A Statistical Assessment of the Performance of FSV

G. Zhang[1], H. Sasse[2], L. Wang[1], and A. Duffy[2]

[1] School of Electrical Engineering and Automation
Harbin Institute of Technology, Harbin, 150001, China
zhang_hit@hit.edu.cn, wlx@hit.edu.cn

[2] School of Engineering
De Montfort University, Leicester, LE1 9BH, UK
hgs@dmu.ac.uk, apd@dmu.ac.uk

Abstract **─** This paper assesses the performance of the feature selective validation (FSV) method by applying probability density functions to the FSV point-by-point analysis. As an augmentation to confidence histograms, probability density functions offer two advantages: they (1) provide the users of FSV with more subtle information about the quality of the data comparison and (2) make a statistical analysis of the FSV results available. The application of probability density functions in the verification of FSV is presented in this paper, which provides a quantitative measure to support the qualitative conclusions drawn in early publications on the FSV method used as a foundation for IEEE Std. 1597.1.

Index Terms **─** Computational electromagnetics, EMC, feature selective validation, FSV, and statistical validity.

I. INTRODUCTION

With the publication of IEEE standard 1597.1 “standard for validation of computational electromagnetics computer modeling and simulation” [1], the feature selective validation (FSV) method has become a *de jure* standard for validation of electromagnetic simulation, particularly focusing on EMC. However, as the technique becomes more widely used [3-6], the need for certain enhancements becomes more apparent. In particular, the original formulation used six ‘natural language’ categories into which the FSV data was binned in order to help the interpretation between purely numerical results and the qualitative approach used by many practitioners. Unfortunately, this histogram approach only provides a coarse level of meta-representation and this lacks sufficient discrimination for more subtle usage, for example when comparing numerical modeling output as part of an optimization exercise. Continuous probability density functions (PDFs) offer the potential for greater precision in analysis over confidence histograms and this was demonstrated in a recent paper [2] but with only little detail. This paper addresses the issues that led to the development of the PDF approach, its implementation and interpretation and provides a more detailed investigation into the use of PDFs in the analysis of FSV comparisons, providing, for the first time a measure to verify the performance of FSV against one of the key design objectives, namely, to perform comparisons in the manner of a group of experts [7].

II. THE FSV METHOD

A. The FSV method

The FSV method was developed to validate electromagnetic models by quantifying the agreement between the reference and the numerical results. The details of this method can be found in [7] and [8]. In the FSV method, datasets under comparison are decomposed into DC, low- and high-frequency components first by use of the Fourier transform. Then three figures of merit are obtained to demonstrate data agreement from

Submitted On: June 26, 2013
Accepted On: Jan. 19, 2014

different perspectives based on these components. The amplitude difference measure (ADM) shows the 'trend' difference, while the feature difference measure (FDM) quantifies the differences of fine details. Then the ADM and FDM are combined to give the global difference measure (GDM). These figures of merit can be further represented in three different ways: point-by-point results (ADMi, FDMi, and GDMi), single value results (ADMtot, FDMtot, and GDMtot), which are obtained by taking the mean of each point-by-point result, and confidence histograms (ADMc, FDMc, and GDMc), which are discussed below.

B. Natural language descriptors and confidence histograms

A very practical characteristic of FSV is that it provides quantitative and qualitative results depicted by six natural language descriptors (excellent, very good, good, fair, poor, and very poor). Table I outlines the relationship between quantitative results and these descriptors. The aforementioned confidence histogram is presented by counting the proportions of the point-by-point results that fall into the six categories (N.B., the process of linearization is discussed in section III).

Table I: FSV interpretation scale [1].

FSV value (quantitative)	FSV interpretation (qualitative)
Less than 0.1	Excellent
Between 0.1 and 0.2	Very Good
Between 0.2 and 0.4	Good
Between 0.4 and 0.8	Fair
Between 0.8 and 1.6	Poor
Greater than 1.6	Very Poor

However, the coarse categorization can mask some subtleties in the distribution of the FSV results. Figure 1 shows some typical results to be compared (for the sake of space, the confidence histogram for the GDM is presented in Fig. 3). Figure 2 shows histograms based on the distribution of the GDMi values of datasets given in Fig. 1 [8]. The first histogram, based on 6 bins, used in FSV as the standard GDMc, suggests only 1 local maximum (mode) at a GDM value of 0.45. While for the second histogram, with 30 bins, more local maxima (around GDM values of 0.3, 0.4, and 0.5) are revealed. So it is necessary to find an alternative indicator to show this information. Further, the histograms show their limitations when confronted with multiple comparisons. Specifically, we lack flexible tools in the cross-comparison of multiple histograms.

