# Numerical Validation Methods

Ricardo Jauregui and Ferran Silva
*Universitat Politècnica de Catalunya (UPC), Barcelona, Spain*

## 1. Introduction

In the last years, numerical simulation has seen a great development thanks to costs reduction and speed increases of the computational systems. With these improvements, the mathematical algorithms are able to work properly with more realistic problems. Nowadays, the solution of a problem using numerical simulation is not just finding a result, but also to ensure the quality. However, can we say that the model results are correct regarding the behaviour of the system? In other words, how could we quantify the similarity between reality and simulations? To answer these questions, it is necessary to establish a validation criterion that allows an objective quantification of the difference between the results and the reality. Another way to say this is, how "true" our results are.

In the case of numerical methods, the main objective is to replicate as closely as possible the behaviour of the "real" world through numbers. Normally, the results of the numerical methods are expressed in terms of graphics, pictures, etc. These results represent the view of reality that the chosen method provides (Oñate, 1998). In order to affirm that the result of a numerical solution is fully consistent with the reality, it must be satisfied that:

a.  The mathematical model must incorporate all aspects of the real world.
b.  The numerical method has to solve exactly the equations of the mathematical modelling.

The problem starts with these two conditions that guarantee the "truth" of the results, since none of them are fully accomplished and it must be admitted that the numerical prediction never completely matches the "real" world behaviour. Then you can only be sure that the numerical solution is a good approximation of the reality. Now, new questions arise: How much does the result obtained by a numerical method resemble the reality? How can we objectively quantify this similarity? The answers to these questions are those that give rise to the validation methods.

## 2. Types of validation

All validation is done through a comparison of a pattern or a reference model with the model under study. There are many ways to make a validation, but in general they are usually classified according to the pattern used in the comparison (Godoy & Dardati, 2001), (Archambeault & Connor, 2008):

a.  **Validation using other numerical solutions.** This technique compares the results to be validated with the results obtained through other numerical methods previously

validated. In other words, one technique has been validated before it can be used as a reference to validate the second method.

Another way to use this technique is using more than one numerical method to solve the problem. If the physics of the problem is properly modelled in all the techniques used, the results should have a clear trend and similarity; therefore, knowing the advantages and disadvantages of each technique it will be possible to perform a validation of our technique.

b.  **Validation using analytical solutions**. This type of comparison can be used when the researcher knows the analytical theory behind the problem and makes a direct comparison of the simulation results with the analytical solution. One of the main problems of this technique is that it can only be used  in extremely simple cases, because trying to find the analytical solution of real problems is almost impossible (in fact, this is the reason of making numerical simulations). However, this technique is useful when you want to validate the code of the numerical method. Through the analytical solution it is possible to obtain the exact value of the problem, thereby reducing the external variables that can affect the results.

c.  **Validation using experimental results.** This technique is the most popular of all; this is due mainly to the fact that the measurement shows the consistency of the model with the reality. However, one cannot forget that whenever you perform a measurement you should introduce a measuring instrument and this directly or indirectly affect the system being measured (Archambeault & Connor, 2008). For this reason, it is essential to have the greatest similarity between the measurement and simulation configurations. For the real environment (measurement) one should take into account the possible limitations of the laboratory and the equipment required to perform the measurement. The most important issue is to narrow down to a minimum any device that cannot be fully simulated such as cables, connectors, etc. On the other hand, for the computing environment (simulation) one should try to model the entire possible setup or at least, include the most important characteristics. Otherwise, it runs the risk that the simulation results do not represent the reality faithfully, causing a validation error.

d.  **Validation using intermediate results.** This technique compares the intermediate results of the numerical model with experimental or theoretical known values, although these results are not the final objective of our comparison. The major drawback of this method is to find an intermediate result that is really close related with the final result under study. On the other hand, it is very easy to lose sight of the factors that could affect the intermediate variable making the comparison of the final result not valid. However, this technique is frequently used to monitor some parameters of the numerical simulations, but it is rarely used alone or as a main validation method.

A good example of this technique can be found in electromagnetic simulations. Imagine that it is required to compare the far-field simulations and measurements produced by a source inside an airplane. In this case, to make far-field measurements in a structure as big as an aircraft can be very expensive and complicated. However, it is possible to measure some near-field values at specific points near the aircraft and to compare them with simulations results calculated at the same points. Based on the direct relationship between the far-field and near–field, the similitude between simulation and existing measurements in the near field will be proportionately the same than in the far-field. As

can be seen in this example, the validation is not done with the final results (far-field), but an intermediate result (near-field) is used to set a criterion of similarity between simulation and measurement.

**Validation using convergence.** This type of validation is based on a comparison of the convergence of the numerical model with the pattern or the reference results. This comparison is done knowing that the solution found is not the best, but assuming that the model results converge.

Another situation where this type of validation can be used is when it is impossible to get a "pattern" to do the comparison. A good example of this type of validation can be found in chaos in structural mechanics area (Awrejcewicz & Krysko, 2008), where the validation is used in two ways; first, using a method to solve the system individually and observing if it converges or not. The second one is to use various methods to analyze the system at the same time and consider whether they all converge towards the same result and which of them do it faster.

In the electromagnetic area this technique of validation is often used when we want to know a general behaviour of the system in a very short time. Normally, a very simple model with very coarse meshes is simulated. Then, the most important resonances and the general behaviour are observed to analyze whether or not we are on track. The only drawback of this technique is that it is not recommended for use as a final method of validation, because convergence of the system cannot be guaranteed.

Regardless of the type of comparison that is performed (analytical, numerical, experimental, etc) at the end, the validation process is reduced to the comparison of the results and, in many cases, to the comparison of a pair of graphs. After that, once the type of validation to be performed in our model is chosen, the problem is how we can compare our results with the pattern in a quantitative mode. In many fields of research, simple visual inspection is used as the validation method when faced with the need to compare their results with the established models or patterns.

Although the visual comparison is used in different environments with apparent reliability, it has potential limitations. Among the most common problems are:

- The eye concentrates on peak positions and ignores the poor correlation of intensities (D. E. Coleby & A. P. Duffy, 2002).
- Due to the potentially subjective nature of the visual comparison, the results produced cannot be used with confidence (D. E. Coleby & A. P. Duffy, 2002),(A. Duffy, D. Coleby, A. Martin, M. Woolfson, & T. Benson, 2003).
- Comparing and quantifying the results objectively between different groups of experts can be difficult (Williams, M. S. Woolfson, T. M. Benson, & A. P. Duffy, 1997), (A. Duffy, D. Coleby, A. Martin, M. Woolfson, & T. Benson, 2003).
- The data may be too large (either a high volume of data or a very complex topography) to be compared visually with ease (Williams et al., 1997).

These limitations force the need to investigate reliable and objective computational techniques to compare the differences of the results and evaluate their quality.

