



Organisation des Nations Unies  
pour l'alimentation  
et l'agriculture



**Indicateur 2.4.1 des ODD :  
lignes directrices  
en matière d'échantillonnage**

**11/08/2021**

## Table des matières

<b>Remerciements</b> .....	3
<b>Introduction</b> .....	4
<b>1 Plan d'échantillonnage de l'enquête consacrée à l'indicateur 2.4.1 des ODD</b> .....	5
<b>1.1 Unités d'observation et population cible</b> .....	5
<b>1.2 Unités d'échantillonnage et base de sondage</b> .....	5
<b>1.3 Unités déclarantes</b> .....	6
<b>1.4 Domaines d'estimation</b> .....	6
<b>1.5 Plan d'échantillonnage et estimateurs</b> .....	6
<b>1.6 Stratification</b> .....	6
<b>1.6.1 Échantillonnage à deux degrés aux fins de stratification</b> .....	7
<b>1.6.2 Post-stratification et calage</b> .....	8
<b>1.7 Taille de l'échantillon</b> .....	8
<b>1.8 Répartition de l'échantillon en strates</b> .....	10
<b>1.9 Absence de base</b> .....	11
<b>2 Collecte de données sur l'ODD 2.4.1 à partir d'une enquête existante</b> .....	12
<b>2.1 Couverture des données</b> .....	12
<b>2.2 Cas des enquêtes agricoles</b> .....	12
<b>2.2.1 Problèmes liés aux unités d'observation et d'échantillonnage</b> .....	12
<b>2.2.2 Taille de l'échantillon</b> .....	13
<b>2.3 Cas des enquêtes non agricoles</b> .....	13
<b>2.3.1 Utilisation d'un échantillon complémentaire</b> .....	13
<b>2.3.2 Échantillonnage indirect</b> .....	14
<b>Conclusion</b> .....	14
<b>Références bibliographiques</b> .....	15

## Remerciements

Ce document, qui a vocation à accompagner la note méthodologique relative à l'indicateur 2.4.1 des ODD, traite plus particulièrement des aspects relatifs au plan d'échantillonnage effectué par l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) en collaboration avec la Stratégie mondiale pour l'amélioration des statistiques agricoles et rurales (GSARS).

L'auteur, Dramane Bako, souhaite remercier toutes les personnes ayant contribué aux différentes sections de ce document, et plus particulièrement Arbab Asfandiyar Khan de la Division de la statistique de la FAO et Amy Heyman et Jean-Marc Faures du Programme stratégique de la FAO sur l'agriculture durable pour leur supervision constante, leurs suggestions concernant le contenu et la structure du présent document et leurs propositions d'amélioration.

## Introduction

La note méthodologique relative à l'indicateur 2.4.1 des ODD (FAO, 2019) aborde la question des sources de données se rapportant à l'indicateur, y compris les enquêtes statistiques classiques. L'évaluation de la plupart des sous-indicateurs à l'échelle nationale suppose que des données spécifiques soient collectées à l'échelle de l'exploitation, dans le cadre par exemple de recensements et d'enquêtes par sondage, ces dernières étant à privilégier de par leur rentabilité et leur fréquence (tous les trois ans).

Il est recommandé de collecter les données sur l'ODD 2.4.1 dans le cadre d'enquêtes nationales préexistantes, afin d'assurer la rentabilité de l'exercice. Au besoin, une enquête ponctuelle peut être menée dans le but précis de collecter les données relatives à l'indicateur. Dans la présente note d'orientation, nous aborderons et approfondirons différentes options, ainsi que les perspectives d'estimation correspondantes concernant les questions d'échantillonnage pour la mesure de l'ODD 2.4.1. Nous parlerons dans un premier temps de la conception de cette enquête ponctuelle, puis fournirons, dans une seconde partie, des conseils sur l'utilisation des enquêtes existantes, agricoles ou non.

## 1 Plan d'échantillonnage de l'enquête consacrée à l'indicateur 2.4.1 des ODD

Nous présenterons dans cette partie les principales caractéristiques de l'enquête spécifique, en tenant compte des besoins de données liés à l'ODD 2.4.1.

