I

Geo-information Science and Remote Sensing
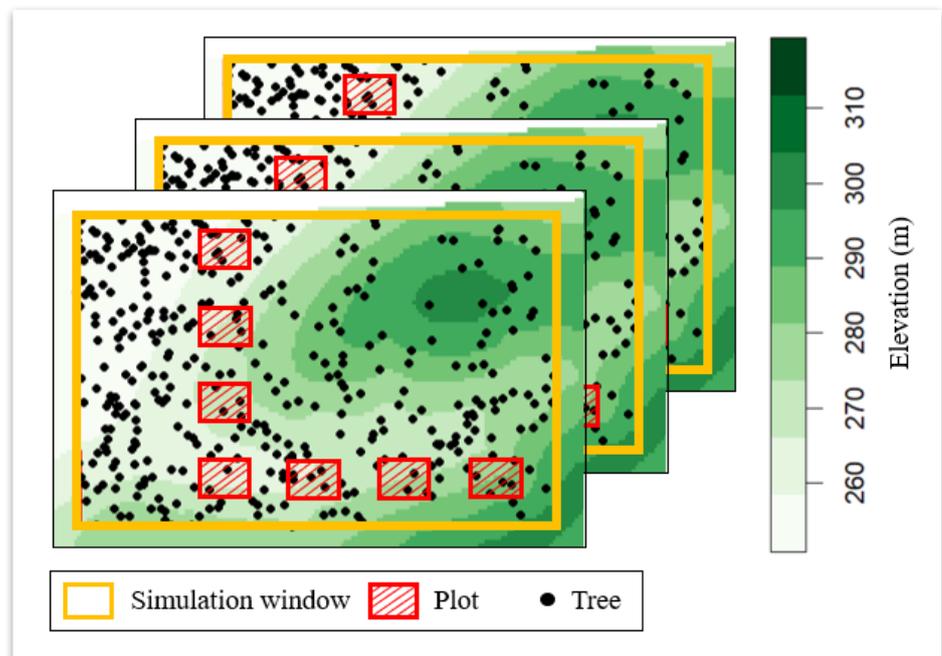
Thesis Report GIRS-2023-02

The Impact of the Spatial Configuration of Forest Inventory plots for Integration with Remote Sensing-based Biomass Products

Kristie Swinkels

02-04-2023

WAGENINGEN
UNIVERSITY & RESEARCH

The Impact of the Spatial Configuration of Forest Inventory plots for Integration with Remote Sensing-based Biomass Products

Kristie Swinkels

Registration number: 1006933

<u>Supervisors</u>:

Natalia Málaga Durán

Sytze de Bruin

A thesis submitted in partial fulfilment of the degree of Master of Science

at Wageningen University and Research,

The Netherlands.

02-04-2023

Wageningen, The Netherlands

Summary

The limited accuracy of Above-ground biomass (AGB) estimates restricts our capacity to put effective carbon emission mitigation strategies into action. Most countries assess information on forest, such as AGB, through surveys called National Forest Inventories (NFIs). NFI's use sampling units, containing plots, ranging between 0.05 and 0.5 ha in size, in which traits of trees that can be used to obtain the AGB are measured. Even though ground surveys are often the primary source of AGB data, remote sensing (RS) -based biomass products can be used to improve maps created with this data. However, the difference in spatial support between the RS-based biomass product pixels and sampling plots can lead to discrepancies in AGB estimation. To deal with this, the mean AGB over a polygon encompassing all plots is compared to the mean AGB of all RS-based biomass product pixels within that polygon. However, how well the plots represent the AGB over that polygon is influenced by their spatial configuration, which raises the research question: what is the best configuration of ground-based plots to capture the AGB within a sampling unit? To answer the main research questions, the following questions were answered:

1. What are the parameter values of HMPP-models for forest stands in different biomes?

2. What are common, plot configurations used for forest inventories?

3. Which plot configuration from question 2 achieves the highest accuracy in estimating AGB within a sampling unit that encompasses the entire plot?

Lister and Leites (2018) laid a basis for simulating AGB using a hierarchical marked point process (HMPP). They used a point based representation of trees and took into account the hierarchical nature of the forest. This was done in this study too. The HMPP framework was parameterized for two biomes, a tropical and a temperate forest. The models of both sites were simulated 100 times and all simulations were samples using six plot configurations. The mean error and the root mean square error were calculated for each plot configuration they were compared to determine which configuration was the most accurate.

On the temperate forest site, elevation influenced the distribution of AGB most. Determining which plot configuration was most accurate was unsuccessful since using a fixed location for every plot configuration introduced a sampling bias. The tree distribution models showed that the point intensity was negatively influenced by elevation (higher elevation – less trees). Therefore, over 100 simulation, fewer trees were simulated in places with high elevation than in places with low elevation. Some plots were placed on locations with higher elevation that other plots, making the comparisons between the plots unfair.

The location of the trees on the tropical forest site was mainly influenced by the presence of other trees and many trees formed clusters. The apparent clustering could be explained by (1) a mistake in the geo-referencing of the tropical forest data. Rather than clustering, indicating that the presence of one tree influence the presence of other trees, the trees could be influenced by the topography of the landscape. (2) Besides, the study site has been repeatedly hit by typhoons, causing high tree mortality in trees with large DBH's. Another mistake, which caused less points to be simulated near the edges, caused the accuracy assessment of the plot configurations to be inconclusive for the tropical study site too.

To gain insight into which plot configuration most accurately estimates the AGB within a sampling unit, the plot configurations should be randomly shifted, to avoid bias caused by topographical (or other) variables.

Acknowledgement

I would like to express my gratitude towards my supervisors, Natalia Málaga Durán and Sytze de Bruin. More specifically, Sytze, I am most grateful for your elaborate and quick responses to my coding problems. Natalia, I really appreciated the positive reinforcements when I was overwhelmed with my project.

Secondly, I would like to thank Andy Lister and Arnan Araza, for providing the scripts that served as a basis for the model development and the data of the Philippine study site respectively.

Lastly, I would like to thank my friends and family for proof reading, and supporting me throughout the process.

Contents

1. Introduction

1.1 Context

 Deforestation and forest degradation in tropical forests causes approximately 8% of the global annual anthropogenic CO_2 emissions (Gibbs et al., 2018). Most of the members of the United Nations Framework Convention on Climate Change (UNFCCC) are parties to the Paris Climate agreement (UN Climate Change, 2022). The agreement aims to mitigate carbon emissions in order to keep the temperature rise below 2 degrees Celsius. The UNFCCC not only aims to lower these emissions by reducing deforestation and forest degradation, but also sets the aim to capture CO_2 by *enhancing* forest carbon stocks; according to Griscom et al. (2017). Reforestation can contribute approximately 25% to climate mitigation potential that nature has to offer in order to keep the temperature rise under 2 degrees Celsius. However, the limited accuracy of forest carbon estimates restricts our capacity to put effective carbon emission mitigation strategies into action (Nesha et al., 2022; Réjou-Méchain et al., 2014).

1.2 General ways to tackle the issues.

Above-ground biomass (AGB), the mass of all the material held in living stems, branches, and leaves of vegetation (Araza et al., 2022), globally accounts for 75-80% of the living biomass (Handavu et al., 2021) and is thus large component of the global carbon cycle (Nesha et al., 2022; Herold et al., 2019). Therefore, accurate and up-to-date information on AGB stocks and changes is key for estimating the climate mitigation potential of forests (Næsset et al., 2020). Traditionally, ground-based sample surveys are used to obtain information on AGB at regional or national scale (Næsset et al., 2020).

 The biomass of a tree can be predicted with allometric models - models that capture the relationship between bio-physical attributes (e.g. tree height and diameter at breast height -DBH) and the tree biomass (Vorster, 2020). But to do this for every tree in a whole forest, let alone a whole country, is prohibitively expensive. Therefore, most countries assess information on forest, such as AGB, through surveys called National Forest Inventories (NFIs). Figure 1 shows an example of a sampling unit of an NFI. Within such a sampling unit, plots, ranging between 0.05 and 0.5 ha in size (Nesha et al., 2022) are established, in a certain configuration. Within every plot, tree locations and some biophysical attributes (i.e. DBH, tree height, tree species) are registered. Such sampling unit arrangement, having several plots within a sampling unit, has been reported to be *operationally*- thus moneywise- more efficient (Yim et al., 2015).
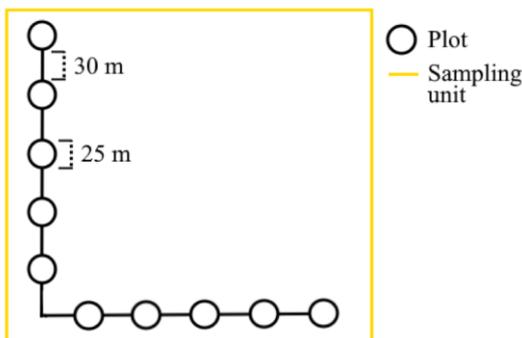
Figure 1: *NFI sampling unit of Peru. The larger circles are plots and lie within the sampling unit: the yellow rectangle. (Edited from* Nesha et al. (2022)*)*

 Even though ground-based surveys are often the primary source of AGB data, the precision of their estimates might be hindered by the completeness and continuity of the surveys (Guitet et al., 2015). Besides, they are expensive and sometimes lack data in inaccessible areas (Naesset et al., 2020). To remedy this, the UNFCCC reports recommend integrating the ground-based sampling data with Remote Sensing (RS) biomass products, as it can improve national AGB estimation. For example, RS-based biomass products are used as an auxiliary source of information to increase the precision of the national estimates or improve the efficiency of

the NFI stratification (Næsset et al., 2020; Herold et al., 2019). Still, such RS-based biomass product to ground-based-plot integration efforts come with harmonization challenges and new sources of uncertainty. For instance, the difference in spatial support between the two can lead to discrepancies in AGB estimates (Guitet et al., 2015; ESA, 2020). NFI plots within a cluster are often smaller in size than RS-based-product pixels (i.e. ~100 m x 100 m spatial resolution), and even when individual plots match the size of RS-based biomass product pixels, the RS-based biomass product pixels and NFI plots will not align (Nesha et al., 2022).

To deal with the difference in spatial support causing discrepancies in AGB estimates, Málaga et al. (2022) defined a sampling unit that encompassed the entire clustered-plot in the form of a polygon (Figure 2), and compared the mean AGB of the pixels within that polygon to the mean AGB observed over the clustered plots located within the sampling unit (Figure 2). Within the study of Málaga et al. (2022), various sources of uncertainty when integrating a RS-based biomass products with NFI data were assessed. They found that within sampling units, spatial variability in biomass was the main source of uncertainty as the ground-based sampling plots were not able to capture the AGB within the square sampling unit. This raises the question: within the boundaries of current ground-based NFI sampling designs, what is the best plot configuration to captures the mean AGB within a sampling unit?
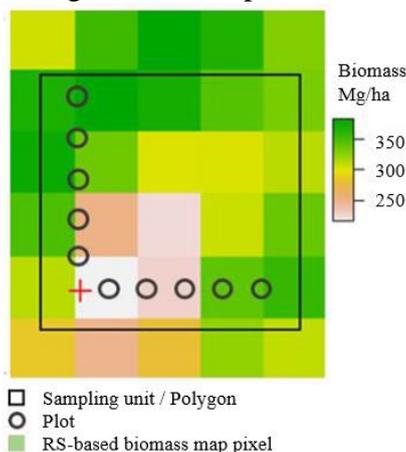

Figure 2: Set-up to integrate RS-based biomass products and NFI plot data presented by Málaga et al. (2022) *(Edited from* Málaga et al. (2022).*)*

1.3   Other research on the topic

When estimating the AGB within a larger plot (i.e. the sampling unit), using single plot can be beneficial over using multiple clustered plots, as using multiple clustered plots can lead to sampling errors (Réjou-Méchain et al. 2014): when the area-to-edge decreases, relatively, there is more edge and consequently, relatively more trees located close to the edge and that leads to error. In NFI field campaigns, a common practice is to consider a tree inside a plot when 50% of its trunk is in the plot (e.g. Mascaro et al., 2011). However, when <50% of a trunk is inside the plot, or only a part of the tree crown is inside the plot, this AGB is not counted as such (Packalen et al., 2022). Therefore, plots might contain much more or less AGB than measured. This sampling error is visualised in Figure 3. This sampling error is a downside to using multiple plots compared to using a single plot, as single plot relatively have more edge length. However, there is also an upside to using multiple plots: when spatial AGB variability is large, such as in the tropics (Réjou-Méchain et al., 2014), multiple plots covering a larger distance can capture the tree data (i.e. tree location and their AGB) more precisely (Yim et al. 2015).
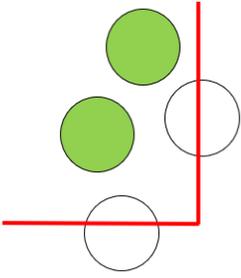
Figure 3: Tree trunks (circles) along a plot edge (red lines). Common practice considers trees to be inside a plot only when their trunk is inside the plot for more than 50%. Therefore, only the green circles are considered to contribute to the AGB inside the plot.

In the scientific field surrounding AGB estimation from ground-based data and RS, AGB is either modelled as a variable continuously in space (RS) (Guitet et al., 2015; Kleinn et al., 2020) or as discrete variable directly linked to tree stems, which are represented by points. To find out which plot configuration best captures the AGB within a sampling unit, it makes more sense to use a discrete bottom-up approach because then the afore mentioned sampling error caused by the edge effect can be accounted for. Using point based representation is not new in forest statistics: the foundation of modern point process statistics was laid in 1960 (Stoyan & Penttinen, 2000). For example, Lister and Leites (2018) used a point based model to model the spatial distribution of trees and their biological interaction. To fully capture the spatial distribution and tree-interaction, they used a Hierarchical Marked Point Process (HMPP) framework, which will be in depth explained in the methodology.

1.4 Overall research aim and questions

The main aim of this research is to study the impact of spatial configuration of NFI plots for integration with RS-based biomass products through modelling. This will be studied by answering the questions below:

1. What are the parameter values of HMPP-models for forest stands in two different biomes?

2. What are common, distinct plot configurations used for forest inventories?

3. Which plot configuration from question 2 achieves the highest accuracy in estimating AGB within a sampling unit that encompasses the entire plot?

2. Theoretical Background

For modelling the AGB distribution within a forest stand, marked point processes can be an effective tool (Lister & Leites, 2018). This chapter provides an overview of some fundamental concepts of point processes and the basics of their application in forest statistics and how then can be practically applied using the R package Spatstat (Baddeley & Turner, 2005). The understanding of these concepts is essential for understanding of the modelling approach presented by Lister and Leites (2018), which is employed in this thesis.

2.1 The Fundaments of Point Processes

A marked point process is a modelling approach that can be used to study spatial patterns in forestry. It uses a point-based representation, where a tree – and its location – is indicated by a point and one of its characteristics (e.g. height or DBH) is indicated by a mark (Stoyan & Penttinen, 2000). When fitted to ground-based data, the model does not only describe the spatial pattern of the tree locations and their marks: it can also be used to *simulate* a pattern. The observed locations of trees within a stand can be seen as a single realisation of a stochastic event-generating process. Each simulation using the model generates another realisation (Lister & Leites, 2018).

