The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific
and Cultural Organization

International Atomic
Energy Agency

H4.SMR/1645-3

"2nd Workshop on Earthquake Engineering for Nuclear Facilities: Uncertainties in Seismic Hazard"

14 - 25 February 2005

Expert Judgement Mehtods

J.-U. Klugel

Kernkraftwerk Goesgen-Daeniken
Daeniken, Switzerland

Expert Judgement Methods

J.-U. Klügel
Kernkraftwerk Goesgen-Daeniken, Kraftwerkstrasse, 4658 Daeniken, Switzerland

Second IAEA-Workshop on Earthquake Engineering for Nuclear Facilities, 14-25 February 2005, Miramare-Trieste, Italy.

ABSTRACT: The lecture presents an general overview on expert judgement methods and procedural guidelines for structured expert judgement. The main focus is given to the incorporation of expert judgement as a scientific source of information into risk-informed applications and seismic hazard analysis. The lecture is based on a review of the state of the art on the use of expert judgement in PRA applications with special emphasis to probabilistic seismic hazard analysis. Key problems of the practical use of expert opinion are identified and addressed by suggesting alternate methods. This lecture is inspired by the experience of a still ongoing review of a large scale seismic hazard analysis performed in Switzerland, which was based on the SSHAC-procedures (SSHAC, 1997) at its most elaborate level – level 4.

1 INTRODUCTION

The use of expert knowledge as well as of expert judgement is an essential part of any practical decision making. This trivial statement is based on some own management experience as well as on classical management theory, showing that most decisions in social, financial and economical life are based on incomplete information and are to a large extend judgmental, heavily based on subjective perception of the final decision makers. The formalized use of expert knowledge – meanwhile called elicitation of expert judgement – is just one of the possible means (besides all other types of scientific, engineering or financial investigations including analysis of the related data (f. e. performance indicators) like in the "management by numbers or management by objectives" approach emphasized) to close such information gaps. The intention of the use of experts by decision makers is

- to incorporate the best available knowledge on the topic into the decision making process
- to provide a more robust basis for the final decision making.

This intention of decision makers shall not be lost from mind, while discussing different formalized methods of expert judgement elicitation. Therefore the results of different methods shall always be judged by compliance with these goals. Due to the complexity of the elicitation process formalized methods are only used for decisions of great significance in terms of political, economical, security or safety consequences and mainly in a context of lack of other sources of information or by other words in conjunction with large uncertainties in the assessment of the topic of interest. The development of simple and transparent methods is therefore a challenge to allow for a more frequent use of expert elicitation processes in practical decision making. Transparency of methodology includes the need of dissemination of information on the limitations of expert judgement to the decision makers. Illusions shall be avoided, that the elicitation of expert opinion is an all healing medicine which allows to avoid other types of information gathering while preparing decisions of large significance. Experts are just human beings and thus subject to cognitive limitations and psychological bias. Their judgements shall be checked and validated and the analyst using the results of expert judgement shall be aware of the potential limitations of the process.

This lecture will provide some general procedural guidance for structured expert judgement and its incorporation into risk –informed applications, with some special emphasis methods to seismic hazard analysis. This lecture is inspired by the experience of a still ongoing review of a large scale seismic hazard analysis performed in Switzerland (the PEGASOS-project, PEGASOS 2004) which was based on the SSHAC-procedures (SSHAC, 1997) at its most elaborate state – level 4 - and which failed to provide results meaningful for decision making requiring a correction of the intermediate study results. It is based largely on some recent work on the use of expert judgement for accident consequence modeling in a PRA framework (Cooke & Goossens, 2000, Bixler et al, 2004, M. McKay & M. Meyer, 2000).

2 CLASSIFICATION AND DEVELOPMENT OF EXPERT JUDGEMENT METHODS

There is a wide variety of expert judgement techniques with different purpose in use in technical applications. The main properties by which the methods can be classified are:

- Purpose of the application
- Degree of formalization
- Used mathematical methods

Expert judgement can be used both to make qualitative as well as quantitative assessments. Qualitative assessments are typically in use when gathering of information is the main purpose of the expert judgement process. In many cases such assessments are accompanied by some categorization of information (different qualities) to rank them, what can be regarded as a first quantification attempt. This paper focuses on methods, which aim on a quantification of properties important for decision making on technical issues.

2.1 *Purpose of application*

Expert judgement is (consciously or not) used in many technical and socio-political fields because it is part of any problem solving technique. Even if sufficient data on a topic of interest is available, expert knowledge is required to interpret this information. A basic differentiation with respect to the purpose of systematic expert judgement can be summarized as follows:

- ***Interpretative application.*** Experts are used to explain empirical information, which is outside the expectations of commonly used theories or beliefs.
- ***Forecasting applications.*** Experts are used to make predictions f. e. in technology or political developments to support decision making in investments (in research and development) or to develop political strategies.
- ***Phenomenological applications.*** Experts are used to describe (quantitatively) the consequences of phenomena (of interest in science or in socio-political life), which are expected to be possible, but cannot investigated directly.
- ***Informative applications.*** Experts are used to complete available, but insufficient information on a topic of interest. Expert opinion is used as an additional complementary source of information.

The results of the expert judgement to some extend represent only an intermediate state of knowledge helping to rank problems, which shall be investigated at a later state to reduce uncertainties.

Expert knowledge cannot be equaled to expert opinion, although the later shall reflect the first and can be regarded as an external (sometimes diffuse, imprecise) expression of expert knowledge.

2.2 *Degree of formalization*

A principal dividing line between different expert judgement methods can be drawn with respect to the degree of formalization of the methods. The simplest way of eliciting expert opinion is the unstructured approach, as it is common practice in brainstorming. Here the purpose of expert elicitation is just the collection of ideas, which at a later stage will be systemized and ranked by a facilitator or a monitoring team.

A similarly simple structured approach consists in direct questioning of experts by the help of questionnaires which later will be processed statistically.

More structured approaches include a systematic approach to expert selection, expert elicitation, the inclusion of normative elements to weigh expert opinion, the use of expert panels and formal consistent mathematical procedures to process the results.

A categorization can also be made with respect to the way how the expert elicitation process is performed – directly, asking straight forward for the parameters and their values of interest or indirectly using intermediate parameters which can be processed using an adequate mathematical model.

2.3 *Used mathematical methods*

Expert judgements methods can be distinguished by the degree of using mathematical methods to process results. This degree can reach from simple intuitive empirical methods to the formalized use of mathematical decision theory or fuzzy set mathematics. The degree of mathematical formalization depends on the purpose of the application. There is a lot of experience available in nuclear technology (human reliability analysis, dependent failure analysis, maintenance frequencies, beyond design basis threat analysis in security issues) demonstrating that reasonable results can be achieved by using simple methods if some basic rules were followed. The most important rules to be complied with are related to the selection of experts (they shall be truly substantive experts in the field) and to the requirement of assuring empirical control of the results setting quantitative constraints to possible extreme assessment results. Another key requirement is the availability of a normative expert, who is able to interpret the substantive experts belief in a consistent formal mathematical sense. It is interesting to note, that the currently in use for PSHA SSHAC-Procedures

(SSHAC, 1997) do not include this important requirement.

2.4 *Development of Structured Expert Judgement Methods*

Structured (formalized) expert judgment procedures have been developed historically in some military applications (the oracle type Delphi-method) and applied in think tanks for scenario-based forecasting (scenario analysis). In western countries the emergence of methods for a systematic use of expert opinion is related to the foundation of the RAND-company (started as a joint project of the U.S. airforce and Douglas aircraft) in the USA. These developments were strongly supported by the US government. Before the Vietnam war about 11'000 independent think tanks worked at charge of the US government (Cooke, 1991). The scenario-based methods can be characterized by the words of some of the founders of the method: "Scenarios are hypothetical sequences of events constructed for the purpose of focusing attention on causal processes and decision-points. They answer two kinds of questions: (1) Precisely how might some hypothetical situation come about, step by step? And (2) What alternatives exist, for each actor, at each step, for preventing, diverting, or facilitating the process." (Kahn & Wiener, 1967, p.6).

The similarity to techniques applied today in probabilistic risk analysis (event trees) or in Probabilistic Seismic Hazard Analysis is obvious. The difference consists just in the assigning of quantitative properties at each "decision" branch of the developed scenario and in the trial to assure completeness of the developed "scenarios". Attention shall be paid to the fact, that the scenario based approach was constructed as a systematic think process and by the trial to imitate the political thinking process. So the weighting factors (probabilities) for the decision points were defined on a political consensus to reflect the opinion of different stakeholders – institutes and organisations – participating in the scenario development. It was essentially a deviation from the formal mathematical decision theory due to the problem to incorporate different opinions into the decision process, what deemed not possible at that time. This can be reflected by the following statement: " The subjective curve of probabilities often seems flat …In order to avoid the dilemma of Buridan's ass, who starved midway between two bales of hay because he could not decide which one he preferred, we must then make arbitrary choices among almost equally interesting, important, or plausible possibilities. That is, if we are to explore any predictions at all, we must to some extent 'make them up'." (Kahn & Wiener,1967, p.8).

It is very interesting to observe that modern expert elicitation processes as formalized in the SSHAC-procedures for seismic hazard analysis lead to results, which are **surprisingly similar to the mentioned Buridan's ass problem** because the obtained subjective probability curves are indeed very flat and under certain circumstances (truncation at a very high number of standard deviations in the attenuation model) do not converge at all. This can be illustrated by figure 1 showing the seismic large early release frequency (LERF) of the Goesgen Nuclear Power Plant (model basis of 2001, Klügel et al, 2004) based on seismic hazard curves which according to proponents of the SSHAC procedures do not "reflect epistemic uncertainties" (indeed in the model only the dispersion of ground motion data was considered resulting in a total standard deviation value of 0.67). If "epistemic uncertainties" will be added – the curve would get even worse.

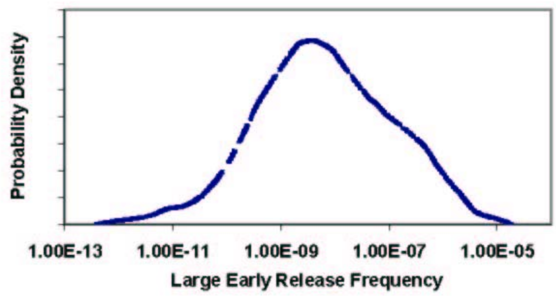


Figure 1 Probability density distribution for the Seismic Large Early Release Frequency – incomplete consideration of "epistemic uncertainties"

It is obvious, that practical decision making is almost impossible under these conditions. The seismic PRA could be replaced by a direct expert based assessment focussing on the upper and lower limit of the large early release frequency and the assumption of a uniform distribution between these boundaries. This would substantially reduce the effort for performing a seismic PRA.

The alternatively used methodology – the Delphi method – was based on systematic questioning of invited, carefully selected, experts, which did not know each other. The results of the questioning was processed statistically to obtain the median values and the interquartile ranges of the responses. In a next iteration the experts were confronted with the results and asked, whether they would modify their responses. Experts giving responses outside of the range between the 25% and the 75% quartiles were asked to justify their responses. These responses were analysed, the statistics were corrected and a next iteration was started. Typically 3 or 4 iterations were performed and the increasing degree of convergence of the experts' results were interpreted as the achievement of a consensus (political consensus!).

It is worth to mention that both methodologies – the scenario-based analysis and the Delphi-method were proven to be substantially wrong in several cases (Cooke, 1991).

Later improvements of the expert judgement methodology involved the systematic use of expert panels and the development of systematic mathematical aggregation methods, which actually means a return to or a combination with elements of mathematical decision theory. Most applications in technical disciplines are based on the scenario analysis approach using expert opinion to develop the "probabilities" at different branches of the scenario sequence event tree. In the nuclear field this development was largely stimulated by the US NRC risk study NUREG-1150 and the subsequent detailed plant-specific analysis. Expert judgement was used extensively to address many problems of severe accident analysis not yet investigated at that time or not possible to be investigated with reasonable efforts at that time (for many cases this is valid even today). It is interesting to note, that experts were selected to ensure a balance of viewpoints (NUREG/CR-4551, Vol. 2, p.3.8) what can be interpreted that the goal of the analysis consisted in achieving a consensus between different organisations. Many basic elements of the procedural guide described in Cooke & Gossens (2000) and presented here in an expanded version were developed in these projects. It is worth to mention, that these elements deviate to a large extend from the approach proposed in the currently in use SSHAC-procedures (SSHAC, 1997). Problems observed with the application of the SSHAC-procedures in seismic hazard analysis can be partially related to these deviations from validated rules of constructing a rational consensus.

3 GENERAL PROCEDURAL GUIDELINES

3.1 *Uncertainty, expert judgement and technical decision making*

The main purpose why the elicitation of expert judgement became a familiar method in PRA and in probabilistic seismic hazard analysis is the need to deal with uncertainties in available information or in their interpretation. A good definition of uncertainty in context of practical decision making is given by Cooke & Goossens (2000):

"Uncertainty is that which is removed by becoming certain. ….certainty is achieved through observation, and uncertainty is that, what is removed by observation. Hence uncertainty is concerned with the results of possible observations."

