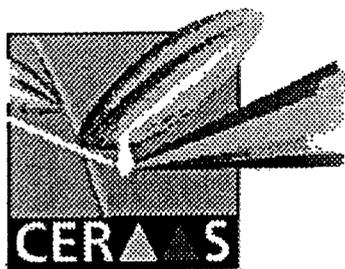
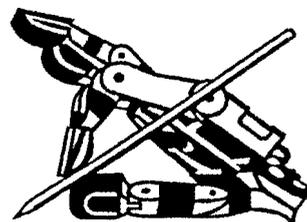


CR001073



Centre d'Etude Régional  
pour l'Amélioration de  
l'Adaptation à la Sécheresse



**I M E R I R**

Institut Méditerranéen d'Etude et de  
Recherche en Informatique,  
Intelligence Artificielle et Robotique

## MEMOIRE D'INGENIERIE

**AUTOPARAMETRISATION DE MODELES DE  
SIMULATION DU DEVELOPPEMENT DES CULTURES**

Réalisé par : Kadidia SAKO.

Promotion Léonard De VINCI 1994 - 1997.

Maître de stage : David BOGGIO.

Du 01 avril au 01 octobre 1997.

## REMERCIEMENTS

Je voudrais témoigner toute ma gratitude à Monsieur Daniel ANNEROSE, Directeur sortant du CERAAS, maintenant Délégué CIRAD Afrique et Océan Indien, d'avoir accepté de m'accueillir au sein de l'équipe CERAAS.

Je tiens à exprimer ma profonde reconnaissance envers Monsieur Harold ROY-MACAULEY, actuel Directeur du CERAAS, avec qui j'ai poursuivi la mise au point de ce projet.

Au CIRAD-CA-URBI, je remercie Monsieur Alain JOLY, Responsable de l'Unité de Recherche Biométrie et Informatique, et Monsieur Pierre MARTIN, informaticien, de m'avoir suivie tout au long de ce stage et d'avoir contribué à la conception de ce document.

Mes remerciements vont particulièrement à Monsieur David BOGGIO, Biométricien, Maître de stage et à Mesdemoiselles COUNA SYLLA et Louise Marie FAYE, Informaticiennes qui n'ont ménagé aucun effort pour la bonne réalisation de ce projet.

Je tiens enfin à rendre hommage à tout le personnel du CERAAS, dont l'accueil sympathique et authentique m'a été d'une grande aide pendant le stage.

## GLOSSAIRE

<b>Autoparamétrisation</b>	Automatisation de la détermination des valeurs des paramètres des modèles de simulation du développement des cultures.
<b>Cultures</b>	Ensemble des opérations propres consistant à tirer du sol les végétaux utiles à l'homme et aux animaux domestiques.
<b>Données d'entrées</b>	Informations qui sont appliquées à un modèle de simulation du développement des cultures.
<b>Données de sortie</b>	Valeurs résultats d'une simulation estimations de composantes du rendement de la culture.
<b>Données de vérification</b>	Valeurs réelles observées aux champs, de même nature que les données de sortie.
<b>Modèle</b>	Représentation simplifiée d'un processus, d'un système.
<b>Modèle empirique</b>	Modèle construit à partir d'équations ajustées aux observations du phénomène au moyen de méthodes statistiques.
<b>Modèle mécaniste</b>	Modèle où le processus est représenté en utilisant les équations exactes qui le définissent.
<b>Modèle sémi-déterministe</b>	Modèle permettant de faire des prévisions, pour ses <b>variables</b> associées, en intégrant des lois de probabilité.
<b>Modélisation</b>	Mise en équation d'un phénomène complexe permettant d'en prévoir <b>les</b> évolutions.
<b>Modèle mathématique</b>	Modèle formé <b>par</b> des expressions mathématiques et destiné à simuler un processus.
<b>Paramètres</b>	Quantité à fixer librement, maintenue constante, dont dépend une fonction de variables indépendantes.
<b>Parcelle</b>	Portion de terrain, identifiée dans l'espace, recevant les mêmes pratiques culturales au même moment.
<b>Point</b>	Jeu de paramètres, combinaison des paramètres d'un modèle.
<b>Simuler</b>	Reproduire à l'aide des caractéristiques du milieu et de la plante, l'évolution d'un processus.

## ABREVIATIONS

<b>Ara.B.Hy</b>	Arachide Bilan Hydrique.
<b>CILSS</b>	Comité Inter-états de Lutte contre la Sécheresse au Sahel.
<b>CIRAD</b>	Centre de coopération Internationale en Recherche Agronomique pour le Développement.
<b>CORAF</b>	Conférence des responsables de recherche agronomique de l'Afrique de l'Ouest et du Centre.
<b>ENSA</b>	Ecole Nationale Supérieure d'Agriculture.
<b>INSAH</b>	<b>INstitut du SAHel.</b>
<b>ISRA</b>	Institut Sénégalais de Recherche Agronomique.
<b>PED</b>	Pays En Développement.
<b>R3S</b>	Réseau de Recherche sur la Résistance à la Sécheresse.
<b>SIG</b>	Système d'Information Géographique.
<b>SNRA</b>	Systèmes Nationaux de Recherches Agronomiques

## SOMMAIRE

INTRODUCTION .....	6
PREMIERE PARTIE ; PRESENTATION DU CERAAS ET CONTEXTE DU PROJET.. .....	.
1. OBJECTIF .....	.
2. ORGANIGRAMME DU CERAAS .....	9
3. ACTIVITES SCIENTIFIQUES .....	10
3.1. <b>Formation par la recherche</b> .....	10
3.2. <i>Recherche</i> .....	10
4. SERVICES D'APPUI .....	12
4.1. <i>Informatique</i> .....	12
4.2. <b>Biométrie</b> .....	12
4.3. <b>Documentation</b> .....	12
4.4. <i>Service d'exploitation</i> .....	13
DEUXIEME PARTIE : ETUDE DU PROJET .....	14
1. EXPRESSION DU BESOIN .....	14
2. SPECIFICATIONS .....	15
3. ANALYSE DU PROBLEME.. .....	17
3.1. <i>Méthode des moindres carrées</i> .....	17
3.2. <b>Détermination</b> de la valeur des paramètres .....	18
4. LA CONCEPTION D'UNE SOLUTION .....	29
4.1. <i>Modification de la base de données des simulations</i> .....	29
4.2. <i>Mise en place de l'application d'autoparamétrisation</i> .....	29
TROISIEME PARTIE : REALISATIONS ET RESULTATS .....	34
1. FONCTIONNEMENT DE L'APPLICATION.. .....	34
2. PRINCIPALES PROCEDURES .....	36
2.1. <i>Procédure RECUPBD</i> .. .....	36
2.2. <b>Procédure ESTSIMP</b> .....	37
2.3. <i>Procédure ESTGRAD</i> .. .....	37
3. RESULTATS OBTENUS. ....	38
QUATRIEME PARTIE : GESTION DU PROJET .....	41
CINQUIEME PARTIE : CONCLUSION .....	43
BIBLIOGRAPHIE .....	44

## INTRODUCTION

Le stage s'est déroulé du 01 avril au 01 octobre 1997 au CERAAS (Bambey-Sénégal). Il a pour thème « Autoparamétrisation des modèles de simulation du développement des cultures ».

Le développement de modèles de fonctionnement des cultures pour les principales espèces vivrières cultivées dans les zones sèches, constitue une des principales activités du CERAAS, dans l'objectif d'améliorer la production agricole. Ces modèles permettent de faire des prévisions de la campagne agricole par les différents intervenants des filières agricoles et alimentaires, les bailleurs de fonds et le gouvernement pour l'aide à la décision. La modélisation offre divers avantages en favorisant l'amélioration de la compréhension du fonctionnement de la plante dans son milieu.

Ara.B.Hy, un modèle développé par le CERAAS, est utilisé pour effectuer des simulations pour la culture d'arachide.

Chaque modèle est défini par une fonction non-linéaire dont les paramètres sont déterminés à partir des mesures effectuées au cours de différentes expériences préliminaires. La précision des modèles de simulations du développement des cultures, dépend des valeurs des paramètres contrôlant leurs réponses aux conditions de l'environnement.

Après une simulation, les données résultats sont comparées à celles recueillies aux champs. Ensuite, suivant un critère de qualité bien défini, une prise de décision s'impose. Dans le cas où le taux d'erreur engendré par la simulation est élevé, il serait intéressant de diminuer l'écart entre les données résultats de simulation et celles observées aux champs.

Ce projet dont l'objectif est de déterminer les valeurs des paramètres des modèles de simulation du développement des cultures (autoparamétrisation) se positionne dans ce cadre. Ces valeurs permettent d'ajuster les données résultats de simulation à celles observées aux champs, et cela indépendamment du modèle de simulation choisi.

Dans la première partie de ce document, le CERAAS et le contexte du projet sont présentés

La deuxième partie est consacrée à l'étude détaillée du projet. Elle montre le besoin exprimé par le CERAAS, les spécifications, l'analyse du problème ainsi que les différentes étapes successives aboutissant à la mise en place d'une solution. Ces différentes étapes de conception sont détaillées en annexe.

La troisième partie de ce document présente le produit logiciel réalisé ainsi que les résultats obtenus.

Le déroulement du projet est décrit dans la quatrième partie du document.

La conclusion et les perspectives de ce travail sont exposées en dernière partie.

