



## **OPTIMISATION DU RESEAU PLUVIOMETRIQUE HODNEEN PAR LA METHODE DES ALGORITHMES GENETIQUES SPATIALISES**

***GUETTOUCHE M.S.<sup>1</sup>, BOUKHETALA K.<sup>2</sup>, AIT-KACI S.<sup>2</sup>***

<sup>1</sup> Laboratoire de Géomorphologie, Faculté des Sciences de la Terre  
USTHB, Alger  
msguetouche61@hotmail.com

<sup>2</sup> Laboratoire de Probabilités et Statistiques, Faculté de Mathématique  
USTHB, Alger

### **RESUME**

Le problème traité est la diminution du nombre de stations d'observations pluviométriques d'un réseau installé dans le bassin du Hodna (Algérie), en choisissant les stations qui rendent le réseau le plus représentatif géographiquement, et cela pour alléger les coûts de fonctionnement et d'entretien du réseau. La formulation du problème ainsi énoncé a donné lieu à la recherche d'un partitionnement optimal d'un ensemble de N individus en k classe. Ce problème combinatoire demande l'application des algorithmes approchés pour le résoudre. Pour ce faire, une méthode de résolution a été élaborée et qui consiste en un algorithme génétique spatial (SGA). Pour finaliser, un logiciel interactif a été établi à l'aide de l'environnement de développement MATLAB. Le résultat de cette méthode est la déduction de cinq stations matérialisant la zone et représentant, en fait, une série de 37 stations pluviométriques implantées actuellement dans le bassin.

**Mots-clés** : Hodna, stations, optimisation, algorithme génétique spatialisé, localisation

## INTRODUCTION

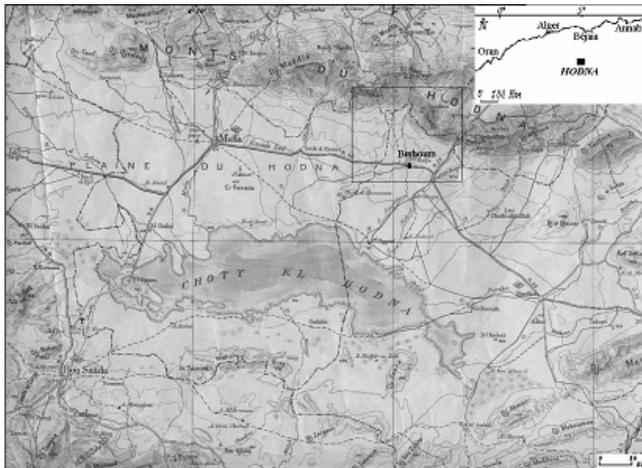
L'Algérie est dotée d'un réseau d'observations météorologiques caractérisé par des stations d'enregistrement manuel, basé sur l'observateur. Les stations sont dans la plupart des cas incomplètes et se résument en un pluviomètre seulement. L'optimisation d'un tel réseau est donc nécessaire et peut être justifiée pour deux raisons. La première raison est d'ordre économique et vise à réduire le temps et le coût en limitant le nombre de stations. Ceci permet également une bonne gestion du réseau et un bon encadrement. La seconde raison est d'ordre pratique et qui consiste à automatiser les stations pour permettre d'une part l'enregistrement aisé de tous les paramètres météorologiques et la diminution des erreurs d'autre part.

Il faut noter aussi que les relevés météorologiques se font par des observateurs non qualifiés; ce qui échappe à la rigueur et au contrôle des observations.

Dans ce contexte, nous avons estimé nécessaire de chercher une méthode d'optimisation du réseau pluviométrique, en prenant comme zone test le Bassin du Hodna.

## PRESENTATION DE LA ZONE D'ETUDE

Le bassin du Hodna est une cuvette qui fait partie des hautes plaines steppiques algériennes. C'est un bassin endoréique, bordé par les monts du Hodna au Nord et les monts d'Ouled Nail (Atlas saharien) au Sud (figure 1).



**Figure 1** : Localisation et situation géographique du bassin hodnéen

La cuvette hodnéenne est fermée au Nord par les monts des Hodna et Titeri, au Sud par l'Atlas saharien, à l'Est par les monts de Belazma et enfin à l'Ouest par



personnes pour entretien du réseau et pour la collecte des données pluviométriques.

