WLTP\ [g/km]$$

NO_x/CO_2 is applied in a default version and in ACEA’s proposal. In ACEA’s proposal for full hybrid and plug-in hybrid vehicles, the outcome is multiplied with the fraction of distance operation on the combustion engine. If the electric distance is half the total distance, the emission results are therefore halved in the ACEA proposal. The difference in results of NO_x/CO_2 and the ACEA proposal exists only for one PHEV vehicle in the RDE test data.

The advantage of the ACEA adaption of the NO_x/CO_2 from RDE3 is the natural treatment of cold start. No bound on the distance of ICE driving is needed, contrary to the hybrid vehicle evaluation NO_x/CO_2 method as described in RDE3.

2.3 Elements of RDE legislation under review

At the start of the study Regulation 2016/427 (RDE1) and 2016/646 (RDE2) had already been published. Regulation 2017/1154 (RDE3) was still not published but only voted on in the technical committee. The comitology text was used to determine the latest status of the RDE legislation.

2.3.1 *RDE packages 1, 2 and 3*

The trip composition and the evaluation methods for conventional ICE vehicles were covered in RDE1. In the final text the same not-to-exceed (NTE) limit was applied to the urban part separate, although the methods were not fully aligned to yield the same definition of the urban part. RDE3 also included cold start, which raised further issues regarding trip composition and trip dynamics at the start of the test. Moreover, there are artefacts in the evaluation methods less suitable to deal appropriately with the cold start. In particular the low weighing of data of the first minutes in the MAW windowing method was identified as problematic. RDE3 also included a hybrid vehicle method based on total measured CO_2.

2.3.2 *Critical elements in the RDE text*

The RDE legislative text of RDE data processing and evaluation makes for difficult reading. In the appendix a list of textual issues are explained. Below we give few examples of critical elements, which affect the outcome.

For example, in MAW, “vehicle ground speed < 1 km/h” “shall not be considered for the calculation” (Appendix 5 3.1), is open for interpretation. Are this data to be excluded, or are windows not to start or end at, or differ by these data, or is another interpretation proper? In the implementation, the statement is more or less ignored as non-sequitur. Moreover, in a second issue with MAW, the law text describes an ‘upper positive primary tolerance 1’, but uses the same variable tol_1 in the equations as describing an upper as well as a lower tolerance. This will be discussed in section 7.3. Likewise in PB, the words “up to Class 5” (Appendix 6 section 3.6) does not clearly indicate if Class 5 is to be included in boundary condition on bin coverage.[4]

These matters of the legislative text are reviewed and discussed in Appendix 1.

Alternative methods like NO_x/CO_2 or raw are much simpler methods, requiring only a few paragraphs of description. However, the method relies on a relation between pollutant emissions and CO_2, which is subject of this study. As far as known there is no ambiguities in the interpretation. Some minor issues have arisen with the ACEA proposal. Initially, in RDE3, plug-in hybrid vehicles (OVC-HEV) were to be evaluated only if more than 12 kilometres were driven with the ICE on. In the ACEA method, with the result weighted with the fraction of ICE distance, such requirement is superfluous. However, a minor numerical issue remains as NO_x/CO_2 can be suffer from rounding errors. If not the same amount of CO_2 is used to establish the ICE distance fraction, the small amount of CO_2 in the denominator of NO_x/CO_2 is not cancelled by the small amount of CO_2 in the fraction of the distance with the combustion engine on. Hence, it is important to define ICE fraction of the distance in terms of minimal rate and duration of CO_2 emissions. For example, the state of combustion ‘engine on' could be defined as the periods of more than three seconds that the CO_2 emission is more than 0.1 g/s. Periods of up to 60 seconds without CO_2 emissions and velocities above 1 km/h in between moments of engine on are to be included in the engine on period in the absent for an engine speed signal. Based on this ACEA proposed the following formula:

[4] A number of items are checked with the developers of the methods. But, as it turns out on a number of essential parts the software implementations differ from the legal text. For example, the criteria of the maximal share in the lowest power bin is 60% in the legal text, and 65% in the CLEAR software of the PB method.

$$NO_x\ [g/km] = (NO_x[g]/CO_2[g]) * CO_2\_WLTP\ [g/km](distance\ ICE/total\ distance)$$

The definition of “distance ICE” should be based on CO_2 in the same manner as the accumulated CO_2.

3 Received RDE data

This chapter discusses the received data of the stakeholders, and review some important properties. Data are treated very much “as is”. All data are held against the same criterions and conditions. The analyses processes are automated to such an extent that in a single run a particular analyses is performed of the full set. Some pre-processing work was needed to make this possible.

Raw data were requested, as to ensure not only “valid” or “repaired” data were received, This allows the analyses of the exclusion of data on the different grounds. It seems that data outside the realm of RDE was included in the sets, which warranted a preselection on very generic grounds such as trip length and velocity distribution.

Data received
Data were obtained following the request of the Commission, to deliver RDE-type data, or ‘normal driving data’. The data were collected by JRC and provided to TNO. Raw measurement data were requested to ensure that data which would fail boundary conditions, or could not be processed by a tool, was not a priori excluded. This would have given a bias towards valid data only, excluding part of the more difficult RDE data.
At same time, however, TNO does not have complete understanding over the way data were gathered, e.g. how tests were performed and under which conditions.

In total about 350 test files were supplied to TNO. Some files were incomplete, other concerned Euro-5 vehicles, not to be included in the study. As no further information of the vehicles was supplied after the initial presentation for the stakeholders in July 2017, some Euro-5 vehicles may still be part of this data. Also some basic requirements on trip length and velocity distributions were applied to exclude trips which were inappropriate as input for the evaluation methods. Eventually, 252 trips were identified as appropriate input for the evaluation methods.

3.1 Raw data, PEMS output files

The data files of the different stakeholders and sources were put in a uniform format. In the output the following columns were retained, converted, or generated:
1. party[-]: stakeholder who supplied the data
2. vehicle[-] unique vehicle number per stakeholder
3. test[-] unique test number per vehicle
4. time[s] from the start of the test
5. vel[km/h] velocity, typically from GPS.
6. acc[m/s2], (midpoint), acceleration according to the midpoint rule
7. NO_x[mg/s], raw NO_x emission rate
8. CO_2[g/s], raw CO_2 emission rate
9. CO[mg/s], raw CO emission rate [optional]
10. lambda[-], air-fuel ration, if provided
11. EMF[g/s], exhaust mass flow
12. T_amb[C], ambient temperature [optional]
13. GPS_alt[m], altitude from GPS

14. GPS_lon[deg], longitude from GPS in decimal degrees [optional]
15. GPS_lat[deg], latitude from GPS in decimal degrees [optional]
16. PN[1/s], particulates number [optional
17. RPM[1/min], engine speed [optional]
18. ECT[C], engine coolant temperature [optional]

In a number of cases not the emission rate but the pollutant concentration was given, together with exhaust mass flow, or exhaust volume flow. Then, the rates were calculated in accordance with RDE1. When the fuel type information was missing, the average standard density for petrol and diesel exhaust gas was used (1.2943+1.2931)/2 = 1.2937 kg/m^3 . The effect is very minor.

Some of the received datafiles have higher sampling rates than 1 Hz, which is recommended as part of the legislation. The equations in RDE were written in such a manner that different, and even fluctuating sampling rates could be used. For example, the midpoint acceleration for varying sampling rates is given by:

$$a_i = (v_{i+1} – v_{i-1})/(3.6*(t_{i+1} – t_{i-1})),$$

where t is the time in seconds, and v is the velocity in km/h. This generalization of the equation in the legal text is used. This acceleration signal can be filtered, smoothed, or resampled. In current study, the original higher sampling rates are retained.

The legal text has some aspects treated at 1 Hz specifically, but in a number of cases, the text accepts sampling rates of 1 Hz and higher. The broader interpretation is used in this study, However, the 8-10 Hz datafiles are very limited and have minor impact on the overall result.

3.2 Conversions

In the conversion from volume flow to mass flow special care is taken in the reference temperature for the volume flow. Both the standard and the normal conditions, with different reference temperatures and pressures, are used in the units of flow. Hence different reference temperature are available. In these conversions the ideal gas law is assumed, e.g.:

$$\text{Volume [at T=273.15 K]} = (273.15/293.15) * \text{Volume [at T=293.15 K]}$$

In a number of cases the volume flow conditions or units were guessed based on the knowledge of the familiar output files of similar PEMS equipment.

3.3 Multiple signals

In a number of cases multiple signals were available, for example, for vehicle velocity. In general, the GPS velocity was used. This signal is sometimes noisy, and sometimes more smooth than one would expect, which might mean that some smoothing took place before submitting the data. In a few cases only OBD velocity is provided. In that case the signal is used, although there were some concerns on the accuracy.

3.4 General criteria for inclusion in the analyses

A number of trips were shorter than 4000 seconds, while the minimum RDE trip duration is 5400 seconds according to the legislative text. Moreover, a number of trips did not have the full range of velocities, from urban driving below 60 km/h up to 110 km/h for motorway driving. In the case the essential signals, velocity, NO_x and CO_2 were missing or the data were not RDE-like by the criteria of minimal length and velocity coverage, the data were excluded from all considerations. This removed about 100 files from the dataset, and it left 252 data files.

3.5 Default input data

In some cases not all the input data to apply the PB and MAW evaluations is provided. In 72% of trips, MAW input data were available, and for PB this value was 50%. In particular the CO_2 emissions in the different WLTC phases is needed, and for PB either the Veline offset and slope for the relation between CO_2 and power at the wheels, or the WLTP road load coefficients and WLTP test mass in combination with a WLTP second-by-second test data are needed.

In the case of limited input data, t/he first option is to use regression formulas for either the total WLTP CO_2 or test mass. If that data are not available as well, default values are used. Both the regression formulas and the default values are determined from the CO_2MPAS database. These regression results were provided by JRC.

The regression formulas are given in terms of the reference CO_2 (Mref), which is related to the WLTP total by: Mref*2/23.266 [g/km]:

WLTC low = 1.3684 * (Mref*2/23.266) - 16.598
WLTC high = 0.8694 * (Mref*2/23.266) - 0.3216
WLTC xhigh = 1.0656* (Mref*2/23.266) - 4.862

If only mass is available, the Mref would have been determined from the default Mref scaled by the ratio of the actual mass and default mass.

CO2MPAS DB	
vehicle_mass [kg]	1760.491
WLTC_low [g/km]	201.4497
WLTC_high [g/km]	139.5868
WLTC_extra_high [g/km]	167.476
Mref [g]	1872.603
f_0 [N]	164.179
f_1 [N/(km/h)]	0.622031
f_2 [N/(km/h)^2]	0.037966

3.6 Data processing

Whether filtering, alignment, or other pre-processing steps should be applied or not, is not fully specified in the RDE legislative text. If no evaluation is applied, such pre-processing steps on velocity data are hardly relevant, as the effect on total distance is expected to be minimal. The more complex and intricate the dependencies of the evaluation data on the velocity data are, the more sensitive the outcome on these processing steps will be.

Little processing is applied to the data that were provided. Some poor quality velocity signals and altitude, both rough and too smooth, were observed. But these input data were used unaltered. In part the results presented in this report include the effects of the variation in data quality as observed across the provided data.

4 Possible bias of the analysis

In this chapter some remarks on a possible bias of the analysis are given.

The data and the evaluation methods are taken “as is”. No adaption is applied to either the method or the data. Consequently, the study can be perceived as an input-output analysis. Simply stated: given the whole dataset, what is the effect of the evaluation by the collective effects, in terms of average, spread, and outliers.

The study is not intended to investigate or exemplify effects by altering data or methods. Therefore, input data could not be changed at will, to highlight certain effects. Instead, a proper representation and interpretation of the results should expound the robustness of the evaluation methods. In these representations three items were identified as essential. These items will be discussed below in more detail. First, the sensitivity for details in the procedure, such as the smoothing of data or accuracy of the signals. Second, the sensitivity for the variation in the input, in particular the reference WLTP values used for the different vehicles. Thirdly, the systematic effects observed, such as a varying correction with execution of the test.

4.1.1 Check of implementation

Due to the difficulty and ambiguous interpretation of the PB and MAW legislation, a comparison was done with the common PB and MAW implementations CLEAR (version 2.0) and EMROAD (version 5.95B01). Two RDE trips were evaluated using TNO’s implementation of CLEAR and EMROAD. Emission results showed a difference smaller than 1%. It should, however, be noted that these implementations showed deviations from the legislative text, as is discussed in Appendix 1.

4.2 Sensitivity for steps in the processing

Some data look more smooth than realistic data would look like. In the case of GPS, the raw data from satellites is usually processed with a Kalman filter, before it is recorded. Consequently, smooth data are not always correct data. On the other hand, noisy data can be smoothed, or not. This mainly affects the $v*a_{pos}(95\%)$ determination, which is sensitive for the smoothness of the velocity signal.

Comparing the raw velocity signal against the filtered velocity signal, the determination of the v*apos(95%) leads on average to similar results as shown in Figure 1. However, some scatter arises from the filtering process. In particular the motorway $v*a_{pos}(95\%)$ shows some scatter, reducing the value in a small number of cases significantly.

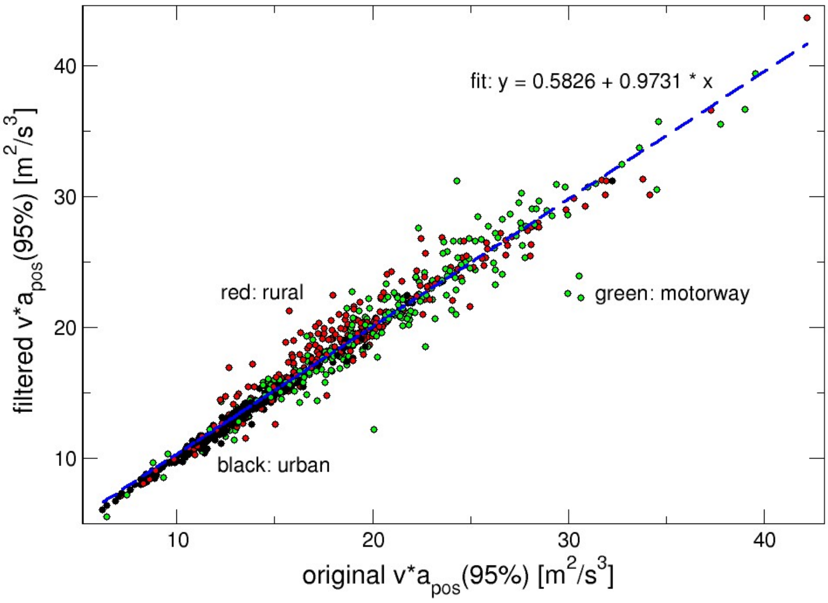


Figure 1 Raw velocity signal against the filtered velocity signal

Another feature of the signal, is the alignment of the emission data with the velocity data. It is not investigated what effects misalignment may entail. It is however observed that a number of test files show a large misalignment between the velocity and the emission data.

4.3 Sensitivity for variation in the input

Apart from the test data from the RDE trip, most of the evaluation methods use some WLTP data as a vehicle reference to compare and weigh the RDE test data against. This WLTP input data refers to certain vehicle tests on a pre-production vehicle, with a fixed test protocol, given mass and road load. These elements may not be fully appropriate for the RDE tests, which can be executed at different ambient temperatures, with a wide variety of payloads, and with or without cold start. Hence, if the dependency on the WLTP input values is strong, valid RDE tests may be excluded or adapted based on the deviation between the WLTP value and the RDE tests. As mentioned in paragraph 3.5, the input data were available for 50% for PB and 72% for MAW. In all other cases, default values and regressions based on these default values are used. This may lead to deviations.

4.4 Systematic effects in the evaluation prone to exploitation

Part of the study is the robustness of the evaluation methods against defeat devices and intentional misuse. Possible artefacts of the evaluation method can only be prone to exploitation if they lead to systematic evaluation corrections upward or downward. If the tests with the same vehicle or vehicle technology show sufficient variation in the raw emissions, but despite the variation all results are corrected in the same direction, then this will be open for abuse.

These artefacts which are open for exploitation can be examined by looking at the test data of several tests with the same vehicle. If the test data vary, it can be assumed that some variation occurs in the test conditions or test execution. It should follow that the corrections by the evaluation methods may remove outliers with respect to the raw test result if the test deviates from the reference. This is the effectiveness of the evaluation.

Moreover, the correction of the raw result would suggest there is an RDE test which may represent the reference test and conditions, at which no correction applies. If the evaluation data, as function of the raw data, lies on a line, the intersection point of this line with the diagonal should be the reference test conditions and execution. If the line is parallel to the axis, it seems there is a systematic correction upward or downward, depending on whether the line is above or below the diagonal. An appropriate correction would lead to an intersection at the reference test conditions. If for the same evaluation method both parallel lines above and below the diagonal occur, it could suggest technology dependent systematic effects.

All these conclusions based on observations of the effects of evaluation methods on the provided data are mere projections. The underlying assumption is the appropriateness of the RDE test data, to draw these conclusions. For example, assume that the emissions of a vehicle depends on the ambient temperature, for which no evaluation correction applies between 0° and 30° C. But if at one temperature the test is executed differently than at another ambient temperature, it may appear that the variation in emissions is the result of the test execution. The results of the evaluation methods may not, or only in a convoluted manner, be affected by the effects of ambient temperature on the emissions. This will be specific for different technologies.

It is assumed the RDE test data are not systematically biased with respect to unknown dependencies, which are not incorporated in the evaluation. A more complex situation would be when the emission characteristics of a vehicle depends on the ambient temperature. In this case a complex dependency of the test results on the RDE test may invalidate the conclusions, because the underlying assumptions of origin of the variation of the emissions are not satisfied.
Apart from ambient temperature, there are a large number of other hidden parameters which may affect the results, yet only indirectly related to the test variations accounted for in the evaluation methods. It is to be expected that for properly functioning emission control, many of the normal variations in driving behaviour are covered by the control strategy. Effects are therefore unexpected and related to more intricate variations and combinations of variations.

5 Signal quality and signal processing

Emissions, velocity, and altitude are the essential signals for processing RDE tests in the evaluation methods. They are defined in the RDE legislation and requirements for the signal quality are given for these signals. PEMS signal quality is defined separately, and is not discussed here. The quality of the velocity and altitude signals are defined in limited terms. GPS is the basis for both signals in most of the data received. GPS signals can vary greatly in accuracy, both between equipment, and in between tests. It is not a priori clear that a GPS signal is poor, unless it disappears completely. Only by cross validation against other signals it can be observed that sometimes GPS signals do not follow the true values, but does produce a broad sense of the trip route and positions.

The need for accurate velocity and altitude is determined by their use in the evaluation methods and the boundary conditions. In many cases these applications are robust, and depend only weakly on GPS signal quality.

5.1 Acceleration

The most sensitive signal in the boundary conditions is the acceleration derived from the velocity by finite difference. Errors and noise in the velocity signal is enhanced in the derivation of the acceleration. Moreover, the result depends strongly on sampling rates and the order of processing. For example, if the quality is too poor because the finite resolution is in the order of 0.1 km/h.

5.2 Altitude gain

Most data have several hundreds of exceedances of the criterion: $|h(t) - h(t-1)| > v(t)*\sin(45^o)/3.6$, because the velocity v may be zero. Figure 2 shows an example of a poor altitude signal which contain many illogical values at zero altitude. Any noise in the altitude signal while the vehicle is stationary leads to an exceedance. This is not necessary a problem in the subsequent processing. Excluding zero velocity data points from the criterion of maximal altitude differences about half the data sets have some exceedance. However, the number of samples with exceedances are typically less than 20. Only a few sets remain poor. If such sets are plotted, it is clear the GPS is not performing well.

The application twice of the linear filter of the finite difference of the altitude signal based on the altitude gains over 400 metres can be simplified to a single equation. The oscillations, or noise, are dampened and shifted by this approach. Undulations in the road below 200 metres are not accounted for in the total altitude gain. Altitude gain is the sum of all height differences from the lowest point to the highest point, discarding oscillations and errors at a short time scale. Determining the precise location and height of maxima and minima in the route by fitting the peaks and valleys with the data collected over 400 metres might lead to a method which can be checked against a repeat test, or with map data.

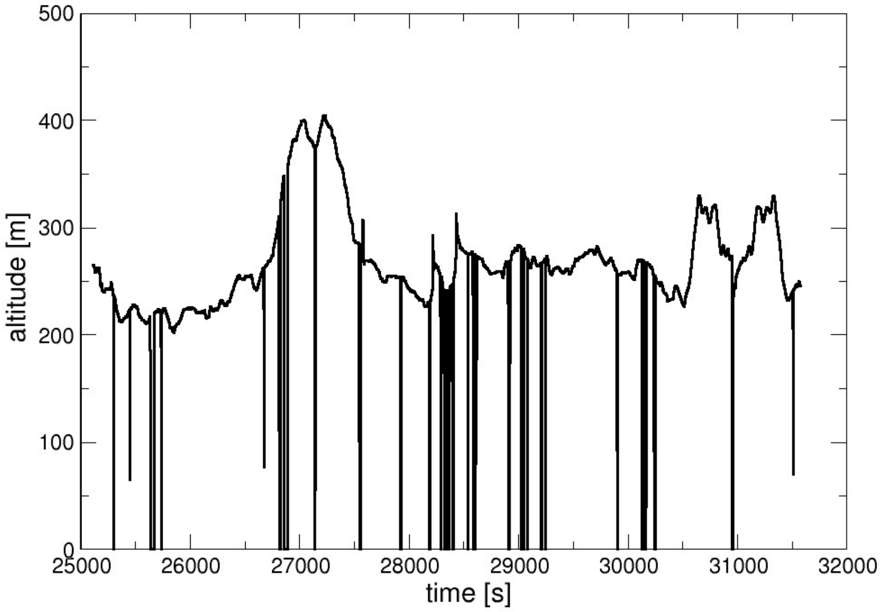


Figure 2: Example of a poor GPS altitude signal Peaks in the altitude signal after signal loss (defaulted to zero) indicate some specific filtering artefacts.

6 Assessed RDE Boundary conditions

In this chapter, the general purpose of boundary conditions is discussed. Also, the obtained data were analysed in the light of these boundary conditions and remarks on these boundary conditions are described.

6.1 Purpose of boundary conditions

Boundary conditions and evaluation methods are sides of the same coin. If the boundary conditions are restrictive, this leads to median testing, which may still contain many aspects of normal driving, but not all possible combinations. In that case, the evaluation method may be unnecessary, as results are close together. On the other hand, if boundary conditions allow too much freedom in testing, evaluation methods may be an essential part to compensate for overrepresentation of certain parts of the test.

In a purist view, one could argue that neither the evaluation method nor the boundary conditions are required. This is, however, easily falsified by a test in which the vehicle remains stationary: any emission at no distance covered will lead to an infinite g/km value. Hence, even though stationary operation can be considered normal for part of vehicle operation, it can never be the basis of a test result, without either (1) boundaries on how much driving should occur, or (2) an evaluation method that limits its impact in the final result, or both.

The relative severity of different boundary conditions can mean that one boundary condition is superfluous, and its purpose is already covered by other, more severe boundary conditions. On the other hand, certain boundary conditions may be complementary to the extent that insisting on all boundary conditions may lead to a very small subset of valid tests.

It is not always clear what the purpose of boundary conditions are beyond producing valid and robust outcomes of the evaluation method. For example, boundary conditions may ensure enough underlying data exist to perform a reweighing. Such criteria are not necessarily proper in the light of the purpose of RDE legislation. In particular if it leads to too many invalid tests, and it puts strain on the execution of the RDE test, which is meant to be any normal and safe driving in normal traffic. Consequently, the boundary condition can be impractical. However, if the boundary condition is essential for the evaluation method, the more far stretching conclusion would be that the evaluation method is impractical.

6.2 Trip length and composition

Trip length and trip composition are part of RDE1. The conditions are a mixture of time related, and distance related conditions. The urban, rural, and motorway distance should each be about a third of the distance, which means that in the 90-120 minutes trip, almost an hour is driving urban, and about 15-20 minutes on the motorway. In general these conditions are satisfied in the data received, once the short trips (below 4000 seconds) were excluded.

6.3 Velocities and stopping

The maximal velocity of 145 km/h can be exceeded by 15 km/h for a small fraction of the time. Hence velocities over 160 km/h would invalidate the test.
There are a number of test supplied in the data where velocities of 170 km/h are driven. These tests are labelled as invalid in most of the evaluation methods as well. On the other end of the spectrum, a number of tests are invalid because of the minimal time of idling of 6% in the urban part is not met. Since all velocities below 60 km/h are included in this evaluation of the idling boundary condition, it may be that the first part of the trip did satisfy the 6% criteria, but rural driving added more data, which invalidated the result.

6.3.1 *Urban and total trip results*

In RDE2 the urban part is evaluated separately to the NTE limit. The urban part, generally defined as all velocities below 60 km/h, is in each evaluation method a separate entity, needed for the reweighing of data to 34%/33%/33% distance split of urban, rural, and motorway. It is not related to the map data, and the urban part can incorporate substantial amount of rural and motorway driving, depending on the test.

According to RDE3, urban driving up to 60km/h needs to be the first part of the test for cold start. Thus, the first time the velocity is higher than 60 km/h can be construed as the end of the cold start and the first non-urban driving. Many tests are therefore invalid according to the latest state of RDE legislation with RDE3, in the case of cold start testing. Comparing the distance up to that point with all driving below 60 km/h, the fractions in the total distance deviate substantially. See Figure 3 below.

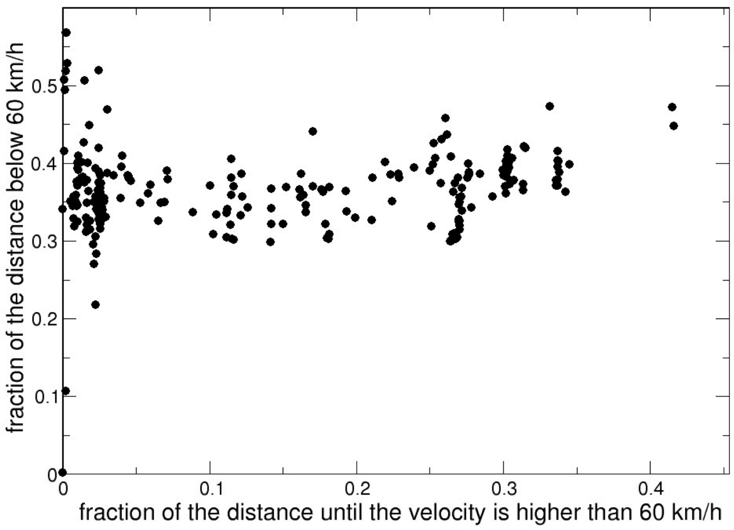


Figure 3 Comparison of the total distance below 60 km/h and the driven distance up to the point that 60 km/h is exceeded on all tests

Hence, it is expected that the results of the urban evaluation deviate more than the results for the total trip. For the urban part, not only the effect of the evaluation method, but also the emission data incorporated in this part vary between the methods.

6.4 Trip dynamics

RPA and $v*a_{pos}$ are the quantitative indicators of trip dynamics, low and high dynamics respectively. In the data provided, the exclusions of trip dynamics are mainly on the basis of $v*a_{pos}$[95%], as shown in Figure 4. On the basis of RPA, or slow driving, there are few invalid tests, see Figure 5. The conditions for trip dynamics are separate for urban, rural, and motorway. Rural $v*a_{pos}$ seems the most demanding condition.

The red dots above the line are invalid trips on one or multiple boundary conditions, where the line represents the boundary condition of $v*a_{pos}$. The red dots under the line represent invalid trips which are valid on $v*a_{pos}$ but invalid on other boundary conditions. Below the relation MAW and PB is discussed.

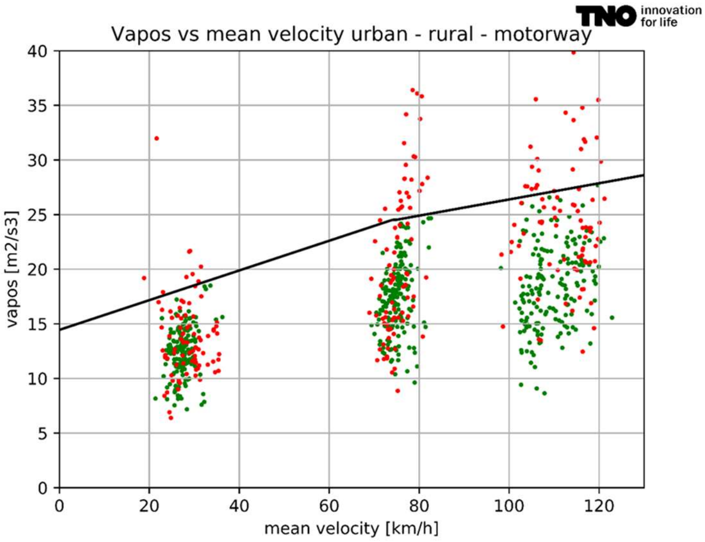


Figure 4 $v*a_{pos}$[95%] versus mean velocity for the three velocity bins. Some tests are invalid based on this condition, for one or more velocity bins.

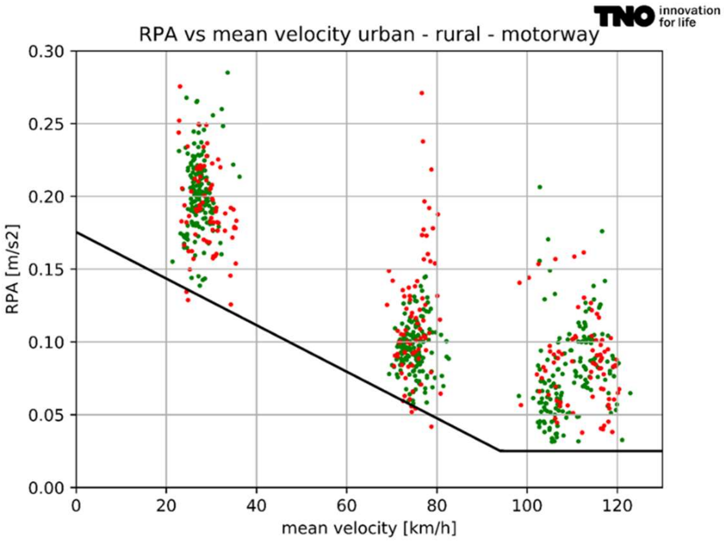


Figure 5 RPA versus mean velocity. Very few tests are excluded on the basis of the RPA boundary.

6.4.1 *Comparing $v*a_{pos}[95\%]$ with MAW and PB*

If for the urban part $v*a_{pos}[95\%]$ is compared with the evaluation methods little trend is observed in the correction as in the test validity. For the urban part a one-to-one comparison is more easily made since a single $v*a_{pos}[95\%]$ condition is applied. For both MAW and PB, valid tests above the $v*a_{pos}[95\%]$ limit occur. Moreover, MAW show a lot of scatter in the corrections, uncorrelated with the $v*a_{pos}[95\%]$ value.

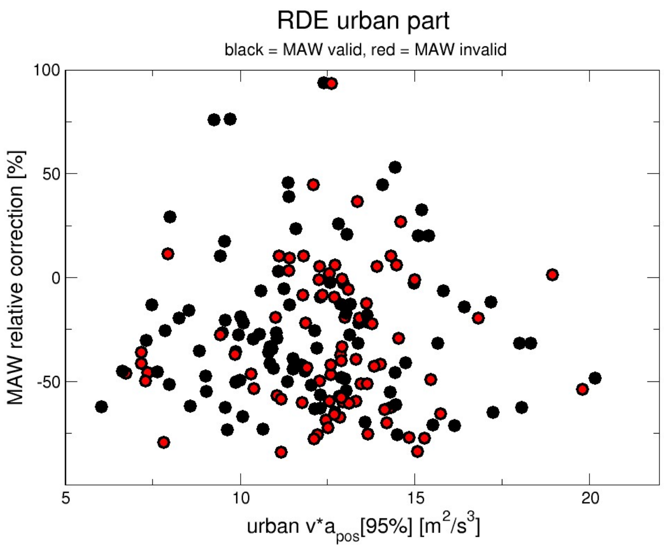


Figure 6 The MAW relative correction plotted a function of the $v*a_{pos[95\%]}$ values of the urban part. The red points are invalid according to MAW boundary conditions. The black points are valid.

It would be natural if the corrections closer to the $v*a_{pos}[95\%]$ limit are downwards. The evaluation methods are intended to correct mainly for driving behaviour.

Hence, reaching the aggressive driving boundary of v*a_{pos}[95%] can be expected to give a downward correction. This is not observed.

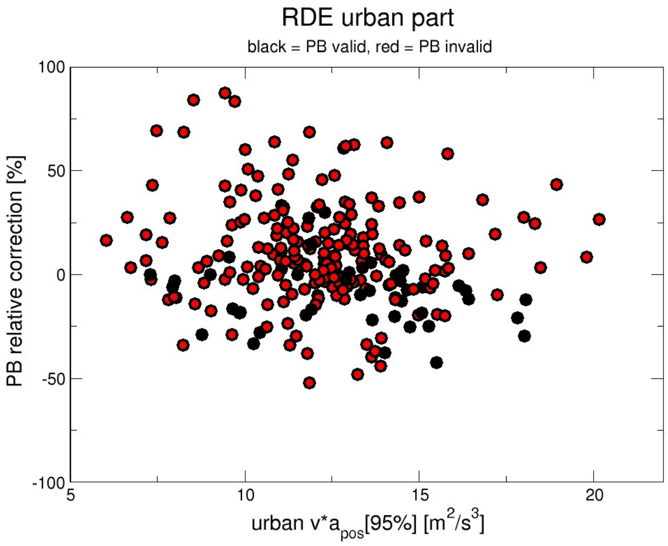


Figure 7 The PB relative correction plotted a function of the v*apos[95%] values of the urban part. The red points are invalid according to PB boundary conditions. The black points are valid.

7 Analysis of Moving Average Windows

In this chapter, the MAW evaluation method is analysed, as described in Appendix 5 of RDE 1. The analysis will start with the MAW boundary conditions which are related to fit-for-purpose, followed by the window velocity, an analysis of the emission result compared to raw emissions, discussing systematic effects and potential abuse, and lastly we will discuss the looped variant of MAW.

