International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report **IR-06-001**

Global Changes: Facets of Robust Decisions

Y. Ermoliev and L. Hordijk

Approved by

Leen Hordijk (hordijk@iiasa.ac.at)
Director, IIASA

January 2006

Abstract

The aim of this paper is to provide an overview of existing concepts of robustness and to identify promising directions for coping with uncertainty and risks of global changes. Unlike statistical robustness, general decision problems may have rather different facets of robustness. In particular, a key issue is the sensitivity with respect to low-probability catastrophic events. That is, robust decisions in the presence of catastrophic events are fundamentally different from decisions ignoring them. Specifically, proper treatment of extreme catastrophic events requires new sets of feasible decisions, adjusted to risk performance indicators, and new spatial, social and temporal dimensions. The discussion is deliberately kept at a level comprehensible to a broad audience through the use of simple examples that can be extended to rather general models. In fact, these examples often illustrate fragments of models that are being developed at IIASA.

Key words: Robustness, decisions, uncertainty, stochastic optimization, discounting, downscaling, catastrophe modeling, extreme events, simulation.

Acknowledgments

The authors are thankful to the participants of the IFIP/IIASA/GAMM workshop on Coping with Uncertainty (December 13-16, 2004, IIASA) and to the IIASA colleagues for their critical and constructive comments which led us to important improvements of this paper.

Contents

Global Changes: Facets of Robust Decisions

Y. Ermoliev and L. Hordijk

1 Introduction

An alarming global tendency is the increasing vulnerability of our society. A thorough scientific policy analysis of related socio-economic, technological and environmental global change processes raises new methodological problems that challenge traditional approaches and demonstrate the need for new methodological developments. A key issue is the vast variety of inherent, practically irreducible uncertainties and "unknown" risks that may suddenly affect large territories and communities [7], [13], [20]. Traditional approaches usually rely on real observations and experiments. Yet, there are no adequate observations for new problems, responses of involved processes may have long term delays, and learning-by-doing experiments may be very expensive, dangerous, or simply impossible.

Large-scale catastrophic impacts and the magnitudes of the uncertainties that surround them particularly dominate the climate-change policy debates [2], [8], [29], [33], [34], [36], [48]. The exact evaluation of overall global climate changes and vulnerability requires not only a prediction of the climate system, but also an evaluation of endogenous socioeconomic, technological, and environmental processes and risks. The main issue is the lack of historical data on potential irreversible changes occurring on large spatial, temporal, and social scales. The inertia of the overall climate change system and the possibility of abrupt catastrophic changes [2] restricts purely adaptive wait-and-see approaches. Moreover, extreme events of heavy consequences playing such a decisive role are, on average, evaluated as improbable events during a human lifetime.

Unfortunately, the evaluation of complex heterogeneous global-change processes on "average" can be dramatically misleading. A 500-year disaster (e.g., an extreme flood that occurs on average once in 500 years) may, in fact, occur next year. However, it is impossible to research all the details connected with such an occurrence in order to achieve evaluations required by the traditional models in economics, insurance, risk-management, and extreme value theory. For example, standard insurance theory essentially relies on the assumption of independent, frequent, low-consequence (conventional) risks, such as car accidents, for which decisions on premiums, claims estimates and the likelihood of insolvency can be calculated via rich historical data. Existing extremal value theory [11] also deals primarily with independent variables quantifiable by a single number (e.g., money). Catastrophes are definitely not quantifiable events in this sense. They have different patterns, spatial and temporal dimensions and induce heterogeneity of losses and gains which exclude the use of average characteristics. Globally, an average resident may even benefit from some climate-change scenarios, while some regions may be simply wiped out.

Under inherent uncertainty and heterogeneity of global processes the role of global change models rests on the ability to guide comparative analysis of the feasible decisions. Although exact evaluations are impossible, the preference structure among decisions can be a stable basis for a relative ranking of alternatives in order to design robust policies.

As we know, finding out which of two parcels is the heavier without having the exact measurements is easier than saying how much heavier that parcel is.

Sections 2 and 3 analyze the known concepts of robustness in statistics, deterministic control theory and classical optimization. Global change decision problems call for new approaches. Sections 3 and 4 show that, contrary to the standard expected utility maximization, stochastic optimization (STO) models allow in a natural manner to represent different endogenous uncertainties and risks, spatial and temporal dependencies, equity constraints and abrupt changes. The ability of STO models to incorporate both anticipative ex-ante and adaptive ex- post decisions induces risk aversion among ex-ante decisions that implicitly depends on input data and goals and that practically cannot be characterized by an exogenous utility function. In particular, even in the simplest linear model (Example 4), the co-existence of ex-ante and ex-post decisions induces VaR and CVaR type risk measures. Section 4 also indicates the misleading character of average characteristics, e.g., hazard maps, which are often used in the analysis of spatial exposures and vulnerability. This emphasizes the importance of stochastic models, distributional aspects, and the use of quantiles instead of average values. Unfortunately, the straightforward application of quantiles destroy additivity and concavity (convexity) of models and it makes the applicability of standard decomposition schemes problematic (Example 2). Section 5 introduces concepts of STO robustness. In particular, it shows that models with quantiles can be equivalently substituted by specific STO models preserving concavity (convexity). Section 6 emphasizes the role played by downscaling and catastrophe modeling to properly represent spatial and temporal distributions and vulnerability. Section 7 outlines the main ideas behind STO methods, especially, fast adaptive Monte Carlo optimization procedures which can be incorporated into catastrophe models and vulnerability analysis in order to evaluate robust strategies. Section 8 discusses the sensitivity of robust strategies with respect to extreme events. It introduces the concept of a stopping time which allows for direct evaluations towards the most distractive extreme events (random scenarios). Combined with the catastrophe modeling, this concept opens up new approaches to spatio-temporal discounting in the presence of extreme events. Section 9 provides concluding remarks.

2 Concepts of robustness.

2.1 Statistical robustness.

The term "robust" was introduced into statistics in 1953 by Box and acquired recognition only after the publication of a path-breaking paper by Huber [27] in 1964. As Huber admits, researchers had long been concerned with the sensitivity of standard estimation procedures to "bad" observations (outliers), and the word "robust" was loaded with many, sometimes inconsistent connotations, frequently for the simple reason of confering respectability on it. According to Huber ([27], pp. 5, 6), "... any statistical procedure ... should be robust in the sense that small deviations from the model assumptions should impair the performance only slightly ..." This concept of robustness, in fact, corresponds to standard mathematical ideas of continuity: when disturbances become small, the performance of the perturbed and initial models also deviate slightly.

2.2 Bayesian robustness.

The concept of robustness was also introduced into Bayesian statistics [28] primarily because of the insensitivity of statistical decisions to the uncertainty of prior probability distribution. A Bayesian sampling model P is often parameterized by a vector θ of unknown parameters. Let ξ be the observable random variables from P with true unknown parameters $\theta = \theta^*$ that have to be recovered from observations of ξ. In contrast to classical statistical models, it is assumed that there is a prior (probability) distribution $\pi(\cdot)$ characterizing the degree of beliefs about true vector θ^*, which in the presence of new information is updated by the Bayesian rule. In this case a statistical decision (estimate) about the true parameter θ^* can be characterized by an expected distance (loss function) $EL(x,\theta) = \int L(x,\theta)\pi(d\theta)$ from x to admissible θ. The efficiency of x is calculated by the posterior expected distance

$$E(L(x,\theta|\xi) = \int L(x,\theta)\pi(d\theta|\xi), \pi(d\theta|\xi) = \frac{P(\xi|\theta)\pi(d\theta)}{\int P(\xi|\theta)\pi(d\theta)}, \quad (1)$$

where ξ is a given sample of data from $P(\xi|\theta^*)$. Bayesian robustness is characterized by the range of posterior expected distance, as the prior $\pi(\cdot)$ varies over the elicited class $\mathcal{P}$. An alternative approach is to choose a hyper-prior on the class of $\mathcal{P}$ and the standard Bayesian model.

2.3 Non-Bayesian minimax robustness.

A probabilistic minimax robustness [28] consists of choosing x with respect to a worst-case distribution: minimize $\max_{\pi \in P} \int d(x,\theta)\pi(d\theta|\xi)$. This type of minimax ranking of x does not correspond to the Bayesian ranking w.r.t. a single distribution in $\mathcal{P}$. The worst-case distribution $\pi \in \mathcal{P}$ depends on x and ξ, i.e., it is a random endogenous distribution.

2.4 Deterministic control theory.

As statistical robustness is similar to the local stability of dynamic systems, the robustness in deterministic control theory [42] was introduced as an additional requirement on the stability of optimal trajectories. In other words, additional constraints were introduced in the form of a stability criterion.

