

Of Disasters and Dragon Kings: A Statistical Analysis of Nuclear Power Incidents and Accidents

Spencer Wheatley,^{1,*} Benjamin Sovacool,^{2,3} and Didier Sornette¹

We perform a statistical study of risk in nuclear energy systems. This study provides and analyzes a data set that is twice the size of the previous best data set on nuclear incidents and accidents, comparing three measures of severity: the industry standard International Nuclear Event Scale, the Nuclear Accident Magnitude Scale of radiation release, and cost in U.S. dollars. The rate of nuclear accidents with cost above 20 MM 2013 USD, per reactor per year, has decreased from the 1970s until the present time. Along the way, the rate dropped significantly after Chernobyl (April 1986) and is expected to be roughly stable around a level of 0.003, suggesting an average of just over one event per year across the current global fleet. The distribution of costs appears to have changed following the Three Mile Island major accident (March 1979). The median cost became approximately 3.5 times smaller, but an extremely heavy tail emerged, being well described by a Pareto distribution with parameter $\alpha = 0.5\text{--}0.6$. For instance, the cost of the two largest events, Chernobyl and Fukushima (March 2011), is equal to nearly five times the sum of the 173 other events. We also document a significant runaway disaster regime in both radiation release and cost data, which we associate with the “dragon-king” phenomenon. Since the major accident at Fukushima (March 2011) occurred recently, we are unable to quantify an impact of the industry response to this disaster. Excluding such improvements, in terms of costs, our range of models suggests that there is presently a 50% chance that (i) a Fukushima event (or larger) occurs every 60–150 years, and (ii) that a Three Mile Island event (or larger) occurs every 10–20 years. Further—even assuming that it is no longer possible to suffer an event more costly than Chernobyl or Fukushima—the expected annual cost and its standard error bracket the cost of a new plant. This highlights the importance of improvements not only immediately following Fukushima, but also deeper improvements to effectively exclude the possibility of “dragon-king” disasters. Finally, we find that the International Nuclear Event Scale (INES) is inconsistent in terms of both cost and radiation released. To be consistent with cost data, the Chernobyl and Fukushima disasters would need to have between an INES level of 10 and 11, rather than the maximum of 7.

KEY WORDS: Nuclear incidents and accidents; quantitative risk analysis; safety analysis

1. INTRODUCTION

The industry-standard approach to the evaluation of the risk of nuclear accidents is a top-down technique called probabilistic safety assessment (PSA). PSA consists of developing fault tree models that allow one to simulate accidents, with different triggers and event paths, and the severity and frequency of such accidents.⁽¹⁾ Furthermore, within a plant, PSA may be an ongoing process where

¹ETH Zurich, Department of Management, Technology and Economics, Switzerland.

²Science Policy Research Unit, University of Sussex, UK.

³Center for Energy Technologies and Department of Business and Technology, Aarhus University, Denmark.

*Address correspondence to Spencer Wheatley, ETH Zurich, Department of Management, Technology and Economics, Switzerland; swheatley@ethz.ch.

both the PSA, and plant operations and technology, evolve together with the purpose of improving plant safety. The basic PSA methodology works as follows. Initiating events, such as component failures, human errors, and external events, are enumerated and assigned probabilities. Next, a (typically deterministic) fault tree is defined to encode the causal links between events, allowing combinations of initiating events to form the ultimate resultant/system-level event. Such a model then allows one to determine the probability of such events, and potentially attach damage/consequence values to the event paths. Thus, a textbook PSA would require the complete and correct definition of initiating events, subsequent cascade effects, and their probabilities and consequences.

It is therefore not surprising that the documentation for a plant-specific PSA often fills a bookshelf, and is a constant “work in progress.” Within PSA, three levels exist, delineating the depth/extent to which events are studied:^(2–4) level 1 concerns core damage events, level 2 concerns radioactive releases from the reactor building given that an accident has occurred, and level 3 evaluates the impact of such releases on the public and the environment. Levels 1 and 2 are required by regulation. Level 3, which is the level considered in this study, is seldom done in PSA. Given that the reliability of PSA depends on the inclusiveness of scenarios, the correct modeling of cascade effects, and the handling of tremendous uncertainties, it is not surprising that PSA has failed to anticipate a number of historical accidents in civil nuclear energy.^(5,6) It has been found that the probability assessments were fraught with unrealistic assumptions,⁽⁷⁾ severely underestimating the probability of accidents. The chairman of the World Association of Nuclear Operators stated that the nuclear industry is overconfident when evaluating risk and that the severity of accidents is often underreported.⁽⁸⁾

Instead of entering this quagmire, several studies have used a “bottom-up” approach, performing statistical analysis of historical data. These studies,^(5,9–12) and others, have almost universally found that PSA dramatically underestimates the risk of accidents. The International Atomic Energy Agency (IAEA) provides the International Nuclear Event Scale (INES) measure of accident severity, which is the standard scale used to measure the severity of nuclear accidents. However, the INES has been censured—for being crude, inconsistent, only available for a small number of events, etc.—not only

in statistical studies, but by the industry itself.^(8,13) As noted by *The Guardian* newspaper, it is indeed remarkable (*sic* astonishing) that the IAEA does not publish a historical database of INES events.⁽¹⁴⁾ However, given that the IAEA has the dual objective of promoting and regulating the use of nuclear energy, one should not take the full objectivity of the INES data for granted. Independent studies are necessary to avoid possible conflicts of interest associated with misaligned incentives.

Presumably for lack of better data sources, a number of statistical studies^(11,12) have used the INES data to make statements about both the severity and frequency of accidents in nuclear energy systems. Here, we also perform a statistical analysis of nuclear incidents and accidents, but we avoid relying on the INES data. Instead, we use the estimated cost value in USD (U.S. dollars) as the common metric that allows one to compare often very different types of events. This database has over triple the number of events compared with most studies, providing a much better basis for statistical analysis and inference, and bringing into question the reliability of the other studies. Moreover, because radiation releases may translate into very different levels and spread of contamination of the biosphere, depending on local circumstances, the quantification of cost is more useful and provides a better comparative tool.

According to PSA specialists, the gaps between PSA-specific results and the global statistical data analysis mentioned above exist in the eyes of observers who ignore the limitations in scope that apply to almost all PSA—e.g., PSA is often restricted to normal operating conditions and internal initiating events. Indeed, PSA is a tool that serves many purposes that do not rely on the accurate absolute quantification of risks. However, PSA *is* used as the tool for discussing risks in nuclear energy systems, and has multiple shortcomings in this regard. That is, PSA applications need to better consider incompleteness, uncertainty,^(6,15–17) and be combined with bottom-up statistical approaches when discussing risks at many levels.⁽¹⁸⁾

Moreover, because of the uniqueness of each reactor, some nuclear experts say that assigning risk to a particular nuclear power plant is impossible.^(3,19) A further argument is that the series of accidents form a nonstationary series; in particular because the industry has been continuously learning from past accidents, implementing procedures, and upgrading each time to fix the problem when a vulnerability was found—especially when revealed by an accident.

For instance:⁽²⁰⁾ the loss of criticality control in the fast breeder reactor EBR-I (1.7MWe), which started operation in 1951 on a test site in the Idaho desert, led to a mandatory reactor design principle to always provide a negative power coefficient of reactivity when a reactor is producing power; the Windscale accident in 1957 catalyzed the establishment of the general concept of multiple barriers to prevent radioactive releases; the Three Mile Island accident in 1979 led to plant-specific full-scope control room simulators, plant-specific PSA models for finding and eliminating risks, and new sets of emergency operating instructions; the Chernobyl accident in 1987 led to the creation of the World Association of Nuclear Operators (WANO) through which participating operators exchange lessons learned, and best practices; the Fukushima-Daiichi accident in 2011 is pushing toward designs that ensure heat removal without any AC power for extended times, etc. As a consequence, each new accident supposedly occurs at a nuclear plant that is not exactly the same as for the previous accident. This leads to the concept that nuclear risks are unknowable because one does not have a fixed reference frame to establish reliable statistics.⁽²¹⁾

In contrast, we propose that it is appropriate—and important—to study the global risk of nuclear energy systems by performing statistical analysis of an independently compiled data set of the global pool of reactors. There is nothing invalid about modeling the overall risk of the heterogeneous global fleet, provided one takes sufficient care to control for nonstationarity, and does not draw inference beyond the “average reactor.” In particular, risk-theoretic stochastic models aiming at describing both the frequency and severity of events, as already used in some studies,^(5,9) offer very useful guidelines for such statistical analyses. This constitutes the standard approach that insurance companies rely upon when quoting prices to cover the risk of their clients, even when the estimation of risk appears very difficult and nonstationary. In this spirit, Burgherr *et al.*⁽²²⁾ write that “the comparative assessment of accident risks is a key component in a holistic evaluation of energy security aspects and sustainability performance associated with our current and future energy system.”

In the next section, we describe the data used in our analyses, how severity of events in nuclear energy systems can be measured, and show that the INES values are a poor measure of severity when compared with the consequences of events measured in USD cost. Section 3 discusses uncertainty quantification in nuclear risks. Section 4 estimates the rate

of events, and proposes simple models to account for the evolution of the nuclear plant industry. Section 5 analyzes the distribution of costs. Section 6 discusses a runaway disaster effect, where the largest events are outliers referred to as “dragon kings” (DK). Section 7 combines the different empirical analyses of previous sections on the rate of events, the severity distribution, and the identification of the DK regime, to model the total future cost distribution, and to determine the expected annual cost. Section 8 concludes and discusses policy implications.

