64/2012

Raport Badawczy

Research Report

RB/60/2012

Robustness testing of model based multiple criteria decisions: fundamentals and applications

A.P. Wierzbicki,
M. Makowski, J. Granat

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. zw. dr hab. inż. Zbigniew Nahorski

Warszawa 2012

INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK

Andrzej P. Wierzbicki[2], Marek Makowski[1,3], Janusz Granat[2]

Robustness Testing of Model Based Multiple Criteria Decisions: Fundamentals and Applications

[1]Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warszawa, Poland
[2]National Institute of Telecommunications, Szachowa 1, 04-894 Warszawa, Poland
[3]International Institute for Applied Systems Analysis, Schlossplatz 1, A2361 Laxenburg, Austria

Praca zaakceptowana do druku w *International Journal of Information Technology and Decision Making*.

Warszawa 2012

ROBUSTNESS TESTING OF MODEL BASED MULTIPLE CRITERIA DECISIONS: FUNDAMENTALS AND APPLICATIONS

ANDRZEJ P. WIERZBICKI*

MAREK MAKOWSKI†

JANUSZ GRANAT‡

Abstract: Robustness or insensitivity is a desirable property of decisions; however, most texts on robustness and/or sensitivity analysis do not define it precisely. A broad literature in this field concentrates on *robust design* (including robust optimization). This paper focuses on *robustness testing*, that is, checking whether a design has actually resulted in robust properties of the system if some of basic assumptions are changed. We propose a general framework of such robustness testing and show that robustness is a property of the relation between three (classes of) models: a basic model of the decision (design) situation, a second model of possible perturbations of the first model, and a third model of the decision (design) implementation, optionally taking into account some measurements of the impact of perturbations. Typical approaches to robustness or sensitivity analysis assume tacitly that the first two models can be combined, and thus analyze parameters deviations in such combined model. However, the role of the first two models can be asymmetric if optimization of the decision is performed on the first model. We extend this framework, intended originally for single criteria (scalar) optimization to multiple criteria (vector) optimization. The proposed approach is illustrated by diverse examples.

Keywords: Robustness; sensitivity analysis; multiple-criteria analysis.

1991 Mathematics Subject Classification: 49K40, 90C29, 90C31

1. Introduction

Robustness is a desired property of decisions, statistical estimates, engineering designs, managerial plans etc. However, what is meant by a robust decision is usually not well defined. A recent paper on the subject of robust decisions (Ermoliev and Hordijk 2006) summarizes many ways of understanding this term, especially in statistics, and concludes that the concept of robustness is context dependent.

There is a broad literature on what we shall call here *robust design,* that is, designing (including optimization) a decision or conclusion that in some sense is robust: robust decisions (starting with Gupta and Rosenhead, 1968), robust optimization (see, e.g., Ben

* *National Institute of Telecommunications, Szachowa 1, 04-894 Warsaw, Poland*
email: A.Wierzbicki@itl.waw.pl

† *International Institute for Applied Systems Analysis, Schlossplatz 1, A2361 Laxenburg, Austria* and *Systems Research Institute of the Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland*
email: marek@iiasa.ac.at

‡ *National Institute of Telecommunications, Szachowo 1, 04-894 Warsaw, Poland*
email: J.Granat@itl.waw.pl

Tal et al. 2009, Bertsimas and Sim 2004), robust conclusion (Roy 1998), robust method (e.g., Vincke 1999). For example, Nikulin (2006) presents an annotated bibliography of almost 90 papers on the issue of robust optimization in the field of combinatorial optimization and scheduling theory, including linear programming. Most earlier approaches to robust optimization relied on the worst case optimization and, as such, were usually overly conservative; later approaches tried to overcome this conservativeness, analyse various issues of computational complexity, and develop the concept of robust optimization in diverse directions. However, only exceptional papers ask the question whether *robust is truly robust,* that is, how to test the results of diverse ways of defining robust optimization.

In a synthesizing paper by (Vincke 2003) the concept of *robustness analysis* is used, but most types of robust solutions listed there are in fact results of *robust design.* However, Vincke correctly draws attention to the fact that *an important feature of robustness is its subjective dimension.* All these various types of robust design are in a sense subjective: a way of designing what is robust is selected by the designer, perhaps accepted by the decision maker, but a test how to analyse whether robust is truly robust is not indicated (and only exceptionally performed). In a sense, these various types of robust design are not subjected to Popperian *falsification* (see Popper 1976). We would like, however, to specify better how robustness of model-based decisions – whether they are results of robust design or not – can be tested in a more objective, falsifying way. *Such an approach might be called robustness testing and is the objective of this paper.* The distinction between robust design and robustness testing was noted before, e.g. (Dias 2007) writes about *ex ante* and *ex post* robustness analysis; *ex ante* would correspond to what we call here robust design and *ex post* to testing robustness. There are several approaches to *ex post* robustness analysis, mostly related to the concept of *robust conclusion* (Roy 1998), but we feel that a general methodology of robustness testing is needed.

The classical definition of robustness in statistics is given by Peter Huber (1981): *"robustness signifies insensitivity to small deviations from the assumptions".* The need of robustness analysis and testing results from the fact that any analysis and design is partly based on measurement and observations, partly on prior assumptions about an underlying situation. Part of such prior assumptions are tacitly made (see, e.g., Polanyi 1966, Wierzbicki and Nakamori 2006, 2007), and the investigator might not be even aware of them. The examples given by Huber correctly stress the need of very precise definition of the conditions of robustness analysis, but tacitly assume the symmetry of two interlinked models that should be considered in this context, i.e., the basic model and the model of possible perturbations.

This tacit assumption is shared by most sensitivity analysis and robust design. Although Huber correctly defines robustness as insensitivity, sensitivity analysis might have broader goals; for example, if applied to learning or parameter estimation, sensitivity analysis might have the goal of finding regions of model parameters with high sensitivity of outputs to inputs. Some authors even claim that the term robustness can be used only in statistical setting. However, most texts in sensitivity analysis, see e.g. (Saltelli et al. 2001), tacitly assume that it is sufficient to analyze one parametrically imbedded family of models, which is equivalent to an assumption that the basic model and its possible perturbations have a symmetric (or anti-symmetric) role in the analysis.

The concepts of a basic model and its perturbations are related to the formalization of a modeling process as described, e.g., in (Rosen 1991), where a "natural system" is encoded into a "formal system" or a model. However, such formalization raises epistemic doubts: we know, what is a *model*, but what is a *natural system* (with all postmodern and post-postmodern doubts about the relation between knowledge and nature, see e.g. Wierzbicki and Nakamori 2007)? If the natural system is a proxy for reality – which, for epistemic clarity, we shall denote by "reality" – how to represent it in analysis?

Works on robust design actually pursue the concept of *procedural robustness* – see, e.g., (Ogryczak 2010, Naseraldin et al. 2010, Bello et al. 2010) – and assume that it is sufficient to include logically or mathematically the concept of robustness and models of some uncertainty parameters, e.g., bounds on possible parameter deviations, etc., into the basic model together with an appropriate robust optimization procedure, and the resulting decision will be robust. While this is certainly better than not addressing robustness at all, the conclusion actually is that the resulting decision will be robust only if the assumptions about the basic model will not change. But what if some tacitly made assumptions will change? For example, the stationarity of underlying stochastic processes is usually tacitly assumed; what if they become non-stationary, as it happened during the last financial and economic crisis? Thus, we assume here that *robustness cannot be defined in an absolute sense, there are always conditions in which it must be experimentally or computationally tested*. Only a few authors writing on robustness perceive this fact and perform such experiments, see, e.g. (Laumanns et al. 2010); when they do it, they use an approach similar to (though usually less general than) that presented here, since we want to present a general methodology of testing robustness by experiments. Thus *we concentrate here on robustness testing: on computational experiments in robustness analysis based on an application of the decision obtained with the help of a basic model* (even if the model is procedurally robust or a basis of robust design) *to a model of possible "reality"*.

In engineering, it was perceived already in (Wierzbicki 1977) that "reality" must be represented by a model or family (a set of model versions) of parameterized models $M(d)$ of possible "realities", while one of such models or model family, say $M(a^*)$ for $d = a^*$ (model parameters a for *assumptions*, d for doubts) is selected as the basis of design or decision and shall be called the *basic model*. We use here the term "model parameter" in the most generic sense, without specifying its many possible detailed meanings (coefficients, constraint values, indexes of model versions, etc.). Clearly, the basic model $M(a^*)$ is not the reality or even "reality" (as noted by Bernard Roy: "a map is not a territory"), but only the best approximation of reality in some defined sense between the models of the family $M(d)$. This family in turn represents how possible "realities" differ from the basic model; for this reason, we shall call the family M(d) the *perturbation model*. We use here the letter d to represent parameters of the perturbation model in order to stress that they represent *doubts* about the validity of the basic model; if $d = a^*$, then we have no doubts.

