

# Validation of simulation models in the context of railway vehicle acceptance

Oldrich Polach<sup>1</sup>, Andreas Böttcher<sup>2</sup>, Dario Vannucci<sup>3</sup>, Jürgen Sima<sup>4</sup>, Henning Schelle<sup>5</sup>, Hugues Chollet<sup>6</sup>, Gernoth Götz<sup>7</sup>, Mayi Garcia Prada<sup>8</sup>, Dirk Nicklisch<sup>9</sup>, Laura Mazzola<sup>10</sup>, Mats Berg<sup>11</sup> and Martin Osman<sup>12</sup>

Proc IMechE Part F:  
*J Rail and Rapid Transit*  
2015, Vol. 229(6) 729–754  
© IMechE 2014  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0954409714554275  
pif.sagepub.com



## Abstract

The evaluation of a reliable validation method, criteria and limit values suitable for model validation in the context of vehicle acceptance was one of the objectives of the DynoTRAIN project. The presented investigations represent a unique amount of testing, simulations, comparisons with measurements, and validation evaluations. The on-track measurements performed in four European countries included several different vehicles on a test train equipped to simultaneously record track irregularities and rail profiles. The simulations were performed using vehicle models built with the use of different simulation tools by different partners. The comparisons between simulation and measurement results were conducted for over 1000 simulations using a set of the same test sections for all vehicle models. The results were assessed by three different validation approaches: comparing values according to EN 14363; by subjective engineering judgement by project partners; and using so-called validation metrics, i.e. computable measures developed with the aim of increasing objectivity while still maintaining the level of agreement with engineering judgement. The proposed validation method uses the values computed by analogy with EN 14363 and provides validation limits that can be applied to a set of deviations between simulation and measurement values.

## Keywords

Validation, railway vehicle model, simulation, running dynamics, vehicle acceptance, certification

Date received: 22 April 2014; accepted: 20 August 2014

## Introduction

Railway vehicle acceptance is one of the significant cost and time drivers during the acquisition of railway rolling stock. Multi-body simulation tools, which are widely used in rolling stock design and development to conduct a wide range of investigations including the prediction of test results, can contribute to reduce the time and cost of the testing for the acceptance of running characteristics. Meanwhile, the reliability of simulations is becoming widely recognised and the opportunity to replace some physical tests by computer simulations has been recently considered in standards and product specifications. However, a reliable validation of the simulation model is the crucial condition when considering the application of simulations in the vehicle acceptance context.

The validation of a computer simulation model is a process of determining the degree to which the model is an accurate representation of the real world from

the perspective of the intended uses of the model.<sup>1</sup> In contrast with the verification, which is primarily dedicated to the checking of the multi-body simulation code and conducted by the code developers, model

<sup>1</sup>Bombardier Transportation (Switzerland) AG, Switzerland

<sup>2</sup>Alstom Transport Deutschland GmbH, Germany

<sup>3</sup>Ansaldobreda, Italy

<sup>4</sup>Siemens AG, Germany

<sup>5</sup>TU Berlin, Germany

<sup>6</sup>IFSTTAR, France

<sup>7</sup>Bombardier Transportation GmbH, Germany

<sup>8</sup>CAF, Spain

<sup>9</sup>DB Netz AG, Germany

<sup>10</sup>Politecnico di Milano, Italy

<sup>11</sup>KTH, Sweden

<sup>12</sup>RSSB, UK

## Corresponding author:

Oldrich Polach, Bombardier Transportation (Switzerland) AG,  
Zürcherstrasse 39, Winterthur 8400, Switzerland.

Email: oldrich.polach@ch.transport.bombardier.com

validation has to be carried out by the model developer and considers the particular model stage and the particular intended application of the model. The validation consists in comparisons with measurements to assess the quantitative accuracy of the simulation model in regard to the intended application, i.e. the simulations using the validated model. Simply said, the validation should check if the model is suitable for the intended simulations, i.e. is it 'fit for purpose'.

The comparison with measurements used for model validation should take into account all uncertainties, errors and scatter of conditions influencing both measurement as well as simulation: the errors of running dynamics measurement, the errors in the measurement of track layout and track irregularities, measurement of rail profiles and wheel profiles, as well as the scatter of the test conditions, e.g. friction coefficient between wheel and rail. The validation assessment should also take into account the number of repeated tests used for validation and their reproducibility.

The surveys dedicated to validation of railway vehicle models by Cooperrider and Law<sup>2</sup> and by Gostling and Cooperrider<sup>3</sup> are both from the advent of modern computer simulation techniques and discuss the verification of the simulation tools/software, rather than model validation. Computer simulations are widely used in the design of railway rolling stock and in research studies; however, progress in validation methodologies is rather limited. A number of publications present particular comparisons between simulation and measurement and document the validation of a particular simulation model, e.g. a validation of tramcar vehicle model,<sup>4</sup> validation of the critical speed of a vehicle as it negotiates a large radius curve<sup>5</sup> or validation of a tilting train.<sup>6</sup> However, no systematic investigations have been presented regarding a validation methodology that considers the simulation of railway vehicles. The state-of-the-art papers by Evans and Berg<sup>7</sup> from 2009 as well as Bruni et al.<sup>8</sup> from 2011 provide some hints regarding the validation of multi-body railway vehicle models.

Experience with the validation of railway vehicle models in the context of the vehicle acceptance process has been gained over many years in the UK and resulted in the Railway Group Standard Guidance Note GM/RC2641.<sup>9</sup> A vehicle model validated against stationary tests based on the protocols in GM/RC2641 can be used in the UK for the assessment of the resistance of railway vehicles to derailment based on the Railway Group Standard GM/RT2141.<sup>10</sup> This model validation method has also been incorporated as recommended practice in the European standard EN 15273-2 that deals with vehicle gauging.<sup>11</sup>

The validation experience gained by dynamics specialists in the UK has been used during the preparation of the model validation process described in UIC 518.<sup>12</sup> Furthermore, two model validation trials were conducted by this committee. The experience

with one of them dealing with the simulations of a locomotive acceptance tests is published in Jonsson et al.<sup>13</sup> The results of the second validation trial concerning a freight wagon with Y25 bogies were presented and discussed in the framework of the DynoTRAIN project.

The recent revision of prEN 14363<sup>14</sup> includes the possibility to use computer simulations under following conditions.

1. Extension of the range of test conditions where the full test programme has not been completed.
2. Approval of vehicles following modification.
3. Approval of new vehicles by comparison with an already approved reference vehicle.
4. Investigation of dynamic behaviour in the case of fault modes.

The requirements specified for the model validation in prEN 14363 originate from the investigations conducted during the preparation of UIC518 as well as from the experience gained with the use of simulations in the UK.

Unfortunately, neither UIC 518 nor prEN 14363 contain a specification of the allowable differences between simulation and on-track test results. Due to the lack of quantitative criteria, an assessment by an independent reviewer is required to ensure that the model provides a sufficient representation of reality for the intended application. To be able to replace this requirement was one of the main objectives for work package 5 (WP5) of the DynoTRAIN project.

Clear, quantitative and measurable criteria and limit values to assess the differences between simulation and measurement (also called matching error limits) in the model validation process represent a crucial requirement when applying simulations to reduce the amount of physical testing during the vehicle acceptance process. Such quantitative limits enable the specialist carrying out simulations to: understand if a particular model fulfils the validation requirements or if it needs an improvement; to visualise the model weaknesses; and to motivate the specialists to improve their model if needed. Unambiguous quantitative validation criteria and limits ensure that all vehicle models used in the vehicle acceptance context have achieved a sufficient level of quality.

The objectives of DynoTRAIN WP5 were as follows.

1. To review the state of the art of building and validation of multi-body railway vehicle models.
2. To test vehicle models by comparisons between simulations and measurements.
3. To specify the requirements for validation of vehicle models in the context of vehicle acceptance.

The DynoTRAIN WP5 investigations were structured into five tasks. The investigations started

with Task 1 dedicated to the state of the art of vehicle modelling and validation. The review of suspension and vehicle modelling was summarised in the state-of-the-art paper presented during the IAVSD Symposium in Manchester in 2011.<sup>8</sup> Questionnaires and presentations about model validation experience showed that the validation is typically carried out as a synthesis of stationary tests and on-track measurements, sometimes combined with validation of component models. Measured track irregularities and rail profiles from along the test route during the on-track tests are often not available. This missing data are usually mentioned as the reason for the observed deviations between simulations and measurements.

Task 2 of DynoTRAIN WP5 was dedicated to investigations about suspension modelling. It provided a variety of comparisons and allowed improved insight in to the modelling of suspension components (rubber components, suspension with friction, viscous dampers, and air springs); see the presentations of some of the results in Mazzola<sup>15</sup> and Mazzola and Berg.<sup>16</sup>

The experience gained in Tasks 1 and 2 was used when modelling the vehicles evaluated in the validation investigations in DynoTRAIN. The preparation of vehicle models and the identification of uncertain or unknown parameters by comparisons with stationary tests was the topic of Task 3. Tasks 4 and 5 were dedicated to validation studies and analyses, which resulted in the proposed new validation approach.

The presented investigations conducted in DynoTRAIN WP5 represent a unique body of work regarding the validation of railway vehicle models in the context of vehicle acceptance. The measurements with a test train with several different vehicle types conducted in four European countries and equipped so as to be able to simultaneously record track irregularities and rail profiles<sup>17</sup> were compared with a large set of simulations. The validation evaluations carried out in the framework of the presented investigations were performed using several vehicle models, built by seven project partners using three different simulation tools. The proposed process, the criteria and the validation limits are based on a large investigation using the state of the art in both modelling and simulation approaches.

The aim of this article is to present the proposed validation method and to explain the investigations that lead to this final proposal. The rest of this paper is structured as follows. The next section presents the tests used for evaluation, simulation models and model configurations with differing input parameters, selection of simulation input parameters and test sections selected for comparisons of simulations and measurements. The section 'Simulation output and comparisons with measurements' describes the comparisons investigated in regard to defining the

model validation approach. The section 'Evaluation of the validation method, criteria and limit values' presents the evaluations related to the selection of a suitable validation method and validation limits (matching errors). The section 'Proposed validation method' presents the proposed method, criteria and limits for validation of vehicle models used for simulations of on-track tests in the context of vehicle acceptance. The 'Discussion' section is dedicated to a discussion about the proposed validation method and about the influence of model adjustments by comparisons with stationary tests. Finally, a summary and conclusions are provided.

## Validation investigations in DynoTRAIN

### *On-track tests used for validation*

The presented model validation investigations used on-track measurements conducted in the framework of DynoTRAIN WP1 as well as some measurement results provided by project partners.

The DynoTRAIN test campaign was conducted in October 2010 with several different vehicles that were equipped with 10 force measuring wheelsets and several acceleration and displacement sensors.<sup>17</sup> The train travelled for a total of 20 days of test runs through Germany, France, Italy and Switzerland at speeds up to 120 km/h with freight wagons connected and up to 200 km/h without freight wagons. A measuring vehicle integrated into the test train continuously recorded the track irregularities and rail profile shapes along all test runs. The test train contained the following vehicles:

- locomotive DB BR 120;
- DB passenger coach Bim;
- empty freight wagon Sgns with Y25 bogies;
- loaded freight wagon Sgns with Y25 bogies;
- Laas freight vehicle unit consisting of two two-axle wagons with UIC link suspension; one empty and one fully loaded; the empty wagon was equipped with measuring wheelsets.

In addition to the vehicles tested in DynoTRAIN, another two vehicles were investigated using measurements carried out during the running dynamic acceptance tests of these vehicles:

- the High-speed EMU for TCDD (Turkey) manufactured by CAF, measurements conducted in 2008;
- DMU IC4 for DSB (Denmark) manufactured by AnsaldoBreda, measurements carried out in 2006.

The uncertainty and error of the measurements used in the described investigations represent the state of the art in the vehicle approval process. There were no investigations in DynoTRAIN WP5 dedicated to

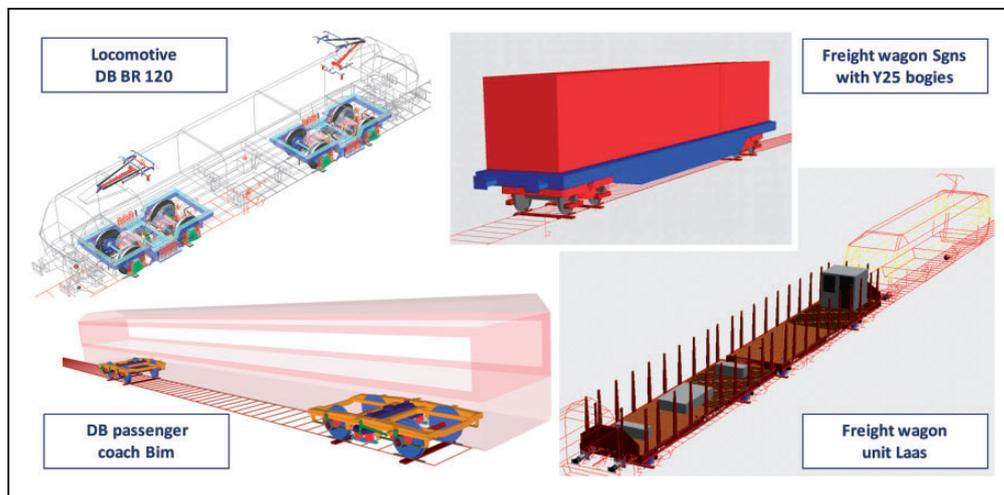
uncertainty of the measured data used for model validation.

### Vehicle models and model configurations

The multi-body vehicle models used for the evaluation of the validation method were prepared by project partners using different simulation tools, see examples of models built using the simulation tool Simpack in Figure 1. Several versions of each vehicle model were prepared using different stages of model parameters, track irregularities, rail and wheel profiles as well as modelling depth. The differing model versions are called ‘model configurations’ in the rest of this paper. An overview of the vehicle models used in the presented investigations is shown in Table 1. The originally proposed set of model configurations exceeded the available time and project budget.

Moreover, some model configurations were not feasible for some vehicles, e.g. if the measurements of track irregularities or rail profiles were not available. These facts resulted in a large variation in the number of vehicle model configurations as can be seen in Table 1.

