EUROPEAN COMMISSION
Community research

PAMINA
Performance Assessment Methodologies in Application to Guide the Development of the Safety Case

(Contract Number: FP6-036404)

AN EXPERT JUDGEMENT PROTOCOL TO ASSESS SOLUBILITY LIMIT DISTRIBUTIONS FOR KEY CHEMICAL ELEMENTS IN A GENERIC SPANISH REPOSITORY IN GRANITE
MILESTONE (N°:M2.2.A.4)

Author: **Ricardo Bolado**
(Joint Research Centre, European Commission)

Date of issue of this report : **20/06/08**

Start date of project : **01/10/2006** Duration : 36 Months

Project co-funded by the European Commission under the Euratom Research and Training Programme on Nuclear Energy within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
RE	Restricted to a group specified by the partners of the [PAMINA] project	
CO	Confidential, only for partners of the [PAMINA] project	

Foreword

The work presented in this report was developed within the Integrated Project PAMINA: **P**erformance **A**ssessment **M**ethodologies **IN** **A**pplication to Guide the Development of the Safety Case. This project is part of the Sixth Framework Programme of the European Commission. It brings together 25 organisations from ten European countries and one EC Joint Research Centre in order to improve and harmonise methodologies and tools for demonstrating the safety of deep geological disposal of long-lived radioactive waste for different waste types, repository designs and geological environments. The results will be of interest to national waste management organisations, regulators and lay stakeholders.

The work is organised in four Research and Technology Development Components (RTDCs) and one additional component dealing with knowledge management and dissemination of knowledge:

- In RTDC 1 the aim is to evaluate the state of the art of methodologies and approaches needed for assessing the safety of deep geological disposal, on the basis of comprehensive review of international practice. This work includes the identification of any deficiencies in methods and tools.

- In RTDC 2 the aim is to establish a framework and methodology for the treatment of uncertainty during PA and safety case development. Guidance on, and examples of, good practice will be provided on the communication and treatment of different types of uncertainty, spatial variability, the development of probabilistic safety assessment tools, and techniques for sensitivity and uncertainty analysis.

- In RTDC 3 the aim is to develop methodologies and tools for integrated PA for various geological disposal concepts. This work includes the development of PA scenarios, of the PA approach to gas migration processes, of the PA approach to radionuclide source term modelling, and of safety and performance indicators.

- In RTDC 4 the aim is to conduct several benchmark exercises on specific processes, in which quantitative comparisons are made between approaches that rely on simplifying assumptions and models, and those that rely on complex models that take into account a more complete process conceptualization in space and time.

The work presented in this report was performed in the scope of RTDC 2.

All PAMINA reports can be downloaded from http://www.ip-pamina.eu.

Proposal/Contract no.:	**FP6-036404**
Project acronym:	PAMINA
Project title:	**PERFORMANCE ASSESSMENT METHODOLOGIES IN APPLICATION TO GUIDE THE DEVELOPMENT OF THE SAFETY CASE**
Instrument:	Integrated Project
Thematic Priority:	Management of Radioactive Waste and Radiation Protection and other activities in the field of Nuclear Technologies and Safety

Task 2.2.A Parameter Uncertainty
Topic 5: Expert judgement techniques for assigning PDFs

An Expert Judgement Protocol to Assess Solubility Limit Distributions for Key Chemical Elements in a Generic Spanish Repository in Granite

Due date of deliverable:	05.30.08
Actual submission date:	06.20.08
Start date of project:	10.01.2006
Duration:	36 months

JRC

Revision: 2

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Dissemination level

PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	X
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

TABLE OF CONTENTS

1.- Introduction

This report contains the description of the protocol that is going to be applied to characterise the uncertainty about the solubility limits for some key chemical elements (Radium, Tin, Selenium, Uranium and Plutonium) in a generic Spanish Radioactive High Level Waste repository in granite. The design of this protocol and its actual application are done within the framework of PAMINA's RTDC2 (treatment of uncertainty), and specifically under Work Package 2.2, task A, topic 5 (task 2.2.A, topic 5).

Expert Judgement (EJ) has been used during roughly the last seventy years in different areas of science, technology, weather forecasting, strategic planning, economy and many other fields as a reasonable way to assess uncertainties about events and variables when the source of uncertainty is lack of knowledge (epistemic uncertainty). Since the design of the pioneering Delphi method, several structured protocols have been proposed and improved thanks to the experience acquired in many applications. Nuclear Safety has been an extremely fertile field for the application and improvement of these processes. During the mid 1980's, researchers from Sandia National Laboratories (SNL) in collaboration with experts in the area of EJ developed a protocol to provide information in large scale risk studies, namely Probabilistic Safety Analyses (PSA) of Nuclear Power Plants (NPPs) and performance Assessments (PA) of Radioactive High Level Waste (HLW) repositories, see Bonano et al. (1990) and Gorham-Bergeron et al. (1991). This protocol is described by Bolado and Badea (2008) and is referred to as the SNL/NUREG-1150 protocol.

The protocol proposed in this document takes as a basis the SNL/NUREG-1150 protocol and modifies it with the intention to adapt it to small scale applications, i.e. a few variables or events to be assessed, and to take advantage of the experience and knowledge acquired by some members of our group after either participating in (Cojazzi et al, 2001, Simola et al., 2005), or coordinating (Bolado & Ibáñez, 1999, Bolado & Gallego, 2000, Bolado & Iglesias, 2001, Bolado et al., 2002, Bolado, 2005) several expert judgment projects, most of them dealing with issues within the area of severe accidents in NPPs. Additionally, it takes into account the opinion of the U.S. Nuclear Regulatory Commission (USNRC) staff about the basic activities that any EJ protocol should include in order be able to deliver high quality information, see Kotra et. al. (1996).

