

Prévision des livraisons de maïs pour une coopérative agricole

D Wallach

INRA, station de biométrie et d'intelligence artificielle, BP 27, 31326 Castanet Tolosan Cedex, France

(Reçu le 24 juin 1991; accepté le 21 juin 1992)

Résumé — La prédiction, dès le mois d'août, de la production et de la quantité de maïs livré en fin de saison à une coopérative agricole est décrite. La méthode adoptée consiste à tirer dans un premier temps un échantillon d'adhérents de la coopérative, puis dans un deuxième temps des parcelles parmi les parcelles de maïs de ces adhérents. Dans chacune de ces parcelles on fait des mesures du nombre de grains. La production est alors prédite par une équation très simple, basée sur le produit du nombre de grains et le poids moyen pluriannuel par grain. La probabilité de tirer un adhérent à chaque tirage est proportionnelle à la quantité de semences qu'il a achetée, tandis que la probabilité de tirer une parcelle est proportionnelle à la superficie de la parcelle. Les erreurs dans les prévisions de la quantité livrée étaient de 17% et de 11% respectivement en 1989 et 1990. La plus grande partie de l'erreur semble venir d'une sous-estimation de la partie de la production qui n'est pas livrée à la coopérative.

prévision / échantillonnage / maïs / coopérative agricole

Summary — **Predicting corn deliveries to an agricultural cooperative.** *The prediction of corn production and corn deliveries to an agriculture cooperative has been described. In August, some two months before harvest, a sample of members is drawn, and for each of these members a sample of corn fields is drawn. Each of these corn fields is visited, and the number of grains determined in randomly selected row sections. The equation for predicting the production of the field is then very simple. Aside from a correction factor, it is just the product of the number of grains and the multi-year average weight per grain. The main emphasis here is on the method of sampling and on the associated equations for predicting total production and deliveries. Member sampling weights each member by the quantity of seed bought at the start of the season. Field sampling weights each field by its area. The predictions presented to the cooperative in 1989 and in 1990 overestimated deliveries by 17% and 11%, respectively. Analysis of the errors indicates that the largest contribution to the error is the underestimate of that part of total production which is not delivered to the cooperative.*

prediction / sampling / corn / agricultural cooperative

INTRODUCTION

Pour gérer au mieux ses ventes d'un produit agricole, une coopérative doit connaître le plus tôt possible la quantité du produit qu'elle aura à vendre. C'est la raison pour laquelle la coopérative du Bassin du Midour (CBM) a cherché à établir une méthodologie pour prévoir les livraisons de maïs dès le mois d'août, c'est-à-dire quelque 2 mois avant la fin de la récolte.

L'approche que l'on a choisie, et que l'on présente dans cet article, est basée sur l'échantillonnage. On utilise les résultats de l'échantillonnage pour estimer d'une part la production totale des adhérents de la coopérative, et d'autre part la quantité de maïs qui ne sera pas livrée à la coopérative (les pertes). La quantité

qui sera livrée est alors simplement la différence de ces deux termes.

L'estimation de la production totale passe par celle de la production dans chacune des parcelles de l'échantillon. Pour cette estimation, on utilise des comptages dans la parcelle du nombre de grains par mètre carré, qui est multiplié par un poids moyen par grain, estimé à partir des résultats d'années précédentes. Cela suppose que l'estimation soit faite suffisamment tard dans la saison pour permettre ces comptages, ce qui représente donc une première limitation à la méthode présentée.

Les données disponibles sur la population étudiée déterminent en grande partie le protocole d'échantillonnage. Dans le cas présent, on dispose essentiellement de la liste des achats de

semences par adhérent en début de saison. On tire donc les adhérents de l'échantillon de cette liste. On peut supposer, et les observations le confirment, que la production (et dans une moindre mesure les livraisons) d'un adhérent est plus ou moins proportionnelle à la quantité de semences achetées. Une façon de profiter de ce lien est de pondérer chaque adhérent, pour le tirage, par la quantité achetée. C'est en effet ce type de tirage (qui s'appelle un tirage de type «ppz»; voir Cochran, 1977), qui est utilisée ici pour choisir les adhérents. Une deuxième limitation à la méthode présentée est donc qu'il faut connaître les quantités de semences achetées.

