Université Montpellier I

Université Montpellier II
Sciences et Techniques du Languedoc

Ecole Nationale Supérieure Agronomique
Montpellier

CIRAD

Prédiction de l'interaction génotype x environnement à partir d'indices variétaux de sensibilité à la sécheresse et de bilan hydrique à l'aide d'un modèle de régression factorielle dans des essais d'arachide au Sénégal

Rapport de stage de DEA de Biostatistique
Option : Théorie et méthodes statistiques

par :

Ibnou DIENG

sous la direction de :

Eric GOZÉ

soutenu le 26 juin 2003 devant le jury composé de

C. ABRAHAM	Maître de conférences, ENSA.M
G. DUCHARME	Professeur, Université Montpellier II
Y. GUEDON	Chargé de recherches, CIRAD
P. MONESTIEZ	Chargé de recherches, INRA

"Quoiqu'on ne puisse parvenir à tout savoir,
il faut néanmoins qu'il soit possible
de tout apprendre"
Talleyrand

à

ma famille
ma belle-famille
mes amis
Malie et Papy

Avant-propos

Ce stage a été effectué au Cirad (Centre de coopération internationale en recherche agronomique pour le développement) sous la direction de M. Eric Gozé.

Les données sont de Danièle Clavel, sélectionneur du Cirad en poste au Ceraas (Centre d'étude régional pour l'amélioration de l'adaptation à la sécheresse) où j'exerce la fonction de biométricien. Ce stage, outre son caractère pédagogique, se propose de développer des outils pour exploiter des informations sur la sensibilité des variétés à la sécheresse et sur les caractéristiques climatiques des environnements qui ne sont pas généralement prises en compte dans des essais multilocaux, pour mieux aider la sélection au Ceraas.

Remerciements

J'exprime mes sincères remerciements

à mon maître de stage Monsieur Eric Gozé, pour son encadrement, ses nombreux conseils et sa grande disponibilité dont il a fait montre tout au long de ce stage

à Monsieur Philippe Letourmy, responsable de l'Unité Mabis, de m'avoir accepté dans son unité et pour ses nombreux conseils durant ce stage

à Messieurs Harold Roy-Macauley et Serge Braconnier, Directeur et Responsable scientifique du Ceraas, de m'avoir donné l'opportunité de m'avoir déchargé provisoirement de mes obligations au Ceraas pour suivre cette formation et d'en avoir financé la première partie

à Madame Danièle Clavel d'avoir mis à ma disposition les données de l'étude et pour toutes les informations qu'elle m'a fournies

à toute l'équipe de l'unité Mabis pour la convivialité

et à tous mes collègues du Ceraas pour le soutien durant cette formation.

Enfin, je ne saurais terminer sans remercier le Ministère des affaires étrangères du gouvernement français d'avoir pris en charge financièrement une partie du stage.

Table des figures

4.1 Répartition des pluies sur la station de Bambey de 1994 à 1997 27

4.2 Résidus des modèles des essais multilocaux de 1996 à 1998 en fonction des valeurs prédites 28

4.3 Droite de Henry des résidus et statistiques de tests d'Anderson-Darling pour les données des essais multilocaux 29

4.4 Résidus des modèles des essais pluriannuels de 1996 à 1998 en fonction des valeurs prédites 32

4.5 Droite de Henry des résidus et statistiques de tests d'Anderson-Darling pour les données des essais pluriannuels 33

5.1 Boîte à moustaches des résidus du modèle de 80J1 36

5.2 Boîte à moustaches des résidus du modèle de 80J2 37

5.3 Boîte à moustaches des résidus du modèle de 80J4 38

5.4 Boîte à moustaches des résidus du modèle de 90J3 39

Liste des tableaux

2.1 Tests d'adéquation fondés sur la fonction de répartition théorique 16

4.1 Tests de rapport de vraisemblance maximale sur modèles emboîtés pour l'homogénéité de variances résiduelles d'essais multilocaux au Sénégal de 1996 à 1998 27

4.2 Tests des effets des interactions d'essais multilocaux au Sénégal de 1996 à 1998 30

4.3 Tests de rapport de vraisemblance sur modèles emboîtés pour des essais d'arachide à Bambey de 1994 à 1997 30

4.4 Tests des effets des interactions pour des essais d'arachide à Bambey de 1994 à 1997 31

Glossaire

ACP	Analyse en composantes principales
AMMI	Additive main effect and multiplication interaction
C	Taux de croissance de culture
CERAAS	Centre d'étude régional pour l'amélioration de l'adaptation à la sécheresse
CIRAD	Centre de coopération internationale en recherche agronomique pour le développement
CORAF	Conseil ouest et centre africain pour la recherche et le développement agricoles
ETR	Evapotranspiration réelle
EVM	Essai variétal multilocal
EVP	Essai variétal pluriannue
ISRA	Institut sénégalais de recherches agricoles
MABIS	Mathématiques appliquées biométrie informatique scientifique
PF	Poids de fanes
PG	Poids de gousses
SARRAH	modèle de simulation de bilan hydrique habillé (détaillé)
T	Durée floraison-récolte

Table des matières

Avant-propos **i**

Remerciements **ii**

Glossaire **v**

1 Introduction **1**

1.1 Problématique 2

1.2 Synthèse bibliographique sur les interactions GxE 3

2 Dispositifs et modèles d'analyse de variance associés **9**

2.1 Essais multilocaux 10

2.2 Essais pluriannuels 11

2.3 Vérification des présuppositions de l'analyse de variance 13

2.3.1 Homogénéité des variances 14

2.3.2 Normalité des résidus 15

3 Régression factorielle **18**

3.1 Présentation du modèle 18

3.2 Description des covariables 23

3.2.1 Covariables associées au facteur lieu 24

3.2.2 Covariable associée au facteur variété 25

4 Vérification des données et des tests des interactions 26

4.1 Résultats des essais multilocaux 27

4.2 Résultats des essais pluriannuels 30

5 Résultats des estimations des modèles de régression factorielle 34

5.1 Essai multilocal de 1996 35

5.2 Essais pluriannuels 35

5.2.1 Essai pluriannuel 80J1 35

5.2.2 Essai pluriannuel 80J2 37

5.2.3 Essai pluriannuel 80J4 38

5.2.4 Essai pluriannuel 90J3 39

6 Conclusion et Perspectives 40

Références citées 42

Annexes 44

A. Liste des variétés 44

B. Modèle de Régression factorielle 45

C. Modèle de simulation de bilan hydrique 49

Chapitre 1

Introduction

Le laboratoire d'accueil

Le Centre de coopération internationale en recherche agronomique pour le développement est un organisme français de recherche agronomique au service du développement durable des pays tropicaux et subtropicaux. Ses activités de recherche s'appliquent aux domaines des sciences du vivant et des sciences sociales appliquées à l'agriculture, à la forêt, à l'élevage, à la gestion des ressources naturelles, à l'agroalimentaire, aux écosystèmes et aux sociétés du Sud.

Le service Mathématiques appliquées biométrie et informatique scientifique (Mabis) est un service du département Cultures annuelles (Ca). Ses activités de travail concernent l'informatique et les mathématiques. L'activité de Biométrie a pour objectif d'améliorer l'utilisation des techniques statistiques. L'activité d'informatique touche la gestion de projets, la programmation d'applications spécifiques sous WINDOWS, les SIG, les bases de données et l'analyse d'image.

1.1 Problématique

Dans les pays du Sahel, l'agriculture est contrainte par des conditions de milieu difficiles, au premier plan desquelles la rareté et l'irrégularité des ressources en eau, ce qui constitue un problème majeur pour les paysans. Dans ces conditions, la recherche de variétés adaptées à la sécheresse est un objectif de nombreux programmes de recherche. Le Ceraas [1] a pour mission de comprendre les mécanismes d'adaptation des plantes à la sécheresse afin de proposer des méthodes de sélection, de suivi des cultures et des itinéraires techniques en fonction des caractéristiques notamment hydriques du milieu ciblé.

La sélection porte sur des caractères phénologiques, physiologiques et moléculaires permettant une production améliorée en conditions de déficit hydrique par rapport aux cultivars actuellement vulgarisés. Elle est validée par des essais en plein champ, réalisés dans des conditions qui reflètent la variabilité du milieu auquel les variétés sont destinées.

Etant donné qu'en milieu sec, le rendement est soumis à une variabilité environnementale accrue par rapport à des conditions de culture favorable, les essais multilocaux ou pluriannuels permettent généralement de constater de fortes interactions entre génotype et environnement, notamment sur les variables de production. Pourtant, à la différence de l'effet principal des variétés qui sert effectivement de critère de sélection, les interactions ne sont pas exploitées judicieusement faute de pouvoir les prédire dans une situation nouvelle. Le but de cette analyse est de prédire les éventuelles interactions d'essais réalisés au Sénégal, entre 1994 et 1998, au moyen de modèles de bilan hydrique et de régression factorielle.

Le chapitre 1 présente l'introduction et la synthèse bibliographique sur les interactions génotype x environnement (GxE). Dans le chapitre 2, nous parlerons des dispositifs des essais et des modèles d'analyse de variance qui leur étaient associés, en rappelant quelques procédures empiriques et de tests de vérification des présupposés de l'analyse de variance. Dans le chapitre 3, nous exposerons le

1. Centre d'étude régional pour l'amélioration de l'adaptation à la sécheresse, placé sous la double tutelle de l'Institut sénégalais de recherches agricoles (Isra) et du Conseil ouest et centre africain pour la recherche et le développement agricoles (Coraf)

modèle de régression factorielle et décrirons les covariables associées aux deux facteurs génotype et environnement. Dans le chapitre 4, nous présenterons les résultats des tests des interactions tandis que le chapitre 5 sera consacré à ceux des estimations des paramètres inconnus de la régression factorielle et l'étude de la qualité de la régression. Enfin, dans le chapitre 6, nous conclurons cette étude et dégagerons des perspectives.

1.2 Synthèse bibliographique sur les interactions GxE

Des essais multilocaux sont utilisés généralement pour comparer des variétés dans plusieurs environnements. Les éventuelles variations des différences des réponses des variétés d'un lieu à un autre sont appelées interaction. Dans ce cas, on s'écarte de la situation d'additivité : l'écart entre les génotypes n'est pas le même selon les conditions du milieu.

Le modèle est le suivant :

$$Y_{ij} = m + a_i + \ell_j + (a\ell)_{ij} + E_{ij} \tag{1.1}$$

où Y_{ij} est la réponse de la variété i de l'environnement j, m représente la moyenne générale, a_i l'effet génotype, ℓ_j l'effet environnement, $(a\ell)_{ij}$ l'effet de l'interaction GxE et E_{ij} l'erreur aléatoire .

Ce modèle est habituellement étudié par une analyse de variance. Or l'analyse de variance classique n'explore pas la structure non additive du modèle, elle se borne à tester si l'effet de l'interaction est significatif. Le cas échéant, l'information contenue dans cette interaction serait inexploitée si aucune analyse supplémentaire n'était faite [4].

Le rôle de la modélisation de l'interaction GxE est dès lors, de tenter d'expliquer au mieux la partie non additive du modèle afin de prédire les réponses des génotypes plus finement et par là, les différences entre génotypes.

D'autre part, d'après le modèle 1.1, les estimations des effets sont :

$$\hat{a}_i = Y_{i.} - Y_{..}$$

$$\hat{\ell}_j = Y_{.j} - Y_{..}$$

$$(\hat{a\ell})_{ij} = Y_{ij} - Y_{i.} - Y_{.j} + Y_{..}$$

Dans l'estimation des termes du modèle qui portent l'indice j, on retrouve $Y_{.j}$ qui est la moyenne des réponses de toutes les variétés du milieu j. Cette moyenne traduit le potentiel du milieu. Or le milieu étant fortement variable et existant rarement plus de 5 ans d'essais pour la même variété, les termes en j ne sont pas bien prévisibles si l'on se cantonne à l'analyse de variance. Cependant si nous considérons la différence entre deux variétés i et i', l'imprévisibilité de l'effet environnement ℓ_j sera gommée lors de l'estimation de cette différence, si le dispositif est complet. En effet, il viendra :

$$Y_{ij} - Y_{i'j} = a_i - a_{i'} + (a\ell)_{ij} - (a\ell)_{i'j}$$

Par contre, le problème demeure pour les interactions. Elles pourront au contraire être expliquées par des données climatiques des milieux étudiés qui sont certes variables mais ont l'avantage d'être nombreuses (plusieurs dizaines d'années).

Plusieurs modèles de l'interaction GxE ont été proposés. Dans ce qui suit, nous en ferons un tour d'horizon.

Régression conjointe : dans ce modèle, l'interaction est liée de façon linéaire à l'effet du milieu j estimé. Pour chaque génotype, la pente calculée caractérise la sensibilité différentielle du génotype i au milieu [6]. Ce modèle a été proposé pour la première fois par Yates et Cochran [23]. L'interaction est de la forme :

$$(a\ell)_{ij} = c_i\hat{\ell}_j$$

où $\hat{\ell}_j$ est l'estimation de l'effet du milieu j issu du modèle additif. Il a ensuite été repris par Finlay et Wilkinson [9] puis par Eberhart et Russel [8] qui considéraient c_i comme la stabilité du génotype i.

Un cas particulier a été développé par Tukey [19] : les coefficients c_i s'écrivent Ka_i et l'interaction se réduit à $Ka_i\ell_j$.

La méthode d'Eberhart et Russel [8], couramment utilisée, paraît peu adaptée à cette forte variation du milieu d'une année sur l'autre ou d'un site à l'autre, car elle suppose un indice de fertilité unique pour le milieu, auquel les variétés sont plus ou moins sensibles. Au contraire, une saison de culture est souvent marquée par plusieurs épisodes secs dont l'influence dépend de l'état de développement de la plante qu'ils touchent. Pour la plupart des espèces, une variété n'est pas caractérisée par une seule sensibilité à la sécheresse, mais par une sensibilité variable selon son stade de développement, autrement dit une fonction de sensibilité.