### 1.1 Unités d'observation et population cible

L'unité d'observation pour la mesure de l'ODD 2.4.1 est l'exploitation agricole, qu'elle relève ou non du secteur des ménages, telle que définie par le Programme mondial de recensement de l'agriculture 2020 de la FAO (FAO, 2015) :

- **Exploitation agricole** : unité économique de production agricole soumise à une direction unique et comprenant tous les animaux qui s'y trouvent et toute la terre utilisée entièrement ou en partie pour la production agricole, indépendamment du titre de possession, du mode juridique ou de la taille. La direction unique peut être exercée par un particulier, par un ménage, conjointement par deux ou plusieurs particuliers ou ménages, par un clan ou par une tribu, ou par une personne morale telle que société, entreprise collective, coopérative ou organisme d'État. L'exploitation peut contenir un ou plusieurs blocs, situés dans une ou plusieurs régions distinctes ou dans une ou plusieurs divisions territoriales ou administratives, à condition qu'ils partagent les mêmes moyens de production tels que la main-d'œuvre, les bâtiments agricoles, les machines ou animaux de trait utilisés pour l'exploitation.
- **Exploitations relevant du secteur des ménages (*exploitations familiales*)** : exploitations agricoles gérées par les membres d'un ménage.
- **Exploitations ne relevant pas du secteur des ménages (*exploitations non familiales*)** : exploitations agricoles telles que les sociétés et les institutions gouvernementales. La définition de ces exploitations peut varier d'un pays à l'autre.

Il convient toutefois de remarquer que les activités agricoles à prendre en compte sont la production végétale et animale. La sylviculture, l'aquaculture et la pêche sortent du champ d'application de l'indicateur 2.4.1. La population cible est donc constituée de toutes les exploitations agricoles générant des produits végétaux ou d'origine animale.

### 1.2 Unités d'échantillonnage et base de sondage

L'unité d'échantillonnage recommandée est l'exploitation agricole, car cela permet d'effectuer une stratification pertinente fondée sur la typologie des exploitations. La base de sondage correspond quant à elle à la liste complète des exploitations agricoles du pays. Cette liste peut être dressée à la suite d'un recensement agricole. La méthode recommandée par la FAO consiste à associer le recensement agricole au recensement de la population, en intégrant aux questionnaires de ce dernier des questions filtres comprenant, aux fins de l'ODD 2.4.1, des variables de stratification pertinentes (voir la section sur la stratification ci-dessous).

### 1.3 Unités déclarants

Toutes les informations qu'il est nécessaire d'obtenir auprès des exploitations peuvent être recueillies à l'occasion d'un entretien. L'unité déclarante est donc toute personne représentant l'exploitation et pouvant fournir les informations requises. Par exemple, l'unité déclarante d'une exploitation familiale peut être le chef du ménage ou tout autre membre du ménage ayant des connaissances solides sur l'exploitation.

### 1.4 Domaines d'estimation

Le domaine d'estimation est le pays. Toutefois, si un pays souhaite obtenir des estimations infranationales de l'ODD 2.4.1, les zones administratives infranationales (région, district, province, etc.) constitueront le domaine d'estimation, ce qui supposera un échantillon de taille plus importante, et donc des frais d'enquête plus élevés.

### 1.5 Plan d'échantillonnage et estimateurs

Un plan d'échantillonnage à deux degrés pourra être adapté aux exploitations du secteur des ménages dans la plupart des pays. De nombreux pays en développement connaissent bien cette méthode, qui présente de nombreux avantages opérationnels et techniques. Un échantillonnage aléatoire simple suffit généralement pour les exploitations non familiales. Pour plus d'informations et des recommandations concernant les plans d'échantillonnage à un ou deux degrés, les estimateurs et la variance, voir GSARS (2017, chapitre 5). Puisque l'indicateur 2.4.1 est un ratio (et est par conséquent non linéaire), sa variance peut être calculée à l'aide d'approximations fondées, par exemple, sur une linéarisation de Taylor.

