A note from the Technical Feature Article Editor

The following paper was originally presented at the International Wire and Cable Symposium in November 2005 (www.iwcs.org). While the focus of the paper was on the cabling industry, it has been included here as a technical feature article because of its relevance to CEM. In particular, if we are comparing data visually, it does not matter whether the origin is simulation, measurements or a combination, the same cognitive processes are in play. We still want to know if one comparison is better than another one and by how much. Those cognitive processes combine tacit and explicit, academic and experiential knowledge; all of which color the decisions we make. This means that a group of experts will probably not all agree on the overall quality of any comparison. But that is not necessarily a bad thing: the more disagreement we have, the more we need to look at the comparison to understand why such a spread of agreement has resulted; the more agreement we have, the more confident we can be of the decision about whether our model has been improved, for example. In terms of trying to understand whether there has been an improvement, there is a clear need to quantify what we see, in a way that we can relate to. This means capturing opinion and looking to put a numerical value to a subjective judgment. "Quantifying experimental repeatability and simulation validation" presents some thoughts on knowledge and how the different types of knowledge contribute to a group's decision making, how to capture and quantify the groups view of the results and how to automatically compare the results using the Feature Selective Validation (FSV) method and the Integrated Error Against Log Frequency (IELF) method. The implications for CEM validation are clear and the work presented in the paper captures a number of the underlying ideas for the IEEE's Standard in development on validating computational electromagnetics (IEEE Project P1597).

Quantifying Experimental Repeatability and Simulation Validation

Alistair Duffy
De Montfort University
Leicester, UK
+44 (0)116 257 7056 apd@dmu.ac.uk

Kenneth Hodge
Brand-Rex Ltd
Glenrothes, Fife, UK
+44 (0)1592 778 555 khodge@brand-rex.com

Abstract

Data bandwidth requirements are continually increasing, stretching the limits of the physical channel. There is no respite in the need to continually reduce design cycles and improve product performance. The result of these two factors is that the volume of information generated during design and conformity assessment is increasing as the requirements and the complexity of the results being compared are increasing but the time available to obtain these results and process them is decreasing. There is clearly a need for tools that can be used to support the decision making of a range of professionals from design engineers to those making technical recommendations on purchasing decisions. That is, a tool that provides some quantification and objectivity for those aspects of the decision making that have, hitherto, relied on subjective judgment. This paper discusses some support tools that may find application in the cabling community.

Introduction

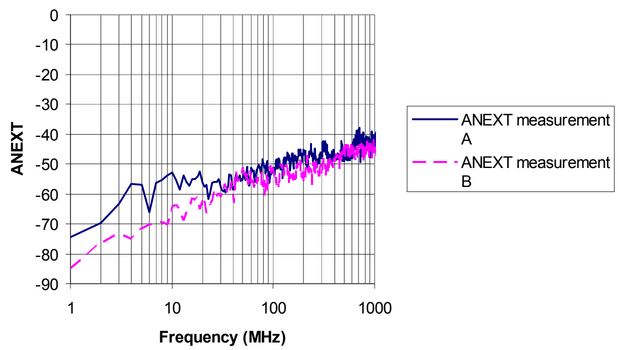
Until recently, cabling design was dominated by incremental changes directed by the increasing knowledge base of a handful of engineers and technicians in a number of individual companies. This picture is changing for a number of reasons:

1. Leaner companies have fewer design office staff to develop a 'group' explicit and tacit knowledge.
2. Consolidation in the industry has opened up the need for distributed design teams.
3. The push for greater bandwidth means that the designs themselves need to be able to operate much closer to much tighter limits.
4. Design cycle times are much reduced compared with a few years ago, with the resulting need to invest in, and rely on, virtual prototyping tools.
5. The higher frequencies in which cabling systems, particularly structured cabling, operate result in the graphical representation of parameters like return loss, cross-talk and attenuation being visually more complex.
6. The increase in general complexity of systems results in less clear-cut technical requirements for purchasing decisions.
7. While all measurements will be subject to errors and measurement artifacts, the fact that distributed teams may be measuring the same product in several sites and comparing their results with third party measurements, the question remains: just how similar are these results?

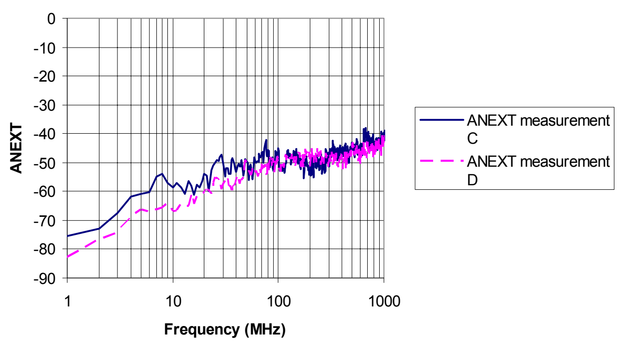
These factors suggest a clear need for technologies that can help to compare visually complex data, allowing a quantification of what is ostensibly a subjective judgment, providing some intelligence into the physics behind the perceived artifacts and allowing a rational basis for discussion and debate. This paper addresses these issues by providing a brief overview of knowledge, how teams make use of different forms of knowledge and then reports on current work to support this interplay of knowledge in terms of visual rating of graphical data and then computer assessment of the data.

