A Spatial Poisson Point Process to classify coconut fields on Ikonos pansharpened images

Teina R.[a] and Béréziat D.[a] and Stoll B.[b]

[a]Université Pierre et Marie Curie, LIP6, 104 avenue du Président Kennedy, 75016 Paris, France;

[b]Université de la Polynésie Française, Laboratoire GePaSud, 98702 Faa'a, Tahiti - French Polynesia;

ABSTRACT

The goal of this study is to classify the coconut fields, observed on remote sensing images, according to their spatial distribution. For that purpose, we use a technique of point pattern analysis to characterize spatially a set of points. These points are obtained after a coconut trees segmentation process on Ikonos images. Coconuts' fields not following a Poisson Point Process are identified as maintained, otherwise other fields are characterized as wild. A spatial analysis is then used to establish locally the Poisson intensity and therefore to characterize the degree of wildness.

Keywords: Poisson Point Process, Ripley's function, quadrant count, nearest neighbor distances, coconut fields typology.

1. INTRODUCTION

French Polynesia government wants to improve their knowledge about the coconut (*Cocos Nucifera L.*) fields and specially to evaluate the coconut fields regeneration of old trees planted in the 1980's. The goal is to enhance the management of Coprah oil extraction as an alternative fuel. The French Polynesia Government acquired in 2003 some RGB pansharpened Ikonos images at 1m pixel resolution on ground for this purpose. Coconuts are detected using a segmentation process based on a marker-controlled watershed algorithm. For each crown, a center of mass, weighted by the values of the pixels in each RGB channels, is computed to estimate the canopies' center.[1] We finally obtain a set of point that we want to characterize as several type according to their spatial distribution.

In this study, we use Poisson Point Process (PPP) to characterize wild (or natural) coconut plantations. If the point pattern does not follow a Poisson Process Point, then the planting is considered as maintained. If the plantations are natural, and therefore identified as a Poisson Point Pattern, then we seek to segment these regions according to a criterion of homogeneity based on the density of the coconut trees. This density is directly related to the parameter of the Poisson distribution. If the density is not uniform, it has to be estimated locally.

We first focus on the Quadrant Count Method. It can be described simply by partitioning the data set into equal sized sub regions. Then some statistics are computed in each quadrant to determine if the observed data are clustered or not.

Another method is to simulate different outcomes of a Poisson Point Pattern, whose the intensity has been estimated from the observation, to compare these achievements with the observation using a suitable distance such as Ripley's K-function,[2,3] L-function, Pair Correlation function (g-function), and to deduce whether the point pattern is spatially uniform or not.

In this paper, we propose to segment regions by analysis of the histogram of the estimated density and according to some theoretical density values for this specific species of coconut trees. We then obtain three regions

Further author information: (Send correspondence to Teina R.)
Teina R.: E-mail: Raimana.Teina@lip6.fr, Telephone: +33 (0) 1 44 27 88 16
Béréziat D.: E-mail: Dominique.Bereziat@lip6.fr

that we can classify as less wild, maintained and wild or natural plantations with the result of a comparison to a Poisson process.

These point pattern analysis methods are applied to coconut field types on the island of Tikehau.

2. DATA AND STUDY AREAS

The study focuses on the atoll of Tikehau, which is located in the north-east of the French Polynesia Territory, specially on the main *motu* (a *motu* is one of the small islands constituting an atoll).
Previous works were done to segment the trees' crown using a marker-controlled watershed approach.[1] Thus, for each crown, a weighted center of mass is computed using informations from the three channel R,G and B of the Ikonos image. Each tree's center is then located by its Cartesian coordinates. Figure 1 shows a picture of the atoll of Tikehau and the main *motu* on which six study regions are located. These regions are representative of the several fields type available on the atoll.

Figure 1. The Atoll of Tikehau and the study areas on the main *motu*.

3. THEORETICAL BACKGROUND

3.1 Spatial Poisson Point Process

A Poisson Point Process is a set of points $X = \{X_i\}$ in a domain $S \subset \mathbb{R}^n$ verifying the following properties:

1. For any $B \subset S$, the number of points belonging to B, denoted $|X \cap B|$ in the remain of this paper, follows a Poisson law with parameter $\lambda\mathcal{A}(B)$, i.e.

$$P(|X \cap B| = n) = e^{-\lambda\mathcal{A}(B)} \frac{(\lambda\mathcal{A}(B))^n}{n!} \quad (1)$$

   where $\mathcal{A}(B)$ stands for the area of B.