Méthode AMMI (Additive main effect and multiplicative interaction) : cette méthode associe l'analyse de variance et l'analyse en composantes principales. Sont d'abord estimés les effets principaux des variétés et des milieux par une analyse de variance du modèle additif. Ensuite, la partie non additive du modèle est étudiée par une analyse en composantes principales [4]. Cette méthode a été introduite par Williams [21] et reprise par Mandel [13] [14] et Gollob [11].

L'interaction est décrite de cette façon :

$$(a\ell)_{ij} = k_1 v_{1i} s_{1j} + k_2 v_{2i} s_{2j} + k_3 v_{3i} s_{3j} + \dots$$

c'est à dire :

$$Y_{ij} = m + a_i + \ell_j + (\sum k_n v_{ni} s_{nj}) + E_{ij}$$

où k_n est la valeur singulière du n^{e} axe (k_n^2 étant la valeur propre), v_{ni} est le vecteur propre du i^{e} génotype pour le n^{e} axe, s_{nj} est le vecteur propre du j^{e} environnement pour le n^{e} axe et $\sum v_{ni} = \sum s_{nj} = 1$.

Le fait de modéliser l'interaction sans utiliser les données météorologiques n'est pas judicieux car, de même que l'analyse de variance, les effets estimés pour les lieux sont imprévisibles. Cette méthode a qu'un grand intérêt descriptif mais ne permet une prédiction pertinente.

Structuration de l'interaction : l'idée ici est de partitionner les variétés en M groupes et les lieux en N groupes, soit par des considérations extérieures aux données en disposant d'informations *a priori*, soit par classification automatique.

Ensuite, d'expliquer la réponse des variétés par la combinaison des MN groupes [7].

Ainsi, l'information relative à l'interaction globale est concentrée dans une interaction entre les groupes engendrés par la structuration [6]. Le modèle est le suivant :

$$Y_{ij}^{s} = m + a_s^B + a_i^W + \ell_t^B + \ell_j^W + E_{st}^{BB} + E_{sj}^{BW} + E_{it}^{WB} + E_{ij}^{WW}$$

où $s = 1, \ldots, S$ avec S le nombre de groupes de lignes ; $t = 1, \ldots, T$ avec T le nombre de groupes de colonnes. Les indices B et W sont utilisés pour *Between* et *Within*. $a_S^B + a_i^W$ représente l'effet de la variété i, $\ell_t^B + \ell_j^W$ l'effet du lieu j et $E_{st}^{BB} + E_{sj}^{BW} + E_{it}^{WB} + E_{ij}^{WW}$ l'effet de l'interaction.

La détermination des groupes dans les deux situations évoquées plus haut ne permet pas de prendre en compte toute l'information apportée par les données météorologiques dans la modélisation de l'interaction. Par la classification automatique, les groupes ne sont définis qu'à partir des données, ce qui en limite la portée surtout dans une perspective de prédiction. Par l'utilisation des pluies par exemple enregistrées dans les lieux pour les caractériser en zones plus ou moins sèches, il y aurait perte d'informations car les données de pluies sont continues alors que les groupes ne seraient que discrets.

Régression factorielle : c'est le modèle additif auquel trois termes de régression sont ajoutés. Si nous faisons l'hypothèse qu'il n'y a qu'une covariable par facteur, le modèle est

$$Y_{ij} = m + a_i + \ell_j + X_i \cdot \gamma \cdot Z_j + \alpha_i \cdot Z_j + X_i \cdot \beta_j + E_{ij} \tag{1.2}$$

- Y_{ij} est la réponse de la variable
- μ est la moyenne générale
- a_i est l'effet principal du facteur variété
- ℓ_j est l'effet principal du facteur lieu
- $X_i \cdot \gamma \cdot Z_j$ est la régression des observations sur le produit des covariables

- $\alpha_i \cdot Z_j$ est la régression sur la covariable Z_j mais le coefficient de régression dépend du facteur variété
- $X_i \cdot \beta_j$ est la régression sur la covariable X_i avec le coefficient de régression qui dépend du facteur lieu
- E_{ij} est tout ce qui reste de l'interaction.

Dans le cas où chaque variété est présente dans chaque lieu, les observations peuvent être mises dans une matrice à I (nombre de variétés) lignes et J (nombre de lieux) colonnes. Alors, la régression factorielle peut aussi être traitée de cette façon : une première régression des observations est faite sur la covariable associée au facteur variété, par exemple. La partie expliquée par cette régression et celle qui ne l'est pas sont ensuite expliquées par la covariable associée au facteur lieu. Et la réponse prédite pour les variétés sera la résultante de ces trois régressions [2].

Un modèle similaire de celui de régression factorielle a été utilisé par Freeman et Perkins [10] qui n'avaient considéré qu'une covariable associée au facteur milieu non calculée sur les données (pluviométrie, température). Wood [22], lui, a utilisé une combinaison linéaire de covariables élémentaires liées au milieu.

Les modèles de l'interaction, pour la plupart, permettent de décrire le phénotype par une fonction multiplicative des effets du génotype et du milieu. Les effets des facteurs, dont cette fonction est le produit, peuvent être estimés à partir des données : c'est le cas, comme nous l'avons vu, pour l'analyse AMMI qui repose sur une ACP des résidus du modèle additif. Le facteur milieu étant variable d'une année à l'autre, cette approche ne permet pas la prédiction des performances des variétés dans un nouveau milieu. Les facteurs, peuvent au contraire être connus *a priori* : c'est le cas pour le milieu, où nous connaissons pour chaque essai, la pluviométrie, l'ensoleillement, le vent, etc. L'inconvénient est que ces variables climatiques sont trop nombreuses. Dans ce cas, l'utilisation de modèle de simulation pour les résumer peut être une alternative.

Justement, au Cirad, un modèle de simulation de culture à l'échelle de la parcelle a été écrit pour prédire localement, à partir de la pluviométrie, la production d'arachide (*Arachis hypogaea* L.), la principale culture de rente au Sénégal, qui est prise comme plante test. Ce modèle SarraH permet également de résumer les

séries journalières de données climatiques en quelques taux de satisfaction des besoins en eau, en prenant en compte l'influence non-linéaire des variables d'entrée. Ses entrées sont des données climatiques disponibles sur les essais de sélection, mais il ne prend pas en compte certaines caractéristiques variables avec le lieu, comme la fertilité.

Ainsi, nous voulons utiliser les sorties du modèle SarraH comme régresseurs associés au milieu, dans la modélisation de l'interaction GxE par régression factorielle. Les variations de fertilité du lieu sont alors prises en compte dans l'effet lieu, et dans ce milieu à pluviométrie très variable, nous pouvons espérer que la majeure partie des interactions GxE sera expliquée par l'interaction génotype x pluviométrie.

D'un autre côté, toujours dans l'optique de mieux interpréter l'interaction GxE, une variable de sensibilité à la sécheresse des variétés peut être substituée à l'effet du génotype dans la fonction multiplicative du modèle des interactions.

Chapitre 2

Dispositifs et modèles d'analyse de variance associés

Des conclusions tirées sur des essais variétaux isolés ne sont valables que dans les conditions particulières de chacun d'eux. Pour espérer produire des recommendations qui doivent s'appliquer à des variétés dans un ensemble de lieux par exemple, il faut multiplier le nombre d'essais dans le temps et/ou dans l'espace. Pour notre cas, la multiplication des essais, d'une part en plusieurs endroits et, d'autre part en plusieurs années sur un même site, permet d'avoir des environnements contrastés du point de vue des caractéristiques climatiques (pluviométriques essentiellement) pour une meilleure prédiction de l'interaction GxE.

Les essais étaient conduits en 1996 sur 3 localités (Maka Fall, Ndieye Ndiaye et Pakhi Kébé), en 1997 et en 1998 sur 4 localités (Maka Fall, Ndieye Ndiaye, Pakhi Kébé et Darou Sam). Ces essais concernant plusieurs sites, sont dits multilocaux et seront notés Essais variétaux multilocaux (EVM). Pour les 3 années, nous avons 3 EVM : EVM96, EVM97 et EVM98.

De plus, 7 autres essais variétaux avaient été menés sur la station expérimentale du Ceraas, située à Bambey (14° 42 N et 16° 28 O), de 1994 à 1997. Ils ne concernaient qu'un seul site et étaient pluriannuels ; ils seront notés Essais variétaux pluriannuels (EVP) [1].

1. pour la liste des variétés utilisées pour les EVP et les EVM, voir Annexe A

2.1 Essais multilocaux

Le but de ces essais était de tester pour chaque année (1996 à 1998), aux différents sites (conditions paysannes), deux ou trois variétés très précoces (cycle de 80 jours) en comparaison avec une ou deux variétés témoin déjà vulgarisées mais qui étaient devenues trop tardives pour ces zones.

Ces essais étaient disposés en blocs complets randomisés. Ce plan d'expérience est tel que chaque traitement (ici variété) est présent dans chaque bloc. Ce type de dispositif est adapté à un terrain d'expérience réputé non homogène comme celui de la station expérimentale de Bambey. Le principe, pour contrôler cette hétérogénéité, est de scinder le terrain en bandes homogènes sur chacune desquelles, l'ensemble des variétés est réparti. On espère ainsi une diminution de la variation résiduelle. S'il n'y a qu'une seule fois la même variété par bloc, le modèle s'écrit :

$$Y_{ij} = m + a_i + B_j + E_{ij}$$

où Y_{ij} est l'observation de la i^e variété et du j^e bloc, m est la moyenne générale, a_i est l'effet variété, B_j est l'effet bloc et les E_{ijk} sont les erreurs résiduelles ; a_i est un effet fixe tandis que B_j est un effet aléatoire. Dans ce qui suit, un effet fixe sera noté par une lettre minuscule et un effet aléatoire par une lettre majuscule.

Chaque EVM avait entre 4 et 6 variétés disposées en 4 répétitions et testées sur 3 à 4 localités. Si Y_{ijk} est la réponse de l'unité expérimentale du k^e lieu, chaque modèle s'écrit :

$$Y_{ijk} = m + a_i + B_{jk} + \ell_k + (a\ell)_{ik} + E_{ijk}$$

où ℓ_k est l'effet lieu, $(a\ell)_{ik}$ est l'effet variété x lieu, B_{jk} est l'effet bloc hiérarchisé dans les lieux.

Ainsi, pour chaque modèle, on s'intéressera prioritairement à l'effet variété x lieu que nous voulons modéliser si toutefois il est significatif.

2.2 Essais pluriannuels

Les EVP ont concerné deux types de variétés : celles à cycle de développement très court (75 à 80 jours) et celles à cycle plus long (90 jours). Les essais avec les premiers types de variétés répondaient à l'objectif de recherche de variétés à cycle court plus adaptées dans ces zones à pluviométrie de plus en plus réduite. Les autres essais, quant à eux, avaient pour but la recherche de variétés physiologiquement adaptées à la sécheresse. Chaque type d'essai était scindé en essais qui se différenciaient par l'origine des variétés. La plupart des variétés (hormis les variétés témoins) avaient été obtenues par croisement et selon les parents dans le croisement, les variétés à cycle 80 jours étaient regroupées dans 4 essais différents nommés 80J1, 80J2, 80J3 et 80J4. Les variétés à cycle 90 jours étaient dans 3 essais : 90J1, 90J2, 90J3.

La technique des dispositifs en blocs complets utilisée pour les EVM présente un inconvénient majeur dès que le nombre de traitements devient élevé. Une taille importante des blocs induit souvent une hétérogénéité notable à l'intérieur des blocs ; le contrôle de l'hétérogénéité devient alors inefficace. Pour pallier cet inconvénient, des plans d'expérience ont été proposés se caractérisant par un nombre d'unités expérimentales par bloc inférieur au nombre de niveaux du facteur étudié. Ces dispositifs sont en "blocs incomplets". Définissons les par ces quelques paramètres :

- v : nombre de traitements
- k : nombre d'unités expérimentales par bloc incomplet
- b : nombre total de blocs incomplets
- r : nombre constant de répétitions de chaque traitement
- λ : nombre de comparaisons de deux traitements différents. Quand ce nombre est constant, on parle de dispositif "équilibré".

On peut s'arranger avec la disposition des blocs incomplets pour avoir des groupes de répétitions à l'intérieur desquels tous les traitements se trouvent le même nombre de fois. Quand chaque groupe de répétition comprend une seule fois chaque traitement, il est appelé réplique et le dispositif est dit lattice. Dans ce cas, le nombre

de traitements est égal à k^2. Les lattices qui ont $k(k + 1)$ traitements sont appelés lattices rectangulaires

Les EVP étaient disposés en lattices rectangulaires 3 x 4. Pour chaque essai, il y avait 12 variétés au total, 3 variétés par bloc incomplet et 4 répliques. Si nous notons Y_{ijq} l'observation du i^e traitement, du j^e bloc et de la q^e réplique, le modèle pour un essai, est :

$$Y_{ijq} = m + a_i + R_q + B_{jq} + E_{ijq}$$

où m est la moyenne générale, a_i est l'effet traitement, R_q est l'effet réplique, B_{jq} est l'effet bloc hiérarchisé dans les répliques et les E_{ijq} sont les erreurs résiduelles.