The vehicle models used in the investigations represent fully nonlinear three-dimensional models, as this is the state of the art in railway engineering and research. Rigid bodies representing the vehicle body, bogie frame, wheelset, axle box, etc. are connected by springs, dampers, friction elements and bump-stops that model the suspension components. Damper models consist of a dashpot together with series stiffness. The nonlinear wheel/rail contact models use the respective contact evaluation method and the respective version of Kalker’s computer code Fastsim implemented in the utilised simulation tool.



**Figure 1.** Examples of multi-body models of the vehicles tested in DynoTRAIN.

**Table 1.** Overview of the multi-body simulation models used for the evaluation of the presented validation methodology.

Vehicle	Project partner	Simulation tool	Number of model configurations	On-track tests used for validation
Locomotive DB BR 120	Siemens	Simpack	24	DynoTRAIN
	IFSTTAR	VOCO	4	DynoTRAIN
DB passenger coach Bim	Bombardier Transportation	Simpack	13	DynoTRAIN
	IFSTTAR	VOCO	4	DynoTRAIN
Freight wagon Sgns, empty	Technical University Berlin	Simpack	8	DynoTRAIN
	IFSTTAR	VOCO	6	DynoTRAIN
Freight wagon Sgns, laden	Technical University Berlin	Simpack	7	DynoTRAIN
Laas freight vehicle, empty	Alstom	Simpack	5	DynoTRAIN
High-speed EMU (TCDD)	CAF	SIDIVE	3	Provided by vehicle manufacturer CAF
DMU IC4, coach T3 (DSB)	Ansaldobreda	Simpack	2	Provided by vehicle manufacturer Ansaldobreda
DMU IC4, coach M4C (DSB)	Ansaldobreda	Simpack	2	Provided by vehicle manufacturer Ansaldobreda

The vehicle models were prepared under the partners' responsibility. The majority of data regarding the vehicles tested during the DynoTRAIN test campaign was provided by DB; the remaining information was estimated or identified from archive material by the partner modelling a particular vehicle. The identification of vehicle model parameters of vehicles tested outside the DynoTRAIN project was fully the responsibility of the respective partner; vehicle manufacturers also provided data obtained in their running tests.

The initial vehicle models were prepared using the available vehicle data without considering the results of stationary tests. Project partners were, however, advised to adjust the mass parameters in their model before starting any comparisons in order to achieve a good agreement between the wheel loads obtained from a static model and the wheel loads measured during the on-track tests. Then, the initial models were adjusted with the aim of improving the agreement between the on-track test results and the simulation results, so that several differing configurations of the same model could be compared. The vehicle models adjusted based on the comparisons with the stationary tests represent other model configurations. In order to assess the effect of using actual measured infrastructure parameters such as track layout, track irregularities and rail profiles, the model configurations with estimated rail profiles (see explanation in the section 'Rail profiles') and estimated track irregularities (see explanation in the section 'Track layout and track irregularities') were also prepared and compared with the on-track measurements.

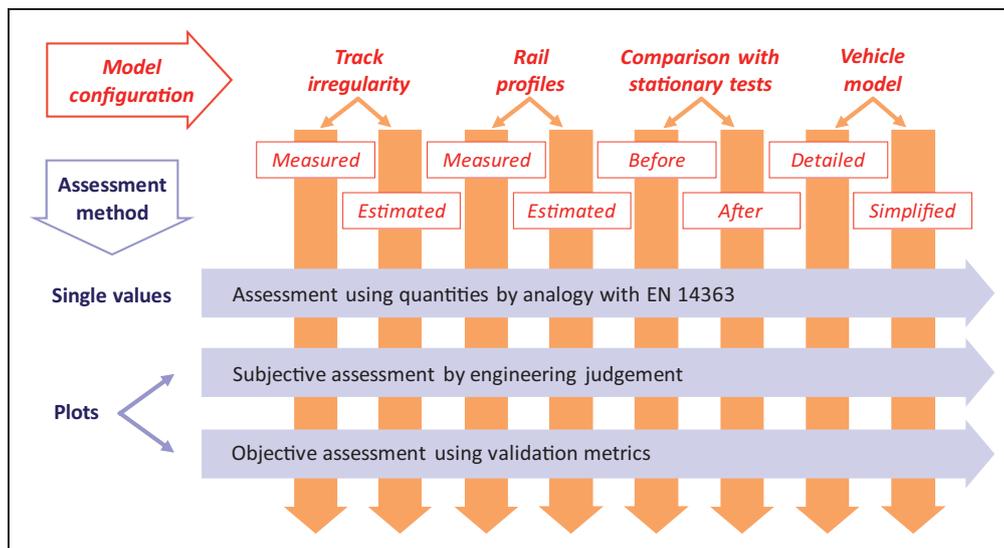
A total of 78 model configurations were investigated, with differing levels of knowledge on vehicle data, input parameters regarding the infrastructure, different usage of stationary tests and applying a

different depth of modelling detail. Moreover, some model configurations of the locomotive BR 120 created by Siemens were varied in implementing the driving torque in test sections where this locomotive was used as a propelling vehicle. Figure 2 shows the variety of investigated model configurations together with the assessed quantities, which are described in more detail in the section 'Simulation output and comparisons with measurements'. The effect of using the results of stationary tests for the model validation in regard to the simulation of the on-track tests, which was investigated by comparing the simulations of the on-track tests using vehicle models before and after the comparisons with the stationary tests, is discussed in the section 'Effect of model adjustment using stationary tests on the simulation of on-track tests'. The effects of measured and estimated wheel and rail profiles, as well as track irregularities data, on the model validation results are not presented in this paper for the sake of brevity; readers interested in those topics are referred to Polach and Bottcher.<sup>18</sup>

**Simulation input parameters**

**Track layout and track irregularities.** The track geometry data were measured during the DynoTRAIN test campaign performed by the DB track recording car 'RAILab I'.<sup>17</sup> The data were obtained at a sampling distance of 0.16 m and stored in binary files.

The manipulation of measured track irregularity data into a format suitable for simulations was performed by DB Netz AG. As the inertial-platform-based RAILab I system uses a special filter algorithm to separate long wavelengths caused by the track layout from the track irregularities to be assessed, the recorded data were de-coloured (transformed



**Figure 2.** Overview of the model configurations and assessment methods evaluated in the framework of the presented investigations.

backward) using corrective filters before they were used in the vehicle dynamics simulations.

For each of the selected track sections, the relevant RAILab I data were transformed into the format used in the multi-body simulation package Simpack. Two input data files were created for each track section; one of them containing the track layout (curvature and cant using high-pass filters above 70 m) and the second describing irregularities (lateral and vertical position of the left and right rails with band-pass filters between 1 and 70 m).

There were no measurements of track irregularities available for the on-track measurements conducted outside of the DynoTRAIN remit. Thus, the simulations with vehicle models DMU IC4 and High-speed EMU Turkey were carried out using estimated track irregularities. This estimated track irregularity data was used not only in the case of missing measured data but also for comparisons regarding the importance of knowledge about the track irregularities. In the following discussions the term 'estimated track irregularities' means either generated data based on the power spectral density as in ORE B176<sup>19</sup> or measured track irregularities from other measurements. The selection of track irregularities to be used instead of the actual measured data was the responsibility of the respective partner.

**Rail profiles.** The rail profiles were measured during the DynoTRAIN test runs by means of an optical measuring device<sup>17</sup> and recorded at a spacing interval of 0.25 m. For the synchronisation of the measured rail profiles with all the other measuring channels, the time stamp and counter signal provided by the track recording car RAILab I was combined with the odometer signal of the rail profile measuring system and both were stored together in an additional synchronisation file.

The implementation of the measured rail profile in multi-body simulations generates several questions. A typical recommendation is to use a 'representative profile'. However, how do you identify this representative profile? As the rail profiles in curves wear differently on the outer and inner rails as well as in a different manner from straight and curved transitions, the use of one profile for each rail along the whole investigated section will obviously provide incorrect results either outside the full curve or in the full curve, unless there is no wear of rails (new or newly ground rails).

Continuously varying rail profile along the track section has been implemented in some of the simulation packages; however, it is still not a state-of-the-art procedure and thus not applied in this paper. After several investigations and discussions regarding this topic, it was finally agreed to calculate averaged rail profiles from the measured rail profiles of the part of the actual track section with constant track curvature (i.e. one profile for the left rail and one for the right

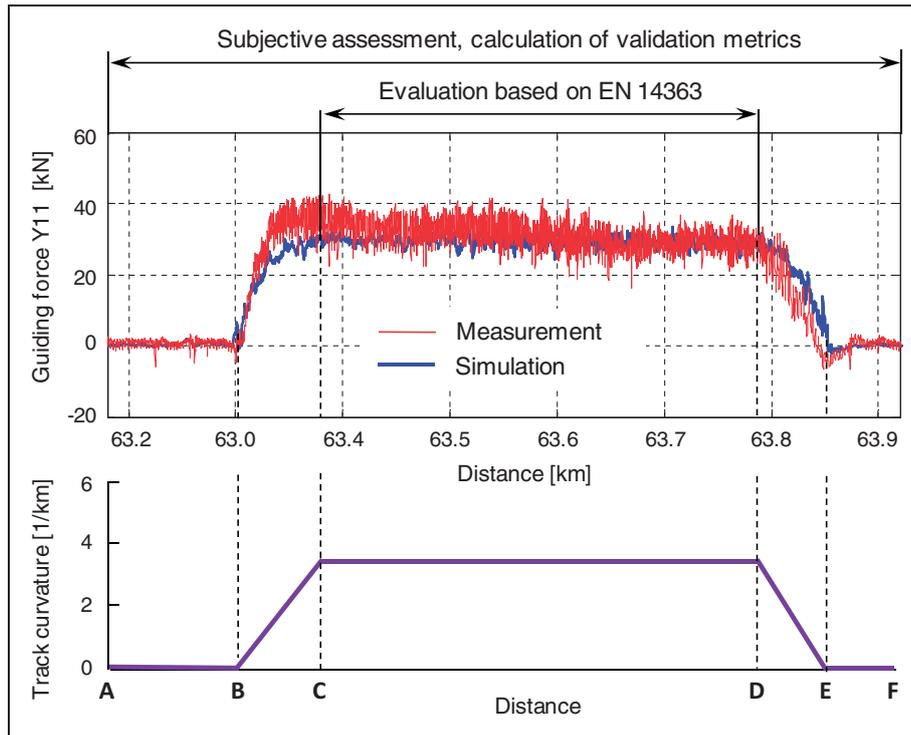
rail) and to use these averaged profiles for simulations of this particular track section. Thus, the used profile may be incorrect in curved transitions and accompanying straight track parts. Moreover, if the actual rail profile changes along the distance, e.g. in some longer sections, the applied averaged rail profiles may not be fully representative.

The preparation of the averaged rail profiles was performed by DB Netz AG. At first the profiles were smoothed and their running surfaces (down to an appropriate profile gradient) were approximated by high-order polynomials. Then all profiles of the same rail within the respective track section were vertically aligned to each other at the rail top and laterally at the gauge measuring point (14 mm below the top of the rail). In order to allow for superposition of measured track irregularities, the resulting rail profiles were shifted in the lateral direction to meet the 1435 mm nominal track gauge.

For each simulation exercise a mean profile for the left rail and a mean profile for the right rail were provided by taking into account all rail profiles of the track section with a constant radius, i.e. section C–D in Figure 3. These single mean profiles for left and right rails were used in simulations of the complete particular section. The model configurations with 'estimated rail profiles' used the nominal rail profile and rail inclination of the particular country under investigation. The simulations of vehicle models DMU IC4 and High-speed EMU Turkey both solely used the respective nominal rail profiles and rail inclinations of the particular country as there were no measurements of rail profiles available.

**Wheel profiles.** The wheel profiles of vehicles tested in DynoTRAIN were measured before and after the test campaign and the measured data were used in the simulations. The details regarding the wheel profile implementation were the individual partner's responsibility. The model configurations with 'estimated wheel profiles' were carried out using the designed wheel profile S1002. There were no measurements of wheel profiles available for the on-track measurements conducted outside of the DynoTRAIN project. Hence, the vehicle models DMU IC4 and High-speed EMU Turkey used the respective nominal wheel profiles: profile S1002 (DMU IC4) or profile GV 1/40 (High-speed EMU Turkey), respectively.

**Friction coefficient between wheel and rail.** The value of the friction coefficient between the wheel and rail represents an uncertain input parameter in the simulation. The selection of this parameter was the responsibility of the partner carrying out the simulation. All test runs selected for validation from the DynoTRAIN measurements were carried out on a dry rail. In their simulations each partner used a wheel/rail friction coefficient of 0.45 or 0.50 to represent those conditions. A few model configurations used a lower



**Figure 3.** Example of validation exercise with the specification of track sections used for different kind of assessments.

friction coefficient than 0.45 or higher than 0.50, respectively, with the aim of testing for an improvement of the agreement with the measured values. The majority of simulations used an identical and constant value of the friction coefficient in the tread and on the flange; only a few simulations used a lower friction coefficient on the flange.

The simulations of the test results provided by vehicle manufacturers used a value of wheel/rail friction coefficient of 0.45 (simulations of DMU IC4 by AnsaldoBreda) or a friction coefficient of 0.35 (simulations of High-speed EMU Turkey by CAF), respectively.