Dans la section suivante, nous présentons les données. Ensuite nous considérons la méthode d'échantillonnage choisie et les formules d'estimation, puis l'évaluation de l'erreur d'estimation. Les résultats et conclusions sont présentés dans la section *Résultats et Discussion*. Les équations pour l'analyse statistique sont présentées dans l'annexe.

DONNÉES

La coopérative compte environ 1 500 adhérents. Les résultats de l'échantillon 1990 donnent une idée de l'organisation de leur production. Les superficies en maïs par adhérent variaient de 1 à 220 ha (correspondant à 1 jusqu'à 10 parcelles), et les rendements de 10 à 117 q/ha suivant la parcelle. La quantité de maïs livrée par adhérent en 1990 allait de 0 à 10 470 quintaux.

Le programme d'échantillonnage choisi pour ce travail se passe en trois temps. D'abord on choisit un échantillon d'adhérents. Ce tirage est fondé sur la quantité de semences achetées par chacun des adhérents en début d'année, information notées en routine pour la comptabilité (les détails sur les tirages d'adhérents et de parcelles sont présentés dans la section «estimation par échantillonnage»). Ensuite on interroge chaque adhérent de l'échantillon. D'une part, on lui demande d'estimer la quantité de maïs qu'il retiendra pour ses propres besoins, la superficie qui fournira de l'ensilage, et la quantité qu'il pense vendre à d'autres organismes, pour pouvoir estimer les pertes de maïs pour la coopérative. D'autre part, on lui demande le nombre et les superficies de ses parcelles de maïs, informations utilisées pour choisir, comme deuxième étape du protocole,

un échantillon de parcelles. Enfin, dans chaque parcelle de l'échantillon on choisit n_1 placettes de 5 m de rang chacune pour compter le nombre d'épis, et n_2 placettes de 5 épis consécutifs chacune pour compter le nombre de grains. Ces placettes sont choisies au hasard dans la parcelle.

En 1989, l'échantillon était composé de 50 adhérents, de 2 parcelles par adhérent, avec $n_1 = 8$ et $n_2 = 4$. Pour tenir compte des résultats de cette première année, l'échantillonnage a été modifié en 1990. On a alors échantillonné 100 adhérents, deux parcelles par adhérent, avec $n_1 = 4$ et $n_2 = 2$. Les mesures dans les parcelles ont été effectuées du 1^{er} au 17 août en 1989, et du 6 au 31 août en 1990.

Au moment de la récolte (qui se termine, dans la région de la coopérative, vers la fin octobre) on a pesé à part la production de la plupart des parcelles de l'échantillon. Par ailleurs on connaît, quelque temps après la fin de la récolte, la quantité totale de maïs livrée à la coopérative par chaque adhérent. Ces informations servent à évaluer et à analyser l'erreur de prédiction.

ESTIMATION

Décomposition de la livraison totale

On peut exprimer la livraison totale de maïs comme une somme de termes, dont certains représentent des quantités mesurables au moment de l'échantillonnage et d'autres des quantités qui ne seront connues qu'à la récolte.

La livraison totale est simplement la somme des livraisons de chacun des n adhérents de la coopérative :

$$LIVRE-TOT = \sum_{k=1}^n LIVRE-ADH_k \quad [1]$$

La quantité livrée par chaque membre est égale à la quantité récoltée mais ce qu'on appellera des pertes, qui correspondent à la somme du maïs gardé à la ferme, ensilé, ou vendu à d'autres organismes stockeurs :

$$LIVRE-ADH_k = REC-ADH_k - PERTE-ADH_k \quad [2]$$

Le maïs récolté par un adhérent k est la somme des récoltes de ses M_k parcelles :

$$REC-ADH_k = \sum_{l=1}^{M_k} REC-PAR_{kl} \quad [3]$$

La récolte dans chaque parcelle peut s'écrire comme le produit de la surface de la parcelle, du nombre de grains par épis au moment de l'échantillonnage, du nombre d'épis par unité de surface au moment de l'échantillonnage, du rapport R_{kl} entre le nombre de grains par unité de surface au moment de la récolte et ce nombre au moment de l'échantillonnage, du poids par grain à la récolte, et du rapport H_{kl} entre le poids de grains récolté et le poids total de grains dans la parcelle au moment de la récolte :

$$REC-PAR_{kl} = M2-PAR_{kl} \times GRAINS-EPI_{kl} \times EPIS-M2_{kl} \times R_{kl} \times POID-GRAIN_{kl} \times H_{kl} \quad [4]$$

