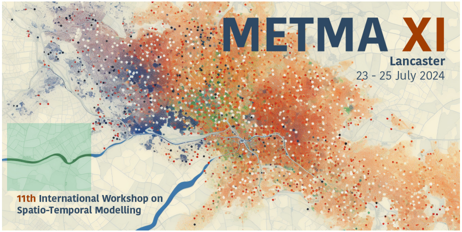
Statistical inference for random T-tessellations models: application to agricultural landscape modeling.

Katarzyna Adamczyk-Chauvat, Mouna Kassa, Julien Papaïx, Kiên Kiêu, Radu S. Stoica

Statistical inference for random T-tessellations models: application to agricultural landscape modeling

Katarzyna Adamczyk-Chauvat[1] Mouna Kassa[2] Julien Papaïx[1]
Kiên Kiêu[1] Radu S. Stoica[3]

[1]INRAE, Dep. MathNum, Jouy-en-Josas, [2]INSA, Rennes, [3]Université de Lorraine-CNRS-Inria, Nancy



Outline

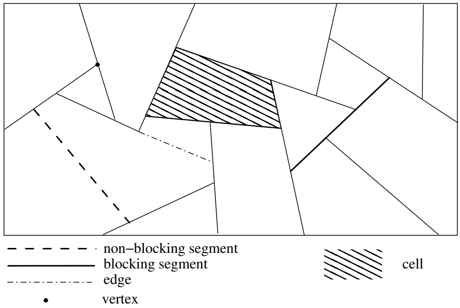
T-tessellation

Definition

A planar tessellation of a bounded window W is called a *T-tessellation of W* if it satisfies two conditions:

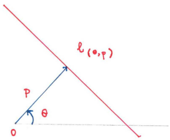
(i) all the vertices in $\mathrm{int}(W)$ are of degree 3,

(ii) two of the three edges connected by the same vertex are collinear.



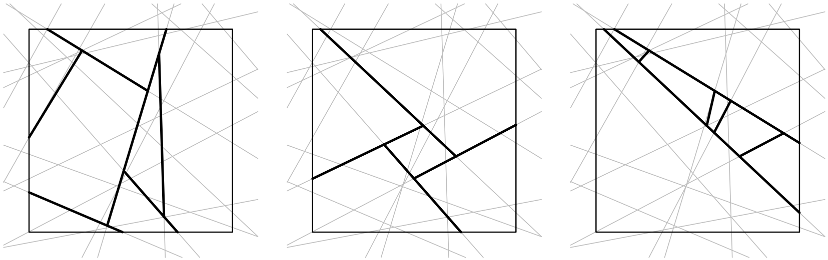
Introducing randomness

$$l_{(\theta,p)} \longleftrightarrow (\theta, p)$$



$$\mathcal{L} = \{(\theta, p) : l_{(\theta,p)} \cap W \neq \emptyset\}$$

T-tessellations supported by the same line set $L \subset \mathcal{L}$:



Poisson lines as a skeleton to define a random T-tessellation

Random T-tessellation models

Definition (Completely Random T-tessellation model)

Let $\mathbf{L}$ be a unit-rate Poisson line process on $\mathcal{L}$. Let $\mathcal{T}(\mathbf{L})$ be the set of $T \in \mathcal{T}$ with segments lying on the lines of $\mathbf{L}$ and such that for each such line $l_i, l_i \cap T$ consists of exactly one segment. The CRTT model is defined by:

$$\mu(A) = P(T \in A) = \frac{1}{Z}\mathbb{E}\left(\sum_{T \in \mathcal{T}(\mathbf{L})} \mathbb{I}_A(T)\right)$$

for $A \subseteq \mathcal{T}$ and Z - normalising constant.

Random T-tessellation models

Definition (Gibbs model)

Let $h : \mathcal{T} \to [0, \infty[$ be the non-negative function, integrable with respect to the measure μ. The Gibbs random T-tessellation model is defined by:

$$P(dT) \propto h(T)\mu(dT).$$

The function $U(T) = -\log h(T)$ is called the energy of the model.

Exponential family:

$$h_\theta(T) = \exp\left(t(T)^T \theta\right)$$

$t(T)$: vector of tessellation statistics, $\theta \in \Theta$: vector of model parameters.

