IAEA
International Atomic Energy Agency

IAEA TECDOC SERIES

No. 2067

Evaluation of Probabilistic Seismic Hazard Analysis (PSHA) for Nuclear Installations Based on Observational Data

IAEA SAFETY STANDARDS AND RELATED PUBLICATIONS

IAEA SAFETY STANDARDS

Under the terms of Article III of its Statute, the IAEA is authorized to establish or adopt standards of safety for protection of health and minimization of danger to life and property, and to provide for the application of these standards.

The publications by means of which the IAEA establishes standards are issued in the **IAEA Safety Standards Series**. This series covers nuclear safety, radiation safety, transport safety and waste safety. The publication categories in the series are **Safety Fundamentals**, **Safety Requirements** and **Safety Guides**.

Information on the IAEA's safety standards programme is available at the IAEA Internet site

www.iaea.org/resources/safety-standards

The site provides the texts in English of published and draft safety standards. The texts of safety standards issued in Arabic, Chinese, French, Russian and Spanish, the IAEA Safety Glossary and a status report for safety standards under development are also available. For further information, please contact the IAEA at: Vienna International Centre, PO Box 100, 1400 Vienna, Austria.

All users of IAEA safety standards are invited to inform the IAEA of experience in their use (e.g. as a basis for national regulations, for safety reviews and for training courses) for the purpose of ensuring that they continue to meet users' needs. Information may be provided via the IAEA Internet site or by post, as above, or by email to Official.Mail@iaea.org.

RELATED PUBLICATIONS

The IAEA provides for the application of the standards and, under the terms of Articles III and VIII.C of its Statute, makes available and fosters the exchange of information relating to peaceful nuclear activities and serves as an intermediary among its Member States for this purpose.

Reports on safety in nuclear activities are issued as **Safety Reports**, which provide practical examples and detailed methods that can be used in support of the safety standards.

Other safety related IAEA publications are issued as **Emergency Preparedness and Response** publications, **Radiological Assessment Reports**, the International Nuclear Safety Group's **INSAG Reports**, **Technical Reports** and **TECDOCs**. The IAEA also issues reports on radiological accidents, training manuals and practical manuals, and other special safety related publications.

Security related publications are issued in the **IAEA Nuclear Security Series**.

The **IAEA Nuclear Energy Series** comprises informational publications to encourage and assist research on, and the development and practical application of, nuclear energy for peaceful purposes. It includes reports and guides on the status of and advances in technology, and on experience, good practices and practical examples in the areas of nuclear power, the nuclear fuel cycle, radioactive waste management and decommissioning.

EVALUATION OF PROBABILISTIC SEISMIC HAZARD ANALYSIS (PSHA) FOR NUCLEAR INSTALLATIONS BASED ON OBSERVATIONAL DATA

The following States are Members of the International Atomic Energy Agency:

AFGHANISTAN
ALBANIA
ALGERIA
ANGOLA
ANTIGUA AND BARBUDA
ARGENTINA
ARMENIA
AUSTRALIA
AUSTRIA
AZERBAIJAN
BAHAMAS
BAHRAIN
BANGLADESH
BARBADOS
BELARUS
BELGIUM
BELIZE
BENIN
BOLIVIA, PLURINATIONAL STATE OF
BOSNIA AND HERZEGOVINA
BOTSWANA
BRAZIL
BRUNEI DARUSSALAM
BULGARIA
BURKINA FASO
BURUNDI
CABO VERDE
CAMBODIA
CAMEROON
CANADA
CENTRAL AFRICAN REPUBLIC
CHAD
CHILE
CHINA
COLOMBIA
COMOROS
CONGO
COSTA RICA
CÔTE D'IVOIRE
CROATIA
CUBA
CYPRUS
CZECH REPUBLIC
DEMOCRATIC REPUBLIC OF THE CONGO
DENMARK
DJIBOUTI
DOMINICA
DOMINICAN REPUBLIC
ECUADOR
EGYPT
EL SALVADOR
ERITREA
ESTONIA
ESWATINI
ETHIOPIA
FIJI
FINLAND
FRANCE
GABON
GAMBIA
GEORGIA
GERMANY
GHANA
GREECE
GRENADA
GUATEMALA
GUINEA
GUYANA
HAITI
HOLY SEE
HONDURAS
HUNGARY
ICELAND
INDIA
INDONESIA
IRAN, ISLAMIC REPUBLIC OF
IRAQ
IRELAND
ISRAEL
ITALY
JAMAICA
JAPAN
JORDAN
KAZAKHSTAN
KENYA
KOREA, REPUBLIC OF
KUWAIT
KYRGYZSTAN
LAO PEOPLE'S DEMOCRATIC REPUBLIC
LATVIA
LEBANON
LESOTHO
LIBERIA
LIBYA
LIECHTENSTEIN
LITHUANIA
LUXEMBOURG
MADAGASCAR
MALAWI
MALAYSIA
MALI
MALTA
MARSHALL ISLANDS
MAURITANIA
MAURITIUS
MEXICO
MONACO
MONGOLIA
MONTENEGRO
MOROCCO
MOZAMBIQUE
MYANMAR
NAMIBIA
NEPAL
NETHERLANDS, KINGDOM OF THE
NEW ZEALAND
NICARAGUA
NIGER
NIGERIA
NORTH MACEDONIA
NORWAY
OMAN
PAKISTAN
PALAU
PANAMA
PAPUA NEW GUINEA
PARAGUAY
PERU
PHILIPPINES
POLAND
PORTUGAL
QATAR
REPUBLIC OF MOLDOVA
ROMANIA
RUSSIAN FEDERATION
RWANDA
SAINT KITTS AND NEVIS
SAINT LUCIA
SAINT VINCENT AND THE GRENADINES
SAMOA
SAN MARINO
SAUDI ARABIA
SENEGAL
SERBIA
SEYCHELLES
SIERRA LEONE
SINGAPORE
SLOVAKIA
SLOVENIA
SOUTH AFRICA
SPAIN
SRI LANKA
SUDAN
SWEDEN
SWITZERLAND
SYRIAN ARAB REPUBLIC
TAJIKISTAN
THAILAND
TOGO
TONGA
TRINIDAD AND TOBAGO
TUNISIA
TÜRKİYE
TURKMENISTAN
UGANDA
UKRAINE
UNITED ARAB EMIRATES
UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND
UNITED REPUBLIC OF TANZANIA
UNITED STATES OF AMERICA
URUGUAY
UZBEKISTAN
VANUATU
VENEZUELA, BOLIVARIAN REPUBLIC OF
VIET NAM
YEMEN
ZAMBIA
ZIMBABWE

The Agency's Statute was approved on 23 October 1956 by the Conference on the Statute of the IAEA held at United Nations Headquarters, New York; it entered into force on 29 July 1957. The Headquarters of the Agency are situated in Vienna. Its principal objective is "to accelerate and enlarge the contribution of atomic energy to peace, health and prosperity throughout the world".

IAEA-TECDOC-2067

EVALUATION OF PROBABILISTIC SEISMIC HAZARD ANALYSIS (PSHA) FOR NUCLEAR INSTALLATIONS BASED ON OBSERVATIONAL DATA

INTERNATIONAL ATOMIC ENERGY AGENCY
VIENNA, 2024

COPYRIGHT NOTICE

All IAEA scientific and technical publications are protected by the terms of the Universal Copyright Convention as adopted in 1952 (Geneva) and as revised in 1971 (Paris). The copyright has since been extended by the World Intellectual Property Organization (Geneva) to include electronic and virtual intellectual property. Permission may be required to use whole or parts of texts contained in IAEA publications in printed or electronic form. Please see www.iaea.org/publications/rights-and-permissions for more details. Enquiries may be addressed to:

Publishing Section
International Atomic Energy Agency
Vienna International Centre
PO Box 100
1400 Vienna, Austria
tel.: +43 1 2600 22529 or 22530
email: sales.publications@iaea.org
www.iaea.org/publications

For further information on this publication, please contact:

External Events Safety Section
International Atomic Energy Agency
Vienna International Centre
PO Box 100
1400 Vienna, Austria
Email: Official.Mail@iaea.org

© IAEA, 2024
Printed by the IAEA in Austria
September 2024
https://doi.org/10.61092/iaea.jfkj-gjuz

IAEA Library Cataloguing in Publication Data

Names: International Atomic Energy Agency.
Title: Evaluation of probabilistic seismic hazard analysis (PSHA) for nuclear installations based on observational data / International Atomic Energy Agency.
Description: Vienna : International Atomic Energy Agency, 2024. | Series: IAEA TECDOC series, ISSN 1011-4289 ; no. 2067 | Includes bibliographical references.
Identifiers: IAEAL 24-01708 | ISBN 978-92-0-130624-1 (paperback : alk. paper) | ISBN 978-92-0-130724-8 (pdf)
Subjects: LCSH: Earthquake hazard analysis. | Nuclear facilities — Earthquake effects. | Hazard mitigation.

FOREWORD

IAEA Safety Standards Series No. SSG-9, Seismic Hazards in Site Evaluation for Nuclear Installations, was first published in 2010 and No. SSG-9 (Rev. 1) was published in 2022. The revision incorporates the changes in the state of the practice in seismic hazard assessments over the previous decade. One of the key changes is the inclusion of a recommendation to use available observations and data from actual seismic events in each step of the seismic hazard assessment process and in the overall evaluation of the results. This publication complements SSG-9 (Rev. 1) by providing detailed technical information on evaluating probabilistic seismic hazard analysis inputs and outputs, both during the development stages of the probabilistic seismic hazard analysis and on the completion of the analysis.

The technical information and practical descriptions provided here will be valuable to nuclear power plant operating organizations, regulatory bodies, vendors, technical support organizations and researchers working in the field of seismic hazard assessment for existing nuclear installations. This publication will also support seismic hazard assessments to be carried out for future nuclear installations.

The contributions of all those involved in the drafting and review of this publication are greatly appreciated, in particular E. Viallet (France) for coordinating the drafting efforts. The IAEA officers responsible for this publication were Y. Fukushima and Z. Gulerce of the Division of Nuclear Installation Safety.

EDITORIAL NOTE

This publication has been prepared from the original material as submitted by the contributors and has not been edited by the editorial staff of the IAEA. The views expressed remain the responsibility of the contributors and do not necessarily represent the views of the IAEA or its Member States.

Guidance and recommendations provided here in relation to identified good practices represent expert opinion but are not made on the basis of a consensus of all Member States.

Neither the IAEA nor its Member States assume any responsibility for consequences which may arise from the use of this publication. This publication does not address questions of responsibility, legal or otherwise, for acts or omissions on the part of any person.

The use of particular designations of countries or territories does not imply any judgement by the publisher, the IAEA, as to the legal status of such countries or territories, of their authorities and institutions or of the delimitation of their boundaries.

The mention of names of specific companies or products (whether or not indicated as registered) does not imply any intention to infringe proprietary rights, nor should it be construed as an endorsement or recommendation on the part of the IAEA.

The authors are responsible for having obtained the necessary permission for the IAEA to reproduce, translate or use material from sources already protected by copyrights.

The IAEA has no responsibility for the persistence or accuracy of URLs for external or third party Internet web sites referred to in this publication and does not guarantee that any content on such web sites is, or will remain, accurate or appropriate.

CONTENTS

1. INTRODUCTION

1.1. BACKGROUND

In order to ensure the safety of nuclear installations against earthquakes, their structures, systems, and components are designed to withstand vibratory ground motions expected to occur at their sites during the life cycle of these facilities. Requirement 16 of IAEA Safety Standards Series No. SSR-1, Site Evaluation for Nuclear Installations [1] states that:

> "**An evaluation of ground motion hazards shall be conducted to provide the input needed for the seismic design or safety upgrading of the structures, systems and components of the nuclear installation, as well as the input for performing the deterministic and/or probabilistic safety analyses necessary during the lifetime of the nuclear installation.**"

Paragraphs 6.8–6.14 of IAEA Safety Standards Series No. SSG-9 (Rev. 1), Seismic Hazards in Site Evaluation for Nuclear Installations [2] recommend probabilistic seismic hazard analysis (PSHA) as one of the methodologies that can be used to estimate vibratory ground motion hazards in site evaluations for nuclear installations. The PSHA methodology explicitly incorporates aleatory and epistemic uncertainties due to imperfect knowledge of the physical processes responsible for the vibratory ground motions, lack of information about the controlling parameters, and incompleteness of data used in the statistical parameterization. The development of a PSHA study utilizes logic trees to accommodate epistemic uncertainties inherent in seismic source characterization and ground motion models. When appropriate, technically justifiable weights are assigned to each branch of the logic tree using objective measures. However, in some instances, subjective weighting schemes based on the consensus of experts' opinions are also needed. An expert elicitation process for a nuclear power plant is thoroughly described in Ref. [3]. The advantages of such an approach are the comprehensive assessments and implementation of epistemic uncertainties and the integration of different experts' judgements in a predictable and formal process.

In PSHA, it is essential to evaluate the input parameters of all logic tree branches and subsequent results for suitability and continued validity. Paragraph 6.13 of SSG-9 (Rev. 1) [2] underlines the importance of verifying individual parameters and models in the integration scheme of the PSHA by stating that:

> "Because of the uncertainties, mainly of an epistemic nature, that are involved at each stage of the hazard assessment process, both the assumptions adopted in previous steps and the overall results obtained from the analysis should be evaluated based on available observational data and data from actual seismic events, with due consideration of the difference between the short period of data availability and the return period usually adopted for seismic design of nuclear installations. This evaluation should be used to check either the consistency of the assumptions or the adequacy of the defined branch of the logic tree, or to assign proper weight in the logic tree."

Similar evaluation procedures are described for the safety assessment of nuclear installations. Requirement 18 of IAEA Safety Standards Series No. GSR Part 4 (Rev. 1), Safety Assessment for Facilities and Activities [4] states that "**Any calculational methods and computer codes used in the safety analysis shall undergo verification and validation.**"

Paragraph 6.23 of IAEA Safety Standards Series No. SSG-2 (Rev. 1), Deterministic Safety Analysis for Nuclear Power Plants [5] states that "**Code to data comparisons are the preferred means to quantify the epistemic uncertainties**" and that "**The preferred means for assessing aleatory uncertainties is the collection of data from nuclear power plants on initial and boundary conditions that are relevant to the events being considered.**"

Paragraphs 5.3 and 5.4 of SSG-2 (Rev. 1) [5] recommend that the predictions of the computer codes used for deterministic safety analysis are confirmed with the experimental data for the significant phenomena modelled. These include the comparisons with 'separate effect tests' and 'integral effect tests'. The definitions of separate effect tests and integral effect tests are in parallel with the elementary and integral evaluation concepts described in this publication.

Similarly, alternative methods for evaluating PSHA studies conducted for non-nuclear facilities using observational data were developed by the scientific and engineering communities in the last decade. A brief discussion of these methods is introduced in Ref. [6]. In general, PSHA studies conducted for non-nuclear projects might not be applicable for nuclear installation sites as they sometimes lack the level of reliability, or the details needed to demonstrate nuclear installation safety. This TECDOC aims to clarify the PSHA evaluation methods for nuclear installation sites and the type of data to be used in the evaluation process both during the development of the PSHA, as well as after it has been completed. This publication also underlines the need for acquiring data from seismic monitoring systems (see paras 3.54-3.59 of SSG-9 (Rev. 1) [2], by describing the ways of using collected data for the evaluation and assessment of PSHA results even after the site characterization stage.

An important prerequisite for a PSHA to be evaluated through the methodology presented in this publication is that the PSHA is conducted using an appropriate level of quality assurance expected for a nuclear installation. In cases where a PSHA study does not comply with the quality standards expected for a nuclear installation, the value of conducting extensive evaluation studies might not provide adequate results. However, an evaluation study can still be conducted for overall comparative purposes.

1.2. OBJECTIVE

The objective of this technical publication is to support the recommendations provided in SSG-9 (Rev. 1) [2] on vibratory ground motion hazard analysis, and the way to meet the relevant requirements in SSR-1 [1], by providing information on state-of-the-practice and detailed technical elements relating to the evaluation of PSHA on the basis of observational data. This publication:

- Defines the methodology and provides practical guidance for evaluating the elementary and integral steps of PSHA with the help of examples from case studies available in the current literature;
- Describes the type of observational data that may be used in the evaluation;
- Discusses ways of using the evaluation results to check the centre, body, and range of the uncertainties (as defined in Ref. [3]), as well as to encourage PSHA developers to update or reconsider their own choices by providing feedback to avoid including any unrealistic assumptions or branches in the logic tree.

This publication does not describe how to perform a probabilistic seismic hazard analysis, which is already addressed in SSG-9 (Rev. 1) [2], as well as in other IAEA and external publications. The methodology described here may be considered as an additional or a complementary step to be applied in the framework of a PSHA, either during its implementation or at a subsequent stage.

1.3. SCOPE

The scope of this publication is the evaluation of inputs and outputs of a PSHA conducted for site evaluation of a nuclear installation based on observational data. The methodology described in this publication may be implemented in the review and assessment of vibratory ground motions for new or existing nuclear power plants and other nuclear installations in any seismotectonic environment. Particular attention is devoted to covering diverse evaluation schemes to reflect the accumulated knowledge and information and the way to process these consistently with the evaluated PSHA.

The technical information and practical descriptions provided here will be valuable to nuclear power plant operating organizations, regulatory bodies, vendors, technical support organizations and researchers working in the field of seismic hazard assessments for existing nuclear installations. This publication will also support seismic hazard assessments to be carried out for future nuclear installations by the IAEA Member States.

1.4. STRUCTURE

Section 2 of this publication discusses the ways that the observational data is utilized in different elementary steps of PSHA, underlining the importance of the consistency of the PSHA inputs with all available observational data. Section 3 overviews the concepts of evaluating PSHA based on observational data. Section 4 presents how to collect and compile the observational dataset needed for the evaluation purposes including elementary and integral evaluations. Section 5 presents details of the evaluation process and its outcomes and provides a sample evaluation review sheet. The Appendix provides further mathematical background and a logical framework in evaluating PSHA studies. The Annex present a few case studies to evaluate the elementary steps or integral results of a PSHA (or of multiple PSHAs).

2. GENERAL CONSIDERATIONS ON THE USE OF AVAILABLE OBSERVATIONAL DATA IN PROBABILISTIC SEISMIC HAZARD ANALYSIS

Seismic source characterization and ground motion characterization models are the two main components of a PSHA study (see Figure 1). The seismic source characterization models include seismic source geometries, earthquake activity rates, and fault rupture models, which incorporate rupture dimensions and orientations, as well as the maximum magnitude estimations for each of the identified seismic sources. The ground motion characterization models define the amplitude of the vibratory ground motions across a range of frequencies at the site, which are based the earthquake scenarios defined by the seismic source characterization models.

The seismotectonic database compiled during the site characterization stage is directly utilized in the development of seismic source and ground motion characterization models as shown in Figure 1. Hence, observational data plays an important role from the beginning. In conducting a PSHA study for a nuclear installation, it is necessary to develop technically defensible seismic source and ground motion characterization models that conform with all available data and knowledge about the region of interest. Thus, prior to conducting a PSHA study for a nuclear installation, great care needs to be taken to collect all available relevant data to be used in the PSHA. In cases where data of a sufficient quality and quantity are used in the development of seismic source and ground motion characterization models, this enhances the reliability of the input models and, in turn, limits the associated epistemic uncertainties.

The difference between the terms ‘available’ and ‘new’ observational data needs to be clearly defined. In this TECDOC, the term ‘new observational data’ is used to describe the information that was not available at the time of site characterization and hence the data was not used in the PSHA. These data sets become critical in assessing the validity of PSHA outputs, such as during the periodic evaluations of seismic hazards at nuclear installations. On the other hand, the term ‘available observational data’ refers to data sets that have already been considered in the PSHA development. It is noteworthy to state that, on specific occasions, not all available and relevant observational data may have been used in the development of a PSHA. In such situations, it is not appropriate to use the PSHA for nuclear installations and a new PSHA will need to be conducted following the recommendations provided in SSG-9 (Rev. 1) [2]. Figure 2 shows a simplified sketch of the timeline for collecting seismotectonic data during the site survey and site evaluation for a nuclear installation. New observational data that is collected after the site evaluation report is finalized may only be used in the evaluation and assessment of an existing PSHA. The evaluation based on new data is defined as a ‘prospective evaluation’.

All data that is collected during the site characterization stage needs to be put in the available observational data category, that can be used in the PSHA in multiple ways. For example, data from macroseismic observations can be used in the assessment of magnitudes and locations of historical earthquakes, which contributes to the development of the seismic source characterization model (Figure 2, red arrow). In this case, the available observational data is used in the PSHA development efforts. The same data can also be useful for other purposes. For example, data from macroseismic observations can be used to evaluate the rates of seismic event occurrences at the site of interest (Figure 2, blue arrows). This type of evaluation, called ‘retrospective evaluation,’ uses available observational data to constrain PSHA outputs.

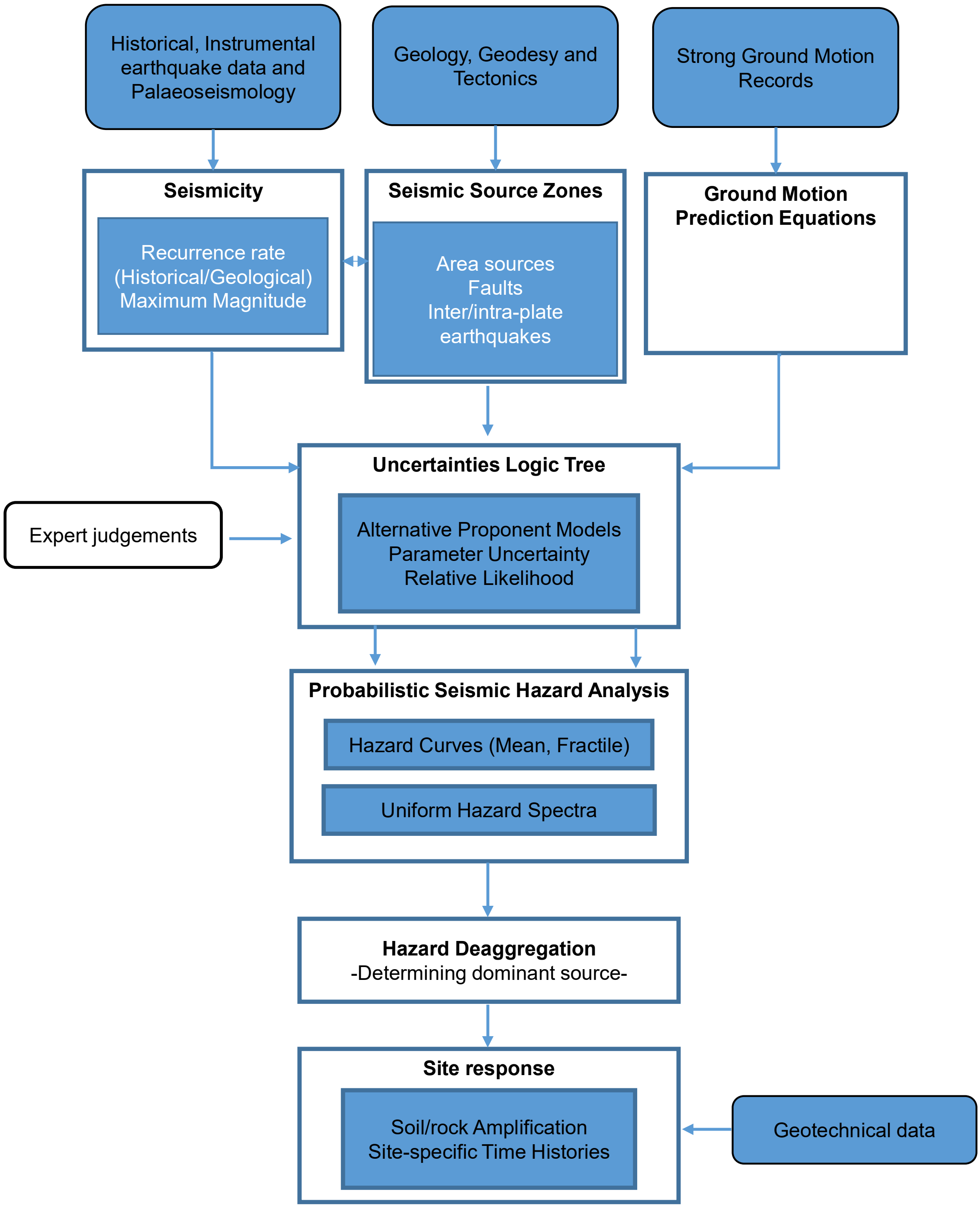

FIG. 1. Components of probabilistic seismic hazard analysis

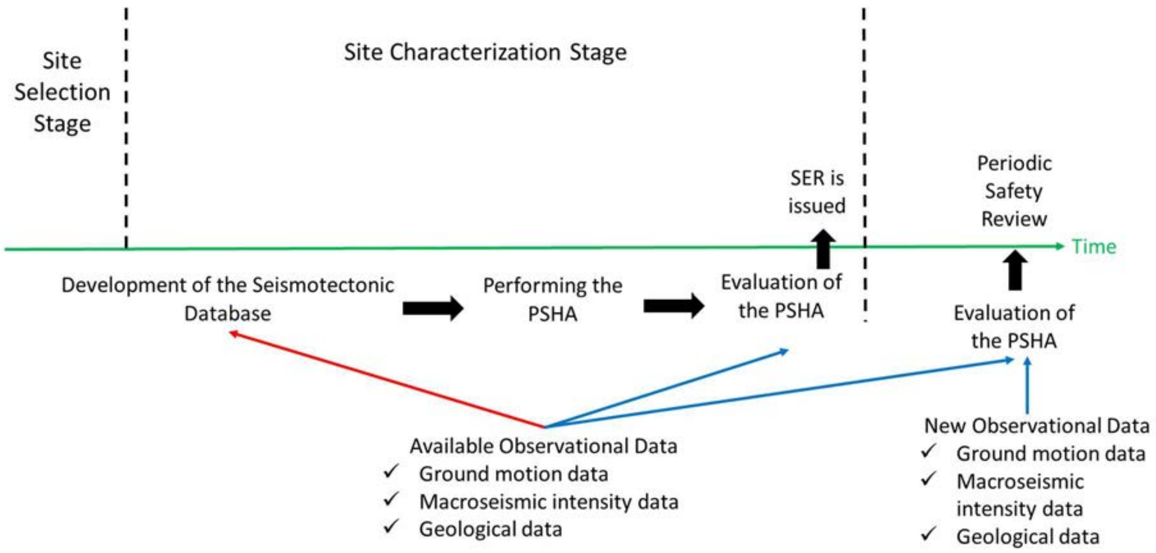

FIG. 2. Timeline of compilation for available and new observational data

Several PSHA input parameters, such as earthquake occurrence rates, maximum magnitude estimations, ground motion models (GMM)[1], etc. have a strong dependence on observational data. Therefore, the current state-of-the-practice for utilizing observational data throughout the model development process during PSHA is elaborated in this section. Later, procedures for using this data in the evaluation process are laid out.

2.1. USE OF OBSERVATIONAL DATA FOR FAULT GEOMETRY AND MAXIMUM MAGNITUDE ASSESSMENTS

Data on historical earthquake occurrences, fault investigations, paleo-seismicity, and seismic records are used to assess the fault geometries and maximum magnitude potential of seismic sources. The key data sets used for this assessment depend on the seismic source type and data availability. Figure 3 presents a flowchart for assessing the maximum magnitude (M_{max}) potential of areal sources, point sources, and 3-dimensional fault sources.

Seismogenic structures as defined in SSG-9 (Rev. 1) [2] are characterized by geological, geomorphological, and geophysical investigations. On the order hand, the assessment of M_{max} values for fault sources typically involves the estimation of the maximum dimensions of a possible rupture, including likely segmentation models, rupture lengths, rupture areas, and displacement per event. Since the rupture dimension parameters are all empirically correlated with earthquake magnitudes (e.g. see Ref. [7]), observational data associated with a particular fault source may be used to evaluate the applicability of magnitude-rupture area scaling relations for that source, especially in high seismicity regions.

