Background information

SATW workshops of the topical platform “Risk”

This PDF document contains background material to the SATW publication “Dealing with Risk – Old and New Challenges”, available as a PDF and in printed form. This background document has not undergone SATW’s usual review process, and SATW therefore assumes no responsibility for the content or the quality of the document. Responsibility for these lies solely with the authors.

SATW Swiss Academy of Engineering Sciences
Gerbergasse 5 | 8001 Zurich | +41 44 226 50 11 | info@satw.ch | www.satw.ch

Table of contents

Introduction

The term **risk** still varies in its scope, and depending on the relevant community may be limited to the consequences of an undesired event or may also include positive aspects and opportunities. There is no universal definition, while the key elements of risk – in particular “probability” and its interplay with other elements – are frequently poorly understood. We are often left perplexed by residual uncertainties which may result from a lack of knowledge or the inherent randomness of events (take earthquake as an example), with the same going for the handling of **residual risk**. The simple **risk concept** with a multiplicative combination of frequency and consequence is often considered inappropriate for events with a high level of damage (despite their admitted rarity), asserting aversion to risk – as is for example common in the nuclear sector.

Furthermore, increasing integration and interdependencies in the field of critical infrastructures have given rise to a complex **“system of systems”** which behaves in a manner heavily dependent on its operational and organisational environment and demonstrates disturbance patterns including cascades which we currently find difficult to understand, let alone anticipate. Traditional quasi-static methods such as fault/event trees are insufficient. **Extreme events** are occurring with increasing regularity in many sectors, and differently than anticipated, falling as they do outside of common distribution functions. Some consider them to be incipient outliers, even when focusing on “heavy tails”; others view them as utterly unpredictable, in other words what are known as “black swans”.

“Resilience” has emerged as a **new paradigm**, expanding the risk concept (previously oriented towards increasing resistance) to include a system’s behaviour after disruptive events until functionality is restored. **Risk management** should for example take a geographically integrated approach, be developed in the direction of **“governance”** for very significant issues, and include key stakeholders in the event of major uncertainty and ambiguity.

These challenges (some old, some new) and touted opportunities were the subject of a workshop in November 2013. The aim of this discussion paper is to present a collection of selected articles to interested parties.

Further information on this workshop and the second workshop in June 2014 can be downloaded from www.satw.ch/risiko.

Definitions and elements of risk, traditional methods of risk analysis

Wolfgang Kröger
ETH Zurich, Risk Center

Risk is a term that has been defined many times and in many ways, most recently by ISO 31000:2009, developed to help SME in their risky endeavour, as the “effect on uncertainty on objectives” while

- an effect is a deviation from the expected – positive and/or negative,
- objectives can have different aspects and can apply at different levels,
- risk is often expressed in terms of a combination of the consequences of potential events and the associated likelihood of occurrence,
- uncertainty is the state, even partial, of deficiency of related information.

Often, as here, the normative risk concept focuses on the negative outcome of undesired events and aims to avoid, reduce and control/manage risks, characterized by persistence (temporal), ubiquity (spatial) and reversibility. Uncertainties play an important role in risk assessment. Three kinds can be distinguished: a) epistemic due to incomplete knowledge, b) aleatoric due to the randomness of the physical process, and c) methodical at the level of plant model due to inadequacy of the approach and boundaries, rare event approximation and “cut-offs” as well as the degree of completeness and validity.

Usually, the elements of risk, i.e. likelihood or better frequency (see Box) and the consequences of an undesired event, are arithmetically multiplied (“Versicherungsformel”) and in case of more than one event summed up afterwards. In order to consider “risk aversion”, where appropriate, the consequences are unequally weighted by an exponent, ranging from 1.2 to 2. To keep risk information separate and transparent in its structure, frequency-consequence diagrams are advisable (Fig. 1).

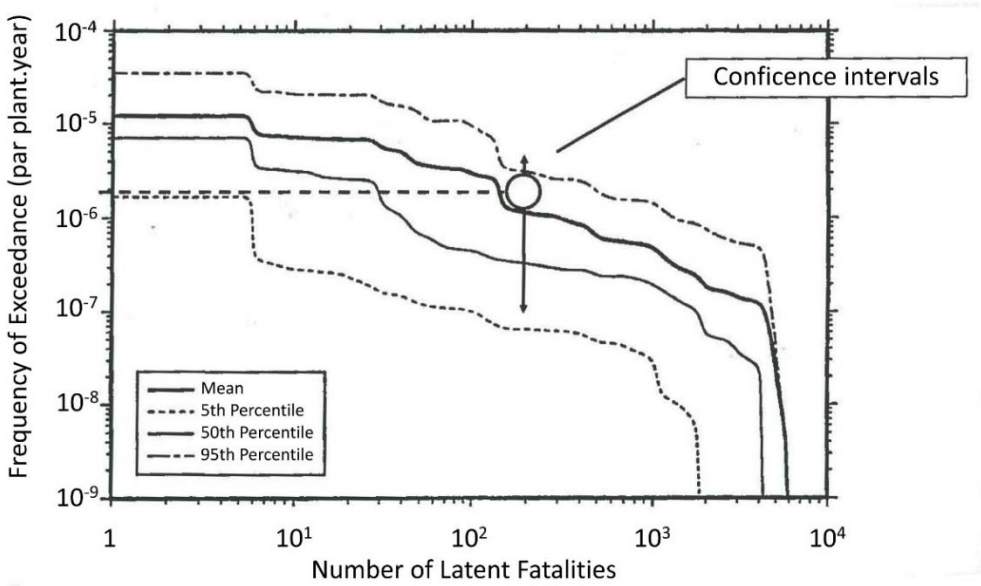

Figure 1: Frequency-Consequence Diagram, exemplarily showing the frequency of events of a certain or higher consequence, here with latent fatalities as damage category, others might be immediate health, environmental, social or economic impact; confidence intervals disclose data uncertainties.[1]

Residual risk is either used as a descriptive term, comprising the remaining risk after implementation of all planned safety measures, arising from deliberately accepted, incorrectly assessed and/or unrecognized risks, or used as normative term meaning the admissible risk following acceptability/tolerability assessment.