These tools are likely to have particular use for engineers who want to go beyond simple pass/fail metrics. For example, getting the right (or at least best) decision can be a little like playing hide-and-seek: it is always more helpful to be told "warmer" or "colder" rather than simply "no". A further example is in the selection of a product for installation or the selection of a modeling tool for analysis: being able to find a way of quantifying the quality of the solution compared to other vendors or the quality in terms of consistency is potentially profitable. A final area where these tools may be applied is in quantifying the difference between a number of (virtual) incremental prototypes and a 'golden' product.

An illustration of this discussion can be seen by reference to Figure 1, which shows two pairs of ANEXT measurements.



(a)



(b)

Figure 1. Two pairs of ANEXT measurements (a) comparison of measurements 'A' and 'B' (b) comparison of measurements 'C' and 'D'

A valid question is which of these two sets of results is better and how can we justify this decision? This may originate from a desire to determine whether two measurement protocols are more repeatable, whether either of two facilities has better repeatability, whether either of two vendors can supply more consistent cabling. In trying to generate a rule-based comparison for this data, one may be tempted to ask the following questions:

1. Is the number of maxima / minima the same?
2. If not, how many differences are there and at what frequencies?
3. What is the amplitude difference between the two data sets for each feature in the traces?
4. How close do the amplitude trends agree across the graphs

As for comparing another iteration of the design with this (or comparing measurements taken by someone else), one may be asking whether the differences have increased or decreased. For the data in Figure 1, that could mean upwards of 20 – 30 individual pieces of information, all of which would need to be weighted differently. The next question then becomes how to identify these individual metrics and their weighting, which would probably be answered by suggesting a detailed survey of a large number of engineers given a substantial set of original data to compare. This brings us on to the next point about the tools used to support the decision making: they must bear some relation to the way in which a group of engineers would approach the comparison.

In comparing the individual traces of Figure 1, anyone so doing would be unlikely to overtly create a personal rule base. The measure of the quality of the comparison will be based on an overall 'feel' for the data based on experience and knowledge of the system. It is likely that two general aspects of the curves will be considered. Firstly, the overall envelope of the data would be considered, so the overall level would be considered to be somewhere between 'good' and 'excellent' but the differences below 10MHz would have an impact on the final decision. Secondly, the location, depth and quantity of the 'high – Q' features would be considered and because there is a fairly close mapping (but some differences around the nulls) it is likely that the assessment would be 'good' or even 'very good', giving a probable overall assessment of 'very good'. Note that the terms 'excellent', 'very good' and 'good' cannot be precisely defined *a priori* but only *a posteriori*, that is through usage and application by a knowledgeable body of analysts.

This concept of definition through usage rather than through declaration can seem confusing or even confused. However, section 2, dealing with knowledge, will show how this concept is regularly used.

Knowledge [1]

Knowledge is the basis on which all decisions are made, whether they are corporate strategy or tactical technical issues. In all cases, not all the important information will be known and it is the ability to make decisions based on imperfect or conflicting information, which almost defines a good engineer.

Knowledge is often considered to be either tacit or explicit, that is, knowledge which is known but cannot be codified and knowledge which can be codified. For example, the experimental procedures are readily committed to paper but the skills of the person who can say, by visual inspection, how good a set of results is are much more difficult to access. Hermenuetic knowledge, as proposed by Heidegger, is the tacit knowledge that underpins individual and collective understanding, it shapes what we do and how and why we do it, but we may not be able to explain to others why this is the case. Hence, it is not only the tacit knowledge that has been learned as part of a job, which needs to be considered, it is also the wider background of the individuals involved. This will also be discussed briefly later in this section when discussing Sensemaking.

One model that is useful to explain how the tacit knowledge of one person can be converted to explicit knowledge and/or used to develop tacit knowledge of another is the SECI model of Nonaka and Takeuchi[2]. This model operates as:

- S Socialization. Tacit to tacit knowledge sharing. The way in which tacit knowledge in one person develops as a result of the tacit knowledge of another. A craft apprenticeship is one example of this.
- E Externalization. Tacit to explicit knowledge sharing. This is the key to knowledge creation and can be undertaken by metaphor and analogy.
- C Combination. Excplicit to explicit knowledge sharing. This takes that which exists in written or other explicit form to generate new information.
- I Internalization. Explicit to tacit knowledge sharing. This is done by practicing what exists in written form by, for example, verbalization or creation of diagrams.

Taking, for an example, the comparison of measurements to assess repeatability, tacit knowledge sharing could come through an experienced engineer showing ‘good’ and ‘bad’ results to a less experienced engineer. This ‘socialization’ allows the experience of the one to pass to the other without being expressly stated. An element of externalization will occur when the less experienced engineer seeks to identify why one comparison is ‘good’ or ‘bad’ and, by query and enquiry, some rules emerge. Combination would then occur when these rules are shared with a wider audience or are built into procedures. Internalization would happen when the less experienced engineer took this set of rules and, through experience integrated them with the tacit knowledge developed from the experienced engineer. In the case of Figure 1 and the associated discussion, the development of the rule-base would be an example of “Combination”, but the more common occurrence of one engineer showing this to a less experienced one and saying simply “this is a good comparison”, with only a few observations, would be “Socialization”

