

Aptitude des modèles des plus proches voisins à la réduction de l'erreur expérimentale en expérimentation agronomique sur champ : Une illustration

Title: Aptitude of the nearest neighbors' models to reduce the experimental error on the field trial

Tahar Sghaier¹

Résumé : Le présent travail illustre par l'exemple l'importance de l'utilisation des modèles des «plus proches voisins» pour réduire l'erreur expérimentale en expérimentation agronomique sur champ où les dispositifs expérimentaux classiques ne peuvent en général que contrôler partiellement l'hétérogénéité du terrain. Cette erreur expérimentale est considérée comme le facteur explicatif principal de la puissance d'une expérience. Les résultats obtenus ont montré que :

- l'effet des itérations du modèle des plus proches voisins (MPPV) utilisé sur la réduction de l'erreur expérimentale est très positif,
- l'analyse des données brutes selon un dispositif en blocs aléatoires complets a permis de réduire le carré moyen résiduel de 38% par comparaison à une analyse des mêmes données selon un dispositif complètement aléatoire,
- l'analyse des données ajustées à l'effet terrain à l'aide du MPPV selon un dispositif complètement aléatoire a permis de réduire le carré moyen résiduel de 77% par comparaison à une analyse des données brutes selon le même dispositif et de 63% par comparaison à une analyse des mêmes données brutes selon un dispositif en blocs aléatoires complets,
- la prise en considération de la structure des blocs n'a aucun effet sur la réduction du carré moyen résiduel, une fois les données ajustées à l'effet terrain à l'aide du MPPV utilisé.

Abstract: The present work illustrates the importance of the nearest neighbors' models to control the experimental error on field experiments where the classical experimental designs can control only partially this experimental error, regarded as the principal explanatory factor of the power of an experiment. The obtained results showed:

- the positive effect of the iterations of the used nearest neighbors model to reduce the experimental error,
- analyses of the original data according to a randomized complete blocks design reduced the mean square error with 38% by comparison to the analyses of the same original data according to a completely randomized design,
- the one way analyses of variance computed on the adjusted data with the nearest neighbors model reduced the mean square error with 77% and 63% by comparison to one and two way analyses of variance computed on the original data respectively.
- after correction of the data to the ground heterogeneity by using the developed nearest neighbors' model, taking into the account of blocks does not have any effect on the reduction of the experimental error.

Mots-clés : Expérimentation, hétérogénéité du terrain, dispositif expérimental, erreur expérimentale, modèle des plus proches voisins.

Keywords: Experimentation, ground heterogeneity, experimental design, experimental error, nearest neighbors model.

Classification AMS 2000 : 62H, 62J, 62K, 62P10, 62P12, 65K

¹ Laboratoire d'Ecologie Forestière, Institut National de Recherches en Génie Rural, Eaux et Forêts, Université de Carthage. BP 10, 2080 Ariana - Tunisie.

Email : sghaier.tahar@iresa.agrinet.tn

1. Introduction

En expérimentation sur champ, l'hétérogénéité du terrain non contrôlée par le modèle d'analyse statistique utilisé affecte directement la fiabilité des résultats et les conclusions qui peuvent être tirées de ces expérimentations. En effet, l'analyse statistique des essais agronomiques au champ a pour but de dissocier l'effet des traitements ou des objets de l'expérience de celui du terrain plus ou moins hétérogène sur lequel l'essai est mis en place. Afin de contrôler cette hétérogénéité du terrain qui entraîne des imprécisions dans l'interprétation des résultats, de nombreuses méthodes ont été développées. Dès 1935, Fisher a tout d'abord émis les premiers fondements d'une planification des expériences permettant de tenir compte de l'hétérogénéité du terrain (Goumari, 1985). Il a ainsi introduit les dispositifs expérimentaux en blocs complets qui sont à la base de nombreux autres types de dispositifs (blocs incomplets, lattices, carrés latins, etc.). Les principes de répétition et de randomisation constituent la base de la construction et de l'utilisation de ces dispositifs. Des détails concernant les principes de construction et/ou les méthodes d'analyse des différents dispositifs permettant un contrôle de l'hétérogénéité expérimentale figurent notamment dans les ouvrages spécialisés de Dagnelie (2011, 2012), Federer (1955), Philippeau (1985) et Cochran et Cox (1950).

L'objectif recherché par l'utilisation des dispositifs expérimentaux, parfois très compliqués à matérialiser et à analyser, est le contrôle de l'hétérogénéité expérimentale qui influence directement les résultats des analyses lorsque cette hétérogénéité n'est pas bien maîtrisée. En effet, la plus grande partie des dispositifs expérimentaux, cités dans la littérature et utilisés dans le domaine agronomique, ne permettent en général qu'un simple, parfois double et rarement triple contrôle de l'hétérogénéité du terrain. Cette inaptitude des dispositifs classiques à bien contrôler l'hétérogénéité du terrain provient du fait que cette dernière ne se présente pas toujours sous forme de gradients réguliers mais plutôt sous forme de gradients irréguliers, le plus souvent discontinus (taches). Dans de telles situations, seules les méthodes d'analyse spatiale telles que les modèles des plus proches voisins (MPPV) initiés la première fois par Papadakis (1937), grâce à un système de lissage sur l'ensemble des parcelles du dispositif expérimental, peuvent rendre compte d'une façon satisfaisante du gradient du terrain (Sghaier, 2005). Il est à signaler également que diverses autres méthodes d'analyse permettant un contrôle local de la variabilité du terrain en expérimentation ont été développées, telles que les modèles autorégressifs (Cullis et Gleeson, 1991), la modélisation des résidus sous forme de processus ARIMA (Azais *et al.*, 1990), les notions de champ aléatoire (Zimmerman et Harville, 1991) et de géostatistique (Chilès et Delfiner, 1999; Goovaerts, 1997) et finalement la méthode de transformation en ondelettes (Lorent, 2002).

L'objectif de ce travail est d'illustrer par l'exemple, sur des données expérimentales empruntées de la publication de (Papadakis, 1937), les différentes étapes suivies pour l'ajustement des données à l'effet terrain à l'aide d'un modèle dit des plus proches voisins élaboré par Sghaier (2005) et de présenter les principales analyses de cette expérimentation avant et après ajustement des données observées à l'effet terrain.

2. Données expérimentales

Les données qui ont servi au présent travail sont tirées de la publication de Papadakis (1937). Il s'agit d'un essai de fertilisation sur pomme de terre réalisé en 1930 par le personnel de la célèbre

		X					
		1	2	3	4	5	6
Y	A			B			
	1	7 (344)	8 (439)	5 (459)	4 (393)	7 (366)	5 (385)
2	6 (388)	4 (360)	3 (392)	3 (437)	2 (444)	6 (434)	
3	2 (293)	9 (406)	1 (372)	8 (438)	9 (439)	1 (334)	
4	5 (339)	7 (332)	9 (479)	2 (385)	3 (338)	7 (297)	
5	1 (234)	2 (291)	8 (413)	5 (382)	9 (421)	8 (365)	
6	6 (297)	3 (279)	4 (295)	1 (262)	4 (335)	6 (367)	
		C			D		

FIGURE 1. Plan de l'essai et données récoltées (Papadakis, 1937)

station expérimentale Rothamsted dans un champ à North Fen Bourne (Lincs) en Angleterre. Le dispositif expérimental adopté est un dispositif en blocs aléatoires complets comportant 9 objets ou traitements et 4 blocs ou répétitions avec des parcelles élémentaires d'une superficie de 58 m² environ. La variable mesurée étant le rendement en tubercules par parcelle élémentaire. La figure 1 présente le plan du dispositif avec les références des traitements et les rendements observés. Dans cette figure, le premier chiffre en gras indique le numéro du traitement et le deuxième nombre (entre parenthèses) indique le rendement de la parcelle en livres ; les lettres A, B, C et D indiquent les blocs tandis que X et Y indiquent les coordonnées des différentes parcelles de l'essai. Ces dernières étant nécessaires pour le programme d'ajustement des données observées à l'effet terrain à l'aide du modèle des plus proches voisins utilisé. Il s'agit d'un programme informatique, rédigé en langage FORTRAN 77, qui permet, grâce un processus itératif, d'ajuster les données observées à l'effet terrain. Ce programme, développé par Sghaier (2005) dans le cadre d'une évaluation d'un essai multi-sites de comparaison de provenances de pin d'Alep (*Pinus halepensis* Mill.) en Tunisie, permet de contrôler l'hétérogénéité du terrain en utilisant différents modèles des plus proches voisins qui varient selon le nombre et la position des voisins les plus proches à considérer dans chaque modèle.

3. Contrôle de l'hétérogénéité spatiale du terrain

3.1. Modèle d'absence de contrôle de l'hétérogénéité spatiale

En considérant l'implantation d'un dispositif complètement aléatoire (CA), une donnée, par exemple le rendement d'un traitement i , situé dans une parcelle de coordonnées (l, c) d'un champ d'essai formé de p lignes ($l = 1, \dots, p$) et q colonnes ($c = 1, \dots, q$) peut être représenté selon le modèle théorique suivant :

$$Y_{i(l,c)} = m + a_i + X_{l,c} \quad (1)$$

avec :

$Y_{l,c}$: valeur individuelle du traitement i ($i = 1, \dots, s$) situé au point de coordonnées (l, c) ,

m : moyenne générale,

a_i : effet dû au traitement i ,

$X_{l,c}$: résidus indépendants, de moyenne nulle et de même écart-type σ .

3.2. Modèle de contrôle de l'hétérogénéité spatiale par blocs aléatoires complets

En considérant l'implantation d'un dispositif en blocs aléatoires complets (BAC), le modèle théorique d'analyse de la variance s'écrit de la façon suivante :

$$Y_{ij} = m + a_i + B_j + X_{ij} \quad (2)$$

avec :

Y_{ij} : valeur d'un individu du traitement i ($i = 1, \dots, s$) situé dans le bloc j ($j = 1, \dots, r$),

m : moyenne générale, a_i : effet dû au traitement i ,

B_j : effet dû au bloc j ,

X_{ij} : résidus indépendants, de moyenne nulle et de même écart-type σ .

