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
Ib: www.iiasa.ac.at

Interim Report **IR-10-015**

Allocation of Resources for Protecting Public Goods against Uncertain Threats Generated by Agents

Chen Wang (cwang37@wisc.edu)

Approved by
Marek Makowski

Leader, Integrated Modeling Environment Project

December 2010

Foreword

This report describes the research the author advanced during her participation in the 2010 Young Scientists Summer Program (YSSP) with the Integrated Modeling Environment Project. The aim of this research is to build up a framework supporting robust decision-making for protecting public goods against uncertain threats generated by agents and chances.

Under increasing interdependencies of globalization processes the protection of public goods is becoming a critical topic, especially against uncertain threats generated by agents. Examples include both direct threats such as terrorist attacks, recent BP oil spill, and financial crisis, and indirect threats associated with natural disasters such as improper land use planning and cascading risk management at a disaster prone area. This work builds up a framework for such a broad class of decision problems with inherent uncertainties and strategic responses.

The framework combines the “leader-follower” game concept with approaches of stochastic optimization and multicriteria analysis. It incorporates both mathematical models and computational algorithms for public goods protection against uncertain and endogenous threats, which makes it ready for realistic applications. In particular, two case studies are presented, including defending urban areas against uncertain intentional attacks and regulating electricity networks with consideration of possible outages.

The research conducted during the three-month YSSP period will be continued by advancing the theoretical framework and its practical applications.

Abstract

This paper analyses a framework for designing robust decisions against uncertain threats to public goods generated by multiple agents. The agents can be intentional attackers such as terrorists, agents accumulating values in flood or earthquake prone locations, or agents generating extreme events such as electricity outage and recent BP oil spill, etc.

Instead of using a leader-follower game theoretic framework, this paper proposes a decision theoretic model based on two-stage stochastic optimization (STO) models for advising optimal resource allocations (or regulations) in situations characterized by uncertain perceptions of agent behaviors. In particular, the stochastic mini-max model and multi-criteria STO model are presented to solve for two different types of protection decisions for public goods security. Furthermore, the use of conditional value at risk (or expected shortfalls) is advanced in the context of quantile optimization for dealing with potential extreme events.

Proposed framework can deal with both direct and indirect judgments on the decision maker's perception about uncertain agent behaviors, either directly by probability density estimation, or indirectly by probabilistic inversion. The quantified distributions are treated as input to the stochastic optimization models in order to address inherent uncertainties. Robust decisions can then be obtained against all possible threats, especially with extreme consequences.

This paper also introduces and compares three different computational algorithms which can be used to solve arising two-stage STO problems, including bilateral descent method, linear programming approximation and stochastic quasi-gradient method. A numerical example of high dimensionality is presented for illustration of their performance under large number of scenarios typically required for dealing with low probability extreme events. Case studies include defensive resource allocations among cities and security of electricity networks.

Acknowledgments

This research was conducted as a part of the author's participation in the Young Scientists Summer Program (YSSP) 2010 at the International Institute for Applied Systems Analysis (IIASA), with support from the US National Academy of Sciences, thanks to NSF Grant Number OISE-738129.

I would like to thank my YSSP supervisor Professor Yuri Ermoliev in the Integrated Modeling Environment project, and his colleagues Dr Marek Makowski and Dr Tatiana Ermolieva, for their advices and help in diverse aspects of my research and professional life. I am grateful to the Director of IIASA Professor Detlof von Winterfeldt for his attention to my work and important advices for its improvement.

Finally, I would also like to thank my PhD supervisor Professor Vicki M. Bier in the Department of Industrial and Systems Engineering, University of Wisconsin-Madison for her advices on the problem formulation, and her support for this research.

About the Author

Chen Wang is currently a PhD candidate in the Department of Industrial and Systems Engineering at the University of Wisconsin-Madison. She holds a Bachelor's degree in Industrial Engineering from Beijing University of Aeronautics and Astronautics, China and a Master's degree in Industrial and Engineering from the University of Wisconsin. Her research interests include quantifying uncertainty from expert judgments and modeling resource allocation problems in the face of such uncertainty, mainly in the context of homeland security.

Table of Contents

List of Tables

List of Figures

Allocation of Resources for Protecting Public Goods against Uncertain Threats Generated by Agents

Chen Wang (cwang37@wisc.edu) [1]

1. Introduction

Decision making for security of public goods should be robust against uncertain threats, especially extreme events generated intentional or unintentional by agents. The standard deterministic analysis usually takes only one single scenario as given, without considering the vast variety of potential scenarios with often non-normal distributions (e.g., heavily tailed or multi-modal) or seemingly irrelevant outliers which may ruin the mean value analysis. Such uncertainties of decision makers arise from their lack of knowledge not only about exogenous factors generated by chances (such as natural disasters), but also about agent behaviors. However, in the context of security analysis, the impact of outliers ("extreme events") is especially significant. Therefore, explicitly quantifying such uncertainties and applying stochastic optimization methods as decision support tools are two important tasks when modeling public good protection.

Examples of extreme events generated by agents include intentional threats such as terrorism. Potential attackers choose from areas (urban areas, military bases, sites on a foreign land, and etc.) and specific places (subways, shopping centers, food supply chains, and etc.) to launch an attack or simultaneous attacks. Choices of targets and means (e.g., improvised explosive device, IEDs) are based on their desirability and accessibility to destinations and resources, on which decision makers can only have partial intelligence. Other examples include social, energy, and financial threats caused by inappropriate agent behaviors. For example, recent BP oil spill[2] has shown its long time ignorance of reliability against such an event with low probability but extreme consequences and the lack of regulations imposing costs on violation of preparedness measures. Some natural disasters may seem less evident to relate to agent actions. However, for example, the catastrophic flood of hurricane Katrina is a combination of natural chances and a failure of levees due to lack of maintenance. The huge amount of losses could also have been avoided if there were regulations against building values close to the levees (Ermoliev and von Winterfeldt, 2010). Another example of improper land use planning is that farmers (agents) in a volcanic prone area tend to move towards the epicenter of eruption for more fertile lands.

Regulations of public goods security have been widely discussed in the context of principal agent models (Laffont and Martimor, 2002), in which a principal regulates all agents to achieve a good overall performance of a system, while agents tend to maximize their individual payoffs. Principal agent (PA) models have feature of the two-stage Stackelberg game, or leader-follower game, where the principal moves first to formulate regulations (or distribute resources) and then all agents make full observation and choose their optimal strategies. It is assumed that the PA has full information about agents and uses their response functions designing his decisions.

[1] Integrated Modeling Environment Project, IIASA, Laxenburg, Austria
Department of Industrial and Systems Engineering, University of Wisconsin-Madison, USA

[2] BP oil spill in 2010 is a massive oil spill in the Gulf of Mexico that is the largest offshore spill in U.S. history. BP is a global energy company headquartered in London, United Kingdom.

Some models allow the agents to have private information, which can be useful in modeling uncertain threats to public goods. However, it is still assumed that the PA is able to evaluate exact expectation functions of agents and there are strong commitments of agents for using these functions and resulting response functions. These unrealistic assumptions may result in unstable misleading solutions. They also create serious computational difficulties. Since the principal and agents usually have different objective functions, even if both objectives are well-defined convex (or concave) functions, the leader-follower structure will generally lose such properties, and solving resulting optimization problem becomes very complicated task. Therefore, instead of the game theoretic framework, this paper follows general approach proposed in Ermoliev and von Winterfeldt (2010) to decision-theoretic modeling based on stochastic optimization (STO) to solve for robust decisions against uncertain threats to public goods. In particular, models of followers and the leader are formulated as STO models with explicit introduction of uncertainties based on the leader's perception of followers behavioral scenarios.

This paper first develops the stochastic minimax model to a problem of defensive resource allocations against intentional attacks. Considering that the agents (intentional attackers) may have private information about their target preferences, the principal (defender) attempts to minimize the expected value of maximal random payoffs to the agent by using PA perception of agents scenarios. This stochastic worst case analysis (stochastic minimax), in fact, corresponds to a decision-oriented extreme events model for regulating public goods security against perceived extreme scenarios of agents. The stochastic minimax model well preserves convexity (or concavity) of the objective functions, which is powerful in developing both analytical and computational results.

In general, dealing with multi-agent problems under uncertainty may lead to rather different multi-criteria STO models (Arthan, 1994). In particular, if the principal and agents share common interests, one can construct the total objective as expected value of the weighted sum of individual objectives. This paper applies the multi-criteria STO model to electricity networks, where the System Operator (SO) determines dispatch of electricity in the electricity network while firms determine generation quantities at each generation facility to gain profits, considering that there are possible outage of power plants and breakdown of transmission lines.

Uncertainties about agent behaviors can be quantified by probability distributions (either by density functions or simulated scenarios) of those uncertain parameters in agent utility functions. However, extreme events generated by agents are generally lack real repetitive observations, so such distributions usually cannot be obtained through standard statistical analysis. In particular, when direct estimations are not available, I can use probabilistic inversion (Kraan and Bedford, 2005; Du et al., 2006) to infer the underlying expert perception about the parameters of interest. For example, in the problem of protecting cities against intentional attacks, if I have expert opinions on attacker rankings of potential targeted cities, I can probabilistically invert their subjective distributions (as simulated scenarios) on the relative importance of city attributes (e.g., expected loss from terrorist attacks, population, national icon, difficulty of launching an attack, and etc), and even the characteristics of un-quantified attributes.

Furthermore, minimizing expected losses (or maximizing expected payoffs) may not be adequate to capture the problems involving extreme events, since mean values are generally not robust to "outliers" (Koenker and Bassett, 1978). For example, when minimizing losses, one may need to focus on the extreme losses beyond a certain critical value. Ermoliev and von Winterfeldt (2010) propose to use the multicriteria version of conditional value at risk (CVaR) or expected shortfall (Uriasev and Rockafellar, 2000; Artzner, 1997; 1999) as the optimization objective to deal with human-related extreme events. The CVaR is defined as the conditional expected loss beyond a certain quantile (value at risk, or VaR). Ermoliev and von Winterfeldt (2010) represent an integrated STO model to simultaneously solve for the quantiles and the optimal CVaR. In this report I present a case study on the discussion of mean value objectives versus conditional value at risk objectives.

The structure of this paper is as follows. Section 2 introduces the STO models that can be used for protecting public goods against uncertain threats generated by agents. The two-stage stochastic minimax model and the two-stage multi-criteria stochastic optimization (STO) model are introduced and two examples of application are presented. Furthermore, this section discusses use of the CVaR as an optimization objective and its relation to quantile optimization. Section 3 explores ways of quantifying decision makers' uncertain perception about agent behaviors. In particular, the technique of probabilistic inversion is applied to elicit indirect expert perceptions about uncertain parameters in the agent utilities. Section 4 discusses in detail an application of defensive resource allocations against intentional attacks. A case study of protecting 47 US urban areas is presented. Section 5 focuses on another application dealing with security of electricity networks. Optimal dispatch decision for the system operator (SO) is discussed for a case study of Belgian high voltage network. Section 6 describes and compares several specific algorithms developed to solve the arising STO problem, including bilateral descent method, linear programming (LP) approximation and stochastic quasi-gradient (SQG) method (Ermoliev, 1983; 2009). Section 7 concludes this paper.

2. Stochastic Optimization (STO) Models

In this section I present two different STO models which can be applied to problems of protecting public goods against uncertain threats generated by agents, including the stochastic minimax model and the multi-criteria STO model. Both models belong to the class of two-stage STO problems. In particular, in both models, the agents are assumed to make strategic decisions in response to the principal actions, and the principal (decision maker) is assumed to have incomplete information about agent behaviors.

