

Point Pattern Analysis

Vitoria Barbosa Ferreira

5/13/2021

Introduction

This assignment focuses on a point pattern analysis using natural and simulated individual tree locations, situated at the alpine tree line in the Passeier valley (South Tyrol, in North Italy). This dataset is part of Gudrun Wallentin's thesis about individual tree modelling.

Nearest Neighbour Point Analysis

In order to compare real-world trees distribution with simulated tree distribution, we will perform a Nearest Neighbor point Analysis, aiming to calculate the distance between individual trees and its nearest neighbor.

Since we want to compare both datasets we will use the mean nearest neighbor. The output shows that the mean distance between the real-world trees is around 6.70m. Simulated trees shows a very similar result, around 6.67m. The nearest neighbor index, which indicates how clustered is the data, shows that the real-world trees tend to be more clustered (nn index less than 1) than the simulated trees (nn index more than 1).

However, the p-value, or the probability of the observed spatial pattern was created by some random process (ESRI), of both datasets shows that it is very unlikely to be completely randomly distributed, which is also proved by the high and very low z-score values.

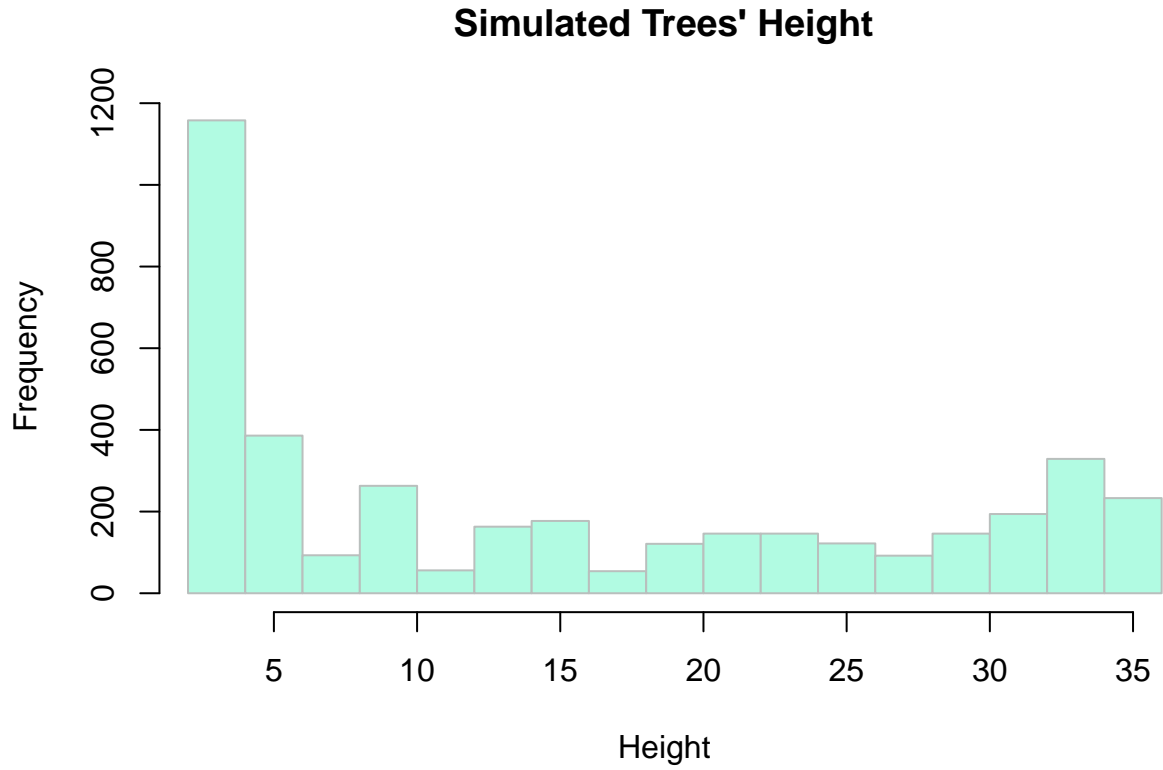
Table 1: Real-world Trees

d.NNI	d.z.score	d.p	d.expected.mean.distance	d.observed.mean.distance
0.9654683	-4.253158	2.11e-05	6.947764	6.707846

Table 2: Simulated Trees

s.NNI	s.z.score	s.p	s.expected.mean.distance	s.observed.mean.distance
1.065133	7.76052	0	6.270697	6.679125

Now that we know that data is not completely random distributed, let's further analyze the simulated trees dataset. For that we will split the data based on its main attribute, the height. Two subgroups were defined based on the median value, which is about 11.88m. The histogram below shows the distribution of the simulated trees' height.



By running the Nearest Neighbor Analysis again for each subgroup it is possible to conclude that the group of trees up to 11.88 meter high tend to me more clustered (nn index = 0.80) than the group of trees with less than 11.88 meters high (nn index = 1.10).