## 3. Validation methods

Numerous studies show that a direct comparison point by point is not feasible when large amounts of data are compared (D. E. Coleby & A. P. Duffy, 2002), (Drozd, 2005), (Archambeault & Connor, 2008). Therefore, this method is not recommended to validate the

results and much less to assign an absolute value of accuracy. This approach makes sense only for simple models, but in the numerical simulations, the results are often very complex. Today there are several methods of validation. Among the most used are:

### 3.1 Correlation

This is a widely used method for its ease implementation and interpretation and it is intended for quantitative variables for which there is a linear relationship. The correlation between two variables is perfect when output value is closest to 1 or -1 and gets worse as it approaches to 0 (D. E. Coleby & A. P. Duffy, 2002). The sign indicates the direction of the association: a value equal +1 indicates a perfect positive linear relationship. When this case happens the relationship between two variables has exactly the same behaviour: when one of them increases, the other increases too. If instead of that, the value is -1, it is said that there is a perfect negative relationship and implies that both signals have a linear relationships, one will decrease as the other increases.

The most popular type of correlation is called the "Pearson correlation coefficient" and is usually used to measure the strength of the relationship between two variables when there is a linear relationship between them.

The Pearson correlation coefficient is defined by the following expression:

$$R_{pe} = \frac{\sum_{i=0}^{n} Y1_{(i)} \cdot Y2_{(i)} - \frac{\left(\sum_{i=0}^{n} Y1_{(i)} \cdot \sum_{i=0}^{n} Y2_{(i)}\right)}{n}}{\sqrt{\left(\sum_{i=0}^{n} Y1_{(i)}^2 - \frac{\left(\sum_{i=0}^{n} Y1_{(i)}\right)^2}{n}\right) \cdot \left(\sum_{i=0}^{n} Y2_{(i)}^2 - \frac{\left(\sum_{i=0}^{n} Y2\right)^2}{n}\right)}} \tag{1}$$

Where $Y1_{(i)}$ is the dataset 1, $Y2_{(i)}$ the dataset 2 and $n$ is the total number of points in both data sets.

The major limitation of this correlation technique is that it only can be used when the relationship between the variables is linear. This means that when variables are closely related, but not linearly, the validation results can not reflect the expert opinion. There are very few cases that have a linear relationship between variables so this method is only used for extremely basic cases.

An additional problem with this method, even when the data sets to compare have a linear relationship, is the interpretation a determined coefficient value. Or how does one know if a value is high or low? The answers to these questions depend largely on the nature of the investigation and the sample size used. For example, a correlation of 0.01 may be significant in a sufficiently large sample and a 0.9 may not be in a small sample. The law of large numbers is fulfilled, being that the weak trends are very unlikely from a null hypothesis and large amounts of data, while strong trends may be relatively likely in the small data size.

### 3.2 Reliability factor

This method is known as the R-Factor (The Reliability Factor) and it is one of the main criteria accepted in the validation area. The R factor could be considered a type of correlation; it is an objective method that provides with a single number, the similarity between two data. This method was created mainly to compare the intensities between the experimental and theoretical results in structural determinations of X-rays. A variation of different R-factors have been proposed: The first of these was introduced by Zanazzi and

Jona (Zanazzi & Jona, 1977), followed by Van Hove (Van Hove, 1977) and Pendry (Pendry, 1980).

### 3.2.1 Zanazzi and Jona R-factor

The R-factor of Zanazzi and Jona is also known as "$R_{ZJ}$-factor" (Zanazzi & Jona, 1977), and was planned to study the similarity of X-Ray diffraction and surface crystallography. Further studies modified their original equations and applied them in the numerical simulations area (Williams et al., 1997). This method was designed to accentuate the maximum slopes rather than the heights. This is accomplished by comparing the gradients of the two signals you want to compare. It basically made the differences between the signals for the first and second derivatives, accentuating the features present; making the R-Factor sensitive to positional changes in the data. The equation used to calculate this R-Factor is as follows:

$$R_{ZJ} = \frac{\sum_{i=0}^{n} W(i).F(i)}{\sum_{i=0}^{n} Y1(i)} \tag{2}$$

$$W(f) = \frac{|Y1''_{(i)} - C.Y2''_{(i)}|}{|Y1'_{(i)}| + |\max(Y1'_{(i)})|} \tag{3}$$

$$F(f) = |Y1'_{(i)} - C.Y2'_{(i)}| \tag{4}$$

$$C = \frac{\sum_{i=0}^{n} Y1_{(i)}}{\sum_{i=0}^{n} Y2_{(i)}} \tag{5}$$

Where $Y1_{(i)}'$ and $Y1_{(i)}''$ is the first and the second derivative of dataset 1. $Y2_{(i)}'$ and $Y2_{(i)}''$ is the first and the second derivative of dataset 2. $Y1$ is normally used with the experimental value $Y2$ obtained and the theoretical value or reference pattern. $C$ is used to adjust the intensity in both dataset.

This technique is useful when comparing sharp signals with peaks where the importance lies in their peaks or valleys, but does not offer the same reliability when you want to compare noisy signals where the peaks are not as marked and the variations between valleys is very fast.

### 3.2.2 Pendry R-factor

The Pendry R-Factor (Pendry, 1980) is used generally to measure the degree of correlation between two signals that have many variations in their maximum positions. This method uses derivatives in place of their intensities; this is attributed to all peaks of the same weight, regardless of the height of each one. The idea of this technique is that any maximum contains structural information due a constructive interference. Thus the maximum occurring at high energies are generally lower than those obtained at low. The equation used to calculate this R-factor is as follows:

$$R_p = \frac{\sum_{i=0}^{n}(Y_{SET1} - Y_{SET2})^2}{\sum_{i=0}^{n}(Y_{SET1}^2 + Y_{SET2}^2)} \tag{6}$$

$$Y_{SET} = \frac{L_{SET}^2}{L_{SET}} \tag{7}$$

$$L = \frac{SET'}{SET} \tag{8}$$

Where $L$ is the intensity and $L'$ is the differentiated intensity in each dataset. $Y1_{(i)}$ is the dataset 1, $Y2_{(i)}$ the dataset 2 and $n$ is the total number of points in both data sets.

Unlike "Zanazzi and Jona" factor where it is necessary to calculate the second derivative, the Pendry R-Factor only requires the first derivative, making it less susceptible to small or rapid changes. This feature makes it a useful tool for analyzing very noisy signals. However, the main problem with the Pendry method is that it requires finding an adjustable parameter that is not constant (Robertson et al., 1990), which seriously restricts the use of this technique.

