

# CONSISTENCY AND STABILITY OF RISK INDICATORS: THE CASE OF ROAD INFRASTRUCTURES

GIULIANO ROSSI<sup>1</sup>, MARA LOMBARDI<sup>2</sup> & PAOLA DI MASCIO<sup>1</sup>

<sup>1</sup>Department of Civil, Constructional and Environmental Engineering, Sapienza University of Rome, Italy.

<sup>2</sup>Department of Chemical Materials Environmental Engineering, Sapienza University of Rome, Italy.

## ABSTRACT

Over the last decade, the World Road Association – PIARC and several European research projects, among which Ecoroads, have encouraged a promising reflection on risk analysis methods, acceptance criteria and safety practices applied to the road system. The goal of this research activity is the definition of best practice for safety analysis and management to be applied to network TERN (Trans European Road Network).

Quantitative Risk Analysis (QRA) provides much information on safety management. Nevertheless, the potential fragility of the method, stochastic uncertainties (both parameters and models), and ethical aspect of criteria must be adequately analyzed. This paper focuses on all these aspects, assessing the reliability of QRA due to modeling errors and statistical errors, and assessing the statistical consistency of Risk Indicators of QRA.

*Keywords: acceptability criteria, Quantitative Risk Analysis (QRA), risk indicators, societal risk*

## 1 INTRODUCTION

Recently, Quantitative Risk Analysis has been increasingly applied in transport engineering. The Trans European Road Network (TERN) standards about road safety are reported in Directive 2004/54/EC [1] applied to tunnels [2] and in Directive 2008/96/EC [3] applied to open road [4]. In the former safety is based on quantitative risk analysis, in the latter on expertise. According to the PIARC *Technical Committee* [5, 6], the aim of technologists and scientists (*i.e. Ecoroads project*) is the harmonization of the safety project of roads and tunnels.

If/when QRA is applied to a transport system, compliance of safety is mainly or exclusively verified on the Social Risk Indicator (F-N design). In the safety tunnel design, the QRA is performed only for the accident scenarios ‘*considered critical in the specific confined space of the tunnels, namely fires, collisions with fire, spills of inflammable substances, releases of toxic and harmful substances*’ [7]. Safety is achieved if all the prescriptions are applied; if any are missing, the minimum level of safety is achieved by increasing the project with the performance approach. According to RAMS approach (Reliability, Availability, Maintainability, and Safety [8]), *safety* is defined as the condition of a system characterized by an acceptable risk of accidents.

Italian Legislative Decree n.264 identifies “*safety targets and risk acceptance criteria for road tunnels [...] represented by the threshold lines shown in the following figure*” [9].

The *Complementary Cumulative Function (CCF)* of the probability density function of fatalities is represented in the plane *F-N*; this function is compared to the compliance criteria (see Fig. 1).

To harmonize these technical methods, it is necessary to choose the best approach to measure system safety specifying the method of Quantitative Risk Analysis (QRA), Risk Indicators and Acceptability Criteria.



Figure 1: F-N verification criteria [10].

Road safety depends on the integrated model between structures and systems. This model may positively influence the behaviour of road users, mitigate the consequences of accidents and facilitate public safety services and self-help.

The QRA consists of Fault Tree Analysis and Event Tree Analysis, organized in the Bow-Tie Model. The Event Tree Analysis calculates the probability density function of random variable fatalities. QRA is a complex inferential analysis. Faber and Stewart showed its reliability limits secondary to the intrinsic variation of phenomena, modelling uncertainty and statistical uncertainty [11]. Under given conditions of epistemic and scientific correctness, we also believe that QRA is the most effective issue for assessing system safety and improving the expected performance.

## 2 QUANTITATIVE RISK ANALYSIS (QRA)

The QRA consists of hazard and fault model, mitigation system reliability and availability analysis, consequences analysis. According to Bow-Tie Model (see Fig. 2), the QRA explores all possible incidents and scenarios.

The hazard analysis has to identify the evolutionary failure scenarios (with Hazop, FMEA, FMECA). The Fault Tree Analysis (FTA) uses combinatorial algebra to explain the evolution of fault (or human error) to the accident (*Top Event, TE*).

