Working Paper 09-60
Statistics and Econometrics Series 19
September 2009

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

CONTROLLING THE INTERNATIONAL STOCK POLLUTANT WITH POLICIES DEPENDING ON TARGET VALUES

Omar J. Casas[1] and Rosario Romera[2]

Abstract

In this paper a stochastic dynamic game formulation of the economics of international environmental agreements on the transnational pollution control, when the environmental damage arises from stock pollutant that accumulates, for accumulating pollutants such as CO2 in the atmosphere is provided. To improve the non-cooperative equilibrium among countries, we propose a different criterion to the minimization of the expected discounted total cost. Moreover, we consider Cooperative versus Non-cooperative Stochastic Dynamic Games formulated as Markov Decision Processes (MDP). We propose a new alternative where the decision-maker wants to maximize the probability that some total performance of the dynamical game does not exceed a target value during a fixed period of time. The task requirements are therefore formulated as probabilities rather than expectations. This approach is different from the standard MDP, which uses performance criteria based on the expected value of some index. We present properties of the optimal policies obtained under this new perspective.

Keywords: Stochastic Optimal Control, Markov Decision Processes, Stochastic Dynamic Programming, Stochastic Dynamic Games, International Pollutant Control, Environmental Economics, Sustainability, Probability Criterion

JEL Classification: C610, C630, C730, C44, D70, Q20

[1] O. Casas, Statistics Department, Universidad Carlos III Madrid, Calle Madrid 126, 28903 Getafe, Spain, e-mail: omar.casas@uc3m.es. Corresponding author.
[2] R. Romera, Statistics Department, Universidad Carlos III Madrid, Calle Madrid 126, 28903 Getafe, Spain, e-mail: rosario.romera@uc3m.es.
The authors acknowledge financial support from the Spanish Ministry of Education and Science, research projects SEJ2004-03303 and SEJ2007-64500.

Controlling the International Stock Pollutant with Policies Depending on Target Values

Omar J. Casas and Rosario Romera *

July 22, 2009

Abstract

In this paper a stochastic dynamic game formulation of the economics of international environmental agreements on the transnational pollution control, when the environmental damage arises from stock pollutant that accumulates, for accumulating pollutants such as CO_2 in the atmosphere is provided. To improve the non-cooperative equilibrium among countries, we propose a different criteria to the minimization of the expected discounted total cost. Moreover, we consider Cooperative versus Non-cooperative Stochastic Dynamic Games formulated as Markov Decision Processes (MDP). We propose a new alternative where the decision-maker wants to maximize the probability that some total performance of the dynamical game does not exceed a target value during a fixed period of time. The task requirements are therefore formulated as probabilities rather than expectations. This approach is different from the standard MDP, which uses performance criteria based on the expected value of some index. We present properties of the optimal policies obtained under this new perspective.

Keywords: Stochastic Dynamic Games, Markov Decision Processes, Probability Criterion.

AMS Classification: 93E03, 91A15, 91B26

---

*Authors address: Statistics Department. Universidad Carlos III de Madrid, c/ Madrid, 126, 28903 Getafe (Madrid), Spain. E-mails: omar.casas@uc3m.es and rosario.romera@uc3m.es. The authors acknowledge financial support from the Spanish Ministry of Education and Science, research projects SEJ2004-03303 and SEJ2007-64500. An earlier version of this paper was presented at the Monte Verità Conference on Sustainable Resource Use and Economic Dynamics, held in Ascona, Switzerland, June 2-5, 2008.

1 Introduction

In the last years, the theory on international environmental agreements (IEA) and the prospect of climate change has motivated many game theoretic studies, often focused on cooperation and core solutions.

The need of cooperation amongst the countries involved, if a social optimum is to be achieved, has already been addressed in the literature in terms of Game Theory concepts; see e.g. Barrett (2003), Finus (2001), Flam (2006) for a review on these topics. With a few exceptions, this literature works with simple static models of pollution despite the fact that many of the important environmental problems, as climate change, the depletion of the ozone layer or the acid rain problem, are caused by a stock pollutant. However, the stock of pollution may change in the course of the game, as a result of a positive rate of natural decay and emissions of the countries. Thus, the presence of a stock pollutant leads to a dynamic game that is not strictly repeated.

In the framework of a deterministic cooperative game with a dynamic, multi-regional integrated assessment model, Eyckmans and Tulkens (2003) calculates the optimal path of abatement and aggregated discounted welfare for each region. They apply the transfer scheme advocated by Chander and Tulkens (1997), the idea of surplus sharing is used for determining the transfer scheme, and they compute all possible partial agreement Nash equilibria. They found that allocation in the full cooperation lies in the core of the emission abatement game under this specific transfer scheme.

The transfer schemes are based on a single year for assigning the permits or shares in the surplus. Such static schemes are also often observed in reality, e.g. the reduction targets in the Kyoto Protocol are designed as reduction compared to 1990 levels. These static schemes, however, do not take into account that the future growth paths of emissions are expected to diverge substantially between regions. This leads to assignments where historically large emitters obtain relatively large shares of the permits/surplus, while fast-growing developing countries, as China or India, obtain relatively small shares. This leads to increasing burdens on these developing countries to reduce theirs emissions; a notion brought forward by many developing countries in their argumentation on why they do not agree on any reduction targets in the Kyoto Protocol.

The role of transfer in the analysis of self-enforcing international environmental agreements (IEA) was developed in Carraro, Eyckmans, and Finus (2006). They propose transfers using internal and external financial resources for making welfare optimal agreements. To illustrate the relevance of their transfer scheme, they use a stylized integrated assessment simulation model of climate change to show how appropriate transfers may induce almost all countries into signing a self-enforcing climate treaty.

The studies by Germain, Toint, Tulkens, and de Zeeuw (2003) have addressed the issue of how many countries will be interested in signing an IEA with stock pollutant, adopting a cooperative game-theory approach. They extend the result established by Chander and Tulkens (1995) and (1997) for flow pollutants to the larger context of closed-loop (feedback) dynamic games with a stock pollutant. In this context, cooperation is negotiated at each period but financial transfers provide incentives to the countries that ensures the implementation of the grand coalition at each period. Their model thus yields a sequence of full cooperative international agreements so that full cooperation is also achieved in a dynamic setting with a stock pollutant.

Another paper related with this issue using a cooperative game-theory approach is Petrosjan and Zaccour (2003). However, in this paper the authors assume that all the countries decide to cooperate at the initial time-consistent decomposition of each player's total cost, as given by Shapley value, so that the countries stick at each moment to the full cooperative solution agreed at initial time, supposing that the global allocation problem has been solved. Nevertheless, there are only a few attempts in the stock pollutant control literature modelling that issue in a stochastic control framework.

Stochastic Programming is considered by Dechert and O'Donnell (2006) in a particular application that explore some fundamental issues of the optimal level of pollution in a lake with competing uses, they show how the model can be interpreted as an open loop dynamic game, where the control variables are the levels of phosphorus discharged into the watershed of the lake, the state of the system is the accumulated level of phosphorus in the lake and the random shock (a multiplicative noise factor on the control variables of the players) is the rainfall that washes the phosphorus in the lake.

The use of stochastic control models to develop climate-economy models has been advocated by Haurie and Viguier (2003) to represent the possible competition between Russia and China on the international market of carbon emissions permits, their model includes a representation of the uncertainty concerning the date of entry of developing countries on this market in the form of an event tree. Also by Bahn, Haurie, and Malhamé (2008), they show how a piecewise deterministic stochastic control model, over an infinite time horizon, can be used as a paradigm for the design of efficient climate policy, their model recognizes the existing uncertainty concerning the true sensitivity of climate, and the fact that the solution to the climate change issue may reside in the introduction of new carbon-free technologies. Keller, Bolker, and

Bradford (2004) have already explored the combined effects of uncertainty and learning about a climate threshold (an uncertain ocean thermohaline circulation collapse) in an economic optimal growth model.

The stability of an International Environmental Agreement among n countries that emit pollutant are studied using differential games, defined in continuous time, by Jorgensen, Martín-Herrán, and Zaccour (2003) and (2004), Rubio and Casino (2005), among others.

As far as we know, none stochastic formulation for the finite and discrete horizon dynamic analysis of international environmental agreements on transnational pollution control has been introduce as an extension of the issues presented in Germain, Toint, Tulkens, and de Zeeuw (2003). We adopt this point of view because to consider randomness on the factors in the model is closer to reality (see Casas and Romera (2005) and Casas and Romera (2009)).

In this paper, we propose an alternative scheme for the non-cooperative game based on probability criteria for which a core property is proved analytically in a stochastic dynamic (closed-loop) game theoretic context. The cooperative and non-cooperative models are formulated as a MDP with constraints for which alternative criteria are considered. Although optimal policies under the standard criteria (expectation optimality criteria) are computationally simple and useful for many real-life problems, the optimal policies obtained are not reliable when considering a simple or a few decision process, since only the average performance over many trials is guaranteed to be optimal. Therefore, our approach considers that the decision-maker wants to maximize the probability that some performance (cost) of the dynamical system does not exceed a target value during the fixed period of time $[0; T]$. It follows that our problem is essentially different from these classical MDP models, see e.g. Puterman (2005), Altman (1999) and Krass and Vrieze (2002). The main results obtained are: (i) the optimal value function are distribution functions of the target value, and (ii) there exists an optimal deterministic Markov policy. These models proposed are directly linked with the Kyoto or post-Kyoto agreement mechanisms.

Probability criterion for MDP is considered by Wu and Lin (1999) that studies the minimizing risk problems with countable state space and reward set, Filar and Petrosjan (2000), Boda, Filar, Li, and Spanjers (2004) for the problem of optimal control of a retirement investment fund over a finite horizon with a target hitting time criteria, among other authors. Stochastic optimization problems should be solved by Stochastic Dynamic Programming Techniques (see Bertsekas (2000)).

In this paper we propose a new alternative model with probability performance criteria, where the decision-maker wants to maximize the probability that some total performance of the dynamical game does not exceed a target value (X) during a fixed period of time. The task requirements are therefore formulated as probabilities rather than expectations. This approach is different from the standard MDP, which use performance criteria based on the expected value of some index. Moreover, we obtain the existence of an optimal policy and we present properties of the optimal policies obtained under this new perspective. In absence of international cooperation, these optimal policies obtained under this new perspective could be an alternative behavior for each country which finally will help reducing the international stock pollutant. The target value (X) could be chosen by each country, according some particular negotiation. Usually the target value (X) should be a quantity ranging between the non-cooperative value function and the cooperative value function. These schemes are also suitable in the context of coalitional rationality.

The paper is organized as follows: In Section 2 we present the international stock pollutant model with its components, the cost functional components and their elements, the description of the modes of countries behaviour, the stock pollutant control models, cooperative and non cooperative, and the underlaying Markov Decision Process (MDP). In Section 3, we report our new proposal of non cooperative model with policies depending on target values, with necessary definitions. In Section 4, we present an algorithm which computes optimal value functions, optimal action sets, and optimal policies for a finite horizon model. In Section 5, we present a numerical example based on real scenarios borrowed from the work by Eyckmans and Tulkens (2003) and Casas and Romera (2009). In Section 6, we present some conclusions and extensions of our work.

