Identification of Protective Actions to Reduce the Vulnerability of Safety-Critical Systems to Malevolent Intentional Acts: An Optimization-Based Decision-Making Approach

Identification of protective actions to reduce the vulnerability of safety-critical systems to malevolent intentional acts: an optimization-based decision-making approach

T. R. WANG,1

1 Chair on System Science and the Energetic Challenge, EDF Foundation,

Laboratoire Genie Industriel, CentraleSupélec/Université Paris-Saclay,

Rue Joliot Curie, 3 - 91190 Gif-sur-Yvette, France

N. PEDRONI,2*

2 NEMO Group, Energy Department, Politecnico di Torino,

Corso Duca degli Abruzzi, 24 - 10129 Torino, Italy

E. ZIO,3,4,5

3 MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France

4 Energy Department, Politecnico di Milano, Via Giuseppe La Masa 34, Milan, 20156, Italy

5 Eminent Scholar, Department of Nuclear Engineering, College of Engineering,

Kyung Hee University, Republic of Korea

V. MOUSSEAU,6

6 CentraleSupélec/Université Paris-Saclay, 3 Rue Joliot Curie, 91190 Gif-sur-Yvette, France

**Address correspondence to Nicola Pedroni, NEMO Group, Energy Department,*

Politecnico di Torino, Corso Duca degli Abruzzi, 24 - 10129 Torino, Italy;

tel: +39-0110904419; nicola.pedroni@polito.it

ABSTRACT: An empirical classification model based on the Majority Rule Sorting (MR-Sort) method has been previously proposed by the authors to evaluate the vulnerability of safety-critical systems (in particular, nuclear power plants) with respect to malevolent intentional acts. In this paper, the model serves as the basis for an analysis aimed at determining a set of protective actions to be taken (e.g., increasing the number of monitoring devices, reducing the number of accesses to the safety-critical system, etc) in order to effectively reduce the level of vulnerability of the safety-critical systems under consideration.

In particular, the problem is here tackled within an optimization framework: the set of protective actions to implement is chosen as the one minimizing the overall level of vulnerability of a group of safety-critical systems. In this context, three different optimization approaches have been explored: (i) one single classification model is built to evaluate and minimize system vulnerability; (ii) an ensemble of compatible classification models, generated by the bootstrap method, is employed to perform a ”robust” optimization, taking as reference the ”worst-case” scenario over the group of models; (iii) finally, a distribution of classification models, still obtained by bootstrap, is considered to address vulnerability reduction in a ”probabilistic” fashion (i.e., by minimizing the ”expected” vulnerability of a fleet of systems). The results are presented and compared with reference to a fictitious example considering nuclear power plants as the safety-critical systems of interest.

200 CHARACTER SUMMARY: A method is proposed that identifies optimal strategies to reduce the vulnerability of safety-critical systems to malevolent intentional acts. An application to nuclear plants shows good results.

KEYWORDS: Risk management, safety-critical system, malevolent intentional attacks, inverse

classification problem, optimization-based approach

1 INTRODUCTION

The vulnerability of safety-critical systems, like nuclear power plants, is of great concern, given the multiple and diverse hazards that they are exposed to (e.g., intentional, random, natural etc.) (Kroger& Zio, 2011) and the potential large-scale consequences. This justifies the increased attention for analyses aimed at (i) the systematic identification of the sources of system vulnerability, (ii) the qualitative and quantitative assessment of system vulnerability (Aven, 2003; Aven, 2010) and (iii) the definition of effective actions of vulnerability reduction.

The issues at stake involve uncertainty given the long time frame, capital intensive investment and large number of stakeholders with different views and preferences, and call for suitable decision analysis (DA) methods (Leroy, Mousseau, & Pirlot 2011) and particularly multiple criteria decision-making (MCDM) (Doumpos & Zopounidis, 2002; Belton & Stewart, 2002).

A number of examples of applications of MCDA approaches to the assessment/ ranking/ prioritization of the vulnerability of safety-critical systems exist. Apostolakis and Lemon (2005) and Patterson and Apostolakis (2007) focus on the identification of critical locations in infrastructures.The vulnerabilities and their ranking according to potential impacts are obtained by Multi-Attribute Utility Theory (MAUT) (Morgan, Florig, DeKay, & Fischbeck, 2000). Koonce, Apostolakis, and Cook (2008) have proposed a methodology for ranking components of a bulk power system with respect to its risk significance to the involved stakeholders. Johansson and Hassel (2010) have proposed a framework for considering structural and functional properties of interdependent systems and developed a predictive model in a vulnerability analysis context. Piwowar, Chatelet, and Laclemence (2009) have proposed a systemic analysis which accounts for malevolence, i.e., the willingness to cause damage. Cailloux and Mousseau (2011) have

proposed a framework to evaluate and compare the threats and vulnerabilities associated with territorial zones according to multiple criteria (industrial activity, population, etc.) by using an adapted ELECTRE method (Corrente, Doumpos, Greco, Słowinski, & Zopounidis, 2017). In (Teng, Thekdi, & Lambert, 2012 and 2013), systemic approaches are proposed to identify and evaluate *priorities* in the business process of risk and safety organizations and to assess the performance of risk, safety, vulnerability and security programs. In particular, with respect to the context of interest to the present paper, the authors answer the following questions: (i) how multiple risk assessment, management and communication procedures, missions or actions should be administered, coordinated and possibly optimized; and (ii) what should be the basis for *resource allocation* to these activities. Based on the proposed approach, the relative priorities among policy *initiatives* and *actions* are quantified, in order to preserve and sustain the compliance of risk, safety and security programs with organizational and administrative guidelines. In (Thorisson et al., 2017), risk is addressed in terms of sensitivities of a multicriteria portfolio optimization, identifying which scenarios are most and least disruptive to multicriteria optimization. In particular, the approach is used to identify those "stressors" (e.g., natural disasters, mismanaged funds, lack of agency cooperation) that most influence a *prioritization* of *initiatives* in the electrical power sector in Afghanistan.

In a previous work (Wang, Mousseau, & Zio, 2013), the authors have proposed an empirical classification framework to tackle issues (i) and (ii) above, considering the analysis of the vulnerability of nuclear power plants to malevolent intentional acts. Specifically, we have developed a classification model based on the Majority Rule Sorting (MR-Sort) method (Leroy et al., 2011) to assign an alternative (i.e., a nuclear power plant) to a given (vulnerability) class (or category). The MR-Sort classification model contains a group of (adjustable) parameters that are calibrated by means of a set of empirical classification examples (also called training set),

i.e., a set of alternatives with pre-assigned vulnerability classes (Leroy et al., 2011; Wang et al., 2013). The performance of the classification-based vulnerability analysis model in terms of accuracy and confidence in the assignments has been thoroughly and systematically assessed in (Wang, Mousseau, Pedroni, & Zio, 2014).

In this paper, we are still concerned with intentional hazards (i.e., those related to malevolent acts) and address issue (iii) above, i.e., the definition of the actions to undertake for reducing the level of system vulnerability. In particular, the empirical classification model developed in (Wang et al., 2013) is tailored to address the corresponding *inverse (classification)* problem (Aggarwal, Chen, & Han, 2010; Aggarwal, Chen, & Han, 2006; Li, Zhou, & Zhang, 2012; Mousseau & Slowinski, 1998; Mousseau, Ozpeynirci, & Ozpeynirci, 2018; Pendharkar, 2002; Ahuja & Orlin, 2001 and 2002; Heuberger, 2004; Mannino & Koushik, 2000; Lin, Kuo, Hsieh, & Wang, 2009), i.e., the problem of determining a set of protective actions (Larsson, 1992; Doumpos & Zopounidis, 2002), which can effectively reduce the vulnerability class of (a group of) safety-critical systems (Aven & Flage, 2009), taking into account a specified set of constraints (e.g., budget limits) (Aggarwal et al., 2010).

The present analysis can be considered part of an encompassing business process of safety management (see, e.g., (Teng et al., 2012 and 2013; Thekdi & Lambert, 2014)), where we seek for the best compromise among risks, costs and benefits in allocating investments in safety-critical systems in the presence of uncertainties (Lambert & Farrington, 2007). Mathematically speaking, the aim is to identify how to modify some features of the input patterns to the classification model (i.e., the attributes of the safety-critical system under analysis) such that the resulting class is changed as desired (i.e., the vulnerability category is reduced to a desired level).

In previous research by the authors, novel sensitivity indicators (Hofmann et al., 2013) have been introduced for quantifying the variation in the vulnerability class of a safety-critical sys-

tem resulting from the application of a given set of protective actions (NWRA, 2002). One single combination of actions is obtained that if applied to each of the alternatives in the group of systems, it allows reducing the overall vulnerability of the group (Wang, Mousseau, Pedroni, & Zio, 2016). However, in practice, under given constraints (e.g., a limited budget), it is more reasonable to find *one* set of protective actions for *each* of the considered alternatives, such that the *overall* vulnerability level of the *group* of safety-critical systems under consideration is *minimized*. To this aim, an *optimization-based* framework is here undertaken. In this context, three different optimization approaches have been sought: (i) *one single* classification model is built to evaluate and minimize system vulnerability, (ii) an *ensemble* of compatible classification models, generated by the bootstrap method, is employed to perform a "robust" optimization, taking as reference the "worst-case" scenario over the group of models; (iii) finally, a distribution of classification models, still obtained by bootstrap, is considered for the vulnerability reduction task, by minimizing the "expected" vulnerability of the fleet of plants. All three optimization problems are numerically solved by CPLEX.

In the framework of interest to the present work, it is known that existing risk assessment methodologies may fail to account for unknown and emergent risks that are typical of large-scale infrastructure investment allocation problems. On the other hand, in modern portfolio theory, it is well known that a diversified portfolio can be very effective to reduce non-systematic risks. The approach of diversification is equally important in choosing robust portfolios of infrastructure projects that may be subject to emergent and unknown risks (Joshi & Lambert, 2011; Thorisson, Lambert, Cardenas, & Linkov, 2017). The proposed methodology is expected to contribute also in this direction of optimal classification of options/investments and combinations of the same.

In summary, the main methodological and applicative contributions of the present paper are the

following:

- the empirical classification model used to assess safety-critical system vulnerability to intentional hazards has been entirely developed by the authors;
- the bootstrap-based robust and probabilistic optimization frameworks here undertaken to address the inverse classification problem have been originally proposed by the authors;
- to the best of the authors' knowledge, it is the first time that an inverse classification problem is formulated and considered for the optimization of the choice of protective actions to reduce the vulnerability of a group of safety-critical systems (e.g., Nuclear Power Plants), taking into account the uncertainty associated to the classification models.

The remainder of the paper is structured as follows. Section 2 recalls the classification model for the assessment of vulnerability to intentional hazards. With reference to that, Section 3 introduces the problem of inverse classification for choosing protective actions and the optimization decision-making approach. In Section 4, case studies are proposed to show the applications of the method. Finally, Section 5 gives the discussion and analysis of the results. The conclusions of this research is drawn in Section 6.

2 CLASSIFICATION MODEL FOR THE ASSESSMENT OF VULNERABILITY TO INTENTIONAL HAZARDS

We limit the vulnerability analysis of a safety-critical system to the evaluation of the susceptibility to intentional hazards. For this, we adopt the three-layers hierarchical model developed in (Wang et al., 2013) (Figure 1). The susceptibility to intentional hazards (layer 1 in Figure 1) is characterized in terms of attractiveness and accessibility (layer 2 in Figure 1). These at-

tributes are hierarchically broken down into factors which influence them, including resilience interpreted as pre-attack protection (which influences on accessibility) and post-attack recovery (which influences on attractiveness). The disaggregation is made in n criteria (layer 3 in Figure 1) described by the n-tuple $MCrit = \{MCrit_1, MCrit_2, ..., MCrit_i, ..., MCrit_n\}$ with $n = 6$ in this case: physical characteristics ($MCrit_1$), social criticality ($MCrit_2$), possibility of cascading failures ($MCrit_3$), recovery means ($MCrit_4$), human preparedness ($MCrit_5$) and level of protection ($MCrit_6$). These six criteria are further decomposed into a layer of $m = 16$ independent basic subcriteria $\{crit^j, j = 1, 2, ..., m = 16\}$(layer 4 in Figure 1), for which data and information are collected in terms of quantitative values or linguistic terms depending on the nature of the subcriterion. The descriptive terms and/ or values of the fourth layer subcriteria are, then, scaled to numerical categories. Finally, to get the value of the six third-layer criteria $MCrit = \{MCrit_1, MCrit_2, ..., MCrit_i, ..., MCrit_n\}, n = 6$, (i) we assign weights to each subcriterion to indicate their importance (e.g., experts assess the contribution of each subcriterion to the corresponding third-layer criterion) and (ii) considering the independence between the subcriteria, we apply a simple weighted sum to the categorical values of the constituent subcriteria $\{crit^j = j = 1, 2, ..., m = 16\}$. These $m = 16$ criteria $\{crit^j = j = 1, 2, ..., m = 16\}$ are evaluated to assess the vulnerability of a given safety-critical system of interest (e.g., a nuclear power plant – NPP).

