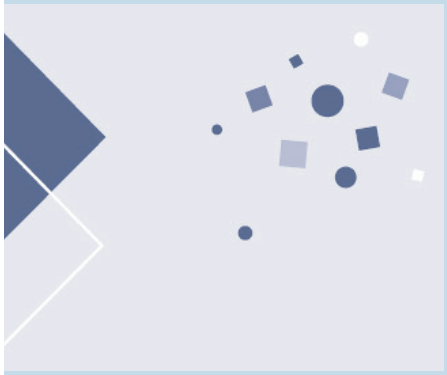
n° 05
août 2012

Série Notes méthodologiques

ExPost



Evaluer l'efficacité de politiques et programmes environnementaux : quel cadrage méthodologique proposer ?

Le cas du Fonds de partenariat pour le carbone forestier

Karine Belna, (karine.belna@engref.agroparistech.fr)
Tiphaine Leménager, (lemenagert@afd.fr)
Laurent Mermet, (laurent.mermet@engref.agroparistech.fr)

Contact : **Tiphaine Leménager**, département de la Recherche, AFD (lemenagert@afd.fr)

Département de la Recherche
Division Évaluation et capitalisation
Agence Française de Développement
5, rue Roland Barthes 75012 Paris - France
www.afd.fr

Cette série Ex Post « Notes méthodologiques » vise à accueillir les travaux relevant de l'analyse des pratiques, des méthodes et des utilisations de l'évaluation des politiques et interventions de développement.

Remerciements :

Nous souhaitons vivement remercier Constance Corbier-Barthaux, Guillaume Ernst, Laurent Fontaine et Jean-David Naudet pour le temps qu'ils ont consacré à la relecture de ce document et pour leurs précieux retours qui ont permis d'apporter des améliorations substantielles à l'analyse menée.

Avertissement :

Les analyses et conclusions de ce document sont formulées sous la responsabilité de ses auteurs. Elles ne reflètent pas nécessairement le point de vue de l'Agence Française de Développement ou de ses institutions partenaires.

Directeur de la publication : Dov ZERAH
Directeur de la rédaction : Laurent FONTAINE
Mise en page : Marcelle LARNICOL

ISSN : 1962-9761
Dépôt légal : 3e trimestre 2012

SOMMAIRE

Introduction 5

1. Le traitement de la question de l'*efficacité* au sein de l'abondante littérature consacrée à l'évaluation de politiques et de programmes 7

1.1. Quelques repères sur le développement de l'évaluation 7

1.2. Terminologie 8

1.2.1. Une définition plurielle de l'évaluation 8
1.2.2. Qu'entend-on par efficacité ? 9
1.2.3. Termes couramment employés en évaluation 10

1.3. Les approches portant sur l'efficacité dans le foisonnement des réflexions théoriques en évaluation 12

1.3.1. Démarche 12
1.3.2. Discussion de quelques approches portant sur l'efficacité 12
1.3.3. Une propension variable à la participation et au débat 14
1.3.4. Un « méta-modèle » de synthèse 17
1.3.5. Conclusion 19

2. L'évaluation de l'efficacité dans le champ de l'environnement : enjeux de cadrage normatif 23

2.1. Les difficultés posées par l'évaluation dans le champ de l'environnement 23

2.1.1. La complexité 24
2.1.2. L'architecture institutionnelle 25
2.1.3. L'échelle spatiale 25
2.1.4. L'échelle temporelle 26
2.1.5. Les enjeux d'équité et de bien public 26
2.1.6. Le rôle de la science, les incertitudes 26
2.1.7. Le caractère conflictuel des problèmes d'environnement 27
2.1.8. Conclusion : l'évaluation environnementale 27

2.2. La définition du référentiel de l'évaluation 28

2.2.1. S'interroger sur les objectifs du programme 28
2.2.2. Définir des objectifs propres à l'évaluation 29
2.2.3. Adopter une posture en termes de *résolution du problème* environnemental en jeu 30

2.3. Les effets à évaluer 34

2.3.1. Des évaluations souvent centrées sur les réalisations et les résultats 34
2.3.2. Comment confronter des effets préliminaires, à un référentiel correspondant aux impacts souhaités *in fine* 36
2.3.3. Effets objectifs ou subjectifs 37
2.3.4. Choisir une échelle spatiale adaptée au programme étudié 37

2.4. L'attribution et l'explication des effets 38

2.4.1. Expériences aléatoires et quasi-expériences 39
2.4.2. Construire des univers contrefactuels 40
2.4.3. Reconstituer les théories d'intervention et analyser les processus de mise en œuvre 40

2.5. Conclusion 43

3. Le Fonds de partenariat pour le carbone forestier (FPCF) : quel cadrage méthodologique proposer en vue d'évaluer son efficacité 45

3.1. Qu'est ce que le FPCF ? 45

3.1.1. La question de la déforestation des forêts tropicales au cœur des négociations sur le régime Climat post-Kyoto 45
3.1.2. Le FPCF, un des principaux programmes internationaux de préparation au REDD+ 46
3.1.3. Une mission principale posant des questions majeures de politique générale : l'examen et le financement des programmes nationaux REDD+ 51
3.1.4. L'importance accordée aux évaluations dans la mise en œuvre du FPCF 53

3.2. Discussion sur le cadre d'évaluation de l'efficacité du FPCF 55

3.2.1. Quel est le but de notre évaluation ? 55
3.2.2. Déterminer le référentiel de l'évaluation 56
3.2.3. Le périmètre de l'évaluation 61
3.2.4. Expliquer et attribuer les résultats : reconstituer les théories du programme FPCF et analyser le processus de mise en œuvre dans une perspective stratégique 63

3.3. Conclusion : comment les difficultés initiales de l'évaluation ont-elles été surmontées ? Quels enjeux demeurent ? 64

Conclusion générale 67

Annexe. Petit glossaire des approches pour l'évaluation de politiques et programmes 71

Liste des sigles et abréviations 89

Bibliographie 91

Introduction

Compte tenu des dégradations environnementales et de leurs impacts sur la société, l'environnement est devenu un champ d'action publique en soi, au même titre que l'éducation, la santé, le développement agricole, etc. Les politiques et programmes qui lui sont dédiés sont aujourd'hui nombreux, chacun tentant d'enrayer tel ou tel problème environnemental diagnostiqué. Pourtant, nombre de problèmes environnementaux restent non résolus, voire même s'accentuent (*Millennium Ecosystem Assessment*, 2005).

La question de l'efficacité de ces programmes se pose donc engendrant par là même toute une série de questions vis-à-vis des processus d'évaluation : comment mener des évaluations dans le champ de l'environnement ? Quels sont leurs enjeux, leurs limites ? Quels enseignements peut-on tirer des travaux portant plus généralement sur l'évaluation de politiques et programmes et de ceux s'intéressant à l'efficacité des régimes internationaux[1] dans le domaine de l'environnement ? Quel cadre d'évaluation peut-on alors proposer pour évaluer l'efficacité environnementale du Fonds de partenariat pour le carbone forestier (FPCF), programme multilatéral qui vise à réduire la déforestation dans le cadre de la lutte contre le changement climatique et auquel contribue l'AFD ? C'est à ces questions que tente de répondre le présent texte. Il 'inscrit dans le cadre d'un travail de recherche mené sur ces problématiques depuis 2009 en partenariat avec l'AFD, AgroParisTech et le Centre international de recherche pour l'environnement et le développement (CIRED)[2].

Elaboré à partir d'une revue de la littérature académique sur l'évaluation des politiques et des programmes, ce document propose tout d'abord de resituer les approches portant sur l'efficacité, au sein du paysage plus large des approches évaluatives existantes. Cette première étape permettra ainsi au lecteur, qu'il soit familier ou non du domaine de l'évaluation, d'appréhender la très grande diversité des approches possibles ainsi que les termes du débat pour l'évaluation de l'efficacité. Fort de ces analyses, ce document aborde ensuite les questions méthodologiques et théoriques que soulève le projet d'évaluer *l'efficacité* des politiques et programmes *environnementaux*. Enfin, dans un dernier temps, ces réflexions de cadrage théorique et méthodologique sont confrontées à un cas concret : le FPCF, programme international issu des négociations sur le changement climatique et la forêt.

[1] Le concept de « régime » est couramment employé dans le domaine des sciences politiques et plus précisément dans celui des relations internationales. La définition de Krasner de 1983, fait référence. Elle définit les régimes comme « *un jeu de principes, de normes, de règles et de procédures de décisions, qui peuvent être implicites ou explicites, et autour desquels les attentes des acteurs convergent, cela dans un domaine donné des relations internationales* ». On peut notamment prendre comme exemples les régimes du Changement climatique ou de la biodiversité. "*Regimes can be defined as sets of implicit or explicit principles, norms, rules, and decision-making procedures around which actors' expectations converge in a given area of international relations*" (Krasner, 1983).

[2] La première phase de ce travail a été consacrée à la construction du cadre d'évaluation du FPCF (objet de ce document). Le cœur de la recherche consiste ensuite à déployer l'évaluation à la fois au niveau international du programme FPCF mais aussi au niveau national, dans deux pays bénéficiaires du FPCF, choisis comme études de cas (Mexique et Congo Brazzaville). Ces analyses pourront faire l'objet d'autres valorisations telles que celle en cours de publication sur le Mexique (Bernard *et al.*, 2012).

1. Le traitement de la question de l'efficacité au sein de l'abondante littérature consacrée à l'évaluation de politiques et de programmes

1.1. Quelques repères sur le développement de l'évaluation

En guise de rétrospective historique, la « brève histoire de la recherche en évaluation », concernant les Etats-Unis, proposée par Rossi et Freeman (1993) dans leur cinquième édition d'*Evaluation* apparaît particulièrement instructive : c'est en effet dans les pays anglo-saxons que la pratique et la recherche en évaluation puisent leurs racines. On y apprend notamment que les premières évaluations de programmes sont antérieures à la première guerre mondiale et qu'ils ont d'abord concerné les domaines de l'éducation et de la santé (Alkin, 2004 ; Rossi et Freeman, 1993). Dans les années 1930, de nombreux chercheurs en sciences sociales ont également plaidé pour l'usage de méthodes de recherche rigoureuses en sciences sociales. C'est plus spécifiquement au cours de la décennie qui a suivi la seconde guerre mondiale que s'est concrétisée la recherche en évaluation. Des programmes publics à grande échelle ont en effet été lancés à cette époque pour répondre à des besoins dans de nombreux domaines (urbanisation, éducation, santé, nutrition, développement rural, etc.). D'importants montants ont été investis, justifiant une évaluation attendue des résultats obtenus. L'essor de l'évaluation dans les années d'après-guerre a aussi été lié au développement des méthodes de recherche et des statistiques ainsi qu'à l'avènement de l'ordinateur, qui ont rendu possible l'analyse de données à plus grande échelle. A la fin des années 1950, les évaluations de programme de grande ampleur sont donc devenues courantes. Les années 1960 voient le nombre de publications académiques sur l'évaluation augmenter considérablement et la première revue consacrée à l'évaluation : « *Evaluation review* » est créée en 1976 (Rossi et Freeman, 1993).

Suite à ces divers développements, la pratique de l'évaluation a véritablement « explosé » à partir des années 1990 et, aujourd'hui, il n'est plus une intervention publique (une politique, un programme, un projet) qui ne requière une évaluation préalable (Vedung, 2010). En s'appuyant sur la métaphore d'une vague et des sédiments qu'elle dépose, Vedung retrace l'histoire de l'évaluation depuis les années 1960 selon une succession de quatre vagues :

- la vague « guidée par la science » des années 1960, qui suppose les décideurs rationnels et leurs décisions fondées sur des analyses scientifiques « objectives » ;
- la vague « orientée par le dialogue » des années 1970, qui s'inscrit dans un paradigme constructiviste et place la participation au cœur de l'évaluation ;
- la vague « néolibérale » à la fin des années 1970, qui est mieux connue sous l'appellation de Nouvelle gestion publique (Naudet, 2012) ;
- la vague de « la preuve » (« *evidence* »), qui s'appuie sur l'expérimentation scientifique quantitative, avec notamment la mise en œuvre d'essais contrôlés randomisés.

L'évaluation apparaît comme un champ de pratique et de recherche transdisciplinaire, de par ses objets et ses approches théoriques. Celles-ci se sont multipliées au fil du temps. Près de 60 approches auraient ainsi été développées entre 1960 et 1990 (Fitzpatrick *et al.*, 2004). Cette diversité reflète la variété des points de vue, des disciplines d'origine des évaluateurs, des conceptions que ceux-ci se font de la rigueur méthodologique ou encore de ce qui constitue on non une « bonne » politique. Ces approches peuvent parfois différer légèrement, mais, par exemple, exiger des preuves chiffrées d'efficacité, tandis que d'autres fondent la légitimité de l'évaluation sur le caractère démocratique de son processus.

1.2. Terminologie

1.2.1. Une définition plurielle de l'évaluation

Les définitions de ce qu'est l'évaluation sont multiples, d'où l'intérêt de connaître les conceptions et débats qui existent avant de chercher à en approfondir certaines approches, comme nous le ferons par la suite.

Une définition simple, couramment citée comme référence, est celle de Scriven (1967) pour qui évaluer, c'est « *juger la valeur de quelque chose* »[3]. La notion de jugement de valeur est en effet souvent placée au cœur de la démarche évaluative. C'est ce qui la distingue de l'*analyse*.

La question de la *production de connaissances* à destination de certaines parties prenantes, ainsi que celle de l'*utilisation* de l'évaluation, sont également reprises dans de nombreuses définitions, même si le public ciblé ou l'objectif de la production de connaissances peuvent varier selon les auteurs. Ainsi, pour certains, « *il y a un besoin d'étendre la définition d'évaluation au-delà du jugement de valeur. […] L'évaluation doit être vue comme une production de connaissances, bornée par le besoin de fournir des informations relatives au programme, à des audiences identifiées qui ont un intérêt légitime dans le programme* » (Alkin, 2004, p. 359)[4]. Pour d'autres, « *l'évaluation est considérée comme une étude conçue et menée pour accompagner un public donné à évaluer la valeur d'un objet* » (Stufflebeam, 2001, p. 11)[5].

Le moment choisi pour l'évaluation est aussi un des éléments discuté. Pour certains auteurs, seules les évaluations *rétrospectives* (réalisées une fois le programme terminé) devraient ainsi être considérées comme des évaluations (Vedung, 2005). Il nous paraît néanmoins important de ne pas exclure les évaluations dites « *ex ante* » (réalisées en amont d'un programme) de la revue de littérature que nous présentons. Ces évaluations, et en particulier les évaluations environnementales et sociales stratégiques (EESS)[6], sont en effet censées être conduites en amont de la réalisation d'un programme, de façon à pouvoir mieux l'orienter, mais bien souvent leur rôle effectif suscite des questions (Slunge *et al.*, 2009). Dans le cas du FPCF, que nous examinerons en partie 3, nous reviendrons sur cette question et sur les diverses considérations de nature procédurales, politiques et stratégiques, en jeu dans la mise en œuvre de ces évaluations. Nous choisissons donc délibérément de conserver dans notre champ d'étude les évaluations *ex ante*, tout en leur reconnaissant un statut particulier (par rapport aux évaluations « chemin faisant » ou rétrospectives), du fait de leur temporalité théorique.

Enfin, la dimension *scientifique* de la démarche, et notamment le recours nécessaire à des « enquêtes empiriques rigoureuses » lorsque l'évaluation est conduite en sciences sociales, sont mentionnés comme des points importants par plusieurs auteurs (Crabbé et Leroy, 2008 ; Owen, 2007). Une bonne part des controverses dans le champ de l'évaluation correspond d'ailleurs aux controverses internes aux sciences sociales, entre différentes disciplines, différentes perspectives épistémologiques et méthodes de référence.

Le schéma 1 reprend les thèmes généralement associés à tout exercice de définition de l'évaluation.

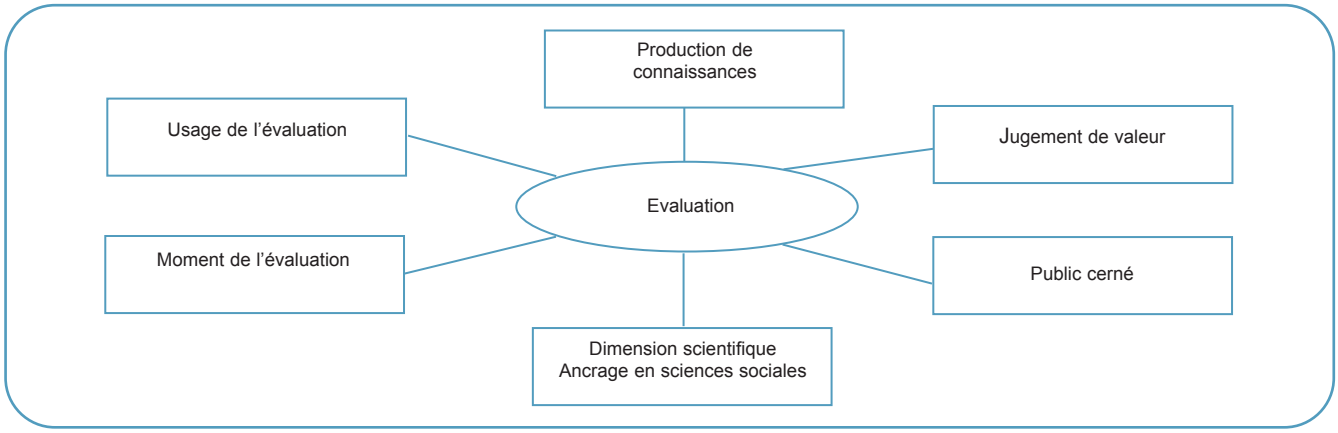
[3] "*Judging the worth or merit of something*".

[4] "*There [is] a need to expand the definition of evaluation over and above the judgement of worth. My view [is] that evaluation should be seen as knowledge production activity, bounded by the need to provide program-related information to identified audiences what had a legitimate interest in that program*".

[5] "*n this study, evaluation means a study designed and conducted to assist some audience to assess an object's merit and worth*".

[6] Ces évaluations sont l'équivalent, pour les programmes et les politiques, des études d'impacts *ex ante* des projets de développement.

Schéma 1. Thèmes associés à la définition de l'évaluation



Source : les auteurs.

1.2.2. Qu'entend-on par efficacité ?

Il est important dans un premier temps d'expliciter la définition d'efficacité que nous adoptons ici, les acceptations de ce terme variant selon les sources.

L'Organisation de coopération et de développement économiques (OCDE, dont les publications servent de référence en matière d'évaluation institutionnelle, définit l'efficacité comme « *le degré auquel les objectifs fixés par l'intervention ont été atteints ou sont censés être atteints […]* » (OCDE, 2002)[7].

La littérature académique, en matière d'évaluation, donne des définitions offrant un cadrage plus large que celui de l'OCDE : ainsi, l'objectif du programme n'est pas toujours celui qui servira pour l'évaluation (Vedung, 2005). Par exemple, les travaux de Young (1994), qui a écrit de nombreux ouvrages de référence dans le domaine de la gouvernance internationale de l'environnement, témoignent de la pluralité des façons de définir l'efficacité, qu'il résume en six dimensions :

- l'efficacité définie en termes de résolution de problème ;
- l'efficacité définie en termes d'atteinte des objectifs (elle rejoint la définition donnée par l'OCDE) ; elle peut différer de la première pour plusieurs raisons (si les objectifs du programme ne sont pas de résoudre le problème environnemental en question, mais par exemple de générer de la connaissance ou de la participation, si les véritables objectifs ne sont pas clairement énoncés, etc.) ;
- l'efficacité « comportementale » ; elle s'intéresse aux changements de comportement des acteurs visés par le programme ;
- l'efficacité « de processus » ; elle s'intéresse à la mise en œuvre du programme et aux instruments mis en place (les règles qui assurent la conformité avec les procédures, par exemple) ;
- l'efficacité « constitutive », caractérisée par la mise en exergue d'une pratique sociale (comme un processus de revue par les pairs, un processus d'apprentissage, etc.), censée pouvoir induire une meilleure résolution du problème en jeu ;
- l'efficacité « évaluative », axée sur l'efficience, l'équité, la durabilité ou la robustesse des résultats.

7 "*The extent to which the development intervention's objectives were achieved, or are expected to be achieved [...]*".

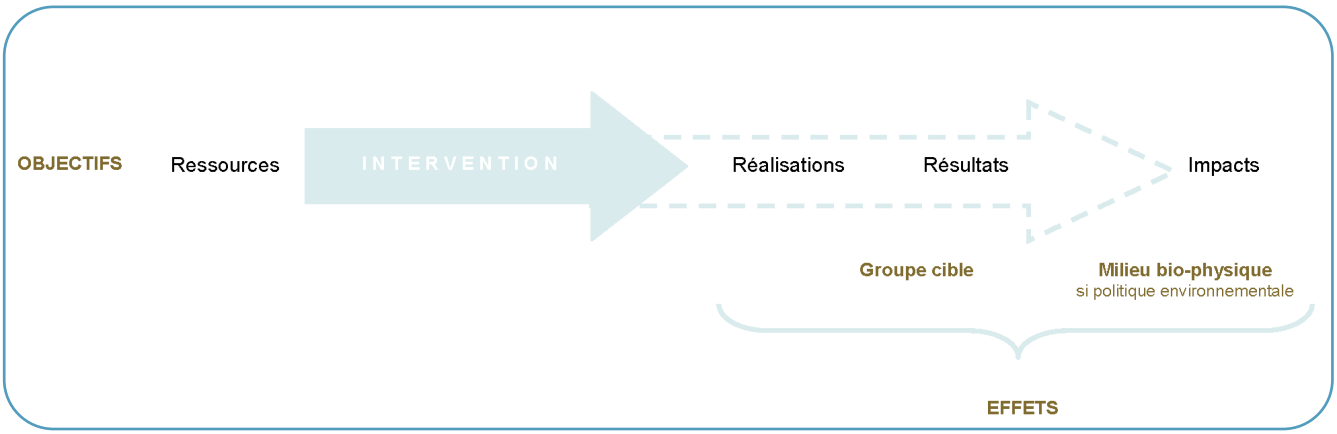
Le terme d'efficacité n'apparaît pas toujours en tant que tel dans les travaux académiques, mais peut être remplacé par l'expression « Evaluation d'impact » (Rossi et Freeman, 1993). L'évaluation d'impact s'intéresse au « *degré auquel le programme produit les résultats désirés* »[8], elle « *mesure le degré auquel le programme induit un changement dans la direction souhaitée* »[9] (Rossi et Freeman (1993), p. 36). Ces « *résultats désirés* » et cette « *direction souhaitée* » ne correspondent pas forcément aux objectifs de la politique étudiée, parce qu'ils sont flous ou peu explicites, ou qu'ils ont un parti pris que souhaite questionner l'évaluateur (Owen, 2007 ; Stufflebeam, 2001 ; Vedung, 2005).

Dans le souci de ne pas endosser un cadrage trop réducteur qui nous conduise à éluder le questionnement du bien-fondé de la politique même (sa pertinence), nous proposons d'adopter une définition de l'efficacité relativement large, qui trouve écho dans les travaux académiques tels que ceux cités *supra*. Nous la définissions comme étant le « *degré auquel l'intervention étudiée atteint les objectifs fixés par l'évaluation* ».

1.2.3. Termes couramment employés en évaluation

Dans l'abondante littérature anglo-saxonne traitant de l'évaluation de politiques et programmes, certains termes clés reviennent fréquemment (Wholey *et al.*, 2004). Présentées dans le schéma 2, nous les définissons en nous référant au glossaire de l'OCDE (2002), aux publications de l'Agence européenne pour l'environnement (AEE), (Guedes Vaz *et al.*, 2001), et à des ouvrages de synthèse en évaluation environnementale (Crabbé et Leroy, 2008 ; Guedes Vaz *et al.*, 2001).

Schéma 2. Logigramme d'une politique



Source : les auteurs.

[8] "*The degree to which a program produces the desired outcomes*" (Rossi et Freeman, 1993).

[9] "*An impact assessment gauges the extent to which a program causes change in the desired direction*" (Rossi et Freeman, 1993).

i) Les *ressources* (*inputs*) correspondent aux moyens financiers, humains et matériels mobilisés par le programme.

ii) Les termes *intervention*, *mécanisme* ou encore *traitement*, désignent l'action mise en œuvre dans le programme.

iii) Les termes *réalisations* (*outputs*), *résultats* (*outcomes*) et *impacts* correspondent à des effets du programme. Ils se distinguent par leur nature et le moment où ils se produisent.

- Les *réalisations* correspondent aux produits immédiats du programme (comme l'organisation d'atelier, la publication de standards de mise en œuvre, etc.). Elles peuvent être définies comme « *la quantité et la qualité des produits et des services délivrés par les décideurs* ».

- Les *résultats* correspondent aux effets du programme en termes de changement de comportement des personnes affectées par le programme, en particulier le groupe cible (*target group*). C'est en quelque sorte « *la réponse du groupe cible aux réalisations* ».

- Les *impacts* correspondent aux résultats finaux, observables à moyen et long termes. Ce sont tous les effets qui influencent l'état de l'objet visé par la politique. Ils peuvent être « *positifs et négatifs, primaires et secondaires, induits directement ou non, intentionnellement ou non* » (OCDE, 2002). Dans le cas d'une politique environnementale, il s'agira des effets bio-géo-physiques sur le milieu.

Pour illustrer les nuances existant entre ces termes, prenons un exemple : une politique qui aurait pour but de limiter la pollution de rivières en restreignant les rejets polluants des industries riveraines par une voie réglementaire. Les *réalisations* de cette politique pourraient être le nouveau règlement limitant les volumes de rejets polluants autorisés ou bien un système de quotas de droits à polluer, accompagnés par les fascicules d'information à destination des cibles, etc. Les *résultats* seraient la modification du comportement des industries face à ces nouvelles règles (réduction de leur pollution, contournement de la règle, délocalisation, etc.). Enfin, les *impacts* se rapporteraient aux nouveaux taux de pollution mesurés dans les rivières suite à la mise en œuvre de la politique (impacts en termes d'efficacité environnementale). Les impacts sociaux pourraient être les modifications des conditions de travail éventuellement engendrées par l'application (ou le contournement) de ces nouvelles réglementations par les usines (comme l'augmentation de la pénibilité du travail, des délocalisations d'emplois, etc.).

Ces termes définis, nous nous intéressons maintenant de plus près aux approches existantes concernant l'évaluation de politiques et programmes, afin de comprendre la place occupée par celles qui portent spécifiquement sur l'efficacité et les débats qu'elles induisent.

1.3. Les approches portant sur l'efficacité dans le foisonnement des réflexions théoriques en évaluation

1.3.1. Démarche

Pour se repérer dans ce domaine marqué par le foisonnement des méthodes, nous proposons deux niveaux de présentation.

- Le premier niveau vise à donner une vue d'ensemble du domaine et à situer l'évaluation portant sur l'efficacité. Pour cela nous nous appuierons sur des travaux de synthèse qui se sont attachés à recenser et classifier les différentes approches théoriques et méthodologiques de l'évaluation (Alkin, 2004 ; Crabbé et Leroy, 2008 ; Fitzpatrick *et al.*, 2004 ; Owen, 2007 ; Rossi et Freeman, 1993 ; Stufflebeam, 2001 ; Vedung, 2005 ; Wholey *et al.*, 2004). Chacun, selon sa perspective propre, propose un panorama des approches d'évaluation. Nous recoupons ces différents panoramas, en soulignant leurs apports respectifs en matière d'évaluation de l'efficacité et en résumant des aspects clés de l'évaluation qu'ils auront permis de mettre en évidence.

- Le second, proposé en annexe, présente succinctement chacune des 50 approches de l'évaluation, évoquées par les travaux de synthèse introduits ici. Pour chaque approche, nous avons indiqué en intitulé sa dénomination dans le texte (la même approche peut être désignée par des termes différents, selon les auteurs) ; nous résumons en quelques lignes ce qui en fait la spécificité et nous indiquons, lorsque cela est possible, une référence bibliographique particulièrement pertinente.

Ce catalogue pourra ainsi montrer à quel point des approches très différentes peuvent chacune avoir un véritable intérêt, selon les objectifs que l'on se fixe et le contexte d'évaluation où l'on se situe. Le lecteur notera aussi, sans doute, le caractère inévitablement hétérogène d'un tel catalogue. Certaines des approches répertoriées sont répandues, influentes et constituent parfois des pans entiers de la littérature sur l'évaluation. D'autres sont plus marginales ou correspondent à des regroupements ou des caractérisations qui sont le fait des auteurs de typologie. Le lecteur percevra les grandes proximités de certaines approches (par exemple, celles qui mettent au centre de l'évaluation l'implication des parties prenantes) et les différences profondes d'autres approches. Nous nous sommes abstenus de les regrouper dans nos propres catégories pour deux raisons. D'une part, notre objectif étant de passer en revue plusieurs typologies proposées par d'autres auteurs, nous nous sommes attachés à respecter la diversité telle qu'elle ressort de l'ensemble de leurs travaux. D'autre part, dans le champ de l'évaluation, des différences de méthode qui pourraient paraître mineures dans le principe, sur le papier, peuvent conduire à des approches bien différentes sur le terrain, tant dans leur conduite que dans les résultats.

1.3.2. Discussion de quelques approches portant sur l'efficacité

- Dans son ouvrage « *Public Policy and Program Evaluation* », Vedung (2005) propose une taxonomie des modèles d'évaluation, construite autour du concept d'efficacité.

  Précisons ici que sa typologie fait suite à une triple réduction. D'une part, il définit l'évaluation comme 1) restreinte au champ des services publics et des affaires gouvernementales et 2) strictement rétrospective (*ex post*), excluant donc de son propos les évaluations dites *ex ante*. D'autre part, 3) Vedung ne s'intéresse qu'aux modèles qualifiés de *substantiels*, centrés sur les résultats du programme, et non aux modèles *procéduraux*, qui portent sur certaines caractéristiques de la mise en œuvre du programme, comme l'équité, la représentativité, la participation ou encore la légalité.

Dans les modèles *substantiels*, les modèles évaluant l'efficacité sont prédominants. Selon la question organisatrice (*organizer*), qui fonde l'évaluation, Vedung distingue six modèles d'efficacité :

- le modèle d'« *atteinte des objectifs* » (*goal attainment model*), qui est l'approche classique ;
- le modèle des « *effets secondaires* » (*side-effects model*), qui invite à prendre en compte les effets non-attendus ;
- le modèle « *libre de tout objectif* » (*goal free model*), qui questionne la fixation initiale d'un cadre d'évaluation et propose de conduire l'évaluation sans aucun objectif *a priori* ; nous détaillerons ces approches section 2.2 ;
- les modèles « *centrés sur les clients* » (*consumer-oriented model*), ou plus largement sur les parties prenantes (*stakeholder model*), qui abordent la question de l'efficacité sous l'angle de *Qui* définit l'objectif et le référentiel de l'évaluation ;
- l'approche « *englobante* » (*comprehensive model*), qui s'intéresse non seulement aux *réalisations* et aux *résultats* mais aussi à la mise en œuvre globale du programme. C'est une forme d'évaluation de processus.

Outre ces six modèles d'efficacité, Vedung identifie deux autres types de modèles *substantiels* :

i) Les modèles évaluant l'efficience. Ils se distinguent des modèles d'efficacité par leur dimension économique. Ils mettent en effet en regard les résultats du programme avec les coûts (modèles de productivité), voire même, attribuent en amont une valeur économique aux résultats (modèles d'efficience).

ii) Les modèles professionnels. Les revues par les pairs (appelées par certains auteurs : « *connoisseurship and criticism* ») représentent les modèles pro-fessionnels par excellence. Il faudrait ajouter les panels d'experts, auxquels Vedung ne fait pas allusion.

Concernant ces modèles professionnels, Vedung apporte une réserve par rapport à la définition des modèles substantiels qu'il avait donnée. Pour lui, les revues par les pairs « *ne donnent pas d'indication sur la question substantielle qui va guider l'évaluation [comme l'efficacité ou l'efficience] mais sur […] comment l'évaluation va être menée. Ce modèle nous dit seulement que l'évaluation va être conduite par des confrères* » (Vedung, 2005)[10]. Ainsi, les modalités de l'évaluation sont abordées par cette approche, et notamment la manière dont le jugement sera formulé (ici par des pairs, mais cela pourrait être, pour d'autres approches, par des experts ou par certaines parties prenantes etc.).

Nous retenons de la typologie de Vedung les trois éléments suivants :

1. le concept d'*organizer* qui correspond au « *point de départ logique de l'évaluation* », à « *la question basique posée dans l'évaluation* » apparaît intéressant pour essayer de distinguer et de regrouper les nombreuses approches existantes ;
2. le cadrage de l'évaluation[11] de l'efficacité fait débat, en témoigne l'existence de divers modèles d'efficacité qui répondent chacun à leur façon à la question « *Qui définit l'objectif de l'évaluation ?* » ;

[10] *"Peer review is an institutional model of evaluation. It does not indicate the substantive question to be asked but how the evaluation is to be organized. The model only tells you that evaluations ought to be carried out by equals"* (Vedung, 2005, p. 88).

[11] On entend par « cadrage » de l'évaluation, la phase de conceptualisation de l'évaluation. C'est l'étape préalable à la conduite de l'évaluation, à proprement parler. Durant cette phase préliminaire se posent les questions : de mise à plat des enjeux de l'évaluation, de définition du critère et du référentiel de l'évaluation (à quelle question herchera à répondre l'évaluation), de choix de la méthode qui sera employée etc.

3. au moment du cadrage de l'évaluation ou lors de sa conduite, le degré d'implication des parties prenantes et de débat varie, y compris pour les évaluations d'efficacité.

- Dans leur ouvrage « *Evaluation, a systematic approach* », Rossi et Freeman (1993) traitent eux aussi explicitement des approches portant sur l'efficacité. Ils les distinguent, d'une part, des évaluations servant à conceptualiser les interventions et, d'autre part, des approches à dominante gestionnaire, qui permettent de *suivre* et de *mesurer* la mise en œuvre du programme dans une perspective de redevabilité (*accountability*).

  Leur typologie met donc l'accent sur le moment où est réalisée l'évaluation, moment qui influence la nature de celle-ci : à un stade précoce, avant la mise en œuvre du programme, il s'agira d'évaluations destinées à influencer sa teneur et sa portée (en particulier des « *analyses de besoins* »); au cours de sa mise en œuvre, les évaluations auront un caractère de suivi ; enfin, à un stade final, les évaluations pourront porter sur les résultats.

  Dans cette typologie, à nouveau les évaluations d'efficacité sont associées aux évaluations d'efficience. Parmi les évaluations d'efficacité, les auteurs recensent les évaluations d'impacts, les expériences et les quasis expériences.

  La typologie de Rossi et Freeman (1993) s'intéresse par ailleurs précisément aux questions de méthode. D'une part, ils développent très largement les méthodes quantitatives que sont les expériences et les *quasi-expériences*. D'autre part, ils s'interrogent, avec les évaluations d'impacts, sur la méthodologie de l'attribution des effets : les effets bruts observés dans les impacts incluraient non seulement les effets propres du programme (« *effets nets* »), mais aussi des facteurs externes liés à d'autres interventions ainsi que des biais dus au dispositif de l'évaluation (Rossi et Freeman, 1993, p. 221). A la différence de la typologie de Vedung, la dimension participative, que ce soit lors du cadrage ou lors de la mise en œuvre de l'évaluation, n'apparaît pas dans le travail de Rossi et et Freeman. Ils mettent en revanche un fort accent sur les méthodes quantitatives pour l'évaluation de l'efficacité.

