

ALGORITHME DU SYSTEME DE KARL PEARSON ET APPLICATION EN TELEDETECTION

PERAKIS Konstantinos
 Laboratoire de Télédétection - C.A.M.S.
 ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
 54, Boulevard Raspail, Paris 75006, FRANCE
 Tel : 49542169, Fax : 49542109

But :

Le développement de l'algorithme de l'application du système de distributions de K. Pearson est le thème essentiel de cette étude. Ayant exposé une analyse sommaire portant sur l'équation différentielle de Karl Pearson, nous classons les solutions (fonctions de densité de probabilité de distributions) issues de celle-ci. L'algorithme adoptée aboutissant à un programme met en oeuvre la méthode de K. Pearson pour l'ajustement des distributions observées à de distributions théoriques (issues de la solution de cette équation différentielle). Une application est testée sur deux images SPOT. Celles-ci permettent d'observer deux différents types de cultures du département de la Macédoine en Grèce du Nord. Les résultats de l'application du programme créé sont aussi exposés.

Abstract :

The purpose of this work is to adress the main aspects of developing an algorithm for the application of the K. Pearson's distribution system. The solutions to the differential equation corresponding to the system (in terms of probability density functions) are classified according to the results issued from hers resolution brief analysis.

The adopted algorithm and the relative software system uses the K. Pearson's method to fit the theoretical distributions (as they result from the solution of the differential equation) to the empirical ones. The algorithm is tested through an application on two SPOT images concerning two different cultivations in Macedonia, North of Greece. The results of this application using the specific software system are also presented.

MOTS CLES : Algorithme, système de K. Pearson, ajustement, distributions, Grèce, agriculture.

1. SYSTEME DE DISTRIBUTIONS DE K. PEARSON

1.1 Equation différentielle de Karl Pearson

La recherche d'une équation différentielle dont les solutions particulières sont des fonctions de densité des distributions se base sur l'interprétation mathématique des formes de courbes correspondant à ces dernières (Elderton, 1906).

En ajustant la distribution hypergéométrique K. Pearson a obtenu l'équation différentielle, suivante :

$$\frac{dy}{y} = - \frac{(x+b)dx}{(ax^2+xc)} \quad (1)$$

Nous l'appelons équation différentielle de K. Pearson (nous la désignerons par la suite E.D.P.).

A partir de l'E.D.P., nous pouvons arriver à une relation de récurrence entre les moments m_r ($r=1$ à 6) de différents ordres (Pérakis 1992) :

$$m_r [(n+1)a-1] + m_{r-1} (n-1)b + m_{r-2} (n-1)c = 0 \quad (2)$$

A partir de la relation (2), nous pouvons extraire des relations entre les paramètres a , b et c du trinôme du dénominateur de l' E.D.P. et les moments centrés (avec $m_1 = 0$) jusqu'à l'ordre quatre.

$$a = \frac{[2m_2 m_4 - 6m_2^3 - 3m_3^2]}{D} \quad (3.a)$$

$$b = \frac{[m_3 (m_4 + 3m_2^2)]}{D} \quad (3.b)$$

$$c = \frac{[m_2 (4m_2 m_4 - 3m_3^2)]}{D} \quad (3.c)$$

$$\text{où } D = 10m_2 m_4 - 18m_2^3 - 12m_3^2$$

Nous pouvons aussi exprimer les coefficients a , b et c , en fonction des caractéristiques de la courbe ; soit :

- la dispersion

$$\sigma = \sqrt{m_2} \quad (4.a)$$

- le coefficient d'asymétrie :

$$k^2 = \frac{m_3^2}{m_2^3} + 4, \quad (4.b)$$

- le coefficient d'aplatissement :

$$\varepsilon = \frac{m_4}{m_2^2} + 3 \quad (4.c)$$

Nous avons :

$$a = \frac{2\varepsilon - 3k^2}{10\varepsilon - 12k^2} \quad (5.a)$$

$$b = \frac{\varepsilon \sqrt{k^2 - 4}}{10\varepsilon - 12k^2} \delta \quad (5.b)$$

$$c = \frac{4\varepsilon - 3k^2}{10\varepsilon - 12k^2} \delta^2 \quad (5.c)$$

Ainsi, nous exprimons les coefficients du trinôme en fonction des caractéristiques de la distribution et, par conséquent, des moments de la distribution.

