

Processus ponctuels marqués et reconnaissance des formes

Radu Stoica
email : Radu.Stoica@avignon.inra.fr

INRA Avignon - Biometrie
Domaine St. Paul, site Agroparc
84914 Avignon, Cedex 9
France

Quelques préliminaires

■ **Définition de la forme :**

- quelque chose de "complexe" : construite à partir des objets "simples" qui interagissent entre eux
- exemple : lego

■ **Question :**

- trouver la forme cachée dans les données
- difficulté : chercher dans des grandes quantités des données \Rightarrow automatisation

■ **Hypothèse clef :**

- la forme que l'on cherche est la réalisation d'un processus ponctuel marqué - processus objet

Processus ponctuels marqués (1)

■ **Définition :** un ensemble aléatoire $\mathbf{y} = \{y_n = (k_n, m_n)\}$ tel que $n \in \mathbb{N}_0$ avec :

- points k_n : emplacement des objets
- marques m_n : caractéristiques des objets (forme géométrique d'un objet aléatoire, mais aussi : espèce, âge, maladie, etc.)

Le plus simple processus ponctuel est *le processus ponctuel de Poisson*.

■ **Cadre probabiliste :** exemple d'une densité basée sur des interactions par paires

$$p(\mathbf{y}) \propto \beta^{n(\mathbf{y})} \prod_{y_i, y_j \in \mathbf{y}; y_i \sim y_j} h(y_i, y_j)$$

avec $h(\cdot, \cdot)$ la fonction d'interaction \Rightarrow : Strauss, hard-core, Widom-Rowlinson, Candy, Bisous, etc.

Processus ponctuels marqués (2)

■ **Processus ponctuels de Gibbs :**

- physique statistique
- la distribution de Gibbs caractérise l'état d'équilibre des particules à l'intérieur de systèmes physiques fermés

■ **Distribution de Gibbs :** une autre forme pour décrire un p.p.m.

$$p(\mathbf{y}) \propto \exp[-U(\mathbf{y})]$$

si par exemple la fonction d'énergie est :

$$U(\mathbf{y}) = - \sum_{i=1}^{n(\mathbf{y})} \log \beta - \sum_{1 \leq i < j \leq n(\mathbf{y})} \log h(y_i, y_j)$$

Et toc ...

- **Réponse** : la forme recherchée est représentée par la configuration d'objets $\hat{\mathbf{y}}$ telle que

$$\hat{\mathbf{y}} = \max_{\mathbf{y}} p(\mathbf{y}) = \min_{\mathbf{y}} U(\mathbf{y})$$

- **Méthodologie** : sous les hypothèses énoncées
- modélisation : construction d'un processus ponctuel marqué
 - simulation du modèle : construction des dynamiques MCMC
 - inférence statistique : détection des positions et des caractéristiques des objets, estimation des paramètres du modèle

Modélisation (1)

■ Observer les données :

