Comparative Assessment of Severe Accidents Risk in the Energy Sector: Uncertainty Estimation Using a Combination of Weighting Tree and Bayesian Hierarchical Models

M. Spada[a]*, P. Burgherr[a] and S. Hirschberg[a]
[a] Laboratory for Energy Systems Analysis, Paul Scherrer Institute (PSI), 5232, Villigen PSI, Switzerland

Abstract: This study analyzes the risk of severe fatal accidents within the full fossil energy chains causing five or more fatalities. The risk is quantified separately for OECD and non-OECD countries. In addition for the Coal chain, Chinese data are analyzed separately because it has been shown that data prior to 1994 were subject to strong underreporting. In order to assess the risk and its uncertainty, a Bayesian hierarchical model was applied. This allows yielding analytical functions for frequency and severity distributions. Furthermore, Bayesian data analysis inherently delivers a measure of a combination of epistemic and aleatory uncertainties, through the *a priori* distribution and likelihood function that compose the Bayes theorem. In this study, in order to reduce the epistemic uncertainty related to the subjective choice of the likelihood function, Bayesian Model Averaging (BMA) is applied. In BMA the final posterior distribution is a weighted combination of the posterior distributions assessed for different likelihood functions (models). The proposed approach provides a unified framework that comprehensively covers accident risks in energy chains, and allows calculating specific risk indicators, including their uncertainties, to be used in a holistic evaluation of energy technologies.

Keywords: Comparative Risk Assessment, Accident Risk, Bayesian Hierarchical Model, Bayesian Model Averaging, Fossil Energy Chains

1. INTRODUCTION

Risk assessment of severe accidents in the energy sector is an important aspect that contributes to improve safety performance of technologies, but is also essential in the broader context of sustainability, energy security and policy formulation by decision makers. Accidents in the energy sector are not only occurring in the production phase, but along the entire energy chain (e.g., [1]). Therefore, a comprehensive analytical framework is needed (e.g., [2]).

The classical approach to assess the risk of severe accidents in fossil energy chains is based on the use of metrics such as aggregated risk indicators focusing on human health impacts, i.e., fatality rates, or frequency-consequence curves (e.g., [1], [2]). However, when dealing with risk, uncertainty estimation is of great importance in order to take into account possible random fluctuations, for example due to the lack of data. Furthermore, uncertainty levels cannot be fully addressed using the aforementioned standard approach (e.g., [3]).

In this study, the risk and its uncertainty levels are assessed through a Bayesian Hierarchical model. In this way separate analytical functions for frequency and severity distributions can be calculated. In Bayesian data analysis, the posterior distribution is given by the product of an *a priori* distribution, describing how the data are distributed before introducing them into the analysis, and a likelihood function (e.g., [4]). The former defines the lack of knowledge and thus is intrinsically related to the epistemic uncertainties. The likelihood function is one of the fundamental parts of the Bayes theorem (e.g., [4]). It describes the randomness of the data and thus defines the aleatory uncertainties. The likelihood function is commonly defined following expert judgment and/or is selected following scientific community agreement. Therefore, besides describing the aleatory uncertainty, the likelihood function is as well a source of epistemic uncertainty due to subjectivity involved in its choice.

The concept of Bayesian Model Averaging (BMA) has been proposed in order to assess posterior distributions and to increase their robustness by considering a set of possible models that could describe a dataset (e.g., [5]). Therefore, BMA can serve as a tool to reduce the subjectivity in the

**matteo.spada@psi.ch*

choice of the model used as likelihood function for the Bayesian analysis. In BMA, the posterior distribution of the parameters of interest is given by the sum of the product of the models belonging to the model space, and the corresponding weights for these models. BMA has been used in different scientific fields, such as for example for dose-response risk assessment (e.g., [6]). Moreover, different methods have been proposed to estimate the weight of each model belonging to the model space to the final posterior distribution, for example using the Bayesian Information Criterion (BIC) (e.g., [7]), or Markov Chain Monte Carlo methods (e.g., [8]).

In this study, BMA is applied to the severity distributions in the risk assessment. The BIC method is used to a set of possible models common in hazard and risk assessment as well as survival analysis (e.g., [9], [10]). Therefore, the final posterior distribution for the parameter of interest, e.g., fatalities, is averaged over the entire model space. The final result is a severity distribution where both aleatory, from the likelihood functions, and epistemic uncertainties, from both *a priori* distribution and likelihood functions, are taken into account. Finally, the estimation of frequencies in the risk assessment is modeled following a common Bayesian analysis, since they can be described by a Poissonian distribution.

The above-described model is applied to the energy sector, and specifically to assess accident risks in fossil energy chains. The current analysis covers severe (≥5 fatalities) accidents in fossil (coal, oil, natural gas) energy chains for the years 1970-2008, which are contained in PSI's Energy-related Severe Accident Database (ENSAD). First, various risk indicators for different energy chains and country groups (e.g., OECD, non-OECD) are calculated. Second, results from the BMA and the standard approach are compared. Finally, a comparative evaluation for average and extreme risk is undertaken across energy chains and country groups.

2. DATA

2.1 ENSAD database

The ENergy-related Severe Accidents Database (ENSAD) (e.g., [11]) comprehensively covers energy related severe accidents worldwide. There exist numerous databases that look at accidents related to various industrial activities (e.g. FACTS online, OSH Update), but in contrast to ENSAD none of them is clearly focused on accidents attributable to the energy sector. Furthermore, ENSAD takes a full-chain approach because accidents can occur at all stages of an energy chain and not only at the actual power generation step. In ENSAD, data on all energy-related accidents is collected and classified into energy chains and activities within those chains. In addition, information on location, accident type, and different types of consequences (e.g. human health, environmental and economic impacts) is coded for to achieve a comprehensive global coverage of severe accidents. Finally, the accidents and severity reported in ENSAD for fossil chains are divided into three major groups, namely the Coal (incl. Lignite) chain, the Oil chain and the Natural Gas chain.