Le facteur bloc, considéré comme aléatoire ($B_j \sim N(0, \sigma_B^2)$), et un facteur accessoire ou auxiliaire introduit en vue de maîtriser les variations résiduelles. Par hypothèse, toutes les parcelles d'un même bloc sont considérées comme identiques. Ainsi, les observations issues d'un même bloc sont corrélées et celles provenant de deux blocs différents sont indépendantes.

Selon Goumari (1985), sur la base des résultats de recensements réalisés sur un grand nombre d'essais (Cox, 1950; Khosla *et al.*, 1973; Claustrioux, 1977), 80 à 85% environ des essais sont réalisés selon des dispositifs en blocs aléatoires complets. L'autre proportion étant partagée entre le carré latin et les dispositifs incomplets. Cette grande proportion d'utilisation des dispositifs en blocs aléatoires complets en expérimentation agronomique est sans doute liée aux multiples avantages dont bénéficient ces dispositifs. En effet :

- les blocs aléatoires complets sont très simples à organiser, même plus simples à organiser que les dispositifs complètement aléatoires.
- les blocs aléatoires complets peuvent être réalisés pour n'importe quel nombre de traitements et n'importe quel nombre de répétitions. La seule restriction étant que le nombre de répétitions doit être, d'une façon générale, identique pour les différents traitements. Cette dernière restriction peut ne pas être respectée entièrement en attribuant à certains traitements (exemple témoin) deux ou plusieurs parcelles expérimentales par bloc, ce qui conduit à adopter pour ces traitements un nombre de répétitions qui est un multiple du nombre de répétitions des autres traitements.
- les blocs aléatoires complets gardent toutes leurs propriétés lorsque, au moment de l'analyse, un ou plusieurs traitements ou un ou plusieurs blocs doivent être éliminés.
- les blocs aléatoires complets sont en général, pour un même matériel expérimental, plus efficaces que les dispositifs complètement aléatoires.

Cependant, l'analyse de la variance des dispositifs en blocs est basée sur l'hypothèse que toutes les parcelles à l'intérieur d'un même bloc sont identiques. Le nombre de blocs ainsi que leurs dimensions et orientations sont déterminés a priori. Lorsque la dimension des blocs augmente, l'hypothèse d'homogénéité des parcelles à l'intérieur des blocs devient irréaliste. Même pour des blocs qui semblent être homogènes de l'extérieur, il existe le plus souvent une hétérogénéité en profondeur invisible, qui se manifeste après un certain temps surtout pour des essais de longue durée comme par exemple ceux relatifs au domaine forestier (Sghaier, 2003).

3.3. *Modèle de contrôle local de l'hétérogénéité spatiale par conditionnel autorégressif*

Quand le terrain est hétérogène, les quantités $X_{l,c}$ du modèle 1 renferment en plus des résidus considérés comme aléatoires, les erreurs dues aux variations locales du terrain. Ces variations constituent l'apport ou le gradient du terrain au point de coordonnées (l,c) . Le principe des modèles contrôlant localement ces variations, en particulier les modèles des plus proches voisins (MPPV), consiste donc à estimer cette contribution (positive ou négative) locale du terrain et de l'éliminer par la suite de la valeur résiduelle observée (Figure 3), de telle sorte que les traitements puissent être comparés comme s'ils avaient été mis en place sur un terrain rigoureusement homogène (Lorent, X). Il s'agit d'une variante de la célèbre méthode proposée par Papadakis (1937) dont l'objectif est d'augmenter la précision des résultats des expériences. A l'origine, cette méthode consistait à calculer pour chacune des parcelles un rendement ajusté ou corrigé par régression, en fonction notamment des rendements des parcelles voisines (Dagnelie, 1987). Afin d'affiner les estimations respectives des facteurs et des effets de terrain, Bartlett (1978) a proposé d'itérer la méthode de Papadakis jusqu'à convergence. Pichot (1993) estime que la convergence est considérée atteinte lorsque la différence maximale de la contribution du terrain estimée au cours de deux itérations successives k et $k-1$ est inférieure à 1/500 de la valeur maximale de la valeur estimée de cette même contribution à l'itération k .

Selon certains auteurs, la méthode des plus proches voisins permet dans certains cas de réduire la variance résiduelle jusqu'à 50% (Azais *et al.*, 1990). D'un autre côté, en testant différentes combinaisons des plus proches voisins pour corriger les données observées, Pearce (1976) a montré que l'efficacité des tests augmentait d'une façon remarquable surtout lorsqu'il y avait un gradient irrégulier du terrain (Aastveit, 1983; Pearce, 1976). Par ailleurs, en comparant l'efficacité de deux modèles des plus proches voisins par rapport aux deux analyses classiques des dispositifs en blocs aléatoires complets et des dispositifs en blocs aléatoires incomplets sur des données simulées, Baird et Mead (1991) ont montré qu'en présence d'hétérogénéité, les analyses des modèles des plus proches voisins sont plus efficaces que les analyses classiques et que la seule exception correspond au cas où cette hétérogénéité est totalement éliminée par la disposition des blocs. Ils ont encore montré que, lorsque l'analyse en blocs incomplets présentait un gain par rapport aux blocs complets, les modèles des plus proches voisins offrent un gain proportionnellement plus important. Finalement, en analysant la hauteur des arbres d'un essai de comparaison d'une trentaine de provenances de pin d'Alep (*Pinus halepensis* Mill.) après 30 ans de croissance sur terrain, Sghaier *et al.* (2004) ont montré que la réduction de la variance résiduelle, par rapport à une analyse sans aucun contrôle d'hétérogénéité, était de l'ordre de 30% pour une analyse en blocs aléatoires complets et varie pour les modèles des plus proches voisins de 66 à 79% selon le nombre et la position des voisins les plus proches pris en considération.

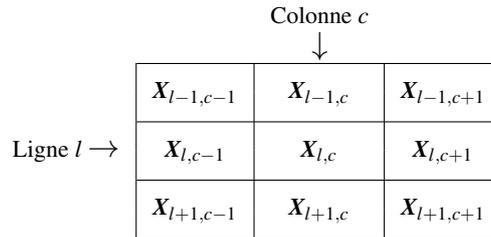


FIGURE 2. Coordonnées des voisins les plus proches utilisés pour l'estimation de la performance propre de chaque parcelle élémentaire.

4. Algorithme itératif par moindres carrés aux plus proches voisins

Le modèle d'analyse spatiale utilisé dans cette étude est un modèle autorégressif d'ordre 2 basé sur le principe d'estimation itérative de la performance propre de chaque parcelle en fonction de celles des 8 parcelles voisines (figure 2).

Les différentes étapes d'ajustement des données observées à l'effet terrain à l'aide du modèle des plus proches voisins sont :

1° A partir du modèle 1, on procède au calcul du résidu propre de chaque parcelle de la façon suivante, \hat{m}_i étant l'estimation de la moyenne de la variété i ou effet de la variété i :

$$X_{l,c} = Y_{i(l,c)} - \hat{m}_i \tag{3}$$

2° Pour chaque parcelle de coordonnées (l,c) , on détermine ensuite une quantité $\hat{X}_{l,c}$ par estimation à l'aide du modèle autorégressif, dit des plus proches voisins (équation 4), correspondant à une régression linéaire à deux variables explicatives passant par l'origine. Les deux variables explicatives considérées correspondent à la moyenne des résidus des quatre voisins et celle des quatre coins. La quantité $\hat{X}_{l,c}$ représente le gradient ou la performance propre du terrain estimé au point de coordonnées (l,c) .

$$X_{l,c} = b_1 w_1 + b_2 w_2 + \varepsilon_{l,c} \tag{4}$$

avec :

$$w_1 = \frac{(X_{l-1,c} + X_{l+1,c} + X_{l,c-1} + X_{l,c+1})}{4}$$

et

$$w_2 = \frac{(X_{l-1,c-1} + X_{l+1,c+1} + X_{l+1,c-1} + X_{l-1,c+1})}{4}$$

où w_1 représente la moyenne des résidus des 4 voisins et w_2 celle des 4 coins, b_1 et b_2 les deux coefficients de régression correspondants et $\varepsilon_{l,c}$ les erreurs de moyenne nulle et de même écart-type σ . Pour les parcelles de bordures comme celles ayant une ou plusieurs données manquantes à côté, le rendement de toute parcelle manquante sera remplacé par le rendement moyen de celles qui existent et qui entourent la parcelle dont on désire estimer le rendement.

3° On calcule alors les données ajustées à l'effet terrain $Y'_{i(l,c)}$ de la manière suivante :

$$Y'_{i(l,c)} = \hat{m}_i + (X_{l,c} - \hat{X}_{l,c}) \quad (5)$$

Après avoir retranché de chaque valeur observée $Y_{i(l,c)}$ le gradient parcellaire estimé correspondant $\hat{X}_{l,c}$, l'analyse de la variance est réalisée selon le modèle :

$$Y'_{i(l,c)} = m + a_i + (X_{l,c} - \hat{X}_{l,c}) \quad (6)$$

Les quantités $(X_{l,c} - \hat{X}_{l,c})$ sont des résidus dans lesquels il ne subsiste plus aucune part de variabilité à expliquer (Goumari, 1985). Ils doivent donc être aléatoires et de moyenne nulle.

4° Pour itérer la méthode, on remplace $Y_{i(l,c)}$ au 1° par $Y'_{i(l,c)}$. Au fil des itérations, on tendra vers un palier correspondant à la stabilisation de la variance résiduelle σ^2 , en éliminant le plus possible de la valeur des résidus observés $X_{l,c}$, la contribution locale estimée du terrain $\hat{X}_{l,c}$. La convergence est considérée atteinte lorsque la différence maximale de $X_{l,c}$ entre deux itérations successives $(k-1)$ et k est inférieure à $1/1000$ de la valeur de $X_{l,c}$ de l'itération k .

La somme des quantités $\hat{X}_{l,c}$, estimées par le modèle au cours des k itérations ($j = 1, \dots, n$), constitue le gradient ou l'effet total du terrain. Cette somme qui correspond également à la différence entre les valeurs observées $Y_{i(l,c)}$ et les valeurs ajustées $Y'_{i(l,c)}$ obtenues à la $k^{\text{ième}}$ et dernière itération peut être désignée par $T_{l,c}$ telle que :

$$T_{l,c} = \sum_{j=1}^k (\hat{X}_{l,c})_j = Y_{i(l,c)} - Y'_{i(l,c)} \quad (7)$$

La figure 3 présente l'organigramme ou l'algorithme de la procédure d'ajustement des données observées à l'effet terrain à l'aide du modèle des plus proches voisins utilisé.

