Incentives and stability of international climate coalitions: An integrated assessment ☆

Valentina Bosetti [a], Carlo Carraro [b], Enrica De Cian [a,*], Emanuele Massetti [a], Massimo Tavoni [a]

[a] Fondazione Eni Enrico Mattei (FEEM) and Euro-Mediterranean Center on Climate Change (CMCC), Italy
[b] University of Venice, Fondazione Eni Enrico Mattei, CEPR, CESifo and CMCC, Italy

HIGHLIGHTS

▶ We analyse climate coalitions with an integrated assessment model.
▶ Coalitions' profitability and stability is analysed under alternative assumptions.
▶ Effective coalitions should include larger emitters (such as India and China).
▶ A coalition that achieves 550 ppm CO_2-eq is not internally stable.
▶ A stable coalition can achieve around 518 ppme in 2050 and 600 ppme in 2100.

ARTICLE INFO

ABSTRACT

This paper analyses the incentives to participate in an international climate agreement and the stability of the resulting climate coalition using the integrated assessment model WITCH. Coalition stability is assessed under alternative assumptions concerning the pure rate of time preference, the aggregation of social welfare, and the severity of climate damages. The profitability, stability, and *strong* potential internal stability of a number of coalitions, those potentially effective in reducing GHG emissions, is explored in the paper. The main conclusion is that only the grand coalition, i.e. a coalition where all world regions cooperate to reduce emissions, can maintain GHG concentration below 550 ppm CO_2-eq. However, this coalition is not internally stable, even when allowing for monetary transfers across world regions. Nonetheless, the paper also shows that *strongly* potentially internally stable coalitions exist, though of smaller size, which can mitigate global warming and limit GHG concentrations to 600 ppm CO_2-eq.

1. Introduction

The global public good nature of climate change control requires cooperation among a large set of emitters for GHG emissions to be effectively reduced. When dealing with international climate agreements, this implies that the climate coalition can be effective only if the major economies, or at least the largest GHG producing regions, participate in the agreement. The events characterising international climate negotiations of the past 20 years show the difficulty of achieving a large consensus on collaborative actions to control climate change. This is not surprising, given the free-riding incentives usually undermining global public good provision.

The theoretical literature on environmental coalitions often stresses that stable coalitions are generally small and might fail to effectively address environmental problems, especially in the case of climate change control (Cf. Barrett, 1994; Carraro and Siniscalco, 1993; Asheim et al., 2006; Bréchet et al., 2011). When the recipients of benefits are not the same of those bearing the costs and the benefits of emissions reductions cannot be appropriated exclusively by those countries paying for the costs, as in the case of GHG emission reduction, a stable agreement can hardly be achieved, thus reflecting the difficulty to provide sufficient participation incentives to widely heterogeneous countries. Only when the benefits from cooperation are small, can a stable coalition succeed at sustaining a large number of signatories (Cf. Barrett, 1994). However, in this case, the climate

agreement would not be effective at achieving large emission reductions.

Despite the slow progresses in global climate policy making, the Durban Action Platform for Enhanced Action, put in place at the end of 2011, has renewed the interest in the assessment of the effectiveness and long-term sustainability of climate coalitions. A successful international climate policy framework should build a coalition of countries with the potential to (i) effectively reduce global GHG emissions; and (ii) to give each of its members sufficient incentives to join and remain in the coalition. Such coalition should be able to meet the required stabilisation target without inducing excessively high mitigation costs and should deliver a net benefit to member countries as a whole.

In order to keep the problem tractable, coalition theory has often made various simplifying assumptions. Some of these assumptions – such as the symmetry of players and the presence of a single externality – represent a major limitations of the analysis undertaken thus far. In this paper, we try to fill this gap by assuming multiple externalities, asymmetric countries and a realistic description of the world economy. Coalition efficiency and stability is then analysed using a numerical approach based on WITCH, a global integrated model of economic and climate systems (see Bosetti et al., 2006 for a detailed description of the model).

To our knowledge, only a few studies have already used a similar approach to assess climate coalitions. Bosello et al. (2003) studies the effects of different equity rules on the incentives to cooperate using the dynamic integrated climate model FEEM-RICE. Carraro et al. (2006) uses the stylized integrated assessment simulation model ClimNeg World Simulator to show how appropriate monetary transfers may induce almost all countries to sign a stable climate treaty. Using the same model ClimNeg World, Bréchet et al. (2011) compares different stability concepts and Bréchet et al. (2010) analyses the negotiation strategy of the European Union, evaluating the incentives of key players to join future climate agreements. Nagashima et al. (2009) uses the STACO model to analyse different transfer schemes and their impact on participation incentives, global welfare and abatement efforts. However, the STACO model does not explicitly model the linkages between economic, energy, and climatic variables, but rather relies on reduced-form cost and benefit functions. The ClimNeg World Simulation model used by Bréchet et al. (2011), Carraro et al. (2006) and Chander and Tulkens (1997), as well as the FEEM-RICE model used by Bosello et al. (2003), are a step ahead as they encompass the economic, climatic, and energy dimensions.

The modelling framework used in this paper represents a further improvement over the just mentioned models. The WITCH model has two major strengths with respect to other models. First, it explicitly incorporates the gains from emission reductions in terms of avoided climate change through a set of regional damage functions that feed climate change back into the economy. Second, the equilibrium is computed as the solution of a game. The 12 model regions and/or coalitions of regions behave strategically with respect to all major economic decision variables, including emission abatement levels, by playing a non-cooperative Nash game. Therefore, when deciding whether or not to cooperate on GHG emission control, countries take into account how their decisions affect all other countries, and whether these countries will cooperate or remain outside the climate coalition. Mitigation options are fully modelled as investment choices in alternative energy technologies, abatement in non-CO_2 gases, and changes in deforestation patterns. Moreover, technological change in energy efficiency and clean technologies is endogenous and reacts to price and policy signals. Technological innovation and diffusion processes are also subject to international spillovers. This implies that the model accounts for multiple externalities, which can be partly internalised when a coalition is formed.

This paper builds on the previous work of Bosetti et al. (2009), where the potential environmental effectiveness of all 4069 coalitions that would result from the 12 regions composing the Witch model was assessed. In particular, all coalitions that could deliver the stabilisation targets commonly discussed in the policy arena were identified. In the present paper, we take stock of our previous work and, within the set of all possible climate coalitions, we focus on those that have the potential to meet an ambitious enough global mitigation target. These are coalitions whose global emission path would be consistent with long-run stabilisation of global GHG concentration at 550 ppm CO_2-eq, despite the Business-as-Usual (BaU) emission pathway of non-participating regions. For this subset of coalitions, we evaluate whether the welfare of each participating country is larger than the welfare it would obtain from withdrawing from the coalition, thus free-riding on other participants' abatement efforts (internal stability). We also check whether there are international monetary transfers that can compensate for the free-riding incentives (potential internal stability).

Given the uncertainties involved in predicting and valuing the future damages and risks from climate change, the analysis is performed under four alternative combinations of damage and discount rate assumptions. A low-damage case is based on the damage assessment in Nordhaus and Boyer (2000), while a high-damage case incorporates the more recent, upward revisions made for instance by Hanemann (2008) and Stern (2007). A low-discounting case assumes a (pure) utility discount rate of 0.1%, in line with Stern (2007), while a high-discounting case assumes the 3% value used in Nordhaus (2007). Finally, we assess the effect of different weighting in the aggregation of regions' welfare and the effect this has on our main findings.

The paper is organised as follows. Section 2 describes WITCH's game-theoretic framework. Section 3 assesses the basic individual incentives for countries to participate in a climate coalition. Section 4 brings together the incentive effects associated with different damages and abatement costs to analyse coalition formation and stability. In particular, we determine under what conditions a subset of world regions find it profitable to sign an international climate agreement (i.e. to join a climate coalition) and whether the resulting coalition is effective and stable (robust to free-riding incentives). In the concluding section, the policy implications of our analysis are discussed.

2. The game theoretic structure of WITCH

Let us start by considering a simplified setting in which countries (players) bargain over the provision of a global public good, namely the control of GHG emissions. On the one hand, reducing GHG emissions is costly because new technologies have to be developed and adopted to replace fossil fuel consumption. On the other hand, lower emissions reduce concentrations and the associated negative economic impacts of climate change.

The difficulty of providing a global public good can be studied using a simple game theoretic set up based on John Nash's early contributions. Using this approach, interactions among countries are modelled as a one-shot simultaneous game and lead either to full cooperation or to free-riding. If the benefits from an higher provision of the global public good are widely dispersed in space and in time and costs are instead high and private as in the case of climate change control, free-riding inevitably prevails and the global common resource is spoiled, leading to the well-known "tragedy of the commons".

However, the strategic interactions outlined above (also named Prisoner's Dilemma) are quite crude and do not correspond to the observed behaviour of countries facing global externalities. Indeed, international cooperation often does exist, albeit at different degrees, on a wide range of issues of common interest. In particular, over the last decades, the emergence of several international treaties to protect global common goods has been observed. How can those treaties come into force when individual countries would benefit from overgrazing the common resource? How is it possible to account for the fact that in the real world countries continuously interact, that they learn from each other's past actions, or that they can commit and design mechanisms to enhance cooperation and deter non-compliance with international norms?

