

Réflexion sur le plan d'échantillonnage appliquée à la quantification des paysages

Exemples d'évaluation des dégâts tempétueux observés dans deux forêts
françaises (Conches-Breteuil, Eure, et Fontainebleau, Seine-et-Marne)

Reflexion on the sampling design applied to the quantification of the landscapes

*Examples of evaluation of the tempestuous damage observed in two French
forests (Conches-Breteuil, Eure, et Fontainebleau, Seine-et-Marne)*

Vincent GODARD – vgodard@univ-paris8.fr

Centre de Biogéographie Écologie (CNRS FRE 2545)
Université de Paris 8 – Maison des Sciences de l'Homme Paris Nord
4, rue de la Croix Faron, 93210 La Plaine Saint-Denis

RÉSUMÉ. La qualification et la quantification des paysages reposent sur des inventaires de terrain. Il n'y a pas, dans l'absolu, de bonnes ou de mauvaises méthodes pour collecter l'information. En s'appuyant sur deux exemples d'évaluation des dégâts liés aux tempêtes de 1999 sur des milieux forestiers, cet article se propose d'une part de faire le point sur un certain nombre de méthodes d'inventaire (rappel des enjeux, nécessité d'évaluer la qualité des estimateurs, ...) et d'autre part, de montrer qu'une enquête relevant de la théorie de l'échantillonnage peut s'avérer précise et légère. L'évaluation du résultat porte principalement sur des aspects, spatiaux et thématiques.

ABSTRACT. Qualifying and quantifying landscapes are based on ground inventories. Actually, there are no good and bad methods to collect data. Using two examples for assessing the damages resulting from 1999 wind storms on forest areas, this paper will on one hand evaluate a number of inventory methods (reminder of objectives, necessity of checking estimators for qualification...) and on the other hand, show that an investigation based on a sampling theory can prove accurate and light. Assessing the result is mainly focused on spatial and thematic aspects, specifically oriented on the interest and reproductibility of the method for checking and following up territory transformations.

MOTS-CLÉS : échantillon, enquête, forêt, France, tempête

KEYWORDS: forest, France, sample, survey, tempest

Introduction

Lorsque la qualification et/ou la quantification des paysages reposent sur des inventaires de terrain, il convient de réfléchir à la méthode de collecte de l'information qui sera la plus adaptée à la problématique développée. Il n'y a pas, dans l'absolu, de bonnes ou de mauvaises méthodes pour collecter son échantillon. En s'appuyant sur deux exemples d'évaluation des dégâts liés aux tempêtes de 1999 sur des milieux forestiers, cet article se propose d'une part de faire le point sur un certain nombre de méthodes d'inventaire (rappel des enjeux, nécessité d'évaluer la qualité des estimateurs, ...) et d'autre part, de montrer qu'une enquête relevant de la théorie de l'échantillonnage peut s'avérer précise et légère. L'évaluation du résultat porte principalement sur des aspects, spatiaux et thématiques. Cette évaluation caractérise l'intérêt et la reproductibilité de la méthode pour l'observation et le suivi des transformations territoriales.

C'est donc une analyse critique puis une argumentation statistique qui sont proposées au débat sur les pratiques d'échantillonnage en géographie lorsqu'il s'agit d'évaluer les caractéristiques des paysages et leur évolution. Bien que les exemples soient pris dans des

domaines relevant plus de « l'identification des composantes du milieu » et de la « Science du paysage », tels qu'ils ont été développés par l'École de Toulouse (*cf.* par exemple Richard, 1989) pour l'analyse des paysages ruraux tropicaux, cette réflexion méthodologique se veut plus générale. Elle espère concerner également les pratiques de collecte d'information en territoires ruraux tempérés mais également en milieu urbain.

1. Contexte méthodologique

1.1 Réflexions sur la constitution de la base de données

Lorsque l'on enquête des paysages, ou du moins certaines composantes du milieu, ceux-ci n'apparaissent généralement pas sous forme d'enregistrements séquentiels dans une base de données. Il n'y a donc pas de listes dans lesquelles sélectionner les individus qui constitueront l'échantillon. Ce sont des portions de paysages qui vont être tirées au hasard (ou non). La base de sondage est donc constituée par une ou des cartes, des photos, de l'imagerie satellitale ou tout autre document comportant l'intégralité de la zone d'étude pour que tous les points de l'espace puissent être, éventuellement, inventoriés. C'est une des grosses différences entre les inventaires ou sondages portant sur des variables socio-démographiques et celles dont l'entrée est le paysage.

Les sondages sur liste et les sondages aréolaires, dont l'individu observable est une surface, diffèrent par la mise à jour de la base de données. Une carte périmée, une photo aérienne ancienne, ne porte plus forcément la représentation actuelle du paysage. Cependant, la portion de territoire qui a évolué est toujours présente, même si son affectation thématique est autre, contrairement à une liste de ménages ou de sièges d'exploitation agricole. Dans ces dernières, un déménagement, une cessation d'activité, une naissance ou une installation ne sont pas toujours signalés et répertoriés. C'est le problème de la dégradation des bases de sondage avec le temps. Celle-ci est donc plus problématique pour les sondages sur liste que pour les sondages spatiaux (aréolaires).

Partant du constat que l'on ne peut pas connaître chaque mètre carré de son terrain d'enquête, ce n'est pas un inventaire au sens premier de terme que l'on va réaliser mais un sondage. Celui-ci va permettre ou non d'extrapoler le peu que l'on a vu à l'ensemble du territoire enquêté. Cette capacité est fonction de la représentativité de l'échantillon observé, donc de sa richesse et de sa fidélité quant à la variabilité des paysages présents dans la zone d'enquête. On accepte, en réalisant un sondage plutôt qu'un recensement exhaustif, de commettre une erreur d'appréciation relative à la (petite) taille de l'échantillon par rapport à « l'immensité » de la zone d'étude.

1.2. Quelques méthodes d'échantillonnage des données de terrain

Quelle que soit l'information que l'on souhaite recueillir sur le terrain, la méthode de collecte relève soit des sondages par choix raisonné soit des sondages aréolaires aléatoires. Elle va faire appel ou non à une stratification (a priori ou a posteriori). L'unité d'observation pourra être un point, une ligne ou une surface. Cette dernière peut être un objet géométrique constant comme la placette d'inventaire ou de taille variable comme la parcelle agricole. Les unités d'observations peuvent être tirées en étant groupées ou non comme des grappes, de façon continue ou discontinue, en segments, transects, etc.

Les combinaisons suivantes sont les plus courantes :

- sondages par choix raisonnés - empiriques - (unités types, quotas, itinéraires, ...)

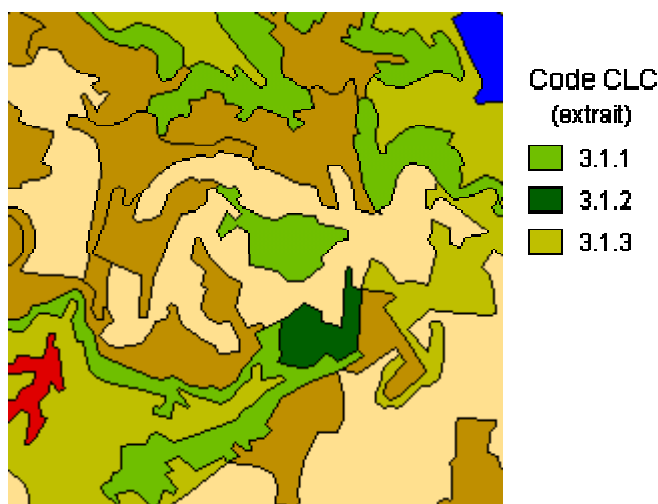
- sondages aréolaires aléatoires (placettes, points, segments, transects, ...).
Chacune a ses avantages et ses inconvénients. Nous allons en détailler quelques-uns.

1.2.1 Méthodes par choix raisonnés (non aléatoires)

Les sondages par choix raisonné sont majoritairement représentés par la méthode des **unités types** qui consiste à diviser la population (le paysage) en un certain nombre de sous-unités homogènes et à sélectionner parmi chacune de ces sous-unités un représentant « moyen » (exemple : un boisement de pin sylvestre « moyen », un ensemble pavillonnaire « type »). Celui-ci devra être représentatif des autres individus constituant le sous-ensemble, pour la variable qui a servi à faire le découpage mais également pour celles que l'on souhaite étudier (niveau d'infestation d'un parasite ou type de scolarisation pour reprendre l'exemple précédent). Comme le souligne P. Ardilly (Ardilly, 1999, p.163), ce choix « ... repose sur un pari audacieux : les individus **semblables** à l'individu moyen selon les variables qui ont servi précisément à définir cet individu moyen seront **également semblables** à l'individu moyen pour **toutes** les variables d'intérêt de l'enquête ». L'avantage réside dans la relative simplicité de détermination de l'échantillon, mais il va de soi qu'aucune estimation du biais ou de la précision du résultat n'est possible.