2. If B_1 and B_2, two subsets of S, are disjoint, then $|X \cap B_1|$ and $|X \cap B_2|$ are independent.

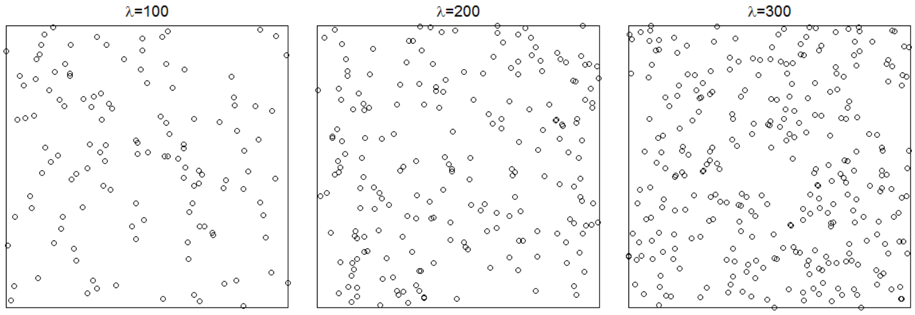
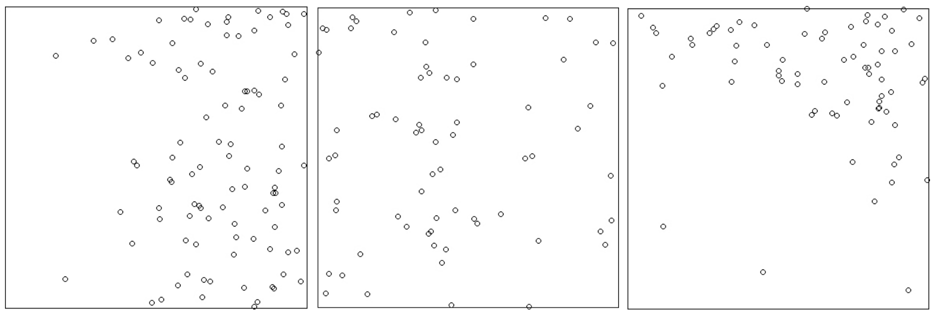


Figure 2. Uniform Poisson point process simulated in the unit square with $\lambda = 100, 200$ and 300.



(a) $\lambda(x, y) = e^{2+5x}$ (b) $\lambda(x, y) = e^{-2x}$ (c) $\lambda(x, y) = e^{2x+5y}$

Figure 3. Inhomogeneous Poisson point process simulated in the unit square using different intensity function $\lambda(x, y)$.

The parameter λ, which fully characterizes the Poisson process, is called intensity. Often, this parameter depends on a spatial coordinate, because the distribution may differ from a region to another one. Such a process is then called Spatial Poisson Point Process (SPPP). A SPPP can be used to characterize, for example, a random distribution of points in space, like a wild coconut field, by building a statistical test: a set of points coordinates are distributed as a Poisson process ($\mathcal{H}_0$) against a non Poisson process (which can characterize a domestic coconut field for instance).

One possible way to characterize a set of point pattern as SPPP is to estimate the intensity in equal area units using an appropriate statistical estimator, such as,[4] and to show how the intensity of SPPP varies. It may be constant ("uniform" or "homogeneous") or may vary depending on location ("spatial" or "inhomogeneous"). A Poisson point process, with a constant intensity λ, is also called Complete Spatial Randomness (CSR) in the literature.

Assuming the intensity of the PPP does not depend on space, the expectation of number of points in a given region B is given by:

$$\mathbb{E}(|X \cap B|) = \lambda \mathcal{A}(B)$$

because $|B|$ follows a Poisson distribution with intensity $\lambda \mathcal{A}(B)$. Let us suppose now the intensity depends on spatial location $\vec{x}$, it can be shown that:

$$\mathbb{E}(|X \cap B|) = \int_B \lambda(x) d\vec{x} \tag{2}$$

Figure 2 shows some simulated realizations of a Poisson Point Process with various values of intensity. Figure 3 shows some simulated realizations of an Spatial Poisson Point Process using various non constant intensity function depending on spatial position.

If the intensity is inhomogeneous, the intensity function can be estimated locally under the assumption of locally constancy. In the following subsection, we present such a method based on an image regular partitioning.

3.2 Quadrant Test

To identify a wild plantation, i.e a PPP, a χ^2 test denoted to test the hypothesis of distribution of two samples, is used on several subregions $B_1, B_2, ...B_m$ of equal area, called quadrants (or quadrats). Then, points belonging to each quadrant is counting: $n_j = n(X \cap B_j)\ for\ j = 1, ..., m$. Under the hypothesis of homogeneous Poisson Point Pattern (or CSR), the $\{n_j\}$ are independent and identically distributed.
For each subregions, we estimate the intensity function λ (under hypothesis of a CSR). To verify their spatial homogeneity, a set of samples pattern is generated using the estimated intensity function λ. Then, we check if the samples are significantly closed to the observation. To perform the comparison, the distance between two samples must be considered. Several distances can be used, such as:

1. the Pearson distance χ^2 goodness-of-fit test;[5]
2. the Clark Evans distance (CE):[6] it compares the mean distance between nearest neighbors to expected distance between nearest neighbors when the points are dispersed in a random pattern;
3. the index of dispersion (ID):[7–9] it is a quadrant based index computed on the frequency of events in each quadrant and the number of quadrants
4. the distance index of dispersion (I):[10] it is a distance based on point-to-event distances. $I = 2$ suggests a random pattern, $I < 2$ a uniform pattern and $I > 2$ a clustered pattern.