To be studied quantitatively, uncertainty must be expressed by a mathematical representation. The most common representation is the use of probability. One principal alternative is the use of fuzzy numbers. Both, probability in its subjective interpretation, as well as fuzzy numbers, are able to express uncertainty as a degree of belief of one person. The degree of belief of persons can be measured indirectly by their choice behavior. This degree of belief is different for different persons and there is no rational mechanism for persuading individuals to adopt the same degree of belief of others.

In practice the degree of belief of single persons as well as of large collectives can differ significantly from reality even then a social or political consensus is achieved among all members of the collective. The fact that many people believe the same is no indication for the truth of the statement in what these people believe.

Technical decision making in the context of a seismic hazard analysis for critical infrastructures shall as far as possible rely on realistic results, on one side to assure the safety of public and environment and on the other side to avoid unnecessary costs to be competitive in the market.

This practical need leads to the consequence, that expert judgement in technical decision making can only be used, if the results of this judgement at least in principle can or could be compared to observations (***validation requirement***) and their judgement will be corrected in case of new evidence (Bayesian approach in theory of probability).

The situation may be different for political decision making, there it is more important, what people believe instead of whether their belief is true.

On the other side it is getting obvious that the intended use of experts as another reliable source of scientific information requires a structured and formalized procedure based on rational consensus principles.

3.2 *Principles for Rational Consensus*

The following principles of building rational consensus were derived by Cooke (Cooke, 1991). Although they sound simple, there is much evidence, that they are not followed rigorously in practical applications.

Principle 1 Reproducibility:

" It must be possible for scientific peers to review and if necessary to reproduce all calculations. This entails, that the calculational models must be fully specified and the ingredient data must be made available."

Reproducibility is an essential element of scientific analysis and has to be regarded as an important quality property of the analysis. In practice this requirement is frequently violated, sometimes for proprietary reasons more frequently because the experts involved simply do not think about the information needs of a potential reviewer. In some cases reviewers got the impression, that studies were set up in a way to exclude the possibility to compare the new results with earlier studies. This was the case for the Swiss PEGASOS project (PEGASOS, 2004), there attenuation laws were developed for mixed soil conditions and later scaled to generic rock using a reference shear wave velocity different from the one used in a comparable study of the Swiss Seismological Service or in similar studies in the US.

Principle 2 Accountability:

"The source of expert subjective probabilities must be identified".

Accountability means in the context of using expert opinion, that the decision maker can trace every subjective probability to the name of the person or institution from which it comes. In cases of public decision making, this information must be made public. In case of a seismic hazard analysis study or any other study which potentially may have very large consequences in a financial or even macroeconomic sense, accountability shall be expanded to legal responsibility. Experts shall be able and ready to defend their positions in court according the national legal system.

Principle 3 Empirical Control:

"Expert probability assessments must in principle be susceptible to empirical control".

Empirical control ensures that the use of subjective probabilities cannot be construed as a license for the expert to say anything whatever. Empirical control is a cornerstone of scientific methodology. More formerly as will be seen below a formal validation process is suggested for the approval of the results of studies incorporating expert judgement.

The principle of empirical control is essential to constrain expert judgements and the resulting probability distributions for the evaluated parameters by observations from the real world.

Principle 4 Neutrality:

" The method for combining/evaluating expert opinion should encourage experts to state their true opinion"

According to Cooke (Cooke, 1991) most methods for forming weighted combinations of expert probability assessments must be criticized of not meeting this requirement. In seismic hazard analysis a typical example is the Lawrence Livermore Study (Bernreuter et al, 1984) a predecessor of the SSHAC-procedures. Here the experts were asked to weight themselves at each question at hand. A high rating is obviously a form of reward – self weights are a means of punishment or reward of the experts themselves and are challenging the principle of neutrality. A self-weight-system does not offer any incentive for performing this task honestly. It also makes it very difficult to satisfy the principles of reproducibility and accountability.

Earlier assumptions of achieving an improved accuracy using self-weights have been challenged by later studies (Cooke, 1991).

Principle 5 Fairness:

"All experts are treated equally, prior to processing the results of observations".

Since empirical control and validation is accepted as the means for evaluating expert opinions, in the absence of any empirical information there is no reason for preferring one expert to the other.

3.3 *Structured Expert Elicitation*

A structured expert elicitation process can be subdivided into four work phases:

- Preparation (detailed task specification)
- Elicitation
- Post-Elicitation
- Validation of results and corrective actions

Phases 3 and 4 often can be combined.

3.3.1 *Preparation of expert elicitation*

This first phase of the working process includes the following steps (based on Cooke & Goossens, 2001 with modifications):

1 Definition of project organization. A clear allocation of tasks shall be made with respect to the role of the involved persons and organizations.
2 Definition of case structures document describing the field of interest for which expert judgement will be required.
3 Identification of target variables – these are the variables whose uncertainty shall be quantified by formal expert judgement.
4 Identification of the query variables: these are the variables to be assessed by the experts. It is important to note, that for a technical decision making these variables have to be observable. Query variables are rarely identical to the target variables, although target variables can be query variables, if they are measurable by established procedures. Target variables for which measurement is not possible cannot be quantified by direct elicitation. For these target variables other derived elicitation variables have to be defined, which allow to obtain the probability distribution for the target variables by probability inversion techniques.
5 Development of the aggregation methodology of expert opinions preferably including a calibration procedure. Such a calibration procedure can be

based on a performance-based approach, using seed variables or based on a scaling of expert judgement by a comparison to known observations using once again probability inversion.

6 Identification of experts in the field.
7 Preliminary selection of experts
8 Definition and preparation of input information with detailed specification of the exact questions and the format of expert elicitations.
9 "Table-top exercise" to test the developed procedures with a few experts or with project team members.
10 Expert training session, describing the procedure of expert opinion elicitation and explaining the final use of the expert judgement results to the experts. This shall include a detailed presentation of the definition of the query variables (their physical meaning) and their relation to the target variables. It is recommended to perform expert qualification tests on the understanding of their tasks. In case that equal weighting procedures will be applied for expert opinion aggregation, the test results can be used to eliminate low score experts from the final aggregation.

This first phase of the whole process can be regarded as a detailed project specification of the elicitation process including a test of the feasibility of the developed procedures and formats.

3.3.2 *Elicitation*

This second phase, is the active working phase of the whole process with direct involvement of the selected experts. The main working steps are:

11 Expert information workshops providing the selected experts with the available technical information to obtain an equally informed state of knowledge of the experts. This step is required in case of complex questions which are to be solved with experts from the field but not familiar with specifics of the question in the specific context. In case of a seismic hazards study such specifics may consist in details of the seismic activity distribution in a certain country or region or the specific knowledge about historical seismic events (macroseismic events), information about known seismic hot spots, fault characteristics (regional fault maps) etc. in the area of interest. It is highly recommended to perform an examination test on the understanding of the technical information provided. Safety analysts of the final customer of the project results shall take part in the information workshops to obtain a similar understanding of the important input variables like the involved experts.

12 Expert elicitation workshops, whereby the individual expert' judgements are discussed in the presence of a normative analyst (experienced in probability issues) and a substantive analyst, experienced in the field of interest. In case of a seismic hazard analysis it is recommended that the substantive analyst is an expert from the region.

3.3.3 *Post-Elicitation*

The post-elicitation process is dedicated to the (preliminary) aggregation of results. The term preliminary is used, because in case of seismic hazard analysis in many cases a correction of intermediate results was found to be necessary to exclude extreme assessments from the final quantification. In case of a performance-based approach or another availabl calibration procedure the need for correction is less, as will be shown later. The main procedural steps in this working phase are:

13 The combination of experts' assessments according the developed in phase 1 aggregation methodology.
14 Sensitivity (or input importance analysis) and discrepancy analysis of the preliminary results. The sensitivity analysis shall show, which input variables from the expert judgement have the largest impact on the final results and on the decision to be made. In case of a performance/ seed variable based approach, the sensitivity analysis can be performed as a robustness analysis, f. e. by replacing the best performer by another randomly picked one and by calculating the resulting information loss. If the information loss is large, than another study using another set of experts and seed variables may lead to different results. Discrepancy analysis shall show the areas of the study, where the largest divergence in expert opinion was observed. In case of the use of a formal calibration procedure this step would include a comparison of the assessed query variables with their observed behavior (which might be in a different range of parameters) and the development of calibration parameters.
15 Feedback workshops, discussing the preliminary results of the study with the experts to provide the opportunity for corrections. In case of large differences observed during calibration on query variables, these results shall be discussed, too. It can be necessary to return to step 11 in case of large deviations to the technical information provided as a support for the study.
16 Post processing analysis for the quantification of the uncertainty distributions of the target values.
17 Documentation of the preliminary results.

3.3.4 *Validation of the results and corrective actions*

The validation of results is an essential requirement for any study intended to be used for decision making on critical issues. This is especially true for such complex items like a probabilistic seismic hazard analysis study based to a large extend on expert judgement. Therefore the results of the expert judgement elicitation process have to be validated. Such a validation is also needed with respect to the communication of the results and their possible consequences to the final decision maker and their safety analysts, who prepare the decision making by performing the final overall safety assessment. They shall be convinced from the trustworthiness of the obtained conclusions and this is not possible without validation.

Practical experience has shown, that probabilistic seismic hazard studies had to be corrected due to lack of plausibility of the results or unexplainable differences between different studies using the same or similar methodologies (Lawrence Livermore study in the USA in comparison to the EPRI study, currently the PEGASOS study in Switzerland). This is to some extend also related to the expert elicitation process itself. Larger issues have been observed in cases, than the expert elicitation process did not include calibration procedures or performance based elements. The formal validation process shall include:

18 The development of validation (benchmark) tests, which are based on observations of target or query variables. This may include some limited amount of additional measurements on query or target variables
19 Extensive testing of the study results against the benchmark tests.
20 Feedback discussions with experts and the substantive analyst (in case of SSHAC procedures this would be the TFI) to explain deviations between study results and the performed tests and implementation of a corrective action plan.
21 Publication of the final results.

4 KEY PROBLEMS OF THE PRACTICAL IMPLEMENTATION OF EXPERT JUDGEMENT AS A SOURCE OF SCIENTIFIC INFORMATION

Based on a review of some recent projects related either to probabilistic seismic hazard analysis (PEGASOS, 2004, Klügel, 2004) or to PRA applications (McKay & Meyer, 2000, Bixler 2004) some key issues were identified, which have caused problems in practical applications of expert judgement elicitation processes.

4.1 *Project organization*

Organizational questions are key issues in any large scientific project or safety evaluation. Problems occurred in probabilistic seismic hazard analysis in the following areas:

- Allocation of responsibilities
- Involvement of the final customer / client
- Specification of results
- Organization of the work process in terms of schedule
- Interference of participatory review teams with the project

4.1.1 *Allocation of responsibilities, involvement of the final customer, specification of results*

Problems associated with the allocation of responsibilities can be explained on an example of the practical use of SSHAC procedures (SSHAC, 1997). It is worth to stress the attribut "practical" because the SSHAC procedures as many other theoretical guidelines are driven by "positive intentions" and seem to be very flexible theoretically. The problem is that in practical applications decisions are to be made at each step of the analysis and the flexibility is lost in the process. Because of the extremely high costs of the project there is no chance to return to earlier decisions to make corrections.

As an example the main goal of SSHAC – procedures can be discussed. This main goal consists in "a representation of the legitimate range of technically supportable interpretations among the entire informed technical community" and the assignment of "the relative importance or credibility that should be given to the different hypotheses across the range".

In practice the representation of the legitimate range of technically supportable information is limited to the selection process of experts involved. Because the number of experts is limited for cost reasons the range of "used in practice" technical information is limited by other reasons than scientific ones namely by the project budget. In practice there are also limitations in the expert selection process itself, because some of the best suitable experts are not available due to other obligations. **So the ideal intention to represent the state of knowledge of the technically informed community is not achievable at all.** In practice the expert selection process is also limited due to the preferred selection of experts which had been involved in PSHA studies previously. This is an undue limitation, because an expert in PSHA-methodology or in quantitative seismology is not necessarily a key expert in other required disciplines to be represented in the study like engineering geology, geodetic research, empirical (field) seismology, fault mechanics. For small countries another limitation occurs – it is not possible to select experts which are sufficiently familiar with the detailed seismology of the region and represent independent scientific views at the same time. In the PEGASOS-project it was attempted to compensate this effect by assigning specialists from the Swiss Seismological Service to each expert group of the subproject 1- which was assigned to develop the seismic source characterization. The final consequence of this approach is that the expert evaluations of the different groups are correlated and cannot be processed as independent sources of information in the final aggregation of results, as it was done in the PEGASOS-project.

On the other hand it is worth to mention that the the SSHAC procedures themselves are contradictious. On one side they require to represent the legitimate range of knowledge on the other side they prescribe a methodology (ergodic assumption, use of a stationary homogenous Poisson process, use of a stationary magnitude-frequency relationship) for the aggregation of the results, which is proven to be mathematically inadequate to assess seismic hazard (Klügel, 2004, Klügel & Groen 2004) .