### Validation exercises

Comparisons between simulation and measurement results were carried out for all vehicle models and model configurations under the same conditions and in the same manner as for selected representative sections of test runs, called validation exercises. One validation exercise consisted of one curve passing scenario including both transitions and parts of straight track as shown in Figure 3. In this context the word ‘section’ means a part of the track; it does not mean section as in the definition in EN 14363.<sup>20</sup>

A total of 17 validation exercises were selected, representing all four track zones in EN 14363: straight track and very-large-radius curves were represented by four sections, large radius curves ( $R > 600$  m) by two sections; five sections were used for small radius curves ( $400 \text{ m} \leq R \leq 600 \text{ m}$ ) and six for very-small-radius curves ( $250 \text{ m} \leq R < 400 \text{ m}$ ). Table 2 shows

the parameters of the test sections selected for vehicles tested in DynoTRAIN in terms of the location, track layout, section length as well as the speed of the test train in the respective section. It should be noted that the number of test sections in each test zone based on EN 14363 reported in this article do not fully comply with the final recommended validation procedure, because the procedure and the conditions to be used were not known at the start of the investigations. Moreover, the test conditions during the DynoTRAIN running tests did not fully comply with EN 14363; see Zacher and Kratochville.<sup>17</sup>

The selection of test sections considered geometrical track quality (irregularities) and wheel/rail contact geometry with the aim of including varying conditions. The track sections for exercises 2, 3 and 5 were included due to a high vertical disturbance in the track irregularities. The properties of the wheel/rail contact geometry were assessed by the calculation of the equivalent conicity and radial steering index over the constant curvature sections using the measured rail profiles averaged over a 100 m distance together with a nominal design wheel profile S1002 and mean track gauge over the respective track section. The definition of a radial steering index was introduced in UIC 518<sup>12</sup> to assess the available rolling radius difference between left and right wheels. Index values lower than a value of one represent a contact geometry that provides a sufficient difference in rolling radius for self-steering wheelsets, whereas values higher than one represent an insufficient rolling radius difference for the considered curve radius. The curve test sections 4, 5, 7, 9, 14 and 15 show a radial steering

**Table 2.** Test runs and parameters of track sections used in the validation exercises performed with vehicles tested in DynoTRAIN.

Exercise number	Line	Country	Test zone according to EN 14363	Curve radius (m)	Cant (mm)	Speed (km/h)	Section length: whole section A–F/constant curvature section C–D (m)
1	Geislingen–Westerstetten	Germany	4	282	120	68	740/400
2	Geislingen–Westerstetten	Germany	4	312	100	68	280/140
3	Geislingen–Westerstetten	Germany	3	572	155	110	1080/320
4	Uffenheim–Ansbach	Germany	3	580	150	110	870/490
5	Uffenheim–Ansbach	Germany	3	581	110	110	1130/680
6	Uffenheim–Ansbach	Germany	2	864	115	120	750/360
7	Uffenheim–Ansbach	Germany	2	694	160	121	690/190
8	Uffenheim–Ansbach	Germany	1	$\infty$	0	120	1760/1760
9	Würzburg–Fulda	Germany	1	5600/6000	75	200	3300/2644
10	Lichtenfels–Bamberg	Germany	1	$\infty$	0	160	3200/3200
11	Lichtenfels–Bamberg	Germany	1	$\infty$	0	160	3200/3200
12	Pisa–Firenze	Italy	4	295	140	76	504/110
13	Pisa–Firenze	Italy	4	292	140	76	968/771
14	Biasca–Göschenen	Switzerland	4	278	150	74	424/280
15	Biasca–Göschenen	Switzerland	4	294	142	74	384/192
16	St. Giovanni–Firenze	Italy	3	442	140	90	510/250
17	St. Giovanni–Firenze	Italy	3	406	150	90	651/426

index below one and thus a good contact geometry regarding curving, whereas sections 1, 2, 3, 6, 12, 13, 16 and 17 give a radial steering index higher than one, i.e. disadvantageous contact geometry conditions regarding self-steering of wheelsets. The equivalent conicities (calculated for a lateral displacement of the wheelset of 3 mm) in section 8 were medium values between 0.20 and 0.25 and in section 9 their values were around 0.1. The sections 10 and 11 were selected because of the occurrence of very high conicities; the conicity calculated per 100 m distance varied from medium values up to a few very high values of around one.

As freight vehicles were included in the test train only at speeds up to 120 km/h, the Laas wagon and the Sgns freight wagons were missing in the runs of the exercises 9, 10 and 11. Each simulation was performed for a part of the test run called ‘part of interest’ (A–F in Figure 3) and some outputs were evaluated over this part, whereas other outputs were solely evaluated over the part of the track with constant curve radius (C–D in Figure 3).

## Simulation output and comparisons with measurements

### Introduction

The simulations of selected on-track tests were evaluated in the same manner by all partners conducting simulations. This required an agreement and specification of the output data and its format.

As the aim of the validation is the application of simulation for vehicle acceptance, a comparison of quantities as they are measured and evaluated according to EN 14363<sup>20</sup> was logically considered as one possible assessment method. Another typical validation assessment is a judgement of the comparison between the time domain signals from simulations and measurements. In contradiction with the quantities based on EN 14363, which are assessed primarily in track sections with a constant curvature, the judgement of time diagrams allows the assessment of the behaviour in transitions as well as the frequency content of the signals. A subjective judgement of time or distance diagrams thus represents another kind of assessment.

However, an engineering judgement is not measurable; the replacement of such an assessment by calculable quantitative criteria is highly preferred. The evaluation of so-called validation metrics conducted recently by the Transportation Technology Center<sup>21</sup> motivated the DynoTRAIN project partners to include the evaluation of the validation metrics as the third kind of assessment. These three kinds of validation assessment were applied to the investigated vehicle models and model configurations as shown schematically in Figure 2. The definition of these assessments and agreed simulation outputs are presented in the following sections.

### Assessment using values based on EN 14363

The comparisons between simulations and measurements were performed using an agreed set of output

**Table 3.** Output quantities used for the assessment by analogy with EN 14363.

Quantity		Filtering	Processing	Notation	Unit
Wheel/rail forces, quasi-static values	Guiding force	Low-pass filter 20 Hz	50th percentile (median)	$Y_{qst}$	kN
	Wheel load	Low-pass filter 20 Hz	50th percentile (median)	$Q_{qst}$	kN
	Ratio $Y/Q$	Low-pass filter 20 Hz	50th percentile (median)	$Y/Q_{qst}$	–
	Sum of guiding forces	Low-pass filter 20 Hz	50th percentile (median)	$\Sigma Y_{qst}$	kN
Wheel/rail forces, dynamic values	Guiding force	Low-pass filter 20 Hz	0.15 percentile, 99.85 percentile	$Y_{max}$	kN
	Wheel load	Low-pass filter 20 Hz	99.85 percentile	$Q_{max}$	kN
	Ratio $Y/Q$	Low-pass filter 20 Hz	Sliding mean (window 2 m, step 0.5 m) 0.15 and 99.85 percentile	$Y/Q_{max}$	–
	Sum of guiding forces	Low-pass filter 20 Hz	Sliding mean (window 2 m, step 0.5 m) 0.15 and 99.85 percentile	$\Sigma Y_{max}$	kN
Bogie frame acceleration, root-mean-square (RMS) values	Lateral acceleration	Band-pass filter 0.4 to 10 Hz	RMS value	$\ddot{y}_{rms}^+$	$m/s^2$
	Vertical acceleration	Band-pass filter 0.4 to 10 Hz	RMS value	$\ddot{z}_{rms}^+$	$m/s^2$
Bogie frame acceleration, dynamic values	Lateral acceleration	Low-pass filter 10 Hz	0.15 percentile, 99.85 percentile	$\ddot{y}_{max}^+$	$m/s^2$
	Vertical acceleration	Low-pass filter 10 Hz	0.15 percentile, 99.85 percentile	$\ddot{z}_{max}^+$	$m/s^2$
Car body acceleration, RMS values	Lateral acceleration	Band-pass filter 0.4 to 10 Hz	RMS-value	$\ddot{y}_{rms}^*$	$m/s^2$
	Vertical acceleration	Band-pass filter 0.4 to 10 Hz	RMS-value	$\ddot{z}_{rms}^*$	$m/s^2$
Car body acceleration, dynamic values	Lateral acceleration	Band-pass filter 0.4 to 10 Hz	0.15 percentile, 99.85 percentile	$\ddot{y}_{max}^*$	$m/s^2$
	Vertical acceleration	Band-pass filter 0.4 to 10 Hz	0.15 percentile, 99.85 percentile	$\ddot{z}_{max}^*$	$m/s^2$

quantities that are used in testing based on EN 14363. The simulation and measurement results were filtered and processed by analogy with the requirements in EN 14363 and compared against each other; this evaluation considers the part of the track with a constant curvature, i.e. section C–D in Figure 3. Table 3 shows the list of output quantities, their filtering, processing as well as the nomenclature and unit. A total of two wheelsets were used in the validation assessment of each vehicle, which resulted in a total of 28 parameters related to wheel/rail forces. The bogie accelerations were measured on the bogie frame above all the wheelsets in the lateral direction and above the wheelsets of one bogie in the vertical direction, resulting in a total of 12 bogie frame acceleration values (not applicable for the two-axle wagon). The vehicle body accelerations were measured at the floor level above both bogie centre pins in the lateral and vertical directions resulting in a total of eight car body acceleration values. Thus, a total of 48 parameters per model configuration and test section (36 for the two-axle wagon) were applied consisting of quasi-static as well as dynamic wheel/rail forces and vehicle body and bogie frame accelerations.

### Subjective assessments

A subjective engineering judgement is based on a visual impression of time history plots and power spectral density (PSD) diagrams. A selected set of quantities consisting of 20 plots per vehicle model configuration and test section (for all vehicles apart

from the Laas freight vehicle that had a lower number of plots) was issued and provided to project partners for the assessment.

The following quantities were displayed and issued in the form of distance or time plots:

- lateral wheel/rail forces ( $Y$ -forces): four diagrams per vehicle model configuration and test section;
- vertical wheel/rail forces  $Q$  (wheel loads): four diagrams;
- ratio  $Y/Q$ : four diagrams;
- lateral accelerations of the bogie frame above wheelsets 1 and 2: two diagrams;
- vertical acceleration of the car body above bogie 1: one diagram;
- lateral acceleration of the car body above bogie 1: one diagram.

The simulation as well as measurement signals were filtered using a 20 Hz low-pass filter, without any other processing, and displayed for the whole investigated test section (section A–F in Figure 3).

Moreover, PSDs of four acceleration signals were also provided as diagrams for subjective assessments:

- lateral acceleration of the bogie frame above wheelset 1;
- vertical acceleration of the bogie frame above wheelset 1;
- vertical acceleration of the car body above bogie 1;
- lateral acceleration of the car body above bogie 1.

These signals were filtered by a 20 Hz low-pass filter for PSDs in the frequency range 0 – 10 Hz.

The project partners were asked to assess the diagrams displaying the comparison of the measurement and simulation signal quantities by a simple binary assessment 'Yes/No'. Assessing a diagram with a Yes means that for the displayed signal quantities of the particular diagram the assessor considers the model as validated and vice versa.

As the form of the diagram (size, number of compared curves, scaling of axis, colours, position of curve in front or background, respectively) can influence the result of this judgement, it was first necessary to select and agree on a suitable form for the diagrams. It was decided to present only two curves in each diagram, comparing measurement and simulation of a quantity's distance or time history. The selection of the scaling of the vertical axis turned out to be a more difficult question. Figures 4 to 7 show examples of comparisons between simulation and measurement data for the following four investigated vehicle models:

- the locomotive DB BR 120 investigated by Siemens;
- the DB passenger coach Bim investigated by Bombardier Transportation;
- the loaded freight wagon Sgns investigated by the Technical University Berlin;
- the Laas freight vehicle investigated by Alstom.

Figure 4 presents the guiding force on the outer wheel of the leading wheelset obtained for test section 1 (curve radius 282 m) using the same scale for all vehicles to illustrate the differences in the level of the investigated values. As can be seen, the position of the signal is not exactly the same in regard to the distance. This may lead to slight differences when calculating the values in the specified interval with a constant curvature. Other effects can be observed, such as a signal offset (locomotive model created by Siemens). For illustration purposes, the same results are displayed in Figure 5 in the original form as submitted for the subjective assessment, together with the percentage of positive assessments by project partners. The diagrams were assessed by 10 partners, i.e. 40% means that four of the 10 partners considered the presented results as documenting a validated model. In addition to a differing scale on the vertical axis, the Laas freight vehicle results are displayed as a time diagram over a longer interval compared with the other vehicles that are presented as distance diagrams. It can be seen that the assessment of the very light two-axle wagon Laas is rather strict compared with the results of the locomotive or loaded freight wagon. Figure 6 shows the ratio  $Y/Q$  at the outer wheel of the leading wheelset for test section 2 (curve radius 312 m) and Figure 7 shows the vertical car body acceleration from test section 8 (straight). Although the  $Y/Q$  ratio

has a similar level for all vehicles, the accelerations significantly vary. This opens the question of the selection of scaling for the presentation of results. When using an equal scaling, the comparison for light vehicles can barely be assessed as they have low vertical as well as lateral wheel/rail forces. Also, the assessment of the acceleration of soft-suspended vehicles is difficult. Alternatively, the use of automatically adjusted scaling leads to the impression of large differences, even if the values are very small. To allow better assessments, it was proposed to the project partners to use a fixed scaling with one of three specified scale groups; however, the final decision was up to the partner conducting the simulation. Consequently, the values presented in the evaluated diagrams are sometimes rather small, whereas in other cases the peaks are outside of the diagram; both effects make the subjective assessment more difficult.