L'une des particularités d'un bloc incomplet est la non indépendance entre les blocs et les traitements. En effet, tous les traitements ne se trouvant pas en même temps dans chaque bloc, une hétérogénéité des blocs induit forcément une différence entre les traitements. D'où la nécessité de réajuster les moyennes des traitements en fonction des blocs. Alors, les tests d'hypothèse fondés sur les carrés moyens ajustés sont réalisés avec une perte d'efficacité. La quantité

$$E = \frac{1-1/k}{1-1/v}$$

est le coefficient d'efficacité du bloc incomplet équilibré et constitue la différence essentielle dans l'estimation des effets traitement entre le bloc complet et le bloc incomplet équilibré ; k étant inférieur à v, ce coefficient E est toujours inférieur à 1 et pour k=2 (valeur minimale pour k) et v devenant très grand, E tend vers 1/2. Ainsi, un bloc incomplet équilibré est toujours moins efficace qu'un bloc complet, à nombre de répétitions par traitement égal [16].

Les lattices étant des blocs incomplets, la propriété d'additivité des sommes des carrés d'écart (SCE), SCE totale = SCE blocs + SCE traitements + SCE résiduelle, n'est plus vraie. Il faut alors soit considérer SCE totale = SCE blocs + SCE traitements ajustés + SCE résiduelle, soit considérer SCE totale = SCE blocs ajustés + SCE traitements + SCE résiduelle. Ce qui conduit à l'égalité SCE blocs + SCE traitements ajustés = SCE blocs ajustés + SCE traitements [3].

Ces essais étant répétés plusieurs années de suite (1994 à 1997) sur un même lieu, si Y_{ijqk} est la réponse de la i^e variété, du j^e bloc, de la q^e réplique et de la k^e année, le modèle de ces séries d'essais s'écrit :

$$Y_{ijqk} = m + a_i + n_k + R_{qk} + B_{jqk} + (an)_{ik} + E_{ijqk}$$

où n_k est l'effet année. S'agissant d'une culture annuelle, on change d'endroit et on rerandomise chaque année.

L'objectif de ce modèle est de tester, entre autres, si l'effet de l'interaction variété x année est significatif.

L'effet bloc étant aléatoire, l'analyse se fait en récupérant l'information entre blocs. Le principe est d'obtenir deux estimations distinctes de la variation entre les traitements : i) d'abord à partir des effets estimés des variétés dans le modèle classique, ii) ensuite en extrayant des différences entre blocs ce qui est dû aux différences entre les traitements, puis de les relier et d'en trouver une estimation unique, pondérée, sans biais qui permette de tester les différences entre les traitements [16].

2.3 Vérification des présuppositions de l'analyse de variance

Pour tester les effets des interactions des essais, des modèles d'analyse de variance sont utilisés. Ce qui prévoit un certain nombre de présuppositions que nous allons rappeler : les erreurs résiduelles de chaque modèle sont indépendantes, de variances égales (ici d'un lieu à un autre en ce qui concerne les essais multilocaux et d'une année à une autre pour les essais pluriannuels) et suivent une loi gaussienne. L'indépendance des erreurs est assurée par la randomisation (affectation au hasard des traitements aux unités expérimentales). La normalité des erreurs et l'égalité des variances seront vérifiées *a posteriori*. Cette vérification porte sur les résidus qui sont les écarts entre les valeurs observées et les valeurs prédites et qui estiment les erreurs résiduelles.

Si les deux dernières présuppositions ne sont pas satisfaites, nous pourrons fonder notre étude sur une analyse de variance non paramétrique. Les conditions d'application des tests sont moins restrictives que celles de l'analyse de variance usuelle. A tout le moins, faut-il que les échantillons soient aléatoires et indépendants entre eux, d'où l'importance de la randomisation. L'idée est d'effectuer une analyse de variance basée sur les rangs, c'est à dire sur les numéros d'ordre des valeurs observées rangées par ordre croissant. Ce qui simplifie les calculs mais provoque une perte d'information notamment les notions d'additivité et d'interaction qui n'ont pas de sens pratique à notre connaissance. Si le modèle est à deux facteurs (génotype et environnement par exemple), le test de Friedman peut aussi être utilisé pour la significativité des effets mais sans possibilité de tester une interaction.

2.3.1 Homogénéité des variances

En général, les tests d'égalité de variance sont peu puissants et peu robustes à la non normalité des erreurs. Le plus usité est celui de Bartlett que nous allons décrire. Soient S_1^{*2} , S_2^{*2} , ... S_g^{*2} les variances de g échantillons indépendants. Si $\sigma_1 = \sigma_2 = \ldots = \sigma_g$ alors la quantité

$$B = (n-g)\ ln\left(\frac{\sum_{i=1}^{g}(n_i - 1)S_i^{*2}}{n-g}\right) - \sum_{i=1}^{g}(n_i - 1)\ ln\ (S_i^{*2})$$

suit approximativement une loi de χ^2 à $g-1$ degrés de liberté.

A l'intérieur d'un même lieu, ce test n'est valable que dans le cas d'un plan d'experience en randomisation totale, ce qui n'est pas notre cas pour les essais. Nous avons donc utilisé pour chaque essai, un test de rapport de vraisemblance sur modèles emboîtés. Nous avons écrit pour chaque essai, un modèle où nous supposions les variances homogènes et avons calculé la vraisemblance. Ensuite pour chaque même essai, la vraisemblance d'un autre modèle où les variances étaient

supposées hétérogènes a aussi été calculée. L'hypothèse nulle de chaque test est alors :

> H0 : le modèle à variances homogènes et celui à variances hétérogènes sont identiques

La statistique de test est la déviance qui suit une loi de χ^2 ayant comme ddl (degrés de liberté) la différence des nombres de paramètres des deux modèles. Par exemple, si nous testons l'homogénéité de la variance résiduelle entre l lieux, le χ^2 aura $l - 1$ ddl.

En outre, il est aussi intéressant, avant de commencer les tests de rapport de vraisemblance, de regarder les graphes des valeurs prédites en fonction des résidus des modèles (en utilisant des modèles où les variances sont supposées égales). Si les variances sont homogènes, le nuage de points obtenu pour chaque modèle ne doit présenter aucune structure particulière. Sinon, il existerait une structure cachée dans les données qu'il faut découvrir.

Aujourd'hui, il existe des logiciels qui peuvent prendre en compte des modèles à variances hétérogènes si toutefois l'effectif de chaque groupe est suffisant. Cela se fait suivant un principe de pondération des variances résiduelles (voir plus loin avec les tests des interactions). Dans le cas particulier du regroupement d'essais, Cochran et Cox [3] recommandent 15 ddl par lieu pour la variance résiduelle si elle doit servir d'inverse de poids.

2.3.2 Normalité des résidus

Pour la vérification de la normalité des résidus, la robustesse des tests F garantit l'utilisation de modèles d'analyse de variance en dehors d'un cadre gaussien à condition de disposer de beaucoup de données (Tomassone et al. [18] suggèrent un nombre d'observations supérieur à la dizaine, chiffre donné de façon empirique, préviennent-ils). Pour les essais multilocaux, nous disposons de 48 observations pour EVM96 et de 80 observations pour EVM97 et pour EVM98. Pour les essais pluriannuels, nous avons une centaine de données. Nous pouvons donc considérer que les éventuelles écarts à la normalité seront sans grande importance, les tests à

effectuer ne servant en pratique qu'à détecter des erreurs de saisie ou de relevé de données.

Nous pouvons donc vérifier cette normalité avec le test d'Anderson-Darling [17]. Ce test appartient à une famille de tests d'adéquation fondés sur la comparaison de la fonction de répartition théorique $F(x)$ et celle empirique $F_n(x)$ obtenue à partir d'un échantillon aléatoire de n observations :

$$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$$

Cette famille de tests comprend essentiellement les tests de Kolmogorov-Smirnov (statistique de test D), Cramèr-von Mises (W^2), Kuiper (V), Watson (U^2) et Anderson-Darling (A). Dans le cas où $F(x)$ est la fonction de répartition de la loi normale, $N(\mu,\sigma)$, μ et σ inconnus, nous posons

$$Z_{(i)} = \Phi\left(\frac{x_{(i)}-\bar{x}}{s}\right)$$

où $\Phi(X)$ est la fonction de répartition de la loi normale $N(0,1)$, $\bar{x}$ est la moyenne de l'échantillon et $s^2 = \sum_i (x_i - \bar{x})^2/(n-1)$. Les statistiques de tests sont présentés dans le tableau 2.1

Test	Statistique
D	$D^+ = \max_{1\leq i\leq n}[i/n - Z_{(i)}]$, $D^- = \max_{1\leq i\leq n}[Z_{(i)} - (i-1)/n]$, $D = \max[D^+,D^-]$
W^2	$\sum_{i=1}^{n}(Z_{(i)} - (2i-1)/(2n))^2 + \frac{1}{12n}$
V	$D^+ + D^-$
U^2	$W^2 - n(\bar{Z} - 0.5)^2$
A	$-n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)(log(Z_{(i)}) + log(1 - Z_{(n+1-i)}))$

TAB. 2.1 – *Tests d'adéquation fondés sur la fonction de répartition théorique*

Pour un test d'adéquation à la loi normale, le test d'Anderson-Darling est le plus puissant suivi de celui de Cramèr-von Mises [15]. Notons enfin que, la statistique de test d'Anderson-Darling peut être modifiée et Stephens [17] la présente ainsi

$$(A - 0.7/n)(1 + 3.6/n - 8/n^2)$$

L'hypothèse nulle de chaque test de normalité peut alors être posée en ces termes :

> H0 : les résidus réduits de chacun des modèles issus des essais multilocaux suivent la loi de Gauss

Les statistiques de tests sont comparées, au seuil de significativité de 5%, à la valeur lue (0.787) sur la table fournie par Stephens.

D'un autre côté, avant de mettre en œuvre ces tests d'ajustement, nous avions utilisé des méthodes empiriques pour avoir une idée sur les formes des distributions. Pour cela, nous pouvions tracer les histogrammes des résidus des modèles. Cependant, le mode de tracé de ces graphiques peut être discutable en ce qui concerne le choix des classes. Aussi, la forme de la courbe de densité de la loi gaussienne n'est pas la seule à être en cloche. C'est le cas des densités des lois de Cauchy et de Student. Dans ces conditions, nous avons préféré la droite de Henry (appelée aussi *qq-plot* pour *quantile-quantile plot*) comme procédure de vérification. Pour chaque essai, les quantiles de l'échantillon des résidus sont représentés en fonction des quantiles théoriques. Si les résidus proviennent d'une variable gaussienne, le nuage de points doit former à peu près une droite.

Chapitre 3

Régression factorielle

3.1 Présentation du modèle

Le but est d'expliquer les observations Y à partir de deux matrices de covariables associées aux deux facteurs étudiés. Nous allons présenté le modèle de régression factorielle de deux manières différentes. Dans la présentation matricielle, la plus simple, nous considérons les observations dans une matrice Y de dimension IxJ avec I lignes (nombre de variétés) et J colonnes (nombre de lieux). Dans la présentation indicielle, la plus générale, les observations Y sont dans un vecteur de longueur IJ.

Nous pouvons disposer en général d'un tableau X de p covariables associées au facteur variété et d'un tableau Z de q covariables associées au facteur lieu. Les matrices X et Z s'écrivent dans ce cas :

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{Ip} \end{pmatrix} \quad Z = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1q} \\ z_{21} & z_{22} & \cdots & z_{2q} \\ \vdots & \vdots & & \vdots \\ z_{J1} & z_{J2} & \cdots & z_{Jq} \end{pmatrix}$$

X est de dimension Ixp et Z de dimension Jxq. Par souci de simplicité dans la présentation, nous ne considérons alors que le cas d'une unique observation

par variété pour une année ou un lieu donné, qui peut être une moyenne ou une moyenne ajustée sur le dispositif de ce lieu.

Alors le modèle 1.2 de la page 6 peut se présenter sous forme matricielle, plus commode à manipuler

$$Y = 1_I m 1'_J + (a_1 + \dot{X}a_2)1'_J + 1_I(\ell'_1 + \ell'_2\dot{Z}') + \dot{X}\Gamma\dot{Z}' + a_3\dot{Z}' + \dot{X}\ell'_3 + Résidus \tag{3.1}$$

- Y est de dimension IxJ, est la réponse de la variable
- m est la moyenne générale
- $(a_1 + \dot{X}a_2)1'_J$ est l'effet principal du facteur variété
- $1_I(\ell'_1 + \ell'_2\dot{Z}')$ est l'effet principal du facteur lieu
- $\dot{X}\Gamma\dot{Z}'$ est la partie de l'interaction expliquée par le produit des deux covariables X et Z
- $a_3\dot{Z}'$ est la partie de l'interaction expliquée par la covariable Z, une fois tenu compte de l'explication fournie par la covariable X
- $\dot{X}\ell'_3$ est la partie de l'interaction expliquée par la covariable X, une fois tenu compte de l'explication fournie par la covariable Z

Ce modèle étant linéaire, les procédures d'estimation usuelles sont employées. Ce qui donne les résultats suivants pour l'estimation des paramètres inconnus [1] :

$$m = 1'_I Y 1_J / IJ$$

$$a_2 = (\dot{X}'\dot{X})^{-1}\dot{X}'Y1_J/J$$

$$\ell'_2 = 1'_I Y\dot{Z}(\dot{Z}'\dot{Z})^{-1}/I$$

$$a_1 = (I_I - \tfrac{1}{I}1_I1'_I - \dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}')\tfrac{1}{J}Y1_J$$

$$\ell'_1 = 1'_I Y(I_J - \tfrac{1}{J}1_J1'_J - \dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}')\tfrac{1}{I}$$

$$\Gamma = (\dot{X}'\dot{X})^{-1}\dot{X}'Y\dot{Z}(\dot{Z}'\dot{Z})^{-1}$$

$$a_3 = (I_I - \tfrac{1}{I}1_I1'_I - \dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}')Y\dot{Z}(\dot{Z}'\dot{Z})^{-1}$$

$$\ell'_3 = (\dot{X}'\dot{X})^{-1}\dot{X}'Y(I_J - \dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}')$$

1. pour le détail des calculs, voir Annexe B

Cependant, cette façon d'écrire le modèle et les calculs qui en découlent sont liés à la structure des données. En effet, dans ce que nous avons présenté, les observations sont supposées être dans une matrice à I lignes et J colonnes, ce qui prévoit cette matrice de données complète. Cela s'interprète dans le cas d'un essai GxE par exemple, par le fait que chaque variété doit être présente le même nombre de fois dans chaque lieu. Or la réalité est souvent autre. Par exemple, dans le cadre d'un programme de sélection par essai pluriannuel, le sélectionneur peut introduire de nouvelles variétés ou retirer certaines au cours du programme, ce qui se traduit sur la matrice des observations par des cases manquantes et donc l'impossibilité de présenter le modèle selon la forme ci-dessus.