Notice that (weighted) hierarchical-tree based structures have been previously used to address the complexity of safety-critical systems: see, e.g., (Courtois, 1985; Haimes, 2012; Larsson, 1992; Lind, 2011a; Lind, 2011b; Ruan, 2000; Zio, 2007). The criteria of the layers are defined and assigned preference directions for treatment in the decision-making process. The preference direction of a criterion indicates towards which state it is desirable to lead it to reduce susceptibility, i.e., it is assigned from the point of view of the defender of an attack who is concerned

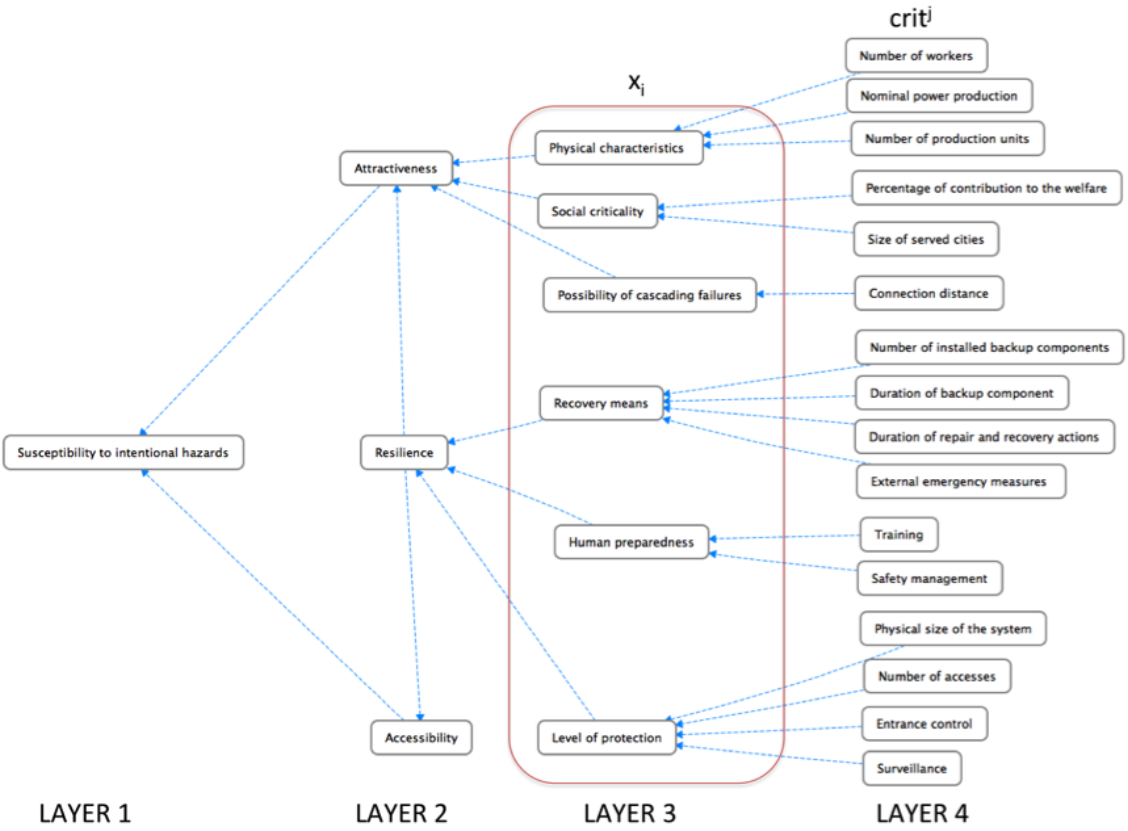

Figure 1: Hierarchical model for susceptibility to intentional hazards

with protecting the system. For the purpose of the present analysis, $M = 4$ levels (or categories) of system vulnerability $\{A^h : h = 1, 2, 3, 4\}$ are considered: A^1 = satisfactory, A^2 = acceptable, A^3 = problematic, A^4 = serious. Then, the assessment of vulnerability corresponds to a classification problem: given the definition of the characteristics of a critical system in terms of the sixteen criteria above, assign the vulnerability category (or class) to which the system belongs. The quantitative model introduced for representing the susceptibility to intentional hazards employs categorized, discretized vulnerability levels: in this sense, the approach may be considered similar to the one underlying the use of risk matrices. However, in our case, the susceptibility to intentional hazards is computed by a simple weighted (linear additive) sum of the six main criteria (with categorized values), enabling effective risk management by mapping ordered categorical ratings of severity into recommended risk management decisions or priorities, as optimal resource allocation may depend crucially on other quantitative information (as reflected in the

following sections, e.g., in terms of costs of different protective actions).

Categorizing the subcriteria may, in certain cases, require subjective judgements. However, since most subcriteria are easily quantified (e.g., the number of workers), the potential for inconsistencies in judgements can be eliminated by consensual agreement among the decision makers. For future work, the criteria may be represented also in other forms, e.g., a continuous scale of measurement, to eliminate the problem completely.

The classification model, indicated as $M(\cdot|(\omega, b))$, is based on the Majority Rule Sorting (MR-Sort) method (Leroy et al., 2011; Roy, 1991; Mousseau & Slowinski, 1998). It is a simplified version of ELECTRE Tri (Corrente et al., 2017), an outranking sorting procedure in which the assignment of an alternative to a given category is determined using a complex concordance non-discordance rule (Roy 1991; Mousseau & Slowinski, 1998). We assume that the alternatives to be classified (in this paper, a safety-critical system or infrastructure of interest, e.g., a nuclear power plant) can be described by an *n*-tuple of elements $x = \{x_1, x_2, ..., x_i, ..., x_n\}$, which represent the evaluation of the alternatives with respect to a set of *n* criteria (by way of example, in the present paper the criteria used to evaluate the vulnerability of a safety critical system may include its physical characteristics, social criticality, level of protection and so on, as presented above) (Almeida-Dias, Figueira, & Roy, 2012; Corrente et al., 2017). We denote the set of criteria by $N = \{1, 2, ..., i, ..., n\}$ and assume that the values x_i of criterion *i* range in the set X_i (for example, in the present paper all the criteria range in [0, 1]) (Rocco & Zio, 2005) . The MR-Sort procedure allows assigning any alternative $x = \{x_1, x_2, ..., x_i, ..., x_n\} \in X = X_1 \times X_2 \times ... \times X_i \times ... \times X_n$ to a particular pre-defined category (in this paper, a class of vulnerability), in a given ordered set of categories, $\{A^h : h = 1, 2, ..., k\}$; as mentioned above, $k = 4$ categories are considered in this work: A^1 = satisfactory, A^2 = acceptable, A^3 = problematic, A^4 = serious. To this aim, the model is further specialised in the following way:

- We assume that X_i is a subset of $\mathbb{R}$ for all $i \in \mathbb{N}$ and the sub-intervals $(X_i^1, X_i^2, ..., X_i^h, ..., X_i^k)$ of X_i are compatible with the order on the real axis, i.e., for all $x_i^1 \in X_i^1, x_i^2 \in X_i^2, ..., x_i^h \in X_i^h, ..., x_i^k \in X_i^k$, we have $x_i^1 > x_i^2 > ... > x_i^h > ... > x_i^k$. We assume furthermore that each interval $x_i^h, h = 2, 3, ..., k$ has a smallest element b_i^h, which implies that $x_i^{h-1} \geq b_i^h > x_i^h$. The vector $b^h = \{b_1^h, b_2^h, ..., b_i^h, ..., b_n^h\}$ (containing the lower bounds of the intervals X_i^h of criteria $i = 1, 2, ..., n$ in correspondence of category *h*) represents the lower limit profile of category A^h.

- There is a weight ω_i associated with each criterion $i = 1, 2, ..., n$, quantifying the relative importance of criterion *i* in the vulnerability assessment process; notice that the weights are normalised such that $\sum_{i=1}^{n} \omega_i = 1$.

  In this framework, a given alternative $x = \{x_1, x_2, ..., x_i, ..., x_n\}$ is assigned to category $A^h, h = 1, 2, ..., k$, iff

$$\sum_{i \in \mathbb{N}: x_i \geq b_i^h} \omega_i \geq \lambda \; and \sum_{i \in \mathbb{N}: x_i \geq b_i^{h+1}} \omega_i < \lambda, \tag{1}$$

  where λ is a threshold ($0 \leq \lambda \leq 1$) chosen by the analyst. Rule (1) is interpreted as follows. An alternative x belongs to category A^h if: 1) its evaluations in correspondence of the *n* criteria (i.e., the values $\{x_1, x_2, ..., x_i, ..., x_n\}$) are at least as good as b_i^h (lower limit of category A^h with respect to criterion *i*), $i = 1, 2, ..., n$, on a subset of criteria that has sufficient importance (in other words, on a subset of criteria that has a weight larger than or equal to the threshold λ chosen by the analyst); and at the same time (2) the weight of the subset of criteria on which the evaluations $\{x_1, x_2, ..., x_i, ..., x_n\}$ are at least as good as b_i^{h+1}(lower limit of the successive category A^{h+1} with respect to criterion *i*), $i = 1, 2, ..., n$, is not sufficient to justify the assignment of x to the successive category A^{h+1}.

  Notice that alternative x is assigned to the best category A^1 if $\sum_{i \in \mathbb{N}: x_i \geq b_i^1} \omega_i \geq \lambda$ and it is

assigned to the worst category A_k if $\sum_{i \in \mathbb{N}: x_i \geq b^{-k-1}} \omega_i < \lambda$. Finally, it is straightforward to notice that the parameters of such a model are the $k \cdot n$ lower limit profiles (*n* limits for each of the *k* categories), the *n* weights of the criteria $\omega_1, \omega_2, ..., \omega_i, ..., \omega_n$, and the threshold λ, for a total of *n(k+1)+1* parameters.

These parameters are calibrated through a disaggregation process by means of a set of empirical classification examples (the training set $D_{TR} = \{(x_p, \Gamma_p^t), p = 1, 2, ..., N\}$, i.e., a set of N alternatives $x_p = \{x_1^p, x_2^p, ..., x_i^p, ..., x_n^p\}, p = 1, 2, ..., N$ together with the corresponding real pre-assigned categories (i.e., vulnerability classes) Γ_p^t (the superscript t indicates that Γ_p^t represents the true, a priori-known vulnerability class of alternative x_p). The number of categories are defined by the experts by assigning a meaning to each of them as "satisfactory", "acceptable", "problematic" and "serious". The definition fixes a part of the structure of the classification model and of the inverse-classification model as well. The operational meaning of the categories may be problem-specific and in practical cases, it may be specified by the experts only after the model is obtained based on the training set.

Further details about the generation of classification models are not reported here for brevity: the interested reader is referred to (Wang, Mousseau, Pedroni, & Zio, 2014).

3 INVERSE CLASSIFICATION PROBLEM FOR PROTECTIVE ACTIONS IDENTIFICATION: AN OPTIMIZATION-BASED DECISION MAKING APPROACH

Classification has been widely studied in the literature because of its applicability to a wide variety of problems (Duda, Hart, & Stork, 2001; James, 1985; Alsabti, Rank,& Singh, 1998; Breiman, Friedman,& Stone, 1984; Bodley & Utgoff, 1995; Breslow, 1997; Friedman, 1977; Gehrke, Ganti, Ramakrishnan, & Loh, 1999; Quinlan, 1993; Zopounidis & Doumpos, 2002).

