MID-TERM REVIEW OF THE GEF RESOURCE ALLOCATION FRAMEWORK

Design of the RAF

Technical Paper #2

6 October 2008

Not edited
Not for citation or distribution

Draft

Contents

See also other Technical Papers:
#1. Methodology and Context
#3. RAF Allocations and Utilization
#4. Implementation of the RAF
#5. Delphi Study of the GEF's Resource Allocation Framework Benefits and Indices

Overview

This paper presents the structure and elements of the design of the RAF, including the content of the constituent indices on benefits and performance and a range of issues associated with the technical quality and practical application of these indices. The analysis of design draws on the Delphi peer expert study, statistical analysis, comparative review of other performance-based allocation (PBA) systems, and expert interviews. Further details on data sources are available in a separate technical paper and the Delphi study report

As in other performance-based allocation arrangements, the GEF RAF is a rules-based system that uses a set of formulas to allocate funds. Virtually all such systems are formula driven, with the formula containing two main components: "potential benefits and needs" and "performance".

The GEF measures *performance* (through the GEF Performance Index, GPI) relative to its mandate in three ways:

- First is measurement of each country's national policies and enabling environment with relevance to the environment;
- Second is the government's performance on environment-related projects, including those funded by GEF grants and, more broadly, by World Bank grants and loans; and
- Third, against a broad measure of policy and institutional performance government-wide.

GEF measures *potential benefits and need* (through the GEF Benefits Index, GBI) according to a country's potential to generate global environmental benefits. The RAF is unique among PBA systems in its direct attention to environmental benefits in assessing "potential and needs". Most other PBA systems use gross national income (GNI) per capita and population, as surrogate measures of "potential and needs". All international donor institutions' PBA systems, take country performance in environmental policy and institutions into account, however, albeit often with a small weight, in their policy and institutional performance index.

A GEF allocation is not an entitlement. It is, in fact, more like a ceiling of available funds to potentially be utilized for projects. A country may receive grants up to a maximum of its allocation during the four years of a particular replenishment period, provided the country submits project proposals through a GEF Agency to the satisfaction of the GEF. Many factors influence the level of the GEF allocation to a particular country. These factors include the number of eligible countries; the total amount of money available for grants in each focal area; and of course, the country's scores on GBI and GPI. In the allocation process, specific rules aim to ensure that extreme allocation results are avoided, though ceilings, pooling and floors.

For each of the three indices constituting the RAF design, this paper addresses (a) the relevance and reliability of indicator data; (b) the related topic of substantive coverage and scope of the indicators; (c) balance and weights among elements as each indicator or index is applied; and (d) data gaps and opportunities for using new data sources.

1 The GEF Benefits Index (GBI)[1]

The GEF Benefits Index measures the scope for producing global environmental benefits in a particular focal area in a given country. It is not designed to measure country intention, capacity or performance.

Given the importance of the indicators in determining allocation, there has been much discussion about the content of the indices. In this case, "content" refers to (a) the variables included in an index; (b) the structural (mathematical) relationships among the variables, including weights; and (c) the quality of data used for applying value to the indicators and arriving at country scores. However, the RAF design cannot respond to all issues of concern to stakeholders regarding GEF support. Some issues are better addressed in implementation, or in a few cases, both.

1.1 The GEF Benefits Index for Biodiversity (GBI-BIO)

The purpose of the GBI for biodiversity is simply "to measure the potential global benefits that can be realized from biodiversity related activities in a country"[2]. It consists of two major parts: terrestrial biodiversity and marine biodiversity (see **Figure** 1), built from indicators with data (from right to left), applying weights in the formula to arrive at a country score.

Figure 1: Components of GBIBIO

GEF Benefits Index For Biodiversity (GBIBIO)

Score	Formula	Indicators	Data
Terrestrial Score WT=0.8	Country-Ecoregion Components (CEC) Biodiversity Score = WT x + WT2 x + WT3 x + WT4 x Where WT1+WT2+WT3+WT4 = 1 (Of total 1,326 CECs)	Represented Species WT1=0.55	Averaging scores for six taxonomic groups (a. mammals, b. birds, c. amphibians, d. reptiles, e. vascular plants, f. freshwater fish)
		Threatened Species WT2=0.20	Species receive credits based on their threat class (for a. mammals, b. birds and c. amphibians)
		Represented Ecoregion WT3=0.15	Each ecoregion has a credit of 1 globally. Then distributed across CECs of ecoregion in proportion to remaining habitat (=uncleared land)
		Threatened Ecoregion WT4=0.10	Threat credits per critical/ endangered, vulnerable, stable; then distributed across CECs in proportion to remaining habitat
Marine Sco WM=0.2	Sum of credits from all marine species in territorial waters (=EEZ) of the country	Then distributed across countries in proportion the estimated habitat for the species in each country	Each evaluated species gets a total credit of 1 globally

First, the **terrestrial score** is weighed 80 percent in the index. For each country, it has four indicators, namely represented species, threatened species, represented ecoregions, and threatened ecoregions. The score is computed from:

- Terrestrial ecoregions within a country. Since ecoregions cross country borders, a country receives the relative share of that region that falls within the country, called Country-Ecoregion-Components (CECs). For instance, if 60 percent of the habitat for a species lie in a particular CEC and the remaining 40 percent are distributed evenly across two other CECs, the three CECs receive credits of 0.6, 0.2, and 0.2 for that species. Each CEC is scored both in terms of representation in the country, and the degree of threat to the ecoregion.
- Species found in the ecoregion in the country. The index measures six types of species, or taxonomic groups, namely mammals, birds, amphibians, reptiles, vascular plants, and freshwater fish. It scores both the representation of the species and how threatened they are[3]. Endemic species, only present in one country, would receive 100% of the score.

Second, the **marine score** is based solely on represented fish species. Each evaluated species receives a total credit of one globally, which is distributed across countries in proportion to the estimated habitat for the species in each country. The marine score for a country is the sum of the credits from all of the marine species located in the territorial waters of the country[4]. The marine score is weighted 20 percent in the GBI for biodiversity.

Issue 1: Reliability and relevance of data

The quality and comprehensiveness of the data are generally satisfactory. Participants in the Delphi expert study expressed strong support for the view that the GBI-BIO data are the most comprehensive and reliable available for the items covered (median score 8 on scale 1-10 with 10 being "to a great extent"). Interviewees also felt that the index is the best possible given currently available data, and that other options considered were not adequate for the index. Biodiversity experts mentioned the use of the IUCN red list of endangered species and biodiversity hotspots as supported by credible recognition within the biodiversity community.

The Delphi panel on biodiversity noted that the three objectives of the Convention on Biological Diversity (CBD) are: conservation of biodiversity; sustainable use of its components; and fair and equitable sharing of benefits arising from the use of genetic resources. The RAF biodiversity index mainly focuses on the first of these - conservation. Delphi experts felt that the emphasis on conservation was about right (average 5.6 on a scale 1-10 in which 5 indicated about right). They also indicated, however, that the index's emphasis was insufficient regarding sustainable use of biological resources and transfer of genetic resources across borders (average rating of 4.0 and 3.9 with 1 being "much too low").

As can be seen from **figure** 1, the GBI contains information on both **ecoregions and species**, with higher emphasis (weight values) on the latter. The index is based on definitions provided by the CBD, which characterizes species diversity with comparable data for many countries. The rationale for the weighting was that ecosystems are a much broader concept, but should be included since species richness (the number of species in a given area) does not capture all aspects of biodiversity in an area. Some country stakeholders were uncomfortable with the emphasis on species richness, as it does not capture all aspects of biodiversity (for example, desert areas) and that data on other aspects are often fragmented. Others want the size of area and habitat complexity to be considered.

Most estimates of the total number of **species** on Earth lie between 5 million and 30 million. Of this total, roughly two million species have been formally described; the others are unknown or unnamed[5]. The RAF index will inevitably reflect some weaknesses in underlying data, such as:

- Data on marine and freshwater species and habitats are more incomplete than for other species and habitats, in general, and may in addition reflect under-representation of mammals, birds, reptiles, amphibians, and fish living in rivers, lakes, or wetland ecosystems.
- Information on arid ecosystems also tends to be less explored. The GEF defines arid and semi-arid lands as "the tropical grassland and savannah/woodland savannah, the warm desert and semi-desert, temperate grasslands, tundra communities, and cold desert biomes. These lands cover over one third of the earth's land surface and are home to over 900 million people"[6].
- Classification of taxa (classifications of plants and animals) depends on national efforts and capacities. Many developing countries do not have a complete mapping of their species. While the RAF index has data for all countries, the underlying data completeness may vary, and some species may be at times underestimated. Likely data gaps were mentioned by source experts for countries such as Papua New Guinea, Angola, and DR Congo.
- Species migrating across borders and ecoregions, especially marine species, birds, and some mammals, pose a challenge to any country-based index. The index cannot monitor migration on a continuous basis, as any index is periodic in that it provides overviews of status at given points in time.

The biodiversity index consists of many different indicators. Statistical analysis found a strong relationship between the separate indicators and both the final index score and the allocations for

countries. Specifically, the statistical correlation between each indicator and the GBI score and the allocation, is relatively high (above 0.9, on a scale ranging from 0-1) for most components (see statistical annex). These high levels of association suggest convergence of measurement among the elements of GBIBIO. It is lowest for amphibian threat (correlation of 0.59 with GBI), and highest for reptile representation, mammal representation and ecoregion representation and threat. The correlation between the marine score and the GBI is 0.79. No indicator dominates the index more than others.

The indices use **data sources** from international organizations specialized in various aspects of biodiversity issues. The respective sources are considered as authoritative in their respective fields within the scientific community - Birdlife for birds, Fishbase for fish, CI and IUCN for species, and WWF for ecoregions. Some country stakeholders raised the concern that wherever possible, national data or data from regional organizations should have been factored into the indexes. The international organizations in fact mainly derive raw data from the national level, from governmental sources or technical experts.

Issue 2: Coverage in the index

The GBI-bio covers six taxonomic groups of species: mammals, birds, amphibians, reptiles, vascular plants, freshwater fish; and marine fish, as well as ecosystems. Other areas are not covered, including habitats (immediate surroundings of species within an ecosystem), and ecosystem services, cultural significance of biodiversity, and sustainable livelihoods and use. While these are among many factors leading to the threatened or protected status of species and ecosystems, they are also intrinsically difficult to measure, or concern effort (ecosystem services) rather than the effect of those efforts. Some Delphi experts objected to measuring biodiversity by counting species and expressed a desire for an ecosystem-based approach. Delphi participants support the view that that data used in the GBI_{BIO} should be expanded to incorporate a broader range of taxa. Specific areas of coverage addressed by the Delphi study include:

- There was no agreement to amend the GBI_{BIO} to give greater weight to **biosafety**. The average response was 4.4, but experts were divided in their opinions. Most answered 1 (not at all) on a scale from 1 to 10 (on scale 1-10, standard deviation 3.2). The comments of those agreeing cited the growing importance of the spread of diseases and genetically modified organisms in the context of biodiversity preservation. The comments of those indicating disagreement indicated a belief that biosafety is separate from biodiversity and that it should be dealt with at the national rather than international level. Biosafety, as currently understood, has strong linkages to sovereign rights, investment policy, health policy, and patents and trade. It was pointed out that it was difficult to envision how one would measure this issue, and that if biotechnology is a threat to biodiversity, the index which is based on threats and richness would reflect this. As per the meeting of parties (COP-MOP) Decision BS-III/14, the deadline for the first national biosafety status reports was September 2007; by May 2008 the reporting rate was around 50 percent. National data is not yet broadly available. Nevertheless, the GEF is the financial mechanism for the Cartagena Protocol, which deals with biosafety, therefore, there is an obligation from the GEF to support initiatives in this area.

- There was a mixture of views on whether the GBI_{BIO} should be amended to give greater weight to **agrobiodiversity**. Answers ranged from 1 (not at all) to 10 (to the maximum extent) with an average of 6.1 (standard deviation 2.9). 32% indicated strong disagreement. This shows that the scientific community has not yet reached consensus on this issue. Agricultural systems interact with, and are parts of, ecosystems, so the presence or absence of agrobiodiversity is part of the larger biodiversity picture. Some experts cited positive high potential benefits of including agrobiodiversity, while others cited disadvantages such as the similarities among agro-ecosystems worldwide, distortions in data, and disadvantaging countries with high biodiversity but little agriculture. There is no clear solution to rewarding, in the indices, both agricultural land and at the same time non-cleared land. The COP 9 Decision IX/1 (May 2008) agreed that the programme of work on agricultural biodiversity continues

to provide a relevant framework to achieve the objectives of the Convention, and also invited "Parties and other Governments and relevant organizations to finance and undertake research as appropriate to further develop and apply methods and techniques for assessing and monitoring the status and trends of agricultural biodiversity and other components of biodiversity in agricultural ecosystems".

- **Invertebrates** are animals lacking a vertebral column, which includes a majority of species (and does not include fish, reptiles, amphibians, birds, and mammals). There is strong expert agreement about the desirability of including marine invertebrates - sponges, jellyfish, corals, mollusks, crabs, shrimp, and lobsters - with an average of 8.5 on the Delphi score of 1-10 (standard deviation 2.1), but moderate support for expansion to terrestrial invertebrates (average 6, standard deviation 3.2). It was cautioned that richness and diversity of invertebrate species are insufficiently documented and that knowledge about invertebrate biodiversity is so patchy that including them would introduce biases.

- The index also does not cover **carbon sequestration** and other ecosystem system values**,** namely the process of increasing carbon pools other than the atmosphere[7] and sinks of ecosystems. Carbon pools can be forest biomass, wood products and soils. The GEF which has a mandate for desertification, biodiversity loss, and climate change, for which issues and policies are strongly interlinked, rendering measurement difficult. While there is usually a positive correlation between high carbon value and high biodiversity, this aspect is also linked to climate change emissions and would have implications for the climate change index and allocation. Even if good data for land use, land use change and forestry and different types of sinks would be readily available, improvements in one area may offset the other. In short, were sinks included, a country with a high forest sequestration would score higher in biodiversity, while its net GHG emissions under the RAF's current GBI for climate change would decrease and it would score lower in climate change.

Issue 3: Balance in the Index of Terrestrial versus Marine Biodiversity

The index includes a number of weights, related to diversity threats, representation, and to ecosystems and species. These do not seem to generate as much debate as other issues, though some stakeholders noted that weighting threats highly appears to reward deterioration of biodiversity.

The main concern on weights relates to the fact that RAF gives a country's **terrestrial** biodiversity a weight of 0.8 and its **marine** biodiversity score a weight of 0.2. The initial design decision on these weights was mainly political, with the aim to mirror historical patterns of allocation, and was partly based on the uncertainties related to marine data. Limitations to existing data disallow worldwide scoring of threatened marine species, so the marine biodiversity score is based solely on represented fish species. While the terrestrial-marine ratio is not based on science, there is also no scientific model that indicates the right balance, as data for the two ecosystems are not comparable.

Delphi experts indicated moderate support for the view that marine biodiversity should be given more weight. Half of the 22 participants responding to this question gave numerical responses of 7 or more, indicating that the weight of terrestrial biodiversity was too high. The average of the numerical responses was 6.3 (on a scale of 1-10, with a standard deviation of 2.8).

There was strong agreement that *threatened/endangered status* of species and ecosystems should be treated the same way for marine biodiversity as terrestrial biodiversity; all but one of the Delphi participants believed they should. There was less agreement about the feasibility of doing so, although participants were somewhat more positive about the feasibility of including threatened/endangered status of ecosystems than of individual species.

Decisions on marine-terrestrial weights mostly affect small island developing states (36, of which 34 are in the group allocation), countries with a large Exclusive Economic Zone (EEZ); *and* on the other side of the spectrum, landlocked countries (35, of which 26 are in the group allocation) with no marine data would be scored on their terrestrial component only.

The mid-term review conducted simulation of the effects of changes in the marine weight (for comparison with the current weighting of 0.2 and terrestrial weight 0.8), by increasing the weight for the marine indicator in increment steps (up to 1). Of the 36 small island developing states, seven now have individual biodiversity allocations. If the marine weight were increased to 0.4 for the initial allocation, there would hardly have been any change to the initial ranking or allocation (one SIDS, Suriname would have moved up from the group allocation, which since happened in the midpoint reallocation). The simulation demonstrated that a 50-50 weight with the current data would bring five SIDS up to individual allocations, while seven countries (including one SIDS and four landlocked countries) would move from individual to group allocations. For SIDS currently receiving individual allocations, amounts would increase for five and decrease for two countries. Other special categories (least developed countries, LDCs; landlocked countries; fragile states) would lose funds. Because countries currently with high individual allocations also have large marine resources, their GBI also increases when modifying the weights. It is not clear if new data on biodiversity ocean resources would change this pattern significantly. The individual allocation countries currently have 85% of the accumulated marine score (and 89% of the accumulated terrestrial score).

Issue 4: Channeling resources for biodiversity to global environmental benefits

The RAF model does channel resources to countries with high global biodiversity environmental benefits, **though not in exact proportion to GBI scores.** Taken together, the 57 indicative countries (i.e. with individual allocation) currently account for the bulk (88%) of GBI-BIO scores of the 150 eligible countries. These countries accumulate 75.3% of the total resources of one billion US$ in the focal area (or 83.69% of 900 M US$, after the set-side is applied). See table 1.

Table 1: Shares of GBI-BIO scores and funding in biodiversity

Country type in biodiversity	# of countries	Share of biodiversity Allocation	max alloc (m$)	min alloc (M$)	max GBI	min GBI	Share of total GBI	Share of Marine score	Share of Terrestrial score
Indicative	57	75%	63.2	3.5256	663.7	19.9	**88%**	**85%**	**89%**
Group allocation	93	15%	3.5255	1	70.5	0.1	12%	15%	11%

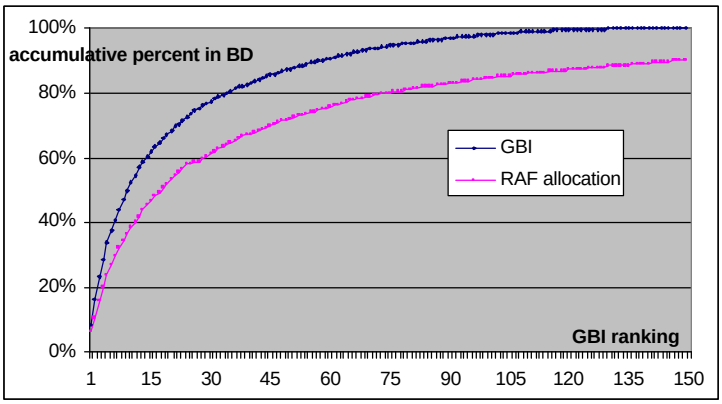


Figure 2: Accumulated percents of GBI and allocations

The share of GBI-BIO scores held by indicative allocation countries is somewhat larger than their share of the total biodiversity allocation after the set-aside; this provides supportive evidence that the RAF model channels resources to countries with high global environmental benefits scores. **Figure** 2 shows this graphically, by plotting countries by accumulated resource allocation versus accumulated GBI. The 20 countries with the highest GBI scores

accumulate 67% of total GBI and 52% of the GEF-4 funds in biodiversity[8]. They also represent 69% of total marine score.