The stochastic minimax problem is used in the cases where the principal and agents have opposite objectives. A typical setting is that the principal wants to minimize the perceived payoffs to the agents. A problem of defensive resource allocations against intentional attackers is demonstrated as an application of the stochastic minimax model. The stochastic multi-criteria model is applicable to the cases where the principal and agents share some common interests, so that they optimize the objective to the same direction. An application of security of electricity network is presented. Moreover, both models can deal with problems of heterogeneous agents or problems where the principal puts different weights on different agents.

At the end of this section, I will discuss the concept of CVaR, which can be appropriate as an optimization objective for a STO problem when modeling threats of extreme events. Based on that, I will introduce quantile optimization.

2.1. Two-Stage Stochastic Optimization (STO) Problems

The two-stage stochastic optimization model (also called the recourse model) can be used for decisions in the face of both adaptive and uncertain agent behaviors. A general two-stage stochastic optimization mode is formulated as

$$\min_{x \in X} Ef(x, y(x,\omega), \omega) \qquad \text{(\textbf{Error! Bookmark not defined.}1)}$$

$$\text{s.t. } f_i(x, y(x,\omega), \omega) \le 0,\; i = 1,...,l, \qquad (2)$$

where $y(x,\omega)$ minimizes (or maximizes) $f(x, y, \omega)$ with respect to y for given x, ω and constraints (2).

An anticipative decision $x \in X$ must be made at stage 1 before the observation of uncertain factor ω is available. At stage 2, for a given $x \in X$ and an observed realization of ω, the adaptive decision y is made according to some response function $y(x,\omega)$. Note that $y(x,\omega)$ can be an implicit function. The main problem is to find first stage variable $x \in X$ so as to minimize the expected value of the function $f(x, y(x,\omega), \omega)$. The first stage variable x corresponds to the principal decision and the second stage variable y to the agent responses. Then the two-stage STO model provides a general framework for the principal-agent models with uncertainty.

Suppose that the probability measure of ω is independent of the decision variables x, that is

$$Ef(x, y(x,\omega), \omega) = \int f(x, y(x,\omega), \omega) dP(\omega), \qquad (3)$$

then the two-stage stochastic optimization problem can be approximately solved by the sample average approximation (SAA) of (1) as given by

$$\min_{x \in X} \frac{1}{N} \sum_{s=1}^{N} f(x, y(x, \omega^s), \omega^s) \qquad (4)$$

where N is the total number of scenarios $s = 1,..., N$.

In this paper, I will mainly focus on cases where the objective function $f(x, y(x,\omega), \omega)$ is convex in x for all feasible ω. Three computational algorithms are proposed to solve such problems, including bilateral descent method, linear approximation (LP) approximation and SQG method. Details of the three algorithms are presented in Section 6.

2.2. Stochastic Minimax Model

Assume $f(x, y, \omega) = \sum_{i=1}^{n} g_i(x, y_i, \omega)$ and for the simplicity of notation, constraints (2) are absent.

If the response function for agents $y(x,\omega) = (y_1(x,\omega),..., y_n(x,\omega))$ in (1) maximizes the individual payoffs for each agent, i.e.,

$$y_i(x,\omega) = \arg\max_{y_i} g_i(x, y, \omega),\; i = 1,...,n, \qquad (5)$$

where n is the total number of agents, then I specify (1) as the stochastic minimax model

$$\min_{x \in X} E \max_{y} \sum_{i=1}^{n} g_i(x, y, \omega). \qquad (6)$$

The adaptive agents are to maximize their individual payoffs given the principal's decision, while the principal's aim is to minimize the summation of their payoffs. In addition, this model is also equivalent to a worst case analysis where the principal attempts to minimize the expected loss from the stochastic worst case. Furthermore, this model also considers the externality between multiple agents, if the agent utility functions $g_i(x, y, \omega)$ are not separable.

Suppose that each individual agent utility function $g_i(x, y, \omega)$ are convex in the 1st stage variable x, the stochastic mini-max model well preserves such convexity, since the summation of convex functions is convex, and the maximum of convex functions is convex. Moreover, the expectation of a stochastic convex function is also convex. Therefore, the entire optimization problem (6) and its sample average approximation (7) are both convex problems, and the local optimum corresponds to the global optimum

$$\min_{x \in X} \frac{1}{N} \sum_{s=1}^{N} \max_{y} \sum_{i=1}^{n} g_i(x, y, \omega^s) \tag{7}$$

For a realized scenario $s = 1,..., N$, the first derivative of $\max_{y} \sum_{i=1}^{n} g_i(x, y, \omega^s)$ is given by

$$\frac{df}{dx} = \frac{\partial}{\partial x} \sum_{i=1}^{n} g_i(x, y, \omega^s) \Big|_{y = y(x,\omega)} \tag{8}$$

for all components of vector X; this property is useful for developing computational algorithms.

2.3. Defensive Resource Allocations against Intentional Attacks

Suppose the defender is faced with potential attacks on a collection of targets (e.g., cities, critical infrastructures, public transportation systems, and etc.). The defender's objective is to minimize the consequences from attacker choices. A Stackelberg game is usually used to model this situation when there is no uncertainty about the attacker preferences. A Stackelberg game is that the defender moves first to decide on an allocation of her defensive resources among a heterogeneous collection of potential targets. The attacker then observes the defensive allocation, chooses whether to attack, and if making an attack, chooses an attack target. However, in reality, the attacker may have private information about his preferences (usually represented by uncertain parameters), which is not fully known to the defender. In the face of such uncertainty, the defender cannot predict the attacker's best response for sure; therefore, the STO model (especially the stochastic minimax model) is needed and the defender is assumed to minimize the expected total consequences.

For simplicity, suppose the defender is faced with one attacker, whose decision is to choose a target i among n targets with the highest payoff to attack. The defender objective is to minimize

$$\min_{x \in X} E \max_{i} g_i(x, \omega) \tag{9}$$

where $x \in X$ is the defensive resource allocation decision among targets and X is a simplex

$$X = \{x \in R^n \mid \sum_{i=1}^{n} x_i = B, x_i \geq 0, i = 1,..., n\} \tag{10}$$

Following the model in (Bier et al., 2007; Wang and Bier, 2010), the attacker utility function on each target $g_i(x,\omega) = p(x_i)u_i(\omega)$ is a product of target vulnerability (success probability)

$$p(x_i) = e^{-\lambda_i x_i} \tag{11}$$

and the attack consequence

$$u_i(\omega) = \sum_{j=1}^{m-1} w_j A_{ij} + w_m \varepsilon_i \tag{12}$$

Note $\omega = (w_1,...,w_m,\varepsilon_1,...,\varepsilon_n)$ is a random vector representing all uncertain parameters in the attacker utility function. I assume that the success probability of an attack on target i is an exponential function of the defender's investment in that target, where λ_i is the cost effectiveness of defensive investment on target i. For example, at the cost effectiveness level of 0.02, if the investment is measured in millions of dollars, then every million dollars of defensive investment will reduce the success probability of an attack by about 2%. I also assume that consequences are valued by the attacker according to a multi-attribute utility function with m attributes (of which $m-1$ are assumed to be observable by the defender). I assume that the attacker's utility is linear in each of the various attacker attributes.

A_{ij} = attacker utility of target i on the j th attribute, where A_{ij} takes values in [0,1], with 1 representing the best possible value and 0 the worst.

ε_i = utility of the unobserved (by the defender) m th attribute of target i.

$(w_1,...,w_m)$ = attacker weights on the m attributes, where $\sum_{j=1}^{m} w_j = 1$ and $w_j \geq 0, j = 1,...,m$.

The defender uncertainty about attacker preferences lies in the random feature of attribute weights $(w_1,...,w_m)$ and unobserved attributes $(\varepsilon_1,...,\varepsilon_n)$. How to quantify uncertainties about these parameters by probability distributions will be discussed in Section 3. A case study of allocating defensive resources among the major US urban areas against intentional attacks will be illustrated in Section 4.

2.4. Two-stage Multi-criteria Stochastic Optimization Model

In some problems, the principal and agents share some common interests. When the principal's main objective is to maximize the social welfare, she also cares about the agent individual profits. At the same time, agents attempt to maximize their individual payoffs without hurting the social welfare. An example is family welfare. The householder is to maximize the family welfare, which is an aggregation of individual welfare for each family member. When the individual family members make decisions, they also care about the total family welfare (Chiappori, 1992).

This class of problems can be modeled in a way of multi-criteria stochastic optimization (STO). The principal is to maximize

$$\max_{x \in X} E \max_{y} \{g_0(x,y,\omega) + \sum_{i=1}^{n} \upsilon_i g_i(x,y,\omega)\} \tag{13}$$

Where $g_0(x, y, \omega)$ is the function of social welfare that the principal mainly cares about, $g_i(x, y, \omega)$ is the agent individual utility function, and υ_i are importance weights the principal puts on different agents. In the example of family welfare, if the householder pays more attention on the utility of elderly family members, she can put higher value on υ_i for them. Again, this model also considers the externality between agents, if the agent utility functions $g_i(x, y, \omega)$ are not separable.

The model (12) is also a special case of the general two-stage stochastic optimization (STO) model (1). Furthermore, if convexity in (x, y) is assumed for each of the utility functions $g_i(x, y, \omega)$, $i = 0,..., n$, the total objective is also convex.

2.5. Security of Electricity Networks

The California energy crisis in 2001 and the collapse of ENRON raise serious concerns about regulations of an electricity network. Leader-follower models have been used to support policy decisions on design and regulation of electricity markets (Ermoliev and von Winterfeldt, 2010). Following the way of modeling by Yao et al. (2008), I consider a system where the independent system operator (ISO) is eligible to control the transmission system and generation firms determine their generation quantities of electricity at each power plant.

An electricity network can be represented by a set of n nodes and a set of L transmission lines. The independent system operator (ISO) determines dispatch (import/export) of electricity at each node. In other words, if the ISO decides on a negative dispatch (export) at a given node of power plant, apart from satisfying its own demand at this node, the power plant needs to produce extra amount of electricity in order to export to other nodes. According to the Kirchhoff's current law that the sum of all current entering a node is equal to the sum of all currents leaving this node, the ISO's decision variables the dispatch quantities $r_i, i = 1,..., n$ at each node must satisfy the balance equation (Yao et al., 2008)

$$\sum_{i=1}^{n} r_i = 0 \tag{14}$$

Moreover, the transfer amount should not exceed the thermal limits of each transmission line $l = 1,..., L$

$$-K_l \le \sum_{i=1}^{n} D_{li} r_i \le K_l, \; l = 1,..., L \tag{15}$$

where D_{li} is the power transfer distribution factor (PTDF) which is an exogenous feature of the electricity network specifying the proportion of flow from a generation node i onto a transmission line l.

Given the ISO's decision on dispatch of electricity r_i at each node $i = 1,..., n$, the electricity producer determines its generation quantity q_i at each node to optimize their profits by maximizing the profit function

$$P_i(q_i + r_i)q_i - C_i(q_i) \tag{16}$$

subject to the capacity limit of each power plant

$$0 \le q_i \le \overline{q}_i \tag{17}$$

at each node $i = 1,...,n$, where $P_i(q)$, $C_i(q)$ are the inverse demand function (wiliness to pay) and generation cost function; $\overline{q}_i$ is the upper bound for generation capacity at node i.