  En nous appuyant sur d'autres travaux de synthèse en évaluation, et du fait de leur prépondérance dans les débats, nous approfondirons deux des points abordés, qui concernent autant les évaluations d'efficacité que les autres types d'évaluations, à savoir :

  - le degré de participation et de délibération, lors du cadrage de l'évaluation ou lors de sa mise en œuvre ;

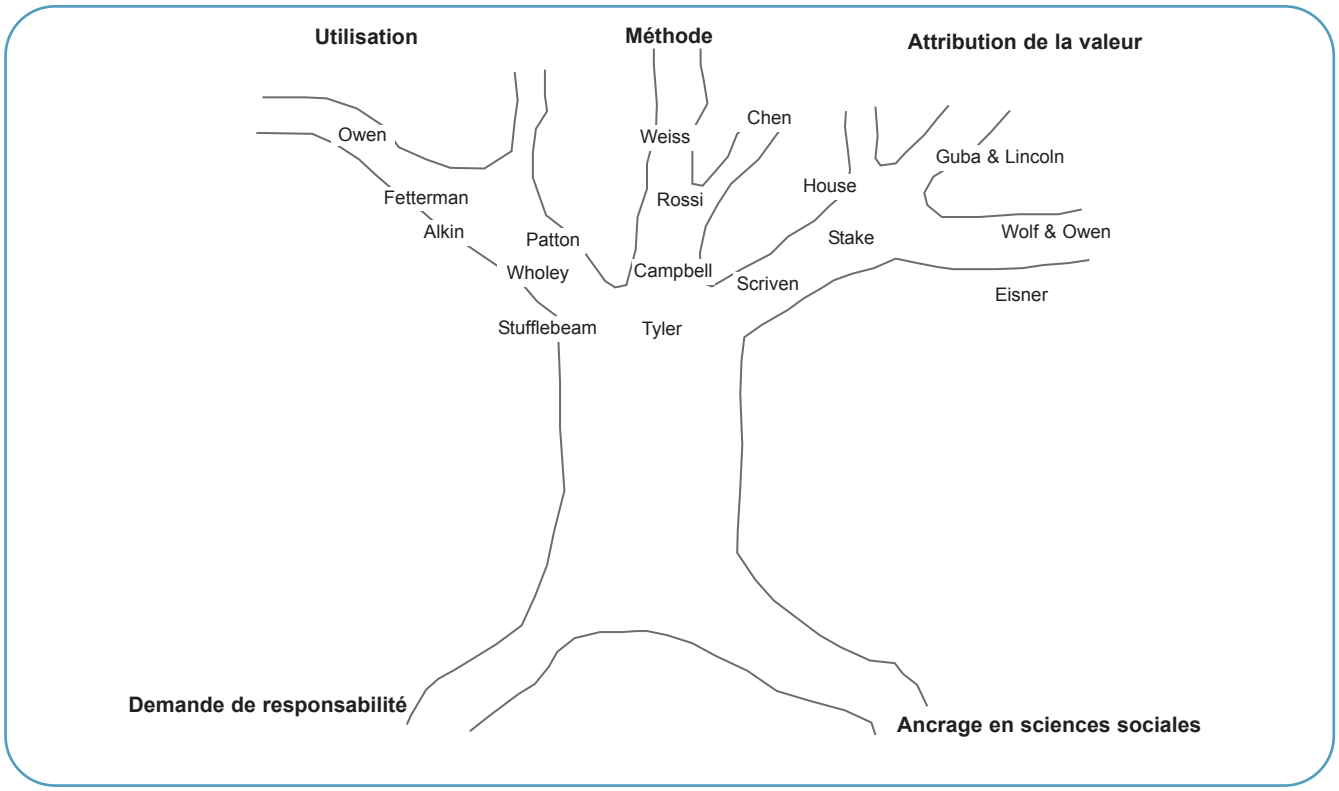
  - le moment de l'évaluation, et en particulier, la question de savoir si les évaluations de l'efficacité sont l'apanage des stades ultimes du programme.

1.3.3. Une propension variable à la participation et au débat

- Dans son ouvrage de référence « *Evaluation Roots. Tracing Theorists' Views and Influences* », Alkin (2004) propose une image vivante de la recherche américaine en évaluation au cours de la seconde moitié du XX^{e} siècle, sous la forme d'un arbre (cf. schéma 3). Cet arbre a deux racines qui caractérisent, selon lui, les fondements de la recherche en évaluation : la demande de responsabilité (*accountability and control*) et l'ancrage en sciences sociales (*social inquiry*). Par ailleurs, cet arbre généalogique de l'évaluation possède trois branches principales qui définissent trois dimensions essentielles pour toute théorie en évaluation :
  - la méthode,
  - l'attribution de la valeur (*i.e.* le jugement formulé sur le programme),
  - l'utilisation de l'évaluation.

L'auteur s'est efforcé de positionner sur ces trois branches les chercheurs américains en évaluation qui ont « *apporté des contributions significatives* » à l'une des trois questions mentionnées *supra*. Pour ce faire, l'ouvrage d'Alkin rassemble les témoignages de ces chercheurs sur les fondements de leur parcours de recherche et leur vision de l'évaluation. Les approches auxquelles ils ont donné naissance sont présentées en annexe.

Schéma 3. La recherche en évaluation aux Etats-Unis depuis 1960



Source : adapté de Alkin (2004, p. 13).

Retenons trois points de ce travail de synthèse :

1. La question de la formation de la valeur (en d'autres termes, du jugement) est un élément clé pour toute évaluation, qu'elle porte ou non sur l'efficacité. Sur ce point, il est intéressant de voir l'hétérogénéité des approches identifiées par l'auteur. On retrouve des approches qui reconnaissent aux experts la légitimité de formuler le jugement, comme les évaluations « *critiques* » (*connoiseurship* et *criticism*), déjà citées. Mais on trouve aussi des approches « *partisanes* » (*advocate-adversary*), que l'on doit à Wolf (1975) et Owens, et qui soulignent la pluralité des constructions possibles du jugement et s'appuient sur la métaphore d'une audience avec plusieurs avocats pour revendiquer des débats contradictoires. Egalement, des approches dites « *naturalistes* » (développées par Guba et Lincoln, 1989), qui reposent sur un paradigme constructiviste fort et plaident pour une large implication des parties prenantes, notamment au moment de la formulation du jugement.

2. La question de l'utilisation de l'évaluation est aussi très importante. Elle fait l'objet de nombreux travaux, ceux de Patton avec son approche « *centrée sur l'utilisation* » (*utilization-focused*) étant les plus connus. Ils partent du constat problématique que les évaluations sont trop souvent lettre morte, si bien que l'on peut se demander si elles sont utiles (Delarue *et al.*, 2009).

3. Certains travaux en évaluation se sont penchés exclusivement sur des questions de méthodes, quantitatives (comme les *expériences* et *quasi-expériences*), mais également qualitatives, comme les approches « *fondées sur les théories* » (*theory-based ou theory-driven* auxquelles Chen (1990) a beaucoup contribué, ou encore les approches politiques portées par Weiss (2004).

- La classification proposée dans « *Program Evaluation, Alternatives. Approaches and Practical Guidelines* » (Fitzpatrick *et al.*, 2004), permet d'éclairer la place donnée au débat dans la conduite de l'évaluation, en particulier au moment de la formulation du jugement. Elle complète en cela le travail d'Alkin (2004). Elle situe pour cela les approches d'évaluation suivantes sur un gradient croissant « utilitaire – délibératif » (cf. schéma 4) :

  - Les approches « *orientées par les objectifs* » (*objective-oriented*), qui s'intéressent avant tout au cadrage de l'évaluation, comme l'approche traditionnelle centrée sur l' « *atteinte des résultats* » (*goal-attaintment*) mais aussi l'approche « *libérée de tout objectif* » (*goal free*) que nous avons déjà présentées.

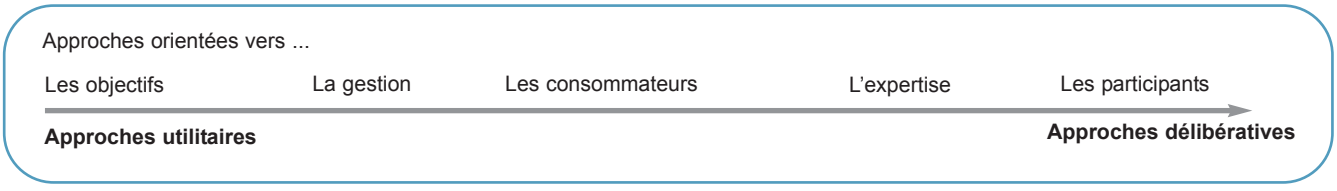
  - Les approches « *orientées vers* la *gestion du programme* » (*management-oriented*), destinées à informer et éclairer les décideurs et les gestionnaires du programme. L'approche emblématique est l'approche intégrative « *contexte, ressources, processus et produits* », (*context, inputs, process and products*), développée par Stufflebeam (2001), qui analyse le programme à tous ses stades.

  - Les approches « *orientées vers les consommateurs* » (*consumer-oriented*), destinées à offrir des connaissances aux consommateurs et destinataires du programme sur les produits de celui-ci.

  - Les approches « *orientées vers l'expertise* » (*expertise-oriented*), qui traitent la participation des experts à l'évaluation.

  - Les approches « *orientées vers les participants* » (*participant-oriented*), pour lesquelles l'évaluateur est envisagé comme facilitateur ; l'implication des parties prenantes dans l'évaluation est vue comme une priorité pour déterminer les critères, les besoins, les valeurs et les conclusions de l'évaluation. Se retrouvent dans cette catégorie : l'évaluation « *réactive* » (*responsive-evaluation*) de Stake (2003), communément utilisée par les organisations internationales lors des évaluations de programme à mi-parcours, l'évaluation «*naturaliste*» que nous avons vue et qui s'inscrit dans un paradigme constructiviste fort, ou encore l'évaluation « *capaci-tation* » (*empowerment evaluation*) de Fetterman (2001), qui plaide pour que l'évaluation serve aussi à renforcer l'influence de parties prenantes du programme, qui sont minoritaires.

Schéma 4. Cinq types d'approches évaluatives selon leur degré de délibération



Source : adapté de Fitzpatrick et al.*, 2004.*

- Dans sa monographie des approches évaluatives académiques anglo-saxonnes, Stufflebeam (2001) regroupe lui aussi les approches qui se singularisent par l'attention accordée aux parties prenantes et à leur adhésion à la démarche d'évaluation, qu'il appelle « *approches de l'agenda social ou de plaidoyer* » (*social agenda/advocacy approaches*). Il y recense aussi l'approche « *délibération et démocratie* » (*deliberative-democratic*), qui prône ces valeurs dans la conduite de l'évaluation, ainsi que l'évaluation « *centrée sur l'utilisation* » (*utilization-focused*), déjà citée.

  Notons, outre le travail considérable d'explicitation et de qualification de chaque modèle d'évaluation, deux autres points d'intérêt de sa monographie. En plus des approches à caractère social ou de plaidoyer, l'auteur définit :

  1. des approches « *orientées vers les questions et les méthodes* » (*questions/methods-oriented*), c'est-à-dire qui s'interrogent sur le cadrage de l'évaluation, que nous avons déjà largement abordé précédemment et qui fondent principalement leur légitimité sur la rigueur de l'acquisition et du traitement des données (comme les évaluations « *fondées sur les théories* » (*theory-based*), les *expériences* et *quasi-expériences*, les méthodes dites mixtes (quantitatives et qualitatives), les évaluations « *critiques* » (*connoisseurship and criticism*) déjà citées, mais aussi les études de cas ;

  2. des approches ayant trait à l'amélioration et à la prise de responsabilité (*improvement/accountability*), c'est-à-dire qui fondent leur légitimité sur leur inscription opérationnelle dans le système de gestion de l'action.

  Par ailleurs, une catégorie concerne les « *pseudo-évaluations* », fortement régies par des enjeux politiques, qui « *proposent des conclusions incomplètes ou invalides* ». Stufflebeam est le seul auteur à mentionner de telles évaluations et nous rappelle ici le caractère souvent politique de nombreuses évaluations et les incidences que cela peut avoir sur la conduite de celles-ci.

- La dernière classification que nous souhaitons aborder ici est celle présentée dans « *The Handbook of Environmental Policy Evaluation* » de Crabbé et Leroy (2008). Ils y proposent « *une revue des approches en évaluation qui ne se veut en aucun cas exhaustive* », qui reprend les approches, selon eux, « *les plus appropriées dans le contexte particulier de l'évaluation dans le champ de l'environnement* ». Leur démarche – un passage en revue du champ de l'évaluation pour en tirer des ressources pour l'évaluation des politiques environnementales – est apparentée à la nôtre. Elle diffère cependant sur deux points importants : la manière de procéder (en mettant en regard plusieurs panoramas présentés par d'autres auteurs, plutôt que de procéder directement à un passage en revue des approches) et le point focal de la démarche : pour nous, la question de l'efficacité, pour eux, l'institutionnalisation de l'évaluation environnementale.

  Sans chercher à commenter leur classification (deux catégories), nous soulignons toutefois un élément important de leur proposition : le fait que des méthodes utilisées pour l'évaluation de projets soient considérées en complément d'un certain nombre des approches mentionnées *supra*. C'est le cas du cadre logique, des évaluations d'impacts *ex ante*, des évaluations multicritères et des méta-évaluations.

1.3.4. Un « méta-modèle » de synthèse

Pour conclure le panorama des approches existantes en évaluation, esquissé au fil des visions données par quelques ouvrages de synthèse, nous nous penchons sur le méta-modèle proposé par Owen (2007) dans son ouvrage « *Program Evaluation, Forms and Approaches* ». Chacune des cinq formes d'évaluation qu'il identifie répond à un objectif spécifique de production de connaissance à un stade donné de la mise en œuvre du programme. Cette nouvelle classification (cf. tableau 1) recoupe en grande partie les typologies précédemment évoquées, et fait donc écho aux

réflexions que nous avons développées sur les évaluations d'efficacité et leur cadrage, ainsi que sur le degré de délibération dans la conduite des évaluations. Elle aborde la question critique du moment de l'évaluation, et mêle explicitement des approches empruntées au domaine de la gestion avec d'autres approches clairement issues des sciences sociales.

i) Les évaluations « *proactives* » (*proactive*) répondent à un souci de synthèse de l'information existante en lien avec le programme, avant la mise en œuvre de ce dernier ou à ses débuts. Le but est d'aider la conception ou la redéfinition de l'intervention, au regard d'expériences passées, ou d'une bonne connaissance du contexte de mise en œuvre. Elles correspondent à la phase de conception de la politique, déjà évoquée, avec notamment les « *analyses de besoin* ».

ii) Les évaluations de « *clarification* » (*clarificative*) s'attachent à rendre explicite l'intervention, dans le courant de sa mise en œuvre. Elles visent à rendre compte de la cohérence des activités menées par rapport aux objectifs fixés et à analyser le processus de mise en œuvre. On trouvera notamment les approches « *fondées sur les théories* » (*theory-based*), qui proposent de reconstituer les hypothèses et liens causaux du modèle, les évaluations *ex ante* ou encore les évaluations qui cherchent à déterminer si le programme est évaluable, avant de lancer toute évaluation (*evaluability assessment*).

iii) Les évaluations dites « *interactives* » (*interactive*) ont pour but d'améliorer les interventions, dans le courant de leur mise en œuvre, à travers le partage d'information. Les évaluations de ce type correspondent aux approches sociales et de plaidoyer déjà largement abordées. Cette forme d'évaluation de type procédural, a davantage pour finalité l'implication des parties prenantes dans l'évaluation, qu'un résultat quelconque d'évaluation. L'apprentissage tiré de la mise en oeuvre de l'intervention et l'amélioration de celle-ci sont envisagés à travers le niveau d'appropriation et de discussion de l'intervention par les parties prenantes.

iv) Les évaluations de « *suivi* » (*monitoring*) reposent sur le suivi de la mise en œuvre de l'intervention comme garantie de qualité. Elles répondent à une demande de responsabilité.

Enfin, les évaluations d'« *impact* » visent à créer de la connaissance sur le succès et les conditions de succès de l'intervention, une fois celle-ci bien avancée. C'est au sein de ce dernier groupe que l'on retrouve les évaluations de l'efficacité déjà évoquées, comme les évaluations « *centrée sur les objectifs* » ou « *libérée des objectifs* », mais aussi des évaluations de l'efficacité fondées sur les « *besoins* » (*need based*), des évaluations du « *processus et des réalisations* » (*process-outcomes*) ou des audits de performance. Notons la mention de l'« *évaluation réaliste* » (Pawson et Tilley, 1997), approche majeure en évaluation – pourtant omise de nombreuses classifications – qui repose sur deux propositions fondamentales :

1. les évaluations doivent adopter une épistémologie réaliste, c'est-à-dire qu'il ne s'agit pas tant de prouver des corrélations entre moyens mis en œuvre et résultats que de comprendre comment fonctionnent, dans la mise en œuvre concrète, les mécanismes qui génèrent les résultats ;

2. il n'est pas possible de porter un jugement décontextualisé sur l'efficacité d'un mécanisme. Il est seulement possible de dire qu'un programme donné conduira à des résultats donnés sous certaines conditions (dans un contexte donné).

Tableau 1. Formes et approches pour l'évaluation de programmes

Nom (numéro de recensement dans l'annexe)		Catégorie
Analyse des besoins (*needs assessment*)	31	proactive
"research synthesis" (*evidence-based practice*)	43	
Revue des meilleures pratiques (*review of best practice*)	45	
Evaluation du caractère évaluable (*evaluability assessment*)	20	clarificative
Evaluation de la logique du programme (*program logic evaluation*)	39	
Evaluation *ex ante*	22	
Evaluation reactive (*responsive evaluation*)	44	interactive
Recherché action (*action research*)	3	
Evaluation recapitulative (*developmental evaluation*)	17	
Evaluation comme capacitation (*empowerment evaluation*)	19	
Revue de qualité (*quality review*)	41	
Analyse de composante (*component analysis*)	7	suivi
Evaluation de la performance (*developed performance assessment*)	18	
Analyse systémique (*system analysis*)	48	
Evaluation fondée sur les objectifs (*objectives-based*)	33	impact
Evaluation fondée sur les besoin (*needs-based*)	31	
Evaluation libérée de tout objectif (*goal-free*)	25	
Evaluation de processus et de réalisation (*process-outcome*)	38	
Evaluation réaliste (*realistic*)	42	
Audit de performance (*performance audit*)	35	

Source : adapté de Owen (2007, p. 1).

1.3.5. Conclusion

Au fil de cette première partie, appuyée de la lecture de l'annexe, nous pouvons prendre la mesure du foisonnement des approches proposées par la littérature académique, pour évaluer des politiques et programmes. Chacune de ces approches porte un regard partiel sur la démarche évaluative fonction de son ancrage disciplinaire, de la posture et du niveau de théorisation qu'elle adopte (cf. encadré 1).

Encadré 1

Les approches académiques et les différents points d'intérêt auxquels elles s'intéressent

Toutes les approches académiques ne s'attachent pas aux mêmes aspects de l'évaluation. Le tableau 2 reprend un certain nombre d'approches et tente de montrer leurs différents centres d'intérêts.

- Certaines approches académiques s'expriment sur la nature du **critère d'évaluation**. L'évaluation portera-t-elle sur les réalisations ou les effets du programme, *i.e.* son efficacité, son efficience (Vedung, 2005), ou plus précisément sur son efficacité environnementale au regard d'un référentiel « centrée sur cette préoccupation » (Mermet *et al.*, 2010) ? Aura t-elle plutôt une visée procédurale, par exemple en s'attachant à favoriser « la délibération » et « la démocratie » (House, 2000), en ayant pour objectif central d'être utile et utilisée (Patton, 2008), ou encore en visant avant tout à « capaciter » certains acteurs minoritaires (Fetterman, 2001) ?

- D'autres approches s'intéressent **aux méthodes (ou théories) mobilisées** par l'évaluateur **pour conduire son évaluation et pour formuler le jugement de valeur**. Seront-elles plutôt quantitatives, *i.e.* les expériences, les quasi-expériences (Rossi et Freeman, 1993) ? Ou bien seront-elles qualitatives, *i.e.* les approches « fondées sur les théories » (Chen, 1990), les approches constructivistes (Guba et Lincoln, 1989), ou encore l'« évaluation réaliste » (Pawson et Tilley, 1997) ?

- D'autres approches encore, réfléchissent **aux modalités-mêmes de l'évaluation** (la façon dont l'évaluation est menée). On trouvera des modalités plus ou moins délibératives, contradictoires et pluralistes, que ce soit au stade du cadrage de l'évaluation, *i.e.* l'approche « d'atteinte des objectifs » et à l'opposé les approches centrées sur les parties prenantes ou les clients (Vedung, 2005) ; ou que ce soit au stade de la réalisation de l'évaluation, *i.e.* comme l'expertise, les revues par les pairs (Vedung, 2005), les évaluations « partisanes » (Wolf, 1975) ou encore les évaluations « réactives » (Stake, 2003).

Tableau 2. Approches pour l'évaluation de programmes

<table>
<tr><td>Nature du critère</td><td colspan="2">Substantiel
- évaluation « fondée sur l'objectif »
- évaluation « libérée de tout objectif » (Scriven, 1972)
- évaluation « centrée sur les effets secondaires » (Vedung, 2005)
- évaluation « centrée sur la préoccupation [environnementale] » (Mermet et al., 2010)
- évaluation « fondée sur les besoins »
- évaluation « coûts-bénéfices »</td><td colspan="2">Procédural
- évaluation « déliberative et démocratique » (House, 2000)
- évaluation « centrée sur la participation »
- évaluation « centrée sur l'utilisation » (Patton, 2008)
- évaluation « de capacitation » (Fetterman, 2001)</td></tr>
<tr><td>Méthode</td><td colspan="2">Quantitatif
Expérience et quasi-expérience (Rossi et Freeman, 1993)</td><td colspan="2">Qualitatif
- évaluation « fondée sur les théories » (Chen, 1990)
- évaluation « réaliste » (Pawson et Tilley, 1997)
- évaluation « naturaliste » (Guba et Lincoln, 1989)</td></tr>
<tr><td rowspan="2">Degré de délibération</td><td colspan="2">Phase de cadrage de l'évaluation</td><td colspan="2">Conduite de l'évaluation (notamment formulation du jugement de valeur)</td></tr>
<tr><td>Pas de délibération
Modèle « d'atteinte des objectifs » (Vedung, 2005)</td><td>Délibération
Modèles « orienté vers les parties prenantes » ou « orienté vers les clients » (Vedung, 2005)</td><td>Délibération restreinte
Expertise et revue par les pairs (Vedung, 2005)</td><td>Délibération plus ouverte
- évaluation « partisane » (Wolf, 1975)
- évaluation « réactive » (Stake, 2003)</td></tr>
</table>

Source : les auteurs.

En conclusion, reprenons les principaux enseignements tirés de cette revue de littérature.

Le caractère plus ou moins participatif et délibératif des approches d'évaluation est un paramètre qui permet de les différencier de manière utile et discriminante. Ainsi, on trouvera des approches qui placent au cœur de leur proposition la question de l'implication des parties prenantes dans l'évaluation (on pourrait qualifier ces approches de *procédurales*). D'autres, en revanche, n'abordent pas du tout cette question, mais portent plutôt sur la définition du référentiel et des effets à mesurer (approches que l'on pourrait qualifier d'approches *substantielles*).

Un autre paramètre distinguant les approches les unes des autres est le *moment de l'évaluation* sur lequel elles portent plus spécifiquement. Globalement, nous avons identifié deux stades importants, qui ressortent en filigrane des travaux de synthèse : le *moment du cadrage* de l'évaluation et de la formulation de la question organisatrice d'une part, et, d'autre part, la phase de conduite de l'évaluation et de formulation du jugement de valeur. Les approches que nous avons évoquées privilégient, en général, l'un ou l'autre de ces moments.

Par ailleurs, le moment où l'évaluation est réalisée par rapport à la mise en œuvre de la politique (en *ex ante*, en *in itinere* ou en *ex post*) est aussi essentiel (cf. encadré 2). Pour certains auteurs, l'évaluation ne peut être envisagée qu'*ex post*, pour d'autres, en revanche, le stade *ex ante* est une opportunité à saisir pour influencer la politique dans ses objectifs ou dans sa mise en œuvre.

En ce qui concerne les approches portant explicitement sur l'efficacité, c'est-à-dire visant à formuler un jugement sur les effets, au regard d'un référentiel donné, nous avons vu qu'elles développaient soit, les enjeux de *cadrage*, avant tout, soit plutôt les aspects ayant trait aux méthodes (méthodes quantitatives ou qualitatives). Nous avons choisi d'examiner en détail *la phase du cadrage de l'évaluation* qui est un moment critique des évaluations d'efficacité. Nous pensons qu'une fois le cadrage posé, bon nombre d'approches traitant des méthodes d'évaluation ou de la formulation du jugement peuvent alors être mobilisées pour une évaluation de l'efficacité. Il revient à l'évaluateur de choisira l'une ou l'autre de ces approches, selon le parti pris quantitatif ou qualitatif et le paradigme dans lequel il souhaite s'inscrire (constructiviste, structuraliste, réaliste…).

Encadré 2

La référence au temps en évaluation

Il convient de ne pas confondre le temps de l'évaluation et l'horizon de l'évaluation (Hilden, 2009).

La référence au temps est double en évaluation. Il y a, d'une part, le moment où se situe l'évaluation par rapport à la mise en œuvre de l'intervention étudiée. On parlera alors d'évaluations *ex ante*, d'évaluations « chemin faisant » (*in itinere*) ou encore d'évaluations *ex post*. Il y a, d'autre part, l'horizon temporel de l'évaluation elle-même : porte-t-elle sur des effets passés, des effets présents ou bien anticipe-t-elle des effets à venir ? Les combinaisons possibles entre ces deux considérations temporelles conduisent à des problématiques évaluatives spécifiques, qui ont leurs propres contraintes et opportunités. Même si toutes les combinaisons ne sont pas exploitées en pratique, il n'en demeure pas moins utile de garder à l'esprit cette double référence au temps, pour mieux comprendre la visée et le positionnement de chaque évaluation.

Dans le cadre du Fonds de partenariat pour le carbone forestier (FPCF), étude de cas que nous traiterons section 3, les pays bénéficiaires sont amenés à réaliser deux types d'évaluations *ex ante* : d'une part, une évaluation de leurs politiques forestières passées afin de tirer des enseignements pour la politique à mener, que financera le FPCF, d'autre part, une étude environnementale stratégique des impacts que pourrait avoir la politique financée par le FPCF.

En lien avec le moment auquel est réalisée l'évaluation, deux conceptions de l'évaluation sont couramment distinguées (Scriven, 1967). Les évaluations « *formatives* » (*endoformatives*) sont le plus souvent effectuées au cours de la phase de mise en œuvre et « *visent à améliorer les performances d'un projet ou d'un programme* ». On les associe souvent aux « *évaluations de processus* ». Les évaluations « *récapitulatives* » (*summatives*) sont conduites en fin d'action (ou à la fin d'une étape de cette action) pour déterminer dans quelle mesure les réalisations escomptées ont été atteintes. L'évaluation « *récapitulative* » vise à fournir des informations sur la validité du programme. On l'associe à l'évaluation d'impact (OCDE, 2002).

Nous nous intéresserons à présent, plus spécifiquement, aux questions méthodologiques et théoriques soulevées par le projet d'évaluer l'efficacité des politiques et programmes environnementaux.

2. L'évaluation de l'efficacité dans le champ de l'environnement : enjeux de cadrage normatif

L'enjeu de cette deuxième partie est de voir comment l'évaluation des politiques environnementales peut traiter la question de l'efficacité. Si cette question n'est que l'un des points d'attention de l'évaluation, elle nous semble toutefois d'une importance toute particulière, alors que la dégradation et la destruction de l'environnement se poursuivent et qu'il apparaît toujours aussi difficile des les enrayer. Nous exploiterons, section 3, ces réflexions méthodologiques et théoriques en les confrontant au projet d'évaluer l'efficacité du FPCF.

Evaluer l'efficacité de l'action publique dans le champ de l'environnement soulève des difficultés liées aux caractéristiques du problème environnemental abordé (déforestation, pollutions diffuses, épuisement des ressources halieutiques, etc.). En outre, poser et cadrer clairement la question de l'efficacité n'est pas une tâche simple.

Nous présenterons tout d'abord les enjeux auxquels est confronté l'évaluateur qui entreprend une évaluation dans le champ de l'environnement (section 2.1). Nous examinerons ensuite les concepts que la littérature de l'évaluation offre, pour aider à poser clairement les questions portant sur l'efficacité d'une action publique. Nous nous intéresserons ici uniquement à la phase de *cadrage* de l'évaluation de l'efficacité, qui mérite en soi des clarifications[12]. Pour cela, nous aborderons successivement trois questions organisatrices (Underdal, 2002) :

- le choix du référentiel de l'évaluation (section 2.2) ;
- la définition précise de l'objet de l'évaluation et l'établissement du périmètre sur lequel porte l'étude (section 2.3) ;
- l'explication et l'attribution des effets (section 2.4).

Pour chacune de ces questions, sont présentés les choix qui s'offrent à l'évaluateur et les arbitrages qu'ils suscitent. Ces arbitrages sont d'autant plus délicats que des considérations liées à la fois à la pertinence conceptuelle, à la faisabilité méthodologique et à la robustesse politique de l'évaluation entrent en jeu (Mermet, 1996). Ces questions étant interdépendantes, nous soulignons également comment ces choix se répercutent d'une question organisatrice à l'autre.

2.1 Les difficultés posées par l'évaluation dans le champ de l'environnement

La pratique de l'évaluation dans le champ de l'environnement est relativement récente. Elle remonte aux années 1980-1990. Dans la littérature sur l'évaluation des politiques publiques, le domaine environnemental n'occupait encore au début des années 2000 qu'une faible place (Mickwitz, 2003). Le développement de l'évaluation dans ce domaine a suivi son développement comme secteur d'activité spécifique (avec la création des ministères en charge de l'environnement, l'apparition de politiques publiques et de programmes dédiés à l'environnement, etc.), mais avec un retard, que Knaap et Tschangho (1998) attribuent à la dominance des sciences naturelles dans les études environnementales, et donc à une certaine distance avec les sciences sociales qui fondent les travaux sur l'évaluation des politiques.

Aujourd'hui, ce retard est en cours de rattrapage. La pratique à peu près systématique de l'évaluation est maintenant un acquis dans le champ de l'environnement, et la littérature académique à ce sujet s'étoffe rapidement. Il est important de noter que le développement de ces travaux se déroule sur un double front : d'un côté, les études environnementales découvrent peu à peu les problématiques de l'évaluation de l'action publique, de l'autre, les spécialistes des méthodes

[12] Nous suggérons qu'une réflexion sur la conduite des évaluations portant sur l'efficacité soit menée dans d'autres travaux.

d'évaluation découvrent progressivement le champ d'application que constitue, pour eux, l'environnement.

Une controverse importante accompagne le développement des évaluations des politiques environnementales. Elle porte sur le caractère spécifique, ou non, de ces évaluations par rapport à d'autres domaines de l'action publique (Crabbé et Leroy, 2008 ; Ferraro, 2009 ; Knaap et Tschangho, 1998 ; Mickwitz et Birnbaum, 2009 ; Vedung, 2009). Notre but n'est pas d'approfondir ce débat. Nous pensons d'ailleurs, que cette controverse peut être attribuée au cantonnement thématique des évaluateurs ou chercheurs en évaluation. L'écueil possible est alors d'imputer à son propre domaine d'application (ici l'environnement) des difficultés méthodologiques rencontrées également dans d'autres domaines (Mickwitz, 2003, p. 416). Nous soulignerons donc simplement ces difficultés dans le champ de l'environnement, domaine qui nous intéresse (sans présager de savoir si ces problèmes se posent identiquement ou non dans d'autres domaines de l'évaluation).

Pour cela, nous nous penchons successivement sur différents attributs considérés par beaucoup d'auteurs comme caractéristiques de nombreux problèmes d'environnement. Suivant Mickwitz[13] dans sa démarche, nous abordons dans un premier temps les caractéristiques associées à l'environnement lui-même, et dans un deuxième temps, celles associées aux connaissances que nous avons des problèmes environnementaux (Mickwitz, 2003, p. 416) (cf. tableau 3). Nous ajoutons à ces considérations une considération relative aux politiques d'environnement qui nous semble très importante : la structure administrative et l'architecture institutionnelle associées.

Tableau 3. Approches pour l'évaluation de programmes

Caractéristiques des problèmes d'environnement	Caractéristiques relatives à la connaissance
- Ils sont complexes - Ils ont des pas de temps longs - Ils concernent des zones géographiquement isolées - Leurs conséquences et leurs causes sont inégalement distribuées	- Ils ont été formulés en tant que problèmes, largement par des scientifiques - Ils mobilisent de fortes incertitudes - Ils impliquent des parties prenantes avec des croyances différentes et des objectifs conflictuels

Source : adapté de Mickwitz (2003, p. 416).

2.1.1. La complexité

La complexité (cf. encadré 3) des problèmes d'environnement est un point mis en avant par tous les auteurs abordant les difficultés liées à l'évaluation dans le champ de l'environnement (Crabbé et Leroy, 2008 ; Ferraro, 2009 ; Gysen *et al.*, 2002 ; Knaap et Tschangho, 1998 ; Mickwitz et Birnbaum, 2009). Elle concerne non seulement les phénomènes naturels sans intervention humaine directe mais aussi les interactions socio-environnementales. Elle est donc liée, d'une part, aux caractéristiques fonctionnelles des écosystèmes et, d'autre part, à la complexité des systèmes anthropiques impliqués (Mickwitz, 2003, p. 416). Cette complexité peut notamment s'illustrer par les effets d'irréversibilité, les effets de seuils et de non-linéarité. On parle ainsi souvent de la *résilience* des écosystèmes, qualité qui, au-delà d'un certain seuil de dégradation, disparaît, entraînant alors l'irréversibilité du phénomène de dégradation, pouvant être brutal (non-linéaire). Citons à ce propos l'exemple d'une

[13] Mickwitz a coordonné le numéro spécial de la revue *New Directions for Evaluation*, l'été 2009, portant sur les enjeux méthodologiques de l'évaluation dans le champ de l'environnement. Ce numéro a été publié à l'issue des trois forums annuels organisés par l'*Environmental Evaluators Network*, entité établie aux Etats-Unis en 2006 sur la base d'une collaboration entre l'Agence de protection de l'environnement (EPA) et la Fondation nationale des poissons et de la vie sauvage.

activité de pêche intensive qui conduit à une raréfaction de certaines populations de poissons jusqu'à un seuil d'irréversibilité : à un moment donné, la population constituée d'un nombre trop réduit d'individus ne peut se renouveler. Elle est alors vouée à disparaître très brutalement. On pourrait prendre d'autres exemples, comme la fragmentation des forêts qui favorise leur destruction, ou encore l'emballement de la « *machine climatique* » pouvant conduire à un accroissement de la fréquence de phénomènes extrêmes, dès lors qu'un certain réchauffement de l'atmosphère est atteint.

On peut aussi considérer que la structure des phénomènes environnementaux génère de la complexité à deux autres niveaux : au niveau des politiques environnementales mises en place et au niveau de l'évaluation de ces politiques (Mickwitz, 2003, p. 416).

Encadré 3

Point de vocabulaire : ce qui est compliqué n'est pas forcément complexe… (Morin, 2004)

Le terme « *complexe* » est employé pour désigner des situations ou des phénomènes variés. A titre d'exemple, Mickwitz et Birnbaum (2009) l'emploient indifféremment pour parler de la non-linéarité des phénomènes environnementaux ou des différences de vision caractérisant les politiques et programmes qui visent à résoudre les problèmes environnementaux.

Il est important de clarifier ici l'emploi de ce concept en évaluation, et en particulier ce qui le distingue du concept de « *compliqué* », qu'il supplante souvent. S'appuyant sur les travaux de Glouberman et Zimmerman (2002) qui portent sur la conceptualisation du « *compliqué* » et du « *complexe* », Rogers (2008) propose une classification distinguant les aspects compliqués des aspects complexes d'une intervention. En résumé, on peut associer « le compliqué » à la multiplicité des composantes, et relier « la complexité » à l'incertitude et aux phénomènes d'émergence (Rogers, 2008, p. 30).

Des interventions pourront donc être qualifiées de *compliquées* si elles ont un ou plusieurs des aspects suivants : 1) une gouvernance et une mise en œuvre reposant sur plusieurs agences, 2) des chaînes causales multiples, 3) des mécanismes causaux dépendant du contexte. Des interventions seront qualifiées de *complexes* si elles présentent un caractère récursif avec des boucles de rétroaction ou si elles présentent des effets émergents (Rogers, 2008, p. 30).

