A Comparison of Risk Measures for Accidents in the Energy Sector and Their Implications on Decision-Making Strategies

Matteo Spada[a], Florentina Parashiv[b,c], Peter Burgherr[a]

[a] *Laboratory for Energy System Analysis, Paul Scherrer Institut, Villigen PSI, Switzerland.*

[b] *Faculty of Economics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.*

[c] *Institute for Operations Research and Computational Finance, University of St. Gallen, St. Gallen, Switzerland*

Corresponding author: Matteo Spada, Laboratory for Energy System Analysis, Technology Assessment Group, OHSA/D19, Paul Scherrer Institut, CH-5232, Villigen PSI, Switzerland; e-mail: matteo.spada@psi.ch; Phone: +41 (0)56 310 56 90

Abstract

Within the broader context of energy security and critical infrastructure protection, the comprehensive assessment of accidents and their related consequences are of high priority for many stakeholders. The risk of accidents is commonly assessed by risk indicators, since they can provide a direct comparison between energy chains and country groups. In this study, Value-at-Risk (VaR), Expected Shortfall (ES) and the Spectral Risk Measures (SRM) are applied within an energy security perspective. In particular, fatality risk indicators are calculated for different characteristics of fossil data sets extracted from the Energy-related Severe Accident Database (ENSAD). The aim is to understand the applicability of such risk measures in a different context and field of application than the financial realm for which they were originally developed. The use of these risk measures could help to facilitate a better understanding of energy accident risks to insurers and other industry stakeholders that normally focus on financial and less infrastructure-related aspects. Furthermore, the pros and cons of these risk measures, and their implications for decision-making strategies are discussed. The results clearly demonstrate the usefulness of VaR, ES and SRM compared to the often used maximum consequence indicator in the evaluation of accident risks in the energy sector.

Keywords: Risk Assessment; Value-at-Risk; Expected Shortfall; Spectral Risk Measures; ENSAD; Energy Security.

Highlights

- VaR, ES and SRM applicability to accident risk assessment in the energy sector has been validated.
- VaR tends to underestimate the energy accident risk with respect to ES (and SRM).
- VaR, ES and SRM could facilitate a better understanding of energy accident risks to the industry.
- VaR, ES and SRM could improve the information on risk that an energy-related business can get.

1 Introduction

In today's modern society, energy is one of the key prerequisites for goods and services production. In the past decades a number of catastrophic events (e.g., Deepwater Horizon oil spill in 2010 or Fukushima nuclear accident in 2011) have influenced the entire energy-related business due to their consequences connected to societal vulnerabilities affecting human health, the environment, and the supply of economic goods and services. Consequently, the assessment of accident risks in the energy sector has become a high priority for many stakeholders within the context of a safe and secure provision of energy.

Comparative risk assessment has been introduced in the 1980s and, since then, it became a central aspect in the comprehensive evaluation of the performance of energy technologies [1-4]. It aims to compare different energy technologies among individual countries or country groups [2, 5, 6]. The two standard methods commonly used are frequency-consequence (F-N) curves and aggregated risk indicators [7]. The former is a common way to express collective and societal risk in a quantitative assessment, since the F-N curves provide the probability of accidents with varying degrees of severity of consequences, including chain-specific maximum damages. On the other hand, risk indicators allow a straightforward comparison between energy chains and country groups, based on aggregated and normalized risk values. However, it is important to consider a variety of risk factors (e.g. average vs. extreme risk) and types of consequences (e.g. fatalities, injuries, etc.) because no single aspect or indicator can provide the full picture [8].

Fatality rates and maximum consequences are often used as risk indicators, since it has been shown to be a good combination in order to assess the risk of accidents in the energy sector [2]. The fatality rate provides a measure of expected fatalities per unit of energy produced, whereas the maximum credible consequences of a single accident can be seen as a measure of risk aversion. The latter plays an important role in risk management for different stakeholders., According to Thomas it is "a measure of the feeling guiding the person who faces a decision with uncertain outcomes" [9].

Although the maximum credible consequence event in a data set of historical accidents is straightforward to determine, one of its major drawbacks is that it is disregarding the specific distribution properties of the accident data [10], limiting the worst-case risk to an unique value. This is particularly relevant for companies in the energy business to better understand the risk associated with high consequences events in the low-frequency domain, also known as extreme events. Furthermore, the non-financial nature of the aforementioned risk indicators could be another obstacle for the industry in order to better understand risks posed by energy accidents, since they normally focus on financial rather than infrastructure-related aspects.

In order to overcome the aforementioned issues, a set of possible risk indicators from the financial realm can be considered. On the one hand they are commonly used by energy companies and, on the other hand, they address the risk at high consequence levels. In particular, the Value-at-Risk (VaR), the Expected Shortfall (ES) and the Spectral Risk Measure (SRM) are considered. While VaR is commonly used by different stakeholders dealing with financial risk [8, 10], ES and SRM have attracted considerable attention in the financial literature in the last two decades [11-18].

VaR is defined as the loss level that will not be exceeded with a certain confidence level during a certain period of time. Therefore, it is based on three components: a time period, a confidence interval and a loss amount. Furthermore, VaR can also be seen as a measure of risk aversion than the maximum credibility interval because it denotes an aversion threshold, which is however only a minimal part of the risk aversion utility function [9]. However, in the last decade, it has been shown that the risk described by VaR, in particular situations, is not the best choice [19], since it is limiting the risk to an unique value without any consideration on what the risk could be beyond it. Therefore, the risk aversion related to VaR displays the attitude of an incoherent investor who is only concerned about the threshold level and neglects all the losses themselves [20].

In this context, since the early 2000s, ES[18] and SRM [14, 20] have been introduced in order to include the risk aversion utility function in the estimation of the risk as well as to be able to consider risk beyond a certain threshold. The SRM is a risk measure given as a weighted average of outcomes, where bad outcomes are, typically, included with larger weights. The ES on the other hand, is a particular case of the SRM, where the risk aversion function is a constant for all the quantiles, meaning that the bad outcomes include the same weights as the good ones. However, among the aforementioned conditions used to introduce ES and SRM, the most important one is related to the definition of a coherent risk measure [19]. A risk measure is considered coherent if the criteria of monotonicity, translation invariance, homogeneity and sub-additivity are fulfilled. The first three conditions give the description of the requirements to make the risk acceptable. For example, in the context of accidents in the energy sector, it refers to increasing safety aspects in order to reduce the risk. On the other hand, the subadditivity states that the diversification helps to reduce the risk. When two risks are aggregated, the combined risk should either decrease or stay the same. For accidents in the energy sector, the latter property is quite important, since in general risk indicators are aggregated measures of accidents triggered by different sources (e.g., man-made, natural events, terrorism) or for different energy chain stages (e.g., transportation, storage, exploration, etc.). Finally, the coherency of a risk measure intrinsically considers the risk aversion. In fact, a coherent risk measure is like that if it assumes larger weights to worst cases [14]. While both ES and SRM have been shown to be coherent risk measures [14, 21], VaR is not since it does not fulfill the subadditivity property[21].