2.5 Robust deterministic optimization.

Optimization theory provides tools for analyzing and solving various decision making problems. A standard deterministic problem is formulated as the maximization (minimization) of a function $f_0(x,\omega)$ subject to constraints $f_i(x,\omega) \geq 0$, $i = 0,1,...,m$, where $x = (x_1,...,x_n)$ is a vector of decisions and ω are fixed variables characterizing the structure of the model, including the input data. Functions $f_i(x,\omega)$, $i = 0,1,...,m$, are assumed to be exactly known and analytically tractable, and ω belongs to an explicitly given set Ω of admissible scenarios, $\omega \in \Omega$. Robustness is defined [4] as the maximization of $\min_{\omega\in\Omega} f_0(x,\omega)$ over solutions x that satisfy all admissible values of uncertainty $f_i(x,\omega) \geq 0$, $i = 1,...,m$, $\omega \in \Omega$. The set Ω is often characterized by a finite number of scenarios or simple sets such as intervals or ellipsoidal uncertainty $\Omega = \{\alpha_l + \sum_k \delta_{lk}\omega_k : \sum_k \omega_k^2 \leq 1\}$. These sets, in a sense, attempt to substitute for normal probability distributions in a simple but inconsistent with statistical analysis manner, which can be misleading (Section 4). It is clear that this type of deterministic worst-case robustness leads to extremely conservative decisions.

3 Decision problems under uncertainty.

Statistical decision theory deals with situations in which the model of uncertainty and the optimal solution are defined by a sampling model with an unknown vector of "true parameters" θ^*. Vector θ^* defines the desirable optimal solution, its performance can be observed from the sampling model and the problem is to recover θ^* from these data. Potential estimates of θ^* define feasible solutions x of the statistical decision problem. It is essential that x does not affect the sampling model so that the optimality and robustness of solutions can be evaluated by posterior distance (1).

The general problems of decision making under uncertainty deal with fundamentally different situations. The model of uncertainty, feasible solutions, and performance of the optimal solution are not given and all of these have to be characterized from the context of the decision making situation, e.g., socio-economic, technological, environmental, and risk considerations. As there is no information on true optimal performance, robustness cannot be also characterized by a distance from observable true optimal performance. Therefore, the general decision problems, as the following Sections illustrate, may have rather different facets of robustness.

3.1 Expected utility maximization.

Standard policy analysis, as a rule, uses a utility (disutility) maximization (minimization) model for the evaluation of desirable decisions. In the presence of uncertainty, any related decision x results in multiple outcomes characterized by functions $g_1(x,\omega)$, ..., $g_K(x,\omega)$ such as costs, benefits, damages, and risks, as well as indicators of fairness, equity, and environmental impacts. They depend on x, $x \in R^n$ and uncertainty from a set of admissible scenarios Ω, $\omega \in \Omega$.

A given decision x for different scenarios ω may have rather contradictory outcomes. In 1738 the mathematician Daniel Bernoulli introduced the concept of expected utility maximization as a rule for choosing decisions under multiple outcomes. It is assumed that all outcomes $g_1(x,\omega)$, ..., $g_K(x,\omega)$ can be summarized in a single index of preferability $q(x,\omega)$, say, a monetary payoff. The standard expected utility model suggests that the choice of decision x maximizing an expected utility function $U(x) = Eu(q(x,\omega)) = \int u(q(x,\omega))P(d\omega)$, where $u(\cdot)$ is a utility associated with an aggregate outcome $q(x,\omega)$. The shape of u defines attitudes to risks. This model presupposes that, in addition to the knowledge of Ω, one can rank the alternative scenarios ω according to weights - objective or subjective probability measure P. The use of a probability measure as a degree of belief was formalized by Ramsey (1926). Savage (1954) published a thorough treatment of expected utility maximization based on subjective probability as a degree of belief (see discussion in [23]). As a result of this work the use of probability measure became a standard approach for modeling uncertainty by using "hard" observations and soft public and expert opinions in a consistent way within a single model.

3.2 Stochastic optimization (STO) model.

The shortcomings of the expected utility maximization model are well known. Generally speaking, it is practically impossible to find a utility function that enables the aggregation of various attributes in one preferability index, including attitudes to different risks, the distributional aspects of gains and losses, the rights of future generations, and responsibilities for environmental protection. It is natural that, for complex problems, nonsubstitutable indicators should exist that have to be controlled separately in the same way as indicators of, say, health (e.g., temperature and blood pressure). Moreover, it is often practically impossible to identify exactly subjective (and objective) probability as a degree of beliefs. Most people cannot clearly distinguish [44] between probability ranging roughly from 0.3 to 0.7. Decision analysis often has to rely [9] on imprecise statements, for example, that event e_1 is more probable than event e_2, or that the probability p_1, p_2 of event e_1 or of event e_2 is greater than 50 percent and less than 90 percent. These statements may be represented by inequalities such as $p_1 \geq p_2$, $0.5 \leq p_1 + p_2 \leq 0.9$. A number of models with imprecise probabilities have been suggested (see, e.g., [43]) and these models were later integrated into classical probability theory.

The expected utility model is a specific case of STO [6], [19], [32], [39], [45] models that use various performance indicators $f_i(x, \omega)$, $i = 1, ..., m$, one of which can be the expected utility (disutility). These indicators depend on outcomes $g_k(x, \omega)$, $k = 1, ..., K$, on x and ω, i.e., $f_i(x, \omega) := q_i(g_1, ..., g_k, x, \omega)$. A rather general STO problem is formulated as the maximization (minimization) of the expectation function $F_0(x) = Ef_0(x, \omega) = \int f_0(x, \omega)P(d\theta)$, subject to constraints $F_i(x) = Ef_i(x, \omega) = \int f_i(x, \omega)P(d\theta) \geq 0$, $i = 1, ..., m$. The choice of proper indicators $f_i(x, \omega)$ and outcomes $g_k(x, \omega)$, $k = 1, ..., K$, is essential for the robustness of x. Globally or regionally aggregated outcomes are less uncertain but they may not reveal potentially dramatic heterogeneities induced by global changes on individuals, governments, and the environment. For instance, an aggregate income or growth indicators may not reveal an alarming gap between poor and rich regions, which may cause future instabilities. By choosing appropriate functions $g_k(x, \omega)$ and $f_i(x, \omega)$, STO models allow in a natural and flexible way to represent various risks, spatial, social, and temporal heterogeneities, and the sequential resolution of uncertainty in time. Often, as in Example 1, $f_i(x, \omega)$ are analytically intractable, nonsmooth, and even discontinuous functions [16], and probability measure P is unknown, or only partially known, and may depend on x (Section 5, 6), which is essential for modeling endogenous catastrophic risks and the effects of increasing returns leading to concentrations of values in risk prone areas. Moreover, decisions x can be composed of anticipative ex-ante and adaptive ex-post components, which allows to model dynamic decision making processes with flexible adaptive adjustments of anticipative decisions when new information is revealed. The main challenge confronted by STO theory is that it is practically impossible in general to evaluate exact values of $F_i(x)$, $i = 0, 1, ..., m$, see, e.g., Example 1. As "deterministic" is a degenerated case of "stochastic", STO methods allow to deal with problems which are not solved by standard deterministic methods.

Example 1. Pollution control. A common feature of most models used in designing pollution-control policies [1] is the use of transfer coefficients a_{ij} that link the amount of pollution x_j emitted by source j to the pollution concentrations $g_i(x, \omega)$ at the receptor location i as $g_i(x, \omega) = \sum_{j=1}^{n} a_{ij}x_j$, $i = 0, 1, ..., m$. The coefficients are often computed with Gaussian type diffusion equations. These equations are solved over all possible meteorological conditions, and the outputs are then weighted by the frequencies of meteorological inputs over a given time interval, yielding average transfer coefficients a_{ij}. Deterministic models ascertain cost-effective emission strategies x_j, $j = 1, ..., n$ subject to achieving exogenously specified environmental goals, such as ambient average standard b_i at receptors $i = 1, ..., m$. These models can be improved by the inclusion of safety constraints that account for the random nature of coefficients a_{ij} and ambient standards b_i to reduce impacts of extreme events:

$$F_i(x) = Prob[\sum_{j=1}^{n} a_{ij}x_j \leq b_i] \geq p_i, i = 1, ...m, \tag{2}$$

namely, the probability that the deposition level in each receptor (region, grid, or country) i will not exceed uncertain critical load (threshold) b_i at a given probability (acceptable safety level) p_i.

Remark 1. The constraints (2) are known as chance constraints [6], [19], [32], [39]. They can be written in the form of the standard STO model with discontinuous functions: $f_j(x,\omega) = 1-p_i$ if $\sum_{j=1}^{n} a_{ij}x_j - b_i \leq 0$ and $f_j(x,\omega) = -p_i$, otherwise. If $p_i = 1$, $i = 1, ..., m$, the constraints (2) are reduced to constraints of deterministic robustness (Section 2.5).

