



**CENTRE D'ETUDES REGIONAL
POUR L'AMELIORATION
DE L'ADAPTATION A
LA SECHERESSE.**

ISRA • CNRA
B.P. 53 BAMBEY SENEGAL
TEL. (221) 73.61.97 - 73.60.50
FAX (221) 73.61.97 - 73. 6052

CR001171

Service Informatique

**Régression non-linéaire du modèle
ARABHY**

par

Yaye COUNA SYLLA

SYLLA

MODE 3

REGRESSION NON-LINEAIRE DU MODELE
ARABHY (Y.C.S.)

1 - INTRODUCTION

L'objectif de ce travail est de développer une application permettant de caler les paramètres d'un modèle de simulation du développement d'une culture à partir de données de terrain.

Les outils disponibles sont :

- un (1) modèle de simulation de culture.
- un (1) module de régression non-linéaire.

2 - ANALYSE PREALABLE

2 .1 - Module de régression non-linéaire

Le module de régression non-linéaire a été extrait du programme STATITCF. Ce programme est écrit en BASIC. Il réalise l'optimisation de n'importe quelle fonction non-linéaire comportant un (1) à dix (10) paramètres .

Le module de régression non-linéaire lit dans un fichier STATITCF les paramètres de la fonction et les observations à comparer aux valeurs simulées.

Il utilise au maximum six (6) observations pour estimer un (1) ou deux (2) paramètres d'une fonction non-linéaire.

2.2 - Modèle de simulation du développement d'une culture

Le module de simulation du développement d'une culture (ARABHY) a été développé au CERAAS (en Quick-BASIC).

Les caractéristiques du sol, des variétés, climatiques et les observations faites à différents stades de développement de la culture sont lues dans des fichiers. La simulation du développement de la culture est faite avec un pas journalier.

Le nombre de paramètres utilisé par ce modèle est supérieur à celui utilisé par le module de la régression non-linéaire.

3 - PRINCIPE GENERAL

L'utilisateur fournit des valeurs approchées des paramètres à estimer.

Le module de régression non-linéaire teste plusieurs valeurs de paramètres et détermine de proche en proche celles qui vont minimiser les écarts entre les valeurs simulées fournies par le modèle et les valeurs observées.

Pour cela il appelle le modèle de simulation autant de fois que nécessaire en lui fournissant les nouvelles valeurs des paramètres à tester.

En retour, le modèle renvoie les valeurs simulées à comparer aux valeurs observées du terrain ou la somme des carrés des écarts entre les valeurs simulées et observées.

Deux interfaces ont été prévues afin de limiter les modifications sur les deux modules.

- Une interface de saisie. Cette interface permet à l'utilisateur de vérifier les informations lues dans des fichiers, de modifier les valeurs initiales des paramètres et de choisir les paramètres à optimiser.
- Une interface de communication placée entre le module de régression linéaire et le modèle de simulation. Elle corrige les valeurs des paramètres retournées à chaque sortie de module.

Les données en sortie correspondent aux :

- valeurs simulées par le modèle,
- valeurs optimisées des paramètres,
- valeurs observées,
- résultats statistiques fournis par le module de régression non-linéaire.

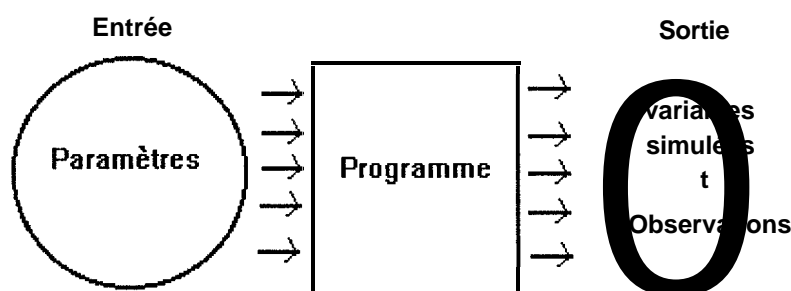


Schéma du principe général

4 - ORGANISATION DE L'APPLICATION

4.1 - Généralités

L'utilisateur fournit les observations de terrain.

Le module de régression non-linéaire envoie les valeurs optimisées (ou en cours d'optimisation) des paramètres et le modèle ARABHY fournit les valeurs simulées à comparer aux valeurs observées.

Pour accélérer l'application, tous les modules vont résider en mémoire et un tableau global sera créé, dans lequel seront stockées toutes les valeurs de paramètres.

Afin de différencier les paramètres modifiables aux paramètres à optimiser, le signe "?" sera ajouté aux paramètres à optimiser.

Le nombre d'observations de terrain doit être supérieur au nombre de paramètres à optimiser.