### 3.2.3 Van Hove R-factor

The technique of Van Hove (Van Hove, 1977) is the most widespread of all the R-Factor ones. This technique uses five different equations (9)-(15) to compare the position and width of the signal peaks; the shape of the peaks and the troughs, their number and their heights (D. E. Coleby & A. P. Duffy, 2002). The different indicators of this method are calculated using the following equations:

$$R_1 = \frac{\sum_{i=0}^n |Y1_{(i)} - C.Y2_{(i)}|}{\sum_{i=0}^n |Y1_{(i)}|} \tag{9}$$

$$C = \frac{\sum_{i=0}^n Y1_{(i)}}{\sum_{i=0}^n Y2_{(i)}} \tag{10}$$

$$R_2 = \frac{\sum_{i=0}^n \left(Y1_{(i)} - C.Y2_{(i)}\right)^2}{\sum_{i=0}^n \left(Y1_{(i)}\right)^2} \tag{11}$$

Where $Y1_{(i)}$ is the dataset 1 and $Y2_{(i)}$ is the dataset 2. Both indicators (R1&R2) show the similarity in positions, heights and widths of peaks and troughs.

$$R_3 = \frac{N^o\ slopes^+(Y1)}{N^o\ slopes^-(Y1)} - \frac{N^o\ slopes^+(Y2)}{N^o\ slopes^-(Y2)} \tag{12}$$

Where $N^o slopes^+$ is the number of positive slopes and $N^o slopes^+$ is the number of negative slopes for each dataset. The R3 indicator compares the number of positive slopes with the negative slopes of the opposite graph.

$$R_4 = \frac{\sum_{i=0}^n |Y1'_{(i)} - C.Y2'_{(i)}|}{\sum_{i=0}^n |Y1'_{(i)}|} \tag{13}$$

$$R_5 = \frac{\sum_{i=0}^{n}\left(Y1'_{(i)} - C.Y2'_{(i)}\right)^2}{\sum_{i=0}^{n}\left(Y1'_{(i)}\right)^2} \qquad (14)$$

Where Y1′ is the first derivate of dataset 1 and Y2′ is the first derivate of dataset 2. In this case the R4 and R5 indicators are used to compare the gradient of the data sets.

Finally, the Van Hove Factor has a very useful indicator that combines all indicators to calculate the total difference between the two graphs, it is called: "$R_T$" (15). This indicator allows to quickly and accurately getting an overall idea of how good a result is with regard to the pattern.

$$R_T = \sqrt{R_1^2 + R_2^2 + R_3^2 + R_4^2 + R_5^2} \qquad (15)$$

### 3.3 Integrated with logarithmic frequency error (IELF)

The idea behind IELF (Integrated against Error Log Frequency) is based on the premise that when comparing data with a very high feature density, the overriding factor to be assessed is a function of the difference between the two traces (Simpson, Jones, MacDiarmid, A. Duffy, & D. Coleby, 2005). Basically, this method is the difference between two traces in logarithmic axis and in frequency domain. Then the result is integrated (summing) to get a single value. The IELF equation is given in (16):

$$IELF = \frac{\sum_{i=0}^{n-1}|error_i|.\left[\ln(f_{(i+1)}) - \frac{\ln(f_{(i-1)})}{2}\right]}{\ln(f_{(n)}) - ln(f_{(0)})} \qquad (16)$$

Where *f* are the frequency points being compared (from point 0 to point n, resulting in n+1 discrete frequencies), $|error|$ is the difference between the two data sets at the nth data point.

There is an improved version of IELF method known as IELF modified (IELF$_{MOD}$). This modification involves summing the elements halfway between the data points in order to improve the approximation of the difference in the measured data. This modification is given in equation (17):

$$IELF_{MOD} = \frac{\sum_{i=1}^{n-1}|error_{(i)}|.\left[\frac{\ln(f_{(i+1)} + f_i)}{2} - \frac{\ln(f_{(i-1)} + f_{i-1})}{2}\right]}{\ln(f_{(n)}) - ln(f_{(0)})} \qquad (17)$$

Today, the ELF$_{MOD}$ method is widely used to validate large volumes of data. This method is used in circumstances in which the data to be compared has a high visual density, ie, it is impossible to differentiate visually (Knockaert, Catrysse, & Belmans, 2006), (A. Duffy, D. Coleby, A. Martin, M. Woolfson, & T. Benson, 2003). The main disadvantage of this method is that it has very few tools to perform a good validation; only one indicator is very little to interpret all the aspects present in the validation process. Another important disadvantage is that it is defined only for the frequency domain in logarithmic axis and there are some weaknesses with abrupt changes in graphics.

**3.4 Feature Selective Validation Method (FSV)**

The method of Feature Selective Validation (FSV) was developed by Anthony Martin and Alistair Duffy in 1999 (A. Martin, 1999) and today is the method most widely used because of its versatility and simplicity. This method is widespread and is currently being developed as a standard of validation for Computational ElectroMagnetics (CEM) within the project by IEEE 1597.1/1597.2 (Standard IEEE, 2008). The Feature Selective Validation (FSV) method was developed with the specific aim to reflect the approach taken by engineers when assessing data presented visually during the validation of computational electromagnetic simulation. Today it is possible to find two FSV free online software, the first one was developed by the Aquila University in Italy (Orlandi, 2006). The second one has been developed by the Electromagnetic Compatibility Group of the Universitat Politècnica de Catalunya (GCEM, 2011). GCEM-UPC has built and developed new tools for the traditional FSV allowing the user to evaluate the graphics in a very quick and easy way (for more information visit: http://www.upc.edu/web/gcem/files/FSV.exe).

The FSV method is based on the decomposition of the results into two groups; the first one discusses the difference in amplitude (Amplitude Difference Measure, ADM) and the second one the difference between the characteristic of the signals (Feature Difference Measure, FDM). The combination of these two indicators (ADM and FDM) is a measurement of the overall difference (Global Difference Measure, GDM) (A. P. Duffy et al., 2006; Orlandi et al., 2006).

All indicators ADM, FDM and GDM have the ability to be configured to perform a point-to-point analysis. The advantage of relying on a point-to-point data is to know which areas of the data sets have the major differences. A subscript "i" is added to consider this point-by-point feature (ADMi, FDMi and GDMi).

Another way to qualitatively analyse the FSV indicators is represented by a probability density function. It is useful for a rapid and comprehensive analysis of the results. This indicator uses a histogram that can be divided into six categories: excellent, very good, good, fair, poor, very poor.

Finally, a technique that has proved useful in presenting and interpreting FSV data, particularly the confidence histograms, is a "Grade and Spread" (G/S) diagram (Archambeault, A. P. Duffy, & Orlandi, 2009; Archambeault & Yu, 2009).The Spread serves a similar purpose to variance or standard deviation in statistical methods and is a measurement of the spread of a distribution. The Grade is a measurement of the quality of the results and serves a similar purpose to skew measurements in statistics. It is important to remember that Grade and Spread must be used together, since if only one is used, the interpretation can be inaccurate.