### Validation metrics

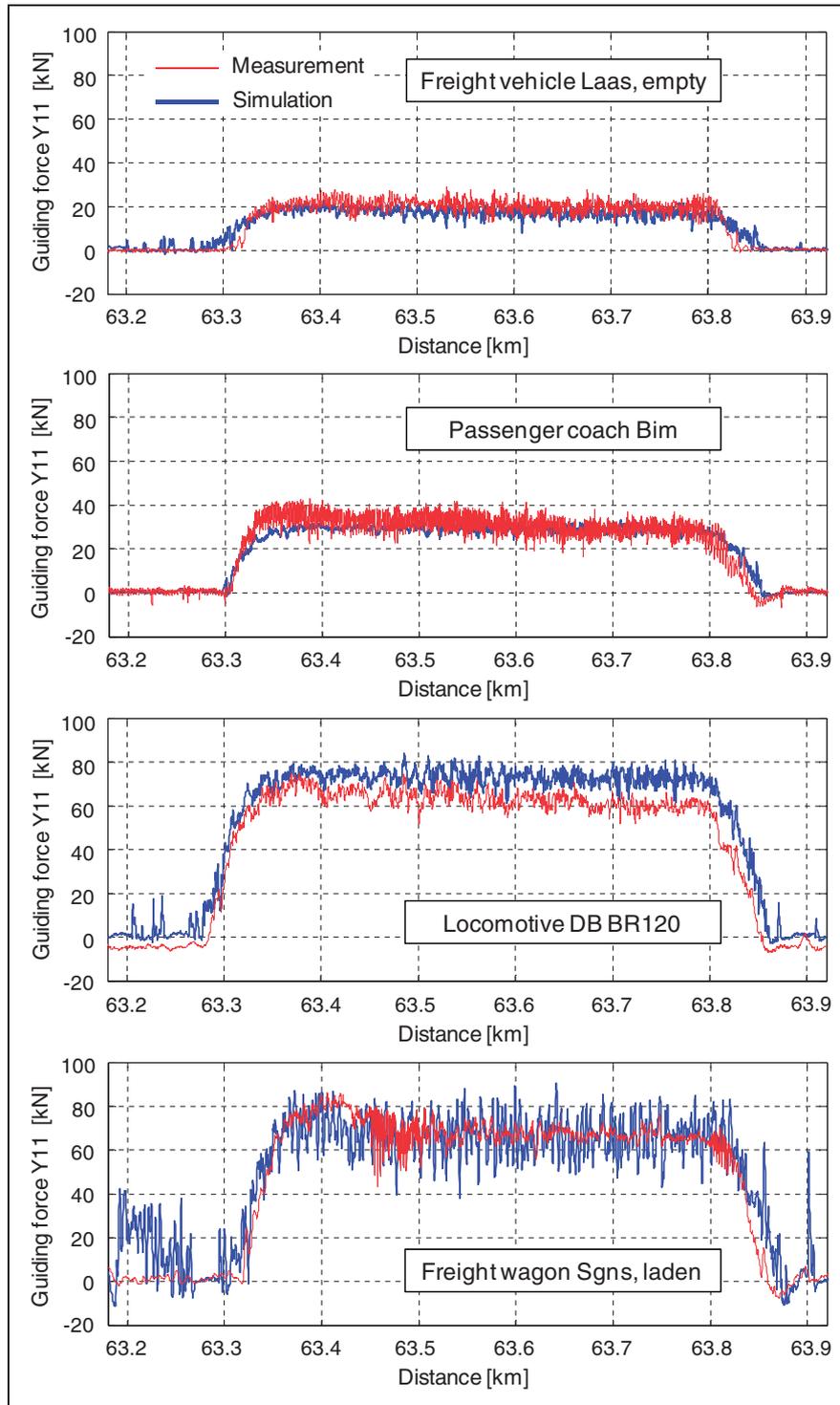
A validation assessment in terms of a comparison of time histories between simulated and measured values generates questions about the subjectivity of this assessment as stated in the previous section. Validation metrics represent an approach to quantifying the comparisons of time history curves with the intent of minimising the subjectivity while still maintaining a correlation with experts' opinion.<sup>22</sup> They are developed and mainly used for comparisons between simulation and measurement in the context of model validation.

A possible metric that could be used to compare the time domain diagrams is the integral approach introduced in 1984 by Geers. Integrals of two wave forms to be compared are computed and used to evaluate the difference in the magnitude and phase of the wave forms expressed in terms of magnitude, phase and comprehensive error factors, with small values of the error factor representing good agreement. The magnitude as well as phase form of the error factors was later adapted by Russell.<sup>23</sup> The new phase form by Russell was combined with the 1984 Geers' metric by Sprague and Geers.<sup>24</sup> By using the same sampling rate and the same length of time or distance interval for the compared measurement and simulation signals, the definitions of error factors proposed in Sprague and Geers<sup>24</sup> can be expressed by the following formulas.<sup>25</sup>

Sprague and Geers magnitude error factor

$$M_{SG} = \sqrt{\frac{\sum_{i=1}^n c_i^2}{\sum_{i=1}^n m_i^2}} - 1 \quad (1)$$

where the  $c_i$  are the simulated values and  $m_i$  are the measured values.



**Figure 4.** Validation examples: guiding force on the outer wheel of leading wheelset, exercise I, Germany, Geislingen–Westerstetten line, curve radius 282 m, cant 120 mm, speed 68 km/h.

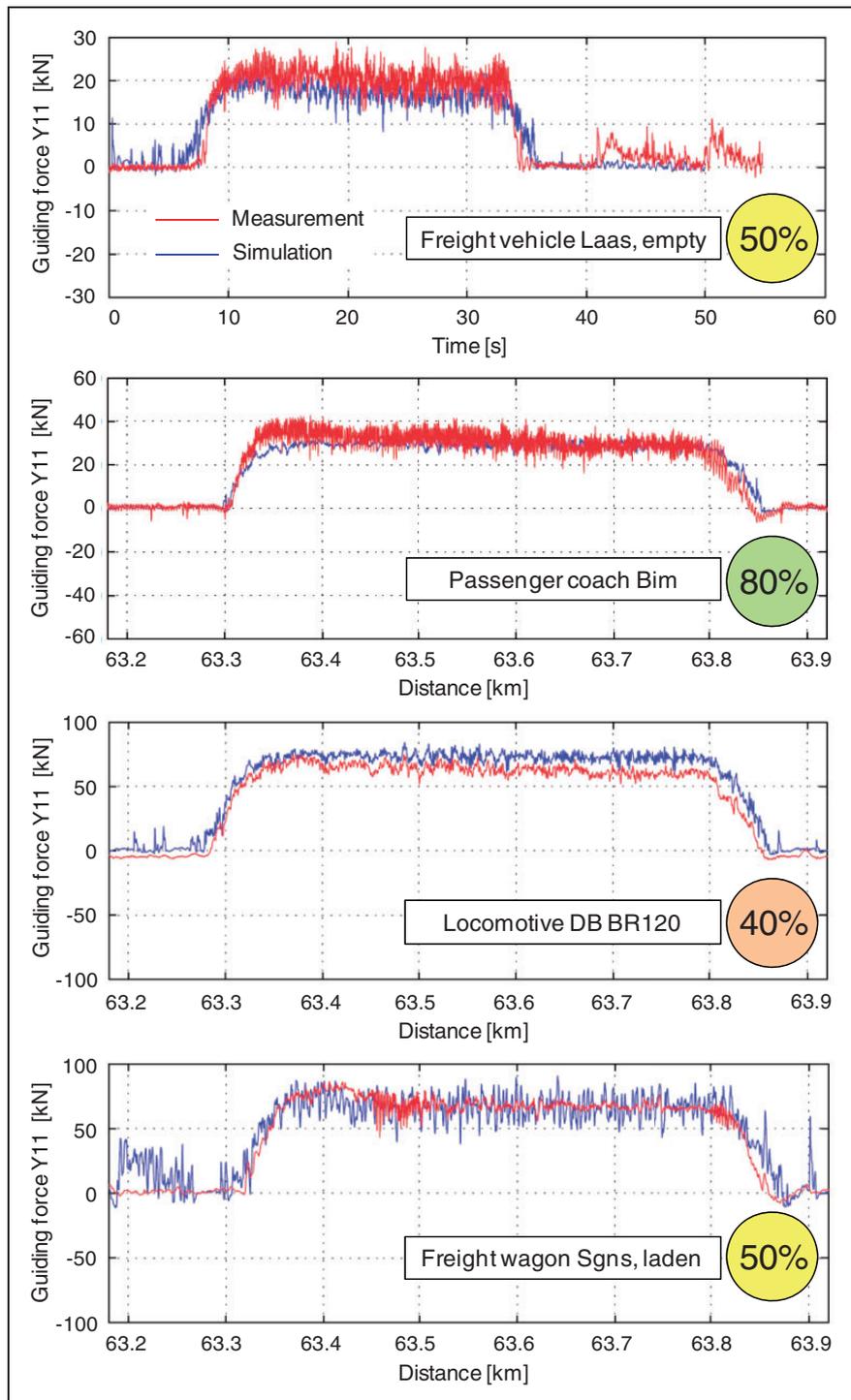
Sprague and Geers phase error factor

$$P_{SG} = \frac{1}{\pi} \cos^{-1} \left( \frac{\sum_{i=1}^n c_i m_i}{\sqrt{\sum_{i=1}^n c_i^2 \sum_{i=1}^n m_i^2}} \right) \quad (2)$$

Sprague and Geers comprehensive error factor

$$C_{SG} = \sqrt{M_{SG}^2 + P_{SG}^2} \quad (3)$$

The error factors of the validation metrics proposed by Sprague and Geers and by Russell were calculated by the project partners to allow comparisons between simulations and measurements provided in the time and distance domain plots and used for the subjective assessment by the partners. The evaluations later focussed on the validation metric by Sprague and Geers which appeared to be more promising.



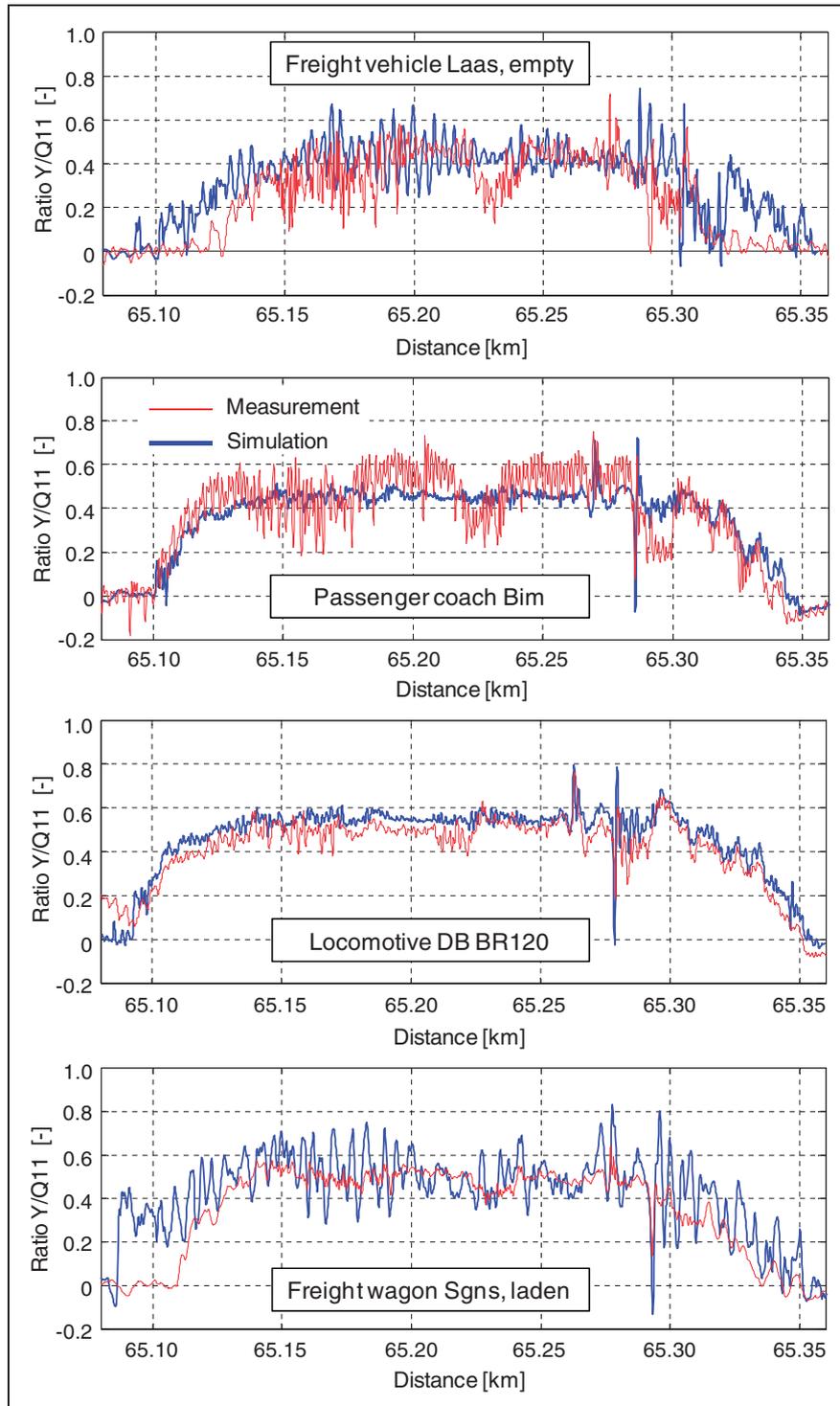
**Figure 5.** Validation examples and subjective assessments. Diagrams from the exercise I (as in Figure 4) in the form used for the subjective assessments by project partners. The values in the circles of each diagram display the percentage of positive assessments.

## Evaluation of the validation method, criteria and limit values

### Evaluation of the assessment based on EN 14363

The assessments based on quantities specified in EN 14363 were carried out using a common preliminary set of validation limits, which were evaluated from the proposals provided by the project partners.

These proposals significantly deviated not only in the proposed limit values but also in principle as shown schematically in Figure 8 that displays the areas fulfilling the proposed validation condition. If the simulated value  $S_v$  and measured value  $M_v$  are identical, the point is on the diagonal line. A deviation from this diagonal line represents a deviation between the simulation and measurement.