- iv. le nombre de stations climatiques complètes permettant l'enregistrement de tous les paramètres climatiques est très insuffisant.

Pour atténuer ces contraintes, nous avons pensé à optimiser ce réseau en complétant et en automatisant cinq stations représentatives qui couvrent tout le territoire du bassin. La méthode d'optimisation adoptée est celle des Algorithmes Génétiques Spatialisés (Boukhetala et al., 1996; Boukhetala, 2004).

## CONSIDERATIONS THEORIQUES

Généralement, les Algorithmes Génétiques (AG) représentent une famille assez riche et très intéressante d'algorithmes stochastiques d'optimisation qui sont fondés sur les mécanismes de la sélection naturelle et génétique. Le principe de ces algorithmes est de procéder par une recherche stochastique sur un espace important et à travers une population de pseudo-solutions. Ces algorithmes sont simples et très efficaces parce qu'ils ne demandent pas d'hypothèse sur la manière d'optimiser, comme la continuité et la différentiation. Ils s'opèrent directement sur l'espace exploré, après un codage de solutions faisables de la fonction d'évaluation. Ils procèdent sur une population de points, au lieu d'un point unique et utilise les valeurs de la fonction étudiée sans aucune autre connaissance ni règle probabilistique de transition.

Pour mieux suivre la dynamique de l'algorithme génétique, nous adoptons les mêmes définitions que celles utilisées dans la littérature. Ainsi :

- i. Individu ou Chromosome: représente une solution potentielle
- ii. Gène : bit ou ensemble de bits codant une information
- iii. Population: ensemble de chromosomes ou de points appartenant à l'espace de recherche
- iv. Environnement: espace de recherche
- v. Ajustement ou fonction d'estimation: fonction recherchée pour maximiser (minimiser).

Le fonctionnement des AG s'appuie sur différents opérateurs de base. Ceux-ci sont inspirés directement du mécanisme de la sélection naturel et du phénomène génétique. Il consiste à faire évoluer la population pour adapter des individus à l'environnement. Techniquement, une nouvelle génération résulte à la fin d'un cycle en utilisant les trois principaux opérateurs standards qui sont les

opérateurs de reproduction, de croisement et de mutation. Pour une raison de mise en oeuvre adaptée aux ordinateurs, une représentation chromatique ou codage des individus (solutions faisables) est nécessaire.

Dans le cas de notre recherche, nous avons utilisé l'Algorithme Génétique adapté au cas spatiale, noté AGS (*Boukhetala et Ait Kaci, 2004*) qui utilise un codage spatial de solution ainsi que deux opérateurs appropriés de croisement et de mutation.

### **Codage de Solution**

On considère un  $n$ -échantillons spatial, dans un espace de recherche de dimension  $L < n$ . Celui-ci prend la forme matricielle suivante:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1L} \\ \cdot & & \dots & \cdot \\ \cdot & & \dots & \cdot \\ x_{n1} & x_{n2} & \dots & x_{nL} \end{bmatrix}$$

Ainsi, une population des solutions dans cet espace est un ensemble de matrices  $M_{(n \times L)}$ . Nous adoptons relativement le codage par rapport à une matrice; donné, pour  $i$  individu, par un vecteur réel de dimension  $n \times L$ :

$$X_i = ((x_{11}, \dots, x_{1L})_i, (x_{21}, \dots, x_{2L})_i, \dots, (x_{n1}, \dots, x_{nL})_i)$$

### **Opérateur de croisement local**

L'opérateur de croisement local sera noté OCL. A partir d'un couple d'individus  $(X_1, X_2)$ , nous choisissons un échantillon  $(x_i^1, x_i^2)$ .

Pour définir l'opérateur local OCL, il faut générer un nouvel échantillon tel que :

$$(x_i^{1'}, x_i^{2'}) \in [a, b] \times [a, b]$$

avec:

$$\begin{cases} x_i^{1'} = P_E \left( \frac{(b-a)}{pas} \times U \right) \times pas + pas/2 \\ x_i^{2'} = x_i^1 + x_i^2 - x_i^{1'} \end{cases}$$

où :

$$a = \max(x_i^1, x_i^2) - \min(x_i^1, x_i^2)$$

$$b = \min(x_i^1 + x_i^2, Sup)$$

Les quantités, dénotées par les  $pas$ ,  $P_E$  et  $U$ , sont respectivement l'incrément (ou pas) de l'espace échantillonné, partie entière et une valeur aléatoire, pris entre  $[0,1]$ .