### 1.6 Stratification

La stratification peut largement contribuer à améliorer la précision des estimations (Fuller, 2009). Les variables de stratification utilisées dans les enquêtes agricoles classiques sont généralement liées aux différentes typologies économiques des exploitations (taille, activité, etc.). Compte tenu de l'importance des dimensions environnementale et sociale pour l'ODD 2.4.1, il serait utile d'envisager une stratification correspondante. De fait, toutes les variables liées aux thématiques de l'indicateur 2.4.1 des ODD pourraient se prêter à une stratification efficace. Par exemple, un échantillon aléatoire composé uniquement d'exploitations situées dans des zones caractérisées par un sol pauvre et affichant par conséquent une faible productivité ne permettra pas d'évaluer correctement les thématiques 1, 2, 4 et 8. Une stratification appropriée permettrait de ne pas sélectionner un tel échantillon. Les variables associées à plusieurs thématiques doivent donc être privilégiées aux fins de stratification et être incluses dans le questionnaire de recensement. Les variables d'intérêt pour cet indicateur aux fins de stratification sont notamment le type d'activité (culture, élevage, mixte) et l'irrigation (exploitations irriguées/non irriguées).

En présence d'un plan d'échantillonnage à deux degrés, il est conseillé de procéder à une stratification des unités primaires d'échantillonnage (UPE). Pour l'indicateur 2.4.1, une stratification fondée sur les zones agroécologiques du pays est adaptée au niveau des UPE.

**Tableau 1** : thématiques proposées aux fins d'évaluation du niveau de durabilité

Dimension	Thématiques	Sous-indicateurs
Économique	Productivité des terres	Valeur de la production de l'exploitation par hectare
	Rentabilité	Revenu agricole net
	Résilience	Mécanismes d'atténuation des risques
Environnementale	Santé des sols	Ampleur de la dégradation des sols
	Utilisation de l'eau	Variation de la disponibilité de l'eau
	Risque de pollution par les engrais	Gestion des engrais
	Risques liés aux pesticides	Gestion des pesticides
	Biodiversité	Recours à des pratiques respectueuses de la biodiversité
Sociale	Emploi décent	Taux de rémunération dans le secteur agricole
	Sécurité alimentaire	Échelle de mesure du sentiment d'insécurité alimentaire (FIES)
	Régime foncier	Garantie des droits fonciers

Idéalement, les informations requises pour la stratification seraient disponibles dans la base de sondage. Si ce n'est pas le cas, il existe deux solutions : i) système d'échantillonnage à deux degrés ; et ii) post-stratification pour étalonner les poids d'échantillonnage après la collecte de données.

### 1.6.1 Échantillonnage à deux degrés aux fins de stratification

Compte tenu des avantages que peut avoir la stratification pour un indicateur comme l'ODD 2.4.1, il est important de disposer de bonnes variables de stratification dans la base de sondage afin de stratifier correctement la population avant échantillonnage. En l'absence de telles variables, il est possible de recourir à un système d'échantillonnage à deux degrés, qui consiste à sélectionner un premier échantillon de taille considérable pour collecter les variables de stratification, puis à effectuer, dans un second temps, un sous-échantillonnage stratifié à l'aide d'un plan d'échantillonnage à un ou deux degrés.

Imaginons par exemple qu'un pays souhaite stratifier la productivité de la terre en fonction des exploitations agricoles. Ces informations ne sont généralement pas disponibles dans les bases de sondage, puisqu'elles sont très difficiles à recueillir dans le cadre d'un recensement. Dans le cadre d'un plan d'échantillonnage à deux degrés, un grand échantillon d'exploitations agricoles est tout d'abord sélectionné pour collecter des données sur la superficie cultivée et la production végétale et animale au cours de la dernière saison agricole, afin de calculer la productivité de chaque exploitation. Cet échantillon peut ensuite être stratifié par niveau de productivité avant qu'un sous-échantillon ne soit sélectionné. Il est possible de procéder à cet exercice sur un laps de temps court si les agriculteurs sont en mesure de fournir des informations fiables sur leur superficie et leur production agricole.

Cochran (1977, p. 327) étudie les estimateurs liés à l'utilisation d'un double échantillon aux fins de stratification.

### 1.6.2 Post-stratification et calage

La post-stratification et le calage sont des méthodes couramment utilisées aux fins de calibrage des poids, qui reposent sur des informations sur la structure réelle de la population après collecte de données. Par exemple, si la structure de l'échantillon d'exploitations agricoles par type d'activité (culture, élevage et mixte) est différente de la structure réelle de la population (obtenue à partir de n'importe quelle source fiable), ces méthodes peuvent être utilisées pour améliorer la précision des estimations. Valliant *et al.* (2013, p. 374-381) fournissent des informations utiles sur ces techniques (et leur application pratique).