There was general agreement among Delphi biodiversity experts that the list of countries qualifying for individual funding was somewhat biased, in methodology as well as in the list of countries, toward conservation, at the expense of sustainable use of biological resources and benefits from transferring genetic resources across borders (median 7 on scale 1-10, standard deviation 2.3)

There was no Delphi consensus as to whether the GBI_{BIO} index should be amended to give greater weight to biodiversity **megadiversity** countries or countries with biodiversity "hot spots" (median 3 on scale 1-10, high standard deviation 3.0).[9] Some participants felt that such countries deserve greater attention; others expressed the view that a focus on such countries has led to neglect of important ecosystems. Of the seventeen countries rich in biological diversity and associated traditional knowledge called Like Minded Megadiverse Countries (LMMC), all have individual allocations for biodiversity. Their cumulative share of the biodiversity index score is 59 %, and they receive $459 million of the focal area funds, or 46% of the total. A recent report to the CBD noted that this amount is an increase of 76% from GEF-3[10].

The picture is more mixed for the biodiversity **hotspots**. Some hotspots fall within or across countries with individual allocations, such as Cerrado in Brazil; the Chilean Winter Rainfall-Valdivian Forests; the Tropical Andes; the Philippine archipelago and the Cape Floristic Region in South Africa. Others cross borders between counties with individual and group allocations, often with the group allocation country being smaller in size. Examples include the Atlantic Forest in Brazil and in Uruguay, a group country; and Caucasus and the Mountains of Central Asia (with Armenia, Georgia, Kyrgyzstan and Tajikistan in the group allocation). The Guinean Forests of West Africa cover Nigeria and Côte d'Ivoire (individual allocation) and Guinea, Ghana and Liberia (group allocation). Country size is, however, not consistently a determining factor; Madagascar and the Indian Ocean Islands (Seychelles and Mauritius) all have individual allocations. Some hotspots fall squarely within group allocation countries, including Polynesia-Micronesia and the Himalayas in Nepal. Of course, allocation to a country with biodiversity hotspots does not imply that GEF funds will be used for the hotspots as such.

Issue 5: New data developments

The *RAF document* indicates that the GBI will be continuously assessed over time to include updated information with assistance from the CBD. While developments are continually ongoing, it is less clear that these efforts have led to identification of a new body of more robust data. Some Delphi experts supported the view that there exist comprehensive, reliable data that would give a more accurate depiction of potential for biodiversity conservation, promotion of sustainable use, and/or transfer of genetic resources across borders (average 5.6 on scale 1-10), and suggested use of agricultural biodiversity data from the Biodiversity International institute in Rome (one of the Centers of the Consultative Group on International Agricultural Research) and trade data from the Convention on International Trade in Endangered Species.

Delphi biodiversity experts were uncertain about the availability of data for assessing threatened/endangered status of **marine species** (4.8 on a scale from 1-10, standard deviation 2.2), though some suggested exploring data on areal extent of coral reefs, mangrove areas, and seagrass beds. A study on the RAF marine biodiversity indicators (Fedder 2007[11]) identified potential new data sources on marine biodiversity that could be relevant to the RAF and warrant monitoring for inclusion in the future:

- **Represented species:** *Ocean Biogeographic Information System* (OBIS) and *KGSMapper*[12], a globally available portal of 77,000 species, 13 million records, and over 80 contributing partners. Other sources do not yet have full availability of data, including the *Sea Around Us Project*

*(*SAUP)with a global distribution of marine fish, mammal, and invertebrates including predictions for ranges of 700 fish species, 100 crustacean species, 100 mollusc species, 100 mammal species, and more; and on various sources with global information systems (GIS) for richness data and national lists of marine species

- **Threatened Species:** The Global Marine Species Assessment (GMSA)[13] will provide assessment of marine species ranges and threat status of 20,000 marine species, mainly fish, but also invertebrates, by IUCN in 2013.
- **Represented Ecoregions:** The Marine Ecoregions of the World (MEOW[14]), a cooperation between WWF and TNC, aim to provide a complete classification of all coasts of the world for 232 marine ecoregions, analogous to the terrestrial ecoregions of WWF.
- **Threatened Ecoregions:** The threat and impact mapping by the National Center for Ecological Analysis and Synthesis (NCEAS[15]) identified a range of 38 threats to the marine environment – and mapped the global impact humans are having on the ocean's ecosystems. An Environmental Vulnerability Index (EVI[16]) by the South Pacific Applied Geoscience Commission (SOPAC) assesses a country's vulnerability using a set of 50 indicators including six on the marine environment. The SAUP is also developing various fisheries related data.

There was little support for the view that there are reliable, comprehensive data that would allow incorporation of the cost of biodiversity conservation (average 4.5 with a standard deviation of 2.3). Comments indicated a general agreement about the lack of a reliable global database of conservation costs, but noted that the methods for evaluating costs and benefits are well established and could be incorporated into the project approval process. Other comments cited recent meta-analyses and papers as sources of cost and cost effectiveness information[17].

1.2 GEF Benefits Index for Climate Change (GBICC)

The purpose of the GBI for climate change is "to measure the potential global benefits that can be realized from climate change mitigation activities in a country". It consists of two major parts (see **Figure** 3):

	Indicator	Definition	Scope	Source
Figure 3: Components of GBI-CC	**Baseline GHG emissions**	o For the year 2000 o In tons of carbon equivalent	Only for: carbon emissions from: o fossil fuel combustion o cement o the emission of other GHG gases	Climate Analysis Indicators Tool (CAIT) of the World Resources Institute
Baseline x Adjustment Factor	**Carbon Intensity Adjustment Factor**	Tons of carbon equivalent emitted by a country per unit of economic activity (GDP)	Computed as the ratio of: o the carbon intensity in 1990 divided by o the carbon intensity in 2000	Climate Analysis Indicators Tool (CAIT) of the World Resources Institute

First, the index measures the country's baseline **greenhouse gas (GHG) emissions** in tons of carbon equivalent, in the year 2000. Only carbon emissions from fossil fuel combustion and cement and the emission of other GHG gases are included in the baseline GHG emissions. Emissions associated with land use changes are not included.

Second, the **carbon intensity** adjustment factor computed as the ratio of the carbon intensity in 1990 to the carbon intensity in 2000[18]. The carbon intensity of a country measures the tons of carbon equivalent emitted by a country per unit of economic activity (GDP). It changes over time because of (a) increased carbon efficiency brought about by changes in fuels or technology or economic growth; and (b) structural shifts in the economy away from carbon intensive activities. Carbon intensity was included in the index

because reducing emissions is assumed to be less costly in countries where there is already a track record to reduce carbon intensity and it rewards these countries for reductions.

Issue 1: Reliability and relevance of data

First, the main component in the index is **annual GHG emissions.** In the approved *RAF document*, the reasons for using GHG as the key indicator were that countries with larger emissions have lower abatement costs, and that projects are likely to have greater demonstration and learning effects in high emitting countries than in countries with smaller levels of emissions. Delphi climate change experts found annual GHG emissions moderately useful as a broad indicator of country mitigation potential (average 5.3 on a scale 1-10, with 5 being "moderately useful", standard deviation 2.0).

Concerns raised in the Delphi study include:

- Even poor countries with a large population and low energy consumption per capita may have large total annual emissions, but yet not be able to mitigate much. Mitigation potential of a country is not only based on GHG emissions but also on the capacity of the country to implement mitigation measures; this capacity is not captured in the GBI-CC index, although measures of more general institutional capacity are included in the GEF Performance Index.
- Annual GHG emissions do not take into account those countries that have low emissions and high forest cover, or if countries have seasonal high emissions.
- Emissions are subject to volatility of external shocks not related to mitigation. Russia, for example, reduced emissions not because of mitigation actions but because of economic collapse in the 1990s.
- For certain major emitters, such as China and India, emissions profiles are changing fairly rapidly. Rather than looking solely at historic emissions, mitigation investments *now* will have considerable impact on eliminating future emissions. The index does not measure growth in emissions since 2000.

As an alternative, **emissions per capita** may give a better view of mitigation potential. Other factors being equal, countries with higher carbon intensity and higher per capita emissions have greater mitigation potential than countries with lower carbon intensity or lower per capita emissions. Delphi participants tended to find the use of *per capita* GHG emissions a useful indicator of countries' mitigation potential (the average response was 6.6 on scale from 1-10, standard deviation 2.5, a somewhat higher mean than for *annual* emissions at 5.3 average).

Per capita emissions make it possible to capture the wide disparity in the levels of emissions which may not be demonstrated from the total of national emissions. It is helpful for estimating mitigation potential, but without accounting for total population, and thus total emissions, it can be difficult to link to needs for support. The benefit of using such as indicator is that it easily identifies countries that are very inefficient energy users. Using per capita emission would be appropriate from a equity perspective, and can be useful for comparison.

However, in contrast, it does not give an estimate of how important that country's emissions are globally. The obvious example is that as of 2007 the USA and China have almost equal total GHG emissions, but the USA emits between 4 and 5 times as much per capita as does China. Emissions also depend on the sectors of economy that contribute most to its GDP. If a country is a service economy its emissions will be lower than if it were a manufacturing economy.

Another alternative may be to use **marginal abatement cost** curves directly if these can be estimated for the different countries. Mitigation potential may be low even if annual GHG emission is large in the case of a country which has already made substantial progress in reducing carbon intensity so that further reduction would be more difficult. Emissions levels are a necessary but not sufficient condition for

measuring mitigation. The actual mitigation potential depends on mitigation costs of the mitigation options available.

Second, in the RAF index, countries' change in **energy intensity** is used to complement the GHG emissions. Climate change experts in the Delphi study found use of energy intensity useful as an indicator of a country's mitigation potential (average response 6.3 with standard deviation of 2.3). Energy intensity was found to be a good indicator due to the significant potential in reducing emissions through improvements in energy efficiency. However, concerns were raised include that this indicator does not identify mitigation opportunities; embeds too many factors to provide clear measurement; and does not reflect the different sources of emissions from different sectors.

Third, the energy intensity adjustment factor takes account of a country's economic growth. The index is constructed with GNI per capita in such a way that a country can increase its GHG emissions at the same pace as the GNI growth without changing its GBI-CC score. Climate change Delphi experts found the GNI per capita useful (average 5.8, standard deviation 2.5), as it provides a measure of economic status; the higher the per capita GDP, the better its capacity to reduce emissions. However, it does not address good mitigation policies in countries with low GNI per capita.

However, the energy intensity adjustment factor plays a small role in the index. The GBI-CC *multiplies* the GHG emissions with energy intensity change (rather than adding the two indicators with relative weights, which is the approach for GBI-BIO and GPI). This means that the GHG emissions dominate the scores, as the energy intensity will only adjust the emissions in the index up or down.

Table 2 indicates the share of allocations per quintile for both GHG baseline and the adjustment factor (the ratio of carbon intensity in 1990 to carbon intensity in 2000). A total of 75% of climate change allocations go to the top 20% emitters, while the energy intensity adjustment factor is more evenly distributed, where quintile 1 includes the top 20% of countries that have improved their energy intensity. Countries with the same total emissions but very different carbon intensities (and thus, implicitly, efficiency) are treated identically. If emissions and energy intensity were added together as separate indictors, this would have produced very different scores with more emphasis on structural economic change, but likely with more volatility as the energy intensity can vary from year to year.

Table 2: GHG Vs. Adjustment Factor

Baseline quintile	Allocation share	Factor quintile	Allocation share
1	75.0%	1	31.5%
2	11.2%	2	27.6%
3	6.0%	3	22.7%
4	4.2%	4	9.9%
5	3.7%	5	8.3%
total	100.0%	total	100.0%

Four, the RAF climate change index uses a **baseline year and a yearly range** for both the GHG emissions and the energy intensity; the year 2000 was used as baseline for the initial allocation. The UNFCCC indicated 1990 is as the base year (see Art. 4.2b). The world in 1990 was fundamentally different (e.g., Soviet Union on the collapse) than it was in 2000 and 2010 will again be very different (e.g. rise of China, India, Brazil), as have the emission profiles of many countries. The geographical distribution of emissions has significantly changed since the year 2000. Delphi experts therefore found that the choice of the baseline years, the 10 year lag, and use of one base year, to be arbitrary. Since annual variability in emissions can sometimes be substantial, one year (2000) at random may not be a good metric.

Experts disagree on the best year for measuring GHG emissions for the index, between (a) the last available year (34% of Delphi participants); (b) year 2000 (29%), and (c) ratio of years 1990 to 2000 (35%). The issue is how to balance current abatement potential of countries versus historical potential of countries. The MTR conducted simulations with three different scenarios with available data (baseline in 1990, 1995, or 2000)[19]. The simulation shows no significant allocation changes across GBI quintiles.

Issue 2: Coverage of data in climate change

In the RAF climate change index, not all GHGs are covered; namely those from **land use change** and from industrial non-CO2 emissions are not included. For many LDCs, non-energy GHG emissions dominate, so that the exclusion of agriculture and land use, land-use change and forestry (LULUCF) emissions may distort the index. Distortion is especially the case for forest-rich countries such as Indonesia and Brazil. While inclusion of emissions from land use should be a component of the index, recent and robust data are not available.

Experts found the representation of **sources of GHGs** (e.g., fossil fuel use and cement production) adequate (69%) to some degree (average 5.2, standard deviation 2.5). Given that the index is supposed to be correlated with mitigation opportunities, the Delphi panel suggested that the index should consider the following alternative sources of GHGs in decreasing order of importance: (a) agriculture and land use; (b) deforestation and forest degradation; (c) gas flaring; (d) industrial non-CO2. However, actual emissions from deforestation and agriculture reflect on a country's policies and interest in dealing with emissions and climate change, but are hard to determine accurately.

About half (51%) of Delphi participants found that the **representation of gases** in the indices is adequate (average 6.0, standard deviation 2.5). Participants were of the view that all types of greenhouse gases and from all sources should be accounted for. Gases other than carbon dioxide need to be weighted by their global warming potentials. In reality, the current data sources used for the climate change index does not provide adequate data on non-carbon sources of GHG or gases.

On the other hand, most (60%) of the Delphi participants found efforts in the clean development mechanism (CDM) **and carbon trading** for the climate change index were not very relevant (average 6.0, standard deviation 2.45). CDM and carbon trading affect the emission reduction targets under the Kyoto Protocol for Annex I countries. National GHG emissions in those countries do not change and thus the index is stable. Carbon trading does offer incentives for reducing emissions, but requires management at a micro level and is best addressed outside an index at national level.

Issue 3: Balance of mitigation and adaptation in the climate change index

The most widely discussed concern for the climate change index is the balance of mitigation and adaptation, which is not reflected by indicators in the formula. **There is high agreement among experts and stakeholders that more should be done to balance funding between adaptation and mitigation in developing countries**. A majority of Delphi experts agreed that more should be done to balance funding (average response 8.5, standard deviation 2.3), because "mitigation produces global benefits, while adaptation produces local benefits".

Experts agree, as do many stakeholders, on channeling most of funding for adaptation to the most vulnerable countries, paralleling the fact that most of the funds under the CBIcc go to mitigation in countries with more emissions. Funding should be country-specific focusing on the country's status of vulnerability when deciding on adaptation funding. However, under the UNFCCC, there is no distinction among developing countries as far as support for adaptation costs from developed countries under Article 4.4 is concerned. From that standpoint, all developing countries, to one extent or another, are particularly vulnerable.

Therefore, there are mixed views on how best to address this, through using vulnerability broadly as a principle to guide funding under RAF; constructing a vulnerability index; or both, if possible. There were also mixed views as to the right grouping or definition of categories of vulnerable countries. In principle, 89% of Delphi respondents agree to some extent to using set categories such as LDCs, SIDS and African countries, but also noted that not all such countries are vulnerable in the same way nor to the same extent. As one panel member stated, "The UNFCCC provisions with respect to vulnerable countries, including those indicated in Art. 4.8, 4.9 and 4.10, should be the foundation for any categorization".

Ultimately, if no scientific evidence exists to determine the right balance, the balance is determined by strategic or political concerns. Donor expectations of the GEF have overall focused on mitigation. In fact core GEF Trust Fund resources have not been used for dedicated adaptation projects until recently, through the Strategic Pilot for Adaptation. Historically, the GEF funding in climate change has focused on the bigger GHG emitters, based among other things on the assumption that projects could catalyze market change and broader impact in such countries. The RAF formula accepts and magnifies this focus.

Issue 4: Channeling resources for climate change to global environmental benefits

The RAF model does channel resources to countries with high global environmental benefits, though not in exact proportion to GBI scores. As can be seen in table 3, the GBI share increase outpaces the resource allocation share increase. The 46 indicative countries receive 75% of the total focal area resources, but cover 89% of the GBI environmental scores of the 161 eligible countries. The indicative countries also cover the most of the underlying indicators (baseline, CO and non-co sources) of the eligible countries in the focal area.

Table 3: Shares of GBI scores and funding in climate change

Allocation climate change type	# of countries	Share of Allocation	max alloc (m$)	min alloc (m$)	max GBI	min GBI	Share of GBI	baseline	co	nonco
indicative	46	75%	150	3.09	2413020.9	8402.8	**89%**	**86%**	**87%**	**85%**
group	115	15%	2.97	1	145958.4	1.0	11%	14%	13%	15%

Forty percent of the Delphi respondents indicated that the GBI-CC makes moderately good sense by giving larger funds to larger emitters and rewarding countries that reduce carbon intensity (average 4.9, standard deviation 2.4).

A few (15%) of the respondents found the GBIcc to make no sense, however, since the larger countries, and those that have already reduced their carbon emissions significantly, have enough experience and capacity regarding GHG mitigation and have generated enough momentum to carry on even with *less* GEF support. Also it is likely that cost-effective GHG mitigation options have already been implemented in these countries. The larger benefits to larger emitters and more rapid economic growth for 1990-2000 does not reward countries that reduce carbon intensity, energy efficient economies and low-carbon emitting countries such as SIDS and LDCs.

Countries with high emissions intensity and low growth, such as countries in transition, are likely to have a higher mitigation potential. Some stakeholders have suggested that the index should concentrate resources in countries in the middle of the range of GHG emissions, not the highest emitters, since the latter most likely can afford projects and pay from their own resources for national consultations. Similarly, the UNFCCC Secretariat pointed out to the mid-term review that the index should concentrate in countries in the *middle* range of the GHG emissions.

It was stated in the Delphi study that neither the over-all size of **emissions** of a country, nor its economic growth, are reliable proxies to getting the most emission reductions for the money spent. Drawbacks of the formula include the focus on energy and industry, when a large share of emissions in developing countries comes from other sources; and use of data that are poor measures of economic output in countries with large informal economic sectors.

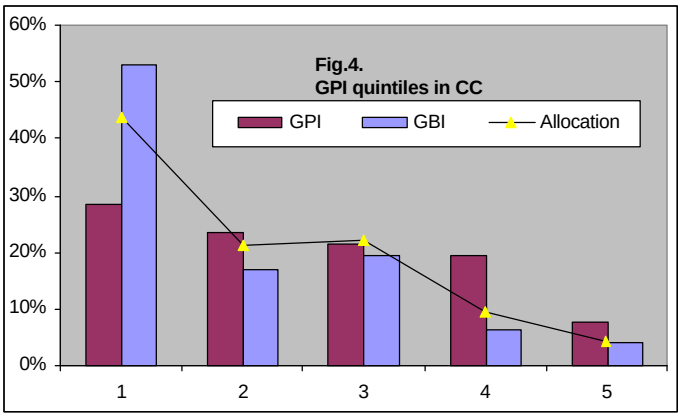


Figure 4 shows the shares of the resource allocation, GBI and GPI in each GPI quintile: The first two quintiles are allocated around 60% of the total resources and accumulate about 50% of the total GPI.

About two-thirds of the experts in the Delphi Panel (62%) thought that there was too great a **concentration of funding** for climate allocated to too few countries (average 7.1, with 10 being "far too much", standard deviation 1.9). But it is not obvious which countries should receive *more*, or *less*[20].