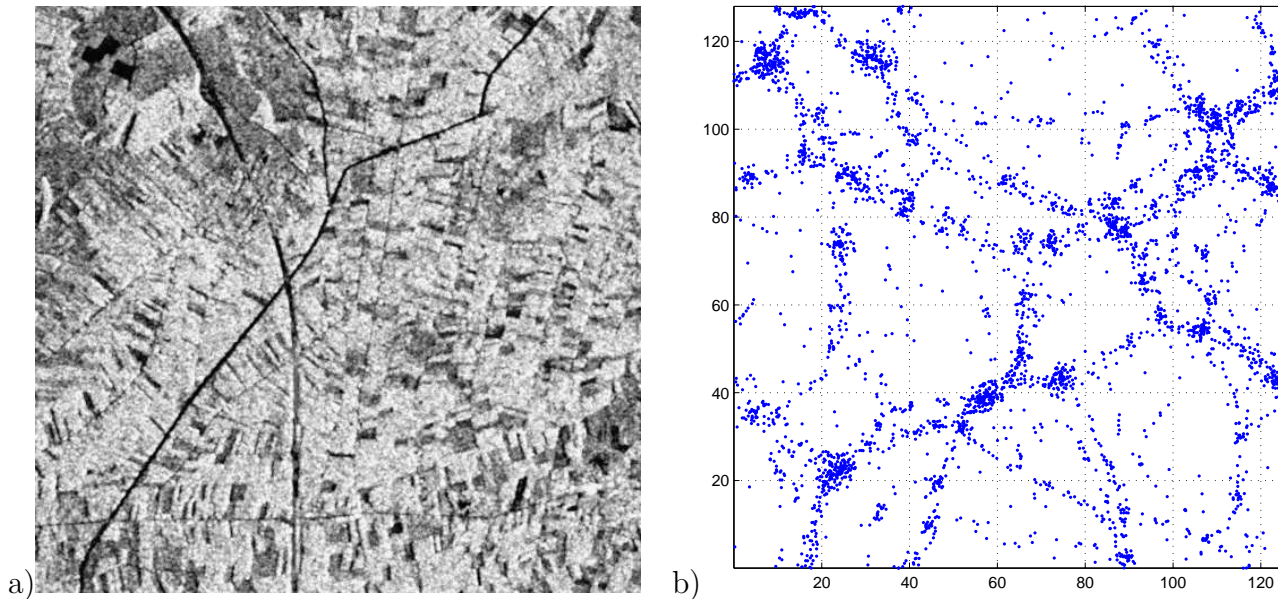


Figure 1: a) Image aérienne - région rurale en Malaisie (<http://southport.jpl.nasa.gov>), b) Simulation 2d des configurations des galaxies (Université de Valence - Espagne et Tartu Observatory - Estonie).

■ Formuler une hypothèse : les réseaux fins sont construites à partir de segments qui se connectent

Modélisation (2)

■ **Caractéristiques d'un segment :** $y = (k, l, w, \theta)$

- $k \in K \subset \mathbb{R}^2$: coordonnées du centre
- $l \in [l_{\min}, l_{\max}]$: longueur
- $w \in [w_{\min}, w_{\max}]$: épaisseur
- $\theta \in [0, \pi)$: orientation

■ **Candy modèle :** processus ponctuel marqué qui simule un réseau de segments permettant

- favoriser les alignements et les connexions entre les segments
- pénaliser les segments non connectés et ceux qui tendent à se superposer
- permettre les croisements

Modélisation (3)

■ Densité de probabilité du Candy modèle :

$$p(\mathbf{y}) \propto \prod_{i=1}^{n(\mathbf{y})} \exp \left[\frac{l_i - l_{\max}}{l_{\max}} + \frac{w_i - w_{\max}}{w_{\max}} \right] \times \gamma_f^{n_f(\mathbf{y})} \gamma_s^{n_s(\mathbf{y})} \gamma_d^{n_d(\mathbf{y})} \gamma_o^{n_o(\mathbf{y})} \gamma_r^{n_r(\mathbf{y})}$$

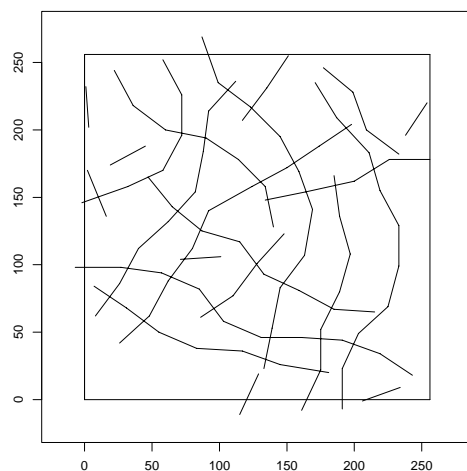
avec $\gamma_f, \gamma_s, \gamma_d > 0$ and $\gamma_o, \gamma_r \in (0, 1)$, les paramètres du modèle et $n_f(\mathbf{y})$, $n_s(\mathbf{y})$, $n_d(\mathbf{y})$, $n_o(\mathbf{y})$, $n_r(\mathbf{y})$ les statistiques suffisantes correspondantes.

■ Energie de Gibbs :

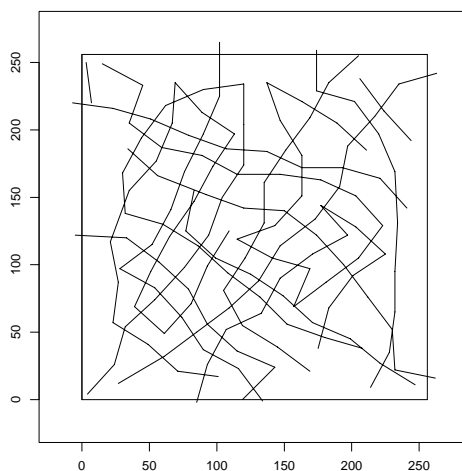
$$U_c(\mathbf{y}) = - \sum_{i=1}^{n(\mathbf{y})} \frac{l_i - l_{\max}}{l_{\max}} + \frac{w_i - w_{\max}}{w_{\max}} - n_f(\mathbf{y}) \log \gamma_f \\ - n_s(\mathbf{y}) \log \gamma_s - n_d(\mathbf{y}) \log \gamma_d - n_o(\mathbf{y}) \log \gamma_o - n_r(\mathbf{y}) \log \gamma_r$$

Modélisation (4)

