

Modèles énergétiques hiérarchiques  
pour la résolution des problèmes inverses  
en analyse d'images  
Application à la télédétection

Annabelle CHARDIN

- laboratoire d'accueil : IRISA/VISTA
- directeur de thèse : Patrick Pérez

## Introduction

Foisonnement de données images



Besoin de

Modèles adaptés

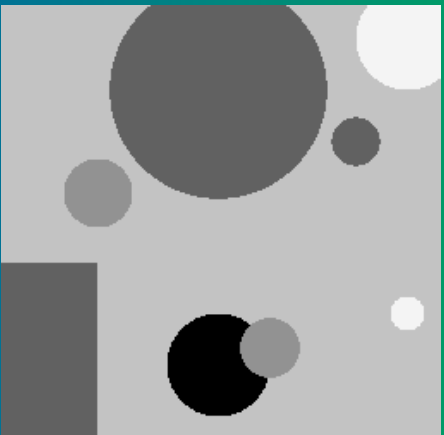
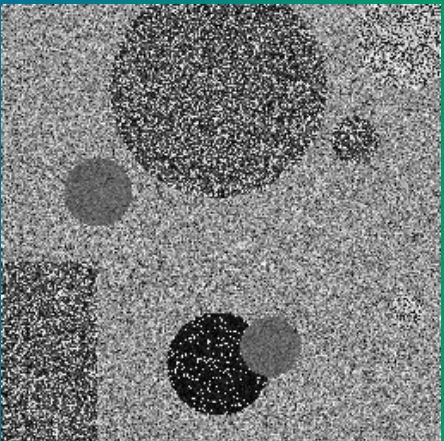
Algorithmes efficaces

Cadre théorique : modèles énergétiques

## Plan

- Modèles énergétiques et structures d'interactions
- Estimation supervisée sur arbre tronqué
- Estimation non supervisée sur arbre tronqué
- Application à la télédétection

## Problèmes inverses et modèles énergétiques



$y$  : variables observées     $x$  : variables inconnues

- Construire une fonction d'énergie paramétrée par  $\theta$   
 $U_\theta(x, y)$  = interactions entre  $x$  et  $y$  + *a priori* sur  $x$
- $P(x, y|\theta) \propto \exp(-U_\theta(x, y))$

Problèmes posés

estimer  $x$   
connaissant  $\theta$



MAP, MPM

échantillonner



Monte Carlo

estimer  $\theta$  ne  
connaissant que  $y$



EM, ICE

$$\hat{x}^{MAP} = \arg \max_x P(X = x | Y = y) = \arg \min_x U_\theta(x, y)$$

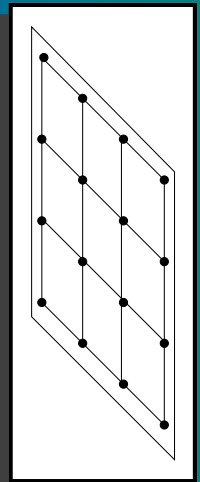
$$V_s, \hat{x}_s^{MPM} = \arg \max_{x_s} P(x_s | y)$$

- $U_\theta(x, y)$  = somme de termes locaux  $\Rightarrow$  exploiter la localité

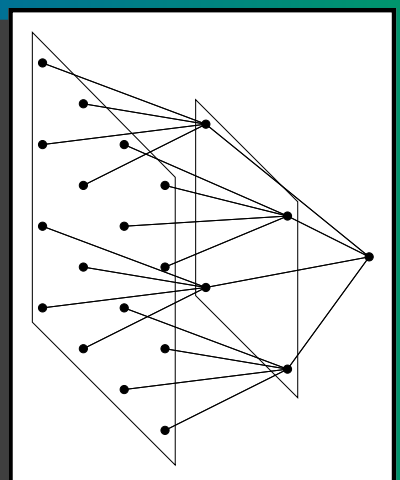
## Hiéarchie

- Techniques par groupe de renormalisation [Gidas]
- Algorithmes multigrilles [Terzopoulos, Enkelmann, Bouman, Pérez, Orchard, ...]
- Modèles sur graphes hiérarchiques : quadarbre [Willisky *et al.*, Bouman *et al.*, Laferté *et al.*, ...], autres graphes [Bouman *et al.*, Kato *et al.*, Fieguth *et al.*, Williams *et al.*, Monfrini *et al.*, ...]

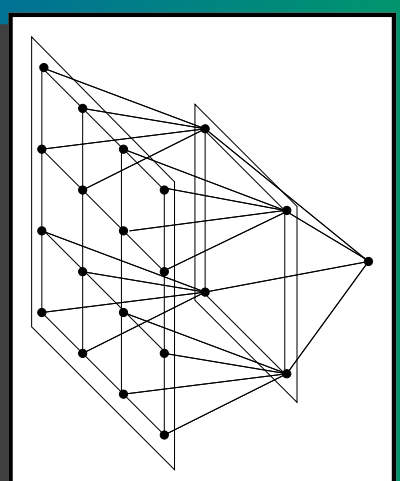
## Les structures d'*a priori* dans les modèles énergétiques



grille 2-D



quadarbre



graphe de Kato

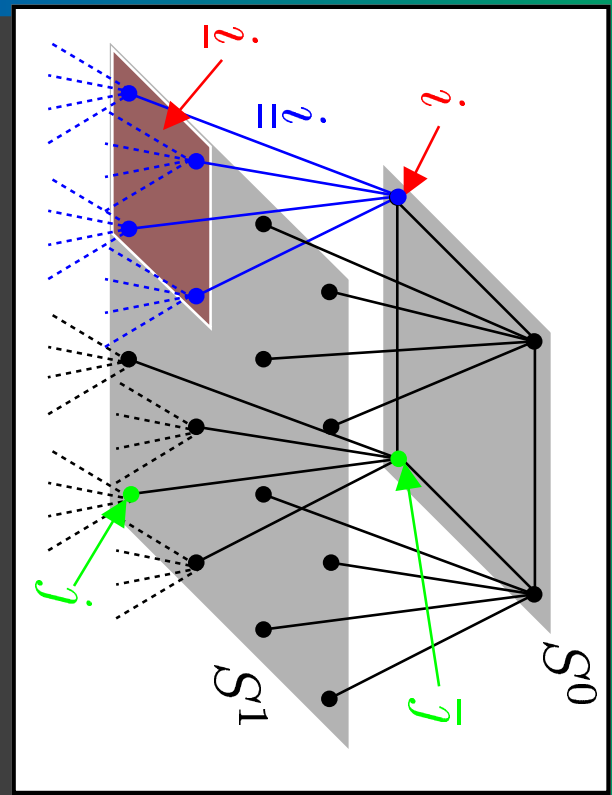


spatiale intuitive  
itératif

hiérarchique  
non-itératif

spatial et hiérarchique  
itératif

## Grappe hiérarchique hybride : un arbre tronqué



$S^0$  : grille rectangulaire (4-connectivité)  
 $S^n$  ( $0 < n \leq N$ ),  $2^n \times 2^n$  plus large  
 $x$  et  $y$  indicés par  $S = \bigcup_{n=0}^N S^n$

$$P(x|y) \propto \prod_{\langle i,j \rangle \in S^0} \underbrace{g_{ij}(x_i, x_j)}_{\text{a priori spatial}} \times \prod_{i \notin S^0} \underbrace{f_i(x_i, x_{\bar{i}})}_{\text{a priori hiérarchique}} \times \prod_{i \in S} \underbrace{h_i(x_i, y_i)}_{\text{mesure ponctuelle}}$$



# Algorithmes d'estimation supervisée

Algorithmes d'estimation supervisée : (1) Passe montante

### Récurrance montante

MAP

$$\left\{ \begin{array}{l} V_i(x_i^-) \triangleq \max_{x_i} [h_i(x_i, y_i) \\ \times f_i(x_i, x_i^-) \prod_{j \in \underline{i}} V_j(x_i)] \\ \text{avec } x_i^*(x_i^-) = \arg \max_{x_i} V_i(x_i^-) \end{array} \right.$$

MPM

$$\begin{aligned} F_i(x_i^-) &\triangleq \sum_{x_i} h_i(x_i, y_i) \\ &\times f_i(x_i, x_i^-) \prod_{j \in \underline{i}} F_j(x_i) \\ \Rightarrow P(x_i | x_i^-, y_i^-) \end{aligned}$$