In this paper, we compare the more recent "coherent" risk metrics such as ES and SRM, with a more "conventional" risk measure, which is VaR. They are purposely used outside the financial realm, i.e., to describe accident risks in the energy sector which is outside their original field of application. On the one hand, ES and SRM include information on the risk beyond a certain threshold (e.g., VaR), and on the other hand, they intrinsically account for the risk aversion utility function, which is of great interest for industry and insurers. Furthermore, since risk indicators in the energy sector are commonly based on the direct use of historical observations [2, 6, 10, 22], in this study, the aforementioned risk measures are calculated directly from historical observations without the use of parametric or semi-parametric modelling [12, 23, 24]. The analysis conducted in this study is based upon data of the Paul Scherrer Institute's (PSI) Energy-related Severe Accident Database (ENSAD) database, which contains information on accidents and related consequences (e.g., fatalities, substance released in metric tons, economic losses) classified into energy chains and activities within them [25].

The current paper is subdivided in the following sections. In Section 2 we provide a detailed description of the data collected for the three fossil energy chains that were analyzed. Section 3 explains the methodology behind the VaR, ES and SMR risk measures. In Section 4 comparative results for the considered risk measures estimated from the historical observations (Section 2) are presented and discussed, including possible implications on the decision-making process. Finally, in Section 5, the conclusions of the study are summarized.

2 Data

ENSAD is a comprehensive collection of accidents related to the energy sector. It was developed at the PSI in the 1990s [25], and since then it has been updated continuously with new information from different sources, such as, specialized databases, technical reports, journal papers, books, etc. In contrast to databases that rely on a single or few information sources, the multitude of sources considered by ENSAD is thoroughly verified, harmonized, and merged to ensure consistent and high quality data. The aim of ENSAD, since its start, is to comprehensively collect information about accidents in all energy chains that are attributable to fossil, nuclear, hydropower and, more recently, new renewables technologies [10]. In ENSAD, data about accidents and related consequences (e.g., human health effects, impacts on environment or economy) are collected and classified into energy chains and activities within those chains, since accidents are not only occurring at the actual power generation step [2].

In the literature no common definition of severe accident exists [2]. ENSAD is primarily focusing on severe accidents, since industries, stakeholders, decision-makers, etc., are more concerned about them. However, accidents with minor consequences are also included in ENSAD for specific energy chains and activities [26]. Naturally, a higher level of reporting and completeness of information can be expected for severe than for small accidents. The emphasis on severe accidents also allows accounting for reporting differences between countries. In ENSAD whenever one or more of seven consequence thresholds is met, an accident is considered to be severe:

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

In this study, selected accidents subsets of ENSAD have been extracted. In particular, coal accidents in China, oil accidents in different country groups, and natural gas accidents in Germany have been chosen. In this way the major fossil energy chains are represented, while at the same time the consideration of different countries and country groups allows to take into account different types of distribution behavior. In the following sections the characteristics of the three accident subsets are described in detail.

2.1 Oil

For the oil chain three country groups, namely OECD, EU28 and non-OECD have been considered in this study. This distinction is made due to the substantial difference in management, regulatory frameworks and safety between different countries. The complete oil chain, i.e., from exploration to end use, and accidents that fall into the period 1971 to 2008 and resulted in at least 5 fatalities are considered. The reason for this choice is that fatalities generally comprise the most reliable consequence indicator with regard to completeness and accuracy of the data [22, 27]. Furthermore, for injured or evacuated persons the severity of an injury or the duration of an evacuation is often not reported precisely [27].

In Fig. 1 the fatality distributions related to accidents in the oil chain for the three considered country groups are shown.

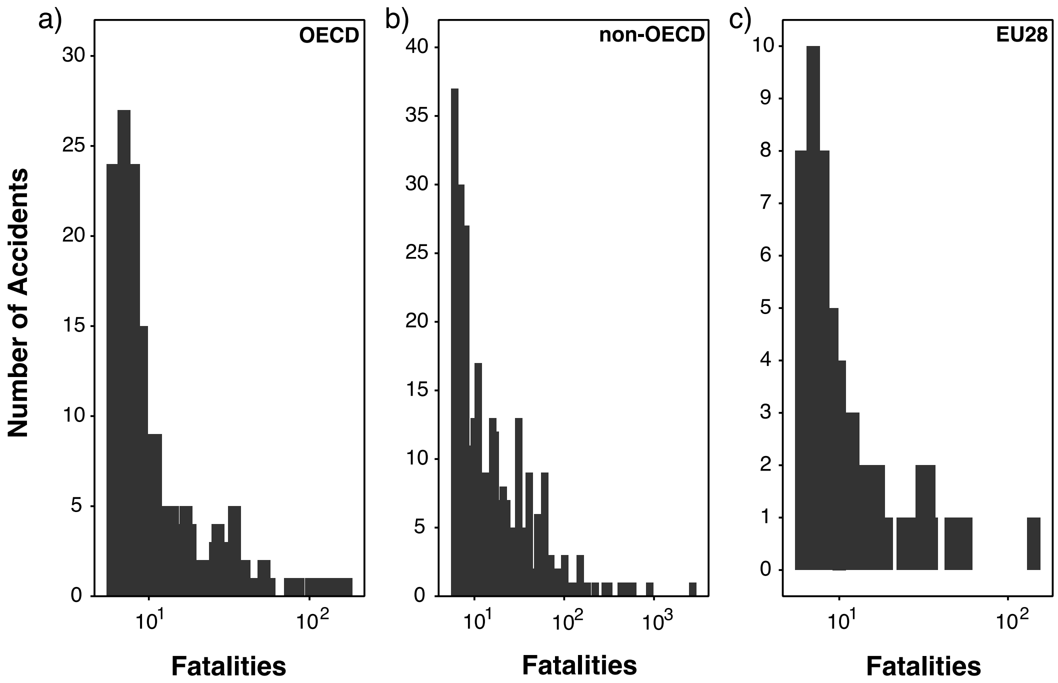

Fig. 1: Number of accidents per fatalities in the oil chain in the time period 1971-2008. a) OECD countries; b) non-OECD countries; c) EU28 countries.

In addition, the descriptive statistics for these datasets are summarized in Table 1.

Table 1: Descriptive statistics for severe accidents in the oil chain in the time period 1971-2008 for the OECD, non-OECD and EU28 country groups.

