Model validation – a random or a reliable result?

G. Götz
Institute of Rail Vehicles, Berlin Institute of Technology, Berlin, Germany

O. Polach
Consultant, Neuhausen am Rheinfall, Switzerland

ABSTRACT: In this paper, the influence of the number and compilation of selected test sections on model validation results is assessed. Millions of model validation results of three different rail vehicles are calculated according to Validation Method 2 of EN 14363:2016 using hundreds of newly created test sections from the DynoTRAIN database. The presented investigation shows that the specified minimum of three test sections per test zone is a reasonable compromise between the effort and reliability of the validation, whereby it is recommended to use the same or similar number of test sections from all test zones. The evaluated validation method seems to be reliable. Once a model revealed to be not validated it is not worthwhile to select other test sections for validation expecting to change the overall validation result.

1 INTRODUCTION

Multi-body dynamics simulation tools are increasingly used as a mean to reduce costs and efforts of rail vehicle certification by reducing the amount of on-track testing. The DynoTRAIN model validation approach [1, 2] as one of the recently proposed validation methods [3], is now integrated as Validation Method 2 in the revision of standard EN 14363:2016 [4]. Contrary to other validation proposals, this validation method provides not only the validation metric but also the specification of the measured signals, the test sections as well as the validation limits (matching errors to be fulfilled for a successful validation). This validation methodology was developed using several models of different rail vehicles developed by different modellers using different simulation tools. However, the number of investigated test sections in the DynoTRAIN project was limited to 17. As in any validation method, the final validation result depends on a set of test sections used for comparisons. Does it mean, the validation result is a pure coincidence or can the validation be manipulated by selecting only “good” simulation results? The authors examined these questions in unique investigations considering millions of model validation results of three rail vehicle models, namely, a passenger coach and a freight wagon in empty and loaded condition.

2 METHODOLOGICAL APPROACH

2.1 *Model validation according to Method 2 of EN 14363:2016*

The Validation Method 2 according to EN 14363:2016 [4] requires comparisons between simulation and test results in at least 12 test sections with the quantities shown in Table 1. It is required to evaluate at least two different measuring signals per quantity. For each validation quantity, the mean value and the standard deviation value are calculated from differences between simulations and measurements. These values are compared with the limit values defined in [1, 2, 4] and presented as non-dimensional values related to the corresponding validation limits. A model is assessed to be validated, if the mean and the standard deviation value of all validation quantities do not exceed their validation limits, see [2] for details.

Table 1. Quantities for model validation according Validation Method 2 of EN 14363:2016 [4]

Quasi-static values of wheel-rail quantities	Guiding force Y_{qst}	Vertical wheel force Q_{qst}	Ratio $(Y/Q)_{qst}$	Sum of guiding forces ΣY_{qst}
Dynamic values of wheel-rail quantities	Guiding force Y_{max}	Vertical wheel force Q_{max}	Ratio $(Y/Q)_{max}$	Sum of guiding forces ΣY_{max}
Vehicle body	Vertical		Lateral	
accelerations	Rms-value	Max. value	Rms-value	Max. value