It is worth to have some closer look on the role of the key players in a PSHA following the SSHAC-procedures. In the basic concept of the SSHAC-procedures there exists only one class of participant in the project. This is the class of expert in geoscience which is assigned to different tasks (performed physically by different persons). Other classes of participants or experts like the required above normative expert responsible for a correct aggregation of the results in the pure mathematical sense, safety analysts, who understand the final purpose of the study, structural engineers, who could interpret the project specifications to the scientists in their practical use are excluded from the process.

The only remaining class of expert in geoscience is executing the following tasks:

- The role of the proponent of a specific scientific position or model
- The role of an evaluator of the various positions in the technical informed community
- The role of a technical consultant providing advice to the technical facilitator/integrator
- The role of the technical facilitator/integrator.

The first three roles are assigned to the same persons, just separated by the time, when these roles are to be fulfilled. The experts may interact in groups representing a group expert or as individuals preparing independent scientific views which are serving as an input for the final aggregation of results. Both approaches have been used in the PEGASOS-project. From a common sense perspective it is rather unlikely, that experts who are strong proponents of a specific model or a specific scientific view agree on a common position of an expert group which is largely different to their own views. The most likely outcome of an expert group process or other forms of expert interactions suggested by the SSHAC- procedures is therefore the development of a conservatively bounding diffuse group position, which allows the participants to retain their own scientific positions after the completion of the project without being "refuted by new scientific results" (without "loosing their face"). The agreed position, of course, will be strongly defended by the experts in the aftermath of the project. This is a typical group dynamical effect long known from teamwork theory (psychological circuit of team building). Unfortunately this does not mean that the agreed position of the group does comply with reality. This expert group consensus process resulting in a bounding envelope of the seismic hazard can be traced easily by sensitivity analysis. If the positions of experts in the different groups or even between groups (due to the organization of joint workshops) do not differ significantly it is very likely that such a "converging on conservative assessments" process occurred. Performing some benchmark validation tests can check whether this indeed took place. It is obvious that once such a process occurred the goal of the SSHAC-procedures to reflect the legitimate range of knowledge is turned into its contrary - into a worst case enveloping analysis provided by a limited amount of experts. For the PEGASOS-project there are clear indications that this phenomenon occurred (Klügel, 2004). The experience from the application of PSHA-procedures similar to SSHAC for other sites, showing a steady increase in the obtained seismic hazard assessments in comparison with earlier studies (in most cases without new technical information justifying this increase), also provides some indication, that such a process takes place.

Because a seismic hazard analysis is an interdisciplinary process it is questionable that a project organization as suggested by the SSHAC-procedures (SSHAC, 1997) can succeed in providing an understandable and accepted seismic hazard analysis. As practice has shown, corrections have been required after the completion of such PSHA projects (Lawrence Livermore study as a precursor , Yucca Mountain study, PEGASOS-project), if the results will not simply be enforced by regulatory bodies, without questioning the results.

The procedural guidelines as outlined above provide some improvement and simplifications to the process of expert elicitation:

- The role of the TFI is replaced by two persons, with a key role assigned to the normative analyst (step 11), who is independent with respect of psychological aspects like defending an own scientific reputation in the field.
- There is no need to achieve a group consensus due to the proposed rated weighting of expert opinion (performance based or by a formal procedure for the elimination of outliers). Interactions besides the processing of the required input

information and the suggested training sessions can be limited minimizing the statistical correlation effects between the different expert opinions.

- The procedures allow for and require an active participation of the final user of the results – the customer, especially in the preparation phase of the project and during the validation of the final results. The active participation of the customer in the preparation phase is very important to submit a correct interpretation of the project output specifications to the involved experts. In the case of the PEGASOS-project the obtained intermediate results did not meet the expectations of the customer because the output specification requiring a seismic hazard analysis as an input for a seismic PRA for a short-lived critical infrastructure was interpreted differently by the experts and the project management (no PRA analyst was assigned to the project management) and the customer including different interpretations of such key parameters like peak ground acceleration (for a seismologist a single spike instrumental value, for a safety analyst an effective ground acceleration to be used as an anchor point for a design spectrum or a probabilistic hazard spectrum). On the other hand the active participation of the final customer is important for assuring, that he can follow the analysis process and accept the methodology used.
- The above procedural guidance requires a detailed project specification including the description of the methodology providing the definition of the target and the query variables, the calibration or validation approach and the aggregation method with active involvement of the final customer up to dry test exercises (table top exercises) and the customers supervision of the starting information workshops (in a certain sense the information submitted to the experts in these workshops are a part of the project specification).
- The procedural guidance also clarifies the role of the ownership of the results. The SSHAC-procedures intend to assure the ownership either by the involved experts (concept of expert interdependence and consensus) or of the involved TFIs. This approach is not suitable for a PSHA study which shall be implemented into practical applications, because neither the experts nor the TFIs bear any responsibility on the consequences of their investigations. Therefore for a meaningful seismic hazard analysis it shall be assured that the final client and decision maker can take over the ownership of the obtained results.

4.1.2 *Organization of the work process, time schedule*

A typical error with respect to many safety analysis projects is the use of an inadequate time schedule. Typically the time required for the final aggregation of the obtained results, their interpretation or for corrective iterations is underestimated. Deadlines are often defined by external requirements or budgetary reasons. Practical experience of the author obtained in a substantial amount of safety analyses performed in different fields of nuclear technology has shown that at least 50% of the time schedule shall be dedicated to the aggregation of the final results and their interpretation. For the phased approach suggested in the procedural guidance given in chapter 3 this means that about 50 % of the project labor time shall be assigned to the post-elicitation and validation phase. This may include some re-elicitation of expert opinion on key input parameters. If the project schedule is too tight to assure this goal it is suggested to downscale the project with respect to the amount of experts involved and to the amount of query variables to be investigated.

4.1.3 *Interference of the participatory review team*

The SSHAC-procedures (SSHAC, 1997) recommend a participatory peer review of a PSHA-study due to the large complexity of the methodology. This recommendation applies especially for the process aspects.

A participatory review of a PSHA-study is generally a meaningful approach. The problem in practical application consists in the selection of independent experts, who are not related to experts involved into the main project. This is a complicate question especially for small countries not possessing a large and independent basis for providing geophysical or seismological services. In the PEGASOS-project the requirement of an independent participatory review could not be met due to the lack of independent experts. So experts from the Swiss seismological service took part as experts in the subproject 1 of the PEGASOS project to assure a reasonable input of swiss specific data into the project. On the other hand the head of the Swiss seismological service was a member of the participatory review team, so he actually was reviewing the work of his own service, what by definition cannot be an independent review. Members of the client (or the sponsor like the SSHAC-procedures prefer to name it) were not allowed to participate in the participatory review team. Other members of the review team were experts assigned by the regulator, some of them commercial consultants, which of course are not free from market interests. Therefore the participatory review turned essentially into a regulatory review during the course of the project not giving the final client of the results any possibility to interfere with the project or to review and comment preliminary results. This led at the end to a deviation of the whole project from the essential intend of the project sponsor – to get direct input data for a site specific seismic PRA.

Based on this experience it is recommended, that in similar projects the participatory review is performed by an interdisciplinary team of independent experts led by safety analysts and engineers, who are staff members of the client and sponsor of the study including independent seismologists. In case of an intended use of the results for a seismic PRA, experienced PSA specialists shall be included into a leading position of the review team.

4.2 *Treatment of uncertainties and validation of results*

The main purpose of using expert judgement consists in the intend to incorporate the best available knowledge and the associated uncertainties into the decision making process. Nevertheless there are different ways how to treat and to incorporate uncertainties into the practical decision making process. This question is closely related to the question of the calibration of expert opinion and the aggregation of the obtained evaluation into the final synthesis of analysis results.

Expert judgement in the most simple way (and many studies are still performed in this way) is performed by directly asking experts for model parameters which then will be applied in the final analysis (McKay & Meyer, 2000). To take into account possible but not investigated calibration errors or simple biases of experts an averaging procedure (or a weighting procedure being a slight progress) is applied to obtain aggregated results. Such an approach is simply based on "blind trust" into expert knowledge and is suitable at the best to make rough estimates for simple problems. The main issues not addressed by these methods are discussed below.

4.2.1 *Correlation between assessed variables and expert judgements*

Assessed variables and model parameters are often correlated. In seismic hazard analysis a simple example consists in the correlation between the assessment of the magnitude of a seismic event and the development of an attenuation law used for the evaluation of ground motion levels at a certain site. If the assessment of magnitudes is performed consistently at the higher level (of a certain uncertainty range) the resulting developed (by experts) attenuation law will show high attenuation for the region of interest (based on the measured ground motions). If the assessment of magnitudes is performed consistently at the lower level the resulting developed attenuation law will indicate low attenuation for the region of interest given the same measured site-specific ground accelerations.

This example shows that the assessment of earthquake magnitudes cannot be separated from the question of the development of an attenuation law. The same applies to the transfer of attenuation laws from other regions to the region of interest because the way how the evaluation of earthquake magnitudes was performed in one region may not be compatible with the way how other experts are solving the same tasks in another region.

From the PEGASOS project an interesting example can be mentioned. The Swiss seismological service performed some reevaluation of the Basel earthquake from 1356 assigning it to a magnitude M_w=6.9. In the corresponding French earthquake catalogue the same event is assigned to a magnitude M_w=5.9. Using the same attenuation law would result in completely different ground motion levels. The peak ground acceleration would differ about a factor of 2 (for Central European Conditions). If such differences were interpreted as uncertainty in attenuation correlations and correspondingly incorporated into a probabilistic seismic hazard analysis without taking into account the correlation between the assessment of magnitudes and the development of attenuation laws the results would be completely skewed. This happened in the PEGASOS-project, which separated the assessment of source characteristics (subproject 1) from the development of an attenuation law (subproject 2) and from site effects (subproject 3) treating the separately obtained model parameters as statistically independent in a huge logic tree. A corrective approach would consist in a selective linking of the obtained results in an event tree providing logical rules which assures that low estimates of earthquake magnitudes f. e. of the used upper limit magnitudes in case of a truncated exponential distribution) are combined statistically with attenuation laws resulting in a weaker attenuation while high estimates of earthquake magnitudes are combined with attenuation laws resulting in a larger attenuation. The blind coupling of independently assessed model parameters in a huge logic tree will result very likely in error.

On the other side it is worth to mention that expert judgements themselves are correlated, if the same expert is asked to assess several model parameters. This puts an additional constraint on the linking of logic trees, because it requires that the results of each expert are treated separately from the results of other experts on the same questions.

4.2.2 *Convergence to reality, validation of results*

The idea of asking experts to provide their judgements is to get the best available knowledge on the topic of interest. This means that the results of expert judgement are expected to be in correspondence with reality, which for some reasons cannot be explored directly by other types of investigations. In a more formal sense with respect to seismic hazard analysis this means that the results of the analysis shall converge to physical reality. It is understood that (especially in contemporarily low seismic areas) it might be difficult to prove this correspondence with reality due to lack of data, but it is possible to apply procedures for expert judgement which assure a higher likelihood that the judgement of experts are trustworthy.

In the simple methods mentioned above no sanity checks are performed to prove the correspondence of the results to reality.

The SSHAC procedures (SSHAC, 1997) also do not require the validation of the obtained results, although they do not exclude it. Therefore compliance with reality cannot be assumed for the results obtained by the SSHAC methodology without additional validation effort.

To assure a check of the obtained results in a systematic way a validation phase as suggested in the procedural guidance on expert elicitation above shall be included into the working process. The need for a validation of results in case of a seismic hazard analysis is meanwhile recognized even by seismologists (Musson, 2004).

The effort for validation of results can be reduced if performance based methodologies (explicit calibration) or techniques deriving model parameters directly from expert judgements on query variables are used. Such methods will be discussed in detail in section 4.3.

4.2.3 *Minimization of the number of random model parameters and aggregation of results*

This requirement to expert elicitation processes is rarely realized in practical applications. The number of random parameters is normally derived from the field of science and therefore excluded from a systematic review before the expert judgement process is started. It is simply assumed, that the experts in the field are able to formulate a minimal set of random parameters to be used as target variables for the expert judgement process. That this question is not as obvious as it might appear can be demonstrated for the case of a probabilistic seismic hazard analysis following the SSHAC procedures. Contemporary probabilistic seismic hazard analysis is based on the use of logic trees for the treatment of uncertainties. Each branch of the logic tree represents one of the random parameters to be assessed by the experts. In the PEGASOS project used here as an example the probability distribution for each of the parameters was described by a discrete probability distribution. This leads to a representation of the seismic hazard model in the format of a logic tree (or an event tree) with multistate branches. Each of the possible states is assigned to a weighting factor representing the discrete probability distribution used for the random parameter asked for at the branch. The seismic hazard for a given frequency of exceedance can be represented then as follows:

$$H(f) = \sum_{i=1}^{k^{n-1}} \prod_{j=1}^{n} w_{ij} F(f, w_{ij}) \qquad (1)$$

where *H(f)* is the hazard level (on-site spectral acceleration) for a given frequency of exceedance and a random variable, *k* is the maximum number of possible states at a single branch in the logic tree (multi- state branches), w_{ji} is the weighting factor used at branch j for the state i- derived from the discrete probability distribution for branch j , $F(f,w_{ji})$ is the assignment law - assigning an on-site ground motion level (spectral acceleration) to the fixed frequency of exceedence f depending on the values of the random weighting factor w_{ij} which is given implicitly by the seismic hazard model, mainly driven by the attenuation law, *n* being the number of branches of the combined logic tree. The function $F(f, w_{ij})$ can be thought of as an assignment law assuring that the input required in the logic tree for branch j+1 is provided at branch j. Running through the whole tree (in practical calculations with some re-ordering) assures that the requested spectral acceleration corresponding to the fixed frequency of exceedance f and to the set of weighting factors w_{ij} is calculated. In case of using expert judgement for the assessment of the distribution parameters at the branch j, w_{ij} itself is a random parameter representing the uncertainty of the expert judgement.

