

# Méthodologie

## L'utilisation du territoire en 2014

### Teruti-Lucas

*Les prémices de l'enquête ont vu le jour en 1946 avec le contrôle des surfaces agricoles basé sur les plans cadastraux. Par la suite, les différents progrès techniques ont conduit à étendre l'enquête au suivi de l'ensemble du territoire agricole et forestier mais aussi naturel et artificialisé. En 2005, les points de la nouvelle enquête Teruti-Lucas ont été géoréférencés. Cette amélioration technique a permis de rapprocher l'échantillon d'autres sources de données géoréférencées. Ainsi, depuis 2012, les points Teruti-Lucas situés en territoire agricole sont renseignés à partir des déclarations de surfaces des demandeurs d'aides PAC.*

#### **De Teruti à Teruti-Lucas**

La connaissance et le suivi de l'occupation du territoire sont des préoccupations anciennes de la statistique agricole. La première enquête sur l'utilisation du territoire agricole remonte à l'année 1946 avec la mise en place d'une enquête « contrôle de surface » reposant sur des contrôles en profondeur de surface et d'occupation du sol à partir de plans cadastraux. En 1962, la photographie aérienne est introduite, non comme support de l'enquête mais comme outil d'actualisation des plans cadastraux. À partir de 1969, l'enquête par photographie aérienne et sondage par points sur le terrain se généralise à l'ensemble des départements.

L'enquête Teruti repose alors sur l'association originale de photographies aériennes constituant la base de sondage et de relevés de terrain effectués par des enquêteurs. À partir de 1982, elle bénéficie d'un atout supplémentaire avec la mise en place d'un échantillon national obligatoire qui permet de stabiliser le système et d'étendre le champ d'analyse, jusque-là orienté vers l'espace agricole, à l'ensemble du territoire.

L'échantillon permanent est renouvelé en 1990 et 1991, à l'exception de neuf départements qui ont conservé l'ancien échantillon, afin de corriger certains biais de représentativité introduits lors du tirage

## Méthodologie

de l'échantillon de 1982. En effet, dans ce dernier, les surfaces agricoles étaient sur-représentées alors que les sols boisés étaient sous-estimés. L'enquête Teruti est réalisée de manière continue jusqu'en 2004.

Deux raisons principales conduisent à faire évoluer l'enquête en 2005. En premier lieu, les progrès techniques en matière de numérisation des documents cartographiques et de géoréférencement des points incitent à redéfinir l'échantillonnage des points Teruti. La localisation de ces derniers, déterminée de façon manuelle sur des photographies aériennes, était en effet difficile à suivre précisément dans le temps, ce qui générait des erreurs d'observation. En second lieu, la perspective de la mise en place de l'enquête européenne Lucas (*Land Use/Cover Area frame statistical Survey*) fondée sur les mêmes principes que Teruti (enquête aréolaire à deux degrés de tirage) incite à procéder à une refonte en profondeur de Teruti afin d'assurer une cohérence sur la nomenclature, la méthode d'observation et la conception de l'échantillon.

L'enquête Teruti évolue donc vers une nouvelle enquête Teruti-Lucas visant à répondre aux nouvelles demandes tout en assurant au maximum la continuité avec Teruti. Le géoréférencement permet ainsi une localisation plus précise et constante des points désormais localisés sur le Scan 25 de l'IGN et sur des orthophotographies. Il permet aussi au statisticien de définir tout zonage « à la demande » au moyen d'un système d'informations géographiques et de faire des recoupements avec toute autre base de données géoréférencée. En outre, la modularité

de l'enquête permet des traitements « flexibles » adaptés aux besoins des différents utilisateurs. Rappelons toutefois que la définition d'un zonage à la demande doit respecter la contrainte du niveau de représentativité de l'enquête et que le respect du secret statistique interdit la communication de données individuelles identifiables.

Depuis 2005, l'enquête n'a pas connu de modifications majeures, sauf l'utilisation depuis 2012 des déclarations de surfaces des demandeurs d'aide PAC pour renseigner les points situés en sols agricoles.