Ce type de méthodes est assez courant en télédétection lorsqu'il s'agit de qualifier des paysages ou de réaliser une couverture d'occupation du sol. Dans un article paru dans le *Courrier de l'environnement* de l'INRA (juin 2003, n°49), traitant de la transformation des paysages africains et des trypanosomoses, Stéphane de La Rocque indique que « 8 types de paysages ont été jugés représentatifs et ont été sélectionnés comme entités d'apprentissage pour une classification paysagère... » (La Roque, 2003, p. 81). Cette méthode d'échantillonnage relève des unités types. Il est dommage que l'auteur ne précise pas dans cet article pourquoi il a retenu cette méthode d'échantillonnage des données de terrain, ni quels en sont, à ses yeux, les avantages et les inconvénients.

Figure 1 – Carte d'occupation du sol - type CLC



Source : extrait d'un fichier Corine Land Cover à 1/100 000, IFEN, 1998.

La méthode des **quotas** est également une méthode empirique. Elle consiste à imposer à l'échantillon de respecter certaines proportions afin de représenter, au mieux, le paysage enquêté. Par rapport à la méthode des unités types, la méthode des quotas « obligera » à

enquêter une plus large combinaison de paysage. Elle suppose que les variables retenues pour déterminer les quotas sont très fortement liées à celles que l'on veut estimer.

A titre d'exemple (sans rapport avec ceux traités en deuxième partie), si l'on dispose d'une carte (*cf.* Figure 1) et d'une liste des postes d'occupation du sol [mode d'occupation des sols (MOS) de l'Institut d'aménagement et d'urbanisme pour la région Île de France (IAURIF), Corine Land Cover de l'Institut français de l'environnement (IFEN)...] et de l'importance de chacun d'eux (*cf.* Tableau 1), ainsi qu'une information altimétrique (*cf.* Tableau 2), il est possible de constituer un échantillon à l'image de la population. Si un poste d'occupation du sol occupe telle proportion dans la population (la zone étudiée) alors il conviendra qu'il occupe la même proportion dans l'échantillon, idem pour la proportion des altitudes.

Ici, les distributions des quotas sont qualifiées de marginales. Elles ne sont pas ventilées selon les deux critères. Les quotas seraient dits croisés si l'on avait un tableau à double entrée « croisant » les postes d'occupation du sol et les altitudes. Cette approche est plus fine, mais l'information est rarement disponible.

Tableau 1 – Liste et importance des postes d'occupation du sol – type CLC

Code	Nomenclature	Surface
---	---	---
3.1.1	Forêts de feuillus	12%
3.1.2	Forêts de conifères	7%
3.1.3	Forêts mélangées	1%
---	---	---
Total	---	100%

Tableau 2 – Part des différents étages de végétation

Étage	Altitude approximative	Surface
montagnard	800 – 1 500 m	75%
collinéen	- de 800 m	25%
total	---	100%

La méthode des **itinéraires** est également une méthode empirique. Elle consiste à enquêter des portions de paysage sur un itinéraire défini « en laboratoire ». Ce sont les mêmes avantages et inconvénients que pour la méthode des quotas avec en outre des inconvénients supplémentaires, à savoir : des problèmes de repérage si les objets observés sur le terrain ne sont pas contigus (cas des continums en particulier) ; des problèmes de biais liés à la surreprésentation des objets qui bordent la route par rapport à ceux qui sont à distance de l'itinéraire [phénomène des lisières (fonction de l'éclairage) qui sont différentes de par leur constitution spécifique et structurelle des bois de l'intérieur de la parcelle, des façades sur rue et des façades sur cour, des cultures intensives et extensives, ...]. Cependant, cette méthode tend à apporter une part d'aléatoire dans la sélection de l'échantillon. Elle est proche des méthodes par transects (*cf.* 1.2.2), sauf que le ou les itinéraires retenus restent choisis, pour une large part, par l'enquêteur.

L'avantage des méthodes par choix raisonnés réside dans la relative simplicité de détermination de l'échantillon. On considère également, qu'elles sont au moins aussi performantes sur de petits échantillons que les méthodes aléatoires (Desabie, 1965, p.54). Cependant, leur caractère non aléatoire ne permet aucune estimation du biais ou de la précision du résultat.

1.2.2 Méthodes aléatoires

Les sondages par **points** et par **placettes** sont des sondages probabilistes dont l'utilisation est recensée dans plusieurs inventaires d'occupation du sol. Un des plus connus, en France, est sans doute TER-UTI, une enquête sur l'utilisation du territoire menée par le Service central des enquêtes et études statistiques (SCEES) du Ministère de l'agriculture depuis le début des années 60 (Fournier, 1972). L'autre inventaire d'occupation du sol qui fait référence est celui de l'Inventaire forestier national (IFN) dont la mise en place remonte également aux années 60 (IFN, 1985). Dans les deux cas, l'Institut géographique national (IGN) réalise des couvertures aériennes, à des échelles comprises entre le 1/17 000 et le 1/30 000, sur lesquelles sont « incrustées » des grilles de points sur la partie « utile » de la photo (zone centrale hors recouvrement pour éviter les doubles comptes et les omissions).

Dans le cas de l'IFN, les points servent à tirer des surfaces. Elles font l'objet d'interprétation et de visites de terrain avec un certain nombre de prélèvements. Les prélèvements se feront sur des surfaces concentriques décroissantes en fonction de la complexité des mesures (Chevroux, 1988, p.32 et p.71). Lors de l'étude d'un département, toutes les photographies sont retenues. La grille de point qui se répète de photo en photo est sans orientation privilégiée et fournit donc un tirage systématique à un degré. On peut assimiler le tirage systématique à un tirage élémentaire ou aléatoire simple (Desabie, 1965, p.98). Donc les catégories de paysages ont une probabilité d'appartenance à l'échantillon proportionnelle à la surface du territoire qu'elles occupent.

Bien que proche de celle de l'IFN, la méthode utilisée pour l'inventaire TER-UTI diffère pour les aspects suivants. Le tirage des points est un tirage à deux degrés car on tire d'abord certaines photographies puis la grille de points sur les photos retenues (on peut aussi l'assimiler à un tirage au premier degré de grappes de points équidistants). Les photographies et les points sont orientés selon des axes nord-sud et est-ouest. L'observation est réalisée sur un point sans dimension, en théorie. Ce sont donc normalement des proportions qui sont calculées et non des surfaces.

Ces deux méthodes (points et placettes d'une manière générale) ont des caractéristiques communes. L'enquête est relativement coûteuse en temps car on se déplace beaucoup entre les points d'enquête par rapport au temps d'observation proprement dit. Elles sont peu adaptées aux objets paysagers compacts et peu couvrants, surtout dans le cadre d'un tirage à deux degrés (Brion, 1982, p.17), mais également pour les vergers, les haies et les peupleraies pour lesquels l'IFN utilise des grilles spéciales (IFN, 1985, pp.28-31). En revanche, la mise en œuvre du tirage est assez simple si l'on dispose de documents numérisés et d'un logiciel, Système d'information géographique (SIG) ou autre, pour effectuer le tirage des échantillons et imprimer le document d'enquête (*cf. infra*). De plus, la méthodologie statistique est bien rodée, le dépouillement est rapide et sensiblement aussi simple qu'un recensement du fait de l'équiprobabilité du tirage (Fournier, 1986, p.4).

Les méthodes de segmentation du paysage en **segments** ouverts, pondérés ou fermés ont également vu le jour dans les années 50-60, aux États-Unis (Meyer-Roux, 1981, p.13). Elles ont été testées en France dans les années 80 (Fournier, 1986) sur des paysages à dominantes agricoles, à des fins de recensement de population en milieu urbain (Dureau, 1990), en zone agropastorale tropicale (Godard, 1996), ou en milieu forestier tempéré (Godard, 1998). Les applications opérationnelles sont venues de l'Union européenne qui, depuis 1992, dans le cadre du programme *Monitoring Agriculture with Remote Sensing* (MARS), produit une partie des statistiques agricoles à l'aide de sondages aréolaires (Gallego, 1999).

Le segment fermé est en général de forme géométrique simple (un carré) dont on a déterminé la taille a priori et dans lequel l'enquêteur relève les limites des formations rencontrées. Il contient en général plusieurs parcelles. Pour le segment ouvert, on enquête non seulement les parcelles incluses dans le segment mais également toutes les terres des propriétaires (ou simples exploitants !) qui ont « quelque chose » dans le segment. On comprend ici qu'il est possible d'enquêter tout autre chose que du foncier ou des formations végétales. Ce peut être, à titre d'exemple, le cheptel des agriculteurs ou le mode de scolarisation des enfants des propriétaires de parcelles incluses dans le segment. Enfin, pour le segment pondéré, l'enquête se déroule chez les propriétaires des parcelles incluses (rendement en blé, cylindrée de la voiture du ménage ou autre). Cette information sera pondérée en fonction du ratio « taille de la parcelle de l'enquêté dans le segment par rapport à la superficie totale du segment ».

Cette méthode d'enquête par segments présente les caractéristiques suivantes :

- en milieu agricole, il est parfois nécessaire de redessiner le parcellaire, celui-ci ne correspond pas toujours au parcellaire foncier ;
- en milieu « naturel » le repérage, comme pour les enquêtes par placettes ou points, peut être long et peu précis sans l'apport d'une localisation de type *Global Positioning System* (GPS) ;
- l'observation d'un ensemble jointif de parcelles rentabilise le passage sur le terrain.