ENSAD has been developed using a wide variety of commercial and non-commercial information sources, ranging from specialized databases to technical reports, journal and newspaper articles, websites, etc. In the literature no commonly accepted definition can be found of what constitutes a so-called severe accident (e.g., [1]). The database ENSAD uses seven criteria to distinguish between severe and smaller accidents (e.g., [2]). Whenever one or more of the following consequences is met, an accident is considered to be severe:

- at least 5 fatalities or
- at least 10 injured or
- at least 200 evacuees or
- extensive ban on consumption of food or
- releases of hydrocarbons exceeding 10,000 (metric) tons or
- enforced cleanup of land and water over an area of at least 25 km^2 or
- economic loss of at least 5 million USD (2000)

We considered the energy-related accidents for the entire chain (from exploration and extraction through processing to end use) of each group.

2.2 Frequency and Severity

Risk can be decomposed into the product of the frequency and severity. The number of accidents per year gives the frequency, while severity measures the extent of the consequences of each accident. In this study, the number of fatalities describes severity. The reason for this choice is that fatalities generally comprise the most reliable indicator with regard to the completeness and accuracy of the data (e.g., [1], [12]). Furthermore, fatality information is superior to injured or evacuated persons information because often the severity of an injury or the duration of an evacuation is not reported (e.g., [1]).

Frequency and fatality distributions in ENSAD exhibit very different statistical behavior in each energy chain (Figure 1). The frequency distribution is influenced by temporal trends within each chain as well as country groups. These trends are commonly related to technological and regulatory differences and changes.

The fatality distribution, on the other hand, follows a very different pattern, and stretches over a broad range (Figure 1). The severity distribution is influenced by a number of parameters such as the material involved in the accident or the different products and the amount of material present, or the number of people in the vicinity of an accident. In addition, the data set is composed of accidents under a large range of different circumstances, meaning that we can thus assume also a large number of drivers for the risk exists (e.g., [3]).

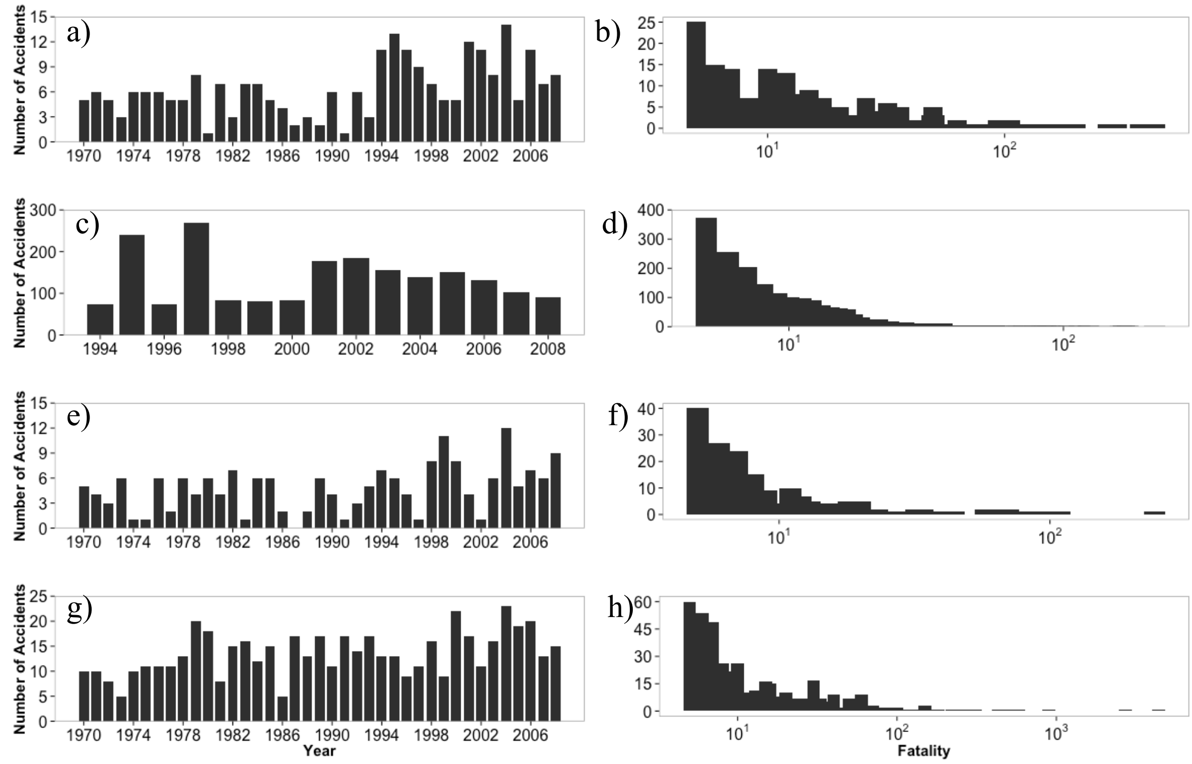

Figure 1 Number of accidents per year (Frequency) and number of accidents per fatalities (Consequence) for Coal (a and b), Coal China 94-08 (c and d), Natural Gas (e and f) and Oil (g and h).

2.3 Country Aggregates

Comparative results can be provided at the level of individual countries or for different country aggregates [12]. Based on the substantial difference in management, regulatory frameworks and general safety culture between industrial countries and developing ones, different energy chains are assigned to two major country groups, namely OECD and non-OECD. In addition for the Coal chain, Chinese data are analyzed separately because it has been shown that data prior to 1994 are subjected to strong underreporting (e.g., [16]). Furthermore, due to different data completeness for Chinese coal in 1994-1999 and 2000-2012, we subdivided the dataset into two subgroups related to the observation period. The summary of severe accidents for each substance energy chain and country group is shown in Table 1.

Table 1 Overview of the analyzed subsets per energy chain and country group for the years 1970–2008. Numbers of severe (≥ 5 fatalities) accidents, corresponding fatalities, and production used for normalization are given.