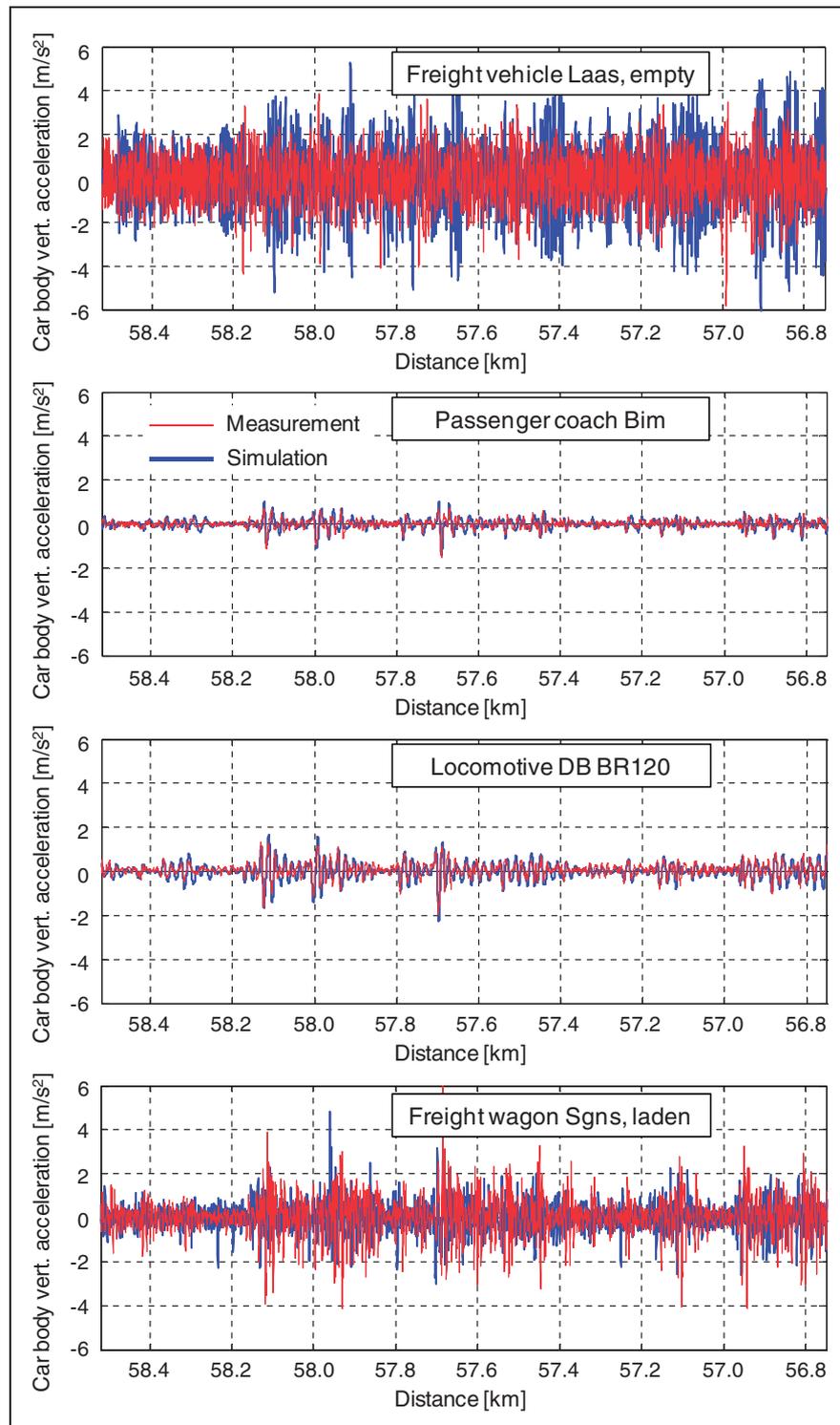
**Figure 6.** Validation examples: the  $Y/Q$  ratio for the outer wheel of the leading wheelset, exercise 2, Germany, Geislingen–Westerstetten line, curve radius 312 m, cant 100 mm, speed 68 km/h.

The following differing definitions of the limit condition were proposed.

1. Deviation limit as a percentage of the measured value (relative deviation limit): see Figure 8(a).
2. Constant deviation limit (absolute deviation limit): see Figure 8(b).
3. Deviation limit decreasing with the measured value increasing towards the limit for vehicle

acceptance based on EN 14363, but not falling below a minimum absolute limit at high measured values, as shown in Figure 8(e).

Some partners proposed combinations of previous principles: a relative limit combined with an absolute deviation limit as shown in Figure 8(c); the addition of an absolute and a relative deviation limit as displayed in Figure 8(d); or an absolute (constant)

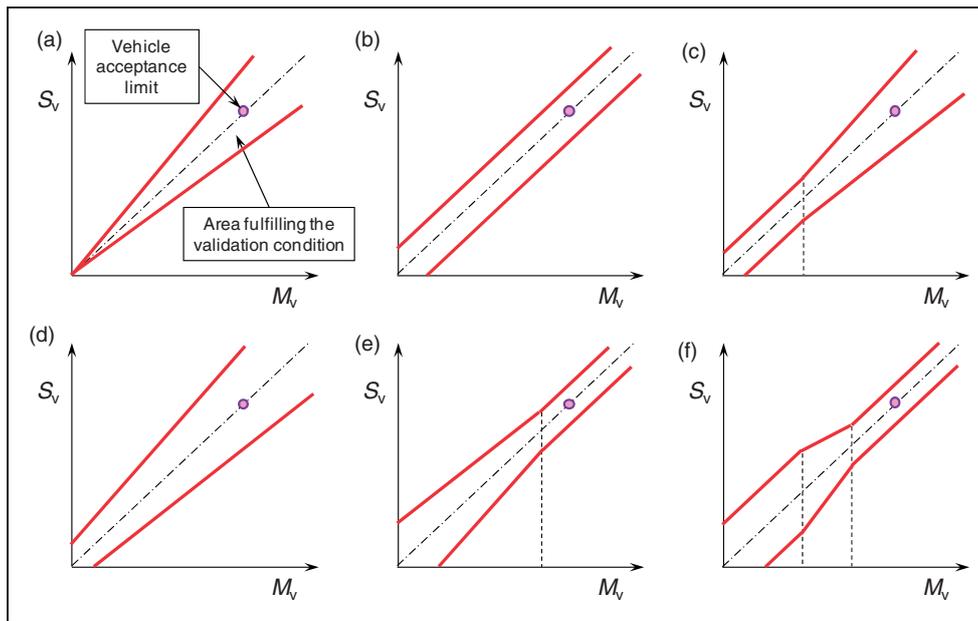


**Figure 7.** Validation examples: vertical acceleration of vehicle body over the leading bogie (wheelset), exercise 8, Germany, Uffenheim–Ansbach line, straight track, speed 120 km/h.

deviation limit that changes with the measured value as shown in Figure 8(f).

A reasonable justification can be provided for each of the different proposals. Any deviation or error is usually considered in regard to relative deviation, thus supporting the approach in Figure 8(a). However, as the vehicle model is intended to be used for simulation of vehicle acceptance tests, it is important to achieve

good agreement especially for values that are close to their limit values for vehicle acceptance, hence supporting the contradicting approach in Figure 8(e). Finally, it was agreed to use constant validation limit values (limits for absolute deviation simulation - measurement), which is quite simple and at the same time the most appropriate compromise for the proposals discussed during the investigations.



**Figure 8.** The main differences in the definitions of the validation limit conditions proposed by project partners.

A set of preliminary validation limits based on the partners' proposals was agreed and then applied for comparisons of model configurations and for the investigation of the possible approach for model validation.

### Evaluation of subjective assessments

Comparisons of measurements and simulations of quantities presented in diagrams were assessed by the project partners using a simple 'Yes/No' method. Due to the large amount of results presented in the form of diagrams, only a part of the results could be assessed by project partners. The following model configurations of vehicles tested in DynoTRAIN were selected for this subjective assessment, all representing the initial vehicle models.

1. Configuration F1 using measured data of wheel and rail profiles as well as measured track irregularities.
2. Configuration D1 using estimated (design) wheel and rail profiles and measured track irregularities.
3. Configuration E1 using measured wheel profiles, estimated (design) rail profiles, measured track irregularities.
4. Configuration C1 using measured wheel and rail profiles, but estimated track irregularities.

These subjective assessments totalled over 6000 diagrams, each assessed by seven to 10 project partners, which resulted in more than 50,000 single assessments. Moreover, a workshop with invited experts dedicated to model validation was held on 7th November 2012, hosted by Siemens AG in Krefeld, Germany. A total of 26 workshop attendees (academics, experts

from industry, railway companies, testing and research institutes, members of the standardisation committee as well as DynoTRAIN project partners) participated in the subjective engineering judgement of diagrams. The assessments questionnaire contained 110 selected time or distance plots and 10 PSD diagrams. The workshop was intended to collect data about the visual assessment of diagrams containing comparisons between simulation and measurement data.

An assessment of a vehicle simulation model requires knowledge about the vehicle itself and about the boundary conditions of the comparison (i.e. kind and quality of available measurement data and parameters of the vehicle model). The information collected in the workshop was intended to be used to investigate the feasibility of replacing the subjective engineering assessment with an objective metric about the degree of similarity between simulation and measurement data. For this reason the workshop procedure stressed the importance of focusing on each single diagram and the workshop attendees were asked to assess each diagram separately by a simple Yes/No method under the following considerations.

1. Assume that a sufficient number of diagrams have already been assessed, each one containing a comparison between simulation and measurement of the particular vehicle.
2. Assume that until the current, last diagram, all previous diagrams were considered as satisfying the validation criteria; some of the previous diagrams, however, did not show a good agreement, so that there are still doubts about whether this model can be confirmed as validated.

- Answer, if the current diagram confirms that the actual vehicle model can be considered as validated or if it confirms your doubts and this vehicle model thus cannot be validated.

It was intended to ask for an engineering judgement based on a pure visual impression from the assessed diagram, so as to not be biased by any consideration about the actual boundary condition of the simulation or any consideration about the reasons why the signals show a particular behaviour. Thus, the requested judgement could be transformed in to a computable measure calculated using the data presented in the diagram without considering any other boundary condition.

The results of the workshop assessments showed strong variation. Only six from a total of 120 plots were assessed unequivocally; an equal assessment by more than 75% of attendees was provided for 54 plots (45%) of diagrams. Although it was not possible to conclude about a replacement of the assessment results by computable values of investigated validation metrics, this workshop provided interesting information. The form of the presentation of diagrams comparing the simulation and measurement (scaling of diagrams, exchange of signals back/front) significantly influenced the assessment result. From six pairs of two plots presenting identical data using a differing scale, only one set received the same assessment for both diagrams. The remaining five diagram pairs were assessed differently, see the example in Figure 9.

Furthermore, the workshop results showed large differences in the 'level' of strictness of the individual assessors. This can be seen in Figure 10 that displays the percentage of positive assessments in each of the six groups of plots provided by a particular attendee. The workshop attendees are ordered from more strict on the left to less strict on the right. No correlation could be identified between the attendee's strictness and any of the considered categories based on their affiliation or experience. Although the workshop

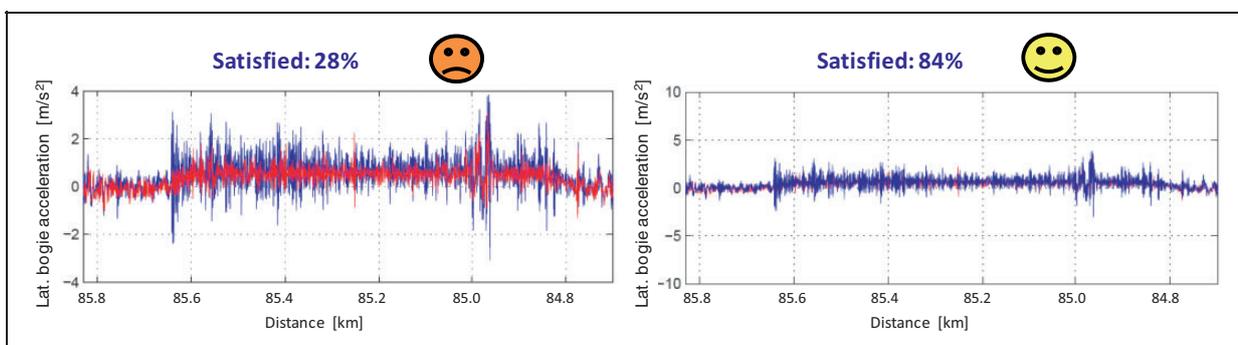
assessments were solely related to diagrams, without any background information about the vehicle type, test conditions and simulation procedure, and thus cannot be considered as representative validation assessments, they illustrate the weakness of subjective judgements. Therefore, it can be concluded, that a subjective assessment using engineering judgement does not ensure the feasibility of an objective model validation.

### Evaluation of validation metrics

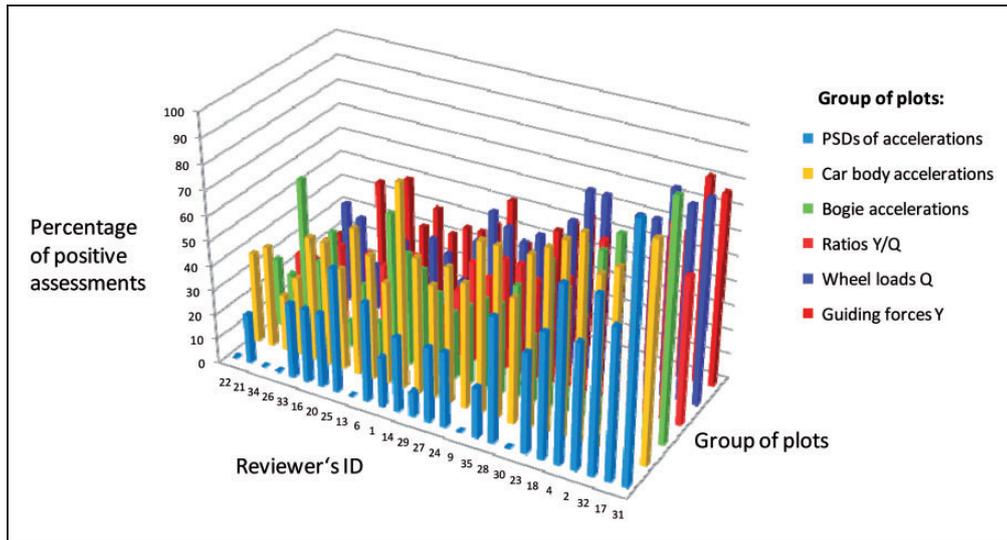
The investigations dedicated to validation metrics were introduced with the aim of replacing a subjective engineering judgement of time or distance plots by a computable and thus objective measure. The previous discussion showed deviations between engineering judgements provided by different assessors, which will surely make a replacement of this judgement more difficult. Moreover, the judgement can further deviate depending on the form and scaling of the diagrams in question as discussed in the section 'Simulation output and comparisons with measurements'. These facts can partly explain the initially surprising effect of a missing correlation between the subjective assessments by project partners and the error factors of the investigated validation metrics.