1.2 Solutions de l'équation différentielle de K. Pearson

1.2.1 Racines réelles de signe opposé.

Soient $-x_1, x_2$ les racines du trinôme. Après un changement d'origine, l'E.D.P. s'écrit :

$$\frac{dy}{y} = - \frac{(x+b)dx}{a(x+x_1)(x-x_2)} \quad (1)$$

Pour intégrer cette équation différentielle, nous introduisons la nouvelle variable $z = x+b$ et les racines du nouveau trinôme $r_1 = x_1 - b$ et $r_2 = x_2 + b$. De cette façon, $x+x_1 = z+r_1$, $x-x_2 = z-r_2$ et $dz = dx$. L'équation différentielle s'écrit :

$$\frac{dy}{y} = - \frac{zdz}{a(z+r_1)(z-r_2)} \quad (2)$$

En intégrant sur le domaine de définition de la variable z , nous avons :

$$y = y_0 \left| \frac{z}{r_1} + 1 \right|^{-vr_1} \left| -\frac{z}{r_2} + 1 \right|^{-vr_2} \quad (3)$$

où
$$v = \frac{1}{a} \frac{1}{r_1 + r_2}$$

L'équation (3) caractérise les différents types de distributions qui appartiennent à la famille des distributions issue des racines réelles du trinôme de l'équation de K. Pearson. Ceux-ci sont différenciés en fonction des valeurs des exposants $-vr_1$ et $-vr_2$. Nous remarquons que la valeur limite inférieure pour les deux exposants

est -1 ; sinon l'intégrale $\int_{-r_1}^{r_2} ydz$ ne convergerait

pas.

D'après l'étude faite par K. Pearson, l'intervalle entre les deux racines $-r_1$ et r_2 : $-r_1 < z < r_2$ est considéré comme le domaine de définition de la variable z (par conséquent pour la variable initiale x nous avons $-x_1 < x < x_2$).

Après avoir effectué le changement de variable $z = x + b$ les nouvelles racines $-r_1$ et r_2 ne sont pas obligatoirement de signe opposé.

Trois cas principaux sont à distinguer (notés I, II et III).

I. Les nouvelles racines $-r_1$ et r_2 sont

toutes les deux négatives : $-r_1 < r_2 < 0$.

L'équation (3) s'écrit :

$$y = y_0 \frac{1}{r_1^{-vr_1} (-r_2)^{-vr_2}} (z+r_1)^{-vr_1} (r_2-z)^{-vr_2} \quad (4)$$

La constante d'intégration se calcule à l'aide

de la contrainte : $\int_{-r_1}^{r_2} y dz = 1$.

Elle est égale à :

$$y_0 = \frac{(r_1 + r_2)^{v(r_1 + r_2) - 1}}{r_1^{vr_1} (-r_2)^{vr_2}} \frac{1}{B(1-vr_1, 1-vr_2)} \quad (5)$$

Pour les exposants $-vr_1$ et $-vr_2$, les cas suivants se distinguent (ils sont codés par les chiffres italiques 1, 2, 3, ...) :

$$-1 < -vr_1 < 0 \quad \text{et} \quad 0 < -vr_2 < 1 \dots \dots \dots \text{cas } 1$$

$$-1 < -vr_1 < 0 \quad \text{et} \quad -vr_2 > 1 \dots \dots \dots \text{cas } 2$$

$$-1 < -vr_2 < 0 \quad \text{et} \quad 0 < -vr_1 < 1 \dots \dots \dots \text{cas } 3$$

$$-1 < -vr_2 < 0 \quad \text{et} \quad -vr_1 > 1 \dots \dots \dots \text{cas } 4$$

II. Les nouvelles racines $-r_1$ et r_2 sont de signe opposé : $-r_1 < 0, r_2 > 0$.

L'équation (3) s'écrit :

$$y = y_0 \frac{1}{r_1^{-vr_1} r_2^{-vr_2}} (z+r_1)^{-vr_1} (r_2-z)^{-vr_2} \quad (6)$$

En effectuant les même transformation de variable que précédemment nous trouvons la constante d'intégration :

$$y_0 = \frac{(r_1 + r_2)^{v(r_1 + r_2) - 1}}{r_1^{vr_1} r_2^{vr_2}} \frac{1}{B(1-vr_1, 1-vr_2)} \quad (7)$$

Pour les exposants $-vr_1$ et $-vr_2$ les cas suivants se distinguent :

$$-1 < -vr_1 < 0 \quad \text{et} \quad -1 < -vr_2 < 0 \dots \dots \dots \text{cas } 5$$

$$0 < -vr_1 < 1 \quad \text{et} \quad 0 < -vr_2 < 1 \dots \dots \dots \text{cas } 6$$

$$0 < -vr_1 < 1 \quad \text{et} \quad -vr_2 > 1 \dots \dots \dots \text{cas } 7$$

$$-vr_1 > 1 \quad \text{et} \quad 0 < -vr_2 < 1 \dots \dots \dots \text{cas } 8$$

$$-vr_1 > 1 \quad \text{et} \quad -vr_2 > 1 \dots \dots \dots \text{cas } 9$$

III. Les nouvelles racines $-r_1$ et r_2 sont positives et de même signe ; l'expression (3) s'écrit :

$$y = \frac{(r_1 + r_2)^{v(r_1 + r_2) - 1}}{B(1-vr_1, 1-vr_2)} (z+r_1)^{-vr_1} (r_2-z)^{-vr_2}$$

Les différents cas possibles pour les exposants $-vr_1$ et $-vr_2$ coïncident avec ceux du

cas où les deux racines $-r_1$ et r_2 sont négatives.

