MET'2015 Croatia
Trogir, October 4 - 7, 2015
12th International Conference
Modern Electrified Transport

SIMPLIFIED VALIDATION OF SIMULATION MODELS OF ELECTRIC TRACTION NETWORKS

Jacopo Bongiorno, Andrea Mariscotti
DITEN – Università di Genova, Italy

Abstract
The validation of an electric traction network simulator is addressed by selecting testing conditions and considering three performance indexes that evaluate amplitude, slope and other curve features. A common interpretation scale is proposed to verify their agreement. Moreover, since the most complex cannot be fully used without a graphical representation of its output, this is simplified, in order to conclude the validation with an estimate of accuracy that is as close as possible to the classical declaration that accompanies instrumentation.

1. Introduction

The use of simulation tools aims at replacing experimental methods and measurement campaigns with significant savings in terms of time and cost. Simulation allows also verifying configurations that cannot be reproduced in practice, but represent exceptional worst-case events and parameter combinations. The evaluation of electrical interoperability by simulation was already accepted and suggested [1-5] for the following phenomena:

- the useful voltage, i.e. the average pantograph voltage available when absorbing traction power, calculated per train or per network area;
- the power factor and displacement factor for ac systems, with the same meaning used in industrial supply networks;
- harmonics and inter-harmonics, caused by the interaction of distorting loads and generators, namely trains during power absorption, and the same trains during braking and electric substations;
- dynamic interaction between trains and the supply network, with possible electrical instability, resonances, supply distortion and considerable reactive power flow.

Especially when safety decisions are taken based on the simulation results, simulation models need to be validated, establishing their adequacy, accuracy, range of validity, as any other instrument [6]. Models of electric networks and electric equipment share a similarity of representation with the physical system, that ensures plausibility of model outcome. Despite the electrical equations of each sub-circuit and cell are in principle simple (e.g. component and Kirchhoff equations), the interaction of the many network elements and parameters is very complex and results in overall non-linear relationships, that are hardly treated in closed form or with analytical methods.

The validation of a simulation model aims at verifying that it meets its intended use, in terms of overall requirements and user's expectations and that it is used within its domain of validity. Besides the verification and validation of single modules during simulator development and the matching of model basic attributes and relations with the physical system, the most relevant part of the validation process is represented by the characterization of the accuracy of model output with respect to reference data. The validation of a simulator using dynamic techniques is performed by executing test runs on reference cases.

Simulated and measured data shall be compared, covering various configurations and parameter combinations, and including a preliminary evaluation of the quality of experimental data (in terms of their metrological characteristics). Data may for clarity assumed as electrical quantities (e.g. voltage, current, impedance) considered as frequency-domain spectra [4][5], normally characterized as amplitude and phase response.

When comparing simulated and experimental data of this kind, several features normally catch the observer's eye and may be used to quantify the degree of similarity [8-10]. The shape of the curves and the relevant distinctive elements (e.g. frequency and amplitude of resonance peaks and anti-peaks, slopes, etc.) orient the choice towards specific performance indexes [10], preferable for several reasons: robustness to noise, adequate response for peaks and slopes, ability to cope with a relatively large uncertainty of experimental data.

More than one performance index is useful to cross-check indexes themselves and avoid biasing and distortion of judgment and validation results. Comparing performance index values is of course not so straightforward, because they have different ranges and different sensitivities to curve characteristics: they were tested extensively on sample curves in [9].

This work, starting from past analyses of performance indexes [9][10], addresses how to obtain a clear simple statement of accuracy of a railway traction network simulator. Despite the extended literature about validation of simulation models, no clear and unique interpretation of results can be identified. The objective of this work is twofold: compare results of a more elaborated index like Feature Selective Validation (FSV) with simpler indexes such as Theil and Mod. Pendry, to verify the degree of agreement; propose a simple method to consider validation results as the outcomes of an experiment indicates an error or deviation in the comparison of the curves, characterized e.g. by a mean value and dispersion, or by a confidence interval.

To this aim in Section 2 the Verification and Validation process is presented, together with a description of indexes used in this work. The followed validation process in the specific case is described in subsection 2.4. The real system which the measurement data refers to is the test ring in Velim described in Section 3. The validation outcomes are reported and discussed in Section 4.

2. Verification, Validation and Accuracy

2.1 Verification

The verification of a model is the evaluation of the correspondence to the requirements, even for single modules during their development. The validation of a simulation model aims at verifying that it meets its intended use, in terms of overall requirements and user's expectations. The verification phase considers intermediate elements by means of static analysis (inspections and reviews) and possibly dynamic analysis (execution of test runs maybe assisted by synthetic data). By the way it is what is usually done when developing new functions and libraries, to be released and integrated in the rest of the simulation program. The verification phase is more related to software development and its behaviour and for this reason is not considered further.