## Algorithmes d'estimation supervisée : (2) À la racine

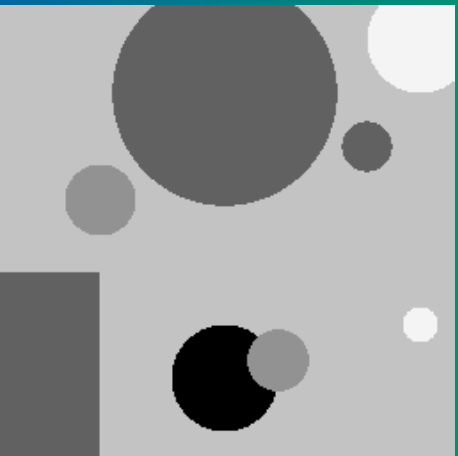
- MAP : Obtenir  $\hat{x}_i$ ,  $\forall i \in S^0$   
⇒ ICM sur les sites de  $S^0$  pour maximiser le terme résultant de la maximisation montante
- MPM : Obtenir  $\forall i \in S^0$ ,  $P(x_i|y)$   
⇒ Approximation des  $P(x_i|y)$  par échantillonnage de  $P(x^0|y)$   
 $\hat{x}_i = \arg \max_{x_i} P(x_i|y)$

## Algorithmes d'estimation supervisée : (3) Passe descendante

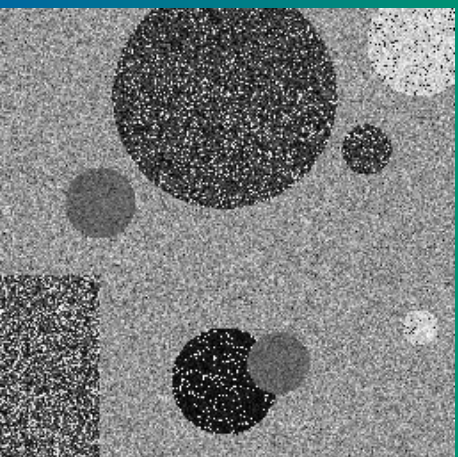
### Récurrance descendante

- MAP : Lecture des tables  $\hat{x}_i = x_i^*(\hat{x}_{\bar{i}})$
- Semi-MPM : Approximation des étiquettes  $\hat{x}_i \approx \arg \max_{x_i} P(x_i | \hat{x}_{\bar{i}}, y_{\bar{i}})$
- MPM : Calcul des marginales *a posteriori*  
$$P(x_i | y) = \sum_{x_{\bar{i}}} P(x_i | x_{\bar{i}}, y_{\bar{i}}) P(x_{\bar{i}} | y)$$
$$\hat{x}_i \triangleq \arg \max_{x_i} P(x_i | y)$$

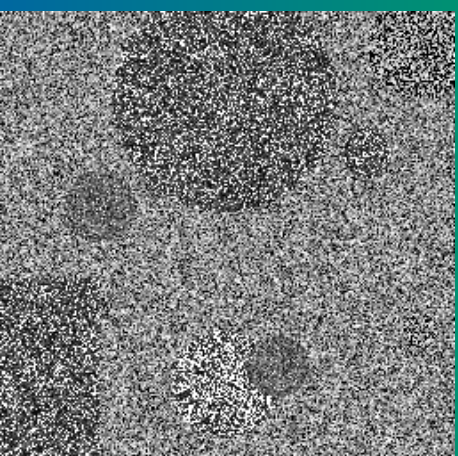
## Classification supervisée



Vérité-terrain

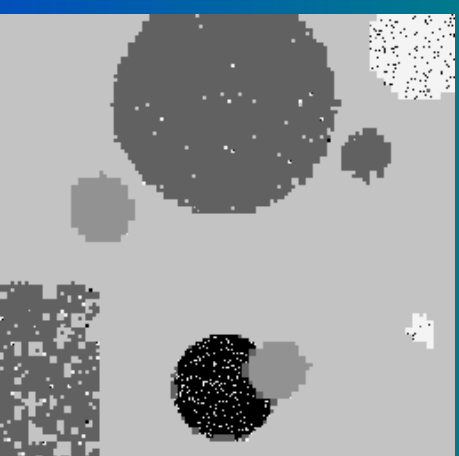
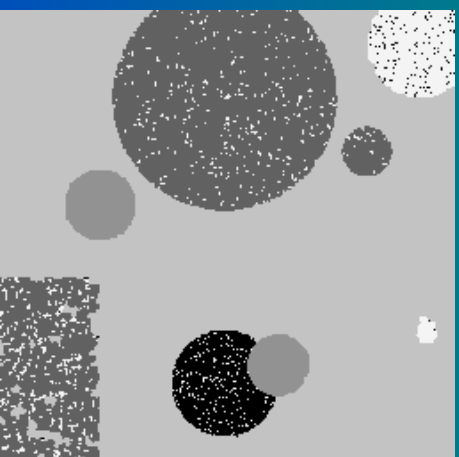
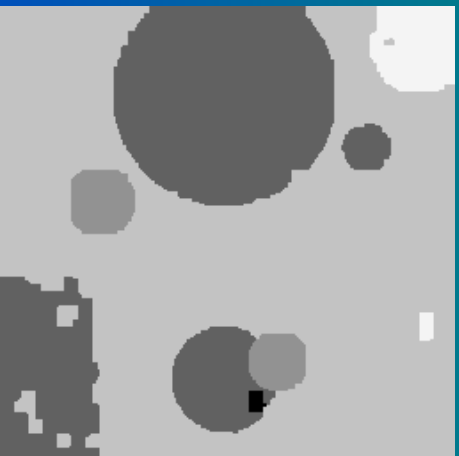
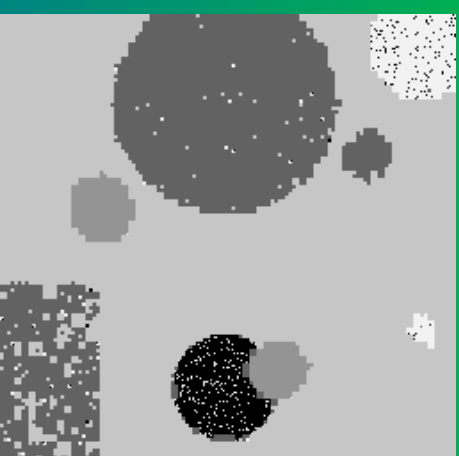
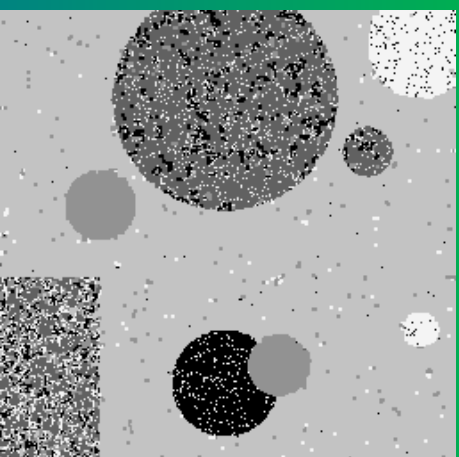
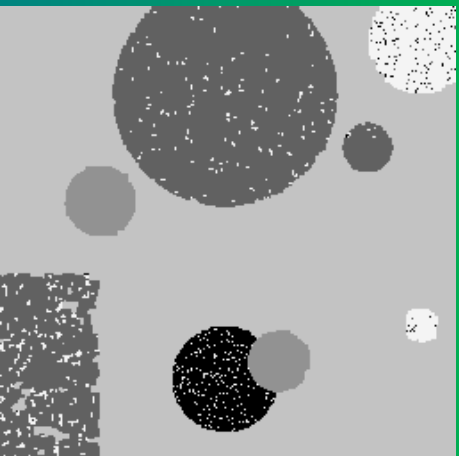


bruit moyen



bruit fort

Bruit gaussien de paramètres dépendant de la classe



Recuit simulé

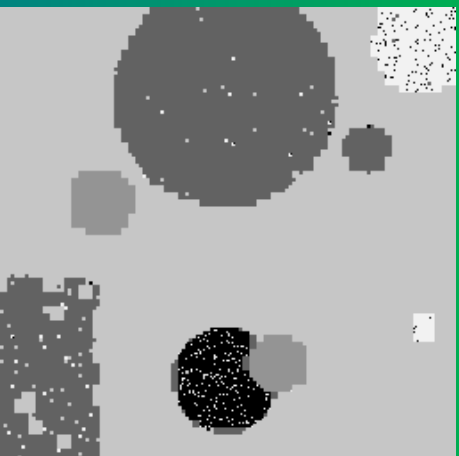
ICM (MAP,  $N=0$ )