The multicriteria inverse classification problem (Aggarwal et al., 2010; Aggarwal et al., 2006; Li et al., 2012) stands on the idea that changes in the independent variables of a system solution are searched so that it can be classified into a more desirable class with respect to the given system criteria (Pendharkar, 2002; Mannino & Koushik, 2000; Mousseau et al., 2018; Lin et al., 2009). Specifically, in this work we aim to identify a set of protective actions to reduce the vulnerability of a (group of) safety-critical system(s) under budget limitations.

In such inverse classification problem, we arrived to determine the action-oriented feature variables for an *incompletely specified* test data set characterising the system vulnerability. The aim is to find possibly different choices of actions so that the feature variables are modified in such a way so as that the test data set belongs to a desired class.

If there is no limitation on the choices of the actions (e.g., number of actions that can be applied), the problem can be formulated as that for the case of the training data set, as both the feature and class variables are completely defined in it. On the other hand, for the case of the test data set, the class variables are completely defined but the feature variables are not. Thus, each test data example has a *desired class label* associated with it. The aim of the inverse classification problem is to choose the test feature variables such that the corresponding classification accuracy with respect to the *desired* test classes is maximized.

If there are action-related constraints (e.g., number of actions that can be applied at the same time, budget limitation on expenditures for actions etc.) then, the problem should be modified. Under the given constraints concerning the choice of the actions, the aim of the inverse classification problem is to choose the ”optimal” set of actions that can modify the feature variables such that the corresponding class variables are brought as close as possible to the ”desired” class label.

The problem is ”inverse” because the usual mapping is from a case to its unknown category

whereas here it is the opposite. Specifically, in the classification problem with missing data, one tries to determine the *unknown* class based on incompletely defined features; on the other hand, in the inverse classification problem, one tries to determine the *action-oriented* missing variables that achieve a *desired* class. The inverse classification problem offers the proper framework for a number of action-driven applications in which the features to define certain actions which drive the decision making towards a *desired* end-result (Aggarwal et al., 2010).

As defined by Belton and Stewart (2002), the problem can be treated also as a "Portfolio Decision Analysis (PDA)" for making multiple informed selections from a discrete set of alternatives through mathematical modeling that accounts for relevant constraints, preferences, and uncertainties (Salo, Keisler, & Morton, 2011; Salo & Hamalainen, 1997; Salo & Hamalainen, 2010). It can also be tackled as a multi-objective combinatorial optimization problem, with discreted decision variables and objective functions and constraints that can take any form (Ahuja & Orlin, 2001 and 2002; Heuberger, 2004; Coello Coello, Dhaenens & Jourdan, 2010; Bornstein, Maculan, Pascoal, & Pinto, 2012).

To illustrate the methodology, we consider a set of N alternatives ($x_p, p \in \{1, 2, ..., N\}$) characterized by $m = 16$ basic features ($crit^j, j \in \{1, 2, ..., m\}$), whose data and information are collected in terms of quantitative values or linguistic terms, depending on the nature of the subcriterion, as mentioned in the previous section. Each vector x_p represents one safety-critical system (in our case, a Nuclear Power Plant - NPP). On the basis of these $m = 16$ features, the NPPs are assigned to $M = 4$ pre-defined categories ($\{A^h : h = 1, 2, 3, 4\}$), where A^1 represents the best situation, i.e., lowest vulnerability, as presented in the previous section. Let $act = \{act^1, act^2, ..., act^F\}$ denote the available set of actions, each of which can influence on one or more basic subcriteria $crit^j, j \in \{1, 2, ..., m\}$ (Figure 2).

The $solid\ lines$ (resp. $dotted\ lines$) of one action towards the related subcriteria indicate that

if such action is applied, the subcriteria limited to it will be improved (resp. worsened). For example, if action "Reduce number of workers" is applied, the "Number of workers" and the "Number of production units" will be reduced and we need less resources for "Training" and "Safety management". On one hand this action is "positive" for the Nuclear Power Plant; on the other hand, due to the reduction of workers, we can only work on a smaller "Number of installed backup components", and the "Duration of repair and recovery actions" will take more time. The "preference" directions of all subcriteria and the detailed illustration of all actions are detailed in previous works by the authors (Wang et al., 2013 and 2016).

The influences upon the 16 subcriteria (before the categorisation and weighted integration to obtain the corresponding main criteria) are of different intensity as measured by a set of coefficients $coeff^{kj}, k \in \{1, 2, ..., F\}, j \in \{1, 2, ..., m\}$. In other words, if we analyze the influential relation of the actions on all the subcriteria considered, then $coeff^{kj}$ represents the corresponding consequence of action k on attribute j: the higher the absolute value of $coeff^{kj}$, the stronger the effect of action k on attribute j. Notice that a positive (resp. negative) coefficient $coeff^{kj}$ means that action k has an ameliorative (resp. deteriorative) effect on attribute j, whereas if $coeff^{kj}$ is equal to zero, then criterion j is not influenced by action k. The determination of such coefficients is a crucial step, since they measure the effectiveness of actions on all subcriteria. Significant efforts have been made to assign numerical values to the impacts of actions, in order to represent the problem as realistically as possible. However, in a non-fictitious situation the task is expected to be complex, in particular for linguistically defined subcriteria. Actually, the relations between the actions and the criteria taking into account the dependencies of different attributes and systems are always difficult to identify: in such cases, resorting to the judgment of real experts and possibly to real historical data is mandatory (Ayyub, 2001). In addition, the inevitable uncertainty associated to these coefficients should be possibly propagated through

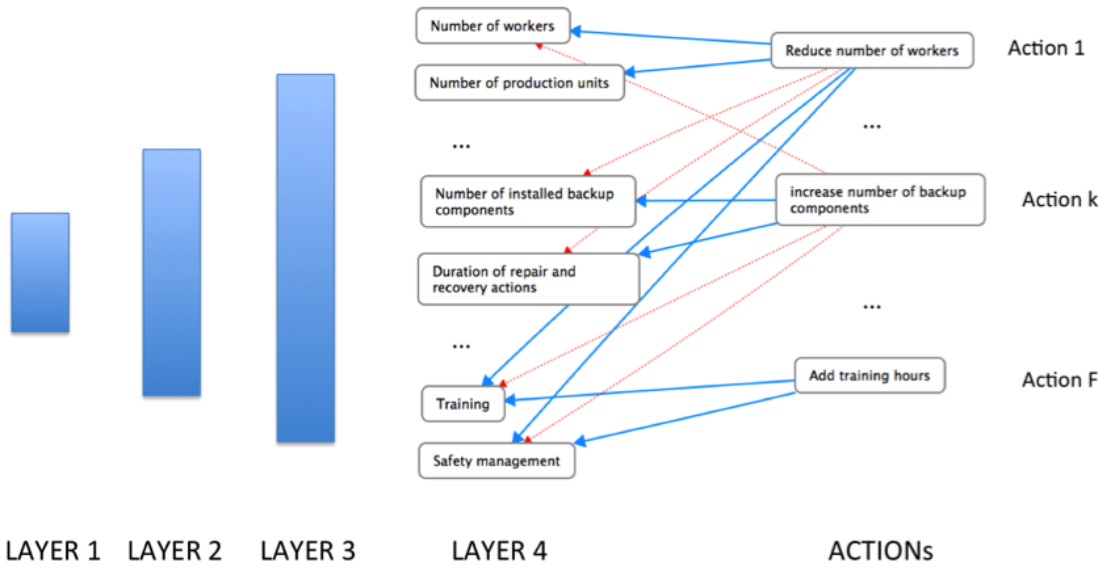

Figure 2: Schema of direct actions for basic criteria

the classification model for a more robust vulnerability assessment and a more reliable (inverse) identification of optimal actions. The implementation of one or more actions modifies the attribute values $crit^j, j \in \{1, 2, ..., m\}$ and as a result, the vulnerability of the system (i.e., the assignment by the classification model) may change. In this paper, we assume that the total effect of the available set of actions $act = \{act^1, act^2, ..., act^F\}$ on criterion j is obtained by a linear superposition of the effects of each action act^k:

$$crit'^j = crit^j + \sum_{k=1}^{F} coeff^{kj} * act^k, k \in \{1, 2, ..., F\}, j \in \{1, 2, ..., m\}. \quad (2)$$

where $crit'^j$ is the value of attribute j after the identified set of available actions has been implemented. The influences of actions are defined in a realistic way, so that after the application of the different combinations of actions upon one NPP, the values of its criteria should still be within a realistic range. In addition, we assume a linear superposition of the effects for the actions. In order to keep the 'after-action' characteristics of one alternative within realistic ranges, the rules of categorisation of the criteria in Layer 3 are properly defined: e.g., if the "number of workers" exceeds a given value, it is always "re-scaled" to the worse categorical value. In more detail, this limitation of the after-action characteristics is done within the optimization process,

i.e., by the use of constraints limiting $crit'^j$. Actually, in the opposite case, the optimization would be misguided to select actions improving some characteristics beyond reasonable levels, thus wasting budget, which could not be used in reality.

Also, let $Cost(x_p, act'), act' \subseteq act$ denote the cost of the combination of actions act' applied to x_p. If $c_k^p (p \in \{1, 2, ..., N\}, k \in \{1, 2, ..., F\})$ is the cost of action k on x_p, then:

$$Cost(x_p, act') = \sum_k aux_k^p * c_k^p, k \in \{1, 2, ..., F\}. \tag{3}$$

where aux_k^p is an auxiliary binary variable that equals 1, if action k is applied to plant x_p, and equals 0, otherwise.

The inverse classification problem can, then, be formulated as follows: given a limited budget B_g for the entire group of NPPs considered, identify, for each NPP, a specifically designed combination of actions that provide the maximal possible reduction in its vulnerability level $A_p^{\lambda'}, A_p^{\lambda'} \in \{1, 2, 3, 4\}, p \in \{1, 2, ..., N\}$ (as presented in the previous Section, the smaller the category value, the less vulnerable the NPP). The combinations of actions for different NPPs may be obviously different. Under certain circumstances, based on the original performance of a given NPP and the performance of the others, one NPP may not need any ameliorative action. In particular, we have chosen the strategy to reduce, under budget constraint, the global vulnerability of a group of alternatives in giving priority to the NPPs that are originally assigned to the worst category; in other words, we try to maximize a properly weighted sum of the ameliorations in the vulnerability categories undergone by all the NPPs.

This can be mathematically represented by a weighted-sum objective function:

$$I^x = \rho_3 * Q_{43} + \rho_2 * Q_{32} + \rho_1 * Q_{21} \tag{4}$$

where $Q_{n(n-1)} (n \in \mathbb{Z})$ represents the overall number of NPPs among the N available ones $\{x | x_p, p \in \{1, 2, ..., N\}\}$ that are ameliorated from category A^n to category A^{n-1} by a given

combination of actions: for example, Q_{43} is the number of plants changing from A^4 to A^3, whereas Q_{32} is the number of plants changing from A^3 to A^2. The constants $\{\rho_i | i \in \{1,2,3\}\}$ represent weights that we assign to the number of ameliorated NPPs $Q_{n(n-1)} (n \in \mathbb{Z})$. There is an operational meaning for the weights. By the choice of the weights, the attitude and strategy of amelioration of the Decision Makers under given constraints can be highlighted. For example, a high value of weight ρ shows a priority of the corresponding "objective" Q: if, for example, ρ_3 is far larger than ρ_1, then, the Decision Makers would like to concentrate preferably on the NPPs that are in danger (i.e., those with higher vulnerability); if all the ρ_i are instead the same, then the Decision Makers are only interested in the overall number of NPPs that are ameliorated. In this paper, we have set:

$$\rho_3 = 100, \rho_2 = 50, \rho_1 = 25. \tag{5}$$

The idea is to pay more attention to the amelioration of the NPPs that are originally assigned to the worse categories, because more critical with respect to being susceptible to attacks. In order to implement it into the algorithm, the set of ρ should be well determined. In the present paper, this has been done by a "reasoned" trial-and-error procedure. The chosen set $\rho_3 = 100, \rho_2 = 50, \rho_1 = 25$ has been found to meet the following requirements: (1) the corresponding objective function is effective in driving the search giving priority to the amelioration of the worst NPPs; (2) the computational time required to solve the corresponding optimisation problem is acceptable (e.g., of the order of few tens of minutes). In this case, by maximizing the objective function I^x, high importance is given to the amelioration of the worst (i.e., most vulnerable) NPPs.