Point interaction

A point process can be used to model the interaction between points representing trees. Baddeley (2010) suggests three ways points can interact, which leads to different patterns. (1) Point do not interact. In that case, the points are independent from each other and randomly distributed (Figure 4a). If so, the point pattern shows complete spatial randomness (CSR) and is often referred to as a realization of a Poisson process. (2) Alternatively, points can repel each other, which leads to a regular point distribution with similar distances between them (Figure 4b). (3) Lastly, the points can attract one another, and form a clustered distribution (Figure 4c). This does not mean that the points are organized in identifiable clusters, but rather that there are patterns of points that are closer together than expected in a Poisson process. (Baddeley, 2010)
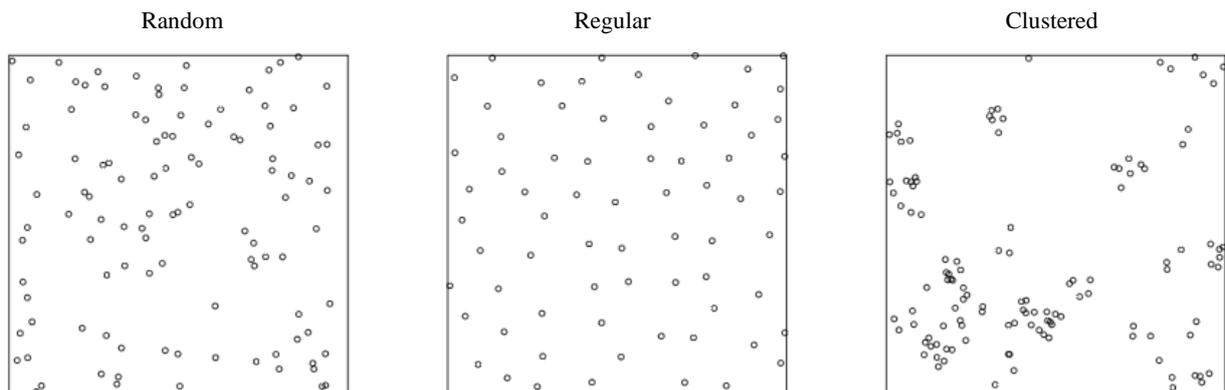

Figure 4: Degrees of clustering in interacting point patterns. Adapted from Baddely (2010).

In addition to points interacting with one another, points can interact with external factors. If so, the intensity of a point pattern can vary from place to place reflecting that a point pattern is inhomogeneous. Conversely, if points do not interact with external factors, the point intensity is constant (i.e. homogeneous - Figure 5a). The simplest point process, a homogeneous Poisson process, is characterized only by the intensity of the points (λ), which is expressed in points per unit area. It is often used as a baseline model, to which other point processes are compared, in order to diagnose either clustering, regularity or CSR (Baddeley, 2010). If the density of a Poisson process is inhomogeneous, it might be worthwhile to research if there are phenomena causing this inhomogeneity. When it comes to trees, for example, the ground water level could be a phenomenon that causes pattern to be inhomogeneous; big trees cannot grow where the ground water level is high, as they do not grow roots below the ground water table. Such phenomena could be used as covariates

for the point process model, which would take the form of Equation 1: the intensity of the point pattern at location u, $\lambda(u)$, is a function of $Z_n(u)$, spatial covariate n at location u, and model parameters θ_0 , ... θ_n . (Baddeley, 2010)

$$\lambda(u) = e^{\theta_0 + \Sigma_{i=1}^{n} \theta_i + Z_i(u)} \tag{1}$$
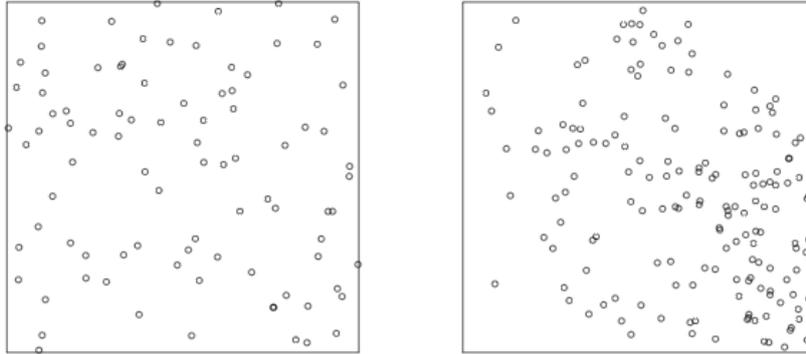

Figure 5: Examples of point density patterns: a)homogeneous; and b) inhomogeneous . (Adapted from Baddeley, 2010).

2.2 Applications of Point Processes in Forest Statistics

Common models

In forest statistics, two families of point processes are most commonly used: Cox processes and Gibbs processes. Cox processes are often used for modelling clustering. One of the earliest cluster models (1960), is the Matérn cluster (Stoyan & Penttinen, 2000). The base of this cluster process is the "parent" points, which are generated by a Poisson process. On a disk centred around these parents, "daughter" points are generated by another Poisson process. A second cluster process, is the Thomas process which is also of the Cox process family. It is very similar to the Matérn cluster, only the intensity function is more continuous. The only difference is that the offspring is not Poisson distributed around the parent point, but according to two zero-mean, symmetrical, independent normal distributions – one for the x- and one for the y-coordinate (Afshang et al., 2017).

As the Matérn and Thomas clusters both consists of two point processes, there are two ways to incorporate inhomogeneity. The first is simply using an inhomogeneous Poisson process for the parent points. This does often not suffice as the radius of the daughter cluster can be large and therefore, reach a large range of the spatial covariate values. For example, take a point pattern representing trees, where a the trees show clustering, the radius of the daughter cluster radius is 40 meters and the hill slope is a model covariate. Within the radius of 40 m around the parent points, the slope can severely change. Therefore, only taking the covariate into account for the parent points does not stop the daughter point from being placed on steep slopes. Therefore, Waagepetersen (2007) proposed a more complex method to also incorporate inhomogeneity into the daughter points by first estimating the overall inhomogeneous intensity of the two combined processes ($\lambda(u)$, Equation 2), and then thinning the parent points (Fabris-Rotelli & Stein, 2018). This overall intensity consists of two components: λ_p, the homogeneous intensity of the parent points, and λ_d, an inhomogeneous cluster density of the daughter points (Baddeley, 2010).

$$\lambda(u) = \lambda_p \lambda_d \tag{2}$$

Gibbs processes are, on the other hand, better for modelling regular patterns. Opposed to the Cox models, where only the parent interacts with the offspring, Gibbs models are based on symmetrical pair-wise point interactions (Stoyan & Penttinen, 2000).

Hierarchical Point Process Framework

For forest stands, however, the simple point processes introduced in the previous paragraphs, are not flexible enough to model the complexity of trees within a forest stand (Lister & Leites, 2018). This is because trees of different sizes in forests compete for space , which is not captured within any of the individual point processes above. The presence of a bigger tree influences the presence of a smaller tree, but not the other way around (Högmander & Särkka, 1999). This can be modelled using a hierarchical model. (Lister & Leites, 2018)

To capture this hierarchy, Lister and Leites (2018) used a hierarchical marked point process modelling (HMPP) framework to model and simulate the spatial pattern of trees. The HMPP framework captures the hierarchy of the forest by assigning trees to different hierarchical levels. The levels are based on competitive strength, for which DBH is an indicator. Therefore, the hierarchical levels are based in DBH. However, not only trees of different levels interact: trees within a level also interact. Therefore, each hierarchical level is assigned a separate point process family (i.e. model type).

Figure 6, a visual representation of a HMPP framework, shows that for every hierarchical level, a different model is built (left side of Figure 6). The hierarchy goes from trees with the biggest DBH (level 4 in Figure 6) to the smallest DBH (level 1 in Figure 6). The density of every higher point pattern is used as an input, in the form of a covariate, to build the model for the lower hierarchical level. Subsequently, the models are used to simulate tree patterns (see the right side of Figure 6). One constraint is, that these models can only simulate a spatial pattern and not marks (the software simply is not written for this (Baddeley, 2010), and therefore, the models cannot simulate biomass.
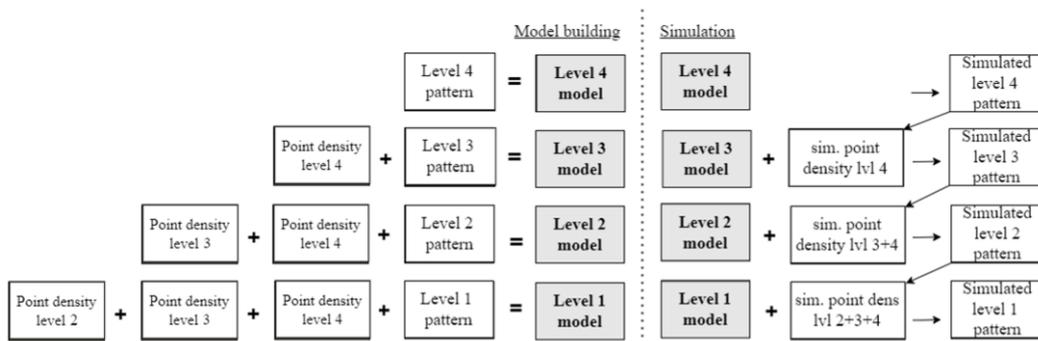

Figure 6: Hierarchical marked point process framework. The left hand side of the diagram shows the model building process. First, a model is built for the trees with the largest diameters: the level 4 pattern. Then, a model level 3 is built, using the point pattern of the level 3 trees and the tree density of the higher level. The same goes for level 2 and 1. Then, within thesimulations, not only the model for the given level is used, but also a simulation of the point density of the higher levels.

2.3 Practical Application of Point Processes in R

Fitting a point process to point data

Various packages are available in R to analyse and generate point patterns, of which Spatstat (Baddeley, 2010) is one. Spatstat uses statistical models to analyse point pattern, of which two are addressed in this study. (1) "`ppm(Y, trend)`", a function that fits both homogeneous and inhomogeneous Poisson models to point data. For a homogeneous Poisson process, the only model input is a point pattern (`Y`) and the final model can be expressed in just a single number: the intensity of the observed point patten (points per unit area). For an inhomogeneous Poisson process, the ppm function not only needs a point pattern as an input, but also a trend. The trend is a covariate on which the inhomogeneity of the point pattern depends; a pixel image containing the values of a spatial covariate suffices as function input. The final model (i.e. the intensity function, see formula 1) is more complex than for a homogeneous point process, as the intensity function of an homogeneous point process is just a single number.

(2) "`kppm(Y, trend, ClusterType)`" is a function that fits cluster models (either type Matérn or Thomas), both homogeneous and inhomogeneous, point data. The model consists of two intensity functions: one for the parent process and one for the daughter process (formula 2). The function "`summary`" can be used to get an insight of parameter values. Appendix I shows how to interpret the parameter values and translate them to equation 1 or 2.

To fit the kppm function, a pattern descriptor based on statistics (i.e. summary function) is used (A.

Baddeley, 2010). This can either be Ripley's *K*-function ($K(r)$), a function that expresses expected number of points within a distance *r* of another point, divided by the point intensity. Alternatively, the pair correlation function *g(r)* can be used, which expresses the probability of finding two points that are a distance *r* apart, divided by the corresponding probability for a Poisson process. The kppm function fits the summary function of the model to the summary function of the data.

Testing the fit of a point process model to point data

Besides fitting cluster models, the summary functions can be purposed for judging the fit of a point process model to observation data (Baddeley, 2010). Another example of a summary function is the nearest neighbour function ($G(r)$), a cumulative distribution function of the distance *r* from a random point of the distribution to the nearest other point, can be used for this purpose. Additionally, the Besags's L function ($L(r)$) is a transformation of Ripley's *K*-function, designed to make visual assessment easier. This transformation is done by dividing *K(r)* by π and thereafter, taking the square root of the outcome.

Judging the fit of a model using summary functions can be done by means of simulation envelopes (Baddeley, 2010). A simulation envelope contains summary functions (e.g. *G(r)*) of *n* model simulations. Figure 7 shows an example of simulation envelope (it's range in grey) of the G-function of a Poisson process model . As the black line lies fully within the simulation envelope, CSR of the observed point pattern is likely and a Poisson process fits the data. Would the line either exceed the upper or lower boundary of the simulation envelope, the pattern is more likely to be either clustered or regular respectively. An envelope can be created using the Spatstat function "`envelope(Y, n, fun)`", where Y is a either a point pattern or a model (either created by the ppm or kppm function) and fun is a summary function (e.g. *G(r)*) and n a number of simulations. If the Y is a point pattern, the envelope function creates an envelope of the summary statistic of *n* simulations of a Poisson model (i.e. the envelope function creates *n* random patterns and their summary statistics) and plots the summary function (`fun`) of the observed point pattern Y on top. Alternatively, if Y is a model, either created by ppm or kppm, the envelope function will simulate the model `n` times and plot the envelope of summary function (`fun`) together with the corresponding summary function of the observed pattern.

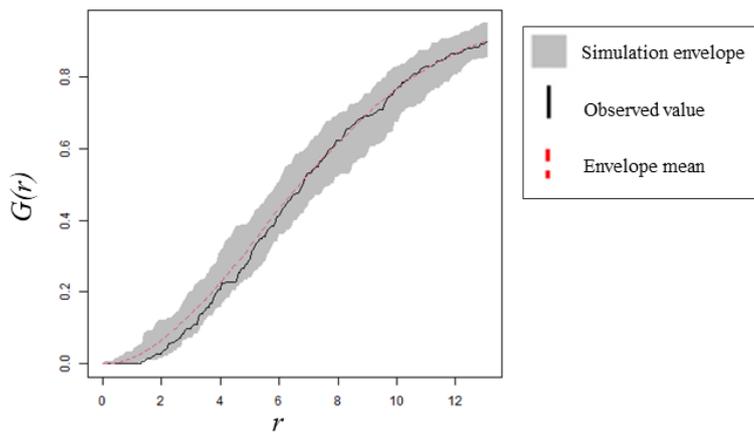*Figure 7: Example of a simulation envelope of the G(r) nearest neighbour function of a Poisson (i.e. random) model, where r represents distance. The grey area indicates the simulation envelope of the G statistic of simulated Poisson patterns.*

3. Methodology

This chapter, first of all, introduces the study sites and then every next subchapter explains how each research question is addressed. A list if all data that was used to research answer the research questions, is found in appendix II. Chapter 3.2 explains the steps that were undertaken to fit models to the tree data of the study sites. First, a preliminary analysis was done, where a large range of covariates and point process families were tested to see how well they fit the data. Based on this pre-selection, the models were further developed. Chapter 3.3 goes into how the plot configurations were selected and the limitation to their selection. Lastly, chapter 3.4 demonstrates the steps that were taken to assess the accuracy of the various plot configurations.

3.1 Study Sites

Datasets of two study sites in two different biomes were used in this study. The first site is an 8 ha plot in a temperate forest located in Pennsylvania, USA, from now on referred to as the temperate forest. The dataset contains the location and DBH values of 2059 trees. The site is located between 240 and 300 m above sea level and a humid continental climate prevails, which is characterised by its four seasons, including warm summers, and a lack of dry spells (Beck et al., 2018).