2.2 Validation

Validation begins with the determination of the model type, its basic attributes and the relations with the system to be modeled. Then, when the model is being formulated and implemented, the model is validated by itself considering the expected model behaviour. Models that incorporate a more or less detailed representation of the physical system (e.g. distributed or lumped parameter equations of an electric network), share with it the overall behaviour. So, the most relevant part of the validation process is represented by the characterization of model accuracy with respect to reference data by executing a series of test runs.

Experimental data themselves are affected by measurement errors and thus characterized by uncertainty. The term "error" is to be intended with a broad meaning, including internal and external noise, offsets and fluctuations, and any relevant system and operating condition that may cause various types of aberrations and outliers. The resulting simulator accuracy values justify that the smaller instrument uncertainty is left aside to simplify the process.

The second relevant aspect is the completeness of the description of the physical system in terms of the selected data. This is related to the accessibility of system variables and to the identification of relevant ones. When selecting test conditions, the job is well done if they represent all plausible system configurations, including transient and exceptional configurations that can however occur in practice, excluding unrealistic ones.

Third, electric power networks, and in particular electrified railway traction systems, are complex systems with nonlinear responses and scarcely known or highly variable parameters. For example for track parameters (e.g. rail-to-earth conductance) different degrees of approximations may be adopted and are expected to be satisfactory, depending on the use of the results (e.g. interference and track circuit tuning, electrical safety, etc.).

2.3 Adequacy and accuracy

Model adequacy and accuracy are evaluated by means of performance indexes that measure the degree of similarity between simulation and experimental data, i.e. the distance between two vectors, ***o*** (simulation output) and ***m*** (measured data). Different types of distances may be applied in general, the most common weighting the amplitude and slope error using absolute value, root mean square, various weighted averages, etc.. Inspecting visually the results and basing the judgment upon this has its strong and weak sides:

- the eye concentrates on peak positions and slopes, ignoring exact values; visual evaluation selects the most relevant behaviour and trend, rejecting many details with adverse influence;
- the amount and organization of data may be too large and complex to be compared visually with ease and in this case selection and feature extraction shall be implemented.

When evaluating the correctness and adequacy of a simulation model, the judgment is based on the comparison of many results, which shall be correctly weighted and combined in order to get a unique answer that represents in the best way model adequacy and accuracy. Since this is hardly possible, combinations of graphical and numerical means are used, such as histograms, mean values, dispersions or spreads.

The proposed performance indexes (or indexes, simply) were evaluated and tested in [9][10: Theil U [11, Modified Pendry R_{PL} [12 and FSV (Feature Selective Validation) [13 with ADM (Amplitude Difference Measure) and FDM (Feature Difference Measure) have increasing complexity, the former evaluating only amplitude error, the second introducing curve slope by means of the normalized prime derivative and the latter fully addressing dc, low and high frequency curve content, evaluating amplitude, slope and convexity and displaying the results also as histograms suitable for graphical interpretation. However, FSV is also prone to some variability due to data noise and in general some implementation ambiguities investigated.

Equations are many and are not all shown here: only the main ones are reported, to support understanding of validation results and related comments. In all the following formulations the data set N-sample long and $\boldsymbol{o}$ represent the simulated data set and $\boldsymbol{m}$ the measured data.

U is the Theil index focusing on amplitude comparison; R_{PL} is the Modified Pendry Index [9][10], where x is the generic i-th element of $\boldsymbol{o}$ or $\boldsymbol{m}$ and x' its first derivative.

$$U = \sqrt{\sum_{i=0}^{N-1} (o_i - m_i)^2} \Big/ \left[\sqrt{\sum_{i=0}^{N-1} o_i^2} + \sqrt{\sum_{i=0}^{N-1} m_i^2} \right] \quad (1)$$

$$R_{PL} = \sum_{i=0}^{N-1} \left| L_{oi} - L_{mi} \right| \Big/ \sum_{i=0}^{N-1} \left(\left| L_{oi} \right| + \left| L_{mi} \right| \right) \qquad L = \frac{x'}{x} \quad (2)$$

FSV indexes ADM and FDM have more complex expressions: a straightforward simplified interpretation is that they compare amplitude and slope, respectively, of $\boldsymbol{o}$ and $\boldsymbol{m}$ data sets. Subscripts "dc", "lo" and "hi" indicates respectively the dc, low-frequency and high-frequency portions in which the data sets are partitioned using a Fourier transform/anti-transform pair, as explained in the IEEE Std. 1597.2 [13].

$$ADM_i = \left| \frac{\left| o_{lo,i} \right| - \left| m_{lo,i} \right|}{\frac{1}{N} \sum_{j=0}^{N-1} \left| o'_{lo,j} \right| + \left| m'_{lo,j} \right|} \right| + \left| ODM_i \right| \exp^{\left| ODM_i \right|} \quad (3)$$

ODM is an index used to consider only the dc behaviour of data sets and included within ADM.