Other writers [3] on the subject suggest that there can be no direct translation between tacit and explicit knowledge. They further suggest that groups as well as individuals may possess tacit knowledge and what an individual does is shaped by their own tacit knowledge, the explicit knowledge to which they have access, the tacit knowledge of the group to which they belong in this instant and the explicit knowledge to which the group has access. The combination of these factors results in "knowing as action"; essentially what is done by the individual and/or the group reflects the combined knowledge, both tacit and explicit, of the individual and the group. The model does help to explain how the outcome of analyzing the measurements by a group of engineers can differ to that when the results are compared by each engineer individually: the group possesses a different level of tacit knowledge to that of the individual. Figure 2 presents this illustratively. In addition to the knowledge of the individual, "Genres" refers to common understandings or behavior by the group, such as ignoring deep nulls in a graph but concentrating on the amplitude of the peaks, this action may not be overtly recognized, but everyone does it. Explicit group knowledge is represented by the standards to which everyone is working or stories emanating from the tacit group behavior.

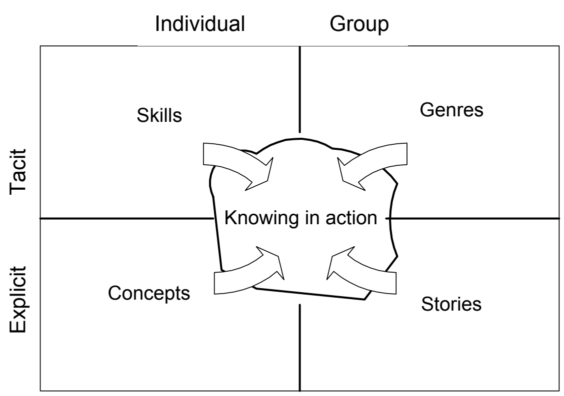


Figure 2 Knowing in action (adapted from [3])

The management of knowledge is the "explicit and systematic management of a vital resource and the process of creation, organization, diffusion, use and exploitation". Put another way, knowledge is a corporate asset, linked to organizational objectives and priorities, which is too important to leave to chance. Knowledge management can be thought of as dividing into three categories:

- Creating and discovering,
- Sharing and Learning,
- Organizing and managing.

The process of data comparison largely falls into the category of sharing and learning as the engineers involved are predominantly discussing their interpretation of the data, trying to identify how others deal with the data and agreeing on the decisions that are made as part of the data comparison. One particularly important 'structure' which can enhance the knowledge sharing is the *Community of Practice*. These are self-organizing, largely informal, networks with a shared purpose; its members may not even formally know its existence. The encouragement of these within an organization and, where possible, across organizations working in the same area, could be one significant means of developing an understanding of how the data is compared, why decisions are made from these comparisons in the way that they are, and how technologies and techniques can be developed to assist in the rigor of the decision making process. As they are informal social networks with shared purpose, it is important that there is a high level of mutual trust which goes beyond the formal structures. In fact, Communities of Practice usually exist in a way which can span organizations and the layers within an organization in a way that would be difficult to create formally. A factor which is particularly important as cabling organizations become more geographically dispersed. The management of the communities involves providing conditions to encourage social links and provide or encourage leadership in learning and enquiry, thus let informal links thrive. Some initiatives to promote these structures [4] include formal and informal events at which possible members can network, the introduction of learning project (providing new knowledge such as new comparison techniques) or managing the artifacts produced by a community (such as software, reports or papers). It is important that this sharing is undertaken at the right time; when engaged in a task *it is better to meet to do it rather than to meet to talk about it*. One technique that does help in the sharing of tacit knowledge is the exchange of 'war stories', the events organized as part of the support of a community of practice would be well placed to encourage this.

Some other approaches to sharing and developing knowledge are:

- Learning Networks. Rather more formal than the communities of practice, these are set up with the express intention of learning from each other and from outside the field of endeavour.
- Sharing Best Practice. It may be appropriate to instigate 'master-classes' to encourage those who are regarded as expert to share their knowledge with others.
- After Action Reviews. These started in the US military as a way of understanding the way in which decisions were made and actions taken. They are non-threatening and non-judgmental and are a positive way of learning from past actions and identifying both best and worst practice. The emphasis is on doing better next time rather than trying to identify how it could have gone better last time.
- Structured Dialogue and Interviews.
- Share-fares. Application of these in this setting may involve bringing examples of good practice to a general meeting and allowing 'delegates' to investigate other approaches to a problem in a more formally organized setting (although the actual interaction could be quite informal). One aim of these could be to encourage access to the many sources of knowledge held within an organization. Such knowledge could be held in the more formal structures such as libraries, within the culture of the organization (culture being "the way we do things round here"), transformations of information or of physical artefacts (referring to processes and procedures), structures within the organisation (roles and responsibilities), ecology (setting that shapes behaviour, such as the way the R&D department is physically laid out).
- Cross-functional Teams. One way to identify how and why someone has done what they did is to ask the naïve questions and challenge the basis on which this was done. One way of doing this is to build cross-functional teams, which, in this case, would not just include high frequency or electromagnetic engineers but may also include those involved in a marketing function. One rule which must be adhered to is that all members of a cross-functional team bring as much as they take; although it should be noted that the form of give and take may be different.
- Decision Diaries. Engineers are generally good at learning by doing and learning from experience. The effect of this is that the background to, and the reasons for, certain decisions being taken may not be initially obvious. This has the disadvantage that it is more difficult for less experienced engineers to learn from them and the development of knowledge technologies can be hampered by limited, or assumed, information. Decision diaries, or a critical analysis in a log book of how decisions were taken, where a personal post mortem is undertaken can help to alleviate this minimal information. It also has the effect of improving self examination, possibly minimizing the effects of progressively developed 'bad habits'.