The main computational complexity confronted by STO methods is the lack of explicit analytical formulas for goal functions $F_i(x)$, $i = 0, 1, , m$. For example, consider constraints (2). If there is a finite number of possible scenarios $\omega = (a_{ij}, b_i, i = \overline{1,m}, j = \overline{1,n})$ reflecting, say, prevailing weather conditions, then $F_i(x)$ are piecewise constant functions, i.e., gradients of $F_i(x)$ are 0 almost everywhere. Hence, the straightforward conventional optimization methods cannot be used.

Ignorance of risks defined by constraints (2) may cause irreversible catastrophic events. Although an average daily concentration of a toxicant in a lake is far below a vital threshold, real concentrations may exceed this threshold for only a few minutes and yet be enough to kill off fish. Constraints of the type (2) are important for the regulation of stability in the insurance industry, known as the insolvency constraints. The safety regulation of nuclear reactors requires $p_i = 1 - 10^{-7}$, i.e., a major failure occurs on average only once in 10^7 years. Stochastic models do not, however, exclude the possibility that a disaster may occur next year.

4 Uncertainty modeling.

As discussed in Section 3, traditional statistical decision theory deals with situations where the model of uncertainty and the performance of optimal solution are given by a sampling model. In general decision problems the uncertainty, decisions and interactions among them have to be characterized from the context of the decision making situation.

Any relevant decision in the presence of essential uncertainty leads to multiple outcomes with potentially positive and negative consequences. A trade-off between them has to be properly evaluated which represents a challenging counterintuitive task. This is often used as a reason to ignore uncertainty with a plea for simple models or for postponing decisions until full information is available. The purpose of this section is to provide important motivations for the appropriate treatment of uncertainty.

4.1 Adaptive control.

Adaptive feedback control is often suggested as a way of dealing with the "unforeseen surprises" (ignored uncertainties) of deterministic models. A feedback control strategy depends on the current state of the system; therefore, when the state is perturbed, the strategy proceeds the control from a new state. The main issue in this approach is the inherent uncertainty, the delayed responses of socio-economic and environmental systems, and irreversibilities. The real consequences of decisions may remain invisible for long periods of time; thus, purely adaptive deterministic approaches can be compared to driving a car in the mountains on a foggy day facing backwards.

4.2 Simple models.

As the assumption of deterministic models about exact input data is often unrealistic, a number of simple models of uncertainty have been used. Simple models that provide an impression of explicit treatment of uncertainty may, in fact, produce misleading or wrong conclusions. One of the most popular ideas is to model uncertainty by a finite number of scenarios or states of the world. All members (agents) of the society know these states and their probabilities, i.e., they know "what-and-when" happens and can thus easily design compensation schemes or securities to spread risks around the world. As Arrow admits [3], catastrophes do not exist in such models (see also discussion in [7], [13]). Moreover, any of these scenarios in reality has the probability of 0.

4.3 Mean-variance analysis.

This analysis substitutes real distributions by normal probability distributions. The following example illustrates its main danger. As discussed in [26], trajectories of the average annual atmospheric CO_2 changes were obtained from various monitoring stations. Analysts suggested characterizing the variability of these trajectories by calculating the sample mean, the standard deviation, and associated 95 percent confidence interval, which, in fact, contains only 13 percent of the observable CO_2 changes. The reason for this is that the histogram of indicated changes has a multimodal character that is fundamentally different from the normal distribution defined by the calculated sample mean and standard deviation. Multimodal distributions are typically used for characterizing the beliefs (opinions) of different political parties or movements and heterogeneities induced by catastrophic events (see Fig. 5, 7 in [13]). In finance, a distribution of portfolio returns can be multimodal due to the contribution of different assets and asset classes.

4.4 Using average values.

Average income, growth, daily pollutant concentration, average losses, expected utility, or expected returns may have a rather misleading character.

The projected global mean temperature changes fall within the difference between the average temperature of cities and their surrounding rural areas. Therefore, global climate change impacts can be properly evaluated only in terms of local temperature variability and related extreme events, in particular, heat waves, floods, droughts, windstorms, diseases, and sea level rise. The proper treatment of indicators with nonnormal, especially multimodal distributions requires special attention. The mean value of a multimodal distribution can be even outside the support of a distribution (the set of admissible values). Still, this value can be reasonably interpreted in the case of frequent repetitive observations. Subjective multimodal probability distributions and rare extreme events call for the use of quantiles, e.g., the median. Unfortunately, this destroys the additive structure and concavity (convexity) of standard models, as (in contrast to the average value) $median \sum_l v_l \neq \sum_l median(v_l)$ for random variables v_l. As a result this makes the applicability of well-known decomposition schemes and optimization methods problematic. Sections 5.2, 5.3 indicate a promising approach for dealing with arising problems.

Example 2. Optimal control problems. Discrete-time optimal control can be viewed as a specific case of STO models. In this case, x is composed of state variables $z(t)$, and control variables $u(t)$, that is,
$x = \{z(t), u(t), t = 0, 1, ..., T\}$, where T is a given time horizon. Functions $f_i(x, \omega)$ are additive: $f_i(x, \omega) = \sum_{t=1}^{T} g_i(z(t), u(t), \omega_t, t)$, where ω_t is a stochastic disturbance at time t. Therefore, the use of $median(f_i(x, \omega))$ destroys the additive structure of optimal control problems essentially utilized in the Pontriagin's Maximim Principle and Bellman's recursive equations.

4.5 Deterministic versus stochastic optimization.

Deterministic decision problems are formulated in two steps. First of all, statistical procedures are used to estimate average values $\overline{\omega}$ of input data ω. After this intermediate task is performed, the deterministic problem with goal functions $f_i(x, \overline{\omega})$, $i = 0, 1, ..., m$ is solved. The use of $\overline{\omega}$ for multimode distributions orients decision analysis even on inadmissible scenarios. As well as for nonlinear $f_i(x, \omega)$, $Ef_i(x, \omega) \neq f_i(x, \overline{\omega})$. For example, if ω is uniformly distributed on $[-1, 1]$, then $\overline{\omega} = 0$ and $E(\omega x)^2 > (\overline{\omega} x)^2 = 0$.

STO methods deal directly with the variability of $f_i(x, \omega)$ affected by the variability of ω and decisions x, i.e., they deal simultaneously with uncertainty and decision analysis. Some decisions x can considerably reduce the variability of indicators $f_i(x, \omega)$, despite significant variability of ω, e.g., decisions $x_1 = 0$, $x_2 = 0$ for function $\omega_1 x_1 + \omega_2 x_2$. Therefore, STO models can significantly reduce requirements on data quality in contrast to disconnected from decisions standard uncertainty analysis (see also Section 4.6).

The use of average values often smoothes the problem, but this may lead to wrong conclusions. The following simple model with abrupt changes shows that the use of average characteristics converts this model to a smooth and even linear deterministic version. Combined with sensitivity analysis, the resulting linear deterministic model is not able to detect abrupt changes: it plays a misleading role and can easily provoke an environmental collapse.

Example 3. Abrupt changes. Global changes with possible dramatic interactions among humans, nature and technology call for nonsmooth models. Nonsmooth and discontinuous processes are typical for systems undergoing structural changes and developments. In risk management, the possibility of an abrupt change is, by its very nature, present in the problem. The concept of nonsmooth and abrupt change is emphasized in the study of environmental and anthropogenic systems by such notions as critical load, surprise, and time bomb phenomena [1], [8], [16]. There are a number of methodological challenges involved in the policy analysis of nonsmooth processes. Traditional local or marginal analysis cannot be used because continuous derivatives do not exist, i.e., a nonsmooth, even deterministic, system cannot be predicted (in contrast to classical smooth systems) outside an arbitrary small neighborhood of local points.

The concentration of a pollutant $r_t = r_0 - xt + \sum_{k=1}^{N(t)} e_k$, where $\{e_k\}$ is a sequence of emissions from extreme episodes in interval $[0, t]$, $N(t)$, $t \geq 0$, is a counting process for the number of episodes in $[0, t]$, x is a rate of emission reduction, and r_0 is an initial concentration. The rate x pushes r_t down, whereas the random flow of emissions pushes r_t up. The main problem is to reduce the probability of a catastrophe associated with crossing a vital threshold ρ by r_t, $r_t > \rho$. Assume that $\{e_k\}$ are independent, identically distributed random variables with mean value $\overline{e}$, $N(t)$ is a Poisson process with intensity α, $EN(t) = \alpha t$ and $\{e_k\}$, $N(t)$ are independent. Then, the expected concentration $\overline{r_t} = r_0 + (\alpha \overline{e} - x)t$, that is, complex random jumping process r_t, is simply replaced by a linear function that decreases in time for $x > \alpha \overline{e}$. The strong law of large numbers for random sums of random variables e_k implies that $xt - \sum_{k=1}^{N(t)} e_k \sim t(x - \alpha \overline{e})$ for large t with probability 1. Thus, deterministic model $\overline{r_t}$ suggests, that if x slightly exceeds the average emission rate $\alpha \overline{e}$, then $\overline{r_t}$ decreases in time, which is the wrong conclusion. This is true only if a catastrophe does not occur before time t. The sensitivity analysis of the linear deterministic model $\overline{r_t}$ under different scenarios for α and $\overline{e}$ produces the same trivial conclusions that robust x has to slightly exceed $\alpha \overline{e}$.