Quadrant based methods are very dependent of the space partition and it therefore does not show necessarily the spatial variation of the intensity function λ between two adjacent regions. Another method, described in the next Section allows us a visual interpretation of the spatial variation of the intensity function λ.

3.3 Kernel Estimation

A second approach is to estimate the intensity function λ using a kernel method such as.[11] This method computes a mass value for each point of the process in order to estimate λ, these values are then smoothed using a Gaussian convolution. This smoothing allows the visualization of λ disparities at various scale σ such as illustrated in Figure 4. Regions obtains by thresholding can provide good candidates for an image partition spatially homogeneous with respect to λ.

In order to analyze the interactions between neighbor points, a second order method should be used. This is described in the next Subsection.

3.4 Second order analysis

Ripley's K-function[2,3] is widely used as a second order descriptive statistic in two-dimensional point pattern analysis. This function analyzes all point-to-point distance, as opposing to a first order analysis like nearest distance. Ripley defines the K-function for a point process such as $\lambda K(r)$ is the expected number of other points of the process within a distance r. Under a CSR assumption, the expected number of points belonging to a circle centered on x with radius r is $\lambda\pi r^2$. The general form of the estimator of K is

$$\widehat{K}(r) = \frac{1}{\widehat{\lambda}^2 \mathcal{A}(S)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \mathbb{1}_{\{||x_i - x_j|| \leq r\}} e(x_i, x_j, r) \tag{3}$$

where $e(x_i, x_j, r)$ is an edge correction used for eliminating biais due to the edge effect. Several edge correction strategies are available in the literature, see[12] for a survey of edge effect corrections. S is the spatial domain and $\mathbb{1}$ is the indicator function defined as $\mathbb{1}_{\{||x_i - x_j|| \leq r\}} = \begin{cases} 1 \text{ if } ||x_i - x_j|| \leq r \\ 0 \text{ otherwise} \end{cases}$.
Under the assumption of CSR, we have $\widehat{K}(r) = \pi r^2$, otherwise values such as $\widehat{K}(r) < \pi r^2$ suggest clustering while values such as $\widehat{K}(r) > \pi r^2$ suggest a regular pattern. The interpretation of the K-function is often difficult. A commonly-used transformation of is the L-function[13] defined as:

$$L(r) = \sqrt{\frac{K(r)}{\pi}} \tag{4}$$

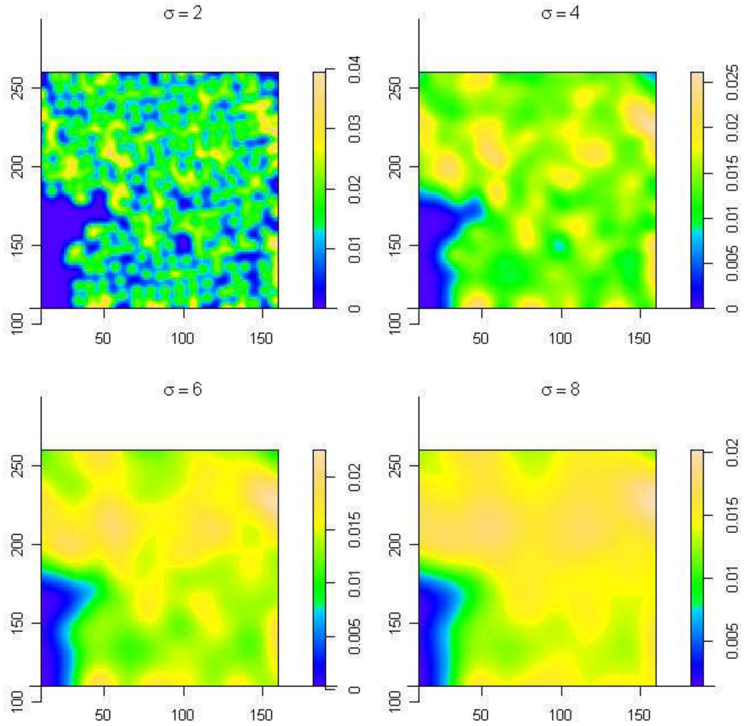


Figure 4. Various estimations of the intensity function $\lambda(u)$ using different values of the Gaussian kernel convolution.

Under the hypothesis of a CSR, we have $L_{pois}(r) = r$.