L'enquête par **transects** est la dernière méthode d'échantillonnage présentée ici. Elle répond surtout à des besoins spécifiques lors d'enquêtes sur des objets qui seraient mal perçus, ou ignorés, par tirage de points ou qui obligerait à un gros effort d'enquête (en termes de temps passé sur le terrain) pour une faible quantité d'informations récoltées. Les objets linéaires se prêtent bien à ce type d'observation. L'analyse des lisières forestières internes et externes, et de toutes autres interfaces, justifie l'utilisation d'observation par transects. Elle nécessite toutefois de repérer dans le paysage les orientations et les régularités qui vont conditionner l'azimut et l'espacement des transects si le tirage est systématique (Galochet, 2002). Là encore, pour suivre un transect sur une assez longue distance, il est préférable d'utiliser un GPS qu'une boussole. Celui-ci tirant l'enquêteur du point de départ au point d'arrivée, il peut se consacrer intégralement à la collecte de l'information. Si les transects ne sont pas tirés au hasard, ce sont des itinéraires (*cf.* 1.2.1).

L'avantage des méthodes aléatoires réside dans la possibilité d'estimer le biais ou la précision du résultat. Elles sont pratiquement reproductibles à l'identique avec l'espérance d'un résultat analogue. Comme nous le verrons en deuxième partie, même avec un relativement petit échantillon, elles peuvent offrir des résultats exploitables, voire précis. Elles nécessitent théoriquement, une connaissance préalable de la zone d'étude, mais celle-ci peut se réduire à l'acquisition d'une base de données (liste ou de préférence document spatial, *cf.* 1.1) support du plan d'échantillonnage. La contrainte majeure réside dans la quasi obligation d'enquêter l'individu - objet - tiré au sort, sous peine de faire perdre au tirage son caractère aléatoire (Desabie 1965 p.55 et 188).

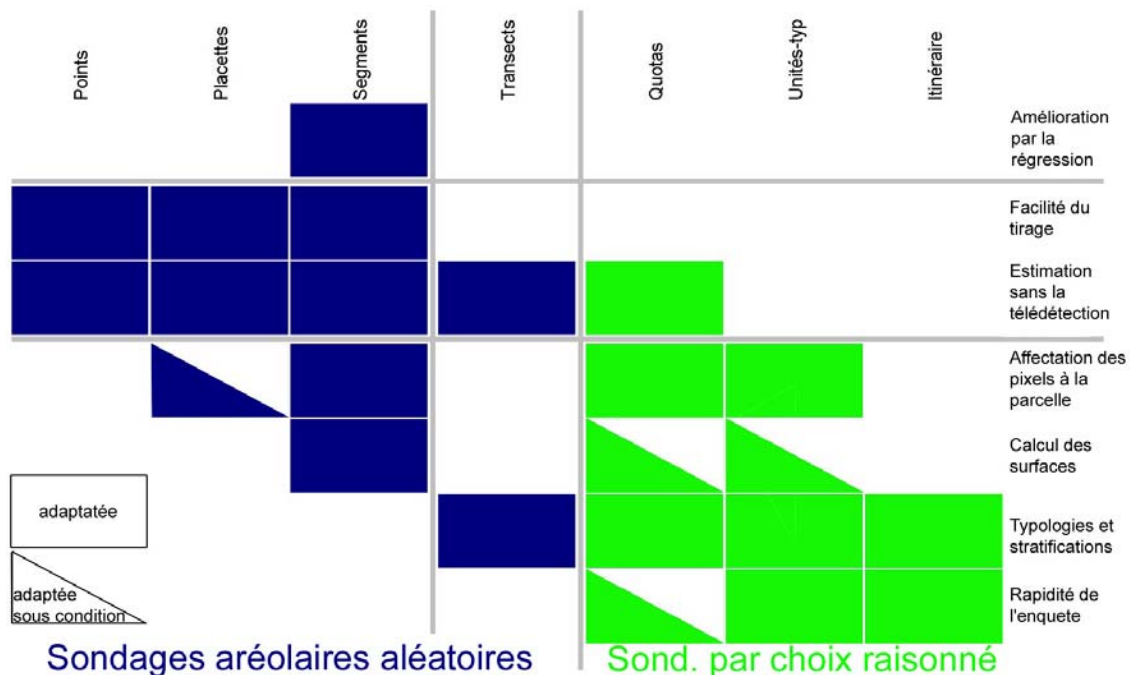
1.3 Méthodes d'échantillonnage et contraintes de mise en œuvre

Si la télédétection aérienne ou spatiale offre un excellent support d'enquête de terrain, elle ne peut pas totalement s'y substituer. Il y a toujours un moment où il est nécessaire d'affecter les pixels à la parcelle, d'un point de vue thématique au moins. On a vu précédemment son intérêt comme base de sondage. Cependant, avec des résolutions spatiales

de plus en plus fines, l'imagerie satellitale vient concurrencer les autres bases de sondage aréolaire et en particulier la photographie aérienne (Hotyat, 2003). Pour peu que l'on ait recours à des capteurs à large champ d'observation, l'exhaustivité de ces données en fait un outil nettement supérieur à tous les autres comme support du plan d'échantillonnage. Il convient d'analyser son intégration dans le processus d'évaluation des paysages.

Pour les sept types d'observations évoqués précédemment, un certain nombre de critères relatifs aux enquêtes et sondages ont été codés en présence - absence puis triés (cf. Figure 2) pour permettre d'identifier, selon les objectifs poursuivis lors de l'enquête, celui ou ceux qui paraissent les plus pertinents. Le résultat présenté (cf. Figure 2) oppose les types d'observation probabiliste à ceux réservés aux méthodes empiriques. Ceci est, en partie, fonction des critères retenus, toutes choses égales par ailleurs. Si les méthodes de l'**unité type** et de l'**itinéraire** (dans une moindre mesure la méthode des **quotas**) bénéficient d'un bref temps d'enquête sur le terrain (*Rapidité de l'enquête*), deux ou trois stations d'enquête par type d'occupation du sol suffisent, il faut, en revanche, que ces types soient déterminés à l'avance. Cela se retrouve dans le temps de préparation en laboratoire (*Typologies et stratifications*) qui signifie que l'on a réalisé des typologies ou stratifications à l'aide de photo-interprétation ou de classifications automatiques de l'image satellitale pour dégager un nombre de thèmes possibles dans la base de sondage (le paysage). L'un ne compensant peut-être pas l'autre, il faut savoir privilégier le temps passé à préparer l'enquête en laboratoire ou le temps passé en observation sur le terrain.

Figure 2 – Types d'observation et contraintes de mise en œuvre



Les **points**, **placettes** et **segments** ont des comportements très différents, quant aux variables retenues. Leur principal avantage est d'être facilement mis en œuvre par tirage aléatoire (*Facilité du tirage*), depuis un logiciel de SIG par exemple. Cela ne nécessite aucune connaissance du terrain d'enquête et peut être automatisé. On prendra toutefois soin d'évaluer l'autocorrélation spatiale pour conserver une certaine indépendance des échantillons. Car, comme le rappelle Ph. Roche (Roche, 1995, p.18), la dépendance des échantillons statistiques (distribution non normale, non stationnarité, ...) induit une invalidité partielle des tests statistiques.

De même, moins sensibles que les autres types d'observation à l'absence de données télédéteectées [(*Estimation sans la télédétection*) nuages sur la zone d'étude ou autres raisons d'absence d'information], les points, placettes et segments, de par leur répartition aléatoire, permettent, quand même, de fournir des estimateurs relatifs à l'évaluation des paysages. Le **transect** et les **quotas** sont également dans ce dernier cas (*Estimation sans la télédétection*), mais ils nécessitent des pré-traitements (*Typologies et stratifications*) pour être efficaces (cf. *supra*) et ne sont pas faciles à mettre en œuvre, surtout les transects.

Si l'on souhaite, dans un processus d'évaluation des paysages, quantifier certains paramètres comme le taux de boisement, le rendement en blé ou le pourcentage de sols nus... il est important de pouvoir affecter sans ambiguïté les mesures effectuées sur une surface au sol au pixel, plus généralement aux pixels qui les représentent dans l'échantillon (*Affectation des pixels à la parcelle*), ce qui n'est possible que sur des types d'échantillons également surfaciques. Même avec un positionnement GPS, l'affectation d'un relevé linéaire ou ponctuel asurfacique à un pixel est trop aléatoire pour être effectué sans risque d'erreur. Ce sont donc les segments, unités types et éventuellement placettes et quotas qui peuvent être le support de mesures nécessitant une certaine emprise au sol pour être calculé. Si la corrélation entre valeur radiométrique des pixels et variables quantifiées est significative, et suffisamment intense, il sera alors possible de réaliser une cartographie de la variable sur l'ensemble de la base de sondage, à l'aide du fichier satellitale.

L'évaluation de la superficie des différentes classes d'occupation du sol est possible avec l'ensemble des types d'échantillon. Ce peut être la proportion de pas faits dans un type de culture par rapport au nombre total de pas faits tout au long de l'itinéraire. Cependant, seuls les segments, et éventuellement les unités types, donnent *pour chacun d'eux* l'extension spatiale et donc la variance de chaque mode d'occupation du sol. Cet avantage comporte une contrainte que ne possèdent pas les autres types d'unités d'enquête. Il faut calculer la surface de chaque mode d'occupation du sol à l'intérieur de chaque segment ou unité type (*Calcul des surfaces*). Ce calcul donne donc l'extension spatiale de chaque thème par unité d'échantillonnage mais également, sous certaines conditions (entre autres de nombre d'échantillons), la variabilité spatiale (géographique) de cette information (concentration ou absence dans tel secteur de tel mode d'occupation du sol). Ce que ne peuvent fournir les échantillons aréolaires « asurfaciques », encore appelés discontinus.

