Using FSV to compare noisy datasets

J. Knockaert[1&2], J. Peuteman[2], J. Catrysse[2], R. Belmans[1]

(1) Katholieke Universiteit Leuven, Departement Electrical Engineering (ESAT)
Research Group ELECTA
Kasteelpark Arenberg 10, B-3001 Heverlee – Belgium
Tel. +32 – (0)16 / 32 10 20

(2) Katholieke Hogeschool Brugge Oostende, Departement IW&T
Zeedijk 101, B-8400 Oostende – Belgium
Tel. +32 – (0)59 / 569 000
Email: jos.knockaert@khbo.be

Abstract: For noisy datasets, as EMC peak measurements, it is difficult to distinguish real features from noise. Validation methods like FSV overestimate the difference when datasets are noisy. This paper discusses two methods to prepare noisy datasets, allowing FSV to give a reliable validation.

Keywords: FSV, feature, noise, EMC measurements

1. Introduction

FSV (Feature Selective Validation) [1-2] is a method for validation of computational electromagnetics [1], with applications in EMC and Signal Integrity. This method has shown its usefulness in the validation of EMC-models [2]. When comparing two EMC-measurements, FSV is partly usable. These measurements are characterized by noisy data, where it is difficult to distinguish a real feature from noise. Practical applications of validation of noisy EMC-measurements are correlation between conducted and radiated emission, comparison of measurement methods and influence of replacing obsolete components on a PCB on the emission.

By preparing the data, which means cancelling out noise, FSV becomes fully usable. The main problem of distinguishing real features from noise is still not 100% fully solved. All methods to cancel noise influence the signal itself. This is a known problem in all types of data processing, e.g. sound- and image processing. A second problem with cancelling noise is that decisions (cut-off borders, attenuation values, …) have to be made. This increases the subjectivity which conflicts with the FSV-philosophy. A benefit is that the area of FSV-applications is increased.

This paper contains four parts. In the first part, the problems with noisy data are shown. A possible solution by weighted values is given in the second part. In the third part, some pre-processing methods of data are discussed. The fourth part validates the pre-processed data with FSV.

2. Conducted emission measurements

Six measurements are performed on a PWM inverter, twice a peak, average and quasi-peak measurement. Equal measurements are done directly after each other, with no changing disturbing factors in the surrounding area. This means that both peak-measurements should be almost identical.

The same can be said for the average and the quasi-peak values. The three measurements are shown in figs. 1, 2 and 3. The right part of each figure shows the FSV-results. The left part shows the probability density function (PDF) of ADM, the right the same for the FDM-values. On the figures, also an overall ADM and FDM value are given. As an additional test, two datasets of white noise and two sets Gaussian noise have been generated and evaluated (fig. 4).

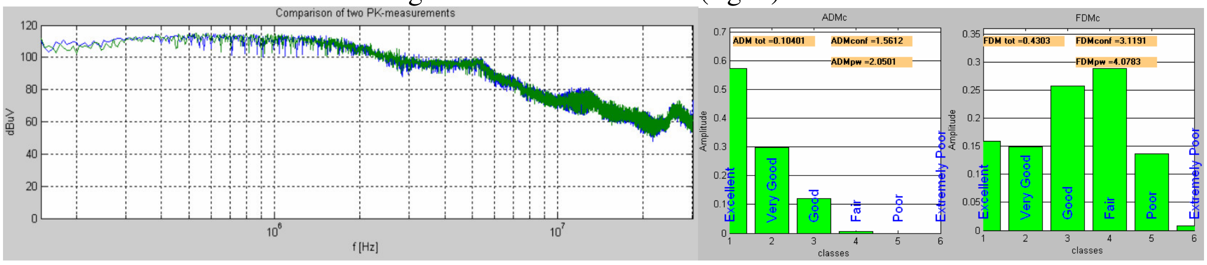

Fig. 1. Two peak measurements (top) and their difference (bottom). Validation (right).