Energy Chain	Country Group	Number of Accident	Number of Fatalities	Production (GWeyr)
Coal	OECD	87	2259	18792
	non-OECD w/o China	162	5788	10071
	China 1994-1999	818	11302	1908
	China 2000-2008	1214	15750	7459
Oil	OECD	181	3430	36606
	non-OECD	350	19334	20524
Natural Gas	OECD	109	1258	17504
	non-OECD	77	1549	13459

2.4 Normalization

To derive a comparable risk measure of accident frequency across energy chains and country groups, the accident frequency should be normalized to a relevant unit of energy. Therefore, the accident frequency is normalized by the amount of production for each energy chain in GWeyr (Table 1). The Gigawatt-electric-year (GWeyr) is chosen because large individual plants have capacities of the order of 1 GW of electrical output (GWe), e.g., [2]. This makes GWeyr a natural unit to use when presenting normalized indicators generated within technology assessment. Furthermore, since we are dealing with data collected for fossil energy chains only, the thermal energy is converted to an equivalent electrical output using a generic efficiency factor of 0.35 (e.g., [2]).

3. METHOD

3.1 Bayesian Analysis

Bayesian inference is an alternative to the classical statistical inference (e.g., [4]). In the latter, also known as frequentist inference, only repeatable events have probabilities, while in the Bayesian inference, probability simply describes both epistemic and aleatory uncertainty (e.g., [4]). In fact, Bayesian analysis combines the information in the data represented by the entire likelihood function with prior knowledge about the parameters, which may come from other data sets or a modeler's experience and physical intuition [9]. Furthermore, the *a priori* distribution describes what is known before observing any data (e.g., [4]). Therefore, this distribution mainly contributes to the lack of knowledge and thus describes the epistemic uncertainty. The likelihood describes the process giving rise to data in terms of unknown parameter (e.g., [4]). It contributes to the random variability of the unknown parameter, and thus describes the aleatory uncertainty. Parameter estimation is made through the posterior distribution, which is computed using Bayes' Theorem:

$$\boldsymbol{p}(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{\boldsymbol{L}(\mathbf{y};\boldsymbol{\theta})\boldsymbol{p}(\boldsymbol{\theta})}{\int \boldsymbol{L}(\mathbf{y};\boldsymbol{\theta})\boldsymbol{p}(\boldsymbol{\theta})\mathbf{d}\boldsymbol{\theta}} \quad (1)$$

where $\boldsymbol{p}(\boldsymbol{\theta} \mid \mathbf{y})$ is the posterior distribution for the parameter $\boldsymbol{\theta}$ given the observed data $\mathbf{y}$, $\boldsymbol{L}(\mathbf{y};\boldsymbol{\theta})$ is the likelihood function, and $\boldsymbol{p}(\boldsymbol{\theta})$ is the *a priori* distribution of the parameter θ. The denominator is a normalizing constant that scales the posterior so that the area under the posterior probability distribution function equals one, i.e. make it "proper" meaning that it must converge (e.g., [4]). The main issue in equation (1) is that computing the integral may not be easy in cases when the parameter vector θ is large (e.g., [9]). In order to overcome this issue, Markov Chain Monte Carlo (MCMC) methods are commonly used (e.g., [13]). In fact, MCMC algorithm samples values of the parameters from the posterior distribution without computing the normalizing constant (e.g., [14]). Therefore, equation (1) can be written as

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto L(\mathbf{y}; \boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{2}$$

Among different type of MCMC algorithms (e.g., [13]), in this study the MCMC Gibbs algorithm is used in the sampling of the posterior distribution (e.g., [14]). This choice is made in order to avoid possible issues related to the incorrect choice of the jumping distribution, which is used for sampling the posterior in the other widely applied sampler, the Metropolis–Hastings algorithm (e.g., [13]).

3.2 Bayesian Model Averaging

The likelihood function in equation (2) is one of the fundamental parts of the Bayes theorem (e.g., [4]). It describes the probability of the evidence, i.e. the data, given the unknown parameter $\boldsymbol{\theta}$. The likelihood function is commonly defined following expert judgment and/or selected following scientific community agreement (e.g., [15], [3]). Therefore, besides describing the aleatory uncertainty, the likelihood function is source of epistemic uncertainty due to the level of subjectivity added in the choice of it. In this context, in order to reduce uncertainties related to the subjective choice of the likelihood, a possible solution, known as Bayesian Model Averaging (BMA), is given by [5] and modified by others, e.g., [8].

The basic idea of BMA is that the distribution of some interested quantity of a model, such as fatalities in our case, is derived over some space of possible models instead of only one, e.g., [5]. In other words, suppose that $\boldsymbol{M}$ is the set of all possible models of interest M, that is $\mathrm{M} \in \boldsymbol{M}$. If $\boldsymbol{\theta}$ is the parameter of interest, and the likelihood corresponding to the model $\mathrm{M_j} \in \boldsymbol{M}$ is given by $\boldsymbol{f}\ (\mathbf{y}|\boldsymbol{\theta}, \mathbf{M_j})$, then the formal Bayesian calculation, as given in equation (2), that summarizes the inference about $\boldsymbol{\theta}$ is given by, e.g. [5]:

$$\boldsymbol{p}(\boldsymbol{\theta} \mid \mathbf{y}) = \sum_{j=1}^{J} \boldsymbol{p}(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{M_j})\, \boldsymbol{p}(\mathbf{M_j} \mid \mathbf{y}) \tag{3}$$

where $\boldsymbol{p}(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{M_j})$ is the posterior density under $\mathrm{M_j}$ and $\boldsymbol{p}(\mathbf{M_j} \mid \mathbf{y})$ is the posterior probability of $\mathrm{M_j}$. The former can be rewritten as:

$$\boldsymbol{p}(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{M_j}) \propto \boldsymbol{f}\ (\mathbf{y}|\boldsymbol{\theta}, \mathbf{M_j})\boldsymbol{p}(\boldsymbol{\theta} | \mathbf{M_j}) \tag{4}$$

where $\boldsymbol{p}(\boldsymbol{\theta} | \mathbf{M_j})$ is the *a priori* density under $\mathrm{M_j}$. Equation (4) is describing a Bayesian analysis. In fact, it has the same structure as equation (2) in case of the sampling of posterior distribution through MCMC algorithms (section 3.1). Therefore, equation (4) in this study is computed through the MCMC Gibbs sampler.