The Ripley's K-function is cumulative, so it becomes problematic to interpret the spatial interaction at larger distances. The pair correlation function $g(r)$[14,15] is better for this purpose and for a stationary point process the $g(r)$ function is defined as:

$$g(r) = \frac{K'(r)}{2\pi r} \tag{5}$$

where $K'(r)$ is the derivative of $K(r)$. With a CSR hypothesis, we have $g(r) = 1$. In the opposite case, values such as $g(r) < 1$ suggest inhibition (or regularity) between points and values such as $g(r) > 1$ suggest clustering (or aggregation).

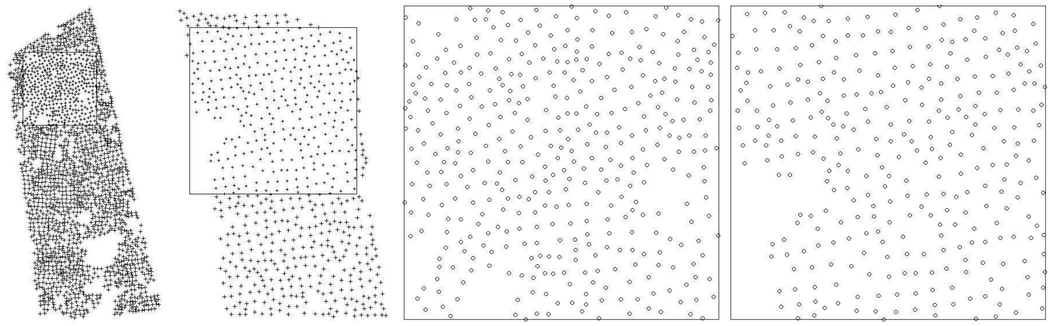
4. CHARACTERIZING COCONUT FIELDS

Our study focusses on two regions, denoted region #01 and #04 and presented in Figure 1. For each region, a subset of the point pattern is extracted within a square of 150m×150m (representing an area of 2.25 Ha). The minimum distance between each trees is $3.6m$ in Subregion #01 and $4m$ in Subregion #02. Then, applying various methods, we characterize the coconut plantations.

Figure 5 shows the full point pattern for regions #01 and #04, and for the corresponding subregions. Table 1 gives summary of the point pattern statistics like areas, number of trees and average intensities.

	Region #01	Subregion #01	Region #04	Subregion #04
Area (in Ha)	12.69	2.25	3.99	2.25
Number of points	1950	399	632	316
Average intensity	0.0154	0.0177	0.0158	0.014

Table 1. Some statistics from studied regions and subregions. The average intensity is the number of trees per square meter.

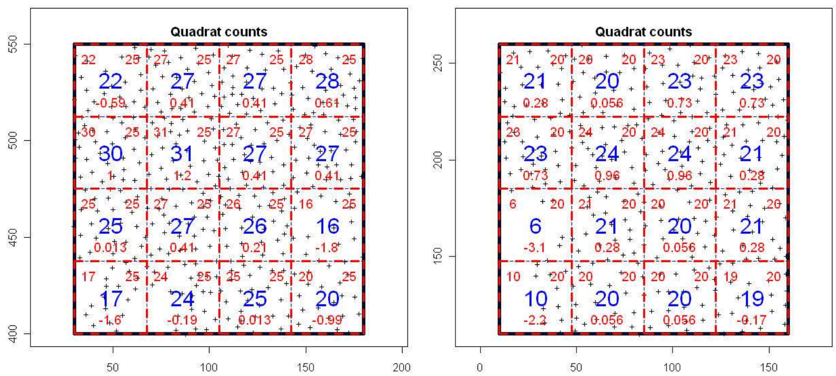


(a) Region #01 (b) Region #04 (c) Subregion #01 (d) Subregion #04

Figure 5. Studied regions and subregions.

The first step of the spatial analysis is to estimate the intensity of PPP. We use the methods described in Sections 3.2, 3.3 and 3.4.

First method is quadrant count. Subregions are divided into 4×4 equals areas and distances cited in Subsection 3.2 are computed. Figure 6 shows the partitioning in equals areas for each Subregions. Table 2 shows values of the various distances computed to determine if the point pattern reject the hypothesis of a CSR. Finally, this technique gives an indication on the intensity function $\lambda(\vec{x})$ in each areas of the partitioning, bit it is highly dependent of the number of quadrants used in the space partition.