MAP,  $N=8$

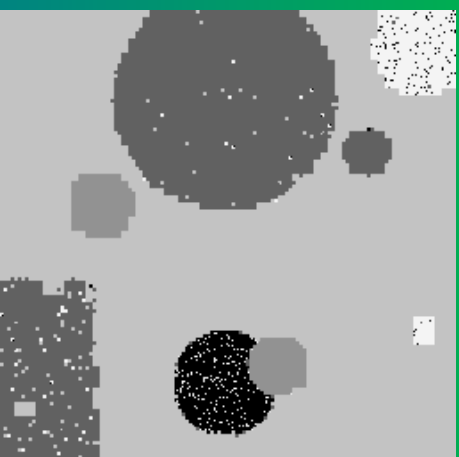
Multigrille

MPM,  $N=0$

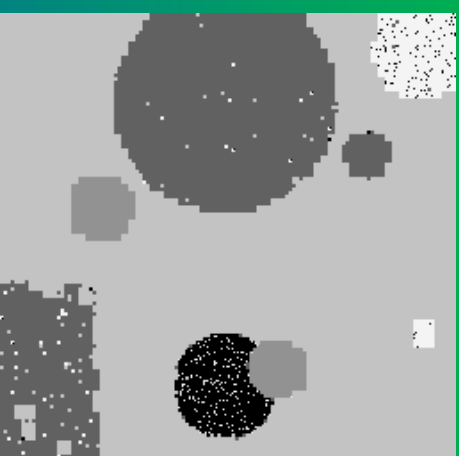
MPM,  $N=8$



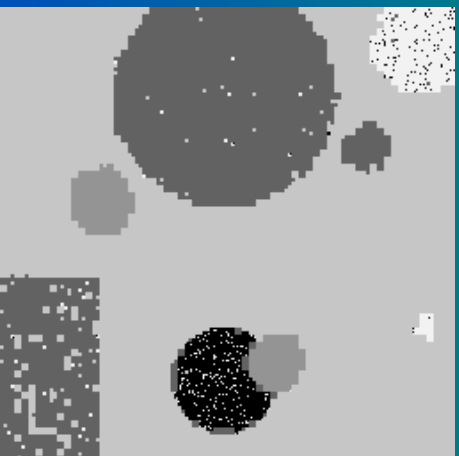
MAP,  $N = 2$



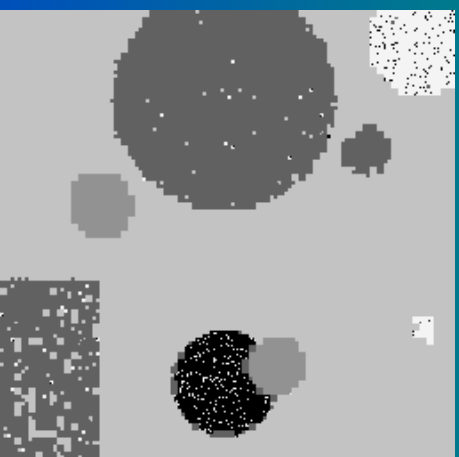
MPM,  $N = 2$



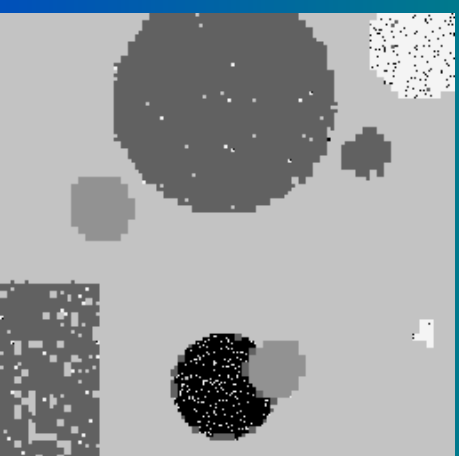
SMPM,  $N = 2$



MAP,  $N = 3$



MPM,  $N = 3$



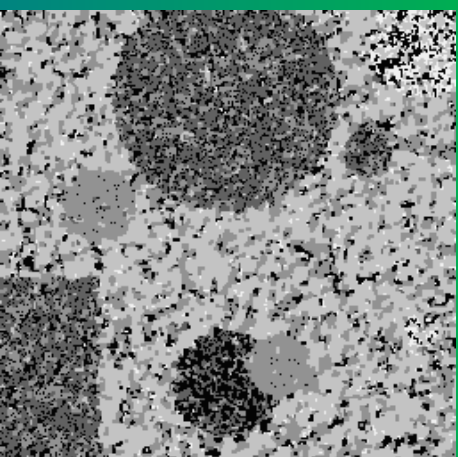
SMPM,  $N = 3$

	MAP			MPM			semi-MPM		
	taux de bonne classif.	temps de calcul	nb. iter.	taux de bonne classif.	temps de calcul	nb. iter.	taux de bonne classif.	temps de calcul	nb. iter.
S	97,2%	200s	134						
G	93,7%	2,7s	2/1/1						
= 0	86%	4s	5	95,2%	187s	33+256	95,2%	187s	33+256
= 2	95,9%	1,3s	2	96,2%	5,1s	5+64	96,2%	3,9s	4+64
= 3	95,8%	1,3s	2	95,8%	3,7s	4+32	95,8%	2,7s	4+32
= 8	95,2%	1,3s	0	95,3%	3,5s	0	95,35%	2,2s	0

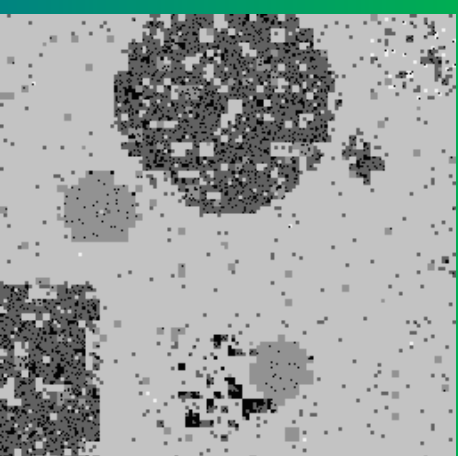




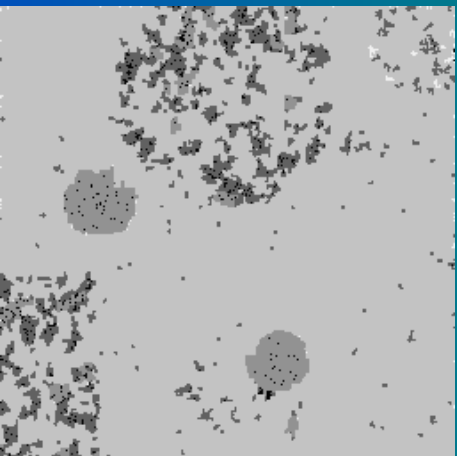
Recuit simulé



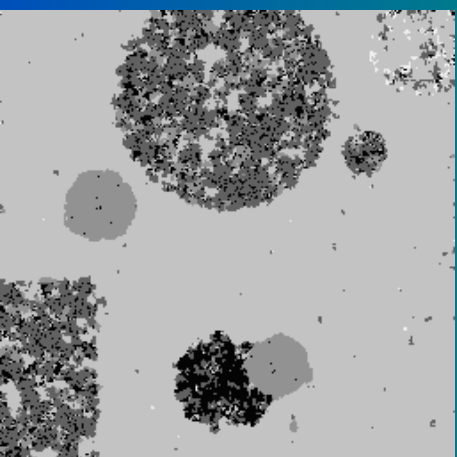
ICM (MAP,  $N = 0$ )