2.2 *Aims of research, database and applied procedure*

The authors investigated already the sensitivity of the Validation Method 2 according to EN 14363:2016 [4] regarding the number of compared test sections, varying section length as well as other parameters on selected examples in reference [5]. The aim of the presented research is to assess the influence of the number and compilation of simultaneously selected test sections from each test zone on the model validation result based on a systematical evaluation of a large number of test sections. The investigated Validation Method 2 requires comparisons between simulation and testing in at least 12 test sections (three per test zone). The maximum number of compared test sections is, however, not limited, and the procedure of selecting test sections for model validation is not specified. The Validation Method 2 was developed in the DynoTRAIN project [1, 2] using 17 test sections from Germany, Switzerland, France and Italy. The authors separated these test sections in 89 EN-sections (parts of test sections with a length according to EN 14363:2016 [4]) and used them for several investigations presented in [5]. Considering the aims of this new research, the authors created hundreds of sections with the length according to EN 14363:2016 [4] (called here EN-sections as well) with measured track geometry and track irregularities from the DynoTRAIN database that have not been applied for investigations in terms of model validation before. Beside other requirements as e.g. constant speed of test train inside the particular EN-section, every EN-section is unique which means that simulation results are not compared to multiple measurements of the same EN-section. At the example of the passenger coach, 149 test sections are created and separated into 611 EN-sections. The authors developed and applied for the presented research the so far unique process shown in Figure 1. Initially, the authors tried to create as many as possible EN-sections for the three rail vehicles. Later, several country-specific compilations of representative 25 EN-sections per test zone were created to fulfil the requirement for evaluating on-track tests according to EN 14363:2016 [4]. For all three vehicles, compilations of 100 EN-sections (25 per test zone) from Germany (test zone 1 to 3) and Switzerland (test zone 4) and 93 Italian EN-sections (25 in test zone 1, 3, 4 and 18 in test zone 2) are created. This results in the following groups of EN-sections used for the presented investigations:

- Case 1: Passenger coach, Germany/Switzerland
- Case 2: Passenger coach, Italy
- Case 3: Freight wagon (empty), Germany/Switzerland
- Case 4: Freight wagon (empty), Italy
- Case 5: Freight wagon (loaded), Germany/Switzerland
- Case 6: Freight wagon (loaded), Italy.

To assess e.g. the influence of the compilation of selected EN-sections on model validation results, the scatter of millions of validation results is analyzed for the first time with boxplot-diagrams, see Figures 3 and 4. The percentage value represents the number of combinations with which the model is assessed to be validated. The initial approach was to calculate the scatter of validation results considering all different combinations of EN-sections when selecting e.g. 12 of 100 EN-sections (three of 25 per test zone). The authors recognized soon that this led to a big data analysis. The number of different combinations increases exponentially with the maximum number of EN-sections per test zone. Despite a programming technique enabling the use of multiple CPUs of one computer, the calculation time of these billions of model validation results would take several months. Considering data processing of the calculated validation results, requirements for RAM increases exponentially, the more validation results are analyzed with boxplot diagrams. Both leads to performance limits of (powerful) office computers. The authors were able to assess approximately 100 million model validation results in eight days but these numbers are extremely far below the numbers considering all different combinations in all cases. Thus, the authors investigated, how the percentage value of randomly created combinations of EN-sections changes with which the model is assessed to be validated. Complementary research with other vehicles and other compilations of EN-sections confirmed that the calculation of 250 thousand model validation results is an accurate estimation of the percentage value that would result from calculating model validation results of all different combinations of EN-sections. The authors selected this number for the following investigations.

The models of the passenger coach, the empty and loaded freight wagons are simulated with SIMPACK 8.905b using measured track geometry and track irregularities, measured wheel profiles, country-specific nominal rail profiles and rail inclination. The wheel/rail friction coefficient is 0.45 which represents dry rail conditions during the on-track tests.