Sampling from the Gibbs model

Metropolis-Hastings algorithm based on splits, merges and flips

---

Require: initial tessellation T_0, density $h(T)$, $(p_\mathrm{s}, p_\mathrm{m}, p_\mathrm{m})$, uniform measures $(q_s^T(S), q_m^T(M), q_f^T(F))$.

The current T-tessellation is T_n.

Draw the update type t from $\{\mathrm{s}, \mathrm{m}, \mathrm{f}\}$ with probabilities $(p_\mathrm{s}, p_\mathrm{m}, p_\mathrm{f})$.

if there is no update of type t applicable to T_n **then**

  $T_{n+1} = T_n$

else

  Draw the update U of type t from the distribution $q_t^{T_n}(U)$.

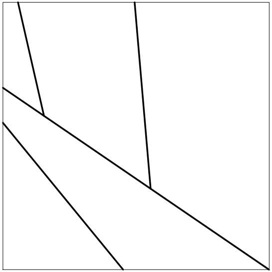
  Compute the Hastings ratio:

$$r_t(T, U) = \frac{h(U(T))}{h(T)} \frac{p_{t^{-1}}}{p_t} \frac{q_{t^{-1}}^{U(T)}(U^{-1})}{q_t^T(U)}$$

  Accept the tessellation $T_{n+1} = U(T_n)$ with probability

$$\min\{1, r_t(T_n, U)\}.$$

end if



$\mathbb{S}_T, \mathbb{M}_T, \mathbb{F}_T$: sets of splits, merges, flips applicable to T

q_S^T, q_M^T, q_F^T: uniform measures on $\mathbb{S}_T, \mathbb{M}_T, \mathbb{F}_T$

Sampling from the Gibbs model

Metropolis-Hastings algorithm based on splits, merges and flips

Require: initial tessellation T_0, density $h(T)$, $(p_{\mathrm{s}}, p_{\mathrm{m}}, p_{\mathrm{m}})$, uniform measures $(q_s^T(S), q_m^T(M), q_f^T(F))$.

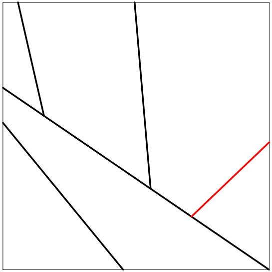
The current T-tessellation is T_n.
Draw the update type t from $\{\mathrm{s}, \mathrm{m}, \mathrm{f}\}$ with probabilities $(p_{\mathrm{s}}, p_{\mathrm{m}}, p_{\mathrm{f}})$.
if there is no update of type t applicable to T_n **then**
    $T_{n+1} = T_n$
else
    Draw the update U of type t from the distribution $q_t^{T_n}(U)$.
    Compute the Hastings ratio:

$$r_t(T, U) = \frac{h(U(T))}{h(T)} \frac{p_{t^{-1}}}{p_t} \frac{q_{t^{-1}}^{U(T)}(U^{-1})}{q_t^T(U)}$$

    Accept the tessellation $T_{n+1} = U(T_n)$ with probability

$$\min\{1, r_t(T_n, U)\}.$$

end if



$\mathbb{S}_T, \mathbb{M}_T, \mathbb{F}_T$: sets of splits, merges, flips applicable to T

q_S^T, q_M^T, q_F^T: uniform measures on $\mathbb{S}_T, \mathbb{M}_T, \mathbb{F}_T$

Sampling from the Gibbs model

Metropolis-Hastings algorithm based on splits, merges and flips

Require: initial tessellation T_0, density $h(T)$, $(p_{\mathrm{s}}, p_{\mathrm{m}}, p_{\mathrm{m}})$, uniform measures $(q_s^T(S), q_m^T(M), q_f^T(F))$.