- La classification des différents types de courbes se base sur la valeur des exposants de l'équation (3). Chacun des exposants $-vr_1, -vr_2$ peut varier dans les intervalles suivants: $[-1,0]$, $[0,1]$ et $[1,+\infty]$.

- Parmi les neuf cas présents, certains se recoupent :

- Le cas 1 coïncide avec le cas 3
- le cas 2 coïncide avec le cas 4
- et le cas 7 coïncide avec le cas 8.

Finalement, six types, appelés désormais S1 à S6, sont retenus :

- S1 : pour les variations $-1 < -vr_1 < 0, -1 < -vr_2 < 0$,
- S2 : pour $0 < -vr_1 < 1, -1 < -vr_2 < 0$,
- S3 : pour $0 < -vr_1 < 1, 0 < -vr_2 < 1$,
- S4 : pour $-1 < -vr_1 < 0, -vr_2 > 1$,
- S5 : pour $0 < -vr_1 < 1, -vr_2 > 1$,
- S6 : pour $-vr_1 > 1, -vr_2 > 1$.

1.2.2 Racines réelles de même signe.

Pour être cohérent avec le cas des racines réelles de signe opposé, nous les appelons $-x_1$ et $-x_2$, avec x_1, x_2 positifs. Sans restreindre la généralité, nous supposons que $-x_1 < -x_2$; le cas $-x_1 > -x_2$ est similaire; l'E.D.P. devient :

$$\frac{dy}{y} = - \frac{(x+b)dx}{a(x+x_1)(x+x_2)} \quad (1)$$

Pour intégrer cette équation, nous effectuons le changement de variable suivant :

$$z = x + x_1,$$

alors $dz=dx$, $x+b=z-x_1+b$ et $x+x_2=z+x_1-(x_1-x_2)=z-c$ où nous avons posé $c = (x_1-x_2) > 0$.

L'équation (1) s'écrit :

$$\frac{dy}{y} = - \frac{(z-x_1+b)dz}{az(z-c)}.$$

Finalement nous avons une équation de la forme :

$$y = k |z|^r |z-c|^{-s}, \quad (2)$$

$$\text{où } r = \frac{1}{a} \frac{b-x_1}{x_1-x_2} \quad \text{et} \quad s = \frac{1}{a} \frac{b-x_2}{x_1-x_2} \quad (3)$$

et la constante d'intégration : $k = \frac{c^{s-r-1}}{B(r+1, 1-s)}$

Selon l'étude faite par K. Pearson, nous allons considérer comme domaine de définition de la variable z l'intervalle $[c, \infty[$ lorsque c est positif et $]-\infty, c]$ lorsque c est négatif. Pour prendre en compte la forme de la courbe représentée par l'équation (2), nous différencions les cas suivants : $-s < 0$, $0 < -s < 1$ et $s > 1$.

Nous établissons les correspondances suivantes :

Dans le cas où $-s < 0$, : type S7,

dans le cas où $0 < -s < 1$: type S8,

dans le cas où $-s > 1$: type S9.

1.2.3 Racines complexes

C'est le cas où le discriminant b^2-4ac du trinôme est négatif. En effectuant le changement de variable $z = x + \frac{b}{2a}$, l'équation différentielle de Pearson devient:

$$\frac{dy}{y} = - \frac{z-t}{z^2+1^2} dz \quad \text{où } t = \frac{b}{2a} (1-2a) \text{ et}$$

$$l^2 = \frac{4ac-b^2}{4a^2} \quad (1)$$

En intégrant, l'équation de la famille des courbes correspondant aux racines complexes du trinôme est obtenue :

$$y = y_0 \left(1 + \frac{z^2}{l^2} \right)^{-\frac{1}{2a}} e^{-\frac{t}{al} \arctg \frac{z}{l}} \quad (2)$$

Pour calculer la constante d'intégration,

nous partons de la contrainte $\int_{-\infty}^{+\infty} y dz = 1$ (3)

La constante d'intégration est :

$$y_0 = \frac{1}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{\frac{1}{a}-2} \varphi e^{\frac{t\varphi}{al}} d\varphi} \quad (4)$$

Le cas de racines complexes est le type S10.

2. ALGORITHME ET PROGRAMMATION DU SYSTEME DE K. PEARSON

L'application du système de K. Pearson nécessite la conception d'un algorithme aboutissant à un programme informatique; cet aspect est présenté dans la suite.

L'étape initiale consiste dans l'enregistrement et la mise en forme des données pour l'application de l'algorithme proposé. La construction d'un tableau caractérisant chaque distribution facilite le calcul de ses paramètres (qui sont des fonctions des moments jusqu'à l'ordre quatre, de la distribution étudiée). Ensuite la courbe étudiée est classée d'après l'examen des différents cas du système de Pearson: des racines réelles de signe opposé ou de même signe des racines complexes.