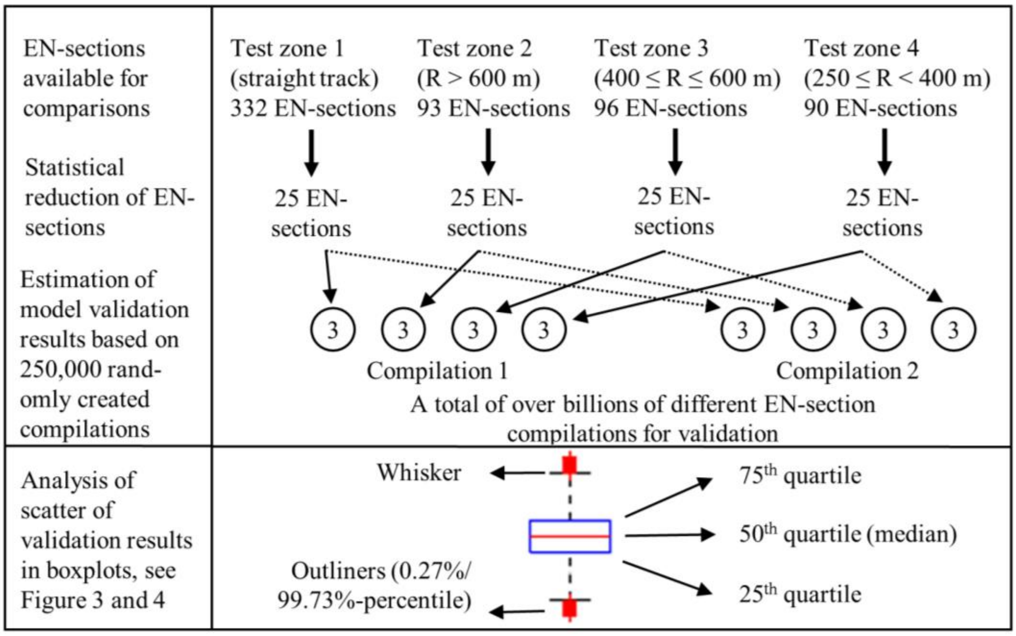

Figure 1. Process to investigate the influence of selected EN-sections on model validation results

3 INFLUENCE OF SELECTED TEST SECTIONS ON MODEL VALIDATION RESULTS

3.1 *Symmetric selection of EN-sections for model validation*

The investigated validation method requires at least 12 test sections (three per test zone) for model validation, but it is allowed to use more than this minimum number of test sections. In this subchapter, the influence of the symmetrical selection of EN-sections on model validation results is assessed. For each vehicle and compilation of EN-sections, the number of randomly selected EN-sections increases successively from three to the maximum available number of EN-sections per test zone. The number of selected EN-sections from each test zone is always the same and thus, symmetric. Figure 2 shows the principal influence at the example of the validation quantity $(Y/Q)_{qst}$. By selecting a smaller number of EN-sections per test zone, the varying differences between simulation and measurement of EN-sections in the compilation increase the scatter of the validation results. A higher number of selected EN-sections attenuate the influence of single EN-sections on validation results reducing the scattering range.

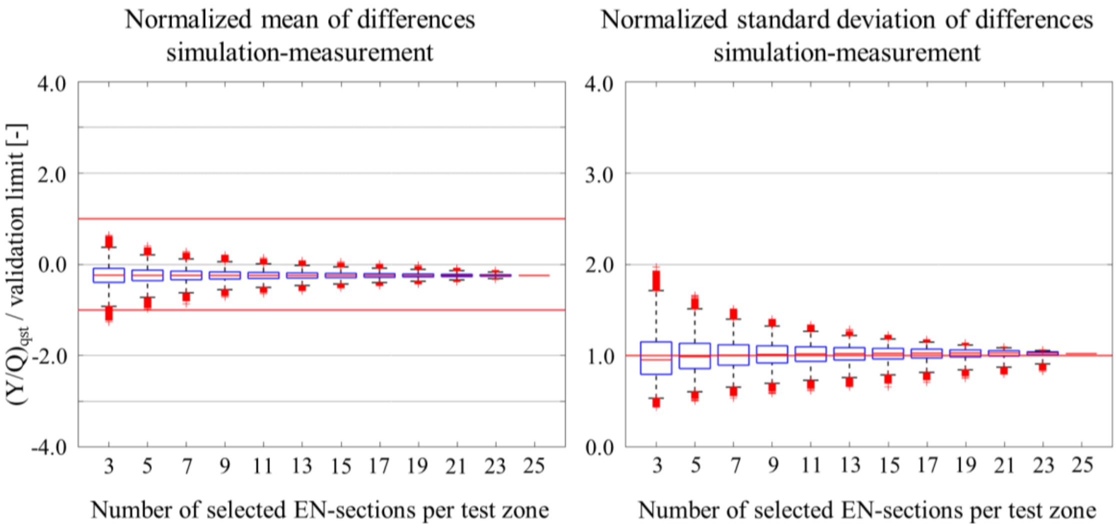

Figure 2. Exemplary influence of selected number of EN-sections on the scatter of validation results