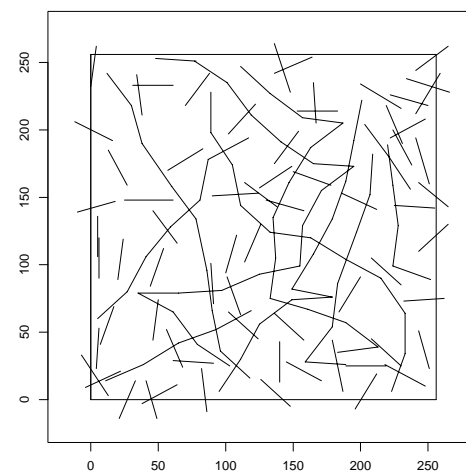
■ Topologies obtenues :



a)



b)



c)

Figure 2: Réalisations du Candy modèle avec des différentes valeurs des paramètres.

Quelques propriétés des processus ponctuels

■ **Avant passer à la simulation :** toujours prendre en compte

– stabilité au sens de Ruelle :

$$p(\mathbf{y}) \leq M^{n(\mathbf{y})}$$

– localement stable :

$$\frac{p(\mathbf{y} \cup \{\eta\})}{p(\mathbf{y})} \leq M$$

– processus de Ripley - Markov \approx la contribution d'un objet à l'énergie totale du système dépend de l'objet même et de l'interaction avec ses voisins seulement

– monotonie ou anti-monotonie : pouvoir imposer une relation d'ordre sur l'espace des configurations (inclusion)

Techniques MCMC pour simuler des processus ponctuels (1)

- **Processus de type naissance-mort** : ajouter (naissance) ou enlever (mort) un objet de la configuration
 - le calcul des rapports naissance/mort demande l'intégration sur tout l'espace des paramètres ; sinon M doit avoir des valeurs raisonnables
 - vitesse de mélangeance : faible
 - convergence : théoriquement à l'infini
- **Simulation exacte** : CFTP, clan des ancêtres, exact Metropolis-Hastings, Gibbs
 - la chaîne simulée indique elle même la convergence
 - M doit avoir des valeurs **trés, très raisonnables**
 - utilisable en pratique pour des valeurs restreintes des paramètres

Techniques MCMC pour simuler des processus ponctuels (2)

■ **Metropolis-Hastings** : un cas particulier des dynamiques à sauts réversibles

- calculs locaux
- noyaux de transitions qui aident le modèle
- grande souplesse versus convergence à l'infini

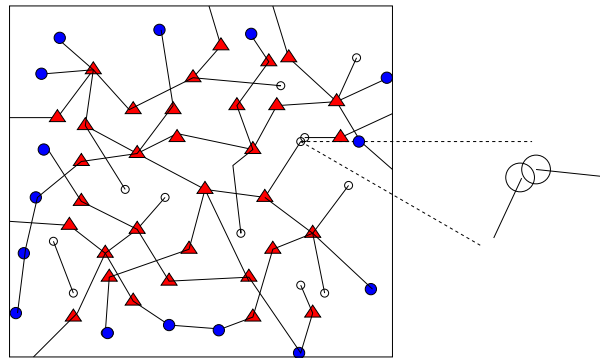


Figure 3: Proposer des segments selon une densité non-uniforme.

Détéction des positions et des caractéristiques des objets (1)

■ Remarques :

- le Candy modèle tout seul ne détecte pas les réseaux dans les données
- il faut compléter le modèle avec un terme d'attache aux données $U_d(\mathbf{y})$:

$$p(\mathbf{y}) \propto \exp[-U_d(\mathbf{y}) - U_c(\mathbf{y})]$$

La configuration des objets recherchés est alors donnée par:

$$\hat{\mathbf{y}} = \min\{U_d(\mathbf{y}) + U_c(\mathbf{y})\}$$

- $U_d(\mathbf{y})$: terme relatif à l'emplacement du réseau dans l'image
- $U_c(\mathbf{y})$: terme jouant le rôle d'un prior
- $\hat{\mathbf{y}}$ presque un MAP : nous sommes pas tout à fait dans le cadre bayésien

Détéction des positions et des caractéristiques des objets (2)

■ Construction du terme d'attache aux données :

- **images aériennes** : tests d'hypothèses considérant que les pixels couverts par un segment suivent une distribution normale