These issues are at the core of the recent developments of coalition theory. Let us consider the simplest case of a simultaneous one-shot game. This game can be ideally decomposed into two stages. In the first step – the coalition game – countries decide whether or not to cooperate. In the second step – the policy game – countries choose the optimal level of carbon emissions. The decision in the first step is influenced by several factors, including what countries perceive to be the optimal strategy of all other countries in the second step of the game.

In this game setting, Carraro and Siniscalco (1993) showed that partial cooperation (i.e. only a subset of negotiating countries join the coalition) is fully rational and can emerge as the equilibrium outcome of the non-cooperative game. They show that the bargaining process may lead to different degrees of cooperation. Full cooperation is achieved when all countries form a coalition to control global emissions, thus fully internalising global damages from climate change. Partial cooperation emerges when only a subset of countries cooperate to control emissions, whereas the remaining countries free-ride. The fully non-cooperative case is yet another possible equilibrium of the game, where each country controls emissions taking into account only its own costs and benefits from GHG emissions, thus neglecting climate damages affecting the other countries.

Non-cooperative coalition theory has shown that countries can endogenously form coalitions of different sizes. In addition, the conditions for a coalition to form and not to break up in sufficient stages of the game (coalition stability) have been explored. Significant effort has also been devoted to studying mechanisms that can foster coalition formation and stability, among which transfers and issue linkage are the most prominent ones.

In this paper, we build on this literature and explore the stability and size of a set of effective coalitions. Instead of using a stylised description of regional costs and benefits, we use a detailed integrated assessment model, which allows us to consider a set of realistic features, such as heterogeneous players, technological change, and the interplay with natural resources.

The numerical analysis of climate coalition is performed with the WITCH model,[1] an energy-economy model that incorporates a detailed representation of the energy sector into an intertemporal growth model of the global economy disaggregated into 12 regions. The emphasis on the energy sector and GHG mitigation options allows technology-related issues to be studied within a general equilibrium framework characterised by environmental (expected future climate change damages), economic (exhaustible natural resources), and technology (knowledge and experience spillovers) externalities (Cf. Bosetti et al., 2006, 2008). As in a Ramsey-Cass-Koopmans model, the central planner of each region maximises intertemporal welfare by optimally choosing investments in a general final-good sector, in power generation technologies, in

R&D, and the expenditure in final energy consumption, given a production function for the final good, a budget constraint and kinetic equations for capital accumulation.

The WITCH model can be used to simulate all degrees of cooperation among the 12 macro-regions in which world countries are aggregated. The model can run in a cooperative mode where global social welfare is maximised. In this case, cooperation internalises all environmental and economic externalities. The model can also be used to compute a decentralised or non-cooperative solution, where each individual region optimises welfare taking as given the other regions' choice. In between these two extremes it is possible to model and analyse all possible combinations of partial coalitions.

It must also be noted that a great deal of cooperation is assumed anyway, because world countries are aggregated into 12 regional blocks that behave as a single negotiating entity. This might be a valid assumption for the European Union and regions that match with a single country (USA and China), but it might be overly optimistic for large regions such as Latin America and South-East Asia.

The climate scenarios obtained from the WITCH model are thus the outcome of a game in which world regions interact in a setting of strategic interdependence. Each region's social planner is a player of the game. The action of each player consists in choosing the investment paths of various energy technologies for power generation, of dedicated energy R&D, and of fuel consumption. Fuels (crude oil, natural gas, coal, and uranium) are needed to feed traditional power plants and to satisfy final consumption in the end use sector.[2] Fuel externalities are driven by the fact that the international prices of oil, gas, coal, and uranium are determined by the equilibrium between global demand and supply and the accounting of resources exhaustibility.[3] International prices endogenously reflect the availability of fuel resources, which is ultimately driven by global consumption. Fuel prices increase with global demand to reflect resource scarcity, similar to Nordhaus and Boyer (2000). For a description of the equations see Bosetti et al. (2006) and Bosetti et al. (2007). A set of reduced-form cost functions accounts for the non-linear effect of both depletion and extraction. Assuming competitive markets, the domestic fuel price is equal to the marginal cost and depends on the total cumulative quantity of fuels extracted. Therefore, even though fuel trade is not modelled, consumption in a region affects consumption and production in other regions, although each region is oblivious of its own effect on fuel prices.

The cost of some low-carbon mitigation options is also endogenous and reflects the effects of innovation and Learning-By-Doing. International innovation externalities occur because technology costs are affected by the global stock of knowledge available at a given point in time, as well as the global stock

---

[1] See www.witchmodel.org for model description and related papers.

[2] Electricity can be generated using fossil-fuel-based technologies and carbon-free options. Fossil-fuel-based technologies include natural gas combined cycle (NGCC), oil- and pulverised coal-based power plants. Integrated gasification combined cycle power plants equipped with carbon capture and storage (CCS) are also modelled. Zero carbon technologies include hydroelectric and nuclear power plants, wind turbines and photovoltaic panels. The end-use sector uses traditional biomass, biofuels, coal, gas, and oil. Together, oil and gas account for more than 70% of energy consumption in the non-electric sector. Instead, the use of coal and traditional biomass is limited to some developing regions and decreases over time. First generation biofuels consumption is currently low in all regions of the world and the overall penetration remains modest over time given the conservative assumptions on their large-scale deployment.

[3] The base year availability of fossil fuel resources for 2005 is calibrated on IEA (2004). The consequences of recent shale gas discoveries are captured by allowing a fast growth of reserves between 2005 and 2025. Our cost function translates this into a decline of prices during 2005–2025. More recent assessments of global shale gas reserves and their long-term impact on gas prices have not been included in this analysis.

of installed capacity, a proxy of Learning-by-Doing. It is assumed that the knowledge in richer countries represents the technology frontier and that each region can use part of this knowledge provided they possess a sufficiently high absorptive capacity. A region's absorptive capacity is also endogenously modelled and depends on R&D investments in the region. Therefore, knowledge spillovers are not a free good, since they require dedicated investments. In contrast, spillovers of experience are completely free, as they do not require domestic investments in order to affect technology costs. It should be mentioned that the model captures technology market failures only in the energy sector, as no general purpose R&D is assumed in this version. In addition, no learning is considered for known, yet potentially improvable technologies, such as nuclear power and carbon capture and storage.

The choice of the energy and fuel mix determines regional GHG emissions. The accumulation of GHG emissions in the atmosphere and the effects on global mean temperature are governed by a reduced form climate module. A climate change damage function identifies the climate feedback on the economy. This feedback is measured in terms of GDP losses. In this setting, a cost-benefit analysis can be performed. Regions choose their optimal investments trading off the costs and the benefits of their decisions, where costs and benefits account for environmental impacts and their feedbacks on the economy.

Since world regions are quite different, both in their climate damages and abatement costs, coalitions are characterised not only by their size but also by their composition. The set of coalitions Γ is therefore composed of 4095 possible combinations among the 12 regions, including the grand coalition γ_{GC}, which comprises all players. When formed, coalitions become players of the game. Regions that do not join a coalition are said to behave as singletons or as free-riders.

The outcome of the game, for each region or group of regions, is a consumption path over the whole simulation horizon. Regions' social planners, indexed by $i = 1, \ldots, 12$, express their preferences over the outcomes of the game using a logarithmic utility function which evaluates per capita consumption. If t denotes time, $c_{i,t}$ is per capita consumption in region i at time t, $L_{i,t}$ denotes population and γ denotes a coalition in the set of all coalitions, $\gamma \in \Gamma$, the payoff of player i is:

$$W_i(\gamma) = \sum_t L_{i,t} \ln(c_{i,t}(\gamma)) D_t \qquad (1)$$

The rate of pure time preference δ_t declines over the century and the utility discount factor D_t is given by:

$$D_t = \prod_{\tau = 1}^{t} \frac{1}{(1 + \delta_\tau)}$$

When a coalition is formed, it acts as a player and chooses actions to maximise the joint welfare of its members. Unfortunately, the joint welfare of the coalition cannot be defined as the sum of the coalition members' utility, as in many theoretical and empirical papers that work with linear or quasi-linear preferences. The choice to characterise preferences by means of a twice continuously differentiable utility function requires the use of a social welfare aggregator, which assigns a social preference to every possible profile of individual preferences. It is indeed important to recall that utility functions can be used to assess if a given consumption level is preferred or not with respect to an alternative consumption level but should not be used to evaluate by how much one consumption level is preferred to the other. Utility is not a cardinal property. This implies that utility cannot be summed across players.

In WITCH we use the following social welfare aggregator S:

$$S(\gamma) = S(W_1(\gamma), W_2(\gamma), \ldots, W_n(\gamma)) = \sum_{i=1}^{n} \Sigma_t \omega_{i,t} L_{i,t} \ln(c_{i,t}(\gamma)) D_t, \qquad (2)$$

where $\omega_{i,t}$ are weights that are used to aggregate regions. Weights, as the discount rate, can be chosen on either normative or positive consideration. We consider weights equal to the inverse of marginal utility and weights proportional to the population share and study how they affect coalition formation and other results.

The first set of weights $\omega_{i,t}$ that we consider equalises marginal utility of consumption across all coalition members in any time period. The weights "linearize" the utility functions. Therefore the social welfare aggregation rule neutralizes differences in marginal utility due to differences on consumption level across coalition members and avoids wealth transfers from wealthy players to poor players and from the future to the present. Abatement effort is distributed with the sole objective of minimising coalition's emission reduction costs, which is equivalent to equalise marginal abatement costs across regions.

