



inbo



Instituut voor
Natuur- en Bosonderzoek

Opmaak van een analysestramien voor de gegevens van de Vlaamse Bosinventarisatie

*Toon Westra, Pieter Verschelde, Hans Van Calster, Els Lommelen,
Thierry Onkelinx, Paul Quataert en Leen Govaere*

Auteurs:

Toon Westra, Pieter Verschelde, Hans Van Calster, Els Lommelen, Thierry Onkelinx, Paul Quataert en Leen Govaere
Instituut voor Natuur- en Bosonderzoek

Het Instituut voor Natuur- en Bosonderzoek (INBO) is het Vlaams onderzoeks- en kenniscentrum voor natuur en het duurzame beheer en gebruik ervan. Het INBO verricht onderzoek en levert kennis aan al wie het beleid voorbereidt, uitvoert of erin geïnteresseerd is.

Vestiging:

INBO Brussel
Kliniekstraat 25, 1070 Brussel
www.inbo.be

e-mail:

Toon.Westra@inbo.be

Wijze van citeren:

Westra T, Verschelde P, Van Calster H, Lommelen E, Onkelinx T, Quataert P en Govaere L (2015). Opmaak van een analysestramien voor de gegevens van de Vlaamse Bosinventarisatie. Rapporten van het Instituut voor Natuur- en Bosonderzoek 2015 (INBO.R.2015.9034827). Instituut voor Natuur- en Bosonderzoek, Brussel.

D/2015/3241/225

INBO.R.2015.9034827

ISSN: 1782-9054

Verantwoordelijke uitgever:

Jurgen Tack

Druk:

Managementondersteunende Diensten van de Vlaamse overheid

Foto cover:

Hallerbos

Dit onderzoek werd uitgevoerd in opdracht van:

het Agentschap voor Natuur en Bos, ANB/KOBE2013/003



Agentschap voor
Natuur en Bos

Opmaak van een analysestramien voor de gegevens van de Vlaamse Bosinventarisatie

**Toon Westra, Pieter Verschelde, Hans Van Calster, Els
Lommelen, Thierry Onkelinx, Paul Quataert, Leen Govaere**

Dankwoord

Onze dank gaat uit naar de opdrachtgever, het Agentschap voor Natuur en Bos, en leidend ambtenaar Leen Govaere voor de goede samenwerking.

Verder wensen we de leden van de stuurgroep te bedanken: Anja Leyman, Leen Govaere, Martine Waterinckx, Bart Roelandt, Paul Quataert, Luc De Keersmaeker en Pieter Verschelde.

Samenvatting

De Vlaamse bosinventarisatie (VBI) is een beleidsondersteunend meetnet om de toestand en evolutie van bos in Vlaanderen op te volgen. Het Agentschap voor Natuur en Bos (ANB) coördineert het meetnet en staat in voor de gegevensverwerking. De eerste inventarisatieronde van de VBI (VBI1) werd uitgevoerd in de periode 1997 – 1999. De tweede inventarisatieronde (VBI2) is van start gegaan in 2009. Vanaf VBI2 wordt er gewerkt met een 10-jarige cyclus, waarbij elk jaar 1/10^{de} van de gegevens wordt ingezameld. Momenteel zit VBI2 dus voorbij halfweg in de tienjarige cyclus. ANB wenst op basis van de reeds verzamelde gegevens een eerste tussentijdse technische rapportage uit te voeren. Daarna wenst ANB deze technische rapportage op periodieke basis te herhalen naarmate er meer gegevens binnenkomen.

Het doel van dit rapport is de ontwikkeling van een stramien om de technische rapportage op een wetenschappelijk onderbouwde manier en met gepaste analysetechnieken uit te kunnen voeren op alle beschikbare gegevens. De rapportage zelf maakt geen deel uit van deze opdracht. Een aandachtspunt is dat de methodiek van de gegevensinzameling in VBI1 en VBI2 verschillend is voor bepaalde onderdelen. Hier moet voldoende rekening gehouden worden bij de berekeningen van de resultaten, zodat een vergelijking mogelijk is tussen VBI1 en VBI2.

Eerst worden de meetvragen van de VBI opgefrist en verduidelijkt waar nodig. Vervolgens worden de meetvragen ingedeeld in groepen op basis van het type gegevens dat nodig is om de meetvraag te beantwoorden. Binnen elke groep zullen de analyses gelijkaardig verlopen. Per groep is er voor een meetvraag een voorbeeldanalyse uitgewerkt via een R-script.

Het analyseproces bestaat uit volgende stappen: (1) een validatieproces op basis van gekoppelde bomen, (2) opmaak van een analysedatabank met analysevariabelen en variabelen die het meetproces beschrijven, (3) selectie van een analyseset uit de analysedatabank in functie van een bepaalde meetvraag, (4) de eigenlijke analyse voor het beantwoorden van een meetvraag en (5) het wegschrijven van de resultaten in een resultatendatabank.

English abstract

The forest inventory of Flanders (VBI) provides information on status and trends of forest and supports forest policy in Flanders. VBI is managed by the Nature and Forest Agency (ANB). ANB is also responsible for analyzing the results of VBI. The first inventory took place during the years 1997 – 1999. The second inventory started in 2009 and is planned to be completed in 2019. ANB will perform a technical reporting based on the data collected until know.

The purpose of this report is the development of a semi-automated analysis process to support the technical reporting of results using appropriate analyzing techniques. The reporting itself is not part of the assignment.

First, we look in detail at the monitoring targets of VBI and define them more precisely if needed. Next, we group monitoring targets according the type of data which is needed to analyze them. For each group we develop a detailed analysis for a specific monitoring target in R. This analysis can be used as an example for the other monitoring targets in the group.

The analysis process consists of following steps: (1) a validation process based on trees measured in both inventories, (2) construction of a database with variables derived from the raw data and variables describing the measurement process, (3) selection of the set of variables which is needed to analyze a specific monitoring target, (4) the actual analysis to answer the questions related to the monitoring target and (5) saving analysis results in a database.

Inhoudstafel

Dankwoord	4
Samenvatting.....	5
English abstract	6
Lijst van figuren.....	10
Lijst van tabellen	10
1 Inleiding.....	11
2 Verduidelijking van de meetvragen	12
2.1 Toestand en evolutie van de karakteristieken van het bosareaal.....	12
2.1.1 Overzicht van de meetvragen	12
2.1.2 Definities/ onduidelijkheden/scherpstelling	13
2.2 Toestand en evolutie van de boomsoortensamenstelling.....	14
2.2.1 Overzicht van de meetvragen	14
2.2.2 Definities/ onduidelijkheden/scherpstelling	14
2.3 Toestand en evolutie van de bestandsopbouw	15
2.3.1 Overzicht van de meetvragen	15
2.3.2 Definities/ onduidelijkheden/scherpstelling	15
2.4 Toestand en evolutie van de indicatoren voor biodiversiteit.....	15
2.4.1 Diversiteit aan houtachtige en kruidachtige vegetatie (voor de volledige doelpopulatie en per bostypegroep)	15
2.4.1.1 Overzicht van de meetvragen	15
2.4.1.2 Definities/ onduidelijkheden/scherpstelling	16
2.4.2 Natuurlijkheid van de houtachtige en kruidachtige vegetatie (voor de volledige doelpopulatie en per bostypegroep).....	16
2.4.2.1 Overzicht van de meetvragen.....	16
2.4.2.2 Definities/ onduidelijkheden/scherpstelling	16
2.4.3 Structuurrijkdom van de bestanden (voor de volledige doelpopulatie en per bostypegroep).....	17
2.4.3.1 Overzicht van de meetvragen	17
2.4.3.2 Definities/ onduidelijkheden/scherpstelling	17
2.4.4 Aanwezigheid van structurele elementen die biodiversiteit bevorderen (voor de volledige doelpopulatie en per bostypegroep)	17
2.4.4.1 Overzicht van de meetvragen	17
2.4.4.2 Definities/ onduidelijkheden/scherpstelling	17
2.4.5 (Sub)Scores van de Authenticiteitsindex (voor de volledige doelpopulatie en per bostypegroep)	18
2.4.5.1 Overzicht van de meetvragen	18
2.4.5.2 Definities/ onduidelijkheden/scherpstelling	18
2.5 Toestand en evolutie van de samenstelling van de bosvegetatie onder invloed van milieuveranderingen	18
2.5.1 Overzicht van de meetvragen	18
2.5.2 Definities/ onduidelijkheden/scherpstelling	18
2.6 Toestand en evolutie van het duurzaam bosbeheer en -gebruik	19
2.6.1 Economisch facet van duurzaam bosbeheer (voor de volledige doelpopulatie en in functie van lidmaatschap bosgroep)	19
2.6.1.1 Overzicht van de meetvragen	19
2.6.1.2 Definities/ onduidelijkheden/scherpstelling	19
2.6.2 Houtkwaliteit van de economisch belangrijkste boomsoorten	19
2.6.2.1 Overzicht van de meetvragen	19
2.6.2.2 Definities/ onduidelijkheden/scherpstelling	20

2.6.3	Milieubeschermend facet van duurzaam bosbeheer (voor de volledige doelpopulatie en in functie van lidmaatschap bosgroep)	20
2.6.3.1	Overzicht van de meetvragen	20
2.6.4	Improved Pan-European Indicators for Sustainable Forest Management (enkel indicatoren die we geheel of gedeeltelijk kunnen beantwoorden met VBI).....	21
2.6.4.1	Overzicht van de meetvragen	21
2.6.4.2	Definities/ onduidelijkheden/scherpstelling	21
3	Indeling van meetvragen/analysevariabelen	23
4	Onderdelen van het analyseproces	25
5	Kwaliteitscontrole van de ruwe meetgegevens op basis van gekoppelde bomen	26
5.1	Opsporen en beoordelen van anomalieën	26
5.2	Validatieproces.....	26
6	Analysedatabanken	28
6.1	Algemeen	28
6.2	Opmaak van de analysedatabank	28
6.3	Meetproces VBI1 en VBI2	29
6.4	Variabelen m.b.t. het meetproces.....	30
6.4.1	Opnamedatum en koppeling van plots.....	30
6.4.2	Type gegevens per plot en onderscheid tussen nulwaarnemingen en ontbrekende waarden	31
6.4.3	Plotgewichten en oppervlaktes van de A2-, A3- en A4-plots.....	33
6.4.4	Verdeling van de individuele bomen over verschillende plotsegmenten.....	34
6.5	De strata en de verklarende variabelen	34
6.5.1	Statische strata	34
6.5.2	Dynamische strata.....	34
6.6	Analysevariabelen	35
6.6.1	Bosoppervlakte	35
6.6.2	Bestandskarakteristieken kwalitatief	36
6.6.3	Bestandskarakteristieken kwantitatief	37
6.6.4	Vegetatie.....	39
6.6.5	Ruimtelijke structuur	42
6.6.6	Dood hout	44
6.6.7	Authenticiteitsindex	45
6.6.7.1	Bosstructuur	46
6.6.7.2	Houtige vegetatie	48
6.6.7.3	Kruidlaag.....	51
6.6.7.4	Dood hout	53
6.6.8	Aanwas	56
6.6.8.1	Aanwasberekening.....	56
6.6.8.2	Analysedatabank	60
6.6.8.3	Bespreking resultaten	60
6.6.9	Houtkwaliteit	61
6.6.10	Boomsoortensamenstelling	62
6.6.11	Vegetatiesamenstelling	63
7	Gegevensanalyse voor het beantwoorden van meetvragen	65
7.1	Overzicht analysescripts.....	65
7.2	Verschillende analysestappen.....	66
7.2.1	Vorbereiding	66
7.2.2	Selectie analyseset	67
7.2.3	Dataverkenning.....	67
7.2.4	Design-based analyse	71

7.2.5	Model-based analyse met mixed models.....	74
7.2.5.1	Achtergrond.....	74
7.2.5.2	Voorbeelden	74
7.2.5.2.1	Mixed model met continue variabele	74
7.2.5.2.2	Mixed-model met binaire variabele	75
7.2.5.2.3	Mixed-model met aantallen.....	77
7.2.5.2.4	Verschil tussen strata met mixed models.....	77
7.2.6	Volledig uitgewerkt voorbeeld: analyse van de houtvoorraad	78
7.2.7	Aandachtspunten bij de analyse	78
8	Opslag van de analyseresultaten	81
9	Overzicht R-scripts en databanken	83
9.1	Data	83
9.2	Output	84
9.3	Scripts	84
	Referenties	87
	Bijlage 1: Indeling van de meetvragen	89
	Bijlage 2: voorbeeld output van geautomatiseerde analyse houtvoorraad	94
	Inleiding 94	
	Gemiddelde houtvoorraad.....	94
	Inleiding 94	
	Invoer analysetabel.....	94
	Aanmaken analyseset.....	99
	Data verkenning	100
	Dotplots 100	
	Onderzoeken van aantal nullen.....	102
	Histogrammen.....	103
	Density Plots	105
	Relaties met covariaten	106
	Boxplots 106	
	Boxplots op basis van segmenten	107
	Grafieken op basis van plots (steekproefcirkels).....	109
	Evolutie houtvoorraad als regressie.....	113
	Analyse 114	
	Design-based schattingen	114
	Model-based schattingen	116
	Model opbouwen in R	116
	Model interpreteren (alle data)	116
	Model interpreteren (enkel reeksen 1 t.e.m 5).....	118

Lijst van figuren

Figuur 1	Overzicht van het analyseproces	25
Figuur 2	Validatieproces van de gekoppelde bomen.....	27
Figuur 3	Schematische weergave van de aanwasberekening voor verschillende situaties (v_A =volume boom op tijdstip t_1 , v_B = volume boom op tijdstip t_2 , v_{23} = volume van een boom met diameter van 7 cm, v_{34} = volume van een boom met diameter van 40 cm, h^* =geschatte hoogte, v^* = geschat volume van een boom op moment van exploitatie/ omvallen, d^* = geschatte diameter van een boom op moment van exploitatie/ omvallen).	58
Figuur 4	Dotplot van het aantal soorten in de vegetatieopname met onderscheid tussen VBI1 (periode=1) en VBI2 (periode=2).....	68
Figuur 5	Histogram van het aantal boomsoorten met onderscheid tussen VBI1 (periode=1) en VBI2 (periode=2).....	69
Figuur 6	Densityplot van het aantal soorten in de vegetatieopname voor VBI2 met onderscheid tussen plots die op een bosrand gelegen zijn en plots die volledig in bos gelegen zijn.	70
Figuur 7	Boxplots van het aantal boomsoorten in VBI1 en VBI2 met onderscheid tussen de verschillende bostypegroepen.	71

Lijst van tabellen

Tabel 3-1	Groepering van meetvragen op basis van type analysevariabelen	23
Tabel 6-1	Bedekking voor de verschillende klassen van de Braun-Blanquet schaal	40
Tabel 6-2	Berekening van de structuurindex = a+b+c+d (maximale score = 20).....	46
Tabel 6-3	Houtige vegetatie-index = e + f + g + h + i (maximale score = 25)	49
Tabel 6-4	Kruidvegetatie-index = j + k + l + m (maximale score = 25).....	51
Tabel 6-5	Doodhout-index = n + o + p + q + r (maximale score =30)	54
Tabel 6-6	Overzicht van de verschillende type bomen en de berekeningswijze van de lopende jaarlijkse aanwas per hectare.....	56
Tabel 6-7	Resultaten van de aanwasberekeningen	61
Tabel 7-1	Overzicht van de verschillende analysescripts	65
Tabel 7-2	Schatting van de kans op aanwezigheid ('wgt.mean'), de gewogen variantie ('wgt.var') en de ondergrens en bovengrens van het 95% betrouwbaarheidsinterval (respectievelijk 'lower limit confidence interval' = 'llci' en 'upper limit confidence interval' = 'ulci') van inheemse, uitheemse en invasieve boomsoorten voor VBI1 en VBI2 en per eigenaarscategorie.	73
Tabel 7-3	Samenvattende analyseresultaten van de kans op aanwezigheid van uitheemse boomsoorten	77

1 Inleiding

De Vlaamse bosinventarisatie (VBI) is een beleidsondersteunend meetnet om de toestand en evolutie van bos in Vlaanderen op te volgen. Het Agentschap voor Natuur en Bos (ANB) coördineert het meetnet en staat in voor de gegevensverwerking. De eerste inventarisatieronde van de VBI (VBI1) werd uitgevoerd in de periode 1997 – 1999. Ter voorbereiding van de tweede VBI (VBI2) herwerkte het Instituut voor Natuur- en Bosonderzoek de werkwijze (Wouters *et al.*, 2008) volgens de principes van goed meetnetontwerp (Onkelinx *et al.*, 2008).

Een eerste vernieuwing was het bepalen van de prioritaire vragen voor het bosbeleid in Vlaanderen. Dit gebeurde op basis een uitgebreide vragenronde bij de verschillende stakeholders. Om de prioritaire beleidsvragen optimaal te beantwoorden, werden vervolgens een aantal wijzigingen doorgevoerd in het steekproefontwerp, de variabelenkeuze en de bemonsteringsmethodiek. Een derde belangrijke wijziging was de omvorming van de VBI van een periodieke inventarisatie naar een continue inventarisatie met een cyclus van 10 jaar, waarbij elk jaar 1/10^{de} van de meetpunten worden opgemeten. VBI2 is gestart in 2009 en zit momenteel voorbij halfweg in de tienjarige cyclus. ANB wenst op basis van de reeds verzamelde gegevens een eerste tussentijdse technische rapportage uit te voeren. Daarna wenst ANB deze technische rapportage op periodieke basis te herhalen naarmate er meer gegevens binnenkomen.

Het doel van dit rapport is de ontwikkeling van een stramien om de technische rapportage op een wetenschappelijk onderbouwde manier en met gepaste analysetechnieken uit te kunnen voeren op alle beschikbare gegevens. De rapportage zelf maakt geen deel uit van deze opdracht. Een aandachtspunt is dat de methodiek van de gegevensinzameling in VBI1 en VBI2 verschillend is voor bepaalde onderdelen. Hier moet voldoende rekening gehouden worden bij de berekeningen van de resultaten, zodat een vergelijking mogelijk is tussen VBI1 en VBI2.

Eerst worden de meetvragen van de VBI opgefrist en verduidelijkt waar nodig. Vervolgens worden de meetvragen ingedeeld in groepen op basis van het type gegevens dat nodig is om de meetvraag te beantwoorden. Binnen elke groep zullen de analyses gelijkaardig verlopen. Per groep is er voor een meetvraag een voorbeeldanalyse uitgewerkt via een R-script.

In § 4 geven we een globaal overzicht van het analyseproces. Vervolgens bespreken we het validatieproces op basis van gekoppelde bomen (§ 5), de opmaak van de databank met analysevariabelen (§ 6), de eigenlijke analyses voor het beantwoorden van de meetvragen (§ 7) en de opslag van resultaten in een analyseresultatendatabank (§ 8). Ten slotte geven we een overzicht van de verschillende R-scripts en Access-databanken die bij dit rapport horen (§ 9).

In dit rapport veronderstellen we dat de lezer voldoende vertrouwd is met de VBI. Moest dit niet het geval zijn, raden we aan om eerst de beleidssamenvatting over het ontwerp van VBI2 te lezen in Wouters *et al.* (2008).

2 Verduidelijking van de meetvragen

In dit hoofdstuk geven we een overzicht van de verschillende meetvragen van de tweede Vlaamse Bosinventarisatie (VBI) zoals geformuleerd in Wouters *et al.* (2008b) en gaan we na of:

- de vragen duidelijk geformuleerd zijn zodat er geen verschillende interpretaties mogelijk zijn;
- de definities van de gebruikte termen gekend zijn;
- alle gegevens voorhanden zijn om de vragen te kunnen beantwoorden.

Het wijzigen, toevoegen of weglaten van meetvragen gebeurde in samenspraak met ANB. Indien er (delen van) vragen wegvallen, dan wordt dit in de tekst doorstreept. Bijkomende vragen of aanvullingen bij bestaande vragen worden cursief weergegeven.

Wat opvalt is dat bij de meeste meetvragen de toestand gevraagd wordt van een bepaalde karakteristiek. ANB wenst echter zowel de toestand en de trend/het verschil tussen VBI1 en VBI2 te weten. Sommige vragen worden als hypothesen geformuleerd. Ook voor deze vragen wenst ANB een parameterschatting te bekomen met bijhorende betrouwbaarheidsintervallen.

2.1 Toestand en evolutie van de karakteristieken van het bosareaal

We geven eerst de meetvragen weer m.b.t. de toestand en evolutie van de karakteristieken van het bosareaal (§ 2.1.1). De meetvragen worden letterlijk overgenomen uit Wouters *et al.* (2008). Vervolgens bespreken we mogelijke onduidelijkheden, leggen we de definities van bepaalde termen vast en verduidelijken we de meetvragen waar nodig (§ 2.1.2)

2.1.1 Overzicht van de meetvragen

Nummering	Meetvraag
1.a.	Wat is het aandeel van de verschillende eigenaarscategorieën?
1.b.	Wat is het aandeel in de verschillende provincies?
1.c.	Wat is het aandeel aan homogene bestanden?
1.d.	Wat is het aandeel aan homogene dennenbossen?
1.e.	Wat is het aandeel aan populierenplantages?
1.f.	Wat is het aandeel aan bestanden gedomineerd door uitheemse boomsoorten?
1.g.	Wat is het aandeel aan gemengde bestanden op basis van inheemse en standplaatsgeschikte boomsoorten? Wordt in bossen met een beheerplan de norm van een oppervlakteaandeel van 20 % gehaald?
1.h.	Wat is het aandeel van de verschillende bostypegroepen?
	Wat is het aandeel aan open plekken? Wordt in openbare bossen boven de doelstelling van 5 – 15 % oppervlakte aandeel open plekken gehaald?
1.i.	Welk aandeel van de steekproefpunten <i>doelpopulatie</i> valt op een bosrand (intern of extern)?

Nummering	Meetvraag
1.j.	Welk aandeel van de steekproefpunten <i>doelpopulatie</i> valt in een overgangssituatie?
	Welk aandeel van de steekproefpunten is niet toegankelijk?

2.1.2 Definities/onduidelijkheden/scherpstelling

- Een aandeel kan zowel in % als in absolute oppervlakte worden uitgedrukt. Voor ANB ligt de focus vooral op het procentuele aandeel van een bepaalde karakteristiek en minder op de absolute oppervlakte.
- Moeten bebossing en ontbossing meegenomen worden als meetvragen? Bebossing en ontbossing zijn moeilijk te definiëren. De ontbossing kan wel afgeleid worden op basis van de registratie 'geen bos na terreincontrole: bestemming gewijzigd'.
- Wat is de definitie van 'homogeen dennenbos'?
 - Grondvlak grove den > 80% of grondvlak Corsicaanse den > 80% of grondvlak andere den > 80%
- Wat is de definitie van een populierenplantage? Zelfde definitie als homogene bestanden (Wouters *et al.*, 2008).
- Aandeel aan bestanden gedomineerd door uitheemse boomsoorten?
 - Definitie gedomineerd door uitheemse boomsoort: een uitheemse boomsoort met > 50% grondvlak.
 - Lijst uitheemse boomsoorten (Van Den Meersschaut *et al.*, 2001)
- Gemengde bestanden op basis van inheemse en standplaatsgeschikte boomsoorten
 - Lijst inheemse en standplaatsgeschikte boomsoorten. Van Den Meersschaut *et al.* (2001) geeft lijst van de inheemse bomen en struiken in Vlaanderen.
 - Norm voor inheemse bestanden: inheemse boomsoorten dienen minstens 90% van het grondvlak van het bestand in te nemen.
 - Norm voor gemengde bestanden: bestanden zijn gemengd zodra er minstens twee verschillende boomsoorten aanwezig zijn en de hoofdboomsoort 80 % of minder van het bestandsgrondvlak inneemt, of 80% of minder van het totale stamtal voor bestanden jonger dan 30 jaar.
 - Wordt in bossen met een beheerplan de norm van een oppervlakteaandeel van 20 % gehaald?
 - De norm van 20% is niet meer relevant. Enkel een schatting van het oppervlakteaandeel aan gemengde bestanden op basis van inheemse en standplaatsgeschikte boomsoorten.
 - De dataaag met bossen met een beheerplan is bij opmaak van dit rapport nog niet beschikbaar.
- Op basis van de gegevensinzameling kunnen open plekken in het bos niet onderscheiden worden. De meetvraag vervalt.
- Randsituaties en overgangssituaties

- Een randsituatie treedt op wanneer we op de grens zitten tussen bos en niet-bos:
 - Bos grenst aan ander landgebruik buiten het bos (weiland, akkerland, water, weg, bewoning, industrie, andere infrastructuur, ...).
 - Binnen bos (volgens boskartering) grenst een bebost gebied aan een permanente open ruimte (bosweg, brandweg, speelweide, vijver, begraasd perceel, heideperceel, infrastructuur,...).
- Een overgangssituatie treedt op wanneer we op de overgang zitten tussen twee bostypes:
 - Loofhout - naaldhout - gemengd loofhout - gemengd naaldhout
 - Hooghout - middelhout – hakhout
 - Verschil in successiestadia: boomhout versus jongwas, dichtwas of staakhout
 - Bebost perceel – kapvlakte of femelslag – verjonging
- Met toegankelijkheid wordt de publieke toegankelijkheid van een bos bedoeld en niet de toegankelijkheid m.b.t. de gegevensinzameling. Op basis van de gegevensinzameling kan de publieke toegankelijkheid niet afgeleid worden. De meetvraag vervalt.

2.2 Toestand en evolutie van de boomsoortensamenstelling

2.2.1 Overzicht van de meetvragen

Nummering	Meetvragen
2.a.	Wat is van de 10 belangrijkste boomsoorten het procentuele aandeel in het totale stamtal van de populatie?
2.b.	Wat is van de 10 belangrijkste boomsoorten het procentuele aandeel in het totale grondvlak/volume van de populatie?
2.c.	Wat is voor de 5 belangrijkste boomsoorten het aandeel aan de homogene en heterogene bestanden?
2.d.	Is het gemiddeld grondvlakaandeel aan inheemse loofbomen in steekproefpunten die tijdens de eerste VBI in homogene bestanden van niet-inheemse boomsoorten lagen, gelijk of groter dan 30%?
2.e.	Is het gemiddeld grondvlakaandeel aan inheemse boomsoorten in de totale populatie gelijk of groter dan 80%?
2.f.	Wat is de kans op voorkomen en de gemiddelde abundantie van NV (als NV voorkomt) van de 5 belangrijkste boomsoorten?
2.g.	Wat is de kans op voorkomen en gemiddelde abundantie van NV invasieve boomsoorten, in het bijzonder Amerikaanse vogelkers (als NV voorkomt)?

2.2.2 Definitie/ onduidelijkheden/scherpstelling

- De belangrijkste bomen worden geïdentificeerd op basis van de gegevensanalyse (stamtal, grondvlak of volume).

- Heterogeen = gemengd
- Wouters et al. (2008) bevat geen lijst invasieve boomsoorten.

2.3 Toestand en evolutie van de bestandsopbouw

2.3.1 Overzicht van de meetvragen

Nummering	Meetvraag
3.a.	Wat is het aandeel aan loofhout, naaldhout, gemengd loofhout en gemengd naaldhout?
3.b.	Wat is het aandeel aan hooghout, middelhout en hakhout?
3.c.	Wat is het aandeel aan jongwas, dichtwas, staakhout en boomhout?
3.d.	Wat is het aandeel van de mengingsvormen stamsgewijs, groepsgewijs en homogeen?
3.e.	Wat is de verdeling van de bestandsleeftijden?
3.f.	Wat is de verdeling van de bestandsleeftijden van de homogene bestanden van de 5 belangrijkste boomsoorten?
3.g.	In welk aandeel van de steekproefpunten is windworp aanwezig?
	Welk aandeel van de steekproefpunten is overwoekerd door bramen of rododendron of ...?

2.3.2 Definities/ onduidelijkheden/scherpstelling

- Overwoekerd door bramen of rododendron: hierover werden geen gegevens ingezameld (hoewel dit wel voorgesteld wordt in Wouters *et al.* (2008)). ANB geeft aan dat dit ook in de toekomst niet zal gebeuren. De meetvraag vervalt.

2.4 Toestand en evolutie van de indicatoren voor biodiversiteit

2.4.1 Diversiteit aan houtachtige en kruidachtige vegetatie (voor de volledige doelpopulatie en per bostypegroep)

2.4.1.1 Overzicht van de meetvragen

Nummering	Meetvraag
4.1.a.	Wat is de oppervlakte bos (ha) geordend volgens het aantal aanwezige boomsoorten?
4.1.b.	Wat is het gemiddeld aantal plantensoorten?
4.1.c.	Wat is de kans op aanwezigheid van 2 of meer zeldzame plantensoorten?
4.1.d.	Wat is de gemiddelde indexwaarde van Simpson?
4.1.e.	Wat is de gemiddelde indexwaarde van Shannon-Wiener?

2.4.1.2 Definities/ onduidelijkheden/scherpstelling

- ANB beschouwt een boomsoort aanwezig vanaf dat ze voorkomt in het A34-proefvlak.
- We beschouwen enkel het aantal boomsoorten aanwezig in het proefvlak en doen dus geen opschaling naar het aantal boomsoorten per hectare, omdat het aantal aanwezige boomsoorten niet lineair toeneemt met de oppervlakte.
- Aantal plantensoorten: zowel houtachtige als kruidachtige planten en in alle vegetatielagen.
- Definitie zeldzame plantensoorten? Dit kan bepaald worden op basis van de frequentie van voorkomen in 4x4 km hokken in Vlaanderen, zoals gespecificeerd in de flora-atlas (Van Landuyt *et al.*, 2006).
- Eerder de exponent van Shannon-Wiener berekenen voor elk proefvlak en daarvan het gemiddelde berekenen. Dit kan direct vergeleken worden met de gemiddelde soortenrijkdom. Indien alle soorten dezelfde bedekking hebben, zijn beide gemiddeldes gelijk.

2.4.2 Natuurlijkheid van de houtachtige en kruidachtige vegetatie (voor de volledige doelpopulatie en per bostypegroep)

2.4.2.1 Overzicht van meetvragen

Nummering	Meetvraag
4.2.a.	Welke bosoppervlakte is gedomineerd door geïntroduceerde boomsoorten?
4.2.b.	Wat is de kans op aanwezigheid en het grondvlakaandeel van inheemse / uitheemse / invasieve boomsoorten?
4.2.c.	Wat is de abundantie aan verjonging van deze soorten (<i>inheemse/uitheemse/invasieve boomsoorten</i>) (cijfers op basis van plaatsen waar verjonging effectief voorkomt)?
4.2.d.	Wat is gemiddelde abundantie van uitheemse plantensoorten in de vegetatielaag?

2.4.2.2 Definities/ onduidelijkheden/scherpstelling

- Geïntroduceerd = uitheems
- Lijst invasieve boomsoorten ontbreekt in Wouters *et al.* (2008).
- Lijst uitheemse plantensoorten ontbreekt in Wouters *et al.* (2008)

2.4.3 Structuurrijkdom van de bestanden (voor de volledige doelpopulatie en per bostypegroep)

2.4.3.1 Overzicht van de meetvragen

Nummering	Meetvraag
4.3.a.	Wat is de gemiddelde Positioning Index van Clark and Evan?
4.3.b.	Wat is de gemiddelde Mixture Index van Von Gadow?
4.3.c.	Wat is de gemiddelde Height Differentiation Index van Von Gadow?
	Wat is de verdeling (aandeel steekproefpunten) van de bestandsleeftijden?
4.3.d.	Wat is de kans op aanwezigheid van de drie vegetatielagen (kruidlaag, struiklaag en boomlaag)?
	Wat is het aandeel ongelijkjarige bestanden?

2.4.3.2 Definities/ onduidelijkheden/scherpstelling

- De vraag ‘Wat is de verdeling (aandeel steekproefpunten) van de bestandsleeftijden?’ is identiek aan vraag 3.e. (§2.3.1) .
- De vraag ‘Wat is het aandeel ongelijkjarige bestanden?’ is onderdeel van vraag 3.e. (§2.3.1) (ongelijkjarig is één van de leeftijdsklassen).

2.4.4 Aanwezigheid van structurele elementen die biodiversiteit bevorderen (voor de volledige doelpopulatie en per bostypegroep)

2.4.4.1 Overzicht van de meetvragen

Nummering	Meetvragen
4.4.a.	Wat is de kans op aanwezigheid van een dood-hout element?
4.4.b.	Wat is het gemiddelde volume dood hout (m ³ /ha) (staand en liggend)?
4.4.c.	Is het gemiddelde aandeel (%) dood hout (staand en liggend) groter dan 4% van het totale volume?
4.4.d.	Wat is de kans op aanwezigheid van een dikke staande levende boom?
4.4.e.	Wat is de kans op aanwezigheid van een dikke staande dode boom?

2.4.4.2 Definities/ onduidelijkheden/scherpstelling

- Een dikke boom wordt gedefinieerd als een boom met diameter ≥ 40 cm.

2.4.5 (Sub)Scores van de Authenticiteitsindex (voor de volledige doelpopulatie en per bostypegroep)

2.4.5.1 Overzicht van de meetvragen

Nummering	Meetvragen
4.5.a.	Wat is de globale Authenticiteitsindex?
4.5.b.	Wat is de score voor de pijler bosstructuur?
4.5.c.	Wat is de score voor de pijler houtige vegetatie?
4.5.d.	Wat is de score voor de pijler kruidlaag?
4.5.e.	Wat is de score voor de pijler dood hout?

2.4.5.2 Definities/ onduidelijkheden/scherpstelling

- Een onderdeel van de score voor de categorie 'kruidlaag' is het aantal mossoorten. In VBI2 worden geen mossen opgemeten tijdens de vegetatieopname. In VBI1 gebeurde dit wel.

2.5 Toestand en evolutie van de samenstelling van de bosvegetatie onder invloed van milieuveranderingen

2.5.1 Overzicht van de meetvragen

Nummering	Meetvragen
5.a.	Wat is de gemiddelde Ellenberg-L, -V, -R en -N indicatorwaarde?
5.b.	Wat is de verdeling van de Ellenberg-L, -V, -R en -N indicatorwaarde?
5.c.	Wat is de gemiddelde abundantie van karakteristieke plantensoorten of groepen van plantensoorten?
5.d.	<i>Wat is het gemiddeld aantal voorkomende soorten binnen een groep van plantensoorten?</i>

2.5.2 Definities/ onduidelijkheden/scherpstelling

- Groepen van plantensoorten: in dit rapport wordt niet gespecificeerd welke groepen van plantensoorten in beschouwing moeten genomen worden. De gebruiker kan zelf groepen definiëren. Dit kan bv. om habitattypische soorten gaan.

2.6 Toestand en evolutie van het duurzaam bosbeheer en – gebruik

2.6.1 Economisch facet van duurzaam bosbeheer (voor de volledige doelpopulatie en in functie van lidmaatschap bosgroep)

2.6.1.1 Overzicht van de meetvragen

Nummering	Meetvraag
6.1.a.	Wat is de totale (m^3) en gemiddelde (m^3/ha) houtvoorraad?
6.1.b.	Wat is voor de homogene bestanden van de 5 belangrijkste boomsoorten de gemiddelde lopende jaarlijkse aanwas per ha? Ligt deze hoger dan $5 m^3/ha/jaar$?
6.1.c.	Wat is de gemiddelde lopende jaarlijkse bestandsaanwas per bostypegroep? Ligt deze hoger dan $5 m^3/ha/jaar$?
	Wat is voor de 5 belangrijkste boomsoorten in homogene bestanden het gemiddelde geëxploiteerde bestandsvolume ($m^3/ha/jaar$)?
	Wat is het gemiddelde geëxploiteerde bestandsvolume ($m^3/ha/jaar$) per bostypegroep?
6.1.d.	Wat is voor de 5 belangrijkste boomsoorten in homogene bestanden het gemiddelde afgestorven bestandsvolume ($m^3/ha/jaar$)?
6.1.e.	Wat is het gemiddelde afgestorven bestandsvolume ($m^3/ha/jaar$) per bostypegroep?

2.6.1.2 Definities/ onduidelijkheden/scherpstelling

- Geëxploiteerde bomen kunnen niet met voldoende zekerheid geïdentificeerd worden.
- Enkel de overgang van levend naar staand dood hout kan vastgesteld worden. De overgang van levend naar liggend dood hout kan niet bepaald worden. Vraag d en e kan dus slechts gedeeltelijk bepaald worden.

2.6.2 Houtkwaliteit van de economisch belangrijkste boomsoorten

2.6.2.1 Overzicht van de meetvragen

Nummering	Meetvraag
6.2.a.	Wat is voor de economisch belangrijkste boomsoorten het totale volume (m^3) aan industrie-, brand- en pulphout enerzijds (klasse C) en kwaliteitsvol (klasse B) en hoog kwaliteitsvol (klasse A) hout anderzijds?
6.2.b.	Wat is de verdeling van de houtvolumesortimenten i.f.v. de bestandsleeftijd voor cultuurpopulier en gelijkjarige beuken-, eiken- en dennenbestanden?
6.2.c.	Wat is per boomsoort voor de klassen A en B het relatieve aandeel van de verschillende defecten ?

Nummering	Meetvraag
6.2.d	<p>Wat is voor de economisch belangrijkste boomsoorten het volume aan:</p> <ul style="list-style-type: none"> • Takvrije stammen van minimaal 6 meter lengte zonder defecten, met een geringe excentriciteit, met een gering verloop en met een minimale onderdiameter. • Takvrije stammen van minimaal 6 meter lengte met maximaal 1 defect en/of een aanvaardbare excentriciteit en/of een aanvaardbaar verloop en een minimale onderdiameter. • Takvrije stammen van minimaal 6 meter lengte die niet aan bovenstaande criteria voldoen. • Takvrije stammen korter dan 6 meter zonder defecten, met een geringe excentriciteit, met een gering verloop en met een minimale onderdiameter. • Takvrije stammen korter dan 6 meter met maximaal 1 defect en/of een aanvaardbare excentriciteit en/of een aanvaardbaar verloop en een minimale onderdiameter. • Takvrije stammen korter dan 6 meter die niet aan bovenstaande eisen voldoen.

2.6.2.2 Definities/ onduidelijkheden/scherpstelling

- De economisch belangrijkste boomsoorten zijn beuk, eik (ongeacht inlandse, Amerikaans of moeras), cultuurpopulier, berk (ongeacht ruw of zacht), grove en Corsicaanse den.
- Vermits enkel de omtrek gemeten wordt, kan de excentriciteit niet bepaald worden.
- Er is nog niet bepaald hoe de verschillende kwaliteitsklassen (A, B en C) onderscheiden moeten worden. Ook normen voor het verloop zijn nog niet bepaald. In dit rapport geven we generiek analysestramien met focus op het berekenen van de analysevariabelen. Op basis van de analysevariabelen kan de gebruiker dan een indeling maken in kwaliteitsklassen en normen vastleggen.

2.6.3 Milieubeschermend facet van duurzaam bosbeheer (voor de volledige doelpopulatie en in functie van lidmaatschap bosgroep)

2.6.3.1 Overzicht van de meetvragen

Nummering	Meetvraag
	Wat is de totale C opslag (ton) voor Vlaanderen en per bostypegroep?
	Wat is de gemiddelde C opslag (ton/ha) voor Vlaanderen en per bostypegroep?

Deze meetvragen vallen buiten de opdracht.