In Equation (3), $\boldsymbol{p}(\mathbf{M_j} \mid \mathbf{y})$ is the posterior probability of $\mathrm{M_j}$, also known as posterior model weight. In fact, it describes the weight of the model $\mathrm{M_j}$ with respect to all the others belonging to the model space $\boldsymbol{M}$ in the posterior distribution of the parameter under interests:

$$\boldsymbol{p}(\mathbf{M_j} \mid \mathbf{y}) = \frac{\boldsymbol{m}(\mathbf{y} \mid \mathbf{M_j})\boldsymbol{p}(\mathbf{M_j})}{\sum_{t=1}^{J} \boldsymbol{m}(\mathbf{y} \mid \mathbf{M_t})\boldsymbol{p}(\mathbf{M_t})} \tag{5}$$

where $\boldsymbol{p}(\mathbf{M_j})$ is the prior probability of the j-th model in the model space reflecting the expert beliefs in the relative correctness of this model. A common choice is $\boldsymbol{p}\ (\mathbf{M_j}) = 1/J$, with $j = 1, \ldots, J$, which means that each model considered is equally likely before the data are observed. Furthermore, $\boldsymbol{m}(\mathbf{y} \mid \mathbf{M_j})$ is the marginal density of the observations under $\mathrm{M_j}$, e.g., [6], that is, the probability computed by integrating the likelihood multiplied by the prior distribution of the parameters over the parameter space:

$$\boldsymbol{m}(\mathbf{y} \mid \mathbf{M}_{\mathrm{j}}) = \int \boldsymbol{f}\,(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{M}_{\mathrm{j}})\boldsymbol{p}(\boldsymbol{\theta} \mid \mathbf{M}_{\mathrm{j}}) \tag{6}$$

Equation (6) is similar to the integral in equation (1), except that the model itself becomes a variable of the problem. Therefore, the integral in equation (6) can be difficult to compute, because a closed form might not be always available. In order to overcome this issue, researchers proposed different methods, from the use of Bayesian Information Criterion (BIC) (e.g., [6]) to MCMC algorithms (e.g., [8]). In this study, the former method is employed to estimate the posterior model weight $\boldsymbol{p}(\mathbf{M}_{\mathrm{j}} \mid \mathbf{y})$, e.g., [10].

3.2.1 BIC Method

The Bayesian Information Criterion has been proposed by researchers to provide an approximation of $\boldsymbol{p}(\mathbf{M}_{\mathrm{j}} \mid \mathbf{y})$, e.g., [8]. Such an approximation is adequate when a non-informative prior is assumed over the model space, e.g.,[10]. In fact, based on equation (1), if the prior is non-informative, the posterior distribution is strongly related to the likelihood function $(\boldsymbol{p}(\boldsymbol{\theta} \mid \mathbf{y}) \sim \boldsymbol{L}\,(\boldsymbol{\theta};\, \mathbf{y}) / \int \boldsymbol{L}\,(\boldsymbol{\theta};\, \mathbf{y})\boldsymbol{d\theta})$. Thus, the introduction of the *a priori* distribution, $\boldsymbol{p}(\boldsymbol{\theta} \mid \mathbf{M}_{\mathrm{j}})$, in equation (6) can be avoided. Under the aforementioned conditions, it has been shown, e.g., [8], that the posterior model weight can be described in terms of BIC as follow:

$$\boldsymbol{p}(\mathbf{M}_{\mathrm{j}} \mid \mathbf{y}) = \frac{exp(-0.5\, BIC_{\mathrm{j}})}{\sum_{t=1}^{J} exp(-0.5\, BIC_{\mathrm{t}})} \tag{7}$$

where $BIC_{\mathrm{j}} = -2\mathrm{L}_{\mathrm{j}} + p_{\mathrm{j}} \log N$. *N* is the sample size of the training set, *p* is the total number of parameters and L is the log-likelihood. Moreover, the lower BIC score the better the model is fitting the dataset.

4 APPLICATION TO THE DATA

4.1 Frequency

Frequency denotes the number of accidents per year (Figure 1). Essentially in the ENSAD database, accidents can be considered rare, independent events so that the frequency can be modeled as a Poisson distribution. Therefore, the frequency is modeled applying the common Bayesian procedure described in section 3.1. In equation (2), the likelihood is described by the Poisson model, while the *a priori* distribution for the parameter of interest, the frequency rate λ, is set to a non-informative, very broad Γ distribution ($\lambda \sim \Gamma\ (\alpha = 0.001, \beta = 0.001)$, with α and β describing the shape and rate of the distribution, respectively). Thus, the posterior distribution would be mainly influenced by the data, since the *a priori* distributions are weak (e.g., [3]).

The MCMC algorithm is run for 30,000 iterations, following a burn-in of 1,000 updates. Furthermore, the latter is also used to train the model. According to the Gelman-Rubin diagnostic (e.g., [16]) the simulated chains converged adequately in the MCMC practice implemented in this study. Once the posterior distribution for the mean frequency is estimated, it is normalized by the corresponding energy production in GWeyr (Table 1).

4.2 Severity

Severity measures the extent of the consequences of each accident (Figure 1). The fatality distribution is right-skewed (skewness > 0) meaning that most of the accidents are located at the left side of the mean, with catastrophic (extreme) events located to the right of the distribution. A unique model possibly describing the fatality distribution is difficult to establish, since different probability distribution functions exhibit right skewness. Therefore, in order to model the fatality distribution, the BMA method is applied (section 3.2).

In this study, the model space is arranged by a group of possible right skewed models that are commonly used in hazard, risk assessment and survival analysis (e.g., [9], [10]). Furthermore, only models described by a maximum of three parameters (location, shape and scale) are considered. This choice is made in order to avoid overfitting due to a high number of parameters in the model. The models used are shown in Table 2.