(a) Subregion #01: $p - value = 0.76$ (b) Subregion #04: $p - value = 0.25$

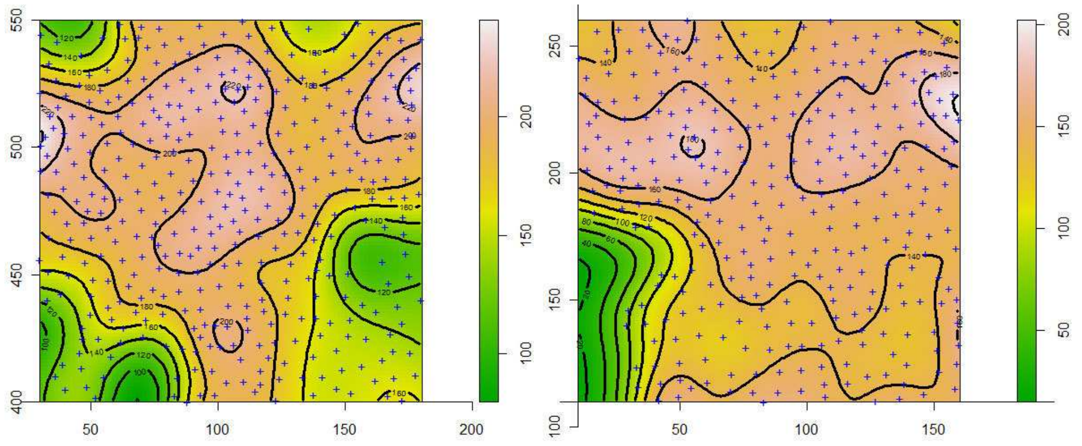
Figure 6. χ^2 test for the two subregions.

	Subregion #01			Subregion #04		
	CE	ID	I	CE	ID	I
Index Value	1.30	3.81	2.53	1.27	5.58	5.95
Test Statistics	8.76	3.81	2.53	7.26	5.58	25.20
Reject Ho	yes	no	yes	yes	yes	yes

Table 2. Determination of Indices of Dispersion: Clark Evans (CE), Index of Dispersion (ID), Distance Index of Dispersion (I).

We can see in Figure 7, the result of the kernel estimation with $\sigma = 12m$. It shows spatial variation of the intensity indicating a spatial PPP.

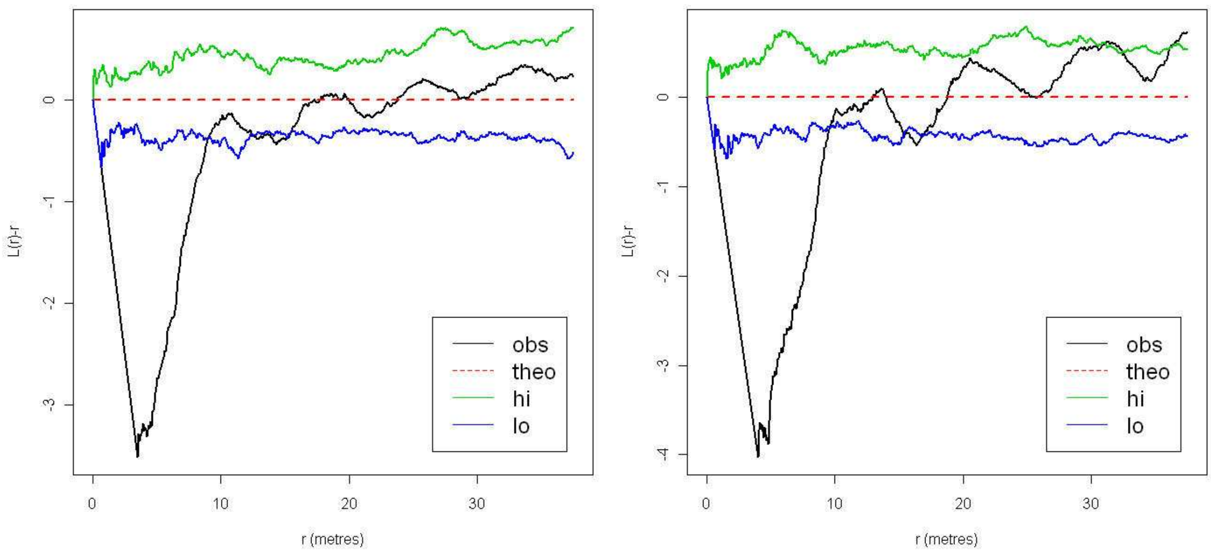
To extend our analysis, we compute the L-function as defined in equation (4) for the two Subregions. Results are shown in Figure 8 as $L(r) - r$. Under the hypothesis of CSR, $L(r) - r = 0$. If the point pattern exhibits aggregation, then $L(r) - r$ tends to be positive, and if it exhibits regularity, then $L(r) - r$ tends to be negative. Points in both Subregions #01 and #04 show a significant regularity for $r < 9.5m$ (the L-function computed with



(a) Kernel estimation on Subregion #01 (b) Kernel estimation on Subregion #04

Figure 7. Kernel estimation of the two subregions #01 and #04. Densities are plotted as number of trees per hectare.

the observation is under the lower confidence bound) with a minimum distance between each tree about $3.5m$ for the Subregion #01 and $4m$ for the Subregion #04. There is also a regularity observed for $r \in]11m, 15.5m]$ for Subregion #01. Finally, points of the Subregion #01 don't show deviation from CSR for $r \in [9m, 11m[$ and for $r > 15.5m$. Points in Subregion #04 exhibit a small clustering for $r \in [31m, 32m]$. For $r \in]9.5m, 31m]$, there is no deviation from CSR (the curve of the observation is between the lower and upper bounds of confidence) at larger scales (larger distances r).