The second site is a 16-ha plot of tropical rainforest located in Palanan, the Philippines, from here on referred to as the tropical forest, which contains 10691 trees (Co et al., 2006). Only 8 out of the 16 ha of the dataset were used to fit the model for the tropical forest due to a lack of processing power. Besides, only trees with a DBH ≥ 10 cm were used to fit the model. Trees with DBHs smaller than 10 cm contribute less than 5% to the overall AGB carbon pool in mature tropical forests (Réjou-Méchain et al., 2014; Chave et al., 2004). The site is located between 90 and 140 m above sea level and a tropical rainforest climate prevails, which is characterized by a monthly precipitation over 60 mm. The tree data of this site was not georeferenced when provided. Only the coordinates corner points of the study site were provided. The trees were referenced relative to the corner points: (0, 0) being the lower left corner point of the provided coordinates, to (400, 400) being the upper right corner point of the provided coordinates (see Appendix III for further explanation).
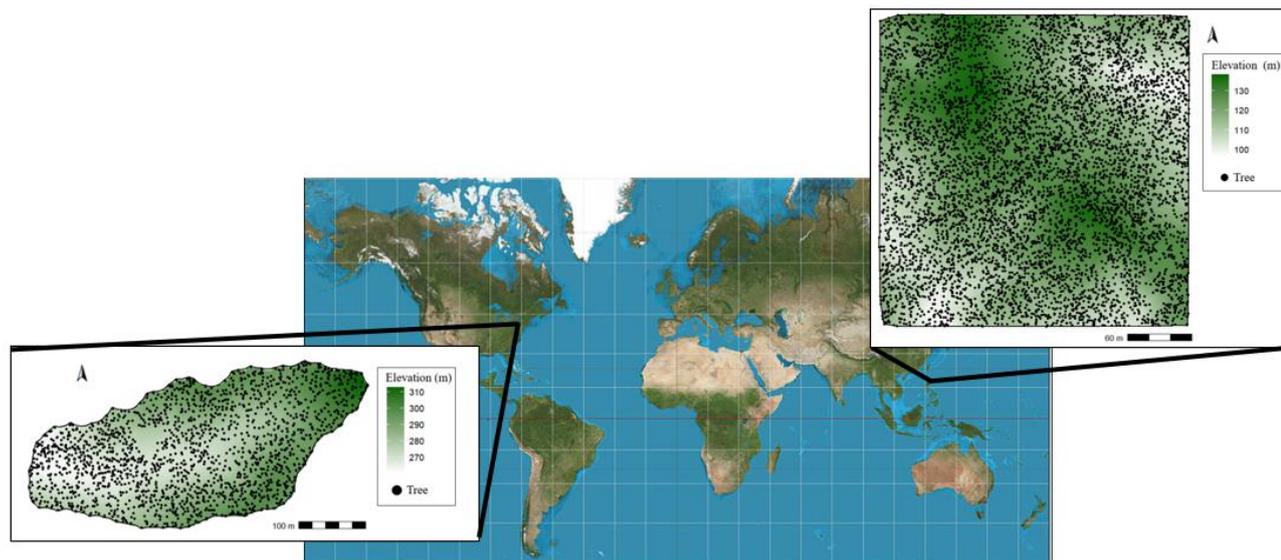
*Figure 8: The study sites: their relative location, trees observed on site and the elevation. Left: the **temperate** forest site in the US. Right: the **tropical** forest site in the Philippines (source basemap: Strebe, 2011).*

3.2 HMPP-models identification

In this study, the AGB of two forest stands was assessed using Lister and Leites (2018) HMPP framework. Lister and Leites (2018) parameterised the framework for a temperate forest site. In this thesis research, models were fitted for both the temperate site and the tropical site using the fitting for the temperate site as a tutorial for the decision making process of the tropical case study.

The HMPP framework uses two types of covariates for model parameterisation: the density of trees of higher hierarchical levels, and other environmental conditions. In this study, the other environmental conditions used for model parameterisation were topographic variables (e.g. elevation or slope). However, topographical variables are not the only environmental conditions that influence the AGB of a forest stand. Besides topography, water availability throughout the year and soil fertility also influence the AGB in a forest stand (Requena Suarez et al., 2021). That is why, in this research, data of proxies of these environmental conditions were considered as model covariates (left column of Table 2).

The framework presented by Lister and Leites (2018) was followed to parametrize and select the best point process families per study site and is split into two sections: (1) a preliminary analysis and (2) model development and selection. The methodology of the HMPP-model parameterisation is further explained in the coming sections.

Preliminary Analysis

The first step to choosing the best model for every hierarchical level, was to define the right point process family for each hierarchical level within the study sites. Therefore, the trees were subdivided into hierarchical levels based on their DBH values using Jenks optimization (step 1a, Figure 9). To assign the right point process family to every hierarchical level, the point process was compared to a homogeneous Poisson process to check for complete spatial randomness (CSR) (step 1b, Figure 9) (Baddeley, 2010). This comparison was done using functions of three summary statistic (i.e. characteristic pattern descriptors) of the point pattern: $G(r)$, $L(r)$ and $g(r)$ (see chapter 2.3).
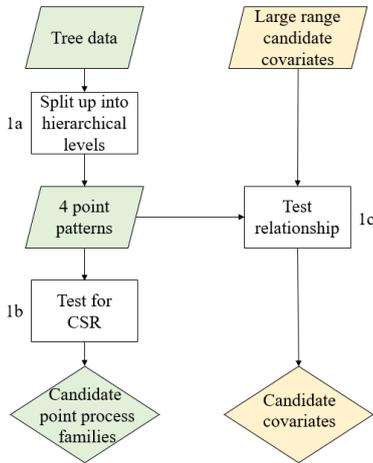

Figure 9: Schematic representation of the preliminary analysis.

39 simulations of a CSR pattern were done, as proposed by (Baddeley et al., 2014), and their summary functions ($g(r)$, $G(r)$ and $L(r)$) were used to create simulation envelopes. Each simulation envelope, consisting of a 39 realisations of summary functions, was compared to the corresponding summary functions of the observed point pattern. See chapter 2.3 for the visual interpretation of the simulation envelope.

The next step to choosing the best model for every level, was to assess the relationship between the point patterns and environmental condition-proxies (step 1c, Figure 9). Two tests were used to do this: a Kolmogorov-Smirnov (KS) test was done to determine if the relationship between the point intensity and the covariate is non-random. Besides, a relative distribution estimate was done to describe the relationship between the point pattern and a possible model covariate (Lister & Leites, 2018). For this purpose, the function f (in Equation 3) was used, which gives an indication of the nature of the relationship between the point intensity and the covariate (whether it is positive, negative, strong, weak or non-existent). f is a function of $Z(u)$, covariate Z at location u and λ is the point intensity at a place u.

$$\lambda(u) = f(Z(u)) \qquad (3)$$

Model Development and Selection

From the models and covariates that followed from the preliminary analysis, the optimal combination was selected for every level (Figure 10). This was done, once again, using summary functions *g(r), G(r)* and *L(r)*. For every combination of possible point process family + covariates, a model was fitted to the data using the Berman-Turner Maximum Likelihood for Poisson processes as proposed in (A. Baddeley & Turner, 2000) and the minimum contrast method was used for cluster processes, as described by Diggle and Gratton (1984) using Ripley's K-function. 39 simulations were done with each model, and their summary functions (*g(r), G(r)* and *L(r)*) were used to create simulation envelopes. Each simulation envelope, consisting of 99 summary functions, was compared to the observed point pattern.

The best combination of the model and covariates of every level were then selected based on the three guidelines as proposed by Lister and Leites (2018). The first guideline was to pick the model where the summary functions of the observed point pattern (the black line in Figure 7) lied mostly within the boundaries of the simulation envelope (the grey area in Figure 7). The second guideline was to pick the model-covariate ensemble where the higher level point density was used as a covariate, in case where two models performed similarly in the envelope-test. The last guideline was to pick the model that was simpler, according to the parsimony principle. In the context of this research, that means that a homogeneous model is simpler than an inhomogeneous, and a Poisson model simpler that a cluster.
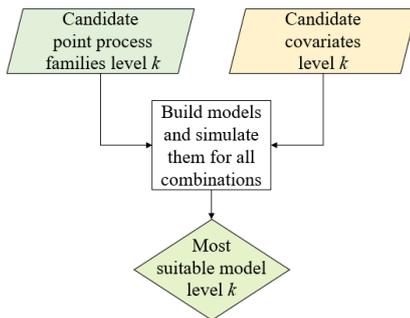

Figure 10: Schematic representation of the selection of optimal point process families + covariates per level

After selecting the best model for every level, another envelope test was done to check whether the chosen models reflected the observed inter-level spatial relationships (Figure 11). The chosen models were implemented in the HMPP framework (similar to Figure 6 – but instead of using all higher level point densities as covariates, only the topographic covariates that were deemed suitable in the last step were used). Simulation envelopes containing the *g* and *L* functions of 39 simulations were created for each possible combination of hierarchical levels (i.e. 4-3, 4-3, 4-1 … 2-1). These simulation envelopes were then compared to the *observed* interlevel relationship (i.e. the *g* and *L* functions of the observed point pattern). Additionally, another set of models was created for every level which were the same as the chosen models (i.e. same family and covariates), except the higher level point intensity-covariates were removed from the model , resulting in a non-hierarchical modelling framework. For each level of the non-hierarchical framework, simulation envelopes were created and again compared to the observed interlevel relationship to test if the hierarchical models more accurately represent the interlevel relationships than the non-hierarchical models.
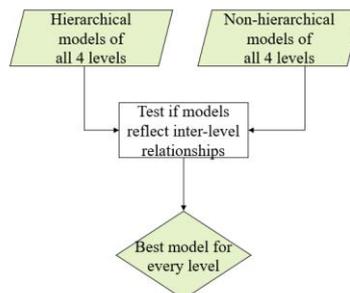

Figure 11: Schematic representation of the test whether the chosen hierarchical model and covariates reflect the observed inter-level spatial relationships more accurately than a set of non-hierarchical models.

3.3 Common plot configurations

NFI plot configurations vary among countries: they vary in number of plots, in shape, size and inter-plot distance. Besides, the shape of the plots themselves, the shape in which the plots are arranged relative to each other is variable too (e.g. L-, cross- shaped or circular). Figure 12 shows a variety of plot configurations among existing NFIs. For this study, plots capturing the most diverse configuration was selected.

The spatial boundary of the simulations of the models within this framework was limited by the extent of the covariate maps acquired for this study (400 x 400 m for the tropical site, and 250 x 550 m for temperate site), as the HMPP frameworks required covariate values as an input. The maximum extent that fitted the covariate maps, both the temperate and the tropical, was a polygon of 400 m x 250 m, from now on referred to as the simulation window (the yellow polygon in Figure 13). Some of the selected plot configurations had to be adapted to fit within the simulation window (i.e. the size of the plots was scaled down to ensure the total plot area equals 0.4 ha). Besides, some plot shapes were simplified as the original NFI sampling strategies were more complex than useful for this research (e.g., they contain nested plots where trees with a DBH < 10 cm were measured, or only trees larger than a certain DBH -Figure 12d).
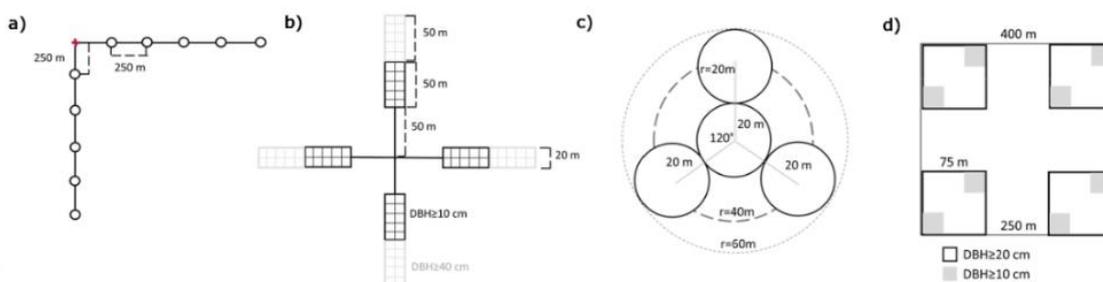
Figure 12: Examples of plot configurations of. a) an L-shaped plot configuration of Peru. b) the cross-shaped plot configuration of Brazil. c) the circular shaped plot configuration of India. d) the square-shaped plot configuration of Democratic Republic of the Congo (Image and text edited from Nesha et al. (2022)*).*

3.4 Accuracy assessment

The first step to finding which plot configuration from question 3 achieves the highest accuracy in estimating the AGB in a RS-based biomass pixel, was to simulate the HMPP frameworks that follow from research question 2. At every hierarchical level in the HMPP framework, the chosen model was used to simulate a tree pattern. Figure 10 shows the multiple simulations of trees (black dots) within a simulation window and a selected plot configuration. The extent of the simulations of the models within this framework was limited by the extent of the covariate maps, as the HMPP frameworks require covariate values as an input. The maximum extent that fit both maps containing covariate information is visualised by the the yellow polygon in Figure 13: Multiple simulations of trees (the black dots) within a simulation window and a plot configuration. The simulation of the HMPP framework was sampled using the plot configurations put forward by research question 2(the red rectangles in Figure 13 are an example). Following a common approach (e.g. Mascaro et al. (2011)), only trees of which the trunk was >50% inside the plot were counted.

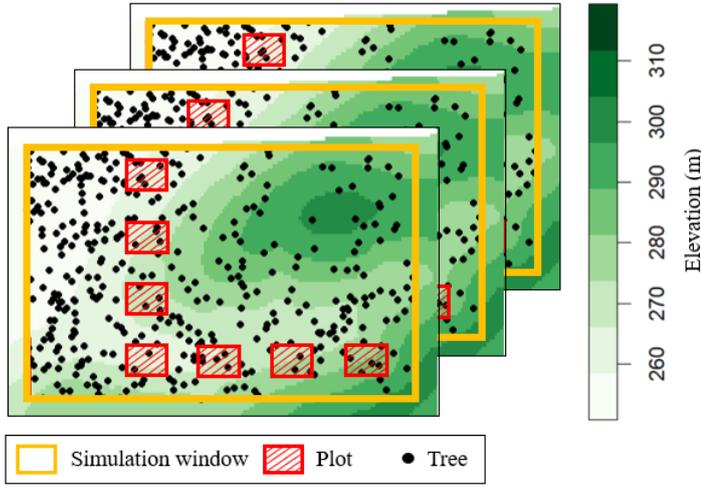*Figure 13: Multiple simulations of trees (the black dots) within a simulation window and a plot configuration.*

As mentioned in the theoretical framework, the models can only simulate a spatial point pattern but no mark representing AGB. Consequently, a tree pattern was simulated for every hierarchical level. Since the research question requires the quantification of the AGB for every simulation, the following method was used to assign AGB values to simulated trees: 1) the AGB values of all *observed* trees within each hierarchical level were estimated using allometric equations, 2) the mean was taken of all estimated AGB values of that hierarchical level, and 3) this mean AGB value was attributed to every *simulated* tree in the corresponding hierarchical level. The allometric equations used for the tropical forest site was an equation presented by Brown et al. (1989) (Equation 4): a generic mixed species tropical allometric equation of for the tropical forest site. A set of species specific allometric equations presented by Jenkins et al. (2003) was used for the temperate forest site, on which besides DBH, also the tree species was noted (Equation 5 shows an example of one of many species-specific equations used).

$$AGB = e^{-2.134 + 2.530*log(DBH)} \qquad (4)$$

$$AGB_{red\ maple\ tree} = e^{-1.9123 + 2.3651*log(DBH)} \qquad (5)$$

For both study sites, a Monte Carlo simulation method was used to identify which plot configurations most accurately represented the AGB within the simulation window. 100 simulations, as recommended by Bonate (2001), of the HMPP framework were done, and each simulation was sampled using all plot configurations. Consecutively, the mean error (ME) and root mean square error (RMSE) of every plot configuration was calculated according to equation 6 and 7 respectively to evaluate their accuracy. In these equations, Y_i represents the simulated mean AGB per hectare of the simulation window in simulation i and is considered the reference AGB while $\hat{Y}_{ij}$ represents the AGB estimated from the plot configuration j, which is obtained by taking the mean AGB per hectare over the plots of plot configuration j for simulation i. The ME is a measure of systematic error or bias and the RMSE includes both the bias and the spread of the differences between plot and simulation window AGB. The closer the ME and the RMSE are to 0, the more accurate the plot configuration.

$$ME_j = \frac{\sum_{i=1}^{n} Y_i - \hat{Y}_{ij}}{n} \qquad (6)$$

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_{ij})^2}{n}} \qquad (7)$$

4. Results

Every subchapter within this chapter presents the results of a research question. Chapter 4.1 shows the results of the tests that were done to select the models that fit the tree data best and presents these models for both study sites. Chapter 4.2 shows which plot configurations were chosen of which the accuracy could be assessed in Chapter 4.3. Chapter 4.3 shows the results of the accuracy tests, and presents the most accurate plot configurations.

