Nuclear Safety
NEA/CSNI/R(2021)10
June 2024
www.oecd-nea.org

Benchmark on External Events Hazard Frequency and Magnitude Statistical Modelling

OECD
BETTER POLICIES FOR BETTER LIVES

NEA
NUCLEAR ENERGY AGENCY

Nuclear Energy Agency

NEA/CSNI/R(2021)10

Unclassified **English text only**

11 June 2024

NUCLEAR ENERGY AGENCY
COMMITTEE ON THE SAFETY OF NUCLEAR INSTALLATIONS

Benchmark on External Events Hazard Frequency and Magnitude Statistical Modelling

This document is available in PDF format only.

JT03545830

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 38 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Türkiye, the United Kingdom and the United States. The European Commission takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

NUCLEAR ENERGY AGENCY

The OECD Nuclear Energy Agency (NEA) was established on 1 February 1958. Current NEA membership consists of 34 countries: Argentina, Australia, Austria, Belgium, Bulgaria, Canada, Czechia, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, Norway, Poland, Portugal, Romania, Russia (suspended), the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Türkiye, the United Kingdom and the United States. The European Commission and the International Atomic Energy Agency also take part in the work of the Agency.

The mission of the NEA is:

- to assist its member countries in maintaining and further developing, through international co-operation, the scientific, technological and legal bases required for a safe, environmentally sound and economical use of nuclear energy for peaceful purposes;
- to provide authoritative assessments and to forge common understandings on key issues as input to government decisions on nuclear energy policy and to broader OECD analyses in areas such as energy and the sustainable development of low-carbon economies.

Specific areas of competence of the NEA include the safety and regulation of nuclear activities, radioactive waste management and decommissioning, radiological protection, nuclear science, economic and technical analyses of the nuclear fuel cycle, nuclear law and liability, and public information. The NEA Data Bank provides nuclear data and computer program services for participating countries.

Committee on the Safety of Nuclear Installations (CSNI)

The Committee on the Safety of Nuclear Installations (CSNI) addresses NEA programmes and activities that support maintaining and advancing the scientific and technical knowledge base of the safety of nuclear installations.

The Committee constitutes a forum for the exchange of technical information and for collaboration between organisations, which can contribute, from their respective backgrounds in research, development and engineering, to its activities. It has regard to the exchange of information between member countries and safety R&D programmes of various sizes in order to keep all member countries involved in and abreast of developments in technical safety matters.

The Committee reviews the state of knowledge on important topics of nuclear safety science and techniques and of safety assessments, and ensures that operating experience is appropriately accounted for in its activities. It initiates and conducts programmes identified by these reviews and assessments in order to confirm safety, overcome discrepancies, develop improvements and reach consensus on technical issues of common interest. It promotes the co-ordination of work in different member countries that serve to maintain and enhance competence in nuclear safety matters, including the establishment of joint undertakings (e.g. joint research and data projects), and assists in the feedback of the results to participating organisations. The Committee ensures that valuable end-products of the technical reviews and analyses are provided to members in a timely manner, and made publicly available when appropriate, to support broader nuclear safety.

The Committee focuses primarily on the safety aspects of existing power reactors, other nuclear installations and new power reactors; it also considers the safety implications of scientific and technical developments of future reactor technologies and designs. Further, the scope for the Committee includes human and organisational research activities and technical developments that affect nuclear safety.

Foreword

The modelling of external hazards encompasses different technical aspects, depending on the type of hazard, though a general feature of hazards is that they produce off-normal conditions that can impact nuclear installations. There is also a coupling of the hazard with its associated risk analysis models. A risk analysis contains a set of scenarios, frequencies and associated consequences, developed in such a way as to inform decisions. A scenario contains an initiating event and (usually) one or more subsequent events leading to an end state that reflects the issue of concern. The objective of this benchmark study is to focus on the initiating event by facilitating an exercise on the statistical modelling for assessing hazard frequency and magnitude for external events risk assessment. This benchmark study report provides details (data and overall objectives) for the benchmarking exercise by specifying synthetic data for a hypothetical external event (e.g. precipitation, extreme temperatures and high winds). The analysis steps and modelling results of the benchmark participants are provided and summarised. Overall conclusions from these submissions are described to gain insights from the activity.

This report was approved by the Nuclear Energy Agency (NEA) Committee on the Safety of Nuclear Installations (CSNI) at the 69th meeting of the CSNI held on 2-3 June 2021 (NEA/SEN/SIN(2021)1, not publicly available).

Acknowledgements

Special thanks are extended to the following individuals for their contributions to this benchmarking exercise – Olle Räty and Marko Laine (Finnish Meteorological Institute); Herve Cordier, Anne-Hélène Dutfoy Lebrun, Jeanne Ady Demgne and Luc Guinard (Électricité de France); Vincent Rebour, Yasser Hamdi and Yann Richet (Institut de Radioprotection et de Sûreté Nucléaire, France); Minkyu Kim, Beomjin Kim and Kun-Yeon Han (Korea Atomic Energy Research Institute); Stef Carelsen (Autoriteit Nucleaire Veiligheid en Stralingsbescherming, Netherlands); and Curtis Smith, Kurt Vedros, Zhegang Ma, Jing Zhang and Andrea Mack (Idaho National Laboratory, United States).

Thanks are also due to all the participants who submitted their analysis for consideration and the writers and reviewers of this document.

Table of contents

List of figures

List of tables

List of abbreviations and acronyms

AM	Annual maxima
ANVS	Autoriteit Nucleaire Veiligheid en Stralingsbescherming (Netherlands)
CSNI	Committee on the Safety of Nuclear Installations (NEA)
EDF	Électricité de France
FMI	Finnish Meteorological Institute
GEV	Generalised extreme value
GPD	Generalised Pareto distribution
IE	Initiating event
INL	Idaho National Laboratory
IRSN	Institut de Radioprotection et de Sûreté Nucléaire (Institute for Radiological Protection and Nuclear Safety, France)
KAERI	Korea Atomic Energy Research Institute
MCMC	Markov Chain Monte Carlo
NEA	Nuclear Energy Agency
OECD	Organisation for Economic Co-operation and Development
POT	Peaks-Over-Threshold
PRA	Probabilistic risk assessment
SSE	Sum square error
TGNEV	Task Group on Natural External Events (NEA)
TNC	Truncated Newton Constrained
TSO	Technical and scientific support organisation
WGEV	Working Group on External Events (NEA)
WGRISK	Working Group on Risk Assessment (NEA)

Executive summary

The March 2011 accident at the Fukushima Daiichi Nuclear Power Plant triggered discussions about natural external events that are low-frequency but high-consequence. To address these issues and determine which events would benefit from international co-operative work, the NEA Committee on the Safety of Nuclear Installations (CSNI) established the Task Group on Natural External Events (TGNEV) at its June 2013 meeting. In June 2014, the CSNI decided to re-organise TGNEV into a Working Group on External Events (WGEV) to improve the understanding and treatment of external hazards and support the continued safety performance of nuclear installations as well as improve the effectiveness of regulatory practices in NEA member countries. The WGEV is composed of a forum of experts for the exchange of information and experience on external events in NEA member countries, thereby promoting co-operation and maintenance of an effective and efficient network of experts.

At its 61st meeting, the CSNI approved the recommended task on the benchmark on external events hazard frequency and magnitude statistical modelling, to be pursued by the WGEV. Modelling of these external events is a common practice in hazard and risk assessments in many countries. Having a valid statistical approach to model these hazards is important. However, current practice indicates a wide variety of approaches being used and a lack of appreciation of the uncertainties inherent to these types of statistical models. The objective of this activity is to provide a benchmark suitable to explore the application of typical approaches to external hazard representation through a data-informed process. This report captured two types of benchmarks, one with data and model provided and one with just data provided (a "blind test").

In the report is a summary of statistical modelling approaches from the organisations Électricité de France (EDF), the Finnish Meteorological Institute (FMI), the Idaho National Laboratory (INL), the Institut de Radioprotection et de Sûreté Nucléaire (IRSN), and the Korea Atomic Energy Research Institute (KAERI).

Several observations can be made related to the approaches used and results from the benchmark for external hazards.

- Different statistical approaches such as regression or probability distribution models can provide reasonable hazard frequency and magnitude estimations for time periods where data exist.
- Predictions for long return periods (e.g. much greater than the existing data time collection) can prove challenging for some types of models and data sets.
- Rather than focusing on predictions of a magnitude for a particular hazard, an alternative approach might be to evaluate the probability of exceeding a critical level in a future time interval.
- If an underlying physical phenomenon that drives an external hazard is unbounded then predictions for the hazard may be underestimated.
- Capturing uncertainties in the hazard predictions is not typically performed.

Based upon the observations made from the benchmark, several potential future WGEV activities are recommended:

- The availability of statistical open-source tools and frameworks offers the potential for standardised approaches for representing the frequency/magnitude hazards and should be investigated for hazard and risk applications.
- Having knowledge of the underlying hazard phenomena could improve the predictions made from models. For example, availability of a maximum upper bound (e.g. a physical limit) could help to improve long-term predictions. Future benchmarks should investigate the application of physical phenomena with hazard modelling.
- The use of "paleo-data" for a period longer than that recorded in actual data sets could have resulted in better predictions from the statistical models. The availability of this type of data, though, is not well understood for some types of hazards, and should be investigated for external hazards of interest to the WGEV.
- Uncertainties inherent in hazard model predictions should be better understood and quantified as a part of validation.

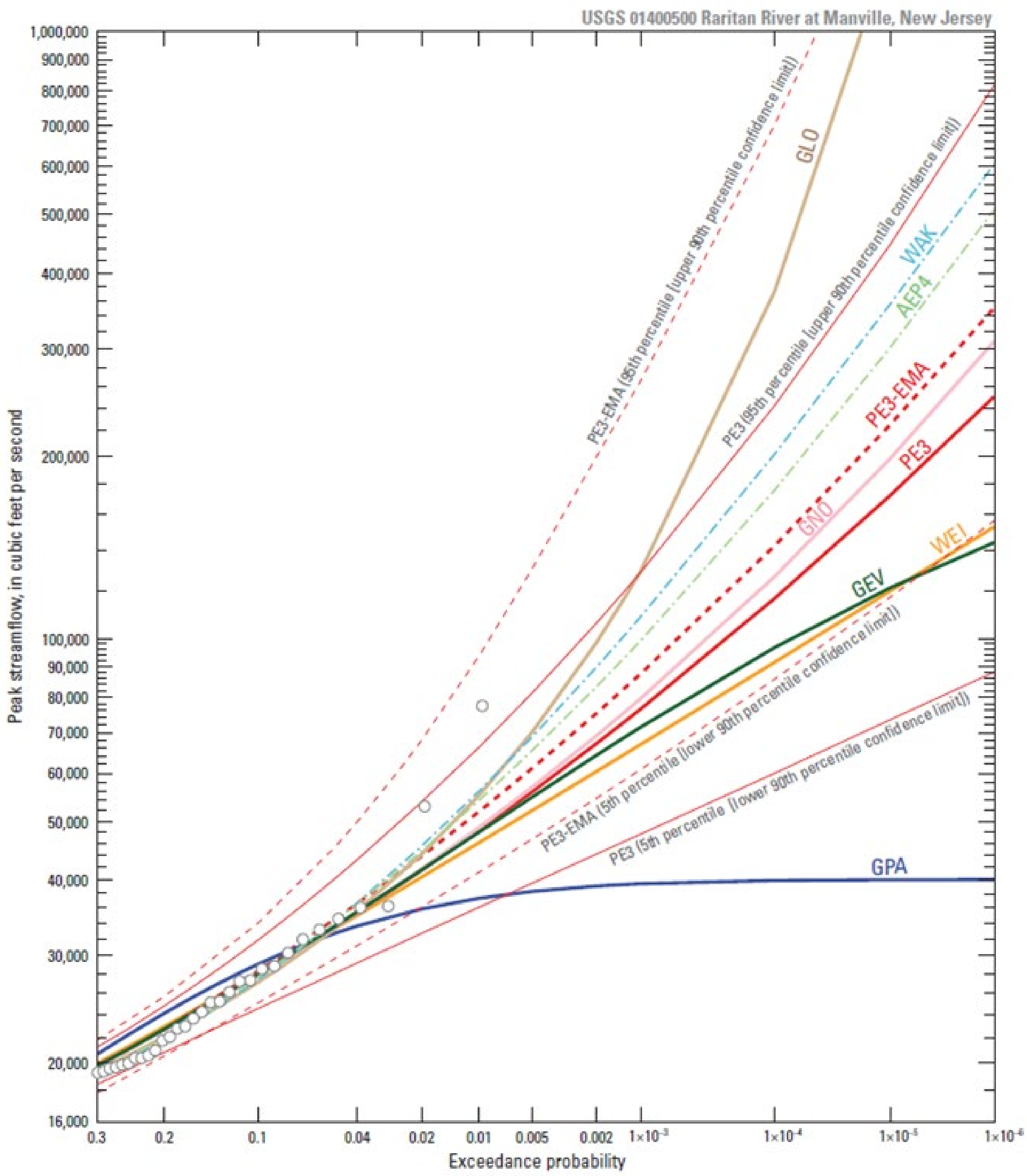
1. Introduction

1.1 Background

As an input into risk analysis modelling and simulation, a hazard initiating event (IE) is typically considered as the starting point for risk models. Since the IE both contributes to the risk quantification results and provides the boundary conditions for the rest of the hazard scenarios, effective modelling of the frequency and magnitude for different external events using data-driven methods can be applied. However, current practice has shown a variety of technical approaches, models, and limitations in validation of these approaches (for example, see Figure 1.1). Consequently, this benchmark study is intended to demonstrate and capture commendable practices in formulating and assessing the quantification of external event IEs when using statistical models.

This benchmark was open to a variety of technical communities including academia, government agencies, industry, research institutes, and technical and scientific support organisations (TSOs). Chapter 2 of this report provides the case studies under consideration of this benchmark. An overview of the results of the participant submissions is presented in Chapter 3. Overall insights and conclusions are provided in Chapter 4. Detailed submissions are listed in the Annexes.

Figure 1.1. Example of the variety in potential modelling choices for magnitude-probability representation of a streamflow initiating event



Source: OECD (2014), OECD Economic Outlook: Statistics and Projections (database), http://dx.doi.org/10.1787/data-00688-en.

1.2 Objective

The objective of this benchmark study was to facilitate an exercise on statistical modelling in order to better understand the quantitative technical analysis steps and processes used for assessing hazard frequency and magnitude in external events risk assessments. This benchmark study report provides details (data and overall objectives) for the benchmarking exercise (in Chapter 2) by specifying synthetic data for a hypothetical external event (e.g. precipitation, extreme temperatures, high winds).

1.3 Overview of the exercises

In terms of scope, the benchmark exercise on the analysis and assessment of the hazard frequency/magnitude (called "the benchmark" in this report) for external events risk assessment covers the description of probabilistic hazard modelling, its associated uncertainty characterisation, and the process used for analysis and assessment. The information that was provided to, and asked of, the benchmark participants encompassed:

- Hypothetical observational data representing an external hazard that have been created from synthetic models (this type of model is used to create synthetic data generated by a computer). Two cases are described: (1) a fully revealed "open" case where both the synthetic data and the synthetic model producing the data are provided, and (2) a "blind-test case" where only the synthetic data are provided.
- Descriptions by the participants of the assumptions made to create the hazard frequency/magnitude model(s), the qualitative and quantitative results of the model(s), the process used to assess the adequacy of the model(s) and the results of the model adequacy assessment. For this benchmark, phenomenologically-based evaluation and modelling will not be considered since only "observational-type" data (derived from a synthetic model) are provided, resulting in statistical-types of models to be considered. However, more complicated modelling situations incorporating phenomena physics and/or spatial considerations may be proposed for a future benchmark.

The specific choice of model(s) to be considered by benchmark participants is not limited a priori but will need to be able to incorporate the data and should, ideally, be able to provide a consideration of prediction uncertainty.

1.4 Glossary

This section contains a list of definitions for key terms used in this benchmark.

Aleatory uncertainty – Pertaining to stochastic (non-deterministic) events, the outcome of which is described by a probability. From the Latin alea (game of chance, die).

Adequacy assessment – The process of judging, both qualitatively and quantitatively, the predictive ability of a model used for decision making.

Data – Distinct observed (e.g. measured) values of a physical process. Generally, data may be subject to uncertainties, such as imprecision in measurement.

External event – An event originating outside a nuclear power plant that directly or indirectly causes an initiating event.

Hazard – Anything that has the potential to cause an undesired event or condition that leads to damage.

Hazard curve – A model that relates the occurrence frequency of a hazard to the magnitude (e.g. intensity of an earthquake, precipitation rate, flood water level, temperature level) of the hazard.

Model – A mathematical construct that converts data and information into knowledge. Two types of models are used for risk analysis purposes, probabilistic (or aleatory) and deterministic.

Statistical model – A model that represents complex phenomena stochastically. Examples of common statistical approaches used in risk assessment include extreme value, exponential, Weibull and Poisson models.

Synthetic data – Data that are generated from computational models instead of from actual observation.

Synthetic model – A computational model that is designed to produce data that mimic actual observed data.

2. Benchmark problem description

2.1 Data generation via synthetic models

Rather than real data gathered from the external hazard community, synthetic examples of a hypothetical physical process (e.g. local precipitation, high winds) are used for the benchmark. The reason for this is that because synthetic cases are used, the underlying data-producing mechanisms are known (exactly), as opposed to real cases, where they are not known. Thus, it is possible to assess how well data-driven models perform from a predictive standpoint. Furthermore, since synthetic models are created and run on a computer to simulate data, it is possible to control the data-generating process, including accounting for elements such as the uncertainty present in the synthetic data.

In general, a synthetic model is an equation that has the functional form of:

Model output = f(inputs) * Uncertainty

For example, a simple linear model with no uncertainty used to represent a hypothetical hazard frequency/magnitude relationship could be expressed as:

Magnitude = $a * t$

where the magnitude could be represented in terms of some observable quantity (e.g. height about a river flood stage, quantity of rain in an hour, velocity of wind, depth of snow in a day), a is a changeable parameter used to control the output of the synthetic model, and t is a time interval – or return frequency – that "sees" the observable quantity described by the magnitude term. In this case, longer return frequencies would produce larger synthetic events in a linear fashion.

2.2 Case 1 – Known model producing the synthetic data

The synthetic model used for the first exercise (Case 1) is:

$M = 0.5 + 0.5 * \log_{10}(a * t)$.

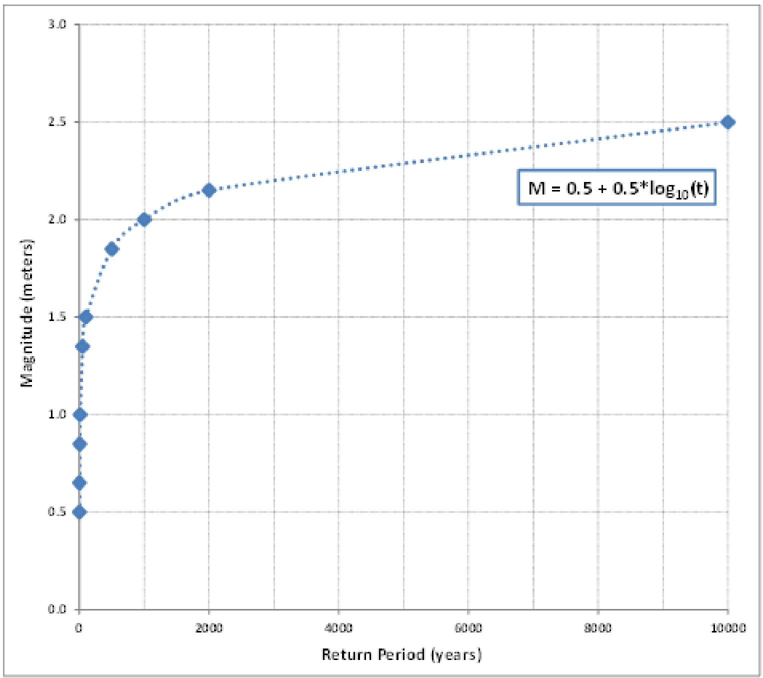
This synthetic example is constructed so that different values of "return time intervals" t produce a hypothetical (but known since it comes from the synthetic model) magnitude M for an annual maxima event. It is possible to use these types of models to produce "synthetic data" where different event outcomes are predicted as a function of time (e.g. producing a flooding hazard curve). As used in this benchmark, the synthetic model serves as surrogate for a complex phenomenological process.

One characteristic in Case 1 is that the data generation process produces "extreme events" such that the larger the event, the less frequently the event is seen – in other words the return interval becomes large as the event magnitude increases. Consequently, the units (italicised) that are present in the Case 1 synthetic model are:

$M\ \textit{metres} = 0.5\ \textit{metres} + 0.5 * \log_{10}(a\ \textit{years}^{-1} * t\ \textit{years})\ \textit{metres}$.

For Case 1, the "a variable" is set to 1. Thus, it is now possible to plot the magnitude of the event that is "produced" by Case 1 for select times up to a return interval of 10 000 years (see Figure 2.1).

Figure 2.1. Case 1 plot up to a return period of 10 000 years



However, what is important for this benchmark is that hypothetical observational data (from a synthetic model) is provided for the hazard frequency/magnitude modelling. For Case 1, this data is shown in Table 2.1.

Table 2.1. Synthetic data for Case 1

Return period (years)	1	2	5	10	50	100	500	1 000	2 000	10 000
Magnitude (metres)	0.50	0.65	0.85	1.0	1.4	1.5	1.9	2.0	2.2	2.5

Participants were asked to use the data for Case 1 and provide a model that best described the frequency/magnitude relationship and the associated analysis and insights. The results of this analysis should include those areas identified in Chapter 3 of this benchmark, including:

- Qualitative aspects and insights
  - assumptions made to create the hazard frequency/magnitude model;
  - the process used to assess the adequacy of the model.
- Quantitative aspects and insights including
  - the type of model describing the hazard frequency and magnitude statistical results;
  - uncertainties of the model - assessing uncertainty is important for both validation and prediction (NRC, 2010);
  - results of the model adequacy assessment or validation.

2.3 Case 2 – Unknown model producing the synthetic data

For Case 2, the synthetic model used was not provided to the participants. However, three parts were provided for this example: part (a), which provides the synthetic data (ten data points) with no uncertainty on the data points provided; part (b), which provides additional synthetic data (26 data points) with no uncertainty on the data points provided; and part (c), which has uncertainty estimates on some of the data. Any of the parts could have been evaluated by participants.

For part (a), the synthetic data output from the unknown model are shown in Table 2.2.

Table 2.2. Synthetic data for Case 2(a)

Return period (years)	1	2	5	10	50	100	500	1 000	3 000	10 000
Magnitude (metres)	0.53	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0

For part (b), the synthetic data output from the unknown model are shown in Table 2.3.

Table 2.3. Synthetic data for Case 2(b)

Return period (years)	1	2	5	10	15	20	25	30	40
Magnitude (metres)	0.53	0.53	0.54	0.55	0.56	0.56	0.57	0.57	0.58
Return period (years)	**50**	**60**	**70**	**80**	**90**	**100**	**125**	**150**	**175**
Magnitude (metres)	0.59	0.60	0.60	0.61	0.62	0.62	0.63	0.65	0.66
Return period (years)	**200**	**300**	**400**	**500**	**750**	**1 000**	**3 000**	**1 0000**	
Magnitude (metres)	0.67	0.71	0.75	0.79	0.87	0.95	1.57	3.97	

For part (c), the synthetic data for long time intervals are presumed to not be known exactly. For these times (500 years and longer), an estimate of the uncertainty on the magnitude has been provided and is shown in Table 2.4.

Table 2.4. Synthetic data for Case 2(c)

Return period (years)		1	2	5	10	50	100	500	1 000	3 000	10 000
Magnitude (metres)	Mean	0.53	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
	SDev.*	-**	-	-	-		-	0.04	0.06	0.15	0.46
	5th***	-	-	-	-		-	0.72	0.85	1.3	3.2
	95th	-	-	-	-		-	0.85	1.1	1.8	4.7

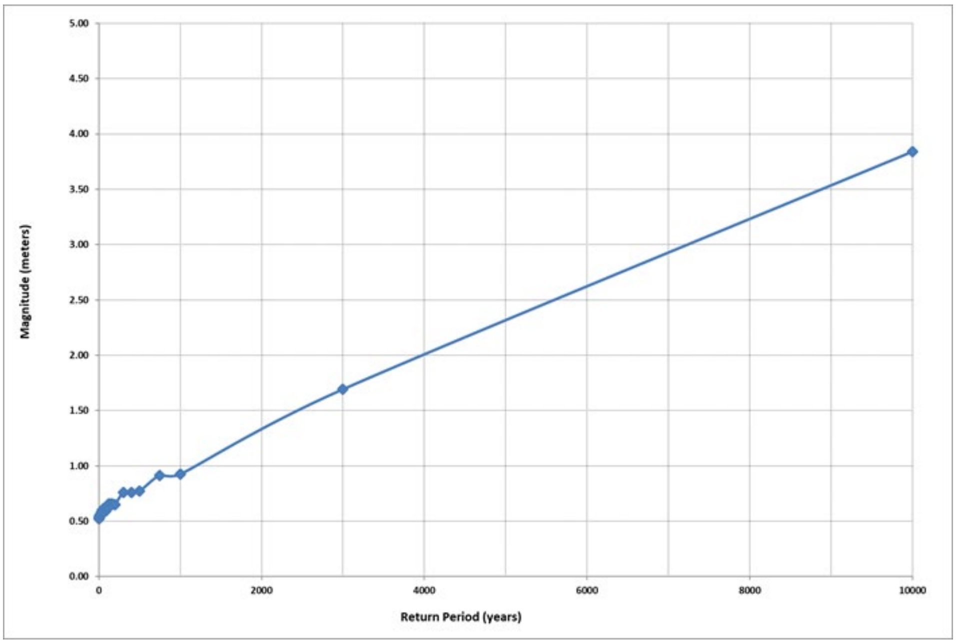
Note:
* SDev. is the standard deviation.
** A "-" indicates information that is not available.
*** The 5th and 95th indicates a 5th percentile and a 95th percentile value, respectively. In other words, there is a 0.05 probability that the magnitude is less than or equal to 0.72 metres for an event with a return period of 500 years.

The synthetic model used for Case 2 was:

$$M\ \textit{metres} = 0.5\ \textit{metres} + 0.01 * b * \exp(1\ \textit{years}^{-1} * t^{0.19}\ \textit{years})\ \textit{metres}.$$

where b is normal distribution with a mean of 1.1 and a standard deviation of 0.15. The use of a normal distribution in this synthetic model provides a degree of "noise" or stochastic variation in the data points that are provided via the model. Using this model, it is possible to plot the magnitude of the event that is "produced" by Case 2 for select times up to a return interval of 10 000 years (see Figure 2.2).

Figure 2.2. Case 2 plot up to a return period of 10 000 years



Lastly, the magnitude of the model prediction was asked of the participants for 500, 5 000, 50 000, and 500 000 years.

Now that the two models are known completely, it is possible to provide the exact results for the different return periods. These results are shown in Table 2.5. Note that since the Case 2 model uses an exponential, for long return periods, the magnitude can become large – this model may not represent physical processes effectively but can provide a challenge for modelling.

Table 2.5. Results for the two cases

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1	1.9	2.4	2.9	3.4
Magnitude (metres) Case 2	0.78	2.2	28	2 000

3. Benchmark outcomes

3.1 Benchmark completion

To complete the benchmark exercise, the following four areas should be addressed:

1. Assumptions
2. Model detail results
3. Model adequacy assessment approach
4. Results of the assessment

While much of the effort spent in responding to this benchmark was on technical analysis, it is important to understand key assumptions behind the modelling approach and the ultimate use of the model for hazard representation. As part of an integrated risk analysis, the hazard characterisation is a key element of effective modelling. Thus, any supporting analyses need to be developed in context with consideration and understanding of the scope, limitations, boundary conditions, complexity and contexts that provide the foundation for the resulting statistical model. These types of assumptions need to be clearly documented to understand how results will be used in risk-informed decision making.

Submissions were provided for consideration by the following: Électricité de France (EDF), the Finnish Meteorological Institute (FMI), the Idaho National Laboratory (INL), Institut de Radioprotection et de Sûreté Nucléaire (IRSN), and the Korea Atomic Energy Research Institute (KAERI).

3.2 Submission by EDF

In its submission, EDF models the frequency-magnitude relation with a generalised extreme value (GEV) distribution. It notes that, under some general hypotheses, the annual extreme value of a process, once normalised, tends to a GEV distribution and the return levels are specific quantiles of the annual extreme value distribution. The GEV limit model makes it possible to estimate large return levels.

In Cases 1 and 2, the fitting is based on the minimisation of a criteria, defined as the squared error, the weighted square error, the maximum error or the weighted maximum error between the data and the model. For Case 1, EDF also used three-points interpolation by solving a system of equations defined by the quantiles of GEV distribution.

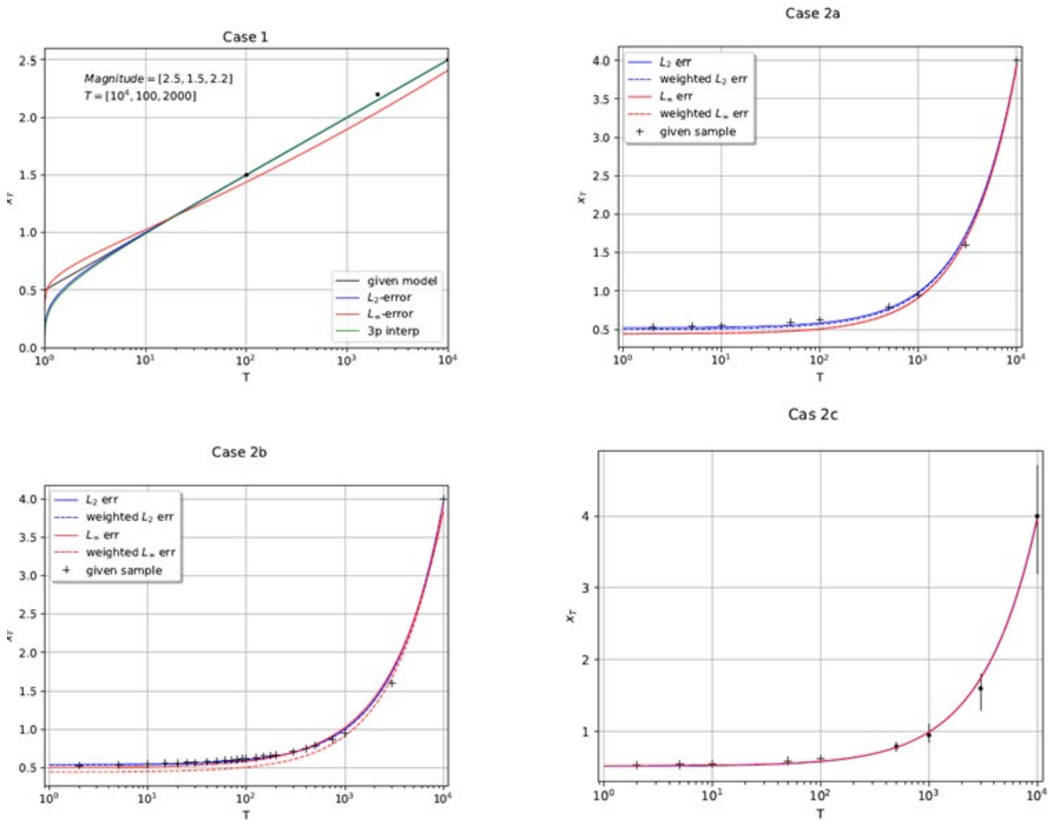
The EDF analysts noted it is possible to improve the precision of the GEV model on large return periods by penalising the errors to give more importance in properly predicting the magnitude associated to large return periods.

In Case 2c, the data are uncertain. EDF uses a GEV model that minimises random features of the error focusing on either the mean or 95th quantile. Examples of analysis results obtained with the L2 criteria are shown in Table 3.1. Instead, the models were used to compare against the data tables that were shown in Chapter 2. Plots of these results are shown in Figure 3.1. The details of their analysis, including tables of numerical results, are shown in Annex A.

Table 3.1. Results for EDF submission (example of L2 criteria results)

Return period (years)	500	5 000	50 000	500 000
Case 1	1.85	2.35	2.85	3.35
Case 2a	0.78	2.40	14.6	105
Case 2b	0.79	2.41	14.3	102
Case 2c	0.79	2.41	14.25	101

Figure 3.1. EDF submission results plotted for the four cases



3.3 Submission by the Finnish Meteorological Institute (FMI)

The FMI submitted a write-up and included the analysis files (in the R tool). All the necessary steps to repeat the exercise (excluding setting up some parts of the R environment) were provided in this notebook. The submission provides the theoretical and technical background for the modelling approach and the assumptions. Then, the results for fitting the selected statistical models are presented for each test case together with an analysis of potential uncertainties.

The model used by FMI is the GEV. They use a Bayesian approach (using a couple of numerical approaches) to the estimation of the GEV parameters. One of the main strengths in the Bayesian approach is that it provides a natural way to estimate both parameter and observational uncertainties. The numerical results of the FMI analysis are shown in 3.1 for Case 1 and 3.2 for Case 2.

Table 3.2. Case 1 results for FMI submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1 Exact	1.9	2.4	2.9	3.4
FMI mean from the Approximate Bayesian Computation approach magnitude (metres)	1.85	2.35	2.84	3.34
FMI mean from traditional Markov Chain Monte Carlo approach (metres)	1.84	2.35	2.88	3.43

Table 3.3. Case 2a results for FMI submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2000
FMI mean from the Approximate Bayesian Computation approach magnitude (metres)	0.76	2.33	16.62	148.36
FMI mean from traditional Markov Chain Monte Carlo approach (metres)	0.76	2.31	17.25	162.27

Table 3.4. Case 2b results for FMI submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2000
FMI mean from the Approximate Bayesian Computation approach magnitude (metres)	0.77	2.32	16.22	142.11
FMI mean from traditional Markov Chain Monte Carlo approach (metres)	0.76	2.30	16.90	156.36

Table 3.5. Case 2c results for FMI submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2000
FMI mean from the Approximate Bayesian Computation approach magnitude (metres)	0.78	2.28	14.16	110.94
FMI mean from traditional Markov Chain Monte Carlo approach (metres)	0.77	2.36	16.95	151.60

3.4 Submission by Idaho National Laboratory

Two independent groups at the INL provided results. The first group used a Bayesian approach for a GEV model, using the OpenBUGS analysis tool. The second group used a regression approach. Specifically, the second group used linear regression with transformed form model to fit the magnitude vs. return period relationship model for Case 1 and non-linear regression for Case 2.

INL Group 1

Group 1 used a Bayesian approach with the GEV model to quantify both Case 1 and 2. The tool used was OpenBUGS. An example of the script used for Case 1 is shown in Table 3.6. The predicted results for Case 1 are shown in Table 3.7. The results for Case 2 are shown in Table 3.8 (Case 2a), Table 3.9 (Case 2b), and Table 3.10 (Case 2c).

Table 3.6. OpenBUGS script for Case 1 from INL Group 1 submission

```
model
{ for(i in 1:N) {
z.p[i] ~ dnorm(mean[i],prec)
y.p[i] <- -log(1 - p[i])
mean[i]<- mu - sigma/xi*(1 -pow(y.p[i],-xi))
}
mu ~ dnorm(0,0.0001)
prec<-pow(sd,-2)
sd~dunif(0,10)
xi ~ dunif(-1,1)
sigma ~ dunif(0,10)
}
data
list(p=c(0.632, 0.393, 0.181, 0.0952, 0.0198, 0.00995, 0.002, 0.001, 0.0005, 0.0002, 0.0001, 0.00002,
0.000002),
z.p=c(0.50, 0.65, 0.85, 1.0, 1.4, 1.5, 1.9, 2.0, 2.2, NA, 2.5, NA, NA), N=13)
list(mu=1.0, sigma=1.0, xi=1.0)
```

Table 3.7. Case 1 results for INL Group 1 submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1 Exact	1.9	2.4	2.9	3.4
INL mean (metres)	1.88	2.37	2.84	3.31

Table 3.8. Case 2a results for INL Group 1 submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
INL mean (metres)	0.76	2.34	16.02	136.20

Table 3.9. Case 2b results for INL Group 1 submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
INL mean (metres)	0.76	2.32	16.26	142.80

Table 3.10. Case 2c results for INL Group 1 submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
INL mean (metres)	0.76	2.34	16.02	136.20

INL Group 2

The hazard frequency/magnitude model here for Case 1 fit the linear regression model with log transformation of the return period to describe the relationship between magnitude and return period since the synthetic model provided is linear. The difference is the residual term in regression equation, which is the vector values of the differences between observed values and predicted values:

$$\widehat{Magnitude_i} = \widehat{\beta_0} + \widehat{\beta_1} log(return\,period_1) + \epsilon_i$$

The estimated equation is:

$$\widehat{Magnitude_i} = 0.503364 + 0.506251 * log(return\,period_1)$$

The estimated values above are the least square estimates of the intercept and slope. They have standard error for the intercept and slope are 0.07006 and 0.03083, respectively. The predicted results are shown in Table 3.11 for Case 1, Table 3.12 for Case 2a, Table 3.12 for Case 2b, and Table 3.14 for Case 2c.

Table 3.11. Case 1 results for INL Group 2 submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1 Exact	1.9	2.4	2.9	3.4
INL mean (metres)	1.87	2.38	2.88	3.39

Table 3.12. Case 2a results for INL Group 2 submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
INL mean (metres)	0.75	2.32	14.83	37.18

Table 3.13. Case 2b results for INL Group 2 submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
INL mean (metres)	0.75	2.28	17.57	170.46

Table 3.14. Case 2c results for INL Group 2 submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
INL mean (metres)	0.71	2.01	13.75	37.09

Group 2 noted that principal component regression has a few technical limitations and uncertainties. First, the uncertainty of mean for short time interval was not provided, even though the unknown standard deviation of each mean magnitude is small. To estimate the standard deviation, the 5th and 95th percentile of the mean magnitude, an imputation must be utilised which produced some uncertainty on those imputed values. Compared with the results estimation obtained from Case 2a, we noticed that the confidence and prediction intervals for Case 2c are all wider than those for Case 2a. Model fitting also does not perform well when there is a long time return period so that the predicted values do not fall into the 95% confidence interval. Lastly, the principal component regression model could not meet the expected condition that the mean magnitude keeps increasing as the return period increases.

3.5 Submission by IRSN

The IRSN submitted its analyses for Case 1, Case 2(a) and Case 2(b) in a write-up that included the analysis files (using the R language).

For Case 1, a set of 100 return periods was first randomly sampled. Then, associated magnitudes were calculated using the synthetic model. Finally, extreme frequency estimations were performed.

The analyst implemented an annual maxima (AM) frequency model in which the distribution of the extreme events converges to a GEV one. The IRSN noticed that the GEV distribution can have the form of the synthetic model only if the shape parameter of the distribution is equal to -1 and the scale parameter is negative. With such conditions, the theoretical upper tail can only be finite and bounded and the GEV distribution cannot be used with a negative scale parameter: thus it was not used to describe the frequency/magnitude relationship for Case 1.

The analyst implemented a Peaks-Over-Threshold (POT) frequency model in which the distribution of the exceedances over the threshold converge to an exponential one (Generalised Pareto Distribution, or GPD). The choice of the threshold value is based on using the synthetic model. The Renext R library (developed by the IRSN and Alpestat) was used for the frequency estimations. The relative difference in estimated magnitudes does not exceed 5%, and all the plotting positions are inside the one sigma confidence interval even though the latter is very narrow (see Figure 3.2).