Thus, the perturbation model or models *express our knowledge about the exactness of the basic model.* We must stress here that a model of uncertainty, e.g., represented by several model versions, if included into the basic model for the purpose of robust design, e.g., robust optimization, does represent uncertainty, but not limited either knowledge or limitations of modeling technology; the latter could be represented, e.g., by additional model versions not included into the robust design. It might seem paradoxical to test uncertainty with the help of additional uncertainty, but we should try our best to effectively cope with the objective modeling limitations, and can use for this purpose any tools well known from uncertainty modeling (scenarios, model versions, regret functions, etc.); we must only remember that these tools are used here not to support a (robust) design, but to support testing how robust are the results of the model analysis when actually applied for design/decisions.

The perturbation model might have a broader character than the basic model; we only assume that at some parameters value we might obtain the basic model from the perturbation model (for example, we can even obtain a deterministic model from a stochastic one by parameter variations). The perturbation model, however, must satisfy one essential condition: *it must be formulated in a way that permits the simulation of results of any decision,* hence it must be a *process model,* not an *operational research* type of model (see the discussion of model types in Wierzbicki et al. 2000); thus, e.g., a way of dealing with inadmissible decisions (either through their projection on the admissible bounds, or through counting penalties associated to constraint violation, or any other way) must be included in perturbation model definition.

Now, a fundamental question in all sensitivity and robustness analysis, in particular robustness testing, is: *are the roles of the basic model and the perturbation model symmetric (or anti-symmetric) in the process of analysis?* Or, in other words, is it correct to analyze simply the parametric dependence of M(d) in a (smaller or bigger) neighborhood of $d = a^*$? The answer, unfortunately, is not an unqualified "yes, of course", as it is tacitly assumed in most writings on this subject. The answer, as we explain in detail in the next section, is "that depends". In some cases the answer might be positive; in other cases, however, such a conclusion is patently wrong.

2. The Asymmetric Relation of Basic and Perturbation Models

Let us assume now that the process of analyzing and deriving conclusions from a basic model does involve optimization: it results in an optimal (in the sense of a specified performance measure, including robustness measures) decision that is derived from the basic model and then applied to reality – or, in analysis, to possible "realities". Now, let us consider two processes: the process of selecting the basic model and the process of testing model's sensitivity or robustness. Between them, of course, is the process of using the basic model for selecting optimal decisions or for robust design or optimization, on which most of literature concentrates; but we consciously concentrate here on the beginning and on the (often omitted) end stage.

When selecting the basic model we consider a family of models *M(a)* – without loss of generality, we might assume that this family is as broad as *M(d)* – and it is natural to assume that we choose such parameters $a = a^*$ that best approximate the "reality" for a given purpose. If the purpose is optimization, we suppose the "reality" is best approximated by $a = a^*$ and we choose a decision that is optimal for this value of basic model parameters. Then, if we choose (by mistake or lack of knowledge) another $a \neq a^*$ as the basis of selecting the optimal decision, then the corresponding decision will be not optimal when applied to the possible "reality" modeled with $d = a^*$, thus the "actually" achieved performance measure can be only worse. This argument will be essential for further considerations; anyway, the dependence of the performance on the selection of a is extremal.

However, the dependence of the performance level on the perturbations d, on possible "realities" needs not to be extremal: we make a decision and then the reality can turn out for the better or worse for us, we can win or lose in a changing market even if it seemed to us that we selected an "optimal" investment. *Thus the roles of the dependences* M(a) *and* M(d) *are not symmetric in decision processes involving optimization,* and a more involved parametric analysis is needed in such cases.

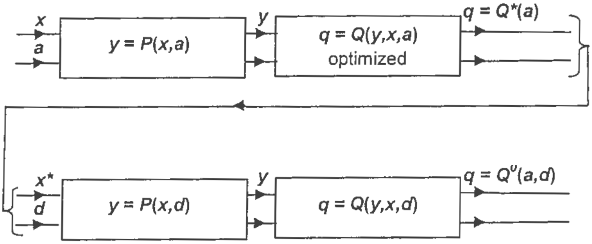

Fig. 1. A block diagram of optimizing a decision and testing its robustness or sensitivity

This fundamental fact was first noted in (Wierzbicki 1977). To illustrate the analysis needed in such a case, we quote here from this book a slightly more detailed form of the model of a decision situation and its sensitivity analysis, see Fig. 1. In the diagram in Fig. 1, a denotes (a vector of) parameters in the basic model, d denotes (a vector of) parameters of the perturbation model, x denotes (a vector of) decisions, y denotes (a vector of) decision outcomes (diverse – including possible criteria or performance measures), q denotes (a scalar in this case – see section 5 for a generalization) the selected performance measure. The models are assumed to have the *process form:* the first block $y = P(x, a)$ specifies outcomes of any decisions x (if the decisions are inadmissible, the way of dealing with inadmissibility must be also specified).

On purpose, we do not specify here the spaces of parameters, decisions, decision outcomes. Parameters might be just scalars, or trajectories (scenarios of future developments), or probability distributions, belonging to mathematically quite abstract spaces. Decisions could be also trajectories (in control engineering) or probability distributions (in mathematical game theory); the same concerns outcomes of decisions. For this reason, whenever we speak about a function mapping, say, decisions into their outcomes, this should be understood as a quite general mapping.

The upper part of the block diagram specifies in more detail the basic model $M(a)$, the lower part the perturbation model $M(d)$. The roles of these models are asymmetric, because we first optimize the performance measure in the basic model, obtain some value of performance measure $Q^*(a)$ and the optimal decision x^* – obviously, both dependent on the selected parameter a. This decision is applied then to the "reality" – represented in analysis by the perturbation model with parameters d. The resulting performance measure $q = Q^0(a,d)$ does not depend only on the difference $a - d$: it has an extremum (minimum, if we minimize the performance measure) with respect to a at $a = d$, but might have no extremum with respect to d. Clearly, if $a = d$, then $Q^0(a,d) = Q^*(a) = Q^*(d)$. If we assume that $Q^*(d) > 0$ for all investigated d, then for minimized performance measures we can use the following *relative sensitivity index,* introduced in (Wierzbicki 1977); later, similar expressions were called *regret functions* – but it should be stressed that we use here the term *regret function* also in the most generic sense, as it was used by (Loones and Sugden 1982) in their paper on *regret theory* stimulated by results of Kahneman and Tversky – thus, it is not necessarily the maxmin regret as used in some approaches to robust optimization, rightly criticised for too conservative results:

$$S(a,d) = (Q^0(a,d) - Q^*(d))/Q^*(d) \qquad (1)$$

$S(a,d)$ has a minimum at all $a = d$, but (unlike $Q^0(a,d)$) both with respect to a and d, since $S(a,a) = 0$, $S(a,d) \geq 0$ for all a and d. The relative sensitivity measure represents relative loss of performance due to the fact that we might be mistaken in assessing the "real" parameters d and use instead the value a of parameters in the optimized basic model; the resulting value of performance measure must be compared with *the optimal value that we would obtain if we knew the "real" value of the parameters d.*

It should be stressed that $Q^0(a,d) - Q^*(d) \neq Q^*(a) - Q^*(d)$, therefore, it is not sufficient to parametrically optimize the model and compare the results; we should first model the "real" situation and compute in the model the corresponding decision, and then apply this decision to something else. It is only our analytical trick that we assume this "something else" can be represented by another model, and that the differences between the models can be represented by differences in parameters. If we wish to test robustness computationally, we must represent possible "realities" by models. Moreover, we can represent even essentially different models by, say, their convex combination, thus by a simple parameter change. However, we must simulate the "real situation" taking into account all contextual aspects.

Only if we did not optimize in the basic model, it would be sufficient to compare $Q(a)$ and $Q(d)$ (without stars for not optimized variables); for small deviations between d and a, we could use $d - a$ as an approximation of $Q(d) - Q(a)$.

As indicated above, the optimization in the basic model can cause also problems with the *feasibility* of obtained solutions for a perturbed model of possible "realities". This happens in particular if the basic model is of linear programming type, see, e.g., (Bertsimas and Thiele 2006). Diverse methods can be used, but the best advice is that even the basic model should be realistic enough, that is, should use (possibly piece-wise linear) penalty factors instead of virtual constraints (constraints that can be violated in reality at some additional cost, but are included as hard constraints into the model just for simplicity).