4.1 Parameter values HMPP-models

Preliminary analysis

Table 1 shows the hierarchical levels to which the trees at both forest sites were assigned to based on their DBH values. Figure 14 and Figure 16 support the visualisation of the tree location per hierarchical level for the tropical and temperate site, respectively. Figure 15 shows a kernel density estimate of the tropical forest site, and for all levels, it shows that the tree density is high in the diagonal. To identify the best point process family to each the hierarchical levels, the summary functions *g, G* and *L* were used, which are visualised in Figure 17 (step 1b, Figure 9) for the tropics as an example.

Table 1: Details of the hierarchical levels the trees were divided into of the tropical forest (left) and the temperate forest (right).

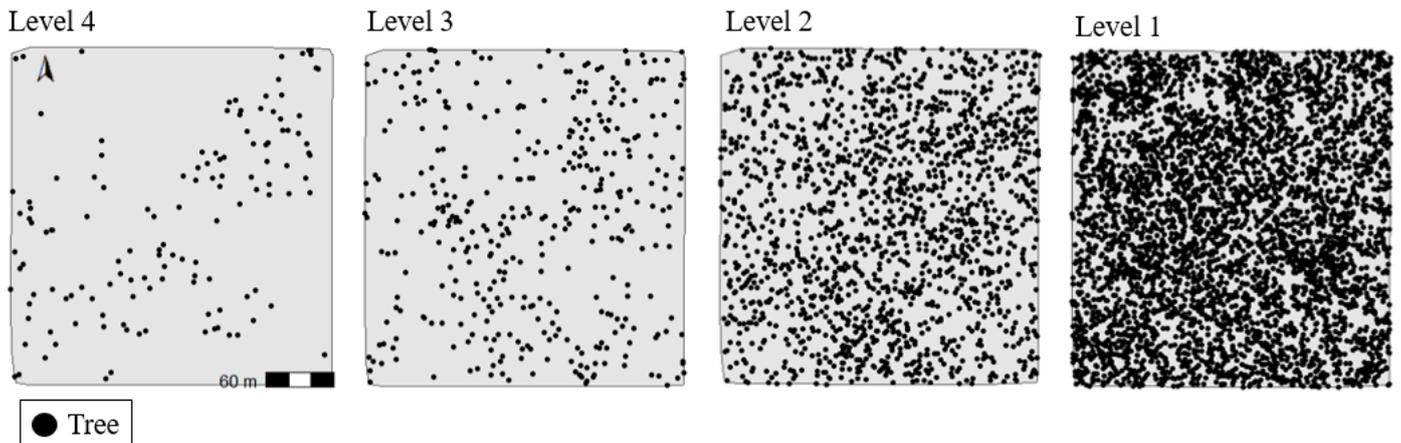| level	Tropical forest		Temperate forest	
	DBH range	No. trees	DBH range	No. trees
1	10.1 - 21.2	4381	16.8 - 29.2	624
2	21.2 - 41.3	1608	29.2 - 37.3	624
3	41.3 - 77.0	378	37.3 - 48.0	415
4	> 77.0	131	> 48.0	188

*Figure 14: Trees in every hierarchical level of the **tropical** forest site.*
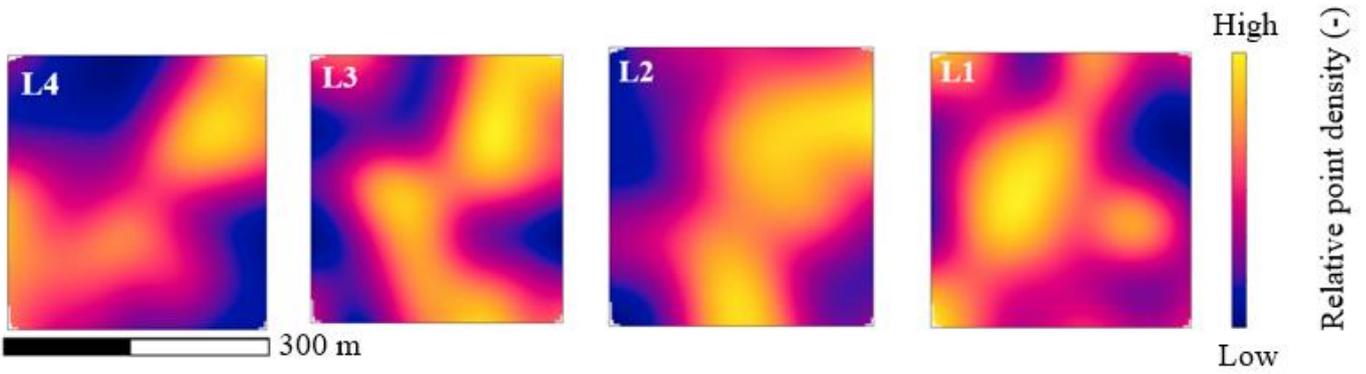
Figure 15: Relative point density (kernel density) of the observed trees for every hierarchical level (L1 to L4) for the tropical forest site.(See Appendix IV for the methodology on how this image was made)
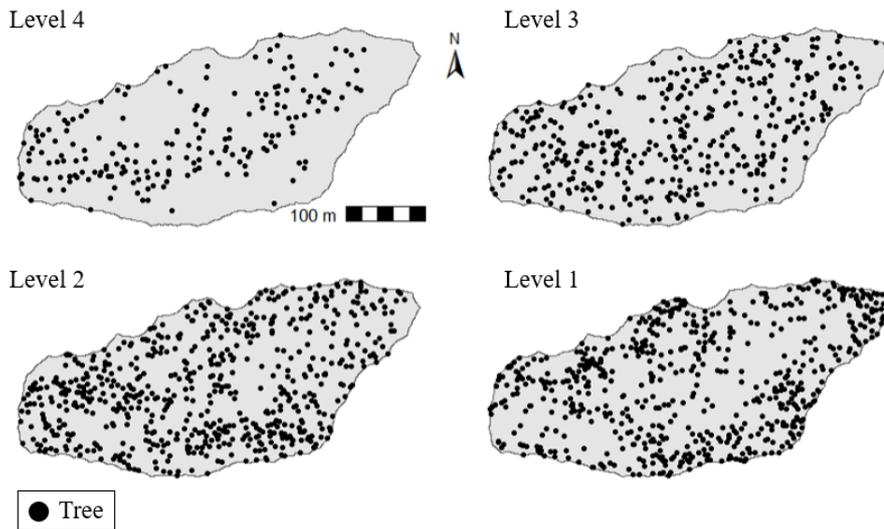
*Figure 16: Trees in every hierarchical level for the **temperate** forest site.*

In Figure 17, the graphs for level 1 and 2 show that the observations (black line) are mainly within the boundaries of the envelope (grey). This indicates that these point patterns show CSR. Therefore, only Poisson process models were used in further model development. However, for level 4, the observation line mostly lies above the simulation envelope of the L-statistic function of a CSR pattern. This indicates that the expected number of points at distance r is higher for the observed pattern than for a Poisson pattern, which indicates clustering. The other summary statistic functions, g and G, also indicate clustering, as the observed lines lie above the simulation envelopes (Baddeley, 2010). Therefore, various versions of cluster models were fitted to the data during the next phase of the model development. The L statistic function for the pattern at level 3 also mostly lied above the simulation envelope. That is why for level 3, both Poisson and various cluster models were used for further model development. The same analysis was done for the temperate forest site. All figures can be found in the files supplied with the thesis report (/output/2_**_model_test).
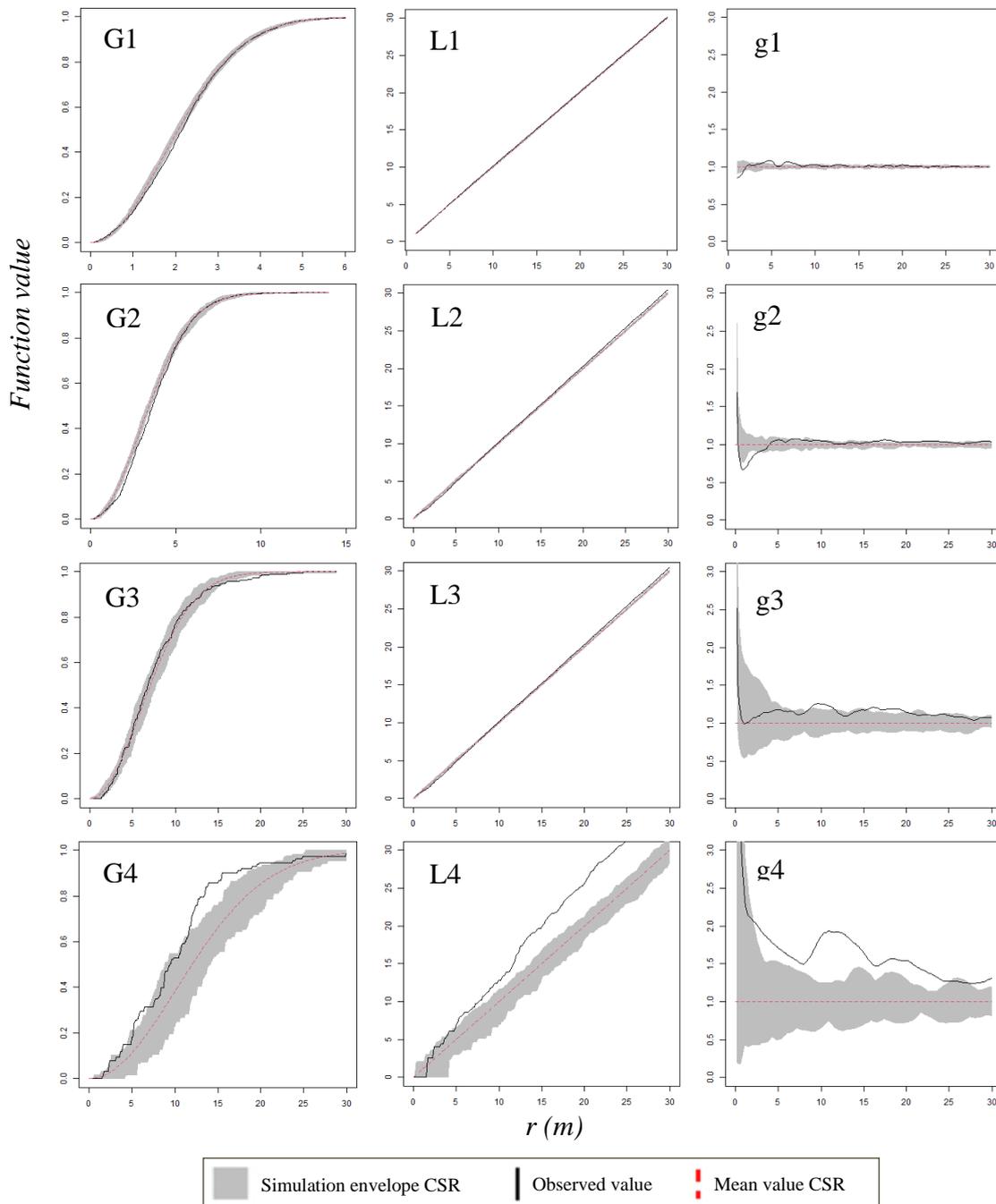
Figure 17: Simulation envelopes of the G, L and g statistic function of 39 Poisson processes (grey), the average of the envelopes (red) and the G, L and g statistic function of the observed point pattern (black). The label of the figure indicate which summary function is used and the numbers indicate the hierarchical levels.

After the preselection of point process families for every level, a pre-selections of model covariates ware made by assessing the relationships between the patterns at each level and the possible model covariates (step 1c, Figure 9). Other than the higher level point intensities, the only covariate datasets that had a resolution high enough, and thus eligible to be used as a covariate, were the digital elevation models of both study sites; with a resolution of 1m x 1m for the tropics, and 3m x 3m for the temperate forest site. Other datasets found, such as soil type and precipitation, did not have the right precision level to be used as model covariates: their resolution was 100m x 100m or coarser.

The relationship between the point pattern of every hierarchical level and covariate was tested for all covariates found in Table 2. The results of the relative distribution estimates of the patterns and the possible covariates of the tropical forest site can be found in Figure 18 as an example. The rest of the results can be found in the files supplementary to this report (/output/ 2_ar_relative_dist_KS_Test). Figure 18a shows the

relationship between the point density at hierarchical level 3 and the point density at hierarchical level 4. It shows a positive relationship, while Figure 18b shows a negative relationship between elevation and tree intensity in level 4. Lastly, Figure 18c shows that a non-monotonic relationship between the presence of level 1 trees and aspect of the terrain: the estimated point intensity is highest when the aspect is π radians (i.e. south facing slopes).

Table 2. Covariates that were used for testing whether they fit the data.

Topographic variables	Higher level point density
Digital elevation model (DEM)	Point density of level n (DENS n)
Slope	Cumulative point density of combinations of higher levels ($DENSn_x$)
Aspect	
Terrain roughness index (TRI)	
Topographic position index (TPI)	
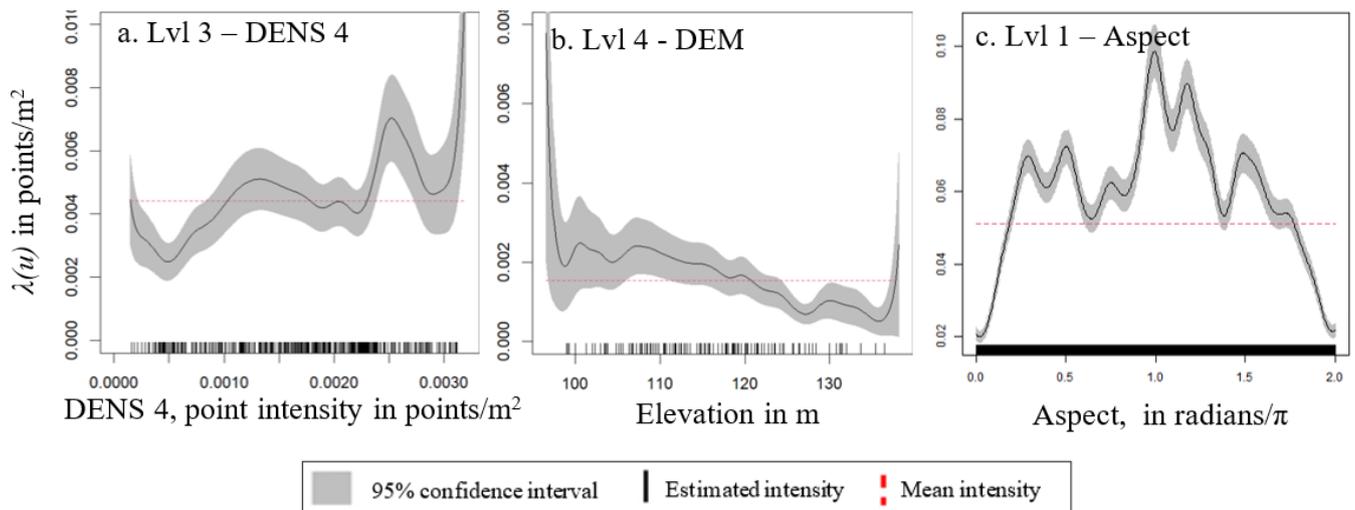
Figure 18: Estimated intensity $\lambda(u)$ $(=f(Z(u)))$, versus candidate covariates. The label of each sub-figure indicates for which hierarchical level the intensity is estimated (Lvl 1,3 or 4) and covariate on which the intensity might depend (DENS 4, DEM, Aspect).

Table 3 shows the *p*-value of the Kolmogorov-Smirnov (KS) test for the tropical forest site, which tests whether the relationship between a pattern and a candidate covariate is non-random. The smaller the *p*-value, the higher the odds that the relationship is non-random. Table 4 shows the results of the KS-test for the temperate forest site.

