HAL
open science

Une contribution a la selection simultanee sur plusieurs caracteres: l'utilisation de la programmation lineaire

Francois-Xavier F.-X. Oury

To cite this version:

Francois-Xavier F.-X. Oury. Une contribution a la selection simultanee sur plusieurs caracteres: l'utilisation de la programmation lineaire. Agronomie, 1990, 10 (7), pp.517-524. hal-02710631

HAL Id: hal-02710631

https://hal.inrae.fr/hal-02710631v1

Submitted on 1 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Agronomie (1990) 10, 517-524
© Elsevier/INRA
MINISTÈRE DE L'AGRICULTURE
INRA - DOCUMENTATION
17, Rue Sully - B.V. 1540
21034 DIJON CEDEX
Tél. 80.63.30.02

Amélioration des plantes

Une contribution à la sélection simultanée sur plusieurs caractères : l'utilisation de la programmation linéaire

FX Oury

INRA, Station d'amélioration des plantes, Domaine de Crouelle, 63000 Clermont-Ferrand, France

(Reçu le 15 décembre 1989; accepté le 27 juin 1990)

Résumé — La programmation linéaire, outil mathématique souvent utilisé en recherche opérationnelle, permet de compléter l'approche par index en introduisant des contraintes sur les caractères qui ne peuvent être intégrés dans l'index. L'optimisation globale ainsi permise paraît donc intéressante pour sélectionner des génotypes lorsque l'information disponible concerne à la fois des caractères quantitatifs, et des caractères qualitatifs ne respectant pas les hypothèses de multinormalité nécessaires à leur prise en compte dans une régression géno-phénotypique.

programmation linéaire / sélection multicaractères / index avec contraintes

Summary — A contribution to multitrait selection: the use of linear programming. *Linear programming, a mathematical tool often used in operational research, enables to complete the selection index approach by imposing constraints on the characters which can not be included in the index (table II). The ensuing global optimization seems to be of interest for the choice of genotypes (table III and fig 1), when the available information simultaneously concerns quantitative traits, and qualitative characters which do not respect the multinormality hypothesis required for their integration in a genotype-on-phenotype regression.*

linear programming / multitrait selection / constrained selection index

INTRODUCTION

En amélioration des plantes, qu'il s'agisse de constituer des populations au départ d'un programme ou de maintenir la variabilité au cours des cycles successifs d'un schéma de sélection récurrente, une phase de description de la variabilité, existant au niveau de caractères aussi bien quantitatifs que qualitatifs, est nécessaire. Actuellement, cette diversité génotypique peut être évaluée au niveau agronomique mais aussi moléculaire (Lefort-Buson *et al*, 1988), et le nombre de données peut donc être très important.

Pour agir simultanément sur l'ensemble des caractères pour lesquels il dispose d'une information, le sélectionneur peut mettre en œuvre une sélection par niveaux indépendants. Cependant l'optimisation de celle-ci apparaît peu évidente lorsque beaucoup de caractères sont en jeu (Sampoux, 1989).

Il peut également utiliser un index statistique bâti à partir des prédictions des valeurs génétiques par régression géno-phénotypique (Vincourt et Gallais, 1983), après estimation des paramètres génétiques (pour les différentes méthodes utilisables à ce niveau, voir Plantevin-Bouchez, 1988). Cette sélection sur index est en théorie au moins aussi efficace que la sélection par niveaux indépendants (Gallais, 1973), mais elle pose le problème de l'attribution d'une valeur économique relative à chacun des caractères (Godshalk et Timothy, 1988; Itoh et Yamada, 1988). Cependant, pour ne pas avoir à fixer de manière plus ou moins arbitraire les pondérations économiques de l'index ou pour pouvoir contrôler l'évolution génétique de caractères ayant un poids économique nul, il est possible

d'introduire des restrictions qui fixent des valeurs (contraintes d'égalité) ou des limites de variation (contraintes d'inégalité), pour les progrès génétiques attendus sur les différents caractères (Mallard, 1972; Itoh et Yamada, 1986, 1987 et 1988; Sampoux, 1989). Il s'agit alors de trouver les pondérations qui procurent les progrès génétiques plus grands, tout en satisfaisant le système de contraintes imposé. Dans sa forme la plus générale, c'est un problème d'optimisation non linéaire avec contraintes, qui peut être résolu en faisant appel à des algorithmes de programmation quadratique.