4.2 - Dictionnaire

4.2.1 - Les modules

Module général: MODGEN

interface de saisie: INTPARA

Module de régression : **NONLIN**

Modèle de simulation : ARABHY

Interface de communication : INTCOMM

4.2.2 - Les fichiers

Libellé	Contenu
FicVar	caractéristiques variétales.
FicSol	caractéristiques d'un sol.
FicSite	caractéristiques d'un site.
FicReg	caractéristiques de la région.
FicPluie	quantité de pluie reçue en un jour dans chaque site.
FicUtil	paramètres de base de la simulation.
FicObs	observations de terrain pour chaque site.

Les paramètres de base sont:

- nomsite (nom du site),
- variete (nom de la variété),
- type de sol (Dior, dior deck, dieri, deck),
- semis (date),
- fin de simulation (date).

4.2.3 - Les tableaux de travail

PARABHY : tableau de tous les paramètres utilisés dans le programme. Sa dimension maximale est égale au nombre de paramètres.

TabVar : tableau des valeurs des variables simulées par le modèle de simulation à comparer aux observations de terrain.

TabRes : tableau des résultats de simulation. Sa dimension maximale égale le nombre de paramètres en sortie demandés par l'utilisateur.

G : tableau des observations de terrain. Sa dimension maximale égale le nombre d'observations sur le terrain.

Q : tableau des dates d'observations (le nombre de jours après semis). Sa dimension maximale égale le nombre d'observations.

INDPAR : tableau des indices des paramètres à optimiser. Sa dimension maximale est égale le nombre de paramètres à optimiser.

R : tableau des valeurs des paramètres à optimiser. Sa dimension maximale égale le nombre de paramètres à optimiser. Ces valeurs sont extraites par le contenu du tableau des indices (INDPAR).

S : tableau des variables en sortie du modèle de régression non linéaire. Sa dimension maximale égale le nombre d'observations sur le terrain.

4.2.4. - Les paramètres et les variables

a) Les paramètres

Les paramètres sont des données en entrée nécessaires à la simulation.

Il existe trois catégories de paramètres:

- 0 Les paramètres de base : ils sont constants, une fois choisis par l'utilisateur ne doivent plus être changés.
- O Les paramètres à optimiser : ces données peuvent être modifiées par le module de régression non-linéaire pendant l'optimisation. Ces paramètres sont nécessaires à la simulation.
- Les paramètres modifiables : ces données peuvent être modifiées par l'utilisateur avant la simulation. Ces paramètres sont nécessaires au fonctionnement du modèle de simulation.

a) Les variables

Les variables sont les données résultats du modèle de simulation et les données observées sur le terrain. Parmi ces variables figurent les variables de comparaison.

Les variables de comparaison sont comparées aux observations de terrain par le module de régression non-linéaire. Ces variables sont stockées dans le tableau TabVar.

4.2.5. - Les données

Impr : drapeau d'impression.

Le drapeau d'impression est baissé par l'interface de saisie (Impr = 0) et est levé par le modèle de régression non-linéaire (Impr=1) quand l'optimisation est complète.

n1 : nombre de paramètres à optimiser

n3 : nombre d'observations de terrain

NatSort : indice de la valeur de la variable simulée à comparer aux valeurs observees

5 - SPECIFICATIONS FONCTIONNELLES

5.1. - Spécification fonctionnelle des modules

Le programme est compose de 5 grands modules :

1. Module général

Le module général effectue les opérations suivantes

- Appel l'interface de saisie
- Appel de paramétrisation
- Sortie

2. Interface de saisie

L'interface assure l'édition des valeurs des paramètres. L'édition consiste à lire, afficher et donner la possibilité à l'utilisateur de modifier les valeurs des paramètres pour la simulation.

L'interface permet aussi de choisir les paramètres à optimiser.

En résumé, l'interface de saisie effectue les opérations suivantes:

- Editer les valeurs des paramètres de simulation
- Editer les valeurs des paramètres de régression non-linéaire
- Retour au module général

3. Module de paramétrisation

Ce module effectue les opérations suivantes

- Appel du module de régression non - linéaire .
- Appel de l'interface de communication
- Appel du modèle de simulation

4. Modèle de simulation du développement de culture
Le modèle effectue les opérations suivantes
 - Simulation du développement d'une culture
 - Sortie des résultats de simulation au cas où le drapeau d'impression est levé par le module de régression non-linéaire
 - Retour à l'interface de communication
5. Interface de communication.
L'interface de **communication** assure le passage des paramètres du module de régression au modèle de simulation. Elle stocke les valeurs optimisées des paramètres en vue d'une **simulation**.
Les opérations effectuées sont les suivantes
 - Stockage des valeurs optimisées pour la simulation
 - Appel du modèle de simulation
 - Retour au module de régression non-linéaire