Enfin, il est possible d'améliorer les estimations (de superficie ou autre) à l'aide la télédétection. Au moment de l'extrapolation des résultats, on va chercher à diminuer les fluctuations aléatoires des estimations en utilisant des informations extérieures à l'échantillon. Déjà connues dans les années soixante, ces techniques de redressement sont les estimations pondérées (stratification à posteriori, Desabie, 1965, p.173), l'estimation par le quotient (Desabie, 1965, p.200), l'estimation par la régression (Desabie, 1965, p.215), etc. Celle qui a donné les meilleurs résultats (fiabilité, mise en œuvre...) semble, au regard des citations bibliographiques postérieures, être l'estimateur par la régression. Celui-ci « ... exige une connaissance minimum du terrain, améliore la variance de l'estimation de superficie obtenue par télédétection en éliminant en particulier le biais qui peut exister sur les observations-terrain. » (Meyer-Roux, 1985, p.5). Ce sont des images de télédétection classifiées qui sont utilisées comme covariables pour redresser les estimations calculées pour les statistiques de l'U.E. (Gallego, 1999, p. 64). Le souci de précision dans l'évaluation des paysages pilote ici clairement le choix des méthodes. Le seul type d'investigations qui permette cette amélioration par la régression (*Amélioration par la régression*) est une méthode de tirage aléatoire, le segment.

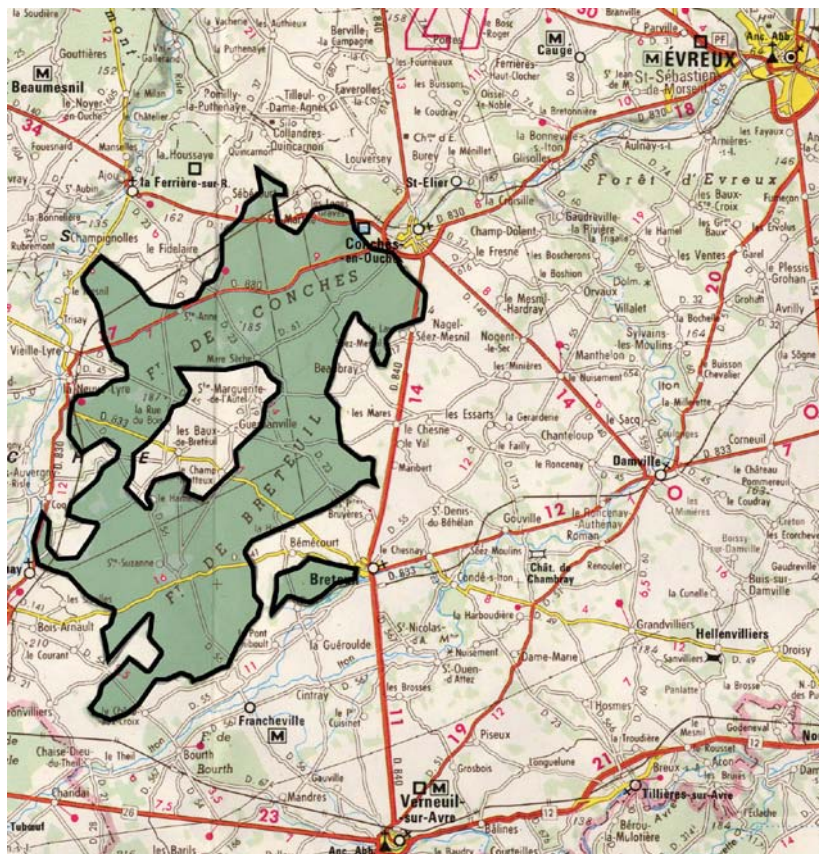
On aura compris, à l'aide de ces quelques présentations de plans d'échantillonnage que, bien qu'il soit possible et parfois justifié de recourir à des méthodes empiriques, la

nécessité d'évaluer, de quantifier, certains aspects des paysages, impose, en toute rigueur, de recourir à des méthodes probabilistes. Un simple calcul de surfaces, quelle que soit la méthode retenue, repose sur une sélection des échantillons qui relève du hasard, sinon le biais inhérent à la sélection est et restera inconnu, quoi que l'on fasse, rendant vaine toute extrapolation à l'ensemble de la zone d'étude. Les exemples présentés ci-après, sont là pour démontrer l'intérêt et la faisabilité, en dehors des Services statistiques, de ces démarches probabilistes.

2. Exemples d'utilisation de sondages probabilistes

Dans les exemples développés ci-après, nous voulions déterminer le taux de dégâts sur l'ensemble des deux zones d'étude (Conches-Breteuil, Eure et Bourron-Marlotte, S.-et-M., deux « univers » au sens statistique), mais également le connaître par unité de gestion pour établir des comparaisons. Subdiviser la zone d'étude en unités de gestion revient à stratifier l'espace.

Figure 3 - Carte de localisation des forêts étudiées : Massif de Conches-Breteuil (Eure)



Source : fond IGN, Série rouge au 1/250 000, n°102

Stratifier, c'est découper l'univers, c'est-à-dire le paysage, en sous-ensembles homogènes (les strates) par rapport à une variable dont on suppose (ou on sait par une étude comparable) qu'elle est fortement corrélée avec celle(s) que l'on cherche à estimer (le taux de dégâts par exemple). Le principal objectif de la stratification est d'augmenter la précision de l'information globale (le taux de dégâts) pour l'ensemble. Ou bien, c'est d'obtenir une précision suffisante au niveau des strates (Desabie, 1965, p.134), ou encore, assurer une

meilleure couverture géographique de l'univers et, là, pas forcément pour améliorer les estimations d'ensemble. Ces objectifs peuvent paraître contradictoires (*cf.* développements en annexe). En fait, seule l'optimisation simultanée de ceux-ci l'est réellement.

Les stratifications existent sous un grand nombre de variantes. Nous en avons pratiqué deux lors de ces inventaires (Amat, 2003). L'une a été le support d'un tirage avec un **taux de sondage équivalent** sur les différentes strates (Conches-Breteuil, Eure), l'autre a reçu un **taux de sondage variable** selon les strates (Bourron-Marlotte, S.-et-M.).

2.1. Exemple de stratification à taux de sondage constant

2.1.1 Présentation de la zone d'étude du massif forestier de Conches-Breteuil

Ce premier exemple porte sur le suivi des dégâts de la tempête de 1999 sur le massif forestier de Conches-Breteuil. Comme les autres sites d'étude, il est situé sur un plateau où les conditions du milieu sont assez constantes. Cela permet une comparabilité stationnelle pour évaluer l'impact des gestions sylvicoles de l'ensemble des sites au regard des dégâts constatés (Amat, 2003). Le secteur étudié se présente en deux parties distinctes, l'une au nord, la forêt des Hospices civils de Lyon (HCL), forêt privée gérée par l'Office nationale des forêts (ONF) et l'autre au sud, également privée et gérée par le Groupement forestier de Souvilly (GFS). Le plan de sondage est un tirage systématique aligné de points (en fait des placettes car dotées d'une surface) mis en place à l'aide d'une grille « jetée » sur la carte topographique (IGN, 1981). Il a servi à enquêter 91 placettes d'environ 2 100 m² (30 m par 70 m), soit 44 pour les HCL et 47 pour le GFS (*cf.* Figure 4). La phase de terrain a eu lieu d'août à novembre 2000.

2.1.2 Résultats de la stratification à taux de sondage constant

Cet article n'a pas pour objet de présenter les résultats des inventaires des dégâts relatifs à la tempête de 1999. Il est essentiellement méthodologique. Toutefois, il s'appuie sur les résultats d'un de ces inventaires pour décrire les calculs d'estimation (*cf.* annexe A.1) de la proportion des dégâts sur le massif de Conches-Breteuil¹.