$$ODM_i = \frac{\left| o_{dc,i} \right| - \left| m_{dc,i} \right|}{\frac{1}{N} \sum_{i=0}^{N-1} \left| o_{dc,i} \right| + \left| m_{dc,i} \right|} \quad (4)$$

$$FDM_i^1 = \frac{\left| o'_{lo,i} \right| - \left| m'_{lo,i} \right|}{\frac{2}{N} \sum_{j=0}^{N-1} \left| o'_{lo,j} \right| + \left| m'_{lo,j} \right|} \quad (5)$$

$$FDM_i^2 = \frac{\left| o'_{hi,i} \right| - \left| m'_{hi,i} \right|}{\frac{6}{N} \sum_{j=0}^{N-1} \left| o'_{hi,j} \right| + \left| m'_{hi,j} \right|} \quad (6)$$

$$FDM_i^3 = \frac{\left| o''_{hi,i} \right| - \left| m''_{hi,i} \right|}{\frac{7.2}{N} \sum_{j=0}^{N-1} \left| o''_{hi,j} \right| + \left| m''_{hi,j} \right|} \quad (7)$$

$$FDM_i = 2 \left| FDM_i^1 + FDM_i^2 + FDM_i^3 \right| \quad (8)$$

All FSV indexes are composed of N elements and indicated with subscript i; their mean value is indicated by the index name (e.g. ADM or FDM), without any subscript.

2.4 Adopted validation procedure

Based on the considerations made in the previous sections, the adopted validation procedure is as follows (see the flowchart in Fig.1):

- experimental data are identified for positions and configurations that are deemed relevant, critical for some parameters, or conversely ideal to remove (or attenuate) some undesirable influence, paying attention to correctly cover as many system states as possible, reaching the widest system representation possible;
- experimental data are collected performing a first check for plausibility and data quality, e.g. deleting data with evident mistakes, inexplicable behaviour, exaggerated noise and outliers;
- remaining data are then evaluated in terms of their statistical properties, deriving an estimate of the amount of outliers, a robust mean (if needed), their dispersion and their uncertainty;
- whenever feasible and advisable experimental data are smoothed to remove noise and small artefacts, with extreme care not to remove relevant features, instead;
- simulations are run for the same configurations and cases, selecting the approach to handle uncertain parameters, e.g. variable or unknown parameters: usually extreme and average values are combined with three points per parameter to keep the number of simulations compact, but more articulated approaches may be followed, e.g. using random number generators, assuming statistical distributions and running Monte Carlo simulations;
- the comparison is first performed visually as a general check, looking for confirmations to the assumptions made when identifying the relevant configurations; whenever weird behaviour or inexplicable differences are found, corrective actions are performed, such as checking again the specific data, the input parameters of simulations, etc.; it is hard to believe how many details may have impact when small differences between curves are relevant [14], such as e.g. wheel slip, small instability or deviations of the fundamental frequency, errors in taking the train position, considering that about 70 m were between the head and the tail;
- then, performance indexes are applied one at a time and the results are discussed and compared, to find agree-

ment between the different indexes and which parts of the curves are responsible for the most significant error, as measured by a specific index.

The attention is focused on the values of the two simpler Theil and Mod. Pendry indexes and on the resultant histograms of the FSV indexes, besides their statistical quantities. Histograms have been conceived to mimic the evaluation of a pool of experts and to represent any possible inconsistency of some of their judgments.

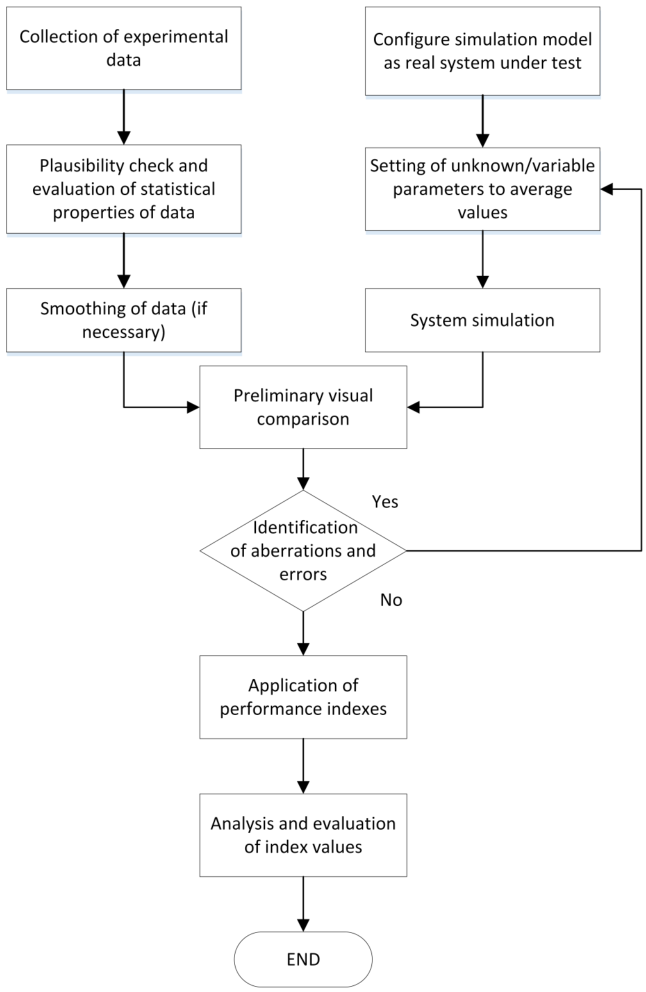


Fig. 1 – Adopted validation procedure.