According to the BIC method, the posterior model weight for the BMA is estimated for all energy chains disaggregated by country groups. Table 2 shows that in all considered datasets, the same two models described the data best, meaning they had the lowest BIC scores. Therefore, the Inverse Gaussian (IG) and the Lognormal (LOGNO) distributions are used to model all the datasets. In addition, in case of Coal China 1994-1999, the Weibull distribution (weight = 0.01) has also to be considered in the assessment of the posterior distribution.

Table 2 Summary of goodness of fit (BIC score) and relative posterior model weight (Weight) for the fatality distributions collected for different fossil energy chains disaggregated by country groups.

Distributions	Coal							
	China 00-08		China 94-99		non-OECD w/o China		OECD	
	BIC	Weight	BIC	Weight	BIC	Weight	BIC	Weight
Logistic	624	0.00	506	0.00	160	0.00	293	0.00
Reverse Gumbel	588	0.00	481	0.00	143	0.00	264	0.00
Generalized Pareto	416	0.00	349	0.00	152	0.00	245	0.00
Lognormal	406	0.09	336	0.97	124	0.30	216	0.05
Weibull	424	0.00	346	0.01	139	0.00	238	0.00
Inverse Gaussian	402	0.91	344	0.02	122	0.70	210	0.95

Distributions	Natural Gas				Oil			
	non-OECD		OECD		non-OECD		OECD	
	BIC	Weight	BIC	Weight	BIC	Weight	BIC	Weight
Logistic	159	0.00	135	0.00	264	0.00	507	0.00
Reverse Gumbel	147	0.00	124	0.00	242	0.00	466	0.00
Generalized Pareto	119	0.00	116	0.00	205	0.00	400	0.00
Lognormal	113	0.08	105	0.22	189	0.13	379	0.01
Weibull	119	0.00	113	0.00	205	0.00	408	0.00
Inverse Gaussian	107	0.92	102	0.78	185	0.87	369	0.99

Once the posterior model weight is estimated (Table 2), the MCMC algorithm is used to assess the posterior distribution, for each model, of the parameters of interest, namely the expected value and the expected extreme value. According to the BIC method applied to BMA, non-informative, very broad prior distributions have to be defined e.g., [8]. For the location parameter (μ) the prior is defined as a normal distribution with mean 0 and standard deviation 0.01. For the shape parameter (σ) the prior is defined as a Γ distribution with shape and rate both equal to 0.001. Finally, for distributions described by three parameters, such as the Weibull distribution, the scale is defined by a Γ distribution with shape and rate both equal to 0.001.

Finally, for each energy chain and country group, the posterior distribution is calculated according to equation (3). Then for each model with posterior model weight > 0 (see Table 2), the posterior distribution is assessed using an MCMC Gibbs sampler, e.g., [14]. The MCMC algorithm is run for 100,000 iterations, following a burn-in of 10.000 updates. Furthermore, the latter is also used to train the model. According to the Gelman-Rubin diagnostic (e.g., [16]) the simulated chains converged adequately in the MCMC practice implemented in this study. The final posterior distribution is then evaluated as the sum of the weighted posterior distributions associated to the different models (equation (3)).

5 RESULTS

The aforementioned models allow us to compare frequencies and severity distribution as well as the total risk, being the product of the two components, between energy chains and country groups (OECD and non-OECD). For each parameter, the mean and the 5 and 95% quantiles are extracted from the posterior distribution.

5.1 Frequency

Figure 2 shows the average accident frequency per GWeyr over the period 1970-2008 for all fossil energy chains, and OECD and non-OECD countries. In addition for the Coal chain, Chinese data are analysed separately because it has been shown that data prior to 1994 were subjected to strong underreporting (e.g., [16]).

The normalized accident frequency is clearly highest for the Chinese coal chain. However, a comparison of the periods 1994-1999 and 2000-2008 indicates that the frequency is decreasing, and thus slowly approaching other non-OECD countries. This result could be possibly explained by the fact that the Chinese government, in the last decade, undertook a large effort to close small private mines in order to move the entire production to large mines, which are under the safety and regulatory policy of the government. Consequently, Coal china should be treated separately at least with regards to analysis of accident frequency. Finally, accident frequencies are generally lower in OECD than non-OECD countries for the coal and oil chains, whereas for natural gas no significant difference is found. The latter could be possibly explained by the lack of data for both OECD and non-OECD country groups. In fact, as shown in Figure 1e, in the natural gas energy chain fewer accidents per year happened with respect to oil and coal energy chains (Figure 1a, c, g).

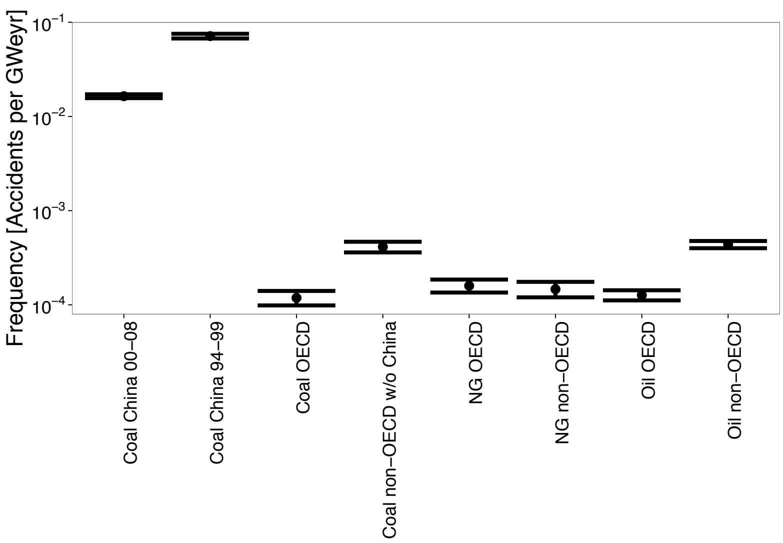

Figure 2 Mean frequencies (accidents per GWeyr), 5% and 95% of the posterior distribution, averaged over time.