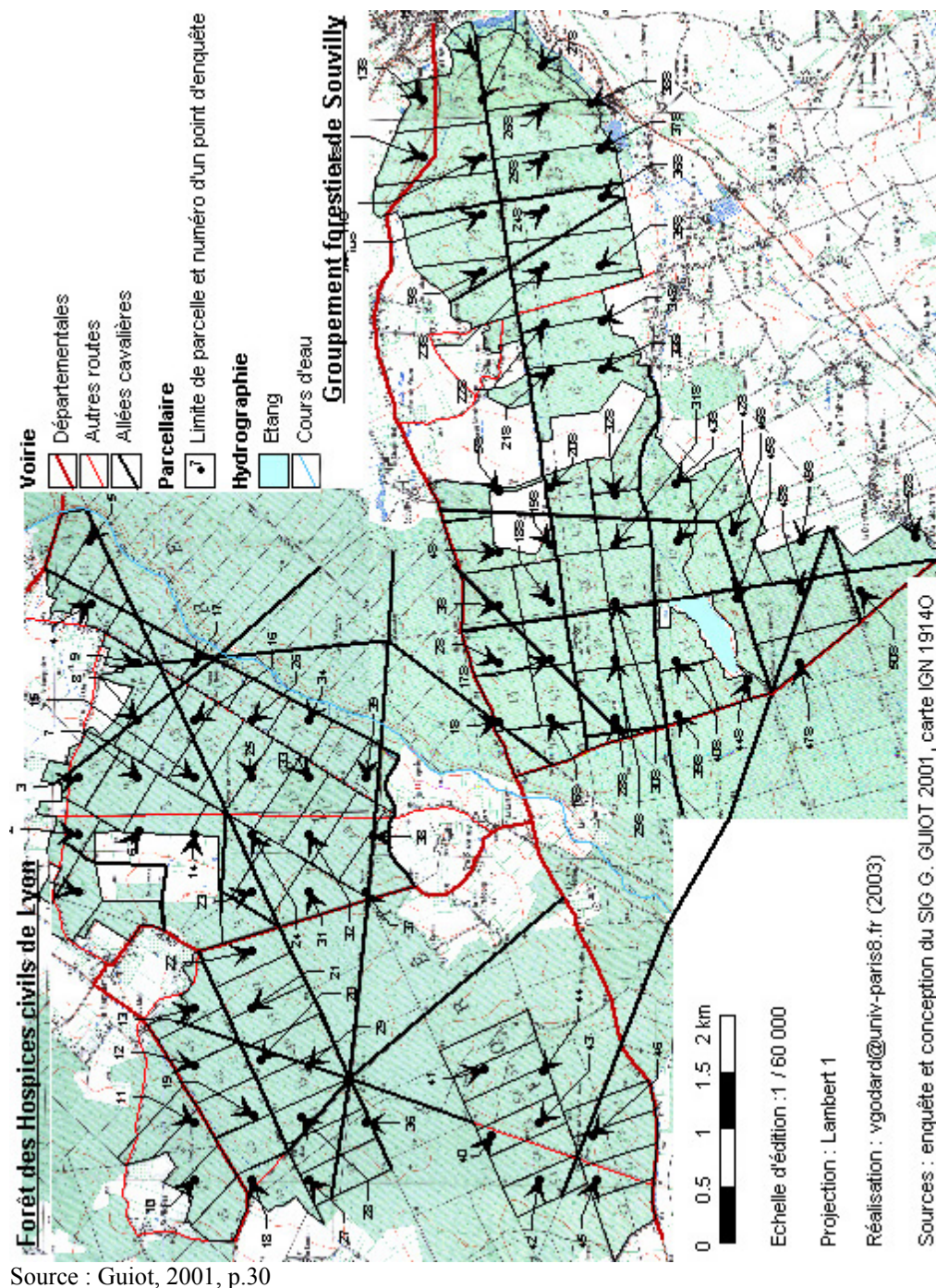
Chaque strate étant considérée comme un sous-univers indépendant où l'on prélève un échantillon, dans l'exemple normand, nous avons retenues deux strates. L'une correspond aux Hospices civils de Lyon (HCL), l'autre au Groupement forestier de Souvilly (GFS). Leurs valeurs sont reprises dans le tableau des paramètres (*cf.* Tableau n°3 et 4).

Ces résultats sont-ils en adéquation avec les objectifs de la stratification présentés en introduction de la 2^{ème} partie ? Rappelons qu'ils doivent, d'une part, présenter une estimation globale de la proportion réelle P à partir des deux secteurs d'étude et, d'autre part, permettre de faire une évaluation strate par strate à des fins de comparaisons.

Pour le premier objectif, il faut se demander si la stratification « géographique » utilisée permet d'améliorer la précision de l'estimation globale. Si tel est le cas, cela veut dire qu'ici la variable *répartition spatiale* (connue) est corrélée à la variable *dégât* (inconnue et recherchée), avec ou sans variables intermédiaires (histoire des peuplements, gestions, caractéristiques aérologiques,...), donc qu'ici la séparation physique vaut strate. Pour ce faire, nous avons juxtaposé dans le tableau n°3 les résultats obtenus avec stratification et ceux que l'on aurait obtenus sans stratification (*cf.* annexe A.1 pour le détail des calculs).

¹ Il apporte également une correction méthodologique par rapport à une précédente publication (Godard, 2004)

Figure 4 - Répartition des placettes d'enquête dans la forêt des Hospices civils de Lyon et le GF de Souvilly



Bien que la différence de dégâts (proportion des dégâts, cf. Tableau n°3) soit nette, la comparaison des résultats avec et sans stratification pointe la faible efficacité de celle-ci quant à l'amélioration globale de la précision de l'estimation de la proportion P des dégâts. Au risque $\alpha = 5$ p.100, la proportion P de dégâts appartient à l'intervalle $23,8 \text{ p.100} \pm 6,7 \text{ p.100}$ si le sondage est stratifié ou à l'intervalle $23,8 \text{ p.100} \pm 7 \text{ p.100}$ s'il ne l'est pas ! Le choix de la variable stratifiante (Hospices civils de Lyon contre Groupement forestier de Souvilly) n'est pas pertinent pour l'amélioration globale de l'estimation des dégâts. L'est-il pour le deuxième objectif (améliorer la précision des estimations sur les sous-ensembles) ?

Tableau 3 – Apport de la stratification dans l'estimations des dégâts en forêt de Conches-Breteil (Eure)

Stratification	Proportion estimée des dégâts	Variance de cette estimation	Ecart-type	CV	Erreur-type ($\alpha = 0,05$)	Intervalle de confiance borne inférieure	borne supérieure	Précision
Avec	23,8%	$1,2 \cdot 10^{-3}$	3,4%	14,3%	6,7%	17,1%	30,5%	28,1%
Sans	23,8%	$1,3 \cdot 10^{-3}$	3,6%	15,0%	7,0%	16,8%	30,7%	29,4%

L'estimation des proportions et de leur variance par strate se dépouille comme un recensement (cf. A. introduction de l'annexe). Sur chaque strate, p_h est l'estimation de la proportion d'arbres touchés et $v(p_h)$ l'estimation de la variance de cette proportion. Le coefficient de variation est noté CV (cf. Tableau n°4).

Tableau 4 – Résultats des estimations par tirage stratifié en forêt de Conches-Breteil (Eure)

Massifs	Proportion des dégâts	Variance apportée par chaque strate	Écart-type	CV	Erreur-type ($\alpha = 0,05$)	Intervalle de confiance borne inférieure	borne supérieure	Précision
Strate 1 HCL	33,5%	$8,96 \cdot 10^{-4}$	3,0%	8,9%	5,9%	27,6%	39,4%	17,5%
Strate 2 GFS	12,2%	$2,77 \cdot 10^{-4}$	1,7%	13,6%	3,3%	9,0%	15,5%	26,7%
Ensemble	22,6%	$1,2 \cdot 10^{-3}$	3,4%	14,3%	6,7%	17,1%	30,5%	28,1%

La différence entre les deux strates est nette. Il n'y a pas de chevauchement des deux intervalles de confiance. Elle est même hautement significative [$\epsilon_U = 3,5$ (significatif au seuil $\alpha < 0,001$)] d'après le test U de Mann et Whitney (Collectif, 1995, p.126). La précision des estimations (erreur-type divisée par proportion des dégâts au seuil de 5 p.100) est cependant moyenne. Elle est de 17,5 p.100 pour la strate 1 et de 26,7 p.100 pour la strate 2. Cela conduit à une précision globale de 28,1 p.100.

Sans apporter une amélioration exceptionnelle, la stratification sur des critères géographiques, de propriété ici, remplit son rôle. La couverture géographique des deux espaces (strates géographiques) est, de fait, correcte car ils ont un nombre d'échantillons suffisant pour tester leur différence de manière significative. Ce résultat doit inciter l'enquêteur à rechercher les variables expliquant les différences de dégâts (essences, structures, conduites, ...).

2.2. Exemple de stratification à taux de sondage variables

2.2.1 Présentation de la zone d'étude de Bourron-Marlotte (S.-et-M.)

La forêt domaniale de Fontainebleau, à une soixantaine de kilomètres au sud-est de Paris dans la Couronne rurale, forme, avec la forêt domaniale des Trois Pignons et une frange de bois privés, un vaste massif de plus de 25 000 hectares. Cette forêt occupe des milieux physiques diversifiés et compartimentés. Les biotopes vont du pôle acide au pôle carbonaté quant aux bilans hydriques stationnels, ils vont du sec à l'hyperhumide. La zone d'étude, dite de Bourron-Marlotte, est située au sud du massif (cf. Figure 5) sur les plateaux sommitaux sous-tendus par le calcaire de Beauce plus ou moins recouvert de dépôts sableux ou limoneux. Ils portent essentiellement des futaies de chênes et de hêtres. Intégralement incluse dans la

forêt domaniale, la zone d'étude comporte deux réserves biologiques « Les Érables et Déluge » et « La Gorge aux Loups ».