Table 2 shows the sensitivity of the model validation results to the number of symmetrically selected EN-sections from the four test zones. In case one, the passenger coach is assessed to be validated in 98.04% of the 250 thousand combinations of three randomly selected EN-sections from the available 25 EN-sections per test zone. In the remaining combinations, a few standard deviation values of $(Y/Q)_{qst}$ are above the limit values, see Figure 3. The differences between simulation and measurement of the passenger coach are on the average in the 100 EN-sections rather small. Thus, the more EN-sections are selected for model validation, the lower the influence of these few EN-sections with higher differences on the model validation results.

Table 2. Sensitivity of the model validation results to the number of symmetrically selected EN-sections from the test zones (250 thousand combinations)

		Number of combinations for a validated model in %					
Number of EN-sections from test zone	Vehicle	Passenger coach		Freight wagon (empty)		Freight wagon (loaded)	
1, 2, 3, 4	Case	1	2	3	4	5	6
3		98.04	25.74	41.44	0.00	2.13	0.00
5		99.84	21.16	36.83	0.00	0.00	0.00
7		100.00	16.55	37.75	0.00	0.00	0.00
9		100.00	12.14	36.77	0.00	0.00	0.00
13		100.00	4.69	36.64	0.00	0.00	0.00
17		100.00	0.49	36.44	0.00	0.00	0.00

A complementary research about the sensitivity of model validation result to the friction coefficient in the wheel/rail contact showed that the flange lubrication system of Italian locomotives, pulling the DynoTRAIN measurement train in Italy, was enabled (cases two, four, six), while there was no lubrication applied in other countries (cases one, three, five). As the presented simulation results are based on a constant wheel/rail friction coefficient, the model of the passenger coach is assessed to be validated in only 25.74% of combinations of EN-sections in Italy (case 2). The percentage value lowers with increasing number of EN-sections, because more EN-sections (from test zone 2 to 4) with higher differences between simulation and measurement are used for model validation. In case three, the percentage values remain for many selections rather constant. The calculated median of the 250 thousand combinations of normalized standard deviation values of $(Y/Q)_{qst}$ is almost independent from the selected number of EN-sections, see Figure 2. When selecting all available EN-sections, the validation result is slightly above the limit value and the model is assessed to be not validated. Figure 4 shows the scattering range of case 5 in which the model is assessed to be validated in only 2.13% of the combinations of EN-sections. The investigation shows that the change of the percentage of validated models by increasing the number of EN-sections from 3 to 5 or to 7 sections is rather small.

The specified minimum of three EN-sections per test zone seems to be a reasonable compromise between the effort and reliability of the model validation process.