2.1.2. L'architecture institutionnelle

L'environnement est un domaine transversal. Les politiques qui s'y rapportent recoupent souvent le champ d'influence d'autres politiques sectorielles (agriculture, santé, transports, énergie etc.). L'architecture institutionnelle et administrative qui en résulte est alors source de difficultés pour l'évaluateur. En effet, même dans les cas où le portage de la politique est clairement attribuée à l'administration en charge de l'environnement, de nombreuses autres administrations sont également concernées, ce qui renforce le caractère compliqué de la politique et donc de l'évaluation entreprise. Des schémas de gouvernance horizontale (sectorielle) et verticale (multi-niveaux) (Stame, 2004) se mêlent étroitement. Face à de telles interactions, il est difficile pour l'évaluateur de savoir *à qui* imputer *quels effets*.

2.1.3. L'échelle spatiale

Les problèmes d'environnement peuvent mettre en relation des régions géographiquement éloignées (c'est le cas des problèmes de pollution diffuse ou de changement climatique, par exemple) (Ferraro, 2009 ; Mickwitz, 2003). Ils ne sont pas non plus cantonnés à des frontières administratives (Crabbé

et Leroy, 2008, p. 31). Les politiques environnementales mises en place sont donc aussi souvent fortement marquées par ces caractéristiques (Ferraro, 2009). Le caractère très global des politiques environnementales, tant en ce qui concerne les processus de décision que les stratégies et la mise en œuvre, rend alors difficile le choix d'un niveau administratif approprié pour l'évaluation. « *Alors que les politiques sont de plus en plus multiniveaux, l'évaluation doit en faire autant ; les questions méthodologiques soulevées sont ardues* » (Crabbé et Leroy, 2008, p. 39)[14].

2.1.4. L'échelle temporelle

Une autre caractéristique de beaucoup de problèmes d'environnement est la différence de pas de temps entre, d'une part, la mise en œuvre de la politique environnementale et ses résultats en termes de changements de comportement sociaux, et, d'autre part, les impacts environnementaux attendus (Mickwitz, 2003). Ce problème est souligné par de nombreux auteurs (Bennear et Coglianese, 2005 ; Crabbé et Leroy, 2008, p. 31 ; Ferraro, 2009). Alors que la phase de mise en œuvre de la politique peut être bien avancée voire même terminée, les impacts sur l'environnement sont parfois encore invisibles. C'est le cas, par exemple, des politiques visant l'amélioration de la qualité de l'eau des nappes souterraines dans les bassins d'alimentation de captage en eau potable. Les effets des mesures limitant les rejets azotés agricoles dans ces zones sont longs à constater du fait de la lenteur de l'infiltration des polluants dans le sol, jusqu'aux nappes.

2.1.5. Les enjeux d'équité et de bien public

Le concept de *bien public*[15] est souvent employé pour parler d'objet environnemental, comme le climat, la qualité de l'air, les océans, ou encore les forêts (même si l'emploi de cette notion est parfois très contesté, comme le fait le Brésil au sujet de l'Amazonie). Les biens publics sont définis comme des objets d'action collective qui concernent *a priori* tout le monde. Les questions d'équité se posent alors fortement lorsque les conséquences dues à leur dégradation sont très inégalement réparties (Mickwitz, 2003). Il est alors difficile d'évaluer la politique en question sans traiter les considérations d'équité qui lui sont étroitement associées.

2.1.6. Le rôle de la science, les incertitudes

Les connaissances sur les problèmes environnementaux sont soumises à de fortes incertitudes (Mickwitz, 2003), ce qui rend les politiques environnementales mises en place fortement sujettes à la controverse. L'insuffisance des données en termes de disponibilité et de qualité est soulignée par plusieurs auteurs (Bennear et Coglianese, 2005 ; Crabbé et Leroy, 2008). Dans certains cas, les séries temporelles (données collectées régulièrement sur une longue période) n'existent pas car les agences en charge de l'environnement et de la collecte d'informations environnementales n'ont été établies que récemment. Dans d'autres situations, c'est la nature de l'information recueillie qui a pu évoluer au cours du temps, ne permettant pas de s'appuyer sur des séries temporelles fiables. C'est le cas notamment des données d'inventaire forestier de l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO), qui ont le mérite de couvrir l'ensemble des pays forestiers mais dont la fiabilité est fortement critiquée, de même que la rigueur des méthodes de mesure d'une session à l'autre (Grainger, 2008).

Par ailleurs, la complexité et la technicité qui s'est emparée de la compréhension de la plupart des phénomènes environnementaux leur confèrent un fort ancrage dans le monde scientifique et les rendent souvent difficilement intelligibles pour des non-experts (Gysen *et al.*, 2006, p. 102). La connaissance et le discours scientifique jouent donc un rôle prééminent dans le développement des politiques environnementales.

[14] "*As policies become more multilevel, so must evaluation, however great the methodological issues this raises*".

[15] Paul A. Samuelson définit un bien public, comme un bien ou un service dont l'utilisation est non-rivale et non-exclusive (Samuelson, 1954).

2.1.7. Le caractère conflictuel des problèmes d'environnement

La dernière caractéristique à examiner concerne le caractère conflictuel des problèmes d'environnement. Ces derniers suscitent en effet la mobilisation de groupes d'intérêts divergents ayant des systèmes de croyance et des objectifs différents (Mickwitz, 2003, p. 416), qu'il s'agisse d'ailleurs de l'intégration des politiques environnementales dans des domaines d'action sectoriels (comme l'industrie, l'agriculture, etc.) (Lafferty et Hovden, 2003) ou de politiques environnementales autonomes, mais qui ne manqueront pas d'interagir avec d'autres secteurs (par exemple, l'assainissement des eaux). « *Les objectifs conflictuels qui sont en jeu dans de nombreux problèmes d'environnement engendrent souvent des processus d'élaboration des politiques longs et exigeants en ressources, processus qui résultent en fait dans une sorte de politique de compromis* » (Mickwitz, 2003, p. 418)[16].

De ce fait, la politique finalement définie peut n'être que lointainement rattachée aux objectifs environnementaux initiaux (Gysen *et al.*, 2006, p. 102), ou alors elle peut être définie en des termes ambigus qui assoient mal l'ambition de résolution du problème environnemental et rendent difficiles l'évaluation (Mitchell, 2008).

2.1.8. Conclusion : l'évaluation environnementale

Nous venons de brosser plusieurs caractéristiques du champ de l'environnement, auxquelles tout évaluateur qui débute son travail d'évaluation se trouve confronté. Soulignons que bon nombre de domaines de l'action publiques, transversaux (comme l'aménagement du territoire), ou même sectoriels (comme les politiques agricoles, d'immigration ou encore d'éducation) peuvent également présenter les caractéristique que nous avons abordées : complexité, imbrication des niveaux, temporalité longue, enjeu de bien public, controverses ou encore conflictualité etc. Mais chacun de ces domaines, l'environnement inclus, possède néanmoins sa spécificité propre.

Au final, l'évaluation des politiques environnementales serait idéalement servie par le croisement entre une compétence approfondie dans le champ de l'évaluation des politiques publiques en général, et une compétence approfondie dans les domaines d'action publique environnementale. Sur le plan opérationnel, cela incite à rechercher des équipes d'évaluation mixtes, associant des spécialistes confirmés des deux domaines. Sur le plan des perspectives de recherche, cela implique *(i)* que les travaux sur l'évaluation des politiques environnementales prennent toute leur place au sein de la littérature spécialisée sur l'évaluation, au même titre que les politiques sociales, de santé ou d'éducation – ce qui est encore loin d'être le cas, et *(ii)* que ces mêmes travaux prennent aussi leur place, et soient débattus dans la littérature sur les études environnementales, ou sur les sciences de l'environnement.

Dans les sections 2.2 à 2.4, nous discutons les choix normatifs qui se posent à l'évaluateur qui entreprend une évaluation de l'efficacité d'un programme. Nous présentons dans quelle mesure l'adoption de telle ou telle option permet de résoudre les enjeux que nous venons de présenter.

[16] *"The conflicting objectives involved in many environmental problems often trigger long and resource-intensive processes of policy formation, finally resulting in some sort of compromise policy".*

2.2. La définition du référentiel de l'évaluation

L'explicitation du référentiel de l'évaluation est un élément essentiel lors du cadrage de l'évaluation. Elle soulève plusieurs questions : faut-il endosser ou questionner les objectifs d'un programme ? Comment définir un référentiel propre à l'évaluation menée ? Envisage-t-on le référentiel en termes de résolution de problème ou en termes d'amélioration ? Nous exposons *infra* les alternatives soulevées par ces différentes questions.

2.2.1. S'interroger sur les objectifs du programme

Dans la section 1, plusieurs conceptions de l'efficacité ont été vues. Si traditionnellement et couramment, l'évaluation de l'efficacité endosse les objectifs du programme (OCDE, 2002 ; Owen, 2007 ; Vedung, 2005), des travaux académiques issus de la recherche en évaluation (Gysen *et al*., 2006 ; Mickwitz, 2003 ; Vedung, 2005) et en gouvernance internationale de l'environnement (Haas *et al*., 1993 ; Young, 1994) proposent néanmoins des postures qui voient en l'évaluation l'opportunité d'adopter un référentiel de valeur plus large.

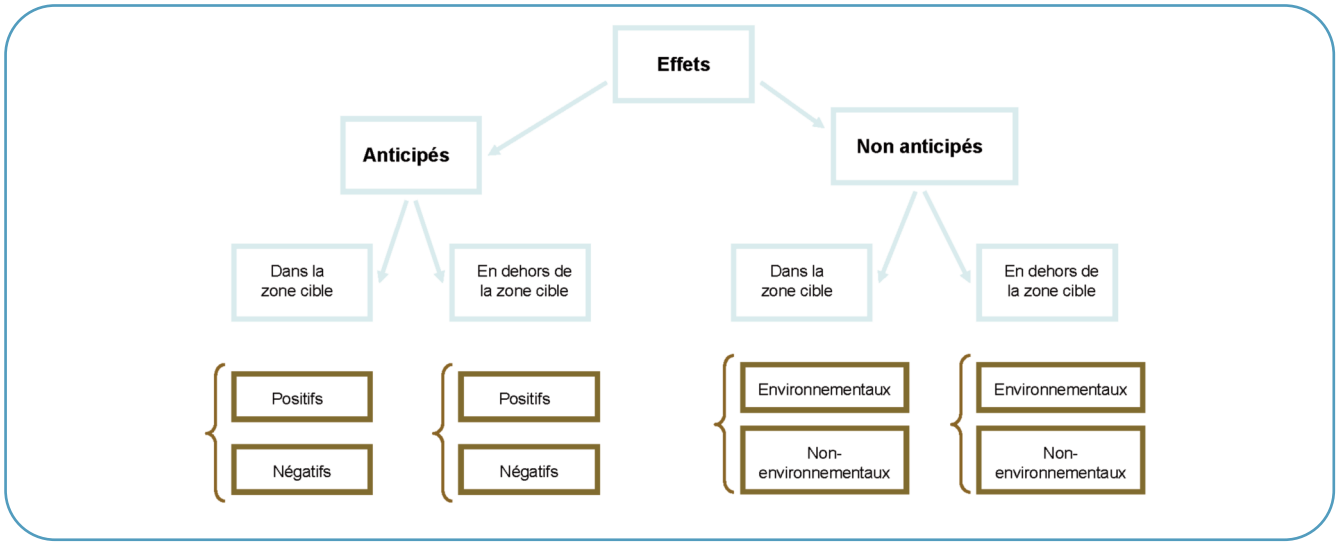
La première approche théorique à avoir profondément remis en question la posture traditionnelle est l'évaluation « *libérée des objectifs* » (*goal-free*), proposée par Scriven (1972). Elle invite à prendre de la distance avec la logique du programme et à l'évaluer délibérément sans *a priori* sur les objectifs à atteindre. Elle répond à deux préoccupations : 1) se référer aux effets réels, tangibles, plutôt qu'aux effets présumés, 2) éviter le problème classique de l'évaluateur confronté à des objectifs manquant de clarté et équivoques (Crabbé et Leroy, 2008 ; Stufflebeam, 2001).

Une autre motivation pour porter un regard critique sur les objectifs du programme à étudier, tient à la « *réduction de complexité* » opérée par les programmes. Dans le domaine de l'environnement, cette réduction de la complexité peut par exemple consister, dans le cadre des politiques de lutte contre le changement climatique, à « résumer » les écosystèmes à des stocks (et des puits) de carbone. Comme le soulignent Crabbé et Leroy (2008) « *une option pour le chercheur est de prendre pour argent comptant cette réduction de la complexité. En faisant ainsi, [il] adopte en fait, la réduction opérée par le décideur, même si c'est dans cette réduction que réside possiblement le motif d'échec de la politique. L'autre option pour le chercheur, est de souligner et d'analyser cette réduction de la complexité, y compris tous les choix subséquents associés* » (Crabbé et Leroy, 2008, p. 33)[17].

Les évaluations centrées sur les « *effets secondaires* » (*side-effects*) (Vedung, 2005) ou « *side-effectiveness* » (Gysen *et al*., 2006) abordent, à leur façon, ce problème de la réduction de la complexité. Comme leurs noms l'indiquent, elles portent leur attention sur les effets secondaires non escomptés par le programme, qui se révèlent au fur et à mesure de sa mise en œuvre et qui sont, d'une certaine façon, hors de son champ (cf. schéma 5). Pouvant être négatifs comme positifs, anticipés ou non, ils questionnent une évaluation qui s'intéresserait uniquement aux effets attendus, mentionnés dans les objectifs du programme.

[17] *"One option for the researcher is to take the complexity reduction for granted. In doing so, the researcher, in fact, adopts the reduction made by the policy-makers even though this is where the reason for policy failure may lie. The other option is for the researcher to emphasize and analyse the complexity reduction, including all subsequent choices".*

Schéma 5. Effets anticipés et non anticipés d'une politique environnementale



Source : adapté de Mickwitz (2003).

S'interroger sur la réduction de la complexité adoptée par le programme à évaluer ou s'intéresser uniquement à ses effets secondaires, n'est pas une position facile à assumer dans le sens où elle est très critique à l'égard du programme à évaluer. De telles évaluations ont par contre l'intérêt de mettre en contexte le programme et de le resituer dans des enjeux plus larges que ses propres objectifs. Elles permettent d'associer implicitement une évaluation de la pertinence et de la cohérence externe à une évaluation de l'efficacité. Pour peu que l'évaluation soit conduite « chemin faisant », il est alors possible de réorienter le programme, de telle sorte qu'il prenne mieux en considération ces questions qui auraient pu être considérées comme secondaires ou annexes.

Nous verrons dans l'étude de cas, section 3, comment les programmes de réduction de la déforestation, développés dans le cadre de la lutte contre le changement climatique et qui s'appuient sur un standard « carbone », peuvent-être examinés de façon critique, au vu de leurs effets potentiels sur la biodiversité et de leur cohérence avec d'autres engagements internationaux relatifs aux forêts, comme la Convention sur la diversité biologique (CDB). Nous verrons alors quel recadrage peut-être proposé pour évaluer des programmes, comme le FPCF.

Qui est alors légitime pour fixer les nouveaux objectifs, le nouveau référentiel au regard duquel évaluer le programme ? Comment porter cette démarche au-delà des controverses relatives aux concepts environnementaux, comme la gestion durable, la dégradation des forêts (Simula, 2009) etc., et au-delà des conflits de vision du monde, qui caractérisent le champ de l'environnement ?

2.2.2. Définir des objectifs propres à l'évaluation

Lorsque l'évaluateur opte pour une distanciation à l'égard des objectifs fixés par le programme, il lui faut définir des objectifs propres à l'évaluation. Quel niveau de délibération va-t'il déployer ? A quelle source va-t-il faire appel pour établir le référentiel ?

Certaines approches conçoivent le référentiel de l'évaluation comme devant refléter des préoccupations sociétales qui dépassent le cadre du programme – c'est le cas de

l'évaluation « *fondée sur les besoins* » (*needs-based evaluation*) (Owen, 2007) ou de l'« *efficacité sociétale* » qui traduit « *le degré auquel le résultat est en accord avec des objectifs sociétaux plus larges* ». « *L'évaluateur va devoir choisir des objectifs et un jeu de critères fondés sur un ensemble de principes et de critères largement reconnus* » (Gysen *et al.*, 2006, p. 100)[18]. De telles approches s'apparentent à une évaluation combinée de la pertinence et de l'efficacité. La question se pose alors de savoir qui est en mesure de définir ces « *objectifs sociétaux* ». D'ailleurs, existent-ils des valeurs qui seraient unanimement et similairement partagées par l'ensemble des groupes constituant la société ou participant au programme ? Cela est discutable, notamment dans les domaines fortement conflictuels, où les divergences de préoccupations parmi les parties prenantes sont élevées, comme c'est le cas dans le champ de l'environnement.

Une option est d'endosser les préoccupations ou les attentes d'un groupe de parties prenantes affectées par le programme, comme l'envisagent les modèles « *orientés vers les clients* » (*client-oriented model*) ou plus généralement « *orientés vers les parties prenantes* » (*stakeholder approach*) (Vedung, 2005).

Une autre option pour construire le référentiel de l'évaluation, consiste à se référer à des travaux académiques et d'expertise ainsi qu'à des engagements généraux (internationaux ou nationaux) (Mermet *et al.*, 2010 ; Underdal, 1992). C'est une façon de combiner des évaluations de la pertinence, de la cohérence externe et de l'efficacité du programme. Dans l'approche de Mermet *et al.* (2010) « *centrée sur la préoccupation* » environnementale (*concern-focused evaluation*), la question de savoir quels groupes d'intérêt ou quelles institutions s'engagent réellement pour rechercher l'efficacité au regard du référentiel établi, n'est pas posée au moment de son établissement. Elle est renvoyée aux étapes ultérieures de l'évaluation et du débat politique. Une telle démarche vise ainsi à ne pas priver d'évaluation des politiques qui, à ce jour, ne sont pas encore assez soutenues d'un point de vue social, administratif, politique ou organisationnel.

Ces différentes options illustrent bien que le critère d'efficacité dépend du point de vue de l'évaluateur et qu'il peut être défini de façon extrêmement différente d'une évaluation à l'autre, selon un mode plus ou moins délibératif. Comme le souligne Young, « *les jugements relatifs à l'efficacité d'un régime peuvent varier selon comment ceux qui les formulent, décident de poser et de cadrer le problème* » (Young, 1994, p. 143)[19]. Il y a là une source majeure de différentiation méthodologique entre approches de l'évaluation des politiques environnementales, mais aussi de différentiation de contenu entre les travaux d'évaluateurs qui auraient adopté un cadre méthodologique similaire, mais choisi des référentiels différents.

Nous proposons de retenir l'option qui consiste à construire le référentiel à partir de travaux académiques en environnement et d'engagements politiques pertinents, ce qui permet, à notre sens, de placer au cœur de l'évaluation le problème environnemental qui a suscité la politique.

2.2.3. Adopter une posture en termes de *résolution du problème* environnemental en jeu

Comme il a été esquissé, une alternative s'offre à l'évaluateur qui n'endosse pas les objectifs du programme comme unique projet d'évaluation (Mitchell, 2008 ; Underdal, 1992, 2002 ; Young, 1994). Il peut chercher à savoir si le programme « fait une différence » avec ce qui se serait passé sinon, autrement dit, si l'intervention a permis d'améliorer la situation. C'est une sorte de comparaison « *situation réelle contre situation contrefactuelle* »[20] (*actual-versus-counterfactual*) (Mitchell, 2008). Mais il peut aussi chercher à savoir si le problème à l'origine du programme est désormais résolu, ou

[18] *"The evaluator will have to choose the goals and the set of criteria based on a set of widely accepted societal principles and goals".*

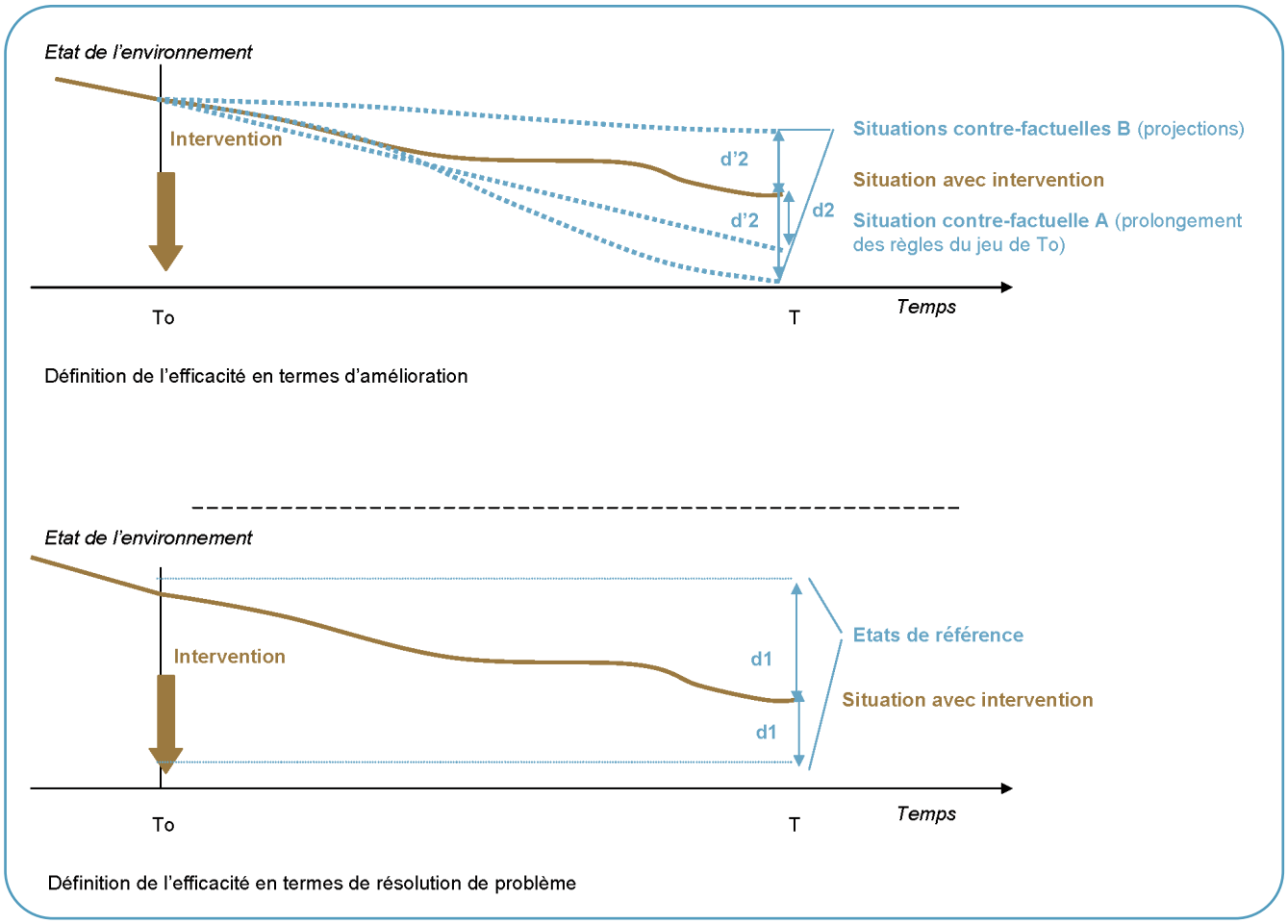
[19] *"Needless to say, judgments regarding the effectiveness of a regime may vary depending on how those making such judgments choose to frame the problem".*

[19] Dans son glossaire, l'OCDE définit comme suit les situations contrefactuelles *: « Situations ou conditions dans lesquelles se trouveraient les personnes, les organisations ou les groupes concernés si l'action de développement [le programme] n'existait pas »* (OCDE, 2002).

s'il est en bonne voie de l'être. Il s'agit alors d'une approche en termes de résolution de problème. On pourrait l'assimiler à une comparaison « *situation réelle contre situation souhaitée* » (*actual-versus-aspiration*) (Mitchell, 2008). La différence entre les deux approches réside dans la référence choisie : dans le premier cas, c'est une situation hypothétique qui aurait eu lieu sans le programme, dans le second, c'est un horizon visé.

Dans l'étude de cas, section 3, cette question est particulièrement importante, car le programme étudié adopte lui-même un de ces deux partis pris. En effet, les pays bénéficiaires du FPCF doivent établir des scénarios de référence de ce que deviendrait leur déforestation sans politique incitative, et c'est précisément par rapport à ces scénarios qu'ils seront rémunérés en fonction des efforts qu'ils justifieront. Il est ainsi utile que l'évaluateur soit en mesure d'expliciter clairement l'alternative évaluative qui se pose à lui. Soit il suit le FPCF dans cette approche de l'efficacité – ce qui ne l'empêche pas d'analyser le bien-fondé des scénarios établis – soit il opte pour une évaluation à horizon fixé. Nous verrons les avantages et inconvénients de ces deux démarches.

Schéma 6. Options possibles pour le choix du standard



Source : les auteurs.

- Si l'on considère le schéma 6, le premier cas (schéma supérieur) illustre un objectif exprimé en termes d'amélioration. Le standard de référence est la situation hypothétique qui aurait eu lieu en l'absence de programme. L'efficacité correspond à la distance mesurée entre la situation observée avec intervention et cette situation contrefactuelle, qui peut être déterminée :

  i) soit en faisant des projections (situations nommées contrefactuelles B) : distances d'2. L'efficacité est ici exprimée en termes absolus ;

  ii) soit en prolongeant les règles du jeu qui prévalaient avant l'intervention (situation nommée contre-factuelle A) : distance d2. L'efficacité est ici exprimée de façon incrémentale.

  Si l'option *i)* semble être la plus séduisante, étant la traduction exacte, en théorie, de ce qui ce serait passé sans le programme, elle est en fait souvent très difficile à mettre en œuvre, en pratique (Mitchell, 2008 ; Underdal, 1992). En effet, selon les sujets considérés, de nombreuses variables peuvent influer et il est très difficile de les envisager toutes et de prédire leur évolution avec certitude (c'est pourquoi nous avons représenté schématiquement deux projections contre-factuelles B, possibles). L'option *ii)* apparaît plus facile, mais le problème réside dans le fait qu'elle ne traduit pas toujours de véritables évolutions (cf. encadré 4). Il y a en effet aucune raison de penser que des tendances passées vont se poursuivre de façon linéaire dans le futur. Passé un certain seuil, la situation peut, parfois empirer jusqu'à « *un effondrement* » (Diamond, 2006) ou parfois se stabiliser, par exemple lorsque les coûts marginaux d'exploitation de la ressource deviennent trop élevés. Notons qu'en posant ainsi la question de l'efficacité et en y répondant en ces termes, l'attribution des effets au programme (qui est une des questions épineuses de l'évaluation) est implicite.

- Dans le second cas (cf. schéma 6, « Définition de l'efficacité en termes de résolution de problème »), l'objectif de l'évaluation est exprimé en termes de résolution de problème. Le standard de référence est alors à définir. Il peut s'agir d'un état idéal recherché, d'un « *optimum collectif* » (Underdal, 1992, p. 230) ou d'un « *état désirable de l'environnement* » (Mermet *et al.*, 2010). L'efficacité ainsi définie correspond à la distance d1, distance entre la situation avec intervention et l'état de référence.

  Dans ce cas, notons que l'on ne peut pas se prononcer quant à la responsabilité du programme vis-à-vis du rapprochement (ou de l'atteinte) de la référence visée. Pour parler d'efficacité du programme ou encore de contribution du programme, il faut se donner les moyens d'inférer la causalité et de reconstituer les enchaînements causaux qui ont conduit aux effets. C'est ce qui permet de justifier que la situation *avec intervention* est bien, avant tout, due au programme. Resituer le programme dans un système de gestion plus large, qui inclut l'ensemble des autres dispositifs impactant l'objet d'étude (positivement comme négativement), permet de mieux appréhender l'action propre du programme et ses interactions avec les autres actions (Mermet *et al.*, 2010).

Encadré 4

Etablir des scénarios de référence de la déforestation : enjeux méthodologiques et théoriques

Nous examinons l'alternative qui se pose pour établir des scénarios de référence de la déforestation : projeter des taux de déforestation futurs ou prolonger des tendances passées.

Les scénarios du premier type sont plus difficiles à reconstituer que ceux du second type car ils impliquent 1) d'identifier les variables influant sur l'évolution du couvert forestier et 2) de prédire leurs évolutions et leur effet conjoint. Or, les facteurs de déforestation sont multiples et impliqués dans de nombreuses chaînes causales. De plus, les dynamiques des écosystèmes forestiers ne sont pas linéaires : par exemple, au-dessous d'une certaine superficie, une forêt soumise à une pression inchangée disparaîtra plus rapidement car ses qualités de résilience auront diminué. Déterminer quelle va être l'évolution du couvert forestier est alors un exercice difficile, controversé et fortement sujet au détournement politique (Karsenty, 2008). C'est pourtant à cet exercice que s'attèlent les pays forestiers engagés dans la réduction de la déforestation dans le cadre de la lutte contre le changement climatique (REDD+)[21]. Si ces scénarios ne sont pas conservateurs, il est à craindre que la comptabilisation de réductions d'émissions fictives (non additionnelles) conduise à une politique en fait inefficace.

Pourquoi alors ne pas opter pour la méthode correspondant au prolongement des tendances, la plus aisée à mettre en œuvre ? Cette méthode induit un biais en surestimant des taux de déforestation futurs, pour les pays ou régions connaissant des taux de déforestation actuellement élevés. En effet, si l'on se fie à la courbe de transition forestière proposée par Rudel (2005), couramment citée, des taux de déforestation élevés vont très probablement décroître pour des questions d'accessibilité et d'accroissement des coûts marginaux d'exploitation des forêts résiduelles. De plus, ce biais ne s'applique pas de façon homogène à tous les pays : ceux qui ont beaucoup déforesté par le passé sont favorisés par rapport à ceux dont les taux de déforestation sont actuellement bas mais qui risquent d'augmenter. C'est pourquoi des approches mixtes avec des facteurs d'ajustement sont envisagées.

Ces deux types de standards (*i.e.* la situation sans programme ou un état visé désirable ou optimal) sont tous deux pertinents pour évaluer l'efficacité (Underdal, 1992). Nous avons souligné leurs avantages et inconvénients respectifs d'un point de vue méthodologique. Il est par ailleurs important de noter que le choix de l'un ou l'autre de ces standards, traduit une posture d'évaluation plus ou moins normative.

Dans la mesure où l'évaluateur a les moyens de reconstituer le processus conduisant aux effets, nous invitons plutôt à adopter une posture en termes de résolution de problème. Et ce, d'une part, car les scénarios contrefactuels sont difficiles à reconstituer et sont donc objet de vives négociations qui remettent largement en question leur caractère prédictif, et, d'autre part, car des évaluations en termes d'amélioration ne permettent pas, selon nous, de jouer un rôle d'alerte suffisant face à l'intensité des processus de dégradation de l'environnement. Pour certains problèmes environnementaux d'ailleurs, une amélioration peut se traduire *in fine* par un échec. C'est le cas lorsqu'il existe des seuils d'irréversibilité, comme pour une espèce en voie d'extinction. A la suite d'une intervention visant à protéger l'espèce, l'augmentation du nombre d'individus peut être insuffisante et le nombre d'individus demeurer au-dessous du seuil de survie, de sorte que l'espèce poursuit sa courbe d'extinction, malgré une amélioration de son effectif. Dans cette situation, seule une mesure drastique visant l'atteinte d'une taille minimum critique de la population peut laisser espérer la résolution du problème d'extinction.

[21] REDD+, pour Réduction des émissions liées à la déforestation et à la dégradation des forêts et activités de gestion forestière durable, de conservation et d'augmentation des stocks de carbone forestiers dans les pays en développement.

Dans une évaluation, en termes de résolution de problème, l'évaluateur sera confronté dans la plupart des cas à des programmes n'atteignant pas l'horizon visé. C'est pourquoi, l'évaluation consistera en fait à apprécier l'évolution depuis la mise en œuvre du programme, et la distance restant à parcourir pour atteindre l'objectif visé. Il est ainsi indispensable de démontrer que ces tendances sont bien le fait du programme. C'est ce sur quoi nous nous pencherons, section 2.4.

2.3. Les effets à évaluer

Si quelques auteurs font allusion à des évaluations de *processus* à des fins d'évaluation d'efficacité (Young, 1994 avec l'« *efficacité constitutive* » et l'« *efficacité de processus* », Vedung (2005) avec l'« *évaluation de processus* »), c'est avant tout dans le but d'expliquer comment la politique a fonctionné, générant les effets observés. Sans être une fin en soi, l'étude du processus de mise en œuvre est le moyen de mieux informer les effets – ce qui constitue la préoccupation spécifique de l'évaluation de l'efficacité – tout en soulevant des questions. Vers quels effets orienter l'évaluation ? Peut-on évaluer séquentiellement les réalisations, les résultats et les impacts ? Quelles circonscriptions spatiale et thématique donner à l'évaluation, alors que les problèmes environnementaux et les politiques qui leur sont associées sont souvent multiscalaires et ont des implications environnementales mais aussi sociales fortes ? Faut-il envisager un cadrage flexible qui soit réceptif à d'éventuels effets secondaires ?

2.3.1. Des évaluations souvent centrées sur les réalisations et les résultats

Les évaluations *ex post*, ou rétrospectives *back end*, ont longtemps été majoritairement pratiquées, puis les évaluations *ex ante* se sont développées (Patton, 2008, p. 334).

Les instruments politiques récemment introduits, attirent aujourd'hui l'attention de chercheurs en évaluation. Leur évaluation apparaît intermédiaire : « *[elle] diffère de l'évaluation traditionnelle ex post car les résultats et les effets non anticipés ne sont pas disponibles, la plupart du temps (Kautto et Similä, 2005). Elle diffère également de l'évaluation ex ante car certains éléments de la mise en œuvre et certaines hypothèses sous-jacentes peuvent souvent être testés empiriquement sur la base des réalisations disponibles* » (Mickwitz, 2003)[22].

Si l'idée d'évaluer l'efficacité de tels instruments en phase de mise en œuvre, dans la perspective de les améliorer, semble bonne, on peut toutefois se demander si le stade précoce le permet. Pour ces auteurs, « *bien qu'une analyse d'impacts […] soit impossible, les résultats ne s'étant pas encore produits, cela ne signifie pas nécessairement que l'usage du critère d'efficacité soit aussi impossible […]* » (Gysen *et al.*, 2006 ; Kautto et Similä, 2005).