3. Velim test ring

After measurements in 2012 and 2013 along high-speed and conventional lines in some countries (e.g. Italy, Belgium, France), a final thorough validation was made using a well-known line that is limited in extension and where the traffic is under control, the ring at the VUZ test centre in Velim, Czech Republic. The Velim test ring is well maintained and the personnel know possible critical deviations from nominal (or normal) values, because tests are done routinely. Additionally, the test circuit was studied extensively for a few days before the tests, thanks to the received support and also to its limited extension.

The test centre features two single-track test rings: the smaller one has a length of 6 km and is surrounded by the bigger one that is about 13.2 km long. The inner ring was sectioned during tests and is not included in the model. The big test ring has a cross section with six conductors. Rails are connected together (transversal bonding) every 300 m, except in track-circuit testing area between chainage km 10.672 and 11.672; the two catenary (positive) feeders are connected with contact line and messenger every 120 m.

The test ring was connected to the 25 kV 50 Hz substation (single-phase transformer, fed by the High Voltage national grid) and a train consist (two wagons trailed by a BB36000 Alstom loco) travelled along the ring with acceleration/ cruising/coasting/braking test patterns in the two straight parts of the ring. Positions with satisfactory current spectrum intensity were selected (see Fig.2): four in section A (A_1=km 8.3, A_2=km 8.0, A_3=km 7.4, A_4=km 7.1) and four in section B (B_1=km 1.6, B_2=km 1.3, B_3=km 1.0, B_4=km 0.4).

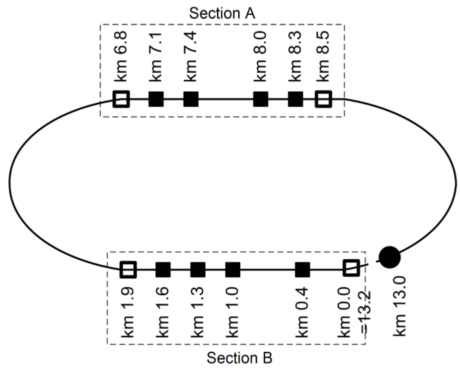


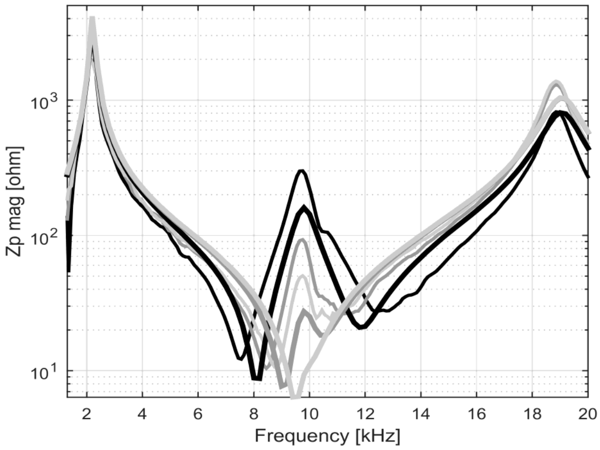
Fig. 2 – Scheme of the Velim test ring with supply substation (circle), test positions (filled squares) and other reference positions (hollow squares).

The line model, as for other study cases [8, includes frequency-dependent inductance and losses of the return circuit elements and stray parameters, in particular rail-to-earth conductance, considering the effect of the interconnection of the outer ring track, the inner ring track, cable screens and the negative pole of transformer secondary.

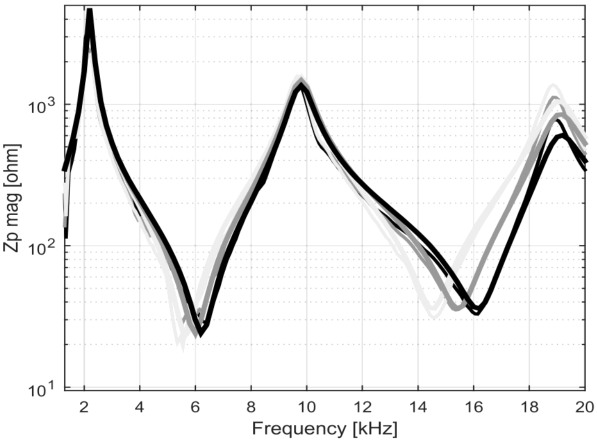
Additionally, the influence of the HV feeder is difficult to model because only the nominal values are known, and a rough estimate of the 110 kV network indicates the possibility of resonance effects in the same frequency range where the supply transformer has its short-circuit resonance (i.e. around 10 kHz).

