

Canada – Mexique – États-Unis

Groupe de concepts de données et de métadonnées agricoles

1. DÉFINITIONS

La statistique agricole est fondée sur des enquêtes et des recensements menés au Canada, au Mexique et aux États-Unis. Chacun des trois pays gère son propre programme de la statistique agricole en s'appuyant sur les données agricoles qu'il recueille par l'entremise d'unités de collecte.

1.1 Canada

Les principales unités agricoles ayant servi à recueillir les statistiques agricoles canadiennes sont les suivantes :

Exploitation : L'exploitation agricole canadienne est le niveau d'exploitation le moins élevé d'une entreprise produisant des denrées pour la vente, et pour laquelle un état indépendant des revenus et dépenses est disponible. Les frontières provinciales sont respectées.

Exploitant : L'exploitant agricole canadien est la personne qui est responsable de la prise des décisions quotidiennes liées à l'exploitation agricole — cette prise de décisions peut se faire en collaboration avec d'autres exploitants.

1.2 Mexique

Les principales unités agricoles ayant servi à recueillir des statistiques agricoles sont les suivantes :

Unité de production : Une unité de production regroupe toutes les parcelles de terrain qui sont exploitées par une seule administration. Il existe deux types principaux de propriété : les propriétés privées et les propriétés communautaires appelées *ejidos*. Les propriétés communautaires renferment de nombreuses unités que les propriétaires exploitent pour leur propre subsistance et non pas pour la vente de produits agricoles.

Producteur : Les personnes qui gèrent l'unité de production et qui sont, par conséquent, responsables de la totalité ou de la plupart des décisions liées aux activités agricoles, telles que l'ensemencement et les récoltes.

1.3 États-Unis

Les principales unités agricoles ayant servi à recueillir les statistiques agricoles américaines sont les suivantes :

Ferme : Toute entreprise qui produit et vend, ou aurait normalement produit et vendu, des denrées d'une valeur d'au moins 1 000 \$US pendant l'année.

Agriculteur : La personne responsable de la totalité ou de la plupart des décisions quotidiennes liées à l'exploitation agricole, telles que l'ensemencement, les récoltes, l'engraissement et la commercialisation. L'exploitant peut être le propriétaire, un gérant engagé, un locataire, un actionnaire ou un associé.

2. PROGRAMMES DE LA STATISTIQUE AGRICOLE

Les programmes de la statistique agricole des trois pays sont fondés sur un ensemble de documents composé d'enquêtes, de recensements et de sources de nature administrative.

2.1 Canada

Le Programme de la statistique agricole du Canada recueille, compile, analyse et publie un large éventail de renseignements sur le secteur agricole dans l'ensemble du pays et dans chaque province. Le Programme comporte des enquêtes bisannuelles, annuelles et saisonnières sur les cultures, l'horticulture, le bétail, les finances agricoles et l'aquaculture, ainsi que le Recensement de l'agriculture, qui est réalisé tous les cinq ans. En outre, pour alléger le fardeau du répondant et réduire les coûts associés aux enquêtes là où c'est possible, on utilise, en remplacement des enquêtes agricoles, des données administratives provenant des gouvernements et organismes fédéraux et provinciaux, d'agences de commercialisation et d'organisations de producteurs.

La taille et le contenu des enquêtes agricoles sont variables. Les grandes enquêtes ordinaires sont les suivantes : a) l'Enquête sur les grandes cultures est réalisée six fois l'an en vue de recueillir des renseignements sur les intentions d'ensemencement, les superficies ensemencées, le rendement, la production et les stocks de céréales; b) l'Enquête sur le bétail, menée en janvier et en juillet de chaque année, fait la collecte des stocks de bovins, de porcs et de moutons; c) l'Enquête financière sur les fermes, réalisée tous les deux ans pendant l'été, recueille des renseignements sur les revenus, les dépenses, les biens et les caractéristiques physiques; d) l'Enquête sur les fruits et légumes est menée en novembre et fournit des données sur la superficie cultivée et la production reliées à 50 fruits et légumes différents; e) l'Enquête annuelle sur les cultures de serre, les pépinières et les gazonnières, menée en janvier, est réalisée dans le but d'obtenir une vue générale de la situation économique dans ces secteurs. Parmi les enquêtes de plus petite envergure figurent : f) l'Enquête sur la superficie et le rendement des pommes de terre; g) l'Enquête sur le prix du foin et de la paille; h) l'Enquête sur les stocks commerciaux de maïs-grain et de soya; i) l'Enquête sur les cultivateurs de champignons; j) l'Enquête sur la production de miel; k) l'Enquête sur les produits de l'érable; l) l'Enquête sur la production de volaille et d'œufs.

De plus, l'Enquête de dépistage des fermes (EDF) est menée régulièrement avec pour principal objectif de repérer les nouvelles fermes potentielles à ajouter au Registre des entreprises et aux bases de sondage. Cette enquête utilise les formulaires de déclaration de revenus et des listes administratives pour identifier les unités qui sont susceptibles d'être de nouvelles fermes.

Enfin, des enquêtes hors série sont réalisées en recouvrement des coûts pour répondre à des besoins particuliers venant d'autres organismes gouvernementaux ou de clients externes. Bien

que la plupart de ces enquêtes ne soient pas répétées, leur portée est rarement négligeable. L'Enquête sur la gestion agroenvironnementale et l'Enquête sur l'eau dans l'agriculture en sont des exemples.

Le Programme de la statistique agricole est enrichi par le Programme des données fiscales, le Programme de statistiques sur le revenu agricole, le recours à la télédétection et à l'analyse géospatiale ainsi que la recherche rurale et agricole.

2.2 Mexique

Le programme mexicain de la statistique agricole recueille, compile, analyse et diffuse un large éventail de renseignements sur le secteur agricole au Mexique. Pour recueillir ces renseignements, le programme utilise les sources et procédés suivants : des registres administratifs pour la plupart des données relatives aux cultures et à l'élevage; des enquêtes saisonnières sur les cultures les plus importantes pour les données relatives à la production, aux prix et aux coûts de production; la télédétection pour l'estimation de la superficieensemencée des principales cultures. À cela s'ajoute un recensement agricole effectué tous les dix ans, le dernier remontant à 2008.

L'information administrative est combinée, analysée et validée par deux systèmes ; le système d'information sur l'agriculture (SIACAP) et le système d'information sur le bétail (SIPCAP). L'information sur l'agriculture comprend notamment la superficieensemencée, la superficie récoltée, la superficie endommagée, les récoltes observées et estimées, la production observée et estimée ainsi que le prix moyen en milieu rural. L'information sur le bétail inclut les stocks, le poids et les prix déclarés des bovins, des porcs, des moutons, des chèvres et de la volaille; la production, le poids et le prix de la viande de bœuf, de porc, de mouton, de chèvre et de volaille; la production et le prix des sous-produits agricoles et des produits d'origine animale comme le lait, les œufs, etc.

Les enquêtes sur échantillon sont réalisées dans les principaux États producteurs dont la production combinée représente 80 % de la production totale. À l'intérieur de chaque État, ces enquêtes sont menées dans les principales zones cultivées dont la production combinée représente 80 % de la production totale de l'État. Les interviews saisonnières avec les producteurs et les mesures physiques dans les unités de production sont utilisées pour estimer la production de ces dernières. Au Mexique, les principaux produits comprennent entre autres le maïs, les haricots, le sorgho et le blé. Les légumes occupent une place importante sur le plan de la valeur de la production.

Enfin, on estime la superficieensemencée au moyen de l'information obtenue sur le terrain et d'images satellites.

2.3 États-Unis

Le National Agricultural Statistics Service (NASS) du United States Department of Agriculture (USDA) réalise des centaines d'enquêtes chaque année et prépare des rapports sur pratiquement tous les aspects de l'agriculture américaine tels que, entre autres, la production et l'approvisionnement en nourriture et en fibres, les prix payés et obtenus par les fermiers, la main-

d'œuvre et les salaires dans le secteur agricole, les finances agricoles, l'utilisation de produits chimiques et les changements dans les données démographiques des producteurs américains.

Le programme comprend des enquêtes annuelles, trimestrielles, mensuelles et saisonnières, de même qu'hebdomadaires dans certains cas. En outre, un recensement de l'agriculture est tenu tous les cinq ans. Des données administratives provenant d'organismes gouvernementaux à aux échelons fédéral et des États, ainsi que de sources privées, sont utilisées en tant que complément à la collecte des données, pour alléger le fardeau du répondant et pour réduire le coût des enquêtes.