Nevertheless, the cases resulting in an unexpected disagreement between the validation metric and subjective assessments (high error factors for diagrams with high percentages of positive assessments and vice versa) were further analysed to understand and possibly modify the validation metrics. These analyses identified the following three possible reasons for disagreement between the subjective assessment and validation metrics as demonstrated on examples in Polach and Böttcher.<sup>18</sup>

- The validation metric error factors are based on a relative deviation, and thus they do not consider the magnitude of the evaluated quantity. A relative deviation between simulation and



**Figure 9.** Example of workshop results displaying the effect of the form of diagrams on the assessment of plots presenting identical data. In the right diagram, the scale of the vertical axis is enlarged and the forward and background signal exchanged (red line = measurement, blue line = simulation; see online version for colours).



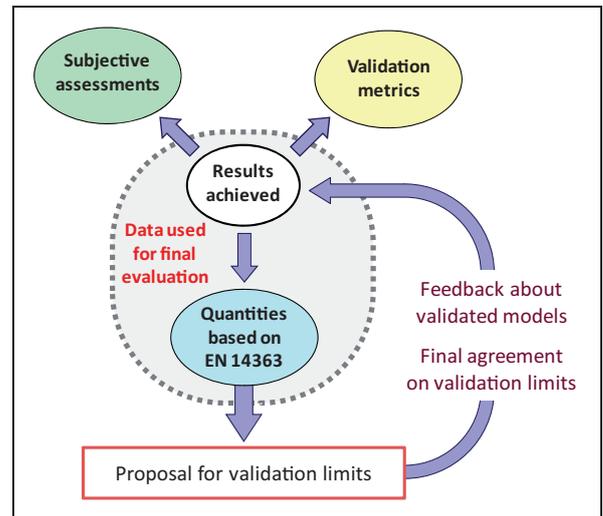
**Figure 10.** Workshop results: percentage of positive assessments provided by any particular attendee in each of the six groups of plots.

measurement at a very low magnitude of a measured quantity is usually neglected in engineering judgement; however, it can provide large error factor values suggesting large disagreements. Although Russell’s definition of the magnitude error factor aims to correct this effect, it is not well suited for the investigated application because of large differences in the magnitudes of different quantities.

2. Another drawback of the validation metric is a strong influence on the phase error factor by the level of synchronisation between simulation and measurement signals, see Polach and Böttcher.<sup>18</sup> A perfect synchronisation is not easy to achieve and is usually not requested, which can lead to high values of the phase error as well as the comprehensive error factors, suggesting disagreement between simulation and measurement in spite of positive visual judgements.
3. The third identified drawback can occur in the case of superposition of dynamic oscillations with a rather high constant quasi-static value. In this case, the resultant integrals will be given by the quasi-static value of the investigated quantity. Thus, if there is agreement in the quasi-static results between simulation and measurement, the error factors will be low and likewise for the case when there is a disagreement in dynamic values. This results in error factors suggesting very good agreement in spite of a low subjective acceptance.

**Evaluation of final validation method, criteria and limits**

The variations of model input data, model adjustments and modelling depth together with variations of track input data resulted in more than 1000



**Figure 11.** Schematic representation of the process used to compare different kinds of validation assessments and to evaluate the final proposal.

simulations of validation exercises. The correlations between the different groups of assessments (EN 14363 quantities, subjective assessments, validation metrics) as well as the relationship between the assessments and the achieved results were investigated as shown in Figure 11.

Summarising the correlation analyses and other results of DynoTRAIN WP5, it is believed that the comparisons of simulation and measurement data using quantities based on EN 14363 represent the best-suited methodology for model validation in the context of vehicle acceptance. Subjective engineering judgement can vary depending on the strictness of the reviewer, and the validation metric, which was considered as suitable for replacement of the subjective judgement, does not show any valuable improvement

compared with the assessment using quantities based on EN 14363 and was therefore not considered in the final proposal.

The validation investigations conducted in DynoTRAIN were carried out under the consideration that the uncertainty of the measurements used for the evaluation represents the state of the art in vehicle approval processes. The analyses of deviations between simulation and measurement data demonstrated that an assessment of a single particular quantity and single pairs of the simulated and measured values do not provide relevant information about the model quality. It is in fact more important to check the overall agreement instead of concentrating on single maximum differences between simulation and measurement data. A single deviation between simulation and measurement data can be related to a particular effect in the measurement or a particular deviation between conditions during the measurement and the input parameters used in the simulations. It is left to chance, if such a single deviation between simulation and measurement will be identified, when selecting the test sections used for comparisons, and the impact of measurement uncertainty on the assessment result increases. Therefore, the model

validation should approve the overall agreement of the deviations between compared pairs of simulation and measurement data.

A statistical approach has been selected to assess this overall agreement; it calculates the mean value and the standard deviation of the differences between the simulation value  $S_v$  and the measurement value  $M_v$  for each of 12 agreed quantities based on EN 14363 (e.g. for all  $Y_{qst}$  values) for a specified minimum of test sections representing the conditions for vehicle acceptance as is described in detail in the next section. The minimum to be used for validation was agreed to be three sections from each test zone according to EN 14363, thus at least 12 sections, and a minimum of two different measuring signals per quantity. Using two force measuring wheelsets to fulfil the later requirement for quasi-static and maximum value of the sum of guiding forces, there are 48 pairs simulation–measurement data points for each of the quantities  $Q$ ,  $Y$  and  $Y/Q$ , which results in a total of 432 compared pairs of simulated and measured values, see Figure 12. The validation evaluations conducted in DynoTRAIN used even more compared pairs. They included 14 sections for freight vehicles and 17 sections for other vehicles, resulting in 504 or 612 compared pairs, respectively.

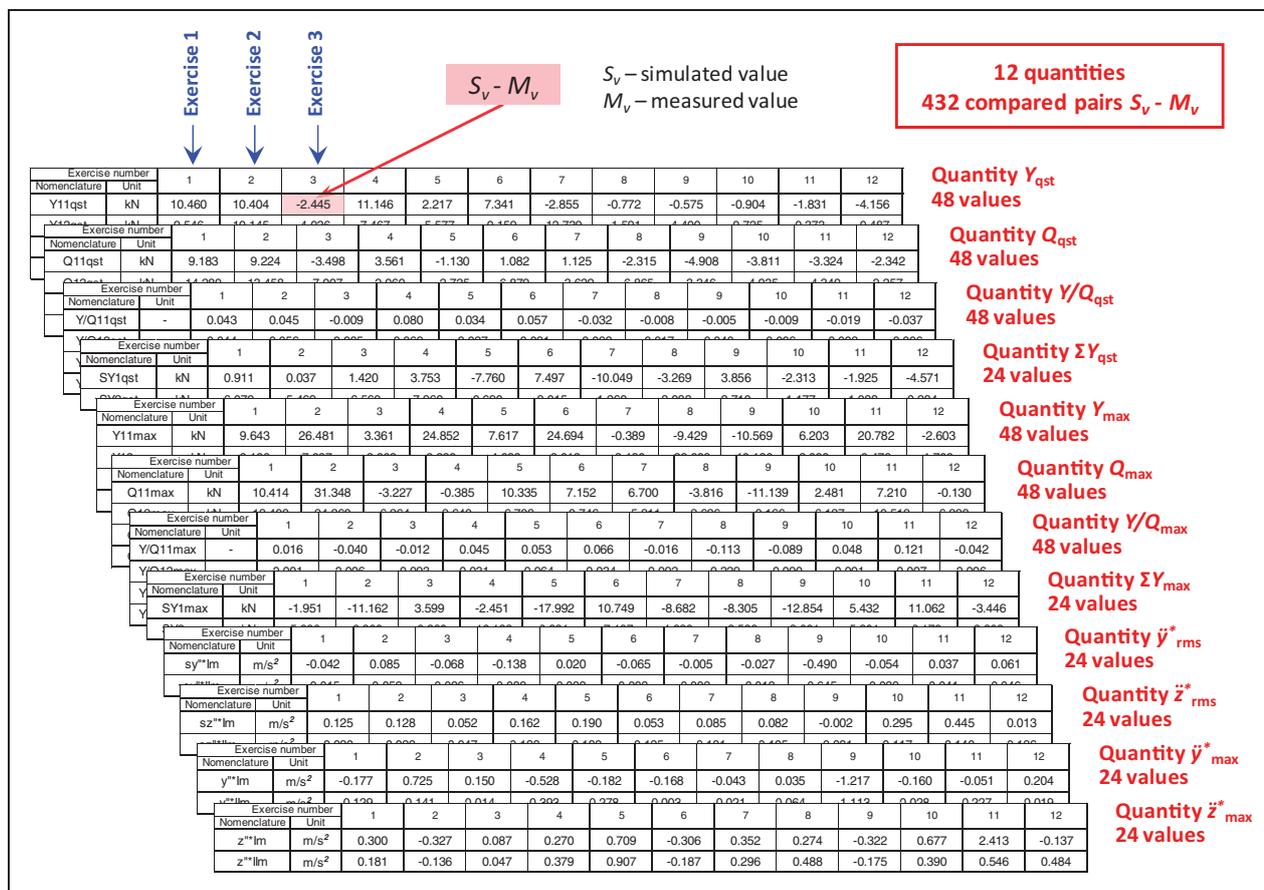


Figure 12. Example of a typical set of comparisons between simulation and measurement values according to the proposed validation method.

The preliminary validation limits agreed in an earlier step of the project were used to assess the validation of all investigated model configurations. The feedback about the validated models was then used for the final adjustment of the validation limits as can be seen in the schematic presentation of this process in Figure 11.

It turned out that the deviations between simulation and measurement values of wheel loads (both quasi-static as well as dynamic) are very sensitive to the static wheel load. Therefore, a validation limit that was dependent on the static wheel load was introduced instead of constant limit value for both quasi-static as well as dynamic wheel loads (see Table 4 in the next section). The constants used in the formulas defining the validation limits for wheel loads were adjusted so that the limits for vehicles with high static wheel loads achieved the range of the originally proposed validation limits, while the validation limits for vehicles with low static wheel loads were smaller.

The level of vehicle body accelerations of freight vehicles and vehicles without bogies or without a secondary suspension is significantly larger than that of vehicles with a typical soft secondary suspension; therefore, the validation limits for the accelerations of the vehicle body of those vehicles were doubled to account for this effect. The accelerations at the bogie frame were evaluated, but not proposed as a

mandatory quantity for model validation. The dynamic behaviour of the bogie or running gear of a particular vehicle model is sufficiently approved by checking the quantities in the wheel/rail contact. Moreover, the investigations carried out showed that the application of the bogie frame acceleration for model validation and the justification of a suitable validation limit are rather difficult and not really necessary as the bogie dynamics is assessed by wheel/rail quantities anyway.

The investigations dedicated to PSD diagrams showed a large variety of results and of deviating assessments by partners as well as during the workshop. Due to limitations of time and resources, a higher priority was put on the evaluation of other criteria. The limited investigations of PSD diagrams did not provide sufficient input for an introduction of criteria and quantitative limits in regard to PSD. This topic needs further investigation.

### Proposed validation method

The proposed validation process is based on a mathematical comparison between the results of on-track tests performed using the normal measuring method based on EN 14363 and the corresponding simulation results. The simulation and measurement results of the specified quantities have to be compared on at

**Table 4.** Quantities and limits for model validation in regard to simulation of on-track tests (from Ref. 18, www.tandfonline.com).

Quantity	Notation	Unit	Filtering	Processing	Validation limit for standard deviation
Quasi-static guiding force	$Y_{qst}$	kN	Low-pass filter 20 Hz	50%-value (median)	5
Quasi-static vertical wheel force	$Q_{qst}$	kN	Low-pass filter 20 Hz	50%-value (median)	4 ( $1 + 0.01 Q_0$ ) $Q_0$ - static vertical wheel force (kN)
Quasi-static ratio $Y/Q$	$(Y/Q)_{qst}$	–	Low-pass filter 20 Hz	50%-value (median)	0.07
Quasi-static sum of guiding forces	$\Sigma Y_{qst}$	kN	Low-pass filter 20 Hz	50%-value (median)	6
Guiding force, maximum	$Y_{max}$	kN	Low-pass filter 20 Hz	0.15%/99.85%-value <sup>a</sup>	9
Vertical wheel force, maximum	$Q_{max}$	kN	Low-pass filter 20 Hz	99.85%-value <sup>a</sup>	6 ( $1 + 0.01 Q_0$ ) $Q_0$ - static vertical wheel force (kN)
Ratio $Y/Q$ , maximum	$(Y/Q)_{max}$	–	Sliding mean (2 m window, step 0.5 m)	0.15%/99.85%-value <sup>a</sup>	0.10
Sum of guiding forces, maximum	$\Sigma Y_{max}$	kN	Sliding mean (2 m window, step 0.5 m)	0.15%/99.85%-value <sup>a</sup>	9
Car body lateral acceleration, RMS-value	$\ddot{y}_{rms}^*$	$m/s^2$	Band-pass filter 0.4 to 10 Hz	RMS-value	0.15 <sup>b</sup>
Car body vertical acceleration, RMS-value	$\ddot{z}_{rms}^*$	$m/s^2$	Band-pass filter 0.4 to 10 Hz	RMS-value	0.15 <sup>b</sup>
Car body lateral acceleration, maximum	$\ddot{y}_{max}^*$	$m/s^2$	Band-pass filter 0.4 to 10 Hz	0.15%/99.85%-value <sup>a</sup>	0.40 <sup>b</sup>
Car body vertical acceleration, maximum	$\ddot{z}_{max}^*$	$m/s^2$	Band-pass filter 0.4 to 10 Hz	0.15%/99.85%-value <sup>a</sup>	0.40 <sup>b</sup>

<sup>a</sup>Absolute values of simulated value  $S_y$  as well as measured value  $M_y$ .

<sup>b</sup>For freight vehicles and vehicles without bogies or without secondary suspension, these limits have to be doubled.

least 12 track sections, called validation exercises. A track section can be either a test section as in EN 14363 or a part of a test track longer than the minimum length specified for track sections in the particular test. Moreover, these sections have to fulfil the other test section requirements in EN 14363 such as constant curve radius, etc. The selected validation exercises have to contain sections from all four test zones, with at least three sections from each test zone. The track geometric irregularities have to represent the conditions of the on-track tests.