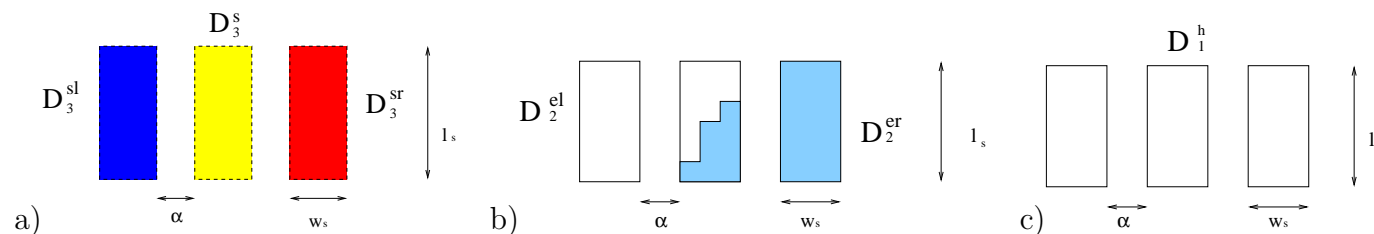


Figure 4: Masques des régions : a) trois régions homogènes, b) deux régions homogènes, c) une région homogène.

la contribution énergétique d'un segment est:

$$v(y_i) = \min \left\{ \log \frac{L(H_3)}{L(H_1)}, \log \frac{L(H_3)}{L(H_2)} \right\} + \text{route blanche/noire}$$

Détéction des positions et des caractéristiques des objets (3)

– données astronomiques : comptage des points

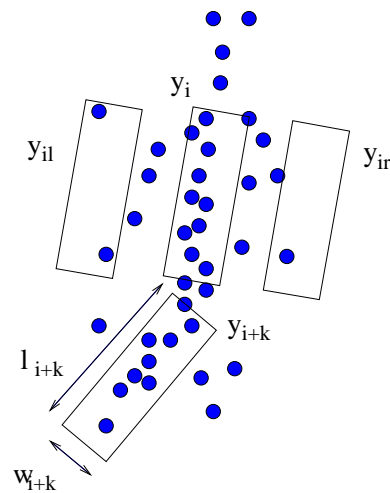


Figure 5: Détecter des segments dans une configuration de points.

la contribution énergétique d'un segment dans le champ de données \mathbf{d} est:

$$v(y_i) = 2n(\mathbf{d} \cap \tilde{y}_i) - n(\mathbf{d} \cap \tilde{y}_{ir}) - n(\mathbf{d} \cap \tilde{y}_{il}) + \text{quelques conditions}$$

Détéction des positions et des caractéristiques des objets (4)

■ **Energie totale** :

$$U(\mathbf{y}) = U_c(\mathbf{y}) + U_d(\mathbf{y}) = U_c(\mathbf{y}) + \sum_{i=1}^{n(\mathbf{y})} v(y_i)$$

■ **Optimisation globale** : recuit simulé

- échantillonner $p(\mathbf{y})^{\frac{1}{T}}$
- pendant que la température $T \rightarrow 0$
- convergence assurée par un schéma de descente logarithmique

Extraction réseaux routiers

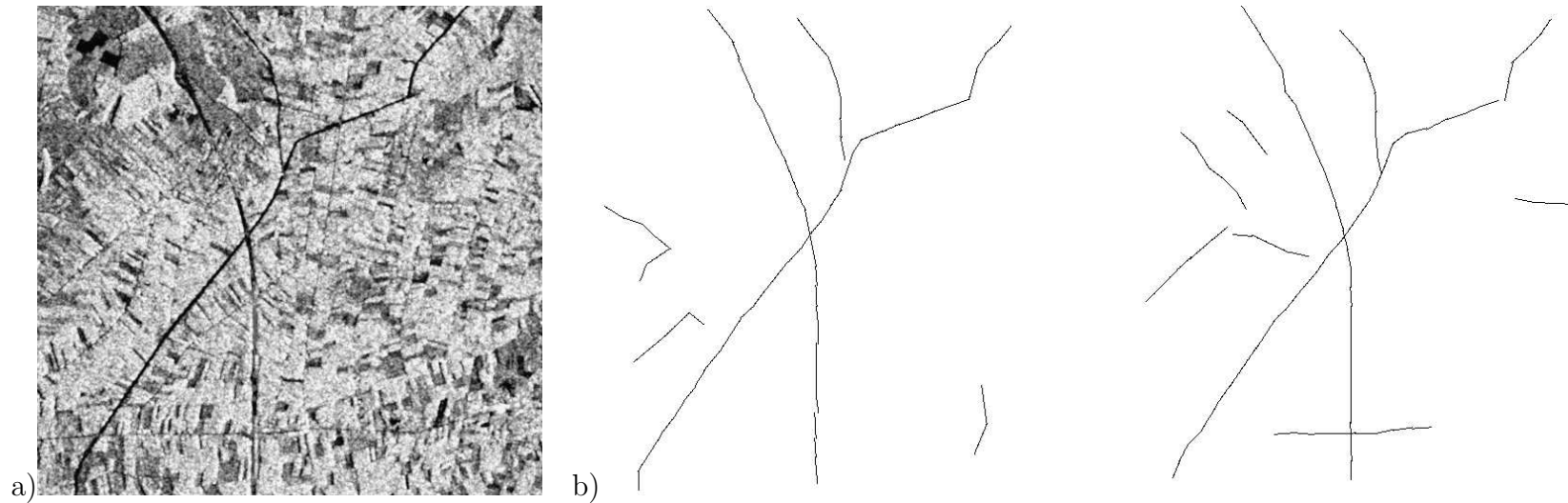


Figure 6: Région rurale en Malaisie: a) image originale; b) résultats obtenus.

