

SPATIAL DISTRIBUTION OF OVERSTOREY TREES ANALYZED BY REPLICATED POINT PATTERN METHOD IN R

Bùi Mạnh Hưng¹, Võ Đại Hải²

¹ Trường Đại học Lâm nghiệp

² Viện Khoa học Lâm nghiệp Việt Nam

SUMMARY

In the forest structure, the spatial distribution of forest trees plays a very important role. The importance of the spatial distribution is to influence the sampling design of forest inventory, timber production, and the need for silvicultural treatments in a stand. The analysis results show that the difference between the two types is significant. Spatial distribution pattern of trees on the ground is more homogeneous in the secondary forest, whereas it is very diverse in the old-growth forest. The G-test and the pair correlation function results show that the spatial distribution of forest trees is random in the regenerating forest. Conversely, the tree distributes more regularly in the old-growth forest. Up to 80% of plots in the old forest show this trend. The Envelope function demonstrates that the spatial distribution variation in young forests is much lower than in the primary forest. In addition, the plants with the same diameter also distribute randomly in stage IIb and regularly in stage IV. The replicated point pattern analysis is very new. It is very useful for small investigated plots with repetitions. It should be applied in Vietnam with the support of R.

Keywords: Forestry data, forest structure, replicated point pattern, R language, spatial distribution

Phân bố không gian của cây tầng cao phân tích bởi phương pháp lặp điểm trong R

Trong phân tích cấu trúc rừng, thì phân bố không gian đóng một vai trò quan trọng. Tầm quan trọng của nó thể hiện ở việc phân bố không gian sẽ ảnh hưởng tới công tác thiết kế rút mẫu trong điều tra rừng, năng suất gỗ và các biện pháp lâm sinh tác động vào rừng. Kết quả phân tích đã cho thấy rằng phân bố không gian rất khác biệt giữa hai trạng thái rừng. Phân bố không gian khá đồng nhất giữa các ô tại rừng thứ sinh, và biến động mạnh ở rừng già. Kết quả hàm G và hàm tương quan cặp cho thấy phân bố là ngẫu nhiên tại rừng IIb. Ngược lại, phân bố có xu hướng đều hơn ở rừng già. Có tới 80% số ô ở rừng già thể hiện xu hướng này. Kết quả của hàm Envelop đã chứng minh rằng biến động phân bố ở rừng thứ sinh là rất thấp so với rừng nguyên sinh. Hơn nữa, các cây rừng có cùng đường kính thường phân bố một cách ngẫu nhiên ở rừng thứ sinh và tương đối đều hơn ở rừng già. Phương pháp phân tích đặc điểm phân bố có lặp là phương pháp rất mới. Nó rất phù hợp trường hợp mà các ô mẫu điều tra nhỏ, có lặp. Và phương pháp này thực sự nên được áp dụng trong nghiên cứu ở Việt Nam với sự hỗ trợ bởi ngôn ngữ R.

Từ khóa: Cấu trúc rừng, ngôn ngữ R, phân tích điểm lặp, phân bố không gian, số liệu lâm nghiệp

I. INTRODUCTION

Forest structure plays an important role in forestry research. Forest structure greatly impacts the habitat of fauna and flora species. Complex forest structures diversify microclimates, niches and habitats for maintaining the majority of terrestrial biodiversity (Pan *et al.*, 2013). Forest structure is the key to understanding and determining ecosystem functions (Spies, 1998; Valbuena, 2015). Understanding forest structure will unlock an understanding of the history, function and future of a forest ecosystem (Spies, 1998, Hai, 2014), assist in forest management planning (Valbuena, 2015), propose silvicultural treatments and enable sustainable use of forest resources (Sau, 1996; Gadow *et al.*, 2011).

In the forest structure, the spatial distribution of forest trees plays a very important role. Spatial distribution of trees is influenced by many ecological processes. Therefore, it reflects the degree of competition among forest trees, density, size distribution, the survival and mortality of forest trees (Wehenkel *et al.*, 2015). The importance of the spatial distribution is to influence the sampling design of forest inventory, timber production, and the need for silvicultural treatments in a stand (Stamatellos and Panourgias, 2005).

Currently, there are many methods and statistical functions to analyze the spatial distribution of forest trees. These methods are based on data coordinates of the tree on the ground, and then distribution features will be determined and analyzed. Methods used include Ripley's K-function, L-function, the nearest-neighbor distance statistic, the pair-correlation function, replicated point pattern

analysis (Fangliang *et al.*, 1997; Condit *et al.*, 2000 and Luo *et al.*, 2009). The replicated point pattern analysis is a new solution, especially for small investigated plots with repetitions. It is now rarely used in ecology, especially in tropical forests and in Vietnam (Ramón *et al.*, 2016).