Country Group	Accidents	Fatalities	Skewness	Kurtosis	Observed Maximum
OECD	187	3495	4.9	36	252
non-OECD	358	19516	13	186	4386
EU28	65	1243	3.8	18.9	167

In all cases, the data show a right skewed distribution (skewness > 0), implying that most of the values are to the left of the mean with extreme values at the right of the distribution. The high value for the kurtosis index for all considered country groups indicates a leptokurtic (kurtosis > 3) data outlay implying fatter tails, i.e. high incidence of events with large consequences. In this context, the non-OECD fatalities distribution exhibits skewness and kurtosis one order of magnitude larger than OECD and EU28 country groups. The difference could be explained by the presence of a large set of accidents that produced more than 100 fatalities in non-OECD countries compared to OECD (only few events) and EU28 (only the observed maximum). Furthermore, the presence of two accidents in the non-OECD country group fatality distribution with number of casualties one order of magnitude larger than the maximum number of fatalities observed in OECD and EU28 could explain this difference. In fact, the two largest events in non-OECD countries, caused by a collision of a ferry and a tanker in the Philippines in 1987 (4386 fatalities) and a collision of a road tanker with an army vehicle in a tunnel and the subsequent explosion in Afghanistan in 1982 (2700 fatalities), are stretching the fatality distribution towards very large consequence events. This is not the case for OECD and EU28 countries, where the observed maxima are located at 252 fatalities (explosion due to a mechanical failure of a pipeline in Mexico in 1992) and 167 fatalities (explosion due to well blowout on a platform in UK in 1988), respectively.

2.2 Coal China

For Coal China, accidents with at least five fatalities are considered for the time period 1994-2008 because data prior to 1994 are subject to strong underreporting [22]. The dataset was subdivided in two groups from 1994-1999 and 2000-2008 because the former period is based on official Chinese data [28], while the second one relies on data collected from freely available Chinese information sources.

In Fig. 2, the fatality distributions for the considered time periods are shown.

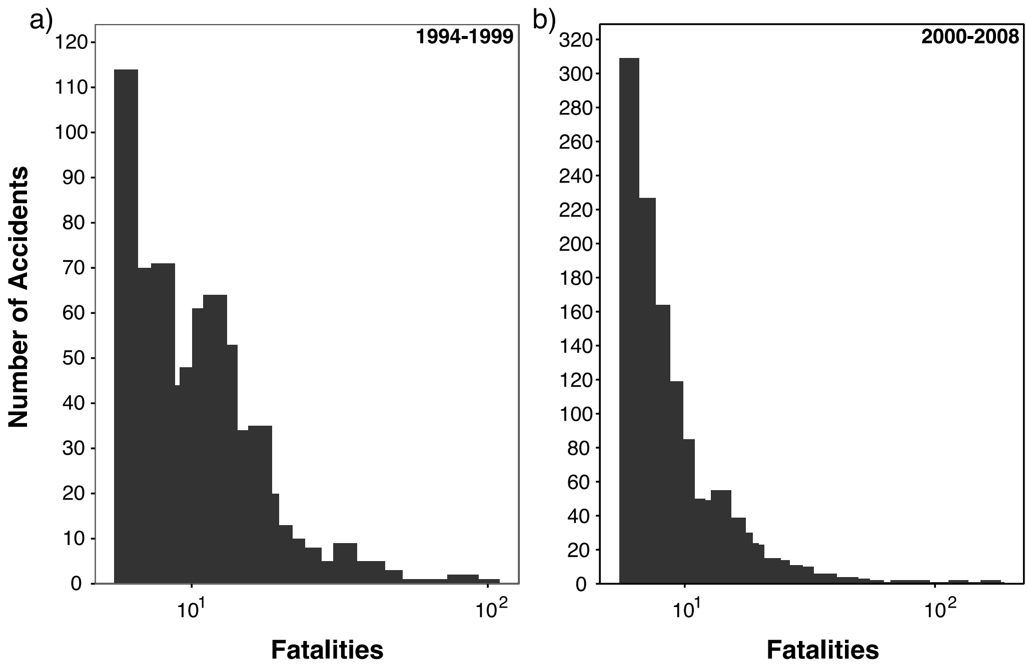

Fig. 2: Number of accidents per fatalities in the Chinese coal chain. a) Time Period 1994-1999; b) Time Period 2000-2008.

Furthermore, in Table 2 the descriptive statistics are given.

Table 2: Descriptive statistics for severe accidents in the Chinese coal chain for the periods 1994-1999 and 2000-2008.

Time Period	Accidents	Fatalities	Skewness	Kurtosis	Observed Maximum
Coal 1994-1999	828	11302	3.6	20.1	114
Coal 2000-2008	1214	15750	6.1	50.7	215

Similar to oil, the two datasets considered for coal china are both right skewed (skewness > 0) with a large kurtosis values (> 3) indicating their leptokurtic nature, and more specifically the high incidence of events with large consequences. However, the two datasets exhibit different skewness and kurtosis, although of the same magnitude. The aforementioned values are larger for the time period 2000-2008 than 1994-1999, due to the presence of few accidents with consequences larger than 100 fatalities in this period compared to only one accident of more than 100 fatalities for the time period 1994-1999. The difference in the descriptive statistics could be also explained by the difference in the observed maximum, which is larger in the time period 2000-2008 (215 fatalities) than in 1994-1999 (114 fatalities). In fact, the larger maximum has the effect to stretch the distribution, increasing both the lack of symmetry (skewness) and the fattened tail (kurtosis).

2.3 Natural Gas Germany

The dataset for accidents in the German natural gas chain covers the time period 1981-2004. However, in contrast to the previously described datasets, the natural gas dataset contains both accidents with at least five fatalities and with minor consequences. Furthermore, this dataset is unique in the sense that complete accident records are not only available for fatalities, but also injured persons and property damages in 2004 monetary value (Euro).

In Fig. 3 the fatality, injury and economic losses distributions are shown.

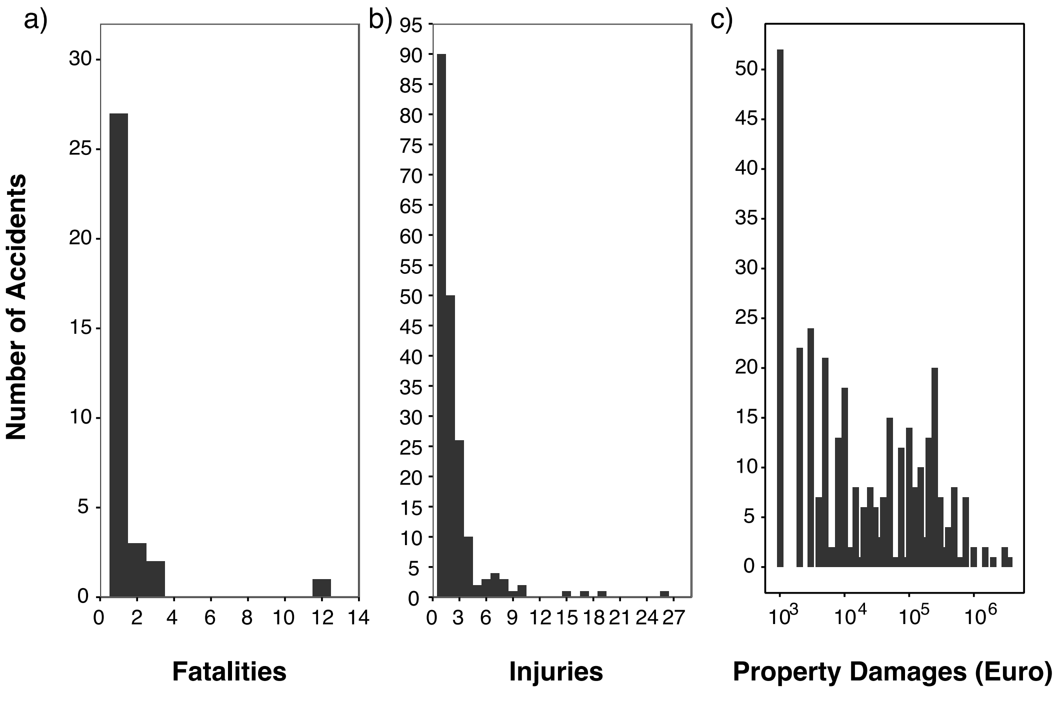

Fig. 3: Number of accidents for different types of consequences in the German natural gas chain for the time period 1981-2004. a) Fatalities; b) Injuries; c) Property Damages (Euro).