2. DATA AND THE MEASUREMENT OF EVENT SEVERITY

We define an “event” as an incident or accident within the nuclear energy system that had material relevance to safety, caused property damage, or resulted in human harm. The nuclear energy system includes nuclear power plants, as well as the facilities used in its fuel cycle (uranium mines, transportation by truck or pipeline, enrichment facilities, manufacturing plants, disposal facilities, etc.). Events are defined to be independent in the sense that one event does not trigger another one. For instance, three reactors melted down at Fukushima in 2011; however, we define this as a single accident due to the fact that the occurrences at the individual reactors were part of a dependent sequence, and linked to a common cause. Statistical changes due to industry responses to past accidents are controlled for in the modeling.

With this definition, we compiled an original database of as many events as possible over the period 1950 to 2014. To be included in the database, an accident had to be verified by a published source, some of them reported in the peer-reviewed literature, but others coming from press releases, project documents, public utility commission filings, reports, and newspaper articles. Such an incremental approach to database building has been widely utilized in the peer-reviewed energy studies literature. Hirschberg *et al.* have constructed the ENergy-related Severe Accidents Database (ENSAD), the most comprehensive database worldwide covering accidents in the energy sector.⁽²³⁾ Flyvbjerg *et al.* built their own sample of 258 transportation infrastructure projects worth about 90 billion USD.^(24,25) Ansar *et al.*⁽²⁶⁾ built their own database of 45 large dams in 65 different countries to assess cost overruns. Also investigating cost overruns, Sovacool *et al.*^(27,28) compiled a database consisting of 401 electricity projects built between 1936 and 2014 in 57 countries,

which constituted 325,515 megawatts (MW) of installed capacity and 8,495 kilometers of transmission lines.

The data set includes three different measures of accident severity: the industry standard measure, INES; a logarithmic measure of radiation release, NAMS (Nuclear Accident Magnitude Scale);⁽¹⁰⁾ and the consequences of accidents measured in 2013 U.S. dollars (USD). The industry standard measure is the discrete seven-point INES, defined as:⁽²⁹⁾ level 0: events without safety significance, level 1: anomaly, level 2: incident, level 3: serious incident, level 4: accident with local consequences, level 5: accident with wider consequences, level 6: serious accident, and level 7: major accident. Levels 1–3 are considered to be “incidents,” and levels 4–7 “accidents.” The distinction between incidents and accidents is not clear, and thus somewhat arbitrary (e.g., see page 152 of the INES user manual⁽²⁹⁾). Incidents tend to concern degradation of safety systems, and may extend to include events where radiation was released and people were impacted. However, when the damage and impact to people and the environment becomes large enough, then the event is deemed an “accident.” But, there are rules about how many fatalities, or how much radiation release, is necessary to qualify for a specific INES level.

The second measure, NAMS, was proposed as an objective and continuous alternative to INES.⁽¹⁰⁾ The NAMS magnitude is $M = \log_{10}(20R)$ where R is the amount of radiation released in terabecquerels. The constant 20 makes NAMS approximately match INES in terms of its radiation level definitions.

Finally, the main measure used here is the USD consequences/costs due to an event. This cost measure is intended to encompass total economic losses, such as destruction of property, emergency response, environmental remediation, evacuation, lost product, fines, court and insurance claims, etc. In the case where there was a loss of life, we added a lost “value of statistical life” of 6 MM USD per death. The 6 MM USD figure is chosen as a lower bound of the value of statistical life reported by various U.S. agencies (e.g., the Environmental Protection Agency, Food and Drug Administration, Transportation Department, etc.).⁽³⁰⁾ Practically speaking, given that the costs are taken from different sources, it is unlikely that the data truly reflect all relevant costs. While imperfect and controversial, this has the advantage of leading to a single USD metric associated to each event that combines all possible negative effects of the accidents. The costs were standardized by using

historical currency exchange rates, and adjusting for inflation to 2013 USD. Adjusting for differing price levels (e.g., because an equivalent event in Switzerland will cost more than one in the Ukraine) was not done because the majority of events belong to countries with a similar price level (the United States, the United Kingdom, Japan, and Western Europe), and because the sample is so heavy tailed that adjusting cost within an order of magnitude has little impact on the statistics.

The result of this effort is a unique data set containing 216 events. Of these events 175 have cost values, 104 have INES values, and 33 have NAMS values.⁽¹⁰⁾ The data sets of Sovacool from the energy studies literature (e.g., a data set⁽³¹⁾ has been studied^(5,9)) provided a starting point of around 100 events with cost values. For our data set, Table I lists the 15 most costly events. The data and severity measures are discussed in the following subsection. The data set has been published online,⁽³²⁾ where the public is encouraged to review and recommend additions and modifications, with the intention of continually expanding and improving the quality of the data. We believe that this is very important for the proper understanding of risk in nuclear energy. The cost and INES scores are plotted over time in Fig. 1. The frequency with which events exceed the threshold of 20MM USD, and the distribution of these excesses, will be studied in Sections 4 and 5, respectively. Further, how these quantities have changed in response to the major accidents of Three Mile Island, 1979, and Chernobyl, 1986, will be studied. There are likely to be changes following the major accident at Fukushima in 2011. However, there has been little time to observe improvements, and the cost data are most incomplete in this area: the cost of 18 of the 29 post-Fukushima events contained in the data set are, as of yet, unknown. Thus, the industry response to Fukushima cannot be quantified in this study.

This data set dwarfs existing data sets from the academic literature, and is mature enough to justify an analysis. However, it is still important to consider the quality of the data, and what methodology is best to handle the limitations. Regarding data completeness, in a rare statistical statement, the IAEA stated: “During the period July 1995–June 1996, the Agency received and disseminated information relating to 73 events—64 at NPPs [nuclear power plants] and nine at other nuclear facilities. Of those 73 events, 32 were stated to be ‘below scale’ (i.e., safety-relevant but of no safety significance) and three to be ‘out of scale’ (i.e., of no safety relevance). Of the remaining 38

Table I. The 15 Largest Cost Events Since 1960 Are Provided with the Date, Location, Cost in MM 2013USD, INES Value, and NAMS Value

Date	Location	Cost (MM USD)	INES	NAMS
1986-04-26	Chernobyl, Ukraine	259,336	7	8.0
2011-03-11	Fukushima, Japan	166,089	7	7.5
1995-12-08	Tsuruga, Japan	15,500	–	–
1979-03-28	TMI, Pennsylvania, United States USA	10,910	5	7.9
1977-01-01	Beloyarsk, USSR	3,500	5	–
1969-10-12	Sellafield, UK	2,500	4	2.3
1985-03-09	Athens, Alabama, USA	2,114	–	–
1977-02-22	Jaslovske Bohunice, Czechoslovakia	1,965	4	–
1968-05-01	Sellafield, UK	1,900	4	4.0
1971-03-19	Sellafield, UK	1,330	3	3.2
1986-04-11	Plymouth, Massachusetts, USA	1,157	–	–
1967-05-01	Chapelcross, UK	1,100	4	–
1982-09-09	Chernobyl, Ukraine	1,100	5	–
1983-08-01	Pickering, Canada	1,000	–	–
1973-09-26	Sellafield, UK	990	4	2.0

Notes: The full data set is provided online.⁽³²⁾ Unknown values are indicated with a dash.

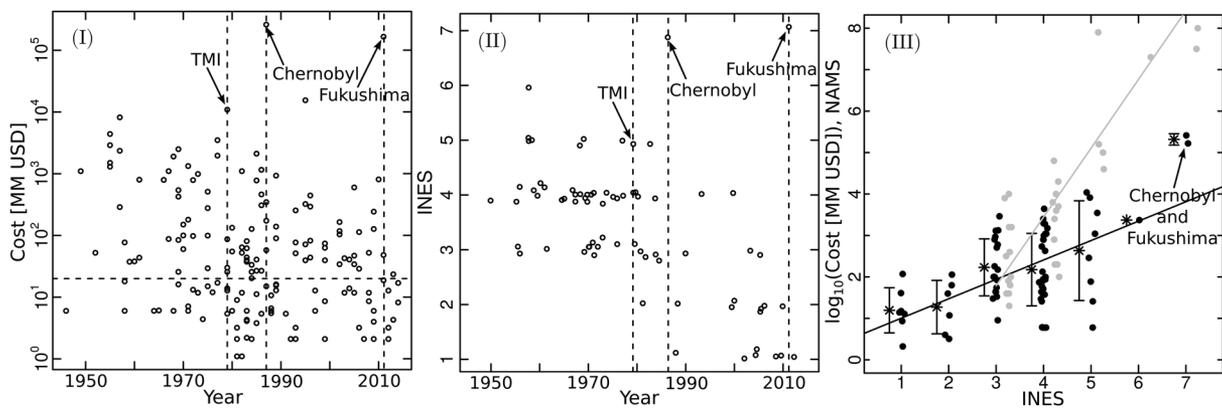


Fig. 1. Figures concerning the raw data. Panel I plots the cost of events over time. The horizontal line is the 20 MM USD cutoff. The vertical lines indicate the occurrence times of TMI, Chernobyl, and Fukushima. Panel II plots the INES scores of events over time. The vertical lines again indicate the occurrence times of TMI, Chernobyl, and Fukushima. Panel III is a scatterplot of the INES scores and cost of events (black), and the INES scores and NAMS values of events (gray). The star and black vertical lines give the mean and standard errors of the logarithmic costs. The points have been shuffled (horizontally) around their integer INES score for visibility.

events, three were rated at INES level 3 and eight at level 2 (i.e., as ‘incidents’), and 27 at level 1 (i.e., as ‘anomalies’).⁽³³⁾ On the other hand, in the data set of this study, only six events fall within this period, none of whose INES values are known, rather than the 38. This statistic tells two important things. First, for increasingly small event sizes, our data are increasingly incomplete—the smaller the event, the less likely it is to be discovered, recorded, reported to the regulator, reported in the media, etc. Second, our data will remain incomplete for small events until the IAEA publishes historical INES data. The shortage of events of INES level 1, and even 2, prohibits a

“near miss” analysis. That is, given that every event develops from having INES score 0, then to 1, and so on, it would be interesting to determine the probability of developing to the next level.