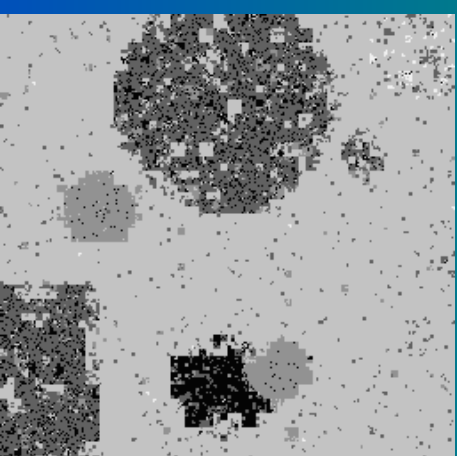
MAP,  $N = 8$



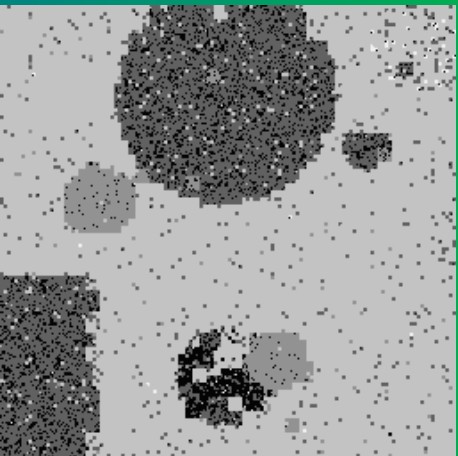
Multigrille



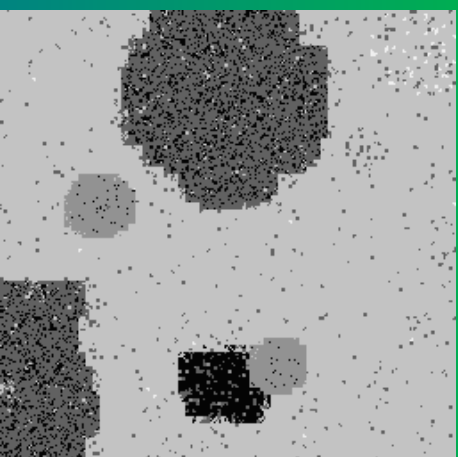
MPM,  $N = 0$



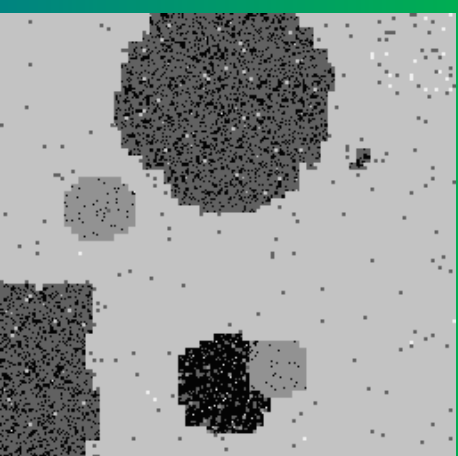
MPM,  $N = 8$



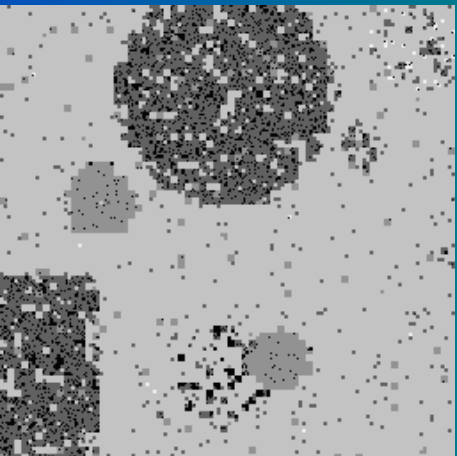
MAP,  $N = 2$



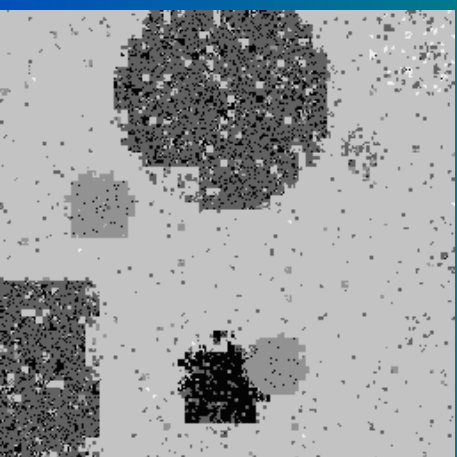
MPM,  $N = 2$



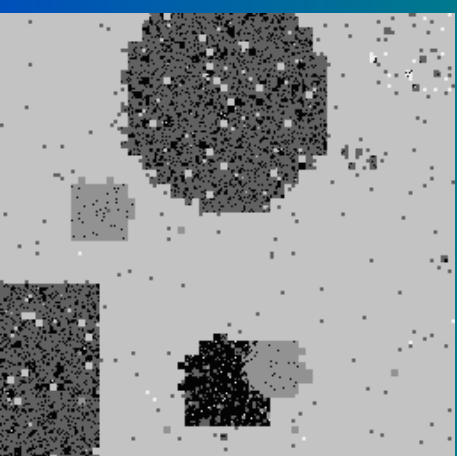
SMPM,  $N = 2$



MAP,  $N = 3$



MPM,  $N = 3$



SMPM,  $N = 3$

	MAP			MPM			semi-MPM		
	taux de bonne classif.	temps de calcul	nb. iter.	taux de bonne classif.	temps de calcul	nb. iter.	taux de bonne classif.	temps de calcul	nb. iter.
1	73,2%	170s	115						
2	71,0%	4s	2/2/2						
0	54,9%	4s	5	80,25%	330s	249+256	80,25%	330s	249+256
2	82,3%	1,5s	3	84%	5,4s	15+64	85,2%	4,3s	17+64
3	78,9%	1,3s	3	81,6%	3,8s	7+32	83,8%	2,6s	8+32
8	78,05%	1,3s	0	80,6%	3,5s	0	83,35%	2,3s	0

# Algorithmes d'estimation non supervisée

## Principe de l'estimation des paramètres sur données incomplètes

- EM (*Expectation-Maximisation*) :
  - maximise  $P(y|\theta)$
  - calcul et maximisation (par rapport à  $\theta$ ) alternatifs de  $Q(\theta|\theta^{(k)}) \triangleq \mathbb{E}[\log P(X, Y = y|\theta)|Y = y, \theta^{(k)}]$
- ICE (*Iterative Conditional Expectation*) :
  - basé sur l'espérance conditionnelle d'un estimateur  $\hat{\theta}(X, Y)$  sur données complètes
  - démarche itérative :  
$$\theta^{(k+1)} = \mathbb{E}[\hat{\theta}(X, Y)|Y = y, \theta^{(k)}]$$

## Paramétrisation sur l'arbre tronqué

- $a(i, \nu) \triangleq \mathbb{P}(X_s = i | X_{\nu_s} \text{ de type } \nu), \forall s \in S^0$
- $f(i, j) \triangleq \mathbb{P}(X_s = i | X_{\bar{s}} = j), \forall s \in S \setminus S^0$
- $h^n(i, l) \triangleq \mathbb{P}(Y_s = l | X_s = i), \forall s \in S^n$

$$\mathbb{P}(x, y | \theta) = \prod_{s \in S^0} \underbrace{\mathbb{P}(x_s | x_{\nu_s}, \theta)}_{=a(x_s, \text{type de } x_{\nu_s})} \prod_{s \notin S^0} f(x_s, x_{\bar{s}}) \prod_{n=0}^N \prod_{s \in S^n} h^n(x_s, y_s)$$