2.6.4 Improved Pan-European Indicators for Sustainable Forest Management (enkel indicatoren die we geheel of gedeeltelijk kunnen beantwoorden met VBI)

2.6.4.1 Overzicht van de meetvragen

Nummering	Meetvraag
6.4.a.	Bosoppervlakte: Wat is de oppervlakte bos (absoluut en relatief) per bostypegroep?
6.4.b.	Houtvoorraad: Wat is de totale (m ³) en gemiddelde (m ³ /ha) houtvoorraad per bostypegroep?
6.4.c.	Leeftijdsstructuur en/of diameter distributie: <ul style="list-style-type: none"> • Wat is de verdeling (aantal steekproefpunten) van de bestandsleeftijden per bostypegroep (aparte cijfers voor gelijkjarige en ongelijkjarige bestanden)? • Wat is het gemiddeld aantal diameterklassen per steekproefpunt per bostypegroep (aparte cijfers voor gelijkjarige en ongelijkjarige bestanden)?
	C-voorraad: Wat is gemiddelde C-voorraad van de houtige vegetatie (ton CO₂-equivalenten/ha)?
	Aanwas en kappingen: Wat is de netto jaarlijkse aanwas per bostypegroep?
6.1.d.	Boomsoortensamenstelling: Wat is de oppervlakte bos (ha) geordend volgens het aantal aanwezige boomsoorten en volgens bostypegroep?
6.1.e.	Verjonging: Wat is voor gelijkjarige en ongelijkjarige bestanden het aandeel (aantal steekproefpunten) en de oppervlakte (ha) waar verjonging voorkomt, gerangschikt per verjongingstype?
	Natuurlijkheid: Wat is de oppervlakte bos voor de bestypes onverstoord door de mens, halfnatuurlijk en aanplantingen?
6.1.f.	Geïntroduceerde boomsoorten: Wat is de oppervlakte aan bestanden gedomineerd door geïntroduceerde boomsoorten?
6.1.g.	Dood hout: Wat is het gemiddeld bestandsvolume (m ³ /ha) van liggend en staand dood hout?

2.6.4.2 Definities/ onduidelijkheden/scherpstelling

- Veel overlap met andere vragen.
- De indeling van de diameterklassen is als volgt: 0 - 19 cm; 20 – 39 cm; 40 – 60 cm; > 60 cm (Wouters et al., 2008).

- Volgende verjongingstypes worden onderscheiden: natuurlijke verjonging (NV), NV verstrekt door aanplantingen, KV, uitlopers van hakhout. Dit onderscheid is niet mogelijk op basis van beschikbare data.
- Vraag m.b.t. C-voorraad valt buiten de opdracht.

3 Indeling van meetvragen/analysevariabelen

In het vorige hoofdstuk werden de meetvragen thematisch ingedeeld. Om de meetvragen te beantwoorden, moeten in eerste instantie een of meerdere analysevariabelen berekend worden. In dit hoofdstuk maken we een indeling van de analysevariabelen op basis van het type meetvariabele dat nodig is om een analysevariabele te berekenen en op basis van de structuur waarin we de analysevariabelen moeten opslaan. Met structuur bedoelen we dat sommige analysevariabelen per plot berekend moeten worden (bijvoorbeeld grondvlak inheemse soorten), anderen op niveau van een individuele boom (bijvoorbeeld een bepaald houtkwaliteitskenmerk) etc. Deze indeling heeft tot doel om de analyses op een praktische en overzichtelijke manier uit te voeren.

Tabel 3-1 geeft een overzicht van de verschillende groepen waarin de meetvragen worden ingedeeld.

Tabel 3-1 Groepering van meetvragen op basis van type analysevariabelen

GroepNr	Beschrijving Groep	Naam Groep
1	Schatting van de oppervlakte bos op basis van alle rasterpunten. Speciale aandacht naar niet toegankelijke punten.	Oppervlakte bos
2	Schatting van de proportie bestanden met een bepaalde karakteristiek (karakteristiek rechtstreeks ingeschat op bestandsniveau).	Bestandskarakteristieken kwalitatief
3	Alle meetvragen die beantwoord worden op basis van stamtal/grondvlak/volume per proefvlak en per soort of groep van soorten.	Bestandskarakteristieken kwantitatief
4	Meetvragen op basis van voorkomen/abundantie van een soort of groep van soorten, of aantal soorten binnen een groep van soorten of een afgeleide index.	Vegetatiekarakteristieken
5	Meetvragen op basis van karakteristieken van de dichtste burens van een boom.	Ruimtelijke structuur
6	Meetvragen op basis van doodhoutgegevens (staand en liggend).	Dood hout
7	Authenticiteitsindex	Authenticiteitsindex
8	Meetvragen i.v.m. aanwas	Aanwas
9	Meetvragen m.b.t. houtkwaliteit.	Houtkwaliteit
10	Stamtal/grondvlak/volume per boomsoort en per plot	Boomsoortensamenstelling

Voor de analysevariabelen geven we ook aan om welk type variabele het gaat, waarbij we onderscheid maken tussen:

- Binaire variabele: een variabele met enkel een waarde 0 of 1 (ja of nee; aan- of afwezig)
- Categorische variabele: niet-numerieke vaste categorieën, bijvoorbeeld bestandstype

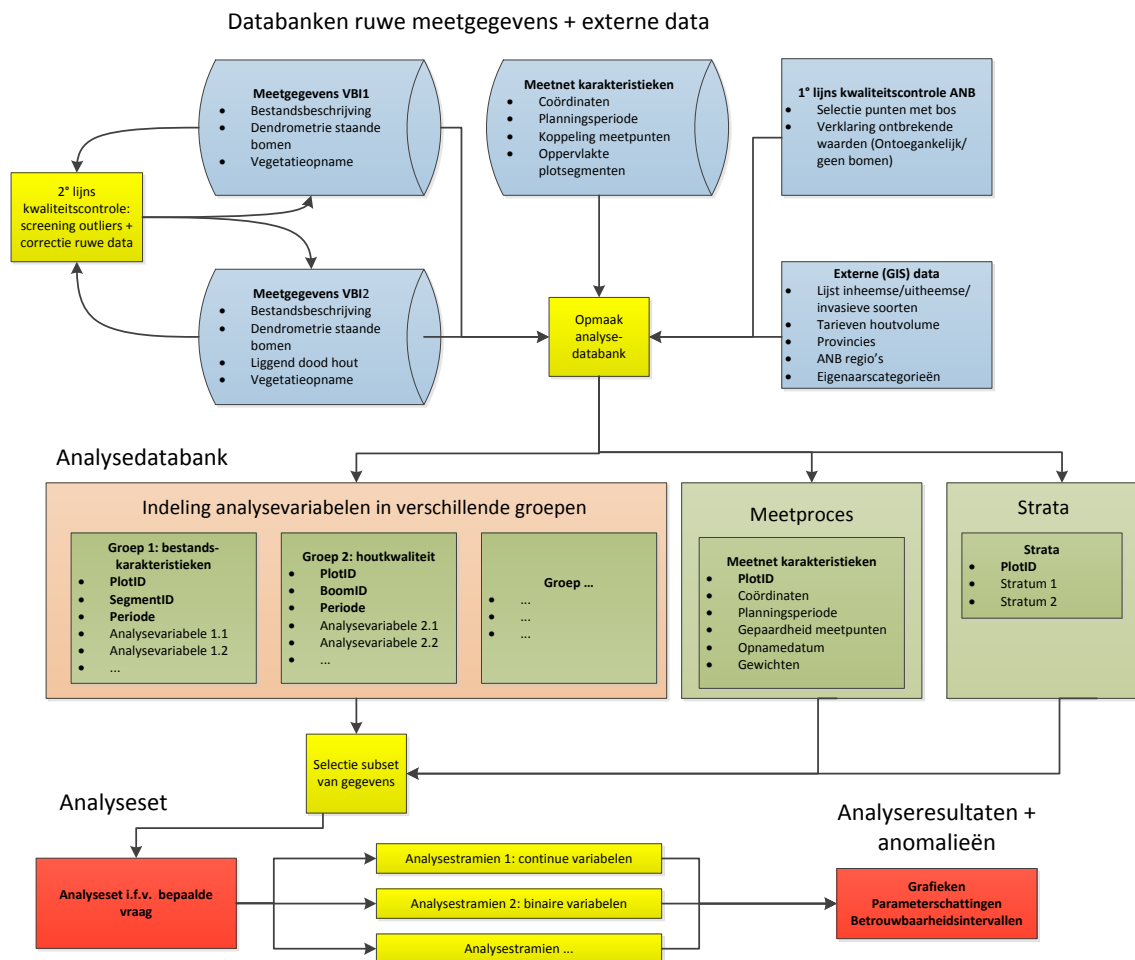
- Continue variabele: (om consequent te zijn ook eerst aangeven wat dit is) we maken een onderscheid tussen volgende subtypen:
 - enkel positieve waarden (bv. grondvlak)
 - een proportie met waarden tussen 0 en 1 (bv. aandeel grondvlak van inheemse bomen)
 - een index met waarden binnen een bepaalde range (bv. een diversiteitsindex)
- Aantallen: discrete variabelen met enkel positieve waarden

Het type variabele bepaalt hoe de verdere analyse moet gebeuren en is in het bijzonder belangrijk bij de keuze van het type model.

De indeling van de meetvragen en analysevariabelen is weergegeven in Bijlage 1.

4 Onderdelen van het analyseproces

Figuur 1 geeft een overzicht van het analyseproces. Het analyseproces vertrekt vanuit de databanken met de basisgegevens. In een eerste stap voeren we een kwaliteitscontrole uit door de meetgegevens van gekoppelde bomen te vergelijken voor twee opeenvolgende meetperiodes (Hoofdstuk 5). Vervolgens worden de verschillende analysevariabelen berekend en opgeslagen in de analysedatabank (Hoofdstuk 6). Uit de analysedatabank wordt een analyseset geselecteerd met de nodige variabelen voor de analyse van een bepaalde meetvraag. Op de analyseset gebeurt de eigenlijke analyse om de meetvraag te beantwoorden. Hiervoor hebben we een aantal voorbeeldscripts ontworpen voor verschillende type variabelen (Hoofdstuk 8). De resultaten van deze analyse worden opgeslagen in een resultatendatabank (Hoofdstuk 9).



Figuur 1 Overzicht van het analyseproces

5 Kwaliteitscontrole van de ruwe meetgegevens op basis van gekoppelde bomen

5.1 Opsporen en beoordelen van anomalieën

Een deel van de plots uit de 1^{ste} VBI werd teruggevonden tijdens de 2^{de} VBI. In deze plots is getracht om opgemeten bomen uit beide inventarisatierondes zoveel mogelijk met elkaar te koppelen op basis van de locatie van de bomen. Op die manier kan de groei van individuele bomen geanalyseerd worden. Deze koppeling laat ook toe om anomalieën te identificeren. We onderscheiden volgende anomalieën:

- shifters: bomen waarvan de soortnaam veranderd is
- zombies: bomen die geëvolueerd zijn van een dode toestand naar een levende
- outliers in verschil in omtrek
- outliers in verschil in hoogte

Een volgende stap bestaat erin de anomalieën te beoordelen en indien nodig (en mogelijk) de ruwe meetwaarden aan te passen. Zo kan de soortnaam, de toestand (levend of dood), de diameter of de hoogte gewijzigd worden. Er kan ook besloten worden dat het om twee verschillende bomen gaat en dat de koppeling fout is.

5.2 Validatieproces

In Figuur 2 geven we een schematisch overzicht van het validatieproces. We kiezen ervoor om wijzigingen aan de ruwe meetgegevens op te slaan in een kopie van de originele databank met de ruwe meetgegevens. De databank van de ruwe meetgegevens is een export uit een 'Fieldmap'-project. Door hiervan een kopie te nemen garanderen we dat we de databank met gewijzigde gegevens opnieuw kunnen inlezen in Fieldmap.

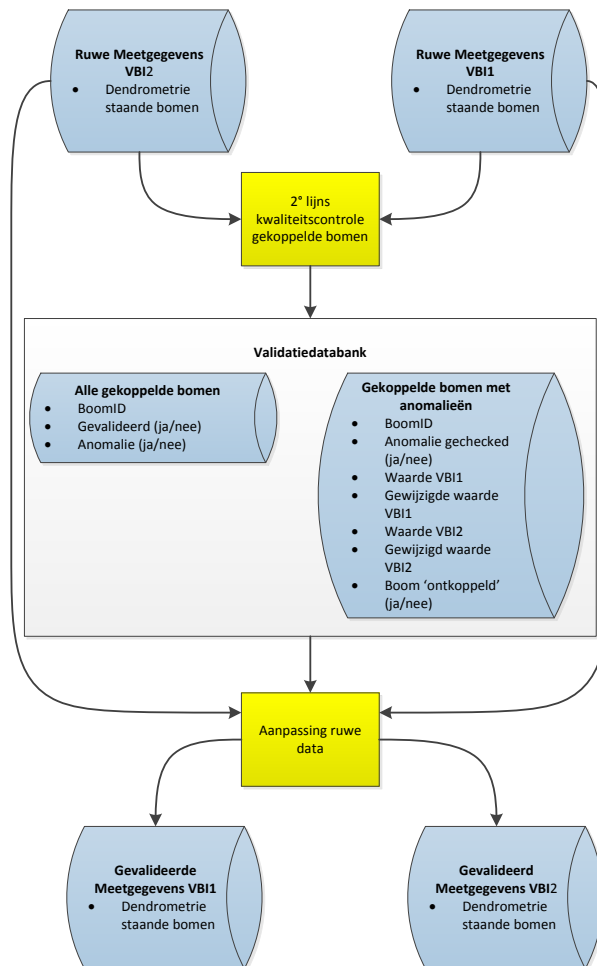
Het validatieproces gebeurt op een semi-geautomatiseerde manier. In de eerste stap genereert een R-script een tabel met anomalieën. De tabel (**tblAnomalies**) bevat de ID van de boom, de meetwaarden van de verschillende perioden die vergeleken worden en een aanduiding van het type anomalie overeenkomstig §5.1. Daarnaast genereert het R-script een overzicht van alle gekoppelde bomen met aanduiding van welke records het validatieproces doorlopen hebben en welke records er als anomalie werden aangeduid (**tblValidationOverview**).

In de tweede stap controleert de validator manueel alle anomalieën. De validator duidt aan als een anomalie werd gecontroleerd en geeft indien nodig een nieuwe waarde in voor een bepaalde meetvariabele of 'ontkoppelt' de boom. Dit gebeurt door het invullen van volgende extra velden in de anomalieëntabel:

- anomalie gevalideerd: ja/nee (default = nee)
- periode aanpassing: één van beide periodes die vergeleken worden en waarin een wijziging wordt doorgevoerd (default = leeg veld)
- veldnaam aanpassing: naam van het veld dat aangepast wordt (default = leeg veld)
- aangepaste waarde: nieuwe waarde voor meetvariabele (default = leeg veld)
- boom ontkoppeld: ja/nee (default = nee)

Na de manuele validatiestap zullen via een R-script alle wijzigingen uit de anomalieëntabel uitgevoerd worden in de kopie van de originele databank.

Bij een update van het validatieproces (wanneer bijvoorbeeld een nieuwe reeks meetgegevens beschikbaar is) wordt er eerst gekeken in **tbValidationOverview** welke bomen reeds gevalideerd werden, zodat enkel voor de nieuwe set gekoppelde bomen anomalieën geïdentificeerd worden.



Figuur 2 Validatieproces van de gekoppelde bomen

6 Analysedatabanken

6.1 Algemeen

Dit hoofdstuk geeft een overzicht van de opbouw van de analysedatabank voor het beantwoorden van de verschillende meetvragen. Een **analysedatabank** geldt als tussenstap tussen de databank met de ruwe meetgegevens en de eigenlijke analyse van de gegevens. Heel wat tussenbewerkingen zijn immers nodig om uit de ruwe meetgegevens variabelen te distilleren die bruikbaar zijn voor de statistische analyse. Vaak moeten we een synthese maken van gegevens uit meerdere bronnen. Het tellen van het aantal exoten in de vegetatieopname bijvoorbeeld, vergt ook een lijst met de exoten. Of we moeten gegevens aggregeren opgemeten in de verschillende concentrische opnamecirkels, bijvoorbeeld om het totale volume hout te berekenen. Daarenboven is de gegevensinzameling van de tweede Vlaamse Bosinventarisatie (VBI2) niet helemaal identiek aan de eerste (VBI1).

Eerst schetsen we het proces van hoe de analysedatabank tot stand komt uit de ruwe gegevens. Vervolgens overlopen we de inhoud van de verschillende tabellen in de analysedatabank. In het bijzonder onderzoeken we de verschillpunten tussen de eerste en tweede bosinventarisatie en gaan we de mogelijke implicaties ervan na op de analyse van de gegevens en de interpretatie van de resultaten. We voorzien de nodige hulpvariabelen om bij de verwerking van de gegevens met eventuele veranderingen rekening te houden. Dit doen we bijvoorbeeld door specifieke selecties te maken uit de dataset en/of informatie in verband met de verandering op te nemen als verklarende variabelen in de statistische modellen.

Ten opzichte van VBI1 zijn er drie belangrijke veranderingen in de methodiek van de gegevensinzameling. De selectie van de meetpunten is bijgestuurd in VBI2. In VBI1 werd gestreefd naar homogene opnameplots. Punten aan de rand van het bos of in open ruimtes in het bos, of met verschillende bostypes werden daarom verschoven. Dat geeft aanleiding tot een vertekening en daarom werden deze punten in VBI2 op de oorspronkelijke locatie teruggelegd. Een tweede belangrijk verschil betreft de vegetatieopnames. In VBI1 werd slechts in de helft van de plots de vegetatie opgenomen. In VBI2 werden in alle plots vegetatieopnames gemaakt, maar er werden geen mossen meer opgemeten. Omdat in VBI2 de kleinste van de vier concentrische opnamecirkels weggelaten werd, wordt de natuurlijke verjonging nu afgeleid uit de vegetatieopname. De derde belangrijke verandering betreft de bepaling van dood hout. Voor VBI2 gebruiken we de *line intersect sampling* (LIS) methode, die veel consistentere resultaten oplevert. Hierdoor zijn de gegevens niet altijd even goed vergelijkbaar meer tussen beide inventarisatierondes.

Daarnaast zijn er nog een aantal subtielere verschillen, waarvoor we zullen moeten corrigeren bij de berekening van bepaalde analysevariabelen. De impact hiervan bespreken we bij het overlopen van de tabellen.

6.2 Opmaak van de analysedatabank

De extractie van de gegevens verloopt geautomatiseerd op basis van R-scripts, zodat we dynamisch en goed gedocumenteerd aanpassingen kunnen doorvoeren. Deze dynamische omgeving is noodzakelijk om, naarmate het project vordert en het inzicht groeit, fouten te corrigeren of variabelen anders te definiëren of bij te voegen.

Voor de verschillende **analysevariabelen** worden de nodige ruwe meetgegevens via access-query's opgevraagd. Vervolgens berekenen R-scripts hieruit, via een aantal tussenstappen, de analysevariabelen. Zo moet er bijvoorbeeld in stap 1 het grondvlak en volume per boom berekend worden, in stap 2 het grondvlak en volume gesommeerd worden per type boom (inheems, uitheems) en per plot of per segment van een plot, en in een laatste stap de variabelen omgezet worden naar een grondvlak en volume per hectare door gebruik te maken van de juiste expansiefactor.

Zoals hoger vermeld, verschilt de gegevensinzameling hier en daar tussen de eerste en tweede bosinventarisatie. Ook de structuur van de databank met ruwe meetgegevens en de naamgeving van de meetvariabelen verschilt. Daarom werken we met aparte scripts voor de het berekenen van analysevariabelen voor VBI1 en VBI2. De analysevariabelen worden vervolgens samengevoegd in één tabel in de analysedatabank, waarbij het gegevensveld 'periode' de inventarisatieronde aangeeft. Het samenvoegen van gegevens uit VBI1 en VBI2 gebeurt dus niet op niveau van de ruwe meetgegevens, maar op niveau van de analysevariabelen.

De analysedatabank bevat ook tabellen die het **meetproces** en/of de meethistoriek beschrijven. Deze informatie is belangrijk om gegevens uit de eerste en tweede inventarisatieronde te koppelen om trends te berekenen. De tabel bevat informatie over welke meetpunten gepaard zijn. Dit wil zeggen dat gemarkeerde plots uit VBI1 werden teruggevonden in de tweede ronde. Informatie i.v.m. het meetproces is ook belangrijk om een onderscheid te maken tussen ontbrekende waarden en nulwaarnemingen. Als we voor een bepaald attribuut (bijvoorbeeld grote bomen) geen opnames vinden, en het proefvlak is wel degelijk geïnventariseerd, dan gaat het om een nulwaarneming.

Ten slotte bevat de analysedatabank ook informatie over allerhande **strata** en/of verklarende variabelen die relevant kunnen zijn in de gegevensanalyse. Daarbij maken we onderscheid tussen strata die onveranderlijk zijn in de tijd (statische strata), zoals provincie of ecoregio, en strata die veranderlijk zijn (dynamische strata) zoals eigenaarscategorie.

Op basis van de informatie over het meetproces en de strata kunnen we heel vlot een specifiek deel van gegevens selecteren uit de analysedatabank. Deze subset van gegevens noemen we de analyseset. De analyseset wordt gebruikt voor de eigenlijke gegevensanalyse.

6.3 Meetproces VBI1 en VBI2

Vooraleer we de verschillende tabellen overlopen, beschrijven we eerst het meetproces en gaan we nader in op de verschillen tussen de eerste en tweede bosinventarisatie. De navigatietechnieken in VBI2 zijn veel nauwkeuriger geworden. Nieuw geïnstalleerde meetpunten zullen veel beter het theoretische raster benaderen dan de oude meetpunten uit VBI1. Daarnaast zijn een aantal ontwerpopties veranderd die een grote impact kunnen hebben. In VBI1 werden punten aan de rand van het bos of een open plek systematisch verschoven naar binnen. Ook meetpunten verspreid over twee bostypes werden verschoven. De bedoeling was om homogene opnameplots te bekommen. Al deze punten worden bij VBI2 teruggelegd op hun oorspronkelijke plaats. Deze meetpunten zijn per definitie niet goed vergelijkbaar tussen beide inventarisatierondes om trends te berekenen.

Voor de VBI is een theoretisch raster van 1000 m x 500 m getekend over Vlaanderen. Alle rasterpunten worden op basis van orthofoto's ingedeeld in drie categorieën: punten met bos, punten zonder bos en twijfelgevallen. De punten met bos en de twijfelgevallen worden vervolgens bezocht op terrein.

In de eerste inventarisatieronde (1997-1999) werden de geselecteerde rasterpunten op het terrein opgezocht via landmetertechnieken, vertrekkend van een punt met gekende coördinaten. Het punt waarop men uitkwam was het meetpunt. De coördinaten van dit meetpunt konden in de praktijk aanzienlijk verschillen van de coördinaten van het (theoretische) rasterpunt. Meetpunten die op de rand van het bos lagen of die verschillende bestandstypen bevatten, werden verplaatst. De meetpunten werden gemerkt met een koperen markering en door markering van bomen. De route waarlangs werd genavigeerd werd bewaard.

Ook in VBI2 worden alle rasterpunten eerst visueel geëvalueerd. Meetpunten die in VBI1 werden opgemeten en waar nog steeds bos aanwezig is, tracht men opnieuw te lokaliseren a.d.h.v. de routegegevens uit VBI1 en een metaaldetector. In het geval dat het meetpunt niet teruggevonden wordt of dat het bos er werd gekapt, wordt naar het theoretische

rasterpunt genavigeerd om daar een nieuw meetpunt te installeren. Meetpunten die in VBI1 werden verschoven t.o.v. het raster, worden in VBI2 teruggeplaatst. Er wordt richting het rasterpunt genavigeerd en een nieuw meetpunt geïnstalleerd. In dergelijke gevallen kunnen het oorspronkelijke en het nieuwe meetpunt niet als gepaarde metingen beschouwd worden. In VBI2 worden de coördinaten van het meetpunt opgemeten.

Elk meetpunt krijgt het unieke ID van de rasterpunt waartoe het hoort. De ID's zijn dezelfde in VBI1 en VBI2. Een meetpunt uit VBI1 kan eenzelfde ID hebben als een meetpunt uit VBI2 maar toch verschillen in locatie. Dit is het geval wanneer het meetpunt uit VBI1 niet werd teruggevonden en een nieuw meetpunt werd geïnstalleerd en wanneer een verplaatst meetpunt werd teruggeplaatst op het raster.

Voor een verder overzicht van de bemonsteringsmethodiek van VBI1 verwijzen we naar Afdeling Bos & Goen (2001). Voor VBI2 verwijzen we naar Wouters *et al.* (2008b).

6.4 Variabelen m.b.t. het meetproces

In deze sectie geven we een overzicht van de verschillende tabellen met variabelen uit de analysedatabank die betrekking hebben op het meetproces. Bij de bespreking maken we een onderscheid tussen de ruwe gegevens/meetvariabelen die rechtstreeks uit de databank met ruwe meetgegevens komen en de analysevariabelen afgeleid uit de ruwe gegevens.

6.4.1 Opnamedatum en koppeling van plots

Deze tabel geeft voor alle opgemeten plots informatie over de opnamedatum en de koppeling van plots tussen verschillende periodes.

tblPlotDetails

- IDPlots: ID van het rasterpunt waar de plot gelegen is. Twee plots kunnen dezelfde waarde voor IDPlots hebben, maar toch niet samenvallen. Dit is het geval wanneer een gemarkeerd plot niet wordt teruggevonden in VBI2 en een nieuw plot wordt geïnstalleerd op of nabij hetzelfde rasterpunt. Om unieke plots te identificeren, wordt een bijkomende ID aangemaakt (zie verder).
- Periode: volgnummer voor de opeenvolgende bosinventarisaties. 1 voor de 1^{ste} bosinventarisatie en 2 voor de 2^{de} bosinventarisatie. Dit veld laat toe om ook gegevens uit volgende inventarisaties in de voorgestelde structuur op te slaan.
- DateDendro: geeft het jaartal aan waarin de dendrometrische gegevensinzameling werd uitgevoerd (A34-plot).
- DateReg: geeft het jaartal aan waarin de doorgroeiende verjonging werd opgemeten (A2-plot).
- DateVeg: geeft het jaartal aan waarin de vegetatieopname werd uitgevoerd. Voor VBI1 ontbreekt de datum wanneer geen vegetatieopname voorzien was.
- Reeks: in VBI2 werden de rasterpunten ingedeeld in 10 reeksen. De aangeleverde databank met VBI2-gegevens bevat gegevens t.e.m. reeks 5. Via het reeksnummer kunnen de overeenkomstige rasterpunten uit VBI1 geselecteerd worden.
- Reloc_cat: beschrijving van de localisatie van gemarkeerde proefvlakken uit een vorige inventarisatieronde a.d.h.v. volgende categorieën:
 - Gekoppelde meting: het gemarkeerd proefvlak werd teruggevonden. We maken daarbij verder onderscheid tussen
 - 'Gekoppelde meting: koperen markering teruggevonden'

- ‘Gekoppelde meting: bomen teruggevonden/ geen koperen markering’
 - Plot op rasterpunt: er wordt een nieuw proefvlak geïnstalleerd, waarbij onderscheid gemaakt wordt tussen
 - ‘Plot op rasterpunt: nieuw plot’: installatie van een proefvlak op een locatie die voorheen geen bos was of het terugplaatsen van een verschoven proefvlak op het theoretische raster.
 - ‘Plot op rasterpunt: plot niet teruggevonden’: indien een plot niet wordt teruggevonden, wordt naar het theoretisch rasterpunt genavigeerd en wordt daar een nieuw meetpunt geïnstalleerd.
 - ‘Plot op rasterpunt: origineel plot gekapt’: indien het originele plot werd gekapt, wordt er genavigeerd naar het theoretische rasterpunt. Indien op de nieuwe locatie bos aanwezig is, wordt daar een nieuw meetpunt geïnstalleerd.
 - ‘Plot op rasterpunt: origineel plot ontbost (bestemming gewijzigd)’: analoog met ‘Plot op rasterpunt: origineel plot gekapt’.
- Reloc_cat_code: code voor localisatiecategorie.
- Plot_cat: via dit veld kan achterhaald worden welke verschoven punten uit VBI1 teruggelegd werden op het raster. Deze teruggeplaatste punten worden aangeduid met ‘geen route – verlegd punt’.
- Plot_cat_code: code voor Plot_cat.
- StartPeriode: periode waarin een plot een eerste maal werd opgemeten. Voor een plot uit VBI1 dat opnieuw werd geïnventariseerd/opgemeten in VBI2 is de startperiode dus 1; voor de andere plots is de startperiode gelijk aan 2. Via dit gegevensveld kan de koppeling van plots ook in de toekomst bijgehouden worden wanneer er gegevens uit VBI3 en verder beschikbaar komen. Een deel van de plots zal dan bijvoorbeeld in de drie periodes opgemeten zijn (startperiode = 1), een deel enkel in VBI2 en VBI3 (startperiode = 2), etc.
- IDGroup: unieke ID per opgemeten plot. Deze ID bestaat uit een samenvoeging van het veld IDPlots en de startperiode. Dit gegevensveld is nodig om groepen te identificeren voor het toepassen van mixed models.
- Replaced: geeft aan welke meetpunten uit VBI1 werden teruggeplaatst naar het rasterpunt in VBI2.

6.4.2 Type gegevens per plot en onderscheid tussen nulwaarnemingen en ontbrekende waarden

Deze tabellen bevatten alle meetpunten waarvoor gegevens werden ingezameld en waarvoor minstens een bestandsbeschrijving werd uitgevoerd. De gegevensinzameling bestaat uit verschillende types van gegevens. Per type geeft de tabel aan of er gegevens aanwezig zijn in de databank. We maken gebruik van aparte tabellen voor VBI1 en VBI2 vermits andere types van gegevens werden ingezameld in beide inventarisatierondes.

Wanneer er voor een bepaald plot geen gegevens van een bepaald type boom (bijvoorbeeld A3- of A4-bomen) aanwezig zijn in de databank, dan zijn daar twee mogelijke oorzaken voor: ofwel zijn er effectief geen A3- of A4-bomen aanwezig, ofwel werden geen gegevens ingezameld (bijvoorbeeld omdat de plot niet toegankelijk was voor het uitvoeren van

dendrometrische opmetingen omdat het te nat is). Het ontbreken van een bepaald type boom (= nulwaarneming) wordt niet op het terrein genoteerd. Daarom maken we een aantal velden aan die per type gegevens aangeven of het om een nulwaarneming gaat of een ontbrekende waarde op basis van een aantal veronderstellingen. Het onderscheid tussen beide is van belang in de eigenlijke analyse.

Een verkenning van de plots uit VBI1 leert dat A3A4-gegevens enkel ontbreken voor plots die in een kaalkapvlakte, in een open ruimte of in zeer jonge bestanden (< 20 jaar) liggen. We veronderstellen hier dus steeds nulwaarnemingen (er zijn geen bomen met diameter groter dan 40 cm). Ook voor A1- en A2-bomen veronderstellen we steeds nulwaarnemingen als gegevens ontbreken. In VBI1 werden slechts in de helft van de meetpunten vegetatieopnames uitgevoerd. De overige helft geeft dus ontbrekende waarden.

In VBI2 wordt er via een regelmatige kwaliteitscontrole nagegaan wat de oorzaak is van ontbrekende gegevens voor bepaalde plots. Hieruit blijkt dan bijvoorbeeld dat in sommige plots wel een bestandsbeschrijving is gebeurd, maar dat het er niet mogelijk was dendrometrische metingen uit te voeren omdat de plots te nat of onvoldoende bereikbaar waren. Via deze informatie kunnen we dus nulwaarnemingen en ontbrekende waarden onderscheiden.

Hieronder geven we een beschrijving van de gegevensvelden van beide tabellen.

tblRecordsVBI1

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- DendroRecord: geeft aan voor welke meetpunten er een dendrometrische gegevensinzameling werd uitgevoerd.
- VegRecord: geeft aan voor welke meetpunten een vegetatieopname werd uitgevoerd.
- A34Record: geeft aan of er bomen werden opgemeten in A34-plot.
- A2Record: geeft aan of er bomen werden opgemeten in A2-plot.
- A1Record: geeft aan of er bomen werden opgemeten in A1-plot.
- vegLayer: geeft aan of de bedekking van de verschillende vegetatielagen werd ingeschat.
- vegSpecies: geeft aan of er soorten in de kruidlaag werden genoteerd tijdens de vegetatieopname.
- A34RecordZero: geeft een nulwaarneming aan voor een bepaalde plot, d.w.z. dat er geen A3- of A4-bomen aanwezig zijn in de plot. Op basis van dit veld kunnen we voor ontbrekende records onderscheiden of het om een nulwaarneming of een ontbrekende waarde gaat.
- A2RecordZero: nulwaarneming voor A2-bomen.
- A1RecordZero: nulwaarneming voor verjonging.
- vegRecordZero: nulwaarneming voor vegetatielaag.

tblRecordsVBI2+

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: dit veld laat toe om ook gegevens van komende inventarisatierondes op te nemen in de analysedatabank.

- StandDiscription: geeft aan voor welke meetpunten er een bestandsbeschrijving is gebeurd.
- DendroRecord: geeft aan of er bomen werden opgemeten in de A34-plot.
- RegRecord: geeft aan of er bomen werden opgemeten in A2-plot.
- VegRecord: geeft aan of er een vegetatieopname werd uitgevoerd.
- LIMRecord: geeft aan of er liggende bomen werden opgemeten.
- DendroRecordZero: nulwaarneming voor A3- en A4-bomen.
- RegRecordZero: nulwaarneming voor A2-bomen.
- LIMRecordZero: nulwaarneming voor liggend dood hout.
- vegRecordZero: nulwaarneming voor vegetatielaag.

6.4.3 Plotgewichten en oppervlaktes van de A2-, A3- en A4-plots

Vanaf VBI2 zijn er meetpunten die slechts gedeeltelijk uit bos bestaan of die verschillende bestandstypen bevatten. Dit was niet het geval in VBI1. In dergelijke gevallen werden de meetpunten verschoven zodat ze steeds uit een homogeen bestandstype zouden bestaan.

Indien een meetpunt gedeeltelijk uit bos bestaat (EN het middelpunt van het meetpunt in bos ligt) of verschillende bestandstypes omvat, wordt de Area Decision Method gebruikt. Het meetpunt wordt daarbij opgesplitst in segmenten. Per segment gebeurt er vervolgens een bestandsbeschrijving.

Bij de gegevensanalyse moeten we er rekening mee houden dat sommige meetpunten slechts gedeeltelijk uit bos bestaan. Dit kunnen we doen door een gepast gewicht toe te kennen aan de meetpunten: het oppervlaktaandeel van de A4-plot (straal 18 m) dat uit bos bestaat. Indien we voor een bepaald meetpunt gegevens uit verschillende segmenten moeten combineren (bv. de gemiddelde score voor 'mengingsvorm' voor een bepaald meetpunt), moeten we ook aan elk segment een gewicht toekennen. Dit gewicht is gelijk aan de verhouding tussen de oppervlakte van het (bos)segment t.o.v. de totale oppervlakte bos in het meetpunt.

Om stamtal, grondvlak en volume per hectare te berekenen, moeten we ook de oppervlaktes van de A4-, A3- en A2-plot kennen. Vermits de A4- en A3-plots worden opgesplitst in segmenten, moeten we de oppervlakte per segment kennen. De A2-plots worden niet opgesplitst, maar ze kunnen wel slechts gedeeltelijk uit bos bestaan. Dit moet dan ook mee in rekening genomen worden wanneer het stamtal van doorgroeidende verjonging berekend wordt per ha.

tblPlotWeights

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: dit veld laat toe om ook gegevens van komende inventarisatierondes op te nemen in de analysedatabank
- PlotWeight: geeft het gewicht van de meetpunten aan. Dit komt overeen met het aandeel bos in de A4-plot (waarde tussen 0 en 1).
- SegmentWeight: geeft het gewicht voor de segmenten aan. Het is de verhouding tussen de oppervlakte van het segment van de A4-plot en de oppervlakte bos in de A4-plot. Dit berekenen we enkel voor de segmenten die uit bos bestaan (en dus per definitie volledig uit bos bestaan).
- AreaA4_m2: oppervlakte van de A4-plot dat in segment j van plot i valt

- AreaA3_m2: oppervlakte van de A3-plot dat in segment j van plot i valt
- AreaA2_m2: oppervlakte van A2-plot dat uit bos bestaat voor plot i .

6.4.4 Verdeling van de individuele bomen over verschillende plotsegmenten

Onderstaande tabel geeft aan in welk segment elke A3- of A4-boom gelegen is. Deze informatie is nodig om per segment de analysevariabelen te berekenen. De tabel bevat enkel gegevens voor VBI2, gezien de plots in VBI1 niet opgesplitst werden in segmenten. De tabel is ook bruikbaar voor opslag van gegevens uit toekomstige inventarisatierondes.

tblTreesSegmentID

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.
- ID: ID van de boom
- IDSegments: ID van het segment binnen de plot.

6.5 De strata en de verklarende variabelen

Enkele meetvragen vereisen een uitspraak over verschillende strata. We maken een onderscheid tussen statische en dynamische strata. De strata kunnen ook als verklarende variabele in de analysemodellen gebruikt worden.

6.5.1 Statische strata

Deze tabel geeft informatie over verschillende strata die als onveranderlijk kunnen beschouwd worden. De tabel bevat alle rasterpunten (dus ook niet-bos).

tblPlotStrataStatic

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- Provincie
- EcoDistrict
- SoilType: bodemtype
- ANBRegion: beheerregio van ANB

6.5.2 Dynamische strata

Deze tabel geeft informatie over verschillende strata die als veranderlijk kunnen beschouwd worden. De tabel bevat alle meetpunten uit de verschillende inventarisatierondes.

tblPlotStrataDynamic

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.
- StandType: bestandstype.
- OwnerType: eigenaarscategorie.
- ForestType: bostype volgens Cornelis *et al.* (2007)
- ForestTypeGroup: bostypegroep volgens Cornelis *et al.* (2007)

6.6 Analysevariabelen

De analysevariabelen worden opgesplitst in verschillende tabellen volgens de indeling van de meetvragen (zie § 4). Per tabel worden de analysevariabelen genummerd zodat op basis van de tabel in Bijlage 1 eenvoudig de overeenkomstige meetvraag kan teruggevonden worden. Naast de analysevariabele bevatten de tabellen de belangrijkste hulpvariabelen die nodig zijn om de analysevariabelen te kunnen berekenen. Om de tabellen overzichtelijk te houden worden niet alle hulpvariabelen weergegeven. Ten slotte bevatten de analysetabellen gegevens over het meetproces (uit de tabellen beschreven in §6.4.1 en §6.4.3) en de strata (uit de tabellen beschreven in §6.5.1 en §6.5.2) die nodig zijn voor de analyse, zodat in de eigenlijke analyse geen tabellen meer moeten samengevoegd worden.

6.6.1 Bosoppervlakte

Deze tabel geeft voor **alle** punten van de steekproef weer of ze al dan niet bos bevatten.

tbl1Bosoppervlakte

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.

Analysevariabelen

- v1_ForestCover:
 - Forest: indien het middelpunt van de plot in bos valt (en de plot dus opgemeten wordt).
 - Non-Forest: indien het middelpunt van de plot niet in bos valt (en de plot dus niet opgemeten wordt).
 - Unknown: indien de plot (volledig) ontoegankelijk is of er geen toestemming is om de plot te bezoeken.

