

# Analyse de dispositifs par approches itératives prenant en compte les performances des plus proches voisins

C Pichot

INRA, centre de recherches d'Orléans, station d'amélioration des arbres forestiers, 45160 Ardon, France

(Reçu le 1<sup>er</sup> septembre 1992; accepté le 4 décembre 1992)

**Résumé** — Quatre méthodes itératives d'analyse de dispositifs agronomiques sont comparées à partir de simulations. Elles prennent en compte, sous forme d'une covariable, les performances des plus proches voisins dans les 2 directions du terrain. Les 4 méthodes diffèrent par la définition des voisins d'une part, et du coefficient associé à la covariable d'autre part. Les résultats d'analyses en blocs complets et en aléatoire total sont également présentés. Les dispositifs simulés sont construits à partir de 4 fonds de carte représentant l'hétérogénéité du terrain, sur lesquels sont répartis aléatoirement des jeux déterminés de valeurs génotypiques associées à des effets résiduels aléatoires. Les méthodes d'analyse spatiale testées, de type Papadakis itéré, s'avèrent plus performantes que l'analyse en blocs lorsque le gradient naturel du terrain est irrégulier. Les variances résiduelles estimées se trouvent largement réduites, mais jamais sous-estimées comme l'indiquaient de précédents travaux. Les estimations des variances génétiques s'avèrent non biaisées. De plus, les analyses spatiales sont moins sensibles au nombre de répétitions de chaque génotype. La méthode Papadakis itérée utilisant comme covariable la moyenne des résidus voisins apparaît la plus adaptée en raison de l'information supplémentaire qu'elle apporte sur l'hétérogénéité du terrain.

**analyse spatiale / plus proche voisin / gradient / simulation**

**Summary** — *Iterated nearest neighbour analysis of field experiments.* Four iterated methods of analysis of agronomic trials were tested by simulation. Complete blocks and totally random analysis were also performed. The experimental trials were simulated with 4 different maps for soil fertility trend, added to controlled genetic values randomly placed within the trial and random values for error. Iterated spatial analysis proved more efficient than block analysis when the natural trend was irregular. Estimated residual variances were far lower than block analysis estimations but at no time did they underestimate the introduced value as previously stated by some authors. It was found that unbiased estimations of genetic variances were obtained. Estimations from spatial analysis were less sensitive to reduction of block number. Among the 4 spatial methods tested, the iterated Papadakis method, using the average of nearest neighbour residuals as covariate, provided more information regarding the heterogeneity trend of the trial.

**spatial analysis / nearest neighbour / trend / simulation**

## INTRODUCTION

L'analyse des dispositifs expérimentaux pour lesquels la performance des individus est influencée par leur position géographique relative, est d'autant plus efficace que la méthode utilisée rend précisément compte des effets de positionnement. Au moins 2 types d'essais sont concernés par ce problème : d'une part l'ensemble des tests agronomiques réalisés au champ (ou en serre) et d'autre part, les tests de

laboratoire pour lesquels les conditions expérimentales (température, éclairage, humidité...) ou les techniques utilisées (inoculation artificielle par pulvérisation...) fluctuent selon des gradients plus ou moins réguliers.

Plusieurs méthodes d'analyse ont été proposées pour limiter le bruit de fond causé par ces effets de «terrain». Tout d'abord, la structuration en blocs (Fisher, 1935) qui est à la base de nombreux types de dispositifs (blocs incomplets, lattice [Yates, 1936], carré latin...), puis la mé-

thode Papadakis (1937) utilisant comme covariable pour corriger les observations, la valeur des résidus des proches voisins; enfin, et beaucoup plus récemment, ont été développées des approches modélisant la covariance entre proches voisins (modèle autorégressif, random walk, etc). Pour plus de détails sur ces différents modèles, on pourra se référer aux revues bibliographique de Azais *et al* (1990) et Gleeson et Cullis (1991). Zimmerman et Harville (1991) considèrent les observations comme étant la réalisation d'un champ aléatoire, et modélisent l'hétérogénéité globale du dispositif sous forme d'une fonction polynomiale des coordonnées géographiques, associée à une modélisation de la covariance résiduelle entre proches voisins.

Parmi les méthodes classiques, la méthode en blocs complets offre l'avantage d'être particulièrement robuste, mais s'avère peu efficace. L'analyse de dispositifs plus élaborés (lattice, carré latin) pose problème dans le cas de données manquantes. Par construction, l'analyse en blocs modélise les effets de terrain en paliers discontinus, et ne peut donc qu'imparfaitement suivre les véritables gradients. Les méthodes récentes d'analyse spatiale peuvent être regroupées dans le cadre très général des processus ARIMA (*Auto Regressive Integrated Moving Average*). Actuellement, la modélisation unidirectionnelle peut être effectivement utilisée (logiciel Proxima, Bardin *et al*), mais l'extension en 2 dimensions proposée par Cullis et Gleeson (1991) se heurte à des problèmes informatiques dans le cas de dispositifs non rectangulaires ou de données manquantes. L'approche de Zimmerman et Harville pose également des problèmes de résolution informatique pour les dispositifs de grande taille. En ce qui concerne la méthode Papadakis, elle ne peut trouver de justification statistique dans la mesure où il n'y a pas indépendance entre variable et covariable, cette dernière étant constituée des résidus issus d'une première analyse (sans covariable). En 1978, Bartlett propose d'itérer la méthode Papadakis jusqu'à convergence, afin d'affiner les estimations respectives des niveaux des facteurs et des effets de terrain (covariable). À partir de tests de simulation réalisés sur des dispositifs sans effet traitement, Wilkinson *et al* (1983) concluent d'une part à une faible efficacité de l'analyse Papadakis non itérée, et d'autre part à une large sous-estimation de la variance résiduelle lorsque l'analyse est itérée, ce qui conduit à une sursignification de l'effet traitement. Le biais est d'autant plus prononcé que les effets de terrain suivent un fort

gradient. Ces résultats ont fait l'objet de nombreux commentaires rapportés dans le cadre des discussions publiées à la suite de l'article. Il apparaît que les estimations de la variance résiduelle ont été comparées à la variance estimée, sans prendre en compte les effets de terrain. La plus faible estimation de la variance résiduelle lors d'une analyse Papadakis est alors prévisible.