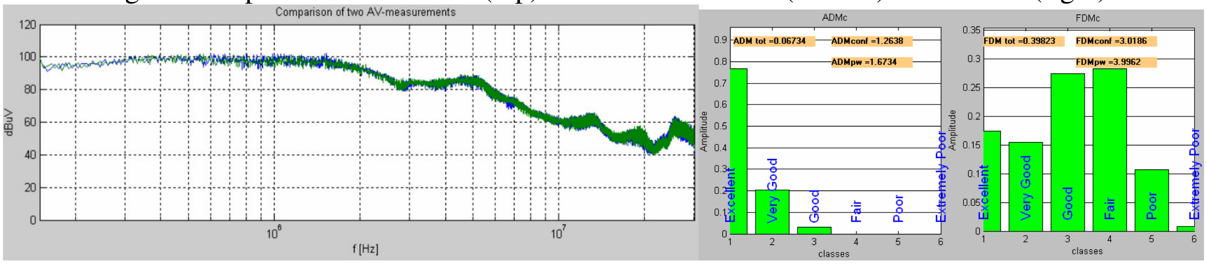

Fig. 2. Two average measurements (top) and their difference (bottom). Validation (right).

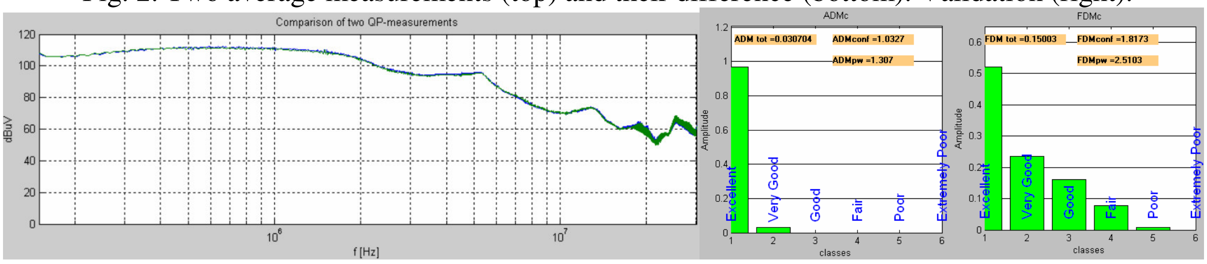

Fig. 3. Two quasi-peak measurements (top) and their difference (bottom). Validation (right).

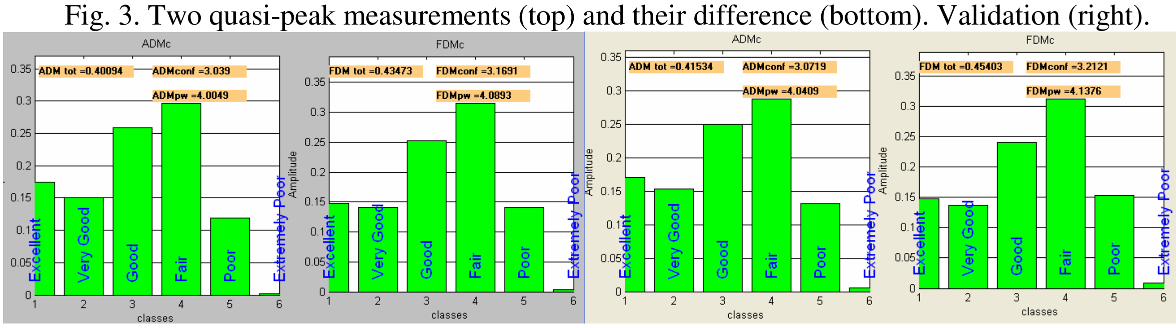

Fig. 4. Evaluation of white noise (left) and Gaussian noise (right).

Table 1. FSV-results of the measurements.

Comparison	Corr%	ADM	FDM	GDM
1. Identical	100	0	0	0
2. Peak	97.53	0.104	0.430	0.461
3. Average	98.83	0.067	0.398	0.414
4. Quasipeak	99.83	0.031	0.150	0.159
5.White noise	-0.42	0.401	0.435	0.658
6. Gaussian noise	-0.0034	0.415	0.454	0.687

Table 1 shows the correlation and results of FSV. The first row shows a comparison of identical data to have an idea of the best possible case. Comparing results, the following decisions can be taken. Correlation has no meaning in this comparison. ADM gives a valuable result, while FDM only provides a valuable result on the QP measurement. FDM makes no difference between PK, AV and both noise types. The PDF's of FDM on fig. 1, 2 and 4 are very similar.