Each quantity has to be evaluated using at least two signals, e.g. vertical acceleration above the leading and trailing bogies, thus, at least 24 simulated values  $S_v$  are compared to the corresponding measured values  $M_v$  of each quantity. Each compared simulated as well as measured quantity has to be filtered and processed based on the requirements in Table 4. The percentiles have to be calculated from the cumulative curve. For the maximum value calculated as 0.15% or 99.85%-value, the higher magnitude of the 0.15%- and 99.85%-values (absolute value) is used. The 50%-values (medians) are applied with their sign to show the agreement of both magnitude and direction of those quantities. The difference  $D_v$  between the simulated value  $S_v$  and the corresponding measured value  $M_v$  has to be evaluated for each value and each quantity; this difference has to be transformed so that, if the magnitude of the simulation value is higher than the magnitude of the measurement (simulation overestimating the measurement), the difference is positive, and vice versa

$$\begin{aligned} D_v &= (S_v - M_v) \frac{M_v}{|M_v|} \quad \text{for } M_v \neq 0 \\ D_v &= S_v \quad \text{for } M_v = 0 \end{aligned} \quad (4)$$

The following values have to be calculated for the whole set of differences  $D_v$  between the simulation and measurement for each quantity:

- the mean of the differences between the simulation value  $S_v$  and the measurement value  $M_v$ ;
- the standard deviation of the same set of differences.

The standard deviation of the set of differences between the simulation value  $S_v$  and the measurement value  $M_v$  for each individual quantity has to be less than or equal to their validation limit shown in Table 4. For each quantity the mean of the set of differences between the simulation value  $S_v$  and the measurement value  $M_v$  should be less than or equal to a validation limit equal to two-thirds of the related validation limit for the standard deviation. The validation limits for accelerations (standard deviation as well as mean of differences) for freight vehicles or vehicles without a secondary suspension are twice the relevant limit values for other vehicles.

As an example, Figure 13 explains the calculation of differences between the simulation value  $S_v$  and the measurement value  $M_v$  for the quasi-static values of the sum of guiding forces between wheelset and track, their transformation, as well as calculation of the mean value and standard deviation, which are used for comparison with the validation limits specified in Table 4.

## Discussion

### *Advantages of the proposed validation method*

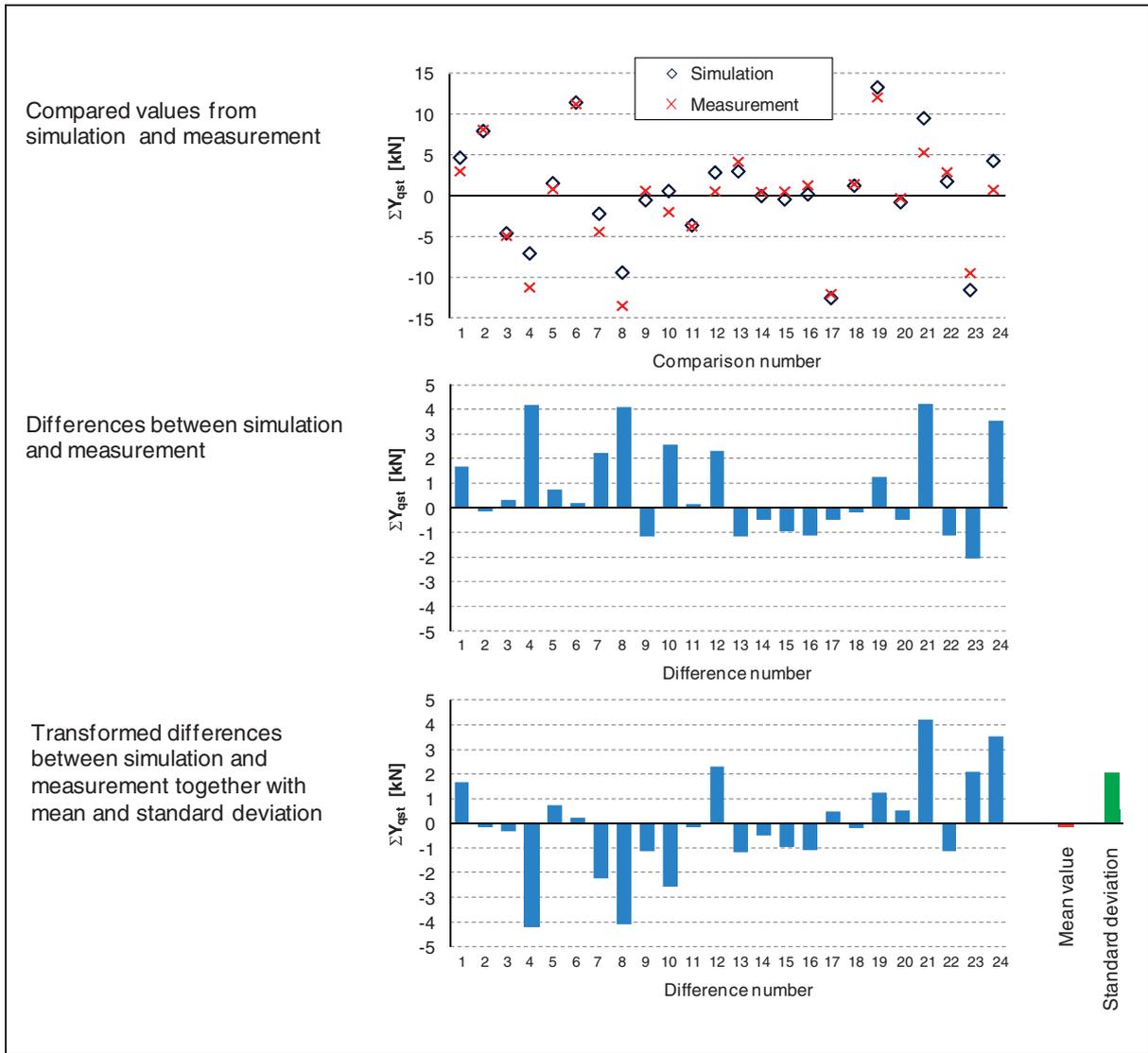
The proposed final set of validation limits was applied to assess the validity of all the investigated model configurations. From a total of 78 model configurations evaluated, only 20 fulfil the proposed model validation limits:

- eight from 24 models of the locomotive BR 120 investigated by Siemens;
- 10 from 13 models of the Bim coach investigated by Bombardier Transportation;
- two from four models of the Bim coach investigated by IFSTTAR.

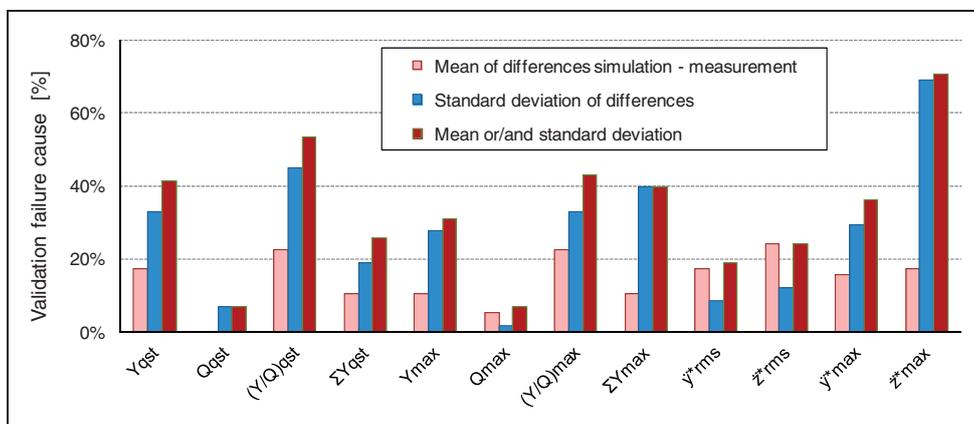
The validated models are the models of vehicles tested in DynoTRAIN and validated using measured track irregularities as well as measured wheel and rail profiles. The only successfully validated models were those of the locomotive BR 120 and the Bim coach.

The contributions of quantities leading to the failure of 58 out of the total of 78 model configurations are displayed in Figure 14. The failure to validate a model could be caused by one quantity or more quantities at the same time; an exceedance could be the result of either the standard deviation of differences, or the mean value of differences or both values at the same time. The most frequent cause was an exceedance of the maximum value of the vertical acceleration of the car body. Other common causes were  $Y/Q$  (quasi-static as well as maximum value),  $Y_{qst}$  and  $\Sigma Y_{max}$ . The wheel loads seldom caused the limits to be exceeded and there was no exceeding of the validation limit for the mean value of the differences of  $Q_{qst}$ . Thus, it seems that on the basis of the proposed validation method, the expected model properties can be easily achieved for the vertical wheel forces whereas it is rather difficult to achieve the expected properties for the ratio  $Y/Q$  or vertical acceleration of the vehicle body.

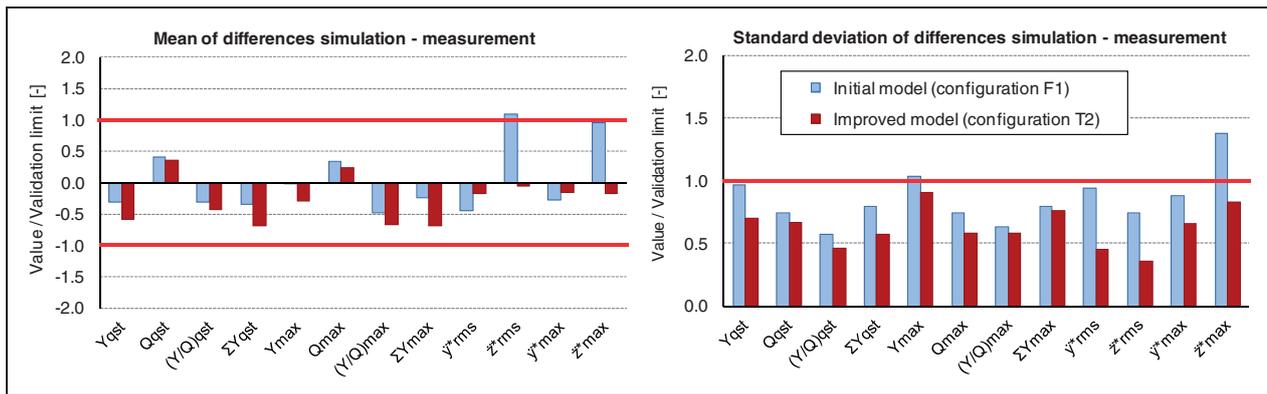
An important advantage of the proposed validation procedure is that this assessment represents an overall assessment of a large amount of data that is impossible to carry out by using engineering judgement of plots, as it is not practically possible to display, check and document the approval of such a large number of plots. The calculation of



**Figure 13.** Example of data evaluation: simulated and measured values, their differences, transformation of differences in regard to the sign of the measured value, and calculation of the mean value and standard deviation of the quasi-static sum of guiding forces  $\Sigma Y$ .



**Figure 14.** Contributions of individual quantities to the validation failure observed for 58 out of the 78 investigated model configurations.



**Figure 15.** Example of validation results using the proposed method. Normalised values of the mean (left) and standard deviation (right) of the differences between simulation and measurement data for two vehicle models of the locomotive BR 120 performed by Siemens: initial model F1 and improved model T2 after comparisons with on-track as well as stationary tests.

characteristic parameters of the mean and standard deviation of the differences between the simulation values  $S_v$  and the measurement values  $M_v$ , and their comparison with the validation limits, however, allows a fast identification of quantities with a large deviation. The data of a particular quantity can be easily checked in detail to identify the validation exercise (section) and the signal (sensor position) that provides a large deviation between the simulation and measurement values.

The specified set of 12 quantities covers the quasi-static as well as dynamic behaviour of a vehicle in regard to vehicle acceptance, which is the intended area of application for a validated model. The signal processing is carried out by analogy with EN 14363 for both the measurement and simulation, thus allowing direct use of the acceptance tests data.

The weakness of the model in question can be identified by a normalisation of the validation criteria, dividing them by the validation limits, as can be seen in Figure 15. The model is validated, if the absolute magnitudes of all displayed values are not higher than one. The vehicle models in Figure 15 were prepared by using the available parameter data, before any model adjustments by comparisons with either stationary or on-track tests. This figure shows normalised values of the mean (left) and standard deviation (right) of the differences between simulation and measurement data for two vehicle models of the locomotive BR 120 performed by Siemens. The initial model F1 does not fulfil the proposed validation limits. This model used measured track irregularities as well as measured wheel and rail profiles, but it was not adjusted based on the measurements. The improved model T2 after adjustments by comparisons with on-track tests and with stationary tests meets the validation limits.

The results confirm the proposed validation criteria and limits as a suitable methodology for the validation of railway vehicle models. The proposed validation method allows not only an objective assessment,

but also a clear identification of the model weaknesses, see also Polach et al.<sup>26</sup>

#### *Effect of model adjustment using stationary tests on the simulation of on-track tests*

Static and low-speed tests can be used to identify missing or uncertain vehicle model parameters and to support vehicle model validation. A comparison of simulations with available stationary tests is required as part of the model validation process in prEN 14363.<sup>14</sup> The simulation and measurements of the stationary tests are compared and the uncertain model parameters adjusted if necessary.<sup>27</sup> However, what is the effect of a model adjustment based on a comparison with stationary tests on the agreement between simulation and measurement of the on-track test (ride test)? This is typically neither presented nor investigated; it is believed that an improved agreement with stationary tests will implicitly improve the exactness of the on-track test simulation. In order to investigate this effect, the validation exercises with on-track tests were repeated with several versions of the same model, either before the comparison with stationary tests or after comparison and adjustment in order to fit the stationary tests results, respectively.

The stationary tests used during the validation evaluations of the simulation models differed depending on the availability of the test results. An overview of the stationary tests used for comparisons and model adjustments of vehicles tested in DynoTRAIN is shown in Table 5. Not all comparisons resulted in a proposal of model adjustment. Particularly the tests performed on a flat curve with a radius of 150 m representing a part of the test of safety against derailment on twisted track according to Method 2 in EN 14363<sup>20</sup> did not provide any suggestion regarding the parameter adjustment of the investigated vehicles.

Figure 16 shows comparisons of the validation results obtained using the proposed validation

**Table 5.** Comparisons and adjustments of vehicle models using stationary tests. The tests used for comparisons are marked with a cross (X); 'no adjustment' is stated if the comparison with the respective stationary test did not provide any suggestion for model adjustment.