Des variances résiduelles estimées sur des dispositifs pour lesquels la vraie valeur résiduelle, indépendante du gradient, est connue, seraient au contraire surestimées lors d'une analyse Papadakis non itérée. Cette surestimation diminuerait sensiblement lors d'une analyse itérative (Laurence, in Wilkinson *et al*, 1983).

Une série de simulations a été effectuée pour 4 méthodes voisines, afin d'évaluer l'incidence d'une approche itérative empirique de type Papadakis, sur l'estimation des variances résiduelles et des traitements, dans le cas de dispositifs présentant différents types de gradients.

Le choix d'une méthode d'analyse prenant efficacement en compte les variations bidirectionnelles du terrain constitue l'objectif final de l'étude. Cette méthode est destinée à être utilisée sur des dispositifs forestiers pour lesquels l'hétérogénéité et mortalité rendent parfois inefficace, voire impossible, l'utilisation des structures en blocs initiales.

## MATÉRIEL ET MÉTHODES

### *Génération des dispositifs*

Les dispositifs simulés ci-après ne comportent qu'un seul facteur traitement, que nous appellerons Génotype, ce qui correspond aux dispositifs agronomiques de sélection. Ils sont constitués en 4 étapes : 1) utilisation d'un «fond de carte» simulant le gradient naturel du terrain (T); 2) génération des valeurs génétiques vraies (G); 3) répartition des génotypes sur le terrain; 4) adjonction d'un effet résiduel (ESP). Un individu Y, appartenant au génotype G<sub>g</sub> et positionné sur le terrain au point T, aura pour valeur :

$$Y_{gt} = G_g + T_t + EPS_{gt}$$

### Fonds de carte

Quatre types de gradient sont utilisés : gradient nul, gradient régulier, gradient lissé par une maille 3x3; noté «gradient 3x3», gradient lissé par une maille 5x5, noté «gradient 5x5» (fig 1).

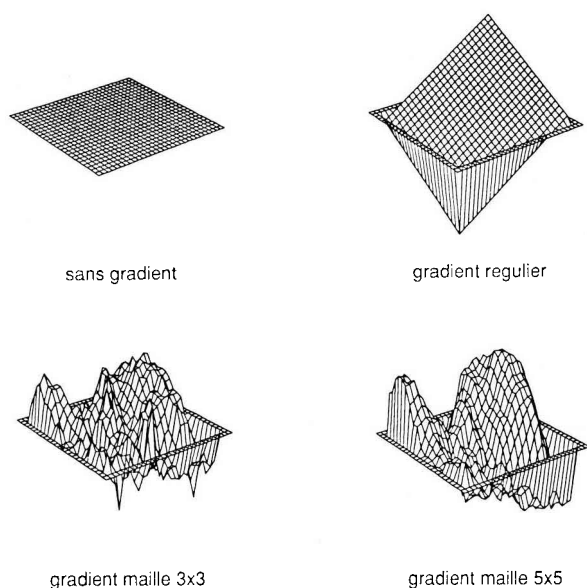


Fig 1. Fonds de carte utilisés pour les simulations.

Les 2 derniers fonds de carte correspondent à des lissages d'observations réalisées sur une partie du dispositif d'un essai de peupliers (*Populus trichocarpa* Torr et Gray) et ajustées aux effets génétiques, pour un seul caractère mesuré : la circonférence des arbres à 1,3 m. Ces 2 gradients ne diffèrent l'un de l'autre que par la taille de la maille servant au calcul de la moyenne mobile. L'individu au centre de la maille étant exclu, 8 (maille 3x3) ou 24 (maille 5x5) individus au plus contribuent à l'estimation de chaque point. Les amplitudes de variation relatives sont celles observées sur le dispositif de peupliers pour la circonférence des arbres. Les 4 fonds de carte ont des moyennes nulles et leurs amplitudes de variation sont décrites dans le tableau I.

**Valeurs génétiques**

Cent jeux de 65 génotypes ont été tirés dans une loi normale de moyenne 100 et de variance 150. Les 50 premiers génotypes de chaque jeu sont les 50 niveaux à estimer lors des simulations et constituent l'intérieur du dispositif. Les 15 derniers génotypes sont placés en bordure de dispositif. Les génotypes de bordure ont

Tableau I. Variabilité des 4 gradients de terrain.

Gradient	Minimum	Maximum	Écart type
Nul	0	0	0
Régulier	-24	24	10
3 x 3	-93	70	36
5 x 5	-49	53	26

été choisis différents des génotypes du dispositif afin de conserver le même effectif par génotype estimé lors des différentes analyses (voir «Méthodes d'analyse»).

**Répartition des génotypes sur le terrain**

Le dispositif expérimental comprend 10 blocs complets de 50 plants soit 1 individu par génotype et par bloc. Les blocs ont été arbitrairement placés, comme l'indique la figure 2. À l'intérieur de chaque bloc, les génotypes sont répartis aléatoirement. La répartition en blocs complets a été choisie dans un souci de simplicité de tirage.

Les génotypes de bordure ont également été répartis aléatoirement.

**Effet résiduel**

Aux effets de terrain et de génotype, a été ajouté un effet résiduel aléatoire tiré d'une loi normale de moyenne 0 et de variance 100. La variance résiduelle est telle que l'héritabilité théorique individuelle est de 60% (variance génétique / (variance génétique + variance résiduelle)). Cette valeur correspond à des estimations classiquement rencontrées lorsque l'hétérogénéité de terrain est relativement bien contrôlée.