2.3.1 Programme des statistiques sur les cultures

La taille et la portée des enquêtes réalisées par le NASS sont variables. Les enquêtes sur les cultures et sur les stocks constituent la principale source de données pour les estimations de la superficie, des rendements et de la production, de même que des quantités de céréales et d'oléagineux entreposées à la ferme. Les enquêtes sont menées trimestriellement en mars, en juin, en septembre et en décembre. Les intentions d'ensemencement des agriculteurs sont recueillies en mars; la superficie ensemencée et les prévisions sur la superficie récoltée le sont en juin; la superficie récoltée et la production de petites céréales, en septembre; enfin, les données sur les cultures en rang et la production de foin sont recueillies en décembre. L'information sur les céréales et les oléagineux entreposés à la ferme pour les principaux produits de base est recueillie tous les trimestres, alors que les données sur les cultures spécialisées sont recueillies une fois l'an dans certains États producteurs.

Une enquête sur les rendements agricoles fournit des données déclarées par les exploitants sur les rendements prévus des récoltes pour l'estimation des niveaux de production végétale tout au long de la saison de croissance. L'enquête est réalisée tous les mois, de mai à novembre. Les données sur les petites céréales sont recueillies de mai à août; celles sur les cultures en rang, d'août à novembre; celles sur les rendements du foin, en août et en octobre, et celles sur les stocks de foin, en mai. Les données sur le tabac sont recueillies de mai à novembre. Des enquêtes sur la mesure objective des rendements sont réalisées aux mêmes périodes pour le maïs, le soya, le blé, le coton et les pommes de terre dans les principaux États producteurs. D'autres enquêtes sont faites plus ou moins fréquemment, notamment l'enquête sur la floriculture et l'enquête sur les agrumes. Les prévisions et estimations portant sur les fruits autres que les agrumes, les noix et les légumes ont été interrompues en 2013.

2.3.2 Programme des statistiques sur le bétail

Les enquêtes sur les cheptels de bovins, de moutons et de chèvres sont menées deux fois l'an, en janvier et en juillet. Les enquêtes sur les bovins d'engraissement ont lieu tous les mois, tout comme les enquêtes sur la production de lait et de produits laitiers. Les enquêtes sur les porcs sont réalisées tous les trimestres, soit en mars, en juin, en septembre et en décembre. Une enquête sur les couvoirs pour les poulets à griller est effectuée hebdomadairement dans les 19 principaux États producteurs et tous les mois dans 11 autres États. Une enquête sur les couvoirs pour les dindons a lieu tous les mois dans

16 États. Des enquêtes sur les poulets et les œufs ont également lieu tous les mois afin de mesurer la production d'œufs, les taux de ponte, le nombre de pondeuses et de poulettes de même que le nombre de mues forcées. Les principales enquêtes sur la production en aquaculture (poissons-chats et truites, notamment livraisons de nourriture pour poissons-chats et transformation des poissons-chats), le rapport de juillet sur le bétail et le rapport sur le vison ont été interrompus en 2013.

2.3.3 Programme des statistiques environnementales, économiques et démographiques

Les données environnementales obtenues au moyen des enquêtes sur l'utilisation de produits chimiques et de pesticides qui étaient menées, pour une sélection de produits, en rotation ou chaque n^{ième} année ont été interrompues en 2013. Des données démographiques sont obtenues grâce à l'enquête semestrielle sur la main-d'œuvre agricole et au Recensement de l'agriculture.

Dans l'optique d'évaluer la santé économique du secteur agricole, le NASS organise tous les ans l'Agricultural Resource Management Study (ARMS). L'ARMS est réalisée en trois phases. Dans la phase I, qui a lieu en mai, on sélectionne les fermes qui se qualifieront pour les autres phases. Dans la phase II, menée de septembre à décembre, on recueille des données sur les pratiques de production agricole, l'utilisation des ressources et les coûts variables de production pour des produits spécifiques. La phase III, qui débute en février et prend fin en avril, permet de recueillir des données sur les finances agricoles, les caractéristiques des exploitants, les ménages agricoles, les dépenses d'exploitation, les améliorations apportées aux immobilisations, les actifs et la dette liée à la production agricole.

Une enquête dont le but est la production d'estimations sur les prix obtenus par les agriculteurs pour leurs produits agricoles est également réalisée chaque mois, et une enquête annuelle sur les prix payés fait l'estimation des coûts des intrants. Enfin, des enquêtes sur les prix des produits laitiers sont faites toutes les semaines.

2.3.4 Recensement de l'agriculture

Un recensement de l'agriculture est réalisé tous les cinq ans. Le suivi du recensement est assuré à l'aide d'enquêtes spéciales, notamment les recensements sur l'horticulture et l'aquaculture, l'enquête sur l'irrigation des fermes et des ranchs de même qu'une enquête sur la propriété des terres agricoles.

3. CYCLE DE TRAITEMENT

Les programmes de la statistique agricole de chacun des trois pays doivent franchir une série d'étapes de traitement avant d'être diffusés.

3.1 Canada

Le Programme de la statistique agricole du Canada est fondé sur un recensement quinquennal de l'agriculture et sur sept grandes enquêtes réalisées tous les ans. Voici certaines des principales étapes de traitement :

3.1.1 Définition et exclusions des bases de sondage

Le Recensement de l'agriculture cible chaque ménage canadien en vue d'identifier les exploitations agricoles et leurs exploitants. En 2006, on a fait parvenir par la poste un questionnaire du Recensement de l'agriculture à un sous-ensemble de fermes, et les autres questionnaires ont été livrés par les agents recenseurs s'il s'avérait qu'un des membres du ménage exploitait une ferme. En 2011, tous les questionnaires ont été envoyés par la poste. En outre, pour repérer les exploitations agricoles manquantes, le questionnaire du Recensement de la population renfermait une question demandant si un des membres du ménage était exploitant agricole. Tous les exploitants agricoles devaient retourner le questionnaire rempli par la poste ou remplir la version électronique en ligne. Les données de 2011 ainsi recueillies ont été utilisées pour la mise à jour du Registre des entreprises du Canada.

Pendant les périodes intercensitaires, le Registre des entreprises du Canada utilise les dossiers d'impôt pour trouver les nouvelles exploitations agricoles potentielles. Les dossiers indiquant un revenu agricole dans la déclaration de revenus sont échantillonnés, et on demande aux exploitants de remplir un questionnaire afin de déterminer si leur dossier représente une exploitation agricole et dans ce cas, quels sont ses produits agricoles.

Dans le cadre du Programme de la statistique agricole du Canada, les enquêtes agricoles utilisent le Registre des entreprises à jour comme base liste. Les bases de sondage excluent habituellement les petites fermes, c'est-à-dire les fermes qui, dans le dernier recensement, ont déclaré un revenu agricole inférieur à 10 000 \$CAN. À l'étape de l'estimation, des ajustements sont faits pour que ces unités exclues soient prises en compte et que toutes les exploitations affichant des ventes positives soient ainsi représentées. En général, ces ajustements comptent pour moins de 1 % des principales estimations visant les produits de base.

3.1.2 Plan d'échantillonnage et sélection de l'échantillon

Le Recensement de l'agriculture utilise toutes les unités actives figurant dans le Registre des entreprises du Canada pour l'envoi initial par la poste du questionnaire du recensement.

Le plan d'échantillonnage des différentes enquêtes agricoles consiste d'abord à repérer toutes les exploitations agricoles de la population d'intérêt d'après les renseignements figurant dans le Registre des entreprises du Canada. Cette population est stratifiée selon la région géographique, la taille et, dans certains cas, le type de ferme. Les fermes les plus petites sont placées dans une strate à tirage nul et ne peuvent pas être sélectionnées aux fins de l'échantillon. À partir des données de la base de sondage qui ont trait aux variables d'intérêt, l'échantillon est réparti selon des méthodes univariées ou multivariées de façon à

ce que les estimations prévues de l'enquête pour ces variables clés aient le degré de précision prescrit. On attribue un nombre aléatoire permanent à chaque unité afin de limiter le chevauchement de l'échantillon d'une occasion d'enquête à l'autre.

3.1.3 Étapes de la collecte

Les recensements et les enquêtes recueillent des données sur les exploitations agricoles. Alors que le recensement a recours aux formulaires papier et à Internet pour la collecte des renseignements, les enquêtes s'appuient principalement sur les interviews téléphoniques assistées par ordinateur (ITAO), qui permettent l'utilisation de contrôles automatisés. Toutes les données recueillies par Statistique Canada correspondent à des périodes de référence ou à des dates spécifiques. Ces périodes, notamment l'année civile, le mois ou la saison de croissance, sont utilisées pour la collecte des données sur la mobilité ou les pratiques agricoles. Les dates sont utilisées pour les données sur les stocks ou l'état des fermes à un point précis dans le temps.