La sélection par index avec estimation de la valeur génétique apparaît donc optimale, lorsqu'on a affaire à des caractères qui respectent les hypothèses théoriques de multinormalité. Elle est particulièrement intéressante lorsque ces caractères sont de type quantitatif, puisque l'estimation des valeurs génotypiques à partir des valeurs phénotypiques est alors cruciale (et cela d'autant plus que l'héritabilité des caractères est faible). Cependant, certaines informations dont dispose le sélectionneur ne peuvent pas être introduites dans une régression géno-phénotypique. C'est le cas pour des données relatives à des caractères qualitatifs dont les distributions sont très éloignées d'une distribution normale, ou pour des données concernant des couples d'individus (coefficient de parenté; divers types de distances). Une manière de traiter cette information est de la prendre en compte par l'intermédiaire de contraintes linéaires. Le problème de sélection simultanée sur plusieurs caractères ainsi formalisé, peut alors être résolu de manière optimale au moyen de la programmation linéaire, outil mathématique souvent utilisé en recherche opérationnelle.

Dans cet article, nous présentons un exemple d'application de cette méthode, qui nous a permis de faire un choix de génotypes en intégrant toutes les informations disponibles.

MATÉRIEL ET MÉTHODES

Données expérimentales

Afin de constituer une population au départ d'un schéma de sélection récurrente, 300 lignées de blé tendre d'hiver ont été évaluées sur plusieurs années et en plusieurs lieux. Ces génotypes ont été étudiés aussi bien au champ (observations en pépinières; essais «rendement» dans des dispositifs parcellaires en blocs complets randomisés), qu'au laboratoire (tests technologiques; électrophorèse des gluténines de hauts poids moléculaires; reconstitution des généalogies).

Les données concernant les sensibilités aux maladies étaient très nombreuses (les 11 lieux d'expérimentation ont fourni, en 2 années d'observation, 23 notations pour l'oïdium, 22 pour la rouille brune, et 23 pour la rouille jaune). Un premier regroupement a donc été réalisé au moyen d'analyses en composantes principales, les individus étant les 300 génotypes, et les variables les différentes notations correspondant à une même maladie. Les représentations des variables dans les plans factoriels définis par les 3 premiers axes principaux, et l'étude des corrélations entre ces variables et les composantes principales, ont permis de distinguer 7 groupes de notations pour l'oïdium, et 6 groupes de notations pour la rouille brune et pour la rouille jaune. Nous avons admis, par la suite, que pour chacune de ces maladies les différents groupes correspondaient à des associations de gènes de résistance différentes.

Un deuxième niveau de synthèse des données a pu être obtenu en créant un indice de résistance aux maladies I_M (valeur regroupant l'information relative aux sensibilités à l'oïdium et aux rouilles brune et jaune), et un indice de qualité technologique I_Q (valeur synthétisant l'information apportée par le test Pelshenke, la teneur en protéines et l'électrophorèse des gluténines (Branlard et Dardevet, communication personnelle). Par ailleurs, les données relatives au rendement ont conduit à la création d'un indice de productivité I_P.

À l'issue des expérimentations et après ces premiers regroupements de données, de nombreuses informations étaient disponibles (tableau I) et l'objectif était de les utiliser au mieux pour choisir les 24 lignées fondatrices de la population parmi les 300 candidates. C'est à ce niveau que la programmation linéaire peut être utilisée.

La programmation linéaire

Cet outil mathématique permet de résoudre les problèmes généraux dans lesquels une fonction linéaire, appelée fonction objectif, doit être optimisée (maximisée ou minimisée) sous un nombre fini de contraintes d'égalités et/ou d'inégalités linéaires.

En fait, ces problèmes correspondent à la recherche d'une combinaison optimale d'activités (ou variables). Il s'agit donc, d'un point de vue mathématique, de trouver un vecteur $(x_1, x_2 \ldots x_n)$ de l'espace à n dimensions, qui maximise (ou minimise) la forme linéaire :

$$z = \sum_{j=1}^{n} c_j x_j$$

soumise aux contraintes $\sum_{j=1}^{n} a_{ij} x_j \leq b_i$ pour $i = 1 \ldots m$

Tableau I. Information disponible pour les 300 génotypes après les 3 années d'évaluation.