**Méthodes d'analyse**

Les différentes méthodes d'analyse spatiale testées sont toutes bâties sur le principe d'estimation itérative des niveaux génétiques et du gradient de terrain jusqu'à convergence. Elles peuvent être regroupées sous le modèle :

$$Y_{gt} = M + G_g + T_t + EPS_{gt} \quad [1]$$

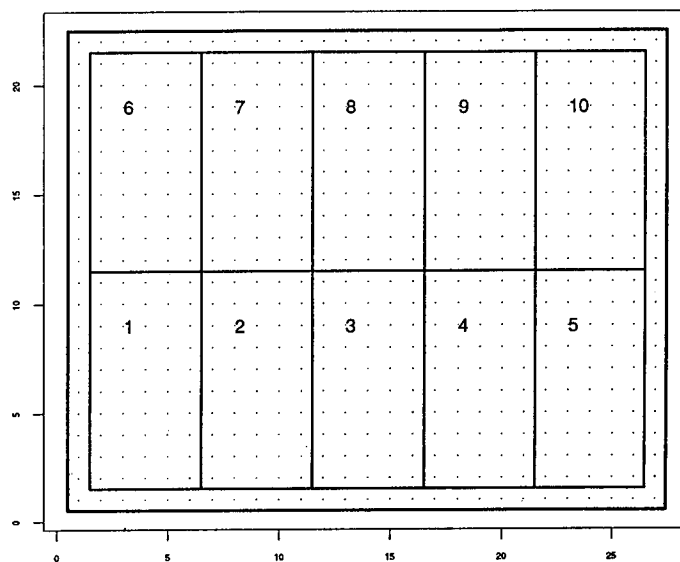


Fig 2. Position des blocs dans le dispositif.

où  $Y_{gt}$  : valeur observée d'un individu de génotype  $g$  à l'endroit  $t$ ;  $M$  : moyenne générale;  $G_g$  : effet du génotype  $g$ ;  $T_t$  : valeur du gradient de terrain au point  $t$ ;  $EPS$  : résiduelle.

Lors du processus itératif, à l'itération  $k$ ,  $T_{t(k)}$  est estimé par :

$$b_{(k)} \times \text{moy}(Y_v - M - G_{v(k-1)})$$

où  $\text{moy}(Y_v - M - G_{v(k-1)})$  est la moyenne calculée sur les génotypes voisins ( $v$ ), et  $b_{(k)}$  un coefficient défini ci-après.

Les méthodes se différencient selon la valeur de  $b_{(k)}$  et selon que l'individu central est ou non pris en compte au même titre que ses voisins dans l'estimation de  $\text{moy}(Y_v - M - G_{v(k-1)})$ .

Méthode A :  $b_{(k)} = 1$  et individu central exclu; B :  $b_{(k)} = 1$  et individu central inclus; C :  $b_{(k)}$  = coefficient de régression de  $\text{moy}(Y_v - M - G_{v(k-1)})$  sur  $(Y - G)$  et individu central exclu; D :  $b_{(k)}$  = coefficient de régression de  $\text{moy}(Y_v - M - G_{v(k-1)})$  sur  $(Y - G)$  et individu central inclus.

La méthode C correspond très exactement à l'approche Papadakis.

Pour tous les modèles, sont considérés comme «voisins» les individus appartenant à une même maille carrée de 3 sur 3, soit 8 ou 9 individus. Un poids identique est attribué à chacun de ces individus. Des simulations non présentées ici, ont montré qu'en l'absence de compétition entre plants, la maille bidirectionnelle la plus petite conduit au meilleur résultat.

### Méthode itérative d'estimation

L'itération est spécifiée en indice et entre parenthèses.

À l'itération 0, on pose :  $G_g(0) = 0$

$$\begin{aligned} \text{itération 1} \quad T_{t(1)} &= b_{(1)} \times \text{moy}(Y_v - M - G_{v(0)}) \\ &= b_{(1)} \times \text{moy}(Y_v - M) \\ G_{g(1)} &= 1/r \times \sum_t (Y_{gt} - M - T_{t(1)}) \end{aligned}$$

$$\begin{aligned} \text{itération } k : \quad T_{t(k)} &= b_{(k)} \times \text{moy}(Y_v - M - G_{v(k-1)}) \\ G_{g(k)} &= 1/r \times \sum_t (Y_{gt} - M - T_{t(k)}) \end{aligned}$$

$$\begin{aligned} \text{à la convergence } G_{g(k)} &= G_{g(k-1)} \\ T_{t(k)} &= T_{t(k-1)} \end{aligned}$$

( $r$  : nombre de répétitions du génotype  $g$ ).

La convergence est considérée atteinte lorsque la différence maximale entre  $T_{t(k)}$  et  $T_{t(k-1)}$  est inférieure au 1/500 de la valeur maximale de  $T_{t(k)}$ .

Enfin, et à titre d'étalonnage, 2 analyses de variances «classiques» ont été effectuées. La première (E) ne prend en compte que l'effet génotype, et la seconde (F) associe les effets bloc (complet) et génotype.

### Comparaison des approches

Pour chacun des 4 fonds de carte, 100 dispositifs correspondant aux 100 jeux de génotypes, ont été simulés. Pour chaque dispositif, les 6 différentes méthodes ont été testées.

La concordance entre valeurs vraies et estimées nous a permis de comparer les méthodes. Les estimations concernent : le carré moyen résiduel (ou variance résiduelle) et la variance génotypique (=  $(CM_{\text{génotype}} - CM_{\text{résiduelle}}) / \text{nombre de répétitions}$ ).

De plus, pour les dispositifs «gradient régulier» et «gradient 3x3», l'incidence du nombre de répétitions sur les estimations des paramètres, a été testée par 3 autres séries de 100 simulations effectuées avec respectivement 6, 4 et 2 blocs. Les blocs ont été éliminés de façon systématique en partant des positions situées les plus à droite sur la figure 2.

### Outil informatique

La création des dispositifs, leur analyse, ainsi que les figures présentées dans cette publication, ont été réalisées avec le logiciel SPLUS (1990) implanté sur un serveur Sun 4/390.