Figure 3.2. Fitting of the SM-Case1 data sets (w=100 years) with GPD distribution

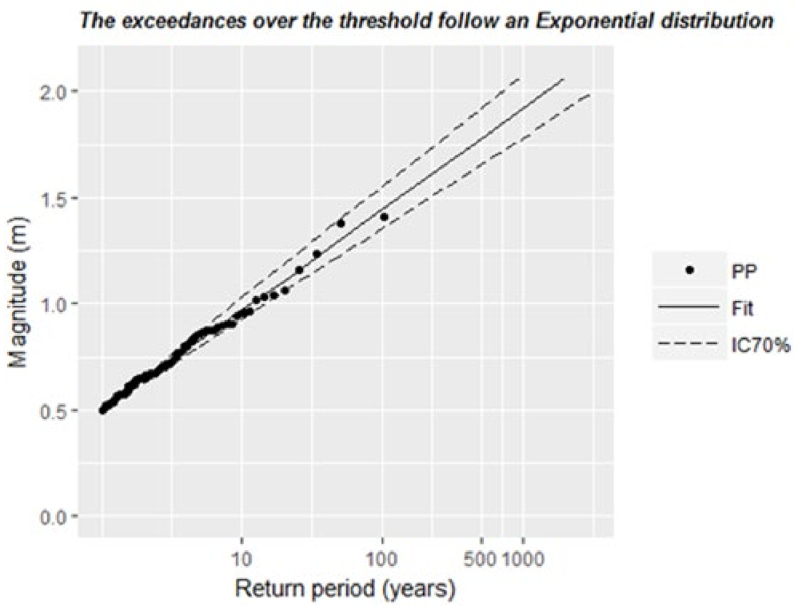


Table 3.15. Case 1 results for IRSN submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1 Exact	1.9	2.4	2.9	3.4
GPD/Exp IRSN mean (metres)	1.78	2.25	2.72	3.19

For both Cases 2a and 2b, non-linear least-squares estimates of the GEV parameters were performed. As shown in Figure 3.3, the fitting with confidence intervals is quite good with heavy tails (very high shape parameter $\xi \approx 0.96$) up to some thousands of years. For longer return periods the fitting is off the curve, perhaps because the proposed cases ignore any physical limits.

Figure 3.3. To the left: Fitting Case2-a synthetic data with a GEV distribution; to the right: Fitting Case2a synthetic data with a GEV distribution

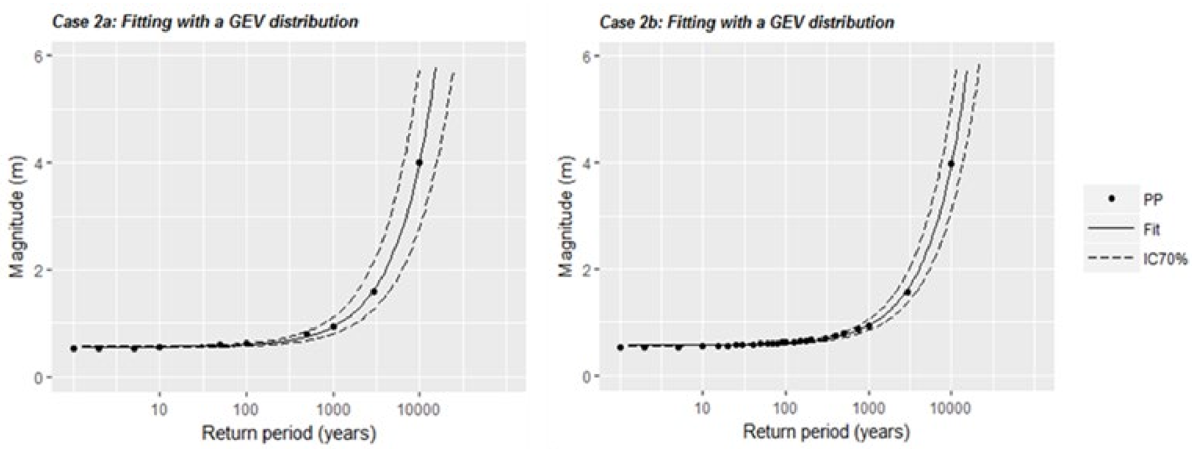


Table 3.16. Case 2a results for IRSN submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
GEV IRSN mean (metres)	0.75	2.32	16.79	150.34

Table 3.17. Case 2b results for IRSN submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
GEV IRSN mean (metres)	0.76	2.31	16.50	146.53

3.6 Submission by KAERI

As a result of fitting the relationship between magnitude and return period in Case 1 and Case 2, the relationship was estimated in log and linear regression. The data fitting is high in R^2 but with a square error. The reason is sensitive to the coefficients of the synthetic model. Therefore, further parameter analysis is required. The regression equation of this study is as shown in the following equations (1) to (2).

For Case 1: $M = A * ln\,(x) + B\ (Parameters\ A\ and\ B)$ (1)

For Case 2: $M = A * (x) + B\ (Parameters\ A\ and\ B)$ (2)

To minimise the square error of parameters A and B, an optimisation method is applied through the square error solver function. As a result, the estimated parameters A and B for Case 1 and Case 2 are shown in the following equations (3) and (4).

For Case 1: $M = 0.219861834 * ln\ (x) + 0.503363549\ (Minimize\ square\ error)$ (3)

For Case 2: $M = 0.000344252 * (x) + 0.565051348\ (Minimize\ square\ error)$ (4)

The predicted results for Case1 and Case2 are shown in Table 3.17 and 3.18. In addition, the re-fitted graph was compared with the existing data and displayed, as shown in Figure 3.4 to Figure 3.5.

Table 3.18. Case 1 results for KAERI submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1 Exact	1.9	2.4	2.9	3.4
Log KAERI mean (metres)	1.870	2.376	2.883	3.389
Optimisation KAERI mean (metres)	1.870	2.376	2.882	3.388

Table 3.19. Case 2 results for KAERI submission

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
Linear KAERI mean (metres)	0.715	2.065	15.565	150.565
Optimisation KAERI mean (metres)	0.737	2.286	17.778	172.691

Figure 3.4. KAERI submission results plotted for Case 1

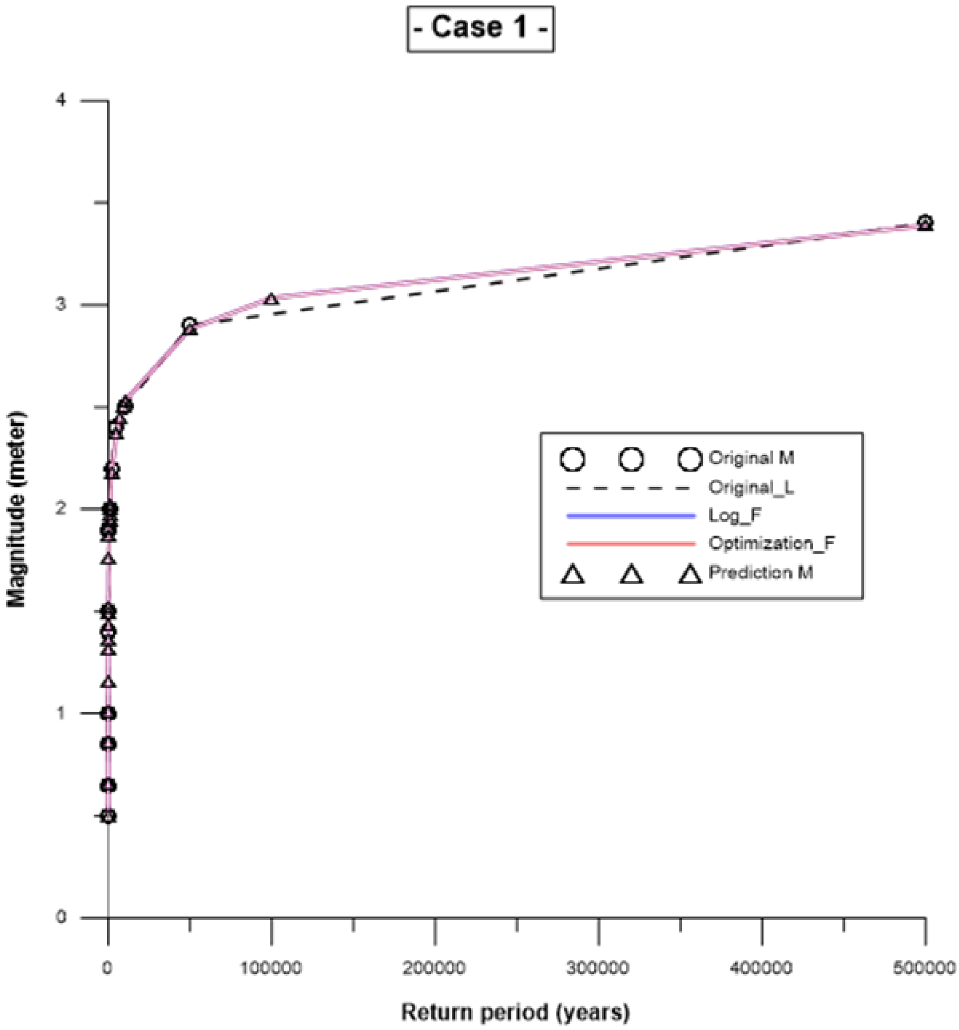
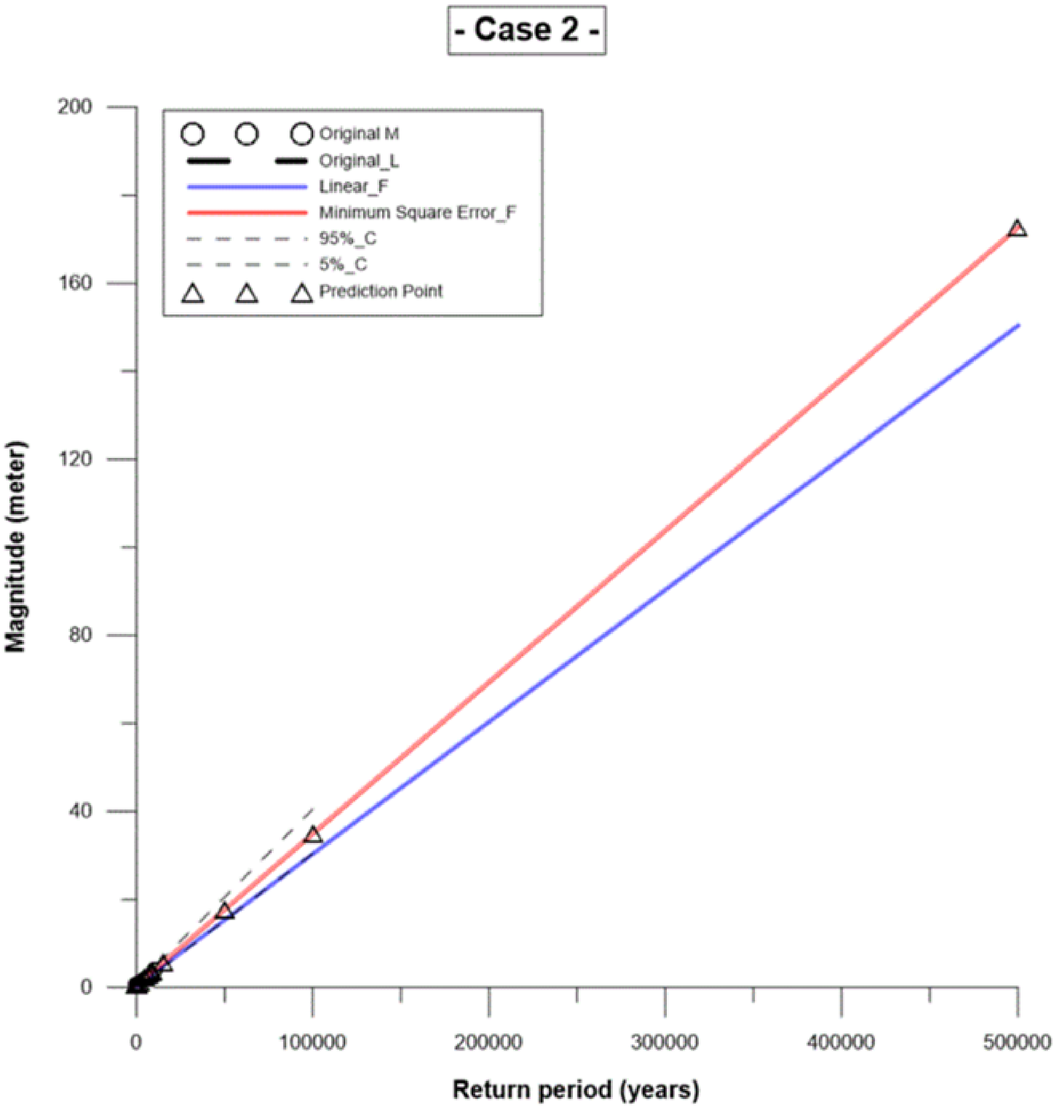


Figure 3.5. KAERI submission results plotted for Case 2



4. Observations

This benchmark provided an exercise focused on the quantitative technical analysis steps and processes used to assess hazard frequency and magnitude for external events risk assessments. The research provided details (data and overall objectives) for two unique benchmarking exercises specific to external events hazard frequency and magnitude modelling by providing synthetic data for a hypothetical external event (precipitation, extreme temperatures, high winds, etc.) to be determined by the participants. The information and results provided by benchmark participants were collected and summarised to gain insights on best practices from the activity.

4.1 Case 1 observations

Figure 4.1 shows the overall results of all participants for Case 1. As can be seen in the figure, the results tended to track well with the exact results and were consistent across all groups. The numerical values used for Figure 4.1 are listed in Table 4.1.

Figure 4.1. Comparison of all submission results for Case 1

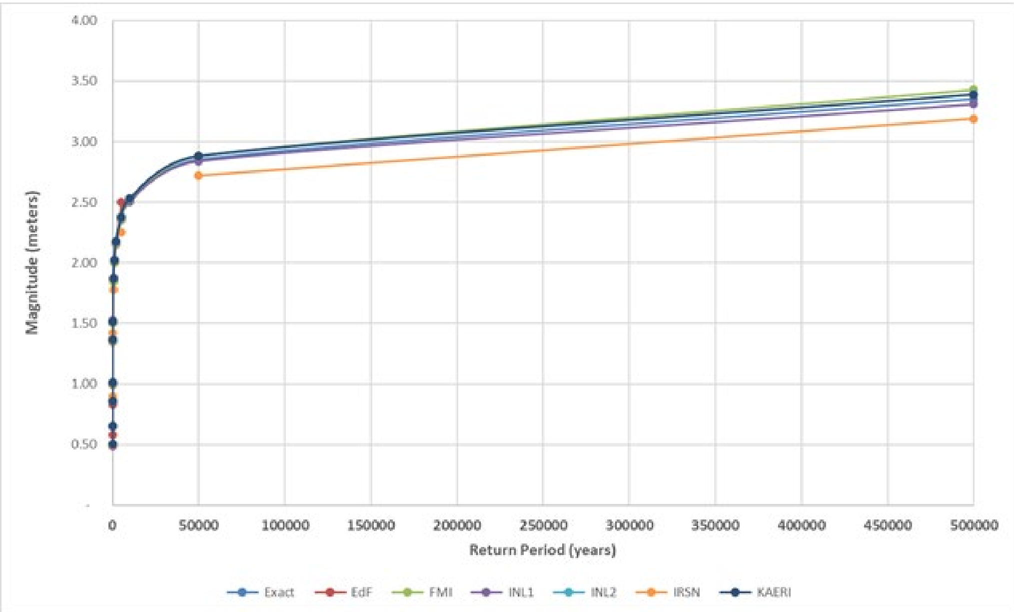


Table 4.1. Case 1 results for all group submissions

Result	1	2	5	10	50	100	500	1 000	2 000	5 000	10 000	50 000	500 000
Exact	0.50	0.65	0.85	1.00	1.35	1.50	1.85	2.00	2.15	2.35	2.50	2.85	3.35
EDF		0.58	0.83	0.99	1.35	1.50	1.85	2.00	2.15	2.35	2.50	2.85	3.35
FMI							1.84			2.35		2.88	3.43
INL1	0.49	0.65	0.86	1.02	1.37	1.53	1.88	2.03	2.17	2.37	2.51	2.84	3.31
INL2	0.50	0.66	0.86	1.01	1.36	1.52	1.87	2.02	2.17	2.38	2.53	2.88	3.39
IRSN	0.50			0.90		1.42	1.78			2.25		2.72	3.19
KAERI	0.50	0.66	0.86	1.01	1.36	1.52	1.87	2.02	2.18	2.38	2.53	2.88	3.39

Note: EDF = Électricité de France (example of results obtained with the L2 criteria), FMI = Finnish Meteorological Institute, INL1 = Idaho National Laboratory Group 1, INL2 = Idaho National Laboratory Group 2, IRSN = Institut de Radioprotection et de Sûreté Nucléaire, KAERI = Korea Atomic Energy Research Institute.
The approach and models used by the different groups ranged from curve fitting logarithm models to performing Markov Chain Monte Carlo (MCMC) calculations using the GEV model. For Case 1, five of the groups used a regression-based approach while two used a MCMC calculation.

4.2 Case 2 observations

The study for Case 2 was much more complicated since the underlying model was not provided. A summary of the numerical predictions that were provided by the submitters is shown in Table 4.2. As can be seen in the table, the predictions for long periods of time (> 10 000 years) started to become problematic due to the underlying synthetic model (based upon an exponential function that was proposed as a strong challenge).

Table 4.2. Case 2 results for all group submissions

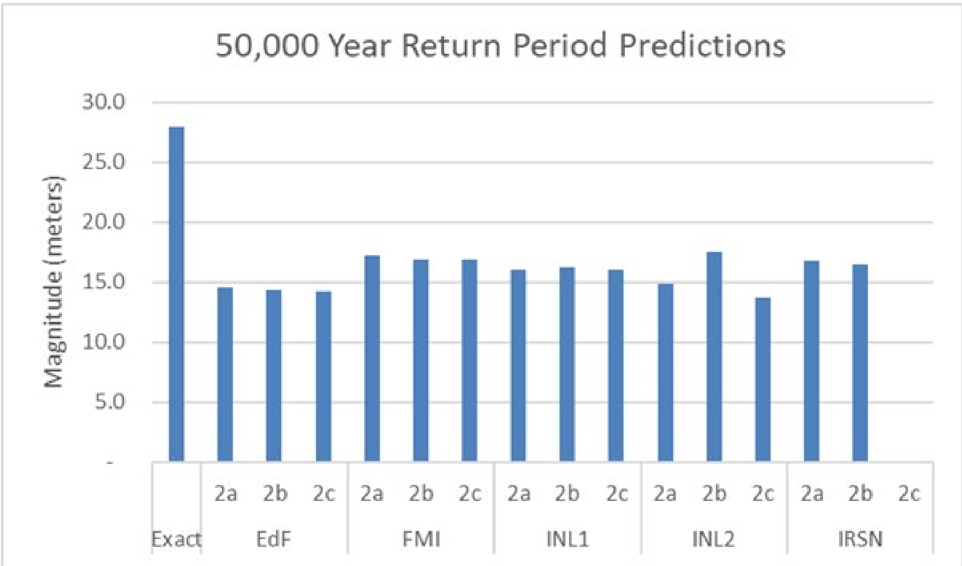
		Return period (years)			
Result	**Case**	**500**	**5 000**	**50 000**	**500 000**
Exact		0.78	2.20	28.0	2 000
EDF	**2a**	0.78	2.40	14.6	105
	2b	0.79	2.41	14.3	102
	2c	0.79	2.41	14.25	101
FMI	**2a**	0.76	2.31	17.3	162
	2b	0.76	2.30	16.9	156
	2c	0.77	2.36	17.0	152
INL1	**2a**	0.76	2.34	16.0	136
	2b	0.76	2.32	16.3	143
	2c	0.76	2.34	16.0	136
INL2	**2a**	0.75	2.32	14.8	37
	2b	0.75	2.28	17.6	170
	2c	0.71	2.01	13.8	37
IRSN	**2a**	0.75	2.32	16.8	150
	2b	0.76	2.31	16.5	147
	2c				

Note: EDF example of results obtained with the L2 criteria

Focusing on just the 50 000-year magnitude predictions (see Figure 4.2), it is possible to see that the predictions were consistent, but too low by about 40%. Again, this under prediction is likely a result of the use of the exponential function for the underlying synthetic model for Case 2. While the models used by the submitters (e.g. GEV, Generalised Pareto, regression) can represent exponential types of behaviour, the data that was provided may represent too short of a period to reliably predict very long return periods (i.e. for periods of time much longer than the data set provided). In Case 2, the "longest" data point provided was for 10 000 years, or five time shorter than the 50 000-year prediction.

Case 2b represented the situation where additional data was provided (over that provided by Case 2a) – 26 data points were provided instead of the original ten. However, in both cases, the longest time period represented was 10 000 years. Looking at the 50 000-year predictions, we can see that increasing the data points (from 10 to 26) resulted in only slightly better predictions (but they were still low by approximately 40%). It appears that having additional data within the seen time did not lead to much of an improvement in the predictions. Alternatively, if we had provided "paleo-data" for a period longer than 10 000 years, this data would have likely resulted in better predictions. This aspect of the analysis might be considered for future benchmark exercises.

Figure 4.2. Comparison of all submission results for Case 2 for the 50 000 year predictions



The participants that submitted information for Case 2 used two basic approaches to the analysis, Bayesian analysis via MCMC or regression to a curve. It appears that the quantity of data and the quantification had little impact on the results shown in Figure 4.2. However, it did appear that the Bayesian approach was able to quantify the uncertainty on the predictions in a straight-forward fashion (for example, the estimate for the INL1 2c result was a mean of 16 m with a 5^{th} percentile of 14 m and a 95^{th} percentile of 17 m for the 50 000-year return period).

When we look at the 500 000-year magnitude predictions, we can see that the predictions were consistent, but very low. This under prediction is (again) likely a result of the use of the exponential function for the underlying synthetic model for Case 2. Note that the use of an exponential function was chosen as a challenging case and most likely does not represent any real physical phenomena.

However, a concern here would be if return periods much greater than what has been experienced (> 10 times the time period) were being predicted and the underlying phenomena that drive the magnitude of the external hazard is unbounded upward like an exponential function. In that case, the predictions for the hazard may be underestimated. Fortunately, most (if not all) natural external hazards would have upper bounds to the magnitude – it would not make sense to assume infinitely large floods, temperatures, rain, snow, earthquakes, etc.

The difficulty in the predictions for long return periods does indicate that knowledge of the underlying phenomena or knowledge of upper bounds (e.g. physical limits) could help to greatly improve the predictions. This aspect of the analysis might be considered for future benchmark exercises.

5. Conclusions

This benchmark study provided a mechanism to develop statistics-based approaches for representing complex physical phenomena such as extreme temperatures, droughts, high winds, floods and extreme snowfall. The benchmark provided the overall objectives of the study, the data to be investigated and the expectations for results reporting.

This report provided two independent benchmark cases to facilitate understanding of statistical modelling that may be used in quantitative technical analysis to assess hazard frequency and magnitude for external events risk assessment:

- Case 1: Based on a data set with a given associated synthetic model (i.e. known function form).
- Case 2: Based on a data set generated from another associated synthetic model that is not given (i.e. a “blind study”).

Participants were asked to provide any key assumptions, the statistical model(s) used, the overall results of the analysis, and the process used to assess the adequacy of the model(s), including uncertainties. The provided data were created using synthetic models for a hypothetical external event (e.g. precipitation, extreme temperatures and high winds). As such, the data used in this benchmark may not be representable for any real physical phenomenon and are solely intended to test different modelling approaches.

One of the findings of this benchmark study is that different statistical approaches, such as regression or probability distribution model application, can provide reasonable hazard frequency and magnitude estimation for time periods where data exist (i.e. interpolation) for both cases. However, the benchmark study did identify potential issues:

- Predictions for long return periods (much greater than the existing data time collection) can prove challenging for some types of data sets.
- Rather than focusing on predicting a magnitude for a particular hazard, an alternative approach might be to evaluate the probability of exceeding a critical level in a future time interval.
- If an underlying physical phenomenon that drives an external hazard is unbounded (e.g. like the exponential function used in Case 2), predictions for the hazard may be underestimated.
- Some participants had a process to quantify uncertainties. However, capturing the uncertainties in the predictions was not typically performed.

While issues and challenges do exist in statistical modelling of hazards, this study pointed to potential improvements in hazard modelling processes. It is recommended that the following items be considered for future WGEV activities:

- The availability of statistical open-source tools and frameworks offers the potential for standardised approaches for representing the frequency/magnitude of hazards applicable to risk applications.
- While not part of this benchmark, having knowledge of the underlying hazard phenomena could improve the predictions made from models. For example, the availability of a maximum upper bound (e.g. a physical limit) could help to greatly improve long-term predictions.
- The use of "paleo-data" for a period longer than that recorded in actual data sets could have resulted in better predictions from the statistical models. The availability of this type of data, though, is not well understood for some types of hazards.
- Uncertainties inherent in hazard model predictions should be better understood and quantified as a part of validation.

References

National Research Council (2010), *Assessing the Reliability of Complex Models*, https://www.nap.edu/read/13395/chapter/1

OECD (2014), OECD Economic Outlook: Statistics and Projections (database), http://dx.doi.org/10.1787/data-00688-en.

Annex A. Submission by EDF

Executive summary

The modelling of external hazards encompasses different technical aspects, depending on the type of hazard. All these hazards impact nuclear installations. Usually, a specific risk analysis is performed for a given hazard to inform a decision. A risk analysis contains a set of scenarios, frequencies and associated consequences. A scenario is a sequence that contains an initiating event and one or more subsequent events. The end state (a consequence) of a scenario is also of interest.

The global objective of the benchmark study launched by the OECD NEA, is to focus on an initiating event induced by an external hazard. The statistical modelling of a hazard frequency and magnitude is of particular interest in this benchmark.

The benchmark study provides synthetic data for a hypothetical external event. Hypothetical data describing the external hazard comes from synthetic models. These data have been generated with a computer code. Two cases are described:

- Case 1: a fully revealed open case where both the synthetic data and the synthetic model producing the data are provided;
- Case 2: a blind-test case where only the synthetic data are provided.

In this report, we model the frequency-magnitude relation with a GEV distribution. Indeed, under some general hypotheses, the annual extreme value of a process, once normalised, tends to a GEV distribution. Furthermore, the return levels are specific quantiles of the annual extreme value distribution. The GEV limit model makes it possible to estimate large return levels.

To fit the GEV model that best predicts the given synthetic model or the given synthetic data, we used several criteria:

- the minimum of the squared error between the data and the model;
- the maximum error over a large range of the return period.

In addition, for Case 1, we used a three-points interpolation by solving a system of equations defined by the quantiles of GEV distribution. We voluntarily present this method only in Case 1 but it can be applied on the other cases without uncertainty.

It is possible to improve the precision of the GEV model on large return periods. To this end, we penalise the errors to give more importance to properly predicting the magnitude associated with large return periods.

If the data are uncertain, we define a GEV model that minimises random features of the error: the mean or its quantile of order 95%.

For each case, we present the relative error and the optimal set of parameters of the GEV model.

Figure A1.1 draws the T-return level obtained with the optimal GEV distribution in each case.

Table A1.1 to Table A1.4 detail the reference magnitudes and the approximated ones, obtained using several error criteria.

Figure A1.1. Frequency - Magnitude obtained with the optimal GEV distribution (in blue, red and green) that best predicts the known model or the synthetic data (in black) with respect to several error criteria

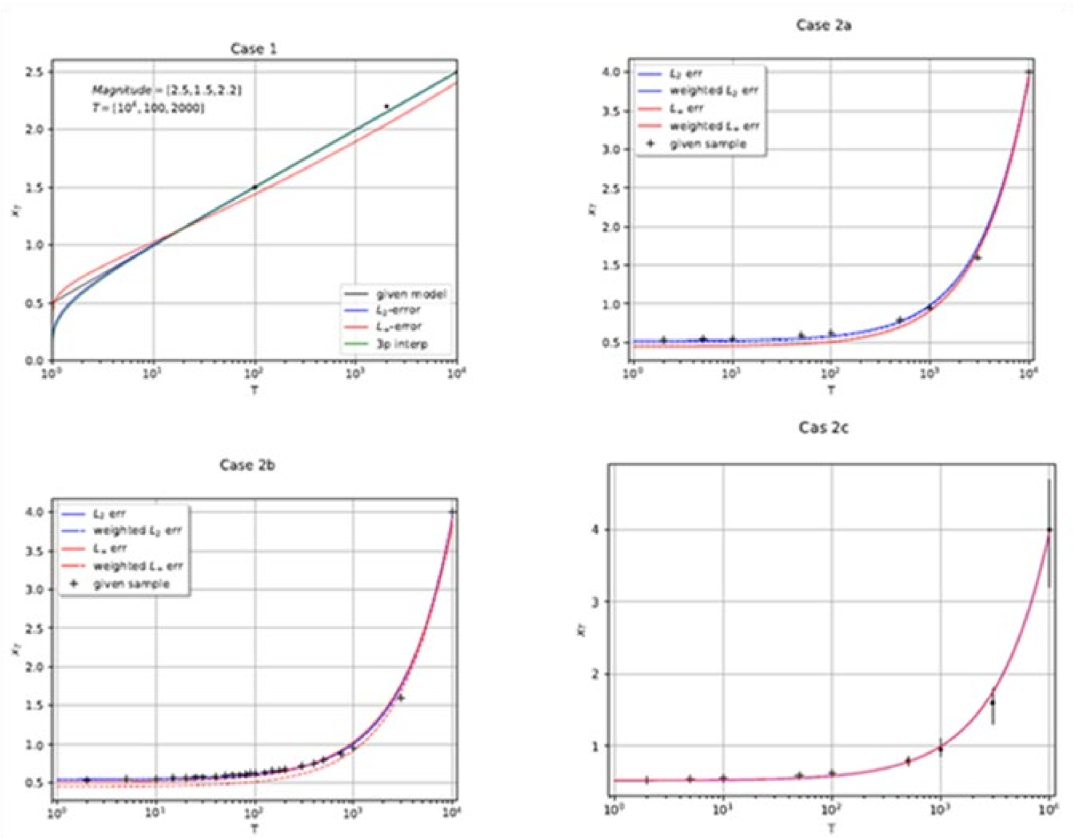


Table A1.1. Case 1: Comparison between the known model g and the optimal GEV model obtained using several error criteria

Return period (years)	2	5	10	50	100	500	1 000	2 000	10 000
Ref Mag (m)	0.650	0.849	1.0	1.349	1.5	1.849	2.0	2.151	2.5
Approx Mag (m) - 3p interp	0.582	0.828	0.99	1.349	1.5	1.85	2.0	2.151	2.5
Approx Mag (m) - L_2-error	0.600	0.841	1.0	1.353	1.503	1.850	2.0	2.150	2.50
Approx Mag (m) - L_8-error	0.724	0.901	1.02	1.31	1.44	1.75	1.90	2.04	2.40

Table A1.2. Case 2a: Comparison between the given synthetic data and the optimal GEV model obtained using several error criteria

Return period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Ref Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m) - L2-error	0.523	0.526	0.530	0.556	0.584	0.776	0.986	1.730	3.959
Approx Mag (m) - weighted L2-error	0.505	0.507	0.511	0.537	0.566	0.759	0.971	1.719	3.969
Approx Mag (m) - L8-error	0.479	0.482	0.486	0.511	0.540	0.732	0.942	1.685	3.915
Approx Mag (m) - weighted L8-error	0.445	0.448	0.452	0.478	0.507	0.701	0.914	1.669	3.940

Table A1.3. Case 2b: Comparison between the given synthetic data and the optimal GEV model using the L2-error, using several error criteria

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Ref Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m) - L2-error	0.543	0.546	0.550	0.576	0.604	0.795	1.00	1.741	3.952
Approx Mag (m) - weighted L2-error	0.538	0.541	0.545	0.571	0.599	0.790	1.00	1.74	3.96
Approx Mag (m) - L8-error	0.479	0.482	0.486	0.512	0.540	0.732	0.942	1.685	3.916
Approx Mag (m) - weighted L8-error	0.448	0.451	0.455	0.481	0.510	0.704	0.917	1.67	3.94
Return Period (years)	15	20	25	30	40	60	70	80	90
Ref Mag (m)	0.56	0.56	0.57	0.57	0.58	0.60	0.60	0.61	0.62
Approx Mag (m) - L2-error	0.554	0.557	0.560	0.564	0.570	0.582	0.587	0.593	0.599
Approx Mag (m) - weighted L2-error	0.548	0.552	0.555	0.558	0.564	0.576	0.582	0.588	0.593
Approx Mag (m) - L8-error	0.489	0.493	0.496	0.499	0.506	0.518	0.523	0.529	0.535
Approx Mag (m) - weighted L8-error	0.459	0.462	0.465	0.469	0.475	0.487	0.493	0.499	0.504
Return Period (years)	125	150	175	200	300	400	750		
Ref Mag (m)	0.63	0.65	0.66	0.67	0.71	0.75	0.87		
Approx Mag (m) - L2-error	0.618	0.631	0.643	0.656	0.704	0.750	0.902		

Table A1.3. Case 2b: Comparison between the given synthetic data and the optimal GEV model using the L2-error, using several error criteria (Continued)

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Approx Mag (m) - weighted L2-error	0.612	0.626	0.638	0.651	0.699	0.745	0.897		
Approx Mag (m) - L8-error	0.554	0.567	0.580	0.592	0.641	0.687	0.840		
Approx Mag (m) - weighted L8-error	0.523	0.537	0.550	0.562	0.611	0.658	0.813		

Table A1.4. Case 2c: Comparison between the given synthetic data and the optimal GEV model minimising several features of the random error

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Ref Mag (mean) (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m) - Mean	0.525	0.527	0.531	0.558	0.587	0.781	0.993	1.73	3.94
Approx Mag (m) - Quantile 95%	0.517	0.519	0.523	0.551	0.581	0.778	0.994	1.75	3.98

A1.1 Introduction

A1.1.1. The OECD context

The modelling of external hazards encompasses different technical aspects depending on the type of hazard, all of them impacting nuclear installations. There is also a link between the hazard and the associated risk analysis. A risk analysis contains a set of scenarios, frequencies and associated consequences, developed in such a way as to inform decisions. A scenario contains an initiating event and (usually) one or more subsequent events leading to an end state that reflects the issue of concern.

The objective of the benchmark study launched by the OECD NEA is to focus on the IE by facilitating an exercise on the statistical modelling for assessing hazard frequency and magnitude for external event risk assessment. The benchmark study provides synthetic data for a hypothetical external event.

The analysis steps and modelling results we obtained are detailed on this report.

We would like to underline that the scope of this benchmark differs from the activities that EDF usually performs regarding extreme natural events characterisation. Indeed, what is usually available is a set of measured data for a given phenomenon, and what is performed is statistical extrapolation of these data to evaluate a magnitude of the natural phenomenon for a high return level period.

A1.1.2. Data

Hypothetical observational data represent an external hazard which has been created from synthetic models (this type of model is used to create synthetic data that have been generated from a computer). Two cases are described:

- Case 1: a fully revealed open case where both the synthetic data and the synthetic model producing the data are provided;
- Case 2: a blind-test case where only the synthetic data are provided.

Case 1 - Known model producing the synthetic data

The synthetic model used for the first exercise is:

$$g : T \rightarrow g(T) = 0.5 + 0.5\log_{10}(T) \quad (1)$$

where T is the return period in years and $M = g(T)$ the associated magnitude given in metres. This model led to the return period / magnitude detailed in Table A1.5.

Table A1.5. Case 1: synthetic data

Return period (years)	1	2	5	10	50	100	500	1 000	2 000	10 000
Mag (m)	0.50	0.65	0.85	1.0	1.4	1.5	1.9	2.0	2.2	2.5

Case 2 - Unknown model producing the synthetic data

In this case, only the synthetic data are provided (the synthetic model used is not provided). The Case 2 presents three subcategories:

- On the first two, the synthetic data have no uncertainty and the last has uncertainty estimates on some of the data.
- The three-point interpolation would also be applicable for the 2a case data set (without uncertainty). In order to focus on presenting the differences between a data set with and without uncertainties, we focused the Case 2 studies on the L_2 and L_∞-error models.

Case 2a: The synthetic data (ten data points) with no uncertainty on the points are provided and gathered in Table A1.6.

Table A1.6. Case 2a: synthetic data

Return period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0

Case 2b: In addition to the synthetic data of Case 2a, 16 data points are provided without uncertainty. The additional data are seven.

Table A1.7. Case 2b: additional synthetic data with respect to Case 2a

Return period (years)	15	20	25	30	40	60	70	80
Mag	0.56	0.56	0.57	0.57	0.58	0.60	0.60	0.61

Return period (years)	90	125	150	175	200	300	400	750
Mag (m)	0.62	0.63	0.65	0.66	0.67	0.71	0.75	0.87

Care: In Table A1.7, we changed the value given in Table 2.3 of the benchmark (benchmark numeration). Case 2b: Comparison between the given synthetic data and the optimal GEV model using the L2-error, using several error criteria of the benchmark (benchmark numeration) at the period T = 3 000 years from the value 1.57 to the value 1.6 and the value given in the benchmark at the period T = 10 000 years from the value 3.97 to the value 4.0, in order to make the data Table A1.7. Case 2b: Comparison between the given synthetic data and the optimal GEV model using the L2-error, using several error criteria (benchmark numeration) coherent with the data of Table A1.6. Case 2a: Comparison between the given synthetic data and the optimal GEV model obtained using several error criteria (benchmark numeration).

Case 2c: Here, the synthetic data for long time intervals (500 years and longer) are presumed to not be known exactly. For these return periods, the uncertainty on the magnitude is provided. The other points ($\leq$ 500 years) have no uncertainty on the associated magnitude. The synthetic data (ten data points) are gathered in Table A1.8.

Table A1.8. Case 2c: known return levels and uncertain ones

Return period (years)	1	2	5	10	50	100	500	1 000	3 000	10 000
Mean	0.53	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
St. dev.	-	-	-	-	-	-	0.04	0.06	0.15	0.46
q0.05	-	-	-	-	-	-	0.72	0.85	1.3	3.2
q0.95	-	-	-	-	-	-	0.85	1.1	1.8	4.7

We can note that the quantiles $q_{0.05}$ and $q_{0.95}$ given in Table A1.11 correspond to those of a normal distribution for which the mean and standard deviation are those given in the table.

Thus, we will model the uncertainty by a normal distribution as follows:

$$X_{500} \sim Normal(\mu = 0.79, \sigma = 0.04) \qquad (2)$$

$$X_{1000} \sim Normal(\mu = 0.95, \sigma = 0.06) \qquad (3)$$

$$X_{3000} \sim Normal(\mu = 1.6, \sigma = 0.15) \qquad (4)$$

$$X_{10\,000} \sim Normal(\mu = 4.0, \sigma = 0.465) \qquad (5)$$

We consider that these four random variables are independent.

Notations

Let us note $G_{\boldsymbol{\theta}}$ the cumulated density function of a GEV distribution para- metered by $\boldsymbol{\theta} = (\mu, \sigma, \xi)$. The Extreme Value Theory shows that the annual maximum of the underlying process follows a GEV distribution.

The T -return level $q_{\boldsymbol{\theta}}(T)$ is by definition the quantile of order $(1 - 1/T)$ of the GEV distribution :

$$G_{\boldsymbol{\theta}}(q_{\boldsymbol{\theta}}(T)) = 1 - 1/T \qquad (6)$$

The T -return level is evaluated from the relation:

$$q_{\boldsymbol{\theta}}(T) = \left| \begin{array}{ll} \mu - \frac{\sigma}{\xi}\left\{1 - [-\log(1-1/T)]^{-\xi}\right\} & \text{for } \xi \neq 0, \\ \mu - \sigma \log(-\log(1-1/T)) & \text{for } \xi = 0. \end{array} \right. \qquad (7)$$

In this report, we want to fit the GEV distribution that best predicts the return levels given either by a known model (Case 1) or by some synthetic data (Case 2).

A1.2 Case 1 - Know model producing the synthetic data

In Case 1, the model g that produced the synthetic data is given by (1).

The objective of the study is to find the optimal parameter $\boldsymbol{\theta} = (\mu, \sigma, \xi)$ so that the GEV distribution parametered by $\boldsymbol{\theta}$ fits the best frequency-magnitude relation (1).

To find the optimal $\boldsymbol{\theta}$, we first considered a 3-points interpolation according to the relation (37). Then, we consider a criteria to evaluate the model error between g and $q_{\boldsymbol{\theta}}$, noted $Err(g, q_{\boldsymbol{\theta}})$ and we find $\boldsymbol{\theta}^*$ defined by :

$$\theta^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, Err(g, q_{\boldsymbol{\theta}}) \quad (8)$$

In our study, we tested different criteria evaluating the model error $Err(g, q_{\boldsymbol{\theta}})$:

- the L_2-error based criteria, which cumulates the errors at each return period T: see Section A1.2.2;
- the L_∞-error based criteria, which focuses on the maximal error over the return period T : see Section A.1.2.3.