The second set of weights that we consider attributes more importance to regions with higher population. With WITCH regional aggregation this implies higher weight on developing countries. According to our definition of regional impact functions developing countries suffer higher negative impacts from global warming. We thus expect that this weighting scheme would yield coalitions that to do more to reduce GHG emissions.

To compute the optimal investment paths for all climate coalitions, we assume a non-cooperative, simultaneous, open membership, Nash game with non-orthogonal free-riding, and we allow for the possibility of international monetary transfers to sustain and possibly enlarge climate coalitions. Using our game-theoretic framework and the WITCH model it is also possible to simulate issue linkage to increase the profitability and stability of coalitions by allowing regions to cooperate on other externalities, e.g. on knowledge spillovers.

The model is solved as a one-shot two stage game. In the first stage, players decide on their participation and a coalition may be formed. In the second stage, coalition members choose their optimal emission levels by internalising the environmental externality only. The other players maximise their own welfare functions taking into account the coalition's and the remaining players' decisions. The game is then solved backward. The equilibrium is found by employing the γ-characteristic function approach (Chander and Tulkens, 1997): in the unique Nash equilibrium, coalition members jointly play their best response to non-coalition members, who individually adopt their best-reply strategies.

The game exhibits positive spillovers. When a new member joins the coalition, countries outside the coalition are better off because they benefit from: (1) a better environment, (2) technology spillovers (knowledge is not a club good) and (3) lower fuel prices.

Let us now introduce some crucial definitions that are later used in the paper to study coalition formation. In order to exist, coalitions must be both *profitable* and *stable* (d'Aspremont et al., 1983; d'Aspremont and Gabszewicz, 1986; Donsimoni et al., 1986; Carraro and Siniscalco, 1991). A coalition $\gamma \in \Gamma$ is said to be *profitable* if coalition members have a higher welfare than in a scenario where the coalition is not formed (Nash equilibrium), $W_i(\gamma) \geq W_i^{\text{Nash}} \quad \forall i \in \gamma$. Where $W_i(\gamma)$ is the welfare of player i that belongs to coalition γ, and W_i^{Nash} is the welfare of player i in the fully non-cooperative Nash equilibrium.

This is a necessary, but not sufficient condition for the coalition to be formed. A second requirement concerns *stability*. A coalition γ is said to be *stable* if it is *internally* and *externally* stable. A coalition is *internally stable* if signatory countries do not have the incentive to defect and to behave non-cooperatively

when other coalition members cooperate, i.e. $\forall i \in \gamma$; $W_i(\gamma) \geq W_i(\gamma \backslash i)$, where $W_i(\gamma \backslash i)$ denotes the welfare of player i when all members but i are cooperating. A coalition is *externally stable* if there is no incentive to enlarge the coalition by including non-signatory countries: $\forall i \notin \gamma$; $W_i(\gamma) \geq W_i(\gamma \cup \{i\})$.

Finally, let us consider how transfers within the coalition can affect stability. We start by defining the set of all self-financed transfers M as follows:

$$M = \left\{ \mu = (\mu_{1,t}, \ldots, \mu_{n,t}) \, \middle| \, \sum_{i \in \gamma} \sum_t \mu_{i,t} R_{i,t} = 0 \right\} \qquad (3)$$

The region's i interest rate[4] $r_{i,t}$ declines over time and the consumption discount factor is given by:

$$R_{i,t} = \prod_{\tau=1}^{t} \frac{1}{(1 + r_{i,\tau})}$$

A coalition γ is potentially internally stable (PIS) if:

$$\exists \hat{\mu} \in M : \sum_t L_{i,t} \ln\left(\frac{C_{i,t}(\gamma) + \hat{\mu}_{i,t}}{L_{i,t}} \right) D_t \equiv W_i(\gamma, \hat{\mu}) \geq W_i(\gamma \backslash i)$$
$$\equiv \sum_t L_{i,t} \ln(c_{i,t}(\gamma \backslash i)) D_t \quad \forall i \in \gamma \qquad (4)$$

Thus, a coalition is PIS if the discounted sum of log utility of per capita consumption with transfers is at least as big as the discounted sum of log utility of per capita consumption when each region behaves as free-rider. The subset of all self-financed transfer schemes for which a coalition is PIS is defined as follows:

$$\hat{M} = \{ \hat{\mu} = (\hat{\mu}_{1,t}, \ldots, \hat{\mu}_{n,t}) \, | \, \hat{\mu} \subset M \ \text{ and } \ W_i(\gamma, \hat{\mu}) \geq W_i(\gamma \backslash i) \, \forall i \in \gamma \}, \qquad (5)$$

Among all transfer schemes that belong to $\hat{M}$, we focus on the subset $\tilde{M} \subset M$ for which each coalition member enjoys a level of discounted consumption with transfer sat least as large as when free-riding:

$$\tilde{M} = \Big\{ \tilde{\mu} = (\tilde{\mu}_{1,t}, \ldots, \tilde{\mu}_{n,t}) \, | \, \tilde{\mu} \subset \hat{M} \ \text{ and } \ \sum_t C_{i,t}(\gamma) R_{i,t}$$
$$+ \sum_t \tilde{\mu}_{i,t} R_{i,t} \geq \sum_t C_{i,t}(\gamma \backslash i) R_{i,t} \forall i \in \gamma \Big\} \qquad (6)$$

In Eq. (6) we use region- and time-specific interest rates because we are discounting wealth and not utility, as opposed to Eq. (4). Coalition members can optimally distribute transfers intertemporally at the country-specific interest rate. If we define with $\vartheta_{i,t}$ the change of consumption of coalition member i at time t induced by the transfer, the following condition must hold to preserve intertemporal budget neutrality:

$$\sum_t \vartheta_{i,t} R_{i,t} = \sum_t \tilde{\mu}_{i,t} R_{i,t} \forall i \in \gamma \qquad (7)$$

Members that expect to receive payments in the future can reduce investment in the present to compensate for the loss of productivity due to climate policy, and vice-versa.

Finally, we show that the transfer scheme $\tilde{M}$ belongs to the set of transfers that make a coalition PIS. Eqs. (6) and (7) imply that in each time period all coalition members can afford a consumption level at least as high as when free-riding:

$$\tilde{M} = \Big\{ \tilde{\mu} = (\tilde{\mu}_{1,t}, \ldots, \tilde{\mu}_{n,t}) \, \Big| \, \tilde{\mu} \subset \hat{M} \ \text{ and } \ \sum_t (C_{i,t}(\gamma) + \vartheta_{i,t})$$
$$R_{i,t} \geq \sum_t C_{i,t}(\gamma \backslash i) \ R_{i,t} \quad \forall i \in \gamma \Big\} \qquad (8)$$

Therefore, with transfers, coalition members can (at least) preserve the same welfare level of free-riding. This leads to the following definition: a coalition is *strongly* PIS if $\tilde{M} \neq \emptyset$. Being *strongly* PIS is a sufficient but not necessary condition to be a PIS coalition.

One final caveat concerns the working assumption that the interest rate does not vary with readjustments of consumption levels financed by transfers. If the change in the distribution of consumption is small, the impact on the rate of return to investments will also be small and the interest rate can be assumed to be fixed with little loss of precision. Our numerical exercises show the loss of accuracy is minimal, because transfers are typically small compared to GDP.

3. Assessing the drivers of participation in international climate coalitions

The incentives for main emitting countries to participate in climate coalitions ultimately depend on a wide range of economic and political factors, not all of which can be captured by model-based exercises. Bearing this caveat in mind, the analysis carried out in this paper covers the major economic drivers of participation incentives, including the expected impacts of climate change, the influence of distant impacts on current policy decisions (i.e. the discount rate), and the costs of mitigation policies. This Section describes how each of these three drivers are captured in the WITCH model, and how participation incentives vary across the main world regions.

3.1. Climate change impacts in WITCH

Adequate knowledge of climate change impacts is a prerequisite for well-informed climate change mitigation policies. Alternative assumptions regarding such impacts can lead to profoundly different policy insights, in terms of the outcome of cost-benefit analyses and the incentives for individual regions to participate in climate coalitions.

Estimating the economic impacts of climate change raises a number of difficult issues. First, the knowledge on the physical impacts of climate is limited, especially in relation to non-market areas or impacts. Second, assigning monetary values to climate change damages is particularly challenging. Third, the need to identify the global cost-benefit optimal emission level requires defining an indicator of the global benefits of emission reduction in terms of avoided damages. Therefore, impacts have to be aggregated across sectors and across regions, which raises equity issues, and over time, which raises intergenerational issues.