Moving Average Window (MAW) plays a central part in the review of evaluations, as it is included in multiple analyses. The method in the legislation is also implemented in an augmented, looped approach which ensure sufficient coverages of the first part of the RDE test, most relevant for the assessment of cold start emissions. Thus, the evaluation of “Baseline EMROAD” and “EMROAD modified” (looped MAW) is discussed in this chapter.

Aspects of the MAW are also used solely as boundary conditions, to restrict the test validity. In this case, the values are not corrected according to MAW, but the boundary conditions of MAW are applied. Two methods use MAW boundary conditions to check for trip validity, namely “In-between (T&E)” and “ACEA new proposal”. These methods will be discussed in chapter 9, but it is good to keep in mind that the boundary conditions of MAW, which will be discussed next, apply to these methods.

7.1.1 ACEA’s proposal

In the case of ACEA’s proposal, to make NO_x/CO_2 applicable to all vehicles, they suggested the tolerances for test normality to change to:

- STD ICE / MHEV from -25% to +30% (consistent with MAW)
- NOVC-HEV[5] from -25% to +50%
- OVC-HEV N/A from-100% to +50% (including electric driving)

The +30% upper positive tolerance for the conventional and mild hybrid vehicles is no change to the current legislative text, as the tolerance is increased from 25% to 30% anyway, if the coverage is initially insufficient. There is no indication that a NOVC-HEV is in the dataset, so this rule is not applied. The only OVC-HEV (PHEV) has therefore adapted MAW boundaries. This difference leads to a much higher inclusion rate than the normal MAW boundary. This will be discussed further in the section on the ACEA proposal.

7.2 MAW boundary conditions

In this subchapter, the MAW boundary conditions which account for the large portion of invalid trips of MAW are analysed. In total 116 tests (75+12+19+16) are valid in MAW from the 252 RDE like tests. From the tests that satisfy trip dynamics and trip composition boundaries (93+75), 93 are excluded on the basis of the MAW boundary conditions. Moreover, a significant of MAW valid trips are considered invalid according to trip dynamics (19), trip composition (12) or not valid at all (16).

[5] PHEV (Plug-in Hybrid Vehicle is the same as OVC-HEV: Off Vehicle Charging Hybrid Electric Vehicle), as opposed to No Vehicle Charging Hybrid Electric Vehicle (NOVC-HEV). Mild Hybrid Electric Vehicle (MHEV), e.g., with stop-start systems, are shared among the conventional vehicles.

On the other hand, MAW seems to include trips which are invalid according to trip dynamics and trip composition. This can be seen in Figure 8.

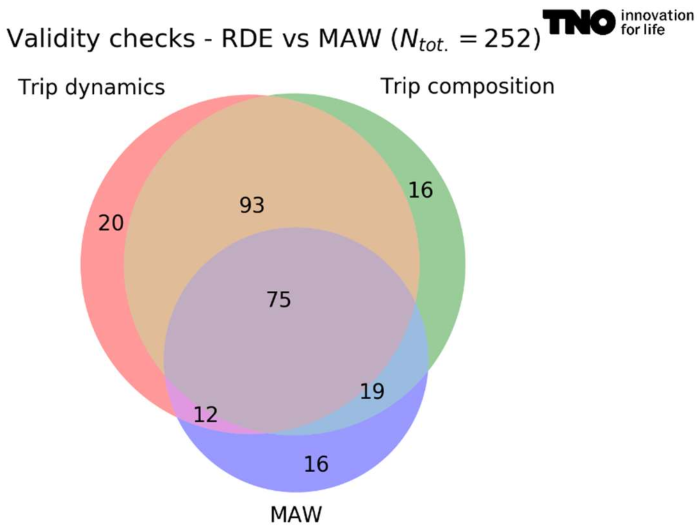


Figure 8 Venn diagram: validity checks: RDE vs MAW, one test was not fully processed.

MAW has two boundary checks: (1) the CO_2 band, where 50% of the MAWs have to fall within the CO_2 band as defined by the primary tolerance 1, and (2) the shares of urban, rural and motorway MAWs from total MAWs. In the following subsection, these two boundary conditions are analysed more thoroughly.

7.2.1 *U,R,M shares*

The first boundary condition is the share of urban, rural and motorway MAWs. As shown in Figure 9, 3 trips are invalid due to rural window share, and 95 tests are invalid due to motorway window share. None were invalid due to urban window share, thus were not plotted in the figure. It can be concluded that the motorway window share boundary is a highly contributing boundary condition. Often, the motorway share is too low for a valid test. These shares are based on the window average velocity, and a large portion is qualified as urban windows, whereas a relatively small portion is qualified as motorway windows, leading to a high number of invalid tests due to the motorway share boundary condition. A stop, or velocity reduction, on the motorway may bring down the average velocity of many windows, and thus invalidate the motorway share.

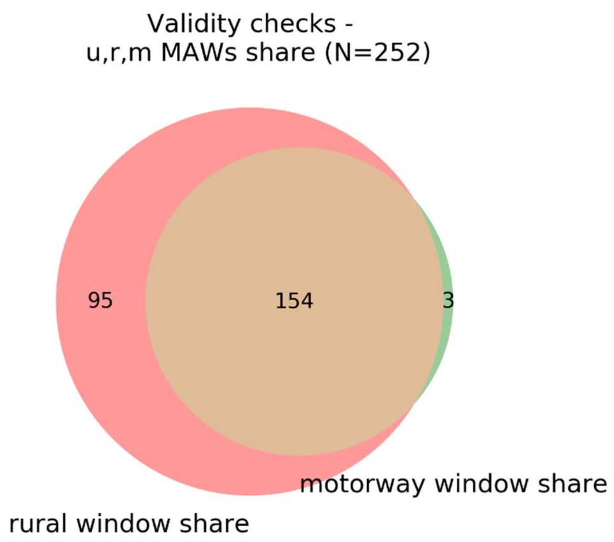


Figure 9 Venn diagram: validity checks: u,r,m shares

7.2.2 *CO_2 band*

In the section '*Test Normality'* which is described in section 5.3 of Appendix 5 in RDE 1, it is stated that the upper positive primary tolerance shall be increased with 1% if the trip normality is not valid.

As shown in Table 8, the effect of this legislative text seems limited for the available data in this study. For 206 trips out of 252, the primary tolerance was not increased. Secondly, for no trips, the primary tolerance was increased and the trip was subsequently found valid. And, for 46 trips out of 252, the primary tolerance was increased but the trip was found still invalid.

Table 8 MAW, invalidation on second count can be for other criteria than the 20%-30% windows.

	primary tolerance 1	#trips
	25%	206
Valid	(25 – 30]%	0
Invalid	(25% -30%+	46

Thus, it can be concluded that the increase of the primary tolerance to 30% has a negligible effect on the outcome of emissions, since these trips are often or always considered invalid. It is also found that the urban part of the primary tolerance is more often exceeded, which is shown below.

U,R,M shares vs primary tolerance

As displayed in Figure 10, it turns out the reason to disqualify RDE trips as invalid according to MAW is not so much the CO_2 band, but mainly the shares of windows in urban, rural and motorway driving. 90 trips which are valid according to the primary tolerance are excluded due to the urban, rural and motorway share boundary. In particular the motorway share is too low for a valid test. These shares are based on the window average velocity, which deviates substantially from the shares on the basis of the instantaneous velocities.

Consequently, adaptions to the CO_2 band will have limited effect on the trip validity. The CO_2 band and the tolerances that define them will not lead in the MAW boundary conditions validity. In the CO_2 band, the urban check remains an important part. Cold start, vehicle mass and braking may affect the CO_2 emissions. In the test data received, the urban tolerance band is as important as the motorway share to disqualify a RDE test.

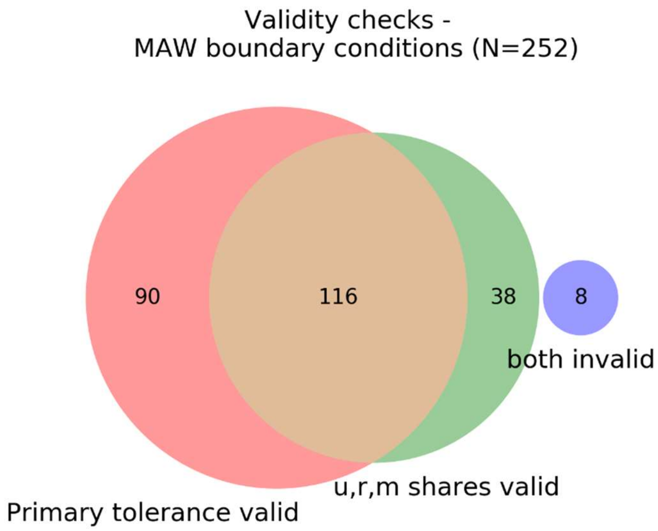


Figure 10 Venn diagram: only U,R,M shares vs primary tolerance (excluding other criteria).

Primary tolerance – urban part
Figure 11 shows a Venn diagram to compare two important boundary conditions: The motorway share and the primary tolerance for the urban part. This Venn diagram should be interpreted differently than earlier diagrams, since “% outside tolerance urban” is not a hard boundary condition. The percentage of MAWs outside the primary tolerance should *overall* be less than 50% in order to be valid. In this case, the trips with the percentage of MAWs outside the *urban* primary tolerance is less than 50%. Here the additional distinction is made of different windows to zoom in on the cause of invalidity. Hence, this is not an explicit boundary, since the percentage outside the primary tolerance *overall* can still be less than 50%.

The Venn diagram is therefore meant to be illustrative. It shows that for 105 out of 252 trips the percentage of MAWs outside the urban primary tolerance is more than 50%. Secondly, the motorway share is a hard boundary, implemented in the legislation, 80 trips are invalid due to the motorway share only.

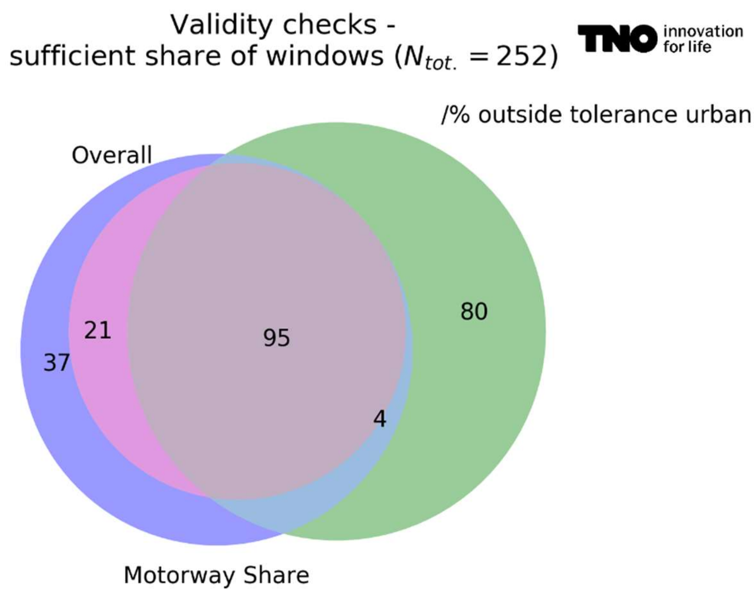


Figure 11 Venn diagram: Primary tolerance – urban part versus motorway share.

To conclude the analyses of the MAW boundary conditions, that two MAW (sub) boundary conditions account for almost all of the invalid trips of MAW.
More than half of valid RDE trips are invalid according to MAW. The boundary conditions of MAW seem to conflict with respect to the other RDE driving behaviour and trip composition boundaries.

Some interesting observations have been found regarding the legal text of the description of the primary tolerance, which are discussed in Appendix 1.

7.3 Instantaneous velocity versus window velocity

As stated in the previous subchapter, the average window velocity and the instantaneous velocity can differ, which results in a high number of urban windows. To analyse this more deeply, the instantaneous velocity and window velocity are compared, by time and distance shares. This comparison is shown in Figure 12 and Figure 13. For these figures the instantaneous velocity shares are compared with the window average velocity shares. Typically MAW shares are more average, with extremes excluded, by virtue of the windowing method. The results are calculated per trip.

The instantaneous velocity and window velocity show a large difference, both in time share and distance share, from which can be concluded that the trip composition does not fully restrict the share in MAW. Both in terms of time and distance the variation in MAW is much larger than in the RDE trip characterization.

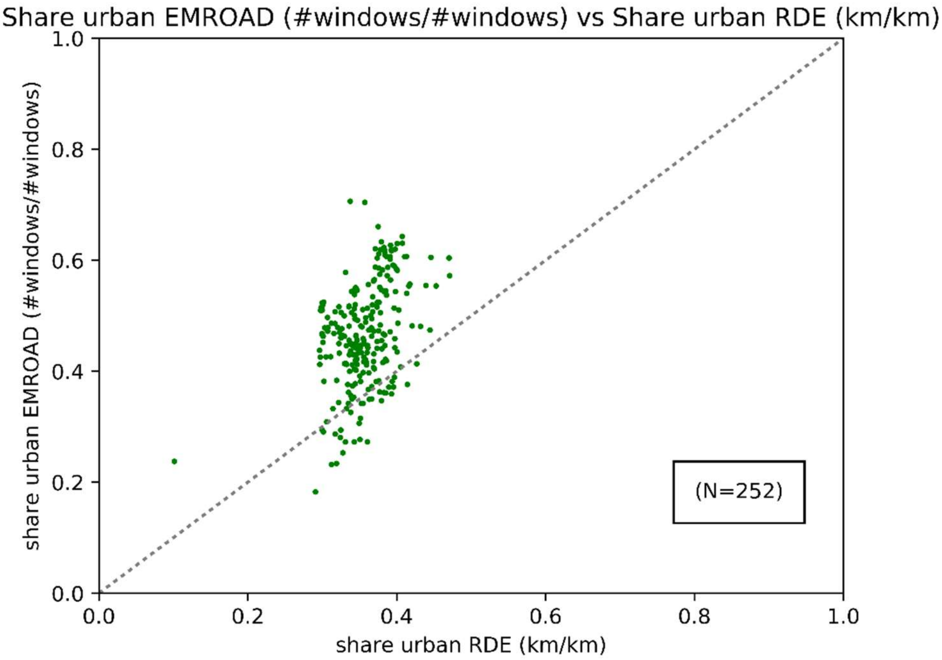


Figure 12 Comparison of the instantaneous velocity and the window velocity by distance shares (urban).

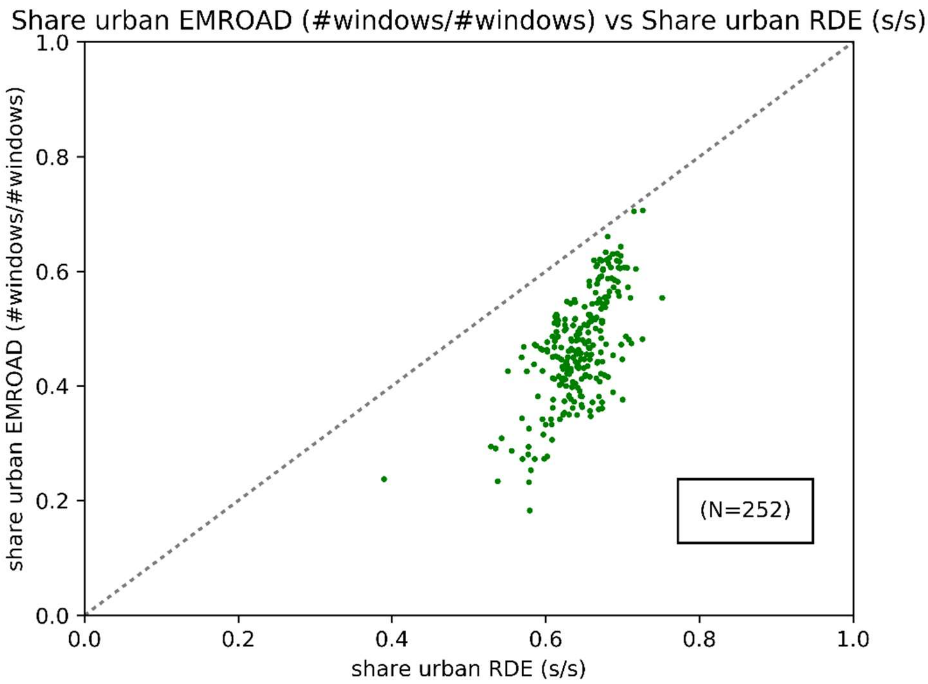


Figure 13 Comparison of the instantaneous velocity and the window velocity by time shares (urban).

Variation in urban, rural, and motorway shares in MAW can be considered an important contributor to the invalidity according to the MAW boundary conditions. If these conditions are plotted in the graph, the centre of the data cloud lies almost at the boundary. In the data there seems to be a mismatch of the window-based, the velocity based, and the map-based shares of urban, rural, and motorway driving. This is one of the root-causes for part of the invalid trips according to the MAW boundary conditions.

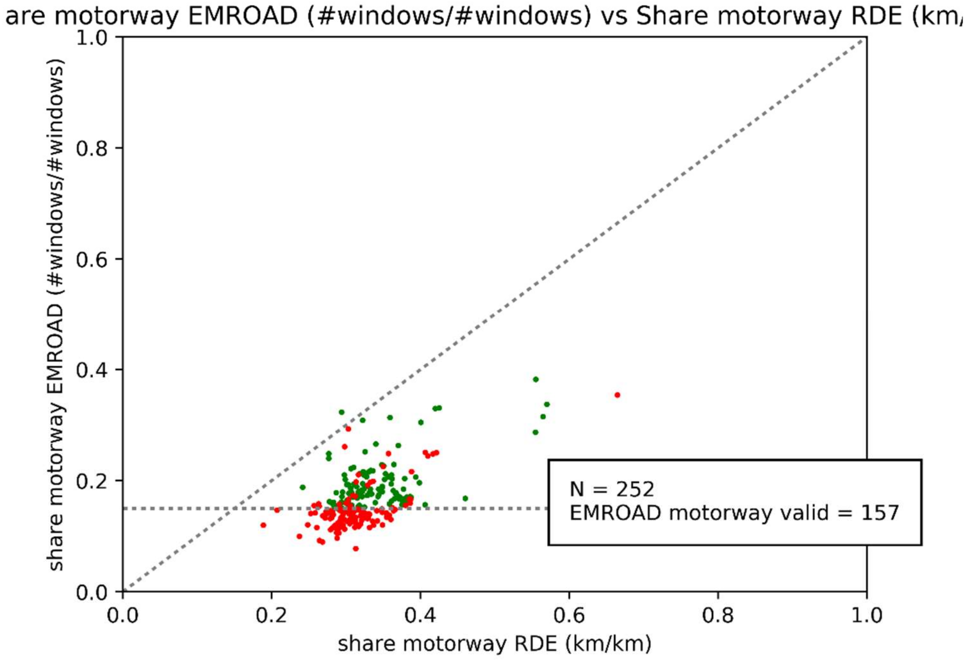


Figure 14 Comparison of the instantaneous velocity and the window velocity by distance shares (motorway).

7.4 Emission result - compared to raw emissions

In this chapter, the evaluation methods are compared with raw emissions. The effectiveness of MAW is analysed in chapter 10. In Figure 15 and Figure 16 this comparison is made. These pictures compared to raw emissions exhibit some scatter. This scatter may be taken as a sign of random variation. This does not have to be the case, as variations may arise as the result of both variation in vehicle technology and testing. Therefore it is important to look at the data from a ‘per vehicle’ perspective, which will be done in chapter 10.
However, with the scatter, still some systematic effects may arise: a general shift upwards or downwards in results, reduced variation for valid tests, etc. are all indications of the functionality of the evaluation methods. This will be discussed here.

The effect of the MAW evaluation on the result is not systematic upward or downward. The MAW evaluation result is slightly higher than the raw, or reference results. For valid tests only the deviation is even smaller. However, some outliers correcting the results by 50% both up and down occur. Some of these results apply to petrol vehicles with low total emissions.

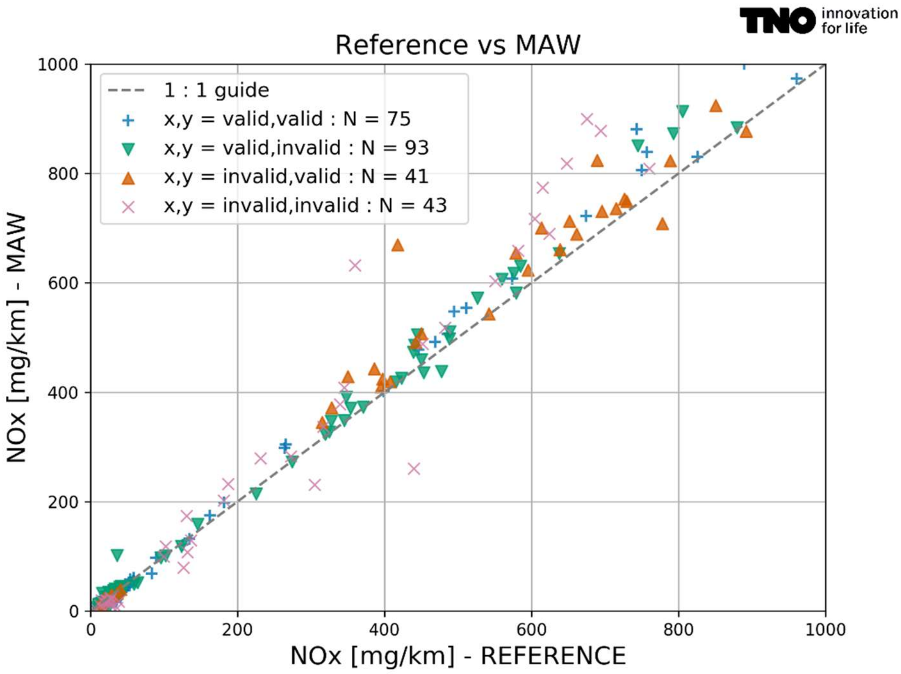


Figure 15 the evaluation method 'MAW' compared with raw emissions with RDE boundary conditions (reference)

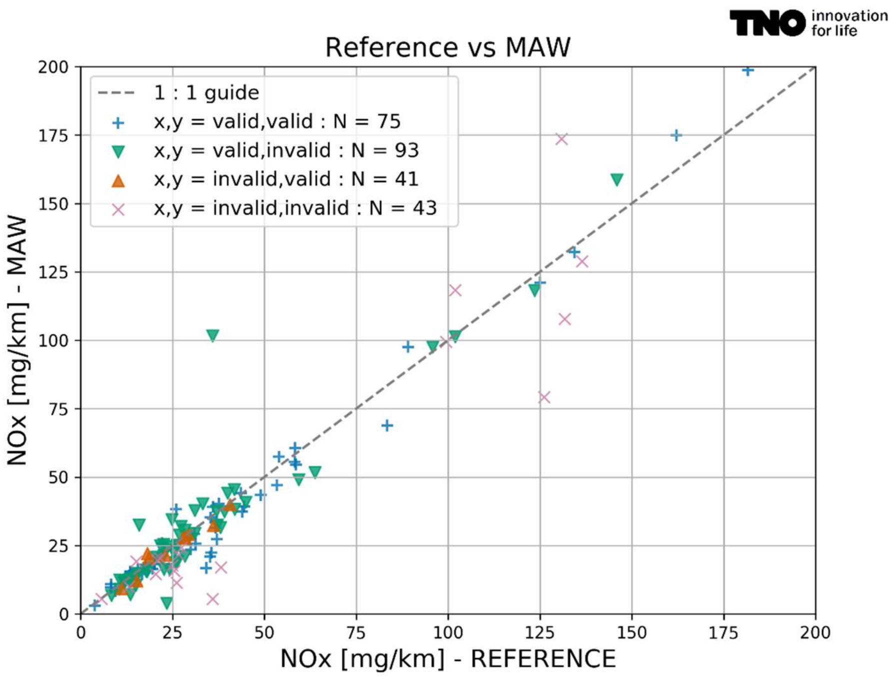


Figure 16 the evaluation method 'MAW' compared with the reference case (RDE valid raw emissions) with a focus on the lower emission results

7.5 Looped variant of MAW

Since windows have a start and an end, the first few seconds are included only in first few windows. Consequently, the emissions in the very first part of the trip, as they may occur during cold start are less represented in MAW. This can be changed, by looping the windows from the end, back to the beginning, starting a window at every second of the test.

This alternative of circular, or looped, MAW is tested as well, and compared against MAW as its alternative, see Figure 17. The differences are very limited in most of the cases, which may be related to limited cold start, or initial emissions. If the driving behaviour and emissions are at a consistent level over the urban part, little effect is to be expected. The effect, on the other hand, is of an unexpected nature. There turns out to be quite a large variation in the validity of the tests between the two methods, but not much on the emission outcome. On some tests, invalid by almost all accounts, some deviation is observed in the emission results. In all other cases the outcome is almost the same. The fact that little result is found in the outcome may be related to the absence of cold start tests or cold start effects in the tests. Cold start effects are expected mainly with SCR technology. Information on cold start or technology was not available in general.

The number of invalid results is lower for MAW compared to looped MAW. Very likely this is related to the fact that boundary conditions are not adapted for looped MAW, but the values are tuned towards giving valid RDE trips in combination with the other boundary conditions.

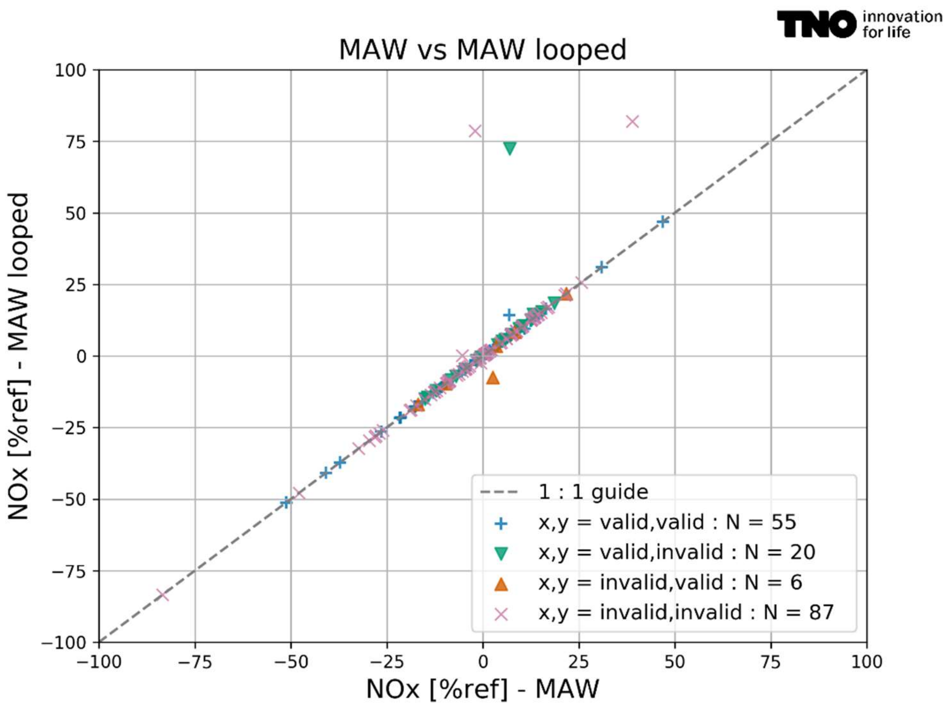


Figure 17 MAW versus MAW looped

8 Analysis of Power Binning

In this chapter, the PB evaluation method is analysed, as described in Appendix 6 of the legislative text of RDE1, with the amendments of RDE3 included. This chapter start with the analysis of the WLTP values, next the boundary conditions are analysed and fit-for-purpose of this evaluation, and finally the emission result, systematic effects and potential for abuse are shown and discussed.

The legal text and the official software tool posed some interpretation problems. The legal text does not match the latest version of the software tool. In particular the maximal amount of data in the lowest power bin is 65% for the tool and 60% in the legal text. Moreover, the minimal amount of data in the higher bins is not clearly stated in the text. “Up to class 5” can indicate it should include class 5, but also exclude this criterion for class 5. Eventually, the most logical interpretation for a number of unclear formulations, and the values for the parameters according to the legal text are taken. Issues regarding the reference velocity were resolved, as they pertained to two versions of PB software: using NEDC data and using WLTP input data. The correct version is with WLTP data, as RDE and WLTP are joint. The impact of different interpretations of the law text is investigated. The law text more thoroughly analysed and discussed in Appendix 1.

The result is a very large number of invalid tests in the power binning method. Out of the 168 valid trips according to trip composition and trip dynamics, only 34 trips are valid in the PB evaluation, while MAW had 75 valid trips.

8.1 Input WLTP values

Power binning reweighs the emission data on the basis of the instantaneous power, derived from the CO_2 emission rate. This requires some additional WLTP data to link power to CO_2. This is linked to the WLTP CO_2 interpolation method on the chassis dynamometer, for which the WLTP test data and chassis dynamometer settings are required. The outcome are an offset D and a slope k. Additionally, the road load values f0, f1 and f2 and rated power are needed.
These values were provided by a number of the stakeholders. So in many cases WLTP input values were modelled for the analyses. This was done in the following way:

1. For trips where absolutely no WLTP data are available, standard values are used, as obtained via the Commission.
2. If the WLTP Mref data is available for a specific trip, the standard values are scaled with respect to Mref to provide the unavailable input value.
3. If Mref was not available but vehicle mass was available, then the standard values would be scaled according to the vehicle mass.

A separate study of the sensitivity for these default values is carried out on the data for which all WLTP input values were provided, this can be found in chapter 12.

8.2 Boundary conditions: Trip validity by fractions in the power bins

The power binning method seems complementary to the trip composition and dynamics boundary conditions. The number of invalid trips in the PB evaluation is in the order of a quarter for each of the subsets of validity on trip composition and dynamics.
In Figure 18 it is shown that 34 trips are valid according to all boundary conditions. Like MAW, PB also exhibits inconsistencies, or complementarity with the trip dynamics and trip composition boundary conditions, since 29 trips are valid according to PB but not valid according to the RDE boundary conditions.

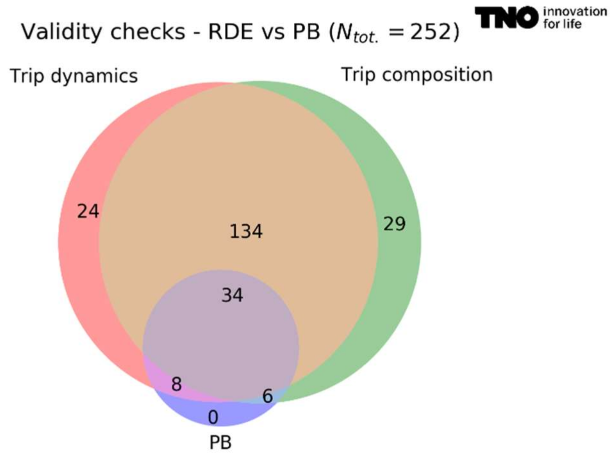


Figure 18 Venn diagram: validity checks: RDE vs PB

If the PB boundary conditions are examined in more detail the coverage of the urban P5 power bin, is the most demanding condition for urban, and P6 for the total trip.

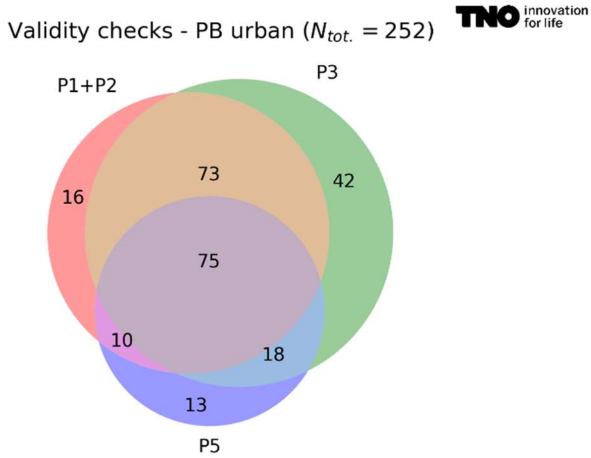


Figure 19 Venn diagram: validity checks: PB urban

It can be concluded that the boundary conditions of PB are contributing significantly to the number of invalid tests, and show complementarity, i.e., seemingly related different aspects, with the other boundary conditions.
Also, urban power bin seems to be a strong boundary condition. However, this boundary condition is open to different interpretations based on the law text, and therefore a more ‘relaxed’ version of PB was implemented resulting in 86 valid tests. Still, this is about half of total RDE valid tests.

8.3 Emission result - compared to raw emissions

In this chapter, the evaluation methods are compared with raw emissions. As was stated in chapter 7.4, the effectiveness of PB is analysed in chapter 10.

Given the fact that only 34 valid test remain, the conclusions on the basis of the subset have limited significance. But overall it can be concluded that the correction of the power binning method on these result is small, and there are little to no systematic effects.

It seems the power binning method is selecting the test data for which a limited correction is needed. The trip normality part of the PB method is therefore an important aspect. The much wider spread in corrections for the invalid data, as shown in Figure 20 and Figure 21, supports this conclusion.

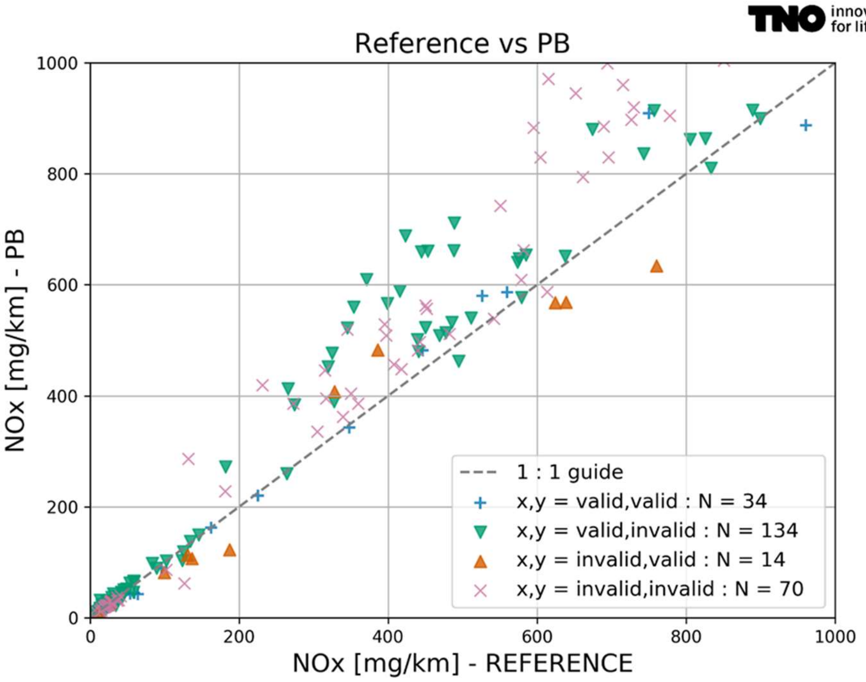


Figure 20 the evaluation method 'PB' compared with raw emissions. Invalid trips are corrected upward somewhat. Valid trips have in general only minor corrections.