The evolving term **systemic risk** depicts the embeddedness of any risk to human health and the environment in a larger context of social, financial and economic consequences and increased interdependencies, both across risks and their various backgrounds (IRGC, 2011). The term **safety** is even more elusive and can be defined in an absolute sense (absence of any danger), in a relative sense (comparatively low or acceptable risk, coping with normative requirements) or as perceived certainty of protection against danger.

Preferably, risk estimates should base on available, directly usable data and on experience with similar events. Often the base for such a **statistical risk estimate** is not given and models for the prediction of rare, not yet occurred failures scenarios and events need to be applied, using empirical data at the components level. Assuming that events including their probability of occurrence can be identified in advance, the Probabilistic Risk/Safety Assessment (PRA/PSA) framework has been developed and applied mainly to nuclear power plants. Resulting **prognostic risk estimates** are used as plant-specific platform for exchanges of safety matters between actors (regulatory body, industry, peers), not thought to become generalized and represent true risks of nuclear power plant operation. Three sequential levels of PSA are distinguished (Fig. 2), traditional methods such as Fault/Event Trees, Human Reliability and Dependent Failure Analysis serve as basis to identify to plant response to assumed initiating events, both internal and external, and unavailability of safety systems or measures needed to handle accidents. Deterministic models are used to investigate physical-chemical effects and containment phenomena.

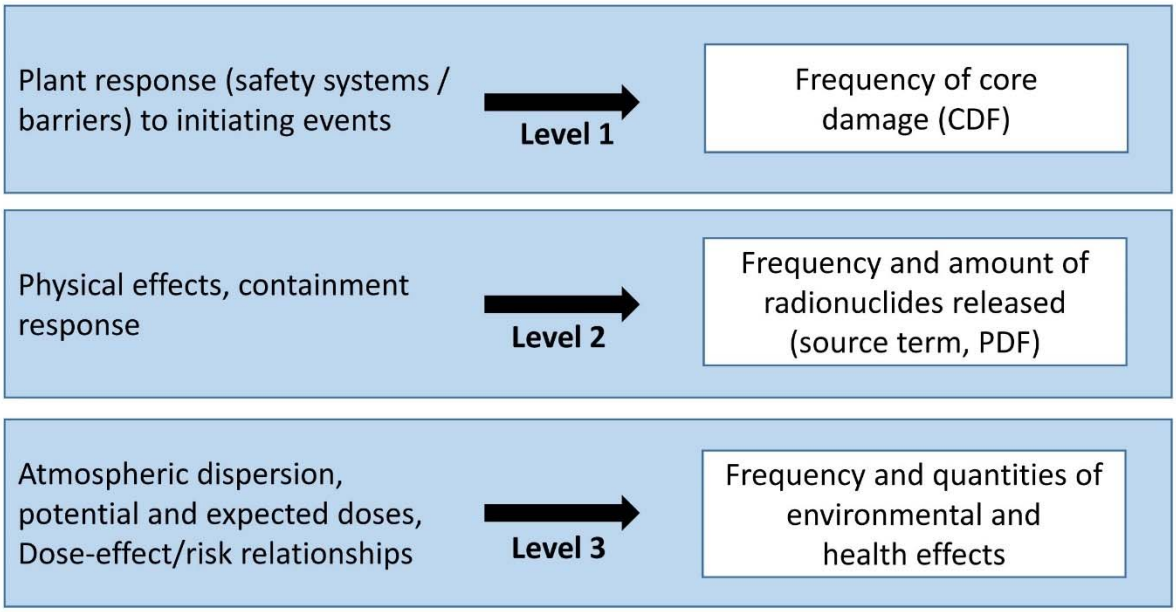

Figure 2: Structure and sequential levels of Probabilistic Safety Assessment for Nuclear Power Plants

In many countries, Switzerland included, PSA level 1 and 2 are required as a regulatory tool (complementary to deterministic analysis) to check whether taken protective measures are sufficiently reliable and balanced and safety goals/target values are met; a methodical framework has been established, high quality standards and peer review processes apply. Only a few PSA level 3 have been carried for academic purposes, often criticized because of running wild uncertainties and too conservative models, e.g. a linear dose-mortality risk relationship without threshold causing unrealistically high numbers for latent fatalities.

Although not thought for, PSA information is in reality often used for decision making in industry or the public sector, contrasted with statistical data: 3 to 5 core damage events (depending how Fukushima Dai-ichi core meltdowns are counted) for about 12'500 to 15'500 operational reactor-years, results in a frequency of slightly more than 10^{-4} per reactor-year while plant specific core damage frequencies obtained by PSA vary between 10^{-4} and 10^{-5} per reactor-year and go down by one order of magnitude for new builds ($2x10^{-6}$ approved for EPR in Finland) or well backfitted plants (e.g. $2x10^{-6}$ per reactor year approved for Leibstadt NPP). These differences are subject of ongoing investigations and discussions about limitations of current statistical and probabilistic analyses. For the latter, standard approaches such as the development of purely linear causal chains, the modelling of the system behaviour by the “sum of the behaviour of its parts” and as a “closed, not interacting with its environment”, also under severe accident conditions, might be too limited.

