



Boomkronen afleiden uit het Actueel Hoogtebestand Nederland

Kwaliteitsaspecten rondom het geautomatiseerd in kaart brengen van bomen op basis van het AHN2-bestand

Marcel Meijer, Frans Rip, Rik van Benthem, Jan Clement en Corné van der Sande



ALTERRA
WAGENINGENUR

Boomkronen afleiden uit het Actueel Hoogtebestand Nederland

Kwaliteitsaspecten rondom het geautomatiseerd in kaart brengen van bomen op basis van het AHN2-bestand

Marcel Meijer¹, Frans Rip², Rik van Benthem³, Jan Clement¹ en Corné van der Sande³

1 Alterra Wageningen UR

2 Wageningen University

3 Neo B.V.

Dit onderzoek is uitgevoerd door Alterra Wageningen UR in opdracht van en gefinancierd door het ministerie van Economische Zaken, in het kader van het Kennisbasis onderzoekthema 'Duurzame ontwikkeling van de groenblauwe ruimte' (projectnummer KB-14-002-037).

Alterra Wageningen UR

Wageningen, november 2015

Alterra-rapport 2671

ISSN 1566-7197

Meijer, M., F. Rip, R. van Benthem, J. Clement en C. Van der Sande, 2015. *Boomkronen afleiden uit het Actueel Hoogtebestand Nederland; Kwaliteitsaspecten rondom het geautomatiseerd in kaart brengen van bomen op basis van het AHN2-bestand*. Wageningen, Alterra Wageningen UR (University & Research centre), Alterra-rapport 2671. 86 blz.; 48 fig.; 31 tab.; 35 ref.

Alom wordt erkend dat bomen belangrijk zijn. Zowel voor de mens, de natuur als het klimaat. Recentelijk is een procedure ontwikkeld om op basis van het nationale Nederlandse hoogtebestand AHN2 een bestand te genereren met alle boomkronen in Nederland, genaamd 'CP'. Een dergelijk bestand kan onder andere het groenbeheer van de gemeenten in Nederland vereenvoudigen en helpen bij het inventariseren van landschapselementen. De vraag is echter: hoe goed is dit bestand?

In dit rapport wordt voor een drietal verschillende gebieden onderzocht wat de kwaliteit is van CP. Verder wordt mede op basis van de ervaringen die tijdens het kwaliteitsonderzoek zijn gedaan een standaard kwaliteitsraamwerk opgezet voor het controleren van nieuwe versies van het boomkronenbestand. Daarnaast is dit document er ook op gericht om de potentiële gebruiker een beter beeld van de kwaliteit te geven.

Everyone recognises that trees are important. For us as humans, but also for nature and the climate. Recently a procedure was developed to extract trees from the Dutch national elevation grid AHN2. The data generated by this procedure can be used by municipalities to simplify the upkeep of their tree data but also as a means to inventory landscape features. The question of course is how accurate and complete this new dataset is? In this report we will try to answer this question for three different landscapes. Based on our experiences we will also present a quality framework to test new versions of the tree dataset. But also give the potential user a better understanding of quality issues in relation to the tree registry.

Trefwoorden: boomkroon, boomkronen, coöperatie, validatie, extractie, methode, dataset, kwaliteit, kwaliteitsaspecten, kwaliteitsraamwerk, boomstam, boomregister.nl, volledigheid, compleetheid, kwaliteitskenmerken, LiDAR, AHN2, objectdetectie, houtopstand, GLB, landschapselementen, park, bomen, stamhoogte, boomkroonoppervlakte, PPS, boombeheer, remote sensing, boomregister

Dit rapport is gratis te downloaden van www.wageningenUR.nl/alterra (ga naar 'Alterra-rapporten' in de grijze balk onderaan). Alterra Wageningen UR verstrekt *geen* gedrukte exemplaren van rapporten.

© 2015 Alterra (instituut binnen de rechtspersoon Stichting Dienst Landbouwkundig Onderzoek), Postbus 47, 6700 AA Wageningen, T 0317 48 07 00, E info.alterra@wur.nl, www.wageningenUR.nl/alterra. Alterra is onderdeel van Wageningen UR (University & Research centre).

- Overname, verveelvoudiging of openbaarmaking van deze uitgave is toegestaan mits met duidelijke bronvermelding.
- Overname, verveelvoudiging of openbaarmaking is niet toegestaan voor commerciële doeleinden en/of geldelijk gewin.
- Overname, verveelvoudiging of openbaarmaking is niet toegestaan voor die gedeelten van deze uitgave waarvan duidelijk is dat de auteursrechten liggen bij derden en/of zijn voorbehouden.

Alterra aanvaardt geen aansprakelijkheid voor eventuele schade voortvloeiend uit het gebruik van de resultaten van dit onderzoek of de toepassing van de adviezen.

Inhoud

	Woord vooraf	5
	Samenvatting	7
1	Inleiding	11
	1.1 Introductie	11
	1.2 Databronnen	13
	1.3 Toepassingsmogelijkheden	13
	1.4 Opzet rapport	13
	1.5 Samenwerking	14
	1.6 Eerdere publicaties	14
	1.7 Begripsbepalingen en afkortingen	15
2	Beschrijving Boomkronen Extractie Methode	16
	2.1 Het extractiealgoritme	16
	2.1.1 Extractie van de boomkroon	16
	2.1.2 Versies van CP	19
	2.2 Het SILVI-STAR-model	21
3	Beschrijving van de validatiemethoden	24
	3.1 Validatiemethoden	24
	3.1.1 Boomlocaties	24
	3.1.2 Ligging en omvang kroonprojecties	25
	3.1.3 Boomhoogte	25
	3.2 Beschrijving testlocaties	26
	3.2.1 Stedelijk gebied	26
	3.2.2 Landelijk gebied	27
	3.2.3 Parkgebied	28
	3.3 Beschrijving veldmetingen/referentiemodel	30
	3.3.1 Stedelijk gebied	30
	3.3.2 Landelijk gebied	34
	3.3.3 Parkgebied	36
4	Resultaten en discussie	43
	4.1 Resultaten kwaliteitscontrole stedelijk gebied	43
	4.1.1 Veldmeting	43
	4.1.2 Boomlocaties	43
	4.1.3 Boomkroonoppervlakte	44
	4.1.4 Hoogte	45
	4.2 Resultaten kwaliteitscontrole landelijk gebied	46
	4.2.1 Boomkroonoppervlakte	46
	4.3 Resultaten kwaliteitscontrole parkgebied	47
	4.3.1 Boomlocaties	47
	4.3.2 Boomkronen	49
	4.3.3 Hoogte	52

5	Conclusies en aanbevelingen	53
5.1	Extractiefouten	53
5.1.1	Stedelijk gebied	53
5.1.2	Landelijk gebied	56
5.1.3	Parkgebied	57
5.1.4	Verschillen tussen de versies van BEM	58
5.1.5	Experiment: verbeteren van BEM	59
5.2	Conclusies	62
5.2.1	Conclusies validatie landelijk en stedelijk gebied	62
5.2.2	Conclusies validatie parkgebied	63
5.3	Conclusie met betrekking tot de kwaliteit	63
6	Integraal kwaliteitsraamwerk	64
6.1	Inleiding	64
6.1.1	Kwaliteit van Open Data	64
6.2	Kwaliteit in het proces	65
6.2.1	Gebruikers van het boomkronenbestand	66
6.2.2	Gebruik van het boomkronenbestand	66
6.2.3	Communiceren over het boomkronenbestand	67
6.2.4	Organisatie	68
6.3	Product- en kwaliteitskenmerken	68
6.4	Toetsen van de bruikbaarheid	73
6.5	Conclusies en aanbevelingen	76
	Literatuur	77
	Bijlage 1 Metadata Boomkronenbestand	80

Woord vooraf

Het belang van bomen wordt alom erkend. Zowel voor mens, dier als ons klimaat zijn bomen onmisbaar. In een stedelijke omgeving vormen ze een levend contrast met de wereld van steen en asfalt. De aanwezigheid van groen in het straatbeeld geeft een natuurlijk en rustgevend karakter aan de stad en verhoogt zo de levenskwaliteit. Daarnaast is uit onderzoek gebleken dat bomen een positief effect hebben op het herstel van patiënten.

Maar bomen hebben niet alleen effect op ons welbevinden als mens, maar verhogen ook de leefbaarheid om ons heen. In steden hebben bomen bijvoorbeeld een grote invloed op het stadsklimaat. De bladeren van de bomen vangen veel stof op en werken daarbij als een filter. Bomen zorgen voor schaduw en koelte in de zomer. Ze dragen bij aan een aangenaam microklimaat en zorgen voor de beleving van de seizoenen. Ook voor dieren zijn bomen belangrijk. Vogels en insecten vinden er bijvoorbeeld hun voedsel en onderkomen.

Jaarlijks geven we alleen in Nederland al vele honderden miljoenen euro's uit aan het beheer van bestaande bomen en de aanplant van nieuwe bomen. Een groot gedeelte van dit geld gaat op aan het monitoren van de status van bomen. Is de boom nog gezond? Veroorzaakt de boom ook overlast? Et cetera.