3. Weighted values

The peak- and average measurements contain a lot of high frequency components which can not be seen as a feature but act as noise. The point-by-point global difference measure is calculated by:

$$GDM_i = \sqrt{ADM_i^{\,2} + FDM_i^{\,2}} \quad (1)$$

In comparing conducted emission measurements a good (small) ADM is combined with a bad (large) FDM [1]. The largest value (FDM) is emphasised. In noisy data, the ADM-part is almost negligible in the GDM result.

To solve this problem, there are three possibilities. To compare conducted emission measurements with FSV, quasi-peak measurements give confidential results, but they are time-consuming (10 times longer than a peak measurement). A second possibility is to change [1]. A new equation emphasising ADM, while not neglecting FDM has shown its usefulness for comparing peak measurements [4-5]. This method has been generalised in [3]. By evaluating the grade and spread of the results, the ADM and FDM are weighted to give a more accurate GDM result. The method has been tested on the previously data mentioned and on other test data as described in [5]. Third possibility is to pre-process the data in order to eliminate noise, while keeping the feature.

4. Pre-processed data

4.1. Introduction

Pre-processing data is a possible method to eliminate noise, while keeping the feature. Eliminating noise and keeping the feature is not fully possible, as this is even by visual inspection difficult. The two datasets under consideration are a dataset with only a trend and noise, but no feature (fig. 1) and a dataset with a trend, noise and an obvious feature (fig. 5).

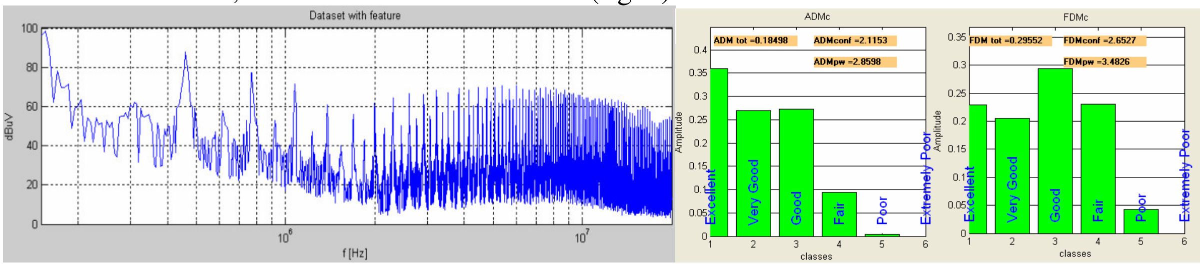

Fig. 5. Dataset with obvious feature (left) and validation (right).

The validation of the data on fig. 5 by FSV gives better results. This means that if data contains more feature, FSV recognises this feature in a better way and quantifies the results. A conclusion is that eliminating the noise obviously improves the validation. For eliminating the noise, two possibilities are investigated. The first is by detrending data and using a histogram on the noise. The second is by doing more than one peak measurement to find real feature in the noise.

4.2. Detrending and histograms

A simple moving average algorithm (SMA) calculates the trend and subtracts this from the data. A detrended dataset with noise and feature is left. EMC-measurements are expressed in dBμV. For a correct moving average of the measured data, data should be expressed linearly first. Nevertheless, applying a SMA to logarithmic data instead of linear data is the same as applying a geometric mean instead of an arithmetic mean, as shown here for an n-point moving average.

$$MA_{dB} = \frac{\sum_{i=1}^{n} A_{idB}}{n} = \frac{\sum_{i=1}^{n} 20.\log_{10}(A_i)}{n} = 20.\log_{10}\left(\sqrt[n]{\prod_{i=1}^{n} A_i)}\right). \tag{2}$$

To approximate a featureless trend, a 100-point SMA gives good results. Fig. 6 shows initial, trend and detrended data and a histogram of this detrended data. This is done for both typical datasets under consideration.