*Table 3: P-values of the Kolmogorov-Smirnov test of the **tropical** forest site to determine if the relationship between the point patterns at every hierarchical level and the candidate covariates are non-random. ***: p < 0.001, **: p < 0.01, *: p < 0.05, ns: p > 0.05.*

Level	DEM	slope	asp	TPI	TRI	DENS 4	DENS 4+3	DENS 4+3+2	DENS 3	DENS 3+2	DENS 2
1	**	ns	***	***	ns	*	ns	*	*	*	ns
2	ns	ns	***	***	ns	*	***	-	***	-	-
3	ns	ns	***	***	ns	**	-	-	-	-	-
4	***	ns	**	ns	ns	-	-	-	-	-	-

*Table 4: P-values of the Kolmogorov-Smirnov test in the **temperate** forest site to determine if the relationship between the point patterns at every hierarchical level and the candidate covariates are non-random. ***: p < 0.001, **: p < 0.01, *: p < 0.05, ns: p > 0.05.*

Level	DEM	slope	asp	TPI	TRI	DENS 4	DENS 4+3	DENS 4+3+2	DENS 3	DENS 3+2	DENS 2
1	***	ns	***	***	ns	***	***	**	***	ns	***
2	ns	ns	***	ns	ns	ns	ns	-	ns	-	-
3	***	ns	***	ns	ns	***	-	-	-	-	*
4	***	ns	***	ns	ns	-	-	-	-	-	

Based on the results of the envelope test for CSR and the KS-test, the following point process families and covariates (Table 5 and 6) were chosen to be further examined for their fit to paired and compared to see how well they fit the observed data. This was done by means of comparing simulation envelopes of summary functions that were constructed with these models, to the summary functions of the observed point pattern.

*Table 5: List of combination of candidate models and covariates to be compared for the **tropical** forest site.*

Level	Point process family	Combination of covariates
4	Thomas	DEM
	Matérn	DEM
3	Poisson	Aspect, TPI, DENS 4, Aspect + DENS 4, TPI + DENS4
	Thomas	Aspect, TPI, DENS 4, Aspect + DENS 4, TPI + DENS4
	Matérn	Aspect, TPI, DENS 4, Aspect + DENS 4, TPI + DENS4
2	Poisson	Aspect, TPI, DENS4, DENS43, DENS3, + all combinations
	Thomas	Aspect, TPI, DENS4, DENS43, DENS3, + all combinations
	Matérn	Aspect, TPI, DENS4, DENS43, DENS3, + all combinations
1	Poisson	Aspect, TPI, DENS4, DENS432, DENS3, DENS32, DENS2 + all combinations

*Table 6: List of combination of candidate models and covariates to be compared for the **temperate** forest site.*

Level	Model type	Combination of covariates
4	Poisson	DEM
3	Poisson	DEM, DENS4
2	Poisson	DENS4 + DENS3 +TRI, DENS43 + TRI
1	Matérn	DENS4 + DENS3 + DENS 2 +slope, DENS234 + slope

Model Development and Selection

Judging the outcomes of these envelope test using the guidelines for model selection described in the methodology, the most suitable point process families and covariates were selected for the tropical forest site (Figure 19 & Table 7) and for the temperate forest site (Figure 20 & Table 8). Appendix V shows an example of how a model was chosen using the guidelines presented in the methodology.

Table 7 shows that the relationship between the point intensity of level 3 and level 4 trees is positive (positive sign of the variable) for the tropical forest. This indicates that level 3 trees are more likely to be found in the vicinity of level 4 trees. Table 7 also shows that the relationship between the intensity of level 2 trees and the combination of level 3 and 4 trees is negative (negative sign of the variable), which indicates that the odds of finding a level 2 tree are smaller when the density of level 3 and 4 trees is higher.
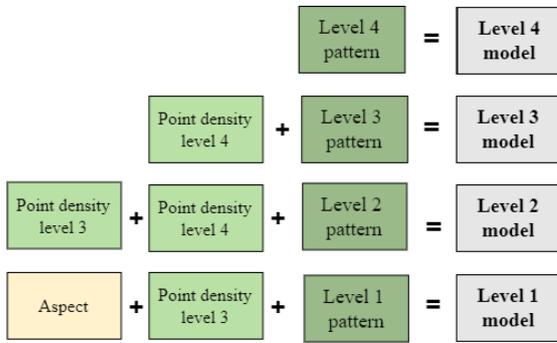
*Figure 19: Selected models and covariates for the **tropical** forest site.*

*Table 7: Selected model and covariates for the **tropical** forest site. The column Model type contains the types of models chosen: they can be either homogeneous or inhomogeneous Matérn indicates a clustered model and PP indicates a Poisson process. In the column model form, λ_p indicates the intensity of the parent process and λ_d indicates the intensity of the daughter process for cluster processes. λ indicates the intensity for Poisson processes.*

Level	Model type	Model form	Radius of daughter cluster
4	Hom. Matérn	$\lambda_p = e^{-8.119}$ $\lambda_d = e^{-6.483}$	39 m
3	Inhom. Matérn	$\lambda_p = e^{-6.070}$ $\lambda_d = e^{-5.783+\,224.867\,DENS\,4}$	28 m
2	Inhom. Matérn	$\lambda_p = e^{-11.091}$ $\lambda_d = e^{-3.863-\,20.343\,DENS\,3,4}$	96 m
1	IPP	$\lambda = e^{-3.031-0.015\,Aspect+\,13.129\,DENS\,3}$	-

Error! Reference source not found. shows that the relationship between point intensity of level 4 trees and the elevation is negative for the temperate forest. This indicates that less trees are found in places with higher elevations and more in places with lower elevations. Besides, it shows that for all levels, a topographical variable influences the tree distribution.
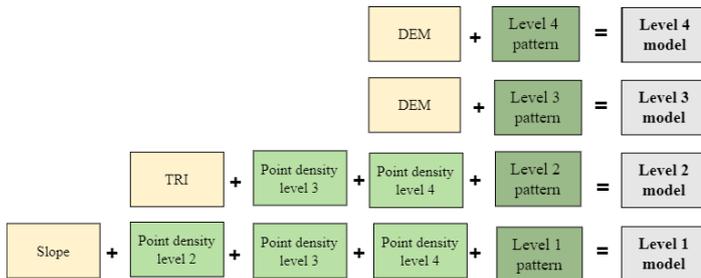

Figure 20: Selected model + covariates for the temperate forest site.

*Table 8: Selected model and covariates for the **temperate** forests site. The column Model type contains the types of models chosen: they can be either homogeneous or inhomogeneous Matérn indicates a clustered model and PP indicates a Poisson process. In the column model form, λ_p indicates the intensity of the parent process and λ_d indicates the intensity of the daughter process for cluster processes. λ indicates the intensity for Poisson processes.*

Level	Model type	Model form
4	IPP	$\lambda = e^{-7.59-0.049\,DEM}$
3	IPP	$\lambda = e^{-1.153-0.015\,DEM}$
2	IPP	$\lambda = e^{-5.650+187.214\,DENS\,3-126.027\,DENS\,4+0.170\,TRI}$
1	Inhom. Matérn	$\lambda_p = e^{-6.536}$ $\lambda_d = e^{-3.951-46.519DENS2,3,4-0.016Slope}$

The last step of the model development, was to check whether the hierarchical models represent the inter-level relationship between the levels better than the non-hierarchical models (the test visualised in Figure 11). Table 9 shows the results: for the tropical forest, the g-function indicates that the hierarchical models represent the inter-level relationships better between level 4-3, 3-2 and 2-1. For the other levels, there is no such indication. For the temperate forest, the L function (Table 9) indicated that the hierarchical models represent the inter-level relationships better only for the level 4-3 and 4-2.

*Table 9: Results of the envelope test to check whether the hierarchical models represented the inter-level relationship better than the non-hierarchical models for the **tropical** forest site (top two figures) and the **temperate** forest site (bottom two figures). "b" means that the hierarchical models represented the inter-level relationship better than the non-hierarchical models (thus the observed summary function was outside the simulation envelope more for the non-hierarchical model then for the hierarchical). "s" indicates that both models represented it similarly (the observed summary functions did not go outside the simulation envelope for either set of models).*

Tropical site

L Level	4	3	2	g Level	4	3	2
3	s	-	-	3	b	-	-
2	s	s	-	2	s	b	-
1	s	s	s	1	s	s	b

Temperate site

L Level	4	3	2	g Level	4	3	2
3	b	-	-	3	s	-	-
2	b	s	-	2	s	s	-
1	s	s	s	1	s	s	s

4.2 Common clustered plot configurations used in NFIs

Figure 21 displays the plot configurations selected for sampling from the model simulations, which are based on existing NFI and research sampling strategies. All of the selected plot configurations, except for the large 1 ha square Figure 21a, cover an area of 0.4 ha (Table 10). The study first evaluated the performance of plot configurations from the study countries, such as the US NFI sampling strategy shown in Figure 21 c (Tomppo

et al., 2010) and a simplified and scaled-down version of the FAO's widely used strategy in the Philippines (Figure 21f) (FAO, 2007). To diversify the inter-plot distance, shape, and number of plots, configurations d and e were added, respectively being a simplified version of Brazil's plot configuration and a simplified and scaled-down version of Peru's plot configuration (Nesha et al., 2022). Figure 21a, the 1 ha square, is a commonly used plot in tropical forestry (FAO, 1998). To facilitate comparison of the large square plot with the other plot configurations, plot b, with an area of 0.4 ha, was added.
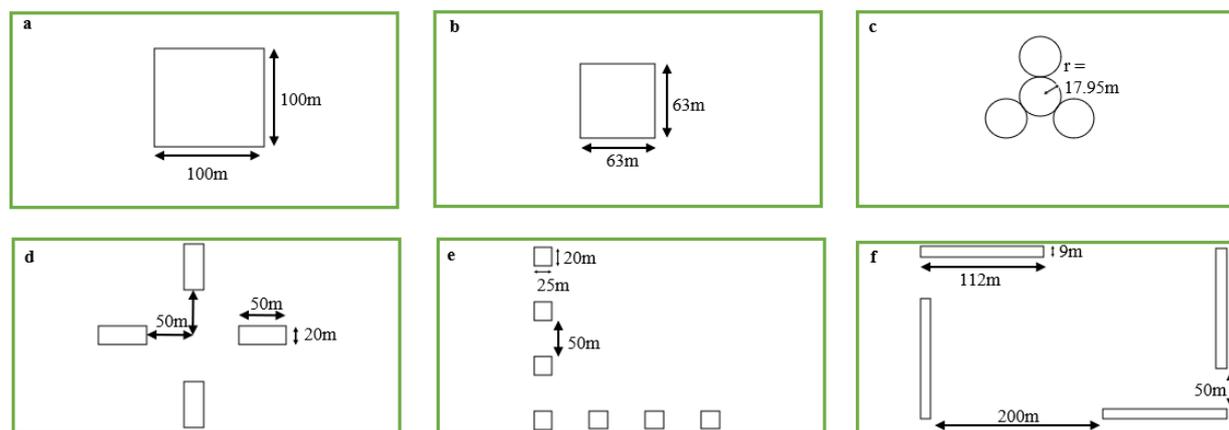

Figure 21. Plot configurations that were used for sampling in this research. a) a 1 ha square plot, b) a 0.4 ha square plot, c) plot configuration of the United States' NFI. b) simplified version of the plot configuration of Brazil's NFI e) plot configuration based on Peru's NFI, f) plot configuration based on the Philippine NFI.

Table 10: Characteristics of the sampling designs that were used to sample simulated forest plots.

ID	configuration	Plot shape	Plot size (ha)	No. Plots	Total area of plots (ha)	Avg distance between plots (m)	Total edge length (m)	Area-to-edge ratio
a	-	Square	1	1	1	-	400	25.0
b	-	Square	0.4	1	0.4	-	252	15.9
c	Y-shape	Circle	0.1	4	0.4	-	450	8.9
d	Cross	Rectangle	0.1	4	0.4	50	560	7.1
e	L-shape	Rectangle	0.057	7	0.4	50	672	6.0
f	Rectangle	Rectangle	0.1	4	0.4	125	968	4.0

4.3 Achieved accuracies

For the tropical forest site, the AGB of the trees that were observed on the study site was 527 Mg/ha, while over 100 simulations, the mean simulated AGB was only 421 Mg/ha. For the temperate forest, the AGB of the trees that were observed on the study site was 179 Mg/ha, while the mean AGB over 100 simulations was 211 Mg/ha. Figure 22 shows examples of simulations for the tropical forest site. Seemingly, all six of them have fewer points along the top- and the right side edges than in the rest of the simulation window.
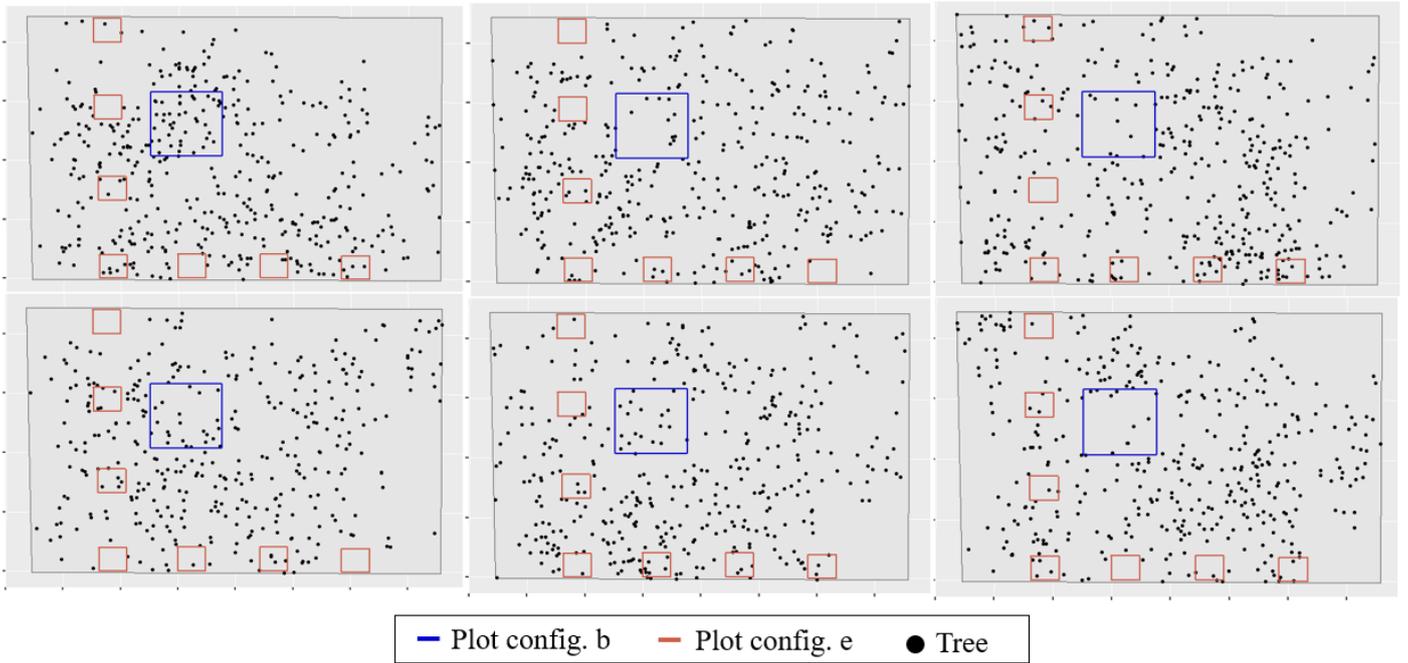
*Figure 22: 6 simulations of trees (level 3 and 4) on the **tropical** forest site. The blue square is plot configuration b. The red rectangles are plot configuration e.*

Table 11 shows the ME and the RMSE of every plot configuration in the tropical forest. Configuration e, the L-shaped plot configuration that has most plots, 7, resulted in the smallest ME and RMSE. The difference between the errors amongst the various plot configurations is large compared to the errors found for the temperate forest (Table 12). In the temperate forest, most plot configurations performed similarly in estimating the AGB within the sampling window. The 0.4 ha square plot, resulted in smaller errors than sampling the 1 ha square plot. Sampling with plot configuration e again results in the lowest bias, however, its RMSE is similar to the one of the other plots. Plot configuration c performed worst in the temperate forest.