Explaining Terms by Examples

When tossing a coin 1000 times, we observe head 470 times (k)

- absolute frequency: 470 (per experiment time)
- relative frequency: ratio of k/n = 0.47

If we increase n to infinity, we shall observe relative frequency around an asymptotic value (measure of certainty) called probability with values between 0 and 1, in this case should be 0.5

Assuming a given frequency value of plant damage of 10^{-4} per plant-year, it denotes a chance of 1 to 10’000 to experience next year such a damage in the respective plant or 1 to 100 in similar plants or one event on average in 10’000 years of (fictitious) operation of that single plant.

If we want to create a story about what might happen in the future, we synoptically collocate event or series of actions and/or events, e.g., by analytical means, called scenarios.

If we want to go by train from A to B, we may be interested in the probability that the train performs the required under stated conditions (mission without maintenance), called reliability. The probability that the train is in working order at a certain point in time, e.g., before departure, is called availability (now including maintenance work).

REFERENCES

[1] Kröger, W. (2005) Risk analyses and protection strategies for the operation of nuclear power plants, chapter 2 in Alkan et al., Landolt-Börnstein. Numerical data and functional relationships in science and technology – new series, advanced materials and technologies, Nuclear Energy, pp. 186-235: Springer (2005)

Integrative Risk Management and Resilience – Promising Concepts to Cope with Risks in a System of Systems Context

Hans R Heinimann
ETH Zurich, Future Resilient Systems at the Singapore-ETH Centre (SEC)

We have been facing new challenges that are mainly driven by two factors, the increasing coupling strength and the decreasing heterogeneity, both within and between systems. These changes results in pushing socio-technical systems to a critical state, at which they are moving into a behavioural regimes domain that we did not observe in the past, and for which the prediction of low-probability/high-consequence events with extreme value distribution does not work appropriately. New approaches are required, and here, we are presenting two emerging concepts that are promising to improve the management of disruptions in a system of systems context.

Integrative risk management (IRM) is an emerging concept aiming at concurrently managing a portfolio of natural, technical, economical, and social risks for a specific geographic region. It is seeking for an optimal balance of risk reduction benefits and costs by collaboratively initiating a set of actions public and private contributors. Traditionally, the coping with and management of hazards and disruptions has been discipline-oriented, organized in different communities. However, in specific geographical areas, a set of hazards that needs to be assessed and managed as a whole. Consequently, the different hazard communities have to come together, aiming to perform a holistic assessment of the hazard portfolio and to propose measures with the best cost-effectiveness in reducing the risk in an area under consideration. IRM requires joint efforts in research, outreach and by industrial partners, and has to bring together several disciplines from engineering, financial, and social sciences. There are three pillars of research building the foundation for IRM: a) measurement and aggregation of different types of risks, b) finding an optimal balance of risk management activities such as prevention, preparation, response, recovery, insurance, and hedging, and (3) designing and implementing purposeful mechanisms for collaborative action. Since most of risk-related activities are still going on in disciplinary silos, a mindset change to IRM is an endeavour that will require joint efforts for a decade, or even longer.

In the last decade, **resilience** gained much attraction, and policymakers, practitioners and academics have been using it liberally and enthusiastically[4]. Resilience as a concept has been around for quite a while[2]. Material scientists use it to describe the ability of an elastic material to recover fully after being deformed, psychologists to characterize the ability of a human to recover after suffering a traumatic injury, and ecologists to characterize the ability of an ecosystem to return to its original state or to a new state after facing major disruptions. There are still different interpretations how the underlying system functions contribute to the resilience characteristics that we are observing at a macro level. Resilience in its essence means the “capacity of a system to absorb disturbances and to reorganize so as to retain essentially the same structure, function, and feedback loops” [7].

This definition embraces two fundamental properties that are interacting symbiotically, system resistance and system resilience. The traditional engineering approach has been to design system’s resistance capacity by “hardening” such that it can withstand characteristic exogenous and endogenous actions. Resilience on the other hand means – going back to the Latin word resilere – to spring and rebound. We are basing our concept on five functions: robustness, resistance, recovery, reconfiguration, and remembering (Figure 1). We are measuring context-specific

performance (MOPs), represented as the y-axis. **Robustness** refers to that range of performance – defined by a lower and an upper limit – that guarantees a continual expected flow of service, although disrupted by small perturbations. **Resistance** is the ultimate limit of a system to withstand actions that are straining it during its lifecycle. Structural engineering resistance often equals the elasticity limit, beyond which non-recoverable deformations will occur. If a disruption is straining a system, its performance will decrease down to a minimum (MOP_{min}), which can become zero in a worst-case. The **recovery** phase consists of two functions, maintaining homeostasis, and repair. Maintaining homeostasis refers to the property of a system to maintain critical functions although facing major disruptions. **Repair** means to re-establish full functionality of a system and to bring the system’s performance back to normalcy. We have to emphasize that system recovery is an active concept that is mobilizing additional resources to get the system back to normalcy, which differs from a "laissez-faire" approach. **Reconfiguration** means to adapt and change systemic properties by introducing or deleting interdependencies, or introducing or deleting components. Our experience shows that reconfiguration rarely happens with man-made systems. It also requires to change or enhance the system boundaries to address the key issue "how should we adapt the topology of the system" to make it more robust and resilient. **Remembering** means the establishment of a memory, enabling a system to respond more rapidly and effectively to disruptions that it encountered previously. We borrowed this property from the immune system, which is a model for system adaptiveness, and which was integrated into the business continuity literature as “corporate immune system” (PWC, 2015). Our resilience approach is similar to concepts recently published[1, 3, 5, 6], but is extending them by reconfiguration and remembering functions. There is still no single approach how to characterize resilience with an appropriate metrics.