Nous avons tenu malgré tout, à écrire d'abord le modèle de régression factorielle de cette façon, parce que nous estimons la présentation matricielle plus simple, même si elle est rarement en adéquation avec des observations d'essais variétaux. Pour les essais où toutes les variétés ne sont pas dans tous les lieux, nous avons gardé la présentation du modèle de la page 6 qui est ici généralisé avec p covariables associées au facteur génotype et q covariables liées au facteur environnement.

$$Y_{ij} = m + a_i + l_j + \sum_{p,q} x_{ijp} z_{jjq} \gamma_{pq} + \sum_{q} z_{ijq} \alpha_q + \sum_{p} x_{ijp} \beta_p + E_{ij} \qquad (3.2)$$

où m, a_i et l_j sont estimés à partir du modèle additif, x_{ip} étant la valeur de la p^{e} covariable associée au facteur variété et z_j^q la valeur de la q^{e} covariable associée au facteur lieu.

Les vecteurs $\gamma = \begin{pmatrix} \gamma_{11} \\ \vdots \\ \gamma_{1q} \\ \vdots \\ \gamma_{p1} \\ \vdots \\ \gamma_{pq} \end{pmatrix}$, $\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_q \end{pmatrix}$ et $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$

sont estimés par régression du vecteur Y des observations, de longueur IJ, respectivement sur le produit semi-tensoriel ligne des deux matrices des covariables, sur la matrice de covariable attachée au facteur lieu et sur la matrice de covariable at-

tachée au facteur variété. Nous définissons ci-après le produit semi-tensoriel ligne et écrivons pour l'estimation des paramètres γ, α et β :

$$Y = X \ominus Z \cdot \gamma + Z \cdot \alpha + X \cdot \beta + Résidus$$

Définition 1 ***Produit semi-tensoriel ligne***

Si A est une matrice rectangulaire (m,r),

Si B est une matrice rectangulaire (m,s),

$A \ominus B$ est dit produit semi-tensoriel ligne de A par B. C'est une matrice rectangulaire (m,rs) qui se présente ainsi

$$\begin{pmatrix} a_{11} \cdot B[1,] & a_{12} \cdot B[1,] & \cdots & a_{1r} \cdot B[1,] \\ a_{21} \cdot B[2,] & a_{22} \cdot B[2,] & \cdots & a_{2r} \cdot B[2,] \\ \vdots & \vdots & & \vdots \\ a_{m1} \cdot B[m,] & a_{m2} \cdot B[m,] & \cdots & a_{mr} \cdot B[m,] \end{pmatrix}$$

où $B[m,]$ est la m^e ligne de B. En clair, $A \ominus B$ est la juxtaposition de tous les produits termes à termes possibles entre une colonne de A et une colonne de B

Nous avons donc présenté le modèle de régression factorielle en deux étapes : d'abord l'estimation des effets principaux du modèle additif auxquels sont ajoutées par la suite la régression des observations sur le produit semi-tensoriel ligne des deux matrices de covariables et celle sur les deux matrices des covariables. Nous présentons ci-dessous un modèle, plus général, où tous les paramètres sont estimables en une seule étape.

$$Y = m1_n + X \cdot a_1 + \mathbb{X} \cdot a_2 + Z \cdot \ell_1 + \mathbb{Z} \cdot \ell_2 + X \ominus Z \cdot (a\ell)_{11} + \mathbb{X} \ominus Z \cdot (a\ell)_{21} + X \ominus \mathbb{Z} \cdot (al)_{12} + \mathbb{X} \ominus \mathbb{Z} \cdot (al)_{22}$$

- Y est le vecteur des observations, de longueur IJ
- $m1_n$ est la moyenne générale des observations, vecteur de même longueur que Y
- Xa_1 est la régression sur les covariables variétés
- $\mathbb{X}a_2$ est l'écart des effets variété à la régression sur les covariables variétés, $\mathbb{X}$ étant la matrice d'incidence des variétés

- $Z\ell_1$ est la régression sur les covariables lieux
- $\mathbb{Z}\ell_2$ est l'écart des effets lieux à la régression sur les covariables lieux, $\mathbb{Z}$ étant la matrice d'incidence des lieux
- $X \ominus Z(a\ell)_{11}$ est l'effet des covariables lieux modulé par les covariables variété
- $\mathbb{X} \ominus Z(a\ell)_{21}$ est l'effet des covariables lieux modulé par les variétés non expliqué par les covariables variétés
- $X \ominus \mathbb{Z}(al)_{12}$ est l'effet des covariables variétés modulé par les lieux non expliqué par les covariables lieux
- $\mathbb{X} \ominus \mathbb{Z}(al)_{22}$ est l'interaction GxE expliquée ni par les covariables variété, ni par les covariables lieux.

Ainsi, les paramètres a_1, a_2, ℓ_1, ℓ_2, $(a\ell)_{11}$, $(a\ell)_{21}$, $(a\ell)_{12}$, $(a\ell)_{22}$, sont estimés par simple régression linéaire sur les matrices et vecteurs adéquats.

Par exemple, dans un essai multilocal de 3 variétés effectué en 2 lieux différents où nous avons 2 covariables associées au facteur lieu et 1 covariable associée au facteur variété, les vecteurs et matrices se présentent comme il suit :

$$Y = \begin{pmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{12} \\ Y_{22} \\ Y_{32} \end{pmatrix} \quad X = \begin{pmatrix} X_{111} \\ X_{211} \\ X_{311} \\ X_{121} \\ X_{221} \\ X_{321} \end{pmatrix} \quad Z = \begin{pmatrix} Z_{111} & Z_{112} \\ Z_{121} & Z_{122} \\ Z_{211} & Z_{212} \\ Z_{221} & Z_{222} \\ Z_{311} & Z_{312} \\ Z_{321} & Z_{322} \end{pmatrix}$$

$$\mathbb{X} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbb{Z} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$$X \ominus Z = \begin{pmatrix} X_{111}Z_{111} & X_{111}Z_{112} \\ X_{211}Z_{121} & X_{211}Z_{122} \\ X_{311}Z_{211} & X_{311}Z_{212} \\ X_{121}Z_{221} & X_{121}Z_{222} \\ X_{221}Z_{311} & X_{221}Z_{312} \\ X_{321}Z_{321} & X_{321}Z_{322} \end{pmatrix}$$

$$\mathbb{X} \ominus Z = \begin{pmatrix} Z_{111} & Z_{112} & 0 & 0 & 0 & 0 \\ 0 & 0 & Z_{121} & Z_{122} & 0 & 0 \\ 0 & 0 & 0 & 0 & Z_{211} & Z_{212} \\ Z_{221} & Z_{222} & 0 & 0 & 0 & 0 \\ 0 & 0 & Z_{311} & Z_{312} & 0 & 0 \\ 0 & 0 & 0 & 0 & Z_{321} & Z_{322} \end{pmatrix} \quad X \ominus \mathbb{Z} \begin{pmatrix} X_{111} & 0 \\ X_{211} & 0 \\ X_{311} & 0 \\ 0 & X_{121} \\ 0 & X_{221} \\ 0 & X_{321} \end{pmatrix}$$

$$\mathbb{X} \ominus \mathbb{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Jusqu'à maintenant les covariables sont supposées continues. Toutefois, cette méthode est encore valable dans le cas où elles sont de type qualitatif. Il suffit de la même manière que l'on passe de la régression à l'analyse de variance, de remplacer chaque colonne de la matrice de covariables par des colonnes indicatrices des niveaux de la colonne de covariable qualitative considérée [5].

Aussi, différentes méthodes pour la selection de covariables dans le cas où elles sont trop nombreuses sont présentées dans Denis [5].

3.2 Description des covariables

Dans notre étude, nous disposons de 3 covariables pour le facteur lieu et d'une seule covariable pour le facteur variété.

3.2.1 Covariables associées au facteur lieu

Comme nous l'avons annoncé plus haut, des variables climatiques sont mesurées quotidiennement sur les essais. Ces variables (pluie, rayonnement solaire, vitesse du vent, etc.) déterminent l'évapotranspiration et se trouvent en fait plus nombreuses que les essais. Ce qui rend leur utilisation dans un modèle de régression factorielle impossible. Ainsi, nous avons utilisé un modèle de bilan hydrique, qui permet de prendre en considération l'action du climat sur les cultures et de resumer ces nombreuses variables. Les modèles de bilan hydrique reposent généralement sur une équation de conservation de la masse d'eau reçue sous forme de pluie évapotranspirée et stockée dans le sol :

$$\text{Pluie}=\text{ETR}+\Delta\ \text{Stock}$$

L'ETR étant fonction de l'état du stock et du développement de la plante, lui même fonction de son état de stress antérieur, il est nécessaire de calculer pas à pas les variations de stock et d'ETR.

$$ETR = f_1(stock/stock\ max,\ stade\ de\ développement)$$

$$Rendement = f_2(\tfrac{ETR}{ETM}\ (k),\ k = 1\ ,\ \ldots\ ,\ longueur\ de\ cycle)$$

Les fonctions f_1 et f_2 sont paramétrées par des contraintes θ qui dépendent de l'espèce et de la variété. Ceci est un problème pour l'utilisation directe d'un tel modèle. En effet, l'estimation des paramètres pour une nouvelle variété est coûteuse car elle demande un expérimentation spécifique.

Notre hypothèse est que les paramètres ne varient qu'assez peu d'une variété à une autre, si toutefois elles ont même longueur de cycle, ce qui est le cas dans tous les essais.

Si tel est le cas, la variation de l'ETR d'une variété à une autre peut être développée en série de Taylor à l'ordre 1,au voisinage du vecteur de paramètres θ_0 d'une variété de référence

$$Y_{ij} = f_2(Z_j, \theta_i) = f_2(Z_j, \theta_0) + \sum_{k=1}^{K} \left(\frac{\partial f_2}{\partial \theta_k}\right)_{\theta=\theta_0} (\theta_{i(k)} - \theta_{0(k)})$$

où $\theta_{(k)}$ désigne la k^e composante du vecteur θ, Z_j le vecteur des variables climatiques du milieu j et θ_j le vecteur des paramètres de la variété i.

Posons $g(j) = \frac{df_2(Z_j, \theta_0)}{d\theta_k}$, c'est une fonction de l'environnement j ; $h(i) = d\theta_k$ est elle une fonction de la variété i. Le produit de $g(j)$ et de $h(i)$ donne les interactions GxE.

En première approximation, nous pouvons donc modéliser toute observation sur une nouvelle variété non calibrée, c'est à dire dont on ne connaît pas les paramètres, comme la même observation sur une variété voisine dans le même environnement plus un terme d'interaction GxE.

Aussi, nous avons simulé une variété de référence dont les paramètres étaient connus. La fonction g est inconnue mais nous avons utilisé les taux de satisfaction des besoins en eau aux différentes phases du cycle végétatif. Ordinairement, ces taux de satisfaction servent de prédicteurs pour le rendement dans des régressions.

3.2.2 Covariable associée au facteur variété

Pour cette covariable, nous pouvons utiliser des indices de tolérance ou de sensibilité qui font intervenir le rendement en conditions de sécheresse et le rendement en conditions optimales. Mais le mode de calcul de ces indices pénalise très fortement les génotypes à rendement potentiel élevé ; des valeurs très élevées caractérisent plutôt des génotypes rustiques à faible productivité que des génotypes à rendement stable et élevé [1]. Nous privilégions le taux de croissance de culture (C) décrit par Turner [20] comme étant :

$$C=(PF+(1.65 \cdot PG))/T$$

où PF est le poids des fanes, PG le poids des gousses et T la durée floraison-récolte. Ce taux permet effectivement de caractériser les variétés selon leur sensibilité à la sécheresse car plus une variété produit de biomasse totale et de gousses en un temps relativement court, mieux elle est adaptée à la sécheresse. Le calcul de C se fait sur les données des essais de la station de Bambey qui ne servent pas à la modélisation de l'interaction GxE : ce sont des estimations indépendantes de la variable à modéliser, bien qu'elles soient calculées sur des données de production. Alors, pour pouvoir prédire ultérieurement le rendement d'une variété pour un nouvel environnement, il sera nécessaire de disposer d'informations sur la biomasse, le poids des gousses et la durée floraison-récolte issues d'essais différents.

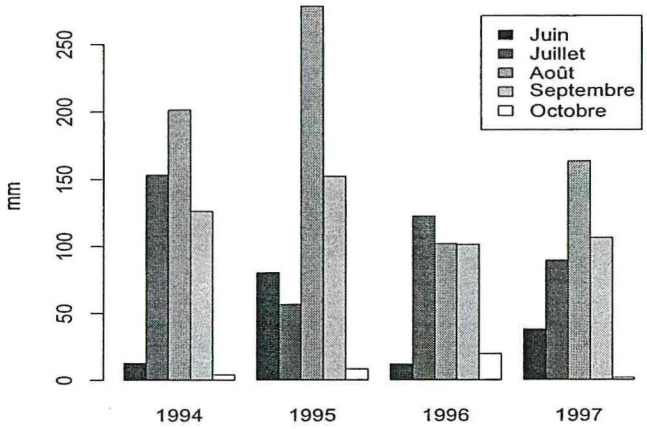
Chapitre 4

Vérification des données et des tests des interactions

La grande variabilité spatio-temporelle des pluies au Sénégal suggère à les intégrer dans la modélisation de l'interaction GxE (les environnements étant souvent différents par leurs niveaux de pluviométrie) pour affiner la recherche de variétés à cycle plus court et de variétés physiologiquement adaptées à la sécheresse. Cette grande variabilité concerne aussi la station de Bambey d'une année à une autre (figure 4.1). Ce qui justifie l'utilisation des données des essais pluriannuels dans cette modélisation. Pour l'analyse, le modèle reste le même et le facteur "année" joue le même rôle que le facteur "environnement".