Habituellement, on constate des taux de non-réponse variant entre 5 % et 20 % dans les enquêtes agricoles canadiennes. Ces taux tiennent compte des taux de refus, qui s'échelonnent entre 2 % et 5 %.

3.1.4 Vérification des données

La vérification des données est effectuée au moment de la collecte, puis plus tard, en lots dans le cadre des activités de traitement. La vérification des données combine la vérification des non-réponses, la comparaison des réponses avec les données antérieures, le contrôle de la cohérence, la vérification statistique et les méthodes de détection des valeurs aberrantes. Tous les enregistrements subissent le même ensemble de vérifications automatisées. En général, les principales unités font en plus l'objet de vérifications manuelles.

3.1.5 Imputation des données

Lorsque des rejets sont constatés au contrôle, les champs erronés sont signalés à des fins d'imputation. Dans le contexte d'une enquête, si les champs les plus importants sont signalés, l'enregistrement est déclaré inutilisable pour les traitements ultérieurs, ce qui a pour effet de réduire le nombre de répondants. Les enregistrements « utilisables » peuvent contenir des éléments rejetés au contrôle, et ces éléments sont habituellement corrigés à l'aide de méthodes d'imputation par enregistrement donneur lorsque des méthodes déterministes ne peuvent être utilisées. Grâce à cette approche, les taux d'imputation pour les produits clés sont presque nuls pour la plupart des enquêtes agricoles canadiennes.

3.1.6 Estimations

Les estimations des enquêtes sont produites au moyen d'un estimateur Horvitz-Thompson ordinaire. Les cas de non-réponse et les autres enregistrements jugés inutilisables suite à l'étape de vérification sont compensés par un ajustement des poids d'estimation. Les techniques d'ajustement des poids ont également un effet positif sur certaines valeurs

aberrantes. Les exclusions des bases de sondage sont ajustées au niveau macro à l'aide des ratios du recensement.

La théorie statistique habituelle produit les indicateurs d'exactitude liés à la variance et au coefficient de variation.

En général, on observe des coefficients de variation se situant entre 1 % et 5 % pour les produits clés faisant l'objet d'estimations dans les enquêtes canadiennes sur l'agriculture.

3.1.7 Contrôle de la divulgation

Les règles de divulgation évaluent les risques de divulgation directe et par recoupements pour les cellules totalisées avant la diffusion. Il peut y avoir divulgation directe si très peu de facteurs apportent une contribution non nulle à l'estimation de la cellule ou si un nombre très limité de facteurs contribuent à un pourcentage très important de l'estimation globale. On utilise des méthodes automatisées pour repérer les cellules qui comportent des risques directs de divulgation. Des méthodes manuelles servent à limiter les risques de divulgation par recoupements. Dans les deux cas, les cellules jugées à risque sont supprimées. On prend soin de tenir compte des tableaux préalablement diffusés pour éviter toute divulgation par recoupements entre plus d'un tableau.

3.1.8 Mise à jour des bases de sondage

Les réponses des enquêtes sont utilisées pour maintenir le Registre des entreprises à jour, et pour être en mesure de fournir des données fiables dans les enquêtes subséquentes. Certaines mises à jour du Registre des entreprises ne concernent que certaines enquêtes, alors que les autres s'appliquent à l'ensemble des enquêtes.

3.2 Mexique

3.2.1 Définition et exclusions des bases de sondage

Pour obtenir la plupart des informations liées aux récoltes et au bétail, le programme mexicain de la statistique agricole utilise des registres administratifs. Pour les données relatives à la production, aux prix et aux coûts de production, il utilise des enquêtes saisonnières sur les cultures les plus importantes. Enfin, il utilise la télédétection pour obtenir une estimation de la superficieensemencée des principales cultures. À cela s'ajoute un recensement agricole effectué tous les dix ans.

3.2.1.1 Données administratives

Les bureaux d'État rassemblent les rapports sur les superficiesensemencées et récoltées pendant les dix premiers jours de chaque mois et continuent à saisir cette information pour pouvoir inclure les données accumulées dans le rapport mensuel suivant. Le bureau central émet alors les règles et le système d'information pour le traitement des données. Le processus principal d'obtention des données administratives est le suivant :

- 1) L'information est saisie dans les systèmes des programmes SIACAP et SIPCAP par les centres d'appui au développement rural (CADER) ou les districts de développement rural (DDR) dans les bureaux d'État;
- 2) L'information est validée par les bureaux d'État;
- 3) Les bureaux d'État transmettent les données au bureau central;
- 4) Le bureau central, qui fournit le soutien informatique, traite l'information et produit la base de données définitive pour la consultation.

3.2.1.2 Enquêtes sur échantillon

Les activités principales qui ont été réalisées dans le cadre des enquêtes sont l'identification des concepts et des méthodes servant à réunir l'information appropriée pour la résolution du problème, la conception des outils pour la collecte de cette information (questionnaires) ainsi que la conception de tableaux de résultats. La conception statistique de l'enquête comporte les activités suivantes : l'identification de l'unité observée, la mise sur pied d'une base d'échantillonnage, le calcul de la taille des échantillons, la mise sur pied de critères de sélection pour l'échantillon ainsi que de critères pour le calcul des estimations et, enfin, la logistique du travail sur le terrain (comprenant l'organisation des activités sur le terrain et l'horaire de ces activités). La superficie qui sera mise à l'étude utilise une base liste telle que le PROCAMPO (Programme d'assistance directe pour le secteur de l'agriculture).

Le bureau central offre des séances de formation au personnel technique de l'État sur la façon d'utiliser et de remplir les questionnaires, en tenant compte des deux méthodes mentionnées précédemment (enquête et recensement). De plus, le personnel coordonne les travaux et les activités pour s'assurer que les méthodes ont été appliquées correctement. Il compile d'autres renseignements provenant d'autres sources pour valider l'information obtenue au moyen des questionnaires, une fois qu'elle a été saisie et emmagasinée dans une base de données au moyen d'un logiciel conçu à cet effet par le bureau central. Par l'entremise d'interviews directes avec les producteurs, ces enquêtes permettent d'obtenir de l'information sur la production, les productions objectives, les coûts de production et les prix payés aux producteurs ainsi que sur d'autres produits.

Le bureau effectue une validation manuelle des données contenues dans les questionnaires et une validation électronique des données contenues dans les bases de données au moyen d'un contrôle des données statistiques. De plus, il détermine la valeur d'une variable lorsqu'il est placé devant plusieurs valeurs différentes résultant de l'utilisation de différentes méthodes. Pour faire l'estimation des récoltes, le bureau réalisera des enquêtes au moyen d'interviews avec les producteurs et en ayant recours de données reliées sur les productions objectives, les coûts de production, les prix payés aux producteurs et sur d'autres domaines.

3.2.2 Plan d'échantillonnage et sélection de l'échantillon

Les enquêtes utilisent toutes les unités actives figurant dans le Recensement de l'agriculture et les registres administratifs en tant que liste d'échantillonnage de la base de sondage. Sont également utilisés les renseignements locaux provenant de programmes gouvernementaux comme la liste des producteurs d'ASERCA. Comme ASERCA est une entité du

gouvernement fédéral qui octroie des subventions aux producteurs agricoles, elle a une liste des producteurs subventionnés et d'autres variables clés qui servent à l'établissement des bases listes.

Le plan d'échantillonnage des différentes enquêtes agricoles consiste d'abord à repérer toutes les exploitations de la population d'intérêt d'après les renseignements provenant des sources accessibles et à créer une liste unique. La population est stratifiée selon la région géographique, la production et le rendement cultural. Toutes les strates sont admissibles.

À partir des données de la base de sondage qui ont trait aux variables d'intérêt, l'échantillon est réparti selon des méthodes multivariées de façon à ce que les estimations prévues de l'enquête pour les variables clés aient le degré de précision prescrit.

On attribue un nombre aléatoire à chaque unité pour limiter le chevauchement de l'échantillon.

D'autres échantillons peuvent être prélevés parmi des populations qui présentent des caractéristiques spéciales. Dans ces cas, la pondération est corrigée au cours d'un processus de poststratification.

3.2.3 Étapes de la collecte

Les enquêtes recueillent des données sur les exploitations agricoles au moyen de formulaires papier et par Internet. Ces données servent à estimer les flux ou les pratiques agricoles pour une période particulière telle que l'année civile, le mois ou la saison de croissance.

Habituellement, les taux de non-réponse varient entre 10 % et 20 %. Ces taux tiennent compte des taux de refus, qui s'échelonnent entre 5 % et 7 %.