## PREMIERE PARTIE : PRESENTATION DU CERAAS ET CONTEXTE DU PROJET

Les catastrophes naturelles et les nombreuses années de grande sécheresse survenues dans le Sahel depuis plus de vingt ans renforcent la nécessité, parmi d'autres, d'approfondir les études sur les stratégies d'adaptation physiologiques adoptées par les végétaux. Une intégration des connaissances acquises dans ce domaine et celui de l'agronomie, permettra de trouver des solutions intéressantes pour stabiliser voire, augmenter la production agricole observées dans les conditions décrites.

Le CERAAS, Centre d'Etude Régional pour l'Amélioration de l'Adaptation a la Sécheresse, créé en 1989, est une base-centre de la CORAF, travaillant sur l'étude de l'adaptation des espèces cultivées à la sécheresse. C'est une structure qui a pris ses origines du programme « l'amélioration de la production de l'arachide en zones sèches » conduit à l'ISRA par ses chercheurs et ceux du CIRAD. Il est devenu un programme régional par le biais du R3S de l'INSAH. Sa thématique « Amélioration de l'adaptation à la sécheresse et création variétale » est l'un des thèmes fédérateurs du R3S.

Le laboratoire principal et les bureaux sont situés dans l'enceinte de l'ENSA à Thiès, à 70 km de Dakar (Sénégal). Un site d'expérimentation et d'autres bureaux sont abrités par le CNBA de Bambey à 50 km de Thiès.

L'équipe du centre est composée de 29 personnes dont 7 chercheurs.

### 1. OBJECTIF

L'objectif général du CERAAS est de contribuer, à travers des actions de recherche et de formation, à l'amélioration de la production agricole vivrière dans les PED des zones sèches.

Les principaux thèmes de recherche abordés sont les suivants :

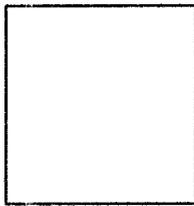
- ⇒ analyse des mécanismes agrophysiologiques d'adaptation à la sécheresse et des espèces cultivées ;
- ⇒ étude de la génétique des caractères afin de proposer des stratégies de sélection et de création variétale adaptées ;

Les connaissances acquises dans une approche pluridisciplinaire du problème, ont été intégrées dans ces thèmes, afin de mettre en place pour chaque situation, des programmes de sélection performants.

Du point de vue des ressources humaines, le projet vise à renforcer les capacités de recherche des équipes des institutions des PED par le développement d'actions conjointes et coordonnées, dans une dimension régionale.

## 2. ORGANIGRAMME DU CERAAS

L'unité de base est le thème. Chaque thème est animé par un cadre et regroupe un ensemble d'activités. Dans le système de gestion comptable, les activités du CERAAS correspondent chacune à un centre d'imputation analytique et sont organisées selon l'organigramme présenté ci-dessous



### 3. ACTIVITES SCIENTIFIQUES

Les activités scientifiques du CERAAS, couvrant à la fois la recherche et la formation à la recherche, sont orientées vers le développement agricole pour l'amélioration de la production dans les régions sèches. L'animation scientifique du projet est assurée par des chercheurs régionaux, nationaux ainsi que par des experts de l'Europe.

L'organisation des activités scientifiques du centre se poursuit autour d'une approche pluridisciplinaire visant à renforcer la maîtrise des concepts et des méthodologies des différents thèmes de recherche sur l'adaptation à la sécheresse (agronomie, physiologie, génétique, sélection et modélisation du développement des cultures) par les chercheurs des PED.

#### 3.1. Formation par la recherche

Le CERAAS a la responsabilité de l'organisation de l'appui aux chercheurs des SNRA.

Cet appui est essentiellement assuré à travers des séjours au **CERAAS** durant lesquels les chercheurs peuvent conduire leurs recherches ou recevoir une formation sur l'étude de l'adaptation à la sécheresse. Ces séjours sont généralement de courte durée :

- ⇒ un à trois mois, pour les chercheurs ayant à résoudre une problématique spécifique dans le cadre de leur programme national de recherche ;
- ⇒ plus de trois mois, pour les chercheurs et étudiants souhaitant obtenir un appui à vocation diplômante (Ingénieur, DEA, DESS, ...). Le **CERAAS** organise également pour les chercheurs des **SNRA**, des séminaires, ateliers afin d'assurer un échange et une bonne diffusion des connaissances.

#### 3.2. Recherche

Les recherches en agronomie, en physiologie et en génétique de l'adaptation à la sécheresse font l'objet de programmes d'amélioration et de création variétale associant plusieurs pays du Sud et du Nord.

De nombreuses cultures sont étudiées. Parmi elles :

- L'arachide** : légumineuse à graine, demeure l'espèce principalement étudiée. Les connaissances acquises sur l'arachide sont actuellement appliquées dans la recherche sur les autres espèces étudiées au CERAAS.
- Le niébé** : légumineuse à graine, est communément étudié par l'ensemble des partenaires.
- Igname-haricot** : légumineuse à tubercules, est une plante utilisée notamment pour l'alimentation humaine, l'alimentation du bétail et l'effet insecticide de ses graines.
- Le mil** : céréale de base pour l'alimentation humaine, notamment dans le Sahel. Elle peut supporter des périodes de sécheresse sans **effet** sur son rendement durant toutes les phases de son cycle à l'exception de la phase de formation et de remplissage des graines.
- Le coton** : culture industrielle. Il existe des variétés de coton particulièrement résistantes à la sécheresse comme le Guazuncho, originaire d'Argentine.

### **Physiologie/biochimie/génétique**

Une meilleure connaissance de la biochimie permet non seulement d'aborder le problème de la création variétale mais aussi de proposer des alternatives modernes comme la création des plantes transgéniques adaptées à la sécheresse. C'est dans cette voie que s'est dirigé le CERAAS, avec ses partenaires notamment l'Université Paris 7.

### **Modélisation : contexte du projet**

Depuis de nombreuses années, le CERAAS a développé des compétences sur la modélisation de croissance et de production de l'arachide. Il est en train d'élargir cette compétence aux autres espèces comme le niébé et le mil. Les connaissances ont été intégrées dans un système de prévision agricole. Ce prototype est testé depuis trois ans dans la région de Diourbel. Ces recherches ont également trouvé des applications en Argentine où le modèle **Ara.B.Hy.** est déjà utilisé pour le pilotage de l'irrigation.

L'« Autoparamétrisation des modèles de simulation du développement des cultures » se situe dans ce domaine.

## 4. SERVICES D'APPUI

### 4.1. Informatique

Le parc informatique du CERAAS se compose :

- ⇒ d'ordinateurs de types 486, Pentium ;
- ⇒ trois serveurs ;
- ⇒ d'un réseau Novell Netware 3.12 sur câble éthernet fin ;
- ⇒ de logiciels de bureautique et de développement informatique ;
- ⇒ trois imprimantes laser Hewlett Packard ;
- ⇒ deux imprimantes Hewlett Packard Deskjet ;
- ⇒ deux imprimantes matricielles.

Les principaux domaines d'activités du service informatique sont :

- ⇒ assistance du personnel aux logiciels bureautiques, analyses de données, animations ;
- ⇒ gestion courante du matériel et des logiciels ;
- ⇒ encadrement des stagiaires ;
- ⇒ développement d'applications internes.

### 4.2. Biométrie

Ce service s'occupe de la mise au point des protocoles expérimentaux et du traitement des données récoltées.

### 4.3. Documentation

Ce service possède trois bases de données. La première, « Current contents », permet la réception de vingt mille nouvelles références sur disquette chaque semaine. Les CD-ROM annuels d'archives bibliographiques facilitent les recherches rétrospectives. La seconde, « CERAAS », rassemble tous les documents scientifiques de la bibliothèque et est mise à jour régulièrement. Enfin la troisième rassemble les renseignements sur les intervenants du centre (chercheurs, stagiaires, missionnaires, partenaires).

La documentation rassemble les différentes éditions réalisées au CERAAS : rapports de mission, publications, comptes-rendus d'ateliers et posters.

#### **4.4. Service d'exploitation**

Ce service est responsable, avec les chercheurs, de la mise en place et la conduite des essais. Ses objectifs sont la rationalisation de l'utilisation des ressources matérielles et humaines et la gestion des parcelles. Il est équipé d'ordinateurs permettant aux techniciens d'effectuer la saisie et le traitement préliminaire des données collectées.

## DEUXIEME PARTIE : ETUDE DU PROJET

### 1. EXPRESSION DU BESOIN

Au Sénégal, les cultures agricoles sont sujettes à de fortes variations inter-annuelles. Les causes en sont : l'irrégularité de la pluviométrie, la baisse de la fertilité des sols, etc.. Ceci entraîne une baisse de la production vivrière, insuffisante par rapport aux besoins de la population.

Une estimation précoce et de manière précise des besoins du pays en produits vivriers, permettra à l'Etat et aux bailleurs de fonds, de prendre des décisions favorables pour amener des solutions améliorantes.

Les modèles de simulation du développement des cultures pour les principales espèces vivrières cultivées dans les zones sèches, élaborés au CERAAS, constituent des moyens de formalisation des connaissances et de diagnostic scientifique, des outils d'aide à la décision. Ils sont réalisés à partir d'informations climatiques, de données agronomiques et physiologiques. Les résultats obtenus sont des estimations de composantes du rendement de la culture, ils sont présentés sous forme de tableaux, de graphiques et de cartes.

Parmi les deux types de modèle qui existent, mécaniste et empirique, celui exploité au CERAAS est du type empirique. Les modèles réalisées au CERAAS permettent une modification de l'ensemble des paramètres utilisés.