In Table 3 the descriptive statistics for each of the analyzed datasets are collected.

Table 3: Descriptive statistics for severe accidents in the German Natural gas Chain in the time period 1981-2004 for different types of consequences, i.e. fatalities, injuries and property damages (Euro).

Consequence Type	Accidents	Total Consequences	Skewness	Kurtosis	Observed Maximum
Fatalities	52	78	5.3	33.2	12
Injuries	340	730	5.3	42.8	26
Property Damages (Euro)	359	5.44E+7	7	62.6	5E+6

For all three consequence indicators, the distributions are right skewed (skewness > 0) with a large kurtosis values (> 3) indicating their leptokurtic nature. Overall, both skewness and kurtosis are slightly different between the different datasets, unless for the skewness for the fatality and injury distributions. However, for injuries the kurtosis is slightly larger than for fatalities. This could be related to the lack of severe accidents for fatalities (≥ 5), where only the observed maximum is observed (12 casualties occurred in 2000), compared to injuries, where the presence of few severe accidents (4 in total with ≥ 10 injuries, including the observed maximum with 26 injuries that happened in 1990) affected the kurtosis (Fig. 3). Furthermore, for the property damages, both skewness and kurtosis are larger than the fatality and injuries distributions. The latter is strongly related to the relative large number of events recorded for property damages with respect to the other two analyzed datasets and the rather large number of events, at least the double, with 1000 euro damages with respect to the other damages levels, which strongly increase both the asymmetry and the fatness of the distribution.

3 Method

In this section an overview of the risk measures is given that were applied to the oil (OECD, EU28, non-OECD), coal China and Natural Gas Germany datasets. Value at Risk (VaR) is compared with the so-called "coherent" risk measures, which in this case are Expected Shortfall (ES) and Spectral Risk Measures (SRM). The concept of coherent was first introduced by [19] in order to present and justify a unified framework for the analysis and structuring of risk measures. The result of the work by [19] was a set of axioms that define the properties, which must be fulfilled by a risk measure to become coherent.

Here the axioms proposed by [19] are briefly summarized since it is not in the aim of this paper to describe them in detail. However, it is important to explain and discuss the concepts and assumptions based on which [19] develop their arguments. According to [19] risk is represented by the change in values between two dates of a financial position, which is the amount of owned or owed items of a trader/investor, or as the variability of the future value of the position due to changes in the market environment. Although these axioms were firstly developed for the financial realm, they could also be applied to non-market risk as well [29].

3.1 Overview of coherent risk measures properties

In order to briefly summarize the 4 axioms described in [19] to define a coherent risk measure, let us denote M as the space of random variables representing portfolio losses over a fixed time interval and L be the loss value for a given portfolio. Furthermore, M is assumed to be a convex cone so that $L_1 \epsilon M$ and $L_2 \epsilon M$, then $L_1 + L_2 \epsilon M$ and $\lambda L_1 \epsilon M$ for every constant $\lambda > 0$. A coherent risk measure $\rho: M \rightarrow \mathbb{R}$ is defined as a function that satisfies the following properties [30]:

- Axiom 1 (Translation Invariance), for all $L \epsilon M$ and every constant $a \epsilon \mathbb{R}$, we have that $\varrho(L + a) = \varrho(L) + a$; which means that adding or subtracting a riskless amount a to the initial amount and investigating it in the reference instrument, simply decreases respectively increases the risk metric by a.

- Axiom 2 (Subadditivity), for all $L_1, L_2 \epsilon M$, we have $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$; which means that the total risk associated to two random variables is equal or lower than the sum of the individual risk of each of the random variables. If an individual participating on an established exchange would wish to take up the risk expressed by the sum of the two random variables $L_1 + L_2$, by employing a risk measure that is not coherent with this axiom, he would simply open two accounts: one for L_1 and another for L_2 since he would benefit from a smaller margin requirement of $\rho(L_1) + \rho(L_2)$.

- Axiom 3 (Positive Homogeneity), for all $L \epsilon M$ and every $\lambda > 0$, we have $\rho(\lambda L) = \lambda\rho(L)$; which means that the risk of a position is proportional to its size [19, 31]. However, by considering Axiom 2, this is somewhat controversial, since it means that for two different portfolio losses, the relation is not the same as for two equal portfolio losses.

- Axiom 4 (Monoticity), for all $L_1, L_2 \epsilon M$ such that $L_2 \leq L_1$, we have $\rho(L_1) \leq \rho(L_2)$; which means that if L_1 has better outcomes than L_2, than the risk associated to L_1 should always be less than the risk associated to L_2.

Once the aforementioned axioms are satisfied, a risk measure is defined to be coherent.

3.2 Value-at-Risk (VaR)

VaR is the most widespread risk measure employed in financial mathematics and financial risk management, as well as within the financial industry [32]. According to [33] the VaR of a portfolio at a specific confidence level a, with $a \epsilon [0,1]$, is formally defined as the smallest number x such that the probability of a random number X to exceed x is not larger than $(1 - a)$. Therefore, the function that formally defines VaR can be described as:

$$VaR_a(X) = \inf\{ x \epsilon \mathbb{R}; P(X > x) \leq 1 - a\} = \inf\{x \epsilon \mathbb{R}; F_x(x) \geq a\} \quad (1)$$

One of the main advantages of VaR is its simplicity in estimation and its probabilistic behaviour. Furthermore, VaR provides a common measure of risk across different exposures and risk factors implying that the risk manager can compare the results of the applied VaR metric computed for different portfolios. On the other hand, the main drawback is that VaR is not a subadditive measure, therefore, could not be considered a coherent risk measure. Furthermore, since VaR is the quantification of a quantile, it does not allow to measure the extent of exceptional losses beyond the quantile of interest

[34]. In other words, in a bad state of the world, e.g., states with extremely low probability, the risk manager might face significantly higher losses than those expressed by the VaR metric.