## Remise à jour des paramètres par l'EM

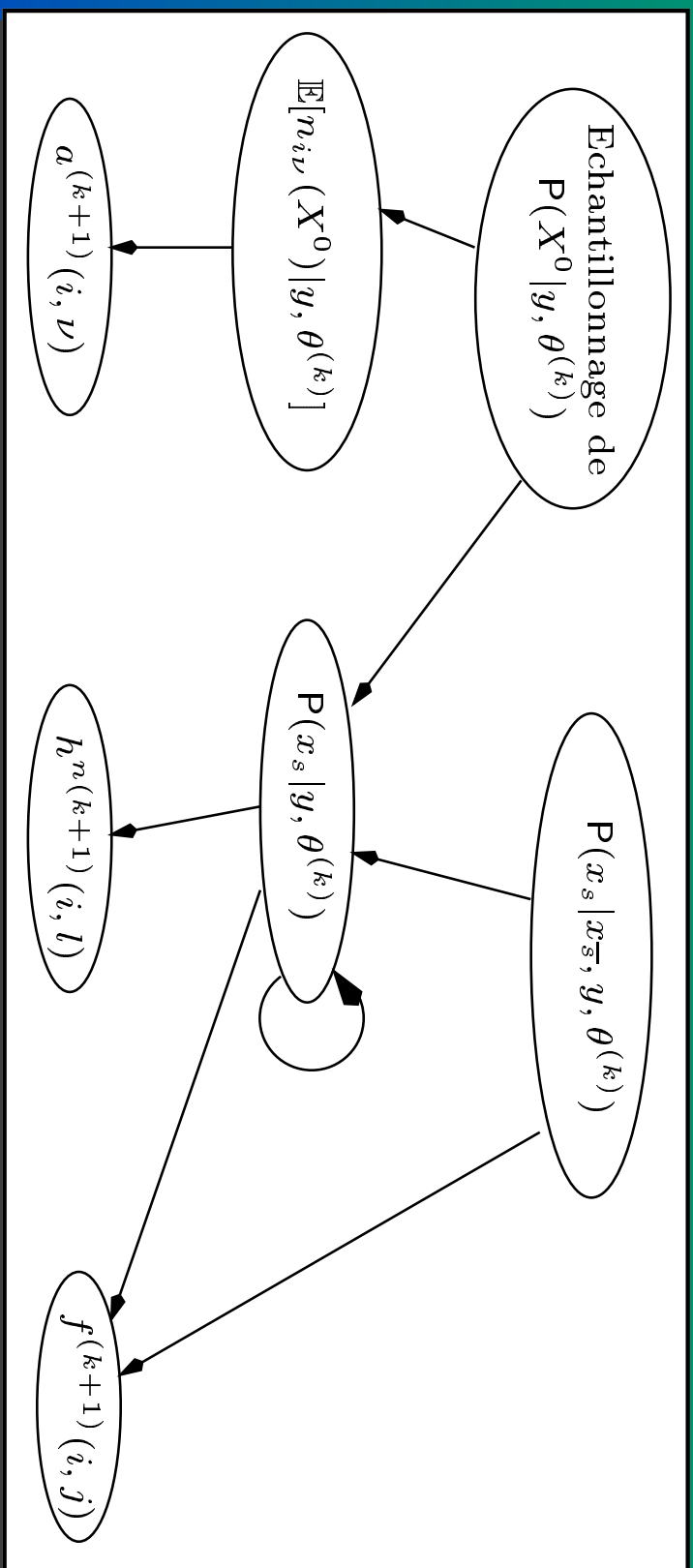
$$a^{(k+1)}(i, \nu) \propto \mathbb{E}[n_{i\nu}(X^0) | y, \theta^{(k)}]$$

$$f^{(k+1)}(i, j) \propto \sum_{s \in S \setminus S^0} \mathbb{P}(X_s = i, X_{\bar{s}} = j | y, \theta^{(k)})$$

$$h^n{}^{(k+1)}(i, l) \propto \sum_{s \in S^n : y_s = l} \mathbb{P}(X_s = i | y, \theta^{(k)})$$

où  $n_{i\nu}(X^0)$  : nombre de sites  $s$  de  $S^0$  tels que  $x_s = i$  et  $x_{\nu_s}$  de type  $\nu$

## Algorithme EM

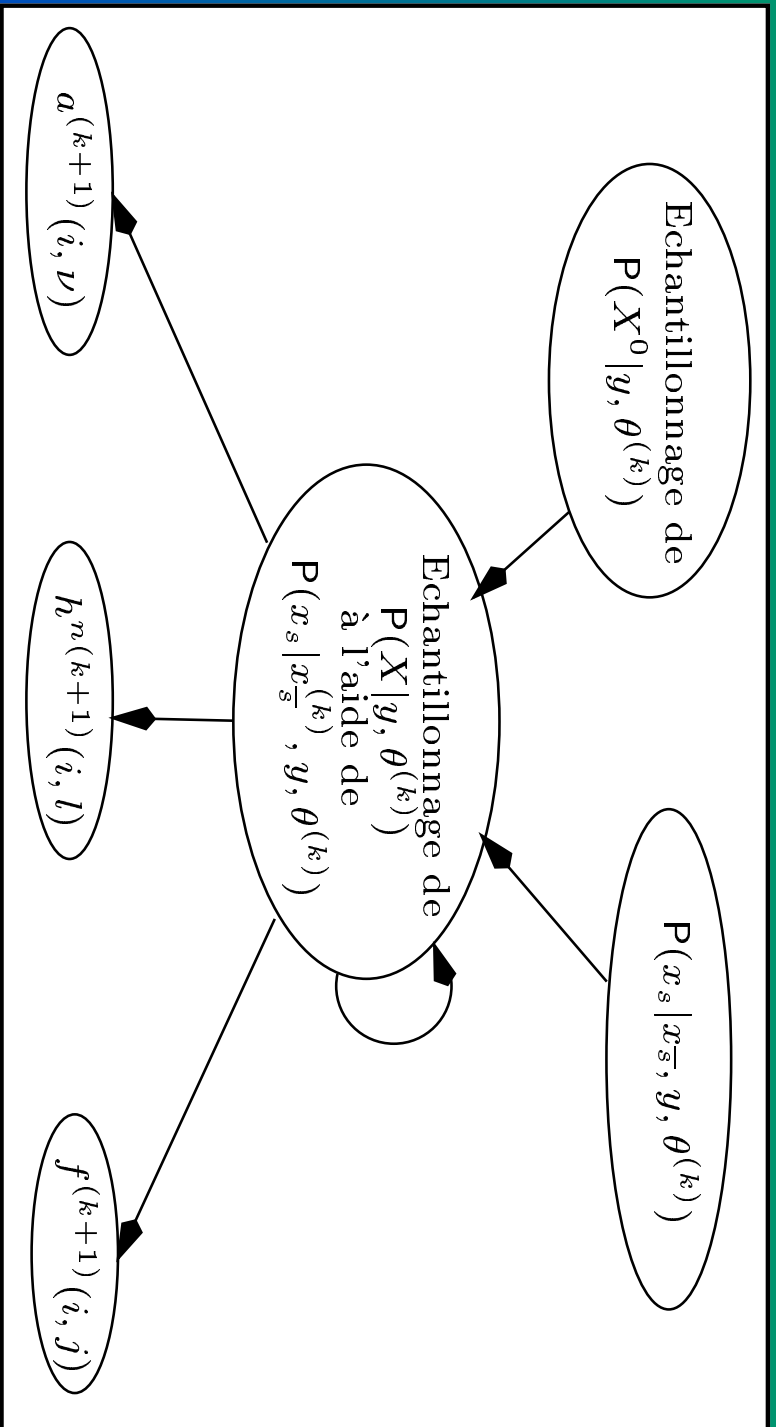




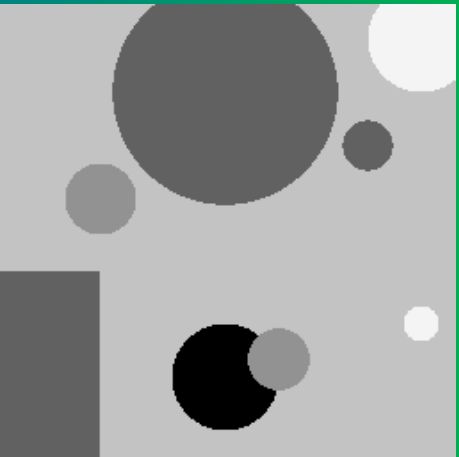
## Remise à jour des paramètres par l'ICE

- Choix des estimateurs : Maximum de Vraisemblance
- Remise à jour : approximation des espérances conditionnelles par moyennes empiriques  
⇒ échantillonnage de  $P(X|Y = y, \theta^{(k)})$