To analyze spatial distribution patterns of forest trees, there are a few programs that can do that. In which R is essential to analyze and understand characteristics of the spatial distribution, especially the spatstat package. From recent decades, R is showing its benefits and strong points in data analysis. R provides a wide variety of statistical such as: linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, LMM, point pattern analysis and so on and graphical techniques (R-project, 2016).

The above explanation has shown the urgency and necessity for applying replicated point pattern analysis, with the support of R for analyzing spatial distribution of tree species. This paper will present how to implement replicated point pattern method by using R command and analyze the real distribution of tree species in Kon Ka Kinh National Park, applied this method.

II. METHODOLOGY

2.1. Data collection method

Data were collected from 10 plots in Kon Ka Kinh National Park, Gia Lai province, Central Highlands of Vietnam. The location of the park is presented in the following figure. In 10 plots, there are 5 plots secondary forests (Type IIb) and 5 plots of old-growth forests (Type IV).

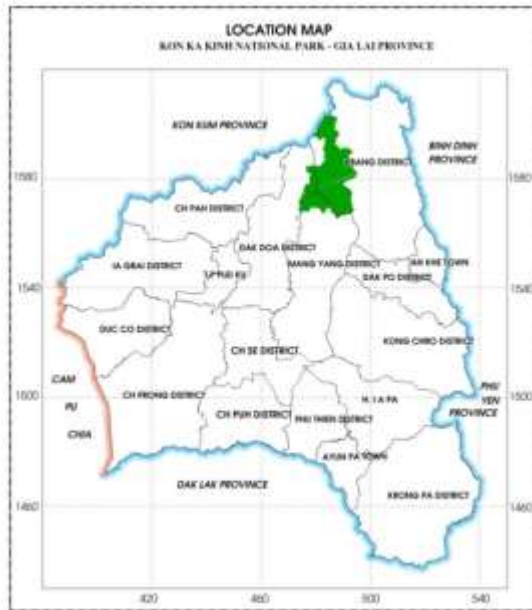


Fig. 01. Kon Ka Kinh National Park location (Hung, 2016)

Stratified random sampling was applied to select plot locations, because the forest resource is not homogeneous (Shiver and Borders, 1996).

When the plot was set up, if a tree had more than or equal to half of the diameter trunk in the plot, it was counted and measured. The diameter at breast height (DBH) and total height of all trees with boles larger than 6cm were measured and recorded by using a caliper and a Blume Leiss. X and Y coordinates of all trees were also recorded. They are the perpendicular distances from the tree center point to the sub-plot sides.

2.2. Data analysis method

2.2.1. Applications

In the case of this research, the coordinates of tree locations in the first were used to analyze point patterns. The object is the overstorey trees. There are some first steps that should be followed (Baddeley, 2008).

- Observation window creation

A rectangular window was generated by using “Owin” command in R. It will provide

information about the region in which the pattern was observed (Baddeley, 2008). To make the window, the following command was used:

```
W <- owin(c(0,50), c(0,10))
```

The window was a rectangle. The length of two dimensions were 10 and 50 units. In this study the unit was meters.

- Point pattern object creation

Point pattern object contained both points and observation window. To create the point pattern object, the ppp command in R was used. For example:

```
# Without a mark.
X1 <- ppp(x, y, c(0, 50), c(0, 10))
# With the mark.
X1 <- ppp(x, y, c(0, 50), c(0, 10),
marks = DBH)
```

- Unit selection

In this study, the unit used to determine the tree location was meters, so the following command was used to notify R program understanding used unit of X and Y coordinates (Baddeley, 2008).

```
unitname(X1) = "metres"
```

2.2.2. Tree density analysis

- Density graph

The number of trees per area unit was calculated based on the density function. A density graph with tree positions was generated by “density” and some other commands in R. That will compute a two-dimensional kernel estimate (Baddeley, 2008).

```
par (mar = c(0, 0, 0, 0))
plot (H$PPP, main = paste ("Tree
distribution"), nrow = 10)
plot(H, quote (plot(density (PPP),
main = "")), main = paste ("Density
graphs"), nrow = 10)
```

- Density difference testing

To test the density difference, the mppm function was called in the spatstat package. A Poisson point process was used (Baddeley, 2008).

```
fit=mppm(PPP ~ Type, data=H, Poisson())
summary(fit) # To obtain information.
```

2.2.3. Testing for randomness

- Nearest-neighbor G function

In term of mathematics, Baddeley (2008) showed that for the Poisson process, the nearest-neighbor distance distribution function is:

$$G_{pois}(r) = 1 - \exp(-\lambda\pi r^2) \quad (1)$$

Where: λ = intensity and r is the distance.