Fôrets galleries

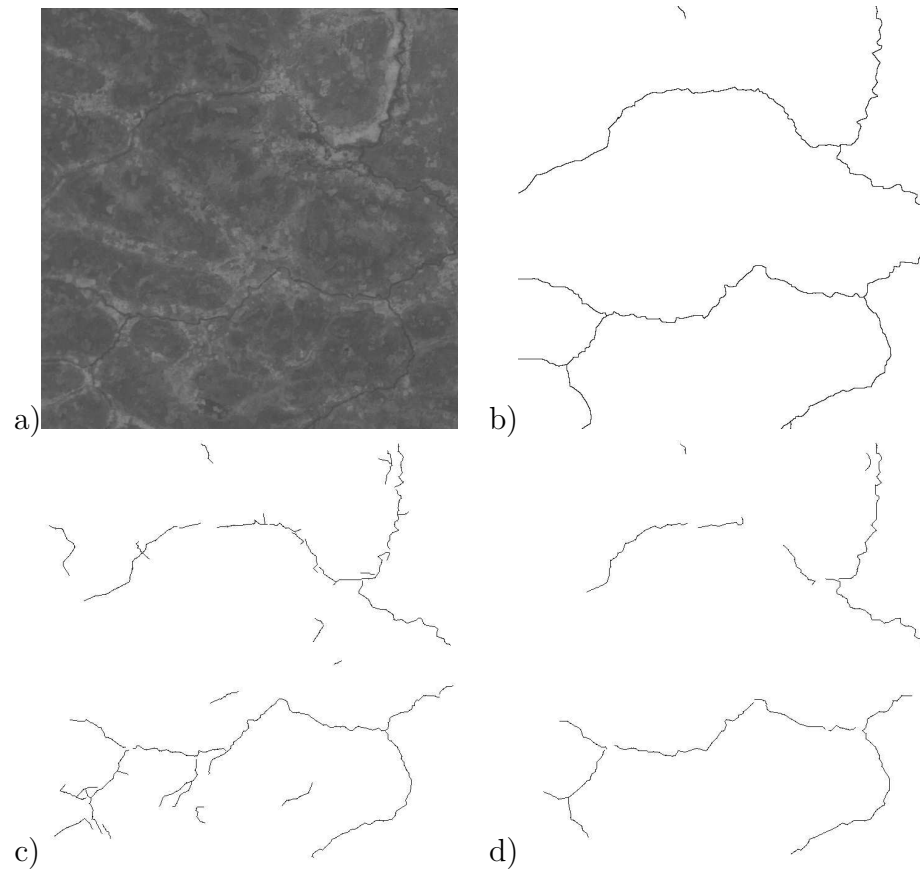


Figure 7: Extraction des fôrets galleries : a) image originale ; b) vérité terrain ; c)-d) resultats obtenus. Données BRGM.