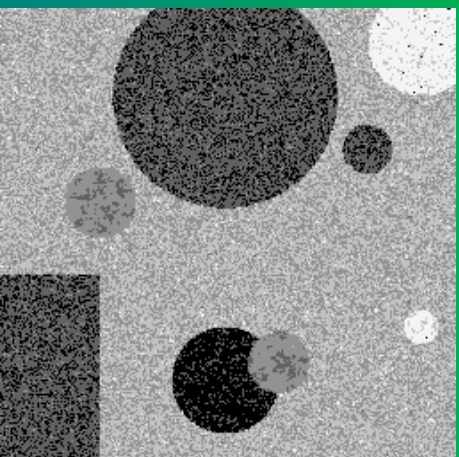
## Algorithme ICE



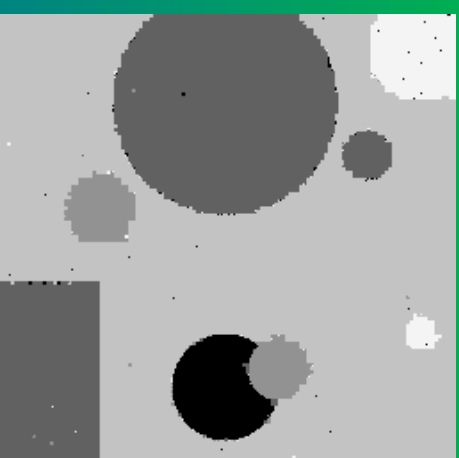
## Classification non supervisée



Vérité-terrain



EM Gibbsien ( $N = 0$ )



EM,  $N = 3$

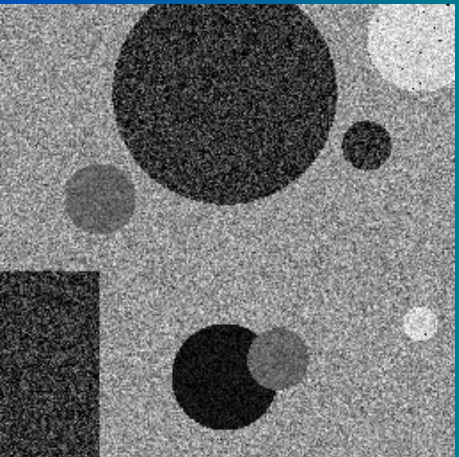
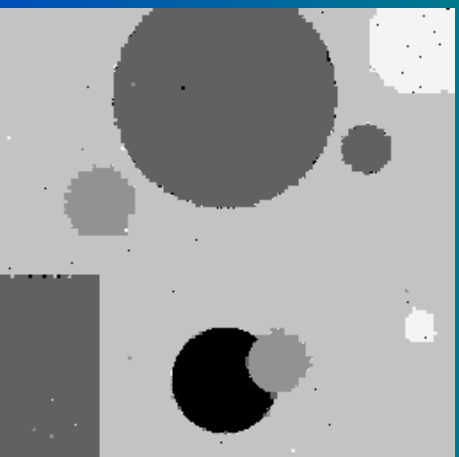
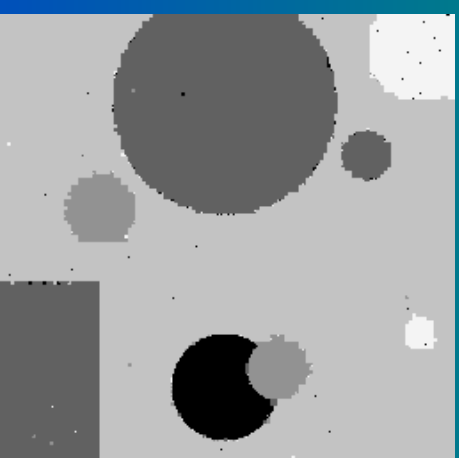


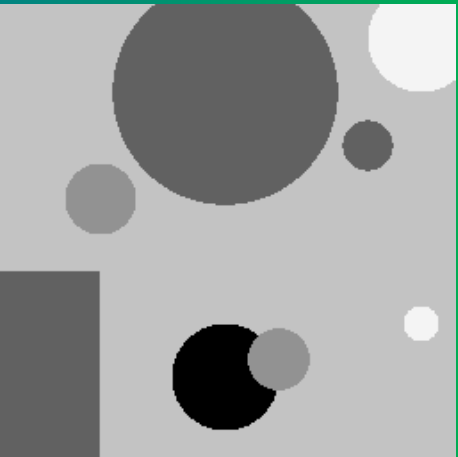
Image bruitée



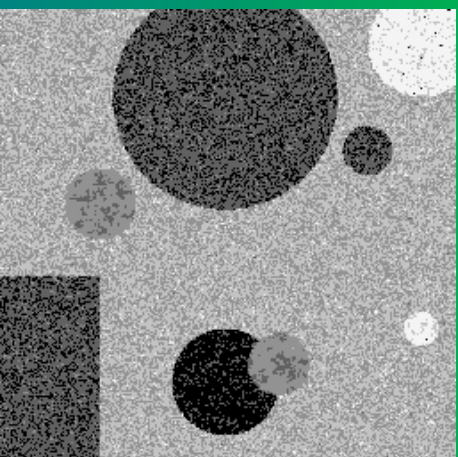
EM sur quadarbre ( $N = 8$ )



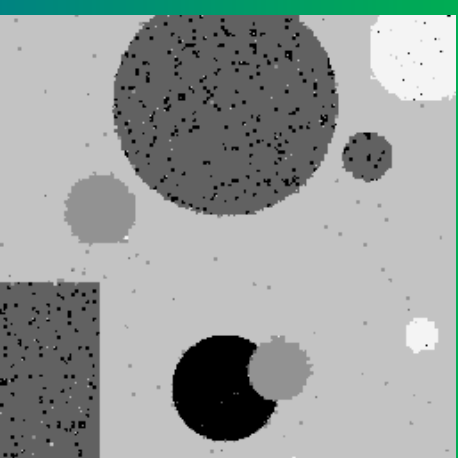
EM,  $N = 4$



Vérité-terrain



ICE Gibbsien ( $N = 0$ )



ICE,  $N = 3$

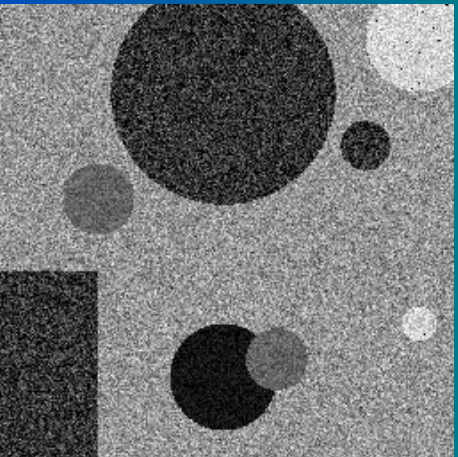
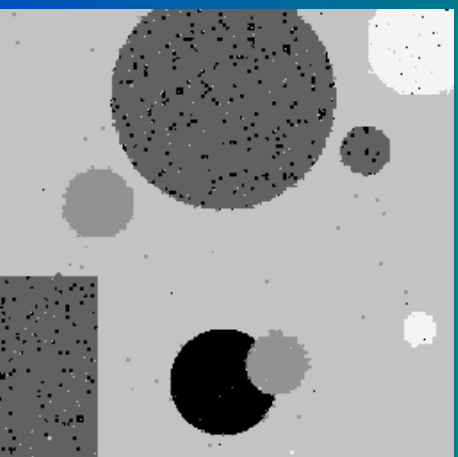
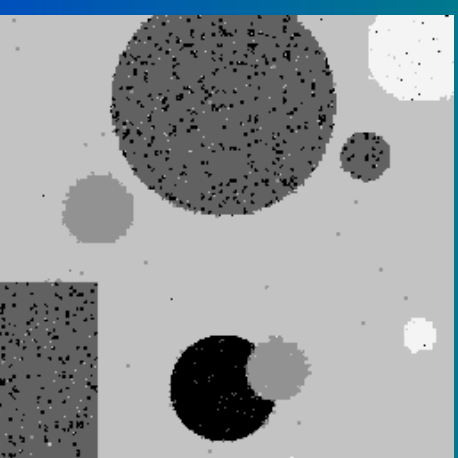


Image bruitée



ICE sur quadarbre ( $N = 8$ )