Notice that in the presented model, an additive aggregation of objectives has been applied (see equations (2) and (4)). Based on the developed hierarchical model (see Figure 2), such additive aggregation is easy for the Decision Makers to understand and use. It has the useful interpre-

tation of the relative compositive factors (e.g., the actions, the number of ameliorated NPPs of corresponding category) as partial values (of the performance of the main criteria and the overall vulnerability level) under the simple interpretation of the weights as conversion factors. The partial values in the additive model are converted by the scaling factors (weights) to commensurate values, which are the summed. Proper interpretations and careful handling of the categorization and normalization used in the model are essential elements to be incorporated properly into the questioning procedures of the Decision Makers, for understanding and informed use (Choo & Wedley, 2008).

In this context, three different optimization approaches have been undertaken: (i) one single classification model is built to evaluate and minimize system vulnerability, (ii) an ensemble of compatible classification models, generated by the bootstrap method, is employed to perform a "robust" optimization, by considering the "worst-case" scenario; (iii) finally, a distribution of classification models, still obtained by bootstrap, is considered to address vulnerability reduction in a "probabilistic" fashion.

Notice that given the linearity of equations (2)-(4) and and the discrete classification and selection of actions, the problem is a mixed-integer linear program.

3.1 *Simple Optimization*

As presented in Section 2, and in more details in (Wang, Mousseau, Pedroni, & Zio 2014), we can construct a classification model as $M^*(\cdot|\omega^*, b^*)$(with ω^* the weights and b^* the lower profiles) compatible with all the pre-assigned alternatives in the training set D_{TR} through a disaggregation process. We name this model the "optimum" classification model. The optimization-based inverse classification process aims at finding an optimal set of actions for each of the

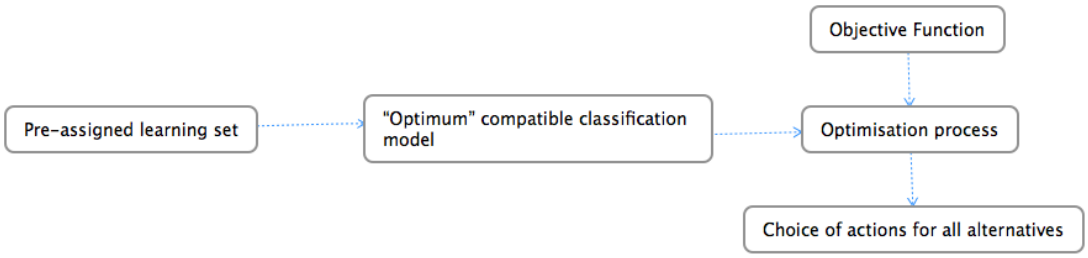

Figure 3: Representation of Simple Optimization

NPPs for which the objective function I^x is maximized: this will improve the performance of the group of NPPs, while giving priority to the worst ones. In more detail, the problem can be formulated as follows:

$$Find\ act'_p = arg\ Max_{\{act'_p, p=1,2,...,N\}}(I^x(act'_p, M^*)), \tag{6}$$

$$s.t.\ \sum_p Cost(x_p, act'_p) \leq B_g, \tag{7}$$

$$\{x|x_p, p \in \{1, 2, ..., N\}\} \tag{8}$$

Under the constraint of budget limitation, we find the combination of protective actions that maximize the value of the objective function I^x, presented above.

3.2 *Robust Optimization*

The optimization approach introduced above provides a choice of protective actions for the NPPs using (only) the "*optimum*" classification model $M^*(\cdot|\omega^*, b^*)$. However, for the training set of pre-assigned alternatives there are a number of compatible classification models. To account for this model uncertainty, we aim at finding the combination of protective actions (for each of the NPPs) that can ameliorate the NPPs to a satisfactorily low level of vulnerability, considering all compatible classification models. In other words, the combination of actions that we obtain should be "*robust*" to the (model) uncertainty arising from the fact that the empirical classification model is trained with a *finite* set of data and, thus, multiple models are

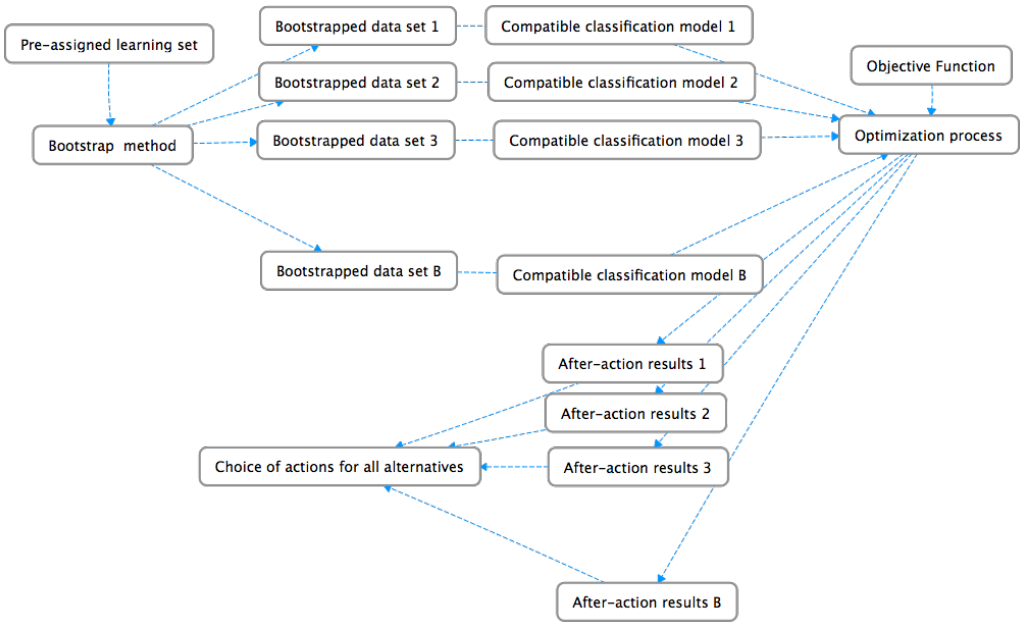

Figure 4: Representation of Robust Optimization

compatible.

To this aim, the bootstrap method (Efron & Thibshirani, 1993) is applied to create an ensemble of classification models constructed on different data sets bootstrapped from the original one (Zio. 2006). The basic idea is to generate different training datasets by random sampling with replacement from the original one (Efron & Thibshirani, 1993): such different training sets are used to build different individual classification models of the ensemble. In this way, the individual classifiers of the ensemble possibly perform well in different regions of the training space. In more detail, the main steps of the bootstrap algorithm are as follows (Figure 4):

a. Generate a bootstrap data set $D_{TR,q} = \{(x_p, \Gamma_p^t) : p = 1, 2, ..., N\}$, by performing random sampling with replacement from the original data set $D_{TR} = \{(x_p, \Gamma_p^t) : p = 1, 2, ..., N\}$ of N input/output patterns. The data set $D_{TR,q}$ is thus constituted by the same number N of input/output patterns drawn among those in D_{TR}, although due to the sampling with replacement some of the patterns in D_{TR} will appear more than once in $D_{TR,q}$, whereas

some others will not appear at all.

b. Build a classification model $\{M_q(\cdot|\omega_q, b_q) : q = 1, 2, ..., B\}$, on the basis of the bootstrap data set $D_{TR,q} = \{(x_p, \Gamma_p^t) : p = 1, 2, ..., N\}$.

Given the bootstrapped ensemble, the mathematical formulation of the robust optimization is as follows:

$$Find\ act'_p = arg\ Max_{\{act'_p, p=1,2,...,N\}}\ Min_q(I^x(act'_p, M_q)), \tag{9}$$

$$s.t.\ \sum_p Cost(x_p, act'_p) \leq B_g, \tag{10}$$

$$\{x|x_p, p \in \{1, 2, ..., N\}\}, \tag{11}$$

$$\{M|M_q \in M, q \in \{1, 2, ..., B\}\}. \tag{12}$$

A large number $B(= 100)$ of compatible classification models $\{M|M_q \in M, q \in \{1, 2, ..., B\}\}$ are typically generated by bootstrap. Correspondingly, the minimum value $Min_M(I^x(act'_p, M_q))$ of objective function $I^x(act'_p, M_q)$ over the B compatible models in correspondence of each set of actions can be gathered. In particular, a distribution of vulnerability classes can be obtained for each NPP. Then, based on the distribution and applying the majority-voting rule, we assign each NPP to its most likely after-action category. Then, the optimization solver aims at finding the optimal combination of actions that robustly and conservatively maximize the worst value of the objective function $I^x(act'_p, M_q)$.

In more detail, the robust optimization algorithm proceeds as follows:

1. The solver proposes a set of actions for each x_p; each bootstrapped classification model $M_q(\cdot|\omega_q, b_q)$ is used to provide an after-action vulnerability class $\Gamma_p^q, q = 1, 2, ..., B$ to each alternative of interest, i.e., $\Gamma_p^q = M_q(x_p|\omega_q, b_q)$;

2. On the basis of the results obtained at step 1 above, a value for function $I^x(act'_p, M_q)$ is computed for *each* compatible model $M_q(\cdot|\omega_q, b_q), q = 1, 2, ..., B$, to obtain an ensemble of values $I^x(act'_p, M_q)$;

3. The minimum (i.e., worst) value among $I^x(act'_p, M_q), q = 1, 2, ..., B$, is taken as the objective function to maximize; in other words, we aim at identifying the set of actions able to improve the "worst-case scenario" over the possible compatible models;

4. We repeat the steps above for different combinations of actions $act'_p, p = 1, 2, ..., N$ in order to find out the combination of actions for each of the considered NPPs that can ameliorate the worst case situation as much as possible.

3.3 *Probabilistic Optimization*

The main steps (Figure. 5) are the same as those of the Robust Optimization presented in Figure 4, but the objective function is changed. Instead of improving the worst case over all the models, we choose to improve the expected value of the probablity distribution of the function I^x. Thus, in this case, we "ignore" some of the "extreme" classification models generated by bootstrap.

The mathematical formulation of the problem is as follows:

$$Find\ act'_p = arg\ Max_{\{act'_p, p=1,2,...,N\}} \frac{1}{B} \sum_{q=1}^{B} (I^x(act'_p, M_q)), \tag{13}$$

$$s.t. \sum_p Cost(x_p, act'_p) \leq B_g, \tag{14}$$

$$\{x|x_p, p \in \{1, 2, ..., N\}\}, \tag{15}$$

$$\{M|M_q \in M, q \in \{1, 2, ..., B\}\}. \tag{16}$$

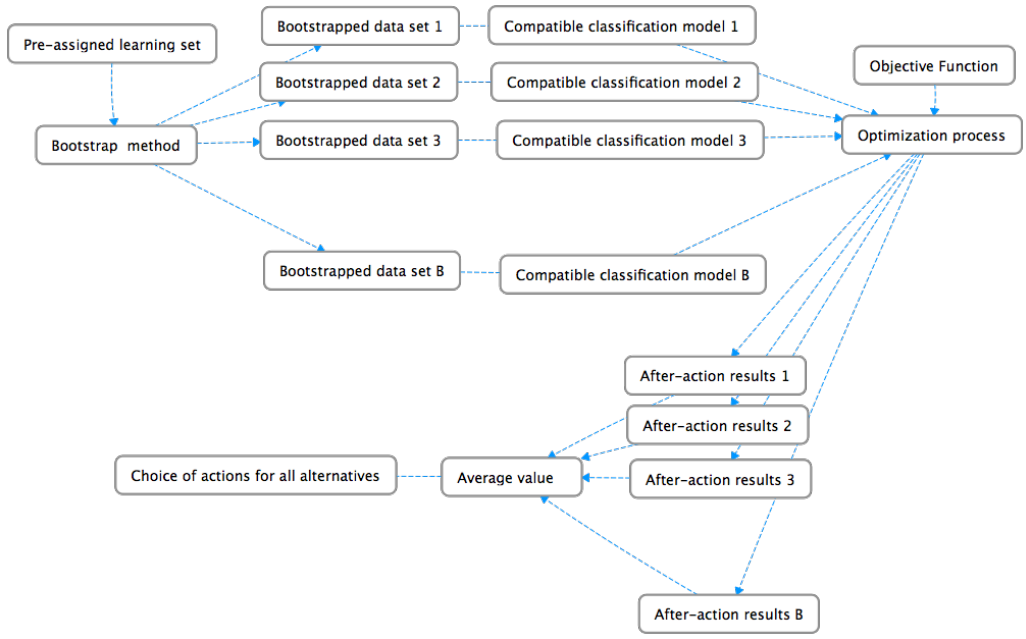

Figure 5: Representation of Probabilistic Optimization

4 APPLICATION

The methods presented in Section 3 are applied on a case study concerning the vulnerability analysis of NPPs (Mousseau & Slowinski, 1998). We identify $n = 6$ main criteria $i = 1, 2, ..., n = 6$ by means of the hierarchical approach presented in (Leroy et al., 2011) (see Section 2): $MCrit_1$ = physical characteristics, $MCrit_2$ = social criticality, $MCrit_3$ = possibility of cascading failures, $MCrit_4$ = recovery means, $MCrit_5$ = human preparedness and $MCrit_6$ = level of protection. Then, these 6 criteria are decomposed into $m = 16$ basic criteria $\{crit^j, j = 1, 2, ..., m = 16\}$ (see Table 1). Finally, $k = 4$ vulnerability categories $A^h, h = 1, 2, 3, 4$ are defined as: A^1 = satisfactory, A^2 = acceptable, A^3 = problematic and A^4 = serious (Section 2).