Contrairement au secteur de Conches-Breteil, le tirage des placettes échantillon y a été réalisé en deux étapes. Un premier tirage, ne concernant que la zone hors réserves, leur a été attribué dans l'échantillon 12 placettes (soit un taux de sondage de 0,6 p.100), un second, sur les réserves, 23 placettes (soit un taux de sondage de 4,4 p.100). Dans les deux cas, le tirage est aléatoire et stratifié. Le logiciel (ici Idrisi 32 v.2) génère des coordonnées X et Y aléatoires (comme pour un tirage élémentaire) à l'intérieur des cases d'une grille pour assurer une bonne répartition des placettes (comme pour un tirage systématique). Pratiquement, les cartes topographiques ont été importées puis géoréférencées, avant de servir de support aux tirages des échantillons. La phase de terrain a eu lieu de juin à septembre 2001.

Figure 5 - Carte de localisation des forêts étudiées : Secteur de Bourron-Marlotte (F. de Fontainebleau)



Source : d'après fond IGN, Série rouge au 1/250 000, n°103

La fiche d'enquête est peu différente de celle utilisée pour le secteur de Conches-Breteil (Amat, 2003, pp.104-109). L'inventaire proprement dit porte sur des surfaces d'environ 3 600 m² (60 m par 60 m).

2.2.2 Résultats de la stratification à taux de sondage variables

Les cas justifiant un tirage avec des taux de sondage différenciés selon les strates trouvent leur application pratique face à des distributions très dissymétriques (Desabie, 1965, p. 136). R. Clairin et Ph. Brion ajoutent que pour obtenir « (...) *la variance la plus faible (c'est-à-dire la meilleure précision au niveau de l'estimation globale sur l'univers).* » il faut utiliser un taux de sondage proportionnel à la dispersion de la variable « (...) *étudiée dans chaque strate : plus une strate est hétérogène vis-à-vis de cette variable, plus on y utilise un taux de sondage important.* » (Clairin, 1997, p.32).

En l'absence de toute information sur les zones de réserve, nous pensons que la forêt gérée par l'ONF était moins susceptible d'être touchée par la tempête (présence de jeunes peuplements, peu d'arbres séniles, etc.) que celle en réserve biologique intégrale (RBI). De ce fait, nous avons estimé qu'il y aurait une disparité importante entre les dégâts constatés en réserve et hors réserve (HR). Nous avons donc sur-enquêté la partie en réserve (cf. Tableau

n°7 en annexe A.2). C'est donc pour éviter de « trop » enquêter un secteur susceptible d'être moins porteur d'information que l'on a appliqué des taux de sondage différents. Ce choix était-il justifié au regard de l'accroissement de complexité de l'inventaire ?

Comme cela a déjà été développé à la section 2.2 pour Conches-Breteuil, les deux objectifs de la stratification sont d'une part de faire une estimation globale (plus fine que sans stratification) de la proportion réelle P à partir des deux secteurs d'étude et d'autre part de faire une évaluation strate par strate à des fins de comparaisons. Si, pour le premier objectif, dans le cadre des tirages à taux de sondages égaux, il est facile de reconstituer un tirage sans stratification, pour comparer l'estimation obtenue à celle issue des stratifications, cette opération est plus complexe dans le cadre des tirages à taux de sondages inégaux. La démonstration ayant été faite que la stratification géographique n'est pas toujours la plus efficace, nous n'y reviendrons pas (on peut toutefois, comme cela a déjà été décrit en introduction de la 2^{ème} partie, espérer un gain sur des secteurs s'y prêtant plus). Dès lors, la stratification à taux de sondage inégaux a-t-elle permis une meilleure connaissance des dégâts survenus en zone de réserve et hors réserve avec une intensité de sondage moindre que pour un tirage à probabilités égales, et donc de répondre au deuxième objectif des stratifications à moindre coût ?

Tableau 5 – Résultats des estimations par tirage stratifié à Bourron-Marlotte (Fontainebleau, S.-et-M.)

Massifs	Proportion des dégâts	Variance de cette proportion	Ecart-type	CV	Erreur-type ($\alpha = 0,05$)	Intervalle de borne inférieure	Confiance borne supérieure	Précision
Strate 1 RBI	18,2%	0,01%	0,9%	4,9%	1,7%	16,4%	19,9%	9,6%
Strate 2 HR	7,7%	0,05%	2,3%	30,0%	4,5%	3,2%	12,3%	58,7%
Ensembles	10,0%	0,06%	2,5%	24,9%	4,9%	5,1%	14,8%	48,8%

Dans l'exemple bellifontain, deux strates ont également été retenues. L'une correspond à « $h = 1$ » pour les parties en réserve (RBI), l'autre à « $h = 2$ » pour les parties hors réserve (HR). Les résultats sont présentés dans le tableau n°5.

Avec une moyenne de 10 p. 100 ($\pm 4,9$ p.100, au risque $\alpha = 0,05$) de dégâts [nombre d'arbres abattus (chablis et volis) enquêtés sur le total des arbres échantillonnés par placette], les dégâts sont globalement moins importants qu'à Conches-Breteuil. Cependant, ces dégâts ne sont pas uniformément répartis. Comme nous le pensions, les réserves biologiques intégrales (RBI) ont été beaucoup plus touchées que le reste du secteur géré par l'ONF. Si les deux RBI enregistrent **18,2** p.100 d'arbres enquêtés et touchés ($\pm 1,7$ p.100, au risque $\alpha = 0,05$), pour le reste du secteur, ce ne sont que **7,7** p.100 ($\pm 4,5$ p.100, au risque $\alpha = 0,05$) qui le sont [$H = 5,26$, différence significative au seuil de 5 p.100, test H de Kruskal et Wallis (Collectif, 1995, p.130)]. Avec un nombre de placettes peu élevé, 35 au total (23 pour les RBI et 12 pour le secteur hors réserve), les intervalles de confiance associés aux dégâts moyens sont variables d'une strate à l'autre. Ils encadrent ceux obtenus sur Conches-Breteuil (cf. 2.1.2).

La précision des estimations (erreur-type divisée par proportion des dégâts au seuil de 5 p.100) est bonne pour la strate 1 et mauvaise pour la strate 2 (respectivement **9,6** p.100 et de **58,7** p.100). La précision globale est de presque 49 p.100 (cf. Tableau 5). L'explication vient du trop faible nombre de placettes dans la strate 2.

Cependant, là encore, la stratification remplit nettement son rôle. La couverture géographique des deux espaces (strates géographiques), bien que déséquilibrée, est correcte car ceux-ci ont un nombre d'échantillons suffisant pour tester leur différence de manière

significative. Un nombre ne placettes plus important hors zone de réserve aurait toutefois permis une meilleure précision globale des estimations.

Que conclure de l'analyse de ces deux exemples ? Globalement, la zone d'étude de Conches-Breteuil (Eure) a été plus affectée par les tempêtes de décembre 1999 [entre 17,1 et 30,5 p.100 des arbres ont été touchés (au risque $\alpha = 0,05$)], que celle de Bourron-Marlotte (S.-et-M.) [entre 5 et 15 p.100 des arbres ont été touchés (au risque $\alpha = 0,05$)]. Au niveau des strates, la précision de la fourchette des résultats informe sur la pertinence et l'efficacité de la stratification. Par ordre décroissant de précision vient la strate des réserves de Bourron-Marlotte (9,6 p.100), puis celle des Hospices (17,5 p.100) et de Souvilly (26,7 p.100) à Conches-Breteuil et enfin la strate hors réserve de Bourron-Marlotte (58,7 p.100). Les explications qu'il est possible d'avancer en tenant compte de la variable étudiée (dégâts sur les arbres) sont les suivantes : dans les réserves de Bourron-Marlotte, la strate est assez homogène et le nombre de placettes d'enquête (23) suffisant ; à Conches-Breteuil, les strates ne sont pas très homogènes mais le nombre important de placettes d'enquête (92) compense en partie l'hétérogénéité de la répartition des dégâts ; contrairement à la strate hors réserve de Bourron-Marlotte, où il y a une forte variabilité des dégâts et peu de points d'enquête (12) pour en rendre compte.

D'une manière générale, au regard de la mise en place et du dépouillement, c'est la stratification avec tirage à probabilités égales, la plus simple à mettre en œuvre. Elle relève des tirages stratifiés représentatifs, où le taux de sondage est le même dans chaque strate. La probabilité d'être dans l'échantillon est la même pour tous. L'autre, à taux de sondage inégaux, est plus complexe à tous points de vue. Elle peut, dans le cas d'une forte hétérogénéité de répartition entre les strates de la variable d'intérêt, apporter un gain de précision dans les estimations. Cependant, elle nécessitera toujours un surcroît d'investissement par rapport aux tirages à probabilités égales dont, on le voit ici, le gain n'est pas toujours au rendez-vous.

Conclusion

À plusieurs reprises, il est apparu nécessaire, dans une démarche d'évaluation des paysages, de quantifier tel ou tel élément d'occupation du sol ou telle ou telle caractéristique du paysage. Il a été rappelé que les méthodes empiriques comme les méthodes probabilistes sont capables de produire une estimation. Cependant, seules les méthodes probabilistes sont capables de fournir la précision de cette estimation.