It is obvious, that the distribution of the calculated hazard presented for a fixed frequency of exceedance in equation (1) depends on the number of branches in the combined logic tree. That means the more random parameters are used to model the seismic hazard, the more diffuse the final distribution will be. In case of separate treatment of "epistemic" and "aleatory" uncertainties (for this separation there is no mathematical justification, occurrence of earthquakes does not represent an ergodic stochastic process) the number of random variables will be increased artificially due to an increased number of branches in the tree.(without reason).

Similar conclusions can be drawn with respect to the use of experts and the aggregation of expert opinion. In the case of a weighted use of multiple experts for the assessment of seismic hazard the number of branches in the logic tree will increase at least by one multistate branch – the expert branch. If the hazard logic tree and the expert judgement process is subdivided into several subprojects the number of additional branches will increase by the number of subprojects used. In the PEGASOS-project 3 subprojects (the fourth dedicated to the calculation can be discarded) were used, meaning that the logic tree was increased by 3 branches. The observed diffuse distribution of the calculated seismic hazard was partially a consequence of this approach. It shall also be noted that the number of experts used in the assessment has some impact on the diffusity of the results. With increasing number of experts the resulting distributions will get more and more wide spread and can get meaningless for practical decision making (or equally - the same decision would be made very likely without performing the study, the effort to perform the study is in vain). The conclusion from this observation is, that alternatives shall be looked for to assure a meaningful incorporation of expert judgement into the analysis. These alternatives shall be aimed at reducing the interface boundaries of the analysis and at providing empirical control of the results.

4.3 *Calibration, performance-based aggregation of expert judgements and inverse probability techniques*

A meaningful alternative for incorporation of expert judgement would consist in a performance – based approach. Based on a calibration procedure the results of expert judgement would be used for the development of a "decision maker –DM-model" (please note the different use of the term decision maker here in comparison to decision makers in terms of management personal) – a performance based model derived from the expert elicitation process based on the results from calibration on seed variables. The most challenging approach would consist in the selection of a single decision maker (a single expert) – the best performer in the calibration tests for each of the evaluated questions. This approach is equally to a qualification of experts with respect to a given problem and therefore controversial. It may be not possible to select a decision maker in the sense of a single person.

Looking at equation (1) it can be seen, that the use of a "decision maker" model can potentially reduce the diffusity of the results of the analysis, by reducing outliers in the assessment. This conclusion can be drawn, because the distribution of ground motion parameters for a fixed frequency of exceedance is lognormal and therefore the mean is largely affected by outliers in the analysis. Removing them leads to a more stable mean hazard assessment. The use of a single performance based "decision maker model" to some extend also accounts implicitly for the correlation between different parameters of the analysis, because the assessment of a single person in different areas of the analysis can be expected as being more consistent (in the sense of balanced) then the "blind" aggregation of the results of evaluations from different experts. It also makes the empirical control and the validation process easier due to the increased simplicity of the approach.

A fundamental assumption of any performance based method is that the future performance of experts can be judged of past performance, as reflected in the evaluation of seed variables. The basic idea of calibration of expert judgement consists in the statistical evaluation of the performance of experts on seed variables. Seed variables are variables from the experts' field whose values are become known to the experts post-hoc. Seed variables serve (Cooke & Goossens, 2000) :

- to quantify experts' performance as subjective probability assessors
- to enable performance – optimised combinations of expert distributions and
- to evaluate and if possible to validate the combination of expert judgement.

Calibration and scoring techniques have been developed by a few authors. An overview can be obtained in the standard text book of Cooke (Cooke, 1991). The author follows to a large extend this breakthrough work, which was tested on a large amount of expert elicitations. The approach was also tested in the US NRC/EU project on uncertainty analysis in accident consequence modelling.

Calibration measures the statistical likelihood that a set of experimental results (or empirical observations) correspond in a statistical sense, with the expert assessments. In the USNRC/EU study on accident consequence modelling as the calibration score the p-value of a standard chi squared goodness of fit test was used. The calibration score can then be regarded as the probability that the divergence between expert's probabilities and the observed values of the seed variables might have arisen by chance. A low score means in the statistical sense that the experts are likely to be wrong. A high score (near 1, at least higher than let's say 0.05) means that the expert's probabilities are statistically supported by the set of seed variables. In the most simple performance-based weighting approach just the calibration scores can be used directly for the development of performance based - weighting factors. In a more elaborated approach usually referred as the classical method (Cooke, 1991, Cooke & Goossens, 2000) and used for assessing uncertainty distributions for input parameters for accident consequence modelling - information in the US NRC/EU projects used as an additional quantitative measure of performance. Hereby the overall information score is the mean of the information scores for each variable. This is proportional to the information in the expert's joint distribution relative to the joint background measure, under the assumption of independence. Independence in the experts' distributions means that the experts would not revise their distributions for some variables after seeing realisations for other variables (learning is excluded). In principle different information measures can be used f. e. from Fishers' information matrix to a direct use of Shannons' or a similarly derived entropy. The use of entropy based information measures is very common in PRA applications and has been extensively used at the Technical University of Delft (Cooke, 1991). For a discrete probability distribution (mass function) P over the integers i=1,…N the associated entropy is:

$$H(P) = -\sum_{i=1}^{N} P(i) \ln P(i) \tag{2}$$

It is obvious that the more wide spread the mass function P is, the larger will be *H(P)*. The information is defined as:

$$I(P) = -H(P) \tag{3}$$

Obviously a high information value (or a low entropy) is a desideratum in expert analysis. While preparing complicated technical decisions it is reasonable to prefer advice of the expert whose probability functions have the lowest entropy or the highest degree of information.

Following Cooke (Cooke, 1991) the idea of developing a calibration score can be explained by the following example. Lets assume that an expert gives the same probability mass function P for a large number n of physically unrelated uncertain quantities. By observing the true values for all these quantities it is possible to generate a sample distribution S with *S(i)* equal to the number of times the value i was observed, divided by n. Comparing statistically the compliance between the sample distribution S with the distribution P one can make conclusions on the quality of calibration. Roughly, we can conclude, that the expert is well calibrated if the true values of the uncertain quantities can be regarded as independent samples of a random variable with distribution P. This entails, that the discrepancy between S and P should be no more than one might expect in the case of independent multinomial variables with distribution P. It is therefore possible to interpret the statement – the expert is well calibrated – as the statistical hypothesis:

$Cal(P)$:= the uncertain quantities are independent and identically distributed with distribution P.

The "discrepancy between S and P" can be measured by the relative information of S with respect to P, *I (S,P)*:

$$I(S,P) = \sum_{i=1}^{N} S(i) \ln\left(\frac{S(i)}{P(i)}\right) \tag{4}$$

I(S,P) may be interpreted as a measure of surprise, which someone would experience if he believed P and subsequently learned S. It is obvious, that large values of *I(S,P)* would be critical for the calibration hypothesis. According to Cooke (Cooke, 1991) the "degree to which the data support the hypothesis *Cal(P)"* as the probability under *Cal(P)* of observing a discrepancy in a sample distribution S' at least as large as *I(S,P,* on n observations:

$$\Pr ob\left\{I\left(S', P\right) \geq I\left(S, P\right) \middle| Cal(P), n_{observations}\right\} \tag{5}$$

This probability can be used to define statistical tests in the classical sense. Of practical interest is that it can be shown, that if P is concentrated on a finite number N of integers that include all observed values, then as the number n of observations gets large, *2nI(S,P)*becomes χ^2- distributed with N-1 degrees of freedom. This allows to use the familiar χ^2 statistic for testing goodness of fit between the sample distribution S and the "theoretical" distribution P to score the calibration hypothesis. Because in principle it can be possible that the calibration score (by directly comparing "the theoretical distribution" obtained from expert judgement with the sample distribution obtained from observations (on seed variables)) is low (below 5%) while the score based on information (entropy) is reasonable (*I(S,P)* is low converging to zero with n going to infinity) it is important to outline as a conclusion, that "Good experts should have good entropy scores and good calibration scores" (Cooke, 1991). Approaches which are based on a combination of calibration and information scores in the way described, are used to be called as the "classical approach" due to the analogy to the calibration of measurement devices (Cooke, 1991).

The example given is not yet very practical since it requires a large number of quantities for which the expert gives the same probability distributions. If a set of random variables X_i have invertible cumulative distribution functions F_i, i=1,2…, it is not difficult to find transformations under which the variables become identically distributed. Under this condition ist suffices to consider the transformed variable, as these all have the uniform distribution on the unit interval. Further simplifications are possible by limiting the elicitation process to the elicitation of percentiles instead of trying to perform the elicitation for the whole mass function.

Such a combined performance based model (calibration and information score based on seed variables) was derived using the following properties (taken from Cooke & Goossens, 2000)

1 Calibration dominates over information, information serves to modulate between more or less equally well calibrated experts.

2 The score is a long run proper scoring rule, that is, an expert achieves his/her maximal expected score, in the long run, by and only by stating his/her true beliefs. Hence, the weighting scheme, regarded as a reward structure, does not bias the experts to give assessments at variance with their real beliefs, in compliance with the principle of neutrality.
3 Calibration is scored as "statistical likelihood with a cut-off". An expert is associated with a statistical hypothesis, and the seed variables enable us to measure the degree to which that hypothesis is supported by observed data. If this likelihood score is below a certain cut-off point, the expert is unweighted (disqualified!!).The use of a cut-off is driven by property (2) above.
4 The cut-off value for (un)weighting experts is determined by optimising the calibration and information performance of the combination (this is essentially a removal of the outliers).

In general weights of experts e (from 1 to N) for a set of evaluated random parameters i (1 to K) associated to this performance based approach can be defined as:

$$w_e = \frac{Cal(e)\Delta_{cal}(Cal(e))}{\sum_{e=1}^{N} w_e} I(e) \qquad (6)$$

where *I(e)* is an average measure for the information degree associated with the judgement of the expert. In case of discrete probability distributions the information measure is just the reciprocal of the average response entropy calculated from M observations for the K parameters (these are the observations for the seed variables)

$$I(e) = \frac{1}{H_e(p)} \qquad (7)$$

while in case of a continuous distribution (assessment based on percentile tests) *I(e)* is the average relative information over the M observations for the cumulative minimal information densities $Q_{i,e}$ meeting the requirement that the fractiles of Q match the fractiles of expert e's assessment for the parameter i with respect to the uniform distribution. The uniform distribution is usually selected as a background measure, because it possesses the property of maximum entropy, thus is usually the most uninformative comparison pair to the assessment of the expert (the expert is judged, how much he is improving knowledge in comparison to the uninformative uniform distribution).

$$\begin{aligned} \Delta_{cal}(x) &= 1 : x \geq \alpha, \\ \Delta_{cal}(x) &= 0 : x < \alpha, \alpha \in (0,1) \end{aligned} \qquad (8)$$

α is the significance measure of the calibration test (χ^2-Test for 2M *I(S,P) for M being large*). From equation (6) it can be seen, that calibration dominates over information (entropy) by assigning a weight 0 to the expert, if his calibration score is low. Cooke (Cooke, 1991) has shown that these scoring measures are proper.

The use of seed variables supports empirical control of any combination schemes, not just those which optimise performance on seed variables. Several studies have been published implementing this methodology and can be used as a reference (Cooke et al, 1988, Goossens et al, 1997, Gosseens et al, 1998, Cooke & Jager, 1998). The advantage of the method consists in rewarding experts for expressing their true belief and the possibility to avoid increased correlation between the expert evaluations due to enlarged discussions as suggested in the SSHAC-methodology which entails the danger of "converging on worst case assumptions". A problem of the performance based method is, that the result of the expert judgement aggregation to some extend depends on the quality of the used seed variables.

With respect to seismic hazard analysis the methodology can be modified. For example, it can be taken into account, that Berril & Davis (1980) have shown that the exponentially truncated Gutenberg-Richter-Correlation corresponds to the maximum Shannon entropy for the probability density distribution of earthquake magnitudes. So the truncated Gutenberg-Richter-Correlation can be used as a background measure with parameters assessed by maximum likelihood estimators (Weichert, 1980) to judge the information gain derived from expert judgement elicited to specify seismic source characteristics of a certain area instead of using a uniform distribution. This might be of special interest in areas with diffuse seismic activity. In this case information or entropy measures can also help to provide empirical control on the results of expert judgement. This can be illustrated on an example.