2.- Protocol steps

The proposed protocol consists of the following steps

1. Selection of the project team.
2. Preparation of supporting material and definition of the questions to be studied.
3. Selection of experts.
4. Training sessions.
5. Refinement of the questions to be studied.

6. First individual work period.
7. Presentation of individual approaches adopted by the experts.
8. Second individual work period.
9. Elicitation sessions.
10. Analysis and aggregation of results.
11. Review.
12. Documentation.

These steps may be grouped into three major phases:

1. pre-process (steps 1 to 3, developed before the actual participation of the experts, where the key players are the organisation interested in getting the opinions of experts and the project team),
2. process (steps 4 to 9, where experts participate actively) and
3. post-process (steps 10 to 12, where the project team plays the key role, post-processing the opinions of experts and delivering the application final report).

The development of this protocol involves three meetings that must be attended by all experts and the project team: 1) the training sessions (step 4) and the refinement of the questions to be studied (step 5) are performed in the first meeting, which takes two days and involves the project team and all the experts simultaneously, 2) the session to describe individual approaches (step 7) requires again the concurrence of the project team and all the experts simultaneously in a second meeting, which takes one day, and 3) the elicitation sessions require a third meeting, which involves the concurrence of the project team and the experts, in this case one by one since opinions are elicited individually in this protocol. These sessions require one day per expert.

Three are the main innovations introduced in this protocol with respect to the SNL/NUREG-1150 protocol and to other protocols: The addition of extra training sessions, the division of the individual work in two phases separated by a meeting where each expert shows the results of his/her preliminary study of the questions addressed and the flexibility in the selection of techniques to aggregate the opinions of the different experts taking into account different aspects.

3.- Protocol description

This protocol has been designed to assess the distributions that best characterise the state of knowledge of our experts about variables of interest (model parameters) and probabilities of events. It hasn't been designed to identify scenarios or to deal with model uncertainties. In what follows is a description of the steps considered in the protocol.

3.1.- Selection of the project team

The project team consists of the analysts and the generalists. The analysts are in charge of managing the whole process. The analysts are the first persons selected and they have to start the process. Usually they are appointed by the organisation interested in getting the results of the study. They should already have participated in other EJ applications since experience is an important point for the correct development of the whole exercise. Additionally, they should have a deep knowledge in probability, statistics, knowledge psychology and techniques to get opinions from experts. They should also be skilful at interacting with people and at moderating group discussions. The number of analysts depends on the magnitude of the exercise to be developed. In this application one analyst is enough.

The generalists have to provide support to the analysts in all matters related to the questions to be assessed, as for example the decomposition of the question in sub-questions and identifying and obtaining information from different sources. Generalists should have general knowledge about the area under study, not being necessarily top researchers in it. Generalists are key persons that know the needs of information in the project where the distributions to be obtained will be used. They know well what is needed, why it is needed, the way it will be used and the benefits of getting it in the most accurate possible way. Additionally, they should also have some management capabilities and be skilful at dealing and interacting with experts since such interaction will happen at different moments along the process. As in the case of analysts, the number of generalists depends upon the questions under study. For a study with the magnitude of this one, one generalist could be enough; nevertheless two generalists will actually participate in this application. One of them is a specialist in PA studies, familiar with both the development of PA models and with the probabilistic framework of the PA. The second one is an expert in geochemistry that could herself have participated as an expert, as in fact she has done in another similar study developed by NDA. This is not a frequent situation, but the project PAMINA has given us the possibility of counting on two generalists, which we think will improve the development of the whole process. The first one will be in charge of focusing on the questions under study from the point of view of the PA, while the second one will deal with information sources and with the definition of the questions from a scientific point of view.

The project team consists of the following persons:

- Analyst: Ricardo Bolado (IE, EC DG-JRC)
- Generalist 1: José Luis Cormenzana (ENRESA)
- Generalist 2: Lara Duro (AMPHOS XXI)

3.2.- Preparation of supporting material and definition of the questions to be studied

Once the team project has been set, analysts and generalists must define the questions to be evaluated by the experts. The organisation interested in performing the study does also play an important role in this step; it has to provide the list of questions to be studied. The starting point for any question to be solved is usually vague. It is necessary to arrive at a *complete definition* of the parameters whose uncertain has to be characterised. *Complete definition* of a parameter means the full definition of the parameter, the initial and boundary conditions to evaluate it and any other implicit hypothesis related to those conditions. The final definition should be very clear and accurate, no ambiguity should be allowed. It should have no problem when undergoing the clairvoyant test.

The complete definition of the question includes the way the experts should provide their answers. Since in PA studies uncertainty is characterised by means of Bayesian probabilities, experts should provide their assessments of uncertainty through this kind of probabilities. So experts should provide probability distributions, either discrete or continuous, depending on the nature of the parameters considered.

Two reports have to be delivered during this phase. The first one puts the questions to be addressed in the context where they make sense, in this case the PA of a generic Spanish radioactive HLW repository in granite. This report describes the system under analysis, the type of study that it is undergoing, and the rationale and works developed that justify the concern about the questions addressed in the EJ exercise. The second report provides the scientific framework for the study. It must provide a scientific description of the questions addressed, factors that must be taken into account and a list with all relevant sources of information (scientific papers in journals and conferences, technical reports, databases, experimental results, etc.). It contains the actual definition of the questions whose uncertainty will be quantified. Potential decompositions of the parameters could be done if deemed necessary. The list of references must show the actual state of knowledge in that area, but independence and reliability of the sources should always be kept in mind. Both reports are supporting material to help experts getting in touch with the problem and will be used along the whole process. Both reports, together with this document, are sent to each expert after step 3 and well in advance to step 4, so that experts have this information available before the first meeting. In this EJ application, generalists 1 and 2 write respectively the first and the second reports.