However, the statistics of the largest costs are much more interesting. For instance, the most costly event (Chernobyl) has cost roughly equal to the sum of all other events together, the two most costly events (Chernobyl and Fukushima) have cost roughly five times the sum of all other events together, and the sum of the 53 costs in excess of 100 MM USD is 99.4% of the total cost of all 175 cost values. This clearly implies that, if one wants to study

the aggregate risk in terms of total cost, then one simply needs data for the largest events. Thus a typical approach taken in such cases is to study events that are in excess of a threshold, above which the data are thought to be reliable and well modeled. As in a former study by Hofert,⁽⁹⁾ we use a threshold of 20MM USD, although results will be similar if a threshold of, e.g., 100 MM were used. For cost, there are 101 values above the threshold of 20 MM. Further, of the 41 events with unknown cost, 31 have known INES scores, of which half are level 2 or higher.

2.1. Comparing Severity Measures and Critiquing INES

There are many ways to quantify the size of a nuclear accident. Following Chernobyl, several authors proposed to use a monetary value of severity to make events comparable, and use a rate measure, normalized by the number of reactor operating years, to consider frequency.^(34–36) This is what we have done. Since the IAEA uses INES, it is instructive to compare the two approaches. First, the INES is a discrete scale between 1 (anomaly) and 7 (major accident). Similarly to the Mercalli intensity scale for earthquakes (which has 12 levels from I [not felt] to XII [total destruction]), each level in the INES is intended to roughly correspond to an order of magnitude in severity (in the amount of radiation released).

The INES has been criticized.⁽¹⁰⁾ Common criticisms include that the evaluation of INES values is not objective and may be misused as a public relations (propaganda) tool; moreover, the historical scores are not published, not all events have INES values assigned, no estimate of risk frequency is provided, and so on. Given confusion over the INES scoring of the Fukushima disaster, nuclear experts have stated that the “INES emergency scale is very likely to be revisited.”⁽¹³⁾ The Nuclear Accident Magnitude Scale (NAMS),⁽¹⁰⁾ a logarithmic measure of the radiation release, was proposed as an objective and continuous alternative to INES. This proposition, to go from the INES to the NAMS, is reminiscent of when the geophysics discipline replaced the discrete Mercalli intensity scale by the continuous Richter scale with no upper limit, which is also based on the logarithm of energy radiated by earthquakes. In the earthquake case, the Mercalli scale was invented more than 100 years ago as an attempt to quantify earthquake

sizes in the absence of reliable seismometers. As technology evolved, the cumbersome and subjective Mercalli scale was progressively replaced by the physically-based Richter scale. In contrast, the INES scale looks somewhat backward from a technical and instrumental point of view, but was created in 1990 by the International Atomic Energy Agency as an effort to facilitate consistent communication on the safety significance of nuclear and radiological events, while more quantitative measures are available.

Here, we perform a statistical back-test of the accuracy of INES values, in relation to costs and NAMS. Indeed, INES is not defined in terms of cost; however, if INES fails to capture the information that costs do, then the cost measure is important. In Fig. 1, we plot both the logarithm of cost, and NAMS versus INES. There is an approximate linear relationship between INES and log cost (intercept parameter at $INES = 0$ is 0.64 [0.3] and slope 0.43 [0.08] by linear regression). This is consistent with the concept that each INES increment should correspond to an order of magnitude in severity. However, cost grows approximately exponentially ($10^{0.43} \approx e^1$) rather than in multiples of 10 with each INES level. Further, the upper category (7) clearly contains events too large to be consistent with the linear relationship. For instance, the largest events (Fukushima and Chernobyl) would need to have an INES score of 10.6 to coincide with the fitted line. In addition, the cost of INES level 3 events does not appear to be statistically different from the sizes of INES level 4. Finally, there is considerable uncertainty in the INES scores, as shown by the overlapping costs. There is an approximate linear relationship between INES and NAMS (at $INES = 3$ the intercept is 1.8 [0.9] and slope 1.7 [0.2] by linear regression). One sees from the points, and from the fact that the slope of the line is greater than 1, that large radiation release events have been given an INES level that is too small. Furthermore, some INES level 3 events should be INES level 2. This illustrates the presence of significant inconsistency of INES scores, in terms of radiation release level definitions.

2.2. The Current Fleet of Reactors

One must judge the number of accidents relative to the so-called volume of exposure; in this case, the number of reactors in operation. These data were taken from the industry data set⁽³⁷⁾ and are plotted in

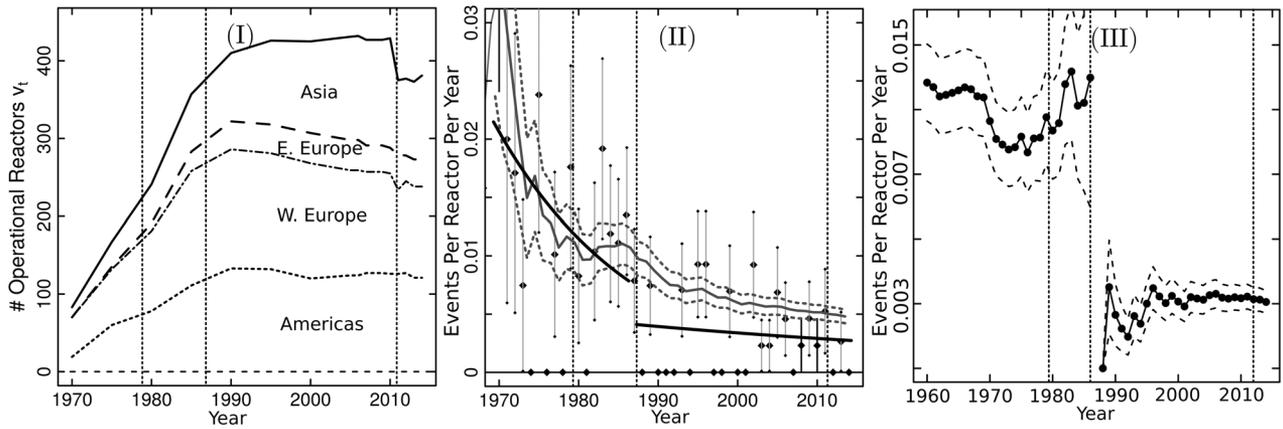


Fig. 2. Figures concerning the rate of events. Panel I: The number of operational reactors over time⁽³⁷⁾ where the bottom layer is the United States, the second layer is Western Europe, the third layer is Eastern Europe, and the top layer is Asia. Panel II: The annual observed frequencies are given by the solid dots, with Poisson standard errors given by the vertical gray lines ($\text{Var}(\lambda_t) = \frac{\lambda_t}{v_t}$). The solid irregular line bounded by dotted lines is the running rate estimate (Equation (1)) from 1970 onwards, with standard errors. The solid black lines are the Poisson GLM regressions (Equation (2)) from 1970 to 1986 and 1987 to 2014, with parameters in Table II. Panel III provides the running rate estimate (Equation (1)) running both backwards and forwards from the Chernobyl event in 1987. In all three panels the vertical dotted lines indicate the occurrence of the TMI, Chernobyl, and Fukushima events.

Panel I of Fig. 2. The number of reactors in operation grew sharply until 1990, after which it stabilized. The stable level has been supported by growth in Asia, compensating for a decline in Western Europe. A steep drop is observed in the Asian volume where, following Fukushima in 2011, all of Japan's reactors were shut down temporarily, until further notice.⁽³⁸⁾ On the topic of reactors, it is important to note that reactors are somewhat informally classified into generations:⁽³⁹⁾ Generation I reactors were early prototypes from the 1940s to 1960s. Generation II reactors were developed and built from the 1960s to the 1990s, of which boiling water reactors (BWR) and pressurized water reactors (PWR) are common. Generation III reactors have been developed since the 1990s. These reactors, such as the advanced BWR and PWR reactors, were improvements upon their Generation II ancestors, replacing safety features that required power with passive ones, and having more efficient fuel consumption. Generation IV reactors concern new technologies, and are still being researched/developed. They have the intention of further improving safety and efficiency, which were deemed as still being inadequate in the Generation III reactors. The vast majority of existing reactors are of Generation II, where most Generation I reactors have been decommissioned, and few Generation III reactors have been constructed. Generation IV reactors are not expected to be deployed commercially until at least 2030 or even 2040.

3. UNCERTAINTY QUANTIFICATION OF RISKS IN NUCLEAR ENERGY SYSTEMS

Prior to moving ahead with data analysis and interpretation, reflection on the degree of uncertainty present, and how it is handled in the analysis, is warranted. Following Aven⁽⁴⁰⁾ two important questions are (i) if (frequentist or Bayesian) probabilities are attainable, and (ii) if the proposed probabilistic model is accurate/valid. Given a lack of relevant data (e.g., when talking about the future), or difficulty with justifying models, the answer to these questions may be no. If the answer to both questions is no, then one can be said to be in a state of *deep uncertainty*. Such considerations are relevant to the study of risk in nuclear energy systems, and are discussed below.