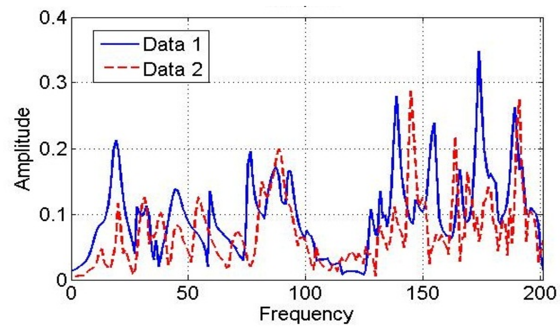


Fig. 1. Data sets for comparison [8].

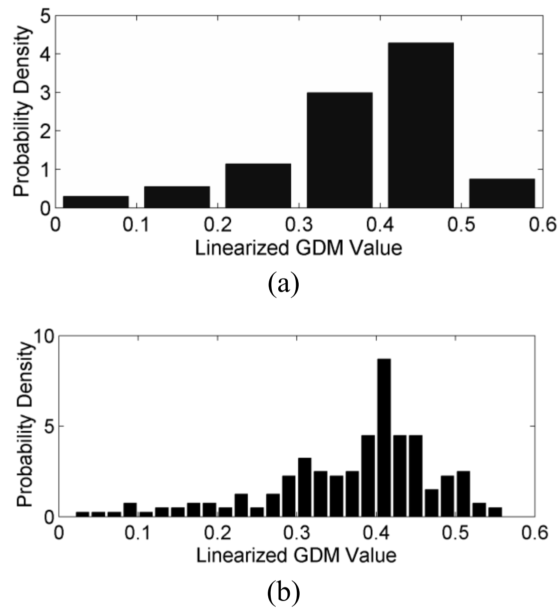


Fig. 2. Histogram of linearized GDMi value based on (a) 6 bins and (b) 30 bins.

III. EXTRACTING PROBABILITY DENSITY FUNCTIONS FROM FSV DATA

To solve the problems of histograms discussed in section II, a PDF, derived from the point-by-point data, is introduced to show the distribution of values in a more general way. As a result, the FSV distribution functions open up opportunities to apply statistical methods to the FSV results, giving rise to potentially revealing meta-analyses.

A. Method

The PDF is estimated based on a normal kernel function [9]. It is known that a probability density function $f(x)$ can be written as,

$$f(x) \equiv \frac{d}{dx}F(x) \equiv \lim_{h\to 0}\frac{F(x+h)-F(x-h)}{2h} \quad (1)$$

where $F(x)$ is the cumulative distribution function of the random variable x, and h is the "bandwidth". For a random sample of size n from the density f, X: $\{x_1, x_2, \ldots ,x_n\}$, its empirical cumulative distribution function (ECDF) has the form,

$$\hat{F}(x) \equiv \frac{N\{X \le x\}}{n} \quad (2)$$

where $N\{X \le x\}$ represents the number of elements less than or equal to x in X. Then the form in equation (1) becomes

$$\hat{f}(x) = \frac{N\{x-h < X \le x+h\}}{2nh}, \quad (3)$$

which can also be rewritten as,

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \quad (4)$$

where

$$K(u) = \begin{cases} \frac{1}{2}, & \text{-1<}u\text{<1}, \\ 0, & \text{otherwise.} \end{cases}$$

The form in equation (4) is that of a Kernel density estimator with uniform Kernel function, K. The choice of Kernel bandwidth h controls the smoothness of the probability density curve, the detail of which can be found in [9]. To obtain smoother PDFs, a Gaussian Kernel function is adopted in this paper,

$$K_{\text{Gaussian}}(u) = \begin{cases} (1/\sqrt{2\pi})\exp(-u^2/2), & \text{-1<}u\text{<1}, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Due to the non-linear relationship between the quantitative results and the qualitative description in Table I, we need to pre-process the point-by-point results according to Table II, thereby reflecting the qualitative information linearly in PDFs.