3.3.- Selection of experts

The objective of this phase is to select the most qualified experts to perform the assessment. Qualified Experts are those that:

1. Have the necessary knowledge and experience to perform the assessment,
2. are willing and available to participate in the assessment, and
3. do not have important motivational biases.

The first step to get the final list of experts is to start with a large list of potential experts. That first list is usually based on the opinion of the generalists and a thorough search in the scientific literature in that area. In some cases, the original list could be not very large due to the scarcity of experts in that specific area of knowledge. A first screening should be done according to the first bullet in the list above. The generic criteria considered in this protocol in order to check if experts have the necessary knowledge and experience are taken from Cooke and Goossens (2000):

- Reputation in the field of interest
- Experimental experience in the field of interest
- Number and quality of scientific and technical publications in the field of interest

- Diversity in background
- Awards received
- Balance of views

In particular, this EJ exercise requires the following specific knowledge from each expert:

- Deep knowledge about the general chemical behaviour of Radio, Tin, Selenium, Uranium and Plutonium.
- Deep knowledge about the possible solid phases that the above-mentioned elements may form.
- Good general understanding about the use of solubility limits in PA studies.
- Knowledge about all the chemical elements under study. All experts must be able to provide estimates for the solubility limits of the five elements considered (necessity to keep small the number of experts).

After performing the screening, a shorter list will be obtained, from which the final selection of experts will be done. In order to arrive at the final list, two ideas will be taken into account: three experts will be selected to participate in this EJ process (some theoretical studies about Bayesian combination of experts' estimates show that, even for moderate correlations among experts, not much benefit is obtained in terms of accuracy when increasing the number of experts; between three and five is the optimum number suggested by some authors), and the selected experts will have as much diversity as possible: different background (theoretical versus experimental, different areas of knowledge, etc.), proceed from different types of organisations (Academia, consultancy, research centres, engineering companies, etc.), etc. Additionally, we will try not to take experts working in other national waste management organisations in order to avoid motivational biases, though this will not be a definitive criterion to disqualify an expert. Finally, one more condition should be taken into account, due to the European wide character of the PAMINA project and the need of taking experts from different countries, being fluent in English is a mandatory requirement for all experts, which does also apply to generalists and analysts.

3.4.- Training sessions (first meeting)

The objective of this phase, as in other protocols, is to show experts normative aspects of EJ elicitation processes. The training sessions are held during the first meeting. The following sub-objectives are pursued, which have associated one training session each:

1. Motivate experts to provide rigorous assessments.
2. Remember basic concepts of Probability and Statistics.
3. Inform experts about basic issues related to cognitive biases.
4. Show experts the kind of statistical support they may expect from the project team.
5. Training in the assessment of Bayesian probabilities.

3.4.1.- Training session 1

In order to motivate experts, a first training session is organised where experts are given information oriented to pointing out the importance of the work they are going to do. Firstly, the project team explains to the experts the framework study where their opinions will be used, stressing the part of the study where their opinions are relevant. Secondly, the necessity of EJ is justified and the concept of *Lack of Knowledge Uncertainty* is introduced. The link between both is also explained. Thirdly, the project team explains that the key issue is not to predict a single value of each parameter under study, but characterising their uncertainty, showing the actual state of knowledge in that area. Finally, experts are informed that their opinions will not be used as a substitute of scientific research, but as a mean of knowing the actual estate of knowledge in the field of interest.

3.4.2.- Training session 2

Next, a session is devoted to the review of main concepts in probability theory, which usually includes some incursions in Statistics. The main topics reviewed in this session are

- Random experiments
- Dependent and independent events
- Probability axioms
- Interpretations of probability (classic, frequentistic, Bayesian, …)
- Total probability theorem
- Bayes' theorem as a tool to update information
- Random variables (continuous and discrete), probability density functions, probability mass function and cumulative probability distribution
- Different probability models of interest (Poisson, Bernouilli, binomial, geometric, negative binomial, uniform, normal, log-normal, exponential, Weibull, gamma, …) and main relations among them.

3.4.3.- Training session 3

The third training session is dedicated to inform experts about basic issues related to cognitive biases and the corresponding consequences on their assessments, and to help them avoiding such problems. Experts are informed about the difficulties that human beings in general and experts in particular encounter when they have to deal with different sources of information and they have to make statements in terms of probability. They are also informed about the main simplifying strategies (heuristics) they usually adopt when they have to make judgements, the problems they find when they have to deal with statistical information and the main biases that this introduces in their judgements. Special attention is dedicated to the most widespread and feared bias: overconfidence. This training session is illustrated with many examples taken from the literature in the area of knowledge psychology, which helps a lot to understand the origin of biases. A calibration exercise is also made to show experts their own vulnerability to biases.

3.4.4.- Training session 4

The fourth training session is dedicated to show experts the kind of statistical support they may expect from the project team. Experts have a lot of knowledge in their own specific professional field, but they do not usually have such a wide experience in Statistics and probabilistic methods. There are two general cases when they could get a lot of help from the project team: when they have to analyse statistical data and when they use a computer code as a tool to get information and to generate their estimates. Sometimes experts have a lot of data but they need help to get some statistical conclusions. The analyst informs the experts about the support they may expect from him/her, as for example:

- Experimental data analysis (use of multivariate analysis techniques with different aims, e.g. checking homogeneity of different data sources, or use of regression techniques to find correlations between data corresponding to different parameters).
- Distribution fitting
- Theoretical support on probability
- Simulation of stochastic problems

Quite frequently experts use computer codes as an aid to make their estimates. In order to use them in a more efficient way, it is convenient to inform them about different methods to get efficiently information from their codes. Experts may expect support from analysts in the following areas:

- Monte Carlo techniques
    - Sampling (simple random sampling, stratified sampling, Latin Hypercube sampling, importance sampling, etc.)
    - Output uncertainty characterisation (use of different statistics to analyse the output variables: means, variances, kurtosis, order statistics, etc.).
    - Sensitivity analysis (use of sampling based techniques, variance based techniques and graphic techniques to identify most relevant inputs).
- Design of experiments (factorial designs, fractional factorial designs, saturated designs, composite designs, etc.).
- Design of computer experiments (maximum entropy and minimum mean squared-error designs).