4. Validation results

Some sample results are shown to support the reasoning that follows on simplification and comparison of the similarity performance results. These results have been extensively reported in [7-10]. The visual comparison between simulated and measured amplitude of pantograph impedances Z_p in different position is shown in Fig. 3



(a)



(b)

Fig. 3– Z_p comparison for test section A and B, at the four different considered locations (1-4 from darker to lighter line), measured (thin line) and simulated (thick line).

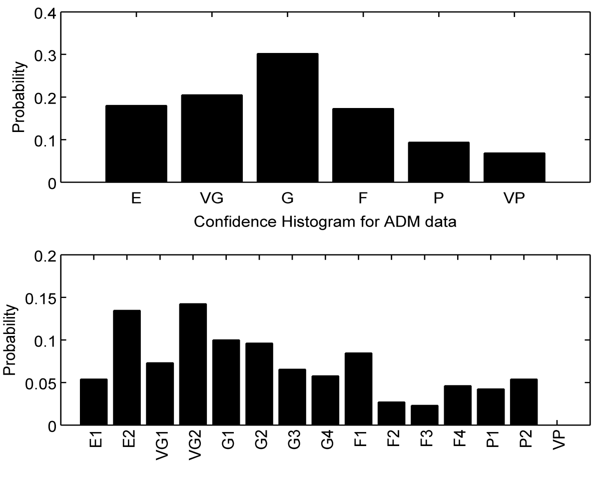
FSV histograms are shown in Fig.4 and Fig.5 for ADM and FDM. The original histograms [13] have six classes; further classes were added for a better representation and these new histograms are visible in the lower part of each figure.
The correspondence between numeric values of ADM and FDM and interpretation categories is shown in Table 1.

Table 1 – Interpretation for FSV, Theil and Pendry indexes

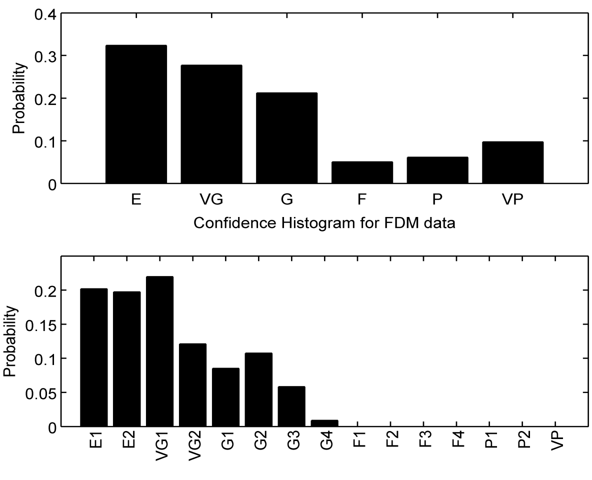
Quality descriptor	FSV		Theil and Pendry	
	Lower bound	**Upper bound**	**Lower bound**	**Upper bound**
Excellent (E)	0.0	0.1	0.0	0.1
Very Good (VG)	0.1	0.2	0.1	0.3
Good (G)	0.2	0.4	0.3	0.5
Fair (F)	0.4	0.8	0.5	0.7
Poor (P)	0.8	1.6	0.7	0.9
Very Poor (VP)	1.6	+∞	0.9	1.0

As confirmed by ADM and FDM mean values in Table 2, the histogram of the former is more disperse and the highest bar is in the Good category, while for FDM the histogram is centered at Excellent and Very Good.
The interpretation of histograms is not easy, with the histogram shape giving an indication of the reliability of the highest bar: nearly flat histograms indicate an unreliable estimate and that the model curve is both well and badly matching experimental data depending on curve portions; in general, to ease interpretation, bars lower than one third of the highest one may be disregarded.
As an additional comment regarding the significance of FSV results, it is observed that their dispersion is always higher than that expressed by evaluations of human experts.

Histograms with more than six categories were created to better support the interpretation of FSV results; following the "one third" criterion, data falling in lower bars of the original histogram are removed, before creating the new denser histogram. Attention is focused on the most relevant classes for model validation, E, VG, G and F, divided into 2, 2, 4 and 4 subcategories, respectively; the remaining P is divided into only two and the VP class is left unaltered. The reason is simply that a model that fall in the Poor and Very Poor categories does not need to be analysed further and shall be corrected or rejected.