2.1 Enregistrement des données.

Le stockage des données (valeurs radiométriques) s'effectue dans un vecteur (variable initiale). Nous calculons ensuite l'effectif total et la moyenne.

2.2 Tri croissant de la variable d'entrée.

La méthode des moments, utilisée par K. PEARSON, ne peut s'appliquer que dans le cas où les réalisations de la variable à étudier sont rangées selon un ordre croissant. C'est la troisième étape du prétraitement informatique.

2.3 Tableau caractéristique, moments observés et coefficients d'asymétrie et d'aplatissement.

A partir du vecteur précédent (cf. §3.1) nous construisons un tableau utilitaire et indicatif de la distribution réelle dont nous calculons les moments. Le tableau contient six colonnes (dans le cas de canaux bruts les dimensions du tableau sont (256x6)) :

- la première représente la modalité de la variable aléatoire.
- la deuxième son effectif.
- la troisième sa valeur centrée.
- les quatrième, cinquième et sixième, sont les valeurs centrées à la puissance 2, 3 et 4.

Schématiquement, ce tableau a la forme suivante :

| | | | | | |
|---|---|-------------|-----------------|-----------------|-----------------|
| . | | | | | |
| x | y | $x-\bar{x}$ | $(x-\bar{x})^2$ | $(x-\bar{x})^3$ | $(x-\bar{x})^4$ |
| . | | | | | |

Une autre procédure informatique qui succède à la précédente effectue la somme en colonnes, des quatre dernières colonnes du tableau créé. En divisant chacune de ces sommes par l'effectif total, nous obtenons la moyenne centrée (donc nulle), la variance et les moments centrés d'ordre

3 et 4. Ce sont les moments centrés de la distribution observée. En considérant que ceux-ci sont égaux à ceux d'une distribution théorique, nous recherchons cette dernière après l'analyse de l'équation différentielle de K. PEARSON. Les caractéristiques d'une distribution sont des fonctions de ses moments. Ainsi, nous calculons

les coefficients d'asymétrie $k^2 = \frac{m_3^2}{m_2^3} + 4$ et

d'aplatissement $\epsilon = \frac{m_4}{m_2^2} + 3$ de la distribution.

Nous établissons un second tableau contenant les observations avec les fréquences relatives. Celui-ci, indicatif de la distribution, met en évidence la répartition des fréquences des valeurs radiométriques et la correspondance de chacune d'entre elles avec sa valeur centrée et réduite. Il contient ou non les modalités observées (selon la réponse de l'utilisateur).

2.4 Algorithme de détermination des courbes Pearsoniennes issues des classifications.

Ayant calculé les moments de la distribution étudiée, nous procédons au calcul des coefficients a, b et c du trinôme du dénominateur de E.D.P.

La valeur du discriminant du trinôme et des coefficients a et b déterminent l'appartenance de la courbe étudiée à la grande classe des racines réelles. Pour cela, il faut que le discriminant soit positif et que les coefficients a et b se trouvent hors d'un petit intervalle défini autour de zéro (par ex. $[10^{-2}, 10^1]$). Si les valeurs de a et b sont incluses dans cet intervalle, nous

classons la distribution en tant que normale ; sinon les racines du trinôme sont calculées.

1. Si le produit des racines est un nombre négatif, nous sommes dans le cas des racines réelles de signes opposées. A la suite du calcul des racines du trinôme du dénominateur $-x_1$ et x_2 ,

nous distinguons deux cas :

a) $|x_1| < |x_2|$ et

b) $|x_1| > |x_2|$.

Sans restreindre la généralité, nous interchangeons les valeurs de x_1 et x_2 , de façon que $|x_1| < |x_2|$ (le trinôme du dénominateur étant symétrique par rapport à $-x_1$ et x_2).

Les racines sont calculées à partir des coefficients a, b et c du trinôme (fonctions des moments). Pour tester le calcul fait jusqu'ici, nous vérifions que le trinôme de coefficients a, b, c s'annule pour les valeurs $-x_1$, x_2 . Ensuite, les racines correspondant $-r_1$ et r_2 de la variable transformée z sont calculées. Afin de classer la distribution étudiée parmi les différents types définis (partie théorique ci-dessus) nous évaluons

les quantités v, vr_1 et vr_2 . Avant de rechercher les ordonnées de la courbe théorique, nous établissons les moments de la distribution de la variable z par rapport au point $z = -r_1$. Les

moments centrés correspondant ne sont pas calculés à ce niveau ; les mêmes formules étant valables dans le cas des racines réelles de même signe (minimisation du coût de calcul). Puisque les deux cas sont exclusifs (racines réelles de signe opposé ou du même signe), les moments théoriques centrés sont établis après le calcul des moments translatés dans le cas de racines réelles de même signe.

L'étape suivante consiste à classer la distribution étudiée parmi les six types S1 à S6. Elle se réalise à l'aide d'un test qui se base sur les valeurs des exposants $-vr_1$ et $-vr_2$.