5.2 Severity

Figure 3 shows the mean and 5-95% error bars for the expected fatalities per accidents as well as the number of fatalities exceeded in 1% of all accidents for various fossil energy chain and country group combinations. Additionally, the figure includes the contribution of each posterior distribution to the final BMA, described by the mean and 5% and 95% quantiles. Overall, the inverse Gaussian (IG) model was the dominant contributor to the final result in most of the cases (Table 2), except for Coal China 94-99 where it was only 2%. In the case of fatalities exceeded in 1% of accidents, the IG model shows a very broad uncertainty range compared to the lognormal and Weibull models. The same effect is to a lesser extent also visible for Coal China 00-08, although IG replaces LOGNO as the dominant contributor. This effect could be possibly related to the lack of events in the tail of the historical distribution of the accidents (Figure 1d). In fact, in case of Coal China 94-99, all events are clustered in the range 5-50 fatalities, with only a data point in the tail at 114 fatalities. Moreover, in case of Coal China 00-08, the maximum fatality is twice as big as the time period 94-99. However, the fatality distribution is similar to the ones for oil, natural gas and coal without China, where more than one observation is located in the tail of the distribution. This lack of data could cause large fluctuations in the inverse Gaussian tail, since IG needs relative high probability, meaning number of observations, in the tail in order to be able to model it (e.g., [17]). Therefore, the large fluctuations would increase the randomness in modeling the tail of the distribution and, thus, increase the aleatory uncertainty.

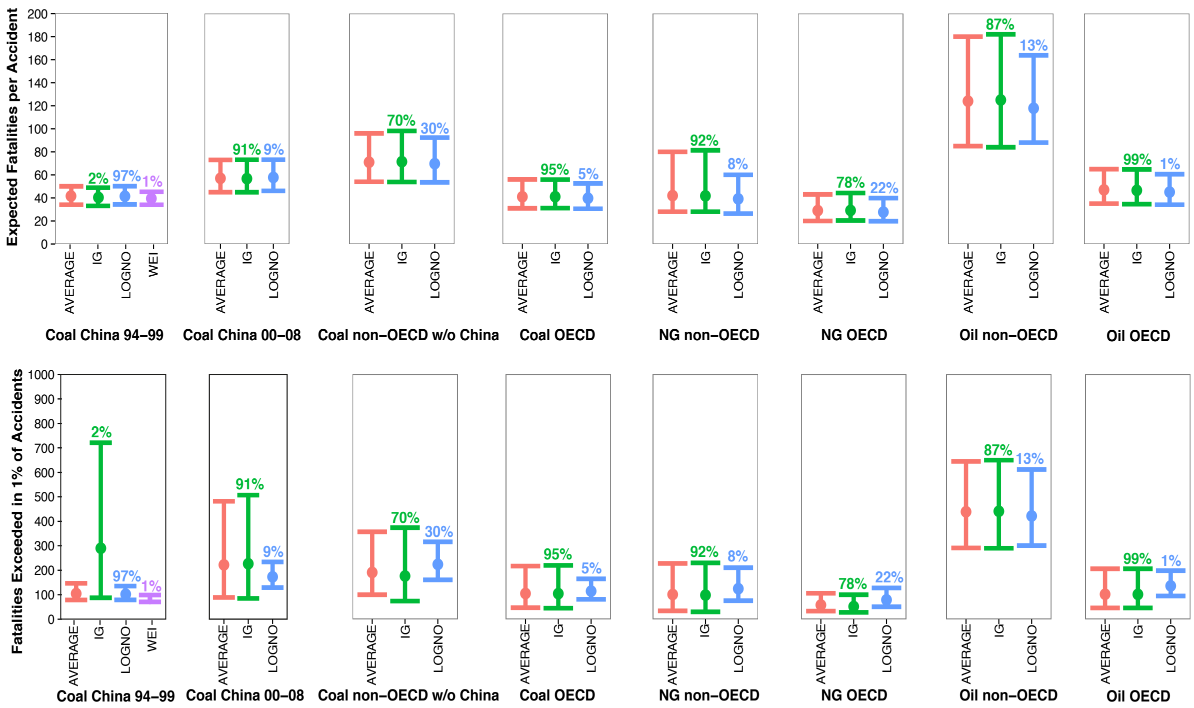

Figure 3 Mean, 5% and 95% quantiles for all the energy chains and country groups analysed. For each case, the BMA (AVERAGE) result is shown for the various distributions including their percent contribution or weight). Results are shown for the expected fatalities and fatalities exceeded in 1% of all accidents

For the expected fatalities per accident (upper panel in Figure 3), the final BMA shows no statistically significant differences between different energy chains and country groups, except in the cases of oil non-OECD and Coal energy chains. In these two cases the expected fatalities per accident are about twice as high for non-OECD compared to OECD countries. Concerning natural gas, similar to the frequency case, there is no significant difference between OECD and non-OECD country groups. However, in the latter, the shape of the major contributor's distribution (IG) clearly affects the final result. In fact, the posterior distribution of the expected value exhibits a long tail, resulting in large values at the 95% quantile. It is interestingly in the case of coal China that both time ranges taken into account are not significantly different with respect to other considered energy chains. However, the mean number of expected fatalities is larger for coal China 2000-2008 than 1994-1999. This could be explained by the fact that moving the production from small private mines to big mines, as was done by the Chinese government in the last decade, the number of accidents (Figure 2) could be reduced, but at the same time the potential consequences can be more severe due to the larger number of workers present in these mines.

Overall, the aforementioned behaviour for the number of expected fatalities well described the behaviour of the number of fatalities exceeded in 1% of the total accidents (lower panel in Figure 3).

However, the main differences are related to the large uncertainty in Coal China 00-08 and to the slightly similar behaviour of Coal non-OECD without China with respect to the other cases. The former can be possibly described by the increase of randomness in modelling the tail of the inverse Gaussian distribution due to the presence of a large number of events in the tail of the historical observations (Figure 1). Therefore, these would increase the aleatory uncertainty and, thus, affect the 95% quantile of the posterior distribution.