The training set D_{TR} is constituted by a group of $N = 18$ NPPs with corresponding a priori-known categories Γ_p^t (all the considered NPPs are pre-assigned to A^2, A^3 *or* A^4, since the alternatives originally assigned to best category A^1 are not taken into account), i.e., $D_{TR} = \{(x_p, \Gamma_p^t) : p = 1, 2, ..., N = 18\}$. On one hand, the alternatives in the training set are "de-

Table 1: Basic Criteria

No.	Basic Criteria
$crit^1$	Number of workers
$crit^2$	Nominal power production
$crit^3$	Number of production units
$crit^4$	Percentage of contribution to the welfare
$crit^5$	Size of served cities
$crit^6$	Connection distance
$crit^7$	Number of installed backup components
$crit^8$	Duration of backup component
$crit^9$	Duration of repair and recovery actions
$crit^{10}$	External emergency measures
$crit^{11}$	Training
$crit^{12}$	Safety management
$crit^{13}$	Physical size of the system
$crit^{14}$	Number of accesses
$crit^{15}$	Entrance control
$crit^{16}$	Surveillance

signed" in a way to be realistic; on the other hand, some criteria of some alternatives are defined to represent extreme situations (e.g., a very large number of workers, taken from existing real NPPs). The training set of plants with the corresponding values of the basic criteria is summarized in Table 2. Taking as reference the 16 basic criteria, 13 actions "directly" impacting on them are defined (Table 3): for example, for the criterion *Number of workers*, the direct action is *Reduce number of workers*. On the other hand, there are certain basic criteria that cannot have a corresponding action: for example, for criterion *connection distance*; it is not possible to physically reduce the distance between sites. Additionally, in our case study, the actions have 3 different influence/impact levels; for example, with reference to action *Reduce number of workers*, the 3 levels imply a reduction of the number of workers of the chosen site by 1) 20%, 2) 25% or 3) 30%. This adds degrees of freedom to the choice of actions. Considering the costs associated to the actions, for the sake of simplicity, we define the cost to be "1" for level 1, "2" for level 2 and "3" for level 3, in relative units, for all actions and NPPs. In what follows, the three optimization-based approaches of Section 3 are applied to obtain the "best" combination

Table 2: Training set with $N = 18$ assigned alternatives

Alternatives, x_p	Vulnerability Class, Γ_p^t
x_1={1600, 3600, 4, 90, 350000, 2, 4, 1, 110, 70, 3, 0, 500, 2, 4, 5}	A^3
x_2={800, 2600, 3, 60, 300000, 10, 2, 3, 90, 25, 7, 3, 300, 2, 4, 4}	A^2
x_3={900, 2600, 3, 70, 250000, 30, 3, 2, 130, 30, 7, 3, 400, 1, 2, 3}	A^3
x_4={1000, 3000, 4, 70, 500000, 12, 2, 1, 60, 20, 3, 3, 400, 1, 3, 4}	A^3
x_5={1400, 3600, 4, 80, 400000, 50, 1, 2, 140, 30, 7, 3, 500, 1, 4, 4}	A^2
x_6={800, 1800, 2, 50, 18000, 10, 1, 2, 85, 10, 7, 7, 250, 2, 3, 3}	A^2
x_7={1600, 5200, 6, 100, 800000, 10, 3, 0, 70, 35, 0, 0, 500, 2, 3, 4}	A^4
x_8={2000, 5400, 6, 100, 1200000, 9, 2, 1, 95, 60, 3, 3, 600, 3, 4, 4}	A^4
x_9={2000, 5400, 6, 100, 1200000, 2, 1, 0, 60, 70, 0, 0, 600, 3, 2, 3}	A^4
x_{10}={830, 2600, 2, 30, 1810994, 70, 2, 3, 135, 15, 7, 3, 280, 1, 5, 5}	A^2
x_{11}={1400, 5200, 4, 100, 100000, 10, 1, 2, 65, 20, 3, 3, 415, 1, 5, 5}	A^3
x_{12}={1000, 1800, 2, 64.2, 556000, 50, 2, 2, 100, 25, 3, 3, 106, 1, 3, 4}	A^3
x_{13}={2200, 5400, 6, 100, 4033197, 35, 2, 3, 130, 55, 15, 7, 152, 1, 5, 4}	A^4
x_{14}={684, 1800, 2, 72, 2538590, 30, 2, 3, 125, 30, 3, 3, 60, 1, 3, 4}	A^3
x_{15}={688, 2600, 2, 100, 2538590, 10, 2, 3, 120, 35, 7, 3, 170, 1, 4, 4}	A^2
x_{16}={2200, 3600, 4, 100, 3258000, 40, 3, 1, 140, 32, 7, 3, 255, 1, 4, 5}	A^3
x_{17}={906, 3000, 2, 100, 1760575, 30, 2, 2, 70, 28, 3, 3, 260, 1, 4, 5}	A^3
x_{18}={1300, 2600, 2, 70, 6272467, 30, 3, 1, 135, 30, 7, 7, 180, 1, 3, 3}	A^2

Table 3: Available protective actions

No.	Direct action effect
act^1	Reduce number of workers
act^2	Decrease nominal power production
act^3	Cut down number of production units
act^4	Decrease percentage of contribution to the welfare
act^5	Add installed backup components
act^6	Add external emergency measures
act^7	Prolong the duration of backup components
act^8	Reduce the duration of repair and recovery actions
act^9	Enhance training
act^{10}	Strengthen safety management
act^{11}	Decrease number of accesses
act^{12}	Enhance entrance control
act^{13}	Strengthen surveillance

of protective actions for each of the NPPs.

4.1 *Simple Optimization*

Two tests are first carried out considering an unlimited or limited budget. The example with unlimited budget aims at showing an ideal case of the inverse classification problem that would lead, in principle, to the best after-action condition. It can be seen that, based on the original dataset of pre-assigned alternatives, there are certain NPPs (i.e., $x_1, x_2\ and\ x_{15}$) that can never be ameliorated to the best category A^1 (Table 4). The identification of the best (i.e., lowest) vulnerability category that one NPP can be assigned to without budget restrictions represents an important base information that provides the decision makers with a global view of the problem goals.

The optimization performed with budget constraints aims at solving the realistic problem of finding out the combination of protective actions for each NPP, that ameliorate the group of NPPs with priority to the most vulnerable ones, managing the "residual" resource to improve the others. With an unlimited budget, most of the NPPs are ameliorated to a lower level of vulnerability. Actually, $x_1, x_2\ and\ x_{15}$ do not change class because of their particular characteristics (e.g., the physical distance between the site and the nearby cities is closer with respect to that of the other plants, and such characteristics cannot be modified by any action). The minimum cost necessary to improve each NPP to the best possible category is $B_{g_{min}} = 78$. It can be seen that, as expected, the largest amount of resources is allocated to the amelioration of plants in class 4 (i.e., x7, x8, x9 and x13). In addition, the amelioration of such plants typically requires the combination of a significant number of different actions (often more than three) with a high influence/impact level (i.e., 2 and 3). In this respect, it is also interesting to note that the actions that are more frequently selected to improve the worst NPPs are those implying quite a radical

Table 4: After-action assignments of the considered NPPs without budget constraint. White cases in the third column indicate unchanged assignment.

	Simple Optimization – No Budget constraint			
Plant	**Original Assignment**	**After-action assignment**	**Actions**	**Resource allocation per plant**
X1	3	3	/	/
X2	2	2	/	/
X3	3	1	Action 3: level 2 Action 13: level 1	3
X4	3	2	Action 7: level 3	3
X5	2	1	Action 1: level 1 Action 2: level 1 Action 3: level 3	5
X6	2	1	Action 13: level 1	1
X7	4	2	Action 2: level 2 Action 3: level 2 Action 8: level 3	7
X8	4	2	Action 2: level 3 Action 3: level 2 Action 4: level 3 Action 5: level 3	11
X9	4	3	Action 2: level 1 Action 3: level 3 Action 4: level 1 Action 5: level 2 Action 9: level 3	10
X10	2	1	Action 2: level 3 Action 4: level 3	6
X11	3	2	Action 3: level 1 Action 5: level 1	2
X12	3	1	Action 2: level 1 Action 4: level 3 Action 8: level 1	5
X13	4	1	Action 1: level 2 Action 2: level 1 Action 3: level 3 Action 6: level 1 Action 8: level 2	9
X14	3	2	Action 2: level 2 Action 4: level 3	5
X15	2	2	/	/
X16	3	2	Action 2: level 1 Action 3: level 1	2
X17	3	2	Action 2: level 3 Action 3: level 1 Action 4: level 3 Action 8: level 1	8
X18	2	1	Action 3: level 1	1

Table 5: After-action assignments of the considered NPPs with budget constraint $B_{g_{min}} = 40$ (simple optimization). White cases in the third column indicate unchanged assignment.

	Simple Optimization - Budget constraint: B_g = 40			
Plant	**Original Assignment**	**Best after-action assignment**	**Actions**	**Resource allocation per plant**
X1	3	3	/	/
X2	2	2	/	/
X3	3	1	Action 1: level 1 Action 3: level 2	3
X4	3	2	Action 7: level 3	3
X5	2	2	/	/
X6	2	1	Action 13: level 1	1
X7	4	2	Action 2: level 2 Action 3: level 2 Action 8: level 3	7
X8	4	3	Action 3: level 1	1
X9	4	3	Action 2: level 1 Action 4: level 1 Action 5: level 2 Action 8: level 3 Action 9: level 3	10
X10	2	2	/	/
X11	3	2	Action 3: level 1 Action 5: level 1	2
X12	3	2	Action 8: level 1	1
X13	4	2	Action 2: level 2 Action 3: level 2	4
X14	3	2	Action 2: level 2 Action 4: level 3	5
X15	2	2	/	/
X16	3	2	Action 3: level 2	2
X17	3	3	/	/
X18	2	1	Action 9: level 1	1

variation in the physical system configuration and/or in the operation and maintenance procedures: Action 2 ("Decrease nominal power production"), 3 ("Cut down number of production units"), 4 ("Decrease percentage contribution to the welfare") and 8 ("Reduce the duration of repair and recovery actions"). This represents an encouraging and positive statement with respect to the "realism" of the model of Section 3 and to the effectiveness of the optimization framework proposed.