### Présentation de l'enquête Teruti-Lucas

#### Finalités de l'enquête Teruti-Lucas

L'enquête Teruti-Lucas poursuit quatre objectifs :

- connaître annuellement les différentes catégories d'occupation et d'usage de l'ensemble du territoire (agricole, naturel et urbanisé), au niveau national, régional et départemental mais aussi à tout autre niveau défini géographiquement,
- suivre et quantifier les changements d'occupation et d'usage des sols et de la structure des paysages au fil du temps,
- constituer une source de données pour d'autres analyses et études, par exemple l'élaboration d'indicateurs agri-environnementaux ou d'indicateurs paysagers,
- dans le cadre d'applications locales ou nationales, pouvoir étendre l'échantillon aussi bien au niveau du nombre de segments que du nombre de points par segments.

### Méthodologie détaillée

Teruti-Lucas reprend les fonctionnalités de Teruti et s'en adjuge quelques autres.

La population étudiée s'étend à la presque totalité du territoire national. Les deux seules restrictions concernent :

- la Guyane où seule la partie cartographiée doit être enquêtée, sur la bande côtière essentiellement (pour des raisons techniques, l'enquête n'a pas encore été réalisée en Guyane).
- Mayotte pour qui la méthode s'adapte mal à la faible taille de son territoire.

C'est une enquête par sondage aréolaire à deux niveaux de tirage, segments et points, ces derniers étant regroupés en grappes (segments) pour rentabiliser les déplacements :

- le segment est le premier niveau de tirage. Il s'agit d'une portion de territoire dont la taille peut varier de 1,5 km sur 600 m à 1,5 km sur 1,5 km ;
- le point est le second niveau de tirage. Il s'agit également d'une portion de territoire, un cercle de 3 m de diamètre (fenêtre d'observation de base) dans le cas général ou 40 m de diamètre (fenêtre d'observation étendue) dans le cas des occupations hétérogènes. Les points sont espacés de 300 m à l'intérieur d'un segment.

Un échantillon à deux niveaux de tirage est un compromis entre le coût de l'enquête et la précision souhaitée pour les résultats. On obtient ainsi en général une moins bonne précision qu'avec un échantillon à un seul niveau de tirage.

Mais ce dernier a un coût supérieur du fait de l'éloignement de chaque lieu d'observation.

## Tirage d'échantillon Échantillon maître

L'échantillon maître a été constitué par intensification du sondage proposé par Eurostat (Direction Générale des Statistiques de la Commission Européenne) dans le cadre des enquêtes Lucas 2001 et 2003 où les segments sont espacés de 18 km. Il inclut donc les points visités lors de ces deux enquêtes européennes.

Une grille systématique, régulière et carrée orientée N-S et E-W de maille 3 km sur 3 km sert de base au tirage d'échantillon. Chaque intersection détermine un segment, portion de territoire carrée de 1,5 km de côté (225 ha). En fait, l'intersection détermine le point 11 de chaque segment, qui se construit à partir de ce point comme indiqué ci-après. Chaque segment est identifié par sa position en ligne et colonne (III-ccc) dans le repère normé dont l'origine est à l'extrême sud-ouest de chaque territoire.

Le tirage d'échantillon a été effectué selon les prescriptions d'Eurostat sur la base de la projection UTM (ellipsoïde GRS80) qui entraîne le moins de déformation, en France métropolitaine UTM31N, aux Antilles UTM20N, en Guyane UTM22N et à La Réunion UTM40S. Il est systématique et sans stratification préalable. En Martinique, Guadeloupe et à La Réunion, la grille de base est de 2 km sur 2 km afin d'obtenir la précision voulue pour chacun des départements pris individuellement. En France métropolitaine, les coordonnées ont été transformées en Lambert II étendu (Lambert IV en Corse) pour permettre la superposition aux supports IGN (Scan 25

et BD Ortho). La transformation des coordonnées métropolitaines a également été calculée afin d'être prêts à utiliser le système géodésique RGF93 devenu la norme légale en France.