---

[1] Historically, empirical prediction of vibratory ground motion was estimated using the ‘attenuation equations’. However, the equation is not only relevant to the seismic wave attenuation but also includes the scaling of ground motions with source and site parameters. Therefore, scientific societies use the term “Ground Motion Prediction Equation (GMPE)” and SSG-9 (Rev. 1) [2] adopted the term GMPE instead of ‘attenuation equation’. Current GMPEs include a large number of parameters to represent the non-ergodicity, and many coefficients in the equations are provided as tables, hence scientific societies use the term “Ground Motion Model (GMM)”. In this publication, the terms of GMPE and GMM are synonymous, and both are used in accordance with the background of the paragraphs.

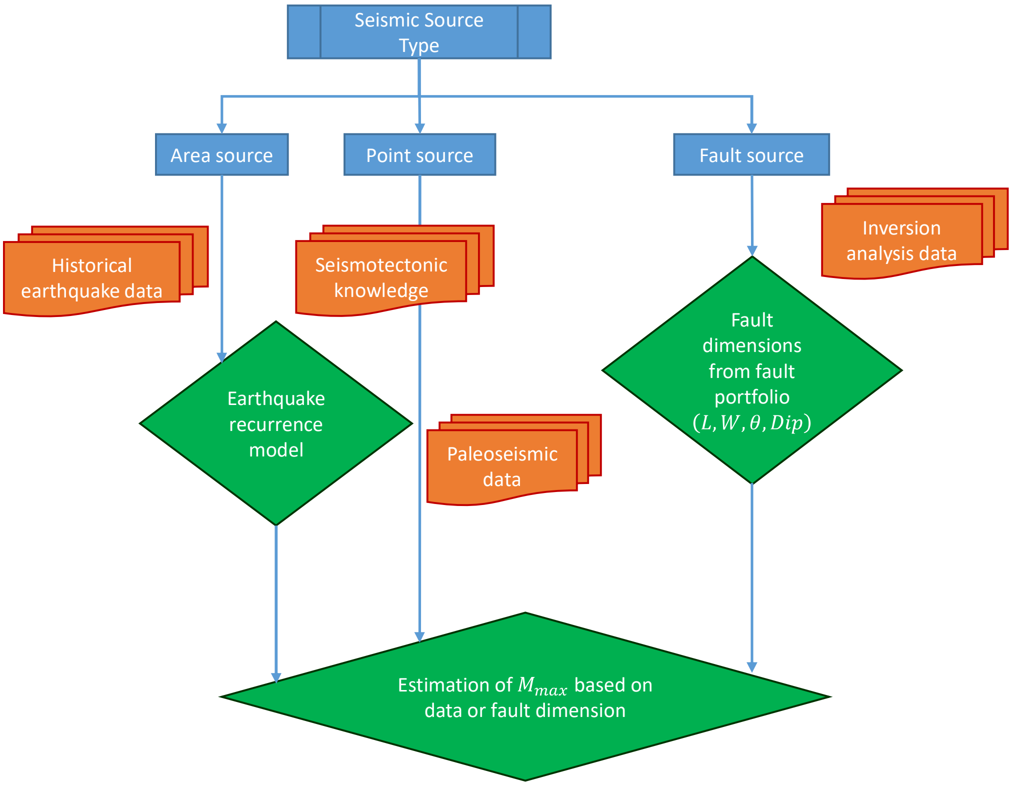

FIG. 3. Flowchart for the M_{max} estimations for different seismic sources

In PSHA studies, zones of diffuse seismicity (as defined in SSG-9 (Rev. 1) [2]) are used in regions where seismogenic fault locations are not well known. Worldwide experience has shown that areal source zones are particularly useful in PSHA studies conducted for stable continental regions where the causative faults are not identified. Data in the historical and instrumental earthquake catalogues are used to develop the earthquake recurrence models for zones of diffuse seismicity. In the absence of specific fault geometries, the estimation of M_{max} values for such sources also becomes challenging. The systematic compilation and examination of large earthquakes within a stable continental region, motivated by the need to assess M_{max} values for the areal sources in the central and eastern United States, resulted in the development of a Bayesian estimation procedure for M_{max} values [8]. This procedure was specifically developed for the seismic source zones within the central and eastern United States and includes prior distributions based on the largest earthquake magnitudes within stable continental regions having similar tectonic characteristics and likelihood functions developed from the largest magnitudes that have been observed within the source zone of interest. Therefore, observational data has a significant contribution to the M_{max} assessments based on such Bayesian estimation techniques. Two example studies that evaluate M_{max} estimations with observational data for areal sources developed for PSHA studies in different Member States are provided in the Annex (Sections A.1.2.5 and A.1.2.6).

The Bayesian M_{max} estimation techniques are more applicable to stable continental regions, which generally have low to moderate levels of seismicity (e.g. central and eastern United States, northern Europe) and where additional constraints on the faults giving rise to the seismicity are not well known. Bayesian methods are typically not applied in active tectonic regions where active tectonics features are dominant because in such regions, the faults giving rise to seismicity are known, and fault characteristics, including their geometries and possible rupture dimensions documented in the fault portfolio, provide direct constraints on the M_{max} assessments.

2.2. USE OF OBSERVATIONAL DATA FOR EARTHQUAKE RECURRENCE PERIOD ASSESSMENTS

Observational data plays a prominent role in the development of earthquake recurrence models, both for time dependent and time independent (Poisson) PSHA approaches. Figure 4 shows a basic flowchart that explains the use of observational data for developing earthquake recurrence models and evaluating the recurrence periods of earthquakes. Alternative models (exponential, characteristic and composite) are available in the literature to define the magnitude recurrence characteristics of a seismic source. Observed seismicity associated with a seismic source, including instrumental and historical earthquake locations, provides the opportunity of testing the validity of the selected magnitude recurrence model and its parameters. The example given in the Annex (Section A.1.2.2) compares the recurrence rates of observed seismicity (with synthetic variations) with the modelled recurrence rate of the seismic source. It is possible to do this check for both areal source zones and fault sources.

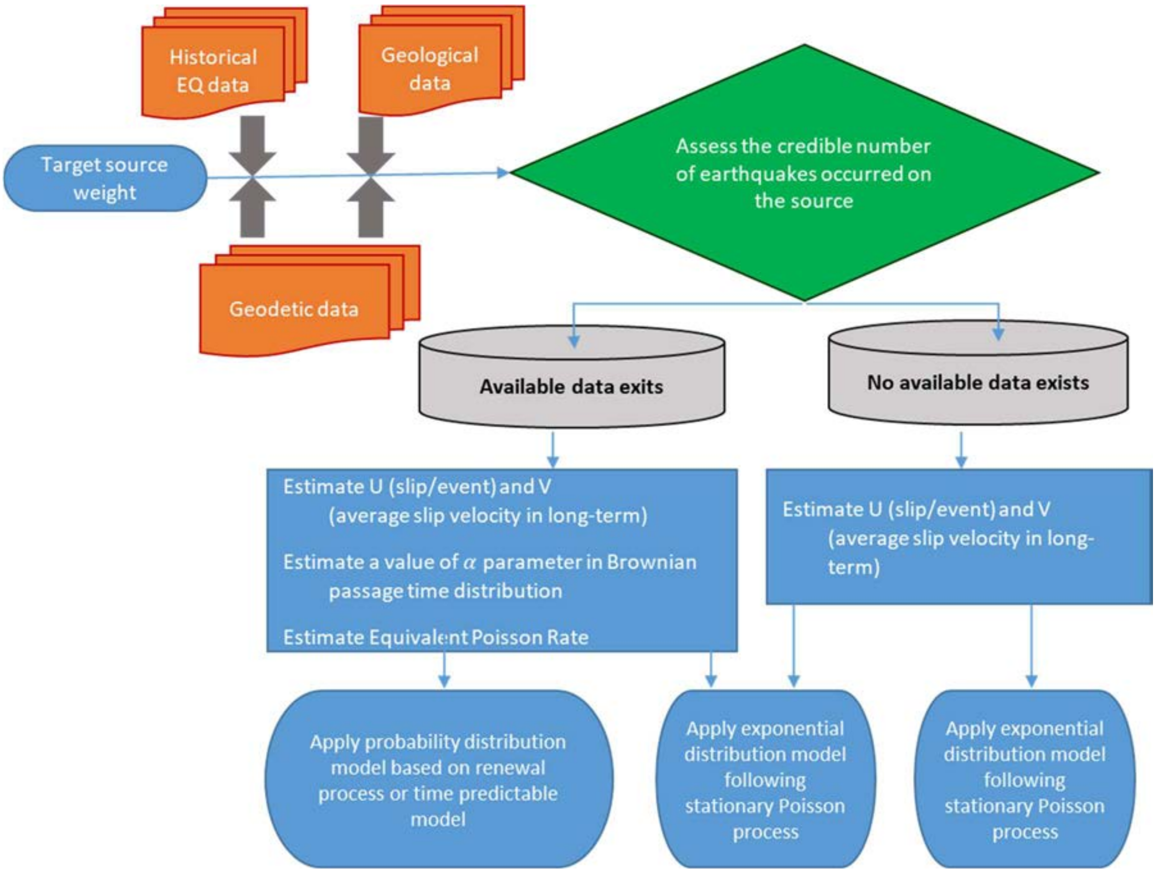

FIG. 4. Flowchart for earthquake recurrence period assessments in PSHA

In some PSHA studies, the earthquake occurrence rates are modelled using time dependent approaches such as Brownian passage time (BPT) (e.g. see Refs [9, 10]). It is often observed that the value of slip per event and the average slip calculated over the whole fault activity period differ depending on the adopted methods and data. Therefore, selection and compilation of observational data is of key importance here as well. It is, however, difficult to judge, if the adopted earthquake occurrence model in the time domain sufficiently conforms with the observational data using a limited number of events because the recurrence period is an average value and the recurrence period itself potentially has significant variability. Section A.1.2.1 of the Annex provides an example for the use of observational data in evaluating the time dependent earthquake probability model.

2.3. USE OF OBSERVATIONAL DATA FOR SPATIALLY SMOOTHED EARTHQUAKE RECURRENCE RATE ASSESSMENTS

Observational data is also a key input to PSHA studies, as smoothed seismicity rates are utilized to calculate future earthquake recurrence rates. Figure 5 shows a flowchart for evaluating the recurrence parameters of spatially distributed earthquakes. For this analysis, seismotectonic knowledge is needed as well as seismicity and paleoseismic data. Seismotectonic structures are developed and assessed based on geologic and seismic data and can be evaluated by analysing a combination of regional seismotectonics, recent instrumentally recorded data, accounts of historical earthquakes, and geomorphological and geologic evidence from various investigations and explorations. An example process described in Ref. [11] develops a spatial earthquake distribution model based on smaller earthquakes in the eastern United States to forecast the locations of larger earthquakes to develop the national seismic hazard maps.

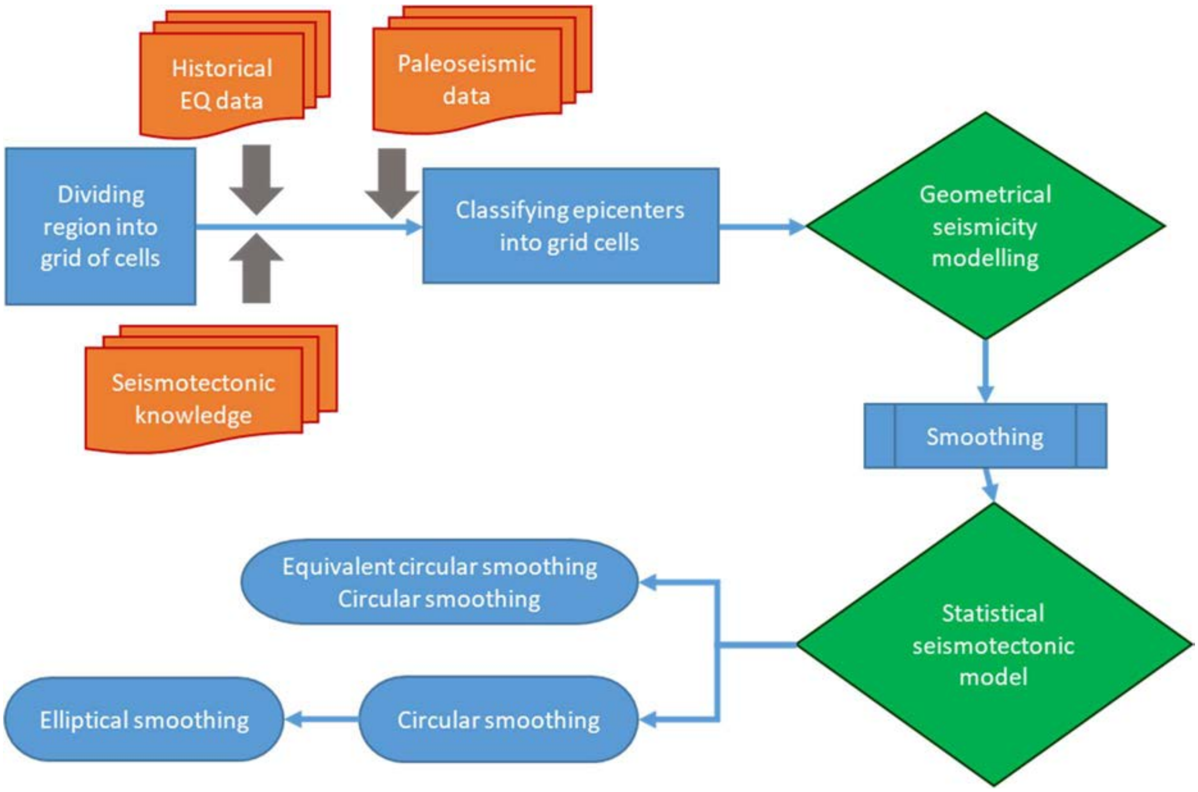

FIG. 5. Flowchart for estimating spatially smoothed earthquake recurrence rates in PSHA

2.4. USE OF OBSERVATIONAL DATA FOR THE ASSESSMENT OF GROUND MOTION MODELS

Figure 6 shows a summary of the GMM selection process in PSHA studies. Selected GMMs are intended to represent the characteristics of ground motions expected by future potential earthquakes affecting the target site (or area). In building ground motion characterization logic trees, GMMs need to be selected objectively, and epistemic uncertainties arising from GMM selections need to be captured quantitatively based on the observational data and other technical considerations. Ideally, GMM selection is conducted based on the site- or region-specific data. However, it is beneficial to acquire additional data that has been recorded in similar tectonic environments, so that epistemic uncertainties can be addressed sufficiently, especially in regions that lack observational data. In selecting GMMs, it may also be necessary to correct or adjust them based on site- or region-specific observational data.

The consistency between modelled and observed values is also a key issue. While it is not always possible to have observed ground motion recordings to be used in such studies, alternative approaches such as utilization of data from other regions with similar tectonic characteristics may be applicable and will need adequate representation of uncertainties.

Figure 6 highlights the criteria regarding the selection methods to be adopted and discussed in detail based on available data. Sections A.1.3.4 and A.1.3.5 of the Annex provide examples of data-driven methods in ranking and performance evaluation of GMMs.

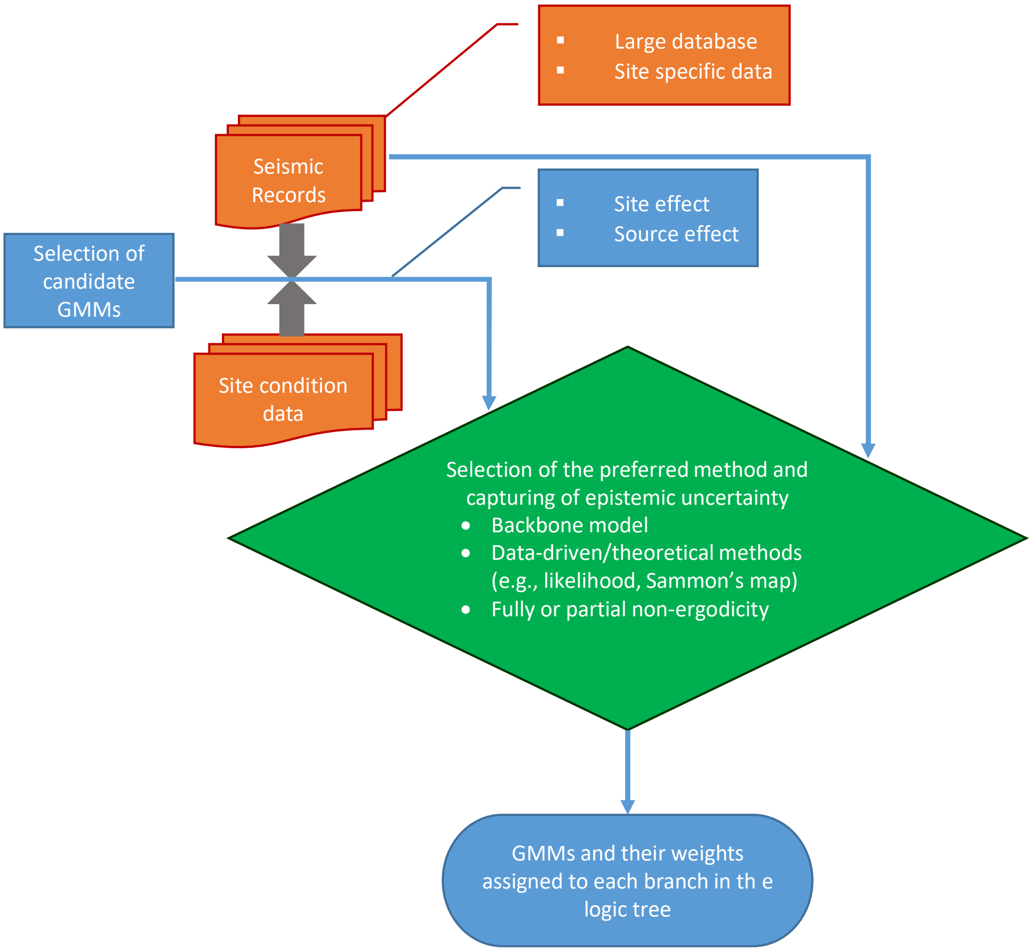

FIG. 6. Selection and weighting of GMMs for the ground motion characterization logic tree.

3. CONCEPTUAL FRAMEWORK FOR EVALUATING PROBABILISTIC SEISMIC HAZARD ANALYSIS

This section describes the conceptual background for evaluating PSHA based on observational data for elementary and integral steps, including the time and geographical scales of the process.

3.1. ELEMENTARY AND INTEGRAL EVALUATION CONCEPTS

Two different types of evaluation process can be described to assess PSHA studies based on the time to apply the evaluation process. The elementary evaluation focuses on the PSHA input parameters (mainly the logic tree branches) during the PSHA development and the integral evaluation focuses on the overall results such as the hazard curves, which incorporate the combined effects of multiple logic tree branches. Figure 7 shows the relationship between these two levels of evaluations. Both evaluation processes provide the PSHA developers with a reproducible path to help them assess their input parameters, models, and assumptions, showing that some of the alternatives might have a low likelihood, and encourage them to reassess or reconsider their choices and assumptions based on the information gained. Therefore, the evaluation process increases the confidence in characterising the centre, body, and range of the technically defensible interpretations.

3.1.1. Evaluating elements of logic trees: the elementary evaluation

The main objective of this evaluation is to check the elementary steps of a PSHA study to assess if they are in a good agreement with the observational data (see Figure 7). For example, when an earthquake occurrence model is developed for the PSHA, an obvious action is to ensure that the model and its parameters are consistent with the observational data. Assuming no procedural errors occurred in the earthquake occurrence model development, the model developed needs to be consistent with the observational data used to create it. Otherwise, the inconsistent initial model might not be suitable for the seismic hazard assessment for nuclear installations. If multiple earthquake occurrence models are used, evaluation of this elementary step would allow the PSHA developer to rank, score, or weight the alternative models for this elementary step (see Section 2.3). Similarly, for the set of alternative GMMs included in the PSHA logic trees, specific evaluation methods exist to rank, score, or weight such alternative models (see Section 2.4). The final goal of the elementary evaluation is similar to data-driven methods used to define the weights of the branches of a logic tree. Nevertheless, the elementary evaluation process could also be used to improve the probabilistic model parameters, adapting some of its underlying assumptions.

3.1.2. Evaluating results of probabilistic seismic hazard analysis: the integral evaluation

In a PSHA study, the full logic tree may include thousands of logic tree branches. Typically, the primary branches of the logic tree are associated with the selection of the seismic sources and the GMMs (see Figure 7). The secondary branches are associated with each primary branch to manage its epistemic uncertainties (e.g. alternative source geometries, maximum magnitudes). Each complete path through the entire logic tree leads to one hazard curve. The final PSHA results include the mean, median and other fractiles hazard curves calculated from thousands of hazard curves obtained for each path. Even when the elementary components of the PSHA are in general consistent with the observational data, some combinations of the logic tree branches may result in hazard estimates that are inconsistent with the observational data.

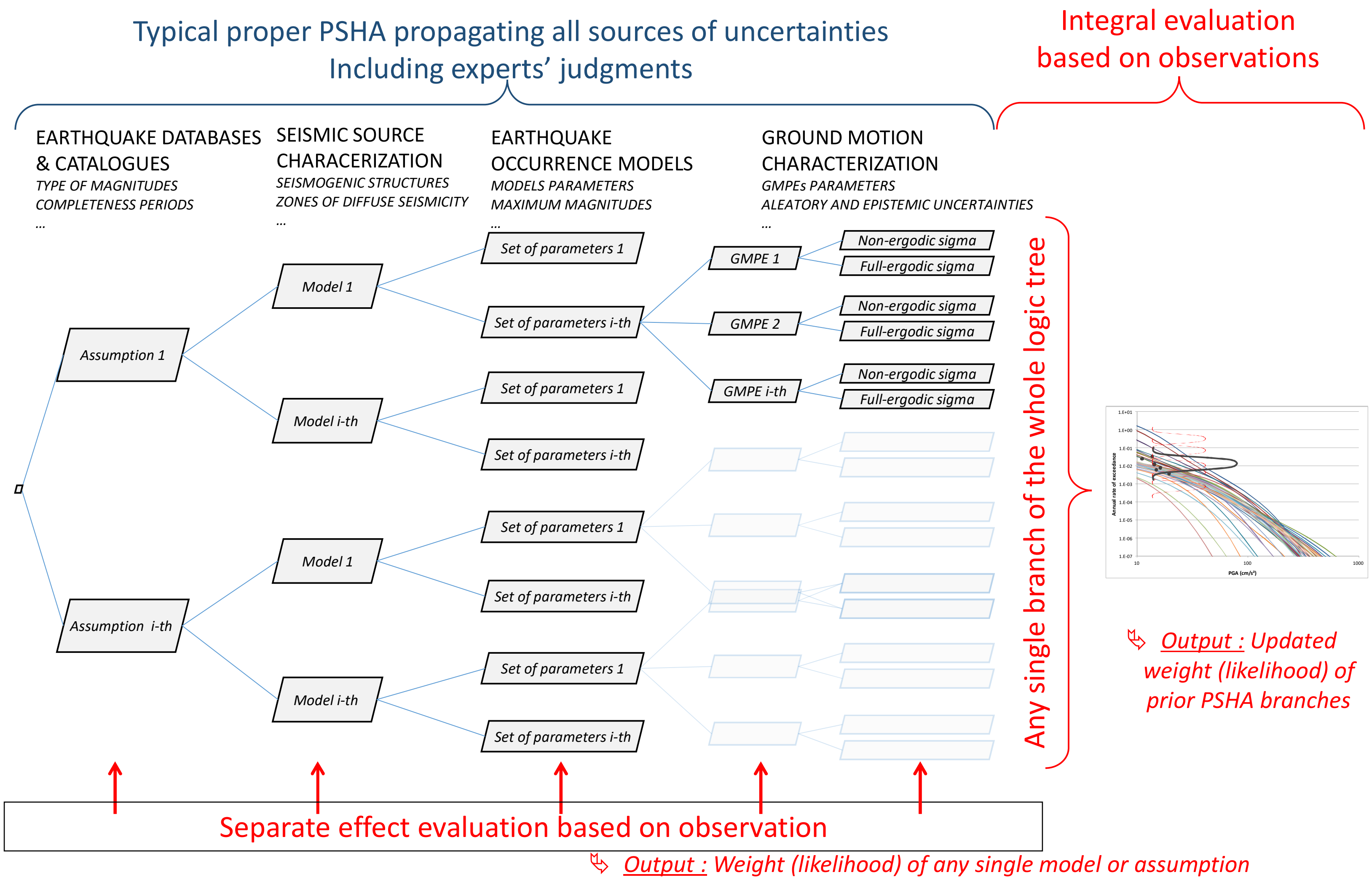

FIG. 7. Overview of the evaluation process of a PSHA including elementary steps evaluation and integral evaluation.

For example, the combination of small biases associated with individual elementary steps may accumulate, or some elementary steps cannot be evaluated based on observations, or certain elementary steps could have been improperly combined. Therefore, the evaluation process based on observational data may also be applied to the full set of final PSHA hazard curves. The integral evaluation results can then be used to understand the models and their weights that may cause inconsistencies and to modify weights or reject models that produce results that are not supported by the observational data.

3.1.3. Key principles for elementary and integral evaluations

The elementary and integral evaluation processes are expected to follow the principles given below:

- Utilize reliable observational data as discussed in Section 4;
- Be consistent with the physical behaviour and assumptions included in the PSHA;
- Propagate aleatory and epistemic uncertainties at each step of the evaluation process;
- Provide an evaluation grade associated with each single model, assumption, or branch of the logic tree;
- Consider the whole distribution of PSHA results (i.e. the full set of hazard curves, and not just median or mean results).

The final objective is to assess the consistency between the calculated PSHA results and the observational data available to determine the adequacy of the seismic hazard calculations and if needed update or change some of the PSHA input parameters to align the outputs to be more consistent with the available observational data.

3.2. TREATMENT OF MODEL UNCERTAINTIES IN EVALUATING PROBABILISTIC SEISMIC HAZARD ANALYSIS

PSHA includes a series of input models and parameters representing a number of technically defensible assumptions. Each model combines available information in a different way and assume different positions about plausible seismogenic processes (e.g. earthquake generation, source geometries, seismic radiation pattern). On the other hand, each model cannot be considered as entirely independent since they may share common information available at the time of their formulation and are calibrated by using the same observational data.

In general, PSHA input models cannot be verified in a standard manner. Only closed formal systems can be actually proven to be providing 'true' statements [12]. On the other hand, a pre-requisite for any model is its validity, meaning that it does not contain logical or technical inconsistencies preventing its application for practical purposes. Thus, it may be assumed that all competing models are valid and scientifically sound, provided that they are developed using sound technical procedures, and include all available data and information. When these models provide different outcomes, they represent the range of epistemic uncertainty included in the process. However, if some of these models produce outcomes that are clearly inconsistent with the data, then they can be excluded from PSHA. Once the unsuitable models are screened out, a comparative quality evaluation may be needed to rank the models or to build a combined model with associated weights that is able to include the best features of all the competing models.

A possible way to study this problem is to adopt a coherent probabilistic formulation in the framework of the Bayesian theory, which provides a good and simple mathematical basis to perform this statistical comparison, although other methodologies could also be applied. In this approach, a confidence measure is associated with each PSHA input model or parameter in the form of a probability value (likelihood). This value can be assessed based on any judgement of the inherent properties of the model ‘prior evaluation’ or from any match with observational data ‘posterior evaluation’. In both cases, confidence assessments may be the result of integral evaluations, concerning the model as a whole, or by considering any combination of elementary evaluations concerning constitutive elements (ground motion prediction equations, seismicity rates, source geometries, etc.). Whatever the procedure used for such an assessment, the Bayesian theory allows combining expert opinion and observational data by considering the first as prior and the latter as posterior probabilities.