The use of any of these three techniques depends upon how expensive is to run expert's code in terms of time. If the code is very fast Monte Carlo techniques are appropriate, design of computer experiments are appropriate for very time consuming codes, while design of experiments is adequate if the computer code is intermediate in terms of computational time demanded.

Experts are not expected to learn all this techniques in a short training session, but they are expected to learn the kind of help that they could expect from the project team The key point in this case is that, if experts know what kind of help they may get from the analysts, they will handle more efficiently the problem proposed. More detailed information about this training session is given in the annex. This training session may be omitted if all experts declare their intention of not using any computer code to support their opinions.

3.4.5.- Training session 5

The last training session is dedicated to train experts providing opinions in probabilistic terms. After attending the previous training sessions, experts will be ready to answer questions in a probabilistic framework, but they will need some practical training. Firstly they are taught some ways to decompose problems in order to make them more tractable: event trees and influence diagrams. The analysts show some examples about decomposing a complex problem into more simple ones. Later on experts are encouraged to provide some estimates for some daily issues in probabilistic terms.

The analyst and the first generalist are in charge of the first training session while the analyst is in charge of the last four training sessions.

3.5.- Refinement of the questions to be studied (first meeting)

This task is done through two interactive sessions celebrated during the first meeting where both, the project team and the experts, participate. The purpose of the first question refinement session is to explain to the experts, in a detailed way, the questions to be assessed and to make a schedule of the activities to be developed by each expert. All the work developed by the project team during the *preparation of supporting material and definition of the questions to be studied* phase (step 2 of the protocol) will be used in this step. The session starts with a presentation by the generalists (generalist 2 in this case) of the parameters to be assessed, including all relevant sources of information previously identified. Experts are encouraged to give their own view of the problem and of the definition of the parameters, pointing out, if needed, further information sources, computations to be made, possible changes in the definition of the questions, etc.

The second question refinement session deals with the possible ways to decompose each parameter. The analyst and the generalist may provide a seminal decomposition that can be discussed with the experts. The objective is to help the experts to develop their own decompositions. Decompositions could be quite different from one expert to another. Experts will have to assess uncertainties of variables in the lowest levels. The analysts will usually do its aggregation. This is the right point to stress the importance of learning about the propagation of uncertainty concept and to show them all the potential variety of tools that the analysts could provide them to pre-process and post-process probabilistic runs of computer codes or of the simple decomposition model developed by experts.

Decomposition, as a further step in the analysis of the problem, can trigger some additional discussion on implicit hypotheses, which could produce a new redefinition and refinement of the questions to be studied. The result of this session is a refined definition of the parameters under study, which must be fully agreed by the project team and all experts. It is extremely important to make sure that the project team and all experts agree about the definition and the interpretation of the different questions to be assessed, otherwise different experts would be solving different problems, which is not a desirable situation.

Protocol steps 4 and 5 are celebrated together during the first meeting, and the different sessions are celebrated according to the following agenda:

1. Training sessions 1 to 3.
2. First question refinement session.
3. Training sessions 4 and 5.
4. Second question refinement session.

3.6.- First individual work period

After setting the refined questions to be assessed, the experts have roughly three or four weeks to start analyzing the problem individually. They are allowed to use whatever means they consider necessary to address the questions under study. They may use any data, computer runs and any other source of information available to them. During this period they may contact the project team to get any support in the areas addressed in the training sessions.

3.7.- Presentation of individual approaches adopted by experts (second meeting)

At the end of the first individual work period, all experts are called to a meeting where each one has to explain to the project team and the rest of the experts the strategy followed to address the questions under study. This is a one-day meeting that starts with a short presentation given by one of the generalists to remind the questions under study. Generalist 2 will do this task. Afterwards, the floor is given to the experts. Each expert gives a presentation where he/she explains the hypotheses taken into account, the sources of information used, the decomposition of the problem considered, computer runs performed, and any other information important to understand his/her approach to the problem. The format of this session is the same as a scientific conference with a moderator (the analyst supported by both generalists).

Two ideas are behind this session. The first one is to get some cross-fertilization among experts. The second one is to uncover implicit hypotheses that did not come up during the refinement phase (protocol step 5). Getting information about how other experts tackle a given problem may be a source of information and ideas to his/her colleagues. This could induce some dependence or correlation among experts, but the benefits of a more clear understanding of the problem and the discussion among experts exceed the shortcomings that dependence could introduce. If deemed necessary by the project team and the experts, a final refinement of the questions can still be done. After this final refinement no further change is allowed in the questions to be assessed.

3.8.- Second individual work period

During a period of roughly one month experts develop their studies taking into account all the information collected along the whole process. As during the first individual work period each expert may use whatever mean he/she wishes and may contact the project team to get support. At the end of this phase each expert has to release a short report summarizing his/her final strategy to solve the problem and whatever information they consider important to understand the way they address the problem. Each expert will forward his/her report to the project team a few days before the formal elicitation sessions to help them preparing the sessions.