The FSV method requires a serie of steps to obtain each indicator, a brief summary of some of them are following described:

a.  The first step is to interpolate the two sets of data to be compared to having the same number of samples for comparison.

b.  Once both datasets have the same number of samples, the Fourier Transformer is applied. Then, a high pass filter is applied in the dataset obtained the "Hi" data. The same procedure with a band pass filter is done to obtain the "Lo" data. An important aspect to consider is that these two new dataset (Hi and Lo) are separated by the breaking point, which is chosen with 40% of all data.

c.  Knowing all values of "Lo" data, the ADM indicator is calculated according to:

$$ADMi_{(i)} = \frac{(|Lo_1(i)| - |Lo_2(i)|)}{\frac{1}{n}\sum_{j=1}^{n}|(|Lo_1(j)| + |Lo_2(j)|)|} + ODM_{(i)} \cdot e^{ODM(i)} \tag{18}$$

$$ODM_{(i)} = \frac{x(i)}{\delta(i)} \tag{19}$$

$$x_{(i)} = (|DC_1(i)| + |DC_2(i)|) \tag{20}$$

$$\delta_{(i)} = \frac{1}{n}\sum_{j=1}^{n}(|DC_1(j)| + |DC_2(j)|) \tag{21}$$

d.  The next step calculated is the FDM. It is composed of three parts based on the derivatives calculated in the last step. The numerical values in the equations are parts of the heuristic and have been determined empirically.

$$FDM_{1(i)} = \frac{(|Lo_1'(i)| - |Lo_2'(i)|)}{\frac{2}{n}\sum_{j=1}^{n}|(|Lo_1'(j)| + |Lo_2'(j)|)|} \tag{22}$$

$$FDM_{2(i)} = \frac{(|Hi_1'(i)| - |Hi_2'(i)|)}{\frac{6}{n}\sum_{j=1}^{n}|(|Hi_1'(j)| + |Hi_2'(j)|)|} \tag{23}$$

$$FDM_{3(i)} = \frac{(|Hi_1''(i)| - |Hi_2''(i)|)}{\frac{7.2}{n}\sum_{j=1}^{n}|(|Hi_1''(j)| + |Hi_2''(j)|)|} \tag{24}$$

$$FDM_{(i)} = 2(|FDM_{1(i)} + FDM_{2(i)} + FDM_{3(i)}|) \tag{25}$$

e.  The GDMi indicator is calculated using the ADM and FDM indicators, as shown in equation (26).

$$GDMi_{(i)} = \sqrt{ADM_{(i)}^2 + FDM_{(i)}^2} \tag{26}$$

f.  Calculation of the mean value (XDMtot). After the ADM, FDM and GDM point-by-point values are calculated, it is possible to find the average value (27). These indicators are very useful to evaluate the quality of the results with one number.

$$XDMtot = \frac{\sum_{i=1}^{n}XDMi(i)}{n} \tag{27}$$

g.  Calculation of the confidence histogram. This is the term that is most often used in the descriptions of the quality of comparisons. The determination of the histogram (Fig 1.) is simply a case of counting the proportion of points that fall into one of the categories, according to the rule base in Table 1.

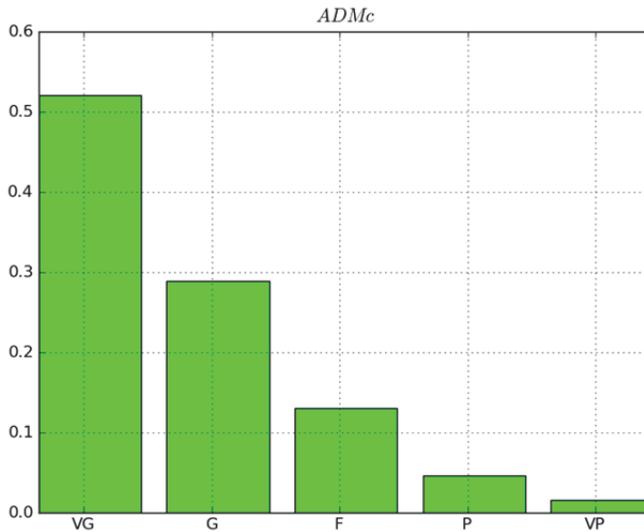| XDMc value (X=A,F,G) | |
|---|---|
| XDMc≤0.1 | Excellent |
| 0.1≤XDMc≤0.2 | Very Good |
| 0.2≤XDMc≤0.4 | Good |
| 0.4≤XDMc≤0.8 | Fair |
| 0.8≤XDMc≤1.6 | Poor |
| 1.6≤XDMc | Very Poor |

Table 1. XDMc interpretation scale.



Fig. 1. Confidence histogram example (ADMc).

It is quite possible that in some cases there is significant spread between the different categories (Excellent (Ex), Very Good (VG), Good (G), Fair (F), Poor (P), and Very Poor (VP)). Often this spread is caused when the datasets are very noisy. A quick and easy solution is to use a wider window width which produces a greater degree of smoothing.

The "Grade" and the "Spread" are calculated based on the ADMc and FDMc. The "Grade" is calculated by taking the number of category, starting from the best (Excellent) to the worst (very poor), which include a user defined amount (named "threshold", an 85 % is recommended) of total samples of data sets to being compared. The "Spread" is similar to a typical standard deviation since it also determines how many categories are required to include 85% of the data, but the starting category is the highest rated (instead of the excellent category, as in the Grade).

Despite all the benefits offered by the FSV, in the validation processes, some studies found problems when trying to use it to analyze transient signals (R Jauregui, Riu, & Silva, 2010; Ri Jauregui, Silva, Orlandi, Sasse, & A. Duffy, 2010). In particular, the main problem was in the ADM indicator.

The problem with the ADM indicator lies in the way it calculates the dataset "Lo" (R Jauregui, Rojas-Mora, & Silva, 2011). This dataset is obtained through the breakpoint (IBP),

which is calculated with the 40% of the signal, because it assumes that most of the signal energy content is within this range. The problem is that many times in a transient signal up to 90% of the energy can be contained in the first peaks. Therefore, it is highly probable that only the first peak of the transient is considered, and the other low level differences are not taken into account when comparing with FSV method. For this reason, before using this method to validate a new type of signal that has not been previously tested, it is recommend making a small review and analyzing the consistency of results.

### 3.5 Validation Transient Signals in Time Domain (VTTD)

One of the most interesting signals from the viewpoint of the numerical simulation is the impulsive noise, also known as transient phenomenon. These types of signals are used in the time domain for analyzing large frequency bands. On the other hand, these same signals offer a challenge in areas such as electromagnetism or resistance of structures in building.

The transient signals could be described as a signal that varies between two consecutive steady states during a short period of time compared to the time scale of interest. In other words, there must be a "momentary" change of the magnitude seen for a very short time in the sense that this short interval of time should be much less than one cycle of the signal. Thanks to the different derivatives performed, it only takes into account the changing intervals in the graphs without giving attention to the level differences affected.     This particular feature makes it difficult to analyze the effect using common methods of validation (R Jauregui, 2009). For this reason, a special method was developed to perform the Validation of the Transient in Time Domain (VTTD). It is proposed to use five indicators to assess the different parameters of the transient data sets:

a.  **Feature Difference Measure (FSV-FDM).** The calculation of this indicator is made using the equations of the FDM specified in the FSV method (see equation (25)). Unlike the amplitude indicator (ADM), the FDM does not present any problem when it is used to analyse the transient in time domain.  Thanks to the different derivatives performed in this indicator, it only takes into account the changing intervals in the graphs without attention to the level differences.

