Summary of the Open Space Technology Discussions

Marija BERTOVIC, Pierre CALMON, Luke CARTER, Jay FISHER, David FORSYTH, Ralf HOLSTEIN, Bernard MCGRATH, Christina MÜLLER, Mato PAVLOVIC, Ulf RONNETEG, Ward RUMMEL, Frank SCHUBERT, Greg SELBY

1. Introduction

The highlight of 5th European American Workshop on Reliability of NDE (5th EAW) was the "Open Space Technology"(OST) discussion. The aim of the 5thEAW and, in particular, of the OST discussion - was to shed light on the reasons, which create a "*delta*" between all the qualifications and reliability models and the real reliability in the field.

In the previous 4 Workshops, the different topics of interest were discussed in the so-called *break-out sessions*, in which a specific goal for each working group was defined beforehand. In order not to move on "the old rails" in search for an answer, the break-out sessions were this year replaced by an *open space* approach. The reason for this was not to restrict the participants to a certain number of topics, but rather to allow the topics to arise from the interests of the participants and the needs of both the researchers and the NDT practitioners. This was achieved through setting up a wall of ideas. Each of the participants was asked to write down as many questions or topics as they wanted, and pin them up on the wall throughout the duration of the entire Workshop. Finally, when the participants gathered for the discussion, those ideas were organized into 6 main topics:

A. New reliability methods: Multiparameter POD, MAPOD, Bayesian
B. Structural health monitoring
C. Applications in industry
D. Human factors
E. Integrated solutions and *"delta"*
F. Basic concepts of reliability of NDE

As in the spirit of the OST, in which the participants themselves determine the topics and decide how the discussion will flow, the participants from the topics C, D and E joined their forces and disused together.

The groups separated and, under the moderation of several experts in different NDT reliability fields, the different issues have been discussed. At the end of the day the groups were asked to present the content and the results of their work. Every group wrote a summary of the discussion.

The purpose of this paper is to present the summaries of the held discussions and, therewith, share with the NDT community the main conclusions.

2. Open Space Technology

Open Space Technology, developed by Harrison Owen in the sixties, is one of the simplest methods to bring a big group of people together to discuss the topics that interest them [1, 2]. The choice of a topic, the duration, the participants and the flow of the discussion are in the participants' hands.

The success of the discussion relies on the commitment and interest of everyone involved and follows simple principles and only one law. The so-called "law of two feet" relies on the premise that everyone carries responsibility for themselves. If the participant is not contributing, not learning from the discussion, then he or she should use their feet and move on to another place, which allows them to be productive.

The principles:

1. *Whoever comes is the right people* (people from different branches, people with different backgrounds, different positions etc. are all invited to join).
2. *Whenever it starts is the right time* (spirit and creativity do not run on the clock)
3. *Wherever it happens is the right place* (in the discussion room or at lunch break, the discussion never stops)
4. *Whatever happens is the only thing that could happen* (the participants should let the discussion take its own flow, without trying to control it in any way).
5. *When it's over, it's over* (there are no rules how long a discussion should last. If it has reached its end, it is finished. Move on.)

Following these simple guidelines, the participants in the OST discussion held during the 5th EAW were instructed the following:

- Every participant is invited to attend any discussion topic he/she likes.
- Participants are free (and encouraged) to move from one session to another and openly discuss the topics with their peers
- Each session will result in a short written summary of each topic discussed
- The summary of each topic will be presented to the whole workshop audience at the end of the day.

As a result, in the following chapters, the summaries of the discussed topics will be presented.

3. Group A: New Reliability Methods (David Forsyth & Pierre Calmon)

The planned purpose of Group A was to be a discussion of "New Reliability Methods: Multiparameter POD, MAPOD, Bayesian". The questions submitted by conference attendees during the week focused on "modelled" and Bayesian methods, and these were the subjects of the discussion.

The discussion began with Bayesian topics. The purpose of this is to combine data from multiple sources, including "modelled", experiments, and in-service inspections. Bayesian methods are one way to do this, and are flexible, robust, and well known. In particular, "Bayesian inference is a method of inference in which Bayes' rule is used to update the probability estimate for a hypothesis as additional evidence is acquired" (see http://en.wikipedia.org/wiki/Bayesian_inference).