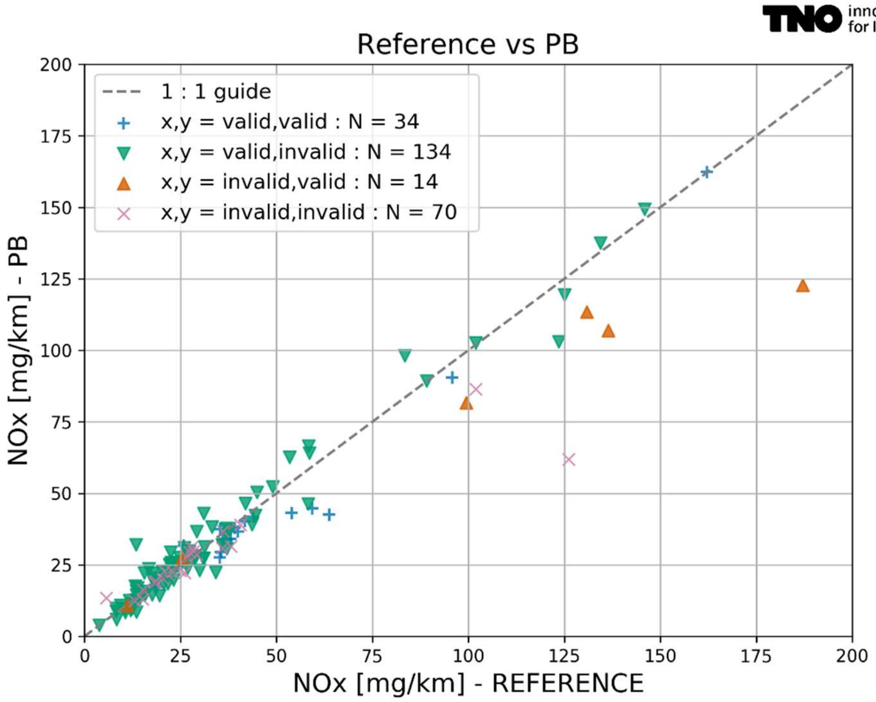


Figure 21 the evaluation method 'PB' compared with raw emissions with a focus on the lower emission results

9 Analysis of NO_x/CO_2 based evaluation methods

NO_x/CO_2 is a method initially designed for plug-in hybrid vehicles, which have electric driving capabilities. ACEA has proposed to apply the method, in an augmented form to all vehicles. Furthermore, they proposed an additional safeguard for trip normality based on somewhat relaxed MAW boundary conditions. This boundary condition must be augmented to allow for the broader range of CO_2 emissions of a plug-in vehicle because of the decoupling of the engine work from the work at the wheels, according to their proposal.

9.1 ACEA proposal for test evaluation

The ACEA proposal has three core elements. First, NO_x/CO_2 as general evaluation method. Second, the fraction of ICE as a method to reduce the emission result proportional with the electric driving distance, which will be discussed in the following subchapter. Third, the validation of test normality according to MAW boundary conditions.

In general the NO_x/CO_2, or more correctly: pollutant/CO_2, does not incorporate boundary conditions. This is shown in Figure 22 and Figure 23 below. It relies solely on the trip composition and trip dynamics boundary conditions to determine if a test is valid. Hence the valid results encompass the largest set of data.

Given the 252 RDE-like tests, 84 are invalid on the basis of trip composition and trip dynamics. These invalid trips have a slightly larger correction, both upward and downward, than the valid trips. Corrections are typically in the order of 10%.

It is very likely that the CO_2 emission in a generic RDE test is higher than the CO_2 emission in a WLTP test. Consequently, the NO_x/CO_2 evaluation method correct the results downward. However, it is possible to drive fuel efficient on a RDE test, while taking care of the RPA driving boundary. Hence, test with a limited correction downward from the raw result are with the realm of RDE testing. The advantage of the ACEA proposal, in terms of the ICE fraction, is the freedom it entails for RDE testing. There are no real limitations of the urban distance on the ICE. Therefore, no special care has to be taken, in terms of battery charging, to avoid invalidated the test from the engine switching on in the last part of the urban test.

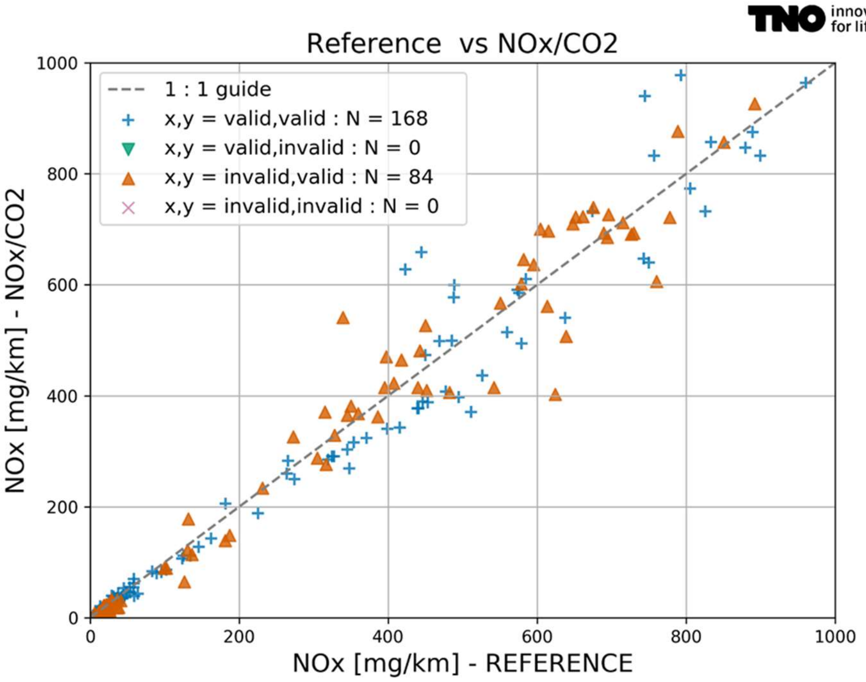


Figure 22 the evaluation method ‘NO_x/CO_2’ compared with raw emissions

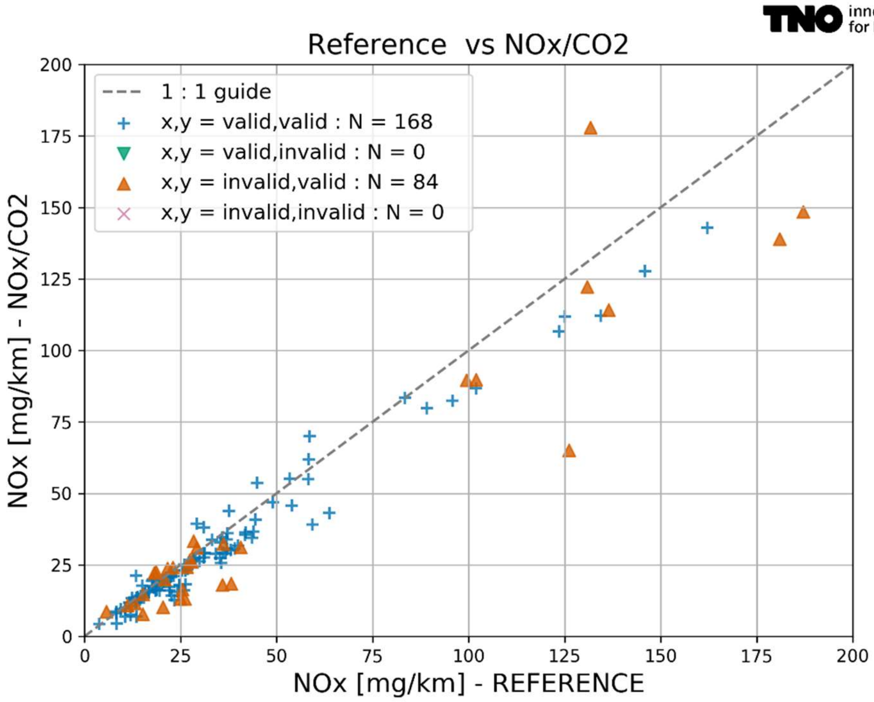


Figure 23 the evaluation method ‘NO_x/CO_2’ compared with raw emissions with a focus on the lower emission results

9.2 Evaluation for hybrids vehicles according to RDE3

Plugin-in hybrid vehicles are evaluated according to RDE3 with the NO_x/CO_2 rate multiplied by the charge sustaining type-approval CO_2 emission, in g/km, on the WLTP. This yields a g/km for NO_x and a #/km for PN, which can be compared to the NTE limit. For the total trip and for the velocities below 60 km/h - the urban part - the total NO_x, PN, and CO_2 are determined. The PN and urban part of the analysis can be found in the Appendix.

Compared to the NO_x/CO_2 the results are reduced by a fraction, smaller than one, when the combustion engine is in operation. 'Engine on' was defined as RPM > 0 and a positive CO_2 emission rate combined, which gave a stable ICE operation and did not fluctuate rapidly. For this vehicle the RPM signal was available.

As one can see in Figure 24 and Figure 25, the evaluated emission results are shifted downwards. This is expected based on the multiplication of the fraction of engine running time.

For the MAW validity checks for PHEV a wider tolerance of -100% to +50% is part of the proposal of ACEA. Initially, 11 PHEV tests, with the same vehicle, were excluded, with the tolerances at -25 to +30%, but with the ACEA proposal of wider tolerances for this vehicle another 8 tests are valid, and only 3 tests remain invalid.

It can be concluded for this chapter that NO_x/CO_2 shows adjustments of about 10%, with no systematic effects. Secondly, the evaluation for PHEV is lowered, as expected.

The NO_xCO_2 method and the ACEA method differ only on the invalid tests and on the treatment of the PHEV for the corrections. The differences are best observed in direct comparison.

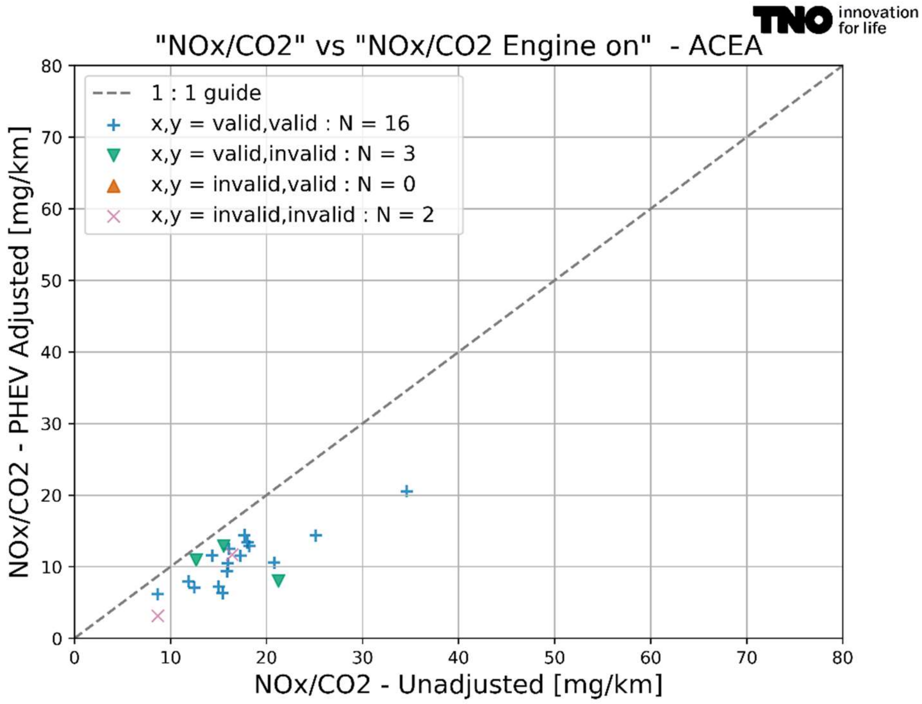


Figure 24 NO_x/CO_2 versus NOx/CO2 with engine on. Since engine on is a fraction of the total distance the correction is always downward.

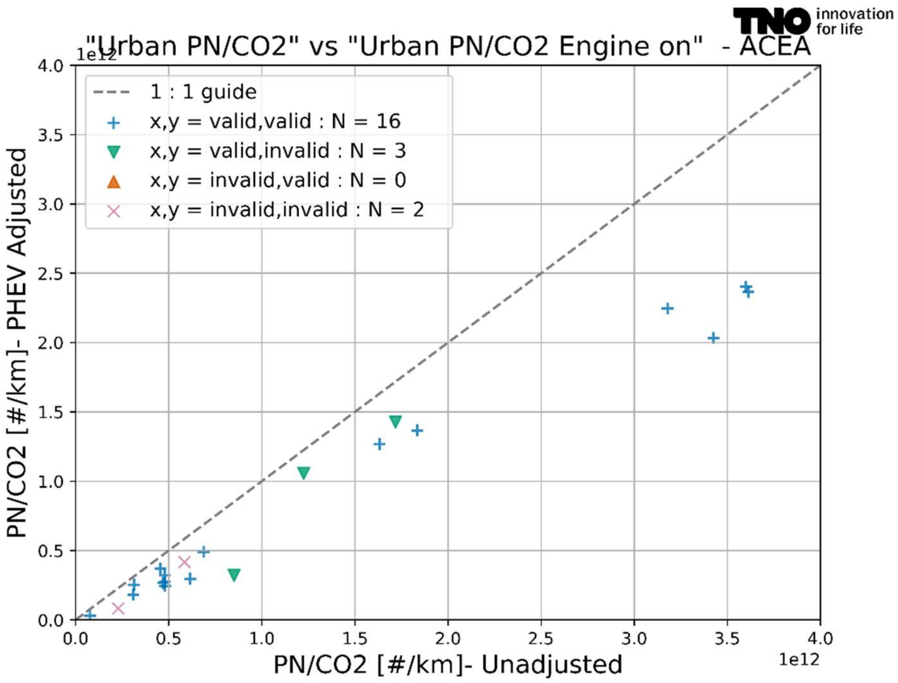


Figure 25 PN/CO_2 versus PN/CO2 with engine on in the urban part

10 Method Effectiveness: systematic effects from multiple tests with the same vehicle

In this chapter, the ‘effectiveness of the evaluation method’ will be defined in this chapter. This is central to the fit-for-purpose of the evaluation methods’ emission result, and whether potential abuse or exploitation can be found. Lastly, it is shown if the evaluation methods are technologically neutral: correct the results for all vehicles in the same manner.

In the plots in the previous chapters, there is large scatter, which may suggest random effects of the evaluation method. If different tests with the same vehicle are compared, it may give a better indication whether systematic effects occur. If both the vehicle and the test conditions are varied at the same time, it is unclear where the variation comes from. Here, the data are examined from a ‘per vehicle’ perspective. When driving with the same vehicle, each RDE test is different, and therefore creates variation. The purpose of the evaluation methods is to correct for deviating test executions with the same vehicle.

Systematic effects may form the basis of exploitation, both in technology to achieve systematic lower emission results after evaluation, but also in test execution to get the largest correction of the result in a desired direction. In order to study systematic effects, data are analysed per vehicle, for the vehicles for which at least 3 tests remain. If these tests show a clear relation, for example, the magnitude of the corrections is proportional to the emissions and the results will fall on the same line. And if the relation between raw results and corrections deviates significantly from the one-to-one relation, e.g., systematically below or above, it will raise concern for exploitation.

It may be so that an evaluation method functions perfectly on a vehicle with fully deterministic emission behaviour, and moreover, the tests differ only in the aspects handled in the evaluation method. In that case the evaluation result may always be the same, although the raw emission vary with the execution of the test. This is a perfectly legitimate situation with a large systematic effect. An unwanted effect would be the case where all data of the same vehicle are corrected downward irrespective of the raw emission results and the execution of the tests. On the other hand, a desired effect would show a limitation of the evaluation result with respect to the raw result, if correction for deviation test executions is needed. Thus, if a vehicle is tested, the variation in emission outcome of the evaluation method would be smaller, regardless of the differences in the test environment, driving behaviour, etc. A quantitative measure to indicate the reduction in variation is defined. This is the ratio of the average variation, or standard deviation, per vehicle in the evaluation result and the variation in the raw result. This ration should ideally be smaller than 1, and show limited variation over the vehicles. The ratio is denoted by σ_y/σ_x.

As displayed in Figure 26 to Figure 29, in a number of evaluation methods there are vehicles which are systematically corrected upward and other systematically corrected downward. In particular PB shows such trends. This conclusion is somewhat weakened by the fact that also invalid tests had to be included to have sufficient data for this analyses.

Given the variation, not all of them decrease variation significantly. For MAW the evaluation method seems to increase the variation in the results for some of the vehicles. These results are opposite to the intended effect.

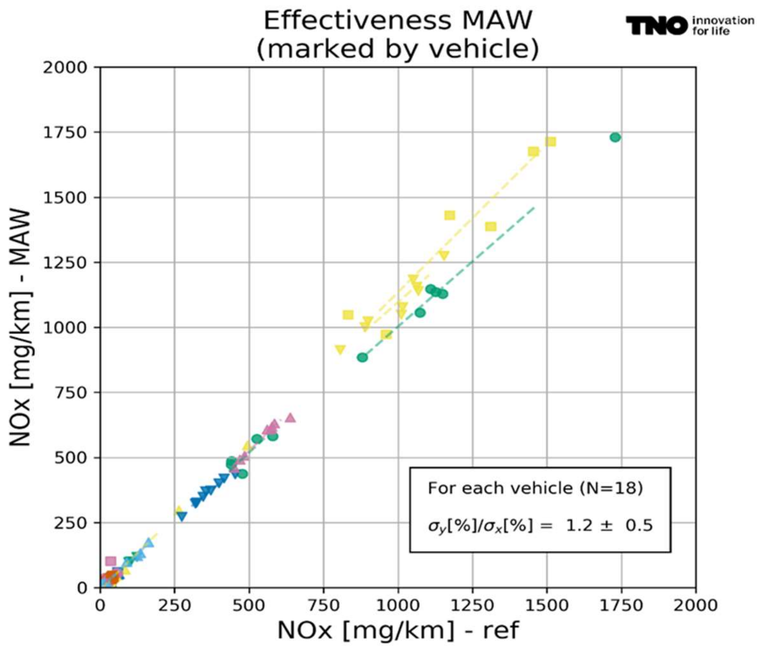


Figure 26 Effectiveness MAW

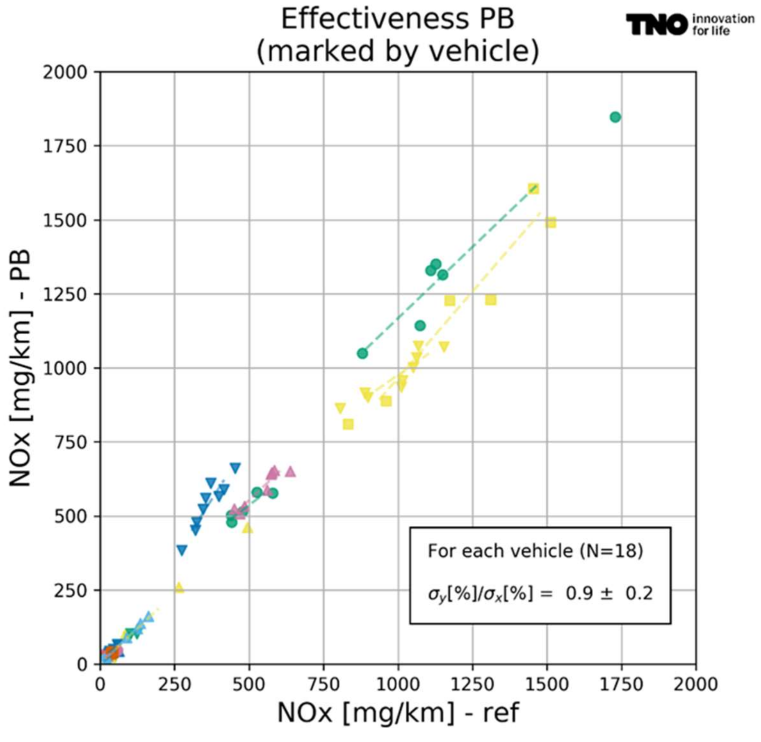


Figure 27 Effectiveness PB

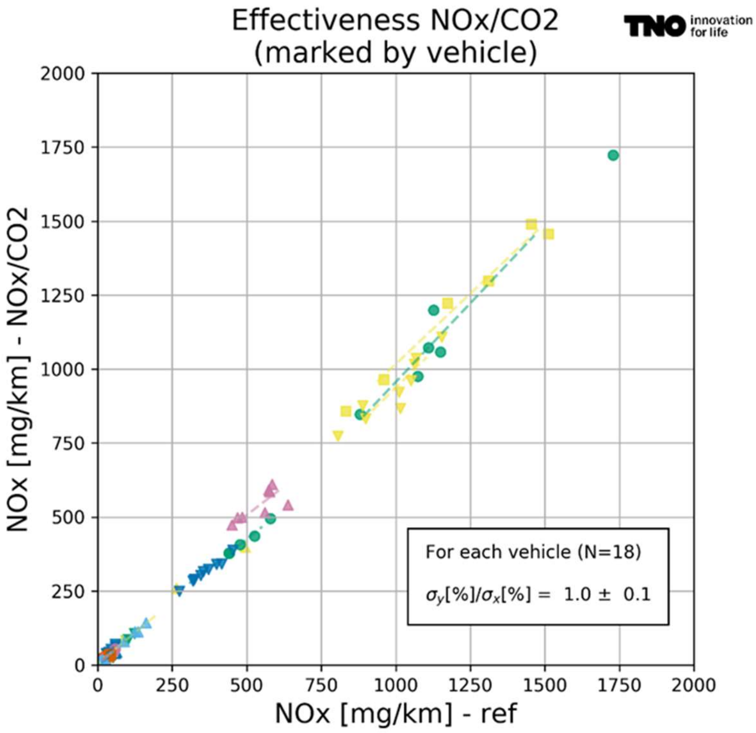


Figure 28 Effectiveness NO_X/CO_2

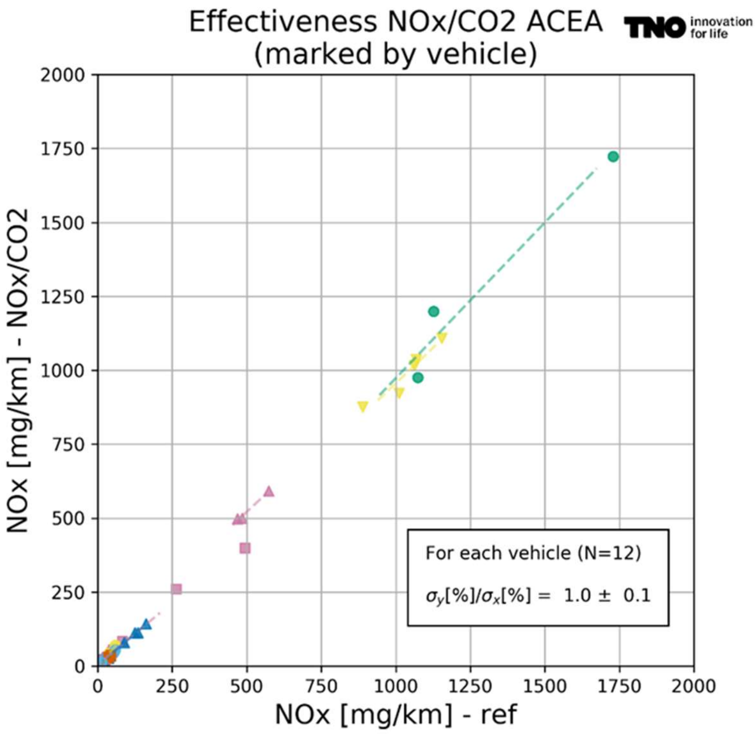


Figure 29 Effectiveness NO_X/CO_2 ACEA

The trend for the total trip and the urban part are not always in line. It seems NO_x/CO_2 mostly reduces variation mainly in the urban part, indicating a wide variety in CO_2 emissions correlated with the pollutant emission. The effect of a 0.8 ± 0.3 for NO_x/CO_2, as shown in Figure 30. For the ACEA proposal this is 0.7 ± 0.3, these graphs can be found in the appendix.

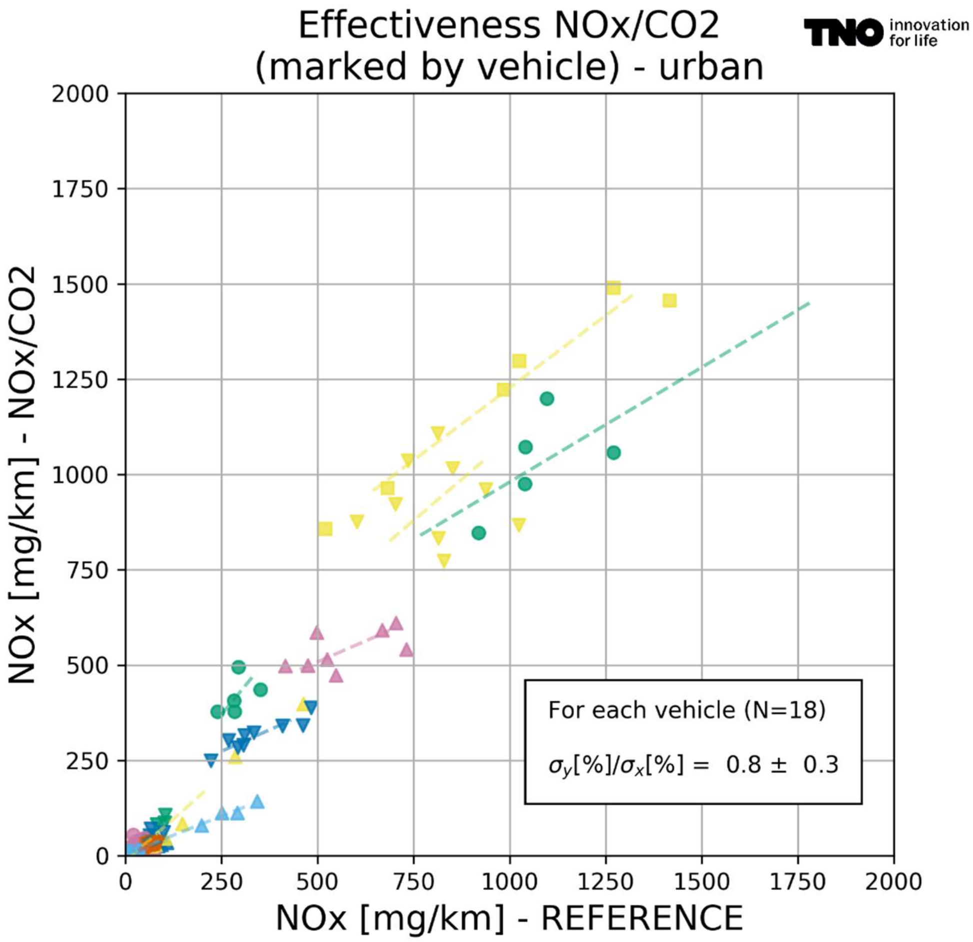


Figure 30 Effectiveness NO_x/CO_2 in the urban part

It may be concluded that none of the evaluation methods are particularly effective. MAW seems to slightly increase in variation. PB shows particular vehicles which seem systematically increased / decreased, where the other evaluation methods did not.

Only if tests were conducted of different routes, with different congestion levels, or different driving styles, it is expected the evaluation methods will correct for the variation. Indirectly, since most evaluation methods rely on the CO_2 emission rate to determine the power demand and the test normality, aspects which will affect the CO_2 emission such as altitude gain, test mass and gear shifting may affect the outcome of the evaluation. The results with the multiple tests with the same vehicle indicate only limited variation in the aspects directly or indirectly covered by the evaluation methods. Dedicated test programs may show more relation between the test variation, the test results and the evaluation results.

11 Comparison and consistency of methods

By looking at data alone a very diffuse picture on the RDE evaluation methods arise. Effects seem random. Few general trends are observed. By looking at the results from different perspectives and angles the scatter in the results generated by the methods over its advantages and systematic effects becomes apparent. The comparisons of different methods is a manner in which the limited effectiveness of the specific purpose is most striking. One would expect different methods correct the same test in a similar manner. Instead the variation is larger than the coherence between the methods.
.
All evaluation methods are supposed to bring down the variation in the results associated with variations in the RDE test executions. Of course, certain effects are not covered, but in the case of a substantial correction downwards, or upwards, in one evaluation method, it is expected that in another method a similar correction is observed. The methods are designed for compensating variation in the test execution.

For that reason it makes sense to compare the corrections of the different methods against each other. Data will appear along the diagonal, if two methods are set against each other, would mean the same correction applies in both cases. This is, in principle, the desired result. This would mean both methods serve the same purpose, albeit in different manners, and likely both methods function according to the general design criteria.

As shown in Figure 31 to Figure 34, in many cases, such results are not observed. The methods pertained to perform the same correction do not seem to do so.
Both MAW and PB give large corrections. This is partly balanced by the fact that the largest corrections are invalid tests. The valid tests show a smaller correction.

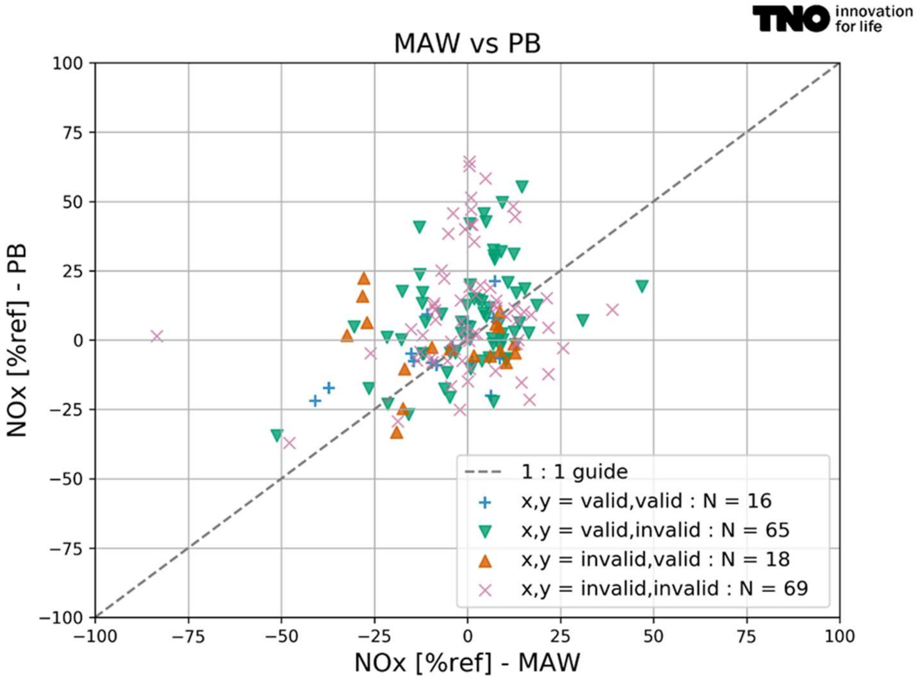


Figure 31 comparison of the corrections of MAW and PB

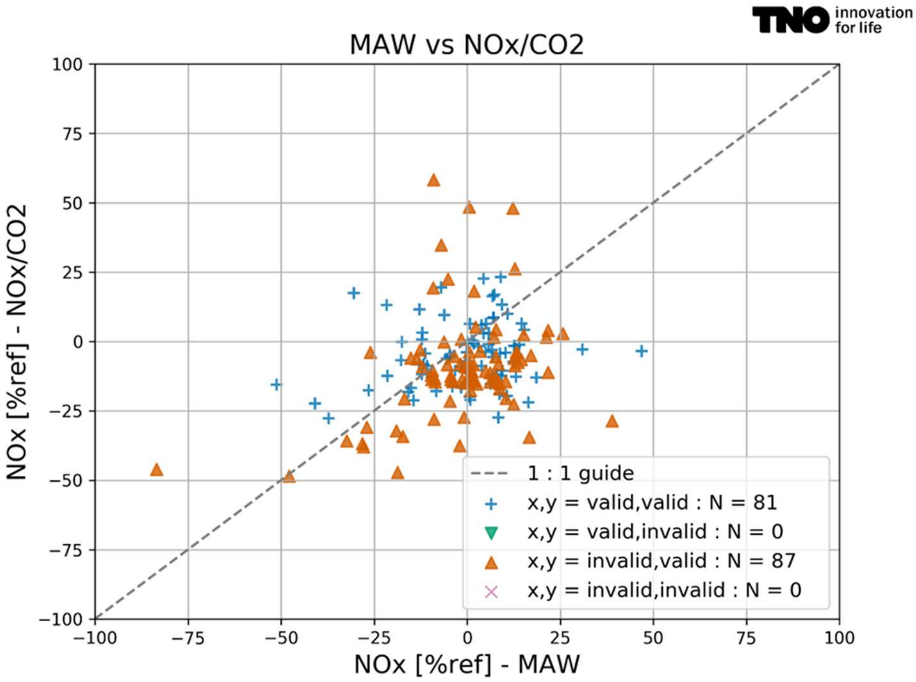


Figure 32 comparison of the corrections of MAW and NO_x/CO_2

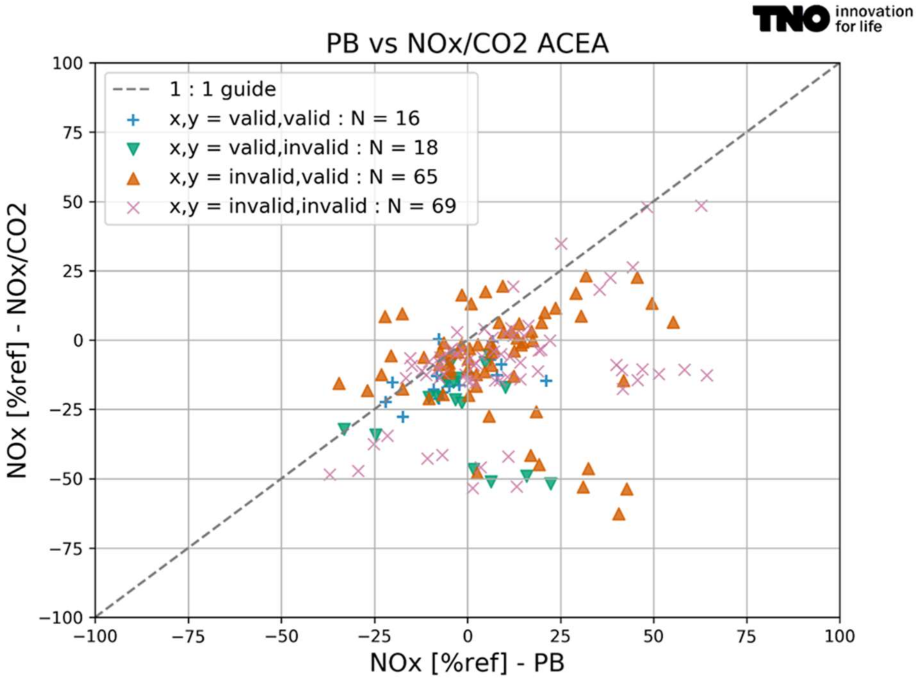


Figure 33 comparison of the corrections of PB and NO_x/CO_2 ACEA

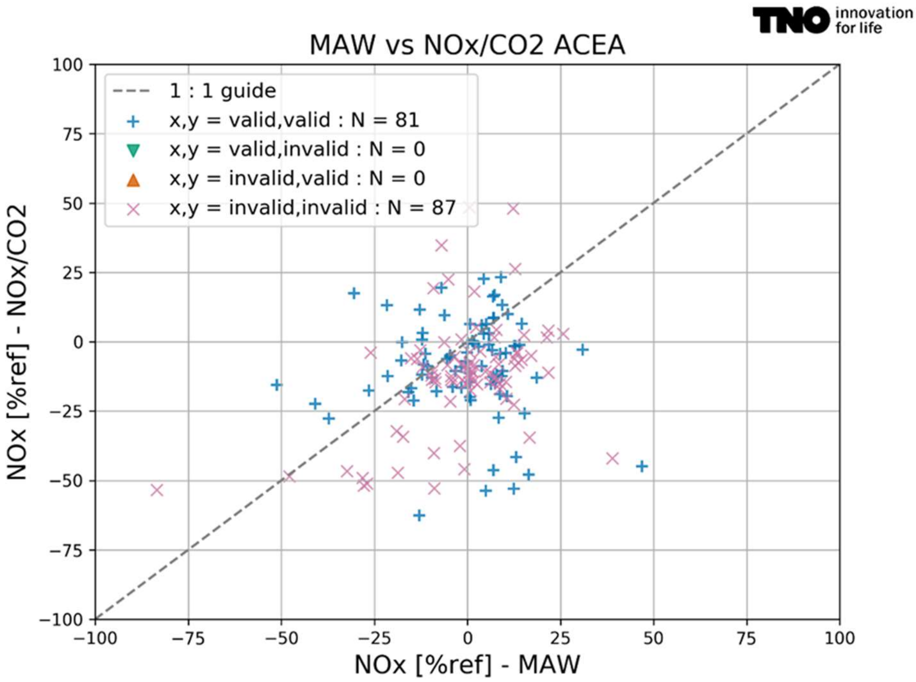


Figure 34 comparison of the corrections of MAW and NO_x/CO_2 ACEA

The graphs show that the methods are not in line which each other. They all have significant corrections of the raw results. The manner in which one evaluation method corrects the results bears little resemblance with the other evaluation method. The intended purpose of the evaluation methods seems to do little to correlate the corrections of the different methods.