    This indicator is applied before taking into account any other indicators, because its value determines whether or not to continue the validation process. Some studies (R Jauregui, Pous, Fernández, & Silva, 2010; R Jauregui, Silva, & Riu, 2007) determined that the optimal limit for a correct interpretation is when the $FDM_{TOT}$ indicator is equal or less than 0.8. This value ensures that the two data sets (numerical simulation and measurement, for example) have a similarity that is within the acceptable margin.

b.  **Amplitude Pulse Level (APL).** This indicator measures the difference between the maximum amplitude of the signals. The maximum level of a transient signal is very important because it can produce several types of problems. Thus, the APL indicator aims to assess the maximum amplitude level difference between the two data sets.

    According to the equations (28) the APL calculates the difference of the maximum of each data set in absolute value to guarantee that the analysis is independent of the polarity.

$$APL = \frac{|\max(Y1) - \max(Y2)|}{|\max(Y1)|} \qquad |\max(Y1)| \geq |\max(Y2)| \qquad (28)$$

$$APL = \frac{|\max(Y2) - \max(Y1)|}{|\max(y2)|} \qquad |\max(Y1)| < |\max(Y2)| \qquad (29)$$

Where max $(Y1)$ is the maximum magnitude for the first dataset and max $(Y2)$ is the corresponding one for the second dataset. An APL range result is from 0 to 1. When the APL is equal to 0 the similarity is perfect, but as they increase, the result moves to 1.

c. **Maximum Rise Time (MRT)**. One important issue in a transient signal is the rise time. The lower it is, the more contents of the disturbance are on the high-frequency band, which is usually a problem in validation analysis. The calculation of this indicator is very similar to the one used in APL; the only difference is that it calculates the first derivative (30) and then applies the equations (31) or (32).

$$D_i^j = \left(\frac{y_i^j - y_{i-1}^j}{x_i^j - x_{i-1}^j}\right) \qquad \begin{matrix} j = \{Y1, Y2\} \\ i = \{1,2,3\dots n\} \end{matrix} \qquad (30)$$

$$MRT = \frac{|\max(D^{Y1}) - \max(D^{Y2})|}{|\max(D^{Y1})|} \qquad |\max(D^{Y1})| \geq |\max(D^{Y2})| \qquad (31)$$

$$MRT = \frac{|\max(D^{Y2}) - \max(D^{Y1})|}{|\max(D^{Y2})|} \qquad |\max(D^{Y1})| < |\max(D^{Y2})| \qquad (32)$$

Where $i$ is the number of the point (from 1 to n). $j$ is the set of the graph that we want to analyze ($Y1$ is the first one and $Y2$ the second one). $D_i$ is the derivative for each point (1 to n) for both dataset. $|max(D^{Y1})|$ is the absolute maximum value of the derivative of the first dataset and $|\max(D^{Y2})|$ is the derivative of the second one. Similar to the APL indicator, the equation (31) or (32) is applied in order to ensure that MTR varies from 0 to 1.

d. **Energy Contained in the Signals (ECS).** This indicator measures the energy contained in the transients. In many cases, a transient energy could be very significantly affecting the behaviour of a system; therefore, it is important to evaluate. Applying the equations (34) or (35), the difference of energy between both datasets can be determined for the same interval of time.

$$E^j = \int_{t1}^{tn} U^j(t)\, dt \qquad \begin{matrix} j = \{Y1, Y2\} \\ t_1 < t_n \end{matrix} \qquad (33)$$

$$ECS = \frac{|(E^{Y1}) - E^{Y2}|}{|E^{Y1}|} \qquad |E^{Y1}| \geq |E^{Y2}| \qquad (34)$$

$$ECS = \frac{|(E^{Y2}) - E^{Y1}|}{|E^{Y2}|} \qquad |E^{Y1}| < |E^{Y2}| \qquad (35)$$

Where "E" represent the energy and "U" is the magnitude recorded for each dataset, both datasets must be defined from t = 0 to t = $t_n$.

e. **The Total Error Average (TEA).** As seen in previous validation methods, it is very useful to have an indicator that reflects the overall quality of results. This indicator

allows a quick and simple way to have a general idea of the quality of results. The indicator TEA meets this objective quickly and easily.

The calculation of TEA is based on finding the squared error of the indicators FDM, APL, MRT and ECS as shown in equation (36). In this equation, a weighting factor for each indicator can be defined by "α", "β", "γ" to highlight the importance of a particular indicator in a particular situation.

$$TEA = \sqrt{\frac{\alpha.(APL^2) + \beta.(MRT^2) + \gamma.(ECS^2)}{\alpha + \beta + \gamma}} \tag{36}$$

The intervals of this indicator may vary depending on the transient and the type of problem being analyzed. It is the user who must define the scales and values that define it. For example, in the case of Electromagnetic Compatibility (EMC) area, a useful scale to analyze the transients is:

**Good:** from 0 to 0.3
**Regular:** from 0.3 to 0.5.
**Bad:** from 0.5 to 1.

This method allows rapid and objective quantification of the simulation results, but it is important to note that this validation method is valid only to study the transient in time domain.

## 4. Validation method application examples

In order to show the application of the different validation methods previously presented, two real cases are chosen to compare their results. Before making any comparison, it is necessary to normalize all validation methods to ensure that all of them are within the same scale (from 0 to 1). As it is usual, we use different categories to help us to identify the quality of the results: excellent (from 0 to 0.16), very good (from 0.17 to 0.34), good (from 0.35 to 0.5), fair (from 0.51 to 0.65), poor (from 0.66 to 0.80), very poor (from 0.81 to 1). Finally, a survey among some experts has been done to compare their opinion with the different validation methods under test.

The VTTD method was not used in these examples, as it is defined only for transient analysis in time domain. If you need more information about the use or implementation of this method, it is recommended to view documents (R Jauregui, 2009; R Jauregui, Riu, & Silva, 2010; Riu, R Jauregui, Silva, & Fernandez, 2007).