3.3 Spectral Risk Measures (SRMs)

The SRM was developed in order to fulfill the axioms to be considered a coherent risk measure [11, 13]. In this context, the general formalization of SRM for a risk measure function M_ϕ is given as follow:

$$M_\phi(X) = \int_0^1 \phi(p) q_p dp \tag{2}$$

where q_p is the p loss quantile and $\phi(p)$ is the weighting function specified by the risk manager and defined over the full range of cumulative probabilities $p \epsilon [0,1]$. This function is also known as risk aversion function, where higher weights should be assigned to less desirable outcomes in order to reflect the risk aversion of the risk manager. Furthermore, the definition of it is a crucial element for SRM, since the admissibility of the risk aversion function is strictly linked to the coherency of an SRM [20]. Concisely summarize the three main conditions for coherency of the risk aversion function $\phi(p)$:

- Non-negative $\phi(p) \geq 0$, which prohibits the existence of negative risk weights;
- Normalization $\int_0^1 \phi(p)dp = 1$, which means that all risk weights should sum up to 1;
- Increasingness $\phi'(p) \geq 0$, where its spirit is to dictate an increase in the risk weight associated with larger, unwanted losses.

In this study, an exponential risk aversion function is chosen for the calculation of SRM, since:

- it is non negative for any quantile;
- the weights are increasing as we move towards the higher loss quantiles, thus exhibiting the manager's degree of risk aversion;
- the area underneath the function is 1 and its derivative is non negative.

In mathematical terms [12, 24]:

$$\phi(p) = \frac{ke^{-k(1-p)}}{1 - e^{-k}} \tag{3}$$

where p is the loss quantile and k is defined as the Arrow-Pratt coefficient of absolute risk aversion (ARA). The most important advantages of SRM are that it is a coherent risk measure and enables the risk manager to incorporate his or her risk aversion, resulting in metrics based on different risk profiles. On the other hand, the main limitations are commonly linked to the fact that they require more computational effort and a clear understanding of the mechanics behind the risk measure [13] . In fact, a risk manager should be very careful in choosing a weighting that fits his or her risk profile. This is a delicate step in a risk analysis since ultimately the results of a risk measurement using SRMs depends on the choice of the power weighting function and its parameters.

3.4 Expected Shortfall

The Expected Shortfall (ES) is a coherent risk measure, which is a particular case of SRM. It is a coherent risk measure like SRM, but it overcomes one of the SRM's disadvantages, which is the definition of a risk aversion function and all its possibly related issues. In fact, the ES is estimated as a SRM, but the risk aversion function is given by a constant, which has an inverse proportionality with the quantile of interest. Furthermore, the ES risk measure is sub-additive and better captures the extent of exceptional losses. In fact, ES can be regarded as a better risk measure than VaR as it does not ignore the losses beyond the specified confidence interval, but it averages over them, while also satisfying the subadditivity criterion that acknowledges the benefits of diversification [35]. However, the fact that it assigns equal weights to the loss quantiles does not necessarily reflect the risk aversion of the risk manager. Formally, ES is defined as the average loss in the $100 * (1 - \alpha)\%$, where α is the quantile of interest, worst cases of our distribution. It averages the events to the right of the specified confidence level value and reports it [18]. In mathematical terms, the ES for a given quantile of interest is given by:

$$ES_{\alpha} = \frac{1}{1-\alpha}\int_{\alpha}^{1} q_p dp \tag{4}$$

Therefore, the risk aversion function is defined as:

$$\phi_{ES_{\alpha}}(p) = \frac{1}{1-\alpha} 1_{\{p \leq (1-\alpha)\}} = \begin{cases} \frac{1}{1-\alpha} & if\ p \leq (1-\alpha) \\ 0 & else \end{cases} \tag{5}$$

3.5 Practical Estimation of VaR, ES and SRM

In this study, the estimation of the different risk measures for the datasets described in section 2 are following the so-called historical method, since it is the most common way used in the past to assess risk indicators for accidents in the energy sector [2, 6, 10, 22].

The estimation of VaR is given by equation (1) for different quantiles of interest. However, for ES and SRM, the estimation is not straightforward as pointed out before. Intuitively, both ES and SRM are weighted averages of the loss tail VaRs. Therefore, the best way to implement these measures is by slicing the tail into n parts and averaging over the values estimated for each slice [35]. For example, for a set of n=100, we got 99 loss quantiles for each of which we calculated the respective VaR and, finally, averaged them in order to get the ES value for the tail section. Furthermore, for SRM, we multiplied each of the 99 VaR values with their respective risk weight given by the exponential risk weighting function $\phi(p)$ in order to get the value of the exponential SRM for the same tail section. In this context, the number of slices should then be defined in order to subdivide the loss tail to estimate the SRM (or ES).

Intuitively, as the number of slices gets large enough, the computed value will converge to the true SRM (or ES) value. In this study, the number of slices (n) has been assessed by estimating the so-called most probable break-point, which is the threshold from where the value of the risk metric remains constant with increasing number of slices. In fact, the distribution of the number of slices with respect to the average risk metric shows two distinct patterns:

- at the low resolution case (i.e., low number of n), the increase of the average risk metric is fast with respect to the increase of the number of slices;
- at the high resolution case (i.e., high number of n), the average risk metric is constant with respect to the increase of the number of slices.

In this study, the break-point is estimated by using an iterative procedure to fit segmented relationships in regression models, which use a bootstrap restarting to avoid sensitiveness issues to the starting values [36]. Furthermore, the use of a resampling method (bootstrap) allowed the estimation of the uncertainty for the number of slices at the break-point [37], see Fig. 4.

This method has been applied to all datasets (section 2) for different quantile levels (0.9, 0.95, 0.99 and 0.999) and for different risk aversion factors (k-factors) between 1 and 2000 for both ES (Fig. 4a) and SRM, (Fig. 4b). In both Fig.s the minimum number of slices varies case by case. Some cases (e.g., Fig. 4a: Oil OECD and Fig. 4b: Germany Natural Gas Property Damages (Euro)) show higher numbers of slices for quantiles or k-factors not at the end of the tail, resulting in a peak of number of slices. Therefore, a conservative value of 30000 slices has been chosen in this study for all the aforementioned risk metrics and datasets in order to avoid these peaks.