3.2.1. Limitations of the evaluation based on observations

In the case of posterior evaluations, the analysis requires some caution. First of all, any comparison to observational data needs to take into account the probabilistic character of the model outcomes. The most straightforward approach accounting for this feature is comparing the exceedance probabilities with the time period of observations. Since the seismic hazard analyses for nuclear installations primarily concern with low annual exceedance probabilities (i.e. long average return periods), this comparison requires long observational records to achieve meaningful results. In many sites such long observational records will not be available. To address this challenge, variability over space may be used as a substitute for variability over time (i.e. the ergodic assumption). This assumption is not concerned with seismicity itself (whose inherent properties are unknown), but the characteristics of the considered PSHA input model. As an example, most PSHA models assume that earthquakes occur independently, as part of a stationary process. This implies that seismicity is assumed to be an ergodic process.

Therefore, the main limitation in the applicability of an evaluation based on observations is the amount of recorded seismic data that are available at a given site. In an ideal case, only the seismic data recorded at the site is used for the evaluation. Nevertheless, very often the recorded data at a specific site is scarce or does not exist. Then, in absence of local data, seismic data recorded in the surroundings of the site could be used. In this situation, data processing will have to follow a specific process. However, the seismic database used to assess PSHA, and observational data (recorded ground motion data) need to be preferably compiled using a region with a similar seismotectonic context to the site.

The assessment is performed by considering the available observational data used to build the model itself. It might be presumed that this ‘sanity check’ may provide a good match between the model outcomes and the observational data, but this is not always the case. In fact, PSHA models are rather complex and include, in the same package, observational data and technical judgements. This combination results in the overall outcome. It is not always obvious that PSHA outputs will be compatible with the original data. This implies that any sanity check is not simply an academic exercise, and its results may be very informative about the overall quality of the model. Of course, it is much better if tests confirm the results of the PSHA model with an independent data set. Depending on the context, the confirmation analysis might not be possible and only a sanity check may need to be performed.

3.2.2. Technical basis of the evaluations based on observations

A single model in the PSHA is expected to provide a set of scenarios (hereafter each scenario is generically indicated by a single representative parameter a), each characterized by a different degree of confidence to be assessed based on available information. This degree of confidence is expressed in the form of a probability $h_\tau(a)$, i.e. by attributing to each scenario a real number in the range [0,1] whose extremes respectively represent the positions of realization. Corresponding scenario for a during the expected operational period of nuclear installation (τ) is impossible if $h(a) = 0$ or certain if $h(a) = 1$. In between, lie all possible nuances of confidence about the actual occurrence of a during τ.

If one assumes that a finite set of n possible scenarios exists, to be coherent, probability values need to be assessed in a way such that the summation equals to 1 as shown in Eq. (1).

$$\sum_{i=1}^{n} h_\tau(a_i) = 1 \tag{1}$$

In the case that the parameter a may assume continuous values in the range [a_{min}, a_{max}], Eq. (1) changes into Eq. (2).

$$\int_{a_{min}}^{a_{max}} h_\tau(a)\,\mathrm{d}a = 1 \tag{2}$$

The set of probabilities $h_\tau(a)$, given in Eq. (2) associated with the possible scenarios, defines a probability density function. More frequently, since establishing a reasonable upper bound for expected value of a is of main concern, the seismic hazard analysis focuses on the 'hazard' function defined as an exceedance probability as shown in Eq. (3).

$$H_\tau(a) = \int_{a}^{a_{max}} h_\tau(a)\,\mathrm{d}a \tag{3}$$

This formulation also includes the unrealistic case of a seismic hazard analysis model characterized by a single possible outcome a_d (it could be considered a prediction about a). In this case, the probability distribution has two values only: $h_\tau(a = a_d) = 1$, $h_\tau(a \neq a_d) = 0$ and the integral given in Eq. (3) assumes the values given in Eq. (4).

$$\begin{cases} H_\tau(a) = 1 & a \leq a_d \\ H_\tau(a) = 0 & a > a_d \end{cases} \tag{4}$$

The general outcome of PSHA in terms of $H_\tau(a)$ is generally hidden behind a very common representation of seismic hazard in terms of any specific quantile of the distribution $H_\tau(a)$, i.e. by focusing on the value a_r corresponding to a fixed expected operational period τ and a fixed exceedance probability H_r. In this representation, a_r is implicitly defined in the form of Eq. (5).

$$H_\tau(a_r) = H_r \tag{5}$$

Therefore, the choice of representing the seismic hazard by a single physical parameter representative of ground shaking may hide the inherent probabilistic form of the forecast.

Two general types of PSHA model input exist. The first is time dependent, and the results only hold for a specific expected operational period interval, $\tau*$ (e.g. a specific time span such as between the years 2020 and 2100). The second is time independent; it only depends on the duration of τ (e.g. 80 years) irrespective to the choice of any specific time interval.

In principle even if all criteria described in the previous section are applied in developing the PSHA input models, due to expert judgements and other different alternative assumptions used in developing the seismic source or ground motion models, PSHA studies conducted by different experts may result in outputs that are different from each other. While a structured expert elicitation approach is used in PSHA studies conducted for nuclear installations, there is always some degree of judgment that may result in different outcomes. These types of observation are more common for PSHA applications conducted for non-nuclear facilities. In such situations multiple PSHA studies result in significantly different outcomes and each PSHA model provides different estimates of $H_\tau(a)$ for the same site. By using the definition of conditional probability to explicitly differentiate among hazard estimates provided by different models, the PSHA provided by the *i-th* model (m_i as shown in Eq. (6)) can be described.

$$H_\tau(a) = H_\tau(a|m_i) \tag{6}$$

Thus, a practical problem arises to evaluate the reliability of each model, or its adequacy to represent the seismogenic processes under the study region. Evaluating the reliability of the model does not correspond to a sensitivity analysis. The latter one aims at defining the relative importance of pieces of evidence included in the model and exploring the impact of relevant uncertainty on the final outcomes. Sensitivity analyses explore properties of a single model without any reference to observational data. Empirical evaluation of the reliability of a model, instead, aims at assessing the degree of confidence to be attributed to each hazard model, comparing the outcomes of each model with observational data.

Each model depends on a set of parameters (θ). In order to keep a coherent description of the problem, a degree of confidence can be attributed to each *i-th* model (m_i) in terms of a probability $Q(m_i)$. For the following discussion it will be useful to separate the uncertainty managed within the computational model (e.g. stochastic earthquake occurrence) and uncertainty concerning the computational model itself. Given the computational model m_i, two probabilities have to be considered: $Q(m_i)$ which expresses the degree of confidence in the m_i model and $H_\tau(a \mid m_i, \theta)$ which expresses the probability that the value a is exceeded when the *i-th* computational model H_i is considered.

As stated in Eq. (7), any model is expected to depend on a number of parameters θ to be inferred from statistical observations or physical constraints. It is also expected that these parameters will be affected by uncertainty (due to measuring tools or statistical variability), which can be represented by a multivariate probability distribution $\pi(\theta)$. This further source of uncertainty can be eventually implemented by considering the marginal probability:

$$H_\tau(a|m_i) = \int_\theta H_\tau(a|m_i,\theta)\pi(\theta)\,\mathrm{d}\theta \tag{7}$$

where integration is extended over to the model parameters space. In the following discussions, the above defined marginal probability will be considered only.

In some cases, it could be preferred to attribute to the model any other quantitative measures of performance $S(m_i)$ which does not represent a degree of confidence (probabilities) but allows to score the models. Such scoring rules need to be 'proper' [13]. A scoring rule is proper if the forecaster maximizes the expected score for an observation drawn from the probability distribution h when the probabilistic forecast is issued.

A fundamental difference exists between $Q(m_i)$ and $S(m_i)$. The first has an intrinsic meaning, being expressive of a degree of confidence associated with the model, while the latter only represents a way to order the models as a function of the relative performances. In other words, $Q(m_i)$ is an absolute quality parameter while $S(m_i)$ is just a relative scoring parameter.

In principle, $Q(m_i)$ could be used for assessing the models and excluding any of them because the associated degree of confidence is below any threshold (e.g. less than 5%). In the case of PSHA models, due to the relatively small dimension of the empirical data sets used in evaluation, the power of these types of test is very low which subsequently increases the probability of accepting an inadequate PSHA model as valid [14]. This implies that when a very low exclusion threshold is considered, the possibility of considering an inadequate model as acceptable remains high. On the other hand, since the exclusion depends on observations available at the time of evaluation, it cannot be guaranteed that future new data sets may change the present evaluation. Thus, any exclusion needs to be performed with great caution [15]. Therefore, testing of PSHA models has not been considered in this publication, which will encourage attributing a low weighting to such a model or assumption instead of providing any exclusion threshold.

The scoring parameter $S(m_i)$ allows to select the best fitting model, irrespective of the eventual lack of confidence affecting it. Another important difference is the possibility offered by $Q(m_i)$ to combine the outcomes of several models to obtain a combined hazard estimate. To this purpose, $Q(m_i)$ values need to satisfy the closure condition as given in Eq. (8):

$$\sum_{i=1}^{M} Q(m_i) = 1 \tag{8}$$

where the summation is extended to all competing models *(m_i)*, which represent the whole set of possible PSHA models. This implies that all possible models considered in the analysis are mutually exclusive and only one of them is actually representative of the underlying seismogenic process. Under this very restrictive condition, given the set of m_i models, a hazard estimate unconditional to model M can be obtained in the form of Eq. (9) by taking advantage of all the models (combined hazard estimate) and providing forecasts possibly more effective than that provided by any single model.

$$H_\tau(a) = \sum_{i=1}^{M} Q(m_i) H_\tau(a|m_i) \tag{9}$$

Equation (9) does not contradict the condition of mutual exclusivity between the models, but only accounts for our inability to identify the one providing reliable outcomes.

However, it could be difficult to state the mutual exclusivity of the models in actual applications. They could share part of the computational procedure or of the data used for tuning some relevant parameters. In these cases, suitable corrections could be defined to reduce the effects the possible overrepresentation of possibly correlated models (e.g. see Ref. [16]).

3.3. BAYESIAN APPROACH FOR EVALUATING PROBABILISTIC SEISMIC HAZARD ANALYSIS MODELS

Both prior or posterior evaluations may be used for $Q(m_i)$ and $S(m_i)$. In case of prior evaluation, the model is evaluated based on its internal coherency or the theoretical plausibility of underlying assumptions with no reference to its outcomes. An example of this kind of approach is the use of the logic tree methodology, which is primarily based on field observations, available data and interpretations including judgments provided by a panel of experts asked to rank the considered models. There are several ways to provide such expert elicitation process (e.g. see Ref. [17]) by considering the model as a whole or by combining piecewise judgments concerning parts of the model (e.g. see Ref. [18]).

In the case of posterior evaluation, the model is evaluated only by comparing its outcomes with seismic occurrences observed during a control period. In theory, a posterior evaluation is more straightforward and objective since it does not rely on debatable expert evaluations. However, it poses a number of problems (e.g. the possibility to identify a suitable control period for evaluations) making the first approach in many cases unavoidable. Furthermore, the physical and logical plausibility of the underlying hypotheses, which are the subject of prior evaluations, cannot be disregarded.

These two approaches are mutually compatible and are not alternatives to one and another. A Bayesian formalization can be considered for making evident the logical links between these two approaches.

Given a PSHA model m_i and a set of seismic occurrences R during the control period, the degree of confidence $Q(m_i|R)$ attributed to the model after a posterior comparison with the observed seismic occurrences will be:

$$Q(m_i|R) = Q^*(m_i)\left(Q(R|m_i) \Big/ \sum_{j=1}^{M} Q^*(m_j) Q(R|m_j) \right) \tag{10}$$

where $Q^*(m_i)$ represents the prior degree of confidence attributed to the *i-th* model before any posterior comparison with observational data and satisfies the closure condition given in Eq. (8). $Q(R|m_i)$ is the likelihood of the realization R in the hypothesis that the model m_i actually represents the underlying seismogenic process. The denominator of Eq. (10) is also known as marginal likelihood. In principle, $Q(R|m_i)$ could be used for screening the models and exclude those are characterized by levels of likelihood that is too low to be considered as unreliable beyond any comparison with other models. In a specific case where all models are completely disconfirmed by the observational data (i.e. all the likelihoods are very low) the resulting degree of confidence $Q(m_i|R)$ relative to any model would result in a relatively high value. This is because the basic assumption behind Eq. (10) is that at least one model among the other mutually exclusive ones is able to represent the seismogenic process correctly, which cannot be the case for PSHA. A preliminary screening phase may be necessary before using Eq. (10) [19]. However, as stated above, since at least one model is necessary to provide the requested forecast, as in the case that all the models present very low likelihoods, at least one of them will be preserved.

Equation (10) presents the Bayesian Model Averaging method, which accounts for model uncertainty, and is theoretically justified by the fact the best predictive model is the one that integrates all known sources of uncertainty [20]. In the case that the likelihoods given in Eqs (7) and (10) cannot be easily computed exhaustively, a numerical procedure may be used, which includes three steps:

(1) Draw any *k-th* model (m_k) at random from the list of M candidate models, selecting each model m in the candidate list with probability $Q(m|a)$;

(2) Simulate the values of the parameters θ_k controlling the *k-th* model by considering the posterior density $\pi(\theta|a,m_k)$, for instance, using Monte Carlo Markov Chain techniques [21];

(3) Simulate $a_{new,k}$ from the conditional probability density function $Q(a_{new}|\theta_k,m_k)$.

This results in a sample $(a_{new,k})_{1<k<K}$ from the model-averaged predictive distribution that can be used, for example, to compute the Monte Carlo estimates of the mean predictive values, predictive credible bounds, scores, or any other relevant quantity.

It is important to note that, contrary to any approach based on the selection of the best performing model, the predictive distributions from all candidate models are explored randomly in Bayesian model averaging, making the inference robust to model uncertainty and also less sensitive to the uncertainty tainting the data. On the other hand, Bayesian model averaging predictions have a higher computational cost than the Bayesian model selection, since they involve running M independent Monte Carlo Markov Chain runs, one for each competing model. By keeping in mind the key criteria for evaluations, pruning branches that have close to zero posterior probability could help in reducing the computational cost of such approaches and provide a good compromise.

If applicable, Eq. (10) implies that a correct evaluation of the *i-th* model involves the knowledge of all the alternative models and respective prior evaluation. A key element of Eq. (10) is the likelihood term $Q(R\text{-}m_i)$, whose value is deduced from the model itself without any consideration about the actual structure of the seismogenic process.

The application of Eq. (10) cannot be extended to the case where any scoring parameter $S(m_i,R)$ is considered instead of $Q(m_i|R)$ because it does not satisfy the closure condition given in Eq. (8). In this case, any re-normalization procedure could be considered in the form of:

$$Q'(m_i|R) = S(m_i, R) \Big/ \sum_{j=1}^{M} S(m_j, R) \tag{11}$$

by keeping in mind, the logical implications of this position (e.g. the fact that mutual exclusivity is warranted).

In the absence of any complete analysis of the all the possible models, the term $Q(R\,|\,m)$ in Eq. (10) can be directly used to compare performances of two alternative models (*i-th* and *j-th*) by considering the Bayes Factor, BF_{ij}:

$$BF_{ij} = \frac{Q(R|m_i)}{Q(R|m_j)} \tag{12}$$

The Bayes Factor only evaluates the relative performance between two models without any evaluation of their actual reliability. Both models, in fact, may be completely characterized by very low likelihoods when compared with observations.

From Eq. (10) it may be seen that any single realization *R* may change the judgment about the reliability of the considered model and prior evaluations play an important role. It is expected that by iterating the application of Eq. (10) to subsequent realizations of the seismogenic process, the importance of initial prior evaluations becomes progressively lower.

3.3.1. Key aspects related with the evaluation of probabilistic seismic hazard analysis

An important aspect to be considered is how PSHA models are evaluated with respect to their internal organization. In general, each model includes several elements, such as earthquake recurrence models, seismic source geometries and ground motion prediction models, combined to provide the outcome $H_\tau(a|m_i)$. It is possible to check the reliability of each element of the model and then provide an overall evaluation by combining the individual reliability estimates. This approach might simplify the task (single elements could be evaluated more easily) but does not determine the role of global organization in which single elements are combined in the model. Alternatively, the model may be considered as a whole and only the outcomes evaluated. In this case, the eventual inadequacy of the model may be detected but no information is provided about the component responsible for this. These two approaches are not alternatives to each other and need to be considered together.

Another important aspect to be considered is to evaluate the outcome of the PSHA model against the observational data. As stated above, the main outcome of PSHA is the hazard function $H_T(a)$ which defines a set of exceedance probabilities associated with the values of *a*. An alternative is to only consider the percentiles, as given in Eq. (5). The first approach is more complete but involves complex comparisons between different observational data available over a relative wide range of *a* values. The second approach is less complete, but more feasible since it focuses on the probability of overcoming *a* single value.

3.3.2. General problems in posterior evaluations

In principle, most reliable models will provide forecasts as close to actual realizations of the seismogenic processes as possible. In the case of PSHA models, it is expected that the hazard estimates are as unbiased as possible, and no systematic under-evaluation or over-evaluation is observed when the outcomes are compared with actual observations. This implies that any measure of bias or misfit among model forecasts and observational data need to take into account the inherent probabilistic character of the estimates. In principle, it is expected that several comparisons of the model outcomes with observational data are needed to define the value of the misfit.

The first key element to consider is the definition of control periods where seismicity data to be compared with the model's estimates are drawn from. As stated above, PSHA for nuclear installations relates to long term estimates. Hence the choice of suitable time periods for posterior evaluations becomes quite difficult. Furthermore, to obtain a reliable evaluation, data to be used with the model's outcomes will not have been considered for the model parameterization (please see Section 2 for the definitions of new and available data). This represents an important constraint for the evaluation of PSHA models that were developed in the last 5–10 years by exploiting all pieces of information available at the time of their formulation. On the other hand, the PSHA estimates need to be applicable to operational periods of the order of several tens of years, through the lifetime of the nuclear installation. Therefore, the evaluation may be performed at a later stage, e.g. during periodic safety reviews. An alternative, in the case of time independent hazard estimates, is considering a backward approach, i.e. only the past occurrences are considered for the analysis. In this case, the evaluations reduce to a consistency evaluation (or sanity check). This cannot effectively evaluate the reliability of the model in anticipating future seismic occurrences. On the other hand, the failure of this consistency evaluation might undermine the credibility of the model.

Another important aspect is the role of uncertainty in the observational data to be assessed against the PSHA results. In some cases, (e.g. when the observational data is the ground acceleration recorded at reference sites) these uncertainties could be considered as negligible. In many other cases, however, (e.g. acceleration data deduced from macroseismic observational data or deduced from non-reference observational data) the relevant uncertainty needs to be considered to avoid biased quality evaluations.

The third key element is the fact that available data will produce few (or just one, in most cases) realizations of the seismogenic process modelled for the PSHA. On the other hand, due to the probabilistic character of model's estimates, eventual biases could be revealed by just considering a number of realizations of the process. The feasibility of this approach (the 'time based' approach) is strongly limited also in the case of a backward evaluation. Since PSHA studies utilize ground motion parameters of engineering interest (response spectral accelerations, peak ground acceleration, etc.), the database of available instrumental observational data covers a relatively short time interval (70 years at most, but 30-40 years in general). Similarly, seismicity databases cover several decades to few centuries. To fill the gap, macroseismic information can also be considered and eventually converted (using empirical relationships) into para-instrumental data to be compared with the model's outcomes. The uncertainty concerning this conversion needs to also be accounted for in the sanity check.

3.3.3. The area based approach for the evaluation of probabilistic seismic hazard analysis

A possible alternative for evaluating PSHA is adopting an 'area based' approach. The basic idea is that a sample of multiple sites can be considered as a multiple realization of the same process. Depending on the characteristics of the model of concern, ergodicity can be assumed or not to evaluate relevant statistics. In this approach, a number of sites (*N*) is considered to define the realization (*R*) of the seismogenic process to be assessed against PSHA estimates. At a *k-th* site, a seismic record exists for a_k values observed during a time span t_k. It is assumed that the record is complete above a threshold value, i.e. all earthquakes at the site with values of *a* above the threshold have been recorded. The seismic records have the form of vector of *l* elements $\{a_{k1}, a_{k2}, \ldots, a_{kl}\}$ reporting single seismic observational data at the *k-th* site. At these locations the *i-th* model provides the exceedance probabilities $H_\tau(a|m_i)$. Commonly, these probability estimates are time independent and have the form of a Poisson distribution:

$$H_\tau(a|m_i) = 1 - e^{-\lambda_i(a)\tau} \quad (13)$$

where $\lambda_i(a)$ is the exceedance rate corresponding to *a* provided by the *i-th* model. Using Eq. (13), one can compute the exceedance probability associated with the time span t_k relative to the seismic record available at the *k-th* site, as shown in Eq. (14).

$$H_{t_k}(a|m_i) = 1 - e^{-\lambda_i(a)t_k} \quad (14)$$

The main part of the area based approach is comparing probabilities provided by Eq. (13) with exceedance frequencies observed at *N* sites. In most cases, PSHA models for nuclear installations provide exceedance probabilities at any site, without any reference to what happens in neighbouring sites. This, however, does not imply that seismic occurrences at different sites are mutually independent. In fact, on most models, GMPEs are used to compute the exceedance probability at any site given that an earthquake occurs somewhere with a specific magnitude. Of course, the same earthquake will shake a number of sites implying that the exceedance probabilities at several sites will not be independent, and correlation is expected (e.g. [21, 22]). This correlation increases when the distance between the considered sites reduces at a rate that depends on the structure of the considered GMPE, the considered ground vibration frequency, etc. By considering mutually distant sites, this correlation is expected to be relatively small. However, in the lack of any estimate of correlation as a function of the site distances, it is unclear how to define the distance over which the effect could be ignored. Spatial correlation of hazard estimates is a necessary step of the area based evaluation procedures.

4. COMPILATION OF THE OBSERVATIONAL DATASET

This section introduces the main types of data that may be used in evaluation of PSHA and describes how to compile the observational dataset needed for the evaluation process. Finally, it gives some guidelines on how to manage observational data, especially on how to propagate uncertainties when a transformation from the PSHA output parameter into observation parameter is necessary.

4.1. TYPES OF OBSERVATIONAL DATA

The most accessible observational ground motion data to be used in PSHA studies consist of:

- Vibratory ground motion data (for the instrumental period): As the response spectrum of vibratory ground motion data is straightforward to calculate and seismic hazard curves are typically given in terms of response spectral accelerations, the evaluation methods may be applied to vibratory ground motion data without any conversion factors. However, the periods of completeness associated with instrumental data are usually short, especially in areas characterized by low-to-moderate seismicity.
- Macroseismic intensity data (for historical period): The use of macroseismic data (i.e. macroseismic intensity) offers the advantage that the periods of completeness can be very long, because in many regions there are macroseismic information for some centuries, at least for large intensities. However, as the comparison is performed with seismic hazard curves given in terms of response spectral accelerations, a transformation from ground motion parameter into intensity (or the reverse) is needed, which introduces an additional factor of uncertainty that needs to be considered.
- Geological and paleo-seismological observations (for pre-historical period): The geological data (i.e. paleo-seismologic observations, brittle speleothems, precarious rocks) offers the advantage to cover even longer periods of observation than macroseismic intensity data. However, to evaluate PSHA results on this basis, an appropriate process has to be followed, which allows the association of any observational data (e.g. permanence of precarious rocks) to a ground motion exceedance threshold.

Due to their availability and certain advantages, macroseismic intensity data (which covers longer observation periods) and vibratory ground motion data (for which no conversion is needed) seem to be the most suitable data for evaluating PSHA models. Geological data may also be used, but they are normally scarce, and their use may be limited and more difficult depending on the context. However, the combination of geological, macroseismic intensity and ground motion data provides a wider evaluation database and therefore, improves the evaluation process.

4.2. COMPILATION OF THE OBSERVATIONAL DATASET

Compilation of a suitable observational dataset is very important for the evaluation of PSHA based on observational data. Without a comprehensive evaluation dataset, a well developed and detailed PSHA might not be properly evaluated. The definition of the applicable observation database for a site could depend on each study and the site specific seismotectonic context.

Compilation of the observational database for the evaluation of PSHA needs to follow the process of the compilation of seismological database for performing the PSHA, as recommended in para. 6.9 of SSG-9 (Rev. 1) [2].

Depending on the type of the data, the information to be included in the evaluation dataset may be very similar to that of the seismological database. For example, the ground motion accelerograms recorded at the site under consideration may be used for performing or evaluating the PSHA. For both purposes, the ground motion dataset needs to include:

- Date, time, magnitude, depth, and location of the earthquakes;
- The peak amplitude parameters (Peak Ground Acceleration, Peak Ground Velocity, etc.) and response spectral accelerations of the recorded ground motions;
- Point and extended source-to-site distance metrics (e.g. rupture distance, Joyner-Boore distance[2]);
- The type of instrumentation and instrument specifications;
- Site characterization of the recording stations for free-field recordings and for in-structure recordings, specifications of the structure where the recordings are made.

For macroseismic intensity data, the information to be included in the evaluation database may be listed as:

- Date, time, intensity, magnitude, and locations of earthquakes;
- The intensity felt at the site and its associated intensity scale.

4.2.1. The seismic hazard curves for other sites

Nuclear installation sites are usually located in regions far from populated areas; therefore, well documented past observational data might not be available. The evaluation process often needs site specific seismic information recorded at some nearby sites (different from the site where the seismic hazard is evaluated). In addition to the data discussed above, the seismic hazard curves need to be computed in order to perform the evaluation process at these sites. The probabilistic model used to characterize the ground motion in a nuclear installation site can also be used to compute the seismic hazard curves in cities, villages or seismic stations situated at some distance from the nuclear installation site (please refer to the area based approach described in Section 3.3.3).

In the selected locations, different observational data could be used for the evaluation. At seismic stations sites, vibratory ground motion data can be directly used (without any conversion) and compared with the seismic hazard curves for integral evaluation. In old cities and villages (or populated areas), it is possible to find observational data (macroseismic intensity data directly reported in historical documents or issued from the interpretation of isoseismal maps) that needs to be converted to the parameter given by the seismic hazard curves.

[2] The Joyner-Boore distance is the shortest horizontal distance to the surface projection of the rupture plane.

4.2.2. The completeness and probability of exceedance calculations

The integral evaluation process is based on the comparison of the probability of exceedance provided by the PSHA and the observation database. To calculate the probability of exceedance, the 'counting approach' (see Appendix) may be used, which generally includes two steps summarized below.

- Step 1 - to count the 'expected number of observational data': In this step, the seismic hazard curves at some selected sites (often cities, villages and seismic stations situated in the proximity of the nuclear installation site) needs to be computed. The analysis of the seismic hazard curves at these sites will allow counting the "predicted number of observational data" on those sites (i.e. the number of exceedances of a fixed acceleration or intensity threshold). In some circumstances (for example for existing sites with instrumental data), the nuclear installation site could be also one of the considered sites.
- Step 2 - to take account of the 'actual number of observational data': In this step, the actual number of observational data need to be calculated, counting the seismic records on the selected sites (i.e. the number of exceedances of a fixed acceleration or intensity threshold). The observational data includes directly recorded or observed data (accurate macroseismic information, isoseismal maps, accelerograms, etc.). In some circumstances (for example for existing sites with instrumental data), the recorded seismicity at the nuclear installation site could be considered as directly recorded or observed data.

The completeness period of observations has to be carefully assessed in order to take into account the observations without any bias, and to use the observations consistently with the PSHA under consideration. Considering that, when time independent estimates are of concern, the seismic hazard curves are mostly defined in terms of annual exceedance rates and the use of incomplete seismicity records may underestimate the actual exceedance rates.

The use of macroseismic intensity in the evaluation process offers a strong advantage in terms of completeness. The time coverage of macroseismic information is significantly longer (historical period) than the period of completeness of vibratory ground motion data (a few decades in the instrumental period). Therefore, even if the use of macroseismic intensity could introduce an additional source of uncertainty, the extension of the period of comparison compensates for this. However, information available about the seismic history of any site may be affected by the incompleteness due to the lack of documentation. The lack of completeness mainly affects smaller earthquakes and sites where macroseismic effects were relatively weak (no damage). The incompleteness period may be quite difficult to define. The most straightforward approach will be the historical analysis of available documentation, but this kind of study is quite complex, expensive and time consuming. When earthquake effects are not documented at the site of interest and are known at any other site for the same earthquakes, possible incompleteness may be addressed by considering suitable GMPEs (in intensity terms) to compute the expected intensity at the site. The larger this expected intensity, the larger the probability of a local incompleteness (e.g. see Ref. [23]). Another possibility is provided by a statistical approach (e.g. see Ref. [24]).

