

Optimisation d'une enquête par sondage dans la recherche système

H. De Groot* & O. Traoré**

Keywords: Optimization - Sample surveys - Farming systems research - Agricultural surveys

Résumé

La collecte des données agro-économiques, essentielle pour la recherche système, est très coûteuse et la qualité des estimations est directement liée aux moyens investis. Il est démontré comment la précision des estimations peut être calculée, à partir de quelques paramètres de base, et comparée avec les coûts nécessaires pour optimiser l'allocation des ressources.

Les paramètres ont été calculés pour les variables clefs à base des données de l'Equipe Systèmes de Production et Gestion des Ressources Naturelles de Sikasso, Mali. La précision de l'estimation a été calculée pour différentes options et les résultats montrent que la précision des enquêtes agricoles est très modeste. La petite taille typique de ces enquêtes (12 villages et 8 exploitations par village) ne permet pas de bonne précision des moyennes: une erreur de 5%, souvent même 10%, est minimale. Une augmentation du nombre de villages a un effet positif important, tandis que l'effet d'augmenter le nombre d'exploitations diminue vite au-delà de six à huit exploitations par village.

Ces enquêtes ne sont donc pas indiquées pour obtenir des statistiques agricoles précises. Elles offrent des indicateurs de faible précision, mais indépendants et facilement accessibles. Elles sont mieux indiquées pour évaluer des relations entre variables, comme les causes de certaines contraintes, et les tendances dans l'évolution du système.

Summary

Optimizing Sample Surveys in Farming Systems Research

The collection of agricultural data, essential for Farming Systems Research, is expensive and the quality of the estimated statistics directly linked to the available budget. This paper shows how to calculate estimator precision of sample surveys from a small number of basic parameters, and how to compare it to a cost function for an optimal resource allocation.

Basic parameters are calculated for key variables of the data base of the Farming System Research Team of Sikasso, Mali, and the estimator precision is derived for a number of options. It is shown that this precision can only be very modest with a minimal mean square error of 5% to 10%, for the typical small sample size (12 villages, each with 8 farms). Increasing the number of villages diminishes this error, but the effect of an increasing number of farms per village decrease quickly above six to eight farms per village.

This type of surveys should therefore not be used for agricultural statistics. They do produce indicators of limited precision, but independent and readily accessible. They are better suited to analyze relationships between variables, to link causes with constraints and reasons with technology adoption, and to follow major changes in the system's evolution.

Introduction

La collecte de données socio-économiques fait partie intégrante de la recherche des systèmes de production. Ces enquêtes se confrontent à plusieurs problèmes. La collecte et la gestion des données sont perçues comme étant lourdes et onéreuses, et l'analyse des données dépasse rarement le stade préliminaire. Des réflexions quant aux objectifs de la collecte des données dans la recherche système, et des moyens à investir pour obtenir ces objectifs s'imposent.

En général, les enquêtes en milieu rural sont exécutées pour obtenir des informations sur son économie,

mieux comprendre le milieu et son évolution, et mieux cibler les actions. Dans les pays en voie de développement ces enquêtes sont confrontées à des problèmes particuliers. Le milieu est peu accessible et souvent il y a un manque des données de base. En plus, la variation est très grande, ce qui rend la collecte des données difficile et coûteuse, surtout vu les moyens très limités de ces pays. L'importance de la qualité des données a récemment reçu une nouvelle attention (7), et l'importance de faire le bilan entre les coûts et les bénéfices des enquêtes est reconnue (6). Néanmoins, des méthodes simples pour faire ce bilan et donc optimiser les enquêtes, manquent.

* IITA, BP 08-0932, Cotonou, Bénin, avant ESPGRN-Sikasso

** ICRISAT, BP 320 Bamako Mali, avant ESPGRN-Sikasso

Adresse de correspondance: Hugo De Groot, IITA, BP 08-0932, Cotonou, Bénin - tél. 229 350 553/350188, fax 229 350 556, E-mail IITA-Benin@cignet.com

Reçu le 31.07.97 et accepté pour publication le 24.03.98.