Having considered a number of related concepts in the management of engineers' knowledge, one of the most important aspects of knowledge in practice is Sensemaking, i.e. turning data into action. One of the difficult aspects of this is to access the tacit knowledge of the members of a group. Some aspects, which may have relevance to the problem being considered, namely communities of practice and the SECI model, have already been outlined. The problem with trying to access tacit knowledge was summed up by Polyani as "We can know more than we can tell". Those doing data comparisons 'do it' but may not be able to describe how, what or why to others – this is a clear example of their tacit knowledge. There is a clear need to access this in order to codify the processes or ensure everyone does the same thing. (We must, however, be aware that "we know more than we are willing to tell.")

In the previous discussions, knowledge and information have been used in a loosely interchangeable manner. One concept, which helps to clarify any differences, and also clearly reviews the relationship of knowledge with, for example, data is the knowledge hierarchy. It also helps, perhaps, to identify at what level the sharing of abilities is required.

Data are the observations made as part of an experiment: the cross-talk readings for example.
Information can be thought of as data with context. Not only do we posses the cross-talk readings, but those readings are referred to the product under test, the test method and the standard's limit lines.

Knowledge is the information with meaning. Knowledge is the interpretation of that information e.g. trying to identify the causes of good / poor performance.

Wisdom is knowledge with insight. Perhaps the interpretation of the data with the details of the measurements known can lead to conclusions about the test method or the expectation of the behavior of, for example, the installation being tested.

If the requirement is only to share data or information, then this need only be supported by information technology, which simply helps in the representation of this data. If, however, knowledge or wisdom is required to be shared, knowledge technologies need to be investigated. The difference between the two is that while information technologies simply support the representation of the data, knowledge technologies support the interpretation of the data. An example of this could be that Excel® can provide an easily customized data representation, the visual rating chart to be discussed next in this paper provides a framework to access the individual's tacit and explicit knowledge without constraining their opinions and the Feature Selective Validation (FSV) method discussed later in this paper supports the interpretation of the differences between two sets of data.

If groups of people from different backgrounds and companies are involved in the comparison, it is essential that differences between the individuals be accounted for. These differences may include background, experience, expectation and expertise. This is important because it is likely that this difference may mean that the 'sender' and 'receiver' of the message will interpret the message in different ways. This is a further example of the importance of accessing the tacit knowledge of the individuals and the groups to which they belong in order to turn received 'data' into useable information.

So far, this paper has discussed some pertinent knowledge management models and has looked at the way in which the original data can be made sense of. The next section addresses some of the key management issues involved in knowledge management.

Visual rating [5]

As discussed in the previous section, quantifying comparisons of data for measurement repeatability or for validation requires the capturing of knowledge, both tacit and explicit. Some aspects of a good or poor comparison can be specifically stated but in many other aspects, it is a case of 'I know a good comparison when I see it'. It is important that any approach used to support quantifying data comparisons is compatible with this concept. That is, the group response and the individual response should both be accessible. To do this, a visual benchmark was developed [6]. The individuals will categorize a comparison according to the binary decisions, resulting in six quantified and qualified categories. The resulting rating scale is shown in Figure 3.

The group response can then be determined by taking the mean (or median as desired) of the collection of individual responses. It is important to appreciate that the various experiences and backgrounds of the individuals will produce a variety of responses and, as there is no 'universal truth' because, for an educated and informed group, all responses are equally valid. The breadth of responses can say as much about the data being compared as the mean response itself. A group response that is highly focused on one category would suggest a high confidence in ascribing that category to the comparison. A group response that is equally spread over several categories gives a low confidence that a single category is sufficient to summarise the comparison. Or put another way, a wide spread of opinion suggests that there is a lot of scope for debate and argument, whereas a smaller spread of opinion is clear-cut. One important aspect of a visual rating tool is that it should not force users into giving an accepted answer. As a consequence, it should maintain the group mean rating (comparing results with and without using the scale). A further consequence of such a rating scale is that because it provides some guidance, outlier opinion can be reduced without altering the overall average response for the group.

The choice of a six point scale is important because it results in a simple structure without excessive detail (as would result with a 10 point scale). Anything other than a binary decision tends to result in the middle option(s) being favored compared with those at the extremes, which introduces an extra (pseudo-)Gaussian confounding factor

A cursory review of the data in Figures 1(a) and 1(b) may suggest that comparison C-D is better than A-B. However, the visual rating scale encourages a more probing review. Using the Visual Rating Scale of Figure 3 on the data of Figure 1, both graphs would probably be regarded as 'Fair'. First, consider graph 1(a), while there is quite a difference in amplitude below about 20 MHz, the region between 20 MHz and 100MHz agrees well, there is a slight offset between 100 MHz and 400 MHz but above this, it is difficult to separate the curves. In general, above approximately 30 MHz, it is difficult to separate individual features. Secondly, consider graph 1(b), despite the offset, the general shape between the two measurements agrees quite well up to approximately 50 MHz, there is a clear difference between about 60 MHz and nearly 100 MHz but some differences in the trends above this as the 'envelopes' oscillate around each other. Thus, it would not be unexpected for each graph to be considered as having 'Reasonable agreement over many portions of the data'. A good case could probably also have been made to say that there is 'generally good agreement across the data' Depending on the background and interests of the engineers involved in assessing the data, some may regard the comparisons as having 'minor variations' and other 'minor agreement'. Hence, on that basis, it is difficult to separate the overall comparisons.