3. The Need of an Implementation Model

The diagram from Fig. 1 misrepresents real processes of decision implementation in one essential respect: it tacitly assumes that the decision will be applied to reality as it was computed, in an *open loop* in terms of control science (the notation $Q^o(a,d)$ corresponds to o for *open loop*). However, decisions are applied often in various forms of a *closed loop*, taking into account some measurements of the impact of perturbations. This is done not only in control engineering, also in social processes when we make *contingent plans*, or in environmental planning with feedback, etc. Therefore, Fig. 1 must be supplemented with a decision *implementation model*, such as in Fig. 2, where the index i denotes a selected implementation rule and its model $F^i(y, x, a)$ (note that the implementation rule, if it does depend on parameters, it should only depend on the parameters a), while $Q^i(a,d)$ denotes the resulting performance index depending on the relation of three factors: the selected implementation rule i, the assumed basic value of parameters a, and the assumed perturbation value of doubts, "real" parameters d.

In (Wierzbicki 1977) it is shown that even if we optimize in the basic model, we can use an infinite number of different implementation rules (or of different *optimal controllers*) with quite different resulting relative sensitivity indexes:

$$S^i(a,d) = (Q^i(a,d) - Q^*(d))/Q^*(d) \geq 0 \qquad (2)$$

and we should select such implementation rules that have the smallest relative sensitivity indexes for α in a neighborhood of a selected $a = a^*$. We might be interested also in an integrated scalar robustness index obtained, e.g., by assuming a discrete probability distribution of perturbation model versions $d_j \in T_d$ in a selected neighborhood T_d of $d = a^*$ and computing the mean value of the relative sensitivity indexes:

$$R^i = \sum_{j \in J} p_j \ (Q^i(a^*,d_j) - Q^*(d_j))/Q^*(d_j) \qquad (3)$$

where p_j is the probability of the scenario that the "real" parameters will have the value d_j, J is the set of such scenarios. Sometimes it is difficult to assign probabilities to generated scenarios, but we can use the rule "when in doubt, use uniform distribution" and assign $p_j = 1/|J|$ where $|J|$ is the number of scenarios used; or we could use worst case scenario, replacing the weighted average in (3) by max operation, as in the maxmin regret indicator in robust optimization, see, e.g. (Kouvelis and Yu 1997); however, we must remember that we use these ways of modeling uncertainty not for robust design, but for modeling our doubts when testing robustness. Similarly, if an implementation rule is independently selected but the optimization results in many seemingly equivalent (optimal or nearly optimal) decisions, we can treat them as options i that should be selected according to their robustness index; thus, robustness computations can be used also to regularize optimization problems with non-unique solutions, see also (Makowski 2000a, b).

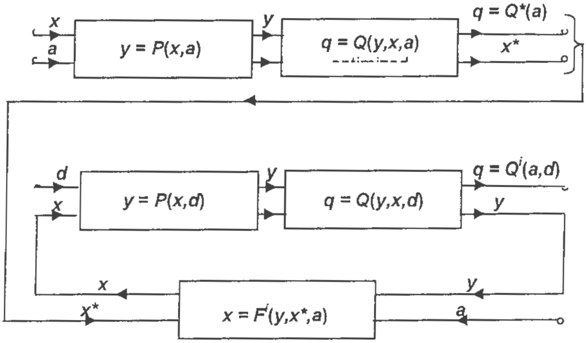

Fig. 2. A block diagram of the procedure of optimizing a decision and testing its robustness or sensitivity while taking into account a decision implementation model

4. A Relational Definition of Robustness Testing of Decisions x

Generally, we define robustness as a property of a relation between a basic model of decision situation (used as a basis of design, to select a decision; including a representation of uncertainty if the design should be robust) and a model of perturbations of such basic model, representing our doubts or possible "realities". The relation is resulting from application of the selected decision into perturbed "reality", taking into account also a rule of implementation of the selected decisions into "reality".

If we have an agreed (scalar valued) performance measure of decisions, then we define as most robust such decisions that – together with their implementation rules – result in the smallest worsening of the value of performance measure (regret function) resulting from the application of the selected decision under the assumed implementation rule and under assumed perturbations, compared to a predicted value of the performance measure. Such a comparison, as shown above, only in the simplest cases (not involving an optimization of decisions) can be based on the value of performance measure computed in the basic model. In the cases when decisions are optimized in the basic model, we must optimize them also for the perturbed models and base the comparison on the optimal value of the performance measure for the perturbed model under assumed perturbation (as if we had perfect foresight; but we actually make only contingent analysis). This is advisable even in cases of procedural robustness: if the basic model assumes, e.g., a robust optimization procedure based on given ranges of uncertainty parameters, we should first simulate the application of the resulting decision to a model with different uncertainty parameter ranges, then re-optimize the decision and compare the results while computing a kind of regret function similar to (1). Some papers on robust optimization (such as Laumanns et al. 2010) perform a similar testing.

The triad: *basic model, perturbation model, implementation model* with their fundamental relations should be the basis of sensitivity and robustness testing in all decision situations. Thus, we agree with (Ermoliev and Hordijk 2006) that the concept of robustness is context dependent, but we stress the need of a general structure to classify such contexts. Another issue is that the simulation and analysis of robustness can be concerned with statistical models, but also with deterministic models of more complex dynamic phenomena, etc. Another dimension of complexity relates to multiple criteria decision analysis.

5. The Issue of Multicriteria Decisions and Their Robustness

Multiple criteria decision analysis uses the concept of an efficient (Pareto-optimal) frontier in a multi-dimensional criteria space. It is well defined as the set of such criteria values that no criterion can be improved without deteriorating the value of another criterion, but is not readily adaptable for a generalization of such definition of robustness index as in Eq. (3).

The concept of robustness is used also in multiple criteria decision analysis with diverse meanings, see, e.g., (Kadziński and Słowiński 2012). (Barrico and Antunes 2006) propose that a robustness indicator should be incorporated in the fitness value in an evolutionary algorithm of finding the efficient frontier, which might result either in a modified efficient frontier (determined at the cost of increasing the robustness indicator, see later comments) or in a classification of the points along the original efficient frontier with an indication of robustness class. For the purpose of such classification, they use a discrete (natural number valued) robustness indicator, defined in a rather ad hoc way and called also *robustness degree*. The points on the efficient frontier can thus have an associated robustness degree, which is an advantage of such ad hoc approach. The disadvantage is the lack of a clear interpretation of parameters of this procedure (relatively clear in a bi-criteria case, but worsening with the increase of dimension of the objective space). Another disadvantage is that the implementation of decisions is tacitly assumed to be in open loop.

Thus, we should ask about a more general definition of a robustness indicator for the multiple objective case. The issue is how to extend the definition of robustness index as in Eq. (3) to analyzing the robustness of a point on efficient, Pareto-optimal frontier? To do this, we need a scalarized performance measure such as $Q^i(a,d)$ function that has a minimum with respect to the selected decisions, and thus to the basic model parameters a at $a = d$. However, such a function can be provided by the reference point approach to multiple criteria optimization, see, e.g., (Wierzbicki, Makowski and Wessels 2000): it is the achievement scalarizing function, e.g., of the form:

$$\sigma(q, q^{\bar{}}, \varepsilon) = \max_{k \in K} (q_k - q^{\bar{}}_k) + \varepsilon \sum_{k \in K} (q_k - q^{\bar{}}_k) \qquad (4)$$

where σ denotes the achievement function, q is the vector of the component criteria q_k, $q^{\bar{}}$ is the vector of the component reference levels $q^{\bar{}}_k$ in criteria space, K is the set of criteria indexes, $\varepsilon > 0$ is a regularization coefficient that should be small enough in order not to exclude too many efficient solutions that are nearly improperly efficient (with trade-off coefficients exceeding $1/\varepsilon$). To be sure that the achievement function has a minimum (equal zero) at $q = q^*$, where q^* is a point at the efficient frontier, it is sufficient to choose $q^{\bar{}} = q^*$; this is the basic check of efficiency in reference point approaches.

An achievement scalarizing function is treated as an *ad hoc* approximation to the value function of the decision maker based on her/his specification of reference levels (aspiration and/or reservation) for component criteria, see (Wierzbicki et al. 2000) and is typically maximized (we shall use such in some further examples). In Eq. (4) we converted its form here to the minimized case, assuming also that all multiple criteria are minimized.

Thus, we can use the achievement scalarizing function $\sigma(q, q^*, \varepsilon)$, interpreted as the performance measure $Q^i(a,d)$ (depending on basic model parameters a, on perturbation model parameters d and on the implementation rule i), to measure the robustness of a point q^* at the efficient frontier of a model with parameters a^* with a corresponding optimal decision x^*. That this function has a minimum also with respect to a results from a reasoning such as presented at the beginning of Section 2: if another $a \neq a^*$ is chosen as the basis of selecting the optimal decision, then the corresponding decision is not necessarily optimal when applied to modeled "reality" with $d = a^*$, thus the achieved performance measure characterized by the value $\sigma(q, q^*, \varepsilon)$ can be only worse. This reasoning, however, does apply only to the selection of basic model parameters a; if we change possible "realities" by selecting $d \neq a^*$, then the entire Pareto frontier can shift, thus the minimal value of the achievement scalarizing function with fixed $q^{\bar{}} = q^*$ can go down below zero (which means in this case that the "reality" turns out to be more advantageous than predicted).