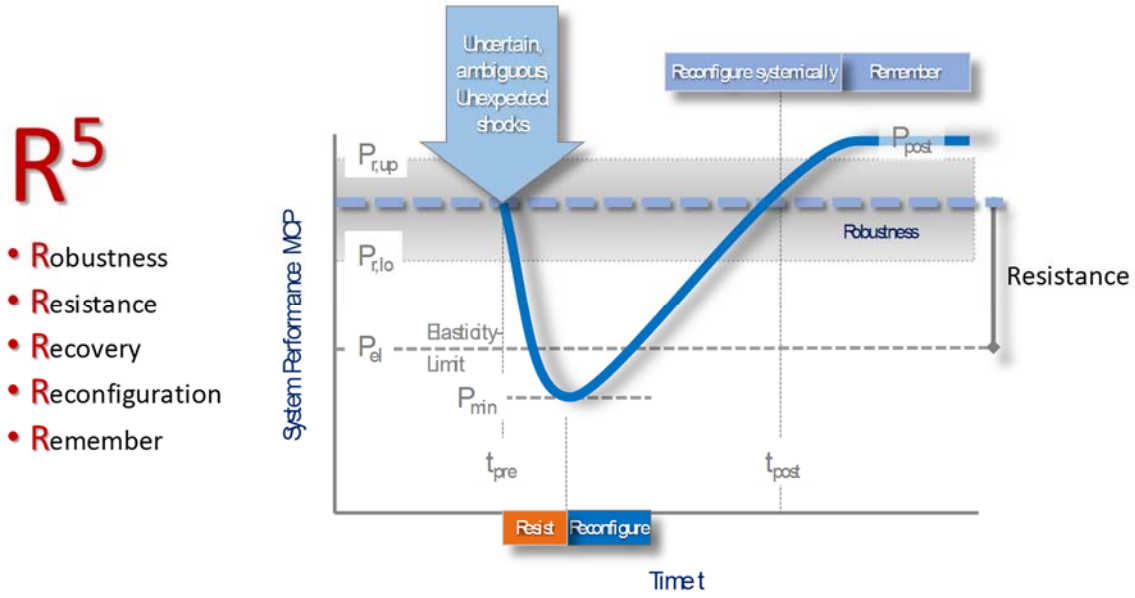

Figure 1: Resilience as a compound of pre-event (robustness, resistance) and post-event (recovery, reconfiguration, remembering) functions of a system. Traditional engineering approaches have mainly been designing pre-event functions, whereas future approaches have to integrate those with post-event functions.

REFERENCES

[1] Ganin, A.A., E. Massaro, A. Gutfraind, N. Steen, J.M. Keisler, A. Kott, R. Mangoubi, and I. Linkov. 2015. Operational resilience: concepts, design and analysis. arXiv preprint arXiv:1508.01230.

[2] Jackson, S. 2015. Overview of Resilience and Theme Issues on the Resilience of Systems. INCOSE Insight. 18 (1): 7-9.

[3] Linkov, I., T. Bridges, F. Creutzig, J. Decker, C. Fox-Lent, W. Kröger, J.H. Lambert, A. Levermann, B. Montreuil, and J. Nathwani. 2014. Changing the resilience paradigm. Nature Climate Change. 4 (6): 407-409.

[4] McAslan, A. 2010. The concept of resilience. Understanding its origins, meaning and utility. Torrence Resilience Institute. Accessed [Feb-02-2016]. [http://www.torrensresilience.org/images/pdfs/resilience%20origins%20and%20utility.pdf]

[5] Needham-Bennett, C. and T. Dalby-Welsh. 2015. Resilience. Analysis-Based Measurement. Continuity. The Magazine of the Business Continuity Institute. 2015 (Q4). Accessed [Dec-10-2015]. [http://www.bcifiles.com/Q4_2015Online.pdf]

[6] Pflanz, M.A. and A.H. Levis. 2015. On Evaluating Resilience in C3 Systems. INCOSE Insight. 18 (1): 29-33.

[7] Walker, B. and D. Salt. 2012. Preparing for Practice: The Essence of Resilience Thinking. In Resilience practice: building capacity to absorb disturbance and maintain function, B.H. WALKER and D. SALT, Editors. Island Press: Washington, D.C.: p. 1-25.

Addressing the nature of extreme events, predictability, and control

Spencer Wheatley, and Didier Sornette
ETH Zurich, Department of Management, Technology and Economics

Extremes and dragon-kings

Extremes dominate the long-term quality and organization of most important natural and societal systems: the largest two nuclear power plant accidents have caused five times more damage than all other (>200) historical accidents together; the largest five epidemics since 1900 caused twenty times the fatalities of all others; etc.[3]. Such statistics are consistent with (extremely) heavy tailed distributions. Furthermore, it is often the case that the largest events are outliers and have special circumstances associated with them.

Despite the importance of extreme events, we often fail to adequately anticipate them: we choose models that are not heavy-tailed enough, and under-appreciate both serial and multivariate dependence of extremes. Such failures in risk assessment include: the failure of economic models to predict the 2007 financial crisis; the under-appreciation of external events, cascades, and nonlinear effects in probabilistic risk assessment; etc. High impact failures (e.g., the 2011 Fukushima disaster) emphasize the importance of the study of extremes.