Il existe plusieurs livres sur la collecte des données en milieu rural dans les pays en voie de développement (1) ainsi que des manuels de statistiques agricoles (5). Ils donnent des informations importantes sur les techniques de base, mais n'approfondissent pas le problème d'optimisation. D'autre part, il y a des manuels d'enquêtes et de statistiques avancées, écrits pour un cadre agricole spécifique (2,3). Ces derniers sont malheureusement très techniques et peu accessibles pour les non-techniciens.

Ces dernières années, un intérêt grandissant pour l'estimation de la production agricole a été constaté (8,9,10, 12). Des efforts ont été faits pour rassembler les méthodes et les résultats d'une analyse coût-bénéfice des enquêtes (6), et des méthodes numériques d'optimisation ont été développées dans ce but (4).

Il reste toutefois à faire un grand effort pour rendre ces techniques statistiques et mathématiques accessibles aux praticiens et aux chercheurs sur le terrain, effort qui doit tenir compte de leur connaissance limitée de la théorie, et de l'utilisation de l'ordinateur qui permet le calcul des formules complexes. Le problème est dans le choix des techniques, des méthodes de calcul et l'interprétation des résultats. L'objectif de cet article est de contribuer à une utilisation optimale de la théorie statistique et d'enquêtes dans le domaine de la recherche système. Sur la base des données du Mali-Sud et des méthodes de calculs simples, la précision d'une enquête peut être calculée d'avance. Ces calculs, combinés avec un calcul simple des coûts, permettent des simulations à l'ordinateur aboutissant à une optimisation du problème et à des recommandations pratiques.

Méthodologie

D'abord un modèle simple est développé pour calculer la précision des estimateurs statistiques sur la base des données secondaires ou des simples suppositions, et les fonctions de coûts pour ces estimateurs. Ce modèle est alors appliqué sur les données de base de l'Equipe Systèmes de Production et Gestion des Ressources Naturelles de Sikasso (ESPRN) de l'Institut d'Economie Rural au Mali qui suit, de façon systématique depuis 1988, 96 exploitations dans 12 villages (voir carte en Figure 1).

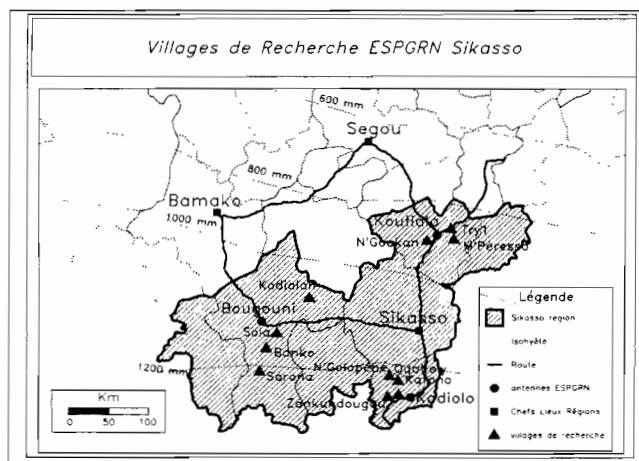


Figure 1

Théorie

En statistique, on appelle population une collection d'unités dont les caractéristiques sont à étudier, comme p.e. toutes les exploitations d'une zone. Une variable est une caractéristique ayant une valeur spécifique pour chaque élément de la population, par exemple la superficie cultivée par exploitation. Les statistiques sommaires comme la moyenne et la variance permettent de décrire une population avec quelques paramètres. On se limite ici aux variables quantitatives. Supposons une population de N éléments dont on veut étudier la variable X . Chaque élément i a une valeur X_i , et la moyenne de la variable X est définie par:

$$\bar{X} = \sum_{i=1}^N \frac{X_i}{N}$$

et la variance de X est:

$$\sigma^2 = \sum_{i=1}^N \frac{(X_i - \bar{X})^2}{N}$$

La racine de la variance est appelée écart-type, et l'écart-type divisé par la moyenne donne le coefficient de variation (CV), intéressant comme nous allons voir parce que indépendant de l'unité de mesure.