In addition, in order to meet the requirement of ISO dispatch, the generation quantity at each node should also satisfy

$$q_i + r_i \ge 0, i = 1,...,n \tag{18}$$

The ISO's main goal is to maximize the social welfare

$$\sum_{i=1}^{n} \int_{0}^{r_i+q_i} P_i(u)du \tag{19}$$

Taking into the ISO uncertainties about the parameters ω in the functions and constraints (14 - 19), the ISO's objective is to determine dispatch of electricity for each node in order to maximize the expected total welfare as the combination of social welfare and individual profits of electricity producers (Ermoliev and von Winterfeldt, 2010).

$$\max_r E[\max_q \{\sum_{i=1}^{n} \int_{0}^{r_i+q_i} P_i(u,\omega)du + P_i(r_i + q_i, \omega)q_i - C_i(q_i, \omega)\}] \tag{20}$$

$$\text{s.t. } \sum_{i=1}^{n} r_i = 0$$

$$q_i + r_i \ge 0, i = 1,...,n$$

$$-K_l(\omega) \le \sum_{i=1}^{n} D_{li} r_i \le K_l(\omega), l = 1,...,L$$

Random parameters ω in the objectives and constraints represent the ISO uncertainties. For example, consider possible outage at a given node i, I can set the cost function to be

$$C_i(q,\omega) = c_i(q) + \omega_i \tag{21}$$

where $c_i(q)$ is the original generation cost function, and ω_i is a random variable. When ω_i takes value 0, it means there is no outage at this node. When ω_i takes value infinity, it means the generation cost at node i is infinity, which is equivalent to an outage. Another example is breakdown of transmission lines. Consider the thermal limit of each transmission line. $K_l(\omega_l) = K_l$ represents no breakdown, while $K_l(\omega_l) = 0$ represents breakdown of transmission line l. More discussions on the quantification of uncertainties are presented in Section 3.

A case study of Belgian high voltage electricity network is shown in Section 5.

2.6. Quantile Optimization

Besides the expected value of the stochastic losses, it is also useful to consider the tail distribution of the stochastic losses, especially when dealing with extreme events of high consequence and low probability. Ermoliev and von Winterfeldt (2010) propose to use the conditional value at risk or expected shortfalls as the optimization objective of general STO models to deal with human-related extreme events. The conditional value at risk is defined as the conditional expected loss beyond a certain quantile. Consider a random variable θ, the conditional value at risk for quantile z_q is given by

$$E[\theta \mid \theta \geq z_q] \tag{22}$$

It can be shown that both the value of (22) and the corresponding quantile z_q can be obtained simultaneously by the quantile optimization model

$$\min_z (1-q)z + E[\max\{\theta - z, 0\}] \tag{23}$$

where $q \in (0,1)$ is the quantile level. In addition, the conditional expected value below a certain quantile can be obtained by another quantile optimization model

$$E[\theta \mid \theta \leq z_q] = \max_z qz + E[\min\{\theta - z, 0\}] \tag{24}$$

If I extend the random variable θ to be the random objective function $f(x, y(x,\omega), \omega)$ in the general two-stage stochastic optimization (STO) problem (1), then by minimizing

$$\min_{x,z} (1-q)z + E[\max\{f(x, y(x,\omega), \omega) - z, 0\}] \tag{25}$$

I can simultaneously get x^* and z^* such that x^* is the optimal decision which solves the minimal conditional value at risk

$$\min_x E[f(x, y(x,\omega), \omega) \mid f(x, y(x,\omega), \omega) \geq z^*] \tag{26}$$

and z^* is inherently the quantile at level q such that

$$\Pr ob\{f(x^*, y(x^*,\omega), \omega) \leq z^*\} = q \tag{27}$$

Note that the model (24) also falls into the general case of two-stage STO

$$\min_{x,z} E[(1-q)z + \max\{f(x, y(x,\omega), \omega) - z, 0\}] \tag{28}$$

and if $f(x, y(x,\omega), \omega)$ is a convex function, the problem (28) is also convex. Therefore, all the computational algorithms this paper will discuss in Section 6 are also applicable to the quantile optimization model (23). I can easily incorporate the quantile optimization model to the stochastic minimax model and the two-stage multi-criteria STO model to extend the application of human-related extreme events.

A setting of chance (safety) constraint (Miller and Wagner, 1965) can also be applied and the original constraints in the two-stage STO model (2) become

$$\Pr ob\{f_i(x, y(x,\omega), \omega) > 0\} \leq \varepsilon_i, i = 1,\ldots,l \tag{29}$$

where ε_i are small positive numbers (safety levels). Furthermore, the reliability-based design optimization is another way to model uncertain constraints. Detailed discussions can be found in (Bordley and Pollock, 2009), (Ermoliev and Winterfeldt, 2010).

3. Quantification of Uncertainty

The inherent and deep uncertainty about agent behaviors is critical to models of protecting public goods. The equilibrium obtained in a deterministic model is usually unstable to even a subtle change in the agent parameters. The STO models are developed to solve for robust decisions against such uncertainties. Therefore, quantifying uncertainty becomes an important task to provide input for the STO models. In this section I present two ways of quantifying uncertainty, from both direct and indirect expert judgments on the agent behaviors.

Uncertainties about agent behaviors can be quantified by probability distributions of uncertain parameters in agent utility functions. I can present the decision maker knowledge directly, if probability densities or simulated scenarios on the parameters of interest are available. Moreover, if additional data related to those parameters are observed, I can apply Bayesian analysis to update the prior distributions to posterior distributions. For example, in the application of electricity network, the decision maker can assign a prior probability of outage at a given power plant, according to his expertise and historical data. He can also change this probability upon availability of new observations related to electricity outages.

However, in some cases direct judgments are not available. Instead the decision maker has expert judgment on some “observable space” that has a relation to the parameters of interest. Probabilistic inversion (PI) is a powerful tool to elicit indirect expert judgments which can infer probability distributions over the parameter space from probability distributions over the observable space. For example, in the application of defensive resource allocations against intentional attacks, if available are only expert opinions about attacker ranking of cities, I can use PI to elicit probability distributions for the relative importance of all the city attributes. Furthermore, it can even infer the characteristics of some unobserved attributes that lead to the expert ranking judgments but the decision maker is not aware of.

3.1. Probability Distributions of Agent Parameters

When direct judgments on the uncertain parameters ω as in (1) are available, the uncertainties can be quantified directly through probability distributions (either density functions or simulated scenarios). In the problem of the defensive resource allocations, I can construct subjective probability distributions to model the decision maker uncertainty about attribute weights and unobserved attributes $\omega = (w_1,..., w_m, \varepsilon_1,..., \varepsilon_n)$ in (12). Appropriate choices of prior distributions may include the Dirichlet distribution[3] for attribute weights $(w_1,..., w_m)$ and independent uniform distribution for unobserved attributes $(\varepsilon_1,..., \varepsilon_n)$. In the problem of electricity network, in order to deal with uncertainties of possible outages, I can assign probabilities on different scenarios of the generation cost function (21). For example, when I assign probabilities on possible values of ω_i as $\omega_i = 0$ with probability $p_i > 0$ and $\omega_i = +\infty$ with probability $1 - p_i$, then the random cost function (21) becomes

[3] See Appendix A

$$C_i(q,\omega) = \begin{cases} c_i(q) & w.p. \quad p_i \\ +\infty & w.p.\, 1-p_i \end{cases} \tag{30}$$

3.2. Bayesian Analysis

Upon availability of new data, I can update the prior probability distributions about agent parameters to posterior distributions by the Bayes' theorem. Let D represent newly observed data, then the posterior distribution for the uncertain parameters ω

$$P(\omega \mid D) = P(\omega)P(D \mid \omega) \tag{31}$$

where $P(\omega \mid D)$ is the posterior distribution after observation of data D, $P(\omega)$ is the prior distribution before observation of data D, and $P(D \mid \omega)$ is the likelihood that data D should happen given the parameters ω. Using (31), I can easily obtain the posterior distribution by simulation, resulting in a set of simulated scenarios representing the uncertainty about ω combining prior judgment $P(\omega)$ and observation of data D. Note that simulated scenarios can be used as direct input to the two-stage STO problem (1).

3.3. Probabilistic Inversion

I consider the case when direct judgments on the uncertain agent parameters $\omega \in \Theta$ are not available, however, I have judgments on some other observables Y which is supposed to have a presumed relation with ω. Note that both observables Y and parameters ω are random vectors, and $G: R^{|\omega|} \to R^{|Y|}$ is a presumably fixed mapping. If

$$G(\omega) \in \{Y \mid Y \in C\} \tag{32}$$

where C is a subset of random vectors on $R^{|Y|}$, then ω is called a probabilistic inverse of G at C (Kraan and Bedford, 2005). ω is sometimes termed the input to model G which are parameters of interest to the decision maker but not observable, and Y the output which is the observable. I usually start with a uniform measure over the parameter space Θ (all feasible scenarios of ω), and drive it by probabilistic inversion to match the available distribution of observable Y. The problem is always feasible if the response of parameter space Θ is broader than the observable space C. When the problem is feasible it may have multiple solutions and I will seek for a preferred solution which elicits as much but no more than all the available information. If the problem is infeasible I seek a random vector ω for which $G(\omega)$ is "as close as possible" to C. More details will be discussed in the following session.

3.4. Probabilistic Inversion of Ranking Judgments

Probabilistic inversion is applicable for arbitrary indirect judgments as long as there exist some relations between the parameters ω and the observables Y. I will consider the process of probabilistic inversion by the example of defensive resource allocation problem. Recall that the attacker utility function on each target is given by

$$u_i(\omega) = \sum_{j=1}^{m-1} w_j A_{ij} + w_m \varepsilon_i ,$$

where the uncertain parameters of interest are $\omega = (w_1,..., w_m, \varepsilon_1,..., \varepsilon_n)$. However, I do not have direct judgments on those parameters. What I have are expert ranking judgments on the top R out of n targets. The ranking judgment is presented as a double stochastic probability matrix

$$P = [p_{ri}]_{R\times n} \tag{33}$$

where p_{ri} represents the proportion of experts who rank target i at the r-th place, $\sum_{r=1}^{R} p_{ri} = 1$, $\sum_{k=1}^{n} p_{ri} = 1$, and $p_{ri} \geq 0, r = 1,..., R; k = 1,..., n$. There are a number of algorithms that can be used for probabilistic inversion such as Iterative Proportional Fitting (Fienberg, 1970), PARUM (Du et al., 2006) and PREJUDICE (Kraan and Bedford, 2005). In this report, I will mainly follow the way of PREJUDICE (Kraan and Bedford, 2005). First, a set of "background" scenarios are generated uniformly on the parameter space $\omega^s = (w_1{}^s,..., w_m{}^s, \varepsilon_1{}^s,..., \varepsilon_n{}^s)$. For each of the scenario $s = 1,..., N$, an ranking indicator matrix J^s is calculated. $J^s{}_{R\times n}$ is a binary matrix with the same dimension as $P_{R\times n}$, where $J^s{}_{ri} = 1$ if target i is ordered at the rth place according to the utility function $u_i(\omega^s) = \sum_{j=1}^{m-1} w_j{}^s A_{ij} + w_m{}^s \varepsilon_i{}^s$ in (12); $J^s{}_{ri} = 0$ if otherwise. Note that $\sum_{i=1}^{n} J^s{}_{ri} = 1$, while $\sum_{r=1}^{R} J^s{}_{ri}$ can be one or zero. The task of probabilistic inversion is to find a measure $\{q \in R^N \mid \sum_{s=1}^{N} q_s = 1, q_s \geq 0, s = 1,..., N\}$ on the "background" samples $\omega^s, s = 1,..., N$ so that

$$T = \sum_{s=1}^{N} q_s J^s \tag{34}$$

is as "close" to the ranking judgment probability matrix P. In particular, I want to minimize the Kullback-Leibler distance between T and P

$$\min_{q_s} \sum_{r=1}^{R} \sum_{i=1}^{n} T_{rk} \ln \frac{T_{ri}}{P_{ri}} \tag{35}$$

subject to (34). If T_{ri} = 0 or P_{ri} = 0, then substitute $\ln \frac{T_{ri}}{P_{ri}}$ with $\ln \frac{T_{ri} + \varepsilon}{P_{ri} + \varepsilon}$ for a small positive number ε. Since (35) is a convex function, all three computational algorithms presented in Section 6 can be used to solve (35). The computational algorithm based on bilateral iterative algorithm is presented in the Appendix B1.