Estimation des quantités non mesurables

Une partie des facteurs qui entrent dans le calcul de $LIVRE-TOT$ n'est pas mesurable au moment de l'échantillonnage. Il s'agit de H_{kl} , $POID-GRAIN_{kl}$, $PERTE-ADH_k$ et R_{kl}

On estime H_{kl} par :

$$H = 0,88(\pm 0,05) \quad [5]$$

calculé à partir de 46 mesures à la station expérimentale de l'INRA, à Auzeville. La quantité entre parenthèses représente deux écarts types, mais cette estimation de l'erreur n'inclut pas d'éventuelles différences entre les parcelles expérimentales et les parcelles de la coopérative.

Comme estimateur du poids par grain, on utilise la moyenne pluriannuelle pour la variété cultivée sur la parcelle, ajustée par une correction qui différencie entre parcelles irriguées et non-irriguées. On note $poide-grain_{kl}$ cette estimation pour la parcelle k de l'adhérent l .

Enfin, on estime R_{kl} par $r = 1$, et $PERTE-ADH_k$ par $perce-adh_k$ (la valeur déclarée par l'adhérent pour la quantité de maïs qu'il ne livrera pas), ce qui revient à supposer qu'il n'y a pas de perte de grains entre l'échantillonnage et la récolte, et que les déclarations des adhérents concernant les pertes sont exactes.

Estimation par échantillonnage

D'autres quantités nécessaires pour estimer $LIV-TOT$ sont mesurables, mais pour des raisons

pratiques ne peuvent être mesurées de façon exhaustive. On est alors amené au deuxième type d'estimation, basé sur l'échantillonnage. Il se distingue du premier type d'une part parce que ces estimations sont sans biais, ce qui n'est pas le cas en général pour le premier type d'estimation, et d'autre part parce que l'on maîtrise la taille de l'erreur due à l'échantillonnage, qui variera en fonction de la taille de l'échantillon. On décrit d'abord le protocole d'échantillonnage, et ensuite l'estimation.

La production de maïs est assez variable entre adhérents de la coopérative. Ainsi, si l'on calcule la production totale comme la moyenne des résultats d'un échantillon aléatoire d'adhérents, la variance sera relativement grande. Par contre, la production par unité de semence achetée à la coopérative en début de saison l'est beaucoup moins. On aura donc intérêt à calculer la moyenne de ce rapport, qui aura une variance relativement faible. Par contre, on ne peut plus alors utiliser un échantillon aléatoire simple. Pour avoir un estimateur sans biais de la production ou livraison totale, il faut pondérer chaque adhérent par la quantité de semence achetée pour l'échantillonnage (ce qui s'appelle un échantillonnage de type «ppz»). Pour tirer alors un échantillon de n adhérents, on fait n tirages indépendants, c'est-à-dire avec remise. À chaque tirage, la probabilité de tirer l'adhérent k est $Z-ADH_k$, où

$$Z-ADH_k = \frac{ACHAT-ADH_k}{ACHAT-TOT} \quad [6]$$

Au deuxième niveau de tirage, pour chaque adhérent choisi au premier tirage, on tire m de ses parcelles de maïs. En général, les parcelles de maïs d'un adhérent peuvent être de tailles assez différentes, avec donc une production par parcelle assez variable. En revanche, la production par unité de surface variera moins. On est alors encore dans une situation où un estimateur de type «ppz» est indiqué. Pour un adhérent k , la pondération de la parcelle l , notée $Z-PAR_{kl}$, est la fraction de la surface totale de l'adhérent représentée par cette parcelle :

$$Z-PAR_{kl} = \frac{M2-PAR_{kl}}{M2-ADH_k} \quad [7]$$

Pour tirer l'échantillon de m parcelles parmi les parcelles de l'adhérent k , on fait m tirages indépendants. À chaque tirage, chaque parcelle a

une probabilité $Z-PAR_{kl}$ d'être tirée. Notons que les surfaces des parcelles ne sont connues qu'après avoir interviewé l'adhérent. C'est la raison pour laquelle il faut procéder par étapes.