B. Statistical analysis

By introducing PDFs, it becomes possible to analyze FSV results using statistics, which is familiar to most engineers. Generally, statistical moments can be applied to a single PDF result. The second moment, variance, provides information on the dispersion of a set of data. The third moment, Skewness, gives a measure of the symmetry of the shape of a distribution. The fourth moment, Kurtosis, is a measure of the flatness, or peakedness, of a distribution,

$$\text{Variance}=\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(GDM(i)-\mu)^2 \quad (6)$$

$$\text{Skewness}=\frac{1}{N}\sum_{i=1}^{N}\frac{(GDM(i)-\mu)^3}{\sigma^3}, \quad (7)$$

$$\text{Kurtosis}=\frac{1}{N}\sum_{i=1}^{N}\frac{(GDM(i)-\mu)^4}{\sigma^4}, \quad (8)$$

where GDM(i) is the point-wise value of the FSV result (note that this also applies to the ADM(i) and FSM(i) results but only GDM(i) is shown because of space limitations). N is the number of points in the GDM, while μ and σ are the mean and standard deviation of GDM, respectively.

Table II: Piecewise linear conversion.

FSV value (point-by-point)	Linearized Value	FSV interpretation
X≤0.1	X	Excellent
0.1＜X≤0.2	X	Very Good
0.2＜X≤0.4	0.2+(X-0.2)/2	Good
0.4＜X≤0.8	0.3+(X-0.4)/4	Fair
0.8＜X≤1.6	0.4+(X-0.8)/8	Poor
1.6＜X≤3.2	0.5+(X-1.6)/16	Very Poor
X＞3.2	0.6	

For multiple comparisons, which is a strong motivation for the introduction of PDFs, statistical tests can provide widely recognized analysis methods. The Kolmogorov-Smirnov test (KS-test) [10] is used here for the following reasons: the KS-test is a non-parametric test, so it has the advantage of making no assumption on the distribution of data (important to ensure the generality of FSV); additionally, the KS-test is a robust test whose result is not affected by scale changing like the aforementioned linearizing procedure.

The KS-test aims to determine if the distributions of two datasets differ significantly. The null hypothesis is that the two datasets are from the same distribution. The alternative hypothesis is that they are from different distributions. The null hypothesis would be rejected if the test statistic, D, is greater than the critical value decided by significance level. The statistic D is determined by the maximum vertical deviation between the two curves of the cumulative distribution functions (CDFs) of the datasets,

$$D = \max\left(\left|CDF_1(x) - CDF_2(x)\right|\right) \tag{9}$$

where $CDF_1(x)$ is the proportion of values less than or equal to x in the first data set and $CDF_2(x)$ is the proportion of values less than or equal to x in the second data set.

The critical value of statistic D [11] for different significance level can be decided by,

$$D_{Critical} = k \cdot \sqrt{(N_1 + N_2)/(N_1 \cdot N_2)} \tag{10}$$

where N_1 and N_2 is the length of datasets being compared. The value of k can be obtained from tables [11]. For 95 % confidence, k is 1.36, for 90 % confidence, k is 1.22.

IV. FSV PERFORMANCE VERIFICATION

A. Method

When FSV was introduced in [7] and [8], its validation was performed by comparing the confidence histograms of a survey of experts and FSV predictions. Figure 3 shows confidence histograms comparison of data sets shown in Fig. 1. By use of PDFs, the comparison can be shown in a more analytical way. Further, the discrepancy between them can be quantitatively represented by the result of the KS-test.

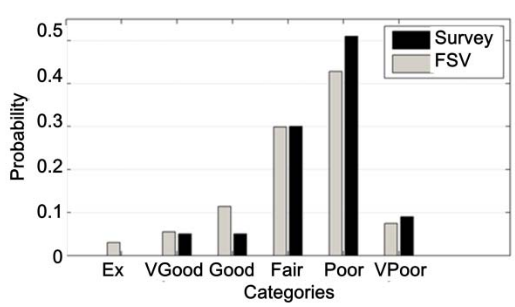


Fig. 3. Comparison of confidence histograms.

Due to the piecewise linear conversion of FSV results in Table II, the PDF results of visual assessment are calculated by transforming the qualitative results of FSV survey to typical quantitative values according to Table III. The survey results come from [8] with 50 experts surveyed.