### 1.7 Taille de l'échantillon

La méthode de calcul de cet indicateur nécessite des estimations fiables des proportions des 11 sous-indicateurs  $n$  ( $SI_n$ ) pour les trois niveaux de durabilité ( $l$ ) : souhaitable ( $d$ ), acceptable ( $a$ ) et non durable ( $u$ ), soit un nombre total de ratios à estimer s'élevant à 33. Pour un sous-indicateur donné cependant, il suffit d'obtenir des estimations précises concernant deux des trois niveaux, le troisième étant égal à un moins la somme des deux autres. Pour conclure, l'enquête relative à l'ODD 2.4.1 doit porter sur un échantillon de taille adéquate pour estimer au moins 22 ratios. Idéalement, s'il existe des informations préalables sur tous les sous-indicateurs, les tailles d'échantillon minimales requises pour obtenir une estimation fiable (fondée sur des précisions attendues) pourraient être calculées pour chacun des 33 ratios et la taille maximale être envisagée pour l'enquête.

En suivant autant que possible les notes de la FAO (2019), prenons, pour un domaine d'estimation donné :

- $SI_{ln}$  : proportion du sous-indicateur  $n$  classée comme étant au niveau  $l$  ;  $n = 1 \dots 11, l \in (d, a, u)$
- $Y$  : superficie agricole
- $Z_{ln}$  : superficie agricole présentant un niveau de durabilité  $l$  pour le sous-indicateur  $n$
- $p_{ln}$  : part d'exploitations ayant une superficie agricole présentant un niveau de durabilité  $l$  pour le sous-indicateur  $n$
- $N$  : nombre total d'exploitations

L'indicateur  $SI_{ln}$  est donc calculé de la façon suivante :

$$SI_{ln} = \frac{Z_{ln}}{Y} \quad (1)$$

Pour une enquête visant à fournir une estimation précise de  $SI_{ln}$  avec une erreur relative  $\varepsilon$  et un intervalle de confiance  $(1 - \alpha)$ , il est nécessaire d'avoir une taille d'échantillon minimale de  $m_{ln}$ . Si  $\widehat{V}(SI_{ln})$  et  $\widehat{CV}_{ln}^2$  sont respectivement la variance estimée et le carré du coefficient de variation (CV) d'une estimation de  $SI_{ln}$ , nous obtenons alors :

$$m_{ln} = \frac{z_{\alpha/2}^2 \widehat{CV}_{ln}^2}{\varepsilon^2} \quad (2)$$



À partir de la formule (1) et de Särndal *et al.* (1992, p. 179), nous obtenons :

$$\hat{V}(S\hat{I}_{ln}) = \hat{V}\left(\frac{\hat{Z}_{ln}}{\hat{Y}}\right) = \frac{1}{\hat{Y}^2} \left[ \hat{V}(\hat{Z}_{ln}) + S\hat{I}_{ln}^2 \hat{V}(\hat{Y}) - 2S\hat{I}_{ln} \widehat{cov}(\hat{Z}_{ln}, \hat{Y}) \right] \quad (3)$$

Il convient de noter que, pour une exploitation agricole donnée,  $Z_{ln} = Y$  si la superficie agricole de l'exploitation est considérée comme présentant un niveau de durabilité  $l$  pour le sous-indicateur  $n$  ; sinon :  $Z_{ln} = 0$ .

Considérons une variable  $X$  suivant une distribution de Bernoulli affichant une probabilité de succès de  $p_{ln}$ . La variable  $Z_{ln}$  peut être modélisée comme un produit de deux variables indépendantes,  $Y$  et  $X$ .

$$Z_{ln} = XY$$

La moyenne de  $Z_{ln}$  est alors :

$$E(Z_{ln}) = E(XY) = E(X)E(Y) = p_{ln}E(Y)$$

La variance de  $Z_{ln}$  est :

$$\begin{aligned} V(Z_{ln}) &= V(XY) = E(X^2Y^2) - E(XY)^2 \\ &= E(X^2)E(Y^2) - E(X)^2E(Y)^2 \\ &= (V(X) + E(X)^2)(V(Y) + E(Y)^2) - E(X)^2E(Y)^2 \\ &= V(X)V(Y) + V(X)E(Y)^2 + V(Y)E(X)^2 \\ &= p_{ln}(1 - p_{ln})V(Y) + p_{ln}(1 - p_{ln})E(Y)^2 + p_{ln}^2V(Y) \\ &= p_{ln}V(Y) + p_{ln}(1 - p_{ln})E(Y)^2 \quad (4) \end{aligned}$$