Au troisième niveau d'échantillonnage, on choisit des placettes dans chaque parcelle de l'échantillon, pour les comptages du nombre d'épis par m^2 et du nombre de grains par épi. Les placettes ont été choisies par les enquêteurs de façon subjective, pour être représentatives de la parcelle. On supposera que les placettes ainsi choisies représentent un échantillon aléatoire. (On gagnerait sans doute en précision par un échantillonnage réellement aléatoire).

L'estimation de la production d'une parcelle de l'échantillon est simplement :

$$rec-par_{kl} = M2-PAR_{kl} \times grains-epi_{kl} \times epis-m2_{kl} \times h \times poid-grain_{kl} \times r \quad [8]$$

où $grains-epi_{kl}$ et $epis-m2_{kl}$ indiquent respectivement les moyennes du nombre d'épis par m^2 et du nombre de grains par épis pour la parcelle l de l'adhérent k .

L'estimateur de la récolte totale de maïs de l'adhérent k est alors :

$$rec-adh_k = \frac{1}{m} \sum_{j=1}^m \frac{rec-par_{kl(j)}}{Z-PAR_{kl(j)}} \quad [9]$$

où l'indice $kl(j)$ indique l'adhérent k , et la parcelle tirée au j^{e} tirage. L'estimateur de la quantité de maïs livrée à la coopérative par l'adhérent k est :

$$livre-adh_k = rec-adh_k - perte-adh_k \quad [10]$$

Enfin, un estimateur de la livraison totale de maïs à la coopérative est :

$$livre-tot = \frac{1}{n} \sum_{i=1}^n \frac{livre-adh_{k(i)}}{Z-ADH_{k(i)}} \quad [11]$$

où l'indice $k(i)$ indique l'adhérent tiré au i^{e} tirage.

L'équation [11] donne l'estimation de la livraison totale à la coopérative. On peut aussi estimer séparément la récolte totale, la quantité de maïs gardée à la ferme, la quantité ensilée ou la quantité vendue ailleurs.

Il faut souligner que le non-biais des estimateurs ci-dessus résulte du système d'échantillonnage. Aucune hypothèse supplémentaire n'est nécessaire.

ERREURS

Il est important d'avoir une idée de l'erreur associée à l'estimation de la livraison totale à la coopérative. D'une part, il faut fournir une estimation de l'erreur à la coopérative avec l'estimation de la livraison, pour permettre d'apprécier la fiabilité de cette dernière. D'autre part, l'analyse des erreurs en fin de saison doit permettre d'identifier les sources principales d'erreur, et donc de focaliser des efforts supplémentaires sur ces points critiques.

Une mesure classique d'erreur est l'erreur quadratique moyenne de prédiction MSEP (sigle anglais pour *mean squared error of prediction*), définie par :

$$MSEP = E \{ (LIVRE-TOT - livre-tot)^2 \} \quad [12]$$

L'espérance ici est calculée sur les différents échantillons possibles, et l'erreur concerne une seule année.

MSEP peut se décomposer comme suit :

$$MSEP = (LIVRE-TOT - E \{ livre-tot \})^2 + E \{ (E \{ livre-tot \} - livre-tot)^2 \} = B^2 + V \quad [13]$$

Le premier terme est le biais au carré, qui mesure globalement l'importance des erreurs de prédiction pour les quantités non mesurables. Le deuxième terme est la variance due aux échantillonnages.

La variance peut être estimée à partir des résultats de l'échantillonnage. On peut distinguer trois contributions à la variance, associées respectivement à l'échantillonnage des adhérents (V_1), à l'échantillonnage des parcelles (V_2), et à l'échantillonnage à l'intérieur d'une parcelle (V_3). Les formules pour ces termes, ainsi que pour des estimateurs correspondants, sont présentées en annexe.

Le biais ne peut être estimé qu'a *posteriori*, une fois la vraie quantité livrée connue. La stratégie sera donc d'estimer ce biais à la fin de chaque année, et d'utiliser la moyenne des valeurs antérieures pour prédire le biais pour l'année en cours. La somme du biais au carré estimé plus la variance estimée donne un estimateur de MSEP pour l'année en cours.

Un estimateur sans biais de B est :

$$\hat{B} = livre-tot'' - livre-tot$$

où $livre-tot''$ est l'estimateur de $LIVRE-TOT$ basé sur les vraies quantités livrées par les adhérents de l'échantillon.