Caractéristiques individuelles	
Caractéristiques agronomiques et morphologiques	
Précocité	
Composition en gènes de demi-nanisme (présence ou absence des gènes *Rht1* et/ou *Rht2*)	
Rendement	indice I_P
Résistance aux maladies	
Sensibilité à l'oïdium (présence ou non de chacune des 7 associations de gènes de résistance)	
Sensibilité à la rouille brune (présence ou non de chacune des 6 associations de gènes de résistance)	indice I_M
Sensibilité à la rouille jaune (présence ou non de chacune des 6 associations de gènes de résistance)	
Sensibilité au piétin-verse (présence ou absence du gène majeur de résistance apporté par *Aegilops ventricosa*)	
Valeur d'utilisation	
Valeur du test Pelshenke	
Teneur en protéines	indice I_Q
Diagramme gluténines	
Caractéristique relative à un couple d'individus	
Coefficient de parenté de Malecot	

et $x_j \geq 0$ pour $j = 1 \ldots n$;

c_j, a_{ij}, et b_i étant des constantes, n est le nombre de variables et m le nombre de contraintes.

Toute contrainte sous forme d'inéquation pouvant être ramenée à une équation par l'introduction d'une variable supplémentaire appelée variable d'écart, on peut également donner du problème la formulation matricielle suivante :

optimiser $z = C.x$ sachant que $A.x = B$; $x \geq 0$

avec z = fonction objectif; x = vecteur-colonne $(n,1)$ des niveaux des variables; c = vecteur-ligne $(1,n)$ des «coûts» (les c_j sont les contributions de chacune des variables à la fonction objectif); A = matrice (m,n) des contraintes (les a_{ij} représentent la part de la variable j dans la réalisation de la contrainte i); B = vecteur-colonne $(m,1)$ des «seconds-membres» (les b_i définissent les niveaux des contraintes).

La méthode de résolution de ces problèmes d'optimisation sous contraintes est de nature itéra-tive : à partir de la donnée d'un point initial x_0, un algorithme engendre une suite $x_1, \ldots, x_k$, qui converge vers l'optimum recherché. Dans le cas des programmes linéaires, qui sont des problèmes convexes, on sait alors que cet optimum est bien l'optimum global, et pas seulement un optimum local.

L'algorithme le plus souvent utilisé en programmation linéaire est l'algorithme du simplexe. En effet, bien qu'il en existe d'autres (Bland, 1981), l'algorithme du simplexe est en pratique le plus efficace : statistiquement, on montre qu'il converge en m à 3 m itérations (Gondran et Minoux, 1979). Pour une étude de son fonctionnement et plus particulièrement des opérations de «pivotage» correspondant à chaque itération) ou des considérations mathématiques qui ont permis sa conception, voir Dantzig (1966), Tirel (1967), Minoux (1983a)).

Outre les niveaux des variables et la valeur de la fonction objectif à l'optimum, l'algorithme du simplexe fourni les coûts marginaux des contraintes (un faible coût marginal indique que la contrainte correspondante a peu d'influence sur la valeur de la fonction objectif), et les coûts de substitution des variables. Ce sont des éléments importants pour l'interprétation de la solution d'un programme linéaire. Par exemple, dans le cas d'un problème de maximisation, les coûts marginaux permettent de savoir quelle augmentation de la fonction objectif serait permise par le relâchement de telle ou telle contrainte, et les coûts de substitution indiquent la baisse de la fonction objectif qui résulterait de l'adoption d'une combinaison de variables différente de celle fournie par la solution optimale.

En ce qui concerne notre recherche d'une combinaison de 24 génotypes choisis parmi 300, les variables sont les 300 lignées candidates, et les contraintes s'exercent au niveau des caractéristiques que l'on veut retrouver dans la population des 24 fondateurs. Nous avons réalisé la modélisation en programmation linéaire de la manière suivante.