3.9.- Elicitation sessions (third meeting)

Elicitation sessions are celebrated on individual basis; each expert is elicited separately. It must be developed in a quiet room, without noise, and avoiding interruptions as much as possible. All project team members will attend each elicitation session. The analyst will coordinate the session and the generalists will help him providing information when requested.

Each session starts with some words from the analyst. He reminds the expert the purpose of the session and the format of the information he/she will be asked (pdfs, probabilities, quantiles, percentiles, etc.). Next the analyst invites the expert to make a short summary of his/her approach to the problem, including the decomposition used, if any. If the decomposition is different from the one expounded by the expert in step 7, the expert is asked to explain the new decomposition, otherwise the expert reminds the definition of the question to be assessed to make sure that both, the expert and the project team, are working on the same variable. If any discrepancy arises, it should be solved before going ahead. After checking the definition, the formal set of questions of the elicitation session starts.

The two main techniques used to assess distributions are the quantile technique and the interval technique. Nevertheless, experts are allowed to give their probability estimates in the format they prefer. If the expert prefers to provide pdfs or probability distribution functions, either by providing estimates of their parameters or by providing drawings, they are allowed to do it. The main criterion is to make sure that the expert feels comfortable when generating his/her estimates. In case they prefer to use indirect techniques to assess their probabilities, they will be allowed to, although this situation will be avoided as much as possible; otherwise the elicitation session could take too long. The project team will continuously check the consistency of expert’s rationale, additionally, from time to time, the analyst will also ask questions to check the consistency of expert’s probability estimates. Measures will be taken if the project team detects important biases (anchoring and adjustment, overconfidence, availability, etc.).

All the project team members will take notes of the session, recording all relevant information in addition to the actual estimates given by the expert. After the session the project team will meet to summarise the main findings of the session and expert’s estimates. One full day will be allocated to the elicitation of each expert. All the days allocated to expert elicitation sessions will be booked during the same week, in consecutive days. Moreover, each expert will be available since the day of his/her elicitation session until the day after the last elicitation session, just in case the project team needs to get in touch with him/her to ask some further clarification about his/her distributions, or if large disagreements among experts arise and a reconciliation session is needed.

3.10.- Analysis and aggregation of results (third meeting)

The analysis of results starts with the review of the report generated by each expert and the notes taken during the elicitation session. The first task is to review again that all the rationale and estimates provided by each expert are consistent. If some problem is detected, the project team gets in touch with the expert affected to discuss about it. Detection of inconsistencies is not expected during this phase since consistency is checked extensively during the elicitation session, but it must be done for sake of redundancy. This task is done after each elicitation session, the same day that the elicitation session takes place.

After eliciting the third expert if, as expected, no problem is detected, the distributions provided by the different experts are compared. The degree of overlap between the different distributions is checked. If the overlap is large, it means that experts agree and provide similar distributions. In that case we can proceed to aggregate the individual estimates in order to get a common distribution for each parameter under study.

Only mathematical aggregation methods are considered in this protocol, we don't consider group combination methods, except in the case of important disagreements among experts. In principle, the two methods considered are the linear pool with equal weights and Bayesian combination. The linear pool with equal weights is the most used method and it is simple and easy to understand. The Bayesian combination allows taking into account in the final estimate the opinion of the organization interested in getting the distributions. Most likely, the method used to aggregate results will be the linear pool with equal weights.

A problem arises if the distributions provided by the experts either do not overlap much or do not overlap at all. This case arises when experts have fundamental disagreements. In this case, a reconciliation session will be celebrated just the day after the last elicitation session. Just before the reconciliation session the project team will analyse again the rationale that supports the distributions provided by each expert. The aim of this task is to find the origin of the large disagreements between experts (different decompositions, different underlying assumptions, etc.). Then the reconciliation session will start. The target of this session is to get a group consensus distribution. The session would be organized as a nominal group in order to avoid group biases. Firstly, the project team shows to the experts the analysis done about their distributions and the origin, in the opinion of the project team, of the disagreement. Next, a debate starts where the experts themselves analyse the origin of such disagreement. The analyst, with the help of the generalists, moderates the debate. Finally, if the project team sees a real convergence of the experts towards a common position, a consensus distribution will be elicited. Then the set of experts will be asked as if they were just one expert. If such a convergence does not happen, the distributions that most overlap will be mathematically aggregated. The other distribution will be kept as an alternative distribution for sensitivity analysis.

3.11.- Review

At the end of the process, after getting all individual and aggregated distribution, the project team will generate the report that summarises all the activities developed and results obtained during the process (next step, documentation). This report will be sent to each expert for review. The objective of this phase is to check that experts agree with the way their individual opinions are described in the report.

3.12.- Documentation

As in any other well-designed protocol, documentation will be as complete as possible, including results and description of the ways to obtain them. The contents of the documentation will follow the order of application of the procedure, recording, in each step, *what* has been done, *why* it has been done, *how* it has been done and *who* has done it. It ought to be always completely clear to the reader what is a result assessed by an expert and what results are the outcome of an aggregation, sensitivity analysis or any other analysis not provided explicitly by an expert.

4.- Preliminary schedule of activities

1. June 20th: Start the process to select experts.
2. July 11th: Experts are appointed.
3. July 18th: Written material to be distributed to experts is sent (System description, supporting scientific report, questions posed for assessment and protocol description – this report).
4. Week 38 (September 15th – 19th), two days to be chosen (most likely 15th and 16th): First meeting with experts. Celebration of training sessions and refinement of the questions under study.
5. Week 42 (October 13th – 17th), one day to be chosen: End of the first individual working period, celebration of the session to describe individual approaches.
6. Week 47 (December 1st – 5th), four days to be chosen (most likely 2nd to 5th): Elicitation sessions, and reconciliation session if needed.
7. December 19th: Edition of the final report draft. It is sent to experts for review.
8. January 30th: Experts send back their comments, if any.
9. February 27th: Final report.