Fortunately for this study, the context is clear, as historical risks, and the current risk level, are being analyzed. Future risk is only being discussed insofar as the current state remains. Thus, the analysis does not need to deal with uncertain futures. Also, by studying risk at a global level, one avoids epistemic uncertainties associated with the specificities of a given plant, type of accident, or technology. Furthermore, relevant data are available and a simple and somewhat justifiable model used. Thus, here a probabilistic approach is valid, where uncertainties include epistemic model uncertainty and aleatoric statistical uncertainty in the parameter estimates, as well as data uncertainty. The epistemic and aleatoric

uncertainties are dealt with by imposing a relatively broad range of parameter estimates (for frequency, severity distribution, and maximum possible severity). Regarding data uncertainty, the data studied here are at a much higher quality level than those of previous studies on nuclear risks. Although the authors are committed to ongoing expansion and refinement of the data, this is a sensible point to provide an analysis. That is, it is unlikely that reasonable modifications of the cost estimates will substantially impact the high-level results provided.

Going beyond this analysis, one can look deeper into nuclear risk. For instance, as regulation requires, PSA is done for each individual plant.^(2,3) This necessitates that the probabilities of initiating events be specified and that the interaction of events be encoded in a fault tree. Further, if one wants to perform level 3 PSA, then the consequences of each event need to be specified.⁽⁴⁾ For this task, at a unique power plant, there is little data and thus the huge number of parameters must be specified based on belief/assumption rather than frequentist estimates. Thus, in addition to the aleatoric sampling uncertainty that is captured by simulating from the model, one should also consider the epistemic uncertainty in the model specification and its parameters.^(17,41) Such an approach has been suggested⁽¹⁶⁾ and a research project considering such methodology in studying the risk of a large loss of coolant is underway.⁽⁴²⁾ However, standard PSA practice and regulations are not yet at this level of uncertainty quantification. Furthermore, needing to encode all possible events in the fault tree implies that the worst case is limited to the one that the modeler can imagine. Finally, the epistemic uncertainty present in the specification of the (typically deterministic) fault tree, which will practically always be incomplete, is not considered at all. These data and epistemological limitations imply that PSA exhibits deep uncertainty. These issues largely inhibit the ability of standard PSA to provide an adequate quantification of overall risk.

Nonetheless, PSA is an important and useful exercise. That is, it is a top-down technique that allows for the generation and prioritization of high-risk events that may not have been observed, and for common causes to be identified. It is thus instrumental in safety improvement, and risk-informed decision making. The statistical approach is a bottom-up technique, whose specificity (e.g., the risk of a specific reactor technology) is restricted by limited historical data, and whose instances from which one can learn are limited to those that have

been observed. It is natural to combine top-down and bottom-up approaches,⁽¹⁸⁾ at least by comparing their results. It is clear that this should be done in nuclear energy systems as well.

Taking a broader perspective, the future risk of nuclear energy should be considered to support decision making, both within the nuclear energy industry, but also within the portfolio of energy source alternatives. The future risk of nuclear energy is deeply uncertain: it depends heavily on developments in reactor and disposal technology, plant build-out and decommissioning scenarios, the emerging risk of cyber-threats and attacks, etc. Furthermore, in making decisions about the holistic plans for future energy systems, the uncertainties of other energy sources also become relevant across multiple criteria. Many risks are reasonably well understood, such as reduced life expectancy, but the evaluation of terrorist threats,⁽⁴³⁾ and the potentially severe environmental impacts of carbon emission, are deeply uncertain.^(44,45) Thus, energy system decisions should be supported by robust multiple criteria decision-making tools with adequate consideration of uncertainty. A probabilistic attempt could be scenario analysis with probability distributions being assigned to the scenarios. Alternatively, nonprobabilistic methods may be warranted, as has been done in large-scale engineering systems with multiple diverse stakeholders.⁽⁴⁶⁾

4. EVENT FREQUENCY

Regarding the frequency of events, we observe $N_t = 0, 1, 2, \dots$ events each year for the v_t nuclear reactors in operation for years $t = 1960, 1961, \dots, 2014$. The *annual observed frequencies* of accidents per operating facility reactor are $\hat{\lambda}_t = \frac{N_t}{v_t}$. The observed frequencies are plotted in Panel II of Fig. 2. The rate of events has decreased, and perhaps stabilized since the end of the 1980s. The running rate estimate,

$$\hat{\lambda}_{t_0, t_1}^{\text{RUN}} = \frac{\sum_{t=t_0}^{t_1} N_t}{\sum_{t=t_0}^{t_1} v_t}, \quad (1)$$

used by Hofert⁽⁹⁾ is plotted for $t_0 = 1970$ and $t_1 = 1970, 1971, \dots, 2014$. In the presence of a decreasing rate of events, such a running estimate overestimates the rate of events for recent times. Furthermore, it is a smoothing method where the estimate

Table II. Parameter Estimate, Standard Error, and p -Value for GLM Estimates of Rate (Equation (2))

Model	β_0	β_1
GLM 1970:1986	$-3.87 (0.3), 10^{-16}$	$-0.06 (0.03), 0.05$
GLM 1987:2014	$-5.53 (0.3), 10^{-16}$	$-0.015 (0.02), 0.45$

Notes: The two rows are two estimates. The first starts at 1970 and ends at 1986. The second starts and 1987 and ends at 2014. The intercept parameter is given for at the starting time.

is taken at the right-most edge of a constantly growing smoothing window, rather than in the center of a window with fixed width. To avoid this bias and to properly evaluate the trend, we consider another approach. We assume that N_t are independently distributed $\text{Pois}(\lambda_t v_t)$. The Poisson model features no interaction between events, which is sensible as separate nuclear events should occur independently, and is compatible with how we have defined our events in Section 2. The changing rate of events is accommodated by a log-linear model for the Poisson rate parameter,

$$\lambda_t^{GLM} = \mathbb{E} \left[\frac{N_t}{v_t} \right] = \exp(\beta_0 + \beta_1(t - t_0)), \quad (2)$$

for given $t_0 < t$ and parameters β_0, β_1 . This is the so-called generalized linear model (GLM) for Poisson counts⁽⁴⁷⁾ and may be estimated by maximum likelihood (using in R: `glm`). The GLM model was estimated from 1970 until 1986 and from 1987 until 2014, with estimated parameters in Table II, and plotted in Panel II of Fig. 2. The first estimate suggests a significantly decreasing rate, which is in agreement with the decreasing running rate estimate. The second (approximately flat) GLM indicates that, from 1987 onwards, the rate has been not significantly different from constant. Clearly, the running rate estimate, starting at 1970, is unable to account for this. To further diagnose this difference, in Panel III of Fig. 2, the running estimate (Equation (1)) is done running both forward and backward from the Chernobyl event of 1987. From here it clear that the rate prior to Chernobyl $\hat{\lambda}_{1960,1986}^{RUN} = 0.013(0.002)$ is larger than the rate after $\hat{\lambda}_{1987,2014}^{RUN} = 0.0032(0.0006)$. Thus it is apparent that there was a significant reduction in the frequency of events following Chernobyl—likely due to a comprehensive industry response.

It is interesting to note that such a change is not apparent following TMI. There are insufficient data to identify a change following Fukushima. That the data are incomplete implies that our rate estimates

Table III. Statistics by Region: Number of Events (N) and Number of Reactor Years (ν) from 1980 through 2014, the Rate of Events per Reactor Year, and the Poisson Standard Error of the Rate

Region	N	ν	$\hat{\lambda}_{1980,2014}^{RUN}$	Std.
Americas	32	4,378	0.0073	0.001
Western Europe	12	4,813	0.0027	0.001
Eastern Europe	5	2,180	0.0023	0.002
Asia	13	2,548	0.0051	0.001

Notes: Russia is included in Eastern Europe.

are underestimates. For instance, even within our data set, there are 40 events occurring after Chernobyl whose cost is unknown. Of these 40 events, 32 have INES values, and 17 of these have INES values of 2 or larger. Based on the known INES and cost values, the median cost of events with INES = 2 is 26 MM USD, i.e., more than half of INES = 2 events exceed the threshold. Thus, based on these statistics, assuming that only nine of the 40 unknown events have cost in excess of 20 MM USD is conservative. With such an assumption, the estimate becomes $\hat{\lambda}_{1987,2014}^{RUN*} = 0.004(0.0006)$. Taking into account the above, and that the rate of events may actually still be decreasing, we consider a conservative range of current estimates $\hat{\lambda}_{2014}$ between 0.0025 and 0.0035. This suggests an average of 1 to 1.5 events per year across the current fleet of reactors. Thus, despite having a data set twice as big as in Hofert,⁽⁹⁾ we find a similar rate estimate. Further, provided that the fleet does not undergo any major changes, we expect the rate to remain relatively stable.

As in Hofert,⁽⁹⁾ a significant difference between the frequency of events across regions is found. In Table III, one sees that the running estimate of the annual rate varies by as much as a factor of 3 across the regions. This is likely to be more due to a difference in reporting rather than a difference in the true rate of events. This provides further evidence that our rate estimates are underestimates.