Een aantal jaren geleden is Alterra gestart met het ontwikkelen van een procedure die op basis van hoogtedata een landsdekkend bestand kan maken van bomenkronen in Nederland. Inmiddels heeft dit geleid tot een samenwerkingsverband met Neo B.V. en Geodan B.V. Samen met deze twee partijen wordt gewerkt aan een dataset met informatie over alle bomenkronen in Nederland die voor iedereen vrijelijk beschikbaar is en online (<http://www.boomregister.nl>) kan worden geraadpleegd. Recentelijk is dit samenwerkingsverband overgegaan in een coöperatie. Naast Neo B.V., Geodan B.V. en de WUR, maakt ook Cobra BV deel uit van deze coöperatie. Het boomkronenbestand wordt daarbij gebruikt als basis voor de ontwikkeling van verdere diensten rondom bomen.

Een belangrijk aspect van het bomenkronenbestand is de kwaliteit. Wil het bestand bruikbaar zijn voor bijvoorbeeld gemeenten, dan is het van belang dat de kwaliteit hoog genoeg is. In dit rapport wordt onderzocht wat de kwaliteit is van het landsdekkende bestand met bomenkronen in Nederland. Voor een drietal type landschappen wordt de kwaliteit gemeten. Hierbij wordt voornamelijk gelet op volledigheid en de compleetheid van het bestand.

Om de kwaliteit van het bomenkronenbestand over een langere tijd te kunnen monitoren, is het van belang dat de kwaliteit steeds op dezelfde manier wordt getoetst. Dit maakt het mogelijk om de verschillende uitkomsten naast elkaar te leggen. Daarom wordt er naast een kwaliteitstoets van verschillende versies van het bestand ook onderzocht in hoeverre het mogelijk is om een standaard kwaliteitsraamwerk op te zetten.

Samenvatting

Zowel Europees als nationaal zien we een toegenomen belangstelling voor bomen en houtopstanden. Dit is ook niet verwonderlijk gelet op de belangrijke rol die bomen spelen in ons dagelijks leven. Maar wil je als gemeente of als landelijke overheid beleid maken, dan heb je informatie nodig over de locatie en het aantal bomen in een bepaald gebied. Recentelijk is een landsdekkend bestand beschikbaar gekomen met informatie over boomkronen in Nederland. Dit landsdekkende bestand is het resultaat van een samenwerkingsverband tussen Neo B.V., Geodan B.V. en Alterra Wageningen UR.

Hoewel het bestand heel Nederland dekt, is onduidelijk in hoeverre alle boomkronen ook volledig en juist in beeld zijn gebracht. Om hier een beter beeld van te krijgen zijn de volgende vragen geformuleerd:

- In hoeverre corresponderen de uit de Boom Extractie Methode (BEM) resulterende kroonprojecties met de werkelijke boomlocaties, d.w.z. de geografische stamposities?
- In hoeverre komen de door BEM bepaalde oppervlakte en ligging van de kroonprojecties overeen met die van de werkelijke boomkronen?
- Hoe correct zijn de door BEM in CP geplaatste waarden van de hoogteattributen van de kroonprojectie?

In dit rapport wordt aan de hand van een tweetal validatiemethoden getracht om voor drie verschillende testgebieden (landelijk, stedelijk en park) antwoord te geven op bovenstaande vragen.

Conclusies validatie landelijk en stedelijk gebied

De verschillende versies van het boomextractie algoritme zijn op meerdere manieren getest om een totaaloverzicht te verkrijgen van de kwaliteit ervan. Het is niet eenvoudig om aan te geven welke versie van het algoritme de beste is. Als die zou worden uitgedrukt in één enkele waarde – de berekende Kwaliteit – dan is direct de vraag wat de onderliggende waarden van Volledigheid en Correctheid waren voor de verschillende testgebieden, omdat die immers nogal verschillen. De prestaties van BEM worden uitgedrukt door een combinatie van de nauwkeurigheid van individuele boomlocaties, van boomkroonoppervlakte en van de afbakening van individuele bomen. Welke beoordeling het belangrijkste is, hangt af van wat de uiteindelijke gebruiker belangrijk vindt.

De eerder beschreven assessments van drie verschillende variabelen is alleen uitgevoerd voor het stedelijk gebied. Voor het landelijk gebied is alleen de nauwkeurigheid van de boomkroonoppervlakte getoetst. Om een beter beeld te krijgen van de prestaties van BEM voor het landelijk gebied is meer onderzoek nodig. Dit neemt niet weg dat de nauwkeurigheid van de boomkroonoppervlakte een goede indicator is voor de prestaties van het algoritme voor het landelijk gebied. Daarnaast maakt het een vergelijking tussen het landelijk en stedelijk gebied mogelijk.

Vanuit een commercieel oogpunt hanteren bedrijven zoals NEO vaak een kwaliteitseis van 95% nauwkeurigheid voor hun classificatieproducten. Aangezien deze waarde door geen enkele versie van het algoritme is gerealiseerd, kunnen we stellen dat het huidige algoritme dit kwaliteitsniveau niet kan realiseren. Daar komt bij dat de resultaten van de verschillende versies van het algoritme laten zien dat alleen het aanpassen van het algoritme weinig impact heeft op de gerealiseerde overall kwaliteit. Sterker nog, het lijkt erop dat wanneer de volledigheid van het resultaat toeneemt, dit ten koste gaat van de correctheid. Waarschijnlijk kunnen we dit toewijzen aan het feit dat de procedure alleen gebruikmaakt van AHN2 rasters als input.

Zoals BEM v10 laat zien, is het mogelijk om op basis van additionele informatie de kwaliteit van de classificering te verbeteren. Met name de volledigheid neemt toe. De correctheid daarentegen blijft nagenoeg gelijk.

Conclusies validatie parkgebied

Referentiemodel

Uit de kartering van de BEM-resultaten met het PlantCover deel van het referentiemodel en de overlap-analyse ervan voor de verschillende versies blijkt dat de kwaliteit van het referentiemodel te wensen overlaat wat betreft het verloop en de ligging van de buitengrens van de PlantCover polygonen. Er lijkt een systematische fout in te zitten, wat blijkt uit het feit dat de False Negatives vaak voorkomen aan de NO-zijde van objecten, en False Positives aan de ZW-zijde ervan. Mogelijk is dat te wijten aan de resolutie van de gebruikte luchtfoto-opnames en aan de omvalling van de daarop zichtbare objecten. Daarnaast valt op te merken dat er in de PlantCover objecten van het referentiemodel geen open plekken zijn opgenomen. Het referentiemodel van het Campusterrein biedt echter voldoende aanknopingspunten om te kunnen concluderen dat er verschillen zijn tussen de bekeken versies van BEM.

Boom Extractie Methode

Uit het rekenwerk van de overlapanalyses blijkt dat versie 9.2 van BEM beter presteert dan versie 5. Dit komt op hoofdlijnen overeen met de conclusies van Neo B.V. over de prestaties van BEM voor het stukje stedelijk gebied in Amersfoort. Een duidelijk zwak punt van BEM is dat bij aaneengesloten boomkronen BEM geen betrouwbare grenslijn tussen de kronen weet te trekken. Ook kan met BEM geen betrouwbare stampositie per kroon worden berekend, omdat als uitgangsmateriaal alleen de hoogtes van de bovenkant van de boomkronen bekend zijn.

Ten slotte valt op dat in versie 9.2 in tientallen gevallen lantaarnpalen door BEM als boom waren aangemerkt. Zij lijken in het winterseizoen qua hoogte en diameter kennelijk op jonge boompjes, speciaal als ze in gras staan.

Uitgangsdata

De problematiek van de extractie van het juiste aantal van de juiste objecten wordt niet alleen veroorzaakt door een tekort aan data over waar vegetatie is, maar ook door de relatieve grofheid van het AHN2-raster. Jonge boompjes, zeker in het winterseizoen, zijn te klein en te dun om in het AHN2-grid terecht te komen.

Er valt te verwachten dat, als boomdetectie in plaats van op 50x50 cm gridcellen wordt gebaseerd op de puntenwolk die aan AHN2 ten grondslag ligt, de detectie succesvoller (grotere correctheid, grotere volledigheid) kan zijn.

Conclusies met betrekking tot de kwaliteit

In dit rapport is de kwaliteit van het landsdekkende boomkronenbestand onderzocht en gerapporteerd, gebaseerd op verschillende versies van de Boom Extractie Methode (BEM). Aan de hand van drie verschillende kleine proefgebieden is de kwaliteit van het bestand beschreven. De gebruikte benaderingen zijn echter niet identiek, waardoor de resultaten niet helemaal vergelijkbaar zijn. Daarnaast zijn bij de beschreven validaties de kwaliteitseisen van mogelijke eindgebruikers onderbelicht gebleven. Voor een goede vergelijking van de kwaliteit is het van belang om te komen met een geïntegreerde benadering van de kwaliteit, waarbij de verschillende aspecten van het gebruik en productieproces een plaats hebben.