Figures 4 and 5 outline the comparison of PDFs and CDFs between FSV prediction and visual assessment of data sets shown in Fig. 1, respectively. The single D value, 0.15, in Fig. 5 indicates the 'accuracy' of FSV when comparing the data sets in Fig. 1. According to the algorithm in equation (10), the $D_{Critical}$ for 90 % confidence is 0.17 with $N_1 = N_2 = 100$. In this case, the null hypothesis is accepted.

Table III: Typical values of FSV categories (mid-points in the linearized categories).

FSV Categories	Typical Values
Excellent	0.05
Very Good	0.15
Good	0.25
Fair	0.35
Poor	0.45
Very Poor	0.55

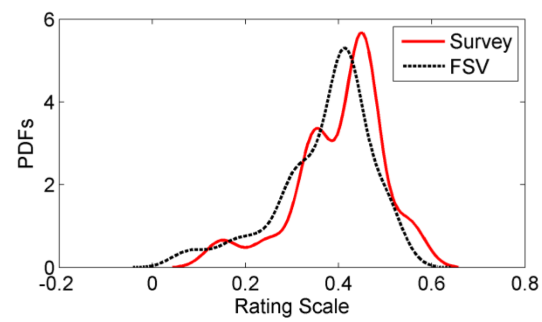


Fig. 4. Comparison of PDFs.

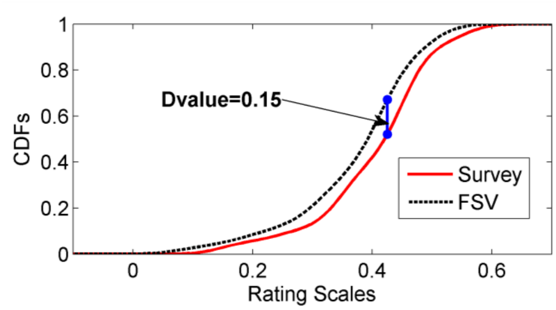


Fig. 5. Comparison of CDFs.

Further, the other 7 pairs of datasets from [8] are also analyzed. They are labeled as Data set 1 to 8 in turn considering data set in Fig. 1 is named Data set 5. The details follow.

B. Comparisons of xDM with survey results

It is known that FSV provides three types of measurement, ADM, FDM, and GDM. All the measurements can be given in the form of PDFs. Consequently, the discrepancy between these measurements and survey results can be shown by use of KS-test. By applying the typical values of FSV categories in Table III, the distributions of GDM of eight datasets are compared with their survey results in [8]. Table IV outlines the comparison result. It can be seen that most D values are smaller than the $D_{Critical}$ value, 0.17, except Data set 8, which means that it can be proposed that FSV results can mirror the assessment of experts in the majority of cases to a given level of accuracy in this case 90 % (or an inaccuracy of 1-in-10, which is about what the survey showed) .

Table IV: D values for different datasets.

Data set	D Value (GDM vs. Survey)
1	0.10
2	0.15
3	0.14
4	0.05
5	0.15
6	0.06
7	0.03
8	0.19

To evaluate the influence of typical value of FSV categories, the D values under different typical values for data sets are shown by boxplots in Fig. 7. As shown in Fig. 6, the typical values of FSV categories are linearly changed based on Table III, i.e., implementing a tolerance on the highly quantized visual results. For instance, the typical value for “Excellent” changes from 0 to 0.1 and, accordingly, the “Fair” value will change from 0.4 to 0.5. Consequently, the estimated PDFs will shift in the tolerance of FSV categories.

It can be seen from Fig. 7 that all the data sets have D values smaller than 0.17 with the change of typical values, including Data set 8. It is suggested that a new subtle survey is necessary to further verify the validity of FSV. Table V shows the comparison of D values when ADM and FDM are separately compared with survey results (using the typical values in Table III). D values of ADM and FDM for most of data sets are close to each. So it is reasonable to take the equal weighting of them when calculate GDM, as shown in equation,

$$GDM = \sqrt{ADM^2 + FDM^2}\ . \qquad (11)$$

It is also observed from Table V that neither ADM nor FDM can represent the discrepancy between data set as experts do not look at just trend or feature differences. But it works well when they are combined using equation (11), as shown in Table IV.