Durant la recherche d'échantillon optimal, il peut y avoir deux échantillons identiques en raison de la nature combinatoire du problème. Pour cela, on introduit un opérateur de diversité qui suit l'augmentation ou la diminution du gène:

$$\begin{cases} x'_i = x_i & Sgn \ pas \\ y'_i = y_i & Sgn \ pas \end{cases} \quad (x_i, y_i) \in [a_1, b_1] \times [c_1, d_1]$$

$a_1, b_1, c_1, d_1$  sont des constantes, et  $Sgn$  est défini par:

$$Sgn = \begin{cases} + si (x_i = a_1 \text{ ou } y_i = c_1) \\ - si (x_i = b_1 \text{ ou } y_i = d_1) \\ (+ \text{ ou } -) \text{ avec la probabilit é } 1/2 \end{cases}$$

Finalement, le croisement local est pour notre problème, une adaptation très intéressante du croisement multi-emplacements classique sur les chromosomes produits par la concaténation de code binaire.

### Opérateur de mutation

L'opérateur de mutation (OM) que nous proposons à notre problème d'échantillonnage spatial procède comme suit :

Nous opérons sur un gène  $S_k = (x_k, y_k)$  que nous amortissons à un nouveau gène  $S'_k = (x'_k, y'_k)$ . Ainsi :

$$\begin{cases} x'_k = x_k & Sgn \ U \\ y'_k = y_k & Sgn \ V \end{cases}$$

tel que:

$$U \in \begin{cases} P_E \left( \frac{U[x_k, b_1]}{pas} \right) \times pas & \text{if } x_k \geq \frac{a_1 + b_1}{2} \\ P_E \left( \frac{U[a_1, x_k]}{pas} \right) \times pas & \text{if } x_k < \frac{a_1 + b_1}{2} \end{cases}$$

$$V \in \begin{cases} U[y_k, d_1] & \text{si } y_k \geq \frac{c_1 + d_1}{2} \\ U[c_1, y_k] & \text{si } y_k < \frac{c_1 + d_1}{2} \end{cases}$$

$U_{[a_1, b_1]}$  est une valeur aléatoire de  $[a_1, b_1]$ .

### Critère d'arrêt

Le nombre des itérations est fixé, après quelques tests, sur la fonction d'évaluation. Il satisfait un compromis entre contraintes de convergence de la population, les temps du CPU et la précision (Boukhetala et Ait Kaci, 2004)

### APPLICATION

Selon les remarques indiquées précédemment, il s'agit de diminuer le nombre de stations du réseau pour alléger les coûts de fonctionnement et d'entretien tout en gardant l'information pluviométrique fournie par le réseau initial.

L'objectif est de fournir au décideur un outil à partir duquel il peut sélectionner un nombre minimum  $k$  ( $k \leq 37$ ) de stations d'observations de telle sorte qu'elles soient réparties de la manière la plus équitable possible sur la région et qu'elles fournissent les mêmes informations sur la pluviométrie que celles du réseau initial.

Nous avons:

$$D = \{(x, y, z) | x \in [540, 800], y \in [150, 400], z \in [456, 1250]\}$$

Les données moyennes annuelles des pluies, observées dans le bassin du Hodna (D), peuvent être approchées par le modèle linéaire suivant (Guettouche, 2003) :

$$Q_{S_i} = \langle \tilde{\alpha}, S_i \rangle + d + m \cdot \varepsilon(S_i) \text{ (en mm)}$$

où :

$$\tilde{\alpha} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

$d$  et  $m$  sont des paramètres réels. Les paramètres  $\tilde{\alpha}$ ,  $b$ ,  $c$  et  $d$  prennent respectivement les valeurs suivantes :

$$\hat{a} = 60,05 \cdot 10^{-3}, \hat{b} = 0,701, \hat{c} = 0,29, \hat{d} = -94,44$$

avec une valeur du coefficient de corrélation  $R^2 = 0,89$ .

Les erreurs des observations sont modélisées par un processus spatial aléatoire  $\varepsilon_s$ , en prenant en considération la caractéristique de variabilité, l'incertitude et l'hétérogénéité du processus pluviométrique.