Kwaliteitsraamwerk

Alle drie de validatiemethoden geven een beeld van de kwaliteit van het boomkronenbestand. Toch is dit beeld nog niet volledig. Deels hangt dit samen met de verschillende manieren waarop de kwaliteit is getoetst, maar deels ook met het feit dat niet alle landschapstypen zijn gecontroleerd. Verder is het ook nog maar de vraag in hoeverre de gemiddelde gebruiker iets heeft aan de uitkomsten van de uitgevoerde validaties en nu kan beoordelen wat de bruikbaarheid is van het boomkronenbestand.

In dit rapport wordt een methodische handreiking gedaan in de vorm van een raamwerk dat de producent kan helpen bij het beschrijven van de dataset die hij heeft geproduceerd. Het raamwerk kan ook de gebruiker helpen bij het beoordelen van de fitness for use, de geschiktheid voor zijn gebruiksdoel. Binnen de scope van dit onderzoek is het niet mogelijk om een volledig kwaliteitsoordeel

te geven over het boomkronenbestand. Toch is er een aantal conclusies te trekken en aanbevelingen te geven.

Op dit moment is er geen platform beschikbaar waar gebruikers met vragen over de open data versie van het boomkronenbestand terecht kunnen. Een mogelijkheid zou kunnen zijn om hier een apart platform voor te creëren, met name gericht op open data gebruikers vergelijkbaar met bijvoorbeeld Open Street Map (OSM). Recentelijk is er een coöperatie opgezet rondom het vermarkten van het boomkronenbestand. Dus een andere optie zou kunnen zijn om dit onder te brengen bij deze nieuwe coöperatie.

Een andere manier om de interactie tussen de open data gebruiker en de producent van het boomkronenbestand te verbeteren, is door alle gebruikers van de dataset zich eerst te laten registreren. Voor de producent wordt dan iets meer inzichtelijk wie de gebruikers zijn. Daarnaast wordt het dan ook makkelijker om met deze groep in contact te treden.

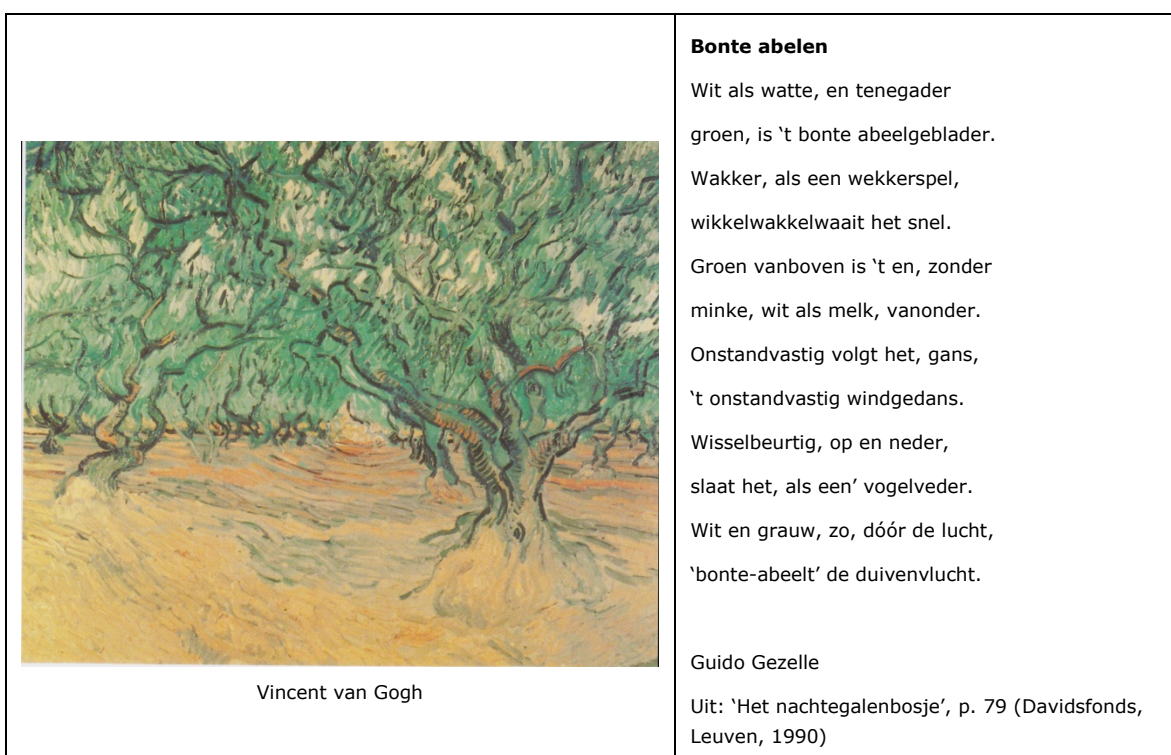
De metadata zijn in hun huidige vorm onvoldoende om voor de gebruiker de fitness for use te bepalen. Los van de inhoud van de data speelt ook mee dat voor een groot aantal gebruikers de terminologie vaak te technisch is en te weinig houvast biedt om te bepalen of de dataset voor hen ook bruikbaar is.

Tot slot: zoals al eerder opgemerkt, is het niet mogelijk om alle vragen te beantwoorden die gebruikers rondom het boomkronenbestand kunnen hebben. Daarvoor is eerst meer onderzoek nodig en moeten meer gebieden worden onderzocht. Voorgesteld wordt om in een vervolgonderzoek het in dit hoofdstuk gepresenteerde kwaliteitsraamwerk als uitgangspunt te kiezen voor verdere controles.

1 Inleiding

1.1 Introductie

Bomen zijn voor mensen een belangrijke inspiratiebron. In tal van schilderijen en gedichten speelt de boom een belangrijke rol. Denk maar aan het schilderij *Olijfbomen in een berglandschap* van Vincent van Gogh (1889) of het gedicht *Bonte abelen* van Guido Gezelle uit 1897. Maar bomen zijn niet alleen een bron van inspiratie. Soms kunnen bomen ook een bron van ergernis (Randrup *et al.*, 2001) zijn of een bron van inkomsten (Oosterbaan *et al.*, 2009).



Vincent van Gogh

Figuur 1.1 Bomen als inspiratiebron voor schilders en dichters.

Zowel Europees als nationaal zien we een toegenomen belangstelling voor bomen en houtopstanden. Dit is ook niet verwonderlijk, gelet op de belangrijke rol die bomen spelen in ons dagelijks leven. Zo zijn er aanwijzingen dat bomen een positieve uitwerking hebben op het herstel na ziekte. Bomen spelen een belangrijke rol bij het verminderen van fijnstof (Oosterbaan *et al.*, 2007), hebben een directe impact op het microklimaat in steden (Bowler *et al.*, 2010) en vormen een belangrijk levend contrast in de wereld van beton en asfalt. Daarnaast vormen bomen een huis voor dieren en insecten.

Bomen zijn niet alleen interessant vanuit maatschappelijk oogpunt of vanuit het perspectief van de natuur. Ook vanuit bedrijfseconomisch perspectief zijn bomen interessant. Bijvoorbeeld als bron van hout en als bron van biomassa ten behoeve van energievoorziening (Kumar *et al.*, 2010; Viana *et al.*, 2010).

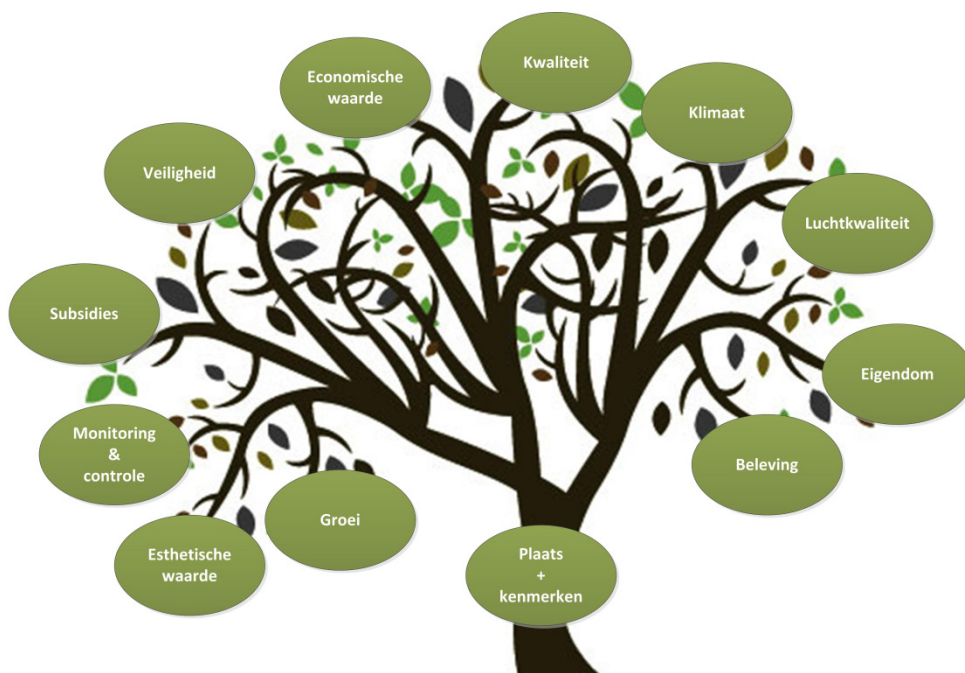
Ook op politiek en beleidsniveau zien we een toegenomen belangstelling voor bomen. Europees is men voornamelijk gericht op het landelijk gebied en streeft men naar het in stand houden van houtopstanden (zie bijvoorbeeld het project SIMWOOD (Sustainable Innovative Mobilisation of Wood)). Dat varieert van bossen tot bomenrijen en solitaire bomen. Een belangrijk Europees instrument hierbij

is bijvoorbeeld het Gemeenschappelijk Landbouw Beleid (GLB), waarin meer dan 40 miljard euro per jaar omgaat.