5.2. - Les algorithmes

52.1 - Module général

- Appel de l'interface de saisie
- Appel du module de paramétrisation
- Stockage des valeurs simulées
- Sortie des résultats

5.2.2 - Interface de saisie

- Lire les données des fichiers FicVar, FicSol, FicSite, FicReg, FicPluie, FicUtil, FicObs
- Editer les valeurs des paramètres modifiables
- Stocker les valeurs des paramètres modifiés
- Editer les valeurs des **paramètres** à optimiser
- Marquer les paramètres à optimiser
- Lecture des paramètres' de simulation
- Retour au menu général

5.2.3 - Modèle de simulation du développement de la culture

Ce module lit les valeurs des paramètres dans le tableau des valeurs des paramètres globaux et simule le développement d'une culture (voir dossier d'analyse).

- Lecture des paramètres de simulation
- Simulation
- **Si** Impr = 1 **ALORS** { drapeau levé }
 - .. Edition des résultats de la simulation
- FIN SI**
- Retour à l'interface de communication

TabVar : tableau des valeurs des variables simulées par le modèle de simulation à comparer aux observations de terrain.

TabRes : tableau des résultats de simulation. Sa dimension maximale égale le nombre de paramètres en sortie demandés par l'utilisateur.

G : tableau des observations de terrain. Sa dimension maximale égale le nombre d'observations sur le terrain.

Q : tableau des dates d'observations (le nombre de jours après semis). Sa dimension maximale égale le nombre d'observations.

INDPAR : tableau des indices des paramètres à optimiser. Sa dimension maximale est égale le nombre de paramètres à optimiser.

R : tableau des valeurs des paramètres à optimiser. Sa dimension maximale égale le nombre de paramètres à optimiser. Ces valeurs sont extraites par le contenu du tableau des indices (INDPAR).

S : tableau des variables en sortie du modèle de régression non linéaire. Sa dimension maximale égale le nombre d'observations sur le terrain.

4.2.4. - Les paramètres et les variables

a) Les paramètres

Les paramètres sont des données en entrée nécessaires à la simulation.

Il existe trois catégories de paramètres:

- Les paramètres de base : ils sont constants, une fois choisis par l'utilisateur ne doivent plus être changés.
- e Les paramètres à optimiser : ces données peuvent être modifiées par le module de régression non-linéaire pendant l'optimisation. Ces paramètres sont nécessaires à la simulation.
- Les paramètres modifiables : ces données peuvent être modifiées par l'utilisateur avant la simulation. Ces paramètres sont nécessaires au fonctionnement du modèle de simulation.

a) Les variables

Les variables sont les données résultats du modèle de simulation et les données observées sur le terrain, Parmi ces variables figurent les variables de comparaison.

Les variables de comparaison sont comparées aux observations de terrain par le module de régression non-linéaire. Ces variables sont stockées dans le tableau **TabVar**.

4.2.5. - Les données

Impr : drapeau d'impression.

Le drapeau d'impression est baissé par l'interface de saisie (Impr = 0) et est levé par le modèle de régression non-linéaire (Impr=1) quand l'optimisation est complète.

n1 : nombre de paramètres à optimiser

n3 : nombre d'observations de terrain

NatSort : indice de la valeur de la variable simulée à comparer aux valeurs observées

5 - SPECIFICATIONS FONCTIONNELLES

5.1. - Spécification fonctionnelle des modules

Le programme est composé de 5 grands modules :

1. Module général

Le module général effectue les opérations suivantes

- Appel l'interface de saisie
- Appel de paramétrisation
- Sortie

2. Interface de saisie

L'interface assure l'édition des valeurs des paramètres. L'édition consiste à lire, **afficher** et donner la possibilité à l'utilisateur de modifier les valeurs des paramètres pour la simulation.

L'interface permet aussi de choisir les paramètres à optimiser.

En résumé, l'interface de saisie effectue les opérations suivantes:

- Editer les valeurs des paramètres de simulation
- Editer les valeurs des paramètres de régression non-linéaire
- Retour au module général

3. Module de paramétrisation

Ce module effectue les opérations suivantes

- Appel du module de régression non - linéaire
- Appel de l'interface de communication
- Appel du modèle de simulation

5.2.4 - Régression non - linéaire

Ce module permet d'optimiser les valeurs de certains paramètres d'un modèle de simulation.