Surveying the literature on the impacts of climate change, it is immediately clear that damage estimates vary widely and uncertainty in the size of economic impact affects mostly non-market areas (Tol, 2005; Jamet and Corfee-Morlot, 2009).[5] Non-market impacts are either not considered or underestimated. Most IAMs are indeed based on out-of-date evidence, and most regional estimates are extrapolations from studies that have been carried out for one or two regions, typically the United States (DICE/RICE, MERGE and PAGE) or the United Kingdom (FUND). Most damage functions used in IAMs have not been updated according to the latest evidence on climate change, except for the PAGE model used in the Stern Review, which takes into account new evidence on more rapid warming and large-scale changes to the climate

---

[4] In a Ramsey-type optimal growth model the interest rate is endogenous and equal to the marginal product of capital. Thus, it depends on the level of development of each region. Regions with low capital accumulation tend to have higher interest rates than those with higher capital accumulation. The interest rate is always higher than the rate of pure time preference, but the differences diminish with capital accumulation, as in a standard Ramsey-type optimal growth framework. We compute the interest rate ex-post, as follows:

$$r_{i,t} = \frac{tf p_{i,t}}{1 + \theta_{1i} T_t + \theta_{2i} T_t^2} [\alpha_{KL} \left(K_{i,t}^\alpha L_{i,t}^{1-\alpha} \right)^\rho + \alpha_{EN} E N_{i,t}^\rho]^{\frac{1}{\rho}-1} \alpha_{KL} \left(K_{i,t}^\alpha L_{i,t}^{1-\alpha} \right)^{\rho-1} \alpha K_{i,t}^{\alpha-1} L_{i,t}^{1-\alpha}$$

[5] For a detailed survey of the literature on climate change impacts we refer to Appendix 3 in Bosetti et al. (2009b).

system ("system" surprise). As a consequence, previous modelling exercises exhibit impacts, which on average, are quite small compared with the results described in the Stern Review and in estimates reported in the UNFCCC report (UNFCCC, 2007) and in the IPCC's Fourth Assessment Report (IPCC, 2007).

Two alternative damage scenarios are considered here: (i) a low-damage scenario, embedded in the basic version of the WITCH model, which in turn is based on the damage assessment provided by Nordhaus and Boyer (2000); (ii) a high-damage scenario, which incorporates more recent, higher damage estimates in the range of Stern (2007) and UNFCCC (2007).

The WITCH model accounts for climate change damages, Ω, by means of regional functions that describe a reduced-form quadratic relationship between temperature change, T, and gross world product, GDP:

$$\Omega_{i,t} = \frac{GDP_{i,t}}{1 + \theta_{1i}T_t + \theta_{2i}T_t^2} \qquad (9)$$

In the low-damage scenario, for an increase in global temperature below 3 °C, climate change impacts on GDP can be either positive or negative, depending on regional vulnerability and geographic location. Above that level, damages are negative throughout the world and increase with a quadratic relationship with temperature change. The resulting pattern of regional damages in a Baseline As Usual scenario shows higher estimated losses in developing countries, in particular South Asia (including India) and Sub-Saharan Africa (Fig. 1). These two regions are expected to lose the most from climate change, especially due to higher damages in agriculture and the increase of vector-born diseases (Sub-Saharan Africa) and because of catastrophic climate impacts (South Asia including India). A recent review (Jamet and Corfee-Morlot, 2009) also indicates Africa, South Asia, and Southeast Asia as the most vulnerable regions, with GDP losses reaching more than 8% for a temperature increase above pre-industrial levels between 2 and 2.5 °C. Damage estimates for agriculture, coastal settlements, and catastrophic climate impacts are significant in Western Europe, resulting in higher damages than in other developed regions. In China, Eastern EU countries, non-EU Eastern European countries (including Russia), Japan–Korea, climate change up to 2.5 °C would bring small benefits, essentially because of a reduction in energy demand for heating purposes (non-EU Eastern European countries including Russia) or positive effects on agricultural productivity (China).

In the higher-damage scenario, impacts become negative for all countries when warming exceeds 1 °C. Global climate damages are, by the end of the century approximately twice as large as in the low-damage scenario.

Fig. 2 illustrates the time profile of the two climate damage scenarios comparing them to climate change damages that can be extrapolated from the IPCC ranges reported in UNFCCC (2007). The WITCH high-damage function follows UNFCCC data quite closely until a 1.5 °C rise in global temperature, and increases more severely beyond that, moving closer to – but remaining lower than –Stern's (2007) estimates.

3.2. Discounting and pure rate of time preference

When analysing the inter-temporal effects of climate change damages, the social discount rate and, in particular, the pure rate of time preference plays a crucial role. There is a longstanding controversy regarding the choice of the latter (Weitzman, 2001). Consistent with a long line of economists (e.g. Ramsey, 1928; Harrod, 1948; Solow, 1974), Stern (2007) argues on ethical grounds for a near-zero value, while others dismiss this assumption on the grounds that it is inconsistent with actual individual behaviour (e.g. Nordhaus, 2007; Weitzman, 2007).

Aggregate discounted impacts are vastly increased if greater weight is assigned to the far future, when damages are expected to be higher. For example, Tol (2005) finds that the median value of the social cost of carbon – an estimate of the marginal impact caused by one additional ton of carbon – increases from 7 to 39 $US per ton of carbon when the pure rate of time preference declines from 3% to 0%, i.e. when it declines from the value used in Nordhaus' DICE/RICE model to that used in the Stern Review.

This implies that cost-benefit considerations would lead to a very different abatement, depending on the value of the pure rate of time preference. Indeed, in our subsequent analysis we show that only if future damages are given enough weight, i.e. a 0.1% pure rate of time preference is adopted, the grand coalition endogenously achieves a stabilisation target that is in line with those discussed in the policy debate, e.g. 550 ppm CO_2-eq (see Section 4).

In order to take into account the existing debate on the choice of the social discount rate, we perform our analyses under two
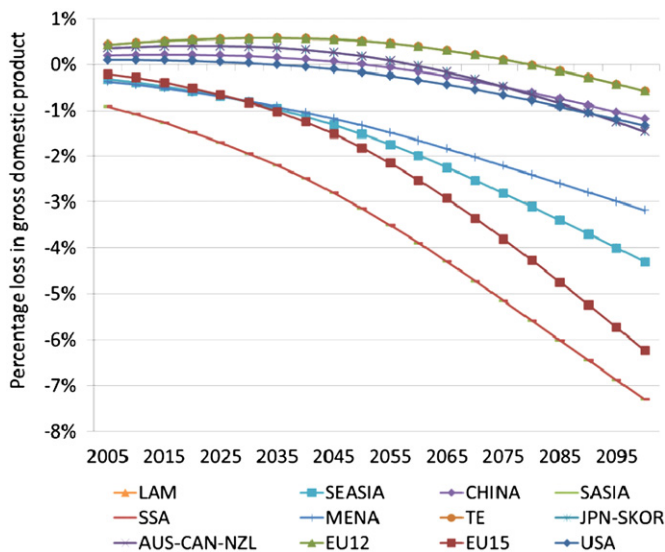

Fig. 1. Regional damage functions in the Business As Usual of the WITCH model. Low damage and high discount rate cases.
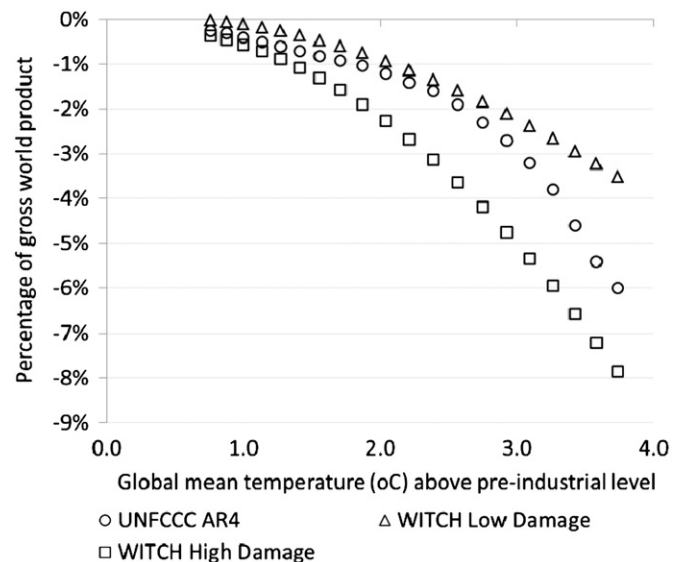

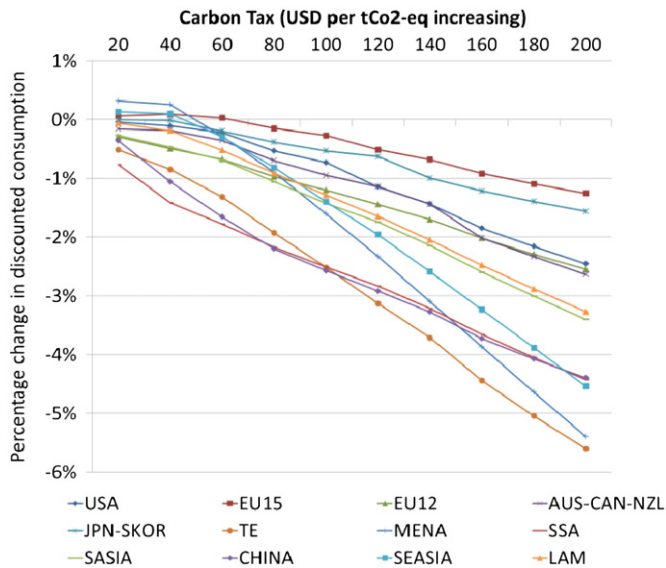**Fig. 2.** Global damage functions in the WITCH model.

Fig. 3. Discounted abatement costs for some representative WITCH regions under a range of world carbon tax scenarios.