## Les segments

L'échantillon de segments a été découpé en 4 sous échantillons inclus les uns dans les autres :

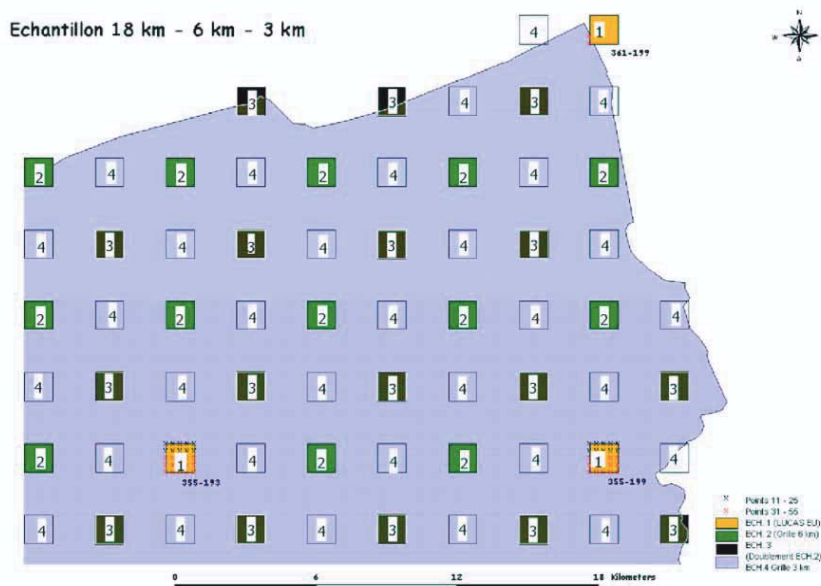
- « 4 » est l'échantillon déterminé par la grille de base la plus fine (3 ou 2 km) ;
- « 1 » est la partie de « 4 » correspondant à la grille Lucas de 18 km ;
- « 2 » est la partie de « 4 », complément de « 1 » pour constituer la grille 6 km ;

- « 3 », appelé « 6 km doublé », correspond au doublement du nombre de segments de « 2 » + « 1 » en quinconces ; « 3 » n'existe ni aux Antilles ni à La Réunion.

Chaque segment se voit affecter le code de l'échantillon le plus « petit » auquel il appartient. Si par exemple on décide de lancer l'enquête sur le sous échantillon « 2 » on devra sélectionner les segments affectés de « 1 », « 2 » ou « 3 » et pour le sous échantillon « 4 » la totalité des segments.

## Nombre de segments par sous-échantillon

Sous-échantillon	1 (18 km)	1+2 (6 km)	1+2+3 (6 km doublé)	1+2+3+4 (maître 3 km)
France	1 764	15 896	31 655	64 492



## Méthodologie

### Les points

Chaque segment contient 5 lignes de 5 points orientés comme les segments N-S, E-W et espacés régulièrement de 300 m.

Les deux premières lignes (points 11 à 25) correspondent aux points des enquêtes Lucas de 2001 et 2003, les trois suivantes ont été ajoutées pour répondre à d'éventuelles extensions.

○	○	○	○	○
11	12	13	14	15
○	○	○	○	○
21	22	23	24	25
●	●	●	●	●
31	32	33	34	35
●	●	●	●	●
41	42	43	44	45
●	●	●	●	●
51	52	53	54	55

### Échantillon national

L'objectif de l'enquête nationale annuelle est de visiter les segments de l'échantillon « 6 km doublé », c'est-à-dire les sous-échantillons 1+2+3 partout, +4 pour 75, 90, 92, 93, 94, 9A, 9B, 9D. Il comprend **32 498** segments de 10 points (11 à 25).

C'est moins de points mais deux fois plus de segments qu'avec Teruti. Le meilleur étalement spatial de l'échantillon permet d'obtenir des résultats fiables jusqu'au niveau départemental pour les occupations du sol les plus importantes et/ou les plus homogènes dans leur répartition.

Cet objectif a été atteint de 2006 à 2010 et de 2012 à 2014, c'est la raison pour laquelle cet échantillon est désigné sous le nom « échantillon

complet ». En 2010, il comprend 322 191 points compte tenu du fait que les segments situés en limite du territoire sont incomplets. En 2005, sous l'effet de contraintes budgétaires, l'enquête n'a pu être réalisée que sur l'échantillon 1+2 partout, +4 pour les départements 75, 90, 92, 93, 94, 9A, 9B, 9D. Cet échantillon est désigné sous le nom « échantillon restreint » ; il comprend 17 598 segments de 10 points (11 à 25) ce qui, pour la même raison que ci-dessus, donne 171 620 points.

En raison de la non-disponibilité de la Bd Ortho, l'enquête n'a pu être réalisée en 2005 en Guadeloupe, et n'est pas réalisée en Guyane.