Hulpvariabelen

- OrthoPhotoClass
 - We onderscheiden volgende klassen:
 - Non-Forest: rasterpunten die via orthofotoherkenning als niet-bos worden geclassificeerd.
 - PotentialForest: punten die via orthofotoherkenning aangeduid zijn als mogelijk bos bevattend.
 - Deze hulpvariabele kan gebruikt worden om de oppervlakte bos in Vlaanderen te schatten op volgende manier: binnen de klasse 'PotentialForest' kan eerst de verhouding berekend worden tussen het aantal punten als bos aangeduid en het aantal punten als niet-bos aangeduid. Deze verhouding kan dan gebruikt worden om het aantal punten met bos te schatten binnen de klasse 'Unknown' van de variabele 'v1_ForestCover'. Vervolgens wordt de verhouding genomen tussen het aantal punten met bos (inclusief het geschatte aantal punten bos binnen de klasse 'Unknown') en het totaal aantal meetpunten. Door deze verhouding te vermenigvuldigen met de oppervlakte van Vlaanderen kan de oppervlakte bos in Vlaanderen geschat worden.

Strata

- Provincie
- EcoDistrict
- SoilType: bodemtype
- ANBRegion: beheerregio van ANB
- OwnerType: eigenaarscategorie
- ForestTypeGroup: bostypegroep volgens Cornelis *et al.* (2007)

Meetproces

- Reeks
- StartPeriode
- IDGroup

6.6.2 Bestandskarakteristieken kwalitatief

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.

Analysevariabelen

- v1_HomogeneStand:
 - TRUE indien een homogeen bestand
 - FALSE indien geen homogeen bestand
- v2_StandType: bestandstype
 - loofhout
 - naaldhout
 - gemengd loofhout
 - gemengd naaldhout
 - kapvlakte
 - open ruimte binnen bos (enkel in VBI1 zijn er plots die onder deze categorie vallen)
- v3_HarvestType: bedrijfsvorm
 - hakhout
 - hooghout
 - middelhout
 - niet van toepassing
- v4_MixType: mengingsvorm
 - homogeen
 - stamsgewijs
 - groepsgewijs
 - niet van toepassing

- v5_StandAge: bestandsleeftijd
 - 1 – 20 jaar
 - 21 – 40 jaar
 - 41 – 60 jaar
 - 61 – 80 jaar
 - 81 – 100 jaar
 - 101 – 120 jaar
 - 121 – 140 jaar
 - 141 – 160 jaar
 - Ongelijkjarig
 - Niet van toepassing
- v6_WindThrow: windworp aanwezig (ja/nee)
- v7_ForestEdge: de plot ligt op een bosrand (TRUE/FALSE). Voor plots in VBI1 is dit steeds FALSE
- v8_ForestOvergang: de plot ligt op een overgangsfase tussen verschillende bestandstypen (TRUE/FALSE)

6.6.3 Bestandskarakteristieken kwantitatief

tbl3BestandskaraktKwant

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.

Analysevariabelen

- v1_HomogeenDen: homogeen dennenbestand
 - ja: het stamtal van de meest voorkomende dennensoort > 80% indien de bestandsleeftijd < 20 jaar, of het grondvlak van de meest voorkomende dennensoort > 80% indien de bestandsleeftijd > 20 jaar of ongelijkjarig
 - nee: het stamtal van de meest voorkomende dennensoort < 80% indien de bestandsleeftijd < 20 jaar, of het grondvlak van de meest voorkomende dennensoort < 80% indien de bestandsleeftijd > 20 jaar of ongelijkjarig
 - nee: de bestandsleeftijd is niet van toepassing (kaalkap, aanplant)
- v2_HomogeenPopulier: populierenplantage
 - ja: het stamtal van populier > 80% indien de bestandsleeftijd < 20 jaar, of het grondvlak van populier > 80% indien de bestandsleeftijd > 20 jaar of ongelijkjarig
 - nee: het stamtal van populier < 80% indien de bestandsleeftijd < 20 jaar of het grondvlak van populier < 80% indien de bestandsleeftijd > 20 jaar of ongelijkjarig
 - nee: de bestandsleeftijd is niet van toepassing (kaalkap, aanplant)

- v3_GedomineerdExotic: bestand gedomineerd door uitheemse soorten. Van Den Meersschaut *et al.* (2001) bevat een lijst met inheemse soorten in Vlaanderen. Alle andere soorten beschouwen we als uitheems.
 - ja: de som van het grondvlak van de uitheemse soorten > 50%
 - nee: alle andere gevallen
- v4_GemengdInheems: inheems gemengd bestand
 - ja: het betreft een gemengd bestand en de som van het grondvlak van de inheemse bomen > 90%
 - nee: alle andere gevallen
- v5_FractionBasalAreaAutocht: grondvlakaandeel van de inheemse soorten
- v6_FractionBasalAreaExotic: grondvlakaandeel van de uitheemse soorten
- v7_FractionBasalAreaInvasive: grondvlakaandeel van de invasieve soorten
- v8_PresenceAutocht: voorkomen van inheemse boomsoorten.
 - aanwezig: minstens één A2-, A3- of A4-boom is een inheemse soort
 - afwezig : alle andere gevallen
- v9_PresenceExotic: voorkomen van uitheemse boomsoorten
 - aanwezig: minstens één A2-, A3- of A4-boom is een uitheemse soort
 - afwezig : alle andere gevallen
- v10_PresenceInvasive: voorkomen van invasieve soorten
 - aanwezig: minstens één A2-, A3- of A4-boom is een invasieve soort
 - afwezig: alle andere gevallen
- v11_TotalVolume: totaal volume in m³/ha op basis van
 - Tabellen met 2 ingangen (Dagnelie *et al.*, 1985; Dik, 1990; Quataert *et al.*, 2011) voor:
 - intacte individuele levende bomen uit VBI1
 - intacte individuele levende bomen uit VBI2 (ontbrekende hoogtes worden vervangen door de mediaanwaarde van de hoogtes per plot)
 - hakhoutloten uit VBI2
 - Tabellen met 1 ingang (Dagnelie *et al.*, 1985) voor:
 - hakhoutloten uit VBI1
 - Volume cilinder:
 - niet-intacte bomen uit VBI2 (intact/niet-intact wordt in VBI1 niet onderscheiden)

Hulpvariabelen

- AutochtBasalArea: grondvlak van de inheemse soorten in m³/ha
- ExoticBasalArea: grondvlak van de uitheemse soorten in m³/ha
- InvasiveBasalArea: grondvlak van de invasieve soorten in m³/ha
- TotalBasalArea: totaal grondvlak in m³/ha

Variabelen meetproces en strata

Zie §6.4.1, §6.4.3, §6.5.1 en §6.5.2.

6.6.4 Vegetatie

tbl4Vegetatie

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.

Analysevariabelen

- v0_NV: aanwezigheid van natuurlijke verjonging (NV)
 - NV is aanwezig wanneer er boomsoorten werden waargenomen in de kruidlaag van de vegetatie
 - Voor VBI1 bepalen we de NV enkel op basis van de vegetatieopname en niet op basis van de A1-plot, zodat de resultaten vergelijkbaar zijn met die van VBI2.
- v1_NbTreeSp_VegOpname: aantal boomsoorten in de vegetatieopname
 - Totaal aantal boomsoorten waargenomen tijdens de vegetatieopname (kruidlaag, struiklaag of boomlaag)
 - De soortenlijst die gebruikt wordt voor de vegetatieopname bevat geen veld dat aangeeft of een soort al dan niet een boom is. Voor de dendrometrische opnamen wordt een aparte lijst met bomen gebruikt. De naamgeving is echter niet volledig compatibel met de soortenlijst van de vegetatieopnamen. Hierdoor kan er geen koppeling tussen beide soortenlijsten gemaakt worden. Daarom werd in de databank 'VBIExterneData' een tabel 'tblSpeciesTreelayerCharacteristics' aangemaakt. Deze tabel bevat alle soorten die bij vegetatieopnames genoteerd werden in de boomlaag, maar waarvan de soortnaam niet voorkomt in de bomenlijst van de dendrometrische opnames. In de vegetatieopnames werden immers ook soorten als klimop, bosrank of maretak genoteerd in de boomlaag. Deze tabel bevat daarom een veld 'Tree' waarin aangegeven wordt of een soort al dan niet een boom is.
 - Inlandse eik of soorten die op genusniveau worden bepaald, zoals bijvoorbeeld berk, worden als één soort meegerekend. Als voor een plot zowel inlandse eik als zomereik werd waargenomen, wordt dit ook slechts als één soort meegerekend. Ook bijvoorbeeld berk (G) in combinatie met zachte berk beschouwen we als één soort.
- v1bis_NbTreeSp_Dendro: aantal boomsoorten in de dendrometrische opmetingen (A2-, A3- of A4-plot)
- v2_NbSpecies: aantal soorten die geen bomen zijn, opgemeten in de boom-, stuik-, of kruidlaag van de vegetatieopname. Mossoorten uit VBI1 worden niet meegerekend. In tegenstelling tot in VBI1, worden mossoorten in VBI2 niet geïnventariseerd. Mossoorten kunnen daarom niet meegenomen worden in de vergelijking tussen beide periodes.

- v3_NVExoticPres: aanwezigheid van natuurlijke verjonging van uitheemse boomsoorten
 - De tabel 'tblSpeciesTreelayerCharacteristics' bevat een veld 'Exotic' dat aangeeft welke bomen uitheems zijn. Deze tabel bevat alle soorten waarvan de soortnaam niet voorkomt in de soortenlijst die gebruikt wordt voor de dendrometrische opnames.
 - Voor de soorten waarvan de soortnaam wel terug te vinden is in de soortenlijst van de dendrometrische opnames, is de informatie over uitheems/inheems terug te vinden in de tabel 'tblTreeSpeciesCharacteristics'.
- v4_NVAutochtPres: aanwezigheid van natuurlijke verjonging van inheemse boomsoorten
- v5_NVInvasivePres: aanwezigheid van natuurlijke verjonging van Amerikaanse vogelkers
- v6_NVExoticAbund: abundantie van de natuurlijke verjonging van uitheemse boomsoorten
 - Als maat voor abundantie nemen we de bedekking. We bepalen de bedekking door aan elke klasse van de Braun-Blanquet schaal een bedekking toe te kennen overeenkomstig het gemiddelde van de klasse (Tabel 6-1). We nemen dan de som van de bedekking van alle uitheemse boomsoorten in de kruidlaag.

Tabel 6-1 Bedekking voor de verschillende klassen van de Braun-Blanquet schaal

Klasse	Code	Bedekking (%)
1	r	0,25
2	+	1
3	1	2,25
4	2m	4
5	2a	8,75
6	2b	18,75
7	3	37,5
8	4	62,5
9	5	87,5

- v7_NVAutochtAbund: abundantie van de natuurlijke verjonging van inheemse boomsoorten
- v8_NVInvasiveAbund: abundantie van de natuurlijke verjonging van Amerikaanse vogelkers

- v9_TreeLayer: aanwezigheid van minstens één soort in de boomlaag van de vegetatieopname
- v10_ShrubLayer: aanwezigheid van minstens één soort in de struiklaag van de vegetatieopname
- v11_HerbLayer: aanwezigheid van minstens één soort in de kruidlaag van de vegetatieopname
- v12_NbRareSpecies: aantal zeldzame soorten
 - Een soort is zeldzaam als het in minder dan 10% van de 4 x 4 km-hokken in Vlaanderen voorkomt.
 - Het voorkomen van een soort wordt bepaald op basis van gegevens uit de periode 1972 – 2004 (Van Landuyt *et al.*, 2006)
- v13_SimpsonIndex: indexwaarde van Simpson (D) berekend op basis van de soorten in de kruidlaag

$$D = 1 - \sum_{i=1}^S p_i^2$$

$$p_i = \frac{n_i}{N}$$

- n_i = bedekking van soort i in kruidlaag
 - N = totale bedekking van alle soorten in de kruidlaag
- v14_ShannonWienerIndex: indexwaarde van Shannon-Wiener op basis van de soorten in de kruidlaag

$$H = - \sum_{i=1}^S p_i \ln p_i$$

- v14bis_ShannonWienerIndexExp: exponent van de Shannon-Wiener index. Dit bevordert de interpreteerbaarheid: $\exp(H)$ kan je interpreteren als het 'equivalent aantal soorten' indien alle soorten met gelijke bedekking voorkomen. $\exp(H)$ is dus maximaal gelijk aan de soortenrijkdom.
- v15_NVSoort1Pres ... v19_NVSoort5Pres: aanwezigheid van natuurlijke verjonging in de kruidlaag van een vegetatieopname voor de 5 belangrijkste boomsoorten. De tabel 'tblFiveMainTreeSpecies' in de databank 'VBIExterneData' geeft de 5 belangrijkste soorten aan en kan indien nodig aangepast worden.
- v20_NVSoort1Abund ... v24_NVSoort5Abund: abundantie van de natuurlijke verjonging in de kruidlaag van een vegetatieopname voor de 5 belangrijkste boomsoorten.
- v25_NbSpeciesList: aantal soorten uit een lijst van (karakteristieke) soorten. Als voorbeeld gebruiken we een lijst van sleutelsoorten voor het Natura 2000 habitatype 9120. Deze lijst is terug te vinden in de databank 'VBIExterneData'.
- v26_AbundSpeciesList: som van de abundantie van soorten in een vegetatieopname uit een lijst van (karakteristieke) soorten.
- v27_L: Ellenbergwaarde voor licht, berekend op basis van soorten in de kruidlaag. We berekenen een gewogen gemiddelde op basis van de bedekking van de soorten.

- v28_V: Ellenbergwaarde voor vocht, berekend op basis van soorten in de kruidlaag. We berekenen een gewogen gemiddelde op basis van de bedekking van de soorten.
- v29_R: Ellenbergwaarde voor reactiegetal (pH), berekend op basis van soorten in de kruidlaag. We berekenen een gewogen gemiddelde op basis van de bedekking van de soorten.
- V30_N: Ellenbergwaarde voor stikstof, berekend op basis van soorten in de kruidlaag. We berekenen een gewogen gemiddelde op basis van de bedekking van de soorten.

Hulpvariabelen

- StandAge: bestandsleeftijd. Deze variabele is nodig voor het beantwoorden van vraag 6.3.f

Variabelen meetproces en strata

- Zie §6.4.1, §6.4.3, §6.5.1 en §6.5.2.

6.6.5 Ruimtelijke structuur

TbI5RuimtelijkeStructuur

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.

Analysevariabelen

- v1_ClarkAndEvan: de Positioning Index van Clark and Evan (CE) drukt uit in welke mate een bosbestand afwijkt van een bestand met een compleet gerandomiseerde ruimtelijke verdeling, een zogenaamd Poisson bestand (Wouters, 2008c).
 - Een gerandomiseerd bestand heeft een *CE*-waarde van 1. Als er clusters voorkomen in het bestand, wordt de *CE*-waarde kleiner dan 1. In regelmatig gepositioneerde bestanden wordt *CE*-waarde groter dan 1. De *CE*-waarde wordt berekend op basis van onderstaande formule.

$$CE = \frac{r_A}{r_E} = \frac{\frac{1}{N} \sum_{i=1}^N r_i}{0.5 \left(\frac{A}{N} \right)^{1/2} + 0.0514 \frac{P}{N} + 0.041 \frac{P}{N^{3/2}}}$$

met:

- r_A : de gemiddelde afstand tussen een boom en zijn naaste buur (m)
- r_E : de verwachte gemiddelde afstand als de bomen random gepositioneerd waren (m)

r_i : de afstand tussen boom i en zijn naaste buur (m)

N : totaal aantal bomen in het proefvlak

A : oppervlakte van het proefvlak (m²)

P : perimeter van het proefvlak (m)

- Voor de berekening worden enkel de bomen binnen een straal van 9 m genomen. Vanaf een straal van 9 m tot en met 18 m worden immers enkel bomen met een omtrek groter dan 122 cm opgemeten.
- Er is gekozen om de berekening enkel op plotniveau te doen en niet op segmentniveau. We doen dit om de vergelijkbaarheid van de index te bevorderen. De oppervlakte en de perimeter hebben immers een impact op de waarde van de index.
- De waarde van de CE-index varieert van 0 tot 2,15.
- v2_ClarkAndEvan_simple: een vereenvoudigde variant van de Positioning Index van Clark and Evan waarbij de noemer uit deze index achterwege gelaten wordt. Dit komt neer op de gemiddelde afstand tussen een boom en zijn naaste buur. De vereenvoudiging maakt deze variant makkelijker interpreteerbaar.
- v3_vonGadowSpeciesMix: de mengingsindex van von Gadow is voor iedere boom i gedefinieerd als de kans dat geen van de drie dichtste buren van dezelfde soort is als boom i :

$$DM_i = \frac{1}{3} \sum_{j=1}^3 V_{ij}$$

met:

$V_{ij} = 0$ als boom i en buur j van dezelfde soort zijn

$V_{ij} = 1$ als boom i en buur j niet van dezelfde soort zijn

- DM_i kan dus slechts vier waarden aannemen: 0 - 0,33 - 0,66 of 1. De mengingsindex van het bestand (DM) wordt berekend als het gemiddelde van alle individuele bomen.
- Sterk vertegenwoordigde of sterk geclusterde soorten zullen aanleiding geven tot lage DM -waarden. Lage DM -waarden zijn een aanwijzing voor het voorkomen van homogene groepen van boomsoorten. Minder voorkomende of regelmatig verspreide soorten zullen hogere DM -waarden krijgen. Hoge DM -waarden zijn een aanwijzing voor complete menging van de soort in het bestand (Wouters, 2008c).
- Enkel bomen binnen een straal van 9 m worden meegenomen in de berekeningen. De berekening gebeurt op plotniveau om de impact van schaal- en vormeffecten op de index te beperken.

- v4_vonGadowHeight: De hoogtedifferentiatie-index van von Gadow beschrijft het verschil (of de gelijkens) in hoogte tussen naburige bomen (Wouters, 2008c).
 - FH_i van een boom i met hoogte H_i in relatie tot zijn dichtste buren is gedefinieerd als:

$$FH_i = \frac{1}{3} \sum_{j=1}^3 \left[1 - \frac{MIN(H_i, H_j)}{MAX(H_i, H_j)} \right]$$

met H_i de hoogte van boom i en H_j de hoogte van j^{de} dichtste buur van boom i .

- De gemiddelde waarde van alle bomen geeft de indexwaarde voor het bestand. De waarden liggen tussen 0 (bestanden met nagenoeg geen hoogtedifferentiatie) en 1 (sterk gedifferentieerde bestanden).
- Enkel bomen binnen een straal van 9 m worden meegenomen in de berekeningen. De berekening gebeurt op plotniveau om de impact van schaal- en vormeffecten op de index te beperken

Variabelen meetproces en strata

- Zie §6.4.1, §6.4.3, §6.5.1 en §6.5.2.

6.6.6 Dood hout

tbl6DoodHout

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.

Analysevariabelen

- v1_VolumeSnags: volume staande dode bomen (m³/ha) per segment op basis van:
 - Tabellen met 2 ingangen (Dagnelie *et al.*, 1985; Dik, 1990; Quataert *et al.*, 2011) voor:
 - intacte individuele dode bomen uit VBI2 (ontbrekende hoogtes worden vervangen door de mediaanwaarde van de hoogtes per plot);
 - hakhoutloten van dode bomen uit VBI2.
 - Tabellen met 1 ingang (Dagnelie *et al.*, 1985) voor:
 - Hakhoutloten van dode bomen uit VBI1
 - Individuele dode staande bomen
 - Volume cilinder:
 - Niet-intacte dode bomen uit VBI2 (in VBI1 wordt geen onderscheid gemaakt tussen intact en niet-intact)
- v2_PresenceSnags: aanwezigheid van staande dode bomen.

- v3_VolumeLogs: volume liggende dode bomen (m³/ha).
 - Voor VBI2 wordt dit afgeleid uit de gegevens van de Line Intersect Sampling (LIS) via volgende formule:

$$V_i = \frac{\pi^2}{8L_i} \sum_1^j d_j^2 \frac{1}{\cos(\alpha_j)}$$

met:

- V_i : Geschat bestandsvolume (m³/ha) liggend dood hout dood in steekproefpunt i
 - L_i : Totale horizontale lengte (m) van de transectlijnen (m) in steekproefpunt i
 - d_j : Diametermeting (cm) van het dood-hout-element j loodrecht op de lengteas waar de transectlijn het dood-hout-element snijdt
 - α_j : Helling van het dood-hout-element j t.o.v. de horizontale lijn
- In VBI1 wordt als onderdeel van de vegetatieopname de lengte liggend dood hout binnen het 16 m x 16 m proefvlak geschat voor 4 diameterklassen (2-7 cm, 7-22 cm, 22 -40 cm, >40 cm). Op basis daarvan maken we een ruwe inschatting van het volume liggende dood hout. Dit doen we door per diameterklasse de gemiddelde diameter te nemen en vervolgens het hout-element als een cilinder te beschouwen. Dood hout in de klasse 2-7 cm rekenen we niet mee, omdat in VBI2 enkel liggend dood hout met diameter > 7 cm wordt opgemeten. Voor de diameterklasse > 40 cm nemen we 60 cm als klassegemiddelde.
 - De berekening van het volume liggende bomen per hectare geeft steeds een waarde op plotniveau. Indien de plot uit verschillende segmenten bestaat, dan kennen we de waarde op plotniveau toe aan elk segment.
- v4_PresenceLogs: aanwezigheid van liggend dood hout
 - Voor VBI2 nemen we binnen een plot een steekproef van liggend dood hout via LIS. Het aantal plots waar er via LIS liggend dood hout wordt opgemeten, is dus een onderschatting van het werkelijk aantal plots met aanwezigheid van liggend dood hout.

6.6.7 Authenticiteitsindex

Vermits de gegevensinzameling in VBI1 en VBI2 op een aantal vlakken verschillend is, is ook de scoreberekening voor een aantal categorieën en subcategorieën verschillend. Daarom kiezen we voor aparte tabellen met analysevariabelen voor VBI1 en VBI2. Voor elke categorie van de authenticiteitsindex bespreken we eerst de tabel voor VBI1 en vervolgens deze voor VBI2. Beide tabellen zijn meestal vrij gelijkaardig. De verschillen tussen beide tabellen worden aangeduid in *italic*. We lichten ook steeds toe welke verschillen in bemonsteringsmethodiek tussen VBI1 en VBI2 er aan de oorzaak van liggen. Per categorie wordt het scoresysteem

weergegeven (Tabel 6-2, Tabel 6-3, Tabel 6-4, Tabel 6-5). Voor meer details verwijzen we naar Van Den Meersschaut *et al.* (2001). Afwijkingen t.o.v. dit scoresysteem worden hieronder toegelicht.

6.6.7.1 Bosstructuur

Tabel 6-2 Berekening van de structuurindex = a+b+c+d (maximale score = 20)

Categorie	Klasse	Score
Sluitingsgraad (a)	> 2/3	2
	1/3 - 2/3	4
	< 1/3	3
Leeftijd (b)	0 jaar	0
	1 - 60 jaar	1
	61 - 100 jaar	2
	101 - 160 jaar	5
	> 160 jaar	7
	ongelijk jarig	5
Verticale structuur (c)	1 etage	2
	> 1 etage	4
Horizontale structuur (d)	homogeen	1
	groepsgewijs	3
	stamsgewijs	5

tblAIForestStructureVBI1

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.

Hulpvariabelen

- CanopyClosure: kroonsluitingscategorieën
- StandAge: bestandsleeftijdcategorieën
- MixtTpe: mengingsvormcategorieën
- LayersN: aantal etages.

Voor elke boom opgemeten in de A3- of A4-cirkel wordt bepaald of deze domineert of gedomineerd wordt. Het aantal etages is gelijk aan 2 wanneer er zowel dominante en gedomineerde bomen zijn. Het aantal etages is gelijk aan 1 wanneer er enkel dominante bomen zijn. Het aantal etages is gelijk aan 0 als er geen bomen zijn (met diameter > 7cm).

Analysevariabelen

- v1_CanopyClosureScore: score voor sluitingsgraad
- v2_StandAgeScore: score voor bestandsleeftijd
- v3_MixTypeScore: score voor gelaagdheid
- v4_LayersScore: score voor boomsoortenmenging
- v5_TotalScore: som van de scores van de subcategorieën voor bosstructuur.
- MaxScore: de potentieel maximale score rekening houdend met ontbrekende waarden. Als er bijvoorbeeld gegevens ontbreken voor sluitingsgraad is de maximaal haalbare score 16 (7 + 4 +5) (zie tabel 6-2). We kunnen de totale score hier eventueel voor corrigeren.

tblAIForestStructureVBI2

De verschillen t.o.v. tblForestStructureVBI1 zijn het gevolg van:

- De bestandsbeschrijving in VBI2 gebeurt per segment.
- De dominantie van bomen wordt niet meer bepaald in VBI2. Als alternatief kijken we naar de aanwezigheid van de struiklaag om de gelaagdheid te beoordelen.

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van segment in plot.

Hulpvariabelen

- CanopyClosure: kroonsluitingscategorieën
- StandAge: bestandsleeftijdcategorieën
- MixtTpe: mengingsvormcategorieën
- CanopyClosureCode: code voor kroonsluitingscategorieën
- StandAgeCode: code voor bestandsleeftijdcategorieën
- MixtTypeCode: code voor mengingsvormcategorieën
- ShrubCover: bedekking van de struiklaag (categorieën)
- ShrubCoverCode: code voor bedekking van de struiklaag

Analysevariabelen

- v1_CanopyClosureScore: score voor de kroonsluiting. *Gezien in VBI2 andere categorieën worden gehanteerd, worden scores als volgt toegekend*
 - 0 – 25% → 3
 - 25 – 50% → 4
 - 50 -75% → 4
 - 75 – 100% → 2
- v2_StandAgeScore: score voor leeftijd
- v3_MixTypeScore: score voor gelaagdheid. *Indien de struiklaag meer dan 5% bedekt, beschouwen we dit als een extra etage.*
- v4_LayersScore: score voor boomsoortenmenging

- v5_TotalScore: som van de scores van de subcategorieën voor bosstructuur.
- MaxScore: de potentieel maximale score rekening houdend met ontbrekende waarden. Als er bijvoorbeeld gegevens ontbreken voor kroonsluiting, is de maximaal haalbare score 16 (7 + 4 +5) (zie tabel 6-2).

6.6.7.2 Houtige vegetatie

tblAIWoodLayerVBI1

IDVariabelen

- IDPlots ID van het rasterpunt waar de plot gelegen is.

Hulpvariabelen

- A3N: aantal levende bomen opgemeten in een A3-plot met een omtrek tussen 22 cm en 120 cm.
- A4N: aantal levende bomen opgemeten in een A4-plot met een omtrek > 120 cm.
- LargeTreesN: aantal zware levende bomen met diameter tussen 40 cm en 80 cm.
- VeryLargeTreesN: aantal zeer zware levende bomen met diameter > 80 cm.
- sdDBH: standaarddeviatie van de diameter op borsthoogte van alle **levende** bomen in de plot.
- treeSpeciesN: aantal boomsoorten in de boomlaag op basis van metingen in de A2-, A3- en A4-plots. Exoten worden enkel meegeteld als hun stamtal- en grondvlakaandeel < 5%.
- PenaltyExoticSpecies: strafpunt voor exoten waarvan het stamtal- of grondvlakaandeel > 50%.
- A2N: aantal bomen in een A2 plot met een omtrek < 22 cm en een hoogte > 2 m)
- SpeciesNatRegN: aantal boomsoorten in de natuurlijke verjonging op basis van gegevens uit de A1-plot.

Analysevariabelen

- v1_LargeTreesNScore: score voor dikke bomen.
- v2_VeryLargeTreesNScore: score voor zeer dikke bomen.
- v3_sdDBHScore: score voor de standaard deviatie van de diameters. Voor plots met 1 of geen enkele boom met een omtrek > 22cm kan geen sd berekend worden. Aan dergelijke plots kennen we een score 0 toe.
- v4_treeSpeciesNScore
- v5_SpeciesNatRegNscore

Tabel 6-3 Houtige vegetatie-index = e + f + g + h + i (maximale score = 25)

Categorie	Klasse	Score
Aantal boomsoorten ⁽¹⁾ (hoogte > 2 m) (e)	1 – 2	1
	3 – 4	2
	5 – 6	3
	7 – 8	4
	> 8	5
Aantal dikke bomen (40 cm < diameter < 80 cm) (f)	1 – 5	1
	6 – 10	2
	11 – 15	3
	15 – 20	4
	> 20	5
Aantal zeer dikke bomen (diameter ≥ 80 cm) (g)	1	3
	2 – 3	4
	≥ 4	5
Aantal inheemse boomsoorten in de natuurlijke verjonging (hoogte < 2 m) (h)	1 – 4	1
	5 – 8	2
	9 – 12	3
	> 12	4
Standaard afwijking van diameter (diameter ≥ 7 cm) (i)	< 10 cm	0
	10 – 15 cm	1
	16 – 20 cm	2
	21 – 25 cm	3
	26 – 30 cm	4
	31 – 35 cm	5
	> 35 cm	6

⁽¹⁾ Exoten met een grondvlucaandeel ≤ 5% of een stamtalaandeel < 50% worden niet meegeteld; voor exoten met een grondvlak- of stamtalaandeel > 50 % wordt een strafpunt (-1) gerekend bij het aantal boomsoorten.

tbIAIWoodLayerVBI2

Verschillen t.o.v. tblWoodLayerVBI1 zijn het gevolg van:

- In VBI2 gebeurt geen gegevensinzameling van natuurlijke verjonging in de A1 plot (met straal van 2,25 meter waarin alle bomen < 2 m worden geteld). Natuurlijke verjonging

wordt in VBI2 ingeschat via de vegetatieopname in een vegetatieplot van 16 x 16 m. Daarbij worden de aanwezigheid en de bedekking van boomsoorten in de kruidlaag geschat.

- Het aantal boomsoorten in natuurlijke verjonging is moeilijk vergelijkbaar tussen beide inventarisatierondes:
 - De oppervlakte van de proefvlakken waarin natuurlijke verjonging bepaald wordt verschilt.
 - De definitie van natuurlijke verjonging verschilt. In VBI1 worden alle bomen < 2 m beschouwd als natuurlijke verjonging, in VBI2 alle bomen in de kruidlaag, waarbij kruidlaag gedefinieerd wordt als alle planten < 0,5 m.
- Bijgevolg moet een ander scoresysteem toegepast worden voor de authenticiteitsindex.

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.

Hulpvariabelen

- A3N: aantal levende bomen *met diameter tussen 7 cm en 39 cm* opgemeten in de A3-plot.
- A4N: aantal levende bomen *met diameter > 39 cm* opgemeten in de A4-plot.
- LargeTreesN: aantal zware levende bomen met diameter tussen 40 cm en 80 cm.
- VeryLargeTreesN: aantal zeer zware levende bomen met diameter > 80 cm)
- sdDBH: standaarddeviatie van de diameter op borsthoogte van alle **levende** bomen in de plot.
- treeSpeciesN: aantal boomsoorten in de boomlaag op basis van metingen in de A2-, A3- en A4-plot. Exoten worden enkel meegeteld als hun stamtaal- en grondvlakaandeel < 5%. In principe zouden hier ook de bomen uit de boomlaag van het vegetatieplot meegeteld moeten worden. Dit is nog niet gebeurd door praktische moeilijkheden bij het koppelen van de soortenlijst gebruikt bij de vegetatieopname en de soortenlijst gebruikt voor de dendrometrische metingen.
- PenaltyExoticSpecies: strafpunt voor exoten waarvan het stamtaal- of grondvlakaandeel > 50%.
- A2N: aantal bomen in de A2 plot *met diameter < 7 cm* en hoogte > 2 m.
- SpeciesNatRegN: *aantal boomsoorten in natuurlijke verjonging op basis van gegevens uit de kruidlaag van de vegetatieopname.*

Analysevariabelen

- v1_LargeTreesNScore
- v2_VeryLargeTreesNScore
- v3_sdDBHScore: voor plots met 1 of geen enkele boom met omtrek > 22 cm kan geen sd berekend worden. Aan dergelijke plots kennen we een score 0 toe.
- v4_treeSpeciesNScore
- v5_SpeciesNatRegNScore

6.6.7.3 Kruidlaag

Tabel 6-4 Kruidvegetatie-index = j + k + l + m (maximale score = 25)

Variabele	Waarde	Score
Aantal soorten vaatplanten (j)	1 – 5	1
	6 – 10	2
	11 - 15	3
	16 – 20	4
	21 – 25	5
	26 – 30	6
	31 - 35	7
	36 – 40	8
	41 – 45	9
	> 45	10
Zeldzaamheid (k)	1 – 5	1
	6 – 10	2
	11 – 15	3
	16 – 20	4
	21 – 50	5
	26 – 30	6
	> 30	7
Aantal mossoorten (l)	1 – 5	1
	6 – 10	2
	11 – 15	3
	16 – 20	4
	> 20	5
Bedekking (m)	< 5%	0
	6 – 25%	1
	26 – 50%	2
	51 – 75%	3
	> 75%	1

tblAIHerbLayerVBI1

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.

Hulpvariabelen

- CoverHerbLayer: totale bedekking van de kruidlaag.
- CoverBryophLayer: totale bedekking van de moslaag.
- PlantSpeciesN: aantal plantensoorten in de kruidlaag.
- RarenessSum: som (per plot) van de zeldzaamheidsscores per plantensoort. De zeldzaamheidsscore is gebaseerd op de frequentie van voorkomen van een soort in de uurhokken (4 x 4 km-hokken) in Vlaanderen. Hiervoor werden atlasgegevens van de periode 1972 - 2004 gebruikt.
- *BryophSpeciesN: aantal mossoorten.*

Analysevariabelen

- v1_PlantSpeciesNScore: score voor het aantal plantensoorten.
- v2_RarenessScore: score voor zeldzaamheid.
- v3_BryophSpeciesNScore: *score voor het aantal mossoorten.*
- v4_CoverHerbLayerScore: score voor de totale bedekking. In principe omvat de totale bedekking zowel de moslaag als de kruidlaag. De bedekking van deze twee lagen werden enkel apart ingeschat en kan niet zomaar opgeteld worden. We beschouwen daarom de bedekking van de kruidlaag als de totale bedekking.

tblAIHerbLayerVBI2

Verschillen t.o.v. tblHerbVBI1 zijn het gevolg van:

- In VBI1 werd slechts in de helft van de meetpunten een vegetatieopname uitgevoerd. In VBI2 gebeurt dit in elk meetpunt.
- In VBI2 gebeurt geen vegetatieopname van de moslaag. Het aantal soorten in de moslaag kan dus niet bepaald worden.

IDVariabelen

- IDPlots ID van het rasterpunt waar de plot gelegen is.

Hulpvariabelen

- CoverHerbLayer: totale bedekking van de kruidlaag. *Bedekking wordt ingeschat a.d.h.v. vaste categorieën.*
- CoverBryophLayer: totale bedekking van de moslaag. *Bedekking wordt ingeschat a.d.h.v. vaste categorieën.*
- *CoverHerbLayerCode: code voor de totale bedekking van de kruidlaag.*
- *CoverBryophLayerCode: code voor de totale bedekking van de moslaag.*
- PlantSpeciesN: aantal plantensoorten in kruidlaag.

- **RarenessSum:** som (per plot) van de zeldzaamheidsscores per plantensoort. De zeldzaamheidsscore is gebaseerd op de frequentie van voorkomen van een soort in de uurhokken (4 x 4 km-hokken) in Vlaanderen. Hiervoor werden gegevens van de periode 1972 - 2004 gebruikt (referentie Atlas van de flora).

Analysevariabelen

- **v1_PlantSpeciesNScore:** score voor aantal plantensoorten.
- **v2_RarenessScore:** score voor zeldzaamheid.
- **v3_CoverHerbLayerScore:** score voor totale bedekking. In principe omvat de totale bedekking zowel de moslaag als de kruidlaag. In VBI2 ontbreekt de moslaag, hiervoor moeten we corrigeren.

6.6.7.4 Dood hout

tblAIDeadWoodVBI1

IDVariabelen

- **IDPlots:** ID van het rasterpunt waar de plot gelegen is.

Hulpvariabelen

- **SnagsBasalAreaHa:** grondvlak van staand dood hout per hectare.
- **SnagsLargeTreesN:** aantal zware staande dode bomen met diameter groter dan 40 cm
- **SnagsTreesN:** aantal staande dode bomen.
- **SnagsSdDBH:** standaard afwijking van de diameter op borsthoogte.
- **LogsLargeTreesLength:** stamlengte van liggende dode zware bomen (met diameter > 40 cm)
- **LogsTreesDiamClass1:** aanwezigheid van liggend dood hout in diameterklasse 2 – 7 cm.
- **LogsTreesDiamClass2:** aanwezigheid van liggend dood hout in diameterklasse 7 – 22 cm.
- **LogsTreesDiamClass3:** aanwezigheid van liggend dood hout in diameterklasse 22 – 40 cm.
- **LogsTreesDiamClass4:** aanwezigheid van liggend dood hout in diameterklasse > 40 cm.
- **LogsTreesDiamClassN:** aantal aanwezige diameterklassen van liggend dood hout. Gezien in VBI2 de aanwezigheid van liggend dood hout in de kleinste diameterklasse niet wordt bepaald, beschouwen we deze klasse altijd als aanwezig.

Analysevariabelen

- **v1_SnagsBasalAreaScore:** score voor het grondvlak van staand dood hout
- **v2_SnagsLargeTreesNScore:** score voor het aantal staande, dode, zware bomen
- **v3_SnagsSdDBHScore:** score voor de standaard afwijking van de diameters van staand dood hout. Indien er 1 of geen staande dode boom aanwezig is, wordt de score gelijkgesteld aan 0.
- **v4_LogsLargeTreesLengthScore:** score voor de stamlengte van zware bomen voor liggend dood hout.

- v5_LogsDiamClassNScore: score voor het aantal diameterklassen van liggend dood hout.

Tabel 6-5 Doodhout-index = n + o + p + q + r (maximale score =30)

Categorie	Variabele	Waarde	Score
Staand dood hout	Grondvlak (n)	< 2 m ² /ha	1
		2 - 3,5 m ² /ha	2
		3,6 - 5 m ² /ha	3
		> 5 m ² /ha	4
	Standaard afwijking diameter (diameter ≥ 7 cm) (o)	< 10 cm	0
		10 – 15 cm	1
		16 - 20 cm	2
		21 - 25 cm	3
		26 - 30 cm	4
		31 – 35 cm	5
		> 35 cm	6
	Aantal dikke bomen (diameter > 40 cm)	1	3
		2 – 3	4
		4	5
	Liggend dood hout	Gesommeerde stamlengte van dikke bomen (diameter > 40 cm) (q)	1 – 10 m
11 – 20 m			5
> 20 m			7
Aantal diameterklassen (r)		1	2
		2	4
		3	6
		4	8
		Klasse 4 aanwezig en aantal diameterklassen = 1, 2 of 3	+ 1

tblAIDeadWoodVBI2

Verschillen t.o.v. tblDeadWoodVBI1 zijn het gevolg van:

- Liggend dood hout wordt in VBI2 bepaald via Line Intersect Sampling (LIS). Enkel liggend dood hout met een diameter > 7 cm wordt opgemeten. Dit betekent dat de

aanwezigheid van liggend dood hout in de kleinste diameterklasse niet kan bepaald worden. We beschouwen deze kleinste diameterklasse daarom steeds als aanwezig. In VBI1 werd de lengte van dood hout ingeschat in het 16 x 16 m vegetatieproefvlak. In VBI2 kan via LIS de lengte van liggend dood hout per ha geschat worden. Hieruit kan dan de lengte per 16 x 16 m proefvlak berekend worden om de vergelijking met VBI1 mogelijk te maken.

