



UMR 1041 INRA – AGROSUP

CESAER



Centre d'Economie et Sociologie
appliquées à l'Agriculture et aux Espaces Ruraux

**L'évaluation des politiques à
l'épreuve des faits :
Apports des débats autour des approches
evidence based policies**

Marielle Berriet-Sollic

Pierre Labarthe

Jacques Baudry

Working Paper

2010/5

**L'évaluation des politiques à l'épreuve des faits :
apports des débats autour des approches *evidence based policies*¹**

Marielle Berriet-Sollic
ENESAD-CESAER, Dijon
Berriet@enesad.inra.fr

Pierre Labarthe,
INRA-SADAPT, Paris
Pierre.labarthe@agroparistech.fr

Jacques Baudry
INRA Rennes
Jacques.baudry@rennes.inra.fr

2èmes journées de recherches en sciences sociales

INRA SFER CIRAD

11 & 12 décembre 2008 – LILLE, France

¹ L'article proposé repose sur un travail de recherche collectif réalisé dans le cadre d'un programme, intitulé EBP-BIOSOC et coordonné par Catherine Laurent de l'INRA-Sadap (Laurent, 2006). Ce programme, qui associe pendant trois ans (2007-2010) des équipes françaises, brésiliennes et sud-africaines, est financé par l'Agence Nationale de la Recherche (ANR) dans le cadre de l'appel d'offre « Agriculture et Développement Durable » (ADD).

Résumé

Les méthodes couramment utilisées pour l'évaluation des politiques publiques rencontrent deux principaux écueils. Le premier est d'analyser les relations de cause à effet uniquement à partir des opinions et avis des parties prenantes (*stakeholders*). Quand ces schémas de causalité mobilisent les théories en sciences sociales (*theory driven evaluation*), le second écueil est qu'ils sont rarement confrontés aux faits. Pour pallier cette défaillance de validité empirique, les approches evidence-based proposent de mobiliser des connaissances scientifiques testées empiriquement. Toutefois, même si ces approches sont répandues dans les domaines de l'éducation ou de la santé dans les pays anglo-saxons, elles s'avèrent encore très rarement utilisées dans les deux domaines d'application étudiés dans ce papier, le conseil en agriculture et l'agri-environnement.

Abstract

The methods for the evaluation of public policies face two major limits. The first one happens when the schemes of causal relations involved in a given public policy are analysed only through the opinions and the knowledge of some stakeholders. When the description of these schemes is rather driven by some social sciences theories, a second limit may occur: the lack of confrontation with field observations. Both methods raise the question of their level of validity. In that respect, the EBP approaches proposes to ground the evaluation of public policies on scientific knowledge that rely on the best available empirical evidences. These approaches and the debates about them are well spread in the cases of health or education policies evaluation in Anglo-Saxon countries. However, they are much less implemented in the two cases studied in this paper: agri-environnement and agricultural extension.

Mots clés : Evaluation, Evidence Based Policies, Conseil agricole, agri-environnement, biodiversité, agriculture

Key words: Evaluation, Evidence Based Decision, Agricultural extension, agri-environment, biodiversity, agriculture

Classification JEL : Q16, Q18, Q59, H50.

En France, dans le cadre de la modernisation des politiques publiques, l'évaluation apparaît comme un des outils pour accroître l'efficacité de l'action publique, mieux maîtriser les dépenses publiques et rendre plus transparent, pour le citoyen, l'utilisation des fonds publics. Toutefois, les pratiques d'évaluation renvoient à des réalités très différentes tant du point de vue conceptuel que méthodologique. En outre, en raison parfois d'une inadéquation des modèles théoriques utilisés, les méthodes d'évaluation ne s'avèrent pas toujours aptes à rendre compte des transformations techniques et socio-économiques. Il en résulte alors de sévères limites des résultats d'évaluation à éclairer le décideur public et à orienter ses choix.

Partant de ce constat, cet article propose, dans un premier temps, de rappeler les principales démarches évaluatives en jetant un regard critique sur les méthodes aujourd'hui mobilisées. Cette analyse conduit à souligner le décalage croissant qui existe, dans les démarches économiques les plus courantes, entre les solutions envisagées et la réalité observée. Cela pose la question de la validation empirique des modèles économiques utilisés pour l'évaluation des politiques publiques.

Ces constats sont étayés par des illustrations dans deux cas d'étude : celui du conseil² agricole d'une part, celui des mesures agro-environnementales d'autre part. L'intérêt de l'exemple du conseil agricole est que derrière la grande diversité des politiques publiques de soutien au conseil agricole se cachent des controverses sur les schémas explicatifs relatifs à l'efficacité de ces politiques (Labarthe 2006, 2009, Carney 1995). Les politiques agri-environnementales, notamment portant sur la biodiversité, ont été mises en place en réponse aux diminutions de la faune et de la flore observées dans les paysages agricoles. Les débats sur leur efficacité, voire même sur leurs objectifs sont nombreux (Kleijn et Sutherland, 2003).

Dans les deux cas, les difficultés d'accès aux données pour évaluer les interventions rendent crucial le développement de méthodes d'évaluation reposant sur une meilleure connaissance des faits et des mécanismes sous-jacents aux interventions à évaluer.

Relativement peu étudiée par les chercheurs du développement économique, cette question a d'ores et déjà fait l'objet de recherches dans le domaine médical et dans celui de l'éducation (Peersman *et al.*, 2001 ; Davies et Nutley, 2001). Elle s'incarne notamment dans les débats autour des approches en termes d'evidence based decision (EBD) conduisant aux evidence based policies (EBP) (Davies, Nutley 2001) que nous proposons de présenter dans une seconde partie de cet article, avant de discuter dans une troisième partie des conditions de mise en œuvre et des difficultés spécifiques à cette approche.