L'analyse de la variance peut aider à améliorer le protocole d'échantillonnage. En particulier, on peut calculer m^* , le nombre de parcelles par adhérent qui minimise la variance pour un coût d'échantillonnage donné. On suppose que le coût total d'échantillonnage s'écrit de la façon suivante :

$$C = c_1 n' + c_2 n' m' \quad [14]$$

où c_1 représente le coût fixe associé à l'inclusion d'un nouvel adhérent dans l'échantillon (il faut le contacter, aller le voir, l'interviewer), tandis que c_2 est le coût d'une parcelle supplémentaire (il faut y aller, et puis faire des mesures). On peut alors démontrer que

$$m^* = \sqrt{\frac{V_1}{V_2} \frac{c_1}{c_2}} \quad [15]$$

L'analyse du biais permettra d'identifier différentes contributions à ce biais. On peut décomposer le biais comme suit :

$$\begin{aligned} \text{LIVRE-TOT-E}[\text{livre-tot}] &= \\ & (\text{LIVRE-TOT-E}[\text{livre-tot}''']) \\ & + (E[\text{livre-tot}'''] - E[\text{livre-tot}']) \\ & + (E[\text{livre-tot}'] - E[\text{livre-tot}]) \\ & = B1 + B2 + B3 \end{aligned} \quad [16]$$

où *livre-tot'* représente la prédiction de la quantité livrée, quand on utilise les vraies valeurs de la production par parcelle à la place des prédictions de ces productions. La première différence $B1$ est égale à zéro, parce que l'espérance sur échantillons de *livre-tot'''* est égale à la vraie quantité totale livrée. Pour estimer les autres différences, on utilisera les mesures en fin de campagne.

Le terme $B3$ est l'espérance de la différence entre *livre-tot*, présenté dans la section précédente, et l'estimateur que l'on aurait si l'on remplaçait la production estimée par la production observée dans chaque parcelle. Ce terme sera alors d'autant plus grand que la production est mal estimée. Le terme $B2$ mesure la différence entre l'estimateur qui utilise la production observée moins la perte estimée pour chaque adhérent, et celui qui utilise la livraison observée pour chaque adhérent. L'importance de ce terme dépend essentiellement des erreurs dans l'estimation des pertes.

RÉSULTATS ET DISCUSSION

Les estimations et quantités totales livrées pour 1989 et 1990 sont présentées dans le tableau I. Il

Tableau I. Estimations basées sur échantillonnage.

Quantité	Valeur estimée (t)	Valeur observée (t)
1989		
1. rec-lot	155 654	
2. perte-lot	17 322	
3. livre-tot	138 332	118 524
1990		
1. rec-tot	106 348	
2. perte-tot	28 510	
3. livre-tot	77 838	69 942

1. Récolte totale de maïs par les adhérents de la coopérative; 2. perte totale; 3. quantité totale de maïs livrée à la coopérative en fin de saison.

y a une surestimation de la quantité totale livrée pour les deux années, de quelque 19 808 tonnes sur 118 524 (17%) en 1989, et de 7 896 tonnes sur 69 942 (11%) en 1990.

En 1989, la quantité livrée représente 89% de la production totale, et en 1990, 73%. Fournir ce pourcentage, difficilement accessible autrement, est en effet un des intérêts de ce travail.

L'examen des erreurs aide à comprendre l'origine des différences entre la prédiction et la quantité de maïs effectivement livrée. Considérons d'abord la variance due à l'échantillonnage (voir tableau II). Le premier constat est que la variance liée à l'estimation du nombre d'épis et de grains dans une parcelle (V_3) est négligeable par rapport aux autres termes. C'est la raison pour laquelle le nombre de placettes par par-

Tableau II. MSEP (*mean squared error of prediction*).

Terme	Estimation 1989 (t ²)	Estimation 1990 (t ²)
Échantillonnage de parcelles (V_1)	10,4 10 ⁶	10,6 10 ⁶
Échantillonnage d'adhérents (V_2)	35,2 10 ⁶	8,3 10 ⁶
Échantillonnage dans une parcelle (V_3)	0,2 10 ⁶	0,1 10 ⁶
Variance totale	45,7 10 ⁶	19,0 10 ⁶
Biais au carré	316,2 10 ⁶	43,4 10 ⁶
MSEP	361,9 10 ⁶	62,4 10 ⁶

celle a été réduit en 1990. Le terme V_1 est petit surtout parce qu'il y a compensation des erreurs entre différentes parcelles.