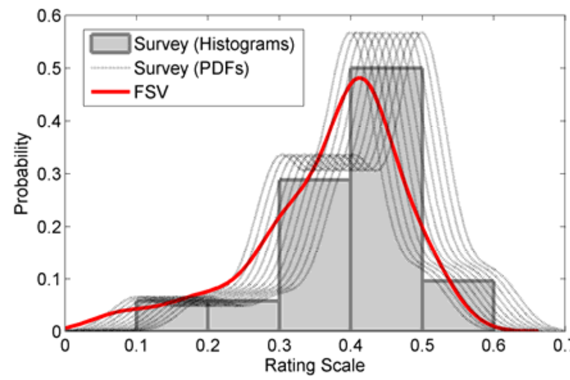


Fig. 6. Comparison of histograms and estimated PDFs given by different typical values of FSV categories.

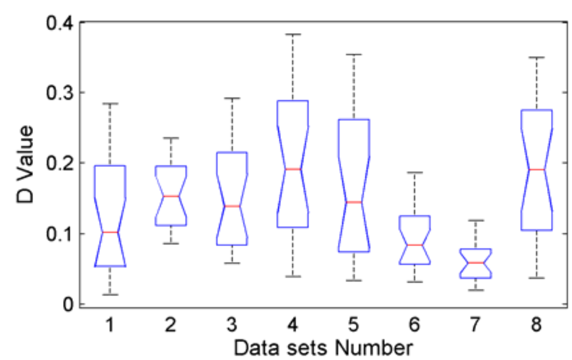
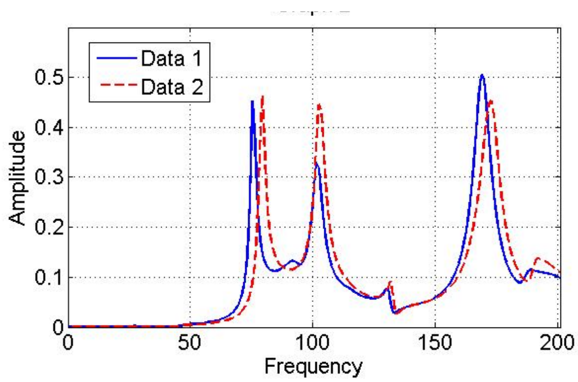


Fig. 7. Boxplots of data sets for different typical values of FSV categories.

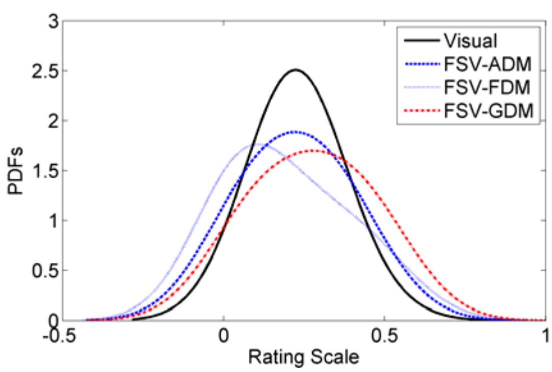
Table V: D values for different comparisons.

Data set	D Value (ADM vs. Survey)	D Value (FDM vs. Survey)
1	0.14	0.12
2	0.07	0.16
3	0.35	0.33
4	0.11	0.77
5	0.31	0.45
6	0.26	0.24
7	0.13	0.10
8	0.34	0.48

Figure 8 (b) shows the results of Data set 2, the ADM is closer to survey results than FDM or GDM. It can be seen from Fig. 8 (a) that the difference between data set is mainly caused by a partial shift in axis X. Experts may visually correct this type of data and just focus on amplitude difference. So we can infer that experts pay little attention to the feature difference when they assess Data sets 2. In this case, how to decide the weight of FDM would be an interesting investigation in the future.



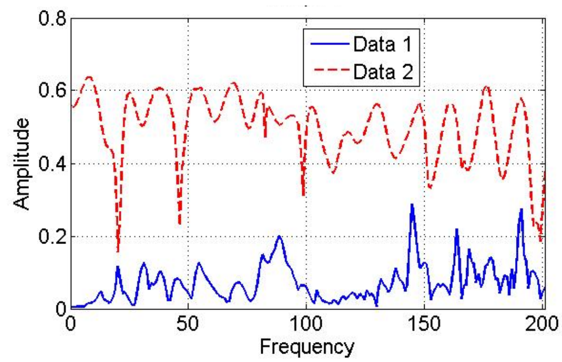
(a)