La covariance mentionnée dans la formule (3) est :

$$\begin{aligned} cov(Z_{ln}, Y) &= E(Z_{ln}Y) - E(Z_{ln})E(Y) = E(XY^2) - E(XY)E(Y) \\ &= E(X)E(Y^2) - E(X)E(Y)E(Y) = E(X)(E(Y^2) - E(Y)^2) \end{aligned}$$

$$cov(Z_{ln}, Y) = p_{ln} V(Y) \quad (5)$$

Nous pouvons déduire des formules (3), (4) et (5) :

$$\begin{aligned} \hat{V}(S\hat{I}_{ln}) &= \frac{1}{\hat{Y}^2} \left[ (S\hat{I}_{ln}^2 - 2\hat{p}_{ln}S\hat{I}_{ln} + \hat{p}_{ln}) \hat{V}(\hat{Y}) + \hat{p}_{ln}(1 - \hat{p}_{ln})\hat{Y}^2 \right] \\ &= (S\hat{I}_{ln}^2 - 2\hat{p}_{ln}S\hat{I}_{ln} + \hat{p}_{ln}) \frac{\hat{V}(\hat{Y})}{\hat{Y}^2} + \hat{p}_{ln}(1 - \hat{p}_{ln}) \frac{\hat{Y}^2}{\hat{Y}^2} \\ \hat{V}(S\hat{I}_{ln}) &= (S\hat{I}_{ln}^2 - 2\hat{p}_{ln}S\hat{I}_{ln} + \hat{p}_{ln}) \widehat{CV}^2(\hat{Y}) + \hat{p}_{ln}(1 - \hat{p}_{ln}) \frac{1}{N^2} \quad (6) \end{aligned}$$

Pour les populations relativement vastes (comme c'est le cas pour les enquêtes nationales), la valeur de  $\frac{1}{N^2}$  serait très faible, et la deuxième partie de l'équation (6) peut être ignorée. Par conséquent, nous obtenons, à partir de l'équation (6) :

$$\widehat{CV}_{ln}^2 = \frac{V(\widehat{SI}_{ln})}{\widehat{SI}_{ln}^2} = \frac{1}{\widehat{SI}_{ln}^2} \left( \widehat{SI}_{ln}^2 - 2\widehat{p}_{ln}\widehat{SI}_{ln} + \widehat{p}_{ln} \right) \widehat{CV}^2(\hat{Y}) \quad (7)$$

En conclusion, les tailles d'échantillon minimales  $m_{ln}$  requises pour estimer les indicateurs  $SI_{ln}$  peuvent être calculées à l'aide d'informations tirées d'une enquête antérieure et des équations (2) et (7) ci-dessus. Les informations qu'il convient de tirer de l'enquête sont les suivantes : i) coefficient de variation estimé de la superficie agricole totale estimée ; ii) proportion  $\widehat{p}_{ln}$  estimée ; et iii) ratios  $\widehat{SI}_{ln}$  estimés. En l'absence d'enquête antérieure, une enquête pilote peut être menée.

La taille de l'échantillon final  $m$  pour l'enquête sera alors :

$$m = \max_{l,n} m_{ln}$$

Les fractions d'échantillonnage devraient être faibles pour la plupart des domaines d'estimation dans les pays. Toutefois, en cas de domaine d'estimation de taille modeste, il est possible de procéder à une correction de la population finie si la fraction d'échantillonnage est supérieure à 5 %. La taille corrigée  $m^*$  de l'échantillon sera alors de :

$$m^* = \frac{m}{1 + \frac{m-1}{N}} \quad (8)$$

Il est suggéré d'élargir la taille de l'échantillon pour tenir compte des éventuelles non-réponses. Par ailleurs, en cas de non-recours à un échantillonnage aléatoire simple, cette taille d'échantillon doit être multipliée par une estimation de l'effet du plan ( $\widetilde{D}_{eff}$ ). Par conséquent, si  $r$  est le taux de réponse attendu, la taille finale de l'échantillon sera comme suit :

$$\widetilde{m}_d = \frac{1}{r} \times \widetilde{D}_{eff} \times m^* \quad (9)$$

En cas de plan d'échantillonnage à deux degrés, la taille de l'échantillon d'UPE correspondra simplement à la taille de l'échantillon d'exploitations divisée par le nombre d'entretiens prévus dans chaque UPE (généralement 10 à 15).