ICE,  $N = 4$

	Classes				
	1	2	3	4	5
VT	20 (15)	50 (40)	100 (20)	150 (35)	210 (30)
Init.	9,7 (9,2)	57,6 (17,9)	121,3 (17,1)	173,2 (16,2)	225,7 (13,5)
M, $N = 0$	12,5 (11,9)	60,9 (26,6)	120,7 (24,0)	166,3 (23,7)	214,6 (22,1)
M, $N = 2$	19,6 (13,9)	51,8 (36,1)	100,6 (21,2)	149,6 (34,8)	208,3 (27,8)
M, $N = 3$	19,6 (14,0)	51,8 (36,1)	100,6 (21,2)	149,6 (34,8)	207,8 (28,2)
M, $N = 8$	19,6 (14,0)	51,8 (36,1)	100,7 (21,4)	149,6 (34,7)	208,0 (28,1)
CE, $N = 0$	12,4 (11,9)	60,9 (26,5)	121,3 (23,9)	167,1 (23,3)	215,6 (21,5)
CE, $N = 2$	13,7 (12,7)	61,5 (30,9)	101,7 (21,0)	151,3 (34,1)	210,4 (26,5)
CE, $N = 3$	16,4 (14,0)	57,1 (33,1)	101,0 (21,1)	150,6 (34,2)	210,2 (26,5)
CE, $N = 8$	17,6 (14,5)	56,3 (34,3)	100,8 (20,3)	150,6 (34,4)	210,2 (26,7)

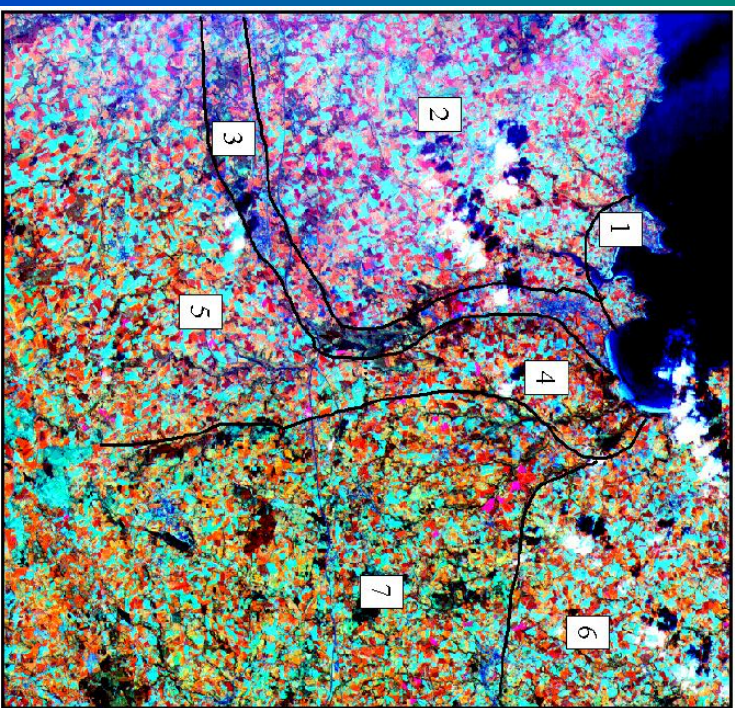
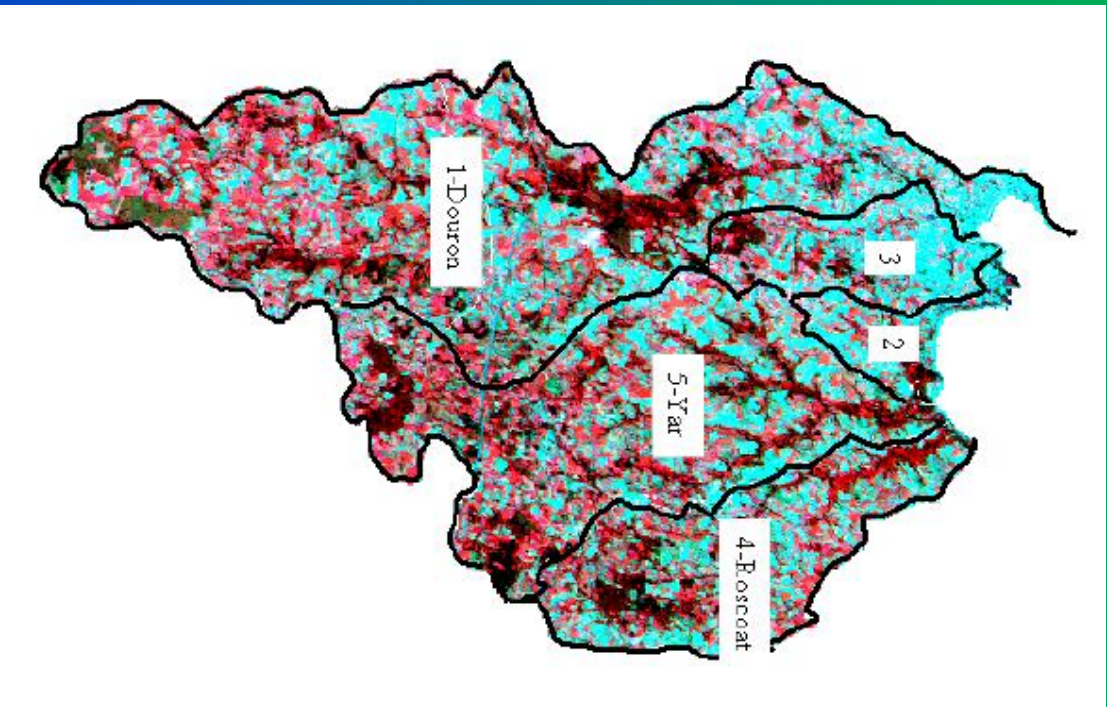
	EM			ICE			
	taux de bonne classif.	temps de calcul	nb. étapes	taux de bonne classif.	temps de calcul	nb. étapes	nb. iter. par étape
$T = 0$	62,8%	33 min	16	50%	430s	13	256
$T = 3$	98,85%	50s	9	96,2%	500s	38	10
$T = 4$	98,95%	45s	9	97,1%	375s	28	10
$T = 8$	98,9%	41s	8	97,3%	390s	30	10

# Application à la télédétection



## Problème réel de classification supervisée en télédétection

- Cadre : GSTB - collaboration avec COSTEL
- Lieu : Baie de Lannion
- Objectifs : comparer les techniques spectrales classiques de classification des géographes et les méthodes contextuelles (spatiales et hiérarchiques)
- Données : scènes multispectrales SPOT (29 déc. 1996 et 15 août 1997)
- Classification supervisée  
⇒ échantillons d'apprentissage



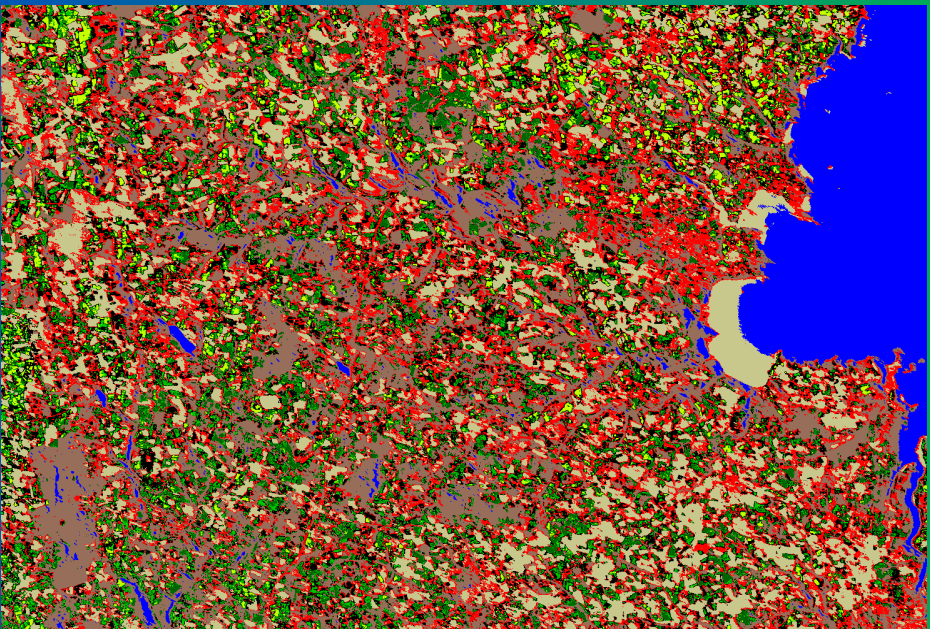
## Techniques spectrales

- Maximum de Vraisemblance :  
$$\hat{x}_s^{MV} = \arg \min_i (y_s - M_i)^T \Sigma_i^{-1} (y_s - M_i) + \log(\det \Sigma_i)$$
- Analyse discriminante pas à pas : maximiser  
différentiation entre classes  
⇒ Distance de Mahalanobis :  
$$\hat{x}_s^{AD} = \arg \min_i (y_s - M_i)^T \Sigma_i^{-1} (y_s - M_i)$$
- Distance minimale :  
$$\hat{x}_s^{DM} = \arg \min_i (y_s - M_i)^T (y_s - M_i)$$