Cet exemple particulier des instruments politiques récemment introduits fait plus largement écho à nombre de politiques, de programmes ou de régimes internationaux qui génèrent des impacts tardifs, souvent de par la nature du problème en jeu. C'est le cas de nombreux problèmes d'environnement, comme nous l'avons vu section 2.1. Le moment auquel est réalisée l'évaluation de l'efficacité est donc un élément déterminant du cadrage de celle-ci (cf. Encadré 1). Si les impacts sont la préoccupation dominante, l'évaluation est toutefois bien souvent décentrée vers les réalisations et les résultats (les changements de comportement des groupes cibles), seuls effets que les évaluateurs sont à même

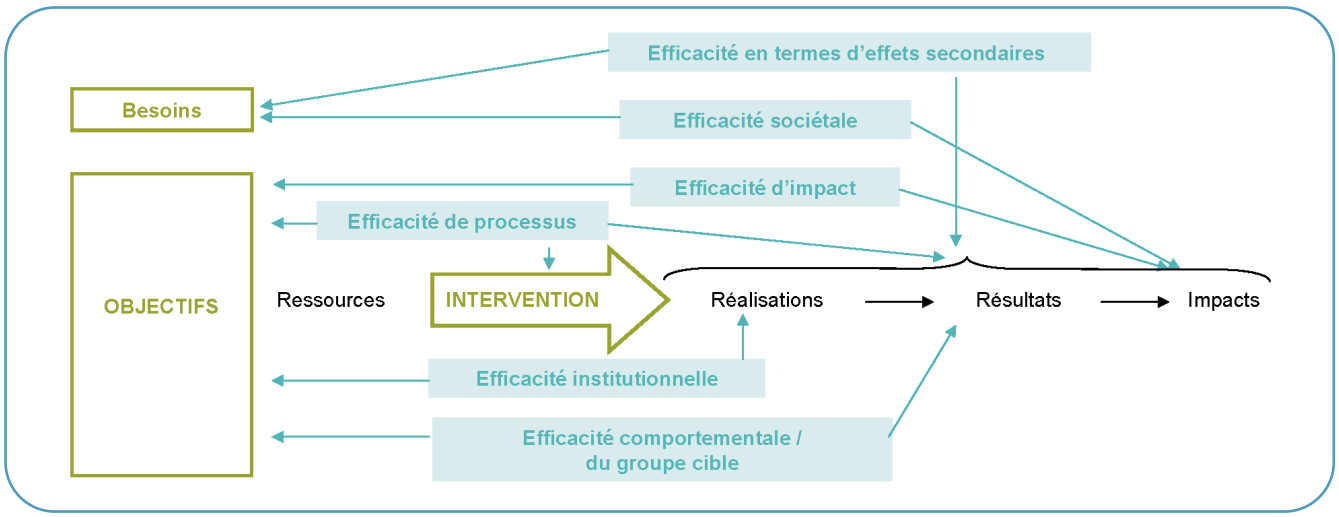
[22] *"RIPI evaluation differs from traditional* ex post *evaluation because outcomes and unanticipated effects are mostly unavailable* (Kautto et Similä, 2005). *It also differs from* ex ante *evaluation because there are often some implementation practices and underlying assumptions that can be empirically tested based on the available experiences".*

d'appréhender. Par exemple, l'efficacité des aires protégées pour conserver la biodiversité pourra être abordée *via* l'analyse de l'évolution des pressions qui s'exercent sur les écosystèmes forestiers considérés (activités agricoles, extraction de bois, feux, etc.) à défaut d'être appréhendée directement sur l'état de l'écosystème (Bruner *et al.*, 2001).

Orienter l'évaluation des impacts vers l'évaluation des réalisations et des résultats, est l'approche retenue par la littérature portant sur l'efficacité des régimes environnementaux internationaux (Haas *et al.*, 1993 ; Miles *et al.*, 2002 ; Mitchell, 2008 ; Tikina et Innes, 2008 ; Underdal, 1992 ; Victor *et al.*, 1998 ; Young, 1994, 1999). Ce choix y est parfois présenté comme un choix par défaut, justifié par l'absence de données suffisantes et satisfaisantes concernant les changements de l'environnement, qui étaient la préoccupation initiale (traduit de Haas *et al.*, 1993, p. 7). Mais il est aussi présenté comme un choix délibéré, et qui plus est « le *plus approprié pour évaluer la performance*» (traduit de Mitchell, 2008, p. 84)[23]. Pour certains auteurs, « *porter son attention sur les changements de comportement a des avantages, y compris pour ceux qui sont convaincus que la qualité de l'environnement est l'unique métrique valide pour mesurer la performance institutionnelle* » (traduit de Mitchell, 2008, p. 84)[24]. Un des arguments mis en avant tient au fait que les changements de comportement sont une condition nécessaire (même si elle est insuffisante) à une amélioration de l'état de l'environnement. « *Pour que des institutions internationales aient un effet en faveur de l'environnement, il faut qu'elles induisent des changements politiques et par conséquent, il est approprié de les juger en fonction de la façon dont elles réalisent cela* » (Haas *et al.*, 1993, p. 396)[25].

Le schéma 7 présente quelques types d'efficacité définis dans la littérature, en fonction, d'une part, des effets auxquels ils se rapportent et, d'autre part, du standard de référence auquel ils se réfèrent.

Schéma 7. Différents types d'efficacité



Source : les auteurs, selon Vedung (2010), Gysen et al. *(2006), Young (1994, 1999).*

[23] *"For various types of institutions, behavioural change may be the most appropriate dimension in which to evaluate performance rather than a second best alternative to environmental quality".*

[24] *"Focussing on behaviour has advantages even for those committed to the view that environmental quality is the only metric of institutional performance".*

[25] *"For international institutions to make a difference environmentally, they must spawn political change, and therefore it is appropriate to judge them according to how well they do so".*

2.3.2. Comment confronter des effets préliminaires, à un référentiel correspondant aux impacts souhaités *in fine*

Un dilemme se pose. Si les impacts ne sont pas visibles, faut-il définir le standard de référence dans les mêmes termes que les effets procéduraux et comportementaux sur lesquels portera l'évaluation ? Dans ce cas, le standard de référence fera allusion à des procédures, à des types de réalisation (la pratique de l'expertise, l'organisation d'ateliers participatifs, etc.) ou à des changements de comportements. Ou bien, faut-il conserver comme aiguillon de l'évaluation, un standard qui a trait aux impacts finaux (sur l'état de l'environnement 1), préoccupation de fond ? Mais il ne sera alors pas de même nature que les effets visibles.

La première option est plus faisable d'un point de vue méthodologique mais la seconde est directement liée à la préoccupation environnementale qui est à l'origine de la politique et de l'évaluation et a donc *a priori* une plus grande pertinence conceptuelle et une robustesse politique (Mermet, 1996).

Cette question est centrale car dans de nombreux cas, il existe bel et bien un risque, en focalisant l'attention sur *les réalisations* d'un programme ou sur les normes créées par un régime : celui d'affaiblir la vigilance portée à l'amélioration de l'état de l'environnement, résultat visé *in fine* (Billé, 2007 ; Young, 1994). Ceux qui peuvent plaider pour une approche portant sur les résultats reconnaissent que ce n'est pas suffisant : « *malgré l'intérêt de se référer aux changements de comportements comme unique dimension pour l'évaluation de la performance institutionnelle, le résultat s'avère souvent insatisfaisant* » (traduit de Mitchell, 2008, p. 84)[26]. En effet, nombreuses sont les évaluations environnementales qui concluent de façon très positive en termes de moyens, comme « un nombre additionnel de *X* aires protégées a été décrété » ou encore « Y ateliers ont été organisés en trois mois », sans que la situation de gestion de ces nouvelles aires ne fasse l'objet de l'évaluation (quel état vise-t-on pour ces habitats à protéger ?) ni que la finalité des ateliers ne soit mentionnée. Le choix des indicateurs (de moyen ou de résultat) est ici critique.

Ce type de conclusion est tellement répandu, que cela nous invite à conserver au cœur du standard de l'évaluation la référence à l'état de l'environnement désiré. Il nous faut voir alors comment dépasser l'incohérence entre les effets étudiés et le standard retenu.

Une façon de contourner cette difficulté est de regarder comment la préoccupation environnementale contenue dans le référentiel, est *intégrée* tout au long du processus de mise en œuvre du programme (Mickwitz et Kivimaa, 2007). Dans quelle mesure les réalisations et les résultats portent-ils les objectifs environnementaux du référentiel ? Quels acteurs tentent de favoriser leur prise en compte (Mermet *et al.,* 2010) ? Si l'on adhère à l'hypothèse que, faire mention et intégrer une préoccupation environnementale à des stades précoces de la mise en œuvre augmente la probabilité de parvenir *in fine* à de tels impacts, alors cette analyse est une bonne manière d'évaluer l'efficacité et de contourner le problème d'incohérence entre notre référentiel et les effets observés.

Une option complémentaire, est de porter une attention particulière sur certains facteurs censés améliorer l'efficacité aux stades des réalisations et des résultats. Les travaux sur l'efficacité des régimes environnementaux qui s'intéressent précisément à cette question peuvent guider l'évaluateur dans cette démarche (cf. encadré 5), même si le nombre de facteurs identifiés dans ces travaux est très élevé, de l'ordre de la dizaine, voire de la cinquantaine (Miles *et al.*, 2002), et peut être déstabilisant (Mitchell, 2008, p. 92).

[26] *"Despite the virtues of using behavior change as the only dimension for evaluating institutional performance, the result often proves unsatisfying".*

Encadré 5

L'approche adoptée par les travaux portant sur l'efficacité des régimes environnementaux
(Haas *et al.*, 1993 ; Miles *et al.*, 2002 ; Mitchell, 2008 ; Tikina et Innes, 2008 ; Underdal, 1992 ; Victor *et al.*, 1998 ; Young, 1994, 1999)

Les travaux portant sur l'efficacité des régimes s'appuient sur une approche commune consistant à mettre en équation l'efficacité du régime (variable dépendante à expliquer : Y) avec un certain nombre de variables indépendantes (variables explicatives : X), selon le modèle simple : Y= f(X). Y est une variable de nature mixte, à la fois état de l'environnement (comme les *impacts*) et changement de comportements (comme les *résultats*). Elle est définie sur dires d'experts (c'est-à-dire que le caractère plus ou moins efficace du régime est fixé). Quand aux variables *explicatives*, elles correspondent à des facteurs intervenant aux stades *réalisations* voire *résultats*, tels : la capacité politique et administrative des gouvernements nationaux et la mobilisation critique de la société civile, l'environnement contractuel favorable à la mise en oeuvre et au maintien des accords, le niveau de préoccupation et de conscience gouvernementale associé à des réseaux actifs d'individus et de groupes, etc. (Haas *et al.*, 1993).

Le but de ces travaux est de déterminer lesquelles de ces variables X expliquent le mieux cette efficacité (ou cette absence d'efficacité). Pour cela, plusieurs études de cas sont considérées et des tests statistiques comparatifs sont effectués.

2.3.3. Effets objectifs ou subjectifs

Que ce soit au stade des réalisations, des résultats ou des impacts, il existe plusieurs façons de « saisir » les effets (Fitzpatrick *et al.*, 2004). Ces approches se distinguent principalement par leur caractère plus ou moins délibératif et constructiviste. D'un côté du spectre, on trouvera des méthodes qui reposent sur la simple observation de l'évaluateur, ou sur l'expertise. De l'autre côté, on trouvera des méthodes participatives où les évaluateurs ne sont que les facilitateurs de l'évaluation et où ils recueillent les perceptions des parties prenantes et leur appréciation des effets. En position intermédiaire, on trouvera les méthodes couramment employées en évaluation, reposant principalement sur les entretiens ; la « construction » des effets est alors le fait des subjectivités de la personne interrogée et de l'évaluateur. Ces quelques exemples volontairement archétypiques montrent qu'une forme de subjectivité est derrière toutes les méthodes, même les méthodes quantitatives, avec la question de savoir qui renseigne l'indicateur et comment.

2.3.4. Choisir une échelle spatiale adaptée au programme étudié

Les politiques environnementales visent en général à réduire des processus dommageables de dégradation de l'environnement (pollutions, destruction d'écosystèmes, surexploitation d'une ressource, etc.). Le risque existe cependant qu'elles ne résolvent pas réellement le problème auquel elles s'attaquent : si les pressions persistent, le problème pourra se déplacer simplement depuis l'endroit où l'action publique est mise en œuvre vers un autre où elle ne l'est pas (on parle d'effet de fuite). Prendre en compte ces questions de portée spatiale est donc un enjeu non seulement pour la politique mais aussi pour l'évaluation qui risque sinon de mal apprécier l'efficacité. Par exemple, étudier l'efficacité d'aires protégées, en ne prêtant attention qu'aux écosystèmes présents dans les limites de ces espaces, peut conduire à surestimer l'efficacité si les pressions se sont déplacées à l'extérieur des zones mise en protection (Andam *et al.*, 2008).

En plus de veiller à une portée géographique suffisante, un enjeu important des évaluations s'intéressant aux politiques qui abordent des problèmes environnementaux à grande échelle (notamment pour essayer de faire face à ce problème de fuites) est de prendre en compte l'ensemble des niveaux de mise en œuvre pertinent quant aux objectifs poursuivis (Stame, 2004). Une évaluation qui ne porterait que sur la dimension internationale d'une politique visant la réduction de la déforestation ne pourrait sans doute pas apporter de conclusions satisfaisantes en termes d'efficacité. A l'inverse, peut-on envisager de réaliser l'analyse du programme uniquement au niveau national ou local où il est mis en œuvre ? Peut-on alors raisonnablement penser englober l'ensemble des effets au niveau local ? Il s'agit donc de choisir la bonne portée de l'évaluation du point de vue des échelles verticales retenues. Ces évaluations confrontées à des aspects compliqués des politiques ne peuvent pas s'appuyer sur des méthodologies standardisées (Crabbé et Leroy, 2008, p. 39) et sont donc amenées à développer des protocoles originaux spécifiques à chaque cas (Rogers, 2008). Le recours aux études de cas est à ce propos une option possible.

2.4. L'attribution et l'explication des effets

Les approches visant à évaluer l'efficacité mettent en regard les résultats observés avec un référentiel. Cette mise en relation soulève une troisième question fondatrice de l'évaluation, déjà évoquée : Est-ce bien l'intervention (la politique, le programme, …) qui est responsable des effets que l'on observe ? Comment expliquer ces effets et comment les attribuer à cette intervention en particulier ?

Contrairement aux évaluations en termes d'amélioration, nous avons vu que l'évaluation en termes de résolution de problème, n'intègre pas, l'inférence de la causalité. L'explication et l'attribution des effets sont donc des enjeux d'autant plus importants pour ce type d'évaluation de l'efficacité de politiques environnementales ; ils sont par ailleurs particulièrement délicats à analyser, ces politiques et leurs effets se déployant en général dans des contextes complexes et interférant souvent avec d'autres interventions publiques qui compliquent encore singulièrement l'évaluation (Mermet *et al.*, 2010).

Pendant longtemps, l'attention des praticiens a porté sur les effets et sur leurs *corrélations* avec les objectifs fixés, sans véritablement utiliser de méthodes permettant de démontrer des liens causaux (Andam *et al.*, 2008 ; Ferraro, 2009). Aujourd'hui, la question de la « boîte noire » (*black box*) ou de l'inférence de la causalité est centrale en évaluation. Et si, « *dans les situations complexes du monde réel, aucune évaluation ne peut répondre à toutes les incertitudes et déterminer de façon non-ambiguë jusqu'à quel point les développements observés sont imputables aux instruments politiques évalués* » (traduit de Mickwitz, 2003, p. 433)[27], il n'est pourtant plus envisageable d'éluder ce point, quelle que soit la démarche évaluative empruntée. De nombreux développements théoriques et méthodologiques s'intéressent spécifiquement à cette question de la causalité.

Trois grands types d'approches se dégagent, chacune de ces approches faisant davantage appel à des aspects quantitatifs ou qualitatifs :

- les expériences et quasi-expériences,
- la construction de scénarios contrefactuels
- la reconstitution des théories d'intervention et l'analyse de processus.

[27] *"In complex real-world situations no evaluation can resolve all the uncertainties and unambiguously determine to what degree the observed development is due to the evaluated policy instruments".*

2.4.1. Expériences aléatoires et quasi-expériences

Une méthode souvent promue pour attribuer les effets observés, s'inspire des sciences expérimentales. Elle consiste à appliquer aléatoirement le traitement (la politique ou le programme)[28] sur un échantillon de sujets éligibles. La conclusion sur l'efficacité est formulée par simple comparaison entre les sujets soumis au traitement et ceux qui n'en ont pas bénéficié et qui constituent donc le « témoin ». L'échantillonnage aléatoire permet en effet, d'induire une variation exogène telle que l'affectation du programme ne soit pas en principe corrélée aux résultats (Ferraro, 2009, p. 80). L'explication des résultats peut alors être recherchée en réalisant des modèles économétriques et en testant le caractère significatif de diverses variables explicatives mobilisées.

Mais l'affectation aléatoire du traitement n'est pas toujours possible, ni d'ailleurs souhaitable. De plus, les politiques et les programmes ne sont pas conçus ou mis en œuvre dans le but d'offrir les conditions optimales d'une évaluation. Ce qu'on pourrait appeler « des biais de sélection » (si la sélection aléatoire est considérée comme idéale), sont fréquents. Par exemple, la création d'aires protégées a rarement lieu sans l'intention initiale de préserver une zone fortement soumise à des pressions anthropiques ou bien présentant un grand intérêt du point de vue de sa biodiversité. Ainsi, la politique de conservation n'est, en principe pas aléatoire, et ce, y compris lorsque la création de l'aire protégée répond à des enjeux qui ne sont pas d'ordre écologique mais politique ou autre (Andam *et al.*, 2008).

La « quasi-expérience » est une méthode alternative. Elle repose sur le même principe que l'expérience aléatoire, mais son protocole est moins strict. L'affectation du traitement n'y est en effet pas aléatoire. En conséquence, les résultats observés sont attribuables à d'autres facteurs que l'intervention évaluée. Aussi, le caractère explicatif de ces facteurs suspectés pouvoir être à l'origine des changements observés est testé, notamment par des modèles économétriques (Ferraro, 2009, p. 80). Une méthode « d'appareillement par jumeaux » est aussi utilisée dans le cadre de ces quasi-expériences. Elle vise à comparer statistiquement deux sujets aux caractéristiques proches, l'un des deux ayant bénéficié du programme à évaluer et l'autre étant le témoin.

Même si leur protocole est simplifié par rapport à celui des expériences aléatoires à proprement parlé, les quasi-expériences restent souvent inexploitables. En effet, l'obtention de données quantitatives adéquates est souvent difficile, tout comme l'identification d'un nombre restreint de facteurs causaux non corrélés. Plus généralement, on peut reprocher à ces méthodes expérimentales de ne pas prendre en compte le contexte comme variable explicative, de se borner à rendre compte de l'efficacité ou de l'absence d'efficacité, mais sans englober l'explication du *pourquoi* (essentielle pour guider l'action en situation), et enfin de ne pas prêter attention aux effets secondaires (Crabbé et Leroy, 2008, p. 67).

[28] Désigné en anglais par *"Randomized Controlled Trial"* (RCT).

Encadré 6

Biais dans le protocole d'une évaluation de l'efficacité des aires protégées

Le protocole d'évaluation suivie par une étude visant à déterminer l'efficacité des aires protégées du point de vue de la conservation de la biodiversité (Bruner *et al.*, 2001) illustre que les quasi-expériences ne sont pas exemptes de biais.

L'évaluation porte sur un échantillon de 93 parcs ayant été sélectionnés dans des zones soumises à des fortes pressions anthropiques sur les forêts. Trois analyses sont conduites : 1) l'évolution de la déforestation au sein des aires protégées, depuis leur établissement, est retracée, 2) la situation dans les parcs est comparée à celles de leurs alentours, 3) des corrélations sont calculées entre le différentiel de déforestation « parc – entourage » et les caractéristiques de gestion associées à une protection effective des parcs (comme le nombre de gardes, le budget de fonctionnement du parc, etc.)[29]. L'article conclut que globalement les aires protégées sont des outils de gestion efficaces pour protéger la biodiversité.

Dans le protocole, les alentours du parc sont envisagés comme la référence contrefactuelle, comme le témoin (ce qui se serait passé s'il n'y avait pas eu d'aire protégée). Mais ce témoin et l'objet de l'intervention (la zone sous statut d'aire protégée) ne sont pas indépendants. Il est ainsi possible que la diminution des pressions dans le parc soit à l'origine d'un accroissement des pressions autour du parc (problème des fuites, traité dans la partie portant sur le périmètre de l'évaluation). Ainsi, globalement, sur des zones plus larges que les seules aires protégées, il n'est pas certain que la déforestation ait diminué. En termes d'enseignements, pour l'évaluation, cet exemple illustre l'importance d'étudier en détail les processus dommageables visés par l'intervention et d'être en mesure de les localiser. En termes d'enseignements pour le projet, cet exemple souligne l'importance du champ géographique choisi. Dans le cas d'une aire protégée, il est essentiel d'étendre la portée du projet à la zone tampon (en périphérie de la zone protégée) avec tout type de mesures d'accompagnement.

2.4.2. Construire des univers contrefactuels

Nous l'avons vu, la reconstitution d'une situation hypothétique, contrefactuelle, est une référence possible. Elle permet d'évaluer l'efficacité en termes d'amélioration car la situation contrefactuelle correspond à la situation hypothétique qui aurait existé s'il n'y avait pas eu de programme. Cette posture est conceptuellement attrayante, dans la mesure où elle permet à la fois de définir le référentiel de l'évaluation et d'attribuer les effets. Mais quelle que soit la situation contrefactuelle prise comme référence (projections ou prolongement des tendances passées), nous avons mis en évidence que les problèmes sont majeurs, à la fois sur les plans méthodologique et conceptuel, s'agissant finalement d'asseoir des décisions aux conséquences importantes, et sur des scénarios ou des projections dont la construction relève inévitablement de choix de construction fortement sujets à controverse, et souvent politiques (Mermet, 1996).

2.4.3. Reconstituer les théories d'intervention et analyser les processus de mise en œuvre

Une autre manière d'aborder l'analyse de l'efficacité est, d'une part, de mettre en œuvre des approches fondées sur l'étude fine, sur le terrain, du fonctionnement et des effets de la politique, et, d'autre part, de confronter ces observations de terrain avec les hypothèses et les principes qui ont guidé le choix de la politique ou qui guident sa mise en oeuvre. Rogers (2008) appelle ainsi à inventer des protocoles innovants qui allient études de cas, visites de terrain et reconstitutions théoriques.

Plusieurs auteurs insistent sur l'importance de reconstituer la théorie d'intervention de la politique étudiée, pour parvenir à

[29] *"We assessed the effectiveness of these parks from three perspectives: land clearing within the boundaries of parks since establishment, current condition of parks compared with the condition of their surroundings, and factors correlated with effective park protection". "We assessed the effectiveness of parks at preventing land clearing by comparing the current extent of clearing with clearing at the time of park establishment".*

attribuer ses effets (Gysen *et al.*, 2006 ; Kautto et Similä, 2005). Sur cette question, les terminologies abondent dans la littérature. Elles ont été proposées puis reprises dans des courants et des contextes de recherche divers (Patton, 2008, p. 336). Toutes ces approches (pour partie évoquées, section 1) renvoient néanmoins à la même idée de base. Rogers *et al.* (2000) proposent une définition commune : « *Les théories de programme, désignées tantôt par « théorie de programme », « logique de programme » […], « évaluation basée sur les théories » ou « théories de changement » […], « évaluation pilotée par la théorie »* (Chen, 1990)*, « théorie d'action » […], « logique d'intervention » […], se rapportent à une multitude de façons de développer un modèle causal, reliant les ressources et les activités à une chaîne de résultats attendus ou observés, et à l'utilisation de ce modèle pour guider l'évaluation* » (traduit de Rogers *et al.*, 2000)[30].

Le but de ces approches est de mettre à jour la théorie d'action de la politique étudiée. Il s'agira de voir comment elle envisage d'atteindre ses propres objectifs. Mais il est aussi intéressant de comprendre quelle(s) théorie(s) d'action lui prêtent les parties prenantes, c'est-à-dire comment ils pensent que la politique fonctionnera (ces théories peuvent être ou non identiques à celle que se donne le programme). Le but de ces approches est également de voir quelle théorie d'action est à l'œuvre en pratique. Une politique reposant sur une théorie d'incitation (que l'on pourrait exprimer simplement ainsi « en incitant économiquement le sujet X, il modifiera son comportement ») pourra par exemple, en pratique, relever exclusivement de théories d'apprentissage : ce ne serait pas la rémunération financière qui ferait changer le comportement du sujet X en question, mais le partage d'expériences avec d'autres parties prenantes. Les analyses de processus sont les plus à même de renseigner sur ce point, ayant pour but d'expliquer la mise en œuvre du programme, et notamment la production des effets, mais aussi d'établir « *un modèle complet des interdépendances causales* » (Vedung, 2005). Comme nous l'avons déjà mentionné, Young (1994) les envisage d'ailleurs comme des évaluations d'efficacité en tant que telles. Elles s'articulent aux approches visant à reconstituer les théories du programme en mettant davantage l'accent sur la création des effets, et notamment les processus de décision sous-jacent. Conduire une analyse stratégique du processus permet à ce propos d'étudier en détail les prises de position des acteurs dans les processus de décision, et de souligner les stratégies des acteurs qui placent la préoccupation environnementale au cœur de l'évaluation (Mermet, 2011).

Pour esquisser les théories de programme, trois types de démarches sont possibles :

1) l'approche *déductive* s'appuie sur des théories de programme développées dans la littérature académique,

2) l'approche *inductive* repose sur le travail de terrain (qui sert de base) et consiste à identifier les hypothèses et principes souvent implicites qui président en pratique à la conduite de la politique,

3) une dernière approche, qui s'appuie largement sur les perceptions des utilisateurs.

Selon la manière dont elle est conduite, l'évaluation d'un même programme peut mettre en évidence plusieurs théories. Là encore, les choix méthodologiques de l'évaluation doivent être réfléchis et explicités.

Dans le domaine de l'environnement, certains auteurs proposent une méthodologie fortement inspirée des approches « *fondées sur les théories* », qu'ils ont appelée le *Modus Narrandi* car il prône la narration de la mise en œuvre du programme pour mieux expliquer et attribuer les effets (Gysen *et al.*, 2006).

30 *"Programme theory, variously referred to as programme theory, programme logic […], theory-based evaluation or theory of change […], theory-driven evaluation […], theory-of-action […], intervention logic […], impact pathway analysis […], and programme theory-driven evaluation science […] refers to a variety of ways of developing a causal model linking programme inputs and activities to a chain of intended or observed outcomes, and then using this model to guide the evaluation".*

Les études de terrain possèdent pour leur part deux avantages qui peuvent s'avérer majeurs. D'une part, leur capacité à voir comment les chaînes de causalité qui mènent de la politique aux effets fonctionnent concrètement, et en particulier comment elles peuvent fonctionner différemment selon les contextes – autant d'informations cruciales pour améliorer la mise en œuvre de la politique étudiée. D'autre part, leur praticabilité : dans nombre de cas où il n'est pas possible d'avoir des mesures statistiquement fondées des effets d'une action publique, les méthodes reposant sur les études de terrain, si elles sont bien conduites, peuvent apporter des résultats à la fois robustes et très utiles. Il est également intéressant de noter que les approches fondées sur le couple : reconstitution de la théorie d'action / études de terrain de la mise en œuvre, se prêtent bien à une dimension participative de l'évaluation. Les parties prenantes ont en général une forte capacité à réfléchir sur les théories d'action alliée à une connaissance du terrain. Les évaluations directement liées aux contextes d'action sont enfin souvent plus faciles à retraduire en propositions d'action concrètes pour les acteurs eux-mêmes, bien que n'étant pas non plus, sans poser problèmes. En effet, elles n'apportent pas de preuves quantitatives à l'appui de leurs conclusions évaluatives, ce qui, dans certains contextes de discussion, peut limiter leur impact. Elles demandent aussi des moyens importants en termes de temps d'étude et de personnel. Leur qualité dépend de manière cruciale de compétences pointues en matière de méthodes de recherche qualitative, compétences à la fois longues et difficiles à acquérir, difficiles à partager de façon simple, et difficiles à évaluer. Enfin, les résultats des études de terrain qualitatives, y compris quand ils sont de très bonne qualité, comportent une part essentielle d'interprétation de situations complexes, et donc sont contingents au choix de l'équipe d'évaluateurs et à la conduite particulière de telle ou telle évaluation.

En guise de conclusion, nous souhaitons évoquer la réflexion menée par Berriet-Solliec *et al.* (2011) concernant *le niveau de preuve* permis par les méthodes quantitatives et qualitatives ainsi que *la validité empirique* de ces preuves.

Pour ces auteurs, la fiabilité des preuves est supérieure lorsqu'il s'agit de preuves de nature quantitative, par rapport à des preuves de nature qualitative. Ils classent les types de preuves selon une fiabilité croissante, comme suit :

- opinions d'autorités compétentes,
- preuves obtenues à partir de comparaison géographiques ou historiques,
- preuves obtenues avec des études de cohortes ou des études de cas contrôlées,
- preuves obtenues à partir de données issues de situations représentatives et après des tests statistiques de robustesse,
- preuves obtenues à partir d'expériences aléatoires contrôlées.

Si une telle conclusion est communément partagée, la réflexion qui en est à l'origine nous paraît d'un grand intérêt. En effet, elle invite l'évaluateur, qu'il opte pour une méthode qualitative ou quantitative, à s'interroger sur le degré de validité des données sur lesquelles il fonde son jugement de valeur. En outre, alors que dans le domaine de l'environnement les méthodes quantitatives sont loin d'être la panacée – les protocoles qu'elles nécessitent étant bien souvent empreints de biais, voire irréalisables – cette réflexion sur le niveau de preuve nous invite à veiller à la meilleure rigueur scientifique possible pour la mise en place de protocoles qualitatifs qui permettent d'attribuer et d'expliquer les effets du programme étudié.

2.5. Conclusion

La phase de cadrage des évaluations d'efficacité est une étape très importante de l'évaluation. Elle pose toutefois certaines difficultés, notamment dans le champ de l'environnement. Des politiques, dont les effets ultimes sur l'environnement ne sont pas toujours appréhensibles dans le pas de temps de l'évaluation, dont les portées spatiales interrogent quant au choix d'une échelle, à la fois réaliste et pertinente, pour l'évaluation, dont la complexité rend difficile la compréhension et la reconstitution des enchaînements causaux qui conduisent aux effets, ou encore dont le champ d'application fait l'objet de nombreuses controverses scientifiques… De telles politiques ne sont en effet pas aisées à évaluer !

Nous avons discuté des enjeux soulevés par les choix de cadrage sur trois aspects clé : le référentiel de l'évaluation, le périmètre de l'évaluation et l'explication / l'attribution des effets, tout en proposant certaines recommandations contextualisées :

- en ce qui concerne le référentiel de l'évaluation, nous avons ainsi vu que le choix à bon escient d'objectifs d'évaluation plus larges ou décalés par rapport à ceux de la politique à évaluer est tout à fait légitime. Nous avons recommandé d'adopter une démarche en termes de résolution de problème plutôt que d'amélioration. Nous avons vu aussi que la question du degré de participation et de délibération pour établir le référentiel était importante à prendre en considération.

- En ce qui concerne le périmètre de l'évaluation, nous avons souligné les difficultés soulevées lorsque les impacts ne sont pas encore appréciables, alors même que le référentiel est défini en termes d'impacts sur l'environnement. Tout en défendant néanmoins cette option, nous avons identifié des pistes pour surmonter ces difficultés, telles qu'analyser la manière dont les préoccupations du référentiel sont intégrées aux stades des réalisations et des résultats. Le caractère multi-scalaire des programmes à évaluer reste un enjeu pour l'évaluateur qui peut difficilement embrasser le programme dans son ensemble. Le recours à des protocoles d'évaluation originaux et aux études de cas sont des façons de répondre à ces contraintes d'échelle spatiale.

- Enfin, nous avons montré que les évaluations stratégiques de processus, assorties d'évaluations se penchant sur les théories du programme, représentent un moyen particulièrement pertinent pour contribuer à expliquer et à attribuer les effets observés.

Pour finir, nous proposons de confronter les réflexions théoriques et méthodologiques ci-dessus au projet d'évaluer de l'efficacité du FPCF.

3. Le Fonds de partenariat pour le carbone forestier (FPCF) : quel cadrage méthodologique proposer en vue d'évaluer son efficacité ?

Forts des analyses précédentes, nous nous intéressons ici à un programme environnemental international, le FPCF, que finance en partie l'AFD, visant à réduire la déforestation et la dégradation des forêts tropicales dans le cadre de la lutte contre le changement climatique. L'objet de cette troisième et dernière partie est de discuter des enjeux de cadrage qui se posent lorsqu'on entreprend d'en évaluer l'efficacité (objet du travail de doctorat dans lequel s'inscrit la présente publication), et non de présenter l'évaluation en tant que telle. Pour cela, nous présentons tout d'abord, de manière concise, ce programme et son contexte d'émergence. Puis, nous abordons successivement chacune des trois questions clés abordées en section 2 : quel référentiel adopter ? A quels effets s'intéresser ? Comment expliquer et attribuer ces effets ?

3.1. Qu'est ce que le FPCF ?

3.1.1. La question de la déforestation des forêts tropicales au cœur des négociations sur le régime Climat post-Kyoto

La question de la déforestation en zone intertropicale est aujourd'hui mise sur le devant de la scène internationale pour son rôle clé dans le changement climatique (*Association for Tropical Biology and Conservation et Society for Tropical Ecology*, 2009) : il est en effet estimé qu'elle représente entre 12 et 20 % des émissions mondiales de gaz à effet de serre (GES) (GIEC, 2007 ; Van der Werf *et al.*, 2009). Dans le même temps, les forêts sur pied contribuent activement à l'atténuation du changement climatique, grâce à leur capacité de stockage et de séquestration de carbone.

Encadré 7

Le protocole de Kyoto et les forêts tropicales

Dans le cadre du protocole de Kyoto, établi en 1997 par les parties à la Convention cadre des Nations unies sur le changement climatique (CCNUCC), à ce jour signé et ratifié par 191 pays[31], les pays développés ont pris des engagements pour diminuer leurs émissions de GES[32]. *Via* le Mécanisme de développement propre (MDP) instauré à cette occasion, ces pays ont la possibilité d'alléger leurs efforts en finançant des projets de réduction d'émissions dans les pays en voie de développement, en comptabilisant ces réductions d'émissions dans leur propre quota. Les projets concernés peuvent être de nature variée. Dans la pratique, ils concernent majoritairement l'amélioration de l'efficacité énergétique d'installations industrielles ou artisanales, mais un certain nombre d'entre eux sont des projets de boisements ou de reboisements en zone intertropicale. Néanmoins, pour de multiples raisons, tenant à la fois à des considérations méthodologiques, politique et d'intégrité environnementale, le secteur forestier tropical occupe en fait une très faible place dans le protocole de Kyoto. D'une part, ces projets de boisements et de reboisement ne sont pas éligibles sur le marché européen de quotas, et, d'autre part, les projets de réduction de la déforestation en zone intertropicale sont mondialement exclus du MDP (ONFI, 2011).

[31] Nombre de pays l'ayant signé et ratifié avant 2011.

[32] Le caractère individuellement et globalement contraignant de ces engagements serait à discuter en profondeur mais ce n'est pas l'objet du présent travail.

Ce n'est qu'en 2007 (cf. encadré 7), lors de la treizième Conférence des parties de la CCNUCC, qu'il fut décidé d'élaborer un mécanisme de Réduction des émissions liées à la déforestation et à la dégradation des forêts (REDD+) pour le régime climatique post-Kyoto (CCNUCC, 2007). Celui-ci devait voir le jour en 2012 mais, compte tenu du retard pris dans les négociations internationales, ce ne sera pas le cas avant 2020. Ce mécanisme incitatif REDD+, est censé rémunérer les pays forestiers tropicaux qui feront des efforts pour diminuer leur déforestation (cf. encadré 8).

Encadré 8

Une référence prise pour rémunérer les efforts de pays forestiers différente de l'option retenue dans le protocole de Kyoto. Une porte ouverte à tous les dérapages ?

Il est important de noter que la référence choisie pour apprécier ces efforts est bien différente de celle qui est actuellement en vigueur dans le protocole de Kyoto pour les pays développés ayant pris des engagements de réduction de GES. En effet, dans le cas de REDD+, il s'agira de mesurer l'amélioration de la situation par rapport à une situation contrefactuelle qui aurait eu lieu sans REDD+, un scénario au fil de l'eau (*Business As Usual* - BAU). L'idée est de récompenser financièrement les efforts faits par les pays. Dans le cas du protocole de Kyoto, la référence prise est la situation qui prévalait dans les pays développés en 1990. Il s'agit donc d'atteindre un objectif X, fixé : X % des émissions de GES de 1990. Cette différence est de taille. Elle illustre tout à fait l'alternative que nous avons détaillée dans la section 2.2, concernant le référentiel de l'évaluation. Cela augure des difficultés auxquelles vont être confrontés les pays REDD+ pour établir des scénarios de référence, plausibles et conservateurs (qui ne surestimeront pas la déforestation qui serait censée avoir lieu, dans le but d'être indemnisé sans faire d'efforts en réalité). En corollaire de ces difficultés d'ordre méthodologique, apparaît un espace de négociation politique et stratégique autour de l'établissement de ces scénarios. De telles considérations confortent notre préférence à choisir un référentiel en termes d' « état désiré de l'environnement », que nous allons justifier *infra*, plutôt qu'un référentiel dit BAU.