À partir des valeurs V_1 et V_2 , on peut calculer m^* , le nombre optimal de parcelles par adhérent (voir section précédente). Ce nombre optimal, calculé à partir des résultats de 1989, est présenté dans le tableau III, en fonction du rapport c_1/c_2 . Dans le cas présent, il semble que le rapport c_1/c_2 soit approximativement 1. D'après le tableau, jusqu'à un rapport de $c_1/c_2 = 5$, on aurait intérêt à ne prendre qu'une seule parcelle par adhérent. Pourtant, avec une seule parcelle on perd la possibilité d'estimer la variance due au tirage des parcelles. Il semble alors judicieux de maintenir l'échantillonnage de deux parcelles par adhérent, mais de ne pas augmenter ce chiffre. C'est pourquoi, en 1990, l'économie d'effort réalisée en réduisant le nombre de placettes par parcelle a été utilisée pour augmenter le nombre d'adhérents de l'échantillon, et non pas le nombre de parcelles par adhérent. Avec d'autres valeurs pour les variances, on tirerait éventuellement d'autres conclusions. Néanmoins, en débutant une nouvelle étude, et en l'absence d'autres informations, ce rapport de 2 parcelles par adhérent fournit un point de départ.

La deuxième contribution à MSEF est le biais au carré. Ce terme domine très largement la variance en 1989, et est encore considérablement plus grand que la variance en 1990. La conclusion pratique est que l'effort pour améliorer les prédictions doit surtout porter sur le biais. Il n'est peut-être pas surprenant que dans une prédiction basée d'une part sur des mesures dans le champ, d'autre part sur une enquête sur les intentions des adhérents, la partie enquête soit le maillon faible.

Les 2 contributions au biais sont présentées dans le tableau IV. Il faut souligner que les estimations ici sont assez approximatives, parce qu'aussi bien B_2 que B_3 sont estimés comme la différence entre deux valeurs qui sont proches. B_2 fournit la contribution prépondérante, surtout en 1990. C'est-à-dire que l'erreur essentielle résulte non pas d'une mauvaise estimation de la production, mais d'une sous-estimation des pertes.

Tableau III. Nombre optimal de parcelles par adhérent.

c_1/c_2	0,1	0,5	1,0	2,0	5,0	10,0
m^*	0,17	0,38	0,54	0,77	1,21	1,72

Tableau IV. Composantes du biais estimées à partir d'observations en fin de saison ¹.

Terme du biais ²	Valeur estimée (t)
B_2	-11 332
B_3	- 6 036

¹ Pour certaines parcelles de l'échantillon, les mesures de fin de saison manquent. Ces parcelles ne sont pas utilisées ici.

² Voir équation [19]. $B_1 = 0$.

Si l'on ne regarde que la production, on a pour une production totale estimée de 155 654 tonnes en 1989 un biais estimé de $B_3 = 6 036$ tonnes, ou 3,9%. En 1990, ce chiffre n'est que de 1,5%. Étant donnée la simplicité de la formule de prédiction de la production, ce résultat est très satisfaisant, et indique que la méthode présentée ici est bien adaptée à la prévision de production régionale. D'ailleurs, une approche analogue est utilisée aux États-Unis pour prédire la production de plusieurs cultures (USDA-SRS, 1983).

Il serait intéressant de comparer cette approche avec d'autres qui ont été suggérées. On pourrait par exemple utiliser une équation agrométéorologique, qui relie la production à des données telles que les températures ou les précipitations décennales ou mensuelles (Dagneaud et Tranchefort, 1985; Feyerherm et Paulsen, 1986). L'intérêt principal de cette méthode est le fait que les données nécessaires sont peu nombreuses et relativement faciles à obtenir. Une autre méthode emploie des modèles dits de «simulation» ou encore «mécanistes», qui prédisent la production d'une parcelle à partir d'informations sur l'environnement (sol, météo) et sur la gestion de la parcelle (Hodges *et al*, 1987). On peut ainsi combiner ces approches, par exemple en utilisant un modèle mécaniste au lieu d'une équation simple pour prévoir la suite du développement de la culture dans une parcelle où l'on a fait des mesures.

Quant à la prédiction de la quantité livrée à la coopérative, on pourrait essayer de traiter les collectes des années passées comme une série chronologique, dont on veut prédire le prochain élément.