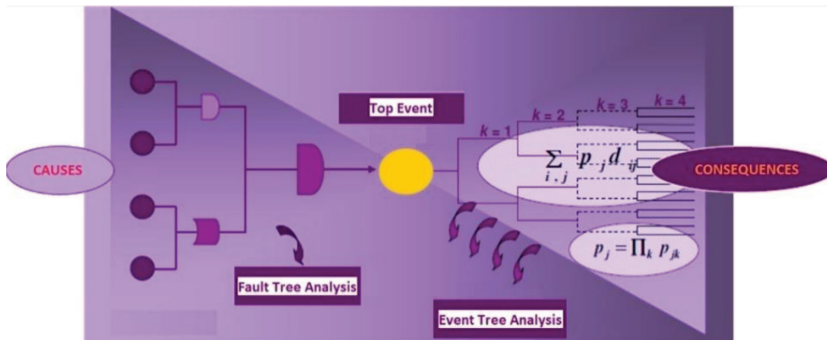


Figure 2: Bow-Tie model of QRA

The set of *Top Events* is a partition of the unsafety of the system, which is complementary to the status safety.

When the accident occurred, the damage to people and exposed goods depends on available energies, mitigation systems, and ambient conditions. The accident is an uncontrolled transformation of potential energy into mechanical and/or thermal energy as defined by the thermodynamic analogy. It is clear that damage  $N$  is a random variable (*r.v.*) dependent on many factors. By focussing on fire in road tunnel, the fire design (firepower, fire duration, combustion products – function of fuel), the number of exposed people, their self-help ability, the emergency lighting (depending on automatic fire detection system, subsystem activation system, manual alarm procedures, and facilities) and the availability of ventilation are all damage determinants.

The RAMS approach for complex systems is used to calculate the reliability and availability of the subsystems (see Fig. 3) according to IEC 61508 [12].

The final accident scenario is the specific outcome between all possible ones. The set of scenarios is a partition of all possible consequences of the accident.

The probability of each end-of-branch scenario is calculated by multiplying the single node probabilities. The final scenario can be simulated (using thermo-fluid-dynamic simulations and exodus) to get the value of the damage. The *ETA*, therefore, calculates the pairs  $\{p_i, d_i\}$  of each scenario.

The fully developed QRA allows the calculation of the following risk indicators:

- *Total Risk (TR) or Expected Value of Damage (EVD)*

$$TR = \sum_{i=1}^n p_i \cdot d_i \tag{1}$$

*TR* is the weighted average of all possible damage in case of accident. Weights consist of the probabilities of scenarios.

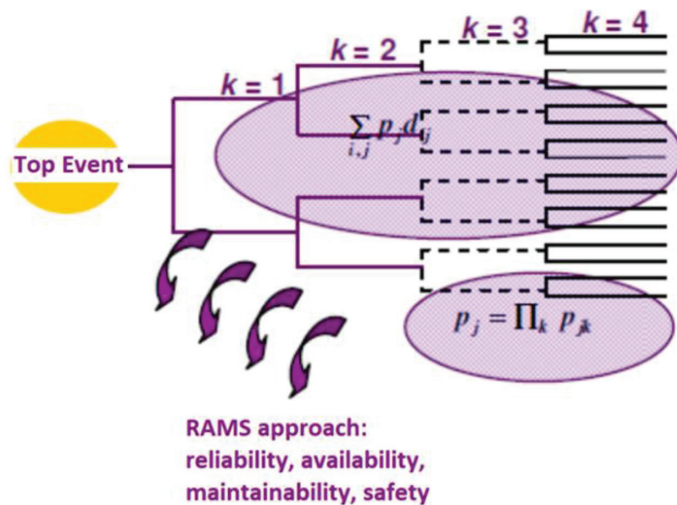


Figure 3: Event tree analysis.

- *Individual Risk (IR)*

$$IR = \frac{TR}{N_{\text{exposed}}} = \frac{\sum_{i=1}^n p_i \cdot d_i}{N_{\text{exposed}}} \quad (2)$$

*IR* is the death probability for the generic user on a single use occasion.

- *Social Risk (SR)*

*Social Risk* is the representation of the function  $F=p(N \geq N^*)$  – *Complementary Cumulative Function (CCF)* of the probability density function of damage. Usually, *SR* measures *N* (*number of fatalities*) and does not consider material damage; so, the *CCF* is represented in the *F-N* plane.

### 3 REVIEW OF SR AND F-N FUNCTION

The technical standards frequently require *F-N* verification criteria. In some occasions, the project compliance is focused on societal risk and individual risk.

The Social Risk shows the distribution of the probability of increasing accidents. It provides useful data to organize prevention and rescue resources. Nevertheless, the Social Risk is an unregulated indicator of exposed subjects to the risk, and it should not be used to assess the acceptability of danger.