2 Stock Pollutant Control Model

We adopt the point of view of the issues presented in Germain, Toint, Tulkens, and de Zeeuw (2003).

In our model, we introduce a stochastic dynamic game formulation, with finite and discrete planing horizon analysis of IEA on transnational pollution control, as an extension of these issues.

Model Components

In our formulation we consider a Markovian Game described by a tuple

$$G = \{J, S, (E_i, r_i)_{i \in J}, p, \mathcal{T}\}$$

with the following elements

- There are n players and $J = \{1, 2, ..., n\}$ denotes the set of countries or regions which we simplify refer to as countries in the sequel.
- S is a Borel subset of some Polish (i.e., complete, separable, metric) countable and non empty space; is the state space of the game, with typical element s. The state transition dynamics is a function of the current state of the system and an additive noise factor on each period of time. The state of the system is the accumulated level of pollution in the atmosphere, given by s_t as stock of pollutant at each period t, $s_t \in S$, according to the state equation

  $$s_t = (1 - \delta)s_{t-1} + \sum_{i=1}^{n} e_{it} + \xi_t \quad , \qquad 1 \leq t \leq T \tag{1}$$

  Where

  - s_0 is the initial stock of pollutant or preindustrial level, given.
  - δ is the pollutant's natural rate of atmospheric absorption of CO_2 between two periods of time, such that $0 < \delta < 1$.
- p specifies the law of motion (or transition probabilities) for the game by associating with each $(s, a) \in S \times E$ a probability $p(\cdot|s, e)$ over the Borel sets of S.
- A finite planing horizon with discrete-time periods t, such that

  $$t \in \mathcal{T} = \{1, 2, ..., T\} \subset \mathbb{Z}^+.$$

- The control variables or action variables are e_{it}, where $e_t = (e_{1t}; e_{2t}; ...; e_{nt})'$ is the vector of the different countries emissions of pollutant at each period t, entailed by economic activity, where $e_{it} \in E$ and E is the countable and non empty overall control space or action space, and

$$E = \bigcup_{s \in S} E(s),$$

where $E(s)$ is the set of *admissible actions* (emissions), when the system is in each state (pollutant level) s. For each $s \in S$ the set $E(s)$ is finite.

- The random disturbance ξ_t is a noise process: a sequence of i.i.d. random variables and independent of the initial state s_0, with

$$\mathbb{E}\left[\xi_t\right] = 0, \qquad \sigma^2 = \mathbb{E}\left[\xi_t^2\right] < \infty, \qquad \forall t = 1, 2, ..., T-1. \tag{2}$$

We consider stock of pollutant in a wide sense, not restricted to the carbon dioxide (CO_2) stock level. Inclusion of manifold pollutants is important. To wit, the 1997 Kyoto Protocol to the Framework Convention on Climate Change limits aggregate emissions of six direct greenhouse gases, such as: carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), hydrofluorocarbons ($HFCs$), perfluorocarbons ($PFCs$), sulphur hexafluoride (SF_6)), as well for the indirect greenhouse gases such as SO_2, NOx, CO and/or micro particles of industrial pollution (between 0.1 y 2.5 μ-meters). The emissions are aggregated and considered as CO_2 equivalents.

Functional Cost Components

Following Jorgensen and Zaccour (2001) among many others, we assume that the emissions are proportional to production. Additionally we consider

- Future costs are discounted by the constant and positive *discount factor* β with $0 < \beta \leq 1$.

- $c_i(e_{it})$: function that measures in monetary terms the total cost incurred by country $i \in J$ at period $t \in \mathcal{T}$ from limiting its own industrial emissions to e_{it}; is a differentiable, decreasing ($c_i' < 0$) and strictly convex function ($c_i'' > 0$).

- $d_i(s_t)$: function that measures in monetary terms the damages caused by the stock of pollutant s_t during the time period t for the i-th country; is a differentiable, increasing ($d_i' > 0$) and convex function ($d_i'' \geq 0$).

- $r_i(e_{it}, s_t) = c_i(e_{it}) + d_i(s_t)$: function that measures in monetary terms the total cost incurred by country $i \in J$ from limiting its emissions to e_{it}, and the damages caused by the stock of pollutant s_t during the time period t for the i-th country; $r_{it} \in R$, where R is the finite cost set and R is a subset of $\mathbb{R}$, is a differentiable and convex function ($r_i'' \geq 0$).
.

We consider that the only way to control the stock of pollution is through the control of emissions, that is reducing pollution is done through the reduction of emissions, and not through the cleaning of the environment. The marginal cost c_i of reducing emissions is higher for lower levels of emissions.

The decreasing character of the cost functions c_i show the evident phenomenon of the increasing costs related to the emissions reduction, i.e. The increasing cost to decrease the emissions could be associated with filter installations or the use of other techniques.

Modes of countries behavior

The damages in each country's environment depend on the emissions of pollutant of all different countries at each time-period t that contribute to a stock s_t.

In cooperative form the countries jointly choose at each period its emissions levels in order to minimize the expected total discount costs, then the resulting trajectories of emissions and stock constitute the international optimum.

In non-cooperative form, each country considers only the damages of the stock of pollutant over itself. In the sense of a Nash equilibrium, the countries minimize, at each period, only its own expected discounted costs, with knowledge of the emissions vector e_{jt}, with $j \neq i$, of the other countries.

These assumptions are similar as those in Germain, Toint, Tulkens, and de Zeeuw (2003) and Eyckmans and Tulkens (2003), where a deterministic model in discrete time of the dynamic of stock pollutant is presented.

2.1 Cooperative Model

In this case, one assumes that the countries behave in an internationally optimal way, i.e. that each of them takes account of the impact of its own industrial pollution not only on itself but on all other countries as well. It is clear that the damages to the environment of country i will depend on the emissions of all countries. We solve the problem of minimize the expected discounted total cost for each period $t \in \mathcal{T}$, where $\mathcal{T}$ is a discrete and finite set, and for all the countries jointly $(P1)$

$$
\begin{array}{lll}
(P1) & \min\limits_{\{e_{it}\}} & \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n}\beta^{t}\left[r_i\left(e_{it}, s_t\right)\right]\right] \\
& \text{s.t.} & s_t = (1-\delta)^t s_{t-1} + \sum_{i=1}^{n} e_{it} + \xi_t \\
& & e_{it} \geq 0 \qquad \forall i = 1, \cdots, n; \quad \forall t = 1, \cdots, T \\
& & s_0 > 0
\end{array}
$$

The convexity of the function r_i suffices to guarantee that the solution exists and is unique. Thus, Problem $(P1)$ has an equilibrium $\{e^*_{it}\}$.

Remark: The resulting family of trajectories of emissions (policies) e^*_{it} for all players $i \in J$ determined together with the resulting stock s^*_t, constitute the international optimum for all periods $t \in \mathcal{T}$ or a cooperative equilibrium (see Dutta and Sundaram (1998)).

The expected value function W satisfies the Dynamic Programming equations for the problem $(P1)$

$$
\begin{array}{ll}
(P1.1) & W(T, s_{T-1}) = \min\limits_{e_{iT}} \mathbb{E}\left[\sum_{i=1}^{n} r_i(e_{iT}, s_T)\right] \\
(P1.2) & W(t, s_{t-1}) = \min\limits_{e_{it}} \mathbb{E}\left[\sum_{i=1}^{n} r_i(e_{it}, s_t) + \beta W(t+1, s_t)\right]
\end{array}
$$

$$\forall t = 1, 2, ..., T-1$$

$$
\begin{array}{lrcl}
\text{s.t.} & s_t & = & (1-\delta)s_{t-1} + \sum_{i=1}^{n} e_{it} + \xi_t \\
& e_{it} & \geq & 0 \qquad \forall t \in \mathcal{T} \quad , \quad \forall i \in J \\
& s_0 & > & 0
\end{array}
$$

If countries cooperate, they jointly solve $(P1.1)$ and country i's expected total discount cost at period T is:

$$W_i(T, s) = r_i(e^*_{iT}, s^*_T)$$

where e^*_{iT} is the optimal emission level (policy) and s^*_T is the optimal stock of pollutant at final period T, given

$$s^*_T = [1-\delta]s + \sum_{i=1}^{n} e^*_{iT}$$

where s is the inherited stock of pollutant at the begin of period T.

In earlier periods, if countries cooperate they jointly solve $(P1.2)$ for $1 \leq t \leq T-1$. Optimal levels of emissions and resulting stock of pollutant are denoted by e^*_{it} and s^*_t respectively.

Then let denotes the country i's expected discounted cooperative equilibrium cost by

$$W_i(t,s) = r_i(e^*_{it}, s^*_t) + \beta W_i(t+1, s^*_t) \qquad \forall t = 1, \cdots, T-1$$

where:

$$s^*_t = [1-\delta]s + \sum_{i=1}^{n} e^*_{it}$$

Let define as τ-expected discounted cooperative total cost by

$$W_i^\tau \equiv \sum_{t=1}^{\tau} W_i(t, s^*_{t-1}), \quad 1 \leq \tau \leq T-1,$$

and the expected discounted cooperative total cost

$$W_i \equiv \sum_{t=1}^{T} W_i(t, s^*_{t-1}). \tag{3}$$

2.2 Non-Cooperative Model

In an alternative mode of behaviour, we describe what would happen if the countries do not sign a voluntary international environmental agreement. One may assume that countries behave non cooperatively in the sense of Nash equilibrium, where each of them minimizes at each period only its own discounted costs, taking given the emissions of the other countries. A Nash equilibrium is a family of strategies, one for each player, that minimize every country i's cost, given the strategies of all other players $j \neq i$. In such an equilibrium, no individual country has an incentive to deviate as long as the other countries stick to their equilibrium strategies.

The considered problem is a dynamic game in discrete time and finite horizon with only one player or country. We can adopt the perspective of an Optimal Control Problem (OCP), where the dynamic model is a system in discrete time $s_{t+1} = \phi(s_t, e_t, \xi_t)$ for all $t \in \mathcal{T}$ with initial condition s_0 and finite horizon $T < \infty$.