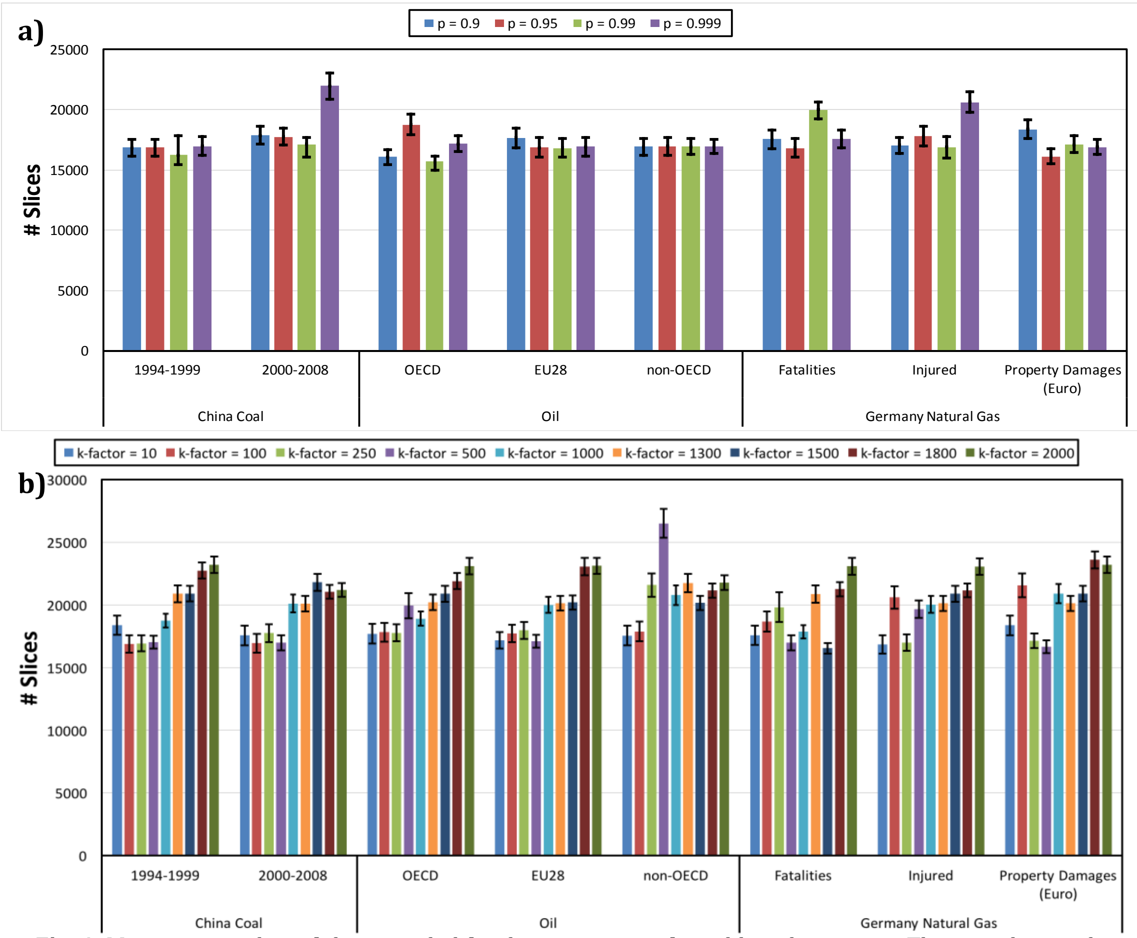

Fig. 4: Minimum number of slices needed for the estimation of a stable risk measure. The error bars indicate the uncertainty assessed for the estimated stable risk measure using a bootstrapping approach. a) ES for different probability levels and different Energy Chains; b) SRM for different risk aversion factors (k-factor) and different Energy Chains.

4 Results & Discussion

An overview of the various datasets for the three different risk measures in comparison with the maximum consequence is given in Fig. 5. In the case of VaR and the ES estimates for different quantiles (i.e., 0.9, 0.95, 0.99 and 0.999) were calculated, in order to better understand the differences between the two. For the SRM, which considers an exponential driven risk aversion function, different k-factors or risk aversion levels were used (i.e., 50, 500 and 1500). This allows comparing different levels of stability along the entire risk aversion function.

In absolute values, by comparing the results for the different probability levels, it is clear that the VaR is lower than the ES for p = 0.9, p = 0.95 and p =0.99, and tends to converge to the maximum consequence and to the ES values the closer p gets to 1. In fact, the VaR is a quantile of the distribution at a given probability level, and ask the question "*how bad can things get?*", while the ES wants to answer the question "*if things do get bad, what is the expected loss?*". These are two different ways to approach the risk. In the former we get the maximum value at which we could be safe from a risk point of view, while in the second we are getting the average, or expected loss, if we are going beyond the VaR. This is somehow expected since the ES is calculated as the average value of the risk after a certain quantile, while the VaR is roughly calculating the quantile value of the distribution. The higher the quantile value of interest becomes (i.e., closer to 1), the more the VaR converges to the ES and to the maximum consequence. This is of course of great interest for decision-makers and other stakeholders dealing with energy risks because the choice of risk measure may influence the decision process. For the most extreme risks (p close to 1) VaR and ES are interchangeable, i.e. the results are not depending on the risk measure, but still lower than the maximum consequence. However, for more probable cases (p between 0.9 and 0.99) the values for ES are clearly larger than for VaR suggesting that VaR tends to underestimate the risk with respect to ES. This is related to the fact that ES includes information on the expectation of consequences in case of an extreme case accident, while VaR is not. Furthermore, ES can provide a more conservative estimate to avoid underestimating the risk [15].

On the other hand, the SRM values at the considered risk aversion levels correspond to the ES results between p=0.95 and p=0.99 for the k-factor = 50, while for the k-factors 500 and 1500, results are similar to the ES and VaR results at p=0.999 and, thus, closer to the maximum consequence. It is interesting to see that the relative difference in all cases of the SRM values is significantly larger between k=50 and k=500 than between k=500 and k=1500. This indicates that the SRM values are strongly driven by the risk aversion function. In fact, by choosing an exponential function as weighting function, this tends to give larger differences at the beginning of the distribution, while the values are more similar at the tail. Therefore, using k=500 or k=1500 as risk aversion level does not change significantly the results. Finally, in all cases, but in particular for Oil non-OECD, the SRM constructed with an exponential weighting function is the risk measure that signals the highest risk. The choice of k>500 for SRM implies an estimation of risk similar to the ES at p=0.999, so it can be employed by stakeholders with a higher risk aversion. An advantage of this risk measure is also that it takes into account the entire distribution of fatalities and it is flexible enough to adapt to different risk aversion levels and stakeholder preferences.

As described before, the absolute values are different, due to the fact that VaR tends to underestimate the risk with respect to ES and SRM. However, from Fig. 5 it is clear that the three risk measures show similar results in terms of their relative ranking for all energy chain datasets. In case of fatalities in Chinese coal mines, for low probability levels (e.g., p=0.9), VaR and ES for the time period 1994-1999 are higher than for 2000-2008. However, on the other hand for larger probabilities (e.g., p=0.999), the fatality indicators for VaR, ES and SRM are higher in the time period 2000-2008 than for 1994-1999 in accordance with the maximum consequence. This result indicates that for extreme events, the China coal mines safety condition has not been improved, but it clearly has for more expected events. 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 potential consequences could have been reduced, but at the same time the potential consequences in case of an extreme event could be more severe due to the larger number of workers present in these mines [27]. This is important on an energy-business point of view, since clearly show were the threshold for the risk willing to take could be considered, meaning more towards more expected accidents than extreme ones.