The QRA was developed in 1960–1970. It defined ethically and technically current risk indicators [13].

Farmer sets out the first compliance criteria of the *F-N* [14]. He used the retrofitting risk function to show the Curie (Ci) radiation effect of Iodio 131. He proposed a limit of unacceptability defined by a straight line in *FN* log-log plan, which is a hyperbole branch on a uniform metric plane. He also introduced the adverse risk criteria, putting a slope greater than 1 (specifically 1.5) in the *F-N* criteria.

Farmer developed the systematization of the social risk represented on the *F-N* plane and the verification criteria to identify acceptable levels of risk in the safety management of the entire sector.

In the 1970s and 1980s Lees used *F-N* functions to assess the territorial risk and to comparison the risk of industrial accidents to natural events [15]. In the 1980s and 1990s, the social risk criteria were used to manage the safety of production sectors and were introduced into the safety policy framework [13]. Recently, *F-N* criteria were used to verify the safety of the individual structures, because of the QRA methods diffusion and of simulation programs and greater calculation possibilities [16–19].

The use of *SR* is a matter of debate inside the scientific community, especially in the statistic one. Ref. [20] showed that the adoption of *SR* criteria equals the adoption of a Minimax criteria. Moreover, they proved that the use of absolute eligibility criteria of social risk indicator caused the selection of inefficient projects compared to Cost-Benefits Analysis (CBA). The theory was debated by Horn and colleagues years later without being refuted. The use of the *F-N* function criteria in technical and engineering setting, that is, the research field in economic risk management indirectly supported this theory [21].

It is of great importance to consider also the ethical significance of risk indicator criteria. Many authors wrote about this subject [17]. Guarascio *et al.* studied the ethical meaning of indicators [22]. The *IR* has an ethical aspect related to the right of safety for the citizen-user [23]. It is the unique societal risk indicator, it is not standardized, and it does not provide data about the right of safety for the users of the service/system.

For example, we can consider these two systems. The first system (called *trap*) is used by 1 user in 1 year; when the system is used, the user dies. The second system (called *road tunnel*) is used by 1,000,000 users in 1 year. There is one deadly accident in annual use. The two systems (*trap* and *tunnel*) have an identical F-N function (the same SR, and the same Expected Value of Damage – EVD, integer of F-N function). The two systems would therefore respond in the same way to SR verification criteria (on F-N plane). The IR, that is a normalized indicator, is the unique risk indicator that highlights the substantial difference between the two systems: from the point of view of the user, the probability of death of the *trap* is 1,000,000 times higher than that of the tunnel.

#### 4 RELIABILITY OF QRA

Taking into consideration their statistical consistency, it is possible to evaluate the indicators. The estimator can be statistically significant if, increasing his number, the distribution probability of the sample is focused on the to be estimated value of the parameter. To qualify the risk indicators by a statistical point of view, it is necessary to consider that the pairs  $\{p_i, d_i\}$  are the estimate  $PDF(N)$ , probability density function of *r.v.* fatalities.

The reliability of the result of QRA depends on the correct estimation of Top Event probability  $p(TE)$  and the correct estimate of  $PDF(N)$ .

The error of  $p(TE)$  is directly transmitted to the result of the analysis. This happens with both the reference to the synthetic indicators (expected value of the damage, individual risk) and the social risk indicator represented on the  $F-N$  plane. The percentage increase of  $p(TE)$  determinates the same percentage increase of  $R$  and  $IR$  and the translation of the accumulated function of the damage on  $F-N$  plane along the y-axis.

The effect on the risk indicators of the  $PDF(N)$  estimation error is less recognizable. The  $PDF(N)$  is estimated by the partition graph (*Event Tree*). Every *i-th* scenario is defined by:

- $p_i$ , multiplication of probability of ETA;
- $d_i$ , evaluation of the scenario *i-th* associated damage (it depends on reliability and availability of mitigation systems and contour and scenario condition).

Therefore, the estimate of the  $PDF(N)$  is due to:

1. the graphic structure ETA *effectiveness* and *completeness*;
2. the *reliability* of damage estimation of each scenario, particularly the *precision* and the *accuracy* of every  $d_i$  estimation.

The concept of precision expresses the level of dispersion of the estimation obtained by sampling. The precision of an estimation is indifferent to systematic errors while decreasing in the case of random errors.

The concept of accuracy expresses the degree of correspondence between the statistical data obtained from sampling and the actual data obtained by the difference between the average value of the sample and the average value of the population. The accuracy of an estimation decreases with systematic errors, while it is indifferent to random errors.