Ara.B.Hy., développé en BASIC, est un modèle semi-déterministe [13], empirique, ajoutant à la simulation du bilan hydrique des cultures d'arachide, celle des réactions de la plante aux différents facteurs du milieu (sol, plante, atmosphère). De plus, il intègre dans son développement la composante génétique qui permet de distinguer le comportement de différentes variétés. Il permet d'estimer le flux de matière (matière sèche fixée, eau) et d'informations (états hydriques, ...) entre différents sous-compartiments interconnectés (feuilles, tiges, racines, ..).

Ce modèle permet d'obtenir le bilan hydrique ainsi que la productivité d'une culture d'arachide avec un taux d'erreur (TE) de 10%. Ce taux d'erreur exprime l'écart entre les données résultats de simulation et celles observées aux champs. L'importance de ce taux fait ressortir le besoin de mettre en place une application permettant de le minimiser. Réduire TE, c'est déterminer une valeur optimale pour chaque paramètre du modèle.

## 2. SPECIFICATIONS

Les modèles de simulation du développement des cultures, réalisés au CERAAS, utilisent au maximum vingt paramètres. Le module d'autoparamétrisation doit estimer la valeur des paramètres à partir de l'équation du modèle ainsi que des valeurs observées sur le terrain.

L'équation du modèle est identifiée par :

- le nom du modèle,
- la variété de culture concernée,
- le type de sol sur lequel est cultivée cette variété,
- le site sur lequel cette variété se trouve,
- le département de ce site.

Le CERAAS étudie plusieurs cultures. Chacune présente plusieurs variétés, Par exemple, 57-422 et 55-437 sont des variétés de la culture d'arachide (*Arachis hypogaea L.*).

Les plantes étudiées sont cultivées et se développent sur des sols de type Dior, Dior-Deck et **Deck** , trouvés au Sénégal.

Le Sénégal est administrativement divisé en régions puis en départements. Chaque département est constitué de plusieurs sites. Les différentes observations sont réalisées sur des sites de la région de Diourbel (Sénégal). Elles sont effectuées de façon périodique et s'orientent autour de trois principaux axes :

- le suivi des cultures,
- les relevés climatiques,
- les relevés des superficies des parcelles agricoles.

Toutes les données nécessaires à la simulation sont stockées dans une base de données appelée « **la base de données des simulations** » [14]. Des modifications sont à apporter à cette base de données pour tenir compte de toutes les valeurs nécessaires à l'application d'autoparamétrisation.

Après avoir choisi les données d'entrées et le modèle de simulation, l'utilisateur doit avoir trois possibilités pour déterminer la valeur optimisée de chacun des paramètres du modèle :

- La méthode du simplex,
- La méthode du gradient,
- La méthode du **simplex** suivie de celle du gradient.

Les méthodes du gradient et du simplex sont employées jusqu'à un seuil indiquant la valeur de la condition d'arrêt, c'est à dire dès que l'ajustement des données résultats de simulation aux données de vérification est jugé satisfaisant. Ce seuil est saisi par l'utilisateur au début de l'optimisation.

Les valeurs optimisées doivent s'afficher sur l'écran. Elles doivent, selon le choix de l'utilisateur, être enregistrées dans la base de données des simulations, **et/ou** dirigées vers une imprimante.

Chaque paramètre d'un modèle donné est limité par une borne inférieure et une borne supérieure. Ces valeurs sont, au préalable, stockées dans la base de données des simulations. L'utilisateur peut à tout moment modifier les bornes inférieure et supérieure des paramètres d'un modèle donné.

Les I.H.M. (Interfaces Homme-Machine) doivent être conviviaux.

Access 2.0 et Visual Basic 4.0 sont des logiciels, utilisés actuellement au **CERAAS** pour les développements informatiques. Les modifications à apporter à la base de données des simulations se feront sous Access 2.0 [Annexes page 31]. L'application sera réalisée avec Visual Basic 4.0 [Annexes page 31].

### 3. ANALYSE DU PROBLEME

Le problème revient à déterminer une valeur optimale pour chaque paramètre dans le but de réduire l'écart (E) entre les données résultats de simulation et celles observées aux champs. Mathématiquement, il s'agit de minimiser une fonction non-linéaire à n variables réelles. Cette fonction traduit E.

Pour  $\eta$  observations réalisées, une fonction exprimant un modèle peut être représentée ainsi :

$$Y_i = f(P, X_i) + e_i \quad 1 < i < n$$

$Y_i$  = données de sortie

$P$  = paramètres du modèle

$X_i$  = données d'entrées

$e_i$  = partie aléatoire non contrôlée

i = indice d'occurrence

**Exemple :** Modèle non-linéaire à plusieurs variables d'entrées

$$Y_i = (a x e^{(1-X_{1i}/b)} + c) / (d x X_{1i} + e x X_{2i})$$

avec :

$$x_i = (X_{11}, X_{12}, \dots, X_{1n}, X_{21}, X_{22}, \dots, X_{2n})$$

$$P = (a, b, c, d)$$

#### 3.1. Méthode des moindres carrés

On cherche à ajuster les valeurs de sortie  $Y_i$  à la matrice des variables d'entrées  $X_i$ , au moyen d'une fonction  $f$  dont les paramètres sont contenus dans la matrice  $P$ . On dispose de n observations.

Optimiser la valeur des paramètres, c'est réduire l'écart entre la valeur observée  $Y_{io}$  et la valeur prédite  $Y_{ip}$  correspondant à une même matrice de données d'entrées X. Cet écart pour l'observation i est :

$$E_i = Y_{io} - f(X_i, P)$$

Sur l'ensemble des observations, on ne peut estimer l'écart total par  $E = \sum_{i=1}^n E_i$

car les écarts opposés s'annuleraient. On estime l'écart total du modèle aux données réelles par :

$$E = S(\mathbf{P}) = \sum_{i=1}^n (E_i)^2 = \sum_{i=1}^n (Y_{i0} - f(\mathbf{X}_i, \mathbf{P}))^2$$

$S(\mathbf{P})$  est la somme des carrés des écarts au modèle. Si l'on minimise  $S(\mathbf{P})$ , alors on détermine la matrice  $\mathbf{P}$  des valeurs des paramètres de  $f$  qui permet un ajustement optimal.

### 3.2. Détermination de la valeur des paramètres

Le problème se réduit maintenant à calculer le jeu de paramètres  $\mathbf{P}$  qui minimise  $S(\mathbf{P})$ . Cette fonction est très complexe et interdit la résolution analytique du problème. C'est pour cette raison, que les méthodes de recherche de minimum sont utilisées. Ces méthodes ne cherchent pas à résoudre l'équation mais à s'approcher le plus près possible de la solution.

Les méthodes du simplexe, du gradient et du gradient suivi du simplexe sont utilisées pour déterminer la valeur des paramètres.

#### 3.2.7. Méthode du simplexe

##### Principe général

Dans le cas d'une optimisation à  $n$  paramètres, on définit un simplexe à  $(n + 1)$  points :  $\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \dots, \mathbf{P}_n$ . Chaque point  $\mathbf{P}_i$  est défini par ses coordonnées dans un espace à  $n$  dimensions, avec  $\mathbf{P}_i = (\mathbf{P}_{1i}, \mathbf{P}_{2i}, \dots, \mathbf{P}_{ni})$

La valeur de la fonction  $S$  pour chacun des points  $S(\mathbf{P}_0), \dots, S(\mathbf{P}_n)$  est déterminée. Parmi les valeurs ainsi obtenues, la plus petite, la plus grande et la plus grande suivante sont recherchées, soient  $S(\mathbf{P}_l), S(\mathbf{P}_h), S(\mathbf{P}_{vh})$ , avec :

- $\mathbf{P}_l$  : point le plus bas,
- $\mathbf{P}_h$  : point le plus haut,
- $\mathbf{P}_{vh}$  : point le plus haut suivant.

L'algorithme du simplexe consiste à remplacer  $\mathbf{P}_h$  par un autre point en réalisant une symétrisation par rapport aux barycentres des autres points [13], à chaque étape. La méthode est itérative. Elle doit être appliquée plusieurs fois. La précision de l'estimation est améliorée chaque fois.

Le critère d'arrêt est exprimé par  $(S(\mathbf{P}_k) - S(\mathbf{P}_{k-1})) / S(\mathbf{P}_k) < \text{Seuil}$ . Le Seuil est fixé par l'utilisateur .

Le barycentre est obtenu par :  $B = \sum_{i=1}^n P_i$  avec  $i \neq h$

### La réflexion

Les coordonnées du point réfléchi sont telles que B est le milieu de [Pr Ph]. Pr est donc obtenu par :

$$Pr = (1+\alpha) \times B - \alpha Ph$$

$\alpha$  est le coefficient de réflexion.  $\alpha = 0.9985$

Après une réflexion, on juge la « qualité de la réflexion » :

si  $S(P_f) < S(P_i)$  alors la réflexion est « très réussie », une extension est tentée,  
 si  $S(Pr) < S(Ph)$  alors la réflexion est « réussie », Pr est conservé,  
 si  $s(Pr) > S(Ph)$  alors la réflexion est « ratée », une contraction est tentée.

### L'extension

Les coordonnées du point d'extension sont obtenues par :

$$Pe = \gamma \times Pr + (1 - \gamma) \times B$$

$\gamma$  est le coefficient d'extension.  $\gamma = 1.95$ .

### La contraction

Les coordonnées du point de contraction sont obtenues par :

$$Pc = \beta \times Ph + (1 - \beta) \times B$$

$\beta$  est le coefficient de contraction.  $\beta = 0.4985$

si  $Pc < Ph$  alors Pc est accepté. Dans le cas contraire, les coordonnées de l'ensemble des points sont modifiées :  $P_i = (P_i + P_l) / 2$

### Explication géométrique :

Soit une fonction ayant deux paramètres  $X$  et  $Y$  (optimisation non-linéaire d'ordre deux).