The typical method proposed to do this is to choose one source of data as the prior, defining the parameters of the POD estimate. Then we use the other sources of data to update the parameters of the POD estimate according to Bayesian methods. This leads to a few key choices for the engineer or scientist responsible:

1. What data shall serve as the prior?
2. How will the new data be weighted?
3. Will the new data be checked for consistency with the original assumptions?

The importance of using Bayesian methods instead of simply pooling the different data was discussed. The advantages of Bayesian methods over pooling are that the data can be weighted differently, and different distributions can be combined.

A key outcome of the discussion was the suggestion that an application guide including worked problems would be a great benefit. This is a potential subject for the next Workshop to undertake.

The next main topic of discussion was the use of models for POD estimation. There is general acceptance of this proposition, assuming that the models have been validated over the range of inputs associated with the specific problem at hand. By providing distributions on inputs to models, and execution of models in Monte Carlo type methods, a distribution of outputs can be found. One topic of discussion is how this modelled distribution of results relates to the confidence bounds on POD determined using empirical methods. It was noted that Bayesian methods may provide a solution to combining these.

Other than directly modelled the complete inspection, it was noted that modelled or simulated data could be used to modify inspection data to get a better POD estimate in cases where it is difficult to make fully representative specimens. Noise due to different material lots or non-relevant signals due to extraneous geometries are examples of this.

In summary, this was a lively discussion marked more by agreements than disagreements. There are many open questions remaining regarding how to execute POD estimation using the different possible data sources and combinations thereof. Further research combined with sharing of experience through development of benchmark problems, guidebooks, worked examples, etc. will be beneficial to the community.

4. Group B: Structural Health Monitoring (Jay Fisher)

The main discussion areas were:

- the ways in which structural health monitoring (SHM) is different from conventional non-destructive evaluation (NDE)
- reliability issues of SHM systems
- issues concerning determination of reliability of SHM

Determination of POD for SHM will be different from that for NDE. In particular, for SHM there may not be a fixed threshold. One reason is that the noise level can vary over time, as well as other conditions. In NDE reliability determination, it is assumed that there is a fixed noise level, sensor performance level, and environmental conditions. These conditions cannot be counted on for realistic SHM. In particular, during the course of monitoring a structure, the mechanical loading and temperature can change. It was also noted that the performance of an SHM system will vary based on the relative location of the SHM sensors and flaws.

The possibility of making SHM fit into the maintenance system was brought up. It might even be possible to use information from SHM sensors to provide operational parameters to supplement existing system diagnostics. For example, an ultrasonic system can be used to obtain temperature information as well as defect information, based on changes in ultrasonic wave velocity.

Issues related to sensor reliability were discussed at length. In order to have practical SHM, there should be a way to check the sensor performance. The sensors should be designed with reliability in mind. Ways to do this include would be to

- design with some level of redundancy,
- minimize the number of single-point failure modes,
- include sensor self-check capability
- design a calibration reflector in the structure with the SHM system
- plan for long term hardware and software support of installed SHM systems

In practice, one difference between SHM and NDE system implementation decisions is that in addition to sensor reliability, the SHM system will also have to balance weight and space against performance and cost.

It was decided that defining POD for various inspection methods are philosophically similar, but the method to introduce and locate flaws will be different.

The role of human factors was discussed. It was agreed upon that human interpretation is still important.

There should be demonstrations of SHM installations that cover small areas but have high payback, as a way to gradually move industry to acceptance of SHM.

What needs to be done?

- Better ways to characterize reliability for SHM are needed.
- We need to determine how to define POD when conditions of the structure may change.
- Methods to qualify SHM systems should be developed.
- General methods to determine where to place sensors should be developed.

5. Group CDE: Definition of Requirements of NDE by Customer vs. Provider (Christina Müller)

The group discussed the question of the definition of requirements on NDE from the customer (end user) point of view in contrast to the requirements seen from the NDE research and provider point of view. A gap was discovered between both positions and means to overcome discussed.