A1.2.1. The 3-points interpolation

The GEV distribution being defined by three parameters, we consider a system based on three equations as follows:

$$\begin{cases} q_{\boldsymbol{\theta}}(T_1) = M_1 \\ q_{\boldsymbol{\theta}}(T_2) = M_2 \\ q_{\boldsymbol{\theta}}(T_3) = M_3 \end{cases}$$

where $(T_i, M_i)_{i=1,2,3}$ represents the return period and its associated magnitude taken in the Table A1.5.

To determine the parameter $\theta = (\mu, \sigma, \xi)$, solution of this system, the parameters μ and σ are defined as a function of ξ : $\mu = f^1{}_{(T_i,M_i)}(\xi)$ and $\sigma = f^2{}_{(T_i,M_i)}(\xi)$, with f^1 and f^2 known. Then, we reach the parameter ξ, solution of the equation $f^3{}_{(T_i,M_i)}(\xi) = 0$, using the Brent's algorithm. Note that the functions f^1, f^2 and f^3 are presented in Annex A. Once the optimal ξ^* is obtained, we deduce $\mu^* = f^1{}_{(T_i,M_i)}(\xi^*)$ and $\sigma^* = f^2{}_{(T_i,M_i)}(\xi^*)$.

The optimal parameter with $(T_i, M_i)_{i=1,2,3} = \{(10\,000, 2.5), (100, 1.5), (2\,000, 2.2)\}$ is:

$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.502162, 0.216915, -1.1 \cdot 10^{-6}) \quad (9)$$

These points were chosen because they lead to a better approximation after several tests performed in the Annex A.

With ξ^* near to 0, we can say that the hazard frequency-magnitude relation follows a Gumbel distribution with parameters $(\mu^*, \sigma^*) = (0.502162, 0.216915)$.

Figure A1.2 draws the T-return level obtained with the optimal GEV distribution and the model g.

Table A1.9 details the T-return levels obtained with the optimal GEV for several periods T and gives the relative error made for each of them (in %) defined by:

$$\varepsilon(T) = 100 \times \left| \frac{g(T) - q_{\boldsymbol{\theta}}(T)}{g(T)} \right| \quad (10)$$

Figure A1.3 draws the relative error function: T → ε(T).

Table A1.9. Case 1: Comparison between the known model g and the optimal GEV model obtained with a 3-points interpolation (in %)

Return period (years)	2	5	10	50	100	500	1 000	2 000	10 000
Ref Mag (m)	0.650	0.849	1.0	1.349	1.5	1.849	2.0	2.151	2.5
Approx Mag (m)	0.582	0.828	0.99	1.349	1.5	185	2.0	2.151	2.5
$\varepsilon(T)$	10.58	2.57	9.7e-1	6.94e-2	0	2.68e-2	2.21e-2	1.55e-2	0

A1.2.2. L2-error

We define the error as the square of the L_2-norm of $(g - q_{\boldsymbol{\theta}})$. In other words, we consider the integrated squared error at each point T between the model $g(T)$ and $q_{\boldsymbol{\theta}}(T)$:

$$Err^2(g, q_{\boldsymbol{\theta}}) = \|g - q_{\boldsymbol{\theta}}\|^2_{L_2} = \int_0^{+\infty} (g(T) - q_{\boldsymbol{\theta}}(T))^2 \, dT \quad (11)$$

To compute the integration, we use the Gauss-Legendre algorithm to approximate the integral with a finite sum of the integrand evaluated on some judicious points. Using the TNC (Truncated Newton Constrained) optimisation algorithm with 100 different starting points, we obtain the following optimal point:

$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.522748, 0.211424, 0.0032985) \quad (12)$$

The associated error is: $Err^* = 0.1356$

As the 3-points interpolation, the hazard frequency-magnitude relation is defined following the Gumbel distribution with parameters $\boldsymbol{\theta}^* = (\mu^*, \sigma^*) = (0.522748, 0.211424)$.

Figure A.1.2. Case 1: Frequency - Magnitude relation obtained with the optimal GEV distribution (in green) that best predicts the known model g (in black) with 3 points: in natural scale (left) and in logscale (right). Zoom on the interval [1, 100] years on the first line.

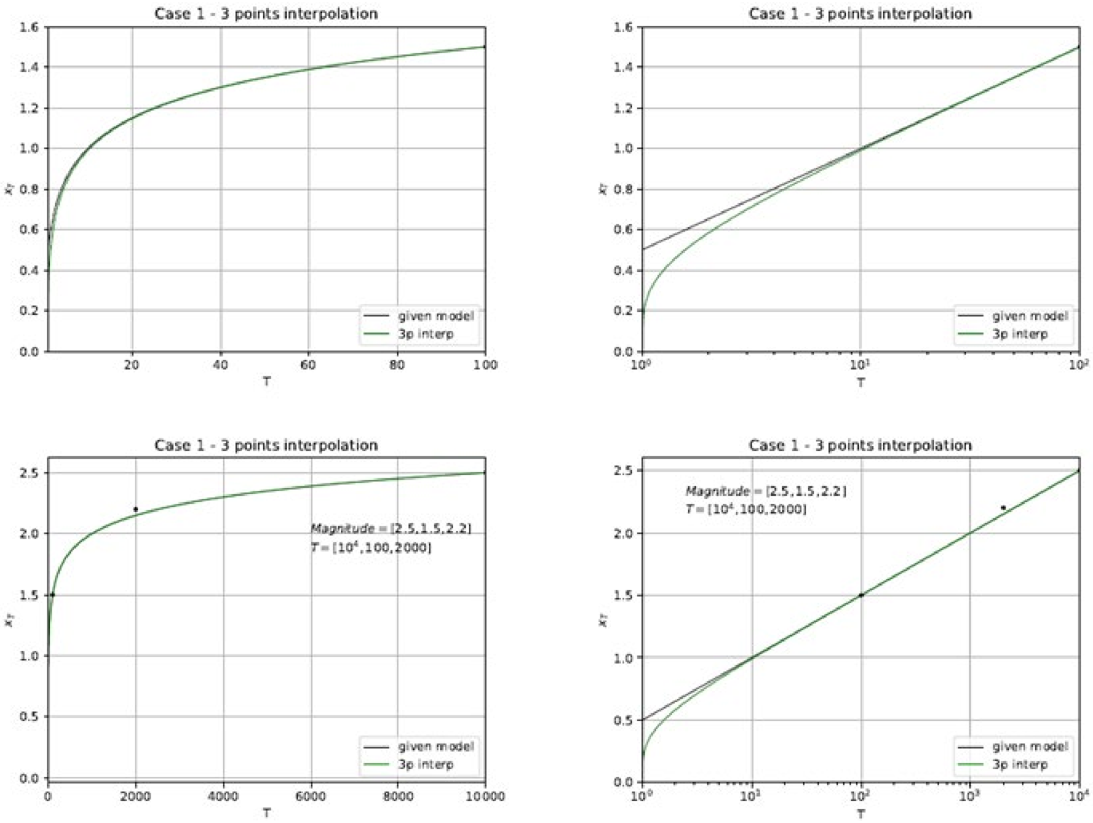
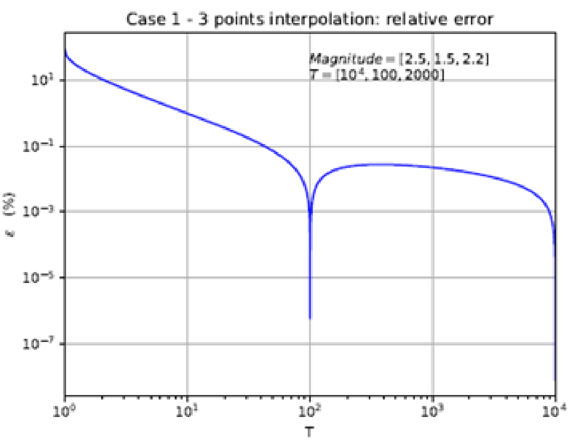


Figure A1.3. Case 1: Relative error between the optimal GEV distribution and the given model g.



optimisation algorithm with 100 different starting points, we obtain the following optimal point:

$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.522748, 0.211424, 0.0032985) \quad (12)$$

The associated error is: $Err^* = 0.1356$

As the 3-points interpolation, the hazard frequency-magnitude relation is defined following the Gumbel distribution with parameters $\boldsymbol{\theta}* = (\mu*, \sigma*) = (0.522748, 0.211424)$.

Figure A1.4 draws the T -return level obtained with the optimal GEV distribution and the model g.

Table A1.5 details the T -return levels obtained with the optimal GEV for several periods T and the relative error made for each of them (in %).

Figure A1.6 draws the relative error function: T → ε(T).

Figure A1.4. Case 1 Frequency - Magnitude relation obtained with the optimal GEV distribution (in red) that best predicts the known model g (in black) with respect to the L2-error: in natural scale (left) and in logscale (right). Zoom on the interval [1, 100] years on the first lines.

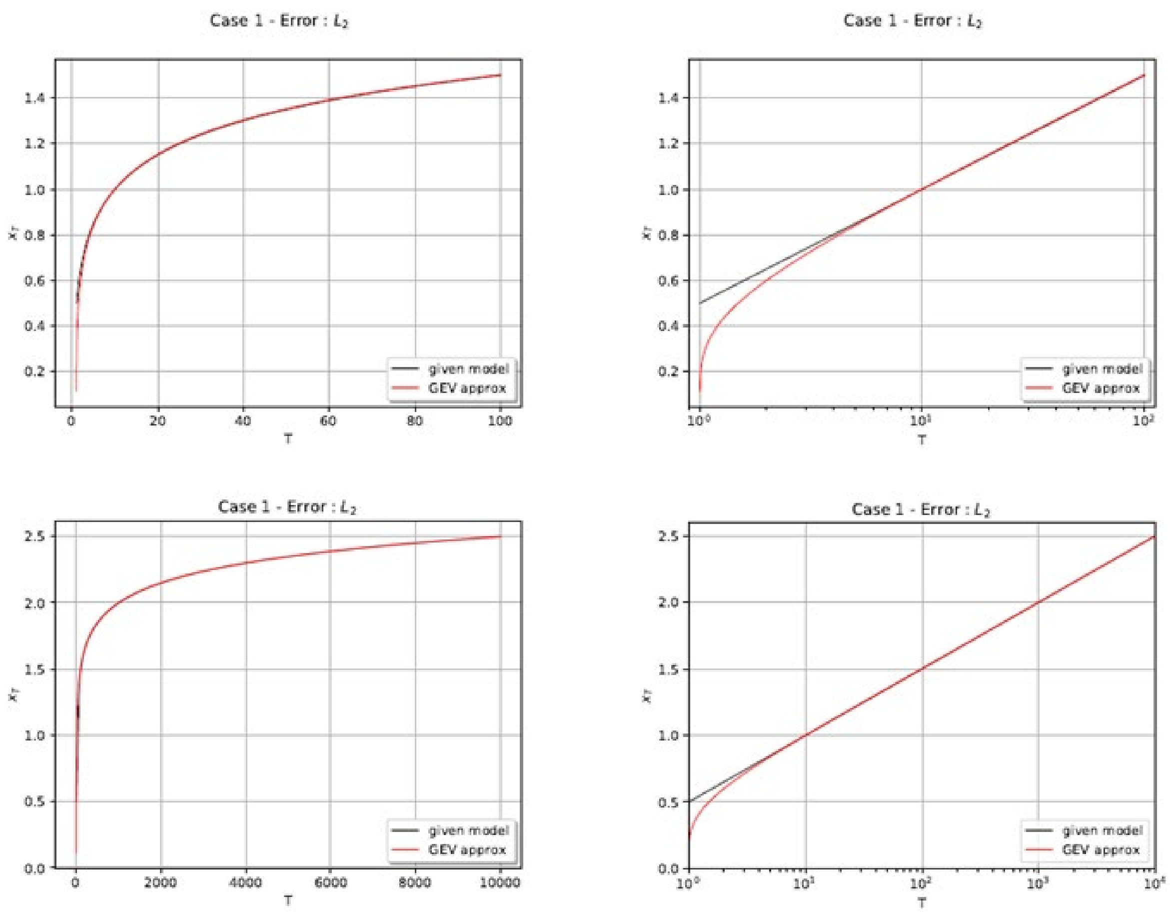
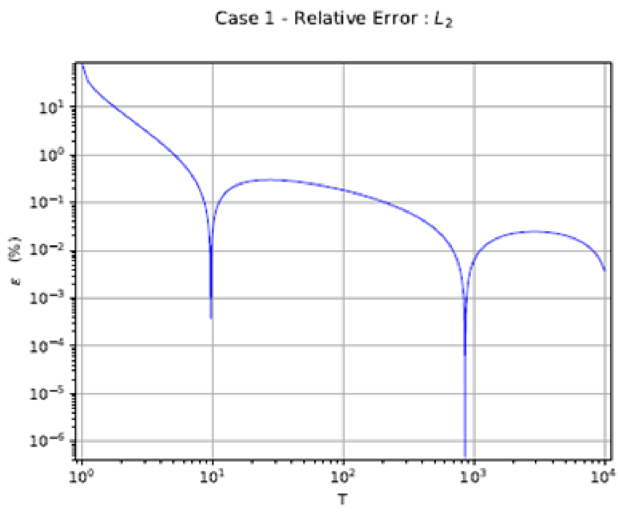


Table A1.10. Case 1: Comparison between the known model g and the optimal GEV model obtained with the relative L2-error (in %)

Return period (years)	2	5	10	50	100	500	1 000	2 000	10 000
Ref Mag (m)	0.650	0.849	1.0	1.349	1.5	1.849	2.0	2.151	2.5
Approx Mag (m)	0.600	0.841	1.0	1.353	1.503	1.850	2.0	2.150	2.50
e(*T*)	7.72	1.04	3.0e-2	2.64e-1	1.83e-1	2.83e-2	6.43e-3	2.29e-2	3.74e-3

Figure A1.5. Case 1: Relative L2-error between the optimal GEV distribution and the given model g.



A1.2.3. L_∞-error

We define the error as the L_∞-norm of $(g - q_\theta)$. In other words, we consider the maximum absolute error between $g(T)$ and $q_\theta(T)$:

$$Err(g, q_\theta) = \|g - q_\theta\|_\infty = \max |g(T) - q_\theta(T)| \tag{13}$$

To solve the optimisation problem, we use the Cobyla algorithm initialised with the optimal point obtained with the L_2-error.

We obtain the following optimal point:

$$\theta^* = (\mu^*, \sigma^*, \xi^*) = (0.668508, 0.149587, 0.0484294) \tag{14}$$

The associated error is:

$$Err^* = 0.107711$$

Figure A1.6 draws the T -return level obtained with the optimal GEV distribution and the model g.

Table details the T -return level obtained with the optimal GEV for several periods T and gives the relative error made for each of them (in %) defined by (10).

Figure A1.7 draws the relative error function: T → ε(T).

Figure A1.6. Case 1: Frequency - Magnitude relation obtained with the optimal GEV distribution (in red) that best predicts the known model g (in black) with respect to the L∞-error: in natural scale (left) and in logscale (right). Zoom on the interval [1, 100] years on the first line.

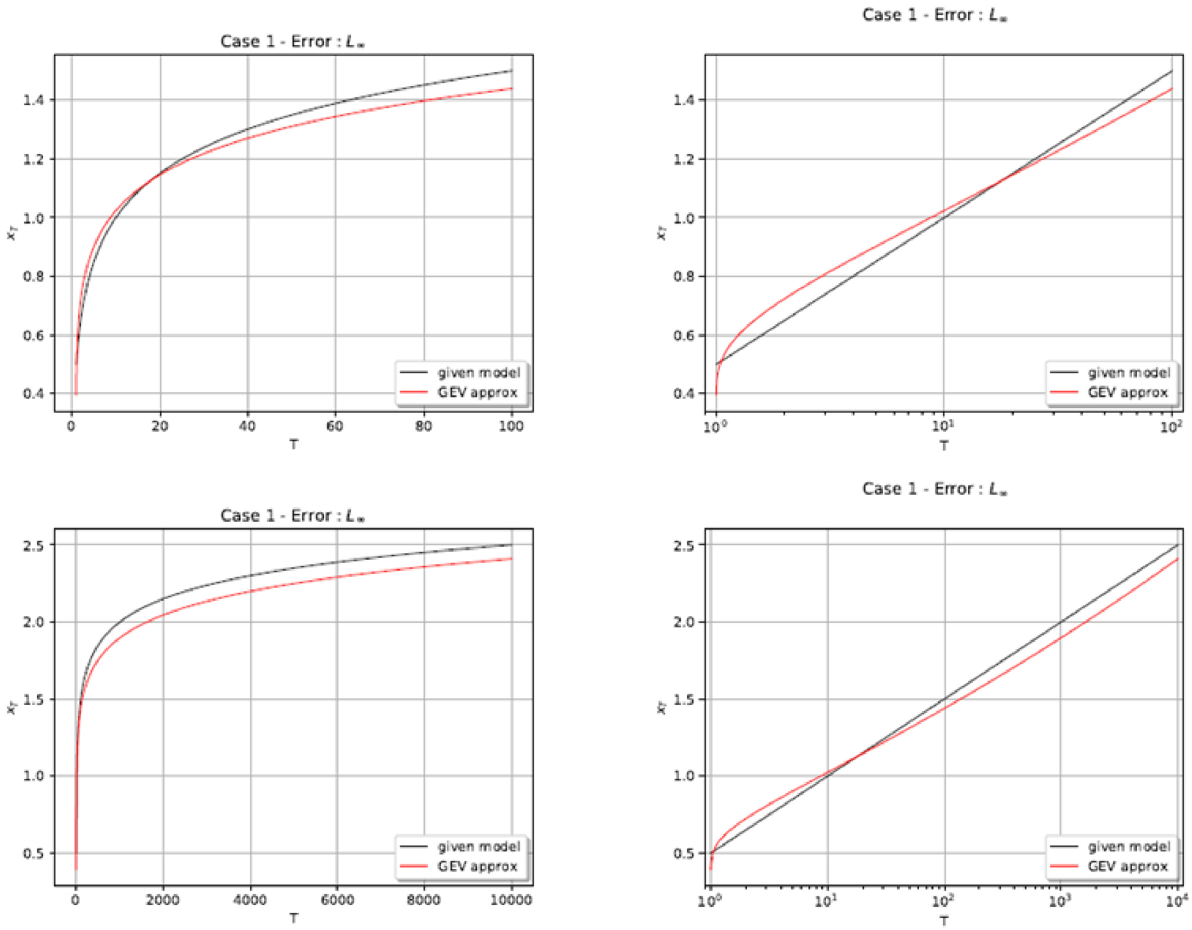


Figure A1.7. Case 1: Relative L∞-error between the optimal GEV distribution and the given model g

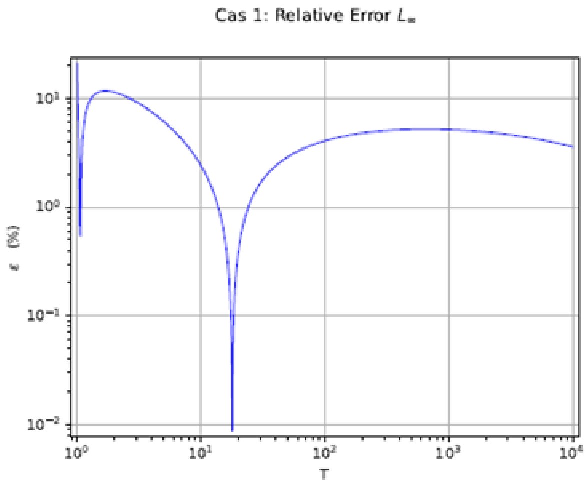


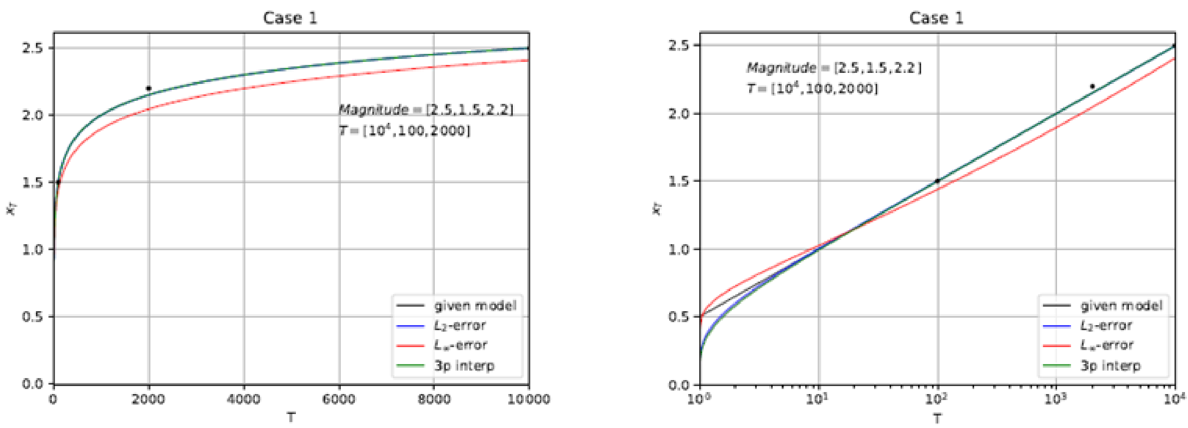
Table A1.11. Case 1: Comparison between the known model g and the optimal GEV model obtained with the L∞-relative error (in %)

Return period (years)	2	5	10	50	100	500	1 000	2 000	10 000
Ref Mag (m)	0.650	0.849	1.0	1.349	1.5	1.849	2.0	2.151	2.5
Approx Mag (m)	0.724	0.901	1.02	1.31	1.44	1.75	1.90	2.04	2.40
e(*T*)	11.3	6.09	2.41	2.85	4.05	5.22	5.22	5.0	3.81

A1.2.4. Models comparison

In order to ease comparison, Figure A1.8 draws both models obtained with 3-points interpolation and with both criteria's (11) and (13).

Figure A1.8. Case 1: Frequency - Magnitude obtained with the optimal GEV distribution (in blue, red and green) that best predicts the known model g (in black) with respect to the L2-error (in blue), the L∞-error (in red) and with 3-points interpolation (in green): in natural scale (left) and in logscale (right).



A1.3 Case 2 - Unknown model producing the synthetic data

For Case 2, the synthetic data is provided without the associated model.

A1.3.1. Case 2a

In Case 2a, we use the synthetic data of Table A1.6.

A1.3.1.1. L2-error

We define the following error criteria:

(15)
$$Err^2(S_a, \boldsymbol{\theta}) = \sum_{i=1}^{10} (x_{T_i} - q_{\boldsymbol{\theta}}(T_i))^2$$

where $S_a = (T_i, x_{T_i})_{1 \leq i \leq 10}$ are the synthetic data of Table A1.6, with T_i the return period and x_{T_i} the associated magnitude.

Using the TNC optimisation algorithm with 100 different starting points, we got the following optimal point:

$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.522859, 0.001, 0.868784) \quad (16)$$

The associated error is:

$$Err^* = 0.1517$$

Figure A1.9 draws the T -return level obtained with the optimal GEV distribution and the sample S_a.

Figure A1.9. Case 2a: Frequency - Magnitude relation obtained with the optimal GEV model (in blue) that best predicts the given synthetic data (in black) with respect to the L2-error: in natural scale (left) and in logscale (right).

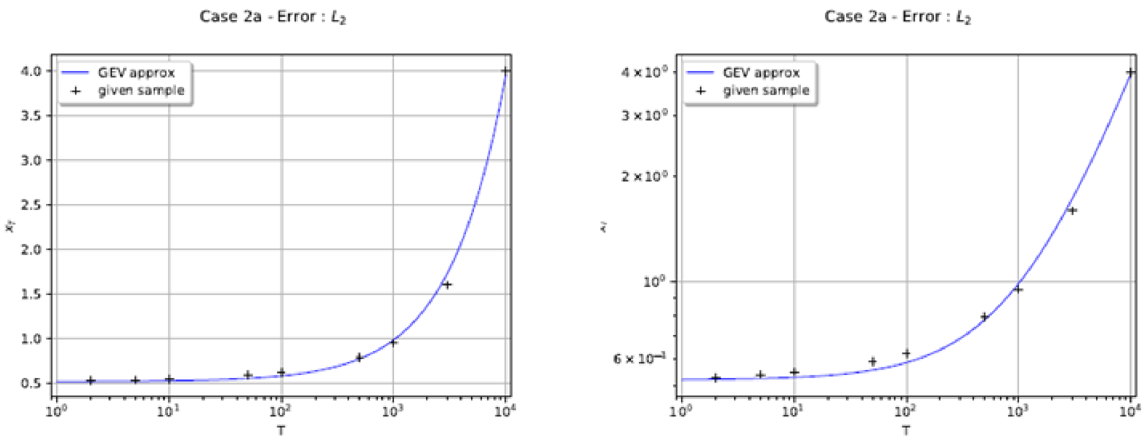


Table details the T -return level obtained with the optimal GEV for several return periods T and gives the relative error made for each of them (in %) defined by:

$$\varepsilon(T_i) = 100 \times \left| \frac{x_{T_i} - q_{\boldsymbol{\theta}}(T_i)}{x_{T_i}} \right| \tag{17}$$

Table A1.12. Case 2a: Comparison between the given synthetic data and the optimal GEV model using the L2-error, with the relative error in %.

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Ref Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m)	0.523	0.526	0.530	0.556	0.584	0.776	0.986	1.730	3.959
e(*T*)	1.27	2.60	3.66	5.79	5.75	1.76	3.84	8.08	1.03

A1.3.1.2. Weighted L2-error

In order to force a best adequation over the large return periods, we consider the weighted least square error defined by:

$$Err^2(S_a, q_{\boldsymbol{\theta}}) = \sum_{i=1}^{10} \log(T_i)(x_{T_i} - q_{\boldsymbol{\theta}}(T_i))^2 \tag{18}$$

where $S_a = (T_i, x_{T_i})_{1\leq i\leq 10}$ are the synthetic data of Table A1.6, with T_i the return period and x_{T_i} the associated magnitude.

Using this weighted error, we penalise the errors on large return periods, considering, for example, that we prefer getting a model that predicts better return levels associated to large return periods even if the model works worse on low return periods.

Using the TNC optimisation algorithm with 100 different starting points, we obtain the following optimal point:

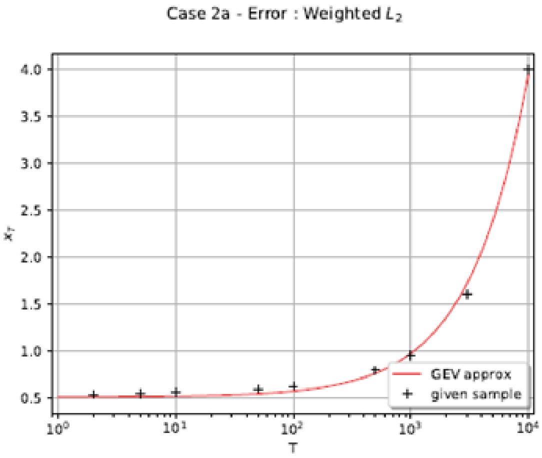
$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.504188, 0.001, 0.869813) \tag{19}$$

The associated error is:

$$Err^* = 0.4018$$

Figure A1.10 draws the T -return level obtained with the optimal GEV distribution and the sample S_a.

Figure A1.10. Case 2a: Frequency - Magnitude relation obtained with the optimal GEV model (in red) that best predicts the given synthetic data (in black) with respect to the weighted L2-error: in natural scale (left) and in logscale (right)



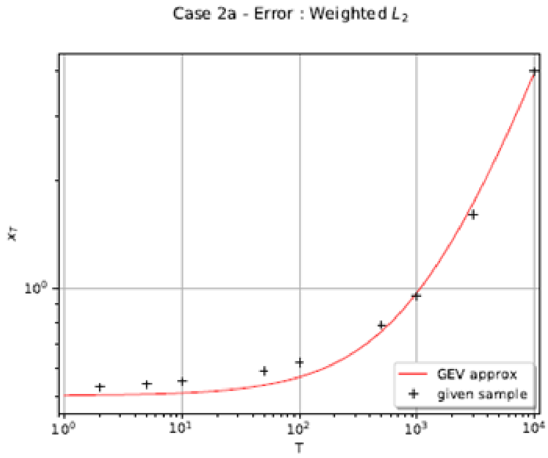


Table A1.13 details the T-return level obtained with the optimal GEV for several periods T and gives the relative error made for each of them (in %) defined by (17). We verify that although the global error is larger, the optimal GEV distribution fits better the data for periods larger than 500 years.

Table A1.13. Case 2a: Comparison between the given synthetic data and the optimal GEV model using the weighted L2-error, with the relative error (%).

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Ref Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m)	0.505	0.507	0.511	0.537	0.566	0.759	0.971	1.719	3.969
$e(T)$	4.79	6.06	7.06	8.94	8.73	3.95	2.17	7.44	0.78

A1.3.1.3. L∞-error

We define the following error criteria:

$$Err(S_a, q_{\boldsymbol{\theta}}) = \max |x_{T_i} - q_{\boldsymbol{\theta}}(T_i)| \qquad (20)$$

where $S_a = (T_i, x_{T_i})_{1\leq i\leq 10}$ are the synthetic data of Table A1.6, with T_i the return period and x_{T_i} the associated magnitude.

Using the TNC optimisation algorithm with the optimal point obtained with the L_2-error as a starting point, we obtain the following optimal point:

$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.478532, 0.00100012, 0.868792) \qquad (21)$$

The associated error is : $Err^* = 0.0851144$

Figure A1.11 draws the T-return level obtained from the optimal GEV distribution and the sample S_a.

Table A1.14 details the T-return level obtained from the optimal GEV for several periods T and gives the relative error made for each of them (in %).

Figure A1.11. Case 2a: Frequency - Magnitude relation obtained with the optimal GEV model (in blue) that best predicts the given synthetic data (in black) with respect to the L∞-error: in natural scale (left) and in logscale (right).

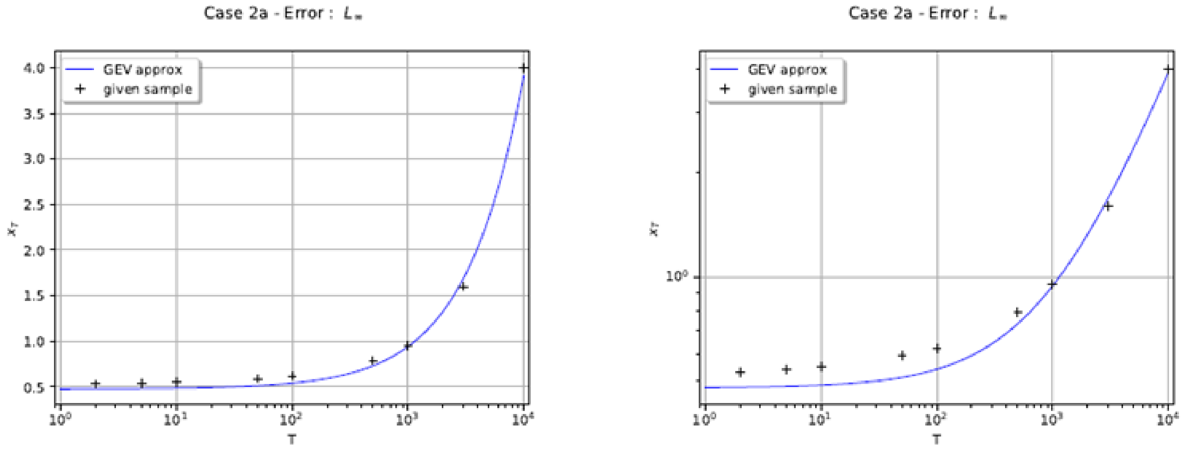


Table A1.14. Case 2a: Comparison between the given synthetic data and the optimal GEV model using the L∞-error, with the relative error (%).

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Ref Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m)	0.479	0.482	0.486	0.511	0.540	0.732	0.942	1.685	3.915
e(*T*)	9.63	10.8	11.7	13.3	12.9	7.36	8.12e-1	5.32	2.12

A1.3.1.4. Weighted L_∞-error

In order to force a best adequation over the large return periods T, we consider the weighted error defined by

$$Err(S_a, q_\theta) = \max |\log(T_i)[x_{T_i} - q_\theta(T_i)]| \qquad (22)$$

where $S_a = (T_i, x_{T_i})_{1\leq i\leq 10}$ are the synthetic data of Table 1.2, with T_i the return period and x_{T_i} the associated magnitude.

Using this weighted error, we penalise the errors on large T, considering, for example, that we prefer getting a model that predicts better large return levels even if the model works worse on little return levels.

Using the TNC optimisation algorithm with the optimal point obtained with the L_2-error as starting point, we obtain the following optimal point:

$$\theta^* = (\mu^*, \sigma^*, \xi^*) = (0.44483, 0.001, 0.870939) \qquad (23)$$

The associated error is: $Err^* = 0.553827$

Figure A1.12 draws the T-return level obtained from the optimal GEV distribution and the sample S_a.

Table A1.15 details the T-return level obtained from the optimal GEV for several periods T and gives the relative error made for each of them (in %) defined by (17). We note that weighting the error for large periods has changed very little the optimal solution: it means that the maximum error was reached for large T , so that the penalisation was not efficient.

Figure A1.12. Case 2a: Frequency - Magnitude relation obtained with the optimal GEV model (in blue) that best predicts the given synthetic data (in black) with respect to the weighted L∞-error: in natural scale (left) and in logscale (right).

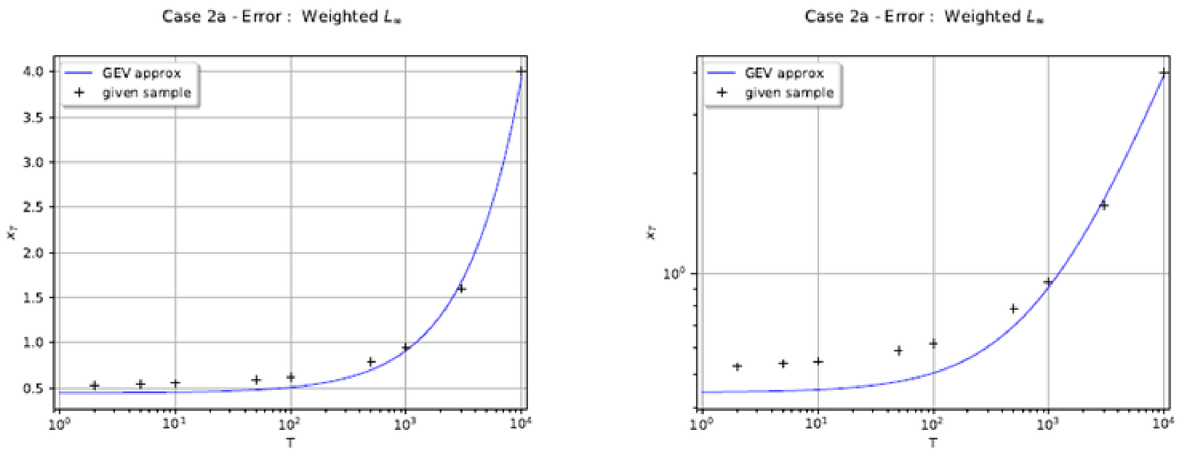


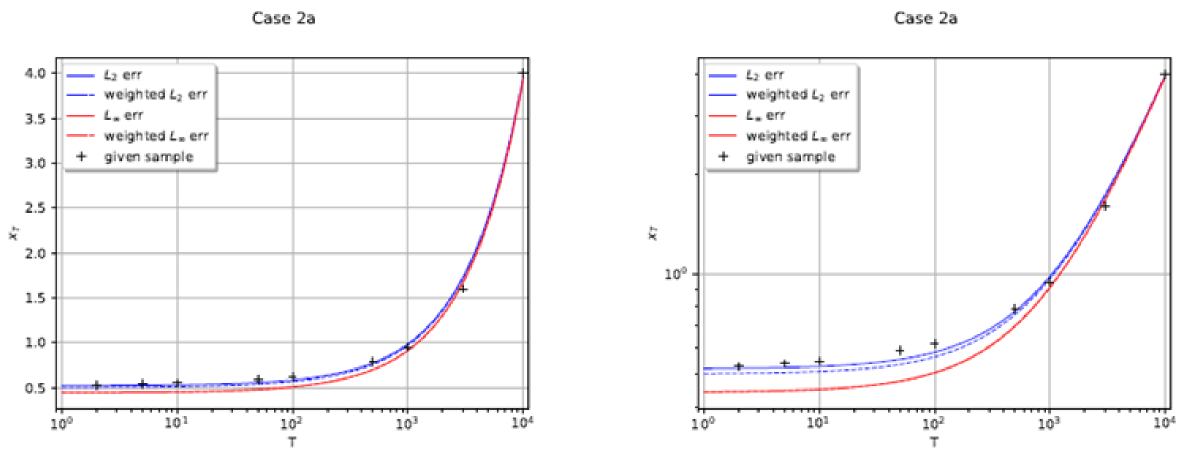
Table A1.15. Case 2a: Comparison between the given synthetic data and the optimal GEV model using the weighted L∞-error, with the relative error (%).

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Ref Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m)	0.445	0.448	0.452	0.478	0.507	0.701	0.914	1.669	3.940
e(*T*)	15.98	17.05	17.85	18.98	18.26	11.28	3.764	4.318	1.479

A1.3.1.5. Models comparison

In order to ease comparison, Figure A1.13 draws both models obtained from both criteria (15), (18), (20) and (22).

Figure A1.13. Case 2a: Frequency - Magnitude relation obtained with the optimal GEV model (in blue and red) that best predicts the given synthetic data (in black): in natural scale (left) and in logscale (right).



A1.3.2. Case 2b

In Case 2b, we use the synthetic data of Table A1.6 and the additional ones of Table A1.7.

A1.3.2.1. L2-error

We define the following error criteria:

$$Err^2(S_b, \boldsymbol{\theta}) = \sum_{i=1}^{26} (x_{T_i} - q_{\boldsymbol{\theta}}(T_i))^2 \qquad (24)$$

where $S_b = (T_i, x_{T_i})_{1 \le i \le 26}$ are the synthetic data of Tables A1.6 and A1.7, with T_i the return period and $x_{T_i}$the associated magnitude.

Using the TNC optimisation algorithm with 100 different starting points, we got the following optimal point:

$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.542937, 0.001, 0.867794) \qquad (25)$$

The associated error is:

$$Err^* = 0.1715$$

Figure A1.14 draws the T -return level obtained from the optimal GEV distribution and the sample S_b.

Table A1.16 details the T -return level obtained from the optimal GEV for several periods T and gives the relative error made for each of them (in %) defined by (17).

Figure A1.14. Case 2b: Frequency - Magnitude relation obtained with the optimal GEV model (in blue) that best predicts the given synthetic data (in black) with respect to the L2-error: in natural scale (left) and in logscale (right).

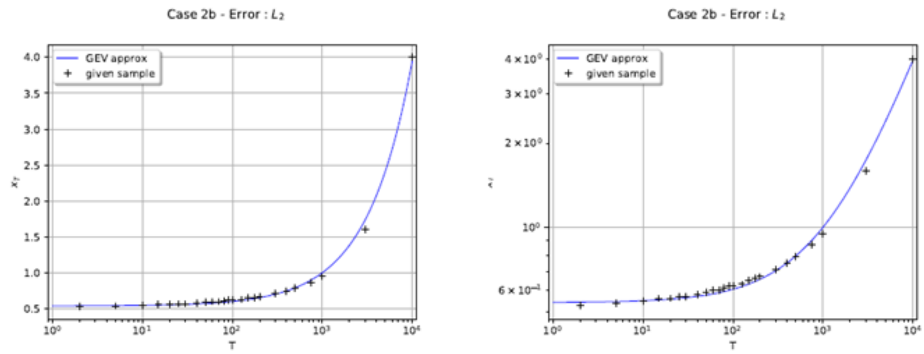


Table A1.16. Case 2b: Comparison between the given synthetic data and the optimal GEV model using the L2-error, with the relative error (%).