4.6 Probabilistic and stochastic models.

There are two fundamental approaches to modeling uncertainty in probability theory, namely, probabilistic and stochastic models. Probabilistic models attempt to characterize processes completely and explicitly in terms of probability distributions or some of their characteristics. If one can evaluate explicitly multidimentional integrals $F_i(x) = Ef_i(x,\omega) = \int_\Omega f_i(x,\omega)P(d\omega)$, then the STO problem is reduced to a standard deterministic optimization model. Even the simplest situations illustrate difficulties. Thus, for two random variables ω_1, ω_2 with known probability distribution functions, the evaluation of probability distribution $\omega_1 + \omega_2$ is already an analytically intractable (in general) task requiring the evaluation of an integral. In addition, the distribution of $f_i(x,\omega)$, say, $\omega_1 x_1 + \omega_2 x_2$ significantly depends on x, e.g., compare $x_1 = 0$, $x_2 = 1$ and $x_1 = 1$, $x_2 = 0$. Exponential increase of computations occurs when one uses probability trees, transition probabilities, and variance-covariance matrices to represent the dynamics of uncertainties. The number of states of even the simplest discrete event systems (see, e.g., [16]) can be too large for explicit representations of them by matrices of transition probabilities. The computational "explosion" of probabilistic models, similar to the well-known "curse of dimensionality" of Bellman's equations, restricts their practical applicability for large scale global change problems.

Stochastic models deal directly with random variables $f_i(x,\omega)$ without an exact evaluation of $F_i(x)$. In combination with fast Monte Carlo simulations, some of the STO methods lead only to a linear increase of computations w.r.t. uncertain variables ω. In this case, goal functions are characterized by random laws (rules) and random processes (e.g., stochastic differential equations) rather than by transition probabilities, variance- covariance matrices, and partial differential equations. In fact, fast Monte Carlo procedures (Example 7) combine probabilistic and stochastic submodels.

5 Robust stochastic optimization.

Although STO models allow to represent interdependencies among decisions, uncertainties and risks, yet inappropriate treatment of the variability of indicators $f_i(x,\omega)$ can be rather misleading for achieving desirable robustness.

5.1 Portfolio selection.

The Nobel prize laureate Markowitz [31] proposed the following mean-variance approach for designing robust portfolios of financial assets (and others, e.g., portfolios of technologies). Assume that $\overline{\omega_j}$ is the expected value of random returns ω_j from asset j, $j = 1, ..., n$, and x_j is a fraction of this asset in the portfolio, $\sum_{j=1}^n x_j = 1$, $x_j \geq 0$, $j = 1, ..., n$. The maximization of expected return $r(x) = \sum_{j=1}^n \overline{\omega} x_j$ from a portfolio $x = (x_1, ..., x_n)$ yields a trivial nonrobust solution: to invest all capital in the asset with the maximal expected return. The main idea [31] to achieve diversified robust portfolio is to consider a trade-off between expected returns and their variability characterized by the variance of returns $Var\rho(x,\omega)$, i.e., to maximize $r(x) - \mu Var\rho(x,\omega)$, $\rho(x,\omega) = \sum_{j=1}^n \omega_j x_j$, where μ is a trade-off (risk) parameter. Let us note that this approach requires that only returns from portfolio $\sum_{j=1}^n \omega_j x_j$ have close to normal distribution, but not returns ω_j .

Remark 2. The most important concerns in the case of more general portfolio selection problems are those related to the overestimation of real returns $\rho(x,\omega)$ by maximizing expected returns $r(x)$, i.e., when $\rho(x,\omega) < r(x)$. This calls for the maximization of a trade-off between expected returns and the risk of overestimation:

$r(x) + \mu E \min\{0, \rho(x,\omega) - r(x)\}$. It is easy to see that when the distribution of random returns $\rho(x,\omega)$ is normal, then the maximization of this function is equivalent to the maximization of the mean-variance criterion, as the absolute values of asymmetric risk function $E \min\{0, \rho(x,\omega) - r(x)\}$ are constant multiples of the standard deviation. Unfortunately, for nonlinear concave function $r(x)$ the mean-variance approach leads to nonconcave optimization. The next section maneuvers this obstacle for rather general optimization problems.

5.2 Robust utility maximization.

Consider the maximization of utility function $U(x) = Eu(q(x,\omega))$, (e.g., returns $r(x)$). If the distribution of random outcome $u(q(x,\omega))$ is not normal, for example, when the policy analysis involves the polarized beliefs of different communities, then, instead of $U(x)$ we can use a quantile $U_p(x)$ of $u(q(x,\omega))$ defined as maximal v such that $Prob[u(q(x,\omega)) \leq v] \leq p$, for $0 < p < 1$. The robust utility maximization problem can be formulated as the maximization of an adjusted to risk utility function $U_p(x) + \mu E \min\{0, u(q(x,\omega)) - U_p(x))\}$, which is not a concave function. As *Remark 2* indicates, for normal distributions and $p = 1/2$, this is equivalent to the mean-variance approach. Similar to Example 4, Section 5.4, one can conclude that the formulated problem is equivalent to the following concave STO optimization problem: maximize w.r.t. (x,z) function $\varphi(x,z) = z + \mu E \min\{0, \beta - z\}$, $\beta = u(q(x,\omega))$, $\mu = 1/p$.

Remark 3. This important fact can be seen from the following simple observations (see also Example 4): $\int_0^z (p - Prob[\beta \leq v])dv = pz + E \min\{0, \beta - z\}$ for a random variable β with density. Let us also notice that for $\mu = 1/p$ we have $U_p(x) + \mu E \min\{0, u(q(x,u)) - U_p(x)\} = (1/p) \int_{u(q(x,\omega) \leq U_p(x)} U(q(x,\omega)dP$, i.e., the adjusted to risk utility function equals to the so-called expected shortfall (see, e.g., [11], [41]).

5.3 General STO model.

Similarly, a robust STO model can be written in the form: maximize w.r.t. (x,z) function $z_0 + \mu_0 E \min\{0, f_0(x,\omega) - z_0\}$ subject to $z_i + \mu_i E \min\{0, f_i(x,\omega) - z_i\} \geq 0$, $i = 1, ..., m$, where μ_i are weights. Components z_i^*, $i = 0, 1, ..., m$, of optimal solution (x^*, z^*) are quantiles of $f_i(x^*, \omega)$. The proof follows from the positivity of the Lagrange multipliers and *Remark 3.* Depending on the case, the robust model can also be formulated by using safety (Example 1) constraints $Prob[f_i(x,\omega) \geq 0] \geq p_i$ in combination, say, with constraints $Ef_i(x,\omega) + \mu_i E \min\{0, f_i(x,\omega)\} \geq 0$, $i = 1, ..., m$ and other possible options [35].

5.4 Flexibility of robust strategies.

The standard expected utility maximization model suggests two types of decisions in the response to uncertainty: either risk averse or risk prone decisions. These two options also dominate the climate change policy debates [33], [36], emphasizing either ex-ante anticipative emission reduction programs or ex-post adaptation to climate changes when full information becomes available. Clearly, a robust policy must include both options, i.e., the robust strategy must be flexible enough to allow for later adjustments of earlier decisions. The so-called (two-stage and multistage) recourse models of stochastic optimization [6], [19], [45] incorporate both fundamental ideas of anticipation and adaptation within a single model and allow for a trade- off between long-term anticipatory strategies and related short-term adaptive adjustments. Therefore, the adaptive capacity can be properly designed ex-ante say, through emergency plans and insurance arrangements. The following example shows that the explicit incorporation of ex-ante and ex-post decisions induces risk aversion measures that cannot, in general, be imposed exogenously by a standard utility function.

Example 4. Mitigation versus adaptation: CVaR Risk measure. A stylized static model of a climate stabilization problem [37] can be formulated as follows: let x denote an amount of emission reduction and let a random variable β denote an uncertain critical level of required emission reduction. Ex-ante emission reductions $x \geq 0$ with costs cx may underestimate β, $x < \beta$. It generates a linear total adaptation cost $az + dy$, where y is an ex-post adaptation, $y \leq z$ with cost dy; z is an ex-ante developed adaptive capacity with cost az.