### 4.1 First case of study
The aim of this first example is to examine the efficiency of each validation method to analyze the similarity between two signals. These signals were obtained by the measurement (Fig. 2-blue) and the simulation (Fig. 2-red) and show the transfer function between two electromagnetically short monopoles inside a resonant cavity. As one can observe, the simulation and the measurement have a similar behaviour for the entire frequency range. However, some minor differences can be found in the negative resonances. In general, the results were classified by a panel of experts with 0.25, which is in the "Very good" category.
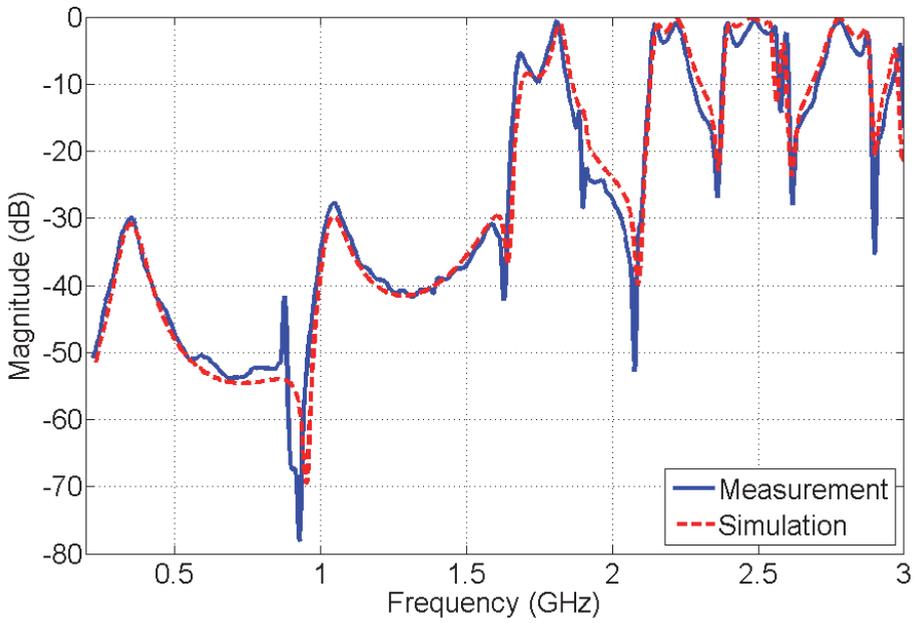
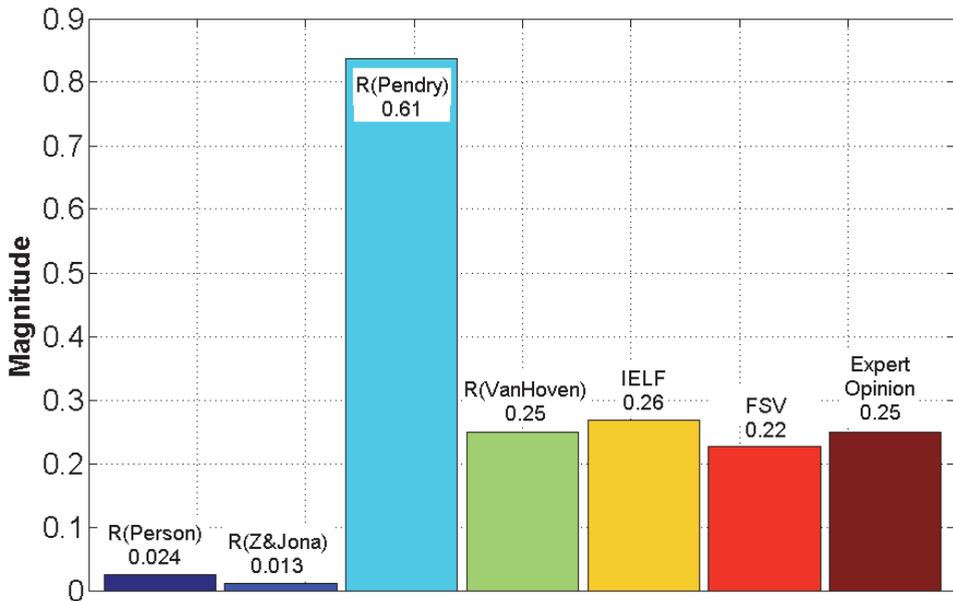Fig. 2. Comparison between measurement and simulation.



Fig. 3. Results from the different validation methods used and the experts opinion.

Applying the different validation methods studied before, we obtain the results shown in the Fig 3. It is possible to observe that each method gives different results that lead to different interpretations. In the cases of the Pearson correlation (0.024) and the $R_{ZJ}$-Factor (0.013) methods, the results show that there is a almost a perfect match between the measurement and the simulation, and this does not agree with the expert's opinion.

The main reason for Pearson and $R_{ZJ}$-Factor results is that the correlation method is unable to evaluate the differences caused by rapid changes in slopes. This method only analyzes the correspondence in amplitude over the signals but no other feature is considered. Because of that reason, those methods are not suitable to analyze signals with abrupt changes or, in particular, with noise.

If the focus is now on the third method ($R_{Pendry}$-Factor), the results show a poor similarity between the simulation and measurement. Again, this result does not match with the expert's opinion. The main limitation of this method is that it is very sensitive to sudden changes in the signals and the indicators are directly affected.

Finally, we have the methods of Van Hoven factor, FSV and IELF which results are very close to expert's opinion. This result is not surprisingly, since these three methods are particularly robust and have been tested against different types of behaviours. Therefore, as it has been explained in the preceding paragraphs, they are ideal for the numerical simulations output validation process.

### 4.2 Second case of study

The aim of this second example is to use the validation methods in a more realistic application for the field of numerical electromagnetic simulations. In this case, we compare a measurement (Fig. 4-Blue) with two different simulations (Fig. 4-red & black). Each
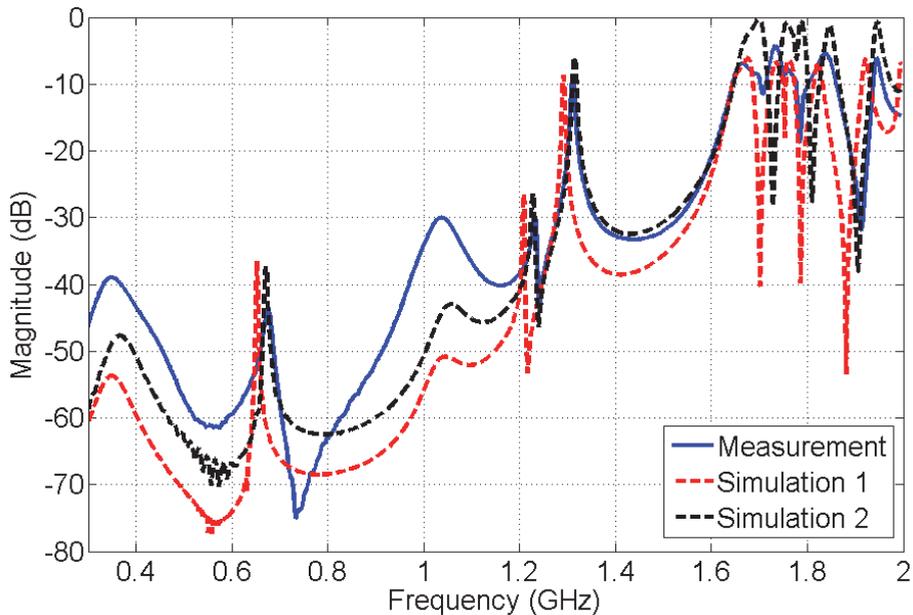


Fig. 4. Comparison between measurement and simulations.

simulation was performed using the same Finite Difference in Time Domain (FDTD) algorithm but with different settings like meshing and time step. The target is to decide, by using IELF and FSV methods, which is the simulation that better fits the measurement from the point of view of a panel of experts in this topic.