Formally, there are n problems to solve. Actually, at each period of time $t \in \mathcal{T}$, each country $i \in J$ solves the following problem $(P2)$

$$
\begin{aligned}
(P2) \qquad \min_{\{e_{it}\}_{t\in\{1,\ldots,T\}}} \quad & \mathbb{E}\left[\sum_{t=1}^{T} \beta^t \left[r_i(e_{it}, s_t)\right]\right] \\
\text{s.t.} \quad & s_t = (1-\delta)s_{t-1} + \sum_{i=1}^{n} e_{it} + \xi_t \\
& e_{it} \geq 0 \quad \forall t \in \mathcal{T}; \quad \forall i \in J \\
& s_0 > 0
\end{aligned}
$$

The *expected value functions* N_i, according to Bellman's principle of optimality, can be found by solving the Stochastic Dynamic Programming equations for $(P2)$

$$
\begin{aligned}
(P2.1) \qquad & N_i(T, s_{T-1}) = \min_{e_{iT}} \mathbb{E}\left[r_i(e_{iT}, s_T)\right] \\
(P2.2) \qquad & N_i(t, s_{t-1}) = \min_{e_{it}} \mathbb{E}\left[r_i(e_{it}, s_t) + \beta N_i(t+1, s_t)\right]
\end{aligned}
$$

$$\forall t = 1, 2, ..., T-1$$

$$
\begin{aligned}
\text{s.t.} \qquad s_t &= (1-\delta)s_{t-1} + \sum_{i=1}^{n} e_{it} + \xi_t \\
e_{it} &\geq 0 \qquad \forall t \in \mathcal{T} \quad , \quad \forall i \in J \\
s_0 &> 0
\end{aligned}
$$

Remark: The resulting family of trajectories of emissions e_{iT}^N thus determined for each $i \in J$, together with the resulting stock s_T^N, constitute a non-cooperative Nash equilibrium for all periods $t \in \mathcal{T}$ (see Dutta and Sundaram (1998))..

The convexity of the function r_i suffices to guarantee that the Nash equilibrium exists and is unique.

In the non cooperative equilibrium the country i's expected total discounted cost at period T is

$$N_i(T,s) = r_i(e_{iT}^N, s_T^N) \qquad ; \qquad s_T^N = [1-\delta]s + \sum_{i=1}^{n} e_{iT}^N$$

where $e_{iT}^N = \{e_{1T}^N, e_{2T}^N, \ldots, e_{nT}^N\}$ is the vector that denotes the emissions equilibrium level and s_T^N denotes the resulting stock of pollutant at final period of time T, where s is the inherited stock of pollutant at the begin of period of time T.

Let define as τ-expected discounted non cooperative total cost by

$$N_i^\tau \equiv \sum_{t=1}^{\tau} N_i(t, s_{t-1}^N), \qquad 1 \leq \tau \leq T-1,$$

and the expected discounted non cooperative total cost by

$$N_i \equiv \sum_{t=1}^{T} N_i(t, s_{t-1}^N). \tag{4}$$

2.3 The underlying MDP Model

We considere by MDP a Markov Decision Process (also called Markov Control Model)together with an optimality criteria. The problems considered in this work are discrete-time, finite-horizon and stationary MDP with expected total reward. Then, we can express the elements of our random scenarios through the following MDP described by a tuple

$$\Gamma = (S, E, R, P, \beta), \tag{5}$$

where the *state space* S and the overall *action space* (or control space) $E = \bigcup_{s\in S} E(s)$ are both countable and nonempty, $E(s)$ is the set of *admissible actions* (emissions), when the system is in each state (pollutant level) s. For each $s \in S$ the set $E(s)$ is finite. The *cost set* R is a bounded countable subset of $\mathbb{R}$. For each $t \geq 1$, let s_t, e_t and r_t denote the state (pollutant level) of the system, the action (emissions) taken by the decision maker (pays), and the cost incurred at period of time t, respectively.

The stationary, single-stage, conditional *transition probabilities* are defined by

$$p_{i,j,r}^e := Prob\left(s_{t+1} = j, r_t = r / s_t = i, e_t = e\right), \tag{6}$$

$$\forall i, j \in S \quad , \quad e \in E(i) \quad , \quad r \in R \quad , \quad t \geq 1,$$

$$\sum_{j \in S, r \in R} p^e_{i,j,r} = 1 \quad , i \in S \quad , \quad e \in E(i).$$

3 Probability Criteria Model

To improve the non-cooperative equilibrium we propose a different criteria to the minimization of the expected total discounted cost with possibly monetary transfers as incentive.

For each country $i \in J$ we consider the problem to find a policy (emission level) which maximize the probability that the expected discounted total cost does not exceed a specified value X, named target. That is, for the finite horizon model $1 \leq t \leq T$, find a policy $\pi^{i*} = \{e^*_{it}\}$, in the sequel $\pi^{i*} = \pi^*$ (we omit the index i), such that

$$(P4) \qquad \max_{\{e_{it}\}_{t \in \mathcal{T}}} \quad \left\{ Prob \left[\left(\sum_{t=1}^{T} \beta^t r_i(e_{it}, s_{t-1}) \right) \leq X \right] \right\}$$

$$\text{s.t.} \qquad s_t = (1-\delta)s_{t-1} + \sum_{i=1}^{n} e_{it} + \xi_t$$

$$e_{it} \geq 0$$

$$s_0 > 0$$

where the total cost is

$$r_i : E \times S \longrightarrow R \subset \mathbb{R}$$

is the total cost defined in section 2 by the addition of the cost and the damage functions.

By considering a new decision-maker state, we formulate a new framework for finding a policy which maximizes the probability that the expected discounted total cost does not exceeds a specified value X, which we name the *target value*.

Definition 1. *The* ***target value*** *(X) is a quantity ranging between the expected discounted non cooperative total cost (4) and the expected discounted cooperative total cost (3).*

Note that the objective function in the model $(P4)$ is equivalent to

$$(P4) \Leftrightarrow \min_{\{e_{it}\}_{t\in\mathcal{T}}} \left\{ -Prob\left[\left(\sum_{t=1}^{T} \beta^t r_i(e_{it}, s_{t-1})\right) \le X\right]\right\} \tag{7}$$

and

$$(P4) \Leftrightarrow \max_{\{e_{it}\}_{t\in\mathcal{T}}} \{\mathbb{E}\left[\mathbf{1}_{R_i \le X}\right]\} \tag{8}$$

where

$$R_i = \left(\sum_{t=1}^{T} \beta^t r_i(e_{it}, s_{t-1})\right)$$

In our formulation, when making a decision and taking an action at each stage s, the decision maker considers not only the state of the original system but also his updated *target* x. Thus, we consider in fact an expanded model of MDP Γ by enlarging the state space. A similar formulation is considered by Wu and Lin (1999) that studies the minimizing risk problems in MDP with countable state space and reward set, and Boda, Filar, Li, and Spanjers (2004) that consider a problem of optimal control of a retirement investment fund over a finite time horizon with a target hitting time criteria.

A new "*hybrid state* $(s, x) \in S \times \mathbb{R}$" is introduced. We refer (s, x) as the *hybrid state* of the decision maker to distinguish it from the system's state s, where x is the target value. The dynamic of the system is now as follows: if the initial state of the decision maker is (s, x) and an action e is taken according to 6, the decision maker's new *hybrid state* transits from (s, x) to $(j, \frac{x-r}{\beta})$ with probability p^e_{sjr}.

Thus, the extended MDP $\tilde{\Gamma}$ has the following structure

$$\tilde{\Gamma} = (\tilde{S}, E, R, P, \beta), \tag{9}$$

where

$$\tilde{S} = S \times \mathbb{R} \quad , \quad E = \bigcup_{(s,x)\in\tilde{S}} E(s, x) = \bigcup_{s\in S} E(s).$$

Note that $E(s, x) = E(s)$, with $(s, x) \in \tilde{S}$, and the extended stationary conditional transition probabilities are simply

$$p^e_{s,j,r} \;:=\; Prob\left(\tilde{s}_{t+1} = \left(j \quad , \quad \frac{x-r}{\beta}\right) / \quad \tilde{s}_t = (s, x) \quad , \quad e_t = e\right),$$
$$\forall s, j \in S \quad , \quad e \in E(s) \quad , \quad r \in R \quad , \quad x \in \mathbb{R}.$$

Note that R and β are the same as in the MDP Γ given by 5.

Remark that the target x is important when making decisions and consequently we must define policies which depend both on the state and the target, that is on the hybrid state $\tilde{s} = (s, x)$.

3.1 Main results

In this subsection we introduce the definition of policy. The policies discussed depend in fact on target values. The main results obtained are: (i) the optimal value function are distribution functions of the target, and (ii) there exists an optimal deterministic Markov policy.

Definition 2. *A* ***decision rule*** π_t *at stage* t *is a conditional transition probability measure on the set of admissible actions* $E(s_t)$ *given the past history* $(\tilde{s}_1, e_1, \cdots, \tilde{s}_{t-1}, e_{t-1}, \tilde{s}_t)$*.*

Definition 3. *A* ***policy*** π *is a sequence of decision rules* $\pi = \{\pi_t\}_{t\geq 1}$*. The set of all policies is denoted by* Π*.*

A policy $\pi = \{\pi_t\}_{t\geq 1} \in \Pi$ is said to be the following.

- **Markov policy**, if each decision rule π_t only depends on the current state at stage t. Moreover π is a Markov policy if each π_t verifies that $\pi_t(\cdot/\tilde{s}_1, e_1, \cdots, \tilde{s}_{t-1}, e_{t-1}, \tilde{s}_t) = \pi_t(\cdot/\tilde{s}_t)$. The set of all Markov policies is denoted by Π_m.

- **Stationary policy**, if the policy π is a Markov policy, and the decision rules of π are all identical, that is, $\pi_t = \pi_1, \quad \forall t > 1$ which is denoted by $\pi = \pi_1^\infty$. The set of all stationary policies is denoted by Π_s.

- **Deterministic policy**, is any policy π such that all of its decision rules π_t are deterministic. The set of all deterministic Markov and deterministic stationary policies are denoted by Π_m^d and Π_s^d, respectively.

- **TI-policy**, a policy π which are independent of targets $x_t, \quad (t \geq 1)$, let Π_0 denote the set of all TI-policies.

Note that a transition law P and a policy π determine the conditional probability measure P_π. Let R_τ^π denote the random variable that is the sum of discounted costs generated by policy π for the τ-stage finite horizon problem. That is

$$R_\tau^\pi = \sum_{t=1}^{\tau} \beta^{t-1} r_t, \qquad \forall \tau \geq 1.$$

Let consider the following *objective function* generated by policy $\pi \in \Pi$

$$F_\tau^\pi(s,x) = P_\pi\left(R_\tau^\pi > x/\tilde{s}_1 = (s,x)\right) \quad , \quad \forall \tau \geq 1. \tag{10}$$

Definition 4. *The following functions are called the* ***optimal value functions***

$$F_\tau^*(s,x) = \inf_{\pi \in \Pi}\left\{F_\tau^\pi(s,x)\right\} \quad , \quad \forall (s,x) \in \tilde{S}, \quad \tau \geq 1. \tag{11}$$

Obviously

$$F_\tau^*(s,x) = \begin{cases} 1 & if \quad x \geq \frac{d(1-\beta^\tau)}{1-\beta} \\ 0 & if \quad x < \frac{b(1-\beta^\tau)}{1-\beta} \end{cases} \tag{12}$$

where we define the lower b and the upper d bounds on the costs by

$$b = \inf\{r : r \in R\} \quad , \quad d = \sup\{r : r \in R\}.$$

Definition 5. *If the policy $\pi^* \in \Pi$ is such that $F_\tau^{\pi^*}(s,x) = F_\tau^*(s,x)$ for all $(i,x) \in \tilde{S}$, $\tau \geq 1$, then π^* is called a* ***τ-stages optimal policy***.