L'enquête n'a pas été réalisée en 2011 (année de collecte du recensement agricole).

Depuis 2012, les déclarations de surface des demandeurs d'aides PAC sont utilisées pour renseigner les points situés en sol agricole. Le croisement de la couche géographique des points Teruti-Lucas avec le Registre Parcellaire Graphique permet de déterminer les points situés dans un îlot PAC. En 2014, 107 380 points ont ainsi été imputés, soit 33 % de l'échantillon complet.

### Extensions locales

Une des caractéristiques essentielles de Teruti-Lucas est sa modularité. Il peut être utilisé comme base de sondage pour organiser une enquête disjointe du passage annuel. Il permet aussi d'envisager des renforcements locaux ou nationaux sous réserve de financement bien sûr. Ceux-ci peuvent prendre deux formes utilisables

séparément ou conjointement :

- augmenter le nombre de segments
- augmenter le nombre de points visités dans les segments

Cette possibilité permet de répondre à des besoins ciblés qu'ils émanent de demandeurs nationaux ou locaux. La zone concernée par une extension peut être déterminée en fonction des besoins du demandeur sans être limitée par les frontières administratives. La seule contrainte est de sélectionner un nombre de points suffisant pour obtenir des résultats fiables. Pour ce faire, la **Base de Données SPOT** a été constituée.

### BdSPOT

Il s'agit d'une base de données contenant tous les segments et points de l'échantillon maître. Elle contient les principales informations relatives aux segments et aux points et est actualisée chaque année après validation de l'enquête.

### Extrapolation des résultats et précision des estimations Cas général

Le coefficient d'extrapolation ou valeur (V) du point Teruti-Lucas est calculé par le rapport entre la superficie (S) planimétrée par l'IGN de chaque département et le nombre (N) de points dans le département. Chaque catégorie h est extrapolée de la manière suivante :

$$\hat{S}_h = V \times N_h$$

$\hat{S}_h$  : superficie de la catégorie h

$N_h$  : nombre de points de la catégorie h

Un point Teruti-Lucas représente approximativement la surface suivante :

- 94 ha pour Paris, les Hauts-de-Seine, la Seine-Saint-Denis, le Val-de-Marne et le Territoire-de-Belfort,
- 178 ha pour les autres départements métropolitains de 2006 à 2014 et 349 ha en 2005,
- 40 ha pour la Guadeloupe, la Martinique et La Réunion.

La formule appliquée pour calculer l'erreur aléatoire sur l'estimation tient compte des deux niveaux de tirage : le tirage systématique, sans remise et équiprobable des segments et le tirage systématique, sans remise et équiprobable des points à l'intérieur d'un même segment.

La **variance** totale  $V_h$  sur l'estimation de surface d'une catégorie  $h$  de territoire s'exprime par la somme de deux sources d'erreur : la variance observée au sein de l'échantillon entre deux segments et la variance observée dans un même segment entre les différents points qui le composent.

À partir de la variance, on calcule l'**écart-type** de la superficie estimée  $\sigma_h$ .

La **précision** exprimée en % s'écrit :

$$\frac{\sigma_h}{\hat{S}_h} \times 100$$

La détermination de la précision des estimations de surface à divers niveaux (France entière, région, département, petite région agricole, zonage spécifique) met en jeu des calculs complexes et implique des traitements informatiques particuliers.

La construction d'un **intervalle de confiance** centré sur la valeur estimée  $\hat{S}_h$  permet de déterminer la plage de valeurs dans laquelle la valeur vraie est comprise avec une certaine probabilité. Au seuil de probabilité de 95 % (95 chances sur 100 que la valeur vraie  $S_h$  soit comprise entre les deux bornes), l'intervalle de confiance est déterminé par la relation suivante :

$$S_h = \hat{S}_h \pm 1,96\sigma_h$$

### Points imputés à partir des données PAC

Depuis 2012, les points situés en sols agricoles sont imputés, dans la mesure du possible, à partir des déclarations PAC.