At a first sight by doing a quick comparison between the signals, anyone can deduce that both simulations can be improved, but more difficult is to decide which one is better. When we want to study the influence of a particular simulation parameter such as time, mesh, etc. it is very important to identify which simulation has a greater similarity with the measurement of the real setup which it is supposed to be the right one.

Some people can realize that first simulation (Fig. 4-red) has a greater similarity than the second (Fig. 4-red); since it seems to have a close behaviour with the measurement at high-frequencies. However, the simulation number one has important differences at the low-frequency band that should not be forgotten because for another set of people this could be an important feature where to focus the comparison. Furthermore, the second simulation seems to have the opposite behaviour to the first one: a closer similarity at lower frequencies, but a significant amplitude difference at higher frequencies. Therefore, in this case, it is not an easy task to take an overall decision without the help of an experts group or an appropriate validation method.

Fig 5 shows the results when the IELF and FSV are applied to compare the simulations 1 and 2 with the measurement. Observing all the indicators of the used methods, one can see how both methods are quite close to the experts. These results show that the worst simulation is the second one (black) or in other words, the first simulation has more similarity with the measurement considering whole plot.
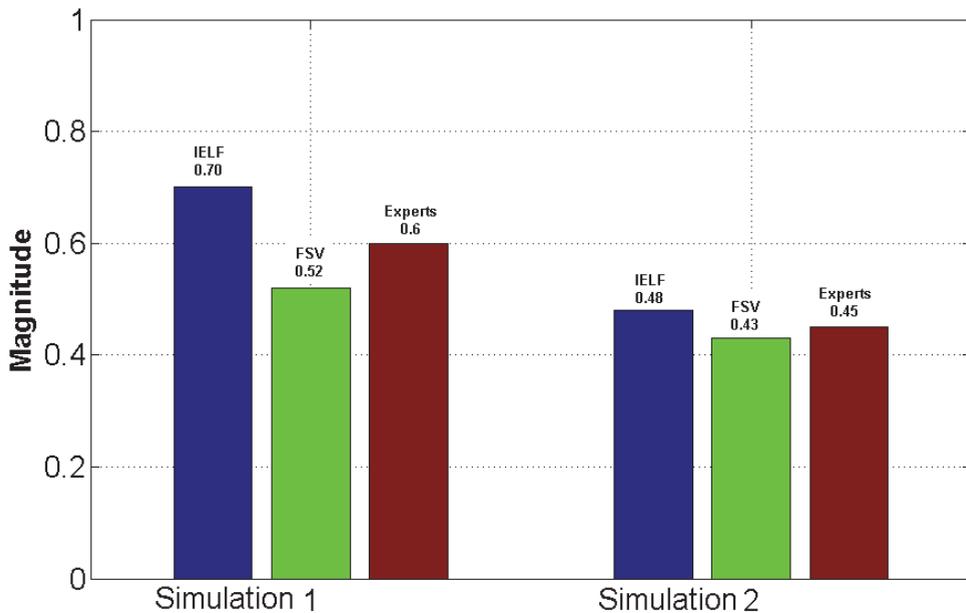


Fig. 5. Results of different validation methods used and the expert panel opinion.

Even though the two validation methods have a unified approach on which simulation is better, each one produces a different value for the final indicator. This is a very common problem when several types of validation methods are compared. It is therefore very important to always use one single method throughout one validation process.

Another point of vital importance, when a validation method is chosen, is to have tools that help us to analyze the different features present in the signals. This is one of the main limitations of the method IELF, because it concentrates all the comparison information in a single number. With this method, it is not possible to obtain more information about the validation process.

The FSV method has, as noted above, several analysis tools that can help to establish a more comprehensive comparison in the validation process. With the mean value of each indicator (Table 2), one can see that the largest difference between the simulations is in the shape indicator ($FDM_{TOT}$) and not in the amplitude ($ADM_{TOT}$) as one might think.

| FSV indicators | | | |
|---|---|---|---|
| Simulation 1 | | Simulation 2 | |
| $ADM_{TOT}$ | 0.28 | $ADM_{TOT}$ | 0.35 |
| $FDM_{TOT}$ | 0.66 | $FDM_{TOT}$ | 0.65 |
| $GDM_{TOT}$ | 0.80 | $GDM_{TOT}$ | 0.82 |

Table 2. FSV mean values indicators.

One way to analyze in greater detail what happens in each comparison is using the point-by-point indicators. It is important to recall that, in this case, the indicators $ADMi$ and $FDMi$ correspond to the point-by-point analysis between each simulation and the measurement. Fig. 6 shows that the indicator most affected in these comparison is the $FDMi$ indicator and of course for the first simulation. Now we can see, very clearly, that the problem is mainly in the shape and not in amplitude.
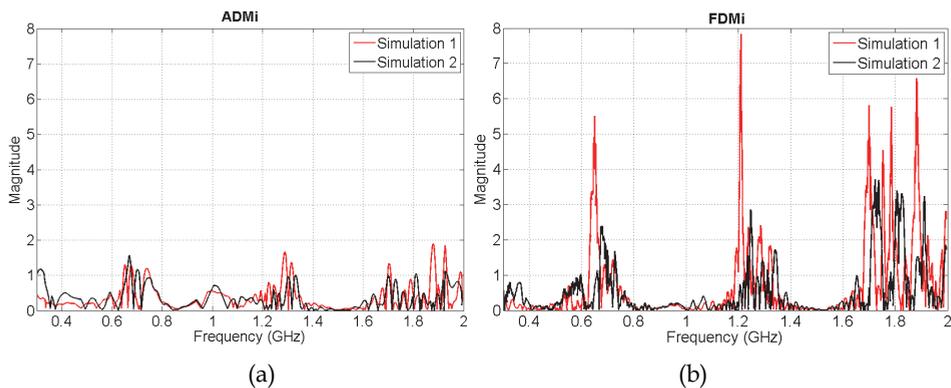


(a)                                                        (b)

Fig. 6. Point-by-point indicators results. (a) ADMi indicator results for simulation 1 & 2 with respect to the measurement. (b) FDMi indicator results for simulation 1 & 2 with respect to the measurement.

Another powerful use of these indicators is to identify where, and the exact value, the major differences over all the data set (frequencies in this case) is produced. For both analysis

made in this example, it is clear that the major difference is found at frequencies of 0.6, 1.2 and 1.8 GHz. In the worst case, we can see (Fig. 6b) that the FDMi indicator reaches values near to 8, showing an important difference between the first simulation and the measurement.

With these two short examples, it became clear that it is very important to choose a proper method of validation that objectively represents the opinion of experts. We have also seen how important it is to have the necessary tools to interpret these results.

It should be noted that all the techniques presented can be used not only to validate the numerical methods and simulation; its use can be extended to other areas that require a quantitative comparison of complex data. The only important thing when a validation method is chosen is that it must provide a similar result to the expert opinion, which implies an objective analysis of the data.