1. Diversité des démarches évaluatives : quelle validité empirique des modèles d'évaluation?

² Le conseil agricole peut être défini comme une activité de service permettant d'accompagner des changements dans les pratiques des agriculteurs (Laurent *et al.* 2006). Ces services sont considérés comme un levier d'action pour encadrer des changements techniques souhaités au niveau des exploitations agricoles vis-à-vis de différents objectifs : productivité, protection de l'environnement, etc. Les investissements publics dans cette forme de soutien technique ont pris une grande diversité de formes et ont subi de nombreuses réformes (Arnon 1989, Rivera et Zijp 2001), aussi bien dans les pays industrialisés que dans ceux en voie de développement (Rivera 2002). Ces réformes concernent les modalités de financement et de gestion de l'offre de services, le type de lien à la recherche ou encore la définition des publics cibles.

La littérature sur l'évaluation des politiques publiques fait référence à une grande diversité de postures méthodologiques et de modèles, à tel point que certains auteurs ont proposé des classifications pour aider les commanditaires à se repérer face à ces différentes pratiques évaluatives. Hansen (2005) distingue ainsi six catégories de modèles d'évaluation dont les modèles économiques *stricto sensu* (analyse coûts-avantages, coût efficacité ou coûts-bénéfices), les modèles systémiques, les modèles d'acteurs ou encore les modèles reposant sur la théorie du programme (*programme theory models*). Notre attention se porte plus particulièrement sur ces modèles, qui mettent l'accent sur l'analyse des mécanismes sous-jacents à l'intervention publique. Après les avoir brièvement présentés (1.1), nous rappelons les principales méthodes permettant d'analyser les relations causales reliant la mise en place d'une mesure et ses répercussions potentielles sur une cible (individu ou zone) (1.2). Mais, dans les cas où une telle analyse est menée, rares sont les travaux qui s'attachent à une confrontation empirique des hypothèses émises dans ces modèles (1.3).

1.1 Les modèles d'évaluation reposant sur la théorie du programme

Les modèles reposant sur la théorie du programme, quand ils sont mobilisés dans les démarches d'évaluation de politique, concernent l'analyse et l'interprétation des relations de cause à effet qui peuvent exister entre la mise en place d'une mesure par les pouvoirs publics et les effets potentiels qui peuvent en résulter. Les effets peuvent concerner directement la cible visée par la mesure ou encore le contexte englobant. Le rôle de l'évaluateur consiste, plus précisément, à émettre des hypothèses sur l'enchaînement des relations causales entre une mesure et ses effets successifs potentiels. Il s'agit de construire ce qui est couramment appelé une théorie d'action ou encore un schéma de causalité (Toulemonde, 1997). Quand il est établi, un tel schéma constitue un référentiel, socle du dispositif évaluatif, à partir duquel l'évaluateur peut proposer des indicateurs ou porter un jugement de valeur sur la base de critères d'évaluation (Berriet-Sollic, 2007). Pourtant, si ces modèles reposant sur la théorie du programme tendent à se développer (Hansen, 2005), force est de constater qu'ils restent encore relativement rares dans les pratiques des évaluateurs.

Dans le cas du conseil, par exemple, il existe peu de travaux d'évaluation des politiques de conseil proposant de telles théories d'action des politiques (Evenson, 1989). La plupart des évaluations se limitent à des recensements des moyens mis en œuvre, dans l'optique de se rendre compte des investissements réalisés auprès de bailleurs de fonds (Richardson et Moore, 2002). Ceci est peut-être lié aux difficultés de construction de ces schémas de causalités. Elles sont bien identifiées dans le cas des services de conseil (Gadrey, 1996) : difficultés de mesure du produit et donc de la productivité des services, complexité de l'identification de leurs effets directs (transformations des pratiques et des performances des exploitations agricoles) ou indirects (apprentissage), etc

Dans le cas des mesures agri-environnementales, deux types d'évaluation existent : une évaluation du nombre d'agriculteurs ayant opté, sur une base volontaire, pour une mesure et une évaluation des effets des mesures sur la biodiversité. Mais il n'existe pas d'objectifs clairement définis permettant de mesurer le succès d'une mesure. Les schémas de causalités auquel il est fait appel sont sommaires et sont souvent limités à des formules telles que « moins d'intrants (fertilisation, pesticide) = plus de biodiversité ». Les mesures étant à l'échelle parcellaire, le contexte paysager ou régional, qui contrôle la possibilité pour de nouvelles espèces de coloniser les parcelles contractualisées, n'est pas pris en compte dans la mise en place. Les travaux évaluant les effets sur la biodiversité tentent de le faire (Kleijn *et al*, 2006).

Quand les évaluateurs s'appuient sur des schémas de causalité, trois constats conduisent à souligner les limites aux pratiques actuelles (Toulemonde, 1997). Tout d'abord, dans la plupart des cas, l'évaluation vise à comprendre les effets, trop rarement à les vérifier et à les confronter aux faits. En deuxième lieu, l'évaluation révèle les facteurs qui influencent l'efficacité sans démontrer les mécanismes qui démontrent comment sont atteints les objectifs. Enfin, l'évaluation rassemble dans

la plupart des cas différents points de vue au sujet d'une théorie d'action et essaie d'arriver à une vision acceptable de la politique ou du programme étudié, à un schéma de causalité qui reflète le mieux possible les avis des parties prenantes (*evaluation-theorisation, theory building evaluation*). Très rarement, l'évaluation dessine un schéma de causalité et le confronte aux faits (*theory driven evaluation*).

1.2 Trois méthodes pour analyser les relations de cause à effet

Parmi les méthodes d'évaluation, trois voies existent pour appréhender les schémas de causalité. La première est l'analyse de politique qui focalise l'attention des évaluateurs sur les réseaux d'acteurs et les relations des différentes parties prenantes, sur la hiérarchisation des objectifs du programme étudié, enfin sur les modalités retenues pour la mise en oeuvre des actions. L'analyse de politique (ou du programme qui la met en oeuvre) conduit, à partir du dépouillement des textes fondateurs et de l'interview des auteurs du programme, à l'établissement d'un référentiel sous la forme de graphe ou diagramme d'objectifs ou encore de diagramme logique d'impacts (Berriet-Sollic, 2007). La deuxième voie d'investigation scientifique est l'analyse micro-économique de ces relations de cause à effet. Il s'agit dans ce cas, dans la lignée des travaux micro-économiques d'Heckman (Heckman *et al.*, 1999), de tester, par les outils de l'économétrie, la robustesse des liens causaux entre variables (Bondonio, 2005). Enfin, les recherches expérimentales se développent avec la mise en place de dispositifs de suivi/évaluation, dès la mise en place d'interventions publiques, qui se traduisent le plus souvent par l'identification de groupes de contrôles et la distinction d'individus avec/sans les soutiens publics. En France, la mise en place du revenu de solidarité active (RSA) a bénéficié, à titre expérimental, d'un tel dispositif.