- De LIS methode is bedoeld om een schatting te bekomen op regionale schaal. Op niveau van een individueel plot zal deze methode slechts tot een zeer ruwe schatting van liggend dood hout komen.
- In VBI1 gebeurde de gegevensinzameling voor liggend dood hout gezamenlijk met de vegetatieopname en daardoor ook slechts in de helft van de meetpunten.

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.

Hulpvariabelen

- SnagsBasalAreaHa: grondvlak van staand dood hout per hectare.
- SnagsLargeTreesN: aantal zware staande dode bomen met diameter groter dan 40 cm.
- SnagsTreesN: aantal staande dode bomen.
- SnagsSdDBH: standaard afwijking van de diameter op borsthoogte.
- LogsLargeTreesLength: stamlengte van liggende dode zware bomen (met diameter > 40 cm). *Deze lengte werd geschat conform de Line Intersect Method. De lengte wordt uitgedrukt per 16 x 16 m oppervlak om vergelijking met VBI1 mogelijk te maken.*
- LogsTreesDiamClass2: aanwezigheid van liggend dood hout in diameterklasse 7 – 22 cm.
- LogsTreesDiamClass3: aanwezigheid van liggend dood hout in diameterklasse 22 – 40 cm.
- LogsTreesDiamClass4: aanwezigheid van liggend dood hout in diameterklasse > 40 cm.
- LogsTreesDiamClassN: aantal aanwezige diameterklassen van liggend dood hout. Omdat in VBI2 de aanwezigheid van liggend dood hout in de kleinste diameterklasse niet wordt bepaald, beschouwen we deze klasse altijd als aanwezig.

Analysevariabelen

- v1_SnagsBasalAreaScore: score voor het grondvlak van staand dood hout.
- v2_SnagsLargeTreesNScore: score voor het aantal staande, dode, zware bomen.
- v3_SnagsSdDBHScore: score voor de standaard afwijking van de diameters van staand dood hout. Indien er 1 of geen staande dode boom aanwezig is wordt de score gelijkgesteld aan 0.
- v4_LogsLargeTreesLengthScore: score voor de stamlengte van zware bomen voor liggend dood hout.
- v5_LogsDiamClassNScore: score voor het aantal diameterklassen van liggend dood hout.

6.6.8 Aanwas

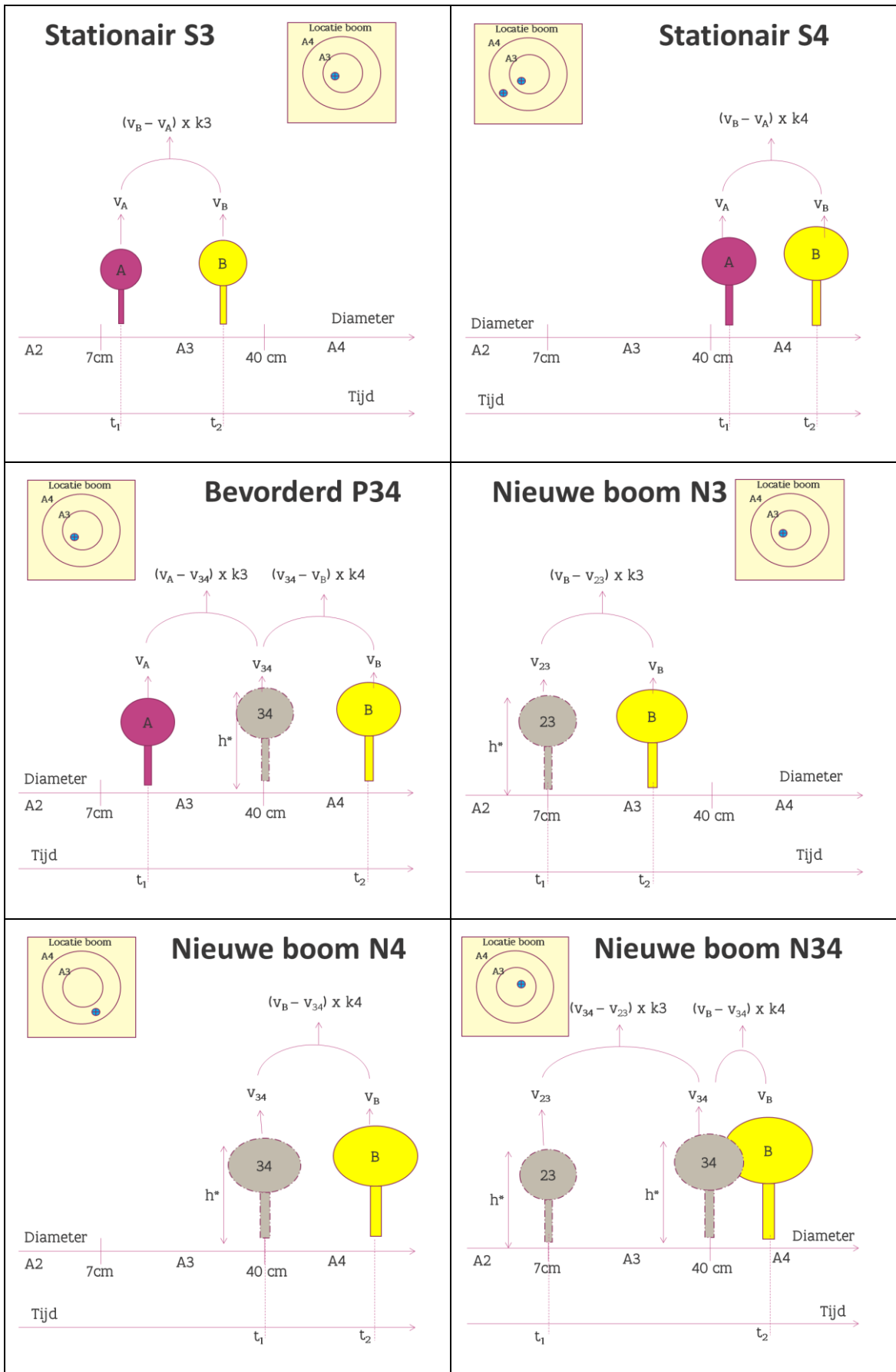
6.6.8.1 Aanwasberekening

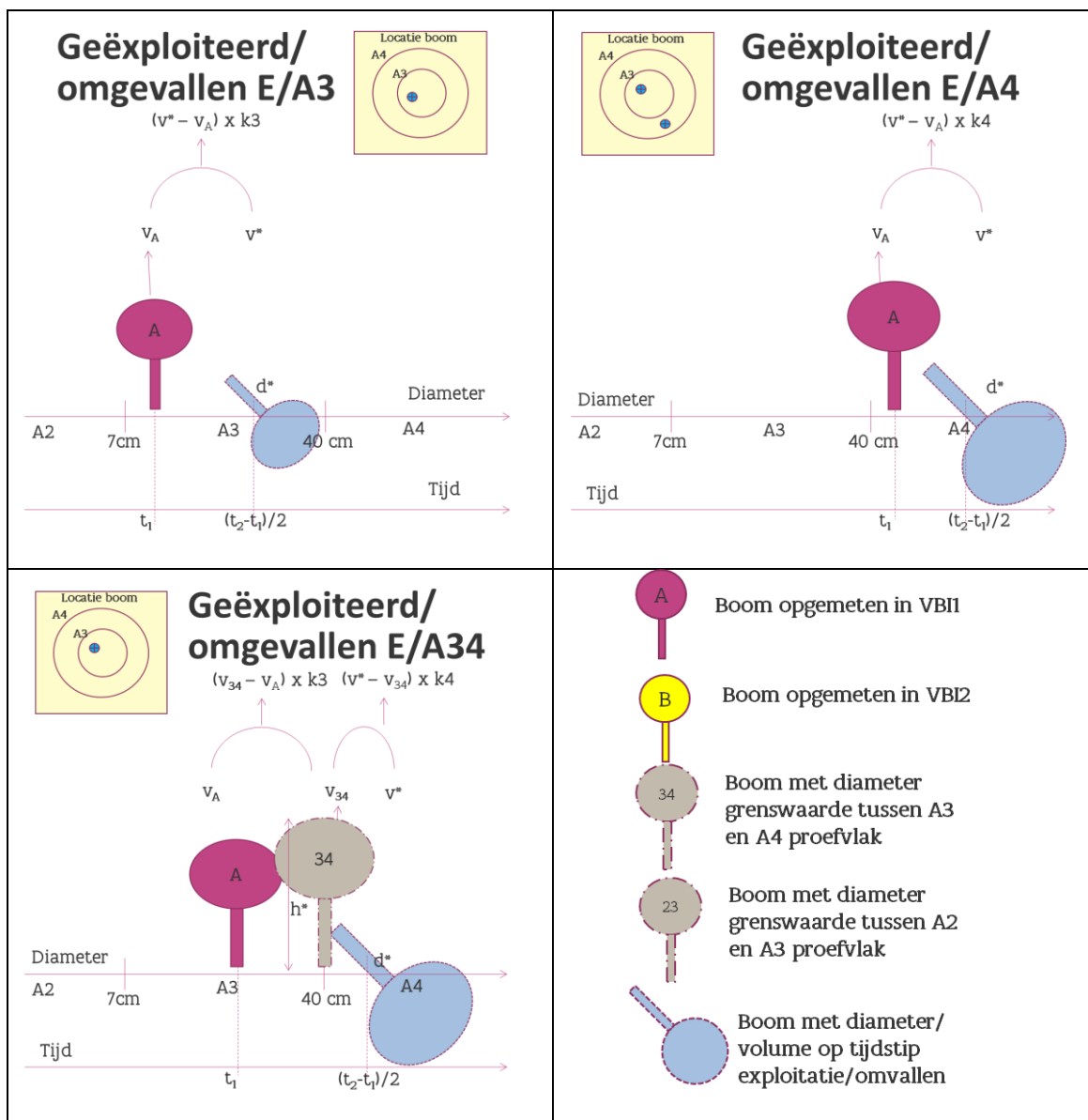
De berekening van de lopende jaarlijkse aanwas per hectare (LJA) gebeurt op basis van alle bomen in de gepaarde plots volgens de methode beschreven in Wouters *et al.* (2008b). Deze methode is samengevat in Tabel 6-6.

Tabel 6-6 Overzicht van de verschillende type bomen en de berekeningswijze van de lopende jaarlijkse aanwas per hectare

Kwalificatie van de boom	Afkorting	Aanwezigheid van de boom		Proefcirkel		Formule	Opmerkingen
		Inv. A	Inv. B	Inv. A	Inv. B		
Stationair	S3	ja	ja	3	3	$k_3(v_B - v_A) / \Delta t$	Boom opgemeten in cirkel 3 in periode A en B
	S4	ja	ja	4	4	$k_4(v_B - v_A) / \Delta t$	Boom opgemeten in cirkel 4 in periode A en B
Bevorderd	P34	ja	ja	3	4	$[k_3(v_{34} - v_A) + k_4(v_B - v_{34})] / \Delta t$	Boom opgemeten in cirkel 3 in periode A en cirkel 4 in periode B
Nieuw opgemeten	N3	neen	ja	-	3	$k_3(v_B - v_{23}) / \Delta t$	Boom binnengekomen in cirkel 3 en in periode B in cirkel 3 opgemeten
	N4	neen	ja	-	4	$k_4(v_B - v_{34}) / \Delta t$	Boom binnengekomen in cirkel 4 en in periode B in cirkel 4 opgemeten
	N34	neen	ja	-	3 vervolgens 4	$[k_3(v_{34} - v_{23}) + k_4(v_B - v_{34})] / \Delta t$	Boom binnengekomen in cirkel 3 en in periode B opgemeten in cirkel 4
Geëxploiteerd/afgestorven	E/A3	ja	neen	3	-(3)	$k_3(v_{est} - v_A) / \Delta t$	Boom opgemeten in periode A (cirkel 3) en de periode van afsterven/kapping geschat in dezelfde cirkel 3
	E/A34	ja	neen	3	-(4)	$[k_3(v_{34} - v_A) + k_4(v_{est} - v_{34})] / \Delta t$	Boom opgemeten in periode A (cirkel 3) en de periode van afsterven/kapping geschat in cirkel 4
	E/A4	ja	neen	4	-(4)	$k_4(v_{est} - v_A) / \Delta t$	Boom opgemeten in periode A (cirkel 4) en de periode van afsterven/kapping geschat in dezelfde cirkel 4

De verschillende situaties en bijhorende formules voor de aanwasberekening zijn schematisch weergegeven in Figuur 3.





Figuur 3 Schematische weergave van de aanwasberekening voor verschillende situaties (v_A =volume boom op tijdstip t_1 , v_B = volume boom op tijdstip t_2 , v_{23} = volume van een boom met diameter van 7 cm, v_{34} = volume van een boom met diameter van 40 cm, h^* =geschatte hoogte, v^* = geschat volume van een boom op moment van exploitatie/ omvallen, d^* = geschatte diameter van een boom op moment van exploitatie/ omvallen).

De berekening gebeurt op basis van volgende stappen:

- Uit de bomen die opgemeten werden in VBI1 worden enkel de levende bomen geselecteerd. Enkel levende bomen kunnen immers aanwas maken. Staande bomen die tussen de periode van VBI1 en VBI2 zijn afgestorven, worden wel meegenomen in de berekening van de aanwas.
- Gekoppelde bomen die in VBI1 een individuele stam hebben en in VBI2 naar hakhout geëvolueerd zijn, worden niet meegenomen in de berekening. Enerzijds is er een grote kans dat ze foutief gekoppeld zijn, anderzijds is de berekening van de aanwas van dergelijke bomen niet eenvoudig. Ook hakhoutbomen uit VBI2 die niet gekoppeld zijn,

worden niet meegerekend. Vermits in VBI1 hakhout enkel in het A2-proefvlak werd bepaald, kunnen we het type boom niet onderscheiden (stationair/bevorderd of nieuwe boom). De berekende aanwas slaat dus enkel op de aanwas van individuele stammen.

- **Onderscheid van het type boom** op basis van de aan-/afwezigheid en de omtrek van de boom in VBI1 en VBI2. Voor het onderscheiden van N4 en N34 wordt de afstand van de boom tot het centrum van het proefvlak gebruikt: een afstand kleiner dan 9 meter komt overeen met een N34 boom, een grotere afstand komt overeen met een N4 boom.
- Schatting van het volume en de diameter van een geëxploiteerde of afgestorven boom op het tijdstip van exploitatie of afsterven (v_{est} en d_{est}). Gezien dit tijdstip niet gekend is, nemen we steeds het midden van de periode tussen beide metingen.

- We modelleren de jaarlijkse toename in volume/diameter in functie van het volume/de diameter tijdens de eerste meting via een 2^{de} graadsvergelijking:

$$\frac{\Delta v}{\Delta t} = a + bv_{t_0} + cv_{t_0}^2$$

$$\frac{\Delta d}{\Delta t} = a + bd_{t_0} + cd_{t_0}^2$$

- Deze modellen stellen we op per boomsoort op basis van alle gekoppelde bomen (de stationaire of bevorderde bomen) voor die boomsoort.
- We stellen ook een algemeen model op voor alle boomsoorten samen. Dit model gebruiken we voor boomsoorten met minder dan 20 gekoppelde bomen. Voor dergelijke boomsoorten is een apart model immers weinig betrouwbaar gezien het beperkt aantal observaties.
- Vervolgens berekenen we voor elke afgestorven/geëxploiteerde boom Δv en Δd op basis van v_{t_0} , d_{t_0} en Δt . Dan is $v_{est} = \Delta v + v_{t_0}$ en $d_{est} = \Delta d + d_{t_0}$.
- d_{est} wordt gebruikt om onderscheid te maken tussen E/A3 en E/A34 (zie Tabel 6-6).
- Schatting van het volume van een boom met een diameter gelijk aan de ondergrens van A3 (v_{23}) en A4 (v_{34}). De diameters van deze bomen zijn respectievelijk 22 cm en 40 cm. De hoogte is niet gekend en wordt geschat.
 - Voor de bevorderde bomen (P34) nemen we aan dat de hoogte evenredig toeneemt met de diameter. De hoogte bij een diameter van 22 cm of 40 cm kan dan bekomen worden via interpolatie.

$$h_{23} = h_{t_0} + \frac{\Delta h}{\Delta d}(d_{23} - d_{t_0})$$

- Voor de nieuw opgemeten bomen kunnen we niet interpoleren omdat er maar één meting is. We kijken dan op plotniveau naar de jaarlijkse toename in hoogte van alle gekoppelde bomen en nemen hieruit de mediaanwaarde. Deze gebruiken we om de hoogte van de nieuw opgemeten bomen te schatten op

het tijdstip van de eerste meting. Vervolgens schatten we de hoogte bij een diameter van 22 cm of 40 cm via interpolatie (op dezelfde wijze als voor P34).

- Op basis van de diameter en de geschatte hoogte wordt het volume geschat op basis van de tarieven met twee ingangen.
- Op basis van het volume van een boom in VBI1 en in VBI2 en op basis van de expansiefactoren k_3 (10000/oppervlakte A3-proefvlak) en k_4 (10000/oppervlakte A4-proefvlak) kan vervolgens per type boom de aanwas berekend worden op basis van de formules in Tabel 6-6. De aanwas wordt vervolgens gesommeerd per plot.

6.6.8.2 Analysedatabank

tbl8Aanwas

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.

Analysevariabelen

- $v1_LJA$: lopende jaarlijkse aanwas in $m^3/ha/jaar$ voor individuele bomen (dus exclusief hakhout).
- $v2_DeadStanding$: afgestorven bestandsvolume (enkel de staande dode bomen) in $m^3/ha/jaar$. Dit komt overeen met het volume van de staande dode bomen van gepaarde plots in VBI2 gedeeld door het aantal jaren tussen beide metingen.
- $v3_CutOrFallenTrees$: bestandsvolume van geëxploiteerde of omgevallen bomen in $m^3/ha/jaar$. Dit komt neer op het volume van de geëxploiteerde of omgevallen bomen (v_{est}) gedeeld door het aantal jaren tussen beide metingen.

Hulpvariabelen

- $LJA_PairedTrees$: LJA voor gekoppelde bomen (stationair of bevorderd).
- $LJA_NewTrees$: LJA voor nieuwe bomen.
- $LJA_EATrees$: LJA voor geëxploiteerde of omgevallen bomen.

Variabelen meetproces en strata

- Zie §6.4.1, §6.4.3, §6.5.1 en §6.5.2.

6.6.8.3 Bespreking resultaten

Tabel 6-7 geeft de eerste resultaten van de aanwasberekeningen. Opvallend is de hoge waarde voor LJA ($12,7 m^3/ha/jaar$). De gekoppelde bomen zorgen voor $7,9 m^3/ha/jaar$ of iets meer dan 60% procent van de totale LJA.

Tabel 6-7 Resultaten van de aanwasberekeningen

Variabele	Gemiddelde (m ³ /ha/jaar)
v1_LJA	12,7
LJA_PairedTrees	7,9
LJA_NewTrees	2,4
LJA_EATrees	2,3
v2_DeadStanding	0,5
v3_CutOrFallenTrees	7,1

Een mogelijke overschatting kan gebeurd zijn doordat bomen in gepaarde plots niet gekoppeld werden. Een boom opgemeten in VBI2 wordt m.a.w. als een nieuwe boom aanzien terwijl de boom al opgemeten werd in VBI1. Hierdoor wordt eenzelfde boom als een E/A-boom en als een nieuwe boom meegerekend en krijgen we een overschatting van de aanwas voor die boom. Om dubbeltelling volledig uit te sluiten, kunnen we de som nemen van de LJA van gepaarde bomen met LJA_NewTrees. Dit geeft dan een minimale inschatting van de totale LJA waarbij we zeker zijn dat er geen dubbeltelling is gebeurd.

Een andere bron van overschatting van de LJA is het niet volledig representatief zijn van de gepaarde plots voor Vlaanderen. Het blijkt immers dat plots in naaldhoutbestanden een grotere kans hebben om teruggevonden te worden dan plots in loofhoutbestanden. Deze naaldhoutbestanden worden gekenmerkt door een hogere aanwas.

Daarnaast zijn er ook bronnen van onderschatting van de LJA. Zo wordt hakhout niet meegerekend in de aanwas. Bovendien wordt bij de berekening van de volumeaanwas van nieuwe bomen steeds een minimale toename in diameter verondersteld. Dit doen we door aan te nemen dat de boom op het tijdstip van de eerste meting een diameter had gelijk aan de ondergrens van de A3- of A4-plot (zie formules in Tabel 6-6).

6.6.9 Houtkwaliteit

Deze tabel bevat enkel gegevens van VBI2. Tijdens VBI1 zijn geen gegevens ingezameld die nodig zijn om de meetvragen te beantwoorden.

tbl9HoutKwaliteit

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.
- IDTrees_2eBosinv: ID van boom

Analysevariabelen

- v1_QualityClass: industrie-, brand- en pulphout (klasse C), kwaliteitsvol hout (klasse B) of hoog kwaliteitsvol hout (klasse A).
 - Er is nog geen classificatie opgesteld om op basis van de ingezamelde kwaliteitskarakteristieken de verschillende kwaliteitsklassen te onderscheiden.

Via de tabel 'tblWoodQualityClassification' in de databank 'VBIExterneData' laten we de gebruiker toe om zelf een classificatie op te stellen. Via deze tabel kunnen de minimumvereisten per kwaliteitskenmerk en per boomsoort weergegeven worden waaraan een bepaalde kwaliteitsklasse moet voldoen. Enkel de minimumvereisten voor klasse A en klasse B worden opgenomen in de tabel. Indien een boom niet voldoet aan de minimumvereisten van een van beide klassen, dan veronderstellen we dat deze tot klasse C behoort.

- De classificatie voor naaldhoutsoorten (grove den, Corsicaanse den, lork en douglas) gebeurt op basis van:
 - aantal stamdefecten (0,1 of meerdere)
 - stamverloop (gering, aanvaardbaar, aanzienlijk)
 - diameter.
- Voor de classificatie van loofboomsoorten (eik, beuk, berk en populier) worden volgende bijkomende criteria gebruikt:
 - stamhoogte van de eerste zijtak van minstens 2 cm diameter
 - takhoek van de eerste zijtak van minstens 10 cm diameter.
- v2_Volume: volume van de boom, berekend zoals beschreven in § 6.6.3.
- v3_StemDefect: type stamdefect.
- v4_StamType:
 - Stamtype 1: takvrije stammen van minimaal 6 m lengte zonder defecten, met een gering verloop en met een minimale onderdiameter.
 - Stamtype 2: takvrije stammen van minimaal 6 m lengte met maximaal 1 defect en/of een aanvaardbaar verloop en een minimale onderdiameter.
 - Stamtype 3: takvrije stammen van minimaal 6 m lengte die niet aan de criteria van stamtype 1 en stamtype 2 voldoen.
 - Stamtype 4: takvrije stammen korter dan 6 m zonder defecten, met een gering verloop en met een minimale onderdiameter.
 - Stamtype 5: takvrije stammen korter dan 6 m met maximaal 1 defect en/of een aanvaardbaar verloop en een minimale onderdiameter.
 - Stamtype 6: takvrije stammen korter dan 6 m die niet aan de criteria van stamtype 4 en stamtype 5 voldoen.
- v5_VolumeTakvrij: volume takvrije stam, berekend als het volume van een cilinder met als hoogte de stamhoogte ter hoogte van de eerste zijtak met een diameter van minstens 2 cm.

Variabelen meetproces en strata

- Zie §6.4.1, §6.4.3, §6.5.1 en §6.5.2.

6.6.10 Boomsoortensamenstelling

Deze tabel laat toe om per boomsoort het stamtal, het grondvlak en het volume per hectare te analyseren.

tbl10Boomsoorten

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- IDSegments: ID van het segment binnen de plot.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.
- IDTreeSp: ID voor boomsoort

Analysevariabelen

- v1_StamtalA3A4_ha: stamtal (1/ha) in A3- of A4-plot per segment/plot en per boomsoort .
- v2_BasalArea_ha: grondvlak (m²/ha) per segment/plot en per boomsoort.
- v3_Volume_ha: volume (m³/ha) per segment/plot en per boomsoort. Voor berekeningswijze zie §6.6.3.

Hulpvariabelen

- StamtalA2_ha
- Stamtal_tot: totaal stamtal in A2-, A3- en A4-plot.
- NameNI: Nederlandse naam van de boomsoort (naamgeving volgens VBI2).
- NameSc: wetenschappelijke naam van de boomsoort (naamgeving volgens VBI2).
- Exotic species: uitheemse boomsoort (1/0)
- Invasive species: invasieve boomsoort (1/0)

Variabelen meetproces en strata

- Zie §6.4.1, §6.4.3, §6.5.1 en §6.5.2.

6.6.11 Vegetatiesamenstelling

Deze tabel laat een meer gedetailleerde analyse toe van de vegetatieopnamegegevens.

Tbl11Vegetatie

IDVariabelen

- IDPlots: ID van het rasterpunt waar de plot gelegen is.
- Periode: volgnummer voor de opeenvolgende bosinventarisaties.
- Layer: boomlaag, struiklaag, kruidlaag of moslaag.
- IDSp: ID voor soort.
 - De soortenlijsten die gebruikt werden voor vegetatieopnames in VBI1 en VBI2 zijn verschillend. De oorspronkelijke ID uit VBI1 en VBI2 per soort komen dus niet overeen.
 - Er werd een gemeenschappelijk soortenlijst aangemaakt voor VBI1 en VBI2. Deze lijst bevat een unieke ID per soort. De lijst is terug te vinden in de tabel 'tblSpeciesListComb' in de databank 'VBIExterneData'. De tabel bevat ook de oorspronkelijke ID uit VBI1 en VBI2.

Analysevariabelen

- CoverClass: Braun-Blanquet klassen (Tabel 6-1).
- Cover: bedekking (%) overeenkomstig Tabel 6-1.
- Tree: de soort is al dan niet een boom, zoals aangegeven in de tabel 'tblSpeciesTreeLayerCharacteristics' in de databank 'VBIExterneData'.
- NameSc: wetenschappelijke naam (volgens VBI2)
- NameNI: Nederlandse naam (volgens VBI2)

7 Gegevensanalyse voor het beantwoorden van meetvragen

Op basis van de analysevariabelen kunnen de verschillende meetvragen beantwoord worden. Als voorbeeld werden enkele analysescripts in R uitgewerkt. De keuze viel op R omdat dit een open softwarepakket is dat iedereen kan gebruiken zonder licentiekosten. Het is tegelijk zeer veelzijdig inzake codeermogelijkheden en wordt zeer veel gebruikt in onderzoekscontext. Er werd gekozen voor verschillende types meetvariabelen (continu, binair, aantallen...) en meetgegevens (dendrometrische gegevens, vegetatieopnamegegevens...). Op die manier komen alle aspecten van de gegevensanalyse aan bod. Hierdoor moet het mogelijk zijn om vertrekkende van de voorbeeldscripts nieuwe scripts te maken voor andere meetvragen.

We geven eerst een overzicht van de verschillende analysescripts die werden uitgewerkt (§7.1). Vervolgens overlopen we de verschillende analysestappen in de scripts (§7.2).

7.1 Overzicht analysescripts

In Tabel 7-1 geven we een overzicht van de verschillende analysescripts, de meetvragen die ermee beantwoord worden en het type variabele dat geanalyseerd wordt.

Tabel 7-1 Overzicht van de verschillende analysescripts

Nr. Meetvraag	Meetvraag	Analysevariabele	Type variabele	Type gegevens	Naam script
3.a.	Wat is het aandeel aan loofhout, naaldhout, gemengd loofhout en gemengd naaldhout?	v2_StandType	Categorisch	Bestandskarakteristieken kwalitatief	AnalyseBestandstype.R
6.1.a.	Wat is de totale (m ³) en gemiddelde (m ³ /ha) houtvoorraad?	v11_TotalVolume	Continu	Bestandskarakteristieken kwantitatief	AnalyseHoutvoorraad.R
6.3.b.	Wat is de totale (m ³) en gemiddelde (m ³ /ha) houtvoorraad per bostypegroep?				
2.e.	Is het gemiddeld grondvlakaandeel aan inheemse boomsoorten in de totale populatie gelijk of groter dan 80%?	v5_FractionBasalArea-Autocht	Continu proportie	Bestandskarakteristieken kwantitatief	AnalyseGrondvlakaandeelInheems.R
4.2.b.	Wat is de kans op aanwezigheid en het grondvlakaandeel van inheemse / uitheemse / invasieve boomsoorten?	v8_PresenceAutocht v9_PresenceExotic v10_PresenceInvasive	Binair	Bestandskarakteristieken kwantitatief	AnalyseKansOpAanwezigheidInheems.R
4.1.b.	Wat is het gemiddeld aantal plantensoorten? (doelpopulatie en per bostypegroep)	v2_NbSp	Aantallen	Vegetatie	AnalyseAantalSoorten.R
4.1.a.	Wat is de oppervlakte bos (ha) geordend volgens het aantal aanwezige boomsoorten? (doelpopulatie en per bostypegroep)	v1_NbTreeSp	Aantallen		
6.1.b.	Wat is voor de homogene bestanden van de 5 belangrijkste boomsoorten de gemiddelde lopende jaarlijkse aanwas per ha? Ligt deze hoger dan 5	v1_LJA	Continu	Aanwas	AnalyseAanwas.R

Nr. Meetvraag	Meetvraag	Analysevariabele	Type variabele	Type gegevens	Naam script
	m ³ /ha/jaar?				
6.1.c.	Wat is de gemiddelde lopende jaarlijkse bestandsaanwas per bostypegroep? Ligt deze hoger dan 5 m ³ /ha/jaar?	v1_LJA	Continu		
6.1.d.	Wat is voor de 5 belangrijkste boomsoorten in homogene bestanden het gemiddelde afgestorven bestandsvolume (m ³ /ha/jaar)?	v2_DeadStanding	Continu		
6.1.e.	Wat is het gemiddelde afgestorven bestandsvolume (m ³ /ha/jaar) per bostypegroep?	v2_DeadStanding	Continu		
2.a.	Wat is van de 10 belangrijkste boomsoorten het procentuele aandeel in het totale stamtal van de populatie?	v1_StamatA3A4_ha	Continu	BoomsoortenSamenstelling	AnalyseMainSpecies.R
2.b.	Wat is van de 10 belangrijkste boomsoorten het procentuele aandeel in het totale grondvlak/volume van de populatie?	v2_BasalArea_ha; v3_Volume_ha	Continu		
2.g.	Wat is de kans op voorkomen en gemiddelde abundantie van NV invasieve boomsoorten, in het bijzonder Amerikaanse vogelkers (als NV voorkomt)?	v5_NVInvasivePres	Binair	Vegetatie	AnalyseVerjonging.R
4.2.c.	Wat is de abundantie aan verjonging van deze soorten (inheemse/uitheemse/invasieve boomsoorten) (cijfers op basis van plaatsen waar verjonging effectief voorkomt)?	v0_NV; v6_NVExoticAbund; v7_NVAutochtAbund; v8_NVInvasiveAbund	Continu	Vegetatie	
4.4.b.	Wat is het gemiddelde volume dood hout (m ³ /ha) (staand en liggend)?	v1_VolumeSnags; v3_VolumeLogs;	Continu	Dood hout	AnalyseDoodHout.R
4.4.c.	Is het gemiddelde aandeel (%) dood hout (staand en liggend) groter dan 4 % van het totale volume?	TotalVolume			

7.2 Verschillende analysestappen

We overlopen de verschillende stappen die aan bod komen in de analysescripts. We geven daarbij voorbeelden van enkele veelgebruikte R-codes, zonder in detail in te gaan op de functionaliteiten van deze codes. Voor meer details verwijzen we naar de BMK-lesserreeks.

7.2.1 Voorbereiding

In een eerste stap worden standaard de bestandsnamen van de verschillende datasets aangegeven (inclusief de map waar de dataset zich bevindt t.o.v. de working directory) en worden alle R-libraries die nodig kunnen zijn, geladen. Dit gebeurt door het script 'Omgevingsvariabelen.R' uit te voeren via volgend commando. Indien nodig moeten de

libraries eerst nog geïnstalleerd worden (via menu 'Tools' en functie 'Install packages...' in RStudio).

```
source("scripts/omgevingsvariabelen.R")
```

Daarnaast worden enkele functies die in verschillende analyseScripts gebruikt worden, ingelezen. Deze functies worden gebundeld in het R-script 'VBI_Functies.R'. Ze worden eveneens ingelezen via het source-commando.

7.2.2 Selectie analyseset

Een tweede stap is het selecteren van de analyseset waarop de analyse zal gebeuren. Daarbij wordt eerst de analysetabel ingelezen die de noodzakelijke analysevariabele(n) bevat (zie Bijlage 1) uit de acces databank 'VBI_AnalyseDatabank'. Als voorbeeld geven we de code voor het inlezen van 'tbl10Boomsoorten'. Hierbij gaan we ervan uit dat dit in het Ms Access formaat is dat vanaf 2007 gebruikt wordt (.accdb).

```
connectieAnalyseData <- odbcConnectAccess2007(dbAnalyseData)

tblData <- sqlFetch(connectieAnalyseData, "tbl10Boomsoorten", stringsAsFactors =
TRUE);print(str(tblData ))

odbcClose(connectieAnalyseData)
```

Vervolgens wordt een selectie gemaakt van variabelen die noodzakelijk zijn voor de analyse:

- Analysevariabelen
- Hulpvariabelen (dit zijn de variabelen die het meetproces definiëren en noodzakelijk zijn in de analyse, zie § 6.4.1 en §6.4.3)
- Strata

We geven hiervan een voorbeeld met volgende code.

```
hulpvariabelen <- c("IDPlots", "IDSegments","Periode", "IDGroup", "Plotweight",
"Segmentweight", "Reeks", "DateDendro","StartPeriode")

respons <- c("v11_TotalVolume")

strata <- c("StandType","OwnerType","ForestTypeGroup")

analyseSet <- tblData[c(hulpvariabelen, respons,strata)]
```

De analyseset bevat de gegevens op segmentniveau. Afhankelijk van de vraagstelling kan voor de vereenvoudiging de analyseset geaggregeerd worden op plot-niveau. Als we bijvoorbeeld het verschil in gemiddelde houtvoorraad willen analyseren op niveau Vlaanderen, volstaat een analyseset op plotniveau. Als we de gemiddelde houtvoorraad per bestandstype willen analyseren, is een analyseset op segmentniveau noodzakelijk. In dat geval wordt het bestandstype op segmentniveau bepaald. De aggregatie gebeurt door een gewogen gemiddelde of gewogen totaal te nemen per plot. Dit kan eenvoudig met de ddply-functie (Wickham, 2015a).

```
analyseSetPerPlot<-ddply(analyseSet,.(IDPlots ,Periode, IDGroup, Plotweight,
Paired_1_2, Periode, Reeks, DateDendro, OwnerType,
ForestTypeGroupCode),summarise,

v11_TotalVolume=sum(SegmentWeight*v11_TotalVolume,na.rm=TRUE)/sum(Segment
weight,na.rm=TRUE) )
```

7.2.3 Dataverkenning

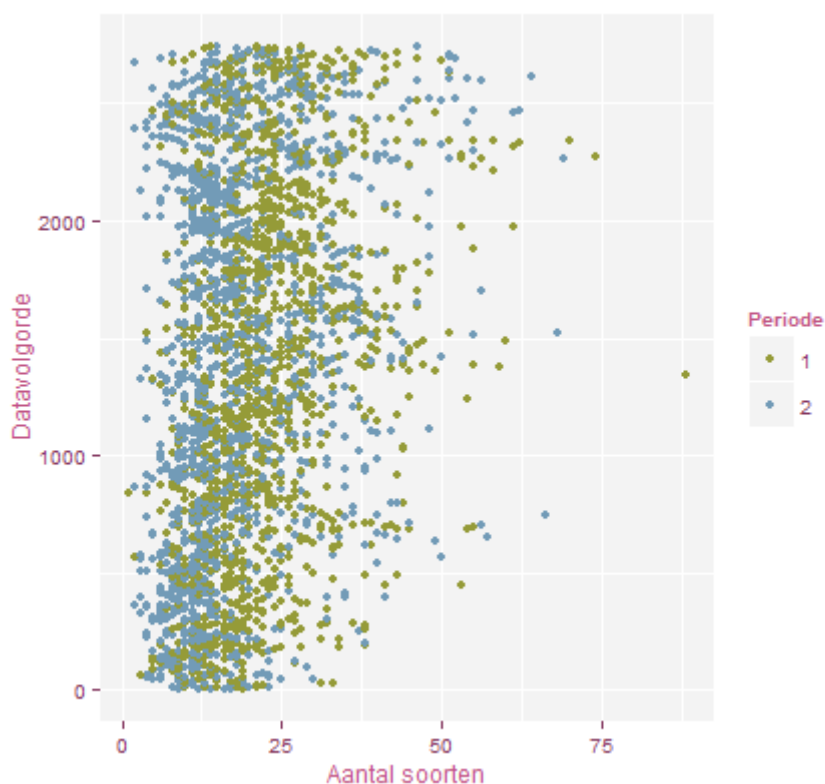
De dataverkenning heeft als doel om mogelijke outliers te identificeren en patronen te ontdekken in de data, zoals bijvoorbeeld verschillen voor een bepaalde variabele tussen de twee inventarisatierondes of tussen verschillende strata. Ook kunnen we verkennen wat de

impact is van het terugplaatsen van plots op de bosrand en of er verschillende patronen zijn voor gepaarde en niet-gepaarde plots.

We geven een overzicht van de grafieken die bruikbaar kunnen zijn in de dataverkenning. Deze grafieken worden aangemaakt via de library 'ggplot2' (Wickham, 2015b). ggplot2 bouwt een figuur sequentieel op. Eerst definieer je de dataset en de variabelen die de assen uitmaken (naast de x- en y-as, kun je ook een kleur-as of punttype-as definiëren). Daarna bouw je de plot laag per laag op.

Dotplots zijn bedoeld om outliers op te sporen voor een bepaalde variabele. In de x-as worden de waarden van de variabele geplott en in de y-as de volgorde van de gegevens. In een dotplot ligt de essentie van de informatie dus in de spreiding ten opzichte van de x-as. Figuur 4 wordt gegenereerd via onderstaande code.

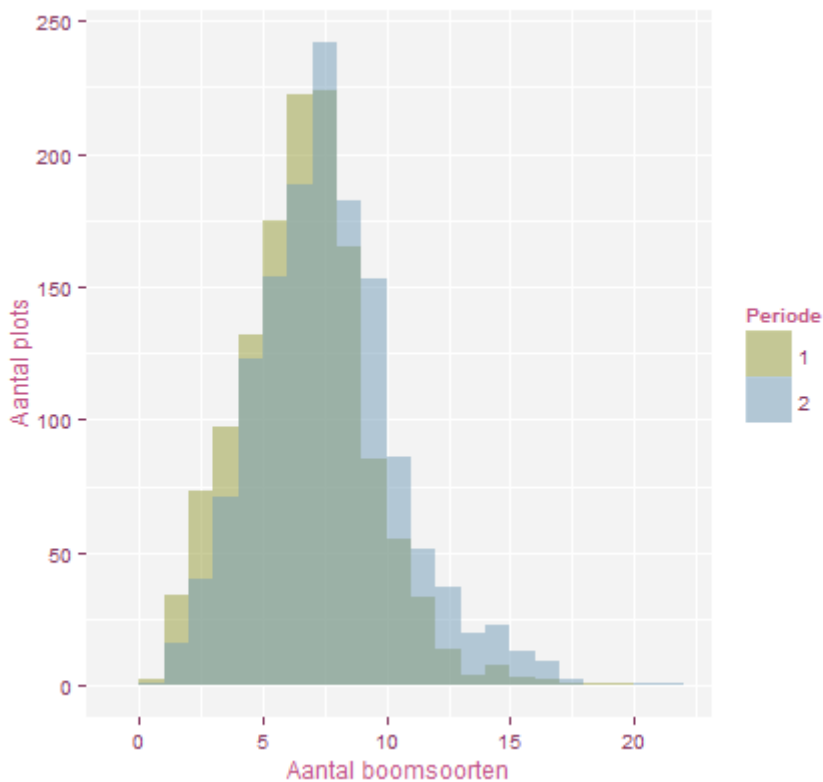
```
ggplot(analyseSetPerPlot, aes(x=v2_NbSp,
y=1:length(analyseSetPerPlot$IDPlots,colour=fPeriode))+geom_point()+labs(x="Aantal soorten", y="Datavolgorde",colour="Periode")
```



Figuur 4 Dotplot van het aantal soorten in de vegetatieopname met onderscheid tussen VBI1 (periode=1) en VBI2 (periode=2).