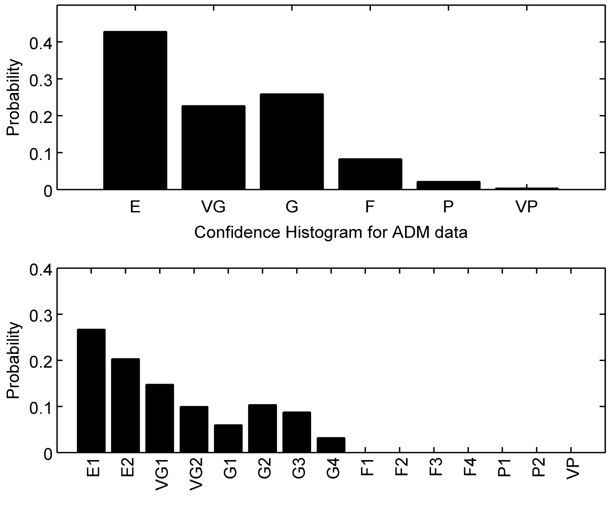


(a)

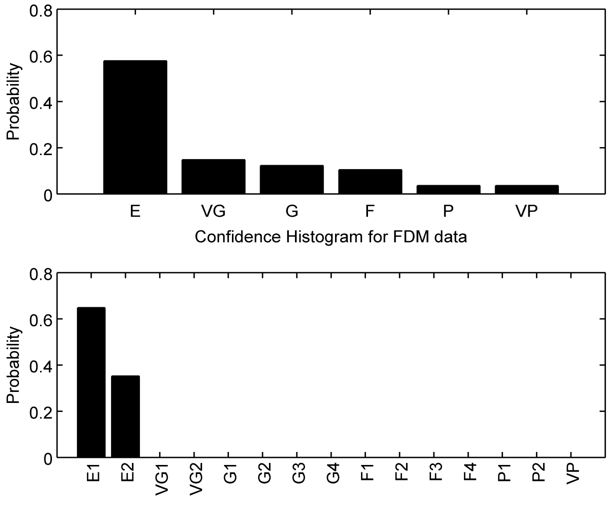


(b)

Fig.4 – Sample confidence histograms of FSV results (a) ADM and (b) FDM in test section A, with basic and refined categories.



(a)



(b)

Fig.5 – Sample confidence histograms of FSV results (a) ADM and (b) FDM in test section B, with basic and refined categories.

Observing Fig.4 and Fig.5 and the original FSV interpretation categories, the new subcategories amount to 5% each for E, VG and G, 10% for F, 20% for P.

As an overview of the complexity of the validation process we synthesize the degrees of freedom and the size of vectors and data sets:

- initially a few parameters of the power supply system (stray parameter of feeding cables and HV grid equivalent circuit) and traction line (rail to earth conductance) underwent sensitivity analysis and were then fixed to average values compatible with all measurements;
- measurement positions are eight, each with a number of collected traces around one hundred, transformed in Fourier spectra of suitable frequency resolution;
- the adopted performance indexes may simply give a similarity value (Theil and Mod. Pendry) or a set of values as for the FSV method, describing amplitude and feature similarity, both numerically and as histograms, conglobated then in the GDM (Global Difference Measure) index.

FSV results are processed for mean value, dispersion and skewness of ADM and FDM; GDM global index was not used because unavoidably it masks the relative contributions of the two. To this aim the mean value of the “pruned” FSV histograms is used. For Theil and Mod. Pendry only one value results for a vector, interpreted as the mean similarity index, without dispersion information.

Besides the statistical characterization of the overall index values, in case of FSV also local considerations and evaluation can be done, identifying which curve portions give best and worst scores, both in amplitude (ADM) and in shape (FDM). In this way, the maximum local error can be estimated e.g. when working on safety-relevant applications, requiring that the error is upper bounded.

Eight measurement positions divided between section A and B are considered: depending on their distance from the substation, they are affected differently by the limited information on the High Voltage grid transfer function, that couldn’t be measured for obvious safety reasons.

ADM may be compared with Theil and Mod. Pendry indexes: the ADM mean value is in general between the Theil and Mod. Pendry values or in a few cases slightly outside, but always much closer than its dispersion, that is within the confidence interval roughly defined by the standard uncertainty (i.e. one standard deviation). All the results are thus in agreement, indicating the similarity between analysed curves as Good or Very Good. Both Theil and Mod. Pendry indexes indicate the higher differences as that in curve A3 and A4; which are affected by the strongest influence of the HV grid for which the model is approximated.

By comparing ADM and FDM values, it appears that the differences between curves are mainly in amplitude, while features (e.g. slopes) are correctly modeled.

Table 2 – Synthesis of performance indexes for positions A1...A4, B1...B4; ADM and FDM are characterized by mean value, (dispersion) and [skewness].