Comme la collecte de mesures pour chaque élément d'une population est fastidieuse et onéreuse, on a souvent recours à des enquêtes par sondage. On sélectionne un sous-groupe représentatif de la population, l'échantillon, et on estime les paramètres statistiques de la population sur la base des valeurs de l'échantillon. La formule pour estimer ces paramètres est appelée estimateur. Par exemple, la moyenne de l'échantillon est un estimateur de la moyenne de la population. On accepte que l'estimateur ne correspond pas au paramètre de la population et qu'il y a donc une erreur, appelée erreur de sondage.

On appelle biais l'erreur systématique, égale à la différence entre l'espérance de l'estimateur et le paramètre qu'on veut estimer. Pour le cas de la moyenne:

$$B = E(\bar{x}) - \bar{X} \quad \text{avec} \quad B = \text{biais}$$

On peut montrer qu'avec le sondage aléatoire simple (SAS), la moyenne de l'échantillon est un estimateur non biaisé de la moyenne de la population, en cas d'absence d'erreurs de mesure.

Le grand avantage d'un échantillon aléatoire est que la précision de l'estimation peut être calculée, en se basant sur la distribution de l'erreur de sondage. On peut prouver que cette distribution tend vers la distribution normale avec une moyenne de zéro et un écart-type, appelé erreur-type, donné par:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Optimisation d'une enquête de Sondage Aléatoire Simple

Optimiser signifie obtenir un maximum de résultat avec des moyens limités. Pour les sondages, on cherche à obtenir la meilleure estimation possible avec un bud-

get minimal et dans le temps disponible. On désire maximiser la précision, ou donc minimiser l'erreur d'estimation. Un indicateur convenable est le carré moyen d'erreur (CME), ou l'espérance de la différence entre l'estimateur et le paramètre qu'on veut estimer, élevée au carré. Pour la moyenne, on peut prouver que:

$$CME(\bar{x}) = E(\bar{x} - \bar{X})^2 = \frac{\sigma_x^2}{n} + B^2$$

Pour mieux juger de cette erreur, il est intéressant de la diviser par la moyenne, ce qui donne le CME relatif ou CMER. De la même façon on définit le biais relatif BR, et pour un sondage aléatoire simple de n éléments, avec un biais B , on peut calculer

$$CMER = \frac{1}{\bar{X}} \sqrt{\frac{\sigma_x^2}{n} + B^2} = \sqrt{\frac{CV^2}{n} + BR^2}$$

La contrainte essentielle d'une enquête est le plus souvent le budget. Les coûts pour un sondage aléatoire simple peuvent se diviser en frais fixes C_0 et en coûts variables, C_1 par élément de l'échantillon. Avec un budget Bud la contrainte peut être exprimée par:

$$C_0 + nC_1 \leq \text{Bud}$$

Dans l'équation (7) il n'y a qu'une inconnue: la taille de l'échantillon n . Elle est déterminée par le budget, ce qui détermine à son tour le CMER ou l'erreur minimale qu'on peut obtenir. En pratique, il convient de calculer le CMER pour différentes options de budget, pour trouver le meilleur ratio de qualité (précision) sur prix (budget). On remarque qu'une estimation du coefficient de variation est nécessaire pour le calcul de CMER.

Une première estimation de l'erreur peut se faire à l'aide de l'abaque en Figure 2. Ceci permet, à partir d'une estimation de l'écart-type relatif, de calculer l'erreur type de l'estimation de la moyenne en fonction de la taille de l'échantillon. A travers l'estimation des deux paramètres de la fonction budgétaire (formule 7), un premier bilan précision-coût peut se faire. Sans compliquer, on peut tenir compte du biais des estimations en ajoutant un pourcentage arbitraire pour cette erreur.