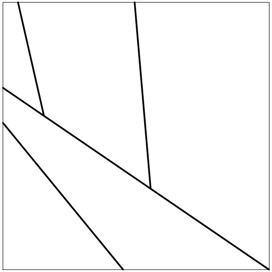
The current T-tessellation is T_n.
Draw the update type t from $\{\mathrm{s}, \mathrm{m}, \mathrm{f}\}$ with probabilities $(p_{\mathrm{s}}, p_{\mathrm{m}}, p_{\mathrm{f}})$.
if there is no update of type t applicable to T_n **then**
    $T_{n+1} = T_n$
else
    Draw the update U of type t from the distribution $q_t^{T_n}(U)$.
    Compute the Hastings ratio:

$$r_t(T, U) = \frac{h(U(T))}{h(T)} \frac{p_{t^{-1}}}{p_t} \frac{q_{t^{-1}}^{U(T)}(U^{-1})}{q_t^T(U)}$$

    Accept the tessellation $T_{n+1} = U(T_n)$ with probability

$$\min\{1, r_t(T_n, U)\}.$$

end if



$\mathbb{S}_T, \mathbb{M}_T, \mathbb{F}_T$: sets of splits, merges, flips applicable to T

q_S^T, q_M^T, q_F^T: uniform measures on $\mathbb{S}_T, \mathbb{M}_T, \mathbb{F}_T$

Sampling from the Gibbs model

Metropolis-Hastings algorithm based on splits, merges and flips

Require: initial tessellation T_0, density $h(T)$, $(p_{\mathrm{s}}, p_{\mathrm{m}}, p_{\mathrm{m}})$, uniform measures $(q_s^T(S), q_m^T(M), q_f^T(F))$.

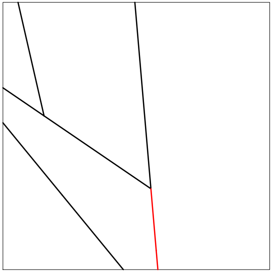
The current T-tessellation is T_n.
Draw the update type t from $\{\mathrm{s}, \mathrm{m}, \mathrm{f}\}$ with probabilities $(p_{\mathrm{s}}, p_{\mathrm{m}}, p_{\mathrm{f}})$.
if there is no update of type t applicable to T_n **then**
    $T_{n+1} = T_n$
else
    Draw the update U of type t from the distribution $q_t^{T_n}(U)$.
    Compute the Hastings ratio:

$$r_t(T, U) = \frac{h(U(T))}{h(T)} \frac{p_{t^{-1}}}{p_t} \frac{q_{t^{-1}}^{U(T)}(U^{-1})}{q_t^T(U)}$$

    Accept the tessellation $T_{n+1} = U(T_n)$ with probability

$$\min\{1, r_t(T_n, U)\}.$$

end if



$\mathbb{S}_T, \mathbb{M}_T, \mathbb{F}_T$: sets of splits, merges, flips applicable to T

q_S^T, q_M^T, q_F^T: uniform measures on $\mathbb{S}_T, \mathbb{M}_T, \mathbb{F}_T$

Sampling from the Gibbs model

Metropolis-Hastings algorithm based on splits, merges and flips

Require: initial tessellation T_0, density $h(T)$, (p_s, p_m, p_m), uniform measures $(q_s^T(S), q_m^T(M), q_f^T(F))$.

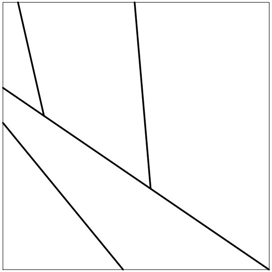
The current T-tessellation is T_n.
Draw the update type t from $\{s, m, f\}$ with probabilities (p_s, p_m, p_f).
if there is no update of type t applicable to T_n **then**
  $T_{n+1} = T_n$
else
  Draw the update U of type t from the distribution $q_t^{T_n}(U)$.
  Compute the Hastings ratio:

$$r_t(T, U) = \frac{h(U(T))}{h(T)} \frac{p_{t^{-1}}}{p_t} \frac{q_{t^{-1}}^{U(T)}(U^{-1})}{q_t^T(U)}$$