There was some degree of opinion among Delphi climate change participants that a more balanced distribution of GEF climate change funding would result in substantially greater GHG emissions mitigation (average 6.5, standard deviation 2.1, with 5 being "about the same", and 10 being "substantially more"). Smaller countries may have less capacity to invest their own human and financial resources, and the small amounts allocated may be below a threshold to carry out meaningful projects. Delphi experts felt that a more balanced distribution would allow some small countries enough to make progress and would encourage action on mitigation in a wider number of countries, and possibly enhance joint benefits with mitigation and adaptation and involvement of many small countries and south-south cooperation.

Issue 5: Emerging new data and data gaps

The data for the climate change index is provided by the WRI's Climate Analysis Indicators Tool (CAIT). Delphi experts agreed with the use of to a limited extent (average 4.6, standard deviation 2.0) over the inventories from national communications. Experts also thought the correlation between the WRI and national communications data was not strong (average 4.5, standard deviation 1.5). However, no consensus emerged on which data to use. Views were mixed on the feasibility of compiling a database with only information from national communications of Annex I countries (average 6.9, standard deviation 2.6). While it was suggested that national communication data could be used to verify the accuracy of global datasets, such verification risks adding subjectivity in judgment when the data differ. Whatever data is used, caution was advised on independence, verification and completeness. Future data should especially be sought on expanding country coverage for other gases; linkages between climate change mitigation and sustainable development; Land use, land-use change and forestry (LULUCF) and Agriculture, Forestry and Other Land Uses (AFOLU).

Most developing countries are still in the process of preparing their GHG inventories based on 1997 IPCC guidelines. Coverage of GHG emissions in **national communications** to the UNFCCC is still too limited to cover all of the countries eligible for GEF support in a consistent manner. Issues include lack of capacity in countries to fully reporting GHG emissions, especially from LULUCF. The data for land use change of 1950-2000 and data for deforestation of years 1990, 1995 and 2000 from the United States Environmental Protection Agency (EPA) could be explored. Most Delphi participants (63%) did not

know of any further **guidance from UNFCCC**/IPCC that should be reflected in the climate change indices (average 6.13 with a standard deviation of 2.4).

Whereas stakeholders and experts indicated areas where improvements are needed to capture the full range of issues related to climate change mitigation and **adaptation**, it is less certain if data would be available for inclusion in an index. Forty seven percent of Delphi respondents did not know of any measures relating to human **vulnerability or social impacts** (e.g., poverty alleviation, employment, etc.) that should be reflected in the climate change indices, or in another index or form of measurement. Future data may emerge from the Nairobi work programme (2005-2010) of the UNFCCC, which aims to "assist all Parties to improve their understanding and assessment of impacts, vulnerability and adaptation to climate change; and to make informed decisions on practical adaptation actions and measures to respond to climate change on a sound scientific, technical and socio-economic basis, taking into account current and future climate change and variability"[21].

The UN Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report found that "While adaptation has a cost, it significantly reduces the residual costs of climate change. However, there is currently very little quantified information on these costs, and further work is urgently needed to build the evidence base [..]"[22]. Other aspects requiring attention are the type of adaptation (autonomous or planned); the level and timing of adaptation (anticipatory or reactive); the types of costs of adaptation (including direct costs and transition costs); the ancillary benefits of adaptation; and the distributional aspects of adaptation. The UNFCCC study attempted to estimate costs of adaptation in a number of sectors, based on (a) costs of specific adaptations in a few sectors ("simple bottom assumptions" for water resources, coastal resources, human health); and percentage of current investment (top-down approaches).

According to the current discussion on *reducing emissions from deforestation and ecosystem degradation* (REDD), this initiative will not only create an enormous sink, but it will have important secondary benefits as well, such as coastal protection through mangrove restoration, improved water supply, and biodiversity conservation. The 2006 IPCC guidelines provide methodology for measuring REDD; however, are not yet being adopted.

2 The GEF Performance Index (GPI)

The likelihood of success of GEF projects and programs depends on, among other things, the capacity of countries' institutions to produce global environmental benefits. The purpose of the GEF Performance Index (GPI) is to "measure each country's capacity to successfully implement GEF programs and projects"[23]. The GPI is a quantitative measure of this capacity, combining data on (see **Figure** 5):

a. Government performance in relevant policy areas, measured by the Country Environmental Policy and Institutional Assessment Indicator (CEPIA),
b. Quality of management in selected areas of the public sector, measured by the Broad Framework Indicator (BFI); and
c. Quality of completed and ongoing environmental projects in the country, measured by the Portfolio Performance Indicator (PPI).

The Performance Index was a central part of the RAF from beginning, and perhaps the most controversial part (see technical paper #2). Interest in including country performance as a criterion for GEF allocations was derived from (a) a concern by Council members for the GEF to focus resources toward "high performing" countries; (b) the presence of an established practice of including institutional capacity and policy performance as part of performance-based allocation systems of multilateral development banks, notably IDA's approach; and (c) broad awareness of recent studies emphasizing importance of country policy environments in effectiveness of development assistance. Diagnostic tools to measure the quality

of countries' policies and institutional arrangement had been applied for decades – the CPIA was developed and initially employed in the mid-1970s – but by the time GEF Council discussions of the RAF had begun, a broad consensus had emerged that country policies and institutions do indeed "matter" for development results. In a 1999 review of development assistance approaches, for example, Eccles and Gwin observed that

"[I]mportant recent research has demonstrated reasonably conclusively that the policy environment into which donors lend largely determines the effectiveness with which aid funds are used. This has been known for some time by many aid practitioners and observers but resisted by others. Now we can be reasonably confident in asserting that however well individual projects may be designed, and even initially executed, they have a high probability of failing, or of not being sustainable, in an inappropriate policy context".[24]

	Index	Source	Ratings	Coverage
Figure 5: Components of GPI	**PPI** - Portfolio Performance Indicator, 10%	Average GEF project ratings in PIR 50%	Based on development objectives (DO) rating for all projects implemented since 1999.	four categories : highly satisfactory, satisfactory, partially satisfactory unsatisfactory
			Implementation progress (IP) ratings for all projects implemented since 1999.	
		Indicator from IEG ratings of ICRs for WB environment projects - 50%	Rating for objectives for all environment projects completed during the last 10 years	six categories –highly successful, partially successful, marginally successful, marginally unsuccessful, partially successful, and highly unsuccessful.
GEF Performance Index (GPI) **GPI = P1 x PPI + P2 x CEPIA + P3 x BFI**	**CEPIA** - Country Environmental Policy and Institutional Assessment Indicator, 70%	From WB Country Policy + Institutional Assessment (CPIA)	**Indicators**: 11. Policies and Institutions for Environmental Sustainability under cluster "Policies for Social Inclusion/Equity"	For policies in **six areas** – air pollution, water pollution, solid and hazardous waste, ecosystem conservation and BD protection, marine and coastal resources, freshwater resources and commercial natural resources.
	BFI - Broad Framework Indicator, 20%	Average of the five indicators under "Public Sector Management + Institutions" CPIA cluster.	**Indicators**: 12. Property Rights and Rule-based Governance, 13. Quality of Budgetary + Financial Management, 14. Efficiency of Revenue Mobilization, 15. Quality of Public Administration, 16. Transparency, Accountability, Corruption in Public Sector	CPIA indicators *not* included: 1. Macroeconomic Management, 2. Fiscal Policy, 3. Debt Policy, 4. Trade, 5. Financial Sector, 6. Business Regulatory Environment, 7. Gender Equality, 8. Equity of Public Resource Use, 9. Building Human Resources, 10. Social Protection and Labor

In the RAF formula, the GPI is a major element (with a weight of 1.0 as compared to the GBI weight of 0.8 for both biodiversity and climate change). At the same time, for technical reasons associated with measurement arrangements supporting the GPI, it is not a driving force in determining individual country ranks for allocations. In the biodiversity focal area, the first quintile (thirty countries) with the highest GPI ratings features thirteen with indicative allocations; while in climate change sixteen of thirty have indicative allocations[25]. That many of the highest-ranked performing countries have a group allocation may be viewed as ironic, since the quality of public policies and institutions is, compared with the level of need for global environmental benefits, an area of relatively high scope of influence by country governments. This in turn has significance for the impact of the RAF on incentives for countries to improve institutional and policy performance: Making notable improvements in a country's policy or institutional environment is unlikely, in itself, to result in a larger RAF allocation.

The Performance Index has technical and institutional foundations in highly regarded PBA systems, such as that of IDA. When asked about the extent to which the GPI uses best practices, the average response of

Delphi participants was 6.86 (standard deviation 1.6), with a higher average for the CEPIA and BFI component than for the portfolio performance indicator (PPI).

2.1 The CEPIA and the BFI

First, government performance is measured through the **Country Environmental Policy and Institutional Assessment Indicator (CEPIA)** from the International Development Association's performance-based allocation system. Data for this indicator are obtained from an IDA indicator called "Policies and Institutions for Environmental Sustainability" and are based on structured and internally reviewed assessments by knowledgeable World Bank staff of country performance in six policy areas (air pollution; water pollution; solid and hazardous waste; ecosystem conservation and biodiversity protection; marine and coastal resources; and freshwater resources and commercial natural resources).[26] This indicator is given a weight of 70 percent in the overall GPI; this means that the numerical value of data for this indicator is multiplied by 0.7 to produce its resulting value for the GPI. Among the three components, the CEPIA "counts for the most" in determining the overall value of the GPI.

Second, quality of public sector management is gauged through the **Broad Framework Indicator** (**BFI**), which carries a weight of 20 percent in the GPI. It is based on the average value for five indicators included under the "Public Sector Management and Institutions" cluster in the IDA CPIA: property rights and rule-based governance; quality of budgetary and financial management; efficiency of revenue mobilization; quality of public administration; and transparency, accountability, and corruption in the public sector. As these two indicators are derived from the same assessment and data source, from the annual World Bank assessment now called the IDA Resource Allocation Index (IRAI), they are analyzed here together.

Issue 1: Relevance and reliability of data

Delphi experts support the overall structure and weight accorded to country policy and institutional performance, with an average response of 7.55 (on a 1-10 scale) regarding the extent to which the CEPIA and BFI make use of best practices in performance measurement. The **CEPIA** indicator arguably involves an assessment of factors closely associated with the likelihood of success in delivering environmental projects. A correlation analysis of strength of association between the GPI and its components reinforces this, showing that the CEPIA correlates with the GPI at 0.99, while the correlations for the BFI and the PPI are 0.83 and 0.43, respectively.

While the possible **ratings** for the CEPIA for a given country range from 1.0 to 6.0, the actual range of scores is more limited in distribution than this. There is a long-standing tendency for ratings to hover considerably around the median, with most ratings in the "3" to "4" range. In the 2007 assessment, CEPIA scores for 75 IDA countries ranged from 2.0 to 4.5. This relatively narrow range of actual ratings has the effect of reducing the influence of the CEPIA on resulting allocations; diminishing the diagnostic power of the indicator, and indicating that methodology improvement may not make a notable difference to scores. As with the CEPIA, the effective range of ratings for the BFI is notably narrower than the possible range. In the 2007 World Bank Country Policies and Institutional Assessment (CPIA), for example, BFI ratings for 75 countries ranged from 2.9 to 4.0 within the possible range of 1.0 to 6.0. In fact, as seen in **Figure** 6, the performance indicators tend to cluster around a few ratings in the mid-range of the scale, contrasting with the skewed distribution of the GBI scores.

The CEPIA aims to "assess the extent to which environmental policies foster the protection and sustainable use of natural resources and the management of pollution"[27]. The relevance of the indicator's substantive coverage to biodiversity is high, but fight against pollution is not the primary GEF mandate. The indicator does not explicitly relate to the GEF mandate within climate change.

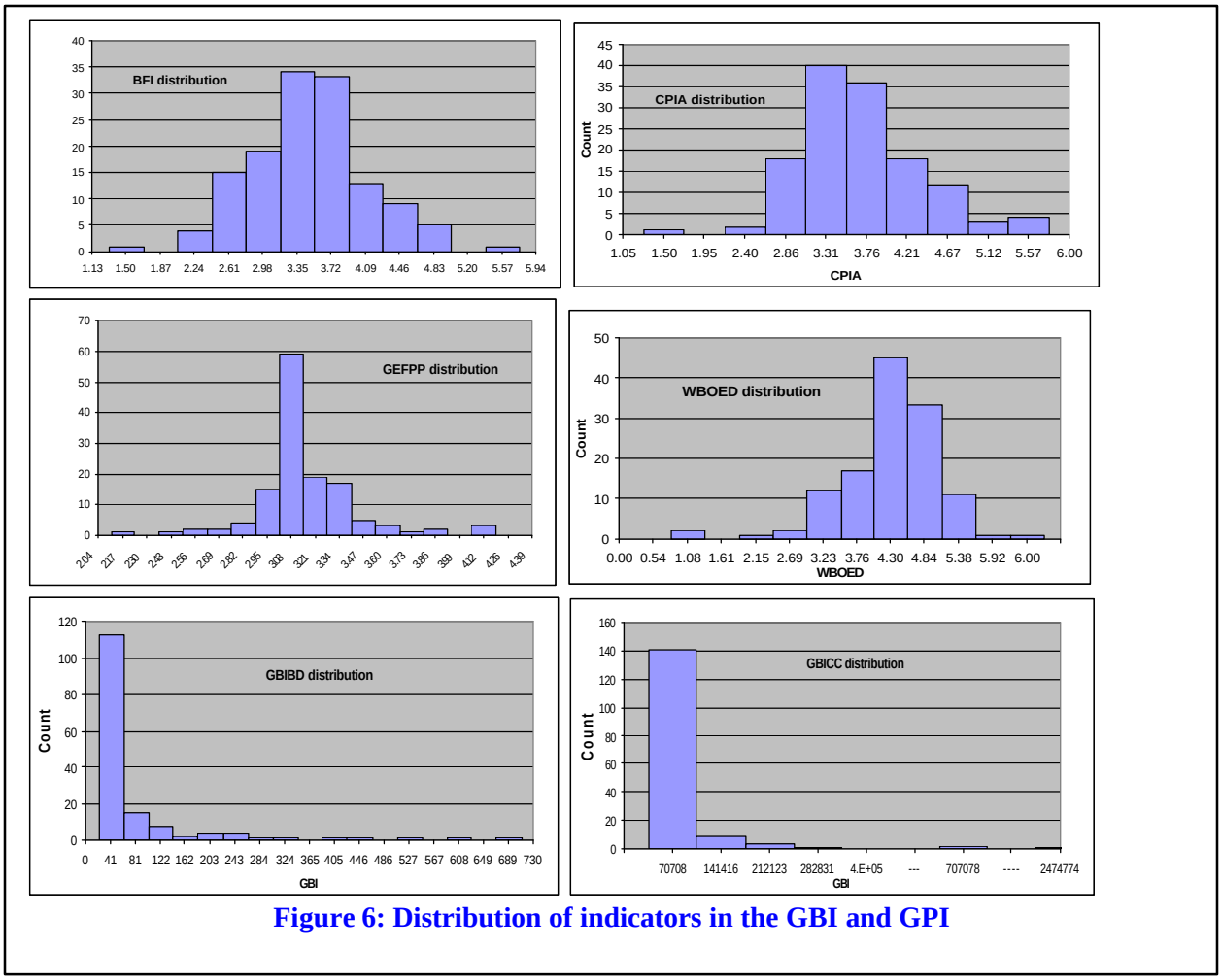


Figure 6: Distribution of indicators in the GBI and GPI

Many dimensions of environmental policy are bundled together in the narrative criteria used to guide scoring. For example, the following is the criterion for a rating of "4" in the CEPIA:
4. For both pollution and natural resource issues: Regulation and policies are in place but important gaps exist. Implementation is weak. Harmful subsidies exist. Very limited public information is available. EA [Environmental Assessment] is applied regularly in selected areas but gaps exist. Priorities are set but only partially adhered to. Sector ministries have basic knowledge of environmental issues.[28]

As a summary of a wide range of relevant issues pertinent to environmental policies and institutions, such a description covers a good deal of substantive ground. At the same time, the diagnostic value of a summary rating of this kind may be limited, since specificity regarding the applicable environment sub-sectors is lacking. In addition, the capacity of key environment-related ministries and other government bodies to formulate and implement well-informed policies is not directly addressed in the narrative.

The CEPIA is focused on environmental policies and institutions, and therefore is sectorally narrow in scope as compared to the BFI. The **Broad Framework Indicator** rating is a simple average of scores for the five indicators described above. Should one or more of these indicators be given relatively more weight, based on their greater relevance to environmental performance? Arguments may be offered that "quality of public administration" or "transparency, accountability, and corruption in the public sector"

are good candidates for added weighting, but on what grounds would these necessarily be more relevant than "property rights and rule-based governance"? The BFI is a simple unweighted average from five indicators because the indicators themselves seem thematically connected and existing knowledge provides no basis for any weights. In this context, an internal assessment by the World Bank in 2006 found that the BFI is "a sound index" for measuring governance to support IDA's resource allocation system.[29]

If the CEPIA and BFI data are not available for a country, a summary score from the Rural Sector Assessment Indicators developed by **IFAD** is used. This group of indicators includes measures in five thematic clusters, and was used for six climate change and seven biodiversity countries countries in the RAF allocation: (a) Strengthening the capacity of the rural poor and their organizations; (b) Improving equitable access to productive natural resources and technology; (c) Increasing access to financial services and markets; (d) Gender issues; and (e) Public resources management and accountability. While these themes may be especially relevant to assessing the health of the rural sector in developing countries, they feature limited or unclear relevance to the quality of environmental policies.

Issue 2: Coverage

On its face, the substantive coverage in the CEPIA indicator is good: Policies in the areas of air pollution, water pollution, solid and hazardous waste, ecosystem conservation and biodiversity protection, marine and coastal resources, freshwater resources and commercial natural resources are given separate sub-scores as this criterion's rating is created. This fine-tuning is lost, however, in the overall rating. In the end, policy status is rated jointly for both pollution and resource issues.

Delphi performance experts did not reflect a consensus on whether there are other available indicators that should be considered for use within the CEPIA or the BFI (average and median 4.0 on 1-10 scale). One panel member stated that the index does not include enough areas specific to climate change such as how countries are developing their energy and transportation sectors. Also, several remaining indicators from IDA's CPIA (IRAI) are *not* included in the overall Performance Index, including:

- **Cluster A: Economic Management** which includes three indicators - Macroeconomic Management, Fiscal Policy; and Debt Policy
- **Cluster B. Structural Policies,** which includes three indicators on Trade; Financial Sector, Business Regulatory Environment
- **Cluster C. Policies for Social Inclusion/Equity** (the CEPIA indicators falls under this cluster) – other not included are four indicators on Gender Equality; Equity of Public Resource Use; Building Human Resources; and Social Protection and Labor.

Overall, however, relevance of these indicators appears limited, and their exclusion from the RAF formula seems appropriate. The macroeconomic criteria address issues such as the quality of the exchange rate and the ability to sustain medium-term economic growth. Causal or operational linkages between macroeconomic policy levers such as these and environmental sustainability are not well documented. In the RAF climate change focal area, 35 of the 40 HIPC countries fall into the group allocation category, with debt policy unlikely to make a difference to country score. Guidelines for the sectoral indicators do not refer to issues important to biodiversity or climate change, such as rules for ecoservices, power concession or subsidies. As the environmental issues are concentrated in the CEPIA indicator, the others do not seem to have high relevance to the GEF mandate. The Delphi expert panel considered that the relationship is minimal between the CEPIA and the ability to implement GEF projects which are sensitive to women and to the poor (average on sensitivity to women was 2.60; to the poor 3.18), stating that these are separate issues and should be dealt with as such.