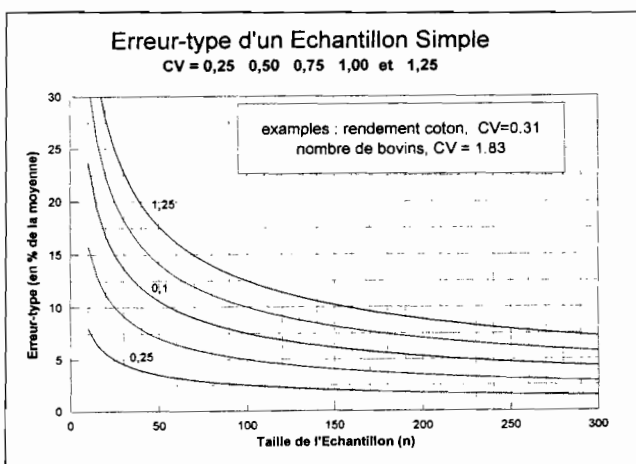


Figure 2

Malheureusement, il existe peu d'études qui ont publié ces estimations.

On souhaite par exemple connaître la surface en céréales par exploitation dans une région donnée. De par les enquêtes précédentes ou dans d'autres régions, on sait que le CV = 60% de la moyenne et on évalue que les paysans surestiment les surfaces avec 5%. Ceci permet de calculer

$$CMER = \sqrt{\frac{CV^2}{n} + BR^2} = \sqrt{\frac{0,60^2}{n} + 0,05^2}$$

Avec un coût fixe pour une telle enquête de 100.000 FCFA par exemple, et un coût variable de 1000 FCFA par exploitation, on peut calculer la précision maximale qu'un budget donné peut obtenir. La figure 2 présente la courbe typique pour cet exemple. Au début, à partir du coût fixe C_0 , l'erreur est très large (60% pour $n=1$), mais diminue très vite avec une augmentation du budget (déjà 27% pour $n=5$, et moins de 10% à partir de $n=50$). Cette amélioration diminue graduellement pour arriver au biais relatif de BR.

Optimiser les enquêtes de sondage dans la recherche système

1. Principes

Dans la recherche système, on utilise surtout le sondage à deux degrés (SDD). Si la population d'intérêt est organisée sur plusieurs niveaux, on peut d'abord tirer un échantillon d'un ensemble d'unités primaires du niveau supérieur. Dans une deuxième étape, on tire dans chaque unité retenue des éléments des unités secondaires (11). Dans la pratique de la recherche système de production, on choisit d'abord les villages comme unités primaires, puis les exploitations dans ces villages. Ce système a des avantages évidents. Les exploitations ne sont pas dispersées dans la zone étudiée, diminuant ainsi les frais de transport. Le nombre de villages est réduit, permettant de mieux connaître le milieu et ses habitants. Dans le cadre de la recherche participative, une meilleure collaboration entre paysans, paysannes et chercheurs est ainsi possible.

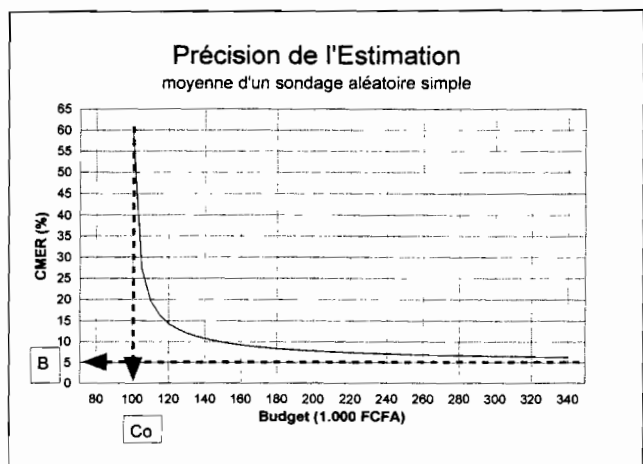


Figure 3

Si les villages et les exploitations sont tous deux tirés au hasard, la précision des estimateurs peut être calculée. Supposons que la zone étudiée contient N villages, dont on tire un échantillon de n. Tous les villages ont approximativement le même nombre d'exploitations M et on en tire m de chaque village.