3.2.4 Vérification des données

La vérification des données est effectuée au moment de la collecte, puis en lots dans le cadre des activités de traitement. Il s'agit de vérifier que les champs critiques sont remplis, de s'assurer de la cohérence interne des données et d'appliquer les méthodes de détection des valeurs aberrantes. On utilise à cette fin des vérifications manuelles et, dans certains cas, des vérifications automatisées.

3.2.5 Imputation des données

Lorsqu'un rejet est constaté au contrôle, le champ erroné est signalé à des fins d'imputation. Un processus de correction et de vérification auprès de la personne ayant saisi les données a lieu, dans la mesure du possible. Si le problème persiste, l'enregistrement est déclaré inutilisable pour les traitements ultérieurs.

3.2.6 Estimations

Les estimations des enquêtes sont produites au moyen d'un estimateur Horvitz-Thompson ordinaire. Les cas de non-réponse et les autres enregistrements jugés inutilisables suite à l'étape de vérification sont compensés par un ajustement des poids d'estimation. Les techniques d'ajustement des poids ont également un effet positif sur certaines valeurs aberrantes.

La théorie statistique habituelle produit les indicateurs d'exactitude liés à la variance et au coefficient de variation.

3.2.7 Contrôle de la divulgation

Les règles de divulgation évaluent les risques de divulgation directe et par recoupements pour les cellules totalisées avant la diffusion.

Il peut y avoir divulgation directe si très peu de facteurs apportent une contribution non nulle à l'estimation de la cellule ou si un nombre très limité de facteurs contribuent à un pourcentage très important de l'estimation globale.

On utilise des méthodes automatisées pour repérer les cellules qui comportent des risques directs de divulgation.

Des méthodes manuelles servent à limiter les risques de divulgation par recoupements. Dans les deux cas, les cellules jugées à risque sont supprimées. On prend soin de tenir compte des tableaux préalablement diffusés pour éviter toute divulgation par recoupements entre plus d'un tableau.

3.3 États-Unis

Le programme américain de la statistique agricole se divise en deux parties. Le programme d'estimations agricoles est constitué d'une série d'enquêtes et de publications portant sur plus de 160 produits agricoles et couvrant le secteur de l'agriculture. Des enquêtes sont réalisées à une fréquence établie en collaboration avec chaque secteur, c'est-à-dire sur des bases hebdomadaire, mensuelle, trimestrielle, semestrielle et bisannuelle. Le recensement de l'agriculture, qui a lieu tous les cinq ans, présente un portrait démographique et économique détaillé du secteur agricole américain. Plusieurs enquêtes de suivi liées au recensement sont organisées pendant les années intercensitaires.

3.3.1 Définition des bases de sondage

La première étape de tous les recensements et de toutes les enquêtes est la définition de la population cible. Dans le NASS, la principale base de sondage, qui s'accompagne de données de contrôle, est un **échantillonnage sur liste** des fermes et des ranchs. L'échantillonnage sur liste permet au NASS d'établir de manière efficiente un échantillon des fermes et des ranchs pour la plupart des produits et pour de nombreuses caractéristiques agricoles. Pour le programme d'estimations agricoles, la liste ne comprend que les fermes et

les ranchs actifs qui correspondent à la définition d'une exploitation agricole. Le recensement de l'agriculture cible toutes les exploitations agricoles connues aux États-Unis figurant sur la liste des fermes et des ranchs du NASS. Contrairement à la base liste des estimations agricoles, la liste postale du recensement (LPR) contient également les fermes potentielles dont l'état d'exploitation n'est peut-être pas confirmé. Bien que de nouvelles sources soient continuellement incorporées à la liste, des efforts intensifs sont consacrés avant chaque recensement pour compléter la liste. L'un des problèmes majeurs associés à la base liste et à la LPR est leur nature incomplète. La base liste est considérée complète pour de nombreux produits spéciaux et pour toutes les entreprises agroalimentaires.

Le NASS a également mis sur pied une **base aréolaire** de toutes les terres américaines, qui inclut, par sa nature même, toutes les fermes. La base aréolaire est divisée en segments de terre compris à l'intérieur de strates définies en fonction du degré de culture. La base aréolaire est complète; cependant, un échantillon de segments établi pour un usage général s'avère inefficace pour un grand nombre de produits. Le NASS fusionne ces deux bases à l'aide de plans d'échantillonnage à bases multiples afin de tirer parti de l'efficacité de la base liste et d'utiliser la base aréolaire pour mesurer le degré d'incomplétude. Presque toutes les grandes enquêtes par sondage réalisées par le NASS ont recours à un échantillon à base aréolaire pour mesurer le degré d'incomplétude de la base liste. Il faut à tout prix que la base liste et la base aréolaire restent statistiquement indépendantes.

3.3.2 Plan d'échantillonnage et sélection de l'échantillon

Le NASS emploie plusieurs plans d'échantillonnage différents. Dans le cas des petites populations cibles, comme les éleveurs de vison et les transformateurs de légumes, un recensement est effectué. Pour les enquêtes portant sur un seul produit, comme le bétail, on utilise des plans stratifiés prévoyant un échantillonnage aléatoire simple à l'intérieur des strates. Dans le cas des enquêtes axées sur de multiples produits, comme les superficies cultivées et les stocks de céréales, on utilise des plans d'échantillonnage multivariés avec probabilité proportionnelle à la taille. Les probabilités de sélection sont calculées pour chaque produit d'intérêt présent dans le profil des données de contrôle d'une ferme. Des échantillons à probabilité proportionnelle à la taille sont sélectionnés pour chaque produit, et chaque enregistrement sélectionné reçoit initialement un poids inversement proportionnel à sa probabilité de sélection maximale. Les poids d'échantillonnage sont étalonnés en fonction des totaux de la base de sondage pour les produits d'intérêt. Pour certaines enquêtes, on utilise des nombres aléatoires permanents pour limiter le fardeau du répondant.

3.3.3 Étapes de la collecte des données

Le NASS a recours à toutes les formes de collecte des données, soit la poste, le téléphone, les interviews en personne et Internet. La majorité des enquêtes sur les estimations agricoles ont recours à une combinaison de moyens de collecte, mais la plupart des données sont recueillies par voie téléphonique. Les formulaires de recensement sont envoyés par la poste, et le suivi des cas de non-réponse est fait par le truchement du téléphone. Un nombre limité d'interviews sur place sont menées pour les fermes de plus grande taille. Les interviews téléphoniques assistées par ordinateur, les interviews sur place assistées par ordinateur et les instruments d'enquête Web permettent une vérification de base en temps réel. Les enquêtes

à base aréolaire sont menées sur place uniquement à l'aide d'interviews. L'ARMS, une grande étude économique, est réalisée au moyen d'interview sur place, et une version postale optionnelle abrégée est acheminée à certains répondants. Toutes les enquêtes entreprises par le NASS comportent une date de référence définie. Cette date peut être une journée donnée ou s'étendre sur une période de temps plus longue. Par exemple, la cueillette des données sur les dépenses et les revenus porte sur une année complète, alors que celle des données sur les prix obtenus pour les récoltes couvre habituellement un mois. Les estimations de la superficie et des rendements définitifs portent sur la saison de croissance. Les éléments appelés à fluctuer constamment, comme les stocks de bétail et les prévisions de rendement, sont mesurés pour une date particulière. Toutes les publications du NASS affichent clairement la période de référence.

3.3.4 Vérification des données

La vérification des données peut être réalisée au moment de la collecte et en lots dans le cadre des activités de traitement. La méthode est constituée d'une combinaison d'activités : la vérification des non-réponses, la comparaison des réponses avec les données antérieures, le contrôle de la cohérence, la vérification statistique et la détection des valeurs aberrantes. En général, les formulaires papier des instruments d'enquête font en plus l'objet de vérifications manuelles.

3.3.5 Imputation des données

Le NASS a plusieurs stratégies d'imputation à sa disposition pour traiter les cas de non-réponse. Il choisit la ou les stratégies jugé(e)s les plus appropriées. Certaines de ces stratégies d'imputation valident les données en leur adjoignant une valeur, alors que d'autres ajustent la pondération de l'échantillon. Certaines stratégies utilisent une ferme déclarante d'une taille et d'un type analogues, alors que d'autres forment des groupes d'imputation d'une taille et d'un type analogues. Les algorithmes d'imputation automatisés sont fondés sur la caractéristique fondamentale de l'homogénéité, peu importe qu'il s'agisse de fermes individuelles ou de groupes de fermes. En général, les fermes très grandes ou uniques ne peuvent être intégrées à des groupes homogènes, ni jumelées à des fermes analogues. Par conséquent, ces fermes doivent être imputées manuellement par des statisticiens.