Dans le cas du conseil agricole, on peut retrouver de façon quasi similaire ces trois méthodes d'évaluation.

- i) Certains travaux s'appuient sur des théories systémiques pour concevoir des méthodes d'évaluation des politiques de conseil agricole reposant sur la participation des différents acteurs impliqués. Ceci s'incarne notamment dans l'utilisation de la *Soft System Methodology* (Checkland, 1981). Par exemple, cette méthode a été utilisée pour construire une évaluation de la politique de formation de conseillers agricoles d'une université aux Philippines (Navarro *et al.*, 2008). Une telle méthode n'inclut pas d'analyse des effets et de l'impact sur le terrain de la politique étudiée.
- ii) D'autres travaux d'évaluation se basent quant à eux sur les résultats de modélisations micro-économiques. Par exemple, l'évaluation des politiques publiques d'investissements dans le conseil a souvent conduit à des réformes de privatisation. Ces évaluations ont pour hypothèse qu'un désengagement total de l'Etat accroîtrait l'efficacité de l'offre de services, qui se trouverait alors tirée par la demande vers des situations d'équilibre (Carney, 1998). Elles s'appuient sur des travaux dont les schémas de causalités s'inscrivent dans la théorie économique standard, appliquée aux services (Dinar et Keynan, 2001). Ces travaux ne sont quasiment pas étayés par des confrontations aux réalités du terrain (Dinar, 1996). Cette distance aux faits empiriques est si importante que certains auteurs discutent de la possibilité même d'appliquer de tels modèles au cas des services de conseil (Hanson et Just, 2001), car leur axiomatique ne permet pas de rendre compte de certaines spécificités fondamentales des services (asymétrie d'information, caractère non exclusif et non rival des connaissances, etc.).
- iii) Il existe enfin quelques travaux fondés sur des méthodes expérimentales quantitatives. Il s'agit d'évaluations reposant sur des outils statistiques pour isoler les effets de politiques de soutien à l'offre ou la demande en conseil technique. La rareté des analyses empiriques quantitatives sur le conseil agricole est en partie due à la difficulté de mesurer son impact sur la performance des exploitations (Evenson, 1997, Esposito, 2002). Récemment, Davis et N'Konya (2008) ont proposé d'utiliser des *méthodes quasi-expérimentales (MQE)* pour

comparer dans le temps les gains de productivité d'exploitations agricoles bénéficiant ou non de services publics de soutien techniques (par exemple, participation à des *farmers field schools*), et d'isoler ainsi l'effet de politiques publiques d'investissement dans le conseil agricole sur la réduction de la pauvreté dans des zones rurales d'Afrique de l'Est. Cependant, de tels travaux nécessitent des protocoles très lourds de collecte de données et demeurent très rares.

Les mesures des effets des mesures agri-environnementales sur la biodiversité reposent sur un échantillonnage aléatoire comparant des parcelles soumises à la mesure à des parcelles non soumises. Comme il n'existe pas d'évaluation au départ de la mesure, on ne peut mesurer d'évolution. Les biais liés au fait que les parcelles soumises à la mesure ont généralement des caractéristiques spécifiques (moindre usage) font qu'il est difficile d'expliquer les différences constatées avec les parcelles non contractualisées. Comme la contractualisation porte sur les pratiques d'usages, il se peut que celles stipulées dans le contrat soient identiques à celles mise en œuvre antérieurement.

1.3 Une quasi inexistence de la confrontation aux faits dans les modèles d'évaluation

Quelle que soit la voie retenue pour mener l'évaluation, la confrontation aux faits occupe une place très relative depuis l'inexistence (raisonnement uniquement à partir de modélisation théorique) à l'empirisme exclusif (construction du raisonnement à partir des faits). Dans la plupart des situations, l'évaluateur est confronté à deux principaux écueils : le manque de conceptualisation théorique pour appréhender les faits nouveaux et les lacunes statistiques pour tester les hypothèses (Laurent, 2007).

En matière d'agri-environnement, le passage de mesures restreintes à des obligations de moyens à des mesures avec obligation de résultats () va obliger d'une part les concepteurs de la mesures à avoir des modèles prédictifs des changements et les évaluateurs à faire des observations de terrain (). De fait il y aura confrontation entre les modèles et les faits. Le débat entre déterminisme et stochasticité dans les dynamiques écologiques sera sans doute renouvelé tant le hasard peut jouer un rôle important ().

Se pose dès lors la question de la validité empirique de ces méthodes d'évaluation : comment les liens de causalité sont-ils testés et validés? Comment concilier la construction de modèles théoriques et la connaissance des réalités observées, dans la mouvance des travaux issus du réalisme critique (Lawson, 2003) ?

Dans ce contexte et face à ces défis méthodologiques, des débats associés aux « d'evidence based decision » (EBD) se développent depuis une vingtaine d'années comme une alternative pour développer, dans les pratiques d'évaluation, un contenu empirique aux connaissances produites et soumises aux décideurs publics.