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Ref Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m)	0.543	0.546	0.550	0.576	0.604	0.795	1.00	1.741	3.952
e(*T*)	2.52	1.11	1.68e-2	2.40	2.55	6.23e-1	5.68	8.82	1.21
Return Period (years)	15	20	25	30	40	60	70	80	90
Ref Mag (m)	0.56	0.56	0.57	0.57	0.58	0.60	0.60	0.61	0.62
Approx Mag (m)	0.554	0.557	0.560	0.564	0.570	0.582	0.587	0.593	0.599
e(*T*)	1.16	5.44e-1	1.71	1.14	1.76	3.04	2.08	2.76	3.43
Return Period (years)	125	150	175	200	300	400	750		
Ref Mag (m)	0.63	0.65	0.66	0.67	0.71	0.75	0.87		
Approx Mag (m)	0.618	0.631	0.643	0.656	0.704	0.750	0.902		
e(*T*)	1.97	2.98	2.51	2.10	8.19e-1	4.17e-2	3.65		

A1.3.2.2. Weighted L2-error

In order to force a best adequation over the large periods T, we consider the weighted least square error defined by:

$$Err^2(S_b, q_{\boldsymbol{\theta}}) = \sum_{i=1}^{26} \log(T_i)(x_{T_i} - q_{\boldsymbol{\theta}}(T_i))^2 \qquad (26)$$

where $S_b = (T_i,\ x_{T_i})_{1\le i\le 26}$ are the synthetic data of Table A1.6 and Table A1.7, with T_i the return period and x_{T_i} the associated magnitude.

Using the TNC optimisation algorithm with 100 different starting points, we obtain the following optimal point:

$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.53762, 0.001, 0.868286) \quad (27)$$

The associated error is:

$$Err^* = 0.6837$$

Figure A1.15 draws the T-return level obtained from the optimal GEV distribution and the sample S_b.

Table A1.17 details the T-return level obtained from the optimal GEV for several periods T and gives the relative error made for each of them (in %) defined by (17). We note that even if the global error is larger, the optimal GEV distribution fits better for periods larger than 500 years.

Figure A1.15. Case 2b: Frequency - Magnitude relation obtained with the optimal GEV model (in blue) that best predicts the given synthetic data (in black) with respect to the weighted L2-error: in natural scale (left) and in logscale (right).

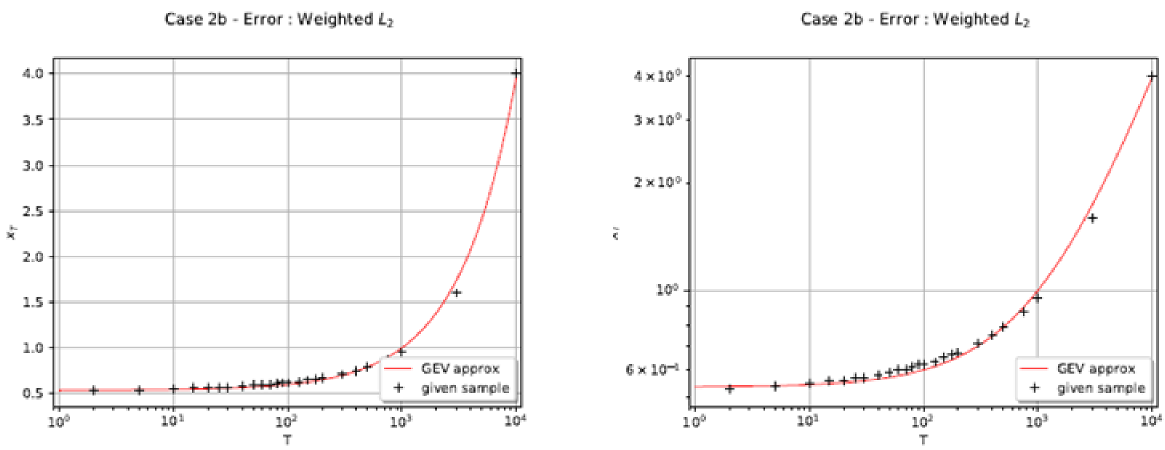


Table A1.17. Case 2b: Comparison between the given synthetic data and the optimal GEV model using the weighted L2-error, with the relative error (%).

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Ref Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m)	0.538	0.541	0.545	0.571	0.599	0.790	1.00	1.74	3.96
e(*T*)	1.52	1.30e-1	9.83e-1	329	3.39	2.99e-2	5.25	8.74	1.00
Return Period (years)	15	20	25	30	40	60	70	80	90
Ref Mag (m)	0.56	0.56	0.57	0.57	0.58	0.60	0.60	0.61	0.62
Approx Mag (m)	0.548	0.552	0.555	0.558	0.564	0.576	0.582	0.588	0.593
e(*T*)	2.11	1.49	2.63	2.07	2.67	3.92	2.96	3.62	4.27
Return Period (years)	125	150	175	200	300	400	750		
Ref Mag (m)	0.63	0.65	0.66	0.67	0.71	0.75	0.87		
Approx Mag (m)	0.612	0.626	0.638	0.651	0.699	0.745	0.897		
e(*T*)	2.79	3.77	3.29	2.86	1.52	6.0e-2	3.15		

A1.3.2.3. L∞-error

We define the error criteria defined in (20).

$$Err(S_b, q_\theta) = \max |x_{T_i} - q_\theta(T_i)| \qquad (28)$$

where $S_b = (T_i, x_{T_i})_{1 \le i \le 26}$ are the synthetic data of Table A1.6 and Table A1.7 with T_i the return period and x_{T_i} the associated magnitude.

Using the TNC optimisation algorithm with the optimal point obtained with the L_2-error as starting point, we obtain the following optimal point:

$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.478596, 0.00100039, 0.868782) \qquad (29)$$

The associated error is:

$$Err^* = 0.0854126$$

Figure A1.16 draws the T-return level obtained from the optimal GEV distribution and the sample Sb.

Table A1.18 details the T-return level obtained from the optimal GEV for several periods T and gives the relative error made for each of them (in %) defined by (17).

Figure A1.16. Case 2b: Frequency - Magnitude relation obtained with the optimal GEV model (in blue) that best predicts the given synthetic data (in black) with respect to the L∞-error: in natural scale (left) and in logscale (right).

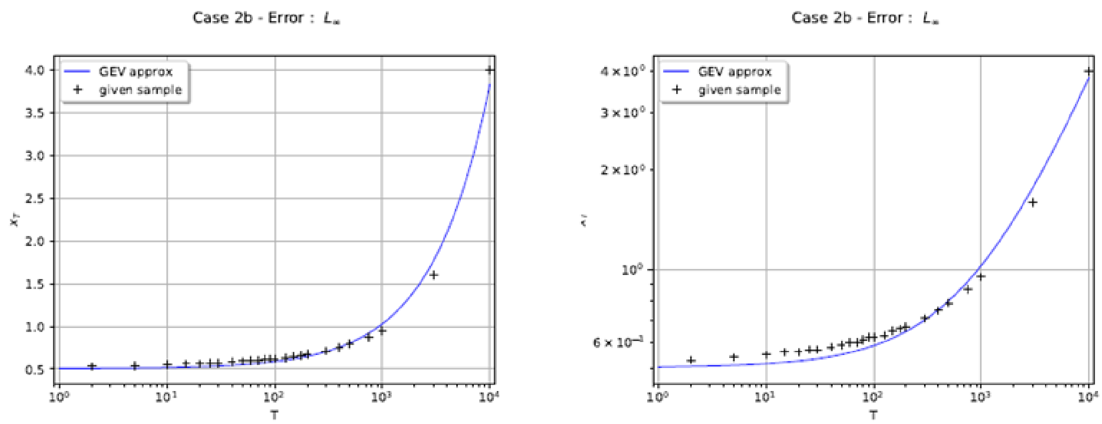


Table A1.18. Case 2b: Comparison between the given synthetic data and the optimal GEV model using the L∞-error, with the relative error (%).

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Ref Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m)	0.479	0.482	0.486	0.512	0.540	0.732	0.942	1.685	3.916
e(*T*)	9.62	10.8	11.71	13.3	12.9	7.35	7.99e-1	5.34	2.10
Return Period (years)	15	20	25	30	40	60	70	80	90
Ref Mag (m)	0.56	0.56	0.57	0.57	0.58	0.60	0.60	0.61	0.62
Approx Mag (m)	0.489	0.493	0.496	0.499	0.506	0.518	0.523	0.529	0.535
e(*T*)	12.6	12.0	13.0	12.4	12.8	13.72	12.8	13.3	13.8
Return Period (years)	125	150	175	200	300	400	750		
Ref Mag (m)	0.63	0.65	0.66	0.67	0.71	0.75	0.87		
Approx Mag (m)	0.554	0.567	0.580	0.592	0.641	0.687	0.840		
e(*T*)	12.1	12.8	12.2	11.6	9.77	8.39	3.50		

A1.3.2.4. Weighted L_∞-error

In order to force a best adequation over the large return periods T, we consider the weighted least square error defined by :

$$Err(S_b, q_\theta) = \max |\log(T_i)[x_{T_i} - q_\theta(T_i)]| \qquad (30)$$

where $S_b = (T_i, x_{T_i})_{1\leq i\leq 26}$ are the synthetic data of Table A1.6 and Table A1.7, with T_i the return period and x_{T_i} the associated magnitude.

Using the TNC optimisation algorithm with the optimal point obtained with the L_2-error as starting point, we got the following optimal point:

$$\theta^* = (\mu^*, \sigma^*, \xi^*) = (00.447999, 0.001, 0.870727) \qquad (31)$$

The associated error is:

$$Err^* = 0.570137$$

Figure A1.17 draws the T-return level obtained from the optimal GEV distribution and the sample S_b.

Table A1.19 details the T-return level obtained from the optimal GEV for several periods T and gives the relative error made for each of them (in %) defined by (17). We note that even if the global error is larger, the optimal GEV distribution fits better for periods larger than 1 000 years.

Figure A1.17. Case 2b: Frequency - Magnitude relation obtained with the optimal GEV model (in blue) that best predicts the given synthetic data (in black) with respect to the weighted L∞-error: in natural scale (left) and in logscale (right).

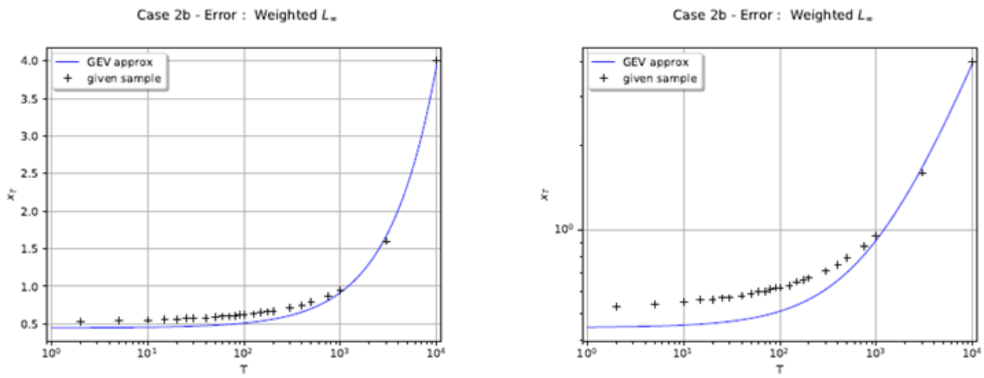


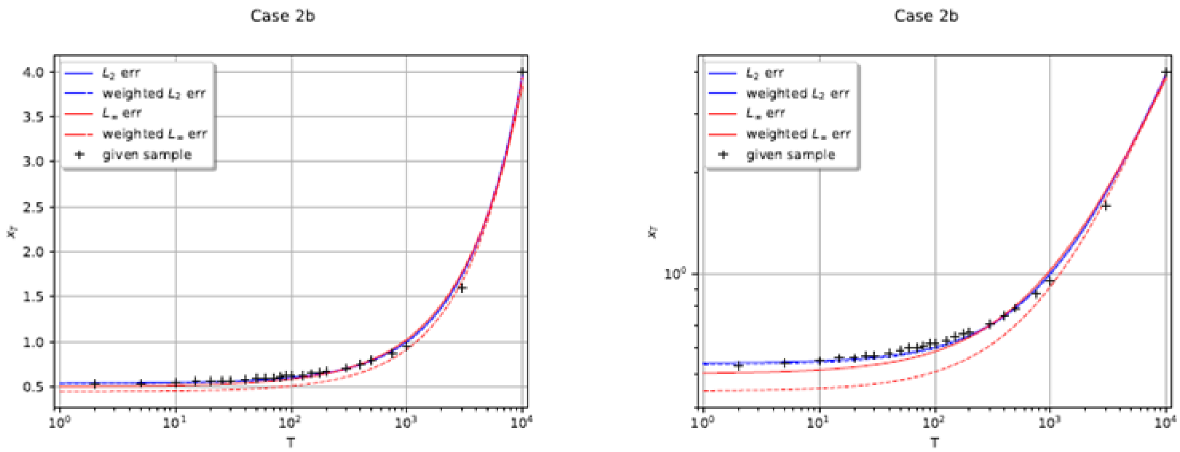
Table A1.19. Case 2b: Comparison between the given synthetic data and the optimal GEV model using the weighted L∞-error, with the relative error (%).

Return Period (years)	2	5	10	50	100	500	1 000	3 000	1 0000
Ref Mag (m)	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m)	0.448	0.451	0.455	0.481	0.510	0.704	0.917	1.67	3.94
e(*T*)	15.4	16.5	17.3	18.4	17.8	10.9	3.49	4.41	1.54
Return Period (years)	15	20	25	30	40	60	70	80	90
Ref Mag (m)	0.56	0.56	0.57	0.57	0.58	0.60	0.60	0.61	0.62
Approx Mag (m)	0.459	0.462	0.465	0.469	0.475	0.487	0.493	0.499	0.504
e(*T*)	18.1	17.5	18.3	17.8	18.1	18.8	17.8	18.2	18.6
Return Period (years)	125	150	175	200	300	400	750		
Ref Mag (m)	0.63	0.65	0.66	0.67	0.71	0.75	0.87		
Approx Mag (m)	0.523	0.537	0.550	0.562	0.611	0.658	0.813		
e(*T*)	16.9	17.4	16.7	16.1	13.9	12.2	6.59		

A1.3.2.5. Models comparison

In order to ease comparison, Figure A1.18 draws both models obtained from all the criteria.

Figure A1.18. Frequency - Magnitude relation obtained with the optimal GEV model (in blue and red) that best predicts the given synthetic data (in black): in natural scale (left) and in logscale (right).



A1.3.3. Case 2c

In Case 2c, we use the synthetic data of Table A1.8.

To fit a GEV distribution on the synthetic data, we define the error function:

$$Err^2(S_c, q_{\boldsymbol{\theta}}) = \sum_{i=1}^{6}(x_{T_i} - q_{\boldsymbol{\theta}}(T_i))^2 + \sum_{i=1}^{4}(X_{T_i} - q_{\boldsymbol{\theta}}(T_i))^2 \qquad (32)$$

where $S_c = (T_i, x_{T_i})_{1\le i\le 6}$ are the known synthetic data and $(X_{T_i})_{1\le i\le 4}$ are the 4 unknown return levels, only described by their distribution.

Then, the error function $Err^2(S_c, q_{\boldsymbol{\theta}})$ is a random variable because of the randomness of the return levels associated to the return periods 500, 1 000, 3 000 and 10 000 years.

We will consider several probabilistic features of the random variable $Err^2(S_c, q_{\boldsymbol{\theta}})$ to find the optimal $\boldsymbol{\theta}$.

A1.3.3.1. Mean feature

In this section, we consider the mean of the random variable $Err^2(S_c, q_{\boldsymbol{\theta}})$ and we find the optimal $\boldsymbol{\theta}$ that minimises the mean error:

$$\boldsymbol{\theta}^* = \operatorname{argmin} \mathbb{E}\ Err^2(S_c, q_{\boldsymbol{\theta}}) \qquad (33)$$

Using the TNC optimisation algorithm with the optimal point obtained in the Case 2b with the L_2-error as a starting point, we obtain the following optimal point:

$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.524142, 0.00104785, 0.862527) \qquad (34)$$

The associated error is: $Err^* = 0.6969$

Figure A1.19 draws the T-return level obtained with the optimal GEV distribution.

Table A1.20 details the T-return level obtained from the optimal GEV for all the periods T. When the return level was unknown, we give to which quantile of the uncertainty distribution the model value corresponds.

Figure A1.19. Case 2c: Frequency - Magnitude relation obtained with the optimal GEV model (in blue) that best predicts the given synthetic data (in black) minimising the random mean squared error E Err2(Sc, qθ): in natural scale (left) and in logscale (right).

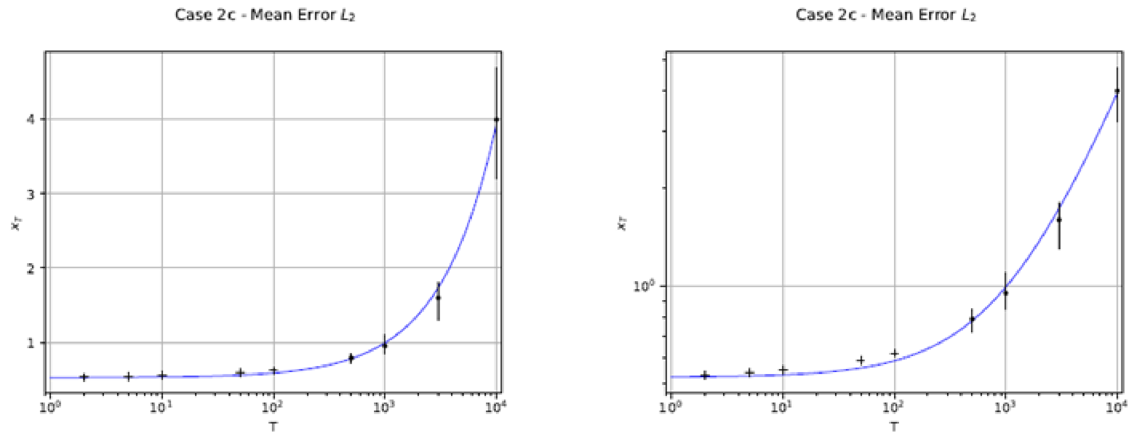


Table A1.20. Case 2c: Comparison between the given synthetic data and the optimal GEV model minimising the random mean squared error E Err2(Sc, qθ), with the relative error (%). For the uncertain values, we give the quantile of the estimated return level.

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Mean	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m)	0.525	0.527	0.531	0.558	0.587	0.781	0.993	1.73	3.94
Quantile	-	-	-	-	-	0.41	0.76	0.82	0.45
e(%)	1.02	2.34	3.38	5.41	5.30	-	-	-	-

A1.3.3.2. Quantile feature

In this section, we consider the quantile 95% of the random variable $Err^2(S_c, q_{\boldsymbol{\theta}})$ and we find the optimal $\boldsymbol{\theta}$ that minimises the quantile error:

$$\boldsymbol{\theta}^* = \text{argmin Quantile}_{0.95}(Err^2(S_c, q_{\boldsymbol{\theta}})) \qquad (35)$$

Using the truncated newton method optimisation algorithm with the optimal point obtained in the Case 2b with the L_2-error as a starting point, we obtain the following optimal point:

$$\boldsymbol{\theta}^* = (\mu^*, \sigma^*, \xi^*) = (0.516115, 0.00108486, 0.859784) \quad (36)$$

The associated error is:

$$Err^* = 1.04575$$

It means that $P[\ Err^2(S_c, q_{\boldsymbol{\theta}}) \geq Err^{*2}\ ] = 0.05$.

Figure A1.20 draws the T-return level obtained from the optimal GEV distribution.

Table A1.21 details the T –return level obtained from the optimal GEV for all the periods T.

When the return level is unknown, we give to which quantile of the uncertainty distribution the model value corresponds.

Figure A1.20. Case 2c: Frequency - Magnitude relation obtained with the optimal GEV model (in blue) that best predicts the given synthetic data (in black) minimising the Quantile$_{0.95}$ $Err^2(S_c, q_\theta)$: in natural scale (left) and in logscale (right).

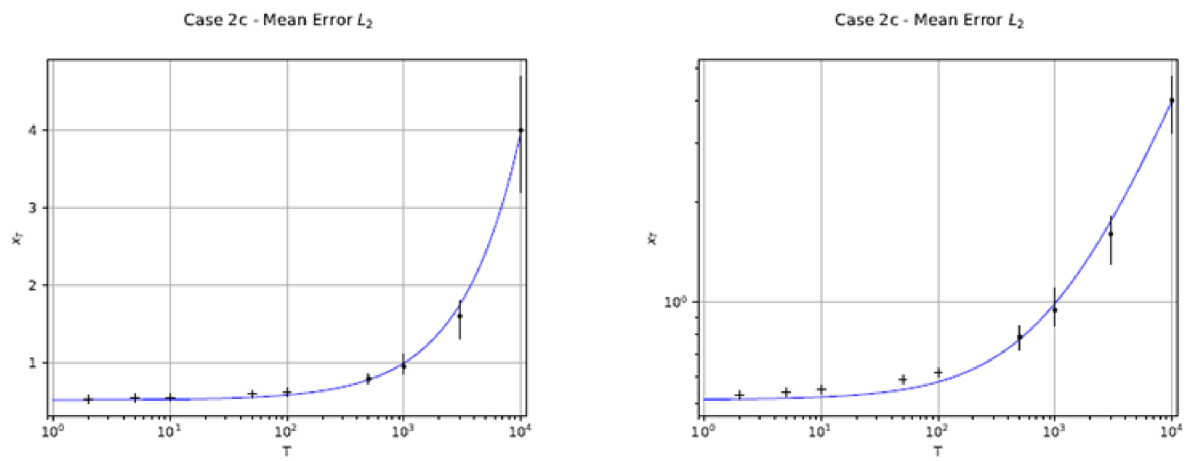


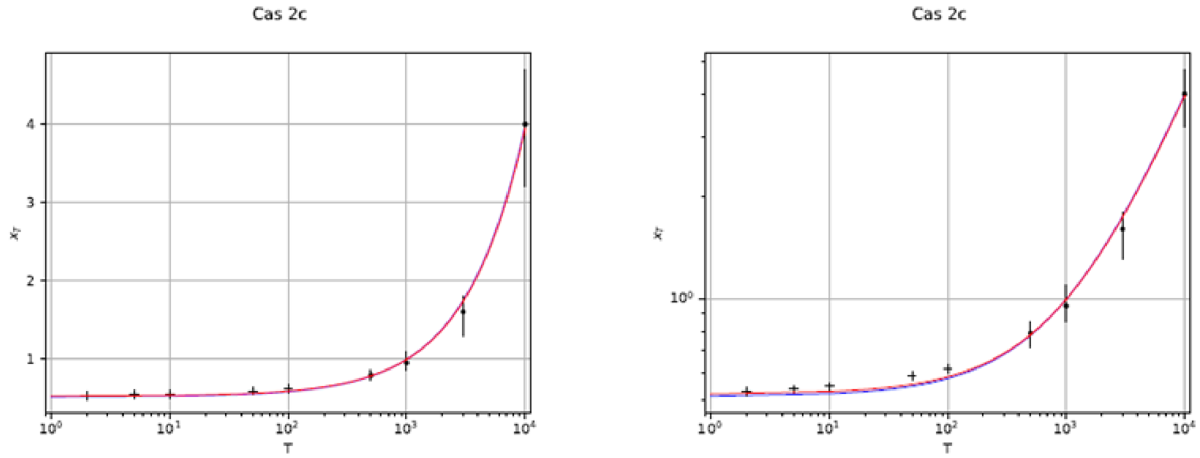
Table A1.21. Case 2c: Comparison between the given synthetic data and the optimal GEV model minimising the Quantile0.95 (Err2(Sc, qθ), with the relative error (%). For the uncertain values, we give the quantile of the estimated return level.

Return Period (years)	2	5	10	50	100	500	1 000	3 000	10 000
Mean	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
Approx Mag (m)	0.517	0.519	0.523	0.551	0.581	0.778	0.994	1.75	3.98
Quantile	-	-	-	-	-	0.39	0.77	0.84	0.48
e(%)	2.53	3.81	4.80	6.61	6.33	-	-	-	-

A1.3.3.3. Models comparison

In order to ease comparison, Figure A1.21 draws both models obtained when minimising the mean error and its quantile of order 95%. We note that both criteria lead to the same optimal solution.

Figure A1.21. Case 2c: Frequency - Magnitude relation obtained with the optimal GEV model (in blue and red) that best predicts the given synthetic data (in black) minimising the mean feature of Err2(Sc, qθ) (in blue) and its quantile 95% feature (in red): in natural scale (left) and in logscale (right).



A1.4. Conclusion

The benchmark study proposed an exercise on the statistical modelling. The main goal is to better understand the quantitative technical analysis steps and processes used for assessing hazard frequency and magnitude for external event risk assessments. To this end, we have the synthetic data for a hypothetical external event.

In this report, we modelled the hazard frequency-magnitude relation with a generalised extreme value distribution. Several criteria to fit the GEV distribution: L_2-error and L_∞-error. For the Case 1, a 3-points interpolation is performed.

For the synthetic data with no uncertainty, we saw that the criterion allowing the best fitting of GEV distribution is L_2-error. For the synthetic data with the uncertainty in some points, minimising the random mean squared error has the best precision.

In this benchmark, whatever the type of data, we managed to approximate the synthetic model/data by a GEV distribution.

Annex B. Submission by Finnish Meteorological Institute

OECD/NEA Benchmark on External Events Hazard Frequency and Magnitude Statistical Modelling

Olle Räty and Marko Laine
Finnish Meteorological Institute

20.6.2019

Introduction

This R notebook contributes to the OECD/NEA benchmark exercise given in the title. The objective of the exercise is stated as the following:

> The objective of this benchmark study is to facilitate an exercise on statistical modelling in order to better understand the quantitative technical analysis steps and processes used for assessing hazard frequency and magnitude for external events risk assessment. This benchmark study report provides details (data and overall objectives) for the benchmarking exercise by specifying synthetic data for a hypothetical external event (e.g., precipitation, extreme temperatures, high winds). The analysis steps, assumptions, insights, and modelling results of the benchmark participants will be collected and summarized to gain insights from the activity.

Synthetic data is provided with only little information on data itself, generating processes or uncertainties. The synthetic data approach is motivated by the fact that it is possible to evaluate the predictive performance of the suggested statistical model and control the data behavior as well as add known uncertainty to it. Two cases are provided: one with a given data generating model and a "blind-test case" where only the data is provided. As requested by the organizer of this exercise the following aspects are documented:

- Model assumptions
- Qualitative/quantitative modeling results
- Model adequacy assessment and the results from this assessment

To make the approach and road to suggested solutions as transparent as possible, all the necessary steps to repeat the exercise (excluding setting up some parts of the R environment) are provided in this notebook. The notebook first provides some theoretical and technical background for the selected approach and the assumptions. Then, results for fitting the selected statistical models are presented for each test case together with an analysis of potential uncertainties.

We first introduce the data provided in the exercise. Then, the underlying theory and crucial assumptions are presented together with the framework for the model adequacy assessment. Next, model- and case-specific results and model adequacy assessment are given. The notebook is concluded with a summary of the most essential results.

Data description

Case 1

In the first case, the functional form of the data generating model is known. The model for the magnitude M with respect to return period t is written as

$$M(t) = 0.5 + 0.5\log_{10}(t). \qquad (1)$$

Figure 1: Data in Case 1.

The data for the Case 1 is plotted in Fig 1. There are a total of 10 values for M corresponding to different return levels, the magnitude increasing from 0.5 m for a 1-year event to up to 2.5 m for an event happening on average once every 10 000 years. Given the unit of M, it could be interpreted as the magnitude of the water level in a water body or river, which gives some hints on the assumed precision and accuracy of the data. Note that here we have plotted x axis on a logarithmic scale to highlight the linear relationship between M and the logarithm of t. The implications of this relationship are discussed in the results section.

Case 2

Case 2 is a blind-test case, where only data has been provided and the data generating model is unknown. There are a total of three sub-tasks in Case 2.

```
case2a <- data.frame(
  x=c(1+1e-5,2,5,10,50,100,500,1000,3000,10000),
  y=c(0.53,0.53,0.54,0.55,0.59,0.62,0.79,0.95,1.6,4.0)
  )
```

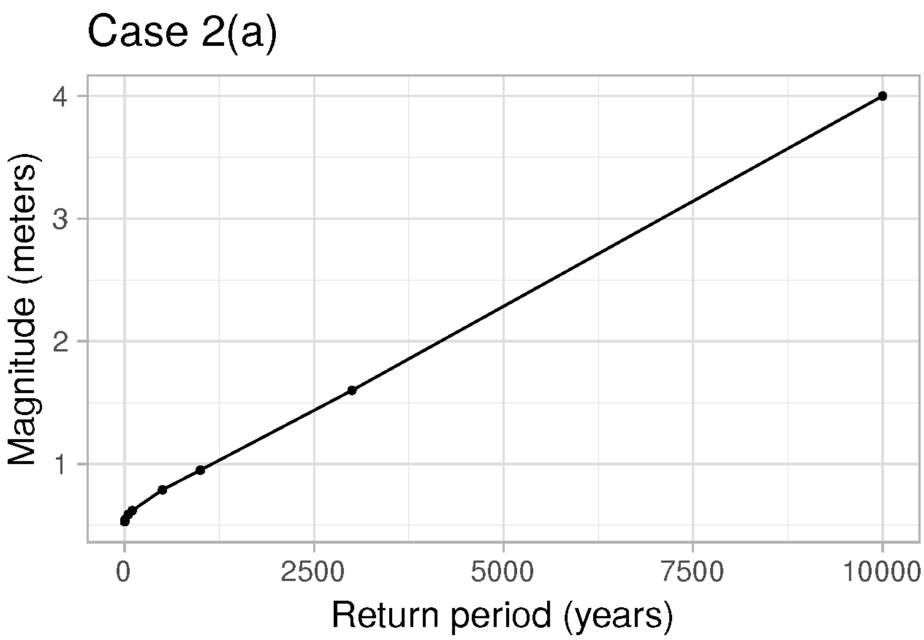


Figure 2: Data in Case 2(a).

In Case 2(a), M exceeds 4 m on average once every 10 000 years (Fig. 2). The main difference to Case 2(b) and 2(c) is the smaller number of data points available for M. Figure 2 shows that the relationship between M and t is quite linear apart from the smallest values for t. Again, the implications of this behavior are discussed in the results section.

As stated above, Case 2(b) differs from 2(a) by having more data points available (Fig. 3). The values of M are the same for short time intervals, while for some reason the two largest values of M differ slightly between between Case 2(a) and 2(b). However, we still interpret that the generating process is the same in both cases and that the aim is just to test how much the increased amount of data affects model fitting.

In Case 2(c), which is illustrated in Fig. 4, the exact values of M are partly unknown and some information on uncertainty is given for the rarest events. The mean estimate is given as black dots, standard deviation marked with red and the range between the

3

Figure 3: Data in Case 2(b).

5[th] and 95[th] percentile with brown dots in the figure. The uncertainty distribution is not exactly normal for any of the return levels, as there is a slightly longer tail towards smaller values for $t \in [500, 3\ 000, 10\ 000]$ and towards larger values when $t = 1\ 000$ years. Based on this, we interpret that the aim is to use this information to obtain a better estimate of modeling uncertainty. For the analysis, where we need to have some uncertainty estimate for all observed quantiles, we have linearly extrapolated the standard deviation to all values of t (not shown).

Assumptions and model adequacy assessment approach

There are plethora of statistical models which could be considered in this exercise. However, as the aim is to model exceedingly rare hazardous events, we restrict the analysis to models arising from classical univariate extreme value theory. We will assume that the underlying data in both Case 1 and 2 describe the annual maxima of a set of independent and identically distributed random variables $(X_1, X_2, ..., X_n)$, where n is the number of observations for one year. The classical extreme value theory states that as $n \to \infty$, generalized extreme value distribution (GEV) is the only possible limiting distribution for the annual maxima. Following the notation of Coles [2001] we write the GEV distribution $G(z; \mu, \sigma, \xi)$ as

Figure 4: Data in Case 2(c).

$$G(z;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \quad (2)$$

where μ, σ and ξ denote the location, scale, and shape parameter with $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$, respectively. GEV is a generalization of three special cases, which are defined by the value of the shape parameter ξ: Fréchet (type II, $\xi < 0$), Weibull (type III, $\xi > 0$) and Gumbel (type I, $\xi \to 0$) distribution.

In all tasks, the data essentially describes the relationship between the specific return levels of M and the corresponding return periods t. Thus, it is convenient to define the model in terms of the quantile function for which GEV has a closed-form expression:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{for } \xi \neq 0 \\ \mu - \sigma\log\{-\log(1-p)\}, & \text{for } \xi = 0 \end{cases} \quad (3)$$

Here $G(z_p) = 1 - p$ so that $t = 1/p$. We use (3) extensively when inferring the optimal model parameters and the corresponding modeling uncertainties in the test cases.

We take a Bayesian approach to the estimation of the GEV parameters. The well-known Bayes' theorem states that the posterior distribution of some model parameters θ given observations y_{obs} can be expressed as $\pi(\theta|y_{obs}) \propto L(\theta|y_{obs})p(\theta)$, where $L(\theta|y_{obs})$ is the likelihood function of θ and $p(\theta)$ is the prior parameter distribution. One of the main strengths in the Bayesian approach is that it provides a natural way to estimate both parameter and observational uncertainties. In this exercise, the underlying likelihood function $L(\theta|y_{obs})$ is unknown, as the observations from which the return levels have been estimated are not available. However, the likelihood function can be assumed to be of a certain well-known form such as Gaussian. The sample quantiles are approximately Gaussian and it would be possible to formulate their exact distribution too, if we had better information on processes that generate the given magnitude values. The discussion in the following sections suggests that Gaussian likelihood function can provide plausible results for the model fitting. We implemented two approaches for Bayesian inference in this notebook: (i) approximate Bayesian computation (ABC) which assumes that we can generate new observations and (ii) MCMC simulation based approach assuming Gaussian likelihood and observation weights that depend on some assumed uncertainty properties and GEV asymptotics.

ABC is a non-parametric approach for parameter inference in cases where the likelihood function is intractable. The main idea in ABC is to (i) sample model parameters from a prior distribution, (ii) simulate the corresponding observations using the sampled parameters, (iii) calculate informative summary statistics from the simulated observations and (iv) then compare the simulated summary statistics against the actual ones with a certain distance metric. To be more specific, a parameter sample θ is drawn from prior parameter distribution $p(\theta)$, which (for example) can assumed to be uniform if no prior knowledge is provided. Using the sampled parameters θ we simulate quantiles z_p from the GEV quantile function (3). They are here used directly as the summary statistics. Euclidean distance d is then calculated between the observed and simulated quantiles. If d is smaller than some pre-specified tolerance ϵ, the parameter sample is stored and sampling repeated until a sufficient number of samples has been obtained.

Parameter sampling can be conducted several different ways, with algorithms ranging from simple rejection sampling to more complex Markov chain Monte Carlo (MCMC) algorithms. While all of these approaches have their specific merits and shortcomings, MCMC is used in this exercise, partly because of personal preference and partly due to more efficient estimation of the approximate posterior distribution. In addition to ABC-based MCMC sampling procedure (ABC-MCMC), we evaluate the parameter uncertainty also directly from the posterior distribution using traditional MCMC. The obtained parameter samples are used to estimate the overall predictive uncertainty of the simulated M.

The details of the ABC-MCMC algorithm can be found from Marjoram et al. [2003]. There are several tools for R (e.g. `EasyABC` and `abctools`) and other programming languages for implementing ABC sampling [Nunes and Prangle, 2015]. However, a set of simple functions was written for this specific exercise which combine a rejection sampling step [Wegmann et al., 2009] with MCMC as in `EasyABC` [Jabot et al., 2013] and which are available together with this notebook.

For the direct likelihood based approach the MCMC sampling was conducted using `FME` package available in CRAN. This package relies on earlier MATLAB implementation of these algorithms. More detailed description of the algorithms implemented in this package can be found from Haario et al. [2006] and Haario et al. [2001].

Case 1

Model details results

Equation (1) shows that there is a linear relationship between M and $-\log(1/t)$ in the generating model. This suggests that the Gumbel distribution is a suitable distribution candidate in Case 1. Gumbel distribution is obtained from (1) by setting $\xi = 0$:

$$G(z;\mu,\sigma) = \exp(-\exp(-\frac{z-\mu}{\sigma})) \tag{4}$$

As mentioned above, to model return levels of M we use the GEV quantile function (3):

$$z_p = \mu - \sigma\log(-\log(1-p)) = \mu - \sigma\log(-\log(1-\frac{1}{t})) \tag{5}$$

Comparing (1) and (5) we see that with suitable values for μ and σ the generating function and the Gumbel quantiles approach asymptotically each other as $t \to \infty$. Thus, setting $\mu = 0.5$ and $\sigma = \frac{0.5}{\log(10)}$ we will get an increasingly good match with the observed M when t increases.

The behavior of this asymptotically optimal model is illustrated in Fig. 5. For $t > 10$ years the fit is extremely good, while for smaller values the asymptotic model unrealistically tends to zero. The underestimation of magnitude M when t is small can be understood by the fact that the extreme value theory considers asymptotic behavior; GEV is a limiting distribution of sample maxima, when the number of observations tends to infinity and therefore is not expected to describe commonly occurring values of M well.

Results of the assessment

Even though the asymptotically optimal GEV model can be inferred directly in Case 1, it is also useful to gain some insights on the predictive uncertainty. This would be of course of interest, were the optimal parameter values unknown. We first use ABC-MCMC for inferring the parameter uncertainty. Uniform prior distributions have been used for all GEV distribution parameters with the restrictions $\mu > 0$ and $\sigma > 0$. While $\xi = 0$ in the asymptotically optimal case, we allow its values to vary slightly around the zero to better emulate the potential parameter uncertainty. The used prior distributions are then

$$\begin{cases} \mu \sim U(0.3, 0.7) \\ \sigma \sim U(0, 0.0.5) \\ \xi \sim U(-0.2, 0.2) \end{cases}$$

The number of samples drawn from the approximate posterior is $n = 10\,000$. We leave out 1-, 2- and 5-year return levels when fitting the model, as these common cases are not of interest and hamper the model fitting. In ABC-MCMC, the maximum allowed distance between the simulated and observed quantiles as well as the proposal range at each sampling step are estimated in advance using naive rejection sampling. A code

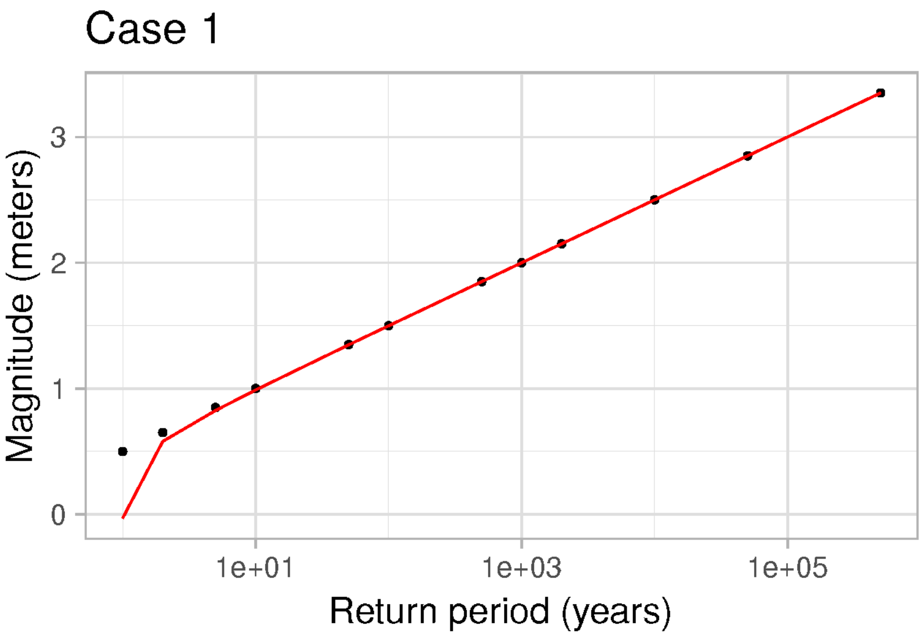


Figure 5: Asymptotic model (red curve) plotted together with the synthetic data in Case 1.

snippet illustrating the parameter settings and most important function calls are provided for Case 1.

```
drop <- c(1,2,3) #Select, which values are dropped
set.seed(100)
pars <- matrix(c(0.3,0,-0.2,
                 0.7,0.5,0.2), nrow = 3)

rejection_sample <- 100000 #Rejection sample
n_sample <- 10000 #MCMC sample
tol <- 0.001 #Tolerance for acceptance
```

```
weights  <- NULL # No weighting
initialisation <- pre_calc(n_sample = rejection_sample, pars = pars, tol = tol,
                           ret = case1$x[-drop], target = case1$y[-drop],
                           weights = weights, calc.weights = F)
posterior_1 <- run_MCMC_ABC(init_vals = initialisation$init_vals, pars = pars,
                            iterations = n_sample, max_dist = initialisation$max_dist
                            ret = case1$x[-drop], target = case1$y[-drop],
                            s_weights = weights[-drop], phi = 1,
                            proposal_range = initialisation$proposal_range)
```

Figure 6 shows the parameter sampling results for ABC-MCMC in Case 1. The optimal parameter values $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ shown in the lower left corner are estimated as posterior means of the marginal distributions. These values provide a very good fit to the observed data apart from the smallest values which were also excluded from the analysis. The posterior distributions of the model parameters show that although there are not much prior information provided for ABC-MCMC, the algorithm is able to identify all parameters well. The bi-variate parameter distributions shown in the middle row illustrate that parameter pairs are correlated with each other and that the correlation is strongest between the scale (σ) and shape (ξ) parameter. The plot also illustrates an important property of the ABC-MCMC sampler: As the acceptance tolerance depends solely on the distance between the observed and modeled summary statistics, bi-variate distributions have a clear boundary defined by the tolerance. This is an important difference to traditional MCMC.