.

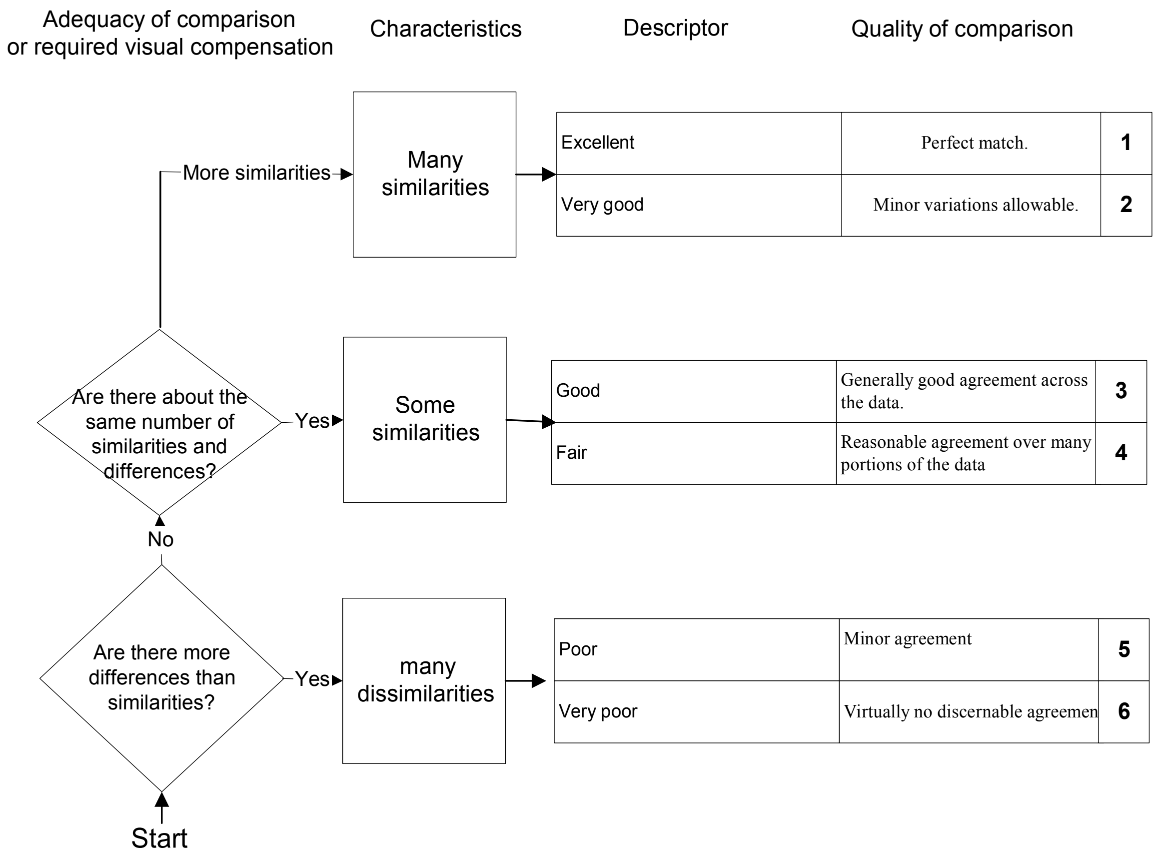


Figure 3 Visual rating scale (from [6])

Computer based assessment [7]
Having seen how individual and group knowledge can be accessed by using a simple rating chart, the next issue is can the data be processed automatically to achieve a similar level of information. Given the large quantities of data to be compared and the option to use optimization techniques in design which requires an objective measure of similarity (or fitness) [8], a computer based approach is very attractive. One such technique which is currently being considered as part of a forthcoming IEEE standard [9] is the Feature Selective Validation (FSV) method [10].

Initial development of the FSV technique was prompted by the need for error determination in the validation of numerical models against experimental data, the desire to assess the effects of incremental design changes on numerical models and the benefits coming from being able to quantify experimental repeatability. The main prompting factor was that there existed no other acceptable way of comparing the data. The common thread that ran through visual inspections was that two aspects of any visual data were considered and combined into an overall judgment. These were the envelope / trend of the data and any resonance like structure (these are referred to here as ‘amplitude’ and ‘feature’). The FSV decomposes the original comparison, by initially Fourier transforming the data sets, into components that contain the amplitude and trend information and components that contain the feature information. These are the Amplitude Difference Measure (ADM) and the Feature Difference Measure (FDM).

These are taken as independent functions and combined into an overall goodness of fit measure, the global difference measure (GDM).