A Copenhague (décembre 2009), puis dans l'accord de Cancun de décembre 2010, le rôle crucial du REDD+ dans le régime climatique post-Kyoto a été rappelé (CCNUCC, 2010*a*, 2011). Le symbole « + » ajouté à l'acronyme REDD entérine désormais l'inclusion additionnelle de trois types d'activités favorisant des changements positifs du couvert forestier (Wertz-Kanounnikoff et Kongphan-Apirak, 2009) : la conservation, la gestion durable des forêts et la promotion des stocks de carbone existant. L'architecture du mécanisme, sa portée, son financement ainsi que certains aspects méthodologiques sont toujours en négociation.

3.1.2. Le FPCF, un des principaux programmes internationaux de préparation au REDD+

En parallèle de l'élaboration du mécanisme REDD+ dans les enceintes de négociation de la Convention Climat, de nombreuses activités de renforcement de capacités sont menées dans les pays forestiers tropicaux depuis 2007 (programmes bilatéraux et multilatéraux, initiatives d'Organisations non gouvernementales (ONG), financements privés, etc.). L'idée est que le mécanisme REDD+ qui rémunèrera les efforts de lutte contre la déforestation, ne pourra fonctionner que si les pays l'ont préparé en amont. La déforestation résultant d'activités mal réglementées ou illégales dans la plupart des pays, voire même, souvent, de politiques gouvernementales sectorielles dommageables pour la forêt (Pirard et Belna, 2012), un certain nombre de mesures et de réformes politiques et institutionnelles sont indispensables à mettre en œuvre avant de passer à la rémunération des réductions d'émissions à proprement parler. Par ailleurs, d'un point de vue technique, l'établissement des

scénarios de référence et des systèmes de suivi du couvert forestier (et des émissions associées) sont aussi promus dans le cadre de cette préparation à REDD+.

Le FPCF[33], lancé en décembre 2007, lors de la Conférence Climat qui s'est tenue à Bali, représente à ce propos le principal programme multilatéral mis en place dans la période de préparation (cf. encadré 9). Il est composé de deux fonds administrés par la Banque mondiale : un Fonds de préparation et un Fonds carbone.

Ses quatre objectifs inscrits sont :

« ***a)*** *assister les pays REDD+ éligibles à obtenir des réductions des émissions dues au déboisement et à la dégradation des forêts, en leur apportant une assistance financière et technique pour le renforcement de leurs capacités à tirer parti de futurs systèmes éventuels d'incitations positives à REDD+ ;*

b) *tester un système basé sur les résultats de réductions d'émissions obtenues grâce aux activités de REDD+, en vue d'un partage équitable et de la promotion d'incitations positives futures à large échelle pour REDD+ ;*

c) *tester dans le cadre de l'approche REDD+ les moyens de maintenir ou d'améliorer les moyens d'existence des communautés locales et de préserver la biodiversité ;*

d) *diffuser largement les connaissances acquises à travers la conception du FPCF, l'exécution des plans de préparation et les programmes de réduction d'émissions.* » (FPCF, 2008).

Encadré 9

Les initiatives de préparation à REDD+

La mobilisation de la communauté internationale sur la préparation à la REDD+ est importante. Outre le FPCF, d'autres initiatives jouent aussi un rôle important dans la préparation à la REDD+. Plusieurs pays investissent des montants substantiels, dans le cadre de leur coopération bilatérale ou de programmes multilatéraux, comme le FPCF (Bernard *et al.*, 2012). Parmi ceux-ci, la Norvège se distingue avec son Initiative internationale Climat – Forêt (ICFI)[34] par les montants dédiés (2,5 milliards d'euros sur 5 ans) et parce qu'elle correspond véritablement à des financements additionnels par rapport à l'aide au développement pratiquée par ailleurs. Le soutien est apporté aux initiatives multilatérales de la communauté internationale ainsi qu'à quelques Etats (les principaux étant le Brésil, le Guyana, l'Indonésie et le Mexique), pour des montants allant jusqu'à 1 milliard d'USD dans le cas du Fonds brésilien pour l'Amazonie, le *Fundo Amazonas*, et de l'Indonésie.

En dehors du FPCF, trois initiatives multilatérales jouent aussi un rôle important dans la préparation à REDD+ :

1) Le programme REDD+ des Nations unies[35] a été lancé en septembre 2008. Il est administré par trois agences des Nations unies : la FAO, le Programme des Nations Unies pour le développement (PNUD) et le Programme des Nations unies pour l'environnement (PNUE) et comprend deux types d'activités : *i)* un soutien direct à un certain nombre de pays récipiendaires pour l'élaboration et la mise en œuvre de leurs programmes nationaux REDD+ ; *ii)* des activités thématiques globales portant sur des aspects clés de REDD+, comme la mesure et le suivi (conduit par la FAO), les « bénéfices annexes » (mené par le PNUE), ou encore l'implication des peuples autochtones. En 2011, 14 pays[36] recevaient un soutien financier direct pour leurs

[33] http://www.forestcarbonpartnership.org/fcp/

[34] http://www.regjeringen.no/en/dep/md/Selected-topics/climate/the-government-of-norways-international-.html?id=548491

[35] http://www.un-redd.org/AboutUNREDDProgramme/tabid/583/Default.aspx

[36] La Bolivie, le Cambodge, la République démocratique du Congo, l'Equateur, les Iles Salomon, l'Indonésie, le Nigeria, le Panama, la Papouasie-Nouvelle-Guinée, le Paraguay, les Philippines, la Tanzanie, le Vietnam et la Zambie.

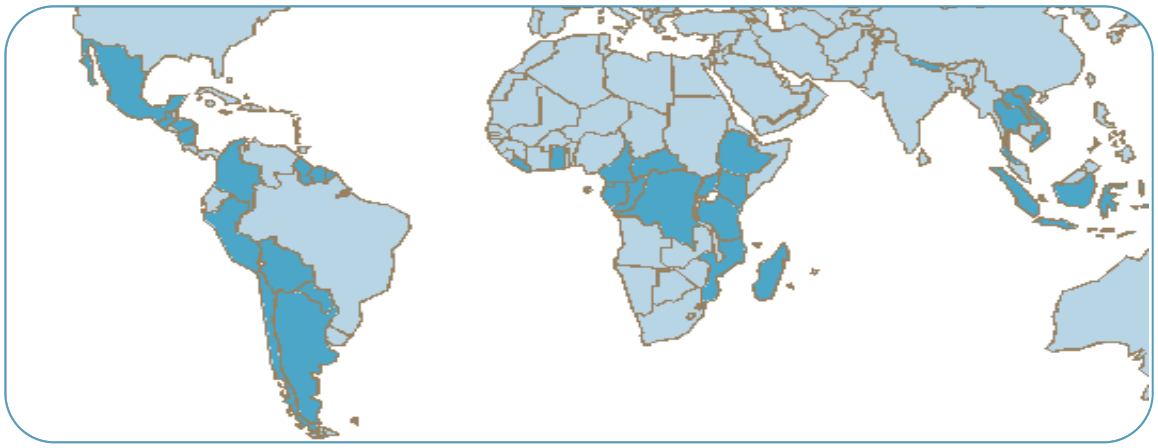
programmes nationaux. Vingt-deux autres, sont partenaires et peuvent donc potentiellement bénéficier du financement d'un aspect de leur préparation à REDD+. Le programme compte aujourd'hui cinq donneurs qui se sont engagés pour un montant total de subvention de l'ordre de 150 millions USD (chiffre de fin 2011) : la Norvège, qui a été l'unique contributeur jusqu'en juin 2009, est toujours le principal bailleur (123,2 millions USD).

2) Le Programme d'investissement pour les forêts[37] (PIF) fait partie des Fonds d'investissement pour le Climat, administrés par la Banque mondiale. Il a été décrété opérationnel le 1er juillet 2008 et vise notamment à financer la mise en œuvre des programmes nationaux REDD+. Il compte à ce jour huit pays pilotes : le Brésil (l'Etat de l'Acre), le Burkina Faso, la République démocratique du Congo, le Ghana, l'Indonésie, le Laos, le Mexique et le Pérou. Sept contributeurs, dont principalement la Norvège, les Etats-Unis et le Royaume-Uni, financent le PIF à hauteur de 599 millions USD[38] : 434 millions comme subvention et 165 millions comme prêts concessionnels.

3) Dans le cadre de sa cinquième période de financement (2010-2014), et plus spécifiquement de sa stratégie "*Sustainable Forest Management/REDD-plus strategy*" (FEM, 2011), le Fonds pour l'environnement mondial (FEM) a mis en place une enveloppe de 250 millions USD visant à favoriser une approche *paysage* pour les projets forestiers en finançant un tiers additionnel à tout projet FEM, qui associera au moins trois thématiques des Conventions de Rio : biodiversité, changement climatique et dégradation des terres.

Le Fonds de préparation du FPCF, qui fait plus spécifiquement l'objet de notre évaluation, a été déclaré opérationnel en juin 2008. Il appuie 37 pays[39] (cf. carte 1) afin que ces derniers élaborent et mettent en œuvre des programmes nationaux REDD+ (appelés *Readiness Preparation Proposals : R-PPs*). Ces programmes comportent non seulement des composantes politiques et institutionnelles (diagnostic de la déforestation, politiques et mesures envisagées, structures institutionnelles associées) mais aussi des éléments plus techniques comme les scénarios de référence et les systèmes de mesure et suivi. Ce Fonds de préparation est financé par 15 bailleurs, dont l'AFD, pour un montant total qui s'élevait au 1er octobre 2011 à 230,5 millions USD (engagés ou versés) (cf. tableau 4).

Carte 1. Pays récipiendaires du Fonds de préparation du FPCF



Source : http://www.forestcarbonpartnership.org/fcp/

[37] http://www.climatefundsupdate.org/listing/forest-investment-program

[38] Chiffres de novembre 2011.

[39] Argentine, Bolivie, Cambodge, Cameroun, Chili, Colombie, Costa Rica, République centrafricaine, République démocratique du Congo, République du Congo, Ethiopie, Gabon, Ghana, Guatemala, Guinée équatoriale, Guyana, Honduras, Indonésie, Kenya, Laos, Liberia, Madagascar, Mexique, Mozambique, Népal, Nicaragua, Ouganda, Panama, Papouasie-Nouvelle-Guinée, Paraguay, Pérou, Salvador, Suriname, Tanzanie, Thaïlande, Vanuatu, Vietnam.

Tableau 4. Contributions au Fonds de préparation du FPCF au 1er octobre 2011

Contributions au Fonds de préparation du FPCF au 1er octobre 2011 (USD m)					
	2009	2010	2011	2012-2015*	Totaux
AFD (France)	4,6	0,6		5,4	10,6
Australie	9,6		8,0		17,6
Canada			41,4		41,4
Danemark		5,8			5,8
Finlande	9,0			5,7	14,7
Allemagne			26,0		26,0
Italie			5,0		5,0
Japon	5,0	5,0			10,0
Pays-Bas	5,0			15,3	20,3
Norvège	5,0	16,4	8,8		30,2
Espagne	7,0				7,0
Suisse	8,2				8,2
Royaume-Uni			5,8		5,8
Etats-Unis	0,5	4,5			5,0
Fonds engagés	53,9	32,3	94,9	26,4	207,6
Commission européenne				5,4	5,4
Allemagne				13,5	13,5
Japon				4,0	4,0
Fonds engagés + fonds promis	53,9	32,3	94,9	49,3	230,5

Source: http://www.forestcarbonpartnership.org/fcp/node/324

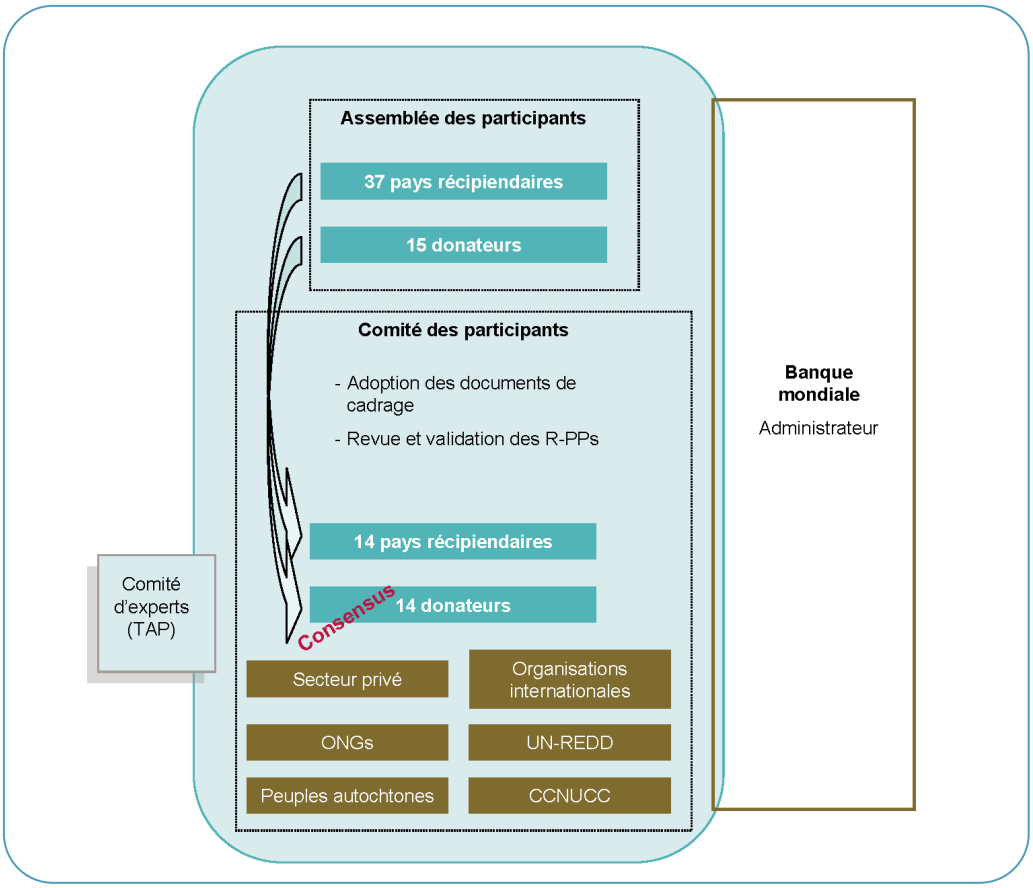
Le Fonds carbone du FPCF, lancé plus récemment en mai 2011, a pour objectif de rémunérer les réductions d'émissions qui seront obtenues par cinq pays pilotes. Il teste en somme l'incitation que devrait développer le mécanisme REDD+ à grande échelle dans le cadre du régime post-Kyoto. Aucun pays pilote n'a encore été choisi à ce jour, aucun pays participant au Fonds de préparation n'ayant fini de mettre en œuvre son programme national REDD+. Le Fonds carbone du FPCF n'est donc pas au cœur de notre évaluation.

Encadré 10

Gouvernance du FPCF

L'instance décisionnelle du FPCF est le Comité des participants (cf. schéma 8), qui se réunit en moyenne trois fois par an. Il compte 14 représentants des contributeurs (dont l'AFD), un nombre identique de représentants des pays récipiendaires et six observateurs, sans droit de vote, représentant la société civile, les peuples autochtones, les organisations internationales, le secteur privé, la CCNUCC, et le programme de préparation à REDD des Nations unies. Outre se prononcer sur les programmes nationaux REDD+, le Comité des participants valide des questions de politique générale comme le budget, le fonctionnement global du FPCF, les documents de cadrage élaborés par l'équipe de gestion de la Banque mondiale. Les décisions se prennent *a priori* par vote aux deux-tiers, mais le consensus est en pratique toujours recherché. L'Assemblée des participants, qui se tient annuellement, valide les décisions du Comité des participants. Deux panels d'expertise *ad hoc* ont été constitués, le deuxième, actuellement en exercice, réalise une revue des programmes nationaux REDD+ proposés par les pays récipiendaires, et fournit des avis scientifiques et techniques au Comité des participants.

Schéma 8. Gouvernance du Fonds de préparation du FPCF



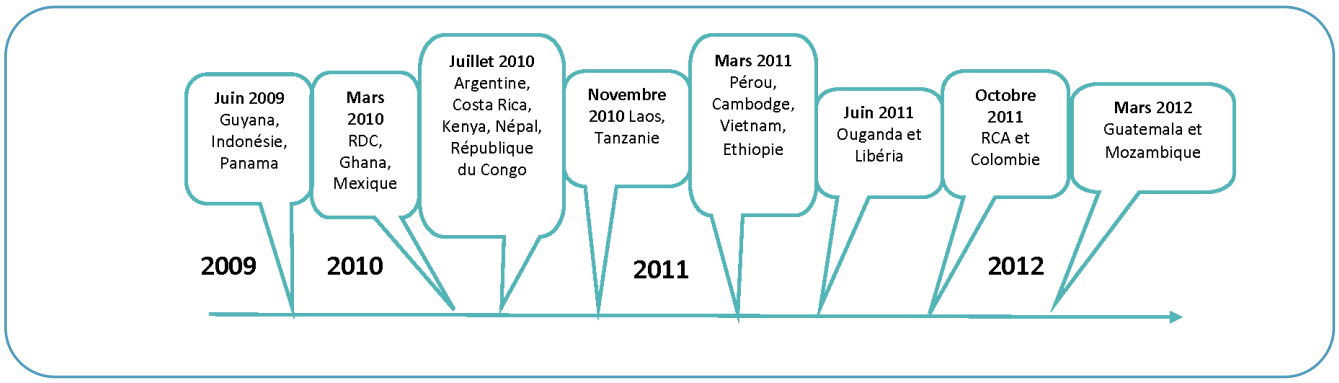
Source : les auteurs.

3.1.3. Une mission principale posant des questions majeures de politique générale : l'examen et le financement des programmes nationaux REDD+

L'activité principale du Fonds de préparation est l'appui à l'élaboration et à la mise en œuvre des programmes nationaux REDD+ (*R-PPs*). Un financement de 200 000 USD est attribué à chaque pays récipiendaire pour la conception de ce programme et une contribution de 3 600 000 USD lui est donnée pour sa mise en œuvre. Ce dernier financement ne couvre pas l'ensemble des dépenses envisagées dans les programmes nationaux REDD+. Les pays sont donc amenés à le compléter dans une large mesure par des financements additionnels, issus de programmes bilatéraux conclus avec des pays ou des institutions internationales, de contrats passés avec des fondations privées ou d'autres financements carbone (comme ceux octroyés par le PIF, autres Fonds Climat de la Banque mondiale ou bien par le Programme REDD des Nations unies).

A ce jour, 24 programmes nationaux REDD+ ont été officiellement évalués par le Comité des participants, à leur stade de plan de travail, et l'accord a été donné à ces pays d'entrer dans la phase de mise en œuvre de ces programmes (cf. schéma 9).

Schéma 9. Soumissions formelles des programmes nationaux REDD+ au Comité des participants du FPCF



Source : les auteurs.

Le processus d'examen de ces programmes par le FPCF se déroule en plusieurs étapes et fait l'objet de multiples procédures notamment, entre la Banque mondiale et les pays soumettant leurs programmes. De nombreuses revues ou évaluations de ces programmes sont réalisées par le panel d'experts du FPCF, par des pairs (les autres membres du FPCF), la Banque mondiale ou des ONG.

En corollaire de ce cœur d'activité du Fonds de préparation (l'appui aux programmes nationaux REDD+), des questions importantes de politique générale sont abordées. Certaines d'entre elles ont été largement débattues tandis que d'autres continuent à l'être :

- les révisions des documents de cadrage (modèles pour rédiger les programmes nationaux REDD+ et critères pour les évaluer),
- l'évaluation environnementale et sociale stratégique que doit obligatoirement mener tout pays récipiendaire du FPCF,
- l'ouverture du FPCF à d'autres partenaires de mise en œuvre (*Multiple Delivery Partners*) (cf. encadré 11),
- les modalités de rémunération des réductions d'émissions par le Fonds carbone du FPCF,
- l'évaluation à mi-parcours du FPCF.

Dispositif pilote, ayant pris comme référence le principe d'apprentissage par l'expérience (FPCF, 2008), le FPCF est dans une dynamique incessante de révision de ses documents de cadrage, en particulier les modèles pour élaborer les programmes nationaux REDD+ et les critères pour leur évaluation par les experts. Chaque nouvelle version fait l'objet d'une consultation avec le Comité des participants avant d'être adoptée par celui-ci. Les lignes directrices données et les exigences retenues pour l'évaluation influent grandement sur le contenu des programmes nationaux soumis. Ce sont des réalisations essentielles à prendre en compte dans l'évaluation du FPCF que nous nous proposons de mener.

L'EESS, est l'un des pré-requis du programme national REDD+, en lien avec les politiques de sauvegarde de la Banque mondiale. L'EESS a été, à l'origine, développée par la Banque mondiale pour des projets ou programmes sectoriels risquant d'impacter l'environnement. Or, le FPCF se démarque fortement de ce type de projets : il s'agit d'un programme environnemental, dont le champ d'application est de plus national. Il n'a pas été aisé pour la Banque mondiale de définir clairement en interne la portée et le niveau d'exigence d'une telle évaluation dans le cadre de la préparation à REDD+. L'enjeu était de ne pas trop contraindre les pays récipiendaires concernés, mais aussi de ne pas risquer la non-conformité avec ses propres politiques de sauvegarde. Ce sujet a occupé une place centrale dans les discussions du FPCF pendant près de deux ans, avant qu'une procédure validée par le Conseil d'administration de la Banque mondiale ne soit proposée et qu'une approche commune en matière de sauvegardes environnementales et sociales ne soit adoptée (FPCF, 2010). Ces discussions s'inscrivaient dans un processus plus global de révision de la politique de la Banque mondiale en matière de sauvegarde. L'évaluation environnementale stratégique est un élément clé de la pertinence des programmes nationaux qui seront élaborés par les pays membres du FPCF. Elle mérite en cela que notre évaluation environnementale du FPCF y attache la plus grande attention.

Les modalités de rémunération des réductions d'émissions par le Fonds carbone est une question actuellement très débattue au sein du Fonds de préparation. Quels seront les critères pour choisir les cinq programmes pilotes ? Quelle sera la garantie apportée pour l'obtention de bénéfices dits « additionnels » (au carbone), c'est-à-dire les bénéfices en matières environnementale (biodiversité et services écosystémiques, etc.), et sociale. Constituera-t-elle un critère discriminant lors de la sélection des programmes ? Ces bénéfices seront-ils rémunérés ? Les vues divergent fortement sur la question[40], qui devrait être tranchée l'été ou l'automne 2012. Il s'agit d'une question cruciale dans laquelle se jouent, *in fine,* la pertinence et l'efficacité environnementale de REDD+.

Un autre sujet qui a largement mobilisé le Comité des participants dans le courant des années 2010-2011 est la réalisation de l'évaluation indépendante à mi-parcours du FPCF, telle que prévue par la charte. Dans le cadre de cette évaluation, à dessein très interactive, le Comité des participants a notamment été sollicité pour réagir aux termes de référence proposés par la Banque mondiale, pour constituer un groupe de travail chargé d'accompagner le travail des deux bureaux d'étude contractés (Baastel et Nordeco) et enfin pour discuter les conclusions de l'évaluation et classer, en termes de priorité les recommandations formulées (Baastel et Nordeco, 2011).

[40] http://www.forestcarbonpartnership.org/fcp/node/486

Encadré 11

L'ouverture à d'autres partenaires de mise en œuvre (*Delivery Partners*) et l'élaboration d'une approche commune en matière de sauvegardes environnementales et sociales

Cette question est très sensible. Les deux principales raisons invoquées par la Banque mondiale pour motiver cet appel à d'autres partenaires pour la mise en œuvre des financements du FPCF dans certains pays, tenaient d'une part à l'absence d'avantage comparatif de la Banque mondiale dans ces pays (parce qu'elle n'y a aucune agence ou qu'elle est en mauvais termes avec le gouvernement, comme c'est le cas en Bolivie) et d'autre part au manque de ressources humaines disponibles. En effet, le FPCF rassemble aujourd'hui 37 pays, alors qu'il était initialement conçu pour n'appuyer que 20 pays. Les agences des Nations unies ainsi que les banques régionales de développement ont été proposées comme partenaires potentiels de mise en œuvre des financements. Le point qui a suscité le plus de questions, soulevé par les ONG et certains donneurs, était celui des clauses de sauvegardes. En effet, les clauses de sauvegardes de la Banque mondiale sont reconnues pour être les plus exigeantes (à tel point qu'elles sont parfois difficiles à mettre en œuvre), alors que celles des agences des Nations unies ou des banques régionales de développement, le sont moins. Une négociation ardue a eu lieu courant 2011 pour établir « une approche commune en matière de sauvegardes » (FPCF et *Readiness Fund*, 2011). A ce jour, le Programme des Nations unies, la Banque interaméricaine de développement (BID), et la FAO ont été adoptés comme partenaires de décaissement pour neuf pays pilotes.

3.1.4. L'importance accordée aux évaluations dans la mise en œuvre du FPCF

Le FPCF a fait l'objet d'un nombre important d'évaluations, à divers niveaux et selon diverses modalités. Nous avons déjà évoqué un certain nombre de ces évaluations. Avant de définir le cadre de notre propre évaluation du FPCF (section 3.2), et afin de mieux la positionner dans le paysage des évaluations existantes, nous recensons ici l'ensemble des évaluations concernant le FPCF (cf. tableau 5).

Ces évaluations portent sur :

- le processus FPCF dans son ensemble (en gras dans le tableau 5), telle l'évaluation à mi-parcours commanditée par le Comité des participants,
- les programmes nationaux REDD+ soumis au FPCF, spécifiquement. Il s'agit alors d'évaluations de ces programmes (par des experts, des pairs ou encore des ONG) (en italiques dans le tableau 5), ou d'évaluations conduites par les pays eux-mêmes, comme partie intégrante de leurs programmes nationaux REDD+. C'est le cas des évaluations environnementales et sociales stratégiques, des évaluations des politiques passées relatives aux forêts, ou encore du suivi de la mise en œuvre des programmes nationaux REDD+.

Le tableau 5 propose un panorama de ces évaluations, selon leur horizon temporel (Hilden, 2009) et leur méthodologie (Owen, 2007).

Tableau 5. Les évaluations conduites dans le cadre du FPCF

Forme d'évaluation / horizon de l'évaluation	Evaluation « proactive »	Evaluation « clarificative »	Evaluation « interactive »	Evaluation de type « suivi »	Evaluation d'impact
Passé					Evaluation de l'efficacité des politiques passées relatives aux forêts (11)
Présent		*Evaluation des programmes nationaux REDD+ > expertise (4), revue par les pairs (5)* *Revue des programmes nationaux REDD+ par les ONG (6)*	**Evaluation indépendante à mi-parcours du FPCF (1) > évaluation « réactive »** **Evaluation à mi-parcours de l'Initiative internationale Climat – Forêt de la Norvège incluant le FPCF (2) > évaluation « de plaidoyer »**	Suivi et reportage de la mise en œuvre du programme national REDD+ (12) : « systèmes de revue de la mise en œuvre » (Victor *et al.*, 1998) **Evaluation à mi-parcours du FPCF réalisée par le Groupe d'évaluation indépendant de la Banque mondiale (3)** *Revue à mi-parcours de la mise en œuvre des programmes nationaux REDD+ (7)*	*Evaluation de la mise en œuvre des programmes nationaux REDD+ > expertise (8), revue par les pairs (9)*
Futur		Evaluation environnementale et sociale stratégique (EESS) (10)			

Source : les auteurs.

Cette classification rend compte à la fois de la multiplicité des évaluations, mais aussi de leur diversité en termes d'horizon temporel et surtout de méthodologie[41]. Chacune des nombreuses évaluations répertoriées joue potentiellement un rôle important dans l'orientation du FPCF, par son contenu ou par l'utilisation qui en est faite. Nous les étudierons en détail dans notre analyse du FPCF, comme des vecteurs potentiels d'une meilleure prise en compte des enjeux environnementaux.

[41] On constate qu'aucune évaluation dite proactive n'est recensée (cf. tableau 1). Cela tient au calendrier de l'étude consacrée au FPCF, dans laquelle s'inscrit cette publication, étude qui a débuté en 2009 et n'a donc pu apprécier directement les évaluations éventuellement réalisées avant le lancement du FPCF (2006-2007). Elle n'a pas non plus trouvé de traces de tels travaux.

3.2. Discussion sur le cadre d'évaluation de l'efficacité du FPCF

Dans cette troisième partie, nous nous proposons de présenter et de justifier le cadre d'évaluation que nous avons choisi pour le FPCF, sans que l'objet soit de dérouler l'évaluation en elle-même.

Même si l'acronyme du FPCF invite à une évaluation d'un point de vue financier ou économique, nous faisons le choix, ici, de nous pencher sur l'efficacité d'un point de vue environnemental, du programme FPCF[42].

3.2.1. Quel est le but de notre évaluation ?

Compléter l'évaluation à mi-parcours commanditée par le FPCF

Comme le prévoyait la charte du FPCF, une évaluation à mi-parcours a été réalisée au cours de l'année 2011, après un peu plus de deux ans de mise en œuvre (Baastel et Nordeco, 2011). L'évaluation s'est appuyée sur trois des cinq principes recommandés par l'OCDE et la Banque mondiale : la pertinence, l'efficacité et l'efficience. Une approche « réactive » a été retenue, selon la pratique de la Banque mondiale (Owen, 2007) et les parties prenantes du FPCF ont été fortement sollicitées. Cette évaluation a donné lieu à des recommandations, élaborées et discutées par les participants, qui préconisent aujourd'hui des réorientations du FPCF (accroissement de la participation non seulement des peuples autochtones mais aussi de « la société civile » des pays REDD+, accélération des décaissements, etc.). Le propos ici n'est pas de faire une analyse exhaustive de cette importante évaluation, ni d'en commenter les retombées positives, mais de souligner l'une de ses limites à laquelle notre évaluation, plus ciblée, tentera de répondre.

L'évaluation à mi-parcours, tout comme l'évaluation globale commanditée par la Norvège dans le cadre de son Initiative internationale Climat-Forêt (CFI)[43] envisagent la *pertinence* du FPCF *dans le contexte des actions REDD+*. Même si ce point est essentiel, il ne nous apparaît pas suffisant en soi. En effet, REDD+ ayant émergé dans un système de gestion internationale des forêts (Smouts, 2008), comprenant déjà de nombreux engagements, de nombreux programmes, comme ceux ayant trait à la protection de la biodiversité forestière, il peut être intéressant et justifié de s'intéresser à une *pertinence* plus large, au regard de tous ces engagements, ce qu'on pourrait aussi appeler *cohérence externe* du programme FPCF.

Toutefois, en évaluant si le FPCF est « en bonne voie » d'atteindre les quatre objectifs qu'il s'est fixés, l'évaluation à mi-parcours touche aux problématiques liées à la biodiversité, puisque le troisième objectif vise à : « *tester dans le cadre de l'approche REDD+ les moyens […] de préserver la biodiversité* ». Cet objectif a été très peu pris en compte, le délai étant insuffisant pour se prononcer sur la question : « *il est clairement trop tôt pour parler de quelconque progrès en termes […] d'amélioration des revenus ruraux et de conservation de la biodiversité* »[44] (traduit de Baastel et Nordeco, 2011, p. 36). Ce troisième objectif a été succinctement abordé, pour conclure que « *depuis son lancement en 2008, le FPCF a fait des progrès notables vers l'atteinte du premier et du dernier objectif (créer les capacités dans les pays et diffuser les leçons apprises au cours de la phase de préparation à REDD+), par contre, en ce qui concerne les deux autres objectifs (piloter un système de paiements sur la base de la performance ; accroître les moyens de subsistance des populations rurales et conserver la biodiversité), des progrès en deçà de ce que l'on pouvait*

[42] A ce titre, il est intéressant de noter que le FPCF est la seule initiative de préparation dans le domaine de REDD+ à porter une dénomination fiduciaire : le programme de préparation à REDD+ des Nations unies s'appelle *programme ONU-REDD+* et le mécanisme financier complémentaire du FPCF au sein de la Banque mondiale se nomme *Programme d'investissement forestier*.

[43] http://www.regjeringen.no/en/dep/md/Selected-topics/climate/the-government-of-norways-international-.html?id=548491

[44] "*Clearly, it is too early to talk of progress in terms of piloting performance based systems and sustaining or enhancing livelihoods or conserving biodiversity*".

escompter à ce stade précoce, ont été réalisés »[45] (Baastel et Nordeco, 2011, p. 36).

Notre projet d'évaluation prétend, au contraire, qu'il est possible de donner d'ores et déjà des éléments d'analyse, notamment en termes de réalisations et de résultats, concernant la prise en compte de la biodiversité dans le processus FPCF, et sur les questions de la pertinence et de l'efficacité environnementale du FPCF, au sens large. Dans un souci de réorientation du programme, il est essentiel de porter un jugement précoce sur la mise en œuvre du programme, la pertinence et l'efficacité environnementale. Ainsi, nous proposons un cadrage d'évaluation qui accorde une place importante à la question des bénéfices environnementaux dits « annexes ».

3.2.2. Déterminer le référentiel de l'évaluation

Recadrer le problème : porter un regard critique sur le prisme « carbone » de REDD+ et du FPCF

Beaucoup de problématiques sont soulevées par la déforestation des forêts tropicales (problématiques écologiques, foncières, sociales, économiques, de droits humains, de genre, etc.) (CCNUCC, 2010*b*). REDD+ et les programmes de préparation comme le FPCF, sont confrontés à ces enjeux, à l'instar de toutes les politiques relatives aux forêts. Si la question de l'efficacité de ces politiques dans l'atténuation du changement climatique se pose, d'autres aspects de ces politiques méritent également d'être évalués. Nous concernant, c'est à une problématique environnementale, au sens large, non restreinte au service de régulation du climat, que s'attache notre évaluation, autour d'une question centrale. En dépit d'une focalisation commune sur le carbone, REDD+ et le FPCF permettent-ils de répondre de façon pertinente aux enjeux environnementaux de la déforestation des forêts tropicales ?

Au lancement de REDD+, d'aucuns pensaient que la biodiversité serait le « passager clandestin » des politiques REDD+. En d'autres termes, que préserver les forêts pour leur stockage ou leur séquestration de carbone serait forcément bénéfique pour la biodiversité. Pourtant, des sauvegardes en matières environnementales et sociales ont été annexées à l'accord de Cancun en 2010, montrant bien que certaines conditions sont nécessaires pour éviter que REDD+ ne cause de préjudices sociaux ou environnementaux (CCNUCC, 2011). Bon nombre de publications, notamment depuis l'année mondiale de la biodiversité en 2010, ont traité de la question des « sauvegardes » et des « bénéfices additionnels » de REDD+, deux facettes de la dimension socio-environnementale de REDD+ (CDB Secretariat, 2009*a*, 2009*b*, 2011 ; CDB Secretariat et GIZ, 2011 ; Karousakis, 2009 ; Pistorius *et al.*, 2010). Les sauvegardes environnementales sont les conditions nécessaires pour éviter toute incidence négative sur les enjeux environnementaux non-carbone. Les bénéfices environnementaux sont les retombées positives, non implicites, que peut avoir REDD+ sur des aspects autres que le carbone. Les premières peuvent-être résumées par « ne pas faire de mal », les seconds par « faire du bien ». Nous nous référons aux publications relatives aux sauvegardes et aux cobénéfices, les plus pertinentes à notre sens pour éclairer notre recadrage et alimenter le référentiel de notre évaluation.

- Pourquoi des « sauvegardes » environnementales ?