L'estimateur des moindres carrés correspond au minimum de la fonction représentée par la surface.

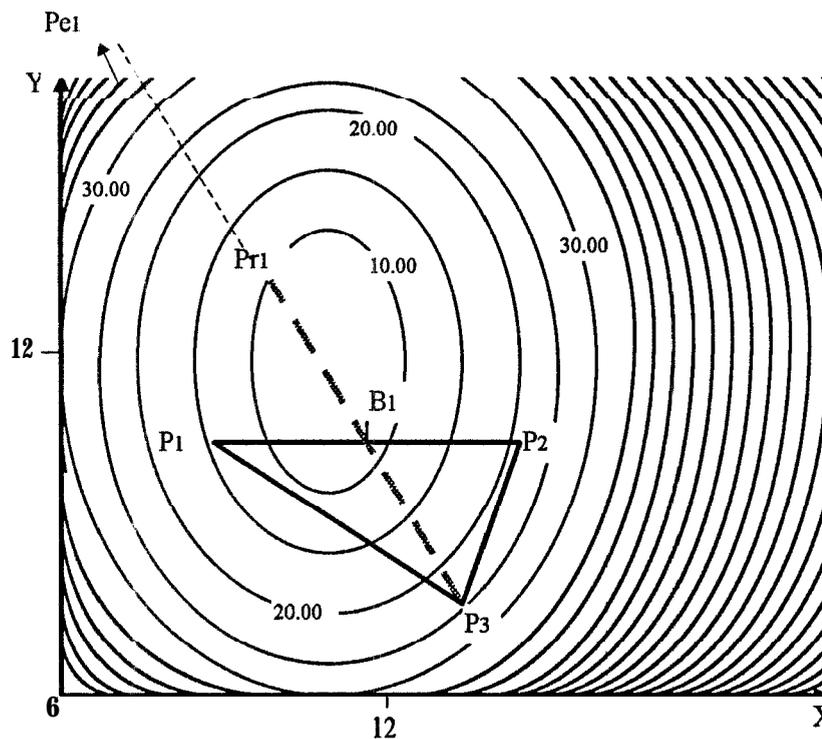
### Première étape

On part initialement de trois points  $P_1$ ,  $P_2$  et  $P_3$  correspondants à trois couples de valeurs  $X$ ,  $Y$  (cf. Figure 1). Ces trois points sont pris au hasard. A chacun d'eux correspond une valeur de  $S(P)$  :

$$S(P_1) \approx 15$$

$$S(P_2) \approx 21$$

$$S(P_3) \approx 25$$



**Figure 1** : Représentation avec courbes de niveaux de la fonction des moindres carrés dans l'espace des paramètres (première étape : réflexion).

Le point ayant le  $S(P)$  le plus élevé (point le plus haut),  $P_3$ , est réfléchi par rapport au barycentre des deux autres. Le point ayant le  $S(P)$  le plus faible (point le plus bas) est  $P_1$ .

$P_{r1}$  est le point réfléchi de  $P_3$  par la symétrie de centre  $B_1$ , milieu de  $[P_1P_2]$ .

Les coordonnées de  $P_{r1}$  sont :

$$P_{r1} = 2 \times B_1 - P_3 = (2 \times X_{B_1} - X_{P_3}, 2 \times Y_{B_1} - Y_{P_3}).$$

$$S(P_{r1}) \approx 10.$$

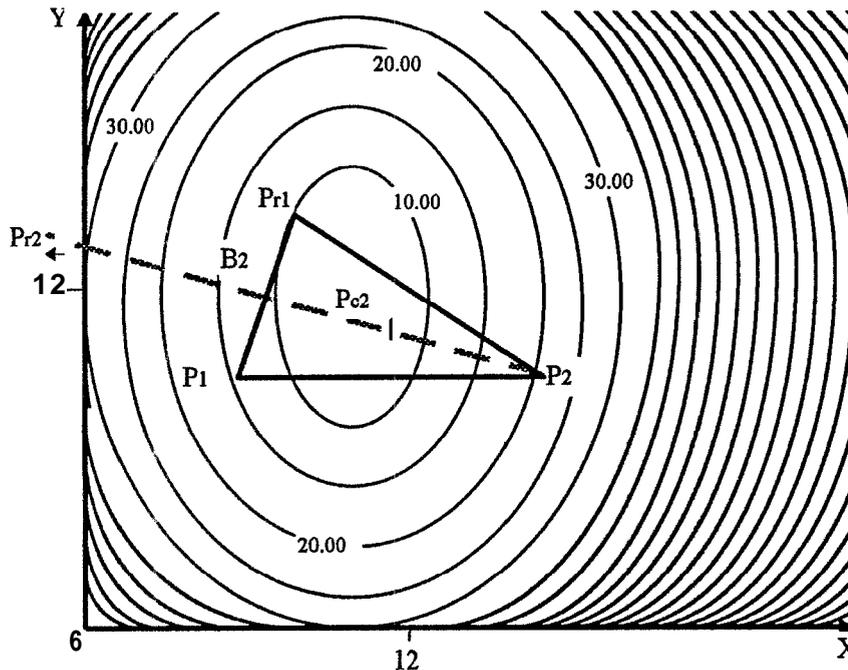
La qualité de la réflexion est jugée.

Si  $S(P_{r1}) < S(P_1)$  : la réflexion est réussie.  $P_{r1}$  est plus bas que  $P_1$ , point initialement le plus bas. On est dans la bonne direction. Une extension est tentée. Les coordonnées du point d'extension sont :

$$P_{e1} = 2 \times P_{r1} - B_1 = (2 \times X_{P_{r1}} - X_{B_1}, 2 \times Y_{P_{r1}} - Y_{B_1}).$$

$$S(P_{e1}) \approx 60$$

$S(P_{e1}) > S(P_{r1})$  donc  $P_{r1}$  est conservé car aucune amélioration n'est apportée par l'extension.

Deuxième étape

**Figure 2** : Représentation avec courbes de niveaux de la fonction des moindres carrés dans l'espace des paramètres (deuxième étape : réflexion + contraction).

P3, point initialement le plus haut est remplacé par  $Pr_1$  (Figure 2). L'opération de l'étape précédente, la réflexion est à nouveau appliquée avec les points  $P_1$ ,  $P_2$  et  $Pr_1$ .  $P_2$  est le point le plus haut.

$Pr_2$  est le point réfléchi de  $P_2$  par la symétrie de centre  $B_2$ , barycentre de  $P_1$  et  $Pr_1$ .

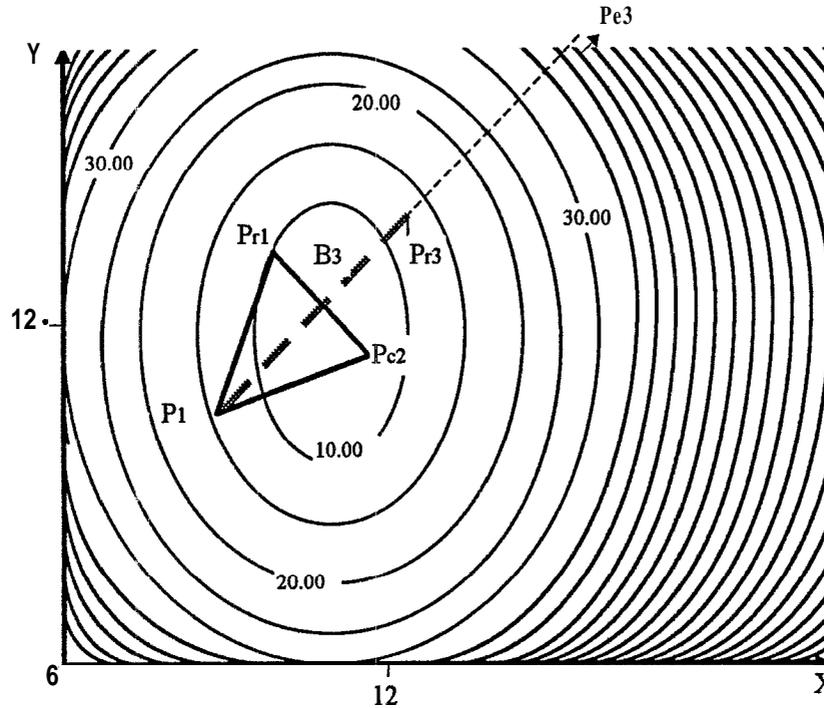
La qualité de la réflexion est jugée.

$S(P_{r2}) > S(P_2)$ , donc, la réflexion n'apporte aucune amélioration.

Une contraction est tentée. Le point de contraction  $P_{c2}$  a pour coordonnées

$$P_{c2} = \frac{1}{2} \times (B_2 + P_2) \approx [\frac{1}{2} \times (X_{B2} + X_{P2}), \frac{1}{2} \times (Y_{B2} + Y_{P2})]$$

$S(P_{c2}) < S(P_2)$ ,  $P_{c2}$  est donc conservé.

Troisième étape

**Figure 3** : Représentation avec courbes de niveaux de la fonction des moindres carrés dans l'espace des paramètres (troisième étape : réflexion + extension).

$P_1$  est maintenant le point le plus haut et  $P_{c2}$  le point le plus bas. La réflexion de  $P_1$  par rapport à  $B_3$  fournit  $P_{r3}$ .

$$S(P_{r3}) \approx 13.$$

La qualité de la réflexion est jugée.

$S(P_{r3}) < S(P_1)$ , la réflexion est réussie.  $P_{r3}$  est plus bas que  $P_1$  (point initialement le plus bas). On tente une extension.