Remark: For any policy π independent of targets $\{x_\tau\}$, at each period of time $\tau \geq 1$, $F_\tau^\pi(s,x)$ is a distribution function of x, but this result does not hold for general policy $\pi \in \Pi$.

In the next step, we introduce some notation necessary to check that dynamic programming operators possess the usual monotonicity properties, e.g. see Puterman (2005).

Let space $\mathcal{D} = \{u/u : \tilde{S} \to [0;1], measurable\}$ be the space of measurable functions on the extended decision maker's state space $\tilde{S}$. For any policy π stationary and $u \in \mathcal{D}$, we define the dynamic programming operators: the average cost G with emissions policies e, the average cost over all emissions policies K^π and the minimum of average cost K, by

$$Gu(s,x,e) = \sum_{j \in S, r \in R} p_{sjr}^e u\left(j, \frac{x-r}{\beta}\right) \quad ; \quad (s,x) \in \tilde{S}, e \in E(s). \tag{13}$$

$$K^\pi u(s,x) = \sum_{e \in E(s)} \pi(e/s,x) Gu(s,x,e) \quad ; \quad (s,x) \in \tilde{S}. \tag{14}$$

$$Ku(s,x) = \min_{e \in E(s)}\left\{Gu(s,x,e)\right\} \quad ; \quad (s,x) \in \tilde{S}. \tag{15}$$

Note that

$$\begin{aligned}
(K^\pi)^0 u &= u, \\
(K^\pi)^\tau u &= K^\pi \left((K^\pi)^{\tau-1} u\right), \\
K^0 u &= u, \\
K^\tau u &= K(K^{\tau-1} u).
\end{aligned}$$

Obviously when π is deterministic stationary policy, that is a non-randomized policy such that

$$\pi_t(\cdot/\tilde{s}_t) = \pi(\tilde{s}_t) \in E(s_t),$$

the decision rule π_t at each $t > 1$, is non-random and the policy is a sequence $\pi = (\pi(\tilde{s}_1), \pi(\tilde{s}_2), \cdots)$, we have

$$K^\pi u(s,x) = Gu\left(s, x, \pi(s,x)\right).$$

In addition, if $r_0 = 0$ and $F_0^\pi(s,x) = P_\pi\left(r_0 \le x, \tilde{s}_1 = (s,x)\right)$ for any policy $\pi = (\pi_\tau, \tau \ge 1)$, then we have

$$F_0^*(s,x) = I_{[0,\infty)}(x) \quad , \qquad \forall (s,x) \in \tilde{S}, \pi \in \Pi, \tag{16}$$

where $I_{[0,\infty)}$ is the indicator function of set $[0,\infty)$. We have the following Lemma, which checked that the operators G, K^π and K defined above possess the usual monotonicity properties of dynamic programming.

Lemma 1.

(i) If $u, v \in \mathcal{D}$, $u \le v$ then $Gu \le Gv$, $K^\pi u \le K^\pi v$, $Ku \le Kv$.

(ii) Let $u \in \mathcal{D}$. If $u(s,x)$ is a non-decreasing and a left continuous function of x for any $s \in S$, then $Ku(s,x)$ is also non-decreasing and a left continuous function of x for each $s \in S$.

(iii) There exists a deterministic stationary policy f such that $K^f u = Ku$.

Proof.

(i) For each $(s,x) \in \tilde{S}$, $\quad e \in E(s)$,

$$\begin{aligned}
\text{since} \quad u\left(j, \frac{x-r}{\beta}\right) &\le v\left(j, \frac{x-r}{\beta}\right) \quad , \quad \forall (j,x) \in \tilde{S}, \\
\text{then} \quad \sum_{j\in S, r\in R} p^e_{sjr} u\left(j, \frac{x-r}{\beta}\right) &\le \sum_{j\in S, r\in R} p^e_{sjr} v\left(j, \frac{x-r}{\beta}\right), \\
\text{and} \quad Gu(s,x,e) &\le Gv(s,x,e).
\end{aligned}$$

Additionally

$$\begin{aligned} K^{\pi}u(s,x) &= \sum_{e\in E(s)} \pi(e/s,x)Gu(s,x,e) \\ &\leq \sum_{e\in E(s)} \pi(e/s,x)Gv(s,x,e) \\ &= K^{\pi}v(s,x). \end{aligned}$$

Finally, for each $(s,x) \in \tilde{S}$

$$\min_{e\in E(s)} \sum_{j\in S, r\in R} p^{e}_{sjr} u\left(j, \frac{x-r}{\beta}\right) \leq \min_{e\in E(s)} \sum_{j\in S, r\in R} p^{e}_{sjr} v\left(j, \frac{x-r}{\beta}\right).$$

$$\text{Thus} \quad Ku(s,x) \leq Kv(s,x).$$

provided that the minimum is taken over a finite set $E(s)$.

(ii) For any $s \in S$, $u(s,x)$ is a non-decreasing function of x

$$\begin{aligned} \lim_{h\to 0} u(s,x-h) &= u(s,x). \\ \text{If} \quad x_1 \leq x_2 \quad \Rightarrow & \quad u(s,x_1) \leq u(s,x_2). \\ \text{From (i)} \quad Gu(s,x_1) &\leq Gu(s,x_2), \\ \min Gu(s,x_1) &\leq \min Gu(s,x_2), \\ Ku(s,x_1) &\leq Ku(s,x_2), \end{aligned}$$

and Ku is a non-decreasing function of x.

$$\lim_{h\to 0} Ku(s,x-h) = \lim_{h\to 0} \min_{e\in E(s)} Gu(s,x-h,e) = \min_{e\in E(s)} Gu(s,x,e),$$

and Ku is a left continuous function of x.

(iii) If $\pi = (\pi_1, \pi_2, \cdots)$ is a deterministic admissible policy, then

$$\begin{aligned} E(s) \equiv e \quad \Rightarrow & \quad \pi(e/s,x) \equiv 1, \\ K^{\pi}u(s,x) = Gu(s,x,e) = & \min_{e\in E(s)} Gu(s,x,e) = Ku(s,x), \\ \text{and} & \quad K^{\pi}u(s,x) = Ku(s,x). \end{aligned}$$

The existence of such deterministic admissible policy is guaranteed by Hernández-Lerma (1989) Proposition D3 p. 130.

□

Below, we establish the "optimality principle" for the target value criterion problem.

Proposition 1.

(i) *The optimal value function* $\{F_\tau^*, \tau \geq 0\}$ *satisfies the optimality equations*

$$F_0^* = I_{[0,\infty)}, \quad F_\tau^* = KF_{\tau-1}^*, \quad \tau \geq 1.$$

(ii) *For all* $\tau \geq 0$ *and* $\tilde{s} \in \tilde{S}$, $F_\tau^*(\tilde{s}, R_i^\tau)$ *is a distribution function of* R_i^τ.

(iii) *For any* $\tau \geq 0$, *there exists a deterministic stationary policy* π *such that* $F_\tau^\pi = F_\tau^*$.

Proof.

Note that if $\tau = 0$ then $F_0^*(s,x) = I_{[0,\infty)}(x)$ holds. Assume that (i) holds if $\tau = \tau_0$. By induction assumption, for any $\tilde{s} \in \tilde{S}$, $F_\tau^*(\tilde{S}, R_i^\tau)$ is a distribution function of R_i^τ. Note that $E(s,x) = E(s)$ is finite for any $(s,x) \in S$. By the measurable selection theorem (see Hernández-Lerma (1989), Proposition D3, p. 130), there exists a measurable mapping δ from S to E such that $\delta(s,x) \in E(s)$ and $GF_\tau^*(s,x,\delta(s,x)) = KF_\tau^*(s,x)$ for all $(s,x) \in S$, that is, $\delta^\infty \in \Pi_s^d$ and $K^\delta F_\tau^* = KF_\tau^*$.

By induction assumption, there exists a policy $\sigma \in \Pi_m^d$ such that

$$F_\tau^\sigma = F_\tau^*.$$

Let $\pi = (\delta, \sigma)$. Then $\pi \in \Pi_m^d$. By Lemma 1 iii) we have

$$F_{\tau+1}^*(s,x) \leq F_{\tau+1}^\pi(s,x) = K^\delta F_\tau^\sigma(s,x) = K^\delta F_\tau^*(s,x) = KF_\tau^*(s,x). \quad (17)$$

On the other hand, for any $\eta \in \Pi$, by let $\pi = (\pi_\tau, \tau \geq 1) \in \Pi$, then

$$F_\tau^\pi(s,x) = \sum_{a \in E(s)} \pi_1(a \backslash s, x) \sum_{j \in S, r \in R} p_{sjr}^a F_{\tau-1}^{\pi(s,x,a)}\left(j, \frac{x-r}{\beta}\right), (s,x) \in S, \tau \geq 1,$$

and F_τ^π is determined by $\pi(\tau)$, we have

$$F_{\tau+1}^*(s,x) \leq F_{\tau+1}^\pi(s,x) = K^\delta F_\tau^\sigma(s,x) = K^\delta F_\tau^*(s,x) = KF_\tau^*(s,x).$$

On the other hand, for any $\eta \in \Pi$, we have

$$F^{\eta}_{\tau+1}(s,x) = K^{\eta} F^{\eta}_{\tau}(s,x) \geq K^{\eta} F^{*}_{\tau}(s,x) \geq K F^{*}_{\tau}(s,x).$$

Hence $F^{*}_{\tau+1}(s,x) \leq K F^{*}_{\tau}(s,x)$. Associating it with 16, we obtain that $K F^{*}_{\tau} = F^{*}_{\tau+1} = F^{\pi}_{\tau+1}$. Thus, by Lemma 1 and 11, $F^{*}_{\tau+1}(s,x)$ is a distribution function of x.

Early results imply that Proposition 1 is also true when $n = \tau + 1$. By induction, for any $n \geq 0$ holds.

This completes the proof of Proposition 1. □

Note that by solving the probabilistic problem $(P4)$, for each country $i \in J = \{1, 2, \ldots, n\}$ we obtain a new strategy, say π^{*}_{i}, which is related to behavior that in somehow fits the gap between cooperative and non-cooperative solutions, e^{*}_{i} and e^{N}_{i} respectively, for each player.

The target value (X) is a crucial element in our setting, and we find that it could be a relevant issue in economic negotiations concerning abatement stock pollutant policies.

4 Algorithm

We present an algorithm which computes optimal value functions, optimal action sets, and optimal policies for a finite horizon model with the probability criteria, using the backward recursion algorithm of dynamic programming adapted to apply to our problem.