Nous supposons que le processus aléatoire spatial des erreurs, associé à chaque emplacement, est un processus gaussien homogène, avec une moyenne  $\mu_\varepsilon$  et une covariance  $C(\varepsilon_i, \varepsilon_j)$ . Par conséquent, on détermine les emplacements optimaux par la recherche de l'échantillon optimal de dimension finie  $k$ , par rapport à un critère d'optimalité tel que :

$$E(\bar{Q} - M)^2$$

avec  $\bar{Q}$  la hauteur moyenne de pluie des  $k$  sites :

$$\bar{Q} = \frac{1}{k} \sum_{i=1}^k Q_{S_i}$$

### Formulation mathématique du problème

Soit :

$$S = \left\{ s_i \in D \subset \mathbb{R}^3, i = 1 \dots k \right\}$$

un ensemble d'échantillons des sites de dimension  $k = |S| \leq n$ , où  $n$  est la dimension du domaine  $D$ . Le critère d'optimalité est donné par :

$$\sigma^2(S^* / \varepsilon) = \min_{S \in D} \left\{ E \left( \sum_{i=1}^k \frac{Q(s_i)}{k} - M \right)^2 \right\}$$

où :

- $k$  : Dimension de  $S$ ,
- $S$  ensemble d'échantillon de dimension  $k$ ,
- $Q_{S_{ii}}$  quantité de pluie observée sur le site  $s_i$ ,
- $s_i^* = (x_i^*, y_i^*, z_i^*)$   $i=1, 2, \dots, k$  coordonnées de l'échantillon spatial optimal  $S^*$ ,
- $\varepsilon = (\varepsilon(s_i))_{i=1, \dots, k}$  vecteur des erreurs correspondantes aux observations dans chaque site. Ce vecteur est un processus spatial Gaussien de type analytique, avec moyenne zéro et fonction de covariance  $\text{Cov}(s_i, s_j)$ .
- $E$  est une moyenne mathématique, définie par rapport à la loi de  $\varepsilon$

**Fonction d'évaluation**

Nous avons:

$$\sigma^2(S^*; \varepsilon) = \text{Min} \left\{ \frac{1}{k^2} \left[ (Cst)^2 + m^2 \cdot \sum_{i=1}^k \sum_{j=1}^k \text{Cov}(\varepsilon(s_i), \varepsilon(s_j)) \right] \right\}$$

tel que:

$$\left\{ s_i \in S : i = 1, \dots, k \right\}, \left\{ s_j \in S : j = 1, \dots, k \right\}$$

*Cst* : constante appropriée préalablement déterminée.

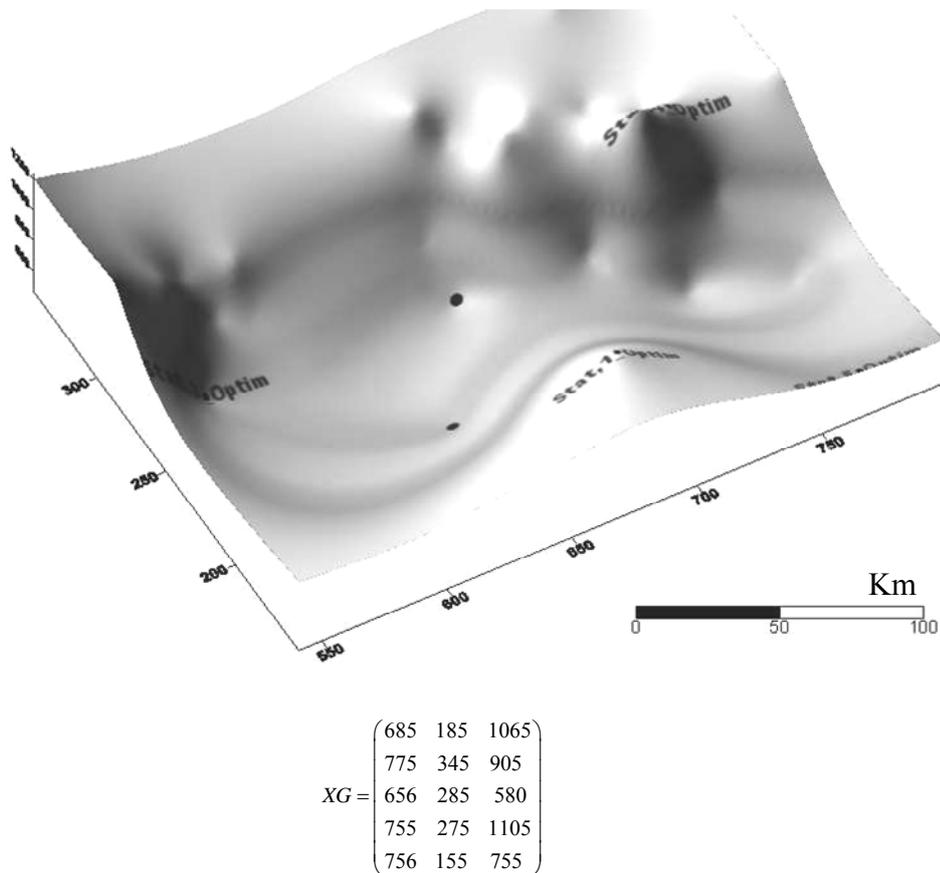
Nous avons appliqué notre algorithme AGS pour les paramètres suivants :