Fig. 4. Mitigation policy costs and free-riding incentives 1. The free-riding incentive is computed as the difference in % between a region's intertemporal welfare if it withdraws from a world coalition of acting countries (the so-called "grand-coalition", see Section 3), and its intertemporal welfare if it participates in the world coalition. A 0.1% annual pure rate of time preference is used to compute the present value of welfare. 2. Cumulative consumption gap relative to BaU in present value terms over 2015–2100, using a 3% annual discount rate.

different assumptions regarding the pure rate of time preference, namely 3% and 0.1%.[6]

3.3. Abatement costs

The incentives to participate in climate coalitions are also shaped by mitigation policy costs. In WITCH, instead of aggregated and static marginal abatement costs curves the model traces the effects of moving towards more costly technologies, of investing in R&D, of switching fuels and changing forest management. In particular, countries differ in their demand for energy, the cost of fuels, the distance from the technological frontier in carbon-free technologies, as well as the potential for renewable penetration.

Fig. 3 shows the discounted consumption loss of alternative global carbon taxes for some representative regions of the WITCH model. Some countries gain slightly for moderate taxes either because of fuel switching or because of their comparative technological position. For higher carbon taxes, fossil fuel producers such as the Middle East and non-EU Eastern European countries (including Russia) are the biggest losers, reflecting their very high energy/carbon intensity. Within the group of developed regions, Western Europe would face smaller costs than the United States, mainly reflecting their lower energy and carbon intensity. Energy demand and large dependence on coal are the main drivers for high costs in China. Latin American can benefit from low costs abatement options in the land use sector and this reduces consumption losses vis a vis other developing countries.

The larger a region's resulting mitigation costs, the smaller its incentives to participate in a climate coalition, ceteris paribus. One possible measure of free-riding incentives that is introduced and discussed below is the difference (in %) between a region's welfare (defined as the discounted sum of the logarithm of future domestic per-capita consumption) if it free-rides on a world coalition of acting countries, and its welfare if it participates in that coalition. As Fig. 4 shows, there is clearly a positive relationship between the indicator of free-riding incentives and the
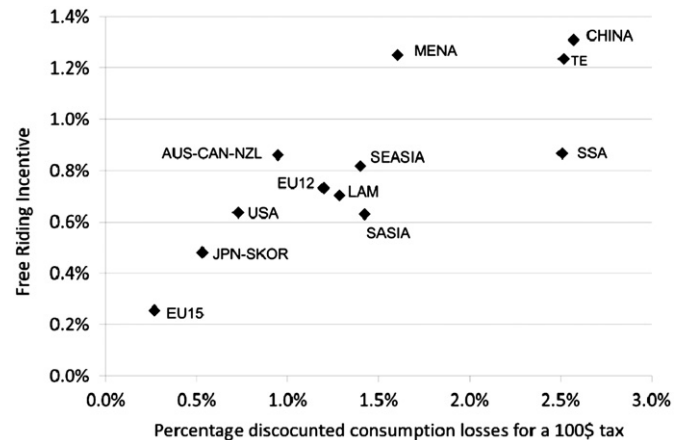
overall consumption loss induced by a given world carbon tax – set here at $US100 per ton of CO_2-eq. This is because climate coalitions are assumed to implement an efficient climate policy, i.e. to equalise marginal abatement costs across all participating regions. As a result, countries that face larger costs from a given world carbon price can expect to gain less from joining an international coalition, and therefore have larger incentives to defect, ceteris paribus.

4. Analysing coalition formation and stability

4.1. Cooperative versus non-cooperative outcomes

We start our analysis from the most environmental effective coalition, namely the grand coalition. A world social planner maximises the aggregate global welfare, which is defined as the weighted sum of regional welfares, using weights that ensure equal marginal abatement costs worldwide. Later we see the effect on coalitions of changing aggregating weights.[7]

We compare this optimal outcome with the non-cooperative outcome where each of the 12 regions is assumed to choose the optimal path of a set of choice variables (investments in physical capital, in different energy technologies, in R&D, etc.) to maximise its own social welfare function. In this framework, each region takes its decisions individually, given the action of the other players. The outcome of this non-cooperative game is an open-loop Nash equilibrium.

Fig. 5 shows the implications of the two solution concepts for global GHG emissions under the different assumptions about the damage and discounting assumptions. In the non-cooperative case, shown in the panel on the left, world emissions grow throughout the century. Little abatement is undertaken with respect to the case where climate damage is not anticipated and

---

[6] .Following Weitzman (2001), the pure rate of time preference is also assumed to be time-declining from 3% in 2005 to 2.35% in 2100 in the high case and from 0.1% to 0.08% in 2100.

[7] .In this analysis we assume cooperation is only on the climate externality. The WITCH model incorporates other economic externalities related to the use of natural resources and to the production and diffusion of knowledge and experience. However, this paper analyses the incentive to form climate coalitions, independently from linkages with other issues. In that context, it is assumed that countries decide whether or not to cooperate only on the environmental externality. Cooperation on technological externalities and on the use of natural resources is not considered.
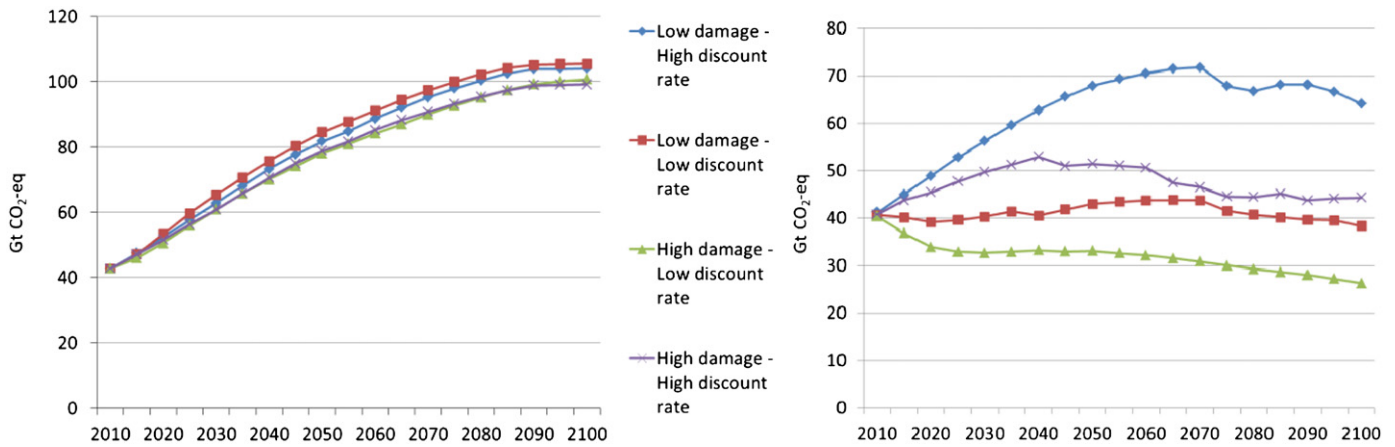
Fig. 5. All GHG emissions: Sensitivity to alternative damage and discount rate assumptions in the non- cooperative (left hand side panel) and cooperative (right hand side panel) solutions.

emissions would grow up to 109 and 113 Gton CO_2-eq by the end of the century in the high- and low-discount cases, respectively. Individual regions do not internalise the negative externality they impose on the other regions and they only take into account the domestic benefits of their climate policy. As a consequence, late century emissions are only slightly reduced to 100 and 105 Gton CO_2-eq, when damages are high and low, respectively. The choice of the pure rate of time preference or of the damage scenario, in this instance, does not make much of a difference. It must be noted that lowering the discount rate has two opposing effects on emissions. On the one hand, a lower discount rate induces lower emissions because climate change negative impacts far in the future are discounted less. On the other hand, a lower discount rate increases the marginal benefit of future consumption with respect to present consumption and stimulates savings (and investments), which lead to a higher BaU emission trajectory. In other words, countries, which see far in the future, not only invest more in environmental protection, but also in all other sectors. With our model specification we find that the "economic growth" effect of lower discounting is important and more than compensates for the "environmental protection" effect in the absence of cooperation, in the low damage case (see Fig. 5, panel on the left).

If countries gather in a grand coalition, the environmental externality is fully internalised and emissions decline drastically (right hand side panel of Fig. 5). In addition, sensitivity to underlying assumptions is far greater than in the non-cooperative case. Under these damage/discounting assumptions, the optimal emission path would be such to stabilise long-run GHG concentration at about 546 ppm CO_2-eq, when the pure rate of time preference is 0.1%, and 672 ppm CO_2-eq when it is 3%. This is equivalent to a global mean temperature increase of 2.4 °C and 2.9 °C in 2100 with respect to the pre-industrial level.[8] When looking at the relative effect of lower discounting and of higher damages, the former has a sizeable greater impact, in particular on short-term emissions.

The difference between the cooperative and non-cooperative outcomes gives an indication of the gains from cooperation, which are substantial. The damage and discounting assumptions affect the size of cooperation gains. A high discount rate reduces the benefits from cooperation, even more so when damages are low.

Table 1
Analysis of the environmental achievements (GHG concentrations in 2100) of potentially effective coalitions, cost-benefit mode, high-damage/low-discounting case.