2. Apports des débats sur les approches evidence based pour l'évaluation des politiques publiques

Cette deuxième partie analyse l'intérêt des débats associés aux approches d'EBD pour l'évaluation des politiques publiques. Encore très rarement mobilisées en France (Laurent, 2006), ces pratiques sont aujourd'hui largement répandues au Royaume-Uni, en particulier depuis 1997 et la modernisation des politiques publiques. Ainsi, dans ce pays, la recherche des « evidences »³

³ Le mot « evidence », difficile à traduire en français, renvoie à la notion d'évidence empirique, de preuve ou encore de données probantes (Laurent, 2007) développées initialement dans le champ de la médecine (Guyatt *et al.* 1992), ces démarches visent à promouvoir l'usage des connaissances scientifiques disponibles par les décideurs publics et ce, de la façon la plus judicieuse et pertinente possible, en prenant en considération le domaine de validité de ces connaissances et la portée relative des résultats.

apparaît comme un élément essentiel dans l'aide à la décision publique, à tel point que le gouvernement préconise « *the better use of evidence and research in policy making* » (extrait du UK-White Paper, 1999, cité par Cartwright, 2007).

C'est pourquoi, les principaux instruments utilisés et débattus dans les démarches EBD sont rappelées ici en mettant l'accent d'une part sur l'intérêt des états de l'art systématiques de la littérature scientifique et d'autre part sur les techniques quantitatives de prise en compte du biais de sélection.

L'intérêt des approches EBD est double dans la mesure où elles permettent à la fois une meilleure connaissance des processus de prise de décision grâce aux enquêtes auprès des praticiens et décideurs, mais également une évaluation reposant sur une validation empirique d'hypothèses portant sur les relations de cause à effet lors de l'application d'un traitement ou d'une intervention publique. Ce travail doit conduire, dans la pratique, à l'utilisation la plus judicieuse possible des connaissances scientifiques effectivement disponibles.

Cette analyse scientifique repose sur plusieurs méthodes pour caractériser une évidence au regard de sa capacité à tester une hypothèse. Ces méthodes peuvent être hiérarchisées selon la qualité des preuves qu'elles produisent (Cartwright, 2007). La première méthode est la réalisation de méta-analyses grâce à des revues systématiques de la littérature (Sutherland, *et al.*, 2004). La seconde est l'utilisation de techniques statistiques comme les essais contrôlés randomisés réalisés sur des groupes de contrôle ou de cohortes. Ces techniques statistiques sont mobilisées dans le domaine de l'évaluation des politiques publiques pour tester la robustesse des relations de cause à effet (Schmitt *et al.*, 2003; Zahm *et al.* 2008). Enfin, les études de cas et les avis d'experts permettent tout au plus de tester des relations de cause à effet mais en contribuant peu à édifier ces relations au statut d'évidence. Nous reprenons donc ici la présentation des deux premiers types de démarches permettant d'ériger une relation de cause à effet au statut d'évidence.

2.1 La méthode des états de l'art réalisés selon des critères explicites (EARC)

La méthode EARC vise à faire le bilan le plus exhaustif possible des études et des preuves existantes sur un sujet donné. Les EARC reposent sur une méthodologie particulière qui se décompose en plusieurs étapes désormais bien définies dans la littérature (Pullin *et al.*, 2004; Davies, Nutley, 2001, Peersman *et al.*, 2001). Parmi ces étapes, une des phases essentielles consistent en un tri des références selon leur degré de validité.

Ainsi, les études sur le sujet étudié ayant été sélectionnées, il s'agit de faire l'analyse de ces références sur la base de critères dit « critères de qualité ». Ces critères sont contingents de la question posée. Par exemple pour une question relative à l'efficacité d'une mesure de protection de la biodiversité Pullin & Knight (2003) proposent une classification en plusieurs niveaux de qualité qui est déterminée par le niveau de validité empirique des preuves produites :

I : Preuves fortes obtenues à partir d'au moins un essai aléatoire convenablement conçu

II-1 Preuves obtenues grâce à des essais (convenablement conçus) non aléatoires

II-2 Preuves obtenues grâce à la comparaison de différences entre sites (lieux) – avec ou sans une espèce ou communauté dont la présence est souhaitée pour l'analyse

II-3 Preuves obtenues à partir d'une série multiple ou d'un résultat désastreux dans des expériences non contrôlées

III Opinions d'autorités légitimes, reposant sur des preuves qualitatives obtenues sur le terrain ou sur les rapports d'experts

IV Preuves inadéquates tenant aux problèmes de méthode (taille de l'échantillon), de compréhension, ou de conflits entre les résultats.

Dans cette phase d'analyse les études et les résultats sont ainsi écartés:

- sur la base de l'incohérence de leur axiomatique (incohérence entre les hypothèses, incohérence entre les corpus théoriques mis en avant et les études empiriques qui en seraient l'application) ;
- ou, pour un certain nombre d'études considérées comme pertinentes, sur la base des problèmes de méthode importants qu'elles posent, études mal faites ou méthodes non explicites, etc.

2.2 Les méthodes quantitatives de mesure des impacts

Lorsque l'aide à la décision publique porte sur des programmes ou des politiques qui ont déjà été mis en oeuvre (cas de situation d'évaluation *ex post*), le test d'une hypothèse et l'obtention d'évidence sur une relation de cause à effet peuvent également émaner du recours à des techniques quantitatives de mesures des impacts de la politique étudiée.

On définit l'impact d'une politique comme la différence entre la situation en présence de la politique et la situation qui aurait prévalu en son absence. Cette dernière situation, hypothétique et inobservée, est appelée situation contrefactuelle. L'évaluateur, suivant les demandes des décideurs, ou adoptant une ou plusieurs visions des objectifs de la politique, sélectionne une ou des variables objectives que la politique devait affecter. C'est la comparaison entre le niveau de ces variables en présence et en l'absence de la politique qui constitue l'impact de la politique (Zahm *et al.*, 2008).

L'objectif est d'estimer l'impact moyen du traitement, ou plus largement de déterminer cette situation contrefactuelle (Foray et Cowan 2002). Se pose alors un problème d'identification, qui se traduit au niveau individuel par une question de données manquantes et plus globalement à un problème qualifié de biais de sélection. Pour surmonter cette écueil méthodologique, deux raisonnements sont possibles : la comparaisons avec/sans l'intervention publique ou le traitement ou la comparaison avant/après.