Nationaal ligt de focus veel meer op de individuele boom als vegetatieobject, al of niet met esthetische waarde. Jaarlijks besteden Nederlandse overheden zo'n 2 miljard euro aan activiteiten rondom het beheer van bomen. Aan individuele bomen buiten bosverband werd door gemeentelijke groenbeheerders in de periode 2008-2010 gemiddeld € 21 per boom per jaar uitgegeven en aan monumentale bomen € 52. Aan een vierkante meter bos werd in diezelfde periode € 0,25 - € 0,46 per jaar besteed (Jong *et al.* 2011, p. 44).

Vanuit het instandhoudingsstreven en de daarmee gepaard gaande verzorgingsnoodzaak zijn met name in de openbare ruimte de boombeheerders veelal geïnteresseerd in praktische aspecten rondom boombeheer. Vragen die daaraan uiting geven zijn bijvoorbeeld: Wat is de kwaliteit van de houtopstand in onze gemeente? Is de boom ziek of gezond? Veroorzaakt de boom overlast? Welke bomen zijn illegaal gekapt?

De vragen en de informatiebehoefte die men heeft ten aanzien van bomen lopen dus nogal uiteen en zijn soms erg complex. In Figuur 1.2 wordt dit geïllustreerd door een aantal aspecten van boom informatie op te sommen. Hoewel de informatiebehoefte zeer gevarieerd is bindt één aspect ze samen, namelijk: waar staan de bomen? Voor effectief bomenbeheer op gemeentelijk niveau is het van groot belang om de juiste locatie van de bomen te kennen. Bomen en hun eigenschappen moeten beter op de kaart komen te staan, letterlijk en figuurlijk.



Figuur 1.2 Bomen-informatiebehoefte.

Uitzonderingen daargelaten gebeurt het inwinnen van gegevens over bomen via veldwerk. Naast het bepalen van de positie van de individuele boom door landmeters is een goed voorbeeld hiervan de periodieke boomveiligheidscontrole. Het resultaat hiervan wordt opgeslagen in een lokaal beheerde database. Overigens blijkt de periodiciteit per beheerder te variëren, evenals de compleetheid. Zo zijn particuliere bomen niet bekend bij de gemeentelijke beheerders. Elke boombeheerder heeft zich vast weleens afgevraagd: kan dat niet anders? Het antwoord op deze vraag is een volmondig 'ja'.

1.2 Databronnen

Recentelijk zien we een aantal ontwikkelingen die het in kaart brengen van bomen sterk kunnen vereenvoudigen. Een van de interessantste ontwikkelingen is de sterke toename van landsdekkende, gedetailleerde ruimtelijke data, afkomstig vanuit de remote-sensinghoek. Remote sensing (RS) gebruiken we hier als een verzamelnaam voor onder meer luchtfoto's, satellietbeelden en LiDAR opnames. Ook het in 2013 voltooide Actueel Hoogtebestand Nederland (AHN2) is met behulp van remote sensing ingewonnen.

Deze nieuwe landsdekkende databron, in combinatie met nieuwe softwareontwikkelingen, maakt het mogelijk om geautomatiseerd bomen in Nederland te karteren. Dit is gedaan door Alterra Wageningen UR in de periode 2011-2013. Uit het AHN2-bestand (0.5x0.5 m gridcellen) zijn alle boomkronen van Nederland geëxtraheerd als polygonen. In dit document wordt naar deze werkwijze verwezen als de 'Boom Extractie Methode' (BEM).

1.3 Toepassingsmogelijkheden

De toepassingsmogelijkheden voor een landsdekkend bestand met alle boomkronen in Nederland zijn legio. Op basis van een dergelijk bestand kunnen gemeenten bijvoorbeeld inventariseren welke bomen niet zijn opgenomen in hun gemeentelijke database en welke bomen inmiddels zijn verdwenen. Hetzelfde bestand kan ook een bijdrage leveren aan het vergroenen van het Gemeenschappelijk Landbouw Beleid (GLB) door een antwoord te geven op de vraag waar alle bomenrijen in Nederland staan. Naast de eerder beschreven toepassingsmogelijkheden van het bestand is de techniek die ontwikkeld is om boomkronen uit het AHN2-bestand te extraheren ook geschikt voor andere toepassingen. Bijvoorbeeld het extraheren van bepaalde landschapselementen. In dit document wordt hier echter niet op ingegaan.



Figuur 1.3 Een fragment van het Boomkronenbestand. Achtergrond: luchtfoto 2010, Kadaster.

1.4 Opzet rapport

Wat is de kwaliteit van het boomkronenbestand? In dit rapport wordt vanuit twee verschillende invalshoeken getracht om de kwaliteit ervan te bepalen. Kwaliteit wordt in beide gevallen opgevat als de mate van overeenkomst van het boomkroonprojectiebestand met een model van de werkelijkheid. Hoe groter de overeenkomst, hoe beter.

Om een idee te krijgen of het in kaart brengen van bomen ook afhankelijk is van het type gebied wordt in dit onderzoek gekeken naar verschillende landschapstypen. Er zijn verschillende definities van het begrip landschap. Sommige hebben een beleidsmatige invalshoek, terwijl andere een meer wetenschappelijke invulling geven aan het begrip landschap. In Bouma (2008) wordt landschap gedefinieerd als een *'gebied zoals dat door mensen wordt waargenomen en waarvan het karakter bepaald wordt door natuurlijke en/of menselijke factoren en de interactie daartussen'*. In dit onderzoek is de blik gericht op landschappen met bomen als beplanting. Er wordt gekeken hoe goed de boomkronen stedelijke straatbomen weergeven, bomen in een parkomgeving en de beplanting in een landelijk gebied.

Wil je de kwaliteit in de tijd gezien monitoren en verschillende onderzoeken met elkaar kunnen vergelijken, is het wenselijk dat voor het controleren van bijvoorbeeld de kwaliteit uitgegaan wordt van dezelfde methodiek. Daarom wordt er ook gekeken in hoeverre beide methodieken met elkaar gecombineerd kunnen worden tot een standaardprotocol, eventueel met nieuwe kwaliteitselementen. Hierbij wordt ook gekeken naar de eisen die gesteld worden aan de basisregisters in Nederland.

De opbouw van dit rapport is naast dit inleidende hoofdstuk als volgt: in hoofdstuk 2 wordt eerst een overzicht gegeven van de achtergronden van de boomkronenextractiemethode. In hoofdstuk 3 wordt een beschrijving gegeven van de verschillende validatiemethoden van de boomkronenextractiemethode. In hoofdstuk 4 worden de resultaten gepresenteerd en bediscussieerd, in hoofdstuk 5 worden de conclusies en aanbevelingen gepresenteerd van de verschillende validatiemethoden en tot slot hoofdstuk 6, waarin een kwaliteitsraamwerk wordt gepresenteerd.

1.5 Samenwerking

Aanleiding voor de uitgevoerde validaties zijn de activiteiten van het Publiek-Private Samenwerkingsproject Open Boomdata. De partners daarvan zijn, behalve Alterra Wageningen UR: Neo B.V. in Amersfoort en Geodan B.V. in Amsterdam. Daarin is toegewerkt naar het als Open Data beschikbaar stellen van de geëxtraheerde boomkronen. Het resultaat hiervan kan men onder andere raadplegen via de website <http://www.boomregister.nl> (maart 2014).

Inmiddels is de PPS overgaan in een coöperatie, getiteld Boomregister.NL. Aan deze coöperatie nemen Neo B.V., Geodan B.V., Cobra BV en Wageningen University and Research (WUR) deel. Een aantal belangrijke doelstellingen van deze coöperatie zijn het vermarkten, onderhouden en verder doorontwikkelen van het boomkronenbestand.

1.6 Eerdere publicaties

De onderzoeksresultaten die in dit document gepresenteerd worden zijn gebaseerd op een tweetal onderzoeken die in 2013 zijn uitgevoerd, validaties van het Boomkronenbestand. De eerste validatie is uitgevoerd door Rik van Benthem, die ten tijde van dit onderzoek stagiair was bij Neo B.V. Zijn validaties zijn geschreven in het Engels ten behoeve van zijn masterstudie bij Universiteit Utrecht. De tweede validatie is uitgevoerd door Frans Rip, medewerker van Wageningen UR GeoDesk. Deze is geschreven in het Nederlands. Om de leesbaarheid van dit rapport te vergroten, is ervoor gekozen om de validatie die door Rik van Benthem is uitgevoerd te vertalen naar het Nederlands. Daarnaast zijn beide validatierapporten zodanig samengevoegd dat de overeenkomstige hoofdstukken bij elkaar staan.