## RÉSULTATS

### Estimation des niveaux génétiques

Quel que soit le fond de carte utilisé, les analyses spatiales (A à D) donnent des résultats semblables (fig 3). Les corrélations entre les valeurs génotypiques vraies et estimées ne décroissent sensiblement que pour le dispositif «gradient 3x3». Dans le cas de gradient nul ou régulier, les estimations issues des analyses E et F qui correspondent simplement aux moyennes par génotype des valeurs observées, sont comparables à celles issues des analyses spatiales. En revanche, des écarts significatifs apparaissent sur les dispositifs «gradient 3x3» et «5x5». Les estimations effectuées à partir des analyses spatiales sont à la fois plus proches des valeurs réelles et moins variables.

### Estimation des carrés moyens

#### Carré moyen de l'effet génotype

Dans le cas de dispositifs à gradient irrégulier, les analyses E et F conduisent à des valeurs plus élevées (fig 4).

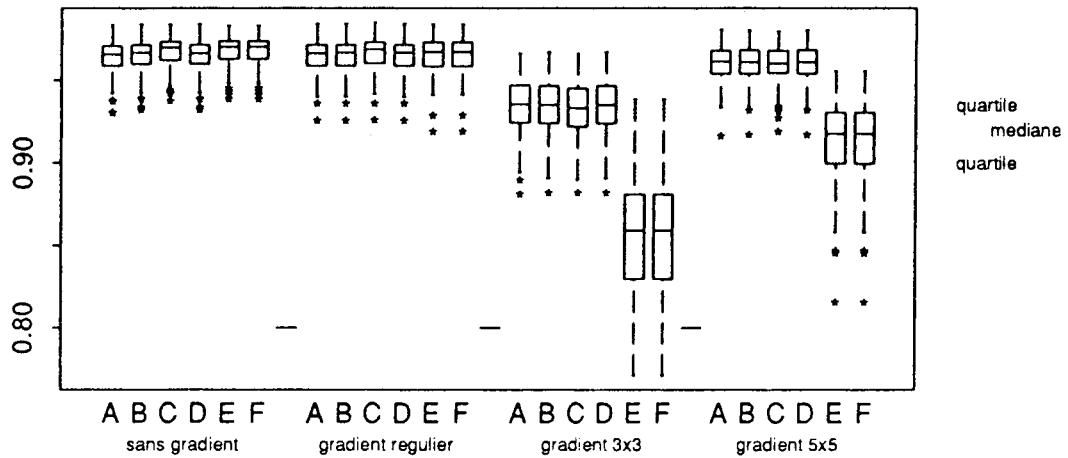


Fig 3. Corrélations entre valeurs génotypiques vraies et estimées.

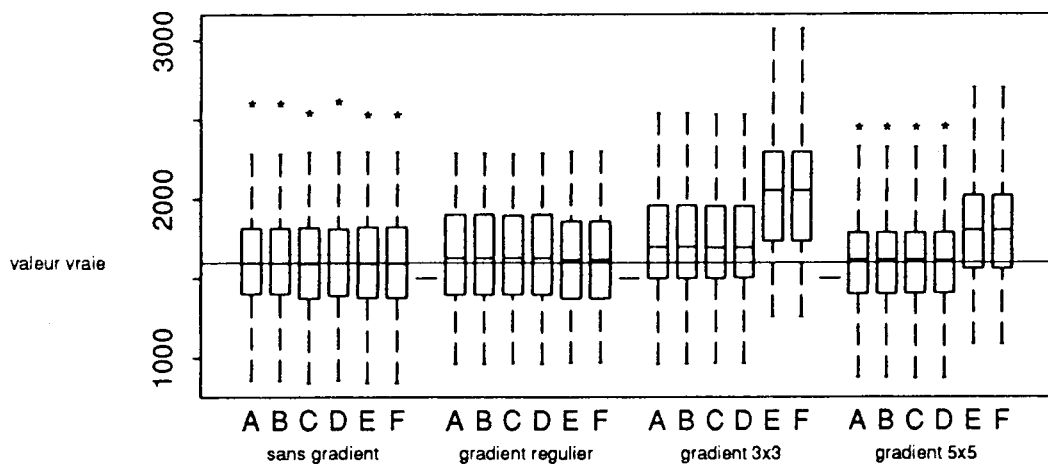


Fig 4. Carré moyen de l'effet génotype.

### Carré moyen résiduel

La prise en compte de l'individu au centre de la maille (analyses B et D), conduit à des variances résiduelles plus petites. Lorsque le dispositif est sans gradient ou à gradient régulier, les estimations de la variance résiduelle sont alors inférieures à celle introduite (fig 5). Les analyses A et C d'une part et B et D d'autre part, donnent les mêmes résultats.

Pour les dispositifs à gradient, les analyses E et F surestiment la variance résiduelle, surtout pour les dispositifs «3x3» et «5x5». La réduction de la variance résiduelle entre les modèles E et F (prise en compte des blocs) est du même ordre que celle observée entre le modèle bloc (F) et les méthodes spatiales (A à D).

### Estimation de la variance génotypique

Les estimations de la variance génotypique apparaissent peu biaisées, sauf celles réalisées par l'analyse E sur les dispositifs à gradient irrégulier, qui conduisent à de fortes sous-estimations (fig 6). Pour l'analyse en blocs, les surestimations des carrés moyens génotypiques et résiduels se compensent et conduisent à des variances génotypiques non biaisées, résultats attendus dans le cadre de la théorie de la randomisation.

La légère surestimation obtenue à partir des analyses B et D est liée à la sous-estimation de la variance résiduelle, comme nous l'avons précédemment remarqué.