On the other hand, it is needed to take into account that a perfect method to validate any kind of result does not exist. Each one of the methods presents advantages and disadvantages depending on the type of data and the type of analysis desired. The most essential thing at the time to apply the validation is to consider the following items:

a.  The implementation of the validation technique should be as simple as possible; this will avoid confusion and data clouding.
b.  The validation method should reflect human opinions. Any technique which leads to conflict with the views of the user will fall rapidly into disuse.
c.  The validation method should provide the possibility to be applied in different environments and/or applications.
d.  The validation method should be commutative. The results of the comparison should always be the same regardless of which is used as a reference or pattern. In other words, the user satisfaction and credibility of the method can be affected if the quality of the technique varies depending on which data is used as a pattern for comparison.
e.  The validation method must analyse the difference between the two data sets and always yield the same result, regardless of the user and number of times the comparison is made.

## 5. Acknowledgment

## 6. References

Archambeault, B., & Connor, S. (2008). Proper model validation is important for all EMI/EMC applications. *Proc. IEEE International Symposium on Electromagnetic Compatibility EMC 2008* (pp. 1-8). doi: 10.1109/ISEMC.2008.4652152.

Archambeault, B., Duffy, A. P., & Orlandi, A. (2009). Using the Feature Selective Validation Technique to Compare Data Sets. *Electrical Engineering*, 248-253.

Archambeault, B., & Yu, Z. (2009). Application of the Feature Selective Validation Method to radio path loss measurements. *2009 IEEE International Symposium on Electromagnetic Compatibility*, 259-263. Ieee. doi: 10.1109/ISEMC.2009.5284666.

Awrejcewicz, J., & Krysko, V. A. (2008). *Chaos in structural mechanics* (1st ed.). Springer. doi: 10.1007/978-3-540-77676-5.

Coleby, D. E., & Duffy, A. P. (2002). Analysis of techniques to compare complex data sets. *The current issue and full text archive of this journal is available*, *21*, 540-553.

Drozd, A. L. (2005). Selected methods for validating computational electromagnetic modeling techniques. *Electromagnetic Compatibility, 2005. EMC 2005. 2005 International Symposium on* (Vol. 1, p. 301--306Vol.1). doi: 10.1109/ISEMC.2005.1513518.

Duffy, A. P., Martin, A. J. M., Orlandi, A., Antonini, G., Benson, T. M., & Woolfson, M. S. (2006). Feature selective validation (FSV) for validation of computational electromagnetics (CEM). part I-the FSV method. *Electromagnetic Compatibility, IEEE Transactions on*, *48*(3), 449-459. doi: 10.1109/TEMC.2006.879358.

Duffy, A., Coleby, D., Martin, A., Woolfson, M., & Benson, T. (2003). Progress in quantifying validation data. *Electromagnetic Compatibility, 2003 IEEE International Symposium on* (Vol. 1, p. 323--328vol.1).

GCEM. (2011). FSV Software. Retrieved from http://www.upc.edu/web/gcem/cat/default.htm.

Godoy, L., & Dardati, P. (2001). Validación de modelos en mecánica computacional. *Mecanica Computacional*, *20*, 663-670.

Jauregui, R. (2009). *Tesis Doctoral Comportamiento Electromagnético en el Dominio del Tiempo de Estructuras Complejas Mediante FDTD*. UPC. Universitat Politècnica de Catalunya.

Jauregui, R, Pous, M., Fernández, M., & Silva, F. (2010). Transient Perturbation Analysis in Digital Radio. *2010 IEEE EMC Symposium, USA, 2010* (pp. 2-7).

Jauregui, R, Riu, P., & Silva, F. (2010). Transient FDTD Simulation Validation. *2010 IEEE EMC Symposium, USA, 2010.*

Jauregui, R, Rojas-Mora, J., & Silva, F. (2011). Study of transient phenomena with Feature Selective Validation Method. *Progress in Electromagnetics Research Symposium (PIERS)* (pp. 2-6).

Jauregui, R, Silva, F., & Riu, P. (2007). FDTD Analysis of the distribution of a transient field inside a car. *EMC Europe 2007 Paris Francia*.

Jauregui, Ri, Silva, F., Orlandi, A., Sasse, H., & Duffy, A. (2010). Factors influencing the successful validation of transient phenomenon modelling. *Asia-Pacific Electromagnetic Compatibility Symposium and Technical Exhibition. Beijing, 2010.* (pp. 2-5).

Knockaert, J., Catrysse, J., & Belmans, R. (2006). Comparison and validation of EMC-measurements by FSV and IELF. *IEEE*.

Martin, A. (1999). *Feature Selective Validation*. De Montfort University.

Orlandi, A., Duffy, A. P., Archambeault, B., Antonini, G., Coleby, D. E., & Connor, S. (2006). Feature selective validation (FSV) for validation of computational electromagnetics (CEM). part II- assessment of FSV performance. *Electromagnetic Compatibility, IEEE Transactions on*, *48*(3), 460-467. doi: 10.1109/TEMC.2006.879360.

Orlandi, A. (2006). stand-alone FSV application. Retrieved from http://ing.univaq.it/uaqemc/public_html/.

Oñate, E. (1998). Limite de los métodos numéricos. *Centro internacional de metodos numericos de ingenieria (CIMNE), N191*.

Pendry, J. B. (1980). Reliability factors for LEED calculations. *Journal Physics C: Solid State Physics*.

Riu, P., Jauregui, R, Silva, F., & Fernandez, M. (2007). Transient Electromagnetic Field Computation in Automotive Environments using FDTD. *Proc. IEEE International Symposium on Electromagnetic Compatibility EMC 2007* (pp. 1-4). doi: 10.1109/ISEMC.2007.8.

Robertson, A. W., Somers, J. S., Ricken, D. E., Bradshaw, A. M., Kilcoyne, A. L. D., & Woodruff, D. P. (1990). "Photoelectron diffraction study of Cu(110)-(2 £ 1)-0. *Surface Science, 227*, 237-45.

Simpson, R. J., Jones, C. R., MacDiarmid, I., Duffy, A., & Coleby, D. (2005). The integrated error against log frequency (IELF) method for CEM validation. *Electromagnetic Compatibility, 2005. EMC 2005. 2005 International Symposium on* (Vol. 1, p. 296--300Vol.1). doi: 10.1109/ISEMC.2005.1513517.

Standard IEEE. (2008, June). IEEE 1597.1 Standard for Validation of Computational Electromagnetics Computer Modeling and Simulations.

Van Hove, M. A. (1977). Surface structure refinement of 2H-MoS2 via new reliability factors for surface crystallography. *Surface Science, 64*, 85.

Williams, A. J. M., Woolfson, M. S., Benson, T. M., & Duffy, A. P. (1997). Quantitative validation of numerical technique results against experimental data. *Antennas and Propagation, Tenth International Conference on (Conf. Publ. No. 436)* (Vol. 1, p. 532--535vol.1).

Zanazzi, E., & Jona, F. (1977). Reliability factor for surface structure determination. *Surface Science, 62*, 61.