4.1 DP - Algorithm

We assume that S and R are both finite sets and we let $R = \{r_1, r_2, \ldots, r_m\}$ with $r_1 < r_2 < \ldots < r_m$. Then, by the Proposition 1 and the finiteness of S, E and R, we have the following conclusions:

(i) For each $s \in S$ and $\tau \geq 1$, $F^{*}_{\tau}(s,x)$ is a step distribution function of x with finite jump points;

(ii) For each $s \in S$ and $\tau \geq 1$, $e^{*}_{\tau}(s,x)$ is a set-valued function from $\mathbb{R}$ to $E(s)$ with finite discontinuity points;

(iii) For each $\tau \geq 1$, there exists an τ stages optimal deterministic Markov policy which k-th decision rule has the structure analogous to that of $F_\tau^*(s,x)$ and $e_\tau^*(s,x)$, $1 < k < \tau$.

The following algorithm is just the proof of the earlier conclusions.

By Proposition 1, we have

$$F_0^*(s,x) = I_{[0,\infty)}(x) \quad , \quad \forall (s,x) \in \tilde{S}, \quad \pi \in \Pi, \quad x \geq 0,$$

$$F_\tau^*(s,x) = \min_{e \in E(s)} \left\{ \sum_{j \in S, r \in R} p_{sjr}^e F_{\tau-1}^* \left(j, \frac{x-r}{\beta} \right) \times I_{[0,\infty)}(x-r) \right\}, \qquad (18)$$

$$\forall s \in S, \quad x \in \mathbb{R}, \quad \tau \geq 1.$$

Then for notational convenience, define

$$\begin{aligned} b_\tau(s,x,e) &\equiv \sum_{j \in S, r \in R} p_{sjr}^e F_{\tau-1}^* \left(j, \frac{x-r}{\beta} \right) \times I_{[0,\infty)}(x-r) \quad , \forall s \in S, e \in E(s), \\ M_\tau(s,x) &= \min_{e \in E(s)} \{ b_\tau(s,x,e) \} \quad , \quad \forall s \in S. \end{aligned}$$

With the Proposition 1, Lemma 1 and Definition 3 of optimal value functions, we obtain the following algorithm.

Step 1. Calculate

$$\begin{aligned} b_1(s,r_k,e) &= \sum_{j \in S} \sum_{r \in R, r \leq r_k} p_{sjr}^e, \quad \forall s \in S, e \in E(s), \\ M_1(s,r_k) &= \min_{e \in E(s)} \{ b_1(s,r_k,e) \} \quad , \forall s \in S, \\ E_1^*(s,r_k) &= \{ e : e \in E(s), b_1(s,r_k,e) = M_1(s,r_k) \} \quad , \quad \forall s, \in S, \end{aligned}$$

and select an action $g_1(s,r_k) \in E_1^*(s,r_k)$, $k = 1,2,\ldots,m-1$, and an arbitrary action $g_1(s,r_m) \in E(s)$. Then by 18 and definition of the optimal action sets

$$F_1^*(s,x) = \begin{cases} 0 & if \quad x < r_1 \\ M_1(s,r_k) & if \quad r_k \geq x \geq r_{k+1}, \quad k = 1,\ldots,m-1 \\ 1 & if \quad x \geq r_m \end{cases}$$

$$E_1^*(s,x) = \begin{cases} E(s) & if \quad x < r_1 \quad or \quad x \geq r_m \\ E_1(s,r_k) & if \quad r_k \geq x < r_{k+1}, \quad k = 1,\ldots,m-1. \end{cases}$$

Let

$$g_1(s,x) = \begin{cases} g_1(s,r_m) & if \quad x < r_1 \quad or \quad x \geq r_m \\ g_1(s,r_k) & if \quad r_k \geq x < r_{k+1}, \quad k = 1,\dots,m-1. \end{cases}$$

Step 2. Assume that F_l^*, E_l^* and g_l have been calculated and all jump points of $F_l^*(i,x)$ ($\forall i \in S$) with $x_1 < x_2 < \cdots < x_\rho$ are known. Calculate the elements of set

$$\{\beta x_k + r_h | k = 1,2,\dots,\rho, \quad h = 1,2,\dots,m\}$$

and denote them by $u_1 < u_2 < \cdots < u_L \quad (L \leq m\rho)$, in an ascending order. Then, for any $j \in S$ and $r \in R$, we have

$$F_l^*(j, \frac{x-r}{\beta}) = \begin{cases} 0 & if \quad x < u_1 \\ F_l^*(j, \frac{u_k - r}{\beta}) & if \quad u_k \geq x \geq u_{k+1}, \quad 1 \leq k \leq L \\ 1 & if \quad x \geq u_L. \end{cases} \tag{19}$$

If $r_1 > u_L$, then $I_{[0,\infty)}(u_k - r) = 0, \quad k = 1,\dots,L$ and, hence, from 17 $F_{l+1}^*(i,u_k) = 0$ for all k. Or, there exists some t such that $u_{t-1} \leq r_1 \leq u_t$, note that if $r_1 < u_1$ we can simply define $u_0 = r_1$ and take $t = 1$.

Calculate

$$\begin{aligned}
b_{l+1}(s,r_1,e) &= \sum_{j\in S} p_{sjr_1}^e F_l^*(j,0), \quad s \in S, e \in E(s), \\
b_{l+1}(s,u_k,e) &= \sum_{j\in S, r\in R, r\leq u_k} p_{sjr}^e F_l^*(j, \frac{u_k - r}{\beta}), \quad s \in S, e \in E(s), k \geq N, \\
M_{l+1}(s,r_1) &= \min_{e\in E(s)} \{b_{l+1}(s,r_1,e)\}, \quad \forall s \in S, \\
M_{l+1}(s,u_k) &= \min_{e\in E(s)} \{b_{l+1}(s,u_k,E)\}, \quad \forall s \in S, k \geq N, \\
E_{l+1}^*(s,r_1) &= \{e : e \in E(s), b_{l+1}(s,r_1,e) = M_{l+1}(s,r_1)\}, \quad \forall s \in S, \\
E_{l+1}^*(s,u_k) &= \{e : e \in E(s), b_{l+1}(s,u_k,e) = M_{l+1}(s,u_k)\}, \quad \forall s \in S, k \geq N.
\end{aligned}$$

Next, select actions $g_{l+1}(s,r_1) \in E_{l+1}^*(s,r_1)$, $g_{l+1}(s,u_k) \in E_{l+1}^*(s,u_k)$, $k = t,\dots,L-1$, and an arbitrary action $g_{l+1}(s,u_L) \in E(s)$. Then by 17, 18 and definition of optimal action sets

$$\begin{aligned}
F_1^*(s,x) &= \begin{cases} 0 & if \quad x < r_1 \\ M_1(s,r_k) & if \quad r_k \geq x \geq r_{k+1}, \quad k = 1,\dots,m-1 \\ 1 & if \quad x \geq r_m \end{cases} \\
E_1^*(s,x) &= \begin{cases} E(s) & if \quad x < r_1 \quad or \quad x \geq r_m \\ E_1(s,r_k) & if \quad r_k \geq x < r_{k+1}, \quad k = 1,\dots,m-1 \end{cases}
\end{aligned}$$

Let the decision rule at the next stage be defined by

$$g_{l+1}(s,x) = \begin{cases} g_{l+1}(s,r_1) & if \quad\quad\quad\quad r_1 < x \le u_N \\ g_{l+1}(s,u_k) & if \quad u_k < x \le u_{k+1}, \quad k = 1,\ldots,m-1 \\ g_{l+1}(s,u_L) & if \quad\quad\quad x \le r_1 \quad or \quad x > u_L, \end{cases}$$

Step 3. Repeat Step 2 until $l+1=\tau$, or $\tau = T$.

We construct the optimal value function F^*_τ and an optimal policy $\pi^* = (g_\tau, g_{\tau-1}, \ldots, g_1)^\infty$. In the process, the corresponding optimal action sets $E^*_1(s,x), E^*_2(s,x), \ldots, E^*_\tau(s,x)$ are constructed as well. By Proposition 1, these sets characterize all τ stages optimal policies.

4.2 Modified DP - Algorithm

The DP Algorithm can calculate the optimal value functions, optimal policies and action sets accurately, however, it can quickly become computationally prohibitive. At each iteration more and more points (u_k) need to be considered. For a large state space, a large action space and a large reward set this will have drastic consequences. The number of points that need to be considered and thereby the time to do this will grow exponentially.

To overcome this problem a new algorithm is presented below. This algorithm approximates the solution found by the DP Algorithm by calculating a fixed number of points at each iteration. However, by taking this number large enough, the approximation will be quite good and the computational time will decrease significantly. We will that all rewards in the problem are positive.

The idea is that (irrespective of the iteration index l) a bounded monotone decreasing function such as $F^*_l(i,x)$ on an interval $[0, v_m]$ can be well approximated by an array of values $\{(v_1, F^*_l(i,v_1)), (v_2, F^*_l(i,v_2)), \cdots, (v_m, F^*_l(i,v_m))\}$ provided that $|v_{i+1} - v_i|$ is sufficiently small. The interpolation between the values $F^*_l(i,v_i)$ and $F^*_l(i,v_{i+1})$ at v_i and v_{i+1} can be carried out in a number of ways. In the implementation below upper end is used. That is,

$$F^*_l(i,v) = F^*_l(i,v_{i+1}) \qquad \forall v \in (v_i, v_{i+1}]$$

The following enhanced dynamic programming algorithm can now be used. For notational convenience, we assume $\beta = 1$ and define

$$\begin{aligned} b_\tau(i,x,e) &\equiv \sum_{j\in S, r\in R} p^e_{ijr} F^*_{\tau-1}(j, x-r) \times I_{[0,\infty)}(x-r) \qquad \forall i \in S, e \in E(i) \\ M_\tau(i,x) &= \min_{e\in E(i)} \{b_\tau(i,x,e)\} \qquad \forall i \in S \end{aligned}$$

Step 1. Initialize:

Choose m points $v_1 < v_2 < \cdots < v_k < \cdots < v_m$ that will represent the target values. The value of v_1 needs to be N_i^T, the expected discounted total cost of non cooperative problem. The value of v_m is the largest target value that will be computed. The larger the m, the more accurate the approximation of the optimal value functions will be. Taking equi-spaced v_k's will have computational advantages. Now by Proposition 1

$$F_0^*(i,x) = \begin{cases} 0 & if \quad x \leq v_1 \\ 1 & if \quad v_{k-1} < x \leq v_k, \quad k = 2,\ldots,m. \end{cases}$$

Step 2. Assume that F_l^* has been calculated. Now calculate

$$\begin{aligned} b_{l+1}(i,v_k,e) &= \sum_{j\in S, r\in R} p_{ijr}^e F_l^*(j, v_k - r), \qquad i \in S, e \in E(i), k \geq 1, \\ M_{l+1}(i,v_k) &= \min_{e\in E(i)} \{b_{l+1}(i,v_k,e)\}, \qquad \forall i \in S, k \geq 1, \\ E_{l+1}^*(i,v_k) &= \{e : e \in E(i), b_{l+1}(i,v_k,e) = M_{l+1}(i,v_k)\}, \qquad \forall i \in S, k \geq 1. \end{aligned}$$

Next, select actions $g_{l+1}(i,v_k) \in E_{l+1}^*(i,v_k)$, with $k = 1,\ldots,m$. Then

$$F_{l+1}^*(i,x) = \begin{cases} 0 & if \quad x \leq v_1 \\ M_{l+1}(i,v_k) & if \quad r_{k-1} < x \leq v_k, \quad k = 2,\ldots,m \end{cases}$$

$$E_{l+1}^*(i,x) = \begin{cases} E(i) & if \quad x \leq v_1 \\ E_{l+1}^*(i,v_k) & if \quad v_{k-1} < x \leq v_k, \quad k = 2,\ldots,m. \end{cases}$$

Let the decision rule at the next stage be defined by

$$g_{l+1}(i,x) = \begin{cases} g_{l+1}(i,v_1) & if \quad x < v_1 \\ g_{l+1}(i,v_k) & if \quad v_{k-1} < x \leq v_k, \quad k = 2,\ldots,m. \end{cases}$$

Step 3. Repeat Step 2 until $l + 1 = \tau$ or $l + 1 = T$.