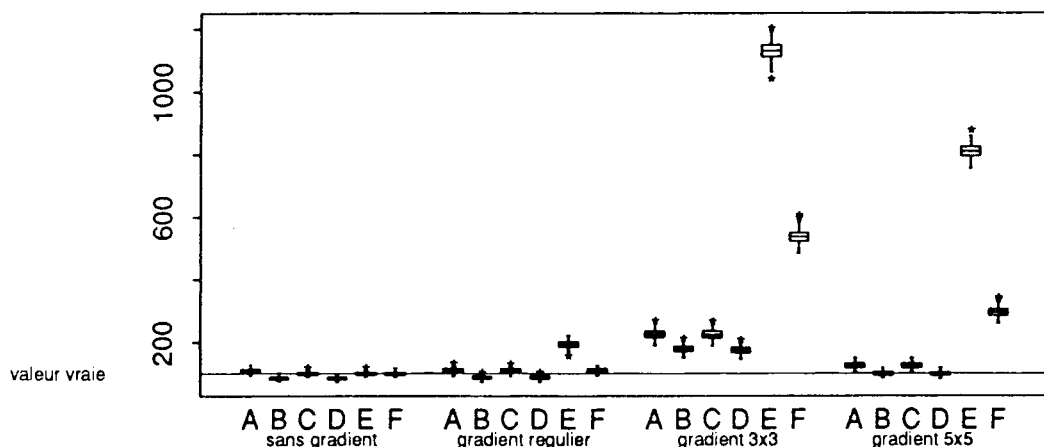


Fig 5. Carré moyen résiduel.

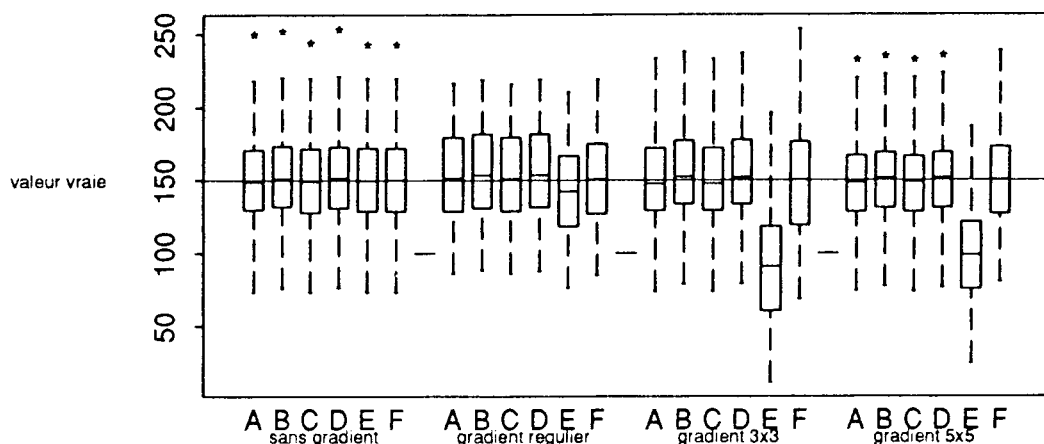


Fig 6. Variance génotypique estimée.

### ***Précision des estimations***

#### **Variance génotypique**

Les 100 simulations effectuées pour chaque dispositif correspondent à 100 jeux différents de génotypes. L'effet tirage ne permet pas d'estimer sans biais la variance d'estimation de la variance génotypique estimée. En revanche, la corrélation entre les variances estimées et les véritables variances de chacun des 100 jeux traduit la fiabilité de l'estimation.

Les 4 types d'analyse spatiale sont équivalents et nettement plus performants que les modèles E et F pour des dispositifs à gradient (fig 7).

#### **Variance résiduelle**

En ce qui concerne les approches spatiales, la variance d'estimation de la variance résiduelle apparaît systématiquement plus faible pour les modèles B et D (fig 8). En valeur absolue, la différence est d'autant plus importante que le gradient est irrégulier.

#### ***Incidence du nombre de répétitions***

#### **Variance génotypique**

La réduction du nombre de répétitions (ou blocs complets) entraîne une chute de précision sur l'estimation de la variance génotypique (fig 9).

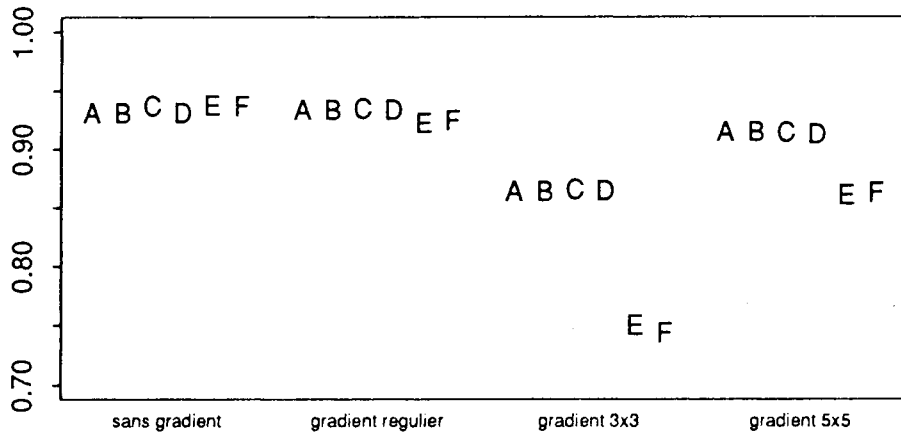


Fig 7. Corrélation entre variances génotypiques vraies et estimées.