References

R. Bolado and E. Gallego (2000). '*El juicio de expertos y su aplicacion a cuestiones de seguridad*'. Fundación para el fomento de la innovación industrial.

R. Bolado and J. Ibáñez (1999). Application of the CTN-UPM expert judgment methodology to an hydrogen DDT issue in a generic evolutionary PWR: Phase 2 of the benchmark exercise on expert judgment techniques in PSA level 2. In: *Safety and Reliability. Vol. 2*. Schuëller, G.I. & Kafka, P. (Eds.). Balkema. 1229-1234.

R. Bolado and R. Iglesias (2001). '*Acuerdo CSN/CTN-UPM sobre juicio de expertos (recombinadores)*'. Technical report of the Nuclear Technology Unit of UPM CTN-09/01.

R. Bolado, R. Iglesias, M.A. Jiménez, J.M. Martín-Valdepeñas, A. Artigao, A. Lantarón, E. Gallego(2002). 'Aplicación de técnicas de juicio de expertos al studio de la necesidad de instalar recombinadores autocatalíticos pasivos en la contención de un reactor nuclear español tipo KWU'. In *Ponencias de la 28[a] reunión anual de la Sociedad Nuclear Española*. SNE (Ed.).

R. Bolado (2005). Acillary tools to help experts providing opinions. In: *Proceeding of the workshop on the Use of Expert Judgment in Decision-making*. Devictor, N. & Bolado, R. (Eds.). EUR 21873 EN.

R. Bolado and A. Badea (2008). Review of expert judgement methods for assigning pdfs. PAMINA report M.2.2.A.3.

E.J. Bonano, S.C. Hora, R.L. Keeney and D. von Winterfeldt (1990). Elicitation and use of expert judgement in performance assessment for high-level radioactive waste repositories. Sandia National Laboratories, SAND89-1821 (NUREG/CR-5411).

G. Cojazzi, D. Fogli, G. Grassini, P. De Gelder, D. Gryffroy, R. Bolado, E. Hofer, R. Virolainen, I.M. Coe, A. Bassanelli, J. Puga, I. Papazoglou, O. Zuchuat, E. Cazzoli, J. Eyink, G. Guida, L. Pinola, U. Pulkkinen, K. Simola, D. von Winterfeldt, A. Valeri (2001). 'Benchmark exercise on expert judgment techniques in PSA level 2'. *Nuc. Eng. & Des*, vol. 209, issue 1-3, 211-221.

R.M. Cooke and L.J.H. Goossens (2000). Procedures guide for structured expert judgement. EUR 18820 EN.

E.D. Gorham-Bergeron et al. (1991). Evaluation of severe accident risks: Methodology for the accident progression, source term, consequence, risk integration, and uncertainty analyses. Sandia National Laboratories, NUREG/CR-4551, Vol. 1, draft revision 1, SAND86-1309.

J.P. Kotra, M.P. Lee, N.A. Eisenberg and A.R. de Wispelare (1996). Branch technical position on the use of expert elicitation in the High-Level Radioactive Waste program. USNRC, NUREG-1563.

K. Simola, A. Mengolini, L. Gandossi and R. Bolado (2005). Application of formal expert judgment to the evaluation of structural reliability. In: *Proceeding of the workshop on the Use of Expert Judgment in Decision-making*. R. Bolado and N. Devictor (Eds.). EUR 21873 EN.

Annex: Training session about Probability and Statistics support and computer code statistical support (training session 4 in the first meeting)

Well-known and well-designed protocols like the SNL/NUREG-1150 protocol include training sessions that inform experts about the biases that can affect them, remind them basic concepts of probability and train them to start providing answers in probabilistic terms. The material used in these training sessions is widely known and may be found elsewhere, see for example Simola et al. (2005). This annex provides some extended information about the main ideas that the project team wants to communicate to experts about the areas where experts may expect some support from the project team. The project team may help experts to analyse more efficiently statistical data available to them and results of computer codes used by experts to make their assessments.

The first idea is to provide the experts information about how to run their computer codes in the most efficient way and to show them that they will have full support from the project team in any issue related to probability and statistics. The training about efficient use of computer codes is implemented through a normal training session where the main concepts about this subject are reviewed. The objective of this session is not to 'teach' them this specific matter, but to show them the set of available techniques, the conditions under which they could or should be used, and the availability of the project team to provide them any support in this direction. Regarding the support in general probability and statistics issues, the target is to show them the kind of support they could get. Both ideas are developed in the following sections.

A.1.- Probability and Statistics Support

The normal training session on probability is just a review of important concepts; no great improvement in probability skills should be expected from experts after such training. It is not reasonable to expect they keep in mind, for example, all the relations between Poisson, binomial, geometric, exponential, negative binomial, uniform and gamma distributions that appear in a Poisson process in one dimension. What is however reasonable is that, after the training, they become able to ask the project team members for specific help when addressing some probability (or statistics) issue that arises in their work.

It is difficult to summarise all the tasks in which the project team may provide help to the experts, due to the great number of different questions that could arise. The following one is a non-exhaustive list:

- Experimental data analysis (use of multivariate analysis techniques with different aims, e.g. checking homogeneity of different data sources, or use of regression techniques to find correlations between data corresponding to different parameters).
- Distribution fitting.

- Testing hypotheses about different data sets.
- Modelling problems via event trees or using influence diagrams.
- Theoretical support on probability.
- Simulation of stochastic problems.

The key idea is that, if the expert knows what kind of help he/she may expect from the analysts, he/she will handle more efficiently the problem proposed.