Sous certaines conditions, la moyenne de l'échantillon \bar{x} est un estimateur non-biaisé de la moyenne de la population, \bar{X} .

La variance de la population σ^2 peut être décomposée en deux parties: la variance entre les villages

$$\sigma_e^2 = \sum_{i=1}^N \frac{(\bar{X}_i - \bar{X})^2}{N}$$

et la variance dans les villages:

$$\sigma_i^2 = \sum_{i=1}^N \sum_{j=1}^M \frac{(X_{ij} - \bar{X}_i)^2}{MN}$$

Ce dernier peut être estimé par la moyenne de la variance de l'échantillon dans les villages:

$$\hat{\sigma}_i^2 = s_i^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(x_{ij} - \bar{x}_i)^2}{n(m-1)}$$

La variance des moyennes des villages, calculée sur l'échantillon comme

$$s_e^2 = \sum_{i=1}^n \frac{(\bar{x}_i - \bar{\bar{x}})^2}{(n-1)}$$

n'est pas un estimateur non-biaisé de la variance entre villages σ_e^2 . Celui-ci peut être calculé avec la formule

$$\hat{\sigma}_e^2 = s_e^2 - \frac{s_i^2}{m}$$

Finalement, ces paramètres peuvent être utilisés pour calculer la précision de l'estimateur, toujours exprimée par l'erreur-type, calculée comme

$$\sigma_{\bar{x}} = \sqrt{\left(1 - \frac{n}{N}\right) \frac{\sigma_e^2}{n} + \left(1 - \frac{nm}{NM}\right) \frac{\sigma_i^2}{nm}}$$

En général, on n'a pas accès à ces paramètres avant d'entamer une enquête, et pour cette raison nous proposons d'utiliser l'erreur relative CMER, qui peut être calculée comme

$$CMER(\bar{x}) = \frac{1}{\bar{x}} \sqrt{\left(1 - \frac{n}{N}\right) \frac{\sigma_e^2}{n} + \left(1 - \frac{nm}{NM}\right) \frac{\sigma_i^2}{nm} + B^2}$$

Cette approche permet d'utiliser deux paramètres relatifs et donc plus facile à obtenir: le CV et la variance relative entre les villages

$$\rho = \sigma_e^2 / \sigma^2$$

On reformule

$$\frac{\sigma_e^2}{\bar{X}^2} = CV^2 \rho \quad \text{et} \quad \frac{\sigma_i^2}{\bar{X}^2} = CV^2(1-\rho)$$

pour obtenir:

$$CMER(\bar{x}) = \sqrt{\left(1 - \frac{n}{N}\right) \rho \frac{CV^2}{n} + \left(1 - \frac{nm}{NM}\right) (1-\rho) \frac{CV^2}{nm} + BR^2}$$

La précision dépend donc de la variance, exprimée en CV, et de la division de cette variance entre villages et dans les villages, la taille de l'échantillon (n et m) et de la population (N et M). Dans la pratique, la proportion des villages sélectionnée (n/N) est souvent petite, et on pourrait négliger les facteurs de correction, ce qui donne la simple formule

$$CMER(\bar{x}) = \sqrt{\rho \frac{CV^2}{n} + (1-\rho) \frac{CV^2}{nm} + BR^2}$$

Maintenant, on n'a plus besoin de connaître le nombre ou la taille des villages. La précision relative de l'estimation de la moyenne peut être calculée en fonction de la taille de l'échantillon, ρ , CV et BR. Ces paramètres peuvent être estimés des enquêtes précédentes ou en faisant des suppositions basées sur des calculs comme Tableau 1 dans la section suivante.