Après extension,  $S(P_{e3}) > S(P_{r3})$ .  $P_{r3}$  est alors gardé car aucune amélioration n'est apportée par l'extension.

La quatrième étape se fera avec  $P_{r1}$ ,  $P_{c2}$  et  $P_{r3}$ . De proche en proche, on détermine ainsi par les coordonnées du point  $P$  le plus bas, la valeur des paramètres  $X$  et  $Y$  qui minimise  $S(X, Y)$ .

**Avantages de la méthode du simplex**

La méthode du simplex est efficace pour une recherche de l'optimum dans un grand espace (les points éloignées de l'optimum). **Elle** est très rapide et demande peu d'itérations dans son fonctionnement. Cette méthode est meilleure pour une première estimation.

**inconvénients de la méthode du simplex**

Au voisinage du minimum, cette méthode est peu rapide car les contractions font perdre du temps.

### 3.2.2. Méthode du *gradient*

#### Principe général

Cette méthode débute avec un seul point.

Le principe général de la méthode du gradient consiste à se déplacer sur la surface représentative du critère des moindres carrés dans une direction qui permet de « descendre » [9], c'est à dire, passer du point  $P_k$  au point  $P_{k+1}$  tel que  $S(P_{k+1})$  soit inférieur à  $S(P_k)$  et d'atteindre ainsi de proche en proche, le point où  $S$  est minimum.

Il s'agit donc d'un procédé itératif, pour lequel quatre questions se posent :

- le choix du premier point,
- le calcul de la direction de descente à chaque étape,
- le choix de la longueur de déplacement (le pas) à chaque étape,
- le choix du critère d'arrêt.

A partir d'un jeu de paramètres  $\mathbf{P}$ , on détermine le gradient  $g$  qui indique la direction de descente.  $g$  exprime la matrice des dérivées partielles d'ordre 1 de  $S(\mathbf{P})$  par rapport aux paramètres  $P_1, \dots, P_n$ . Cette matrice est déterminée par :

$$g = \begin{pmatrix} \frac{\partial S}{\partial P_1} \\ \frac{\partial S}{\partial P_2} \\ \dots \\ \frac{\partial S}{\partial P_n} \end{pmatrix}$$

$\lambda$ , le pas de déplacement sur le vecteur pente, est déterminé à partir du vecteur transposé de  $g$ ,  $g^T$ , ainsi que de la matrice hessienne. Cette matrice exprime les dérivées partielles d'ordre 2 de la fonction des moindres carrés  $S(\mathbf{P})$  par rapport à chaque paramètre :  $P_1, \dots, P_n$ .

$$H = \begin{pmatrix} \frac{\partial^2 S}{\partial P_1 \partial P_1} & \dots & \frac{\partial^2 S}{\partial P_1 \partial P_j} & \dots & \frac{\partial^2 S}{\partial P_1 \partial P_n} \\ \vdots & \dots & \frac{\partial^2 S}{\partial P_i \partial P_j} & \dots & \vdots \\ \frac{\partial^2 S}{\partial P_n \partial P_1} & \dots & \dots & \dots & \frac{\partial^2 S}{\partial P_n \partial P_n} \end{pmatrix}$$

$$g^T = \left[ \frac{\partial S}{\partial P_1} \quad \frac{\partial S}{\partial P_2} \quad \dots \quad \frac{\partial S}{\partial P_n} \right]$$

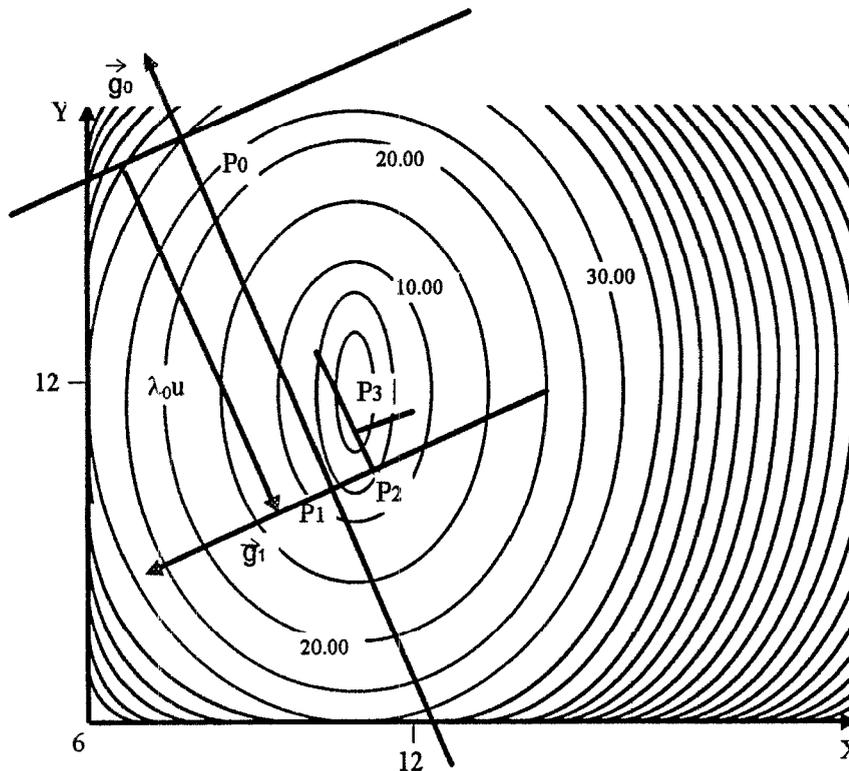
$$\lambda = \frac{|g|^3}{g^T \cdot H \cdot g}$$

$\lambda$  rend S minimal.

Le critère d'arrêt est exprimé par  $(S(P_k) - S(P_{k-1})) / S(P_k) < \text{Seuil}$ . Le Seuil est fixé par l'utilisateur.

**Explication géométrique :**

Soit une fonction ayant deux paramètres X et Y.



**Figure 4 :** Représentation avec courbes de niveaux de la fonction des moindres carrés dans l'espace des paramètres.

$g_0$  est le « vecteur pente » en  $P_0$ . Sa direction indique la ligne de la plus grande pente et son sens indique le sens de la pente. La norme de  $g_0$  est proportionnelle à l'intensité de la pente.

Dans le sens opposé de  $g_0$ , « on descend ». En parcourant la droite de plus grande pente, il faut éviter de « remonter ». Après avoir parcouru la distance  $\lambda$  sur la droite, on atteint un point dont la projection sur la surface est noté  $P_1$ , point bas de la surface dans la direction de  $g_0$ . Au point  $P_1$  on calcule le gradient  $g_1$ .  $g_1$  indique la nouvelle direction ainsi que le nouveau sens de la plus grande pente. On détermine alors  $P_2$ , point bas de la surface dans la direction de  $g_1$ .

### **Avantages de la méthode du gradient**

Au voisinage du minimum, cette méthode permet une convergence rapide et les résultats sont précis. Le jeu de paramètres est amélioré à chaque itération, La méthode du gradient est efficace pour une deuxième estimation”

### **Inconvénients de la méthode du gradient**

Cette méthode est très lente.

### ***3.2.3. Méthode du simplex suivie de celle du gradient***

La meilleure stratégie pour estimer la valeur des  $n$  paramètres permettant de minimiser la fonction des moindres carrés revient à :

1. déterminer les  $n + 1$  points,
2. déterminer un premier jeu de paramètres,  $J_1$ , en appliquant la méthode du simplex (après avoir fixé une condition d'arrêt),
3. déterminer, à partir de  $J_1$ , le meilleur jeu de paramètres,  $J_2$ , en appliquant la méthode du gradient (après avoir fixé une condition d'arrêt).

La résolution du problème d'estimation des paramètres est plus aisée si le domaine de variation des paramètres est restreint.

## **4. LA CONCEPTION D'UNE SOLUTION**

### **4.1. Modification de la base de données des simulations**

Pour modifier la base de données des simulations, la méthode d'analyse MERISE est utilisée. Cette méthode permet une analyse complète d'une base de données relationnelles, en faisant ressortir, entre autre, les différentes données regroupées en tables ainsi que les relations existantes entre elles.

Le modèle conceptuel, représentant les relations existantes entre les différentes tables utilisées dans l'application est détaillé en annexe [Annexes page 18].

### **4.2. Mise en place de l'application d'autoparamétrisation**

#### **4.2.1. Environnement matériel**

Un ordinateur de type Pentium est utilisé pour le développement. Une des imprimantes du centre est utilisée pour l'impression des résultats de l'autoparamétrisation.

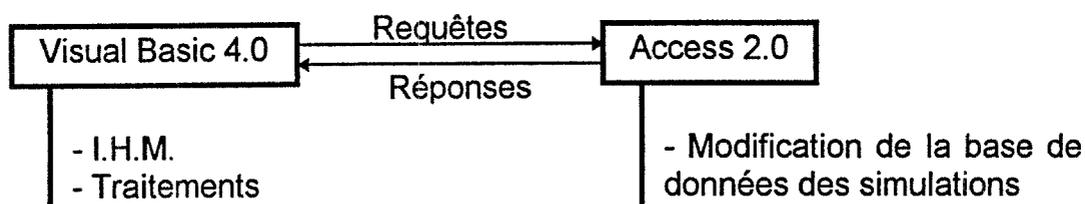
#### **4.2.2. Environnement logiciel**

Les logiciels utilisés sont :

**Access 2.0** : permet d'apporter aisément des modifications à la base de données des simulations, à savoir, ajouter, modifier des tables, et des relations. Ce logiciel permet de bien gérer les tables contenues dans une base de données de type relationnelles. Il fournit facilement le modèle conceptuel des données (MCD de MERISE).