12 Robustness: Sensitivity for input variation

For legislative purposes, the evaluation methods should robust and should not give unexpected effects with a large or systematic deviation. Robustness can be summarized as small differences in the input leading to large differences in the output, or systematic differences in the output for different vehicles. Moreover, not only the results should not deviate greatly, but also the classification of valid and invalid should not depend strongly on minor aspects of the method.

Most methods need some additional vehicle information in order to operate. In general that would be WLTP emission values and WLTP test parameters. For NO_x/CO_2 the average CO_2 emission on the WLTP is needed, for MAW the results per phase are needed, and for PB more details of either the WLTP test, or the intermediate derived parameters are needed. In principle, such values are not available, because the CO_2 for a specific vehicle model is determined from the interpolation method in the WLTP with emission data only available for a high and low vehicle models of the same type-approval family. But some input data will exist.

The input data may vary and the outcome of the evaluation method may vary with the input, for the same RDE test data. The variation of the outcome should not be large with a limited variation of the input. For NO_x/CO_2 the variation in the results is simply inversely proportional with the input CO_2 value: a 5% higher WLTP CO_2 [g/km] will lead to a 5% lower NO_x [g/km] result on the same RDE test data. Low temperature and high payload will yield higher CO_2 emissions, correcting RDE results downward.

For other methods it is more complex and the method need to be run with alternative input values, which may come in a variety of combinations. A variation of 10% in WLTP values is considered within the range of expected variations for a specific vehicle. This should not lead to much larger changes than 10% in the results.

Moreover, such variations should not change the number of valid and invalid tests. The latter seems to be the main problem. Evaluation methods show a strong sensitivity of the boundary conditions of the WLTP input parameters in terms of invalids tests.

12.1 PB sensitivity to WLTP input

A typical variation one might expect with the PB input would be in the Veline, i.e., the relation between power at the wheels and CO_2 rate. This line may vary in its slope, leaving the average unchanged. This means changing the offset and the slope in opposite direction. As displayed in Figure 35, this leads to an effect on the result of 10% as well, with a large variation. Moreover, it increases dramatically the number of valid tests with PB, from 34 to 48 tests of all 252 tests. It does raise concerns of the default input values used for the PB results.

Therefore, the effect of the WLTP input data were also investigated on a subset of the dataset. In this subset, all WLTP input data were available. By comparing the outcome of the evaluation method with default WLTP values as input versus the actual WLTP values as input, shows a similar picture, see Figure 36.

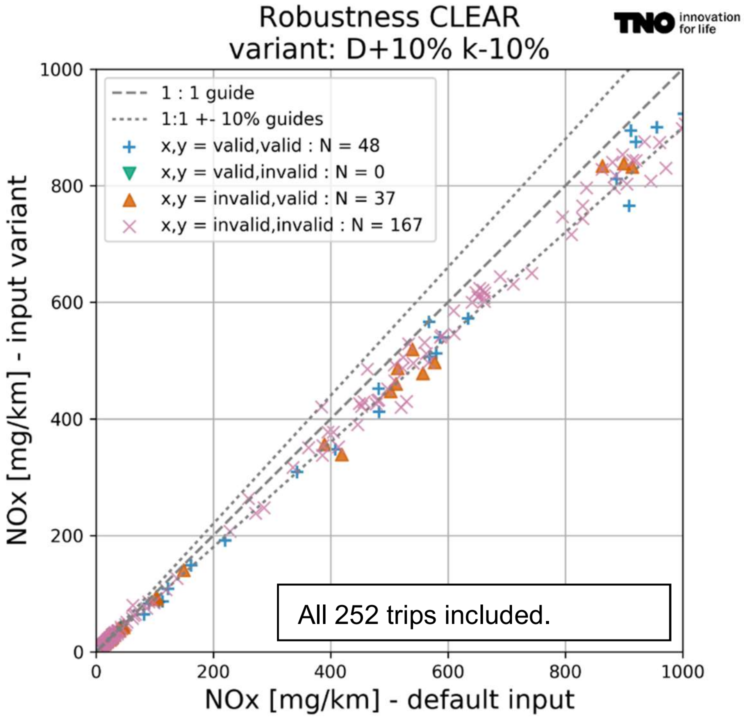


Figure 35 Robustness PB by modified WLTP input data

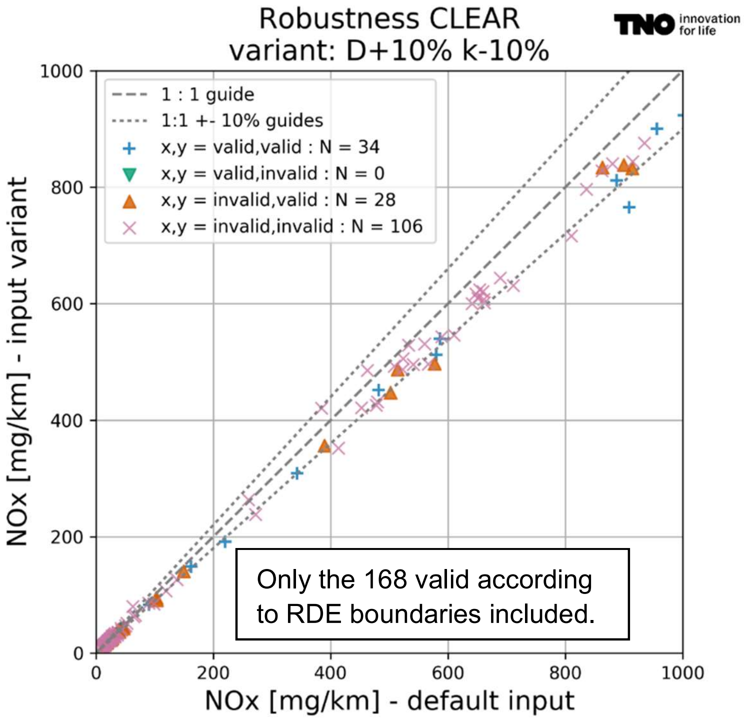


Figure 36 Robustness PB by modified WLTP input data (valid tests only)

As shown in Figure 37, valid input values bring the evaluation result down in most cases, and up in a few. Moreover, many tests remain invalid, but a total of 39 tests flip. Using PB with the correct WLTP input data give more valid tests.

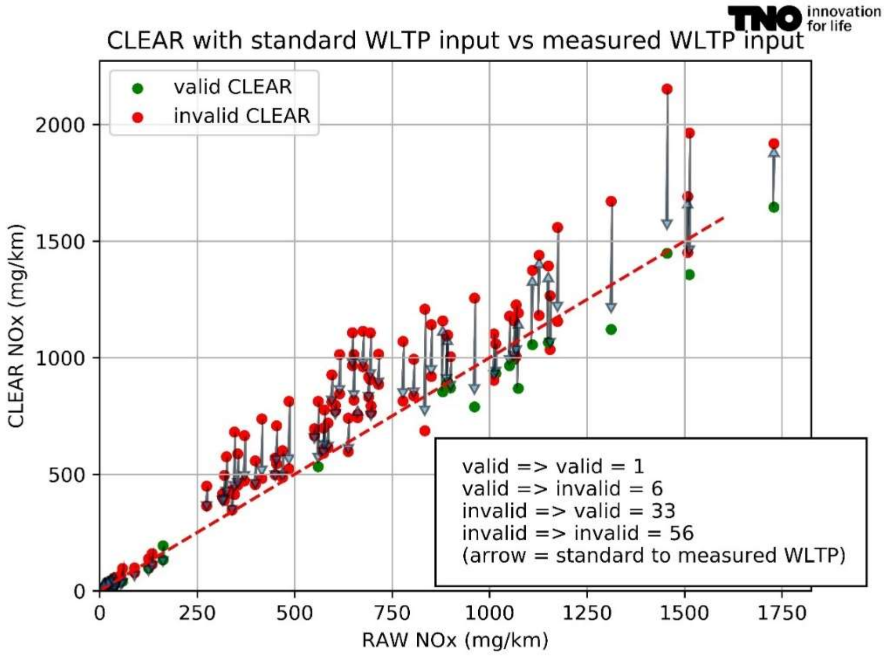


Figure 37 Direction of correction

12.2 MAW sensitivity to WLTP input

For MAW the sensitivity of the emission result with the change of WLTP input values is much weaker. In this case the correct input data were provided for most of the vehicles. Hence the effect of default input does not have to be investigated, like it was for PB. It was already concluded from the limited disqualifications on the basis of the tolerances in the CO_2 band, that the WLTP has a limited influence there. A second aspect where the WLTP input plays a role is the window length. The combined effect from the input variation is limited.
WLTP input dependence is strong for test validity, or test normality. The test validity flips around a lot, in both directions. A number of tests re valid with WLTP + 10%, 127 versus 116 of all trips.
Also, 33 tests flip, from valid to invalid or vice versa. This is also an unwanted feature. This signals a lack of robustness on the MAW boundary conditions. From the study of these boundary conditions in the chapter above, it was already clear that the current data lie close to two boundaries: the fraction of motorway data and the urban tolerance. The change in the input value is likely to tip the data on both conditions: a longer window includes more data in a window bring the average velocity down, invalidating more tests on the basis of a lack of motorway windows.. The urban CO_2 band will shift upward with a higher WLTP CO_2 making more trips valid. Precisely this combination is observed in the data.

It can be concluded that the emission result of MAW is not highly sensitive to input values. However, the check for test normality of MAW is sensitive for boundary conditions: a substantial number of tests 'flips' from valid to invalid and vice versa when input values are changed.

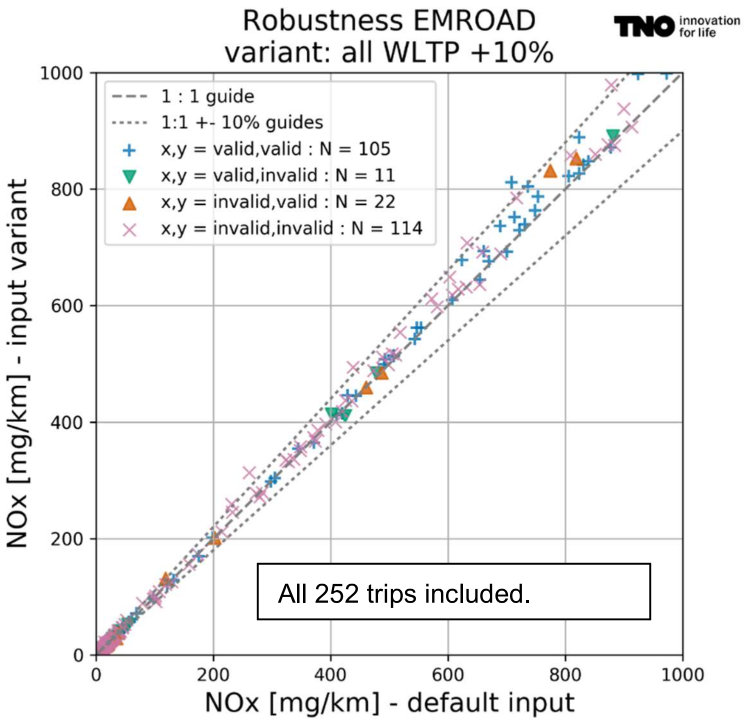


Figure 38 Robustness MAW by modified WLTP input data

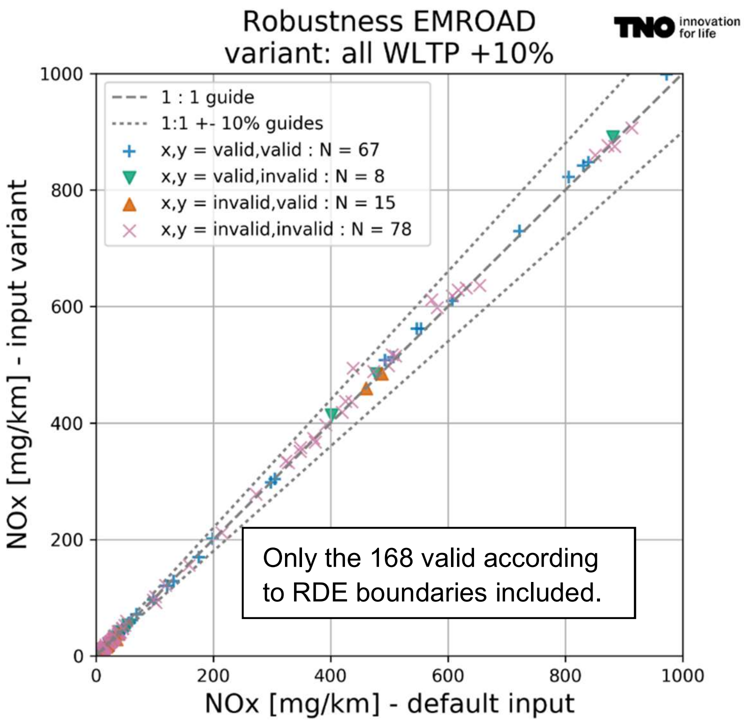


Figure 39 Robustness MAW by modified WLTP input data (valid tests only)

13 Transparency, simplicity and unambiguous understanding

More text and more definitions are a sign of complex procedures. In that respect both MAW and PB, but also the altitude gain procedure are complex. With complexity comes ambiguity and opacity. On purpose no attempt is made in this study to explain the results of the evaluation methods on the basis of the underlying mechanistic principles of the evaluation methods. Such an attempt would be heavily biased by the effort it took to implement the legal text into a validated computer code. The seemingly organic development of the original ideas for the evaluation methods to the final results by the many years of discussions and evaluations did the final result probably no good. The initial room to manoeuvre, to tweak, the evaluation methods to serve specific purposes and satisfy new criteria introduced later allowed the methods to wander into directions no longer fitting the equations. A rewrite of the methods, based on the key surviving elements, should be considered if further developments are planned. A possible development could be to make methods less restrictive, such that they do not invalidate so many tests.

13.1 Unambiguous understanding

An important criteria for each method is also its simplicity and ease of use.
Here, we found that PB and MAW are relatively hard to implement. Where PB and MAW take weeks to implement correctly, NO_x / CO_2 took only three lines of code, and it can also be done in Excel or any other simple mathematical programs.
This leads to less discussion about interpretation of the legislative-text and less implementation differences among institutions. Therefore, it makes it easier for each institution to do their own RDE tests and evaluations, which is beneficial for all parties.

13.2 Transparency

Also, the emission results of the evaluation method are easier to interpret. It is often not clear why a specific trip is evaluated upwards or downwards by PB and MAW because a lot of mathematical proceedings are applied to the emissions, leaving it unclear why certain trips are evaluated upwards or downwards. NO_x/CO_2 leaves less discussion about the evaluation of the emission results, and manufacturers can directly focus on lowering the NO_x emissions to obtain a lower NO_x/CO_2 outcome. Though NO_x/CO_2 does have advantages in this point, on clarity and unambiguity, the most simple and easy to use evaluation method is no evaluation method at all: raw emissions.

13.3 Simplicity - WLTP

Another difference between the evaluation methods is the dependence on the WLTP values. For MAW and NO_x/CO_2, the WLTP values are easy to obtain. However, for PB, a number of extra values are necessary which are not always easy to obtain, such as f0, f1, f2. This makes it uneasy to use for all parties unless all data are provided in a transparent way. This concerns not only the WLTP test data, but also the data of each particular vehicle sold.

14 Discussion

In this review the evaluation methods are taken “as is”, and the RDE test data were the material to work with. The discussion is limited to the experiences with implementing and running the methods, and the analyses of the outcome. The motivation for, and underlying principles of, the evaluation methods have little consequences for these experiences. Simply said, this report is a “consumers’ review” of the evaluation tools: How does it work? What does it do? What are the results?

The current report confirms earlier reporting in the RDE-LDV group on the experiences with the evaluation methods. The term “randomization” of the results has been coined before. This is not confirmed by current study, but a large scatter and little systematic effects, together with a large fraction of invalidation on unclear grounds is central to the result of this study. Randomization, i.e., results as by a crapshoot, would mean the methods are not prone to exploitation, or abuse. There are some indications some systematic effects occur, which can be studied to achieve systematic corrections upward or downward of the measured result by either designing the technology or the test to that specific purpose, unrelated to the actual purpose of RDE legislation.

14.1 Normal driving and vehicle behaviour according to the WLTP input

The use of WLTP type-approval values as input to assess vehicle emission performance in RDE can be too restrictive, as the RDE test can greatly vary in many aspects, while the WLTP data are obtained in the laboratory under a strict protocol. For example, in a cold start and the low phase, i.e., urban part, of the WLTP, additional emission may occur which affect the urban evaluation which may not have a cold start or, may have an enhanced cold start effect in low ambient temperatures. Moreover, the test mass of the WLTP, and the refence mass of the vehicle may be far off the range of test masses allowed in the RDE test. Hence, putting too much focus on WLTP input values may be unwarranted, because these values are related to testing and conditions which may deviate substantially from the conditions in the RDE test. On the other hand, in terms of emissions the WLTP limit is the standard, also for the development of RDE legislation.

14.2 Is driving behaviour strongly correlated to emissions?

Certain effects of the evaluation methods rely on the underlying assumption that somehow the emissions are directly linked to driving behaviour in the way that the evaluation methods are linked to driving behaviour. In that case the emissions change proportionally with the change in driving behaviour. However, in practice high emissions may occur for a specific event, which may not even be related to driving behaviour. Consequently, the emissions may change stepwise with the number of these types of events that occur. Hence, the variation in emission may seem rather random when they are evaluated according to the variation in driving behaviour.

For CO_2, it is a fairly safe assumption that driving behaviour is correlated to CO_2 emissions, however, for NO_x emissions this might not be the case. This could be the reason why emissions results in the NO_x/CO_2 method exhibit scatter when compared to raw emissions. The pollutant emissions for a petrol vehicle with a warm three-way catalyst are probably the foreland of NO_x emissions of a RDE-compliant diesel vehicle. The emissions of a petrol vehicle are incidental and often related to very specific problems in the emission control strategy, under real-world driving conditions.

One could question if driving behaviour is strongly correlated to NO_x emissions in the way these evaluation methods assume. The normal driving conditions should be part of RDE such that incidences of high emissions are assessed during RDE tests. In this regard, no evaluation method at all seems to have an advantage. Using proper RDE boundaries to define normal driving.

14.2.1 *Raw emissions disadvantages*

The use of raw emissions and no evaluation method does have some disadvantages. Raw test results can deviate highly from the average emissions, for example, by repetitiously mimicking a particular situation which proved difficult for a particular control strategy, thus leading to an unfair result in RDE tests. The RDE boundary conditions, on driving behaviour and trip composition, may not invalidate such results.

14.3 Maximum corrections of evaluation method

Another way of assessing the evaluation methods is by looking at the maximum deviation of the emission results from raw emissions. These deviations are calculated for trips that are valid on all accounts. This is shown in Table 9.

Table 9 Corrections and valid trips for each method

METHOD	MAX CORRECTION (%) (COMPLETELY VALID)	MIN CORRECTION (%) (COMPLETELY VALID)	MAX CORRECTION (%) (ALL)	MIN CORRECTION (%) (ALL)	NUMBER OF COMPLETELY VALID TRIPS (FROM ALL 252 TRIPS)
OPTION 1: ONLY TRIP COMPOSITION	0%	0%	0%	0%	217
REFERENCE (RAW WITH RDE BOUNDARY CONDITIONS)	0%	0%	0%	0%	168
RAW WITH MAW BOUNDARY CONDITIONS (T&E)	0%	0%	0%	0%	75
MAW (EMROAD)	47%	-51%	87%	-51%	75
LOOPED MAW (WINDOWS GOING ROUND TO THE BEGINNING)	47%	-51%	112%	-51%	61
PB (CLEAR)	25%	-34%	139%	-51%	34
NO_X/CO_2	58%	-49%	59%	-52%	168
NO_X/CO_2 * ICE WITH MAW BOUNDARY (ACEA)	48%	-63%	59%	-52%	84

However, this does give a slightly distorted view. On the one hand, one would like to look at the maximum correction of the evaluation method applied to all trips. This shows large deviations in PB, MAW and Looped MAW. On the other hand, one would like to see how large the deviations are on only valid trips. When considered in such a way, PB shows the smallest deviations. Therefore, deviations should be considered in light of the number of valid trips according to this method. PB shows a relatively small deviation, but this is on a very small subset of the dataset of 34 trips. Trips that deviate from raw are apparently considered as invalid according to PB boundary conditions, which is not necessarily a desired property.

In summary, the evaluation methods lead to a large variation in corrections both up and down. The trip validity checks within the evaluation methods seem to be the most restrictive part of the different evaluation methods. The magnitude of the corrections in the different methods should not be seen independently from the number of valid trips it applies to.

14.4 Complexities with valid urban driving conditions

Already the range of average velocity of 15 km/h to 40 km/h and stopping time of 6% to 30% of the urban part are indications of the large range of valid driving conditions in the urban part. Consequently, the CO_2 emissions may vary greatly in normal urban driving and urban boundary conditions can be restrictive. Indeed, both MAW and PB lead to a large number of invalid tests on the basis of the urban boundary conditions. The strict protocol for the WLTP test, and thereby generated input for the evaluation methods, do not seem to cover the bandwidth of collected RDE test data. Urban driving will come in a large variation, related to cold start, braking events, gear shifting, and velocity distribution.

From the trip boundary conditions, allowing more than factor two in the average velocity and stopping time, it is to be expected that the CO_2 emission can also vary about factor two in different trip composition and trip dynamics executions of RDE tests. For PHEVs this factor is much larger. The implicit bandwidth allowed in both MAW and PB is smaller than this factor of two. Hence, the invalidity due the deviating urban test results in the evaluation methods was already to be expected.

Hence, both MAW and PB lead to restrictive boundary conditions on the urban part of the RDE test. In the case of hybrid vehicles, for which little test data are available, it seems that MAW boundary conditions, with the wider tolerances, as proposed by ACEA, remain restrictive, with a large fraction of invalid tests. Part of the problem of the use of the evaluation methods as boundary condition on urban driving is their dependence on the WLTP input data. It may lead to the fact that with an empty vehicle and a hot start a different driving style has to be adopted to arrive at a valid RDE test, than with a heavy payload and a cold start. On the other hand, it is not expected that normal driving styles vary greatly between these two cases, and, instead, driving behaviour is restricted beyond the bounds of normality.

One could argue whether an appropriate coverage of different urban driving is enforced by the evaluation methods. However, neither the trip composition boundary conditions nor the MAW evaluation method seem to be designed to do so. For PB evaluation method it could be argued the overall coverage is part of the design. But this argument may go against the spirit of RDE legislation which is intended to cover “all” normal driving, rather than “average” normal driving.

15 Concluding remarks

The main results, presented in the table below, reflect the main conclusions of the study. The different evaluations methods give corrections of the results, but also they invalidate tests on the basis of internal boundary conditions, which check for test normality and ensure stability of the method. Applying evaluation methods may lead to large corrections, both upward and downward. The correction of one trip may not be consistent between the different evaluation methods. The number of invalid tests according to the evaluation method is significant. The possible need for these boundary conditions in the evaluation method can be deduced from the fact that for valid tests the corrections are smaller than for invalid tests.

METHOD	MAX CORRECTION (%) (COMPLETELY VALID)	MIN CORRECTION (%) (COMPLETELY VALID)	MAX CORRECTION (%) (ALL)	MIN CORRECTION (%) (ALL)	NUMBER OF COMPLETELY VALID TRIPS (FROM ALL 252 TRIPS)
OPTION 1: ONLY TRIP COMPOSITION	0%	0%	0%	0%	217
REFERENCE (RAW WITH RDE BOUNDARY CONDISTIONS)	0%	0%	0%	0%	168
RAW WITH MAW BOUNDARY CONDITIONS (T&E)	0%	0%	0%	0%	75
MAW (EMROAD)	47%	-51%	87%	-51%	75
LOOPED MAW (WINDOWS GOING ROUND TO THE BEGINNING)	47%	-51%	112%	-51%	61
PB (CLEAR)	25%	-34%	139%	-51%	34
NO_X/CO_2	58%	-49%	59%	-52%	168
NO_X/CO_2 * ICE WITH MAW BOUNDARY (ACEA)	48%	-63%	59%	-52%	84

The problems with test validity and the nature of the disqualification of tests in MAW and PB was not a priori clear. Special code was developed to track all the separate conditions stipulated in the legal text. In this respect only the study interpreted the method and added some new definitions. Since there are many conditions, the union of passes on all these conditions is small. The cause is probably twofold: Firstly, it is difficult to translate the conditions in actual test instructions or protocol, to ensure a valid test. The methods PB and MAW are not transparent in this respect. Secondly, the conditions are in many cases complementary; they do not reflect the same general principles of normal driving. Many complementary conditions decrease the number of valid tests.

Another issue with the invalidity of many of the tests lies in the sensitivity of the evaluation methods PB and MAW for the WLTP-associated input data. Tests may change around from valid to invalid, and vice versa, for a 10% change in input data like WLTP CO_2 emission. It is unlikely the RDE test lies in such a narrow CO_2 band, therefore a strong dependence on such reference values reflects a strong sensitivity for these test parameters.

One of the modifications to the current evaluation methods proposed by T&E and ACEA is to drop the MAW correction and just keep the normality check of MAW and PB. Since the boundary conditions of MAW and PB invalidates so many of the tests, it is probably not proper to retain especially this aspect alone. Without major adaption, it is very likely that many RDE tests, valid on trip composition and driving behaviour are disqualified on the MAW and PB test normality. One important aspect of test normality is the share of motorway windows. Very likely this share is influenced by a moment of congestion, or velocity drop, on the motorway, driving the window-average velocity down. In normal traffic this may occur. Hence the RDE test operators may start avoiding normal traffic, to avoid the risk of many invalid tests and a high test burden.

The NO_x/CO_2 method for hybrids was extended to include conventional vehicles. The procedure is simple and it seems natural. More CO_2 is related to more severe testing in many ways, such as payload, uphill driving, battery charging, etc.. However, on the basis of the current data the weighing of pollutant with CO_2, as in NO_x/CO_2, performs only marginally better than PB and MAW. In particular, on the urban part, the NO_x/CO_2 method brings some benefit on the review criteria set out in this study. Specifically, it seems to correct for the variation in the emissions with the tests a little. Given the fact that the large test variation in the urban part is the main problem observed with the evaluation methods, this benefit is relevant. The deviation from the WLTP reference is here the largest, leading also to many invalid tests.

In the case of a large part electric driving, the cold start can be disproportionate in the NO_x/CO_2 method, therefore a minimum distance of 12 km ICE driving was introduced in RDE3. The ACEA proposal, of weighing results with the fraction of ICE driving, avoids this issue of a minimal distance of 12 km by taking effectively the total distance driven as reference.

The raw emissions is the simplest approach, and it is most directly related to the purpose of RDE legislation. Retaining the trip composition and driving dynamics boundaries reduces the number of tests from 252 to 168, this reduction is minor compared to the other conditions which invalidates the test. However, it is very likely the provided test data did not seek out the corners of the RDE test variations. Very likely, most stakeholders were getting acquainted with the RDE procedure, and found it already hard to drive valid RDE tests, given the fact so many tests are invalid. Hence, if most of these tests are initial attempts “to drive within the RDE lines”, it is very likely in the future RDE tests could be closer to the important boundaries. Hence, dropping the evaluation methods altogether on the basis of the current data may be premature.

The experiences with the current RDE legislation may be a reason to simplify RDE legislation. Current study gives ground to do so, for a number of reasons. Firstly, large scatter is observed, with both outliers up and down, entailing unwarranted risks for all parties. Secondly, the current evaluation methods make it hard to execute a valid test. The reasons for RDE test invalidation are often obscured by the complexity of MAW and PB, and the sensitive relation with WLTP input values.

Thirdly, some systematic effects are observed, but not fully understood. Systematic effects can be exploited by OEMs to lower the final result, but also by independent, yet accredited, testers, who seek to expose the worst emission performance within the RDE boundaries, can exploit systematic upwards corrections..

Looped MAW and the raw emissions with MAW boundary conditions (T&E proposal) are two alternative methods which do not retain the best, but rather the worst of the MAW evaluation method. Moreover, also the more relaxed MAW boundary conditions in the ACEA proposal still retain most of this problem of invalidating tests, as observed with the T&E proposal.

The raw results and the NO_x/CO_2 with ICE distance fraction, without additional MAW boundary conditions, do not pose the problems with invalid tests, as above. This would be preferred. The two approaches have their own advantages and disadvantages, but not according to the aspects studied here. Raw results are the measured emissions, and therefore most simple and most direct. Current data, with current vehicle technologies and current test regimes, show little advantages of NO_x/CO_2 over raw emissions. However, with the changing technologies and other RDE testing, NO_x/CO_2 may show some benefit. Given the fact that real-world CO_2 is in many cases higher than the type-approval CO_2 a downward correction is to be expected in most cases. But on the other hand, seeking out corners of the RDE regime with high pollutant emissions are likely to be associated with higher CO_2 emissions. Consequently NO_x/CO_2 may be the right trade-off between the risk for the OEM with the freedom of the independent tester.

16 Signature

The Hague, 25 August 2017

Sam van Goethem
Project leader

Norbert Ligterink
Author

A Remarks on the legislative text

This appendix contains the review of the legislative text of "*Commission Regulation (EU) 2016/427 of 10 March 2016 amending Regulation (EC) No 692/2008 as regards emissions from light passenger and commercial vehicles (Euro 6)*" with amendment *2016/646 of 20 April 2016* and *2017/1154 of 7 June 2017*.

The legal text reviewed are the RDE boundaries, which is to be found in Appendix 7a and 7b, and the legislative texts of the evaluation methods MAW and PB, which are Appendix 5 and 6.

RDE boundaries

MAW – Appendix 5

v < 1km/h
In A5.3.1 it is stated that

"*The following data shall not be considered for the calculation of the CO_2 mass, the emissions and the distance of the averaging windows:*

- *(…)*
- *vehicle ground speed < 1km/h*"

This can be interpreted as "*data points where the vehicle ground speed is smaller than 1 km/h should be adjusted to 0 km/h.*", to allow for an error in the vehicle ground speed. However, it can also interpreted as "*data points where the vehicle ground speed is smaller than 1 km/h should not be considered (for emissions, averaging windows, etc.*", which seems more likely from the law text, but also seems strange since no emissions during vehicle stops would then be used for analysis.
The newest version of EMROAD (EMROAD_5_95B2 used as reference in this report) seems to be following the later.
We have used the interpretation above. We have considered the second interpretation, and have seen that the effect is usually negligible. However, there could be a large effect in some cases, especially for cold-start and idling emissions.

Primare tolerance 1
As was stated in Chapter 2, there is some confusion about primary tolerance 1.