Pour faire le bilan, on a besoin des coûts. Les coûts d'un sondage à deux degrés peuvent être décomposés en un coût fixe C_0 , un coût variable C_1 par unité primaire (village) et un coût variable C_2 par unité secondaire (exploitation). La contrainte budgétaire devient maintenant

$$C_0 + nC_1 + nmC_2 \leq Bud$$

Il y a maintenant donc deux variables de choix, le nombre d'unités primaires n et le nombre d'unités secondaires par unité primaire, m. La taille de l'échantillon est alors donné par nm. En calculant les précisions de l'estimateur avec le CMER d'un côté, et les coûts avec la formule précédente, on peut faire le bilan et optimiser l'enquête. Saisir les formules dans un logiciel tableur et simuler différents scénarios donnant très vite des résultats. Alternativement (la théorie est expliquée en référence 2), on peut calculer la taille optimale de l'échantillon avec la formule:

$$m = \sqrt{\frac{c_1 \hat{\sigma}_i}{c_2 \hat{\sigma}_e}} = \sqrt{\frac{c_1 (1-\rho)}{c_2 \rho}}$$

2. Application: l'ESPGRN de Sikasso

Prenons un exemple, la superficie moyenne par exploitation dans la zone Mali-Sud. Pour les villages de l'ESPGRN-Sikasso, on a accès aux données des enquêtes précédentes qui permettent de calculer la taille moyenne (9,56 ha), le CV (0,53) et la variance entre villages relative ρ (0,19). La variance entre villages ne représente donc que 19% de la variance totale, ce qui diminue l'effet d'augmenter le nombre de villages. A

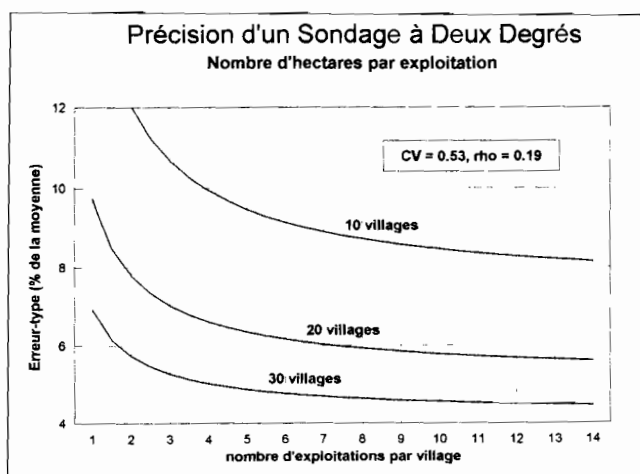


Figure 4

partir de l'équation (15) le bilan entre nombre de villages et nombre d'exploitations peut être calculé. Ces calculs sont démontrés dans la Figure 4.

La précision de l'estimation de la surface moyenne par exploitation est calculée pour un nombre d'options. La ligne la plus haute représente la précision qu'on peut obtenir avec un échantillon de 10 villages ($n=10$), en fonction du nombre d'exploitations par village (m). On constate qu'avec 4 exploitations par village on obtient une erreur de plus de 10% (en négligeant ici les erreurs de mesures), et l'erreur peut être réduite jusqu'à 8% en augmentant le nombre d'exploitations par village jusqu'à 10, mais il n'est pas possible d'aller plus loin. En général, l'effet d'augmenter le nombre d'exploitations diminue vite au-delà de 6 ou 8 exploitations par village. Augmenter le nombre de villages, par contre, a de fortes répercussions sur l'erreur, même avec le petit ρ . Dix villages de plus diminuent l'erreur de 2-3% (de la moyenne), et encore 10 villages avec 1%. Il est important d'étudier la limite des estimations. même avec 30 villages et 10 exploitations par village l'erreur reste supérieure à 4%, avant même d'inclure les erreurs de mesure, ou un intervalle de confiance de $\pm 8\%$ (niveau de confiance de 95%).