Il est à signaler qu'en raison de la dépendance entre les variables, des degrés de liberté doivent être retranchés du facteur résiduel comme pour l'analyse de la covariance. Bartlett (1978), dans Kempton et Howes (1981), recommande pour obtenir une estimation conservatrice du carré moyen résiduel ajusté, de soustraire deux degrés de liberté dans le cas d'un modèle à une dimension avec un seul coefficient de régression. Pearce et Moore (1976) suggèrent également d'enlever trois degrés de liberté dans le cas d'un modèle avec deux coefficients de régression. Pearce (1998) propose quant à lui de retirer deux degrés de liberté par covariable. En l'absence de formulation mathématique exacte en la matière, Goumari (1990) a proposé de réduire le nombre de degrés de liberté d'autant d'unités qu'il y a de parcelles prises en considération dans le modèle des plus proches voisins utilisé. Concernant la présente étude et en se basant sur la proposition de Pearce (1998), nous avons réduit deux degrés de liberté par covariable du facteur résiduel, soit un total de quatre degrés de liberté selon le modèle autorégressif utilisé contenant deux coefficients de régression.

Finalement, si la méthode de Papadakis a fait couler beaucoup d'encre ces dernières années (Dagnelie, 1987, 1989; Gill, 1991; Goumari, 1985, 1990; Sghaier *et al.*, 2004; Sghaier, 2005), elle n'est pas tout à fait exempte de reproche sur le plan théorique (Bartlett, 1978; Wilkinson *et al.*, 1983). En effet, cette méthode se montre extrêmement proche du modèle utilisant comme

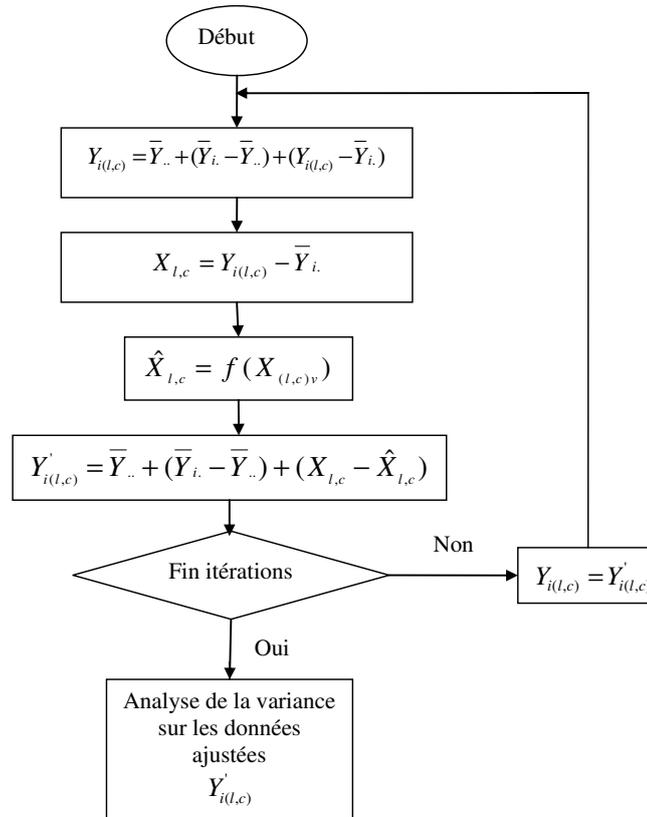


FIGURE 3. Organigramme d'ajustement des données observées à l'effet terrain à l'aide du modèle des plus proches voisins.

covariable des témoins répartis d'une façon systématique sur l'ensemble du dispositif, ce qui signifie que la covariable de Papadakis n'est pas indépendante de la variable à expliquer.

5. Critères d'évaluation et vérification des conditions d'application

Pour évaluer la capacité du modèle des proches voisins utilisé à réduire l'erreur expérimentale sur champ et vérifier les conditions d'application de l'analyse de la variance, des analyses numériques et graphiques basées essentiellement sur la valeur et la distribution des résidus ont été réalisées. Il s'agit des résidus issus de l'analyse de la variance à un critère et à deux critères de classification réalisée d'une part sur les données brutes et, d'autre part sur les données corrigées à l'effet terrain à l'aide du modèle des plus proches voisins développé. Les critères utilisés pour cette évaluation et vérification sont les suivants :

1° Test d'égalité des variances : l'égalité des variances avant et après ajustement des données à l'effet terrain a été vérifiée à l'aide du test de Bartlett (Dagnelie, 2013).

2° Test de normalité de la distribution des résidus : la normalité de la distribution des résidus a été vérifiée à l'aide du test de Ryan-Joiner (Ryan et Joiner, 1976) et la présentation graphique des

quantiles normaux (QQ-probability plots) : le test de normalité de Ryan-Joiner est une alternative simple au test de normalité de Shapiro-Wilk (de Smith, 2015) utilisé le plus souvent dans les logiciels statistiques. Le principe de ce test repose sur le calcul du coefficient de corrélation entre les résidus (e_i) classés par ordre croissant et les scores normaux ou quantiles normaux (z_i), s^2 étant la variance des ces résidus :

$$\rho_{obs} = \frac{\sum_{i=1}^n e_i z_i}{\sqrt{s^2(n-1) \sum_{i=1}^n z_i^2}} \quad (8)$$

Les scores normaux ou quantiles normaux (z_i) sont calculés comme suit :

$$z_i = \Phi^{-1} \left(\frac{i - 3/8}{n + 1/4} \right) \quad (9)$$

où $\Phi^{-1}(z)$ correspond à l'inverse de la fonction de répartition cumulative de la distribution normale réduite.

Une valeur de ρ_{obs} proche de l'unité indique que la distribution des résidus peut être considérée comme normale. La normalité des résidus doit être rejetée au niveau de confiance $(1 - \alpha)$ lorsque $\rho_{obs} < \rho_\alpha$. Les valeurs ρ_α étant reprises dans une table qui est fonction du nombre d'observations n (Looney et Gullledge, 1985). Une représentation graphique des z_i et des e_i , permettant de juger le caractère linéaire de la relation, constitue une première approche de ce test.

3° Test d'indépendance des résidus : le test du quotient de Von Neumann (Morice et Chartier, 1954) a pour but de vérifier l'absence de corrélation ou l'indépendance entre les résidus successifs (autocorrélation d'ordre 1) lorsque ceux-ci sont classés selon un ordre logique (rendement dans ce cas). Ce test, basé sur le caractère asymptotiquement normal du quotient :

$$q = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{2 \sum_{i=1}^n e_i^2} \cong 1 - \hat{\rho}_1 \quad (10)$$

avec

$$\hat{\rho}_1 = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2}$$

consiste à comparer la valeur :

$$z_{obs} = \frac{|q - 1|}{\sqrt{\frac{n-2}{n^2-1}}} \quad (11)$$

à la valeur théorique $z_{1-\alpha/2}$ et de rejeter le caractère aléatoire (indépendance) des résidus lorsque $z_{obs} \geq z_{1-\alpha/2}$, une valeur de q inférieure à l'unité correspond à une corrélation positive et une valeur supérieure à l'unité à une corrélation négative entre les résidus consécutifs.

4° Test du caractère aléatoire de la répartition spatiale des valeurs et des signes des résidus sur la parcelle d'essai : Deux tests ont été envisagés.

(i) Valeurs des résidus (Indice de Moran) : L'indice de Moran (1950), remanié par Cliff et Ord (1969, 1973), basé sur le calcul de l'autocorrélation spatiale, consiste à vérifier s'il existe une relation entre la proximité spatiale des parcelles élémentaires et leur degré de ressemblance. Une autocorrélation spatiale significative indique l'absence d'une répartition spatiale aléatoire des résidus sur la parcelle d'essai. Il s'agit de mesurer l'autocorrélation spatiale à l'aide d'un indice qui est le rapport entre la covariance des régions contiguës et la variance totale :

$$I_{Moran} = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (12)$$

où i unité spatiale ou parcelle élémentaire ; j voisins des entités i ; n = nombre d'unités spatiales ou nombre total de parcelles de l'essai ; x_i est la valeur de la variable (gradient du terrain) dans l'unité i ; \bar{x} est la moyenne de x et w_{ij} sont les éléments de la matrice d'interactions spatiales (matrice carré de n lignes et n colonnes), définie sous la forme de la contiguïté, dont les éléments prennent, par exemple, la valeur 1 pour les (i,j) voisins et 0 autrement.

Selon la valeur obtenue de cet indice, l'autocorrélation spatiale peut être interprétée comme suit :

Si $I_{Moran} > 0$: l'autocorrélation spatiale est positive, ce qui signifie que deux parcelles proches se ressemblent plus que deux parcelles éloignées.

Si $I_{Moran} < 0$: l'autocorrélation spatiale est négative, ce qui signifie que deux parcelles proches se ressemblent moins que deux parcelles éloignées.

Si $I_{Moran} = 0$: l'autocorrélation spatiale est nulle, ce qui signifie que deux parcelles proches se ressemblent ni plus ni moins que deux parcelles éloignées.

Pour le test d'hypothèse statistique nulle (pas d'autocorrélation spatiale), l'indice I de Moran peut être transformé en z-scores (les valeurs critiques) et p-value (la significativité). L'espérance mathématique de l'indice de Moran (hypothèse de non autocorrélation spatiale) est (Sawada, 2004) :

$$E(I_{Moran}) = -\frac{1}{n-1} \quad (13)$$

et la variance est égale à :

$$V(I_{Moran}) = E(I_{Moran}^2) - (E(I_{Moran}))^2 \quad (14)$$

Selon la méthode de l'approximation normale, la variance de l'indice de Moran pour une variable surfacique continue se calcule de la façon suivante (Thomas-Agnan, 2012) :

$$V(I_{Moran}) = \left(\frac{1}{S_0^2(n^2 - 1)} (n^2 S_1 - n S_2 + 3 S_0^2) \right) - (E(I_{Moran}))^2 \quad (15)$$

avec :

$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$, la somme des valeurs de la matrice de contiguïté.