To assess the predictive uncertainty in the simulated magnitude, Fig. 7 shows the *predictive distribution* obtained from Eq. (3) using the sampled GEV parameter values for return periods up to 500 000 years (see also Table 1). When looking at the mean prediction (blue curve), it is seen that it matches the asymptotically optimal values (red curve) almost exactly apart from the shortest return periods. Furthermore, the uncertainty range in the posterior predictions is narrow even for the largest values of t. We conclude that using GEV provides a very good fit to the synthetic data in Case 1 and has a good predictive skill as well as low predictive uncertainty for the long return periods.

Traditional MCMC

As the second approach to parameter inference, we use traditional MCMC sampling assuming Gaussian likelihood. Prior distributions for μ, σ and ξ are assumed to be also

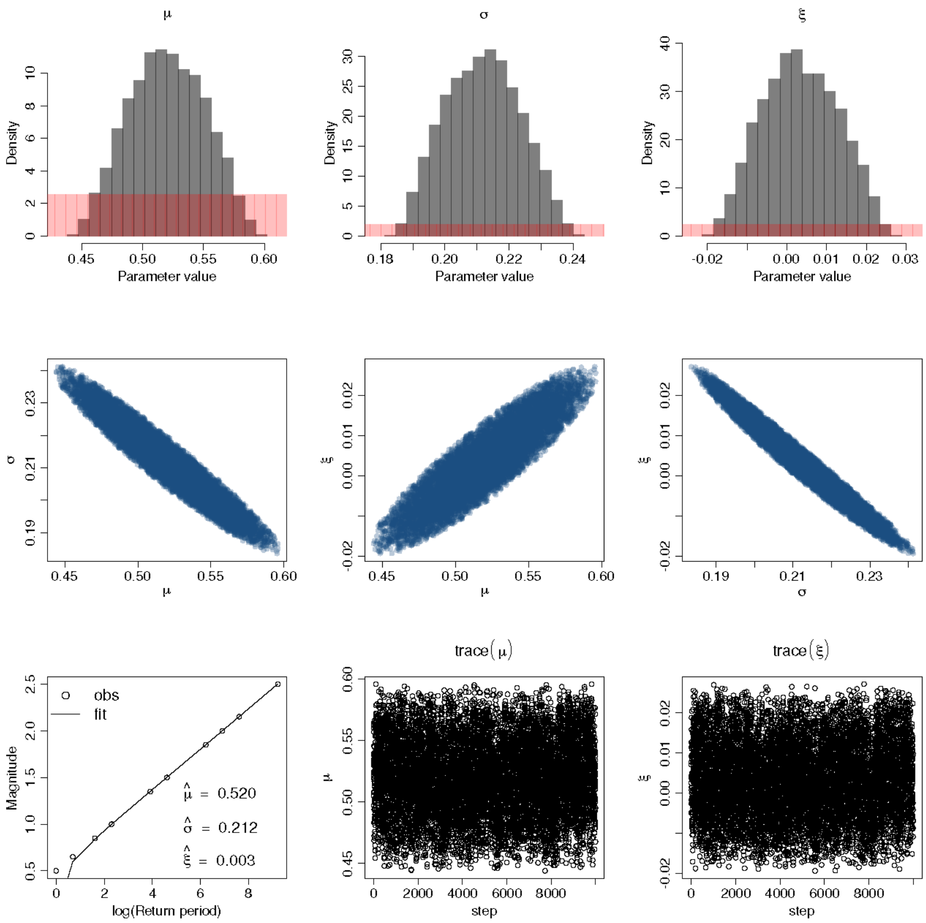


Figure 6: Panel showing different aspects of the posterior parameter distributions obtained with ABC-MCMC in Case 1. The first row shows the posterior marginals for location (left), scale (middle) and shape (right) parameter together with the prior distributions (red shading). The middle row shows the bi-variate posterior densities for each of the parameter pairs. Bottom row illustrates the distribution mean fit (left) with respect to the synthetic observations and also shows the sampling traces for the location (middle) and shape parameter (right), respectively.

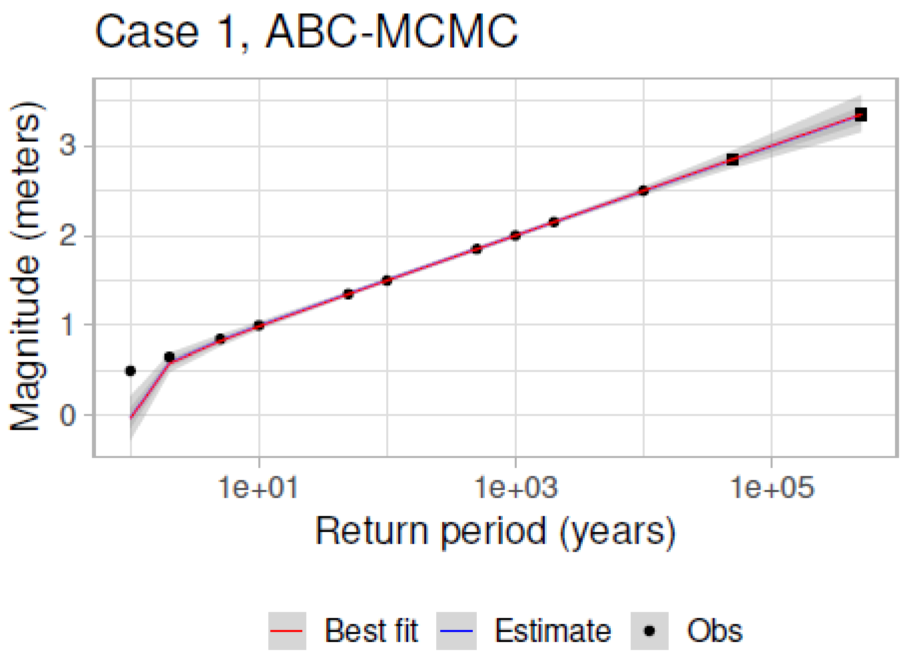


Figure 7: Predictive distribution for magnitude M obtained with ABC-MCMC. The mean prediction (blue) and the asymptotically best fit (red) are shown. The interquartile range and the range between the 5th and 95th percentile of the predictive distribution is indicated by the grey shadings. Observed synthetic values are also shown with two synthetic values to be predicted marked with squares.

Table 1: The estimated mean return level shown together with the 5th and 95th percentile of the predictive distribution for return periods between 500-500 000 years, obtained with ABC-MCMC in Case 1.

	500	5000	50000	500000
Mean	1.85	2.35	2.84	3.34
5%	1.82	2.31	2.75	3.15
95%	1.88	2.38	2.95	3.57

Gaussian with

$$\begin{cases} \mu \sim N(0.4, 10^2), \\ \sigma \sim N(0.5/\log(10), 10^2), \\ \xi \sim N(0, 2^2). \end{cases}$$

The prior means are close to the optimal values, while the spread is large, which leads to rather uninformative priors. One of the major differences with respect to ABC-MCMC is the inclusion of weights for different values of M. We use the following weight function

$$w_i = 1/(1 - \exp(-t_i/10))\sigma_i, \quad i = 1, \ldots, N,$$

where σ_i is the (potentially unknown) spread for the i^{th} observed magnitude and N is the number of synthetic observations. This weighting aims to reduce the effect of small values of M to the parameter estimation and simultaneously takes the potential increase in the observation uncertainty into account when M increases. Doing this there is no need to remove small observed values of M when sampling from the posterior distribution. We provide a code snippets also for MCMC, which illustrate how the parameter priors are defined and how to use the most important function calls.

```
case1$w <- 1/(1-exp(-case1$x/10))*0.05

paramdef <- list(
 parnames = c('μ', 'σ',  'ξ'),
 par0     = c(0.4,  0.5/log(10), 0.0),
 par.low  = c(0,    0,         -Inf),
 pri.mu   = c(0.4,  0.5/log(10), 0.0),
 pri.sig  = c(10,   10,   2)
)

mcmcopts <- list(
 nsimu = 12000,
 burn  =  2000,
 qcov  =  diag(pmax(0.01,(paramdef$par0*0.01)^2))
)

chain_1 <- mcmcrun(list(modelfun=modelfun),case1,paramdef,mcmcopts)
```

```
## number of accepted runs: 3122 out of 12000 (26.01667%)
```

The posterior distribution obtained with MCMC is shown in Fig. 8. As one would expect, the marginal aspects as well as the optimal parameter values are similar to ABC-MCMC.

The predictive distribution and the requested return levels are shown in Fig. 9 and Table 2, respectively. The predictive uncertainty is larger for MCMC than for ABC-MCMC, which is probably caused by the addition of (somewhat hypothetical) weights encompassing observational uncertainty.

12

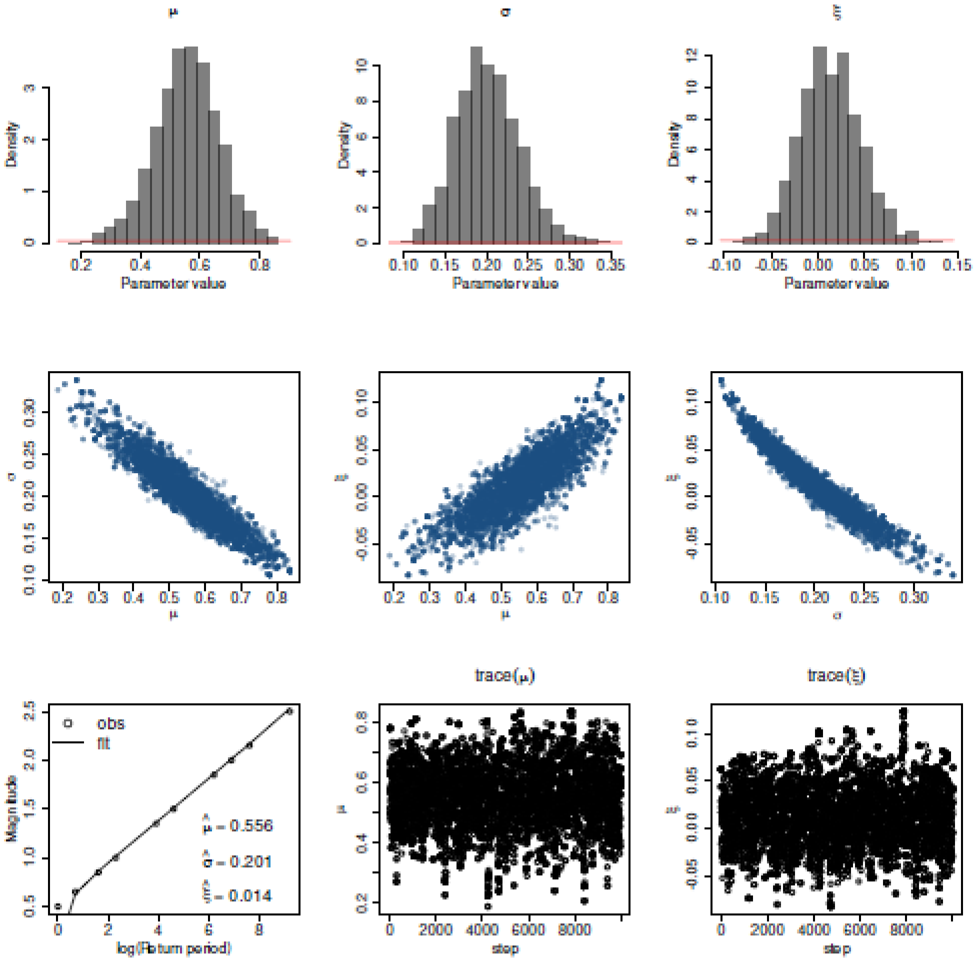


Figure 8: Same as in Fig. 6 but for MCMC in Case 1.

Table 2: The estimated mean return level shown together with the 5th and 95th percentile of the predictive distribution for return periods between 500-500 000 years, obtained with MCMC in Case 1.

	500	5 000	50 000	500 000
Mean	1.84	2.35	2.88	3.43
5%	1.80	2.29	2.73	3.13
95%	1.88	2.41	3.03	3.78

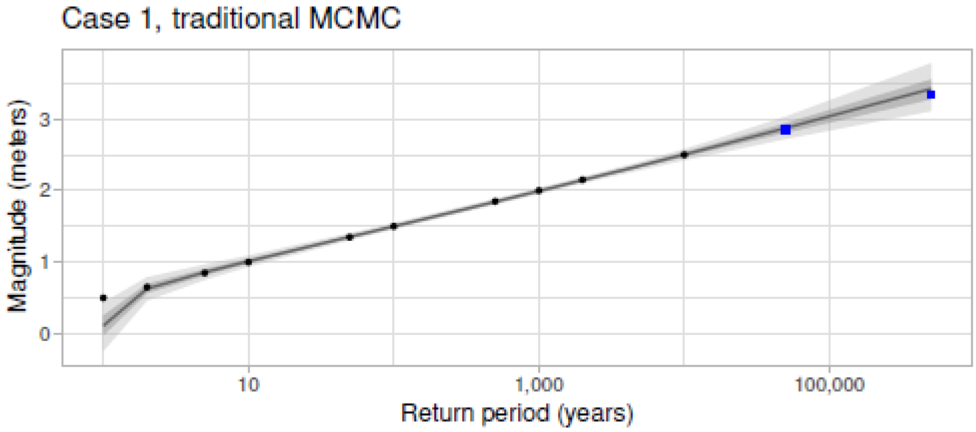


Figure 9: Same as in Fig. 7 but for MCMC in Case 1 with the grey line showing the mean prediction.

Case 2

In Case 2, the underlying generating model is unknown, which means that the direct evaluation of the predictive performance of the chosen statistical model is impossible. We use the full form of GEV and the corresponding quantile function given by Equations (2) and (3). We compare both ABC-MCMC and MCMC algorithms and provide the requested results in all three cases.

Model details results

We first inspect the data to obtain some prior insights into it. The synthetic data for the magnitude M and return period t suggests a rather linear relationship between them, particularly as t increases. GEV distribution provides a similar behavior, when ξ is close to unity and $p \to 0$. This information can be facilitated when constructing prior distributions for the model parameters, as μ should be close to the intercept, σ is proportional to the slope of the fit and $\xi \sim 1$.

Case 2(a)

ABC-MCMC

We first present the results for ABC-MCMC. The selected uniform priors are

$$\begin{cases} \mu \sim U(0.3, 0.7), \\ \sigma \sim U(0, 0.8), \\ \xi \sim U(0.7, 1.2). \end{cases}$$

The same prior distributions are used in all the following sub-tasks. While the selection of suitable prior distributions requires some hand tuning, the results were not found to be

14

overly sensitive to the specific choice of prior values, as long they were within a reasonable range. The acceptance ratio in the rejection sampling had to be set to 0.0005%, which is a very small number and requires a sample size of $n_{init} = 1\ 000\ 000$. In all cases, values for $t < 10$ years have been dropped from parameter estimation to ensure a better model fit to large M.

```
drop <- c(1,2,3) #Select, which values are dropped
pars <- matrix(c(0.3,0,0.7, #lower boundary
                 0.7,0.8,1.2), nrow = 3) #upper boundary
rejection_sample <- 1000000 #Rejection sample
n_sample <- 10000 #MCMC sample
tol <- 0.000005
```

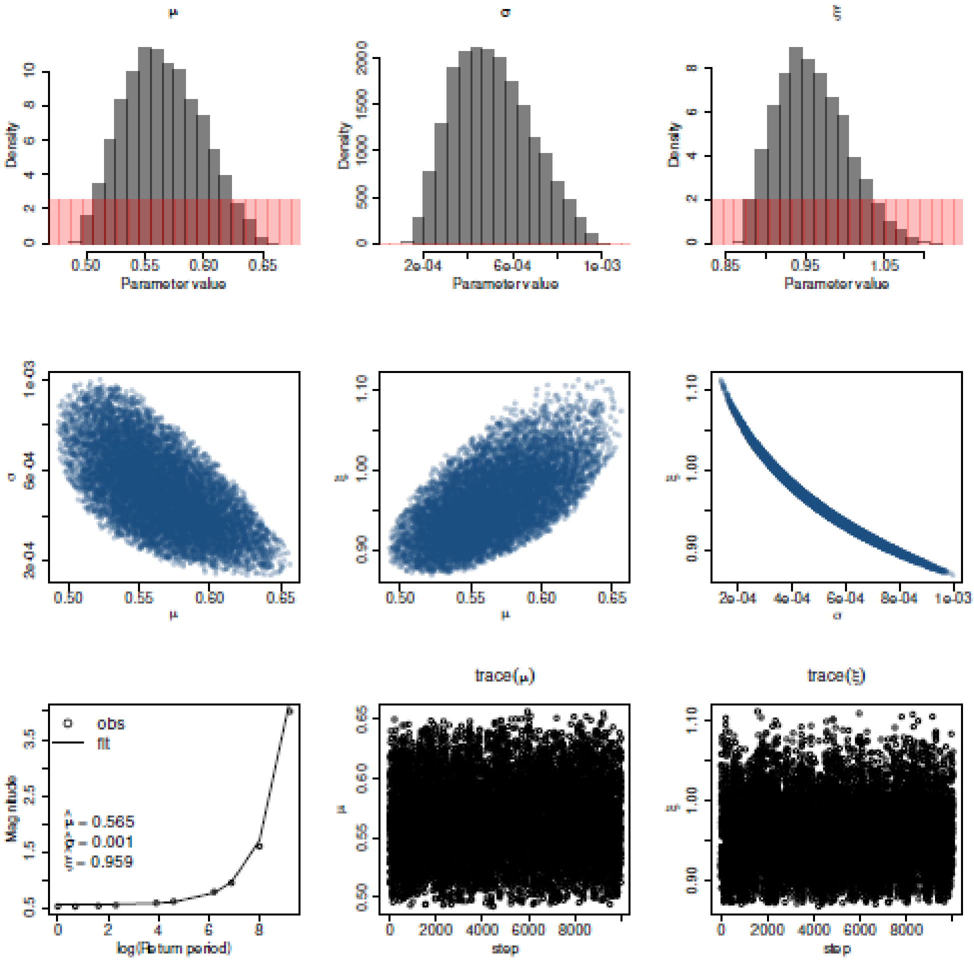


Figure 10: Same as in Fig. 6 but for ABC-MCMC in Case 2(a).

The optimal values for the GEV distribution parameters are again given in the lower left

15

Table 3: The estimated mean return level shown together with the 5th and 95th percentile of the predictive distribution for return periods between 500-500 000 years obtained with ABC-MCMC in Case 2(a).

	500	5000	50000	500000
Mean	0.76	2.33	16.62	148.36
5%	0.72	2.24	14.93	113.51
95%	0.80	2.40	18.77	196.99

corner of Fig. 10. The fit obtained with these values follows quite closely the observations. As can be seen from the figure, the uninformative priors do not provide significant amount of information but the ABC-MCMC is still able to identify posterior distribution relatively well. In addition, Bi-variate posterior marginals show a strong correlation between the scale and shape parameter.

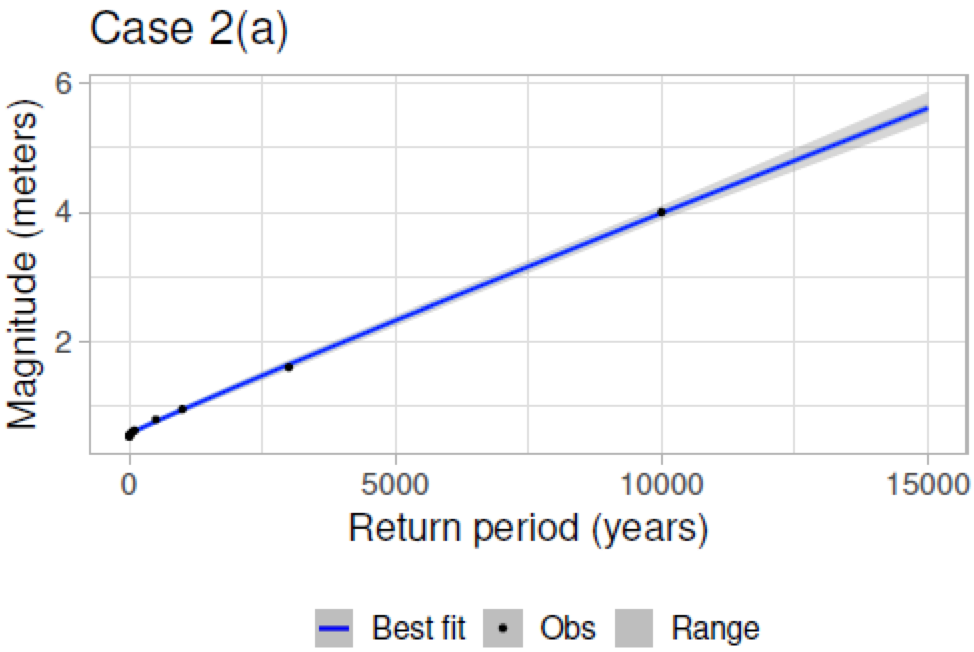


Figure 11: Same as in Fig. 7 but for ABC-MCMC in Case 2(a).

The predictive distribution obtained with ABC-MCMC is illustrated in Fig. 11. The requested return levels are provided in Table 3. These results show that the mean estimate grows substantially large when t increases. Whether the magnitude of these results is reasonable, is briefly speculated later.

Table 4: The estimated mean return level together with the 5th and 95th percentile of the predictive distribution for return periods between 500-500 000 years obtained with MCMC in Case 2(a).

	500	5 000	50 000	500 000
Mean	0.76	2.31	17.25	162.27
5%	0.72	2.25	15.74	128.58
95%	0.80	2.37	18.70	197.36

Traditional MCMC

We then compute an alternative parameter estimates using traditional MCMC. As in Case 1, we assume Gaussian likelihood and use Gaussian priors for the model parameter. The same priors are used throughout Case 2.

$$\begin{cases} \mu \sim N(0.5, 10^2), \\ \sigma \sim N(0.0003, 10^2), \\ \xi \sim N(1, 0.05^2). \end{cases}$$

The first 2000 samples are left out as burn-in values after which the number of samples drawn from the posterior is $n = 10\ 000$.

```
paramdef <- list(
 parnames = c('μ', 'σ',        'ξ'),
 par0     = c(0.64, 0.000281, 1.02),
 par.low  = c(0,    0,         0.1),
 pri.mu   = c(0.5,  0.0003,    1),
 pri.sig  = c(10,   10,        0.05)
)

mcmcopts <- list(
 nsimu = 12000,
 burn  =  2000,
 qcov  =  diag((paramdef$par0*0.01)^2)
)
```

```
case2a$w <- 1/(1-exp(-case2a$x/10))*0.05
chain_2 <- mcmcrun(list(modelfun=modelfun),case2a,paramdef,mcmcopts)
```

```
## number of accepted runs: 1577 out of 12000 (13.14167%)
```

Figures showing the different aspects of the posterior parameter distributions (Fig. 12) as well as the predictive distribution (Fig. 13) indicate only small differences in comparison to ABC-MCMC. The largest difference is the slightly narrower uncertainty range and the higher mean return level estimates for MCMC (Table 4).

Case 2(b)

The technical steps taken in Case 2(b) are practically identical to Case 2(a) and therefore are not repeated here. The only difference in comparison to Case 2(a) is the larger number

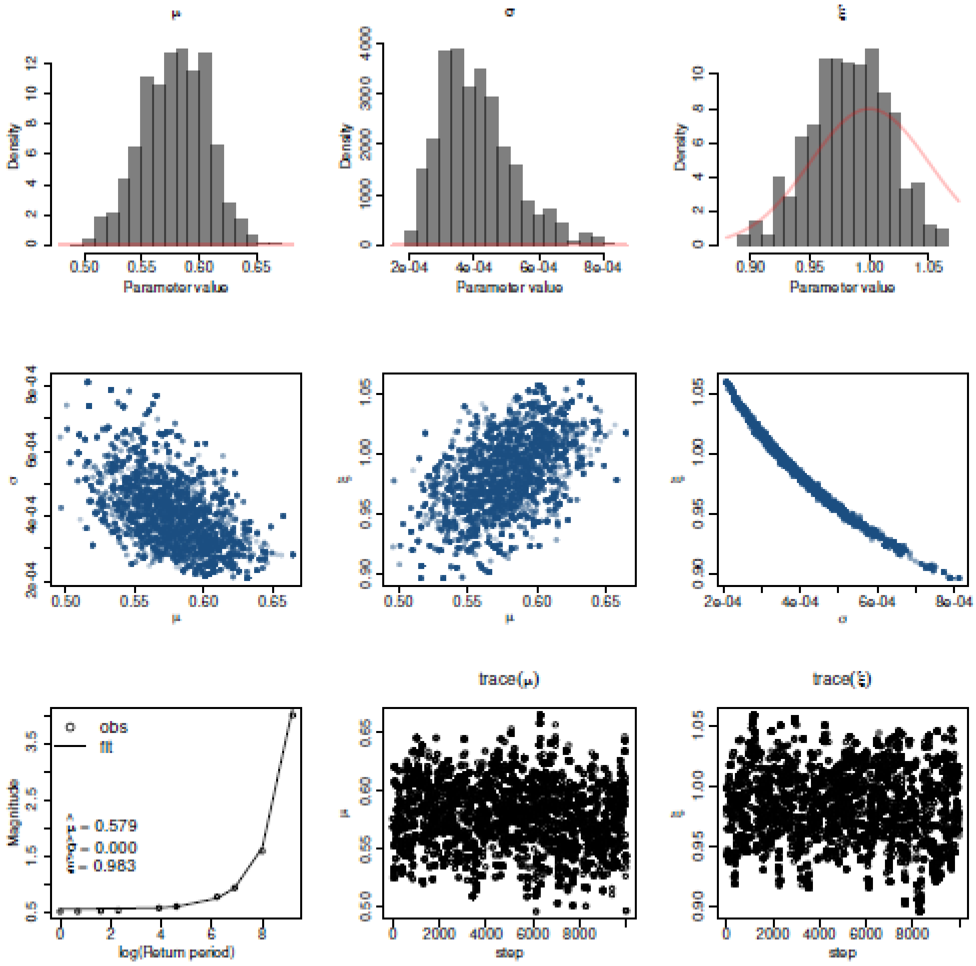


Figure 12: Same as in Fig. 8 but for MCMC in Case 2(a).

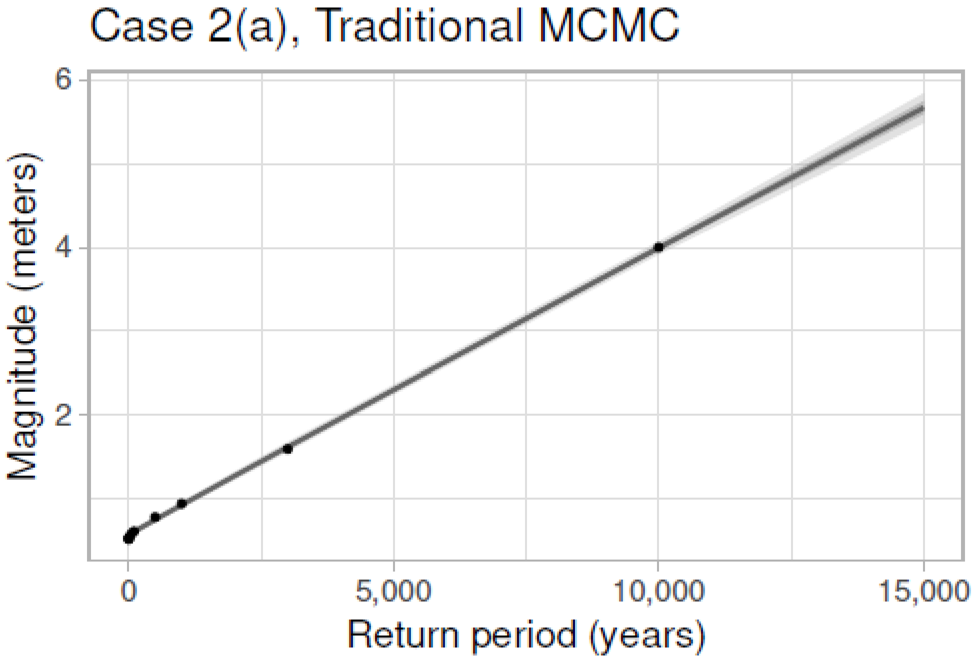


Figure 13: Same as in Fig. 9 but for MCMC Case 2(a).

Table 5: The estimated mean return level together with the 5th and 95th percentile of the predictive distribution for return periods between 500-500 000 years obtained with ABC-MCMC in Case 2(b).

	500	5000	50000	500000
Mean	0.77	2.32	16.22	142.11
5%	0.73	2.22	14.41	106.88
95%	0.80	2.41	18.55	192.92

of quantiles available for parameter estimation.

For ABC-MCMC, the best fit obtained in Case 2(b) is very similar to case 2(a) and differences in the posterior parameter distributions are small (Fig. 14). The most notable difference in comparison to Case 2(a) is seen in the predictive distribution (Fig. 15 and Table 5), which shows smaller mean values for the predicted magnitude as well as a slightly broader uncertainty range particularly for longer return periods.

Traditional MCMC

```
case2b$w <- 1/(1-exp(-case2b$x/10))*0.05
chain_2b <- mcmcrun(list(modelfun=modelfun),case2b,paramdef,mcmcopts)
```

```
## number of accepted runs: 2069 out of 12000 (17.24167%)
```

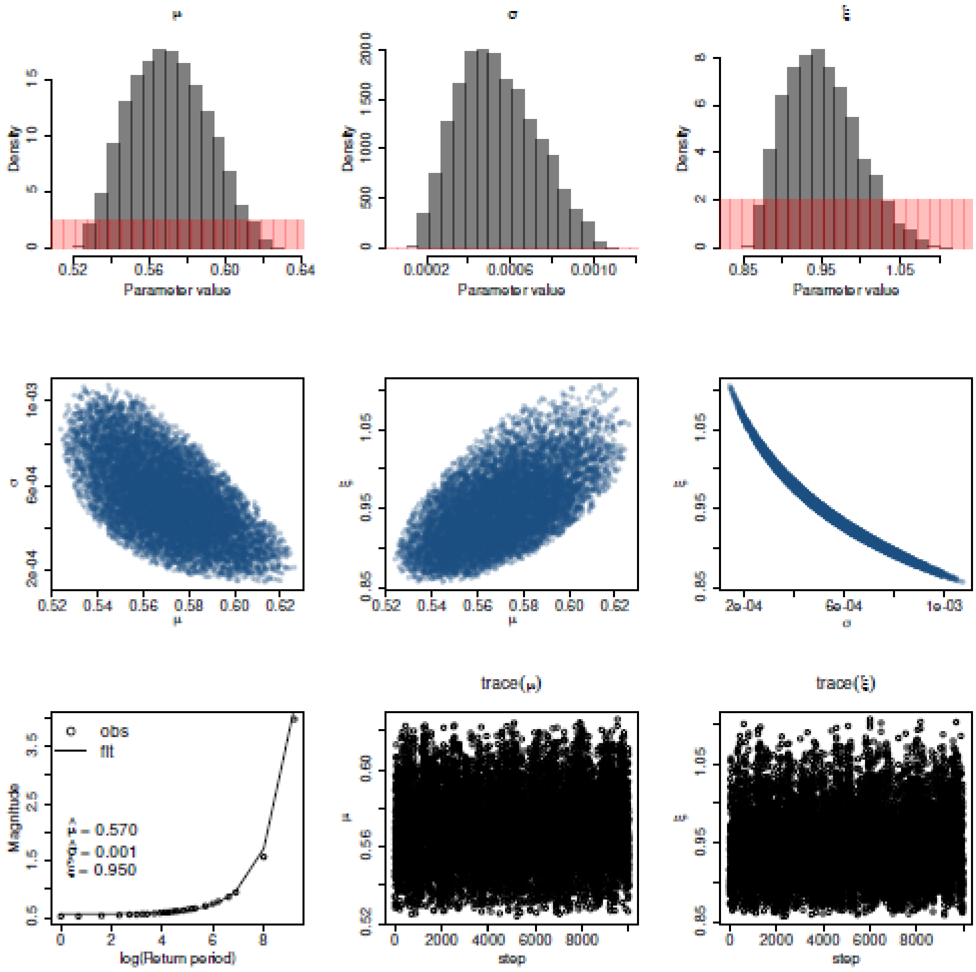


Figure 14: Same as in Fig. 6 but for ABC-MCMC Case 2(b).

Table 6: The estimated mean return level together with the 5th and 95th percentile of the predictive distribution for return periods between 500-500 000 years obtained with traditional MCMC in Case 2(b).

	500	5 000	50 000	500 000
Mean	0.76	2.30	16.90	156.36
5%	0.74	2.24	15.68	130.85
95%	0.78	2.35	18.26	186.96

20

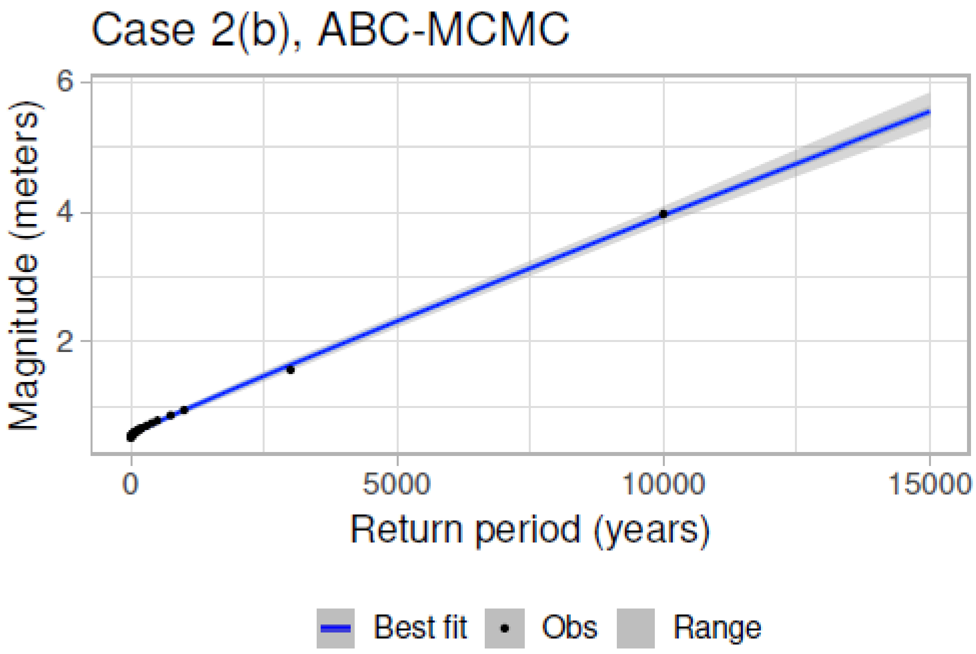


Figure 15: Same as in Fig. 7 but for ABC-MCMC Case 2(b).

As in Case 2(a), differences between the results for ABC-MCMC and MCMC are modest. (Figs. 16 and 17). For the longest recurrence intervals, MCMC provides slightly higher values for the estimated mean return levels, while the uncertainty range is smaller (Table 6).

Case 2(c)

ABC-MCMC

In ABC-MCMC, observational uncertainty is emulated by adding random noise to the observed summary statistics using the observed standard deviation when estimating the distance between the simulated and observed quantiles. Otherwise, the technical implementation is similar to Case 2(a) and 2(b).

Fig. 18 shows that the inclusion of random noise to the summary statistics leads to slightly smaller posterior values for ξ. Even though the uncertainty range is larger than in Cases 2(a) and 2(b), ABC sampling is not able to reproduce the given uncertainty range (Fig. 19). On the other hand, the uncertainty range in the predictive distribution extends to much smaller values, which means that incorporating observational uncertainty to the parameter estimation tends to produce somewhat smaller return level estimates compared to the previous cases (Table 7).

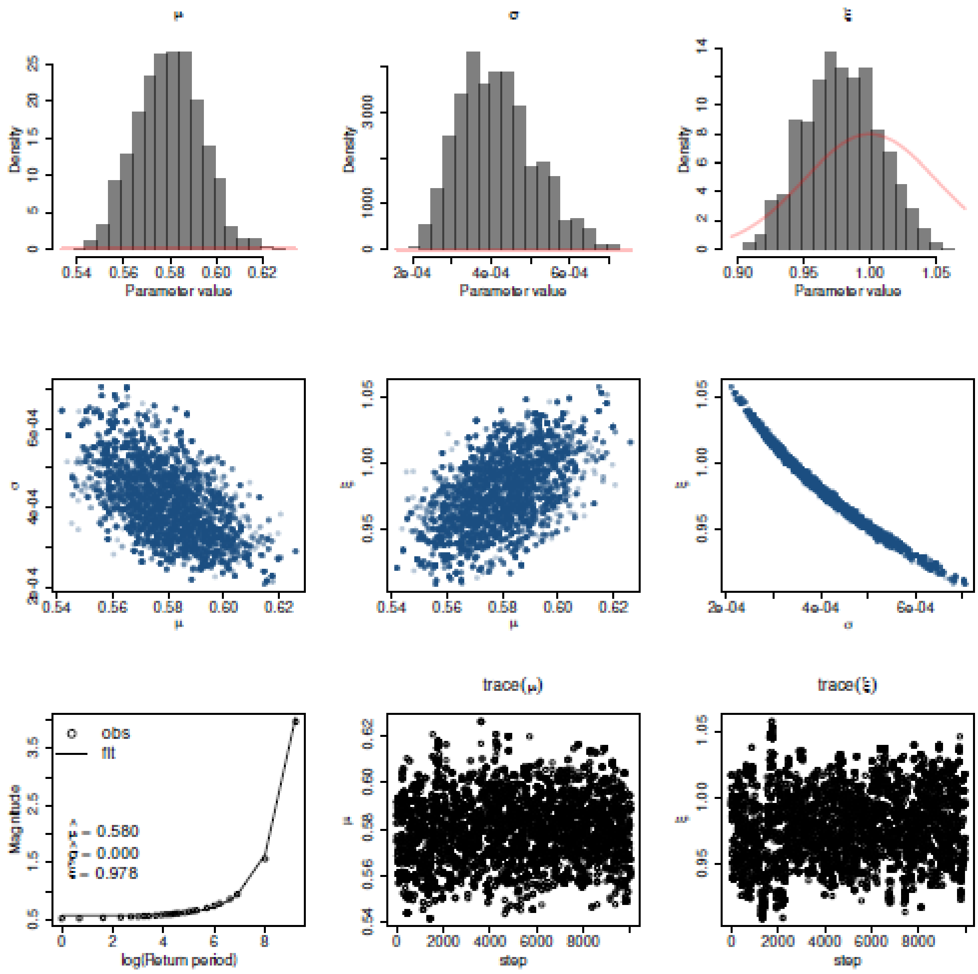


Figure 16: Same as in Fig. 8 but for MCMC in Case 2(b).