5. EVENT SEVERITY

For the quantitative study of event severity, costs (measured in MM 2013 USD) are considered to be i.i.d. random variables $X_i, i = 1, 2, \dots, n$ with an unknown distribution function F . Here, we estimate the cost distribution. A common heavy-tailed model for such applications is the Pareto CDF,

$$F_P(x; u_1) = 1 - (x/u_1)^{-\alpha}, x > u_1 > 0, \alpha > 0, \quad (3)$$

which may be restricted to a truncated support as,

$$F(x|u_1 \leq X \leq u_2) = \frac{F(x) - F(u_1)}{F(u_2) - F(u_1)}, 0 < u_1 < u_2, (4)$$

where u_1 and u_2 are lower and upper truncation points that define the smallest and largest observations allowed under the model. Extending further, truncated distributions may be joined together to model different layers of magnitude,

$$F_{2P}(x|u_1 \leq X) = F_P(x|u_1 \leq X \leq u_2)\Pr\{u_1 \leq X \leq u_2\} + F_P(x|u_2 \leq X)\Pr\{u_2 \leq X\}. \quad (5)$$

In Fig. 3, the severity measures are plotted according to their empirical complementary cumulative distribution functions (CCDFs). In Panel I, the sample of costs in excess of 20 MM USD is split into pre- and post-TMI periods, with 42 and 62 events, respectively. The distributions are clearly different. Indeed, the KS test,⁽⁴⁸⁾ with the null hypothesis that the data for both subsets come from the same model, gives a p -value of 0.015. As can be seen from the lower inset of the first panel of Fig. 3, this p -value is much smaller than the p -values obtained for testing other change-times. For instance, there was no apparent change between the pre- and post-Chernobyl periods. The pre-1979 data, having median cost of 283 MM USD, have a higher central tendency than the post-1979 data, having a median cost of 77 MM USD. However, the post-1979 distribution has a heavier tail, whereas the pre-1979 distribution decays exponentially. It is a rather well-known observation that improved safety and control in complex engineering systems tends to suppress small events, but often at the expense of more and/or larger occasional extreme events.^(49–51) This raises the question of if the observed change of regime belongs to this class, as a result of the improved technology and risk management introduced after TMI.

Thus, we focus on estimating the left-truncated Pareto (Equation (3)) for the post-1979 data. The estimate $\hat{\alpha}(u_1)$ fluctuates in the range of 0.5–0.6 for lower threshold $20 < u_1 < 1,000$ MM USD, indicating that the data are consistent with the model. For $u_1 < 20$, the estimate of α is smaller, as is typical for data sets where small events are underreported. In the former study of Sornette,⁽⁵⁾ the estimated value was larger ($\alpha = 0.7$), while Hofert⁽⁹⁾ also found values between 0.6 and 0.8. With our more complete data set, the smaller value α is qualitatively consistent with previous studies, but further emphasizes

the extremely heavy tailed model ($\alpha \leq 1$) where the mean value is mathematically infinite. In practice, this simply means that the largest event in a given catalog accounts for a major fraction ($\sim 1 - \alpha$) of the total dollar cost of the whole.⁽⁵²⁾ That is, the extremes dominate the total cost.

6. RUNAWAY DISASTERS AS “DRAGON-KING” OUTLIERS

In a complex system with safety features/barriers, once an event surpasses a threshold, it can become uncontrollable, and develop into a “runaway disaster”—causing disproportionately more damage than other events. This is the type of phenomenon that is considered here. In Panel I of Fig. 3 one can see that (at least) the two most costly events since TMI (Chernobyl and Fukushima) lay above the estimated Pareto CCDF, and that Chernobyl, TMI, and Fukushima form a cluster of outliers in the sample of NAMS data. The NAMS data cannot be split into pre- and post-TMI samples due to insufficient data. The NAMS distribution is well described by the exponential distribution for values between 2 until 5, as reported by Smythe.⁽¹⁰⁾ That is, for NAMS values above $u_1 = 3.5$, the estimate is $\hat{\alpha}_{NAMS} = 0.72(0.3)$ with sample size $n = 15$, where the three suspected outliers were censored to avoid biasing the estimate.⁽⁵³⁾ Since NAMS is a logarithmic measure of radiation, this corresponds to a Pareto distribution for the radiation released, which is valid over three decades.

We relate this concept of runaway disasters to the concept of DK. The term DK has been introduced to refer to such situations where extreme events appear that do not belong to the same distribution as their smaller siblings.^(49,50) DK is a double metaphor for an event that is both extremely large in size or impact (a “king”) and born of unique origins (a “dragon”) relative to other events from the same system. For instance, a DK can be generated by a transient positive feedback mechanism. The existence of unique dynamics of DK events gives hope that they may be “to some extent” predictable. In this sense, they are fundamentally different from the *a priori* unpredictable “black swans.”^(54,55) Statistically speaking, given that extreme events tend to follow heavy-tailed distributions, DK can be specifically characterized as outlying large extremes within the population of extreme events. That is, DK live beyond power law tails, whereas black

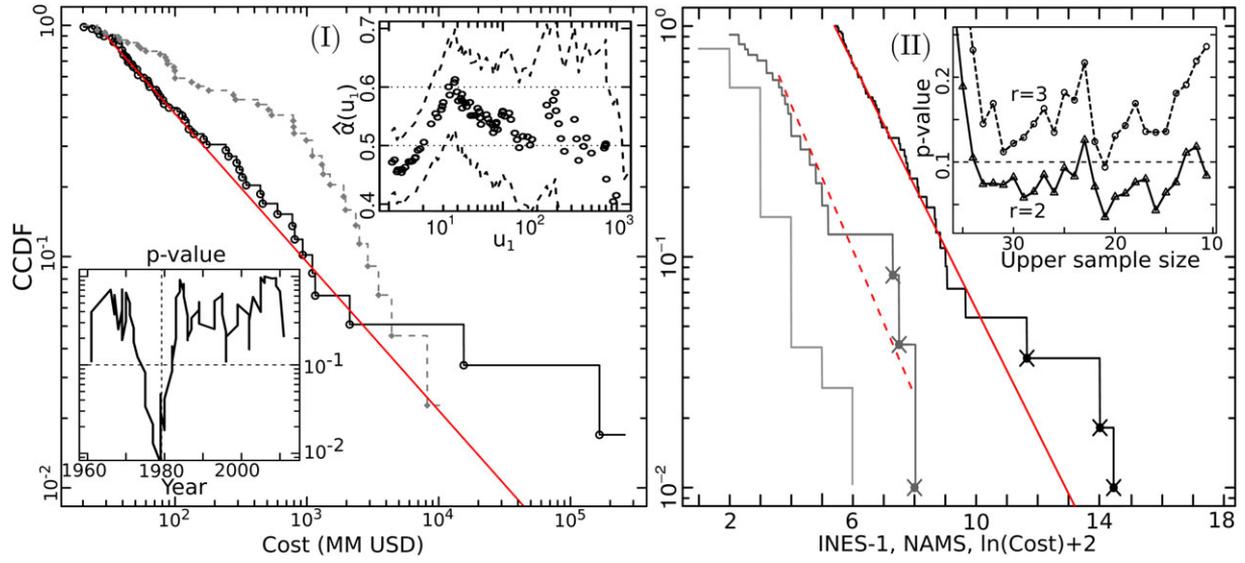


Fig. 3. Figures concerning the severity of events. Panel I: The main frame plots the cost of events for the pre- and post-TMI periods according to their CCDFs, in gray and black, respectively. The lower inset figure shows the p -value of a segmentation test of the cost data, identifying TMI (1979) as the change-point in the cost distribution. The upper inset figure shows the estimated parameter α (with standard deviation) of a Pareto distribution (Equation (3)), for the post-TMI cost data, for a range of lower thresholds (u_1). The fit for $u_1 = 30$ (MM USD) is given by the red solid line in the main frame (color visible in on-line version). Panel II: In the main frame, from left to right, are the CCDF of INES scores above 2 (shifted left by 1), the CCDF of NAMS scores above 2, and the CCDF of the natural logarithm of post-1980 costs (shifted right by 2). For the center and right CCDFs, the dots with x marks indicate suspected outliers/dragon kings. The dashed and solid red lines are exponential fits to the CCDFs. The inset figure provides the p -value for the outlier test for the upper sample above a growing threshold. The upper curve is for $r = 3$ outliers, and the lower curve for $r = 2$ outliers.

swans are often thought of as being generated by (unanticipated) heavy power law tails.

Let us now make the above observations more rigorous by testing the apparent DK points as statistical outliers. There are many tests available to determine if large observations are significantly outlying relative to the exponential (or Pareto) distribution.^(56–59) A suitable approach to assess the NAMS outliers is by estimating a mixture of an Exponential and a Gaussian density,

$$f_{\text{NAMS}}(x|x > 3.5) = \pi\alpha\exp\{-\alpha x\} + (1 - \pi)\phi(x; \mu, \sigma), \quad \alpha, \sigma > 0, \quad (6)$$

where the Gaussian density $\phi(x; \mu, \sigma)$ provides the outlier regime, and $0 \leq \pi \leq 1$ is a weight. The test is done for the 14 points in excess of 3.5, where the exponential tail is valid. The maximum likelihood estimation of this model (Equation (6)) is done using an expectation maximization algorithm.^(60,61) The estimates of this (alternative) model are ($\hat{\pi} = 0.8, \hat{\alpha} = 0.78, \hat{\mu} = 7.8, \hat{\sigma} = 0.21$). We also consider a null model with no DK regime ($\pi = 1$). For this the MLE is $\hat{\alpha} = 0.66$. The alternative model has a significantly superior log-likelihood (the p -value of the

likelihood ratio test⁽⁶²⁾ is 0.04). Thus there is a statistically significant DK regime relative to the exponential, with $(1 - \hat{\pi}) \cdot 14 \approx 3$ outliers expected.

That the amount of cost is related to the amount of radiation released suggests testing for a DK regime in cost. Not every runaway radiation release disaster produces commensurate financial damage (see Three Mile Island in Table I). But, given that the majority of nuclear power installations have surrounding population densities higher than Fukushima,⁽²¹⁾ the DK regime in radiation should amplify cost tail risks. From Panel II of Fig. 3 as many as the three largest points could be outlying. For this we consider the *sum-robust-sum* (SRS) test statistic,

$$T_r^{\text{SRS}} = \frac{\sum_{i=1}^r x_{(i)}}{\sum_{i=r+1}^n x_{(i)}}, \quad m \geq 1, \quad (7)$$

for the ordered sample $x_{(1)} > x_{(2)} > \dots > x_{(n)}$, which compares the sum of the outliers to the sum of the nonoutliers.⁽⁵⁹⁾ This test was performed for $r = 2$ and $r = 3$ outliers for a range of upper samples—i.e., the

sample in excess of a growing lower threshold. For $r = 2$, the p -value fluctuates between 0.05 and 0.1, for samples ranging from the 10 to the 40 largest points. For $r = 3$, the test fluctuates between 0.1 and 0.2. Thus, there is evidence that the two largest events are indeed outliers, both in terms of radiation and cost.