	Overall GHG concentration (ppm CO_2 eq)	
	2050	2100
Non-participating regions		
None (Grand coalition)	507	546
Africa	518	603
Africa, Latin America	532	612
Africa, Non-EU Eastern Europe	531	603
Africa, Middle East and North Africa	529	609
Africa, South East Asia	526	598
Africa, South East Asia, Non-EU Eastern Europe	529	603

4.2. Environmental effectiveness, profitability and stability of climate coalitions

With 12 regions the number of possible coalitions is very large (4095 coalitions). However, we consider only a smaller number of coalitions that are politically meaningful and have the potential to generate emission reductions that lead to the 550 ppm CO_2-eq 2100 atmospheric targets (approximately 2.4 °C increments in global mean temperature with respect to the pre-industrial level).

A coalition is said to be politically meaningful if it includes at least high-income countries (United States; European Union; Japan and Korea; Australia, Canada and New Zealand). A coalition is said to be potentially effective (PEC) if participating regions' lower bound emission levels added to the BaU emissions of non-participating countries yield a 550 ppm CO_2-eq GHG concentration (or below) at the end of the century. More precisely, PECs are identified by comparing, for each coalition structure, the path of singletons' BaU emissions and the world emission path required to meet the target both in 2050 and 2100. This comparison is repeated four times for each coalition structure, corresponding to the four different non-cooperative baselines derived from alternative damage and discount rate assumptions.

The politically meaningful PECs at the end of the century are listed in Table 1. These PECs include all industrialised countries and both China and India. This is a first important result. No coalition is effective (i.e. can achieve the 550 ppm target) if it does not include both China and India.

We perform a cost-benefit analysis of each of these coalitions under different assumptions on the pure rate of time preference,

---

[8] The model assumes a climate sensitivity equal to 3.

damage scenario, and weighting assumptions.[9] Let us first discuss the environmental effectiveness of PECs and then look at their profitability and stability.

Among PECs, only the grand coalition is found to keep GHG concentration below 550 ppm CO_2-eq by 2100. This is illustrated in Table 1, which shows the environmental performance of the grand coalition, together with six other large coalitions that have the technical potential to meet the target. The composition of these coalitions suggests that large emitters such as China and India should always be included. By contrast, Sub-Saharan Africa is always excluded from these sub-coalitions as it is realistic to assume that, if any other region stays out, then Sub-Saharan Africa will necessarily follow, invoking equity arguments such as the right to grow. Even if only Sub-Saharan Africa behaves as a singleton, the 550 ppm CO_2-eq target is exceeded in 2100. Leaving an additional region out of the coalition raises GHG concentration above the 600 ppm CO_2-eq target.

Two main factors explain the failure of smaller coalitions to achieve the 550 ppm CO_2-eq concentration target:[10]

1. Smaller coalitions internalise only a fraction of the environmental externality. The fraction of damage that is not internalised drives a gap between the coalition and the global marginal benefit of emission reduction. The larger this gap, the smaller the emission cut that the coalition achieves. However, there is not a clear relationship between coalition size, coalition marginal damage, and emission reductions, because damages are not evenly distributed among world regions. This implies that the inclusion in the coalition of regions facing larger damages from climate change increases coalition's marginal damage, and the optimal level of emission cuts, ceteris paribus.
2. As coalitions become less inclusive, the number of free-riders obviously increases; these countries might simply keep emissions unchanged, or even undertake some small abatement due to the lower price of carbon-free technologies, but most likely they increase emissions and undermine some of the coalition's emission reduction.

These two factors can be illustrated by looking at a coalition structure under which only Sub-Saharan Africa does not participate and behaves as a singleton (Fig. 6). First, by leaving out Sub-Saharan Africa, a large share of the global damage is not internalised (see Fig. 1, where South Asia and Sub-Saharan countries would suffer the largest damage). As a consequence, such coalition would achieve a lower abatement effort compared to the grand coalition. The reaction of the other countries is a moderate increase in emissions, at most about 6% (1.3 GtCO$_2$-eq). Second, and more important, the emissions of Sub-Saharan Africa itself dramatically increase when it behaves as a singleton. The global increase in emissions is thus mostly due to the behaviour of Sub-Saharan African countries, which more than double emissions by the end of the century.

It is worth noting that, although the emissions of the Sub-Saharan African region are larger than what they would be in the
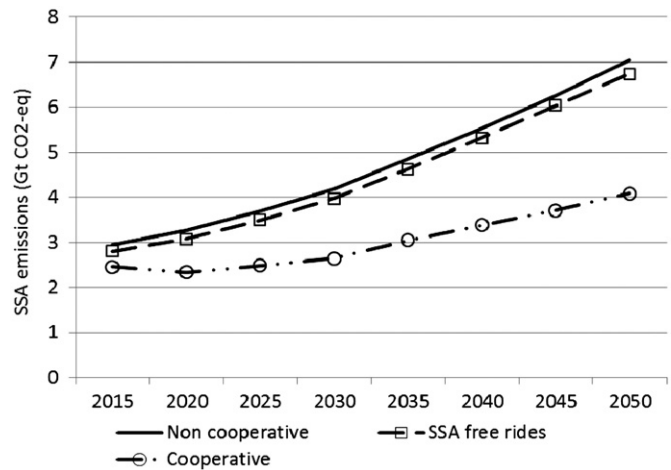

Fig. 6. An illustration of free-riding incentives: the case of Africa. GHG emissions in Sub-Saharan Africa.

grand coalition, they are still lower than in the non-cooperative scenario. Although Sub-Saharan Africa could emit as much as in the non-cooperative scenario or even more, it is not optimal for these countries to do it. The reason is the presence of international technology transfers from coalition members that reduce the costs of carbon-free technologies outside the coalition itself, thus inducing some emission reductions in the Sub-Saharan African region as well.[11]

An important factor in shaping the effectiveness of the coalition is the weight given to future damages (i.e. the pure rate of time preference). A higher rate implies, for the grand coalition, an addition of GHG concentration in the atmosphere equal to 125 ppm CO_2-eq compared to the low discounting case (approximately equivalent to an additional warming of 0.5 °C). Smaller coalitions, already reaching more than 600 ppm CO_2-eq with a low discount rate, would lead to an increase of 80 ppm CO_2-eq or more.

Finally, the coalition effectiveness is obviously affected by the weights used to aggregate different regions when maximising aggregate welfare. The results discussed so far are based on normative weights equal to the inverse of marginal utility, which imply a cost-efficient allocation of abatement within the coalition. Table 2 illustrates the grand coalition and compares the solution based on normative weights with that based on a weighting scheme more favourable to developing countries, e.g. proportional to the population share. This aggregation scheme increases the environmental effectiveness of the coalition in the short and medium term, and also in the long term with a sufficiently high discount rate. However, with a low discount rate the population social welfare aggregator yields higher GHG emissions concentrations in 2100.

International climate coalitions need not only to be environmentally effective, but should also be self-enforcing. In other words, a coalition should be profitable and stable, or at least potentially stable. As noted in Section 2, a coalition is profitable if each cooperating player achieves a welfare larger than in the non-cooperative scenario. A coalition is internally stable if the welfare of each participating region is larger or equal to the welfare it would obtain from staying out of the coalition thus free riding on coalition members' abatement efforts. As an example, when

---

[9] In what follows, we only report results for the high damage scenario which is most conducive to significant emission reductions. If a coalition is not environmentally effective in this case, it cannot be effective under lower damages.

[10] One assumption is crucial here and may alter these results: the absence of negative emissions technologies, or any other technology that might alter the climate (i.e. geo-engineering). If one assumed that by means of bio-energy and carbon capture and sequestration, or direct CO_2 capture, or other technologies that alter the incoming solar radiation, cooperating countries could dramatically change the climate, than the requirements on the dimension and composition of a coalition to endogenously produce the 550 ppm CO_2-eq by 2100 target could be substantially different.

[11] Depending on the stringency of the target and the characteristic of the free-riding countries either the energy market or the innovation effect may prevail, see Bosetti and De Cian (2011) for a throughout discussion of these competing effects.

Table 2
Analysis of the environmental achievements of potentially effective coalitions, cost-benefit mode, high-damage/low and high discounting using different weights.

	Overall GHG concentration (ppm CO_2-eq)		
	2030	**2050**	**2100**
Normative weight (equal to the inverse of marginal utility) – 0.1% PRTP	480	507	546
Population weight (equal to the inverse of marginal utility) – 0.1% PRTP	473	502	556
Normative weight (equal to the inverse of marginal utility) – 3% PRTP	506	574	672
Population weight (equal to the inverse of marginal utility) – 3% PRTP	489	538	622

Table 3
Profitability, stability and strong potential internal stability of potentially effective coalitions.

Non-participating regions	Profitability	Internal stability	Strong potential stability	Concentrations in 2100 (ppm)
Low PRTP (0.1%) – Normative Weighted				
None (Grand coalition)	NO (Africa, China)	NOT STAB (all)	NOT PIS	546
Africa	NO (China)	NOT STAB (All)	NOT PIS	603
Africa, Latin America	NO (China)	NOT STAB (All)	PIS	612
Africa, Non-EU Eastern Europe	NO (China)	NOT STAB (all)	PIS	603
Africa, Middle East and North Africa	NO (China)	NOT STAB (all)	PIS	609
Africa, South East Asia	NO (China)	NOT STAB (all)	PIS	598
Africa, South East Asia, Non-EU Eastern Europe	NO (China)	NOT STAB (all)	PIS	603
High PRTP (3%) – normative weights				
None (Grand coalition)	YES	NOT STAB (all but China and Latin America)	PIS	672
Low PRTP (0.1%) – Population Weighted				
None (Grand coalition)	NO (all but Africa, South Asia, South-East Asia)	NOT STAB (all but Africa)	NOT PIS	556

In parenthesis countries for which coalitions are not profitable or internally stable.

checking the stability of the grand coalition, we need to run simulations in which any of the 12 regions is assumed to deviate, while the others continue to cooperate. Let us recall that, although not stable, a coalition might be Potentially Internally Stable (PIS) if there is a transfer scheme that gives each member at least its free-riding pay-off and shares the remaining surplus (McGinty, 2007).