Table 7: The estimated mean return level together with the 5th and 95th percentile of the predictive distribution for return periods between 500-500 000 years obtained with ABC-MCMC in Case 2(c).

	500	5000	50000	500000
Mean	0.78	2.28	14.16	110.94
5%	0.73	2.02	9.75	54.46
95%	0.84	2.56	19.25	189.19

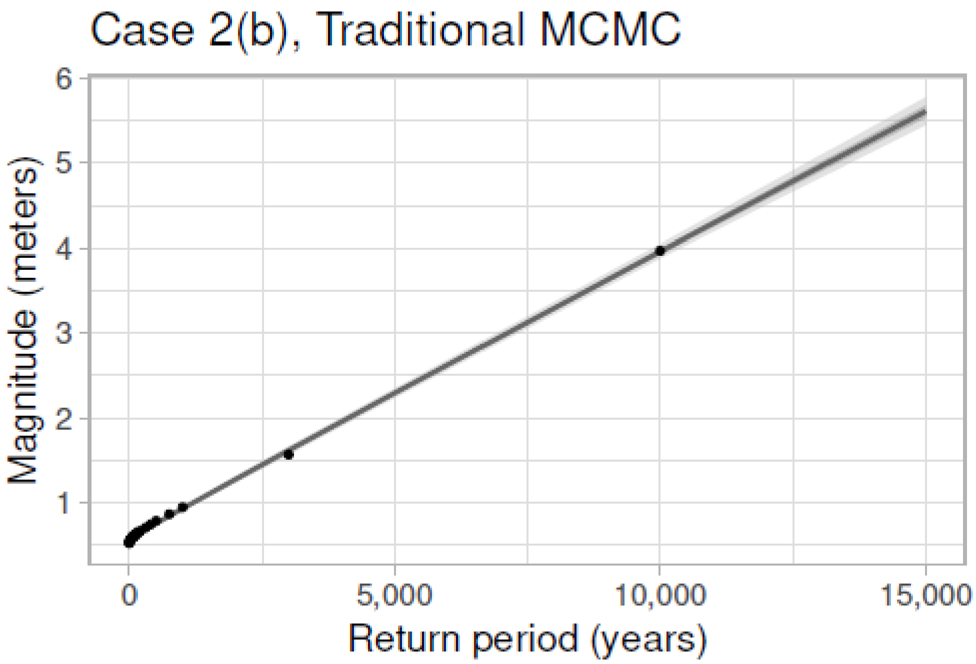


Figure 17: Same as in Fig. 9 but for MCMC Case 2(b).

Traditional MCMC

We apply traditional MCMC sampling using the given standard deviation with a suitable weighting function to take the observational uncertainty into account when estimating the joint posterior parameter distribution.

```
case2c$w <- case2c$sd2 * 1/(1-exp(-case2c$x/10))
chain_2c <- mcmcrun(list(modelfun=modelfun),case2c,paramdef,mcmcopts)
```

```
## number of accepted runs: 3154 out of 12000 (26.28333%)
```

As can be seen from Fig. 20 the obtained joint posterior distribution is relatively similar to the one for ABC-MCMC. The main difference between the two approaches is revealed by looking at the predictive distribution (Fig. 21), which shows a much better correspondence to the mean observation for MCMC, although the range still does not cover the full observed uncertainty range. In particular, the lower boundary of the predicted uncertainty range is too high in comparison to the observed one. On the other hand, the predictive uncertainty extends to larger values at the 500 000-year return level for MCMC than for ABC-MCMC (Table 8).

Results of the assessment

The upper bound of uncertainty range for the simulated return level reaches 200 m in Case 2 for the longest requested recurrence interval, which is roughly two orders of magnitude

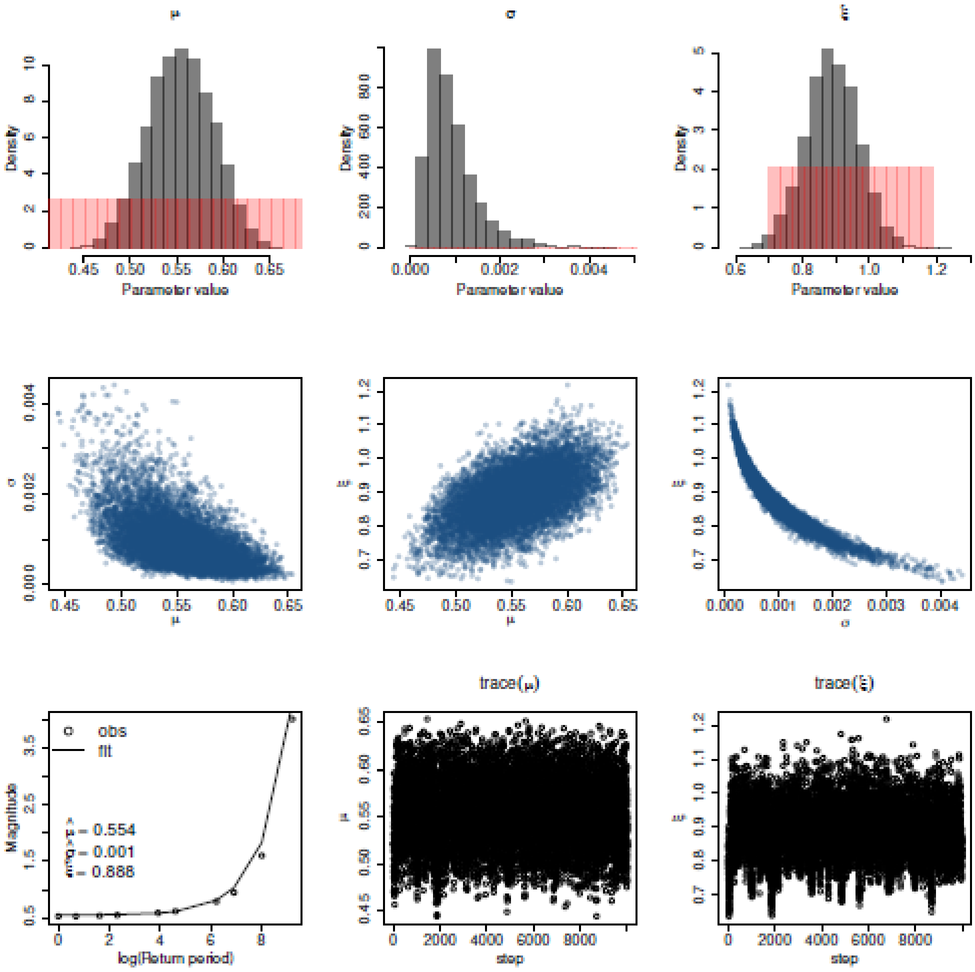


Figure 18: Same as in Fig. 6 but for ABC-MCMC in Case 2(c).

Table 8: The estimated mean return level together with the 5th and 95th percentile of the predictive distribution for return periods between 500-500 000 years obtained with traditional MCMC in Case 2(c).

	500	5 000	50 000	500 000
Mean	0.77	2.36	16.95	151.60
5%	0.73	2.12	13.33	101.76
95%	0.80	2.61	21.12	215.07

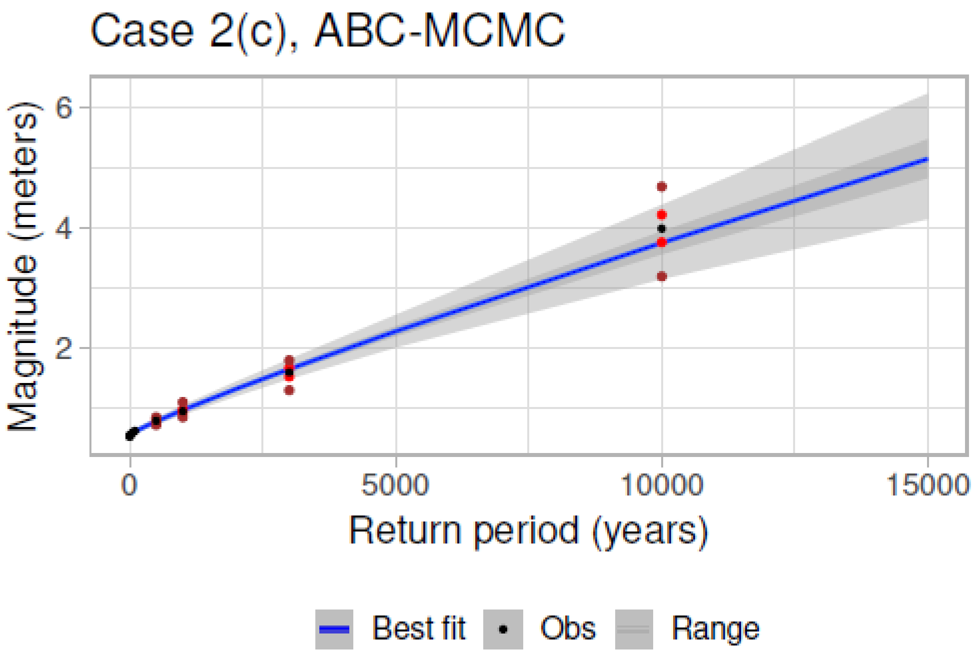


Figure 19: Same as in Fig. 7 but for ABC-MCMC in Case 2(c).

larger than the 10 000-year return level given in the synthetic data set. While this value sounds unrealistically large regardless of potential physical phenomenon the synthetic data tries to emulate, without further information on the actual nature of the extremes generating process it is difficult evaluate the plausibility of the results. In Case 2(c), uncertainty estimates provided for the predicted distribution quantiles suggest that M could be substantially smaller than what would be inferred based on Cases 2(a-b) only. Thus, while GEV is the optimal distribution for describing the asymptotic behavior of sample maxima, it possible that it overestimates the return levels for very large recurrence intervals in this case.

An obvious shortcoming of ABC-MCMC is the need for tuning of several parameters and proper selection of summary statistics. In our case, it was natural to use quantiles directly as the summary statistic. In many cases, however, the selection of descriptive summary statistics is not straightforward and requires substantial physical insights to the problem itself. Even in our case it might have been beneficial to use some other statistics such as some optimal combination of distribution quantiles. We used an automatic selection of acceptance distance and proposal range in our ABC implementation before running the MCMC sampler, which required a substantially large sample size (10^6) to work properly in Case 2. However, the overall computation benefits are substantial and the results were much better than what would have been obtained without the initialization step.

Implementation of traditional MCMC was based on the assumption on Gaussian likelihood. The results are very similar to ABC-MCMC, which suggests that this assumption is feasible. Furthermore, as only 3 parameters were estimated MCMC turned out to be

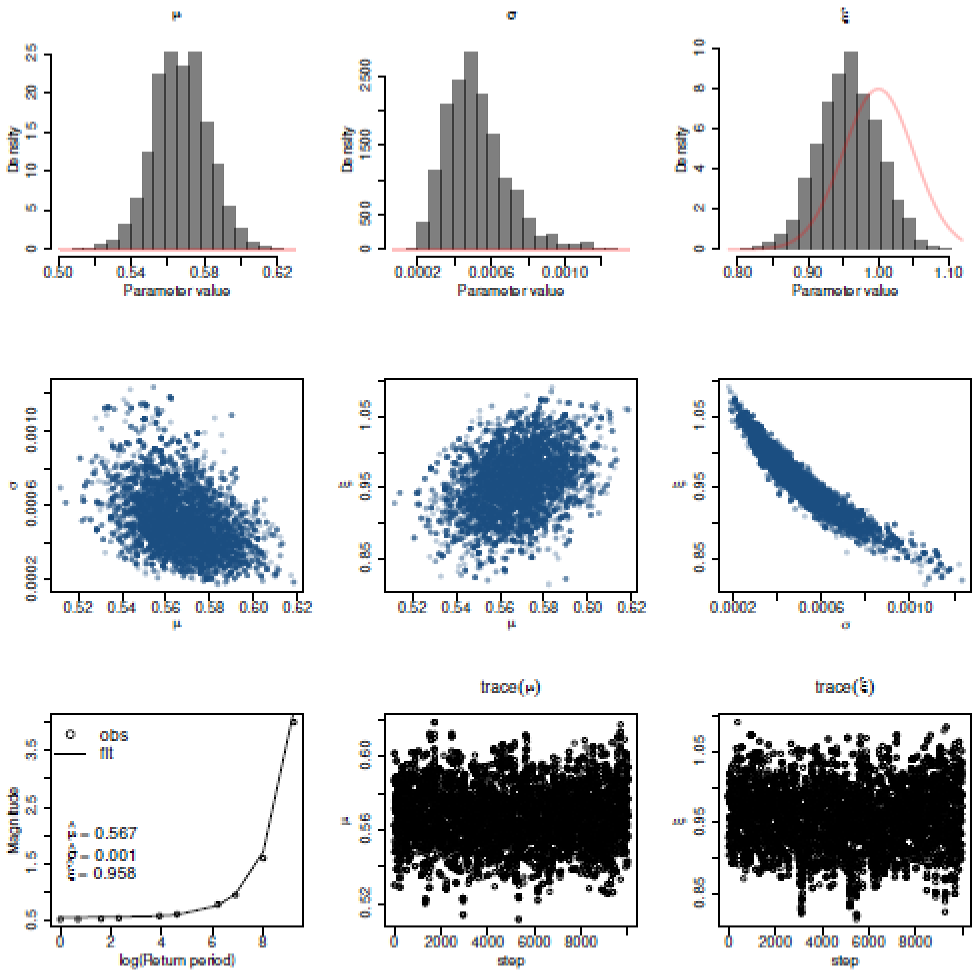


Figure 20: Same as in Fig. 8 but for MCMC in Case 2(c).

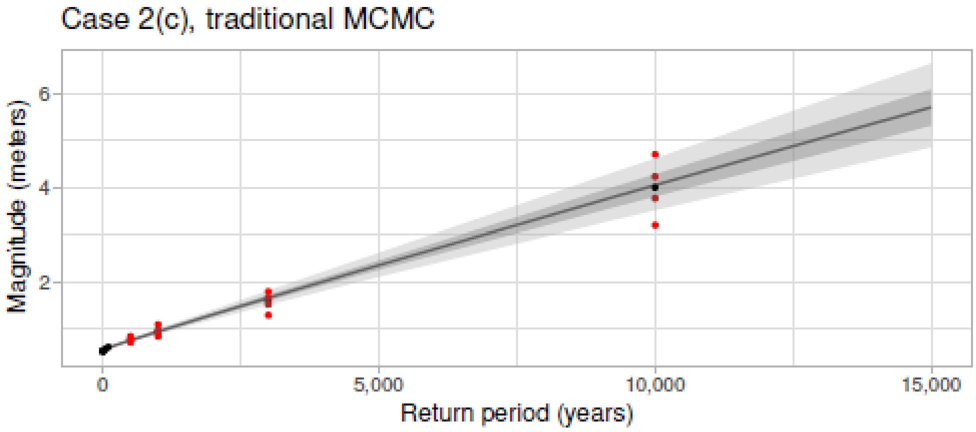


Figure 21: Same as in Fig. 9 but for MCMC in Case 2(c).

Table 9: Summary of the obtained mean return levels (M) with respect to the requested return periods (t) in each of the cases (ABC-MCMC/MCMC).

	500	5000	50000	500000
Case 1	1.85/1.84	2.35/2.35	2.84/2.88	3.34/3.43
Case 2(a)	0.76/0.76	2.33/2.31	16.62/17.25	148.36/162.27
Case 2(b)	0.77/0.76	2.32/2.30	16.22/16.90	142.11/156.36
Case 2(c)	0.78/0.77	2.28/2.36	14.16/16.95	110.94/151.60

computationally much efficient than ABC-MCMC. One source for uncertainty in MCMC is the selection of proper observational weights for M. There is no single preferred way to do this and our approach is based on heuristic reasoning how the weights could behave.

Conclusions

This notebook provides a solution to the OECD/NEA exercise, where the aim is "to describe the assumptions made to create the hazard frequency/magnitude model(s), the qualitative and quantitative results of the model(s), the process used to assess the adequacy of the model(s), and the results of the model adequacy assessment". We selected generalized extreme value distribution as our model candidate due to its well-know theoretical background and the relatively straightforward parameter estimation. Two Bayesian approaches for parameter estimation were tested: (i) Approximate Bayesian Computation (ABC-MCMC) which does not require information on the underlying likelihood function and (ii) pure MCMC sampling assuming Gaussian likelihood. Best estimates for the GEV parameter values and the corresponding mean return level estimates are provided as tables for both ABC-MCMC and pure MCMC approaches (Table 9 and 10).

The mean return level estimates are qualitatively similar for both methods. MCMC tends to produce slightly larger mean values for the 500 000-year return levels in all cases.

Table 10: Best estimates of the distribution parameters for all test cases (ABC-MCMC/MCMC).

	μ	σ	ξ
Case 1	0.507/0.556	0.218/0.2010	-0.000/0.014
Case 2(a)	0.565/0.579	0.001/0.0004	0.959/0.983
Case 2(b)	0.570/0.580	0.001/0.0004	0.950/0.978
Case 2(c)	0.554/0.567	0.001/0.0005	0.888/0.958

Comparison of the posterior mean parameter values between ABC-MCMC and MCMC shows that differences are not very large. MCMC tends to give slightly larger values for μ and ξ, while the opposite is seen for σ. Furthermore, the parameter values are close to each other in Case 2(a) and (b), while in Case 2(c) ξ is slightly smaller particularly for ABC-MCMC, which is also seen as smaller values of simulated M in the predictive distributions.

References

Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values.* Springer, 2001. doi: 10.1007/978-1-4471-3675-0.

Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001. doi: 10.2307/3318737.

Heikki Haario, Marko Laine, Antonietta Mira, and Eero Saksman. DRAM: efficient adaptive MCMC. *Statistics and Computing*, 16(4):339–354, 2006. doi: 10.1007/s11222-006-9438-0.

Franck Jabot, Thierry Faure, and Nicolas Dumoulin. Easyabc: performing efficient approximate bayesian computation sampling schemes using r. *Methods in Ecology and Evolution*, 4(7):684–687, 2013. doi: https://doi.org/10.1111/2041-210X.12050. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12050.

Paul Marjoram, John Molitor, Vincent Plagnol, and Simon Tavaré. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26): 15324–15328, 2003. doi: 10.1073/pnas.0306899100.

Matthew A. Nunes and Dennis Prangle. abctools: An R Package for Tuning Approximate Bayesian Computation Analyses. *The R Journal*, 7(2):189–205, 2015. doi: 10.32614/RJ-2015-030.

Daniel Wegmann, Christoph Leuenberger, and Laurent Excoffier. Efficient approximate bayesian computation coupled with markov chain monte carlo without likelihood. *Genetics*, 182(4):1207–1218, 2009. ISSN 0016-6731. doi: 10.1534/genetics.109.102509. URL https://www.genetics.org/content/182/4/1207.

Annex C. Submission by Idaho National Laboratory Group 1

Introduction

This paper responds to an OECD Nuclear Energy Agency (NEA) report that calls for a benchmark exercise on the analysis and assessment of the hazard frequency and magnitude for external events risk assessments. The objective of this benchmark study is to focus on the statistical modelling of an external hazard initiating event (IE), assessing its frequency and magnitude, that could be used in probabilistic risk assessment (PRA) of external hazards. The NEA report specifies two cases of hypothetical observational data created from synthetic models (which are used to create synthetic data that have been generated from a computer) for a hypothetical external event (e.g. precipitation, extreme temperatures, high winds): (1) a fully revealed "open" case where both the synthetic data and the synthetic model producing the data are provided, and (2) a "blind-test case" where only the synthetic data are provided. This paper documents the author's analysis steps, assumptions, insights and modelling results for the exercise.

Case 1 – Known model producing the synthetic data

The following synthetic model is provided in Case 1 to create hypothetical observational data for the hazard frequency/magnitude modelling:

$$M = 0.5 + 0.5*log_{10}(a*t)$$

The synthetic model serves as surrogate for a complex phenomenological process. It is of the form that different values of "return time intervals" t produce a hypothetical (but known since it comes from the synthetic model) magnitude M for an annual maxima event. These types of models can be used to produce "synthetic data" and predict different event outcomes as a function of time (e.g. producing a flooding hazard curve).

The hypothetical observational data for the magnitude M (in metres) and return period t (in years) from the synthetic model with the "a variable" set to 1 is shown in Table C.1.

Table C.1. Synthetic data for Case 1.

Return Period (years)	1	2	5	10	50	100	500	1 000	2 000	10 000
Magnitude (metres)	0.50	0.65	0.85	1.0	1.4	1.5	1.9	2.0	2.2	2.5

Participants are asked to use the data for Case 1 and provide a model that *best* described the frequency/magnitude relationship and the associated analysis and insights. The results of this analysis should include those areas identified in Chapter 3 of this benchmark, including:

- Qualitative aspects and insights.
  - Assumptions made to create the hazard frequency/magnitude model.
  - The process used to assess the adequacy of the model.

- Quantitative aspects and insights including:
  - The type/form of the model describing the hazard frequency and magnitude statistical results.
  - Uncertainties of the model. Assessing uncertainty is important for both validation and prediction [NRC, 2010].
  - Results of the model adequacy assessment or validation.

Case 2 – Unknown model producing the synthetic data

In Case 2, the synthetic model used is not provided. Instead, three sets of synthetic data output from the unknown model are provided.

- Case 2a provides the synthetic data (ten data points) with no uncertainty associated with the data points (see Table C.2).
- Case 2b provides additional synthetic data (26 data points) with no uncertainty associated with the data points (see Table C.3).
- Case 2c has uncertainty estimates on some of the data (see Table C.4).

Table C.2. Synthetic data for Case 2a

Return Period (years)	**1**	**2**	**5**	**10**	**50**	**100**	**500**	**1 000**	**3 000**	**10 000**
Magnitude (metres)	0.53	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0

Table C.3. Synthetic data for Case 2b

Return Period (years)	**1**	**2**	**5**	**10**	**15**	**20**	**25**	**30**	**40**
Magnitude (metres)	0.53	0.53	0.54	0.55	0.56	0.56	0.57	0.57	0.58
Return Period (years)	50	60	70	80	90	100	125	150	175
Magnitude (metres)	0.59	0.60	0.60	0.61	0.62	0.62	0.63	0.65	0.66
Return Period (years)	200	300	400	500	750	1 000	3 000	10 000	
Magnitude (metres)	0.67	0.71	0.75	0.79	0.87	0.95	1.57	3.97	

Table C.4. Synthetic data for Case 2c

Return Period (years)		**1**	**2**	**5**	**10**	**50**	**100**	**500**	**1 000**	**3 000**	**10 000**
Magnitude (metres)	Mean	0.53	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
	Sdev.*	**						0.04	0.06	0.15	0.46
	5th***							0.72	0.85	1.3	3.2
	95th							0.85	1.1	1.8	4.7

Assumptions

A generalised extreme value (GEV) model was used in this benchmark exercise to predict the magnitude values for long return periods using the synthetic data provided in Case 1 and Case 2 (Table C.2., C.3. and C.4.). OpenBUGS was used to implement the GEV model. It is believed by many analysts that observed data from relatively short-term return periods should not be used to predict the magnitude values for much longer-term periods (e.g. use observed data up to 5 000 years to predict for 50 000 years). However, when such needs arise, for example for the use of external hazard frequency in probabilistic risk assessment (PRA) models, various approaches and engineering judgement have to be used to provide such estimates, and the GEV model is believed to be such a tool to provide best estimates, with the modelling uncertainty being in mind when using the results from the model.

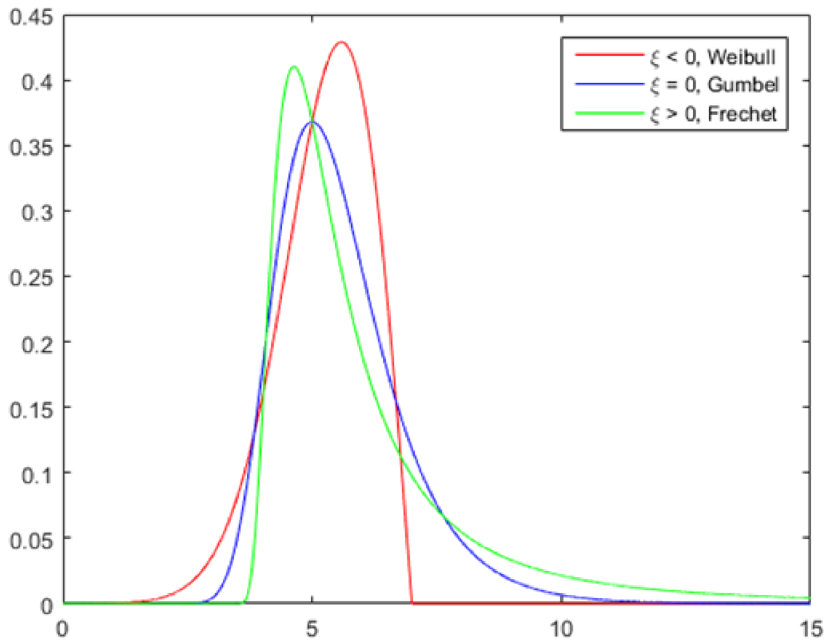
Assessment approach

The GEV is a single family of limiting distribution that combines three other limit distributions: Gumbel, Fréchet and Weibull. It has the following cumulative distribution functional form:

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \qquad (1)$$

where ξ is a shape parameter, μ is a location parameter, and σ is a scale parameter. The shape parameter ξ determines the distribution type and the tail behaviour. $\xi = 0$ corresponds to the Gumbel distribution, which is unbounded and has exponential tail. $\xi > 0$ corresponds to the Fréchet distribution, which has lower bound and long tail. $\xi < 0$ corresponds to the Weibull distribution, which has an upper bound and short tail (Figure C.1).

Figure C.1. Generalised extreme value distribution family



The inverse distribution function, or quantile function, is often used to calculate the extreme values:

$$z_p = G^{-1}(1-p) = \mu - \frac{\sigma}{\xi}\{1 - [-\log(1-p)]^{-\xi}\}, \xi \neq 0 \qquad (2)$$

The return level z_p is exceeded by the annual maximum in any particular year with the annual exceedance probability p, when the measuring time is in years. Or one can say that the return level z_p is exceeded, on average, once during the return period $1/p$.

The GEV model shown in Eq. (1) was applied to the synthetic data in this benchmark exercise using OpenBUGS. An example of the script used for Case 1 is shown in Table C.5.

Table C.5. OpenBUGS script for Case 1 using GEV model

```
model
{ for(i in 1:N) {
z.p[i] ~ dnorm(mean[i],prec)
y.p[i] <- -log(1 - p[i])
mean[i]<- mu - sigma/xi*(1 -pow(y.p[i],-xi))
}
mu ~ dnorm(0,0.0001)
prec<-pow(sd,-2)
sd~dunif(0,10)
xi ~ dunif(-1,1)
sigma ~ dunif(0,10)
}
data
list(p=c(0.632, 0.393, 0.181, 0.0952, 0.0198, 0.00995, 0.002, 0.001, 0.0005, 0.0002, 0.0001, 0.00002,
0.000002),
z.p=c(0.50, 0.65, 0.85, 1.0, 1.4, 1.5, 1.9, 2.0, 2.2, NA, 2.5, NA, NA), N=13)
list(mu=1.0, sigma=1.0, xi=1.0)
```

Results of the assessment

The predicted results for Case 1 are shown in Table C.6. The results for Case 2 are shown in Table C.7(Case 2a), Table C.8(Case 2b), and Table C.9(Case 2c).

Table C.6. Case 1 results

Return Period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1 Exact	1.9	2.4	2.9	3.4
INL mean (metres)	1.88	2.37	2.84	3.31

Table C.7. Case 2a results

Return Period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
INL mean (metres)	0.76	2.34	16.02	136.20

Table C.8. Case 2b results

Return Period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
INL mean (metres)	0.76	2.32	16.26	142.80

Table C.9. Case 2c results

Return Period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
INL mean (metres)	0.76	2.34	16.02	136.20

Annex D. Submission by Idaho National Laboratory Group 2

Introduction

The aim of the benchmark is to investigate the frequency/magnitude relationship for external events. Two cases represent different external hazard from synthetic data. The first case provides both the synthetic data and synthetic model. The second case introduces three parts with only the synthetic data, where the first and second parts do not consider the uncertainty, whereas the third part of the magnitude has been provided with the uncertainty. The magnitude from those two cases is modelled as a function of return period. The return period is also known as recurrence interval, which is the estimated average time between floods events occurring.

Linear regression with a transformed form model would be considered to fit the magnitude vs. return period relationship model for case1. Cases using ordinary least squared method would have the same assumptions and similar model adequacy assessment approach. A linear model has the following general form:

$$Y_i = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2 \ldots + \epsilon_i$$

where the Y is the response variable, x are predictor variables, ϵ is the residuals derived from the differences between tabulated values and predicted values.

Assumptions

For a linear regression model with ordinary least squared method, there are several assumptions:

1. residuals are approximately independent, homoscedastic and normally distributed with zero mean; and
2. there is linearity between the response and predictors.

Model adequacy assessment approach

Linear least squared was used to estimate the parameters in the regression. The model will be evaluated using the coefficient of determination R^2 and F-test to assess the model adequacy. R^2 could give the strength of the linear relationship between the predictor and the response variable. F-test provides the test result of whether the linear regression model provides a better fit to the data than a model only with intercept. The F-statistics is: $F^* = (\frac{SSE(R) - SSE(F)}{d_{fR} - d_{fR}}) / (\frac{SSE(F)}{d_{fF}})$, where the $SSE = \sum(observed - fitted)^2$, SSE(F) is the predicted full model, and SEE(R) is the reduced model with only the intercept term, df(Residual) = n – (k+1) where k is number of parameters being estimates and n is the sample size. The $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$, where the $SS_{res} = \sum_i {\epsilon_i}^2$, is the residual sum of squares, and the $SS_{tot} = \sum_i (y_i - \bar{y})^2$, is the total sum of squares.

Four different types of diagnostic plots will assess each assumption of a linear regression model:

1. Residuals vs. fitted: is used to check the residual homoscedasticity assumption. If a linear model is correctly specified, the residual vs. fitted plots should not have any systematic features with a flat line. Studentised Breusch-Pagan tests would be used if this graph does not show a visible flat line with separated dots around it. The ξ^* is the studentised test statistics and it is calculated by finding out the R^2 when running the auxiliary regression equation of $\hat{\epsilon}^2 = \gamma_0 + \gamma_1 x + v$; examined by regressing the squared residuals on each independent variable.
2. Normal QQ: is used to check whether the residuals are normally distributed.

   A Shapiro-Wilk normality test is another option used to check this assumption. The W statistic is:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\left(\sum_{i=1}^{n} x_i - \bar{x}\right)^2}$$

   Where the $x_{(i)}$ are the ordered sample values, a_i are constants generated from the means, variances and covariance's of the order statistics of the sample from a normal distribution.
3. Residuals vs. leverage: is used to check the outliers and identify the influential cases, since extreme values may affect the regression results if those values are excluded or included from the analysis. Outliers could be identified if they are outside of the Cook's distance dashed line.

Case 1

The hazard frequency/magnitude model here for Case 1 fits the linear regression model with log transformation of return period to describe the relationship between magnitude and return period since the synthetic model provided is linear. The difference is the residual term in regression equation, which is the vector values of the differences between observed values and predicted values

$$\widehat{Magnitude_i} = \widehat{\beta_0} + \widehat{\beta_1} log(return\ period_1) + \epsilon_i$$

Model adequacy assessment approach

The estimated equation is:

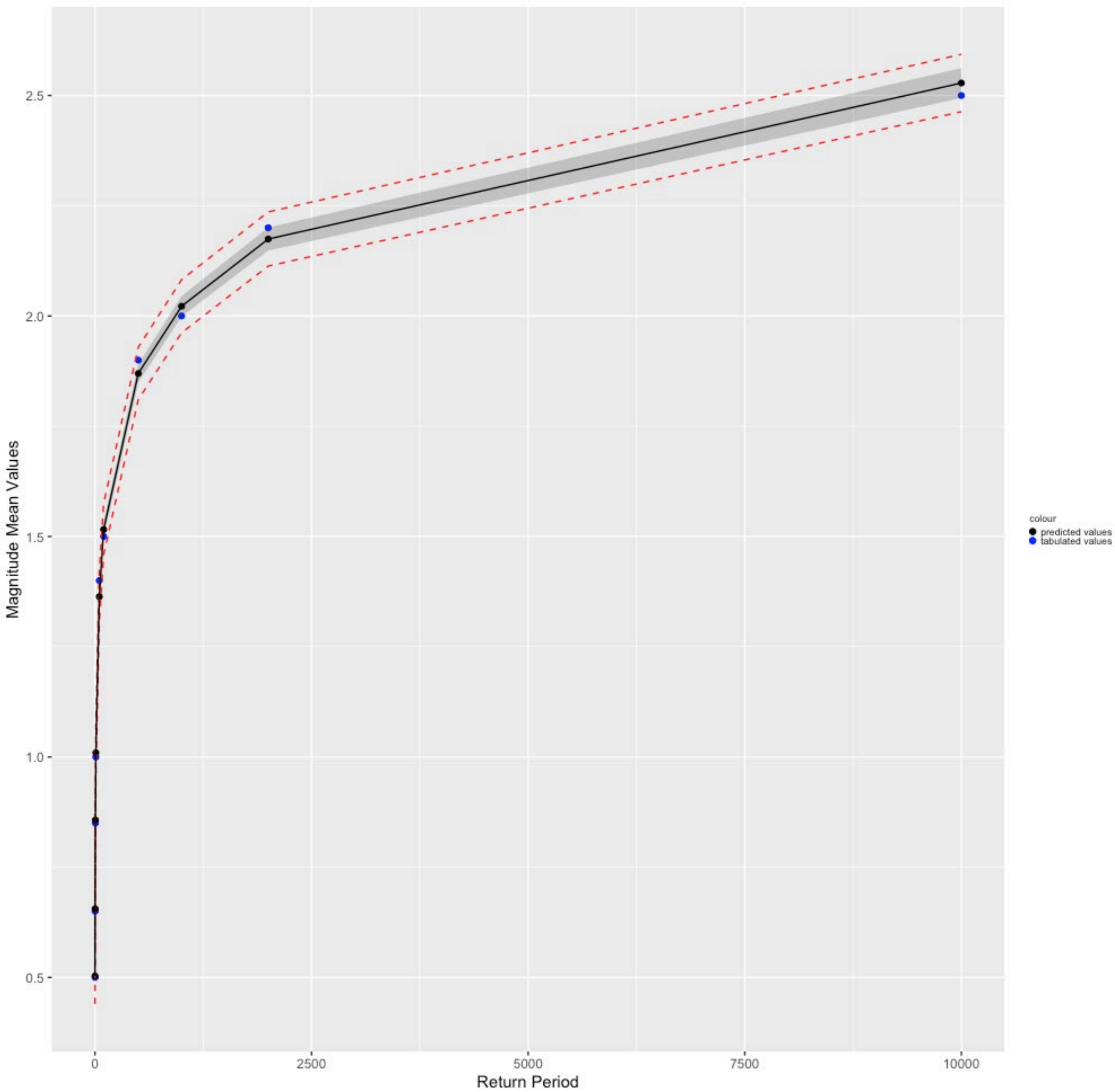
$$\widehat{Magnitude} = 0.503364 + 0.506251 * log(\text{return period}).$$

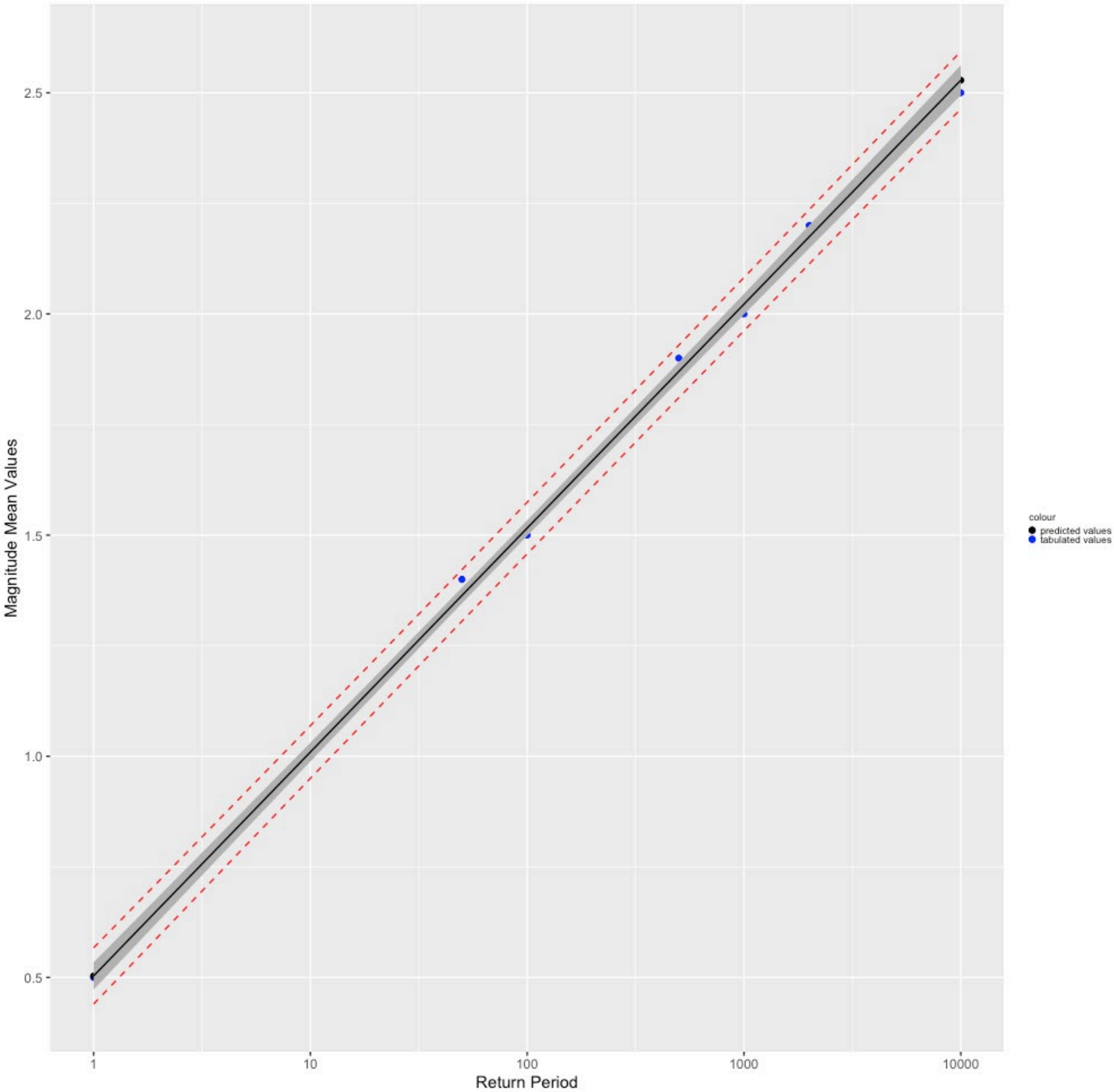
The estimated values above are the least square estimates of the intercept and slope. They have standard error for the intercept and slope are 0.07006 and 0.03083, respectively.

The coefficient of determination $R^2 = 0.9988$, which means 99.88% of the uncertainty in the predicted values can be explained by the straight-line regression model. The assessment of model uncertainties is essential to the practical application of the models to inform the decision makers. The small uncertainties of the models will provide the confidence on decision making.

Figure D.1 shows the tabulated/predicted data and the regression model. There is a narrow range of 95% confidence band. Two tabulated values out of ten not underlying on the 95% confidence band, but they all fall into the prediction intervals.

Figure D.1. Return period versus mean magnitude





The following, Figure D.2 to Figure D.5, are the four different diagnostic plots for the regression model:

Figure D.2 is the plot of residuals vs. fitted values. The residuals do not change a lot as the fitted values increase or decrease. Here, an approximate horizontal line without any patterns shown in Figure D.2 indicates a linear relationship. Figure D.3 is the normal QQ plot. Although there are a few points that do not follow the straight dashed line, the normality of assumption is not violated. A Shapiro-Wilk normality test would be used to check this assumption. A p-value (0.161) of this test suggests that no evidence rejected the normality assumption. Figure D.4 is the plot of Scale-Location.