Issue 3: Balance and Weight

Within the GPI, 90% of the formula weight is accorded to country policy and institutional performance. This weighting arises from the view that the quality of policies and institutions (governance) is crucial to the success of GEF objectives. Delphi participants support the overall structure and weight accorded to country policy and institutional performance, finding that the CEPIA and BFI make use of best practices in performance measurement (7.55 average response, standard deviation 1.4). Experts were somewhat less convinced about the exact weight of 90% accorded to country policy and institutional performance within the GPI (6.10 average response).

2.2 The Portfolio Performance Indicator (PPI)

The third component of the GEF performance Index is the **Portfolio Performance Indicator** (PPI), which measures the quality of environmental projects. Compared to the CPIA which focuses on national policies and institutions, this indicator rates the quality of both *ongoing* and *closed* projects. Data for the PPI are derived in an equally weighted split between (a) an indicator summarizing ratings contained in Project Implementation Reviews (PIRs) of GEF projects and (b) an indicator that summarizes ratings by the World Bank Independent Evaluation Group (IEG) of Implementation Completion Reports (ICRs) from World Bank environment-related projects in the country. The PPI counts for 10 percent of the calculation of the GPI.

Issue 1: Relevance and Reliability

In principle, an indicator that captures performance of projects would be highly relevant to assessment of the capacity to produce global environmental benefits through GEF projects. Both the ICR and the PIR are based on self-assessments by project management, but internal verification processes have been established to ensure consistency in ratings. The GEF Evaluation Office established a verification process for ICRs through its Annual Performance Report and is comfortable overall with the ratings[30].

There was a high level of support of Delphi performance experts for the proposition that PIR scores and World Bank project implementation scores are likely to be a useful partial measure of a country's ability to successfully implement GEF projects (average 7.11 for the PIR and 6.22 for the World Bank ICR scores, standard deviation 1.4). The PIR is highly relevant in that it rates GEF projects only, implemented by all Agencies. It covers ratings for all GEF projects, in all focal areas, going back the last ten years. The long time horizon means, however, that the performance being assessed is relatively dated.

The ICR rating covers environmental projects, whether funded by GEF or not, implemented by the World Bank. The classification of "environmental" is quite broad, and the decision of which to projects to include is based on judgment by the GEF Secretariat. On the other hand, experts interviewed have argued that for portfolio and project management, the broader the coverage the better, as a country's ability to manage projects is relatively consistent whatever the sector. A broad coverage will reduce the risk of a few less successful projects unduly influencing the country score.

The PPI does not cover terminal evaluations from Agencies other than the WB. When the RAF was developed, the Evaluation Office had only reviewed a limited number of terminal evaluations, so there was no possibility yet of including the ratings of completed GEF projects as verified by the Evaluation Office in the performance of project indices. The number of completed projects with verified outcome ratings has increased and potentially this could be included in the next version of the RAF. At present the Evaluation Office has outcome ratings for 143 projects, and ratings are expected for more than 200 projects by the end of this fiscal year, and for about 270 projects by the end of the fourth Overall Performance Study.

Possible measurement issues in both the PIR and ICR indicators include:

- **Small portfolio effects**. A major concern regarding portfolio indicators for performance-based allocation systems is the small portfolio problem. Many countries do not have enough environmental projects ongoing to allow for statistical stability in measurement of their quality. Some PBA systems use various methods of "smoothing" data (ten-year moving averages of ratings, for example) to address this problem. The PPI summarizes data for all projects implemented in a country since 1999 (for the GEF PIRs) or during the last ten years (for the WB IEG ICR reviews). This does help, but it does not fully address the challenge faced by countries with small portfolios.

  For the PIRs, 28 countries have only one project rated; 19 have two countries, fourteen countries have three projects rated, while 19 countries have four projects rated. Twenty-nine countries have a medium portfolio between five and ten projects. Only eight countries have more than ten different projects rated. This does not imply that countries with a small portfolio have worse ratings than larger countries; they are, however, more vulnerable to changes in ratings. In the 2008 reallocation dataset, there is only one *highly unsatisfactory* rating (for an Asian country with three projects rated).

  The coverage is better for the ICRs, for which 59 countries have ratings for more than ten projects, and 35 countries have between five and ten projects. Eleven countries have only one ICR; ten have two projects rated; seven have three and ten have four projects rated (in the reallocation data).

- **Timeliness**. Data from terminal evaluations can be made available some time after project completion, causing the resulting indicator to be a measure of past performance, or of projects no longer of high relevance to the country or to the GEF. Of the 42 terminal evaluations submitted to the 2007 APR, for all Agencies, 36% were submitted more than six months after project closure. On average, there is a delay in submission of the final evaluations to the Evaluation Office of 7.9 months (APR 2007). Past performance may signal future performance, but if the indicator is to contribute to stronger country incentives for improved performance the data should be reasonably recent.

Issue 2: Coverage of issues in the PPI

The ICR review component of this indicator is both too broad and overly narrow in its coverage: It summarizes the quality of all World Bank environment projects in the country, which may include projects in water service delivery, other areas of infrastructure and additional areas that are not closely relevant to biodiversity or climate change. World Bank environment projects, even if they do fall into a RAF focal area, do not necessarily feature generation of global environmental benefits as objectives. The ratings for closed projects only include World Bank activities, not projects from other Agencies. The slow moving project cycle implies that relatively few GEF projects have closed, but numbers have increased as of late and there is now 143 terminal evaluations with reviews and ratings by the GEF Evaluation Office. Of the 143 terminal evaluations, the World Bank has submitted 63 (44%); UNDP 56, UNEP 21 and other Agencies three.

To overcome the bias of smaller portfolio and provide stability, the PPI covers ratings over the last ten years. Delphi experts did not reach consensus whether to shorten this period (average 4.45 and standard deviation of 2.7); though one commenter did suggest that the period be shortened from 10 years to 5 years. PIR ratings cover yearly ratings for the same projects, though it is relatively rare that there are major fluctuations in ratings for one project over time. On average, PIRs give ratings 0.83 points higher than those given by the Evaluation Office. Of the 21 projects that have been rated in the unsatisfactory range by the Evaluation Office, only one had been so rated in the PIR.

Another approach to counter bias is the inclusion of several ratings. For the PIR, both development objective (DO) and implementation progress (IP) are rated. Implementation progress addresses how well project has progressed as per the schedule and to what extent it is meeting its major milestones (relates mainly to outputs). The likelihood of achievement of development objective refers to the chances of achieving the expected project outcomes and objectives. It incorporates an element of uncertainty related to inherent risks (so even though project implementation progress has been satisfactory it may have a lower likelihood of achieving its development objectives due to various factors that affect conversion of outputs into outcomes and impacts). The ICRs may contain many ratings[31], but only one is used in the index, namely for the overall outcome.

The PIRs and the ICRs assess, among other things, the quality of the implementing organization in reaching planned project accomplishments. In the case of GEF projects, this organization is a GEF agency. Historically, governments have executed 70% of GEF fullsize projects (FSPs) and 36% of medium-size projects (MSPs) (Source: Joint evaluation), and in most cases the government also plays a role in projects executed by NGOs and private sector partners. Under the banner of national execution as established by the Economic and Social Council of the United Nations (ECOSOC), it is reasonable to consider non-governmental sources among national capacities. It is less certain how the government influences projects managed by foundations, multilateral or bilateral entities (28% historically of FSPs, though many of these are global projects). Government influence on the ability of a project to successfully meet its objectives or to make implementation progress may therefore be indirect. Several country stakeholders raised concerns that Agency performance is equally, or maybe more, critical than country capacity for project performance. To the extent that data are available for both Agency and country performance in the ICR, the mid-term review did not find that these ratings notably differed, or favored Agency performance over the borrower's.

Also, in relation to the GEF Performance Index, countries consulted asked if they had been disadvantaged by limited capacity and experience, and whether cultural expectations play a role. For example, does SIDS management of government fit the World Bank criteria? As indicated by the ADB, it can be difficult to compare governance systems of very different countries and cultures. Some countries, such as those in the Caribbean, have limited experience with the World Bank, and have instead worked with UNEP and UNDP whose ratings are not used if available. Countries having benefited from mainly regional programs are not rated. This would be a disadvantage if the performance is high but not rated. Country stakeholders expressed the fear that the indices might "over-magnify or undermine" the actual levels of performance of a country.

Issue 3: Balance and Weight of Portfolio Performance

Within the GPI, a total weight of 10% is accorded to the Portfolio Performance Indicator (PPI). Half of this weight (5%) is derived from the average GEF project ratings for each country in the PIR and the other half (5%) is based on ratings by the World Bank Independent Evaluation Group of environmental projects completed during the past ten years.

Delphi participants indicated general support for the **inclusion of portfolio performance** in the GPI although that support is not as strong as support for the CEPIA and BFI. One commenter referred to portfolio performance as a "reality check". There was neutral support for the current weighting of portfolio performance within the GPI, with an average response of 5.30, though some experts would support a higher weight for the PPI up to 30%.

Statistical simulations were undertaken to change the PPI weight up to 30% while keeping constant other GPI ratios (CPIA/BFI = 2/9 and CPIA + BFI + PPI = 1). The results in both focal areas show that the pattern of allocations are not significantly sensitive to increased PPI weight.

Issue 3: Channeling resources to high performing countries

The 57 countries with individual allocations in biodiversity accumulate 41% of the GPI scores and 75% of the funds, while the group allocation countries obtain 59% of the total GPI scores (see table below). In climate change, the shares are more spread; 35% versus 65% of GPI from individual and group allocation countries, respectively. See **table** 4.

Table 4: Shares of GPI scores and funding

Allocation type in BD	# of countries	Allocation in BD	max alloc (m$)	min alloc (M$)	GPI	max GPI	min GPI	mean GPI	med GPI
indicative	57	75%	63.2	3.5256	**41%**	**3.9**	1.9	3.1	3.1
Group	93	15%	3.5255	1	**59%**	**4.4**	1.3	2.9	2.8
Allocation type in CC	**# of countries**	**Allocation in CC**	**max alloc (m$)**	**min alloc (m$)**	**GPI**	**max GPI**	**min GPI**	**mean GPI**	**med GPI**
indicative	46	75%	150	3.09	**35%**	**4.3**	2.3	3.2	3.1
Group	115	15%	2.97	1	**65%**	**4.4**	1.3	2.8	2.8

Within the GPI for the first 20 highest ranked countries; eight out of twenty are indicative countries allocated with M$169.27, and the others have higher GPI but lower GBI, and are allocated to the group. However, the high GPI indicative countries accumulate 16.9% of the total resources. In general the countries with highest GPI are found in the group allocation, and the group allocation countries has more GPI scores collectively. The individual allocation countries spread across the quintiles described in **table** 5.

Table 5: Distribution of GPI and GBI scores across GPI quintiles

GPI Quintile	GPI	GPI %	Alloc M	Allocation %	Max GPI	Min GPI	GBI	GBI %
1	113.83	26.7%	$296.99	33.00%	4.43	3.37	2366.42	30.99%
2	95.04	22.3%	$246.54	27.39%	3.36	3.05	2011.54	26.34%
3	87.58	20.6%	$152.95	16.99%	3.05	2.74	1180.36	15.46%
4	79.84	18.7%	**$94.55**	10.51%	2.73	2.49	694.76	**9.10%**
5	49.72	11.7%	**$108.98**	12.11%	2.49	1.27	1384.14	**18.12%**
Total	426.01	100.0%	$900.00	100%	4.43	1.27	7637.20	100%

The GPI has a small range of only 1-5, and its values spread more evenly cross the recipient countries. The six countries with highest GPI score are all group allocation countries, mainly in Europe[32]. As seen in **figure** 6, some of the highest GPI score countries are fixed at one million US$, some with low performance scores receive high allocations; and the rest is spread unevenly. The figure is for biodiversity; patterns in climate change are the same.

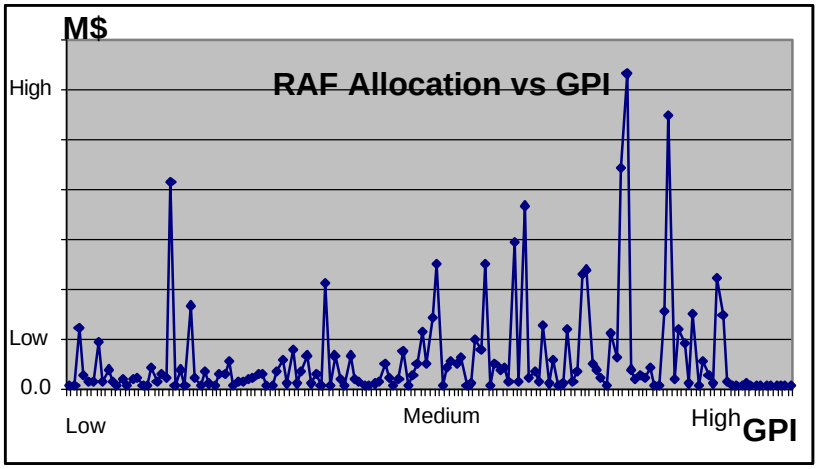


Figure 6: RAF Allocation and GPI. All countries in biodiversity are plotted (because of CPIA confidentiality the axis are neutral).

It is fair to say that the RAF channels resources to high GBI countries - but adjusted with their performance. However, the formula could provide encouragement for a group allocation country with relatively high GBI to increase its performance rating to be lifted out of the group, and for an indicative country to increase its country share of resource allocations. However, a country with an already high performance score will have less room to improve the country's score as measured by the GPI ranging from 1 to 5, or conversely reduce its GPI. For example, China has considerable room to lower its performance effort to keep the ceiling allocation while still retaining 150 M$ in climate change.

Issue 4: New data developments in performance

When Delphi participants expressed doubt whether there were other available indicators that should be considered for use within the CEPIA or the BFI (average 4.0). There was also no consensus on the question of moving to a scoring system based on the percentage of projects at risk in the portfolio (average 5.00). Delphi participants support the current construct of the PPI as a reasonable indicator of countries' abilities to implement GEF projects.

More in-depth environmental assessments are undertaken on a periodic basis by various donors. Country stakeholders mentioned the Country Environment Assessment (CEA). In countries where they are conducted, the findings of the CEA may influence the World Bank's annual CPIA data on the environment sector policies and institutions. The GEF Country Portfolio Evaluations provide an informative overview of the environmental framework in which the GEF support is given, although they do not contain performance ratings.

3. Other Design Elements

This section goes into further detail on specific design elements related to key questions on the indices. These elements include (a) weights used in the formula, (b) flexibility and volatility, and (d) exclusions to the formula, together with floors and ceilings to allocations. The section also refers to the practice of other PBA systems on these issues. Further information of the *effect* of this design is covered in technical paper #4.

3.1 Weights in the RAF Formula

This section addresses the key questions on the weights of performance within the indices, and the possible impact on allocations if the weights in the allocation equation were different.

Definition and Importance of Weights

To apply a weight in an allocation formula is to use some numerical multiplier (or other coefficient) to data for a component in the formula in order to control its importance or influence in relation to other components. Elements given a relatively high weight contribute more to the calculation result than do elements given a lower weight[33].Weighting components in a performance allocation formula is a method of ensuring that the various components relate statistically to each other in conformance to the goals and standards of the formula's design.

There are two ways in which organizations have chosen to apply weights to factors in their PBAs. First, one way is to assign weights to different indicators and then *add* them up. The GEF's calculation of its Performance Index (GPI) is an example of this approach. The GPI is the sum of three different country scores: the Portfolio Performance Indicator (PPI); the Country Environmental Policy and Institutional

Assessment Indicator (CEPIA), and the Broad Framework Indicator (BFI), each with its own weight in the GPI formula – of 10 percent, 70 percent and 20 percent:

$$GPI = 0.1\ PPI + 0.7\ CEPIA + 0.2\ BFI$$

This is a common, simple approach. The IDB uses it throughout its PBA. Other organizations (AfDB, CDB, IDB, IFAD and the World Bank IDA) do not use it throughout the PBA, but they do tend to use it in calculating the performance component.

Second, an organization may assign weight is the use of *exponents* in a *multiplication* formula. The overall GEF RAF formula is an example:

$$\text{Country Score} = GBI^{0.8} \times GPI^{1.0}$$

The GBI and GPI exponential weights are 1 for the performance index and 0.8 for the benefit index. The weight = 1 indicates no changes to any GPI values when applying the formula. Virtually all organizations that use PBAs, with the single exception of the IDB, apply this form of allocation formula. There seem to be two reasons why this rather complex mode of assigning weights is popular. First, it was the format of the pioneering allocation formula of the World Bank in the 1990s. Second, there is a common belief that using the exponents in a multiplication formula gives greater weight to performance. It is correct, in general, that a larger exponent means a higher weight to that variable. Because of its exponential nature, however, it is not a straightforward link. More precisely, the sensitivity of the country's allocation to changes in its score on a certain factor in a PBA (statistically referred to as the "beta" of that factor) is determined by several things, including the mix of other factors in the equation, their arithmetic weights and exponents (if any), and the volatility[34] of the scores on various variables in the equation.

Simulation of Weights

In a multiplication formula, with a different mix of variables with different exponents, there are many things that affect the resulting pattern of country scores. The relationships between allocations and variables in such a formula are complex. The complexity of this type of formula, and common misunderstandings about how it works, can be serious arguments against it. While multiplication arrangements continue to be found useful in the various PBA formulas, at times some institutions have refined their formulas to exclude certain multiplication or exponential aspects to reduce volatility in scores and enhance predictability of results.

Because of its exponential nature, the effect of changing the weights differs greatly depending on which weight is increased or decreased. The mid-term review tested three scenarios using various modifications to the RAF formula, including quantitative relationships between the GPI and the GBI. Key results are as follows:

Scenario A: Increasing GPI weight from 0.5 to 4 (by small steps), while keeping the GBI weight constant at 0.8. In both focal areas, increasing the weight of performance would reduce the number of countries with indicative allocations, while increasing the maximum allocation to the group countries as well as the indicative countries. In biodiversity the rankings among top-scoring countries would change. The pattern is the same in both focal areas, though because of the 15% ceiling, China would maintain its current allocation in climate change. See table 6.

Table 6: Scenario A: Increasing weight of GPI	GBI-GPI (0.8-1) Actual case	GBI-GPI (0.8-2)	GBI-GPI (0.8-3)	GBI-GPI (0.8-4)
Climate change				
Number of indicative countries	46	44	42	40
indicative top (M$)	China M$150	China M$150	China M$150	China M$150
indicative low (M$)	Uganda M$3.09	Uganda M$3.152	Morocco M$3.37	Bangladesh M$3.47
Number countries moving to group	None	1	2	2
Number of countries up to indicative	None	3	6	8
group ceiling (M$)	Tunisia M$2.97	Croatia M$3.149	Uganda M$3.14	Ethiopia M$3.46
Biodiversity				
Number of indicative countries	57	53	51	49
Countries moving to group	None	5	7	10
Countries moving up to indicative	None	1	1	2
indicative top (M$)	Brazil M$63.22	Brazil M$69.12	Mexico M$74.04	Mexico M$83.45
indicative low (M$)	Afghanistan M$3.526	Korea Rep M$3.54	Nigeria M$3.91	Samoa M$3.54
group ceiling (M$)	Suriname M$3.525	Suriname M$3.51	Malawi M$3.86	Nicaragua M$3.54

Scenario B: Keeping the GPI weight equal to 1.0, while decreasing GBI weight from 1.0 (by step = 0.1). This case would have a much larger effect on the current pattern of allocations. As the GBI is applied below its current exponent, the number of countries changing from indicative to group allocations, or in the other direction, may be a few or more than a dozen, depending on the particular size of the GBI weight. See table 7.