Le tableau 3 donne encore quelques variables clés de la base de données de l'ESPGRN-Sikasso avec les différentes composantes de la variance. La précision est calculée pour un échantillon de 12 villages, et 8 exploitations par village. On voit que les variables dont la variance est petite et située surtout dans les villages même, comme "nombre de personnes dans l'exploitation" peuvent être estimées avec une précision relativement bonne. Les variables liées à la zone agro-écologique (superficie ou production de coton et de maïs) varient beaucoup entre les villages ($\rho > 0.30$) tandis que les variables de taille (nombre de personnes, bovins ou hectares cultivées) varient surtout dans les villages ($\rho < 0.20$). Le nombre de membres ou d'hectares varie peu entre les exploitations ($CV = 0.54$ et 0.52) tandis que la taille des troupeaux est très variable ($CV = 1.82$).

Tableau 1
Moyenne et variance relative de quelques variables clés

variable	unités	\bar{X}	C.V.	ρ	$CMER_{\bar{X}}$ *
personnes	nbre/exploitation	15.11	0.51	0.06	0.06
bovins	nbre/exploitation	9.53	1.73	0.13	0.24
bœufs de labour	nbre/exploitation	2.39	0.87	0.29	0.15
superficie en maïs	ha/exploitation	1.28	0.99	0.48	0.21
superficie en coton	ha/exploitation	2.20	0.82	0.37	0.16
superficie totale	ha/exploitation	10	0.53	0.19	0.08
maïs produit	kg/exploitation	1538	1.01	0.46	0.21
coton produit	kg/exploitation	1763	1.14	0.42	0.23
rendement coton	kg/ha	1125	0.29	0.25	0.05
rendement maïs	kg/ha	1360	0.68	0.29	0.12
production céréalière	kg/membre	284	0.55	0.43	0.11

* ($n=12, m=8$)

Ces résultats nous permettent d'estimer l'ordre de grandeur d'un bon nombre de paramètres et ainsi calculer la précision approximative des enquêtes futures.

Jusqu'ici nous avons traité seulement des sondages aléatoires. Dans la réalité de la recherche système, malheureusement, les éléments sont souvent tirés de façon raisonnée. Puisque le raisonnement est nécessairement subjectif, il n'existe pas de calcul objectif pour estimer la précision. En général, on espère que l'erreur d'une estimation de la moyenne d'un choix (bien) raisonné ne devrait pas dépasser celle d'un tirage au hasard. Par contre, comme on évite généralement les extrêmes, la variabilité sera sous-estimée.

Conclusions

Dans cet article une nouvelle approche est testée pour optimiser des enquêtes de sondage dans la recherche système. Il a été démontré qu'à la base de quelques paramètres clés sur la structure et la variation de la population, un indicateur pratique des erreurs, le Carré Moyen d'Erreurs Relatif, peut être calculé. Comme l'erreur est ici exprimé en pourcentage de la moyenne de la population, son calcul ne nécessite que des paramètres relatifs, qu'on peut obtenir assez facilement. De la même façon on calcule les coûts avec quelques paramètres de base. Le calcul sous différents scénarios avec un simple modèle de tableur permet de faire le bilan entre précision désirée et moyens disponibles.

Cette approche s'applique facilement sur les sondages typiques à la recherche système: le sondage aléatoire simple et à deux degrés, et permet d'inclure de simples suppositions sur les erreurs de mesures et le biais. La méthode ne nécessite qu'une connaissance de base des statistiques et la manipulation d'un tableur sur un ordinateur de base.

Une application sur les données de l'ESPGRN de Sikasso montre bien les faiblesses et les avantages de ces enquêtes. La taille, typiquement petite, ne permet pas de bonne précision des moyennes, avec des erreurs rarement inférieures à 5%, même 10%. Ces enquêtes ne se prêtent donc pas à des statistiques agricoles comme la comparaison des zones. De l'autre côté, les données sont d'une qualité élevée, et les résultats sont indépendants et vite accessibles aux cher-

cheurs. Toujours, ce ne sont que des indicateurs, et il est donc indiqué de se limiter aux variables jugées nécessaires pour comprendre et suivre le système. Ici, la recherche système profiterait de son réseau d'agents de terrain déjà basés dans les villages de recherche.