En conclusion, ce travail montre dans un cas concret la précision que l'on a atteint dans la prédiction des livraisons, en utilisant une approche basée sur l'échantillonnage. L'analyse des résultats a permis d'identifier la source essentielle de

l'erreur, ce qui devrait permettre par la suite d'améliorer ces prédictions.

$$V_3 = \sum_{k=1}^N \sum_{l=1}^{M_k} V[f_{kl}]$$

ANNEXES

On présente ici les équations de la variance pour un échantillonnage hiérarchique à trois niveaux, qui peuvent se déduire des résultats dans Cochran (1977). Les 2 premiers niveaux sont des échantillonnages de type «ppz». On tire d'abord n adhérents parmi N , la probabilité de tirer l'individu k étant Z_k , et puis on tire m parcelles parmi M_k , la probabilité de tirer la parcelle l de l'adhérent k étant V_{kl} . Au troisième niveau, on tire des placettes par simple échantillonnage aléatoire dans chaque parcelle de l'échantillon.

On note la quantité à estimer F , où :

$$F = \sum_{j=1}^N \sum_{l=1}^{M_k} F_{kl} \quad [17]$$

On suppose que, pour chaque parcelle de l'échantillon, on a un estimateur sans biais de F_{kl} , noté f_{kl} . Un estimateur sans biais de F est alors :

$$\hat{F} = \frac{1}{n} \sum_{i=1}^n \frac{1}{Z_{k(i)}} \frac{1}{m} \sum_{j=1}^m \frac{f_{k(i)l(j)}}{V_{k(i)l(j)}} \quad [18]$$

où les indices $k(i)$ et $l(j)$ indiquent respectivement l'adhérent choisi au i^{e} tirage et la parcelle choisie au j^{e} tirage.

La variance de \hat{F} est :

$$V[\hat{F}] = V_1 + V_2 + V_3 \quad [19]$$

où

$$V_1 = \frac{1}{n} E \left[\sum_{k=1}^N \frac{1}{Z_k} \frac{1}{m} \sum_{l=1}^{M_k} V_{kl} \left(\frac{F_{kl}}{V_{kl}} - F_k \right)^2 \right]$$

$$V_2 = \frac{1}{n} E \left[\sum_{k=1}^N Z_k \left(\frac{F_k}{Z_k} - F \right)^2 \right]$$

Comme estimateurs de ces termes, on a :

$$\hat{V}_1 = \frac{1}{n^2} \sum_{i=1}^n \frac{1}{Z_{k(i)}^2} v_m[\hat{F}_{k(i)}] \quad [20]$$

avec :

$$v_m[\hat{F}_{k(i)}] = \frac{1}{m(m-1)} \sum_{j=1}^m \left(\frac{F_{k(i)l(j)}}{V_{k(i)l(j)}} - \hat{F}_{k(i)} \right)^2$$

$$\hat{V}_2 = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{\hat{F}_{k(i)}}{Z_{k(i)}} - F \right)^2 - \frac{1}{n^2} \sum_{i=1}^n \frac{1}{Z_{k(i)}^2} v_m[\hat{F}_{k(i)}] \quad [21]$$

$$\hat{V}_3 = \frac{1}{n} \sum_{i=1}^n \frac{1}{Z_{k(i)} m'} \sum_{j=1}^{m'} \frac{v[f_{k(i)l(j)}]}{V_{k(i)l(j)}} \quad [22]$$

où $v[f_{kl}]$ est un estimateur de la variance de f_{kl} .

RÉFÉRENCES

Cochran WG (1977) *Sampling techniques*. John Wiley and Sons, New York, 3rd edn, 428 p

Dagneaud JP, Tranchefort J (1985) Modèles de prévision de rendements. Bilan de quatre années de prévisions ITCF sur blé. *Perspect Agric* 92, 57-62

Feyerherm AM, Paulsen GM (1986) Development of a weather-yield function for winter wheat. *Agron J* 78, 1012-1017

Hodges T, Botner D, Sakamoto C, Hays Hang C (1987) Using the CERES-maize model to estimate production for the US cornbelt. *Agric For Meteorol* 40, 293-303

USDA-SRS (United States Department of Agriculture – Statistical Reporting Service) (1983) Scope and methods of the statistical reporting service. *Misc Publ* 1308