All of the ADM, FDM and GDM are usable as point-by-point analysis tools or as a single, overall, measurement. The point by point analysis allows clear identification of the aspects of the initial data which are primarily responsible for the degradation of the comparison in a way which is objective and can be readily communicated. The overall comparison value gives a similarly objective interpretation of the overall agreement.
The ADM and FDM are obtained using the following equations.

$$ADM(f) = \left| \frac{\left(\left|Lo_1(f)\right| - \left|Lo_2(f)\right|\right)}{\frac{1}{N}\sum_{i=1}^{N}\left(\left(\left|Lo_1(i)\right| + \left|Lo_2(i)\right|\right)\right)} \right|$$

(1)

$$FDM_1(f) = \frac{\left|Lo_1'(f)\right| - \left|Lo_2'(f)\right|}{\frac{2}{N}\sum_{i=1}^{N}\left(\left(\left|Lo_1'(i)\right| + \left|Lo_2'(i)\right|\right)\right)} \qquad (3)$$

$$FDM_2(f) = \frac{\left|Hi_1'(f)\right| - \left|Hi_2'(f)\right|}{\frac{6}{N}\sum_{i=1}^{N}\left(\left(\left|Hi_1'(i)\right| + \left|Hi_2'(i)\right|\right)\right)} \qquad (4)$$

$$FDM_3(f) = \frac{\left|Hi_1''(f)\right| - \left|Hi_2''(f)\right|}{\frac{7.2}{N}\sum_{i=1}^{N}\left(\left(\left|Hi_1''(i)\right| + \left|Hi_2''(i)\right|\right)\right)} \qquad (5)$$

$$FDM(f) = 2\left(\left|FDM_1(f) + FDM_2(f) + FDM_3(f)\right|\right) \qquad (6)$$

where Lo_1 *and* Lo_2 are the intensities of the low frequency components of the data sets 1 and 2 at data point f. This is obtained by Fourier transforming the data and inverse transforming the lowest 40% of the data.

Hi_1 and Hi_2 are the high pass component of the data sets, obtained by Fourier Transforming the data sets and inverse transforming the highest 60%. The single primes (') indicate the first derivative of the inverse Fourier transformed data sets with respect to the x-axis and the double primes (") indicate the second derivative of the inverse Fourier transformed data. A simple central-difference-based scheme has been used to determine the first and second derivatives.

It should be noted that summing for the ADM and the FDM point-by-point values overall quantitative values of these components can be obtained. The single values allow ready overall assessment and the point-by-point values allow easy identification of regions of poor comparison, which dominate the overall similarity rating. While this is not terribly important for simple structured results, it is very useful for busy data. The main benefit of the point-by-point results is that these can help to identify regions where attention needs to be focused during validation of the model or in the post-mortem phase.

The Global Difference Measure (GDM) is then obtained as either a single figure of merit

$$GDM = \sum_{f_{\min}}^{f_{\max}} \sqrt{(ADM(f))^2 + (FDM(f))^2}$$

or as a point-by-point result:
or

$$GDM(f) = \sqrt{(ADM(f))^2 + (FDM(f))^2} \quad (7)$$

Natural language descriptors have been assigned to the output from this technique and are useful in communicating the results in a meaningful way. It should be noted that there is not theoretical maximum value for the FSV technique, although most comparisons fall into the range 0 (there is no difference between the data sets) and about 2 (excursions much beyond this are possible, but rarely for an overall comparison). The relationship between the numerical values (for ADM, FDM and GDM) and the natural language descriptors is given in Table I.

Table 1 FSV interpretation scale

FSV value (quantitative)	FSV interpretation (qualitative)
difference ≤0.1	Excellent
0.1 < difference ≤ 0.2	Very good
0.2 < difference ≤0.4	Good
0.4 < difference ≤ 0.8	Fair
0.8 < difference ≤1.6	Poor
difference >1.6	Very poor

It will be noted that the natural language descriptors used here correspond to the categories used in the visual rating scale. Probability density functions for the ADM, FDM and GDM, essentially the proportion of the point-by-point analyses within a particular qualitative category, which appear to mimic the spread that would be obtained by a group of engineers. Initial results for the FSV are very encouraging [5].

The FSV is particularly useful where the results to be compared are visually complex (that is, they could not be described to a third party with a single paragraph of text) but were individual features can be clearly discerned. Where individual features cannot be clearly discerned, but the envelope can, the Integrated Error against Logarithmic Frequency (IELF) method may be more suitable.

The (IELF) method [11] is based on the premise that in comparing data with a very high feature density, the overriding factor to be assessed is a function of the difference between the two traces. A single figure is obtained for comparison purposes by integrating (summing) the difference over the frequency range taken on a logarithmic axis. The same technique can also be applied to frequency bands (sub-frequencies), where there is a particular physical or system-dependent driver for doing this. The basic IELF equation is given in equation 8.

$$IELF = \frac{\sum_{i=1}^{n-1} |error_i| . \{\ln(f_{i+1}) - \ln(f_{i-1})\}/2}{\ln(f_n) - \ln(f_0)}$$

(8)

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

A small modification to the IELF involves summing the elements halfway between the data points in order to improve the approximation to the difference in the measured data. This modification is given in equation 9.

$$IELF\ \mathrm{mod} = \frac{\sum_{i=1}^{n-1} |error_i| . \{\ln(\frac{f_{i+1}+f_i}{2}) - \ln(\frac{f_i+f_{i-1}}{2})\}}{\ln(f_n) - \ln(f_0)}$$

(9)

As with FSV, in the IELF method, a value of zero indicates a perfect comparison. There is no upper limit on the quality factor produced, which enhances its ability to discriminate differences in overall quality for poor comparisons as well as for close comparisons.