Here a special type of outlying extreme event is discussed: Dragon-king (DK) is double metaphor for an event that is both extremely large in size or impact (a “king”) and generated from a unique process/origin (a “dragon”) relative to other events from the same system.

It is worth contrasting the DK with the “black swan” theory popularized by Taleb: a metaphor for an event that is surprising, has a major effect, and, after being observed, is rationalized in hindsight. Black swan events are claimed to be unpredictable, and in practice encourage one to “prepare rather than predict”, and limit one’s exposure to extreme fluctuations. However, in a wide range of physical systems, many extreme events are predictable to some degree; provided that one has a sufficiently deep understanding of the structure and dynamics of the focal system, and the ability to monitor it[2].

Statistics of extremes and beyond

When dealing with crises and extremes, power law tails are the normal case. Power laws have a unique property, implying that all events – both large and small – are generated by the same mechanism, and thus there will be no distinct precursors by which the largest events may be predicted. However, in a variety of studies it has been found that the largest events are significantly outlying[3]. Such events are interpreted as DK as they indicate a departure from the generic process underlying the power law. That is, DK are statistical outliers that are highly informative, and should be the focus of much statistical attention. Thanks to a key result from EVT (Extreme Value Theory), a general outlier test is available for detecting DK and assessing their significance[3].

Mechanisms for dragon-kings

DKs may be associated with the regime changes, bifurcations, and tipping points of complex out-of-equilibrium systems[1]. A catastrophe (fold bifurcation) of the global ecology – where incremental loading has little impact, but loading beyond a threshold results in a dramatic change that is difficult to reverse – could be considered as an example. Secondly, positive feedback, e.g. where in

a stampede the number of cattle running increases the level of panic which causes more cattle to run, can cause DK in crowds and stock markets. It could also be the case that DKs are created as a result of system control or intervention. That is, trying to suppress the release of stress or death in dynamic complex systems may lead to an accumulation of stress or a maturation towards instability. For instance, bush/forest fires may be unnaturally suppressed, allowing for a build-up of dead wood, and resulting in a huge uncontrollable fire. An analogue to this could be quantitative easing and low interest rate policies, leading to bubbles and perhaps a massive systemic instability in the making.

Modeling DK requires dynamic models that are complex and/or non-linear, and which need to be fitted to data provided by the continual monitoring of the focal system. For instance, in non-linear systems with phase transitions at a critical point, it is well known that a window of predictability occurs in the neighborhood of the critical point due to precursory signs. Regarding prediction, given the relevant model estimates, one may compute quantities such as the probability of an extreme in a future time interval, related risk measures, the most probable occurrence time of an event, etc. An optimal decision will then balance the cost of false negatives/false positives and misses/false alarms according to a specified loss function. It has been proposed that the more homogenous and connected the system, the more predictable its behavior will be, as presented in Fig. 1.

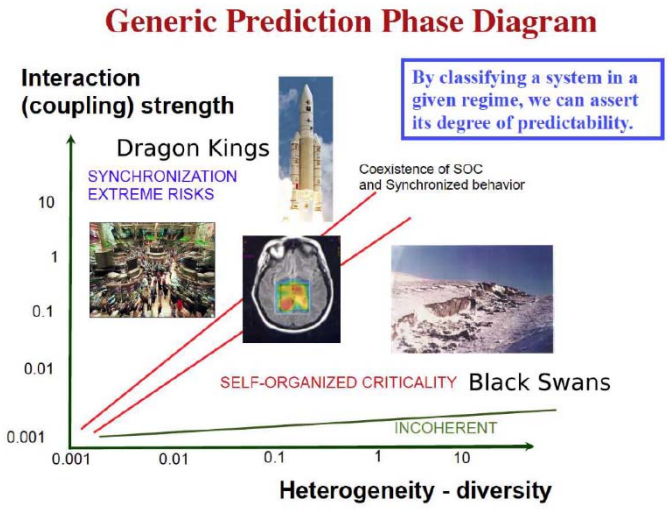

Fig. 1. Schematic for the predictability of a system, characterized by the strength of interaction between, and diversity of its component parts. SOC (self-organized critical) systems exhibit the spontaneous emergence of complexity and extremes due to simple local interactions, which are unpredictable and could be called Black Swans. Avalanches are an example. In systems with stronger interaction more coherent extremal dynamics may be identified, and Dragon Kings may be present. Financial markets provide an example.

Time-at-risks

In a dynamic setting, the dataset will grow over time, and the model estimate, and its estimated probabilities, will evolve. Tests for DK will likely be weak most of the time (e.g., when the system is around equilibrium), but as one approaches a DK, and precursors become visible, the true positive

rate should increase. An important point to mention and ponder is the risk and potential harm resulting from well-intended control.

To mention some critical statistical issues: Second only to the selection of the proper model is selection of the relevant variables. Next, in any such non-trivial models there is bound to be substantial uncertainty that must be accounted for. Thirdly, in the absence of data, simulation is a powerful tool effectively allowing one to conduct rigorous quantitative thought experiments, and providing a valuable complement to the analysis.

The impact of extremes, and DK in particular, urges that extremes be studied and monitored. Future designs should be robust and resilient, acknowledging the potential for such extremes, within a dynamical risk management approach that we call “time-at-risks” (in contrast with the standard static procedures such as value-at-risk, expected shortfalls and the like).

REFERENCES

1 Sornette, D. and Ouillon, G. Dragon-kings: mechanisms, statistical methods and empirical evidence. The European Physical Journal Special Topics 205.1 (2012): 1-26.