Pour cette famille de courbes, il ne reste plus qu'à calculer les ordonnées de la courbe théorique dont l'équation est :

$$y = \frac{(r_1+r_2)^{v(r_1+r_2)-1}}{B(1-vr_1, 1-vr_2)} (r_1+z)^{-vr_1} (r_2-z)^{-vr_2}$$

Puisque la variable z est issue de la variable centrée initiale, x nous calculons pour toutes les modalités de la variable x les valeurs correspondantes de la variable z ($z=x+b$). Notons que pour le calcul informatique les variables x, z et y sont des tableaux de dimension lxl où l est le nombre des modalités de la variable x. La fonction Beta, ayant comme arguments $1-vr_1$ et $1-vr_2$, est calculée grâce à sa propriété :

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

La fonction Gamma qui intervient dans cette dernière formule fait partie des fonctions standard de la librairie mathématique du langage C (langage C sous UNIX Norme ANSI.). Sa valeur, pour un argument donné, peut être obtenue directement.

Ainsi ayant déterminé toutes les grandeurs nécessaires, nous procédons au calcul des

ordonnées de la courbe théorique. Puisqu'il s'agit d'une courbe continue, il ne faut pas se contenter de calculer les ordonnées correspondant aux différentes valeurs radiométriques observées (abscisses). Après avoir effectué différents essais afin d'obtenir une représentation graphique acceptable, nous avons utilisé comme pas pour la variable le millième de l'intervalle de son

domaine de définition, c.a.d. $\frac{\max - \min}{1000}$.

2. Si le produit des racines est un nombre positif, nous sommes dans le cas des racines réelles du même signe. Les racines étant $-x_1$ et

$-x_2$ et, après avoir interchangé les valeurs des racines de telle sorte que $-x_1 < -x_2$ (sans restreindre la généralité en supposant que x_1, x_2 positifs), nous calculons les quantités $c=x_1-x_2, r$ et s . Ensuite, la translation $x_1-c=x_2$ est

conservée pour une utilisation ultérieure. Nous calculons ensuite les moments par rapport au point c . Les moments centrés correspondants ne sont pas calculés ici (même raison que pour la famille S1-S6).

Finalement, pour la recherche des ordonnées de la courbe théorique, nous calculons d'abord la constante d'intégration k . Pour évaluer la fonction Beta, nous utilisons sa relation avec la fonction Gamma. Afin d'éviter la création et le calcul de variables intermédiaires, nous déterminons, pour toutes les modalités de la variable x , les quantités $x+x_1 (=z)$ et $x+x_2 (=z-c)$.

3. Si le discriminant du trinôme est négatif, nous sommes en présence de racines complexes (courbe en forme de cloche et infinie de deux cotés).

Pour toutes les modalités de la variable x , nous calculons les valeurs correspondantes de la variable transformée z . Nous évaluons les quantités t et l en fonction des coefficients a, b et c . Le calcul de la constante d'intégration y_0 ne se fait pas de la même façon pour les familles précédentes du système. Il s'effectue par intégration numérique. A cause de sa relative simplicité, nous avons choisi pour cela la méthode des parallélogrammes.

2.5 Conclusion sur l'algorithmie.

Etant donnée la multitude des cas de distributions intégrés dans le système de K. Pearson, l'outil informatique facilite beaucoup l'application de sa méthode.

De plus, il permet la manipulation de grandeurs intervenant dans le calcul de la courbe théorique pouvant atteindre de très grandes valeurs (p. ex. de l'ordre de 10^{250}).

Dans l'algorithme proposé, des adaptations amenées par certains éléments ont été effectuées :

1. La similitude de certains cas (racines réelles de signe contraire, ou de même signe) réduit le nombre des cas réellement distincts. Il s'ensuit une simplification du programme proposé.

2. Deux orientations sont envisagées pour le calcul de certaines fonctions telles que la fonction Gamma :

- Soit nous utilisons la fonction fournie par la bibliothèque interne au langage de programmation adopté (le langage C) et nous perdons ainsi la portabilité du programme,

- Soit nous approximations la valeur de la fonction par des méthodes d'intégration numériques. Dans ce cas, nous diminuons la précision du calcul. Une comparaison entre les

valeurs de la fonction Gamma trouvées par la méthode de parallélogrammes et celles fournies par la bibliothèque mathématique du langage C a fourni une concordance de l'ordre de 10^{-3} .

3. L'application de méthodes d'approximation numérique est inévitable lorsque les fonctions ne sont pas incorporées aux bibliothèques mathématiques : c'est par exemple le cas de la

fonction $F(a,b) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^a \phi e^{b\phi} d\phi$ nécessaire pour

le calcul de la constante d'intégration dans le cas de racines complexes.

3. APPLICATION DE LA METHODE

Dans le cadre de la mise en oeuvre opérationnelle de la méthode vue précédemment, une application est testée. Elle s'opère sur un échantillon de parcelles représentatif d'une région agricole située dans la plaine d'Axios (Macédoine) en Grèce du Nord. Les documents de télédétection font partie des sources d'information privilégiées et les différentes longueurs d'onde (canaux) constituent les variables analysées.