Fixing a limited budget to $B_{g_{min}} = 40$, the optimization of the actions leads to the ameliorations reported in Table 5. Obviously, x_1, x_2 *and* x_{15} still do not change class as in the case with unlimited budget. Moreover, since the budget is lower than that necessary to ameliorate all NPPs to their best category ($B_{g_{min}} = 78$), there are other NPPs (x_5, x_{10} *and* x_{17}) whose vulnerability category is not changed. On the contrary, all NPPs originally assigned to the "worst" category A^4 improve after action(s); then, the rest of the budget is distributed to ameliorate the other

NPPs as much as possible. For example, x_8 and x_{12} are improved by one category, whereas they can be improved by two categories in the case of unlimited budget (Table 4). In relation to the selected actions, similar considerations can be done with respect to the results obtained in Table 4. As before, the largest amount of resources is allocated to the amelioration of plants in class 4 by means of Actions 2, 3, 4 and 8. Also, it can be seen that Action 1 "Reduce the number of workers" (typically, at level 1) and Action 13 "Strengthen surveillance" (typically, at level 1) frequently intervene in the effective amelioration of NPPs from Categories 2-3 to the best one, i.e., Category 1.

In the next two subsections, we present the results of the other two optimizations approaches considering only the realistic case of limited budget.

4.2 *Robust Optimization*

The results in the case of limited budget, $B_g = 40$, are shown in Table 6 and compared to the original categories (obtained by majority-voting over the B compatible bootstrapped classification models). There are only 4 NPPs that are ameliorated: x_{13} is ameliorated from A^4 to A^3; x_2, x_3 and x_{18} are ameliorated from A^3 to A^2. There are changes in the bootstrapped distributions of the categories of the other NPPs, but not consistent enough to change their final assignments by majority-voting. In comparison with the results obtained in the previous subsection, there are less NPPs that are ameliorated. This is reasonable for a "robust" solution, since "extreme" (worst-case) compatible classification models affect the optimization. The resource allocation may seem weird: actually, the algorithm tries to associate actions also to plants that eventually do not improve their category (see x4, x5 and x6 among the others): as a result, only a relatively small amount of the entire budget is fruitfully employed to ameliorate the situation (16 over Bg = 40). As said above, this is because the actions are actually used to modify the

Table 6: After-action assignments of the considered NPPs with budget constraint (robust optimization). White cases in the third column indicate unchanged assignment. MV = majority-voting

	Robust Optimization - Budget constraint: $B_g = 40$			
Plant	**Original Assignment by MV**	**After-action assignment by MV**	**Actions**	**Resource allocation per plant**
X1	3	3	/	/
X2	3	2	Action 1: level 1 Action 3: level 1 Action 5: level 2 Action 6: level 1	5
X3	3	2	Action 1: level 1 Action 3: level 2	3
X4	3	3	Action 2: level 1 Action 3: level 3	4
X5	3	3	Action 13: level 1	1
X6	2	2	Action 1: level 1 Action 3: level 1 Action 6: level 3 Action 9: level 3 Action 11: level 1 Action 12: level 1	10
X7	4	4	Action 3: level 1	1
X8	4	4	/	/
X9	4	4	Action 3: level 1	1
X10	2	2	Action 2: level 3 Action 4: level 3	6
X11	3	3	/	/
X12	3	3	/	/
X13	4	3	Action 2: level 1 Action 3: level 1	2
X14	3	3	Action 4: level 1	1
X15	4	4	/	/
X16	3	3	/	/
X17	4	4	/	/
X18	3	2	Action 2: level 2 Action 3: level 1 Action 8: level 1 Action 13: level 2	6

bootstrapped distributions of the categories of the NPPs (in the attempt to reduce the worst-case assignments), but such changes are not consistent enough to modify their final assignment by majority voting (MV). For example, taking as reference plant x5, the initial situation in terms of assignments provided by the distribution of B = 100 bootstrapped models is the following: Category 1 = 0; Category 2 = 13; Category 3 = 87 and Category 4 = 0. According to MV, this results in final assignment to Category 3. After the robust optimization procedure, the situation of x5 is the following: Category 1 = 0; Category 2 = 36; Category 3 = 64 and Category 4 = 0, still leading to a final MV assignment to Category 3.

Finally, as before Actions 2, 3, 4 and 8 are the most effective in reducing the vulnerability, in particular of the worst plants (i.e., those initially belonging to Category 4); Actions 1 and 13 are also shown to have a relevant, selective role in improving the condition of NPPs in Categories 2-3.

4.3 *Probabilistic Optimization*

The probabilistic case is a variation of the Robust Case of Section 4.2. Instead of maximizing $Min_M(I^x(act'_p, M_q))$(i.e., the worst after-action objective function value), we choose to maximize the expected value of the bootstrapped probability distribution of the weighted objective function $I^x(act'_p, M_q)$.

The results are shown in Table 7, in comparison with the original majority-voting category of each NPP. There are 8 NPPs that are ameliorated: $x_8, x_{13}, x_{15}\ and\ x_{17}$ are changed from category A^4 to A^3; $x_2, x_3\ and\ x_{12}$ are changed from A^3 to A^2. In comparison with the results of Section 4.1, there are less NPPs that are ameliorated; in addition, not all the NPPs that were originally assigned to the worst category (A^4) are improved. On the other hand, with respect to

Table 7: After-action assignments of the considered NPPs with budget constraint (probabilistic optimization). White cases in the third column indicate unchanged assignment. MV = majority-voting

	Probabilistic Optimization - Budget constraint: B_g = 40			
Plant	**Original Assignment by MV**	**After-action assignment by MV**	**Actions**	**Resource allocation per plant**
X1	3	3	/	/
X2	3	2	Action 7: level 1	1
X3	3	2	Action 1: level 1 Action 2: level 1 Action 3: level 2 Action 4: level 1 Action 13: level 1	6
X4	3	3	Action 3: level 1	1
X5	3	2	Action 2: level 2 Action 4: level 2 Action 8: level 2 Action 9: level 1	7
X6	2	2	Action 2: level 1 Action 3: level 1 Action 12: level 3	5
X7	4	4	Action 3: level 1	1
X8	4	3	Action 3: level 1	1
X9	4	4	/	/
X10	2	2	/	/
X11	3	3	/	/
X12	3	2	Action 2: level 3 Action 4: level 3	6
X13	4	3	Action 3: level 3	3
X14	3	3	/	/
X15	4	3	Action 4: level 1	1
X16	3	3	Action 2: level 2	2
X17	4	3	Action 2: level 2 Action 3: level 1 Action 8: level 1 Action 12: level 2	6
X18	3	3	/	/

the results of the robust optimization (which also considers an ensemble of different compatible models), the group of NPPs is globally improved. The results of the probabilistic case are more satisfactory since most of the NPPs that were assigned to the worst category (A^4) are improved; then, the rest of the resources is used to ameliorate those plants that were assigned to the second worst category (A^3). The results of the resource and action allocation show a more rational use of the budget with respect to the robust case: actually, very few actions are associated to NPPs that do not eventually improve their category. Instead, the largest part of the budget (31 out of Bg = 40) is fruitfully employed to ameliorate the overall ("mean") situation. For the sake of clarity and for illustration purposes, let us consider again the case of plant x5. In the robust optimization, Action 13 (level 1) is allocated, still without any improvement in the plant category according to majority voting (see above). After the application of the probabilistic optimization framework, instead, the situation of x5 is the following: Category 1 = 21; Category 2 = 45; Category 3 = 34 and Category 4 = 0, leading to a final MV assignment to Category 2.

5 DISCUSSION AND ANALYSIS OF THE RESULTS

Based on the representation of the intentional hazards and ameliorative actions (Section 2 and 3), and considering a limited budget as constraint, we have managed to find one set of protective actions for $each$ of the alternatives considered through an optimization-based framework, such that the $overall$ vulnerability level of the $group$ of safety-critical systems (in our case, the Nuclear Power Plants - NPPs) under consideration is $minimized$. The group of combinations of actions for different alternatives is obtained through three different methods: "Simple Optimization" with one single (optimal) classification model; "Robust Optimization" with an ensemble of classification models constructed on different data sets, bootstrapped from the original one; "Probabilistic Optimization" still based on an ensemble of models, but aiming at minimizing

the "expected" vulnerability of a group of plants. These three optimizations considered provide conceptually and practically different solutions to the choice of protective actions for the NPPs. The simple optimization provides a quite specific and limited indication of the amelioration capability of a set of actions with reference to a single classification model with given characteristics. In this case, the classification model is fixed (generated through a disaggregation process based on the number of real-world classification examples available): the number n of main criteria, the number m of basic criteria and the number M of categories (given by the analysts according to the characteristics of the systems at hand). On this basis, the space of all possible combinations of actions for each of the NPPs of the group (and consequently the space of all possible objective functions with the structure mentioned above, i.e., n criteria and M categories) are exhaustively explored by the optimization solver. The weighted objective function defined fulfills the original purpose of ameliorating the NPPs group overall performance, giving preference to those NPPs that were originally assigned to the worst categories.

The robust optimization is inevitably more conservative, given the uncertainty in the compatible models. By the bootstrap method applied on the training set available, an ensemble of B compatible models is built. By so doing, we explore the space of all the classification models compatible with that particular training set. In this view, the bootstrap serves the purpose of accounting for the uncertainty related to using a specific and finite (training) data set for building a classification model of given structure (i.e., with given numbers n and M of criteria and categories, respectively). In addition, the objective function I^x for the optimization represents the "worst-case scenario" over all models and this injects additional conservatism in the choice of the protective actions for the NPPs.

The probabilistic optimization approach applied to the same set of B compatible models aims at the objective of maximizing the expected value of the weighted function I^x. The overall im-

provement of the NPPs turns out to be satisfying. Comparisons and thorough discussions are presented in the following subsections.

5.1 *Comparison of the assignments of the NPPs after protective actions*

Three different perspectives of optimization have been carried out under a limited budget ($B_g = 40$), where the simple case considers one "optimum" classification model, whereas the robust and probabilistic cases consider $B(= 100)$ compatible models obtained by the bootstrap method. For fair comparison of the after-action assignments, an adaptation of the results of the simple case is needed.

The set of protective actions generated in the simple case is now applied to the group of alternatives for all B compatible models of the robust and probabilistic cases. Then, the assignments are obtained by the majority-voting rule. This shows the effect that a set of action "optimistically" obtained by resorting to one single "optimum" model has on the NPPs, when applied to an ensemble of compatible models in light of the uncertainties. The results are listed in Table 8. First, we compare the data of the first and second columns. In the first column, there are the original assignments for all the NPPs, evaluated by the single "optimum" classification model; in the second column, there are the assignments for the same group of NPPs obtained by majority voting based on the $B(= 100)$ models. It can be seen that there are some differences of assignments for some NPPs ($x_2, x_5, x_{15}, x_{17}\ and\ x_{18}$): with the single "optimum" model, the vulnerability of these NPPs is "underestimated". This shows the importance of adopting the robust and probabilistic approaches.

Then, we compare the results of the following three columns, which represent the three cases employing $B(= 100)$ compatible classification models. For the simple case results, there are some ameliorations in the group, whereas there is one NPP (x_{16}) that is assigned to a worse

Table 8: Resume of after-action assignments of the considered NPPs with budget constraint and $B = 100$. White cases of the third to the sixth comlumns indicate unchanged assignments.

Budget constraint: Bg=40	original assignments by single "optimum" model	original assignment by MV	simple optimization assignment by MV	Robust optimization assignment by MV	Probabilistic optimization assignment by MV
x1	3	3	3	3	3
x2	2	3	3	2	2
x3	3	3	2	2	2
x4	3	3	3	3	3
x5	2	3	3	3	2
x6	2	2	2	2	2
x7	4	4	4	4	4
x8	4	4	4	4	3
x9	4	4	4	4	4
x10	2	2	2	2	2
x11	3	3	3	3	3
x12	3	3	3	3	2
x13	4	4	3	3	3
x14	3	3	3	3	3
x15	2	4	4	4	3
x16	3	3	4	3	3
x17	3	4	4	4	3
x18	2	3	3	2	3

category than before. This is explained as follows. In the procedure of majority voting, the number of models that originally assign x_{16} to A^3 is slightly lower than that to A^4. With the actions obtained by the simple optimization, some models that originally assigned x_{16} to A^3, now evaluate it in category A^2; at the same time, the number of models that assign it to A^4 does not change, becoming the majority. This further calls for the adoption of robust and probabilistic approaches. Indeed, the robust case gives a better result than the simple one. No NPPs are assigned to a worse category as before (x_{16}). More NPPs are improved (x_2 *and* x_{18}) but there is still only one NPP that is ameliorated from the worst category (A^4).