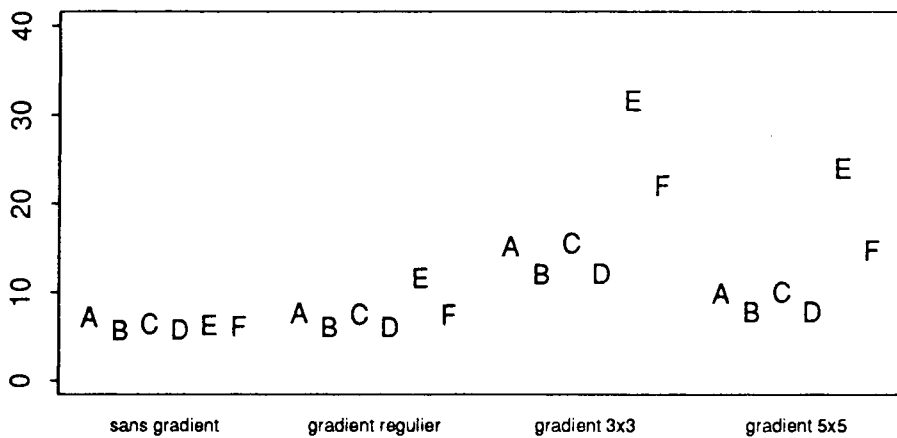


Fig 8. Écart type de la variance résiduelle.

### Variance résiduelle

Le passage de 10 à 2 blocs a peu d'incidence sur l'estimation de la variance résiduelle pour le dispositif à gradient régulier. En revanche, pour le «gradient 3x3», les analyses E et F conduisent à des estimations beaucoup plus fluctuantes que celles issues des analyses spatiales, et toujours largement supérieures à la variance introduite (fig 10). Il faut toutefois souligner que les estimations dépendent directement des blocs analysés, dans la mesure où ceux-ci n'ont pas été tirés aléatoirement, mais systématiquement (voir «Comparaison des approches»).

La réduction du nombre de blocs semble accentuer la sous-estimation pour les analyses B et D.

Enfin, la variance d'estimation augmente beaucoup moins pour les méthodes A à D que pour E et F (fig 11). Les analyses spatiales permettraient

donc de réduire le nombre de répétitions par traitement.

### DISCUSSION

Le reproche majeur et rédhibitoire fait par Wilkinson *et al* (1983) à l'analyse Papadakis itérée, était la sous-estimation de la variance résiduelle. Les résultats de nos simulations ne mettent en évidence aucune sous-estimation pour les méthodes A et C, mais concluent au contraire à une surestimation, d'autant plus forte que le gradient est irrégulier. Ces résultats concordent avec ceux obtenus par Laurence (*in* Wilkinson *et al*, 1983). Toutefois, la prise en compte de l'individu central au même titre que ses voisins, lors de l'estimation du gradient, diminue la variance résiduelle. Cette observation peut être facilement expliquée : soient  $v$  le nombre de voisins

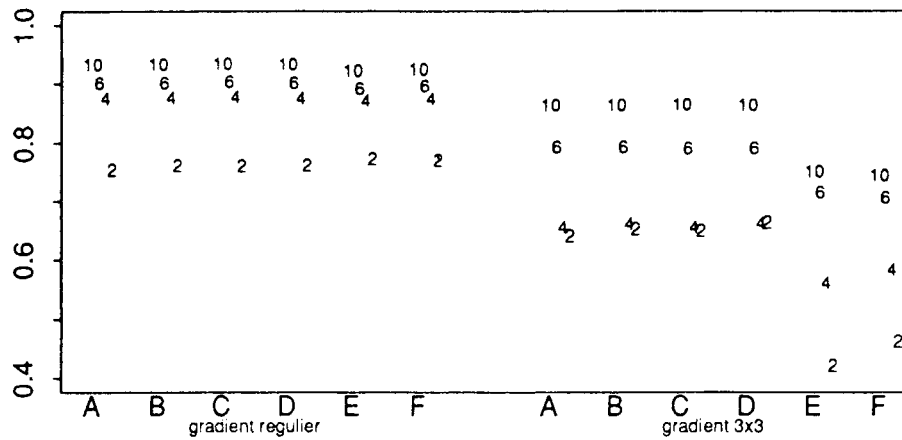


Fig 9. Incidence du nombre de répétitions sur la corrélation entre variances génotypiques vraies et estimées.

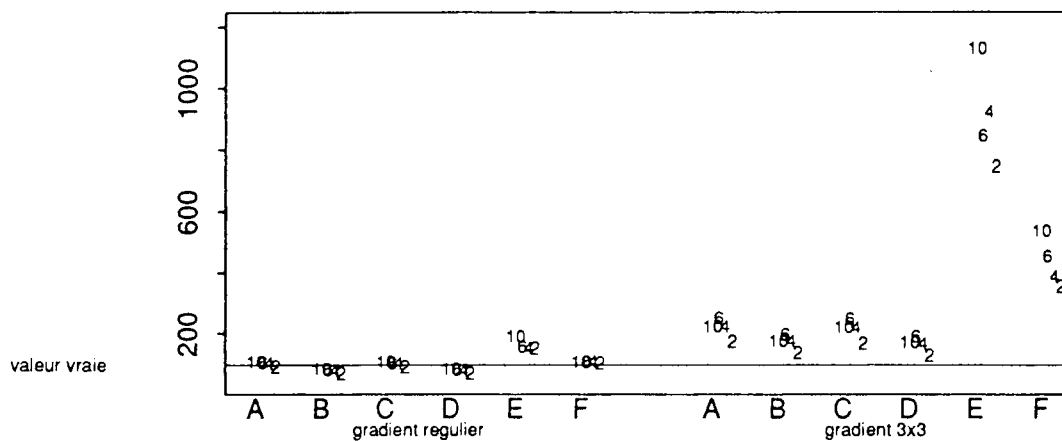


Fig 10. Incidence du nombre de répétitions sur l'estimation de la variance résiduelle.