3.1 Les données de télédétection.

Des scènes SPOT (Système Probatoire d'Observation de la Terre) sont disponibles pour les dates du 11 Mai et du 9 Juin 1987 pour la plaine d'Axios.

3.2 Les données terrain.

Sur la plaine d'Axios, 3 grands sites et 5 grands champs isolés ont été sélectionnés selon un choix raisonné (Silleos N. 1990).

3.3 Le fichier statistique.

Il associe la modalité de la variable thématique (culture) et les valeurs observées des variables radiométriques recueillies pour chaque parcelle. Un échantillon qualitatif, mais néanmoins représentatif des grandes cultures, est établi pour la plaine d'Axios.

Dans la présente étude, nous nous intéressons à la culture du riz.

3.4 Applications sur l'agricole.

La méthode de K. Pearson permet d'approfondir l'information des variables (canaux) portant sur les cultures ; ceci révient à fournir la loi statistique régissant l'échantillon de la culture étudiée.

3.4.1 Le comportement du riz de la plaine d'Axios de la Macédoine.

(Tableau 1 - Figures 1 à 6).

CALENDRIER AGRICOLE.

SEMIS APRES INONDATION DES PARCELLES : du 5 Avril au 5 Mai.
DEBUT DE MONTAISON : du 5 Mai au 5 Juin.

DEVELOPPEMENT : fin Juillet.
 FLORAISON : jusqu'au fin Septembre.
 RECOLTE : du fin Septembre au fin Octobre.

Mois de Mai.

L'amplitude de l'intervalle des valeurs est du même ordre de grandeur pour les 3 canaux. Cependant, les minima et les maxima varient. Le canal 1 montre des valeurs faibles mais son maximum d'amplitude reste normal, comparé aux autres canaux. Ce fait s'explique par la sensibilité de cette longueur d'onde aux variations d'épaisseur de la nappe d'eau et aux matières en suspension.

Dans le cas analysé, le canal 1 présente plusieurs modes (de l'ordre de 7) traduisant des différences d'irrigation et de travail de la terre. Le canal 2, moins contrasté, admet le même nombre de modes. Mais sa forme de distribution diffère et l'amplitude des modes est hétérogène; les propriétés mises en évidence par le canal 2 de SPOT se référant aux activités biologiques du type zooplanctonique et d'oxygénation. La petite bosse en bas (à gauche ou à droite de la figure X) indique une forte activité chlorophyllienne. Celle-ci traduit la présence de rizières avec une certaine épaisseur et durée d'état d'eau.

L'autre extrême de la courbe traduit une activité chlorophyllienne peu marquée. Les rizières présentent une épaisseur d'eau faible ou nulle; le sol est alors seulement humide.

Entre les deux extrêmes, la grande masse des rizières à tendance gaussienne se regroupe. Le canal 3 est asymétrique et présente moins de modes, voir cinq.

Le canal 3 recouvre la bande de longueur d'onde dite d'absorption d'eau.

Les modes délivrent une information sur le rapport eau/végétation.

Ayant le moment d'ordre 3 négatif, la loi théorique pour le canal 1 est en forme de J; elle est du type S2. Etant donnée la multimodalité assez marquée l'ajustement n'est pas satisfaisant.

En se présentant sous forme de cloche, les lois théoriques qui ajustent les distributions empiriques de canaux 2 et 3 sont de type respectivement S6 et S10; pour le canal 2 la distribution ayant des valeurs des coefficients d'asymétrie et d'aplatissement ($k^2=4.04$ et $\epsilon=5.91$) proche de celles correspondant à la loi Normale et des pics locaux relativement petits, admet un ajustement qui n'est pas satisfaisant; de même pour le canal 3.

Mois de Juin

Dans un intervalle d'un mois, les réponses spectrales montrent une certaine uniformisation.

L'origine de ce fait réside dans le taux de recouvrement de 50% ou plus atteint par la plante. La multimodalité observée au mois de Mai disparaît; exception faite du canal 2; les deux modes de ce dernier correspondant à deux taux de recouvrement différents. Ils sont généralement asymétriques et seul un rapport du type eau + plantes existe.

Les ajustements dans les trois canaux se présentent sous forme de cloche. L'ajustement du canal 2 est satisfaisant pour les seuils 5% et 1%; il n'en est pas de mêmes pour les canaux 1 et 3. Les classes pearsonniennes pour les trois canaux sont S6, S6 et S5. La présence de petits groupes de pixels vers les valeurs fortes du domaine de définition de la distribution observée (36,71) fait que le domaine de définition de la distribution théorique est très allongé sur la droite de la distribution (39.1,213.1).