*Table 11: Mean error and root mean square error of the various plot configurations of the plots in the **tropical** forest site in Mg/ha. In the row containing the ME values, a negative sign indicates overestimation of AGB ($Y_i < \hat{Y}_i$) and a positive sign indicates underestimation of AGB ($Y_i > \hat{Y}_i$).*

	a	b	c	d	e	f
ME (Mg/ha)	-65.5	-92.7	-96.4	-45.4	-11.0	+89.1
RMSE(Mg/ha)	105.4	143.1	134.6	104.3	82.3	105.5

*Table 12: Mean error and root mean square error of the various plot configurations of the plots in the **temperate** forest site in Mg/ha. In the row containing the ME values, a negative sign indicates overestimation of AGB ($Y_i < \hat{Y}_i$) and positive sign indicates underestimation of AGB ($Y_i > \hat{Y}_i$).*

	a	b	c	d	e	f
ME (Mg/ha)	+23.1	+15.1	+38.7	+7.9	-1.7	+19.0
RMSE (Mg/ha)	23.8	21.5	41.2	20.8	22.7	26.8

Figure 23 displays the variety of elevation that was found within plot configurations a, b and c: the small square covers more of the lower elevation values while the circular plots cover more of the higher elevation values. This is also confirmed by Figure 24, which shows the elevation values of the pixels within the plot configurations. It shows that the elevation distribution and mean elevation within plot b is most similar

to the elevation window and that the elevation distribution and mean elevation within plot c is the least similar to the simulation window.
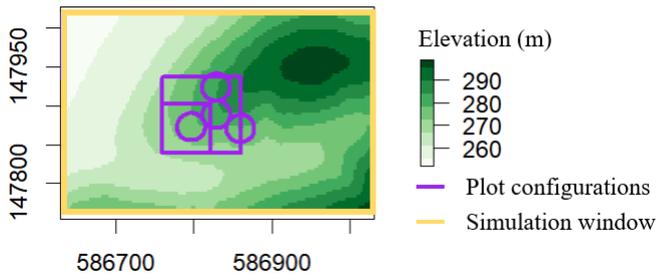
*Figure 23: Elevation map of the **temperate** forest site, overlayed with the simulation window (in blue), the 1ha square sampling plot (ID: a in Table 10), the 1 ha square sampling plot (ID: b) and plot configuration c.*
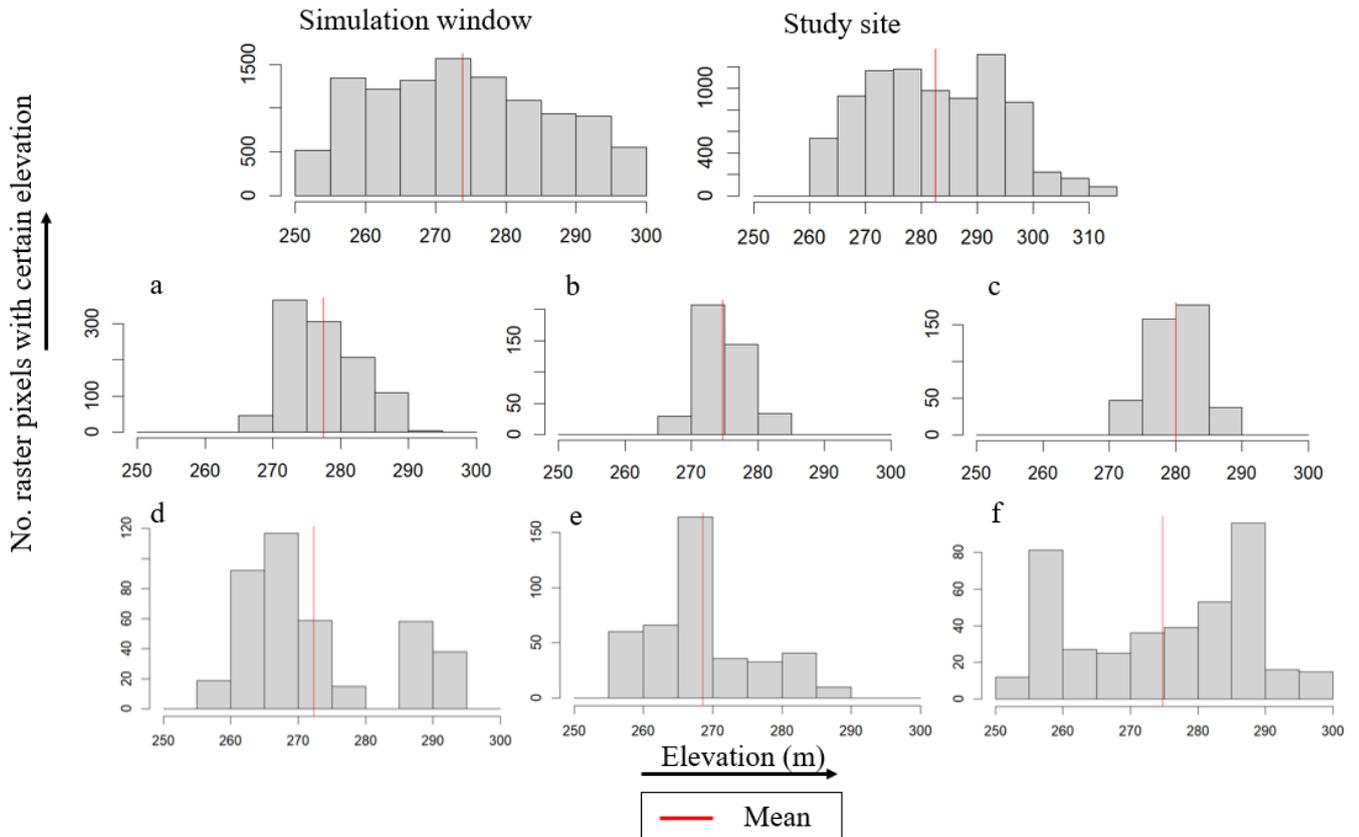
Figure 24: Distribution of the elevation within the sampling plots. The means of the elevation within the simulation window is 273.8, 282.5 for the study site, 277.4 for plot configuration a, 274.7 for plot configuration b and 280.0 for c, 272.2 for d, 268.6 for e and 274.7 for f.

5. Discussion

In this section, the obtained results are interpreted and compared to scientific literature. Besides, the limitations of the study are discussed. The first section discusses the results of the model identification for the forests separately. Subsequently, the accuracy of the plot configurations is discussed in the second section, also first for one forest site, and then for the other.

5.1 HMPP-models identification

Tropical forest site

With regards to the selected covariates of the models, topography was expected to influence the AGB distribution, according to the findings of Lister and Leites (2018). Nevertheless, in the current study the topography does not seem to influence the tree distribution much: a topographical covariate was only used to model level 1 trees (Table 7). However, the tree data was wrongly geo-referenced: appendix II describes how the all trees were referenced relative to the lower left and upper right corner point. Conversely, the trees should have been referenced relative to the lower right corner and the upper left corner. This mistake caused the tree data to be mirrored. As a result, the tests done to check the relationships between elevation and tree location might have wrongly indicated that there was no relationship at most hierarchical levels. Nevertheless, topography not significantly influencing the tree distribution is not unheard of: Baraloto et al. (2011) and Guitet et al. (2015) found that topographical factors only weakly influence the AGB in tropical forests while the structure of the forest stand is much more influential. Besides, even if the trees would have been georeferenced correctly, and the topography would have influenced the tree distribution, the parsimony principle (used for model selection) can cause a covariate not to be taken up into the model, even if it could influence the tree distribution. This is illustrated by Figure 30 in Appendix I. It shows how the two models, a Matérn cluster including the DEM as covariate, and the model that was eventually chosen: a Matérn cluster not including the DEM as covariate. Even though the envelope test in Figure 30 shows that the models perform similarly, the parsimony principle implied that the model that did *not* include the topographical variable DEM was chosen. Although the topography could have been of influence to some extent, it was not taken into account in these models due to the guidelines that were used for choosing the models.

Besides topography, also the tree density of higher hierarchical levels was used as model covariates. For example, a negative relationship between the density of level 2 and level 3+ 4 trees (Table 7) was found. This indicates that level 2 trees are less likely to be found in the vicinity of level 3 and 4 trees. This is in line with scientific literature about tree interaction, which states that trees tend to compete for space and are therefore expected to be regularly distributed rather than clustered (Stoyan & Penttinen, 2000). Conversely, the positive relationship between level 3 and 4 trees is positive (Table 7) – which indicates that level 3 trees are more likely to be found in the vicinity of other large trees – is opposite to this literature on tree interaction. Furthermore, all higher levels point patterns show clustering (Table 7), which is also in contrast with this literature on tree interaction. The clustering of trees indicates positive spatial autocorrelation of AGB of which the range is different for the various levels (Table 7- cluster radius): the level 3 trees show positive spatial autocorrelation until 38 meters, while the level 2 trees show spatial autocorrelation up until 100 meters. This is different to what Saatchi et al. (2011) found: no spatial autocorrelation after 11 m.

Both the clustering of trees within each level, and the positive relationships between level 3 and 4 trees that were discussed in the previous paragraph, are thus not easily explained by literature on tree interactions. However, there are some phenomena that could explain the unexpected tree patterns. First of all, there could be environmental factors (i.e. covariates) of influence on the tree distribution that were not considered. In fact, due to the mirroring of the tree data, the statistical methods probably missed the correlation of the points with the topographic variables, causing the bigger trees being close together being misinterpreted as clustering. A second phenomenon that could explain the tree patterns that are left unexplained by literature on tree interactions, is cyclonic wind storms (Hogan et al., 2018). The study site is hit with 13-15 typhoons on average per decade (Cinco et al., 2016). Sandra et al. (2016) found that typhoons in the Palanan (the region of the study site) have led to high tree mortality among large diameter trees. Thus perhaps the larger trees at the site were regularly distributed once, but the cyclones have taken out the trees at some locations, leaving only patches of trees.

The observed AGB (i.e. the AGB calculated from the DBH of the trees that were measured on site) in the tropical forest was 527 Mg/ha, while the mean AGB over 100 simulations of the model was only 421 Mg/ha. This indicates that too few trees were simulated, which is confirmed by Figure 22. As it turns out, there is an error in the code that was used to simulate the trees, which stops level 1, 2 and 3 trees from being simulated over the entire simulation window.

Temperate forest site

When it comes to the model parameters of the temperate forest, they are close to expectations as they are almost similar to the models chosen by Lister and Leites (2018). The largest difference, however, is the level 1 model Table 8. shows the combined density of level 2, 3 and 4 trees (DENS234) is a model covariate of the level 1 model and that the relationship between level 1 trees and the combined tree density of level 2, 3 and 4 is a negative relationship. This indicates that level 1 trees are less likely to be found when the combined density of level 2, 3 and 4 trees is high. This is different to what Lister and Leites (2018) found: instead of the combined density of level 2, 3 and 4 trees (DENS234), they used the separate tree density maps as covariates (DENS2 + DENS 3+ DENS4). This resulted in a positive relationship between level 1 and the density of level 2 trees, and negative relationships between level 1 and the densities of level 3 and 4 trees. Not all of the relationships are negative, which indicates that by merging the tree densities of different levels, information on forest structure was lost. The model chosen by Lister and Leites (2018) thus represented the forest structure better than the model that was chosen in this study. However, the guidelines that were presented in the methodology for choosing the best models for every level, dictated that if two models performed similarly, the simples model should be chosen. This was the case in this study for level 1 model: the model including the combined densities performed similarly to the model including the separate density. Consequently, as including a single covariate (DENS234) is simples than including three covariates (DENS2 + DENS 3+ DENS4), the model using DENS234 was chosen, despite it being less representative for the forest structure.

The models of the temperate forest systematically overestimated the AGB within the simulation window: the AGB of the observed trees on site was 179 Mg/ha while the mean AGB over 100 simulations was 211 Mg/ha. This overestimation of AGB could have been caused by a difference in elevation of the study area and the simulation window, which can be seen in Figure 24: the mean elevation at the study site is 282.5 m, while the mean elevation in the simulation window is only 273.8 m. Because the level 3 and 4 models negatively depend on elevation (Table 8) (which means the higher the elevation, the less trees) the tree density in the simulation window was higher than the tree density on the study site and therefore, and therefore, so was the AGB. A second reason why the models could wrongly estimate the AGB, is a lack of the model fit. The goodness-of-fit of all models was only tested through second order methods (envelope test of the G-, g- and L-function). However, Baddeley & Silverman (1984) stress that the second order methods (the summary statistics used in this study) only show the summary functions of the underlying model and cannot detect differences in point patterns of different models with identical summary statistics. Therefore, even though the models chosen in this study were the best fit, there are differences in point patterns that summary functions cannot capture. This could mean that the spatial structure of the simulated points may match the required summary function values while the number of trees per level and thus the overall intensity might still differ from the observed values.

5.2 Accuracy Assessment

Tropical forest site

Table 11 indicates which plot configurations most accurately estimate the AGB over the simulation window. The L-shaped plot configuration (e) had the lowest ME and RMSE and therefore, most accurately estimated the AGB, which was as expected. This can be illustrated by Figure 22, which shows high spatial variability of the AGB: in some places, there are clusters with many points, while some places of the map are almost empty. If many plots are spread out over the sampling unit, they have higher chance of capturing this variability of AGB and therefore give more accurately represent the mean AGB within the simulation window. Conversely, if only a single plot is used, the chances are that it is either contains an entire cluster and overestimate the mean AGB of the window, or that it barely contains any points as many areas of the map are quite empty, and

underestimates the mean AGB of the sampling window. This is also in line with what Yim et al. (2015), who found that spread out plots should be able tree data (i.e. tree location and their AGB).

The results presented in Table 11, however, are unreliable, due to a phenomenon mentioned earlier: Figure 22 shows that fewer points were simulated along the edges of the simulation window, which caused the 1 ha plot (a) (and all the other plots in the middle (b & c)) to systematically overestimate the AGB. Conversely, configuration f places most subplots along the edges, which is why this plot configuration systematically underestimated the AGB within the simulation window.

Temperate forest site

For the temperate forest site, the 1 ha square plot (Figure 21a), was expected to perform better than the 0.4 ha square plot (Figure 21b). Zolkos et al. (2013) and Mascaro et al. (2011) both compared AGB estimates based on plot measurements to AGB estimates based on remote sensing measurements, and found that the error in AGB estimates tends to decline when the area-to-edge ratio declines. Therefore, as the area-to-edge ratio was larger for plot configuration a than for b, plot configuration a was expected to more accurately estimate the AGB. However, plot configuration a did not perform better than b: Table 12. The ME and the RMSE of plot configuration b, the 0.4 ha plot, is smaller than the RMSE and ME of plot configuration a, the 1 ha plot. Another unexpected outcome, is that plot configuration c, which was used for forest inventories in the US, least accurately estimates the AGB. These unexpected outcomes can be explained by the plots being fixed at the same geographical location for all simulations. The plots being in the same location for every simulation caused a bias as the models that were used for simulation used topographic variables that are also spatially fixed (Figure 23). Table 12 table shows that, for example, for the level 3 and level 4 models, the point intensity of these levels have a negative relationship with elevation. Therefore, over 100 simulation, more trees were simulated in places with high elevation than in places with low elevation. Because the plots configurations were in the same geographic location every simulation, and the topographical variables too, more trees were systematically simulated in some plots configurations than in others.

Consequently, the distribution of elevation values within the pixels the plot configuration influences the size of the bias. Figure 24 shows the distribution of elevation values of the pixels within the plot configurations and the simulation window. Because elevation influences the AGB, the plot configuration that has elevation values more similar to the simulation window, more accurately estimates the AGB. This is shown in Figure 24: the distribution of elevation values of plot configuration b are more similar to the elevation values of the simulation window than the values of plot configuration a & c and therefore, plot configurations b has a smaller ME.