La portée de REDD+ s'est progressivement élargie. Alors que le mécanisme initial, discuté en 2005, concernait uniquement la réduction des émissions liées à la déforestation (RED) , il englobe aujourd'hui la réduction de la dégradation (REDD) ainsi que des activités de conservation, de gestion durable et d'accroissement des stocks de carbone existant, désignées par le « + » (CCNUCC, 2009*a*). L'ajout d'un second « + » à l'acronyme REDD+, représentant plus largement la prise en compte du carbone stocké grâce aux activités

[45] *"Overall, the evaluation found that since its inception in 2008, FCPF has made significant progress in meeting the first and last objectives (building in-country capacity and disseminating lessons learned in readiness), but less progress has been made on the two other objectives as would be expected at this early stage (piloting a performance-based system of payments; enhancing livelihoods and conserving biodiversity)".*

agricoles, est actuellement en discussion. Par ailleurs, la terre forestière est définie comme une terre « à vocation forestière », qu'elle soit boisée ou temporairement non boisée[46]. L'accroissement des stocks de carbone peut donc dans certains cas être synonyme de reforestation (dans d'autres cas, il correspond à un enrichissement de couvert forestier). Au final, si RED a pu être initialement synonyme de conservation des forêts tropicales, REDD+ ne l'est plus. La préoccupation centrale axée sur le stockage de carbone, peut laisser craindre que des activités de reforestation, permettant de séquestrer du carbone dans les arbres en croissance, soient privilégiées aux dépends d'efforts de réduction de la déforestation des forêts naturelles, plus difficiles. « Bien que la distinction entre les forêts et les plantations soit parfois floue, comme dans les cas où des espèces natives sont plantées, il est néanmoins nécessaire de différencier ces deux usages de la terre, sous peine que beaucoup des valeurs des forêts, la biodiversité en particulier, ne soient perdues sans qu'il n'y ait eu de changement en termes de couvert « forestier »[47](Putz et Redford, 2009).

Des fuites de la déforestation vers les forêts primaires ne sont pas non plus identifiables dans une comptabilité de ce type. « *En général, réduire la déforestation et la dégradation va impacter positivement la conservation de la biodiversité, mais cela sera réduit à néant si la déforestation et la dégradation sont déplacées d'une zone de faible valeur de conservation vers une autre zone de plus forte valeur de conservation ou vers d'autres écosystèmes natifs* »[48] (CDB Secretariat, 2009*b*, p. 12). Ainsi, la déforestation des forêts primaires pourrait se poursuivre alors même qu'un Etat serait récompensé pour une réduction de son taux net de déforestation.

Les sauvegardes de Cancun apportent un cadre réglementaire qui prend en compte ces risque et invite à mettre en place des « *mesures qui soient compatibles avec la préservation des forêts naturelles et de la diversité biologique […] [et] ne se prêtent pas à une conversion des forêts naturelles* » (CCNUCC, 2011). Mais la mise en œuvre de ces sauvegardes n'est pas obligatoire. De plus, elles ne s'interrogent pas sur la définition de la forêt utilisée ni ne se positionnent sur les modalités de prise en compte des plantations dans la comptabilité (CCNUCC, 2010*b*). Actuellement, ce sont les pays qui décident d'adopter des cadres plus ou moins ambitieux en matière de sauvegarde. Quelques rares pays – comme la Colombie (*Ministerio de Ambiente y Desarollo Sustenible*, 2011) – choisissent délibérément d'écarter les plantations de leur comptabilité REDD+. D'autres (Brésil, République démocratique du Congo, Equateur, Tanzanie, etc.) élaborent des sauvegardes nationales en s'appuyant sur les principes et critères développés par le programme REDD des Nations unies (Programme ONU-REDD, 2012) ou/et ceux du *Standard Social et Environnemental* (CCBA, 2010).

• Comment promouvoir les « cobénéfices » environnementaux ?

Comme le mentionne le rapport sur l'Economie des écosystèmes et de la biodiversité (TEEB, 2009) ainsi que le rapport de la Banque mondiale (2009), il existe des jeux de compensation entre services écosystémiques. En d'autres termes, maximiser un service, de production par exemple (comme la production de biomasse et donc la séquestration de carbone) peut se faire aux dépends d'un autre service. « *Pour que REDD contribue avec succès à combattre le changement climatique, il va falloir que les pays ciblent des*

[46] La définition de la FAO fait état de considérations liées à l'occupation des sols (une surface supérieure à 0,5 ha, avec une couverture arborée d'au moins 10 %, des tiges pérennes de plus de 5 m) et au mode d'utilisation des terres http://www.fao.org/DOCREP/005/Y2328F/y2328f03.htm. C'est sur ce dernier point que se joue le glissement de définition vers l'inclusion de terres nues à vocation forestière.

[47] *"Perhaps the most insidious and persistent threat to forest biodiversity results from confounding plantations with forests. Although the forest-plantation distinction is sometimes blurred, such as where native tree species are planted, the two land uses nevertheless need to be differentiated lest many of the values of forests, biodiversity in particular, be lost with no change in "forest" cover".*

[48] *"In general, reducing deforestation and degradation will positively impact biodiversity conservation, but this will be negated if deforestation and degradation is displaced from an area of lower conservation value to one of higher conservation value or to other native ecosystems".*

forêts en danger, avec de grands volumes de carbone dans la biomasse ou dans le sol. Mais il se pourrait que les sites prioritaires pour réduire la déforestation dans le but de réduire les émissions ne reflètent pas toujours d'autres valeurs de la forêt, comme la conservation de la biodiversité, des bénéfices en termes de moyens d'existence des communautés locales, ou encore la distribution d'eau »[49] (Banque mondiale, 2009).

La métrique carbone masque la complexité des enjeux écologiques et l'importance de la localisation spatiale. En effet, un des avantages de la « monnaie carbone », souvent mis en avant, est son équivalence en tout lieu. C'est à dire qu'une tonne de carbone non émise en un lieu donné équivaut à une tonne non émise en un autre lieu. Cependant, dès lors que l'on considère l'écosystème qui supporte cet attribut carbone, les limites de ce raisonnement apparaissent évidentes : deux forêts, aussi riches soient-elles en carbone, sont rarement équivalentes. Leur localisation et leurs propriétés fonctionnelles sont déterminantes. Cette équivalence spatiale des tonnes de carbone évitées masque l'hétérogénéité spatiale des enjeux écologiques.

Aussi, le choix des zones à préserver et de celles à reboiser est déterminant pour inciter, comme le mentionne l'accord de Cancun, « *à protéger et à conserver [les] forêts [naturelles] et les services rendus par leurs écosystèmes, ainsi qu'à renforcer d'autres avantages sociaux et environnementaux* »[50] et pour « *veiller à ce que les activités viennent en complément des objectifs des programmes forestiers nationaux et des conventions et accords internationaux pertinents ou soient compatibles avec ces objectifs* », notamment la Convention sur la diversité biologique (CCNUCC, 2011). Le Secrétariat de la CDB invite à réaliser des cartographies des écosystèmes et des planifications spatiales, pour considérer en priorité les forêts primaires, les forêts intactes, les forêts à haute valeur de conservation, les zones favorisant la connectivité (CDB Secrétariat, 2011).

[49] *"For REDD to make a successful contribution to combating climate change, countries would need to target threatened forests with a high volume of carbon in their biomass and soils. Priority sites for tackling deforestation to reduce emissions may not always reflect other forest values* e.g. *biodiversity conservation, livelihood benefits, or water delivery".*

Adopter une posture en termes de résolution de problème

Comme nous l'avons vu, section 2.2, on peut évaluer l'efficacité d'un programme en termes d'amélioration (absolue ou relative) ou en termes de résolution de problème. Nous avons mis en évidence les intérêts et les limites de chacun de ces partis pris, et avons plutôt recommandé de conduire l'évaluation selon la deuxième posture. Du fait de la complexité du phénomène de déforestation et de la politique FPCF, nous écartons effectivement l'option de définir une situation hypothétique, contrefactuelle, « sans FPCF », séduisante en termes conceptuels mais peu robuste en pratique (Underdal, 1992) (cf. encadré 7).

Nous posons donc la question en termes de résolution de problème : celui de la destruction des forêts naturelles tropicales. Nous identifions plusieurs enjeux environnementaux : perte de couverture forestière, destruction d'habitats et de biodiversité, perte de fonctions éco systémiques et destruction des lieux de vie des peuples autochtones (Guéneau, 2011 ; Leroy, 2006 ; ONU, 1992) que nous déclinerons en objectifs.

S'il peut sembler déplacé d'évaluer l'efficacité du FPCF au regard de cette large tâche, alors même que le problème de déforestation et d'érosion de la biodiversité persiste depuis plus de 30 ans et que le FPCF n'a pas la vocation ambitieuse de le résoudre, il nous semble alors important d'amener à considérer deux arguments :

1) le FPCF est inséré dans un régime forestier qui s'est fixé des objectifs environnementaux, notamment dans le cadre de la révision de la Stratégie internationale pour la biodiversité (CDB, 2010). Evaluer la cohérence externe du FPCF avec ces autres engagements est par conséquent légitime.

[50] *Compte tenu de la nécessité de moyens de subsistance durables pour les peuples autochtones et les communautés locales et de leurs liens d'interdépendance avec la forêt dans la plupart des pays, comme le reconnaît la Déclaration des Nations unies sur les droits des peuples autochtones, tout comme la Journée internationale de la Terre nourricière.*

2) Le cadre de mise en œuvre du FPCF (documents de cadrage notamment) est en reformulation permanente. Evaluer le FPCF au cours de sa mise en œuvre, selon un point de vue normatif décalé, est une façon d'apprécier dans quelle mesure le FPCF se réoriente vers ces préoccupations environnementales, annexes aux questions de carbone. Analyser « chemin faisant » les actions stratégiques et l'influence des acteurs qui partagent ces préoccupations environnementales est aussi un moyen de favoriser une meilleure prise en compte de ces questions et de favoriser cette dynamique.

A titre d'exemple, mentionnons que le troisième objectif du FPCF (traitant de la biodiversité) a été rajouté, à la suite notamment de critiques d'ONG concernant l'orientation trop « carbone » du FPCF (Greenpeace International, 2007). Le modèle de programme national REDD+ a été enrichi d'une composante (composante 4b) portant spécifiquement sur la mesure et le suivi de bénéfices autres que le carbone (comme la biodiversité) suite à une consultation du Comité des participants et aux suggestions de certaines ONG, apportées à cette occasion (Global Witness, 2010).

A partir de quelles sources établir le référentiel ?

Se pose maintenant la question des sources pour établir le référentiel de l'évaluation, et en corollaire celle du degré de délibération associé.

Alors que la divergence des positions dans le champ de l'environnement est particulièrement forte (Lafferty et Hovden, 2003 ; Mickwitz, 2003) et que des intérêts sectoriels (industriels, agricoles, etc.) s'opposent souvent à ceux visant la préservation de l'environnement, une approche stratégique, au moment de la définition des objectifs et du référentiel de l'évaluation, apparaît nécessaire (Mermet *et al.*, 2010). Dans le cas des politiques internationales visant la réduction de la déforestation, ce problème se pose de façon critique. Quel état de référence des forêts définir, sachant que chaque bassin forestier a ses spécificités écologiques (essences d'arbres, qualité de sol, de biodiversité, etc.), et sachant que les dynamiques des écosystèmes forestiers sont encore insuffisamment connues (notamment les processus de dégradation, les seuils de régénération, etc.) ?

Face à ces difficultés, il ressort que l'établissement du référentiel combine nécessairement une dimension stratégique (pour lui assurer une certaine robustesse politique dans le débat évaluatif) et une dimension technique (pour fournir un référentiel méthodologiquement exploitable dans la conduite de l'évaluation). Cette définition des standards devra articuler plusieurs échelles de précision, depuis des standards de principe assez généraux pour fonder l'évaluation jusqu'à des référentiels plus détaillés, par exemple pour conduire les différents types d'analyses (études de cas, statistiques, etc.) qui alimentent l'évaluation.

Nous nous référons, d'une part, à la littérature académique traitant des thèmes forêts tropicales, REDD+ et biodiversité, qui permet de fonder scientifiquement notre référentiel (CDB Secrétariat, 2009*a*, 2009*b*, 2011 ; Harvey *et al.*, 2010 ; Pistorius *et al.*, 2010*b* ; Sasaki et Putz, 2009), et, d'autre part, aux engagements internationaux en faveur de l'environnement, en particulier le plan stratégique pour 2020 de la Convention sur la diversité biologique (CDB, 2010), la CCNUCC (CCNUCC, 2009*b*)[51] et les objectifs du millénaire pour le développement, qui justifient les références quantitatives de notre référentiel et plaident pour une cohérence de l'action de la communauté internationale (Leroy, 2006 ; Underdal, 1992). Par ailleurs, nous nous appuyons sur le travail considérable réalisé par Guéneau (2011) pour établir un référentiel d'un état désirable des forêts tropicales et sur les standards environnementaux et sociaux REDD+ (CCBA, 2010 ; Programme ONU-REDD, 2012).

[51] Texte de négociation provisoire de la CCNUCC, même s'il n'a finalement pas été retenu dans l'accord de Cancun (qui a l'intérêt d'exprimer un objectif en termes absolus) : « Toutes les parties doivent chercher collectivement à au moins stopper la perte de couvert forestier dans les pays en voie de développement d'ici 2030 et à y réduire de moitié la déforestation brute, par rapport aux taux actuels, d'ici 2020 » (UNFCCC, 2010*c*) : "*[1. Objectives and Scope. 1. [All Parties should collectively aim at halting forest cover loss in developing countries by 2030 at the latest and reducing gross deforestation in developing countries by at least 50 per cent by 2020 compared to current levels.]*"

Encadré 12

Plan stratégique 2011-2020 et Objectifs d'Aichi relatifs à la diversité biologique (décision X/2) pertinents pour REDD+ (CDB, 2010)

Quatre objectifs d'Aichi, définis lors de la Conférence de Nagoya, en décembre 2010, sont pertinents pour REDD+ (CDB Secrétariat, 2011) :

« • *Objectif 5 : d'ici 2020, le rythme d'appauvrissement de tous les habitats naturels, y compris les forêts est réduit de moitié au moins et si possible ramené à près de zéro, et la dégradation et la fragmentation des habitats sont sensiblement réduites ».*

• *Objectif 7 : d'ici 2020, les zones consacrées à l'agriculture, l'aquaculture et la sylviculture sont gérées d'une manière durable, afin d'assurer la conservation de la diversité biologique ».*

• *Objectif 11 : d'ici 2020, au moins 17 % des zones terrestres et d'eaux intérieures et 10 % des zones marines et côtières, y compris les zones qui sont particulièrement importantes pour la diversité biologique et les services fournis par les écosystèmes, sont conservées au moyen de réseaux écologiquement représentatifs et bien reliés d'aires protégées gérées efficacement et équitablement et d'autres mesures de conservation effectives par zone, et intégrées dans l'ensemble du paysage terrestre et marin ».*

• *Objectif 15 : d'ici 2020, la résilience des écosystèmes et la contribution de la diversité biologique au stock de carbone sont améliorées, grâce aux mesures de conservation et restauration, y compris la restauration d'au moins 15 % des écosystèmes dégradés, contribuant ainsi à l'atténuation des changements climatiques et l'adaptation à ceux-ci, ainsi qu'à la lutte contre la désertification ».*

Si cette démarche n'est pas délibérative à ce stade, nous prenons le référentiel ainsi établi comme base pour initier l'évaluation, et cherchons à le confronter aux référentiels (préoccupations) des différentes parties prenantes au processus REDD+, tout au long de l'évaluation, afin d'en tester la portée et le degré d'adhésion, mais aussi d'y apporter des changements éventuels.

Les principes directeurs généraux qui guideront l'évaluation sont les suivants :

- Principe 1. La déforestation des forêts naturelles est réduite au moins de moitié à l'horizon 2020, par rapport à son niveau actuel.
- Principe 2. Le couvert forestier est restauré, reconstitué, et la connectivité entre écosystèmes est favorisée.
- Principe 3. Les habitats forestiers tropicaux essentiels au maintien de la diversité biologique sont conservés.
- Principe 4. En plus du service de régulation du climat, les services écosystémiques, de régulation, de soutien et les services écosystémiques culturels particulièrement importants pour les zones concernées sont maintenus.
- Principe 5. Les lieux et conditions de vie des populations faisant un usage traditionnel et durable de la forêt sont préservés.

Conclusion

Nous avons mis en évidence l'importance d'évaluer le FPCF dans une problématique environnementale plus large que celle strictement focalisée sur le carbone, prenant d'ailleurs au sérieux un objectif corollaire que le FPCF s'est lui-même

donné. En somme, notre posture revient à inclure dans l'évaluation de l'efficacité une évaluation de la pertinence et de la cohérence externe du FPCF, au regard d'enjeux écologiques dépassant le carbone et au regard de plusieurs engagements internationaux relatifs aux forêts (notamment la Convention sur la diversité biologique).

Nous avons vu qu'aborder l'évaluation en termes de résolution de problème permettait d'éviter les débats méthodologiques difficiles posés par l'approche par scénario de référence. Alors que c'est justement dans ces délicats méandres politico - méthodologiques que se lance la politique REDD+, notre positionnement « de côté » permet de clarifier les tenants et aboutissants de cette politique REDD+ et du FPCF, en fixant un objectif écologique explicite, comme pilier de l'évaluation. Néanmoins, notre posture n'est pas exempte de difficultés. Celles-ci sont liées à l'échelle spatiale de la politique étudiée et à l'état des connaissances en matière d'écosystèmes forestiers. La déclinaison des enjeux environnementaux en objectifs d'évaluation nous plonge dans les débats et les controverses en écologie, comme ceux de la définition des zones prioritaires pour la biodiversité (Guéneau, 2011). Nous avons donc essayé de prendre un référentiel robuste face à la controverse, qui consiste en des objectifs généraux que l'on déclinera plus précisément dans les pays choisis comme études de cas.

3.2.3. Le périmètre de l'évaluation

Analyser les résultats intermédiaires (réalisations et résultats) et rendre compte des impacts potentiels sur l'environnement

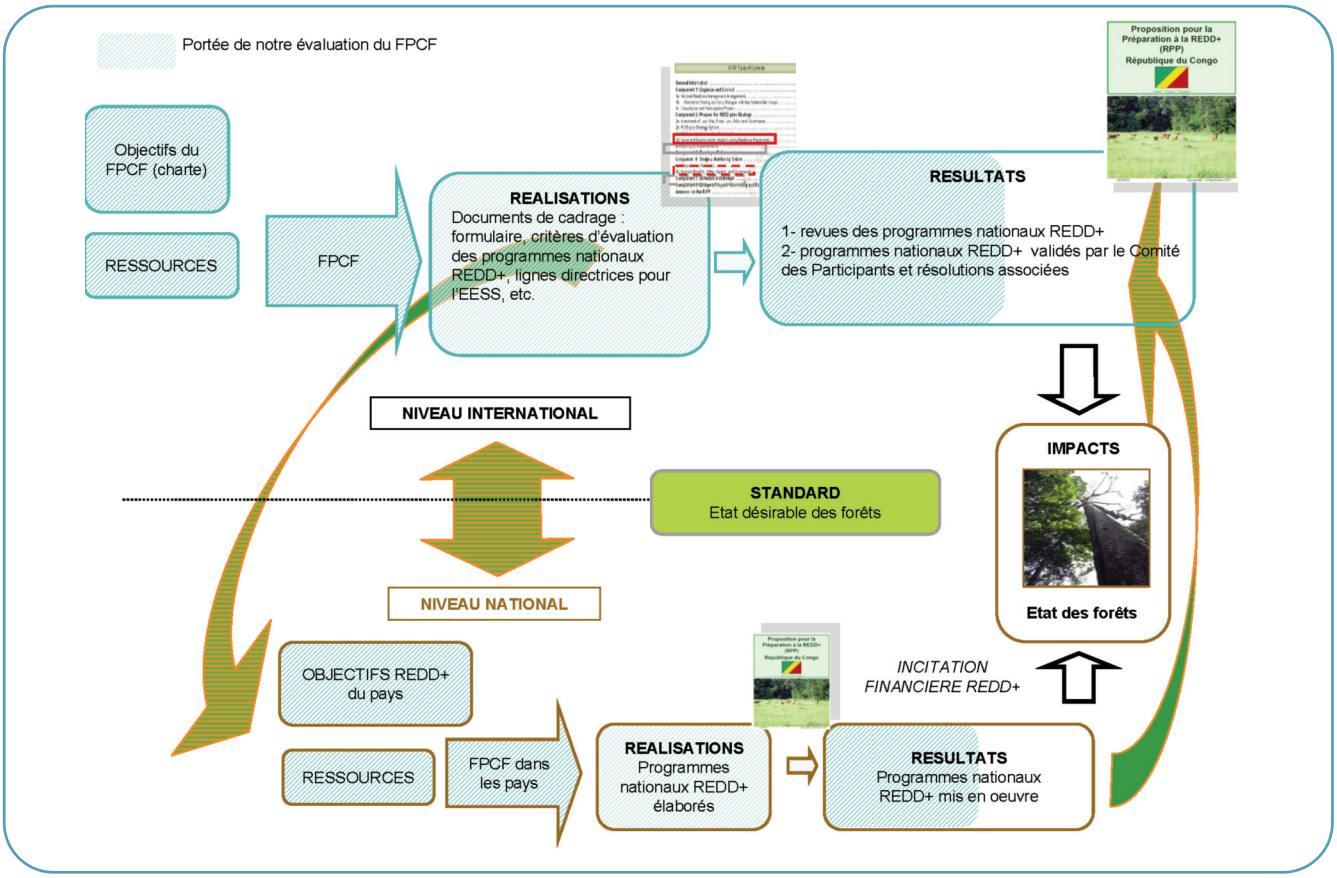
En suivant la théorie d'action de REDD+, nous pouvons dire que les résultats sur l'environnement ne seront effectifs et visibles qu'au moment où le mécanisme REDD+ incitera financièrement les acteurs à réduire concrètement la déforestation. Le FPCF se situe à un stade plus amont de « renforcement de capacité », de « préparation », qui s'illustre par des réformes et des mesures préalables, estimées comme essentielles.

Comme le représente le schéma 10, en suivant la mise en œuvre du FPCF, nous avons accès à deux niveaux :

- **niveau international** : aux objectifs exposés dans la charte, aux réalisations qui sont les documents de cadrage pour guider les pays dans l'élaboration de leurs programmes nationaux REDD+ et aux résultats qui correspondent aux revues des programmes nationaux REDD+ par les experts et par les pairs ainsi qu'aux programmes validés par le Comité décisionnel du FPCF (à leur stade de document programmatique et à leur stade final, après leur mise en œuvre) associés aux résolutions correspondantes qui ont été prises.

- **niveau national** : aux objectifs des pays qui sont présentés dans les documents de candidature au FPCF et éventuellement explicités dans la programme national, aux réalisations qui correspondent aux programmes nationaux REDD+ et aux résultats qui correspondent à la mise en œuvre de ces programmes et aux modifications concrètes des comportements des acteurs cibles concernés générés (administrations sectorielles, acteur local pratiquant la déforestation, etc.).

Schéma 10. Processus FPCF vu en termes de diagramme logique multiscalaire



Source : les auteurs.

Nous déroulerons l'évaluation tout au long de la mise en œuvre du FPCF, en suivant « à la trace » le processus (Leroy, 2006). Les réunions du Comité des participants, durant lesquelles se négocie le programme, constituent en quelques sortes les jalons de l'évaluation, même si les négociations décisives ont souvent lieu dans d'autres enceintes plus informelles (comme les groupes de travail) (Laurans et Dubien, 2000).

Concernant le contenu et les décisions stratégiques sous-jacentes, notre évaluation du FPCF nous amène donc à examiner les objectifs du FPCF, les programmes nationaux REDD+ soumis au Comité des participants, les résolutions afférentes adoptées et les différentes revues et évaluations conduites dans le cadre même du FPCF. A tous ces niveaux de la mise en œuvre, nous proposons alors de décrypter les intentions environnementales et les stratégies qui portent cette préoccupation.

Comment envisager l'évaluation multiscalaire ?

La difficulté qui se pose est la suivante : l'efficacité globale du programme FPCF est fonction non seulement du cadrage et des décisions prises au niveau international mais aussi et surtout de l'efficacité de chaque programme national REDD+ élaboré et mis en œuvre au niveau national. Or, il est impossible de considérer la mise en oeuvre du FPCF dans l'ensemble des 37 pays qu'il rassemble. Si l'on rejoint Pawson et Tilley (1997) dans leur *évaluation réaliste*, « *un mécanisme donné fonctionne dans un contexte donné pour conduire à un résultat donné* ». Rien ne peut donc être présagé sur l'efficacité du FPCF dans un pays dont le contexte n'est pas analysé précisément. Nous proposons par conséquent d'analyser la pertinence d'un échantillon représentatif des 37 programmes nationaux REDD+ (réalisations) soumis au FPCF, mais de restreindre l'analyse de la mise en œuvre à quelques pays traités sous forme d'études de cas. Dans ces pays, une analyse détaillée des contextes de la mise en œuvre, du processus national REDD+ et des interactions avec d'autres initiatives REDD+ sera conduite. Nous nous intéresserons ainsi aux systèmes de gestion puis à la place de la politique FPCF dans ces systèmes (Billé, 2007). Nous proposons de réaliser des contrepoints local – national en analysant en détail comment un enjeu de conservation local est intégré ou non au programme national REDD+.

Le dernier point que nous souhaitons aborder afin d'illustrer la réflexion menée, section 2, concerne l'attribution des effets et son explication.

3.2.4. Expliquer et attribuer les résultats : reconstituer les théories du programme FPCF et analyser le processus de mise en œuvre dans une perspective stratégique

Les méthodes qualitatives sont ici retenues : d'une part, la reconstitution des théories du programme (pour attribuer les effets, il est important de voir quelle est la logique causale assumée par le FPCF lui-même, et comment les parties prenantes au programme se représentent le FPCF), et d'autre part, l'analyse stratégique du processus.

Des réserves quant à la théorie économique incitative qui fonde REDD+

Une approche « fondée sur les théories » du programme cherche à comprendre les effets obtenus, en distinguant ce qui relèverait des hypothèses sous-jacentes au programme (des présupposés théoriques), de ce qui relèverait de la mise en œuvre même du programme (Chen, 1990).

La théorie d'action de REDD+, selon laquelle « *les pays vont réduire leur déforestation devant une incitation économique* », repose sur le présupposé que les agents de la déforestation agissent en acteurs rationnels dans leur réponse à l'incitation financière, et sur le fait que la compensation financière sera suffisamment attractive pour générer des changements de stratégies de développement économique et compenser les moindres revenus de secteurs particulièrement rémunérateurs (comme l'agro-industrie ou l'exploitation minière, etc.). Ces deux présupposés sont-ils néanmoins vérifiés ?

L'hypothèse de rationalité économique est bien souvent contestable, les choix étant éminemment politiques et répondant d'ailleurs souvent à des agendas privés (Karsenty et Fournier, 2008). Par ailleurs, compenser les revenus qui auraient été générés par l'exploitation des forêts nécessite, dans de nombreux cas, des financements très élevés, contrairement à ce qui a été suggéré aux débuts des négociations sur REDD+ (avec le rapport Stern). C'est le cas de la compensation des plantations de palmiers à huile par exemple (Sheil *et al.*, 2009). Or, ces financements ne sont pour l'instant ni disponibles ni même promis.

Les réserves que l'on peut émettre concernant la théorie qui fonde REDD+ sont donc fortes. Un véritable engagement politique environnemental apparaît nécessaire à la base. Toutefois, les financements de préparation à REDD+, abondants et dirigés vers des acteurs institutionnels porteurs de changement environnemental, peuvent atténuer certains rapports de force financiers en faveur de secteurs responsables de la déforestation, comme au Mexique (Bernard *et al.*, 2012).

Que nous dit la théorie d'action du FPCF sur ses potentiels effets ?

En ce qui concerne le FPCF qui s'inscrit dans le cadre fixé par REDD+, mais dans une phase amont de préparation, *i.e.* du renforcement de capacités, sa théorie d'action pourrait s'énoncer comme suit : « *les pays forestiers vont, d'une part, engager des politiques et des réformes leur permettant, dans le futur, de réduire leur déforestation face à une compensation financière et d'autre part, se doter de méthodes et de techniques leurs permettant de comptabiliser les efforts consentis en termes de réduction d'émissions* ». Cependant, de nombreuses zones d'ombre persistent au sujet du futur mécanisme REDD+, sur des aspects aussi centraux que les montants investis par les pays développés, l'architecture du mécanisme en tant que tel (marché ou fonds) ou encore les modes de comptabilité qui seront adoptés. Peut-on alors s'attendre à ce que les pays consentent à des dépenses de préparation (politiques, économiques, etc.) élevés, alors même qu'ils sont dans l'incertitude de la compensation future qu'ils recevront, en termes de montant et de modalités de rétribution (Karsenty, 2008) ? En outre, le FPCF ne finance les pays qu'à hauteur de 3,8 millions USD alors même que le coût moyen des stratégies REDD+ est de 14 millions USD (moyenne calculée pour les dix-neuf premiers pays ayant soumis leurs programmes nationaux REDD+ au FPCF) (Simula, 2010). Rares sont les pays qui ont trouvé d'autres bailleurs pour parvenir à financer l'ensemble de leurs coûts de préparation. Enfin, les réserves émises dans le paragraphe précédent, quant à une volonté politique suffisante pour porter les réformes supposées par REDD+, restent posées.

L'autre volet de l'approche « théorie de programme » consiste à s'intéresser à la mise en œuvre du programme et aux théories d'action sous-jacentes, que chaque acteur se représente ou que l'évaluateur reconstitue en analysant le processus. L'idée est que ce sont les parties prenantes qui font le programme ; la façon dont elles envisagent qu'il peut marcher ou pas, influence grandement la mise en œuvre du programme et ses effets. Il s'agit de mettre en perspective ce qui était planifié avec ce qui a effectivement lieu (Chen, 1990). Le pas de temps de l'évaluation « *chemin faisant* » est en cela un atout puisqu'il offre l'opportunité d'accéder à la « *boîte noire* » – *i.e.* aux mécanismes sous-jacents aux effets observés.

Dans le cas du FPCF, dont la mise en œuvre est fortement négociée (aux niveaux international et national), l'analyse stratégique du processus et plus particulièrement des processus de décision est essentielle pour comprendre et expliquer les effets générés par le programme. Chaque réalisation et chaque résultat est étudié dans son contenu et dans son processus de formulation à l'aulne du standard d'évaluation : comment intègre-t-il la préoccupation d'efficacité environnementale ? (Mickwitz et Kivimaa, 2007). Dans quelle mesure ? Qui intervient dans cette direction ?

3.3. Conclusion : comment les difficultés initiales de l'évaluation ont-elles été surmontées ? Quels enjeux demeurent ?

Au terme de cette troisième partie, le cadre d'évaluation que nous avons dessiné pour le FPCF nous permet de répondre aux difficultés méthodologiques associées à l'évaluation environnementale, présentées section 2 :

- *la complexité*. Nous avons traité cette question, en partie au moins, à travers :
  1) le choix d'un recadrage en termes de résolution de problème, et un référentiel basé sur la littérature académique en écologie, qui permet de prendre en considération les enjeux d'irréversibilité et de non-linéarité des phénomènes environnementaux et évite de s'engager dans l'établissement hautement spéculatif de scénarios contrefactuels pour l'évolution des processus de déforestation ;
  2) la conduite d'études *in situ*, au niveau national et sur des terrains locaux, pour analyser, dans leurs contextes socioécologiques spécifiques, les processus dommageables compliquées et complexes conduisant à la déforestation.

- *L'architecture institutionnelle*. A travers l'analyse stratégique du processus FPCF, et notamment des processus de décision, conduite à différents niveaux de gouvernance, notre évaluation aborde les enjeux liés à l'architecture institutionnelle et à l'intersectorialité.

- *La portée spatiale*. L'analyse multiscalaire proposée, l'étude documentaire d'un échantillon conséquent de programmes nationaux REDD+ ainsi que la conduite de deux études de cas, visent à prendre en compte l'étendue du problème de déforestation et du programme FPCF.

- *La question du temps*. Le Fonds de préparation du FPCF est un programme de renforcement de capacités, qui vise à changer les comportements, les structures, les réglementations etc. en amont de la rémunération financière des efforts de réduction d'émissions, rémunération supposée être suffisamment incitative pour conduire à des impacts positifs sur la préservation des forêts. Le Fonds de préparation intervient au stade des réalisations et des résultats, mais pas à celui des impacts. En outre, l'évaluation entreprise dans le cadre de l'étude à l'origine de cette publication, a débuté aux premiers stades de la mise en œuvre du FPCF, stade auquel seuls les réalisations et résultats précoces sont appréhensibles. En évaluant comment ces réalisations et résultats intègrent la préoccupation d'efficacité environnementale définie dans notre référentiel d'évaluation, *i.e.* dans quelle mesure elles se font les porte-paroles et relaient ces préoccupations environnementales, nous résolvons d'une certaine façon la difficulté posée par le moment de l'évaluation.

- *Le rôle de la science, les incertitudes*. En définissant un référentiel simple, en termes généraux ? qui sera décliné au niveau national, selon les contextes spécifiques et les études disponibles, et en nous appuyant sur des engagements internationaux qui le légitiment, tout en évitant l'écueil méthodologique des approches par scénarios de référence, nous visons à contourner au maximum les controverses scientifiques et les critiques d'ordre politique.

- *Le caractère conflictuel des problèmes d'environnement*. L'analyse stratégique permet de traiter le caractère conflictuel des dossiers environnementaux abordés. En outre, la référence aux engagements internationaux (et nationaux) pour l'établissement du référentiel d'évaluation lui donnent une certaine légitimité.

Pour terminer, soulignons deux enjeux auxquels se trouve confronté notre cadre d'évaluation. En premier lieu, l'isolement des effets propres du FPCF, dans les études de cas nationales.

Le FPCF étant seulement l'une des politiques REDD+, et, plus largement, l'une des nombreuses politiques ayant trait aux forêts tropicales, il interagit dans un même pays avec plusieurs politiques et programmes visant la conservation des forêts. Fournissant des financements limités aux pays membres, il est alors d'autant plus difficile de lui attribuer des effets finaux pouvant découler de financements complémentaires. On se demandera notamment si la meilleure prise en compte des aspects relatifs aux « co-bénéfices » environnementaux est le fait des sauvegardes ajoutées en annexe à l'accord de Cancun, si elle est due au travail en collaboration de la Convention sur la diversité biologique et de la Convention Climat, ou si elle est due, en amont, à l'intégration de ces aspects dans le FPCF, programme pilote. L'analyse du processus que nous proposons et la conduite de nombreux entretiens, associés à une observation des processus, est nous semble-t-il l'une des meilleures façons de répondre à cet enjeu.

En second lieu, il apparaît que notre évaluation s'oriente vers une évaluation de l'*intégration de la préoccupation environnementale aux stades amont* de la préparation à REDD+. En considérant qu'il s'agit d'une évaluation de l'efficacité, nous supposons implicitement que, plus une préoccupation environnementale est portée en amont, plus les impacts positifs, du point de vue de cette préoccupation, sont supposés être importants *in fine*. Bien que ce postulat semble aller de soi, gardons à l'esprit qu'il s'agit bien d'une hypothèse.

Conclusion générale

Alors que les enjeux en matière de conservation de la biodiversité restent immenses et qu'une nouvelle stratégie mondiale de la biodiversité a été adoptée lors de la conférence de Nagoya en 2010, assortie d'objectifs ambitieux relatifs notamment aux forêts tropicales, les mises en cohérence et en synergie des actions de la communauté internationale (Convention sur la biodiversité, Convention sur le changement climatique), apparaissent plus que nécessaires.

Tel est l'objet de l'analyse proposée ici, qui s'intéresse à la prise en charge des enjeux de conservation de la biodiversité par les politiques et programmes de préparation au mécanisme REDD+, et en particulier par le FPCF, programme multilatéral actuellement en cours de mise en œuvre. Est plus spécifiquement présentée une discussion approfondie du cadre d'évaluation qu'il apparaît pertinent de retenir pour évaluer l'efficacité environnementale de ce Fonds de préparation à REDD+, l'enjeu étant de pouvoir l'améliorer « chemin faisant ».