Les c_j sont données par la valeur d'un index I, construit de manière empirique et prenant en compte les critères de résistance aux maladies, de qualité technologique, et de productivité : $I = I_M + I_Q + I_P$.

La valeur de la fonction objectif à l'optimum correspond donc à la somme des valeurs de l'index I, pour les 24 lignées retenues dans la solution.

Comme contraintes, nous avons imposé :

– une représentation égale des différentes classes de précocité (la solution devra donc comprendre 8 génotypes précoces, 8 génotypes intermédiaires, et 8 génotypes tardifs);

– chacune des associations de gènes de résistance mises en évidence par ACP pour l'oïdium, la rouille brune, et la rouille jaune, doit être représentée par au moins 2 individus;

– les allèles gluténines correspondant aux bandes 13-16 et 17-18 liées à la qualité, doivent être apportés à

la population par l'intermédiaire d'au moins un fondateur;

– l'absence d'apparentement entre les lignées fondatrices (pour cela, une étude de la matrice des coefficients de parenté nous a permis de distinguer 22 groupes constitués chacun par au moins 2 génotypes fortement apparentés; la contrainte correspond alors à l'impossibilité, pour 2 lignées retenues dans la solution, d'appartenir à un même groupe d'apparentement);

– une certaine proportion, à définir, des génotypes possédant le gène de demi-nanisme *Rht2*, ou le gène majeur de résistance au piétin-verse apporté par *Aegilops ventricosa*.

Le tableau II indique les différents éléments qui entrent dans la construction de la matrice de programmation linéaire. On remarque qu'un type supplémentaire de contrainte peut exister dans un programme linéaire : il s'agit des bornes (inférieures ou supérieures) qu'il est possible d'imposer pour chacune des variables. Notre problème apparaît alors comme un cas particulier où toutes les variables ne peuvent avoir pour niveau que 0 ou 1. En effet, si ces bornes n'existaient pas, le calcul algorithmique pourrait conduire à une solution où les génotypes contribuant le plus fortement à la fonction objectif seraient représentés plus d'une fois.

Les contraintes sur le non-apparentement des lignées fondatrices pourraient suffire à éviter la répétition des génotypes, car, tout individu étant apparenté à lui-même, aucun génotype ne pourrait alors être retenu plus d'une fois. Cependant, au niveau de la construction de la matrice de programmation linéaire, il est beaucoup plus simple de déclarer une borne supérieure pour l'ensemble des lignées que d'introduire autant de contraintes de non-apparentement qu'il y a de génotypes. En procédant ainsi, nous avons pu nous limiter à la définition de 22 contraintes de non-apparentement.

Comme l'algorithme du simplexe résoud les problèmes à variables continues, il est possible que la solution optimale comporte des «fractions» de génotypes. Cependant, dans la mesure où tous les coefficients a_{ij} et b_i sont des nombres entiers, le problème aura généralement une solution entière. Il est donc important que la formalisation du problème conduise à des coefficients a_{ij} et b_i entiers. De toutes façons, en cas de solution partiellement non entière, il est possible d'utiliser la programmation linéaire à variables mixtes. Cela ne nécessite pas de grosses modifications de la matrice de programmation linéaire (il suffit de déclarer les variables qui seront astreintes à ne prendre que des valeurs entières), car la plupart des logiciels de programmation linéaire possèdent un module de programmation en nombres entiers (pour l'étude théorique des algorithmes mis en œuvre, voir Attonaty et Tirel, 1969; Minoux, 1983b). Cependant, il faut savoir que le temps de calcul augmente très vite avec le nombre de variables entières : en effet, la solution optimale d'un problème à variables mixtes ne se déduit absolument pas de la solution en variables continues par simple arrondi des valeurs.

Pour notre application, les différentes optimisations ont été réalisées à l'aide du logiciel SCICONIC.

RÉSULTATS

Nous avons procédé à plusieurs interrogations du modèle, correspondant à différents jeux de contraintes.