For the use of instrumental data in the evaluation, it is important to define the actual period of recording of the seismic station (taking due consideration of maintenance periods for instance). Furthermore, it is very important to evaluate the trigger threshold, which may vary with time due to the change of recorder or other features (e.g. the trigger threshold may be increased due to the increase of environmental noise).

In this case, reconstructing recording history for each accelerometric site is of paramount importance. In the absence of documentation, the approach based on the use of GMPE could be of help (e.g. see Ref. [23]).

The same arguments could be applied to the use of pre-historical data. Its use has the advantage of even longer periods of observation, but it also implies the introduction of additional sources of uncertainty. In some cases, the pre-historical data is used to study the maximum events (i.e. precarious rock) rather than the number of events. In this case, the period of completeness is less important than when dealing with numbers of observational data with values higher than a given (generally large) threshold. In any case, the completeness period of observation needs to be clearly defined and documented.

4.2.3. Limitations relating to the compiled observational dataset

A limitation of the evaluation process may come from the amount of available observational data. On the other hand, for evaluating PSHA results, 'failed observation' may be a useful piece of information and 'no exceedance' of any given acceleration threshold needs to be considered. It is usually observed that when comparing a very large number of observation sites with predictions, the efficiency of the evaluation process is high. Similarly, when comparing a very scarce number of observation sites with predictions, the efficiency of the evaluation process is lower, but the comparison is still possible and useful, especially when the predictions are far away from the observational data.

Another important issue to be considered is the ground motion level(s) to be used in the evaluation process. The comparison between predictions and observational data can be very different for low ground motion levels and high ground motion levels. Ideally, the evaluation needs to be performed using ground motion thresholds similar to the ground motion levels associated with the return periods that are expected to be calculated. Nevertheless, the number of exceedances associated with high ground motion levels are very low. In such cases, the efficiency of the evaluation process is lower. However, it remains necessary to perform evaluation of PSHA results in the time period when observations are available or, in the case of time independent estimates, for a period of time of the same duration.

4.3. HARMONIZATION OF THE OBSERVATIONAL DATABASE

The observational dataset may include data from different sources, as explained in Section 4.1. For datasets that include indirect estimates, such as macroseismic intensity, an appropriate transformation relation has to be applied to transform PSHA output into observation (or the reverse) suitable for the comparison. Ground motion intensity conversion equations facilitate the conversions between ground motion amplitudes (typically peak ground acceleration or velocity) and shaking intensity. Global or regional ground motion intensity conversion equations have been developed for intensity data measured on several scales including modified Mercalli intensity and the European Macroseismic Scale [25]. Two example studies on the ground motion intensity conversion equations are given in Sections A.1.3.1 and A.1.3.2 of the Annex.

The transformation needs to consider all sources of uncertainties when its outcomes are compared with the PSHA results. This can be achieved in two ways: by taking into account the relevant uncertainty in the evaluation phase or by including the transformation (and related uncertainty) inside the PSHA model. Both the epistemic and aleatory uncertainties associated with the relations used to transform the PSHA outputs (e.g. peak ground acceleration) into the observed parameter (e.g. macroseismic intensity) need to be considered.

The epistemic uncertainty is typically associated with the different conversion relations that can be found in the literature. These relations are also associated with their own uncertainties (i.e. standard deviation).

The general formalization of this procedure has two assumptions [23]. First, it is assumed that the PSHA outcome has the form of the probability density distribution *h(a)* and that observational data are in the form of any other parameter *b*. Then it is assumed that an empirical probability distribution $C(b\,|\,a)$ exists, which determines the probability that the value b is exceeded when the shaking parameter is a. In this case, the hazard function *H(b)* is determined as:

$$H(b) = \int_{a_{min}}^{a_{max}} h(x)C(b|x)\,\mathrm{d}x \tag{15}$$

In this way, the epistemic uncertainty associated with the conversion from a to b is implemented in the new hazard evaluation provided by PSHA to be compared with observational data.

It is of great importance to ensure that observational data have been collected in the same site conditions expected in the PSHA development. PSHA estimates generally refer to a reference soil condition (rigid soil flat outcrop), not necessarily matching the site conditions where the observed data have been obtained. In these cases, corrections have to be applied to PSHA outcomes or to observational data before any comparison is made. For example, if the evaluated PSHA is performed considering reference soil conditions and observation are available for sites located on other soil conditions (stiff soil, soft soil, etc.) an appropriate transformation needs to be implemented in order to transpose PSHA results into observation parameters, considering the response of the site, and propagating uncertainties appropriately. This can be achieved by suitable deconvolution approaches based on a detailed knowledge of the station subsoil layering (e.g. see Ref. [26]) or by applying empirical correction factors (e.g. see Ref. [24]).

In other cases (mainly in which seismic stations are installed in facilities such as, nuclear installations, large dams, oil facilities, etc.), the seismic stations might not be installed at free field but inside a facility. If this is the case, it is necessary to transpose the PSHA results (usually free field) into the appropriate observation location and propagate uncertainties appropriately (especially soil–structure interaction).

Finally, when macroseismic intensity is used as an observation, it is necessary to take into account that the macroseismic intensity is an average parameter on a wide area (versus an accelerometric record which is one point), which could also include site effects.

5. PROCESS AND KEY ASPECTS FOR EVALUATION OF PROBABILISTIC SEISMIC HAZARD ANALYSIS BASED ON OBSERVATIONAL DATA

This section describes the overall process that needs to be followed in order to evaluate the PSHA results based on observational data. As stated in Section 1, the scope of this publication is not to describe how to perform a PSHA, which is outlined in other IAEA publications (e.g. see Refs [2, 27]), but to focus on the evaluation process that may be considered as an additional or complementary step to be implemented in the framework of a PSHA either during its development (typically included in project development tasks or peer reviews) or in a posterior stage. Flowcharts and a sample evaluation review sheet are given in this section in order to help the user to implement the procedure (Figures 8–12).

5.1 EVALUATION OBJECTIVES AND GRADES

The final target of the evaluation process is to improve the confidence in PSHA outputs. The objective is not to prove that a model or an assumption is true, which would be too presumptuous. Therefore, the outcomes of the evaluation could provide the PSHA developers with a reproducible process to update the weights of the prior models or assumptions (and possibly reject some of them considering their likelihood) or even encourage them to reassess or reconsider their choices, assumptions or PSHA relying physics, based on the lessons learnt from the evaluation process.

The practical implementation of the evaluation procedure includes the description of the evaluation objectives, which provides guidelines for managing the dependency on observational data used for the evaluation and the data used to perform the PSHA. There are two main objectives in evaluating PSHA results based on observational data and, depending on the objective, the evaluation process will imply some specific actions, as follows:

- Objective #1 (O1) - Consistency checking: The goal of O1 is to check and confirm that the evaluated PSHA outputs are consistent with the available observational data. Even if the evaluated PSHA followed the state of practice based on applicable standards or guidelines, the outputs (or a part of them) might not be consistent with observational data due to the high complexity of the process and often large range of epistemic uncertainties. In other words, O1 aims to answer the following question: are the assumptions, models or branches of the evaluated PSHA intrinsically consistent with observational data?
- Objective #1 (O2) - Weighting: The goal of O2 is to attribute a weight to any single branch or model of the evaluated PSHA, relying on its likelihood compared to the observational data. This objective can specifically address some of the epistemic uncertainties that are included in the evaluated PSHA. It may be implemented in addition to O1, but in that case specific care relating to the independency between observational data and evaluated PSHA input data is needed. In other words, O2 aims to answer the following question: how likely the prior assumptions, models or branches to be correct?

Different grades in evaluating PSHA results based on observational data are described in Figure 8. These grades, depending on the type of the observational data that are available and the type of output (qualitative or quantitative) that are obtained, may be estimated at different phases of the PSHA and may be improved by including more observations or getting a higher level of evaluation as the PSHA project moves into a more detailed phase. The grading is divided into three categories, G1: ranking; G2: scoring; G3: weighting, based on the nature of the evaluation that can be qualitative or quantitative. G3 is the highest grade that is specifically adapted to fulfil the evaluation objective O2.

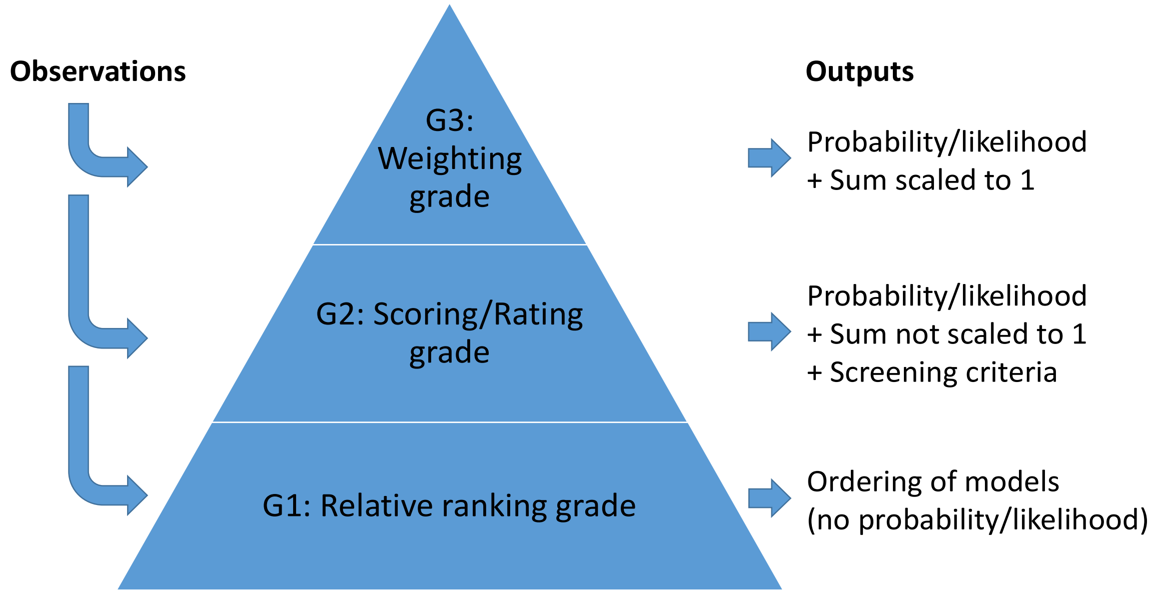

FIG. 8. Different grades in evaluating PSHA models based on observational data.

The first level (G1, lower level) is the relative ranking grade. This level leads to the classification of the logic tree branches and allows to identify which model or assumption better reproduces observational data. This grade level does not provide the quantitative likelihood of each parameter, model, or assumption in accordance with its capability to reproduce observational data.

The second level (G2, intermediate level) may be implemented for the scoring or rating. This level leads to the classification of logic tree branches through the quantitative likelihood of each parameter, model, or assumption in accordance with its capability to reproduce observational data. In the process, some models or assumptions may be screened out when their likelihood appears to be negligible with respect to observational data. If the results indicate that none of the evaluated PSHA models or branches appears to be credible, then the basic general assumptions need to be reconsidered.

The third level (G3, highest level) is the weighting level. This grading leads to weighting of each assumption, model, or branch, based on their consistency with the observational data. The total sum of the weights for the possible models or assumptions under consideration is equal to 1. The output is a posterior weight that will be used to substitute the prior weight.

G1 may be implemented by the PSHA developers (or peer reviewers) as a preliminary step but it is considered neither as sufficient nor necessary in this publication, considering its low informative outcome. G2 is specifically adapted to fulfil the evaluation objective O1 (consistency checking) and G3 is specifically adapted to fulfil evaluation objective O2 (weighting).

5.2 TIMESCALES AND EFFICIENCY OF THE EVALUATION

The timescales relating to the PHSA evaluation process depend on the context and availability of data and have to be considered as a relative value in any case. Depending on the seismic activity of the area under consideration, the amount of observational data of a given ground motion amplitude observed during a certain time period may significantly differ.

Figure 9 shows two sample seismic hazard curves representing sites with low and high seismic hazard levels. The information provided in Figure 9 may be considered in two different ways: (i) looking at the vertical axis provides the change in annual rate of exceedance given the 'constant ground motion level', or (ii) looking at the horizontal axis helps the reader to understand the change in ground motion level, given a 'constant annual rate of exceedance rate'. Compared with a high seismicity area, the hazard curve for a low seismicity area indicates that a relatively lower number of observational data of a given amplitude is expected for future seismic events. In both low seismicity and high seismicity areas, a comparable number of observational data will be collected during a similar time period, each of them with the same exceedance rate, but not the same amplitude. Collecting a comparable number of observations in low seismicity areas with those in high seismicity areas is currently challenging due to lack of observations, but an increase in the number of seismometers will increase the number of observations in future.

Figure 9 illustrates that: instrumental seismicity (with a period of observation of some decades) may typically address the range of a hazard curve from about 10^{-3} to 10^{-2} annual rate of exceedance; historical seismicity (with a period of observation of some centuries) may typically address a range of approximately 10^{-4} to 10^{-3} annual rate of exceedance; and pre-historical seismicity (with a period of observation of some thousands of years, or even more) may typically address a range from about 10^{-5} to 10^{-4} annual rate of exceedance. These values are estimated based on the context and may be adjusted depending also on the geographical scale of the evaluation (especially by gathering multiple sites of observation). Figure 9 also demonstrates that combining multiple types of observation will allow the assessment of a wider range of evaluated PSHA hazard curves, with possible cross and complementary checks.

In principle, the efficiency of the evaluation needs to be quantified by comparing the range of PSHA results and the amount of available observational data. The main parameters to be considered to assess how efficient the evaluation process may be are as follows:

- The median exceedance rate of the ground motion parameter under consideration according to the PSHA;
- The body and range of the distribution of the ground motion parameter under consideration according to the PSHA;
- The actual number of observational data of the ground motion parameter under consideration.

Alternative results are expected, from a very high efficiency of the evaluation (e.g. the body and range of the whole distribution of the PSHA gives consistent predictions, which are also consistent with observational data), to very low efficiency (e.g. the body and range of the whole distribution of PSHA leads to a wide range of predictions, with only a few of them consistent with observational data). Some odd (but possible) cases may also be faced, such as in which the body and range of the PSHA do not give consistent predictions with observational data. In such a case, the outcome of the evaluation process will strongly encourage the PSHA developers to reassess their choices and assumptions.

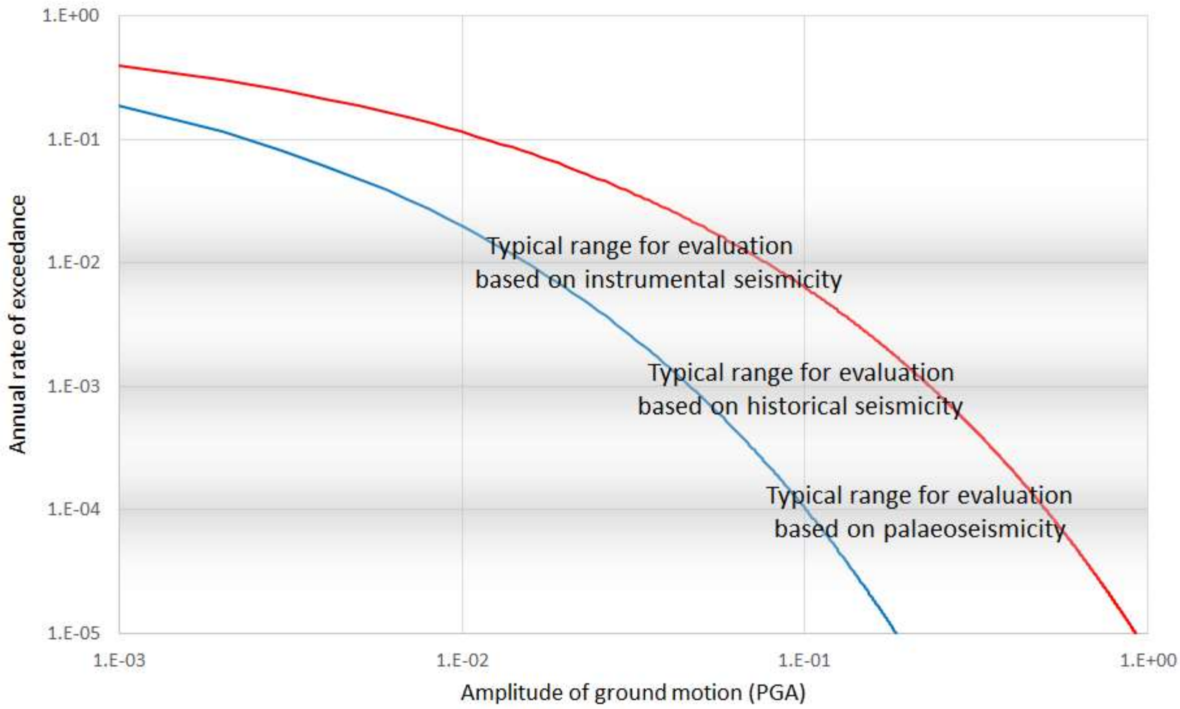

FIG. 9: Illustration of possible timescales for evaluation of PSHA results based on observation (blue curve is a sample hazard curve for low seismicity area and red curve is a sample hazard curve for high seismicity area).

It may be difficult to estimate, a priori, the efficiency of the evaluation. However, the indicator of the informative capability of the evaluation process may be calculated by gathering all available observation windows and calculating the effective cumulated time of observation. A possible indicator is the cumulated period of independent available observations (CPIO) that is calculated based on the number of observation sites and their corresponding periods of observation. The higher the cumulated period of observation is the higher the efficiency of the evaluation may be. A range of possible situations is given below.

- Example case #1 - One observation site with instrumental seismicity based on recording stations with 20 years of operation: In this example, the CPIO is equal to 20 years. This example underlies that relying on only one site of instrumental seismicity will not be informative due to the low period of observation.
- Example case #2 - 40 independent observation sites with instrumental seismicity based on recording stations with 20 years of operation: In this example, the CPIO is equal to 20 x 40 = 800 years. This example indicates that the cumulative period of observations from multiple sites may allow the user to assess the range of the logic tree branches, especially the extreme branches of the logic tree.
- Example case #3 - 10 independent observation sites with historical seismicity records with 300 years of completeness: In this example, the CPIO is equal to 10 x 300 = 3000 years. This example shows that the cumulative period of observations may allow the user to assess the body and the range of the PSHA. However, the transformation relationship between PSHA results and observation data parameter may introduce additional uncertainty and reduce the efficiency of the evaluation process.
- Example case #4 - One observation site with paleo-seismicity records reaching to 5000 years: In this example, the CPIO is equal to 5000 years.

- Example case #5 - Combination of the example cases #2-4: The cumulative CPIO for this case is equal to 8800 years. This example presents the most efficient situation that may allow the user to assess the body and the range of the PSHA with a high degree of confidence.

The examples given above underline that, for instrumental and historical seismicity, it is of great interest to gather multiple observation sites in order to increase the efficiency of the evaluation. If paleo-seismicity observation sites are available (e.g. precarious rocks, brittle stalagmites), the efficiency of the evaluation will be significant, even if only one observation site is available. In any case, the combination of multiple sources and sites of observation is encouraged in order to increase the efficiency of evaluation.

5.3 GEOGRAPHICAL SCALES

The examples given in Section 5.2 emphasize that the efficiency of the evaluation increases with the size of the area and timescales included in the evaluation. Therefore, two different evaluation scales may be introduced (the site area and the regional scale around the site) as shown in Figure 10.

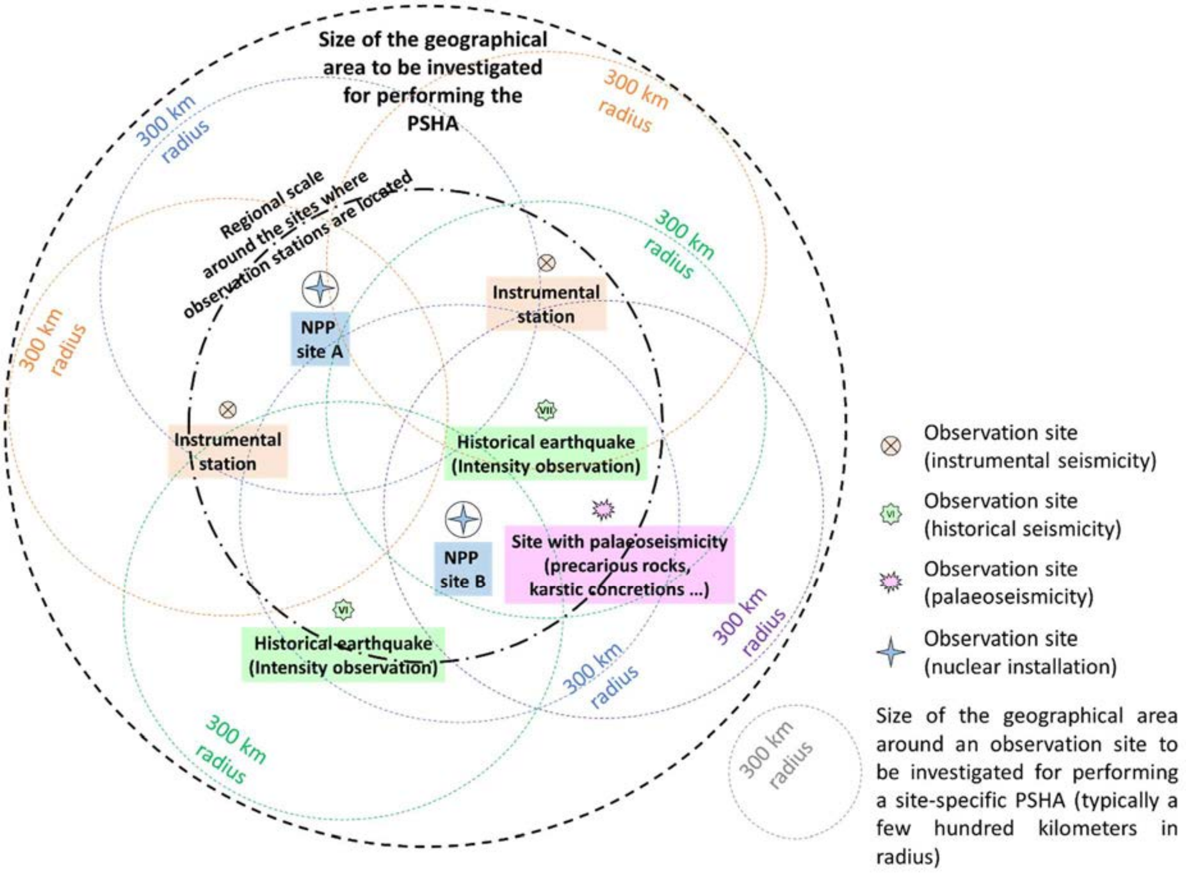

FIG. 10. Different geographical scales to be considered for evaluation.

Regional scale evaluation gathers information from a large area (covering possibly the locations of multiple nuclear installations). The regional scale is especially useful for evaluating the seismic source characterization models and their underlying assumptions. For this purpose, the regional scale would be as large as necessary, depending on the tectonic context and the scope of the evaluated PSHA (typically a few hundreds of kilometres around a site). This allows capturing all relevant seismic sources affecting the seismic hazard calculations at the site of interest. The site area scale evaluation process will address the site specific data and associated uncertainties. This scale is particularly important in the evaluation of elementary steps of PSHA such as the site response models and assumptions, and corresponding uncertainties.

The regional scale evaluation process is particularly suitable for the integral evaluation of PSHA models. It is often the case that multiple nuclear installations are located at relatively short distances (tens or few hundreds of km) from each other compared to the site region to be studied in order to perform a site specific PSHA (which is typically 300 km according to Ref. [8]). The size of the regional scale considered may also be of several hundreds of kilometres. The additional resources needed to develop such a large scale PSHA, covering a wider area than is necessary for a site specific PSHA, may be shared by multiple operating organizations. It is expected that in most cases, this effort would be less costly than the sum of the efforts that would be necessary to develop site specific PSHAs for the nuclear installations under consideration (Figure 10). In the case of a regional scale study, it is encouraged to take all the benefits of observations gathered on multiple sites in order to evaluate the overall PSHA results. In this perspective, the distribution of sites has to be homogeneous, and the possible correlation between the hazard estimates at the sites have to be considered through an appropriate method.

5.4 MANAGEMENT OF INDEPENDENCE OF OBSERVATIONAL DATA COMPARED WITH EVALUATED PROBABILISTIC SEISMIC HAZARD ANALYSIS INPUT DATA

As stated in Section 2, the term ‘available observational data’ refers to the data that was used in the development of a PSHA. However, the term ‘new observational data’ refers to data that was not available at the time of the development of the PSHA study but is now available to assist in the evaluation phase. It is usually necessary that the observational data used for the evaluation is independent from the data used in the PSHA. If the same data were to be used both in the development and evaluation stages, this would give the available observational data an inappropriate weight compared to the new observational data.

The question of dependency between the new observational data and the evaluated PSHA’s input data (available observational data) will depend on the objective of the evaluation process. Due to its goal, the question of dependency between the new observational data and the evaluated PSHA input data has no reason to be considered when objective O1 is targeted. In other words, any observational data can be used to achieve objective O1 without any limitation. In contrast, the question of dependency between the new observational data and the evaluated PSHA input data needs to be considered when the objective O2 is targeted, to avoid any over-weighting or over-fitting. The approach to be applied in this case is described in Figure 11 in a sequential and interrogative way.

The use of new observational data compared to the ones used to perform the PSHA is possible and encouraged. In regions of low seismicity, often the GMPEs used in the evaluated PSHA are adopted from strong motion data from similar tectonic environments.

In those cases, the use of local macroseismic data or even instrumental data would be fully compatible with the above principle (even if they are used, their prior weight would be negligible). In addition, paleo-seismicity observations such as precarious rocks or brittle stalagmites are almost never used in PSHA input data due to their rare occurrences and because these kinds of observational data are evidence of no occurrence of earthquake events or motions (no occurrence information is rarely explicitly taken into account in PSHA development that mainly focuses on characterizing the occurrence of earthquakes).