L'extrapolation des résultats dépend du type d'îlot dans lequel est situé le point :

- l'îlot contient une seule culture : le coefficient d'extrapolation ( $V$ ) du point est affecté en totalité à la culture  $h$ .
- l'îlot contient plusieurs cultures : la structure des données PAC ne permet pas de savoir dans quelle parcelle est situé le point. En effet les exploitants déclarent d'une part le contour des îlots composant leur exploitation (Registre Parcellaire Graphique), et d'autre part la liste (et la surface) des cultures implantées dans chaque îlot, mais l'implantation géographique des parcelles au sein de l'îlot n'est pas connue. Dans ce cas, le choix a été fait d'attribuer au point Teruti-Lucas autant d'occupations que de cultures dans l'îlot. Le coefficient d'extrapolation du point est alors partagé entre les différentes occupations, au prorata de leur surface respective dans l'îlot.

Par exemple, un point situé dans un îlot contenant une parcelle en blé,

une parcelle en maïs et une parcelle en prairie, se voit attribuer trois occupations et son coefficient d'extrapolation se décompose de la façon suivante :

La formule permettant de calculer

$$V = V_{blé} + V_{maïs} + V_{prairie}$$

$$= V \times \frac{S_{blé}}{S_{îlot}} + V \times \frac{S_{maïs}}{S_{îlot}} + V \times \frac{S_{prairie}}{S_{îlot}}$$

- $S_{blé}$  : surface de la parcelle en blé
- $S_{maïs}$  : surface de la parcelle en maïs
- $S_{prairie}$  : surface de la parcelle en prairie
- $S_{îlot}$  : surface totale de l'îlot

la précision des estimations peut être appliquée uniquement si les points sont associés à une seule occupation. Ainsi, pour les catégories agricoles, les précisions ne sont calculées que sur l'agrégat « sols cultivés et toujours en herbe ».

### Résultats

Pour la France métropolitaine, le niveau de précision des estimations de surface est satisfaisant puisque les postes de la nomenclature d'occupation de synthèse qui enregistrent une erreur relative inférieure à 5 % couvrent 99 % du territoire métropolitain. Les postes à plus de 5 % d'erreur sont ceux de moindre importance qui couvrent moins de 130 000 ha chacun. Il convient toutefois de préciser qu'il ne s'agit que de l'aléa lié à l'observation d'un échantillonnage de points. D'autres sources d'erreur existent, principalement des erreurs d'observation sur le terrain, pouvant affecter certains types d'occupation plus difficiles à déterminer par l'enquêteur (par exemple la distinction entre prairies permanentes et temporaires).

### Précision des estimations de surfaces en 2014 pour certains postes de la nomenclature d'occupation de synthèse

Nature	Surface (1 000 ha)	Précision (%)
Sols cultivés et surfaces toujours en herbe	28 029,9	0,3
Forêts de feuillus	9 480,4	0,7
Sols de forme linéaire revêtus ou stabilisés	1 632,1	1,1
Forêts de résineux	3 211,9	1,2
Landes, friches, maquis, garrigues, savanes	2 509,0	1,2
Sols enherbés artificialisés	1 590,9	1,3
Volumes construits bas	862,1	1,7
Sols de forme aréolaire revêtus ou stabilisés	853,2	1,8
Superficies enherbées naturelles	383,6	2,6
Sols nus artificialisés	144,7	4,6
Dunes, plages	42,6	9,0
Glaciers, neiges éternelles	35,0	14,3

#### Le questionnaire

Il comporte trois parties :

#### Les informations relatives aux conditions d'observation du point

- mode d'observation du point.
- distance entre l'enquêteur et le point observé.
- diamètre de la fenêtre d'observation du point.
- direction observée en cas de point limite.

#### L'occupation et l'usage

L'occupation ou couvert (utilisation physique) et l'usage ou utilisation (utilisation fonctionnelle) étaient les informations relevées par Teruti-Lucas. Elles restent au centre de Teruti-Lucas.

En chaque point d'observation, deux codes occupation et deux codes usage peuvent être saisis : un premier code obligatoire dans tous les cas, un deuxième à saisir lorsqu'il y a deux occupations ou deux usages.

Pour les points imputés à partir des déclarations de surfaces PAC, l'usage attribué est « agriculture », car les données administratives ne permettent pas de distinguer les usages « production végétale » et « élevage ».

#### Les informations environnementales

Elles sont toujours observées dans la fenêtre étendue, c'est-à-dire dans le cercle de 40 m de diamètre :

- coupe ou plantation en forêt (coupe rase, coupe partielle ou plantation).