En particulier, une interrogation avec relâchement total des contraintes illustre bien l'intérêt de la méthode. En l'absence de contraintes, l'optimum consiste à retenir les lignées correspondant aux 24 valeurs les plus fortes de l'index *I*. On constate alors (tableau III), que la solution optimale aurait conduit à une représentation insuffisante de l'association de gènes de résistance à l'oïdium n° 4, à la non-représentation des allèles gluténines correspondant aux bandes 13-16 et 17-18 liées à la qualité, à une sur-représentation des génotypes de précocité intermédiaire, et à une population pour laquelle plusieurs fondateurs auraient été apparentés.

Par ailleurs, nous avons représenté sur la figure 1 l'évolution de la fonction objectif, lorsqu'on soumet la population des fondateurs à une contrainte de plus en plus forte au niveau de sa composition en gènes *Rht2*, pour une proportion variable de génotypes possédant le gène majeur de résistance au piétin-verse (pour toutes ces interrogations, les autres contraintes concernant la précocité, les résistances à l'oïdium et aux rouilles brune et jaune, les allèles gluténines, et l'absence d'apparentement, sont réalisées). Les différentes courbes obtenues permettent alors de fixer des niveaux raisonnables pour les contraintes étudiées, en fonction de la diminution de la fonction objectif que l'on peut tolérer. Par exemple, si on accepte une baisse de 10% de la fonction objectif, une des solutions possibles est de retenir une population de fondateurs comprenant 18 génotypes *Rht2* et 1/4 de génotypes résistants au piétin-verse.

On peut remarquer que la contrainte sur *Rht2* n'est effective qu'au-delà d'un certain seuil. Par exemple, dans le cas où il n'y a pas de contrainte sur le nombre de génotypes résistants au piétin-verse, la maximisation de la fonction objectif conduit, par la simple réalisation des contraintes autres que celle sur *Rht2*, à une solution pour laquelle 13 lignées possèdent le gène *Rht2* (fig 1).

De même, il faut noter qu'il n'est pas possible d'obtenir une population constituée uniquement

Tableau II. Éléments nécessaires à la construction de la matrice de programmation linéaire.

Définition des contraintes	*Contrainte*	*(type)*
Caractéristiques soumises à contrainte		
Précocité		
variété précoce	PREC	(≤)
variété intermédiaire	INT	(≤)
variété tardive	TAR	(≤)
Présence du gène de demi-nanisme RhT2	RTH2	(≥)
Résistance de l'oïdium		
association de gènes de résistance n° p	OIDp	(≥)
Résistance à la rouille brune		
association de gènes de résistance n° q	RBq	(≥)
Résistance à la rouille jaune		
association de gènes de résistance n° r	RJr	(≥)
Résistance au piétin-verse	PV	(≥)
Présence des allèles gluténines liés à la qualité		
allèle correspondant aux bandes 13–16	G13–16	(≥)
allèle correspondant aux bandes 17–18	G17–18	(≥)
Appartenance à un groupe de génotypes apparentés		
groupe d'apparentement n° s	APPs	(≤)

Définition des variables

Variable	*Fonction objectif*	c_j	*Variable*	*Fonction objectif*	c_j
lignée y	F0	22	lignée z	F0	17

Variable	*Contrainte*	a_{ij}	*Variable*	*Contrainte*	a_{ij}
lignée y	PREC	1	lignée z	PREC	0
–	INT	0	–	INT	0
–	TAR	0	–	TAR	1
–	RHT2	1	–	RHT2	0
–	OIDp	1	–	OIDp	0
–	RBq	1	–	RBq	0
–	RJr	0	–	RJr	1
–	PV	1	–	PV	0
–	G13–16	0	–	G13–16	0
–	G17–18	1	–	G17–18	0
–	APPs	1	–	APPs	1

Définition des seconds membres

Contrainte	b_i
PREC	8
INT	8
TAR	8
RHT2	à déterminer
OIDp	2
RBq	2
RJr	2
PV	à déterminer
G13–16	1
G17–18	1
APPs	1

Définition des bornes

Variable	*Type de borne*	*Valeur de la borne*
lignée y	supérieure	1
lignée z	supérieure	1