Lets assume, that for a seismic hazard analysis to be performed an earthquake catalogue is available. This catalogue is also the basis for the expert elicitation process to be performed on the characterisation of the seismic sources in the relevant area. Experts are asked to provide their assessments on the probability distribution of the parameters a and b in the format of a truncated Gutenberg-Richter-Correlation. From the parameters of the correlation an statistical assessment of the upper magnitude is performed. This means that the results of the experts are presented in a discrete probability distribution for the triplets (a,b, M_{upper}) for all individual seismic sources considered for the region of interest for one of the zonation schemes developed by the experts. The discrete probability distribution is associated by a corresponding set of discrete probabilities (or weighting factors) $p_{i,e}$ which are used at the corresponding branch of the logic tree. The index i is related to the source i and the index e is a shortcut for the expert (expert group) e. Each expert or expert group had the task to develop a few, different zonation schemes (total number L, with a total maximum number of K sources for each zonation scheme) to reflect the experts assessment of the uncertainty of the location of seismic sources. This task would be a standard procedure, f. e. in a formal expert elicitation procedure like SSHAC (in PEGASOS this was a task of the subprojekt SP1). With respect to the results it can be expected, that for each expert the following relations are valid:

$$\sum_{l=1}^{L} w_l \sum_{i=1}^{K} \left(N \middle| m \geq M_{low}\right)_i^2 p\left(N \middle| m \geq M_{low}\right)_i - 10^{2a_{catal}} \leq \sigma_{a_{catal}}^2 \quad (9)$$

$$H_e\left(N\right) \leq S_{catal}\left(N\right) \quad (10)$$

Correlation (9) requires that the calculated weighted mean of the total number of earthquakes with a magnitude m larger or equal M_{low} in the region of interest obtained from the elicitation of expert e deviates from the calculated mean number of the earthquake catalogue not more than corresponds to the variation of catalogue data. This correlation puts an empirical constraint on the number of earthquakes.

Correlation (10) puts an empirical constraint on the entropy (on the information content) of the experts e results by requiring, that the weighted entropy of his results shall be lower than the entropy calculated from the truncated Gutenberg-Richter-correlation of the complete earthquake catalogue. This correlation is obviously justified, because the goal of expert elicitation is to incorporate additional knowledge into the decision making process in addition to already available information. Correlation (10) is once again related to the means. The weighted entropy for the expert assessment is calculated as:

$$H_e(N) = -\sum_{l=1}^{L} w_l \sum_{i=1}^{K} S_i\left(a_i, b_i, M_{upper,i}\right) p_{i,e} \quad (11)$$

The entropy S_i is to be calculated for each of the triplets (a_i, b_i, $M_{upper,i}$) as well as for the catalogue by the following equation:

$$S = -\int_{M_{lower}}^{M_{upper}} p(M) \ln\left(M\right) dM \quad (12)$$

where *p(M)* is the truncated exponential version of the Gutenberg-Richter-correlation.

$$p\left(M\right) = \frac{\beta}{1 - e^{-\beta M}} e^{-\beta M} \quad (13)$$

with

$$\beta = \ln\left(10\right) b \quad (14)$$

Relations (9) and (10) shall be fulfilled together. The results of experts, not meeting these requirements shall be removed from the final aggregation of the results, because their results do not correspond or do not add additional knowledge to the already available technical information.

The principal alternative to a performance based aggregation of expert judgement results would consist in the use of inverse probability techniques (Cooke & Goossens, 2000, Kraan & Cooke, 2000). If the query variables can be related directly to observable parameters this technique assures that the resulting distributions will be constrained in a very natural way by in principle measurable data. This technique has been applied as part of the project for the analysis of uncertainties in accident consequence modelling. The basic idea consists in an indirect development of the required probability distributions of model parameters by eliciting expert opinion on directly observable parameters. Returning to the example of performing an expert elicitation on seismic source characteristics this means, that experts would not be asked to provide probability distributions to the parameters a or b of the Gutenberg-Richter correlation directly, but rather would be asked how many earthquakes above a certain magnitude they would expect in a certain area (given the technical information on seismic activity in the region of interest, fault map, historical events in the format of a declustered catalogue) for a certain period of time. It is also possible to obtain information on return periods (earthquake recurrence periodicity). This information than will be used to develop the parameters of the model by probabilistic inversion. Because the solution of the problem is not unique, the shape of the resulting distributions shall be derived on physical considerations or where this is not possible by using the principle of maximum entropy. For this purpose the expert assessment can be combined with measured data – f.e. with the results of a declustered catalogue. Statistical techniques to develop probability distributions by inversion are available. With respect to the SSHAC-procedures this methodology provides more flexibility, because also Non-Poissonian models could be used (f. e. Lognormal model for return periods or Markov models).

5 CONCLUSIONS

Expert opinion can be a valuable source of scientific information, if obtained by a systematically structured expert elicitation process. Such process shall be based on a rational consensus approach. An expanded set of procedural guidance rules has been developed based on guidelines originally developed by the Institute of Applied Mathematics of the Technical University in Delft (Prof. Cooke and his co-workers). The currently popular among proponents of probabilistic seismic hazard analysis SSHAC-procedures do not fully comply with these rules and shall be improved by:

- Providing empirical control on the results of expert judgement
- Introducing performance -based expert opinion calibration
- Adding a normative experts as a participant of the expert elicitation meetings besides the TFI, whose role shall be reduced to the role of the substantive expert
- Reducing methodological dependencies between the experts (by reducing the number of common workshops, providing more freedom in the selection of the mathematical models etc.) to allow for diverse scientific opinions
- Taking into account physical dependencies by an appropriate project organisation
- Utilizing the principle of minimisation of the number of random model parameters
- And last but not least by providing an extended participation of the client of the studies preferable in a project management roles as well as in the role of a participating observer in the elicitation process.

REFERENCES

Berrill, J.B. & Davis, R.O. 1980. Maximum entropy and the magnitude distribution. Bulletin of the Seismological Society of America, Vol. 70, No.5, pp. 1823-1831.

Bernreuter, D.L., Savy, J.B., Mensing, R.W. & Chung, D.H.,1984. Seismic Hazard Characterization of the Eastern United States Methodology and Interim Results for Ten Sites. NUREG/CR 3756, 1984.

Bixler, N.A., Morrow, C., Phillips, J. & Fatenejad, M. 2004.Methodology for Assessment of Uncertain Input Data for a Level –3 PRA Analysis of A Nuclear Reactor Accident MACCS2", In: C. Spitzer, U. Schmocker and V.N. Dang (eds), Probabilistic Safety Assessment and Management, pp.3685 –3690. Berlin: Springer.

Cooke, R.M. 1991. Experts in Uncertainty. Opinion and Subjective Probability in Science. New York, Oxford: Oxford University Press.

Cooke, R.M. & Goossens L.H.J. 2000. Procedures Guide for Structured Expert Judgement in Accident Consequence Modeling. Radiation Protection Dosimetry, Vol. 90 (3), 303-309.

Cooke, R.M., Mendel, M. & Thys, W. 1988. Calibration and information in Expert Resolution: a Classical Approach. Automatica 24, 87-94.

Cooke, R.M. & Jager, E. 1998. Failure Frequency of Underground Gas Pipelines: methods for Assessment with Structured Expert Judgement. Risk. Anal. 18(4), 511-527 .

Cornell, A. 1968. Engineering seismic risk analysis, Bull. Seism Soc. Am;58:1583-1606.

Gill, J. Bayesian Methods. 2002. A Social and Behavioral Sciences Approach. Boca Raton: Chapman & Hall CRC.

Goossens,L.H.J., Cooke, R.M., Woudenberg, F. & van der Tom, P. 1998. Expert Judgement and Lethal Toxicity of Inhaled Chemicals, J. Risk res. 1(2), 117-133.

Goossens, L.H.J. & Cooke,R.B. 1997. Applications of Some Risk Assessment techniques: Formal expert Judgement and Accident sequence Precursors. Saf. Sci. 26(1/2), 35-48.

Kahn, H. & Wiener, A.J. 1967. The Year 2000, A Framework for Speculation. New York: Macmillan.

Kraan, B.C.P. & Cooke, R.M. 2000. Processing Expert Judgements in Accident Consequence Modeling. Radiation Protection Dosimetry, Vol. 90(3), 311-315.

Klügel, J.-U. & Groen, F. 2004. Uncertainty analysis performed for the review of a large scale probabilistic seismic hazard analysis. Paper B1P3 submitted to ESREL 2005. Accepted for publication October 30th,2004. .

Klügel, J.-U. 2004. Problems in the Application of the SSHAC Probability Method for Assessing Earthquake Hazards at Swiss Nuclear Power Plants, accepted for publication in Engineering Geology.

Krinitzsky, E.L.,1993. Earthquake probability in engineering – Part 1: The use and misuse of expert opinion, Eng. Geology. 33, 257-288

McKay, M. & Meyer, M. 2000. Critique and Limitations on the Use of Expert Judgements in Accident Consequences Uncertainty Analysis. Radiation Protection Dosimetry, 90(3), 325-330.

Musson, R.M.W. 2004. Objective Validation of Seismic Hazard Source Models, 13th World Conference on Earthquake Engineering, Vancouver, August 1-6, Paper No. 2492.

Noubary, R. 2000. Statistical Methods for Earthquake Hazard Assessment and Risk Analysis. New York: Nova Science Publishers, Inc.

NUREG/CR-4551, Vol.2 1991. Evaluation of Severe Accident Risks: Quantification of Major Input Parameters. Experts determination of Containment Loads and Molten Core Containment Interaction Issues.

PEGASOS. 2004. Probabilistic Seismic Hazard Analysis for Swiss Nuclear Power Plant Sites (PEGASOS Project), Vol. 1 –6. NAGRA.

Senior Seismic Hazard Analysis Committee (SSHAC). 1997. Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts. NUREG/CR-6372.

Weichert, D.H. 1980. Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. Bulletin of The Seismological Society of America, Vol. 70, No. 4,pp. 1337-1346.

IAEA/ICTP Workshop on
Earthquake Engineering for Nuclear Facilities - Uncertainties in Seismic Hazard Assessment

"Expert Judgement Methods"

Trieste, Italy, 14 – 25 February 2005

(Unit 26) - (Dr. Jens-Uwe Klügel)

Contents

- Introduction
- Classification and historical development of expert judgement methods
- General Procedural Guidelines
  - Uncertainty, expert judgement and technical decision making
  - Principles of rational consensus
  - Procedural steps of a structured expert elicitation process
- Key problems of practical applications
- Modern techniques of aggregation of results
- Conclusions

Introduction - I

- Use of expert knowledge is an essential part of any practical decision making
- “Expert knowledge NOT EQUAL to expert opinion, expert opinion shall reflect expert knowledge (at least a mirror of expert knowledge)
- Goals of a Decision maker (Manager)
  - Incorporate the best available knowledge on the topic into the decision making process
  - Provide a more robust basis for final decision making

→ Improve information basis for decision making

Introduction - II

- Expert judgement methods shall be judged against these goals
  - What are the additional benefits from expert elicitation in comparison to the available information before the start of expert elicitation?
- Experts are human beings and thus subject to cognitive limitations and psychological bias
  - Expert judgement results shall be checked and validated or corrected, once new information became available
  - Many examples, that experts „were wrong“, but nevertheless their judgement helped towards the resolution of critical topics by later additional research

Introduction III

- Lecture will present some general procedural guidelines with special focus to seismic hazard analysis
- Focus on structured processes including a quantification process
- Based on a Methodology developed at the Technical University in Delft (Prof. Cooke et al)
  - Used in a common US NRC & EU project on uncertainty analysis of consequences of severe accidents (Level 3 PRA)
  - Approved in about 11000 expert elicitations

Classification of Expert Judgement Methods -I

- Possible Classification
  - Purpose of application
  - Degree of formalization
  - Used mathematical methods

Classification of Expert Judgement Methods -II

- Classification by purpose of application
  - Interpretative application – to explain empirical information contradicting existing theories or beliefs
  - Forecasting applications – predict technological or socio-political developments, develop strategies
  - Phenomenological applications –quantitative description of (possible) phenomena, which cannot be investigated directly
  - Informative applications – insufficient information, expert opinion provides an additional complementary source of information
    - Expert Judgement in Seismic Hazard Analysis – combines phenomenological, informative and interpretative elements

Classification of Expert Judgement Methods -III

- Classification by degree of formalization
  - Unstructured (mostly qualitative) approach
    - Brainstorming (collection of ideas)
    - Structured Brainstorming – Ishikawa's fish-bone diagrame
  - Structured
    - Simple f.e. questionnaire techniques (marketing)
    - Complex – formal expert selection, expert elicitation including normative elements, formalized weighting procedures
  - Direct and indirect Methods