**Visual Basic 4.0** : permet de créer facilement les différents I.H.M., de gérer aisément les événements se rapportant à ceux-ci. L'utilisation du moteur jet avec les objets d'accès aux bases de données (DAO) de Visual Basic (Database, Tabledef, Recordset, Workspace, Index, Field), permet de structurer et manipuler les données de format Access et autres, dans du code.

Les relations entre eux sont les suivantes :



Les requêtes réalisées avec Visual Basic 4.0 sont interprétées par Access 2.0. Les résultats de ces requêtes reviennent à Visual Basic 4.0. Les I.H.M. [Annexes page] ainsi que les différents traitements sont réalisés avec Visual Basic 4.0.

Les différents traitements se répartissent comme suit :

1. localiser la base de donnée des simulations,
2. localiser le modèle de simulation choisi par l'utilisateur,
3. déterminer la valeur optimale de chaque paramètre,
4. modifier la valeur minimale et maximale de chaque paramètre,
5. gérer la nouvelle valeur de chaque paramètre.

Les requêtes qui suivent dans ce document sont détaillées en annexe.

### SITUER LA BASE DE DONNEES DES SIMULATIONS

L'I.H.M. n°1 [Annexes page 24], permet la saisie du chemin d'accès ainsi que du nom de la base de données des simulations. Cette base de **données** est de type Access 2.0 ou compatible.

### SITUER LE MODELE DE SIMULATION CHOISI PAR L 'UTILISATEUR

L'I.H.M. n°2, [Annexes page 25] est utilisée pour cela. Toutes les données nécessaires pour cette localisation sont répertoriées dans la base de données des simulations, Différentes requêtes sont employées.

**Requête 1** : Sélection du nom de tous les départements répertoriés dans la base de données,, Cette requête est réalisée après avoir situé la base de données des simulations.  
Tables **concernées** : TDépartement.

**Requête 2** : Sélection du nom de tous les sites du département choisis par l'utilisateur. Cette requête est réalisée après la requête 1.  
Tables **concernées** : TDépartement, TSite.

**Requête 3 :** Sélection du nom de tous les types de sol, du site, du département choisis par l'utilisateur. Cette requête est réalisée après la requête 2.

Tables **concernées** : TDépartement, TSite, TSol.

**Requête 4 :** Sélection du nom des variétés de cultures se trouvant sur le type de sol, du site, du département choisis par l'utilisateur. Cette requête est réalisée après la requête 3.

Tables **utilisées** : TDépartement, TSite, TSol, TVariété.

**Requête 5 :** Sélection du nom des modèles de la variété de culture se trouvant sur le type de sol, du site, du département choisis par l'utilisateur. Cette requête est réalisée après la requête 4.

Tables **utilisées** : TDépartement, TSite, TSol, TVariété, T Fonction.

#### DETERMINER LA VALEUR OPTIMALE DE CHAQUE PARAMETRE

L'I.H.M. n°2, [Annexes page 24] est utilisée pour déterminer la valeur optimale de chaque paramètre.

Les algorithmes détaillés des méthodes utilisées pour la détermination de la valeur optimale des paramètres; se trouvent en annexe [Annexes page 32].

Les méthodes statistiques utilisées pour l'optimisation de la valeur des paramètres se servent de la fonction des moindres carrés ainsi que de la valeur minimale et maximale de chaque paramètre du modèle. La fonction des moindres carrés est exprimée en fonction :

- d'observations de terrain à la date saisie sur l'I.H.M. n°2,
- des entrées fixes, des entrées variables,
- de la fonction de modélisation.

Ces données sont récupérées dans la base de données grâce à des requêtes,

**Requête 6 :** Récupération du code du département correspondant au nom de département choisi par l'utilisateur. Cette requête ne peut pas être utilisée avant la requête 5.

Tables **utilisées** : TDépartement.

**Requête 7 :** Récupération du code du site correspondant au nom de site choisi par l'utilisateur. Cette requête ne peut pas être utilisée avant la requête 5.

Tables **utilisées** : TSite.

**Requête 8 :** Récupération du code du type de sol correspondant au nom de sol choisi par l'utilisateur. Cette requête ne peut pas être utilisée avant la requête 5.

Tables **utilisées** : TSol.

**Requête 9** : Récupération du code de la variété de culture, du code et de la formulation de la fonction correspondante, aux noms choisis par l'utilisateur. Cette requête ne peut pas être utilisée avant la requête 5.

**Tables utilisées** : TFonction.

**Requête 10** : Détermination du nombre des entrées fixes et variables du modèle choisi par l'utilisateur. Cette requête ne peut pas être utilisée avant la requête 9.

**Tables utilisées** : Tvariables calculées.

**Requête 11** : Récupération du nom des entrées fixes, variables et de la table contenant les valeurs des entrées variables du modèle choisi par l'utilisateur. Cette requête ne peut pas être utilisée avant la requête 10.

**Tables utilisées** : Tvariables calculées.

**Requête 12** : Récupération de la valeur de chaque entrée fixe du modèle choisi par l'utilisateur. Cette requête ne peut pas être utilisée avant les requêtes 11.

**Tables utilisées** : TVariété.

**Requête 13** : Détermination du nombre total des observations réalisées aux champs pour le modèle choisi par l'utilisateur. Cette requête ne peut pas être utilisée avant la requête 11.

**Tables utilisées** : TObservations.

**Requête 14** : Récupération de la valeur des entrées variables du modèle choisi par l'utilisateur. Cette requête ne peut pas être utilisée avant les requêtes 7, 8 et 9.

**Tables utilisées** : Le nom est désigné dans le champ « Type variable » de la table TVariables calculées.

**Requête 15** : Détermination du nombre de paramètres du modèle choisi par l'utilisateur. Cette requête ne peut pas être utilisée avant la requête 9.

**Tables utilisées** : Tparamètres.

**Requête 16** : Récupération de la valeur minimale et maximale de chaque paramètre du modèle choisi par l'utilisateur. Cette requête ne peut pas être réalisée avant la requête 9.

**Tables utilisées** : Tparamètres, TFonction.

**Requête 17** : Récupération de la position du champ contenant les valeurs des observations réalisées aux champs pour le modèle choisi par l'utilisateur. Cette requête ne peut pas être réalisée avant la requête 5

**Tables utilisées** : TObservations.

**Requête 18** : Récupération des valeurs des observations réalisées aux champs pour le modèle choisi par l'utilisateur. Cette requête ne peut pas être réalisée avant les requêtes 7, 8, 9, 17  
**Tables utilisées** : TObservations.

### MODIFIER LA VALEUR MINIMALE ET MAXIMALE DE CHAQUE PARAMETRE

L'I.H.M. n°3, [Annexes page 26] est utilisée pour cela. Les limites inférieure et supérieure des paramètres, initialisées au départ dans la base de données sont modifiables. La requête 16 est utilisée pour réaliser ce traitement.

### GERER LA NOUVELLE VALEUR DE CHAQUE PARAMETRE

Les valeurs optimisées sont directement **affichées** sur l'I.H.M. n°2' Elles peuvent être :

- diriger vers une imprimante.
- enregistrer la valeur des paramètres dans la base de données,

**Requête 19** : Stocker la valeur de chaque paramètre optimisée du modèle choisi par l'utilisateur dans la table Tparamètres. Cette requête ne peut pas être réalisée avant l'exécution d'une des méthodes d'optimisations proposées.  
**Tables utilisées** : Tparamètres.

## TROISIEME PARTIE : REALISATIONS ET RESULTATS

### 1. FONCTIONNEMENT DE L'APPLICATION

Après la mise en place de la fonction des moindres carrés, l'application d'autoparamétrisation détermine la valeur des paramètres permettant de minimiser l'écart entre les données résultats de sortie de simulation et celles observées aux champs.

L'application d'autoparamétrisation fonctionne avec la base de données des simulations. Avant de la lancer, certaines tables de cette base de données doivent obligatoirement être renseignées :

1. **TDépartement**, **TSite**, **TSol**, **TVariété**, **TFonction**, permettent de situer le modèle dont on veut optimiser les paramètres,
2. **TObservations**, **TVariables** calculées, permettent de situer les valeurs observées aux champs ainsi que les entrées du modèle ; cela dans le but de mettre en place la fonction des moindres carrés.

Ces tables répondent à des contraintes d'intégrité.

L'utilisateur a la possibilité de quitter l'application à tout moment en choisissant le bouton Quitter.

Après avoir lancer l'application, la base de données des simulations est demandée à l'utilisateur. Elle doit être de type MS Access 2.0 ou compatible sinon l'erreur **E1** se produit : « Ce fichier *n'est pas une base de données de type MS Access* ».

La structure de la base de données des simulations doit être conforme à celle attendue. Elle doit contenir toutes les tables nécessaires à l'autoparamétrisation. Dans le cas contraire, après recherche de la première table T attendue, l'erreur **E2** se produit : « La table *T n'existe pas dans cette base de données* »

Une fois la base de données saisie, l'utilisateur doit situer le modèle de simulation dont il veut calibrer les paramètres. Cela se traduit par une saisie du nom du département, du site, du type de sol, de la variété de culture et enfin du modèle lui-même.

Ensuite, l'utilisateur doit saisir la date de paramétrisation. Par défaut, la date du jour est utilisée. Cette date permet de déterminer les valeurs d'observations à prendre en compte. Il s'agit des observations effectuées avant cette date. Si la date saisie n'est pas dans le format attendu, l'erreur **E3** se produit : « *DATE INCORRECTE* ». Le format attendu est alors indiqué sur l'écran.