Organisée selon trois étapes, cette discussions nous a tout d'abord conduits à passer en revue des travaux majeurs menés dans le champ de l'évaluation, qui inventorient et classifient les nombreuses approches d'évaluation existantes. Ce premier temps de réflexion nous a permis de prendre conscience de la diversité des méthodes existantes, des questions qu'elles génèrent et de resituer plus précisément l'évaluation dite de « *l'efficacité* ». Nous nous sommes alors penchés sur le champ de l'environnement, domaine de l'action publique qui motive notre étude, et avons examiné les difficultés qui se posent en matière d'évaluation. Nous avons étudié les choix de cadrage et d'arbitrage que suppose toute évaluation de l'efficacité, en particulier dans le champ de l'environnement, puis nous avons examiné les options méthodologiques disponibles pour attribuer les effets observés, étape qui constitue le maillon clé de l'évaluation de l'efficacité. Enfin, troisième et dernière étape de cette discussion, nous avons illustré les réflexions menées par un exemple concret, en proposant un cadrage méthodologique qui permette de mener à bien l'évaluation de l'efficacité environnementale du FPCF.

Outre le fait que cette réflexion a permis de poser les bases de l'évaluation du FPCF, que nous réalisons en partenariat avec l'AFD, elle nous conduit à souligner les conclusions suivantes.

Il apparaît essentiel de connaître, de reconnaître et d'assumer la multiplicité des approches de l'évaluation et les différences profondes qui existent entre elles. Cette diversité constitue la réponse que le champ de recherche sur l'évaluation a apportée, au fil des décennies, à la très grande variété des contenus à évaluer et des situations où les évaluations ont vocation à être conduites, discutées et utilisées. Quel que soit l'enthousiasme avec lequel telle ou telle approche exprime sa conviction d'avoir trouvé la démarche imparable et généralisable attendue, il nous semble que toute approche doit être appréhendée au mieux comme une nouvelle ressource pertinente dans un répertoire riche et ouvert à des évolutions permanentes, pour s'adapter à des situations qui se renouvellent sans cesse. La connaissance de ce répertoire de ressources et la conscience de la spécificité de situations d'évaluation différentes, constituent, selon nous, la base fondamentale de toute compétence sur la problématique de l'évaluation – à des niveaux d'exigence différents, selon que l'on exerce comme participant à

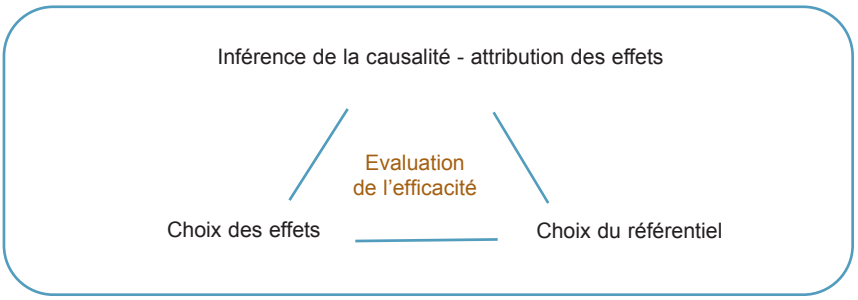
l'évaluation, commanditaire ou expert réalisant des évaluations.

Le champ de l'environnement présente quant à lui des caractéristiques marquantes liées à la nature des phénomènes en jeu et des contextes sociaux (complexité, larges échelles temporelles et spatiales, questions d'équité, etc.), et à la connaissance que nous en avons (expertise, incertitudes, conflictualité). Le développement de l'évaluation de l'efficacité des politiques environnementales gagnera donc à être alimenté et ancré dans la dynamique des recherches spécialisées sur l'environnement (*environmental studies* et *environmental science*), et à partager pleinement les enjeux, les débats, les ressources théoriques et méthodologiques du champ de l'évaluation de politiques en général.

Dans le domaine de l'environnement, la question de l'efficacité s'impose aujourd'hui face à une crise écologique croissante. Elle est pourtant loin d'être au centre de toutes les démarches d'évaluation : nombreuses sont les évaluations qui se concentrent, par exemple, sur la bonne mise en œuvre des moyens. Notre analyse de la littérature sur l'évaluation montre que celle-ci recèle des ressources importantes pour bien aborder le traitement des questions d'efficacité. Il ressort en particulier que toute évaluation de l'efficacité gagnera à reposer sur un traitement clair et cohérent de trois questions structurantes (cf. schéma 11) :

- le choix du référentiel qui fonde le jugement,
- la délimitation du périmètre des effets pris en compte,
- l'explication des effets observés et leur attribution à la politique étudiée.

Schéma 11. Trois questions organisationnelles à aborder pour l'évaluation de l'efficacité



Source : les auteurs.

Pour répondre à chacune de ces questions structurantes de l'évaluation de l'efficacité, l'évaluateur a le choix entre des options très différentes. Nous avons pu établir des recommandations quant à chacune de ces questions sachant que les choix à effectuer doivent tenir compte non seulement de préférences *a priori* de participants à l'évaluation, mais également des spécificités de l'action à évaluer et du contexte de l'évaluation. De plus, soulignons que l'option choisie pour chacune des questions structurantes influe sur les autres et met en jeu un arbitrage entre des approches alternatives. Il est alors important de réaliser que cet arbitrage se joue entre des approches qui pourraient être soit optimales d'un point de vue conceptuel, soit idéales d'un point de vue pratique, soit meilleures du point de vue de leur résistance à la critique politique, qualité non négligeable étant donné les fortes pressions dont les évaluations font l'objet (Mermet, 1996 ; Stufflebeam, 2001 ; Owen, 2007). L'optimisation simultanée de ces trois qualités sera rarement possible dans le champ de l'environnement. Il apparaît ainsi d'autant plus essentiel, avant toute évaluation de l'efficacité, de bien poser les enjeux du

cadrage, de considérer les options possibles et de justifier la posture normative et méthodologique finalement adoptée. Pour aider en cela le lecteur, nous nous sommes efforcés, à chaque étape de ce document, de mettre à sa disposition les ressources les plus pertinentes trouvées dans la littérature portant sur l'évaluation des politiques et programmes.

Fort de cet outillage méthodologique, nous espérons et nous appelons à de futurs exercices menés en ce sens sur le FPCF ou tout autre programme environnemental, afin de valider, compléter et enrichir cette démarche, de manière, *in fine*, à faire progresser nos modes d'actions sur l'environnement.

Annexe. Petit glossaire des approches pour l'évaluation de politiques et programmes

Tableau récapitulatif des approches en évaluation mentionnées par quelques auteurs d'ouvrage de synthèse de référence : Alkin, Crabbé et Leroy, Owen, Rossi, Stufflebeam et Vedung

N° de référencement dans l'annexe	Nom de l'approche	Classe dans laquelle l'approche est recensée	Auteur de l'ouvrage de synthèse mentionnant l'approche
1	Accountability (payment by results)	Questions and/or methods oriented approaches	Stufflebeam
2	Accreditation/certification approaches	Improvement/accountability oriented approaches	Stufflebeam
3	Action research	Interactive	Owen
4	Adversary evaluation	Valuing (Thomas Owens & Robert Wolf)	Alkin
	Advocate-adversary evaluation	Approach for designing evaluation research	Crabbé et Leroy
5	Case study evaluation	Approach for evaluation	Crabbé et Leroy
		Questions and/or methods oriented approaches	Stufflebeam
	Clarification hearing	Questions and/or methods oriented approaches	Stufflebeam
6	Client-oriented model	Effectiveness. Client concerns.	Vedung
	Consumer-oriented studies	Improvement/accountability oriented approaches	Stufflebeam
7	Component analysis	Monitoring	Owen
8	Comprehensive evaluation	Effectiveness. System components	Vedung
9	Concern-focused evaluation	Effectiveness	
10	Connoisseurship and criticism	Questions and/or methods oriented approaches	Stufflebeam
		Valuing (Elliot W. Eisner)	Alkin
	Pear review	Professional	Vedung
11	Constructivist evaluation	Social agenda/advocacy approaches	Stufflebeam
		Approach for designing evaluation research	Crabbé et Leroy
		Valuing (Egon G. Guba et Yvonna S. Lincoln)	Alkin
12	Context inputs process & product	Approach for designing evaluation research	Crabbé et Leroy
		Use (Daniel L. Stufflebeam)	Alkin

N° de référencement dans l'annexe	Nom de l'approche	Classe dans laquelle l'approche est recensée	Auteur de l'ouvrage de synthèse mentionnant l'approche
13	Context-adapted utilization	Use (Marvin C. Alkin)	Alkin
14	Cost effectiveness	Economic. Productivity	Vedung
	Cost-benefit analysis	Approach for evaluation	Crabbé et Leroy
	Benefits-costs analysis	Questions and/or methods oriented approaches	Stufflebeam
15	Decision/accountability-oriented studies	Improvement/accountability oriented approaches	Stufflebeam
16	Deliberative democratic evaluation	Social agenda/advocacy approaches	Stufflebeam
	Deliberative democratic evaluation	Approach for designing evaluation research	Crabbé et Leroy
17	Developmental	Interactive	Owen
	Developmental/formative evaluation	Approach for evaluation	Crabbé et Leroy
18	Developped performance assessment	Monitoring	Owen
19	Empowerment evaluation	Interactive	Owen
		Approach for designing evaluation research	Crabbé et Leroy
		Use (David M. Fetterman)	Alkin
20	Evaluability assessment	Clarificative	Owen
		Approach for designing evaluation research	Crabbé et Leroy
		Use (Joseph S. Wholey)	Alkin
21	Evaluation forms	Use (John M. Owen)	Alkin
22	*Ex ante*	Clarificative	Owen
23	Experiment and quasi experiment	Approach for evaluation	Crabbé et Leroy
	Experimental and quasi-experimental designs and internal validity	Method (Thomas D. Cook)	Alkin
	Experimental studies	Questions and/or methods oriented approaches	Stufflebeam
	Generalizability theory and construct validity	Method (Lee J. Cronbach)	Alkin
	Randomized field experiment	Method (Robert F. Boruch)	Alkin
	Randomized field experiment and quasi experimental design	Method (Donald Campbell)	Alkin
24	Goal attainment	Effectiveness. Objectives	Vedung
25	Goal free evaluation	Effectiveness. Results	Vedung
		Impact	Owen
		Approach for evaluation	Crabbé et Leroy
		Valuing (Michael Scriven)	Alkin

N° de référencement dans l'annexe	Nom de l'approche	Classe dans laquelle l'approche est recensée	Auteur de l'ouvrage de synthèse mentionnant l'approche
26	Logframe method	Approach for evaluation	Crabbé et Leroy
27	Management information systems	Questions and/or methods oriented approaches	Stufflebeam
28	Meta-evaluation	Approach for designing evaluation research	Crabbé et Leroy
29	Mixed-method evaluation	Approach for designing evaluation research	Crabbé et Leroy
		Questions and/or methods oriented approaches	Stufflebeam
30	Multicriteria analysis	Approach for evaluation	Crabbé et Leroy
31	Needs analysis	Approach for evaluation	Crabbé et Leroy
32	Needs assessment	Proactive	Owen
	Needs-based	Impact	Owen
33	Impact assessment	Approach for evaluation	Crabbé et Leroy
	Objective testing programs	Questions and/or methods oriented approaches	Stufflebeam
	Objective-based studies (side effects, process, goals…)	Questions and/or methods oriented approaches	Stufflebeam
	Objectives-based	Impact	Owen
	Objectives-oriented evaluation	Method (Ralph W. Tyler)	Alkin
34	Outcome evaluation / value-added assessment	Questions and/or methods oriented approaches	Stufflebeam
35	Performance audit	Impact	Owen
36	Performance testing	Questions and/or methods oriented approaches	Stufflebeam
37	Politically controlled studies	Pseudo-evaluation	Stufflebeam
38	Process-outcome	Impact	Owen
39	Program logic evaluation	Clarificative	Owen
	Program theory-based evaluation	Questions and/or methods oriented approaches	Stufflebeam
	Programme theory evaluation	Approach for evaluation	Crabbé et Leroy
	Theory - driven evaluation	Method (Huey-Tsy Chen)	Alkin
		Method (Peter H. Rossi)	Alkin
	Theory - driven evaluation and political context and theory of change	Method (Carol Hirschon Weiss)	Alkin
40	Public relations-inspired studies	Pseudo-evaluation	Stufflebeam
41	Quality review	Interactive	Owen

N° de référencement dans l'annexe	Nom de l'approche	Classe dans laquelle l'approche est recensée	Auteur de l'ouvrage de synthèse mentionnant l'approche
42	Realistic evaluation	Impact	Owen
		Approach for evaluation	Crabbé et Leroy
43	Research synthesis (evidence-based practice)	Proactive	Owen
44	Participatory evaluation	Use (Jean A. King)	Alkin
	Practical participatory evaluation	Use (J. Bradley Cousins)	Alkin
	Responsive evaluation	Interactive	Owen
		Approach for designing evaluation research	Crabbé et Leroy
		Valuing (Robert Stake)	Alkin
	Responsive evaluation (client-centered studies)	Social agenda/advocacy approaches	Stufflebeam
45	Review of best practice	Proactive	Owen
46	Side effects	Effectiveness. Objectives	Vedung
47	Stakeholder model	Effectiveness. STK concerns	Vedung
48	System analysis	Monitoring	Owen
49	Utilization-focused evaluation	Social agenda/advocacy approaches	Stufflebeam
		Approach for designing evaluation research	Crabbé et Leroy
		Use (Michael Q. Patton)	Alkin

1. Accountability Evaluation [Stufflebeam]

[Questions and/or Methods Oriented Approaches]

- Lessinger, L. M. (1970), *Every Kid a Winner: Accountability in Education*, Simon and Schuster, New York.

Cette approche s'intéresse aux résultats du programme, mais à la différence des évaluations de type « *objective-based* », souvent internes à l'organisation qui met en œuvre le programme, elle prône une évaluation de type externe, avec un jugement impartial. Ce type d'évaluations a émergé dans les années 1970 aux Etats-Unis, en lien avec les faibles résultats rapportés dans le domaine des grands programmes d'investissement public (la santé et l'éducation). Lessinger est souvent reconnu comme le pionnier dans cette démarche de « responsabilité ».

2. Accreditation / Certification approaches [Stufflebeam]

[Improvement/Accountability Oriented Approaches]

Stufflebeam désigne sous ce nom les évaluations d'institutions ou d'organisations conduites dans le but de s'assurer qu'elles remplissent bien les exigences de la profession ou du domaine de service qui leur sont rattachés et qu'elles sont bien en mesure d'assurer leurs fonctions spécifiques. Une méthode classiquement utilisée dans ce type d'approche, est le rapportage mené par les organisations et institutions concernées, elles-mêmes.

3. Action Research [Owen]

[Interactive]

Non restreinte au champ de l'évaluation, il s'agit d'une méthode de recherche très pratiquée en sciences sociales. Elle consiste en une collaboration et une interaction forte entre l'évaluateur et un groupe de parties prenantes impliquées dans l'évaluation. Elle vise à susciter un apprentissage et à améliorer les pratiques. La perception du programme par les différents membres du groupe joue un rôle essentiel.

4. Adversary Evaluation [Stufflebeam, Alkin, Crabbé et Leroy], voir : advocate – adversary evaluation, clarification hearing

[Questions and/or Methods Oriented Approaches, T. Owens et R. Wolf. Valuing, Approach for Designing Evaluation Research]

- Wolf, R. L. (1975), "Trial by Jury: A New Evaluation Method", *Phi Delta Kappan*, 57(3), pp. 185-187.

Cette approche s'appuie sur une métaphore juridique. Elle repose sur l'hypothèse que tout évaluateur induit un biais dans la conduite de l'évaluation. Elle propose donc de créer un débat contradictoire entre plusieurs évaluateurs, chacun présentant les arguments supportant son point de vue.

5. Case-Study Evaluation [Stufflebeam, Crabbé et Leroy]

[Questions and/or Methods Oriented Approaches, Approach for Evaluation]

- Yin, R. (2003), *Case Study Research: Design and Methods,* Thousand Oaks, Sage.

L'étude de cas est avant tout une méthode générale de recherche en sciences sociales. Déclinée au champ de l'évaluation, elle permet d'évaluer des politiques et programmes dans leur cadre spécifique, d'apprécier comment ces politiques fonctionnent et pourquoi, dans ce contexte précis. Elle s'intéresse aux contextes géographique, culturel, organisationnel et historique des politiques et programmes dans la perspective de mieux en comprendre les résultats. Les pionniers à avoir utilisé cette méthode en évaluation de programmes sont : Campbell, 1975 ; Stake *et al*., 1978 ; Lincoln et Guba, 1985.

6. *Client-Oriented Model* [Stufflebeam, Vedung], voir : Consumer-Oriented Studies

[Improvement/Accountability Oriented Approaches, Effectiveness. Client concerns]

- Scriven, M. (1967), "The Methodology of Evaluation", *in* Stake, R.E., *Curriculum Evaluation. AERA Monograph Series on Curriculum Evaluation*, Rand Mc Nally, Vol. 1, Chicago.

Dans cette approche visant l'évaluation de l'efficacité d'un programme, ce sont les buts, les attentes, les préoccupations, voire même les besoins du groupe cible qui sont pris comme référence, et non ceux des développeurs du programme. Scriven fut pionnier dans l'application de cette approche de l'évaluation.

7. Component Analysis [Owen]

[Monitoring]

Owen désigne sous ce nom l'évaluation d'une composante sélectionnée d'un programme, évaluée d'une part en tant que telle (au regard de ses propres objectifs) et d'autre part, du point de vue de sa contribution à la mission générale et aux objectifs globaux du programme.

8. Comprehensive Evaluation [Vedung]

[Effectiveness System Components]

- Kaufman, R.A. et S. Thomas (1980), *Evaluation without Fear,* New Viewpoints, New York.

Vedung désigne sous ce nom l'évaluation de l'efficacité, qui ne se limite pas à l'analyse de l'atteinte des objectifs et des résultats, mais porte également sur la mise en œuvre du programme voire même sur la phase de conception. Aux différentes phases de la conception et de la mise en œuvre du programme, elle consiste d'une part à comparer les observations aux attendus, et d'autre part à porter un jugement à partir de critères de valeur explicités.

9. *Concern-Focused Evaluation*

- Mermet, L., R. Billé et M. Leroy (2010), *"Concern-Focused Evaluation for Ambiguous and Conflicting Policies: An Approach from the Environmental Field*", *American Journal of Evaluation*, 31(2), pp. 180-198.

Cette approche, développée par un groupe de chercheurs français en sciences de gestion, est ciblée sur l'évaluation de l'efficacité des politiques et programmes dans le domaine de l'environnement. Elle replace les objectifs de l'action environnementale au cœur du système de gestion plus large dans lequel la politique à évaluer s'insère. Elle fonde l'évaluation sur la mise en regard de la performance globale de ce système de gestion de fait, et de la contribution propre de la politique aux résultats environnementaux de ce système de gestion.

10. Connoisseurship and Criticism [Stufflebeam, Alkin, Vedung, Crabbé et Leroy], voir : *Pear Review*

[Questions and/or Methods Oriented Approaches, *E.W. Eisner, Valuing*, Professional Evaluation, Approach for Designing Evaluation Research]

- Eisner, E.W. (1986), "Educational Connoisseurship and Criticism: their Form and Functions in Educational Evaluation", *in* Madaus, G.F., M. Scriven et D. Stufflebeam, *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, Kluwer-Nijhoff, Boston, MA.

Cette approche, largement reconnue en évaluation, est inspirée des méthodes de critiques d'art et de littérature. Elle fait porter l'évaluation par certains experts du domaine d'action publique considéré, dont on considère qu'ils sont capables de faire des analyses de fond et des évaluations. Dans cette perspective, l'expertise sur le domaine d'action est plus décisive que les cadres de méthodes qui guident l'évaluation.

11. Constructivist Evaluation [Stufflebeam, Alkin, Crabbé et Leroy], voir : Naturalistic Enquiry

[Social agenda/advocacy approaches, *E.G. Guba & Y. S. Lincoln. Valuing*, Approach for Designing Evaluation Research]

- Guba, E.G. et Y.S. Lincoln (1989), *Fourth Generation Evaluation*, Sage, Newbury Park, CA.

Cette approche, largement reconnue en évaluation, a été développée par Guba et Lincoln dans les années 1980. Reposant sur le paradigme constructiviste, elle s'inscrit en réaction aux approches rationnelles positivistes de l'évaluation qui souhaitent éliminer les « biais » liés à la pratique humaine de l'évaluation. L'approche constructiviste valorise au contraire les différences de perception. Elle soutient qu'il n'existe pas de réponse absolue ni unique à une question d'évaluation et cherche à rendre compte de la pluralité des perceptions des parties prenantes à travers une forte implication de leur part dans le processus d'évaluation.

12. Context, Inputs, Process and Product [Alkin, Crabbé et Leroy]

[*D.L. Stufflebeam. Use*, Approach for Designing Evaluation Research]

- Stufflebeam, D.L. (1983), "The CIPP Model for Program Evaluation", *in* Madaus, G.F., M. Scriven et D. Stufflebeam, *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, Boston, MA.

Cette approche a été développée par Stufflebeam. Il s'agit d'un cadre d'évaluation systémique, englobant, qui vise à assister les décideurs dans les processus de décision. Il invite à évaluer quatre aspects du programme : le contexte, les « *inputs* », le processus et les produits.

13. Context-Adapted Utilization [Alkin]

[Marvin C. Alkin. *Use*]

Cette approche développée par Alkin, est centrée sur le contexte d'utilisation de l'évaluation. Une étape importante, préliminaire, de l'évaluation est le cadrage de celle-ci par rapport au contexte d'utilisation. Puis suit une phase de négociation des critères de l'évaluation, préalable à la conduite de l'évaluation à proprement parler.

14. Cost-Benefit and Cost-Effectiveness Analysis [Stufflebeam, Vedung, Crabbé et Leroy], voir : Benefits-Costs Analysis, Cost-Efficiency Analysis

[Questions and/or Methods Oriented Approaches, Economic. Productivity & Efficiency, Approach for Evaluation]

- Kee, J.E. (2004), "Cost-Effectiveness and Cost-Benefit Analysis", *in* Wholey, J.S., H.P. Hatry et K.E. Newcomer, *Handbook of Practical Program Evaluation*. Second Edition, Wholey Hatry Newcomer Editors, San Francisco.

Ces deux approches, largement reconnues dans le champ de l'évaluation économique, visent à rendre compte de l'efficience d'un programme. Dans l'analyse coût-bénéfices (ou coût-efficience), les bénéfices du programme sont exprimés dans la même unité que les coûts (souvent en termes monétaires), et leur sont comparés. L'analyse coût-efficacité, quant à elle, met en regard les bénéfices du programme, cette fois exprimés dans leur unité physique, avec les coûts (en termes monétaires). Au-delà du calcul de l'efficience ou de l'efficacité d'un programme donné, le but de ces évaluations est de permettre la comparaison avec des programmes alternatifs.

15. Decision / Accountability-Oriented Studies [Stufflebeam]

[Improvement/Accountability Oriented Approaches]

Stufflebeam désigne par cette approche les évaluations visant à générer de l'information sur un programme dans le but de l'améliorer et de rendre les développeurs responsables vis-à-vis des parties prenantes. Comme dans plusieurs autres approches participatives, ce sont les parties prenantes concernées *par* et impliquées *dans* le programme qui définissent les questions qui vont être traitées par l'évaluation. L'évaluation, dans sa mise en oeuvre est également proactive.

16. Deliberative Democratic Evaluation [Stufflebeam, Crabbé et Leroy]

[Social Agenda/Advocacy Approaches, Approaches for Designing Evaluation Research]

- House, E.R. (2000), "Deliberative Democratic Evaluation", *New Directions for Evaluation*, n° 85, pp. 3-12.

Cette approche procédurale a été développée récemment par House et Howe. Pour que l'évaluation parvienne à des conclusions justifiables, elle préconise l'utilisation de trois principes démocratiques : la participation, le dialogue et la délibération. « *It aspires to arrive at unbiased conclusions by considering all relevant interests, values, and perspectives; by engaging in extended dialogue with major stakeholders; and by promoting extensive deliberation about the study's conclusions […]* ».

17- Developmental / Formative Evaluation [Owen, Crabbé et Leroy]

[Interactive, Approach for Evaluation]

- Patton, M.Q. (1994), "Developmental Evaluation", *Evaluation Practice*, 15(3), pp. 311-319.

Désignée sous le nom de « *Formative Evaluation* » ou de « *Developmental Evaluation* », cette approche, largement reconnue en évaluation, est utilisée lors de la mise en œuvre et de l'opérationnalisation des politiques et programmes. Elle vise à formuler des recommandations qui pourront permettre d'en ajuster la réalisation. Dans ce type d'évaluation, une attention importante est accordée au contexte de mise en œuvre de la politique et les divergences entre la politique envisagée « sur le papier » et la politique telle qu'elle est mise en œuvre sont étudiées précisément.

18. Developed Performance Assessment [Owen]

[Monitoring]

Owen désigne sous ce nom les approches de management qui prévoient une évaluation de la performance de toutes les composantes d'un programme (voir « *Component Analysis* »). Le but est d'apprécier la contribution de chacune des composantes à la performance globale du programme.

19. Empowerment Evaluation [Alkin, Owen, Crabbé et Leroy]

[*D. M. Fetterman.* Use, Clarificative, Approaches for Designing Evaluation Research]

- Fetterman, D.M. (2001), *Foundations of Empowerment Evaluation*, Sage, Thousand oaks, CA.

Cette approche, largement reconnue en évaluation, a été développée par Fetterman dans les années 1990. Il la décrit en ces termes « *it is designed to help people help themselves and improve their programmes using a form of self-evaluation and reflection* ». L'hypothèse sous-jacente est la suivante : les changements seront davantage favorisés si l'attention est portée en priorité sur les processus de changement (et non sur les résultats attendus). L'évaluation est conduite par un évaluateur qui a un rôle de facilitateur auprès d'un groupe de parties prenantes concernées par le programme (appartenant à l'équipe de mise en œuvre et / ou au groupe cible). L'accent est mis sur l'autodétermination et l'auto-évaluation. Le but est qu'au sein de ce groupe se forme une vision commune de l'évaluation, autour des objectifs de l'évaluation, des stratégies, des méthodes et des produits attendus. L'auto-évaluation se veut institutionnalisée dans le processus de décision.

20. Evaluability Assessment [Alkin, Owen, Crabbé et Leroy]

[*J.S. Wholey. Use*, Clarificative, Approaches for designing Evaluation Research]

- Smith, M.F. (1989), *Evaluability Assessment: A Practical Approach*, Kluwer Academic Publishers, Clemson, NC.

Cette évaluation a lieu avant toute évaluation, comme une sorte d'évaluation préliminaire de vérification. Son but est d'apprécier si un projet peut être évalué et quels types de barrières vont se présenter lors de l'évaluation. Elle consiste d'une part à explorer et préciser la logique du programme, en explicitant les hypothèses sous-jacentes et les relations causales attendues, et d'autre part, à appréhender les connaissances et les intérêts qu'ont les parties prenantes de ce programme, à travers leurs perceptions.

21. Evaluative Inquiry for Learning in Organizations [Alkin]

[*Hallie Preskill. Use*]

- Preskill, H. et R.T. Torres (1999), *Evaluative Inquiry for Learning in Organizations*, Sage, Thousand Oaks, CA.

Cette approche vise à accroître le rôle de l'évaluation au sein des organisations, en termes d'apprentissage. Elle repose sur un certain nombre d'hypothèses relatives à l'objet de l'évaluation, à la nature de la pratique évaluative, au contexte du travail d'évaluation et au partage des résultats de l'évaluation. Si elle ressemble à de nombreux égards, à d'autres approches évaluatives, qui accordent une place centrale à la dimension participative et à l'implication des parties prenantes, elle s'en distingue toutefois par quelques aspects importants : la réflexivité et l'implantation dans l'organisation, le travail avec les parties prenantes y compris aux derniers stades de l'évaluation, l'apprentissage tiré du processus d'évaluation lui-même et la reconnaissance de la complexité du système englobant l'évaluation et de l'influence de la structure organisationnelle sur l'action publique.

22. *Ex ante* Evaluation [Owen, Crabbé et Leroy], voir : Impact Assessment

[Clarificative, Approach for Evaluation]

- Barrow, C.J. (1997), *Environmental and Societal Impact Assessment: An Introduction*, Wiley and Sons, New York, NY.

Les évaluations *ex ante* visent à évaluer par anticipation, avant la mise en oeuvre du programme, les effets qui pourront être générés. Les évaluations d'impacts sociaux, ou environnementaux, sont des exemples de telles évaluations. Conduites préalablement à la mise en œuvre d'un programme, les évaluations *ex ante* questionnent la validité de ses objectifs et les hypothèses sur la manière de les atteindre.

23. Experiment and Quasi-Experiment [Stufflebeam, Alkin, Crabbé et Leroy], voir : Experimental Studies, Randomized Field Experiment, Generalizability Theory

[Questions and/or Methods Oriented Approaches, *T.D. Cook, L.J. Cronbach, R.F. Boruch, D. Campbell. Method*, Approach for Evaluation]

- Campbell, D. et J. Stanley (1968), *Experimental and Quasi-Experimental Designs for Research*, Rand McNally and Company, Chicago, IL.

Campbell, Stanley puis Cronbach et Cook sont les pionniers dans le développement de ces méthodologies statistiques en

évaluation. Ce type d'évaluation est largement inspiré des sciences expérimentales. Le but est de tester la validité d'une hypothèse causale. Autant que faire se peut, le traitement (l'intervention) est assigné de façon aléatoire à un certain nombre de bénéficiaires. La validité de l'hypothèse est déterminée par la comparaison des résultats obtenus sur ce groupe expérimental (les bénéficiaires de l'intervention) d'une part, et sur le groupe témoin d'autre part.

24. Goal Attainment [Vedung]

[Effectiveness. Objectives]

Ce modèle correspond à la forme la plus ancienne et la plus classique de l'évaluation de l'efficacité. Il consiste à mettre en regard les effets d'un programme avec les objectifs fixés par celui-ci. C'est la conception de l'efficacité retenue par l'OCDE aujourd'hui, de même que par de nombreuses institutions internationales. Il comprend deux volets : 1) l'adéquation entre les effets et les objectifs et 2) l'attribution des effets à telle ou telle action publique précise.

25. Goal-Free Evaluation [Alkin, Vedung, Owen, Crabbé et Leroy]

[*M. Scriven. Valuing*, Effectiveness, Results, Impact, Approach for Evaluation]

- Scriven, M. (1972), "Pros and Cons about Goal-Free Evaluation, Evaluation Comment", *The Journal of Educational Evaluation*, 3(4), pp. 1-7.

Cette approche, largement reconnue en évaluation, s'inscrit en contradiction avec les nombreuses approches fondées sur l'évaluation au regard des objectifs propres de la politique. Dans « l'évaluation libérée des objectifs », l'évaluateur fait comme s'il ne connaissait pas le propos et les objectifs du programme. Son attention se porte simplement vers les impacts effectifs du programme, de quelque nature qu'ils soient. Ces impacts sont ensuite évalués en fonction de leur intérêt ou non pour la société, indépendamment des objectifs initialement affichés.

26. Logframe Method [Crabbé et Leroy]

[Approach for Evaluation]

- SIDA (2004), *The Logical Framework Approach: A Summary of the Theory behind the LFA Method*, SIDA, Stockholm.

Le Cadre logique a été introduit dans les années 1960 aux Etats-Unis, par le département de la Défense, avant d'être repris et approfondi par l'Agence d'aide internationale américaine (USAID). Aujourd'hui, les institutions de coopération et de développement nationales (NORAD, SIDA, DANIDA, CIDA, AusAID) et internationales (Banque mondiale, Commission européenne) l'utilisent largement. Depuis sa création comme cadre de formatage des objectifs de projets, cette approche a évolué vers un véritable outil de rationalisation de la mise en œuvre des projets, permettant d'impliquer de nombreuses parties prenantes à différents stades de la mise en œuvre et de l'évaluation du programme. Elle invite à décliner pour chaque programme, les objectifs stratégiques et opérationnels, les résultats attendus et les activités envisagées. Pour chacun de ces

critères, les hypothèses sous-jacentes, les indicateurs de mesure, les sources d'information et les risques contextuels sont exprimés.

27. Management Information Systems [Stufflebeam]

[Questions and/or Methods Oriented Approaches]

Il s'agit de fournir aux gestionnaires de programmes l'information dont ils ont besoin pour bien conduire leurs programmes et être en mesure d'en rendre compte. Le statut évaluatif d'une telle approche est discutable.

28. Meta-Evaluation / Analyse [Crabbé et Leroy]

[Approaches for Designing Evaluation Research]

- Boruch, R et A. Petrosino (2004), "Meta-Analysis, Systematic Reviews, and Research Syntheses", *in* Wholey, J.S., H.P. Hatry et K.E. Newcomer, *Handbook of Practical Program Evaluation. Second Edition*, Wholey Hatry Newcomer Editors, San Francisco.

Crabbé et Leroy désignent par ces deux noms des évaluations qui s'appuient sur des évaluations : ces évaluations utilisent comme « *inputs* » les résultats d'autres évaluations. La méta-évaluation est l'évaluation d'une évaluation. La méta-analyse est une analyse qui repose sur les données (souvent quantitatives) issues d'évaluations ou d'études qui se sont accumulées sur les mêmes projets ou programmes, visant à tirer davantage d'information et d'objectivité de la capitalisation de ces travaux.

29. Mixed-Method Evaluation [Stufflebeam, Crabbé et Leroy]

[Questions and/or Methods Oriented Approaches, Approaches for Designing Evaluation Research]

- National Science Foundation (1997), *User-Friendly Handbook for Mixed-Method Evaluation: Science, Mathematics, Engineering, and Technical Education*, National Science Foundation, Washington, D.C.

En proposant de combiner, dans une approche mixte, les deux types de méthodes – quantitative (qui s'appuie sur des données numériques mesurables) et qualitative (qui s'appuie sur des descriptions et des analyses non numériques) – cette approche est une réponse au vif débat qui oppose les tenants de chacune d'elles. Tyler est l'un des pionniers de cette approche « mixte ».

30. Multicriteria Analysis [Crabbé et Leroy]

[Approach for Evaluation]

- Janssen, R. et G. Munda (2002), "Multi-Criteria Methods for Quantitative, Qualitative and Fuzzy Evaluation Problems", pp. 831-854, *in* J.C.J.M. van den Bergh, *Handbook of Environmental and Resource Economics*, Edward Elgar, Cheltenham.

Cette approche vise à comparer par des méthodes formalisées et quantitatives plusieurs politiques alternatives les unes aux autres, dans un contexte politique complexe, pour parvenir à identifier la meilleure d'entre elles. L'analyse multicritères cherche à prendre en compte simultanément plusieurs critères. Il existe plusieurs méthodes pour conduire une telle analyse : permutations, matrice de l'atteinte des objectifs, sommes pondérées... A la différence des analyses coûts-bénéfices, les effets ne sont ici pas exprimés en unité monétaire.

31. Needs Analysis / Assessment [Owen, Crabbé et Leroy], voir : Needs-Based

[Proactive. Impact, Approach for Evaluation]

- Reviere, R. (1996), *Needs Assessment: A Creative and Practical Guide for Social Scientists*, Taylor and Francis, Washington, D.C.

L'évaluation des besoins correspond à une évaluation de la pertinence du programme. « *By confronting policy and its results or expected « outcomes » with policy needs, the evaluator assesses policy for its relevance* ». L'explicitation des « besoins » pris en considération (besoins qui correspondent à la différence entre un état désirable et l'état actuel) est un des points clé de cette approche. En effet, ils seront différents selon, d'une part, qui définit cet état « désirable », et d'autre part, quelle est la nature de cet état : état idéal, état normatif ou encore état minimum acceptable.

32. Objective testing programs [Stufflebeam]

[Questions and/or Methods Oriented Approaches]

Stufflebeam définit ce type d'évaluation uniquement dans le champ de l'éducation. Il s'agit de tester les réussites des élèves ou de groupes d'élèves en comparaison avec des normes ou des standards fixés à plus grande échelle (Stufflebeam, 2001, p. 20).

33. Objective-based Studies [Stufflebeam, Alkin, Owen], voir : Objective-Oriented Evaluation

[Questions and/or Methods Oriented Approaches, *R.W. Tyler, Method*, impact]

- Tyler, R. (1950), *Basic Principles of Curriculum and Instructions: Syllabus for Education*, University of Chicago Press, Chicago, IL.