To illustrate the main idea, let us assume that ex-post adaptive capacity is unlimited, $z = \infty$, and $c < d$. A two-stage stochastic optimization model is formulated as the minimization of expected total cost $cx + dEy$ subject to the constraint $x + y \geq \beta$. This problem is equivalent to the minimization of function $F(x) = cx + E \min \{dy | x + y \geq \beta\}$ or $F(x) = cx + dE \max\{0, \beta - x\}$, which is a simple stochastic minimax problem. Optimality conditions for these types of STO minimax problems show (see, e.g., [15], [16], [19], pp. 107, 416, [41] see also *Remark 3*) that the optimal ex-ante solution is the critical quantile $x^* = \beta_p$ satisfying the safety constraint $Prob[x \geq \beta] \geq p$ for $p = 1 - c/d$. This is a remarkable result: highly non-linear and even often discontinuous safety or chance constraint of type (2) is derived (justified) from an explicit introduction of ex-post second stage decisions y. Although the two stage model is linear in variables (x, y), the ex-post decisions y induce strong risk aversion among ex-ante decisions characterized by the critical quantile β_p.

Remark 4. If $c/d < 1$, then $x^* > 0$, i.e., it calls for coexistence of ex-ante and ex-post decisions. The optimal value $F(x^*) = dE\beta I(\beta > x^*)$, where $I(\cdot)$ is the indicator function. Again, according to Remark 3, this is the expected shortfall or Conditional Value-at-Risk (CVaR) risk measure [11], [41].

Remark 5. In more general two-stage models [13], [37], the risk aversion of an ex-ante decision is not necessarily induced in the form of the critical quantile and CVaR risk measure. Despite this, the structure of robust policy remains the same. Only partial commitments are made ex-ante whereas other options are kept open until new information is revealed. In a sense, such flexible decisions incorporate both risk-averse and risk-prone components according to different "slices" of risks.

5.5 Uncertain probability distributions.

Models of Section 3 assume that $P(d\omega)$ is known exactly. However, only some of its characteristics may be known. The elicited class $\mathcal{P}$ for admissible P is often given by constraints $\int \varphi_k(\omega)P(d\omega) \geq 0$, $k = \overline{1, K}$, $\int P(d\omega) = 1$; for example, constraints on joint moments $c_{s_1,\dots,s_l} \leq \int \omega_1^{s_1} \dots \omega_l^{s_l} P(d\omega) \leq C_{s_1,\dots,s_l}$, where $c_{s_1,\dots,s_l}$, $C_{s_1,\dots,s_l}$ are given constants. The robust STO problem can be formulated similar to Section 2.3 as a probabilistic maximin problem: maximize $F_0(x) = \min_{p \in \mathcal{P}} \int f_0(x, \omega)P(d\omega)$ subject to general constraints of Section 3.2. This probabilistic maximin approach was first initiated in STO in [12], [14], [49]. For specific sets $\mathcal{P}$, the solution of the inner minimization problem has a simple analytical form [28], [30]. For example, it is concentrated only in a finite number ([12], [14], [30], and Example 5) of admissible scenarios from Ω. Numerical methods for general problems were developed in [12], [14], [24], [25], [28].

Example 5. Robust stabilization and CVaR. The simple emission stabilization problem is defined (Example 4) by the minimization of

$cx + dE\max\{0, \beta - x\} = z + d\int_z^\infty (\beta - x)P(d\beta)$. A robust CVaR measure can be defined by minimization $cx + d\max_{p\in\mathcal{P}} \int_x^\infty (\beta - x)P(d\beta)$. To illustrate this possibility, suppose that β is a scalar random variable, $\Omega = [a, b]$, and an additional condition that defines the class $\mathcal{P}$ is $E\beta = \mu$. It is easy to see that the worst-case distribution is concentrated only in points a, b, with the probability mass $p(a) = \frac{b-\mu}{b-a}$, $p(b) = \frac{\mu-a}{b-a}$. Hence, the robust model is reduced to replacing the set of all admissible scenarios Ω by only two extreme scenarios a and b with probabilities $p(a)$, $p(b)$.

Probabilistic maximin robustness may not be sufficient to properly address the effects of extreme events (Section 8.2). A more general approach would be the combination of a probabilistic and a stochastic maximin model with $F_0(x) = \min_{p\in\mathcal{P}} E\min_{z\in Z} f_0(x, y, z, \xi)$, where ω is represented by variables y, z, ξ, $\omega = (y, z, \xi)$. Z is a set of variables z which are there to take into account potential extreme random scenarios, as in the extremal value theory [11]; the x variables are themselves decision variables; the y, $y \in Y$ variables correspond to uncertainty ranked by an objective or subjective probability measure P from $\mathcal{P}$; and ξ variables are ranked by a fixed probability measure as in the basic STO models. Thus in this model the worst case situation is evaluated with respect to the worst-case distribution for some uncertain variables y, whereas for other uncertain variables z it is evaluated from potential extreme random scenarios. In particular, this class of models includes purely stochastic maximin models with $F_0(x) = E\min_{y\in Y} f_0(x, y, \xi)$ as well as models with $F_0(x) = \min_{y\in Y} Ef(x, y, \xi)$ combining the worst-case and the Bayesian approaches of Sections 2.2, 2.5 (see also discussion in [18], [19], pp. 105-106).

6 Temporal, spatial and social heterogeneities.

The significance of extreme events arguments in global climate changes has been summarized in [47] as follows: Impacts accrue ... not so much from slow fluctuations in the mean, but from the tails of the distributions, from extreme events. Catastrophes do not occur on average with average patterns. They occur as "spikes" in space and time. In other words, the distributional aspects, i.e., temporal and spatial distributions of values and risks are key issues to capture the main sources of vulnerability for designing robust policies.

6.1 Temporal heterogeneity.

Extreme events are usually characterized by their expected arrival time, for example, as a 1000-year flood, that is, an event that occurs on average once in every 1000 years. Accordingly, these events are often ignored as they are evaluated as improbable during a human lifetime. In fact, a 1000-year flood may occur next year. For example, floods across Central Europe in 2002 were classified as 1000-, 500-, 250-, and 100-year events. Another tendency is to evaluate potential extreme impacts by using so-called annualization, i.e., by spreading losses from a potential, say, 30-year catastrophe, equally over 30 years. In this case, roughly speaking, a potential 30-year crash of an airplane is evaluated as a sequence of independent annual crashes: one wheel in the first year, another wheel in the second year, and so on, until the final crash of the navigation system in the 30th year. The main conclusion from this type of deterministic analysis is that catastrophes do not exist. Section 8.1 introduces the notion of stopping time and related new approaches to discounting that allow for properly addressing the temporal variability of extreme events.

6.2 Spatial and social heterogeneity.

A similar common tendency is the ignorance of real spatial patterns of catastrophes. A general approach is to use so-called hazard maps, i.e., maps showing catastrophe patterns that will never be observed as a result of a real episode, as a map is the average image of all possible patterns that may follow catastrophic events. Accordingly, social losses in affected regions are evaluated as the sum of individual losses computed on a location-by-location rather than pattern-by-pattern basis w.r.t. joint probability distributions. This highly underestimates the real impacts of catastrophes, as the following simple example shows.

Example 6. Social and individual losses. In a sense, this example shows that $100 \gg \overbrace{1+1+\ldots+1}^{100}$. Assume that each of 100 locations has an asset of the same type. An extreme event destroys all of them at once with probability 1/100. Consider also a situation without the extreme event, but with each asset still being destroyed independently with the same probability 1/100. From an individual point of view, these two situations are identical: an asset is destroyed with probability 1/100, i.e., individual losses are the same. Collective (social) losses are dramatically different. In the first case 100 assets are destroyed with probability 1/100, whereas in the second case 100 assets are destroyed only with probability 100^{-100}, which is practically 0. This example also illustrates the potential exponential growth of vulnerability from increasing network-interdependencies.

6.3 Downscaling, upscaling and catastrophe modeling.

So-called downscaling (see discussion in [5], [22]) and catastrophe modeling [46] are becoming increasingly important for estimating spatio-temporal vulnerability and catastrophic impacts. The designing of a catastrophe model is a multidisciplinary task requiring the joint efforts of environmentalists, physicists, economists, engineers and mathematicians. To characterize "unknown" catastrophic risks, that is, risks with the lack of historical data and large spatial and social impacts, one should at least characterize the random patterns of possible disasters, their geographical locations, and their timing. One should also design a map of values and characterize the vulnerabilities of buildings, constructions, infrastructure, and activities. Catastrophe models allow to derive histograms of mutually dependent losses for a single location, a particular hazard-prone zone, a country, or worldwide from fast Monte Carlo simulations rather than real observations [13], [46].