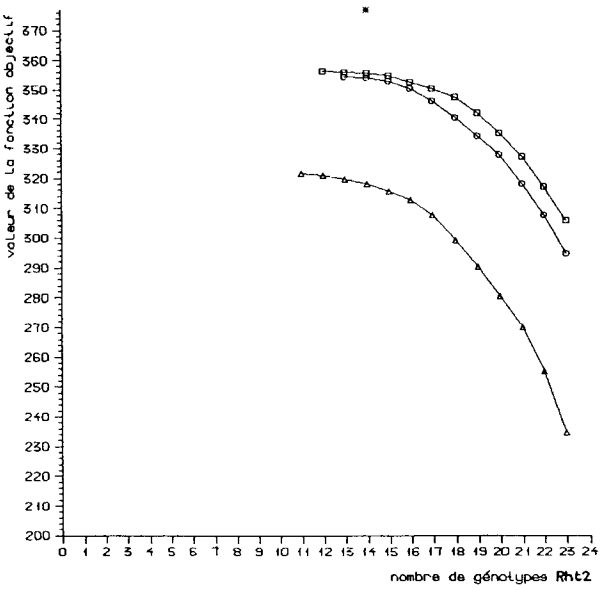

Fig 1. Évolution de la valeur de la fonction objectif à l'optimum, pour différentes interrogations du modèle de programmation linéaire : * interrogation sans contraintes; ❑ contrainte croissante sur *Rht2*, et absence de contrainte sur le nombre de génotypes résistants au piétin-verse; ○ contrainte croissante sur *Rht2*, et nombre de génotypes résis-tants au piétin-verse ≥ 6; Δ contrainte croissante sur *Rht2*, et nombre de génotypes résistants au piétin-verse ≥ 12.

de génotypes *Rht2* (fig 1). Ceci n'est pas dû à un nombre insuffisant de candidats possédant ce gène de demi-nanisme, mais au fait que le problème devient infaisable (il y a incompatibilité entre la réalisation des différentes contraintes).

Une autre manière de raisonner le choix des niveaux des contraintes consiste à étudier leurs coûts marginaux. Par exemple, dans le cas de l'interrogation où l'on impose 18 génotypes *Rht2* et 6 génotypes résistants au piétin-verse, les coûts marginaux des contraintes relatives à la précocité sont de 15 pour la contrainte PREC, 20,1 pour la contrainte INT, et 15,7 pour la contrainte TAR. Une plus forte disparité de ces coûts marginaux aurait pu nous conduire à imposer une contribution plus forte des génotypes intermédiaires. En effet, le désavantage dû au déséquilibre entre les différentes classes de précocité, aurait alors été compensé par une valeur globale de la population nettement supérieure.

De même, si les valeurs de ces coûts marginaux avaient été toutes les 3 très faibles, nous aurions pu décider de réduire la taille de la population pour que celle-ci ne comprenne pas de génotypes de mauvaise valeur propre.

Tableau III. Caractéristiques des populations de fondateurs correspondant à 2 interrogations contrastées du modèle de programmation linéaire.

Contraintes *	*Interrogation du modèle avec relâchement total des contraintes*	*Interrogation du modèle avec des contraintes*
PREC	7	8
INT	10	8
TAR	7	8
RHT2	14	18
OID1	19	17
OID2	16	14
OID3	5	4
OID4	1	2
OID5	2	3
OID6	9	6
OID7	4	4
RB1	24	24
RB2	21	20
RB3	17	18
RB4	3	2
RB5	18	19
RB6	16	16
RJ1	23	20
RJ2	16	15
RJ3	18	16
RJ4	12	10
RJ5	10	9
RJ6	17	16
PV	7	6
G13–16	0	1
G17–18	0	1
APP6	2	1
APP7	3	1
APP12	2	1
APP13	3	1
APP16	2	1

* Les chiffres correspondent au nombre de fondateurs réalisant les contraintes.

DISCUSSION

La modélisation en programmation linéaire nous a permis de prendre en compte simultanément un index (celui-ci représentant la contribution de chaque individu à la fonction objectif), et une information relative soit à des caractères n'entrant pas dans la construction de cet index, soit à des groupes d'individus (ces informations intervenant alors au niveau de la réalisation de contraintes). Nous avons pu ainsi obtenir, d'une part que des gènes majeurs intéressants soient représentés parmi les fondateurs, et d'autre part que ces fondateurs ne soient pas apparentés.