- i.  $M = 337,9$  mm: hauteur moyenne des pluies, dans le bassin du Hodna,
- ii.  $k = 5$  : nombre de stations à choisir,
- iii. Sélection par rangement (Pression de sélection = 1,4),
- iv. Croisement: OC et OCL sont utilisés, avec les probabilités adéquates  $P_{OC}$  et  $P_{OCL}$ , respectivement,
- v. Mutation: OM est utilisé avec une probabilité adéquate POM,
- vi. Covariance spatiale :

$$\text{Cov}(s_i, s_j) = \exp\left(-\theta \|s_i - s_j\|^2\right), \text{ avec } \theta = 0,05$$

Le choix initial d'individu a un effet sur la précision de solution. L'effet est positif quand le choix concerne un individu mauvais.

Pour  $\sigma^2(S^* / \varepsilon) = 0,62$ , la configuration spatiale optimale correspondante est donnée par la figure 3.



**Figure 3 :** Représentation 3D des sites optimaux obtenue par les AGS

Le tableau 1 donne les résultats obtenus par l'AGS. Ces résultats, qui concernent les paramètres optimums par rapport à un échantillonnage régulier, résultent après une répétition de 80 temps de l'AGS.

Pour réduire la variabilité de l'estimateur  $\bar{Q}$  et améliorer la performance de l'AGS, nous avons utilisé un opérateur quantitatif. Le résultat de cette opération est donné par le tableau 2 :

Enfin, l'implémentation de l'AGS a donné 5 sites à référence spatiale que nous estimons optimums. Ces sites sont géoréférencés par leurs coordonnées géographiques (tableau 3).

**Tableau 1:** Résultats obtenus de l'application de l'AGS

$P_{CG}$	$P_{CL}$	Erreurs absolues (EA): $ \bar{Q} - M $	Intervalle de Confi. (AE)
0,9	0,01	4,43; 3,98; 0,52; 1,81; 2,02; 1,4; 0,3; 1,26; 0,76; 0,11; 5,10; 2,41; 3,46; 9,07; 1,81; 0,651; 0,34; 0,41; 2,19; 0,00; 0,94; 0,75; 5,17; 2,80; 7,14; 2,79; 0,19; 1,78; 0,03; 0,19; 8,28; 0,63; 1,44; 7,59; 8,84; 1,42; 2,14; 0,58; 4,09; 2,46; 7,54; 1,42; 4,62; 2,02; 4,14; 0,39; 11,85; 0,11; 0,26; 1,11; 0,56; 1,81; 2,39; 0,84; 2,01; 0,01; 0,98; 1,58; 7,71; 3,28; 1,07; 2,16; 0,07; 0,21; 0,95; 4,56; 0,92; 0,03; 0,43; 1,22; 1,46; 13,74; 7,81; 1,27; 5,51; 1,64; 0,90; 5,31; 1,18; 0,99	[4,02; 9,96]

**Tableau 2:** Résultat final des AGS

$P_{CG}$	$P_{CL}$	Erreurs Absolut (AE): $ \bar{Q} - M $	Intervall de Confi. (AE)
0,9	0,01	0,35; 2,09; 1,31; 0,22; 2,57; 0,7; 0,56; 0,26; 0,43; 0,61; 0,56; 2,38; 5,58; 0,93; 5,53; 2,70; 0,32; 0,80; 0,441; 0,47; 6,63; 1,95; 1,03; 1,17; 1,67	[ 2,43; 5,79 ]