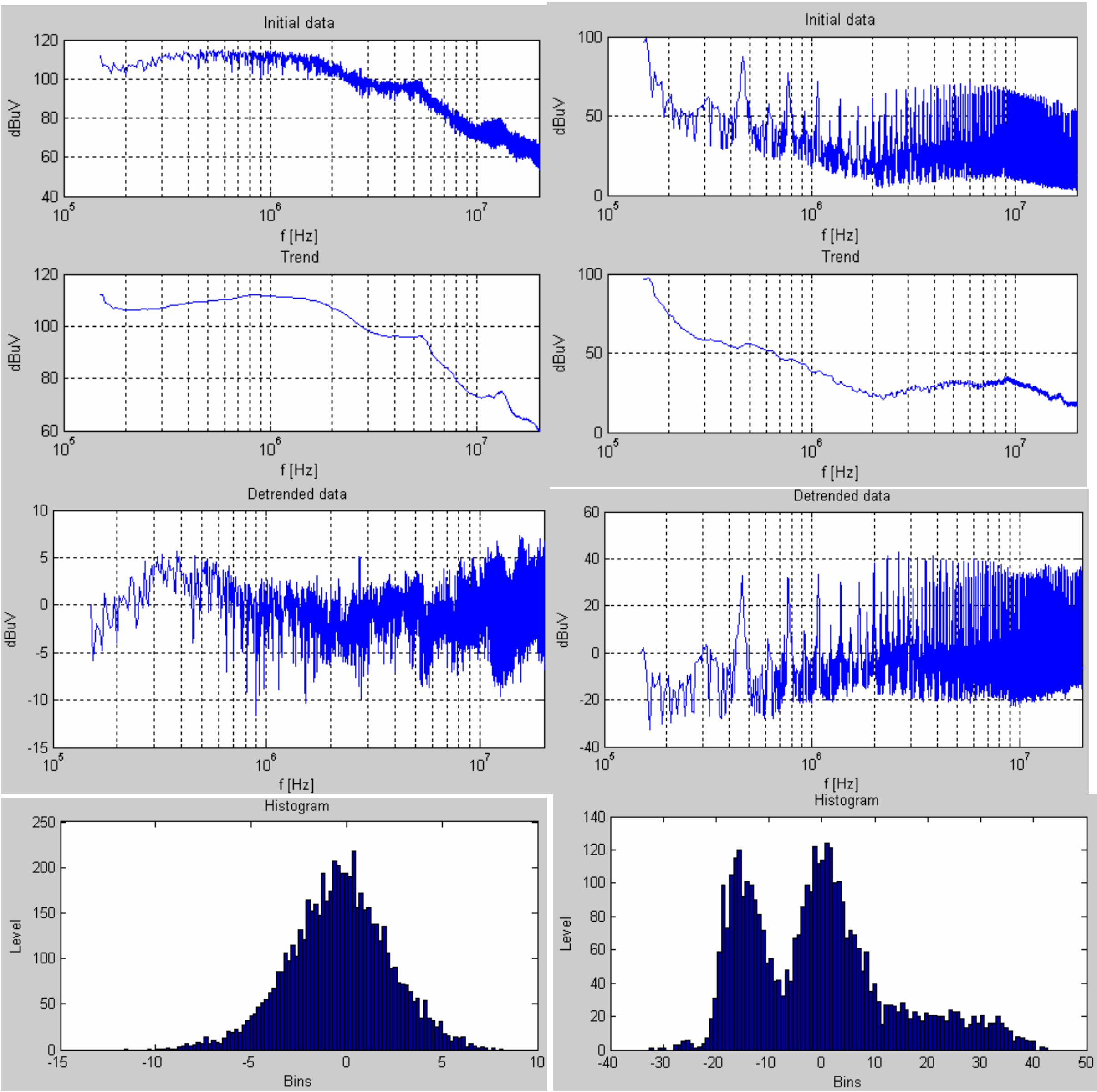

Fig. 6. Two typical datasets (left and right): top: initial data and trend; middle: detrended data; bottom: histogram of detrended data (bins in dBμV).