3.3.6 Estimations

La plupart des sommaires d'enquête produisent des estimations ponctuelles multiples, souvent une combinaison de mesures directes et d'estimations de ratios. Le calcul des écarts-types et des coefficients de variation est fondé sur la théorie statistique habituelle. Les sommaires sont exécutés au niveau des États et, pour les enquêtes nationales, les résultats des États sont regroupés au niveau national.

Les bureaux régionaux sont chargés de l'examen minutieux des données, de l'interprétation des résultats des enquêtes et de la soumission de recommandations au bureau central. Le NASS a recours à une approche descendante pour l'établissement des estimations

officielles, l'adoption de l'estimation nationale et l'ajustement des recommandations d'États sélectionnés afin d'assurer la participation des États au total national.

Pour les produits clés, les enquêtes sont conçues pour atteindre des coefficients de variation de 1 % à 3 % pour les estimations nationales et de 5 % à 10 % pour les estimations faites au niveau des États.

3.3.7 Contrôle de la divulgation

La législation fédérale oblige le NASS à retenir toute estimation qui donnerait lieu à la divulgation de données agricoles individuelles ou qui permettrait au public de connaître, ne serait-ce que d'une manière approximative, les caractéristiques d'une ferme. Le NASS a recours à une règle de divulgation double pour déterminer les suppressions primaires. La première règle exige qu'au moins trois fermes produisent une denrée pour que la publication des données sur cette denrée soit publiée. La seconde règle, qui est celle de la prédominance, protège la confidentialité des données des fermes de très grande taille. Le programme d'estimations agricoles utilise la règle (n,k) pour déterminer la prédominance. Cette règle supprime une valeur si la ferme n la plus importante produit un pourcentage k ou plus du total. Le programme du recensement a recours à la règle p-pour cent, qui supprime la valeur si le reliquat, après le retrait des données des deux plus grandes fermes, n'assure pas une protection p-pour cent pour la ferme la plus importante. Le NASS ne divulgue pas les paramètres utilisés dans la règle de prédominance. On appelle suppression primaire la suppression d'une estimation qui ne respecte ni l'une ni l'autre des règles de divulgation. Il arrive souvent qu'une estimation faisant l'objet d'une suppression primaire fasse partie d'un total publié et qu'une seconde suppression (complémentaire) soit requise pour protéger la suppression primaire.

3.3.8 Mise à jour des bases de sondage

La mise à jour de l'échantillonnage sur liste et de la base aréolaire du NASS est réalisée en continu dans le cadre de la procédure standard d'enquête. En général, les mises à jour sont effectuées dès la fin de l'enquête ou du recensement.

4. MÉTHODES D'AJUSTEMENT

Pendant l'étape des estimations, les programmes de la statistique agricole administrés par les trois pays doivent prendre en compte le sous-dénombrement, entre autres problèmes, et ce, avant de diffuser les résultats.

4.1 Canada

4.1.1 Couverture

La couverture des enquêtes canadiennes dépend fortement de leur base de sondage et donc du Recensement de l'agriculture, qui alimente le Registre des entreprises. En dépit des efforts que les agents recenseurs déploient pour localiser et dénombrer toutes les

exploitations agricoles au Canada, le Recensement de l'agriculture oublie certaines fermes, principalement en raison des difficultés liées à l'identification d'une exploitation agricole dont aucun des exploitants ne vit sur la ferme ou à proximité de celle-ci. Pour réduire le sous-dénombrement, depuis 1991, une question-filtre visant les exploitants agricoles figure sur le questionnaire du Recensement de la population. Cette question sert à identifier les exploitants agricoles qui auraient été oubliés lors de la livraison du questionnaire. Les agents de l'enquête de suivi des fermes manquantes peuvent alors appeler les répondants qui ont répondu « oui » à cette question, afin de compléter le questionnaire du Recensement de l'agriculture.

Contrairement aux Recensements de l'agriculture antérieurs, le Recensement de 2011 ne comportait aucune Enquête d'évaluation de la couverture explicite conçue pour les ménages qui n'ont pas fait l'objet d'un appel téléphonique dans le cadre de l'exercice de suivi des fermes manquantes, de même que pour les exploitations de la base du recensement qui n'ont pas retourné le questionnaire pendant la période d'activité du Recensement de l'agriculture. On a plutôt calculé un taux estimatif de sous-dénombrement en fonction des résultats observés du processus de suivi des fermes manquantes. Par conséquent, le Recensement de 2011 a affiché un taux de sous-dénombrement de 1,8 % pour les fermes et un taux de sous-dénombrement inférieur à 1 % pour la superficie agricole totale et les ventes agricoles totales.

Dans le cadre du Programme de la statistique agricole, les résultats des enquêtes sont comparés à ceux des recensements. Par conséquent, les résultats de la couverture du Recensement de l'agriculture sont recyclés pour les enquêtes.

4.1.2 Autres ajustements pour les concepts

Comme il a été mentionné au point 3.1.1, les bases de sondage ciblent toutes les exploitations qui affichent un revenu agricole positif, bien que les exploitations dont le revenu est inférieur à 10 000 \$CAN soient exclues de l'échantillon. Ces exploitations sont plutôt prises en compte grâce à des facteurs d'ajustement, sans mesure des erreurs.

4.2 Mexique

4.2.1 Couverture

La couverture thématique s'applique aux secteurs de l'agriculture, de l'élevage, de l'aquaculture et de la pêche.

La couverture selon le type d'unité de production (tous les biens et services et les terrains gérés par une même administration) regroupe toutes les unités de production, et ce, peu importe la taille de l'exploitation, le type de propriété (communautaire, privée ou publique) ou le montant des revenus.

Depuis 2009, le SAGARPA développe un système d'information qui repose sur des méthodes objectives telles que l'échantillonnage et la géomatique et qui utilise comme intrants les matériaux et les résultats du recensement de l'agriculture et du bétail. Ce

système sera appliqué sur les principales cultures et zones de production. Pour que l'enquête couvre toutes les cultures, on complètera avec des informations de nature administrative pour les cultures restantes.

4.2.2 Autres ajustements

Les deux systèmes (registres administratifs et recensements) tiennent compte du même nombre de producteurs sans faire de distinction entre les groupes.

Par conséquent, il est inutile d'utiliser une méthode d'ajustement pour estimer une éventuelle sous-estimation ou surestimation.

4.3 États-Unis

4.3.1 Couverture

Des mesures d'ajustement de la couverture sont utilisées pour résoudre le problème d'incomplétude des listes, lesquelles servent également à s'assurer de la représentation adéquate de l'échantillon de répondants. L'incomplétude des listes est de loin le défi le plus important, et le NASS intègre des mesures de l'incomplétude dans ses enquêtes et ses recensements. Chaque année, en juin, le NASS procède à un dénombrement complet des segments échantillonnés de la base aréolaire et identifie les personnes qui exploitent les fermes pour chaque segment. Les noms de ces personnes sont comparés à ceux de la liste afin de déterminer les fermes qui figurent dans les populations échantillonnées (chevauchement) et celles qui n'y figurent pas (non-chevauchement). Les fermes non chevauchantes sont sondées en vue de mesurer directement ce qui n'est pas couvert par la liste. La représentativité est un aspect important des enquêtes et des recensements qui portent sur divers secteurs agricoles (bovins, grandes cultures, fruits). Après les ajustements pour non-réponse et le sous-dénombrement, certains secteurs peuvent être surreprésentés, alors que d'autres sont sous-représentés, ce qui donne lieu à des estimations ponctuelles biaisées. Les ajustements de la couverture sont réalisés en étalonnant les poids d'échantillonnage de façon à ce qu'ils correspondent aux cibles externes. En termes de procédure, l'étalonnage trouve une solution à un système d'équations linéaires (contraintes) définies par les cibles.

Estimations agricoles

Pour toutes les enquêtes nationales, l'incomplétude des listes est mesurée directement à l'aide des secteurs de non-chevauchement. Pour les enquêtes sur les cultures et le bétail, les sommaires traitent l'échantillon comme étant représentatif, ce qui ne nécessite aucun ajustement. En ce qui concerne l'enquête annuelle sur l'économie, les poids sont étalonnés en fonction de cibles de production connues.