Cette approche, largement reconnue en évaluation, a longtemps été la référence pour les évaluations de l'efficacité. Elle vise à déterminer dans quelle mesure les objectifs fixés par le programme lui-même sont atteints. Typiquement, il s'agit d'une évaluation interne faite par le gestionnaire du programme. Ce type d'évaluation peut être complété par une étude du processus et des « outcomes » du programme (voir « process - outcome » Evaluation) dans la perspective de mieux comprendre dans quelle mesure et comment les objectifs ont été atteints (Stufflebeam, 2001, p. 17).

34. Outcome Evaluation / Value-added Assessment [Stufflebeam]

[Questions and/or Methods Oriented Approaches]

Stufflebeam définit ce type d'évaluation uniquement dans le champ de l'éducation. Il s'agit de déterminer la contribution d'un programme donné aux réussites des élèves.

35. Performance Audit [Owen]

[Impact]

- Chelimsky E. (1985), "Comparing and Contrasting Auditing and Evaluation: Somes Notes on their Relationship", *Evaluation Review*, 9(4), pp. 483-503.

Il s'agit d'une méthode d'évaluation externe, basée sur les résultats. Selon Owen, elle se singularise des autres approches de ce type par le fait que l'évaluation puise être conduite sans la permission de l'agence ou du développeur de programme. Un fort accent est mis sur la vérification et le focus est donc sur les variables de respect de la conformité et sur les variables de gestion. Chelimsky a comparé l'audit de gestion et l'évaluation en s'intéressant à leur histoire et à leur développement ainsi qu'aux méthodologies utilisées par les praticiens des deux domaines.

36. Performance Testing [Stufflebeam]

[Questions and/or Methods Oriented Approaches]

Stufflebeam définit ce type d'évaluation uniquement dans le champ de l'éducation. Il s'agit d'évaluer les compétences développées par les élèves dans des domaines définis, par rapport à des normes données. Ce type d'évaluation ressemble à l'*outcome Evaluation*.

37. Politically Controlled Studies [Stufflebeam]

[Pseudo-Evaluation]

Stufflebeam définit ce type d'évaluation par le fait qu'elle fasse l'objet, à un moment ou à un autre du processus d'évaluation, d'une manipulation de la part du commanditaire: il peut s'agir d'une rétention d'information, d'une promulgation partielle des conclusions de l'évaluation ou encore d'une falsification de ces conclusions (Stufflebeam, 2001, pp. 13–14).

38. Process-Outcome Evaluation [Vedung, Owen], voir : Process Evaluation

[Process, Impact]

L'évaluation de processus peut être envisagée de plusieurs façons ; comme une évaluation procédurale qui porte sur certaines

qualités du processus ou comme une évaluation substantielle qui vise à expliquer les effets du programme (Vedung, 2005, p. 210). Vedung cite six facteurs qui permettent, selon lui, d'expliquer les résultats de l'intervention et qui doivent donc être abordés dans une évaluation de processus : le contexte historique de l'intervention, la conception de l'intervention, sa mise en œuvre, la réponse des personnes ciblées par l'intervention, les autres interventions ou agences interagissant avec l'intervention étudiée et enfin les questions de réseau et d'environnement (Vedung, 2005, p. 210).

39. Program Logic Evaluation [Stufflebeam, Alkin, Owen, Crabbé et Leroy], voir : Theory-Based Evaluation, Program Theory Evaluation, Theory-Driven Evaluation

[Questions and/or Methods Oriented Approaches, *Huey-Tsy Chen. Peter H. Rossi. Carol Hirschon Weiss. Method*, Clarificative, Approach for evaluation]

- Rossi, P.H. et H.E. Freeman (1993), *Evaluation: A Systematic Approach*, 5e éd., Sage, Londres.
- Chen, H. (2004), "The Roots of Theory-Driven Evaluation, Currents Views and Origins", pp. 132-152, *in* Alkin, M.C., *Evaluation Roots, Tracing Theorists' Views and Influences*, Sage, Londres.
- Rogers, P., A. Petrosino, T.A. Huebner et T.A. Hacsi (2000), "Program Theory Evaluation: Practice, Promise and Problems", *New Directions for Evaluation*, n° 87, pp. 5-13.

Cette approche, largement reconnue en évaluation, porte sur la dimension causale de l'évaluation de l'efficacité d'un programme. « *Program Theory Evaluation consists of an explicit theory or model of how the programme causes the intended or observed outcomes and an evaluation that is at least partly guided by this model* » (Rogers *et al.*, 2000). Cette approche a deux dimensions : une dimension normative qui vise à expliciter la théorie d'action sous tendant le programme, et une dimension descriptive qui vise à reconstituer les enchaînements causaux lors de la mise en œuvre. En questionnant le bien fondé et la validité des hypothèses supportant le programme et en reconstituant les processus causaux conduisant aux effets observés, des deux dimensions contribuent grandement à l'évaluation de l'efficacité. Parmi les chercheurs ayant largement participé au développement de cette approche on peut citer C. Weiss avec le concept de «*Theory of Change* », P. Rossi et H-T. Chen.

40. Public Relations-Inspired Studies [Stufflebeam]

[Pseudo-Evaluation]

Stufflebeam désigne sous ce nom des approches dont le but est de convaincre de l'efficacité et de la qualité d'un programme, en mobilisant des données à l'appui. L'accent est mis sur les forces du programme, et dans une bien moindre mesure sur ses faiblesses. On peut qualifier une telle démarche de propagande. Elle est souvent pratiquée par des organisations qui souhaitent donner des gages de leur excellence et s'assurer d'un soutien public (Stufflebeam, 2001, pp. 13–14).

41. Quality Review [Owen]

[Interactive]

Owen désigne par ce nom un type d'évaluations issu du domaine de la gestion, conduites au sein de « sous-groupes logiques » d'une organisation, comme les services d'un établissement public ou privé. Les gestionnaires de ces entités ont la responsabilité de l'évaluation, tout en suivant des lignes directrices définies au niveau macro.

42. Realistic Evaluation [Owen, Crabbe et Leroy]

[Impact, Approach for Evaluation]

- Pawson, R. et N. Tilley (1997), *Realistic Evaluation*, Sage, Etats-Unis.

Cette approche, largement reconnue en évaluation, a été développée par Pawson et Tilley dans les années 1990. Elle repose sur deux propositions fondamentales : 1) les évaluations doivent adopter une épistémologie réaliste, c'est-à-dire qu'il ne s'agit pas tant de prouver des corrélations entre moyens mis en œuvre et résultats que de comprendre comment fonctionnent, dans la mise en œuvre concrète, les mécanismes qui génèrent les résultats et 2) il n'est pas possible de porter un jugement décontextualisé sur l'efficacité d'un mécanisme, il est seulement possible de dire qu'un programme donné conduira à des résultats donnés sous certaines conditions (dans un contexte donné). « *Realistic evaluation is a method for evaluating the effects of policy. The basic assumption underlying the method is that effects are the product of the interaction between a mechanism and a context* ». « *Realistic evaluation is based on the concept of 'generative causation', that assumes that there is an internal explanatory dimension to causality* ».

43. Research Synthesis (Evidence-Based Practice) [Owen]

[Proactive]

Cette pratique repose sur la capitalisation d'expériences passées : les leçons apprises par le passé doivent être réutilisées dans la conception et la mise en œuvre de nouvelles interventions. Cette approche peut être vue comme une « réponse scientifique » à la conception de programmes dans la mesure où elle vise à compléter l'expérience des praticiens et les savoirs locaux par des résultats obtenus « scientifiquement ».

44. Responsive Evaluation [Stufflebeam, Alkin, Owen, Crabbé et Leroy], voir : Participatory Evaluation et Practical Participatory Evaluation

[Social Agenda/Advocacy Approaches, *R. Stake. Valuing*, Interactive, Approaches for Designing Evaluation Research]

- Stake, R.E. (2003), *Standards-Based and Responsive Evaluation*, Sage, New York.

Cette approche, largement reconnue en évaluation, a été développée par Stake au début des années 1970. Elle vise à améliorer un programme à travers l'implication des parties prenantes, le partage et la discussion de l'information relative à celui-ci. Les questions traitées dans l'évaluation sont avant tout des questions qui intéressent les parties prenantes : l'évaluation « répond » à leurs interrogations. Mais les parties prenantes ne conduisent pas l'évaluation directement ; elles la suivent de près, participent à la sélection des méthodes et sont la principale source d'information et de formation du jugement. L'évaluation est davantage orientée vers les activités du programme (la réalité de la mise en œuvre du programme) que vers les intentions du programme (ce qui était envisagé formellement). Ce type d'évaluation est couramment pratiqué par la Banque mondiale.

45. Review of best practice [Owen]

[Proactive]

- Bendell, T., L.Boulter et J. Kelly (1993), *Benchmarking for Competitive Advantage*, Pitman Publishing, Londres.

Owen désigne sous ce nom la pratique de « *benchmarking* » couramment employée en gestion depuis les années 1990. Elle consiste à comparer les processus mis en œuvre par plusieurs organisations pour des finalités similaires, dans le but d'identifier les meilleures pratiques et d'améliorer la performance.

46. Side-Effects Evaluation [Vedung]

[Effectiveness. Objectives]

Vedung définit cette approche en réaction à l'approche classique de « *goal attainment* » qui ne se préoccupe que de l'atteinte des objectifs fixés par le programme. L'évaluation de l'efficacité dite « *side-effects* » a une portée plus large, puisqu'elle prend en compte également des effets inattendus générés par le programme.

47. Stakeholder Model [Vedung]

[Effectiveness. Stakeholder concerns]

- Guba, E.G. et Y.S. Lincoln (1989), *Fourth Generation Evaluation*, Sage, Newbury Park, CA.

Vedung désigne sous ce nom les évaluations de l'efficacité qui prennent comme référence les préoccupations de parties prenantes intéressées ou affectées par le programme. Ces évaluations se distinguent des évaluations « *client-oriented* » de par leur portée : elles ne s'adressent pas à un unique groupe de parties prenantes (les cibles du programme), mais à des groupes d'intérêts plus nombreux.

48. System Analysis [Owen]

[Monitoring]

- Rossi, P.H. et H.E. Freeman (1993), *Evaluation: A Systematic Approach*, 5e Ed., Sage, Londres.

Owen désigne sous ce nom les évaluations de l'efficacité et de l'efficience d'un certain type de programmes : les programmes mis en œuvre sur plusieurs sites mais qui sont définis de façon centralisée. Ce type d'évaluation vise à comparer la performance sur les différents sites en utilisant les mêmes critères opérationnels.

49. Utilization-Focused Evaluation [Stufflebeam, Alkin, Crabbé et Leroy]

[Social agenda/Advocacy Approaches, Michael Q. Patton. Use, Approaches for Designing Evaluation Research]

- Patton, M.Q. (2008), *Utilization-Focused Evaluation*. 4th Edition, Sage, Etats-Unis.

Cette approche, largement reconnue en évaluation, est issue des travaux de Patton dans les années 1980. Elle considère que ce sont avant tout des critères d'utilisation et d'utilité des évaluations qui doivent guider les choix de méthode. Cette approche accorde donc une grande importance à l'identification des utilisateurs de l'évaluation et à leur implication (qu'il s'agisse des décideurs, des réalisateurs du programme ou d'autres parties prenantes). « *The ambition is to enhance the target-oriented utilization of the evaluation by its target group. It does not stand for a specific research approach, but rather enhances the situational responsiveness, implying that, depending upon the circumstances, a choice is made for a particular [evaluation] research method and techniques* » (Patton, 1980).

Liste des sigles et abréviations

AEE	Agence européenne pour l'environnement
BAU	*Business As Usual*
BID	Banque interaméricaine de développement
CDB	Convention sur la diversité biologique
CFI	Initiative internationale Climat-Forêt (Norvège)
CIRED	Centre international de recherche pour l'environnement et le développement
CCNUCC	Convention cadre des Nations unies sur le changement climatique
EESS	Evaluation environnementale et sociale stratégique
EPA	*Environmental Protection Agency*
FAO	*Food and Agriculture Organisation*
FPCF	Fonds de partenariat pour le carbone forestier
FEM	Fonds pour l'environnement mondial
GES	Gaz à effet de serre
ICFI	Initiative internationale Climat – Forêt
MDP	Mécanisme de développement propre
ONG	Organisation non gouvernementale
PIF	Programme d'investissement pour les forêts
PNUD	Programme des Nations unies pour le développement

PNUE	Programme des Nations unies pour l'environnement
RCT	*Randomized Controlled Trial*
RED	Réduction des émissions liées à la déforestation
REDD+	Réduction des émissions liées à la déforestation et à la dégradation des forêts

Bibliographie

Alkin, M.C. (2004), *Evaluation Roots, Tracing Theorists' Views and Influences*, SAGE, Londres.

Andam, K.S., P.J. Ferraro, A. Pfaff, G. A. Sanchez-Azofeifa et J. A. Robalino (2008), "Measuring the Effectiveness of Protected Area Networks in Reducing Deforestation", *Proceedings of the National Academy of Sciences*, 105(42), pp. 16089 -16094.

Association for Tropical Biology and Conservation et Society for Tropical Ecology (2009), "The Marburg Declaration. The Urgent Need to Maximize Biodiversity Conservation in Forest Carbon-Trading", communiqué conjoint, Marbourg, pp. 26-29, juillet.

Baastel et Nordeco (2011), *First Program Evaluation for the Forest Carbon Partnership Facility. Evaluation Report,* p. 174, Forest Carbon Partnership Facility, Washington, D.C.

Banque mondiale (2009), *Convenient Solutions to an Inconvenient Truth: Ecosystem-based Approaches to Climate Change,* p. 91, Environment Department of the World Bank, Washington, D.C.

Barrow, C.J. (1997), *Environmental and Societal Impact Assessment: An Introduction*, Wiley and Sons, New York, NY.

Bendell, T., L. Boulter et J. Kelly (1993), *Benchmarking for Competitive Advantage*, Pitman Publishing, Londres.

Bennear, L.S. et C. Coglianese (2005), "Measuring Progress. Program Evaluation of Environmental Policies", *Environnement*, 47(2), pp. 22-39.

Bernard, C., K. Belna et T. Leménager (2012), « La préparation du mécanisme REDD+ au Mexique et ses perspectives environnementales pour les forêts tropicales », *Focales,* n° 13, Agence Française de Développement, Paris.

Berriet-Solliec, M., P. Labarthe, C. Laurent et J. Baudry (2011), "Empirical Validity of the Evaluation of Public Policies: Models of Evaluation and Quality of Evidence", présenté au colloque : Evidence-Based Agricultural and Rural Policy Making: Methodological and Empirical Challenges of Policy Evaluation, Ancona.

Billé, R. (2007), "A Dual-level Framework for Evaluating Integrated Coastal Management beyond Labels", *Ocean & Coastal Management*, n° 50, pp. 796-807.

Boruch, R et A. Petrosino (2004), "Meta-Analysis, Systematic Reviews, and Research Syntheses", *in* Wholey, J.S., H.P. Hatry et K.E. Newcomer, *Handbook of Practical Program Evaluation, Second Edition,* Wholey Hatry Newcomer Editors, San Francisco.

Bruner, A.G., R.E. Gullison, R.E. Rice et G.A.B. da Fonseca (2001), "Effectiveness of Parks in Protecting Tropical Biodiversity", *Science*, n° 291, pp. 125-128.

Campbell, D. et J. Stanley (1968), *Experimental and Quasi-Experimental Designs for Research,* Rand McNally and Company, Chicago, IL.

CCBA (2010), *Standards sociaux et environnementaux pour REDD+, version du 1er juin,* p. 18, CCBA, Washington, D.C.

CCNUCC (2007), *FCCC/CP/2007/6/Add.1, Decision 2/CP.13, Reducing Emissions from Deforestation in Developing Countries: Approaches to Stimulate Action*, pp. 8-11, UNFCCC, Bonn.

CCNUCC (2009*a*), *FCCC/CP/2009/L.7, Draft Decision -/CP.15, Proposal by the President, Copenhagen Accord,* UNFCCC, Bonn.

CCNUCC (2009*b*), *FCCC/AWGLCA/2009/14, III.C. Policy approaches and positive incentives on issues relating to reducing emissions from deforestation and forest degradation in developing countries; and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries (paragraph 1 (b) (iii) of the Bali Action Plan,* UNFCCC, Bonn.

CCNUCC (2010*a*), *FCCC/CP/2009/11/Add.1. Copenhagen Accord. Dec.2/CP.15*, p. 7, UNFCCC, Bonn.

CCNUCC (2010*b*), *Advance Unedited Version, Draft decision -/CP.16. Outcome of the Work of the Ad Hoc Working Group on Long-Term Cooperative Action under the Convention*, p. 29*,* UNFCCC, Bonn.

CCNUCC (2010*c*), *FCCC/AWGLCA/2010/14:Ad Hoc Working Group on Long-Term Cooperative Action under the Convention. Twelfth Session. Tianjin, 4-9 October, Negotiating Text,* UNFCCC, Bonn.

CCNUCC (2011), FCCC/CP/2010/7/Add.1, *Decision 1/CP.16 The Cancun Agreements: Outcome of the Work of the Ad Hoc Working Group on Long-term Cooperative Action under the Convention*, p. 31*,* UNFCCC, Bonn.

CDB (2010), *COP 10 Decision X/2 X/2, Strategic Plan for Biodiversity 2011-2020*, CBD, Montreal.

CDB Secretariat (2009*a*), "Forest Resilience, Biodiversity and Climate Change: A Synthesis of the Biodiversity/Resilience/Stability Relationship in Forest Ecosystems", *CDB Technical Series*, *43*, p. 67, Montreal.

CDB Secretariat (2009*b*), "Connecting Biodiversity and Climate Change: Mitigation and Adaptation, Report of the Second *Ad Hoc* Technical Expert Group on Biodiversity and Climate Change", *CDB Technical Series*, n° 41, p.127, Montreal.

CDB Secretariat (2011), "REDD+ and Biodiversity", *CDB Technical Series,* n° 59, p. 65, Montreal.

CDB Secretariat et GIZ (2011), *Biodiversity and Livelihoods. REDD-Plus Benefits,* p. 44, Montreal et Eschborn, Montreal.

Chelimsky E. (1985), "Comparing and Contrasting Auditing and Evaluation: Somes Notes on their Relationship", *Evaluation Review,* 9(4), pp. 483-503.

Chen, H. (2004), “The Roots of Theory-Driven Evaluation, Currents Views and Origins”, *in* Alkin, M.C., *Evaluation Roots, Tracing Theorists' Views and Influences*, pp. 132-152, Sage, Londres.

Chen, H.-T. (1990), *Theory-Driven Evaluations,* Sage Publications, California.

Crabbé, A. et P. Leroy (2008), *The Handbook of Environmental Policy Evaluation*, Earthscan, Londres.

Delarue, J., J.-D. Naudet et V. Sauvat (2009), « Les évaluations sont-elles utiles ? Revue de littérature sur « connaissances et décisions », *Notes méthodologiques*, n° 3, Agence Française de Développement, Paris.

Diamond, J. (2006), *Collapse: How Societies Choose to Fail or Succeed,* Penguin, New York.

Eisner, E.W. (1986), "Educational Connoisseurship and Criticism: their Form and Functions in Educational Evaluation", *in* Madaus, G.F., M. Scriven et D. Stufflebeam*, Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, Kluwer-Nijhoff, Boston, MA.

FEM (2011), *GEF 5. Focal Area Strategies,* p. 116, GEF, Washington, D.C.

Ferraro, P.J. (2009), "Counterfactual Thinking and Impact Evaluation in Environmental Policy", *New Directions for Evaluation*, n° 122, pp. 75-84.

Fetterman, D.M. (2001), *Foundations of Empowerment Evaluation*, Sage, Thousand Oaks, CA.

Fitzpatrick, J.L., J.R. Sanders et B.R. Worthen (2004), *Program Evaluation, Alternative Approaches and Practical Guidelines,* Pearson education, Etats-Unis.

FPCF (2008), *Forest Carbon Partnership Facility, Information Memorandum*, p. 121, International Bank for Reconstruction and Development, Washington, D.C.

FPCF (2010), *Incorporating Environmental and Social Considerations into the Process of Getting Ready for REDD plus. Revised DRAFT, March 7,* Note FMT No. 2010-9, World Bank, Washington, D.C.

FPCF, Readiness Fund (2011), *Common Approach to Environmental and Social Safeguards for Multiple Delivery Partners,* Washington, D.C.

Georgieva, K. (2007), Re: Forest Carbon Partnership Facility (FPCF), World Bank, Washington, D.C.

GIEC (2007), *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel of Climate Change,* Cambridge.

Global Witness (2010), *Comments on Component 4 of the R-PP Template v. 5 Draft, October 30,* Washington, D.C.

Glouberman, S. et B. Zimmerman (2002), *Complicated and Complex Systems: What would Successful Reform of Medicare Look Like?,* Commission on the Future of Health Care in Canada*,* Discussion Paper 8.
Available at http:// www.hc-sc.gc.ca/english/pdf/romanow/ pdfs/8_Glouberman_E.pdf

Grainger, A. (2008), "Difficulty in Tracking the Long-Term Global Trend in Tropical Forest Area", *Proceedings of the National Academy of Sciences*, *105*(2), pp. 818-823.

Greenpeace International (2007), *Letter to Kathy Sierra, Vice President, Sustainable Development, November 20*, The World Bank, Washington, D.C.

Guba, E.G. et Y.S. Lincoln (1989), *Fourth Generation Evaluation*, Sage, Newbury Park, CA.

Guedes Vaz, S., J. Martin, D. Wilkinson et J. Newcombe (2001), *Reporting on Environmental Measures: Are we being effective?*, n° 25, European Environment Agency, Luxembourg.

Guéneau, S. (2011), *Vers une évaluation des dispositifs de prise en charge du problème du déclin des forêts tropicales,* AgroParisTech, Paris.

Gysen, J., K. Bachus et H. Bruyninckx, (2002), "Evaluating the Effectiveness of Environmental Policy: An Analysis of Conceptual and Methodological Issues", p. 17, présenté à la Conférence de Séville de la Société européenne d'évaluation, Séville.

Gysen, J., K. Bachus et H. Bruyninckx (2006), "The Modus Narrandi: A Methodological for Evaluating Effects on Environmental Policy", *Evaluation*, 12(1), pp. 95-118.

Haas, P. M., R.O. Keohane et M.A. Levy (1993), *Institutions for the Earth. Sources of Effective International Environmental Protection*, The MIT Press, Cambridge.

Harvey, C.A., B. Dickson et C. Kormos (2010), "Opportunities for Achieving Biodiversity Conservation through REDD", *Conservation Letters*, n° 3, pp. 53-61.

Hilden, M. (2009), "Time Horizons in Evaluating Environmental Policies", *New Directions for Evaluation*, n° 122, pp. 9-18, University Press, Cambridge, UK et New York.

House, E.R. (2000), "Deliberative Democratic Evaluation", *New Directions for Evaluation*, n° 85, pp. 3-12.

Janssen, R. et G. Munda (2002), "Multi-Criteria Methods for Quantitative, Qualitative and Fuzzy Evaluation Problems", *in* J.C.J.M. van den Bergh, *Handbook of Environmental and Resource Economics,* pp. 831-854, Edward Elgar, Cheltenham.

Karousakis, K. (2009), "Promoting Biodiversity Co-Benefits in REDD", *OECD Environment Working Papers*, n° 11, p. 25.

Karsenty, A. (2008), "The Architecture of Proposed REDD Schemes after Bali: Facing Critical Choices", *International Forestry Review*, 10(3), pp. 443-457.

Karsenty, A. et P. Fournier (2008), « États défaillants : le secteur forestier en Afrique centrale », *Monde en développement*, 3(143), pp. 43-56.

Kaufman, R.A. et S. Thomas (1980), *Evaluation without Fear*, New Viewpoints, New York.

Kautto, P. et J. Similä (2005), "Recently Introduced Policy Instruments and Intervention Theories", *Evaluation*, n° 11, pp. 55-68.

Kee, J.E. (2004), "Cost-Effectiveness and Cost-Benefit Analysis", *in* Wholey, J. S., H.P. Hatry et K.E. Newcomer, *Handbook of Practical Program Evaluation, Second Edition,* Wholey Hatry Newcomer Editors, San Francisco.

Knaap, G.J. et J.K. Tschangho (1998), *Environmental Program Evaluation, a Primer*, University of Illinois Press, Etats-Unis.

Krasner, S.D. (1983), "Structural Causes and Regime Consequences: Regimes as Intervening Variables", *in* Krasner, S.D., *International Regimes,* pp. 1-21, Ithaca: Cornell University Press.

Lafferty, W.M. et E. Hovden (2003), "Environmental Policy Integration: Towards an Analytical framework", *Environmental Politics*, 12(3), pp.1-22.

Laurans, Y. et I. Dubien (2000), *Nature et place des arguments sanitaires dans les négociations autour des implantations d'incinérateurs*, ASCA, Paris.

Leroy, M. (2006), *Gestion stratégique des écosystèmes du fleuve Sénégal*, L'Harmattan, Paris.

Lessinger, L.M. (1970), *Every Kid a Winner: Accountability in Education,* Simon and Schuster, New York.

Mermet, L. (1996), « Les études d'évaluation, entre stratégie et méthodologie - L'exemple des politiques publiques en matière de zones humides », *Gérer et Comprendre, Annales des mines*, n° 46, pp. 55-64.

Mermet, L. (2011), "Strategic Environmental Management Analysis: Adressing the Blind Spots of Collaborative Approaches", *Idées pour le débat*, IDDRI, pp. 1-32.

Mermet, L., R. Billé et M. Leroy (2010), "Concern-Focused Evaluation for Ambiguous and Conflicting Policies: An Approach from the Environmental Field", *American Journal of Evaluation*, 31(2), pp. 180-198.

Mickwitz, P. (2003), "A Framework for Evaluating Environmental Policy Instruments: Context and Key Concepts", *Evaluation*, 9(4), pp. 415-436.

Mickwitz, P. et M. Birnbaum (2009), "Editors' Note", *New Directions for Evaluation*, n° 122, pp. 1-8.

Mickwitz, P. et P. Kivimaa (2007), "Evaluating Policy Integration: The Case of Policies for Environmentally Friendlier Technological Innovations", *Evaluation*, 13(1), pp. 68-86.

Miles, E.L., A. Underdal, S. Anderson, J. Wettestad, J.B. Skjaerseth et E.M. Carlin (2002), *Environmental Regime Effectiveness, Confronting Theory with Evidence*, The MIT Press, Cambridge, Massachusetts.

Millennium Ecosystem Assessment (2005), *Ecosystems and Human Well-Being: Synthesis*, Island Press, Washington, D.C.

Ministerio de Ambiente y Desarollo Sustenible (2011), *Propuesta de preparation para REDD+*, p. 236, Ministerio de Ambiente y Desarollo Sustenible, Bogota.

Mitchell, R.B. (2008), "Evaluating the Performance of Environmental Institutions: What to Evaluate and how to Evaluate it?", *in* Young, O.R., L.A. King et H. Shroeder, *Institutions and Environmental Change: Principal Findings, Applications and Research Frontier*, pp. 79-114, MIT Press, Cambridge, MA, Etats-Unis.

Morin, E. (2004), *Introduction à la pensée complexe*, Seuil, Paris.

National Science Foundation (1997), *User-Friendly Handbook for Mixed-Method Evaluation: Science, Mathematics, Engineering, and Technical Education,* National Science Foundation, Washington, D.C.

Naudet, J.-D. (2012), « Les réformes de l'aide au développement en perspective de la nouvelle gestion publique », *Document de travail*, n° 119, p. 34, Agence Française de Développement, Paris.

OCDE (2002), *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Paris.

ONFI (2011), « Le mécanisme REDD+ de l'échelle mondiale à l'échelle locale. Enjeux et conditions de mise en œuvre », *Document de travail*, n° 109, p. 85, Agence Française de Développement, Paris.

ONU (1992), *Convention sur la diversité biologique*, Rio de Janeiro.

Owen, J.M. (2007), *Program Evaluation. Forms and Approaches*, The Guilford Press, New York et Londres.

Patton, M.Q. (1994), "Developmental Evaluation", *Evaluation practice, 15*(3), pp. 311-319.

Patton, M.Q. (2008), *Utilization-Focused Evaluation. 4th Edition*, Sage, Etats-Unis.

Pawson, R. et N. Tilley (1997), *Realistic Evaluation*, Sage, Etats-Unis.

Pirard, R. et K. Belna (2012), "Agriculture and Deforestation: Is REDD+ Rooted in Evidence?", *Forest Policy and Economics,* n° 21, pp. 62-70.

Pistorius, T., C.B. Schmitt, D. Benick et S. Entenmann (2010*a*), *Greening REDD+, Challenges and Opportunities for Forest Biodiversity Conservation,* Policy Paper, p. 43, University of Freiburg, Fribourg.

Pistorius, T., C.B. Schmitt, D. Benick et S. Entenmann (2010*b*), *Greening REDD+, Challenges and Opportunities for Forest Biodiversity Conservation, Executive Summary,* Policy Paper, p. 5, University of Freiburg, Fribourg.

Preskill, H. et R.T. Torres (1999), *Evaluative Inquiry for Learning in Organizations*, Sage, Thousand Oaks, CA.

Programme ONU-REDD (2012), *Principes et critères sociaux et environnementaux du Programme ONU-REDD*, Genève.

Putz, F.E. et K.H. Redford (2009), "Dangers of Carbon-Based Conservation", *Global Environmental Change*, n° 19, pp. 400-401.

Reviere, R. (1996), *Needs Assessment: A Creative and Practical Guide for Social Scientists*, Taylor and Francis, Washington, DC.

Rogers, P. (2008), "Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions", *Evaluation*, 14(1), pp. 29-48.

Rogers, P., A. Petrosino, T.A. Huebner et T.A. Hacsi (2000), "Program Theory Evaluation: Practice, Promise and Problems", *New Directions for Evaluation*, n° 87, pp. 5-13.

Rossi, P.H. et H.E. Freeman (1993), *Evaluation: A Systematic Approach,* 5th Ed., Sage, Londres.

RUDEL, T.K. (2005), "Forest Transitions: Toward a Global Understanding of Land Use Change", *Global Environmental Change*, n° 15, pp. 23-31.

SAMUELSON, P.A. (1954), "The Pure Theory of Public Expenditure", *Review of Economics and Statistics,* 36(4), pp. 387-389.

SASAKI, N. ET F.E. PUTZ (2009), "Critical Need for New Definitions of « Forest » and « Forest Degradation » *in* Global Climate Change Agreements", *Conservation Letters*, 2(5), pp. 226-232.

SCRIVEN, M. (1967), "The Methodology of Evaluation", *in* Stake, R.E., *Curriculum Evaluation, AERA Monograph Series on Curriculum Evaluation,* Rand Mc Nally, Vol. 1, Chicago.

SCRIVEN, M. (1972), "Pros and Cons about Goal-Free Evaluation, Evaluation Comment", *The Journal of Educational Evaluation,* 3(4), pp. 1-7.

SHEIL, D., A. CASSON, E. MEIJAARD, M. VAN NOORDWIJK, J. GASKELL, J. SUNDERLAND-GROVES, K. WERTZ ET M. KANNINEN (2009), *The Impacts and Opportunities of Oil Palm in Southeast Asia - What do we Know and what do we Need to Know?*, Occasional Paper n° 51, p. 80, CIFOR, Bogor.

SIDA (2004), *The Logical Framework Approach: A Summary of the Theory behind the LFA Method,* SIDA, Stockholm.

SIMULA, M. (2009), *Vers une définition de la dégradation des forêts - Analyse comparative des définitions existantes,* p. 66, FAO, Rome.

SIMULA, M. (2010), *Analysis of REDD+ Financing Gaps and Overlaps*, p. 99, REDD+ Partnership.

SLUNGE, D., S. NOOTEBOOM, A. EKBOM, G. DIJKSTRA ET R. VERHEEM (2009), *Conceptual Analysis and Evaluation Framework for Institution-Centered Strategic Environmental Assessment, Final Draft*, *revised version, June 23*, p. 64.

SMITH, M.F. (1989), *Evaluability Assessment: A Practical Approach*, Kluwer Academic Publishers, Clemson, NC.

SMOUTS, M.-C. (2008), "The Issue of an International Forest Regime", *International Forestry Review*, 10(3), pp. 429-432.

STAKE, R.E. (2003), *Standards-Based and Responsive Evaluation,* Sage, New York.

STAME, N. (2004), "Theory-based Evaluation and Types of Complexity", *Evaluation*, 10(1), pp. 58-76.

STUFFLEBEAM, D. (2001), "Evaluation Models", *New Directions for Evaluation*, n° 89, pp. 7-98.

STUFFLEBEAM, D.L. (1983), "The CIPP Model for Program Evaluation", *in* Madaus, G.F., M. Scriven et D. Stufflebeam, *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, Boston, MA.

TEEB (2009), *Chapter 2: Biodiversity, Ecosystems and Ecosystem Services, the Economics of Ecosystems and Biodiversity: The Ecological and Economic Foundations, Vol. 2,* p. 84, UNEP, Nairobi.

TIKINA, A.V. ET J.L. INNES (2008), "A Framework for Assessing the Effectiveness of Forest Certification", *Canadian Journal of Forest Research*, n° 38, pp. 1357-1365.

Tyler, R. (1950), *Basic Principles of Curriculum and Instructions: Syllabus for Education,* University of Chicago Press, Chicago, IL.

Underdal, A. (1992), "The Concept of Regime `Effectiveness'", *Cooperation and Conflict*, 27(3), pp. 227-240.

Underdal, A. (2002), "One Question, Two Answers", *in* Miles, E.L., A. Underdal, S. Anderson, J. Wettestad, J.B. Skjaerseth et E.M. Carlin*, Environmantal Regime Effectiveness, Confronting Theory with Evidence,* pp. 3-45, MIT Press, Cambridge, Massachusetts.

Van der Werf, G.R., D.C. Morton, R.S. DeFries, J.G.J. Olivier, P.S. Kasibhatla, R.B. Jackson, G.J. Collatz et J.T. Randerson (2009), "CO_2 Emissions from Forest Loss", *Nature Geoscience*, n° *2*, pp. 737-738.

Vedung, E. (2005), *Public Policy and Program Evaluation*, Transaction Publishers, New Brunswick et Londres.

Vedung, E. (2009), "Book Review: The Handbook of Environmental Policy Evaluation, Ann Crabbe and Pieter Leroy", *Evaluation*, *15*(4), pp. 483-486.

Vedung, E. (2010), "Four Waves of Evaluation Diffusion", *Evaluation*, *16*(3), pp. 263-277.

Victor, D. G., K. Raustiala et E.B. Skolnikoff (1998), *The Implementation and Effectiveness of International Commitments, Theory and Practice,* IIASA, the MIT Press, Cambridge, Massachusetts.

Weiss, C.H. (2004), "Rooting for Evaluation: A Cliff Notes Version of My Work", *in* Alkin, M.C., *Evaluation Roots, Tracing Theorists' Views and Influences*, pp. 153-169, SAGE, Londres.

Wertz-Kanounnikoff, S. et M. Kongphan-Apirak (2009), *Emerging REDD+. A Preliminary Survey of Demonstration and Readiness Activities*, Working Paper n° 46, p. 58, CIFOR, Bogor.

Wholey, J. S., H.P. Hatry et K.E. Newcomer (2004), *Handbook of Practical Program Evaluation,* 2nd Edition*,* Wholey Hatry Newcomer Editors, San Francisco.

Wolf, R. L. (1975), "Trial by Jury: A New Evaluation Method", *Phi Delta Kappan*, *57*(3), pp. 185-187.

Yin, R. (2003), *Case Study Research: Design and Methods*, Thousand Oaks, Sage.

Young, O.R. (1994), *International Governance: Protecting the Environment in a Stateless Society,* Cornell University Press, New York.

Young, O.R. (1999), *The Effectiveness of International Environmental Regimes: Causal Connections and Behavioural Mechanisms*, the MIT Press, Cambridge, Massachusetts.