Pos.	rms error	Theil	Mod. Pen-dry	ADM orig.	ADM	FDM
A1	0.438	0.186	0.279	0.126	0.233 (0.186) [1.391]	0.037 (0.026) [0.469]
A2	0.430	0.152	0.275	0.117	0.212 (0.185) [1.572]	0.035 (0.025) [0.421]
A3	0.399	0.259	0.305	0.115	0.347 (0.340) [1.643]	0.073 (0.050) [0.598]
A4	0.380	0.269	0.330	0.116	0.366 (0.346) [1.679]	0.077 (0.055) [0.573]
B1	0.181	0.135	0.174	0.086	0.138 (0.099) [0.986]	0.029 (0.022) [0.707]
B2	0.198	0.280	0.199	0.104	0.252 (0.199) [1.087]	0.068 (0.056) [0.746]
B3	0.189	0.245	0.190	0.097	0.238 (0.192) [1.118]	0.037 (0.027) [0.408]
B4	0.202	0.249	0.207	0.101	0.244 (0.187) [0.948]	0.043 (0.027) [0.245]

The objective stated at the beginning is to propose a procedure that, despite using complex and multi-faceted performance indexes, is able to give a simple answer as for measuring instruments of medium complexity: two different approaches are considered.

First, in order to use performance index results to establish the degree of accuracy of the simulator, agreement between indexes that have different underlying criteria is verified by using a common ground for comparison: the interpretation scales. To this aim, the values of ADM and FDM indexes and Theil and Pendry indexes are classified using the interpretation scale of Table 1; for better accuracy the subintervals resulting from refined categories may be used. Then classifications are compared in order to, first, determine the degree of agreement between indexes and, second, derive a more or less quantitative evaluation of simulation error. FSV is considered the preferred method, with Theil and Pendry indexes are used for confirmation.

The maximum ADM error is, as expected, in location A4, where the effect of the approximate model of the HV network is significant. The mean value is 0.366 with a spread of local error values between 0.02 and 0.71. For all other locations the improvement is by about a factor of 2, so that the overall average mean ADM is 0.254, that is between Good and Very Good.

FDM error estimates are much better and slopes and other features are substantially well modeled: the worst location is again A4 with a mean FDM of 0.08 and a spread of local error values between 0.021 and 0.2. In nearly 50% of locations the improvement of FDM is again by a factor of 2. The overall average mean FDM is 0.05 that is Excellent.

ADM is larger than Theil index and comparable to Mod. Pendry; the reasons are many and are still under investigation: ADM (and FDM) is not calculated on the original data vector but on transformed and anti-transformed version, that is affected by leakage and aberrations; Mod. Pendry includes slope with the first derivative. Direct comparison is thus difficult. To this aim Table 2 reports a column called “ADM orig.”, where the ADM operator is applied to the original data ignoring FSV procedure: mean values are all smaller and give a more reliable indication of the true error in classical sense as interpreted by ADM operator.

In another work under review it is shown that these performance indexes, when evaluating hypothetically perfect data in the Excellent category, tend to overestimate differences when measurement data noise and uncertainty are considered, thus causing a systematic error, that may be as large as a few %. This systematic error reduces the effective interval of the Excellent class, resulting in worse judgment and slightly pessimistic results of cases falling in the best categories.

The use of interpretation scales allows the comparison, but misses the objective of comparing and interpreting index values in a quantitative way with a close relationship to classical error metrics. Thus the problem is attacked by analyzing the structure of the index expressions.

For Theil index the denominator is the sum of the rms of the two data sets except for the missing normalization by the number of points; the numerator corresponds to the rms difference of data sets again without the normalizing number of points, that compensate.

ADM is similar, but is a linear operator and applies to the low frequency and dc portion of vectors identified by “lo” (excluding for simplicity the contribution of “dc”). The absolute deviation between curves at numerator is divided by the mean absolute value of the two curves obtaining mean absolute deviation (MAD).

Indexes compare the curves without selecting one for reference, so the correct normalization is the half-sum of rms or mean absolute value of the two curves $\boldsymbol{o}$ and $\boldsymbol{m}$. This introduces a further multiplicative factor of 2 for Theil and ADM indexes to translate them into estimates of rms and absolute deviation (and to ADM dispersion too).

However, two important facts shall be underlined, that prohibit the direct application of the two indexes (Theil and ADM) for a calculation of a classical error, that is rms or mean absolute deviation:

- the correct formulation of rms error requires that the difference (o_i-m_i) is normalized by the local value (o_i+m_i)/2, not that the whole sum of squares of errors is divided by the overall square values of ***o*** and ***m***: in this case Theil is only an approximation of the rms criterion;

- as said, ADM calculates the absolute deviation, not of original ***o*** and ***m*** vectors, but of o_{lo} and m_{lo}, that are affected by inaccuracy due to leakage and other aberrations during the transformation and anti-transformation process; the direct application of ADM as a MAD operator to the original vectors ***o*** and ***m*** gives values equal to or smaller than 50% of ADM value.