Après la saisie de la date, deux choix s'offrent à l'utilisateur :

- Modifier la valeur des paramètres du modèle : un choix du bouton **Paramètres** permet de modifier chaque paramètre du modèle par l'intermédiaire de l'I.H.M. n°3 [Annexes page 26]. Pour être prise en compte, chaque paramètre modifié doit être enregistré dans la base de données.
- Sélectionner une des méthodes de détermination de la valeur des paramètres : A chaque méthode est associée un seuil permettant de mettre en place la condition d'arrêt. Ce seuil est demandé à l'utilisateur. Sa valeur par défaut est 0.01 pour la méthode du simplex et 0.001 pour celle du gradient. Une validation de ce seuil entraîne l'exécution de la méthode. Pour celle du simplex suivi du gradient, la valeur du seuil du gradient doit être inférieure ou égale à celle du simplex. Dans le cas contraire, l'erreur **E4** se produit : « Seuil Simplex *doit être inférieur à Seuil Gradient* ».

Les résultats obtenus, affichés à l'écran, peuvent être enregistrés dans la table Tparamètres de la base de données ou dirigées vers une imprimante, à la demande de l'utilisateur.

### Gestion des principales erreurs

**E1** : L'utilisateur doit choisir une base de données du type MS Access 2.0 ou compatible.

**E2** : La base de données saisie n'est pas celle attendue, c'est à dire la base de données des simulations.

**E3** : La date d'autoparamétrisation doit avoir le format suivant : le jour (deux positions) suivi d'un séparateur suivi du mois (deux positions) suivi d'un séparateur suivi de l'année (quatre positions). Les séparateurs sont : ., :, /, -, **espace**. Le mois doit être compris entre 1 et 12. Le jour est compris

- entre 1 et 31 pour les mois : 1, 3, 5, 7, 8, 10, 12 ;
- entre 1 et 30 pour les mois : 4, 6, 9, 11.
- entre 1 et **28/29** pour le mois 2.

**E4** : Il faut choisir un seuil pour la méthode du gradient de telle sorte que celui-ci soit inférieur à celui du simplex.

## 2. PRINCIPALES PROCEDURES

Elles sont au nombre de trois : RECUPBD, ESTSIMP et ESTGRAD. Ces procédures sont détaillées en annexe [Annexes page 32]

### 2.1. Procédure RECUPBD

Cette procédure permet la récupération de la base de données de simulations, toutes les données nécessaires à l'application. Elle permet également la mise en place de la fonction des moindres carrés.

Elle prend en entrée, les caractéristiques du modèle de simulation dont on veut déterminer la valeur des paramètres ainsi que le nom et le chemin d'accès à la base de données des simulations.

#### Principale contrainte

Elle concerne l'expression de la fonction de modélisation contenue dans le champ Fonction de la table **TFonction** de la base de données des simulations. Cette fonction peut être exprimée avec des conditions multiples « si », elle peut être, aussi complexe que possible, à condition de bien respecter le format.

Exemple : le modèle poids sec feuille de la culture de mil

```

si TempCum<=TempSéne alors
    PS feuille = Psmax/(X(0)+(X(1)*exp(X(2)*SomTemp)))
sinon
    PS feuille = Psmax/(X(0)+(X(1)*exp(X(2)*SomTemp)))-(Tempcum-
    TempSéné)*ss)
fin si

```

Cela sera représenté dans la table **TFonction** de la base de données des simulations comme suite :

```

(si TempCum<=TempSéne;Psmax/(X(0)+(X(1)*exp(X(2)*SomTemp)));
Psmax/(X(0)+(X(1)*exp(X(2)*SomTemp)))-(Tempcum-TempSéné)*ss)

```

X(0), X(1), X(2) sont les paramètres du modèle.

TempCum est une entrée variable du modèle.

Psmax, ss, TempSéne sont des entrées fixes du modèle.

## 2.2. Procédure ESTSIMP

Cette procédure permet de déterminer une valeur optimale pour chaque paramètre d'un modèle de simulation donné en utilisant la méthode du simplexe. C'est une procédure itérative, qui donne des résultats rapidement moyennant une condition d'arrêt d'estimation.

Elle prend en entrée, la fonction des moindres carrés, le seuil saisi à l'écran, les points de départ, le nombre total de paramètres du modèle ainsi que la valeur minimale et maximale de chaque paramètre.

## 2.3. Procédure ESTGRAD

Cette procédure permet de déterminer une valeur optimale pour chaque paramètre d'un modèle de simulation donné en utilisant la méthode du gradient. C'est une procédure itérative, lente, qui donne des résultats moyennant une condition d'arrêt d'estimation.

Elle prend en entrée, la fonction des moindres carrés, le seuil saisi à l'écran, le point de départ, le nombre total de paramètres du modèle ainsi que la valeur minimale et maximale de chaque paramètre.

### 3. RESULTATS OBTENUS

L'exemple suivant porte sur le modèle PS feuille (Poids sec feuille) de la culture se trouvant sur un sol de type Dior du site de Saint-Louis, dans le département de Dagana au Sénégal. 10 observations sont prises en compte. Ce modèle comporte trois paramètres P1, P2, P3. Les résultats obtenus avec EXCEL 5.0, sont détaillés en annexe [Annexes page 104]. Pour  $P1 = 1$ ,  $P2 = 926.63$ ,  $P3 = -0.0085$  ; à la 51<sup>ème</sup> observation, les valeurs de sortie du modèle diffèrent de celles observées aux champs.

l'application d'autoparamétrisation utilise les valeurs minimale et maximale de chacun des paramètres. Le tableau ci-dessous détaille ces valeurs.

Nom du paramètre	Valeur minimale	Valeur maximale
P1	0	1.2
P2	926	929
P3	-0.008	-0.005

L'I.H.M. ci-dessous montre les résultats obtenus après optimisation par la méthode du simplex avec comme valeur de seuil : 0.01.

AUTOPARAMÉTRISATION DES MODÈLES DE SIMULATION DU DÉVELOPPEMENT DES CULTURES

Département:

Site:

Type de sol:

Variété de culture:

Modèle:

Date:

Méthode d'optimisation:  Simplex  Gradient  Simplex + Gradient

Seuil:

Détermination de la valeur des paramètres permettant d'ajuster les données résultats de simulation avec celles observées aux champs

RESULTATS D'OPTIMISATION

CAS DU SIMPLEX

Valeurs estimées:

p1 = 1  
p2 = 926.06294233554  
p3 = -5.0000003129029E-03

Coefficient de détermination: 1.00000

L'I.H.M. ci-dessous montre les résultats obtenus après optimisation par la méthode du gradient avec comme valeur de seuil : 0.001.

AUTOPARAMETRISATION DES MODELES DE SIMULATION DU DEVELOPPEMENT DES CULTURES	
Département	Dagana
Site	Saint_louis
Type de sol	DIOR
Variété de culture	M#
Modèle	PS feuille
Date	09/09/1997
<p>Méthode d'optimisation</p> <p><input type="radio"/> Simplex</p> <p><input checked="" type="radio"/> Gradient</p> <p><input type="radio"/> Simplex + Gradient</p>	
Seuil	0.001
Paramètres	
Enregistrer	
Imprimer	
Quitter	
<p>Détermination de la valeur des paramètres permettant d'ajuster les données résultats de simulation avec celles observées aux champs.</p>	
<p><b>RESULTATS D'OPTIMISATION</b></p> <p><b>CAS DU GRADIENT</b></p> <p>Valeurs estimées</p> <p>p1 = 1</p> <p>p2 = 327</p> <p>p3 = 5.000000000000000E-03</p> <p>Coefficient de détermination: 1.000</p>	

L'I.H.M. ci-dessous montre les résultats obtenus après optimisation par la méthode du simplex suivi de celle du gradient. La valeur du seuil pour la méthode du simplex est 0.01 et pour celle du gradient 0.001.

AUTOPARAMETRISATION DES MODELES DE SIMULATION DU DEVELOPPEMENT DES CULTURES	
Département	Dagana
Site	Saint_louis
Type de sol	DIOR
Variété de culture	M#
Modèle	PS feuille
Date	09/09/1997
<p>Méthode d'optimisation</p> <p><input type="radio"/> Simplex</p> <p><input type="radio"/> Gradient</p> <p><input type="radio"/> Simplex + Gradient</p>	
Seuil	0.01
Paramètres	
Enregistrer	
Imprimer	
Quitter	
<p>Détermination de la valeur des paramètres permettant d'ajuster les données résultats de simulation avec celles observées aux champs.</p>	
<p><b>RESULTATS D'OPTIMISATION</b></p> <p><b>CAS DU SIMPLEX SUIVI DU GRADIENT</b></p> <p>Valeurs estimées</p> <p>p1 = 1</p> <p>p2 = 326.01</p> <p>p3 = 0.005000000000000E-03</p> <p>Coefficient de détermination: 1.000</p>	

Dans cet exemple, le meilleur résultat est obtenu avec la méthode du simplex suivi de celle du gradient.

Pour des fonctions de modélisation très complexe, le temps d'exécution de l'application est élevé.

## QUATRIEME PARTIE : GESTION DU PROJET

Les premiers jours de travail ont été consacrés à l'étude des documents présentant le CERAAS. Cette étude a été réalisée dans le but de bien situer le contexte du stage.

Ensuite, une **étude** du modèle de simulation Ara.B.Hy. a été nécessaire pour comprendre le fonctionnement des modèles de simulations réalisés au CERAAS.