La présence d'arbres isolés (de un à trois arbres), et les semis de prairies temporaires ont été observés sur l'ensemble du territoire jusqu'en 2010. Ces informations n'étant pas disponibles pour les points renseignés avec les données PAC depuis 2012, elles ne sont plus collectées.

#### Les nomenclatures

Les principes de base qui ont guidé le SSP pour l'établissement des

nomenclatures de collecte sont les suivants :

- la compatibilité avec la nomenclature des enquêtes Lucas 2001 et 2003 : on part de celle-ci et on la subdivise en fonction de nos besoins,
- l'harmonisation de la nomenclature forestière (classe 3 de Teruti-Lucas) avec celle de l'Inventaire forestier national et avec celle de la FAO (FRA 2005),
- l'harmonisation, dans la mesure du possible, avec les autres nomenclatures utilisées par le SSP et par Eurostat : Statistique agricole annuelle, enquête Structures, ZPA1, etc.
- l'harmonisation, dans la mesure du possible, de la nomenclature d'usage avec la nomenclature d'activités française (Naf rev 2),
- l'inclusion des occupations des départements d'outre mer dans une nomenclature française.

L'application de ces principes se traduit par une nomenclature de collecte comprenant 4 niveaux et

122 postes pour l'occupation, 2 niveaux et 38 postes pour l'usage. Un niveau supplémentaire facultatif est prévu pour un code local. À partir de 2012, la nomenclature d'occupation est aménagée pour la rendre compatible avec celle des déclarations de surfaces : la nouvelle nomenclature d'occupation qui permet de passer des codes cultures PAC aux codes Teruti-Lucas comprend 88 postes. La nomenclature d'usage, elle aussi aménagée, comporte désormais 36 postes.

Pour assurer au maximum le maintien de la comparabilité des résultats de Teruti-Lucas avec Teruti, une nomenclature d'occupation de synthèse combinant occupation et usage avait été créée en 2005. Après adaptation aux données PAC, la nouvelle nomenclature d'occupation de synthèse comporte 50 postes.

Les résultats publiés (tableaux présentés sur le site Agreste et cubes Disar) utilisent cette nouvelle nomenclature de synthèse. Compte tenu de la méthode d'imputation des points en sols agricoles qui conduit dans certains cas à attribuer plusieurs cultures à un même point de l'échantillon, les postes agricoles sont regroupés en un seul poste « sols cultivés et STH » dans

les matrices de changement d'occupation du sol entre deux dates. Des nomenclatures agrégées ont également été créées lorsque la précision sur certains postes de la nomenclature détaillée est insuffisante.

### Collecte terrain

Un segment est visité par un enquêteur du Service régional de l'information statistique et économique correspondant à la localisation départementale du segment 25 points (affectation à la majorité de la surface). Après localisation du point sur le terrain, l'enquêteur attribue à la fenêtre d'observation l'occupation et l'usage qu'il observe en référence aux nomenclatures correspondantes. Les points situés en altitude, en forêt primaire, en zone interdite ou inaccessible peuvent être photo-interprétés.

### Plan d'assurance qualité

Le passage de Teruti à Teruti-Lucas a donné l'occasion de mettre en place un véritable plan d'assurance qualité inspiré de Lucas. Ses principes sont les suivants :

- le recrutement, la formation et l'organisation du travail des enquêteurs restent gérés au niveau régional. Chaque département comprend ainsi une équipe de 10 à 15 enquêteurs.

Les documents de formation et les instructions aux enquêteurs restent préparés au niveau national (SSP) qui recourt le cas échéant à des compétences particulières (IFN, ONF).

Des contrôles de terrain « en double aveugle » sont organisés sur 5 % à 10 % des segments. Contrairement à Teruti, les contrôles sont saisis et conservés dans la base de données, ce qui permet la quantification des divergences.

Des contrôles additionnels orientés sont réalisés en cas d'évolution aberrante de l'occupation ou de l'usage.

La possibilité de modifier la précédente observation lors de l'observation en cours est ouverte. Il peut en résulter un écart entre deux publications successives des données d'une même année.

À partir de 2009, la saisie directe des questionnaires par les enquêteurs sur tablet-Pc à l'aide d'un logiciel de saisie-contrôle adapté et la télétransmission sécurisée des données sur un serveur central participent au processus d'amélioration continue de la qualité des résultats.