If the minimal distance at optimality for (35) is not zero, then the probabilistic inversion problem (34) is infeasible, and the optimal measure q makes the response matrix T the "closest" to the expert ranking judgment probability matrix P. On the contrary, if the minimal distance at optimality for (35) is zero, then the probabilistic inversion problem (34) is feasible[4], and it may probably have multiple solutions. Among all feasible solutions, it is preferred to have the minimal Kullback-Leibler distance between the measure q and the uniform measure $\frac{1}{N}$

$$\min_{q_s} \sum_{s=1}^{N} q_s \ln(N q_s) \tag{36}$$

$$\text{s.t. } P = \sum_{s=1}^{N} q_s J^s \ . \tag{37}$$

An intuitive explanation is that if two scenarios map to the same response, then there should be no difference between the probability measures for them.

4. Defending US Urban Areas against Intentional Attacks

4.1. Problem Setting

I now apply the model developed in Section 2.3 to the sample data of 47 US urban areas. I consider four attributes of attractiveness, including expected property losses from terrorism according to Willis et al. (2005), fatalities, populations and population densities Willis et al. (2005). Note that the values are scaled into [0, 1] (See Appendix C1). Following Wang and Bier (2010), I assume that the defender knows the values of those attributes, but is uncertain about how much weight the attacker puts on each one. Moreover, the defender may have ignorance about some unobserved attributes that can be important to the attacker.

In the base case, I assume that the random attribute weights follow the Dirichlet distribution (see Appendix A) with equal expected values (0.2 for each of the attributes, including the unobserved attribute). Changing the value of a single spread parameter in the Dirichlet distribution (see Appendix A), while holding the expected values of the weights constant, enables us to vary the extent of the defender's uncertainty. Furthermore, I allow the existence of unobserved attributes to the attacker, and assume that the unobserved attributes are identically and independently uniform distributions between 0 to 1 for all cities.

4.2. Optimal Defensive Resource Allocations

For the spread parameter in the Dirichlet distribution set at a relatively high uncertainty level $\alpha_0 = 1$, the optimal allocations as a function of the cost effectiveness are presented in Figure 1.

[4] There is another way to show whether the probabilistic inversion problem is feasible or not based on linear programming (LP). The algorithm is presented in the Appendix B2.

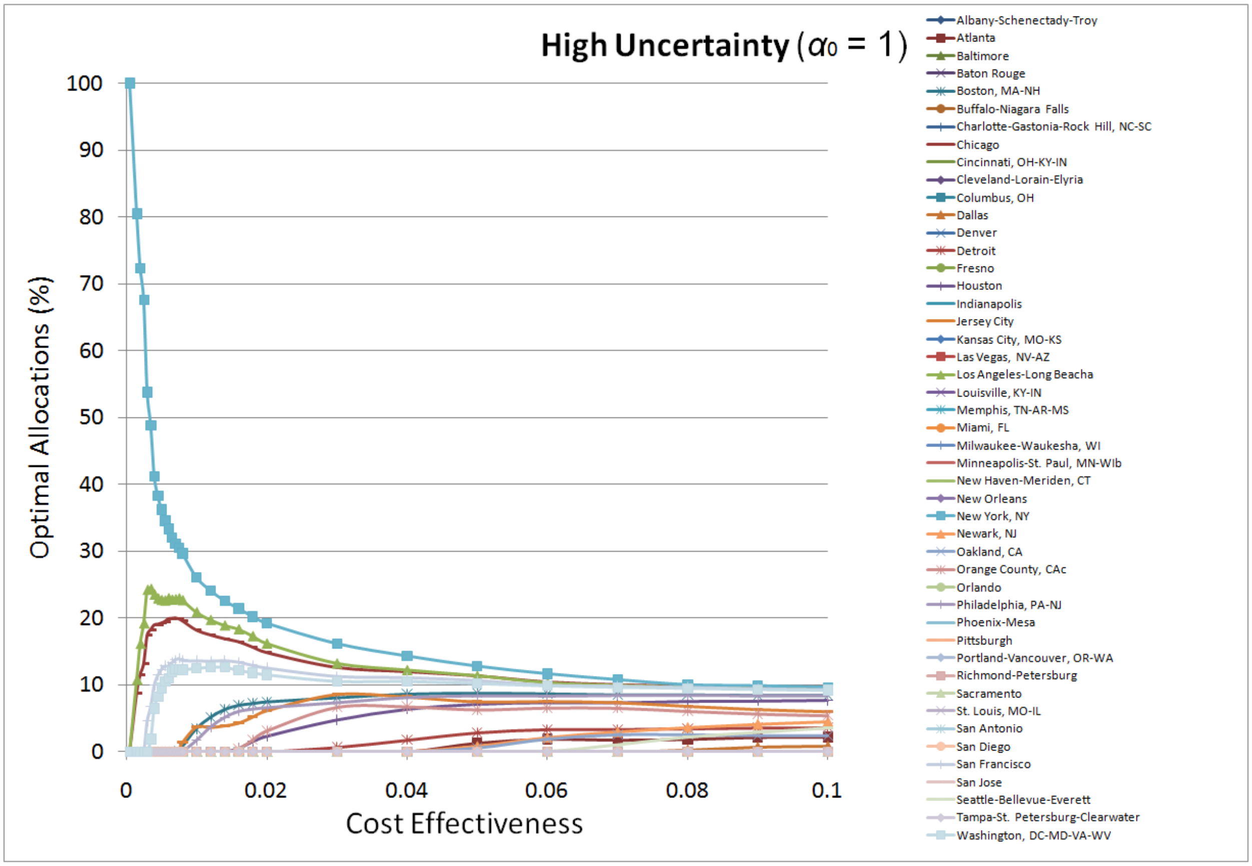

Figure 1 Optimal Defensive Allocations for the Case of Dirichlet Weights

At lower levels of cost effectiveness, the optimal allocation is more spread among all cities than in the case of higher level of cost effectiveness, i.e., more resources will be distributed to small cities. It is due to the fact with higher level of cost effectiveness, less money is adequate to protect the big targets well and more money can be spared to protect the relatively small targets.

4.3. Elicitation of Expert Judgments

As stated in Section 3.3, the method of probabilistic inversion can be used to construct a reasonable prior distribution to represent the decision maker's (PA) uncertainty about attacker preferences among various attributes, including explicit treatment of unobserved attributes that may be important to the attackers, but are not known by the decision maker. Due to the challenges of expert elicitation including the reluctance of experts to provide quantitative estimates, one way is to elicit only rank orders. In particular, the experts are asked to rank the top 20 in terms of attractiveness out of the total 47 urban areas. From the expert rankings, the attacker weights on various attributes and information about the unobserved attributes (e.g., high values on Los Angeles may imply an importance on entertainment industry) are inferred, using probabilistic inversion.

In this case study, I make up data for three groups of 50 experts. Presumably, Group 1: All think that A1 (property) is more important than all other attributes. Furthermore, there should be no important un-quantified attributes.

Group 2: All think that A4 (population density) is more important, and there should be an un-quantified attribute related to entertainment industry.

Group 3: This group is a combination of Group 1 and Group 2, presenting disagreement in judgments.

The joint distributions (simulated scenarios) of attribute weights and unobserved attributes are elicited by the probabilistic inversion. For illustration, only marginal probability distributions (histograms) for the attribute weights are presented:

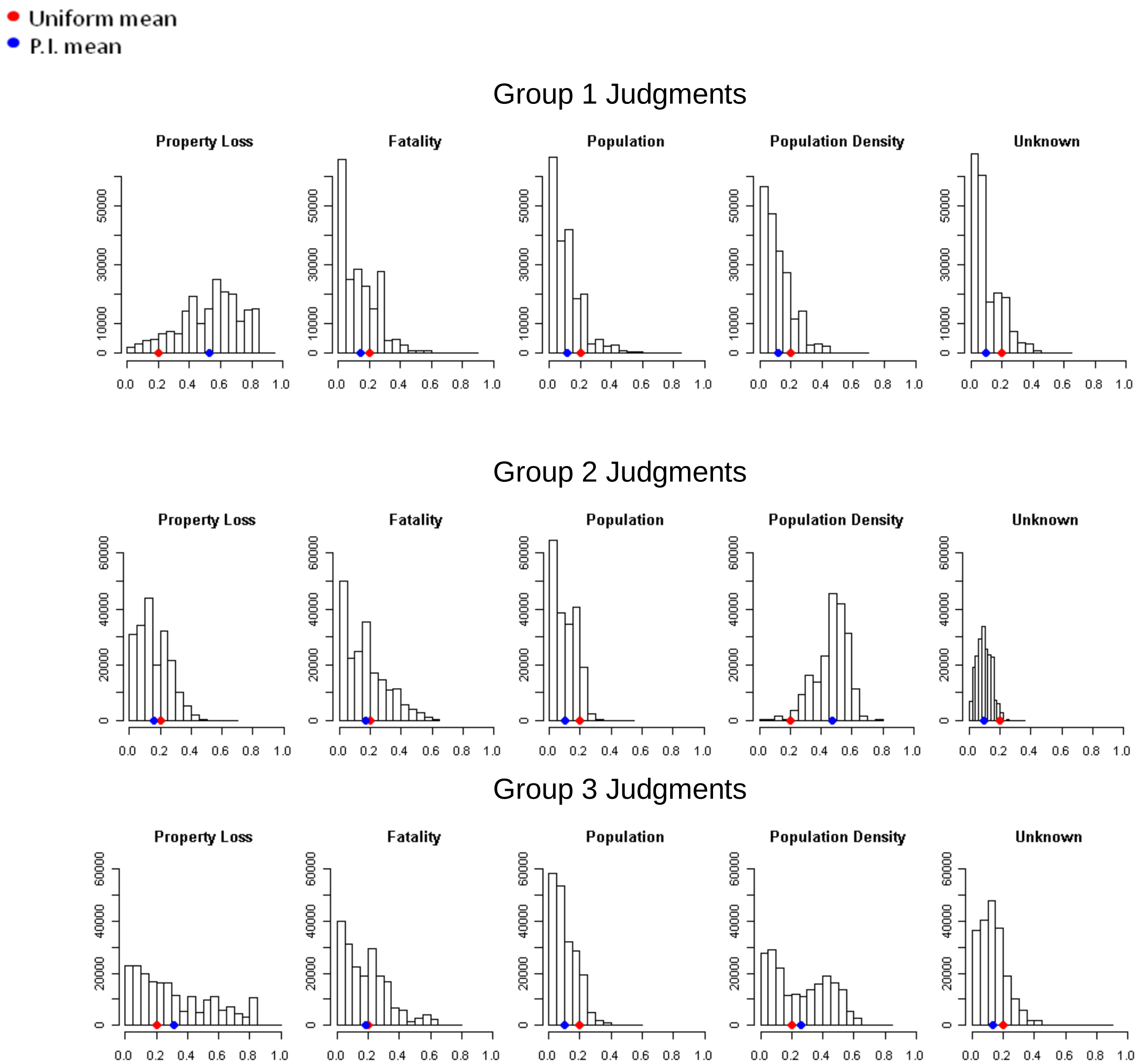

Figure 2 Elicitation of Expert Judgments on Attribute Weights

The results obtained by probabilistic inversion match the presumed assumptions of the make-up data. For Group 1 judgments, the probabilistic inversion weight for the attribute property loss is likely to be high (higher than the uniform mean 0.2). For Group 2 judgments, the probabilistic inversion weight for the attribute population density is highly likely to be around 0.5. For Group 3 judgments, the probabilistic inversion weights may have multi-modal distributions (see especially the weight for population density) and higher variance due to disagreement between experts.