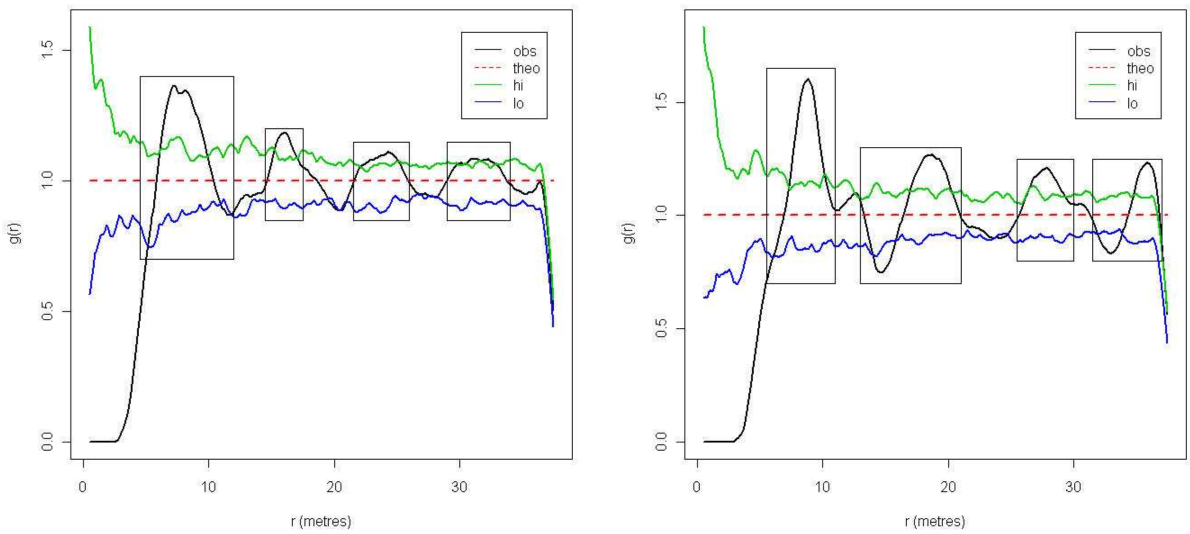


(a) L-function on Subregion #01. (b) L-function on Subregion #04.

Figure 8. L-function plotted as $L(r) - r$ with the upper (green line) and lower (blue blue) 95% confidence bounds, the theoretical curve (red dashed line) and the observation (black line).

The pair correlation function g for a point process defined as in equation (5) is also computed for the two subregions. Figure 9 displays such a function with lower (green line) and upper (blue line) confidence bounds plotted. Theoretical curves (red dashed line) and observed curves (black line) are also plotted. Under the hypothesis of CSR, $g(r) = 1$, values of $g(t) < 1$ suggest a regularity between points and values $g(r) > 1$ suggest clustering.

We observe that trees in Subregion #01 are clustered around distance values of $r = 6-10, 15-17, 23, 24-25m$, and those in the Subregion #04 are aggregated around distance values of $r = 7-10.5, 17-21, 26-29m$. Comparing Figures 8 and 9, we see that L-function and pair correlation function detect the same patterns but with different distance values.



(a) g function on Subregion #01.

(b) g function on Subregion #04.

Figure 9. Pair correlation functions (g function) plotted with the upper (green line) and lower (blue line) 95% confidence bounds, the theoretical curve (red dashed line) and the observation (black line).

For this species of coconut trees (*Grands Locaux*), spacing between each tree should be $8.5m$, which implies a density of 140-150 trees per hectare.[16] According to these informations, we examine the density histogram and we define three classes of coconut plantations taking account the theoretical density and the standard deviation of the result of the kernel estimation (σ_{KD}), described as:

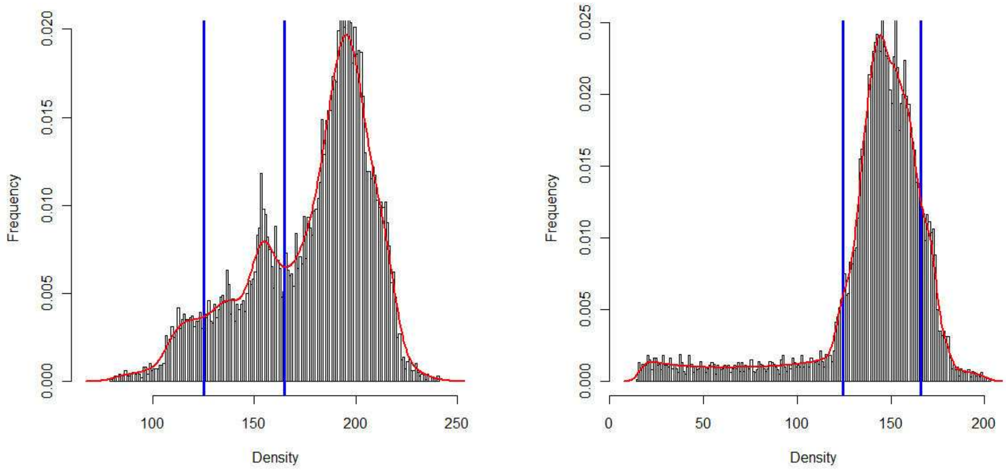
$$\begin{cases} values < 140 - \sigma_{KD} & \text{less wildness region: class 1} \\ 140 - \sigma_{KD} \leq values \leq 150 + \sigma_{KD} & \text{maintained fields: class 2} \\ values > 150 + \sigma_{KD} & \text{wild or natural fields: class 3} \end{cases} \tag{6}$$

Figure 10 shows the histogram of the estimated density for Subregions #01 and #04. The threshold values are plotted in the blue vertical line. Low and High threshold values for Subregion #01 are 125 trees/hectare and 165 trees/hectare. Those for Subregion #04 are 124 trees/hectare and 166 trees/hectare.

Figure 11 shows the result of the histogram threshold in three regions with over plotted trees.

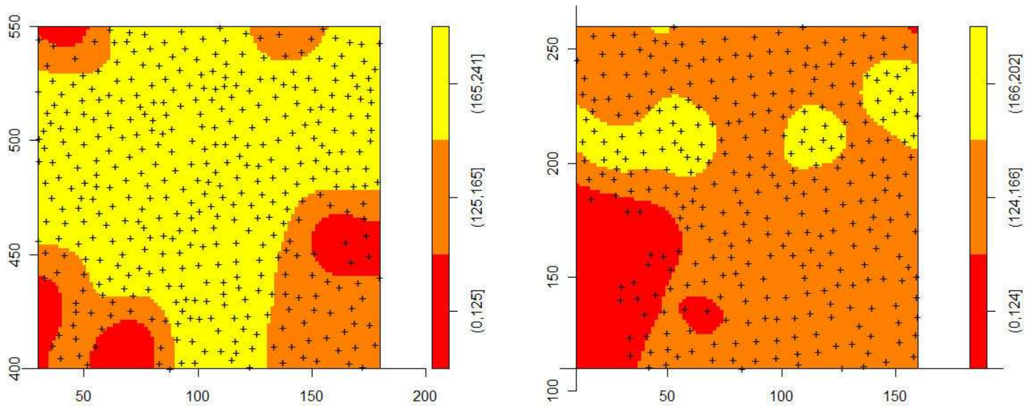
Table 3 lists statistics computed for the three classes, these classes being obtained by a histogram thresholding using values described in (6). Even if the area and the number of trees in class 3 (representing then the wild or natural plantation) of each subregion (Subregions #01 and #04) are different, the average intensity is the same. We observe the same conclusion for class 2 (classified as maintained plantations). Addressing class 1 in Subregion #01, there is only 7 trees detected inside. Even if the average intensity is closed to the intensity of class 1 in Subregion #04, we can not conclude for this region because the number of trees is not relevant. But it also indicates that maintained and wild plantations are mainly the two types of plantation in the Subregion #01.

In order to take a decision on the spatial properties of each segmented region, we use a Q-Qplot of residual. A Q-Qplot of residual is a diagnostic tool for checking some assumptions about the distribution. For example, a Q-Qplot of the (smoothed) residuals from a fitted model is a useful way to check the inter-point interaction.[17] Figure 12 shows the Q-Qplot of smoothed residuals for the three segmented regions on Subregions #04. For each plot, intervals with 95% of confidence are plotted in red dashed line and the model to fit in black dotted line.



(a) Kernel density histogram on Sample 01 (b) Kernel density histogram on Sample 04

Figure 10. The histogram of the kernel density estimation of Subregions #01 and #04. The vertical blue line are the threshold values.



(a) Thresholded kernel density of Subregion #01 (b) Thresholded kernel density of Subregion #04

Figure 11. Thresholded kernel density of Subregions #01 and #04.