Les principales méthodologies utilisables pourront être (Schmitt *et al.*, 2003) :

- des estimations sur données de panel (cohortes) si on souhaite mesurer de performances sur données agrégées ;
- des méthodes de *matching* (ou *appariement*) avec et/ou sans remise ainsi que des estimations en doubles différences éventuellement combinées qui reposent sur le principe des groupes de contrôle ;
- enfin, si les données disponibles se « limitent » à des caractéristiques initiales et à des performances finales des individus et/ou zones ciblées (bénéficiaires et non bénéficiaires), on devra se « contenter » d'utiliser la méthode du *matching* avec ou sans remise.

Dans tous les cas, la mesure des effets propres passe par une bonne connaissance du programme et par une identification des effets attendus de chaque mesure. Cette étape est indispensable à plusieurs titres. D'une part, elle permet de choisir les indicateurs de performance le plus pertinents à retenir pour identifier les effets propres du programme. D'autre part, c'est dans la confrontation des processus d'évolution des situations socio-économiques potentielles ou observées, identifiées dans la phase d'analyse de politique avec les résultats économétriques de l'analyse des effets propres que sont à rechercher les effets réels du programme. En effet, la grande difficulté à la caractérisation des relations de causalité en particulier dans le cas de programmes de développement aux objectifs larges comme par exemple les politiques territorialisées, ne permettent pas aux outils statistiques d'apporter de réponses univoques et certaines quant aux effets propres de la politique évaluée. Par

contre, l'explicitation des résultats par les effets potentiels ou observés à travers la théorie d'action permet de conclure de façon plus robuste

Par conséquent, les résultats statistiques obtenus permettent un test de robustesse des relations de cause à effet testées. Mais, même si les résultats obtenus sont probants statistiquement, ils ne peuvent rien prouver sur le sens même de la relation (Schmitt *et al.*, 2003). C'est pourquoi, il est essentiel de relativiser le recours à de telles démarches, ou tout au moins de les coupler avec des analyses qualitatives complémentaires (Berriet-Sollicie, 2007).

Au final, l'utilisation des méthodes d'EBD permet l'identification de mécanismes socio-économiques récurrents, testés statistiquement. En ce sens, elles constituent un outil essentiel pour consolider, dès le démarrage d'un programme d'intervention publique, les théories d'action des différentes mesures de ce programme, et les leviers d'actions qui sous-tendent chacune de ces mesures grâce à la mobilisation de données quantitatives et qualitatives. Malgré tout des lacunes méthodologiques subsistent qui conduisent à dépasser un certain « effet de mode » et à relativiser le recours à ces méthodes.

3. Mise en œuvre et difficultés d'application des approches EBP

L'utilisation des méthodes d'évidence based pour l'aide à la décision publique nécessite de consacrer des financements spécifiques pour soutenir et développer les liens entre instituts de recherche, universités et instances gouvernementales. Une telle structure existe depuis 2000 au Royaume-Uni : the Evidence for Policy and Practice Information and Co-ordination Center (EPPI-Center) (Olivier *et al.*, 2005). De telles pratiques n'existent pas en France. Une des raisons est sans doute une méconnaissance des démarches evidence-based (Laurent *et al.*, 2009).

3.1 la nécessité de structures de transfert

Les états de l'art réalisés selon des critères systématiques sont des outils importants dans les méthodes d'EBD dans la mesure où ils permettent de confronter les conclusions de nombreuses études et d'éviter de retenir des propositions ayant conduit à des résultats peu probants. Largement répandues ces vingt dernières années, ces techniques sont pourtant aujourd'hui sujettes à des critiques en raison du relatif manque de rigueur dans l'utilisation de ces méthodes et du manque de prise en compte du contexte dans lequel ont été réalisées les études.

Face à ces difficultés, l'établissement d'organisations spécifiques pour mettre en œuvre des méthodes EBD peut s'avérer utile voire nécessaire de différents points de vue. D'une part, les méthodes quantitatives d'évaluation de l'impact des politiques créent des besoins en recueil de données qui rendent nécessaire un travail conséquent et centralisé. De même, l'établissement d'états de l'art selon des critères explicites (EARC) nécessite des collaborations et des échanges directs entre scientifiques aptes à évaluer des travaux de recherche inscrits dans différentes disciplines ou différents champs théoriques. D'autre part, de telles organisations peuvent favoriser des liens et un accès plus direct des politiques aux résultats de démarches scientifiques s'inscrivant dans la démarche EBP.

Par exemple, *the Social Science Research Unit (SSRU)* de l'Université de Londres, en lien avec le centre de coordination et d'information sur les pratiques et les politiques reposant sur les évidences (EPPI-Center), propose de nouveaux modèles pour l'utilisation des revues systématiques de la littérature à partir de démarches couplant données quantitatives et qualitatives (Oliver *et al.*, 2005). L'originalité est d'aborder d'emblée les états de l'art en conjuguant deux objectifs (i) obtenir des renseignements, pour les décideurs publics, sur la mise en œuvre des interventions publiques (ii) mettre en place un dispositif d'évaluation de ces interventions. Dans le protocole pour réaliser l'état

de l'art, les auteurs distinguent donc les études relatives aux points de vue exprimés par les parties prenantes et d'autres relatives aux impacts de la politique étudiée. La valeur ajoutée de cette démarche des chercheurs de l'EPPI-Center est plus globalement de conjuguer cet intérêt méthodologique avec une dimension théorique (en l'occurrence il s'agissait d'avancées théorique en matière de santé publique relatif aux jeunes).

Dans le cas de l'agriculture et des services de conseil, des organisations spécifiques ont été mises en place pour favoriser l'accès des décideurs publics aux connaissances scientifiques pour la conception et l'évaluation de politiques agricoles. C'est par exemple le cas du *Consultative Group on International Agricultural Research (CGIAR) International Food Policy Research Institute (IFPRI)*⁴, qui propose des recherches pour orienter les politiques de réduction de la faim dans le monde. On peut retrouver au sein de cet institut des débats sur la possibilité de mieux connecter recherche et évaluation des politiques agricoles à travers les méthodes EBD. Dans le cas du conseil agricole, cet institut propose de combiner des approches institutionnelles et théoriques des schémas de causalité de réformes du conseil agricole déjà mises en œuvre et de faire la synthèse de leurs efficacités dans différents contextes ("*from best practices to best fit*", cf. Birner et al. 2006) d'une part ; et une évaluation quantitative ex-post des effets de ces réformes (Davis et N'Konya 2008) d'autre part. Cependant, cette démarche ne s'appuie pour l'instant pas explicitement sur les méthodes EBD, et sur des méthodologies validées d'états de l'art systématiques.