In the legal text, A.5.5.3 (Verification of test normality), the following is written:

"If the specified minimum requirement of 50% is not met, the upper positive tolerance tol_1 may be increased by steps of 1 % until the 50 % normal windows target is reached. When using this mechanism, tol_1 shall never exceed 30%."

It became clear to us that this legal text probably should be interpreted as the following: The primary, or lower, tolerance has a positive and a negative bound, which are both 25% by default. However, if a trip contains windows which fall outside of the primary tolerance of the CO_2 characteristics curve for more than 50%, then the *upper bound* of the primary tolerance shall be exceeded to 30% with steps of 1 %, and the *lower bound* of the primary tolerance shall be kept constant at 25%.

However, in all formulas following A.5.5.3, only one variable called tol_1 is used, without specifying if the positive lower bound or negative lower bound is meant. The positive bound could be 29%, for instance, and the negative bound could be 25%, resulting in incorrect formula's.

Since tol_1 seems to consist of a bound above the mean and a bound below and this leads to mathematical incorrectness and could lead to misinterpretations, it is recommended to update the law text and split the name of tol_1 to $tol_{1,lower}$ or $tol_{1,upper}$, and to discontinue using the variable name tol_1 in all instances.

Cold start weighting
In the amending text of 2017 the following is written in Annex A.5.6.1:
The following paragraph is added: "For all averaging windows including cold start data points, as defined in point 4 of Appendix 4, the weighting function is set to 1."
This is interpreted as follows: all moving averaging windows containing at least one cold start data point, should be given a weight of 1. The first 180 seconds after cold start are considered cold start data points in the legislative text. It is not unusual for a moving average window to have a length of 500 seconds. Therefore, the outcome of the test emissions could be influenced through the driving behaviour in the first 180 to roughly 700 seconds, since these moving averaging windows will not be weighted.
In this study, we were not able to analyse the severity of this effect. It is recommended to review of this line or a study to determine the severity of this effect.

Typing Errors
A5.6.1

$k = u, r,m$
Mathematical error: k should be on a new line.
Also: K_{22} should be k_{22}.

A5.7.1

M_{CO}
Error: Should be M_{CO2}.

Amending text of 2017: A5.5.3
'When testing a NOVC-HEV and only if the specified minimum requirement of 50 % is not met, the upper positive tolerance tol_1 may be increased by steps of 1 percentage point until the 50 % of normal windows target is reached. When using this approach, tol_1 shall never exceed 50 %.'
Firstly, if tol_1 is increased to 50%, this will lead to numerical singularities since tol_2 - tol_1 will be equal to 0, as used in the denominator. 49% would be a better option. Also, as was noted previously, it is recommended to change the naming convention of tol_1 to a $tol_{1,lower}$ or $tol_{1,upper}$ convention.

PB – Appendix 6

Up to class No.
As was stated in Chapter 2, there is some confusion regarding the check of power class normality.

In the updated law text of 2017, it is written in A.6.3.6, '*Check of power class coverage and of normality of power distribution',* that: (bold face font added)

"A minimum coverage of 5 counts is demanded for the total trip in each wheel power class ***up to class No. 6*** *or up to the class containing 90% of the rated power whatever gives the lower class number. If the counts in a wheel power class* ***above number 6*** *are less than 5, the average class emission value ($m_{gas,3s,k}$) and the average class velocity ($v_{3s,k}$) shall be set to zero."*
and
"A minimum coverage of 5 counts is demanded for the urban part of the trip in each wheel power class ***up to class No. 5*** *or up to the class containing 90% of the rated power whatever gives the lower class number. If the counts in a wheel power class* ***above number 5*** *are less than 5, the average class emission value ($m_{gas,3s,k}$) and the average class velocity ($v_{3s,k}$) shall be set to zero.".*

However, it is unclear to us if power class 6 and urban power class 5 should be included or not. For instance, since it is stated that the emissions for urban powerclasses 'above number 5' should be neglected, one could argue that all others *including* powerclass 5 should not be considered for this text. On the other hand, since it is stated that each powerclass 'up to class No. 5' should contain at least 5 windows, one could argue that powerclass 5 *should* be considered for this text.
In our default implementation, powerclass 6 and urban powerclass 5 are included in the requirement of at least 5 windows to be valid. This leads to a higher number of invalid trips.

Additionally, another boundary condition of PB was interesting. This was the urban powerclass 1+2 upper bound of 60%, *i.e.*, PB requires that less than 60% of the urban trip windows fall under urban powerclass 1 or 2. This restriction also leads to a large number of invalid PB trips.

The impact of changing these requirements is investigated through an implementation of a more relaxed version of PB boundary conditions. Here powerclass 6 and urban powerclass 5 did not have to contain at least 5 windows, and changed urban powerclass 1+2 upper bound to 65% (up from 60%). This lead to 82 PB valid trips and 86 PB invalid trips, from the database of 168 valid RDE trips. This is considerably more than the 34 valid PB trips that are identified in our default implementation. Still, more than half of the valid RDE trips are considered invalid according to PB boundary conditions.

Power at the wheel
In A6.4, *'Assessment of the wheel power from the instantaneous CO_2 mass flow'*, it is explained how the power at the wheel can be calculated from the CO_2 mass flow.

However, another formula is given with the same variable naming convention to calculate the power at the wheel from acceleration. This could lead to confusion. We recommend to update this text to more clearly differentiate between formulas used for calculation of the WLTP power phases and the formulas used for calculation of power at the wheel from CO_2 mass flow.

Typing Errors
A6.4

$P_{w,j}$ shoud be $P_{w,i}$.

Differences CLEAR and MAW tools
The newest CLEAR tool, version 2.0 which is used as a reference in this report, uses slightly different values than is written in the law text. For instance, v_{ref} and a_{ref} are defined as 66 km/h and 0.44 m/s² in the tool, but are defined as 70 km/h and 0.45 m/s² in the law text. Also, boundaries differ slightly.

B Additional analyses

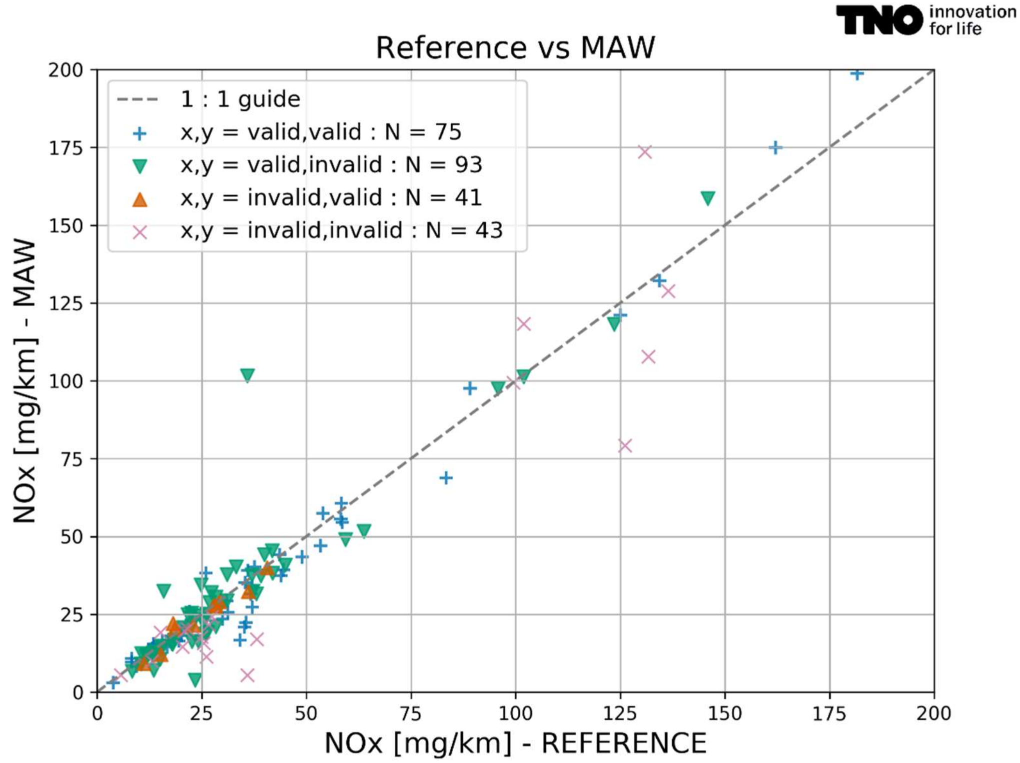
In this part, additional analyses of the evaluation methods are presented. This material is included for completeness. It can serve to address new questions, and as additional evidence for the findings in the report.

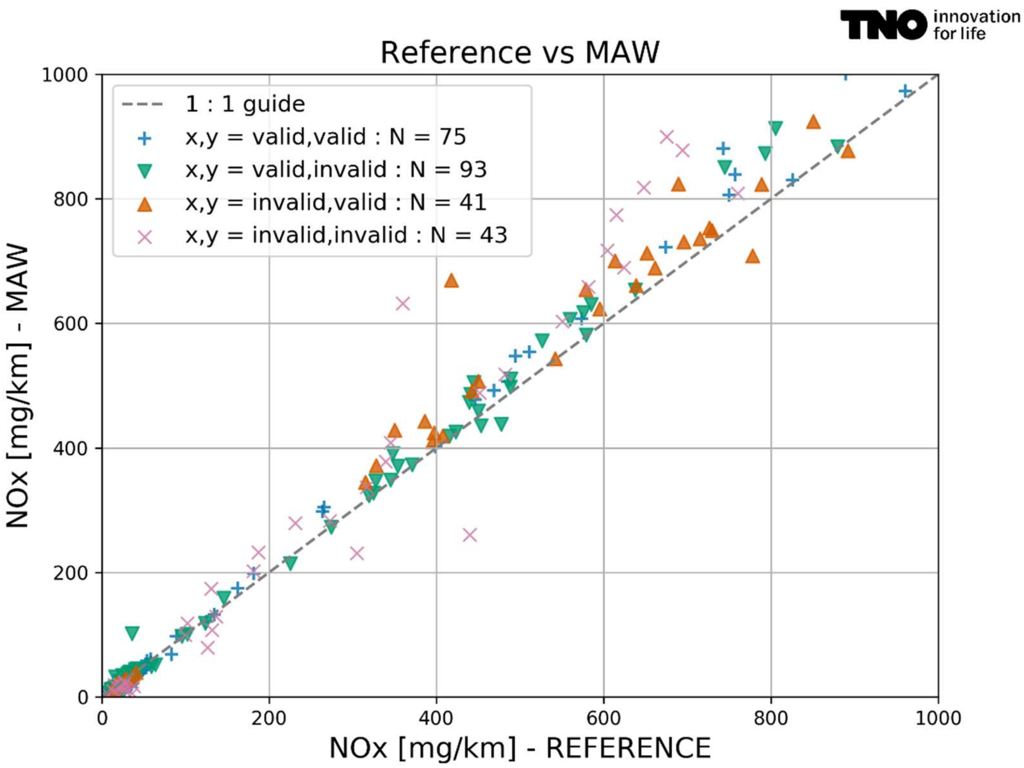
The additional comparison with raw analysis will be presented first, followed by the additional analysis on the robustness will be shown, following with the effectiveness, and lastly the analysis of the relaxed version of PB is included, based on the software and the alternative interpretation of the legal text.

B.1 Comparison with RAW emission values

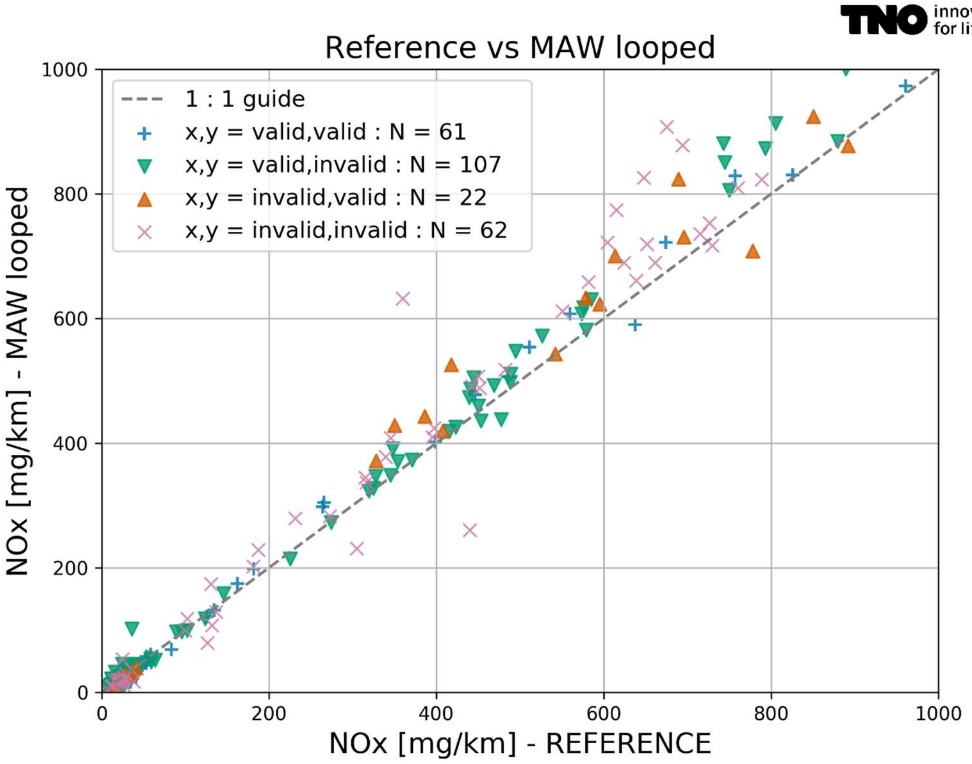
Here, the comparison of the evaluation methods' output with RAW emissions values are presented, with RDE validity. This is the reference case.
A large number of pictures are shown. MAW is firstly reported, containing the normal as well as looped version, and the urban analysis. Next, PB analyses are shown, containing urban as well. Lastly, the NOx/CO2 methods are shown, containing the ACEA and urban analyses. Multiple pictures are shown with different axes to provide more insight.

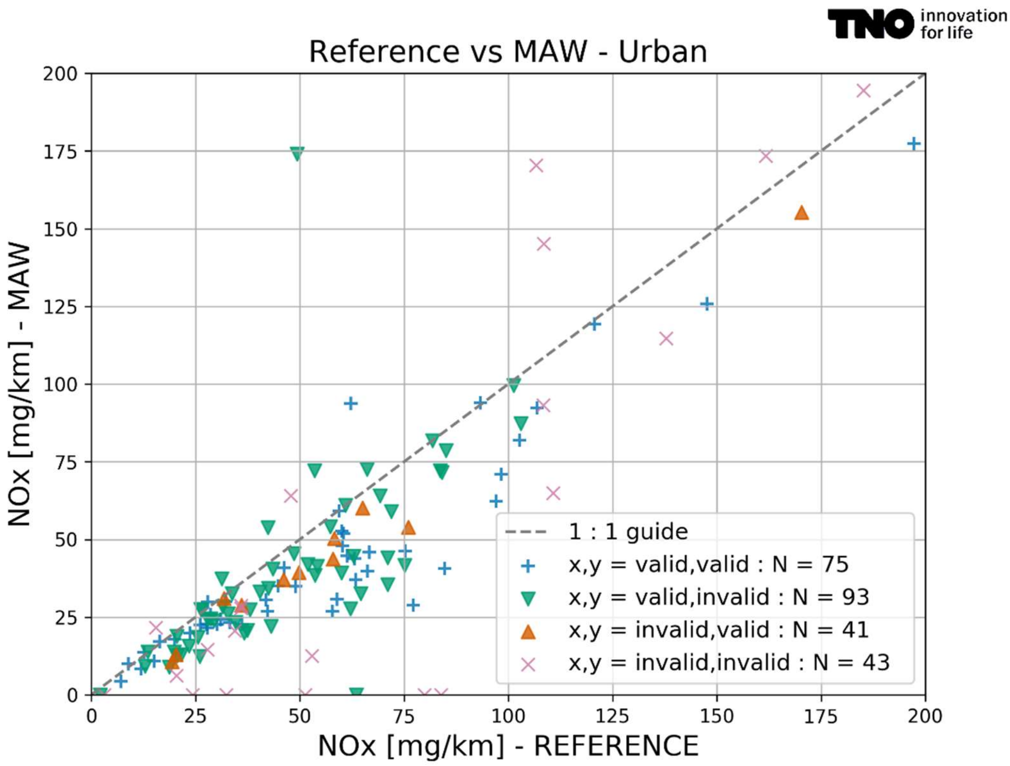
B.1.1 MAW

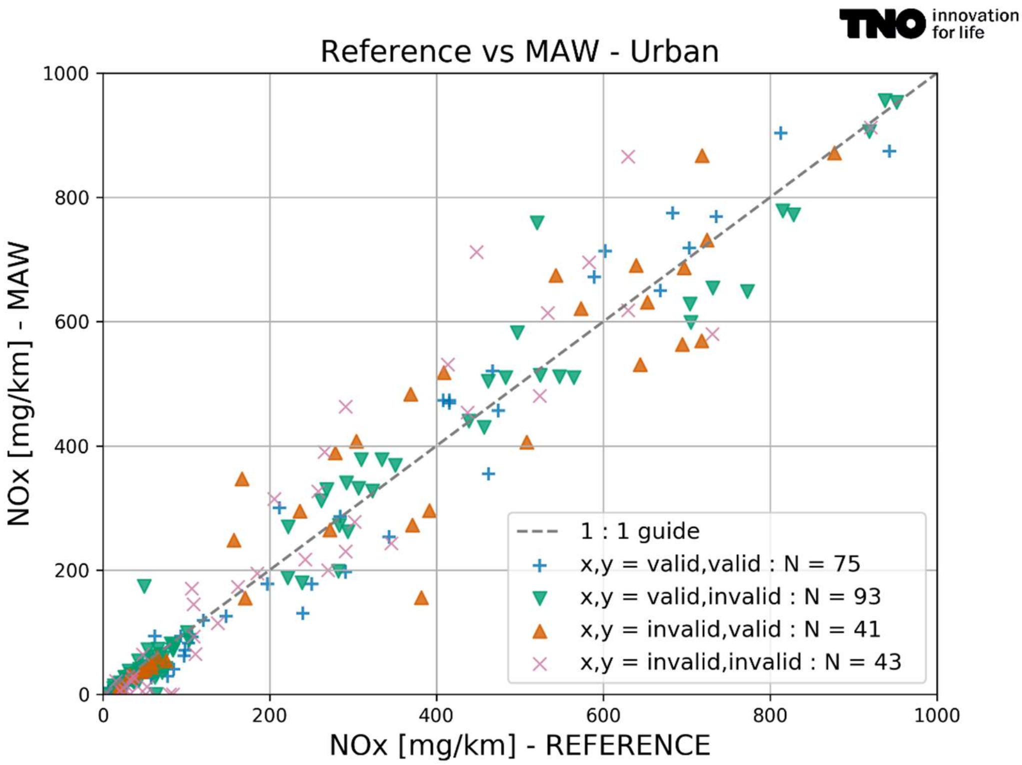

Reference vs MAW
1 : 1 guide
x,y = valid,valid : N = 75
x,y = valid,invalid : N = 93
x,y = invalid,valid : N = 41
x,y = invalid,invalid : N = 43
NOx [mg/km] - MAW
NOx [mg/km] - REFERENCE
0
25
50
75
100
125
150
175
200

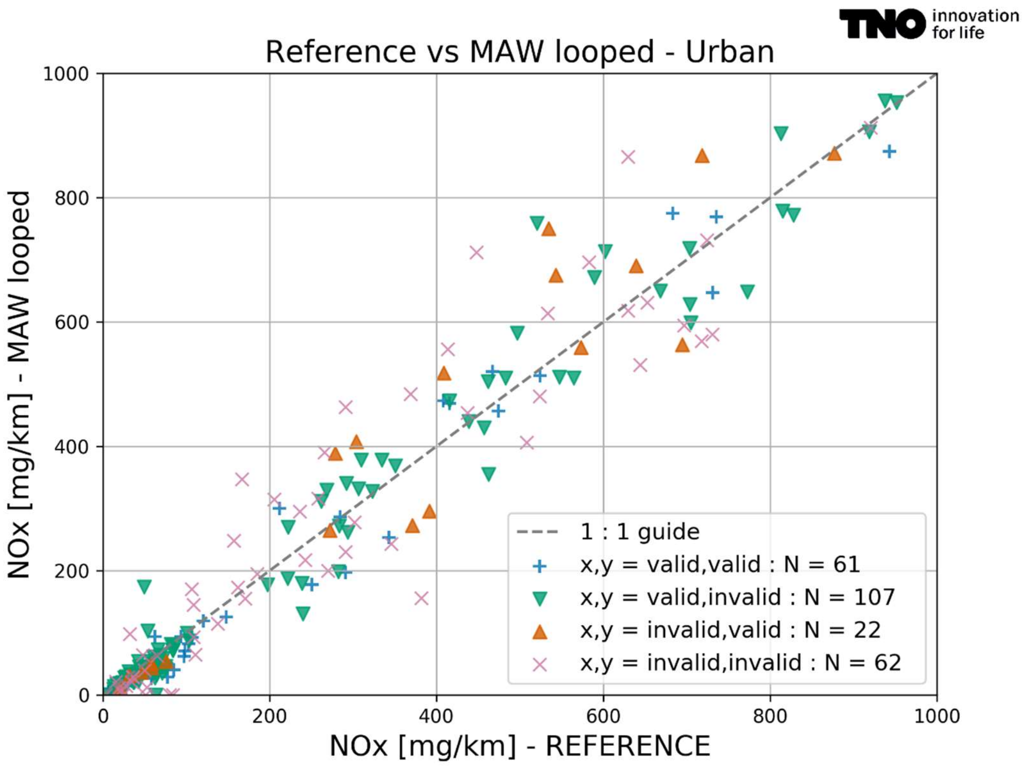

Reference vs MAW
1 : 1 guide
x,y = valid,valid : N = 75
x,y = valid,invalid : N = 93
x,y = invalid,valid : N = 41
x,y = invalid,invalid : N = 43
NOx [mg/km] - MAW
NOx [mg/km] - REFERENCE
0
200
400
600
800
1000

B.1.2 Looped

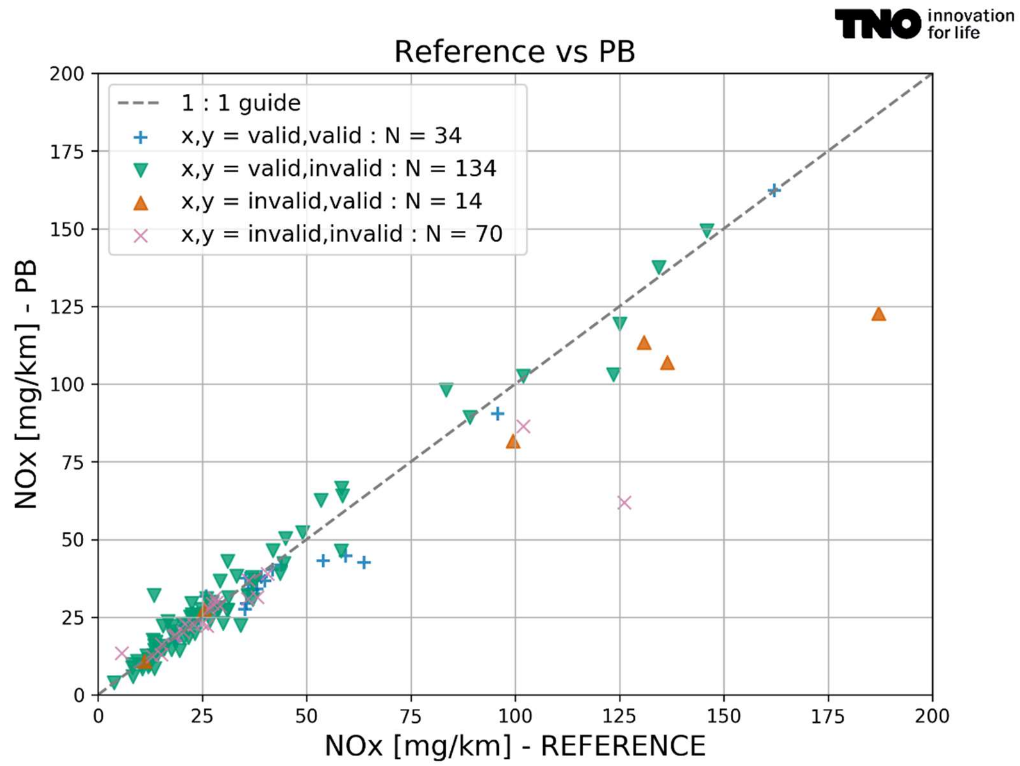

Reference vs MAW looped
1 : 1 guide
x,y = valid,valid : N = 61
x,y = valid,invalid : N = 107
x,y = invalid,valid : N = 22
x,y = invalid,invalid : N = 62
NOx [mg/km] - MAW looped
NOx [mg/km] - REFERENCE
0
200
400
600
800
1000

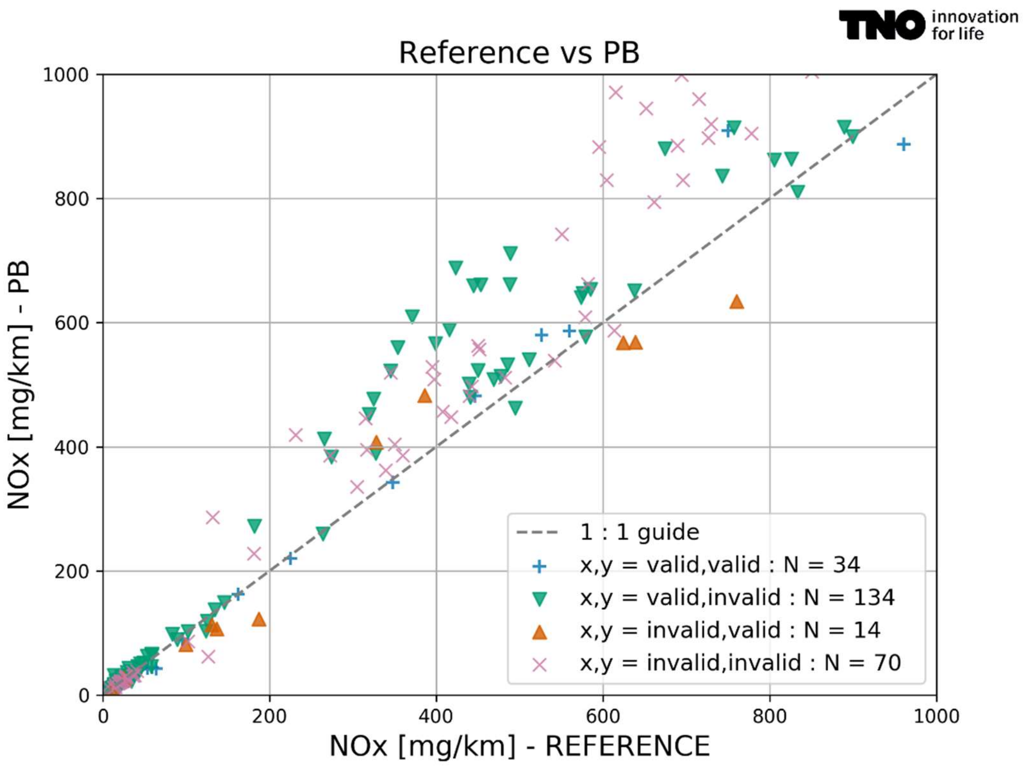

Reference vs MAW - Urban
NOx [mg/km] - MAW
NOx [mg/km] - REFERENCE
0
25
50
75
100
125
150
175
200
1 : 1 guide
x,y = valid,valid : N = 75
x,y = valid,invalid : N = 93
x,y = invalid,valid : N = 41
x,y = invalid,invalid : N = 43


TNO innovation for life
Reference vs MAW - Urban
NOx [mg/km] - MAW
NOx [mg/km] - REFERENCE
0
200
400
600
800
1000
1 : 1 guide
x,y = valid,valid : N = 75
x,y = valid,invalid : N = 93
x,y = invalid,valid : N = 41
x,y = invalid,invalid : N = 43

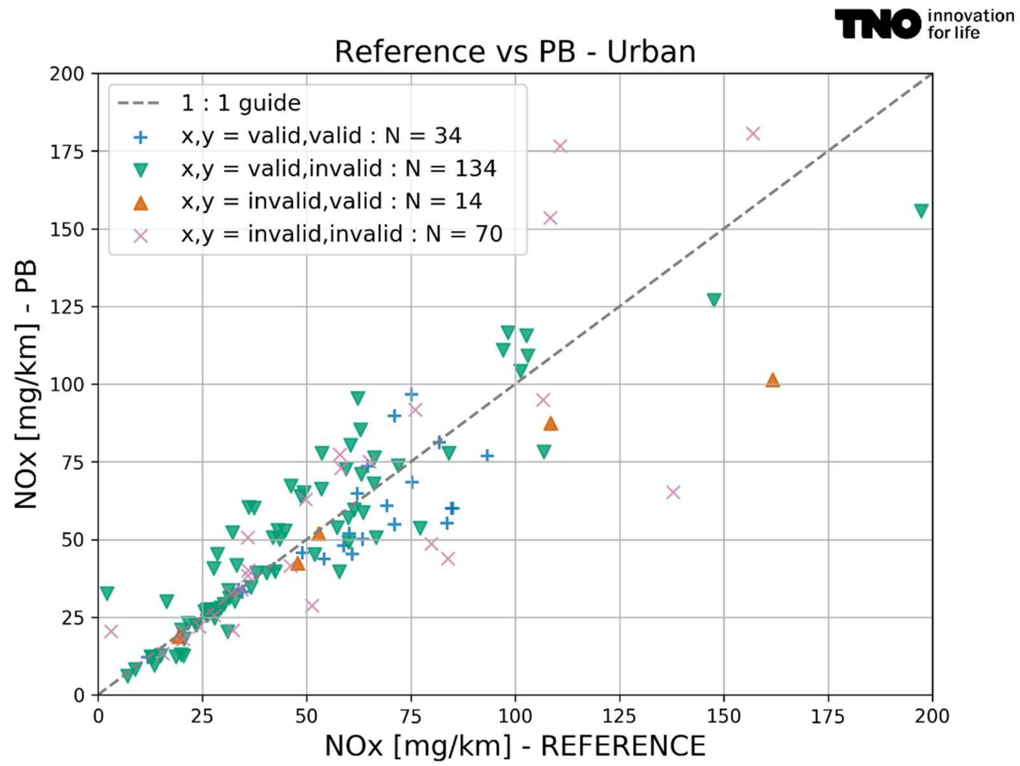

TNO innovation for life
Reference vs MAW looped - Urban
NOx [mg/km] - MAW looped
NOx [mg/km] - REFERENCE
0
200
400
600
800
1000
1 : 1 guide
x,y = valid,valid : N = 61
x,y = valid,invalid : N = 107
x,y = invalid,valid : N = 22
x,y = invalid,invalid : N = 62

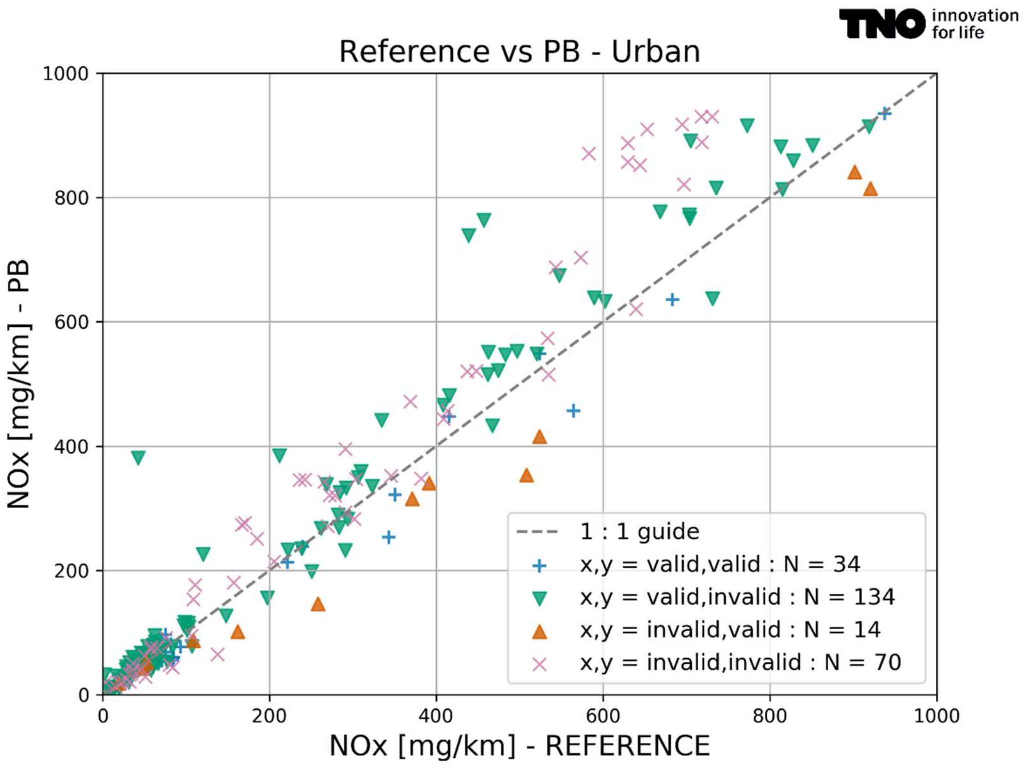
B.1.3 PB


TNO innovation for life
Reference vs PB
1 : 1 guide
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 134
x,y = invalid,valid : N = 14
x,y = invalid,invalid : N = 70
NOx [mg/km] - PB
NOx [mg/km] - REFERENCE
0
25
50
75
100
125
150
175
200