A.2.- Session on Computer Code Statistical Support

This session is oriented to provide general information to the experts about the kind of support they may expect from the project team regarding the efficient use of their computer codes. Certainly, the kind of support depends essentially on the speed of the computer code used by the expert. Conditional on the number of runs the expert will be able to get during the study period, she/he will find useful different tools and will get different types of information. In some cases, the number of input parameters considered in the computer code could also be relevant regarding tools to be used and expected benefits.

In principle, the tools and techniques to be proposed and shown to the experts may be divided into three types:

- Tools to be used when fast or relatively fast computer codes are used (from several tens of runs obtainable during the study period, say above 50 runs).
- Tools to be used when the computer codes available are neither fast nor slow (a few tens of runs may be obtained, say between 20 and 50),
- Tools to be used when the computer codes available are slow (just a few runs are reasonably obtainable, say less than 20).

According to this classification (the numbers shown in the previous list are given as an indication, they should not be considered as absolute limiting values), the session should be divided into three parts, one dedicated to each case. Explicitly, in an almost one to one relation with the previous list, the following approaches could be addressed in the three consecutive parts of the session:

- Monte Carlo
- Design of experiments
- Design of computer experiments/design of experiments

A description of the methods to be shown in the three parts of the session is given as follows.

A.2.1.- Training on Monte Carlo techniques

When the expert is using a fast running code, the Monte Carlo approach is the best option to get a lot of information about the model itself and the results it provides. One of the most attractive characteristics of this approach for experts is the fact that, in many cases, the Monte Carlo simulations will provide in a straightforward way the answers to the questions asked by the project team in the right probabilistic framework. In other cases some further work will be needed. A full Monte Carlo analysis may be divided into five steps:

1. Identification of relevant input parameters and characterisation of their uncertainty/likely range.
2. Sampling.
3. Uncertainty propagation (running the code as many times as needed and possible).
4. Characterisation of the uncertainty of the outputs and interpretation of results.
5. Sensitivity analysis.

In addition to a short conceptual introduction to the method, some notions should be introduced regarding steps 1,2,4 and 5 in the previous list.

A.2.1.1.- Identification of relevant input parameters and characterisation of their uncertainty/likely ranges.

Experts know that the uncertainty they have about the questions brought for their analysis has two sources: 1) The uncertainty in the inputs and 2) the way that uncertainty is propagated by the model/computer code. Part 2 is done automatically by Monte Carlo method, but the expert must estimate the uncertainty in the inputs. Experts' attention is drawn in this case on the classical training sessions. Nevertheless, analysts should tell to the experts that the characterisation of uncertainty is not so crucial in this step. Three ideas should be taken into account:

- Full distributions are needed in this step, not just percentile estimates (if the expert is able to provide just percentile estimates, the project team should fit cumulative distribution functions to the points provided by experts).
- What is really important is to determine very well the ranges of variation
- In case of doubt, ranges should be widened in order not to exclude any possible combination of inputs. Independence between input parameters should also prevail in case of doubt because of the same reason.

A.2.1.2.- Sampling

Analysts present main notions to experts about different ways to get a sample from a set of input distributions, including some variance reduction techniques to get more accurate estimates. In principle three techniques should be shown to the experts:

- Random sampling.

- Latin Hipercube sampling (LHS).
- Stratified sampling.

Additionally, the concept of dimension reduction in propagation of uncertainties could also be introduced; it could be useful for some experts.

Some notions should also be introduced regarding sample size selection. It is interesting to introduce the concept of tolerance interval (one sided and two sided). Conditional on the speed of the computer code, this concept will be of help to select the sample size and to know what can be expected from the sample in terms of coverage of the output variable domain.

A.2.1.3.- Characterisation of the uncertainty of the outputs and interpretation of results

After running their codes, experts will get samples from the output variables, and usually they will not know how to handle these samples to extract as much information as possible. Usually they will be able to compute the most common statistics such as the mean, the variance and the standard deviation, but they seldom know how to compute other statistics and to interpret then. Additionally, some graphics could also help to understand more precisely what is the behaviour of their outputs. In this step of the training session, the main statistics used could be introduced and their meaning could be explained. Those that deserve more attention are:

- Mean.
- Variance and standard deviation.
- Skewness coefficient.
- Kurtosis.
- Order statistics and their confidence intervals, especially the most relevant ones like 1%, 5%, 10%, 25%, median (50%), 75%, 90%, 95% and 99%.
- Histogram.
- Box plots.
- Empirical (cumulative) distribution function with Kolmogorov confidence bands (and its complementary curve).

Among the items in this list, it is important to stress the importance of order statistics and their confidence intervals. Order statistics are asymptotically unbiased distribution-free statistics (independent of the distribution) of the corresponding variable quantile (David and Nagaraja, 2003), whose exact confidence intervals may be easily computed from the sample. This means that, directly from the sample obtained by experts through simulation, they may get an estimate for any quantile and a confidence interval (with the limitations imposed by the sample size). All the most useful details related to the interpretation of the results should be explained to the experts. Information regarding the difficulties to provide confidence intervals for means and variances under non-normal distributions should also be shown explicitly to the experts.

The empirical cumulative distribution obtained from a sample is a graphic estimator of the whole cumulative distribution function. The Kolmogorov band available for these plots also provides very interesting information that could be much appreciated by experts.

A.2.1.4.- Sensitivity analysis

Experts should be informed in the session about all the information they could get from an adequate sensitivity analysis from the sample. The main techniques to be introduced to experts are

- Regression based techniques (Partial Correlation Coefficients, Standardised Regression Coefficients and their corresponding rank based versions).
- Techniques based on non-parametric statistics, for example the Mann-Withney statistic or the Kruskal-Wallis statistics.
- Variance based statistics (mainly correlation ratios, since other more powerful techniques like FAST or Sobol's sensitivity indices are applicable only to very fast computer codes).
- Distribution sensitivity techniques.
- Graphical techniques (mainly scatterplots and Cobweb plots).