Structure filamentaire dans des données astronomiques

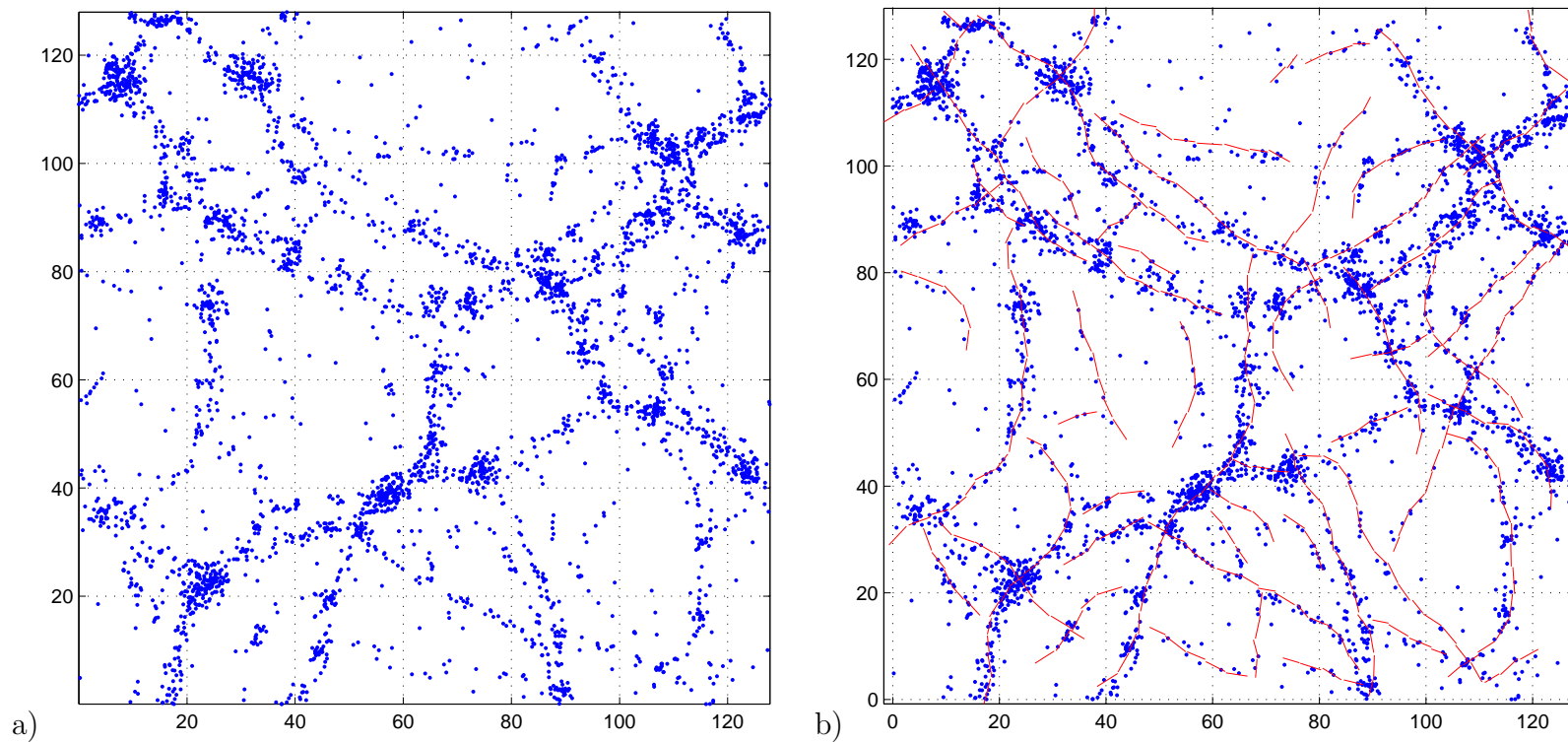


Figure 8: Filaments formés par des galaxies : a) données simulées, b) résultat obtenu.

Estimation des paramètres du Candy modèle (1)

■ **Données complètes :** le Candy modèle peut s'écrire

$$p_{\omega}(\mathbf{y}) = \alpha(\omega) \exp [t(\mathbf{y})^T \omega] h(\mathbf{y})$$

avec la constante de normalisation $\alpha(\omega)$, le vecteur des statistiques suffisantes

$$t(\mathbf{y}) = (n_f(\mathbf{y}), n_s(\mathbf{y}), n_d(\mathbf{y}), n_o(\mathbf{y}), n_r(\mathbf{y}))^T,$$

le vecteur des paramètres

$$\omega = (\log \gamma_f, \log \gamma_s, \log \gamma_d, \log \gamma_o, \log \gamma_r)^T,$$

et :

$$h(\mathbf{s}) = \prod_{i=1}^n \exp \left[\frac{l_i - l_{\max}}{l_{\max}} + \frac{w_i - w_{\max}}{w_{\max}} \right].$$

Estimation des paramètres du Candy modèle (2)

Suivant l'approche de [Geyer, 1999], le rapport des constantes de normalisation peut s'écrire :

$$\alpha(\omega_0)/\alpha(\omega) = E_{\omega_0} \exp [t(Y)^T(\omega - \omega_0)]$$

ce qui donne pour la log-vraisemblance par rapport à un paramètre fixé ω_0 :

$$\begin{aligned} l(\omega) &= \log \frac{p_\omega(\mathbf{y})}{p_{\omega_0}(\mathbf{y})} \\ &= t(\mathbf{y})^T(\omega - \omega_0) - \log E_{\omega_0} \exp [t(Y)^T(\omega - \omega_0)]. \end{aligned} \quad (1)$$

avec l'espérance dans (??) qui peut être calculée par des approximations de type *Monte Carlo* :

$$\frac{1}{n} \sum_{i=1}^n \exp [t(Y_i)^T(\omega - \omega_0)]$$

Estimation des paramètres du Candy modèle (3)

Il est alors possible de calculer, le gradient

$$\nabla l(\omega) = t(\mathbf{s}) - E_{\omega} t(S)$$

et le hessien

$$I(\omega) = -\nabla^2 l(\omega) = \text{Var}_{\omega} t(S)$$

de la fonction de log-vraisemblance.