De plus, la comparaison des résultats fournis par plusieurs interrogations du modèle et l'examen des coûts marginaux des contraintes, nous ont permis de quantifier la diminution de la valeur globale des génotypes sélectionnés (perçue à travers la valeur de la fonction objectif) correspondant à tel ou tel paramétrage des se-conds membres. Nous avons pu disposer ainsi d'éléments pour retenir le système de contraintes le plus adéquat.

Dans notre exemple, nous avons utilisé un index phénotypique, et l'optimisation que nous avons réalisée ne portait donc pas sur les valeurs génétiques. Cependant, il est possible de prendre, comme coefficients *cj,* les valeurs d'un index statistique. On obtient ainsi une optimisation portant sur les valeurs génotypiques, avec la possibilité d'imposer des contraintes sur ces dernières. En effet, on peut remarquer que les contraintes imposées par la programmation linéaire ne s'exercent pas uniquement sur des caractères non pris en compte dans l'index. C'est ainsi que les contraintes relatives à la représentation des différentes associations de gènes de résistance pour l'oïdium et les rouilles brune et jaune, por-taient sur des caractères entrant dans la construction de l'indice I_M, et que les contraintes concernant les allèles gluténines liés à la qualité portaient sur une information intégrée dans la construction de l'indice I_Q. En cas d'utilisation des valeurs d'un index statistique comme cœfficients de la fonction objectif, les contraintes que l'on exerce sur des caractères entrant dans l'index sont alors équivalentes à des gains génétiques désirés. Cependant, ces contraintes ne sont pas de même nature que celles que l'on trouve au niveau des index de sélection avec restrictions. En effet, pour ce type d'index, on a des contraintes *a priori* sur les gains génétiques souhaités et ce sont ces contraintes qui permettent de déterminer les coefficients de l'index qui conduiront aux progrès génétiques les plus satisfaisants, qu'il s'agisse de contraintes de stricte proportionnalité (Itoh et Yamada, 1986 et 1987), ou de contraintes moins restrictives de type inégalités (Itoh et Yamada, 1988). En programmation linéaire, au contraire, nous avons affaire à des contraintes *a posteriori,* puisqu'elles s'exercent une fois les valeurs de l'index connues, et donc quand les coefficients de cet index sont parfaitement définis. Néanmoins, les résultats d'un choix d'individus par programmation linéaire ou par utilisation d'index avec restrictions, sont nécessairement convergents.

Cependant, ce n'est pas l'analogie avec les index avec restrictions qui représente l'aspect le plus intéressant de l'utilisation de la programmation linéaire. En effet, la méthode par index permet un classement des individus (sur leur valeur d'index) indépendant du nombre d'individus à retenir. En programmation linéaire, le classement des individus est également possible à partir des valeurs des coûts de substitution, mais ce classement dépend du jeu de contraintes employé, et donc en particulier de celles concernant la taille de la population à sélectionner. Il peut alors être nécessaire de procéder à plusieurs interrogations du modèle pour tester différents jeux de contraintes. Ceci est aisément réalisable, car le paramétrage des seconds membres est très simple, et l'analyse des valeurs duales (coûts marginaux et coûts de substitution) rend possible une interprétation très fine de la solution optimale, qui permet de ne pas trop multiplier les interrogations. Néanmoins, il y a là un inconvénient par rapport à une calcul unique d'index avec restrictions.

En fait, le véritable intérêt de l'utilisation de la programmation linéaire pour une sélection multicaractères est de permettre de compléter l'approche par index par une prise en compte des informations qui ne peuvent être introduites dans une régression géno-phénotypique. La démarche optimale consiste alors à prendre comme coefficients de la fonction objectif les valeurs d'un index statistique (avec ou sans contraintes) intégrant les caractères quantitatifs (et éventuellement certains caractères qualitatifs dont les distributions sont proches d'une distribution normale), et à introduire des contraintes linéaires sur les caractères qualitatifs à distribution très éloignée de la loi normale. À ce niveau, on peut se demander s'il n'est pas préférable de procéder à des transformations de variables permettant une normalisation de ces caractères, afin de pouvoir les introduire dans l'index. Mais outre le fait que cette normalisation n'est pas toujours possible (par exemple lorsque l'information correspond à la présence d'un gène de résistance ou à l'appartenance à un groupe d'apparentement), il faut également prendre en compte le fait que les problèmes d'estimation des valeurs génétiques deviennent importants lorsque trop de variables sont introduites dans la régression géno-phénotypique. Cette limitation du nombre de caractères pouvant être pris en compte n'existe pas en programmation linéaire, car c'est une technique de recherche d'optimum très puissante qui permet un niveau de désagrégation des modèles très élevé. C'est ainsi que dans notre application nous avons pu introduire