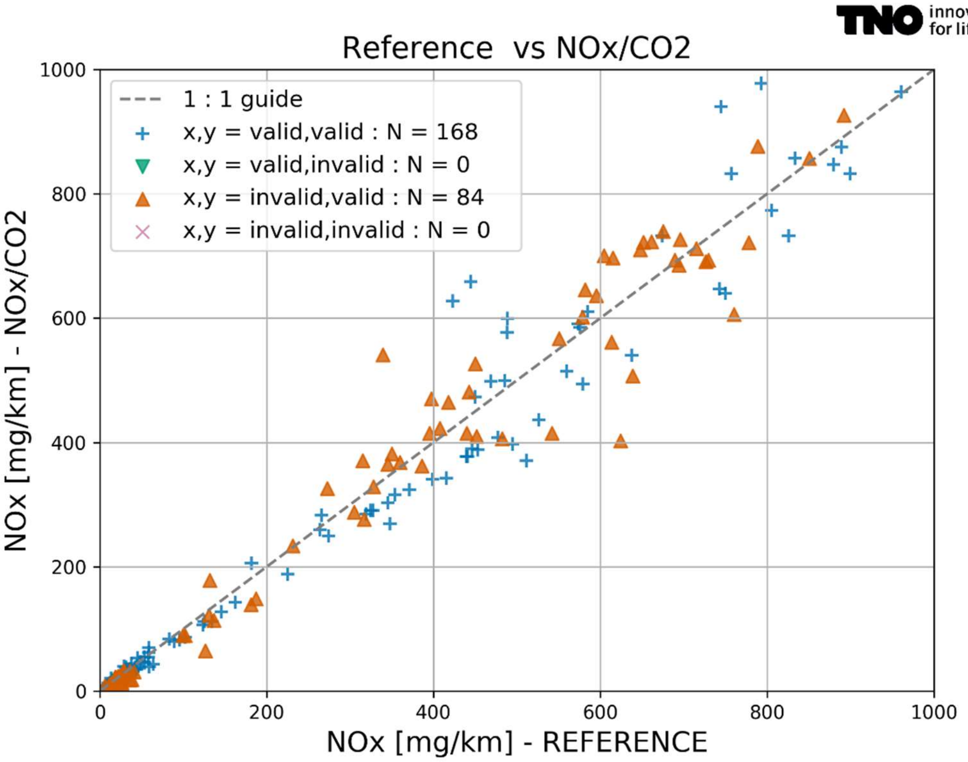

TNO innovation for life
Reference vs PB
1 : 1 guide
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 134
x,y = invalid,valid : N = 14
x,y = invalid,invalid : N = 70
NOx [mg/km] - PB
NOx [mg/km] - REFERENCE
0
200
400
600
800
1000

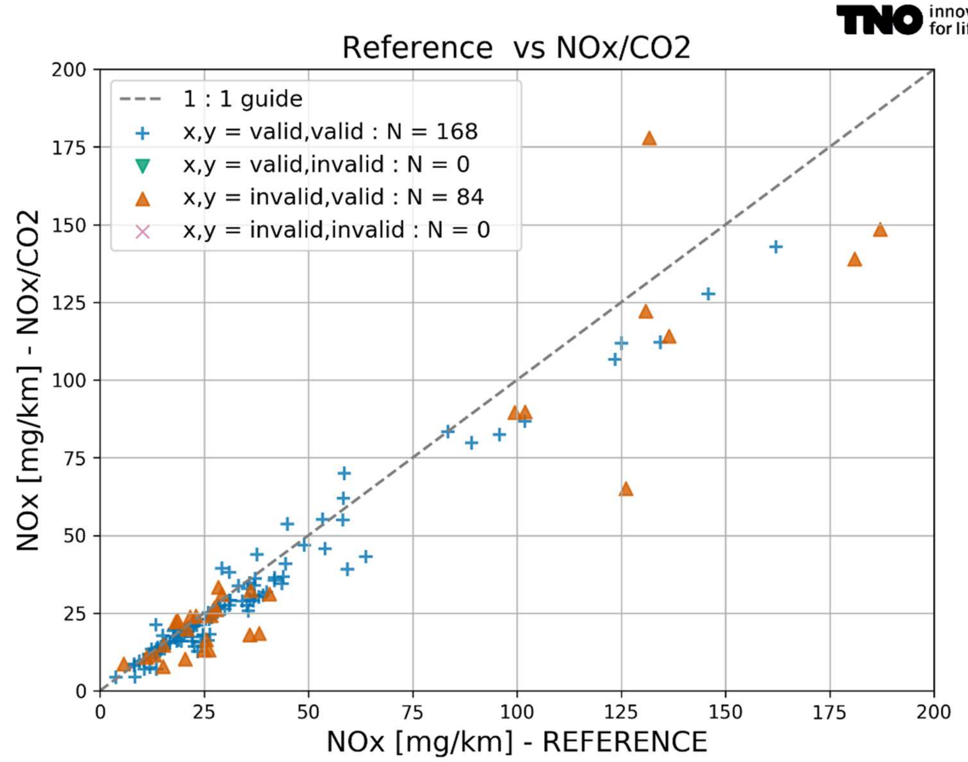
B.1.3.1 Urban


TNO innovation for life
Reference vs PB - Urban
1 : 1 guide
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 134
x,y = invalid,valid : N = 14
x,y = invalid,invalid : N = 70
NOx [mg/km] - PB
NOx [mg/km] - REFERENCE
0
25
50
75
100
125
150
175
200

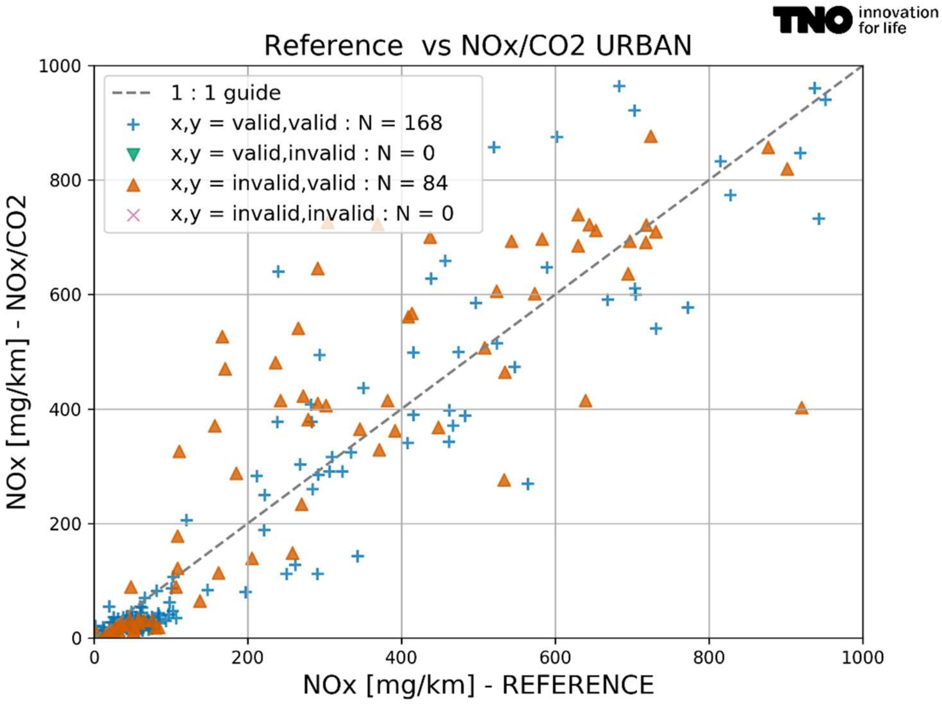

TNO innovation for life
Reference vs PB - Urban
1 : 1 guide
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 134
x,y = invalid,valid : N = 14
x,y = invalid,invalid : N = 70
NOx [mg/km] - PB
NOx [mg/km] - REFERENCE
0
200
400
600
800
1000

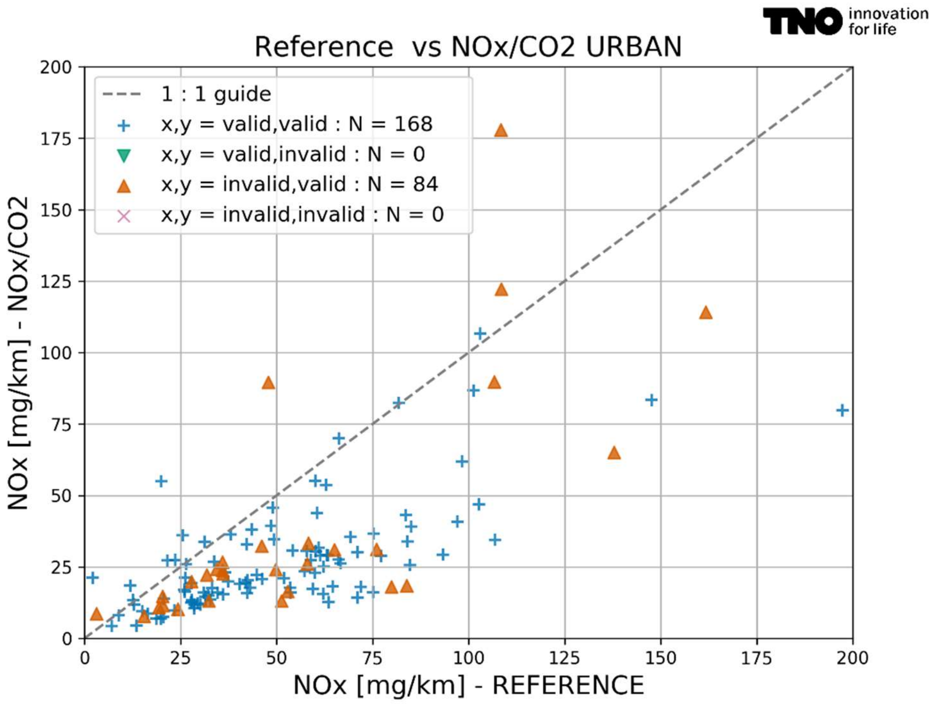
B.1.4 NOx/CO2


Reference vs NOx/CO2
1 : 1 guide
x,y = valid,valid : N = 168
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 84
x,y = invalid,invalid : N = 0
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
0
200
400
600
800
1000

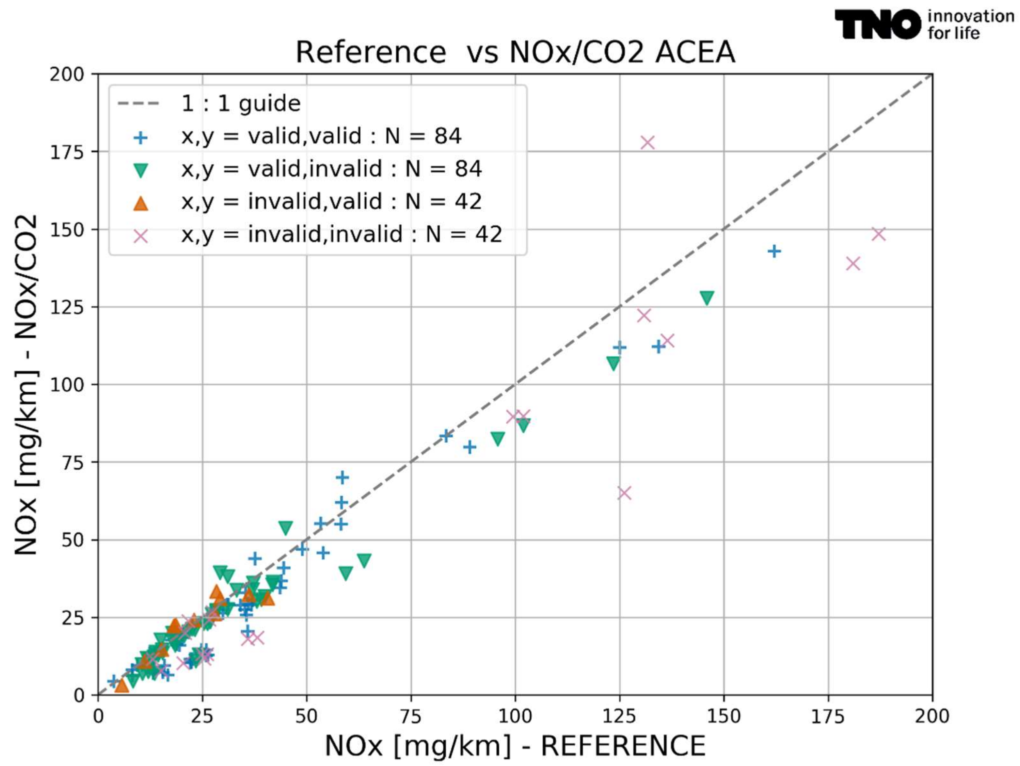

Reference vs NOx/CO2
1 : 1 guide
x,y = valid,valid : N = 168
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 84
x,y = invalid,invalid : N = 0
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
0
25
50
75
100
125
150
175
200

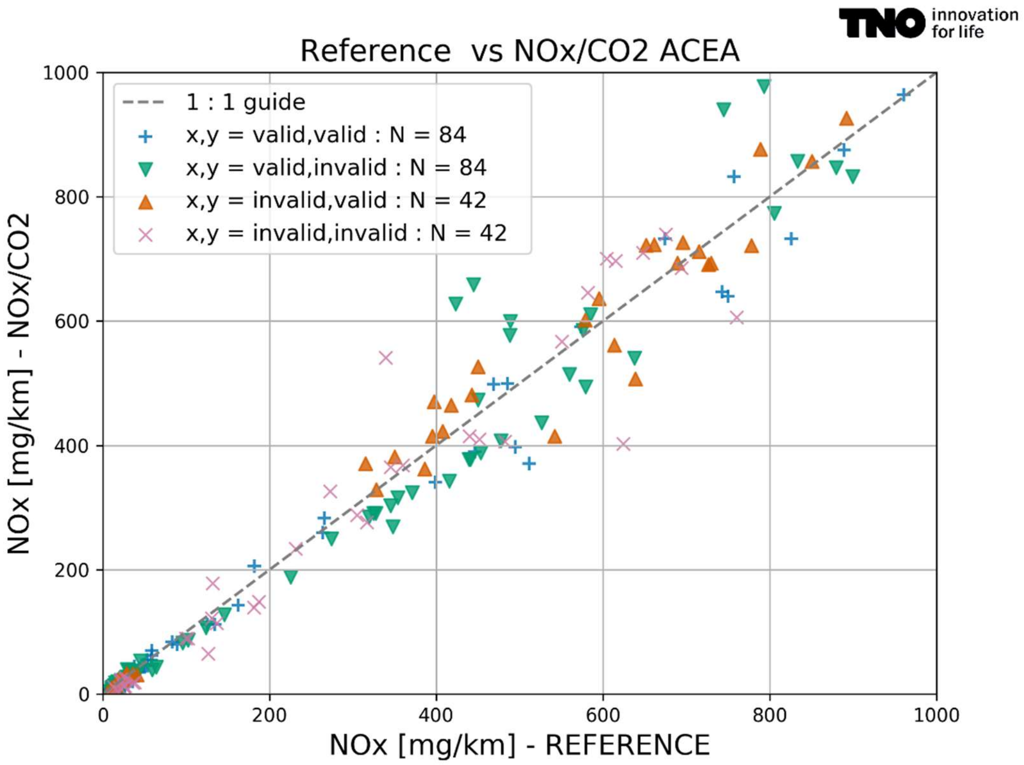
B.1.5 NOx/CO2 urban



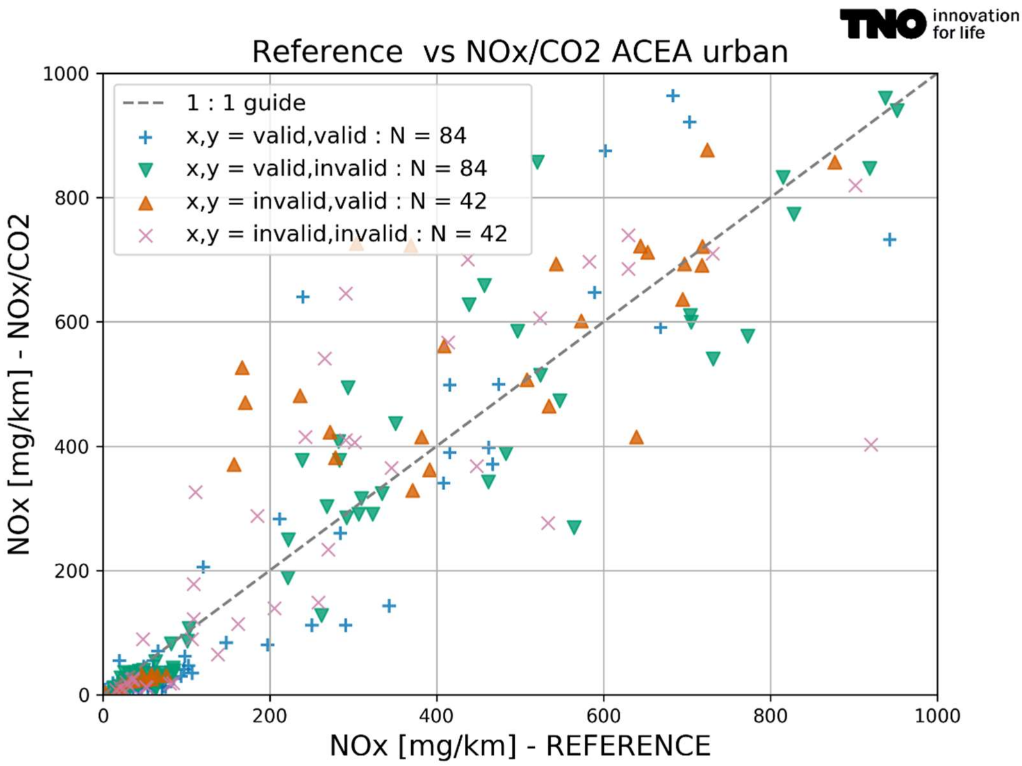


B.1.6 ACEA

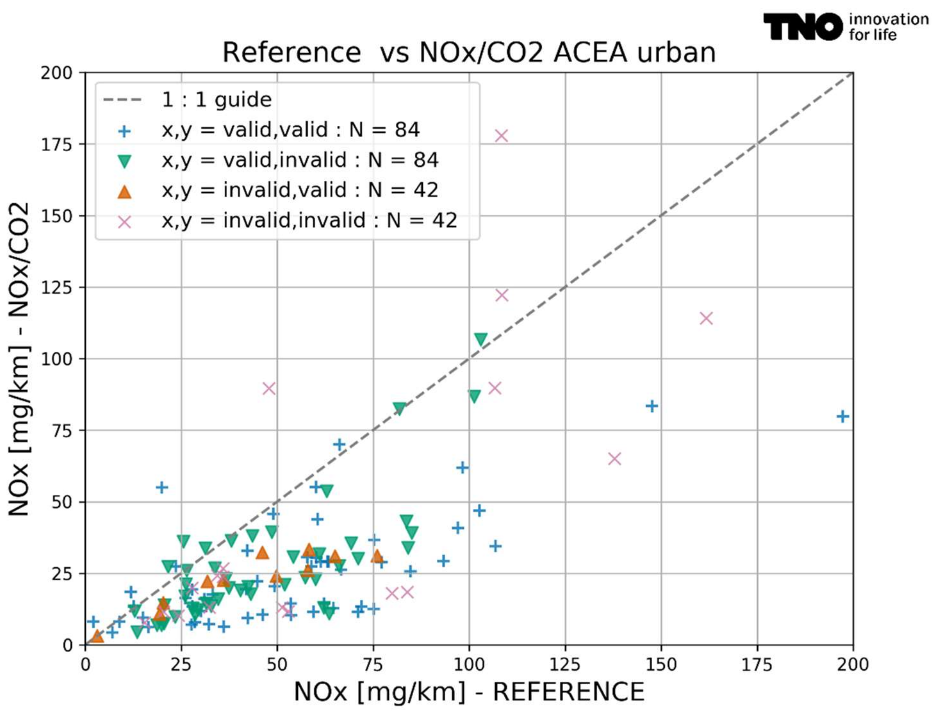

TNO innovation for life
Reference vs NOx/CO2 ACEA
1 : 1 guide
x,y = valid,valid : N = 84
x,y = valid,invalid : N = 84
x,y = invalid,valid : N = 42
x,y = invalid,invalid : N = 42
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
0
25
50
75
100
125
150
175
200


TNO innovation for life
Reference vs NOx/CO2 ACEA
1 : 1 guide
x,y = valid,valid : N = 84
x,y = valid,invalid : N = 84
x,y = invalid,valid : N = 42
x,y = invalid,invalid : N = 42
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
0
200
400
600
800
1000

B.1.6.1 *Urban*



Reference vs NOx/CO2 ACEA urban
1 : 1 guide
x,y = valid,valid : N = 84
x,y = valid,invalid : N = 84
x,y = invalid,valid : N = 42
x,y = invalid,invalid : N = 42
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
0
200
400
600
800
1000

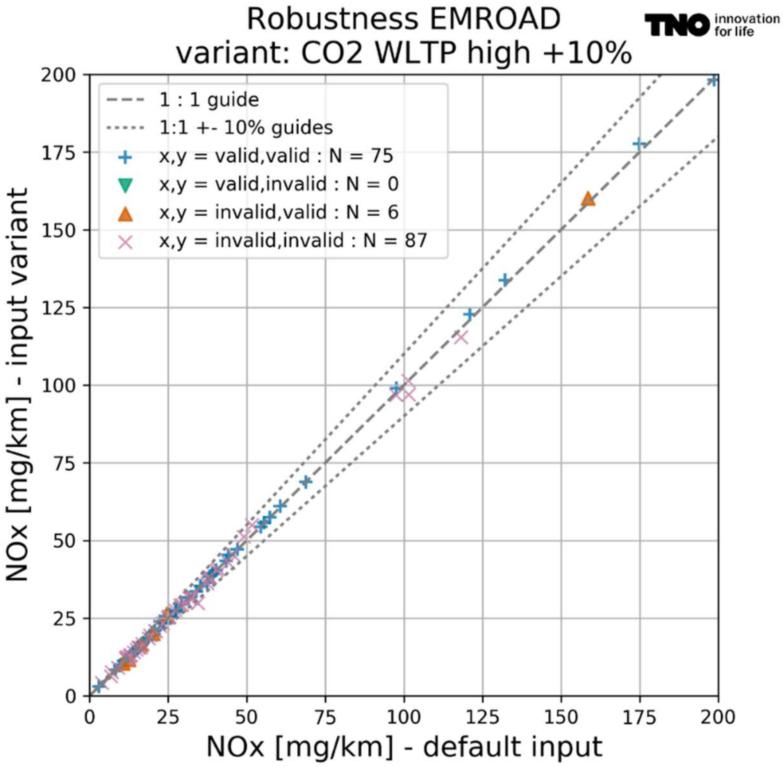


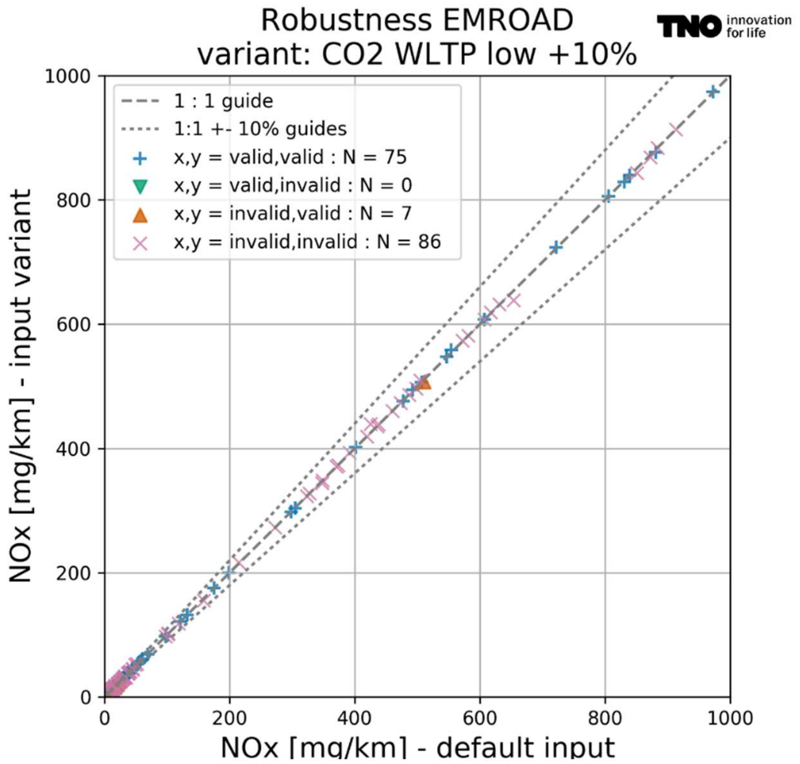
Reference vs NOx/CO2 ACEA urban
1 : 1 guide
x,y = valid,valid : N = 84
x,y = valid,invalid : N = 84
x,y = invalid,valid : N = 42
x,y = invalid,invalid : N = 42
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
0
25
50
75
100
125
150
175
200

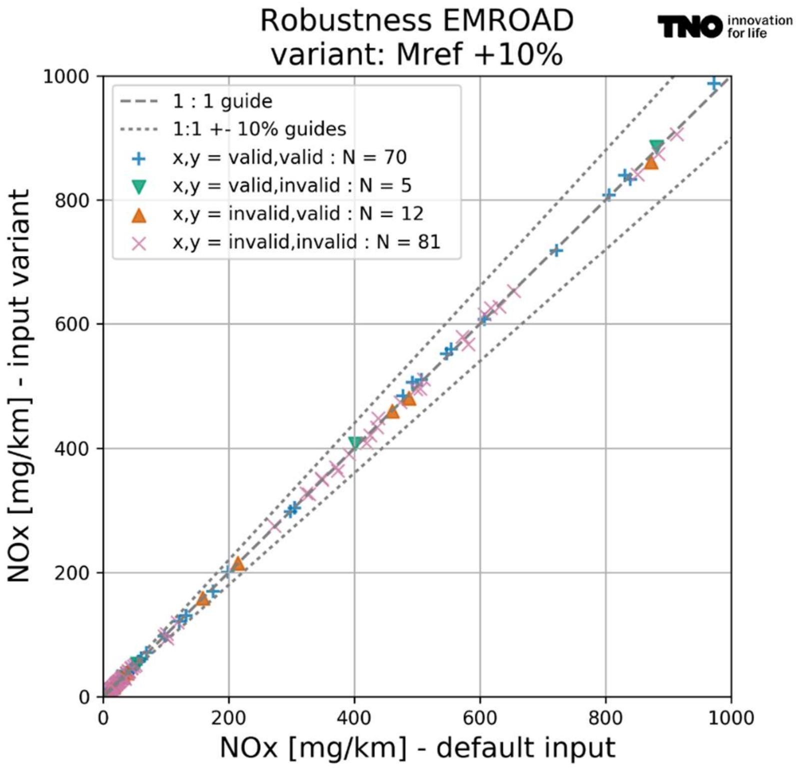
B.2 Robustness

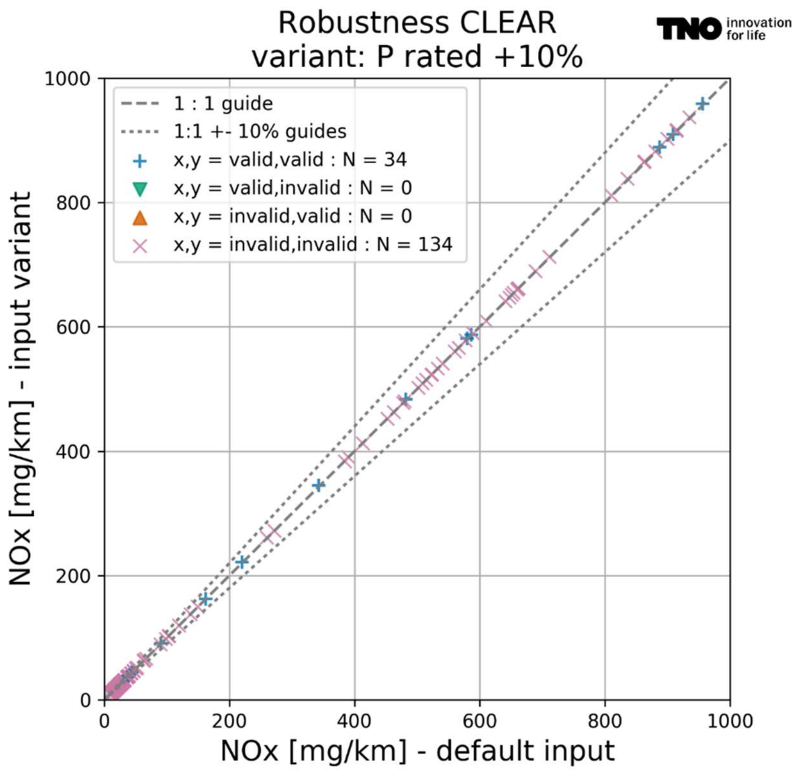
As was mentioned, robustness is defined as an analysis of the change in output of the evaluation methods when the WLTP input values are changed. Firstly, the robustness of MAW is presented, and secondly the robustness of PB.

B.2.1 *MAW*

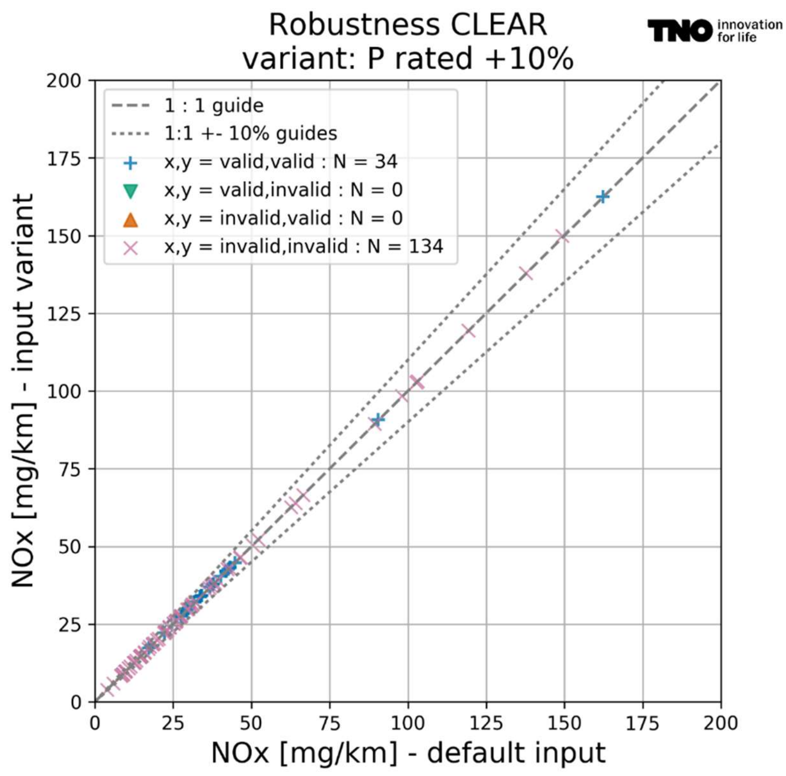

Robustness EMROAD
variant: CO2 WLTP high +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 75
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 6
x,y = invalid,invalid : N = 87
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
25
50
75
100
125
150
175
200

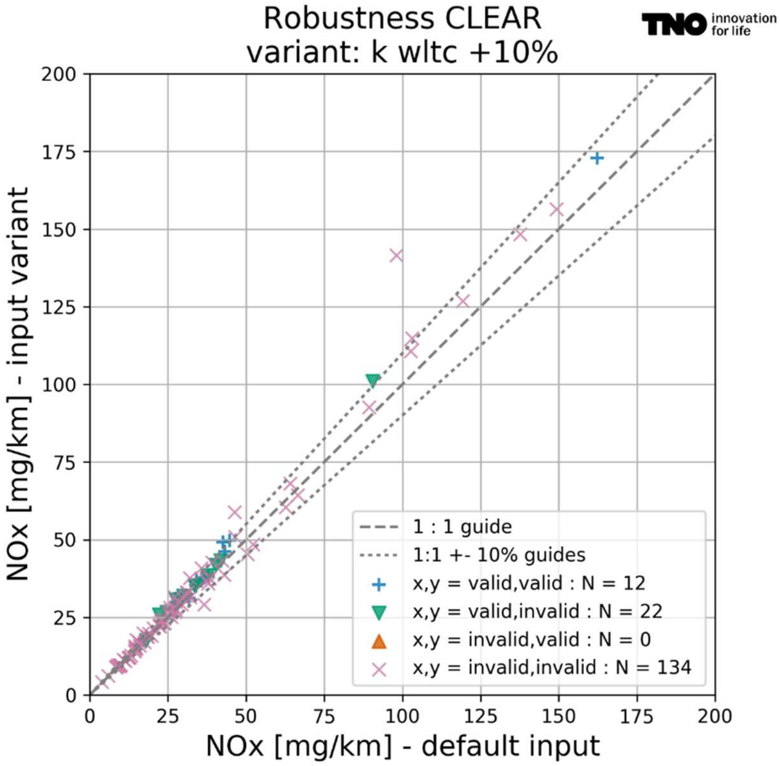

Robustness EMROAD
variant: CO2 WLTP low +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 75
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 7
x,y = invalid,invalid : N = 86
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
200
400
600
800
1000

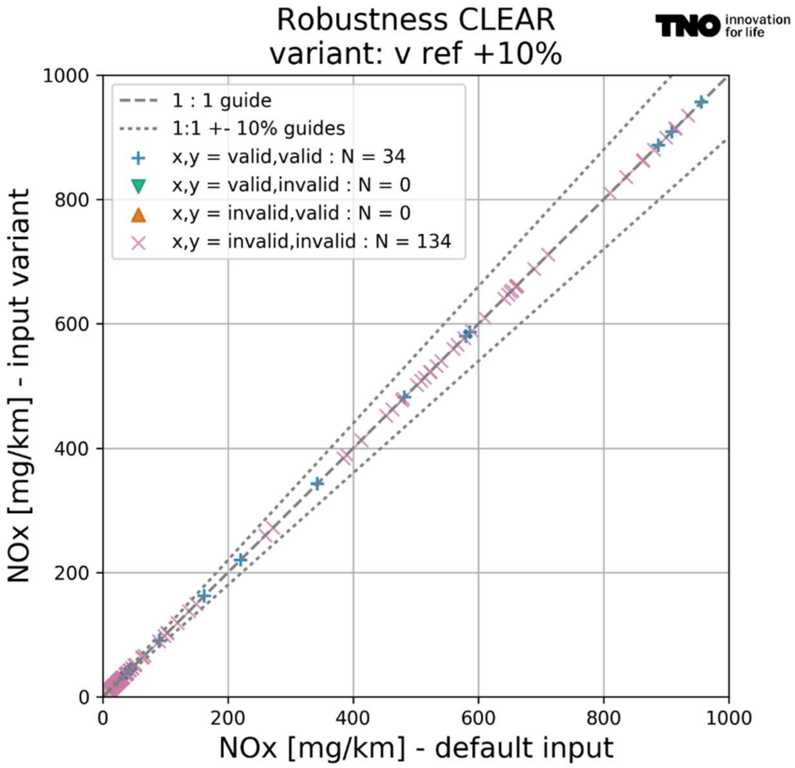

Robustness EMROAD
variant: Mref +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 70
x,y = valid,invalid : N = 5
x,y = invalid,valid : N = 12
x,y = invalid,invalid : N = 81
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
200
400
600
800
1000