Visual inspection of the histograms learns that the detrended data of the left set resembles the histogram of Gaussian noise. This data can be seen as a trend line with noise and no particular feature. To eliminate noise, a left and right boundary have to be chosen wherein all data can be eliminated. For the histogram of the right dataset, it is far more difficult to choose boundaries. An obvious point is that the upper boundary can be seen. The noise goes approximately up to bin 10 or 12. The real peaks give values between 15 and 43. For the lower boundary, we can see two Gaussian curves. It is difficult to say what boundary we have to choose, -10 or -20. For EMC-measurements, the problems are related to the upper parts. The boundary of the lower parts is less important to be chosen correctly. Fig. 7 shows data after noise elimination. It is obvious that choosing a boundary is the weak spot of this method. A way to help the interpretation is by processing detrended data using a linear scale (μV) instead of a logarithmic scale (dBµV). By this the histogram becomes smaller, with some values for bins far from the histogram, indicating real feature.

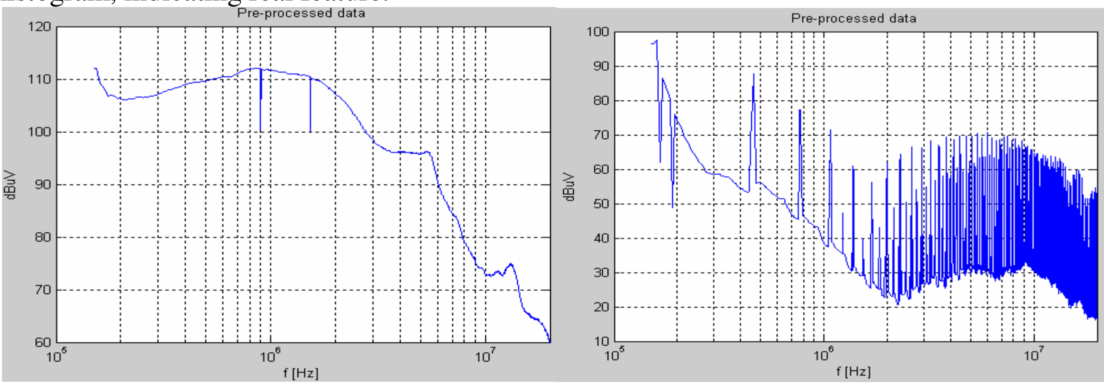

Fig. 7. Data with elimination of the noise.

4.3. Multiple peak measurements

The previous method cancels the entire noise band. This can lead to errors, as smaller peaks can be hidden in that noise band. The peaks are even by visual inspection not noticeable. A method to see such small but real features is by performing a time consuming quasi-peak measurements. As can be seen on fig. 3, the quasi-peak measurement has a very small noise band. The question arises whether it is possible to perform a peak measurement and conclude what the quasi-peak result would be? The answer is negative, as there is a lack of information on the repetitive nature of the signal. However, multiple peak measurements can give an idea of this repeatability and can cancel out part of the noise. The benefit here is that performing two or three successive peak measurements is still less time consuming than one quasi-peak measurement.

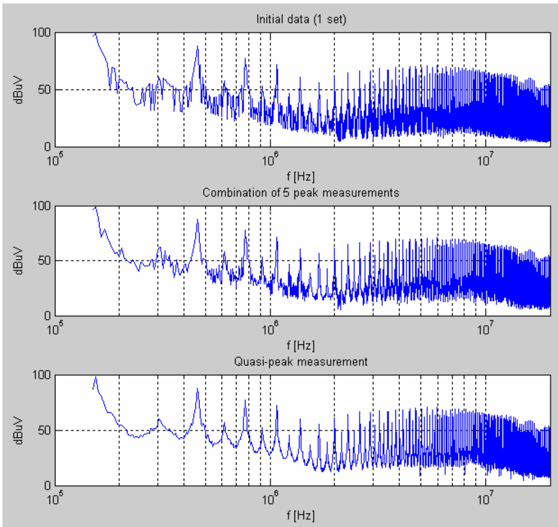

Fig. 8. Top: One peak measurement; Middle: Combination of 5 peak measurements, Bottom: Quasi-peak measurement

With pure Gaussian noise the following can be proven. With n measurements, the standard deviation decreases by a factor $1/\sqrt{n}$. This means there is indeed some noise cancellation, but this is not enough for evaluation using FSV. Applying this to real data results in the figure 8. Combining several peak measurements (five on fig. 8) it is obvious that the more combinations are used, the better the quasi-peak measurement is approximated. The combination method is a simple mean value calculation at each frequency. The advantage in comparison with the histogram-method is that no decision on e.g. borders have to be made. The major disadvantage is that for evaluation by validation methods this data is still to noisy, even with five combined measurements. Further investigation is needed.