L'observation de terrain reste indispensable pour évaluer (quantifier) les paysages. Dans ce processus, il est nécessaire que l'échantillon de terrain soit représentatif, à savoir, d'une manière ou d'une autre, proportionnel à chaque type d'occupation du sol. Les sondages probabilistes s'opposent aux sondages empiriques dans la mesure où on ne connaît pas a priori, dans ces derniers, la probabilité d'appartenance d'un individu à l'échantillon. En général, ils sont utilisés lorsqu'il y a absence de base de sondage. Ce n'est donc pas un argument opposable lorsque l'on dispose de cartes, photographies aériennes ou autres images satellitaires du paysage à inventorier. En revanche, ils sont plus rapides et moins chers à mettre en œuvre que les sondages probabilistes (Ardilly, 1994, p.153), « *Le coût du sondage, qui est alors faible, peut compenser le manque (inconnu) de précision.* » (Gourieroux, 1981, p.156) et souvent plus simples à gérer pour les cas de non-réponses (terrains engrillagés inaccessibles, par exemple). Là encore, une analyse objective du but poursuivi doit permettre

de trancher entre les méthodes de collecte pour retenir la plus appropriée à la problématique (en coûts et en précision).

Remerciements

Que Philippe Brion pour ses explications dévouées et indispensables, Michelle Giboire pour ses relectures attentives et sa traduction, Gérard Vallée pour son accueil et ses conseils, les étudiants de Paris 8 ainsi que tous ceux qui ont collaboré à la collecte des données de terrain : Sultan Aralioglu, Antoine Bouffandeau, Josselin Burhig, Quentin Burhig, Anna Godard, Lisa Godard, Alexandra Grandmoujin, Stéphane Guillez, Gwenaëlle Guiot, Kim Dae Young, Mickael John-Pierre, Brice Oyharçabal trouvent ici la marque de mes remerciements.

Bibliographie

- Amat J.-P., Godard V., Hotyat M. (2003) *Bilan des dégâts : Milieu, gestion, histoire et scénarios de reconstitution dans les sylvosystèmes touchés par les tempêtes de décembre 1999*. GIP-ECOFOR, Min. agriculture, 132 p.
- Ardilly P. (1994) *Les techniques de sondage*. Paris, Éd. Technip, 393 p.
- Brion Ph., Bertin M., Labat Ph. (1982) *Sondages*. Paris, Ministère de l'agriculture, SCEES, Série S. Col. : Méthodes et Applications Scientifiques n°5, 40 p.
- Chevrou R.B. (1988) *Inventaire forestier national*. Méthodes et procédures. Montpellier, Ministère de l'agriculture, IFN, 105 p.
- Clairin R., Brion Ph. (1997) *Manuel de sondages. Applications aux pays en développement*. Paris, CEPED, 2ème éd., Documents et Manuels du CEPED n°3, 108 p.
- Collectif (1995) *Aide-Mémoire statistique*, Montreuil, CISIA.CERESTA, 286 p.
- Desabie J. (1965) *Théorie et pratique des sondages*, Dunod, Paris, 481p
- Deville J.-C., Roth N. (1986), « La précision des enquêtes sur l'emploi », *Économie et statistique*, n° 193-194, p. 127-134.
- Dureau F., Barbary O., Michel A., Lortic B. (1990) *Sondages aréolaires sur image satellite pour des enquêtes socio-démographiques en milieu urbain. Manuel de formation*. Paris, ORSTOM, 15 fiches.
- Fournier Ph. (1972) *Étude sur l'utilisation du territoire. Méthodologie. Résultats 1969-1970-1971*. Paris, Ministère de l'agriculture, SCEES, Série Études, supplément du n°104, 112 p.
- Fournier Ph. (1986) *Enquête sur l'utilisation du territoire effectué en 1985 par la méthode des segments*. Paris, Ministère de l'agriculture, SCEES, Série S, Col. : Méthodes et Applications Scientifiques n°13, 63 p.
- Gallego F.J. (1999) Télédétection et estimation des superficies. Ed. Sc. G. Brossier et A.-M. Dussaix in : *Enquêtes et sondages. Méthodes, modèles applications, nouvelles approches*. Paris, Dunod, p. 60 – 68.
- Galochet M., Godard V., Hotyat M. (2002) Land Units and the Biodiversity of Forest Islets: From Satellite Images to Ground Analysis. Ed. by Lech Ryszkowski in : *Landscape Ecology in Agroecosystems Management*. Boca Raton, CRC Press, p. 317-330.
- Godard V. (2004) L'évaluation des paysages : empirisme ou probabilisme, question d'objectif - Une réflexion sur le plan d'échantillonnage accompagnée d'un exemple d'évaluation des dégâts tempétueux pris dans la forêt de Conches-Breteuil (Eure). in : *L'évaluation du paysage : une utopie nécessaire ? À la recherche d'indicateurs / marqueurs pluridisciplinaires*. Montpellier 15 et 16 janvier 2004, p.249-259.
- Godard V. (1996) « Evaluation of the natural surfaces by remote sensing and ground survey in a dry tropical environment ». *Revue électronique européenne de Géographie CYBERGEO*, n°13, mel 14/12/96, 22 p.
- Godard V., Simon L. (1998) Analyse des structures forestières des forêts de Saint-Gobain et Coucy-Basse (Aisne, France) à l'aide de l'imagerie satellitaire et de l'enquête de terrain : choix d'un pas d'échantillonnage. VII^{es} journées scientifiques du Réseau Télédétection de l'AUPELF-UREF. Sainte-Foy, 13-17 octobre 1997. Éd. par DUBOIS J.M.M. et al. in : *La réalité de terrain en télédétection : pratiques et méthodes*. Montréal (Qué., Canada), AUPELF-UREF, Col. Universités francophones, Actualité scientifique, p. 45-53.
- Gourieroux Ch. (1981) *Théorie des sondages*. Paris, Économica, 272 p.

- Guiot G. (2001) *Analyse de l'impact de la tempête de décembre 1999 à l'aide d'un SIG sur deux forêts du département de l'Eure en Haute Normandie, la forêt des Hospices civils de Lyon et le groupement forestier de Souvilly*. Université Paris VIII, mémoire de maîtrise de géographie, 154 p.
- IFN (1985) *But et méthodes de l'inventaire forestier national*. Paris, Ministère de l'agriculture, Service des forêts. 67 p.
- IGN (1981) 1914O, La-Neuve-Lyre, Carte au 1/25 000
- Hotyat M., Godard V. (2003) Paysages et modifications de l'espace appréhendés grâce aux nouvelles sources d'information. Photographies aériennes et données satellitales. A. Corvol (dir.), *In Les sources de l'Histoire de l'Environnement*, Tome III : le XXe siècle, Paris, L'Harmattan : 143-153.
- La Roque S. de (2003) « Épidémiologie des trypanosomoses africaines. Analyse et prévision du risque dans les paysages en transformation ». *Le Courrier de l'environnement de l'INRA*. Juin 2003, n°49, p. 80-86.
- Meyer-Roux J. (1981) « Estimation des surfaces cultivées à partir des données LANDSAT et des enquêtes de terrain ». *Cahiers de Statistique Agricole*, n°1-6, janvier-février, p. 13 – 16.
- Meyer-Roux J. (1985) « Observations terrain et télédétection ». *Cahiers de Statistique Agricole*, n°1-6, janvier-février, p. 5-11.
- Richard J-F., (1989), *Le paysage. Un nouveau langage pour l'étude des milieux tropicaux*. Coll. Initiations - Documentations Techniques n°72, ORSTOM, Paris, 217 p. (19 planches et cartes h.t., 7 planches photos couleur, 1 livret-guide de terrain h.t. de 37 p.)
- Roche Ph. (1995) « L'analyse spatiale des données écologiques : théorie et application à la végétation en Provence calcaire ». *Ecologie*, t.26 (1) : 17-32.

Annexe - Rappels et formulaires de stratification

A.1 - Que peut apporter une stratification dans un processus d'évaluation des paysages ?

Stratifier, c'est découper le paysage, en sous-ensembles homogènes (les strates) par rapport à une variable dont on sait ou suppose qu'elle est fortement corrélée avec celle(s) qu'on cherche à estimer. Usuellement, on stratifie avant le tirage de l'échantillon, puis on tire de façon indépendante dans les strates. En pratique, on stratifie sur une variable « principale » qui est corrélée avec la plus grande partie des variables étudiées. Le principal objectif de la stratification est, en général, de réduire la variabilité de l'estimateur au niveau de l'univers, c'est-à-dire augmenter la précision de l'information globale, pour l'ensemble. On peut aussi se satisfaire du niveau de précision donné par le tirage non stratifié et diminuer la taille de l'échantillon par rapport à celui-ci. Le deuxième objectif est d'obtenir une précision suffisante au niveau des strates (Desabie, 1965, p.134), c'est-à-dire que l'on souhaite, en stratifiant, améliorer la précision des estimations sur certains sous-ensembles ou bien encore, assurer une meilleure couverture géographique de l'univers et là pas forcément pour améliorer les estimations d'ensemble. Cependant, comme le rajoutent R. Clairin et Ph. Brion (Clairin, 1997, p. 28) « *Ces deux objectifs ne doivent pas être confondus.* ». Il y a un choix à faire quant à la répartition de la base de sondage et donc de l'échantillon entre les strates. Ce choix est fonction de l'objectif que l'on juge prioritaire.