Table 3 reports profitability, internal and potential internal stability for different coalitions. As expected, all large coalitions considered in the analysis are not profitable, as there is always at least one region, namely China and/or Sub-Saharan Africa, which is worse off than in the non-cooperative case.

The second column from the right in Table 3 reports results on whether a coalition can be stabilised through a set of transfers $\tilde{M}$, i.e. whether the coalition is *strongly* PIS. By aggregating Eq. (6) over all coalition members and by recalling that $\sum_{i \in \gamma} \sum_t \tilde{\mu}_{i,t} R_{i,t} = 0 \ \forall \ \tilde{\mu} \in \tilde{M}$, we test if $\sum_{i \in \gamma} \sum_t C_{i,t}(\gamma) R_{i,t} \geq \sum_{i \in \gamma} \sum_t C_{i,t}(\gamma \backslash i) R_{i,t}$.

First, the grand coalition is not strongly PIS. The aggregate, discounted surplus from cooperation is equal to 477 USD trillions over the 2005–2100 time horizon, while the sum of the gains from free-riding is equal to 680 USD trillions. What about the coalition that includes all regions but Africa, i.e. the most environmental effective of all partial coalitions? Table 3 shows that this coalition is not potentially internally stable, but the gap is now only 2% of the aggregate discounted consumption gain of the coalition.

The smaller PECs are strongly PIS. Therefore, the lowest level of GHG concentration that can be achieved by a stable coalition is slightly above 518 ppm CO_2-eq in 2050 and around 600 ppm CO_2-eq in 2100 (see Table 3 again).

To provide greater insight on internal stability, and how this interplays with crucial assumptions on the aggregation and discounting assumptions, Fig. 7 reports free-riding incentives across regions, e.g. the difference in consumption between free riding regions and regions participating in the grand coalition.

Let us start by looking at the light grey bars, which refer to the central case of 0.1% discounting and normative weights. The Middle East–North Africa region, China, the rest of Africa, and non-EU Eastern European countries are estimated to have the largest incentive to free ride. By contrast, developed countries have the lowest free-riding incentives, with the exception of the Australia–Canada–New Zealand region.

When the pure rate of time preference used in the analysis is 3% instead of 0.1% (see second row in Table 3) each region's welfare in the grand coalition is larger than in the non-cooperative case. Therefore, the grand coalition is profitable. In addition, the incentives to free ride are now much smaller, although the coalition is neither stable nor internally stable. We find that the discounted surplus, in consumption terms, generated by cooperation (133 USD trillions) is four times greater than the discounted aggregate surplus from free-riding (35 USD trillions).

This result confirms a well-established conclusion of coalition theory. When gains from cooperation are large, as in the 0.1% case, free-riding incentives are also likely to be high. On the contrary, when gains from cooperation are small, as in the 3% case, free-riding incentives are reduced significantly. In Fig. 7, the black bars reiterate this conclusion by showing the incentive to free ride when the discount rate is higher. A clear message emerges from our analysis: cooperation is indeed possible, profitable, and potentially stable, only if the environmental target is moderate (at least compared to what currently discussed in the policy debate), i.e. around 600 ppm CO_2-eq by 2100.

Finally, we test whether a different set of aggregation weights affects the results (see the last row in Table 3). When using
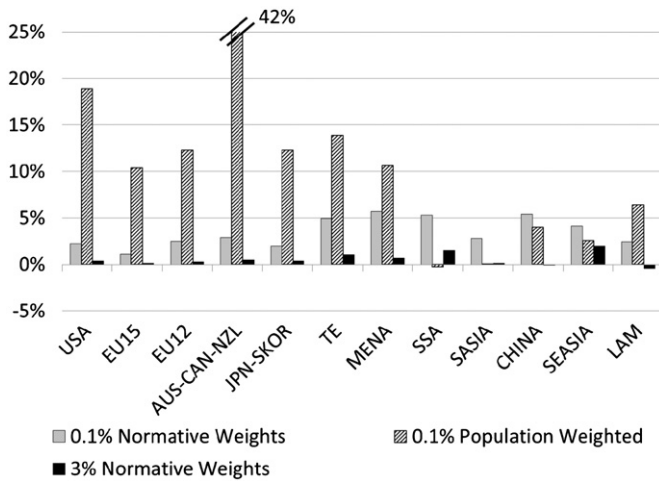
Fig. 7. Estimated ranking of free-riding incentives across regions % Difference in consumption between free-riding on and participating in the grand coalition.

population-based weights, the grand coalition is profitable only for Africa, South Asia, and South-East Asia and it is not internally stable (only Africa sees welfare gains from cooperation that compensate incentives to deviate). Fig. 7 (upward diagonal bars) shows that when population weights are used, the regional pattern of free-riding incentives is reversed with respect to the cost-effective abatement allocation. First, with population weights the abatement allocation is no longer efficient, as it does not equalise marginal abatement costs across regions. In addition, perceiving a larger damage, the grand coalition abates more. The effort is shared across all members, which all abate more compared to the normative aggregation, with the exception of Africa, South Asia, and South-East Asia. These regions have two characteristics. First, due to their high population they receive the largest weights in the social welfare function, together with China. Second, they have the largest benefits from cooperation, and in fact, cooperation becomes profitable for them. These characteristics lead to a high benefit-cost ratio, which explains their very low free riding incentives.

5. Conclusions

The Durban Action Platform has resuscitated the hope that an international climate policy agreement might be signed in the next few years. By replacing the long standing divide between Annex 1 and non Annex 1 countries, it has opened up a new negotiation phase in which new coalitions of major emitters could be formed. This policy advancement calls for new explorations of which coalitions can and should emerge from the next negotiation rounds. The integrated assessment modelling community, which plays an influential role in the IPCC process, has mostly focused on the analysis of the effectiveness of coalitions (mostly the grand coalition) in attaining climate stabilisation. Considerably less attention has been devoted to the assessment of the conditions under which such coalitions would form in the first place, and whether they would be able to stick together for a sufficiently long period. Given the policy relevance of this question, more research should be devoted to this issue.

This paper contributes to this debate by analysing the incentives and the stability of international climate coalitions using the integrated assessment model WITCH. The nature of the model allows us to provide a richer and more exhaustive description of the many forces at play than currently done in the literature,

including multiple externalities (related climate change, technological innovation and fuel prices), investments in mitigation technologies and knowledge, and regional climate damages.

The WITCH model has been used to perform an analysis of coalitions' profitability and stability under alternative combinations of climate damage and social time rate of preference, and assessed the effect of two different schemes for aggregating welfare across countries. As expected, the scenario with high climate damages and low discount rate is shown to be the most conducive to cooperation for emissions control as it increases the size of the expected present value of climate change damages. The weighting scheme proportional to population size also increases the effort of the coalition by giving greater weight to developing countries, where most of climate damages are projected to occur.

Our cost-benefit analysis suggests that only the grand coalition finds it optimal *as a whole* to stabilise total GHG concentration below 550 ppm CO_2-eq (in the high-damage/low-discounting case). Smaller coalitions, including the grand coalition excluding Africa only, achieve somewhat less ambitious targets, above 600 ppm CO_2-eq. This is because they do not fully internalise the global environmental externality and allow a larger number of (non-participating) countries to free ride.

Although the grand coalition *as a whole* has an incentive to achieve the 550 ppm CO_2-eq target, it is not internally stable. Most regions gain more from non-participation than from participation to the grand coalition. This is also true for smaller coalitions. The grand coalition is not even *strongly* potentially internally stable (PIS) – i.e. there is not a set of international monetary transfers that would make discounted consumption at least as high as when free-riding, for all coalition members. This is because the overall welfare gain from the grand coalition relative to the non-cooperative outcome is not large enough to give each country its free-riding pay off. After compensating all losers in the coalition, the remaining coalition surplus is too small to offset free-riding incentives in pure consumption terms.

The coalition that includes all regions except Africa is also not *strongly* PIS, but discounted consumption after the redistribution of cooperation gains is only 2% lower than discounted consumption when regions do not cooperate. All other coalitions that are able to stabilize GHG concentrations to about 600 ppm CO_2-eq in 2100 are instead *strongly* PIS.

Although not strongly PIS, the grand-coalition and the coalition composed by all regions but Africa could be *weakly* PIS – i.e. discounted utility, rather than per capita consumption, with the self-financed transfer scheme is higher than discounted utility when free-riding, for all coalition members. Weakly PIS coalitions might in fact exploit international and intertemporal differences in the marginal utility of consumption of coalition members to devise a transfer scheme that neutralizes free-riding incentives. Lower transfers of consumption would be necessary to compensate losses of welfare due to cooperation and the set of PIS coalitions might be larger. With transfer schemes less strict than those examined in this paper it might be possible to attain a level of GHG concentrations lower than 600 ppm CO_2-eq in 2100. Therefore, there is certainly space for further research in this area.