L'indépendance des erreurs résiduelles supposée pour l'analyse de variance, se trouve en fait difficile à verifier, une fois les données collectées. En effet, cette indépendance est assurée par la randomisation qui se fait au moment de la mise en place des essais. Deux cas peuvent se présenter : soit une randomisation indépendante de chaque essai dans le cas d'essais multilocaux est effectuée et l'indépendance des erreurs en découle, soit la randomisation n'a pas été réalisée et alors, l'analyse de variance n'est pas la méthode adaptée pour étudier les données. Pour notre étude, la responsable des essais, D. Clavel, a fourni l'assurance d'une randomisation correcte des essais et nous n'avons relevé aucune systématisation de l'affectation des variétés aux unités expérimentales.

FIG. 4.1 – *Répartition des pluies sur la station de Bambey de 1994 à 1997*



4.1 Résultats des essais multilocaux

Sur la figure 4.2 sont présentés les graphiques des résidus des modèles des essais multilocaux de 1996 à 1998 en fonction des valeurs prédites.

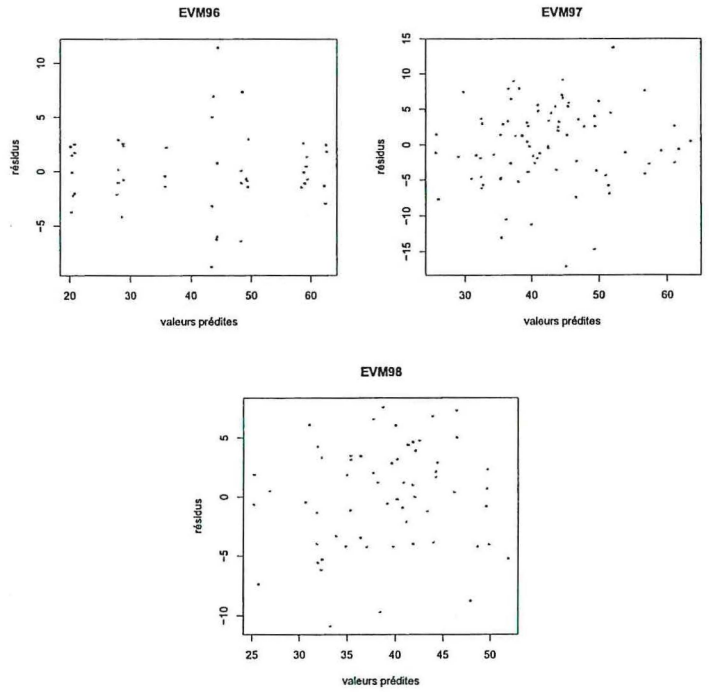
L'examen de ces trois graphiques semble montrer des variances résiduelles constantes pour EVM97 et pour EVM98 tandis que celles de EVM96 ne paraissent pas homogènes.

Pour les tests d'égalité des variances par rapport de vraisemblance, le tableau 4.1 donne les niveaux de signification des différents essais.

Essai	Niveau de signification
EVM96	**0.033**
EVM97	0.203
EVM98	0.977

TAB. 4.1 – *Tests de rapport de vraisemblance maximale sur modèles emboîtés pour l'homogénéité de variances résiduelles d'essais multilocaux au Sénégal de 1996 à 1998*

FIG. 4.2 – *Résidus des modèles des essais multilocaux de 1996 à 1998 en fonction des valeurs prédites*

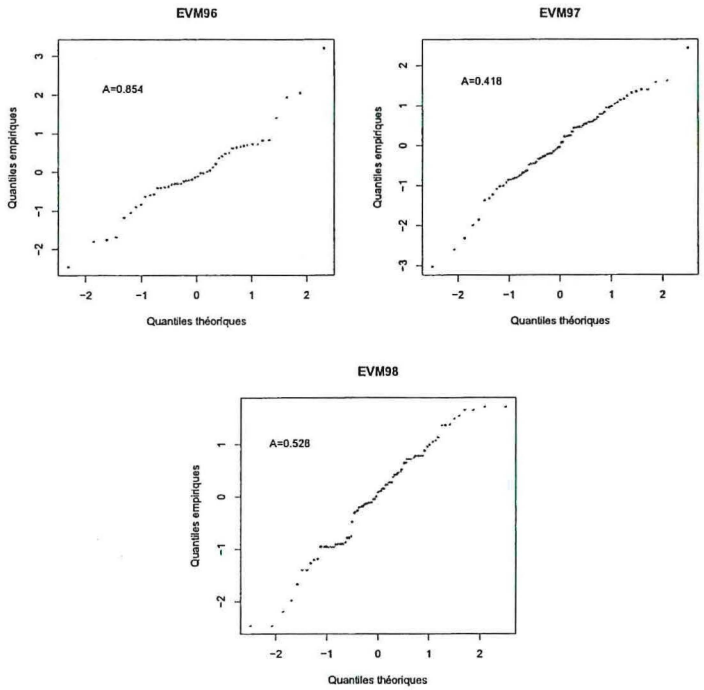


Ainsi, pour EVM96 où il semblait exister une relation entre les résidus et les valeurs prédites, nous rejetons l'hypothèse selon laquelle les variances sont homogènes. Pour EVM97 et EVM98, nous ne pouvons pas rejeter cette hypothèse.

Les droites de Henry des résidus des modèles des essais multilocaux sont présentées à la figure 4.3. Les résidus des modèles de EVM97 et de EVM98 semblent être tirés d'un échantillon gaussien alors que EVM96 montre une distribution plus aplatie, vraisemblablement le résultat du mélange de plusieurs distributions gaussiennes de variances différentes ; A est la statistique de test d'Anderson-Darling.

D'après les tests d'Anderson-Darling, nous refusons l'hypothèse de la normalité pour EVM96 ($A > 0.787$) ; pour EVM97 et EVM98, nous ne pouvons pas refuser cette hypothèse. Cependant, la robustesse des tests F couplée à une taille importante des données font que l'analyse de variance reste adaptée au modèle de EVM96.

FIG. 4.3 – *Droite de Henry des résidus et statistiques de tests d'Anderson-Darling pour les données des essais multilocaux*



Une fois vérifiées toutes les hypothèses du modèle, il reste à voir les effets des interactions. Pour EVM97 et EVM98, les tests des interactions sont faits normalement puisque la présupposition d'égalité des variances a été validée. Par contre, pour EVM96, il a fallu tenir compte dans le calcul de la statistique de test de l'effet de l'interaction GxE, de l'hétérogénéité des variances résiduelles des différents lieux. Puisque les variances résiduelles sont différentes d'un lieu à un autre, il a été nécessaire de pondérer les données par l'inverse de la variance résiduelle estimée séparément dans chaque lieu. Cela a été fait avec la procédure mixed d Sas avec l'instruction repeated/group=lieu. Le tableau 4.2 donne les niveaux de signification des tests des interactions GxE.

Seul le modèle issu de EVM96 présente un effet des interactions GxE significatif. Pour pouvoir prédire alors efficacement les rendements des variétés, il faut étudier

Essai	Niveau de signification
EVM96	**0.0001**
EVM97	0.3828
EVM98	0.5655

TAB. 4.2 – *Tests des effets des interactions d'essais multilocaux au Sénégal de 1996 à 1998*

plus en détail ces interactions qui correspondent aux résidus du modèle additif, c'est à dire du modèle sans interaction.

4.2 Résultats des essais pluriannuels

Sur la figure 4.4 sont représentés les graphiques des résidus des modèles des essais pluriannuels en fonction des valeurs prédites.

Si pour chacun des essais 80J1, 80J2, 80J3, 90J1 et 90J3, le nuage de points ne présente aucune structure particulière, il n'en est pas de même pour les essais 80J4 et 90J2. En effet, ces essais semblent avoir des variances hétérogènes.

Pour l'égalité des variances résiduelles, nous avons présenté dans le tableau 4.3 les niveaux de signification des tests de rapport de vraisemblance pour les essais pluriannuels.

Essai	Niveau de signification
80J1	0.212
80J2	0.659
80J3	0.179
80J4	**0.003**
90J1	**0.008**
90J2	**0.000**
90J3	0.438

TAB. 4.3 – *Tests de rapport de vraisemblance sur modèles emboîtés pour des essais d'arachide à Bambey de 1994 à 1997*

Les tests qui ont un niveau de signification inférieur à 0.05 (80J4, 90J1 et 90J2) sont significatifs c'est à dire, nous pouvons conclure à la non égalité des variances.

Essai	Niveau de signification
80J1	**0.0006**
80J2	**0.0473**
80J3	0.1612
80J4	**0.0142**
90J1	0.8829
90J2	0.2648
90J3	**0.0252**

TAB. 4.4 – *Tests des effets des interactions pour des essais d'arachide à Bambey de 1994 à 1997*

Pour les autres, nous ne pouvons pas rejeter l'hypothèse d'homogénéité de la variance. Il faut alors pondérer les données par les variances résiduelles des lieux.

Pour la présupposition de la normalité, les droites de Henry des résidus des modèles des EVP sont présentés à la figure 4.5 ; A étant la statistique du test d'Anderson-Darling.

Ainsi, pour les modèles des essais 80J1, 80J3, 80J4 et 90J3 nous ne pouvons pas rejeter l'hypothèse selon laquelle les résidus suivent la loi normale (la statistique de test est inférieure à la valeur lue c'est à dire à 0.787). Pour les autres essais, l'hypothèse de normalité est rejetée. En comptant sur le nombre des données, nous pouvons néanmoins tester les effets des interactions par analyse de variance tout en sachant que les niveaux de signification seront imprécis [3].

Avec ces considérations, les tests des effets des interactions pour les différents modèles sont présentés au tableau 4.4. Les effets des interactions des essais 80J4, 90J1 et 90J2 ont été estimés en pondérant les données de tests par les variances résiduelles des différentes années.

Ainsi, pour les essais 80J1, 80J2, 80J4 et 90J3, les modèles d'analyse de variance ne sont pas additifs. Pour ces essais, la réponse de la variable, en l'occurrence le rendement, sera modélisée comme la somme des effets du facteur variété, du facteur lieu et des interactions.

Pour la suite, l'objectif de l'étude est de tenter d'expliquer les interactions. Cette analyse concerne uniquement les essais pluriannuels 80J1, 80J2, 80J4 et 90J3 et l'essai multilocal EVM96 en l'absence d'interaction sur les autres essais.

FIG. 4.4 – *Résidus des modèles des essais pluriannuels de 1996 à 1998 en fonction des valeurs prédites*