5.3 Risk Indicators: Mean and Exceedance

By definition, risk is the product of the frequency and severity. In order to compare the risk between different energy chains and country groups, two risk indicators are used in this study. The first one addresses the mean risk and is expressed as the expected number of fatalities per accident. The second one represents the extreme risk, defined by the threshold exceeded for a specific return frequency, and is given by the total number of fatalities exceeded at 1% frequency per accident (e.g., [5]). Table 3 summarizes the results for these risk indicators. In addition, results from the standard approach, non-normalized mean and 1% exceedance per accident based on frequency-consequence curves, are also shown in order to compare the results.

Generally, the results for the mean aggregate indicators show a good accordance between BMA and the standard approach. The main difference concerns the Oil non-OECD case. This could be explained by the fact that the historical observations are distributed with a very long tail (see Figure 1), due to an extreme event happened in 1987 in the Philippines, where the tanker *Victor* collided with the Ferry *Dona Paz* resulting in 4386 fatalities (e.g., [1]). Such extreme events can have a strong impact on the mean value in the case of the standard approach, resulting in a large difference to the expected value in BMA, where posterior distributions are more resistant to outliers (e.g., [6]). In addition, the 5 and 95% quantiles are significantly different between the standard approach and the BMA model. This can be explained by the fact that in the former case no uncertainty analysis is included, resulting in a broad range in the entire fatality space of the dataset, while in BMA both aleatory and epistemic uncertainties are modeled (e.g., [7]).

Table 3 Results for the full risk, expected fatalities per accident and 1% exceedances per accident. Each value is given by the mean with 5 and 95% intervals. In addition, the mean and 1% exceedance is calculated for frequency and production level in the time range 1970-2008 following the standard approach.

Country Group	BMA Model		Standard Approach	
	Mean per Accident	1% Exceedance per Accident	Mean per Accident	1% Exceedance per Accident
Coal China 00-08	57(45; 73)	222(89; 482)	55(8; 124)	60(5; 211)
Coal China 94-99	41(34; 50)	112(84; 153)	39(8; 88)	44(9; 112)
Coal non-OECD w/o China	71(54; 96)	191(100; 357)	67(8; 199)	183(120; 413)
Coal OECD	41(31; 56)	105(47; 217)	39(7; 93)	123(29; 231)
NG non-OECD	42(28; 80)	101(34; 228)	38(6; 97)	88(81; 215)
NG OECD	29(20; 43)	58(33; 106)	26(6; 87)	37(10; 104)
Oil non-OECD	124(85; 180)	439(291; 645)	173(9; 554)	419(28; 3812)
Oil OECD	47(35; 65)	102(46; 206)	44(7; 139)	48(37; 226)

For the risk indicator 1% exceedance per accident, the comparison between the BMA and standard approach results shows a different behavior for the 5 and 95% quantiles due to the fact that in the second case no uncertainty analysis is assessed. In case of the average, the results differ in all the cases. In most of them the average values estimated using the standard approach are lower than the modeled extreme. This is related to the fact that a significant number of the historical observations have small consequences. Furthermore, they compensate the presence of few data points, such as the extreme value, resulting in a shift of all the quantiles towards small number of fatalities. Therefore, the value for 1% exceedances is close to the mean of the distribution, resulting in a different value with respect to the BMA result, which is accounting for the uncertainty and the outliers that strongly affect the standard approach. In cases where the standard approach shows larger averages compared to BMA, the former's results is strongly affected by the presence of outliers. These extreme values strongly affect, in terms of number, the distribution, shifting the higher quantiles toward them. This results in a larger value of the fatalities exceeding 1% of accidents in the standard approach with respect to BMA. In fact, in the latter, the posterior distribution is resistant to outliers (e.g., [6]), while in case of the standard approach the result is strongly affected by them.

In Figure 4 the visualization of the risk is shown for the average risk (Figure 4a) and for the risk of extreme events (Figure 4b). The overall highest risk is found for coal China 94-99 with an expected number of 41 fatalities per accident at current consumption levels and a 1% probability that an accident with more than 112 fatalities takes place. However, as described above, this result is different from all other energy chains and country groups due to its much larger historical dataset than in any other case (e.g., [16]). Furthermore, in all other cases, Oil non-OECD clearly shows the highest risk. In fact for Oil non-OECD 124 fatalities are expected per year at current consumption levels, and at a 1% probability per year an accident with more than 439 fatalities is expected. It is important to note that the result for Coal non-OECD w/o China in terms of number of fatalities or fatalities exceeding 1% of the accidents is comparable with Coal China. However, based on the frequency it is not, since in case of China, many more accidents occurred. Overall, Natural gas is the least risky energy chain and, more specifically, the Natural gas OECD group performs best. Finally, OECD generally exhibits lower risk levels than non-OECD, and even more pronounced than coal China.