1.7 Begripsbepalingen en afkortingen

In deze paragraaf wordt een overzicht gegeven van alle in dit document gebruikte termen en afkortingen.

Afktoring / Begrip	Omschrijving
AHN2	Het Actueel Hoogtebestand Nederland (AHN2) is een bestand met voor heel Nederland gedetailleerde en precieze hoogtegegevens. Een digitale hoogtekaart als het ware. Voor heel Nederland is van elke vierkante meter bekend wat de hoogte is. De waterschappen en Rijkswaterstaat laten het AHN maken voor hun dagelijks werk, met name voor waterbeheer en waterkeringbeheer. Maar ook voor andere toepassingen wordt het AHN gebruikt (bron: www.ahn.nl).
BAG	Basis Administratie Gebouwen
BEM	Boomkronen Extractie Methode (om boomkronen uit AHN2 te halen)
CP	Boomkroonprojecties bestand (Canopy Projections dataset). De recentste versie is Clement J., 2013: Canopy Projections [dataset, version 9.2]. Alterra Wageningen UR
DEM	Digitaal Elevatie Model
DTM	Digitaal Terrein Model
LiDAR	Light Detection And Ranging of Laser imaging Detection And Ranging is een techniek die de afstand tot een object of oppervlak bepaalt door middel van het gebruik van laserpulsen. De techniek is vergelijkbaar met radar, dat echter radiogolven gebruikt in plaats van licht. De afstand tot het object of oppervlak wordt bepaald door de tijd te meten die verstrijkt tussen het uitzenden van een puls en het opvangen van een reflectie van die puls.
NAP	Nieuw Amsterdams Peil
NDVI	De N ormalized D ifference V egetation I ndex is een simpele grafische indicator die gebruikt wordt om remote sensing-data te analyseren. De indexwaarde geeft aan wat het aandeel is van levende vegetatie op een geobserveerde plek. De waarde varieert tussen de -1 en de +1, waarbij positieve waarden vanaf ongeveer 0,2 duiden op levende vegetatie.
PPS	Publiek-Private-Samenwerking: Publiek-private samenwerking kunnen we losjes omschrijven als 'min of meer duurzame samenwerking tussen publieke en private actoren waarin gemeenschappelijke producten en/of diensten worden ontwikkeld en waarin risico's, kosten en opbrengsten worden gedeeld' (Klijn/Teisman 2000).
RMSE	Root Mean Square Error
RS	Remote Sensing gebruiken we hier als een verzamelnaam voor onder meer luchtfoto's, satellietbeelden en LiDAR-opnames.
Volledigheid	Onder volledigheid verstaat men de mate waarin boomobjecten, die in werkelijkheid voorkomen, in het bestand zijn opgenomen.
VTA	Visual Tree Assessment

2 Beschrijving Boomkronen Extractie Methode

Dit hoofdstuk beschrijft de door Alterra ontwikkelde methode om informatie over boomkronen uit het AHN2-bestand af te leiden. In de laatste fase van de ontwikkeling is van de kant van Neo B.V. waardevolle feedback geleverd.

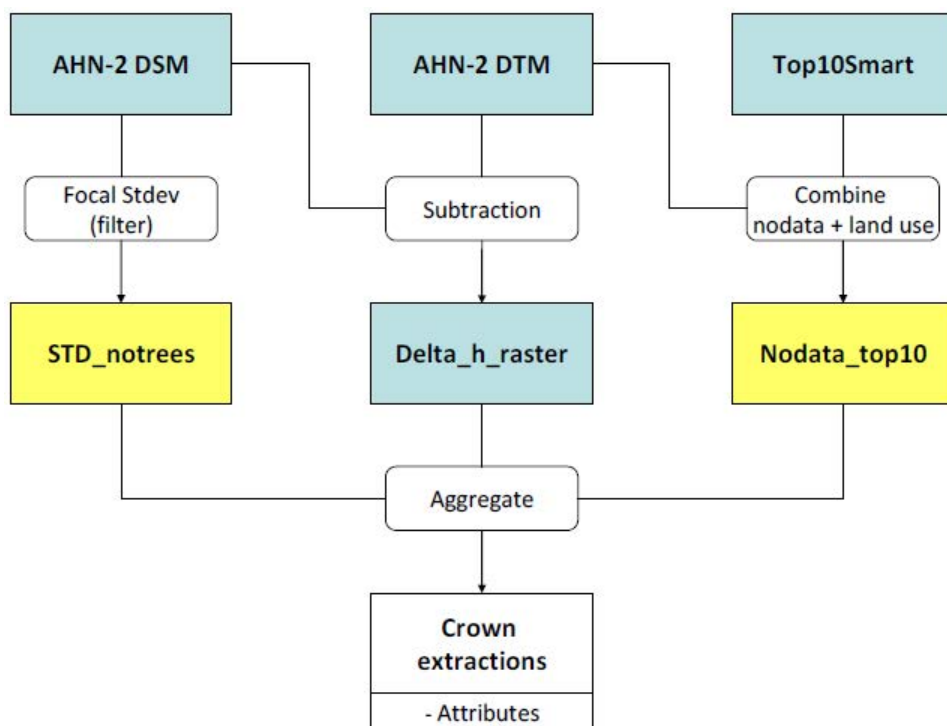
2.1 Het extractiealgoritme

De Boomkronen Extractie Methode (BEM) bestaat uit twee aparte trajecten.

1. In het eerste traject worden uit het AHN2-hoogteraster de boomkronen geëxtraheerd als polygoon met attributen. Zo'n polygoon representeert de projectie van de boomkroon op het grondvlak. Het resulterende bestand wordt hierna aangeduid als 'CP', het Canopy Projection-bestand.
2. In het tweede traject worden deze polygoon en hun attributen gebruikt om de attributwaarden volgens het Silvi-Star-model (Koop 1989) te berekenen. Daarmee wordt de overgang gemaakt naar een boommodel als driedimensionaal object.

Deze twee trajecten worden in volgende paragrafen toegelicht. De logistieke aspecten van BEM, zoals het verwerken van het omvangrijke AHN2-bestand, blijven hier buiten beschouwing.

2.1.1 Extractie van de boomkroon



Figuur 2.1 Flowchart van het boomkroonextractiealgoritme.

Het extractiealgoritme gebruikt het AHN2 ('Actueel Hoogtemodel Nederland') hoogtemodel om bomen te vinden en te classificeren. AHN2 is een Digitaal Elevatie Model (DEM) dat heel Nederland dekt. Het werd ingewonnen in de periode 2007-2012 onder directie van het Waterschapshuis. Het bestand bestaat uit 36027 'bladen', die elk 1000 x 1250 meter groot zijn en 5 miljoen rastercellen bevatten.

De elevatie werd vanuit de lucht ingewonnen met behulp van Laser altimetrie, ook aangeduid als LiDAR: Light Detection And Ranging. Er zijn verschillende AHN2-producten verkrijgbaar. Hier zijn relevant om te noemen: LiDAR puntenwolken, Digitale Terrein Modellen (DTM) en 'Digital Surface Models' (DSM). De puntenwolk is de collectie van alle verzamelde punten. Die worden gegeneraliseerd naar 1 punt per horizontale rastercel van 50x50 cm. DTM (AHN-aanduiding: 'Int') geeft het uit de puntenwolk gefilterde maaiveld weer, terwijl DSM (AHN aanduiding: 'Ruw') de 'bovenkant' van objecten weergeeft (bouwsels, vegetatie, enzovoort).

De puntenwolk bevat gemiddeld 8 punten/m². DTM en DSM zijn rasterdata met een resolutie van 0.5 m, waarbij het DTM een verticale nauwkeurigheid van 5 cm¹ heeft. Deze bestanden geven de hoogte per rastercel boven het nul-niveau (Nieuw Amsterdams Peil). Het DTM-bestand, het maaiveld, wordt afgeleid uit DSM door bouwsels, objecten en vegetatie uit DSM weg te filteren.

Een eigenschap van inwinning van data met LiDAR is dat wateroppervlakken (meren, rivieren, plassen op het land) weinig of geen punten opleveren. Zulke plekken leveren cellen zonder waarde in het rasterbestand, hierna aangeduid als 'nodata'. Ook onder dichte begroeiing, en uiteraard onder gebouwen en auto's, bevat het DTM-raster nodata.

Om de Canopy Projection (CP) te maken, worden zowel AHN2 DTM als AHN2 DSM gebruikt. De eerste stap is om de celwaarden in DTM af te trekken van de celwaarden in DSM. Daardoor ontstaat een genormaliseerd hoogteraster, aangeduid als het 'delta-h-raster'. Dit representeert per rastercel de hoogte boven het maaiveld².