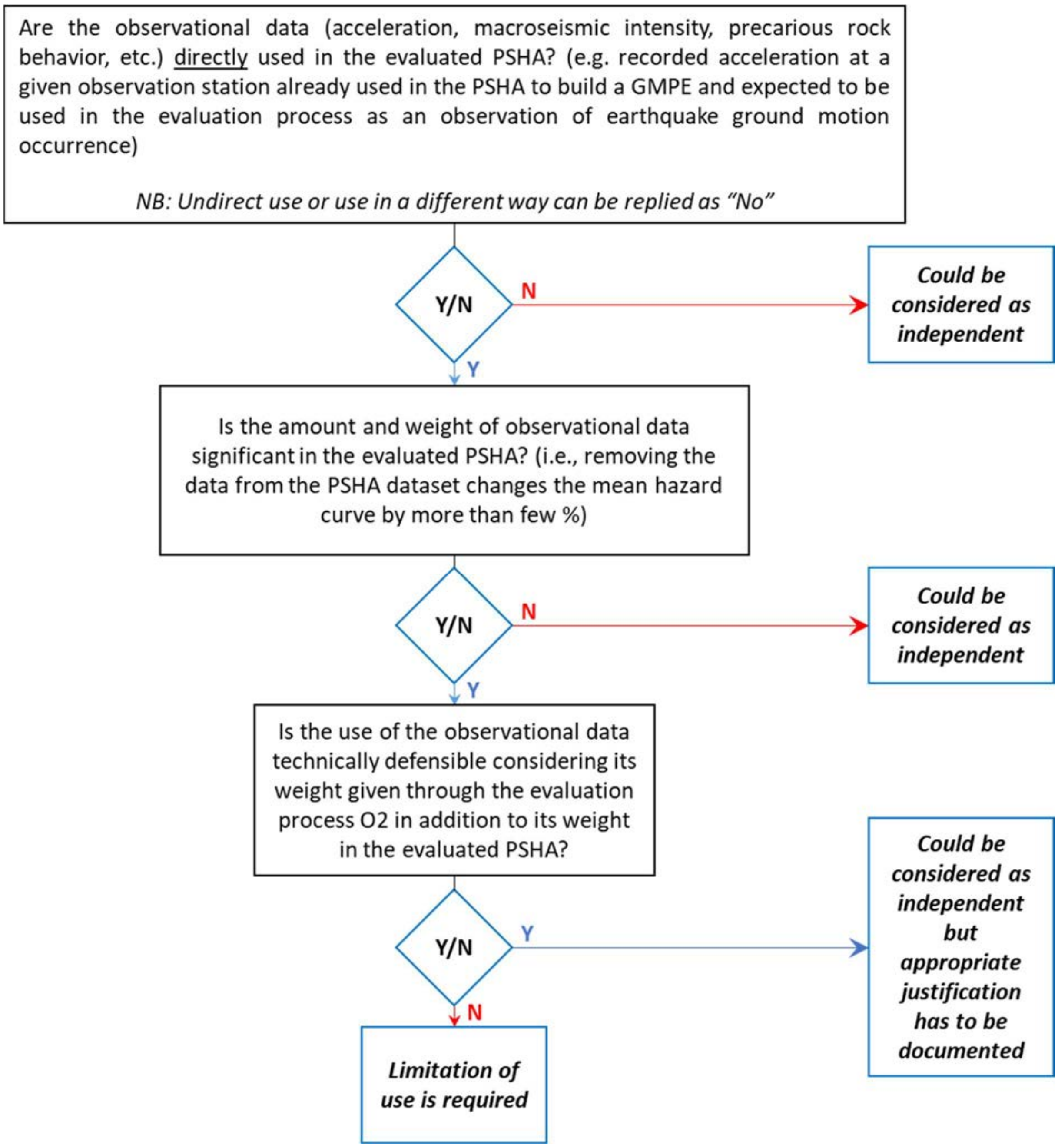

FIG. 11: Management of independence between observation data and the data used in the evaluated PSHA

It is possible to use available observational data in different ways. For example, historical data may be used as an observation of an earthquake ground motion occurrence at a special location for the evaluation process; alternatively, these historical data may contribute to the assessment of the magnitude and location of a given historical earthquake event in the development of the PSHA, which is a completely different approach.

If local seismic data was used in the evaluated PSHA and it is also desired for evaluation, care is needed to avoid double counting or overfitting.

For instance, after providing a well-developed justification that the weights in the evaluated PSHA are very low compared to other data, such input data could also be compatible with the above principle (some dozens of data are usually used for the evaluation compared to thousands of similar data used in the evaluated PSHA).

5.5 EVALUATING ELEMENTS OF LOGIC TREES: THE ELEMENTARY EVALUATION

Most of the elementary steps of the PSHA may be evaluated based on observational data. These elementary steps could be either a physical model, an assumption, or a set of input data. This step comes first in the evaluation process. The elementary steps are typically the ones listed below:

(1) Selection of earthquake databases and catalogues (type of magnitude, possible magnitude conversion, completeness periods, etc.);
(2) Identification of seismotectonic models (geometry of the seismogenic structures, zones of diffuse seismicity, etc.);
(3) Definition and selection of earthquake occurrence models (model parameters, maximum magnitudes, etc.);
(4) Selection of vibratory ground motion parameters (selection of GMPEs, propagation of aleatory and epistemic uncertainties, etc.);
(5) Any other steps that include epistemic uncertainty.

For these elementary steps, any aleatory or epistemic uncertainty need to be handled through a weighting or ranking process following the principles given in Section 5.2 (using G1 or G2, or ideally G3). A priority can be attributed to the actions described in this section based on the results of a sensitivity analysis. Several examples for elementary evaluation are provided in Section A.1 of the Annex.

5.6 EVALUATING RESULTS OF PROBABILISTIC SEISMIC HAZARD ANALYSIS: THE INTEGRAL EVALUATION

The overall results of the PSHA need to be covered by the integral evaluation phase based on observational data in order to verify at the end of the PSHA process that the results obtained are consistent with observational data (including the centre, body and range of the evaluated PSHA). This step is at the end of the evaluation process and is considered as a higher level compared with the elementary step of evaluation. The integral evaluation follows the principles listed below:

(a) Applying the evaluation methods and tools defined in Section 3;
(b) Relying on all available observation, as defined in Section 4;
(c) Assessing separately any single epistemic branch of the PSHA logic tree.

At the end of the integral evaluation, depending on the objective of the evaluation, a grade is attributed to any of the evaluated PSHA models, assumptions or branches based on the observational data. Ideally, the weight of each single branch of the PSHA is updated based on its likelihood in accordance with its consistency with observational data (O2, G3). Examples for integral evaluation are provided in Section A.2 of the Annex.

5.7 FEEDBACK TO PROBABILISTIC SEISMIC HAZARD ANALYSIS DEVELOPERS

As already stated, when conducted during the PSHA development process, PSHA evaluations provide a feedback mechanism to the developers to allow them to assess the weights in their models or assumptions. This will also allow them to reject some of the input parameters considering their low likelihood or even to encourage them to reassess or reconsider their choices, and assumptions.

This feedback mechanism is ideally implemented during the project development or peer review process. Even if the integral evaluation may also be performed after the PSHA project, it is valuable to perform the evaluation process in close relationship with the PSHA development project.

The consistency checking objective O1 may conclude that some parts of the body and range of the PSHA gives inconsistent results (or more precisely, the results are too unlikely to be considered as technically defensible) compared with observational data. This may encourage the developers of a PSHA to reassess or reconsider their choices, assumptions or the PSHA relying on physics, especially when only a small part of the PSHA body and range may appear to be consistent with observational data. This stage is not finalized until a significant part of the body and range of the PSHA gives consistent results compared with observational data.

In addition to O1 (which is a necessary preliminary step of the process), weighting objective O2 may provide with new weights to assumptions, inputs or branches of the PSHA in accordance with their likelihood compared with observational data. These new weights may directly be attributed to the PSHA logic tree, or may support PSHA developers to re-assess their initial choices. This stage is not finalized until the outcomes of O2 have been considered by the PSHA developers and corresponding decisions have been documented.

Using a review sheet would help the PSHA developers or reviewers to document the results in a structured way. A sample PSHA evaluation review sheet is provided in Figure 12.

Evaluation of PSHA based on observation check-list										
Type of observation used			Geographical scales involved		Management of uncertainties			Grade of evaluation		
Pre-historical	Historical	Instrumental	Site area	regional scale	Conversion equations I / PGA	Different acceleration thresholds	Other transformations	Relative ranking	Scoring / Rating	Weighting

Section	Item	Detail	Y	N	N/A	Evaluation check-list
Evaluated PSHA description						
	Name and reference: ...					
	Scale (site-specific, territory ...): ...					
	Other description items: ...					
	The evaluated PSHA is performed following state of practice, based on applicable IAEA standards, and well documented ?		Y	N	N/A	
	NB: If not, evaluation process might be biased: Consider redoing the PSHA					
Objective of the evaluation (section 4)			Y	N	N/A	
	Consistency checking (O1): to be systematically implemented		✓			
	Weighting objective (O2): to be implemented in case of independent data only					
Observation data used for evaluation (cf. section 3)			Y	N	N/A	
	Type of available dataset (cf. section 3.1)	Pre-historical?				
		Historical?				
		Instrumental?				
	Observation database (cf. section 3.2)	*Documented database available?*				
		Completeness checked?				
		Applicability checked?				
	Management of observations (cf. section 3.3)	*Management of correlation between sites included?*				
		Management of site response included?				
Geographical scale (cf. section 4.2)			Y	N	N/A	
	The site area	*To be detailed as necessary by user*				
	The regional geographic scale around the site(s)	*To be detailed as necessary by user*				
Management of uncertainties (cf. section 4.4)			Y	N	N/A	
	Conversion equations used to convert intensity in accelerations (if historical information is used)?	*To be detailed as necessary by user*				
	Different acceleration thresholds used to evaluate PSHA results?	*To be detailed as necessary by user*				
	Other transformations depending on the nature of observations?	*To be detailed as necessary by user*				
Management of independence of observation data (section 4.7)			Y	N	N/A	
	Dependency between observation data and evaluated PSHA input data checked (to be assessed only if O2 objective is envisaged)?	*To be detailed as necessary by user*				
Separate effect (elementary) evaluation based on observations (cf. section 4.5)			Y	N	N/A	✓ or blank
	Evaluation of seismotectonic models?	*To be detailed as necessary by user*				
	Evaluation of earthquake occurrence models?	*To be detailed as necessary by user*				
	Evaluation of vibratory ground motion parameters?	*To be detailed as necessary by user*				
	Evaluation of any other steps?	*To be detailed as necessary by user*				
Integral (overall) evaluation based on observations (cf. section 4.6)			Y	N	N/A	✓ or blank
	Overall description of the integral evaluation					
	Evaluation performed on any single epistemic branch of the PSHA logic tree (whole body and range of the PSHA)?	*To be detailed as necessary by user*				
	Evaluation performed only on some branches of the PSHA (partial evaluation of the body and rnage of the PSHA), to be detailed?	*To be detailed as necessary by user*				
	Evaluation performed on some centiles of the PSHA (partial evaluation of the body and range of the PSHA), to be detailed?	*To be detailed as necessary by user*				
	Detailed description of the integral evaluation					
	Application of evaluation methods and tools defined in section 2.2?	*To be detailed as necessary by user*				N/A
	Evaluation consistent with physical models and assumptions of the PSHA under consideration?	*To be detailed as necessary by user*				N/A
	Propagation of aleatory and epistemic uncertainties at each step of the testing or ranking process?	*To be detailed as necessary by user*				N/A
Feedback to PSHA assumptions, inputs and weithing process (section 4.7)			Y	N	N/A	
	Status of the evaluation in the overall PSHA development process (should reply "YES" at least once to the following questions)					
	Is the evaluation done by the PSHA developpers ?	*To be detailed as necessary by user*				=> If no "Y" to any of the 3 questions, the evaluation process has to be stoped and considered as uncompleted
	Is the evaluation done by the Peer review?	*To be detailed as necessary by user*				
	Is the evaluation done by a dedicated independent team?	*To be detailed as necessary by user*				
	Main outcomes and feedback from the evaluation process					
	Did consistency checking stage (O1) conclude that at least a significant part of the body and range of the PSHA gives consistent results compared with observational data?	*To be detailed as necessary by user*				=> If answer is "N", the whole PSHA logic tree & assumptions have to be reassessed and evaluation process has to be stoped and considered as failed => If answer is "Y", next steps can be processed
	Did consistency checking stage (O1) concludes that all assuptions or branches of the PSHA are considered as consistent with observational data?	*To be detailed as necessary by user*				=> If answer is "N", unconsistent assumptions and branches have to be removed from the PSHA (or their weight drastically reduced through O2) => If answer is "Y", next steps can be processed
	Did weighting objective (O2) take into consideration the feedback of the consistency checking (O1)?	*To be detailed as necessary by user*				=> If answer is "N", evaluation process has to be stoped and considered uncompleted => If answer is "Y", new weights coming from the output of O2 can be taken into consideration
	Did the feedback from stages O1 and O2 explicitely taken in consideration and documented by PSHA actors?	*To be detailed as necessary by user*				=> If answer is "N", evaluation process has to be stoped and considered uncompleted => If answer is "Y", evaluation process can be considered as completed

FIG. 12. Sample review sheet for PSHA evaluation with observational data.

APPENDIX

THE MATHEMATICAL BACKGROUND

This Appendix includes the necessary mathematical background for performing PSHA to ensure that any evaluation based on observations will be consistent with this mathematical background. In addition, mathematical tools that can be used to perform both the integral evaluation and elementary evaluations are discussed.

A.1. EVALUATION TOOLS

The comparison between the results of a PSHA study and available observational data can be performed in several ways (see also Refs [28, 29]), some of which will be outlined below.

A.1.1. The counting methods

The simplest evaluation is computing for a threshold value of a_t, which has been overcome at least once during the respective control period at a number $w(a_t)$ of sites out of N considered. This is considered as the realization R of the seismogenic process. In PSHA models, the number of sites w where the threshold is exceeded is a random variable. Based on the *i-th* PSHA model, the probability distribution relative to this variable can characterized in terms of expectation and its standard deviation. The expected number of sites $\mu(w)$ where the threshold a_t has been exceeded has the form given in Eq. (16):

$$\mu(w) = \sum_{k=1}^{N} H_k \tag{16}$$

(for the sake of simplicity, the dependence of H on the site location k is analysed only) and the respective standard deviation $\sigma(a_t)$ is given in Eq. (17):

$$\sigma(w) = \sqrt{\sum_{k=1}^{N}[H_k(1-H_k)] + 2\sum_{k<q}^{N-1}\left[H_{kq} - H_k H_q\right]} \tag{17}$$

where the probability H_{kq} is the joint probability (correlation) that at the *k-th* and *q-th* sites the threshold a_t has been jointly overcome during t_k. The term H_{kq} is computed (or estimated) on the basis of the model and it depends on the geographical configuration of the N sites considered for the empirical evaluation.

The central limit theorem suggests that the random variable w is normally distributed and thus z is a random variable with a standard Gaussian distribution (Eq. 18).

$$z = \left|\frac{w^0 - \mu(w)}{\sigma(w)}\right| \tag{18}$$

where w^0 is a realization of the variable w. If some kind of uncertainty is associated with the observational data used to compute w^0,, (e.g. when the values considered for the evaluation are converted by empirical rules from direct measurements) this last value will be considered a random variable with expectation $\mu(w^0)$ and standard deviation $\sigma(w^0)$. In this case, Eq. (18) can be re-written as Eq. (19).

$$z = \left| \frac{\mu(w^0) - \mu(w)}{\sqrt{\sigma^2(w^0) + \sigma^2(w)}} \right| \tag{19}$$

Here the key aspect is to compute $\mu(w^0)$ and $\sigma(w^0)$. This can be achieved by using Equations (16) and (17) by substituting H_k with the probability O_k that at the *k-th* site, the threshold a_t has been exceeded. This probability can be computed from the probability distribution modelling uncertainty affecting the *k-th* observation.

The above formulas allow a direct computation of the likelihood $Q(R | m_i)$ in the form given in Eq. (20):

$$Q(R|m_i) = P(z > z^o) \tag{20}$$

where the probability P is computed from the Z standardized Gaussian distribution. Equation (17) emphasizes the role of correlation between exceedances at different sites. When this component is ignored, $\sigma(w)$ results are reduced significantly (and unrealistically). The effect of this bias is to increase the z values and thus the degree of confidence of the model considered. Moreover, since the correlation among the site exceedances depends on the model under consideration, ignoring this effect could severely bias any comparison among different models. In Eq. (18), the possible uncertainty affecting observational data have not been considered. A similar approach can be developed by considering the number of exceedances observed at any number of sites. Further details concerning this approach can be found in Refs [19, 30]. Other approaches are also possible to take into account the mutual dependence among exceedances at different sites in the counting approach (e.g. See Refs [31–34]).

Another procedure based on the counting approach has been provided by Refs [32, 33]. In this case, the counting process includes the following:

1) Some sites (i.e. *1* to *k* sites) are selected where the seismic hazard curves have been defined by the probabilistic models included in the analysis (*i-th* models).
2) For each site, given a fixed a_t, the number of exceedances observed from number of exceedances in station 1 (N_{1obs}) to number of exceedances in station *k* (N_{kobs}), is counted during a fixed period of time, per range of magnitudes used to count the observed exceedances corresponding to the period of completeness defined in the PSHA.
3) The exceedances counted, N_{1obs} to N_{kobs}, were produced only by earthquakes with a magnitude equal to or greater than the minimum magnitude M_{min} used in the PSHA. Otherwise, the predictions (obtained considering M_{min}) and observational data would not be comparable.
4) For each site, from *1* to *k* sites, the predicted number of exceedances of the same acceleration threshold, N_{qpred} to N_{kpred} is counted using the seismic hazard curves defined at each site (which give annual exceedance rates of acceleration thresholds).
5) The addition of observed exceedances at different sites, as shown in Eq. (21),

$$N_{obs} = \sum_{i=1}^{i=k} N_{iobs} \tag{21}$$

is compared with the addition of predicted exceedances at same sites, as shown in Eq. (22).

$$N_{pred} = \sum_{i=1}^{i=k} N_{ipred} \tag{22}$$

This comparison between N_{obs} and N_{pred} can be used as a scoring rule. Roughly, the score is higher when the values of N_{obs} and N_{pred} are closer. If normalized, these scores can also be used to define the weights of the different branches of the logic trees.

A.1.2. The likelihood method

The key element of this approach is the direct estimation of the probability $Q(R|m_i)$ that the set of observational data *R* is a possible realization of the *i-th* PSHA model m_i. As stated in Eq. (10), this element is just a part of the assessment and, in principle, no decision can be made in the lack of comparative estimates (via the denominator of Eq. 10). When $Q(R|m_i)$ results are quite low, the model m_i could be considered as unconfirmed by the observational data and the relevant degree of confidence can be lowered. This approach can be applied in several ways. The simplest one is the following:

- The set of observational data R is collected at a number N of sites where seismicity has been monitored during a control period spanning over t_o years.
- It is observed that at a number of sites $w^o(a_t)$ out of the *N* considered, any ground motion threshold a_t, has been exceeded at least once.
- In the assumption that exceedances at the N sites are mutually independent, the likelihood $Q(R|m_i)$ can be estimated in the form given in Eq. (23):

$$Q(R|m_i) = \left\{\frac{N!}{(N-w^0)!\,w^0!}\right\}\left\{\prod_{k=1}^{w^0(a_t)} H_k\right\}\left\{\prod_{k=w^0(a_t)+1}^{N} (1-H_k)\right\} \tag{23}$$

where H_k represents the hazard value relative to threshold a_t associated with the *k-th* site by the m_i PSHA. The first factor accounts for the possible combinations of $w^o(a_t)$ and N-w^o sites and makes comparable likelihood values computed over different dataset sizes. The second factor includes the sites where the threshold a_t, has been exceeded at least once during the control period and the third factor includes the remaining sites.

Suitable evaluating procedures can be based on this methodology (e.g. see Refs [30, 35]). The basic limitation of this apparently straightforward approach is the assumption of mutual independence between the *N* realizations of the random process. This may be the case when the sites under consideration are sufficiently far away one from each other. The amount of this distance depends on the GMPEs considered and the topology of the accelerometric network and needs to be determined on a case-by-case basis (e.g. Ref. [36]). Otherwise, a correction term needs to be considered; however, this cannot be easily defined by a generalized approach.

A.1.3. Scoring rules

Scoring rules have a long lasting tradition in evaluating weather forecasts (e.g. see Ref. [37]). Scoring rules depend on the specific form assumed by the forecast (binary, alarm levels or probabilistic by following the classification proposed by Ref. [38]). Recently, a class of strictly proper scoring rules [13], the 'gambling scores', has been proposed for seismological applications [38].

To compute the score $S(m_i)$ relative to any *i-th* model m_i, a reference model m^* is first considered. At any *k-th* site, out of an overall number of *N* sites, the m^* model provides the exceedance probability $p_0 = H_t(a|m^*)$ for the control period *t*. At the same site, the *i-th* model provides the exceedance probability $p_i = H_t(a|m_i)$. If an earthquake occurred at the *k-th* site during *t*, the *i-th* model gains the score given in Eq. (24).

$$S_i = -(1 - p_i) + p_i \frac{(1 - p_0)}{p_0} \tag{24}$$

Otherwise, the gain is given by Eq. (25).

$$S_i = (1 - p_i) \frac{p_0}{(1 - p_0)} - p_i \tag{25}$$

The overall score $S(m_i)$ (the fixed-odd score in Ref. [38]) is obtained by summing all the scores obtained by the *i-th* model for the *N* considered sites. The reference model here plays the role of the "house" for the betting game and the score is the "gain" relative to the *i-th* layer. It can be demonstrated that the score rule is strictly proper [10], $S(m^*) = 0$ and the higher is the score the better the model performs. In the case that no reference model exists, one can assume the score is given by Eq. (26).

$$p_0 = \frac{1}{M} \sum_{i=1}^{M} p_i \tag{26}$$

In this case, $S(m_i)$ is defined as the "parimutuel gambling score' and it may demonstrate how to optimize the properties [39].

A.1.4. Elementary evaluation

As a preliminary to integral evaluation, elementary evaluation can have several advantages. It offers a precious insight into the different components of the PSHA logic tree and can help better interpret its final parameters. For instance, if a certain PSHA model m_i receives a very weak posterior weight $Q(m_i|R) << 1$, as defined in Eq. (10), elementary evaluations can help to explain why, by identifying the assumptions or sub-models defining m_i that are not supported by the data.

In order to evaluate any element ψ_i of the *i-th* model, it is necessary to evaluate the degree of belief $Q(\psi_i)$ associated with that element. For this purpose, a data set *R* is considered, which includes a number of observed realizations of the stochastic process described by the probability distribution $Q(\psi_i)$. Thus, the likelihood $Q(R|\psi_i)$ can be computed, and this can be used to compute the posterior degree of belief in the element ψ_i by using Eq. (10) with ψ_i instead of m_i. In this case, any prior estimate of $Q(\psi_i)$ will also be necessary. Some examples concerning GMPEs, seismicity rates and maximum magnitudes can be found in Refs [40–42]. Other examples are provided in the Annex. A promising perspective would be to combine both approaches. As shown in Ref. [42], integral evaluation of the full PSHA could in theory be deduced from the elementary evaluations of each step, the integral weight of a complete PSHA branch being obtained as a combination (e.g. the product) of the weights of each element.

REFERENCES

[1] INTERNATIONAL ATOMIC ENERGY AGENCY, Site Evaluation for Nuclear Installations, IAEA Safety Standards Series No. SSR-1, IAEA, Vienna (2019).

[2] INTERNATIONAL ATOMIC ENERGY AGENCY, Seismic Hazards in Site Evaluation for Nuclear Installations, IAEA Safety Standards Series No. SSG-9 (Rev. 1), IAEA, Vienna (2022).

[3] NUCLEAR REGULATORY COMMISSION, Updated Implementation Guidelines for SSHAC Hazard Studies, Rep. NUREG-2213, Office of Standards Development, Washington, DC (2018).

[4] INTERNATIONAL ATOMIC ENERGY AGENCY, Safety Assessment for Facilities and Activities, IAEA Safety Standards Series No. GSR Part 4 (Rev. 1), IAEA, Vienna (2016).

[5] INTERNATIONAL ATOMIC ENERGY AGENCY, Deterministic Safety Analysis for Nuclear Power Plants, IAEA Safety Standards Series No. SSG-2 (Rev. 1), IAEA, Vienna (2019).

[6] COMMITTEE ON THE SAFETY OF NUCLEAR INSTALLATIONS, NUCLEAR ENERGY AGENCY, ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, Workshop on Testing Probabilistic Seismic Hazard Analysis Results and the Benefits of Bayesian Techniques, Pavia (2015).

[7] WELLS, D.L., COPPERSMITH, K.J., New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, Bull. Seismol. Soc. Am. 84 4 (1994) 974–1002.

[8] JOHNSTON, A.C., KANTER, L.R., COPPERSMITH, K.J., CORNELL, C.A., The Earthquakes of Stable Continental Regions, Vol. 1, Assessment of Large Earthquake Potential, Final Report, EPRI-TR-102261-V1, Electric Power Research Inst., Palo Alto, CA (1994).

[9] MATTHEWS, M.V., ELLSWORTH, W.L., REASENBERG, P.A., A Brownian Model for Recurrent Earthquakes, Bull. Seismol. Soc. Am. 92 6 (2002) 2233–2250.

[10] FUJIWARA, H. et al., Statistical Comparison of National Probabilistic Seismic Hazard Maps and Frequency of Recorded JMA Seismic Intensities from the K-NET Strong-motion Observation Network in Japan during 1997–2006, Seismol. Res. Lett. 80 3 (2009) 458–464.

[11] FRANKEL, A., Mapping Seismic Hazard in the Central and Eastern United States, Seismol. Res. Lett. 66 4 (1995) 8–21.

[12] ORESKES, N., SHRADER-FRECHETTE, K., BELITZ, K., Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences, Science 263 5147 (1994) 641–646.

[13] GNEITING, T., RAFTERY, A.E., Strictly Proper Scoring Rules, Prediction, and Estimation, J. Am. Stat. Assoc. 102 477 (2007) 359–378.

[14] MAK, S., CLEMENTS, R.A., SCHORLEMMER, D., Validating Intensity Prediction Equations for Italy by Observations, Bull. Seismol. Soc. Am. 105 6 (2015) 2942–2954.

[15] HACKING, I., Logic of Statistical Inference, Cambridge Philosophy Classics, Cambridge University Press, Cambridge (2016).

[16] MARZOCCHI, W., ZECHAR, J.D., JORDAN, T.H., Bayesian Forecast Evaluation and Ensemble Earthquake Forecasting, Bull. Seismol. Soc. Am. 102 6 (2012) 2574–2584.

[17] SAATY, T.L., The Analytic Hierarchy Process in Conflict Management, Int. J. Confl. Manag. 1 1 (1990) 47–68.

[18] KULKARNI, R.B., YOUNGS, R.R., COPPERSMITH, K.J., "Assessment of Confidence Intervals for Results of Seismic Hazard Analysis", Proc. of Eighth World Conference on Earthquake Engineering, Vol. 1, San Francisco (1984) 263–270.

[19] ALBARELLO, D., D'AMICO, V., Scoring and Testing Procedures Devoted to Probabilistic Seismic Hazard Assessment, Surv. Geophys. 36 2 (2015) 269–293.

[20] HOETING, J.A., MADIGAN, D., RAFTERY, A.E., VOLINSKY, C.T., Bayesian Model Averaging: A Tutorial, Stat. Sci. 14 4 (1999) 382–401.

[21] ROBERT, C.P., CASELLA, G., Monte Carlo Statistical Methods, Springer Texts in Statistics, Springer New York, New York, NY (2004).

[22] ALBARELLO, D., PERUZZA, L., Accounting for spatial correlation in the empirical scoring of probabilistic seismic hazard estimates, Bull. Earthq. Eng. 15 6 (2017) 2571–2585.

[23] ALBARELLO, D., PERUZZA, L., D'AMICO, V., A scoring test on probabilistic seismic hazard estimates in Italy, Nat. Hazards Earth Syst. Sci. 15 1 (2015) 171–186.

[24] ALBARELLO, D., Detection of Space and Time Heterogeneity in the Completeness of a Seismic Catalog by a Statistical Approach: An Application to the Italian Area, Bull. Seismol. Soc. Am. 91 6 (2001) 1694–1703.

[25] GALLAHUE, M., ABRAHAMSON, N., New methodology for unbiased Ground-Motion Intensity Conversion Equations, Bull. Seismol. Soc. Am. 113 3 (2023) 1133-1151

[26] CASTELLARO, S., ALBARELLO, D., Reconstructing seismic ground motion at reference site conditions: the case of accelerometric records of the Italian National Accelerometric Network (RAN), Bull. Earthq. Eng. 15 6 (2017) 2377–2391.

[27] INTERNATIONAL ATOMIC ENERGY AGENCY, Diffuse Seismicity in Seismic Hazard Assessment for Site Evaluation of Nuclear Installations, IAEA Safety Report Series No. 89, IAEA, Vienna (2016).

[28] RHOADES, D.A., VAN DISSEN, R.J., DOWRICK, D.J., On the handling of uncertainties in estimating the hazard of rupture on a fault segment, J. Geophys. Res. Solid Earth 99 B7 (1994) 13701–13712.

[29] SCHORLEMMER, D., GERSTENBERGER, M.C., RELM Testing Center, Seismol. Res. Lett. 78 1 (2007) 30–36.

[30] SCHORLEMMER, D., GERSTENBERGER, M.C., WIEMER, S., JACKSON, D.D., RHOADES, D.A., Earthquake Likelihood Model Testing, Seismol. Res. Lett. 78 1 (2007) 17–29.

