



**Méthodologie d'imputation et de
validation pour le domaine de production de
FAOSTAT**

Groupe Technique Régional – UEMOASTAT

Moussa Kaboré

3-7 Juin 2013

Abidjan, Côte d'Ivoire

Sommaire

- Contexte et justification
- Pourquoi des données manquantes?
- Pourquoi a-t-on besoin de faire des imputations au niveau national ?
- Comment les institutions internationales comme la FAO procèdent?
- Quelles sont les forces et les faiblesse?
- Quelles alternatives au niveau des Pays?



- ❖ Les services de statistiques agricoles des Pays n'ont une habitude d'élaboration des cadres macro (bilans alimentaires, comptes économiques de l'agriculture) ;
- ❖ Les services de statistiques agricoles n'ont pas une pratique courante des imputations ;
- ❖ Les Pays ne produisent pas toutes les statistiques dont la FAO ou l'UEMOA a besoin pour l'établissement des bilans alimentaires ou renseigner des indicateurs de performance de l'agriculture et suivi du secteur rural;
- ❖ Il y a un besoin de données exhaustives et complètes pour l'agrégation des données à des niveaux supranational;
- ❖ Les instructions internationales procèdent à des estimations et à des imputations pour combler les données manquantes ;



L'absence de données dans CountrySTAT est due à divers causes dont :

- ❖ Les données sont produites mais il manque un mécanisme de centralisation et dans un cadre unique;
- ❖ Les données sont produites mais le service responsable de la centralisation l'ignore ;
- ❖ Les données sont produites mais les services ne coopèrent pas dans l'échange des données;
- ❖ Les données sont produites mais ne sont pas exhaustives (ex : pêche industrielle, pêche maritime artisanale, pêche continentale);
- ❖ Les données ne sont pas produites parce qu'il n'y a pas d'enquête courante ou système statistique administratif qui permette de les élaborer ;
- ❖ Les données ne sont pas produites faute de méthodologie, de ressources technique ou financière pour les établir;

Les imputations sont rendues nécessaires pour les raisons suivantes :

- ❖ Reconstituer les bilans alimentaires complètes et exhaustives sur des séries historiques;
- ❖ Etablir des indicateurs de performances de l'agriculture et de sécurité alimentaire sur des séries historiques;
- ❖ Impossibilité de réaliser des enquêtes rétrospectives pour reconstituer les données manquantes, non exhaustives ou partielles;
- ❖ Les méthodes d'estimations et d'imputation faites au niveau international ne tiennent pas compte des spécificités locales des Pays;
- ❖ Les Pays disposent de paramètres spécifiques qui permettent d'améliorer la qualité des imputations de données lorsqu'elles sont manquantes;
- ❖ Les services statistiques nationaux connaissent mieux le contexte national pour le faire;
- ❖ La responsabilité officielle non assumée au niveau national contraint les institutions internationales et les autres usagers à établir l'état de l'agriculture et de l'alimentation des pays à la place des Etats.



Cas des imputations sur les productions, superficies, rendements

L'objectif de l'imputation est, en incorporant l'ensemble des informations fiables utilisables, de fournir les meilleures estimations de la quantité d'aliments disponible pour permettre le calcul du bilan alimentaire.

La relation entre les variables du domaine de la production peut être exprimée ainsi :

$$P_t = A_t * Y_t;$$

Où P, A et Y représentent respectivement la production, la surface cultivée et le rendement, tous indexés par le temps t.;

De fait le rendement est inobservable, il est déduit de la production et de la superficie.



Deux catégories de méthodologies ont été proposées par le passé pour évaluer les données manquantes dans le domaine de production :

- la première catégorie utilise les séries historiques et applique des méthodes d'interpolation et de régression sur les tendances;
- la seconde catégorie fait recours aux taux de croissance des produits et/ou sur des agrégations par région;

L'imputation est menée de manière indépendante à la fois sur la surface cultivée et sur la production, tandis que les rendements sont calculés de manière implicite.

A la suite de l'analyse de la robustesse des estimations, la seconde méthode d'agrégation a été préférée car elle permet d'atteindre un taux de couverture élevé pour l'imputation, et semble extrêmement performante.

Dans un premier temps, on calcul le taux de croissance agrégée (r) de la production et de la surface par produit et de la région sur le groupe de pays ayant renseignés la donnée des deux années consécutives :

$$r_{s,t} = \frac{\sum_{c \in S} X_{c,t}}{\sum_{c \in S} X_{c,t-1}}$$

Le taux de croissance est ensuite appliqué a la dernière valeur exhaustive observée dans la série concernée du pays pour obtenir une valeur projetée de l'année manquante :

$$\hat{X}_{c,t} = X_{c,t-1} \times r_{s,t}$$

Forces :

- ❖ Permet d'atteindre un taux de couverture élevé pour l'imputation par rapport à l'imputation selon la méthode linéaire ;

Faiblesses :