G-test was applied and if the G(r) is greater than $G_{pois}(r)$, so the nearest-neighbor distances are shorter than for the Poisson process. Therefore, the distribution has a clustering pattern. In contrast, if G(r) is smaller than $G_{pois}(r)$, the distribution is regular. Some commands were implemented to generate the K-test graphs.

Commands for one-point patter (X1):

```
plot(Gest(X1, legend = FALSE,
        ylim = c(0, 2), xlim = c(0.5, 2.5)))
```

- The pair correlation function

The pair correlation function will calculate all distances between any two points. It will use the random pattern as a reference. Illian et al. (2008) also indicated that the pair correlation function is proportional to the derivative of K(r) with respect to r. In the 2 - dimensional case, it is:

$$g(r) = \frac{K'(r)}{2\pi r} \quad \text{for } r \geq 0 \quad (2)$$

When r reaches infinity, then the limit of g(r) will be equal to 1, so in the Poisson process

case, g(r) is 1. The distribution will be clustering if g(r) is greater than or equal to 1. On the contrary, the distribution has regular pattern, if g(r) is smaller than or equal to 1 (Illian et al., 2008).

To carry out the pair correlation function in the spatstat package, the following commands were applied. The “pcf” is the function in the package for the pair correlation analysis.

Commands for one-point pattern (X1):

```
plot(pcf(X1, correction=c("Ripley")),
      legend = FALSE,
      ylim = c(0, 2), xlim = c(0.5, 2.5))
```

- The mark correlation function

The mark correlation function ($k_{mm}(r)$) of a marked point process is a tool to measure the dependence between the marks of two points of the process a distance r apart (Baddeley *et al.*, 2015). It includes summary statistics used for quantitatively marked patterns when the mark is quantitative (Wiegand and Moloney, 2014). The mark was tree diameter (DBH). The following commands were called in R to run markcorr.

Commands for one-point pattern (X1):

```
plot(markcorr(X1,
              correction=c("Ripley")), main="",
      legend = FALSE,
      ylim = c(0, 2), xlim = c(0.5, 2.5))
```

2.2.4. Comparing point pattern variation

Envelope was implemented to understand the variation. The intention of generating envelopes is to compare empirical estimates and simulative estimates in the same window. The statistics of simulations lead to an estimate of S(r), where $S_{min}(r)$ and $S_{max}(r)$ are the most extreme values of the statistics of the simulation runs. As these extreme values can be determined for any distance r, an environment of extremes, i.e. the “envelope” is generated. After that, estimates of S(r) and extreme values are plotted (Illian *et al.*, 2008).

- For the pair correlation

The “envelope” function was used in the package in order to know the variation in term of the pair correlation. To do that, the following commands were run.

```
window <- owin(c(0,50), c(0,10))
```

After that, a list of point patterns in the dataset was generated by using the following command.

```
SList <- list(X1, X2, X3, X4, X5)
```

Finally, the envelope function was used in the package to generate an acceptance interval which shows a range of variation. The commands below are examples for Type IIb. Similar commands were conducted for Type IV.

```
E <- envelope(X, pcf, nsim=5,
simulate=SList)

plot(E, main = "Envelope graph for Type
IIb", xlim = c(0.5, 2.5), ylim = c(0.0, 2.5),

legend = FALSE, lwd=2, col="Red")
```

- For the Markcorr

The following commands were applied.

```
window <- owin(c(0,50), c(0,10))
```

And then a list of point patterns in the dataset was generated. Here is an example for Type IIb.

```
SList <- list(X1, X2, X3, X4, X5)
```

The final step was to run envelope in order to depict the variation of the mark correlation function. For instance:

```
E <- envelope(X, markcorr, nsim=5,
simulate=SList)

plot(E, main = "Envelope graph for
Type IIb", xlim = c(0.5, 2.5),
ylim = c(0.0, 2.5),

legend = FALSE, lwd=2, col="Red")
```

III. RESULT AND DISCUSSION**3.1. Density testing results**

Coordinates of trees were used to analyze and understand spatial distribution on the ground of the plot. The received distribution and density graph with the diameter mark is shown in the following charts (Fig.02). These charts show that the density of trees is significantly higher in the young than in old stands which are shown in the following table. Comparison results by using mppm command in R will better prove this, because the p-value is less than 0.05. This test was applied for replicated point patterns, but without the mark because spatstat does not provide this test for point patterns with the mark. Comparison of stand density of two forest types by using mppm:

Tab. 01. Density comparison between two type

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.07147	0.07834	-26.443	< 2e-16 ***
TypeType IV	-0.51613	0.12814	-4.028	5.67e-05 ***

Signif. codes:				
0	'****'	0.001	'***'	0.01
	'**'	0.05	'.'	0.1
	' '	' '	' '	1

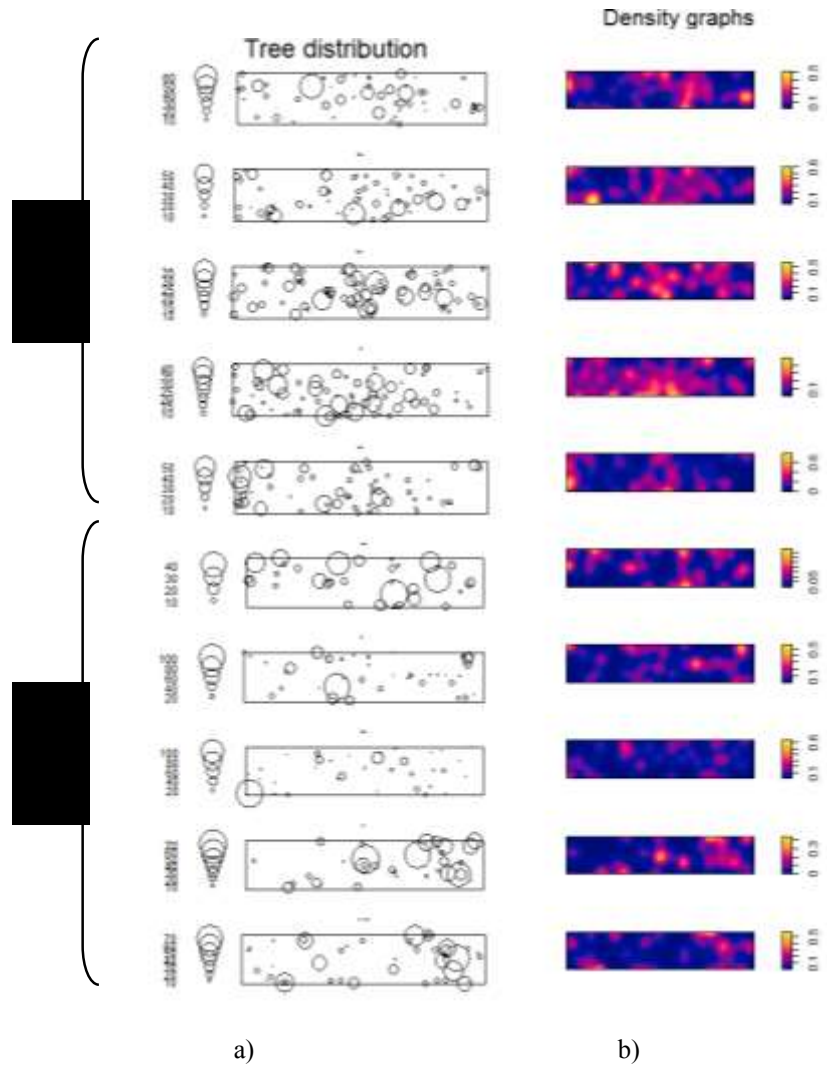


Fig.02. Tree distribution (a) and tree density (b) of two types

3.2. Randomness checking results

The next figures illustrate the results of G-function and pair correlation function to

understand more about the spatial distribution of the tree.