##### 4.1 Effectiveness and completeness of ETA

The only guarantee for the effectiveness of the analysis is the formalization quality of ETA.

Every level of ETA represents a variable that has an influence on the evolution of the incident, therefore on the damage [24]. Every variable can reach one of  $n$  state;  $n$  states have to constitute a complete partition and respect the principles of completeness and synthesis.

The estimated damage  $d_i$  for every scenario is an *r.v.*;  $N$  is an *r.v.* damage referred to the whole ETA. It is an *r.v.* obtained by the aggregation of *r.v.*: its *PDF* is the collection of  $d_i$ . The frequency of every  $d_i$  is equal to  $p_i$ , multiplication of ETA probabilities. Very structured ETA (high number of scenarios) maintains low the variance of  $d_i$  (intergroup variance), and so the accuracy of the estimated  $d_i$  is improved. The presence of a high number of ETA generated scenarios also improves the likelihood of the whole analysis. This aspect is more important if the phenomenon is unstable (singularity). An effective reconstruction of the scenarios is so obtainable and the analysis likely.

It is not possible to formulate ETA completeness rules. However, it is possible to recognize evidence of incompleteness. ETA is certainly incomplete if it not includes:

1. A level of analysis (of reliability or availability) for every factor that affects the damage (rather that has a significant effect on the damage), including emergency procedures;
2. All possible exposure scenarios (crowding ranging between 0 and maximum capacity);
3. The most fortunate injury case (accident with no damage);
4. The case of apocalyptic damage (all values are destroyed).

This means ETA is certainly incomplete if the  $N$  variation range is different from  $[0; N_{max}]$ .

A continuous *r.v.* can be discretized in a finite number of *modality* or *condensation intervals* (complete partition of the range of the variable): this approach is consistent with Vapnik's theory [25] about *PDF* sample. In certain cases, it is possible to synthesize the variable in two modalities only (dummy or Boolean variable). Other cases require more modalities.

To discrete continuous variables, the number of condensation intervals should grow as the irregularity *PDF* of *r.v.* increase. The choice of few and central condensation intervals causes a distribution queues censorship. Good condensation intervals on continuous *r.v.* are the deciles of distribution.

#### 4.2 Reliability of estimation of $d_i$

The statistical reliability of estimation of damage  $d_i$  associated in the final scenario depends on the reliability of the simulation (*i.e.* fire simulation under confinement, or explosion, or spread of toxic). The reliability of the estimation depends on the *precision* and *accuracy*. This fragility of QRA is well-known to analysts and obvious to everyone [26].

The problem of simulation precision is less worrying. The simulation of a scenario with high turbulence is inaccurate; similarly, a simulation with high and prolonged interaction between simulated objects is inaccurate (agents, danger, performance of the subsystems of protection ...). To get sufficient precision can be increased the number of simulations: the estimate improves with square root of simulation number if output distribution is normal.

The solution to the problem of *low accuracy* of the simulation is difficult; the accuracy depends on the ability of a simulator to return a likely result of the phenomenon [27]. Can be obtained a measure of the accuracy of a simulator with back-analysis of case studies or tests. Again, in this case, the reliability of the inferences of the good outcome of a specific validation case should always be considered.

### 4.3 The consistency of risk indicators

For a conscious use of indicators and verification criteria, it is important to measure the statistical consistency and stability of QRA indicators.

$TR$  is the expected value of  $r.v. N$ ;  $N$  is an aggregation of  $d_i$ . Every  $d_i$  is  $r.v.$  as it is independently estimated through simulation. The expected value of  $N$  is the weighted average value of  $d_i$  (linear combination of expected value of  $d_i$  with  $p_i$  weights):

$$E(N) = \sum_{i=1}^n p_i \cdot E(d_{i,x}) = \sum_{i=1}^n p_i \cdot d_i = TR \tag{3}$$

The variance of  $N$  will instead be:

$$s(N) = E[(N - E(N))^2] = \sum_{i=1}^n p_i (d_i - E(N))^2 = \sum_{i=1}^n p_i \left( d_i - \sum_{i=1}^n p_i \cdot d_i \right)^2 \tag{4}$$

By increasing ETA structure, the  $E(N)$  and  $s(N)$  estimate improve. It is particularly demonstrable that as the numbers of scenarios grows the  $s(N)$  goes from above to  $s(N)$ .

If we consider  $TR$  as the expected value of  $r.v. N$ , this implies that as the completeness of the ETA increases the standard error of  $TR$  decreases; this demonstrates that  $R$  is a robust, statistically consistent and centred indicator.