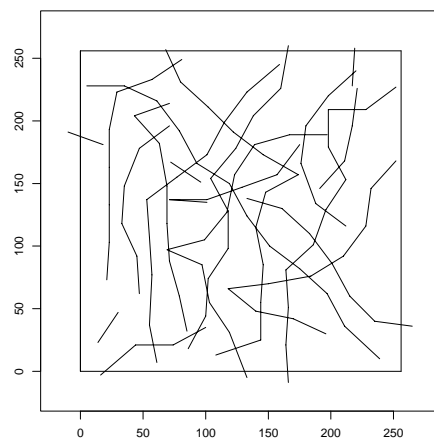
■ **Méthode de gradient à pas optimal** : pour obtenir une valeur initiale suffisamment proche de l'estimateur de maximum de vraisemblance

$$\begin{cases} l_n(\omega_k + \rho(\omega_k) \nabla l_n(\omega_k)) = \max_{\rho \in \mathbb{R}} l_n(\omega_k + \rho \nabla l_n(\omega_k)) \\ \omega_{k+1} = \omega_k + \rho(\omega_k) \nabla l_n(\omega_k) \end{cases}$$

avec $\rho(\omega_k)$ le pas optimal.

Estimation des paramètres du Candy modèle (4)

■ **Modèle de référence:**



Paramètres du modèle
$\gamma_f = 0.0002$
$\gamma_s = 0.05$
$\gamma_d = 12.2$
$\gamma_o = 0.08$
$\gamma_r = 0.08$

Statistiques suffisantes
$n_f = 4$
$n_s = 34$
$n_d = 63$
$n_o = 12$
$n_r = 9$

Figure 9: Réalisation du modèle à estimer.

Estimation des paramètres du Candy modèle (5)

■ Résultats :

Conditions Initiales	Méthode de gradient	Monte Carlo MLE
$\omega_f^i = -9.5$	$\hat{\omega}_f^0 = -8.37$	$\hat{\omega}_f^n = -8.32$
$\omega_s^i = -4.0$	$\hat{\omega}_s^0 = -2.74$	$\hat{\omega}_s^n = -2.73$
$\omega_d^i = 1.5$	$\hat{\omega}_d^0 = 2.46$	$\hat{\omega}_d^n = 2.47$
$\omega_o^i = -3.5$	$\hat{\omega}_o^0 = -2.13$	$\hat{\omega}_o^n = -2.17$
$\omega_r^i = -3.5$	$\hat{\omega}_r^0 = -2.42$	$\hat{\omega}_r^n = -2.42$

Figure 10: Estimation des paramètres du modèle de référence.

■ Etude des erreurs :

Ecart type asymptotique $\hat{\omega} - \omega_{vrai}$	MCSE $\hat{\omega} - \hat{\omega}^n$
0.51	0.002
0.23	0.003
0.17	0.001
0.30	0.002
0.31	0.005

Figure 11: Erreurs d'estimation.

Estimation des paramètres du Candy modèle (6)

■ Approximation Monte Carlo de la fonction de log-vraisemblance :