In the case of the Oil chain, for more expected accidents (e.g., p=0.9), the three risk measures show a similar behavior among the analyzed country groups. The non-OECD country group results to be of higher risk with respect to OECD and EU28, which are comparable. However, it is interesting to see that the relative difference between non-OECD and EU28 countries with non-OECD is larger in the case of ES (or SRM) with respect to VaR. This is possibly related to the fact that, as discussed before, VaR tends to underestimate the risk with respect to ES. Furthermore, for more extreme cases (p=0.999), while the non-OECD country group show a relative higher risk compared to OECD and EU28, the former results to be larger than the latter in accordance with the maximum consequence indicator. This result indicates that for the Oil chain, as larger the p level for an accident is, the larger is the accident risk for OECD with respect to EU28. In other words, while for more expected accidents the risk between OECD and EU28 is similar, for extreme cases the OECD countries performs worst due to the relative higher OECD maximum consequence with respect to the EU28 one. In fact, the former tend to increase the skewness of the distribution more than for the EU28 case affecting the considered risk measures. Furthermore, in the non-OECD cases, it is clear that, although they always perform worse than OECD and EU28 indicating a need of safety regulation improvements, the larger relative difference is shown for extreme case scenarios. This result could be of great interest for energy-related business willing to understand the acceptable risk for fatalities related accidents in the Oil chain. Furthermore, it is clear that an acceptable risk moves toward a more expected accident (p=0.9, 0.95) than an extreme one, mostly in the case of non-OECD countries were the relative difference between different p-levels is significant.

Finally, for the Natural Gas case, a direct comparison cannot be made, since the indicators are of different scales. Therefore, no conclusions related to the relative ranking could be made.

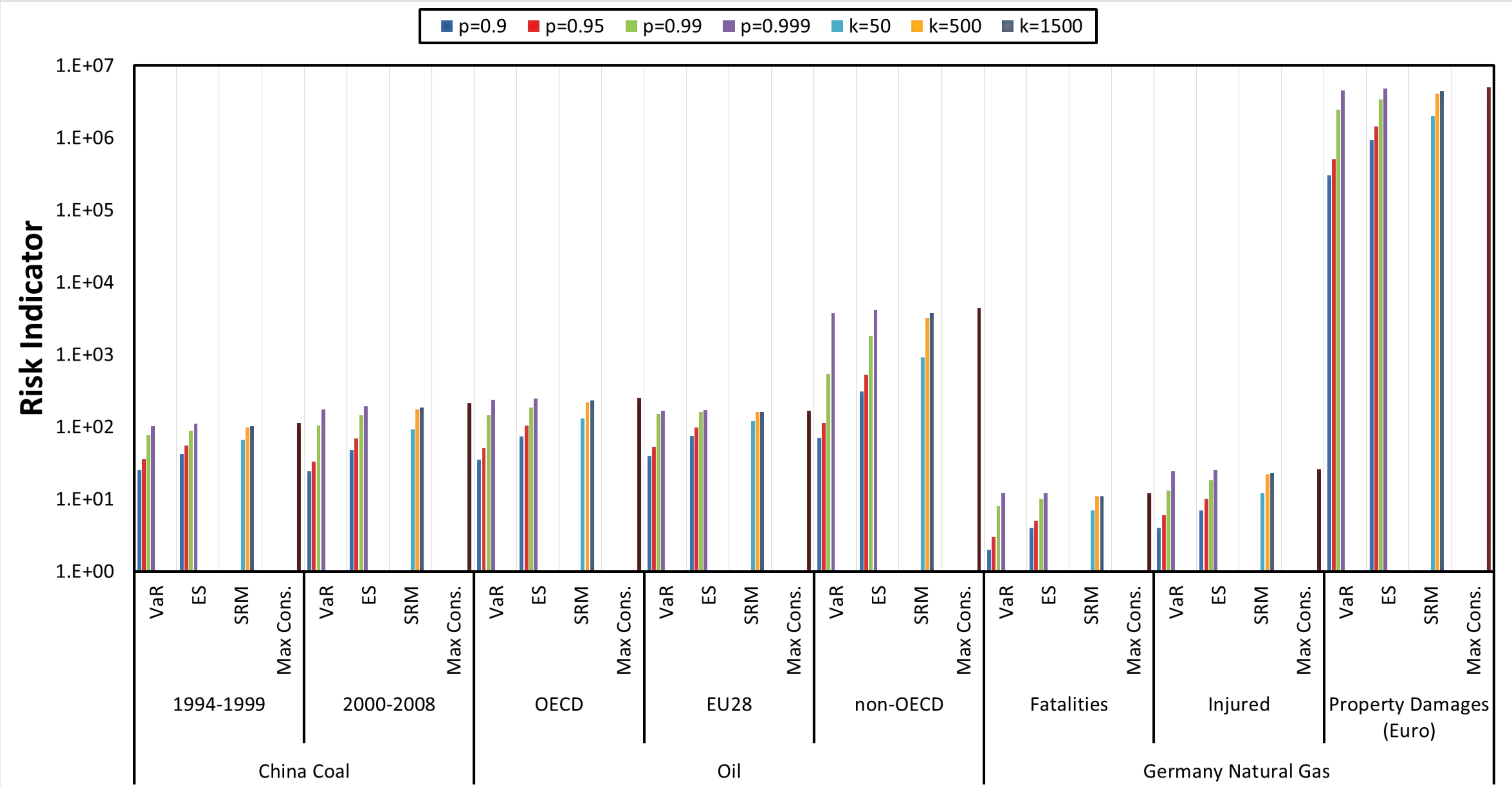

Fig. 5: Maximum Consequence, Value-at-Risk (VaR) and Expected Shortfall (ES) for different probability levels and Spectral Risk Measures (SRM) for different risk aversion factors (k-factor) for the various energy chain datasets analysed in this study.

5 Conclusions

In this paper we applied and compared coherent (ES, SRM) and non-coherent (VaR) risk measures, commonly used in the financial realm, for accident risk in the energy sector. For this purpose, the so-called historical method has been used, since it is a common practice to estimate risk indicators for accidents in the energy sector based on historical observations collected in the PSI's ENSAD database [2, 6, 10, 22].

The VaR, ES and SRM risk metrics analyzed in this study indicate stable relative rankings in terms of risk measures for a given probability level (or k-factor), validating the applicability of these risk measures in accident risk in the energy sector. Although the relative ranking results are similar, the absolute values are not. In general, VaR tends to underestimate the risk with respect to ES (and SRM), since the latter includes information on the expectation of consequences in case of an extreme accident. This is of great interest for stakeholders and decision makers in general. In fact, in the energy sector and in particular in the context of energy security, underestimating the risk of an accidents at a critical infrastructure could lead to large consequences in terms of human health, environmental impact and economic losses in case something will happen. In this context, these might be mitigated by the use of ES (or SRM), since they will help in the improvement, for example, of the safety measures in preparation of a larger event than what expected by the VaR, or by the increase of investments at an energy critical infrastructure that enhance prevention and response induced by the insurance industry. However, while ES can be considered a good alternative with respect to VaR, the SRM is strongly affected by the risk aversion, and thus, to the definition of it. This risk measure is a tool flexible enough to assess risk and can be adapted to individual preferences and risk aversion levels of stakeholders. However, this risk metric implies a higher degree of sophistication and poses more challenges for the implementation, which still limits its application in practice.