On average, the ME is smaller for the plot configurations that have distance between the plots (d-f), compared to the plot configurations that have no distance between the plots (a-c)(Table 12). However, out of the plot configurations that have distance between the plots (d-f), plot configuration f has the largest bias, and underestimates the AGB most. This could have been caused by relatively high amount, a peak, of high elevation values in plot f (Figure 24f) compared to the simulation window and the other plot configurations with space between the plots. The relationship between point intensity (at level 3 and 4) and elevation is exponential (Table 8), which means a lot of high elevation values have a relatively large influence on the point intensity, and therefore, the AGB. Therefore, the peak in high elevation values in plot f could have caused a the higher ME.

6. Conclusion and Recommendations

The goal of this research was to study the impact of spatial configuration of forest inventory plots for integration with RS-based biomass products through modelling. First, the a conclusion to each research question is given. Subsequently, some recommendations are made.

The parameters for the models of the tropical forest site indicated that the largest trees in the study area were clustered and that the distribution of most trees did not depend on topographic variables. This can be explained by the mistake that was made when geo-referencing the data. Perhaps there was a relationship between the tree locations and topographic variables but owing to wrong geo-referencing, the distribution of the trees was misinterpreted as clustering, rather than it being influenced by topography. A second reason for a clustered distribution of the larger trees could be typhoons that hit the area regularly, and cause high mortality in trees with large diameters. The model parameters for the temperate forest site showed that the distribution of trees greatly depends on topography.

The second research question – What are common, distinct plot configurations used for forest inventories? – yielded six plot configurations and does not require a further conclusion.

Which plot configuration achieves the highest accuracy stays unanswered for both study sites. For the tropical forest site, an error was made in the execution of the simulation, which caused fewer points to be simulated near the edges. This caused a bias for almost all of the plot configurations and therefore no conclusion about the most accurate plot configuration can be made. For the temperate forest site, the topography caused a bias, as in this site the AGB depends on the topography. Because of model decisions, configurations systematically sampled locations with higher elevations, leading to systematic underestimation of AGB. Therefore, no valid conclusion about the accuracy of plot configurations can be drawn.

The first recommendation for better model parameterisation to correctly geo-reference the data and re-investigate the relationship between the tree locations and the topographic variables. Besides, an additional guideline could be applied when selecting the best model for every level. When the guidelines for choosing a model presented by Lister and Leites (2018) were followed, the chosen model for the smallest trees (level 1) of the temperate forest site, used a single covariate to describe the interaction with all higher hierarchical levels of trees (DENS234), while biologically, it could easily be that the relationship between level 1 and 2 trees is positive, while the relationship between level 1 and 4 trees is negative. Therefore, between the second and third guideline, an additional guideline could be given: pick the model that represents proven ecological relationships best.

To gain insight into which plot configuration most accurately estimates the AGB within a larger area, the plot configurations should be randomly relocated, to avoid bias caused by topographical (or other) variables on which the models are based.

References

Afshang, M., Saha, C., & Dhillon, H. S. (2017). Nearest-Neighbor and Contact Distance Distributions for Thomas Cluster Process. *IEEE Wireless Communications Letters*, *6*(1), 130–133. https://doi.org/10.1109/LWC.2016.2641935

Araza, A., de Bruin, S., Herold, M., Quegan, S., Labriere, N., Rodriguez-Veiga, P., Avitabile, V., Santoro, M., Mitchard, E. T. A., Ryan, C. M., Phillips, O. L., Willcock, S., Verbeeck, H., Carreiras, J., Hein, L., Schelhaas, M. J., Pacheco-Pascagaza, A. M., da Conceição Bispo, P., Laurin, G. V., … Lucas, R. (2022). A comprehensive framework for assessing the accuracy and uncertainty of global above-ground biomass maps. Remote Sensing of Environment, 272. https://doi.org/10.1016/j.rse.2022.112917

Baddeley, A. (2010). Analysing Spatial Point Patterns in R. Version 4.1. Perth: CSIRO and University of Western Australia.

Baddeley, A., Diggle, P. J., Hardegen, A., Lawrence, T., Milne, R. K., & Nair, G. (2014). On tests of spatial pattern based on simulation envelopes. In *Source: Ecological Monographs* (Vol. 84, Issue 3).

Baddeley, A. J., & Silverman, B. W. (1984). *A Cautionary Example on the Use of Second-Order Methods for Analyzing Point Patterns* (Vol. 40, Issue 4). https://www.jstor.org/stable/2531159

Baddeley, A., & Turner, R. (2000). PRACTICAL MAXIMUM PSEUDOLIKELIHOOD FOR SPATIAL POINT PATTERNS (with Discussion). In *Z. J. Statist* (Vol. 42, Issue 3).

A. Baddeley and R. Turner. (2005). Spatstat: an R package for analyzing spatial point patterns. Journal of Statistical Software, 12(6):1–42. URL: www.jstatsoft.org, ISSN: 1548- 7660.

Baraloto, C., Rabaud, S., Molto, Q., Blanc, L., Fortunel, C., Hérault, B., Dávila, N., Mesones, I., Rios, M., Valderrama, E., & Fine, P. V. A. (2011). Disentangling stand and environmental correlates of aboveground biomass in Amazonian forests. *Global Change Biology*, *17*(8), 2677–2688. https://doi.org/10.1111/j.1365-2486.2011.02432.x

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*. https://doi.org/10.1038/sdata.2018.214

Bonate, P. L. (2001). A Brief Introduction to Monte Carlo Simulation. In *Clin Pharmacokinet* (Vol. 40, Issue 1).

Bormann, F. H., & Likens, G. E. (1979). Pattern and Process in a Forested Ecosystem. In *Pattern and Process in a Forested Ecosystem*. Springer New York. https://doi.org/10.1007/978-1-4612-6232-9

Brown, S., & Iverson, L. R. (1992). BIOMASS ESTIMATES FOR TROPICAL FORESTS. In *World Resource Review* (Vol. 4, Issue 3).

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., Duque, A., Eid, T., Fearnside, P. M., Goodman, R. C., Henry, M., Martínez-Yrízar, A., Mugasha, W. A., Muller-Landau, H. C., Mencuccini, M., Nelson, B. W., Ngomanda, A., Nogueira, E. M., Ortiz-Malavassi, E., … Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, *20*(10), 3177–3190. https://doi.org/10.1111/gcb.12629

Cinco, T. A., De Guzman, R. G., Ortiz, A., Delfino, R. J. P., Lasco, R. D., Hilario, F. D., Juanillo, E. L., Barba, R., & Ares, E. D. (2016). Observed trends and impacts of tropical cyclones in the Philippines. *International Journal of Climatology*, 36(14), 4638–4650. https://doi.org/10.1002/joc.4659

density function - RDocumentation. (z.d.). consulted on 22/03/2023. https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/density

Diggle, P. J., & Gratton, R. J. (1984). Monte Carlo Methods of Inference for Implicit Statistical Models. In *Source: Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 46, Issue 2).

Fabris-Rotelli, I., & Stein, A. (2018). Inhomogeneous spatial modelling of DPT pulses for marine images. *Spatial Statistics*, *28*, 257–270. https://doi.org/10.1016/j.spasta.2018.08.004

FAO. (1998). *Guidelines for the management of tropical forest 1*. The Production of Wood. FAO Forestry Paper 135. Rome, Italy: Forest Resource Division FAO Forestry Department.

FAO. (2007). *BRIEF ON NATIONAL FOREST INVENTORY, NFI PHILIPPINES*. www.fao.org/forestry

Gibbs, D., Harris, N., & Seymour, F. (2018, October 4). *By the Numbers: The Value of Tropical Forests in the Climate Change Equation*.

Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamäki, J. v, Smith, P., Woodbury, P., Zganjar, C., Blackman, A., Campari, J., Conant, R. T., Delgado, C., Elias, P., Gopalakrishna, T., Hamsik, M. R., … Leavitt, S. M. (2017). Natural climate solutions. *Source*, *114*(44), 11645–11650. https://doi.org/10.2307/26488842

Guitet, S., Hérault, B., Molto, Q., Brunaux, O., & Couteron, P. (2015). Spatial structure of above-ground biomass limits accuracy of carbon mapping in rainforest but large scale forest inventories can help to overcome. *PLoS ONE*, *10*(9). https://doi.org/10.1371/journal.pone.0138456

Handavu, F., Syampungani, S., Sileshi, G. W., & Chirwa, P. W. C. (2021). Aboveground and belowground tree biomass and carbon stocks in the miombo woodlands of the Copperbelt in Zambia. *Carbon Management*, *12*(3), 307–321. https://doi.org/10.1080/17583004.2021.1926330

Herold, M., Carter, S., Avitabile, V., Espejo, A. B., Jonckheere, I., Lucas, R., McRoberts, R. E., Næsset, E., Nightingale, J., Petersen, R., Reiche, J., Romijn, E., Rosenqvist, A., Rozendaal, D. M. A., Seifert, F. M., Sanz, M. J., & de Sy, V. (2019). The Role and Need for Space-Based Forest Biomass-Related Measurements in Environmental Management and Policy. In *Surveys in Geophysics* (Vol. 40, Issue 4, pp. 757–778). Springer Netherlands. https://doi.org/10.1007/s10712-019-09510-6

Hogan, J. M., Zimmerman, J. K., Thompson, J., Uriarte, M., Swenson, N. G., Condit, R., Hubbell, S. P., Johnson, D., Sun, I., Chang-Yang, C., Su, S., Ong, P. S., Rodriguez, L. J., Monoy, C. C., Yap, S. L., & Davies, S. J. (2018). The Frequency of Cyclonic Wind Storms Shapes Tropical Forest Dynamism and Functional Trait Dispersion. Forests, 9(7), 404. https://doi.org/10.3390/f9070404

Jenkins, J. C., Chojnacky, D. C., Heath, L. S., & Birdsey, R. A. (n.d.). *National-Scale Biomass Estimators for United States Tree Species*.

Lister, A. J., & Leites, L. P. (2018). Modeling and simulation of tree spatial patterns in an oak-hickory forest with a modular, hierarchical spatial point process framework. *Ecological Modelling*, *378*, 37–45. https://doi.org/10.1016/j.ecolmodel.2018.03.012

Málaga, N., de Bruin, S., McRoberts, R. E., Arana Olivos, A., de la Cruz Paiva, R., Durán Montesinos, P., Requena Suarez, D., & Herold, M. (2022). Precision of subnational forest AGB estimates within the Peruvian Amazonia using a global biomass map. *International Journal of Applied Earth Observation and Geoinformation*, *115*. https://doi.org/10.1016/j.jag.2022.103102

Mascaro, J., Detto, M., Asner, G. P., & Muller-Landau, H. C. (2011). Evaluating uncertainty in mapping forest carbon with airborne LiDAR. *Remote Sensing of Environment*, *115*(12), 3770–3774. https://doi.org/10.1016/j.rse.2011.07.019

Molto, Q., Hérault, B., Boreux, J.-J., Daullet, M., Rousteau, A., & Rossi, V. (2013). Preticting troical tree height Climate of the Past Discussions Earth System Dynamics Earth System Dynamics Discussions Geoscientific Instrumentation Methods and Data Systems Geoscientific Instrumentation Methods and Data Systems Geoscientific Model Development Predicting tree heights for biomass estimates in tropical forests Preticting troical tree height Preticting troical tree height. *Biogeosciences Discuss*, *10*, 8611–8635. https://doi.org/10.5194/bgd-10-8611-2013

Moorcroft, P. R., Hurtt, G. C., & Pacala, S. W. (2001). A method for scaling vegetation dynamics: The ecosystem demography model (ED). *Ecological Monographs*, *71*(4), 557–586. https://doi.org/10.1890/0012-9615(2001)071[0557:AMFSVD]2.0.CO;2

Næsset, E., Bollandsås, O. M., Gobakken, T., Solberg, S., & McRoberts, R. E. (2022). Exploring characteristics of national forest inventories for integration with global space-based forest biomass data. *Science of the Total Environment*, *850*. https://doi.org/10.1016/j.scitotenv.2022.157788

Næsset, E., McRoberts, R. E., Pekkarinen, A., Saatchi, S., Santoro, M., Trier, Ø. D., Zahabu, E., & Gobakken, T. (2020). Use of local and global maps of forest canopy height and aboveground biomass to enhance local estimates of biomass in miombo woodlands in Tanzania. *International Journal of Applied Earth Observation and Geoinformation*, *89*. https://doi.org/10.1016/j.jag.2020.102109

Packalen, P., Strunk, J., Maltamo, M., & Myllymäki, M. (2022). Circular or square plots in ALS-based forest inventories—does it matter? *Forestry: An International Journal of Forest Research*. https://doi.org/10.1093/forestry/cpac032

Réjou-Méchain, M., Muller-Landau, H. C., Detto, M., Thomas, S. C., le Toan, T., Saatchi, S. S., Barreto-Silva, J. S., Bourg, N. A., Bunyavejchewin, S., Butt, N., Brockelman, W. Y., Cao, M., Cárdenas, D., Chiang, J. M., Chuyong, G. B., Clay, K., Condit, R., Dattaraja, H. S., Davies, S. J., … Chave, J. (2014). Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. *Biogeosciences*, *11*(23), 6827–6840. https://doi.org/10.5194/bg-11-6827-2014

Saatchi, S., Marlier, M., Chazdon, R. L., Clark, D. B., & Russell, A. E. (2011). Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. *Remote Sensing of Environment*, *115*(11), 2836–2849. https://doi.org/10.1016/j.rse.2010.07.015

Stoyan, D., & Penttinen, A. (2000). Recent Applications of Point Process Methods in Forestry Statistics. In *Statistical Science* (Vol. 15, Issue 1).

Strebe, D. R. (2011). Winkel triple projection SW.jpg, Wikipedia. Consulted on 15-03-2022. https://commons.wikimedia.org/wiki/File:Winkel_triple_projection_SW.jpg

Tomppo, E., Kuusinen, N., Mäkisara, K., Katila, M., & McRoberts, R. E. (2017). Effects of field plot configurations on the uncertainties of ALS-assisted forest resource estimates. *Scandinavian Journal of Forest Research*, *32*(6), 488–500. https://doi.org/10.1080/02827581.2016.1259425

UN Climate Change. (2022, September 17). *Introduction to Mitigation*.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, *63*(1), 252–258. https://doi.org/10.1111/j.1541-0420.2006.00667.x

Yim, J. S., Shin, M. Y., Son, Y., & Kleinn, C. (2015). Cluster plot optimization for a large area forest resource inventory in Korea. *Forest Science and Technology*, *11*(3), 139–146. https://doi.org/10.1080/21580103.2014.968222

Appendix I

This appendix shows how to write the parameter values given by the "summary" function in the form of equation 1 or 2, when a point process model is fitted to a point pattern (using either PPM or KPPM of Spatstat).

$$\lambda(u) = e^{\theta_0 + \theta_n Z_n(u)} \tag{1}$$

$$\lambda(u) = \lambda_p \lambda_d \tag{2}$$

PPM

Homogeneous Poisson Process
When no trend is entered into the PPM function, the function will fit a homogenous Poisson process to the point pattern and the intensity will be a single number. This number is given by "uniform intensity" Figure 25. To translate this parameter value in the form of formula 1, the natural logarithm of this number is taken to obtain θ_0. $\theta_n Z_n(u)$ is equal to 0, as there is no covariate influencing the point distribution. Alternatively, θ_0 can read under "estimate"

```
---------------------------------------------------------------------------
FITTED MODEL:

Stationary Poisson process

---- Intensity: ----


Uniform intensity:
[1] 0.00152903

            Estimate      S.E.    CI95.lo   CI95.hi Ztest      Zval
log(lambda) -6.483122 0.08737041 -6.654364 -6.311879  *** -74.20272
```

Figure 25: Output of the summary function when a homogeneous Poisson process is fitted to point data.