The development of catastrophe models can be considered as a key risk management policy providing information for decision analysis in the absence of historical observations, in particular, on potential extreme events that have never occurred in the past. This raises new estimation problems. Traditional statistical methods are based on the ability to obtain observations from unknown true probability distributions, whereas new problems require information to be recovered from only partially observable or even unobservable variables. Rich data may exist on occurences of natural disasters, incomes, or production values on global and national levels. Downscaling and upscaling methods in this case must - by using all available objective and subjective information - make plausible evaluations of local processes consistent with available global data, as well as, conversely, with global implications emerging from local data and tendencies.

7 STO methods for robust solutions.

7.1 Scenario analysis.

Outcomes of Monte Carlo simulations for a STO model are random sample functions $f_0(x,\omega)$, $f_1(x,\omega)$, ..., $f_m(x,\omega)$, that depend on the simulation run ω and a given vector of decisions x. Therefore, for a given x, outcomes vary at random from one simulation to another. The estimation of their mean values, variances, and other moments or histograms is time consuming in the presence of rare extreme events that require developments of specific fast Monte Carlo-type sampling procedures. Moreover, a change in policy variables x affects the probabilistic characteristics of outcomes and requires a new sequence of Monte Carlo simulations to estimate their new values. If functions $f_i(x,\omega)$, $i = 0, 1, ..., m$, have well defined analytical structure with respect to x for each simulated ω, then the following scenario analysis is often used. The Monte Carlo simulations generate scenarios ω^1, ω^2, ..., ω^N for each of which optimal solutions $x(\omega^1)$, $x(\omega^2)$, ..., $x(\omega^N)$ of the deterministic optimization model are calculated. Any of these solutions calculated for one scenario may not be feasible for other scenarios. The number of possible combinations of potential scenarios ω and decisions increases exponentially. Thus, with only 10 feasible decisions, for instance, levels of emission reductions in a given region, 10 regions and 10 possible scenarios for all of them, the number of "what-if" combinations is 10^{11}. The straightforward evaluation of these alternatives would require more than 100 years if a single evaluation takes only a second. Besides, the probability of each scenario ω^l, $l = 1, ..., K$, is in general, equal to 0. Therefore, the choice of final robust decisions is unclear and is not explicitly addressed.

7.2 Sample-mean approximations.

STO models of Sections 3.2, 5 are able to explicitly characterize robustness by using proper indicators of different risks, flexible decisions and various equity and fairness constraints as goals of desirable policy. The main challenge is to design a search procedure that enables to find policy decisions specified by these goals. STO methods, in particular, adaptive Monte Carlo (AMC) optimization methods [13], avoid exact evaluations of all feasible alternatives. The problem confronted by STO methods is to estimate the maximum $F_0(x^*)$ of $F_0(x)$ subject to constraints $F_i(x) \geq 0$, $F_i(x) = Ef_i(x,\omega)$, $i = 1, ..., m$, by making use of only random outcomes from simulations $f_i(x,\omega)$, $i = 0, 1, ..., m$. Standard Monte Carlo methods can be regarded as estimating the value of multidimensional integrals $F_i(x) = \int f_i(x,\omega)P(d\omega)$, $i = 0, 1, ..., m$, for fixed x. In particular, this can be done by using a sample mean $F_i^N = 1/N \sum_{k=1}^{N} f_i(x,\omega^k)$. If functions $f_i(x,\omega)$ are analytically tractable w.r.t. x, then $F_i^N(x)$ can be used to find an approximate solution of the STO problem, assuming that $F_i^N(x)$ sufficiently approximates $F_i(x)$, $i = 1, ..., m$. Although in this case the original STO model is approximated by a deterministic optimization problem, its solution often requires new deterministic large-scale optimization methods (see, e.g., [6], [19], [32], [39], [45]), as well as the sample size N reduction techniques and fast Monte Carlo simulations. A principle complexity (Sections 5, 8) is that the measure P is often analytically intractable, that it may depend on x as in Section 5.5, and that samples are affected by current x and rare catastrophic events. In this case, in general, only AMC optimization is applicable.

7.3 Adaptive Monte Carlo optimization.

An "Adaptive Monte Carlo" simulation [40] is a technique that makes online use of sampling information to sequentially improve the efficiency of the sampling itself. The notion

"Adaptive Monte Carlo" optimization is used [13], [20] in a rather broad sense, where improvements of the sampling procedure with respect to the variability of estimates may be only a part of the improvements with respect to other goals of robust decisions.

Remark 6. A counterintuitive fact is that the estimation of a robust solution x^* and $F_0(x^*)$ starting from an initial solution x^0 often requires approximately the same (or an even smaller) number of simulations than the estimation of only $F_0(x^0)$ for fixed x^0. This is because of two forces. First of all, robust solutions x^* reduce risks and, hence, the variability of $F_0(x)$; therefore, movements toward $F_0(x^*)$ according to STO methods are themselves a variance-reducing process (see, e.g., numerical calculations in [20]). In contrast, $F_0(x^0)$ may have considerable variability due to the effects of extreme events; therefore, its estimation requires large samples. Secondly, the variance reductions can also be achieved by deliberate switches in the importance sampling.

Example 7. Environmental collapse. Let us illustrate the main idea of fast sample mean approximations and AMC optimization by a modification of Example 3. The concentration of a global pollutant at time t is calculated as $r_t = r_0 + \sum_{t=0}^{t} x_t e_t$, where x_t is the rate of global emission e_t reduction, $0 \leq x_t \leq 1$, and e_0, e_1, ... are random dependent variables. At a random time moment τ, the critical threshold β for r_t is revealed and a collapse occurs when $r_t > \beta$. Assume that β is characterized by a probability distribution $B(z) = Prob[\beta < z]$ and $Prob[\tau = t] = p(1-p)^t$, $t = 0, 1, ...$, where probability p is characterized by a probability distribution in an interval $[p_*, p^*]$. The probability of a collapse $\Psi(x) = E\sum_{t=0}^{\infty} I(\beta < r_t)$, where $I(\beta > r_t) = 1$ or 0 if $\beta > r_t$ or $\beta \leq r_t$, respectively. Equivalently,

$$\Psi(x) = E\sum_{t=0}^{\infty} E[p(1-p)^t] B(r_0 + \sum_{t=0}^{t} x_t e_t) = E\sum_{t=0}^{\tau} EB(r_0 + \sum_{t=0}^{t} x_t e_t). \quad (3)$$

The probabilistic model is described by the analytically intractable function $\Psi(x)$. Moreover, an emission path e_0, e_0, ... is usually generated by solving a global energy/economy model, and e_t is a complex function of an emission reduction policy x. The stochastic model in this example is described by the right hand side of (3) including the process r_t, the probability distribution for τ, and a stochastic generator of uncertainties and dependent emission path, (e.g., using global energy/economy model).

It is possible to use a straightforward Monte Carlo simulation to estimate $\Psi(x)$ for a fixed x. A simulation run s, $s = 1, 2, ...$ consists of sampling $p^s \in [p_*, p^*]$; $\tau = \tau_s$; a path e_t^s, $t = 0, 1, ..., \tau_s$ and β^s. The value $\Psi(x)$ is estimated as $\Psi^N(x) = \sum_{s=1}^{N} I(\beta^s < r^s)/N$. If p^s is a small probability then this straightforward approach requires large N. A stochastic model (3) allows much faster sample mean evaluations of $\Psi(x)$ and fast AMC optimization procedures [17], [20]. Conceptually, AMC optimization involves the following steps. An initial solution x^0 is fixed; p^0, τ^0, e_0^0, e_1^0, ..., $e_{\tau^0}^0$ are simulated. On this basis, a so-called stochastic gradient is calculated allowing for adaptive adjustment of x^0 to x^1. For x^1, a new sample p^1, τ^1, e_0^1, e_1^1, ..., $e_{\tau^1}^1$ is calculated, and x^1 is adaptively adjusted in the same manner as x^2, and so on. It is important that evaluation of robust strategy in this manner proceeds with simulations $s = 1, 2, ...$ without intermediate evaluations of $\Psi^N(x)$. Details of this solution technique for rather general risk processes are discussed in [13], [17]. In parallel with adjustments of solutions x^s, the AMC optimization is able to change the sampling procedure [20] itself (importance sampling).

8 Sensitivity of robust strategies.

Robust strategies for global changes require a proper focus on potential extreme events. As a result, the robust strategy with a small $\varepsilon > 0$ probability of extreme events can be significantly different from the policy that ignores these events by using $\varepsilon = 0$. Formally speaking, this is evident from Section 5.3, when $\varepsilon > 0$ results in shifts of ranges $f_i(x, \omega)$ to include potential catastrophic impacts (say, ranges of required emission reductions β in Example 4) that suddenly disappear for $\varepsilon = 0$. Informally speaking, the explicit introduction of extreme events with $\varepsilon > 0$ requires new sets of feasible decisions, new spatial, temporal, and social dimensions which suddenly disappear for $\varepsilon = 0$. This Section shows that a key issue is the proper treatment of discounting and random time horizons of extreme events.