We might thus use a robustness index such as defined by Eq. (3) except for one basic difference: because the achievement scalarizing function can be positive or negative (which corresponds to $Q^*(d)$ changing its sign) we cannot normalize the index by this value. Therefore, we might use as the robustness index of a vector q^* of criteria values on the Pareto frontier, corresponding to a decision x^*, just the value:

$$R^i_m = \sum_{j \in J} p_j \; (\varsigma^i(a^*,d_j) - \varsigma^i(d_j,d_j)) \tag{5}$$

where $\varsigma^i(a^*,d_j) = \sigma(Q^i(a^*,d_j), q^*, \varepsilon)$ is a corresponding value of achievement scalarizing function (4), $Q^i(a^*,d_j)$ denotes the vector of criteria values resulting from the application of the decision x^* modified by an implementation rule $F^i(y, x^*,a^*)$ to the model of a possible „reality" with parameters d_j, selected (or generated by a pseudo-random generator) from a chosen neighborhood of the parameter value $d = a^*$, constituting a family of scenarios $j \in J$. However, we must know also $\varsigma^i(d_j,d_j)$ that denotes the value of achievement scalarizing function (3) also with $q^{\sim} = q^*$, but re-optimized on the model of possible "reality" with parameters d_j (estimating the loss of optimality due to our imperfect foresight).

Such an approach to robustness testing of efficient solutions on Pareto-optimal frontier inherits the advantages of relative simplicity and uniformity of scalar and vector optimization as well as consistency with well tested methods of vector (multiple criteria) optimization. As commented above, we can use, instead of expected value (5) of the loss of optimality due to uncertainty, the maximal value due to the most pessimistic scenario; we must only remember that this representation of uncertainty has the goal of robustness testing, not of robust design. Thus, if the decision was based on a maximal regret analysis for multiple criteria case, as in, e.g., (Dias and Climaco 2003), then we must consider not only the set of assumed parameter variations or model versions $a_j \in T_a$ used to determine a robust decision, but also our doubts about the validity of these assumptions, a different set $d_j \in T_d$ which will be used in robustness testing, e.g., checking how the robust decisions will behave if we test them on model versions other than initially assumed.

6. Examples

We present here several diverse examples just to show the role of various aspects of robustness testing: starting with a very simple, academic multiple criteria example, quoting a single-criteria but more complex example with diverse implementation rules, following with more advanced multiple criteria managerial example, and indicating a complex example of air quality modeling. The diversity of such examples indicates the generic nature of the proposed methodology of robustness testing.

6.1. *Simple multiple criteria example*

We start with a very simple example illustrating the principles of robustness analysis on an elementary model M of multiple criteria decision problem with a simple scalarizing function. The problem has two criteria q_1 and q_2:

$$\max q_1 = x_1 \tag{6a}$$

$$\max q_2 = x_2$$

with constraints set X defined as follows

$$x_1 + x_2 \leq 1 \tag{6b}$$

$$x_1, x_2 \geq 0$$

We assume a simple perturbation model $M(d)$

$$\max q_1 = d\, x_1 \tag{6c}$$

$$\max q_2 = x_2$$

with the same constraints set X defined as in (6b); thus, $a^* = 1$. We also assume a simple scalarizing function (which has minima at Pareto solutions for the problem, as it is easy to check):

$$\sigma(q, \bar{q}) = \max_{i=1,2} (q_i - \bar{q}_i). \tag{6d}$$

If $\bar{q}_1 = 0.5$ and $\bar{q}_2 = 0.5$, then the minimum of the scalarizing function with respect to $x \in X$ occurs at $x_1 = 0.5$, $x_2 = 0.5$ for the problem (6a, b), while $\sigma(q, \bar{q}) = 0$ at this point. If we make a simple sensitivity analysis by changing d and computing the resulting values of σ, we obtain a graph such as in Fig. 3 (e.g., if $d > 1$, then we obtain $\sigma < 0$). However, such a graph does not inform us about robustness of the solution.

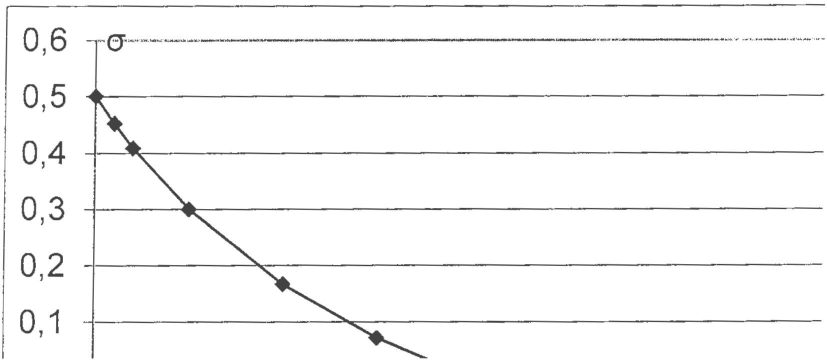

Fig. 3. Sensitivity graph

The robustness of the solution can be characterized by the simple robustness index (5). In this case, it suffices to compute the regret function $\varsigma(a^*,d) - \varsigma(d, d)$, where $a^*=1$, $\varsigma(a^*, d) = \sigma(Q(a^*,d), q^{\sim})$, while $Q(a^*, d)$ denotes the vector of objective functions (q_1, q_2) obtained when applying the solution $x_1 = 0.5$, $x_2 = 0.5$, optimal for the basic model (6a, b), to the perturbation model (6c, b), and $\varsigma(d, d) = \sigma(Q(d, d), q^{\sim})$, while $Q(d, d)$ denotes the vector of objective functions (q_1, q_2) obtained when applying a solution of (6c, b, d) re-optimized with respect to $x \in X$ to the perturbation model (6c, b). The re-optimization gives result $x_1 = 1/(1+d)$, $x_2 = d/(1+d)$. The regret function is presented in Fig. 4.

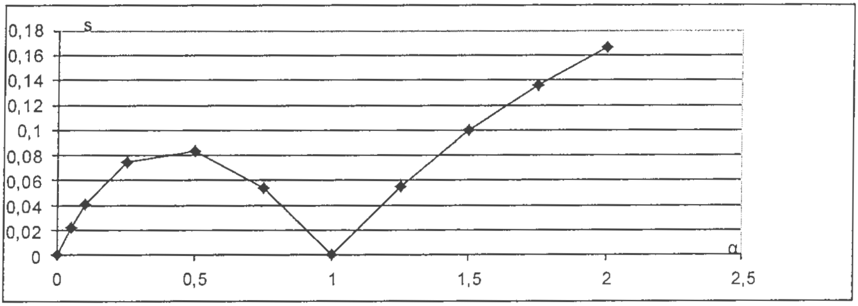

Fig. 4. Regret function

Note that in this simple example we assumed an open-loop implementation of decisions, without any complex decision implementation model.

6.2. *Example of time-optimal control*

As a more complex example of robustness testing, we quote here first from (Wierzbicki 1977) the analysis of decisions concerning time-optimal control, because they illustrate the importance of implementation model. In a basic case, we decide about the control trajectory $x(t)$ that influences the position of a servomotor with the basic model $dy_1/dt = y_2(t)$, $dy_2/dt = a\,x(t)$ in order to bring the model from a given initial position $y_1(0) = y_{10}$, $y_2(0) = y_{20}$ to the final position $y_1(T) = 0$, $y_2(T) = 0$ treated as a goal in the shortest time T treated as the performance measure, with a constraint $|x(t)| \leq 1$. The optimal decision is just to switch, say, from $x(t) = -1$ to $x(t) = +1$ at a suitable time $t_s(a)$; say, $t_s(a) = 1/a^{0.5}$ if $y_{10} = 1, y_{20} = 0$.

Three diverse rules of implementing this optimal decision were investigated in (Wierzbicki 1977): one is an *open loop control* $x^0(t) = sign\ (t - t_s(a))$, second is *closed-loop control* based on a switching manifold in the phase space $x^1(t) = -\,sign(ay_1(t) + 0.5\,(y_2(t))^2 sign\ y_2(t))$, third is a *trajectory following control* in which the control $x^2(t)$ is selected to follow the pre-computed optimal trajectory $y_1{}^*(t,a)$, $y_2{}^*(t,a)$.