Table 3: Simulated Trees - Trees'Height < 11.88

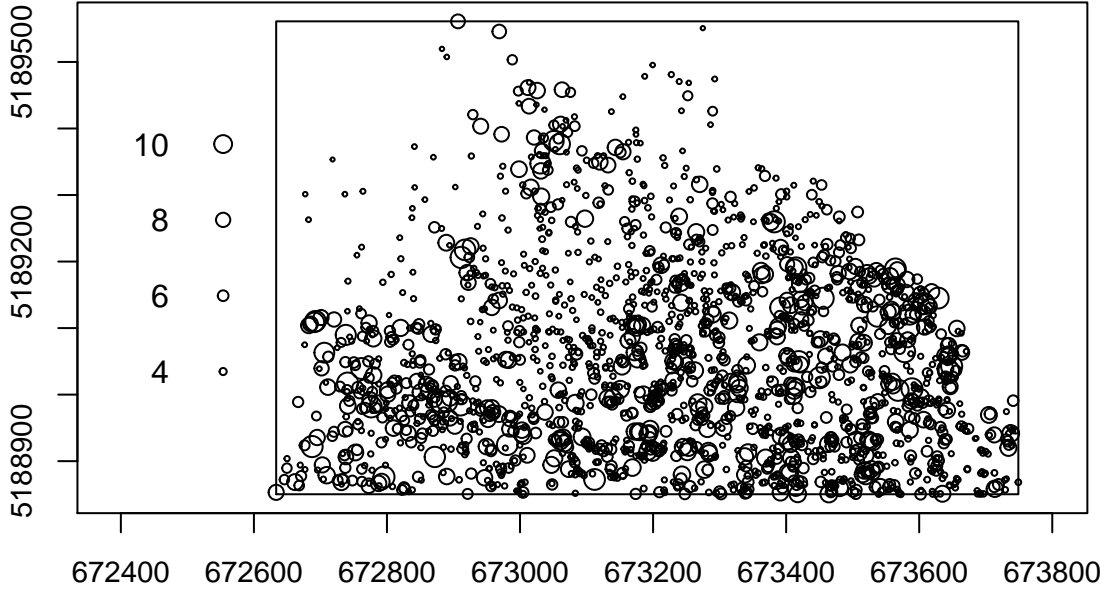
s.NNI	s.z.score	s.p	s.expected.mean.distance	s.observed.mean.distance
0.8000159	-16.84671	0	8.824089	7.059412

Table 4: Simulated Trees - Trees'Height > 11.88m

s.NNI	s.z.score	s.p	s.expected.mean.distance	s.observed.mean.distance
1.10117	8.522558	0	7.612068	8.382178

In the sequence, we decided to further split the simulated data with trees'height less than 11.88m into clustered, dispersed and random by visual inspection of the following plot.

Simulated Trees (<11.88m high)



Next, the Nearest Neighbor Analysis were run one last time. As expected, the Nearest Neighbor Index of the Clustered group is the lowest indicating it is the most clustered group. Even though we set the group random based on visual interpretation, the NNA proved it is not completely random, although less clustered than the cluster group. We can also see that mean distance between the random trees group is more than 3 times higher than the Clustered group.

Table 5: Clustered

c.NNI	c.z.score	c.p	c.expected.mean.distance	c.observed.mean.distance
0.815331	-10.72148	0	7.009955	5.715434

Table 6: Dispersed

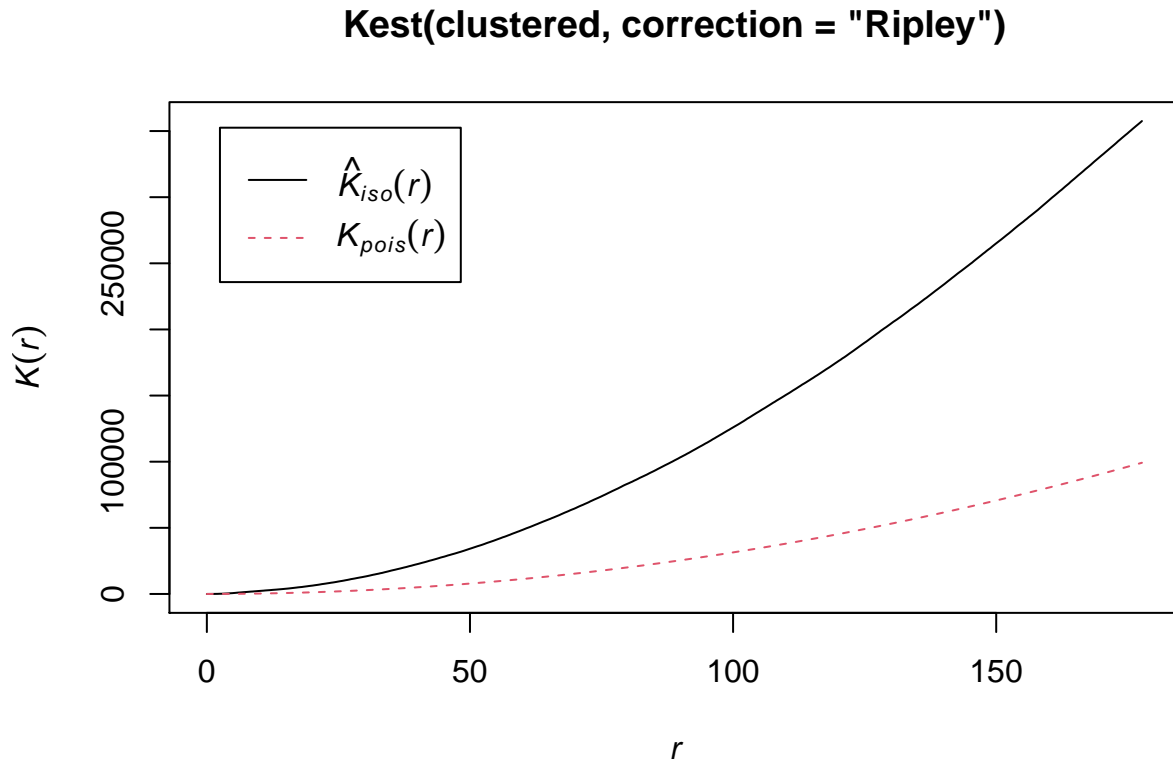
d.NNI	d.z.score	d.p	d.expected.mean.distance	d.observed.mean.distance
0.869654	-7.460024	0	8.492633	7.385652