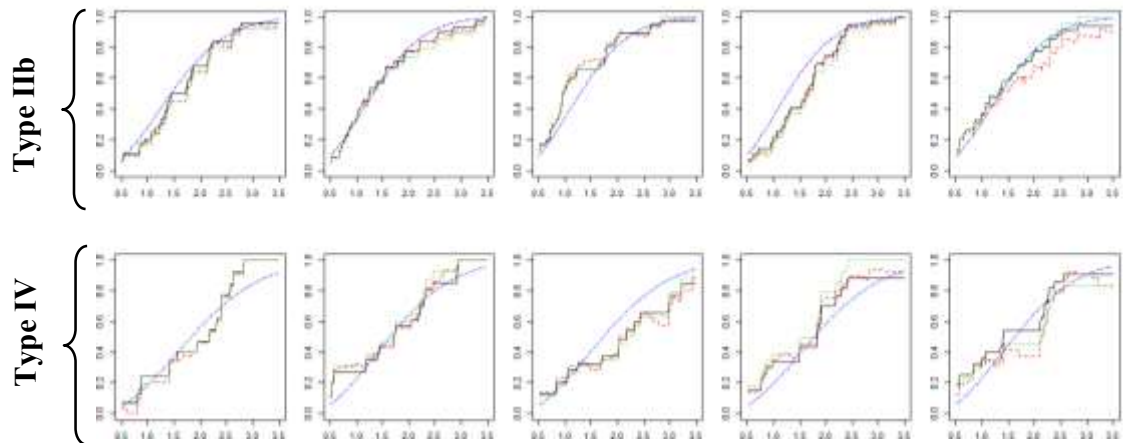


Fig. 03. G-function graphs

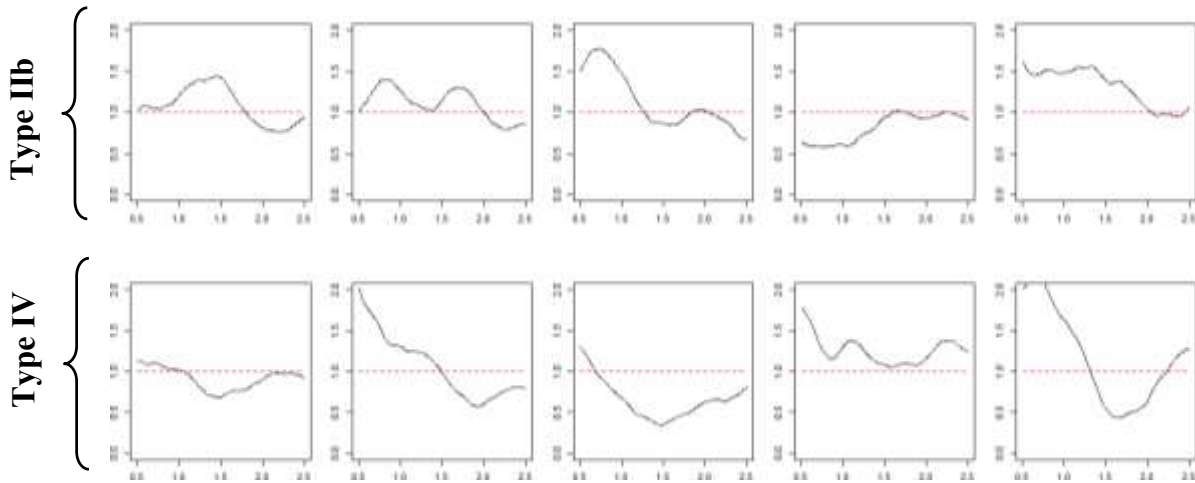


Fig. 04. Pair correlation function results

The results from G-function and the pair correlation function indicate the following conclusions. For the regenerating forest, trees distribute mainly randomly and clusteringly on the ground. Distribution with the scale smaller than 1.5, 80% of plots tend to be random and cluster. This type of distribution is typical in tropical forests, especially young forests. This has also been concluded and presented in previous studies such as Fangliang *et al.* (1997), Condit *et al.* (2000), Luo *et al.* (2009), Rejou-Méchain (2011) and Hung (2013).

Another difference found in the old-growth forest is that the spatial distribution tends to resemble a regular one. This is clearly expressed in the pair correlation function

results. Up to 60% of old forest plots have regular distributions, while 40% of the plots are randomly distributed. Therefore, the tree spatial distribution tends to shift from random to regular distributions when the forest gets mature. This trend has also been demonstrated in both Vietnam and other countries through the study of Christensen (1977), Sau (1996) and Fangliang *et al.* (1997). One explanation for this is the competition of the big trees. Big trees will push other individuals out further and generate a more evenly distributed form.

The mark correlation function is also applied to distribution of trees with same mark values. The results of these analyses are indicated in following graphs.