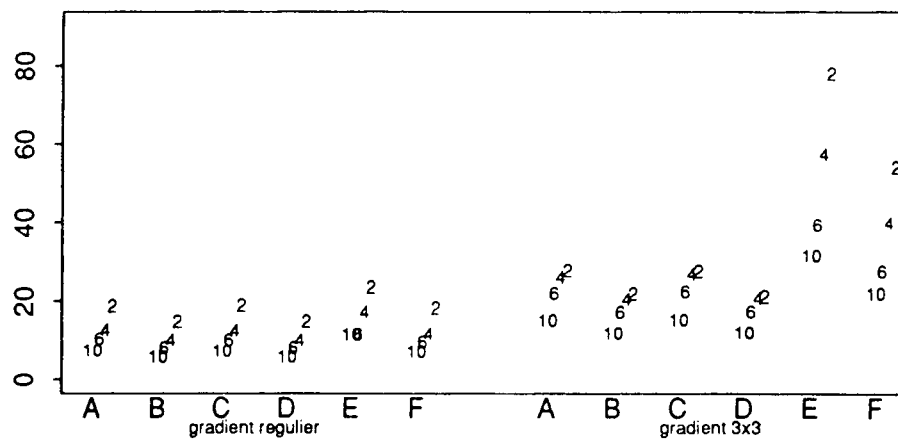


Fig 11. Incidence du nombre de répétitions sur l'écart type de la variance résiduelle.

de l'individu situé au lieu  $t$ ,  $EPS_{gt}$  la valeur résiduelle associée aux estimations de  $G$  et  $T$  (génotype et terrain). En considérant la moyenne générale nulle, la résiduelle du modèle (I) s'écrit :

$$EPS_{gt} = Y_{gt} - G_g - T_t$$

À la convergence :

$$EPS_{gt} = Y_{gt} - G_g - b \times (1/v \times \sum_v (Y_v - G_v))$$

Méthode A

$$EPS_{gt}(a) = Y_{gt} - G_g - 1/v \times \sum_v (Y_v - G_v)$$



Méthode B (individu central inclus)

$$T_t = 1/(v + 1) \times (\sum_v(Y_v - G_v) + (Y_{gt} - G_g))$$

d'où :

$$EPS_{gt}(b) = v/(v+1) \times (Y_{gt} - G_g - 1/v \times \sum_v(Y_v - G_v))$$

en faisant l'hypothèse, confirmée par les simulations (tableau II), que les estimations de  $G_g$  et  $T_t$  sont égales à celles de A.

L'équation est alors celle du modèle A excluant l'individu central, avec pour seule différence le coefficient multiplicatif  $v/(v + 1)$ .

D'où

$$EPS_{gt}(b) = (v/(v + 1) \times EPS_{gt}(a))$$

Le rapport entre les estimations A et B de la résiduelle est donc de  $(v + 1)/v$  soit 9/8 dans notre cas.

Méthodes Papadakis (C et D)

Soient  $b(c)$  et  $b(d)$  les coefficients de régression intervenant dans l'estimation du gradient.

On a alors sous l'hypothèse d'égalité des estimations des G et T :

$$EPS_{gt}(d) = ((v + 1 - b(d)) / (v + 1)) \times \{(Y_{gt} - G_g) - (b(d) / (v + 1 - b(d))) \times \sum_v(Y_v - G_v)\}$$

$$EPS_{gt}(d) = ((v + 1 - b(d)) / (v + 1)) \times \{EPS_{gt}(c) + (b(c) / v - (b(d) / (v + 1 - b(d)))) \times \sum_v(Y_v - G_v)\}$$

Contrairement au passage de A à B, la réduction d'estimation de la variance résiduelle lors du passage de C à D ne dépend pas seulement de la taille de la maille. Elle est également fonction des coefficients de régression. Dans ce cas, le rapport des variances estimées dépend du type

**Tableau II.** Moyenne et écart type de la corrélation entre les valeurs génétiques estimées par les méthodes A et B.

Gradient	Corrélation ( $G_g(a), G_g(b)$ )	
	Moyenne	Écart type
Nul	0,999 93	8.10 <sup>-5</sup>
Régulier	0,999 99	3.10 <sup>-6</sup>
3 x 3	0,999 74	1.10 <sup>-4</sup>
5 x 5	0,999 94	2.10 <sup>-4</sup>

de dispositif (fond de carte), du nombre de répétitions et du tirage aléatoire du dispositif. Les simulations expérimentales confirment ces observations (tableau III).

Le passage de C à D réduit plus les variances résiduelles que les écarts types de celles-ci.

Les simulations effectuées en l'absence de gradient ou pour un gradient régulier, concluent à une sous-estimation de la variance résiduelle lorsque l'individu central est pris en compte. Les méthodes A et B sont similaires dans la mesure où cette sous-estimation peut être corrigée *a posteriori*. En revanche, comme il n'est pas possible d'évaluer le niveau de sous-estimation de la variance résiduelle pour le modèle D, il est donc préférable d'utiliser le modèle C. Nos simulations ne mettent en évidence aucune différence entre les méthodes A et C. L'analyse A ne nécessite pas de calcul du coefficient de régression b. Toutefois, la valeur des corrélations entre résidus voisins est très informative sur le gradient de terrain, ce qui donne la préférence à l'approche Papadakis itérée (C).

L'approche spatiale s'avère largement supérieure à l'analyse en blocs lorsqu'il existe un gradient de terrain, et conduit à des résultats équivalents en l'absence de gradient. Bien entendu, la structure en blocs complets n'est pas la plus performante des structures en blocs. Il serait plus intéressant de comparer les résultats à ceux obtenus avec des blocs incomplets. En raison des travaux de Wilkinson *et al* concluant à la sous-estimation de la variance résiduelle,

**Tableau III.** Rapport des variances résiduelles et des écarts types de ces variances lors du passage des modèles A à B et C à D ( $1,266 = (9/8)^2$ ).