Given the suggestive evidence that the extreme tail of cost is heavier than the rest of the tail, perhaps due to a runaway disaster effect, it is important to include this in the risk modeling. For simplicity, we continue with the Pareto model. The MLE for the top 5 points is $\hat{\alpha}(u_1 = 1, 100) = 0.36(0.15)$. To pursue a pleasant but nonrigorous argument, this appears to be consistent with the run-away effect that propels NAMS values from $M \approx 5$ to $M \approx 8$. That is, transforming back from log scale, this same effect on the Pareto model would transform the parameter α to $\alpha(u_1 = 1, 100) \approx \frac{5}{8} \times 0.61 = 0.375$.

7. MODELING AGGREGATE ANNUAL DAMAGE

We now characterize the annual risk of nuclear events measured by cost, with three quantities: quantiles, return periods, and expected values. These characterizations are relevant for the current state of the operating nuclear fleet—excluding any potential improvements following Fukushima—and do not consider scenarios for the transition to more advanced reactor technologies, such as Generation III and beyond.

7.1. Quantiles and Return Periods

Having estimated the rate of events in Section 4, the severity distribution in Section 5, and identified the DK regime in Section 6, we here combine these models in a compound Poisson process (CPP)⁽⁶³⁾ to model the annual total cost in MM USD,

$$Y_t = \sum_{i=1}^{N_t} X_{i,t} \sim \text{CompPois}(v_t \lambda_t, F), \quad (8)$$

where, for each year $t = 1980, 1981, \dots, 2014$, there are a random number of events N_t , modeled by a Poisson process with annual rate $v_t \lambda_t$, and each event has a random size (in MM USD) $X_{i,t} \stackrel{i.i.d.}{\sim} F$, $F(20) = 0$, $i = 1, \dots, N_t$, where the condition $F(20) = 0$ ensures that we consider only events with costs larger than 20 MM USD.

There are a range of statistically valid parameter estimates that should be considered. From Section 4,

rate estimates ranging between 0.025 and 0.035 were suggested as conservative underestimates. From Section 5, the distribution of cost in excess of 20 MM was found to be well described by a Pareto distribution, with parameter between 0.5 and 0.6. Further, in Section 6 it was found that the largest cost values are significantly larger than what would be expected under the Pareto model. An attempt to account for this was made by including a heavier tail (a DK regime) with $\alpha \approx 0.4$ for the top 10% of the mass (with lower threshold $u_1 = 1, 100$ MM USD).

First we characterize the risk level with *return periods*, defined within the CPP model by considering,

$$\begin{aligned} \Pr [\{\text{\#events with size} \geq x_{(j)} \text{ in } \tau \text{ years} > 0\}] \\ = 1 - \exp[-\lambda v \tau \Pr\{X \geq x_{(j)}\}], \end{aligned} \quad (9)$$

which is the probability of observing at least one event, at least as large as some size (e.g., given by an order statistic $x_{(j)}$), in a given time period τ . One sets the equation (Equation (9)) to a given probability p and solves for the return period $\tau_j(p)$ of the j th largest event. Setting $p = 1 - e^{-1}$, one obtains the standard return period $\tau_j(p) = \frac{1}{\lambda v \Pr\{X \geq x_{(j)}\}}$. In Table IV, median and quartile estimates of $p = 0.5$ return period estimates are given for combinations of the above range of parameter values. For each given set of parameter values, 100,000 samples of the data were simulated, parameters reestimated on these samples, and the return period computed. The median and quartiles were taken over these 100,000 return period estimates. In the lowest risk model ($\lambda = 0.00275$, and $\alpha = 0.6$), the median $p = 0.5$ return periods for TMI and Fukushima are 17 and 154 years, respectively. In the most conservative case ($\lambda = 0.0035$, and $\alpha = 0.4$ in the extreme tail), these values are 12 and 62. It is clear that including the DK effect does not inordinately amplify the risk. For instance, the return periods with high frequency risk ($\lambda = 0.0034$) and low risk severity ($\alpha = 0.6$) are similar to those of the low frequency risk ($\lambda = 0.00275$) and high severity risk ($\alpha = 0.4$ in the extreme tail).

Next we provide quantiles. For F , we take the estimated Pareto cost distribution (Equation (3)) with $u_1 = 20$ and $\hat{\alpha} = 0.55$. We also consider F_{DK} , which is a two-layer model (Equation (5)) where the upper layer is for the DK regime. The first layer, from $u_1 = 20$ to $u_2 = 1,100$, is Pareto with $\hat{\alpha}_1 = 0.55(0.15)$ estimated by MLE. The second layer, from $u_2 = 1,100$ onwards, is also Pareto with heavier tail $\hat{\alpha}_2 = 0.4$. Given v_t , λ_t , and F , we can calculate the “aggregate” distribution G for annual cost Y_t . We do this

Table IV. Median and Quartile $p = 0.5$ Return Periods (Equation (9)) for Fukushima ($x_{(2)}$) and TMI ($x_{(4)}$) for Different Rate Parameters λ and Parameters α for the Pareto Distribution Above Lower Threshold u

λ	$x_{(j)}$	$\alpha = 0.6, u = 20$	$\alpha = 0.55, u = 20$	$\alpha = 0.4, u = 1, 100$
0.00275	$x_{(2)}$	(97,154,259)	(64,98,157)	(37,82,274)
0.00275	$x_{(4)}$	(13,17,24)	(10,13,18)	(10,15,30)
0.0035	$x_{(2)}$	(80,128,214)	(50,80,130)	(31,66,198)
0.0035	$x_{(4)}$	(10,13,19)	(8,10,14)	(8,12,22)

Notes: The last column corresponds to the “dragon-king” regime. The median and quartiles are computed over 100,000 estimates of the parameter values, computed on data simulated using the parameter values provided in the table.

Table V. The Estimated 0.95 and 0.99 Quantiles, as Well as the Probability of the Annual Cost Exceeding the Cost of Fukushima $x_{(2)} = 166,089$ MM USD, Are Given for the Aggregate Distribution G

Model	λ	$q_{0.95}$	$q_{0.99}$	$\Pr\{Y_t \geq x_{(2)}\}$
Pareto	0.002	2,950	54,320	0.0054
	0.0025	4,440	82,440	0.0068
	0.003	6,200	115,780	0.0082
Pareto DK	0.002	2,180	120,730	0.0088
	0.0025	3,720	220,510	0.011
	0.003	5,880	331,610	0.013

Notes: The Pareto model is with $u_1 = 20, \alpha = 0.55$, and the Pareto DK model is with $u_1 = 20, u_2 = 1, 100, \alpha_1 = 0.55, \alpha_2 = 0.4$. The volume (number of active nuclear reactors) is taken to be $v_{2014} = 388$. The quantiles are given in MM 2013 USD.

for the year 2014 with the Panjer algorithm by Monte Carlo.⁽⁶³⁾ Quantiles of the estimated G are in Table V. The 0.99 quantile is highly sensitive to the choice of λ and distribution F : for very low rate $\lambda = 0.002$, and without considering the DK effect, the 0.99 quantile is 54,320 (MM USD), which is five times the cost of TMI. For $\lambda = 0.0025$, we obtain a similar estimate to Hofert,⁽⁹⁾ who obtained 81,000 (MM USD). Considering $\lambda = 0.003$, with the DK effect, this quantile is 331,610 (MM USD), which is double the estimated cost of Fukushima.

7.2. Expected Annual Damage

So far we have considered models without a limiting cost, in which the mean cost is mathematically infinite, since our various estimations of the Pareto exponent α all have values less than 1.⁽⁵²⁾ Of course, the Earth itself is finite, thus there is an upper cutoff, u_2 , to the maximum possible cost. But this upper cutoff could be exceedingly large, and there is—as of yet—no evidence of a maximum being reached thus far (i.e., no accumulation of observations at an upper limit in Fig. 3). Think, for instance, of the real-estate

value of New York City, USA or Zurich, Switzerland, both of which are rather close to a nuclear plant, and would become inhabitable in a worst-case scenario. Here, we would be speaking of up to tens of trillions of dollars of financial losses, not to speak of human ones. Thus, insurance and re-insurance companies introduce a maximum loss for their liabilities, which for them works as if there is a genuine upper cutoff: u_2 . Everything above such a cutoff is then the responsibility of the government(s) and society; for the truly extreme catastrophes, only the state can be the insurer of last resort.

It is useful to put hard numbers behind these considerations by using scenarios. For the CPP (Equation (8)), the mean and variance of the annual cost are:

$$E[Y_t] = \lambda_t v_t E[X], \text{Var}(Y_t) = \lambda_t v_t E[X^2]. \quad (10)$$

Given lower and upper truncations, u_1 and u_2 , the first two moments for the Pareto are,

$$E[X] = \frac{\alpha}{\alpha - 1} \left[\frac{u_1^{1-\alpha} - u_2^{1-\alpha}}{u_1^{-\alpha} - u_2^{-\alpha}} \right],$$

$$E[X^2] = \frac{\alpha}{\alpha - 2} \left[\frac{u_1^{2-\alpha} - u_2^{2-\alpha}}{u_1^{-\alpha} - u_2^{-\alpha}} \right]. \quad (11)$$

Thus the mean grows in proportion to $u_2^{1-\alpha}$ (and the variance faster as $u_2^{2-\alpha}$). In Table VI, we compute these moments of the costs X when the maximum value, u_2 , is equal to the present estimate of the cost of Fukushima, 10 times greater, and 100 times greater. Since the expected annual number of events $\lambda_{2014} v_{2014}$ is approximately 1, these values provide a rough estimate of the mean and standard deviation of annual cost in 2014 (Equation (10)).