4.4. Minimizing Conditional Value at Risk

This section presents the optimal allocations by minimized CVaR for different quantiles. Figure 3 shows the results using Group 1 expert judgments as input for uncertainties in the quantile optimization model.

In this case the decision maker cares about the extreme values of losses (say, losses greater than 75% quantile), and the corresponding optimal defensive allocations are off the mean value results (corresponding to the CVaR for 0% quantile). In particular, if caring about tail losses, the defender should spend more on the least valuable target among those with positive investment (Houston in this particular problem setting), because such a city is the most likely to be attacked (and therefore the main source of losses) at optimality for the mean value optimization. When dealing with decisions involving extreme events, it is important to further consider the conditional value at risk besides the expected loss as the optimization objectives.

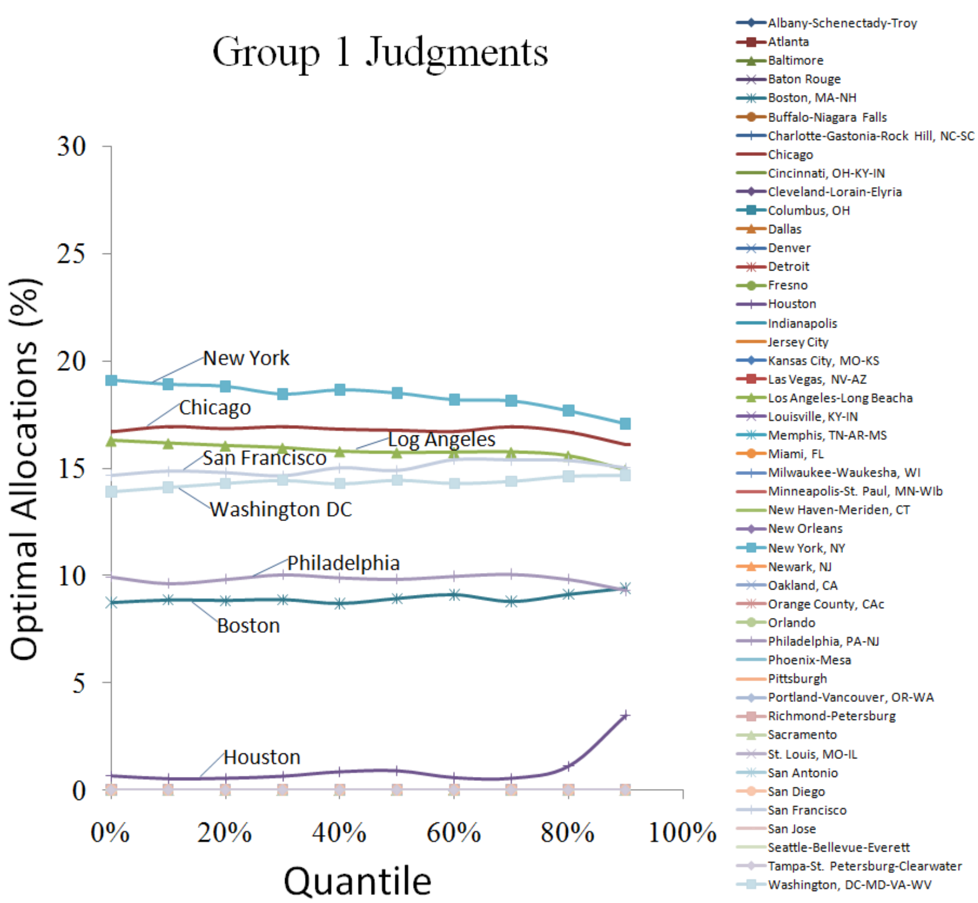

Figure 3 Minimized CVaR for Different Quantiles for Group 1 Expert Judgments

5. Dispatch of Belgian High Voltage Electricity Network

5.1. Problem Setting

Based on the model developed in Section 2.5, this section uses a case study on the Belgian electricity network to illustrate the analysis (Yao et al., 2008). Figure 4 copies the Belgian high voltage electricity network from Yao et al. (2008). All nodes are numbered from 1 through 53. Among them 19 nodes are power plants whose generation capacity is positive; other nodes are transmission nodes. A System Operator (SO) determines how to dispatch electricity at each node of the electricity network to gain maximal social welfare; electricity firms determine generation quantities at each node of power plant to maximize their individual profits. Uncertain shocks may happen from outage of power plants or breakdown of transmission lines. The related parameters in the demand inverse function $P_i(q)$, cost generation function $C_i(q)$, generation capacity $\overline{q}_i$ and thermal limit K_l (copied from Yao et al., 2008) are listed in Appendix C2.

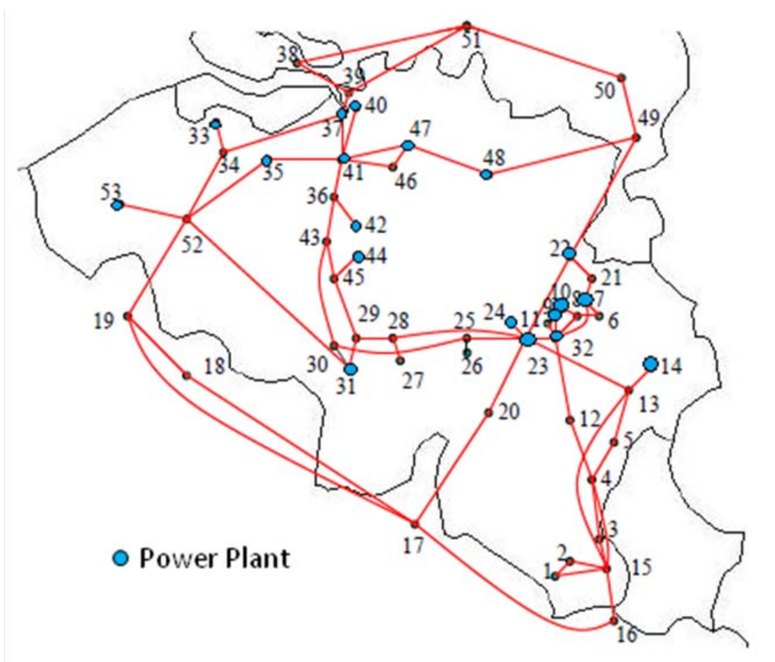

Figure 4 Belgian High Voltage Electricity Network

5.2. Dispatch of Electricity against Outages

In this section I consider only uncertainties about possible power plant outages. Table 1 shows the outage probability of four major power plants.

Table 1 Outage Probability of Generation Nodes

Node	Outage Probability
10	0.03
14	0.03
24	0.04
41	0.04

Taking into account the uncertain outages, then I consider the impact of price caps on the optimal ISO decisions in the electricity network as in (Yao et al., 2008). In the first case there is no price cap, and in the second case a price cap of 400 is imposed at all nodes. Figure 5 shows the optimal ISO decisions on electricity dispatch with and without price caps. Note that the negative dispatch means besides satisfying its own demand, a generation node needs to produce extra amount of electricity to transmit to other nodes. On the other hand, the positive dispatch at a node means that electricity is transmitted into this node to satisfy its demand.

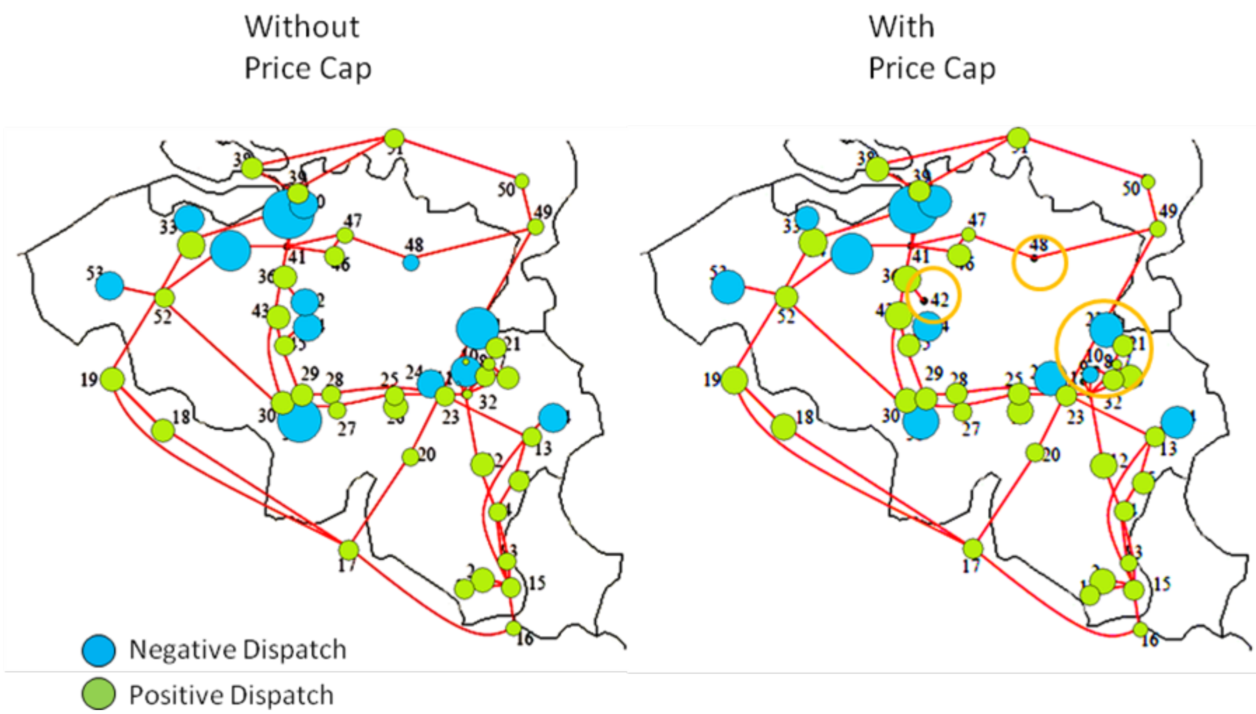

Figure 5 Optimal Dispatch of Electricity against Uncertain Outages

With a price cap, the ISO decision is less powerful in the sense that it can determines on less electricity quantities to transmit over the network. Some power plants which are determined less negative dispatch than in the case without price cap, which means they have to produce less extra electricity for other nodes. It is because price cap leads to higher demand at each node. The generation nodes need to satisfy their own demand first, and leaves less flexibility for the ISO to decide on

5.3. Dispatch of Electricity against Breakdown of Transmission Lines

Breakdown of transmission lines needs to be formulated as random parameters in the constraints. For example, consider the thermal limit of each transmission line. $K_l(\omega_l) = K_l$ represents no breakdown, while $K_l(\omega_l) = 0$ represents breakdown of transmission line l. Unfortunately, computational algorithms for the chance constraints fall out of the class of two-stage STO models and beyond the scope of this paper.