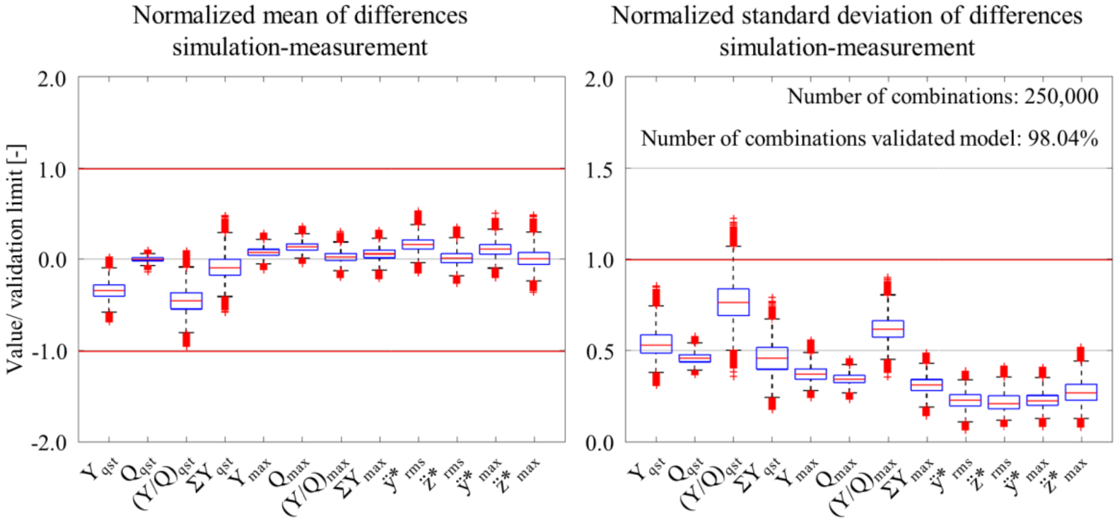

Figure 3. Scatter of validation results: Passenger coach (case 1)

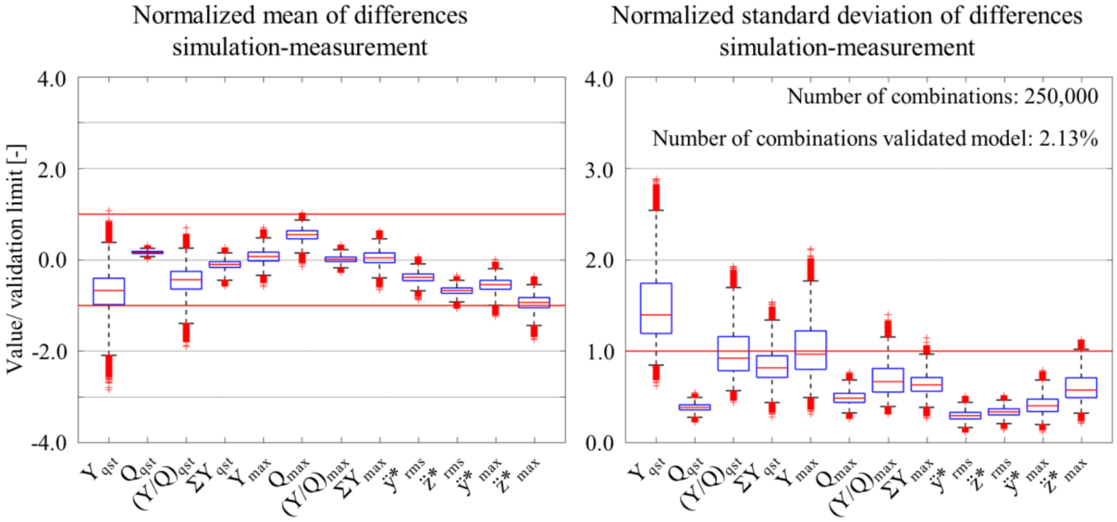

Figure 4. Scatter of validation results: Loaded freight wagon (case 5)

3.2 *Asymmetric selection of EN-sections for model validation*

As the Validation Method 2 [4] does not specify the maximum number of test sections per test zone, the user can also select a different number of test sections in each test zone. Table 3 shows the results while increasing the number of randomly selected EN-sections in the test zone 1 up to 15. In the other test zones, three EN-sections are always randomly selected. The already good result in case 1 is improved further, the more EN-sections from test zone 1 are selected for model validation. Despite the mentioned difference of the wheel/rail friction coefficient in simulation and measurements due to flange lubrication in case 2 (Italian compilation of EN-sections), the percentage values are increasing. Differences between simulation and measurements remain small because there is no flange contact in straight EN-sections (test zone 1). An opposite tendency is shown in case 3, in which the empty freight vehicle partly ran unstable while the simulation results shows no instability. The results of cases 4 to 6 demonstrate that the model validation results with clearly exceeded limit values are not sensitive to a higher number of EN-sections from test zone 1. The models remain to be assessed as not validated.