If these implementation rules are applied to a model of possible "reality" $dy_1/dt = y_2(t)$, $dy_2/dt = \alpha\,x(t)$ with $\alpha \neq a$, then the open-loop implementation is just infinitely sensitive, cannot achieve the goal. The closed loop control is the most robust of the three, with possible oscillations or even sliding motion if $\alpha \neq a$, but always reaching the goal. The trajectory following control (trying to keep to a pre-computed optimal trajectory despite of parameter variations) is slightly less robust for $\alpha > a$, but infinitely less robust for $\alpha < a$, because in that case the pre-computed trajectory cannot be followed when $|x(t)| \leq 1$. The results are illustrated in Fig. 5.

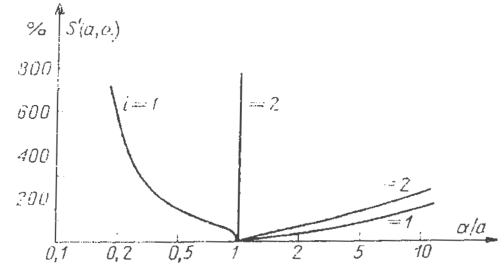

Fig. 5. Relative sensitivity index $S^i(a,d) = (Q^i(a,d) - Q^*(d))/Q^*(d)$ for the time-optimal control decisions with a scalar parameter α of possible "reality"; $i = 1$ – optimal closed loop; $i = 2$ – optimal trajectory following

The above is just an illustrative quote showing, however, that an implementation rule of pre-computed "optimal" as well as "robust" decisions matters significantly; thus, decision analysis should pay more attention to implementation rules. On the other hand, the same example in a time-discrete form might be a demanding testing example for evolutionary algorithms of optimization; therefore, we give here the discretized form:

$$y_1[t+1] = y_1[t] + y_2[t];\ y_2[t+1] = y_2[t] + a\,x[t];\ |x[t]| \leq 1;\ y_1[0] = 100;\ y_2[0] = 0;$$
$$|y_1[T]| \leq \delta;\ |y_2[T]| \leq \delta \qquad (6)$$

with $t = 0, 1, 2, \ldots T$ (square brackets stress the discrete nature of t), the goal slightly less demanding, e.g. $\delta = 1$ (in order to bring the end state sufficiently close, not necessarily equal to zero) and the minimized performance measure T. A bicriteria version of this test example might be including δ, T as two minimized criteria. Robustness testing might concern the same model with $d \neq a$ in place of a as a model of possible "reality".

6.3. *Example of ranking of discrete alternatives*

In (Wierzbicki 2008) the concept of so-called *objective ranking* of discrete alternatives was introduced (no ranking is absolutely objective, but in many situations we need a ranking that is as objective as possible) based on reference point approach with statistically defined reference points or aspiration and reservation levels. The following example (from an actual application, but distorted for commercial reasons) was analyzed.

Suppose an international corporation consists of six divisions A,...F. Suppose these units are characterized by diverse data items, such as name, location, number of employees etc. However, suppose that the CEO of this corporation is really interested in ranking or classification of these divisions taking into account the following attributes used as criteria:

1) profit (in % of revenue),

2) market share (m.share, in % of supplying a specific market sector, e.g. global market for a type of products specific for this division),

3) internal collaboration (i.trade, in % of revenue coming from supplying other divisions of the corporation), and

4) local social image (l.s.i., meaning public relations and the perception of this division – e.g., of its friendliness to local environment – in the society where it is located, evaluated on a scale 0-100 points).

All these criteria are maximized, improve when increased. An example of decision table of this type is shown in Table 1 (with data distorted; any similarity to an actual corporation is accidental), while Pareto-nondominated divisions are distinguished by mark *.

Table 1. Data for an example on international business management

Division	Name	Location	Employees	q_1: profit	q_2: m.share	q_3: i.trade	q_4: l.s.i.
A	Alpha	USA	250	11%	8%	10%	40
B*	Beta	Brasil	750	23%	40%	34%	60
C*	Gamma	China	450	16%	50%	45%	70
D*	Delta	Dubai	150	35%	20%	20%	44
E*	Epsilon	C. Europe	350	18%	30%	20%	80
F	Fi	France	220	12%	8%	9%	30

The CEO obviously could propose an intuitive, subjective ranking of these divisions – and this ranking might be even better than a rational one resulting from the table above, if the CEO knows all these divisions in minute detail. However, when preparing a discussion with her/his stockholders, (s)he might prefer to ask a consulting firm for an objective ranking.

In order to obtain such ranking, we compute first statistical averages of criteria values q_k^{av} that would be used as basic reference levels, a modification of these values to obtain aspiration levels a_k, and another modification of these values to obtain reservation levels r_k; these might be defined (for the case of maximization of criteria) as follows:

$$q_k^{av} = \Sigma_{l \in L}\, q_{lk}/|L|;\; r_k = 0.5(q_k^{lo} + q_k^{av})\;;\; a_i = 0.5(q_k^{up} + q_k^{av}) \qquad (7)$$

where $|L|$ is the number of alternative decision options, hence q_k^{av} is just an average value of k-th criterion between all alternatives, and aspiration and reservation levels – just averages of these averages and the lower and upper bounds, respectively. However, there are no essential reasons why we should limit the averaging to the set of alternative options ranked; we could use as well a larger set of data in order to define more adequate (say, historically meaningful) averages, or a smaller set – e.g., only the Pareto-nondominated alternatives.

The objective ranking approach uses a nonlinear aggregation of criteria by an achievement function that is performed in two steps:

1) We first count achievements for each individual criterion or satisfaction with its values by transforming it (monotonically and piece-wise linearly) e.g. in the case of maximized criteria as shown in Eq. (2) below. For problems with a continuous (nonempty interior) set of options, for an easy transformation to a linear programming problem, such a function needs additional specific parameters selected to assure the concavity of this function, see (Wierzbicki et al. 2000). In a discrete decision problem, however, concavity is not important for optimization performance; therefore we can choose these coefficients to have a reasonable interpretation of the values of the *partial (or individual) achievement function*. Since the range of [0; 10] points is often used for eliciting expert opinions about subjectively evaluated criteria or achievements, we adopted this range in Eq. (8) below for the values of a partial achievement function $\sigma_k(q_k, a_k, r_k)$:

$$\sigma_k(q_k, a_k, r_k) = \begin{cases} \mu\,(q_k - q_k^{lo})/(r_k - q_k^{lo}), & \text{for } q_k^{lo} \le q_k < r_k \\ \mu + (\nu - \mu)\,(q_k - r_k)/(a_k - r_k), & \text{for } r_k \le q_k < a_k \\ \nu + (10 - \nu)\,(q_k - a_k)/(q_k^{up} - a_k), & \text{for } a_k \le q_k \le q_k^{up} \end{cases} \tag{8}$$

where q_k^{lo} and q_k^{up} denote correspondingly the lower and upper bounds on criteria values. The parameters μ and ν, $0 < \mu < \nu < 10$, in this case denote correspondingly the values of the partial achievement function for $q_k = r_k$ and $q_k = a_k$. The value $\sigma_{kl} = \sigma_k(q_{kl}, a_k, r_k)$ of this achievement function for a given alternative $l \in L$ signifies the satisfaction level with the criterion value for this alternative. Thus, the above transformation assigns satisfaction levels from 0 to μ (say, $\mu = 3$) for criterion values between q_k^{lo} and r_k, from μ to ν (say, $\nu = 7$) for criterion values between r_k and a_k, from ν to 10 for criterion values between a_k and q_k^{up}.

2) After this transformation of all criteria values, we might use then the following form of the overall achievement function:

$$\sigma(q, a, r) = \min_{k \in K} \sigma_k(q_k, a_k, r_k) + \varepsilon/|K|\, \Sigma_{k \in K}\, \sigma_k(q_k, a_k, r_k) \tag{9}$$

where $q = (q_1, \dots q_k, \dots q_K)$ is the vector of criteria values, $a = (a_1, \dots a_k, \dots a_K)$ and $r = (r_1, \dots r_k, \dots r_K)$ are the vectors of aspiration and reservation levels, while $\varepsilon > 0$ is a small regularizing coefficient and $|K|$ is the number of criteria.

We use then the achievement values to rank the alternatives, as illustrated in Table 2, where $\varepsilon = 0.4$ was used ($|K| = 4$), for two types of averaging: of all alternatives or, more demanding, of Pareto-optimal alternatives.