Other uncertain shocks on transmission lines include their “blocking” by big electricity firms. For example, Hogan (1997) discusses a type of “market power” which dominates the stream of electricity on transmission lines. In such cases, externality between agents should be considered in the two-stage STO models.

6. Computational Algorithms

In this section, I explore three algorithms that can be useful in designing effective computational methods for solving the two-stage STO problems, including bilateral descent method (Ermolieva et al., (2010), linear programming (LP) approximation and stochastic quasi-gradient (SQG) method. The first two methods are based on sample average approximation (SAA) of the objective function, each calculation involving the entire set of scenarios. However, the SQG method is an “adaptive Monte Carlo optimization method” (Ermoliev, 2009). When calculating the searching direction, the SQG method just picks one or two scenarios (depending on which format of SQG is used) at a time, which saves computing time and reduces complexity.

Consider a simple version of the two-stage STO problem. The decision maker is to minimize function (1) where X is a simplex defined by (9). Note that this class of problems includes the stochastic minimax, the multi-criteria STO and the quantile optimization. For simplicity, I introduce some assumptions on $f(x, y(x,\omega),\omega)$, so that the optimized objective is well-defined, and has the property that enables all three algorithms to find a global optimum.

Assumption 1 $f(x, y(x,\omega),\omega)$ is a convex function in $x \in X$ for all feasible ω.

Under assumption 1, both the optimization objective function (1) and the sample average approximation (SAA) function (3) are convex. Therefore, local minima obtained by these computational algorithms are also global.

Assumption 2 $f(x, y(x,\omega),\omega)$ is a separable function in x and y, i.e.,

$$f(x, y(x,\omega),\omega) = \sum_{i=1}^{n} f_i(x_i, y_i(x_i,\omega),\omega) \tag{38}$$

This assumption is essential for the bilateral descent method and LP approximation.

Assumption 3 $f(x, y(x,\omega),\omega)$ is continuously differentiable in $x \in X$.

This assumption is very important for the bilateral descent method. Table 2 summarizes the three different algorithms in terms of assumptions and convergence speeds.

Table 2 Comparison of Three STO Algorithms

Bilateral Descent	LP Approximation	Stochastic Quasi-Gradient
Require A1, A2, and A3	Require A1 and A2	Require A1
Converges to SAA optimal solution; Fast convergence to optimal neighborhood dependent on *N*; No limit on the number of scenarios	Converges to SAA solution; Slow convergence to optimal neighborhood dependent on *N*; Constrained by the number of scenarios	Asymptotically converges to real solution; Fast convergence to optimal neighborhood; No limit on the number of scenarios; Applicable to functions with no closed form (e.g., implicit functions)

In the following section I introduce the three algorithms in detail, and present a numerical example for the case study of US defensive resource allocation. Note that the SQG methods are applicable even for discontinuous functions (Ermoliev 2009).

6.1. Bilateral Descent Method

Consider the optimization problem (3), i.e., the minimization of function (3), i.e.,

$$F^N(x) = \frac{1}{N}\sum_{s=1}^{N} f(x, y(x, \omega^s), \omega^s).$$

Suppose that the assumptions 1, 2 and 3 are all satisfied. It follows from the convex analysis that the Lagrange function can be written as

$$L(x,\alpha,\mu) = \frac{1}{N}\sum_{s=1}^{N}\sum_{i=1}^{n} f_i(x_i, y_i(x_i,\omega^s),\omega^s) - \alpha(\sum_{i=1}^{n} x_i) - \sum_{i=1}^{n}\mu_i x_i \quad (39)$$

where $\mu_i \geq 0, i = 1,...,n$. The KKT conditions are given by

$$\frac{1}{N}\sum_{s=1}^{N}\frac{\partial}{\partial x_i} f_i(x_i, y_i, \omega^s)\Big|_{y_i = y_i(x_i,\omega^s)} - \mu_i = \alpha, i = 1,...,n \quad (40)$$

$$x_i \mu_i = 0, i = 1,...,n \quad (41)$$

$$\sum_{i=1}^{n} x_i - B = 0 \quad (42)$$

$$\mu_i \geq 0, i = 1,...,n \quad (43)$$

The bilateral descent method works as the following:

1. Start from a feasible solution x^0 such that $\sum_{i=1}^{n} x_i^{\,0} - B = 0$ and $x_i^{\,0} \geq 0, i = 1,...,n$.

2. At a feasible x^t, identify the index set $I^t{}_+$ of targets where $x_i^t > 0$. Find

$$i^t{}_{\max} = \arg\max_{i \in I^t{}_+} \frac{1}{N} \sum_{s=1}^{N} \frac{\partial}{\partial x_i} f_i(x_i, y_i, \omega^s) \Big|_{y_i = y_i(x_i, \omega^s)} \quad (44)$$

$$i^t{}_{\min} = \arg\min_{i} \frac{1}{N} \sum_{s=1}^{N} \frac{\partial}{\partial x_i} f_i(x_i, y_i, \omega^s) \Big|_{y_i = y_i(x_i, \omega^s)}$$

3. Let Δ be a positive number. Find Δ_t minimizing w.r.t. $\Delta \ge 0$, $x_i \ge 0$, $x_k \ge 0$ function $F^N(x_1^t, ..., x_i^t + \Delta, x_{i+1}^t, ..., x_k^t - \Delta, x_k^t, ..., x_n^t)$, where $i = i_{\min}^t$, $k = i_{\max}^t$. Optimal $\Delta_t \le x_k^t$ defines new approximate solution x^{t+1}. (45)

4. Stop when according to the KKT conditions (39)-(43), $\frac{1}{N} \sum_{s=1}^{N} \frac{\partial}{\partial x_i} f_i(x_i^t, y_i^t, \omega^s) \Big|_{y_i^t = y_i(x_i^t, \omega^s)}$ are equal for all $i \in I^t{}_+$.

The convergence of this procedure can be proven by the convergence of cyclic coordinate descent method (Zangwill, 1969), if assumptions A1, A2 and A3 are satisfied. In practice, the bilateral descent method converges very fast to the optimum. A numerical example will be presented in Section 6.4. Increase of the number of scenarios will not dramatically change the complexity, since it only affects the time of function $F^N(x)$ evaluation.

If the function is not continuously differentiable, then in (37) sub-gradients instead of gradients should be used in calculating the searching direction. Usually the set of sub-gradients are given implicitly, therefore the choice of appropriate sub-gradients is a difficult task; an appropriate choice is also very important for the performance of the algorithm.

6.2. Linear Programming (LP) Approximation

Consider the problem (3):

$$\min_{x \in X} \frac{1}{N} \sum_{s=1}^{N} f(x, y(x, \omega^s), \omega^s)$$

Suppose that the assumptions A1 and A2 are satisfied. The function $f(x, y(x, \omega^s), \omega^s)$ is separable so that $f(x, y(x, \omega^s), \omega^s) = \sum_{i=1}^{n} f_i(x_i, y_i(x_i), \omega^s)$. Since $f_i(x_i, y_i(x_i), \omega^s)$ are convex, then their piece-wise linear approximations are given by

$$f_i(x_i, y_i(x_i, \omega^S), \omega^S) \sim= \max_k \{l_{ik}(x_i, \omega^S)\} \quad (45)$$

for some linear functions

$$l_{ik}(x_i,\omega^s) = a_{iks}x_i + b_{iks}\,. \tag{46}$$

The original SAA problem (3) can be written as a linear programming problem in the extended space

$$\min_{x\in X}\frac{1}{N}\sum_{s=1}^{N}\sum_{i=1}^{n} z_i^{\ s} \tag{47}$$

$$z_i^{\ s} \ge a_{iks}x_i + b_{iks}, \forall i,s,k \tag{48}$$

This method reduces the nonlinear optimization to the linear programming (LP) problems. The assumption of smoothness is not necessary for this reformulation and the global optimum can be obtained through proper LP procedure. Note that the optimal solutions of problems involving inherent and deep uncertainties are not very sensitive to the quality of linear approximation.

However, the number of scenarios is a very critical constraint of the LP reformulation since it increases the number of both variables and constraints, and will affect complexity of the linear program dramatically. A numerical example is presented in Section 6.4.

6.3. Stochastic Quasi-Gradient (SQG) Method

The stochastic quasi-gradient (SQG) method is a type of adaptive Monte Carlo optimization method (Ermoliev, 2009). Instead of using all scenarios at every step, the stochastic quasi-gradient (SQG) method picks just one or two scenarios to calculate the searching direction. For simplicity, here I also assume convexity (A1) so that the local optimum obtained by SQG is also the global optimum.

If the sub-gradient of $f(x,y(x,\omega^s),\omega^s)$ is easy to get, then at each step t of the SQG algorithm updates the current solution can be defined by:

$$x^{t+1} = \Pi_X\{x^t + \rho_t \partial f(x^t, y(x^t,\omega^t),\omega^t)\}\ , \tag{49}$$

where Π_X is the projection onto the feasible region X , and ρ_t is an adaptive step size which changes according to an oscillation measure of the objective values.

If the sub-gradient of $f(x,y(x,\omega^s),\omega^s)$ is hard to get (e.g., implicit functions, black box functions), I can approximate the stochastic quasi-gradient by:

$$\xi^t = \frac{f(x^t+\gamma_t\eta^t, y^t, \omega^{t1}) - f(x^t, y^t, \omega^{t2})}{\gamma_t}\eta^t \tag{50}$$

where η^t is a random unit vector, and γ_t is a small positive number which can be adapted as the algorithm proceeds. Then at each step of the SQG algorithm the update of the current solution is given by:

$$x^{t+1} = \Pi_X\{x^t + \rho_t \xi^t\} \tag{51}$$

Besides (3), the stochastic quasi-gradient (SQG) method can also deal with situations where the probability measure depends on the decision variables, that is

$$Ef(x,y(x,\omega),\omega) = \int f(x,y(x,\omega),\omega)dP(x,\omega) \tag{52}$$

The convergence of the stochastic quasi-gradient (SQG) method is shown in (Ermoliev, 1983; Ermoliev and Wets, 1988). There is no limit on the number of scenarios, because at each step only one or two scenarios are used to calculate the sub-gradient or stochastic quasi-gradient. Compared to the bilateral descent method, the use of sub-gradient here is less strict. At every step, I do not have to choose the "best" sub-gradient in terms of the function value reduction, because of its stochastic characteristics.

One problem of this algorithm is that it usually converges asymptotically very fast to the optimal neighborhood, but takes longer time to converge to the optimal solution. In fact, slow asymptotic convergence with respect to the sample size N is a common feature for any other general STO methods. Steering the procedure by adaptively changing the step size ρ_t will help the convergence. However, the steering is problem-specific, depending on the shape of the objective function and the structure of the probability measure.

A numerical example is presented in Section 6.4.

6.4. A Numerical Example

Three computational methods are implemented to solve the defensive resource allocation problem for 47 US urban areas. (All computational algorithms are presented in Appendix B.) For a selected case where the cost effectiveness is fixed at the level of 0.02, the performances of three methods for a case of 100,000 scenarios (common random numbers are used for comparison) are compared in Table 3.