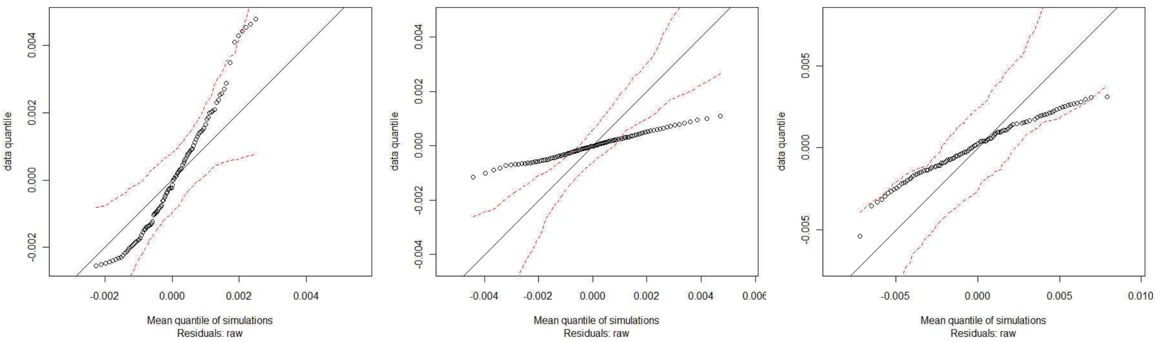
	Subregion #01			Subregion #04		
	Area 1	Area 2	Area 3	Area 1	Area 2	Area 3
number of trees	7	78	314	20	243	53
Average intensity (# trees/m^2)	0.00414	0.0154	0.02	0.00631	0.0148	0.0185
area (in m^2)	1692	5080.5	15727.5	3170.5	16465.5	2864.25

Table 3. Resume of each segmented regions for both Area 01 and Area 04.

Sub-figures 12(a) and 12(c) suggest that the uniform Poisson model is appropriate for points for classes 1 and 3 and we conclude that the coconuts' fields in the corresponding regions are wild. Class 3 has a higher degree of wildness than class 1 because its average intensity within class 3 is higher. Figure 12(b) indicates that the Poisson model is inappropriate for class 2 and can be characterized as a maintained plantation.

5. CONCLUSION

In this study, we classified a coconut trees plantation using a Spatial Poisson Point Pattern approach. Methods used allow us to characterize the spatial distribution of coconuts. The last part of this work consisted to segment the regions into three classes using a histogram analysis and a threshold on the result of kernel density estimation according to theoretical values of density of trees per hectare for a specific species of coconut trees. The analysis of these segmented regions gives information of the spatial distribution of trees.



(a) Q-Qplot for Area 1 (b) Q-Qplot for Area 2 (c) Q-Qplot for Area 3

Figure 12. Q-Qplot for the segmented regions of Subregion #04.

In a future work, the threshold values should be estimated using histogram approximation with a mixture of Gaussian.

REFERENCES

[1] Teina, R., Béréziat, D., Stoll, B., and Chabrier, S., "Toward a global Tuamotu archipelago coconut trees sensing using high resolution optical data," IGARSS08 (July 2008).

[2] Ripley, B. D., "Modelling spatial patterns," *Journal of the Royal Statistical Society, Series B* **39**, 172–212 (1977).

[3] Ripley, B. D., [*Spatial Statistics*], Wiley, New York, NY (1981).

[4] Diggle, P. J., [*Statistical Analysis of Spatial Point Patterns*], Academic Press (1983).

[5] Pearson, K., "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonable be supposed to have arisen from random sampling," *Philosophical Magazine* **50**, 157–175 (1900).

[6] Clark, P. J. and Evans, F. C., "Distance to nearest neighbor as a measure of spatial relationships in populations," *Ecology* **35** (1954).

[7] Ludwig, J. A. and Reynolds, J. F., [*Statistical ecology: a primer on methods & computing*], John Wiley & Sons, Inc., New York, NY, USA (1988).

[8] Pielou, E., [*Mathematical Ecology*], John Wikey & Sons, New York (1977).

[9] Jayaraman, K., *A statistical manual for forestry research.* Forestry Statistics and Data Collection - DCA/MISC/01, Bangkok, Thaïland (Oct 2000).

[10] Johnson, R. B. and Zimmer, W. J., "A more powerful test for dispersion using distance measurements," *Ecology* **66**(5), 1669–1675 (1985).

[11] Diggle, P., "A kernel method for smoothing point process data," *Applied Statistics* **34**(2), 138–147 (1985).

[12] Baddeley, A. J., "Spatial sampling and censoring," in [*Stochastic Geometry: Likelihood and Computation, chapter 2*], 37–78, Chapman and Hall (1998).

[13] BESAG, J., "Contribution to the discussion of dr ripley's paper," *Journal of the Royal Statistical Society, Series B* **39**, 193–195 (1977).

[14] Stoyan, D. and Stoyan, H., "Random shapes and point fields, methods of geometrical statistics," (1994).

[15] Stoyan, D. and Stoyan, H., "Estimating pair correlation functions of planar cluster processes," *Biometrical Journal* **38**(3), 259–271 (1996).

[16] Chantry, G.; Cabannes, Y. W. V., [*Le cocotier: production et mise en oeuvre dans l'habitat*], Ministère de la Coopération, Paris (France) (1983). ISBN: 2-11-084847-2.

[17] Baddeley, A., Turner, R., Moller, J., and Hazelton, M., "Residual analysis for spatial point processes (with discussion)," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 617–666 (November 2005).