Table 3. Sensitivity of the model validation results to the number of asymmetrically selected EN-sections from the test zones (250 thousand combinations)

			Number of combinations for a validated model in %					
Number of EN-sections from test zone		Vehicle	Passenger coach		Freight wagon (empty)		Freight wagon (loaded)	
1	2, 3, 4	Case	1	2	3	4	5	6
3	3		98.04	25.74	41.44	0.00	2.13	0.00
5	3		99.73	59.14	40.14	0.00	3.62	0.00
7	3		99.98	86.66	36.58	0.00	4.85	0.00
9	3		100.00	97.64	31.59	0.00	5.85	0.00
11	3		100.00	99.72	28.90	0.01	6.14	0.00
13	3		100.00	99.97	24.53	0.04	5.54	0.00
15	3		100.00	100.00	20.04	0.14	4.19	0.00

4 CONCLUSIONS

The validation approach proposed in DynoTRAIN project and integrated as Validation Method 2 in EN 14363:2016 [4] provides not only the validation metric but also the specification of the measured signals, the test sections as well as the validation limits to be used in the validation process. Nevertheless, as in any validation method, the final validation result depends on the selection of test sections used in the validation process. This paper presents outcomes of unique investigations of millions of possible test section combinations regarding the sensitivity of the final validation result to the selection of the test sections applied in the validation process. The presented investigation shows that the specified minimum of three test sections per test zone is a reasonable compromise between the effort and reliability of the validation. The selection of test sections can certainly change the overall validation result. However, if the validation criteria show a sufficient margin to the validation limits, then the model can be considered as unambiguously validated, independently of the selection of test sections. The evaluation of asymmetric selection of test sections using different number of test sections from different test zones shows that the probability of a successful validation can occasionally change significantly if the number of test sections from one test zone is much higher than from other test zones. Thus, it is recommended to use the same number or at least similar number of test sections from all test zones. Once a model revealed to be not validated, the exceedance of the limit values should be analysed to assess the required improvement of the applied simulation model and parameters. It is not very likely and worthwhile to select other test sections for validation expecting that the overall validation result would change from “not validated” to “validated”.

5 ACKNOWLEDGEMENT

This investigation used data from the research project DynoTRAIN (2009-2013) supported by the European 7th Framework Program, contract number: 234079. The research presented in this paper received no specific grant from any funding agency or industrial company.

6 REFERENCES

[1] Polach O. and Böttcher A. 2014. *A new approach to define criteria for rail vehicle model validation*. Vehicle System Dynamics; 52:sup1: 125-141.

[2] Polach O., Böttcher A., Vannucci D., Sima J., Schelle H., Chollet H., Götz G., Garcia Prada M., Nicklisch D., Mazzola L., Berg M. and Osman M. 2015. *Validation of simulation models in context of railway vehicle acceptance*. Proc IMechE Part F: J Rail and Rapid Transit, 229(6): 729–754.

[3] Götz, G. and Polach, O. 2017. *Verification and validation of simulations for rail vehicle certification*. First International Conference on Rail Transportation, Chengdu, China, July 10-12, 2017

[4] EN 14363 2016. *Railway applications – Testing and simulation for the acceptance of running characteristics of railway vehicles – Running Behaviour and stationary tests.* CEN, Brussels.

[5] Götz, G. and Polach, O. 2017. *Influence of varying the input parameters on model validation results*. Proc IMechE Part F: J Rail and Rapid Transit, 231(6): 598-609 (Special Issue of the Stephenson Conference 2015)