**Tableau 3 :** Emplacement des 05 stations optimales

X	Y	Altitude	Station
685	185	1065	<i>Stat.1 Optim</i>
775	345	905	<i>Stat.2 Optim</i>
565	265	580	<i>Stat.3 Optim</i>
755	275	1105	<i>Stat.4 Optim</i>
765	155	755	<i>Stat.5 Optim</i>

## DISCUSSION

Il faut noter que les AGS sont implémentés par le langage MATLAB. La procédure de ce langage est utilisée dans la génération des erreurs gaussiennes de la mesure spatiale et par les opérateurs de simulation OC, OCL et OM.

L'application de cet algorithme aux données de la zone d'étude a permis de situer les cinq postes demandés, car, du point de vue morphologique, on estime

que le bassin peut être cadré par un maximum de 5 stations représentant les différents dispositifs morphoclimatiques.

En observant la figure 3, nous remarquons que :

- i.* les postes matérialisent bien la zone.
- ii.* deux postes représentent les montagnes au Nord, morphologiquement très massives et condensées.
- iii.* deux postes au centre Est et à l'Ouest caractérisant une morphologie ouverte : la plaine hodnéenne.
- iv.* une station au sud caractérisant les montagnes du sud de massivité moyenne et de morphologie large.

## **CONCLUSION**

L'application d'un Algorithme Génétique Spatialisé, pour optimiser un réseau pluviométrique, est proposée. Les résultats d'échantillonnage obtenus sont améliorés par un opérateur quantitatif des erreurs, pour des échantillons de petite dimension. Ceci se justifie par le fait que cet opérateur a un effet sur la régularité de la fonction d'ajustement, en créant une configuration de choix, souhaitable pour un algorithme de type génétique.

L'approche proposée dans ce travail peut résoudre tout autre exemple de ce type de problème, c'est à dire le choix des stations d'observations pluviométriques représentatives pour un réseau dense et ayant une redondance de l'information pluviométrique. Dans ce cas, il faut introduire de nouvelles données de paramètres géographiques tels que l'exposition, l'angle de vent dominant, etc... qui, par leur variabilité, ont une grande influence sur les précipitations de la région à étudier.

Les travaux en perspective seront consacrés, en terme mathématique, à l'amélioration des paramètres et des opérateurs de l'Algorithme Génétique Spatialisé. Cette amélioration permettra de diminuer l'erreur et d'obtenir des intervalles de confiance réalistes plus adaptés aux applications pratiques.

Mais pour les praticiens, ce sont les paramètres intervenant dans la caractérisation du site qui doivent être mieux choisis, car ils ont une influence sur le phénomène à étudier que ce soit, hydrologique, climatique, sismologique ou autre.

## **REFERENCES BIBLIOGRAPHIQUES**

- BENHENNI K., CAMBANIS S. (1998). The effect of quantization on the performance of sampling design, IEEE transactions on information theory 44, 5, 109-121.
- BOUKHETALA K., AIT-KACI S. (2004). Finite Spatial Sampling Design and quantization, COMPSTAT'2004, Symposium, Physica Verlag/Springer .
- BOUKHETALA K., HABIB F.Z. (2002). Finite Sampling design with quantization for a pharmacokinetic problem. The 6<sup>th</sup> WMSCI, Computer Science, Vol XI, edition IIS, USA, 167-170.
- BOUKHETALA K., MEHASSOUEL N. (2000). Optimal Spatial Sampling for an estimating problem, based on correlated observations, Compstat'2000, edition Statistics Netherlands, 1, 157-158.
- BOUKHETALA K., BENHENNI K., BENAMARA S.(1996). Optimal Sampling for estimating integral of function from observations *with correlated measurement errors*, Compstat'96, Edition UPC, A. Prat and E. Ripoll, 161-162.
- GUETTOUCHE M.S.(2003) . Du terrain à l'analyse numérique essai d'une typologie systémique de l'espace géographique hodnéen. (Algérie nord orientale). Thèse de Doctorat d'Etat en Géomorphologie, FSTGAT, USTHB, Alger, 230 p.