- ❖ N'incorpore pas d'autres informations, comme les indices de végétation, de précipitations, ou de température qui peuvent apporter une information précieuse et aider à améliorer la précision de l'imputation ;
- ❖ N'incorpore pas des informations socio-économiques connexes comme les prix, la main d'œuvre, le potentiel en ressources naturelles, etc.
- ❖ la production et la surface sont estimées de manière indépendante;
- ❖ Du fait des données manquantes, le panier calculé peut ne pas être comparable au cours du temps, induisant ainsi des erreurs dans le calcul de la croissance de la production;
- ❖ Les paniers permettant de calculer les changements de production ou de surface cultivées peuvent être considérablement différents ;
- ❖ la méthodologie ne donne aucun aperçu des facteurs sous-jacents déterminant la production, qui sont pourtant nécessaires à une meilleure compréhension des phénomènes en jeu et donc à l'interprétation

Domaine : production **Indicateur :** Quantité de production de cultures primaires **Produit :** Sorgho

Option 1 :

Paramètres de base	Valeurs de base des paramètres	Source	Estimation pour imputation
R1: production moyenne par habitant (ou par actif agricole)		Recensement agricole, enquête agricole spécifique, recherche	QS1 = $R1 * S1 * E1 * x1 * (1+a)$
S1: Superficie moyenne par habitant (ou par actif agricole)		RGPH	
E1: Population agricole (ou effectif de la population active agricole)		RGPH ou RNA	
x1: Indice de qualité de la campagne (taux de variation de l'indice de végétation, de la pluviométrie ou du taux d'humidité)		satellite : à estimer chaque année	
a: taux de croissance de la population agricole ou de la population active agricole		RGPH	

Domaine : production **Indicateur** : Quantité de production de cultures primaires **Produit** : Gombo en pluvial

Parametres de base	Valeurs de base des paramètres	Source	Estimation pour imputation
R1 : rendement par hectare		Recensement agricole, enquête agricole spécifique, recherche	$QS1=R1*S1*E1*(1+tx1)*x1$
S1 : Superficie moyenne par exploitation agricole		RNA, enquête spécifique	
E1 : Nombre d'exploitation agricole		RNA, RGPH	
tx1 : taux de croissance de la population agricole		RGPH, RNA	
x1: Indice de qualité de la campagne (taux de variation de l'indice de végétation, de la pluviométrie ou du taux d'humidité)		satellite : à estimer chaque année	

Domaine : production **Indicateur :** Quantité de production de cultures primaires **Produit :** Mangue

Parametres de base	Valeurs de base des paramètres	Source	Estimation pour imputation
R1 : Rendement par pied d'arbre productif	137	Recensement agricole, enquête agricole spécifique, recherche	QM1 = $R1 * S1 * E1 * (1 + tx1) * x1$
S1 : Nombre total de pieds t-1	5000000	Recensement agricole, enquête agricole spécifique, recherche	
E1 : Pourcentage de pieds productifs	97%	RNA, enquête spécifique	
Tx1 : Taux d'accroissement de la plantation de manguiers	2,10%	RNA, enquête spécifique	
dx1 : Densité de pieds par hectare		RNA, enquête spécifique, Recherche	
x1: Indice de qualité de la campagne (taux de variation de l'indice de végétation, de la pluviométrie ou du taux d'humidité)	98%	satellite ou service de la météo : à estimer chaque année	

Domaine : production	Indicateur : Effectif des animaux vivants	Produit : bovin
-----------------------------	--	------------------------

Parametres de base	Valeurs de base des paramètres	Source	Estimation pour imputation
S1 : Effectif de l'espèce à l'année t-1	8000000	Recensement agricole, enquête spécifique sur le cheptel	$QM1 = S1*(1t+tx1)$
Tx1 : Taux de croît du troupeau	3,10%	RNA, enquête spécifique	

Domaine : production	Indicateur : Production de viande	Produit : viande de bovin
-----------------------------	--	----------------------------------

Parametres de base	Valeurs de base des paramètres	Source	Estimation pour imputation
S1 : Effectif de l'espèce à l'année t-1	8000000	Recensement agricole, enquête spécifique sur le cheptel	$QM1 = S1*(1t+tx1)*R1$
Tx1 : Taux de croît du troupeau	3,10%	RNA, enquête spécifique	
R1 : rendement moyen carcasse en kg/tête	80	enquête spécifique sur le cheptel, recherche	

Intégrer et systématisé la production des bilans alimentaires et des comptes économiques de l'agriculture dans le processus normal de l'activité des services statistiques agricoles et alimentaires;

Inventorier les paramètre par indicateur et par produit;

Mettre en place un sous groupe national de CountrySTAT chargé du Bilan alimentaire. Ce Groupe existe déjà dans certains pays : « Comité national du bilan alimentaire »

Intégrer les cadres services des comptes nationaux dans le comité national du bilan alimentaire

Produire des statistiques nationales au-delà des statistiques de sources

MERCI!