Table 7: Random

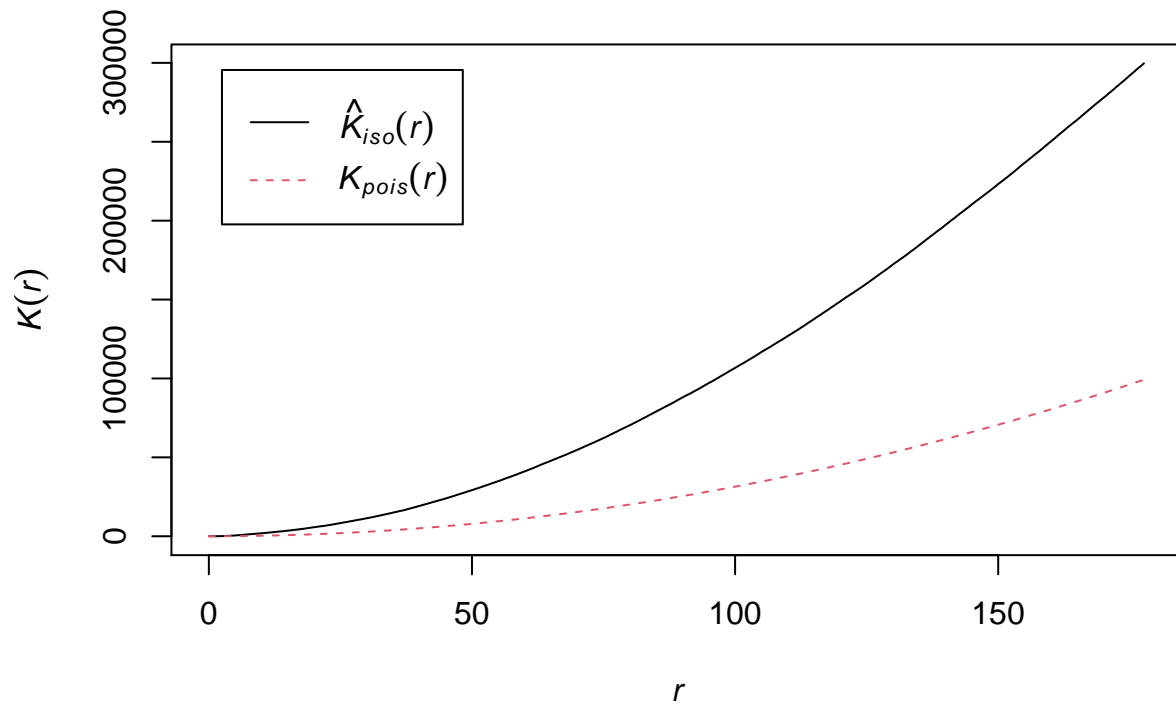
r.NNI	r.z.score	r.p	r.expected.mean.distance	r.observed.mean.distance
0.919903	-1.699416	0.0892408	17.33828	15.94954

Ripley's K-function

Lastly, we will use the Ripley's k-function to analyze the degree of clustering of the subgroups: clustered, dispersed and random. Instead of measuring the nearest distance between the points, the Ripley's K-function method considers all the distances between events in the study area. In the plots below, the dashed line represents the expected value under complete spatial randomness (CSR) and the black solid line indicates the actual estimated values. In all cases, since the estimated values varies significantly from CSR, it indicates the tendency of clustering, emphasizing the previous outputs.



Kest(dispersed, correction = "Ripley")



Kest(random, correction = "Ripley")