5. Conclusions

This work has considered the validation of a simulator for electric traction network against experimental data by using and comparing three different performance indexes, Theil, Modified Pendry and FSV. The effort is twofold: trying to express the similarity scores of the indexes in a way so that they are comparable and quantifying the similarity error between model and measurement data so that it may be used to characterize the accuracy of the simulator, as if it were a measuring instrument.

Theil and Mod. Pendry indexes have a 0-1 output scale, but evaluate different characteristics of the curves, being based on the amplitude comparison the former and the normalized derivative comparison the latter. FSV method does not fit a 0-1 scale in principle but observing the very bad similarity cases are of no use, the output scale may be mapped in a 0-1 range. Interpretation categories may be used for comparison.

A direct relation of indexes to classical error metrics (absolute deviation, rms, etc.) was considered and some inconsistencies identified, so that the problem is still under investigation. Using ADM on the original data, rather than on transformed data (as dictated by IEEE 1597.2 standard) indicates smaller differences. Regarding the validation of *traXsim* simulator FDM index that measures slope and concavity confirms an excellent correspondence with experimental data. On average the classical error may be estimated around 15-20% over the 1.6-20 kHz range (about 1% at fundamental, as shown years ago in a publication based on measurements on the Italian Alta Velocità [16], but a quantitative correspondence with the proposed performance indexes is still under investigation.

Acknowledgments

The authors wish to warmly thank the support of VUZ (Výzkumný Ústav Železniční a.s.) personnel at the Velim Test Centre, Czech Republic, and in particular Mr. Karel Peška and Mr. Aleš Stopka.

The test campaign was performed within the EU 7th Framework Program Project EUREMCO, grant agreement n. 285082 [15.

References

[1] Railway applications – Fixed installations – Requirements for the validation of simulation tools used for the design of traction power supply systems, CENELEC C20 working document, May 2012.

[2] European Union, Technical Specifications for Interoperability – Energy System, 2002/733, 30-05-2002.

[3] Railway applications. Power supply and rolling stock. Technical criteria for the coordination between power supply (substation) and rolling stock to achieve interoperability, CENELEC EN 50388 Std., Aug. 2005.

[4] A. Mariscotti, “Direct Measurement of Power Quality over Railway Networks with Results of a 16.7 Hz Network,” *IEEE Trans. on Instrumentation and Measurement*, Vol. 60, n. 5, May 2011, pp. 1604-1612.

[5] A. Mariscotti, “Measuring and Analyzing Power Quality in Electric Traction Systems”, *International Journal of Measurement Technologies and Instrumentation Engineering*, Vol. 2 n. 4, Oct.-Dec. 2012, pp. 21-42.

[6] A. Mariscotti, “On the validation of models of large complex electrical systems,” *International Journal of Measurement Technologies and Instrumentation Engineering*, Vol. 5, n. 1, Jan.-March 2015.

[7] J. Bongiorno and A. Mariscotti, “Recent Results on the Power Quality of the Italian 2x25 kV 50 Hz Railway”, 20th Imeko TC4 Intern. Symp., Benevento (Italy), Sept. 15-17, 2014.

[8] J. Bongiorno and A. Mariscotti, “Experimental validation of the electric network model of the Italian 2x25 kV 50 Hz railway,” 20th Imeko TC4 International Symposium, Benevento (Italy), Sept. 15-17, 2014.

[9] J. Bongiorno and A. Mariscotti, “Performance of indexes used for model validation,” Proceedings of CSECS 2014, Lisbon, Portugal, Oct. 30–Nov. 01, 2014.

[10] J. Bongiorno and A. Mariscotti, “Evaluation of performances of indexes used for validation of simulation models based on real cases,” *International Journal of Mathematical Models and Methods in Applied Sciences*, Vol. 9, 2015, pp. 22-43.

[11] H. Theil, *Economics and Information Theory*, Amsterdam, North-Holland, 1967.

[12] J.B. Pendry, “Reliability factors for LEED calculations,” *Journal of Physics – Part C: Solid State Physics*, Vol. 13, pp. 937-944, 1980.

[13] IEEE Std. 1597.2, *Recommended Practice for Validation of Computational Electromagnetics Computer Modeling and Simulations*, 2010.

[14] J. Bongiorno, A. Mariscotti, “Variability of pantograph impedance curves in DC traction systems and comparison with experimental results”, *Przegląd Elektrotechniczny*, Vol. 90, n. 6, pp. 178-183, 2014.

[15] European Project EUREMCO, FP7, www. euremco.eu, grant agreement n. 285082.

[16] R. Cella, G. Giangaspero, A. Mariscotti, A. Montepagano, P. Pozzobon, M. Ruscelli and M. Vanti, “Measurement of AT Electric Railway System currents and validation of a Multiconductor Transmission Line model”, *IEEE Trans. on Power Delivery*, vol. 21 n. 3, July 2006, pp. 1721-1726.