Classification of Expert Judgement Methods -IV

- Classification by mathematical methods
  - Simple – direct parameter assessment (in engineering applications)
  - Complex
    - Use of subjective probability / formalized decision theory
    - Use of fuzzy set mathematics
- Engineering experience
  - Often good results for simple approaches if empirical control on the results is maintained (setting constraints on results by data)
  - Often poor results for complex methods if empirical control on results is not assured

Development of Structured Expert Judgement -I

- Origin in some military applications
  - Oracle type Delphi-Method
  - Think tanks for scenario-based forecasting
- In the US – foundation of the RAND corporation (US-airforce and Douglas aircraft)
  - Before Vietnam war – 11000 think tanks working for the US government

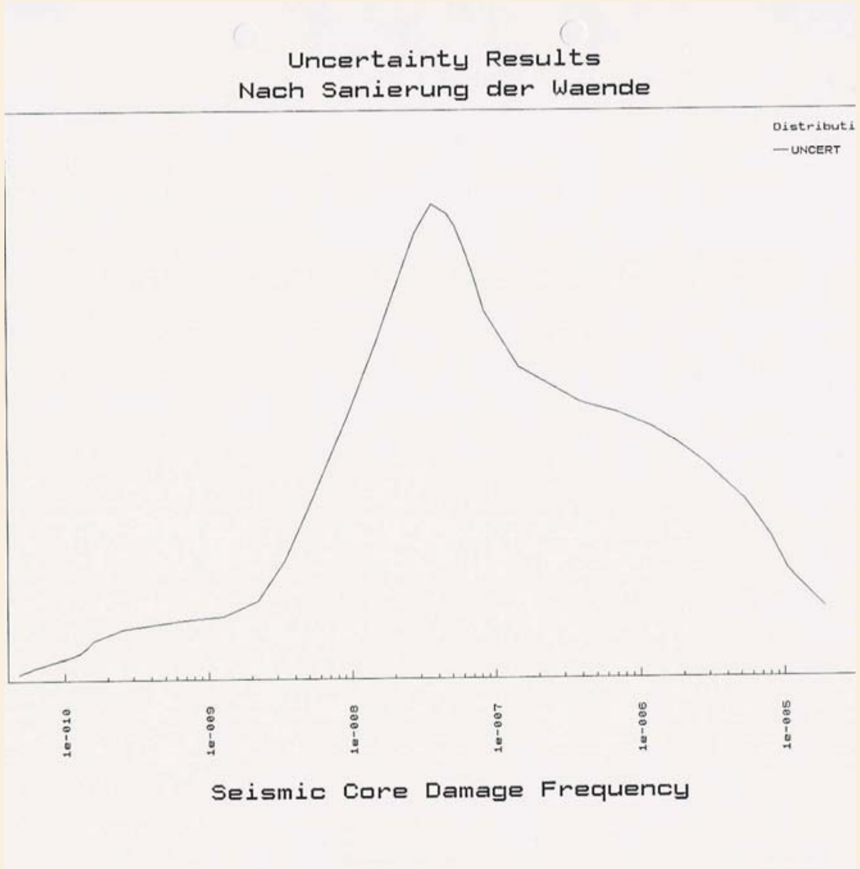
Development of Structured Expert Judgement –II – Scenario-based Methods

- „Scenarios are hypothetical sequences of events constructed for the purpose of focusing attention on causal processes and decision-points.They answer two kinds of questions: (1) Precisely how might some hypothetical situation come about, step by step? And (2) What alternatives exist, for each actor, at each step, for preventing, diverting, or facilitating the process.“ (Kahn & Wiener, 1967)
- Large similarity to the event tree approach in PRA (a decision tree is just another event tree)
- Observation- The scenario based approach was constructed as a systematic think process by a trial to simulate political thinking
  - In practical applications = a political consensus approach to reflect the opinion of different stakeholders, weighting factors/ probabilities at decision points were reflecting this political consensus approach
  - Diversion from subjective probability and formal decision theory because it deemed impossible to incorporate different opinions and interests

Development of Structured Expert Judgement –III – Scenario-based Methods

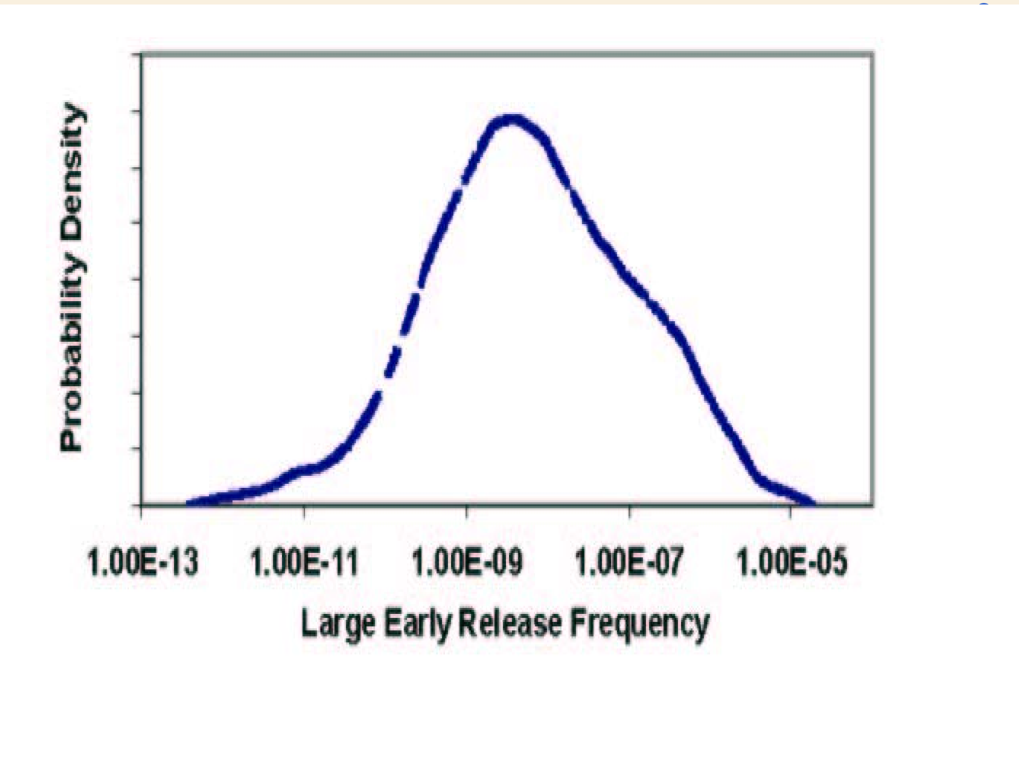
- Formal critic to mathematical decision theory due to the observed very flat probability distributions, making possible decisions undistinguishable (no difference what decision shall be made)
- „The subjective curve of probabilities often seems flat...In order to avoid the dilemma of Buridan‘s ass, who starved midway between two bales of hay because he could not decide which one he preferred, we must then make arbitrary choices among almost equally interesting, important, or plausible possibilities. That is, if we are to explore any predictions at all, we must to some extent ‚make them up‘. (Kahn&Wiener, 1967).

Development of Structured Expert Judgement –IV – Scenario-based Methods



- Interesting observation
  - Modern expert elicitation processes combined with subjective probabilities lead again to very flat distributions – Seismic CDF-distribution for the Goesgen plant (Klügel, 2004)
  - Buridan‘s ass dilemma reoccured

RE-Occurence of Buridan‘s ass Dilemma in Probabilistic Seismic Risk Assessment



- Even improved techniques for reduction of uncertainties at the plant logic level does not improve the situation very much

What shall a decision maker do?

Go for the median – no measures required, risk is low, Go for the mean – some effort for upgrades may be reasonable, Go for the max value – shutdown the plant for upgrades

Development of Structured Expert Judgement –IV – Delphi-Method

- Systematic questioning of carefully selected experts, which did not know each other
  - Special questionnaires
  - Results processed statistically to obtain the median, and the quartiles
  - Experts giving answers outside the range of the 25%/75% quartiles were asked to provide justification or to provide corrections
  - Several iterations were performed to get an increasing range of convergence

Development of Structured Expert Judgement –V – PRA-applications

- NUREG-1150 and subsequent reports
  - Analysis and quantitative description of severe accident phenomena (basis of modern containment event trees)
  - Systematic expert elicitation process
      - Contained many elements of a rational consensus approach
    - Systematic expert selection
    - Use of a substantive and normative facilitator
    - Training and dry exercise sessions
    - Formalized mathematical aggregation procedures
  - Incomplete empirical control, no calibration
  - But – results not intended to stand forever – gave directions for later research efforts to improve knowledge by data and facts

Development of Structured Expert Judgement –V – PSHA-applications, SSHAC-procedures

- A separate (side path) of the development of expert judgement methodology consists in the development of the SSHAC-procedures
  - Two predecessor studies (but differ in details)
    - Lawrence Livermore
    - EPRI
  - 4 different levels
    - Level 1 to 3 resemble a team moderation process known from classical team work psychology, experts are used to provide information for the final model of the team facilitator and acting as „supporting utilities“
    - Level 4 – experts/ or the facilitator are the „owner“ of the results, introduces large interdependence between expert judgements with a tendency to overly conservative results

Development of Structured Expert Judgement –V – PSHA-applications, SSHAC-procedures

- The SSHAC procedures deviate to a large extend from the methodological development of expert elicitation in PRA applications as well as from the procedural guidance presented here
- Main deviations
  - Lack of a normative expert (specialist in treating subjective probability issues) in the process
  - Different interchanging roles of experts and team facilitators/integrators
  - Lack of empirical control although it would be allowed to be performed
  - No calibration, although different weighing procedures may be allowed
  - Aggregation of results is generally not performance-based
  - Lack of accountability for level 4 procedures (if all experts own the results, nobody is responsible)
  - Narrow focus to one mathematical methodology for earthquake recurrence (assumption of an ergodic stochastic process, characteristic magnitudes (is just the contrary of an ergodic process)) limits the freedom of experts to use other methods

General Procedural Guidelines –I, Uncertainty, expert judgement and technical decision making

- Expert elicitation is getting popular in PRA applications because of the need to deal with uncertainties
- „Uncertainty is that which is removed by becoming certain ...certainty is achieved through observation, and uncertainty is that, what is removed by observation. Hence uncertainty is concerned with the results of possible observations.“(Cooke & Goossens,2000)

→ Expert Judgement can and shall be an intermediate state in the quest for true answers, helping to rank problems by their importance, results shall be updated based on new information

General Procedural Guidelines –II, Uncertainty, expert judgement and technical decision making

- Uncertainty in conjunction with expert judgement represents a different degree of belief in the truth of some statements
- The degree of belief is different for different persons, there is no way to persuade a person to adopt the degree of belief of other persons
- The degree of belief can be expressed mathematically by (subjective) probability or by fuzzy set theory (possibility)
- Technical decision making is attempting to be rational, based on facts, therefore there is the objective requirement to validate the quantified belief of „experts“ and to apply a set of rules leading to a rational consensus

General Procedural Guidelines –III, Principles for Rational Consensus (Cooke, 1991)

- Principle 1 – Reproducibility
  - „It must be possible for scientific peers to review and if necessary to reproduce all calculations. This entails, that the calculational models must be fully specified and the ingredient data must be made available“.
  - Essential element of any scientific analysis
  - Often not fulfilled
    - Proprietary, Confidentiality, some times impression that reproducibility and comparability is intentionally avoided (fear of scientists to be compared with others)

General Procedural Guidelines –IV, Principles for Rational Consensus (Cooke, 1991)

- Principle 2 Accountability:
  - „The source of expert subjective probabilities must be identified“.
  - Accountability means in the context of using expert opinion, that the decision maker can trace every subjective probability to the name of the person or institution from which it comes.
  - In case of decisions of great importance accountability shall be expanded to legal responsibility.
    - Can be difficult in case of using foreign experts, which are “leaving“ the country after the project.