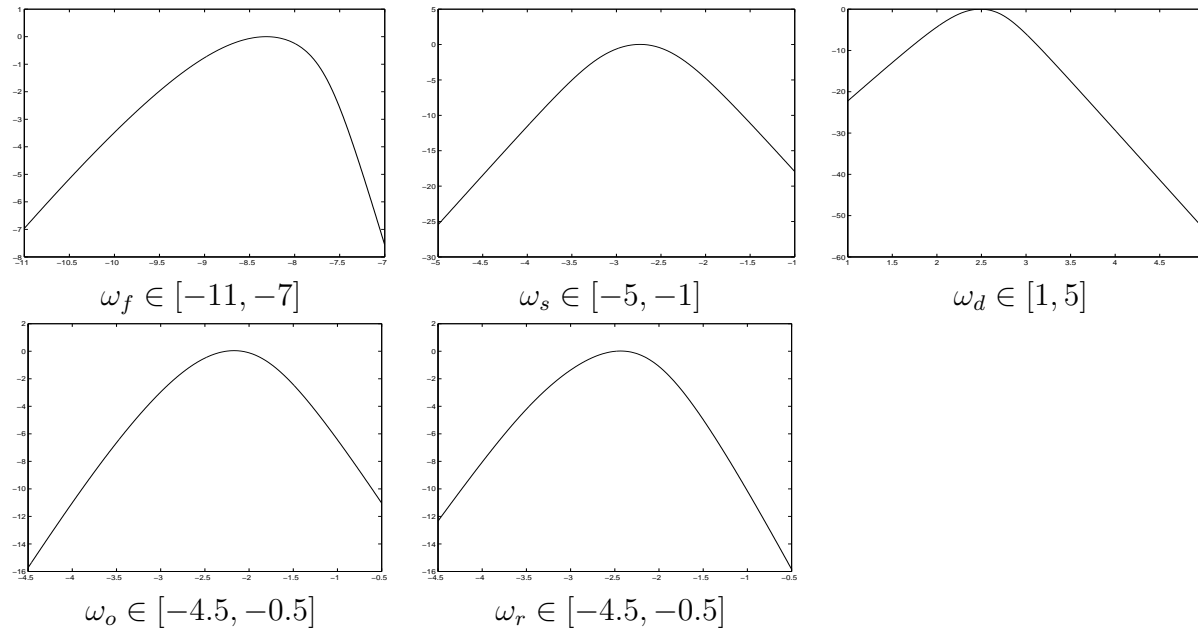


Figure 12: Représentation de la fonction de log-vraisemblance par rapport à chacune des variables prises individuellement.

Les gens ...

- **INRIA** : Xavier Descombes, Josiane Zerubia

- **IGN et BRGM** : Hervé Le Men et Nicolas Baghdadi

- **CWI** : Marie-Colette van Lieshout

- **UJI** : Pablo Gregori, Jorge Mateu

- **OA-UV et TO** : Vicent Martinez et Enn Saar

... et la liste n'est pas encore finie ...

... et là tu fais quoi, toi ?

■ **INRA :**

- détection d'agrégats (Emilie et Rachid)
- modélisation paysage (Kassia, Hervé et Kien)
- discussions (André)
- discussions (Joël)

■ **Processus ponctuels :**

- applications à des données $3d$: Bisous modèle
- estimation des paramètres : données complètes et incomplètes
- simulation exacte : problème ouvert

... des Bisous ... juste pour voir ...

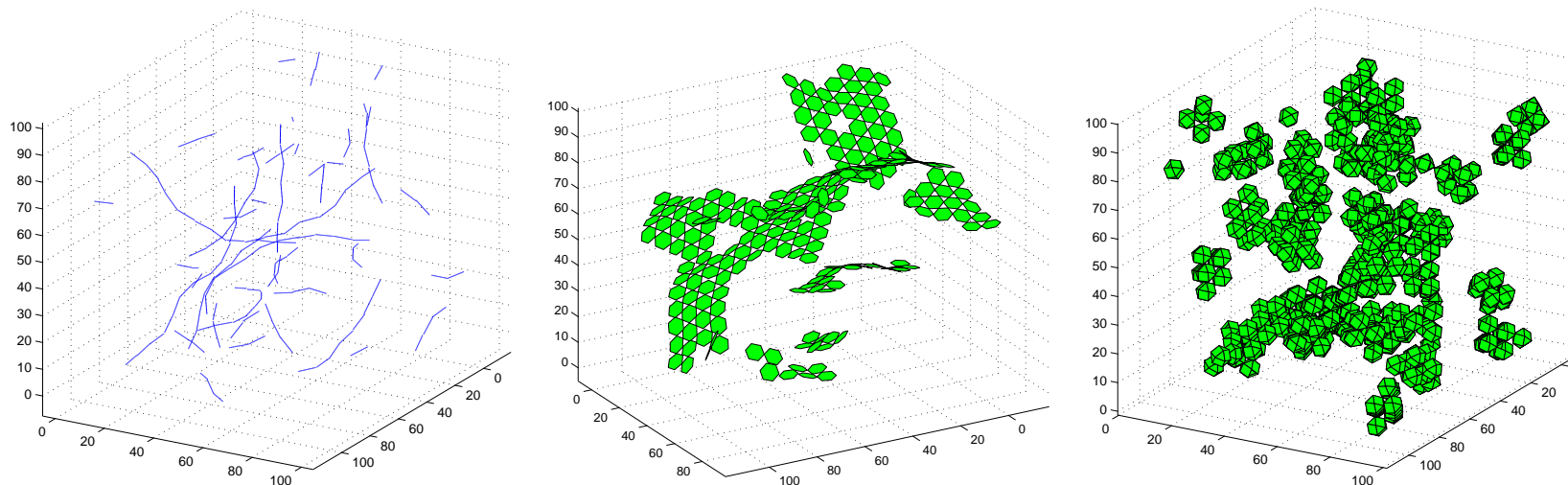


Figure 13: Formes aléatoires générées avec Bisous modèle.