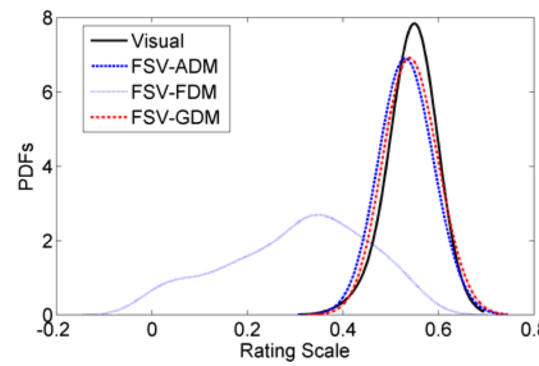
(b)

Fig. 8. (a) Data set 2 and (b) comparisons of its PDFs.

Figure 9 gives Data set 8 and the comparison of PDF results. It is indicated that the discrepancy between data sets is mainly caused by offset difference. So the ADM is much greater than the FDM and will dominate the GDM through equation (11). The equal weighting of ADM and FDM is also reasonable in this situation.



(a)



(b)

Fig. 9. (a) Data set 4 and (b) comparisons of its PDFs.

C. Comparison of statistical analysis

Table VI gives the comparison of the statistical analysis. The slight difference between mean values demonstrates that the FSV method can mirror the process of expert assessment. The mean value differences also indicate that, overall, experts tend to give more pessimistic assessment than FSV method for these data sets, which may provide a direction to the improvement of FSV. Data set 2 is an exception, as discussed in sub-section B, because experts intuitively correct the distortion caused by partial shift and give an optimistic assessment.

The comparison of variance shows that experts' assessment has tighter dispersion than FSV prediction. It means that FSV is more pessimistic in this. Again, comparisons of Skewness and Kurtosis show the general agreement between experts' assessment and FSV predication. It is noted that the variance of Data set 1's survey result is 0, so its Skewness and Kurtosis cannot be obtained.

To sum up, the PDFs and corresponding statistical analysis provide solid evidence to the validity of FSV method. And the introduction of PDFs makes it possible for the FSV to be a pre-statistical analysis method.

Table VI: Comparison of statistical analysis between FSV and experts assessment.

	Methods	1	2	3	4	5	6	7	8
Mean (GDMtot)	Survey	0.05	0.23	0.25	0.54	0.40	0.29	0.17	0.47
	FSV	0.03	0.27	0.20	0.54	0.37	0.28	0.16	0.41
Variance	Survey	0.0000	0.0048	0.0129	0.0006	0.0090	0.0138	0.0072	0.0037
	FSV	0.0009	0.0256	0.0069	0.0009	0.0106	0.0164	0.0140	0.0062
Skewness	Survey	-	1.0195	1.1188	-3.7500	-0.9285	0.1033	1.8939	-0.1585
	FSV	2.8684	0.0082	0.3778	0.4443	-1.0226	0.0385	0.9600	-1.0131
Kurtosis	Survey	-	4.9669	3.5826	15.0625	3.8481	2.5132	7.3729	2.4719
	FSV	14.6380	1.7572	3.0547	2.2734	3.9211	2.1639	3.0704	4.3879

V. CONCLUSION

This paper has discussed the introduction of PDFs as an enhancement to the presentation of FSV results. The performance of this indicator has been demonstrated in the verification of the performance of the FSV method itself. As the result of the introduction of PDFs, statistical analysis is employed to give quantitative results. By use of these analysis, the performance of FSV is verified when it is compared with experts' assessment. Further exploration with more data sets and more subtle rating scales would increase the statistical power of this observation, which needs more people to test this. Furthermore, the FSV method can be classified as a pre-statistical analysis method (or a data pre-conditioning method), which can provide greater flexibility for a wider range of computational electromagnetics practitioners.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundational of China under Grant No. 51077016.

REFERENCES

[1] *IEEE Standard for Validation of Computational Electromagnetics Computer Modeling and Simulations*, IEEE STD 1597.1-2008, pp. 1-41, 2008.

[2] G. Zhang, A. Duffy, H. Sasse, and W. Lixin, "The use of probability density functions to improve the interpretation of FSV results," *IEEE International Symposium on EMC*, U.S., pp. 685-689, 2012.