[31] IERVOLINO, I., GIORGIO, M., CITO, P., The effect of spatial dependence on hazard validation, Geophys. J. Int. 209 3 (2017) 1363–1368.

[32] HUMBERT, N., PSHA Updating Technique with a Bayesian Framework: Innovations, CSNI Workshop on Testing PSHA Results and Benefit of Bayesian Techniques for Seismic Hazard Assessment, Eucentre Foundation, Pavia (2015).

[33] SECANELL, R., MARTIN, C., VIALLET, E., SENFAUTE, G., A Bayesian methodology to update the probabilistic seismic hazard assessment, Bull. Earthq. Eng. 16 6 (2018) 2513–2527.

[34] WOO, G., The Mathematics of Natural Catastrophes, World Scientific (1999) 306 pp.

[35] ALBARELLO, D., D'AMICO, V., Testing probabilistic seismic hazard estimates by comparison with observations: An example in Italy, Geophys. J. Int. 175 3 (2008) 1088–1094.

[36] ESPOSITO, S., IERVOLINO, I., PGA and PGV Spatial Correlation Models Based on European Multievent Datasets, Bull. Seismol. Soc. Am. 101 5 (2011) 2532–2541.

[37] JOLLIFFE, I.T., STEPHENSON, D.B., Forecast Verification: A Practitioner's Guide in Atmospheric Science, John Wiley & Sons (2012) 304 pp.

[38] ZHUANG, J., Gambling scores for earthquake predictions and forecasts, Geophys. J. Int. 181 1 (2010) 382–390.

[39] ZECHAR, J.D., ZHUANG, J., A parimutuel gambling perspective to compare probabilistic seismicity forecasts, Geophys. J. Int. 199 1 (2014) 60–68.

[40] KELLER, M. et al., A Bayesian Methodology Applied to the Estimation of Earthquake Recurrence Parameters for Seismic Hazard Assessment, Qual. Reliab. Eng. Int. 30 7 (2014) 921–933.

[41] BERTIN, M., MARIN, S., MILLET, C., BERGE-THIERRY, C., Using Bayesian model averaging to improve ground motion predictions, Geophys. J. Int. 220 2 (2020) 1368–1378.

[42] KELLER, M., MAYOR, J., DUVERGER, C., SENFAUTE, G., Optimal Use of Bayes' Method for the Computation of Seismic Hazard Curves and Application to PSHA Testing, Workshop on the Evaluation of Probabilistic Seismic Hazard Analysis on the Basis of Observations, Saclay, Paris (2018).

ANNEX

EXAMPLE APPLICATIONS

The purpose of this Annex is to present several examples of PSHA evaluations conducted in some Member States that are generally consistent with the methods that are described in this publication. Information is presented for specific case studies that give useful examples of application of the evaluation approach. Case studies do not focus on one single approach but give different ways to achieve the objective of evaluating PSHA based on observational data. This section first presents some case studies relating to the elementary step evaluation and afterwards presents some case studies relating to the integral step evaluation. In this latter case, separation is made between areas with low to moderate seismicity and areas with high seismicity. Each case study is presented through a similar format including a short summary, an overall description of the case study in relation to the related section of this publication, and an illustration.

A–1. CASE STUDIES ON ELEMENTARY STEPS

A–1.1. Seismotectonic models

A–1.1.1. Seismotectonic models: example 1

Overall description of the case study:

- Type of observational data used: instrumental and historical data.
- Geographical scale: regional scale around the site under consideration.
- Implementation of uncertainties: random uncertainties propagated through generation of synthetic catalogues.
- Other effects: N/A.
- Considerations of correlation and independence issues: N/A.
- Evaluated studies:13 possible seismic prior source models are evaluated.
- Corresponding evaluation objectives and grades referring to this publication: O2, G3.
- Main outcomes: 6 posterior seismic source models are selected and weighted.

Specific details are provided in Ref. [A–1]. In summary [A–1] states:

> “We developed a fully non-linear method for a quantitative assessment of seismic source characterization models based on the Metropolis-Hastings Algorithm, one of the Markov-Chain Monte Carlo methods, and the Bayesian inference. We refer to it as the Bayesian Metropolis-Hastings approach. This allows us to sample many potential models compatible with the available data, model the key components of the source model jointly rather than individually, and capture their uncertainties. It also provides an objective and reproducible way to define the source zone model weighting used for the PSHA. Therefore, it can be used to both retrospectively test the seismic source characterization model used for a PSHA or to develop new seismic source characterization models. Expert judgements are still necessary for the seismic source characterization model, but the Bayesian Metropolis-Hastings approach will help to allow rapid considerations of the suitability of various source zone models to the observed seismicity in the region under investigation. We test the Bayesian Metropolis-Hastings approach by applying this to a series of possible seismic source characterization models developed for the Wylfa

Newydd nuclear site, one of several proposed sites for new nuclear power plants in the United Kingdom.”.

An illustration of the results is given in FIG. A–1.

This work		Arup	
SZM	Weight	SZM	Weight
1	0.17	7	0.25
4	0.18	8	0.20
5	0.16	9	0.25
6	0.17	10	0.25
7	0.18	11	0.05
8	0.14		

FIG. A–1. Thirteen possible seismic prior source models (left) and the final six posterior models and weights (right) (reproduced from Ref. [A–1] with permission).

A–1.1.2. Seismotectonic models: example 2

Overall description of the case study:

- Type of observational data used: instrumental and historical data.
- Geographical scale: United Kingdom.
- Implementation of uncertainties: random uncertainties propagated through generation of synthetic catalogues.
- Considerations of correlation and independence issues: N/A.

- Evaluated studies: 2 possible seismic prior source models are evaluated.
- Corresponding evaluation objectives and grades referring to this publication: O1, G2.
- Main outcomes: 1 prior source model is rejected.

Specific details are provided in Ref. [A–2]. In summary, Ref. [A–2] states:

> “In this paper a method is demonstrated in which large numbers of synthetic earthquake catalogues that match the completeness thresholds of the historical catalogue, are generated. The study area can be divided into a grid of uniform cells, and the number of earthquakes in each cell in both the historical catalogue and each simulated catalogue are then counted. Comparison of the historical pattern and a set of 1,000 simulated patterns, using a χ^2 test, shows if the historical pattern is credibly a member of the set of outcomes obtainable from the seismic source model. A second method is to chart the distribution of a large sample of simulated catalogues in terms of magnitude frequency and observe whether the historical catalogue is comfortably within this distribution, or an outlier…A worked example is presented here for the UK, using a source model that was used in Global Seismic Hazard Map, compared to one that was artificially constructed to be credible but flawed. Two tests find the Global Seismic Hazard Map model to be an acceptable representation of the pattern of seismicity in the United Kingdom, while the artificial model is conclusively rejected.”

An illustration of the results is given in Figure A–2.

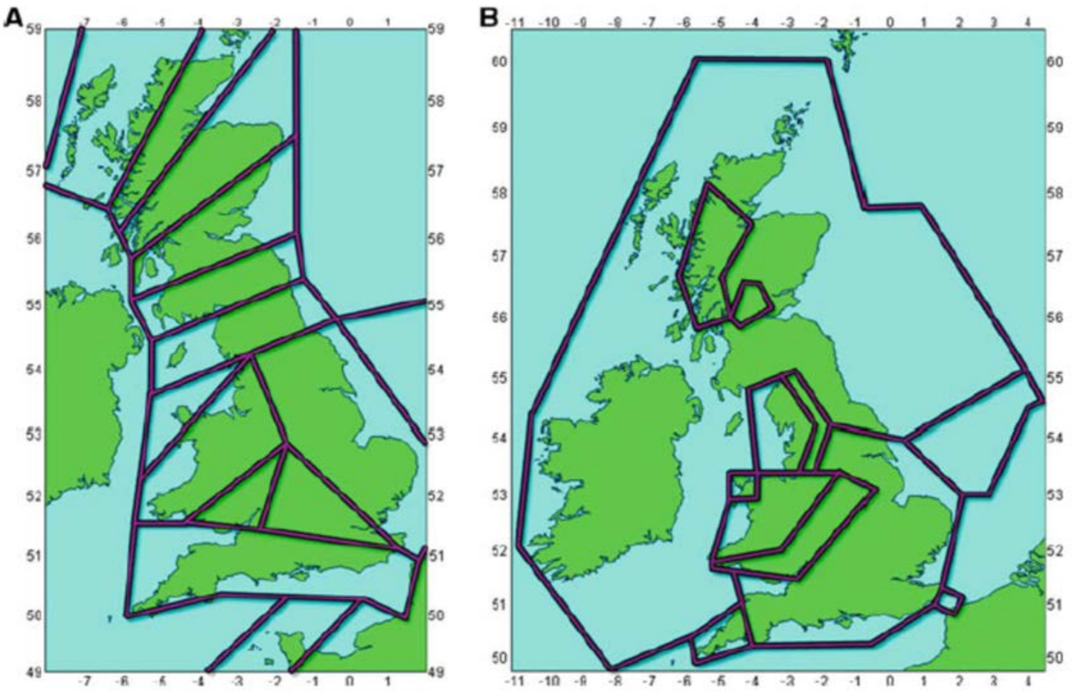

FIG. A–2. Two possible seismic source models for United Kingdom were tested. Model A was rejected after an evaluation process while model B was judged as acceptable (reproduced from Ref. [A–2] with permission).

A–1.2. Earthquake occurrence models

A–1.2.1. Updating the earthquake occurrence model in time domain based on Bayesian estimation (based on Refs [A–3 to A–7])

Overall description of the case study:

- Type of observational data used: macroseismic observational data.
- Geographical scale: Japan.
- Implementation of uncertainties: aleatory variabilities and epistemic uncertainties affecting both observational data and estimations of the future events occurrences.
- Considerations of correlation and independence issues: correlation among parameters for earthquake occurrence models in time domain.
- Corresponding evaluation objectives and grades referring to this publication: O2, G3.
- Main outcomes: weighting earthquake occurrence models based on Bayesian predictors.

This study uses a Bayesian predictive model to calculate occurrence probability in time domain. The model developed is characterized by using event slip and geological slip rate data combined with historical earthquake occurrence data. This study utilized fault activity data on more than 110 active faults in Japan where the Brownian Passage Time model is adopted. The coefficient of variation is updated by using a Bayesian approach. This method makes it possible to compute earthquake occurrence probabilities in a more unbiased and stable way.

A–1.2.2. Evaluation of Gutenberg-Richter occurrence models using observed and synthetic catalogues

Overall description of the case study:

- Type of observational data used: seismic catalogue.
- Geographical scale: zone where the seismic distribution law is considered (i.e. France in Ref. [A–8]).
- Implementation of uncertainties: epistemic uncertainty is mainly associated with the method used to define the seismic distribution law ([A–9, A–10]) and to the Monte Carlo random sampling if it is used (i.e. Ref. [A–8]).
- Considerations of correlation and independence issues: correlation between seismic parameters needs to be considered (a and b parameters in Gutenberg Richter relation, for example, are correlated).
- Corresponding evaluation objectives and grades referring to this publication: O1/O2, G1/G2/G3.
- Main outcomes: ranking and weighting based on a comparison between the predicted seismic activity and observational data.

This study uses methods that have been used in the past to fit the Gutenberg-Richter law, least squares, maximum likelihood methods (i.e. Refs [A–9, A–10]) or Bayesian methods (i.e. Ref. [A–11]). Several methods exist to model the seismic parameters of the characteristic model. Sometimes, for zones of diffuse seismicity, a combination of Gutenberg-Richter distribution for low magnitudes and the characteristic model for large magnitudes are used. Independently of the seismic distribution used to characterize the seismicity of a seismic source and independently of the method used to define the seismic parameters of the model, the observations (real occurrence rates observed) and predictions (annual occurrence rates predicted by models) need to be compared and the method that provides the better adjustment between predictions and observational data is retained. In some cases, the Monte Carlo random sampling is used in order to generate synthetic catalogues considering the uncertainty on magnitude and location of earthquakes. A Gutenberg-Richter law can be fitted with each synthetic catalogue (i.e. Ref. [A–8]). In other cases, the Monte Carlo random sampling can be used to generate couples of parameters a and b (considering the correlation between both parameters). Independently of the method used, the Gutenberg-Richter laws developed need to be checked and compared with observed data in order to avoid the consideration of unrealistic synthetic data.

Figure A–3 shows an example of the Gutenberg-Richter law defined using Monte Carlo random sampling and its comparison with the observed data. The Gutenberg-Richter seems to have a visually good agreement with the data and therefore, no unrealistic Gutenberg-Richter relationships will be considered in the PSHA. To avoid the use of unrealistic models is an objective of the evaluation of elementary steps of a PSHA.

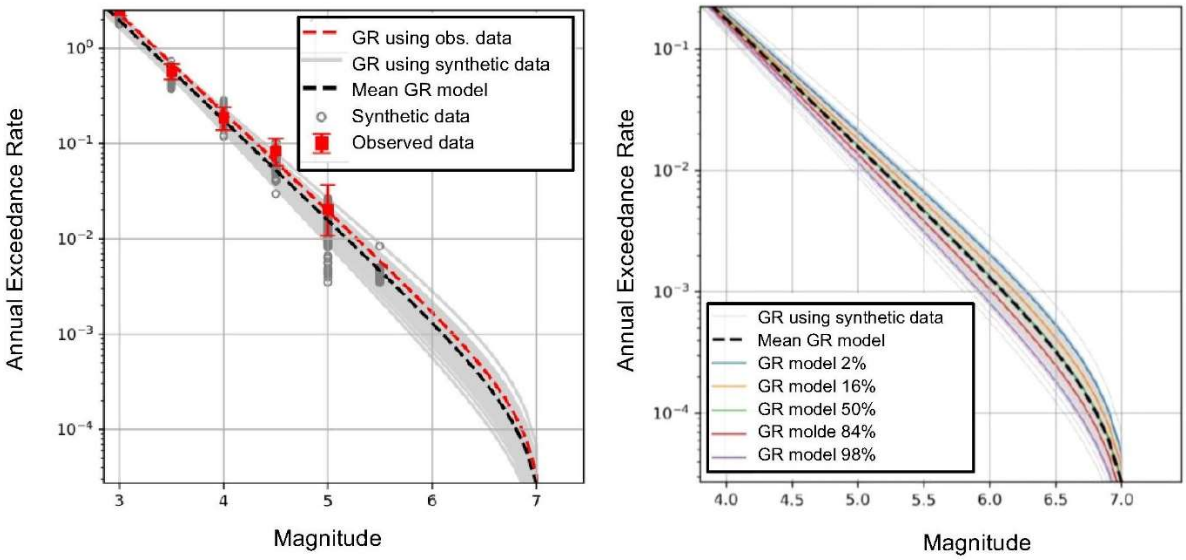

FIG. A–3. Right: Example of observed and synthetic annual exceedance rates (red squares and dark grey circles) with Gutenberg-Richter fit to observed and synthetic data (dashed red and grey lines) and mean Gutenberg-Richter fit from synthetic data (dashed black line). Left: Example of Gutenberg-Richter fit to synthetic data (grey lines) with associated statistics (adapted from Ref. [A–8]).

A–1.2.3. Prospective evaluation of earthquake occurrence models based on instrumental seismicity: example 1

Overall description of the case study:

- Type of evaluation: truly prospective 5-year test of seismicity forecasts.
- Type of observational data used: instrumental seismicity.
- Geographical scale: all of California including a margin of approx. 1 degree around it.
- Implementation of uncertainties: all results were computed with simulations for each target earthquake's position and magnitude considering the respective uncertainties.
- Corresponding evaluation objectives and grades referring to this publication: O1, G2.
- Main outcomes: detailed test results documenting the forecasting power of each model.

The objective of this study was to formalize earthquake forecasts as prospective earthquake rates on a grid for California, to test them in the five years following the project, and to inform the next hazard modelling efforts with the test results to include one or more of the tested models depending on their performance. The models were forecasting earthquakes of magnitudes 4.95 and larger in a well-defined California testing region. Part of the experiment design were unambiguous descriptions of the data source, the data processing, and the evaluation tests. The tests defined encompass several consistency tests: the overall likelihood L-test, the spatial S-test, the magnitude M-test, and number N-test, and the conditional likelihood CL-test [A–6, A–12] as shown in Figure A–4. The three best performing models in this experiment were based on seismicity data only. This leads to the conclusion that on the short term, the seismicity distribution may be a better predictor for future seismicity than other data.

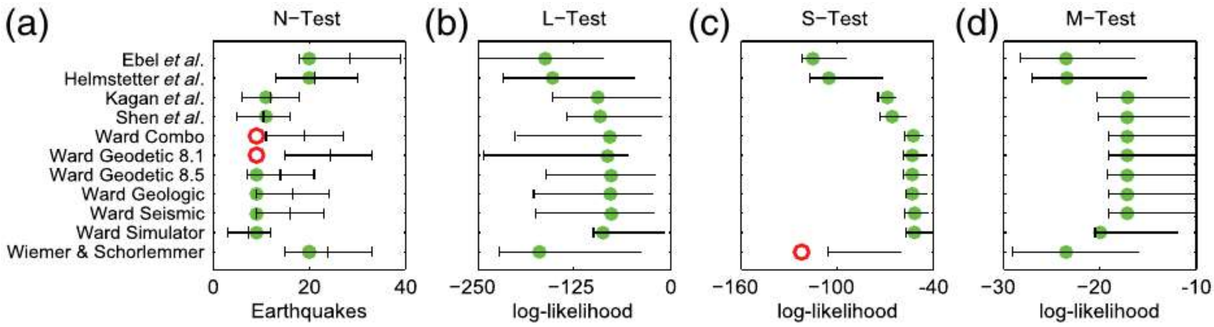

FIG. A–4. Results of consistency tests. Hollow circles indicate that the forecast failed the test, and filled circles indicate that the forecast passed the test. Horizontal black lines delimit the 95% confidence (pass) region. (a) Results of the two-sided N-test, where the forecast number of target earthquakes is indicated as the middle vertical dash on each horizontal line, and the observed number of target earthquakes is indicated by the circle. (b–d) Results of the one-sided L- (S-, M-) test; circles represent the observed space–rate–magnitude (space, magnitude) log-likelihood for each forecast (reproduced from Ref. [A–13] with permission).

A–1.2.4. Prospective evaluation of earthquake occurrence models based on instrumental seismicity: example 2

Overall description of the case study:

- Type of evaluation: truly prospective 5-year test of seismicity forecasts and 40-year retrospective test.
- Type of observational data used: instrumental seismicity.
- Geographical scale: all of California including a margin of approx. 1 degree around it.
- Implementation of uncertainties: all results were computed with simulations for each target earthquake's position and magnitude considering the respective uncertainties.
- Corresponding evaluation objectives and grades referring to this publication: O1, G2
- Main outcomes: detailed test results documenting the forecasting power of each model.

This study focuses on two models, the Uniform California Earthquake Rupture Forecast 2 and the forecast of the National Seismic Hazard Mapping Project. Both models are important because they provide forecasts for the California and national seismic hazard map that inform the governmental agencies and are used to determine the building codes. To complement the truly prospective tests with a longer-term investigation, a 40-year retrospective test was included. The results in Ref. [A–14] show that both models pass all consistency tests for 5-year periods but underestimate the number of observed earthquakes in the retrospective 40-year test as shown in Figure A–5. The model in Ref. [A–15] is outperforming all models. This overall good performance of the model in Ref. [A–15] lead to its inclusion in the next generation of the California forecast (Uniform California Earthquake Rupture Forecast 3).

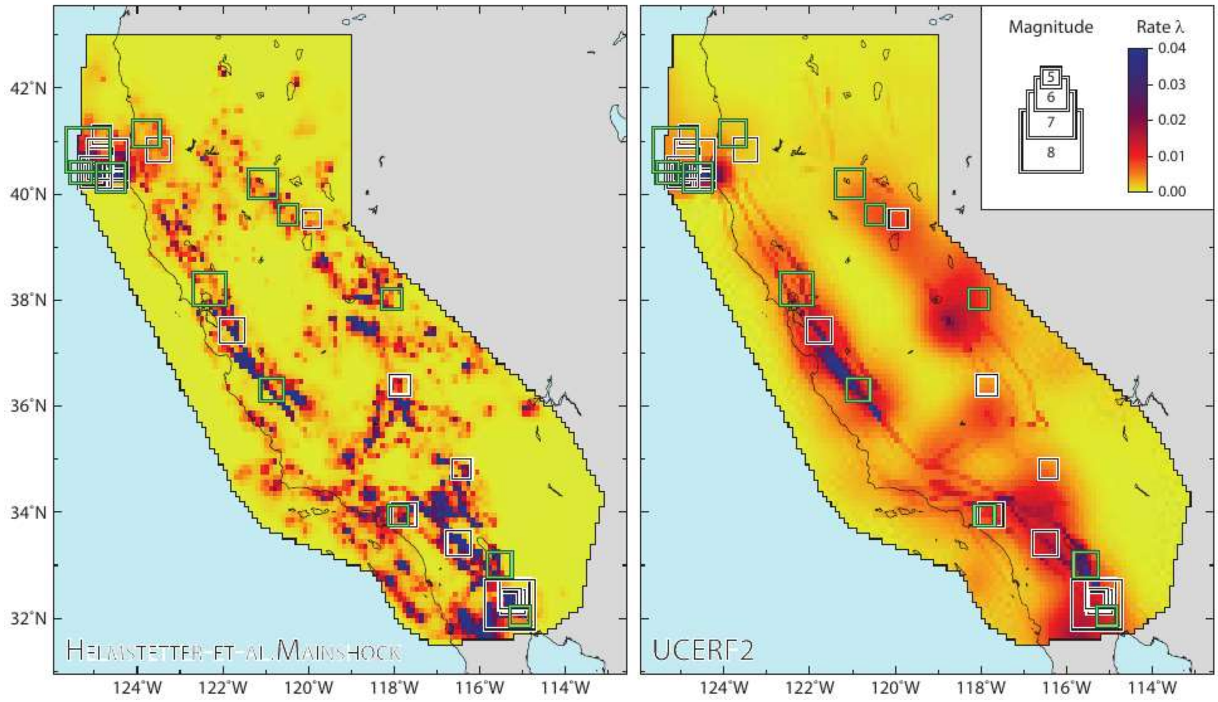

FIG. A–5. Five-year seismicity forecasts of the model in Ref. [A-15] and Uniform California Earthquake Rupture Forecast 2. Purple and dark red regions have elevated seismicity rates, whereas seismicity is lower within yellow areas. Black and green squares indicate the locations of observed earthquakes during both 5-year periods (reproduced from Figure 1 of Ref. [A–14] with permission).

A–1.2.5. Evaluation of maximum magnitude based on historical seismicity

Overall description of the case study:

- Type of evaluation: elementary evaluation of maximum magnitude.
- Type of observational data used: historical seismicity.
- Geographical scale: France.
- Implementation of uncertainties: at every stage of the process.
- Corresponding evaluation objectives and grades referring to this publication: O2, G3.
- Main outcomes: posterior distribution of maximum magnitude.

The work developed in Ref. [A–16] proposes a new approach, based on Bayesian updating and extreme value statistics to determine the maximum magnitudes for truncated magnitude–frequency distributions such as the Gutenberg-Richter model in the framework of PSHA. Only the maximum observed magnitude and the associated completeness period are needed so that the approach is easy to implement and there is no need to determine and use the completeness periods for smaller events. The choice of maximum magnitudes can have a major impact on hazard curves when long return periods, as needed for safety analysis of nuclear power plants, are considered. In this case, not only a singular value but a probability distribution accounting for prior information, data and uncertainty is provided. Moreover, uncertainties relating to magnitude–frequency distributions, including the uncertainty relating to the maximum observed magnitude are discussed and accounted for. The accuracy of the approach is validated based on simulated catalogues with various parameter values. Then the approach is applied to French data for a specific region characterized by high seismic activity in order to determine the maximum magnitude distribution and to compare the results to other approaches.

Figure A–6 shows an example of maximum magnitude distribution using a prior maximum magnitude distribution and a likelihood function. Then, the Bayesian approach is compared with other usual approaches ([A–8, A–17, A–18]) in order to illustrate the necessity of propagating all sources of uncertainties in the process.

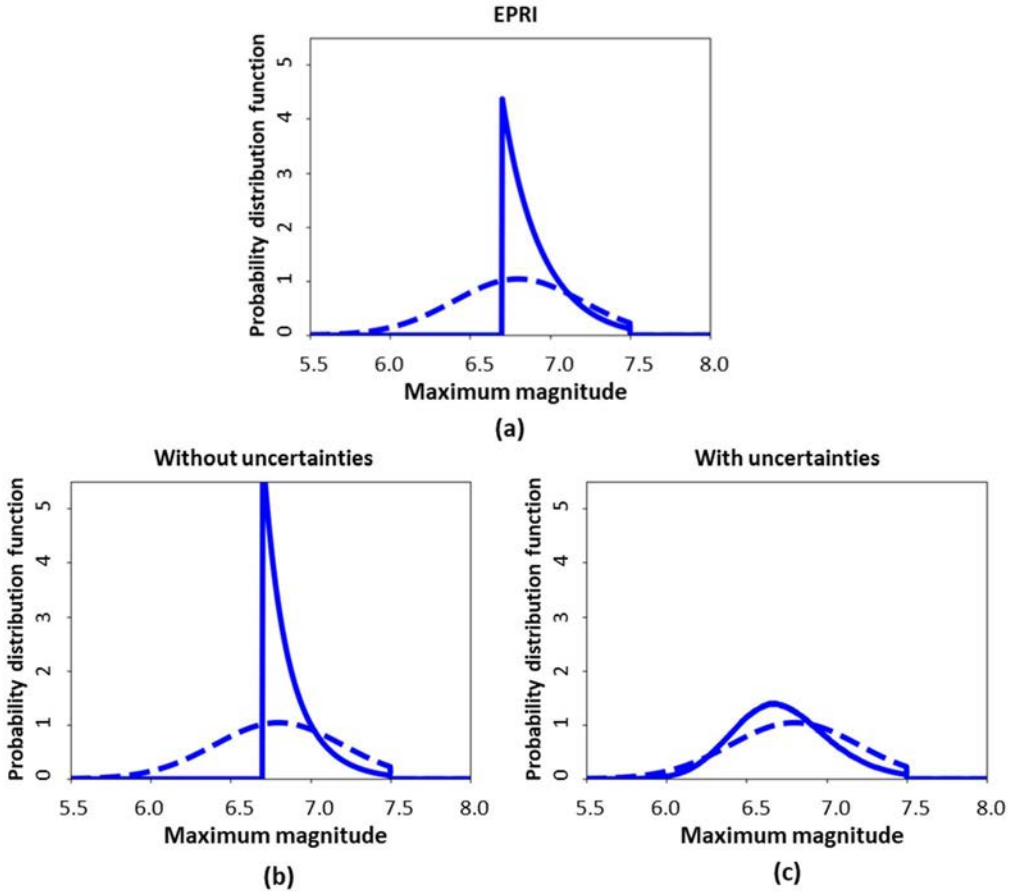

FIG. A–6. Prior and posterior (after Bayesian update) maximum magnitude distributions for a seismotectonic source (a) highlighting the importance of propagating all sources of uncertainties in the process (right) relating to usual approaches with and without uncertainties (b, c) (adapted from Ref. [A–16]).

A–1.2.6. Evaluation of maximum magnitude based on observational data

Overall description of the case study:

- Type of evaluation: retrospective test and analysis about testability of maximum magnitude.
- Type of observational data used: instrumental seismicity.
- Geographical scale: Japan and Switzerland.
- Implementation of uncertainties: standard testing procedure.
- Corresponding evaluation objectives and grades referring to this publication: O1, G2.
- Main outcomes: maximum magnitudes are not testable in a reasonable time frame or with tolerable errors, making them a completely free parameter.