Table 7: Scenario B: Decreasing weight of GBI	GBI-GPI (1-1) Neutral	GBI-GPI (0.9-1)	GBI-GPI (0.8-1) Actual case	GBI-GPI (0.7-1)	GBI-GPI (0.6-1)
Climate Change					
Indicative countries	35	40	46	53	61
indicative top (M$)	China M$150	China M$150	China M$150	China M$150	China M$118.37
indicative low (M$)	Syria M$3.44	Morocco M$3.18	Actual case	Nepal M$2.92	Sri Lanka M$3.31
up to indicative				7	15
down to group	11	6			
group ceiling (M$)	Tanzania M$3.37	Kenya M$2.86	Tunisia M$2.97	Mali M$2.84	Burkina Faso M$3.30
Biodiversity					
indicative	44	50	57	64	71
indicative top (M$)	Brazil M$85.73	Brazil M$74.30	Brazil M$63.22	Brazil M$52.70	Brazil M$43.00
indicative low (M$)	Jamaica M$3.80	Egypt M$3.55	Afghanistan M$3.526	Paraguay M$3.77	Ghana M$3.70
up to indicative				7	14
down to group	13	7			
group ceiling (M$)	Zambia M$3.80	Cape Verde M$3.39	Suriname M$3.525	Korea Rep M$3.60	Botswana M$3.56

Scenario C: Neutral weight: both GBI and GPI with weight 1 (seen in column in table 7 above). This approach would reduce the number of countries with individual allocations from 46 to 35 in climate change, and from 57 to 44 in biodiversity. The top ranked country in biodiversity would get 22.6 US$ million more than currently.

After viewing multiple analytical perspectives, it becomes clear that while changes in the relative weights of GBI and GPI can potentially produce significant shifts in country allocations, under the current RAF weights it is GBI and not GPI that is the driving force behind the distribution of allocations for both focal areas. Is this the way things "should" be? Do existing good practices in design of PBAs inform us further about the weights in the RAF? These are not particularly easy questions to answer since there are no formal criteria for weights in PBAs. The Delphi panel of experts indicated a positive (though not

overwhelmingly positive) assessment on the appropriateness of the relative weights of the GBI and the GPI appropriate (average score of 6.25 of 10). There was recognition among participants on the panel that the weighting is a "judgment call".

Exponential Weights

Exponential weights encourage volatility in allocations. With exponential weights, relatively minor changes in a country's score on that variable can radically affect its dollar allocation, particularly if there is a large exponent on a variable. See a theoretical case in Box 3.1.

A multiplication formula with exponential weights leads to a volatile system if the variables in the formula change their values significantly from year to year. Exponential weights and change in performance scores can be a rather explosive combination. Issues of volatility and flexibility are discussed in the next section.

Box 3.1: Illustration: Exponential Weights and Performance Change

As an example, consider allocations to two countries whose scores in one year are the same on all variables. Suppose both score "10" on the potential/needs variable and "4 out of 6" on the performance variable. Let us say that the exponent on the performance factor is 5 in this formula (as it is, for example, in the IDA performance assessment system). Therefore both countries have the same performance score of $(4)^5$ and the same total score of $10\ x(4)^5$, which is 10 x 1,024 or 10,240. They receive equal allocations of funds.

Let us imagine that in the next year nothing changes except one of the two countries increases its performance score from "4 out of 6" to "5 out of 6". Now its total score is $10\ x(5)^5$, which is 10 x 3,125 or 31,250. Instead of receiving 50% of the money to be allocated, it will receive $10\ x(5)^5/[10\ x(5)^5 + 10\ x(4)^5]$, or 31,250/41,500 which equals 75%. To push the point further, if two countries are separated by 2 points on the 6-point performance scale then the better performing country will receive 93% of the funds. This seems an unduly discriminating outcome, considering the imprecision of the performance scoring upon which the ratings are based.

From the point of view of the country whose scores remain the same, one year it receives 50% of the funds. The next year its performance is the same but it receives only 7% of the funds.

3.2 Flexibility and Volatility in Allocations

This section addresses the key questions on how flexible the RAF formula is, how the scores and allocation fluctuate (volatility), and how the formula takes account of changes the underlying indicators.

Flexibility tools

In a lending or granting program, formula-based resource allocations must be balanced with effective demand from eligible member countries. Effective demand is partly a matter of a member government's interests and priorities and partly a matter of absorptive capacity. Since these vary, and since a program's impact will be significantly affected by the timely utilization of its resources, most organizations have found it important to build flexibility, or readiness to adapt to changing circumstances, into their resource allocation frameworks. There are several ways in which this is done:

- **Reserves**. Some organizations have traditionally maintained a significant portion of their concessionary funds in an unallocated reserve. The use of reserve funds is unfettered, subject to Board approval of specific projects. The IDB, for example, has a reserve of $100 million.

- **Set-Asides**. Most organizations have set-asides of funds (taken out of the overall pot of available funds before the country-wise allocation is made) for special purposes. The special purposes have

included, for example, regional projects, emergency/disaster response, post-conflict or otherwise fragile states, or high priority public goods such as the control of AIDS, and other specifically targeted areas such as the reduction of extreme poverty or the preservation of endangered species. Funds set-aside are normally unrestricted geographically. The portion of funds set-aside varies from about 5% (for example, the IDA allocation for regional projects) to as much as 25% of the total concessionary funds being allocated (for example, Caribbean Development Bank (CDB) set-asides for regional projects and special purposes). The GEF RAF has a 10% exclusion of the climate change and biodiversity funds (including 5 % for global and regional activities and 5 % for cross-cutting issues, the Small Grants Programme and capacity building)[35].

- **Reallocations**. If reallocations are sufficiently frequent they can make a major contribution to flexibility. This can be done in two ways. First, the formal allocations can be done more frequently. World Bank (IDA) allocations are done annually, for example, rather than every two years. Second, a single allocation can have several iterations. That is, the first iteration can produce allocations for all eligible countries. Then the "excess" allocations above expected country effective demand can be "re-pooled" and allocated again using the standard allocation formula. This is approximately how IFAD conducted its 2006 allocation exercise. The key concept was that allocation iterations must be *de novo* – that is, the re-pooled funds are reallocated using the same performance-based formula as in the original allocation.

  The World Bank (IDA) operates a three-year rolling allocation system, with reallocation annually. In the third year of the replenishment period there is provision for a limited amount of additional ad hoc reallocation of funds. Regional Vice-Presidents can request the reallocation of funds from specific low-demand countries to specific high-demand countries, as long as funds flow only from lower-performing countries to higher-performing countries.

- **Special Pools of Funds**. GEF has a special pool of 15% of the total funds that it has available for biodiversity and climate change grants (the group allocation). Countries that would otherwise have very small allocations are eligible for grants from the pool. Flexibility is enhanced by the fact that the maximum grant for any country in the pool is fixed by the highest allocation that any country in the pool would receive if they all received formula-based individual allocations. On the other hand, the pool has a rigid ceiling which cannot be changed, so that not all countries can access the maximum.

  The ADB also operates a pool of funds for its Pacific Region, which, with the exception of New Guinea, comprises small island states, at 4.5% of total funds. However there is no obvious gain in flexibility. The allocation within the pool uses the standard ADB allocation formula resulting in individual allocations (which are, on average, larger than they would be if the small states competed directly with all ADB members for concessionary funds, rather than having their own pool).

- **Front-Loading** within a replenishment period. Flexibility is enhanced in most organizations by front-loading provisions within the replenishment period. Under such rules, a country may seek access to funds which normally would be available only in a later time segment of the replenishment. However most organizations do impose limits on front-loading. The rationale for limiting carry-forwards and carry-backs is to discourage governments from using a whole allocation and returning for more, or neglecting to use an allocation until the last minute. It is debatable, however, whether this rationale outweighs the practical difficulties that are created by constraints on how quickly an allocation can be drawn down – especially the practical difficulties faced by small countries trying to put together a viable sized project in face of an already small allocation.

  An example of provisions for front-loading is given by the World Bank (IDA), which, in year 1 of its rolling three-year allocation period, can increase an individual country's allocation by up to 30%.

Similarly the first year share of a country's AfDB allocation can be increased by up to 50%, and the ADB first year allocation can be increased by different amounts depending on country size, up to 75% for the smallest countries. The GEF first year allocation can be doubled, but in that case nothing remains for the second year, whereas in the other organizations cited the second year allocation is generally unaffected by the front-loading in year 1. Front-loading from a two-year perspective provides greater flexibilities than from a one-year perspective. The MTR Delphi panel on performance indicated moderate to strong agreement with enabling access to 100 % of their allocations to fund viable projects (average 7.27 on a 10-point scale).

- **Back-Loading** within a Replenishment Period. In general organizations do not limit back-loading (waiting till late in the replenishment period) until the final year. In the final year allocations that have not been taken up tend to be subject to reallocation.

- **Waivers, Exceptions and Ad Hoc Adjustments**. Information on waivers, exceptions and ad hoc adjustments to allocations is difficult to obtain. Nevertheless they seem to contribute substantially to the flexibility of resource allocation systems in some organizations. For example, actual amounts approved by the CDB for some countries, during Special Development Fund (SDF) 5, were in some instances twice the formal allocation or greater. At one time, before disclosure of allocations became the norm, the World Bank (IDA) management of the performance-based allocation system assumed that actual approvals might reasonably vary between plus 20% and minus 20% of the indicative allocation. The World Bank, in all its communications on the system, emphasized that the allocation was indicative, not an entitlement.

- **Final Year Adjustments**. Most organizations allow more flexible adjustments to allocations than usual in the final year of the replenishment period. For example, the World Bank (IDA) allows shifting of allocated monies from one country to another in the final year of the allocation period, on a paired case-by-case basis. Such case-by-case reallocations must be from a country with a lower performance score to a country to country with a higher performance score. The RAF includes no provision for final-year adjustments.

Flexibility in response to socioeconomic changes and crises

Another aspect of flexibility is a PBA's capacity to respond to special country circumstances that may be relevant to need or performance, such as significant economic downturns, natural disasters or extensive civil conflict. The allocation systems of various donor institutions vary in their approaches to such situations. In some, such as the RAF, there is no provision for such circumstances in the PBA itself. Historically, countries in conflict situations did not tend to access funding for GEF projects.

In other PBAs, particular approaches are applied. Fragile states receiving African Development Bank assistance, for example, may receive a "top-up" (increase) to their country allocations as fragile states have their own special window for support. In the ADB, set-asides for conflict and post-conflict situations account for 10-15% of the Asian Development Fund. For example, Afghanistan is given special consideration, through a 6-year transition, for post-conflict environment. Some IFIs, such as the CBD, use their PBA system flexibly so as to respond to disasters. The Caribank also has a Vulnerability Index which takes into account both economic shocks and natural disasters. In this way Grenada, for example, moved from a medium to a high allocation following hurricane damage. As an organization based regionally, CDB is expected to respond to such needs. International agencies may be able to adopt a more rigid and anonymous type of allocation system. Of course, other organizations normally also have crisis and emergency support as part of their mandate. The World Bank has a separate assessment for post-conflict countries to replace the CPIA scores. This assessment, not used by the GEF, includes other

indicators such as demobilization of militia. Then, as circumstances allow, a transition period begins using a mixture of the PCPIA and the CPIA.

One issue regarding climate change activities is the significance of crisis events for energy intensity. Civil conflicts or natural disasters could result in major and sudden increases or decreases to energy intensity which would not be captured through GEF's current system for scoring GBI in climate change. The Delphi panel of experts developed only moderate agreement on the importance of crisis events for energy intensity (average 4.81 of 10). A sudden crisis would be likely to abruptly - but temporarily - increase the carbon intensity until the crisis is ameliorated. The effect would be largest and most lasting in countries with the least capacity (financial, governance, expertise) to deal with environmental crises such as earthquakes, storms and widespread flooding from storm surges (such as experienced in Myanmar in 2008).

For biodiversity, some sources have noted that the GBI-BIO includes no measure of the *urgency* of needed actions. Since it is sometimes the case that response to urgent biodiversity challenges is critical, it may be helpful to have such a component in the index, if technically feasible.

Relatedly, some have suggested that GEF develop a "rapid response mechanism" for addressing immediate effects of environmental disasters such as extended drought, flooding, and similar event.[36] It is widely recognized that lower-income countries tend not to have adequate systems of national response to environmental crises. The RAF does not feature a rapid response mechanism, and in addition the climate change funds are restricted for use in mitigation rather than adaptation. But a set-aside or other flexibility mechanism to support rapid response may become especially relevant to the RAF if and when its scope is expanded to focal areas beyond biodiversity and climate change.

Crises in various countries have nevertheless drawn project assistance from the GEF, sometimes via RAF funding. The GEF was able to quickly approve support to the 2008 China earthquake, for example. This included, under POPs, the project *Rapid Assessment of Chemical Contamination of the Wenchuan Earthquake in Sichuan Province* (ID3702). Also, an MSP was approved; *CBPF: Emergency Biodiversity Conservation Measures for the Recovery and Reconstruction of Wenchuan Earthquake Hit Regions in Sichuan Province* (ID3706). This project came under the China programmatic approaches and was funded from the county RAF allocation.

Volatility

Given its weight in the formula, changes in a country's GBI can readily influence the resulting allocation. What about changes in the GPI? Currently, the updates are conducted every two years, and for the reallocation at midpoint not all the biodiversity GBI-BIO data were available. The changes indicate that the RAF is more stable than volatile as currently designed, though a few countries may experience fluctuations in scores and allocations.

The mid-term review conducted simulations of changes in data, from updating the CEPIA and BFI scores from the World Bank's 2005 data set, which were used for the GEF-4 allocations, with the data from 2006 to the data set used in the 2008 reallocation. (The PPI dataset is relatively stable as it uses the least ten years or ratings). Overall, the changes in both biodiversity and climate change at a macro level (in number of individual allocation countries, for example) are not great, but at the country level there are significant changes for a minority of countries.

The simulations revealed that a simple one-year updating of performance data results in only minor changes in the overall pattern of allocation, but for a relatively small number of countries such an update can be significant. Also, as the foundational data are updated, a country's ranking may shift even if its

own data have not changed. Changes observed are (a) changes in the number of individual allocation countries and group allocation countries (a couple up and down); (b) small shifts in group ceilings; (c) decrease or increase in allocation of more than 10 percent (two countries only).

3.3 Group allocation, supplements, floors and ceilings

This section considers the key question of the impact on allocations of the various exclusions to the allocation formula based on indices, as well as the impact of the floors in allocations. (The 10% exclusion for the SGP and global and regional projects is discussed in technical paper #4.)

The Group Allocation and Targeted Supplement

Unusually among IFIs[37], the RAF created two categories of countries that are treated differently in regard to allocations. The RAF document specifies that the countries with highest score accumulating to 75% of all focal area funds will receive individual allocations. The remaining countries can access collective funds, amounting to total funds minus the exclusion (or set-aside) of 10% and minus the funds given to individual countries. Before the adjustment of 75% is made, each group country has a preliminary allocation that corresponds to its score from the RAF formula. This results in two categories of countries:

- Category 1: The indicative countries. This category of countries is defined by the smallest number of countries whose allocations sum to 75% of the funds available for the focal area. Each country in Category 1 received an individual allocation in GEF-4.[38]
- Category 2: The group allocation countries; all countries not in Category 1.[39] A country in this category can access a maximum amount[40] that is equal to the highest individual allocation that any country in this group would have received if there had been no grouping.[41] That is, all countries in category 2 have the same maximum amount that they can access (subject to a cash constraint for the whole group).

For each country whose *preliminary country allocation* is less than $1 million, *The RAF document* specifies that these countries will be lifted up to 1M US$, so that the country will have a minimum *adjusted allocation* of $1 million. The 1 M US$ becomes the floor. The additional amount needed to bring these countries up to 1 million US$ is called a *targeted supplement*, in effect a kind of subsidy to these countries. In total it amounts to $ 41.3 million, or only 2% of the resources for the two focal areas.

In **biodiversity**, of the 93 countries in group allocation, their preliminary allocations vary before they were pooled in the group funds:

- Fifty-three countries were allocated in a range between M$ 1.02 and M$ 3.53.
- Ten countries were set to the minimum allocation of one M$, in accordance with paragraph 26 in *The RAF document*, if they lack historical allocations, are not GEF Participants, or lack basic data for the RAF indicators.
- Thirty countries were increased to the minimum allocation of one M$ as their preliminary allocations were found below M$ 1. The amount needed for the targeted supplement for the biodiversity group was M$ 15.37.

In **climate change**, for the 115 countries in group allocation:

- 41 countries were allocated in a range between M$ 1.03 and M$ 2.97 in preliminary allocations.
- 33 countries were set to the minimum allocation of 1 M$ due to paragraph 26 guidance in the *RAF document*.
- 41 countries were increased to the minimum allocation of 1 M$ as their preliminary allocations were found below M$1. The amount needed for the targeted supplement for the climate change group was M$ 25.9.

One reason for the group allocation is to give each smaller country access to the possibility of a larger amount than it would have had otherwise. This was not an entitlement; but the possibility of the larger grant could provide Category 2 countries with greater flexibility and a greater incentive - the possibility of a grant from each focal area large enough for a substantial viable project. The group countries share around 15% of all the focal funds.

There was an implicit assumption in design that many small countries would not access funds during every replenishment period, as their access to funds had been intermittent in the past. Therefore there would not be much risk that the group allocation would run out of cash because all countries would submit projects for the maximum ceilings[42].

The *RAF document* is ambiguous in some important respects in its explanation of how the group country allocations are to be calculated[43]. Each country in the group allocation is limited in the funds for which it can apply for in two ways. First, it cannot apply for more than its standard maximum allocation. Second, there is a limit on the total funds available to the group. The mid-term review also looked at an alternative scenario for group allocation countries in the future, with only one constraint on their grant applications. See Box 3.1 for scenarios for group allocation countries. This situation seems unnecessarily complex. It is clear that the RAF needs simplification in regard to the group allocation for smaller countries.

Box 3.1: Group allocation scenarios

Current situation. In the two-constraint scenario, the small countries in the group allocation bear the full costs of all exclusions, without the larger countries bearing a share of the costs of exclusions for regional and global projects and small grants. In this scenario the Category 2 country allocations are constrained in a way that the Category 1 allocations are not. In Category 1 the sum of the individual country allocations is the same as the total allocation to the group. In Category 2, in contrast, the sum of the country maximums is larger than the total allocation to the Group. That is, all Category 2 countries could not receive approval for grants that are each equal to their maximum without the total exceeding the limit of the funds available to Category 2.

Case 1: Competition for higher amounts. To see the implications of this second scenario, consider the situation of a small Category 2 country "Ruritania". Its preliminary individual allocation (which was not disclosed by GEF as is standard practice for Category 2 countries) was, say, $700,000, but it can apply for grants up to the limit of the largest indicative allocation on any country in the group allocation, say $2.9 million. However the group allocation could run out of funds if other countries are quicker than Ruritania to apply and receive approval for grants. In that case Ruritania would have access to less than $2.9 million, but how much less will depend on how much money is left in the pot at the time.