$S_1 = \frac{\sum_{i=1}^n \sum_{j=1}^n (w_{ij} - w_{ji})^2}{2}$, si la matrice de contiguïté est symétrique on a : $S_1 = 2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}$

$S_2 = \sum_{i,j=1}^n (w_i + w_j)^2$, si la matrice de contiguïté est symétrique on a : $S_2 = 4 \sum_{i,j=1}^n w_{ij}^2$.

Ainsi, la valeur critique z-score est égale à :

$$z_{obs} = \frac{|I_{Moran} - E(I_{Moran})|}{\sqrt{V(I_{Moran})}} \quad (16)$$

L'hypothèse nulle (absence d'autocorrélation) doit être rejetée au niveau de signification α lorsque :

$$z_{obs} \geq z_{1-\alpha/2} \quad (17)$$

L'indice de Moran a été calculé dans notre cas d'une part sur les résidus obtenus à partir des données brutes et ajustées issus des analyses de la variance à un critère (Figure A2) et à deux critères (Figure A3) de classification et, d'autre part sur le gradient ou l'effet total du terrain $T_{l,c}$ estimé pour chaque parcelle élémentaire à l'aide de l'équation 7 (figure 7).

(ii) Signes des résidus (Test des suites) : tester le caractère aléatoire de la répartition des signes des résidus sur la parcelle d'essai.

Pour tester si la répartition des signes des résidus sur le plan de l'essai peut être considérée comme une répartition aléatoire ou non, nous avons eu recours au test de suites (Palm, 1986, 1987). Il s'agit de vérifier si le nombre k de suites de résidus de même signe est compatible avec l'hypothèse d'une répartition aléatoire des signes des résidus, compte tenu de fait que les nombres de résidus positifs et négatifs sont théoriquement identiques en moyenne. Dans le cas de séries comportant un nombre suffisant d'observations ($n > 20$), la distribution de k est approximativement normale. On calcule la valeur :

$$z_{obs} = \frac{|k - \frac{n+2}{2}|}{\sqrt{\frac{n-1}{4}}} \quad (18)$$

et on rejette le caractère aléatoire de la succession des signes plus et moins si z_{obs} est supérieur ou égal à $z_{1-\alpha/2}$ (Morice et Chartier, 1954).

6. Ajustement des données à l'effet terrain et choix du nombre d'itérations à retenir

Pour déterminer le nombre d'itérations nécessaires à retenir pour l'ajustement des données observées à l'effet terrain et de vérifier s'il reste justifié de tenir compte de la structure des blocs pour l'analyse de ces données ajustées, une analyse de la variance à un critère (essai conduit selon un dispositif complètement aléatoire) et à deux critères (essai conduit selon un dispositif en blocs aléatoires complets) de classification a été effectuée d'une part sur les données brutes et, d'autre part sur les données ajustées issues des différentes itérations tout en suivant :

- l'évolution des carrés moyens résiduels et des valeurs du test F de Snédécour : pour l'analyse des données ajustées à l'effet terrain à l'aide du modèle des plus proches voisins utilisé, 4 degrés de liberté ont été retranchés du facteur résiduel.

- l'évolution du carré moyen relatif au facteur blocs pour l'analyse de la variance à deux critères de classification.
- l'évolution de la corrélation entre résidus voisins : la corrélation entre les résidus des différentes parcelles élémentaires et la moyenne des résidus de toutes les parcelles voisines correspondantes.

Les lattices de la figure A1 (Annexes) présentent les données ajustées à l'effet terrain obtenus pour les cinq premières et la treizième itération où la convergence totale est atteinte. Cette figure montre comment évoluent les données ajustées, d'une itération à une autre, vers la stabilité. Pour cet essai de petite taille et par comparaison aux valeurs ajustées obtenues à la treizième itération, la convergence des données ajustées vers la stabilité semble être atteinte vers la cinquième itération.

Le tableau 1 présente les résultats des analyses de la variance réalisées selon un dispositif complètement aléatoire (un critère de classification) et un dispositif en blocs aléatoires complets (deux critères de classification), d'une part sur les données brutes et, d'autre part sur les données ajustées à l'effet terrain obtenues au cours des différentes itérations effectuées. Pour le calcul des carrés moyens résiduels ajustés (analyse des données ajustées à l'effet terrain), les nombres de degrés de liberté du facteur résiduel ont été réduits de 4 unités à raison de deux degrés de liberté par coefficient de régression du modèle autorégressif utilisé pour corriger les données en tenant compte de la fertilité locale du terrain. L'égalité des variances des différents traitements avant et après ajustement des données à l'effet terrain obtenues à la 5^{ème} itération étant vérifiée préalablement à l'aide du test Chi2 de Bartlett (Dagnelie, 2013). Les valeurs obtenues sont $\chi^2_{obs} = 4,898$ ($p = 0,678$) pour les données brutes et $\chi^2_{obs} = 7,513$ ($p = 0,482$) pour les données ajustées.

TABLE 1. Résultats des analyses de la variance à un et à deux critères de classification, réalisées sur les données brutes et les données ajustées des différentes itérations.

Itér.	Analyse selon un dispositif complètement aléatoire			Analyse selon un dispositif en blocs aléatoires complets			
	CM _a	CM _r	F _{obs}	CM _a	CM _b	CM _r	F _{obs}
0	6859,632	2751,019	2,49	6859,632	11044,630	1714,317	4,00
1	4946,132	730,969	6,77	4946,132	10,809	838,993	5,90
2	5454,412	651,463	8,37	5454,412	68,572	738,896	7,38
3	5289,573	636,181	8,31	5289,573	15,896	729,223	7,25
4	5352,085	633,431	8,45	5352,085	29,904	723,960	7,39
5	5321,881	632,805	8,41	5321,881	22,082	724,414	7,35
...
13	5330,675	632,631	8,43	5330,675	24,204	723,895	7,36

CM_a : Carré moyen factoriel; CM_b : Carré moyen blocs, CM_r : Carré moyen résiduel, F_{obs} : Valeur observée du test F de Snédécour.

D'après les résultats du tableau 1, nous constatons ce qui suit :

- la réduction très importante du carré moyen résiduel dès la première itération et sa stabilisation environ vers la cinquième itération et ceci pour les deux types d'analyse de la variance, sans tenir compte et en tenant compte de la structure des blocs,
- par comparaison à une analyse de la variance des données brutes selon un modèle complètement aléatoire, l'analyse des mêmes données brutes à l'aide d'un modèle en blocs aléatoires complets a permis une réduction du carré moyen résiduel d'environ 38%,
- en prenant toujours l'analyse de la variance des données brutes selon un modèle complètement aléatoire comme source de comparaison, l'analyse de la variance réalisée sur les données ajustées, obtenues à la cinquième itération, a permis des réductions du carré moyen résiduel de l'ordre de 77% pour le modèle complètement aléatoire et de l'ordre de 63% pour le modèle en blocs aléatoires complets,
- pour les différentes itérations, le carré moyen résiduel issu de l'analyse de la variance des données ajustées à l'effet terrain selon un modèle complètement aléatoire présente des valeurs légèrement inférieures à celles du carré moyen résiduel issu de l'analyse de la variance des mêmes données selon un modèle en blocs aléatoires complets,
- dans le cas d'une analyse de la variance réalisée selon un modèle en blocs aléatoires complets sur les données ajustées à l'effet terrain, la réduction du carré moyen relatif au facteur blocs est spectaculaire dès la première itération (de l'ordre de 99%) et ceci par rapport au même carré moyen résiduel obtenu à partir des données brutes.

Il ressort d'après ces constatations, qu'après ajustement des données à l'effet terrain par l'intermédiaire du modèle autorégressif des plus proches voisins, la prise en compte de la structure des blocs ne se justifie plus. Au contraire, le fait de tenir compte des blocs dans notre exemple a augmenté légèrement l'erreur expérimentale. Des degrés de liberté sont affectés au facteur blocs pour rien, ce qui diminue la puissance du test relatif au facteur étudié.

Le tableau A1 (Annexes) présente les résultats de la comparaison multiple des moyennes réalisée par la méthode de Newman et Keuls (Dagnelie, 2011), d'une part sur les données brutes et, d'autre part sur les données ajustées issues des itérations 1, 2, 3, 4, 5 et 13. Nous remarquons d'après ce tableau que les rangs occupés par les traitements (objets) se stabilisent dès la première itération, par contre les moyennes de ces derniers ne deviennent stables qu'à partir de la quatrième ou la cinquième itération. En plus du changement dans les rangs de quelques traitements (5 et 6 ; 2 et 3), le nombre et le contenu des groupes homogènes de traitements ont également changé après ajustement des rendements des différents traitements testés à l'effet terrain.

La figure 4 présente l'évolution en fonction du nombre d'itérations, de la corrélation entre les résidus des différentes parcelles de l'essai et la moyenne des résidus des parcelles voisines. Les résidus sont issus d'une analyse de la variance à un critère de classification sans tenir compte du facteur bloc. La corrélation entre résidus voisins qui indique l'importance de l'hétérogénéité du terrain non contrôlée par le modèle d'analyse de la variance adopté et qui était de l'ordre de 0,80 pour les données brutes, baisse après ajustement des données d'une façon très rapide dès la première itération à environ 0,20 pour devenir stable et négligeable vers la cinquième itération.

En se basant sur l'évolution des carrés moyens résiduels en fonction du nombre d'itérations, des valeurs du test F observé relatif au facteur traitement (Tableau 1), de la comparaison multiple des moyennes des traitements (Tableau A1) et de la corrélation entre résidus voisins (Figure 4),

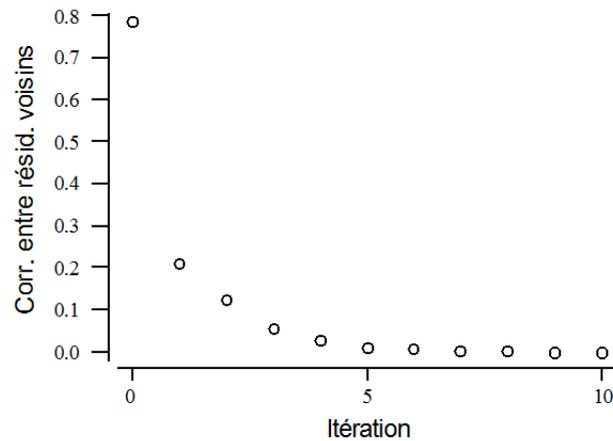


FIGURE 4. Coefficients de corrélation (en valeurs absolues) entre les résidus des différentes parcelles et la moyenne des résidus des parcelles voisines pour les 10 premières itérations.

les données ajustées après 5 itérations ont été donc retenues pour démontrer l'intérêt du modèle autorégressif des plus proches voisins utilisé à réduire l'erreur expérimentale et améliorer les conditions d'application de l'analyse de la variance en expérimentation agronomique sur champ.