Finally, the probabilistic case shows a more promising way to choose the set of protective actions. Actually, 8 out of 18 NPPs are ameliorated. Among these, 4 were originally assigned to A^4, whereas the other 4 were originally assigned to A^3. This matches well our "expected use" of the limited resources ($B_g = 40$), implicitly defined by the weighted objective function I^x.

Table 9: Qualitative highlights and lessons learned from the analyses.

	Type of optimization		
	Simple	**Robust**	**Probabilistic**
Classification Models	- single classification model - uncertainty not accounted for in model construction	- multiple classification models (bootstrap) - uncertainty accounted for in model construction	- multiple classification models (bootstrap) - uncertainty accounted for in model construction
Objective function and optimization criteria	- Weighed objective function - Amelioration of the NPP fleet overall performance - Preference given to NPPs originally assigned to worst categories (by means of proper weights)	- Maximization of the minimum value of the weighed objective function(s) over the different bootstrapped models - Amelioration of the "worst-case" scenarios	- Maximization of the mean value of the bootstrapped probability distribution of the weighed objective function(s) - Amelioration of the NPP "expected" performance over the different classification models
After-action assignments	- "Optimistic" assignments and underestimation of the NPP fleet vulnerability - Some plants assigned to categories that are worse than the original one	- Significant improvement with respect to simple optimization: no NPPs are assigned to worse categories	- More promising way of choosing the set of protective actions - Large part of the NPPs are ameliorated (most of them originally included in the worst categories)

A summary of the qualitative highlights and lessons learned from the analyses is reported in Table 9.

5.2 *Sensitivity analysis on the weights adopted in the objective function of the probabilistic optimization*

As presented in the previous sections, the set of protective actions generated from the probabilistic case shows a more satisfactory result than the others in the present application. A sensitivity of the performance of the probabilistic approach to the weights included in the objective function is, thus, carried out for completeness.

We consider the whole group of NPPs and the $B = 100$ compatible classification models. The weights used to represent the importance of changes from A^4 to A^3, A^3 to A^2 and A^2 to A^1 are $100, 50, 25$. Now, we change to a uniform set of weights $(100, 100, 100)$ and count the cumulative number of changes from each category to its adjacent improved category for all NPPs and all models. The changes are compared for each NPP between its original and the corresponding after-action category evaluated independently by each compatible model. A change

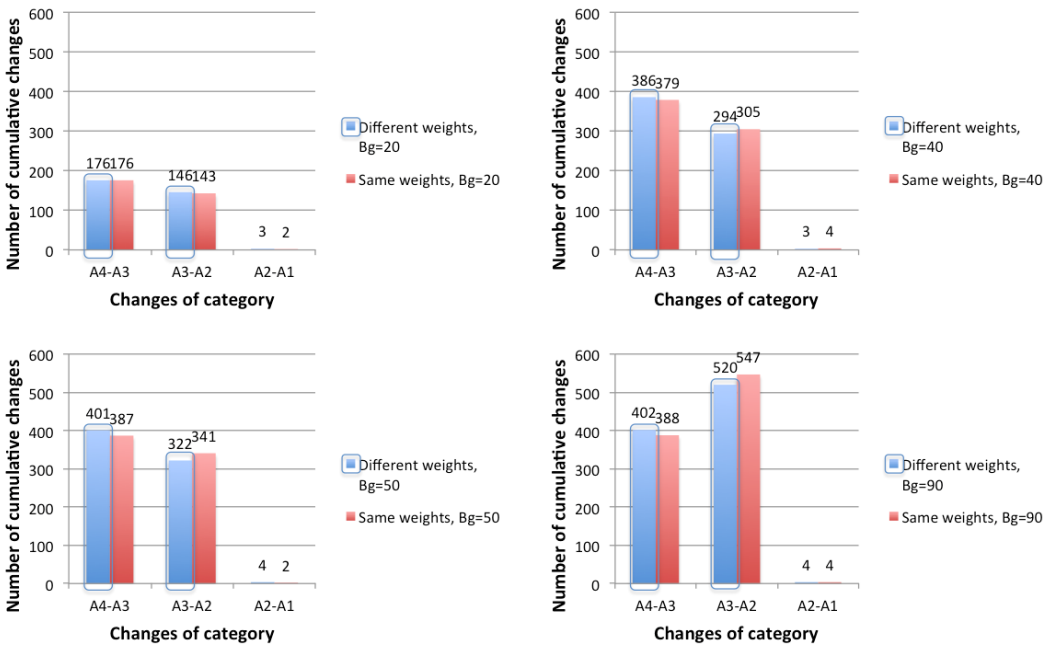

Figure 6: Distribution of number of cumulative changes of category for x in considering $B = 100$ classification models, two sets of weights of objective function and 4 different budgets

of categories more than one for a NPP is considered as the combination of single changes from different "original" assigned categories (e.g., for one NPP that is ameliorated from A^4 to A^3, it is counted as one change from A^4 to A^3, one from A^3 to A^2 and one from A^2 to A^1).

We discover that, reasonably, with a very small amount of budget (e.g., $B_g < 10$), the algorithm cannot find any action, since resource is not enough to make any amelioration. With a very large budget (e.g., $B_g > 702 = total\ cost\ of\ all\ highest\ level\ actions\ applied\ on\ all\ NPPs$), since the resource is adequate, there is no need for the weights to steer its allocation: for the different weights set, the final ameliorations are the same. Focusing on realistic cases of limited and relatively small budgets, we consider 4 different budgets, $B_g = 20, 40, 50\ and\ 90$ and run two optimizations with the weights of the objective function set at $(100, 50, 25)$ and $(100, 100, 100)$, respectively. The results are shown in Figure 6. It is obvious that, the bigger the budget, the bigger the number of cumulative changes after actions. For budgets $B_g = 40, 50\ and\ 90$, the number of ameliorated NPPs originally assigned to the worst category (A^4) is bigger in the

case of the objective function with weights $(100, 50, 25)$ than in the case of uniform weights $(100, 100, 100)$. For the budget $B_g = 20$, the number of ameliorated NPPs originally assigned to the worst category (A^4) is the same with both weights sets. But with the set $(100, 50, 25)$, a larger number of NPPs originally assigned to the second worst category (A^3) is ameliorated than with uniform weights $(100, 100, 100)$.

The results of this sensitivity study show and confirm that the use of priority weights $(100, 50, 25)$ in the objective function makes the probabilistic optimization framework able to steer the allocation of the linked budget on protective actions that can ameliorate the performance of the NPPs, with preferential consideration to those originally assigned to worse categories.

6 CONCLUSIONS

We have addressed the issue of selecting a set of protective actions for minimizing the vulnerability of safety-critical systems (in the case study, nuclear power plants), within an optimization framework based on an empirical classification model. In particular, an MR-Sort model trained by means of a small-sized set of data representing a priori-known classification examples has been used.

Three optimization approaches have been developed and investigated: (i) one single classification model is built to evaluate and minimize system vulnerability; (ii) an ensemble of compatible classification models, generated by the bootstrap method, is employed to perform a "robust" optimization, taking as reference the "worst-case" scenario over the group of models; (iii) a distribution of classification models, still obtained by bootstrap, is considered to address vulnerability reduction in a "probabilistic" fashion (i.e., by minimizing the "expected" vulnerability of a fleet of systems). To the best of the authors' knowledge, it is the first time that an inverse classification problem is formulated and considered for the optimization of the choice of

protective actions to reduce the vulnerability of a group of safety-critical systems (e.g., Nuclear Power Plants), taking into account the uncertainty associated to the classification models.

From the results obtained, it can be concluded that a combination of protective actions can be obtained using only a single classification model, but this set of actions is not robust with respect to the uncertainty of the classification model. The robust optimization may, then, be used for a more conservative set of actions, coping with model uncertainty. Eventually, the probabilistic optimization seems most practical for real cases, for the following reasons: (i) as for the robust case, it handles the uncertainty coming from the finite data set available and the compatible models; (ii) by maximizing the expected value of the bootstrapped probability distribution of the objective function, some " extreme" compatible models of the bootstrapped ensembles are "neglected", which is reasonable and more realistic.

The proposed methodological framework provides a powerful tool for systematically and pragmatically evaluating the safety and vulnerability as well as other characteristics of critical systems.

For future research, the following issues will be considered. Since one set of weights is usually an insufficient basis for giving priorities, the sensitivity of investment priorities to the weights of criteria can be tackle: for example, in (Martinez, Lambert, & Karvetski, 2011; Karvetski & Lambert, 2012; Thekdi & Lambert, 2014) a "scenario" is introduced that reflects a set of weights for each stakeholder, such as emphasis on particular aspects of safety in the aftermath of a major nuclear incident.

As presented in (Hamilton, Lambert, Keisler, Linkov, & Holcomb, 2013), an influential set of weights can suggest R&D priorities in protection of energy systems.

Moreover, a set of weights can also be brought by other stakeholders, such as owners, operators and users etc.: each set of weights presumably leads to variation in the preferred safety invest-

ments (Rogerson & Lambert, 2012).

In addition, the proposed methodology could be easily integrated in a business process of safety management, where we seek for the best compromise among risks, costs and benefits in allocating investments in safety-critical systems in the presence of uncertainties, as suggested in (Thekdi & Lambert, 2014). In this line of thoughts, our approach could be used, e.g., by agencies to preserve and sustain the compliance of risk, safety and security programs with organizational and administrative guidelines, by identifying the relative priorities among policy initiatives and actions (e.g., multiple and possibly competing risk assessment, management and communication procedures and missions) (Teng et al., 2012 and 2013). Since the method basically relies on a general and flexible classification model, it might be adopted in several different contexts, e.g., health, environment, risk communication, cost-benefit analysis, etc.

Finally, although in this work significant efforts have been made to assign numerical values to the costs and impacts of actions (in order to represent the problem as realistically as possible), in a non-fictitious situation the task is expected to be much more complex. Actually, the relations between the actions and the criteria taking into account the dependencies of different attributes and systems are always difficult to identify: in such cases, resorting to the judgment of real experts/users/decision makers and possibly to real historical data will be mandatory.

ACKNOWLEDGMENTS

The authors wish to thank the reviewers and the Area Editor, Prof. Seth Guikema, for their constructive comments that helped improve significantly the quality of the article.

7 REFERENCES

Aggarwal, C. C., C. Chen, & J. W. Han (2006). On the inverse classification problem and its applications. Data Engineering, 2006. ICDE '06. Proceedings of the 22nd International Conference. IEEE, 111–113.

Aggarwal, C. C., C. Chen, & J. W. Han (2010). The inverse classification problem. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY, 25, 458–468.

Ahuja, R. K., Orlin, J. B. (2001). Inverse optimization. Operations Research, 49(5), 771–783.

Ahuja, R. K., Orlin, J. B. (2002). Combinatorial algorithms for inverse network flow problems. Networks, 40(4), 181–187.

Almeida-Dias, J., Figueira, J. R., Roy, B. (2012). A multiple criteria sorting method where each category is characterized by several reference actions. European Journal of Operational Research, 217, 567–579.

Alsabti, K., S. Rank, & V. Singh (1998). Clouds: A decision tree classifier for large datasets. Electrical Engineering and Computer Science, 41. https://surface.syr.edu/eecs/41.

Apostolakis, G. E. & D. Lemon (2005). A screening methodology for the identification and ranking of infrastructure vulnerabilities due to terrorism. Risk Analysis, 25(2): 361-76.

Aven, T. (2003). Foundations of Risk Analysis. Chichester, England: John Wiley & Sons, Ltd.

Aven, T. (2010). Some reflections on uncertainty analysis and management. Reliability Engineering & System Safety, 95, 195–201.

Aven, T. & R. Flage (2009). Use of decision criteria based on expected values to support decision-making in a production assurance and safety setting. Reliability Engineering & System Safety, 94, 1491–1498.