If we accept that the Fukushima or Chernobyl events represent roughly the largest possible cost then (see Table VI) the mean annual cost is approximately 1.5 billion USD, with a standard error of 8

Table VI. The First Moment and the Square Root of the Second Moment of X Are Given by the First and Second Value, Respectively

Model	$x_{(2)}$	$10 \times x_{(2)}$	$100 \times x_{(2)}$
Pareto	1,513, 8,253	5,367, 54,590	20,488, 349,736
Pareto DK	1,404, 8,466	3,982, 45,267	11,250, 240,810

Notes: The Pareto model is with $u_1 = 20$, $\alpha = 0.55$, and three values for the maximum value u_2 given by the column labels. The Pareto DK model is with $u_1 = 20$, $u_2 = 1, 100$, $\alpha_1 = 0.55$, $\alpha_2 = 0.4$ and three values for the maximum value u_3 . The maximum values are 1, 10, and 100 times the cost of Fukushima, $x_{(2)} = 166,089$ MM USD. All units of costs are in MM USD.

billion USD. This brackets the construction cost of a large nuclear plant, suggesting that about one full equivalent nuclear power plant value could be lost each year, on average. However, the heavy tailed severity implies that most years, there is little cost, and once in a while an extreme hits, driving the total cost up considerably. If we assume that the largest typical possible cost is about 10 times that of the estimated cost of Fukushima, then the average annual cost is about 5.5 billion USD, with a very large dispersion of 55 billion USD. Indeed, the outlook is even more dire for larger possible upper cutoffs. Such numbers do not appear to be taken into account in standard calculations on the economics of nuclear power.⁽⁶⁴⁾ To be fair, we should also note that the long-term effect on, say, lung cancer risks and other particle pollution induced deaths, are not taken into account in evaluating the cost-benefits of alternative sources of energy such as coal.

8. DISCUSSION AND POLICY CONCLUSIONS

Our study makes important conclusions about the risks of nuclear power. Regarding event frequency, we have found that the rate of incidents and accidents per civil nuclear installation decreased from the 1970s until the present time. Along the way, there was a significant drop in the rate of events after Chernobyl (April 1986). Since then, the rate has been roughly stable, implying a rate between 0.0025 to 0.0035 events per reactor per year in 2015. It is worth noting that the decrease in risk due to the reduced accident frequency per reactor from the 1960s onwards has been somewhat offset by an increasing number of reactors in operation.

Regarding event severity, we found that the distribution of cost underwent a significant regime change shortly after the Three Mile Island major ac-

cident. Moderate cost events were suppressed, but extreme ones became more frequent, to the extent that the costs are now well described by the extremely heavy tailed Pareto distribution with parameter $\alpha \approx 0.55$. We noted in the introduction that the Three Mile Island accident in 1979 led to plant-specific full-scope control room simulators, plant-specific PSA models for finding and eliminating risks, and new sets of emergency operating instructions. The change of regime that we document here may be the concrete embodiment of these changes catalyzed by the TMI accident. We also identify statistically significant runaway disaster (“dragon-king”) regimes in both NAMS and cost, suggesting that extreme events are amplified to values even larger than those explained under the Pareto distribution with $\alpha \approx 0.55$.

In view of the extreme risks, the need for better bonding and liability instruments associated with nuclear accident and incident property damage becomes clear. For instance, under the conservative assumption that the cost from Fukushima is the maximum possible, annual accident costs are on par with the construction costs of a single nuclear plant, with the expected annual cost being 1.5 billion USD with a standard deviation of 8 billion USD. If we do not limit the maximum possible cost, then the expected cost under the estimated Pareto model is mathematically infinite. Nuclear reactors are thus assets that can become liabilities in a matter of hours, and it is usually taxpayers, or society at large, that “pays” for these accidents rather than nuclear operators or even electricity consumers. This split of incentives improperly aligns those most responsible for an accident (the principals) from those suffering the cost of nuclear accidents (the agents). One policy suggestion is that we start holding plant operators liable for accident costs through an environmental or accident bonding system,⁽⁶⁵⁾ which should work together with an appropriate economic model to incentivize the operators.

Third, looking to the future, our analysis suggests that nuclear power has inherent safety risks that will likely recur. With the current model—which does not quantify improvements from the industry response to Fukushima—in terms of costs, there is a 50% chance that (i) a Fukushima event (or larger) occurs in 62 years, and (ii) a TMI event (or larger) occurs in 15 years. Further, smaller but still expensive (≥ 20 MM 2013 USD) incidents will occur with a frequency of about one per year, under the assumption of a roughly constant fleet of nuclear plants. To curb these risks of future events would require sweeping

changes to the industry, as perhaps triggered by Fukushima, which include refinements to reactor operator training, human factors engineering, radiation protection, and many other areas of nuclear power plant operations. To be effective, any changes need to minimize the risk of extreme disasters. Unfortunately, given the shortage of data, it is too early to judge if the risk of events has significantly improved post-Fukushima. We can only raise attention to the fact that similar sweeping regime changes after both Chernobyl (leading to a decrease in frequency) and Three Mile Island (leading to a suppression of moderate events) failed to mitigate the very heavy tailed distribution of costs documented here.

A separate conclusion of our article concerns the nature of data about nuclear incidents and accidents. We found that the INES scale of the IAEA is highly inconsistent, and the scores provided by the IAEA incomplete. For instance, only 50% of the events in our database have INES scores. Further, for the costs to be consistent with the INES scores, the Chernobyl and Fukushima disasters would need to be between an INES level of 10 and 11, rather than the maximum level of 7. The INES scale was compared to the antiquated Mercalli scale for earthquake magnitudes, which was replaced by the continuous physically-based Richter scale. Clearly, an objective continuous scale such as the NAMS would be superior to the INES. However, while using INES, scores should be made available for all accidents. When such a framework is established, and data on incidents and accidents made more rigorous and transparent, accident risks can be better understood, and perhaps even minimized through positive learning.

Finally, our study opens a number of avenues for future research. Our results have been obtained for the current fleet, dominated in large part by Generation II reactors. Future research directions could be to investigate how much of the specific risks for each reactor type or design can be inferred from statistical analysis, with the goal of identifying which of the reactors are the safest. In addition to the role of technology, another natural extension would be to correlate accidents to the type of market or form of regulatory governance, restructured versus monopoly/state run, or limited liability versus no limited liability.

Speaking in comparative terms, our focus on the risks of civil nuclear power plants might give the impression that this technology is riskier than other competing technologies, such as coal or wind energy. However, due to the more diluted nature of the costs, and the quasi-hysteric focus on nuclear

risks following the Fukushima disaster, an insidious villain may be hidden: it has been estimated that fine particle air pollution causes about 7 million premature deaths globally each year, including more than 1 million in China, and about 60,000 in Europe. Considering the value of life to be in the millions (10^6), these deaths alone account for a cost on the order of a trillion USD (10^{12}), not taking into consideration the billions also being required for healthcare. Coal, whose global use has soared by 50% from 2000 to 2010, is the leading source of fine particles, which are embedded in the lungs, causing cancers. Between 2010 and 2012, European coal consumption jumped 5%, or 50 million tons. Thus, performing a rigorous empirically-based comparative analysis of the risks of nuclear versus other forms of energy providers is absolutely essential to avoid falling in the traps of media hypes and availability biases, in the goal of a better steering of our societies. Furthermore, such an analysis should, on one hand, take into consideration the costs of the disposal of nuclear wastes, while on the other hand recognize that humankind is confronted with a “nuclear stewardship curse,” whereby existing nuclear byproducts/waste need to be securely managed over immense time scales.⁽⁶⁶⁾

REFERENCES

1. Lee JC, McCormick, NJ. Risk and Safety Analysis of Nuclear Systems. Newark, NJ: John Wiley & Sons, 2012.
2. International Atomic Energy Agency, Development and Application of Level 1 Probabilistic Safety Assessment for Nuclear Power Plants. IAEA Safety Standards No. SSG-3, Specific Safety Guide, International Atomic Energy Agency, Vienna, 2010.
3. International Atomic Energy Agency, IAEA Safety Standards for Protecting People and the Environment, Development and Application of Level 2 Probabilistic Safety Assessment for Nuclear Power Plants. Specific Safety Guide, No. SSG-4, International Atomic Energy Agency, Vienna, 2010.
4. International Atomic Energy Agency, Procedures for Conducting Probabilistic Safety Assessments of Nuclear Power Plants (Level 3), Off-site Consequences and Estimation of Risks to the Public. Specific Safety Guide, No 50-P-12, International Atomic Energy Agency, Vienna, 1996.
5. Sornette D, Maillart T, Kröger W. Exploring the limits of safety analysis in complex technological systems. International Journal of Disaster Risk Reduction, 2013; 6:59–66.
6. Kröger W, Sornette D. Reflections on limitations of current PSA methodology. ANS PSA 2013 International Topical Meeting on Probabilistic Safety Assessment and Analysis, Columbia, SC, USA, September 22–26, 2013.
7. Lochbaum D. Nuclear plants risk studies: Failing the grade. Union of Concerned Scientists Report, 2000.
8. Butler D. Nuclear safety chief calls for reform. Nature, 2011. Available at: <http://www.nature.com/news/2011/110418/full/472274a.html>, Accessed April 1, 2015.
9. Hofert M, Wüthrich MV. Statistical review of nuclear power accidents. Asia-Pacific Journal of Risk and Insurance, 2013; 7(1):1–18.