2 Sornette, D. Dragon-Kings, Black Swans and the Prediction of Crises, International Journal of Terraspace Science and Engineering 1(3), 1-17 (2009) (http://arXiv.org/abs/0907.4290) and (http://ssrn.com/abstract=1470006)

3 Wheatley, S. and Sornette, D. Multiple Outlier Detection in Samples with Exponential & Pareto Tails: Redeeming the Inward Approach & Detecting Dragon Kings. arXiv preprint arXiv:1507.08689 (2015).

Complex socio-technical engineered systems: Challenges to modelling

Giovanni Sansavini
ETH Zürich, Department of Mechanical and Process Engineering

Decades of massive system integration, strategy changes, and scarce investments have caused large-scale/wide-area technical networks, such as engineered critical infrastructures (CI), to become increasingly interdependent and to operate in the proximity of system limits. Highly integrated energy-carrier networks (e.g. coupled electric power and gas), energy supply with high penetrations of renewable energy sources, cyber-physical systems that rely on pervasive use of communication, and other physically networked systems may serve as examples. Humans are essential components of these complex socio-technical engineered systems because they interact with them both individually, e.g. as operators and managers, and collectively, e.g. as users. As known from the past, human-organizational social systems play a major role in severe accidents occurred in infrastructure sectors, highlighting the importance of properly assessing the performance of the social system together with the technical systems. Resulting “systems of systems” (SoS) have shown the emergence of unprecedented complex behaviours with negative impacts, e.g. cascading failures with widespread consequences. Understanding and characterizing such complex behaviours is vital for thorough reliability and vulnerability analysis.

Research on large-scale networks is characterized by a dilemma: a) it is not feasible to build lab-scale experimental setups to reproduce systemic failures involving the entire system; b) empirical observations of systemic failures are available but sparse because these systems are designed and operated to minimize such occurrences. Therefore, we have to resort to models and simulations that can adequately capture the system behavior and underlying physics to reproduce the consequences of the disruptive event, taking advantage of the few empirical observations. These approaches must be balanced; on the one hand, the creation of artefacts in the models may postulate systemic hazards, which cannot occur in real-world infrastructures, while on the other hand, oversimplifications may fail in capturing the actual system behavior and misrepresent potential severe consequences at the system level.

The modelling and simulation of CI and SoS, in particular, is challenged by the key characteristics of these systems: a) coexistence of multiple time scales, from infrastructure evolution to real-time contingencies; b) multiple, mutually dependent levels and lack of fixed boundaries as CI are made of multiple layers (management, information & control, energy, physical infrastructure); c) broad spectrum of hazards and threats; d) different types of physical flows, i.e. mass, information, power, vehicles; e) organizational and human factors. Furthermore, the resolution required to fulfil the objectives of the analysis guides the level of abstraction and aggregation in the models.

Failure behaviours can emerge from interactions among the topology, physics, and operational procedures that underpin these intricate systems, which makes standard risk-assessment tools insufficient in evaluating the levels of vulnerability, reliability, and risks. Complex Network Theory (CNT) has become pervasive in the study of complex socio-technical systems and offered a number of surprising discoveries. CNT has been applied for the development of infrastructure system models and interdependency-related assessments by representing relationships established through connections among system components. CNT is based on mapping physical configuration of the components of CI and their physical or logical interconnections, which defines the system topology or structure.

The analysis of the topological properties of the network is able to reveal **useful information** about the structural properties, topological vulnerability, and the level of system functionality resulting from the connectivity pattern of its components. The underlying assumption is that connectivity properties can approximate and capture system functionality. For example, the size of the largest connected component of a graph is used as a proxy for the level of system performance. In this respect, catastrophic transitions in system operations are connected to the sudden disappearance of a giant connected component when a fraction of links is removed, i.e. the so-called percolation transition. However, CNT lacks the ability to capture uncertain and dynamic characteristics of infrastructure systems and system properties when time-dependent processes, acting on the network, occur. For instance, the underlying physics of voltage collapse or frequency instability in power grids are overlooked by CNT-models.

At this stage, CNT is extended to account for the **propagation of perturbations** through the system connections and for the knock-on effect that emerges from system connectivity and results in cascading failures. Such cascades are triggered by an initiating local disturbance, e.g. a failure of a component due to local overload, and can affect the whole network via overload propagation and, possibly, disrupt the service it provides. Since network topology has a strong influence on the spreading mechanism, the analysis of the cascade failure evolution must deal with the mutual interplay between the system dynamics and structural complexity. Cascading failure simulations offer insights into the operations of complex socio-technical systems. Operators are aware that large cascades may outbreak, if systems operate in stressful conditions and close to safety margins. Furthermore, there is a “right balance” between need for connectivity and vulnerability to cascade spreading. Poorly connected networks can work under larger stresses before cascades occur; yet, when critical loading condition are achieved, the failure cascade propagates abruptly and, therefore, it is difficult to mitigate. Conversely, if the connectivity increases, the network becomes more vulnerable to cascade spreading due to its numerous links. Nevertheless, such a network can resist the outbreak of cascades for increasing loading conditions. System managers and operators know that if the system routinely approaches critical loading conditions, its connectivity should be larger for ensuring graceful cascade propagation and room for mitigating actions. However, if they want to ensure operating conditions far away from threat of cascades, they should trade some connectivity for decreasing the vulnerability to the spreading of cascading failures.