Ayyub, B.M. (2001), Elicitation of Expert Opinions for Uncertainty and Risks. Boca Raton, Florida: CRC Press, 328 Pages, ISBN 9780849310874.

Belton, V. & T. Stewart (2002). Multiple Criteria Decision Analysis: An Integrated Approach. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bodley, C. E. & P. E. Utgoff (1995). Multivariate decision trees. Machine Learning, 20, 63–94.

Breiman, L., J. Friedman, C. J. Stone & R.A. Olshen (1984). Classification and Regression Trees. Boca Raton, Florida: Chapman and Hall/CRC.

Breslow, L. (1997). Simplifying decision trees. Knowledge Engineering Review, 12, 1–40.

Cailloux, O. & V. Mousseau (2011). Parameterize a territorial risk evaluation scale using multiple experts knowledge through risk assessment examples. In: C. Berenguer, A. Grall, C. Guedes Soares (Eds.), Advances in Safety, Reliability and Risk Management, ESREL 2011, Troyes, France, 18-22 September 2011 (pp. 2331–2339). London, UK: Taylor and Francis Group.

Coello Coello, C., Dhaenens, C., Jourdan, L. (2010). Multi-Objective Combinatorial Optimization: Problematic and Context, Chapter 1. Advances in Multi-Objective Nature Inspired Computing. Berlin Heidelberg: Springer.

Choo, E. & W. Wedley (2008). Comparing fundamentals of additive and multiplicative aggregation in ratio scale multi-criteria decision making. The Open Operational Research Journal, 2: 1-7.

Claudio, T., M. Nelson, P. Marta, & L. L. P (2012). Multiobjective combinatorial optimization problems with a cost and several bottleneck objective functions: An algorithm with reoptimization. Computer & Operations Research, 39, 1969–1976.

Corrente, S., M. Doumpos, S. Greco, R. Słowiński, C. Zopounidis (2017). Multiple criteria hierarchy process for sorting problems based on ordinal regression with additive value functions. Annals of Operations Research, 251, 117–139.

Corrente, S., S. Greco, R. Słowiński (2016). Multiple Criteria Hierarchy Process for ELECTRE Tri methods. European Journal of Operational Research, 252, 191–203.

Courtois, P. J. (1985). On time and space decomposition of complex structures. Communications of the Acm, 28, 590–603.

Doumpos, M. & C. Zopounidis (2002). Multicriteria Decision Aid Classification Methods. Dordrecht, The Netherlands: Kluwer Academic Publishers, ISBN 1- 4020-0805-8.

Duda, R., P. Hart, & D. Stork (2001). Pattern Classification (Second ed.). New York, NY: John Wiley and Sons.

Efron, B. & R. Thibshirani (1993). An introduction to the bootstrap. New York, NY: Chapman & Hall.

Friedman, J. H. (1977). A recursive partitioning decision rule for non-parametric classifiers. IEEE Transactions on Computers, C26(4), 404–408.

Gehrke, J., V. Ganti, R. Ramakrishnan, & W. Y. Loh (1999). Boat: Optimistic decision tree construction. In: SIGMOD '99 Proceedings of the 1999 ACM SIGMOD international conference on Management of data, Philadelphia, Pennsylvania, USA — May 31 - June 03, 1999 (pp. 169-180). New York, NY, USA: ACM.

Haimes, Y. Y. (2012). Modeling complex systems of systems with phantom system models. Systems Engineering, 15, 333–346.

Hamilton, M.C., J.H. Lambert, J.W. Keisler, I. Linkov, and F.M. Holcomb (2013). Research and development priorities for energy islanding of military and industrial installations. ASCE Journal of Infrastructure Systems, 19(3): 297-305.

Heuberger, C. (2004). Inverse combinatorial optimization: A survey on problems, methods, and results. Journal of Combinatorial Optimization, 8, 329–361.

Hofmann, M., Kjølle, G., Gjerde, O., Hernes, J.G., Kjølle, G.H. & Foosnæs, J.A. (2012). Development of indicators to monitor vulnerability of power lines — Case studies. In: Proceedings of the 22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013),

Stockholm, Sweden, 10-13 June 2013. The Institution of Engineering and Technology (IET).

James, M. (1985). Classification Algorithms. New York, NY, USA: Wiley-Interscience.

Johansson, J. & H. Hassel (2010). An approach for modelling interdependent infrastructures in the context of vulnerability analysis. Reliability Engineering & System Safety, 95, 1335–1344.

Joshi, N.N. and J.H. Lambert (2011). Diversification of engineering infrastructure investments for emergent and unknown non-systematic risks. Journal of Risk Research, 14(4): 1466-4461.

Karvetski, C.W., and J.H. Lambert (2012). Evaluating deep uncertainties in strategic priority-setting with an application to facility energy investments. Systems Engineering, 15(4): 483-493.

Koonce, A. M., Apostolakis, G.E., Cook, B.K. (2008). Bulk power risk analysis: Ranking infrastructure elements according to their risk significance. Electrical Power & Energy Systems, 30(3): 169-183.

Kroger, W. & E. Zio (2011). Vulnerable Systems. London, UK: Springer-Verlag.

Lambert, J.H. and M.W. Farrington (2007). Cost-benefit functions for the allocation of security sensors for air contaminants. Reliability Engineering and System Safety. 92(7): 930-946.

Larsson, J. E. (1992). Knowledge-based methods for control systems. PhD dissertation, Lund Institute of Technology, Department of Automatic Control.

Leroy, A., V. Mousseau, & M. Pirlot (2011). Learning the parameters of a multiple criteria sorting method based on a majority rule. In: Brafman, R. (Eds.), Proceedings of the Second International Conference on Algorithmic Decision Theory, Piscataway, NJ, USA, October 26-28, 2011 (pp.219-233). Heidelberg, Germany: Springer.

Li, A. G., x. Zhou, & J. L. Zhang (2012). Performance analysis of quantitative attributes inverse classification problem. JOURNAL OF COMPUTERS, 7, 1067–1072.

Lin, T.C., Kuo, C.C., Hsieh, Y.H., Wang, B.F. (2009). Efficient algorithms for the inverse sorting problem with bound constraints under the l-norm and the Hamming distance. Journal of

Computer and System Sciences, 75(8), 451-464.

Lind, M. (2011a). An introduction to multilevel flow modeling. Nuclear safety and simulation, 2, 22–32.

Lind, M. (2011b). Reasoning about causes and consequences in multilevel flow models. In: C. Berenguer, A. Grall, C. Guedes Soares (Eds.), Advances in Safety, Reliability and Risk Management, ESREL 2011, Troyes, France, 18-22 September 2011 (pp. 2359–2367). London, UK: Taylor and Francis Group.

Mannino, M. V. & M. Koushik (2000). The cost-minimizing inverse classification problem: a genetic algorithm approach. Decision Support Systems, 29, 283–300.

Martinez, L.J., J.H. Lambert, and C. Karvetski (2011). Scenario-informed multiple criteria analysis for prioritizing investments in electricity capacity expansion. Reliability Engineering and System Safety, 96, 883-891.

Morgan, M., H. K. Florig, M. L. DeKay, & P. Fischbeck (2000). Categorizing risks for risk ranking. Risk Analysis, 20(1), 49-58.

Mousseau, V., Özpeynirci, O., Özpeynirci, S. (2018). Inverse multiple criteria sorting problem. Annals of Operations Research, 267(1), 379-412.

Mousseau, V. & R. Slowinski (1998). Inferring an electre tri model from assignment examples. Journal of Global Optimization, 12, 157–174.

NWRA, N. W. R. A. (2002). Risk assessment methods for water infrastructure systems.

Patterson, S. A. & G. E. Apostolakis (2007). Identification of critical locations across multiple infrastructures for terrorist actions. Reliability Engineering & System Safety, 92(9), 1183-1203.

Pendharkar, P. C. (2002). A potential use of data envelopment analysis for the inverse classification problem. Omega, 30(3), 243–248.

Piwowar, J., Chatelet, E., Laclémence, P. (2009). An efficient process to reduce infrastructure

vulnerabilities facing malevolence. Reliability Engineering & System Safety, 94(11), 1869-1877.

Quinlan, J.R. (1993). Chapter 4.5. Programs for Machine Learning. California, USA: Morgan Kaufmann Publishers.

Rocco, C. & E. Zio (2005). Bootstrap-based techniques for computing confidence intervals in monte carlo system reliability evaluation. In: Proceedings of the Annual Reliability and Maintainability Symposium, Alexandria, VA, USA, USA, 24-27 Jan. 2005 (pp. 303–307). IEEE.

Rogerson, E.C. and J.H. Lambert (2012). Prioritizing risk via several expert perspectives with application to airport runway safety. Reliability Engineering and System Safety, 103, 22-34.

Roy, B. (1991). The outranking approach and the foundations of electre methods. Theory and Decision, 31, 49–73.

Ruan, D. (2000). Logic-based hierarchies for modeling behavior of complex dynamic systems with applications. Volume 1 of Fuzzy systems and soft computing in nuclear engineering, Chapter 17, pp. 364–395. Berlin Heidelberg, Germany: Springer-Verlag.

Salo, A. & R. Hamalainen (1997). On the measurement of preferences in the analytic hierarchy process. Journal of Multi-Criteria Decision Analysis, 6, 309–319.

Salo, A. & R. Hamalainen (2010). Multicriteria decision analysis in group decision processes. In: Kilgour D., Eden C. (eds), Handbook of Group Decision and Negotiation. Advances in Group Decision and Negotiation, vol 4. (pp. 269–283). Dordrecht, The Netherlands: Springer.

Salo, A., J. Keisler, & A. Morton (2011). An Invitation to Portfolio Decision Analysis. In: Salo A., Keisler J., Morton A. (eds), Portfolio Decision Analysis. International Series in Operations Research & Management Science, vol 162 (pp. 3–27). New York, NY: Springer.

Teng, K., S.A. Thekdi, and J.H. Lambert (2012). Identification and evaluation of priorities in the business process of a risk or safety organization. Reliability Engineering and System Safety,

99, 74–86.

Teng, K., S.A. Thekdi, and J.H. Lambert (2013). Risk and safety program performance evaluation and business process modeling. IEEE Transactions on Systems, Man, and Cybernetics: Part A, 42(6), 1504-1513.

Thekdi, S.A., and J.H. Lambert (2014). Quantification of scenarios and stakeholders influencing priorities for risk mitigation in infrastructure systems. ASCE Journal of Management in Engineering, 30(1), 32-40.

Thorisson, H., J.H. Lambert, J.J. Cardenas, and I. Linkov (2017). Resilience analytics for power grid capacity expansion in a developing region. Wiley journal Risk Analysis, 37(7), 1268-1286.

Wang, T. R., V. Mousseau, N. Pedroni, & E. Zio (2014). Assessing the performance of a classification-based vulnerability analysis model. Risk Analysis, 35(9), 1674-1689.

Wang, T. R., V. Mousseau, N. Pedroni, & E. Zio (2016). Identification of protective actions to reduce the vulnerability of safety-critical systems to malevolent intentional acts: a sensitivity-based decision-making approach. Reliability Engineering & System Safety, 147(C), 9-18.

Wang, T. R., V. Mousseau, & E. Zio (2013). A hierarchical decision making framework for vulnerability analysis. in: R.D.J.M. Steenbergen, P.H.A.J.M. van Gelder, S. Miraglia and A. C.W.M. Ton. Vrouwenvelder (Eds.), Safety, Reliability and Risk Analysis, Beyond the Horizon, Proceedings of the European Safety and RELiability Conference (ESREL) 2013, Amsterdam, The Netherlands, 29 September-2 October 2013 (pp. 1157–1165). London, UK: Taylor and Francis Group.

Zio, E. (2006). A study of the bootstrap method for estimating the accuracy of artificial neural networks in predicting nuclear transient processes. IEEE Transactions on Nuclear Science, 53, 1460–1470.

Zio, E. (2007). An Introduction to the Basics of Reliability and Risk Analysis. Singapore, Sin-

gapore: World Scientific Publishing Co.

Zopounidis, C., Doumpos, M. (2002). Multicriteria classification and sorting methods: A literature review. European Journal of Operational Research, 138(2), 229-246.