La répartition des échantillons va donc se faire selon un taux de sondage égal entre les strates, ou proportionnellement à un paramètre que l'on estime corrélé à celui recherché. Si les strates sont homogènes au regard de la variabilité du thème étudié (variance très proche pour la variable qui lui est corrélée dans chaque strate), alors il est suffisant d'utiliser un tirage identique pour toutes les strates. En revanche, si les strates sont hétérogènes, il est préférable d'utiliser un taux de sondage proportionnel à leur dispersion. En tout cas, c'est la répartition dont on peut espérer qu'elle fournira la meilleure estimation sur l'ensemble de l'univers. Elle est évidemment un peu plus complexe !

Lorsque le taux de sondage est le même pour toutes les strates, on parle de tirage stratifié représentatif (ou encore de sondage autopondéré). Ce tirage est le seul qui se

dépouille comme un recensement (Desabie, 1965, p.145). Sa moyenne simple est calculée sur l'échantillon et permet d'estimer la moyenne sur l'univers. La variance de cet estimateur stratifié représentatif est « ... *toujours inférieure ou égale à celle du tirage aléatoire simple et d'autant plus inférieure que les moyennes calculées sur les strates diffèrent de la moyenne générale.* » (Clairin, 1997, p.31). On a donc toujours intérêt à stratifier. Au pire, cela n'apporte rien (mais nécessite du travail en plus car il faut connaître un minimum son terrain pour préparer le tirage), au mieux un gain de précision. En terme de bilan, « *Le sondage autopondéré simplifie le dépouillement, conduit à des calculs faciles et ne réserve pas de mauvaises surprises en cas d'erreur, mais il risque d'aboutir à des résultats peu significatifs pour les petites strates et il ne donne pas – en général - la précision globale la meilleure.* » (Clairin, 1997, p.35).

A.2 - Formulaire pour la stratification à taux de sondage constant

D'une manière générale, on a k strates. Chaque strate est considérée comme un sous-univers indépendant où l'on prélève un échantillon. L'univers des points enquêtés est noté M, celui de la strate h est noté M_h. La proportion des arbres touchés par la tempête (cassés ou tombés) est notée P. Cette proportion est inconnue, on va l'estimer sans biais par l'estimation du ratio « arbres enquêtés tombés sur arbres enquêtés total (tombés plus debout) » calculée sur l'échantillon stratifié P_s (cf. formule n°1) :

$$P_s = \frac{\widehat{X}_1}{\widehat{X}_2} = \frac{\sum_{h=1}^k \sum_{i=1}^m \frac{M_h}{m_h} * [x_1]_{ih}}{\sum_{h=1}^k \sum_{i=1}^m \frac{M_h}{m_h} * [x_2]_{ih}} \quad \text{formule n°1 (Clairin, 1997, p.90, d'après Deville, 1986)}$$

Avec

- X₁ = estimation du nombre d'arbres touchés sur l'ensemble des strates (l'accent circonflexe indique que c'est une estimation) ;
- X₂ = estimation du nombre total d'arbres sur l'ensemble des strates ;
- La double sommation (ΣΣ) indique que l'on fait la somme du nombre d'arbres x d'abord par placette (de « i = 1 » la première à « i = m » la dernière de l'échantillon), puis par strate (de « h = 1 » à « h = k ») ;
- M_h/m_h = coefficient d'extrapolation par strate. C'est une pondération qui tient compte du nombre total de placettes par strate (M_h) et du nombre de placettes de la strate dans l'échantillon (m_h).

Dans l'exemple normand, deux strates (k) ont été retenues. L'une correspond à « h = 1 » pour les Hospices civils de Lyon (HCL), l'autre à « h = 2 » pour le Groupement forestier de Souvilly (GFS). Leurs valeurs sont reprises dans le tableau des paramètres (cf. Tableau n°6). On obtient donc une proportion de dégâts estimée sur les échantillons de (cf. résultat n°1) :

$$P_s = \frac{(122,6 * 880) + (121,02 * 272)}{(122,6 * 2626) + (121,02 * 2223)} = 23,8\% \quad \text{résultat n°1}$$

On remarque qu'il y a beaucoup plus d'arbres enquêtés et touchés aux Hospices qu'à Souvilly (respectivement 880 contre 272, soit 3,2 fois plus) alors qu'ils ne sont que 1,2 fois plus nombreux à être enquêtés (2 626 contre 2 223). On remarquera également, que les coefficients d'extrapolation sont quasi identiques dans la mesure ou les taux de sondage le sont.

Tableau 6 – Paramètres du plan d'échantillonnage Conches-Breteuil (Eure)

	Nb. d'échantillons possibles		Nb.d'échantillons enquêtés		tx sondage
Ensemble	M =	11 082	m =	91	0,82%
Strate n°1 HCL	M ₁ =	5 394	m ₁ =	44	0,82%
Strate n°2 GFS	M ₂ =	5 688	m ₂ =	47	0,83%

Lorsque les statistiques sont complexes [ici, elles se présentent sous la forme d'une fonction de plusieurs totaux (total des arbres enquêtés et touchés, total des arbres enquêtés)], l'estimation de la variance de l'estimateur du ratio P_s est simplifiée en passant par l'estimation d'une variable intermédiaire linéarisée. Celle-ci (Z), pour chaque placette « i » de l'échantillon prend la forme suivant (cf. formule n°2) :

$$Z_i = \frac{1}{\hat{X}_2} * X_{i1} - \frac{\hat{X}_1}{\hat{X}_2^2} * X_{i2}$$

formule n°2 (Clairin, 1997, p.90)

Où X_{i1} est le nombre d'arbres touchés par placette et X_{i2} le nombre d'arbres total par placette. La variance de l'estimation du total de Z (cf. formule n°3) fournit une estimation de la variance de l'estimation du ratio (total des arbres touchés sur total des arbres).

$$V(\hat{T}(Z)) = \sum_{h=1}^k M_h^2 * \left(1 - \frac{m_h}{M_h}\right) * \frac{1}{m_h} * \frac{1}{m_h - 1} * \left(\sum_{i_h=1}^{m_h} (Z_{i_h} - \bar{Z}_h)^2\right)$$

formule n°3

Cette formule d'estimation de la variance d'une variable stratifiée revêt une forme courante si ce n'est le calcul de la somme des écarts au carré entre la valeur de la variable linéarisée placette par placette « i » puis strate par strate « h » avec la moyenne de cette même variable linéarisée calculée strate par strate (cf. résultat n°2).

$$V(\hat{T}(Z)) = \left[5\,394^2 * \left(1 - \frac{44}{5\,394}\right) * \frac{1}{44} * \frac{1}{44-1} * 5,87 * 10^{-8}\right] + \left[5\,688^2 * \left(1 - \frac{47}{5\,688}\right) * \frac{1}{47} * \frac{1}{47-1} * 1,82 * 10^{-8}\right]$$

résultat n°2

Soit :

$$V(\hat{T}(Z)) = 1,2 * 10^{-3}$$

résultat n°3

Dans la mesure où le tirage est uniforme sur les deux strates, le calcul de la variance de l'estimateur v(p) sur l'ensemble de la zone d'étude, considérée comme une strate unique (donc sans stratification) s'effectue à l'aide des formules n°1, 2 et 3 en remplaçant M_h et m_h (paramètres sur les strates 1 et 2) par M et m (paramètres sur l'ensemble, cf. Tableau n°6). Les résultats sont comparés dans le tableau n°4 (cf. 2.1.2).

A.3 - Formulaire pour la stratification à taux de sondage variables

Dans l'exemple bellifontain, deux strates (k) ont également été retenues. L'une correspond à « h = 1 » pour les parties en réserve (RBI), l'autre à « h = 2 » pour les parties hors réserve (HR). Leurs valeurs sont reprises dans le tableau des paramètres (cf. Tableau n°7). On obtient donc une proportion de dégâts estimée sur les échantillons de :

$$P_3 = \frac{(22,96 * 306) + (163,48 * 67)}{(22,96 * 1\ 682) + (163,48 * 868)} = 10\%$$

résultat n°4

On remarque qu'il y a beaucoup plus d'arbres enquêtés et touchés en RBI que hors réserve (respectivement 306 contre 67, soit 4,6 fois plus) alors qu'ils ne sont que 2 fois plus nombreux à être enquêtés (1 682 contre 868). On remarquera également, que les coefficients d'extrapolation varient dans un rapport de un à sept. Ce que nous n'avions, bien évidemment pas pour Conches-Breteil où les taux de sondage étaient identiques sur les deux strates.

$$v(\hat{T}|Z_i) = \left[528^2 * \left(1 - \frac{23}{528} \right) * \frac{1}{23} * \frac{1}{23-1} * 1,51 * 10^{-7} \right] + \left[1\ 962^2 * \left(1 - \frac{12}{1\ 962} \right) * \frac{1}{12} * \frac{1}{12-1} * 1,85 * 10^{-8} \right]$$

résultat n°5

Soit :

$$v(\hat{T}|Z_i) = 6,2 * 10^{-4}$$

résultat n°6

Tableau 7 – Paramètres du plan d'échantillonnage pour le secteur de Bourron-Marlotte (S.-et-M.)

	Nb. d'échantillons possibles		Nb.d'échantillons enquêtés		tx sondage
Ensembles	M =	2 490	m =	35	1,4%
Strate n°1 RBI	M ₁ =	528	m ₁ =	23	4,4%
Strate n°2 HR	M ₂ =	1 962	m ₂ =	12	0,6%