7. Analyse des résidus et vérification des conditions d'application

Valeurs et signes des résidus :

Les figures A2 et A3 présentent la répartition sur le plan de l'essai des valeurs et des signes des résidus observés issus d'une analyse de la variance à un critère (CA) et à deux critères (BAC) de classification réalisée sur les données brutes (itération 0) et les données ajustées à l'effet terrain (5^{ème} itération).

Normalité des résidus :

La figure 5 qui présente les quantiles normaux (QQ-probability plots) et les résultats relatifs au test de normalité de Ryan-Joiner montre que la normalité de la distribution des résidus issus des analyses de la variance à un critère et à deux critères de classification réalisées d'une part sur les données brutes (Figure 5a) et, d'autre part sur les données ajustées à l'effet terrain obtenues à la cinquième itération (Figure 5b) peut être acceptée.

Indépendance des résidus :

Le tableau 2 présente les résultats du test du quotient de Von Neumann relatifs à l'indépendance des résidus. Il s'agit des résidus issus des analyses de la variance à un (Figure A2) et à deux critères de classification (Figure A3) réalisées d'une part sur les données brutes (itération 0) et, d'autre part sur les données ajustées à l'effet terrain (itération 5) et classés selon les rendements

observés des différentes parcelles de l'essai. D'après ce tableau, l'hypothèse d'indépendance des résidus est rejetée uniquement pour les résidus issus d'une analyse de la variance à un critère de classification (CA) réalisée sur les données brutes. Une fois les données sont ajustées à l'effet terrain par le modèle autorégressif utilisé, ces mêmes résidus deviennent indépendants.

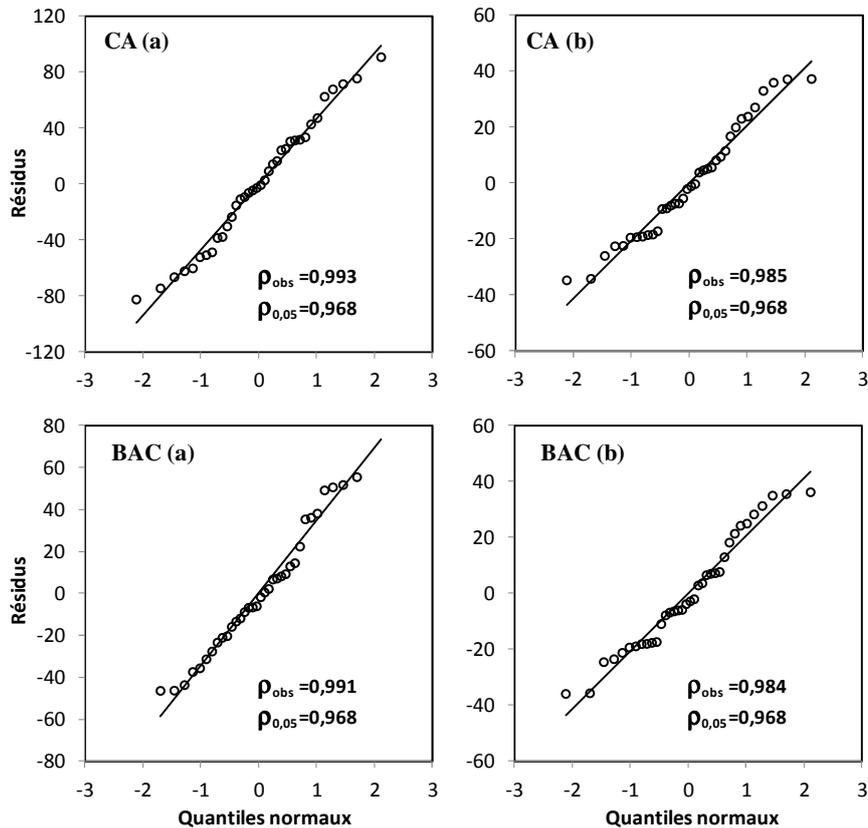


FIGURE 5. *QQ-probability plots et test de normalité de Ryan-Joiner relatifs à la distribution des résidus issus de l'analyse de la variance à un critère (CA) et à deux critères (BAC) de classification réalisée sur les données brutes (a) et les données ajustées (b) obtenues à la 5^{ème} itération à l'aide du modèle des plus proches voisins utilisé.*

Caractère aléatoire de la répartition spatiale des résidus sur le terrain :

1° Valeurs des résidus (Indice de Moran) :

Le tableau 3 présente les résultats de l'indice de Moran calculé sur les données des figures A2 et A3 relatives aux résidus issus des analyses de la variance à un critère (CA) et à deux critères de classification (BAC) réalisées d'une part sur les données brutes et, d'autre part sur les données ajustées à l'effet terrain obtenues à la 5^{ème} itération. Comme pour l'indépendance des résidus, l'absence d'autocorrélation spatiale entre les résidus peut être rejetée uniquement dans le cas de l'analyse des données brutes en considérant un dispositif complètement aléatoire. Ce

TABLE 2. Résultats du test du quotient de Von Neumann relatifs à l'Indépendance des résidus.

Dispositif	Données analysées	Nombre d'itérations	Z_{obs}	Probabilité (α)
CA	Brutes	0	3,29	0,0010
	Ajustées	5	1,47	0,1413
BAC	Brutes	0	1,51	0,1317
	Ajustées	5	1,27	0,2025

TABLE 3. Résultats du Test d'autocorrélation spatiale des résidus.

Dispositif	Données analysées	Nombre d'itérations	Indice de Moran	Z_{obs}	Probabilité (α)
CA	Brutes	0	0,5465	6,76	<0,0001
	Ajustées	5	0,0054	0,40	0,6891
BAC	Brutes	0	0,1106	1,64	0,1010
	Ajustées	5	-0,0106	0,21	0,8337

résultat prouve qu'il existe une hétérogénéité du terrain non contrôlé par le modèle d'analyse de la variance utilisé et que cette hétérogénéité s'ajoute aux résidus, supposés aléatoires et indépendants. Une fois les données ajustées, les résidus obtenus ne présentent plus d'autocorrélation spatiale. Il s'agit bien des résidus aléatoires dans lesquels il ne subsiste plus aucune part de variabilité liée à l'hétérogénéité du terrain.

2° Signes des résidus (Test des suites) :

Pour tester le caractère aléatoire de la répartition des signes des résidus sur le terrain de notre essai avant et après ajustement des données à l'aide du modèle des proches voisins (parties droites des figures A2 et A3), nous avons transformé les parcelles de l'essai en séquence linéaire de parcelles, une fois dans le sens horizontal de gauche à droite et de droite à gauche, et une fois dans le sens vertical du haut vers le bas et de bas vers le haut comme indiqué dans le schéma de la figure 6. Les résultats de cette transformation et du test préconisé figurent au tableau 4.

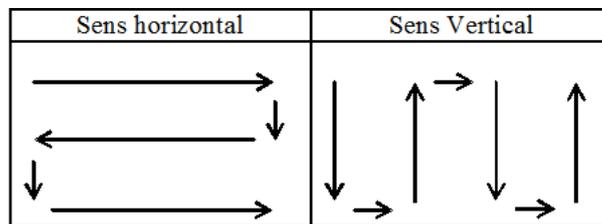


FIGURE 6. Sens suivi pour transformer les parcelles de l'essai en une séquence linéaire de parcelles.

Par comparaison à la valeur critique de $z_{1-\alpha/2}$ qui est égale à 1,96 au seuil de probabilité de 5%, les valeurs de z_{obs} présentées dans le tableau 4 montrent que, seule la répartition des résidus

TABLE 4. Test du caractère aléatoire de la succession des signes des résidus issus des analyses de la variance à un critère (CA) et à deux critères (BAC) de classification, réalisées d'une part sur les données brutes et, d'autre part sur les données ajustées (5^{ème} itération).

Dispositif	Sens suivi	Nombre d'itérations	Nombre de suites	Z_{obs}	Probabilité (α)
CA	Horizontal	0	8	3,72	0,0002
	Vertical	0	8	3,72	0,0002
	Horizontal	5	17	0,68	0,4990
	Vertical	5	19	0,00	1,0000
BAC	Horizontal	0	15	1,35	0,1763
	Vertical	0	18	0,34	0,7353
	Horizontal	5	16	1,01	0,3105
	Vertical	5	19	0,00	1,0000

issus de l'analyse de la variance réalisée sur les données brutes selon un modèle complètement aléatoire ne peut pas être considérée comme aléatoire et ceci pour les deux directions, horizontale et verticale. Ce résultat est attendu, puisqu'en regardant la répartition spatiale des signes des résidus de la figure A2 (cas de l'analyse de la variance des données brutes selon un dispositif complètement aléatoire), nous constatons que la parcelle de l'essai est divisée en deux parties selon les signes des résidus, une partie supérieure avec des signes positifs et une partie inférieure avec des signes négatifs. La prise en compte de la structure des blocs pour l'analyse des données brutes (Figure A3) comme l'utilisation du modèle des proches voisins pour corriger les données observées à l'effet terrain et leur analyse par la suite selon un modèle complètement aléatoire (Figure A2), donne un caractère aléatoire à la répartition des résidus sur les différentes parcelles de l'essai.

Le fait que la prise en compte de la structure des blocs rend la répartition sur le terrain des résidus issus des données brutes aléatoire, montre que la parcelle d'essai et la disposition des blocs sur le terrain sont bien choisies (Figure A3). Il s'agit bien en fait d'une station de recherche dont le terrain expérimental est bien étudié. L'analyse des données ajustées à l'effet terrain selon un modèle en blocs aléatoires complets, améliore d'avantage le caractère aléatoire de la répartition des résidus sur le terrain (Figure A3 et Tableau 4). Toutefois, la prise en compte de la structure des blocs dans ce dernier cas n'apporte pas de contribution significative quant à la réduction du carré moyen résiduel (Tableau 1).