Case 2: The 50% rule. There are also constraints on the timing of grants. Ruritania may access 50% of $2.9 million in the first two years ($1.45 million). Alternatively, if the other group allocation countries were quicker off the mark, total grants to all Category 2 countries in the first two years might begin to exhaust the actual funds for Category 2 in that period. In that case Ruritania would be able to access less than its full 50% of $2.9 million. In short, if Ruritania moves quickly it could get grants up to $1.45 million in the first two years of GEF4 and if it does not it might be limited to $500,000, or something in between. If it is slow to apply for a grant and others in the category are fast it might in fact be able to access nothing.

Case 3: Moving up to individual allocation. What happens to Ruritania if it "graduates" into indicative allocations of Category 1 at mid-cycle because its performance improves? Probably not much. It already had potential access to funds equal to the top of Category 2 (which is the same as the bottom of the group allocation). This happened to Suriname in biodiversity; it received 3.6 million US$ in allocation for the four years at mid-term (though it could in principle have accessed 3.5 million US$ under the group). In climate change, Croatia and Turkmenistan obtained 3.250 M US$, while Serbia and Tunisia were allocated US$ 3.5 M at midpoint, higher than the 3.1 M US$ group maximum. None of these countries had managed to access anything in the first half of the RAF. If a country in this situation has prepared to ask for 1 M, it would now have to turn around to add a smaller project or expand its current proposal. If a country had accessed group in the first part of RAF, with limits of 50% to the group allocation of about US$ 1.5 M, it would now find itself left with another 1.5-2 million US$ or so, for another small project.

Case 4: Alternative scenario: One constraint. One alternative could be to limit the constraint for group allocation countries to the individual limit equal to the largest hypothetical allocation of the top scored country in the group. If the effective demand of group countries did in fact exceed the group's total allocation, they could use un-utilized funds from the indicative allocation countries.

The mid-term review Delphi panel on performance exhibited mixed viewpoints on these issues. Participant responses did not indicate a consensus on the advisability of exclusions, the group allocation and targeted supplements. In response to the question, "To what extent does the exclusion of some resources impair the achievement of GEF objectives?" the average rating was 4.47 out of a possible rating of 10, with a relatively high 2.4 standard deviation. In their comments, several of the participants expressed their view that as much money as possible should be put through the formula. Some participants also noted these processes are complicated and difficult to understand.

On the other hand, a number of participants expressed the view that the GEF has multiple objectives for these programs, and that not all of those objectives are intended to be met by the RAF. Such participants supported the need to build capacity, deal with multi-country problems, and generally to level the playing field so all countries can compete for funding.

Ceilings

This section considers whether a limit (ceiling) should be placed on the maximum allocation that a country can receive in a single allocation period. Ceilings are limits on country allocations, to ensure a more equal distribution of funds. The "capped" amounts are smaller than what would have been provided to the countries if they received their full allocation as per the formula.

In GEF-4, GEF limited its individual member countries to a maximum allocation of 10% of the total funds available in the biodiversity focal area, and 15% of the funds available in climate change. Other organizations impose ceilings as well. These ceilings are formulated in various ways. For example:

- IFAD has a ceiling of 5% of its total funds that can be allocated to any one country during a single replenishment period.

- The World Bank (IDA) limits the allocation of any individual country to a maximum of special drawing rights (SDR) 19.8 per capita and, in IDA 15, two credit-worthy blend countries (countries with both grants and loans) were capped below IDA's operational cut-off (India and Pakistan).

- The Asian Development Fund operates what it calls a "soft cap", whereby blend countries whose individual allocation is greater than 14% of the total funds available receive only half of the allocation above that threshold.

- The African Development Bank has a maximum allocation of 10% of the total AfDB funds (which effectively means a cap on allocations to African countries with the largest population, such as Nigeria).

- The IDB Fund for Special Operations (FSO) limits individual allocations to a maximum of US$54 million per year. The CDB has a fixed allocation for Haiti, a country that would receive by far the lion's share of that Bank's concessionary funding if allocations were strictly according to the PBA formula.

Since ceilings are often used as methods to prevent especially large-population or low-income countries from being granted major proportions of the total available funding, some may argue that ceilings are a move away from maximizing potential benefits and performance.[44] On the other hand, there may be three reasons for establishing ceilings: (1) political equity and mandate concerns to protect the interests of small countries; (2) practical concerns about absorptive capacity in countries that may be given the largest allocations; and (3) concerns about marginal return to scale of investment.

When there are decreasing returns to scale in investment a ceiling may make sense. To illustrate, consider three countries that are each striving to save an endangered species, say the South American condor. Let us suppose that each country, given $1 million, can save 20 condors, for a total of 60 condors. A ceiling cap ensures that each country gets $1 million and 60 condors are saved. If we remove the cap let us suppose that one country is now allocated by formula the whole $3 million but with $3 million can only save 24 condors. This type of case is interesting; but rare. In most cases, in both biodiversity and climate change, decreasing returns to scale of investment are probably real but small.

Ceilings may have practical advantages. Without ceilings, for example, a very small country that performs very well might receive an allocation that exceeds its absorptive capacity. Similarly a large country might receive such a large allocation as to crowd out access to resources by some smaller countries entirely – or at least reduce their allocations below the threshold size for a single viable project.[45] This would be particularly problematic for organizations that have a mandate to pay particular attention to the needs of small or low-income countries.

The choice of a ceiling depends on the degree to which allocations would be skewed towards large countries if the ceiling did not exist. In the GEF biodiversity focal area, for example, no country had received grants historically that exceeded about 4% of the total funds available. Therefore there was no pressing need for a ceiling. Nevertheless a ceiling was set at 10%, with no effect on allocations.

In contrast, in the GEF climate change area, a single country (China) had historically received as much as 17% of total funds. A climate change ceiling of 15% of total funds was established. In this case the shift for the largest-allocation country was substantial; without the ceiling China would have received US$ 224 million and with the ceiling it received US$ 150 million, a reduction of 33 %.

Furthermore, the GEF ceiling is applied to *total* focal area funds including exclusion (1 billion US$), not as a ceiling on country funds (900 million US$), *after* the exclusion. If the latter were the case, China would have received 135M US$ in climate change, not 150 M US$.

Depending on their placement in the overall ranking and the proportion of the total funding they account for, different countries' sensitivity to the ceilings vary. For example, if the biodiversity ceiling had been set at 5% in GEF 4, the allocations of two countries (Brazil and Mexico) would have been reduced by 21% and 8% respectively. The increment would have been redistributed among other countries in proportion to their base allocations. Since the increment would have been only 2.4% of the total monies available in the biodiversity resource envelope, the impact of the ceiling would have been small on countries other than the two directly affected.

In contrast, because a higher proportion of total funding in climate change goes to a few large-allocation countries, a ceiling of 5% in the climate change area would have resulted in a redistribution of about 28% of the total resources in this area. China's allocation would have fallen by 67%, India's by 33% and Russia's by 31%. As a secondary effect, five countries (Ecuador, Tunisia, Turkmenistan, Croatia and Madagascar) would have had individual allocations rather than being in the general pool of small countries.

We might consider an intermediate position. If the climate change ceiling for any individual country had been 10% of total funds, approximately 7.4% of total funds in this focal area would have been redistributed to other countries with modest but discernable affects on their allocations. Countries that would have lost funds would have included China (- 67 %), India (-33 %) and Russia (-31%). Two more countries would have been added to the indicative countries.

Establishing the right level of ceiling is ultimately a strategic decision, but can have notable effect on the distribution of funds. In simulations of alternative levels of country caps, it is found that *lower* ceilings mainly benefit the individual allocation counties. An argument for lowering ceilings in the climate change area is that, at present, the countries with largest GHG emissions are receiving large GEF funds. In addition, the current formula addresses only the potential costs of climate change mitigation without addressing adaptation to the effects of climate change. This suggests the possible utility of lower ceilings, perhaps combined with reforms in the nature of the allocation formula, in the climate change area. The current system is focused on the cost of adopting cleaner technologies for mitigation purposes, and not adapting to climate change.

Floors

Floors aim to secure countries a minimum amount, and avoid allocations of tiny amounts that are not practicable to program. The RAF document specifies that "for each country whose *preliminary country allocation* is less than $1 million, a targeted supplement will be provided so that the country will have a minimum *adjusted allocation* of $1 million". This is in effect a 'floor' in allocations. In practical application, this "supplement" becomes part of the group allocation pool, to be shared with other group countries that were under or above the 1M US$ floor. There is no guaranteed minimum amount. In biodiversity, thirty countries are subject to the floor, in climate change 41 countries.

There are three different scenarios with regard to floors. First, if a floor is set lower than M$1, the number of indicative countries falls below 57. The number in the group allocation grows and the group ceiling becomes higher. The overall allocation to the indicative countries increases.

Second, in the extreme case without *any* floor, the allocations are equivalent to the country's preliminary allocation (not made public), however small. No supplement is needed to raise the group countries to M$1, and the 75% cutoff point moves up. The number of the indicative countries decreases from 57 to 51 in biodiversity and from 46 to 31 in climate change.

Third, if a RAF floor in either focal area is set at a level higher than M$1, the number of indicative countries increases, the group shrinks, the group ceiling falls and the overall allocation to the indicative countries decreases. For example, with a floor at 2 million US$, the number of the indicative countries would rise to 75 in biodiversity and to 86 in climate change.

These changes may perhaps seem counterintuitive, in that if the floor were higher, the supplementary funding needed would normally increase and there would be less money for individual country allocations. The pattern above is mainly because of the rule that 75% of resources go to indicative allocations for the top-scored countries. The more funds available, the more goes to the top 75%. The floors help distribute resources across countries more than would be the case without floors.

4 The RAF and Recognizing Country Achievements

This section discusses the way that the current design of the RAF recognizes countries' achievements. The key question is: "To what extent does the RAF provide incentives for countries to improve their

performance, in each focal area?" The mandate of this mid-term review is, in part, to analyze whether the incentive aspects of the RAF formula are reasonable and appropriate, given its objectives.

Resource allocation systems like the RAF have several objectives. One is to address needs and potential benefits. Another is to recognize good performance. By "performance"[46] GEF means (a) project performance and (b) policy/institutional performance. Both are important; but the longer-term effects on policy/institutional performance are potentially the most important because the whole country would be affected by improved policies and institutions. Recognition of this is partly intended to place grants or loans where they are likely to be effective; and partly to provide an incentive for improved performance. This means favoring governments that have both the opportunity to produce global benefits and the capability of doing so. They have demonstrated that they can use funding well (measured by their portfolio performance, PPI), and that they have policies and institutions in place that ensure country-wide effectiveness (measured by their policy and institutional performance, CEPIA and BFI). The focus is on the size of each member country's allocation.

In the longer term, the aim is to recognize that member countries have improved their practices so that their RAF scores improve and this improves their access to grants. The question is how much a country's RAF score (and therefore its allocation) changes when its practices change. The incentive depends on how much recognition a country would realistically get when it has improved its practices. Of course it also depends on the government understanding the link between its performance, its scores and the grants it receives.

It is useful to be as precise as possible what "providing incentives" means in the RAF context. Incentives are partly a matter of how much money is at stake. There is a wide range of RAF allocations. It is reasonable to assume that larger allocations receive more attention from their recipient governments and that larger GEF budgets would exert greater influence. (See technical paper #4 on this.) This section asks whether, by instituting the RAF, the GEF has obtained an incentive effect for its money. Does the RAF formula reward good performance rather than accommodating poor performance? If it does reward good performance, is this effect clear enough to be communicated by the GEF in policy dialogues?

A resource allocation framework can aim to reward performance at many levels – and with different timeframes - by national policies and institutions (CPIA and BFI), of ongoing projects (PPI) – and in producing global environmental benefits. The increases in GBI will naturally be more long-term, so that the shorter term perspective to improve scores is through the performance index.

The relative weight of "performance" in various allocation formulae

The effects of incentive weights are often not easy to understand because of the complexity of the allocation formulae used by MDBs and funds. Most organizations[47], including GEF, use a complex type of formula that contains several variables, each variable raised to a power (exponent). The weighted scores on the variables are multiplied together to give a country score and, in general, resources are allocated in proportion to country scores. Nevertheless, the effective weight of "performance" is difficult to calculate. It depends on the number of variables, the exponent on each, and the nature and variability of the underlying data. However, to simplify, a larger exponent indicates a greater weight to that variable.

Table 8 shows the various allocation formulae and the variety of variables and weights in those formulae. The balance between the weight of "needs/potential" variables and "performance" variables is particularly pertinent to incentives. Any variable can have incentive (or disincentive) effects but the performance variable is the one normally designed with the closest eye to providing incentives.

Table 8: Variables[48] in Resource Allocation Formulae

Institution	Allocation Formula "Needs and Potential" Factors		Performance Factors	Result
AfDB	$POP^{1.0}$ x $GNPPC^{-0.125}$	x	$(0.26CPIA_{A-C} + 0.58\ CPIA_{D} + 0.16PORT)^{4.0}$	= allocation share
AsDB (AFD)	$POP^{0.6}$ x $GNPPC^{-0.125}$	x	$[(ES\text{-}CPIA^{0.7}$ x $PORT^{0.3}) \times GOV]^{2.0}$	= allocation share
CDB	$LogPOP \times GNPPC^{0.9} \times VUL^{2.0}$	x	$(0.7\ CPIA + 0.3\ PORT)^{2.0}$	= allocation share
EC (ACP)	$LogPOP \times 0.2\ GNPPC^{-1.0} \times 0.2\ HDI^{-1.0} \times DEBT \times VUL$			= allocation share
GEF	$Benefits\ Index\ Number^{0.8}$	x	$[0.2CPIA + 0.10PORT + 0.70CPIA]^{1.0}$	= allocation share
IDB (FSO)[49]	(0.22FUND x POPSHARE) + (0.133FUND x GNPSHARE)	+	(0.6FUND x PERFORMANCESCORESHARE)	= $ allocation
IFAD	$POP^{0.75}$ x $GNPPC^{-0.125}$	x	$(0.2CPIA + 0.35PORT + 0.45RuralCPIA)^{2.0}$	= allocation share
World Bank (IDA)	$POP^{1.0}$ x $GNPPC^{-0.125}$	x	$(0.24CPIA_{A-C} + 0.68CPIA_{D} + 0.08\ PORT)^{5.0}$	= allocation share

Source: Inter-Organizational Technical Meeting on Performance-Based Allocation, IFAD, Rome, 2008

The reasons for different weights in the allocation formulae

The World Bank (IDA) was the first of the IFIs and Funds to adopt a formal PBA, with a rules-based allocation formula. Its use dates back to 1977, although during the first decades it was confidential with access restricted to management. Towards the end of the 1990s the IDA Board became increasingly interested in two things – first, the importance of the recipient country's policy and institutional context in determining whether aid would be effective; and, second, the importance of providing positive incentives (and avoiding perverse incentives) when providing aid. From 1999 to 2007 all of the IFIs and Funds shown above in Table 4.4.1 adopted the concept of a performance-based allocation system for concessionary funds, generally during replenishment negotiations. The designers of the new systems, and the Boards or committees that approved the systems, had several objectives in mind:

- To provide a strong incentive for improved performance.
- To avoid radical shifts in traditional levels of allocations.
- To harmonize approaches with other IFIs and Funds.
- To express the special values and priorities of the particular organization.

Of course not all of these objectives can be maximized at the same time. The result was variations among the allocation formulae, but basic similarities as well. All formulae contain at least one "needs/potential" variable. In many cases population (POP)[50] and gross national income per capita (GNIPC) are surrogates for "need and potential". As one can see in Table 4.4.1, the exponents on POP vary from about 0.6 to 1.0. The smaller exponents on POP are more advantageous to smaller countries. GEF's choice of 0.8 for its "need/potential variable" (GBI) puts it squarely in the middle of this customary range for the "needs/potential" variable.

The weight given to performance

As noted above one cannot easily tell from a formula how much weight is given to performance because it requires complex calculations of elasticities. However, as a rule of thumb, other organizations, including IDA[51], IFAD and the regional banks, have generally sought to have about 60% of the variance of the country allocations determined by the performance variables in the formula. The exponents on variables in the GEF RAF formula are approximately in balance with this idea. Performance is more heavily weighted than "need/potential" but not by a lot; country score = $GBI^{0.8}$ x $GPI^{1.0}$.

To explore the influence of performance in the RAF the mid-term review undertook computer simulations to show what happens to country allocations when the weight of GPI vary in each focal area. The RAF formula produces very different incentive effects in the two focal areas. In one (biodiversity) it produces

modest but clear incentives; in the other (climate change) it does not. To show this, simulations were undertaken to explore what happens to allocations when more weight is given to GBI or more weight to GPI, in each focal area (simulation tables in statistical annex).

Scenario A. What happens if the weight of GBI (emissions and change in carbon intensity) is increased in the Climate Change focal area? In climate change, the RAF formula results in a high concentration of allocated monies in the top fifth (quintile) of member countries. In GEF-4 the top quintile by GBI was allocated 76% of climate change funds. The bottom quintile was allocated only 3%. The resource concentration coefficient (RCC)[52] is 23.5 to 1. When increasing the GBI weight, nothing much happens. Climate change resources are already very concentrated in the top fifth of countries. Increasing the weight of GBI in the allocation formula strengthens this pattern a little while, leaving the fundamental concentration on a few countries undisturbed. See table 9.

Table 9: Climate Change quintile shares of funds as the weight of GBI increases

GBI Quintiles	The GBI Exponent							
	0.80	0.88	0.97	1.06	1.17	1.29	1.42	1.56
1	76%	78%	80%	82%	84%	85%	85%	85%
2	11%	9%	8%	7%	6%	5%	4%	4%
3	6%	5%	4%	4%	4%	4%	4%	4%
4	4%	4%	4%	4%	4%	4%	4%	4%
5	3%	3%	3%	3%	3%	3%	3%	3%
Total:	100%	100%	100%	100%	100%	100%	100%	100%
RCC	23.5	24.3	25.0	25.6	26.0	26.2	26.4	26.5

Scenario B. What happens if the weight of GPI (policy and institutional performance, and portfolio performance) is increased in the Climate Change focal area? The resource allocations for climate change are not much sensitive to changes in country performance unless the weight of GPI in the allocation formula is greatly increased. The pattern of allocations is not much affected if the weight of GPI is increased modestly by 10% seven times. The top quintile of countries by performance increases its share from 44% to 48%, and the resource concentration ratio rises from 11.2 to 13.2. If the weight were increased to 5.0 then the top quintile of countries by performance would receive a considerably higher proportion of allocations.

In summary, the GEF climate change allocations are not very responsive to changes in country performance. Part of the reason for this is that China's allocation does not change (since it is at the ceiling already and no amount of improvement is going to push its allocation past that ceiling). Delphi climate change experts agreed that the RAF did not provide adequate incentives for countries to improve their mitigation performance (average 4.4 on a scale 1-10, standard deviation 2.2). 34% thought the RAF did to a limited extent. Participants thought that incentives were given for those countries where fossil fuel use and cement production are dominant in the economy. In other countries there is less recognition of achievements.

Scenario C. What happens if the weight of GBI is increased in the Biodiversity focal area? Like climate change, biodiversity allocations are heavily concentrated in the top quintile and only modestly responsive to increases in the weight of GBI in the allocation formula. It makes little difference to the three bottom quintiles, essentially because many countries in the bottom quintiles have "fixed allocations" at a minimum indicative $1 million. This does not change for most of them as the weight of GBI

changes. In summary, increasing the weight of GBI increases the recognition of achievements of countries to improve biodiversity, where that is possible.