General Procedural Guidelines –V, Principles for Rational Consensus (Cooke, 1991)

- Principle 3: Empirical Control
  - „Expert probability assessments must in principle be susceptible to empirical control“
  - Empirical control ensures that the use of subjective probabilities cannot be construed as a license for the expert to say anything whatever.
  - For critical decision making a formal validation of the results shall be required.
  - For seismic hazard analysis such approaches have been developed

General Procedural Guidelines –VI, Principles for Rational Consensus (Cooke, 1991)

- Principle 4 Neutrality
  - „The method for combining/evaluating expert opinion shall encourage experts to state their true belief“
  - Experts shall not be rewarded to join a common group opinion
  - Expert assessments can be performed independently – reduces the effort of the expert elicitation process (smaller number of workshops)
  - Implies the use of calibration or performance based methods for the aggregation of the results, each expert will be judged how he performed objectively

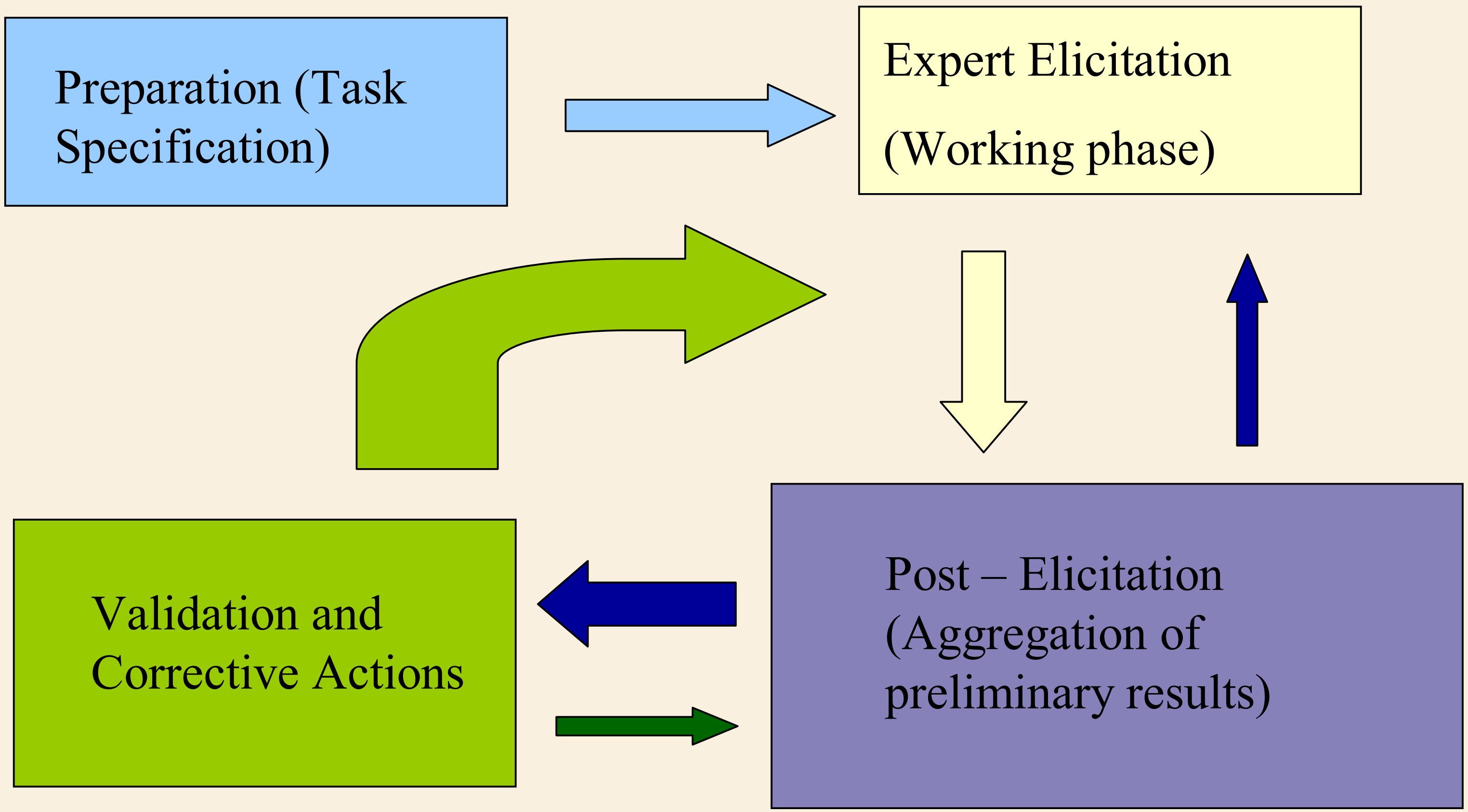
General Procedural Guidelines –VII, Principles for Rational Consensus (Cooke, 1991)

- Principle 5 Fairness:
  - „All experts are treated equally, prior to processing the results of observations“.
  - Since empirical control and validation of results is accepted as the means for evaluation expert opinions, there is no justification to prefer one expert to another (in absence of empirical information) besides his true performance

General Procedural Guidelines –VIII – a Structured Expert Elicitation Process

- A structured expert elicitation process can be subdivided into four phases
  - Preparation (detailed task specification)
  - Elicitation (active working phase)
  - Post-Elicitation
  - Validation of results and corrective actions
- In dependence of the scale of the project phase 3 and 4 can be combined

Four Phases of a Structured Expert Elicitation Process



Phase I – Preparation of Expert Elicitation -I

- Step 1 – Definition of project organisation
  - Clear allocation of tasks required
  - Decision on involvement of the client/sponsor of the study to be made
- Step 2 – definition of a case structures document describing the field of interest for which expert judgement is required
- Step 3 Identification of target variables
- Step 4 Definition of query variables
  - Necessary if target variables are not measurable (observable), rarely identical with target variables

Phase I – Preparation of Expert Elicitation -II

- Step 5 – Development of the methodology for the aggregation of expert opinion
  - Preferably include a calibration procedure/ or a performance based weighing procedure
  - Alternative – Use of Inverse Probability Functions
  - Bayesian techniques
- Step 6 – Identification of experts in the field
  - Nomination procedure
- Step 7 – Preliminary selection of experts

Phase I – Preparation of Expert Elicitation -III

- Step 8 – Definition and preparation of input information
  - Detailed specification of the exact questions and the format of expert elicitations
- Step 9 – „Table-top exercise“ to test the developed procedures with a few experts not involved in the project
- Step 10 – Expert training session
  - Detailed description of query variables and used procedures communicated to the experts

Phase II – Elicitation

- Step 11 Expert Information Workshops
  - Assure an equally informed state of knowledge of involved experts
  - Examination test for experts
    - May lead to disqualification of experts
  - Participation of safety analysts of the client !!
    - To communicate the methodology and the basic information also to the client

Phase II – Elicitation - II

- Step 12 Expert Elicitation workshops
  - Presence of a normative and substantive expert
  - For seismic hazard analysis – substantive expert shall be from the region
  - Use of prepared and carefully checked questions
  - Avoid strong interdependence between experts and group effects
    - Experts shall express their own opinion/belief on the issue
    - They shall be ranked preferably by their performance
    - Substantive expert shall only moderate the process and assure that the background information is understood correctly

Phase III – Post - Elicitation

- Step 13 – Combination of experts‘ assessments according the agreed aggregation procedure
  - Develop weights from calibration and/or performance tests (on seed variables)
  - Bayesian combination of the results
  - First Quantification of target parameters
- Step 14 Sensitivity /robustness studies, discrepancy analysis
  - Identify the parameters with the largest impact on the results

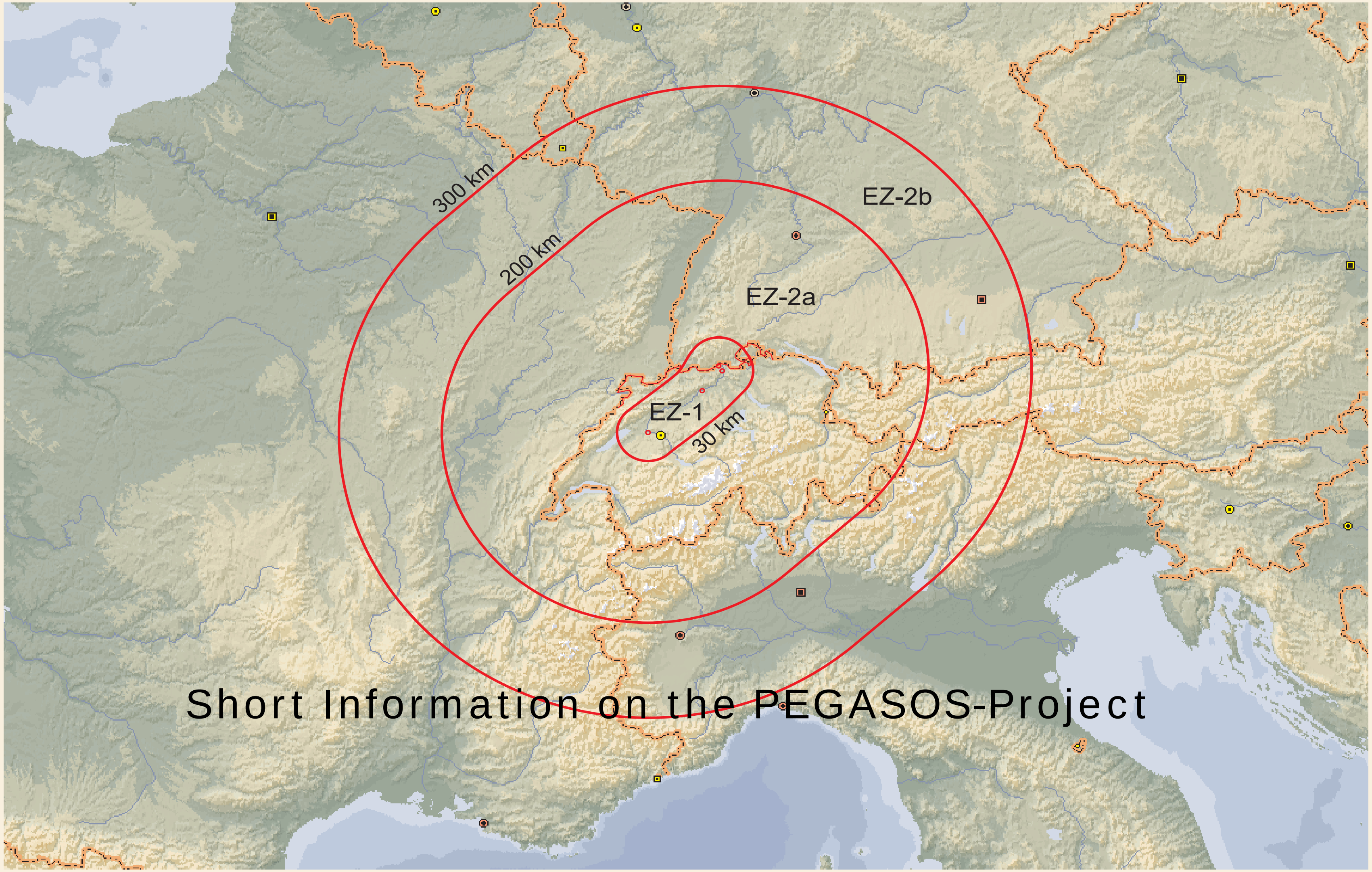
Phase III – Post Elicitation II

- Step 15 – Feedback Workshops,
  - Discussing preliminary results of the study with experts
  - Explanatory discussions on items, where the largest deviations were observed
  - Resolution of possible misinterpretation of technical background information
  - Reduced scope for a performance based elicitation process
- Step 16 – Post processing analysis (uncertainty distributions of target variables)
- Step 17 – Documentation of the preliminary results

Phase IV – Validation (Task of the Client)

- Step 18 – Development of validation procedures (benchmark tests)
- Step 19 – Extensive testing of the study results against the benchmark tests
- Step 20 – Feedback discussions with experts and the substantive analyst
- Step 21 Publication of the final results

PSHA - PEGASOS



Short Information on the PEGASOS-Project

Key Problems of Practical Implementation – Project Organisation

Project organisation

- Allocation of responsibilities
- Involvement of the final customer/client
- Specification of results
- Organisation of the work process /Schedule
- Interference of participatory review teams with the project

Key Problems– Project Organisation II – Role of experts, Example SSHAC

- SSHAC - Intention to represent the state of knowledge of the technically informed community
  - Practical limitations – not all best –suitable experts are available for such a project (other obligations)
  - Not sufficient independent experts in small countries – foreign experts may not be accountable for their assessments
  - Preferable selection of PSHA-experts (closed family), other disciplines underestimated

Key Problems– Project Organisation III – Role of experts, Example SSHAC

- SSHAC – only one class of expert (in geoscience, no normative expert) with interchanging roles
  - Role of the proponent of a specific scientific position or model
  - Role of an evaluator of the various positions in the technically informed community
  - The role of a technical expert providing advice to the TFI
  - The role of the TFI (eliciting uncertainties, building the model)

Key Problems– Project Organisation IV – Role of experts, Example SSHAC

- Interchanging role of experts causes a psychological difficult situation
  - Scientists tend to defend their models
  - SSHAC is oriented towards team consensus
  - Leads to a tendency for convergence on enveloping conservative assumptions
    - People don‘t want to ‚loose their face‘

→ Team position shall envelope the position of each expert = conservative envelope

Key Problems– Project Organisation V – Participation of Customer, Project Organisation, Specification

- It is difficult to communicate final results to the customer without his active participation
  - Total documentation of the PEGASOS-project consists of 60 Gbyte of data, time to review - several years
- Need to include the final customer in all main technical issues and decisions
- Example – PEGASOS – critical decisions were made by TFIs
  - Reference rock shear wave velocity, generic shear wave velocity profiles used for scaling, use of crude scaling laws to adjust spectral ordinates to other conditions
  - Candidate attenuation equations (majority not from Europe)

without discussions with the client – decisions now regarded as questionable

Key Problems– Project Organisation VI – Participation of customer, Project Organisation, Specification