Table 3 Numerical Comparison of Three STO Algorithms

	Bilateral Descend	LP Approximation	STO Quasi-Grad.
Iteration	191	1,237,000	463
CPU Time	116s	10,710s	160s
Opt. Function Value*	659.9	663.3	660.7

* For a selected case

The bilateral descent method and stochastic quasi-gradient (SQG) are fast, while the LP approximation method takes much more time. It is because for the LP Approximation method the number of variables and constraints grows dramatically along with increasing number of scenarios.. For example, for this particular problem of 100,000 scenarios, the LP approximation problem has 18,800,002 rows, 100,048 columns and 37,700,048 non-zeroes. After the complex presolver it still has 100,047 rows, 18,800,001 columns, and 37,600,047 non-zeros.

Furthermore, the three developed methods get very close optimal solutions. See Figure 6 for the resulting optimal resource allocations as a function of cost effectiveness obtained by three computational methods for a case of 10,000 scenarios (when computing time is bearable for all three methods). The non-smoothness for SQG algorithm is because that the SQG method converges to the asymptotically optimal neighborhood according to the original random process, rather than the optimal value for the SAA problem (used in the bilateral descent method and LP approximation).

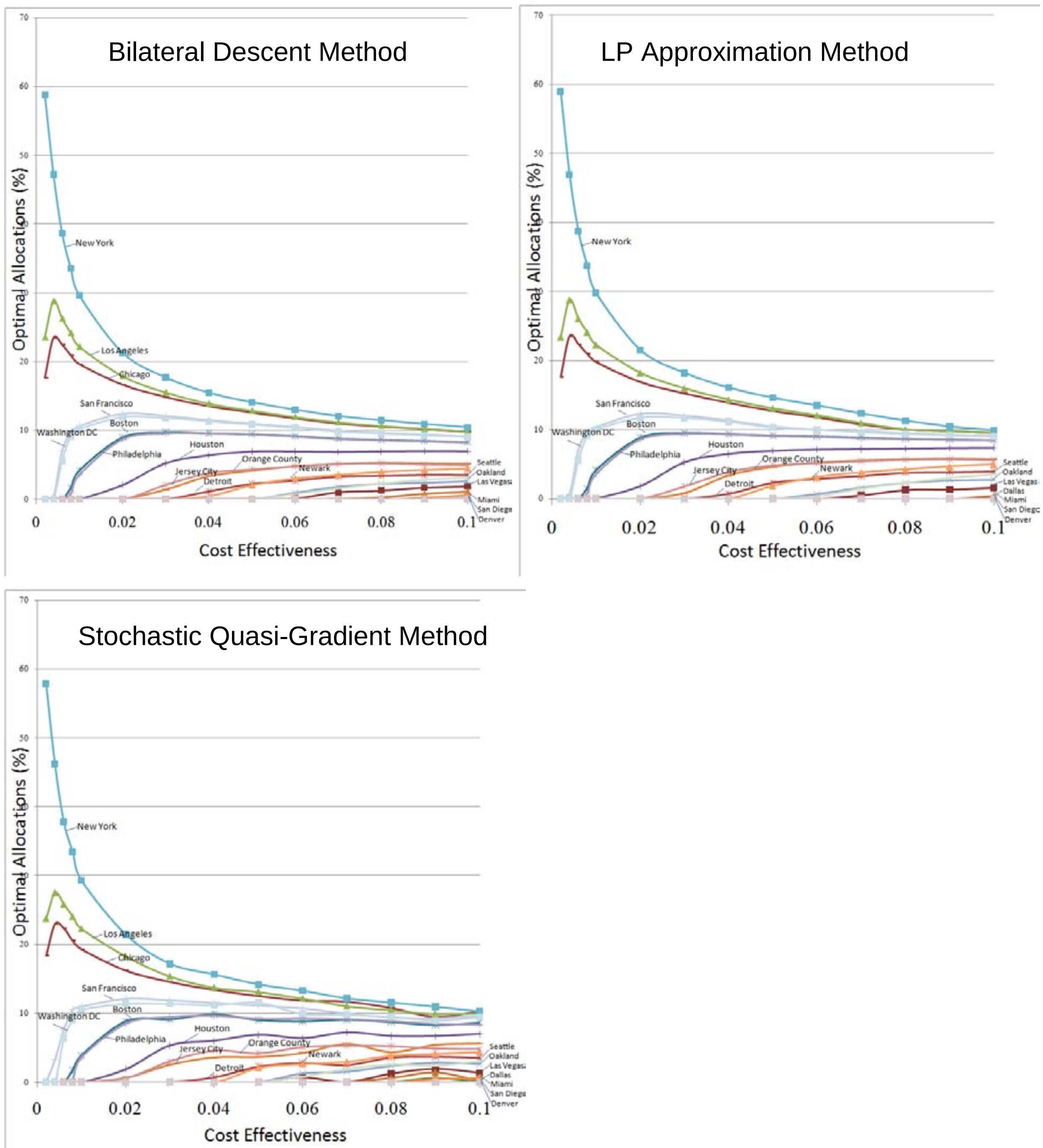

Figure 6 Optimal Solutions Obtained by Three STO Algorithms

7. Conclusion

This paper documents a powerful framework for designing robust decisions against uncertain threats to public goods generated by multiple agents. Two case studies from very different application areas show that this framework can be applied to a quite broad class of problems with decisions against uncertain and adaptive agent responses.

This framework can deal with both direct and indirect judgments on the decision maker's perception about uncertain agent behaviors, either directly by probability density estimation, or indirectly by probabilistic inversion. The quantified distributions are treated as input to the stochastic optimization (STO) models. Robust decisions can then be obtained against all possible consequences, especially extreme consequences.

The available computational methods are explored and implemented successfully, showing that the framework is ready for a wide range of actual applications.

References

Artzner, P., F. Delbaen, J. Eber, D, Heath, 1997. Thinking Coherently. Risk 10 (11), 68-71.

Artzner, P., F. Delbaen, J. Eber, D, Heath, 1999. Coherent Measures of Risk. Mathematical Finance 9 (3), 203-228.

Athan, T.W., 1994. A Quasi-Monte Carlo Approach to Multicriteria Optimization. Doctoral Dissertation, Department of Mechanical Engineering and Applied Mechnics, University of Michigan, Ann Arbor, Michigan.

Bier, V. M., S. Oliveros, L. Samuelson. 2007. Choosing what to protect: Strategic defensive allocation against an unknown attacker. Journal of Public Economic Theory 9 (4) 563-587.

Bordley, R.F., S.M. Pollock, 2009. A Decision-Analytic Approach to Reliability-Based Design Optimization. Operations Research 57 (5), 1262-1270.

Chiappori, PA., 1992. Collective Labor Supply and Welfare. The Journal of Political Economy 100 (3).

Du, C., D. Kurowicka, R.M. Cooke, 2006. Techniques for Generic Probabilistic Inversion. Computational Statistics and Data Analysis 50 (5), 1164-1187.

Ermolieva T, Ermoliev Y, Fischer G, Jonas M, Makowski M, Wagner F (2010).

Carbon emission trading and carbon taxes under uncertainties. Climatic Change, 103(1-2).

Ermoliev, Y., 1983. Stochastic Quasigradient Methods and Their Application to Systems Optimization. Stochastics 9, 1–36.

Ermoliev, Y., R.J.Wets(Eds.), 1988. Numerical Techniques for Stochastic Optimization. Springer-Verlag, Berlin.

Ermoliev, Y., 2009. Stochastic Quasigradient Methods: Applications. C.A. Floudas, P.M. Pardalos (Eds.) Encyclopedia of Optimization.

Ermoliev, Y., 2009. Stochastic Quasigradient Methods in Minimax Problems. C.A. Floudas, P.M. Pardalos (Eds.) Encyclopedia of Optimization.

Ermoliev, Y., D. von Winterfeldt, 2010. Risk, Security and Robust Solutions, working paper.

Fienberg, S. E., 1970. An Iterative Procedure for Estimation in Contingency Tables. Annals of Mathematical Statistics 41 (3): 907–917.

Hogan, W., 1997. A Market Power Model with Strategic Interaction in Electricity Networks. The Energy Journal 18 (4), 107-142.

Kraan, B., T. Bedford, 2005. Probabilistic Inversion of Expert Judgments in the Quantification of Model Uncertainty. Management Science 51 (6), 995-1006.

Koenker, R., G. Bassett, 1978. Regression Quantiles. Econometrica 46 (1), pp. 33-50.

Laffont, JJ., D. Martimor, 2002. The Theory of Incentives: The Principal-Agent Model.

Michelot, C., 1986. A finite algorithm for finding the projection of a point onto the canonical simplex of $\propto^n$. Journal of Optimization Theory and Applications 50 (1), 195-200.

Miller, B.L., H.M. Wagner, 1965. Chance Constrained Programming with Joint Constraints. Operations Research 13 (6), 930-945.

Wang, C., V.M. Bier, 2010. Impact of Intelligence on Target-Hardening Decisions Based on Multi-Attribute Utility Functions. Submitted to Decision Analysis.

Willis, H. H., A. R. Morral, T. K. Kelly, J. J. Medby. 2005. Estimating terrorism risk. RAND Corporation MG388.

Yao, J., I. Adler, S.S. Oren, 2008. Modeling and Computing Two-Settlement Oligopolistic Equilibrium in a Congested Electricity Network. Operations Research 56 (1), 34-47.

Appendix A: Dirichlet Distribution

The probability density function of the Dirichlet distribution is given by

$$g(x_1, \ldots, x_m) \sim \text{Dirichlet}\,(\alpha_1, \ldots, \alpha_m) = \frac{\Gamma(\alpha_0)}{\prod_{j=1}^{m}\Gamma(\alpha_j)}\prod_{j=1}^{m} x_j^{\alpha_j - 1}$$

where $\Gamma(z) = \int_0^{\infty} t^{z-1}e^{-t}dt$, $\alpha_j > 0$ for $j = 1,\ldots,m$, $\alpha_0 = \sum_{j=1}^{m}\alpha_j$, $x_j > 0$, and $x_m = 1 - \sum_{j=1}^{m-1} x_j$. In this case, the means and variances of the attribute weights x_j are given by $E[x_j] = \frac{\alpha_j}{\alpha_0}$ and $Var[x_j] = \frac{\alpha_j(\alpha_0 - \alpha_j)}{\alpha_0{}^2(\alpha_0 + 1)}$, respectively.

Appendix B: Computational Algorithms

B1 Probabilistic Inversion: Convex Optimization

The optimization objective is to minimize

$$\min_{q \in X} F(q) = \frac{1}{N}\sum_r \sum_i \left(\sum_{s=1}^{N} q_s J_{ri}^{s}\right)\log\left(\frac{\sum_{s=1}^{N} q_s J_{ri}^{s}}{P_{ri}}\right)$$

where $X = \{q \in R^N \mid \sum_{s=1}^{N} q_s = 1, q_s \geq 0, s = 1,\ldots,N\}$.

The computational algorithm is as follows.