  Accept the tessellation $T_{n+1} = U(T_n)$ with probability

$$\min\{1, r_t(T_n, U)\}.$$

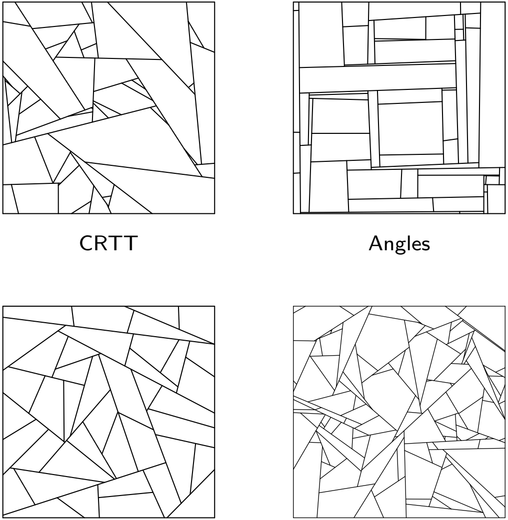
end if



$\mathbb{S}_T, \mathbb{M}_T, \mathbb{F}_T$: sets of splits, merges, flips applicable to T

q_S^T, q_M^T, q_F^T: uniform measures on $\mathbb{S}_T, \mathbb{M}_T, \mathbb{F}_T$

Examples of model simulations



CRTT   Angles

Areas   ACS

Additional tools

Definition

Let $\mathscr{C}_s = \{(s, T) : T \in \mathcal{T}, s \in \mathbb{S}_T\}$. The split Campbell measure of $\mathbf{T} \sim P$ is defined by:

$$C_s^!(B) = \mathbb{E} \sum_{s \in \mathbb{M}_{\mathbf{T}}} \mathbb{I}_B(s, \mathbf{T} \setminus \{s\}) \quad \forall B \subset \mathscr{C}_s$$

$$C_s^!(ds, dT) = \underbrace{\frac{h(T \cup \{s\})}{h(T)}}_{\text{split Papangelou intensity}\lambda_s(s,T)} dsP(dT)$$

Analogous definitions hold for flips.

Outline

Minimum contrast estimation

$$\rho(\theta; T) = -\sum_{m \in \mathbb{M}_T} \log \lambda_{s,\theta}(m^{-1}, m(T)) + \int_{\mathbb{S}_T} \lambda_{s,\theta}(s, T) ds - \sum_{f \in \mathbb{F}_T} \log \lambda_{f,\theta}(f^{-1}, f(T)) + \sum_{f \in \mathbb{F}_T} \lambda_{f,\theta}(f, T)$$

Proposition

If $\mathbf{T} \sim P_{\theta^*}$ *then for every* $\theta \in \Theta$*:*

$$\mathbb{E}_{\theta^*} \rho(\theta; \mathbf{T}) = -\int_{\mathscr{C}_s} \log \lambda_{s,\theta}(s, T) C^!_{s,\theta^*}(ds, dT) + \int_{\mathscr{C}_s} \lambda_{s,\theta}(s, T) ds P_{\theta^*}(dT) - \int_{\mathscr{C}_f} \log \lambda_{f,\theta}(s, T) C^!_{f,\theta^*}(df, dT) + \int_{\mathscr{C}_f} \lambda_{f,\theta}(f, T) df P_{\theta^*}(dT)$$

$\mathbb{E}_{\theta^*} \rho(\theta; \mathbf{T})$ *has a global minimum at* $\theta = \theta^*$.

$$\hat{\theta}_{MPL} = \arg\max_{\theta}(-\rho(\theta; T))$$

Statistical tools extended to T-tessellations

- Monte Carlo Maximum Likelihood to refine the pseudolikelihood estimation

$$\underbrace{\hat{L}(\theta; T)}_{\text{MC log likelihood}} = - < t(T), \theta > - \log \frac{1}{n} \sum_{i=1}^{n} \exp\left(- < t(T_i), \theta - \theta_0 >\right) + \text{const}$$

- Model assessment based on global envelope tests:

$$H_0 : \text{observed T-tessellation} \sim P_\theta$$

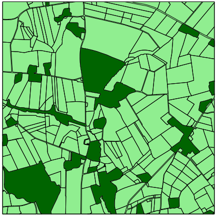
$$F = (F(r_1), \ldots, F(r_d)) : \text{functional statistic of tessellation}$$

$$\underbrace{F_1}_{\text{data}}, \underbrace{F_2, \ldots, F_s}_{\text{model simulations}}$$