An approximate flat line shown in this figure demonstrates a constant variance of the residuals. Figure D.5 is the residual vs. leverage plot. There is a data point in this figure that exceeds Cook's distance, which is probably influential to the regression results. Excluding this outlier data point and re-running the log-transformed variable model, we see that there is not a huge difference between the two models. The sample size is small and there is a large time interval corresponding to each observation. The return period for expected magnitude prediction on this benchmark is larger than 10 000, which means excluding this point is not appropriate.

Figure D.2. Residuals vs fitted

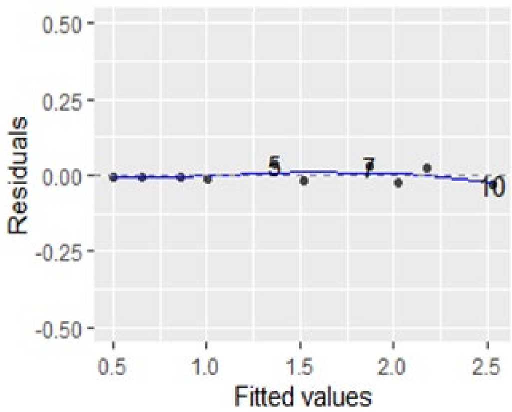


Figure D.3. Normal Q-Q

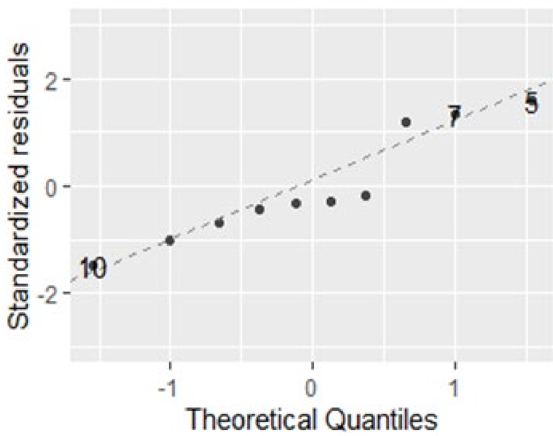


Figure D.4. Scale-location

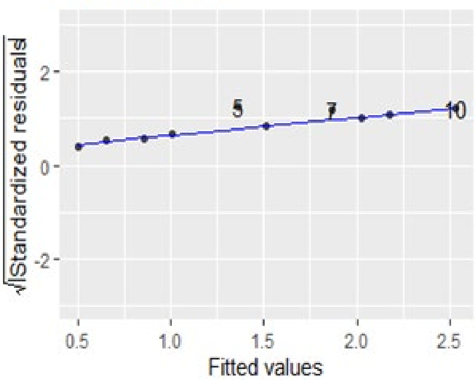
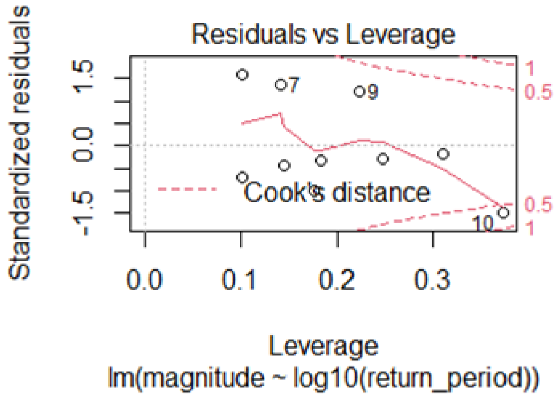


Figure D.5. Residuals vs leverage



Results

The following table shows the magnitude prediction for Case 1 for the return period of:

Table D.1. Resulting magnitude prediction for Case 1

Return Period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1	1.87	2.38	2.88	3.39

The model for Case 1 indicates that for 1% increase in the return period variable, the magnitude increases by about 0.506*log (1.01) = 0.005.

Results of Assessment

Utilising the linear regression to predict magnitude is simple and efficient, there are some technical basis and limitations. For example, the given range of magnitudes tabulated for up to 10 000 return period years and limited observations cause high confidence and model uncertainty. If a linear regression is correctly specified, all of its assumptions must be met including the linearity of data, normality of residuals, independence and homogeneity of residuals. For most of the cases, the data does not simply follow the linearity assumption, we need to transform their forms to meet the assumptions. Besides, even though our model for Case 1 has high value of R^2, the small size and outliers happened may influence the model results so that there are a few points not underlying on the confidence intervals. The biggest difference between the synthetic model and the regression we fit is the random error term associated with the model, which makes the regression model be a stochastic model. However, the trend of the linear regression is constant, so that an exact relationship between variables is determined without considering the random error component. The linear regression model hypothesises a probabilistic relationship between magnitude and return period.

Case 2a

The hazard frequency/magnitude model here for Case 2 does not have a synthetic model provided. For the first case, Non-linear regression would be considered to fit the magnitude vs. return period relationship for Case 2a. A non-linear regression model has the following general form:

$$Y_i = \widehat{f(\theta, \mathrm{x})} + \epsilon_i$$

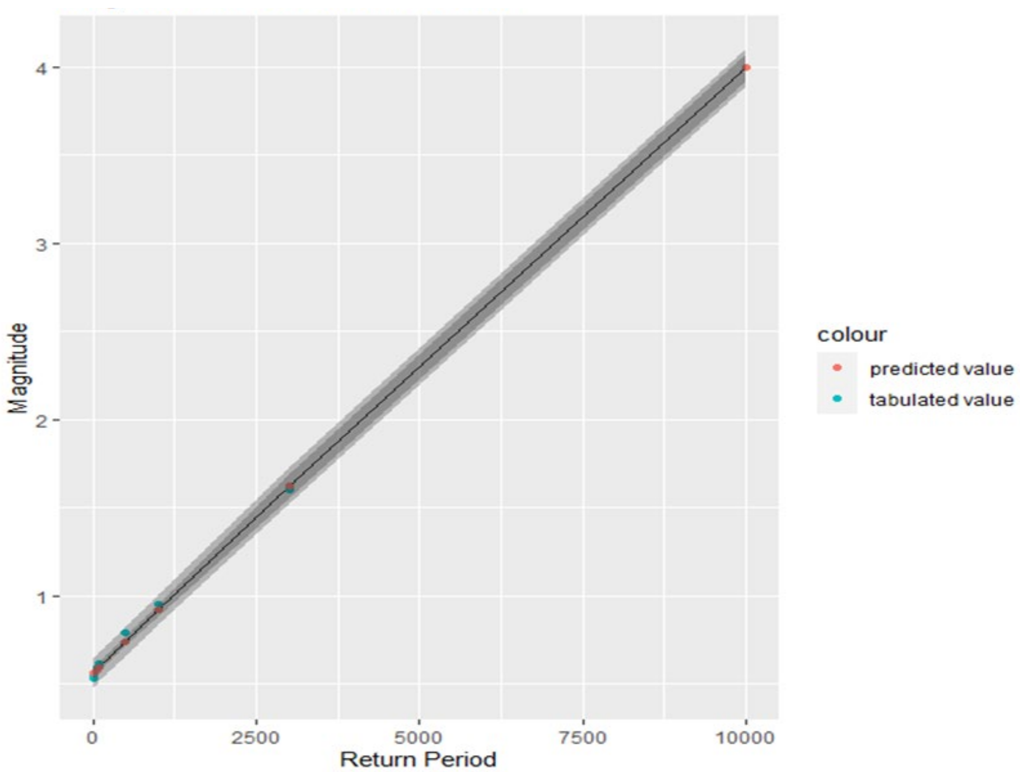
Where Y is the response variable, x is the independent variable, θ is the vector of model parameterthe relationship between x and y through the function $f(\theta, x)$. and ϵ is the residual error term.

Case 2a has an exponential form relationship between outcome interest and predictor, the estimated equation is:

$$\widehat{Magnitude_i} = -36.9e^{-0.000009781*return\ period_i} + 37.46$$

Figure D.6 shows the tabulated/predicted data and the regression model with 95% confidence interval and prediction interval. There is a narrow range of 95% confidence band when the return period is below 1 000 years, and there are a few tabulated values do not fall on the 95% confidence band.

Figure D.6. Magnitude vs return period



Assumptions

For a non-linear regression model with least squared method also has the same residuals assumption:

Residuals are approximately independent, homoscedastic and normally distributed with zero mean.

Model adequacy assessment approach

Unlike the linear regression, the R^2 could explain the goodness of fit for a model, R^2 is invalid for non-linear regression. To access the goodness of fit for non-linear regression, this could be done by looking that the correlation between the actual observed values and values predicted by a model. A high correlation coefficient (0.99) indicates that observed and predicted values are very close to each other.

Since the non-linear regression has the same residuals assumption as the linear regression has, the following Residuals vs. Fitted, normal Q-Q plot, Autocorrelation plots would check the residual assumption separately.

Figure D.7. Residuals

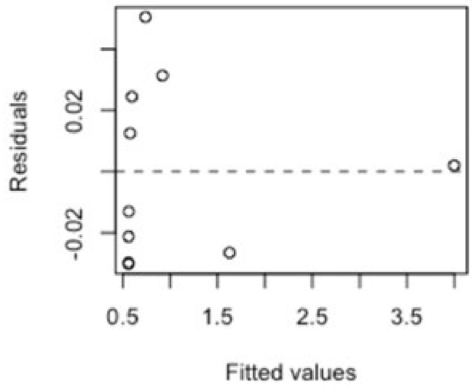


Figure D.8. Autocorrelation

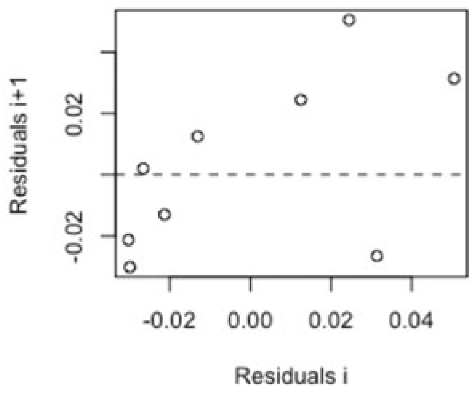
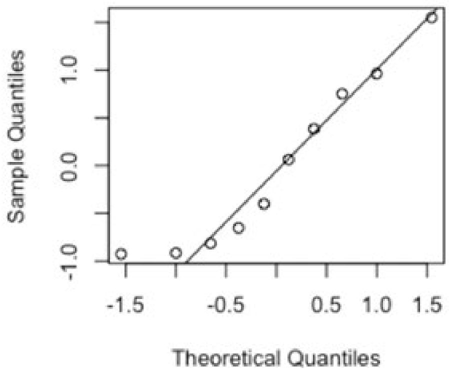


Figure D.9. Normal Q-Q plot of standardised residuals



For the Residual vs Fitted plot, each corresponding residuals from fitted values does not change a lot between each other, which indicates the assumption of residual homoscedastic is met. However, it seems like the autocorrelation plot show an increasing trend. A Runs test was used to determine the autocorrelation exists. The test statistics of Runs test is $Z = \frac{r - \mu_r}{\sigma_r}$, where $\mu_r = \frac{2n_0n_1}{N} + 1$ is the expected number of runs, $\sigma_r = \frac{2n_0n_1(2n_0n_1 - N)}{N^2(N-1)}$ is the standard deviation of the number of runs, r is the observed number of runs, n_0 is the number of observations below the threshold, n_1 is the number of observations above the threshold, and $N = n_0 + n_1$. A p-value of 0.18 indicates autocorrelation assumption is not violated. The last graph is normal Q-Q plot with a p-value (0.27) of Shapiro-Wilk normality test indicates a normality assumption is met.

Results

The following table shows the magnitude prediction for Case 2-1 for the return period of:

Table D.2. Resulting magnitude prediction for Case 2-1

Return Period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1	0.74	2.32	14.83	37.18

Results of assessment

An exponential non-linear regression is a good option to fit the non-linear relationship between predictors and outcome and to meet the desired condition that the magnitude would be stabilised when the return period is large enough. However, this technique may have some technique limitations for Case 2a. First of all, the limited observed values and their trend corresponding to the return period is dramatically increasing, which makes it hard to estimate the maximum mean magnitude when the return period is large enough. Due to the small sample size, a non-significant p-value of runs test could not detect the residuals autocorrelation issues for our non-linear regression model. The power for our residual correlation is only 0.52. If we want to achieve our power to 0.8, the sample size of residuals must be at least 22. Besides, the confidence and prediction intervals for a return period after 2 500 years has wider bands than those intervals for return period before 2 500 years. The small sample size with a range of magnitudes is tabulated from long-term return periods, which demonstrates a larger uncertainty of a model. In addition, compared with the log-transformed linear regression for Case 1, exponential non-linear regression has larger predictions for the long-term return period, but has similar predictions for the short-term return period.

Case 2b

For Case 2b, a simple linear regression with the Cochrane-Orcutt Procedure would be considered. The estimated equation is:

$$Magnitude_i = 0.586522 + 0.000339738(retur\ period_i)$$

The Cochrane-Orcutt Procedure could adjust the serial correlation of errors for a linear regression model. The residuals generated from the simple linear regression is first-order autoregressive structure based on the following two graphs, which could be written as $\epsilon_t = \rho\epsilon_{t-1} + \epsilon_t$, $|\rho| < 1$, and the model could be transformed as $y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + \epsilon_t$.

Figure D.10. Series redis(mod) [1:26]

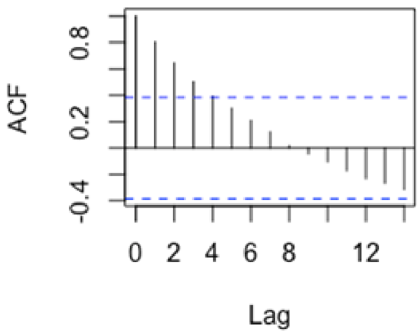
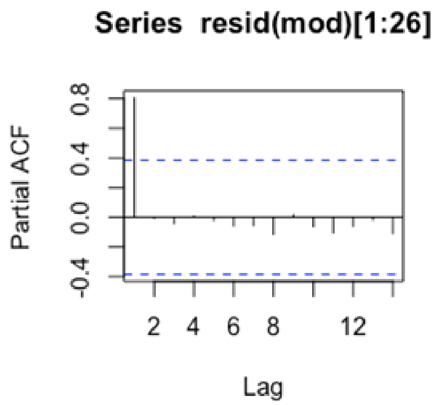
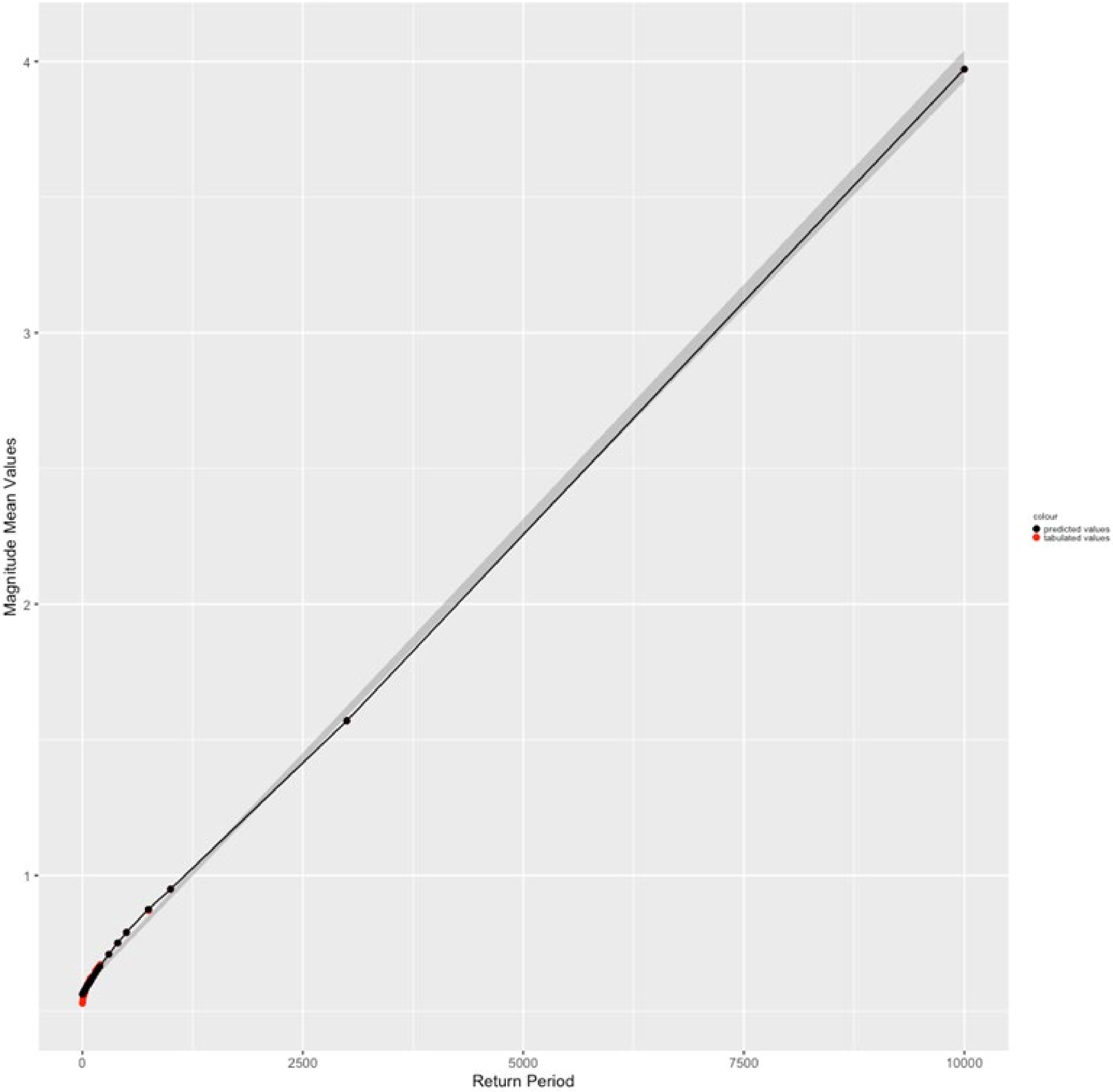


Figure D.11. Series redis(mod) [1:26]



The coefficient of determination is 0.9994, which means 99.94% of the uncertainty in the predicted values can be explained by the linear regression model. Figure D.12 shows the tabulated/predicted data and the regression model. It seems like a few points do not fall into the 95% confidence band.

Figure D.12. Return period vs mean magnitude



The following figures are diagnostic plots for the regression model.

The first plot is residuals vs. fitted values. In this plot, it seems like the line has some patterns shown below. However, each of corresponding residuals from each fitted value only changes a very small range, which means no large variance between the fitted values and tabulated values. The second graph, the normal Q-Q plot, noticed that the first and the last residuals do not follow the referenced dashed line. The last figure is checking the residuals correlation assumption. There is no trend shown on this graph.

Figure D.13. Residuals versus fitted

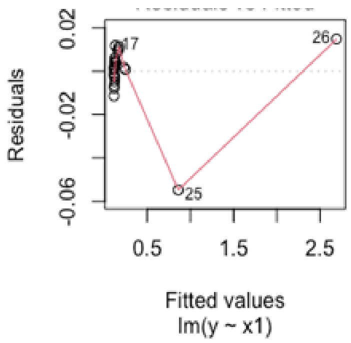


Figure D.14. Normal Q-Q

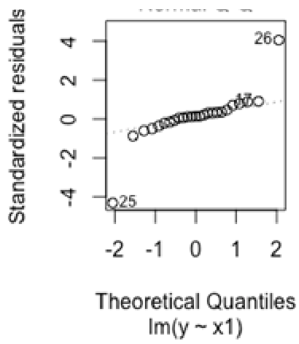
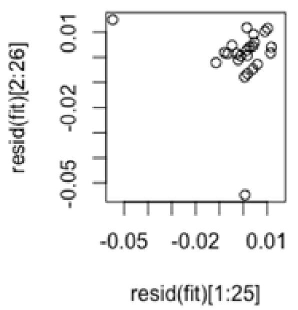


Figure D.15. Autocorrelation



Results

The following table shows the magnitude prediction for Case 2b for the return period of:

Table D.3. Resulting magnitude prediction for Case 2b

Return Period (years)	500	5 000	50 000	500 000
Magnitude (metres)	0.75	2.28	17.57	170.46

Results of assessment

Although this simple regression with the Cochrane-Orcutt Procedure solved the residuals correlation issue, the prediction of magnitude has larger differences for long-term periods compared with other cases, and the normality assumption is violated due to the first and the last residuals points. If the sample size is small, the Type I error rates will not be far from the target significance level, but the non-normality residuals would generate the inaccurate prediction interval. Besides, the model could not obtain an ideal and stabilised estimated magnitude for a long time period. If we could have more observations, especially for long time-period, we could probably estimate a model for ideal conditions.

Case 2c

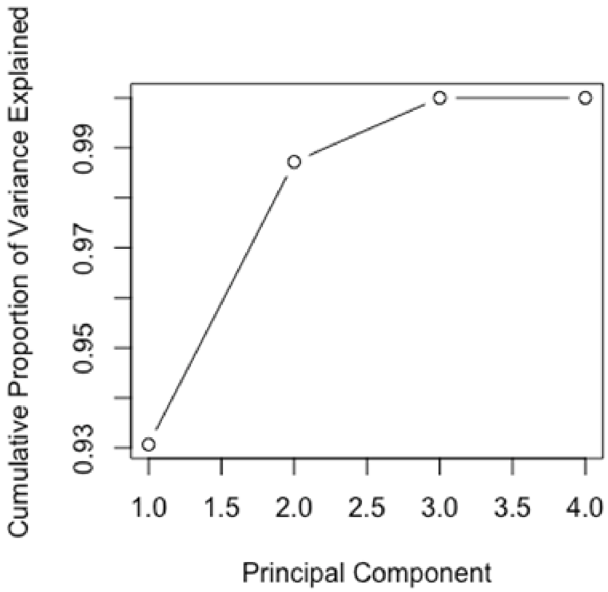
The hazard frequency/magnitude model here for Case 2c has the estimation of the uncertainty on the magnitude for long time intervals. However, the uncertainty of short time intervals has not been provided for Case 2c. Therefore, the first step is using simple linear regression to simulate those unknown uncertainties for short time intervals. Based on the known mean magnitude, the standard deviation, 5th percentile and 95th percentile values for long time intervals used three different linear regression models to estimate the coefficient of relationship between the mean magnitude and those uncertainty characteristics. A complete table is shown below:

Table D.4. Benchmark data provided for Case 2-3

Return Period (years)		1	2	5	10	50	100	500	1 000	3 000	10 000
Magnitude (metres)	Mean	053	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4.0
	SDev	0.0067	0.00677	0.00808	0.00939	0.01462	0.01854	0.04	0.06	0.15	0.46
	5th	0.508	0.5082	0.5160	0.5237	0.5546	0.57787	0.72	0.85	1.3	3.2
	95th	0.5564	0.5564	0.5683	0.5803	0.6279	0.66376	0.85	1.1	1.8	4.7

To create the predictive model based on the return period and the uncertainty of mean magnitude characteristics, the standard deviation, and the 5th and 95th percentile values, the multicollinearity of those predictors is a big issue. A principal component analysis is to find a new linear combination of those variables that contains much information by looking at the highest variance. The new variables derived by the principal component analysis are orthogonal, where the first axis contains the most information, and the second axis contains the second most information, and so forth. The following plot shows that only one component explains around 93% variance, and two components explain around 99% variance in the data set.

Figure D.16. Principal component analysis results



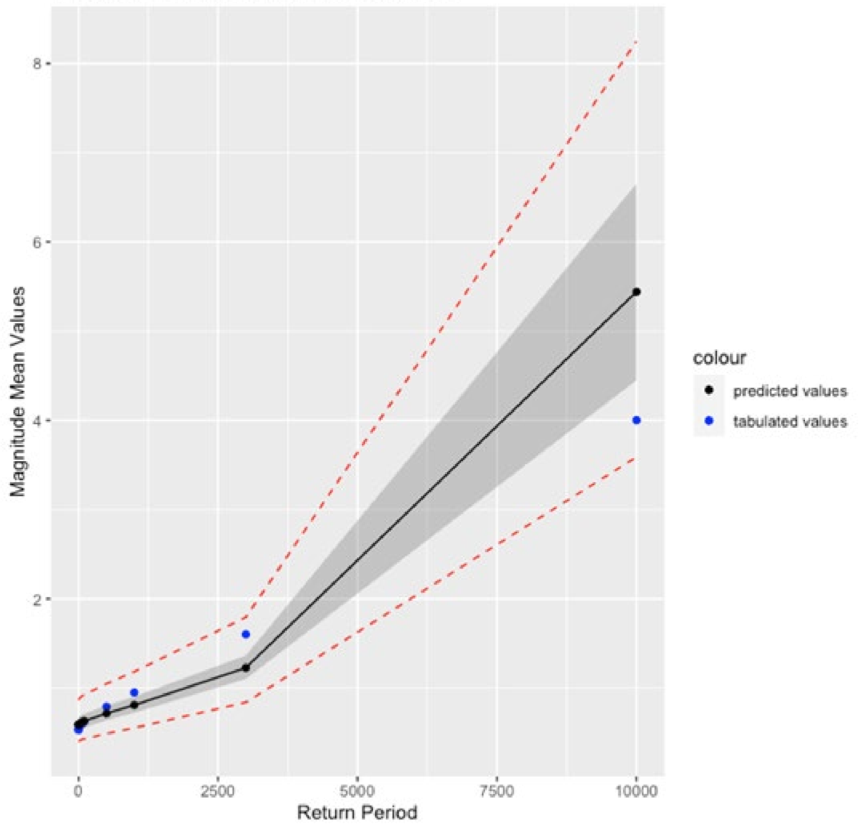
The next step is to regress the response variable mean magnitude in a set of principal components obtained from the return period, standard deviation, and 5th and 95th percentiles values of the mean magnitude. The principal component regression with a few components often could explain most of the variability of all predictors and consequently with a relationship with outcome interest. The first component from the principal component analysis would be considered to be used in the regression model, and a polynomial term of the first component would be added in order to meet the normality assumption. The principal components regression model for Case 2c has the following form:

$$\log\,(\widehat{Magnıtdue_i}) = 1.04862 + 1.56278PC_1 - 0.18800{PC_1}^2$$

The coefficient of determination $R^2 = 0.9862 R^2 = 0.9996$, which means 98.62% of the uncertainty in the predicted values can be explained by the squared regression model. Besides, F-statistics (428.7) and significant p-value (<0.05) points out the model provides better fit than the null model.

Figure D.17 shows the tabulated/predicted data and the regression model. Predicted values for long time periods could not fall into the 95% confidence interval but fall into the prediction interval.

Figure D.17. Return period versus mean magnitude



Since principal component regression still uses the linear regression to predict the response variable by using the components as predictors, residuals vs. fitted and normal Q-Q plots could be utilised to check the residuals' homoscedasticity and normality assumption. Another Durbin Watson test would check the measurement of autocorrelation in residuals from regression model. The test statistic is $DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$, where the e_t are residuals. Diagnostic plots for the regression model are shown below:

Figure D.18. Residuals versus fitted

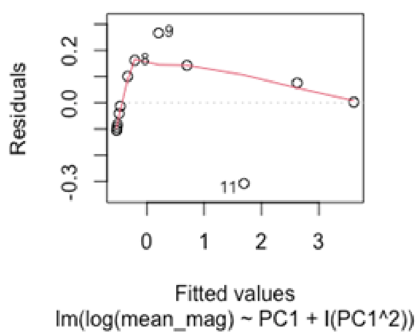
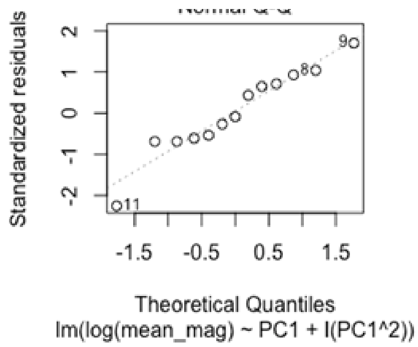


Figure D.19. Normal Q-Q



For the residuals vs. fitted values plot, it seems like the line has some patterns shown between fitted values and residuals. However, residual values do not change a lot between each other, which means there are small variances between the fitted values and tabulated values. For the normal Q-Q plot, a few residuals points do not completely follow the referenced dashed line. Using the Shapiro-Wilk test again to ensure the normality assumption test, a p-value with 0.18 (>0.05) suggests that the null hypothesis of normality assumption could not be rejected. Besides, the Durbin Watson test with p-value 0.09 also indicates the autocorrelation assumption is not violated.

Results

The following table shows the magnitude prediction for Case 2-3 for return periods of:

Table D.5. Resulting magnitude prediction for Case 2-3

Return Period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1	0.71	2.01	13.75	37.09

Results of assessment

This principal component regression has a few technical limitations and uncertainties. First, the uncertainty of mean magnitude for a short time interval has not been provided in the table, even though the unknown standard deviation of each mean magnitude is small. In order to estimate the standard deviation, the 5th and 9th percentile of mean magnitude, an imputation has to utilised, which produced some uncertainty for those imputed values. To co-operate that uncertainty and return period, we used the principal component regression that only one component of those combinations represents 93% of variances in the dataset, inducing larger uncertainty in our model. Compared with the results estimation obtained from Case 2a, we noticed that the confidence and prediction intervals for Case 2c are all wider than those for Case 2a. Model fitting also does not perform well when there is a long time return period so that the predicted values do not fall into the 95% confidence interval. Besides, this principal component regression model could not meet the expected condition that the mean magnitude would keep increasing as the return period increases.

Annex E. Submission by IRSN

E.1.1. Case 1 – Known model producing the synthetic data

Table 1. Synthetic data for Case 1

t	M
1	0.50
2	0.65
5	0.85
10	1.00
50	1.40
100	1.50
500	1.90
1000	2.00
2000	2.20
10000	2.50

The synthetic model (SM-Case1):

$$M = 0.5 + 0.5 \times \log_{10}(a \times t)$$

The synthetic model serves as surrogate for a complex phenomenological process.

Inputs:

Return time intervals (or periods) t;

Outputs:

Magnitude M for an annual maxima (AM) event. Variable a equal to 1 in cas1: $M = 0.5 + 0.5 \times \log_{10}(t)$).

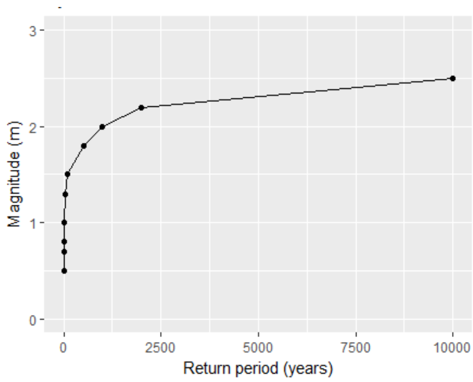


Figure 1. **Hazard curve (SM-Case1).**

Benchmark completion: Provide, using the data for Case 1, a model that best describes the frequency/magnitude relationship and the associated analysis and insights.

The randomness, homogeneity and stationarity of data are necessary conditions to conduct a frequency analysis. As a magnitude M produced by the synthetic model is supposed to be for annual maxima (AM) events (mean rate of events $\lambda = 1$, this parameter gives the number of events per year), the hypothetical data set should be independent. They should also be homogeneous since the magnitudes are produced by one synthetic model (the same statistical population). The sample size is a prerequisite for a frequency analysis, as well. This last-mentioned condition can be very easily satisfied since we know the synthetic model producing the data (we can produce as much data as necessary).

Assumption 1: The magnitudes are assumed to be stationary, independent and homogeneous.

As mentioned earlier, a magnitude M is produced by the synthetic model with a rate of events equal to one ($\lambda = 1$). Return periods can then be estimated with one of the two following frequency models:

1. An annual maxima (AM) frequency model in which the distribution of the AM events converges to a GEV one.
2. A Peaks-Over-Threshold (POT) model in which the distribution of the exceedances over the threshold u converges to an exponential one (of parameter θ).

Let x_T be the event with the return period T and F the non-exceedance probability function. In its general definition, x_T corresponds to the quantile of probability of exceedance equal to $1/\lambda T$ (i.e. equal to $1/T$ when $\lambda=1$): $\Pr(X \geq x_T) = \overline{F}(x_T) = 1/\lambda T$ ($\overline{F}$ is the survival function).

In the 1st case (SM-Case1), a set of 100 return periods is randomly sampled. Associated magnitudes are calculated with the synthetic model and a classic frequency estimation is then performed with a GEV distribution. To take this issue one step further, a simple analogy of the synthetic model with GEV and GP distribution functions is performed.

1. AM/GEV frequency model:

The GEV distribution is the limiting distribution for the AM of independent and identically distributed (iid) random variables.

$$F(x) = \begin{cases} e^{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}} & \xi \neq 0 \\ e^{-e^{-(x-\mu)/\sigma}} & \xi = 0 \end{cases}$$

Where μ, $\sigma > 0$ and ξ are the location, scale and shape parameters, respectively. A sample of 100 magnitudes (AM events with an effective duration equal to 100 years) is generated and analysed. The hypothetical observational data set (from a synthetic model) is selected in such a way that their empirical distribution looks as natural as possible (close to the natural behaviour and variability of external hazards such as floods, extreme temperatures or high winds, etc.). More concretely, the sampling is done as follows: the R “sample” function is used to randomly sample 100 values of “return times” t from 1 to 100 years. In all, 95 values have empirical return periods less than 100 years. It is worth noting that the case with aleatory sampling in a period of 10 000 years is not realistic and cannot represent natural hazards.

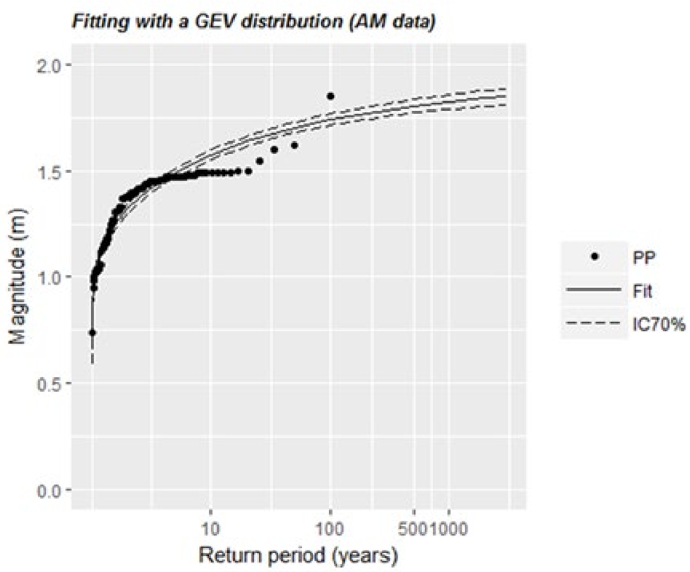
```
library("evd")
library("ggplot2")
# Hypothetical data from synthetic model
# A sample of 100 AM in which the 100-year return level was not exceeded
w <- 100
tt1 <- sample(1:100, size=w, replace=TRUE);
xt <- round(0.5+0.5*(log10(tt1)),2)
# Plotting positions - empirical probabilities (Weibul)
Px <- 1:length(xt)/(length(xt)+1)
# associated empirical return periods
Tx <- 1/(1-Px)
# Theoretical fitting
fit.GEV <- fevd(xt, type="GEV")
# Estimated parameters: # location  scale   shape
#               1.2589728  0.1924330 -0.2943914
ttic <- c(1.25, 1.5, 2, 3, 5, 10, 20, 50, 100, 500, 1000, 2000, 3000)
RLs <- ci(fit.GEV, type = "return.level", return.period=ttic, method =
"normal", alpha=0.30)
```

Plot the fitting . . .

```
data1 <- data.frame(xx = Tx, yy = sort(xt))
data2 <- data.frame(xxx = ttic, yyy = RLs [,2])
data3 <- data.frame(xxxx = ttic, lb = RLs [,1])
data4 <- data.frame(xxxx = ttic, ub = RLs [,3])
sp <- ggplot(data.frame("pp"=NULL, "tf"=NULL))
sp + geom_point(aes(x=data1$xx, y=data1$yy, colour="PPcol" )) +
geom_line(aes(data2$xxx, data2$yyy,colour="Fitcol",linetype="Fitline1")) +
geom_line(aes(data3$xxxx, data3$lb, colour="ICcol", linetype="Fitline2"))+
geom_line(aes(data4$xxxx,data4$ub,colour = "ICcol",linetype="Fitline2")) +
scale_color_manual("",labels   =    c("PP",    "Fit",    "IC70%"),breaks=c("PPcol",
               "Fitcol","ICcol"),values=c("black","black","black")) +
scale_linetype_manual("",breaks = c("fitline1","fitline2"),
                           values = c("solid","longdash")) +
guides(fill = guide_legend(order = 1), colour = guide_legend(order = 2, keywidth =
        2.5,  override.aes  =  list(linetype  =   c("blank",  "solid",  "longdash"),
        shape=c(16,NA,NA)))) +
scale_x_log10(breaks=c(10,  100,  500,  1000),  labels=as.character(c(10,  100,  500,
               1000)), name = "Return period (years)") +
scale_y_continuous(name = "Magnitude (m)", limits = c(0,2)) +
ggtitle("Fitting with a GEV distribution (AM data)") +
theme(
plot.title = element_text(color = "Black", size=10, face = "bold.italic"))
```

The fitting is bounded to a final value equal to $\mu - \sigma/\xi$ (equal to 1.91 m). The 500 000-year return level is equal to this end value (1.90 m) which is much lower than the 3.35 m calculated with the synthetic model for the same return period. Paradoxically, it can be seen from Figure E.1 that the adequacy of the GEV distribution is quite good and the uncertainty (in term of confidence intervals) is very low for high return levels.

Figure E.1. Fitting with a GEV distribution with the SM-Case1 data sets ($\mu = 1.26; \sigma = 0.19; \xi = -0.29$**)**



The GEV distribution combines three distributions, identified by Fisher and Tippet (1928) into a single form. Indeed, depending on the value of the shape parameter ξ , the GEV can take the form of the Gumbel ($\xi = 0$), Fréchet ($\xi > 0$) or the Reverse Weibull distributions ($\xi < 0$). On the other hand, and as mentioned earlier in this section, the scale parameter of the GEV distribution should be positive ($\sigma > 0$). The probability and quantile functions for the GEV distribution can be written as:

$$\begin{cases} F(x) = \begin{cases} e^{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}} & \xi \neq 0 \\ e^{-e^{-(x-\mu)/\sigma}} & \xi = 0 \end{cases} \\ x_F = \mu - \dfrac{\sigma}{\xi}\left[1 - \left\{-\log(1-F)\right\}^{-\xi}\right] \end{cases}$$

A linear relationship between magnitudes M and the logarithm of return periods t could be obtained if the x-axis was plotted in a logarithmic scale. Note that this linear relationship cannot be obtained with all the asymptotic extreme value distributions and only the exponential form (the Gumbel distribution) might be suitable (but with the limitation of bad fitting of small magnitudes associated to small return periods up to 50-100 years).

As can be seen from the probability equation, the GEV distribution can as well have the form of the synthetic model when the shape parameter is equal to -1. With such a value, the theoretical upper tail can only be finite and bounded (as may be useful for estimates of specific cases of extreme values which may have an upper bound, as is the case here). This hypothesis is in line with the bounded theoretical fitting presented in Figure E.1.

$$x_F = \mu + \sigma\left[1 + \log(1-F)\right] \rightarrow x_T = \mu + \sigma + \sigma\log\left(\frac{1}{T}\right) \rightarrow x_T = \mu + \sigma - \sigma\log(T)$$

$$x_T = \mu + \sigma - \sigma\log(10) \times \log_{10}(T)$$

This last equation is similar to the proposed synthetic model as follows:

Frequency model $\longrightarrow$ $x_T = \underbrace{\mu + \sigma}_{} - \underbrace{\sigma\log(10)}_{} \times \log_{10}(T)$

Synthetic model $\longrightarrow$ $M = 0.5 + 0.5 \times \log_{10}(t)$

$$\Rightarrow \begin{cases} \mu + \sigma = 0.5 \\ \sigma\log(10) = -0.5 \end{cases}$$

It can be concluded from these two conditions that, with a shape parameter equal to -1, the scale parameter can only be negative. As just noted, a GEV distribution obviously cannot be used with a negative scale parameter: then it cannot describe the frequency/magnitude relationship for the SM-Case1.