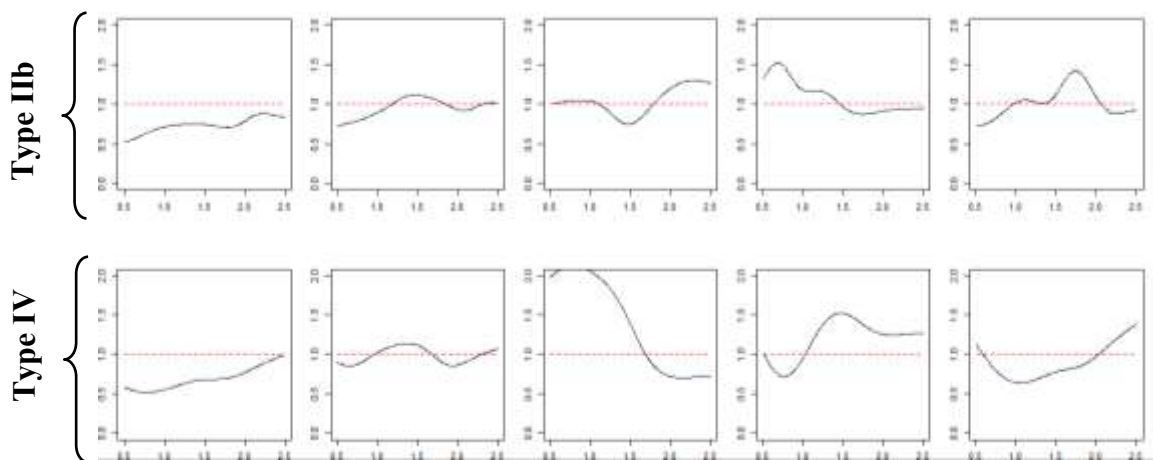


Fig. 05. Mark correlation function results

The mark correlation function results show a difference between the two forest types. For the regenerating forest, the trees are more randomly distributed in terms of diameter. In other words, the same size trees are distributed randomly on the ground. In contrast to the jungle, the forest trees are more regularly distributed also with regard to the diameter. This means that trees with different diameters may be neighbors while plants with the same size are spread evenly over plots. Clustering of trees with similar size is not common. This is the result of competition and self-thinning.

All above results indicate that the spatial distribution pattern of trees on the ground is more homogeneous in the secondary forest, whereas it is very diverse in the old-growth forest.

3.3. Variation difference between two types

Point pattern datasets of each forest type were listed and used to generate variation charts. Envelope graphs were created for the pair correlation function (pcf in R) and mark correlation function (markcorr in R). The results are presented in next figures. For both functions, the gray area of the young forest is much smaller and narrower than the old forest.