8. Cartographie de la fertilité locale du terrain et test d'autocorrélation spatiale

La figure 7 présente la projection sur le plan de l'essai de la contribution du terrain aux rendements observés, correspondant à la fertilité locale ou la performance du terrain au niveau de chaque parcelle élémentaire ($T_{l,c}$) estimé par le modèle autorégressif utilisé. Il s'agit des différences entre les rendements observés et les rendements ajustés obtenus à la cinquième itération.

La figure 8, obtenue à l'aide de la procédure « Contour Plot » du logiciel MINITAB, schématise cette fertilité sur le même plan de l'essai.

Les valeurs de la figure 7 et le schéma de la figure 8 sont en concordance avec la figure A2 correspondant à la répartition des signes des résidus issus d'une analyse de la variance à un

		X					
Y		1	2	3	4	5	6
1		7 (19,00)	8 (28,09)	5 (36,05)	4 (47,86)	7 (37,64)	5 (33,35)
2		6 (-10,26)	4 (10,73)	3 (58,10)	3 (51,75)	2 (48,59)	6 (45,01)
3		2 (-30,44)	9 (0,07)	1 (17,61)	8 (40,91)	9 (27,71)	1 (6,11)
4		5 (-38,81)	7 (-25,53)	9 (27,50)	2 (6,91)	3 (-6,95)	7 (-14,00)
5		1 (-65,18)	2 (-44,76)	8 (-17,11)	5 (-8,92)	9 (-24,21)	8 (-22,23)
6		6 (-74,54)	3 (-66,00)	4 (-41,45)	1 (-30,48)	4 (-16,05)	6 (-34,61)

FIGURE 7. Valeurs de la fertilité locale du terrain estimée à l'aide du modèle des plus proches voisins utilisé.

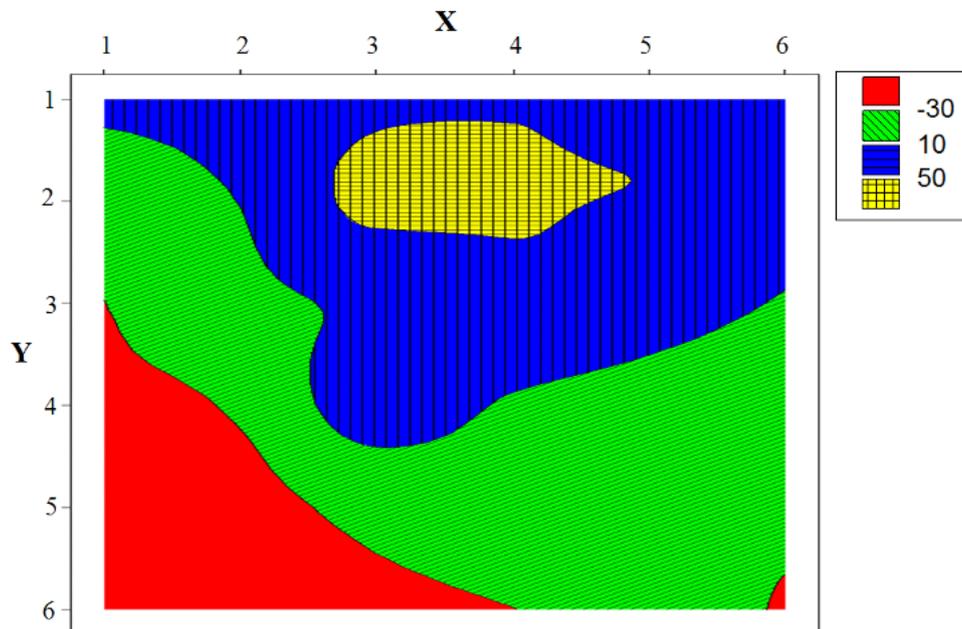


FIGURE 8. Cartographie de la fertilité locale du terrain estimée à l'aide du modèle des plus proches voisins utilisé.

critère de classification (CA) réalisée sur les données brutes. Cette concordance signifie que dans cette dernière analyse, les résidus aléatoires, supposés indépendants, de distribution normale, de moyenne nulle et de même écart-type σ , renferment en plus une composante terrain non aléatoire et dont le modèle d'analyse adopté ne peut pas en tenir compte. Le fait d'ajuster les données à l'effet terrain par le modèle adéquat, cette composante terrain non contrôlée disparaît et le même modèle d'analyse de la variance, en plus de la réduction remarquable de l'erreur expérimentale, donne des résidus dans lesquelles il ne reste plus rien à expliquer et dont la répartition sur la parcelle de l'essai devienne parfaitement aléatoire (Figure A2 : répartition spatiale des signes des résidus des données ajustées). La présence de cette hétérogénéité spatiale de la fertilité du sol sur la parcelle d'essai (Figure 7) a été confirmée par le test d'autocorrélation spatiale de Moran. La valeur de l'indice de Moran calculée à partir des données de la figure 6 est de l'ordre de 0,696 avec une variance de l'ordre de 0,007244 ce qui donne une valeur transformée (z-scores) du même indice de l'ordre de 8,51. Cette dernière valeur qui dépasse largement la valeur théorique 1,96 correspondant à la valeur critique $z_{1-\alpha/2}$ de la distribution normale réduite au seuil de confiance de 95%, permet de rejeter largement l'hypothèse d'absence d'autocorrélation spatiale.

9. Conclusion

Les données analysées dans ce travail proviennent d'un essai de petite dimension, conduit dans une station expérimentale en Angleterre selon un dispositif en blocs aléatoires complets. Malgré que l'hétérogénéité du terrain dans lequel l'essai a été installé est supposée bien connue, la disposition des traitements en blocs n'a permis de réduire l'erreur expérimentale que de 38% par comparaison à une disposition complètement aléatoire des mêmes traitements. Après ajustement des données à l'effet terrain à l'aide du modèle des plus proches voisins utilisé, l'analyse de ces dernières en considérant un dispositif complètement aléatoire a permis de réduire l'erreur expérimentale de 77% par rapport à une analyse des données brutes selon le même dispositif et de 63% par rapport à une analyse des mêmes données brutes selon un dispositif en blocs aléatoires complets. La prise en compte de la structure des blocs lors de l'analyse des données ajustées à l'effet terrain n'a pas amélioré la réduction de l'erreur expérimentale par comparaison à une analyse des mêmes données ajustées selon un dispositif complètement aléatoire. Le carré moyen du facteur blocs dans ce dernier cas a enregistré une réduction de 99,8% par comparaison au même carré moyen d'une analyse réalisée sur les données brutes. Ce dernier résultat montre qu'il n'est plus justifié de tenir compte de la structure des blocs dans l'analyse des données expérimentales une fois ces dernières sont corrigées à l'effet terrain à l'aide d'un modèle des plus proches voisins adéquat. Cela ne veut pas dire pour autant qu'on doit abandonner la répartition des traitements selon une structure en blocs, car un plan en blocs peut être utilisé rien que pour un besoin d'organisation.

Grâce à leur efficacité démontrée dans le contrôle de l'hétérogénéité expérimentale, les modèles des plus proches voisins peuvent avoir un grand rôle à jouer dans l'analyse des résultats des essais agronomiques sur champ. En effet, pour un grand nombre d'essais surtout de grandes dimensions, installés le plus souvent sur des terrains accidentés tels que les terrains forestiers, les analyses des données selon les dispositifs expérimentaux classiques sont peut efficaces voir inappropriées à cause de l'hétérogénéité du terrain qui peut varier même à l'échelle de la parcelle élémentaire. D'un autre côté et pour des essais de longues durées tels que les essais relatifs aux espèces arboricoles ou forestières, même si le terrain sur lequel ces essais sont installés semble

être homogène à la surface, une hétérogénéité de profondeur peut exister. Cette hétérogénéité de profondeur non visible au moment de la mise en place des essais et dont l'effet se manifeste de plus en plus avec l'âge de développement des arbres, ne sera jamais contrôlée par le modèle de départ qui est conçu sur la base de l'hétérogénéité superficielle du terrain. Selon les résultats d'une étude sur le comportement des provenances de Pin d'Alep (*Pinus halepensis* Mill.) après une quarantaine d'années de croissance en sites de reboisement, les zones de fertilité sur le terrain varient selon les stades de développement des arbres et ne se stabilisent qu'à un certain âge (Sghaier, 2005). Les zones les plus fertiles au jeune âge ne sont pas certainement les mêmes qu'à un âge avancé. Finalement, les modèles des plus proches voisins sont faciles à utiliser quel que soit le plan d'expérience adopté. En plus de la souplesse qu'ils présentent quant au choix des proches voisins à utiliser pour l'estimation de la fertilité locale du terrain, ils ne présupposent aucune structure expérimentale particulière, ce qui offre moins de contraintes pour l'expérimentateur lors de l'implantation de ses essais sur le terrain.