Dans le domaine de la biologie de la conservation, auquel sont rattachées les mesures agri-environnementales, un groupe travaille activement pour produire des EARC (Pullin, op. cit.). Les limites de la démarche sont importantes, du fait du manque de données (problème d'échelles spatiales et temporelles, en particulier). Ceci conduit certains à rejeter ces méthodes pour l'instant (Ferraro et Pattanayak, 2006) ou à souligner les risques d'impasse dus à l'étroitesse des situations analysées (Ludwig et al, 2001).

3.2 Deux obstacles majeurs : manque de théorisation et de données

Si l'engouement pour les pratiques « evidence-based » est notoire, ces méthodes, relativement récentes tant dans les pratiques scientifiques que dans les prises de décision politiques, souffrent de plusieurs lacunes. Nous en identifions deux principales. La première est l'absence de concept stabilisé sur ce qu'est précisément une « evidence » (Cartwright, 2007; Kirsch et Laurent, 2007). La seconde est le manque de précision sur l'utilisation des différents types d'évidence et sur leur domaine de validité pour tester une hypothèse émise par la communauté scientifique. Cette dernière remarque renvoie indirectement à la qualité des données disponibles.

En effet, les méthodes mobilisées en évaluation de politique pour mesurer les impacts des politiques sont intimement liées à la disponibilité de données concernant non seulement le groupe des bénéficiaires mais aussi un groupe de comparaison. Ces données doivent comprendre des caractéristiques descriptives des individus notamment celles qui ont initialement déterminé la sélection au programme, des indicateurs de performances dont le choix doit être conditionné par les objectifs attendus du programme ainsi que ses effets potentiels et, dans la mesure du possible un ensemble de variables explicatives des indicateurs de performance retenus. De façon évidente, plus l'information disponible est abondante en coupes temporelles et plus la qualité des estimations sera améliorée en termes de correction de biais potentiels et de précision des estimateurs. *A minima*, trois dates de recueil des données avant la mise en place du programme et trois dates postérieures sont requises pour obtenir un niveau de fiabilité acceptable. Disposer d'un suivi temporel régulier des

⁴ L'objectif de l'IFPRI est "To Provide Policy Solutions That Reduce Poverty and End Hunger and Malnutrition" (www.ifpri.org)

différentes variables pour l'ensemble des individus et qui débute au moins trois ans avant la mise en place du programme correspond à la situation « idéale ».

Dans le cas des mesures agri-environnementales, les informations sur les relations de cause à effet entre pratiques agricoles et espèces cible sont manquantes et les échelles spatiales d'intervention sont généralement inadéquates. En ce sens, le recours aux approches EBD apparaît comme un des outils pour consolider les pratiques évaluatives et affiner les préconisations qui en découlent. Ceci suppose la mise en place d'expérience d'évaluation prenant en compte l'ensemble des facteurs pouvant expliquer les variations de biodiversité, depuis les fluctuations naturelles des populations animales et végétales jusqu'à l'histoire des pratiques dans les parcelles et l'histoire du paysage qui peuvent expliquer les facilités ou impossibilités d'arrivée de nouvelles espèces dans les parcelles contractualisées. Dans un premier temps, il est nécessaire de 1) définir l'échelle spatiale à laquelle seront évalués les résultats et 2) construire un modèle reliant explicitement les pratiques agricoles à la biodiversité. Le premier point est crucial ; en conclusion de leur EARC sur "*the effectiveness of land-based schemes (incl. agri-environment) at conserving farmland bird densities within the U.K.*", Roberts et Pullin (2007) écrivent: "Whether species are simply redistributing between the available resources by aggregating in fields under agri-environment management and deserting conventional fields (no change in population trend), or experiencing increased breeding success or overwinter survival rate (positive population trend) giving evidence of species recovery, is unknown ». On voit là qu'il y a à la fois un problème de définition des politiques et un problème de limites aux EARC.

Dans leur synthèse sur les relations entre agriculture et biodiversité, Le Roux *et al* (2008) mettent en évidence la synergie entre pratiques à la parcelles et structures du paysage dans les dynamiques de la biodiversité. Ils soulignent aussi le manque de recherches traitant explicitement de l'intégration de la biodiversité dans les exploitations agricoles.

Discussion: quelles perspectives de développement des méthodes evidence-based dans les pratiques évaluatives ?

Par conséquent, le développement des « evidence-based policies » est intéressant à divers points de vue pour l'évaluation des politiques publiques. D'un point de vue politique, il contraint les décideurs publics à davantage intégrer les résultats de la recherche et incitent à réaliser des évaluations *ex ante* qui s'appuient sur des théories des programmes étudiés corroborés aux faits. D'un point de vue scientifique, il encourage les chercheurs à prêter davantage attention aux mécanismes et impacts des décisions publiques et à développer les méthodes de mesures d'impacts raisonnées en fonction d'une hiérarchie des objectifs initialement visés. De tels développements mettent en lumière la nécessité de mieux préciser les cadres conceptuels et de poursuivre la construction de nécessaires bases de données pour pallier les données manquantes.

Alors qu'on entend désormais parler d'evidence-based transport policy (Cartwright, 2007), il n'est pas encore question d'evidence-based agricultural policy. Et pourtant, dans un contexte européen de réforme en profondeur de la politique agricole et de développement rural à l'horizon 2013, il serait judicieux de davantage mobiliser de telles analyses qui mettent à la fois l'accent sur la synthèse des connaissances existantes et plus largement sur l'intérêt de mobiliser des productions scientifiques reposant sur une connaissance des réalités observées. Ainsi, au-delà de la recherche des « evidences », les approches EBD pointent l'importance de mobiliser les approches positives de l'aide à la décision publique, dans un contexte où prédomine le recours à des approches normatives et l'utilisation de simulations au détriment des données d'observation.