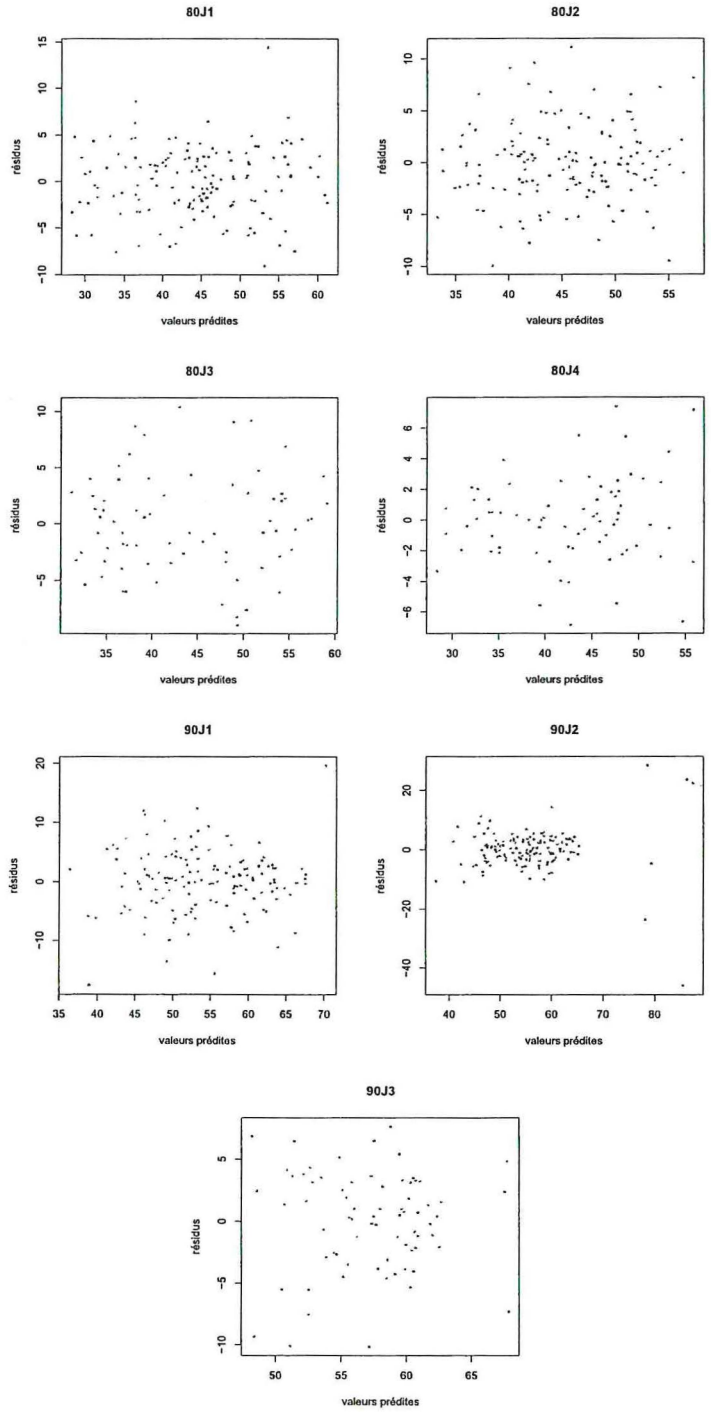
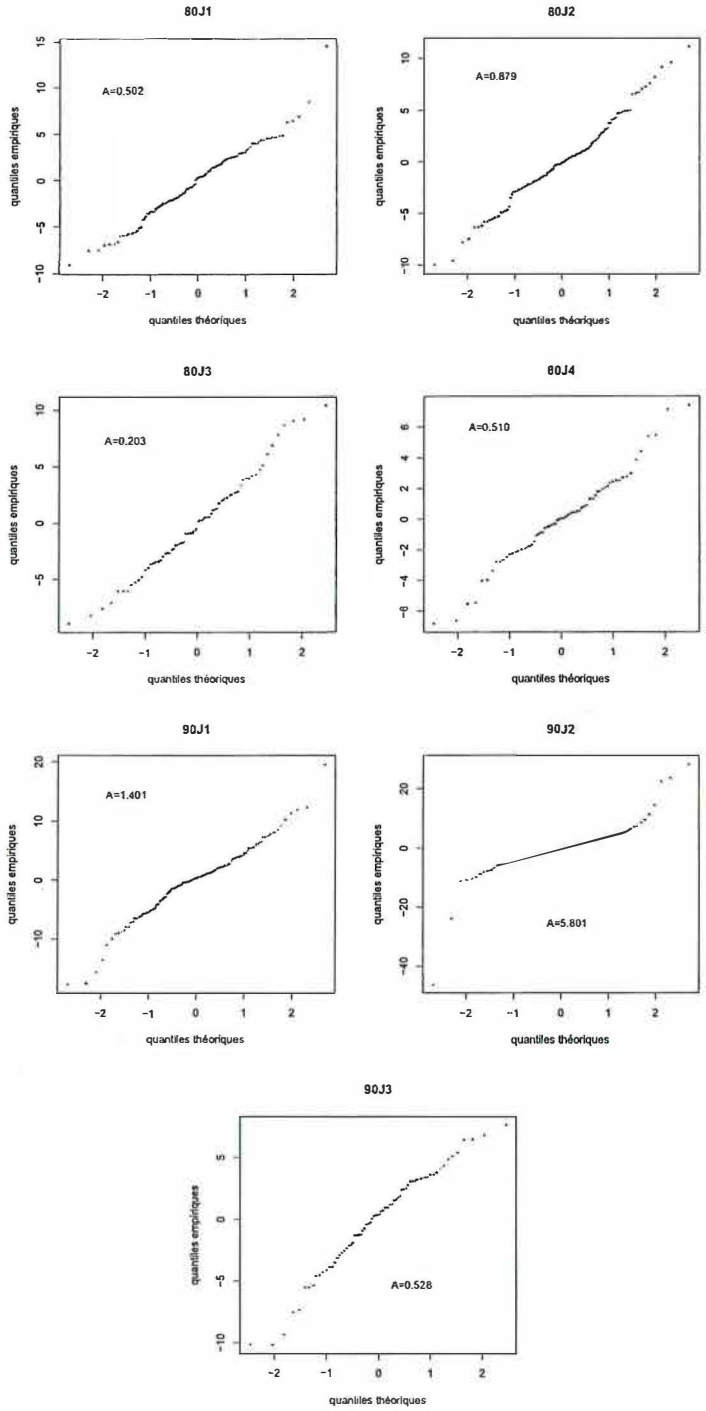


FIG. 4.5 – *Droite de Henry des résidus et statistiques de tests d'Anderson-Darling pour les données des essais pluriannuels*



Chapitre 5

Résultats des estimations des modèles de régression factorielle

Seuls les modèles des essais qui présentent une interaction GxE significative sont étudiés : EVM96 pour les essais multilocaux, 80J1, 80J2, 80J4 et 90J3 pour les essais pluriannuels. Il s'agit alors d'expliquer l'interaction, pour chaque modèle, par régression factorielle, à l'aide de covariables liées aux facteurs. Pour le facteur variété, la covariable utilisée est le taux de croissance de culture et pour le facteur lieu ou année, nous avons cinq taux de satisfaction des besoins en eau calculés sur les cinq phases de développement de l'arachide : phase de semis-levée,phase végétative, phase reproductive, phase de maturité 1 et phase de maturité 2. Les taux de satisfaction sont estimés par simulation à l'aide de SarraH qui utilise des variables climatiques et pluviométriques en entrée.

Nous n'étudierons pas en détail les résidus de la régression factorielle comme nous l'avons fait pour ceux des modèles d'analyses de variance car en régression factorielle, les résidus n'ont pas de particularité caractéristique [7]. Cependant, nous tracerons des boîtes à moustaches (*box-plot* en anglais) pour les résidus pour regarder leur variation.

5.1 Essai multilocal de 1996

Cet essai fut conduit sur 3 localités : Pakhi Kébé, Maka Fall et Darou Sam. Cependant, pour ces localités, nous ne disposons pas de données de pluies et de climat. Une idée aurait été de regarder les données des postes pluviométriques et climatiques les plus proches et attribuer ses valeurs aux localités. Mais dans cette zone à forte variation surtout pluviométrique, cette approche n'est pas pertinente. Car il serait erroné de prédire la réponse des variétés d'une zone par des données de pluies d'une autre du fait que les pluies sont fortement variables dans l'espace [12]. En effet, rappelons que l'efficacité d'un modèle de régression est principalement liée à la pertinence des variables explicatives. Dans ces conditions, nous n'avons pas prédit le rendement des variétés dans ce modèle (EVM96) faute de covariables liées au facteur lieu appropriées.

5.2 Essais pluriannuels

Ces essais ont été conduits sur 4 années pour 80J1 et 80J2 et sur 2 années pour 80J4 et 90J3. Comme il y a 5 covariables liées au facteur année, nous avons choisi la phase de maturité 1, la plus sensible du cycle végétatif de l'arachide, et calculé le taux de satisfaction des besoins en eau correspondant. Ce qui donne pour nos modèles de régression factorielle une seule covariable liée au facteur année et une seule associée au facteur variété.

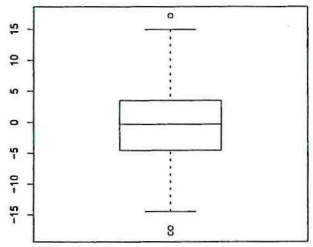
5.2.1 Essai pluriannuel 80J1

Pour cet essai, il y avait eu 25 variétés. Nous pouvons adopter le modèle suivant

$$Y_{ij} = m + a_i + l_j + 1.4\, x_{ij1} z_{ij4} + \mathbb{X} z_{ij4} + x_{ij1} \mathbb{Z} + \mathbb{X}\mathbb{Z} \tag{5.1}$$

où m, a_i et l_j sont les estimations habituelles de la moyenne générale et des effets principaux ; x_1 est le taux de croissance de culture ; z_4 est le taux de satisfaction

FIG. 5.1 – *Boîte à moustaches des résidus du modèle de 80J1*



des besoins en eau de la quatrième phase de développement de l'arachide. $\mathbb{X}$ est la matrice d'incidence des variétés et $\mathbb{X}$ est la matrice d'incidence des années.

Dependent Variable: rendement

Source	DF	Type I SS	Mean Square	F Value	Pr > F
variete	19	3382.632278	178.033278	5.99	<.0001
an	3	3563.368385	1187.789462	39.97	<.0001
c*tsat4	1	251.148272	251.148272	8.45	0.0045
c*an	3	212.983762	70.994587	2.39	0.0732
tsat4*variete	10	643.475166	64.347517	2.17	0.0259
variete*an	6	355.300036	59.216673	1.99	0.0735

La somme des carrés variete*an est tout ce qui reste de l'interaction une fois que nous avons utilisé les covariables liées aux facteurs pour l'interpréter. Cet effet n'étant pas significatif au seuil de 5% (0.0735), nous pouvons donc penser que l'interaction a été réduite par nos covariables jusqu'à ne plus être significative.

L'examen de la figure 5.1, pour les résidus, révèle une médiane autour de 0 et trois valeurs déviantes.

5.2.2 Essai pluriannuel 80J2

Il y avait 32 variétés pour cet essai. Le modèle suivant peut être postulé

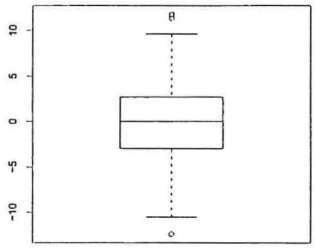
$$Y_{ij} = m + a_i + l_j - 0.1\, x_{ij1} z_{ij4} + \mathbb{X} z_{ij4} + x_{ij1} \mathbb{Z} + \mathbb{X}\mathbb{Z} \tag{5.2}$$

Dependent Variable: rendement

Source	DF	Type I SS	Mean Square	F Value	Pr > F
variete	31	3365.474583	108.563696	4.17	<.0001
an	3	624.692333	208.230778	8.00	<.0001
c*tsat4	1	32.463538	32.463538	1.25	0.2669
c*an	2	303.394905	151.697453	5.83	0.0041
tsat4*variete	8	181.908560	22.738570	0.87	0.5417
variete*an	3	69.956080	23.318693	0.90	0.4464

L'interaction variete*an, une fois que nous avons utilisé les covariables, n'est plus significatif. En effet, cette interaction est interprétée, pour une grande partie, par le taux de satisfaction des besoins en eau de la quatrième phase de développement végétatif de l'arachide calculé pour chaque année et par le taux de croissance de culture calculée pour chaque variété.

FIG. 5.2 – *Boîte à moustaches des résidus du modèle de 80J2*



La figure 5.2 présente les mêmes tendances que 80J1 avec cependant un intervalle interquartile plus réduit.

5.2.3 Essai pluriannuel 80J4

Nous pouvons retenir comme modèle pour 80J4

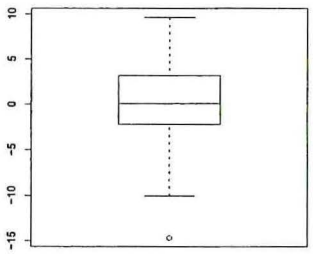
$$Y_{ij} = m + a_i + l_j - 15.8\, x_{ij1} z_{ij4} + \mathbb{X} z_{ij4} + x_{ij1} \mathbb{Z} + \mathbb{X}\mathbb{Z} \tag{5.3}$$

Dependent Variable: rendement

Source	DF	Type I SS	Mean Square	F Value	Pr > F
variete	12	1816.373611	151.364468	6.42	<.0001
an	1	1058.801856	1058.801856	44.90	<.0001
c*tsat3	1	131.983682	131.983682	5.60	0.0221
c*an	0	0.000000	.	.	.
tsat3*variete	9	269.320712	29.924524	1.27	0.2781
variete*an	0	0.000000	.	.	.

Nous n'avons eu que deux années pour ce reseau d'essais. Le seul ddl qui existe pour le facteur année est consommé par celui-ci. Donc il est impossible d'estimer les effets c*an et variété*an, une fois tenu compte des covariables.

FIG. 5.3 – *Boîte à moustaches des résidus du modèle de 80J4*



La figure 5.3 montre la variation des résidus de ce modèle avec un point déviant autour de -15.

5.2.4 Essai pluriannuel 90J3

Nous pouvons opter pour le modèle suivant pour 90J3

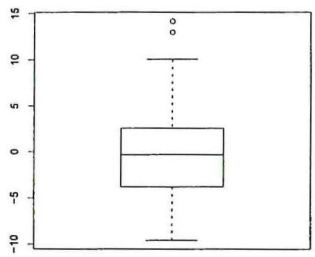
$$Y_{ij} = m + a_i + l_j + 14.81\ \dot{x}_i^1 \dot{z}_i^1 + \mathbb{X} z_{ij4} + x_{ij1}\mathbb{Z} + \mathbb{X}\mathbb{Z} \tag{5.4}$$

Dependent Variable: rendement

Source	DF	Type I SS	Mean Square	F Value	Pr > F
variete	14	387.7437500	27.6959821	1.14	0.3506
an	1	450.0890741	450.0890741	18.52	<.0001
c*tsat3	1	50.2322524	50.2322524	2.07	0.1570
c*an	0	0.0000000	.	.	.
tsat3*variete	7	473.9570068	67.7081438	2.79	0.0163
variete*an	0	0.0000000	.	.	.

De même pour 80J4, il n'y a eu que deux années pour cet essai.

FIG. 5.4 – *Boîte à moustaches des résidus du modèle de 90J3*



La figure 5.4 donne la variation des résidus du modèle de 90J3.

Chapitre 6

Conclusion et Perspectives

Nous voulions modéliser au mieux les interactions GxE éventuellement constatées sur des essais variétaux multilocaux sur des sites au Sénégal et sur des essais variétaux pluriannuels sur Bambey, station expérimentale du Ceraas.

Nous disposions à cet effet de longues séries de données climatiques. Ces données relevées sur les sites d'essais nous ont permis d'estimer en partie les interactions. Cette approche ayant été préférée à celle de les estimer à partir des observations vu le caractère très variable du facteur milieu d'une année à l'autre. Ceci, d'autant plus généralement pour une variété, il n'existe pratiquement pas d'essais sur plusieurs années. Nous avons aussi calculé des indices de tolérance à la sécheresse pour chaque variété sur des essais indépendants. Ces indices sont fonction du poids de gousses, du poids de fanes et de la durée floraison-récolte et permettent de caractériser les variétés selon leur sensibilité à la sécheresse.