Een **histogram** wordt gebruikt om de verdeling van de data te visualiseren. Dit is bruikbaar voor de selectie van een gepast datamodel. Het laat ook toe om verschillen in distributie te verkennen tussen inventarisatierondes, strata of type plots (gepaard/niet-gepaard, randplot/plot volledig in bos). Figuur 5 wordt aangemaakt via onderstaande code. In de figuur zien we dat de distributie van het aantal boomsoorten tussen VBI1 en VBI2 lichtjes verschuift richting hogere waarden.

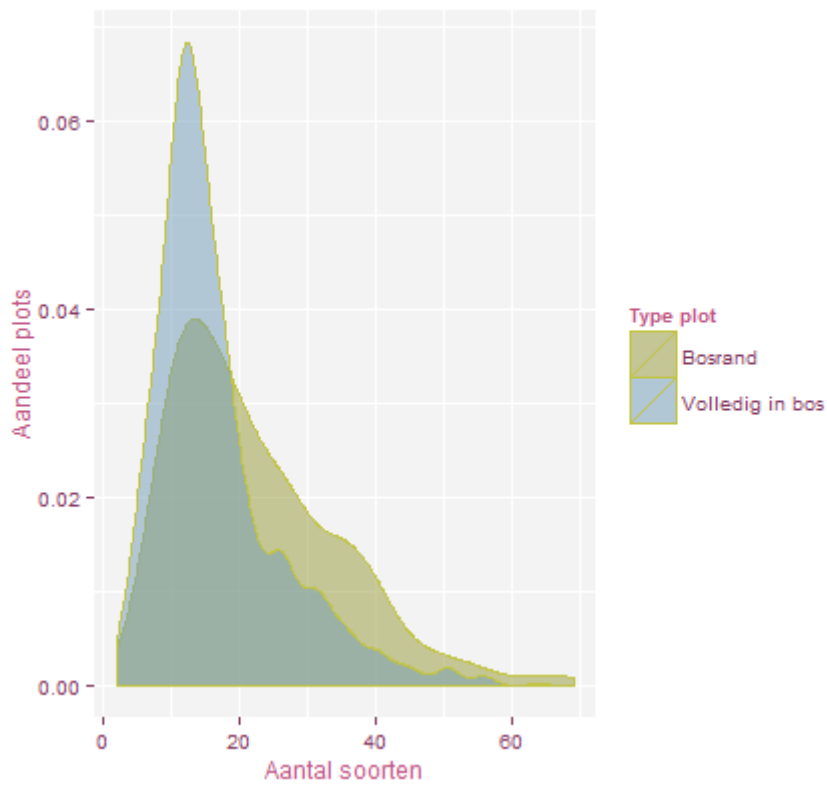
```
ggplot(analyseSetPerPlot, aes(x=v1_NbTreeSp, fill = fPeriode)) +
geom_histogram(alpha=.5, position="identity",binwidth=1) + labs(x="Aantal boomsoorten", y="Aantal plots",fill="Periode")
```



Figuur 5 Histogram van het aantal boomsoorten met onderscheid tussen VBI1 (periode=1) en VBI2 (periode=2).

Een **densityplot** is gelijkaardig aan een histogram, maar geeft het relatieve aandeel observaties voor elke waarde(klasse) van de analysevariabele. De histogramfunctie geeft steeds het absoluut aantal observaties per waardeklasse. Een densityplot is bijvoorbeeld bruikbaar wanneer je twee groepen wilt vergelijken met een sterk verschillend aantal observaties. Onderstaande code genereert Figuur 6 waarin de distributie van het aantal soorten vergelijkt wordt tussen plots op een bosrand en plots die volledig in het bos gelegen zijn. We zien dat een groter aandeel van de plots op een bosrand een groter aantal soorten hebben.

```
ggplot(analyseSetPerPlot[analyseSetPerPlot$Periode==2,],
  aes(x=v2_NbSp, fill=Rand)) + geom_density(alpha=.5, position="identity") +
  labs(x="Aantal soorten", y="Aandeel plots")
```

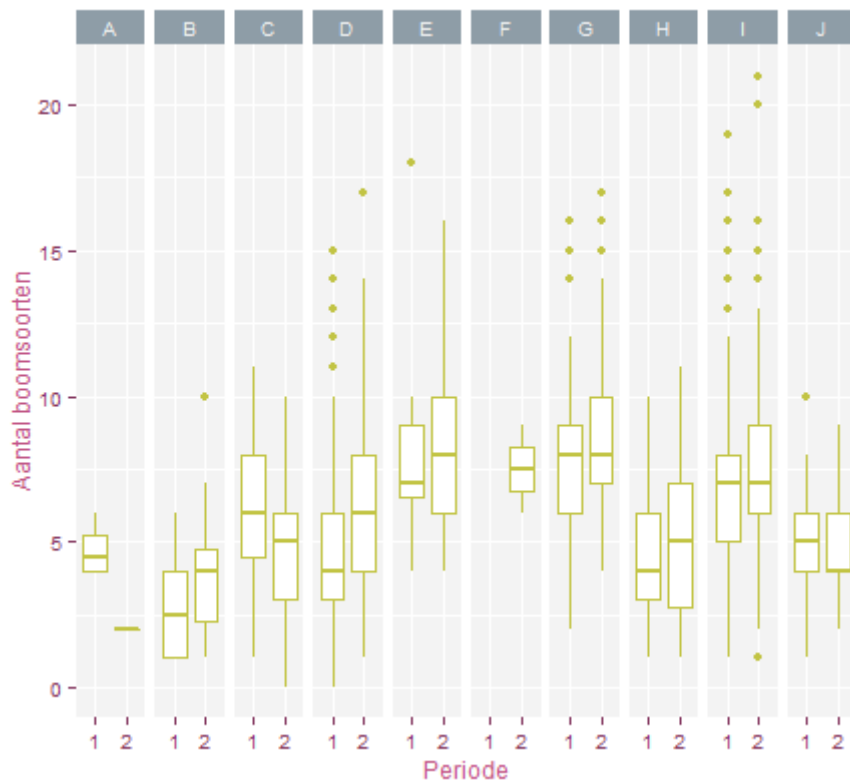


Figuur 6 Densityplot van het aantal soorten in de vegetatieopname voor VBI2 met onderscheid tussen plots die op een bosrand gelegen zijn en plots die volledig in bos gelegen zijn.

Een **boxplot** is een vereenvoudigde weergave van de distributie van een variabele en laat toe om verschillen te identificeren tussen periodes of strata.

Onderstaande code genereert Figuur 7.

```
ggplot(analyseSetPerPlot[!is.na(analyseSetPerPlot$ForestTypeGroup),], aes(y =
v1_NbTreeSp, x = fPeriode)) + geom_boxplot() + facet_grid(.~ForestTypeGroupCode)
+ labs(x="Periode",y="Aantal boomsoorten")
```



Figuur 7 Boxplots van het aantal boomsoorten in VBI1 en VBI2 met onderscheid tussen de verschillende bostypegroepen.

Een boxplot bevat volgende onderdelen:

- De rechthoek wordt gevormd door het 1^{ste} en 3^{de} kwartiel.
- De horizontale lijn in de rechthoek is de waarde van de mediaan.
- De afzonderlijke punten zijn outliers. Deze liggen op een afstand t.o.v. de mediaan die groter is dan 1,5 keer de kwartielaafstand (hoogte van de rechthoek).
- De verticale lijn verbindt de rechthoek met de hoogste/laagste waarde die niet als outlier beschouwd wordt.

7.2.4 Design-based analyse

Met design-based analyse worden parameterschattingen volgens de klassieke (steekproef)statistiek bedoeld. In de context van de bosinventarisatie gaat het meestal om de schatting van het gewogen steekproefgemiddelde en het 95% betrouwbaarheidsinterval rond dit gemiddelde. Een gewogen gemiddelde is noodzakelijk voor de analyse van plots die gedeeltelijk uit bos bestaan en/of plots die uit verschillende types bos bestaan. Voor een analyseset op segmentniveau wordt het gewicht gelijk gesteld aan het product van de PlotWeight (fractie bos binnen een plot) en de SegmentWeight (fractie bos binnen het plot ingenomen door een segment). Voor een analyseset op plotniveau wordt de PlotWeight als gewicht gebruikt.

Het gewogen steekproefgemiddelde wordt berekend via onderstaande formule.

$$\bar{x}_{Gewogen} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i}$$

met x = analysevariabele, w = gewicht en N = de steekproefgrootte.

De gewogen steekproefvariantie wordt via onderstaande formule berekend.

$$s^2 = \frac{\sum_{i=1}^N w_i (x_i - \mu^*)}{V_1 - (V_2/V_1)}$$

$$V_1 = \sum_{i=1}^N w_i$$

$$V_2 = \sum_{i=1}^N w_i^2$$

Vervolgens kan de standaard error berekend worden:

$$SE = \frac{s}{\sqrt{N}}$$

Het 95% betrouwbaarheidsinterval wordt dan begrensd door $\bar{x}_{Gewogen} - 1,96 * SE$ en $\bar{x}_{Gewogen} + 1,96 * SE$.

Er werd een functie in R aangemaakt ('My.WgtParEstimation') die het gewogen steekproefgemiddelde en bijhorend betrouwbaarheidsinterval berekent voor een analysevariabele, waarbij optioneel een gemiddelde per stratum kan berekend worden. De functie laat ook toe om een subset uit de volledige dataset te selecteren waarop de berekening wordt uitgevoerd. De selectie gebeurt op basis van de Periode, de Reeks en/of het jaartal.

De functie heeft volgende inputparameters:

- Data: de volledige dataset.
- VariableName: naam van de te analyseren variabele of een vector van verschillende variabelenamen indien men meerdere analysevariabelen wenst te verwerken.
- Periode: periode van de subset die men wenst te analyseren.
- MinReeks: minimumreeks van de subset die men wenst te analyseren.
- MaxReeks: maximumreeks van de subset die men wenst te analyseren.
- MinJaar: minimumjaar van de subset die men wenst te analyseren.
- MaxJaar: maximumjaar van de subset die men wenst te analyseren.
- UseStrata: TRUE als men een parameterschatting per stratum wenst en FALSE als men dit niet wenst. Indien men meerdere analysevariabele wenst te analyseren, moet een vector van TRUE/FALSE waarden ingegeven worden.
- Strata: naam van de variabele die de strata definieert. Indien men meerdere analysevariabele wenst te analyseren, moet een vector van stratanamen ingegeven worden.

Hieronder geven we een voorbeeld van het gebruik van de functie. Eerst wordt het gewicht gedefinieerd en het jaartal. Vervolgens worden de namen van de variabelen gedefinieerd. Vervolgens wordt de functie `My.WgtParEstimation` viermaal uitgevoerd en worden de resultaten samengevoegd via `rbind`. Eerst gebeurt er een schatting voor VBI2, vervolgens voor VBI1 (enkel voor de plots tot en met reeks 5), dan per eigenaarscategorie voor VBI2 en ten slotte per eigenaarscategorie voor VBI1.

```
analyseSet$weight<-analyseSet$SegmentWeight*analyseSet$Plotweight
analyseSet$Year<-analyseSet$dateDendro
respons <- c("v8_PresenceAutocht","v9_PresenceExotic","v10_PresenceInvasive")
results<- rbind(My.WgtParEstimation(analyseSet,respons,Periode = 2, MaxReeks = 5),
               My.WgtParEstimation(analyseSet,respons,Periode = 1, MaxReeks = 5),
               My.WgtParEstimation(analyseSet,respons,Periode = 2, MaxReeks = 5,
               UseStrata = c(TRUE,TRUE,TRUE),Strata = c("OwnerType","OwnerType","OwnerType")),
               My.WgtParEstimation(analyseSet,respons,Periode = 1, MaxReeks =
               5,UseStrata = c(TRUE,TRUE,TRUE),Strata = c("OwnerType","OwnerType","OwnerType")))
```

De output van bovenstaand script wordt weergegeven in Tabel 7-2.

Tabel 7-2 Schatting van de kans op aanwezigheid ('wgt.mean'), de gewogen variantie ('wgt.var') en de ondergrens en bovengrens van het 95% betrouwbaarheidsinterval (respectievelijk 'lower limit confidence interval' = 'lci' en 'upper limit confidence interval' = 'ulci') van inheemse, uitheemse en invasieve boomsoorten voor VBI1 en VBI2 en per eigenaarscategorie.

variabele	strata	stratumNaam	periode	minYear	maxYear	minReeks	maxReeks	nbObservaties	wgt.mean	wgt.var	lci	ulci
v8_PresenceAutocht			2	2009	2014	1	5	1517	0.955	0.043	0.944	0.966
v9_PresenceExotic			2	2009	2014	1	5	1517	0.584	0.243	0.557	0.611
v10_PresenceInvasive			2	2009	2014	1	5	1517	0.269	0.197	0.245	0.294
v8_PresenceAutocht			1	1997	1999	1	5	1328	0.903	0.088	0.887	0.919
v9_PresenceExotic			1	1997	1999	1	5	1328	0.516	0.250	0.489	0.543
v10_PresenceInvasive			1	1997	1999	1	5	1328	0.214	0.168	0.192	0.236
v8_PresenceAutocht	OwnerType	Private	2	2009	2014	1	5	894	0.958	0.040	0.943	0.972
v8_PresenceAutocht	OwnerType	Public	2	2009	2014	1	5	610	0.950	0.047	0.932	0.969
v9_PresenceExotic	OwnerType	Private	2	2009	2014	1	5	894	0.577	0.244	0.542	0.613
v9_PresenceExotic	OwnerType	Public	2	2009	2014	1	5	610	0.587	0.243	0.545	0.629
v10_PresenceInvasive	OwnerType	Private	2	2009	2014	1	5	894	0.300	0.210	0.267	0.333
v10_PresenceInvasive	OwnerType	Public	2	2009	2014	1	5	610	0.217	0.170	0.182	0.252
v8_PresenceAutocht	OwnerType	Private	1	1997	1999	1	5	828	0.917	0.076	0.898	0.936
v8_PresenceAutocht	OwnerType	Public	1	1997	1999	1	5	500	0.880	0.106	0.851	0.909
v9_PresenceExotic	OwnerType	Private	1	1997	1999	1	5	828	0.492	0.250	0.457	0.526
v9_PresenceExotic	OwnerType	Public	1	1997	1999	1	5	500	0.556	0.247	0.512	0.600
v10_PresenceInvasive	OwnerType	Private	1	1997	1999	1	5	828	0.219	0.171	0.190	0.247
v10_PresenceInvasive	OwnerType	Public	1	1997	1999	1	5	500	0.206	0.164	0.171	0.241

Een design-based analyse is geschikt voor het bepalen van een toestand voor een bepaalde periode, maar minder geschikt voor het bepalen van een verschil tussen periodes. Dat komt omdat een design-based analyse moeilijk kan omgaan met de (gedeeltelijk) gepaarde plots. Daarom raden we mixed models (§7.2.5) aan voor het analyseren van verschillen of trends.

7.2.5 Model-based analyse met mixed models

7.2.5.1 Achtergrond

Een model-based analyse is beter geschikt dan een design-based analyse om verschillen of trends te schatten. Dit is zeker het geval indien we te maken hebben met deels gepaarde metingen. Klassieke regressiemodellen veronderstellen dat alle observaties onafhankelijk van elkaar zijn. Dit is niet het geval bij gepaarde plots, vermits de observaties in opeenvolgende periodes gecorreleerd zijn. Mixed effect regressiemodellen kunnen wel met dit soort gecorreleerde data overweg. Net zoals een klassiek regressiemodel, modelleert een mixed effect regressiemodel (of mixed model, MM) het verband tussen een responsvariabele enerzijds en een of meerdere verklarende variabelen anderzijds. Anders dan bij een klassiek regressiemodel worden bij een MM de responsvariabelen onderverdeeld in 'fixed effects' die het globale patroon verklaren en 'random effects' die lokale verschillen t.o.v. globale patronen verklaren. Voor meer achtergrond over het gebruik van MM verwijzen we naar Jansen & Quataert (2013).

De MM techniek kan een grote verscheidenheid aan responsvariabelen modelleren. Wanneer de respons (binaire of ordinale variabelen, proporties, tellingen) niet normaal verdeeld is, kunnen we vrij eenvoudig bovenstaand model veralgemenen (Zuur *et al.*, 2009). Bates *et al.* (2015) werkten tools uit in R die MM fitten voor continue variabelen, tellingen en binaire variabelen.

7.2.5.2 Voorbeelden

7.2.5.2.1 Mixed model met continue variabele

Als voorbeeld geven we de R-code voor een mixed model dat het grondvlakaandeel van inheemse boomsoorten modelleert in functie van de periode (VBI1 versus VBI2) met de plot als random effect (gepaarde plots hebben eenzelfde waarde voor IDGroup). Dit kan via de lme-functie of via de lmer-functie. Het voornaamste verschil tussen beide functies is de wijze waarop de gewichten worden ingegeven. Daarnaast toont de lmer-functie geen p-waarden voor de geschatte parameters. De uitkomst van de parameterschatting is identiek voor beide functies. De lme functie zit vevat in het nlme package (Pinheiro *et al.*, 2015) en (g)lmer zit vevat in het lme4 package (Bates *et al.* 2015).

```
Model.lme <- lme( v5_FractionBasalAreaAutocht ~ fPeriode, random = ~1|IDGroup, data = analyseSetPerPlot, weights=varFixed(~1/Plotweight))

Model.lmer <- lmer( v5_FractionBasalAreaAutocht ~ fPeriode + (1|IDGroup), data= analyseSetPerPlot, weights=Plotweight )
```

In bovenstaande modellen is de periode een factorvariabele. Dit betekent dat het een categorische variabele is met vaste categorieën (hier '1' en '2') en dat deze categorieën dus niet als numerieke waarde geïnterpreteerd worden. Dit is van belang bij de interpretatie van de uitkomst van de modellen. Om te verduidelijken dat het om een factorvariabele gaat kiezen we als naam 'fPeriode' voor de variabele.

Via de functie `summary(Model.lme)` krijgen we volgende resultaten:

```
Linear mixed-effects model fit by REML
Data: analyseSetPerPlot
      AIC      BIC    logLik
1598.774 1623.855 -795.3871

Random effects:
Formula: ~1 | IDGroup
      (Intercept)  Residual
StdDev:  0.3516443  0.06643779

Variance function:
Structure: fixed weights
Formula: ~1/Plotweight

Fixed effects: v5_FractionBasalAreaAutocht ~ fPeriode
      value  Std.Error  DF  t-value p-value
```

```
(Intercept) 0.7719191 0.006305438 3353 122.42117 0.000
fPeriode2 -0.0061406 0.003858014 553 -1.59164 0.112
Correlation:
  (Intr)
fPeriode2 -0.203

Standardized within-Group Residuals:
      Min       Q1       Med       Q3      Max
-6.5134495 -0.1101796  0.1072214  0.1183217  6.4072788

Number of Observations: 3908
Number of Groups: 3354
```

De belangrijkste resultaten zijn te vinden onder de rubriek 'fixed effects'. In de kolommen vinden we achtereenvolgend:

- de waarde van de geschatte parameter
- de standaardfout
- het aantal vrijheidsgraden
- de t-waarde (waarde/standaardfout die verdeeld is volgens de t-distributie)
- de p-waarde

In de eerste rij met naam '(Intercept)' vinden we de schatting voor het eerste niveau van de variabele 'fPeriode'. Dit komt dus neer op de 1^{ste} periode (hier gelijk aan VBI1). Het grondvlakaandeel van inheemse boomsoorten wordt dus geschat op 77,2% voor VBI1. De volgende rij geeft het verschil weer van het tweede t.o.v. het eerste factorniveau (hier VBI2 t.o.v. VBI1). Dit verschil wordt geschat op 0,6 %. We zien echter dat de p-waarde voor het verschil gelijk is aan 0,112 en dus groter is dan 0,05, waaruit we kunnen aannemen dat het grondvlakaandeel in beide periodes niet significant verschillend is bij een significantieniveau van 5%.

Via de standaardfout kunnen we het betrouwbaarheidsinterval berekenen op dezelfde wijze als in §7.2.4. Op basis van de modeloutput kunnen we hier echter ook eenvoudig het betrouwbaarheidsinterval op het verschil berekenen. In dit voorbeeld is dit echter weinig zinvol vermits het verschil niet significant is.

7.2.5.2.2 Mixed-model met binaire variabele

Via de glmer-functie in R kunnen ook mixed-models voor andere type variabelen gefit worden. Als voorbeeld geven we hier voor een MM de R-code van de kans op aanwezigheid van uitheemse boomsoorten in functie van de periode (VBI1 en VBI2) met proefvlak als random effect. Het type verdeling specificeren we via het argument 'family'. Aanwezigheid van uitheemse boomsoorten is een binaire variabele. Deze volgt een binomiale verdeling (met 1 herhaling).

```
Model.PresExot <-glmer(data=analyseSetPerPlot,v9_PresenceExotic ~
  fPeriode + (1|IDGroup), family="binomial",weights = PlotWeight)
```

Het model geeft volgende uitkomst:

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
['glmerMod']
Family: binomial ( logit )
Formula: v9_PresenceExotic ~ fPeriode + (1 | IDGroup)
Data: analyseSetPerPlot
weights: weight

      AIC      BIC    logLik deviance df.resid
 5331.3  5350.2 -2662.7  5325.3    3998

Scaled residuals:
  Min       1Q   Median       3Q      Max
-1.1921 -0.8533  0.6221  0.8306  1.0166

Random effects:
```

```

Groups Name Variance Std.Dev.
IDGroup (Intercept) 0.8298 0.9109
Number of obs: 4001, groups: IDGroup, 3443

Fixed effects:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.03233 0.04676 0.691 0.489
fPeriode2 0.38449 0.08035 4.785 1.71e-06 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr)
fPeriode2 -0.544

```

Het model geeft aan dat er een significant verschil is in de kans op aanwezigheid van uitheemse boomsoorten tussen beide periodes (VBI1 en VBI2), vermits de p-waarde voor 'fPeriode2' nul benadert. De parameterschattingen van een mixed binomiaal model staan in de logit-schaal. De kans op aanwezigheid in lineaire schaal wordt bekomen via de inverse-logit functie: $\exp(x)/(1+\exp(x))$. Dit kan ook eenvoudig via de R-functie `plogis()`.

Via onderstaande R-code bekomen we de schatting per periode en de schatting van het verschil. In de variabele 'summary' slaan we eerst de modeloutput op. Vervolgens kunnen we hieruit de verschillene parameterschattingen selecteren.

```

samenvattingExot<-summary(Model.PresExot)

Mean.Periode1<-plogis(summary$coefficients["(Intercept)","Estimate"])

Mean.Periode2<-plogis(summary$coefficients["(Intercept)","Estimate"]+summary
$coefficients["fPeriode2","Estimate"])

Verschil<- Mean.Periode2 - Mean.Periode1

```

De kans op aanwezigheid van uitheemse boomsoorten bedraagt 0,51 in VBI1 en 0,60 in VBI2, een verschil van 0,09. De betrouwbaarheidsintervallen op de schattingen kunnen als volgt berekend worden:

```

Mean.Periode1_llci<-plogis(summary$coefficients["(Intercept)","Estimate"]-
1.96*summary$coefficients["(Intercept)","Std. Error"])

Mean.Periode1_rlci<-
plogis(summary$coefficients["(Intercept)","Estimate"]+1.96*summary$coefficients[
"(Intercept)","Std. Error"])

Mean.Periode2_llci<-
plogis(summary$coefficients["(Intercept)","Estimate"]+summary$coefficients["fPer
iode2","Estimate"]-1.96*summary$coefficients["fPeriode2","Std. Error"])

Mean.Periode2_rlci<-
plogis(summary$coefficients["(Intercept)","Estimate"]+summary$coefficients["fPer
iode2","Estimate"]+1.96*summary$coefficients["fPeriode2","Std. Error"])

Verschil_llci<-plogis(summary$coefficients["(Intercept)", "Estimate"] +
summary$coefficients["fPeriode2", "Estimate"] - 1.96 *
summary$coefficients["fPeriode2", "Std. Error"]) -
plogis(summary$coefficients["(Intercept)", "Estimate"])

Verschil_rlci<- plogis(summary$coefficients["(Intercept)", "Estimate"] +
summary$coefficients["fPeriode2", "Estimate"] + 1.96 *
summary$coefficients["fPeriode2", "Std. Error"]) -
plogis(summary$coefficients["(Intercept)", "Estimate"])

```

De resultaten zijn samengevat in Tabel 7-3.

Tabel 7-3 Samenvattende analyseresultaten van de kans op aanwezigheid van uitheemse boomsoorten

Periode	Kans op aanwezigheid inheemse boomsoort	Onderlimiet 95% betrouwbaarheids-interval	Bovenlimiet 95% betrouwbaarheids-interval
VBI1	0,51	0,49	0,53
VBI2	0,60	0,56	0,64
Vershil	0,09	0,06	0,13

7.2.5.2.3 Mixed-model met aantallen

Aantallen kunnen gemodelleerd worden via een Poisson mixed model. Als voorbeeld geven we de code voor een model van het aantal boomsoorten in functie van de periode, met proefvlak als random effect.

```
M1bis<-glmer(data=analyseSetPerPlot,v1_NbTreeSp ~ fPeriode + (1|IDGroup),
family="poisson")
summary(M1bis)
```

Dit geeft volgende output voor de fixed effects:

```
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.80599    0.01227  147.17  <2e-16 ***
fPeriode2    0.14025    0.01606   8.73  <2e-16 ***
```

De resultaten staan deze keer in de log-schaal en kunnen naar een lineaire schaal getransformeerd worden via de exp() – functie. De schattingen van het aantal boomsoorten per periode, het verschil in aantal boomsoorten en de betrouwbaarheidsintervallen gebeuren op een analoge manier als in § 7.2.5.2.2. Een verschil daarbij is dat exp() moet gebruikt worden i.p.v. plogis().

7.2.5.2.4 Verschil tussen strata met mixed models

Via het onderstaand model kunnen we nagaan of het grondvlakaandeel van inheemse boomsoorten verschilt tussen privé- en openbare bossen (via de term 'OwnerType') en of het verschil in grondvlakaandeel tussen VBI1 en VBI2 verschilt tussen privé- en openbare bossen (via de interactieterm 'fPeriode*OwnerType').

```
Model.lme.OwnerType <- lme( v5_FractionBasalAreaAutocht ~fPeriode + OwnerType +
fPeriode*OwnerType, random = ~1|IDGroup,, data =
analyseSetPerPlot,weights=varFixed(~1/Plotweight) )
summary(Model.lme.OwnerType)
```

Volgende output wordt bekomen:

```
Fixed effects: v5_FractionBasalAreaAutocht ~ fPeriode + OwnerType + fPeriode *
OwnerType
              Value      Std. Error    DF  t-value  p-value
(Intercept)  0.7895642  0.008732436  2118  90.41741  0.0000
fPeriode2    -0.0137305  0.005290375   545  -2.59537  0.0097
OwnerTypePublic -0.0284540  0.010663296   545  -2.66840  0.0078
fPeriode2:OwnerTypePublic 0.0170679  0.008440462   545   2.02215  0.0436
```

Hieruit kunnen we het grondvlakaandeel schatten per periode en per eigenaarscategorie op volgende wijze:

- Periode 1 en privé-bos: grondvlakaandeel = (Intercept) = 0,789
- Periode 2 en privé-bos: grondvlakaandeel = (Intercept) + fPeriode = 0,789 – 0,014 = 0,775
- Periode 1 en openbaar bos: grondvlakaandeel = (Intercept) + OwnerTypePublic = 0,789 – 0,028 = 0,761
- Periode 2 en openbaar bos: grondvlakaandeel = (Intercept) + fPeriode2 + OwnerTypePublic + fPeriode2:OwnerTypePublic = 0,789 – 0,014 – 0,028 + 0,017 = 0,764

Wat betreft p-waarden mogen we hier enkel naar de p-waarde van de interactieterm kijken. De interactieterm is hier licht significant want net onder 0,05. Er is dus aanwijzing dat het verschil tussen beide periodes meer positief is voor openbare bossen (-0,014 voor privé-bossen en -0,014 + 0,017 = + 0,003). De verschillen zijn echter zeer klein en de conclusies zijn dan ook onder voorbehoud. Bovenstaand voorbeeld is voornamelijk ter illustratie bedoeld van hoe om te gaan met strata.

7.2.6 Volledig uitgewerkt voorbeeld: analyse van de houtvoorraad

In Bijlage 2 worden de verschillende stappen van de analyse in detail doorlopen voor de meetvraag m.b.t. de houtvoorraad.

In Bijlage 2 is daarnaast ook een voorbeeld opgenomen van een geautomatiseerde rapportage via de R Markdown tool. Dit laat toe om tekst te combineren met resultaten uit R (figuren, berekeningen). Het voordeel hiervan is dat, indien de brongegevens veranderen, het script opnieuw kan uitgevoerd worden. De R-resultaten worden daardoor automatisch aangepast. In de praktijk zal de rapportage van een meetvraag waarschijnlijk niet zo uitgebreid gebeuren en zal i.f.v. de meetvraag een selectie gemaakt worden van de belangrijkste resultaten.

De bestandsnaam van het script dat de tekst in Bijlage 2 genereert is 'AnalyseHoutvoorraad.Rmd'.

7.2.7 Aandachtspunten bij de analyse

- **Definitie periodes**
In bovenstaande voorbeeldanalyses wordt periode 1 steeds gedefinieerd als VBI1 en periode 2 als VBI2. Periodes kunnen door de gebruiker ook anders gedefinieerd worden op basis van de reeks of opnamedatum. Wanneer er meer gegevens beschikbaar komen in de volgende jaren, kan er bijvoorbeeld een periode gevormd worden met de gegevens uit de laatste 10 jaar en vergeleken worden met de periode van 10 jaar daarvoor, ongeacht de inventarisatieronde waarin de gegevens werden ingezameld. Eventueel kunnen ook twee kleinere periodes van bijvoorbeeld 5 jaar vergeleken worden met elkaar, maar in dat geval ga je geen gepaarde plots hebben en uiteraard ook minder gegevens, waardoor het moeilijker wordt om veranderingen te detecteren.
- **Periode als factor of als continue variabele**
In de voorbeeldanalyse wordt de periode steeds als factor gebruikt en worden er twee periodes vergeleken. Naarmate meer gegevens beschikbaar komen, kunnen er

meerdere periodes onderscheiden worden waarop de analyse kan gebeuren. Wanneer bij meerdere periodes de periode als factor beschouwd wordt, dan zullen de verschillende periodes onderling vergeleken worden. Wanneer we een globale (minimale) trend willen schatten, moeten we de periode als continue variabele ingeven in het model. We hebben hiervoor wel een voldoende aantal periodes nodig om tot een goede schatting van de trend te komen (drie periodes is onvoldoende). Eventueel kunnen gegevens in kleinere periodes gegroepeerd worden. Dit heeft dan wel als keerzijde dat er minder waarnemingen per periode zijn en dat bijgevolg het model minder makkelijk veranderingen zal kunnen detecteren. Momenteel zijn er nog onvoldoende gegevens om dergelijke zaken uit te testen.

- **Distributie van data niet volledig conform modelassumpties**

Het zal regelmatig voorkomen dat de distributie van continue data sterk afwijkt van de normale verdeling. Kunnen we dan vertrouwen op de uitkomst van lineaire mixed models? In principe zijn parameterschattingen en p-waarden van een model niet correct als er afgeweken worden van de modelveronderstellingen. In de praktijk hebben we in het geval van de VBI te maken met een groot aantal observaties, waardoor we ervan uit kunnen gaan dat de MM redelijk robuust zijn en dat de vertekeningen op de parameterschattingen en p-waarde in de meeste gevallen minimaal zijn. We stellen daarom voor om op volgende wijze om te gaan met variabelen waarvan de distributie sterk afwijkt van de normale distributie:

- Indien de p-waarde voor het verschil tussen periodes duidelijk lager of hoger is dan 0,05, dan kan er besloten worden dat met voldoende zekerheid geconcludeerd kan worden dat de variabele respectievelijk significant of niet significant verschilt tussen de periodes.
- Indien de p-waarde rond de 0,05 ($\pm 0,02$) ligt dan is er onvoldoende zekerheid om te concluderen of er al dan niet een verschil is tussen de periodes.
- In principe is het ook mogelijk om gebruik te maken van complexere modellen die wel kunnen omgaan met afwijkende distributies van continue variabelen. In de praktijk ondervinden we dat in het geval van een groot aantal observaties (zoals bij de VBI) de resultaten van dergelijke complexere modellen meestal slechts in beperkte mate verschillen van de eenvoudigere modellen.

- **Van gemiddeldes naar totalen voor Vlaanderen**

Veel meetvragen hebben betrekking op een gemiddelde per hectare (bv. houtvolume/ha) of een proportie (bv. % punten dat uitheemse boomsoorten bevat). Daarnaast zijn er ook heel wat meetvragen die betrekking hebben op totalen (bv. totaal volume hout in Vlaanderen of oppervlakte bos met uitheems boomsoorten in Vlaanderen). Om tot het totaal te komen, volstaat het om het gemiddelde of de proportie te vermenigvuldigen met de oppervlakte bos in Vlaanderen. Het betrouwbaarheidsinterval op het totaal wordt dan als volgt berekend:

- Indien de totale oppervlakte bos als een constante wordt beschouwd (c_{bos}), moet de standaardfout op het totaal (SE_{T_x}) vermenigvuldigd worden met de wortel van de oppervlakte.

$$\hat{T}_x = c_{bos}\bar{x}$$

$$\Rightarrow s_{T_x}^2 = c_{bos}^2 s_x^2$$

$$\Rightarrow SE_{T_x} = \sqrt{c_{bos}} SE_x$$

- Indien de totale oppervlakte bos als een variabele beschouwd wordt (\hat{T}_{bos}), dan is de standaardfout van het totaal van variabele x af te leiden uit de kwadratische som van de relatieve fout van de totale oppervlakte bos (\hat{T}_{bos}) en de relatieve fout van de variabele x .

$$\hat{T}_x = \hat{T}_{bos}\bar{x}$$

$$\Rightarrow \left(\frac{s_{T_x}}{\hat{T}_x}\right)^2 = \left(\frac{s_{T_{bos}}}{\hat{T}_{bos}}\right)^2 + \left(\frac{s_x}{\bar{x}}\right)^2$$

$$SE_{T_x} = \frac{s_{T_x}}{\sqrt{N}}$$

8 Opslag van de analyseresultaten

We raden aan om resultaten van analyses op te slaan in een analyseresultatendatabank. Dit is in de eerste plaats bedoeld voor analyses die op punt staan en periodiek herhaald worden.

Per analyse wordt er een record (rij) toegevoegd, die naast de eigenlijke resultaten ook informatie bevat die het moet mogelijk maken om de analyse te reproduceren. We voorzien twee datastructuren: een voor de schatting van de toestand in een bepaalde periode en een voor het verschil tussen twee periodes.

Volgende kenmerken worden opgeslagen voor analyses m.b.t. de schatting van de toestand voor een bepaalde periode.

- Naam analysevariabele
- Kenmerken analyse
 - Design-based/ model-based
 - Datum analyse
 - Optioneel beschrijving van het model
- Kenmerken dataset
 - Minimumjaar in dataset
 - Maximumjaar in dataset
 - Minimumreeks in dataset
 - Maximumreeks in dataset
 - Aantal observaties
- Stratum
 - Ja/nee
 - Stratumnaam
- Eigenlijke resultaten
 - (Gewogen) gemiddelde
 - (Gewogen) variabiliteit
 - Benedengrens 95% betrouwbaarheidsinterval
 - Bovengrens 95% betrouwbaarheidsinterval

Volgende kenmerken worden opgeslagen voor analyses m.b.t. de schatting van het verschil tussen twee periodes.

- Naam analysevariabele
- Kenmerken analyse
 - Model-based
 - Datum analyse
 - Optioneel beschrijving van het model
- Kenmerken periode 1
 - Minimumjaar in dataset
 - Maximumjaar in dataset
 - Minimumreeks in dataset
 - Maximumreeks in dataset

- Aantal observaties
- Kenmerken periode 2
 - Minimumjaar in dataset
 - Maximumjaar in dataset
 - Minimumreeks in dataset
 - Maximumreeks in dataset
 - Aantal observaties
- Stratum
 - Ja/nee
 - Stratumnaam
- Eigenlijke resultaten
 - Verschil
 - Variabiliteit verschil
 - Benedengrens 95% betrouwbaarheidsinterval op verschil
 - Bovengrens 95% betrouwbaarheidsinterval op verschil
 - p-waarde

De bestandsnaam van de resultatendatabank is 'VBI_ResultatenDatabank'.

9 Overzicht R-scripts en databanken

9.1 Data

- Ruwe meetgegevens
 - 'bosinv1_2011.accdb': VBI1 dendrometrische opnames.
 - 'Bosveg4.mdb': VBI1 vegetatieopnames.
 - 'FieldMapData.mdb': VBI2 dendrometrische opnames en vegetatieopnames.
 - 'A2_idFULLID.shp': shapefile van segmenten in A2-plots VBI2.
 - 'A3_idFULLID.shp': shapefile van segmenten in A3-plots VBI2.
 - 'PointOnAreaOverlayTreesStandDescr.shp': shapefile met alle boomlocaties en met per boom de ID van het segment waarin de boom gelegen is.
- Gevalideerde meetgegevens
 - 'bosinv1_2011_val.accdb': VBI1 dendrometrische opnames.
 - 'Bosveg4_val.mdb': VBI1 vegetatieopnames.
 - 'FieldMapData_val.mdb': VBI2 dendrometrische opnames en vegetatieopnames.
- Meetproces
 - 'invb2_ID.shp': shapefile met alle steekproefpunten en overeenkomstige
 - IDPlots
 - Reeks
 - 'tblTreeDEF.accdb': koppeling tussen doorlopende boom-ID 'Old ID' in VBI2 en boom-ID per plot in VBI1.
 - 'VBI2_bos.csv': lijst van de plots waar bos werd vastgesteld na terreincontrole. De lijst is afgeleid uit het door ANB aangeleverde bestand 'QC_2020_INBO.xls'.
 - 'VBI2_geenBos.csv': lijst van de plots waar geen bos werd vastgesteld na terreincontrole. De lijst is afgeleid uit het aangeleverde bestand 'QC_2020_INBO.xls'.
 - 'OntbrekendeWaardenDendro.csv': lijst van de plots met ontbrekende waarden voor dendrometrische opnames en vegetatieopnames. Deze lijst geeft ook de oorzaak van ontbrekende waarde aan: ontoegankelijk (= ontbrekende waarde s.s.), geen bomen of geen vegetatie (= nulwaarneming). De lijst is afgeleid uit het aangeleverde bestand 'QC_2020_INBO.xls'.
- Externe data: gegevens nodig om analysevariabelen te berekenen, gebundeld in de databank 'VBI_ExterneData'. Deze databank bevat volgende tabellen:
 - inbo_vw_Kenmerken_Florabank1: kenmerken van de flora uit de floradatabank, waaronder Ellenberggetallen.
 - tblFiveMainTreeSpecies: lijst met de vijf belangrijkste boomsoorten.

- tblForestType: bostype en bostypegroep afgeleid uit de vegetatieopname via het script 'VBI_Bostype' en via bostypologie-software.
 - tblKarakteristiekeSoorten: lijst met sleutelsoorten van habitattype 9120 (als voorbeeld).
 - tblSpeciesListComb: gemeenschappelijke soortenlijst voor vegetatieopnames in VBI1 en VBI2.
 - tblSpeciesTreeLayerCharacteristics: eigenschappen van de soorten die in de boomlaag van de vegetatie voorkomen, maar niet in de lijst met boomsoorten van de dendrometrische opnames staan.
 - tblTariefgroepBoomsoort: geeft per boomsoort aan met welk tarief het volume berekend moet worden.
 - tblTarieven_1ing: tarieven met 1 ingang.
 - tblTarieven_2ing: tarieven met 2 ingangen.
 - tblTreeSpeciesCharacteristics: eigenschappen van de boomsoorten die in de soortenlijst van de dendrometrische opnames voorkomen.
 - tblZeldzaamheid: voorkomen van de soort in de periode 1972 – 2004, uitgedrukt als proportie van de 4 x 4 km hokken in Vlaanderen waar de soort aanwezig is.
- Strata
 - 'invb2_ecod_bodem_BHR.shp': shapefile met alle steekproefpunten en gegevens over:
 - Ecodistrict
 - Provincie
 - ANB-regio
 - Bodemtype
 - Eigenaarscategorie anno 2012

9.2 Output

- 'VBI_Meetproces.accdb' (§ 6.4)
- 'VBI_Strata.accdb' (§ 6.5)
- 'VBI_ValidatieDatabank.accdb' (§ 5)
- 'VBI_AnalyseDatabank.accdb' (§ 6.6)
- 'VBI_ResultatenDatabank.accdb' (§ 8)

9.3 Scripts

- Overzichtscript dat andere scripts uitvoert:
 - 'VBI_Main.R'
- Validatie

- 'Validatie_Anomalien.R'
- 'Validatie_UpdateDatabank.R'
- Analysedatabank
 - 'AnalyseDatabank_meetproces_inlezen.R'
 - 'AnalyseDatabank_meetproces_berekening.R'
 - 'AnalyseDatabank_strata_inlezen.R'
 - 'AnalyseDatabank_strata_berekening.R'
 - 'AnalyseDatabank_Dendro_inlezen.R'
 - 'AnalyseDatabank_Dendro_berekening.R': genereert volgende tabellen:
 - tbl2BestandskaraktKwal
 - tbl3BestandskaraktKwant
 - tbl6DoodHout
 - tbl10Boonsoorten
 - 'AnalyseDatabank_Vegetatie_inlezen.R'
 - 'AnalyseDatabank_Vegetatie_berekening.R': genereert volgende tabellen:
 - tbl4Vegetatie
 - tbl11Vegetatiesamenstelling
 - 'AnalyseDatabank_StructuurIndices_inlezen.R'
 - 'AnalyseDatabank_StructuurIndices_berekening.R'
 - 'AnalyseDatabank_Aanwas_inlezen.R'
 - 'AnalyseDatabank_Aanwas_berekening.R'
 - 'AnalyseDatabank_Houtkwaliteit_inlezen.R'
 - 'AnalyseDatabank_Houtkwaliteit_berekening.R'
 - 'AnalyseDatabank_AIWoodLayer_inlezen.R'
 - 'AnalyseDatabank_AIWoodLayer_berekening.R'
 - 'AnalyseDatabank_AIHerbLayer_inlezen.R'
 - 'AnalyseDatabank_AIHerbLayer_berekening.R'
 - 'AnalyseDatabank_AIBosstructuur_inlezen.R'
 - 'AnalyseDatabank_AIBosstructuur_berekening.R'
 - 'AnalyseDatabank_AIDeadWood_inlezen.R'
 - 'AnalyseDatabank_AIDeadWood_berekening.R'

- Functies
 - 'VBI_Functies.R'
- Eigenlijke analyse (zie Tabel 7-1)

Referenties

- Afdeling Bos & Groen, 2001. De bosinventarisatie van het Vlaamse Gewest. Resultaten vande eerste inventarisatie 1997-1999. Ministerie van de Vlaamse Gemeenschap.
- Bates D., Maechler M., Bolker B. and Walker S (2015). lme4: Linear mixed-effects models using S4 classes. R package version 1.1-8.
- Cornelis J., Hermy M., De Keersmaecker L. & Vandekerckhove K. (2007). Bosplantengemeenschappen in Vlaanderen. Een typologie van bossen op basis van de kruidachtige vegetatie. Rapport INBO.R.2007.1. Instituut voor Natuur- en Bosonderzoek en K.U.Leuven, afdeling Bos, Natuur en Landschap in opdracht van de Vlaamse Overheid, agentschap voor Natuur en Bos, Brussel.
- Dagnelie, P., Palm, R., Rondeux, J., & Thill., A. (1985). Tables de cubage des arbres et des peuplements forestiers. Gembloux: Les presses agronomiques de Gembloux.
- Dik, E. (1990). De schatting van volumes en werkhoutlengten bij populier. Wageningen: De Dorschkamp.
- Jansen I. & Quataert P. (2013), Een intuïtieve inleiding op de mixed model regressietechniek, studie uitgevoerd in opdracht van de Vlaamse Milieumaatschappij, MIRA, MIRA/2013/03, INBO.
- Onkelinx T., Verschelde P., Wouters J., Bauwens D., & Quataert P. (2008). Ontwerp en evaluatie van meetnetten voor het milieu- en natuurbeleid. Dimensionering en kosteneffectiviteit. Rapporten van het Instituut voor Natuur- en Bosonderzoek (INBO.R.2007.27), Brussel.
- Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2015). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-121
- Quataert, P., Van der Aa, B. Verschelde, P. (2011). Opstellen van tarieven voor Inlandse eik en Beuk in Vlaanderen ten behoeve van het berekenen van houtvolumes. Statistische evaluatie van de regressiemodellen en overzicht van de resultaten (technisch rapport deel III).. Rapporten van het Instituut voor Natuur- en Bosonderzoek 2011 (18). Instituut voor Natuur- en Bosonderzoek, Brussel.
- Van Den Meersschaut D., Vandekerckhove K., Van de Kerckhove P., Delbecque F., & Van Slycken J. (2001). Selectie en evaluatie van indicatoren en uitwerking van een praktisch bruikbare methodologie voor de beoordeling van biodiversiteit in bossen. Eindrapport project Vlaams Impulsprogramma Natuurontwikkeling VLINA/C96/04. Rapporten van het instituut voor bosbouw en wildbeheer (IBW Bb R.2001.009), Brussel.
- Van Landuyt, W., Hoste, I., Vanhecke, L., Van den Bremt, P., Vercruyse, W. & De Beer, D. 2006. Atlas van de Flora van Vlaanderen en het Brussels Gewest. Instituut voor natuur- en bosonderzoek, Nationale Plantentuin van België & Flo.Wer.
- Wickham (2015a). plyr: Tools for Splitting, Applying and Combining Data. R package version 1.8.3.
- Wickham (2015b). ggplot2: An implementation of the grammar of graphics. R package version 1.0.1.
- Wouters J., Onkelinx T., Bauwens D. & Quataert P. (2008a). Ontwerp en evaluatie van meetnetten voor het milieu- en natuurbeleid. Leidraad voor de meetnetontwerper.

Vlaamse Overheid, Departement Leefmilieu, Natuur en Energie, Instituut voor Natuur- en Bosonderzoek, Brussel.

Wouters J., Quataert P., Onkelinx T. & Bauwens D. (2008b). Ontwerp en handleiding voor de tweede regionale bosinventarisatie van het Vlaamse Gewest. Rapporten van het Instituut voor Natuur- en Bosonderzoek 2008 (INBO.R.2008.17). Instituut voor Natuur- en Bosonderzoek, Brussel.

Wouters J. (2008c). Discussienota structuurindices. Instituut voor Natuur- en Bosonderzoek, Brussel.

Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A. & Smith G.M. (2009). Mixed effects models and extensions in ecology with R: Springer.

Bijlage 1: Indeling van de meetvragen

Groep	Nr. meetvraag	Meetvraag	Analysevariabele	Type	Naam variabele in analysedatabank
1. Bosoppervlakte	1.a.	Wat is het aandeel van de verschillende eigenaarscategorieën?	bos/niet-bos	Binair	v1_ForestCover
	1.b.	Wat is het aandeel in de verschillende provincies?	bos/niet-bos	Binair	v1_ForestCover
	6.3.a.	Bosoppervlakte: Wat is de oppervlakte bos (absoluut en relatief) per bostypegroep?	bos/niet-bos	Binair	v1_ForestCover
2. Bestandskarakteristieken kwalitatief	1.c.	Wat is het aandeel aan homogene bestanden?	homogeen/gemengd	Binair	v1_HomgeneStand
	1.i.	Welk aandeel van de doelpopulatie valt op een bosrand (intern of extern)?	Bosrand (ja/nee)	Binair	v7_ForestEdge
	1.j.	Welk aandeel van de steekproefpunten valt in een overgangssituatie?	Overgangssituatie (ja/nee)	Binair	v8_ForestOvergang
	3.a.	Wat is het aandeel aan loofhout, naaldhout, gemengd loofhout en gemengd naaldhout?	Bestandstype	Categorisch	v2_StandType
	3.b.	Wat is het aandeel aan hooghout, middelhout en hakhout?	Bedrijfsvorm	Categorisch	v3_HarvestType
	3.c.	Wat is het aandeel aan jongwas, dichtwas, staakhout en boomhout?	Ontwikkelingsfase	Categorisch	
	3.d.	Wat is het aandeel van de mengingsvormen stamsgewijs, groepsgewijs en homogeen?	Mengingsvorm	Categorisch	v4_MixType
	3.e.	Wat is de verdeling van de bestandsleeftijden?	Bestandsleeftijd	Categorisch	v5_StandAge
	3.f.	Wat is de verdeling van de bestandsleeftijden van de homogene bestanden van de 5 belangrijkste boomsoorten?	Bestandsleeftijd	Categorisch	v5_StandAge
	3.g.	In welk aandeel van de steekproefpunten is windworp aanwezig?	Windworp	Binair	v6_WindThrow
3. Bestandskarakteristieken kwantitatief	1.d.	Wat is het aandeel aan homogene dennenbossen?	homogeen dennenbos (ja/nee)	Binair	v1_HomogeenDen
	1.e.	Wat is het aandeel aan populierenplantages?	populierplantage (ja/nee)	Binair	v2_HomogeenPopulier
	1.f.	Wat is het aandeel aan bestanden gedomineerd door uitheemse boomsoorten?	bestand gedomineerd door uitheemse boomsoorten (ja/nee)	Binair	v3_GedomineerdExotic
	4.2.a.	Welke bosoppervlakte is gedomineerd door geïntroduceerde boomsoorten?	bestand gedomineerd door uitheemse boomsoorten (ja/nee)	Binair	v3_GedomineerdExotic
	1.g.	Wat is het aandeel aan gemengde bestanden op basis van inheemse en standplaatsgeschikte boomsoorten?	gemengd, inheems bestand (ja/nee)	Binair	v4_GemengdInheems

Groep	Nr. meetvraag	Meetvraag	Analysevariabele	Type	Naam variabele in analysedatabank
	2.e.	Is het gemiddeld grondvlakaandeel aan inheemse boomsoorten in de totale populatie gelijk of groter dan 80 %?	grondvlakaandeel inheemse boomsoorten	Continu_Proportie	v5_FractionBasalAreaAutocht
	4.2.b.	Wat is de kans op aanwezigheid en het grondvlakaandeel van inheemse / uitheemse / invasieve boomsoorten?	voorkomen en grondvlakaandeel van inheemse, uitheemse en invasieve boomsoorten	Binair en Continu_proportie	v5_FractionBasalAreaAutocht; v6_FractionBasalAreaExotic; v7_FractionBasalAreaInvasive; v8_PresenceAutocht; v9_PresenceExotic; v10_PresenceInvasive
	6.1.a.	Wat is de totale (m ³) en gemiddelde (m ³ /ha) houtvoorraad?	Volume (m ³ /ha)	Continu_positief	v11_TotalVolume
	6.3.b.	Houtvoorraad: Wat is de totale (m ³) en gemiddelde (m ³ /ha) houtvoorraad per bostypegroep?	Volume (m ³ /ha)	Continu_positief	v11_TotalVolume
4. Vegetatie	2.f.	Wat is de kans op voorkomen en de gemiddelde abundantie van NV (als NV voorkomt) van de 5 belangrijkste boomsoorten?	Voorkomen en abundantie van natuurlijke verjonging (NV) inlandse eik, grove den, Corsicaanse den, beuk en berk	Binair; Continu_positief	v15_NVInlandseeikPres; v16_NVGrovedenPres; v17_NVCorsicaansedenPres; v18_NVBeukPres; v19_NVBerkPres; v20_NVInlands eikAbund; v21_NVGrovedenAbund; v22_NVCorsicaansedenAbund; v23_NVBeukAbund; v24_NVBerkAbund
	2.g.	Wat is de kans op voorkomen en gemiddelde abundantie van NV invasieve boomsoorten, in het bijzonder Amerikaanse vogelkers (als NV voorkomt)?	Voorkomen en abundantie van NV van invasieve boomsoorten	Binair; Continu_positief	v5_NVInvasivePres; v8_NVInvasiveAbund
	4.1.a.	Wat is de oppervlakte bos (ha) geordend volgens het aantal aanwezige boomsoorten?	Aantal boomsoorten	Aantal	v1_NbTreeSp_Vegopname
	4.1.b.	Wat is het gemiddeld aantal plantensoorten?	Aantal plantensoorten	Aantal	v2_NbSp
	4.1.c.	Wat is de kans op aanwezigheid van 2 of meer zeldzame plantensoorten?	Aantal zeldzame plantensoorten	Aantal	v12_NbRareSp
	4.1.d.	Wat is de gemiddelde indexwaarde van Simpson?	Indexwaarde van Simpson	Continu_positief	v13_SimpsonIndex
	4.1.e.	Wat is de gemiddelde indexwaarde van Shannon-Wiener?	Indexwaarde van Shannon-Wiener	Continu_positief	v14_ShannonWienerIndex
	4.2.c.	Wat is de abundantie aan verjonging van deze soorten (<i>inheemse / uitheemse / invasieve boomsoorten</i>) (cijfers op basis van plaatsen waar verjonging effectief voorkomt)?	Abundantie van NV van inheemse, uitheemse en invasieve boomsoorten	Continu_positief	v6_NVExoticAbund; v7_NVAutochtAbund; v8_NVInvasiveAbund
	4.2.d.	Wat is gemiddelde abundantie van uitheemse plantensoorten in de vegetatielaag?	Abundantie uitheemse plantensoorten	Continu_positief	Geen lijst met uitheemse plantensoorten beschikbaar
	4.3.d.	Wat is de kans op aanwezigheid van de drie vegetatielagen (kruidlaag, struiklaag en boomlaag)?	Aanwezigheid van de drie vegetatielagen	Binair	v9_TreeLayer; v10_ShrubLayer; v11_HerbLayer
	5.a.	Wat is de gemiddelde Ellenberg-L, -V, -R en -N indicatorwaarde?	Ellenberg-L, -V, -R en -N indicatorwaarde	Continu_positief	v27_L; v28_N; v29_R; v30_V

Groep	Nr. meetvraag	Meetvraag	Analysevariabele	Type	Naam variabele in analysedatabank
	5.b.	Wat is de verdeling van de Ellenberg-L, -V, -R en -N indicatorwaarde?	Ellenberg-L, -V, -R en -N indicatorwaarde	Continu_positief	v27_L; v28_N; v29_R; v30_V
	5.c.	Wat is de gemiddelde abundantie van karakteristieke plantensoorten of groepen van plantensoorten?	Abundantie van een (groep van) soort(en)	Continu_positief	v26_AbundSpeciesList
	5.d.	Wat is gemiddeld aantal voorkomende soorten binnen een groep van plantensoorten?	Aantal van een groep van soorten	Aantal	v25_PresSpeciesList
	6.3.e.	Boomsoortensamenstelling: Wat is de oppervlakte bos (ha) geordend volgens het aantal aanwezige boomsoorten en volgens bostypegroep?	Aantal boomsoorten	Aantal	v1_NbTreeSp
	6.3.f.	Verjonging: Wat is voor gelijkjarige en ongelijkjarige bestanden het aandeel (aantal steekproefpunten) en de oppervlakte (ha) waar verjonging voorkomt gerangschikt per verjongingstype?	Voorkomen NV	Binair	v0_NV
5. Ruimtelijke structuur	4.3.a.	Wat is de gemiddelde Positioning Index van Clark and Evan?	Positioning Index van Clark and Evan	Continu_index	v1_ClarkAndEvan; v2_ClarkAndEvan_Simple
	4.3.b.	Wat is de gemiddelde Mixture Index van Von Gadow?	Mixture Index van Von Gadow	Continu_index	v3_vonGadowSpeciesMix
	4.3.c.	Wat is de gemiddelde Height Differentiation Index van Von Gadow?	Height Differentiation Index van Von Gadow	Continu_index	v4_vonGadowHeight
6. Dood hout	4.4.a.	Wat is de kans op aanwezigheid van een dood-hout element?	Voorkomen dood hout	Binair	v2_PresencesSnags; v4_PresenceLogs
	4.4.b.	Wat is het gemiddelde volume dood hout (m ³ /ha) (staand en liggend)?	Volume dood hout	Continu positief	v2_VolumeSnags; v4_VolumeLogs
	4.4.c.	Is het gemiddelde aandeel (%) dood hout (staand en liggend) groter dan 4 % van het totale volume?	Volumeaandeel dood hout	Continu proportie	v2_VolumeSnags; v4_VolumeLogs
7. Authenticiteitsindex (AI)	4.5.a.	Wat is de globale Authenticiteitsindex?	Score AI totaal	Discreet	v1_AITotal
	4.5.b.	Wat is de score voor de pijler bosstructuur?	Score bosstructuur	Discreet	v2_AIStructure
	4.5.c.	Wat is de score voor de pijler houtige vegetatie?	Score boomlaag	Discreet	v3_AIWoodlayer
	4.5.d.	Wat is de score voor de pijler kruidlaag?	Score kruidlaag	Discreet	v4_AIHerbLayer
	4.5.e.	Wat is de score voor de pijler dood hout?	Score dood hout	Discreet	v5_AIDeadWood
8. Aanwas	6.1.b.	Wat is voor de homogene bestanden van de 5 belangrijkste boomsoorten de gemiddelde lopende jaarlijkse aanwas per ha? Ligt deze hoger dan 5 m ³ /ha/jaar?	Lopende jaarlijkse aanwas	Continu	v1_LJA

Groep	Nr. meetvraag	Meetvraag	Analysevariabele	Type	Naam variabele in analysedatabank
	6.1.c.	Wat is de gemiddelde lopende jaarlijkse bestandsaanwas per bostypegroep? Licht deze hoger dan 5 m ³ /ha/jaar?	Lopende jaarlijkse aanwas	Continu	v1_LJA
	6.1.d.	Wat is voor de 5 belangrijkste boomsoorten in homogene bestanden het gemiddelde afgestorven bestandsvolume (m ³ /ha/jaar)?	Afgestorven bestandsvolume	Continu	v2_DeadStanding
	6.1.e.	Wat is het gemiddelde afgestorven bestandsvolume (m ³ /ha/jaar) per bostypegroep?	Afgestorven bestandsvolume	Continu	v2_DeadStanding
9. Houtkwaliteit	6.2.a.	Wat is voor de economisch belangrijkste boomsoorten het totale volume (m ³) aan industrie-, brand- en pulphout enerzijds (klasse C) en kwaliteitsvol (klasse B) en hoog kwaliteitsvol (klasse A) hout anderzijds?	Kwaliteitsklasse; volume	Continu positief	v1_QualityClass; v2_Volume
	6.2.b.	Wat is de verdeling van de houtvolumesortimenten i.f.v. de bestandsleeftijd voor cultuurpopulier en gelijkjarige beuken-, eiken- en dennenbestanden?	Kwaliteitsklasse; volume	Continu positief	v1_QualityClass; v2_Volume
	6.2.c.	Wat is per boomsoort voor de klassen A en B het relatieve aandeel van de verschillende defecten ?	Defecttype	Categorisch	v1_QualityClass; v3_StemDefects
	6.2.d.	Wat is voor de economisch belangrijkste boomsoorten het volume aan: <ul style="list-style-type: none"> • takvrije stammen van minimaal 6 m lengte zonder defecten, met een gering verloop en met een minimale onderdiameter; • takvrije stammen van minimaal 6 m lengte met maximaal 1 defect en/of een aanvaardbaar verloop en een minimale onderdiameter; • takvrije stammen van minimaal 6 m lengte die niet aan bovenstaande criteria voldoen; • takvrije stammen korter dan 6 m zonder defecten, met een gering verloop en met een minimale onderdiameter? 	Stamtype; takkenvrij stamvolume	Continu positief	v4_StemType; v5_VolumeTakvrij
10. Boomsaortensamenstelling	2.a.	Wat is van de 10 belangrijkste boomsoorten het procentuele aandeel in het totale stamtaal van de populatie?	stamtaal soort	Continu positief	v1_Stamtaal_A3A4
	2.b.	Wat is van de 10 belangrijkste boomsoorten het procentuele aandeel in het totale grondvlak/volume van de populatie?	Grondvlak soort; volume soort	Continu positief	v2_BasalArea; v3_Volume

Groep	Nr. meetvraag	Meetvraag	Analysevariabele	Type	Naam variabele in analysedatabank
	2.c.	Wat is voor de 5 belangrijkste boomsoorten het aandeel aan de homogene en heterogene bestanden?	Mengingsvorm; voorkomen soort; hoofdboomsoort	Continu positief	v1_StamtaI_A3A4

Bijlage 2: voorbeeld output van geautomatiseerde analyse houtvoorraad

Houtvoorraad

Toon Westra & Pieter Verschelde

19 juni 2015

Inleiding

Dit document beschrijft een voorbeeld van een geautomatiseerde rapportage via de R Markdown tool.

Dit document is geen typische rapportage en probeert ook wat de basisprincipes van de interpretatie te illustreren. Dit rapport overloopt de stappen in de analyse van de meetvraag met betrekking tot de houtvoorraad.

De code van de R scripts wordt niet getoond in het document, maar kan je achterhalen via het .Rmd bronbestand. Je kan ervoor kiezen de R code ook te laten tonen door in de chunks `echo=TRUE` te gebruiken.

De geautomatiseerde rapportage kan nog uitgebreid worden door stijlen te wijzigen en inline text dynamisch te laten genereren op basis van een R script.

Gemiddelde houtvoorraad

Inleiding

Deze analyse beantwoordt volgende vragen:

- **Vraag 6.1.a.:** Wat is de totale (m^3) en gemiddelde (m^3/ha) houtvoorraad?
- **Vraag 6.3.b.:** Houtvoorraad: Wat is de totale (m^3) en gemiddelde (m^3/ha) houtvoorraad per bostypegroep?

Invoer analysetabel

Op basis van onderstaande tabel kan je verifiëren of je data goed is ingelezen:

- Zijn numerieke variabelen echt wel numeriek?
- Liggen de variabelen in de verwachte range?
- Hoeveel NA-waarden vind je per variabele?

IDPlots	IDSegments	Periode	v1_HomogeenDen
Min. : 17144	Min. :1.00	Min. :1.000	Min. :0.0000
1st Qu.:156121	1st Qu.:1.00	1st Qu.:1.000	1st Qu.:0.0000
Median :221014	Median :1.00	Median :1.000	Median :0.0000

Mean :209399	Mean :1.06	Mean :1.368	Mean :0.2683
3rd Qu.:263133	3rd Qu.:1.00	3rd Qu.:2.000	3rd Qu.:1.0000
Max. :424004	Max. :6.00	Max. :2.000	Max. :1.0000
NA	NA's :2	NA	NA's :15

Table continues below

v2_HomogeenPopulier	v3_GedomineerdExotic	v4_GemengdInheems
Min. :0.00000	Min. :0.0000	Min. :0.0000
1st Qu.:0.00000	1st Qu.:0.0000	1st Qu.:0.0000
Median :0.00000	Median :0.0000	Median :0.0000
Mean :0.07777	Mean :0.2019	Mean :0.1509
3rd Qu.:0.00000	3rd Qu.:0.0000	3rd Qu.:0.0000
Max. :1.00000	Max. :1.0000	Max. :1.0000
NA's :15	NA's :15	NA's :13

Table continues below

v5_FractionBasalAreaAutocht	v6_FractionBasalAreaExotic
Min. :0.0000	Min. :0.00000
1st Qu.:0.6624	1st Qu.:0.00000
Median :0.9945	Median :0.00551
Mean :0.7690	Mean :0.23099
3rd Qu.:1.0000	3rd Qu.:0.33756
Max. :1.0000	Max. :1.00000
NA's :245	NA's :245

Table continues below

v7_FractionBasalAreaInvasive	v8_PresenceAutocht
------------------------------	--------------------

Min. :0.0000	Mode :logical
1st Qu.:0.0000	FALSE:427
Median :0.0000	TRUE :3773
Mean :0.0154	NA's :20
3rd Qu.:0.0000	NA
Max. :1.0000	NA
NA's :245	NA

Table continues below

v9_PresenceExotic	v10_PresenceInvasive	v11_TotalVolume
Mode :logical	Mode :logical	Min. : 0.0
FALSE:1989	FALSE:3262	1st Qu.: 127.0
TRUE :2211	TRUE :938	Median : 221.6
NA's :20	NA's :20	Mean : 236.0
NA	NA	3rd Qu.: 319.2
NA	NA	Max. :2426.1
NA	NA	NA's :13

Table continues below

AutochtBasalArea	ExoticBasalArea	InvasiveBasalArea
Min. : 0.000	Min. : 0.000	Min. : 0.0000
1st Qu.: 6.122	1st Qu.: 0.000	1st Qu.: 0.0000
Median : 20.167	Median : 0.000	Median : 0.0000
Mean : 19.749	Mean : 7.232	Mean : 0.3448
3rd Qu.: 29.724	3rd Qu.: 7.057	3rd Qu.: 0.0000
Max. :262.059	Max. :95.243	Max. :21.5326

NA's :15

NA's :15

NA's :15

Table continues below

TotalBasalArea	MaxStamtalSpeciesName	MaxBasalAreaSpeciesName
Min. : 0.00	Grove den : 543	Grove den :1119
1st Qu.: 17.75	Berk : 408	Populier : 599
Median : 27.21	Populier : 356	Corsicaanse den: 373
Mean : 26.98	Zomereik : 340	Zomereik : 297
3rd Qu.: 35.50	Corsicaanse den: 251	Berk : 296
Max. :273.89	(Other) :2194	(Other) :1413
NA's :13	NA's : 128	NA's : 123

Table continues below

Reeks	StartPeriode	IDGroup	DateDendro
Min. : 1.00	Min. :1.000	173067_1: 4	Min. :1997
1st Qu.: 3.00	1st Qu.:1.000	173155_1: 4	1st Qu.:1998
Median : 4.00	Median :1.000	263105_1: 4	Median :1999
Mean : 4.71	Mean :1.232	96126_2 : 4	Mean :2003
3rd Qu.: 7.00	3rd Qu.:1.000	136077_2: 3	3rd Qu.:2011
Max. :10.00	Max. :2.000	138028_2: 3	Max. :2014
NA	NA	(Other) :4198	NA

Table continues below

Replaced	PlotWeight	SegmentWeight
Mode :logical	Min. :0.0620	Min. :0.000
FALSE:3819	1st Qu.:1.0000	1st Qu.:1.000
TRUE :401	Median :1.0000	Median :1.000
NA's :0	Mean :0.9751	Mean :0.968
NA	3rd Qu.:1.0000	3rd Qu.:1.000

NA Max. :1.0000 Max. :1.000

NA NA's :3 NA's :3

Table continues below

LandUse	StandType	OwnerType
bos :4126	loofhout :2144	Private:2582
bos - kapvlakte : 33	naaldhout :1385	Public :1622
open ruimte binnen bos: 45	gemengd naaldhout : 360	NA's : 16
NA's : 16	gemengd loofhout : 237	NA
NA	open ruimte binnen bos: 45	NA
NA	(Other) : 33	NA
NA	NA's : 16	NA

Table continues below

ForestTypeGroupCode	ForestTypeGroup	Province
I :1929	Dennen-Eikenbos:1929	ANT :1388
D : 450	Essen-Elzenbos : 450	LMB :1377
G : 119	Essen-Eikenbos : 119	OVL : 512
H : 114	Eiken-Beukenbos: 114	VBR : 751
J : 85	Berken-Elzenbos: 85	WVL : 188
(Other): 166	(Other) : 166	NA's: 4
NA's :1357	NA's :1357	NA

Table continues below

EcoDistrict	SoilType
Ecoregio van de Kempen :2494	Vochtig zand: 944
Ecoregio van de zuidoostelijke heuvelzone : 506	Droog zand : 717

Ecoregio van de Pleistocene rivier valleien: 350	Nat zand : 397
Ecoregio van de cuesta's : 254	Antropogeen : 335
Ecoregio van de zuidwestelijke heuvelzone : 228	Nat zandleem: 285
(Other) : 387	(Other) :1538
NA's : 1	NA's : 4

Table continues below

ANBRegion

Hoge Kempen : 480

Schelde - Neteland : 386

Netebronnen - Zuiderkempen:
378

West-Limburg : 362

Turnhoutse Kempen : 343

(Other) :2270

NA's : 1

Aanmaken analyseset

In een volgende stap wordt de analyseset (standaard op segmentniveau) en de analyseset per plot gemaakt, specifiek voor alle berekeningen omtrent houtvoorraad. Deze set bestaat uit volgende deelvariabelen (de set per plot bevat natuurlijk de segmentinformatie niet meer):

	Type	AantalNA
IDPlots	integer	0
IDSegments	numeric	0
Periode	numeric	0
IDGroup	factor	0
PlotWeight	numeric	0
SegmentWeight	numeric	0
Reeks	integer	0
DateDendro	numeric	0
StartPeriode	numeric	0
v11_TotalVolume	numeric	0

v1_HomogeenDen	numeric	2
StandType	factor	0
OwnerType	factor	0
ForestTypeGroupCode	factor	1301
fPeriode	factor	0
Paired_1_2	logical	0

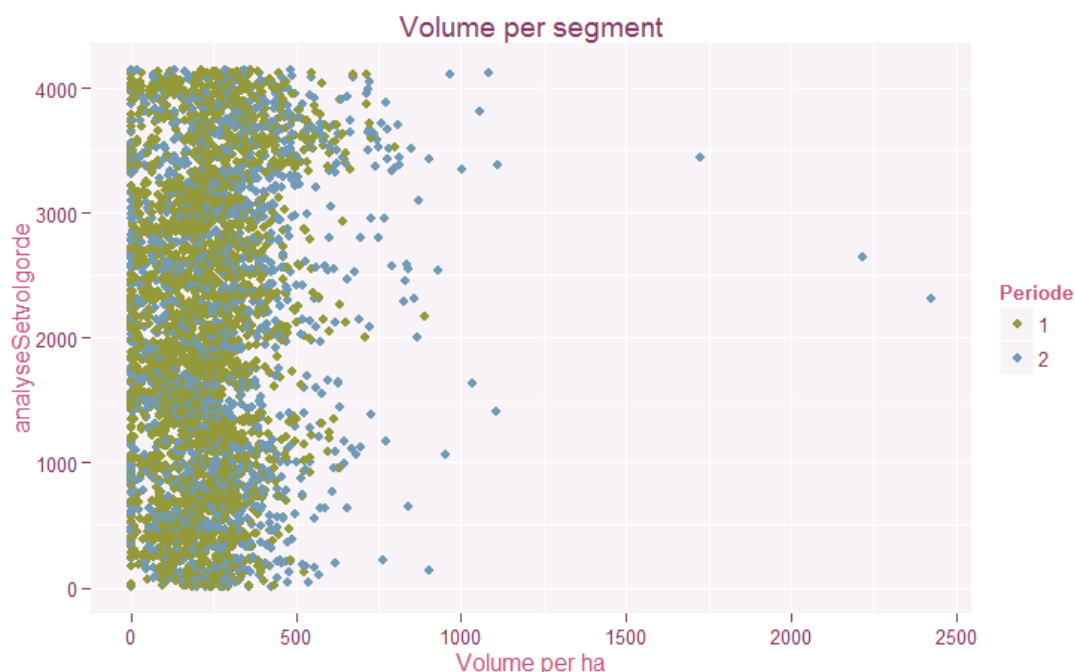
Variabelen in de analyseset

In bovenstaande tabel zie je veel NA's voor ForestTypeGroupcode. Dit komt omdat de bostypegroep bepaald wordt op basis van de vegetatie-opname en in VBI1 slechts in de helft van de plots een vegetatie-opname werd uitgevoerd.

Data verkenning

Dotplots

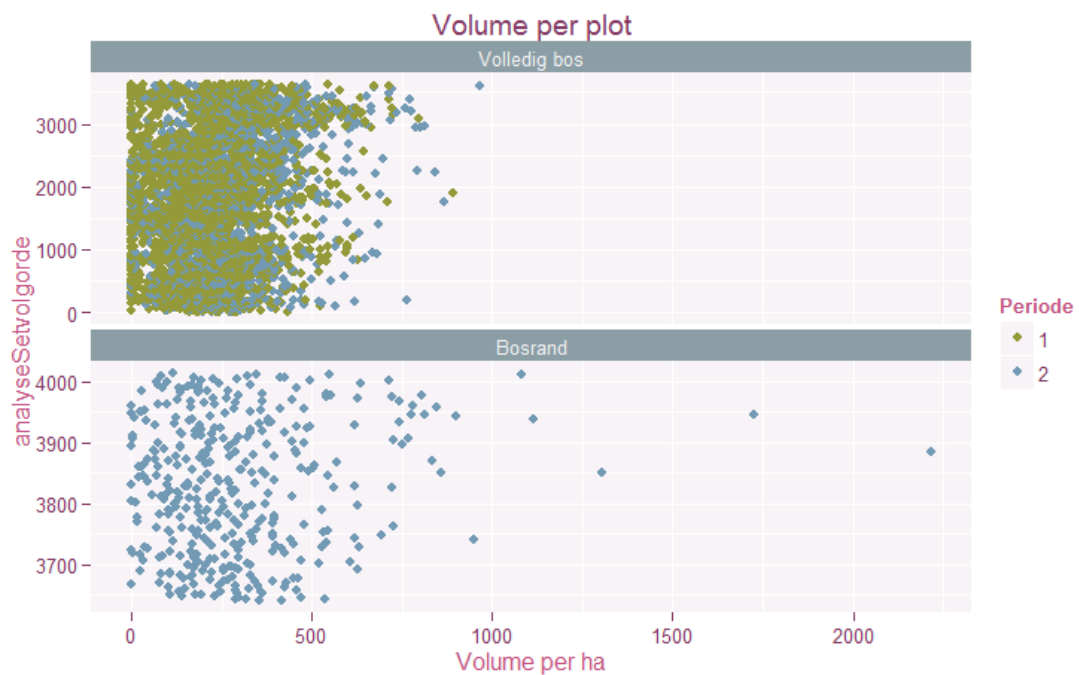
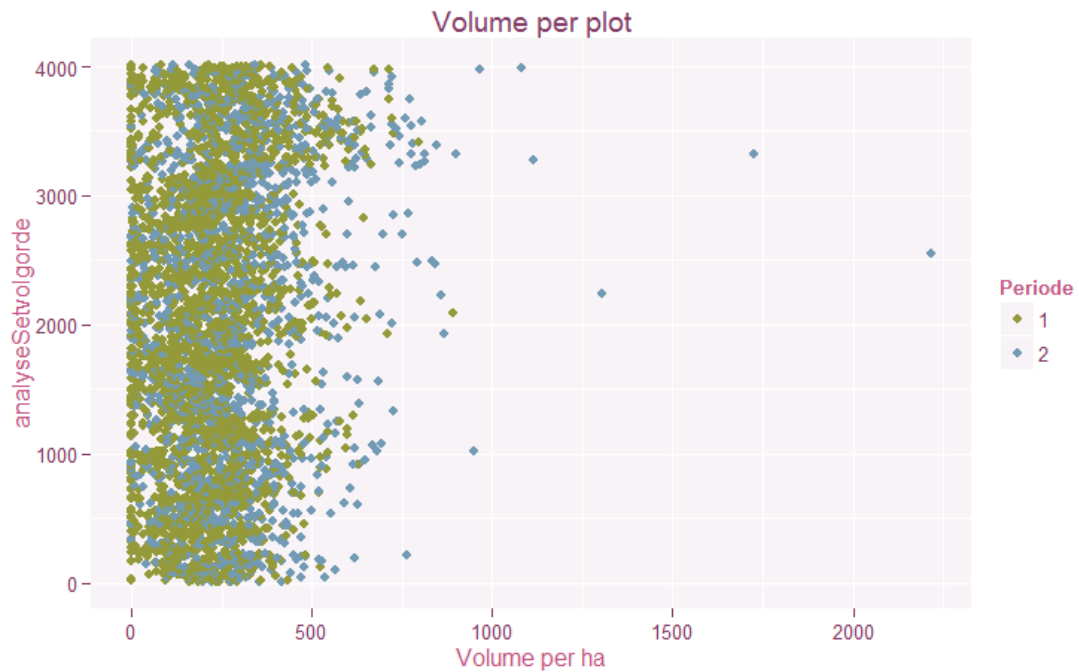
Dotplots hebben over het algemeen niet zoveel informatie-inhoud, behalve dat je een zeer snelle kijk kan hebben op de aanwezigheid van outliers. Zo zie je in onderstaande figuur dat er in sommige segmenten zeer hoge volumes >1000 kubieke meter per hectare geschat worden. Verder blijken deze zeer hoge volumes zich vooral voor te doen bij de gesegmenteerde plots.



Hier zie je de datarijen met de outliers:

```
## IDPlots IDSegments Periode IDGroup PlotWeight SegmentWeight Reeks
## 1437 172011 1 2 172011_2 0.701 0.35092725 3
## 1661 175129 2 2 175129_2 1.000 0.14200000 5
## 2362 242155 1 2 242155_2 0.673 0.39970282 1
## 2696 252142 1 2 252142_2 0.659 1.00000000 2
## 3409 282150 2 2 282150_2 1.000 0.28500000 2
```

##	3455	296073	1	2	296073_2	0.062	1.00000000	5
##	3504	304009	1	2	304009_2	0.581	1.00000000	3
##	3880	325023	3	2	325023_2	0.932	0.09871245	2
##	4189	384034	1	2	384034_2	0.542	1.00000000	3
##	DateDendro	StartPeriode	v11_TotalVolume	v1_HomogeenDen	StandType			
##	1437	2011	2	1105.074	0 loofhout			
##	1661	2013	2	1032.048	0 loofhout			
##	2362	2010	2	2426.118	1 naaldhout			
##	2696	2010	2	2217.936	0 loofhout			
##	3409	2011	2	1002.384	0 loofhout			
##	3455	2014	2	1112.838	0 loofhout			
##	3504	2012	2	1724.866	0 loofhout			
##	3880	2010	2	1055.849	0 loofhout			
##	4189	2012	2	1082.909	0 loofhout			
##	OwnerType	ForestType	GroupCode	fPeriode	Paired_1_2			
##	1437	Private	I	2	FALSE			
##	1661	Private	I	2	FALSE			
##	2362	Private	I	2	FALSE			
##	2696	Private	I	2	FALSE			
##	3409	Public	I	2	FALSE			
##	3455	Private	G	2	FALSE			
##	3504	Private	E	2	FALSE			
##	3880	Private	D	2	FALSE			
##	4189	Private	D	2	FALSE			



Onderzoeken van aantal nullen

In VBI1 zijn er meer plots zonder (opgemeten) houtvolume dan in VBI2.

Percentage nullen:

Periode 1 segmentniveau: 4%

Periode 2 segmentniveau: 5.5%

Periode 1 plotniveau : 4%

Periode 2 plotniveau : 1.3%

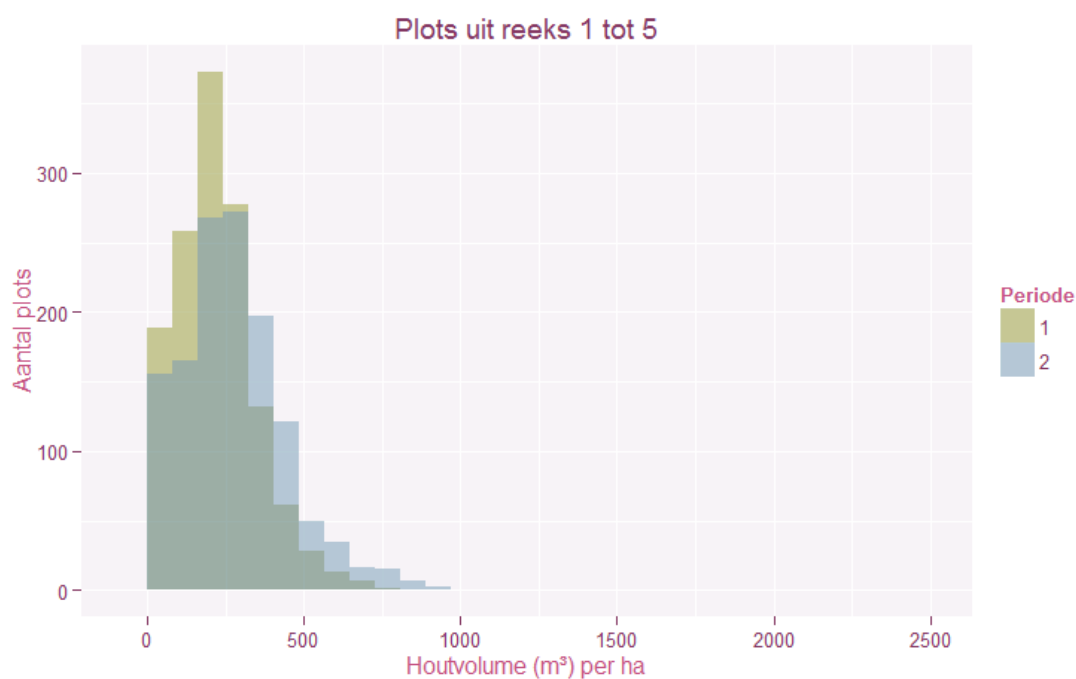
Histogrammen

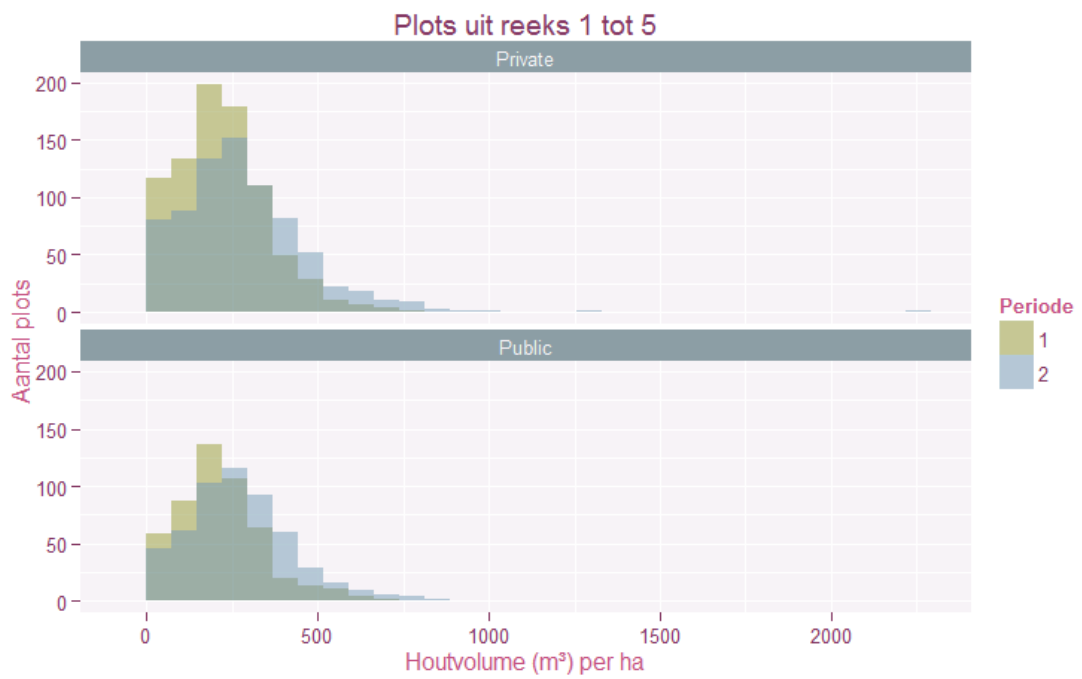
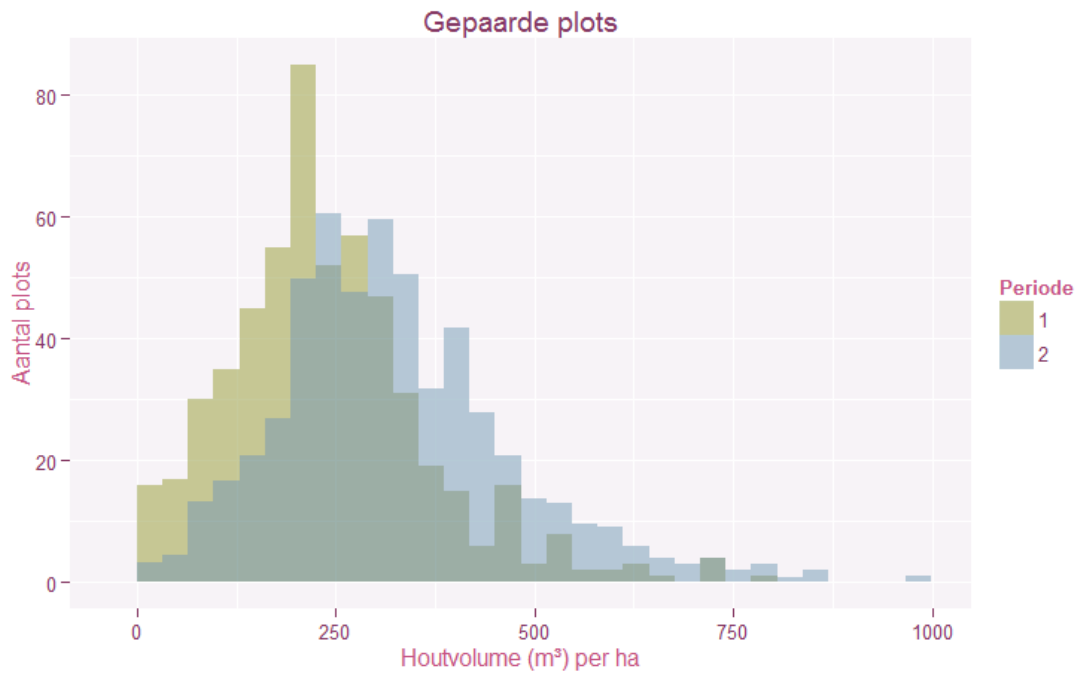
De histogrammen tonen de volledige distributies van het houtvolume voor de geselecteerde groepen. Deze zijn semitransparant zodat je de verdelingen volledig kan zien, en een indicatie hebt van hoe sterk ze overlappen. Indien echter het ene aantal veel hoger is dan het andere, dan zal dat grafisch moeilijker te interpreteren zijn.

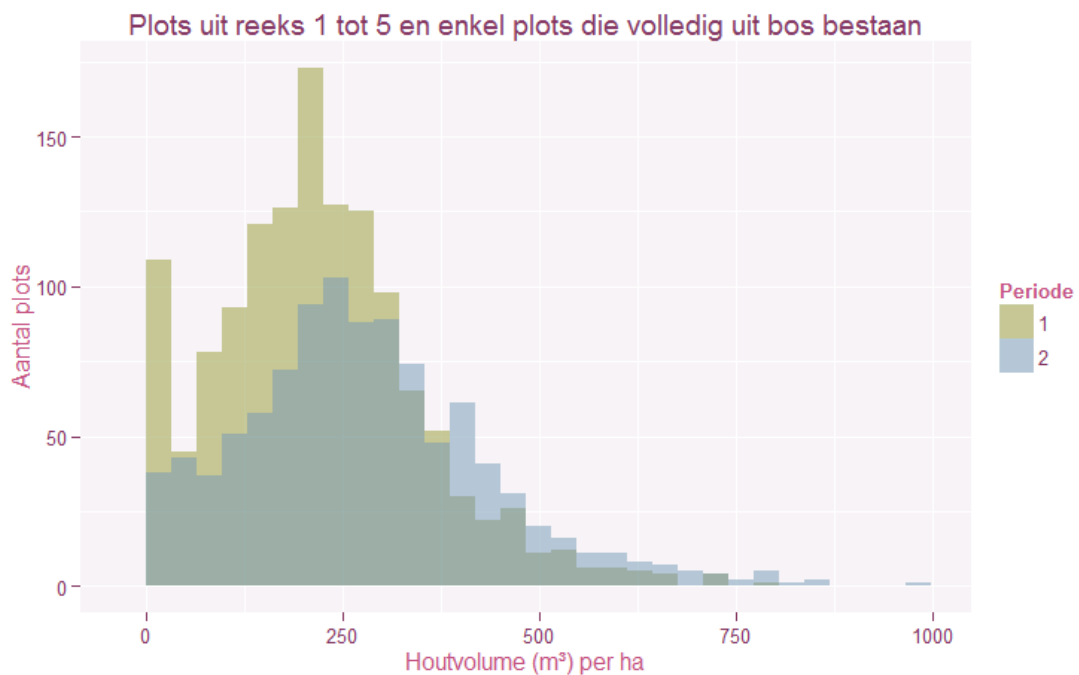
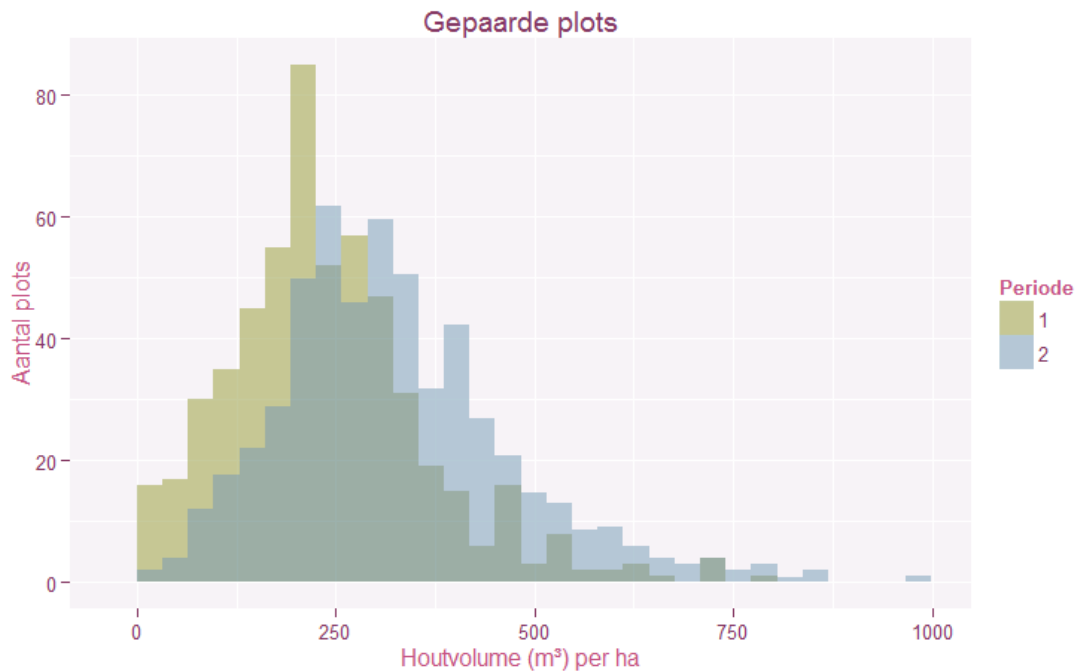
In de histogramma hieronder worden de aantallen daarnaast ook nog eens gewogen volgens hun plot of segmentgewicht, zodat plots met weinig bos of een klein segment minder doorweegt.

Wat zien we in onderstaande figuren:

- In periode 2 komen hogere houtvolumes meer voor en lijkt er een algemene verhoging in het volume, dit geldt zowel voor gepaarde als voor niet-gepaarde plots
- In de publieke en private bossen zien we een gelijkaardig patroon in evolutie in houtvolume per periode
- Voor plots die volledig uit bos bestaan lijkt dit groeieffect minder uitgesproken







Density Plots

Een alternatief voor de histogrammen zijn density-plots. Dit zijn histogrammen die gewogen worden zodat de oppervlakte onder de curve altijd 1 is. op die manier kunnen meerdere density-plots op 1 grafiek beter met elkaar kunnen vergeleken worden.

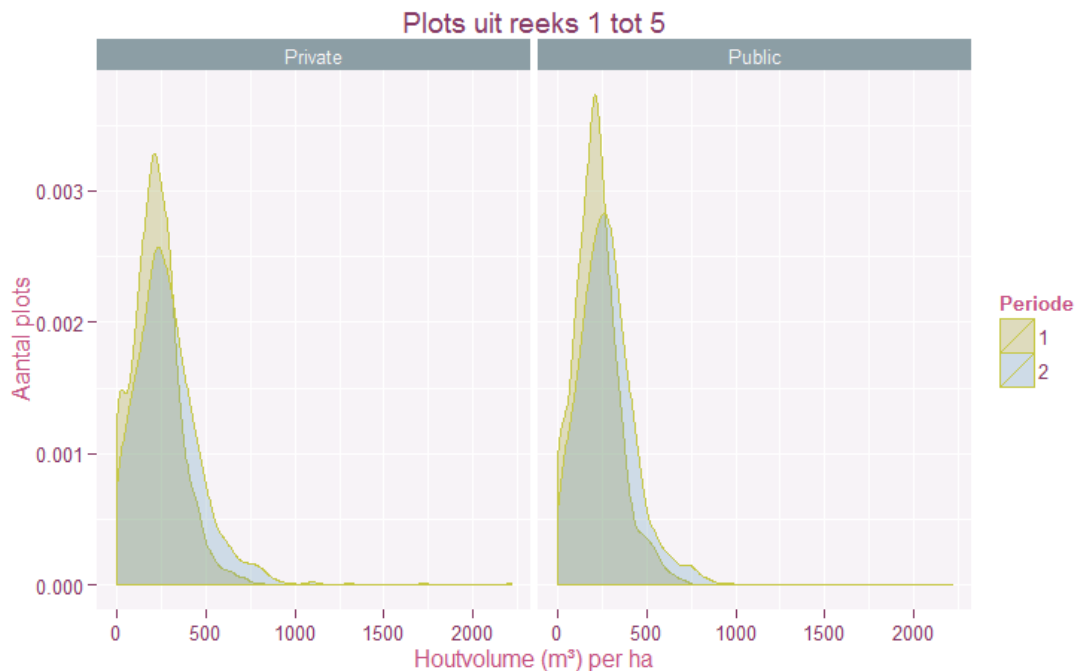
De voornaamste zaken die je uit de grafieken kan afleiden zijn:

- de breedte van (variabiliteit) een distributie is
- de scheefheid van beide distributies

- een verschuiving van de meest voorkomende waarde in de distributie

Een groot verschil met de histogrammen hierboven is dat geen rekening gehouden wordt met het plotgewicht. Doordat immers de totaaloppervlakte niet 100% bos is per plot, zouden de gewichten niet sommeren tot 1 en zou de densiteit niet volledig correct ingeschat kunnen worden.

Je zou hier dezelfde plots kunnen maken zoals bij de histogrammen, de conclusies zullen vrij gelijkaardig zijn. Hieronder staat een voorbeeld van hoe je een histogram omzet naar een density plot.



Relaties met covariaten

Boxplots

Boxplots tonen iets minder informatie over de distributie dan de histogrammen, maar kunnen bij niet al te scheve distributies wel een stuk helderder zijn om verschillen tussen periodes of eigenaarscategorieën te identificeren.

Karakteristiek aan een boxplot:

- De boxen zelf bevatten de middelste 50% van de waarden van een variabele
- De lijn door de box komt overeen met de mediaan
- De snorharen (whiskers) komen overeen met de laatste datapunten die minder ver dan de lengte van anderhalve box verwijderd liggen van de mediaan.

Wat kan je aflezen in een boxplot:

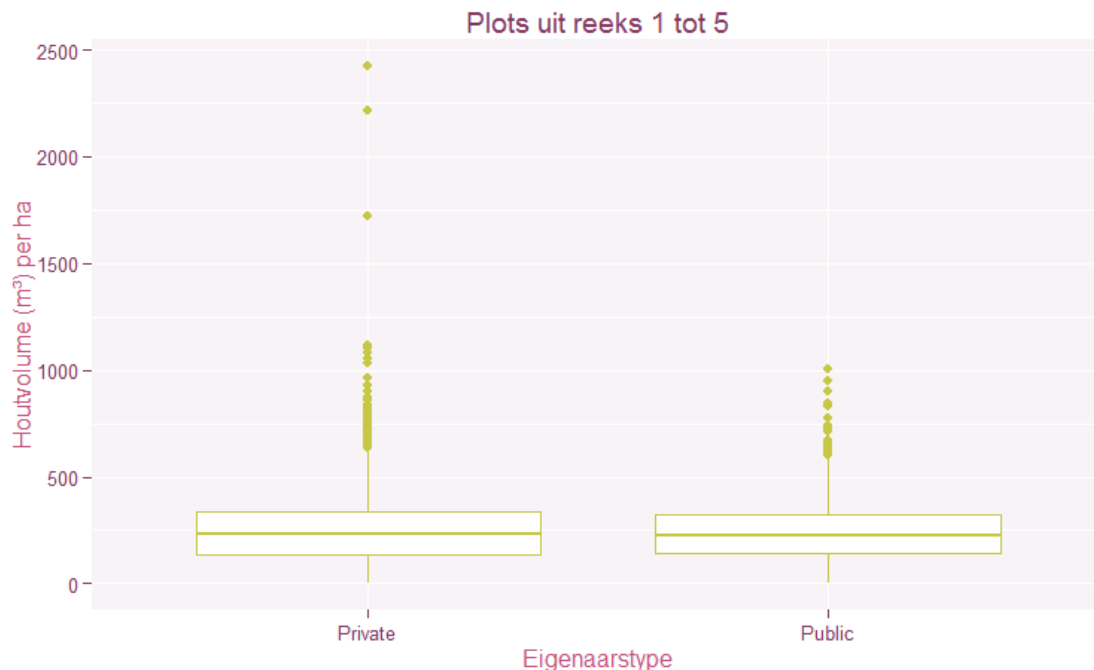
- Verschillen de medianen?
- Zit er meer variatie in sommige variabelen?
- Zijn er grote uitbijters?
- Is de distributie scheef?

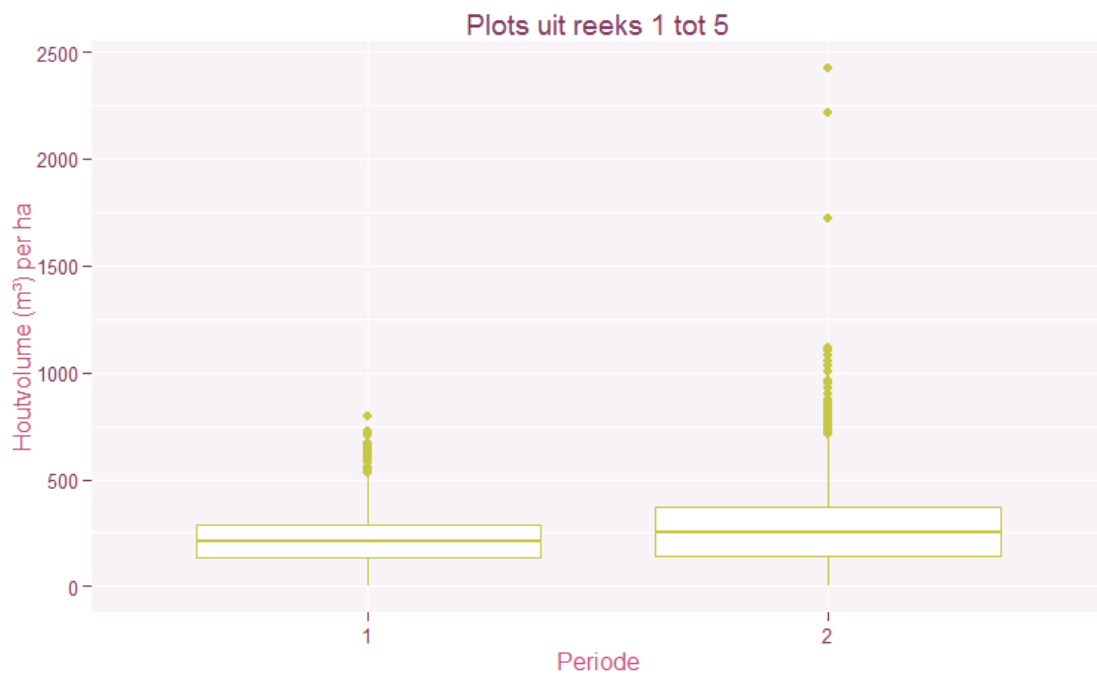
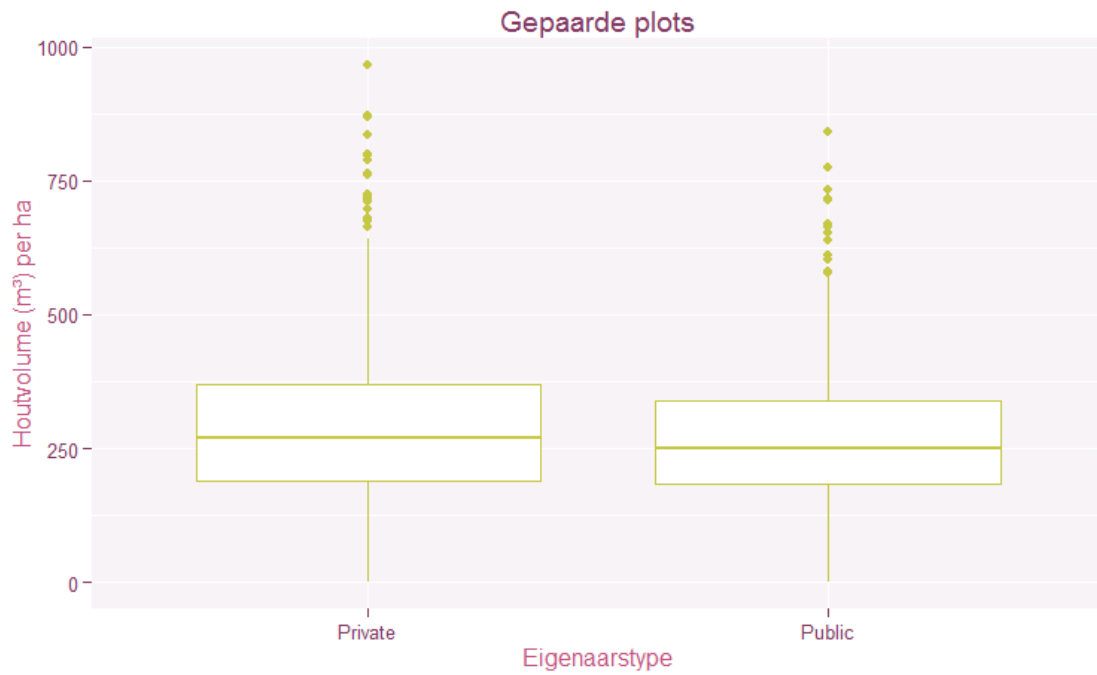
- Hoe waarschijnlijk is het dat de variabelen significant verschillen (hoe minder overlap tussen de boxen, hoe meer kans, afhankelijk van de aantallen per variabele)?
- Zijn de variabelen op een gelijkaardige manier verdeeld? Hebben de plots een gelijkaardige vorm, of is de box zeer smal, of ligt de mediaan ver van het midden van de box?

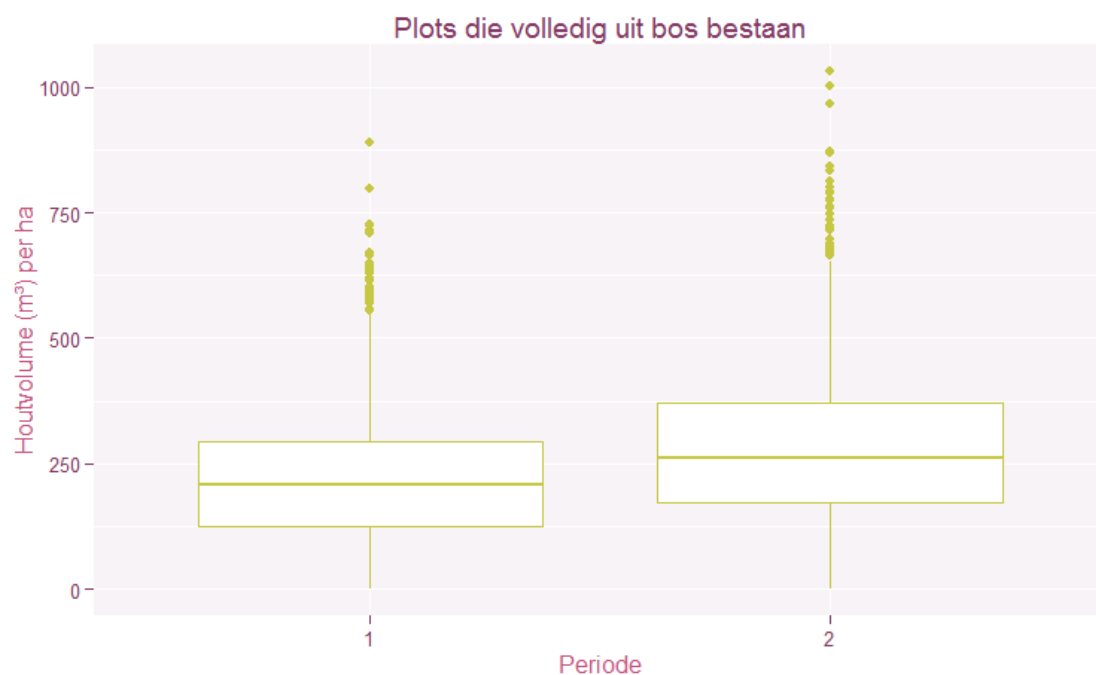
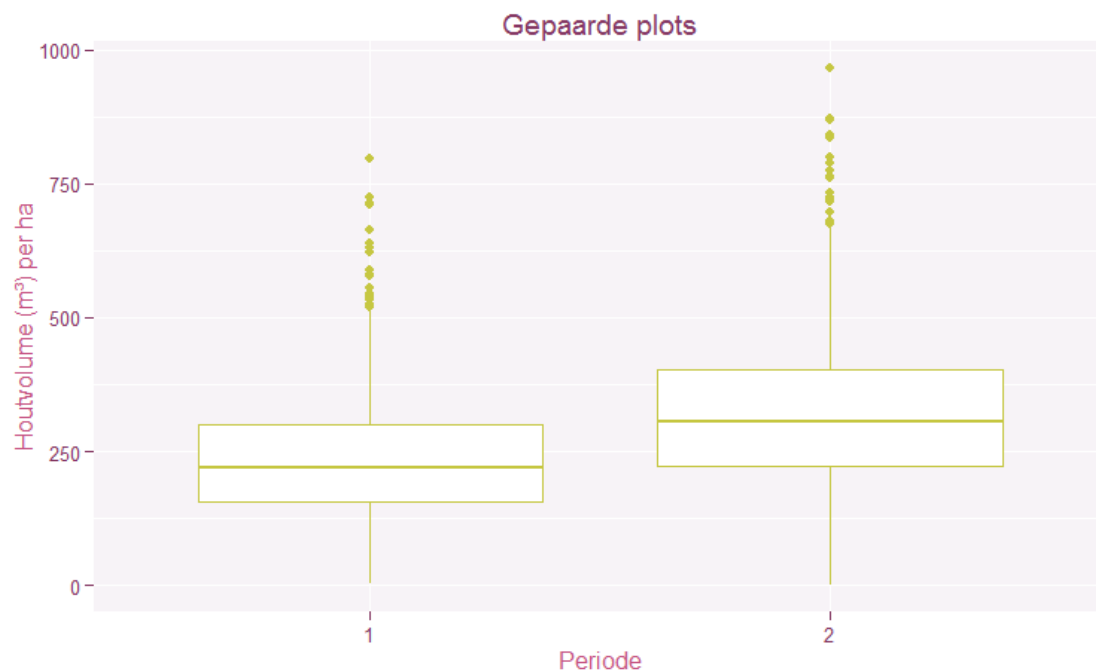
Interpretatie onderstaande figuren:

1. Op een paar uitbijters na lijken de houtvolumes vrij gelijk per eigenaarstype en op dezelfde manier verdeeld
2. Bij de gepaarde plots blijkt ook de distributie vrij gelijk per eigenaarstype, maar lijkt dit volume toch een klein beetje lager te liggen, maar vermoedelijk niet significant
3. Als de periodes vergeleken worden blijkt de 2e periode iets breder verdeeld te zijn en een iets hoger volume te hebben
4. Bij de gepaarde plots blijkt dit effect van periode nog een stuk duidelijker, en hoogst waarschijnlijk ook significant verschillend.
5. Ook bij plots die volledig uit bos bestaan blijkt het effect van periode vrij uitgesproken en is er meer hout in de tweede periode.

Boxplots op basis van segmenten



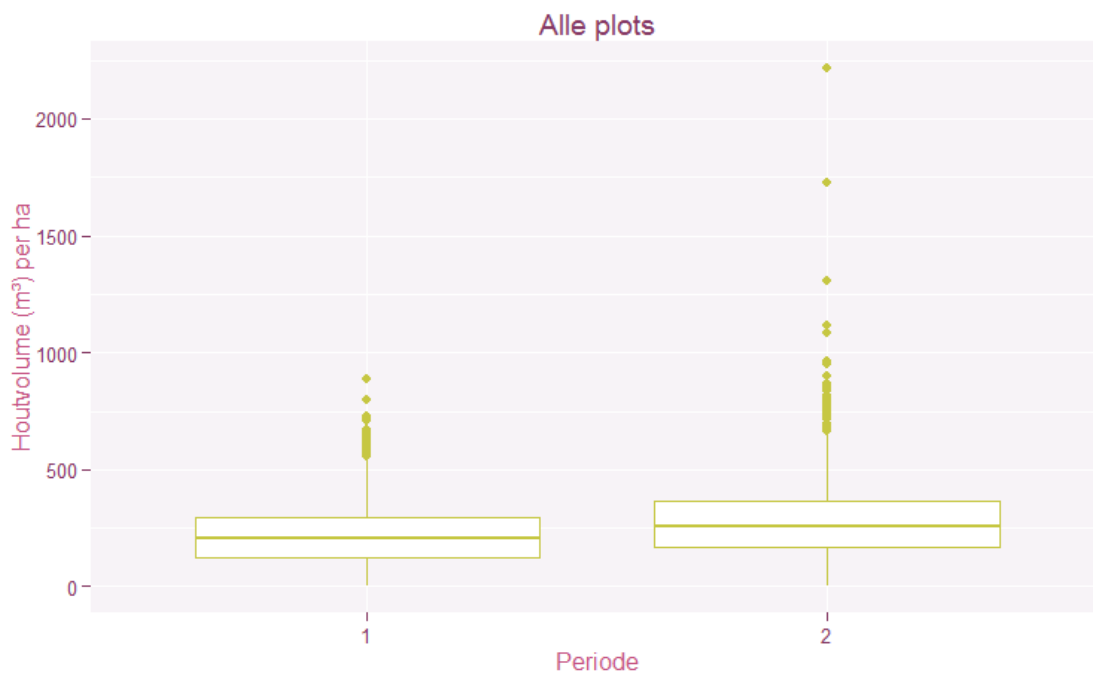
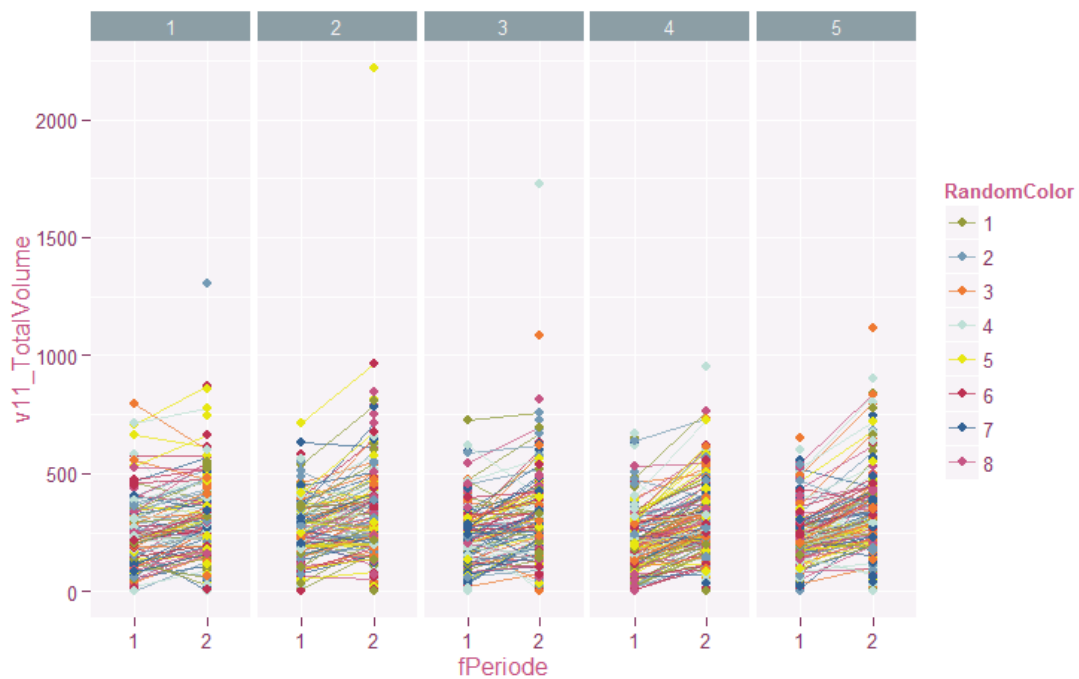




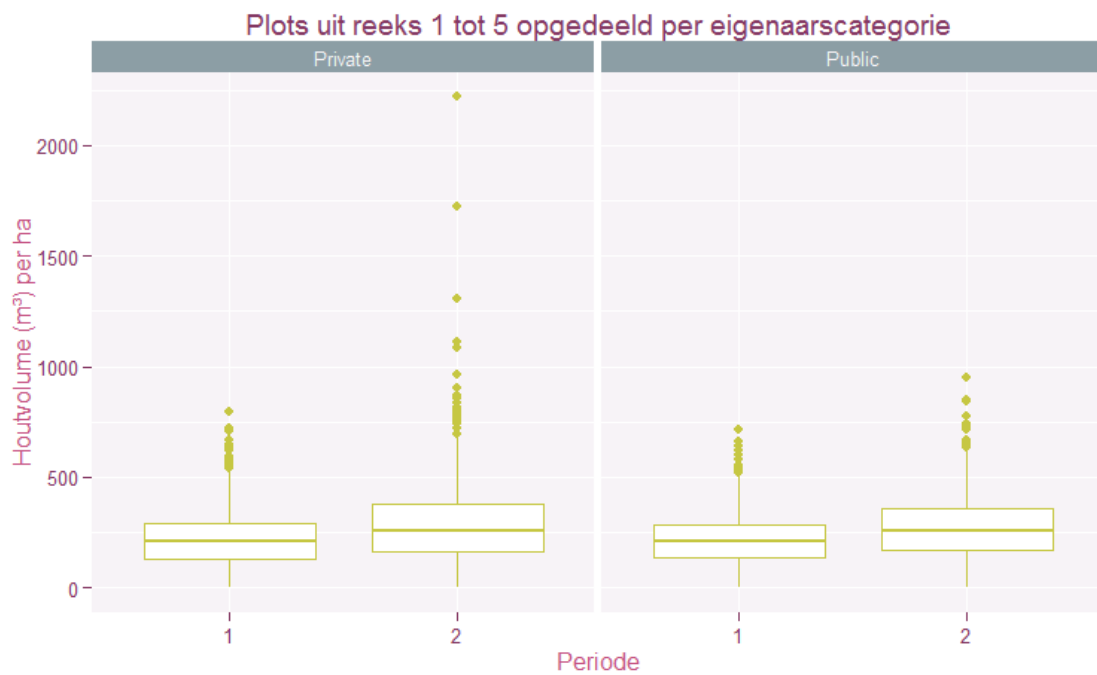
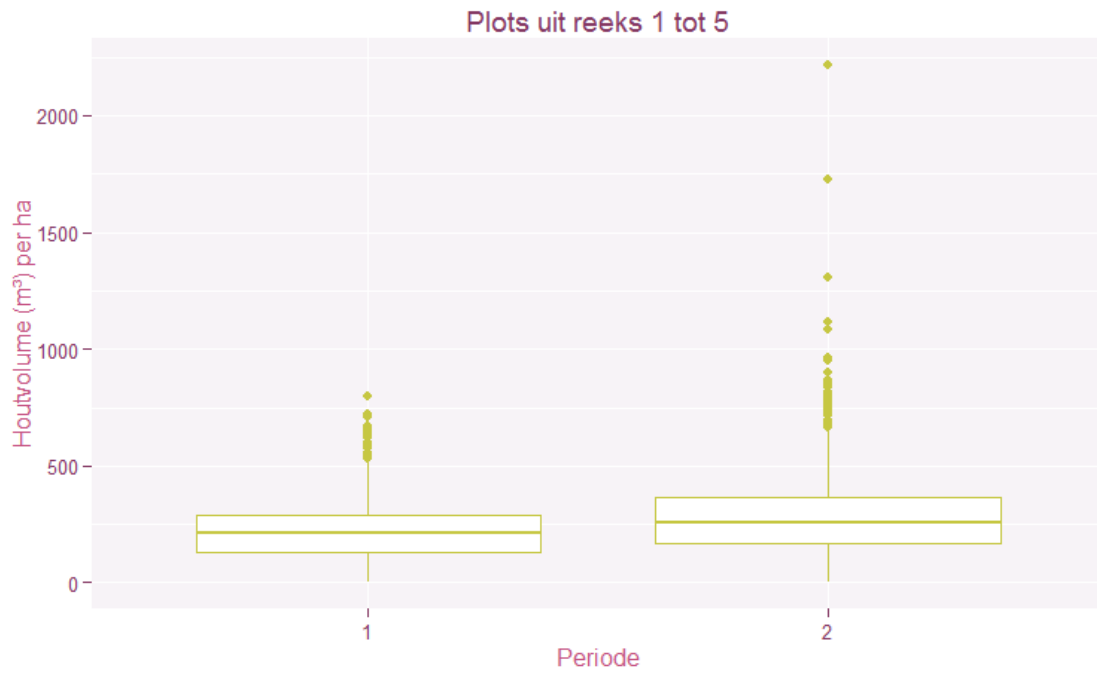
Grafieken op basis van plots (steekproefcirkels)

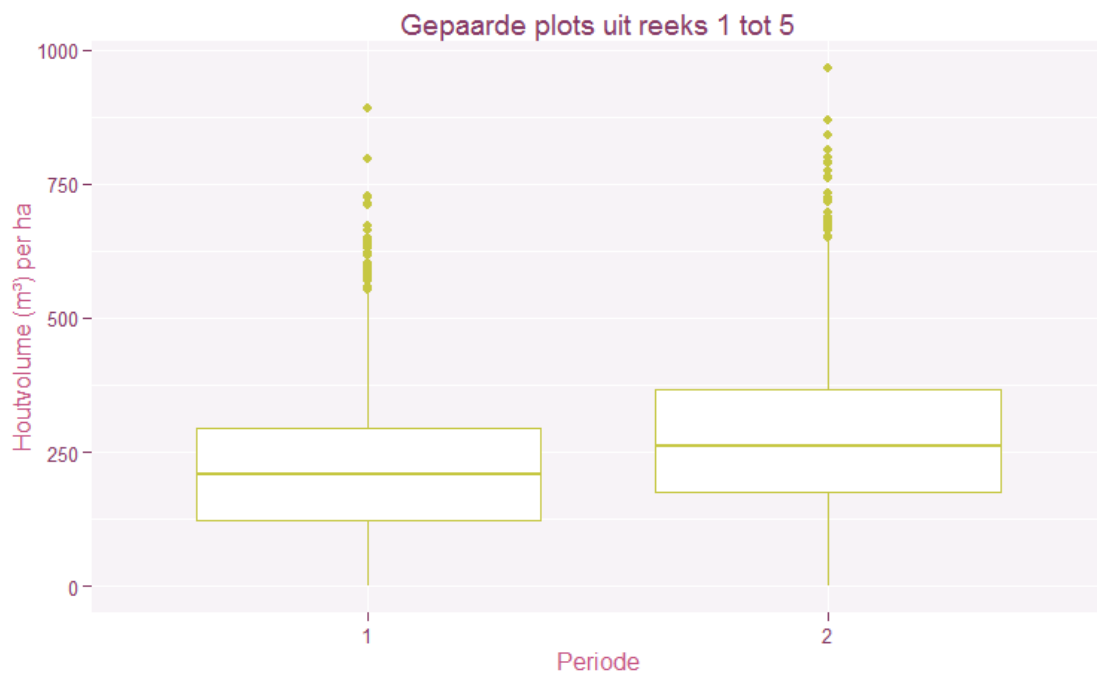
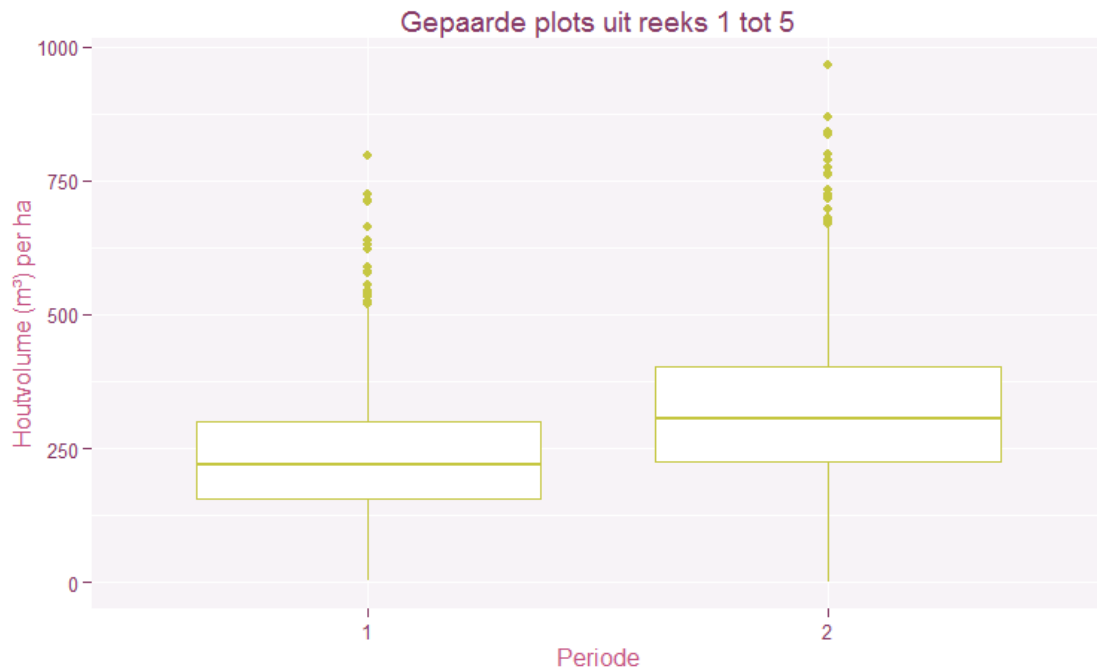
Een gelijkaardige analyse kan gebeuren op plotniveau in plaats van segmentniveau.

Onderstaande plot toont hoe het houtvolume per plot evolueert over de beide periodes. Het merendeel stijgt in houtvolume, maar er zijn er toch ook heel wat die dalen in volume of gelijk blijven. De kleuren hebben geen betekenis en dienen enkel om het spaghetti-patroon duidelijker te tonen. Dus de globale stijging doet zich niet voor op ieder plot. Er staan ook losse punten op, dit zijn niet-gepaarde plots.

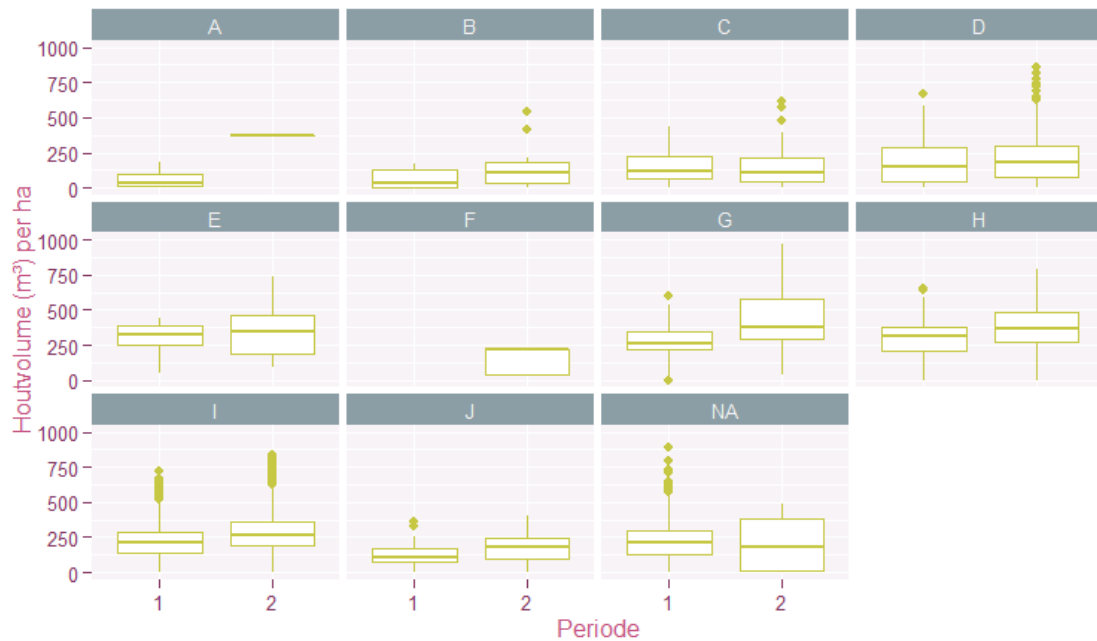


De boxplots op plotniveau tonen in grote mate dezelfde evoluties als op segmentniveau, zowel voor alle plots of enkel de gepaarde plots.





Voor de 10 bostypegroepen staat hier de evolutie over de periode (voor de helft van de plots in VBI1 werd het bostype niet bepaald). Algemeen lijkt ieder bostype ongeveer gelijk te blijven of te stijgen, zo lijkt er wel een duidelijke stijging in B, G, H, I en J. Bostype F kwam niet voor in periode 1.



Evolutie houtvoorraad als regressie

Onderstaande plot geeft de evolutie weer van de houtvoorraad voor de gepaarde plots. Door de puntenwolk wordt een smoother getrokken. De smoother bevindt zich duidelijk boven de de bissectrice die de lijn weergeeft waarbij het volume van periode 1 gelijk is aan dat van periode 2. De smoother lijkt wat naar boven verschoven te zijn t.o.v. de bissectrice, maar toch vrij lineair. De schatting in periode 2 is dus doorgaans wat hoger dan in periode 1 zoals eerder ook al opviel.



Analyse

Design-based schattingen

Design-based schattingen zijn schattingen waarbij je aanneemt dat je steekproef een goede weerspiegeling is van de werkelijkheid. Omdat je hiermee uitspraken doet over het totaal over een hele regio en geen uitspraak doet per plot, hoeft er geen rekening gehouden worden met dat plots al dan niet gepaard zijn. Je kan de totale bostoestand in beide periodes inschatten en gemakkelijk het verschil hiertussen berekenen. Een modelgebaseerde benadering daarentegen houdt zowel rekening met hoe de gepaarde plots zijn geëvolueerd als hoe het houtvolume tussen beide periodes is toegenomen.

Aan elke plot wordt een gewicht toegekend gelijk aan het aandeel bos binnen de plot. We gebruiken de functie `My.WgtParEstimation` om het gewogen gemiddelde, de gewogen variantie en het 95% betrouwbaarheidsinterval te berekenen.

	1	2
variabele	v11_TotalVolume	v11_TotalVolume
strata		
stratumNaam		
periode	1	2
minYear	1997	2009
maxYear	1999	2014
minReeks	1	1
maxReeks	10	5
nbObservaties	2619.000	1307.345
wgt.mean	216.7285	279.3598
wgt.var	17715.76	30197.40
llci	211.6309	269.9399
ulci	221.8262	288.7797

Voor alle plots in de dataset

	1	2
variabele	v11_TotalVolume	v11_TotalVolume
strata		
stratumNaam		
periode	1	2
minYear	1997	2009
maxYear	1999	2014
minReeks	1	1
maxReeks	5	5
nbObservaties	1339.000	1307.345
wgt.mean	219.3735	279.3598

wgt.var	17133.85	30197.40
llci	212.3623	269.9399
ulci	226.3847	288.7797

Enkel voor de reeksen 1 tot 5 voor alle plots in de dataset

	1	2
variabele	v11_TotalVolume	v11_TotalVolume
strata		
stratumNaam		
periode	1	2
minYear	1997	2012
maxYear	1999	2014
minReeks	10	5
maxReeks	10	5
nbObservaties	256.000	316.733
wgt.mean	233.7121	294.2469
wgt.var	16336.57	27084.01
llci	218.0548	276.1225
ulci	249.3694	312.3714

Enkel voor plots met volume > 0

	1	2
variabele	v11_TotalVolume	v11_TotalVolume
strata		
stratumNaam		
periode	1	2
minYear	1997	2013
maxYear	1999	2014
minReeks	5	5
maxReeks	5	5
nbObservaties	106.000	105.435
wgt.mean	234.1915	343.3200
wgt.var	11708.15	21564.32
llci	213.5925	315.2895
ulci	254.7906	371.3506

Enkel voor gepaarde plots

	1	2
variabele	v11_TotalVolume	v11_TotalVolume
strata		
stratumNaam		
periode	1	2
minYear	1997	2012
maxYear	1999	2014
minReeks	10	5
maxReeks	10	5
nbObservaties	268	245
wgt.mean	223.2474	289.2837
wgt.var	17947.32	24921.61
llci	207.2080	269.5158
ulci	239.2868	309.0516

Enkel niet-rand plots

Model-based schattingen

De modelgebaseerde schatting houdt rekening met de gepaardheid van de plots. Hiervoor wordt een mixed model gebruikt, waarbij de gepaardheid in het model wordt opgenomen als een random intercept model. Hierbij wordt rekening gehouden met de correlatie tussen beide periodes binnen eenzelfde plot, immers als de houtvoorraad in de eerste periode hoog is, zal dit in de tweede periode ook meestal zo zijn, tenzij er iets ingrijpend is gebeurd.

De gepaarde metingen geven dan een beter beeld hoe binnen een plot de houtvoorraad is toegenomen tussen beide periodes, en gebruikt de variatie op deze metingen om een inschatting te geven voor toename in houtvoorraad in de niet-gepaarde plots.

Model opbouwen in R

Dit model schatten we met een functie in de lme4 bibliotheek in het programma R. We gebruiken de functie lmer om dit model te schatten. Het random effect die de gepaardheid weergeeft wordt gecodeerd als (1 | IDGroup), waarbij IDGroup de identificatie van het plot bevat. Het fixed (interesse) effect is de relatie tussen de respons v1_TotaalVolume t.o.v de periode (fPeriode). Verder wordt er ook een variabele doorgegeven die de plotweging behelst via het argument weight = Weight, waarbij Weight de kolom in het model is die het plotgewicht meetelt.

Model interpreteren (alle data)

De summary functie toont de meest essentiële informatie in het model. De belangrijkste output hierin zijn het deeltje random effects en het deeltje fixed effects.

Fixed effects: Hierin zie je 2 schattingen staan, namelijk (Intercept) en fPeriode2. R gebruikt de eerste periode als referentie, dus de houtvoorraad in de eerste periode komt overeen met de *Estimate* voor (Intercept).

In onderstaand geval die alle plots bevat is dit 209.23. Daarnaast staat een standaard error. Als je deze vermenigvuldigd met 1.96 en optelt en aftrekt van de estimate heb je het 95% betrouwbaarheidsinterval op de schatting.

Op de lijn eronder staat fPeriode2 en die geeft de schatting hoeveel de tweede periode afwijkt van de referentieperiode (dus de eerste periode). In het eerste voorbeeld is dit 73.03, dus dit wil zeggen dat de schatting voor de tweede periode gelijk is aan $209.23 + 73.03 = 282.26$.

Ook nu kan je het betrouwbaarheidsinterval op dit verschil schatten door $1.96 * \text{de standaard error van fPeriode2}$ op te tellen en af te trekken van de Estimate van 73.03, dus het verschil tussen beide periodes zal liggen tussen 65.96 en 80.10. De significantie kan je afleiden uit de t-waarde waarbij je als vuistregel kan nemen dat als de t-waarde groter is dan 2 (of kleiner dan -2) dat het verschil tussen beide periodes significant is.

Random effects: Hierin zie je ook 2 lijnen verschijnen, namelijk (IDGroup en Residual). Vooral de kolom Std.Dev is hier belangrijk. De stdev van IDGroups (hier 129.92) toont in welke mate de houtvoorraad tussen de plots verschillen van elkaar.

Met een waarde van 129.92 verwacht je dat 95% van de plots liggen tussen -1.96 en + 1.96 keer deze waarde met retrospect tot het fixed effect (Intercept), dus verwacht je dat de houtvoorraad van de plots in de eerste periode tussen de -45 en 464 zullen liggen ($209.23 - 1.96 * 129.92$ en $209.23 + 1.96 * 129.92$).

De variatie in houtvoorraad binnen een plot over de periodes wordt ingeschat op de waarde die hoort bij de Residual in het random effect, dus binnen de plots los van de periode hebben we een standaardafwijking van 74.15.

Grafiek: De onderstaande grafiek is een korte samenvatting van hoe de model fit. De Pearson residu's van het model worden uitgezet ten opzichte van de gefitte waarden. Als we een goed model hebben verwachten we hier geen patronen. Ongewenste patronen kunnen zijn: - De spreiding op de residu's stijgt naarmate de fit hoger wordt - Er is een systematische afwijking te zien tussen residu's en gefitte waarden, zoals bv. een lineaire of kwadratische of schommelende trend. - Grote outliers: waarden die zeer sterk afwijken van de rest van de observaties. Als dit er niet teveel zijn kan dit vaak geen kwaad, maar die worden best wel eens onderzocht.

In dit model zijn er 3 zeer grote outliers aanwezig voor hoge gefitte waarden. Deze mogen enkel uit het model verwijderd worden als daar gegronde redenen voor zijn.

De rest van de residu's lijken geen groot patroon meer te vertonen (hierin mag je je niet laten misleiden door punten die op een lijn lijken te liggen, dat is vaak een artefact doordat er maar 2 periode's zijn en niet alle plots gepaard zijn (daardoor krijg je alsof 2 stijgende lijnen te zien)).

Onderstaande code illustreert het hierboven beschreven model en toont de modelresultaten voor de dataset waarbij we alle plots meenemen in de analyse:

```
Model.lmer_all <- lmer(v11_TotalVolume ~ fPeriode + (1|IDGroup), analyseSetPerPlot, weight=Weight)
summary(Model.lmer_all)
```

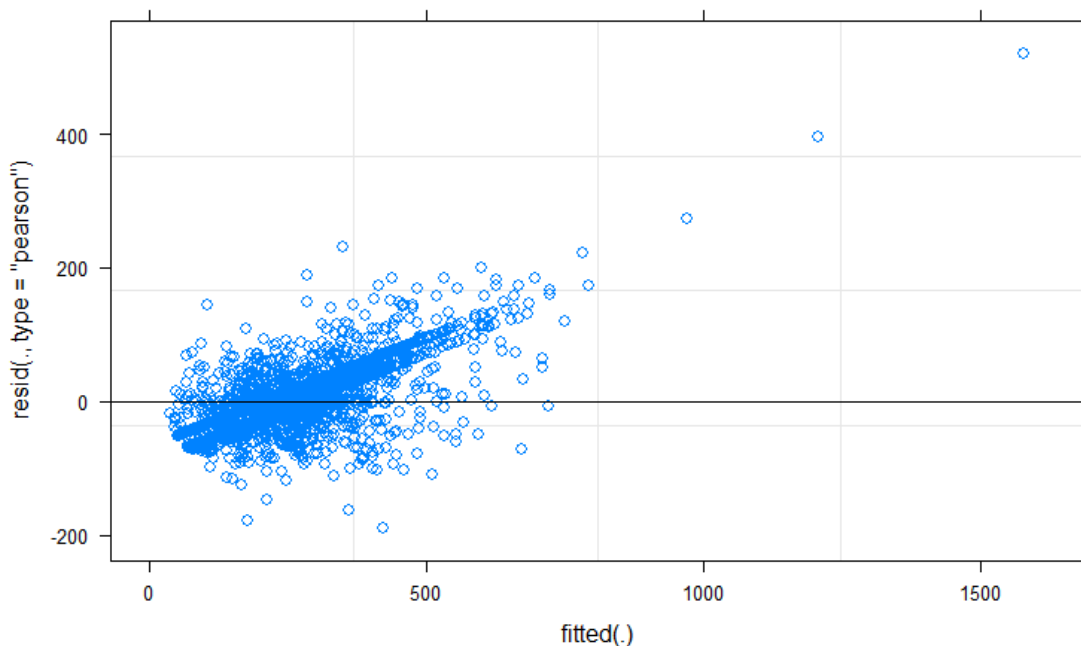
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: v11_TotalVolume ~ fPeriode + (1 | IDGroup)
## Data: analyseSetPerPlot
## Weights: Weight
##
## REML criterion at convergence: 51172.3
##
## Scaled residuals:
```

```

##   Min   1Q Median   3Q   Max
## -2.5484 -0.3756 -0.0414  0.2968  7.0100
##
## Random effects:
## Groups Name      Variance Std.Dev.
## IDGroup (Intercept) 16879  129.92
## Residual          5499   74.15
## Number of obs: 4015, groups: IDGroup, 3465
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept) 209.230    2.777  75.34
## fPeriode2   73.029    3.605  20.26
##
## Correlation of Fixed Effects:
##           (Intr)
## fPeriode2 -0.417

```

plot(Model.lmer_all)



Model interpreteren (enkel reeksen 1 t.e.m 5)

Omdat het gebruik van alle plots in periode 1 de hele oppervlakte dekt, en in periode 2 enkel de tot dan toe doorlopen stroken, herhalen we dit model voor enkel de plots van reeks 1 tot 5. Voor de rest blijft het model identiek.

De resultaten zijn niet zeer verschillend van de resultaten met alle plots. De variatie tussen de plots in periode 1 is wat gestegen naar 139 ((IDGroups) in random effects), de variatie binnen de plots is vrij gelijk gebleven.

In de fixed effects kan je zien dat de geschatte waarde in periode 1 wat gedaald is naar 205, en dat het verschil tussen de periodes iets hoger wordt ingeschat (76), maar deze afwijkingen vallen ruim binnen het 95% betrouwbaarheidsinterval van bovenstaand model, dus in essentie geven de beide datasets zeer gelijkaardige resultaten en is er een duidelijke zeer significante stijging in houtvoorraad in de tweede periode.

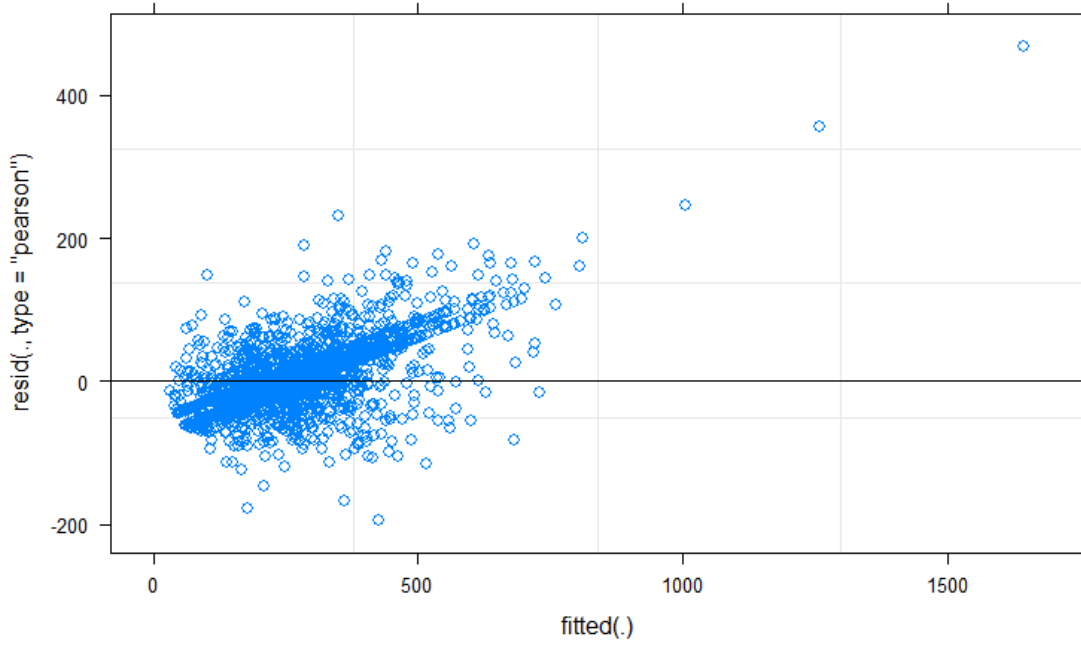
De plot tussen de residu's en de gefitte waarden is bijna identiek aan dat van het model met de gehele dataset. Daarnaast hebben we ook nog een qqplot toegevoegd die toont hoe dicht de residu's de normale distributie (de lijn) benaderen. In de staarten kan je altijd wel wat afwijking verwachten (in de ene staart naar beneden, in de andere naar boven) zonder problemen, zolang de afwijkingen niet te hoog zijn. Als er echter een duidelijk patroon te zien is (zoals een afvlakkende curve, waar de staarten beiden boven of onder de lijn liggen), dan wijken we redelijk sterk af van de normaliteit.

Iets afwijken van de normaliteit kan geen kwaad, zeker als de significantie zeer duidelijk is, maar hoe minder normaal de residu's verdeeld zijn, hoe minder betrouwbaar de t-waarden en dus de significanties zijn in de summary-output van het model voor het periode-effect.

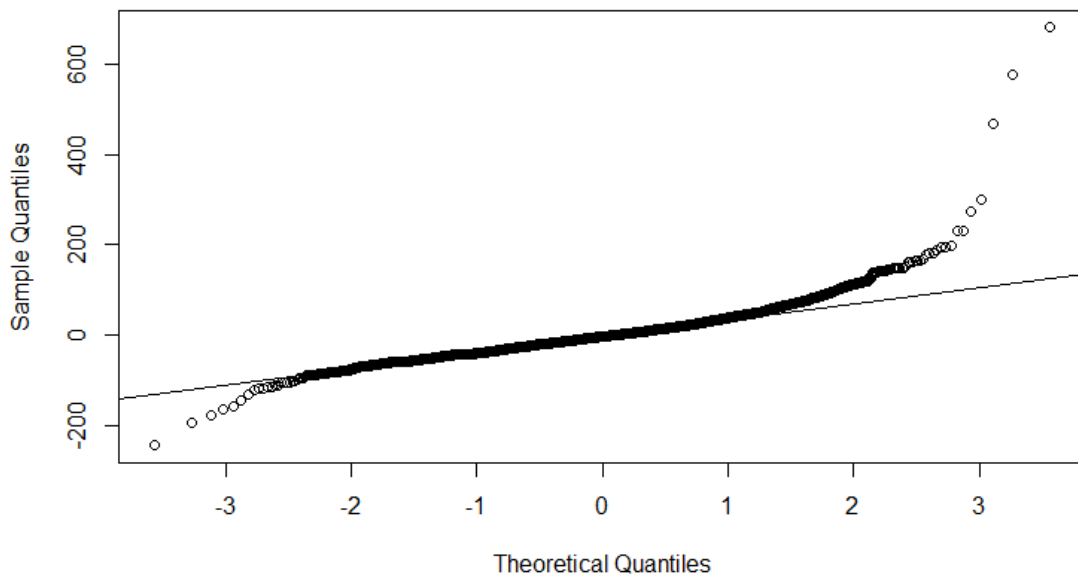
De boxplots geven de verdeling van de residuen weer per periode. De variantie per periode moet ongeveer gelijk zijn. Dit is hier het geval.

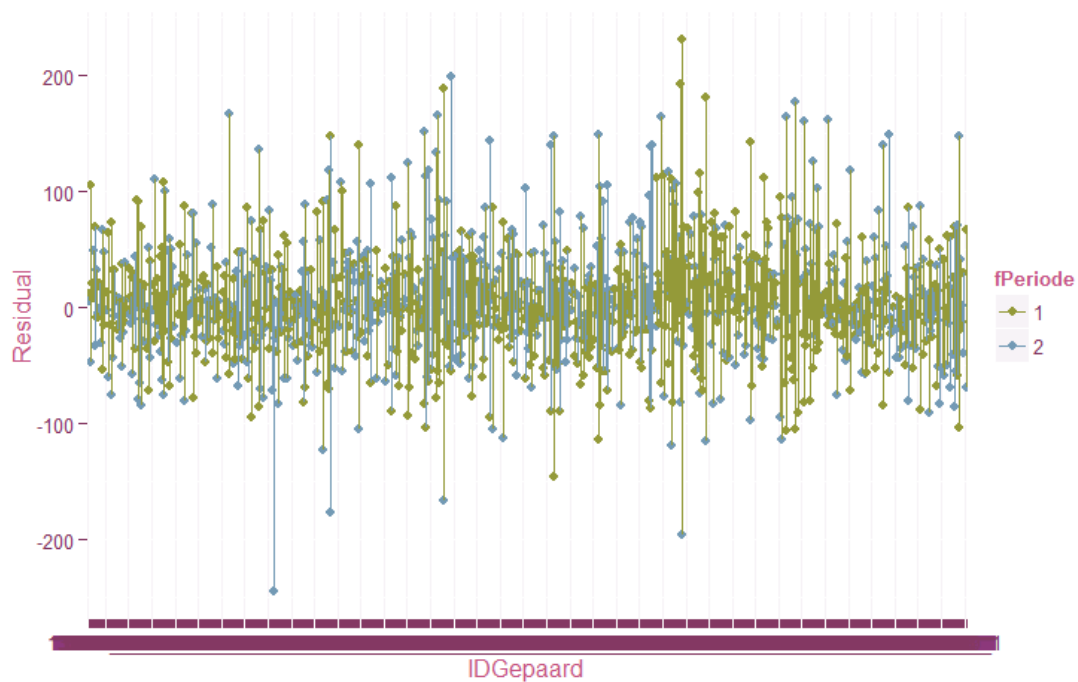
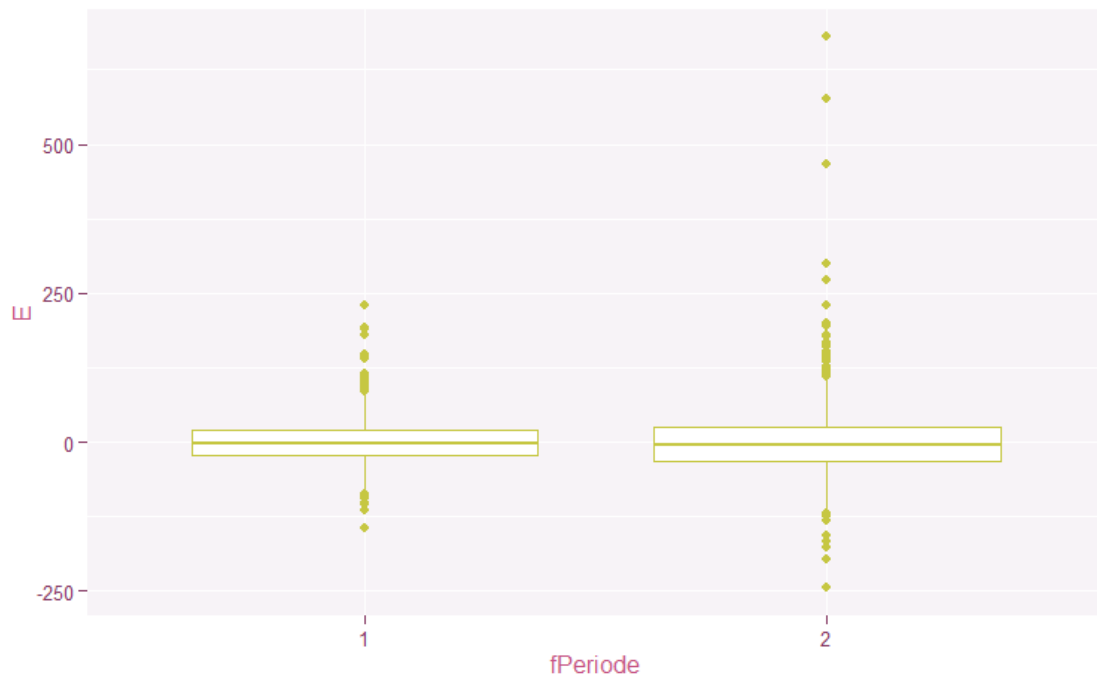
In de laatste plot wordt er getoond wat het random effect eigenlijk doet met de data. Hierin zie je immers duidelijk hoe dat de residu's van de gepaarde plots in beide periode's rondom 0 gecentreerd worden, waarbij de ene periode telkens onder 0 en de andere periode telkens boven nul ligt in een specifiek plot, waardoor dus eigenlijk de ruis binnen een plot over de periode's gecentreerd wordt wat ten gunste komt van de betrouwbaarheid op de schatting van het periode-effect.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: v11_TotalVolume ~ fPeriode + (1 | IDGroup)
## Data: analyseSetPerPlotReeks5
## Weights: Weight
##
## REML criterion at convergence: 34937.7
##
## Scaled residuals:
##  Min   1Q Median   3Q   Max
## -2.6279 -0.3697 -0.0473  0.2837  6.3546
##
## Random effects:
## Groups Name      Variance Std.Dev.
## IDGroup (Intercept) 19395   139.27
## Residual           5396    73.46
## Number of obs: 2735, groups: IDGroup, 2185
##
## Fixed effects:
##      Estimate Std. Error t value
## (Intercept) 204.864    3.869  52.95
## fPeriode2   76.037    3.867  19.67
##
## Correlation of Fixed Effects:
##      (Intr)
## fPeriode2 -0.506
```



Normal Q-Q Plot



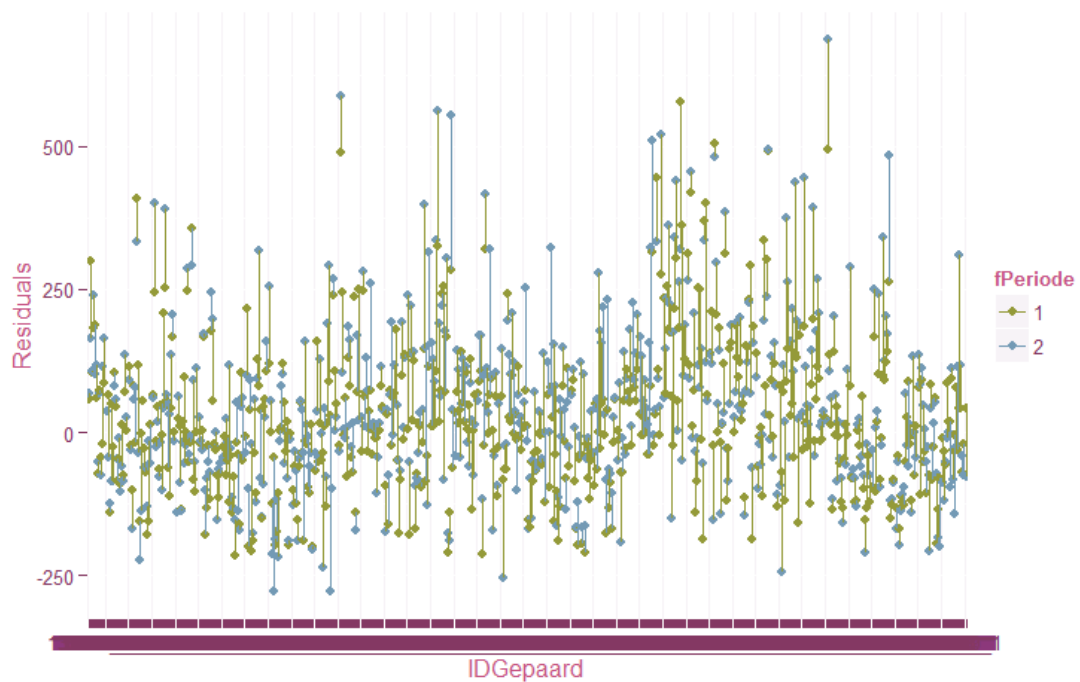
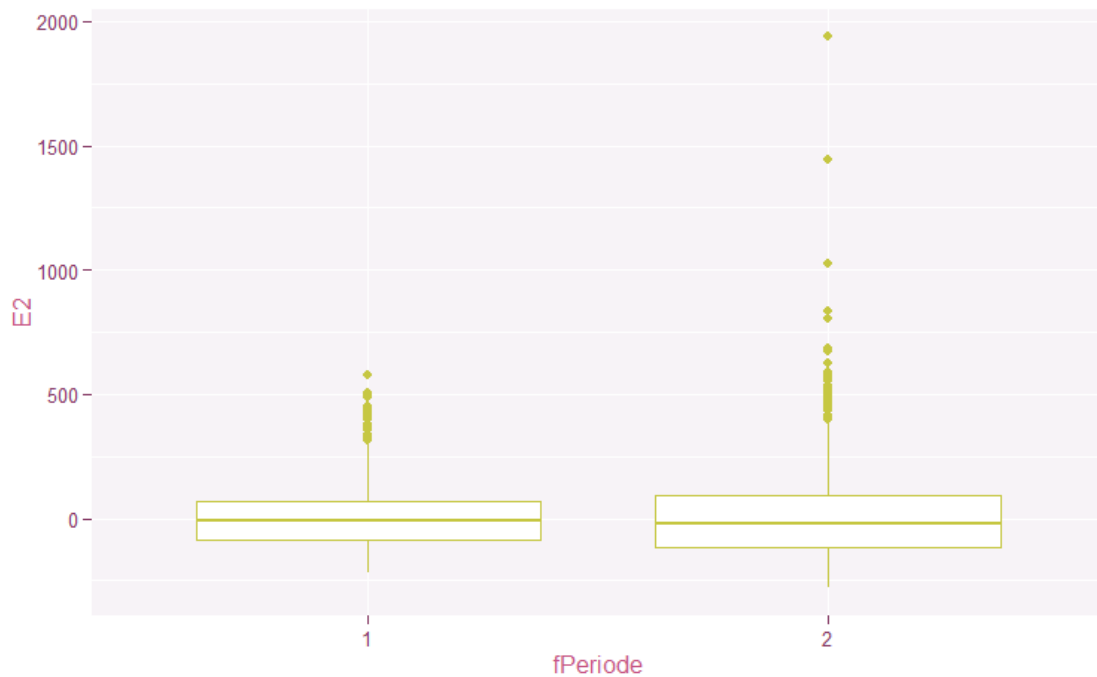


Tot slot bekijken we nog eens wat het effect zou zijn moesten we gewoon een naïef standaard lineair model gebruiken in plaats van een mixed effect model.

De output is nu een stuk eenvoudiger, aangezien er enkel nog fixed effects aanwezig zijn. De interpretatie van deze is echter dezelfde als die van het random effect. In tegenstelling tot het random effect model wordt er geen rekening gehouden met de gepaardheid van de plots en heeft dit in dit geval het gevolg dat de schatting voor periode 1 een stuk hoger ligt en het verschil tussen Periode 1 en 2 een stuk lager wordt ingeschat, meer in lijn met de design based schatter.

In de boxplots is er weinig verschil met voorgaande, maar het grote verschil is te merken in de residu's van de gepaarde plots tussen de 2 periode's. Hier worden deze verschillen niet mooi gecentreerd rond 0, waardoor de houtvoorraad in sommige plots in beide periode's sterk overschat (beide residu's van een plot onder 0) worden en andere plots sterk onderschat (beide residu's van een plot boven 0). Er zit dus nog een duidelijk ploteffect in de schattingen dat niet in rekening is gebracht met dit model.

```
##  
## Call:  
## lm(formula = v11_TotalVolume ~ fPeriode, data = analyseSetPerPlotReeks5,  
## weights = Weight)  
##  
## Weighted Residuals:  
## Min 1Q Median 3Q Max  
## -279.36 -97.54 -12.90 77.64 1573.71  
##  
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 219.373 4.129 53.14 <2e-16 ***  
## fPeriode2 59.986 5.874 10.21 <2e-16 ***  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 151.1 on 2733 degrees of freedom  
## Multiple R-squared: 0.03676, Adjusted R-squared: 0.03641  
## F-statistic: 104.3 on 1 and 2733 DF, p-value: < 2.2e-16
```



Dus de design-based schatter (of het gewone lineair model) geven misschien iets exacter het verschil in de totale houtvoorraad, terwijl de model-based schatter eerder een juistere inschatting geeft van de toename in houtvoorraad gemiddeld genomen in een plot.