10. Smythe D. An objective nuclear accident magnitude scale for quantification of severe and catastrophic events. *Physics Today: Points of View*, 2011. Available at: <http://scitation.aip.org/content/aip/magazine/physicstoday/news/10.1063/PT.4.0509>, Accessed April 1, 2015.
11. Ha-Duong M, Journe V. Calculating nuclear accident probabilities from empirical frequencies. *Environment Systems and Decisions*, 2014; 34(2):249–258.
12. Escobar-Rangel L, Leveque F. How Fukushima Dai-ichi core meltdown changed the probability of nuclear accidents. *Safety Science*, 2014; 64:90–98.
13. Brumfiel G. Nuclear agency faces reform calls. *Nature*, 2011; 472:397–398.
14. Rogers S. Nuclear power plant accidents: Listed and ranked since 1952. *Guardian*, March 18, 2011. Available at: <http://www.theguardian.com/news/datablog/2011/mar/14/nuclear-power-plant-accidents-list-rank>, Accessed April 1, 2015.
15. Mosleh A. Delivering on the promise: PRA, real decisions, and real events. PSAM11 - ESREL 2012 Helsinki, Finland, Plenary Talk, June 29, 2012.
16. Durga Rao K, Kushwaha HS, Verma A, Srividya A. Quantification of epistemic and aleatory uncertainties in level-1 probabilistic safety assessment studies. *Reliability Engineering & System Safety*, 2007; 92(7):947–956.
17. Ferson S, Ginzburg LR. Different methods are needed to propagate ignorance and variability. *Reliability Engineering & System Safety*, 1996; 54(2):133–144.
18. Linkov I, Anklam E, Collier ZA, DiMase D, Renn O. Risk-based standards: Integrating topdown and bottomup approaches. *Environment Systems and Decisions*, 2014; 34(1): 134–137.
19. Idaho National Laboratory. Next Generation Nuclear Plant Probabilistic Risk Assessment White Paper, INL/EXT-11-21270. Idaho National Laboratory, Next Generation Nuclear Plant Project Idaho Falls, ID, 2011.
20. Laaksonen J. Thoughts in the aftermath of accident at the Fukushima Daiichi NPP. PSAM11 - ESREL 2012 Helsinki, Finland, June 26, 2012.
21. Butler D. Nuclear neighbours. *Nature*, 2011. Available at: <http://www.nature.com/news/2011/110421/full/472400a/box/3.html>, Accessed April 1, 2015.
22. Burgherr P, Eckle P, Hirschberg S. Comparative assessment of severe accident risks in the coal, oil and natural gas chains. *Reliability Engineering and System Safety*, 2012; 105:97–103.
23. Hirschberg S, Burgherr P, Spiekerman G, Dones R. Severe accidents in the energy sector: Comparative perspective. *Journal of Hazardous Materials*, 2004; 111(1):57–65.
24. Flyvbjerg B, Holm M, Buhl S. Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association*, 2002; 68(3):279–295.
25. Flyvbjerg B, Holm M, Buhl S. What causes cost overrun in transport infrastructure projects? *Transport Reviews*, 2004; 24(1):3–18.
26. Ansar A, Flyvbjerg B, Budzier A, Lunn D. Should we build more large dams? The actual costs of mega-dam development. *Energy Policy*, 2014; 69:43–56.
27. Sovacool BK, Nugent D, Gilbert A. An international comparative assessment of construction cost overruns for electricity infrastructure. *Energy Research & Social Science*, 2014; 3:152–160.
28. Sovacool BK, Gilbert A, Nugent D. Risk, innovation, electricity infrastructure and construction cost overruns: Testing six hypotheses. *Energy*, 2014; 74:906–917.
29. International Atomic Energy Association. INES User Manual, 2013. Available at: <http://www-pub.iaea.org/MTCD/Publications/PDF/INES2013web.pdf>, Accessed April 1, 2015.
30. Appelbaum, B. As U.S. agencies put more value on a life, businesses fret. *New York Times*, Feb 16, 2011.
31. Sovacool BK. The costs of failure: A preliminary assessment of major energy accidents, 1907 to 2007. *Energy Policy*, 2008; 35(5):1802–1820.
32. Wheatley S, Sornette D, Sovacool B. Nuclear events database, 2016. Available at: https://innovwiki.ethz.ch/index.php/Nuclear_events_database, Accessed April 1, 2015.
33. International Atomic Energy Association. Measures to Strengthen International Cooperation in Nuclear, Radiation and Waste Safety. Specific Safety Guide, No 50-P-12. 1996. Available at: https://www.iaea.org/About/Policy/GC/GC40/GC40InfDocuments/English/gc40inf-5_en.pdf, Accessed April 1, 2015.
34. Hsü K. Nuclear risk evaluation. *Nature*, 1987; 328:22.
35. Sengör AMC. Evaluating nuclear accidents. *Nature*, 1988; 335.6189:391.
36. Sengör AMC. Predicting the threat of nuclear disasters. *Nature*, 2011; 472.7342:169–169.
37. International Atomic Energy Association. Nuclear power reactors in the world 2010. Reference data series no. 2. 2010 Edition, 2010.
38. Schneider M, Froggatt A. World Nuclear Industry Status Report 2013. Mycle Schneider Consulting, 2013.
39. European Commission. The different generations of nuclear technology, 2015. Available at: http://ec.europa.eu/research/energy/euratom/index_en.cfm?pg=fission§ion=generation, Accessed April 1, 2015.
40. Aven T. On how to deal with deep uncertainties in a risk assessment and management context. *Risk Analysis*, 2013; 33(12):2082–2091.
41. Ferson S, Ginzburg L, Akcakaya R. Whereof one cannot speak: When input distributions are unknown. *Risk Analysis*, 1996. Available at: www.ramas.com/whereof.pdf, Accessed January 1, 2016.
42. IRSN. BEMUSE project, 2015. Available at: <http://www.irsn.fr/EN/Research/Research-organisation/Research-programmes/BEMUSE/Pages/BEMUSE-project-4481.aspx>, Accessed April 1, 2015.
43. Hirschberg S, Bauer C, Burgherr P, Cazzoli E, Heck T, Spada M, Treyer K. Health effects of technologies for power generation: Contributions from normal operation, severe accidents and terrorist threat. *Reliability Engineering and System Safety*, 2015; 145:373–387.
44. Kandlikar M, Risbey J, Dessai S. Representing and communicating deep uncertainty in climate-change assessments. *Comptes Rendus Geoscience*, 2005; 337(4):443–455.
45. Hallegatte S, Shah A, Brown C, Lempert R, Gill S. Investment decision making under deep uncertainty: Application to climate change. World Bank Policy Research Working Paper 6193, 2012.
46. Karvetski CW, Lambert JH. Evaluating deep uncertainties in strategic priority-setting with an application to facility energy investments. *Systems Engineering*, 2012; 15(4):483–493.
47. Ohlsson E, Johansson B. Non-Life Insurance Pricing with Generalized Linear Models. Berlin: Springer Science & Business Media, 2010.
48. Stephens MA. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 1974; 69.347:730–737.
49. Sornette D. Dragon-kings, black swans and the prediction of crises. *International Journal of Terraspace Science and Engineering*, 2009; 2(1):1–18.
50. Sornette D, Ouillon G. Dragon-kings: Mechanisms, statistical methods and empirical evidence. *European Physical Journal Special Topics*, 2012; 205:1–26.
51. Kovalenko T, Sornette D. Dynamical diagnosis and solutions for resilient natural and social systems. *Planet Risk, Global Risk Forum Davos*, 2013; 1(1):7–33.

52. Sornette D. *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*. Berlin: Springer Science & Business, 2006.
53. Deemer W, Votaw D. Estimation of parameters of truncated or censored exponential distributions. *Annals of Mathematical Statistics*, 1955; 26(3):498–504.
54. Taleb NN. *The Black Swan: The Impact of the Highly Improbable Events*. New York: Random House, 2010.
55. Aven T. On the meaning of a black swan in a risk context. *Safety Science*, 2013; 57:44–51.
56. Balakrishnan K. *Exponential Distribution: Theory, Methods and Applications*. Amsterdam: CRC Press, 1996.
57. Subramaniam L, Kumar N. Multiple outlier test for upper outliers in an exponential sample. *Journal of Applied Statistics*, 2012; 39.6:1323–1330.
58. Pisarenko V, Sornette D. Robust statistical tests of dragon-kings beyond power law distributions. *European Physical Journal-Special Topics*, 2011; 205(1):95–115.
59. Wheatley S, Sornette D. Multiple outlier detection in samples with exponential & pareto tails: Redeeming the inward approach & detecting dragon kings. Submitted to *Journal of Applied Statistics*, 2015. Available at: <http://arxiv.org/pdf/1507.08689.pdf>, Accessed April 1, 2015.
60. Redner R, Walker H. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 1984; 26.2:195–239.
61. Bilmes J. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, 1998; 4.510:126.
62. Wilks S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 1938; 9.1:60–62.
63. Mikosch T. *Non-Life Insurance Mathematics*, 2nd printing. Berlin: Springer, 2006.
64. World Nuclear Association. *Economics of Nuclear Power*, 2015. Available at: www.world-nuclear.org/info/Economic-Aspects/Economics-of-Nuclear-Power, Accessed April 1, 2015.
65. Sovacool BK, Dworkin MH. *Global Energy Justice: Problems, Principles, and Practices*. Cambridge University Press, 2014.
66. Sornette D. A civil super-apollo project in nuclear research for a safer and prosperous world. *Energy Research & Social Science*, 2015; 8:60–65.