Une synthèse bibliographique sur les interactions GxE a été faite et nous avons choisi le modèle de régression factorielle. Ce modèle a permis, dans notre cas, de prédire le rendement des variétés des essais à interactions GxE significatives à l'aide de covariables associées aux deux facteurs. Celles liées au facteur environnement (ou année) sont calculées à partir des données climatiques avec un modèle de simulation de bilan hydrique (SarraH). La covariable de sensibilité à la sécheresse des variétés étant associée au facteur génotype.

Mais avant cela, pour chaque modèle de chaque essai, nous avons testé si l'effet de l'interaction GxE était significatif. Ainsi, EVM96 pour les essais multilocaux

et 80J1, 80J2, 80J4 et 90J3 pour les essais pluriannuels, avaient des interactions GxE significatives. Mais la prédiction par régression factorielle n'a concerné que les essais pluriannuels à Bambey en l'absence de données climatiques sur les sites de l'essai EVM96.

La prédiction du rendement des variétés passe par la modélisation des interactions GxE à l'aide de covariables associées aux facteurs génotype et environnement. Mais, comme nous l'avions annoncé, un modèle n'est pertinent que par ses régresseurs. Dans notre parti-pris de vouloir expliquer des variables de production telles que le rendement par des données climatiques relevées sur les champs d'essais, une absence de celles-ci rend l'exercice caduc. Ainsi, si nous voulons ultérieurement prédire la réponse d'une variété dans un nouveau lieu, devrons-nous veiller à collecter les données climatiques sur le lieu.

Références citées

[1] Belhassen E., This D. et Monneveux P. (1995) L'adaptation génétique face aux contraintes de sécheresse. *Cahiers agricultures*, vol. 4, pp. 251-261.

[2] Bergonzini J. C. 1989. Analyse des données sur tableau d'interaction. Regroupements d'essais et essais en milieu paysan, méthodologie et traitements statistiques. Séminaire, Montpellier du 31 août eu 1er septembre 1989. Délégation Biométrie, Cirad.

[3] Cochran W. G. et Cox G. M. 1957. Experimental Designs. Wiley, New-York, 2e éd.

[4] Crossa J. xxxx. Statistical analyses of multilocation trials. *Advances in agronomy*, vol. 44, pp. 55-85.

[5] Denis J. B. 1980. Analyse de régression factorielle. *Biométrie praximétrie*, vol. 1 pp. 1-34.

[6] Denis J. B. et Vincourt P. 1982. Panorama des méthodes statistiques d'analyse des interactions génotype x milieu. *Agronomie*, vol. 2, pp. 219-230.

[7] Denis J. B. 1992. Analyse statistique de l'interaction entre deux facteurs. Support de cours. Laboratoire de biométrie, Inra.

[8] Eberhart S. A. and Russel W. A. 1966. Stability parameters for comparing varieties. *Crop Science*, vol. 6, pp. 36-40.

[9] Finlay K. W. and Wilkinson G. N. 1963. The analysis of adaptation in plant-breeding programme. *Aust. J. Agric Res.*, vol. 14, pp. 742-754.

[10] Freeman G. H. and Perkins J. M. 1971. Environmental and genotype-environmental components of variability. VIII. Relations between genotypes grown in different environments and measures of these environments. *Heredity*, vol. 27, pp. 15-23.

[11] Gollob H. F. 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*, vol. 33, pp. 73-116.

[12] Lebel T., Amani A, Taupin J. D. 1996 . La pluie au Sahel : une variable rebelle à la régionalisation. IN : Guillot B. - Problèmes de validation des méthodes d'estimation des précipitations par satellite en Afrique intertropicale ORSTOM, Paris , pp. 103-119.

[13] Mandel J. 1961. Non-additivity in two-way analysis of variance. *J. Am. Stat. Assoc.*, vol. 56, pp. 878-888.

[14] Mandel J. 1971. A new analysis of variance model for non additive data. *Technometrics*, vol. 13, pp. 1-18.

[15] Pearson, E. S. and Hartley, H. O. (1976) Tables for statisticians. Biometrika. Charles Griffin and Company LTD for the Biometrika trustees. pp. 117-121.

[16] Phillipeau G. 1989. Théorie des plans d'expérience, application à l'agronomie. Institut technique des céréales et des fourrages, Paris.

[17] Stephens M. A. 1974. EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, vol.69, pp. 730-737.

[18] Tomassone R., Audrain S., Lesquoy de Turkheim E., et Millier C. 1992. La régression, nouveaux regards sur une ancienne méthode statistique. Inra et Masson, Paris, ISBN 2-225-82696-X.

[19] Tukey J. W. 1949. One degree of freedom for non additivity. *Biometrics*, vol. 5, pp. 232-242.

[20] Turner N. C., Wright G. C. and Siddique K. H. M. 2002. Adaptation of grain legumes (pulses) to water-limited environments. *Advances in agronomy*, vol. 71, pp. 193-231.

[21] Williams E. J. 1952. The interpretation of interactions in factorial experiments. *Biometrika*, vol. 39 pp. 65-81.

[22] Wood J. T. 1976. The use of environmental variables in the interpretation of genotype-environment interaction. *Heredity*, vol. 37, pp. 1-7.

[23] Yates F. and Cohran W. G. 1938. The analysis of groups of experiments. *J. Agric. Sci. Camb.*, vol. 28, pp. 556-580.

Annexe A : Liste des variétés

Pour les EVP, 100 variétés ont été utilisées au total. Il y avait 7 essais disposés en lattices rectangulaires, chaque essai étant conduit sur 3 ans (1994 à 1997). Le tableau 1 présente ces variétés

45-437	55-11	55-112	55-113	55-114
55-115	55-116	55-118	55-121	55-122
55-123	55-128	55-131	55-132	55-135
55-138	55-139	55-14	55-140	55-142
55-147	55-148	55-15	55-16	55-17
55-21	55-210	55-211	55-212	55-213
55-214	55-215	55-22	55-24	55-26
55-27	55-28	55-29	55-31/21-1	55-32/21-13
55-33/33-15	55-33/38-4	55-34/34-4	55-35/36-3	55-36/36-10
55-37/38-2	55-38/38-3	55-437	57-111	57-115
57-120	57-123	57-125	57-126	57-14
73-34	73-35	73-39	73-41/1-7	73-42/3-8
73-43/4-10	78-936	AHK-85-19	AHK-85-3	CHICO
FLEUR11	GC-8-13	GC-8-31a	GC-8-35	GC-8-36
GC-8-37	GC-8-38	ICGS-11	Pi-165-317	S-34
S-43	S-45	S-46	SR-1-1	SR-1-11
SR-1-12	SR-1-2	SR-1-22	SR-1-23	SR-1-27
SR-1-30	SR-1-31	SR-1-4	SR-1-6	SR-1-9
SRV-1-15/38-4	SRV-1-18/19-1	SRV-1-19/31-3	SRV-1-3/24-1	SRV-1-38/1-2
SRV-1-42/14-11	SRV-1-45/23-11	SRV-1-46/20-12	SRV-1-47/24-9	US-83

TAB. 1 – *Variétés testées dans les EVP à Bambey entre 1994 et 1997*

Pour les EVM, il y avait 5 variétés : 55-114, 55-138, 55-437, FLEUR11 et GC-8-35. Elles étaient testées en 1996 sur 3 localités et en 1997 et 1998 sur 4 localités.

Annexe B : Modèle de Régression factorielle

Présentation matricielle

On dispose de trois matrices : X de dimension Ixp est la matrice des covariables associées au premier facteur (facteur variété par exemple), Z de dimension Jxq est la matrice des covariables associées au deuxième facteur (facteur lieu) et Y de dimension IxJ est la matrice des observations. X, Y et Z s'écrivent ainsi :

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{Ip} \end{pmatrix} \quad Z = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1q} \\ z_{21} & z_{22} & \cdots & z_{2q} \\ \vdots & \vdots & & \vdots \\ z_{J1} & z_{J2} & \cdots & z_{Jq} \end{pmatrix}$$

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1J} \\ y_{21} & y_{22} & \cdots & y_{2J} \\ \vdots & \vdots & & \vdots \\ y_{I1} & y_{I2} & \cdots & y_{IJ} \end{pmatrix}$$

L'idée est d'utiliser les deux matrices de covariables X et Z attachées aux deux facteurs variété et environnement pour expliquer la matrice Y.

Pour cela, une première régression de Y sur X est faite :

$Y = X\alpha \quad \Rightarrow \quad \hat{Y}_X = X(X'X)^{-1}X'Y$

et $\quad R_X = Y - \hat{Y}_X = (I_I - X(X'X)^{-1}X')Y$

avec I_I représentant la matrice identité d'ordre I et R_X la matrice des résidus de la régression de Y sur X qui est de dimension IxJ.

Une deuxième régression de $\hat{Y}'_X$ sur Z est faite.

$\hat{Y}'_X = Z\beta \quad \Rightarrow \quad \hat{Y}'_{XZ} = Z(Z'Z)^{-1}Z'\hat{Y}'_X$

Ainsi, $\hat{Y}_{XZ} = \hat{Y}_X Z(Z'Z)^{-1}Z' = X(X'X)^{-1}X'YZ(Z'Z)^{-1}Z'$

D'où $R_{\hat{Y}_X} = \hat{Y}_X - \hat{Y}_{XZ} = X(X'X)^{-1}X'Y - X(X'X)^{-1}X'YZ'(ZZ')^{-1}Z$

$\Rightarrow \quad R_{\hat{Y}_X} = X(X'X)^{-1}X'Y(I_J - Z'(ZZ')^{-1}Z)$

Enfin, une dernière régression de R_X sur Z est faite.

$R'_X = Z\gamma \quad \Rightarrow \quad \hat{R}'_{XZ} = Z(Z'Z)^{-1}Z'R'_X$

$\hat{R}_{XZ} = R_X Z(Z'Z)^{-1}Z'$ d'où $\hat{R}_{XZ} = (I_J - X(X'X)^{-1}X')YZ(Z'Z)^{-1}Z'$

Ainsi, $Y = \hat{Y}_{XZ} + R_{\hat{Y}_X} + \hat{R}_{XZ} + Résidus,$

Posons

$$\hat{Y}_{XZ} = X(X'X)^{-1}X'YZ(Z'Z)^{-1}Z' = XMZ' \qquad \text{(B.1)}$$

$$R_{\hat{Y}_X} = X(X'X)^{-1}X'Y(I_J - Z(Z'Z)^{-1}Z') = X\beta' \qquad \text{(B.2)}$$

$$\hat{Y}_{XZ} = (I_J - X(X'X)^{-1}X')YZ(Z'Z)^{-1}Z' = \alpha Z' \qquad \text{(B.3)}$$

X et Z peuvent être mis sous la forme $X = (\ 1_I \quad \dot{X}\ )$ $\ Z = (\ 1_J \quad \dot{Z}\ )$

$\Rightarrow \quad X' = \begin{pmatrix} 1'_I \\ \dot{X}' \end{pmatrix} \quad et \quad Z' = \begin{pmatrix} 1'_I \\ \dot{Z}' \end{pmatrix}$ avec $\dot{X}$ se déduisant de la matrice X par centrage des colonnes (de même pour $\dot{Z}$)

$$\dot{X}'X = \begin{pmatrix} 1'_I \\ \dot{X}' \end{pmatrix}(\ 1_I \quad \dot{X}\ ) = \begin{pmatrix} I & 1'_I\dot{X} \\ \dot{X}'1_I & \dot{X}'\dot{X} \end{pmatrix}$$

Or $\begin{pmatrix} A & B \\ B' & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + FE^{-1}F' & -FE^{-1} \\ -E^{-1}F' & E^{-1} \end{pmatrix}$

avec $E = D - B'A^{-1}B$ et $F = A^{-1}B$

Ici $A = I$, $B = 1_I'\dot{X} \Rightarrow B' = \dot{X}'1_I$ et $D = \dot{X}'\dot{X}$

D'où $A^{-1} = \frac{1}{I}$, $F = \frac{1}{I}1_I'\dot{X} = 0$ car $\dot{X}$ centré

et $E = \dot{X}'\dot{X} \Rightarrow E^{-1} = \left(\dot{X}'\dot{X}\right)^{-1}$.

