MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION

Liberté
Égalité
Fraternité

Refonte de la base de sondage des exploitations agricoles : vers une utilisation accrue des fichiers administratifs

Thierry GUILLAUME (*)

(*) Service de la Statistique et de la Prospective, Bureau des Méthodes et de l'Information Statistique

JMS 2022

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Plan de la présentation

Définition de l'exploitation agricole

Quelles variables et quels sources utilisées dans la base de sondage ?

La nécessité d'une nouvelle base de sondage

Le programme historique d'appariement

Validation des appariements

Appartenance au champ de base

Un objectif : la mise en place d'une démographie des exploitations agricoles

Conclusions et perspectives

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Définition de l'exploitation agricole

Une exploitation agricole est définie comme une unité économique et de production répondant simultanément aux trois conditions suivantes :

1. avoir une activité agricole,

2. atteindre ou dépasser une certaine dimension (en pratique, atteindre au moins un seuil minimal de surfaces cultivées, d'animaux ou de production),

3. être soumise à une gestion courante indépendante (l'existence d'un Siret est considérée comme une présomption suffisante d'autonomie).

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Quelles variables dans la base de sondage?

La base de sondage, dite BALSA, contient plusieurs types de variables :

- d'identification : identifiant de l'exploitation agricole, raison sociale, Siret (si celui-ci existe et est connu), identifiants administratifs, adresse de l'établissement,
- de contacts : nom et prénom des répondants aux enquêtes, adresses de contact, téléphones et mails,
- des variables servant à la stratification des enquêtes du SSP: surfaces des cultures, effectifs animaux, effectifs salariés, production brute standard (PBS) qui est un proxy du « chiffre d'affaires » et l'orientation technico-économique (OTEX) qui se rapproche conceptuellement de la notion d'APE.

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Quelles sources pour alimenter la base?

En complément des fichiers Sirene, la base est alimentée avec les fichiers administratifs ayant un rapport avec l'activité agricole :

- des déclarations (surfaces de cultures et effectifs bovins, ovins et caprins) dans le cadre de la PAC,
- du casier viticole informatisé (CVI) sur les surfaces de vignes, des mouvements et des détenteurs d'animaux (bovins, ovins, caprins et porcins) provenant des bases de données de la Direction Générale de l'Alimentation (DGAL),
- de la mutualité sociale agricole (MSA) sur les salariés et les cotisants,
- des fichiers de l'AgenceBio et de l'institut national de l'origine et de la qualité (INAO),
- des fichiers de bénéfices et micro-bénéfices agricoles,
- des fichiers sur des activités spécifiques tels l'horticulture, l'apiculture, les équidés,…

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

La nécessité d'une nouvelle base de sondage

- Une base, initialisée à partir du recensement agricole 2010, alimentée d'abord par les enquêtes du SSP
- Une mise à jour depuis 2018 par Sirene et l'ensemble des fichiers administratifs « agricoles » disponibles
- Des opérations de qualité pour mettre à jour la base pour préparer le recensement agricole 2020 mais avec une approche « prudente » sur la cessation d'unités
- Un recensement agricole 2020 dont les résultats vont permettre notamment d'affiner les approches sur l'intégration ou la cessation d'unités

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

La nécessité d'une nouvelle base de sondage

L'objectif de la refonte est :

- de permettre des mises à jour automatisées à partir des données administratives, de façon régulière (dans la mesure de ce que les sources permettent de faire)
- améliorer la qualité des données
- diminuer la charge de réponse des répondants par la prise en compte plus systématique des données administratives
- de réduire la charge de travail de l'administrateur de la base qui intègre aujourd'hui ces données via des traitements manuels.

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Les modalités de mise en place de la nouvelle base de sondage

Utilisation du recensement agricole 2020 (initialisation de la base à partir des données du recensement) et mise à jour avec les fichiers administratifs « agricoles » et les enquêtes du SSP

La mise à jour de la base à partir des fichiers administratifs présente un certain nombre de limites dans l'appariement des fichiers:

- le Siren ou Siret n'est pas utilisé comme identifiant principal dans l'ensemble des fichiers administratifs « agricoles »

- des fichiers administratifs « agricoles » avec des taux de Siren (ou Siret) variant de 25 % à 98 %.

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Les appariements

2 étapes:

1. Appariement Fichier administratif / SIRENE (sirétisation)

2. Appariement Fichier administratif / Balsa

Le programme historique d'appariement

Utilisation des caractéristiques du fichier Sirene pour la normalisation.