Références

- Aastveit, H. (1983). On the effect of correlation between plots in randomized block experiments. *Biom. J.*, 25, No. 2, 129-153.
- Azaïs, J.M., Denis, J. B., Dhorne, T., and Kobilinsky, A. (1990). Neighbour analysis of experiments : a review of the different approaches. *Biom. Parxim.*, 30, 15-39.
- Baird, D., and Mead, R. (1991). The empirical efficiency and validity of two neighbour models. *Biometrics*, 47, 1473-1487.
- Bartlett, M.S. (1978). Nearest neighbour models in the analysis of Field experiments. *J. R. Statist. Soc. B*, 40, No. 2, 147-174.
- Chilès, J.P., and Delfiner, P. (1999). *Geostatistics : modeling spatial uncertainty*. New York, Wiley, 695 p.
- Claustrioux, J.J. (1977). *Etude de la robustesse de l'analyse des résultats de dispositifs expérimentaux à répartition non conforme des objets*. Gembloux(Belgique), Facultés des Sciences agronomiques, 171 p.
- Cliff, A.D., and Ord, K.J. (1969). The Problem of Spatial Autocorrelation. (in Scott, A.J.,(ed.)), *Papers in Regional Science, Volume 1, Studies in Regional Science*, Pion, Londres, pp. 25-55.
- Cliff, A.D., and Ord, K.J. (1973). *Spatial autocorrelation*. London, Pion.
- Cochran, W.G., and Cox, G.M. (1950). *Experimental designs*. New York, Wiley, 454 p.
- Cox, G.M. (1950). A survey of types of experimental designs. *Biometrics* 6, 301-302.
- Cullis, B.R., and Gleeson, A.C. (1991). Spatial analysis of field experiments : an extension to two dimensions. *Biometrics* 47 (4), 1449-1460.
- Dagnelie, P. (1987). La méthode de Papadakis en expérimentation agronomique : considérations historiques et bibliographiques. *Biom. Praxi.*, 27, 49-64.
- Dagnelie, P. (1989). The method of Papadakis in agricultural experimentation : an overview. *Biuletyn Oceny Odmian. Zrzeszt* 21-22, 111-121.
- Dagnelie, P. (2011). *Statistique théorique et appliquée*. Tome 2. *Inférence statistique à une et à deux dimensions*. Bruxelles, De Boeck, 736 p.
- Dagnelie, P. (2012). *Principes d'expérimentation : planification des expériences et analyse de leurs résultats*. Gembloux, Presses agronomiques, 413 p.
- Dagnelie, P. (2013). *Statistique théorique et appliquée*. Tome 1. *Statistique descriptive et bases de l'inférence statistique*. Bruxelles, De Boeck, 517 p.
- de Smith, M.J. (2015). *Statistical analysis handbook : A comprehensive handbook of statistical concepts, techniques and software tools*. <http://www.Statsref.com/HTML/index.html?ryan-joiner.html>
- Federer, W.T. (1955). *Experimental design : theory and application*. New York, Macmillan, 544 + 47 p.
- Gill, P.S. (1991). A bibliography of nearest neighbour methods in design and analysis of experiments. *Biom. J.* 33(4), 455-459.
- Goumari, A. (1985). *Contribution à l'étude et à la maîtrise de l'hétérogénéité des expériences en champ*. Thèse de doctorat. Faculté Universitaire des Sciences agronomiques de Gembloux., 150 p.

- Goumari, A. (1990). Analyse comparative des résultats d'essais en champ selon les techniques des blocs aléatoires complets, des lattices et des plus proches voisins. *Biom. Praxim.*, 30, 91-105.
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York, Oxford University Press, 483 p.
- Kempton, R.A., and Howes, C.W. (1981). The use of neighbouring plot values in the analysis of variety trials. *Appl. Statist.* 30(1), 59-70.
- Khosla, R.K., Sardana, M.G., Saksena, M.P. and Sahani, M.L. (1973). A review of agricultural experimentation in India. *J. Indian Soc. Agric. Statist.* 25(1), 136-139.
- Lorent, F. (2002). Utilisation de la transformation en ondelettes pour le contrôle de la variabilité du terrain en expérimentation : Comparaison avec l'analyse en blocs et la méthode de Papadakis. *Actes des XXXIV^{es} Journées de Statistique, Bruxelles et Louvain-la-Neuve*, 13-17 mai, 283-284.
- Lorent, F. (X). Méthode Papadakis : SIMUWAVE. <http://francois.lorent.free.fr/index.php?affiche=Book&chapter=conclusions-fr>
- Looney, S.W., and Gullidge, T.R. (1985). Use of the Correlation Coefficient with Normal Probability Plots. *The American Statistician*, 39, 1,76-79.
- Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, Vol. 37, n°1/2, pp. 17-23.
- Morice, E., et Chartier, F. (1954). *Méthodes statistiques* (vol. 2). Paris, Imprimerie Nationale, 555 p.
- Palm, R. (1986). *Etude des résidus de régression : principes et application*. Gembloux (Belgique), Faculté des Sciences agronomiques, 13 p.
- Palm, R. (1987). *Etude des séries chronologiques par les méthodes de décomposition*. Gembloux (Belgique), Faculté des Sciences agronomiques, 25 p.
- Papadakis, J.S. (1937). Méthode statistique pour les expériences sur champ. Institut d'amélioration des plantes à Salonique (Grèce). *Bulletin scientifique*, no. 23.
- Pearce, S.C. (1976). An examination of Fairfield Smith's law of environmental variation. *J. Agric. Sci.*, 87, 21-24.
- Pearce, S.C. (1998). Field experimentation on rough land : the method of Papadakis reconsidered. *J. Agric. Sci.* 131(1), 1-11.
- Pearce, S.C., and Moore, C.S. (1976). Reduction of experimental error in perennial crops, using adjustment by neighbouring plots. *Expl. Agric.* 12, 267-272.
- Pearson, E.S., and Hartley, H.O. (1966). *Biometrika tables for statisticians*(Vol. I). University Press, Cambridge, 264 p.
- Philippeau, G. (1985). *Théorie des plans d'expérience : application à l'agronomie*. Paris, Institut Technique des Céréales et des Fourrages, 205 p.
- Pichot, C. (1993). Analyse de dispositifs par approches itératives prenant en compte les performances des plus proches voisins. *Agronomie*, 13, 109-119.
- Ryan, T.A., and Joiner, B.L. (1976). Normal probability plots and tests for normality. Pennsylvania State University. 12 p.
- Sawada, M. (2004). Global Spatial Autocorrelation Indices - Moran's I, Geary's C and the General Cross-Product Statistic. Research paper from the Laboratory for Paleoclimatology and Climatology at the University of Ottawa. (as posted online at : <http://www.lpc.uottawa.ca/publications/moransi/moran.htm>)
- Sghaier, T. (2005). *Contrôle de l'hétérogénéité spatiale dans des essais comparatifs de provenances de pin d'Alep (Pinus halepensis Mill.) installés dans les arboreta de Tunisie*. Institut National Agronomique de Tunisie, (Thèse de doctorat), 126 + 41 p.
- Sghaier, T., Claustriaux, J.J., and Béji, M.A. (2004). Intérêt des modèles des plus proches voisins pour le contrôle de l'hétérogénéité spatiale : application à un essai de provenances de pin d'Alep (*Pinus halepensis* Mill.) en Tunisie. *Revue de l'INAT*, vol. 19, 2, 5-22.
- Sghaier, T. (2003). Contrôle local de l'hétérogénéité spatiale et analyse dynamique d'un essai multi-site de provenances de pin d'Alep (*Pinus halepensis* Mill.) en Tunisie. *Revue Internationale GEO-ECO-TROP*, 27, 1-2 : 9-28.
- Thomas-Agnan, C. (2012). Analyse statistique des données spatiales. Séminaire international de statistique. Institut Basque de statistique (EUSTAT), Novembre 2012, 73 p. http://www.eustat.eus/productosServicios/datos/Seminario_54.pdf
- Wilkinson, G.N., Eckert, S.R., Hancock, T.W., and Mayo, O. (1983). Nearest neighbour (NN) analysis of field experiments. *J. R. Statist. Soc., Ser. B*, 45 (2), 151-211.
- Zimmerman, D.L., and Harville, D.A. (1991). A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics* 47 (1), 233-239.

Annexes

TABLE A1 : Comparaison multiple de moyennes à l'aide de la méthode de Newman et Keuls : Données brutes analysées selon un modèle complètement aléatoire (CA) et un modèle en blocs aléatoires complets (BAC) et données ajustées analysées selon le modèle complètement aléatoire (CA).

	Rang	1	2	3	4	5	6	7	8	9
<i>Données brutes (CA)</i>	Moy.	436,25 ^a	413,75 ^{ab}	391,25 ^{ab}	371,50 ^{ab}	361,50 ^{ab}	353,25 ^{ab}	345,75 ^{ab}	334,75 ^{ab}	300,50 ^b
	Obj.	9	8	5	6	3	2	4	7	1
<i>Données brutes (BAC)</i>	Moy.	436,25 ^a	413,75 ^{ab}	391,25 ^{abc}	371,50 ^{abc}	361,50 ^{abc}	353,25 ^{abc}	345,75 ^{abc}	334,75 ^{bc}	300,50 ^c
	Obj.	9	8	5	6	3	2	4	7	1
<i>Itération 1 (CA)</i>	Moy.	424,43 ^a	403,17 ^a	391,93 ^{ab}	386,56 ^{ab}	356,87 ^{bc}	350,99 ^{bc}	344,26 ^{bc}	330,83 ^c	321,14 ^c
	Obj.	9	8	6	5	2	3	4	7	1
<i>Itération 2 (CA)</i>	Moy.	430,04 ^a	407,85 ^a	388,26 ^{ab}	386,16 ^{ab}	358,48 ^{bc}	353,00 ^{bc}	346,19 ^{bc}	330,47 ^c	317,65 ^c
	Obj.	9	8	6	5	2	3	4	7	1
<i>Itération 3 (CA)</i>	Moy.	427,97 ^a	405,86 ^a	390,60 ^{ab}	385,73 ^{ab}	357,98 ^{bc}	352,06 ^{bc}	345,30 ^{bc}	330,42 ^c	318,57 ^c
	Obj.	9	8	6	5	2	3	4	7	1
<i>Itération 4 (CA)</i>	Moy.	428,90 ^a	406,74 ^a	389,64 ^{ab}	385,92 ^{ab}	358,30 ^{bc}	352,46 ^{bc}	345,65 ^{bc}	330,50 ^c	318,36 ^c
	Obj.	9	8	6	5	2	3	4	7	1
<i>Itération 5 (CA)</i>	Moy.	428,48 ^a	406,34 ^a	390,10 ^{ab}	385,83 ^{ab}	358,18 ^{bc}	352,27 ^{bc}	345,48 ^{bc}	330,47 ^c	318,49 ^c
	Obj.	9	8	6	5	2	3	4	7	1
<i>Itération 13 (CA)</i>	Moy.	428,61 ^a	406,46 ^a	389,97 ^{ab}	385,86 ^{ab}	358,22 ^{bc}	352,33 ^{bc}	345,53 ^{bc}	330,48 ^c	318,45 ^c
	Obj.	9	8	6	5	2	3	4	7	1