Ainsi $(\dot{X}'X)^{-1} = \begin{pmatrix} I & 1_I'\dot{X} \\ \dot{X}'1_I & \dot{X}'\dot{X} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{I} & 0 \\ 0 & (\dot{X}'\dot{X})^{-1} \end{pmatrix}$

De même $(\dot{Z}'Z)^{-1} = \begin{pmatrix} \frac{1}{J} & 0 \\ 0 & (\dot{Z}'\dot{Z})^{-1} \end{pmatrix}$

D'après l'équation (B.1) $M = (X'X)^{-1}X'YZ(Z'Z)^{-1}$

$$M = \begin{pmatrix} \frac{1}{I} & 0 \\ 0 & (\dot{X}'\dot{X})^{-1} \end{pmatrix}\begin{pmatrix} 1_I' \\ \dot{X}' \end{pmatrix} Y \begin{pmatrix} 1_J & \dot{Z} \end{pmatrix}\begin{pmatrix} \frac{1}{J} & 0 \\ 0 & (\dot{Z}'\dot{Z})^{-1} \end{pmatrix}$$

$$M = \begin{pmatrix} \frac{1}{I}1_I'Y \\ (\dot{X}'\dot{X})^{-1}\dot{X}'Y \end{pmatrix}\begin{pmatrix} 1_J & \dot{Z} \end{pmatrix}\begin{pmatrix} \frac{1}{J} & 0 \\ 0 & (\dot{Z}'\dot{Z})^{-1} \end{pmatrix}$$

$$M = \begin{pmatrix} \frac{1}{I}1_I'Y1_J & \frac{1}{I}1_I'Y\dot{Z} \\ (\dot{X}'\dot{X})^{-1}\dot{X}'Y1_J & (\dot{X}'\dot{X})^{-1}\dot{X}'Y\dot{Z} \end{pmatrix}\begin{pmatrix} \frac{1}{J} & 0 \\ 0 & (\dot{Z}'\dot{Z})^{-1} \end{pmatrix}$$

D'où

$$\boxed{M = \begin{pmatrix} \frac{1}{IJ}1_I'Y1_J & \frac{1}{I}1_I'Y\dot{Z}(\dot{Z}'\dot{Z})^{-1} \\ \frac{1}{J}(\dot{X}'\dot{X})^{-1}\dot{X}'Y1_J & (\dot{X}'\dot{X})^{-1}\dot{X}'Y\dot{Z}(\dot{Z}'\dot{Z})^{-1} \end{pmatrix} = \begin{pmatrix} m & m_2' \\ m_1 & \dot{M} \end{pmatrix}}$$

D'après l'équation (B.2) $\beta' = (X'X)^{-1}X'Y(I_J - Z(Z'Z)^{-1}Z')$

$$(\dot{X}'\dot{X})^{-1}\dot{X}'Y = \begin{pmatrix} \frac{1}{I}1_I'Y \\ (\dot{X}'\dot{X})^{-1}\dot{X}'Y \end{pmatrix}$$

$$Z(Z'Z)^{-1} = \begin{pmatrix} 1_J & \dot{Z} \end{pmatrix}\begin{pmatrix} \frac{1}{J} & 0 \\ 0 & (\dot{Z}'\dot{Z})^{-1} \end{pmatrix} = \begin{pmatrix} \frac{1}{J}1_J & \dot{Z}(\dot{Z}'\dot{Z})^{-1} \end{pmatrix}$$

Donc $Z(Z'Z)^{-1}Z' = \begin{pmatrix} \frac{1}{J}1_J & \dot{Z}(\dot{Z}'\dot{Z})^{-1} \end{pmatrix}\begin{pmatrix} 1_J' \\ \dot{Z}' \end{pmatrix} = \frac{1}{J}1_J1_J' + \dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}'$

D'où $\beta' = \begin{pmatrix} \frac{1}{I}1'_I Y \\ (\dot{X}'\dot{X})^{-1}\dot{X}'Y \end{pmatrix}(I_J - \frac{1}{J}1_J 1'_J - \dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}')$

$$\boxed{\beta' = \begin{pmatrix} \frac{1}{I}1'_I Y(I_J - \frac{1}{J}1_J 1'_J - \dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}') \\ (\dot{X}'\dot{X})^{-1}\dot{X}'Y(I_J - \frac{1}{J}1_J 1'_J - \dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}') \end{pmatrix} = \begin{pmatrix} b' \\ B' \end{pmatrix}}$$

D'après l'équation (B.3) $\alpha = (I_I - X(X'X)^{-1}X')YZ(Z'Z)^{-1}$

$X(X'X)^{-1}X' = (\ 1_I \quad \dot{X}\ )\begin{pmatrix} \frac{1}{I}1'_I \\ (\dot{X}'\dot{X})^{-1}\dot{X}' \end{pmatrix} = \frac{1}{I}1_I 1'_I + \dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}'$

D'où $\alpha = (I_I - \frac{1}{I}1_I 1'_I - \dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}')(\ \frac{1}{J}Y1_J \quad Y\dot{Z}(\dot{Z}'\dot{Z})^{-1}\ )$

$$\boxed{\begin{aligned} \alpha &= (\ (I_I - \tfrac{1}{I}1_I 1'_I - \dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}')\tfrac{1}{J}Y1_J \quad (I_I - \tfrac{1}{I}1_I 1'_I - \dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}')Y\dot{Z}(\dot{Z}'\dot{Z})^{-1}\ ) \\ &= (\ a \quad A\ ) \end{aligned}}$$

Ainsi $Y = \hat{Y}_{XZ} + R_{\hat{Y}_X} + \hat{R}_{XZ} + Résidus = XMZ' + X\beta' + \alpha Z' + Résidus$

D'où $\hat{Y} = (\ 1_I \quad \dot{X}\ )\begin{pmatrix} m & m'_2 \\ m_1 & \dot{M} \end{pmatrix}\begin{pmatrix} 1'_J \\ \dot{Z} \end{pmatrix} + (\ 1_I \quad \dot{X}\ )\begin{pmatrix} b' \\ B' \end{pmatrix} + \ a \quad A\ )\begin{pmatrix} 1'_J \\ \dot{Z} \end{pmatrix}$

$\hat{Y} = (\ 1_I m + \dot{X}m_1 \quad 1_I m'_2 + \dot{X}\dot{M}\ )\begin{pmatrix} 1'_J \\ \dot{Z}' \end{pmatrix} + 1_I b' + \dot{X}B' + a1'_J + A\dot{Z}'$

D'où

$$\boxed{\hat{Y} = 1_I m 1'_J + \dot{X}m_1 1'_J + 1_I m'_2\dot{Z}' + \dot{X}\dot{M}\dot{Z}' + 1_I b' + \dot{X}B' + a1'_J + A\dot{Z}'}$$

Annexe C : Modèle de simulation de bilan hydrique

Les covariables associées au facteur lieu ont été calculées à l'aide d'un modèle de simulation de bilan hydrique, SarraH.

Ce qui suit est extrait du manuel d'utilisateur et de la documentation technique de SarraH :

(...) *Le modèle SarraH est une version "habillée" (plus détaillée) de Sarra, un bilan hydrique dynamique très simple utilisé pour estimer l'impact d'un scénario climatique sur une culture annuelle. Pour Sarra, la plante n'est qu'un conduit couplant une demande atmosphérique (puits) avec une réserve hydrique "utile" dans le sol (source) - un conduit à résistance variable, fonction de la contrainte physique (stress). Sarra suppose que la performance d'une culture est une fonction simple des freins hydriques cumulés au cours d'un cycle végétal.*

L'évapotranspiration est liée au sol et à la plante et peut être décomposée en processus de transpiration et en processus d'évaporation. Ces processus sont fonction du rapport offre sur demande en eau. Cette demande peut être décrite en deux processus que l'on distingue ou non : l'évaporation du sol et la transpiration de la plante ou encore l'évapotranspiration (sol-plante). Ces deux processus varient en fonction de la demande atmosphérique (Eto), du taux de couverture et des phases phénologiques de la plante. Ces processus s'effectuent de façon simultanée. On distingue la demande potentielle de la consommation réelle, pour chaque processus en fonction de l'eau disponible pour la partie évaporation et la partie transpiration. L'évaporation potentielle est contrainte par l'eau disponible pour ce processus (notion de réservoir de surface), la transpiration potentielle est contrainte par la fraction d'eau disponible aux racines (notion de réservoir des

racines). Ces indices de contrainte sont calculés de différentes manières. Le rapport entre la consommation réelle et la demande potentielle de la plante donne un taux de satisfaction hydrique.

Evaluation de l'évapotranspiration

Hypothèse : le calcul de l'évapotranspiration est fonction de la demande atmosphérique définie par l'équation de Penman-Monteith modifiée Eto (Ref FAO 56). Une partie de la demande est lié à l'évaporation du sol, l'autre à la transpiration de la plante. Pour l'évaporation, on considère qu'elle s'effectue sur un réservoir dimensionné (Ex : 20 cm de profondeur) que l'on définie comme réservoir de surface. Un coefficient d'auto mulching ou de mulch permet de réduire l'évaporation du sol. Pour la transpiration la demande dépend du stade et/ou l'état de développement de la plante. Elle s'effectue uniquement sur la profondeur de sol colonisée par les racines. Cette profondeur de sol dépendant de la vitesse d'enracinement de la plante.

Sous Hypothèses : 1) non séparation de la transpiration et de l'évaporation, 2) séparation. Ce choix influe sur la série d'étapes de calculs correspondant à l'évaluation de la demande et de la consommation.

Evaluation de l'offre

fraction d'eau disponible pour la plante dans le réservoir racinaire, et fraction d'eau disponible dans le réservoir de surface (relation sol atmosphère)

Evaluation de l'évapotranspiration réelle

Hypothèse : On simule une demande potentielle. Pour évaluer l'évapotranspiration ou la transpiration réelle, on fait intervenir un indice de contrainte hydrique (cstr) fonction de la fraction d'eau disponible (FTSW fraction d'eau accessible par la plante), on simule aussi l'évapotranspiration d'un sol nu ou l'évaporation réelle du sol en fonction de FESW (fraction d'eau en relation plus directe avec l'atmosphère, se rapporte à la notion d'un réservoir de surface).

Evaluation de la demande potentielle

Hypothèse : on évalue la demande potentielle à travers des coefficients culturaux (Kc) et de la demande climatique ETo. On peut dissocier ou non l'évaporation de la transpiration.

1) Non séparation de la demande d'évaporation et de transpiration : elle est basé sur la notion des coefficients culturaux (Kctot). Le Kctot est un indice qui intègre l'évaporation du sol et la transpiration de la plante. On simule donc directement l'évapotranspiration du jour. Le Kc est obtenu par mesures expérimentales (lisimètres) qui intègrent alors l'évapotranspiration. Celle ci est déduite de la demande potentielle qui peut être réduite par un indice de contrainte hydrique sol-plante-atmosphère (cstr) fonction de la fraction d'eau disponible. (cf les différentes méthodes de calcul de cet indice).

2) Pour mieux tenir compte du processus d'évaporation on peut ajouter une estimation de l'évapotranspiration sur un réservoir de surface ou l'on considère un Kc constant de 1 (cas Sarra).

3) En considérant que le facteur le plus important de l'évolution des demandes (pour une même demande atmosphérique) est lié à l'évolution du taux de couverture du sol par la plante, on peut aussi dériver cette méthode en séparant la transpiration de l'évaporation au potentiel. Cela revient à considérer que ces processus sont fortement gouvernés par le taux de rayonnement. Dans ce cas le Kc est dissocié en : i) Kcp qui représente la partie couverte par la plante, il est déduit d'une relation avec le taux de transmission de la lumière (1-ltr, fn(lai)) et le Kc-Max de la variété (espèce de fait) et la transpiration potentielle est obtenue en multipliant l'indice par la demande atmosphérique, ii) Kce qui est déduit d'une relation avec le ltr et d'un coefficient de mulch ou auto mulching (même approche Kcp défini par la FAO).

Evaluation de la contrainte hydrique sol-plante-atmosphère

Hypothèse : on considère qu'en fonction de la demande climatique (Eto) et de la fraction d'eau disponible, la plante va avoir une dynamique de consommation différente. Plusieurs méthodes de calculs existent : pFactor notion de réserve facilement utilisable et difficilement utilisable défini par un point d'inflexion (FAO), Eagleman ce rapportant à une relation cubique entre Eto et la fraction d'eau disponible (FTSW) ou Dans toutes ces équations il n'y a pas de différentiation lié à des effets potentiels de la structure des sols, on se rapporte de fait à une variable

calculée, fraction d'eau disponible FTSW qui indirectement en tient compte. Pour lors ces équations semblent insuffisantes dans le cas des plantes isohydriques qui ont une faible variation de la teneur en eau des feuilles. Elles ont une régulation stomatiques très sensible à la demande atmosphérique en plus de la fraction d'eau disponible du sol.

Evaluation de la consommation réelle

Hypothèse : la consommation en eau s'effectue de façon homogène sur chaque réservoir.

Sans séparation : l'évapotranspiration est retranchée sur chaque réservoir . On évalue l'indice de satisfaction hydrique de la plante.

Avec séparation : l'évaporation ne s'effectue que sur le réservoir de surface, alors que la transpiration s'effectue sur le réservoir des racines qui pour partie, ou complètement intègre le réservoir de surface. On considère aussi que l'évaporation est le processus le plus rapide, on le retranche en premier sur le réservoir de surface. Comme le réservoir racinaires intègre tout ou partie du réservoir de surface, il nous faut aussi calculer sur le réservoir ayant des racines la part déjà extraite pour l'évaporation, de même pour le réservoir de surface la part extraite par transpiration au prorata des stocks. Les estimations d'évaporation et de transpirations sont effectuées séparément, on peut ainsi avoir une consommation légèrement supérieure à l'eau disponible. On diminuera donc la transpiration en conséquence sous l'hypothèse que le processus d'évaporation est plus rapide (...)

Résumé

Des modèles de régression factorielle sont proposés pour prédire les interactions GxE pour des essais d'arachide (*Arachis hypogaea* L.) au Sénégal. Au facteur variété est attachée un indice de sensibilité à la sécheresse qui fait intervenir le poids de gousses, le poids de fanes et la durée floraison-récolte des variétés. Au facteur environnement ou année, est associé un taux de satisfaction des besoins calculé sur la première phase de maturité du cycle végétatif de l'arachide. Ce taux a été estimé par simulation à l'aide d'un modèle de bilan hydrique, SarraH, qui utilise en entrée les caractéristiques climatiques du milieu. Sur ces covariables, nous faisons la régression factorielle des observations afin de prédire l'interaction GxE dans de nouveaux environnements. Après avoir replacé la régression factorielle parmi les autres méthodes de modélisation de l'interaction, le modèle présenté par Denis et Vincourt est réécrit afin d'être ajusté sur des données même incomplètes à l'aide d'un logiciel ordinaire de modèle linéaire.

Sur l'exemple d'un essai pluriannuel, ce modèle permet de rendre compte de l'interaction sans écrats significatifs. Pour qu'il puisse être appliqué plus largement, la collecte de données climatiques devra être généralisé en essais multilocaux.