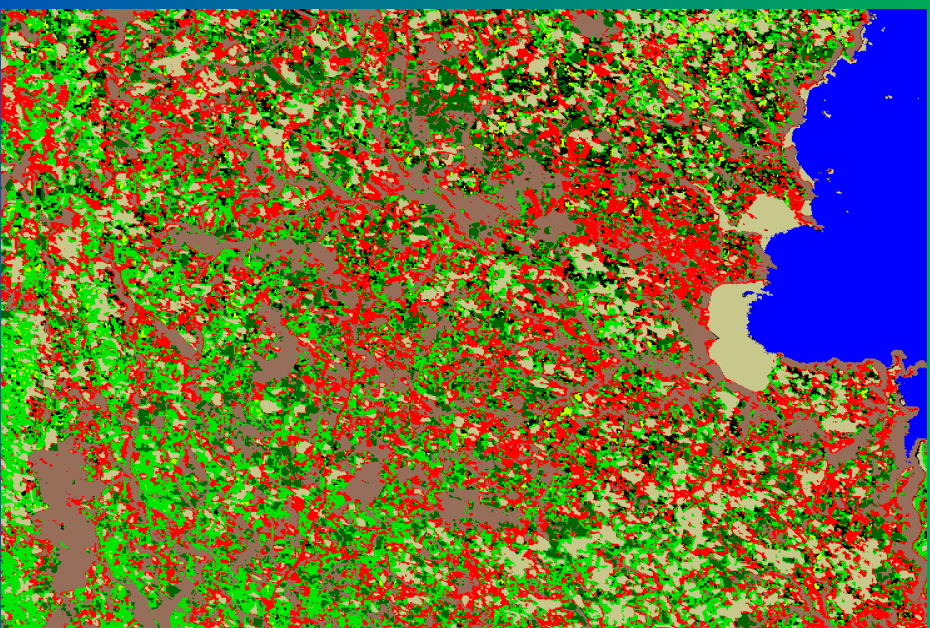
où  $M_i$  vecteur des moyennes de classe  $i$ ,  $\Sigma_i$  matrice de covariance de la classe  $i$  et  $\Sigma$  matrice de variance de l'échantillon complet

## Résultats

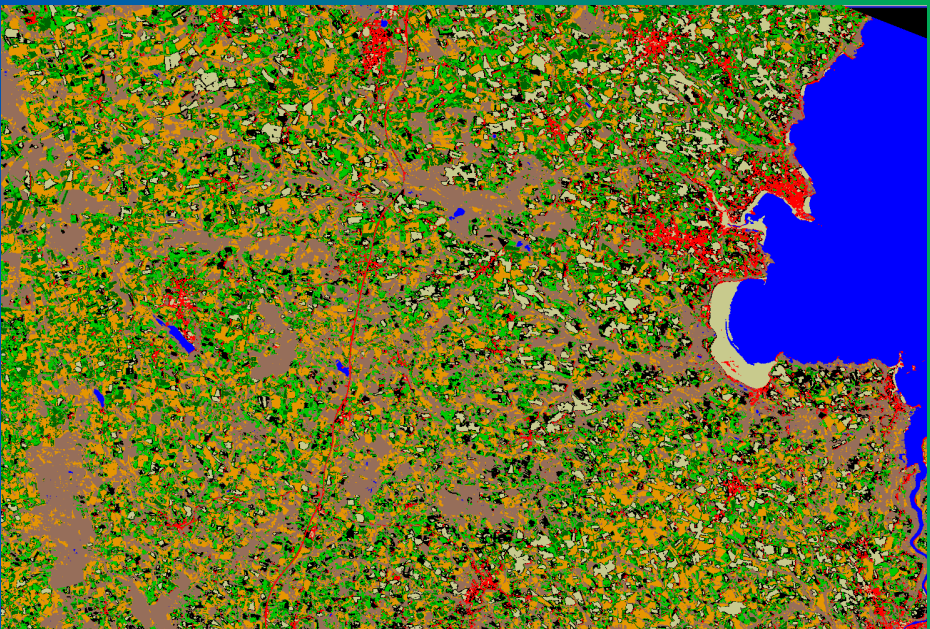
Méthodes	Hiver		Été	
	$\tau$	$\kappa$	$\tau$	$\kappa$
Analyse Discriminante	88,8%	83,6%	78,8%	73,8%
Distance Minimale	82,1%	74,0%	75,9%	70,5%
Max. Vraisemblance	89,0%	84,1%	79,3%	74,3%
ICM	89,2%	84,5%	81,1%	76,5%
MAP arbre tronqué	89,0%	84,1%	81,6%	77,1%
MPM arbre tronqué	89,3%	84,5%	83,5%	79,5%



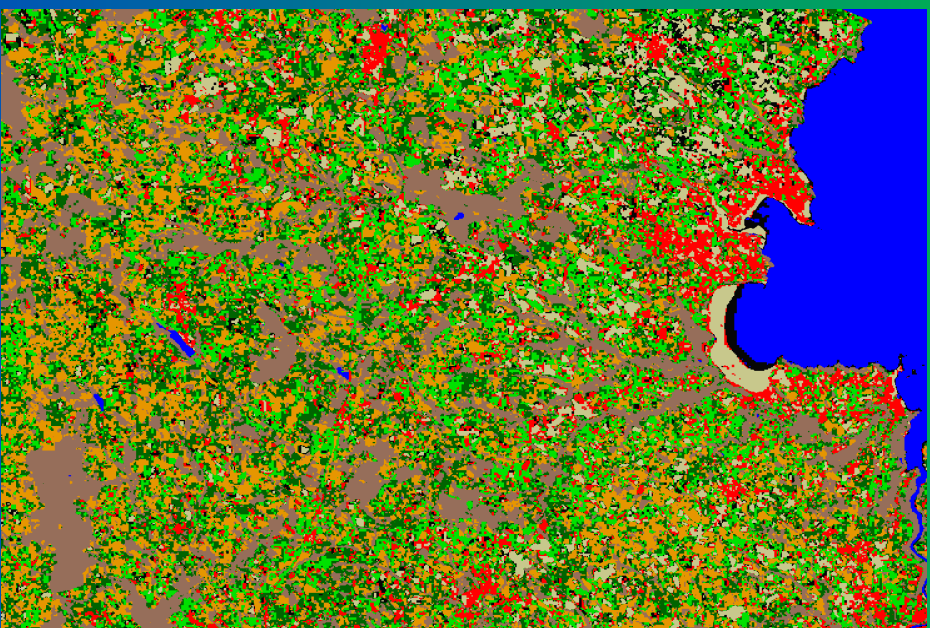
Analyse discriminante



MPM sur arbre tronqué



Analyse discriminante



MPM sur arbre tronqué

## Conclusion

- Conclusions sur le problème de télédétection
  - choix de la meilleure méthode dépendant de la date, du type de paysage et des thèmes à discriminer
  - faire mieux connaître les méthodes contextuelles auprès des thématiciens
- Introduction d'un nouveau type de graphe hiérarchique et des algorithmes associés
  - *a priori* spatial et en échelle
  - meilleurs résultats que sur grille 2-D et sur quadarbre, pour 3 ou 4 niveaux dans l'arbre tronqué
  - algorithmes semi-itératifs de coût calculatoire relativement équivalent aux algorithmes non-itératifs sur quadarbre
  - plus grande souplesse sur la taille des images traitées

## Perspectives

- Perspectives pratiques :
  - Estimation des paramètres d'*a priori* pour aider la manipulation des méthodes contextuelles par les thématiciens
  - Application à des données aéroportées de type CASI
- Perspectives théoriques :
  - Étudier l'impact d'une paramétrisation sur les algorithmes d'estimation de paramètres
  - Rendre la structure adaptative
  - Contrepartie continue des modèles sur arbre



tronqué et étude d'un couplage continu-discret