Is it possible to apply the same considerations to the *Individual Risk* indicator, derived from the division of  $TR$  by the number of exposed people.

The reliability of  $CCF$  can be analysed from a quality point of view. The  $F-N$  function (in conformity with the analytic definition of the *Social Risk* indicator) is calculated by many scenarios that decrease when the  $N$  increases. The result is that as the  $N$  grows, the reliability of  $F$  estimates decreases.

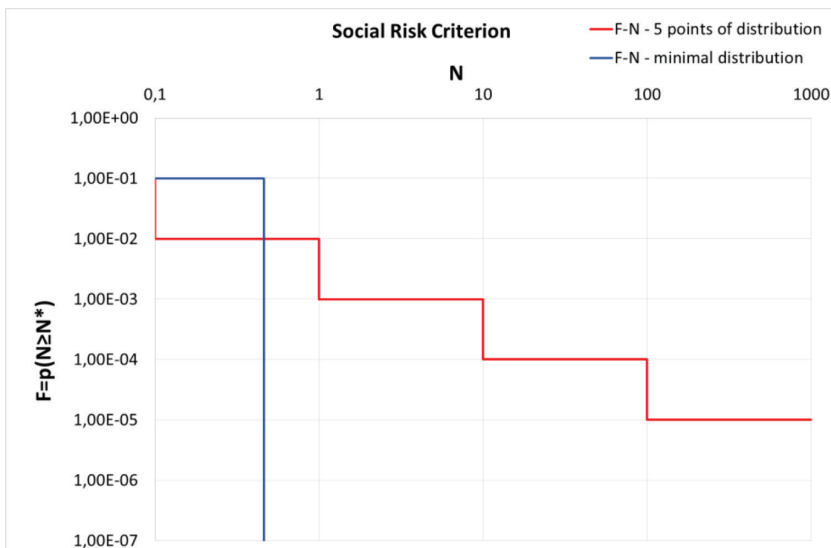


Figure 4: Comparison of F-N: same technical system, same TR, different synthesis of ETA.



If we represent same system  $F-N$  obtained from different partitions (keeping unaltered the range), it is possible to point out the low statistical consistency of the  $SR$  indicator.

Figure 4 shows two  $SR$  functions of the same technical system; both  $F-N$  has the same  $TR$  (integral of  $CCF$ ). Because of distortion of the *log-log graph*, the equivalence of the areas does not appear. The Fig. 4 shows a general characteristic of the  $F-N$  function: if the statistical complexity is too much synthesized, the risk is to obtain an underestimation of  $F-N$  distribution in the right part of the graph and an overestimation of  $F-N$  distribution in the left part of the graph. This aspect is quite serious: the use of more approximate models may lead to an underestimation of  $F-N$  function in case of high  $N$ . Therefore, the shape of  $SR$  is unstable, and it may be unreliable to establish on its form a design criterion.

It should be noted that in the analysis the statistical synthesis does not affect *Total Risk (TR or EVD)* or *Individual Risk (IR)*.

The  $F-N$  form is useful to highlight completeness and reliability limits: if the  $F-N$  has got few points, the ETA is poorly developed and the  $F-N$  form is very unreliable.

## 5 CONCLUSION

This analysis showed that  $SR$  is characterized by instability, and it significantly changes its form when there is frequent completeness of the analysis errors. Due to the analysis incompleteness, the low statistical consistency may underestimate  $F-N$  distribution in the right part of the graph and overestimate  $F-N$  distribution in the left part of the graph.

The use of  $F-N$  as safety indicator brings to a deviating from the effective allocation of the investments. Indeed, the economic resources are used not to reduce the risk but to improve the geometrical compliance of the function  $F-N$  to the verification criterion. The use of criteria for  $SR$  to measure the safety of the single systems leads to a distortion in the  $CBA$  applications. The applications become even more serious as the methods are used separately, applied to single structures or systems.

The chosen criteria on  $SR$  impose a condition of  $F-N$  distribution form and only indirectly a limit to the risk, whereas the fixed limit values on  $IR$  impose a limit to the risk to be achieved through the best investments for  $CBA$ .

Concerning what we said, an aware use of  $SR$  on  $F-N$  plane is to prefer:  $SR$  is useful to represent what happened in large sectors and to observe the safety of whole sectors. The use of  $SR$  for the single system should not be combined with verification criterion. In this case, the  $F-N$  is appropriate to determinate the reliability limits of the analysis in order to make *effectiveness* and *completeness* of Event Tree.