With reference to the data of Figure 1, the IELF method, using equation 8 gives a value for Figure 1(a) of 7.8 and for 1(b) of 5.5. Clearly suggesting that 1(b) is better – although the lack of scaling does not allow a more absolute level to be gauged nor a conclusion drawn as to how much it is better. Figure 4 compares the differences on which the IELF values are based.

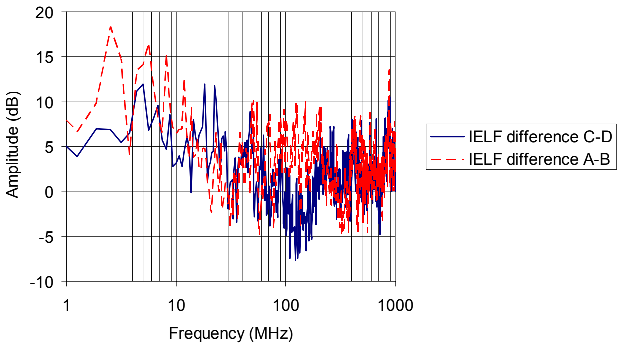


Figure 4 IELF differences

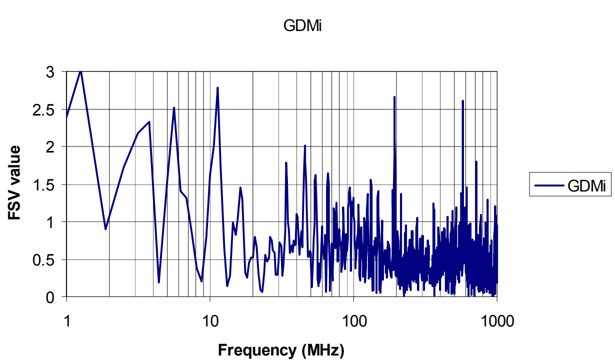
One observation from the differences is that the comparison C-D has less error below approximately 20 MHz and between 50 MHz to 200 MHz than comparison A-B.

Applying the FSV routine to these comparisons gives the values for the ADM, FDM and GDM as in Table 2

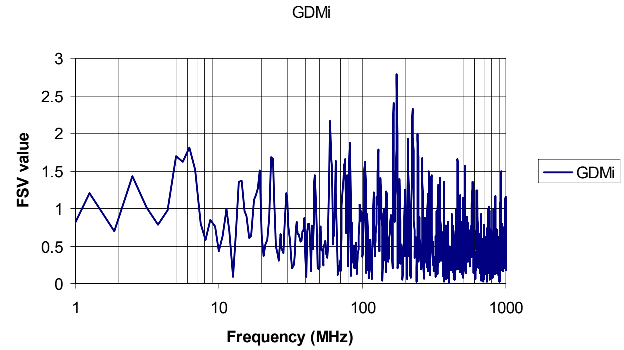
Table 1 FSV values for the data of Figure 1

	Comparison A-B	Comparison C-D
ADM	0.24 (Good)	0.26 (Good)
FDM	0.42 (Fair)	0.43 (Fair)
GDM	0.52 (Fair)	0.55 (Fair)

So the FSV routine suggests that there is not much difference in the original comparisons overall, but A-B has a slight advantage. The reasons can be seen with reference to Figure 5, the point by point analysis of the Global Difference Measure.



(a)

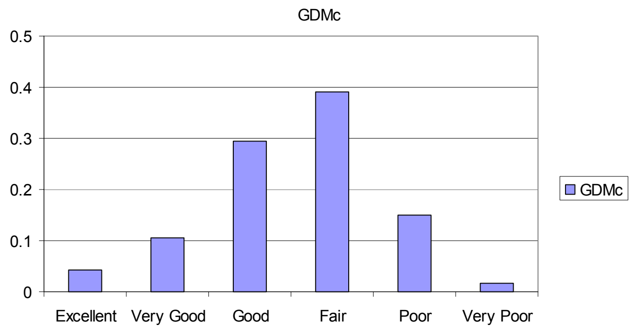


(b)

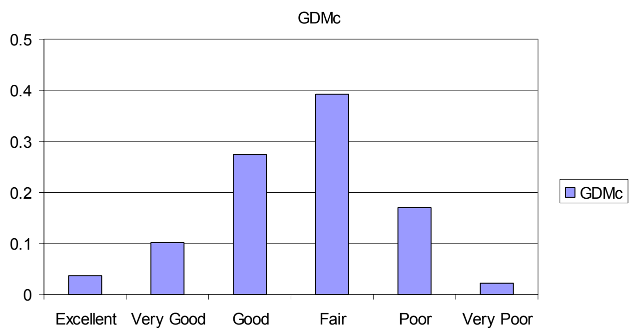
Figure 5 FSV Global Difference Measures (a) for Figure 1(a) and (b) for Figure 1(b).

This is particularly important in that it shows that above approximately 70 MHz comparison A-B (figure 5(a) is more consistently between 0.5 and 1 – 1.5 but the comparison C-D alternates between regions of very low GDM to regions actually higher than for comparison A-B. An important factor is that comparison C-D is better below approximately 20 MHz than comparison A-B. A benefit of the point-by-point analyses is that they can help to direct efforts intended to improve the comparisons, measurements, or models and thereby marshal limited resources. Considering Figure 5(a), FSV is drawing attention to the lower frequency differences (<10 MHz) and to two features at ~ 200 MHz and 600 MHz: it would be helpful to investigate these first. For Figure 5(b), interest in the differences around 200 MHz should be considered.