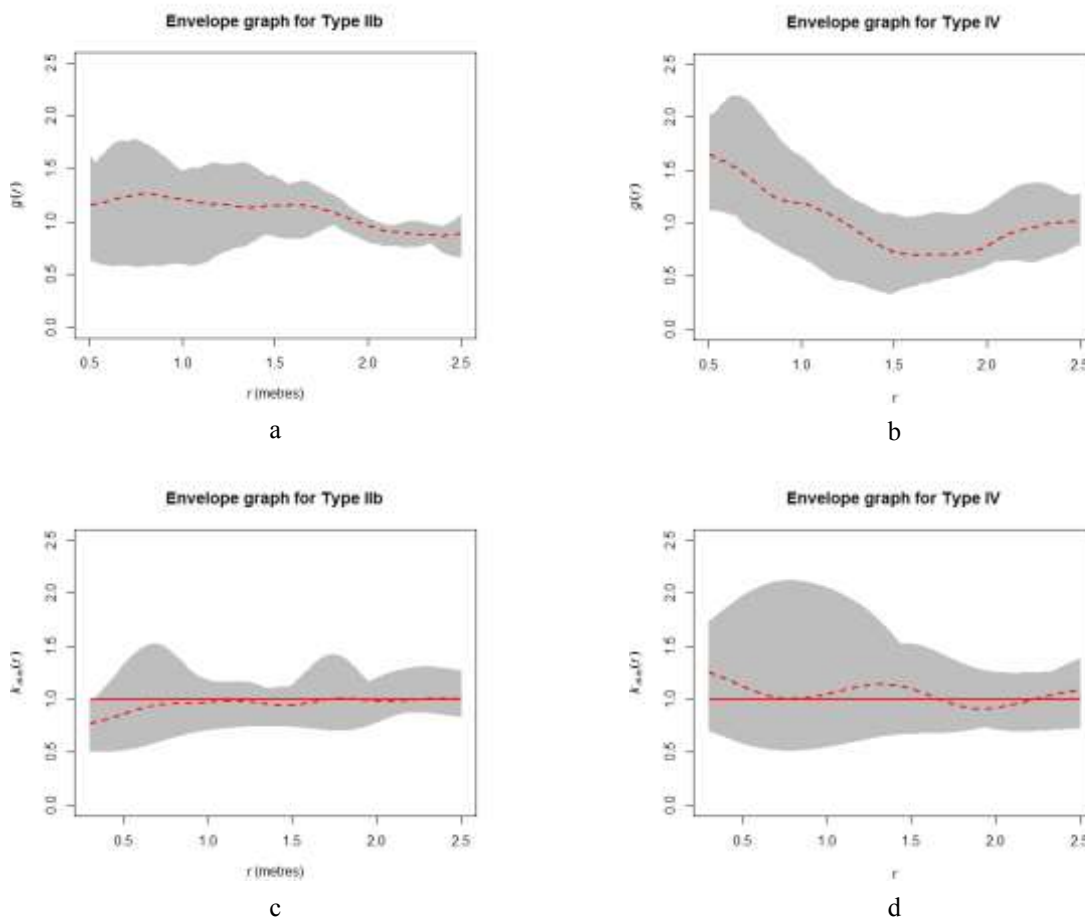


Fig.06. Envelope graphs: a and b for pcf; c and d for markcorr

The results of the envelope function express that the variation is much greater than the young forest. This is a consequence of long-term development of forest trees. During the growth process, the survival of trees is affected

by many environmental factors such as light, nutrients, competition and so on (Mabberley, 1992; Fangliang *et al.*, 1997; Hai, 2009, Wagner *et al.*, 2011, Hung, 2011). As a result, the distribution of trees is altered dramatically

and unpredictably. Another reason leading to the higher variation and more regular distribution in the jungle is an observed area. The area of investigation in each plot is $10 \times 50\text{m}$. This area is small compared to many other studies. Low density combined with the growing competition are a cause of the trend, because trees tends to keep others away, leading to the regular distribution. The heterogeneity of tree distribution patterns on the grounds and investigated plot area also influence the results.

IV. CONCLUSION

The analysis results show that the difference between the two types is significant. Spatial distribution pattern of trees on the ground is more homogeneous in the secondary forest, whereas it is very diverse in the old-growth forest.

Regarding the spatial distribution of overlayer trees, the G-test and the pair correlation function results show that the spatial distribution of forest trees is random in the regenerating forest. Conversely, the tree distributes more regularly in the old-growth forest. Up to 80% of plots in the old forest show this trend. The spatial point pattern difference is significant. The Envelope

function demonstrates that the spatial distribution variation in young forests is much lower than in the primary forest. In addition, the plants with the same diameter also distribute randomly in stage IIb and regularly in stage IV. This is the conclusion drawn from the mark correlation function.

Nowadays, many methods and statistical functions to analyze the spatial distribution of forest trees such as: Ripley's K-function, L-function, the nearest-neighbor distance statistic, the pair-correlation function, replicated point pattern analysis (Fangliang *et al.*, 1997; Condit *et al.*, 2000; Luo *et al.*, 2009; Rejou-Méchain, 2011; Hung, 2013). The replicated point pattern analysis is very new. It is very useful for small investigated plots with repetitions. It should be applied in Vietnam (Ramón *et al.*, 2016, Hung, 2016).

The previous studies, especially in Vietnam, researchers often use Excel, SPSS, Stata, Sas ect. However, currently, R is showing advantages in scientific data analysis in general, forestry data in particular. Firstly, R is free and analysts can add, install and use many packages for specific targets. Finally, R can perform analysis that Excel, SPSS, Stata or Sas cannot conduct or hard to conduct, for example point pattern analysis.

REFERENCES

1. Baddeley, A., 2008. Analysing spatial point patterns in R, School of Mathematics and Statistics, University of Western Australia, Crawley, 6009 WA, Australia. Available from: <http://umaine.edu/computingcoursesonline/files/2011/07/SpatstatmodelingWorkshop.pdf> (Accessed 16 August, 2016).
2. Baddeley, A., 2016. Analysing replicated point patterns in spatstat, School of Mathematics and Statistics, University of Western Australia, Crawley, 6009 WA, Australia. Available from: <https://cran.r-project.org/web/packages/spatstat/vignettes/replicated.pdf> (Accessed 25 April, 2016).
3. Baddeley, A., 2015. Spatial Point Patterns: Methodology and Applications with R, CRC Press, Taylor & Francis Group, 6000 Broken Sound Parkway, Boca Raton, Florida, USA.
4. Christensen, N. L., 1977. "Changes in Structure, Pattern and Diversity Associated with Climax Forest Maturation in Piedmont, North Carolina". *The American Midland Naturalist* 97(1): 176 - 188.
5. Condit, R., 2000. "Spatial Patterns in the Distribution of Tropical Tree Species." *Science* 288: 1414 - 1418.
6. Fangliang, H., 1997. "Distribution patterns of tree species in a Malaysian tropical rain forest." *Journal of Vegetation Science* 8: 105 - 114.