What was discussed? Specifically the representatives of end users (e.g. power plants, aero-space industry etc.) listed the aspects they are urged on when defining the requirements on NDE as:

- Safety margins
- Codes & regulations
- 100% detection (of critical defects)
- Dependent on the number of defects and their characterization
- Difference between Structural Integrity and NDE
- A POD including the Human Factors would be desirable for risk assessment

The NDE community (provider and research organizations) would define the NDT reliability as:

- The detection capability with respect to parameters (researchers)
- Being ahead of what is actually required by end user (researchers)
- Being a product of the application of procedures and standards (provider)

What needs to be done? The gap between both positions could be found in an **agreement on reasonable targets**. This requires in its turn an adequate information management between both parties.

6. Group CDE: What Value of POD is Good Enough? (Luke Carter)

The general consensus of the group on the topic of what value of POD is good enough was that the necessary POD should be decided jointly between the structural integrity and NDT groups. Other comments and suggestions include:

- A higher POD value brings higher cost in demonstration and implementation.
- Refinement of procedures until required POD is achieved (Recursive POD).

7. Group CDE: Human Factors (Marija Bertovic & Luke Carter)

Another discussion topic that received a lot of attention during the Workshop was the human factors (HF) topic. It appears that there is a lot of interest in the influence of human factors on the reliability of NDE, at least from the research community. However, there is a gap in the communication between the utilities and the service providers, causing problems in the transfer of knowledge and, hence, posing a difficulty to implement the findings in the field.

The first issue discussed was the shared responsibility of dealing with human factors between the NDT service provider and the customer. It was concluded that the human factors do not always receive the appropriate attention in the field, and that the communication between the customer and NDT provider is a big part of the problem. The customer frequently wants low cost and does not always include human factors into the consideration. The inspection vendor, in contrast, tries to fulfil the requirements of the customer, sometimes at the expense of human factors and NDT reliability in general. The question is: is the customer even aware of the human factors influence on the POD and whose responsibility is to inform them? The existence of standards, codes, requirements for training and experience were acknowledged, as was the existence of HF tools (computer – based training, pre- and post-debriefing, self-checking, etc.), especially in the nuclear field, well spread in the USA and slowly followed by the Europeans.

Second issue discussed was of a more practical nature, i.e. how to keep an inspector vigilant, even when he never sees a flaw. Examples from different industries were given, e.g. railway axle inspection, where in the course of 5 years only 87 defects have been found. This was recognized as an important issue in different application branches and the need for solutions was expressed.

Third topic was related to the practical needs for future research in the field. It seems there is a communication gap between what is known about human behaviour under difficult working conditions in psychology and what is known by the engineers.

What needs to be done? A broad discussion and raising of awareness is needed between the customer and the NDT service providers. Considering that the customer is no NDT expert, it is up to the NDT community to spread the word. However, the customer should take over the responsibility and consider the benefits and the costs of considering NDE reliability and human factors.

Re-qualification to refresh operator knowledge and skills for detection and characterisation, recording geometric indications to keep operators occupied during their task, yearly practice on test components with realistic defects, engagement with staff in non-outage time (developing procedures, performing open trials) were some of the suggested solutions for the vigilance problem.

Relying only on the experience and qualification of the inspecting personnel has shown to be a flawed approach. Influences of quality and recency of experience, as well as loss of skill between certifications, have to be acknowledged. Further topics of interest for the future research include:

- the influences of working conditions on the inspectors (shift work, night shifts, vigilance, fatigue, heat, noise etc.), or the transfer of that knowledge from the social sciences to the engineers and the inspecting personnel;
- the extent of the effects of social loafing on human redundancy and whether human redundancy is at all an appropriate approach;
- improvement of the interaction between the customer and the service provider
- deeper insights into the human factors issues in mechanized NDT ,especially during data evaluation
- HF aspects of the management and the organization, in general.

8. Group CDE: Manual vs. Automated Inspection (Ulf Ronneteg)

The group discussed, among a number of topics the question –What advantages and disadvantages regarding the reliability can be foreseen in automated, mechanized vs. manual inspection?

First of all it was determined that there is a clear difference between automated and mechanized inspections. In the automated inspections, data collection, data evaluation and even decisions are performed automatically while in the mechanized inspections normally the data evaluation and the decision making are done by the human inspector.