Recensement de l'agriculture

Les cibles d'incomplétude de la LPR sont définies en ajoutant un compte des fermes manquantes mesuré directement au compte des répondants du recensement ajusté pour la

non-réponse pour plus de 60 variables. Les cibles de production qui ne proviennent pas de sources du recensement sont utilisées pour évaluer la représentativité. Les deux ensembles de cibles sont fusionnés, et l'ajustement des poids se fait grâce à l'étalonnage. Ainsi, l'incomplétude et la représentativité sont prises en compte simultanément dans chaque État.

4.3.2 Imputation

Estimations agricoles

Pour les enquêtes sur le bétail, la vérification identifie les cellules dont les éléments doivent être imputés et renvoie ces cellules à un analyste, qui procédera à la résolution manuelle. La non-réponse des unités est réalisée par une repondération au sein des groupes. Les enquêtes sur le bétail constituent des plans stratifiés traditionnels. Les groupes de pondération sont donc intégrés au plan. Les enquêtes sur la superficie et la production ont recours à un algorithme personnalisé pour les cas de non-réponse partielle et totale, et pour les produits clés qui font l'objet d'une imputation en raison de données manquantes. L'algorithme a recours à des groupes de pondération fondés sur la taille et le type, subdivisés par région, dans le but de créer des groupes plus homogènes pour les modes de culture agricole et le rendement des cultures. L'enquête économique annuelle utilise des groupes de pondération selon la taille et le type. Les groupes dont la taille de l'échantillon ne respecte pas la taille minimale sont combinés à d'autres groupes, jusqu'à ce qu'un nombre suffisant de rapports soit obtenu.

Recensement de l'agriculture

L'imputation des éléments est réalisée grâce à la validation des données. Lorsque les données manquantes ne peuvent être déterminées à partir d'autres données déclarées et que l'élément n'a pas déjà été déclaré dans une autre enquête du NASS, un algorithme du « plus proche voisin » est utilisé. On définit une ferme analogue comme étant la distance euclidienne entre le destinataire et les donneurs éventuels (rapports propres) à l'aide de variables sélectionnées dans la LPR. Chaque section du formulaire de rapport possède des variables d'appariement spécifiques, qui comprennent la latitude et la longitude du principal comté opérationnel. Le donneur le plus près est considéré comme le plus proche voisin et devient ainsi la source de l'imputation. Étant donné que les imputations sont réalisées indépendamment pour chaque occurrence, les rapports qui nécessitent des imputations multiples sont fondés sur des donneurs multiples. Les réserves initiales de donneurs sont établies avant l'exécution des vérifications du premier lot. Ces réserves de donneurs reçoivent des données tirées des recensements antérieurs; ces données sont « mises en correspondance » pour qu'elles ressemblent aux données actuelles de recensement, elles sont assujetties à la vérification en cours pour enfin être complétées par les données recueillies pendant la mise à l'essai du contenu du recensement. Si l'algorithme d'imputation ne peut fournir de solution acceptable, l'enregistrement est transféré à un analyste.

Des ajustements pour non-réponse sont effectués pour des fermes entières à l'aide d'estimations modélisées de fermes comprises à l'intérieur du champ au sein de groupes de pondération au niveau des États. La LPR est conçue de manière à pouvoir contenir

beaucoup plus de noms qu'il n'y a de fermes admissibles. Par conséquent, de nombreux non-répondants ne sont pas des fermes. Les groupes de pondérations sont circonscrits au moyen de techniques d'exploration des données et d'une méthode progressive. Toutes les fermes qui figurent sur la LPR sont classées en fonction d'une liste de variables prédéterminées. Des tests de signification sont réalisés sur les taux de réponse de chaque groupe. Si l'écart est significatif, le groupe est établi de façon permanente et la variable suivante est testée pour les noms de chaque groupe. Si l'écart n'est pas significatif, les noms sont combinés de nouveau, et la variable suivante est testée. Lorsque toutes les variables ont été testées et que les groupes d'imputation définitifs ont été définis, le taux de répondants à l'intérieur du champ est calculé, puis appliqué au compte de non-répondants afin d'estimer le nombre de fermes de la LPR qui n'ont pas été prises en compte. Les poids des répondants du groupe sont rajustés en vue de représenter les fermes manquantes.

5. PROCESSUS D'ASSURANCE DE LA QUALITÉ ET INDICATEURS DE LA QUALITÉ DES DONNÉES

Des méthodes d'assurance de la qualité sont mises en œuvre tout au long des diverses phases du processus de recensement ou d'enquête. En outre, des indicateurs de la qualité des données servent à décrire de manière quantifiable la qualité des données ou du processus.

5.1 Canada

5.1.1 Pratiques d'assurance de la qualité des données

Les enquêtes et les recensements comportent tous un certain degré d'assurance de la qualité. Dans le cas du Recensement de l'agriculture mené au Canada, des plans d'assurance de la qualité spécialement conçus et bien définis sont en place pour les étapes les plus importantes. Pour d'autres enquêtes, des pratiques plus générales sont utilisées.

5.1.1.1 Conception des questionnaires

Le Centre de ressources en conception de questionnaires de Statistique Canada examine et met à l'essai tout nouveau questionnaire ou tout questionnaire ayant fait l'objet de changements importants. Il peut s'agir d'un simple examen des questions ou d'un processus plus complexe qui consiste à mettre à l'essai les questions, leur ordre et leur enchaînement auprès d'un échantillon de répondants dans le cadre d'un groupe de discussion. Les gestionnaires d'enquête reçoivent un rapport sur les constatations du Centre, y compris les changements suggérés au besoin.

5.1.1.2 Bases de sondage

Les enquêtes et le Recensement de l'agriculture ont une relation mutuellement avantageuse pour ce qui est de maintenir la qualité des bases de sondage. Le recensement fournit une liste complète des unités qui sert de point de départ aux bases de sondage. Les mises à jour consécutives aux enquêtes font en sorte que la base de sondage reste pertinente et à jour. De plus, une enquête de dépistage des fermes utilise les renseignements provenant de sources

externes comme les associations de producteurs pour repérer d'éventuelles nouvelles fermes. On envoie ensuite un questionnaire d'enquête à ces fermes pour recueillir des renseignements sur leurs activités. À partir de ces renseignements, la ferme peut être ajoutée ou pas à la base de sondage.

5.1.1.3 Plan d'échantillonnage, sélection de l'échantillon et fardeau de réponse

Parce que les enquêtes agricoles sont propres aux activités, on peut rendre les plans d'enquête très efficaces en utilisant les renseignements de la base de sondage sur ces activités. Il est ainsi possible d'obtenir des estimations justes avec un échantillon de taille réduite. On utilise des méthodes statistiques pour limiter le chevauchement de l'échantillon à même les enquêtes et entre celles-ci afin de réduire le fardeau de réponse. Pour la plupart des enquêtes, les outils logiciels généralisés de Statistique Canada servent à la sélection de l'échantillon.

5.1.1.4 Collecte des données

Les interviews aux fins des enquêtes sont menées à partir des bureaux régionaux de Statistique Canada. Les intervieweurs suivent une formation particulière relative aux enquêtes agricoles. En outre, on confie à des intervieweurs spécialisés les cas plus difficiles tels que la conversion des refus. Un système de collecte intégré permet d'importer l'échantillon, d'attribuer les cas particuliers aux intervieweurs et de renvoyer au bureau central les données recueillies. Les gestionnaires des bureaux régionaux reçoivent chaque jour des renseignements sur le déroulement des enquêtes qui les aident à gérer la charge de travail. Pour la plupart des enquêtes, les données sont recueillies au moyen d'une application d'interview téléphonique assistée par ordinateur à laquelle sont intégrées d'importantes vérifications. Il est ainsi possible de relever les erreurs et les incohérences pendant l'interview et de les régler immédiatement. Dans le cas des enquêtes plus complexes, un exemplaire du questionnaire est envoyé d'avance au répondant pour qu'il puisse l'examiner et préparer les réponses avant l'interview.

5.1.1.5 Traitement des données

Des procédures sont en place pour le contrôle de la qualité des données qui doivent être saisies ou codées manuellement. Un système propre à l'agriculture est utilisé pour le traitement de la presque totalité des enquêtes. Les analystes peuvent examiner des enregistrements particuliers et comparer les données déclarées avec celles qui proviennent d'occasions d'enquête précédentes. Les enregistrements incomplets font l'objet d'une imputation, soit à partir des connaissances des analystes, soit selon des méthodes automatisées. Il existe un certain nombre de méthodes d'imputation automatisées, dont l'imputation historique, l'imputation par enregistrement donneur ou l'imputation par la moyenne.