Table 2. An example of objective ranking and classification for the data from Table 1

Criterion	q_1	q_2	q_3	q_4			
Upper bound	35%	50%	45%	80			
Lower bound	11%	8%	9%	30			
Reference *A* (average)	19.2%	26%	23%	54			
Aspiration *A*	27.1%	38%	34%	67			
Reservation *A*	15.1%	17%	16%	42			
Reference *B* (Pareto average)	23%	35.0%	29.7%	63.5			
Aspiration *B*	29%	42.5%	37.4%	71.7			
Reservation *B*	17%	17%	19.4%	46.7			
Ranking *A*: Division	σ_1	σ_2	σ_3	σ_4	σ	Rank	Class
A	0.00	0.00	0.37	2.50	0.29	5	III
B	5.63	7.50	7.00	5.88	8.23	1	I
C	3.30	10.0	10.0	7.62	6.39	2	II
D	10.0	3.57	3.89	3.32	5.40	4	II
E	3.97	5.48	3.89	10.0	6.30	3	II
F	0.73	0.00	0.00	0.00	0.07	6	III
Ranking *B*: Division	σ_1	σ_2	σ_3	σ_4	σ	Rank	Class
A	0.00	0.00	0.29	1.80	0.21	5	III
B	5.00	6.61	6.24	5.13	7.30	1	I
C	2.50	10.0	10.0	6.73	5.42	2	II
D	10.0	3.47	3.13	2.51	4.42	4	II
E	3.33	5.04	3.13	10.0	5.28	3	II
F	0.50	0.00	0.00	0.00	0.05	6	III

In Table 2, the column Class means a rough classification of divisions A,...F into three classes (required by the user in actual application); we see that ranking A and B give slightly different results but this does not influence the rough classification.

6.3.1. *Robustness of ranking*

Now we can illustrate robustness testing in two variants. First is the *robustness of ranking*, say, of ranking *A* (based on full averages). For this purpose, we assume two scenarios involving parameters that might influence the ranking and are likely to change; say, let $q_{4B} = 60\, d_1$ and $q_{1B} = 23\, d_2$ % with $d^*_1 = d^*_2 = a^*_1 = a^*_2 = 1$ (cf. Table 1). What is the robustness of this ranking if the parameters d_1 and d_2 change below 1? We must establish first reasonable ranges of parameter changes; let us assume that $d_1^{lo} = 0.5$ and $d_2^{lo} = 0.478$ (in both cases, criteria values reach their lower bounds – we can also continue the analysis below these criteria values by accordingly modifying lower bounds, but we omit here the precise description since it changes the character of computations). We compute then the dependence of the losses of achievement (if we maintain the first place of alternative B while another alternative becomes actually first because of parameter changes) $R_1(\alpha_1)$ and $R_2(\alpha_2)$ as the differences of attainable achievements (cf. Eq. 5 which we modify for this case just by assuming originally that the probability of examined scenario is equal 1).

The character of achievement functions used here is piece-wise linear and an essential problem is to compute the value of parameter change when linear pieces meet. For example, as long as the alternative B remains first in ranking, the loss of achievement either $R_1(\alpha_1)$ or $R_2(\alpha_2)$ remains equal zero (the achievement might change, but after re-optimization it is the same achievement). This leads to the problem how to find the critical values of d_1 or d_2, where the losses exceed zero. We illustrate the solution only for the case of d_1.

For d_1 close to 1, σ_{4B} has the form (resulting from Eq. 8 with $\mu = 3$ and $\nu = 7$) $\sigma_{4B} = 9.6d_1 - 3.72$; it becomes the lowest of σ_{kB} when $9.6d_1 - 3.72 = 5.63$ (the value of σ_{1B} in Table 2), hence for $d_1 = 0.978$, and becomes equal to 3.0 (at the reservation level) when $9.6d_1 - 3.72 = 3.00$, hence for $d_1 = 0.70$. However, for $d_1 < 0.978$, the overall achievement σ_B takes the form (Eq.9 with $\varepsilon = 0.4$ and $|K| = 4$) $\sigma_B = 10.56\ d_1 - 2.079$. Since the second alternative in ranking is C with $\sigma_C = 6.39$, alternative B loses its first rank when $10.56\ d_1 - 2.079 = 6.39$, at $d_1 = 0.802$; etc. Results of such computations are given in Table 3 and Fig. 6. Similar computations for d_2 give results illustrated in Table 4 and Fig. 7; in this case, alternative B loses its first rank at $d_2 = 0.782$, etc.

Table 3. Results of analysis of ranking robustness, scenario 1 with $q_{4B} = 60\ d_1$

d_1	1.0	0.9	0.802	0.8	0.7	0.6	0.5
q_{4B}	60	54	48.12	48	42	36	30
σ_{4B}	5.88	4.92	3.98	3.96	3.00	1.50	0.00
σ_B	8.23	7.425	6.39	6.37	5.313	3.663	2.013
σ^*	8.23	7.425	6.39	6.39	6.39	6.39	6.39
$R_1 = \sigma^* - \sigma_B$	0.00	0.00	0.00	0.02	1.077	2.727	4.377

Table 4. Results of analysis of ranking robustness, scenario 2 with $q_{1B} = 23\ \alpha_2$ %

d_2	1.0	0.9	0.8	0.782	0.7	0.657	0.6	0.478
q_{1B}	23%	20.7%	18.4%	18.0%	16.1%	15.1%	13.8%	11%
σ_{1B}	5.63	4.86	4.10	3.96	3.33	3.00	2.11	0.00
σ_B	8.23	7.39	6.54	6.39	5.70	5.33	4.33	2.03
σ^*	8.23	7.39	6.54	6.39	6.39	6.39	6.39	6.39
$R_2 = \sigma^* - \sigma_B$	0.00	0.00	0.00	0.00	0.69	1.06	2.06	4.36

We can see that the robustness index changes are quite similar for both scenarios, they differ in minor details. The ranking is quite robust, alternative B loses its first rank (to alternative C) when data change more than 20% of their initial value.

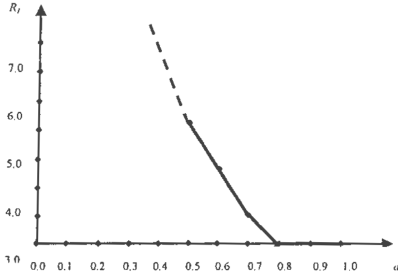

Fig. 6 Regret function change for a simple example, scenario 1 ($a_1^* = 1$)

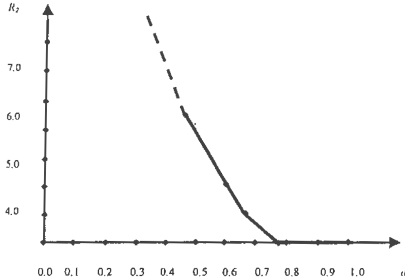

Fig. 7 Regret function change for a simple example, scenario 2 ($a_2^* = 1$)

6.3.2. *Robustness of Pareto-optimal solutions*

On the same example, say, for the second ranking with more demanding aspirations (ranking *B* in Table 2), we can study the question: how to compare the *robustness of several Pareto-optimal solutions*? Let us assume that we want to compare the robustness of two Pareto-optimal solutions, alternatives B and C, with respect to four scenarios of parameter change: $q_{4B} = 60\, d_1$, $q_{1B} = 23\, d_2$ %, $q_{4C} = 70\, d_3$, $q_{1C} = 16 d_4$ %, with $a^*_1 = a^*_2 = a^*_3 = a^*_4 = 1$. For simplicity of presentation, we shall consider here a composite scenario jointly changing $d = d_1 = d_2 = d_3 = d_4$ and assuming equal probabilities (we could, of course, assume much more complex scenarios with parameters influencing all data of this example and diverse probabilities, but we present here only this simple case for illustrative purposes).

In order to test the robustness of Pareto-optimal solutions, we must apply the method presented in Section 5: we can compute the losses of achievement functions due to parameter changes, but with reference points shifted to the criteria values at the corresponding Pareto-optimal solutions. Since we used in this example achievement function (9) with double-valued reference point (composed of aspirations and reservations), we shall keep this form of achievement function, only assume that the aspiration levels are shifted to the criteria values at the corresponding Pareto-optimal solutions, while reservation levels are averages of lower bounds and aspiration levels. These assumptions result in the following table of aspiration and reservation levels:

Table 5. Aspiration and reservation levels for robustness analysis of Pareto-optimal solutions

Criterion	q_1	q_2	q_3	q_4
Upper bound	35%	50%	45%	80
Lower bound	11%	8%	9%	30
Analysis of alternative B				
Aspiration a_B	23%	40%	34%	60
Reservation r_B	17%	24%	21.5%	45
Analysis of alternative C				
Aspiration a_C	16%	50%	45%	70
Reservation r_C	13.5%	29%	27%	50

Now we can compute the partial achievements (8) and overall achievement (9) together with their dependence on *d*, because we assume joint changes $q_{4B} = 60d$, $q_{1B} = 23d\%$, $q_{4C} = 70d$, $q_{1C} = 16d\%$; we consider $d \geq 0.687$, since q_{1C} reaches then its lower bound (as noted above, we can perform also the analysis with changing lower bounds, but we omit these details). We must only remember to compare the overall achievement for alternative B or C with other Pareto-optimal alternatives and compute the loss of achievement if some other alternative starts to outrank B or C; this is equivalent to re-optimization assumed in Eq. (5)). Therefore, we must compute achievement values for all Pareto-optimal alternatives (according to the properties of achievement scalarizing functions, the dominated alternatives might be neglected). In Table 6 we present the resulting achievement values for the analysis of alternative B.