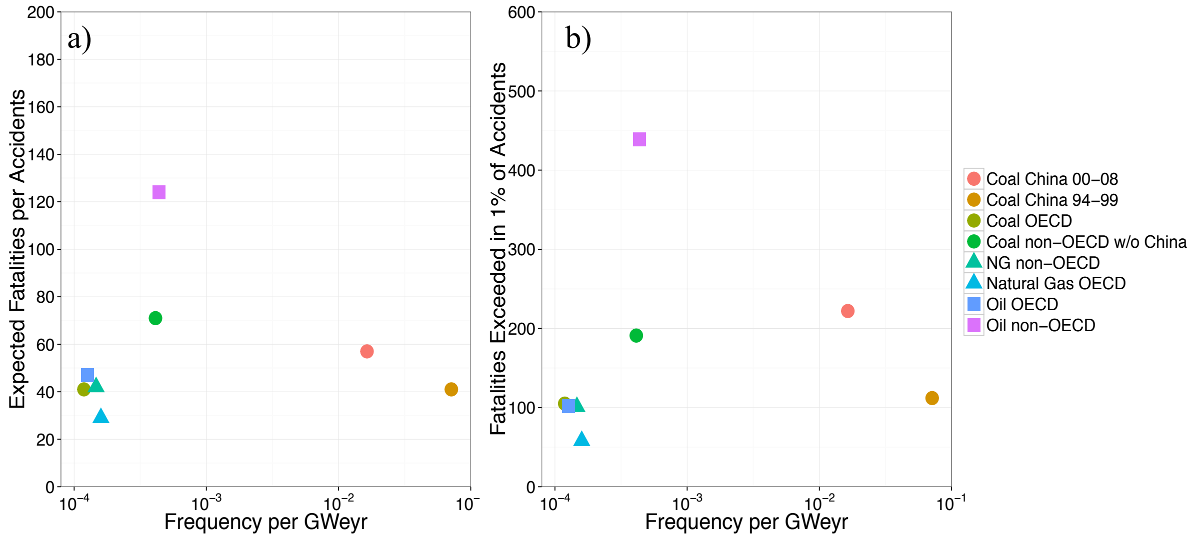

Figure 4 a) Mean number of fatalities per severe accident versus frequency of severe accidents per GWeyr. b) Fatalities exceeded in 1% of accidents versus total accident frequency per GWeyr.

6 CONCLUSIONS

This study presented a first-of-its-kind implementation of the Bayesian Model Averaging (BMA) method in a comparative risk assessment framework to comprehensively quantify the risk of severe accidents in fossil energy chains. This framework allows estimating uncertainty and dealing with lack of data and lack of knowledge by averaging the posterior distribution over a pre-defined model space. This "top down" approach can also be useful to complement conventional, detailed "bottom-up" models of risk quantification that are conducted for individual plants with specific physical processing and site conditions. Therefore, the proposed approach provides a unified framework that comprehensively covers accident risks in energy chains, and allows estimating specific risk indicators, including their uncertainties. This information provides an essential element in a holistic sustainability and energy security evaluation of energy technologies. The overall risk is found to be highest in Coal China for the time range 94-99. Among Coal China, non-OECD country groups for all energy chains show higher risk in terms of expected number of fatalities per accident as well as for extreme cases. Furthermore, results show that Natural gas is the least risky energy chain. In future work based on the database ENSAD, both the scope of this model will be expanded towards incorporating other types of consequences (e.g. injured) and other energy chains (e.g. hydropower), and the resolution of risk will be increased, that is, to differentiate the risk for more activities or regions.

Acknowledgements
This study was partially performed using data collected within the Collaborative Project SECURE (Security of Energy Considering its Uncertainty, Risk and Economic implications), Contract No. 213744 of the 7th Framework Programme of European Commission.

References

[1] P. Burgherr and S. Hirschberg. *A Comparative Analysis of Accident Risks in Fossil, Hydro, and Nuclear Energy Chains.* Human and Ecological Risk Assessment: an International Journal, 14, pp. 947–973, (2008).

[2] S. Hirschberg, G. Spiekerman and R. Dones. *Severe accidents in the energy sector*. PSI Report No. 98-16, Paul Scherrer Institut, (1998), Villigen PSI, Switzerland.

[3] P. Eckle and P. Burgherr. *Bayesian Data Analysis of Severe Fatal Accident Risk in the Oil Chain.* Risk Analysis, 33, pp. 146–160, (2013).

[4] A. O'Hagan. Bayesian statistics: principles and benefits. *Bayesian Statistics and Quality Modelling in Agro-Food Production Chain*, Dordrecht: Kluwer Academic Publishers, (2003).

[5] L. R. Pericchi and M. E. Pérez. *Posterior robustness with more than one sampling model.* Journal of Statistical Planning and Inference, 40, pp. 279–294, (1994).

[6] K. Shao and J. S. Gift. *Model Uncertainty and Bayesian Model Averaged Benchmark Dose Estimation for Continuous Data*. Risk Analysis, 34, pp. 101–120, (2013).

[7] A. E. Raftery. *Bayesian model selection in social research*. Sociological Methodology, 25, pp. 111–163, (1995).

[8] L. Wasserman. *Bayesian Model Selection and Model Averaging*. Journal of Mathematical Psychology, 44, pp. 92–107, (2000).

[9] D. S. Reis and J. R. Stedinger. *Bayesian MCMC flood frequency analysis with historical information*. Journal of Hydrology, 313, pp. 97–116, (2005).

[10] A. J. Bailer, R. B. Noble and M. W. Wheeler. *Model Uncertainty and Risk Estimation for Experimental Studies of Quantal Responses*. Risk Analysis, 25, pp. 291–299, (2005).

[11] P. Burgherr, P. Eckle and S. Hirschberg. *Final Report on Severe Accident Risks including Key Indicators*. Deliverable No 5.7.2a SECURE Project, 7th Framework Programme of European Commission, (2010).

[12] P. Burgherr, P. Eckle and S. Hirschberg. *Comparative assessment of severe accident risks in the coal, oil and natural gas chains*. Reliability Engineering and System Safety, 105, pp. 97–103, (2012).

[13] S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. Taylor and Francis Group, LLC, (2011).

[14] C. Andrieu, N. de Freitas, A. Doucet and M. I. Jordan. *An introduction to MCMC for machine learning*. Machine Learning, 50, pp. 5–43, (2003).

[15] P. J. Birrell, G. Ketsetzis, N. J. Gay, B. S. Cooper, A. M. Presanis, R. J. Harris, A. Charlett, X.-U. Zhang, P. J. White, R. G. Pebody and D. De Angelis. *Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London*. Proceeding of the National Academy of Sciences, 108, pp. 18238–18243, (2011).

[16] A. Gelman and D. B. Rubin. *Inference from Iterative Simulation Using Multiple Sequences*. Statistical Science, 7, pp. 457–511, (1992).

[17] I. J. Good. *The Population Frequencies of Species and the Estimation of Population Parameters*. Biometrika, 40, pp. 237–264, (1953).