Vehicle model	Wheel unloading test (twist test) based on EN 14363	Test in flat curve $R=150$ m based on EN 14363	Bogie rotational resistance test based on EN 14363	Bogie lateral resistance test to measure the characteristic of the lateral suspension	Sway test - measurement of roll coefficient
Locomotive DB BR 120, Siemens	X	X (no adjustment)	X	X	X
Locomotive DB BR 120, IFSTTAR	—	—	X	X	X
DB passenger coach Bim, Bombardier Transportation	X	X (no adjustment)	X	X	—
DB passenger coach Bim, IFSTTAR	—	—	X (no adjustment)	X	—
Freight wagon Sgns, empty, Technical University Berlin	—	X (no adjustment)	X	—	—
Freight wagon Sgns, empty, IFSTTAR	—	—	X	X	—
Freight wagon Sgns, laden, Technical University Berlin	—	X (no adjustment)	X	—	—

method for six models of vehicles tested in DynoTRAIN WP1. The figure presents comparisons of the initial model configurations F1 using measured track irregularities and measured wheel as well as rail profiles, however, without any model adjustments based on comparisons with stationary or on-track tests, and model configurations T1 after adjustments based on comparisons with stationary tests stated in Table 5 (in the case of laden freight wagon Sgns by TU Berlin the compared configurations are F2 and T2 with modified suspension modelling). The parameters adjusted to improve the models can be illustrated by considering the example of the locomotive model performed by Siemens, where the modifications considered the vertical and lateral stiffness of the secondary suspension, the characteristics of the secondary lateral bump stop and the height of the centre of gravity of the vehicle body.

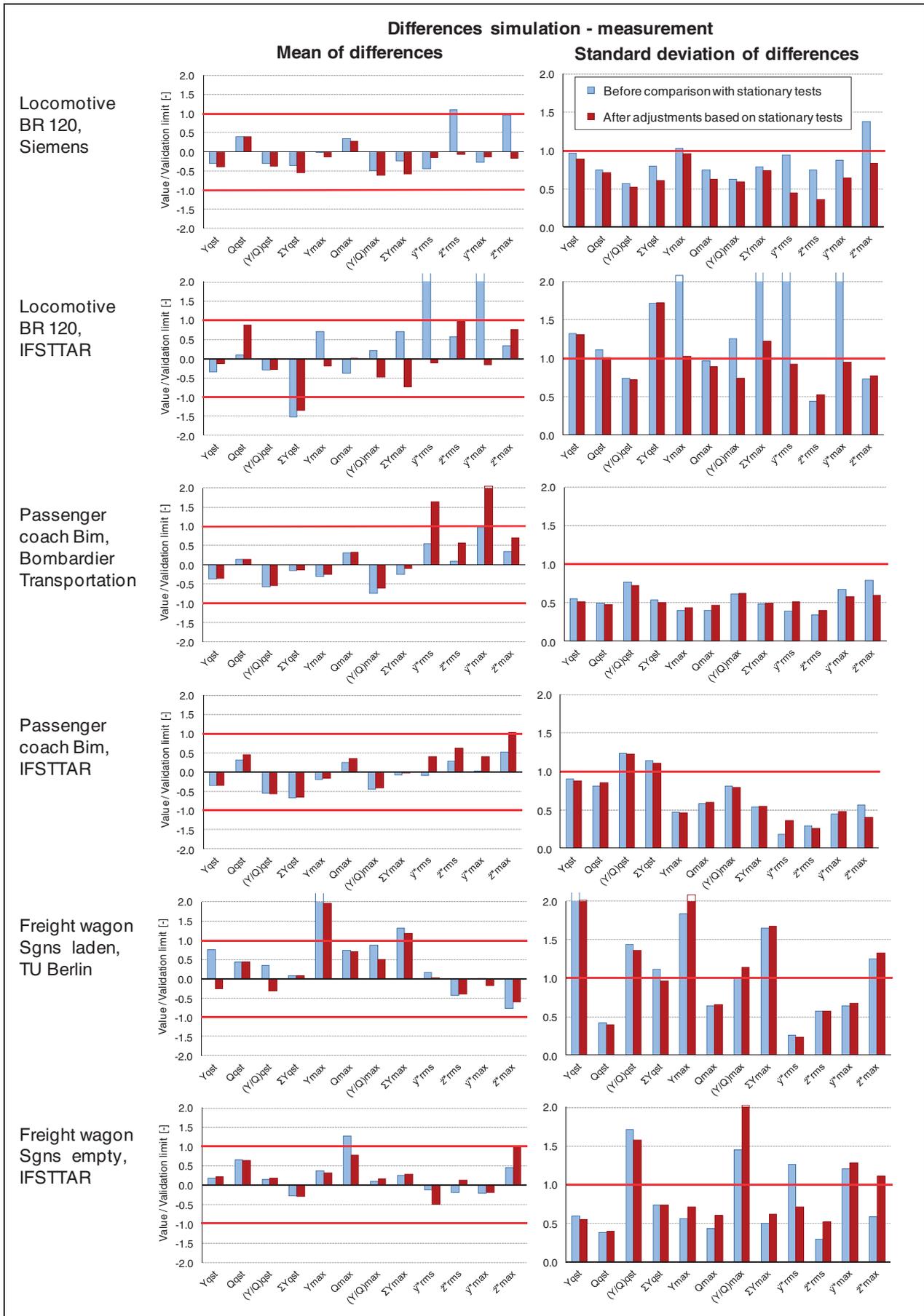
The model adjustments by comparisons with stationary tests did not lead to expected improvements of the results regarding the simulations of the on-track tests. Only the investigations by Siemens and IFSTTAR regarding the locomotive BR 120 clearly provided better results for the models after the comparison and adjustment due to the stationary tests. In other cases, the model adjustments introduced using the stationary tests did not significantly affect the agreement between the simulation and measurement concerning the on-track tests or provided even worse results. For example, in the model of the Bim coach by Bombardier Transportation, an implementation of friction elements intended to model a rather small

hysteresis in the secondary lateral suspension resulted in the failure of the model validation due to significantly higher lateral car body accelerations compared with the values measured during the on-track test.

These investigations did not confirm the traditional opinion regarding the positive effect of model adjustments by comparisons with stationary tests on the simulation of on-track tests. A possible explanation is that focussing on the static and low-speed behaviour can result in model adjustments that are less accurate in regard to dynamic behaviour. The stationary tests can support the identification of model parameters that are unknown or uncertain. A good agreement between simulation and measurement of stationary tests, however, does not guarantee a good agreement between simulation and measurement of on-track tests. An adequate number of comparisons between simulations and on-track measurements is the only suitable and reliable model validation method in regard to the simulation of on-track tests.

## Summary and conclusions

The presented part of the investigations in the framework of the DynoTRAIN project was dedicated to the evaluation of a validation method suited for simulations in the context of vehicle acceptance. It represents a unique activity of complex testing, simulations, comparisons with measurements and evaluations. The on-track measurements included several vehicles, tested using 10 force measuring wheelsets in four European countries and a test train



**Figure 16.** Effect of comparisons with stationary tests on the validation results. Normalised values of mean (left) and standard deviation (right) of the differences between simulation and measurement for the initial vehicle models and models after adjustments based on comparisons with stationary tests.

equipped for the simultaneous recording of track irregularities and rail profiles. The simulations were performed using several vehicle models, built with the use of different simulation tools by different partners. The comparisons between simulation and measurement were conducted in a large number of simulations using a set of the same test sections. The results were assessed by three different validation approaches: by comparisons based on values according to EN 14363; by subjective engineering judgement; and by using computable measures, so-called validation metrics.

The proposed model validation criteria and limits are based on 12 quantities evaluated by analogy with EN 14363, covering quasi-static and dynamic wheel/rail force measurements and vertical as well as lateral vehicle body accelerations. For each quantity, a set of at least 24 comparisons between simulation and measurement are evaluated using values based on EN 14363 from at least 12 sections that represent all four test zones as required in EN 14363 from straight track to curves with a very small radius. The agreement between simulation and measurement is assessed by comparing the mean value and standard deviation for a set of differences between simulated and measured values of each quantity with the proposed validation limit.

The investigations neither confirm nor deny the traditional opinion about the positive effect of the model adjustments by comparisons with stationary tests on the simulation of on-track tests. This topic would need further investigation. According to the presented investigations, comparisons between simulations and on-track measurements represent the only suitable and reliable model validation method in regard to the simulation of on-track tests. The proposed method, criteria and limits represent a suitable methodology for the validation of railway vehicle models. It represents an overall assessment of a large number of data, which is impossible to carry out by using engineering judgement, as it is not practically possible to display, check and document the approval of such a large number of plots. This validation process not only allows an objective assessment, but also supports an identification of the weaknesses of the model. The presented methodology is proposed for implementation in a revised standard EN 14363. Feedback from future applications of this method in allied projects will help to further improve and develop the model validation, which is the crucial condition for successful use of simulation to reduce the amount and cost of physical testing in the railway vehicle acceptance process.

### Funding

This article describes work undertaken in the context of the DynoTRAIN project, Railway Vehicle Dynamics and Track Interactions: Total Regulatory Acceptance for the Interoperable Network ([www.triotrain.eu](http://www.triotrain.eu)). DynoTRAIN is

a collaborative project – medium-scale focused research project supported by the European Seventh Framework Programme, contract number: 234079 and is led by UNIFE.

### References

1. ASME. *Guide for verification and validation in computational solid mechanics*. New York, NY: ASME Press, 2006.
2. Cooperrider NK and Law EH. A survey of rail vehicle testing for validation of theoretical dynamic analyses. *J Dyn Syst, Meas Control* 1978; 100: 238–251.
3. Gostling RJ and Cooperrider N. Validation of railway vehicle lateral dynamics models. *Veh Syst Dyn* 1983; 12: 179–202.
4. Cheli F, Corradi R, Diana G and Facchinetti A. Validation of a numerical model for the simulation of tramcar vehicle dynamics by means of comparison with experimental data. *J Comput Nonlinear Dyn* 2007; 2: 299–307.
5. Alfi S, Bruni S and Mazzola L. Numerical methodology for the evaluation of high speed vehicle stability in curved track. In: Z Istavan (ed.) *The eighth international conference on railway bogies and running gears*, Budapest, Hungary, 13–16 September 2010, pp.247–256. Budapest: BME.
6. Kuka N, Ariaudo C and Verardi R. Modelling and simulation of tilting trains. In: *The 22nd international symposium on dynamics of vehicles on roads and tracks*, Manchester, UK 14–19 August 2011, paper no. 135. Manchester: Manchester Metropolitan University.
7. Evans J and Berg M. Challenges in simulation of rail vehicle dynamics. *Veh Syst Dyn* 2009; 47: 1023–1048.
8. Bruni S, Vinolas J, Berg M, et al. Modelling of suspension components in rail vehicle dynamics context. *Veh Syst Dyn* 2011; 49: 1021–1072.
9. GM/RC2641, Issue 2: 2009. Recommendations for vehicle static testing.
10. GM/RT2141, Issue 3: 2009. Resistance of railway vehicles to derailment and roll-over.
11. EN 15273-2: 2013. Railway applications - gauges - part 2: rolling stock gauge.
12. UIC Code 518: 2009. Testing and approval of railway vehicles from the point of view of their dynamic behaviour – safety – track fatigue – running behaviour.
13. Jönsson LO, Nilstam N and Persson I. Using simulations for approval of railway vehicles: a comparison between measured and simulated track forces. *Veh Syst Dyn* 2008; 46: 869–881.
14. prEN 14363: 2013. Railway applications—testing and simulation for the acceptance of running characteristics of railway vehicles—running behaviour and stationary tests.
15. Mazzola L. The influence of modelling of the suspension components on the virtual homologation of a railway vehicle. In: Pombo J (ed.) *The first international conference on railway technology: research, development and maintenance*, La Palma, Spain, April 2012, paper no. 75. Stirling, Scotland: Civil-Comp Press.
16. Mazzola L and Berg M. Secondary suspension of railway vehicles - air spring modelling: performance and critical issues. *Proc IMechE, Part F: J Rail Rapid Transit* 2014; 228: 225–241.

17. Zacher M and Kratochwille R. Stationary and on track tests with different vehicles. *Proc IMechE, Part F: J Rail Rapid Transit* 2015; 229(6): 668–690.
18. Polach O and Böttcher A. A new approach to define criteria for rail vehicle model validation. *Veh Syst Dyn* 2014; 52(Suppl. 1): 125–141.
19. ORE B 176: 1989. Bogies with steered or steering wheelsets. Part 1: specifications and preliminary studies. Part 2, specification for a bogie with improved curving characteristics.
20. EN 14363: 2005. Railway applications—testing for the acceptance of running characteristics of railway vehicles—testing of running behaviour and stationary tests.
21. Fries R, Walker R and Wilson N. Validation of dynamic rail vehicle models. In: *The 23rd international symposium on dynamics of vehicles on roads and tracks*, Qingdao, China, 19–23 August 2013, paper no. 10.1.
22. Schwer LE. Validation metrics for response histories: perspectives and case studies. *Engng Comput* 2007; 23: 295–309.
23. Russell DM. Error measures for comparing transient data: part I: development of a comprehensive error measure, part II: error measures case study. In: *The 68th shock and vibration symposium*, Hunt Valley, MD, 3–6 November, 1997, pp.175–198.
24. Sprague MA and Geers TL. A spectral-element method for modelling cavitation in transient fluid-structure interaction. *Int J Numer Methods Engng* 2004; 60: 2467–2499.
25. Mongiardini M, Ray MH and Anghileri M. Development of software for the comparison of curves during the verification and validation of numerical models. In: *The seventh European LS-DYNA conference*, Salzburg, Austria, 14–15 May, 2009. Stuttgart: DYNAmore GmbH.
26. Polach O, Böttcher A, Vannucci D, et al. Validation of multi-body models for simulations in authorisation of rail vehicles. In: Istavan Z (ed.) *The ninth international conference on railway bogies and running gears*, Budapest, Hungary, 9–12 September, 2013, pp.187–196. Budapest: BME.
27. Evans J. Validation of vehicle dynamic modelling – some practical experience. In: *The 23rd international symposium on dynamics of vehicles on roads and tracks*, Qingdao, China, 19–23 August 2013, paper no. 2.3.