La méthodologie de travail adoptée a été celle du cycle en V [Annexe page 13]. Les différents documents réalisés tout au long du projet, validés par D. BOGGIO (agronome-biométricien) et C. SYLLA (informaticienne), ont permis d'aboutir au produit final.

Les suggestions apportées par D. BOGGIO et C. SYLLA ont permis de bien positionner le cahier des charges dans le contexte de la simulation.

La programmation a été faite de manière ascendante. En effet, il a fallu d'abord voir l'efficacité des méthodes statistiques utilisées avant de mettre au point les I.H.M. ainsi que les traitements liés à la base de données des simulations, La non efficacité de ces méthodes statistiques aurait orienté le choix vers d'autres outils d'estimation et d'optimisation. Il est plus aisé de développer avec Visual Basic de manière descendante plutôt qu'ascendante, car, c'est un logiciel de type événementiel. Cette procédure a entraîné un ajout de 20% au temps initialement prévu pour l'intégration des I.H.M. et de la base de données de simulations au reste de l'application.

Les méthodes statistiques utilisées ont été réalisées par étape. Elles ont été validées par D. BOGGIO. La méthode du simplexe a été réalisée en premier. Après validation, celle du gradient a été mise au point. Celle-ci a été plus longue à réaliser à cause de la complexité des fonctions comme celle permettant la mise en place de la matrice hessienne.

Une fois les deux premières méthodes validées, celle du simplexe suivi du gradient à été rapidement mise en place.

Les différents tests de validation ont été effectués avec le modèle de calcul du poids sec des feuilles de la culture de mil, une fonction mise au point par le responsable du stage, et des **exemples** de fonctions des livres de statistiques,

Au cours du stage, des **éclaircissement** apportés par des chercheurs du centre, ont permis de tenir compte de certaines contraintes liées aux paramètres d'un modèle.

Le planning réalisé au début du stage a été suivi du mieux que possible. La marge établie a été bien utilisée. Il n'y a pas eu de décalage par rapport à la durée initialement prévue pour le stage. Le tableau suivant montre le décalage dans le **temps** intervenu au niveau des différentes tâches du projet

Tâches	Durée prévue	Durée réelle
Etude du besoin exprimé par le CERAAS	5j	14j
Mise en place du plan de développement:	3j	3j
Mise en place du dossier de spécifications fonctionnelles	3j	5j
Mise en place du dossier de conception	18j	20j
Codage des différentes fonctions	35j	45j
Tests de validation	10j	17j
Tests d'intégration	10j	20j
Réalisation du rapport de stage	3j	4j
Livraison de l'application	2j	1j

Au départ, le travail était prévu pour 89 jours de travail, soit 74% du temps reparti sur les six mois. La marge obtenue était de 26%.

A la fin du stage, la marge dégagée est nulle. Les six mois prévus ont été suffisants pour réaliser le projet.

## CINQUIEME PARTIE : CONCLUSION

L'objectif principal, c'est à dire, déterminer la valeur des paramètres permettant de minimiser l'écart entre les données résultats de simulation et celles observées aux champs, a été atteint. Ces paramètres stockés dans la base de données des simulations sont utilisés pour la réalisation des simulations. En effet, ces paramètres servent à renseigner le modèle pour que les résultats de sorties de simulations soient assez proche de la réalité observée aux champs.

Les tests effectués sur le **modèle** de calcul du poids sec des feuilles de la culture du mil sont satisfaisants.

L'application d'autoparamétrisation est facile à utiliser. Les I.H.M. sont simples et peu nombreuses.

La modularité du code et les commentaires qui y sont associés facilitent la maintenance.

Plus la fonction de modélisation est complexe, et le nombre d'observations élevé, plus le temps d'exécution du programme est long.

Les méthodes statistiques d'optimisation et d'estimation, utilisées, à savoir la méthode de simplex et celle du gradient, sont efficaces pour des modèles de simulation ne présentant pas beaucoup de paramètres (maximum 20). Dans le cas contraire, il faudra utiliser des méthodes comme les algorithmes génétiques.

Le meilleur procédé pour déterminer la valeur optimale de chaque paramètre est d'utiliser la méthode du simplex suivi de celle du gradient.

La méthode du gradient, seul, est satisfaisant pour des modèles ne demandant pas une grande exploration de l'espace défini.

La méthode du simplex est **très** exploratrice. Elle doit donc être utilisée pour des modèles dont l'espace défini par les paramètres est très étendu. Mais cette méthode est moins précise aux alentours de la valeur minimale recherchée.

Ce stage réalisé au CERAAS, avec les moyens mis à ma disposition m'a permis de comprendre ce que fait la particularité d'une telle société.

Je retiens trois enseignements essentiels :

1. Approfondissement des connaissances en Visual Basic et Access : l'interface entre les deux logiciels est très facile à utiliser.
2. Rôle et fonction de la communication en entreprise : la réalisation **d'un** projet informatique ne peut se faire sans une échange permanente avec le client.
3. Discipline et responsabilité : respect des horaires, gestion du temps de travail.
4. L'importance de l'informatique dans une structure de **recherche**.

**BIBLIOGRAPHIE**

- [1] ANNEROSE D. J. M. (1990), Recherches sur les mécanismes physiologiques d'aptation à la sécheresse. Application au cas de l'arachide (*Arachis hypogaea* L.). Thèse, Université Paris VII, 282 p.
- [2] ANNEROSE D. J. M., DIAGNE Madiagne, Modélisation et prévision agricole, Bambey, Sénégal.
- [3] ANNEROSE D. J. M., Modélisation et applications pour le développement agricole, CERAAS, Bambey, Sénégal.
- [4] BEYALOU M. (1994), Etude comparative du développement agrophysiologique de 5 variétés d'arachide (*Arachis hypogaea* L.) cultivées au Sénégal durant les 2 premières phases de leur cycle en vue d'alimenter en données le modèle Ara.B.Hy.. Mémoire ENCR, Bambey - Sénégal.
- [5] BERGONZONI J. C., DUBY C., Analyse et planification des expériences les dispositifs en bloc.
- [6] CERAAS, Rapport annuel 1996. Projet de renforcement du CERAAS.
- [7] FAYE Omar Ndaw (1995), Découverte d'une structure de recherche : le CERAAS. Rapport de stage, Bambey - Sénégal.
- [8] GUEYE M. (1993), Modélisation du rendement et méthodes d'estimation des superficies en cultures d'arachide. Mémoire de fin d'études, INDR/Thiès, Bambey - Sénégal.
- [9] HUET S., JOLIVET E., MESSEAN A. (1992), la régression non-linéaire, méthodes et applications en biologie, INRA, Paris - France.
- [10] JOLIVET E. (1983), Introduction aux modèles mathématiques en biologie, Masson, Paris - France.
- [11] LAOUMADJI K. (1994), Modélisation des cultures. Application au modèle ARA.B.HY. au suivi de la campagne arachidière 1994 : cas de la région de Diourbel au Sénégal. Mémoire ENCR, Bambey - Sénégal.
- [12] LOUKOUBAR Thierno (1990), Modélisation du développement de l'arachide (*Arachis hypogaea* L.) au Sénégal INDR/Thiès, Bambey - Sénégal.
- [13] MEAD R. NELDER J. A. (1965), A simplex method for function minimisation, Computer Journal, 308-313.
- [14] ORSATTI Xavier (1997), Mise en place d'une base de données associée à la modélisation des cultures et génération automatique des cartes. Mémoire DESS, Bambey Sénégal.
- [15] SAPORTA G. (1990), PROBABILITES ANALYSE DES DONNEES ET STATISTIQUE, Technip, Paris - France.
- [16] SAUTEREAU O. (1994), Suivi du bilan hydrique et du développement agrophysiologique de 5 variétés d'arachide en vue d'alimenter en données nouvelles le modèle de croissance Ara.B.Hy. Mémoire ISTOM, Cergy Pontoise - France.
- [17] SPRENT Peter (1992), Pratique des statistiques nonparamétriques - TECHNIQUES ET PRATIQUES, Paris - France.
- [18] SYLLA C (1993), Analyse du modèle de simulation Ara.B.Hy version 01.

# AUTOPARAMETRISATION DE MODELES DE SIMULATION en DEVELOPPEMENT DES CULTURES

## Résumé

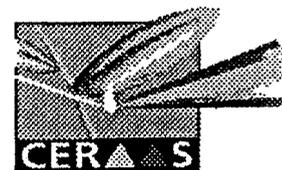
Ce mémoire présente les résultats de six mois de stage effectué au CERAAS (Centre d'Etude Régional pour l'Amélioration de l'Adaptation à la Sécheresse). L'objectif du stage est de mettre en place une application permettant de déterminer la valeur des paramètres des modèles de simulation du développement des cultures, afin d'ajuster les données résultats de simulation à celles observées aux champs. Pour cela, des méthodes d'estimation et d'optimisation statistiques sont utilisés. Ce sont les méthodes du simplexe et du gradient. Les résultats obtenus permettent d'ajuster les données résultats de simulation à celles observées aux champs avec un taux d'erreur assez faible. Ces résultats sont utilisés par les modèles de simulation du développement des cultures, réalisés au CERAAS.

Réalisé par : Kadidia SAKO.  
Promotion Léonard De VINCI 1994 - 1997.  
Maître de stage : David BOGGIO.  
Lieu : CERAAS.

*Stage effectué du 01 avril au 01 octobre 1997.*



Institut Méditerranéen d'Etude et de  
Recherche en Informatique,  
Intelligence Artificielle et Robotique



Centre d'Etude Régionale  
pour l'Amélioration de  
l'Adaptation à la Sécheresse