## 1.8 Répartition de l'échantillon en strates

Compte tenu du nombre d'indicateurs devant être estimés à partir de l'enquête aux fins de l'évaluation de l'ODD 2.4.1, il serait adapté ici de procéder à une répartition proportionnelle de l'échantillon en strates, afin de garantir la réalisation des objectifs d'estimation. Une répartition multivariée de l'échantillon peut aussi être envisagée dans un tel contexte.

## 1.9 Absence de base

En cas d'absence de base actualisée et adaptée dans le pays pour sélectionner un échantillon dans le cadre d'une enquête relative à l'ODD 2.4.1, il est possible de sélectionner, pour les exploitations familiales, un échantillon d'UPE (districts de recensement, comtés, villages, etc.) puis de dresser une liste complète des exploitations qui s'y trouvent en vue de sélectionner l'échantillon final d'exploitations.

Pour les exploitations non familiales, les données administratives (p. ex., les registres d'entreprises agricoles) peuvent servir à élaborer une base à des fins d'échantillonnage.

## 2 Collecte de données sur l'ODD 2.4.1 à partir d'une enquête existante

Lorsque des enquêtes sont menées régulièrement dans le pays, la solution rentable consiste à en tirer parti pour collecter des données à des fins d'estimation de l'ODD 2.4.1. Dans cette partie, nous étudierons les principales difficultés pouvant survenir concernant les besoins en matière de données pour l'ODD 2.4.1 et proposerons des solutions, destinées notamment à résoudre les problèmes d'estimation correspondants.

### 2.1 Couverture des données

Si une enquête existante est utilisée, il conviendra d'évaluer la couverture des données. Il est important de vérifier que les questionnaires recueillent toutes les informations requises pour l'ODD 2.4.1. Il pourra être nécessaire d'ajouter au questionnaire un module spécifique intégrant les informations manquantes.

### 2.2 Cas des enquêtes agricoles

Les pays disposant d'un système national de statistiques agricoles bien établi peuvent utiliser les enquêtes agricoles existantes pour collecter des données relatives à l'ODD 2.4.1. Certains ajustements seront toutefois potentiellement nécessaires pour améliorer la méthode existante, en particulier en cas de recours à une base aréolaire.

#### 2.2.1 Problèmes liés aux unités d'observation et d'échantillonnage

Il est important de s'assurer que l'unité d'observation de l'enquête est effectivement l'exploitation agricole, telle que définie à la section 1.1, en particulier lorsqu'une base aréolaire est utilisée. Dans ce cas, l'unité d'échantillonnage est généralement un segment ou un point. L'unité d'observation est parfois la parcelle située dans le segment ou autour de la parcelle dans laquelle se trouve le point échantillonné.

Il est recommandé de s'assurer que les données sont collectées sur l'ensemble de l'exploitation à laquelle appartient la parcelle. Il sera donc nécessaire de localiser l'exploitant et d'organiser un entretien sur l'exploitation aux fins de l'indicateur 2.4.1. D'une manière générale, dans tous les cas, les informations devront être collectées à l'échelle de l'exploitation agricole, auprès d'une unité déclarante capable de les fournir.

Des problèmes d'estimation peuvent survenir lorsque les unités d'échantillonnage diffèrent des unités d'observation requises, à savoir les exploitations agricoles. C'est le cas, déjà décrit ci-dessus, des échantillons sélectionnés à partir d'une base aréolaire. Sinon, dans certains pays, les unités d'échantillonnage des enquêtes agricoles sont les ménages, en particulier lorsque l'échantillon final est directement tiré des données de recensement de la population. Toutefois, il peut arriver qu'une exploitation appartienne à plusieurs ménages, ou qu'un seul ménage possède plusieurs exploitations. Dans ces cas, le calcul du poids des exploitations doit prendre en compte les multiplicités reliées à la complexité de leurs liens avec les unités d'échantillonnage. La GSARS (2015a et 2017) propose dans ces situations de recourir à la méthode généralisée du partage des poids (MGPP) mise au point par Lavallée (2007). Cette méthode est illustrée dans la section 2.3.2 ci-dessous.