Inomogeneous Poisson Process
When a trend is entered into the PPM function (the DEM "shdem" in Figure 26) , the function will fit a inhomogeneous Poisson process to the point pattern and the intensity function will look like equation 1. The fitted trend coefficients are θ_0 & θ_n. For θ_0 , the number found under *intercept (*Figure 26) can be filled out and θ_1 is the number below "shdem" (-0.0488943). (Baddeley, 2010, p.p. 97)

```
---------------------------------------------------------------------
FITTED MODEL:

Nonstationary Poisson process

---- Intensity: ----

Log intensity: ~shdem
Model depends on external covariate 'shdem'
Covariates provided:
        shdem: im

Fitted trend coefficients:
(Intercept)        shdem
 7.59147690 -0.04888943

             Estimate       S.E.     CI95.lo     CI95.hi Ztest
(Intercept) 7.59147690 1.913511668  3.84106295 11.34189085   ***
shdem       -0.04888943 0.006921962 -0.06245622 -0.03532263   ***
```

Figure 26: Output of the summary function when an inhomogeneous Poisson process is fitted to point data.

KPPM

Inhomogeneous Cluster process

The intensity function of a cluster process consists of 2 parts: the homogeneous intensity of the parent process and the inhomogeneous intensity of the daughter process (equation 2). The intensity of the parent process is both homogeneous and Poisson, and therefore, a single number: kappa (κ) (Baddeley, 2010), highlighted in yellow in Figure 27. The intensity of the daughter cluster is an inhomogeneous Poisson process and the intensity function therefore has the form of equation 1. In the case of Figure 27, two model covariates were used: shdem (a DEM) and dens4 (a density map of trees). θ_0 is again the intercept. θ_1 is the "Esitmate" (Figure 27) besides "shdem" and θ_1 is the "Esitmate" besides "dens4" (Baddeley, 2010, p.p. 153).

```
----------- CLUSTER MODEL -----------
Model: Thomas process

Fitted cluster parameters:
      kappa        scale
 0.003398586 22.970912940
Mean cluster size:  [pixel image]

Final standard error and CI
(allowing for correlation of cluster process):
             Estimate        S.E.       CI95.lo       CI95.hi Ztest
(Intercept) -6.584351086    3.43918724 -13.32503422   0.15633205
shdem         0.002751998   0.01129496  -0.01938571   0.02488971
dens4       217.028097788 118.12453870 -14.49174376 448.54793934
             Zval
(Intercept) -1.9145079
shdem        0.2436484
dens4        1.8372821
```

Figure 27:Output of the summary function when an inhomogeneous Thomas cluster process is fitted to point data.

Appendix II

All data can be found in the "data" folder, that is supplementary to this thesis. For the tropical forest site, was provided by Arnan Araza. It consisted of an excel file containing the tree data (location + DBH) and a digital elevation model. The data for the temperate forest site, was provided by Andy Lister, and consists of an excel file, containing the tree data, and a digital elevation model. The plot configurations were created in ArcGIS and can be found in the "data" folder too.

Appendix III

This appendix explains how the dataset of the tropical forest site was georeferenced. The position information of the trees were as follows in the original dataset: they ranged between 0-400 in the X direction and between 0-400 in the Y direction. The only other information given were the corner points of the plot:

Table 13: Corner points of the tropical forest site.

Corner	Longitude	Latitude
Corner 1 (0-0)	122°23'14.3"	17°02'19.9"
Corner 2	122°23'13.5"	17°02'32.4"
Corner 3	122°23'01.0"	17°02'19.7"
Corner 4	122°23'00.5"	17°02'31.8"

The coordinates of the dataset (0-400; 0-400) indicate that the plot rectangular, as the X and Y axis are perpendicular Figure 28a. However, if lines are drawn between the corner points that are given in Table 13, a parallelogram will show (Figure 28b). For the purpose of simplification, it is assumed that the plot is a perfect rectangle so a simple rotation will suffice to geo-reference the data. Because of the position of the rectangle shape, the X and Y coordinates do not only have to be translated in the X and Y direction, they also have to be rotated. This rotation was done using an R script called ar_0_rotate_data_points and can be found in the files supplementary to this thesis. The points are rotated around the centre of the corner points (200, 200) by α degrees (Figure 28). Consecutively, the points are linearly translated in the X and Y direction. The excel file where this is done can be found in the files supplementary to this thesis and is called "./data/Arnan/Palanan_rotated_shifted_sub.csv". The translation in X direction was done according to equation 8, where X_{orig} are the coordinates of the rotated points, X_1 is the X-value of corner 1 (Figure 28b): 122°23'13.5". The slope is defined by equation 9, where X_4 is the X-value of corner 4 (122°23'00.5"). The translations in the Y directions were done using the same formula's, only replacing every X with a Y.

$$X_{translated} = X_{orig} * slope + X_1 \qquad (8)$$

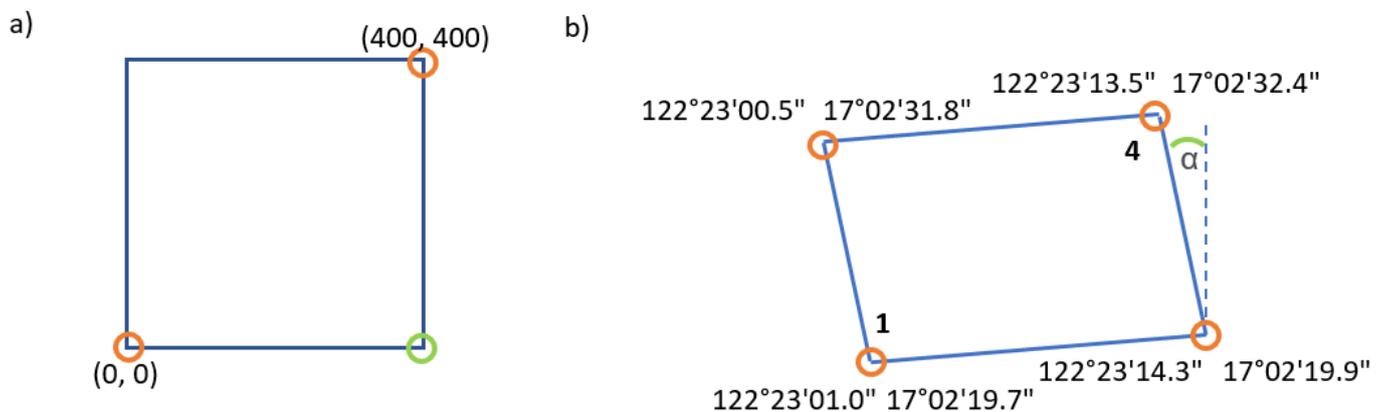$$slope = \frac{X_4 - X_1}{400} \qquad (9)$$

Figure 28: a) the tropical forest site as a square plot with coordinates (0-400; 0-400), b) the tropical forest site as a diamond shaped plot with the coordinates of the corner points as presented in table 13.

Appendix IV

The images of the kernel density were made using the function `density` in R. The kernel density is a probability density function of the data in question and the plots in the figures, the heat maps, are meant to visualise the underlying distribution of the data. (density function - RDocumentation, z.d.)

7. Appendix V

Figure 29 shows two examples of the summary statistic functions for level 3. The top row show the summary statistic functions an IPP, using level 4 density as a covariate and the bottom row shows the summary statistic functions of a Matérn cluster model, also using the level 4 density as a covariate. From these two, the Matérn cluster model was chosen, based on the observed values of the *L* and *g* functions lying fully within the simulation envelope for the Matérn cluster model (

Figure 29e & f), while for the IPP, the *L* functions lies almost fully outside the simulation envelope and the *g* function leaves the envelope twice.
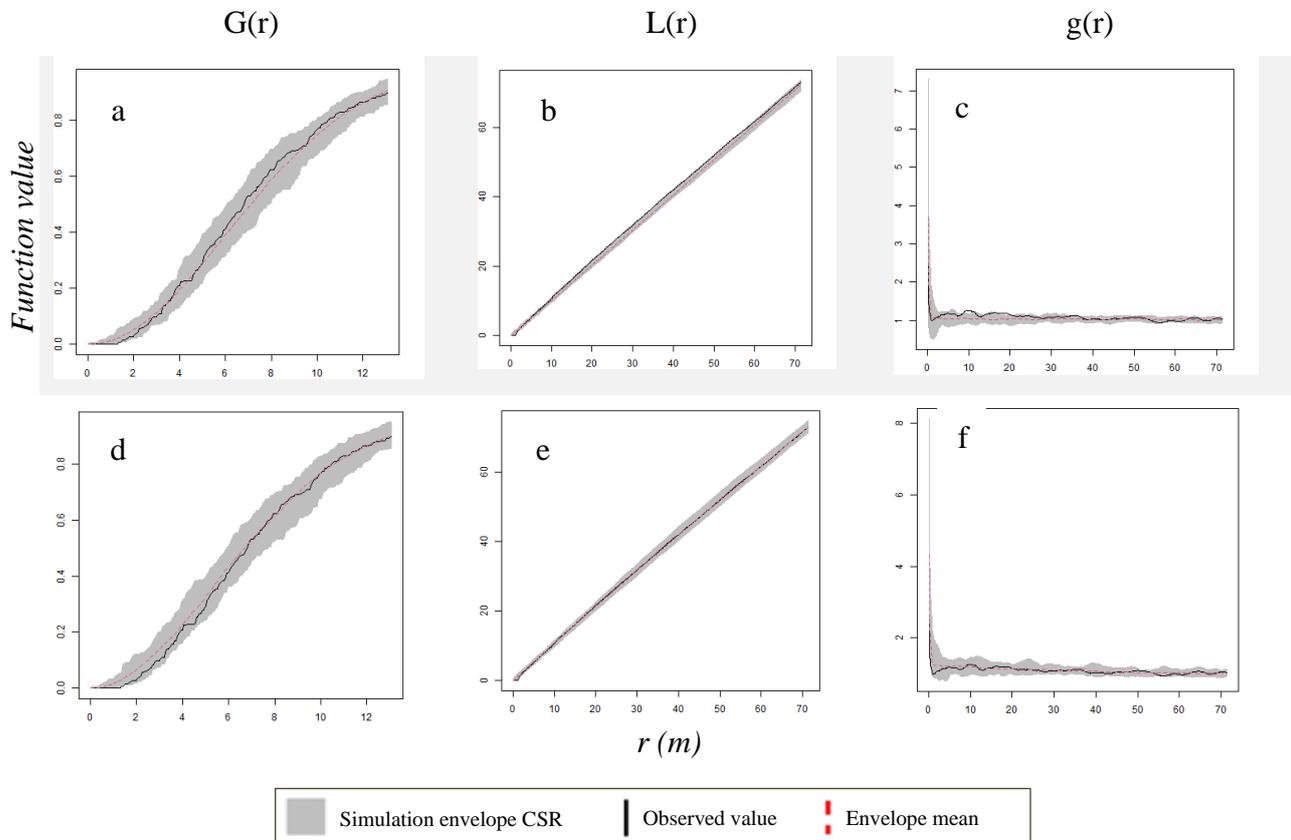

Figure 29: Example of the summary statistic functions for level 3. Figures a-c show the G, L and g function for an inhomogeneous poisson process using the higher level point intensity (level4) as a covariate. Figure d-f show the these summary functions for the chosen model: a Matérn cluster model, also using the higher level point intensity (level4) as a covariate.

Figure 30 shows an example of the application of the second guideline: choose the simpler models. In the top three figures, the summary functions of a cluster model including a DEM is shown. The summary functions of the same model are shown, only not including a DEM. Because the models perform similarly, the simples model, therefore, the model not including the DEM is chosen.
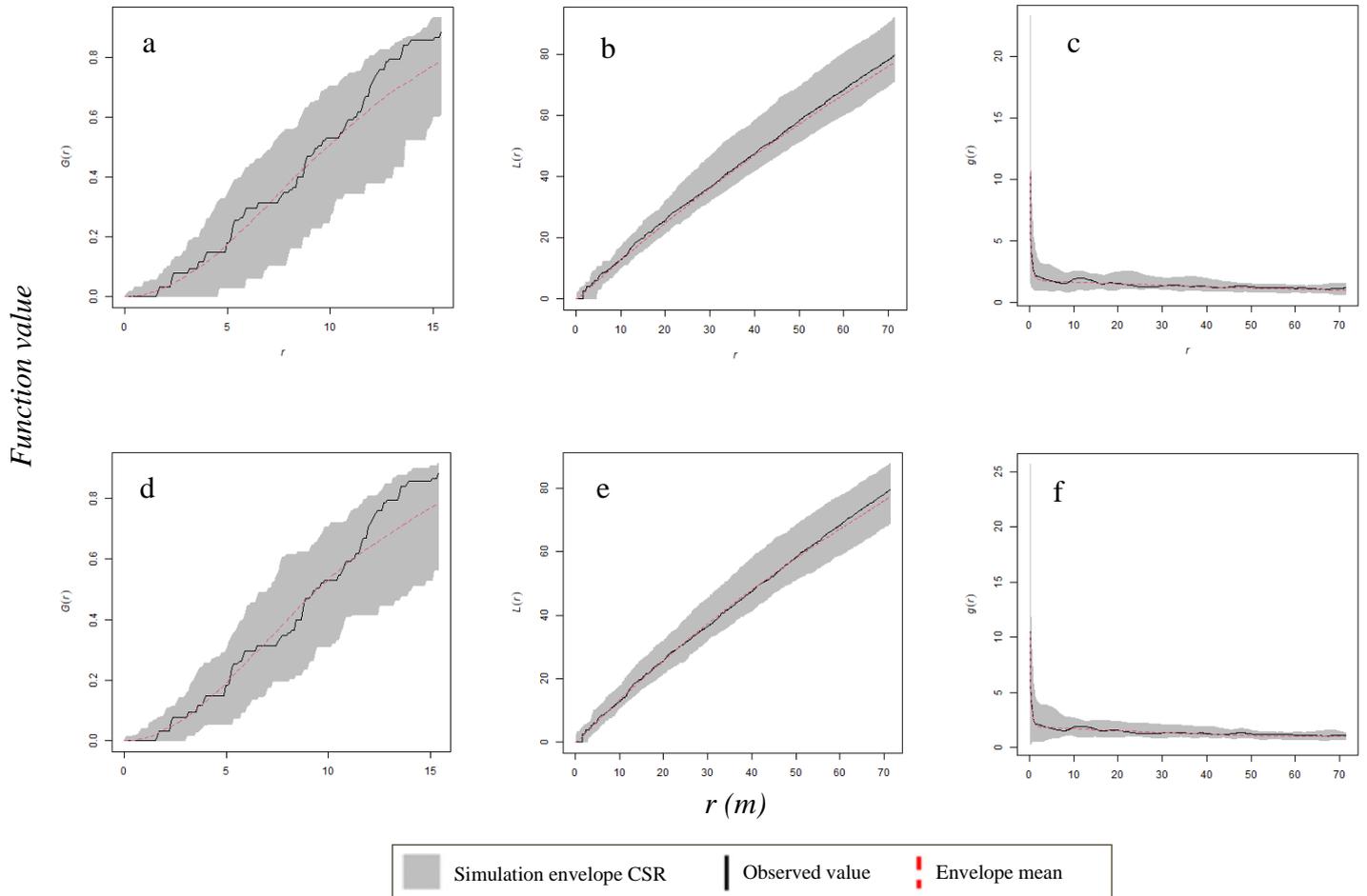
G(r)   L(r)   g(r)

Figure 30: Example of the summary statistic functions for level 4. Figure a-c show the G, L and g function for an inhomogeneous Matern Cluster not including the DEM. Figure d-f show the these summary functions for the chosen model: a Matérn cluster model, not including the DEM.