Découpage de l'adresse du fichier administratif en plusieurs variables qui sont le numéro de voie, le type de voie, le libellé de voie et l'indice de répétition

Nettoyage et normalisation des données (par exemple: les caractères spéciaux, les accents, les chiffres (sauf pour le numéro de voie) sont exclus)

adresse	numeroVoieEtablissement	typeVoieEtablissement	libelleVoieEtablissement
35 HAMEAU DE RIBEAUVILLE	35	HAM	RIBEAUVILLE
1 RUE DE CHARLES BRUYERES	1	RUE	CHARLES BRUYERES
3 ROUTE DE WALLERS	3	RTE	WALLERS
11 CARRIERE ETREUX EN BAS	11		CARRIERE ETREUX BAS
11 LA HAIE LONGPRE	11		HAIE LONGPRE

Le programme historique d'appariement

Processus basé sur une démarche séquentielle de relâchement successif de clés de moins en moins robustes…. :

- Adresse complète (numéro de voie, type de voie, libellé de voie) + code commune,
- Nom complet (les prénoms et nom de famille) + code commune,
- Adresse complète sauf le numéro de voie + code commune,
- Premier prénom et le nom de famille + code commune,
- Premier prénom et le nom de famille inversé + code commune,
- Adresse sauf le type de voie + code commune.

…. et en deux temps:

- Sur unités avec une APET « agricole »
- Sur unités avec une APET non « agricole »

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Les appariements envisagées (ou en cours)

Méthodes d'appariement : séquentiel vs probabiliste

Services testés, ou à l'étude:

- Le service d'identification automatique de masse (SIAM) de Sirene4 (Insee)
- Un outil (« Random forest ») développé par l'Observatoire du Développement Rural de l'INRAE
- matchID
- Rapsodie
- Moteur d'appariement d'InserJeunes

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Validation des appariements

Sélection automatique d'un appariement à partir de plusieurs échos

Expertise humaine nécessaire lorsque plusieurs Siret ont été appariés

Principal problème: la volumétrie des validations par les gestionnaires (par exemple, le fichier annuel de la PAC : 5 000 à 8 000 unités et celui des micro-bénéfices agricoles: 30 000 unités)

Jusqu'à présent, le traitement manuel est effectué uniquement sur les fichiers considérés comme stratégiques en appliquant des seuils (pour se focaliser sur les exploitations les plus importantes)

Pour réduire la volumétrie, des réflexions en cours pour utiliser les résultats non pas d'un seul fichier mais de plusieurs fichiers administratifs.

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Appartenance au champ de la base

Deux conditions à remplir: **« actif » ET « être une exploitation agricole »**

Avant le RA2020, une exploitation agricole était active si :

- elle était présente dans un fichier administratif « agricole » récent
- elle disposait d'un Siret « actif » dans Sirene avec une APET « agricole ».

Cependant, des règles de dérogation (absence de Siret, APET « non agricole ») si fichier administratif « stratégique ».

Les règles ont évolué au cours de la précédente décennie pour éviter un défaut de couverture dans la perspective du RA 2020.

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Appartenance au champ de la base

Des approches supervisées en cours pour définir si une unité est (ou non) une exploitation agricole.

Première idée : un modèle unique utilisant la présence ou non dans chaque fichier administratif

Deuxième idée : plusieurs modèles (un par source) en fonction des variables de chaque fichier administratif puis de pondérer ces probabilités de présence en fonction de la qualité de la source (une réflexion sur les règles de pondération est en cours)

Des approches à affiner, notamment car les motifs d'être hors champ au recensement agricole sont multiples (cessations, auto-consommation, doublons, autres raisons,…)

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

Un objectif : la mise en place d'une démographie des exploitations agricoles

Démarche exploratoire entre les recensements de 2010 et de 2020

<u>Approche:</u>

- Partir du stock d'une année N
- Enlever les unités qui ont cessé dans Sirene cette année N
- Ajouter dans Sirene en amont les « vraies » créations du domaine agricole (créations donnant suite à une activité économique réelle c'est-à-dire se retrouvant dans un fichier administratif agricole)
- Retirer en aval ce qui semble être des « faux-actifs » (sortants d'un fichier administratif).

<u>Résultats:</u>

- Une baisse continue des exploitations agricoles, cohérente avec celle observée dans la réalité
- Mais suivi difficile des Siret (une part non négligeable des Siret ne se retrouvent pas au recensement de 2020)

Conclusions et perspectives

Mise à jour plus efficace et régulière de la base de sondage

Diminution de l'intégration de nouvelles unités sans réalité économique (inscription dans un fichier sans activité réelle par la suite)

Augmentation des cessations d'unités (réduire le maintien des « faux-actifs »)

Réaliser à court terme une approche rigoureuse de la démographie d'exploitations et d'entreprises agricoles qui soit comparable aux travaux de l'Insee sur le champ de l'industrie, du commerce et des services