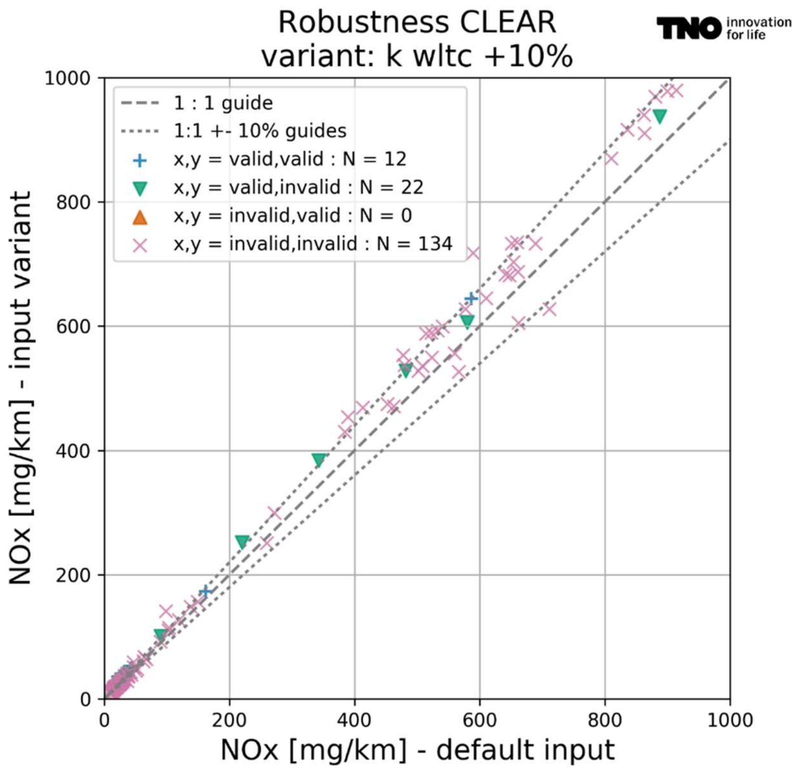

Robustness CLEAR
variant: P rated +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 0
x,y = invalid,invalid : N = 134
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
200
400
600
800
1000

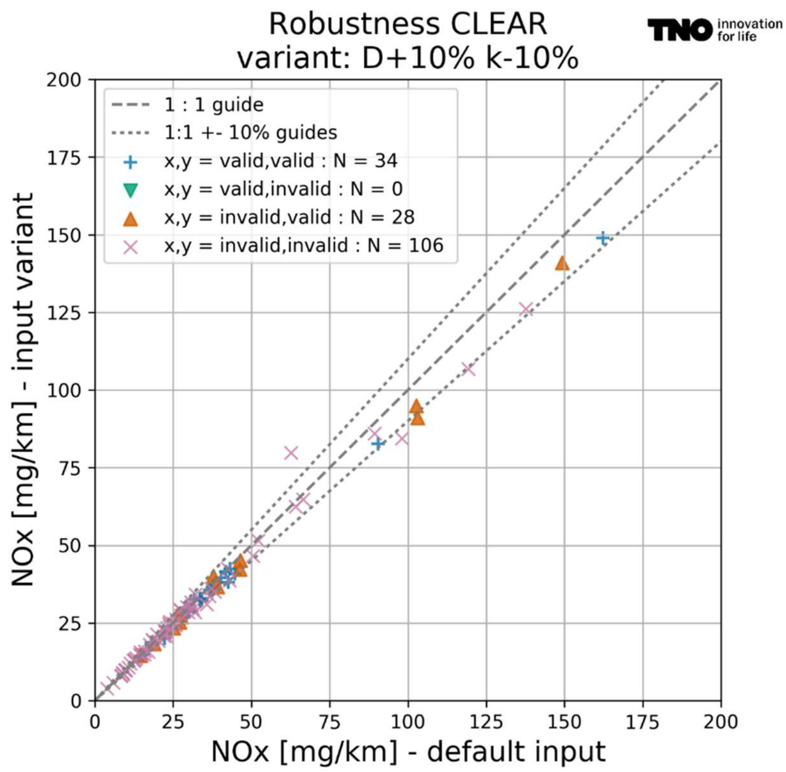
B.2.2 PB

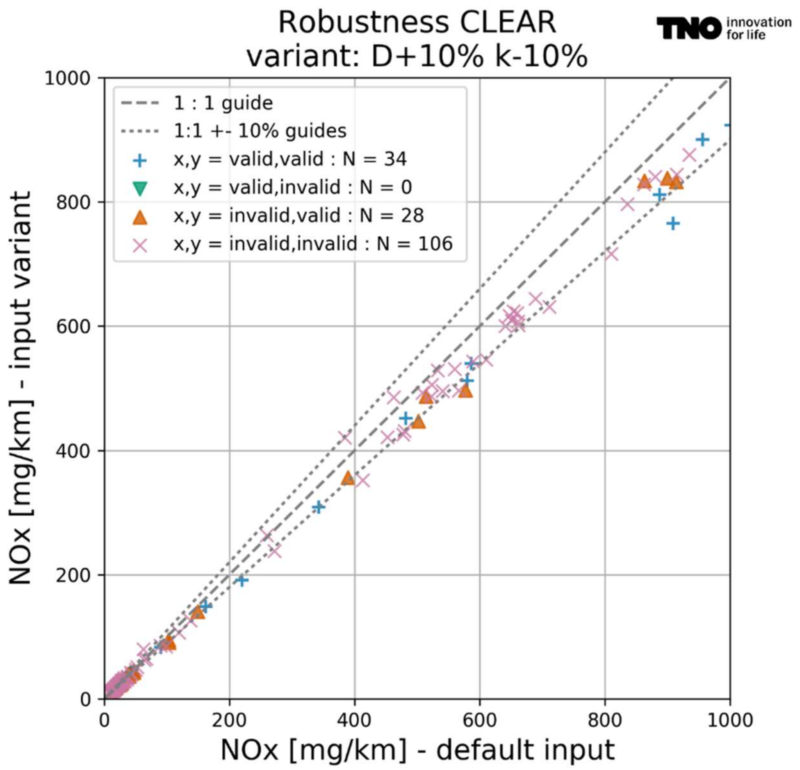

Robustness CLEAR
variant: P rated +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 0
x,y = invalid,invalid : N = 134
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
25
50
75
100
125
150
175
200

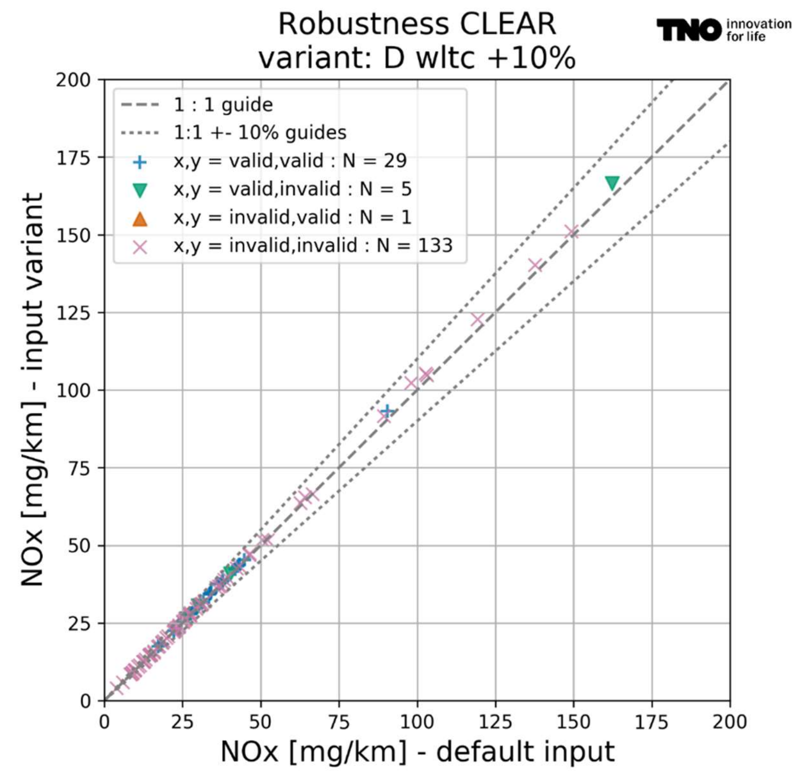

Robustness CLEAR
variant: k wltc +10%
TNO innovation for life
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
25
50
75
100
125
150
175
200
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 12
x,y = valid,invalid : N = 22
x,y = invalid,valid : N = 0
x,y = invalid,invalid : N = 134

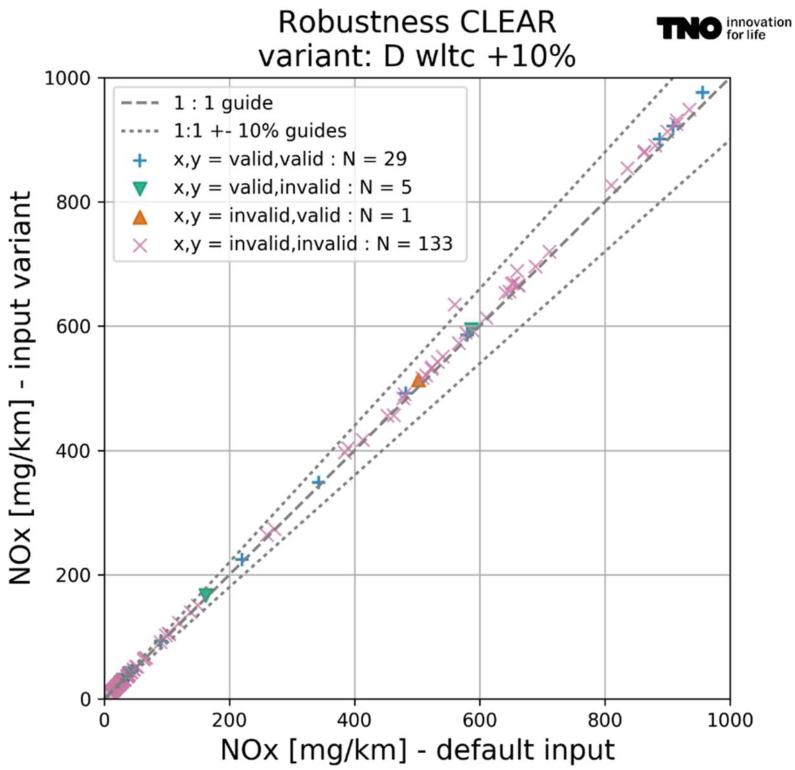

Robustness CLEAR
variant: v ref +10%
TNO innovation for life
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
200
400
600
800
1000
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 0
x,y = invalid,invalid : N = 134


Robustness CLEAR
variant: k wltc +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 12
x,y = valid,invalid : N = 22
x,y = invalid,valid : N = 0
x,y = invalid,invalid : N = 134
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
200
400
600
800
1000

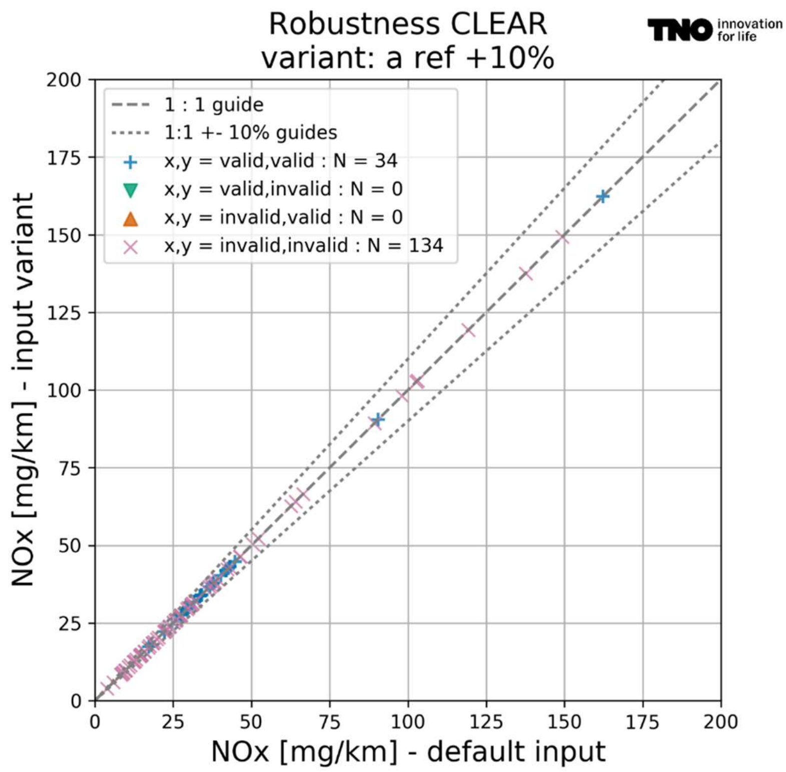

Robustness CLEAR
variant: D+10% k-10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 28
x,y = invalid,invalid : N = 106
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
25
50
75
100
125
150
175
200


Robustness CLEAR
variant: D+10% k-10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 28
x,y = invalid,invalid : N = 106
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
200
400
600
800
1000

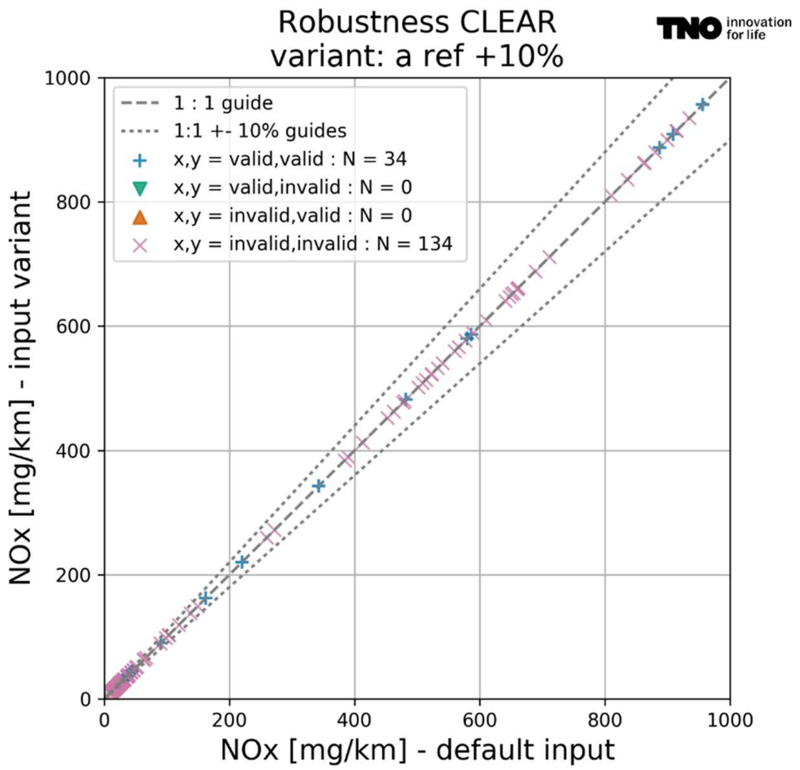

Robustness CLEAR
variant: D wltc +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 29
x,y = valid,invalid : N = 5
x,y = invalid,valid : N = 1
x,y = invalid,invalid : N = 133
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
25
50
75
100
125
150
175
200

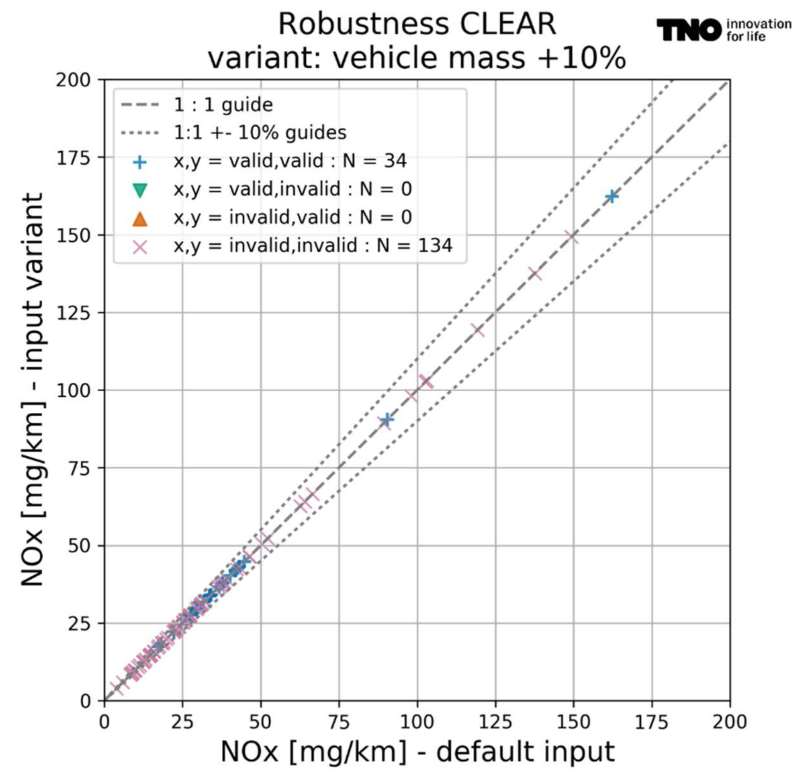


Robustness CLEAR
variant: D wltc +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 29
x,y = valid,invalid : N = 5
x,y = invalid,valid : N = 1
x,y = invalid,invalid : N = 133
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
200
400
600
800
1000



Robustness CLEAR
variant: a ref +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 0
x,y = invalid,invalid : N = 134
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
25
50
75
100
125
150
175
200

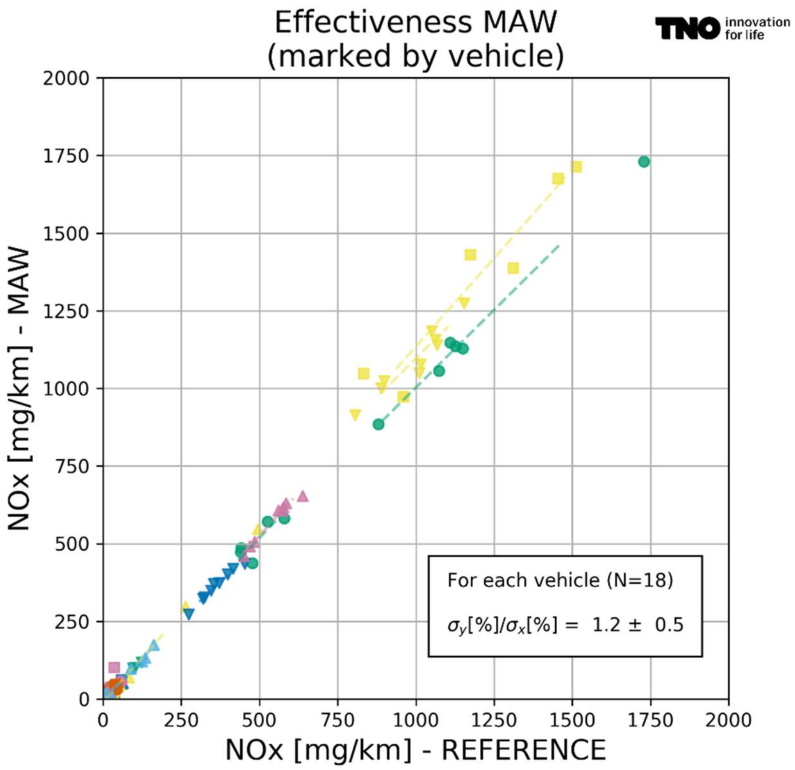

Robustness CLEAR
variant: a ref +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 0
x,y = invalid,invalid : N = 134
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
200
400
600
800
1000

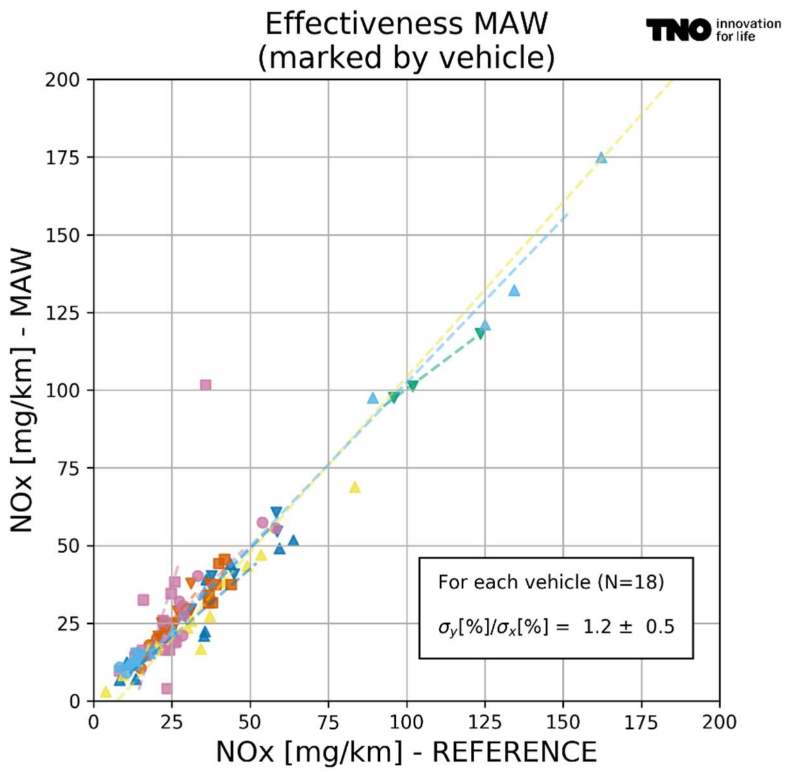

Robustness CLEAR
variant: vehicle mass +10%
TNO innovation for life
1 : 1 guide
1:1 +- 10% guides
x,y = valid,valid : N = 34
x,y = valid,invalid : N = 0
x,y = invalid,valid : N = 0
x,y = invalid,invalid : N = 134
NOx [mg/km] - input variant
NOx [mg/km] - default input
0
25
50
75
100
125
150
175
200

B.3 Effectiveness

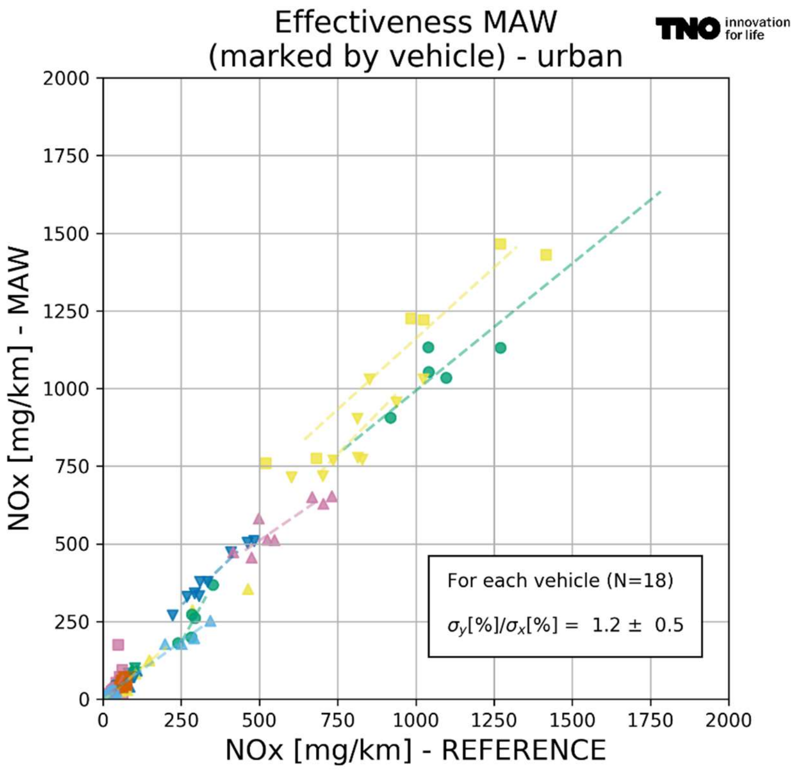
The definition of effectiveness analysis was defined in chapter 7. Here, additional analyses for the effectiveness of MAW, PB, NOx/CO2, ACEA and their urban parts are presented.

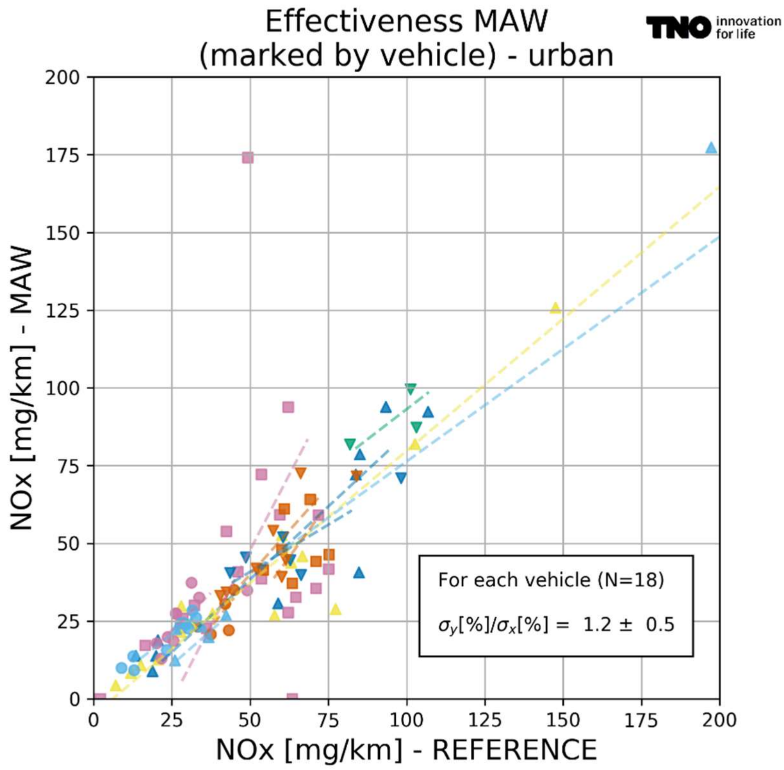
B.3.1 MAW


Effectiveness MAW
(marked by vehicle)
TNO innovation for life
NOx [mg/km] - MAW
NOx [mg/km] - REFERENCE
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%]$ = 1.2 ± 0.5
0
250
500
750
1000
1250
1500
1750
2000

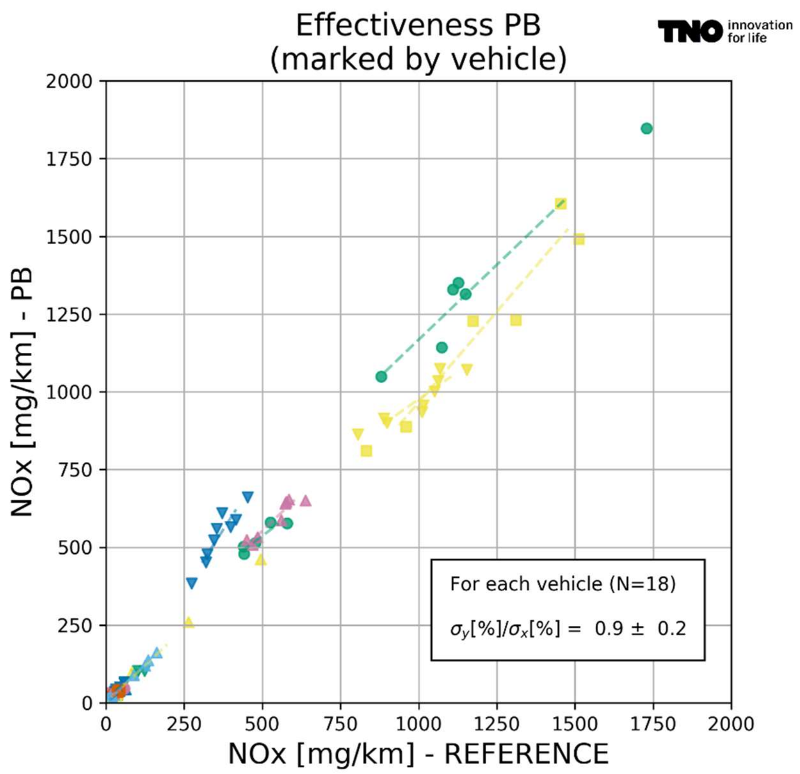

Effectiveness MAW
(marked by vehicle)
TNO innovation for life
NOx [mg/km] - MAW
NOx [mg/km] - REFERENCE
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%]$ = 1.2 ± 0.5
0
25
50
75
100
125
150
175
200

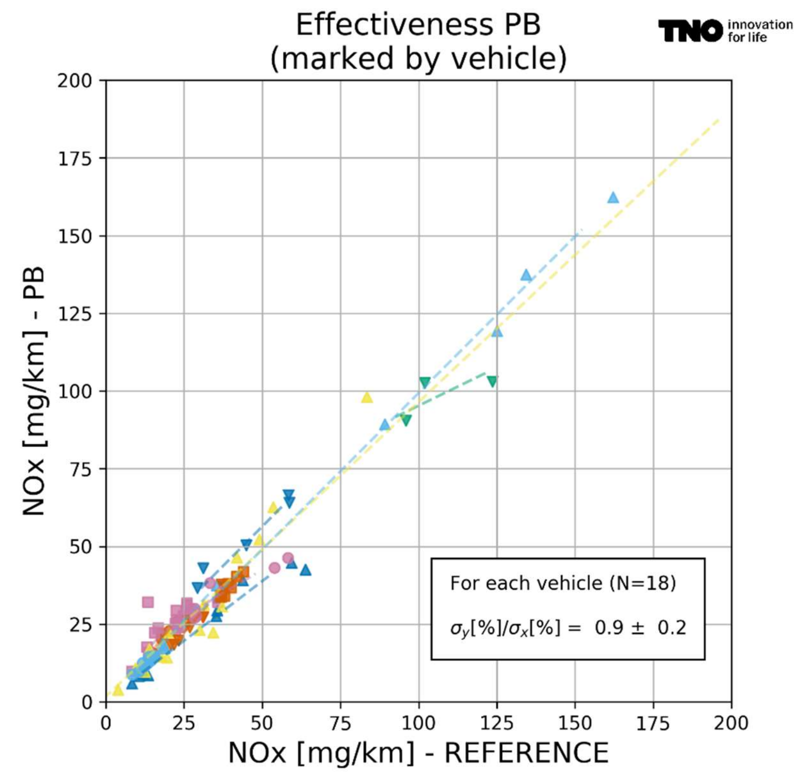
B.3.2 Urban


Effectiveness MAW
(marked by vehicle) - urban
TNO innovation for life
NOx [mg/km] - MAW
NOx [mg/km] - REFERENCE
2000
1750
1500
1250
1000
750
500
250
0
0
250
500
750
1000
1250
1500
1750
2000
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%]$ = 1.2 ± 0.5

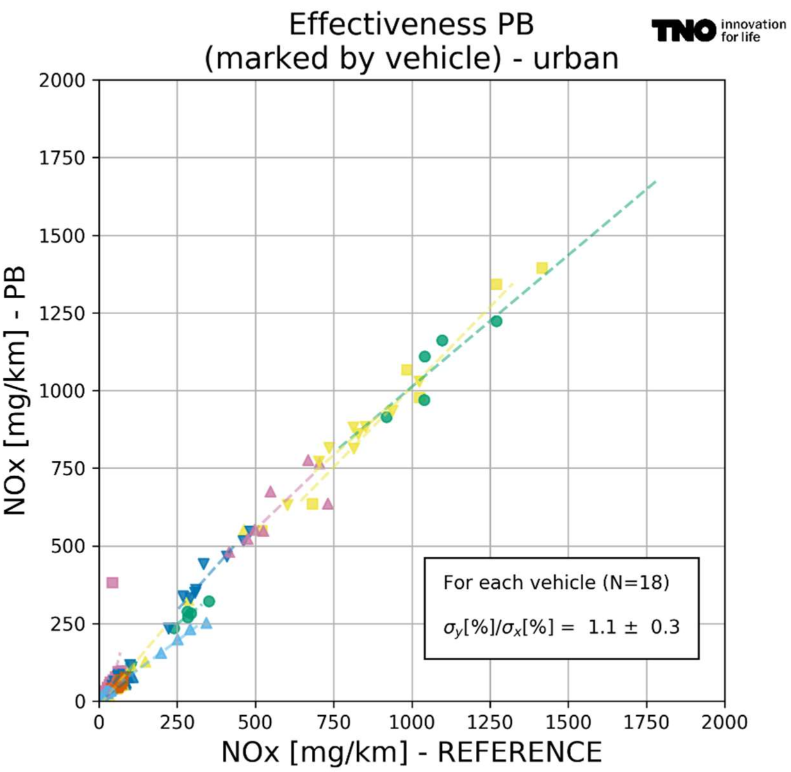

Effectiveness MAW
(marked by vehicle) - urban
TNO innovation for life
NOx [mg/km] - MAW
NOx [mg/km] - REFERENCE
200
175
150
125
100
75
50
25
0
0
25
50
75
100
125
150
175
200
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%]$ = 1.2 ± 0.5