[3] A. Hastings, G. Palafox, and K. Tamhankar, "Assessment of feature selective validation (FSV) method for comparing modeled antenna patterns against measurements," *27th Annual Review of Progress in Applied Computational Electromagnetics (ACES)*, pp. 885-890, Virginia U.S., March 2011.

[4] A. Denton, A. Martin, and A. Duffy, "Quantifying EMC measurement accuracy using feature selective validation," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 23, no. 1, pp. 104-109, March 2008.

[5] A. Drozd, I. Kasperovich, C. Carroll, Jr., and A. Croneiser, "Antenna siting sensitivity study using the feature selective validation," *26th Annual Review of Progress in Applied Computational Electromagnetics (ACES)*, pp. 573-580, Tampere, Finland, April 2010.

[6] L. Hiltz and B. Archambeault, "Comparison of the modelled and measured antenna radiation pattern of a parabolic reflector using FSV," *25th Annual Review of Progress in Applied Computational Electromagnetics (ACES)*, pp. 173-177, California, U.S., March 2009.

[7] A. Duffy, A. Martin, A. Orlandi, G. Antonini, T. Benson, and M. Woolfson, "Feature selective validation (FSV) for validation of computational electromagnetics (CEM). Part I - The FSV method," *IEEE Transactions on Electromagnetic Compatibility*, vol. 48, pp. 449-459, Aug. 2006.

[8] A. Orlandi, A. Duffy, B. Archambeault, G. Antonini, D. Coleby, and S. Connor, "Feature selective validation (FSV) for validation of computational electromagnetics (CEM). Part II - Assessment of FSV performance," *IEEE Transactions on Electromagnetic Compatibility*, vol. 48, pp. 460-467, Aug. 2006.

[9] J. Simonoff, *Smoothing Methods in Statistics*, New York: Springer Verlag, 1996.
[10] F. Massey, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, pp. 68-78, 1951.
[11] N. Smirnov, "Tables for estimating the goodness of fit of empirical distributions," *Annals of Mathematical Statistics*, vol. 19, pp. 279-281, 1948.

Gang Zhang received the B.Sc. in Electrical Engineering from China University of Petroleum, Dongying, China, in 2007, and M.Sc. degrees in Electrical Engineering from Harbin Institute of Technology (HIT), Harbin, China, in 2009. He is currently working toward the Ph.D. degree in HIT and Department of Engineering, De Montfort University, U.K., and involving in the research of Feature Selective Validation (FSV) method. His research interests include analysis of electromagnetic compatibility (EMC), electromagnetic simulation, and the validation of CEM.

Hugh Sasse graduated from The University of York in 1985. He has since has been working in the Machine Vision Group and later the Applied Electromagnetics Group at De Montfort University, Leicester and was awarded his PhD in 2010. His main interests are computer modeling of novel antenna structures and other communication channel components, with an emphasis on optimization of such structures.

Lixin Wang received the B.Sc. degree in Electrical Engineering from Nankai University, Tianjin, China, in 1988, and the M.Sc. and D.Sc. degrees in Electrical Engineering from Harbin Institute of Technology (HIT), Harbin, China, in 1991 and 1999, respectively.

He is currently a Professor of Power Electronic and Electric Drives at the HIT. He conducts research with Faults Online Monitoring and Diagnosis Laboratory, HIT, on a wide variety of topics including electromagnetic compatibility at the electronic system level, aircraft electromechanical fault diagnosis expert system and prediction and health management (PHM) of Li-ion battery.

Alistair Duffy was born in Ripon, U.K., in 1966. He received the B.Eng. (Hons.) degree in Electrical and Electronic Engineering and the M.Eng. degree from the University College, Cardiff, U.K., in 1988 and 1989, respectively. He received the Ph.D. degree from Nottingham University, Nottingham, U.K., in 1993 for his work on experimental validation of numerical modeling and the MBA from the Open University in 2003.

He is currently a Reader in electromagnetics at De Montfort University, Leicester, U.K. He is the author of over 150 articles published in journals and presented at international symposia. His research interests include CEM validation, communications cabling, and technology management.

Dr. Duffy is a Fellow of the Institution of Engineering and Technology (IET) and a Member of the International Compumag Society and the Applied Computational Electromagnetics Society. He is a member of the IEEE EMC Society Board of Directors and is the Chair of the IEEE EMC Society Standards Development and Education Committee.