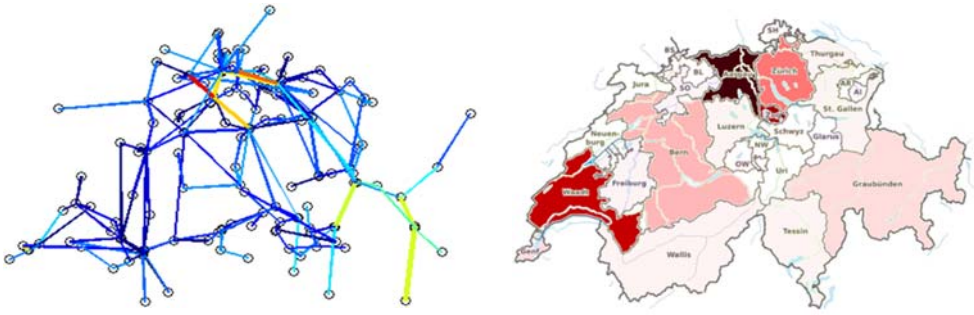

Figure 1. Left pane: Graph theory is the fundamental machinery underlying Complex Network Theory (CNT). This graph maps the Swiss electric transmission system into a set of vertices (nodes) and the set of edges (links) connecting them. Each node represents a power substation and links represent overhead lines. The colour code visualizes the systemic impact of the line failure; warmer colours map larger systemic disruptions to power supply. Right pane: Economic losses caused by cascading failures in the Swiss electric transmission system. Dark colours indicate large losses.

Risk Governance: Concept and Application to Technological Risk

Ortwin Renn, University of Stuttgart, Center of Risk and Innovation Research

The term **risk governance** refers to the various ways in which multiple actors, individuals and institutions, public and private, deal with risks surrounded by uncertainty, complexity and/or ambiguity[1]. It includes formal institutions and regimes and informal arrangements. It refers to the totality of actors, rules, conventions, processes and mechanisms concerned with how relevant risk information is collected, analysed and communicated, and how regulatory decisions are taken. One of the concepts for risk governance has been develop by the International Risk Governance Council in Geneva[5]. This framework provides guidance for constructing comprehensive assessment and management strategies to cope with risk. The framework integrates scientific, economic, social and cultural aspects and includes the engagement of stakeholders.

Klinke & Renn[6] have proposed some alterations to the original concept, as it appeared too rigid and standardized for being applied to complex technological risks, and extended the model to additional adaptive and integrative capacity. The modified framework consists of the following interrelated activities: pre-estimation, interdisciplinary estimation, characterization, evaluation and management of risk. This requires the ability and capacity of risk governance institutions to use resources effectively (see Figure 1).

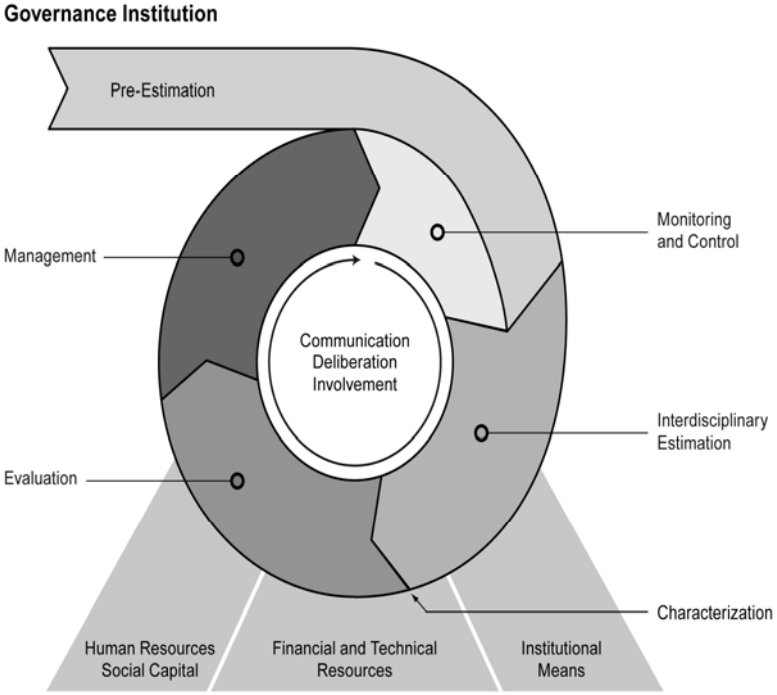

Fig. 1: Adaptive and integrative risk governance model

Within the domain of technological risks, the **pre-estimation** phase includes the choice of an appropriate frame and the establishment of institutions and procedures to deal with emerging threats or events. Aven & Renn[2] claim that many decisions to site hazardous facilities and install adequate safety devices depend on the underlying frames of the actors involved.

[11] Ambiguity is referred to the plurality of legitimate viewpoints for evaluating decision outcomes and justifying judgements about their tolerability and acceptability, so to the existence of multiple values and perspectives.

The interdisciplinary risk estimation comprises two activities: a) *risk assessment*: producing the best estimate of the physical harm that a risk source may induce; and b) *concern assessment*: identifying and analysing the issues that individuals or society as a whole link to a certain risk. When applied to technological risks, this phase includes four major steps[7, 8]: first, risk analysis need to develop scenarios that lead to plausible sequences of accidents or other pathways of harm (pollution, waste production). Second, these scenarios need to be augmented with assumptions about human behavior that one can expect in such situations including crisis management, domino effects, perception driven responses and human errors. It is important that these behavioural components are integrated into the technical analysis because the interaction of both the technical and the human sphere creates the risk to human health and the environment. Thirdly, each scenario needs to be assessed according to its probability of occurrence within the uncertainty ranges in which these estimates are embedded. Fourthly, these scenarios need to be tested for stakeholder and public concerns with respect to their consequences and its implications. There may be equity violations involved or special symbolic meanings affected. These four steps of generating knowledge and insights provide the data and information base for the next step.