Nonetheless, a clear message emerges from our analysis: cooperation is possible and profitable, but it is unlikely to ensure the stabilization of GHG concentrations at levels consistent with the 2 °C threshold. It is nonetheless possible to attain an intermediate stabilisation objective (such as 600 ppm CO_2-eq), provided substantial financial transfers across regions are implemented.

Our findings are subject to a number of limitations. Even though extensive sensitivity analysis has been carried out to assess the robustness of the main results, it should be acknowledged that our model-based analysis relies on strong assumptions. In particular, there are wide uncertainties surrounding

future emission trends,[12] the market and non-market impacts from climate change, the likelihood and effect of catastrophic risks, and the cross-country distribution of these damages and risks.

Furthermore, the analysis focuses on immediate, irreversible and self-enforcing participation to mitigation action, thereby abstracting from other possible bargaining options including e.g. delayed participation, renegotiation, sanctions or joint negotiation in multiple areas (e.g. linking climate and international trade negotiations). For instance, a major emitting country may have greater participation incentives than found here if it expects its withdrawal to prevent the formation of *any* coalition. The use of the internal–external stability concept might thus be somewhat too restrictive when dealing with large and influential regions.

The co-benefits from mitigation action, e.g. in terms of human health, energy security or biodiversity, are also not taken into account. Other studies suggest that such co-benefits are large, although the participation incentives they provide are dampened by the fact that some of these co-benefits could be reaped through direct policy action – in particular, local air pollution might be reduced at a lower cost through direct policy action than through reductions in GHG emissions (Bollen et al., 2009; Burniaux et al., 2008). Removal of fossil fuel subsidies, one of the few policies to yield potentially both climate and economic benefits, is also omitted from the analysis. Insofar as phasing out subsidies would bring an economic gain and lower the carbon intensity of a number of (mainly developing) countries, incentives to participate in international mitigation action could improve.

Another potential limitation of the analysis is to assume that even if a country benefits from an international coalition relative to a BaU scenario, it will always prefer to free-ride if that option is even more profitable. While this assumption merely derives from individual welfare maximisation, current international redistributive policies such as official development aid point instead to some degree of altruism. Against this background, there might be a possibility for some countries to sign an agreement even if they could in principle gain more from free-riding on other countries' abatement efforts. We test this possibility by computing the cost for developed countries of using additional resources (additional to the coalition surplus) to stabilise the grand coalition, i.e. to give each other region its free-riding pay off. These calculations show that with a 3% loss in the discounted value of their consumption levels, industrialised countries could stabilise the grand coalition in the high-damage/low-discounting case, i.e. all other participating regions could be fully compensated for their free-riding incentives through financial transfers, thereby bringing them into an agreement.

Finally, two crucial assumptions affect the results presented in the paper: the absence of negative emissions technologies or any other technology that might alter the climate (i.e. geo-engineering), and the absence of adaptation policies.

If cooperating countries could unilaterally change the climate by using bioenergy with carbon capture and sequestration, direct CO_2 capture, or other technologies that alter the incoming solar radiation, then the requirements on the dimension and composition of a coalition to endogenously produce the 550 ppm CO_2-eq by 2100 target could be substantially different. Adaptation policies, by providing benefits that are local, at least within the boundary of macro-regions considered in this model, could also change free-riding incentives and thus the willingness to cooperate on climate change mitigation. These topics are left for future research.

References

Asheim, G., Froyn, C.B., Hovi, J., Menz, F.C., 2006. Regional versus global cooperation on climate control. Journal of Environmental Economics and Management 51, 93–109.

Barrett, S., 1994. Self-Enforcing International Environmental Agreements. Oxford Economic Papers 46 pp. 878–894.

Bollen, J., Guay, B., Jamet, S. and Corfee-Morlot, J., 2009. Co-benefits of Climate change Mitigation Policies: Literature Review and New Results, OECD Economics Department Working Papers, No. 692, OECD, Paris.

Bosello, F., B. Buchner, Carraro, C., 2003. Equity, development, and climate change control, Journal of the European Economic Association 1, 601–611.

Bosetti, V., Carraro, C., Galeotti, M., Massetti, E., Tavoni, M., 2006. WITCH: a world induced technical change hybrid model. The Energy Journal, Special Issue. Hybrid Modeling of Energy-Environment Policies: Reconciling Bottom-up and Top-down, 13–38.

Bosetti, V., Carraro, C., Massetti, E., Tavoni, M., 2008. International technology spillovers and the economics of greenhouse gas atmospheric stabilization. Energy Economics 30 (6), 2912–2929.

Bosetti, V., Carraro, C., Duval, R., De Cian, E., Massetti, E. and M..Tavoni, 2009. The Incentives to Participate in and the Stability of International Climate Coalitions: a Game Theoretic Approach Using the WITCH Model, OECD Economics Department Working Papers No. 702, June 2009.

Bosetti, V., Massetti, E. and Tavoni, M., 2007. The WITCH Model. Structure, Baseline, Solutions, FondazioneEni Enrico Mattei Working Paper No. 10.2007.

Burniaux, J.-M., Chateau, J., Duval, R. and Jamet, S., 2008. The Economics of Climate Change Mitigation: Policies and Options for the Future, OECD Economics Department Working Papers, No. 658, OECD, Paris.

Bréchet, T., Eyckmans, J., Gérard, F., Marbaix, P., Tulkens, H., van Ypersele, J.P., 2010. The impact of the EUs carbon emissions reduction proposals on the stability of global climate agreements. Climate Policy 10, 148–166.

Bréchet, T., F., Gerard, and H., Tulkens, 2011. Efficiency vs. Stability in climate coalitions: a conceptual and computational appraisal, The Energy Journal 32 (1), 49–76.

Carraro, C., Siniscalco, D., 1993. Strategies for the international protection of the environment. Journal of Public Economics 52, 309–328.

Carraro, C., Siniscalco, D., 1991. Transfers and Commitments in International Negotiations, Paper prepared for the ESF Task Force 3 on Environmental Economics. In: Maler, K.G. (Ed.), International Environmental Problems: An Economic Perspective. Kluwer Academic Publishers, Dordrecht.

Chander, P. and H. Tulkens, 1997. The core of an economy with multilateral environmental externalities. International Journal of Game Theory 26.

Carraro, C. and J. Eyckmans and M. Finus, 2006. Optimal transfers and participation decisions in international environmental agreements. The Review of International Organizations 1.

d'Aspremont, C.A., Gabszewicz, J.J., 1986. On the stability of collusion. In: Matthewson, G.F., Stiglitz, J.E. (Eds.), New Developments in the Analysis of Market Structure. Macmillan, New York.

d'Aspremont,C.A., A. Jaquemin, J. Gabszewicz and J. Weymark, 1983. On the stability of collusive price leadership. Canadian Journal of Economics 16.

Donsimoni, M.P., Economides, N.S., Polemarchakis, H.M., 1986. Stable cartels. International Economic Review 27, 317–327.

Hanemann, W.M., 2008. What is the Cost of Climate Change?, CUDARE Working Paper No. 1027, University of California, Berkeley.

Harrod, R., 1948. Towards a Dynamic Economics: Some Recent Developments of Economic Theory and Their Application to Policy. MacMillan, London.

IEA, 2004. World Energy Outlook 2004. OECD, IEA Paris.

IPCC, 2007. Climate change 2007: impacts, adaptation and vulnerability. In: Parry, M., Canziani, O., Palutikof, J., van der Linden, P., Hanson, C. (Eds.), Contribution of Working Group II to the Fourth Assessment Report on Climate Change. Cambridge University Press, Cambridge.

Jamet, S., Corfee-Morlot, J., 2009. Assessing the Impacts of Climate Change: a Literature Review of the Impacts of Climate Change, OECD Economics Department Working Papers, No. 691, ECD, Paris.

McGinty, M., 2007. International environmental agreements among asymmetric nations. Oxford Economic Papers 59, 45–62.

Nagashima, M., Dellink, R., van Ierland, E., Weikard, H.P., 2009. Stability of international climate coalition – a comparison of transfer schemes. Ecological Economics 68, 1476–1478.

Nordhaus, W., 2007. A review of the stern review on the economics of climate change. Journal of Economic Literature 45 (3).

Nordhaus, W.D., Boyer, J.G., 2000. Warming the World: the Economics of the Greenhouse Effect. MIT Press, Cambridge, Massachusetts.

Ramsey, F., 1928. A mathematical theory of saving. Economic Journal 38 (152).

Solow, R.M., 1974. The economics of resources or the resources of economics. American Economic Review 64 (2).

Stern, N., 2007. The Economics of Climate Change: The Stern Review. Cambridge University Press, Cambridge.

Tol, R.S.J., 2005. The marginal damage costs of carbon dioxide emissions: an assessment of uncertainties. Energy Policy. (33).

---

[12] . For instance, projected world BaU emission growth is somewhat higher in WITCH than in the OECD model ENV-Linkages as featured in Burniaux et al. (2008) (100% versus 85% over the period 2005-50).

UNFCCC, 2007. Investments and Financial Flows to Address Climate Change, Background Paper on Analysis of Existing and Planned Investments and Financial Flows Relevant to the Development of Effective and Appropriate International Response to Climate Change.

Weitzman, M., 2007. A review of the stern review on the economics of climate change. Journal of Economic Literature 45 (3).

Weitzman, M., 2001. Gamma discounting. American Economic Review 91 (1).