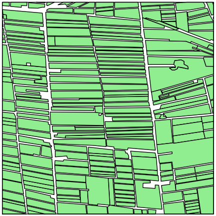
Outline

Landscapes approximated by T-tessellations

L_1: Selommes



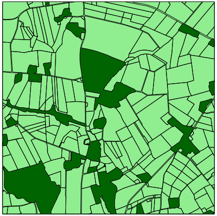
L_2: Kervidy



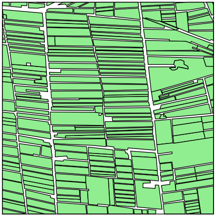
L_3: BVD

Landscapes approximated by T-tessellations

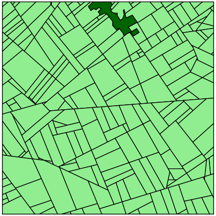
L_1: Selommes

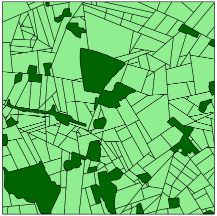


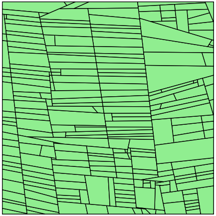
L_2: Kervidy



L_3: BVD







Model construction

Guidelines:

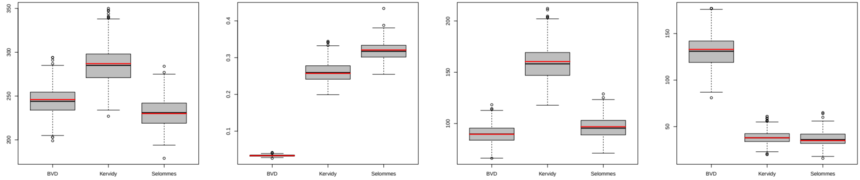
- accounts for typical landscape features
- highlights the main differences between patterns

$$
\begin{aligned}
U_\theta(T) \;&=\; \theta_1 \text{ number\_of\_segments}(T) && \leftarrow \textit{scale statistic } t_1(T) \\
&+\; \theta_2 \sum_{\text{faces } f} \text{area}(f)^2 && \leftarrow \textit{area statistic } t_2(T) \\
&+\; \theta_3 \sum_{\text{vertices } v} \left(\frac{\pi}{2} - \alpha(v)\right) && \leftarrow \textit{angle statistic } t_3(T) \\
&+\; \theta_4 \sum_{\text{faces } f} g\left(\frac{\text{length}}{\text{width}}(f)\right) && \leftarrow \textit{elongation statistic } t_4(T)
\end{aligned}
$$

Estimation results

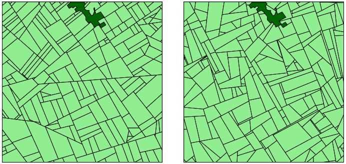
Model parameter	Landscape	$\hat{\theta}_{MPL}$	$\hat{\theta}_{MCML}$
θ_1 (scale)	Selommes	1.62	1.99 (0.12)
	Kervidy	1.43	1.70 (0.11)
	BVD	2.32	2.11 (0.13)
θ_2 (areas)	Selommes	68.36	181.19 (66)
	Kervidy	23.73	79.19 (56)
	BVD	1361.46	2561.46 (679)
θ_3 (angles)	Selommes	0.95	2.23 (0.15)
	Kervidy	0.77	1.26 (0.12)
	BVD	1.72	2.21 (0.16)
θ_4 (elongation)	Selommes	0.48	0.28 (0.17)
	Kervidy	0.31	0.07 (0.15)
	BVD	-0.64	-0.63 (0.08)

Model checking



$t_1(T)$ $t_2(T)$ $t_3(T)$ $t_4(T)$

Distributions of energy statistics calculated over 500 model runs.

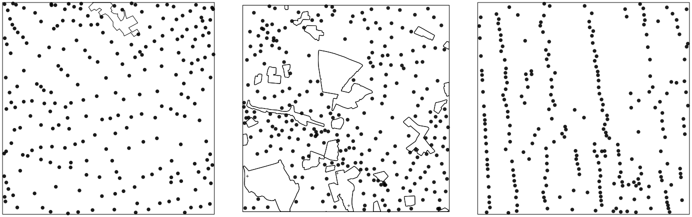


Observed and simulated landscape pattern.