Fond de carte	Rapport des variances estimées		Rapport des écarts types des variances	
	A/B	C/D	A/B	C/D
Sans gradient	1,266	1,154	1,266	1,101
Gradient régulier	1,266	1,254	1,266	1,230
Gradient 3x3	1,266	1,283	1,265	1,274
Gradient 5x5	1,266	1,275	1,266	1,271
Rég 6 blocs	1,266	1,256	1,265	1,216
Rég 4 blocs	1,266	1,264	1,266	1,243
Rég 2 blocs	1,266	1,289	1,266	1,277
3x3 6 blocs	1,266	1,294	1,266	1,280
3x3 4 blocs	1,265	1,300	1,266	1,288
3x3 2 blocs	1,266	1,301	1,265	1,299

l'analyse Papadakis a été rarement itérée. Néanmoins, les travaux publiés sur le sujet font état de la supériorité de l'approche Papadakis (Pearce, 1980). Kempton (1985 in Gleeson and Cullis, 1991) estime à 14% le gain de précision sur la différence entre traitements pour un dispositif de blé. L'analyse Papadakis conduit à des meilleurs résultats que ceux obtenus avec des modèles spatiaux unidirectionnels et s'avère très proches d'autre modèles bidirectionnels (Zimmerman et Harville, 1991). De même, Bisoffi (1992) estime que l'approche Papadakis est au moins aussi performante que l'analyse en lattice équilibrée pour des dispositifs de peupliers en pépinière.

### **Perspectives**

L'analyse Papadakis itérée prend un intérêt tout particulier lorsque :

- la conception initiale du dispositif ne permet pas de prendre en compte les gradients de terrain (randomisation totale, blocs complets, confusion entre un effet traitement et l'hétérogénéité de terrain...);
- des accidents rendent inutilisable la structuration initiale (forte mortalité...);
- le nombre de répétitions par niveau de traitement est faible (et/ou très hétérogène).

Les simulations n'ont ici porté que sur une variable, mais il serait très intéressant d'évaluer l'indidence de l'analyse Papadakis sur l'estimation des corrélations entre caractères pour lesquels les gradients liés au terrain seraient différents.

Pour un dispositif de peupliers répété sur 2 sites, les corrélations intersites entre performances de croissance sont plus élevées pour des analyses Papadakis itérées que pour des analyses en blocs incomplets. Les estimations de l'interaction génotype x environnement s'en trouvent alors réduites (non publié).

De plus, l'approche permet une visualisation du gradient naturel qui peut être décrit d'une part par la valeur de la corrélation entre résidus voisins et d'autre part par la dispersion globale sur l'ensemble du dispositif. Des mesures successives, par exemple pluriannuelles, sur un même essai peuvent ainsi faire apparaître des déplacements de «zones fertiles» et des variations de corrélation entre voisins généralement dues aux effets de compétition. La concurrence entre indi-

vidus induit des corrélations négatives entre proches voisins et pose alors le problème de la modalité d'estimation du gradient. Un élargissement de la maille (5x5, voire 7x7) s'avère parfois efficace (résultats non publiés). Il serait très certainement préférable d'utiliser une pondération des résidus voisins en fonction de la distance, voire de l'orientation. Une étude de la structure entre voisins, de type corrélations sérielles, pourrait être associée à l'analyse Papadakis itérée. Des simulations supplémentaires sont nécessaires pour évaluer l'intérêt d'une telle approche.

### **CONCLUSION**

La puissance des méthodes d'analyse de dispositifs est très souvent limitée par des effets de gradients naturels ou induits qui ne sont que partiellement contrôlés par la structure du dispositif.

La méthode Papadakis s'avère un bon outil d'analyse pour ces cas de figures. Les simulations effectuées sur différents types de gradients infirment les résultats de Wilkinson *et al* (1983) concernant la sous-estimation de la variance résiduelle lors de l'itération du processus. D'un point de vue théorique, l'analyse Papadakis itérée est nécessairement inférieure aux approches spatiales reposant sur de véritables modèles statistiques. D'un point de vue empirique, les résultats s'avèrent très satisfaisants. Informatiquement, l'approche de type Papadakis est la seule méthode bidimensionnelle utilisable, quel que soit le dispositif. Elle mérite sans doute une plus large utilisation, notamment pour les dispositifs mal structurés.

### **REMERCIEMENT**

Nous remercions messieurs JB Denis et H Monod pour leur lecture critique du manuscrit. Ces travaux ont été réalisés à la station d'amélioration des arbres forestiers (INRA Orléans), dans le cadre d'une thèse sur l'amélioration des peupliers, cofinancée par le conseil régional de la région Centre et le fond régional de la Maîtrise de l'énergie-région Centre.

### **RÉFÉRENCES**

- Azais JM, Denis JB, Dhome T, Kobilinsky A (1990) Neighbour analysis of plot experiments: a review of the different approaches. *Biom Praxim* 30, 15-39

- Bardin A, Azais JM, Durier C. Proxima version 1.0. Notice d'utilisation. INRA, laboratoire de biométrie, Versailles, 38 p
- Bartlett MS (1978) Nearest neighbour models in the analysis of field experiments (with discussion). *JR Statist Soc B*, 40, 147-174
- Bisoffi S (1992) Nearest-neighbours and check-plots in poplar nursery trials. Proceedings IUFRO Meeting S4.11, September 10-13, 1991, London (in press)
- Cullis BR, Gleeson AC (1991) Spatial analysis of field experiments – An extension to two dimensions. *Biometrics* 47, 1449-1460
- Fisher RA (1935) *The design of experiments*. Oliver and Boyd, London, 244 p
- Gleeson AC, Cullis BR (1991) Recent developments in the analysis of plants breeding trials
- Papadakis JS (1937) Méthode statistique pour des expériences au champ. *Thessaloniki Plant Breed Instit Sci Bull* 23, 30 p
- Peace SC (1980) Randomized blocks and some alternatives: a study in tropical conditions. *Trop Agric* 57, 1-10
- Splus (1990) Manuel de référence. Version 2.3. Statistical Sciences, Inc, Seattle, Washington
- Wilkinson GN, Eckert SR, Hancock TW, Mayo O (1983) Nearest neighbour (NN) analysis of field experiments. *JR Statist Soc B* 45(2), 151-211
- Yates F (1936) Incomplete randomized blocks. *Ann Eugenics* 7, 121-140
- Zimmerman DL, Harville DA (1991) A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics* 47, 223-239