The approximate optimal value function F_τ^* and an optimal policy $\pi^* = (g_\tau, g_{\tau-1},\ldots,g_1)^\infty$ have now been constructed. In the process, the corresponding approximate optimal action sets $E_1^*(i,x), E_2^*(i,x),\ldots,E_\tau^*(i,x)$ have been constructed as well. By Proposition 1, these sets characterize all τ stages optimal policies.

5 Numerical Results

In this section, we show some numerical results obtained of solver the probabilistic problem ($P4$), the first applied to an illustrative example with linear reward functions r_i, and the second example applied to a real scenario. The simulations are made for a time horizon of 100 years, we give the results only up to 2030, in order to avoid boundary problems. All computations were made by use of the software Matlab 7.3.0 (R2006b).

We have implemented the equivalent formulation of problem ($P4$) given in Section 3, following the objective function equivalents (7) and (8). Thus, we have developed specific code for our example. All the tables are included in the Appendix.

5.1 The Linear Case

One assumes, by definition of reward function, that

$$r_i(e_{it}, s_t) = c_i(e_{it}) + d_i(s_t)$$

where $c_i(e_{it}) = ae_{it}$, and $d_i(s_t) = bs_t + c$, $\forall a, b, c \in \mathbb{R}$.

Following (7) and (8) the objective linear function of the probabilistic model ($P4$) has the following form

$$\max_{\{e_{it}\}_{t\in\mathcal{T}}} \quad \left\{ Prob\left[\left(\sum_{t=1}^{T} \beta^t \left(c_i(e_{it}) + d_i(s_t) \right) \right) \leq X \right] \right\}$$

Then the objective function of model ($P4$) is equal to the objective function of the following linear programming

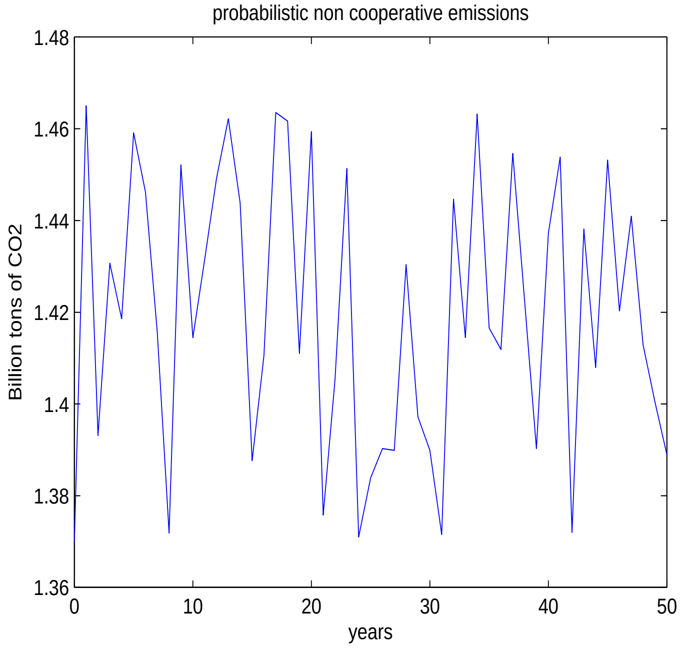
$$\max_{\{e_{it}\}_{t\in\mathcal{T}}} \quad \left\{ Prob\left[\left(\sum_{t=1}^{T} \beta^t \left(ae_{it} + bs_t + c \right) \right) \leq X \right] \right\}$$

We use as target value $X = 220000$, and the initial conditions following

$$e_{1990} = [1.37, \quad 0.29, \quad 0.872, \quad 0.805, \quad 1.066, \quad 3.43]$$

the initial CO_2 vector of emissions e_{1990}, in absence of any control are taken from the RICE model and these emissions are measured in billion tons of carbon, the initial stock or preindustrial level of the CO_2 atmospheric stock, is taken as 590 billion tons of carbon equivalent ($s_0 = 590$). Finally the

Figure 1: Optimal probabilistic non cooperative emissions e^l_{it} for country i at each period of time t in billion tons of carbon equivalent.



discount factor per year, that appears in the objective function of problem $(P4)$ is taken as

$$\beta = \frac{1}{(1+\rho)^1} = 0.98$$

where the annual discount rate is chosen as $\rho = 0.02$.

The random disturbance ξ_t is a noise process as in (2), i.e. sequence of i.i.d. random variables and independent of the initial state s_0, with normal distribution and

$$\mathbb{E}\left[\xi_t\right] = 0, \qquad \sigma^2 = \mathbb{E}\left[\xi_t^2\right] = 10, \qquad \forall t = 1, 2, ..., T-1.$$

In our simulations we have estimate the expectation of the damages functions and its probability correspondences, over 1000 runs carried out after the corresponding 1000 values of the standard normal disturbance ξ_t. We obtain the probability value of $Prob = 0.8150$ with the target value equal to 220000.

Table 1 column 1 gives the optimal cooperative emissions e^*_{it} in billion tons of CO_2 equivalent for each country during each period of time t. These results are related with problem $(P4)$. The last row gives the cumulated emissions per country until the end of the horizon T in billion tons of carbon.

Figure 2: Optimal Cooperative Value Function P^l_{it} per country i for each period of time t in billions of 1990 USA dollars.

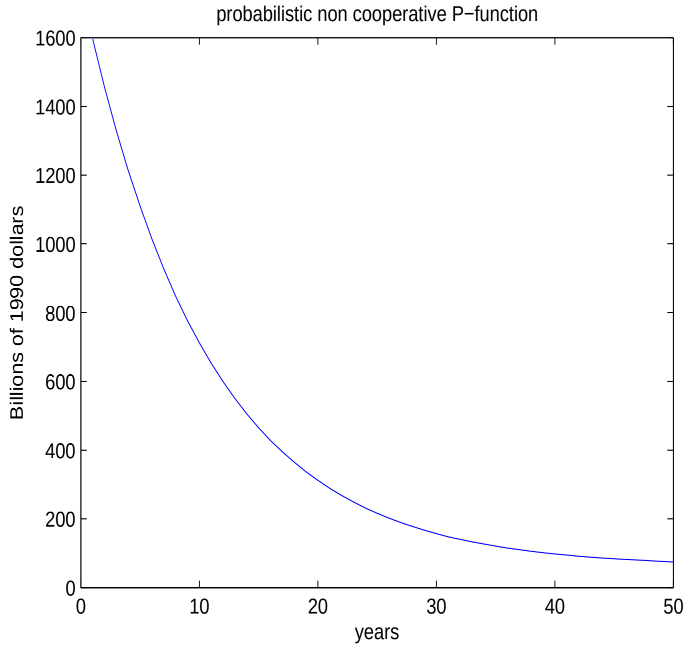


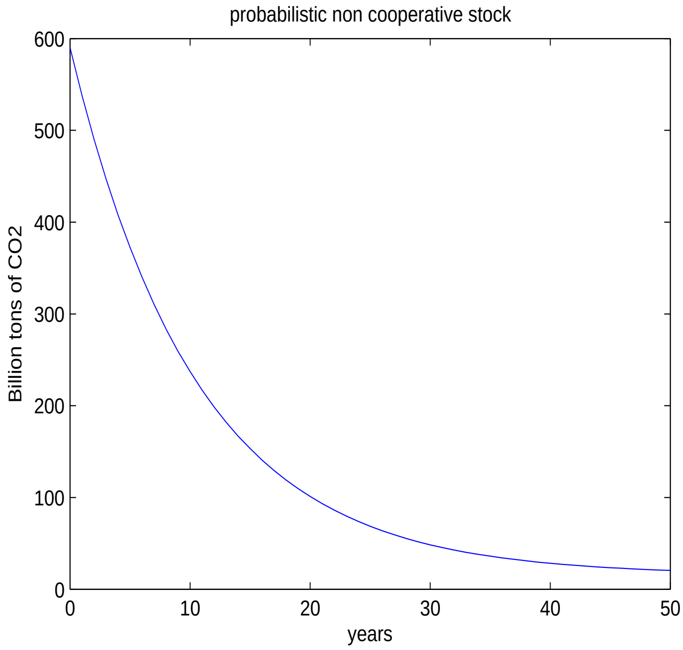
Figure 1 shows the optimal cooperative emissions e^*_{it} for country i and per each period of time t.

Table 1 column 2 gives the optimal probabilistic value function P^l_{it} for the country i during each period of time t in billions of 1990 USA dollars. These results are related with problem $(P4)$. The last row gives the cumulated value function per country and the total of the world at the end of the final period T, measured in billions of 1990 USA dollars. Figure 2 shows the optimal cooperative value function P^l_{it} for the country i and per each period of time t in billions of 1990 USA dollars.

Table 1 column 3 gives the probabilistic non cooperative optimal stock of pollutant, s^l_t at each period of time t in billion tons of carbon.

Figure 3 depicts the optimal probabilistic and non-cooperative stocks of pollutant, s^l_t for each period of time t in billion tons of carbon equivalent.

Figure 3: Optimal probabilistic non-cooperative stocks



5.2 The Real Scenario Case

A complete overview of the equations and parameter values used, can be found in Casas and Romera (2009).

The division of the world is the same as in the RICE model. There are 6 countries or regions: USA, Japan, European Union (EU), China, Former Soviet Union (FSU) and Rest of the World (ROW). The time is divided in years, the initial period (period $t = 0$) refers to year 1990. To take account on the long term impacts of stock pollutant, we take a long planning horizon of 100 years, but we will only consider results until 2030 in order to avoid boundary problems.

One assumes, by definition of target value, that

$$X_i = [55200, 27300, 63650, 10250, 6680, 124500]$$

where x_i, is the target value for each country i. Each target x_i was found by simulation, country per country. And with this target values, we obtain, for each country and their respective target values, the following optimal probability values

$$Probr = [0.9120, 0.9180, 0.8980, 0.8830, 0.8920, 0.8860]$$

Figure 4: Optimal probabilistic non cooperative emissions e^r_{it} for country i at each period of time t in billion tons of carbon equivalent.