des contraintes sur 19 variables relatives aux différentes associations de gènes de résistance à l'oïdium et aux rouilles brune et jaune, sur 22 variables concernant les différents groupes de génotypes apparentés.

CONCLUSION

La programmation linéaire permet de prendre en compte l'ensemble de l'information disponible, aussi bien celle qu'il est possible d'intégrer dans un index que celle relative à des caractères qualitatifs ne respectant pas les hypothèses de multinormalité. L'optimisation globale ainsi permise paraît donc intéressante dans le cadre d'une sélection simultanée sur plusieurs caractères. De plus, l'algorithme du simplexe est facile à mettre en œuvre. Des logiciels très performants sont disponibles et, bien que le fichier informatique correspondant à la matrice de programmation linéaire puisse être assez volumineux, sa construction est aisée dans la mesure où les données existent déjà sous une forme structurée (par exemple dans une base de données). Par ailleurs, la taille des matrices correspondant aux problèmes couramment rencontrés en amélioration des plantes est très faible par rapport à celle des plus gros programmes linéaires traités en recherche opérationnelle (plusieurs milliers de lignes et de colonnes), ce qui autorise la résolution de ces problèmes sur un simple micro-ordinateur (sauf dans les cas où il faudrait faire appel à la programmation à variables mixtes).

RÉFÉRENCES

Attonaty JM, Tirel JC (1969) Programmation linéaire à variables mixtes : principes de résolution par la procédure SEP. INRA ESR, Grignon, 49 p

Bland RG (1981) Gestion et Programmation Linéaire. *Pour la Science* 46, 106-118

Dantzig GB (1966) Applications et prolongements de la programmation linéaire. Dunod, Paris, 434 p

Gallais A (1973) Sélection pour plusieurs caractères : synthèse critique et généralisation. *Ann Amélior Plant* 23, 183-208

Godshalk EB, Timothy DH (1988) Factor and principal component analyses as alternatives to index selection. *Theor Appl Genet* 76, 352-360

Gondran M, Minoux M (1979) Graphes et algorithmes. Eyrolles, Paris, 469-488

Itoh Y, Yamada Y (1986) Re-examination of selection index for desired gains. *Génét Sél Évol* 18, 499-504

Itoh Y, Yamada Y (1987) Comparisons of selection indices achieving predetermined proportional gains. *Génét Sél Évol* 19, 69-81

Itoh Y, Yamada Y (1988) Selection indices for desired relative genetic gains with inequality constraints. *Theor Appl Genet* 75, 731-735

Lefort-Buson M, Hebert Y, Damerval C (1988) Les outils d'évaluation de la diversité génétique et phénotypique. *Agronomie* 8, 173-178

Mallard J (1972) La théorie et le calcul des index de sélection avec restrictions : une synthèse critique. *Biometrics* 28, 713-735

Minoux M (1983a) Programmation mathématique : théorie et algorithmes tome I. Dunod, Paris, 294 p

Minoux M (1983b) Programmation mathématique : théorie et algorithmes tome II. Dunod, Paris, 1-54

Plantevin-Bouchez A (1988) Contribution à la construction et à la validation des index de sélection : application à la sélection du soja. Thèse, INP, Toulouse, 97 p

Sampoux JP (1989) L'index de sélection sur plusieurs caractères : une synthèse sur les pondérations économiques, le choix de contraintes et de caractères associés. *Agronomie* 9, 993-999

Tirel JC (1967) Aspects méthodologiques élémentaires de la programmation linéaire. *Bull Tech Inf* 216, 3-36

Vincourt P, Gallais A (1983) Sur la recherche de critères de sélection : la régression géno-phénotypique. *Agronomie* 3, 827-830