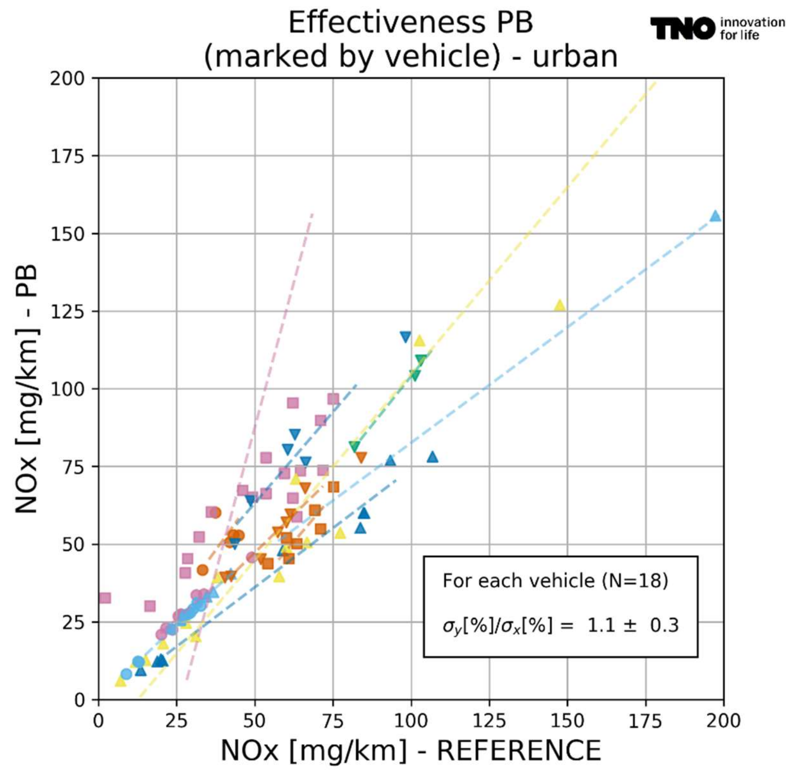
B.4 PB

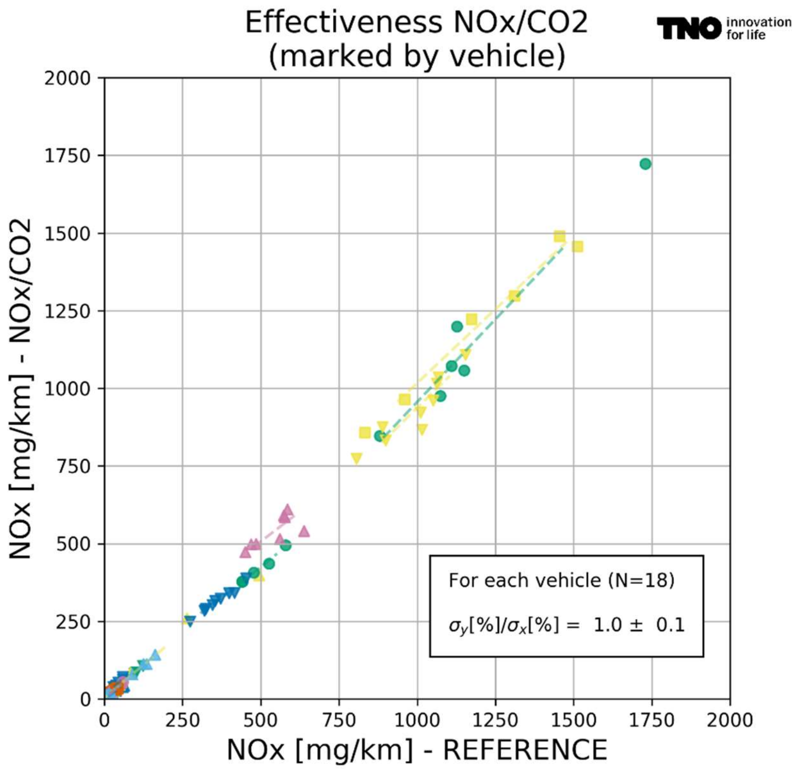

Effectiveness PB
(marked by vehicle)
innovation for life
NOx [mg/km] - PB
NOx [mg/km] - REFERENCE
0
250
500
750
1000
1250
1500
1750
2000
For each vehicle (N=18)
σ_y[%]/σ_x[%] = 0.9 ± 0.2

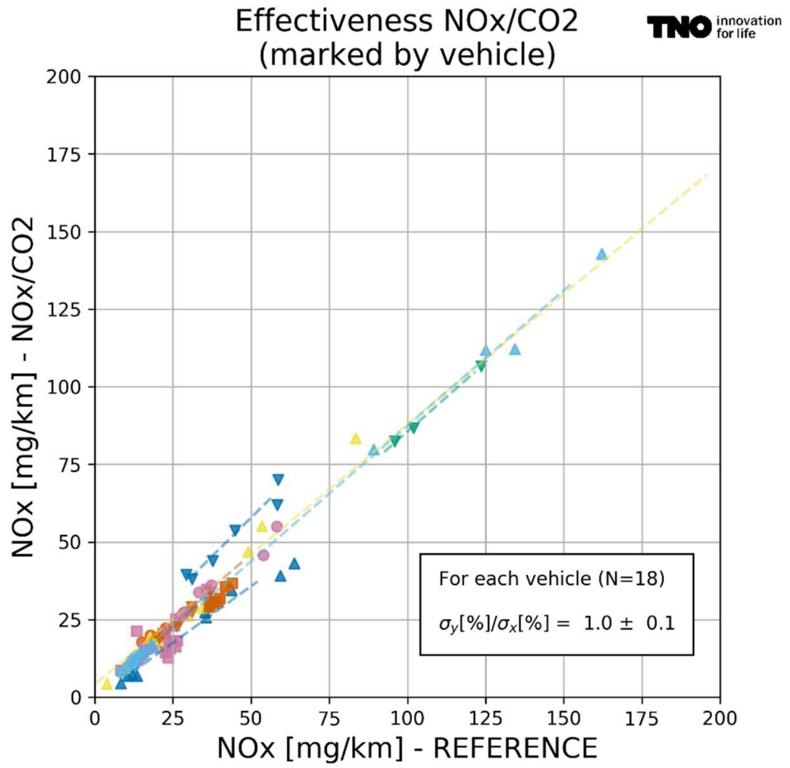

Effectiveness PB
(marked by vehicle)
innovation for life
NOx [mg/km] - PB
NOx [mg/km] - REFERENCE
0
25
50
75
100
125
150
175
200
For each vehicle (N=18)
σ_y[%]/σ_x[%] = 0.9 ± 0.2

B.4.1 Urban

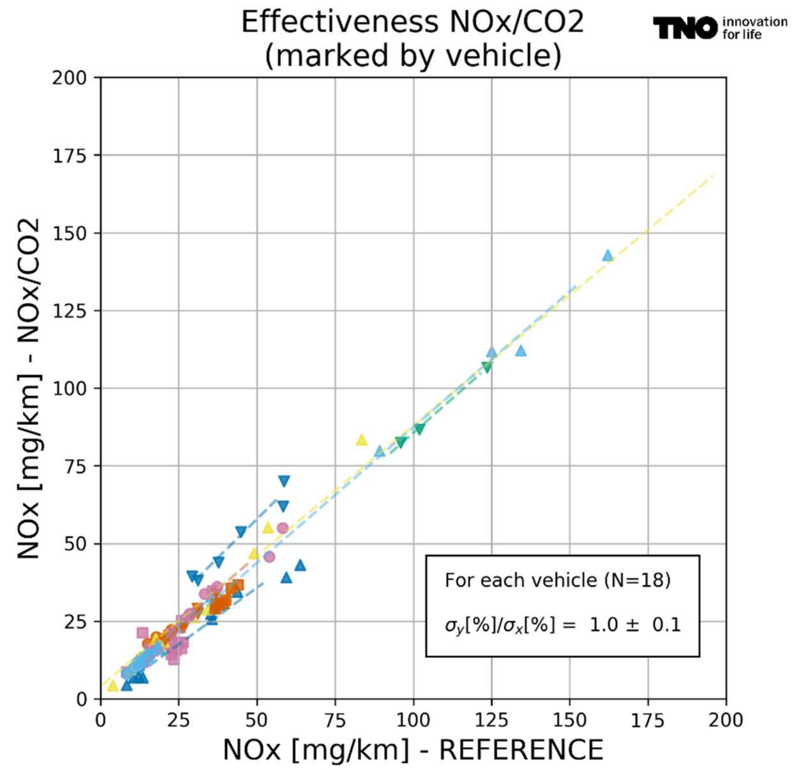

Effectiveness PB
(marked by vehicle) - urban
TNO innovation for life
NOx [mg/km] - PB
NOx [mg/km] - REFERENCE
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%]$ = 1.1 ± 0.3
0
250
500
750
1000
1250
1500
1750
2000

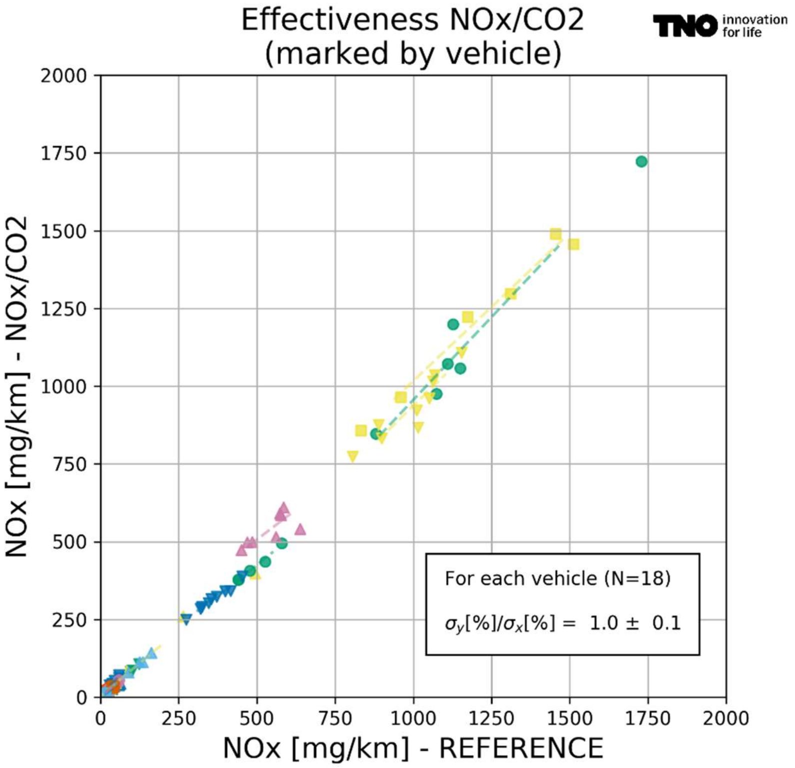

Effectiveness PB
(marked by vehicle) - urban
TNO innovation for life
NOx [mg/km] - PB
NOx [mg/km] - REFERENCE
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%]$ = 1.1 ± 0.3
0
25
50
75
100
125
150
175
200


Effectiveness NOx/CO2
(marked by vehicle)
TNO innovation for life
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
0
250
500
750
1000
1250
1500
1750
2000
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%]$ = 1.0 ± 0.1

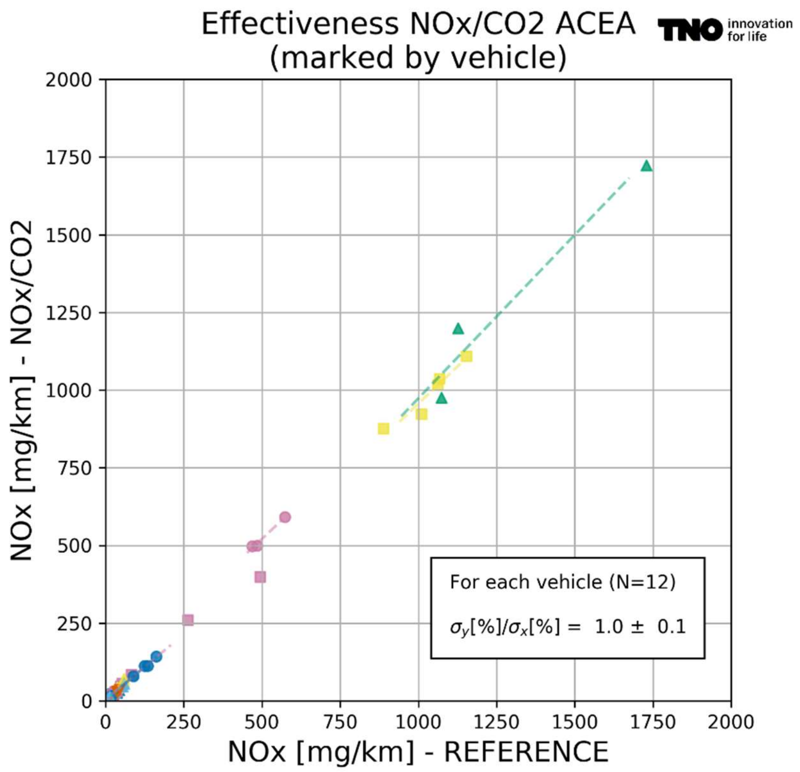

Effectiveness NOx/CO2
(marked by vehicle)
TNO innovation for life
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
0
25
50
75
100
125
150
175
200
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%]$ = 1.0 ± 0.1

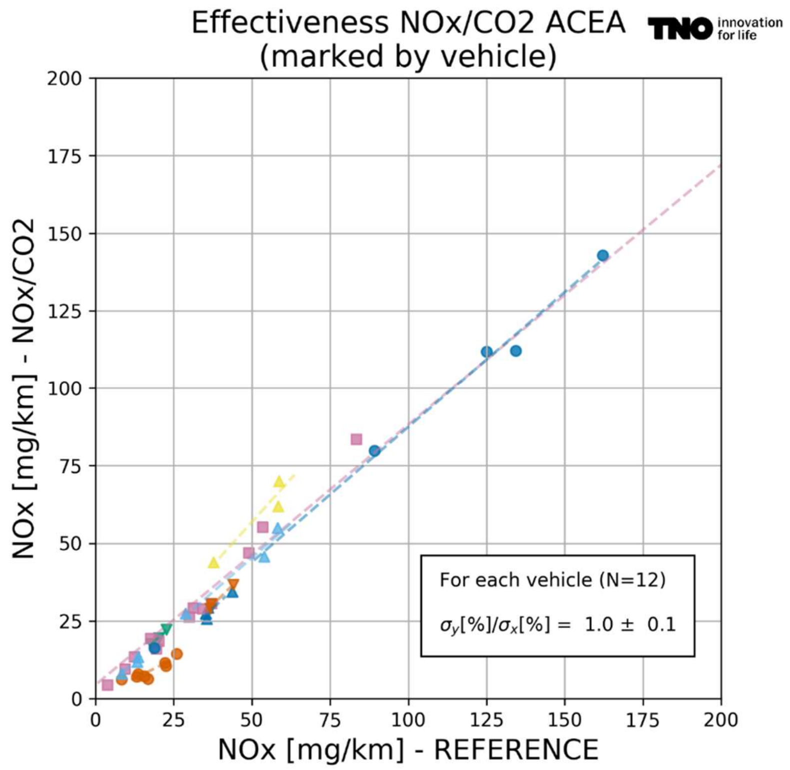
B.5 NOx/CO2

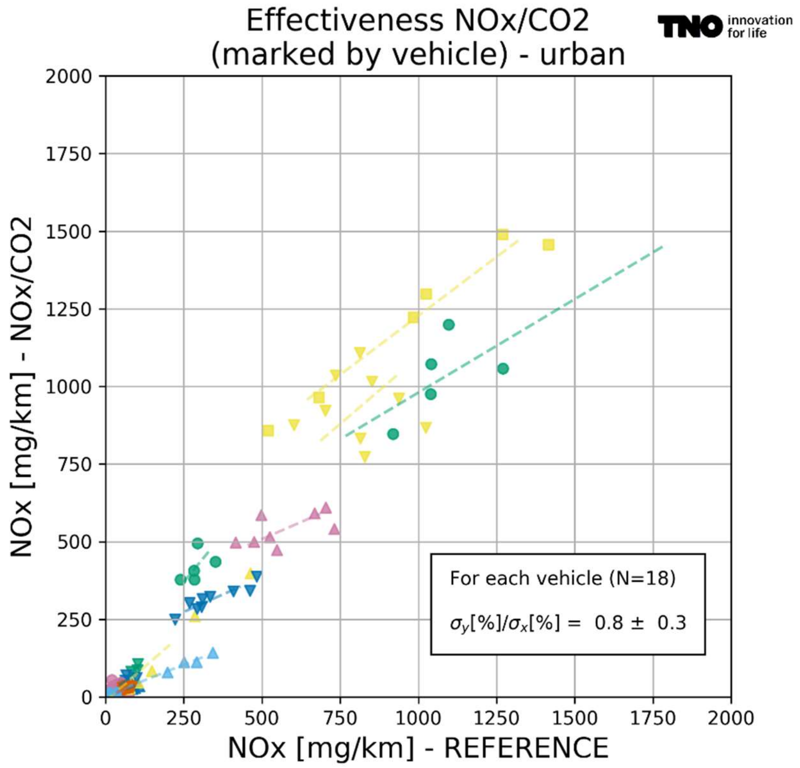

Effectiveness NOx/CO2
(marked by vehicle)
TNO innovation for life
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%] = 1.0 \pm 0.1$

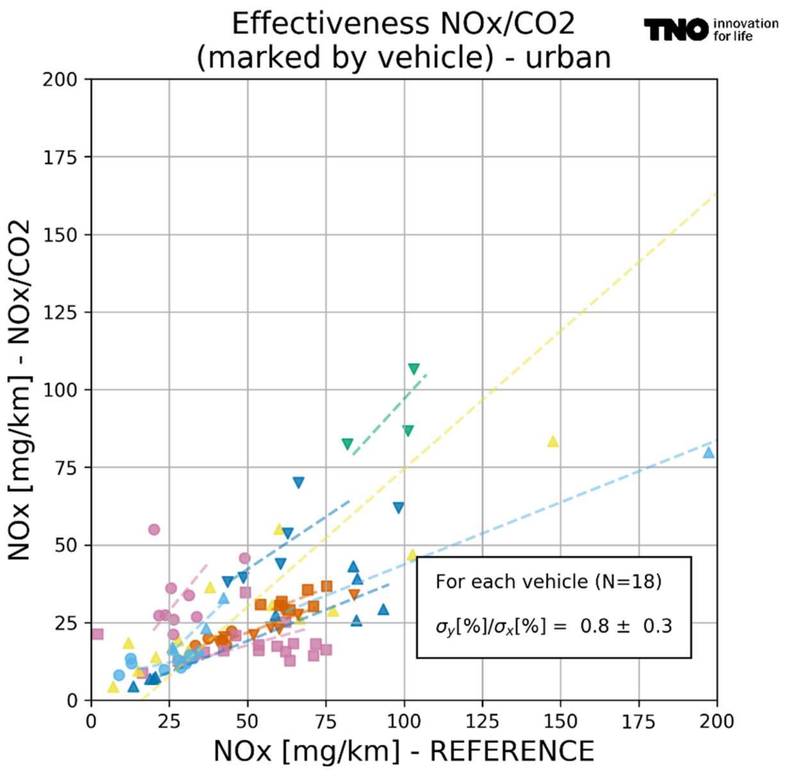

Effectiveness NOx/CO2
(marked by vehicle)
TNO innovation for life
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%] = 1.0 \pm 0.1$

B.6 ACEA

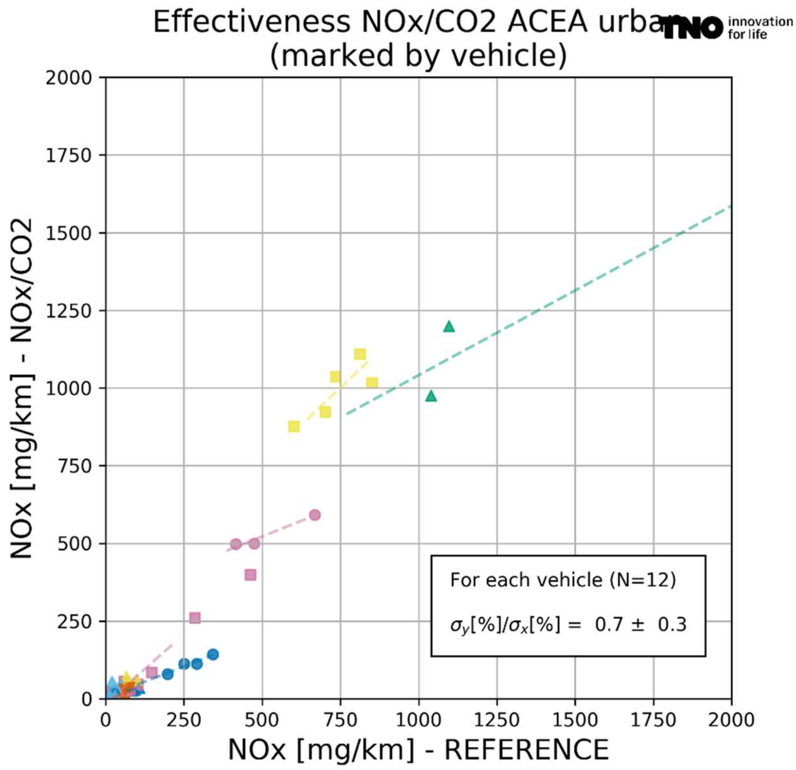

Effectiveness NOx/CO2 ACEA
(marked by vehicle)
TNO innovation for life
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
For each vehicle (N=12)
$\sigma_y[\%]/\sigma_x[\%] = 1.0 \pm 0.1$

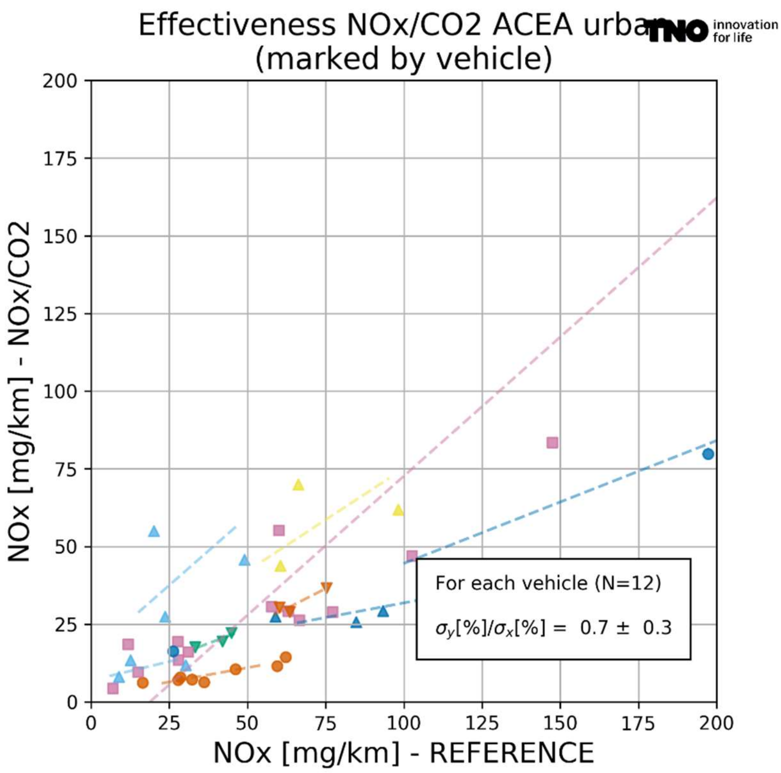

Effectiveness NOx/CO2 ACEA
(marked by vehicle)
TNO innovation for life
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
For each vehicle (N=12)
$\sigma_y[\%]/\sigma_x[\%] = 1.0 \pm 0.1$


Effectiveness NOx/CO2
(marked by vehicle) - urban
TNO innovation for life
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%] =$ 0.8 ± 0.3
0
250
500
750
1000
1250
1500
1750
2000


Effectiveness NOx/CO2
(marked by vehicle) - urban
TNO innovation for life
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
For each vehicle (N=18)
$\sigma_y[\%]/\sigma_x[\%] =$ 0.8 ± 0.3
0
25
50
75
100
125
150
175
200

B.6.1 ***U**rban*


Effectiveness NOx/CO2 ACEA urban
(marked by vehicle)
TNO innovation for life
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
For each vehicle (N=12)
$\sigma_y[\%]/\sigma_x[\%] = 0.7 \pm 0.3$
0
250
500
750
1000
1250
1500
1750
2000

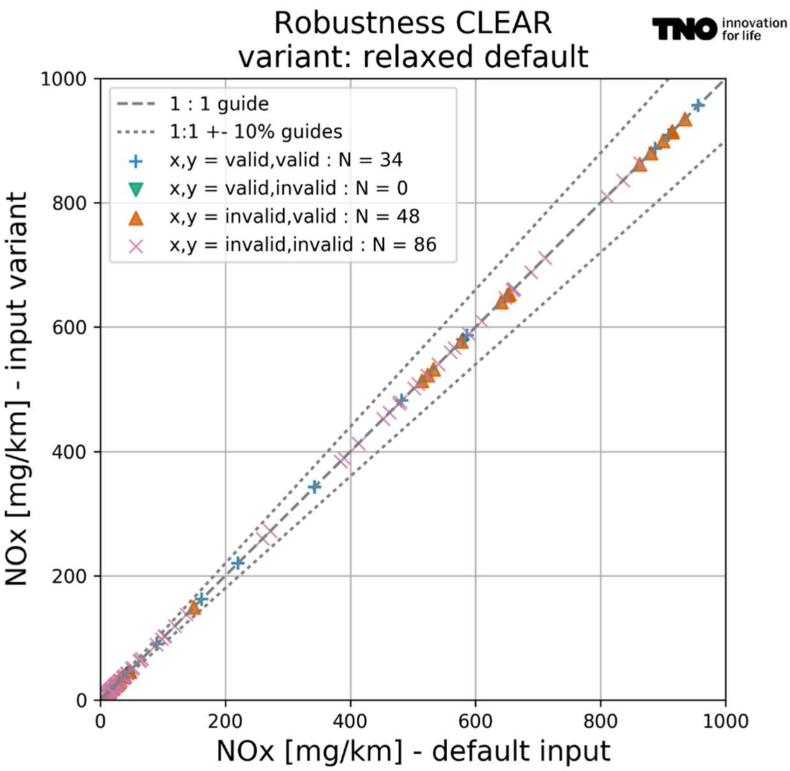

Effectiveness NOx/CO2 ACEA urban
(marked by vehicle)
TNO innovation for life
NOx [mg/km] - NOx/CO2
NOx [mg/km] - REFERENCE
For each vehicle (N=12)
$\sigma_y[\%]/\sigma_x[\%] = 0.7 \pm 0.3$
0
25
50
75
100
125
150
175
200

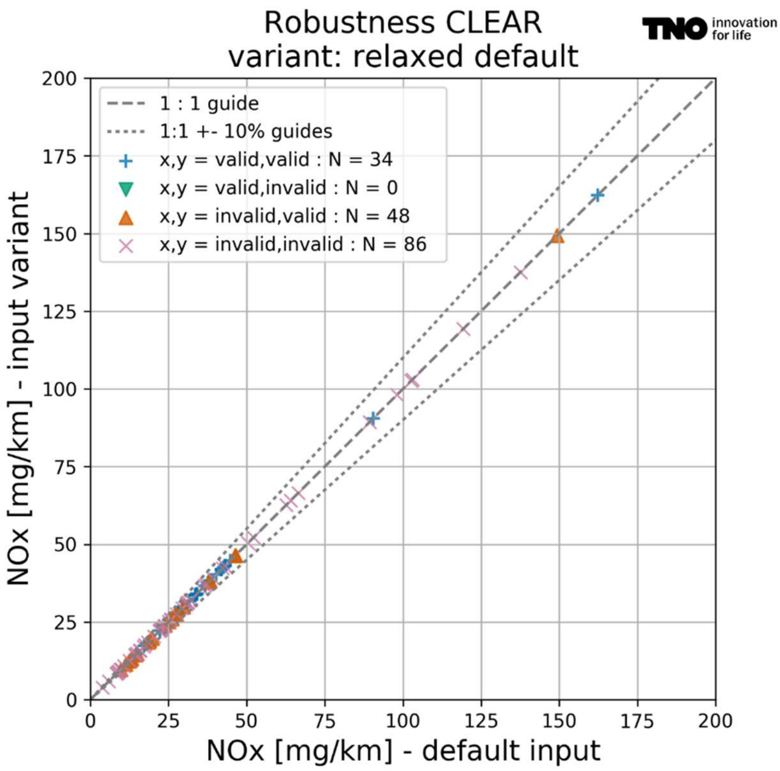
B.7 PB relaxed

Here, the robustness, effectiveness and analysis of the PB relaxed version that was implemented is presented. The PB relaxed version consisted of an increase from 60% to 65% of the P1+2 boundary threshold, and neglecting the requirement for P6 and urban P5 to have at least 5 counts in that bin.

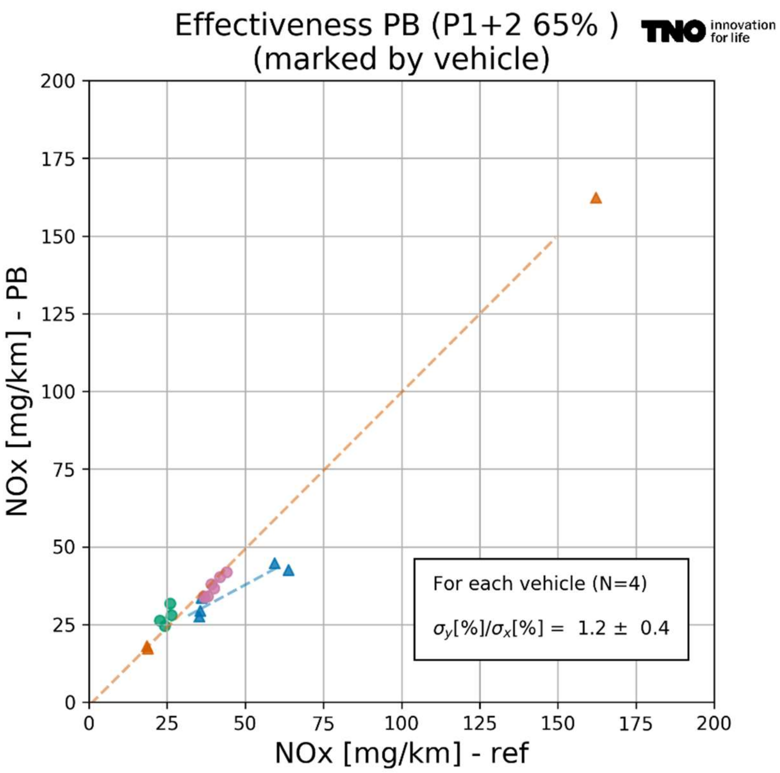
Firstly, the robustness is presented, following with the effectiveness and concluding with the comparison with raw.

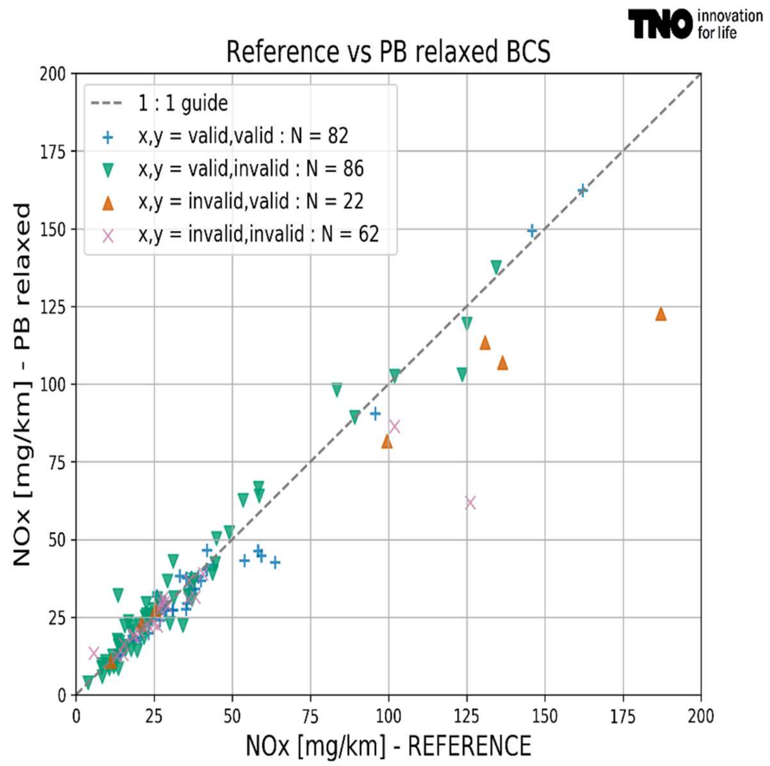
B.7.1 Robustness

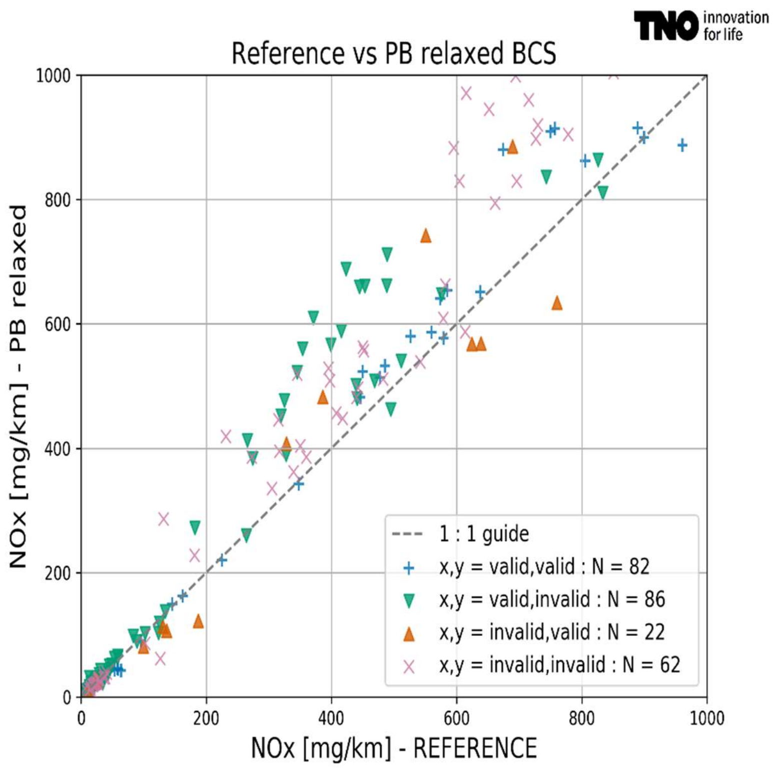




B.7.2 Effectiveness







Comparison with RAW

Note that the number of invalid trips is greatly decreased in this implementation, however still consists of 86 invalid trips.