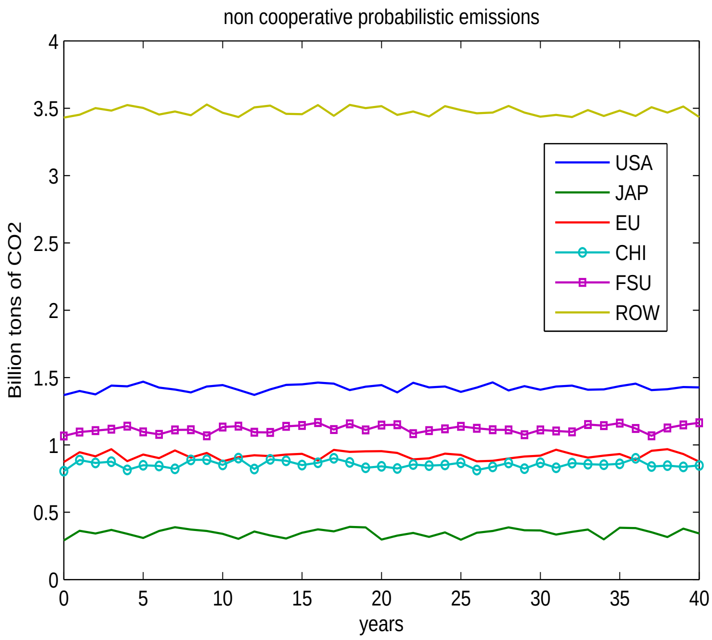


Table 2 gives the optimal cooperative emissions e^r_{it} in billion tons of CO_2 equivalent for each country during each period of time t. These results are related with problem $(P4)$. The last row gives the cumulated emissions per country until the end of the horizon T in billion tons of carbon. Figure 4 shows the optimal cooperative emissions e^r_{it} for country i and per each period of time t with target X.

Table 3 gives the optimal probabilistic value function P^r_{it} for each country i during each period of time t in billions of 1990 USA dollars with target X. These results are related with problem $(P4)$. The last row gives the cumulated value function per country and the total of the world at the end of the final period T, measured in billions of 1990 USA dollars. Figure 5 shows the optimal cooperative value function P^l_{it} for the country i and per each period of time t in billions of 1990 USA dollars with target X.

Figure 5: Optimal Cooperative Value Function P^r_{it} per country i for each period of time t in billions of 1990 USA dollars.

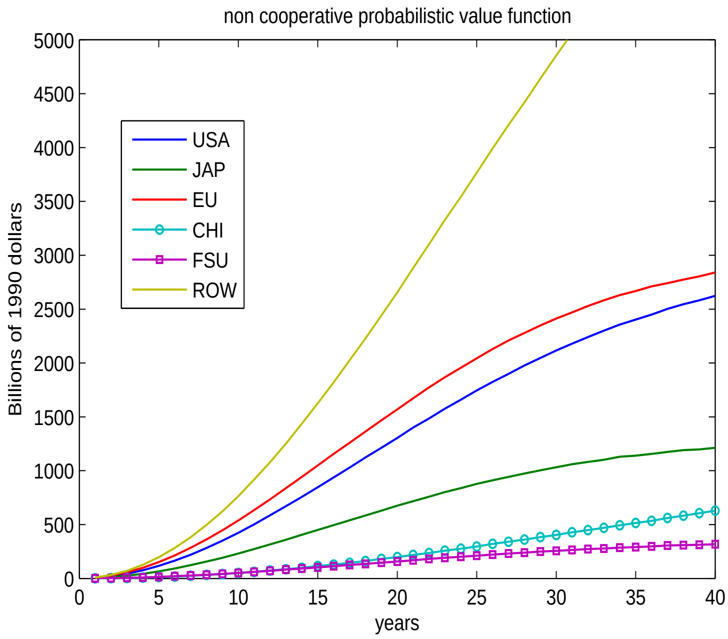
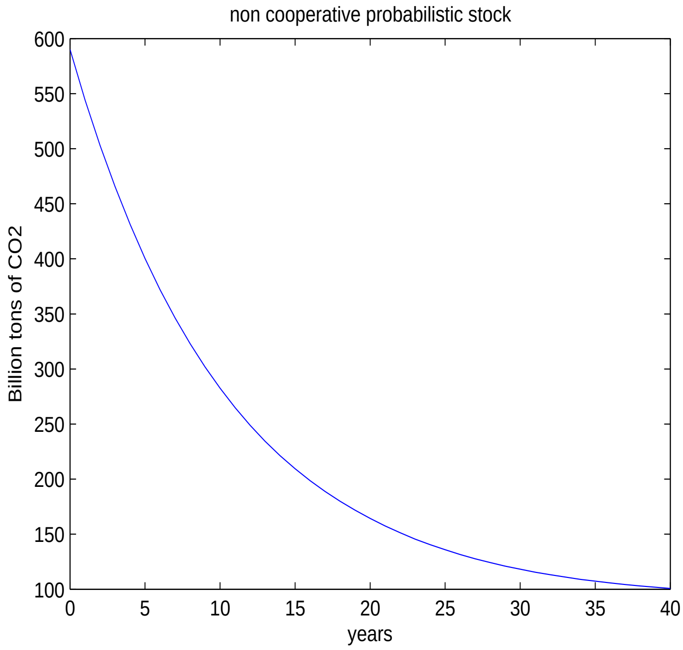


Table 4 gives the probabilistic non cooperative optimal stock of pollutant, s^r_t at each period of time t in billion tons of carbon with target X.

Figure 6 depicts the optimal probabilistic and non-cooperative stocks of pollutant, s^r_t for each period of time t in billion tons of carbon equivalent with target X.

Figure 6: Optimal probabilistic non-cooperative stocks



6 Final Remarks

In this paper we propose a new model with probability performance criteria and we obtain the existence of an optimal policy. In absence of international cooperation, these optimal policies obtained under this new perspective could be an alternative behavior for each country which finally will help reducing the international stock pollutant. Note that the target value (X) could be chosen by each country, according some particular negotiation. Usually the target value (X) should be a quantity ranging between the non-cooperative value function and the cooperative value function. These schemes are also suitable in the context of coalitional rationality.

Summarizing our results, for each country $i \in J$ and each period $t \in \mathcal{T}$ we obtain the optimal stocks pollution $\{s_t^r\}$, emissions $\{e_{it}^r\}$, the probability values and values functions $\{P_i(t, s_{t-1}^r)\}$ with a target value given.

We find of interest to consider stochastic performance criteria based on bounds of probability, i.e., MDP with percentile performance criteria where the decision-maker wants to find a policy that achieves a specific value (target) at a specified probability level α, or α-percentile criteria.

Further research could be done if we consider uncertainty about the random perturbation (the variance of the i.i.d. sequence). We propose to estimate the parameter recursively and to include the estimation in the stochastic control problem.

References

Altman, E. (1999). *Constrained Markov Decision Processes*. Stochastic Modeling. Chapman & Hall/CRC.

Bahn, O., A. Haurie, and R. Malhamé (2008). A stochastic control model for optimal timing of climate policies. *Automatica 44*, 1545–1558.

Barrett, S. (2003). *Enviroment and Statecraft: The Strategy of Environmental Treaty-Making*. Oxford: Oxford Univertity Press.

Bertsekas, D. (2000). *Dynamic Programming and Optimal Control* (Second ed.), Volume I. Belmont, Massachusetts: Athena Scientific.

Boda, K., J. Filar, Y. Li, and L. Spanjers (2004). Stochastic target hitting time and the problem of early retirement. *IEEE Transactions on Automatic Control 49 No. 3 March*, 409–419.

Carraro, C., J. Eyckmans, and M. Finus (2006). Optimal transfers and participation decisions in international environmental agreements. *Review of International Organisations 1(4)*, 379–396.

Casas, O. and R. Romera (2005, November). The international stock pollutant control: A stochastic formulation. In *Actas Seminarios Internacionales Complutenses. Madrid*. Madrid. Seminarios Internacionales Complutenses. Madrid.

Casas, O. and R. Romera (2009, February). The international stock pollutant control: A stochastic formulation. Working Paper 09-08, Universidad Carlos III de Madrid, Departamento de Estadística. Statistics and Econometrics Series 04.

Chander, P. and H. Tulkens (1995). A core-theoretic solution for design of cooperative agreements on transfrontier pollution. *International Tax and Public Finance 2*, 279–294.

Chander, P. and H. Tulkens (1997). The core of an economy with multilateral enviromental externalities. *International Journal of Game Theory 26*, 379–401.

Dechert, W. and S. O'Donnell (2006). The stochastic lake game: A numerical solution. *Journal of Economic Dynamics and Control 30*, 1569–1587.

Dutta, P. and R. K. Sundaram (1998). *The Equilibrium Existence Problem in General Markovian Games. In: Organizations with Incomplete Information: Essays in Economic Analysis. A tribute to Roy Radner*, Chapter 5, pp. 159–207. Cambridge and New York: Cambridge University Press.

Eyckmans, J. and H. Tulkens (2003). Simulating coalitionally stable burden sharing agreements for the climate change problem. *Resource and Energy Economics 25*, 299–327.

Filar, J. and L. Petrosjan (2000). Dynamic cooperative games. *International Game Theory Review 2 (1)*, 47–65.

Finus, M. (2001). *Game Theroy and International Environmental Cooperation*. Edwar Elgar, Cheltenham, UK and Northampton, USA.

Flam, S. D. (2006). Balanced environmental games. *Computer and Operations Research 33*, 401–408.

Germain, M., P. Toint, H. Tulkens, and A. de Zeeuw (2003). Transfers to sustain dynamic core-theoretic cooperation in international stock pollutant control. *Journal of Economic Dynamics and Control 28*, 79–99.

Haurie, A. and L. Viguier (2003). A stochastic dynamic game of carbon emission trading. *Environmental Modelling and Assessment 8*, 239–248.

Hernández-Lerma, O. (1989). *Adaptive Markov Control Processes*. Springer-Verlag, New York.

Jorgensen, S., G. Martín-Herrán, and G. Zaccour (2003). Agreeability and time-consitency in linear-state differential games. *Journal of Optimization Theory and Applications 119*, 49–63.

Jorgensen, S. and G. Zaccour (2001). Incentive equilibrium strategies and welfare allocation in a dynamic game of pollution control. *Automatica 37*, 29–36.

Keller, K., B. M. Bolker, and D. F. Bradford (2004). Uncertain climate thresholds and optimal economic growth. *Journal of Environmental Economics and Management 48*, 723–741.

Krass, D. and O. Vrieze (2002). Achieving target state-action frequencies in multichain average-reward markov decision processes. *Mathematics of Operations Research 27 No. 3 August*, 545–566.