7 (321,07)	8 (406,41)	5 (427,64)	4 (335,31)	7 (329,31)	5 (350,90)
6 (403,19)	4 (352,88)	3 (322,48)	3 (387,16)	2 (395,40)	6 (387,89)
2 (323,31)	9 (401,38)	1 (358,92)	8 (389,40)	9 (408,13)	1 (332,03)
5 (381,44)	7 (361,01)	9 (440,84)	2 (378,18)	3 (343,99)	7 (311,94)
1 (303,30)	2 (330,60)	8 (432,04)	5 (386,27)	9 (447,36)	8 (384,85)
6 (375,45)	3 (350,31)	4 (337,19)	1 (290,30)	4 (351,67)	6 (401,17)
<i>1^{ère} itération</i>					
7 (326,10)	8 (412,87)	5 (423,80)	4 (349,99)	7 (330,00)	5 (353,71)
6 (395,60)	4 (348,70)	3 (338,56)	3 (387,11)	2 (396,05)	6 (389,44)
2 (322,35)	9 (407,22)	1 (354,81)	8 (400,98)	9 (413,11)	1 (326,51)
5 (374,69)	7 (356,13)	9 (455,56)	2 (379,60)	3 (344,65)	7 (309,63)
1 (296,05)	2 (335,92)	8 (430,20)	5 (392,45)	9 (444,27)	8 (387,35)
6 (367,30)	3 (341,66)	4 (335,19)	1 (293,24)	4 (350,86)	6 (400,72)
<i>2^{ème} itération</i>					
7 (324,49)	8 (410,38)	5 (422,89)	4 (343,66)	7 (328,03)	5 (350,89)
6 (399,17)	4 (349,51)	3 (332,51)	3 (384,91)	2 (395,51)	6 (389,00)
2 (323,40)	9 (405,41)	1 (354,36)	8 (395,90)	9 (410,67)	1 (328,25)
5 (378,81)	7 (357,90)	9 (450,27)	2 (377,61)	3 (345,01)	7 (311,27)
1 (299,76)	2 (335,38)	8 (430,10)	5 (390,34)	9 (445,52)	8 (387,04)
6 (372,74)	3 (345,82)	4 (336,81)	1 (291,91)	4 (351,20)	6 (401,50)
<i>3^{ème} itération</i>					

FIGURE A1 : (1/2) Données ajustées à l'effet terrain pour les itérations 1, 2, 3, 4, 5 et 13.

7 (325,36)	8 (411,39)	5 (423,09)	4 (346,41)	7 (328,72)	5 (352,26)
6 (397,52)	4 (349,09)	3 (335,09)	3 (385,64)	2 (395,42)	6 (389,02)
2 (323,33)	9 (406,32)	1 (354,47)	8 (398,11)	9 (411,79)	1 (327,56)
5 (376,96)	7 (357,19)	9 (452,55)	2 (378,50)	3 (344,88)	7 (310,71)
1 (298,54)	2 (335,96)	8 (430,14)	5 (391,38)	9 (443,95)	8 (387,34)
6 (370,48)	3 (344,22)	4 (336,13)	1 (292,85)	4 (350,97)	6 (401,56)
4 ^{ème} itération					
7 (325,00)	8 (410,91)	5 (422,95)	4 (345,14)	7 (328,36)	5 (351,65)
6 (398,26)	4 (349,27)	3 (333,90)	3 (385,25)	2 (395,41)	6 (388,99)
2 (323,44)	9 (405,93)	1 (354,39)	8 (397,09)	9 (411,29)	1 (327,89)
5 (377,81)	7 (357,53)	9 (451,50)	2 (378,09)	3 (344,95)	7 (311,00)
1 (299,18)	2 (335,76)	8 (430,11)	5 (390,92)	9 (445,21)	8 (387,23)
6 (371,54)	3 (345,00)	4 (336,45)	1 (292,48)	4 (351,05)	6 (401,61)
5 ^{ème} itération					
7 (325,12)	8 (411,05)	5 (422,98)	4 (345,53)	7 (328,47)	5 (351,84)
6 (398,03)	4 (349,21)	3 (334,27)	3 (385,36)	2 (395,40)	6 (389,00)
2 (323,42)	9 (406,05)	1 (354,41)	8 (397,40)	9 (411,45)	1 (327,79)
5 (377,54)	7 (357,43)	9 (451,82)	2 (378,22)	3 (344,93)	7 (310,91)
1 (299,00)	2 (335,84)	8 (430,12)	5 (391,07)	9 (445,13)	8 (387,27)
6 (371,22)	3 (344,77)	4 (336,35)	1 (292,61)	4 (351,02)	6 (401,61)
13 ^{ème} itération					

FIGURE A1 :(2/2) Données ajustées à l'effet terrain pour les itérations 1, 2, 3, 4, 5 et 13.

7 (9,25)	8 (25,25)	5 (67,75)	4 (47,25)	7 (31,25)	5 (-6,25)	+	+	+	+	+	-
6 (16,5)	4 (14,25)	3 (30,5)	3 (75,5)	2 (90,75)	6 (62,5)	+	+	+	+	+	+
2 (-60,25)	9 (-30,25)	1 (71,5)	8 (24,25)	9 (2,75)	1 (33,5)	-	-	+	+	+	+
5 (-52,25)	7 (-2,75)	9 (42,75)	2 (31,75)	3 (-23,5)	7 (-37,75)	-	-	+	+	-	-
1 (-66,5)	2 (-62,25)	8 (-0,75)	5 (-9,25)	9 (-15,25)	8 (-48,75)	-	-	-	-	-	-
6 (-74,5)	3 (-82,5)	4 (-50,75)	1 (-38,5)	4 (-10,75)	6 (-4,5)	-	-	-	-	-	-
Dispositif CA : Itération 0						Dispositif CA : Itération 0					
7 (-5,47)	8 (4,58)	5 (37,12)	4 (-0,34)	7 (-2,11)	5 (-34,18)	-	+	+	-	-	-
6 (8,16)	4 (3,79)	3 (-18,38)	3 (32,98)	2 (37,23)	6 (-1,11)	+	+	-	+	+	-
2 (-34,74)	9 (-22,26)	1 (35,90)	8 (-9,25)	9 (-17,19)	1 (9,41)	-	-	+	-	-	+
5 (-8,03)	7 (27,06)	9 (23,02)	2 (19,91)	3 (-7,33)	7 (-19,47)	-	+	+	+	-	-
1 (-19,30)	2 (-22,41)	8 (23,78)	5 (5,09)	9 (16,73)	8 (-19,10)	-	-	+	+	+	-
6 (-18,56)	3 (-7,28)	4 (-9,03)	1 (-26,01)	4 (5,58)	6 (11,51)	-	-	-	-	+	+
Dispositif CA : 5 ^{ème} itération						Dispositif CA : 5 ^{ème} itération					

FIGURE A2 : Répartition sur le plan de l'essai des valeurs et des signes des résidus observés issus d'une analyse de la variance à un critère de classification (essai conduit selon un dispositif complètement aléatoire CA) réalisée sur les données brutes (itération 0) et les données ajustées à l'effet terrain (5^{ème} itération).

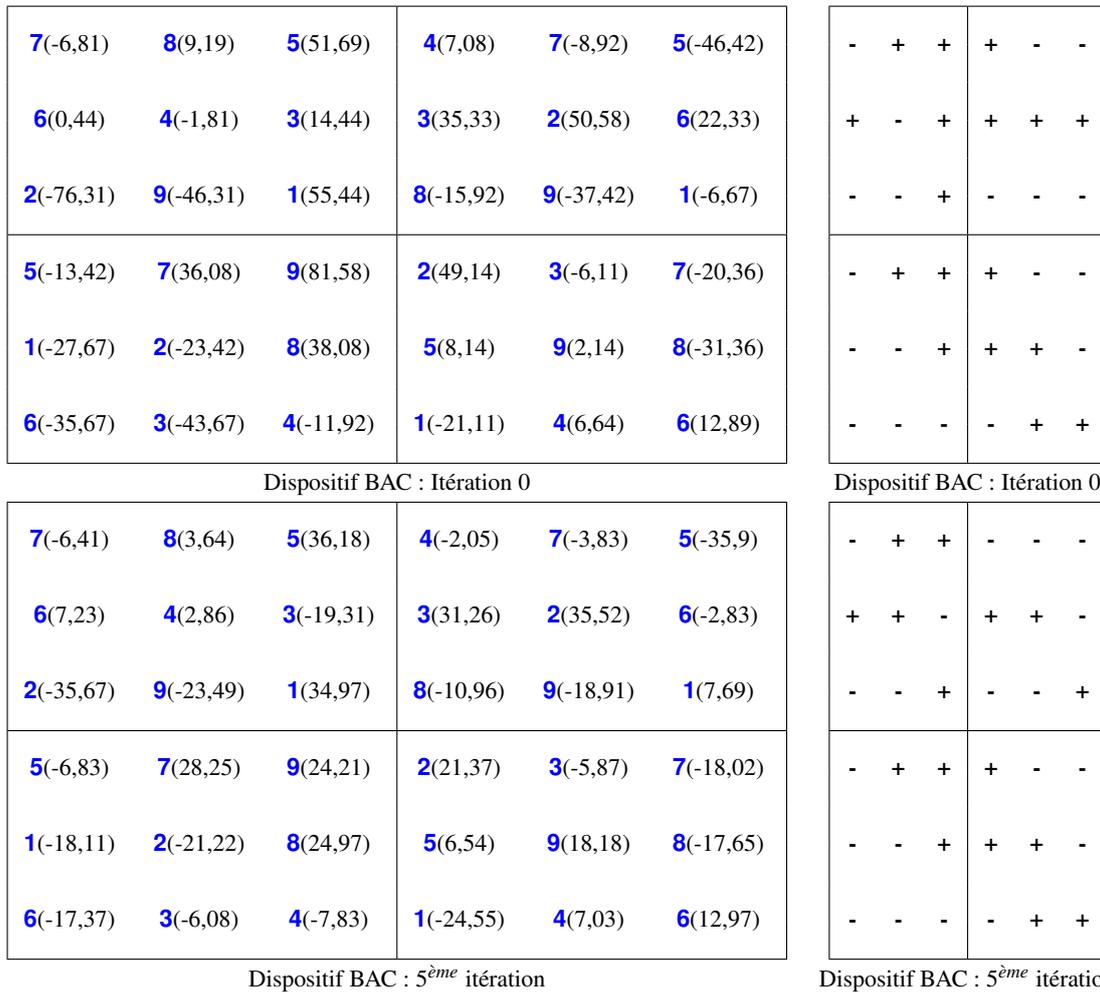


FIGURE A3 : Répartition sur le plan de l'essai des valeurs et des signes des résidus observés issus d'une analyse de la variance à deux critères de classification (essai conduit selon un dispositif en blocs aléatoires complets - BAC) réalisée sur les données brutes (itération 0) et les données ajustées à l'effet terrain (5^{ème} itération).