Om het zoekgebied van het algoritme te verkleinen tot de voor CP relevante objecten, de bomen, worden twee uitsluitingsmaskers gemaakt:

- Het **nodata-top10 masker** Dit masker dient om objecten in het zoekgebied die geen boom zijn op maaiveld-niveau uit te sluiten. Het bevat de no-data-gebieden van de AHN2-DTM (water, terrein onder auto's, locaties van uitgefilterde gebouwen). Daarnaast bevat het de gebouwomtrek. In eerdere versies van BEM werd de gebouwomtrek ontleend aan het Top10Smart³ bestand. Dat is een rasterversie van de Top10NL bestand⁴ met een celgrootte van 2,5 m. In versie 9.2 en latere werden hiervoor de nauwkeurigere polygonen uit het BAG bestand (Basisregistratie Adressen en Gebouwen⁵) gebruikt.
- Het **STD_notrees masker**. Dit masker is gemaakt door een standaardafwijkingsfilter toe te passen op het AHN2-DSM. Dit masker bevat een ondergrens voor de standaardafwijking van de hoogtewaarden in een gebied van 3x3 cellen. Rastercellen met een standaardafwijking (STD-waarde) lager dan de drempel worden uit het zoekgebied verwijderd. Dat komt neer op verwijdering van egaal vlakke gebieden. Bomen en andere vegetatie, die als gevolg van hun textuur een relatief hoge standaardafwijking vertonen, blijven over.

Nadat het zoekgebied voor het algoritme door toepassing van de maskers is ingeperkt, vindt de extractie van boomkronen plaats op basis van het resterende deel van het delta-h-raster. Uitgangspunt is dat een boomkroon in een rasterdataset herkenbaar is als een gebied rondom een piekwaarde (de hoogte van de boomtop). De rastercellen met de piekwaarden worden opgezocht door te kijken welke celwaarden het meest afwijken (naar boven toe) van de gemiddelde genormaliseerde hoogte.

De individuele kronen in een gebied met meerdere pieken worden bepaald door rond elke piek zones te bepalen: de fragmentatie. De zone rond een piek bestaat uit de verzameling cellen waarvoor geldt dat ze dichterbij de piek in die zone liggen dan bij een andere piek. De zone rond een piek bevat dus de cellen met de kleinste euclidische afstand tot die cel met de piekwaarde.

¹ http://www.ahn.nl/wat_is_het_ahn/techniek_van

² Dit landsdekkende bestand is sinds 2013 beschikbaar als zelfstandig product: Kramer, H., Clement, J., Mûcher, C.A., 2013: dataset OHN – Object Hoogten Nederland [computer file]. Wageningen UR – Alterra

³ Zie <http://www.wageningenur.nl/en/show/TOP10Smart.htm>

⁴ Zie <http://www.kadaster.nl/web/artikel/productartikel/TOP10NL.htm>

⁵ Zie <http://www.kadaster.nl/web/Themas/Registraties/BAG-1.htm>

In een iteratief proces worden daarna correcties aangebracht op de grootte van de zones. Te kleine cel-clusters worden gecombineerd met al bestaande zones. Het laatste onderdeel van het proces is de omzetting van een cluster cellen in een polygoon met de buitenste hoekpunten van de buitenste cellen als polygoon-knikpunt. Het algoritme legt ook attributen van de geëxtraheerde boomkroon vast, zoals te zien is in Tabel 2.1.1.

Tabel 2.1.1

De attributen van de geëxtraheerde boomkronen.

Attribuut naam	Beschrijving
BLAD_ID	Nummer van het corresponderende AHN2-blad
BOOM_ID	Uniek boomkroonnummer
BM_X	X-coördinaat van het kroon zwaartepunt (pseudostamlocatie)
BM_Y	Y-coördinaat van het kroon zwaartepunt (pseudostamlocatie)
GEM_MV	Gemiddelde maaiveldhoogte onder de kroon (AHN2 DTM)
BOOM_HOOGT	Maximale genormaliseerde hoogte (delta_h_raster) binnen kroonprojectie
GEM_HOOGTE	Gemiddelde genormaliseerde hoogte (delta_h_raster) binnen kroonprojectie
SHAPE_AREA	Oppervlak van de kroonprojectie in m ²



Figuur 2.2 Voorbeeld van het resultaat van de boomkroonextractie. De boomkroon is weergegeven in groen.

Figuur 2.2 toont de kroonprojectie zoals het algoritme die produceert. Omdat de resulterende kroonontrek enkel is gebaseerd op de hoogtegegevens van de bovenkant van de kroon, kunnen er geen conclusies worden getrokken over de boom daaronder. Daarom moet in gedachten worden gehouden dat de geëxtraheerde boomkronen in werkelijkheid soms uit meerdere bomen bestaan. De plaatsaanduiding door de waarden van de attributenduo BM_X en BM_Y kan dan ook niet als precieze stamlocatie worden gezien.

2.1.2 Versies van CP

Er zijn in de periode 2011-2013 meerdere experimentele versies van het algoritme gemaakt, die elkaar soms snel opvolgden. De extractie van kronen uit AHN2 was met versie 6 voltooid. Daarna lag de nadruk op het verbeteren van de input voor de filters: datasets en uitsluitingscriteria.

De toepassing van versie 3 op de bomen in de Duifpolder werd al medio 2012 beschreven (Schouten, 2012)⁶ en versie 5 stamt uit oktober 2012. De versies 6 tot en met 11 kwamen daarna in de periode tot mei 2013 tot stand.

Aanleiding voor de opeenvolging van versies waren de commentaren van incidentele gebruikers op testversies. Een stand in het Innovatiepaviljoen op de GIS Conferentie van ESRI in september 2012 leverde tussen 10 en 20 belangstellenden op, afkomstig van waterschappen, gemeenten, boomadviseurs en andere bij boombeheer betrokken organisaties, die een specifiek proefgebied van 1x1 km met boomkronen aanvroegen. Als ze feedback gaven, ging dat vaak over het ontbreken van wel aanwezige bomen, of delen daarvan.

Het leidende thema bij de opeenvolging van versies was het vinden van zo goed mogelijke criteria voor welk extractieresultaat wel of niet als boomkroon wordt beschouwd. Dit was – en is – een proefondervindelijk proces. Daarbij spelen zowel de uitsluitingsmaskers een rol als de criteria om uit de gebieden met hogere standaardafwijking de boomkronen te onderscheiden van niet-boomkronen. De verschillen in de versies na V6 zijn hoofdzakelijk gelegen in de manier van filtering van de output, het maskeren van gebouwen en het opvullen van nodata. Daarbij ging het om experimenten. Niet elke volgende versie was beter dan de vorige.

Tabel 2.1.2 geeft een overzicht. Daarna worden die aandachtspunten afzonderlijk toegelicht.

Tabel 2.1.2

Overzicht van de verschillende versies van het boomkroon extractiealgoritme.

Naam	Gebouwmasker	No data	Filter output	Overig
V5	Top10smart	No data uit AHN2 DTM (gefilterde en opgevulde versie)	Kroon > 1 m hoog, opp. > 1 m ²	
V6	Top10smart	No data uit AHN2 DTM (gefilterde en opgevulde versie)	Kroon > 1 m hoog, opp. > 1 m ²	
V7	Filter gebaseerd op no data AHN2 DTM Gebouwen uit Top10NL + 1 m buffer	Opgevuld a.d.h.v. het AHN2 5 m grid (alle no data verwijderd)	Kroon > 4 m hoog Kroon opp. > 5 m ²	
V8				Extractie als in v6 Omval-correctie toegevoegd
V9.1, v9.2	Gebruik BAG IR Lufo2008 NDVI >0.2			Extractie als in v6 Omval-correctie
V10				Extractie als in v6 Omval-correctie
V11a	Formosat NDVI BGT			Extractie als in v6 Omval-correctie
V11b	IR Lufo2008 NDVI			

⁶ L. Schouten *et al.*, 2012: Bomen modelleren met laseraltimetrie. Geo-Info 2012-7, p.3-7

Filtering van output

Als criteria voor de 'output filtering' van de gevonden objecten zijn gebruikt:

- de hoogte. Het gaat erom binnen de gevonden opgaande vegetatie onderscheid te maken tussen bomen en lager struikgewas, hagen e.d.
- de oppervlakte. Objecten met een kleine oppervlakte kunnen jonge boompjes zijn, maar ook lantaarnpalen, verkeersborden, vlaggenmasten of nog andere ruis, zoals gebouwranden.

Verbetering van het gebouwmasker

Als verbetering in de manier om vegetatie van niet-vegetatie te onderscheiden, is na versie 7 geëxperimenteerd met het gebruiken van NDVI (Normalized Difference Vegetation Index)⁷. Dit is een indexwaarde tussen -1 en +1, afgeleid uit de hoeveelheid reflectie van infrarood door vegetatie⁸. Een waarde groter dan +0,2 wijst op levende beplanting.

De uitsluiting van rastercellen met een lagere NDVI-waarde biedt een betrouwbaarder filter dan het gebruik van categorieën uit de topografische kaart.