2. The POT GPD/Exponential frequency model:

The GP distribution for the return levels (with an exponential distribution for the exceedances over the threshold u). The exceedances over the threshold u follow an exponential distribution of parameter ρ .

$$F_u(X) = 1 - \exp\left[-\rho(x-u)\right]$$

$$x_T - u = -\frac{1}{\rho}\log(1-F) \rightarrow x_T = u - \frac{1}{\rho}\log\left(\frac{1}{\lambda T}\right) \rightarrow x_T = u + \frac{1}{\rho}\log(\lambda T)$$

Therefore, the $T-year$ return level can be written as:

$$x_T = u + \frac{1}{\rho}\ln(\lambda) + \frac{1}{\rho}\log(10)\log_{10}(T)$$

This last equation is similar to the proposed synthetic model as follows:

Frequency model → $x_T = \underbrace{u + \frac{1}{\rho}\ln(\lambda)}_{\uparrow} + \underbrace{\frac{1}{\rho}\log(10)}_{\uparrow} \times \log_{10}(T)$

Synthetic model → $M = 0.5 + 0.5 \times \log_{10}(t)$

$$\Rightarrow \begin{cases} u + \frac{1}{\rho}\ln(\lambda) = 0.5 \\ \frac{1}{\rho}\log(10) = 0.5 \end{cases}$$

Considering the case in which the rate of events is equal to one ($\lambda = 1$), the first condition gives a threshold $u = 0.5$. It can also be easily concluded from the second condition that $\rho = \log(10)/0.5 = 4.605$.

> **Assumption 2:** The number of events is randomly sampled following a Poisson Process. It is also assumed that the rate of events is equal to 1.

In the first step, the developed frequency model is used to describe the frequency/magnitude relationship. Magnitudes are then sampled on a period of time w, say 100 years.

```
set.seed(1248)
# effective duration
w <- 100
# Threshold
u <- 0.5
# annual rate of events
lambda <- 1
# Parameter of the exponential distribution
rho <- log(10)/0.5
# Assuming the Nbre of arrivals happen randomly following a Poisson Process
N <- rpois(n = 1, lambda = lambda * w)
# Return levels (exceedances follow an exponential distribution)
xt <- u + rexp(n = N, rate = rho)
```

In the next step, the fitting with the theoretical distribution (exponential) is performed. The fitting of the SM-Case1 data set (w=100 years) with a GPD distribution ($u = 0.5$ & $\lambda = 1$) for the return levels (the exceedances over u are exponential) is presented in Figure E.2. The results are also shown in Table E.1. A good and visually adequate fitting is obtained. Indeed, all the observed probabilities are in the 70% confidence interval. As shown in Table E.1, the 500 000-year return level is equal to 3.19 m which is close to the 3.35 m calculated by the synthetic model for the same return period. On the other hand, the uncertainty (in term of confidence intervals) is reasonably low for high return levels. All the simulations were carried out within the R environment (open-source software for statistical computing: http://www.r-project.org/). The Renext library (IRSN and Alpstat, 2013), developed by the French Institute for Radiological Protection and Nuclear Safety, was used for frequency estimations. The Renext package was specifically developed for flood frequency analyses using the POT method.

```
library("Renext")
library("ggplot2")
fit.exp <- Renouv(xt, effDuration = w,                    distname.y =
                  "exponential",                          threshold = u,
                  Tlim= c(1, 10000),                      pct.conf = 70,
                  plot = TRUE)
```

Figure E.2. Fitting of the SM-Case1 data sets (w=100 years) with GPD distribution (u = 0.5 & λ = 1) for the return levels (the exceedances over u are exponential).

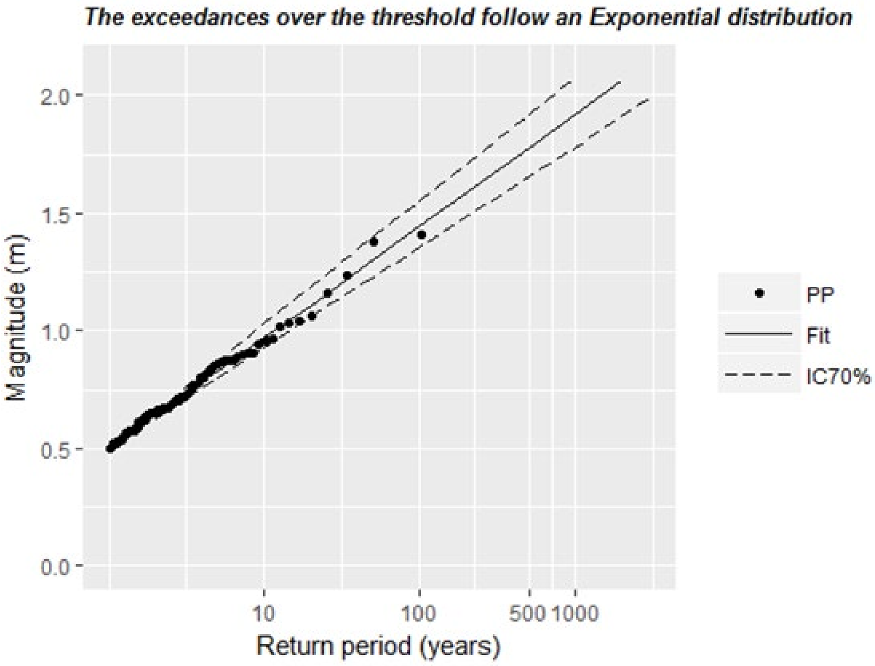


Table E.1. Comparison of magnitudes (m) calculated with the synthetic data for Case 1. The values in brackets correspond to the 70% confidence intervals.

Return Period (years)	500	5 000	50 000	500 000
Synthetic data for Case 1	**1.85**	**2.35**	**2.85**	**3.35**
POT model (GPD/Exp)	1.78 (1.66-1.92)	2.25 (2.08-2.45)	2.72 (2.51-2.98)	3.19 (2.94-3.50)

3. Case 1 – Conclusion

In this first case, both the hypothetical data and the synthetic model are known. It can be concluded that the GEV distribution is not very suitable to describe all the frequency/magnitude relationships for the SM-Case1 data. These data are best described with a POT frequency model in which the distribution of the return levels converges to a GPD (threshold equal to 0.5 and rate of events equal to 1) with an exponential behaviour of the exceedances over the threshold. Indeed, the desired high return levels are estimated with this frequency model and compared to the hypothetical observational data from the synthetic model. The relative difference in magnitudes did not exceed 5% (7 cm, 10 cm, 13 cm and 16 cm for the 500-, 5 000-, 50 000- and 500 000 year return levels, respectively). Moreover, all the plotting positions are inside the confidence interval despite the fact that the latter is very narrow (low uncertainty).

E1.2. Case 2 – Unknown model producing the synthetic data

For Case 2, there is no synthetic model provided. Furthermore, three parts to this example are provided.

Case 2(a): provides the synthetic data (10 data points) with no uncertainty provided;

Case 2(b): provides additional synthetic data (26 data points) with no uncertainty provided;

Case 2(c): provides the synthetic data (10 data points) with uncertainty estimates on some of the data.

We need to provide, using the data for Case 2, a model that best describes the frequency/magnitude relationship and the associated analysis and insights.

Non-linear least-squares estimates of the distribution parameters are performed using the “nls” R function. Confidence intervals are then calculated and plotted. The same assumptions are used for this second case.

Assumption 1: The magnitudes are assumed to be stationary, independent and homogeneous.

Only parts 2a and 2b are evaluated herein. As in Case 1, return periods can then be estimated with one of the two following frequency models:

1. The AM/GEV frequency model: Annual maxima (AM) frequency model in which the distribution of the AM events converges to a GEV one.
2. POT/GPD frequency model: Peaks-Over-Threshold (POT) model in which the distribution of the exceedances converge to a GPD.

For both Cases 2a and 2b, non-linear least-squares estimates of the GEV and GPD parameters are performed. It is worth noting that the GPD must give almost the same parameters and fitting (results not presented hereafter). The fitting with confidence intervals is presented in Figure E.5 and Figure E.8. The adequacy of the theoretical distribution in both Cases 2a and 2b is visually quite good with heavy tails (very high shape parameter $\xi \approx 0.96$). A comparison of T-year return levels (corresponding to 500-, 5 000-, 50 000- and 500 000-year return periods) for both Case 2a and Case 2b are presented in Table E.2 (the values in brackets correspond to the absolute and relative widths of the 70% confidence intervals). The results in Table E.2 indicate that the relative widths of confidence intervals in Case 2b (with 26 synthetic data points) are 1.5 times narrower than those obtained in Case 2a (with 10 synthetic data points) for the GEV distribution.

Table E.2. The T-year quantiles with absolute and relative widths of their 70% confidence intervals

Return Period (years)	500	5 000	50 000	500 000
GEV for Case 2-a	0.75 (0.68-0.84) (21.3%)	2.32 (1.72-3.21) (64.2%)	16.79 (10.75-26.44) (93.4%)	150.34 (89.35-254.16) (109.6%)
GEV for Case 2-b	0.76 (0.71-0.82) (14.5%)	2.31 (1.88-2.87) (42.8%)	16.50 (12.17-22.46) (62.4%)	146.53 (102.70-209.56) (73.0%)

Cases 2a and 2b – Conclusion

In these two cases, only the synthetic data are known (no synthetic model provided). It can be concluded that both the MA/GEV and POT/GP distributions best describe the magnitude/frequency relationship with heavy tails (very high shape parameter, then there is no physical limit!). On the other hand, as more data are provided in Case 2b, the confidence intervals are narrower (reduced uncertainty).

It is obvious that, with such a high shape parameter, these data reflect what can be observed for natural hazards up to a return period of 1 000 - 10 000 years. Beyond this, return magnitudes increase much more quickly to very high values. This is the main characteristic of models with very heavy tails (as is the case here).

Case 2a: the MA/GEV frequency model:

10 synthetic data points are provided with no uncertainty provided;

```
#  Benchmark  on  External  Events  Hazard  Frequency  and  Magnitude  #
Statistical Modelling
# Case 2 – Unknown Model Producing the Synthetic Data
library("evd")
library("ggplot2")
# Case 2-a synthetic data
t <-c(1,2,5,10,50,100,500,1000,3000,10000)
M <- c(0.53,0.53,0.54,0.55,0.59,0.62,0.79,0.95,      1.60,4.00)
# Plot data - Hazards curve
qplot(t, M, geom=c("point", "line"), xlim = c(0, 10000), ylim = c(0.5, 4),
       main = "Data  (with  no  synthetic  model  provided)  -  Cas2-a",  xlab  =
       "Return period (years)",ylab = "Magnitude (m)")
```

Table E.3. Synthetic data for Case 2a

t	M
1	0.53
2	0.53
5	0.54
10	0.55
50	0.59
100	0.62
500	0.79
1 000	0.95
3 000	1.57
10 000	3.97

Figure E.3. Hazard curve (Case 2a synthetic data)

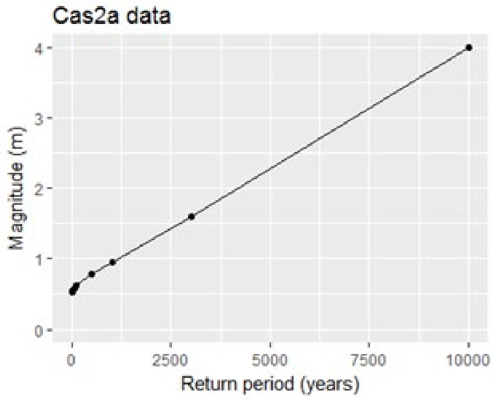
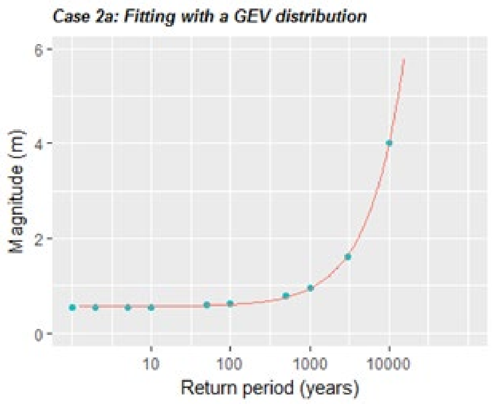
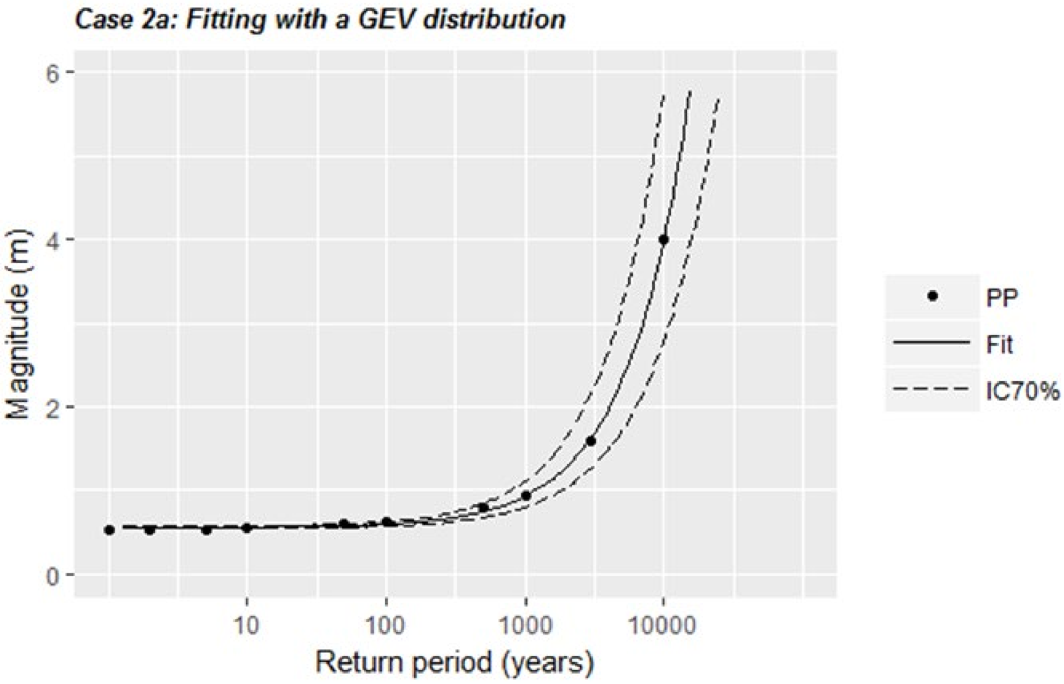


Figure E.4. Fitting Case 2a synthetic data




```
df1 <- data.frame(t = t, xt = M)
# Fitting with a Non-linear least-squares estimates of the GEV parameters
fit <- nls(xt ~ qgev(1-1/t, loc = mu, scale = sigma, shape = xi), data=df1[-1,], start
           = list(mu = 0.12, sigma = 0.2, xi = 1))
coef(fit)
# mu       sigma     xi
# 0.5593335970 0.0004576594 0.9649876346
# Plot fitting (code not presented herein)
```

Figure E.5. Fitting Case 2a synthetic data (GEV ($\mu = 0.5593; \sigma = 0.0005; \xi = 0.9650$))



Case 2b: the MA/GEV frequency model:

26 synthetic data points are provided with no uncertainty provided;

```
library("evd")
library("ggplot2")
# Case 2-b synthetic data
t <-c(1,2,5,10,15,20, . . . ,750,1000,3000,10000)
M <- c(0.53,0.53,0.54,0.55, . . . ,0.95,1.57,3.97)
# Plot data - Hazards curve
qplot(t, M, geom=c("point", "line"), xlim = c(0, 10000), ylim =
      c(0.5, 4), main = "Data (with no synthetic model provided)
      - Cas2-b", xlab = "Return period (years)",ylab =
      "Magnitude (m)")
```

Table E.4. Synthetic data for Case 2b

t	M
1	0.53
2	0.53
5	0.54
10	0.55
15	0.56
20	0.56
25	0.57
30	0.57
40	0.58
50	0.59
60	0.60
70	0.60
80	0.61
90	0.62
100	0.62
125	0.63
150	0.65
175	0.66
200	0.67
300	0.71
400	0.75
500	0.79
750	0.87
1 000	0.95
3 000	1.57
10 000	3.97

Figure E.6. Hazard curve (Case 2b synthetic data)

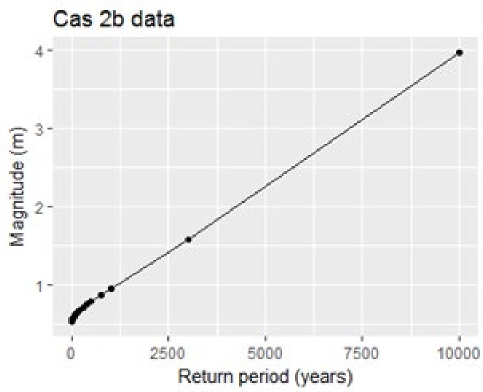
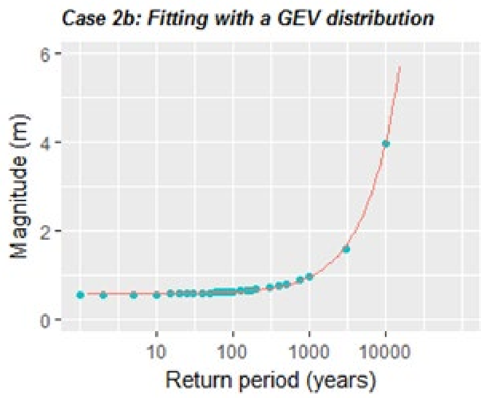
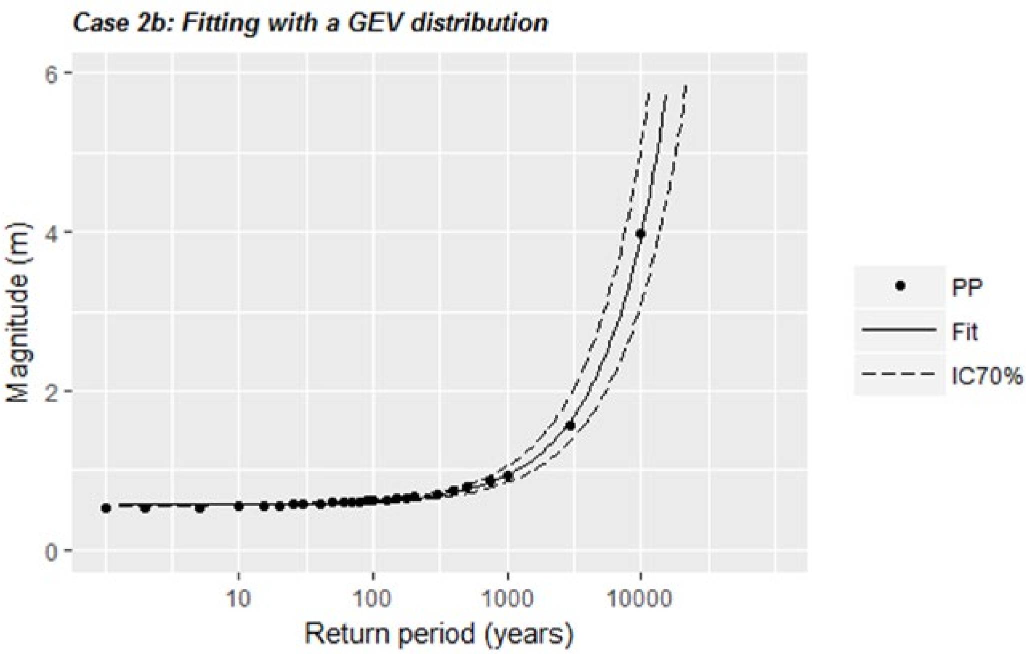


Figure E.7. Fitting the Case 2b synthetic data




```
df1 <- data.frame(t = t, xt = M)
# Fitting
fit <- nls(xt ~ qgev(1 - 1/t, loc = mu, scale = sigma, shape = xi),
       data = df1[-1, ],
       start = list(mu = 0.12, sigma = 0.2, xi = 1))
coef(fit)
# mu       sigma     xi
# 0.5707180614 0.0004624945 0.9619867644
# Plot fitting (code ggplot not presented herein)
```

Figure E.8. Fitting Case 2a synthetic data with a GEV distribution ($\mu = 0.5709; \sigma = 0.0005; \xi = 0.9619$ **)**



Annex F. Submission by KAERI

Introduction

This benchmark study aims to apply statistical modelling for frequency and magnitude estimation based on data for external event hazard assessment. Based on the results of this study, it is believed that an approach to the quantification of external event IEs can be formulated and evaluated through the application of an effective statistical model. In this study, analysis was based on two cases that considered benchmarks provided by the OECD NEA. Each case was given a magnitude according to the return period. Based on this data, an appropriate statistical model was applied through regression analysis for each case. Based on the results, the magnitudes of 500, 5 000, 50 000, and 500 000 years were predicted and presented.

Synthetic data analysis

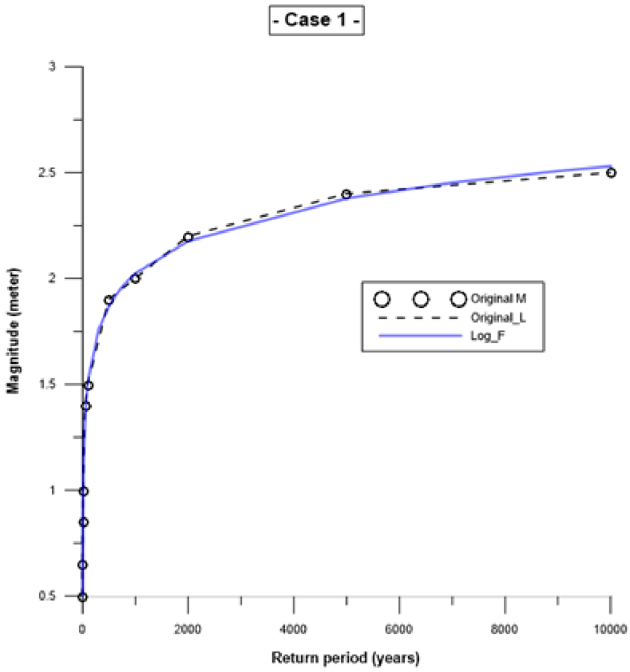
- Study for Case 1

Table F.1 shows the synthetic data for Case 1 as provided by the OECD NEA. Using regression analysis, log regression showed appropriate fitting results for the relationship between magnitude and return period, as shown in Figure F.1.

Table F.1. Synthetic data for Case 1

Return period (years)	1	2	5	10	50	100	500	1 000	2 000	10 000
Original M (metres)	0.50	0.65	0.85	1.00	1.40	1.50	1.90	2.00	2.20	2.50

Figure F.1. Log regression fitting



As a result of the regression analysis on the magnitude of Case 1, the log regression equation including variables A and B is shown in Eq. (1). The magnitude of the return period from 1 to 10 000 years was estimated by Eq. (1); Table E.2 compares the values to the original magnitude values proposed by the OECD NEA.

Case 1: $M = 0.2199 * ln\,(x) + 0.5034\ (parameters\ A\ and\ B)$ (1)

Table F.2. Regression results for Case 1

Return period (years)	1	2	5	10	50	100	500	1 000	2 000	10 000
Original M (metres)	0.50	0.65	0.85	1.00	1.40	1.50	1.90	2.00	2.20	2.50
Log Magnitude (metres)	0.503	0.656	0.857	1.010	1.364	1.516	1.870	2.022	2.175	2.529

An error analysis was then performed using the SUMXMY2 function to verify the statistical justification of the estimated magnitudes. The SUMXMY2 function squares and sums the difference between two corresponding values, and therefore, the closer to 0, the smaller the error between the two variables, and the more statistically valid the estimation can be considered. The SUMXMY2 function is expressed as Eq. (2).

$SUMXMY2 = \sum(x - y)^2$ (2)

Here, x is the value of the original magnitude, and y is the value of the estimated magnitude. As a result of error analysis using the SUMXMY2 function, the sum square error (SSE) value was calculated as 0.005, which is very close to zero. It was therefore judged that the estimated magnitude values were very similar to the original values and valid. However, to minimise SSE and more precisely estimate the magnitude values, a solver function was used. The target of the SSE value was set to 0, and an optimisation analysis was performed on parameters A and B of Eq. (1). The results are shown in Table F.3.

Table F.3. Optimisation for parameters

Parameter	Original	SSE_Solver
A	0.2199	0.219861834
B	0.5034	0.503363549

These optimised parameters are then used in the regression equation for Case 1, as shown in Eq. (3) below. Table F.4 compares the magnitude values estimated by Eq. (3) with those estimated by Eq. (1) and the original magnitude values.

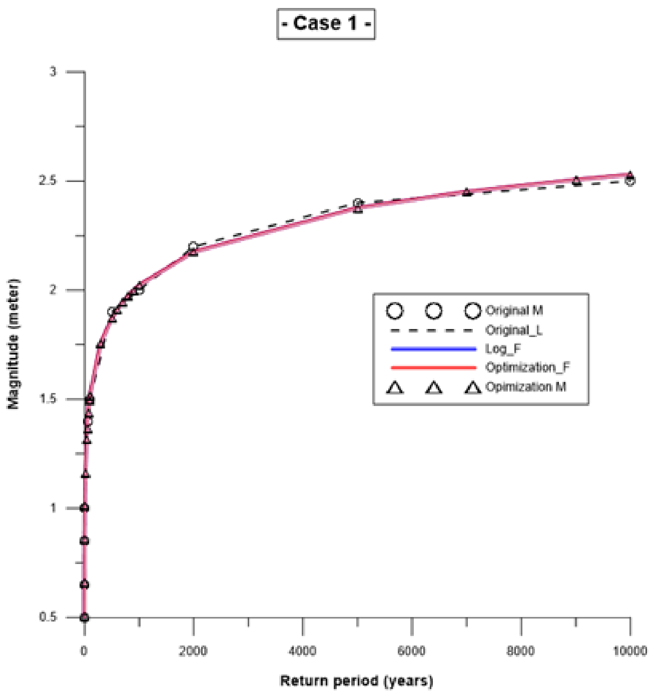
Case 1: $M = 0.219861834 * ln\ (x) + 0.503363549\ (parameters\ A\ and\ B)$ (3)

Table F.4. Comparison of magnitudes

Return period (years)	**1**	**2**	**5**	**10**	**50**	**100**	**500**	**1 000**	**2 000**	**10 000**
Original M (metres)	0.500	0.650	0.850	1.000	1.400	1.500	1.900	2.000	2.200	2.500
Log Magnitude (metres)	0.503	0.656	0.857	1.010	1.364	1.516	1.870	2.022	2.175	2.529
Optimised magnitude (metres)	0.503	0.656	0.857	1.010	1.363	1.516	1.870	2.022	2.175	2.528

For Case 1, the optimised parameters A and B (Table F.3) were similar to the existing values. Likewise, the SSE value of 0.0049 was also similar to the existing value of 0.005. Fitting was then performed based on the optimised magnitude values; results are shown in Figure F.2.

Figure F.2. Fitting result for Case 1



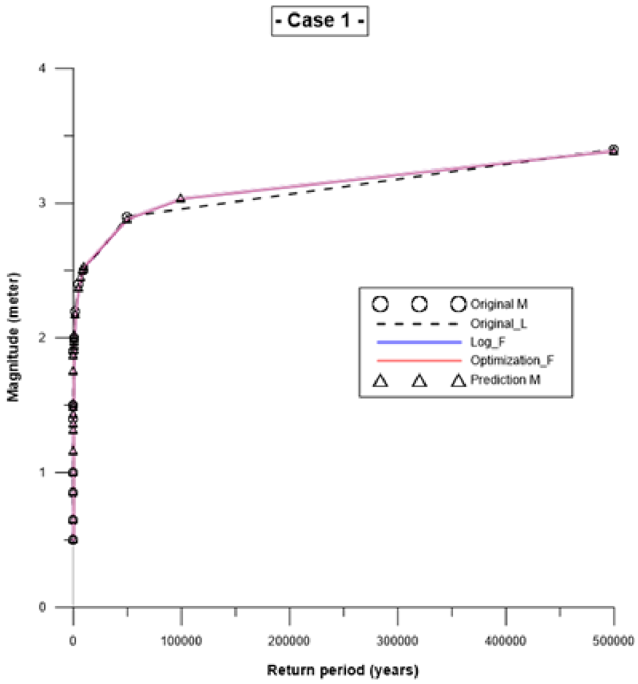
F.1.1. Case 1 result assessment

The results of the optimised fit for Case 1 were estimated to be similar to the size values presented by the OECD NEA. The trend lines calculated from the estimated magnitude values (Figure F.1 and Figure F.2) were also similarly estimated. Therefore, it can be judged that the fitting result for Case 1 in this study is valid. Additionally, based on the regression analysis estimates in Case 1, magnitude values were predicted for 500, 5 000, 50 000, and 500 000 years return period. The results are presented in Table F.5 and Figure F.3.

Table F.5. Magnitude prediction by return period (Case 1)

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 1 Exact	1.9	2.4	2.9	3.4
Log KAERI mean (metres)	1.870	2.376	2.883	3.389
Optimised KAERI mean (metres)	1.870	2.376	2.882	3.388

Figure F.3. Magnitude prediction fitting for Case 1



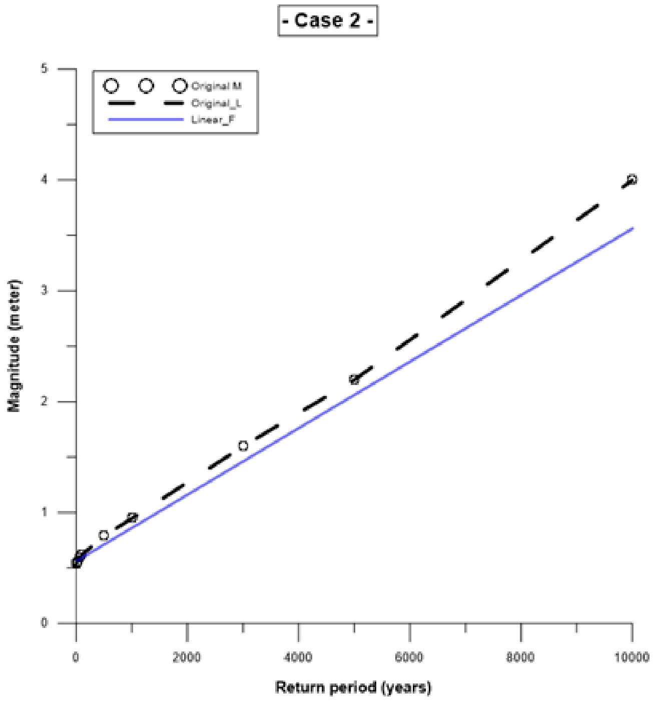
F.1.2. Study for Case 2

Table F.6 shows the synthetic data for Case 2 provided by the OECD NEA. In this case, regression analysis found linear regression to give the best fit between magnitude and return period, as shown in Figure F.4.

Table F.6. Synthetic data for Case 2

Return Period (years)	1	2	5	10	50	100	500	1 000	3 000	10 000
Original M (metres)	0.53	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.60	4.00

Figure F.4. Linear regression fitting



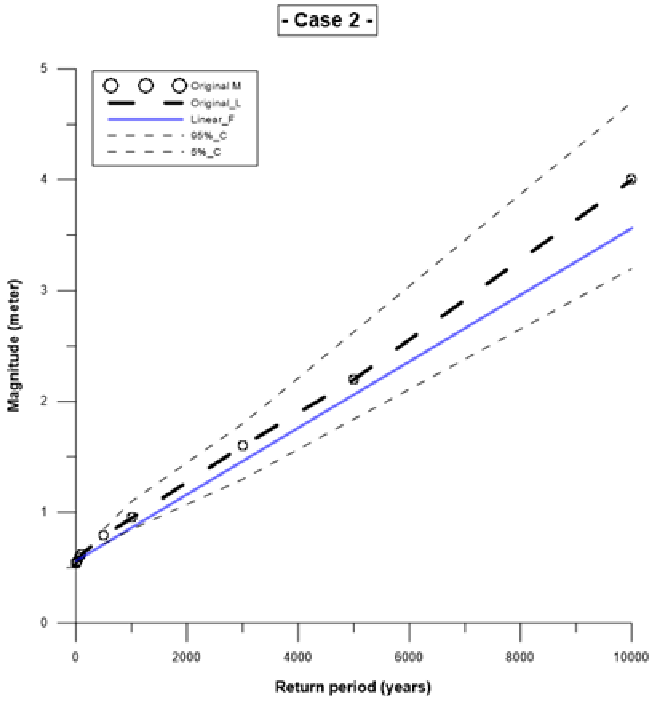
Equation (4) below is the linear regression equation including variables A and B as a result of the regression analysis on the magnitude of Case 2. Magnitudes of return periods from 1 to 10 000 years were estimated by Eq. (4) and compared to the values of the original magnitude proposed by the OECD NEA (Table F.7). In addition, 95% and 5% confidence interval magnitude provided by the OECD NEA are also given in Table F.7, with fitting results plotted in Figure F.5.

Case 2: $M = 0.0003 * x + 0.5651\ (parameters\ A\ and\ B)$ (4)

Table F.7. Regression results for Case 2

Return period (years)	1	2	5	10	50	100	500	1 000	3 000	10 000
Original M (metres)	0.53	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.60	4.00
Linear Magnitude (metres)	0.565	0.566	0.567	0.568	0.580	0.595	0.715	0.865	1.465	3.565
Original M 95% (metres)	—	—	—	—	—	—	0.85	1.1	1.8	4.7
Original M 5% (metres)	—	—	—	—	—	—	0.72	0.85	1.3	3.2

Figure F.5. Linear regression fitting according to confidence interval



As a result of the regression analysis, the estimated magnitude values fell within the confidence interval and cannot be judged as inappropriate. The estimated magnitude for the initial return period was similar to those proposed by the OECD NEA. However, as the return period increases, errors in the magnitude values were found to occur. Error analysis in this case using the SUMXMY2 function gave an SSE value of 0.22443. In order to minimise this error in Case 2, the solver function was used, where again the target of the SSE value was set to 0 and an optimisation analysis was performed on parameters A and B from Eq. (4). The results are shown in Table F.8.

Table F.8. Optimisation for parameters

Parameter	Original	SSE_Solver
A	0.0003	0.000344252
B	0.5651	0.565051348

Based on the optimised parameters, linear regression analysis for Case 2 was re-estimated via Eq. (5). Table F.9 compares the magnitude values from the three sources [original, Eq. (4), and Eq. (5)] along with the confidence intervals.

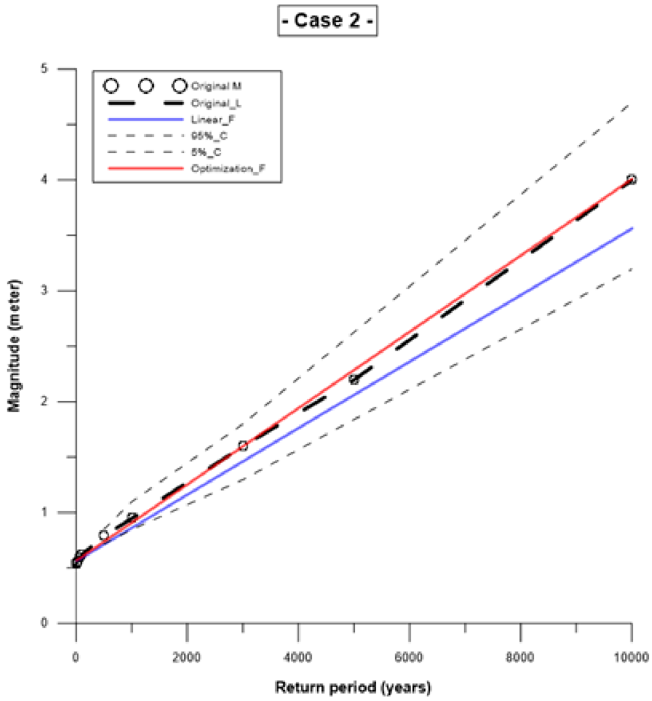
Case 1: $M = 0.000344252 * x + 0.565051348\ (Parameters\ A\ and\ B)$ (5)

Table F.9. Compare for magnitude

Return period (years)	1	2	5	10	50	100	500	1 000	2 000	10 000
Original M (metres)	0.53	0.53	0.54	0.55	0.59	0.62	0.79	0.95	1.6	4
Linear Magnitude (metres)	0.565	0.566	0.567	0.568	0.58	0.595	0.715	0.865	1.465	3.565
Optimised magnitude (metres)	0.565	0.566	0.567	0.568	0.582	0.599	0.737	0.909	1.598	4.008
Original M 95% (metres)	—	—	—	—	—	—	0.85	1.1	1.8	4.7
Original M 5% (metres)	—	—	—	—	—	—	0.72	0.85	1.3	3.2

In Case 2, the SSE of the optimised model was calculated to be 0.01. Accordingly, it was judged that the parameters A and B were significantly improved compared to the existing parameters. Fitting was then performed based on the optimised magnitude values. The results are shown in Figure F.6.

Figure F.6. Fitting result for Case 2



F.1.3. Case 2 result assessment

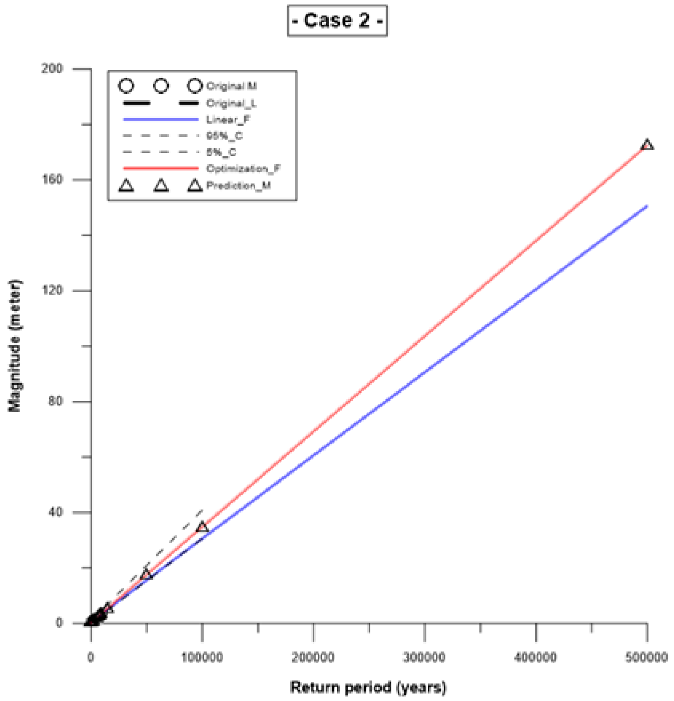
It was found that when the optimisation technique was applied to Case 2, the model performance further improved, as seen in Figure F.5. In other words, the optimised fitting was able to estimate values similar to the magnitude values proposed by the OECD NEA. Comparing Figure F.4 and Figure F.5, the trend lines calculated from the estimated magnitude values were also similarly estimated. Therefore, like Case 1, it can be judged that the fitting result for Case 2 in this study is valid.

Then, based on the regression analysis estimates in Case 2, magnitude values for return periods of 500, 5 000, 50 000, and 500 000 years were predicted. The results are shown in Table F.10 and Figure F.7.

Table F.10. Magnitude prediction by return period (Case 2)

Return period (years)	500	5 000	50 000	500 000
Magnitude (metres) Case 2 Exact	0.78	2.2	28	2 000
Linear KAERI mean (metres)	0.715	2.065	15.565	150.565
Optimised KAERI mean (metres)	0.737	2.286	17.778	172.691

Figure F.7. Magnitude prediction fitting for Case 2



Conclusion

In this study, statistical analysis was applied to the estimation of two cases presented by the OECD. In any statistical analysis, it is important to understand the characteristics of the data set. For the given problems here, the range of the return period was 10–10 000 years, while that of the magnitude was only 0.4–5.0 metres. Therefore, the coefficient of the synthetic model had a great influence on the analysis results. This study demonstrates that employing the full extent of the significant figures is important to handle the different ranges of data values. In the future, it is expected that data-based statistical values can be better estimated through various verified statistical models.