The safety project of a system should be done according to  $IR$  – quantity normalized indicator – both for its ethic consistency (the expression of citizen-user right to safety) and for the better statistic property.

## REFERENCES

- [1] Directive 2004/54/EC of the European parliament and of the council of 29 April 2004 on minimum safety requirements for tunnels in the trans-European road network.
- [2] Italian Legislative Decree 264:2006.
- [3] Directive 2008/96/EC of the European Parliament and of the Council of 19 November 2008 on road infrastructure safety management.
- [4] Italian Legislative Decree 35:2011.
- [5] PIARC *Technical Committee 3.3 Road tunnel operation*.
- [6] PIARC *Technical Committee 3.2 design and operation of Safer Road Infrastructure*.



- [7] Italian Legislative Decree 264:2006 Annex 3, §1.
- [8] IEC 61508 *Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems*.
- [9] Italian Legislative Decree 264:2006 Section 4, Annex 3
- [10] Italian Legislative Decree 264:2006 Section 4, Annex 2
- [11] Faber, M.H., & Stewart, M.G., Risk assessment for civil engineering facilities: critical overview and discussion. *Reliability Engineering & System Safety*, **80**(2), pp. 173–184, 2003.  
[https://doi.org/10.1016/s0951-8320\(03\)00027-9](https://doi.org/10.1016/s0951-8320(03)00027-9)
- [12] IEC 61508 *Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems*
- [13] Ball, D.J., & Floyd, P.J., *Societal risks*. London: HSE, 1998.
- [14] Farmer, W.R., *The synoptic problem: A critical analysis*. Mercer University Press, Macon, GA, 1976.
- [15] Lees, F.P., *Loss Prevention in the Process of Industries* (Vols. 1–3, 2nd ed.). Butterworth Publication, London, 1996.
- [16] Jonkman, S., van Gelder, P. & Vrijling, J., An overview of quantitative risk measures for loss of life and economic damage. *Journal of Hazardous Materials*, **99**(1), pp. 1–30, 2003.  
[https://doi.org/10.1016/s0304-3894\(02\)00283-2](https://doi.org/10.1016/s0304-3894(02)00283-2)
- [17] Trbojevic, V., *Risk criteria in e.u. Presented at Conference on European Safety and Reliability* (ESREL'05), Tri-city, Poland, 2005.
- [18] Jonkman, S., *Loss of life estimation in flood risk assessment theory and application*. Doctoral Thesis, Technical University of Delft, Netherlands, 2007.
- [19] Proske, D., *Catalogue of Risks Natural, Technical, Social and Health Risks*. Springer, Berlin, 2008.
- [20] Evans, A.W. & Verlander, N.Q., What is wrong with criterion FN-lines for judging the tolerability of risk? *Risk Analysis*, **17**(2), pp. 157–168, 1997.  
<https://doi.org/10.1111/j.1539-6924.1997.tb00855.x>
- [21] Abrahamsen, E.B. & Aven, T., On the consistency of risk acceptance criteria with normative theories for decision-making. *Reliability Engineering & System Safety*, **93**(12), pp. 1906–1910, 2008.  
<https://doi.org/10.1016/j.ress.2008.03.021>
- [22] Guarascio, M., Lombardi, M. S., Rossi, G. & Sciarra, G., *Risk Analysis and Acceptability Criteria*. WIT Transactions on The Built Environment, 2007.
- [23] Hartford, D.N.D., Legal framework considerations in the development of risk acceptance criteria. *Structural Safety*, **31**(2), pp. 118–123, 2009.  
<https://doi.org/10.1016/j.strusafe.2008.06.011>
- [24] Guarascio, M., Lombardi, M. S., Rossi, G. & Sciarra, G. *Road Tunnel Safety Rules in Italy: The Tunnel Country*. WIT Transactions on The Built Environment, 2007.
- [25] Vapnik, V. & Chapelle, O., Bounds on error expectation for support vector machines. *Neural Computation*, **12**(9), pp. 2013–2036, 2000.  
<https://doi.org/10.1162/089976600300015042>
- [26] Mona, K.R., *Global Risk Assessment of Natural Disasters: New Perspectives*, Thesis, University of Waterloo, Ontario, Canada, 2014.
- [27] Vanem, E., Ethics and fundamental principles of risk acceptance criteria. *Safety Science*, **50**(4), pp. 958–967, 2012.  
<https://doi.org/10.1016/j.ssci.2011.12.030>