Goodness-of-fit

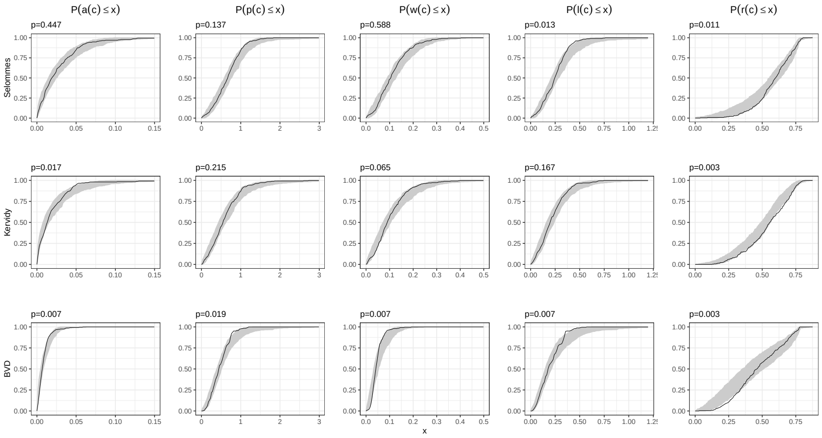
GET for tessellation and/or landscape metrics c.d.f.

- features of a typical cell: the area $a(c)$, the perimeter $p(c)$, the width $w(c)$, the length $l(c)$ of the smallest rectangle encompassing c and the isoperimetric quotient $r(c)$;
- spatial pattern of cell centroids : $G(r)$, $F(r)$, $L(r)$, $D(o)$ (density of the nearest neighbor vector orientation)

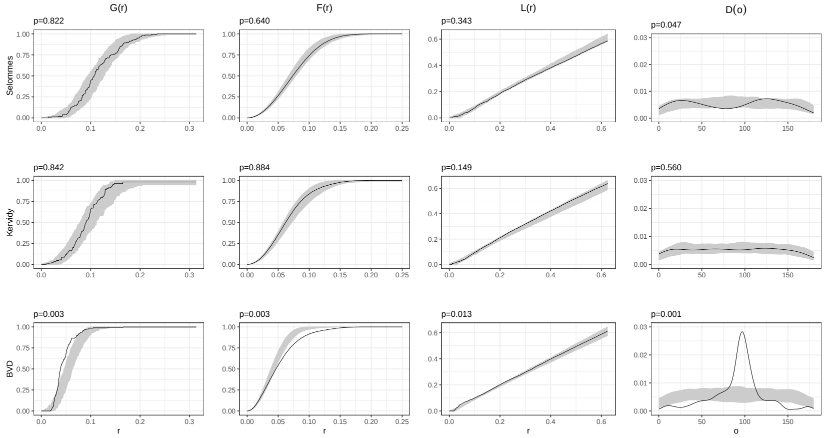


Cell centroid pattern

GET for the features of a typical cell



GET for the summary statistics of cell centroids



Concluding remarks

- Gibbs T-tessellation model: feature-driven model built on Poisson line skeleton;
- statistical methods: MPL estimation, adaptation of point process methods;
- application : characterization of agricultural landscapes... and other patterns that can be approximated by T-tessellations.

References

[1] K. Kiêu, K. Adamczyk-Chauvat, H. Monod, R.S. Stoica.
A completely random T-tessellation model and Gibbsian extensions.
Spatial Statistics, 6, 118-138 (2013).

[2] K. Kiêu and K. Adamczyk-Chauvat.
Pseudolikelihood inference for Gibbsian T-tessellations... and point processes.
Preprint HAL, hal-01234119, 2015.

[3] K. Adamczyk-Chauvat and K. Kiêu
http://kien-kieu.github.io/lite.

[4] K. Adamczyk-Chauvat, M. Kassa, J. Papaïx, K. Kiêu, R.S. Stoica.
Statistical inference for random T-tessellations models. Application to agricultural landscape modeling.
Ann Inst Stat Math, 76, 447–479 (2024).