5. Validation of pre-processed data by FSV

The various types of pre-processed data are now validated by FSV. When pre-processing, a lot of choices have to be made. Which type of moving average, linear or logarithmic, what boundaries for the histogram? Most decisions do not make a significant difference on the results. Only the boundaries have to be chosen accurately. To simplify this decision, detrended data is expressed linearly. Table 2 shows that the histogram methods give better results. There is no overestimation of the feature. Another conclusion can also be made. The more particular feature data has, the better FSV is doing the validation.

Table 2. Evaluation by FSV

	Classic FSV			Weighted FSV (grade-spread)			Pre-processed data with histograms, logarithmic noise elimination and FSV			Pre-processed data with histograms, linear noise elimination and FSV		
	ADM	FDM	GDM	weight ADM*	weight FDM*	GDM	ADM	FDM	GDM	ADM	FDM	GDM
Pk fig 1	0.104	0.430	0.461	1	0.5	0.447	0.034	0.232	0.241	0.034	0.223	0.233
Av fig 2	0.067	0.398	0.414	1	0.5	0.407	0.049	0.205	0.224	0.049	0.225	0.243
Qp fig 3	0.033	0.150	0.160	1	0.33	0.154	0.024	0.172	0.183	0.024	0.172	0.182
Pk fig 5	0.185	0.295	0.382	1	0.75	0.367	0.114	0.233	0.275	0.100	0.136	0.190
Qp fig 5	0.041	0.080	0.098	1 (0.060)	0.5	0.098	0.065	0.067	0.104	0.067	0.070	0.109
G. Noise	0.415	0.454	0.688	1 (1.569)	0.75	1.499	1.561	0.733	1.981	1.599	0.537	1.816

**Values for ADM and FDM are the same as for classic FSV, except for some particular cases, were the new value is mentioned between brackets. In the other cases, only the weight is specified in the columns.*

6. Conclusions

FSV is a valuable method for validation of model-data and measured data. To compare EMC peak measurements, FSV overestimates the difference due to the noisy nature of data. Three methods have been investigated, to overcome this problem. First is by changing the weight of the ADM and FDM when calculating the GDM. This method has shown its benefit, however, being difficult to generalize. Second method is by detrending data and using histograms to cancel noise. This method gives the best results. By expressing the feature in a linear amplitude scale, the method can be used more accurately. The third method is by using more than one measurement to cancel noise. This method has its benefit, but data stays too noisy. This method still needs further investigation.

References

[1] A. Duffy, A. Martin, G. Antonini, A. Orlandi, C. Ritota, The feature selective validation method (FSV) method, *Proc. of IEEE 2005 EMC Int. Symp., Chicago*, USA, 8-12 August, pp.272-277, 2005.

[2] G. Antonini, A. Ciccomancini Scogna, A. Orlandi, C. Ritota, Applications of FSV to EMC and SI data, Proc. *of IEEE 2005 EMC Int. Symp., Chicago*, USA, 8-12 August, pp.278-283, 2005.

[3] A. Duffy, G. Antonini, A. Orlandi, C. Ritota, Enhancing feature selective validation (FSV) interpretation of EMC/SI results with Grade-Spread, *Proc. of IEEE 2006 EMC Int. Symp., Portland,* USA, 14-18 August, 2006.

[4] J. Knockaert, J. Catrysse, R. Belmans, Comparison and validation of EMC-measurements by FSV and IELF, *Proc. on EMC Europe 2006, Barcelona*, 4 – 8 September, pp. 324-329, 2006.

[5] J. Knockaert, J. Catrysse, R. Belmans, Comparison of alternative conducted emission measurement methods by using FSV and IELF algorithms, *Proc. on EMC Europe 2006, Barcelona*, 4 – 8 September, pp. 718-722, 2006.