### 2.2.2 Taille de l'échantillon

Si la taille de l'échantillon de l'enquête agricole a été calculée en vue d'une estimation précise de la superficie agricole située dans le domaine d'estimation (comme c'est le cas dans de nombreux pays en développement), elle peut être supérieure ou inférieure à la taille d'échantillon requise pour estimer l'indicateur 2.4.1. Si la taille est supérieure à celle requise pour l'ODD 2.4.1, il est possible de sélectionner un sous-échantillon aux fins de l'indicateur. Dans les autres cas, l'échantillon devra soit être pris en compte dans son intégralité, soit être élargi aux fins de l'indicateur, notamment en cas d'écart important avec la taille requise.

En fait, pour un sous-indicateur  $SI_{ln}$  donné, un sous-échantillon pourrait être sélectionné si  $\widehat{CV}_{ln}^2 \leq \widehat{CV}^2(\hat{Y})$ . En partant de la formule (8), cela équivaut à :

$$\frac{1}{\widehat{SI}_{ln}^2} \left( \widehat{SI}_{ln}^2 - 2\widehat{p}_{ln}\widehat{SI}_{ln} + \widehat{p}_{ln} \right) \leq 1 \Leftrightarrow \widehat{SI}_{ln}^2 - 2\widehat{p}_{ln}\widehat{SI}_{ln} + \widehat{p}_{ln} \leq \widehat{SI}_{ln}^2 \Leftrightarrow \widehat{SI}_{ln} \geq \frac{1}{2} \quad (10)$$

L'équation (10) fournit alors une condition nécessaire au recours au sous-échantillonnage pour un sous-indicateur  $n$  classé comme ayant un niveau de durabilité  $l$ . En pratique, cette condition doit être vérifiée pour tous les ratios  $SI_{ln}$  avant d'opter pour le sous-échantillonnage. Dans les autres cas, l'échantillon tout entier devra être pris en compte pour l'indicateur 2.4.1.

### ***Échantillon complémentaire***

Au cas où l'échantillon utilisé pour l'enquête existante est jugé non adapté à l'ODD 2.4.1, par exemple parce qu'il ne couvre pas de façon adaptée un type spécifique d'exploitations ou les exploitations d'une zone spécifique, et ne peut être complètement renouvelé, un échantillon complémentaire pourra être sélectionné pour améliorer la couverture.

## 2.3 Cas des enquêtes non agricoles

Certains pays en développement ne mènent pas d'enquêtes agricoles, mais mènent certaines enquêtes auprès des ménages, dans le cadre desquelles certaines données agricoles sont parfois collectées. Ces enquêtes ne sont généralement pas conçues pour établir des statistiques agricoles ; elles ne peuvent donc produire des estimations fiables des données agricoles de base, telles que la superficie plantée ou la production végétale. Ces enquêtes tiennent parfois compte du secteur agricole, mais les ménages continuent d'être les unités d'échantillonnage et d'observation. Citons par exemple l'étude sur la mesure des niveaux de vie – enquêtes intégrées sur l'agriculture (LSMS-ISA) de la Banque mondiale.

Ces enquêtes ne sont donc pas conseillées pour un indicateur comme l'ODD 2.4.1. S'il n'est pas possible de mener une enquête spécifique, deux autres approches sont toutefois proposées : sélection d'un échantillon complémentaire et échantillonnage indirect d'exploitations agricoles.

### 2.3.1 Utilisation d'un échantillon complémentaire

Encore une fois, un échantillon complémentaire peut être sélectionné pour être sondé, en plus de l'échantillon de l'enquête non agricole. La première étape consiste à évaluer la part d'unités qui pratiquent l'agriculture au sein de l'échantillon existant, ce qui donnera une idée de la couverture des unités d'intérêt dans chaque domaine d'estimation. Pour les domaines dans lesquels la couverture est faible, la sélection d'un échantillon complémentaire peut être envisagée pour collecter des données.