5.1.1.6 Estimation et diffusion

La pondération et l'estimation sont fondées sur des méthodes reconnues. On utilise des données auxiliaires du recensement pour ajuster les estimations de l'enquête afin de tenir

compte de la population non sondée. Des comparaisons sont établies entre les résultats de l'enquête et d'autres sources de données accessibles telles que les estimations provenant des associations de producteurs. Chaque sujet est accompagné de documents décrivant les méthodes et les approches utilisées afin que l'utilisateur comprenne la façon dont les estimations ont été produites.

5.1.2 Indicateurs de la qualité des données

Des indicateurs quantitatifs de la qualité des données sont généralement produits à deux étapes du cycle de production, à savoir pendant la collecte des données et à l'étape du traitement et de l'estimation.

5.1.2.1 Indicateurs de la qualité qui se rattachent à la collecte des données

Des taux de réponse sont utilisés tout au long du processus d'enquête. À mesure que la collecte se déroule, ces taux sont surveillés et comparés aux taux de réponse cibles. Dans les cas où les taux observés sont nettement inférieurs aux cibles, la période de collecte peut être prolongée. Deux types de taux de réponse sont calculés : des taux non pondérés, ou bien des taux pondérés selon une valeur du Registre des entreprises qui se rapporte aux variables d'intérêt de l'enquête. Une distinction est également établie entre les taux de réponse de collecte et les taux de réponse définitifs. Ces premiers représentent le pourcentage des cas considérés comme des répondants à la suite de la collecte des données, tandis que ces derniers représentent le pourcentage des répondants une fois que les données ont été nettoyées et traitées.

Les raisons de la non-réponse font également l'objet d'un suivi au fil du temps. Si l'on observe une augmentation d'un certain type de non-réponse comme les refus, il se peut que les consignes données aux intervieweurs soient révisées. On s'appuie également sur les raisons de la non-réponse pour apporter les ajustements appropriés aux poids d'échantillonnage afin de tenir compte des non-répondants.

Dans le cadre du recensement, l'erreur de couverture est déterminée en fonction des résultats du processus de suivi des fermes manquantes décrit à la section 4.1.1. Des estimations du sous-dénombrement sont établies non seulement pour le nombre de fermes, mais également pour le nombre de principaux produits.

On utilise de plus en plus des parodonnées, ou des données liées à la collecte même des données, pour fournir des indicateurs relatifs au processus de collecte des données tels que le temps qu'il faut pour réaliser une interview ou le fardeau imposé au répondant.

5.1.2.2 Indicateurs de la qualité qui se rattachent aux estimations

Pendant la conception de l'échantillon, un ensemble de cibles de précision sont établies pour chaque variable d'intérêt importante. Ces cibles sont exprimées sous forme de coefficients de variation (CV). On répartit l'échantillon de manière à limiter le plus possible le nombre d'unités avec lesquelles un contact doit être établi, tout en respectant ces cibles. Pendant l'estimation, les CV véritables de l'enquête sont calculés. Les CV observés sont

comparés aux CV théoriques, et l'on fait enquête sur les différences importantes afin de les expliquer. Pour la publication des CV d'estimations particulières, on utilise souvent une cote alphabétique plutôt que la valeur réelle du CV. À chaque estimation est attribuée une lettre de A à F qui en indique la qualité.

Les principales estimations sont comparées d'un cycle d'enquête à l'autre. En cas de différences importantes, on fait enquête pour en déterminer la raison. Les estimations sont également comparées à celles qui proviennent d'autres sources de données, comme une enquête, un recensement ou des sources administratives. Les différences importantes font l'objet d'une enquête.

5.2 Mexique

5.2.1 Pratiques d'assurance de la qualité des données

Les enquêtes et les recensements comportent tous un certain degré d'assurance de la qualité. Certaines pratiques générales telles que la cohérence interne et les filtres statistiques sont utilisées, mais elles dépendent de l'enquête.

5.2.2 Indicateurs de la qualité des données

Des procédures sont en place pour le contrôle de la qualité des données qui doivent être saisies ou codées manuellement. Un système propre à l'agriculture est utilisé pour le traitement des enquêtes.

5.2.2.1 Indicateurs de la qualité qui se rattachent à la collecte des données

Aucun indicateur de la qualité n'est expressément utilisé ou consigné au cours de ce processus, mais seulement le nombre d'enregistrements qui ont été consignés et modifiés.

Des taux de réponse sont utilisés tout au long du processus d'enquête.

Deux types de taux de réponse sont calculés : des taux non pondérés, ou bien des taux pondérés selon une valeur de la base liste qui se rapporte aux variables d'intérêt de l'enquête.

Une distinction est également établie entre les taux de réponse de collecte et les taux de réponse définitifs. Ces premiers représentent le pourcentage des cas considérés comme des répondants à la suite de la collecte des données, tandis que ces derniers représentent le pourcentage des répondants une fois que les données ont été nettoyées et traitées.

Les raisons de la non-réponse font également l'objet d'un suivi au fil du temps. Si l'on observe une augmentation d'un certain type de non-réponse comme les refus, il se peut que les consignes données aux intervieweurs soient révisées. On s'appuie également sur les raisons de la non-réponse pour apporter les ajustements appropriés aux poids d'échantillonnage afin de tenir compte des non-répondants.

5.2.2.2 Indicateurs de la qualité qui se rattachent aux estimations

Pendant la conception de l'échantillon, un ensemble de cibles de précision sont établies pour chaque variable d'intérêt importante.

Ces cibles sont exprimées sous forme de coefficients de variation (CV). On répartit l'échantillon de manière à limiter le plus possible le nombre d'unités avec lesquelles un contact doit être établi, tout en respectant ces cibles.

Les principales estimations sont comparées d'un cycle d'enquête à l'autre. En cas de différences importantes, on fait enquête pour en déterminer la raison. Les estimations sont également comparées à celles qui proviennent d'autres sources de données, comme une enquête, un recensement ou des sources administratives. Les différences importantes font l'objet d'une enquête.

5.3 États-Unis

5.3.1 Pratiques d'assurance de la qualité des données

Des mesures d'assurance de la qualité sont mises en œuvre pour l'ensemble des enquêtes et des recensements menés par le NASS. L'établissement d'un ensemble de politiques et de normes permet d'assurer une application judicieuse et uniforme des procédures aux activités fréquentes. Une norme est établie pour toute activité statistique qui est exécutée fréquemment dans des conditions semblables, que ce soit au fil du temps, d'une enquête à l'autre, ou dans l'ensemble des unités organisationnelles.

5.3.3.1 Conception des questionnaires

Le NASS procède à l'essai préalable des instruments d'enquête pour veiller à ce que chaque question de chaque instrument d'enquête permette de saisir de manière efficace et exacte les données appropriées aux estimations qu'elles appuient.

Tous les instruments d'enquête pour le recensement et les autres enquêtes fédérales sont coordonnés par la division des recensements et des enquêtes du bureau central du NASS et sont diffusés aux bureaux régionaux des États. La conception et l'essai des questionnaires sont coordonnés par les méthodologistes d'enquête de la division de la recherche et du développement du NASS. Les instruments ITAO et IPAO et les questionnaires Web fournis aux bureaux régionaux doivent être conformes au questionnaire papier de chaque enquête. La conception et le contenu des questionnaires respectifs tiennent compte des différences qui existent entre les modes de collecte des données (questionnaire à remplir soi-même, par interview sur place ou au téléphone). Tous les instruments d'enquête Web, sur papier et assistés par ordinateur pour les enquêtes fédérales sont assujettis à l'examen et à l'approbation de l'Office of Management and Budget des États-Unis.

Un certain nombre de grandes enquêtes suivent un processus officiel de spécifications pour le traitement des demandes de changement de libellé ou d'ajout ou de suppression de questions.

5.3.3.2 Bases de sondage

Il incombe au NASS de tenir une liste à jour des exploitations agricoles aux États-Unis, conjointement avec des indications de la taille et du type de l'exploitation. Ces « données de contrôle » sont importantes pour l'élaboration de plans d'échantillonnage efficaces. **Tous les types et toutes les tailles d'exploitation agricole** doivent être représentés sur la liste. Pour assurer l'utilisation la plus efficace possible des ressources affectées à cette tâche, le NASS fixe des priorités relatives à l'établissement de la couverture des différents sous-groupes du secteur agricole.

Le conseil des bases listes, la division des statistiques et le bureau des opérations sur le terrain se consultent afin d'établir les objectifs de couverture des bases listes à l'échelon national et au niveau des États, selon le cas, pour l'ensemble des exploitations agricoles, d'importants sous-groupes démographiques et du secteur agricole et la production de produits importants.