Finally, with respect to the commonly used maximum credible consequence indicator, the risk metrics considered in this study could improve the information that an energy-related business can get. In fact, VaR, ES and SRM could help to facilitate a better understanding of energy accident risks to the industry, due to the fact that they normally focus on financial and less infrastructure-related aspects. Furthermore, although both the maximum credible consequence and VaR are unique values for the risk, they have a different nature. The former is defined by the maximum historical observation, thus no complete conclusions could be made from it (e.g., is it an extreme event? Could an event overcome it? Can be considered as an acceptable risk threshold?). VaR defines a threshold, which does not tell a stakeholder what is the risk beyond it, but could help him during decision-making processes by considering a maximum risk that he is willing to take. In case of ES and SRM, the discussion is different. In both cases, the improvements with respect to the information given by the maximum credible consequence are significant. In fact, on the one hand, both measures intrinsically include the risk aversion and, on the other hand, while ES measures the risk beyond a certain threshold, SRM is able to model the entire spectrum of it, giving the full picture.

References

[1] Fritzsche AF. The Health Risks of Energy Production. Risk Analysis. 1989;9(4):565-77.

[2] Burgherr P, Hirschberg S. Comparative risk assessment of severe accidents in the energy sector. Energy Policy. 2014;74, Supplement 1:S45-S56.

[3] Greenberg M, Truelove HB. Energy Choices and Risk Beliefs: Is It Just Global Warming and Fear of a Nuclear Power Plant Accident? Risk Analysis. 2011;31(5):819-31.

[4] Rasmussen NC. The Application of Probabilistic Risk Assessment Techniques to Energy Technologies. Annual Review of Energy. 1981;6(1):123-38.

[5] Eckle P, Burgherr P. Bayesian Data Analysis of Severe Fatal Accident Risk in the Oil Chain. Risk Analysis. 2013;33(1):146-60.

[6] Burgherr P, Hirschberg S. A Comparative Analysis of Accident Risks in Fossil, Hydro, and Nuclear Energy Chains. Hum Ecol Risk Assess. 2008;14(5):947-73.

[7] Jonkman SN, van Gelder PH, Vrijling JK. An overview of quantitative risk measures for loss of life and economic damage. Journal of hazardous materials. 2003;99(1):1-30.

[8] Burgherr P, Eckle P, Hirschberg S. Severe accidents in the context of energy security and critical infrastructure protection. In: Ale BJM, Papazoglou IA, Zio E, editors. Reliability, Risk and Safety - Back to the Future. London, UK: Taylor & Francis Group; 2010. p. 456-64.

[9] Thomas PJ. Measuring risk-aversion: The challenge. Measurement. 2016;79:285-301.

[10] Burgherr P, Hirschberg S, Spada M. Comparative Assessment of Accident Risks in the Energy Sector. In: Kovacevic RM, Pflug GC, Vespucci MT, editors. Handbook of Risk Management in Energy Production and Trading: Springer US; 2013. p. 475-501.

[11] Wächter HP, Mazzoni T. Consistent modeling of risk averse behavior with spectral risk measures. European Journal of Operational Research. 2013;229(2):487-95.

[12] Precht CM. Using spectral risk measures for determining capital requirements for banks' trading books. St. Gallen, Switzerland: University of St. Gallen, 2012.

[13] Dowd K, Cotter J, Sorwar G. Spectral Risk Measures: Properties and Limitations. J Finan Serv Res. 2008;34(1):61-75.

[14] Acerbi C. Spectral measures of risk: A coherent representation of subjective risk aversion. J Bank Financ. 2002;26(7):1505-18.

[15] Dick A. Value at Risk and Expected Shortfall: Traditional Measures and Extreme Value Theory Enhancements with a South African Market Application. Matieland, South Africa: University of Stellenbosch, 2013.

[16] Hoogerheide LF, van Dijk HK. Bayesian Forecasting of Value at Risk and Expected Shortfall using Adaptive Importance Sampling. Amsterdam, Netherlands: Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, and Tinbergen Institute; 2008. p. 38.

[17] Risk Magazine. VAR versus expected shortfall. Risk. http://www.risk.net/risk-magazine: Incisive Risk Information (IP) Limited; 2007.

[18] Acerbi C, Tasche D. Expected Shortfall: A Natural Coherent Alternative to Value at Risk. Economic Notes. 2002;31(2):379-88.

[19] Artzner P, Delbaen F, Eber J-M, Heath D. Coherent Measures of Risk. Mathematical Finance. 1999;9(3):203-28.

[20] Acerbi C. Coherent representations of subjective risk-aversion. In: Szegö G, editor. Risk Measures for the 21st Century. Chicester, UK: John Wiley & Sons; 2004. p. 147.

[21] Acerbi C, Tasche D. On the coherence of expected shortfall. J Bank Financ. 2002;26(7):1487-503.

[22] Burgherr P, Eckle P, Hirschberg S. Comparative assessment of severe accident risks in the coal, oil and natural gas chains. Reliability Engineering & System Safety. 2012;105:97-103.

[23] Paraschiv F, Mudry P-A, Andries AM. Stress-testing for portfolios of commodity futures. Economic Modelling. 2015;50:9-18.

[24] Rif AS. Risk Measures for Extreme Events [Master Thesis]. St. Gallen, Switzerland: University of St. Gallen, 2014.

[25] Hirschberg S, Spiekerman G, Dones R. Severe accidents in the energy sector. Villigen PSI, Switzerland: Paul Scherrer Institut; 1998. p. 340.

[26] Burgherr P, Hirschberg S. Comparative assessment of natural gas accident risks. Villigen PSI, Switzerland: Paul Scherrer Institut; 2005. p. 53.

[27] Spada M, Burgherr P, Hirschberg S. Comparative Assessment of Severe Accidents Risk in the Energy Sector: Uncertainty Estimation Using a Combination of Weighting Tree and Bayesian Hierarchical Models. Conference Comparative Assessment of Severe Accidents Risk in the Energy Sector: Uncertainty Estimation Using a Combination of Weighting Tree and Bayesian Hierarchical Models, Honolulu, HI, USA.

[28] Burgherr P, Hirschberg S. Assessment of severe accident risks in the Chinese coal chain. International Journal of Risk Assessment and Management. 2007;7(8):1157-75.

[29] Delbaen F. Coherent Risk Measures. Pisa, Italy 2000. p. 36.

[30] Haugh M. Risk Measures, Risk Aggregation and Capital Allocation. IEOR E4602 Quantitative Risk Management. New York, NY, USA: Columbia University; 2010.

[31] Artzner P, Delbaen F, Eber J-M, Heath D. Thinking Coherently. Risk. 1997;10:68-71.

[32] Jorion P. Value at Risk. New York, USA: McGraw-Hill, 2006.

[33] McNeil AJ, Frey R, Embrechts P. Quantitative Risk Management: Concepts, Techniques and Tools. Revised Edition ed. Princeton, NJ, USA: Princeton University Press, 2015.

[34] Alexander C. Market Risk Analysis IV: Value at Risk models. Chicester, UK: John Wiley & Sons, 2008.

[35] Dowd K. Measuring Market Risk. Chicester: John Wiley & Sons, 2005.

[36] Muggeo VMR, Adelfio G. Efficient change point detection for genomic sequences of continuous measurements. Bioinformatics. 2011;27(2):161-6.

[37] Wood SN. Minimizing Model Fitting Objectives That Contain Spurious Local Minima by Bootstrap Restarting. Biometrics. 2001;57(1):240-4.