Au cas où l'enquête non agricole est une enquête auprès des ménages, il est obligatoire de sélectionner un échantillon complémentaire d'exploitations non familiales.

### 2.3.2 Échantillonnage indirect

La deuxième option consiste à utiliser un cadre d'échantillonnage indirect pour estimer les poids des unités d'intérêt (exploitation agricole). Cette solution est envisageable si chaque exploitation du pays est liée d'une façon ou d'une autre à au moins une unité d'échantillonnage de l'enquête non agricole. L'idée consiste à collecter des données relatives à l'ODD 2.4.1 sur toutes les exploitations liées aux unités échantillonnées de l'enquête. Le poids des exploitations peut être estimé à l'aide de la méthode généralisée du partage des poids (MGPP) de Lavallée (2007).

Supposons par exemple que l'unité d'échantillonnage de l'enquête non agricole est le ménage. Le lien entre un ménage et une exploitation peut être défini de la façon suivante : « *un ménage est lié à une exploitation si le ménage détient tout ou partie de l'exploitation* ». Prenons  $L_{if}$  comme indicateur du lien entre l'exploitation  $f$  et le ménage  $i$ . La variable  $L_{if}$  est égale à 1 s'il existe un lien et à 0 en l'absence de lien.  $L_f = \sum_i L_{if}$  correspond donc au nombre total de liens entre l'exploitation  $f$  et la population de ménages.

Partons de l'hypothèse qu'une exploitation  $f$  donnée est liée à  $s^f$  ménages de l'échantillon de l'enquête non agricole ( $s^f \neq 0$ ) et liée à  $t^f$  autres ménages qui n'ont pas été sélectionnés dans ledit échantillon. Si  $w_i$  correspond au poids d'échantillonnage du ménage  $i$ , le poids  $w_f$  de l'exploitation  $f$  sera, selon la méthode MGPP :

$$w_f = \frac{\sum_i L_{if} w_i}{\sum_i L_{if}} = \frac{\sum_i L_{if} w_i}{L_f} = \frac{\sum_{i=1}^{s^f} w_i}{s^f + t^f} \quad (11)$$

## Conclusion

Pour conclure, ce travail fournit des conseils aux pays sur l'échantillonnage et l'estimation en vue de la production de statistiques sur l'ODD 2.4.1, qui correspond à la proportion des zones agricoles exploitées de manière productive et durable. L'indicateur peut être estimé en utilisant une enquête existante ou en menant une enquête spécifique. La recommandation rentable consiste cependant à utiliser les enquêtes agricoles ou enquêtes auprès des ménages collectant des données agricoles qui existent déjà. Ce document traite des problèmes d'échantillonnage et d'estimation rencontrés dans tous les cas et fournit des recommandations pour une estimation fiable de l'indicateur.

## Références bibliographiques

**Cochran, W.G.**, 1977, *Sampling Techniques*. 3<sup>e</sup> édition. John Wiley & Sons : New York, États-Unis.

**FAO**, 2015, *Programme mondial du recensement de l'agriculture 2020. Volume 1 : Programme, concepts et définitions*. Publications de la FAO : Rome.

**FAO**, 2019, *Indicateur 2.4.1 des ODD – Proportion des zones agricoles exploitées de manière productive et durable : note méthodologique. Quatrième révision*. FAO, Rome.

**Fuller, W. A.**, 2009, *Sampling statistics*. Hoboken, NJ : Wiley.

**GSARS**, 2015a, *Directives sur les cadres d'enquêtes intégrées*. Rapport technique GSARS : Rome.

**GSARS**, 2015b, *Manuel sur les bases de sondage principales pour les statistiques agricoles. Développement de la base de sondage, conception du plan d'échantillonnage et estimation*. Manuel de la Stratégie mondiale : Rome.

**GSARS**, 2017, *Manuel AGRIS sur les enquêtes agricoles intégrées*. Manuel de la Stratégie mondiale : Rome.

**Lavallée, P.**, 2007, *Indirect Sampling*. Springer : Ottawa.

**Särndal, C.E., Swensson, B. et Wretman, J.**, 1992, *Model-Assisted Survey Sampling*. Springer-Verlag : New York, États-Unis.

**Valliant, R., Dever, J. et Kreuter, F.**, 2013, *Practical Tools for designing and weighting survey samples*. New York, NY : Springer.