This study investigated whether or not it is possible to at least reject a maximum magnitude within reasonable errors if it cannot be determined. A testing framework for maximum magnitude tests was developed and two case studies were conducted, spanning the possible range of cases in terms of earthquake data availability. The first case, Japan, was chosen because of the best data quality and availability and the rather frequent large earthquakes. The test was carried out based on the maximum available dataset, containing 180 earthquakes with magnitude equal or larger than 7, recorded in the years 684–2012. Even assuming the true maximum magnitude being infinity, the alternative hypothesis of maximum magnitude is equal to 9 is not rejected in 7% of the cases. The second case study covers the north west of Switzerland where four nuclear power plants are located. For this source zone, the study provided the maximum magnitude as a sequence of values with individual probability weights that were tested against each other. The results show that for a significance level of 0.05 and a power of 0.95, the test duration needs to be longer than 3000 years. This is further increased to more than 7000 years for a more powerful test (0.99).

A–1.3. Vibratory ground motion

A–1.3.1. Evaluation of GMPEs in intensity

Overall description of the case study:

- Type of observational data used: macroseismic observational data.
- Geographical scale: whole Italian territory.
- Implementation of uncertainties: aleatory uncertainties affecting both observational data and estimates from attenuation relationships.
- Other effects: incompleteness of the documentation relative to lower intensity values.
- Considerations of correlation and independence issues: possible correlation between stations has not been taken into consideration.
- Corresponding evaluation objectives and grades referring to this publication: O1, G2.
- Main outcomes: sanity check of intensity attenuation relationships.

This study (Ref. [A–20]) provides a procedure for the sanity check relative to an intensity attenuation relationship proposed for Italy in Ref. [A–21]. To this purpose, the counting approach has been considered by accounting for the probabilistic character of the attenuation relationship and the presence of ill-defined intensity attributions both at the epicentre (epicentral intensity) and at the site (site intensity). The database considered for the analysis was the same used for the empirical parameterization of the attenuation relationship under study and included more than 50,000 Mercalli-Cancani-Sieberg scale intensity observational data. The homogeneity and significance of the intensity assessments and derived-source parameters were warranted by selecting only data relative to earthquakes that occurred from 1801 until 1990 whose epicentral locations were determined by the analysis of a macroseismic field including intensity data relative to at least 20 localities. To reduce eventual biases induced by the incomplete reporting of lowest intensities and eventual source directionality effects, the procedure proposed in Ref. [A–22] has been applied. The outcomes of this analysis are summarized in Table A–1. In Table A–1, the expected number of times that an intensity threshold given in the first column has been observed is reported along with the corresponding standard deviation. The corresponding values obtained by considering the attenuation relationship proposed in Ref. [A-21] are reported in the following two columns. In the last column the standardized deviation value computed is reported for each intensity threshold.

TABLE A–1. OUTCOMES OF THE SANITY CHECK PERFORMED ON THE ATTENUATION RELATIONSHIP PROPOSED IN REF. [A-21] FOR THE ITALIAN AREA (ADAPTED FROM REF. [A–20]).

Intensity Threshold	Observed Number	Expected Number	Observed St. Dev.	Expected St. Dev.	Standardized Deviation
VI	11896	11206	16	54	12
VII	6804	6772	22	50	1
VIII	2258	3284	18	41	-23
IX	391	1189	8	29	-27
X	68	292	4	16	-14
XI	2	43	1	6	-6

A–1.3.2. Prospective evaluation of intensity prediction equations

Overall description of the case study:

- Type of evaluation: prospective and retrospective tests.
- Type of observational data used: Italian macroseismic database, “Did You Feel It?”, and “Hai sentito il terremoto?”.
- Geographical scale: Italy.
- Corresponding evaluation objectives and grades referring to this publication: O1, G2.
- Main outcomes: global intensity prediction equations have higher forecasting power than local ones and physics based intensity prediction equations perform better than statistics based intensity prediction equations.

This study (see Ref. [A–23]) selects eight intensity prediction equations for Italy based on Italian data only and one global intensity prediction equation based on global data. Several metrics were developed to score the forecasting performance of intensity prediction equations on several different Italian datasets (see Figure A–7). Some tests were purely prospective while others also included retrospective data (which was partly used in the intensity prediction equation generation).

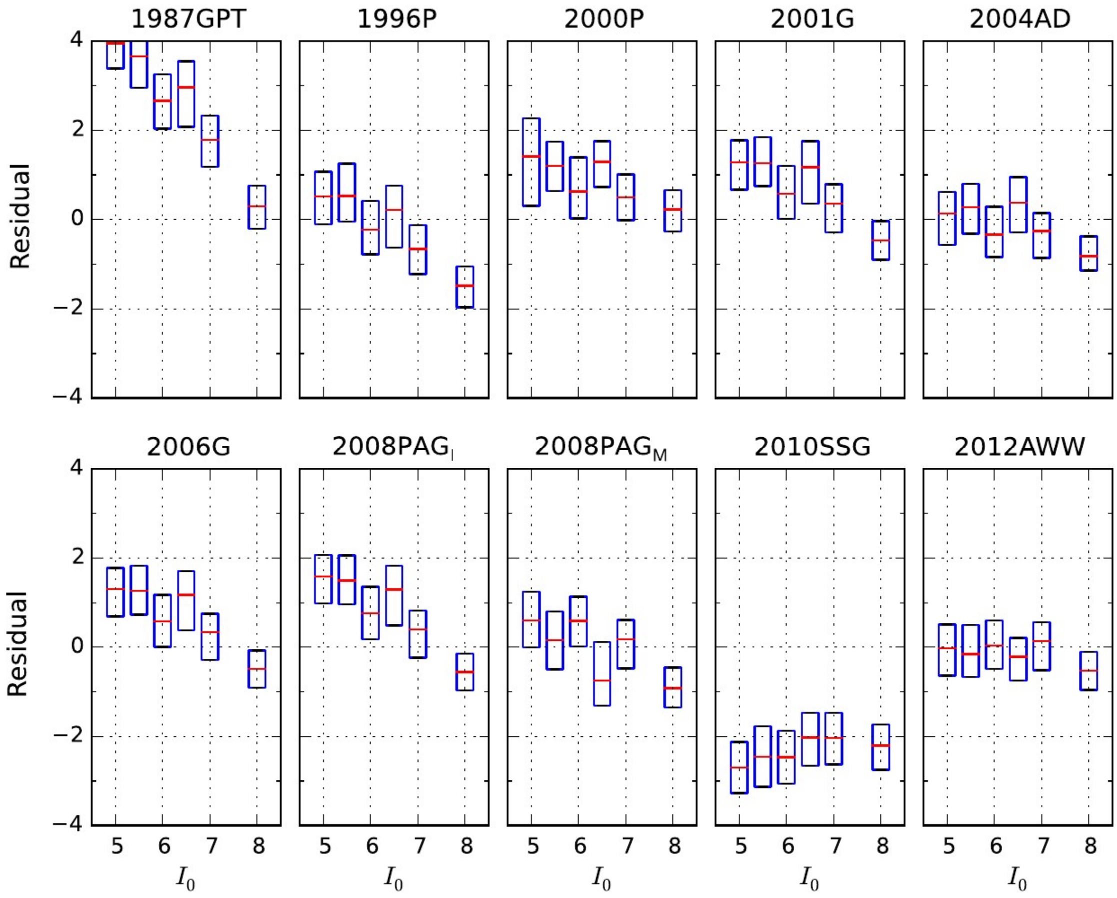

FIG. A–7. Box plots for residuals versus the measured intensity in "Hai sentito il terremoto?" database. Each box represents data of the same I_0 value. The box extends from the lower- to upper-quartile values of the data group, with a horizontal line at the median. The model 2012AWW is the global one showing the smallest residuals (reproduced from Ref. [A–23] with permission).

A–1.3.3. Evaluating GMPEs and their impact on PSHA

Overall description of the case study:

- Type of observational data used: accelerometric data provided by the Italian accelerometric network.
- Geographical scale: whole Italian territory.
- Implementation of uncertainties: aleatory uncertainties affecting attenuation relationships.
- Other effects: site response.
- Considerations of correlation and independence issues: possible correlation between stations has not been taken into consideration.
- Corresponding evaluation objectives and grades referring to this publication: O2, G2.
- Main outcomes: scoring GMPEs proposed for Italy corresponding to a number of average return times.

This study (Ref. [A–24]) empirically evaluated the effectiveness of 12 GMPEs (in peak ground acceleration and response spectra at the periods 0.15s,1.0s and 2.0s) proposed for Italy by jointly considering their impact on PSHA. For this purpose, accelerometric data relative to 56 sites distributed along the whole Italian area that have been operated for at least 25 years have been considered. Shallow subsoil of all these sites has been characterized to classify each of them in terms of the Eurocode 8 soil classification system [A–25]. The maximum peak ground acceleration values recorded at each site during the considered period have been considered for the scoring analysis. In order to score performances of the considered GMPEs, these have been implemented into the same computational procedure based on the Cornell-McGuire approach used for computing the seismic hazard map of Italy [A–6]. Outcomes of PSHA estimates provided by considering the different GMPEs have been scored by the likelihood approach. On this basis, respective performances have been scored by considering several ground motion parameters and mean return times. Outcomes of this analysis (Figure A–8) suggest that recent research concerning more advanced GMPEs parameterization provided a significant improvement of PSHA performances in Italy.

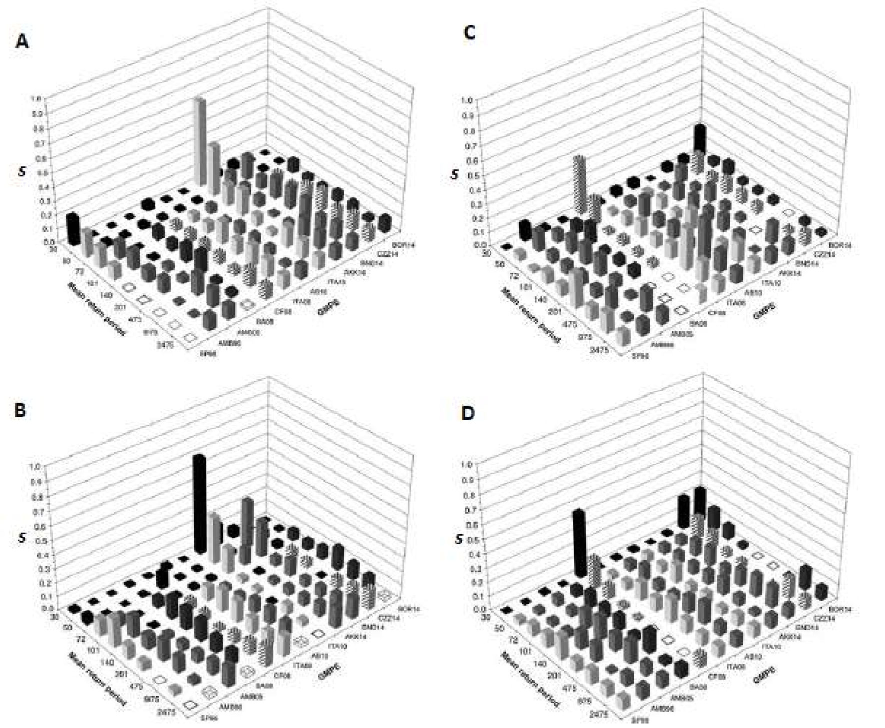

FIG. A–8. Scores (S) assigned to the PSHA models corresponding to the GMPEs marked along the horizontal axis for short period motions: A) peak ground acceleration; spectral acceleration at B) 0.15s spectral period, C) 1s spectral period), D) 2s spectral period. GMPEs pertinent to the same area (or developed by the same working group) are displayed using different patterns. Bins in black and white indicate that models over-predict and under-predict observational data, respectively (reproduced from Ref. [A–24] with permission).

A–1.3.4. Ranking GMPEs based on observational data

Overall description of the case study:

- Type of evaluation: retrospective test and analysis about testability of maximum magnitude.
- Type of observational data used: instrumental seismicity in Refs [A–27 to A–29]; historical data in Ref. [A–23].
- Geographical scale: France, Germany, and Switzerland in Ref. [A–27] but the methodology has been applied to many other regions.
- Implementation of uncertainties: standard testing procedure.
- Corresponding evaluation objectives and grades referring to this publication: O1/O2, G1/G2/G3.
- Main outcomes: ranking scheme and weighting for each GMPE used in a PSHA.

In 2009, the log-likelihood approach developed in Ref. [A–28] supersedes the likelihood technique proposed in Ref. [A–27]. The log-likelihood method is based on the measure of the distance between two probability density functions associated with observational data and predictions of the GMPE. This method was used for the selection and ranking of GMPEs in several studies (i.e. Ref. [A–30]). In some large-scale scientific projects, e.g. the seismic hazard maps for Europe, the log-likelihood method was used in order to define the weights of the different GMPEs selected for the PSHA.

More recently, other authors proposed similar methodologies based on the analysis of residuals to rank and select GMPEs for a PSHA. For example, Ref. [A–29] proposed a new method called Euclidean distance-based ranking. It modifies the Euclidean Distance concept, similar to the residual analysis, for ranking of GMPEs with a given set of observed data.

The application of these methodologies can provide objective criteria to define the weights of the different GMPEs used in a PSHA. However, these methodologies need to be used carefully and their application needs to be based on a good comparison between observational data and predictions. The conventional residual analysis could be yet helpful as a visual tool and can be also useful for the GMPE selection process. An alternative approach to modelling the epistemic uncertainties in PSHA, called the backbone approach, is described in Ref. [A–31] and it has been used in many recent PSHAs.

In the past years, the evaluating/validating methods of GMPEs based on acceleration became popular and different methodologies can be found in the bibliography. However, the validation of intensity prediction equations has attracted less attention. Nevertheless, some publications (i.e. Ref. [A–23]) tried to fill this gap. The evaluation methods of a PSHA described in this publication encourages the use of real data to select the best adapted model for each elementary step in a PSHA. In this sense, the selection of a set of GMPEs is one of the most important decisions in a PSHA. This decision is preferably based on one of the cited ranking methods rather than based on subjective criteria.

Finally, we note that the integral evaluation methods, which are based also in a comparison between observational data and predictions provide an alternative methodology mainly based on Bayesian theorem to define the weights of the GMPEs retained for a PSHA.

A–1.3.5. Other examples

Overall description of the case study:

- Type of observational data used: empirical ground motions, mainly recorded in Japan before the 1995 Hyogoken Nanbu (Kobe) Earthquake, complemented with data from California and other regions for near fault scaling.
- Geographical scale: Japan.
- Implementation of uncertainties: residuals between the predicted and observed peak ground accelerations from the Kobe earthquake have been discussed relevant to site effects. The regression coefficients of the GMPE were updated in 1992 from the publication in 1990 using observed data during this period.
- Consistency checking the GMPE with the observation after development.
- Considerations of correlation and independence issues: possible correlation between stations has not been taken into consideration.
- Corresponding evaluation objectives and grades referring to this publication: O1, G1.

- Main outcomes: major difference has not been seen before and after the update, so that the stability of the result has been validated.

This study (Ref. [A–32]) compared the predicted peak ground accelerations with the collected observed data from the Kobe earthquake as shown in Figure A–9. Then, the authors evaluated the discrepancies between observed and predicted peak ground accelerations and confirmed the consistency although the GMPE was developed without observed data from the Kobe earthquake.

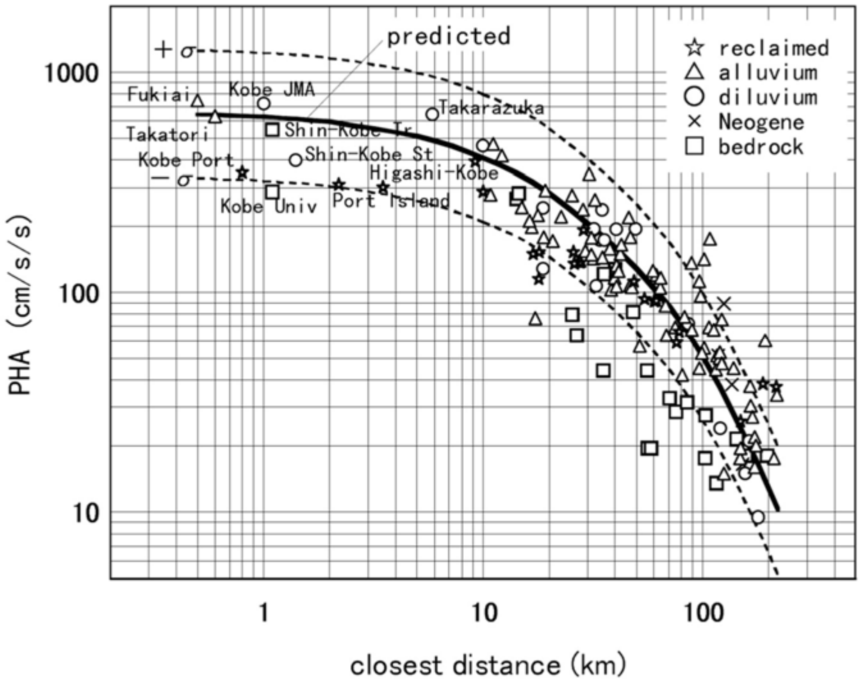

FIG. A–9. Comparison between observed peak horizontal acceleration during 1995 Kobe, Japan earthquake and predicted ones by the GMPE developed before the earthquake (reproduced from Ref. [A–32] with permission).

A–2. CASE STUDIES ON INTEGRAL STEP

A–2.1. Areas with low to moderate seismicity

A–2.1.1. Case 1

Overall description of the case study:

- Type of observational data used: instrumental and historical seismicity.
- Geographical scale: Eastern France.
- Implementation of uncertainties: epistemic uncertainties on data are taken into consideration.
- Other effects: historical seismicity is complemented by undirect observational data.
- Considerations of correlation and independence issues: possible correlation between stations is taken into consideration.
- One PSHA results is assessed and 24 epistemic branches are considered.
- Corresponding evaluation objectives and grades referring to this publication: O2, G3.
- Main outcomes: posterior weight (Bayesian likelihood) based on observational data.

Reference [A-33] provides specific details. In summary, Ref. [A–33] states:

"This paper presents a Bayesian methodology for updating the seismic hazard curves. The methodology is based on the comparison of predictive exceedance rates of a fixed acceleration level (given by the seismic hazard curves) and the observed exceedance rates in some selected sites. The application of the methodology needs, firstly, the definition of a prior probabilistic seismic hazard assessment based on a logic tree. Each main branch corresponds to a probabilistic model of calculus of seismic hazard. The method considers that, initially (or a priori), the weights of all branches of the logic tree are equivalent. Secondly, the method needs to compile the observational data in the region. They are introduced in a database containing the recorded acceleration data (during the instrumental period). Nevertheless, the instrumental period in stable zones (as France) shows only very low acceleration levels recorded during a short observation period. Then, a method to enlarge the number of observational data is presented considering the historical data and defining synthetic accelerations in the sites of observation. The synthetic data allows to expand the period of observation and to increase the acceleration thresholds used in the Bayesian updating process."

An illustration of the results is given in Figure A–10.

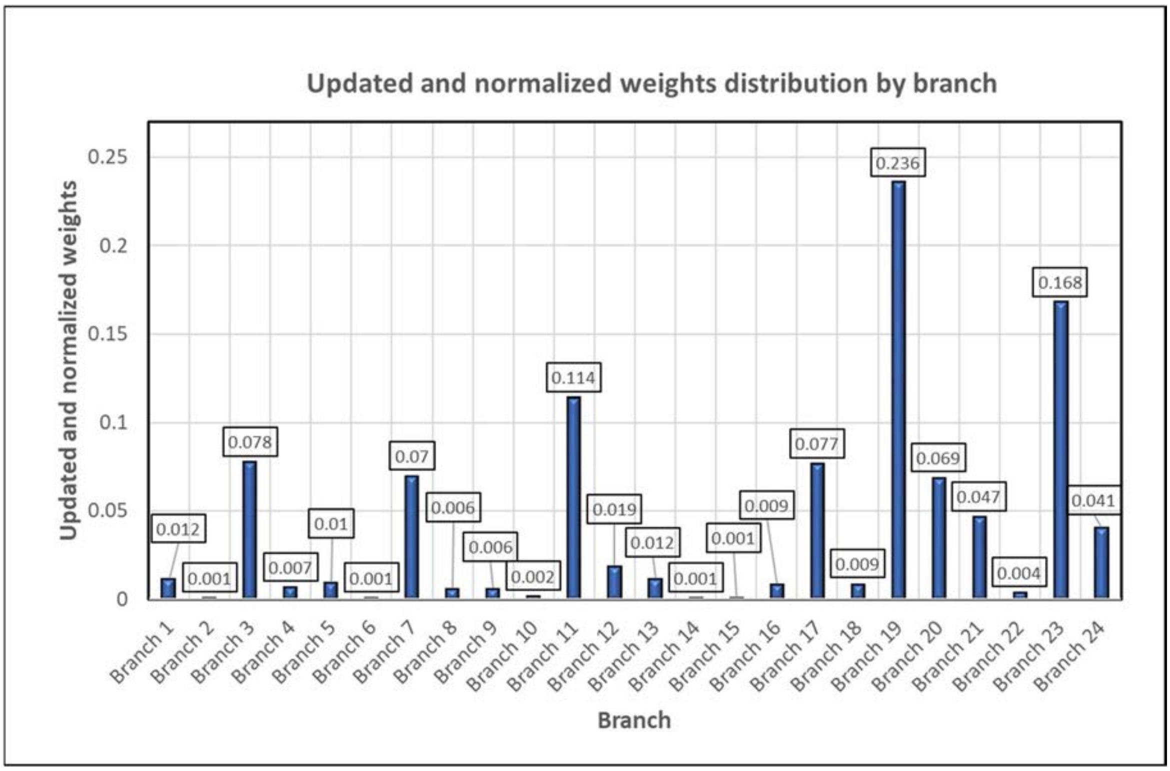

FIG. A–10. A posteriori weights predicted by the Bayesian methodology after comparison of observational data and predictions (24 prior weights were equivalent) (adapted from Ref. [A–33]).

A–2.1.2. Case 2

Overall description of the case study:

- Type of observational data used: instrumental seismicity and historical seismicity (macroseismic Intensity observational data).
- Geographical scale: French metropolitan territory.
- Implementation of uncertainties: epistemic and aleatory uncertainties.
- Other effects: peak ground accelerations to intensity transformation relationships are used, propagating aleatory and epistemic uncertainties.

- Considerations of correlation and independence issues: possible correlation between stations is taken into consideration.
- One PSHA results is assessed: 42 epistemic branches are considered.
- Corresponding evaluation objectives and grades referring to this publication: O2, G3.
- Main outcomes: posterior weight (Bayesian likelihood) based on observational data.

This study (Ref. [A–34]) presented a methodology that can be used to take into consideration instrumental and historical observations in order to improve confidence in PSHA results. The method developed is based on a Bayesian inference technique used to quantify the likelihood of the prior estimation. The period of observation under consideration is the completeness period for each set of observations used. The updating process is developed at a regional scale, over a significant number of stations. The potential correlation between points of observation is also discussed and accounted for. Finally, a case of application to the French metropolitan territory is proposed to demonstrate the efficiency of this updating method and draw perspectives for further applications. The outcome of this approach is to weight the whole PSHA branches in accordance with instrumental and historical observations.

An illustration of the results is given in Figure A–11.

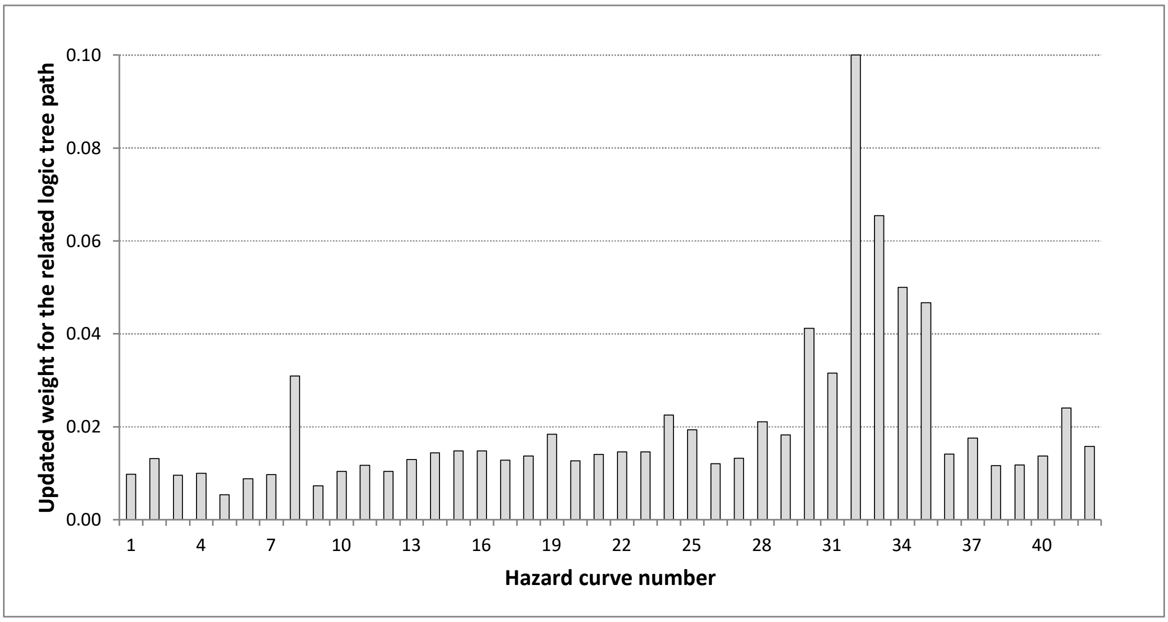

FIG. A–11. Full logic tree including instrumental seismicity and historical seismicity (vs. final likelihood (updated weight) (adapted from Ref. [A–34]).

A–2.1.3. Case 3

Overall description of the case study:

- Type of observational data used: historical earthquakes and intact vulnerable stalagmites in natural caves.
- Geographical scale: a stalagmite in the Plavecka priest cave located in the vicinity of Vienna and Bratislava.
- Implementation of uncertainties: epistemic uncertainties.
- Other effects: the thickness of the rock overburden in the caves.
- Corresponding evaluation objectives and grades referring to this publication: O1, G1.

- Main outcomes: horizontal ground acceleration in the cave and at the surface in different times into the past for comparing with prehistorical earthquakes in the region. The most valuable result is the information for evaluating the maximum intensities (or magnitudes) occurring for a very long timescale.

This study (Ref. [A–35]) proposed a methodology to estimate the upper limit of horizontal peak ground acceleration generated by prehistoric earthquakes using an intact and vulnerable stalagmite. The objective was to determine a critical horizontal ground acceleration curve in different times into the past.

An illustration of the results is given in Figure A–12.

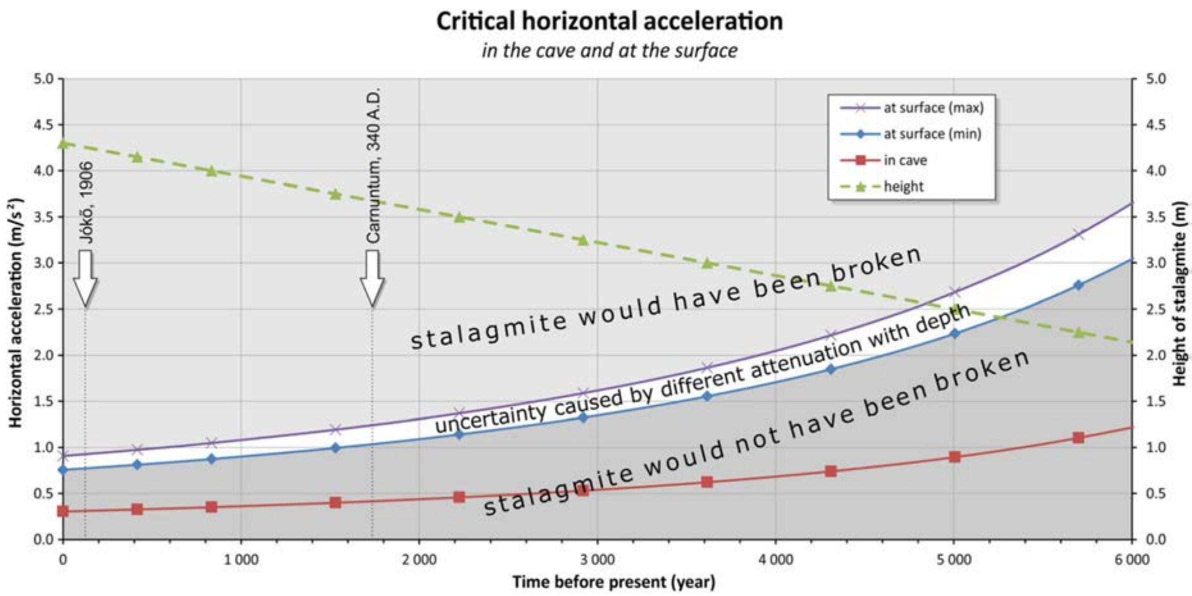

FIG. A–12. Horizontal ground acceleration at the surface and in the cave. Ages of the real historical earthquakes in the area are indicated (reproduced from Ref. [A–35] with permission).