Table 6. Achievement values for the analysis of alternative B

Analysis of alternative B Division	σ_1	σ_2	σ_3	σ_4	σ	Rank $d = 1$
B	$7.00 \geq \sigma_{1B}$	7.00	7.00	$7.00 \geq \sigma_{4B}$	$9.80 \geq \sigma_B$	1
C	$2.50 \geq \sigma_{1C}$	10.0	10.0	$8.50 \geq \sigma_{4C}$	$5.60 \geq \sigma_C$	2
D	10.0	2.25	2.64	2.80	3.98	4
E	3.67	4.50	2.64	10.0	4.72	3

After that we can compute the dependence of σ_B and σ_C on *d*; the results are presented in Table 7.

Table 7. Results of robustness testing of alternative B, composite scenario

d	1.0	0.9	0.8	0.7
q_{4B}	60	54	48	42
q_{1B}	23%	20.7%	18.4%	16.1%
q_{4C}	70	63	56	49
q_{1C}	16%	14.4%	12.8%	11.2%
σ_{4B}	5.88	4.92	3.96	3.00
σ_{1B}	7.00	5.40	3.80	2.40
σ_{4C}	8.50	7.45	5.93	4.07
σ_{1C}	2.50	1.70	0.9	0.1
σ_B	8.23	7.43	6.37	5.31
σ_C	5.60	4.72	3.84	2.96
σ^*	8.23	7.43	6.37	5.31
$R_B = \sigma^* - \sigma_B$	0.00	0.00	0.00	0.00

We see that alternative B is very robust as a Pareto-optimal solution: it does not lose its first place even with parameter changes exceeding 30%. When we analyze the robustness of the alternative C as a Pareto-optimal solution, we must change accordingly aspiration and reservation levels, as presented in Table 5. The achievement values change accordingly as in Table 8.

Table 8. Achievement values for the analysis of alternative C

Analysis of alternative B Division	σ_1	σ_2	σ_3	σ_4	σ	Rank at $d = 1$
B	$8.11 \geq \sigma_{1B}$	5.10	4.56	$5.00 \geq \sigma_{4B}$	$6.79 \geq \sigma_B$	1
C	$7.00 \geq \sigma_{1C}$	10.0	10.0	$7.00 \geq \sigma_{4C}$	$10.4 \geq \sigma_C$	2
D	10.0	2.25	2.64	2.80	3.98	4
E	3.67	4.50	2.64	10.0	4.72	3

After that we can again compute the dependence of σ_B and σ_C on d; the results are presented in Table 9.

Table 9. Results of robustness testing of alternative C, composite scenario

d	1.0	0.9	0.8	0.7
q_{4B}	60	54	48	42
q_{1B}	23%	20.7%	18.4%	16.1%
q_{4C}	70	63	56	49
q_{1C}	16%	14.4%	12.8%	11.2%
σ_{4B}	5.00	3.80	2.70	1.80
σ_{1B}	8.11	7.25	7.13	7.01
σ_{4C}	7.00	6.30	5.60	4.90
σ_{1C}	7.00	4.44	2.16	0.24
σ_B	6.79	5.87	4.65	3.65
σ_C	10.4	7.51	4.93	2.75
σ^*	10.4	7.51	4.93	4.72
$R_C = \sigma^* - \sigma_C$	0.00	0.00	0.00	1.97

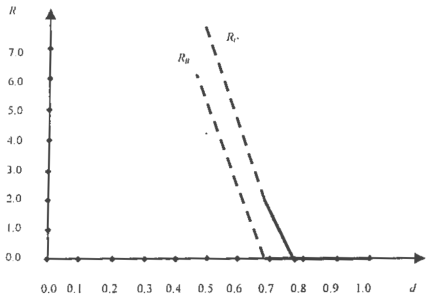

Fig. 8 Regret functions for comparing robustness of Pareto-optimal alternatives B and C

Again, alternative C as a Pareto-optimal solution is quite robust, but loses its first rank (to alternative E) at about $\alpha = 0.79$, thus it is less robust than alternative B as a Pareto-optimal solution. These results are illustrated in Fig. 8.

We presented here the computations related to robustness testing in some detail for illustrative purposes. The conclusions are that even if robustness testing is computationally expensive, and even if the assumptions of data perturbation scenarios are quite subjective, belong to the art of modeling, it is possible nevertheless to compare robustness of diverse Pareto-optimal solutions.

6.4. *Example of the air quality modeling*

Models of complex environmental problems, such as European air quality, have not only very large dimensions, but also an infinite number of near-optimal solutions (Makowski 2000a); if we formulate them as multiple criteria problems, they also have infinitely many Pareto-optimal solutions. Such solutions are similar in the sense of the objective function (or criteria) values, but have usually very different compositions of values of decision variables, which in turn imply substantial differences in obligations of partners of international agreements. This property can be treated as advantageous, if one uses an additional criterion for selecting between a large number of suboptimal solutions a specified solution that has an additional property, for example being robust. This can be achieved by defining a robustness indicator and including such indicator in the analysis of the model results driven by decision maker preferences: the most robust solutions would then be preferred over all less robust solutions having similar values of the original objective function. Such a technique is known as *regularization* and its application to the European air quality model by using a classical regularizing term (e.g. minimization of a norm, or of a distance from a preferred point is the subspace of decision variables) is discussed in (Makowski 2000b). Here we propose an extension of this approach by replacing the classical regularizing term by a robustness indicator.

The key problem then is to define a robustness indicator. (). For using the robustness index defined by Eq. (3) or (5), we shall select a reference (trial) decision, e.g., one of near-optimal solutions, or a Pareto-optimal one, and develop a number $|J|$ of possible "reality" scenarios characterized by parameters α_j and probabilities p_j. For each of such scenarios, we not only compute the value of $Q^j(a^*,d_j)$ or $\varsigma^i(a^*,d_j)$, estimating the performance obtained through the decision optimized for a basic value of parameters a^*; , but we also re-optimize the decisions to determine the value $Q^*(d_j)$ or $\varsigma^i(d_j,d_j)$ for different representations of the "real" value denoted by α_j thus estimating the performance obtained if we knew beforehand how the "reality" would change and optimized perfectly. The number of scenarios is practically limited by computational resources; for larger models the computations are usually dispersed over a computational grid, therefore this limitation can easily be softened. To construct such scenarios, we typically select a subset of parameters of selected constraints considered crucial; such a selection can be done during the post-optimization analysis of the basic solution focused on identification of those binding constraints that have key impact, e.g., either environmental or economical. For the selected subset of parameters a manageable set J of scenarios is generated, i.e. $|J| = n$, where n is the number of scenarios, which depends mainly on computational resources. The variations of the stochastic parameters are context-spacific. They are uniformly distributed over a given range of values, if there is no sufficient knowledge about their variation. If the probability distributions of selected parameters are known, then one can use a sampling technique to generate scenarios.

One should also note that the multiple-criteria analysis copes in a natural way with the issue of price of robustness, see e.g., (Bersimas, Sim, 2004). The main performance indicator of the air quality problem discussed here is the cost-effectiveness; therefore the typical reference decision is the one minimizing the total cost. Other indicators include diverse measures of air pollution impacts. A set of selected indicators can be used as criteria for which the aspiration-reservation methodology of multiple-criteria analysis can be used (see e.g., Wierzbicki et al. 2000). Therefore a natural approach to analysis of the price of robustness it to add a criterion representing the robustness measure, and to perform a truly integrated multiple-criteria analysis of the whole problem; such an approach provides the decision-makers with an intuitive way to analyze diverse trade-offs between conflicting criteria. In such a way also the price of robustness can be analyzed.