1. Start with a feasible $q^0 \in X$
2. Given the current $q^t \in X$,

   $q^{t+1} = \Pi_X\{q^t - \rho\nabla F\,|_{q=q^t}\}$

   where Π_X is the projection onto the simplex X, which can be realized according to (Michelot, 1986), and the gradient

$$\nabla F = \left[\sum_r \sum_i J_{ri}^s \left\{ 1 + \log\left(\frac{\sum_{s'} q_{s'} J_{ri}^{s'}}{P_{ri}} \right) \right\} \right]_{s\times 1}$$

3. Stop when the objective converges.

B2 Probabilistic Inversion: Linear Programming (LP)

The linear programming designed to check feasibility of the probabilistic inversion problem is given by

$$\begin{aligned}
&\min_{q,z} z \\
&z \geq \sum_s q_s J_{ri}^s - P_{ri}, \forall r, i \\
&z \geq P_{ri} - \sum_s q_s J_{ri}^s, \forall r, i \\
&q_s \geq 0, \forall s \\
&\sum_s q_s = 1
\end{aligned}$$

B3 Bilateral Descent Method for Case Study I

The optimization objective is to minimize

$$\begin{aligned}
&\min_{x,z} \frac{1}{N} \sum_{s=1}^{N} z_s \\
&z_s \geq g_i(x_i) u_i(\omega^s), \forall i, s \\
&x_i \geq 0, \forall i \\
&\sum_i x_i = B
\end{aligned}$$

where $g_i(x_i) = e^{-\lambda_i x_i}$. This is a convex optimization problem. The Lagrange function is given by

$$L(x, z, \alpha, \mu) = \frac{1}{N} \sum_{s=1}^{N} z_s + \frac{1}{N} \sum_i \sum_{s=1}^{N} \alpha_{is} \left\{ g_i(x_i) u_i(\omega^s) - z_s \right\} + \mu \left(\sum_i x_i - B \right)$$

The KKT conditions are

$$\sum_i \alpha_{is} = 1$$
$$\alpha_{is} \geq 0$$
$$\alpha_{is} \left\{ e^{-\lambda_i x_i} u_i(\omega^s) - z_s \right\} = 0$$
$$x_i \left\{ \lambda_i \left(\frac{1}{N} \sum_s \alpha_{is} u_i(\omega^s) \right) e^{-\lambda_i x_i} - \mu \right\} = 0$$
$$\sum_i x_i = B$$
$$x_i \geq 0$$

The computational algorithm is as follows.

1. Start with feasible x^0 such that $x_i^0 \geq 0$ and $\sum_i x_i^0 = B$.

2. Given the current x^t, calculate

   $z_s^t = \max_i e^{-\lambda_i x_i^t} u_i(\omega^s)$ for all s.

   Let $I_{\max}^{s,t} = \{i \mid e^{-\lambda_i x_i^t} u_i(\omega^s) = z_s^t\}$, and

   $$\alpha_{is}^t = \begin{cases} \dfrac{1}{\left|I_{\max}^{s,t}\right|} & i \in I_{\max}^{s,t} \\ 0 & o.w. \end{cases}$$

3. Let $I_{positive}^t = \{i \mid \sum_s \alpha_{is}^t > 0\}$. If $i \notin I_{positive}^t$, then $x_i^{t+1} = 0$.

   Find $i^+ = \arg\max_{i \in I_{positive}^t} \lambda_i \left(\frac{1}{N} \sum_s \alpha_{is}^t u_i(\omega^s) \right) e^{-\lambda_i x_i^t}$ and $i^- = \arg\min_{i \in I_{positive}^t} \lambda_i \left(\frac{1}{N} \sum_s \alpha_{is}^t u_i(\omega^s) \right) e^{-\lambda_i x_i^t}$.
   Then

   $x_{i^+}^{t+1} = x_{i^+}^t + \Delta$ and $x_{i^-}^{t+1} = x_{i^-}^t - \Delta$.

4. Go back to 2.

5. Stop when $\lambda_i \left(\frac{1}{N} \sum_s \alpha_{is}^t u_i(\omega^s) \right) e^{-\lambda_i x_i^t}$ = constant for all $i \in I_{positive}^t$.

B4 Linear Programming (LP) Approximation for Case Study I

Approximate $g_i(x_i) = e^{-\lambda_i x_i} \sim= \max_k a_{ik} x_i + b_{ik}$, then the linear programming (LP) approximation problem is given by

$$\min_{x,z} \frac{1}{N}\sum_{s=1}^{N} z_s$$
$$z_s \geq (a_{ik}x_i + b_{ik})u_i(\omega^s), \forall i, s, k$$
$$x_i \geq 0, \forall i$$
$$\sum_i x_i = B$$

B5 Stochastic Quasi-Gradient (SQG) Method for Case Study I

The optimization objective is given by

$$\min_{x \in X} F(x) = E \max_i g_i(x_i)u_i(\omega^s)$$

where $X = \{x \in R^n \mid \sum_{i=1}^{n} x_i = 1, x_i \geq 0, i = 1,..., n\}$. The stochastic quasi-gradient method is as follows.

1. Start with a feasible $x^0 \in X$.
2. Given the current $x^t \in X$, sample a new scenario ω^t, calculate
   $$x^{t+1} = \Pi_X \left\{x^t - \rho_t \partial F^t(x)|_{x=x^t}\right\}$$
   where Π_X is the projection onto the simplex. The sub-gradient
   $$\partial F^t(x) = \begin{cases} \frac{d}{dx_i} g_i(x_i)u_i(\omega^t) & i = \arg\max g_i(x_i)u_i(\omega^t) \\ 0 & o.w. \end{cases}$$
3. Stop when $\frac{1}{t}\sum_{s=1}^{t} \max_i g_i(x_i^t)u_i(\omega^s)$ converges.

B6 Stochastic Quasi-Gradient (SQG) Method for CVaR

The optimization objective is to minimize

$$\min_{x \in X, z} E\left[(1-q)z + \max\left\{f(x,\omega) - z, 0\right\}\right]$$

The computational algorithm is given by

1. Start with a feasible $x^0 \in X$ and z^0.
2. Given the current $x^t \in X$ and z^t, sample a new scenario ω^t
   $$x^{t+1} = \Pi_X \left\{x^t - \rho_t \xi^t\right\}$$

where $\xi^t = \begin{cases} \partial f(x,\omega^t)|_{x=x^t} & f(x^t,\omega^t) \le z^t \\ 0 & o.w. \end{cases}$

$$z^{t+1} = z^t - \rho_t \begin{cases} 1-q & f(x^t,\omega^t) \le z^t \\ -q & o.w. \end{cases}$$

3. Stop when $\frac{1}{t}\sum_{s=1}^{t}(1-q)z^t + \max\{f(x^t,\omega^s) - z^t, 0\}$ converges.

Appendix C: Data for Case Studies

C1 Attribute Values for US Urban Areas

No.	Urban Area	Property Loss	Fatality	Population	Population Density
1	Albany-Schenectady-Troy	0.0392	0.0002	0.1966	0.1845
2	Atlanta	0.2055	0.2442	0.6423	0.3959
3	Baltimore	0.2134	0.1131	0.4718	0.5201
4	Baton Rouge	0.0198	0.0000	0.1399	0.2480
5	Boston, MA-NH	0.8347	0.9093	0.5733	0.7174
6	Buffalo-Niagara Falls	0.0952	0.0582	0.2536	0.4289
7	Charlotte-Gastonia-Rock Hill, NC-SC	0.1042	0.0198	0.3126	0.2832
8	Chicago	1.0000	1.0000	0.8736	0.7064
9	Cincinnati, OH-KY-IN	0.0861	0.0198	0.3374	0.3091
10	Cleveland-Lorain-Elyria	0.2592	0.0952	0.4303	0.4642
11	Columbus, OH	0.0676	0.0100	0.3196	0.3075
12	Dallas	0.1894	0.2592	0.5851	0.3474
13	Denver	0.2212	0.1975	0.4098	0.3434
14	Detroit	0.3430	0.3161	0.6706	0.5747
15	Fresno	0.0198	0.0001	0.2060	0.0819
16	Houston	0.6671	0.8347	0.6481	0.4111
17	Indianapolis	0.0676	0.0198	0.3309	0.2897
18	Jersey City	0.3560	0.3297	0.1412	0.9999
19	Kansas City, MO-KS	0.1042	0.0100	0.3585	0.2187
20	Las Vegas, NV-AZ	0.3363	0.1813	0.3235	0.0296
21	Los Angeles-Long Beacha	0.9666	0.9666	0.9074	0.8276
22	Louisville, KY-IN	0.0582	0.0001	0.2262	0.3101
23	Memphis, TN-AR-MS	0.0488	0.0001	0.2472	0.2469
24	Miami, FL	0.2366	0.0952	0.4307	0.5804

25	Milwaukee-Waukesha, WI	0.1042	0.0198	0.3128	0.5374
26	Minneapolis-St. Paul, MN-WIb	0.2366	0.0769	0.5239	0.3075
27	New Haven-Meriden, CT	0.1042	0.0040	0.1268	0.6116
28	New Orleans	0.0769	0.0198	0.2843	0.2558
29	New York, NY	1.0000	1.0000	0.9026	0.9978
30	Newark, NJ	0.5181	0.5507	0.3985	0.6197
31	Oakland, CA	0.3297	0.1813	0.4502	0.7081
32	Orange County, CAc	0.3093	0.3297	0.5091	0.9331
33	Orlando	0.0582	0.0198	0.3371	0.2976
34	Philadelphia, PA-NJ	0.8775	0.8347	0.7206	0.6293
35	Phoenix-Mesa	0.1730	0.0198	0.5565	0.1540
36	Pittsburgh	0.0952	0.0198	0.4455	0.3178
37	Portland-Vancouver, OR-WA	0.1813	0.0198	0.3809	0.2485
38	Richmond-Petersburg	0.0392	0.0002	0.2205	0.2239
39	Sacramento	0.0676	0.0198	0.3344	0.2586
40	St. Louis, MO-IL	0.1894	0.1131	0.4784	0.2631
41	San Antonio	0.0392	0.0100	0.3284	0.3018
42	San Diego	0.2442	0.1813	0.5051	0.3950
43	San Francisco	0.9967	0.9918	0.3513	0.7216
44	San Jose	0.1563	0.0769	0.3434	0.6239
45	Seattle-Bellevue-Everett	0.4883	0.5507	0.4532	0.3360
46	Tampa-St. Petersburg-Clearwater	0.0861	0.1813	0.4506	0.5051
47	Washington, DC-MD-VA-WV	0.9727	0.9970	0.7079	0.4328

C2 Parameters for Belgian High Voltage Electricity Network

Node	Demand Slope	Marginal Cost	Capacity	Node	Demand Slope	Marginal Cost	Capacity
1	1		0	28	1		0
2	0.82		0	29	0.93		0
3	1.13		0	30	0.85		0
4	1		0	31	1	180	712
5	0.93		0	32	1	580	95
6	0.85		0	33	0.88	20	496
7	1	450	70	34	0.5		0
8	1		0	35	1	250	1053
9	0.88	180	460	36	0.73		0
10	0.9	180	121	37	1	100	1399
11	1	200	124	38	0.85		0

12	0.73		0	39	1		0
13	1		0	40	1.15	100	1378
14	0.85	130	1164	41	1	210	522
15	1		0	42	0.79	180	385
16	1.3		0	43	0.68		0
17	1		0	44	1.03	200	538
18	0.79		0	45	1		0
19	0.68		0	46	1		0
20	1.05		0	47	1	100	32
21	1		0	48	0.73	220	258
22	1.1	190	602	49	1.2		0
23	1		0	50	1.5		0
24	0.75	100	2985	51	1		0
25	1		0	52	1		0
26	0.8		0	53	1	200	879
27	1.13		0				

This table is copied from Yao et al. (2008).