7. Fangliang, H., 1997. "Distribution patterns of tree species in a Malaysian tropical rain forest." *Journal of Vegetation Science* 8: 105 - 114.
8. Gadow, K. v., 2011. *Forest Structure and Diversity. Continuous Cover Forestry*. T. Pukkala and K. v. Gadow, Springer.
9. Hai, V. D., 2009. "Research on natural regeneration characteristics of *Schima wallichii* Choisy in rehabilitation forests in Luc Ngan and Luc Nam districts, Bac Giang province." *Vietnam journal of forest science* 7(1): 1149 - 1156.
10. Hai, V. D., 2014. "Research on structure of high trees of forest status IIA in protection forest of Yen Lap reservoir, Quang Ninh province." *Vietnam journal of forest science* 3(2004): 3390 - 3398.
11. Hung, B. M., 2011. The influence of light, planting methods and densities on the growth of *Amoora gigantea* and *Pygeum arboreum* in the North of Vietnam. Fenner School, The Australian National University. Master thesis.
12. Hung, B. M., 2016. Structure and restoration of natural secondary forests in the Central Highlands, Vietnam. Chair of Silviculture, Institute of Silviculture and Forest protection, Faculty of Environmental Sciences, Dresden University of Technology. Doctoral thesis.
13. Hung, N. V., 2013. Research on spatial distribution of over storey trees in evergreen broadleaf forests in Ba Vi and Cuc Phuong National Parks, Vietnam National University of Forestry. Master thesis.
14. Illian, J., 2008. *Statistical Analysis and Modelling of Spatial Point Patterns*, John Wiley & sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.
15. Luo, Z., 2009. "Distribution patterns of tree species in an evergreen broadleaved forest in eastern China." *Frontiers of Biology in China* 4(4): 531 - 538.
16. Mabberley, D. J., 1992. *Tropical rain forest ecology*, Chapman and Hall, Inc. 29 West 35th Street, New York, NY 10001 - 2291, USA.
17. Pan, Y., 2013. "The Structure, Distribution, and Biomass of the World's Forests." *The Annual Review of Ecology, Evolution, and Systematics* 44(2013): 593 - 622.
18. Ramón, P., 2016. "A new non-parametric method for analyzing replicated point patterns in ecology." *Ecography* 39: 001 - 009.
19. Réjou-Méchain, M., 2011. "Spatial aggregation of tropical trees at multiple spatial scales." *Journal of Ecology* 99: 1365 - 2745.
20. R-Project, 2016. What is R?, Available from: <https://www.r-project.org/about.html> (Accessed 09 February, 2016).
21. Sau, L., 1996. Research on forest structure characteristics and propose some economic technical criteria for selective harvest and sustainable forest use in Kon Ha Nung - Central Highlands., Vietnam National University of Forestry. Doctoral thesis.
22. Shiver, B. D. and B. E. Borders, 1996. *Sampling techniques for forest resources inventory*, John Wiley & Sons, Inc. Canada.
23. Spies, T. A., 1998. *Forest Structure: A Key to the Ecosystem. Proceedings of a workshop on Structure, Process, and Diversity in Successional Forests of Coastal British Columbia*. J. A. Trofymow and A. MacKinno, Northwest Science, Washington State University Press.
24. Stamatellos, G. and G. Panourgias, 2005. "Simulating spatial distributions of forest trees by using data from fixed area plots." *Forestry* 78(3): 305 - 312.
25. Valbuena, R., 2015. Forest structure indicators based on tree size inequality and their relationships to airborne laser scanning. Faculty of Science and Forestry, University of Eastern Finland. Doctoral thesis.
26. Wagner, S., 2011. "Canopy effects on vegetation caused by harvesting and regeneration treatments." *European Journal of Forest Research* 130(2011): 17 - 40.
27. Wehenkel, C., 2015. "Spatial Distribution Patterns in the Very Rare and Species-Rich *Picea chihuahuana* Tree Community (Mexico)." *Plos* 10(11): 1 - 19.
28. Wiegand, T. and K. A. Moloney, 2014. *Handbook of Spatial Point Pattern Analysis in Ecology*, CRC Press, Taylor & Francis Group, 6000 Broken Sound Parkway, Boca Raton, Florida, USA.

Email của tác giả chính: buimanhhungvfu@gmail.com

Ngày nhận bài: 20/07/2017

Ngày phản biện đánh giá và sửa chữa: 02/10/2017

Ngày duyệt đăng: 03/10/2017