Des enregistrements de base liste représentant les établissements agricoles pour lesquels une activité agricole a été déclarée dans un recensement ou un instrument d'enquête récent sont tenus à jour en tant qu'enregistrements actifs dans la base liste. Cela comprend les enregistrements qui ne satisfont pas nécessairement au critère de ventes de 1 000 \$, mais qui font état d'une activité agricole.

5.3.3.3 Plan d'échantillonnage, sélection de l'échantillon et fardeau de réponse

Le NASS a recours à plusieurs procédures d'échantillonnage et d'estimation pour réduire la fréquence des contacts multiples avec les exploitations échantillonnées. On utilise l'échantillonnage séquentiel par intervalles de Poisson pour sélectionner des échantillons aux fins de la phase III de l'enquête sur la gestion des ressources agricoles (ARMS) de façon à limiter le plus possible le chevauchement entre l'échantillon de l'ARMS de l'année courante, l'échantillon de l'ARMS III de l'année précédente et l'échantillon de l'enquête agricole trimestrielle de l'année courante. On se sert aussi de l'échantillonnage séquentiel par intervalles de Poisson pour éviter tout chevauchement entre l'échantillon de l'enquête sur le rendement agricole des petites céréales et l'échantillon de l'enquête sur le rendement agricole des cultures en rangs dans le programme de l'enquête agricole trimestrielle.

L'échantillonnage répété sert généralement à limiter le chevauchement des échantillons prélevés dans la même population. On utilise la méthode des nombres aléatoires permanents (NAP) pour attribuer un NAP à chaque membre de la population échantillonnée admissible de certaines enquêtes. La sélection des échantillons dans des fourchettes de NAP données permet de limiter le chevauchement des échantillons. La modélisation du non-chevauchement avec la liste (NCL) consiste à utiliser les données déclarées des exploitations de taille comparable de la liste pour modéliser la partie NCL de l'estimateur de façon à éviter tout contact avec des exploitations NCL pendant les périodes d'enquête subséquentes des enquêtes à bases multiples.

Toutes les enquêtes dont les périodes de collecte des données se chevauchent sont coordonnées de manière que la collecte des données soit effectuée en un seul contact, à moins que l'exploitant ne demande des contacts distincts. Les enquêtes non mensuelles dont les dates de référence sont séparées de tout au plus un mois sont coordonnées de façon que la collecte des données en un seul contact soit effectuée le plus tard possible au cours de la période de collecte des données de l'enquête qui a lieu le plus tôt.

5.3.3.4 Collecte des données

Les activités de collecte des données de la plupart des enquêtes fédérales sont coordonnées par la division des recensements et des enquêtes du NASS. La collecte des données est menée par les bureaux régionaux des États et les centres régionaux de collecte des données.

Aux termes d'un accord de collaboration avec le NASS, la National Association of State Departments of Agriculture (NASDA) fournit un corps de recenseurs qualifiés aux fins de la collecte des données agricoles. Les deux organismes visent à offrir un programme de formation pratique et efficace aux employés de la NASDA. Un programme de formation efficace contribue à uniformiser les méthodes de collecte des données, à limiter les erreurs de collecte des données et à assurer des statistiques agricoles de grande qualité.

Comme le prévoit la politique du NASS et de la NASDA, les recenseurs mènent des activités de collecte des données uniquement après avoir suivi une formation appropriée. Il incombe au coordonnateur de la NASDA des bureaux régionaux (statisticien principal d'enquêtes) et aux recenseurs principaux de mettre en œuvre la formation générale et propre aux différentes enquêtes.

5.3.3.5 Traitement des données

Un système sommaire généralisé SAS (GenSumm) permet de regrouper les données d'enquête, de calculer les niveaux de précision, de fournir des données analytiques supplémentaires et d'afficher les résultats dans un format facile à lire. Il est fondé sur le système sommaire SPS que le NASS utilise depuis le début des années 1990. Les deux outils ont recours aux mêmes programmes sous-jacents pour calculer les indications et les données s'y rattachant et offrir une analyse cohérente et statistiquement valable.

5.3.3.6 Estimation et diffusion

Les estimations relatives aux produits sont établies au moyen d'un processus statistique prédéterminé, qui part des données tirées des enquêtes auprès des producteurs et des négociants agricoles. Toutes les données d'enquête sont validées grâce à un processus détaillé de vérification et d'analyse. Dans la mesure du possible, on compare ces données d'enquête aux données administratives de diverses sources et aux connaissances acquises au fil des voyages et des contacts personnels fréquents avec des participants à la production, au marketing et à l'administration des produits agricoles pour veiller à ce que les données publiées soient raisonnables et fiables et qu'elles représentent les tendances actuelles en matière d'agriculture.

À peu d'exceptions près, les données sont recueillies, vérifiées, analysées et résumées au niveau des États. Les estimations pour les États sont préparées en vue d'être présentées à l'Agricultural Statistics Board (ASB). Ce groupe d'experts en produits examine les recommandations particulières et combinées des États et il adopte et publie les estimations étatiques, régionales et nationales.

5.3.4 Indicateurs de la qualité des données

Des analyses quantitatives des données sont produites pendant la collecte, le traitement et l'estimation des données. L'Office of Management and Budget (OMB) exige une analyse quantitative de toutes les activités d'enquête fédérales qui sont assujetties à son approbation.

5.3.4.1 Indicateurs de la qualité qui se rattachent à la collecte des données

Pour chaque effort de collecte des données, une ou plus d'une procédure d'assurance de la qualité est utilisée. Il incombe aux recenseurs principaux et aux statisticiens d'assurer la qualité des données en accomplissant les tâches suivantes : (1) établir un nouveau contact avec un sous-ensemble des répondants pour vérifier les données; (2) examiner une partie ou la totalité des questionnaires papier remplis pour s'assurer qu'ils sont dûment remplis, que l'enchaînement des questions est approprié et que les notes le sont aussi; (3) surveiller les interviews pour s'assurer que les recenseurs utilisent des procédures d'interview efficaces et pour déterminer les secteurs où une formation supplémentaire se révèle nécessaire. À des fins de surveillance, les recenseurs sur le terrain sont périodiquement accompagnés d'un recenseur principal à une ou plusieurs interviews. Les recenseurs travaillant par téléphone sont observés à l'aide de matériel électronique (le cas échéant) qui permet au superviseur de surveiller les écrans ITAO et de suivre la conversation pour s'assurer que les intervieweurs suivent les procédures établies.

L'Office of Management and Budget fixe les taux de réponse cibles des enquêtes menées par des organismes fédéraux. Les taux de réponse et le caractère adéquat de la couverture font l'objet d'un contrôle tout au long du processus de collecte des données. On calcule des taux de réponse pondérés aussi bien que non pondérés.

Un rapport contenant une analyse détaillée des taux de réponse entre les périodes d'enquête et au fil du temps est établi pour chaque grand programme d'enquête probabiliste. Ces analyses détaillées, qui sont réalisées au niveau des États et à l'échelon national, comprennent les raisons de la non-réponse telles que les refus ou les répondants inaccessibles, la non-réponse selon les modes de collecte des données, ainsi que la non-réponse selon le type de répondant (exploitant, conjoint, teneur de livres, etc.).

Une enquête d'évaluation de la couverture du Recensement de l'agriculture permet de fournir une mesure de l'erreur de couverture. Des estimations du sous-dénombrement sont établies pour le nombre de fermes, les principaux produits et certaines catégories démographiques.

5.3.4.2 Indicateurs de la qualité qui se rattachent aux estimations

Le NASS établit des normes de précision statistique des indications découlant des grandes enquêtes probabilistes nationales. Les normes de précision statistique des indications sont exprimées sous forme de coefficients de variation (CV).

Des niveaux de précision sont établis pour des indications d'enquête particulières. La détermination des valeurs cibles tient compte de six facteurs, à savoir (sans ordre particulier) : 1) l'utilisation prévue de l'estimation par les utilisateurs de données; 2) les incidences sur les mesures de rendement stratégique de l'organisme; 3) la compatibilité entre les cibles régionales et étatiques et les cibles américaines; 4) l'ampleur relative de l'estimation; 5) la réalisabilité étant donné les ressources budgétées de l'organisme, sa charge de travail, le fardeau de réponse, ainsi que la qualité des données d'échantillonnage; 6) l'expérience historique.

Les responsabilités relatives à l'évaluation du rendement de chaque enquête sont confiées aux unités compétentes de l'organisme. On évalue les éléments particuliers du CV (c.-à-d. les éléments de la base liste et de la base aréolaire) et on formule des recommandations pour permettre à chaque enquête d'atteindre régulièrement le CV cible. Ces recommandations peuvent comprendre des modifications du plan d'échantillonnage et de la taille de l'échantillon.

Dernière mise à jour : Juin 2013