3.5 Conclusions sur les applications

Pour le cas de la culture du riz de la plaine d'Axios il est difficile de tirer de conclusions (pour les distributions des échantillons) concernant les relations pouvant exister entre les canaux de la même date. Pourtant, il est assez fréquent que les canaux 1 et 2 suivent la même loi théorique (cas des distributions du riz au mois de Juin). En définitif, pour obtenir des conclusions concernant les distributions des échantillons des cultures, il faut réaliser, si possible, un suivi diachronique (interannuel) en utilisant toujours les mêmes dates et les mêmes échantillons.

4. TABLEAUX ET FIGURES

4.1 Tableau des caractéristiques des distributions

Le tableau suivant présente les caractéristiques principales des distributions théoriques selon les ajustements effectués sur tous les échantillons disponibles de la plaine d'Axios en Macedoine (Grèce).

Pour chaque ajustement réalisé les résultats sont les suivants :

- μ : la moyenne de la distribution
 - σ : l'écart-type de la distribution
 - mod ou md : le nombre des modalités de la distribution
 - D1 : le domaine de définition de la distribution réelle
 - m3 : le moment d'ordre 3
 - k^2 : le coefficient d'asymétrie
 - ϵ : le coefficient d'aplatissement
 - r1 et r2 : les racines du trinôme de l'E.D.P.
 - D2 : le domaine de définition de la loi théorique
 - e1 et -e2 : les exposants de la loi théorique (ce sont les quantités $-vr_1$ et $-vr_2$ pour les classes S1 à S6)
 - r et -s : les exposants de la loi théorique (pour les classes S7 à S9)
 - S1 à S10 : la classe d'appartenance de la loi théorique par rapport au système de K.Pearson
 - d°lb : les degrés de liberté après regroupement
 - χ^2 : la valeur de la quantité de χ^2
- Il faut noter que si les degrés de liberté sont inférieurs à 1 la quantité du χ^2 n'est pas calculée.
- Pour éviter toute confusion sur les nombres réels la "," est notée ".".

TABEAU 1

LE RIZ - MACEDOINE

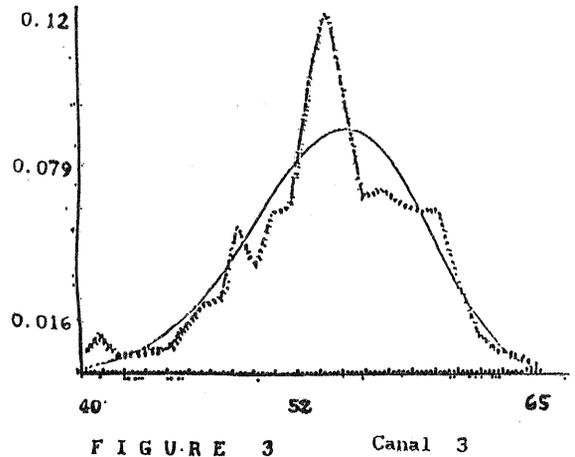
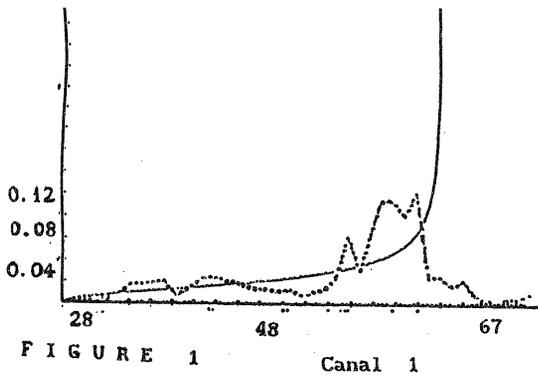
NOMBRE DE PIXELS : 940

| | CANAL 1 | CANAL 2 | CANAL 3 |
|-------------------|---|--|---|
| 11 MAI 1987 | $\mu=52.02, \sigma=8.58$ mod:41, D1=[29,69] m3=-539.372 $k^2=4.73 \quad \varepsilon=5.68$ r1=-24.0 r2=9.27 D2=]28.0, 61.29[-e1=0.46 -e2=0.44 type S2 $d^{\circ}lb:25, \chi^2=171.37$ | $\mu=64.28, \sigma=3.68$ mod:24, D1=[53,76] m3=-14.77 $k^2=4.09 \quad \varepsilon=6.5$ RACINES COMPLEXES D2=]-\infty, \infty[type S10 $d^{\circ}lb:15, \chi^2=45.404$ | $\mu=53.99, \sigma=4.68$ mod:27, D1=[40,66] m3=-47.57 $k^2=4.22 \quad \varepsilon=6.36$ r1=23.45 r2=148.8 D2=]-\infty, 202.83[r=-191.5-s=29.03 type S9 $d^{\circ}lb:17, \chi^2=85.529$ |
| 09 JUN 1987 | $\mu=54.77, \sigma=2.15$ mod:15, D1=[50,64] m3=7.36 $k^2=4.54 \quad \varepsilon=6.52$ r1=-4.56 r2=22.07 D2=]50.32, 76.84[-e1=2.39 -e2=15.8 type S6 $d^{\circ}lb:6, \chi^2=20.104$ | $\mu=40.5, \sigma=2.41$ mod:16, D1=[35,50] m3=2.73 $k^2=4.04 \quad \varepsilon=8.61$ r1=-6.21 r2=8.61 D2=]34.37, 49.19[-e1=2.44 -e2=3.77 type S6 $d^{\circ}lb:7, \chi^2=10.246$ | $\mu=46.98, \sigma=5.45$ mod:36, D1=[36,71] m3=207.997 $k^2=5.66 \quad \varepsilon=8.31$ r1=-7.88 r2=166.2 D2=]39.1, 213.14[-e1=0.95 -e2=40.2 type S5 $d^{\circ}lb:19, \chi^2=46.772$ |