A heavily disputed task in the risk governance process concerns the procedure of how to evaluate the societal acceptability or tolerability of a risk. Risk evaluation involves the deliberative effort to characterize risks in terms of acceptability and tolerability. With respect to technological risks, the judgment of acceptability or tolerability is usually related to occupational safety, routine emissions of waste into air soil or water, and accidents with sudden emission of energy and/or material. For all three aspects of technical risks there are normally regulatory standards that need to be adhered to. For sudden events such as accidents often deterministic (safety provisions) and probabilistic (safety goals) standards are in effect; for controlling emissions maximum tolerance levels for certain time intervals (daily, yearly) are specified[1].

Risk management for technological systems requires technological, organizational, and behavioural measures for reducing risks that are not regarded as acceptable in the first place[4]. Technological measures relate to the inclusion of active and passive safety constructions, inclusion of filters and purifiers, and waste handling technology. Organizational measures include emergency and contingency plans, guidelines for daily operations and safety checks, monitoring requirements and provisions for assuring accountability and competence. Behavioural measures extend to all educational and training efforts to improve personal performance, increase sensibility for safety issues and strengthen the feeling of responsibility and accountability among the staff (safety culture). The historic record about technological accidents and failures has shown that the lack of alignment between these measures was often the main reason for disastrous events[3].

Effective communication among all relevant interests is one of the key challenges in risk governance. It is not a distinct stage (in contrast to how it is often treated in the literature), but central to the entire process and at the core of any successful risk governance activity.

In essence, the risk governance concept argues for a broader, paradigmatic turn from government to governance. In the context of risk, the idea of governance is used in both a descriptive and normative sense: as a description of how decisions are made and as a normative model for improving structures and processes of risk policy-making. Risk governance draws the attention to the fact that many risks, particular pertaining to large technological systems, are not simple; they cannot all be calculated as a pure function of probability and effect/consequence. Many risks

embed complex trade-offs of costs and benefits. Risk governance underscores the need to ensure that societal choices and decisions adequately address these complicating features.

REFERENCES

1 Aven, T. & Renn, O. (2010) Risk Management and Governance. Springer, Heidelberg and New York.

2 Aven, T. & Renn, O. (2012) On the Risk Management and Risk Governance for Petroleum Preparations in the Barents Sea Area. Risk Analysis, 32 (9), 1561-1575.

3 Cohen, A. V. (1996) Quantitative Risk Assessment and Decisions about Risk: An Essential Input into the Decision Process, in C. Hood and D. K. C. Jones (eds) Accident and Design: Contemporary Debates in Risk Management. UCL Press, London, pp.87–98.

4 Hood, C., Rothstein, H., and Baldwin, R. (2002) The Government of Risk: Understanding Risk Regulation Regimes. Oxford University Press, Oxford.

5 IRGC (2007) An Introduction to the IRGC Risk Governance Framework. Policy Brief, IRGC, Geneva.

6 Klinke, A. & Renn, O. (2012) Adaptive and Integrative Governance on Risk and Uncertainty. Journal of Risk Research, 15 (3), pp. 273-292.

7 Kröger W. (2008) Critical Infrastructures at Risk, A Need for a New Conceptual Approach and Extended Analytical Tools. Reliability Engineering & System Safety, 93(12), pp.1781-1787.

8 Renn, O. (2014) A Generic Model for Risk Governance: Concept and Application to Technological Installations’ in: P.H. Lindoe, M. Baram, and O. Renn (eds) Risk Governance of Offshore Oil and Gas Operations. Cambridge University Press: New York, pp. 9-33.

9 Renn, O. & Klinke, A. (2016) Complexity, Uncertainty and Ambiguity in Inclusive Risk Governance, in: T. J. Andersen (ed) The Routledge Companion to Strategic Risk Management. Routledge: New York and London, pp. 13-30.

Outlook

In the future (and perhaps to an increasing extent) we expect to continue to be confronted with developments and issues presenting truly serious risks for society, but longstanding methods and approaches examining potential risks will no longer be sufficient to deal with or respond to them. These developments include the integration of socio-technical systems and their control via digital information and communication systems through to the “internet of everything” and autonomous mobility on the one hand, and “pioneering, smart” energy supply systems or traffic and construction concepts on the other.

In light of the generally associated complexity of system design and problem definition, our ability to understand, simulate and evaluate in advance is reaching its limit. New methods and approaches are required and to be applied when available.

There are some promising ideas in this area, but also lingering doubt regarding whether we will succeed. The authors’ articles demonstrate approaches which could potentially be developed into viable methods.

The most promising avenue for making our systems and processes more robust and resilient is often said to be simplifying them, for example via decoupling and decentralisation in fields such as electricity supply and transfer control functions from central units to the end-users. The extent to which this is in fact possible, feasible or desirable remains to be seen.

New, primarily cyber-induced risks are correspondingly worthy of attention, as our ability to imagine the potential dangers arising from malicious system manipulations and to manage them is still limited.

Establishing “good” handling (managing) of risk requires ongoing effort. For far-reaching risks with social relevance featuring high levels of uncertainty and ambiguity, the “risk governance concept” is sufficiently developed and definitely worth a try.

The answer to the “old” question “How safe is safe enough?” – and thus in particular the level to which damage is tolerable – must ultimately be decided by society, which must also understand that there is no such thing as “zero risk”. The most that scholarship can do is provide the basis for the assessment.