Un des attendus du développement des approches evidence-based serait de capitaliser les résultats de nombreuses évaluations menées jusqu'à ce jour par la Commission Européenne. Les pratiques

évaluatives sont en effet largement répandues en Europe et en particulier pour les interventions communautaires. L'évaluation a lieu à différents stades de la mise en place de la politique (*ex ante*, intermédiaire, finale, *ex post*) : les Directions de la Commission sont donc en possession de nombreux résultats dans les différents Etats-membres. Des meta-analyses sont certes réalisées mais les synthèses systématiques grâce à des techniques éprouvées sont rares.

Une telle perspective appelle la mobilisation de compétences spécifiques et la création d'une agence spécialisée dédiée à la mise en oeuvre et en débat des approches d'évaluation evidence-based. Le Royaume-Uni a d'ores et déjà misé sur de telles structures dans le cadre de sa réforme gouvernementale amorcée en 1999. Cette expérience permet de mettre en exergue le nécessaire développement conjoint des recherches dans le domaine. Consolidation des équipes de recherche sur ces questions d'Evidence-based et institutionnalisation de la démarche au travers de structures de transfert spécialisées dans ces méthodes s'avèrent ainsi deux facteurs dont la conjugaison est une condition *sine qua non* au développement raisonné d'évidence-based policies.

Références bibliographiques

ARNON, I. 1989. *Agricultural research and technology transfer*. Londres: Elsevier Applied Science.

BERRIET-SOLLIEC, M. 2007. *L'analyse de politique : une démarche essentielle pour l'évaluation. L'exemple des politiques territoriales*, Ecole - chercheurs INRA « Evaluation des politiques publiques » -18-20 Juin 2007.

BIRNER R., DAVIS K., PENDER J., NKONYA, E., ANANDAJAYASEKERAM P., EKBOIR J., MBABU, A. SPIELMAN D., HORNA D., BENIN S.,; KISAMBA-MUGERWA W. 2006. *From "best practice" to "best fit": a framework for designing and analyzing pluralistic agricultural advisory services*. 2006. IFPRI Research Brief

BONDONIO D., 2002. Evaluating Decentralized Policies: A Method to Compare the Performance of Economic Development Program across Different Regions or States. *Evaluation*, 8(1): 101-124

BOWEN, W. 2002. An analytical Review of Environmental Justice Research: What Do We really know ? *Environmental Management*, Vol 29, N°1, pp 3-15.

CARNEY, D. 1998. *Changing public and private roles in agricultural services provision*. Londres: Overseas Development Institute.

CARNEY, D. 1995 *The changing public role in services to agriculture: a framework for analysis*. *Food Policy*, 20, pp. 521-528.

CARTWRIGHT N. with help from Goldfinch A and Howick J., 2007. *Evidence-based policy: Where is our theory of evidence?* Working paper, LSE and UCSD.

CHECKLAND, P.B., 1981. *System thinking, system practice*. New-York: John Wiley.

DAVIES H.T.O., NUTLEY S.M. 2001. Evidence-based policy and practice: moving from rhetoric to reality. *Proceedings of the Third international Interdisciplinary Evidence-based Policies and Indicator Systems Conference*, 86-95.

DINAR, A. 1996. *Extension commercialization: how much to charge for extension services*. *American Journal of agricultural economics*, 78, pp. 1-12.

DINAR, A., KARAGIANNIS, G., TZOUVELEKAS, V. 2007. Evaluating the impact of agricultural extension on farms' performance in Crete: a nonneutral stochastic frontier approach. *Agricultural Economics*, 36:2, pp. 135-46.

DINAR, A. et KEYNAN, G. 2001. Economics of paid extension: lessons from experiences in Nicaragua.

American Journal of agricultural economics, 83 (3), pp. 768-776.

ESPOSTI, R. 2002. Public agricultural R&D design and technological spill-ins: A dynamic model. *Research policy*, 31:5, pp. 693-717.

EVENSON, R. E., 1998. *Economic issues in agricultural extension policy*. New Haven, CT: Yale University, Economic Growth Center.

FERRARO, P. J. et S. K. PATTANAYAK .2006. Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments *PLoS Biology* 4.(4): e105.
<http://dx.doi.org/10.1371%2Fjournal.pbio.0040105>

FORAY, D., COWAN, R., 2002. Evolutionary economics and the counterfactual threat: on the nature and role of counterfactual history as an empirical tool in economics, *Journal of Evolutionary Economics*, Springer, 12 (5), 539-562.

GADREY, J. 1996. *Services: la productivité en questions*. Paris: Desclée de Brouwer, 1996.

GUYATT ET AL. (EVIDENCE-BASED MEDICINE WORKING GROUP) 1992. Evidence-Based Medicine. A new approach to teaching the practice of Medicine. *JAMA*. vol 268, N°17, pp.2420-5

HANSEN, H.F. 2005. Choosing Evaluating Models : a discussion on Evaluation Design. *Evaluation*, Vol 11 (4) : 447-462.

HANSON J. C., JUST R. E. 2001. *The potential for transition to paid extension: some guiding economic principles*. *American Journal of agricultural economics*, 2001, 83 (3), pp. 777-784.

HECKMAN J., LALONDE R. ET SMITH J., 1999. The Economics and Econometrics of Active Labor Market Programs. In O.Ashenfelter and D.Card (eds), *Handbook of Labor Economics*, vol.III, North Holland, pp.1865-2097.

KIRSCH M., LAURENT C. 2007. *Scientificity, policy making and empirical validity of knowledge : new issues ?* Communication pour le colloque mondial de philosophie des sciences. Pekin. Août 2007. 10 p.

KLEIN, D.; SUTHERLAND, W.J. 2003. How effective are European agri-environment schemes in conserving and promoting biodiversity ? *Journal of Applied Ecology*, 40, pp 947-969.