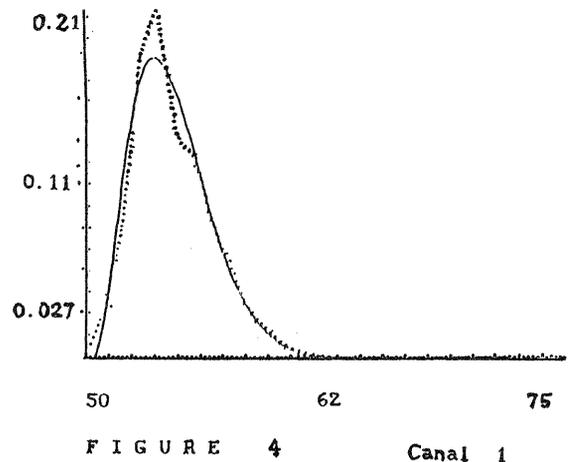
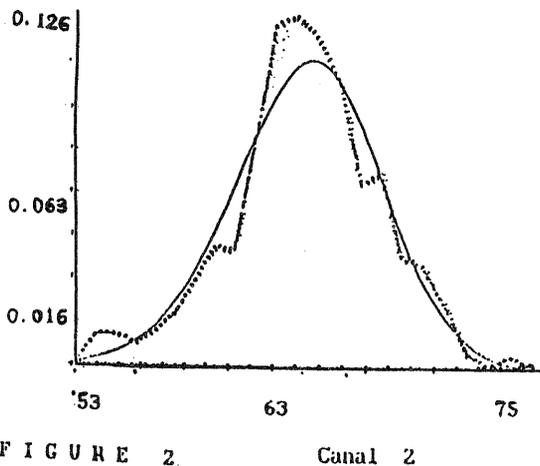
4.2 Figures des ajustements des distributions réelles par des lois théoriques de K. Pearson.

GRECE - MACEDOINE - PLAINE D'AXIOS
ECHANTILLON DU RIZ

11 MAI 1987



09 JUN 1987



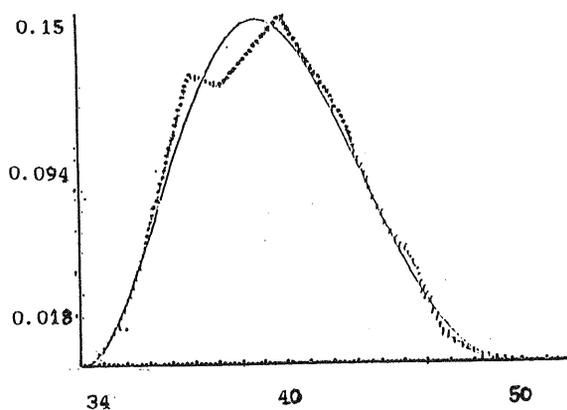


FIGURE 5 Canal 2

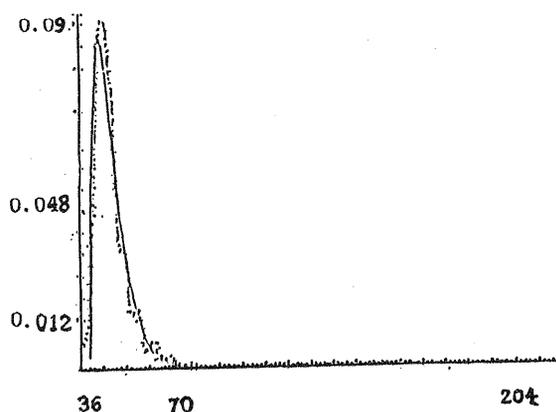


FIGURE 6 Canal 3

BIBLIOGRAPHIE

Elderton, P., 1906. Frequency curves and correlation. Charles and Edwin Layton, London, pp. 1-40.

Perakis, K., 1992 Méthodes statistiques de Karl Pearson en télédétection satellitaire. Thèse de doctorat. Ecole des Hautes Etudes en Sciences Sociales - Centre d'analyse et de mathématique sociales.

Silleos, N., 1990. Evaluation des données de la deuxième génération des satellites pour la construction rapide des cartes thématiques. Laboratoire de pédologie appliquée - Ministère de l'agriculture - Université d'Aristote de Salonique.