8.1 Extreme events and discounting.

How can we justify strategies that may possibly turn into benefits over long and uncertain time horizons in the future? For example, how can we justify investment, say, in a flood defense system to cope with foreseen extreme 100-, 250-, 500- and 1000- year floods? A common approach is to discount future costs and benefits using a geometric (exponential) discount factors with the prevailing market interest rate as $V = \sum_{t=0}^{\infty} d_t V_t$, where $d_t = (1+r)^{-t}$, r is a discount rate. An infinite deterministic stream of values V_t, $t = 0, 1, ...$, can represent a cash-flow stream of a long-term investment activity. In economic growth models and integrated assessment models (see, e.g., [36]) the value V_t represents utility $U(x^t)$ of an infinitely living representative agent with consumptions x^t .

The infinite time horizon in V creates an illusion of truly long-term analysis. The choice of discount rate r as a prevailing interest rate within a time horizon of existing financial markets is well established. Uncertainties, especially related to extreme events, challenge the possibility of markets to offer proper rates. The following simple fact shows [21] that the standard discount factors obtained from markets orient policy analysis only on few decades, what precludes to properly address catastrophic impacts.

Let $p = 1 - d$, $d = (1+r)^{-1}$, $q = 1 - p$, and let τ be a random variable with the geometric probability distribution $P[\tau = t] = pq^t$. It is easy to see [21] that

$$\sum_{t=0}^{\infty} d_t V_t = E \sum_{t=0}^{\tau} V_t, \tag{4}$$

where $d_t = d^t$, $t = 0, 1, ...$ This is also true for general discounting $d_t = (1 + r_t)^{-t}$ with increasing positive r_t, where the stopping time τ is defined as $P[\tau \geq t] = d_t$.

That is, the discounted sum can be viewed as an expected value of the undiscounted sum within a random interval $[0, \tau]$. We can think of τ as a random "stopping time" associated with the first occurrence of an extreme stopping time (killing) event. The expected duration of τ, $E\tau = 1/p = 1 + 1/r \approx 1/r$ for small r. Therefore, for the interest rate of 3.5 percent, $r \approx 0.035$, the expected duration is $E\tau \approx 30$ years, i.e., this rate orients the policy analysis on an expected 30-year time horizon with the standard deviation $\sqrt{q}/p$, i.e., approximately another 30 years. The bias in favour of the present in discounting with the rate of 3.5 percent is easily illustrated [38]. For a project with long-run benefits or costs, 1 Euro of benefits or costs in years 50, 100, and 200, has a present value respectively of 0.18, 0.003, and practically 0 Euros. Definitely, this rate has no correspondence with how society has to deal with a 300-year flood with the standard deviation of another 300 years.

Example 8. Catastrophic risk management. The implications of (4) for long-term policy analysis are rather straightforward. It is realistic to assume [38] that typical cash-flow investment in a new nuclear plant has the following average time horizons: without a disaster, the first six years of the stream reflect the costs of constructions and commissioning, followed by 40-years of operating life when the plant is producing positive cash flows and, finally, a 70-year period of expenditure on decommissioning. The flat discount rate of 5 percent, according to (4), orients the analysis on a 20- year time horizon. It is clear that a lower discount rate places more weight on distant costs and benefits. For example, the explicit treatment of a potential 200- year disaster would require a discount rate of at least 0.5 percent instead of 5 percent. Similar examples are investments in mitigations to cope with climate change related extreme events. A rate of 3.5 percent, as is often used in integrated assessment models [36], [48] is definitely not appropriate.

Example 9. Time varying discounting. Multipliers $E[p(1-p)^t]$ in (3) with random p can be viewed as time-varying discount factors. It is easy to see that the asymptotic of these multipliers are dominated by the least-probable extreme events. Indeed, assume that there is only a finite number of scenarios $p_1 < p_2 < ... < p_L$ ranked by probability weights v_1, v_2, ..., v_L. Then $E[p(1-p)^t] = (1-p_1)^t[v_1 p_1 + \sum_{s=2}^{L}(\frac{1-p_s}{1-p_1})^t] \sim v_1 p_1 (1-p_1)^t$.

Therefore, the ignorance of the least-probable extreme events can significantly shrink internal stopping time horizons of evaluation V. Definitely, a given exogenous standard discount rate cannot match the expected time horizons of rather different sets of extreme events. This calls for the explicit introduction of stopping time τ and the use [21] of undiscounted evaluation $V = E\sum_{t=0}^{\tau} V_t$ instead of $V = \sum_{t=0}^{\infty} d_t V_t$. As (4) shows, this approach includes any standard exogenous discounting. Its advantage is the ability to introduce proper endogenous discounting in the presence of catastrophic risks. As decisions affect the occurrence of extreme events (τ) in space, this approach, in fact, is equivalent to using implicit spatio-temporal endogenous discounting dependent on goals and input data of the decision problem, such as the incomes of agents, risks, equity, and fairness constraints. This approach allows also to treat distributional aspects by using distributions of random sum $\sum_{t=0}^{\tau} v_t$, $V_t = Ev_t$, e.g., its quantiles instead of mean value.

8.2 Stopping time and stochastic minimax.

As Section 8.1 shows, the concept of stopping time allows to orient the analysis on the least-probable and the most destructive (killing) extreme events. There are strong connections [17] between the stopping time- and stochastic maximin type-problems defined in Section 5.5 that can be used for designing optimization methods.

The stopping time is often associated with the likelihood of some processes crossing "vital" thresholds. Consider a random process $R_t(x)$ and the threshold defined by a random β. Let us define the stopping time τ as the first time moment t when $R_t(x)$ is above β, that is,
$\tau(x) = \max\{t \in [0,T] : R_s(x) \le \beta, 0 \le s \le t\}$. For example, climate change mitigations x deal with preventing the global temperature, say, R_t, from crossing its critical level β. In this case, the safety constraint can be defined by probability $Prob[\tau(x) \ge T]$, where T is a given horizon. Explicit analytical evaluation of this probability is practically impossible even for the simplest insurance risk processes [11]. This precludes the use of standard optimization methods. A promising idea is to use connections with stochastic minimax problems (see, e.g., [17]. Assume that r_t and β are one dimensional random variables, β is independent of r_t, $H(y) = Prob[\beta \ge y]$, and the performance indicators of the general STO problem depend on t, $f_i(t, x, \omega)$, $i = 0, 1, ..., m$. The robustness can be defined as in

[13] by functions $Ef_i(t, x, \omega)$ at $t = \tau(x)$, $F_i(x) = Ef_i(\tau(x, \omega), x, \omega)$. Functions $F_i(x)$ can be written [17] as $F_i(x) = E\sum_{t=0}^{T} f_i(t, x, \omega)H(\max_{0\leq s\leq t} R_s(x, \omega))$, i.e., a stopping time problem with implicit and, in general, discontinuous random function $\tau(x)$ is equivalently transformed into a stochastic minimax problem that can be solved by different methods [16].

9 Concluding Remarks.

In the absence of sufficient information, models play a key role in comparative analysis of alternative solutions for designing robust policies. Any policy analysis focuses attention on situations where processes can be changed by decisions that should be selected in the best possible manner. In this paper we discussed various facets of robustness assuming that the policy analysis includes optimization models with given sets of goals and feasible decisions. In reality these sets are also uncertain and they can be specified through a dialogue of users with models, where optimization models create only some blocks of the overall decision support system. Advances in modeling and computational methods allow us to create a "laboratory world" [10], where we can test new policies never implemented in reality. This "learning-by-modeling" dialogue with models requires specific robust optimization methods which are able to maintain a consistency of outcomes under the changing environment of the "laboratory world" where goals and sets of feasible solutions are subject to modifications by users, new information and gained experience. In particular, the evaluation of robust policies often requires specific robust optimization methods that are able to correctly detect the effects of rare extreme events. A discussion of these is beyond the scope of this paper. At least, they require the development of specific fast Monte Carlo procedures (see, e.g., [17]). The use of quantiles, thresholds, and stopping times requires, in general, specific non- smooth stochastic optimization methods [16], [17]. Since the notion of robustness depends on the nature of decision problems, it is hopeless to provide a complete overview of all its feasible facets. Therefore, in this paper we have primarily focused on issues relevant to on-going modeling of global change processes at IIASA.

References

[1] Alcamo, J., Shaw, R., Hordijk, L. (Eds.) (1990): The RAINS Model of Acidification, Science and Strategies in Europe. Kluwer Academic Publishers, Dordrecht/Boston/London.