Scenario D. What happens if the weight of GPI (policy and institutional performance, and portfolio performance) is increased in the Biodiversity focal area? In the biodiversity focal area, allocations are considerably less concentrated in the upper quintiles of performance than was the case with dimate change. In biodiversity the resource concentration ratios are in the range 2.7 to 10.7, whereas in climate change they were in the range of about 10.6 to 18.1. Modest changes in the weight of GPI make only a small difference to allocations. Larger changes to the weight of performance do accumulate to substantial differences.

In summary, incentives in the climate change area need improvement. Although the format of the allocation formula for GBI and GPI is the same for climate change and biodiversity the incentives provided to countries in each focal area are different in some respects. In biodiversity, the incentives are positive. In climate change, where one might have expected the greatest scope for incentives because there is wide scope for policy interventions to lessen emissions, the incentive effects of the allocation formula are weak or negative. In effect the largest emitters receive the largest grants, without a balancing reward for improvement.

The importance of portfolio performance (PPI)

The Portfolio Performance Indicator (PPI) measures each country's average performance in environmental projects over the past ten years. It gives equal weight to two things: (a) the average of GEF project ratings contained in the Project Implementation Review, and (b) the average World Bank Independent Evaluation Group (IEG) rating of environment-related completed projects.

All other organizations that use a performance-based allocation system have a formula that includes a measure of portfolio performance. However it has tended to be a controversial variable. Its weight varies a great deal from one formula to another, for example from 8% (IDA) to 35% (IFAD). The weight GEF gives to project performance is at the lower end of the range (10%).

There are arguments for and against a high weight for PPI. Some believe that PPI is worth considerable weight because it is a good indicator of likely performance under a new grant, is relatively objective, and provides an incentive for performing well on GEF projects. Others believe that it is unwise to give much weight to PPI because project performance is a result of many factors including donor behavior. Average project performance can also be somewhat open to manipulation. For example, consider a country that has only two projects, one well performing and one poorly performing. If it closes the poorly performing project early its portfolio now contains only the well performing project and its PPI score, say, doubles. Scores can be unstable.

Another problem with PPI is that scores can be *too* stable. If one takes a very long-term view (as GEF does with its ten year perspective on portfolio performance) then the PPI can be out-of-date and slow to change. If it is, there might be little incentive for a country to try to improve it.

Since the PPI is the only part of the performance factor that is partly based directly on GEF judgments (the other GPI performance data is lifted from other organizations that calculate their own scores for policy and performance assessments) there is perhaps an additional argument for giving more weight to PPI. However there is inevitably some tradeoff between stability (by being averaged over a long period of time) and responsiveness and predictive accuracy (by emphasizing recent performance).

Do the different measurement scales of different variables affect the allocation outcomes?

The variables in the allocation formula are measured on different scales. Some are on a scale of 1 to 6, and some on much broader scales – say, 1 to 100. This has no effect on the allocation outcomes as long as scoring is true to scale, because it is relative scores not absolute scores that determine the allocations among countries. That is, getting 1 out of 5 or 20 out of 100 makes no difference as long as all countries are scored on the same scale - their relative scores will be unchanged and therefore their allocations will be unchanged. However if scoring is *not* true to scale (say the average score changes from '4 out of 5' to '5 out of 6' when the scale is changed) then there is a problem. No such issues have arisen thus far in the experience of the RAF. The main difference is that the scores in the GBI are more unevenly distributed as compared to the GPI.

The GBI variable

The formulae used by performance-based allocation systems all have two components. The first component measures "potential and needs"; this shows the scale of the problem to address. Countries that have many very poor people, or high carbon emissions, or a great deal of biodiversity that is under threat, receive a higher allocation. The scale of the problem determines the scale of funding. (Of course the scale of funding is then modified by a performance factor, as seen in the section above.

The "scale" variable often includes measures of population and income (GNP per capita). It can also include other variables meant to signal the scale of the problem. For example, the Caribbean Development Bank uses a measure of vulnerability to natural disasters and external economic shocks (VUL) and the European Commission's Africa, Caribbean and Pacific Program uses five variables to indicate the scale of "potential and needs". These include POP, $GNP_{\text{per capita}}$, vulnerability, indebtedness and rating on the U.N. Human Development Index.

In its biodiversity focal area, GEF measures "potential and needs" by a combination of terrestrial data (represented and threatened species) and marine data (represented species). In its climate change area, GEF measures its scale variable, "potential and needs", by data on greenhouse gas emissions.[53]

Do problems at a larger scale indicate cost effective opportunities for intervention?

It is sometimes argued that the scale variable also indicates where the most cost effective opportunities are for intervention. However this confuses two ideas. The first idea is that there are often diminishing returns to investment in any given set of potential investments. Any government faces a set of opportunities for carbon emission abatement, for example, and it is well advised to pursue those opportunities first that produce the greatest bang for the buck (more formally, the greatest marginal benefit for a marginal unit of investment). The second idea is that larger countries, having larger problems, will have better investment opportunities – that is, that investments at the margin in larger countries will have better marginal returns than in smaller countries. This is not necessarily correct, as it is not possible to generalize by country size where the best marginal returns to investment in carbon emission abatement will be, for instance.

In fact the rationale for giving a "scale" variable considerable weight in the allocation formula is implicitly that the safe assumption is 'constant returns to scale'. If a country has twice as much carbon emission, it will take twice as much investment to fix the problem, and the benefits of doing so will be twice as large. Of course this is not necessarily so either; but in the absence of specific information to the contrary it is the best assumption, more sensible and fair to countries regardless of size, than contrary assumptions.

The scale variable should have a separate weight

Ideally the "scale" variable should have a single weight. If the scale variable has several component factors, then these factors should not receive individual weights. They should be combined first and then given a single weight. When "scale" has a single visible weight and "performance" has a single visible weight, the task of assigning weights is much easier and more intuitive. This is not a trivial point because all of the IFIs and Funds that use an allocation formula have a history of Boards and committees changing the weights of various factors in the formula without a steady grasp of the mathematical implications of their choices. Best practice in this regard is the simple arithmetic weights used throughout their formula by the Inter-American Development Bank.[54]

The GEF comes close to best practice by having a single scale factor (GBI) and a single performance factor (GPI). Each has a weight and the two weights add to 100%. In the biodiversity focal area this works well. In the climate change focal area, however, the scale factor (greenhouse gas emissions) is multiplied by a performance factor (change in the carbon intensity of the economy) before a weight is applied. This makes it difficult to grasp just how much weight overall is being applied to the scale factor and how much to performance factors.

Summary of findings on performance weights

Delphi experts on performance were not confident about the impact of both the **CEPIA and the Broad Framework Indicator (BFI)** concerning the enabling environment in providing an incentive to countries to improve their performance in the future (average 4.6 for CEPIA and 4 for BFI, standard deviation 2.6 and 2.1 with 10 being 'to a large extent'). In comments, participants cited the relatively modest level of GEF funding as one reason why the performance impact may not be great. Participants also mentioned the need for a clear difference in funding between well and less well performing countries; and the need to publicize performance results and use them in policy dialogues.

As with the CEPIA and the BFI, Delphi participants do not believe that either the **PIR** or the World Bank final project evaluation performance scores provide a strong incentive to improve performance in the future (average response 4.90 for PIRs, 4.20 for **ICRs,** on a scale 1-10). Operational focal points noted that ratings for regional projects would not be taken into account, of concern to those countries that have mainly benefited from regional GEF funds.

Some key assumptions underlying the use of the GEF Performance Index (GPI) include:

- Linking country performance to resources, along with other International Financial Institutions (IFIs), and making performance scores visible and transparent, will provide a significant incentive to improve performance.
- The process of scoring country performance is tractable. For example, if country performance, either portfolio performance or policy/institutional performance, were unstable and inconsistent, it might be impossible for governments to see the link between their performance and the level of GEF funds made available to them; and there might be insurmountable practical difficulties in managing unpredictable country program budgets.

When the RAF was first presented to GEF country focal points in 2006, concerns were expressed regarding the apparent disincentives being created by RAF, and that RAF "does not include any mechanism or recommendation as to how a country may improve its performance and country allocation in the future" (Dakar subregional report, April 2006). These concerns seem to be materialized in implementation.

5 Synergies and inter-relationships

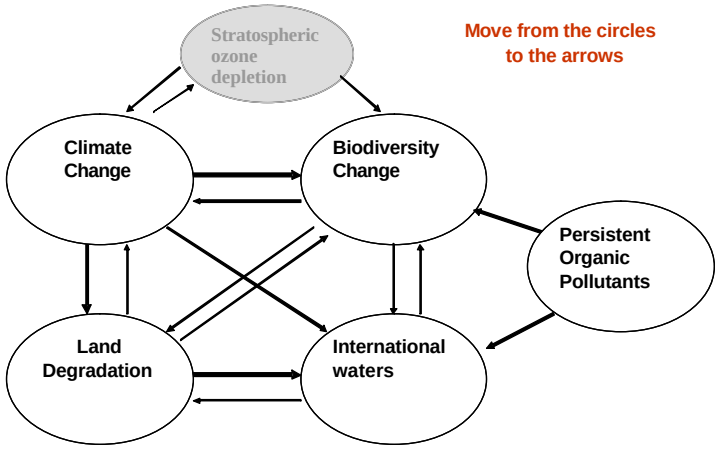
This section covers an evaluative assessment of relationships among and between the indices, based on key question on "how the RAF provides opportunities for synergies between climate and biodiversity work", or with other focal areas.

Synergies occur when two or more discrete influences or agents acting together create an effect greater than the separate effects of the individual agents[55]. Under the RAF, synergies can emerge both in the index design and in implementation. As the GEF Scientific and Technical Advisory Panel (STAP) pointed out to the Council: "There has been an emerging recognition of important links between various global environmental issues such as loss of biodiversity, climate change, land freshwater and coastal systems degradation. From the GEF perspective, an additional challenge is to ensure that these interlinkages are properly reflected in the design and implementation of its projects.[56]"

There is no positive assessment of synergies *before* the RAF. So far, GEF has particularly pursued synergies through operational program 12: *Integrated ecosystem management*, which is intended to be *multifocal*, dealing with two or more focal areas; and *synergistic*, where achievement of benefits in one focal area leads to increased benefits in another.[57] A review of the GEF OP12: Integrated ecosystem management (IEM) by the GEF Evaluation Office (2005), concluded that "very few projects convincingly presented potential synergies among focal areas". Issues raised included careful assessment of "win-wins" versus trade-offs; risks that holding projects responsible for multifocal outcomes could be beyond project capabilities and budgets; lack of strategic guidance of the operational program; and unclear guidelines for designing and achieving successful IEM projects. The review also found that no project proposed a convincing model to measure synergies between or among focal areas, and that "It was implicit in many documents that, because there were two focal areas written into the project, synergies would automatically occur". More than 52 percent of projects reviewed scored moderately unsatisfactory or less for synergy, while only a little over 25 percent scored satisfactory or better and only 5 percent scored highly satisfactory. A STAP review concurs that "…GEF project documents do not reveal evidence of a systematic approach to incorporating these [linkages] explicitly in project design".

Synergies can be particularly important for smaller states. STAP "recognizes that the GEF has been active in SIDS through all of its focal areas. However, [..] the range of GEF-assisted activities may be more effective if they are better linked in concept and in project interventions, and through activities on the ground in any such State[58]".

Fig. 7: A possible framework for interlinkages



The STAP developed a design tool for focal area interlinkages (see **Figure** 7). Climate change largely acts on the other areas of interest (arrows flowing out), whereas biodiversity is affected by other factors (arrows flowing in).

There was no clear goal or assumption that the RAF would lead to synergies. By itself, RAF *design* does not ensure synergies. The RAF was not assigned goals to promote synergies, and by its firm distinction between climate change and biodiversity focal area funding, does not ease work between the two areas.

According to the framework in figure 7, areas that are important for synergies, such as adaptation, carbon sequestration, CDM and biomass are not covered in RAF design. It is, however, not apparent how these aspects might be reflected in indices. For example, global environmental benefits in terms of climate change can accrue from the sequestration of carbon and from the reduction of greenhouse gases. Carbon sequestration is achieved from maintaining or increasing plant biomass through changes in land use and management. A country that loses carbon stocks through burning forest vegetation, would through the rationale of the RAF index merit more funding in climate change (higher emissions) – but less in biodiversity (threat to ecosystems, reduced forest cover). On linkages in climate change, "It is very difficult to demonstrate globally significant climate change benefits by increasing soil carbon stocks through land use change" (EO OP12 review).

Projects funded under the RAF must correspond to the GEF-4 focal area strategies. These strategies do not explicitly aim for links or synergies. In the GEF-4 climate change focal area strategy, the only mention of links is the *Strategic Program 6: Management of Land Use, Land-Use Change and Forestry (LULUCF) as a Means to Protect Carbon Stocks and Reduce GHG Emissions*[59]. The program aims to define a methodology for estimating avoided GHG emissions as a result of GEF support. The tools to be developed will link forest and satellite data to better monitor progress to reduce emissions from undesirable land-use changes. Resources may be devoted from biodiversity, climate change GRE, land degradation focal area, and country allocations.

In the biodiversity GEF-4 focal area strategy, links and synergies are not mentioned explicitly in the focal area strategy, but may occur in new strategic programs of sustainable forest management framework strategy "Management of LULUCF as a Means to Protect Carbon Stocks and Reduce GHG Emissions"; and "Promoting Sustainable Energy Production from Biomass."

In implementation, synergies under GEF-4 are not linked directly to the RAF. Synergies are more likely to emerge from implementation. Under GEF-4, there has been a growth in programmatic approaches as well as multifocal approaches (see technical paper #4). In part multifocal projects are now established to overcome rigid walls between limited focal area funding under RAF. There are some examples of countries, with a smaller or group allocation, obliged to change the pipeline, whereby projects are 'merged' to fit allocations without synergy as a primary objective. At the biodiversity COP-9, the Africa representative pointed out that "multi-focal area funding [..] may constitute a risk for biodiversity in that such activities may be diluted". A STAP review of PIFs in the November 2007 work program came across only one project (of 50), in international waters, that would develop novel forms of intervention and linkages to deliver global environmental benefits[60].

The GEF also puts emphasis on synergies with other activities at the country level, by financing incremental costs of global environmental benefits. The 'principle' of incremental cost was originally envisaged to ensure that GEF funds do not substitute for existing development finance. In response to a key evaluative question, the RAF design does not take the **actions of governments and other donors** on global environmental benefits into account. Obviously, improvement in the RAF indices, both for benefits or performance, would stem from support and actions taken by the government, its development partners, and other national stakeholders, but this is not possible to capture in an index.

In implementation, the improved predictability of RAF funding may in principle facilitate joint programming, but the fixed amounts for smaller countries may also make this more challenging. In such cases, the RAF de facto fixes limits on the increment for the country to baseline that government and other donors finance. Delphi biodiversity experts supported the view that the costs of biodiversity conservation and/or sustainable use should be taken into account, to encourage greater conservation of biodiversity and/or sustainable use of biological resources. However, this does not indicate if more funds

should be given to countries managing biodiversity efficiently; or to countries with high biodiversity but also high costs.

Few examples were found of key environmental donors being involved in RAF pipeline priority setting at the country level. The GEF does not have specific mechanisms, such as country programs, for donor or stakeholder consultation at national level. At the project level, cooperation continues as before, though there is some indication that the RAF puts more time pressure on project proponents and may make it more difficult to work with cofinancing and other development actions in a synergistic manner. For example, UNIDO interested the Swiss Cooperation (SECO) in South Africa in an energy efficiency project for co-funding. After back and forth related to RAF pipeline prioritization, it was decided to proceed only with SECO with 10M US$ in contribution and drop the GEF involvement.

Synergies between indices and programming

Conceptually, it might also be possible to promote synergies between the index design and project design. Currently, there is no clear relationship between the information used to construct the GBI (and thus determine funding allocations) and the expected benefits of specific proposed projects in the focal area (use of allocation). Several different stakeholders indicated to the mid-term review that "allocations do not seem to translate into projects for which the money was allocated in the first place". Certainly this is not helped by the fact that the underlying indicators data have not been disclosed. Other PBAs use their assessments for dialogue with select countries on relative weaknesses for attention in policy and programming. The GEF has put considerable investment into its indices and data development. The lack of policy dialogue and the lack of knowledge of the indicator data represent a lost opportunity for better targeting and effectiveness.

In biodiversity, most indicators provide a score on both representation and threat – for ecosystems, mammals, birds, and amphibians. The index consists of many indicators and weights, so that it is difficult for one or few indicators to dominate the scoring. The underlying data has not been disclosed, so that it is not possible for countries and Agencies to develop projects that are consistent with the threat or representation for which the country obtained its allocation. For example, see the annex with the relative ranking of the top 25 countries per biodiversity RAF indicator, as well as allocations. Put simply, a country may derive much of its allocation from amphibians and their threat – yet spend its allocation on another species, or on other GEF priorities not directly related to this aspect.

In climate change, the link is less obvious. Delphi respondents' average opinion was that it would be useful, to a limited extent, to use a country's GBIcc in guiding the selection of specific climate change projects. Comments indicated that "The existing GBIcc methodology is too narrow [….]" ; "the only criteria to judge specific mitigation projects would be to see if a project likely lead to a reduction in the country's overall emissions (Baseline GHG emissions 2000); and would a project likely lead to a reduction in the carbon intensity of the country".

Biodiversity Delphi Participants were neutral on whether the information contained in a country's GBI_{BIO} is relevant for guiding the selection of biodiversity projects (average 4.7 on a scale 1-10, standard deviation 1.8). Experts did not agree on the extent to which extent using indicative allocations influences the coherence of GEF biodiversity funding portfolios. There was higher agreement of taking cost of conserving biodiversity into account for greater conservation of biodiversity and/or sustainable use of biological resources (average 6.8, standard deviation of 2.6). This would point to a closer attention to a country's ability to support and maintain proposed biodiversity conservation, sustainable use delivery, or implementation of biosafety safeguards.

However, most Delphi climate change experts (62%) thought that the RAF should provide more opportunities for interactions between climate and biodiversity work, within a context of sustainable development. It was stated that incentives for avoided deforestation are needed on the national or sub-national scale. Other areas for synergies are land use change to prevent loss of biodiversity, climate change and adaptation capacity, as well as adaptation-mitigation interactions. It was noted by experts that the rationale of RAF is based on mitigation, while the linkages between climate and biodiversity are more obvious in adaptation.

Delphi participants agreed that good policies are the key to successful results. However, the linkages are not clear between country performance and GBI as described in the RAF indices. Delphi climate change participants found the scoring gives too much importance to good government policy. This type of scoring would mean that countries that are poorly scored for performance are less likely to receive much funding because they will have a low GPI; and a lower GBI if their energy intensity also rose significantly in the 1990s. The CEPIA measures environmental policy but that is much different from the climate change policy.

Obviously, there are different incentive structures in the RAF for indicative allocation and group allocation countries. Delphi participants were not convinced that using indicative allocations to guide funding decisions would affect the quality of project proposals. Indicative allocations may allow countries to plan better and give more leverage to negotiate, but if allocations are seen as entitlements, proponents may take less care in selecting and preparing proposals. The effect from crowding out of NGO and civil society proposals due to caps on funding is also a concern.