Given the observation that the confidence histograms show broad agreement with the overall opinions of groups of engineers, it is instructive to review the FSV data and compare back to the initial observations of the visual rating scale. Figure 6 shows the confidence histograms for the data of Figure 5.



(a)



(b)

Figure 5 FSV Global Difference Measure Confidence Histograms (a) for Figure 5(a) and (b) for Figure 5(b).

The initial observation is that there is marginally more of Figure 5(a) that would be regarded as being 'Good' and slightly less 'Poor' and 'Very Poor'. The important point here is that the broadness of the histograms suggest very little emphasis should be placed on the second decimal place.

Discussion

This paper has described a number of approaches to quantifying what may normally be a subjective assessment of data using a visual rating scale, the Integrated Error against Log Frequency (IELF) method and the Feature Selective Validation (FSV) method. The advantages of each have been introduced. Testing the approaches involved a difficult challenge of identifying which of two pairs of data has a better comparison. A brief visual review of the data would suggest Figure 1(b) to be better. However, a more detailed visual review, supported by the visual rating scale, made the decision much more difficult, i.e. it would be difficult to decide between the two using the visual rating scale (although this would only be a proper test if it involved a statistically significant group size) but it helps to justify why the decision is difficult. The IELF method said that C-D was a better comparison but there is no absolute scale, so it is difficult to judge by how much. FSV suggested that A-B is a better comparison, but by very little.

It should be noted that FSV works on a linear rather than logarithmic scale in its current implementation but the results were plotted on a logarithmic scale for consistency; IELF is naturally based on a logarithmic scale.

What this paper has shown is that there are tools available that can help in quantification, and therefore in objective decision making. There is no single 'right' approach but a significant benefit of using any / all of these approaches is that it encourages an objective discussion about the data which fosters a combination of the individual's tacit and explicit knowledge coming to play in the individual's contribution to the overall opinion of a group.

Acknowledgements

The work reported in this paper is the subject of international interest, particularly within the EMC community. It gives the authors pleasure to acknowledge the ongoing support, contributions and critical dialogue of Professor Antonio Orlandi (University of L'Aquila, Italy), Dr Anthony Martin (MIRA Ltd, UK), Dr Dawn Coleby (University of Leicester, UK), Dr Bruce Archambeault (IBM, NC, USA), Dr Andrew Drozd (ANDRO Computational Solutions, Rome, NY, USA), Dr Christopher Jones (BAE Systems, Warton, UK) and Professor Robert Simpson (Simtronics Ltd, Preston, UK)

References

[1] AP Duffy, "EMC Comparison: A technical problem or a knowledge

management issue?", Zurich EMC Symposium, 2003

[2] I. Nonaka and H. Takeuchi, "The knowledge-creating company: How Japanese companies create the dynamics of innovation", 1995, Oxford University Press.

[3] SDN Cooke and JS Brown "Bridging epistemologies: the generative dance between organizational knowledge and organizational knowing", Organizational Science, 1995

[4] E Wenger, "Communities of Practice", 2000, Cambridge University Press.

[5] B Archambeault, S Connor, A Duffy, "Comparing FSV and human responses to data comparisons", IEEE International Symposium on EMC, Chicago, 2005.

[6] DE Coleby and AP Duffy, "A visual interpretation rating scale for validation of numerical models", International Journal for Computing and Mathematics in Electrical and Electonic Engineering (COMPEL) in press.

[7] A Duffy, A Martin, G Antonini, A Ciccomancini Scogna, A Orlandi, "Issues in validation of complex valued simulations for signal integrity analysis" IEEE International Symposium on EMC, Santa Clara, 2004

[8] C Dole, H Sasse, A Duffy "Particle swarm optimization for broadband coaxial cabling", 2004 IWCS, Philadelphia.

[9] http://grouper.ieee.org/groups/emc/emc/ieee_emcs_-_sdcom/p1597/ieee_emcs_-_sdcom/p1597/ieee_emcs_-_p1597_main.htm

[10] AJM Martin, "Quantitative data validation (Automated visual evaluations), PhD Thesis, De Montfort University, UK, 1999.

[11] RJ Simpson, CR Jones, I MacDiarmid, A Duffy, D Coleby, "The Integrated Error against Log Frequency (IELF) method for CEM validation", IEEE International Symposium on EMC, Chicago, 2005.

Author biographies

Alistair P. Duffy was educated at Conyers School, Yarm; University College, Cardiff, and Nottingham University. He is currently a Principal Lecturer in Communication Engineering at De Montfort University, Leicester. Particular research interests include experimental, numerical and analytical applications for electromagnetics. Particular emphasis for this in recent years has been in communication cable design and analysis. Other aspects of this interest include microwave and rf heating and novel antennas. A substantial interest is in the formal comparison of complex data sets.

Kenneth G. Hodge holds a BEng from Liverpool University and an MSc from Warwick University. He is currently a representative to BSI and CENELEC committees working on LAN cabling. Ken has been working in the cable industry since 1982, he was initially involved with the development of optical and copper data communication and telecommunication cables, then transferred to the R&D facility as Design Manager Optical Cables. He is now Technical Manager at Brand-Rex (Part of Honeywell).