- Customer shall provide a clear project specification
  - Shall be discussed with the experts
  - Shall not be modified during the project course
- Example PEGASOS
  - Lack of communication on the specification
    - Experts (including TFIs) did not understand, that a PSHA input for a PRA for a short-lived structure is different from an input for the definition of a design hazard for a long-term repository (Yucca Mountain)

$$P\left(A\right) \neq P\left(A \middle| B\right)$$

    - Experts did not understand that pga for an engineer (this was specified) has the meaning of an EGA and not of a spike instrumental spectral acceleration (zero period acceleration)

Key Problems– Project Organisation VII – Participatory Review Team

- SSHAC allows for a participatory review, reasonable idea
  - Lesson from PEGASOS - This review shall be performed by independent experts of the customer, not by regulators
  - In small countries lack of independent experts
    - PEGASOS project – head of the Swiss seismological service was a member of the review team of the regulator, his staff members did participate in the project
    - Critical question if results are looking not reasonable

Key Problems– Project Organisation VIII – Ownership of the results

- SSHAC-procedures level IV
  - Ownership by the experts or by the TFIs
  - PEGASOS – by experts, very strong interdependence (strong correlation)
- Insufficient approach
  - Goal – Ownership shall be taken over by the final client of the study results, he shall be able not only to accept the results but to achieve a personal identification with the outcome of the study
  - Client may have to invest hundreds of millions Euro due to the results of the study

Key problems – Treatment of uncertainties and validation of results

- Key problems in this area are:
  - Treatment of correlation
  - Convergence to reality, validation procedures
  - Seismic Hazard Analysis – Minimisation of the number of random model parameters
- Correlation
  - Physical parameters and their uncertainty are themselves correlated
  - Expert assessments are correlated

Key Problems– Treatment of uncertainties and validation of results

- Example for Correlation:
  - PEGASOS – Catalogue calibrated with respect to the Basel earthquake from 1356
  - Assessment of PEGASOS –experts, M_W=6.9
  - Earlier assessments – M_w=6.5 (6.4)
  - French catalogue – M_w=5.9 – Factor 2 difference in ground motion
  - Attenuation laws in PEGASOS „checked“ against a French earthquake St. Die – M_w=4.8 (M_L=5.1), first Swiss assessment was M_L=5.5
  - Attenuation laws are not independent from the assessment of the magnitude of the event, attenuation laws developed based on low assessments of the magnitudes shall not be combined with source characteristic models leading to high assessments of magnitudes

Key Problems– Treatment of Uncertainties and Validation of Results, Convergence

- Modern PSHA is based on logic trees,
- Results depend on the number of branches
- Result can easily be skewed up by just adding uncertain (random) parameters

$$H(f) = \sum_{i=1}^{k^{n-1}} \prod_{j=1}^{n} w_{i,j} F\left(f, w_{i,j}\right)$$

H(f) – spectral acceleration for a fixed frequency of exceedance f , (one ordinate of an UHS), w_{ij} -weighting factor (probability), k maximum number of possible branch states, n- number of branches, F – assignment law

Key Problems– Treatment of Uncertainties and Validation of Results, Convergence II

- Adding experts (more than 1) into the analysis process leads to at least one additional branch
- PEGASOS – 3 additional branches due to the separation of the project into three subprojects with separate expert judgement
- Separate treatment of epistemic and aleatory uncertainties leads to additional branches
- Some experts even developed distributions for the uncertainty of their uncertainty assessments
  - Additional branches

Key Problems– Treatment of Uncertainties and Validation of Results, Convergence III

- Conclusion of the review analysis– the hazard model shall be based on a minimal set of random parameters,
- the number of „effective“ experts shall be reduced f.e. by developing a performance –based decision maker model
- Alternative - the use of Bayesian techniques
- Use of inverse probability techniques (avoid logic trees)

Calibration / Performance-Based Weighting

- Methodology developed by Prof. Cooke (1991), Technical University of Delft
  - Different forms in use, calibration was considered as a possible way to develop weighting factors during the NUREG-1150 ff. studies, but not used
  - Based on the concept of measuring information (entropy measure) content (classical approach)
  - General assumption – The performance of experts on seed variables (test cases) is a measure of their performance on query/target variables
  - Very strong assumption

Classical Method

„The expert is well calibrated if the true values of the uncertain quantities can be regarded as independent samples of a random variable with distribution P“ – problem approximated by χ^2-test of goodness of fit for RI=2NI(S,P)

Relative information of S with respect to P

Measure of surprise, which someone would experience expecting P and obtaining S

$$I(S,P)=\sum_{i=1}^{N} S(i)\ln\left(\frac{S(i)}{P(i)}\right)$$

$$\mathrm{Pr}ob\left\{I(S',P)\geq I(S,P)\,\middle|\,Cal(P),n_{observations}\right\}$$

Calibration / Performance-Based Weighting II

- Calibration alone could be used to develop weighting factors, the classical approach combines it with an additional performance measure

$$w_e = \frac{Cal(e)\Delta_{cal}\left(Cal(e)\right)}{\sum_e w_e} I(e)$$

$$\Delta_{cal}(x) = 1 : x \geq \alpha$$

$$\Delta_{cal}(x) = 0 : x < \alpha, \alpha \in \in (0,1)$$

$$I(e) = 1 / H_e(p)$$

$$H(P) = \sum_{i=1}^{B} n_i H(P_i)$$

$$H(P_i) = -P_i \ln(P_i)$$

Average information I_e = inverse response entropy (discrete probability distribution associated with B bins, normalized, n the number of items in each bin)

Measuring Information - Gain of Expert Judgement

- Weighted Expert Mean of the total number of earthquakes shall deviate from catalogue data not more then by the variance of the catalogue data

$$M_e^2 - 10^{2a_{cat}} \leq \sigma_{a,cat}^2$$

The entropy of the expert assessments shall be below/equal the entropy of the catalogue data, otherwise the expert judgements shall be removed from analysis

$$H_e\left(N\right) \leq S_{cat}\left(N\right)$$

Alternative approaches – Bayesian Techniques for Aggregation

- Suitable tool for Probabilistic Seismic Hazard Analysis
  - Distribution of spectral ordinates (for 5% damping,pga) for a fixed frequency of exceedance is approximately lognormal (Klügel&Groen, 2004)
  - This knowledge can be utilized to use Bayesian aggregation techniques developed for Probabilistic Risk Analysis
    - S.Kaplan (1983) – Two stage Bayesian Approach
    - Expert Judgement results can be combined with other information
  - Elimination of „improper“ branches in the logic tree

Alternative approaches – Bayesian Techniques for Aggregation

$$H(f) = \sum_{i=1}^{k^{n-1}} \prod_{j=1}^{n} w_{i,j} F\left(f, w_{i,j}\right)$$

PEGASOS – Subdivision into 3 subprojects led to 3 additional branches for experts in the compiled logic tree

Knowledge of the shape of the distribution allows to consider a reduced set of combinations – splitting the large logic tree into a subset of 80 logic trees and correspondingly into 80 expert opinions with 80 * n (number of spectral ordinates) lognormal distributions

Case Study – PSHA for an NPP, pga for SL2 (10-4 frequency of exceedance)

Expert	Median	Mean	Range Factor (K_95%)
1	0.15	0.187	3.0
2	0.19	0.289	4.5
3	0.25	0.292	2.5
4	0.27	0.361	3.5
5	0.8	1.14	4.0
6	0.35	0.437	3.0
7	0.27	0.328	2.8
8	0.23	0.307	3.5
Prior Study 1	0.167	0.21	3.0
Study 2	0.32	0.4	3.0

Case Study – PSHA for an NPP, pga for SL2 (10-4 frequency of exceedance) -II

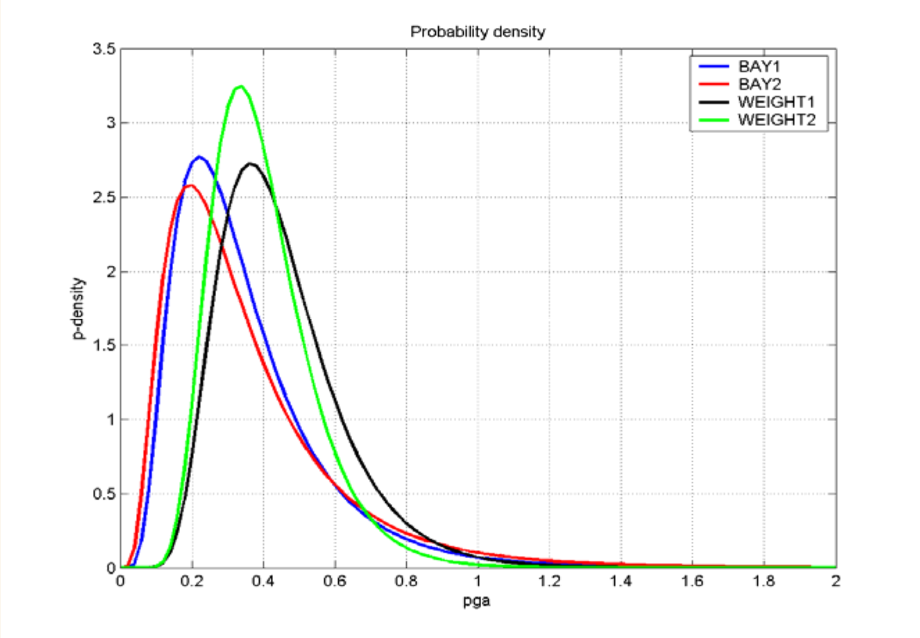
- Comparison of following procedures
  - Direct Bayesian aggregation – BAY1
  - Equal weight for experts only-WEIGHTI
  - Equal weight for experts and previous studies-WEIGHT II
  - Bayesian aggregation after elimination of outliers (out of the interquartile range – IQR (0.25,0.75))-BAYII
- Bayesian Aggregation with RISKMAN®

Case Study – PSHA for an NPP, pga for SL2 (10-4 frequency of exceedance) -III

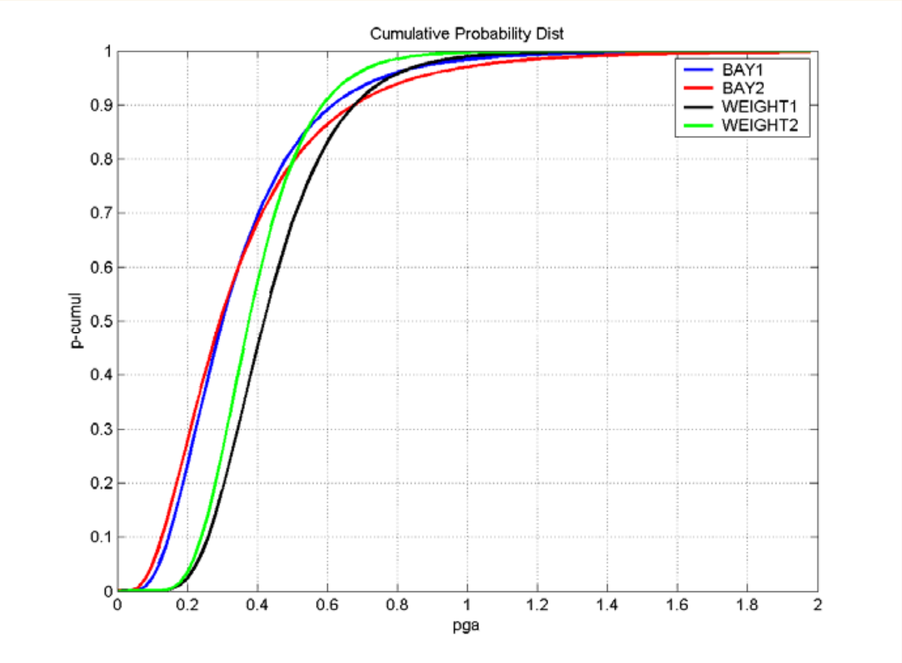
Method	5% [g]	50% [g]	95% [g]	Mean [g]
WEIGHT-I	0.232	0.384	0.711	0.419
BAY-I	0.101	0.259	0.649	0.301
WEIGHT-II	0.217	0.35	0.619	0.376
BAY-II	0.0779	0.242	0.703	0.294

Case-Study -Comparison Of Results

Probability Density Distribution



Cumulative Distribution



Bayesian approach leads to smaller mean values maintaining proper uncertainty bounds

Conclusions -I

- Expert opinion can be a valuable source of scientific information if obtained in a structured process
- Expert elicitation shall be based on a rational consensus approach
- Guidelines developed by the TU Delft (Prof. Cooke) and expanded by this paper are helpful to organise an structured expert elicitation project
- SSHAC-procedures (SSHAC, 1997) do not comply with these guidelines and potentially lead to irrational results (Return of ‚Buridans‘ ass dilemma)

Conclusions -II

- Possible improvements to SSHAC-procedures
  - Provide empirical control
  - Use performance-based/ calibration based weighting or Bayesian aggregation with elimination of outliers
  - Introduction of an additional normative expert
  - Reduce interdependence between experts (less workshops)
  - Take into account physical dependencies between model variables by an appropriate project organisation (avoid „split-of" of the analysis task)
  - Utilize the principle of minimisation of the number of random model parameters

References

- Complete List of references in the „long“ paper version of the lecture
- Excellent Book of R. Cooke „Experts in Uncertainty“, 1991.