In general it was concluded that the reliability is higher in mechanized inspection compared to manual inspection, but it was stated that “the mechanical system is not always as good as the best manual inspector but at least better than the average manual inspector”

Regarding mechanized inspections some advantages and some disadvantages where discussed and the most important is summarized below:

+ Normally better data coverage and especially better traceability of the data and coverage
+ The data evaluation can be done by different experts in the future (good if new design criteria arise)
– The data evaluator can miss the feeling from doing the inspection, i.e. if something unexpected occurs during inspection that the operator not document or if the actual inspected part differ from the drawings, this might not be captured by the data evaluator

There were also some discussions about fully automated inspections and the general remark about this is that the need for correct input data is extremely important. Otherwise the inspection could be very unreliable if something is wrongly set.

The final conclusion from the group is that independent of the level of automation, the human is to some extent always involved, i.e. human factors always exists with the risk of human errors.

What needs to be done? In general it needs to be spread out that there still are needs to investigate the reliability of more or less automated inspection systems. Especially as it has been concluded that the errors that might occur often are different from errors in manual inspection. One of the most important tasks in mechanized inspection is the evaluation of data and as it often includes interpretation of complex signals and/or images, there is a need for high quality procedures.

The general conclusion for what needs to be done is recommendations/guidelines for written procedures regarding format, structure and content. It should be emphasized what should be considered in writing and reviewing of the procedure, involving the user in the process (user-centred approach). It should also be pointed out that the inspector should be informed about the scope of the inspection and that the procedure is well understood and accepted. Finally, this information needs to be available, well known and accepted by the whole NDT "world".

9. Group F: Basic Concepts of Reliability of NDE (Mato Pavlovic & Ward Rummel)

The group discussed, as its name suggests, the basic concepts of reliability of NDE. The number of questions posted on the wall, as well as the number of participants in the discussion, suggests that even if the NDE reliability has gone a long way, there is still a need to discuss the fundamental principles that form the basis for all advanced concepts.

The first topic discussed were the confidence bands on the POD curve. It was stressed that the width of the confidence band comes from the experiment only and is determined by the sample size. Only the lower confidence band is of interest for determining the a90/95. Also, it has been noted that there is a difference in calculation of the confidence bands when calculation is performed as described by Berens or according to MIL 1823. Consequently, this results in different a90/95. It has been concluded that calculations are based on different mathematical model, hence the difference.

The question was raised why there in no more publicly available POD data so that one can test newly developed models. Even if, due to the frequently sensitive nature of the data, there is limited amount of data, there is still some available. For example a data-book updated regularly, published by Ward Rummel.

Next topic was the applicability of POD curves on different kinds of defects. It was concluded that NDE reliability engineers have to work closely together with stress and materials engineers.

On the question of possibility to compare POD data from different sensors, it was concluded that it is indeed possible and that POD can be used as a tool to compare performances of different sensors.

The question was followed by the discussion about how is risk included in the POD analysis. As described in the MIL 83444, first the crack growth rate is calculated. The time of the failure is derived from this rate and then the inspection interval is set at the half of this time.

It has been also discussed the role of the threshold and who decides on it. As it is known, every system can have 100% (low threshold) or 0% (high threshold) detection rate, but what is changing is the false calls rate. It has been pointed that even though stress engineers would like to have the threshold as low as possible, they should not be those that decide.

What needs to be done? From the wide spectrum of questions one can conclude that there is a need for learning materials in form of the text book and possibly organization of courses and tutorials on the topic of reliability of NDE. Interested party would have the opportunity to learn in an easy and understandable way the elementary concept of the reliability. These materials will give a jump start for those entering the field of NDE. On the other hand, as one hears in informal talks with the participants of the Workshop, the materials would be a great help also for those that already have an experience in the NDE, but missed to understand properly some of the basic concepts of reliability.

10. References

[1] Seliger, R. "Einführung in Großgruppen-Methoden", Heidelberg: Carl-Auer Verlag, 2011

[2] Harrison, O., "Open Space Technology: Ein Leitfaden für die Praxis", Schäffer-Poeschel Verlag, 2011.