KLEIJN, D., R. A. BAQUERO, Y. CLOUGH, M. DÍAZ, J. ESTEBAN, F. FERNÁNDEZ, D. GABRIEL, F. HERZOG, A. HOLZSCHUH, R. JÖHL, E. KNOP, A. KRUESS, E. J. P. MARSHALL, I. STEFFAN-DEWENTER, T. TSCHARNTKE, J. VERHULST, T. M. WEST et J. L. YELA .2006.. Mixed biodiversity benefits of agri-environment schemes in five European countries *Ecology Letters* 9 (3), pp 243-254.

LABARTHE, P. 2006. *La privatisation du conseil agricole en question. Evolutions institutionnelles et performances des services de conseil dans trois pays européens (Allemagne, France, Pays-Bas)*. Thèse de doctorat en Sciences Economiques, Université Paris XII, 320 p. (Sous la dir. de C. Laurent)

LABARTHE P. 2009, Extension services and multifunctional agriculture, Lessons learnt from the French and Dutch contexts and approaches, *Journal of environmental management*, 90(2), pp. 193-202.

LAURENT C., BAUDRY J., BERRIET SOLLIEC M., KIRSCH M., PERRAUD D., TINEL B., TROUVE A., N.ALLSOPP, BONNAFOUS P., BUREL F., CARNEIRO M.-J ., GIRAUD, C., LABARTHE P., MATOSE F., RICOCH A. 2009. Pourquoi s'intéresser à la notion d'Evidence-based policy ? *Revue Tiers-monde*, n°200, pp. 853-873

LAURENT L. (DIR.) 2006. *Validité des connaissances scientifiques et intervention publique : le cas de l'agriculture dans le développement durable*. Projet ANR « EBP-BIOSOC. 2007-2010» 17 p. + annexes.

LAURENT L. 2006. *Inégalités sociales, pauvreté et développement durable. Nouveaux enjeux autour des modèles de connaissances qui sous tendent l'action politique*. Communication au colloquio international sobre Desenvolvimento Territorial Sustentavel. Florianopolis. Santa Catarina. Brasil. 22-25 de agosto de 2007.

LAURENT C., CERF M., LABARTHE P. 2006. Agricultural extension and market regulation : learning form a comparison of six EU countries. *Journal of Agricultural Education and Extension*. 2006, 12 (1) : 5-16

LAWSON, T. (2003) *Reorienting Economics*, Routledge, London.

LE ROUX, X., R. BARBAULT, J. BAUDRY, F. BUREL, I. DOUSSAN, E. GARNIER, F. HERZOG, S. LAVOREL, R. LIFRAN, J. ROGER-ESTRADE, J. P. SARTHOU et TROMMETER, M., (Eds.) .2008. *Agriculture et biodiversité. Valoriser les synergies*. Expertise scientifique collective, synthèse du rapport. Paris, INRA .France 112 p.

LUDWIG, D., M. MANGEL et B. HADDAD .2001. Ecology, conservation, and public policy *Annual Review of Ecology and Systematics* **32** (1), pp 481-517.

NAVARRO M., ROCHS F., PRUSSIA S.E., KANLAYANARAT S., MCGLASSON B. 2008. Using soft-systems methodology to plan advanced academic programs in post-harvest thechnology in thailand. 24TH annula conference of the association for international agricultural and extension education. Guacimo (costa-rica), 10-15 mars 2008.

OLIVER S. ; HARDEN A. ; REES R. ; SHEPERD J. ; BRUNTON G. ; GARCIA J. ; OAKLEY A. 2005. An emerging framework for including different types of evidence in systematic reviews for public policy., *Evaluation*, Vol 11 (4), pp 428-446.

PEERSMAN, G. ; OLIVER, S. ; OAKLEY, A. 2001. Systematic reviews of effectiveness, in S.Oliver and G.Peersman (eds) *Using research for effective Health Promotion*, pp 96-108, Buckingham : Open University Press.

PIERSON, P. 1993. When effect becomes cause. Political feedback and political change, *World Politics* 45 : 595-628.

PULLIN A.S., KNIGHT T.M., STONE D.A., CHARMAN K. 2004. Do conservation managers use scientific evidence to support their decision-making? *Biological conservation* 119, 245-252.

RICHARDSON J. G., MOORE C. L. 2002. *Determining extension program economic value*. North Carolina State University, Working paper, 2002, AEE-01-02.

RICROCH et BONNAFOUS 2008. *Scientific knowledge used in three French agro-environment schemes related to European agriculture and biodiversity policies*. Communication au séminaire EBP-BIOSOC, Rio, 28-29 août 2008.

ROBERTS, P. D. et A. S. PULLIN .2007. The Effectiveness of Land-Based Schemes .incl. Agri-environment. at Conserving Farmland Bird Densities within the U.K. *Systematic Review No. 11*. Collaboration for Environmental Evidence <http://www.environmentalevidence.org/CEE%20Reviews.htm>

RIVERA, W. R. 2000. Confronting global market: public sector agricultural extension reconsidered. *Journal of extension systems*, 16, pp. 33-54.

RIVERA, W., ZIJP, W. 2002. *Contracting for agricultural extension. International case studies and emerging practices*. Cambridge (USA): CABI Publishing.

SCHMITT, B.; LÉPICIER, D.; BERRIET-SOLLIEC, M., 2003 - *Economic impact evaluation of the european program for rural development in Burgundy : allowance for selection bias*. Dijon : INRA-ENESAD UMR CESAER, 24 p. Working paper 6 .

TOULEMONDE, J. 1997. Faut-il libérer l'évaluation de ses liens de causalité ? *Annales des mines*, Mars 1997.

ZAHM, F., CHABE-FERRET, S. GASSIAT, A., SUBERVIE, J. 2008. *Proposition d'un cadre méthodologique pour évaluer les effets propres des mesures agro-environnementales européennes – application au dispositif français*, Communication au colloque de la Société Française d'Evaluation, Strasbourg, 3-5 juillet.