Les données ainsi collectées sont bien adaptées pour des calculs statistiques autres que les moyennes. La précision des calculs des tendances et les analyses de régression multiple sont moins dépendantes de la taille de l'échantillon et sont donc plus indiquées.

Finalement, la base de donnée de la recherche système se laisse facilement combiner avec d'autres enquêtes sur des thèmes spécifiques, formelles aussi bien qu'informelles. Le réseau des agents de terrain permet de réaliser très vite des enquêtes informelles, en groupes ou individuelles, sur des thèmes actuels, par exemple le taux d'adoption d'une nouvelle variété

ou technologie. Ces enquêtes montrent facilement les facteurs clefs et les variables nécessaires à ajouter dans la base pour approfondir les connaissances ou pour tester formellement certaines hypothèses. Cette sorte d'étude peut intéresser un bon nombre de services et de bailleurs de fonds, ce qui peut générer des revenus, sans trop gêner la vocation centrale de la recherche système.

Remerciements

Nous voulons remercier nos collaborateurs à l'Equipe Systèmes de Production et Gestion des Ressources Naturelles-Sikasso, particulièrement le chef d'équipe Dr. Demba Kébé, nos collègues économistes Hamady Djouara et Johan Brons, les enquêteurs, les superviseurs et les chefs d'antenne pour leur support dans ce travail; ainsi que deux lecteurs anonymes pour leurs commentaires critiques et constructifs.

Références bibliographiques

1. Casley Dennis J. & Denis D. Lury, 1981. Data collection in developing countries. Cambridge MA: Oxford University Press.
2. Cochran William G., 1977. Sampling Techniques. New York NY: John Wiley and Sons.
3. Dagnelie Pierre, 1973. Théorie et méthodes statistiques, applications agronomiques. Les Presses Agronomiques de Gembloux, Gembloux (Belgique).
4. De Groote H., 1996. "Optimal Survey Design for Rural Data Collection in Developing Countries." Quarterly Journal of International Agriculture Vol. 35, No.2, pp. 163-175.
5. FAO, 1982. Estimation of crop areas and yields in agricultural statistics, FAO Economic and Social Development Paper 22. Rome: Food and Agriculture Organization.
6. Groves Robert M., 1989. Survey Errors and Survey Costs. New York: John Wiley & Sons.
7. Kelly V., J. Hopkins, T. Reardon & E. Crawford, 1995. Using Micro Data to Improve the Measurement and Analysis of African Agricultural Productivity, Michigan State University, Dept. of Agricultural Economics Staff Paper No. 95-24.
8. Murphy, Josette, Dennis J. Casley & John J. Curry, 1991. Farmers' Estimations as a Source of Production Data: Methodological Guidelines for Cereals in Africa. Washington, D.C.: World Bank.
9. Poate D., 1988. A review of methods for measuring crop production from smallholder producers. Experimental Agriculture, 1988, 24 (1).
10. Poate C.D. & Casley Dennis J., 1985. Estimating crop production in development projects, methods and their limitations. World Bank, Washington.
11. Snedecor George & William Cochran, 1957. Méthodes Statistiques. Association de Coordination Technique Agricole, Paris.
12. Verma Vijay, Marchant Tim and Scott Chris, 1988. Evaluation of crop-cut methods and farmer reports for estimating crop production, results of a methodological study on five African countries. London: Longacre Agricultural Development Centre Limited.

Dr. Ir. H. De Groote: Belge. Ingénieur agronome de l'université de Gand, Doctorat en agro-économie de University of Wisconsin-Madison, travaille actuellement pour l' International Institute of Tropical Agriculture (IITA) au Bénin, avant à l'Equipe Systèmes de Production et Gestion des Ressources Naturelles-Sikasso de l'Institut d'Economie Rurale, Mali.

Ir. O.F. Traoré: Malien. Ingénieur agronome de l'Institut Polytechnique Rural de Katibougou, Mali, travaille pour l' International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) au Mali, avant agro-économiste pour l'Equipe Systèmes de Production et Gestion des Ressources Naturelles-Sikasso de l'Institut d'Economie Rurale, Mali.