Note that there is also a trade-off between the number of selected stochastic parameters and the corresponding scenarios for each trial decision on one side, and the number of trial decisions compared for robustness on the other side; , the computational effort quickly becomes excessive unless at least one of these numbers is small. Usually, the computation of the values $Q^j(a^*,d_j)$ or $\varsigma^i(a^*,d_j)$ through simulation takes much less computational effort than the re-optimization needed to determine the value $Q^*(d_j)$ or $\varsigma^i(d_j,d_j)$ (which is equivalent to repeating many times the computation of the values

$Q^i(a^*,d_j)$ or $\varsigma^i(a^*,d_j)$). This necessity of repeating many times the re-optimization ("what if we knew beforehand what would happen in reality") and the related large computational effort is the main obstacle in such applications of robustness analysis for large scale models.

A selection of a robustness indicator is problem specific, and belongs to the modeling art; thus, the definition provided above is just an example. The robustness indicator used as an advanced regularizing term is typically used for achieving numerical stability, and in such situations the original goal function is the primary objective for selecting a solution. However, a robustness indicator can play a role of a criterion (or a component of an objective function in traditional single-criterion optimization approaches), and then support analysis of trade-offs between the original objective (or a set of objectives in multi-criteria analysis) and robustness of a selected solution. In such traditional approaches, for example, assuming the single objective to find a cost-effective solution for achieving agreed environmental standards. It might be rational to consider (possibly slightly or substantially) more expensive (than thecost-minimizing) solutions that are more robust than the cheapest one. This requires analyzes of the trade-offs between cost and robustness. In a simplistic approach such analysis can be done using a composite objective function defined as a weighted sum of the cost and robustness indicator. However, such approach has a number of pitfalls (see, e.g. Wierzbicki et al., 2000), therefore one should use a truly multi-criteria analysis outlined above. A critical element for the robustness analysis remains however the selection of the robustness indicator, which is by far a more difficult issue than a selection of an outcome variable to serve as a criterion. The methodology of specification of the robustness indicators is still an open research issue.

7. Conclusions

There are diverse concepts and approaches to robustness and sensitivity analysis serving different purposes. The paper deals with a specific issue of robustness testing focused on multiple-criteria decision-making support; therefore it concentrates on robustness of decisions. The concept of robustness of decisions is context dependent; we proposed to distinguish clearly between *robust design* (*ex ante* robustness analysis, including *robust optimization* etc.) and *robustness testing* (*ex post* robustness analysis); the paper concentrated on the latter case and we proposed a general structure to classify contexts of robustness testing. The structure relates to a triad: *basic model, perturbation model,* and *implementation model* with their fundamental relations explained in the paper. The roles of the basic model and the perturbation model in decision processes involving optimization are not symmetric; therefore one cannot just analyze parametrically one family of models, a more involved sensitivity and robustness testing isthus proposed. The simulation and testing of robustness can be concerned with statistical models, but also with deterministic models of more complex dynamic phenomena. Another dimension of complexity relates to multiple criteria decision analysis. We proposed an approach to robustness testing of efficient solutions on Pareto-optimal frontier that has the advantage of relative simplicity and uniformity of scalar and vector optimization as well as consistency with well tested methods of vector optimization, especially effective for decision-making support through interactive multiple-criteria model analysis.

A general conclusion of this paper is that robustness testing should be included into good practice of model-based decision analysis and support. Rational decision-making support requires a process that results in a design of a robust decision; it is however not enough to base it on a robust optimization using a selected model it is necessary to test robustness of such decions on the implementation model.

Acknowledgement

The comments of two anonymous reviewers contributed greatly to an essential clarification of the goal and contribution of this paper.

References

1. Barrico C,, Antunes C.H. Robustness Analysis in Multi-Objective Optimization Using a Degree of Robustness Concept. *Proceedings of the 2006 IEEE Congress on Evolutionary Computation,* 6778-6783, Vancouver, Canada, July 2006
2. Barrico C., Antunes C.H. Robustness Analysis in Evolutionary Multi-Objective Optimization. Working Paper of INESC Coimbra (2006)
3. Bello L., Blanquero R., Carrizosa E. On Minimax-regret Huff Location Models. In C.H. Antunes, D. R. Insua, L. C. Dias (Eds.), Proceedings of the 25th Mini-EURO Conference on *Uncertainty and Robustness in Planning and Decision Making*, Coimbra, Portugal, April 2010.
4. Ben-Tal, A., Ghaoui, El, L., and Nemirovski, A. 2009. *Robust Optimization*, Princeton University Press
5. Bertsimas D., Sim M. The Price of Robustness, *Operations Research* 52 (2004) 35-53
6. Bertsimas D., Thiele A. A Robust Optimization Approach to Inventory Theory. *Operations Research* 54: (2006) 1: 150-168
7. Ermoliev Y., Hordijk L. Facets of Robust Decisions. In Marti K., Ermoliev Y., Makowski M., Pflug G. (eds) *Coping with Uncertainty* (Springer Verlag, Berlin-Heidelberg 2006)
8. S. Gupta, J. Rosenhead, Robustness in sequential investment decisions, *Management Science* 15 (2) (1968) 18–29
9. Huber J.P. *Robust Statistics.* (J. Wiley & Sons, New York 1981)
10. Kadziński M., Słowiński R. Interactive Robust Cone Contraction Method For Multiple Objective Optimization Problems. *International Journal of Information Technology and Decision Making (IJITDM)* 11 (2012) 2:327-357
11. Kouvelis P. and Yu G. *Robust discrete optimization and its application.* (Kluwer Academic Publishers, Norwell, M.A 1997).
12. Laumanns M., Zenklusen R., Schüpbach K., Hendricks M., Lefeber E., Udding J. Robust Adaptive Resource Allocation in Container Terminals. In C.H. Antunes, D. R. Insua, L. C. Dias (Eds.), Proceedings of the 25th Mini-EURO Conference on *Uncertainty and Robustness in Planning and Decision Making*, Coimbra, Portugal, April 2010.
13. Loones G., Sugden R. (1982) Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal* 92: 805-824
14. Makowski M. Modeling Techniques for Complex Environmental Problems, in: Makowski, M and Nakayama H, eds., *Natural Environment Management and Applied System Analysis,*

(IIASA, Laxenburg, Austria 2000a); the paper is also available at: http://www.iiasa.ac.at/~marek/ftppub/MM/nemasa_mm.pdf

15. Makowski M. Modeling paradigms applied to the analysis of European air quality, *European J. Oper. Res.* 122: 2 (2000b) 219-241
16. Naseraldin H., Milner J., Baron O. Facility Location: A Robust Optimization Approach. In C.H. Antunes and D.R. Insua, eds.: *Uncertainty and Robustness in Planning and Decision Making,* proceedings of a conference in Coimbra, April 2010.
17. Nikulin Y. Robustness in Combinatorial Optimization and Scheduling Theory: An Extended Annotated Bibliography. Working Paper of Christian-Albrechts-Universität zu Kiel, Institut für Betriebswirtschaftslehre, Olshausenstr 40, 24118 Kiel, Germany (2006)
18. Ogryczak W. Tail Mean and Related Robust Solution Concepts. In C.H. Antunes and D.R. Insua, eds.: *Uncertainty and Robustness in Planning and Decision Making,* proceedings of a conference in Coimbra, April 2010.
19. Polanyi M. *The Tacit Dimension* (Routledge and Kegan, London 1966)
20. Popper K.R. *Objective Knowledge* (Oxford University Press, Oxford 1972)
21. Rosen R. *A Comprehensive Inquiry into the Nature, Origin and Fabrication of Life* (Columbia University Press, New York 1991)
22. Roy B. A missing link in OR-DA: robustness analysis, *Foundations of Computing and Decision Sciences* 23 (1998), 141-160
23. Saltelli A., Chan K., Scott E.M., eds. *Sensitivity Analysis* (J. Wiley & Sons, Chichester 2000)
24. P. Vincke, Robust solutions and methods in decision-aid, *Journal of Multicriteria Decision Analysis* 8 (1999) 181–187.
25. P. Vincke, About Robustness Analysis, *EWG-MCDA Newsletter,* Series 3, No.8, Fall 2003
26. Wierzbicki A.P. *Models and Sensitivity of Control Systems.* (WNT-Elsevier, Warsaw-Amsterdam, 1977, English version in 1984).
27. Wierzbicki A.P. The problem of objective ranking: foundations, approaches and applications. *Journal of Telecommunications and Information Technology* (3/2008) 15-23
28. Wierzbicki A.P., Makowski M., Wessels J., eds. *Model-Based Decision Support with Environmental Applications* (Kluwer, Dordrecht 2000)
29. Wierzbicki, A.P. and Nakamori Y. *Creative Space: Models of Creative Processes for the Knowledge Civilization Age* (Springer Verlag, Berlin-Heidelberg 2006)
30. Wierzbicki A.P., Nakamori Y., eds. *Creative Environments: Issues of Creativity Support for the Knowledge Civilization Age* (Springer Verlag, Berlin-Heidelberg 2007)