The first set of techniques is suitable to detect linear and monotonic relations between inputs and outputs. The second set of techniques is suitable to detect specific relations between specific regions of input and output variables, as for example a relation between the 10 % largest values of one variable and the 20% of intermediate values of another variable. Cobweb plots (Cooke and van Noortwijk, (2000)) provide very good graphic complementary support to the results obtained with non-parametric statistics. Variance based techniques are model independent techniques that allow to compute what fraction of the variance of a given output variable is due to the uncertainty in any input variable.

Distribution sensitivity techniques are extremely useful since they allow estimating the impact of changing the distributions of the input parameters on the output distributions without running again the code. This is why a very accurate assessment of the input distributions (see section A.2.1.1) is not really necessary. Using these techniques, the impact of changing slightly the input distributions may be assessed very fast.

All these techniques may provide very useful pieces of information to experts, which could be used to modify their assessments based only on the output uncertainty characterisation.

A.2.2.- Training on design of experiments

Design of experiments is a branch of Statistics dedicated to study what input parameters have a significant impact on the results of a process and what parameters do not have it. Though originally developed to study industrial processes affected by random perturbations, most of the ideas developed in this field may be applied to computer code simulations.

When thinking about the applicability of these techniques to help experts getting information from their computer codes, three main areas come up: 1) Exploratory runs to get some idea about the output values obtainable conditional on the support of the input parameters, 2) screening of input parameters (identifying important and non important parameters) and 3) model building (response surfaces). The first two cases are suitable when the number of runs the experts may get during the study period is small, and both objectives may be achieved using the same strategies (samples), so that experts may get information about the possible range of output values and, at the same time, discard non relevant input parameters. Both pieces of information could be of use when providing their final opinions to the project team. The third case demands more computing capabilities, but provides more detailed and valuable information about the model. In fact, response surface techniques (Box and Draper, (1998)), which are a mixture of design of experiments concepts and regression analysis techniques, are used to create polynomial meta-models that capture the essential features of the models studied. The meta-models could be used to replace the real codes in further extremely cheap Monte Carlo analyses. They could also be used as information sources to apply some variance reduction techniques, i.e. control variates, in a Monte Carlo analysis.

The main concepts to introduce to the experts in the training session are the following:

- Factorial designs (2^k and 3^k).
- Fractions (2^{k-p} and others).
- Resolution of a design and alias matrix.
- Saturated designs.
- Composite designs.

Again, the leading idea is to make understandable to the experts the meaning of these concepts. Experts could need weeks, even months, to fully understand all the ideas behind these bullets and to apply them by themselves, but in less that half an hour they may understand main concepts, what the benefits are of using these techniques and what help they may expect from the project team. The project team should be able to provide, after some discussion with the experts, the design values in the input space and to post-process the results provided by the experts. Depending on the approach adopted (screening, exploratory analysis or response surface), experts will get different types of information as a result of the post-process done by the project team. Post-processing includes interpretation of the results and potential uses.

A.2.3.- Training on design of computer experiments

Design of computer experiments is a new branch of Statistics started approximately fifteen years ago (Currin et al. (1991), Santner et al. (2003)). The main innovation introduced by this new branch is the assumption of smoothness of the response. The results provided by the model are considered just a realisation of the random field, and the main objective is to estimate the parameters of the random field. As soon as the parameters of the random field are estimated, the random field model may be used to predict the output of the computer code.

In principle, there is no restriction on when computer design of experiments could be used. Nevertheless, using these techniques when the computer code is very fast is not very convenient since, in that case, the Monte Carlo approach is very efficient and, when combined with sensitivity techniques, provides much information. When the computer code is neither very fast nor very slow, the information provided about the model is probably not as easily understandable as the information provided by response surfaces. The reason for this is the non-straightforward assumptions of these models, the concept of random field on which they rely, in addition to correlation structures and other related concepts, are usually far beyond the common basic probability knowledge expected from an expert in another area of knowledge. The main advantage comes up when the model is very time consuming. In that case, the sampling strategies inherent to these techniques could be very efficient in an exploratory analysis.
The main concepts to be introduced to experts are just related to the different strategies to sample in the input space. Though other strategies are available (based on number theory and grids), the most useful and best-supported (Bayesian) strategies to create different designs are:

- Maximum entropy designs
- Minimum mean squared-error designs

The main idea to show to experts is related to the differences between both strategies to allocate sampling points. The first strategy is based on maximising the expected posterior entropy, which is equivalent to maximising the prior entropy. This strategy favours the edges of the input space instead of the interior. The second strategy tries to minimise the mean squared error of the predictions, favouring interior points instead of points on the edges of the input space.

References

G.E.P. Box and N.R. Draper (1987). Empirical model-building and response surfaces. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons.

C. Currin, T. Mitchell, M. Morris and D. Ylvisaker (1991). *Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments*. JASA, vol. 86, No 416, 953-963.

H.A. David and H.N. Nagaraja (2003). Order statistics. Third edition. Wiley Series in Probability and Statistics. John Wiley & Sons.

R.M. Cooke and J.M. van Noortwijk (2000). Graphical methods. In: Sensitivity analysis. A. Saltelli, K. Chan and E.M. Scott (Eds.). Wiley Series in Probability and Statistics. John Wiley & Sons.

T.J. Santner, B.J. Williams and W.I. Notz, (2003). The design and analysis of computer experiments. Springer Series in Statistics. Springer.

K. Simola, A. Mengolini, R. Bolado and L. Gandosi (2005). Training material for formal expert judgement. JRC's Technical Report EUR 21770 EN.