[2] Alley, R.B., Marotzke, J., Nordhaus, W.D., Overpeck, J.T., Peteet, D.M., Pielke Jr., R.A., Pierrehumbert, R.T., Rhines, P.B., Stocker, T.F., Talley, L.D., Wallace, J.M. (2003): Abrupt Climate Change. Science **299**.

[3] K.J. Arrow (1996): The theory of risk- bearing: small and great risks. Journal of Risk and Uncertainty **12**, 103-111.

[4] Ben-Tal, A., Nemirovski, A. (1998): Robust Convex Optimization. Mathematics of Operation Research **23**, 769-805.

[5] Bierkens, M.F.P., Finke, P.A., de Willigen, P. (2000): Upscaling and Downscaling Methods for Environmental Research. Kluwer, Dordrecht, The Netherlands.

[6] Birge, J., Louveaux, F. (1997): Introduction to Stochastic Programming. Springer-Verlag, New York.

[7] Chichilnisky, G., Heal, G. (1993): Global Environmental Risks. Journal of Economic Perspectives **7(4)**, 65-86.

[8] Clark, W.C., Munn, R.E. (Eds.) (1985): Sustainable Development of the Biosphere. Cambridge University Press, Cambridge, UK.

[9] Danielson, M., Ekenberg, L. (1998): A Framework for Analysing Decisions under Risk. European Journal of Operation Research **104/3**, 474-484.

[10] Dantzig, G. (1979): The Role of Models in Determining Policy for Transition to a More Resilient Technological Society. IIASA distinguished lecture series **1**, 1979.

[11] Embrechts, P., Klueppelberg, C., Mikosch, T. (2000): Modeling Extremal Events for Insurance and Finance. Applications of Mathematics, Stochastic Modeling and Applied Probability. Springer Verlag, Heidelberg.

[12] Ermoliev, Y. (1970): On Some Stochastic Programming Problems. Kibernetika **1**, 3-9.

[13] Ermoliev, Y., Ermolieva, T.Y., MacDonald, G., and Norkin, V. (2000): Stochastic Optimization of Insurance Portfolios for Managing Exposure to Catastrophic Risks. Annals of Operations Research **99**, 207-225.

[14] Ermoliev, Y.E., Gaivoronski, A., Nedeva, C. (1985): Stochastic Optimization Problems with Incomplete Information on Distribution Functions. SIAM J. Control and Optimization **23/5**, 697-708.

[15] Ermoliev, Y., Leonardi, G. (1982): Some Proposals for Stochastic Facility Location Models. Mathematical Modeling **3**, 407-420.

[16] Ermoliev, Y., Norkin, V. (1997): On Nonsmooth and Discontinuous Problems of Stochastic Systems Optimization. European Journal of Operation Research **101**, 230-244.

[17] Ermoliev, Y., Norkin, V. (2004): Stochastic Optimization of Risk Functions. In: Marti, K., Ermoliev, Y., Pflug, G. (Eds.): Dynamic Stochastic Optimization. Springer Verlag, Berlin.

[18] Ermoliev, Y., Nurminskiy, E. (1980): Stochastic Quasigradient Algorithms for Minimax Problems in Stochastic Programming. In: Dempster M.A.H. (Ed.): Stochastic Programming. Academic Press, London.

[19] Ermoliev, Y., Wets, R. (Eds.) (1988): Numerical Techniques for Stochastic Optimization, Computational Mathematics, Springer Verlag, Berlin.

[20] Ermolieva, T.Y. (1997): The Design of Optimal Insurance Decisions in the Presence of Catastrophic Risks. IIASA Interim Report IR-97-068, Web: www.iiasa.ac.at.

[21] Ermolieva, T., Ermoliev, Y., Hepburn, C., Nilsson, S., Obersteiner, M. (2003): Induced Discounting and Its Implications to Catastrophic Risk Management. IIASA Interim Report IR-03-029, Web: www.iiasa.ac.at.

[22] Fischer, G., Ermolieva, T., Van Veltuizen, H., Yermoliev, Y. (2004): On Sequential Downscaling Methods for Spatial Estimation of Production Values and Flows. Proceedings of the Conference on Data Assimilation and Recursive Estimation: Methodological Issues and Environmental Applications, Venice, Italy.

[23] Fishburn, P. (1981): Subjective Expected Utility: A Review of Normative Theories. Theory and Decision **13**, 139-199.

[24] Gaivoronski, A.A. (1986): Linearization Methods for Optimization of Functionals Which Depend on Probability Measures. Mathematical Progamming Study **28**, 157-181.

[25] Golodnikov, A.N., Stoikova, L.S. (1978): Numerical Methods of Estimating Certain Functionals Characterizing Reliability. Cybernetics **2**, 73-77.

[26] Hudz, H., M. Jonas, T. Ermolieva, R. Bun, Y. Ermoliev and S. Nilsson (2003): Verification times underlying the Kyoto Protocol: Consideration of risk. Background data for IR-02-066. International Institute for Applied Systems Analysis, Laxenburg, Austria. Available on the Internet: http://www.iiasa.ac.at/Research/FOR.

[27] Huber, P.J. The Case of Choquet Capacities in Statistics. Bulletin of the International Statistical Institute **45**, 181-188.

[28] Insua, D.-R., Ruggeri, F. (2000): Robust Bayesian Analysis. Springer Verlag, New York.

[29] IPCC (2001). Climate Change 2001: The Scientific Basis. Technical Report. Intergovernmental Panel on Climate Change.

[30] Kall, P., Ruszczynski, A., Frauendorfer, K. (1988): Approximation Techniques in Stochastic Programming. In: Ermoliev, Y., Wets, R. (Eds.): Numerical Techniques for Stochastic Optimization, Computational Mathematics, Springer Verlag, Berlin.

[31] Markowitz, H.M. (1987): Mean Variance Analysis in Portfolio Choice and Capital Markets. Blackwell, Oxford.

[32] Marti, K. (2005): Stochastic Optimization Methods. Springer, Berlin, Haidelberg.

[33] Manne, A.S., Richels, R.G. (1995): The Greenhouse Debate: Economic Efficiency, Burden Sharing and Hedging Strategies. The Energy Journal **16/4**, 1-37.

[34] Morgan, M.G., Kandlikar, M., Risbey, J., Dowlatabadi, H. (1999): Why Conventional Tools for Policy Analysis are Often Inadequate for Problems of Global Change: An Editorial Essay. Climatic Change **41/(3-4)**, 271-281.

[35] Mulvey, J.M., Vanderbei, R.J., Zenios, S.A. (1995): Robust Optimization of Large Scale Systems. Operations Research **43**, 264-281.

[36] Nordhaus, W.D., Boyer, J. (2001): Warming the World: Economic Models of Global Warming, MIT Press, Cambridge, Mass.

[37] ONeill, B., Ermoliev, Y., Ermolieva, T. (2006): Endogenous Risks and Learning in Climate Change Decision Analysis. In: Marti, K., Ermoliev, Y., Pflug, G., Makovskii M. (Eds): Proceedings of the IFIP/IIASA/GAMM Workshop On Coping with Uncertainty, Springer-Verlag, Heidelberg.

[38] OXERA (2002): A Social Time Preference Rate for Use in Long-term Discounting. OXERA Press, 1-74.

[39] Prekopa, A. (1995): Stochastic Programming. Kluwer Academic Publishers, Dorbrecht, Netherlands.

[40] Pugh, E.L. (1966): A gradient technique of adaptive Monte Carlo. SIAM Review **8/3**.

[41] Rockafellar, T., Uryasev, S., Optimization of Conditional-Value-at-Risk, The Journal of Risk **2**, 21-41.

[42] Reithmeier, E., Leitmann, G. (1996): Robust Constrained Control for Vibration Suppression of Mismatched Systems. Applied Mathematics and Computation **78**, 245-257.

[43] Shafer, G. A. (1976): Mathematical Theory of Evidence. Princeton University Press.

[44] Shapira, Z. (1995): Risk Taking: A Managerial Perspective. Russel Sage Foundation.

[45] Wallace, S.W., Ziemba, W.T. (2005): Applications of Stochastic Programming. MPS-SIAM Books Series on Optimization, **5**.

[46] G.R. Walker (1997): Current Developments in Catastrophe Modeling, in: Financial Risk Management for Natural Catastrophes, eds. N.R. Britton and J. Oliver, Aon Group Australia Limited, Griffith University, Brisbane, 17-35.

[47] Wigley, T.M.L. (1985): Impacts of Extreme Events. Nature **286**, 106-107.

[48] Wright, E.L., Erickson, J.D. (2003): Incorporating Catastrophes into Integrated Assessment: Science, Impacts, and Adaptation. Climate Change **57**, 265-286.

[49] Zackova, J. (1966): On Minimax Solutions of Stochastic Linear Programming. Casopis pro Pestovani Matematiky **91**, 430-433.