Annex: Relative ranking of the top 25 RAF countries per biodiversity RAF indicator

	Mammals	Mammal THREAT	Ambhibian	Ambibian THREAT	Birds	Bird THREAT	Reptile	Plants	Fish	Ecoregion	Ecoregion threat	Marine	Allocation
1	Indonesia	Indonesia	Brazil	Colombia	Indonesia	Brazil	Brazil	Brazil	Brazil	Mexico	Mexico	Indonesia	Brazil
2	Brazil	China	Colombia	Mexico	Brazil	Indonesia	Mexico	China	China	China	China	Mexico	Mexico
3	Mexico	Mexico	Mexico	Ecuador	Colombia	Colombia	China	Russia	Indonesia	Brazil	Brazil	Philippines	China
4	China	Brazil	Peru	Peru	Peru	Peru	Indonesia	Colombia	India	Canada	India	South Africa	Indonesia
5	Congo DR	Madagas.	Ecuador	Honduras	Philippines	Philippines	Madagas.	Indonesia	Colombia	Russia	Indonesia	Chile	Colombia
6	PNG	PNG	China	Venezuela	Ecuador	Mexico	Russia	Mexico	Peru	Indonesia	Russia	Brazil	India
7	Madagasr	India	Madagascr	Guatemala	China	Ecuador	India	Venezuela	Venezuela	India	Colombia	PNG	Russia
8	Peru	CongoDR	Indonesia	India	India	Australia	Colombia	Peru	Mexico	Colombia	Venezuela	India	Peru
9	India	Philippines	PNG	Haiti	Madagascr	Madagasr	Argentina	Madagasr	Zambia	Venezuela	Philippines	Russia	Madagascr
10	Argentina	Kenya	Venezuela	Cuba	Mexico	India	Peru	India	Russia	Argentina	S. Africa	China	Ecuador
11	Colombia	Cameroon	India	Costa Rica	PNG	China	Philippines	Bolivia	Cameroon	S.Africa	Argentina	Colombia	S. Africa
12	Russia	Malaysia	Malaysia	Madagascr	Tanzania	Cuba	Congo DR	Ecuador	Malawi	PNG	Kazakhstan	Oman	Philippines
13	Malaysia	Viet Nam	Congo, DR	China	Congo DR	Venezuela	S. Africa	Congo,DR	Thailand	Peru	Ecuador	Ecuador	Venezuela
14	Bolivia	Ethiopia	Costa Rica	Cameroon	Venezuela	Kenya	Ecuador	Malaysia	Myanmar	Myanmar	Turkey	Bahamas	Chile
15	Ethiopia	Tanzania	Cameroon	Panama	Bolivia	Tanzania	Venezuela	Philippines	Mozambique	Philippines	Madagascr	Panama	Malaysia
16	Philippines	Russia	Bolivia	Domin Rep	Argentina	Bolivia	Cuba	Argentina	Lao	Mongolia	Iran	Mozambique	Cuba
17	S. Africa	Peru	Tanzania	Tanzania	Russia	Congo DR	Tanzania	Cuba	Malaysia	Kazakhstan	Peru	Micronesia	Argentina
18	Myanmar	S. Africa	Guatemala	Argentina	Cuba	Russia	Malaysia	Chile	Viet Nam	Turkey	PNG	Seychelles	Tanzania
19	Venezuela	Colombia	Argentina	Jamaica	Fiji	Argentina	Iran,	Turkey	Congo	Ecuador	Nigeria	Cape Verde	PNG
20	Tanzania	Sri Lanka	Panama	Malaysia	S. Africa	Mauritius	PNG	PNG	Uganda	Tanzania	Tanzania	Madagascar	Costa Rica
21	Kenya	Sudan	S.Africa	S. Africa	Ethiopia	Cameroon	Sri Lanka	S.Africa	PNG	Pakistan	Viet Nam	Mauritius	Cameroon
22	Cameroon	Iran	Viet Nam	Bolivia	Kenya	Viet Nam	Sudan	Iran	Angola	Iran,	Myanmar	Peru	Bolivia
23	Sudan	Solomon I	Cuba	Indonesia	Viet Nam	Fiji	Bolivia	Guyana	Guyana	Afghanistan	Mongolia	Costa Rica	Panama
24	Angola	Myanmar	Honduras	Chile	Dom. Rep	S. Africa	Chile	Thailand	Ecuador	Madagascar	Cuba	Maldives	Vietnam
25	Ecuador	Bolivia	Philippines	Philippines	Solomon I	PNG	Somalia	Tanzania	Tanzania			Venezuela	Congo DR

Endnotes:

[1] For the exact language describing the index as approved, see the RAF Document (GEF/C.27/Inf.8/Rev1). For ease of understanding, the text in this chapter attempts to describe the index in simple, non-technical terms.
[2] The GEF Resource Allocation Framework. October 17, 2005.
[3] Threats use the IUCN Red List definitions for hierarchical classification of causes of species decline. Threats can occur from habitat loss/degradation (human induced) , invasive alien species (directly affecting the species) , harvesting [hunting/gathering] , accidental mortality , persecution , pollution (affecting habitat and/or species) , natural disasters , changes in native species dynamics , intrinsic factors , and human disturbance.
[4] The GEF Resource Allocation Framework. October 17, 2005. p. 15
[5] Priorities and indicators for Global Environment Benefits from Biodiversity: The current international architecture, *Impact Evaluation Information Document No. 5, Divya Nair, GEF Evaluation Office*, September 2007
[6] GEF Operation Program 1, from Classification of Biogeographical Provinces of the World (Udvardy, 1975).

[7] *Land Use, Land-Use Change, and Forestry*, Special report: IPCC, 2000 - Robert T. Watson, Ian R. Noble, Bert Bolin, N. H. Ravindranath, David J. Verardo and David J. Dokken (Eds.)

[8] Another way to show the RAF has indeed been designed to channel resource to target the higher GBI countries is to rank the eligible countries by their GBI scores in decreasing order and place them in quintiles. See statistical annex.
[9] The **megadiverse countries** are a group of countries that harbor the majority of the earth's species and are therefore considered extremely biodiverse. The World Conservation Monitoring Centre, an agency of the United Nations Environment Programme, has identified 17 megadiverse countries, most located in the tropics. http://en.wikipedia.org/wiki/Megadiverse_countries. A biodiversity hotspot must contain at least 1,500 species of vascular plants as endemics, and it has to have lost at least 70% of its original habitat. Around the world, at least 25 areas qualify under this definition, with nine others possible candidates.
[10] UNEP/CBD/COP/9/INF/20, 26 February 2008, REVIEW OF IMPLEMENTATION OF ARTICLES 20 AND 21 *Review of the effectiveness of the financial mechanism: the report of the independent evaluator commissioned to carry out the third review of the effectiveness of the financial mechanism*
[11] M.Sc. Thesis in International Studies in Aquatic Tropical Ecology: *Assessment of Global Marine Biodiversity Indicators for the Global Environment Facility Resource Allocation Framework* (GEF RAF), by Gorch Detlef Bevis Fedder, presented to the University of Bremen, Faculty for Biology & Chemistry, Bremen, August 2007
[12] OBIS (http://www.iobis.org/); KGS (http://drysdale.kgs.ku.edu/website/Specimen_Mapper/), http://www.seaaroundus.org/distribution/search.aspx), WCMC (http://www.unep-wcmc.org/), Veron (N/A), Jefferson, (http://nlbif.eti.uva.nl/bis/marine_mammals.php?menuentry=atlas), Reefbase (http://www.reefbase.org/global_database/default.aspx?section=r5), FishBase (www.fishbase.org)
[13] (http://www.sci.odu.edu/gmsa/)
[14] (http://www.worldwildlife.org/MEOW/)
[15] http://www.nceas.ucsb.edu/GlobalMarine
[16] (http://www.vulnerabilityindex.net/)
[17] See Bode, Balmford, and Naidoo.

[19] Other scenarios were also simulated: keeping year 2000 for nonco but considering 2 years after and before 2000 for co^2.
[20] Delphi average response 6.6 for "group countries that should qualify for individual funding", and 5.1 for "individual countries that should qualify for group allocation", standard deviation 2.1 and 2.2..
[21] FCCC/SBSTA/2008/L.13/Rev.1, Subsidiary body for scientific and technological advice, June 2008, Nairobi work programme on impacts, vulnerability and adaptation to climate change.
[22] Climate change: the cost of inaction and the cost of adaptation, European Environment Agency, EEA Technical report, 2007
[23] *The GEF Resource Allocation Framework*. October 17, 2005. p. 1
[24] Eccles and Gwin (1999). See also World Bank (1998) and Easterly (2001).
[25] Some of these countries have since graduated from GEF support; three in climate change are subject to paragraph 26 in *The RAF document* that fixes the country in group allocation due to historical and other reasons.
[26] The indicator source is included under IDA's "Policies for Social Inclusion/Equity" cluster of indicators within its Country Policy and Institutional Assessment (CPIA).
[27] 2007 CPIA guidelines, WB.
[28] World Bank (2007), p. 31.
[29] World Bank (2006).
[30] GEF/ME/C.33/3 (March 25, 2008) *Management Response to the GEF Annual Performance Report 2007*.

[31] ICR may include ratings on Risks, Sustainability, Impact, Quality at Entry, Bank Supervision, Overall Bank Performance, Borrower Preparation, IEG Borrower Implementation, IEG Borrower Compliance, Overall Borrower Performance, and ICR Quality.
[32] Several of these have since graduated from GEF assistance.
[33] W. Paul Vogt (1993). *Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences. Newbury Park, CA, p.245 ("Weighted Data").* In statistics, for example, weights are commonly applied to various subgroups of a sample in order to ensure that the resulting analysis reflects the relative proportions of subgroups in the population at large: Suppose we want to generalize about the attitudes of all 75-year-olds. We have a sample of 100 men and 100 women, all in their 75th year. We might want to give the women's attitudes more weight, because there are many more 75-year-old women than men and the goal is to generalize to the attitudes of all 75-year-olds.
[34] Volatility refers to the tendency for scores or rankings to fluctuate in value sharply and often. In a PBA, which typically is constructed in part to build regularity and predictability into allocations, volatility is something to be avoided.
[35] In addition, the GEF manages other resources such as the GEF Trust Fund for the four other focal areas, the adaptation funds and the funds for climate change (LDCF, SCCF). They existed before the RAF was developed and are not a set-aside, though they function as such because they have to be accessed and managed separately.
[36] This has been mentioned, for example, in GEF Sub-Regional Workshops.
[37] Some other IFIs use pools of funds, by country groups, but to a much smaller extent. For example, AsDB operates a pool of funds for its Pacific Region, which, with the exception of New Guinea, comprises very small island states. That Bank puts aside fifty million dollars into a pool that is allocated exclusively among the approximately 15 Pacific member states. Each obtains an individual allocation that is larger than would be the case if these states were to compete for allocations on an equal footing with all member states.
[38] Each Category 1 country was guaranteed an individual allocation (although no fixed amount) through the whole period. That is, its indicative allocation might change after mid-point reallocations but it would remain an individual allocation. Any country that is in Category 1 for the first half for GEF4 stays in that group for the whole of GEF4.
[39] In GEF4 about two-thirds of all countries were in Category 2 (71% of countries in the climate change focal area; and 62% in the biodiversity focal area).
[40] Only the highest individual allocation of any country in Category 2 is made public. The initial individual allocations that are the basis for the division into Groups 1 and 2 are not made public.
[41] GEF. The GEF Resource Allocation Framework. GEF Council Nov. 8-10, 2005. GEF/C.27/Inf.8/Rev.1, Washington DC. October 17, 2005. The RAF Resource Allocation Framework (2005) describes the formation of the two groups as follows: "Step 5. Indicative Allocations to countries and the group. 14. For each focal area, all eligible countries are listed in decreasing order of adjusted allocations. The highest-ranked countries whose cumulative adjusted allocations equal 75 percent of the total resources in the focal area will receive country specific indicative allocations equal to their respective adjusted allocations. 15. The remaining countries will be placed in a group with collective access to the indicative allocations for countries in the group for each focal area. The indicative allocation of the group for each focal area will consist of the resources available for the focal area that are not excluded from the RAF as specified in paragraph 22 and are not allocated to individual countries as specified in paragraph 14. For each focal area, the upper limit on approved projects for any country in the group will be equal to the adjusted allocation of the highest-ranked country in the group.
[42] Another way of coping with the same small-country-long-projects-cycle problem is that adopted by IFAD – quick iterative reallocations to delete "inactive" countries from the allocations list.
[43] The RAF document is ambiguous in some important respects in this regard. Specifically the RAF document requires two incompatible calculations. First, in paragraphs 9 to 15, it says that exclusions (essentially the 10% for global and regional and small grants) are made as the first step in calculating Category 2 allocations. That is all countries bear the cost of these exclusions. After exclusions, the remaining 90% of funds (the adjusted allocations) is split 75/25. Therefore Category 2 receives 25% of 90% - that is 22.5% of total funds in each focal area. However, in contradiction, paragraph 16 says that three items cannot add up to more than 25% of total funds. Those three items are: (1) The Group adjusted allocations (22.5% in the example above). (2) Exclusions (10%), and (3) Targeted supplements (defined as the difference between the total unadjusted allocation for all countries under $1 million and the total adjusted allocations for these countries – that is $1 million each). Targeted supplements is in GEF4 amount to $$25.9 million (Climate Change) and $15.3 million (Biodiversity). Clearly if all three items above must sum to less than 25% of total funds in the focal area, then something has to give. If Category 2 countries bear the full weight of exclusions as well as targeted allocations their share drops from 22.5% to around 15%. In fact the 75% is applied to the full focal area amount, not to the 90% going to countries.
[44] The RAF Delphi Panel was not asked a separate question on the appropriateness of ceilings on the maximum size of a single country allocation.
[45] Ceilings normally redistribute monies proportionally among all countries too small to reach the ceiling. One result is normally an increase in the median size of allocations, with a positive effect in terms of providing budget coverage of projects of at least minimum size.
[46] *The 2003 council paper defined"performance" as follows : What is "Performance" in the GEF context?* 5. The second issue is what "performance" means and how it is to be accounted for in the framework. There are three possible interpretations of performance: (a) Overall (macro-level) country policy, capacity, and institutional framework; or (b) Country policy, capacity, and

institutional framework relevant to delivery of global environmental benefits; or (c) Country policy, capacity, and institutional framework relevant to success of the project under consideration. 16. The first interpretation, commonly employed in multilateral agencies, requires an assessment of the current policy, capacity, and institutional framework of a country in terms of how conducive that framework is towards effective use of development assistance. Such an assessment usually rates factors that include, inter-alia, economic management, sector policies, social policies, public sector management and general governance, and portfolio performance. While such assessments could be readily available off-the-shelf for most countries from other multilateral institutions, their relevance to performance of GEF-supported projects is indirect. 17. The second interpretation would require the assessment of current policy, capacity, and institutional framework of a country in terms of how conducive that framework is towards effective delivery of global environmental benefits. Such assessments could be undertaken periodically, and used as a basis during review of projects; these assessments could also provide guidance towards the choice and size of GEF financing vehicles approved for each country. However, such assessments are not readily available, and would require some additional analytical work by the GEF Secretariat and the Agencies. 18. The third interpretation would require an assessment of "the ability of the country to successfully implement GEF-financed projects and deliver global environmental benefits." The current GEF programming framework embodies such an approach that assesses country/sectoral policy, capacity, and institutional framework that is appropriate for ensuring success of the project under consideration. As already mentioned, the GEF depends upon the policies and procedures of the Implementing and Executing Agencies to undertake this analysis, and the GEF project review criteria assesses the adequacy of such analysis.10 The current programming framework could be further enhanced by establishing, in agreement with the Council, specific country-level policy, capacity, and institutional assessment criteria to be used by Agencies in their due diligence assessment during project preparation."

[47] One exception is the IDB, which has a relatively simple allocation formula. It allocates 60% of its concessionary funds (Fund for Special Operations, FSO) solely according to member countries' relative scores on a performance index. Nothing other than country performance is taken into account in allocating this pot of funds. Separate pots of concessionary funds are allocated separately, each according to other variables.

[48] Variables: CIPE= Country Institutional and Policy Evaluation (IDB); CPIA= Country Policy and Institutional Assessment; DEBT = Debt service ratio; ES_CPIA= Economic and Social Performance Criteria in CPIA (for ADB); FSO=Fund for Special Operations (IDB); Fund= Size of IFF and FSO Envelope; GOV= Average of the five criteria in the "public sector management cluster" for ADF; average of the six criteria in the Governance and Public Sector Performance for AfDF; average of the five criteria in the public sector management cluster (Cluster D) for IDA; GNPPC = GNP per capita; Log = logarithm; HDI = Human Development Index; PCEF= Post-conflict Enhancement Factor (AfDB);POP = Population; PORT= Portfolio rating; Rural CPIA= Performance rating on policies and institutions for rural development (IFAD); VUL = Country Vulnerability (EU ACP).

[49] $$(0.22FUND)\left(\frac{POP}{\sum POP}\right)+(0.133FUND)\left(\frac{\frac{1}{GNPPC}}{\sum\frac{1}{GNPPC}}\right)+(0.6FUND)x\left[\frac{0.7CIPE+0.3PORT}{\sum(0.7CIPE+0.3PORT)}\right]$$

[50] It is interesting to note the use of LogPOP by two organizations (IFAD and the EC). The effect is to make a non-linear (exponential) distribution of country sizes closer to linear. This is useful when the organization has many small country members and, in addition, one or two relatively very large country members (CDB, for example, with its micro island states + Guyana and Haiti). By moderating the influence of country size, the "Log transformation" of the scale variable may help avoid explic it and arbitrary caps on the allocations of the largest countries. It would be possible to use LogGBI in the same way and for the same purpose.

[51] The amount of weight given to the performance variable changes frequently even within a single organization. For example, the traditional World Bank (IDA) allocation formula had an exponent of 2.0 on the performance variable. As one can see from Table 4.4.1 above, the most common exponent on the performance variable is still 2.0 (AsDB, CDB and IFAD). However the World Bank first added a separate "governance" variable, double counting part of the CPIA, and then, during the recent IDA 15 negotiations, dropped the separate governance variable, split the CPIA performance variable into two, and raised the exponent on the performance variable to 5.0. We are told that these changes, with the selection of new weights for all three components[51] of the performance variable, and the simultaneous increase of the exponent on the performance variable from 2.0 to 5.0, did not have a significant effect on the initial IDA 15 allocations. That is, the weights were selected to keep the allocations much as they would have been under the old formula. This may have been only partly successful and, anyway, there is more than one combination of weights that will achieve this result. The important point is that the sensitivity of IDA allocations to changes in performance in future has been substantially increased. The GEF RAF, with an exponent of 1.0 on the performance variable, is apparently less sensitive to changes in performance than is the case with IDA.

[52] The "resource concentration coefficient" is defined as the ratio of the monies allocated to the top quintile relative to the bottom quintile.

[53] GEF confuses the issue somewhat by including a performance variable, the change in carbon intensity of the economy, in its Global Benefits Index rather than in its Performance Index.

[54] However even in the IDB case there are complexities. There are in fact two scale variables in the IDB formula for allocating the IDB Fund for Special Operations. 22% of the Fund is allocated by share of population alone, and 13.3% is allocated by

relative levels of GNP $_{\text{per capita}}$ alone. These percentage weights are arbitrary, as all judgmental weights ultimately are, but, in addition, they lack a visible link to priorities and reasonableness.

[55] wiki
[56] GEF/C.24/Inf.10, November 10, 2004; *A conceptual design tool for exploiting interlinkages between the focal areas of the GEF*
[57] Integration and synergies has been also attempted a the institutional level with the creation of the Natural Resources Management Team (a merged of Bio, IW and Land Degradation).
[58] GEF/C.31/Inf.8, May 21, 2007, *Workshop report: STAP workshop on Small island developing states (SIDS), groundwater and interlinkages*
[59] LULUCF is not yet reflected in the indices.
[60] Report to the GEF on the quality of applications received on Project Identification Forms, October 16, 2007