Dat brengt echter ook nieuwe problemen met zich mee:

- Het is van belang met welke resolutie de NDVI-gegevens beschikbaar zijn. Als die resolutie te grof is in verhouding tot de AHN2-rastercelmaat van 50x50cm, worden bomen met een geringe kroonvang gemist. Bij een experiment met Formosat-data (met resolutie van 2 m) bleek dat 2 m een te grove resolutie is, die voor het stedelijk gebied geen meerwaarde biedt. De bruikbaarheid van een NDVI-dataset uit 2008 met een 25 cm-resolutie bleek groter.
- De introductie van NDVI brengt een nieuwe uitdaging met zich mee: omvalling. Luchtfoto's hebben, als ze schuin ('oblique') zijn gemaakt in plaats van verticaal ('orthofoto'), door de relatief geringe hoogte last van zijwaartse vertekening van de ligging van objecten, aangeduid als 'omvalling' of parallax. Daardoor liggen de gridcellen met NDVI-waarden van 0.2 en groter, die onderdeel van een boomkroon zouden kunnen zijn, niet op dezelfde plaats als de AHN2-hoogtes van hetzelfde object. Om die rastercellen in het NDVI-bestand dan te combineren met de aan AHN2 ontleende boomkronen waar ze bij horen, zijn extra filterbewerkingen nodig. Daarvoor is de omvalcorrectie-procedure ontwikkeld. Die doet het volgende:
 - Als er minstens een kleine overlap is tussen object in AHN2 (STD > 1.5) en object in NDVI (NDVI > 0.7), dan
 - bufferzone aanbrengen rond het NDVI object ter grootte van 3 rastercellen.
 - Waar deze bufferzone het aangrenzende AHN2-object overlapt is kennelijk het AHN2-object incompleet. De overlappende rastercellen worden aan het AHN2-object toegevoegd. Een andere mogelijkheid is om omvalling te vermijden. Dit kan door orthobeelden te construeren uit overlappende luchtfoto's.
- De datum waarop de infraroodopnames zijn gemaakt, is van groot belang om twee redenen.
 - In de eerste plaats moet die datum zo dicht mogelijk liggen bij de opnamedatum van AHN2 vanwege de gewenste zekerheid dat het om dezelfde objecten gaat.
 - Daarnaast is het wenselijk om de NDVI te baseren op opnames in het groeiseizoen, bij voorkeur zo veel mogelijk op dezelfde datum. Beelden van buiten het groeiseizoen zijn vanwege bladval – dus minder infraroodreflectie, dus een ander NDVI-patroon – niet bruikbaar voor het hier benodigde vegetatiefilter. Maar landsdekkende opnames in het groeiseizoen zijn schaars. In Nederland zijn in 2008 en 2011 landsdekkende opnames gemaakt. De schaarste is het gevolg van voorkeur bij de opdrachtgevers voor inwinning waarbij de boomkronen zo min mogelijk het beeld op het aardoppervlak in de weg staan. Dus worden de opnames bij voorkeur buiten het groeiseizoen gemaakt. De combinatie van wensen voor een hoge resolutie, groeiseizoen-opnames en een inwinningsdatum zo dicht mogelijk bij de AHN2-opnamedatum is voor Nederland als geheel moeilijk te realiseren.

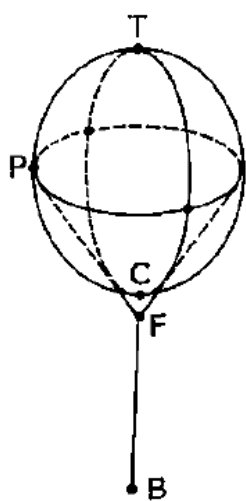
⁷ Zie http://en.wikipedia.org/wiki/Normalized_Difference_Vegetation_Index

⁸ Bronbestand: Eurosense 2009. DKLN2008-CIR, 2009-07-22. Eurosense B.V.

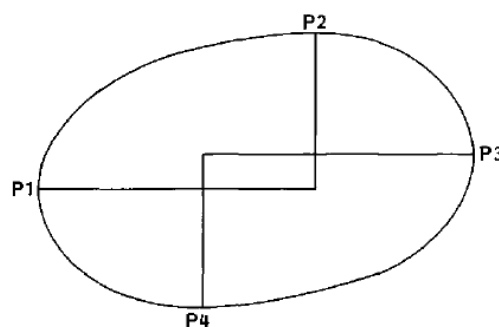
2.2 Het SILVI-STAR-model

De resultaten van het extractie-algoritme worden opgeslagen als polygonen (ESRI-shape), met attributen volgens het SILVI-STAR-boommodel (Koop 1989). SILVI-STAR is ontworpen om boomattributen op te slaan als een geparametriseerd 3D-model. Opslag van bomen als 3D-objecten heeft verschillende voordelen: het faciliteert bijvoorbeeld berekeningen van het kroonvolume, biomassaschattingen en de opslagcapaciteit voor CO₂ of fijnstof, maar het maakt ook een betere en meer realistische visuele weergave van bomen mogelijk.

SILVI-STAR beschrijft bomen aan de hand van de X,Y,Z-coördinaten van 8 punten. Die punten vertegenwoordigen: de stambasis (B), de hoogte van de 1^e vertakking (F), de top van kroon (T), de onderkant van de kroon (C) alsmede de 4 periferiepunten van de kroonomtrek (P1-P4). Het resulterende elementaire 3D-boommodel is te zien in Figuur 2.3. De vorm van de kroon is benaderd door ellipsvergelijkingen te gebruiken in zowel het horizontale als het verticale vlak. De horizontale projectie (het periferie vlak) wordt benaderd door de 4 kwarten van de ellips die door de 4 buitenste punten gaat (zie Figuur 2.4). De combinatie van deze vier ellipsen vormt de projectie.

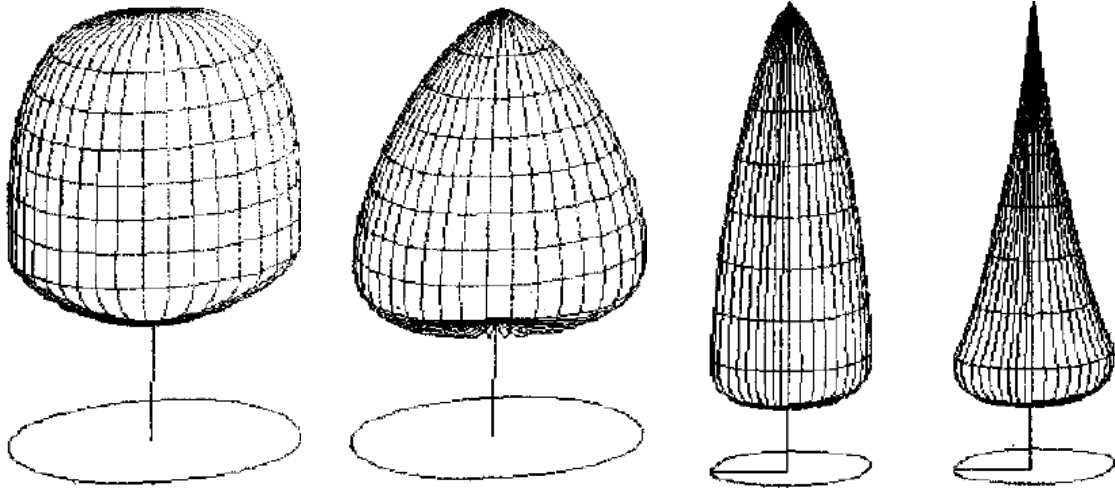


Figuur 2.3 Het 3D boommodel met stambasis B, 1^e vertakking F, kroontop T, kroonbasis C en periferiepunt P (bron: Koop, 1989).



Figuur 2.4 Horizontale projectie van de boomkroon. De periferiepunten (P1-P4) zijn met elkaar verbonden door kwart-ellipsen (bron: Koop, 1989).

De verticale projectie (het transverse vlak) wordt benaderd met soortgelijke vergelijkingen, waarbij de ellipsen worden gebruikt om zowel de top (T) en de basis (C) van de boomkroon te verbinden met elk van de periferiepunten. De ellipsvergelijkingen zijn licht aangepast om de best passende dwarsomtrek van de kroon te verkrijgen. Afhankelijk van de boomsoort zijn de ellipsvergelijkingen bijgesteld om hollere of bollere kroonvormen mogelijk te maken (zie Figuur 2.5).



Figuur 2.5 Twee bomen met dezelfde attribuutwaarden, getoond met twee verschillende kroonvormen (bron: Koop, 1989).

Van CP naar SILVI-STAR

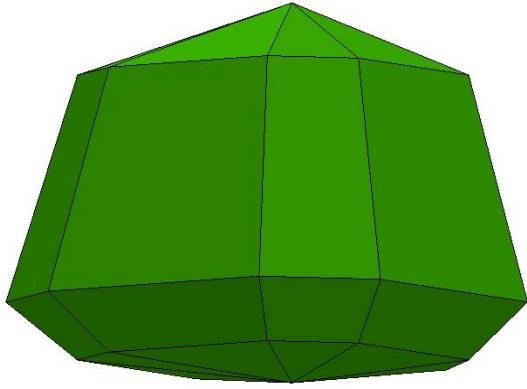
Bij de productie van CP berekent het extractiealgoritme alleen de top van de kroon en de periferiepunten (d.w.z. de kroonprojectie).