- .0 Lecture des variables en entrées
- Optimisation des paramètres
- .0 Appel de l'interface de communication

Tableaux des résultats statistiques

1) Tableau des valeurs initiales

observations
valeurs initiales des paramètres
somme des carrés des valeurs des paramètres initiaux
somme des carrés des y

2) Tableau des valeurs au cours des itérations

numéro d'itération
valeurs des déterminants
*somme des carrés au minimum
coefficient de détermination R^2
variance des résidus
ecart-type des résidus
nombre de degrés de liberté
valeur des paramètres estimés
estimation de l'écart-type de chaque valeur des paramètres
intervalle de confiance asymptotique de chaque valeur des paramètres
estimation de la matrice de corrélation de chaque valeurs des paramètres

* C'est la somme des carrés des termes d'erreur quand les erreurs sont minimales (voir "STATITCF Comment interpréter les résultats d'une régression linéaire R.TOMASSONE I.N.A. • P.G. Mathématique-Informatique-Biométrie Février 89").

3) Tableau des valeurs ajustées

Il s'agit d'un tableau des valeurs observées et ajustées avec un intervalle de confiance approximatif au niveau de confiance de 95%.

Nom variable	Signification
n°	d'itération
JAS	date d'observation
yobs	observations
vest	estimation
résidu	estimation des termes d'erreur
vinf	valeur inférieure
ysup	valeur supérieure

52.5 - Interface de communication

L'interface de communication doit connaître les valeurs des paramètres en cours d'estimation, et les valeurs des paramètres simulés à comparer aux observations.

Pour cela, deux tableaux et certaines variables vont être créés.

- Un tableau des indices des paramètres dont les valeurs doivent être optimisées.
- Un tableau des observations de terrain.
- Des variables intervenant dans le fonctionnement du module de régression non-linéaire. Ces variables sont :
 - le nombre de paramètres à optimiser
 - l'indice de la valeur simulée à comparer aux observations
 - le drapeau d'impression

En résumé l'interface effectue les opérations suivantes:

{Recopie les n1 paramètres contenus dans le tableau R vers le tableau PARABHY aux endroits désignés par IndPar}

```
•• Pour Natsort variant de 1 à n1
    j ← ← IndPar( Natsort )
    PARABHY( j ) ←← R( NatSort )
```

•• Fin boucle

•• Appel du modèle de simulation

•• Réinitialisation du tableau S

• Retour au module de régression non - linéaire

6 - LES ECRANS

Deux écrans principaux sont ajoutés à ceux du modèle de simulation:

- Ecran de visualisation des valeurs des paramètres lus par le module d'interface de saisie. Il permet à l'utilisateur de modifier les valeurs des paramètres de simulation.
- Ecran d'affichage des valeurs des paramètres optimisables. Celui ci permet à l'utilisateur de choisir ses paramètres à optimiser.

7 - CAS PARTICULIER

Pour une première version :

L'indice maximal du tableau PARABHY sera 20, cependant cette valeur peut être changée s'il existe des paramètres omis ou en doublon.

La valeur maximale du nombre de paramètres à optimiser (nl) sera 10. Cependant pour cette version, le nombre de paramètres à optimiser est 2. Les éléments de tableau seront les valeurs du taux de couverture simulé et du poids sec simulé

La taille du tableau des indices des paramètres (INDPAR) à optimiser est 2.

L'indice des valeurs simulées, (NatSort), à comparer aux valeurs observées est soit égal à 1 s'il s'agit du pourcentage de sol couvert soit égal à 2 s'il s'agit du poids sec total.

Le nombre d'observations (n3) est limitée au maximum à 6.

NB : il faut pas confondre les paramètres à modifier (exemple densité de semis, date de semis...) et les paramètres à optimiser (exemple: taux de couverture).

8 - COMMENT PARAMETRISER UN AUTRE MODELE OU UNE AUTRE FONCTION

Pour paramétrer une autre fonction non-linéaire, il faut :

- Disposer de la fonction à paramétrer.
- Disposer du module de régression.
- Remplacer le modèle de simulation de culture par la fonction ou par le modèle à paramétrer.

- Définir les paramètres dont les valeurs sont à optimiser (nl) c'est à dire les variables résultats de la fonction à paramétriser. Ces données vont servir d'éléments de comparaison des modules.
- Définir le nombre d'observations (n3), et les observations.
- Remplir le tableau des indices de paramètres (INDPAR).
- Définir l'indice de la variable résultat du modèle à comparer aux observations (NatSort).
- Initialiser les tableaux G (nombre d'observations), R (nombre de paramètres à optimiser).

Après initialisation des variables et des tableaux ci-dessus, le module de régression peut être appelé. Il envoie en sortie le tableau S dont les éléments sont les valeurs optimisées (ou en cours d'optimisation) des paramètres.