Petrosjan, L. and G. Zaccour (2003). Time-consistent shapley-value allocation of pollution cost control. *Journal of Economic Dynamics and Control 27*, 381–398.

Puterman, M. (2005). *Markov Decision Processes:Discrete Stochastic Dynamic Programming*. John Wiley and Sons, New Jersey.

Rubio, S. and B. Casino (2005). Self-enforcing international enviromental agreements with a stock pollutant. *Spanish Economic Review 7*, 89–109.

Wu, C. and Y. Lin (1999). Minimizing risk models in markov decision processes with policies depending on target values. *Journal of Mathematical Analysis and Applications 231*, 47–67.

A Appendix-Tables

Table 1: Optimal probabilistic non cooperative emissions e^l_{it}, Value Function P^l_{it} and stock s^l_t for country i at each period of time t in billion tons of carbon equivalent for linear case with target X.

t	e^l_{it}	P^l_{it}	s^l_t
1	1.4650	1596.1820	537.8340
2	1.3931	1456.0505	490.3380
3	1.4307	1329.4736	447.1970
4	1.4186	1214.0498	407.9654
5	1.4591	1109.5905	372.3404
6	1.4462	1014.2473	339.9409
7	1.4156	927.3518	310.4559
8	1.3719	848.1165	283.6073
9	1.4521	777.1436	259.2796
10	1.4145	711.7523	237.1255
11	1.4315	652.7073	217.0024
12	1.4492	599.0960	198.7260
13	1.4622	550.3752	182.1240
14	1.4438	505.8222	167.0128
15	1.3876	464.8828	153.2189
16	1.4106	428.2401	140.7019
17	1.4635	395.3043	129.3756
18	1.4617	365.0139	119.0771
19	1.4110	336.9918	109.6640
20	1.4594	312.3067	101.1549
21	1.3758	288.7449	93.3357
22	1.4053	268.1352	86.2568
23	1.4513	249.6661	79.8674
24	1.3710	231.8015	73.9784
25	1.3839	216.1882	68.6377
26	1.3903	201.9763	63.7888
27	1.3899	189.0125	59.3802
28	1.4304	177.6265	55.4130
29	1.3972	166.6980	51.7731
30	1.3899	156.8977	48.4568
31	1.3715	147.8542	45.4236
32	1.4447	140.4641	42.7393
33	1.4145	132.9939	40.2688
34	1.4632	126.8678	38.0716
35	1.4166	120.5386	36.0275
36	1.4119	115.0288	34.1644
37	1.4546	110.4684	32.5135
38	1.4225	105.7413	30.9805
39	1.3903	101.3280	29.5547
40	1.4372	97.9771	28.3054

Table 2: Optimal probabilistic non cooperative emissions e^r_{it} for each country at each period of time t in billion tons of carbon equivalent with target X.

t	USA	Japan	EU	China	FSU	ROW	Total
0	1.3700	0.2920	0.8720	0.8050	1.0660	3.4300	7.8350
1	1.3997	0.3617	0.9453	0.8862	1.0946	3.4522	8.1397
2	1.3749	0.3411	0.9142	0.8660	1.1054	3.5004	8.1021
3	1.4393	0.3676	0.9681	0.8751	1.1163	3.4822	8.2488
4	1.4350	0.3389	0.8792	0.8142	1.1382	3.5233	8.1289
5	1.4683	0.3086	0.9273	0.8475	1.0966	3.5013	8.1497
6	1.4253	0.3601	0.9012	0.8426	1.0772	3.4528	8.0591
7	1.4100	0.3883	0.9578	0.8216	1.1103	3.4750	8.1630
8	1.3899	0.3707	0.9056	0.8883	1.1127	3.4472	8.1143
9	1.4325	0.3604	0.9400	0.8889	1.0675	3.5269	8.2161
10	1.4433	0.3385	0.8773	0.8502	1.1324	3.4656	8.1073
11	1.4076	0.3015	0.9077	0.9007	1.1384	3.4349	8.0907
12	1.3710	0.3565	0.9218	0.8197	1.0942	3.5055	8.0687
13	1.4120	0.3265	0.9154	0.8920	1.0922	3.5195	8.1576
14	1.4454	0.3040	0.9282	0.8819	1.1368	3.4586	8.1550
15	1.4494	0.3467	0.9337	0.8494	1.1444	3.4551	8.1786
16	1.4620	0.3723	0.8833	0.8671	1.1646	3.5233	8.2726
17	1.4545	0.3574	0.9618	0.9002	1.1133	3.4431	8.2303
18	1.4068	0.3899	0.9475	0.8690	1.1563	3.5241	8.2935
19	1.4321	0.3862	0.9511	0.8297	1.1111	3.5002	8.2104
20	1.4431	0.2959	0.9535	0.8403	1.1465	3.5148	8.1940
21	1.3894	0.3260	0.9390	0.8238	1.1489	3.4509	8.0780
22	1.4605	0.3449	0.8921	0.8541	1.0826	3.4755	8.1096
23	1.4269	0.3162	0.8993	0.8459	1.1054	3.4381	8.0318
24	1.4332	0.3497	0.9346	0.8514	1.1181	3.5151	8.2021
25	1.3934	0.2949	0.9257	0.8661	1.1378	3.4862	8.1042
26	1.4249	0.3471	0.8780	0.8121	1.1229	3.4619	8.0469
27	1.4632	0.3601	0.8809	0.8364	1.1121	3.4675	8.1201
28	1.4035	0.3862	0.8991	0.8658	1.1105	3.5168	8.1820
29	1.4356	0.3651	0.9129	0.8225	1.0748	3.4672	8.0780
30	1.4092	0.3640	0.9194	0.8671	1.1103	3.4374	8.1074
31	1.4327	0.3332	0.9629	0.8296	1.1026	3.4500	8.1110
32	1.4399	0.3534	0.9316	0.8637	1.0963	3.4349	8.1199
33	1.4097	0.3703	0.9049	0.8556	1.1512	3.4867	8.1784
34	1.4114	0.2984	0.9198	0.8515	1.1419	3.4422	8.0652
35	1.4355	0.3845	0.9317	0.8591	1.1610	3.4822	8.2541
36	1.4538	0.3816	0.8881	0.8992	1.1218	3.4417	8.1862
37	1.4072	0.3502	0.9549	0.8392	1.0674	3.5070	8.1259
38	1.4125	0.3154	0.9676	0.8452	1.1256	3.4675	8.1338
39	1.4295	0.3773	0.9316	0.8358	1.1476	3.5123	8.2341
40	1.4266	0.3413	0.8749	0.8462	1.1637	3.4347	8.0873
Total	57.0005	13.9325	36.8693	34.2008	44.7516	139.0818	

Table 3: Optimal Probabilistic Non Cooperative Value Function P^r_{it} for each country i for each period of time t in billions of 1990 USA dollars.

t	USA	Japan	EU	China	FSU	ROW	Total
1	4.949	5.366	6.843	0.544	0.605	7.952	26.258
2	19.517	11.779	25.862	1.904	2.373	31.761	93.196
3	43.159	26.372	57.065	4.296	5.242	71.152	207.286
4	75.575	42.815	99.157	7.786	9.170	126.127	360.630
5	116.150	65.099	151.599	12.175	14.107	196.526	555.656
6	164.690	92.268	213.768	17.743	20.044	282.556	791.069
7	220.261	123.562	284.116	24.530	26.683	382.268	1061.420
8	282.720	156.028	362.521	31.667	34.202	496.857	1363.995
9	349.980	191.995	446.924	40.284	42.706	622.404	1694.293
10	423.282	230.876	538.664	50.521	51.195	763.889	2058.427
11	502.264	272.354	634.226	60.595	60.609	917.211	2447.260
12	585.973	315.002	734.297	73.960	70.972	1078.946	2859.150
13	670.194	359.128	837.257	85.197	81.373	1251.438	3284.587
14	757.295	404.690	942.082	99.059	91.512	1435.908	3730.545
15	847.112	449.534	1048.226	114.474	102.285	1626.273	4187.904
16	937.201	494.782	1155.903	129.261	112.955	1818.940	4649.043
17	1029.170	540.184	1259.400	144.330	124.807	2025.260	5123.152
18	1122.824	584.754	1364.138	162.018	135.112	2227.238	5596.084
19	1213.448	629.149	1467.929	181.034	147.083	2440.439	6079.082
20	1304.186	675.464	1569.906	198.810	157.396	2655.098	6560.859
21	1399.364	717.210	1671.883	218.505	168.415	2880.092	7055.469
22	1484.428	758.421	1772.513	236.578	181.007	3100.321	7533.267
23	1575.614	800.894	1868.132	256.985	191.180	3328.057	8020.862
24	1659.194	837.317	1954.929	276.758	201.032	3542.063	8471.293
25	1745.600	878.296	2042.489	296.672	210.554	3766.427	8940.038
26	1824.538	910.501	2130.305	320.548	220.910	3991.662	9398.464
27	1898.579	943.259	2208.315	340.465	230.684	4209.940	9831.241
28	1976.696	973.152	2279.085	360.122	239.658	4419.745	10248.458
29	2045.803	1003.790	2348.330	384.243	249.861	4639.845	10671.872
30	2116.553	1031.572	2412.971	403.657	256.846	4854.243	11075.842
31	2179.239	1059.499	2470.066	427.740	265.139	5061.795	11463.478
32	2240.120	1081.922	2528.721	447.933	273.004	5268.084	11839.784
33	2300.321	1102.261	2581.613	470.361	277.833	5461.790	12194.179
34	2357.107	1130.231	2629.724	493.240	285.203	5667.494	12562.999
35	2403.557	1139.425	2668.665	514.610	290.514	5850.514	12867.285
36	2449.690	1155.875	2711.899	534.613	298.142	6042.108	13192.327
37	2501.945	1173.885	2741.848	560.813	306.571	6224.744	13509.806
38	2545.572	1191.953	2773.867	583.294	309.397	6413.434	13817.516
39	2582.507	1197.174	2804.503	606.212	313.324	6584.189	14087.908
40	2623.725	1212.716	2840.040	629.036	317.748	6774.550	14397.814
Total	52580.100	25970.554	60639.783	9802.571	6377.453	118539.337	273909.797

Table 4: Optimal probabilistic stocks of pollutant non cooperative s_t^r for each period t in billion tons of carbon equivalent with target X.

t	s_t^r
0	590.0000
1	544.5887
2	503.2917
3	465.7098
4	431.6674
5	400.5691
6	372.1760
7	346.5303
8	323.1425
9	301.9700
10	282.6366
11	265.1027
12	249.1796
13	234.5776
14	221.3860
15	209.4455
16	198.5148
17	188.5776
18	179.5473
19	171.2710
20	163.9237
21	157.1870
22	150.8389
23	145.2260
24	140.0085
25	135.2929
26	131.0180
27	127.2437
28	123.7602
29	120.5500
30	117.7626
31	115.1078
32	112.8148
33	110.6808
34	108.6996
35	106.9226
36	105.3961
37	103.9672
38	102.7126
39	101.6355
40	100.5178