A–2.2. Areas with high seismicity

A–2.2.1. Case 4: Empirical scoring of PSHA estimates [A–36]

Overall description of the case study:

- Type of observational data used: instrumental data from the Italian accelerometric network.
- Geographical scale: Italy.
- Implementation of uncertainties: epistemic and aleatory uncertainties.
- Other effects: site response has been accounted for.
- Considerations of correlation and independence issues: possible correlation between stations is taken into consideration. PSHA results have been assessed by considering different average return times.
- Corresponding evaluation objectives and grades referring to this publication: O2; G2.
- Main outcomes: scoring the PSHA estimates provided by the considered models.

In this study (Ref. [A–36]) hazard outcomes for peak ground accelerations were compared with accelerometric observational data from 76 accelerometric stations operated in Italy for at least

25 years. The counting approach has been used. The likelihood relative to a number of time independent models has been computed. The outcomes of this analysis are reported in Table A–2 and indicates that nearly all the outcomes of hazard estimates provided so far are compatible with observational data; the likelihood associated with these models is always above 5%. However, the performance of the models is significantly different. The performance is different when different average return periods or spectral ordinates are of concern.

In Table A–2, the reference GMPE relative to each model is also reported. On the left q(z*) values (Eq. [A–17]) are reported as a function of the average return time. On the right, corresponding z* values (Eq. [A–16]) are reported (negative values indicate that the model outcomes tend to overestimate actual observational data).

TABLE A–2. RESULTS OF THE SCORING PROCEDURE RELATIVE TO THE SET TIME INDEPENDENT MODELS CONSIDERED FOR HAZARD IN THE ITALIAN PENINSULA. (ADAPTED FROM REF. [A–36]).

PSHA	Likelihood					Standardized Deviation				
Models	73-94	284	475-484	984	2475	73-94	284	475-484	984	2475
Model 1	0.16	0.45	0.62			-1.41	-0.75	0.49		
Model 2			0.93					0.09		
Model 3	0.19		0.62		0.11	-1.32		0.49		1.61
Model 4			0.90	0.52				0.13	0.64	
Model 5			0.90	0.93				0.13	0.09	
Model 6	0.03		0.49		0.28	-2.23		-0.69		-1.08

A–2.2.2. Case 5

Overall description of the case study:

- Type of observational: historical earthquakes and intact or broken speleothems in several natural karstic caves.
- Geographical scale: stalagmites and stalactites in four caves in the Pollino Range (Southern Italy) and one cave in Central Italy.
- Implementation of uncertainties: epistemic uncertainties are considered.
- Considerations of correlation and independence issues: others cause (e.g. ice flow, flooding, animal or anthropic passage, collapses) that can break speleothems are excluded.
- Corresponding evaluation objectives and grades referring to this publication: O1, G2.
- Main outcomes: modelling of stalactite vulnerability is consistent with a fault based seismic hazard model. An ~1 g peak ground acceleration threshold for collapsed speleothems achieved every 5.6 ka during strong and close earthquakes. Capable fault sources responsible for the collapse of a tall stalagmite and for two additional paleo-events are identified.

This study (Ref. [A–37]) carried out a multidisciplinary study to constrain the ground shaking threshold and seismogenic sources in the Pollino Range (Southern Italy), an area with a lack of significant earthquakes in the last ~ 500 years. Implementing this methodology, Ref. [A–37] examined and validated the effectiveness of an ad hoc earthquake source and seismic hazard model (Figure A–13) developed for this particular region of Italy.

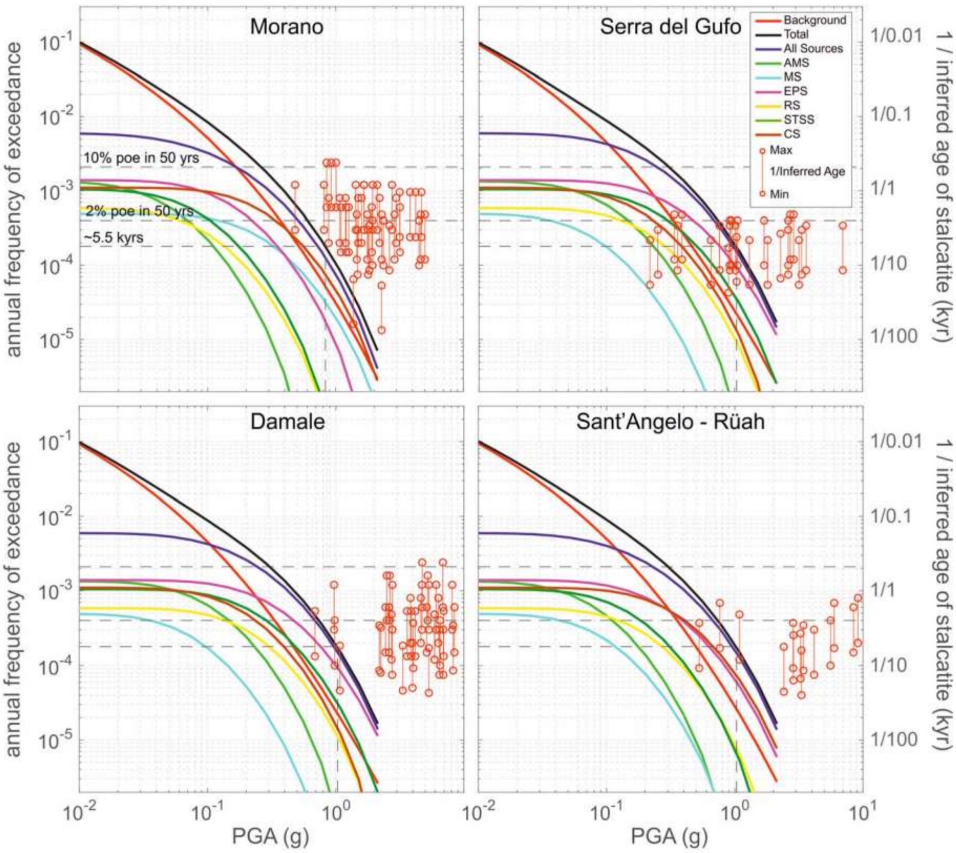

FIG. A–13. Annual frequency of exceedance vs. peak ground acceleration for the four studied caves (Morano, Serra del Gufo, Damale and Sant'Angelo-Ruah). The black curve is the total seismic hazard, obtained summing the contribution of seven earthquake source models. The red circles are the modelled peak ground acceleration of intact speleothems (reproduced from Ref. [A–37] with permission).

A–2.2.3. Statistical comparison of probabilistic seismic hazard maps and frequency of recorded seismic intensity

Overall description of the case study:

- Type of observational data used: instrumental intensities records from more than 1000 seismometers in 30 years.
- Geographical scale: Japan.
- Implementation of uncertainties: aleatory variabilities and epistemic uncertainties affecting both observational data and estimations of the future events occurrences.
- Considerations of correlation and independence issues: correlation among parameters for earthquake occurrence models in time domain.
- Corresponding evaluation objectives and grades referring to this publication: O1, G1.
- Main outcomes: sanity check of earthquake occurrence models.

This study (Ref. [A–4]) conducted a statistical comparison of national seismic hazard maps and the frequency of instrumentally observed seismic intensities as a validation of the maps. The comparison was performed as follows: (i) statistics of all recorded seismic ground motions were calculated, (ii) the probability of exceedance for certain ground motion levels were calculated for 10-year intervals since the maps focused on the upcoming 30 and 50 years, and (iii) the results were compared with the national hazard maps.

Figure A–14 shows the exceedance probabilities that the observed seismic intensity exceeds 5 or lower and 6 or lower.

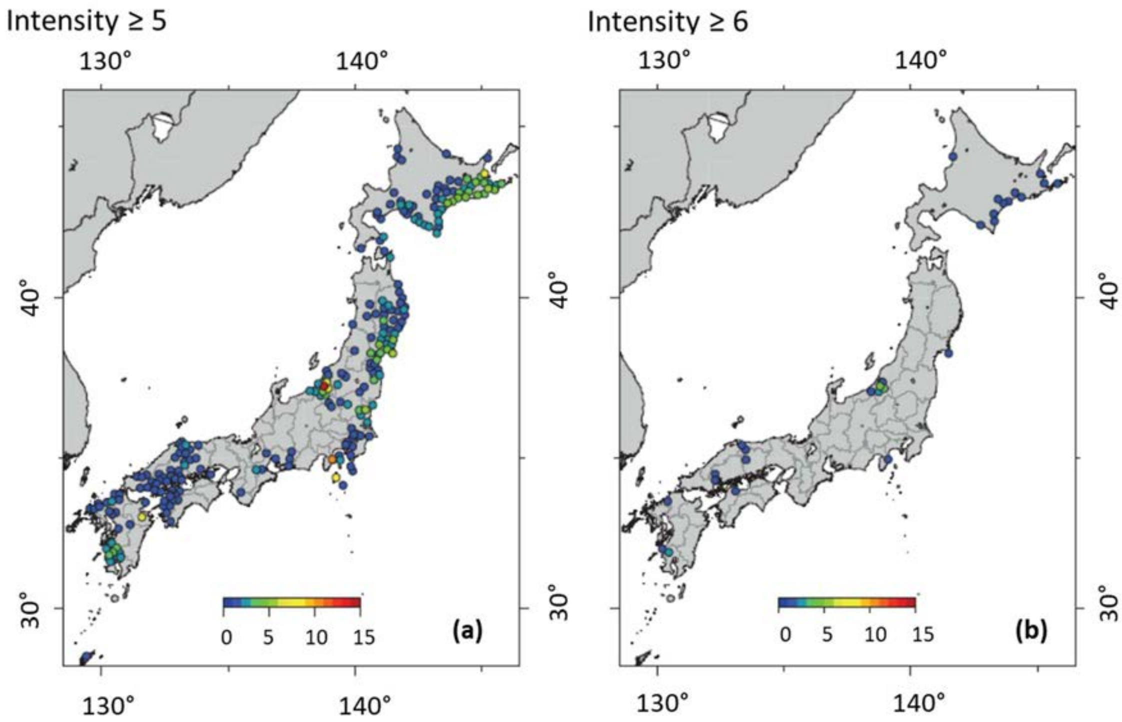

FIG. A–14. Observation stations where the recorded intensities exceeded (a) Japan Meteorological Agency seismic intensity of 5 and (b) Japan Meteorological Agency seismic intensity of 6 from 1 January 1997 to 31 December 2006. Colours indicate the number of times the frequency exceeded certain intensities. Dark dots indicate more than 15 times (reproduced from Ref. [A–4] with permission).

REFERENCES TO ANNEX

[A–1] MOSCA, I., BAPTIE, B., VILLANI, T., COURTNEY, T., Objective Quantification of the Seismic Source Model for Nuclear Sites, SMiRT-25, Charlotte, NC (2019).

[A–2] MUSSON, R.M.W., WINTER, P.W., Objective assessment of source models for seismic hazard studies: with a worked example from UK data, Bull Earthquake Eng **10** 2 (2012) 367–378.

[A–3] MATTHEWS, M.V., ELLSWORTH, W.L., REASENBERG, P.A., A Brownian Model for Recurrent Earthquakes, Bulletin of the Seismological Society of America **92** 6 (2002) 2233–2250.

[A–4] FUJIWARA, H. et al., Statistical Comparison of National Probabilistic Seismic Hazard Maps and Frequency of Recorded JMA Seismic Intensities from the K-NET Strong-motion Observation Network in Japan during 1997–2006, Seismological Research Letters **80** 3 (2009) 458–464.

[A–5] NOMURA, S., OGATA, Y., KOMAKI, F., TODA, S., Bayesian forecasting of recurrent earthquakes and predictive performance for a small sample size, Journal of Geophysical Research: Solid Earth **116** B4 (2011).

[A–6] RHOADES, D.A. et al., Efficient testing of earthquake forecasting models, Acta Geophys. **59** 4 (2011) 728–747.

[A–7] OGATA, Y., Slip-size-dependent renewal processes and Bayesian inferences for uncertainties, Journal of Geophysical Research: Solid Earth **107** B11 (2002) ESE 1-1-ESE 1-14.

[A–8] DROUET, S., AMERI, G., LE DORTZ, K., SECANELL, R., SENFAUTE, G., A probabilistic seismic hazard map for the metropolitan France, Bull Earthquake Eng **18** 5 (2020) 1865–1898.

[A–9] KIJKO, A., SMIT, A., Extension of the Aki-Utsu b-Value Estimator for Incomplete Catalogs, Bulletin of the Seismological Society of America **102** 3 (2012) 1283–1287.

[A–10] WEICHERT, D., Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes, Bulletin of the Seismological Society of America **70** (1980) 1337–1346.

[A–11] KELLER, M. et al., A Bayesian Methodology Applied to the Estimation of Earthquake Recurrence Parameters for Seismic Hazard Assessment, Quality and Reliability Engineering International **30** 7 (2014) 921–933.

[A–12] SCHORLEMMER, D., GERSTENBERGER, M.C., WIEMER, S., JACKSON, D.D., RHOADES, D.A., Earthquake Likelihood Model Testing, Seismological Research Letters **78** 1 (2007) 17–29.

[A–13] ZECHAR, J.D. et al., Regional Earthquake Likelihood Models I: First-Order Results, Bulletin of the Seismological Society of America **103** 2A (2013) 787–798.

[A–14] STRADER, A., SCHNEIDER, M., SCHORLEMMER, D., Prospective and retrospective evaluation of five-year earthquake forecast models for California, Geophysical Journal International **211** 1 (2017) 239–251.

[A–15] HELMSTETTER, A., KAGAN, Y.Y., JACKSON, D.D., High-resolution Time-independent Grid-based Forecast for M ≥ 5 Earthquakes in California, Seismological Research Letters **78** 1 (2007) 78–86.

[A–16] ZENTNER, I., AMERI, G., VIALLET, E., Bayesian Estimation of the Maximum Magnitude mmax Based on the Extreme Value Distribution for Probabilistic Seismic Hazard Analyses, Pure Appl. Geophys. **177** 12 (2020) 5643–5660.

[A–17] JOHNSTON, A.C., KANTER, L.R., COPPERSMITH, K.J., CORNELL, C.A., The Earthquakes of Stable Continental Regions. Vol. 1, Assessment of Large Earthquake Potential, Final Report, EPRI-TR-102261-V1, Electric Power Research Inst. (EPRI), Palo Alto, CA; Memphis State Univ., TN. Center for Earthquake Research and Information; Geomatrix Consultants, Inc., San Francisco, CA; Cornell CA, Portola Valley, CA (1994).

[A–18] DEPARTMENT OF ENERGY, ELECTRIC POWER RESEARCH INSTITUTE, UNITED STATES NUCLEAR REGULATORY COMMISSION Central and Eastern United States (CEUS) Seismic Source Characterization (SSC) for Nuclear Facilities Project, Rep. DOE/NE-0140, Washington, DC. Rep. EPRI 1021097, Palo Alto, CA. Rep. NUREG-2115, Washington, DC. (2012).

[A–19] ZÖLLER, G., HOLSCHNEIDER, M., The Earthquake History in a Fault Zone Tells Us Almost Nothing about mmax, Seismological Research Letters **87** 1 (2015) 132–137.

[A–20] ALBARELLO, D., Validation of Intensity Attenuation Relationships, Bulletin of the Seismological Society of America **95** 2 (2005) 719–724.

[A–21] CARLETTI, F., GASPERINI, P., Lateral variations of seismic intensity attenuation in Italy, Geophysical Journal International **155** 3 (2003) 839–856.

[A–22] GASPERINI, P., The Attenuation of Seismic Intensity in Italy: A Bilinear Shape Indicates the Dominance of Deep Phases at Epicentral Distances Longer than 45 km, Bulletin of the Seismological Society of America **91** 4 (2001) 826–841.

[A–23] MAK, S., CLEMENTS, R.A., SCHORLEMMER, D., Validating Intensity Prediction Equations for Italy by Observations, Bulletin of the Seismological Society of America **105** 6 (2015) 2942–2954.

[A–24] BARANI, S., ALBARELLO, D., MASSA, M., SPALLAROSSA, D., Influence of Twenty Years of Research on Ground-Motion Prediction Equations on Probabilistic Seismic Hazard in Italy, Bulletin of the Seismological Society of America **107** 1 (2017) 240–255.

[A–25] EUROPEAN COMMISSION, EUROCODE 8; Design of Structures for Earthquake Resistance - Part 1: General Rules, Seismic Actions and Rules for Buildings, EN 1998-1:2004, Comite Europeen de Normalisation, Brussels, (2004).

[A–26] BARANI, S., SPALLAROSSA, D., BAZZURRO, P., Disaggregation of Probabilistic Ground-Motion Hazard in Italy, Bulletin of the Seismological Society of America **99** 5 (2009) 2638–2661.

[A–27] SCHERBAUM, F., On the Use of Response Spectral-Reference Data for the Selection and Ranking of Ground-Motion Models for Seismic-Hazard Analysis in Regions of Moderate Seismicity: The Case of Rock Motion, Bulletin of the Seismological Society of America **94** 6 (2004) 2164–2185.

[A–28] SCHERBAUM, F., DELAVAUD, E., RIGGELSEN, C., Model Selection in Seismic Hazard Analysis: An Information-Theoretic Perspective, Bulletin of the Seismological Society of America **99** 6 (2009) 3234–3247.

[A–29] KALE, Ö., AKKAR, S., A New Procedure for Selecting and Ranking Ground-Motion Prediction Equations (GMPEs): The Euclidean Distance-Based Ranking (EDR) Method, Bulletin of the Seismological Society of America **103** 2A (2013) 1069–1084.

[A–30] DELAVAUD, E. et al., Toward a Ground-Motion Logic Tree for Probabilistic Seismic Hazard Assessment in Europe, Journal of Seismology **16** 3 (2012) 451.

[A–31] ATKINSON, G.M., BOMMER, J.J., ABRAHAMSON, N.A., Alternative Approaches to Modeling Epistemic Uncertainty in Ground Motions in Probabilistic Seismic-Hazard Analysis, Seismological Research Letters **85** 6 (2014) 1141–1144.

[A–32] FUKUSHIMA, Y., IRIKURA, K., UETAKE, T., MATSUMOTO, H., Characteristics of Observed Peak Amplitude for Strong Ground Motion from the 1995 Hyogoken Nanbu (Kobe) Earthquake, Bulletin of the Seismological Society of America **90** 3 (2000) 545–565.

[A–33] SECANELL, R., MARTIN, C., VIALLET, E., SENFAUTE, G., A Bayesian methodology to update the probabilistic seismic hazard assessment, Bull Earthquake Eng **16** 6 (2018) 2513–2527.

[A–34] VIALLET, E., HUMBERT, N., MOTTIER, P., Updating a probabilistic seismic hazard assessment with instrumental and historical observations based on a Bayesian inference, Nuclear Engineering and Design **350** (2019) 98–106.

[A–35] GRIBOVSZKI, K. et al., Estimating the upper limit of prehistoric peak ground acceleration using an in situ, intact and vulnerable stalagmite from Plavecká priepast cave (Detrekői-zsomboly), Little Carpathians, Slovakia-first results, J Seismol **21** 5 (2017) 1111–1130.

[A–36] ALBARELLO, D., PERUZZA, L., Accounting for spatial correlation in the empirical scoring of probabilistic seismic hazard estimates, Bull Earthquake Eng **15** 6 (2017) 2571–2585.

[A–37] FERRANTI, L. et al., Speleoseismological Constraints on Ground Shaking Threshold and Seismogenic Sources in the Pollino Range (Calabria, Southern Italy), Journal of Geophysical Research: Solid Earth **124** 5 (2019) 5192–5216.

DEFINITIONS

The definitions given below might not necessarily conform to definitions adopted elsewhere for international use.

Bayesian updating. Formal framework for modifying an existing probability estimate (prior probability) by accounting for information not considered in advance.

ergodicity. A stochastic process is said to be ergodic if its statistical properties can be deduced from a single, sufficiently long, random sample of the process.

likelihood. In a formal sense, it is the probability that a set of observational data is the realization of the probability model of concern. Likelihood is thus estimated by considering the inherent formal properties of the model. In the frame of a Bayesian approach, likelihood is one component of the posterior evaluation of the degree of confidence associated with a model after a confirmation analysis.

macroseismic observational data. Documentary data relating to effects (damages, effects on people, etc.) during earthquakes. These data are usually codified in terms of an intensity value by the use of specific codification prescribed by any macroseismic scale.

observation database. The observational data database is the database of observations that are used in the evaluation process. Its construction has to follow the same procedure as the one used in the evaluated PSHA but may also contain a different set of proxies than the seismological database used for the evaluated PSHA.

posterior evaluation. Assessing the degree of confidence associated with a model by matching model outcomes with observational data. Posterior evaluation corresponds to a sanity check when the match is with data used (at least partially) for calibrating the model and to confirm when the match is with independent data. In Bayesian statistics, the posterior probability is the conditional probability that is assigned to the prior one after taking into consideration the observation.

power of a test. The power of a test in common statistics is defined below:
Consider two alternative hypotheses (H_0 and H_1) to be compared by considering any statistical parameter G. Given a set of observations the statistics G* is computed along with the probabilities $P(G > G^* | H_0)$ and $P(G > G^* | H_1)$, which are representative of the likelihood that G* may occur if the hypotheses H_0 and H_1 are respectively true. A very small threshold P_r (significance level) is assumed and the hypothesis H_0 is rejected if $P(G>G^* | H_0) < P_r$ because the likelihood of corresponding hypothesis is low. In this way, one implicitly assumes that the corresponding probability $P(G | H_1)$ is much larger than P_r. However, this could not be the case. The 'power' of the test is equal to $1- P(G > G^* | H_1)$ and thus low power implies that one is accepting as true an alternative hypothesis characterized by a low likelihood.

prior evaluation. Assessing the degree of confidence associated with a model by considering its internal structure (e.g. the physical soundness of the underlying elements of the model, reliability of the data used for tuning the model, etc.). In Bayesian statistics, a prior probability is the probability distribution that would express beliefs about a parameter of interest before taking into consideration the observation though the Bayesian Theorem.

probability. How likely something is to happen. It is generally represented as a numerical value in the range [0,1] expressing the degree of confidence in the truth of any statement. Probability 0 implies that the relevant statement is considered 'false/not likely at all' and probability 1 implies that it is considered 'true/very likely (certain)'. Any value in between indicates the relative propension towards one of these extreme positions. Probabilities can be assessed in several ways (including subjective attitudes), but to be coherent, these estimates need to conform to a set of formal axioms (probability axioms).

sanity check. A sanity check is usually considered as a basic or simple test whose objective is to assess whether an assumption, a statement or the result of a calculation could possibly be true. This kind of test may allow to screen out obviously false assumptions or models, will not catch every possible error or bias.

scoring. Providing a quantitative measure of performance to the model based on the forecast.

weighting. Is a part of the process inside or outside of logic tree(s), aiming at providing a degree of confidence of an individual assumption or of a set of assumptions relying on observational data. The overall sum of weights is 1. The weighting process needs to be performed at each elementary step and at the integral step of the PSHA.

MATHEMATICAL NOTATION

a	Generic ground motion parameter representative of a single earthquake scenario
a_r	Reference seismic scenario corresponding to a reference exceedance probability and a fixed expected operational period
τ	Operational period of nuclear installation
τ_I	Specific time interval whose duration corresponds to the expected operational period
$h_\tau(a)$	Probability that the scenario *a* occur during the expected operational period τ
$H_\tau(a)$	Unconditional exceedance probability relative to the scenario *a* during the expected operational period τ
H_r	Reference exceedance probability
$H_\tau(a \vert m_i)$	Exceedance probability relative to the scenario *a* computed by the m_i PSHA model
$Q(m_i)$	Degree of confidence (or confidence) associated with the *i-th* PSHA model m_i
$Q(R \vert m_i)$	Probability of the realization *R* in the case that the model m_i actually represents the seismogenic process (likelihood)
$Q(m_i \vert R)$	Degree of confidence in the associated PSHA model m_i in the case after that the data set R of observational data has been matched with the forecasts provided by the model
$Q^*(m_i)$	Prior degree of confidence associated with the model m_i before any match with the set *R* of observational data
R	Data set of observational data to be considered for empirical evaluations of PSHA models
$S(m_i)$	Score associated with the PSHA model m_i
$\lambda_t(a)$	Unitary (annual) exceedance rate corresponding to the ground motion scenario *a*
N	Number of sites considered for evaluations of the PSHA models
M	Number of competing PSHA models
m_i	PSHA model
π	multivariate probability distribution representative uncertainty affecting the parameters of a PSHA model
θ	Parameters of the PSHA model
ψ	Element of the logic tree branch relative to any PSHA model

CONTRIBUTORS TO DRAFTING AND REVIEW

Abrahamson, N.	University of California, Berkeley, USA
Ake, J.	United States Nuclear Regulatory Commission, USA
Albarello, D.	University of Siena, Italy
Contri, P.	International Atomic Energy Agency
Fukushima, Y.	International Atomic Energy Agency
Graizer, V.	United States Nuclear Regulatory Commission, USA
Gulerce, Z.	International Atomic Energy Agency
Nakajima, M.	Central Research Institute of Electric Power Industry, Japan
Schorlemmer, D.	GFZ - German Research Centre for Geosciences, Germany
Seber, D.	United States Nuclear Regulatory Commission, USA
Secanell, G. R.	Fugro France SAS, Spain
Senfaute, G.	Électricité de France, France
Valentini, A.	International Atomic Energy Agency
Viallet, E.	Électricité de France, France

Consultants Meetings

Vienna, Austria: 24 May 2017, 19–20 October 2017, 22–23 February 2018, 5–6 July 2018, 2–3 April 2019, 5–6 September 2019, 13–14 January 2020, (Virtual), 9 February 2021, (Virtual), 16 February 2021, 04–05 July 2022

Technical Committee Meetings

Saclay, France, 3–5 December 2018 (Workshop)

IAEA
International Atomic Energy Agency

No. 27

ORDERING LOCALLY

IAEA priced publications may be purchased from the sources listed below or from major local booksellers.

Orders for unpriced publications should be made directly to the IAEA. The contact details are given at the end of this list.

NORTH AMERICA

Bernan / Rowman & Littlefield

15250 NBN Way, Blue Ridge Summit, PA 17214, USA
Telephone: +1 800 462 6420 • Fax: +1 800 338 4550
Email: orders@rowman.com • Web site: www.rowman.com/bernan

REST OF WORLD

Please contact your preferred local supplier, or our lead distributor:

Eurospan

1 Bedford Row
London
WC1R 4BU
United Kingdom

Trade Orders and Enquiries:

Tel: +44 (0)1235 465576
Email: trade.orders@marston.co.uk

Individual Customers:

Tel: +44 (0)1235 465577
Email: direct.orders@marston.co.uk
www.eurospanbookstore.com/iaea

For further information:

Tel. +44 (0) 207 240 0856
Email: info@eurospan.co.uk
www.eurospan.co.uk

Orders for both priced and unpriced publications may be addressed directly to:

Marketing and Sales Unit
International Atomic Energy Agency
Vienna International Centre, PO Box 100, 1400 Vienna, Austria
Telephone: +43 1 2600 22529 or 22530 • Fax: +43 1 26007 22529
Email: sales.publications@iaea.org • Web site: www.iaea.org/publications

24-03246E

International Atomic Energy Agency
Vienna
www.iaea.org