Omdat het algoritme rasters gebruikt die alleen informatie over de bovenkant van de kroon geven, is het onmogelijk om daaruit het punt van eerste vertakking en de kroonbasis af te leiden. Daarnaast is de stampositie gebaseerd op een simpele zwaartepuntberekening en daarom niet betrouwbaar. Dit betekent dat de combinatie van het extractiealgoritme met SILVI-STAR alleen een modellering van de bovenkant van de kroon mogelijk maakt.

Niettemin biedt het de mogelijkheid om boomkronen uit CP-bestanden als eenvoudige bomen te visualiseren. Daartoe kan het CP-bestand worden omgezet naar een collectie 3D-objecten volgens het SILVI-STAR-model. Voor dat doel worden de onderstaande attributen toegevoegd:

minX_x	P1x (het x-coördinaat van periferiepunt P1)
minX_y	P1y
maxY_x	P2x
maxY_y	P2y
maxX_x	P3x
maxX_y	P3y
minY_x	P4x
minY_y	P4y
deltaX	lengte van de oost-westas van de projectie polygoon
deltaY	lengte van de noord-zuidas van de projectie polygoon
gemB	gemiddelde as-lengte binnen de projectie polygoon
ShpIndex	hoogte-breedteverhouding van de kroon

Met behulp van deze parameters en de waarde van de boomhoogteparameter kan van een boomkroonprojectie een eenvoudig 3D-boommodel worden gemaakt. Het algoritme hiervoor is aanvankelijk geïmplementeerd in Python door J. Clement. In het najaar van 2012 heeft A. Rafiee van Geodan B.V. op zijn verzoek het programma vertaald naar C# voor een snellere werking. Daarbij wordt voor elke boomkroon een gesloten 3D-vorm gemaakt, bestaande uit 26 knooppunten ('nodes'), met elkaar verbonden door 26 vlakken ('faces').



Figuur 2.6 De stamloze 3D-boomkroon op basis van de boomkroonprojectieparameterwaarden.



Figuur 2.7 Een perspectiefbeeld van een beplantingspatroon, gemaakt in 2012 met ArcScene door H. Kramer (Alterra) op basis van de eenvoudige 3D-kronen.

De resulterende 3D-modellen kunnen met ESRI software worden gevisualiseerd (zie Figuur 2.7). De visualisatie zelf is echter een vervolgtraject dat hier buiten beschouwing blijft.

Naast visualisatie is berekening van het kroonvolume een toepassingsmogelijkheid. Een voorbeeld daarvan is gegeven door Rafiee (Rafiee *et al.*, 2013) waarbij in een deel van Middelburg het volume van gebouwen wordt vergeleken met het volume van boomkronen.

3 Beschrijving van de validatiemethoden

Om een idee te krijgen van de kwaliteit van de Boomkronen Extractie Methode (BEM) moet je het resultaat van BEM valideren. Een dergelijke validatie kan op verschillende manieren. In dit hoofdstuk wordt een beschrijving gegeven van een aantal validatiemethoden die tot nu toe zijn gebruikt om het boomkronenbestand te valideren. Uitgangspunten daarbij waren de volgende toepassingsvragen:

- In hoeverre corresponderen de uit BEM resulterende kroonprojecties met de werkelijke boomlocaties, d.w.z. de geografische stamposities?
- In hoeverre komen de door BEM bepaalde oppervlakte en ligging van de kroonprojecties overeen met die van de werkelijke boomkronen?
- Hoe correct zijn de door BEM in CP geplaatste waarden van de hoogteattributen van de kroonprojectie?

Om deze vragen te kunnen beantwoorden is gekozen voor drie verschillende soorten testgebieden:

- stedelijk gebied
- landelijk gebied
- parkgebied

De combinaties van toepassingsvragen en testgebieden komen in de volgende paragrafen van dit hoofdstuk elk kort aan de orde. Elke paragraaf begint met een onderzoeksvraag die vervolgens per testgebied wordt uitgewerkt en toegelicht.

3.1 Validatiemethoden

3.1.1 Boomlocaties

Vraag: In hoeverre is het BEM-algoritme in staat om uit het AHN2-rasterbestand de stamlocaties van individuele bomen te identificeren?

De validatie voor het stedelijk gebied is gedaan door de geëxtraheerde bomen te vergelijken met metingen die ter plaatse zijn uitgevoerd, waarvoor de gemeentelijke bomendatabase van Amersfoort is gebruikt. Om de gemeentelijke bomendatabase van Amersfoort te verbeteren, is aanvullend beeldmateriaal gebruikt. Voor het deel van de database dat betrekking heeft op het studiegebied zijn bomen bevestigd, verwijderd, verplaatst of toegevoegd aan de database. Bovendien zijn ook bomen op privégrond toegevoegd aan de database. Het resultaat is een bomendatabase voor het testgebied waarbij alle bomen zijn vastgelegd als puntlocatie. Deze puntlocaties zijn vervolgens vergeleken met het boomkronenbestand.

Voor het landelijk testgebied is geen bestaande referentiedataset beschikbaar die gebruikt kan worden om de kwaliteit van de door BEM geëxtraheerde boomkronen te bepalen. Gezien de omvang van het landelijk testgebied en het aantal daar aanwezige bomen was het handmatig creëren van een referentiebestand met boomlocaties, zoals voor het stedelijk testgebied, geen optie.

Voor het parkgebied is op basis van luchtfoto's een referentiemodel vervaardigd van de situatie ten tijde van de inwinning van de AHN2-data. Het model bevat twee soorten objecten: vrijstaande en aaneengesloten boomkronen. Een bestand met stamposities uit 2012, verstrekt door de terreinbeheerder, bleek niet goed te matchen met de beschikbare luchtfoto's. Er is daarom van afgezien om dat bestand bij de validatie te betrekken. Stamlocaties zijn bij de validatie voor het parkgebied niet verder onderzocht.

3.1.2 Ligging en omvang kroonprojecties

Vraag: In hoeverre komen de met BEM geëxtraheerde kroonprojecties overeen met de werkelijke kroonprojectie?

Om deze vraag te beantwoorden is de voor het stedelijk testgebied aangevulde en verbeterde gemeentelijke bomendatabase gebruikt als toets voor de boomlocaties en kroonumfang in CP. Voor beide is de nauwkeurigheid getoetst door na te gaan in welke mate er overeenstemming is met de referentiedataset. Volledigheid en correctheid zijn berekend voor drie verschillende versies van BEM. Daarnaast is er onderscheid gemaakt tussen de bomen op privégrond en die op gemeentegrond. Van de uitkomsten worden ook de RMSE (Root Mean Square Error) en de determinatie coëfficiënt berekend.

Voor het landelijk testgebied is met een semi geautomatiseerde, object gebaseerde benadering een referentiedataset met boomkroonprojecties gemaakt. De gebruikte techniek en databronnen lieten het niet toe om onderscheid te maken tussen individuele bomen. De kwaliteit van het BEM-algoritme is bepaald door de geëxtraheerde boomkroonoppervlakte te vergelijken met de boomkroonoppervlakte van de referentiedataset. Aangezien de vervaardigde referentiedataset geen individuele bomen of boomlocaties bevat, is dit de enige evaluatie die is uitgevoerd voor het landelijk testgebied.

De in het vervaardigde referentiemodel van het parkgebied onderscheiden objectsoorten (vrijstaande en aaneengesloten boomkronen) zijn apart beoordeeld. Voor de aaneengesloten boomkronen is getoetst op de oppervlakte in m² ten opzichte van die objectsoort in het model. Voor de vrijstaande bomen is getoetst op het aantal geëxtraheerde boomkronen ten opzichte van het aantal gemodelleerde kronen. Daarnaast is getoetst hoe goed BEM in het geval van aaneengesloten boomkronen het juiste aantal boomkronen onderscheidt. Voor een drietal in het model onderscheiden objecten is in het veld het aantal stammen geteld.

3.1.3 Boomhoogte

Vraag: In welke mate zijn de door BEM in CP geplaatste hoogtegegevens correct?

Ten behoeve van de validatie in het stedelijk testgebied is een apart bestand gemaakt met hoogteparameters op basis van de AHN2-rasterdata. Daarna zijn de waarden in het referentiebestand vergeleken met de waarden in CP.

De voor het landelijk testgebied vervaardigde referentiedataset bevat geen gegevens over individuele bomen. Validatie van de door BEM geproduceerde boomhoogtes is daarom niet uitgevoerd.

Voor de validatie in het parkgebied is de juistheid van het hoogteattribuut van de met BEM geëxtraheerde kroonprojectie niet getoetst. De uit AHN2 verkregen gridcelwaarden zijn als correct beschouwd.

3.2 Beschrijving testlocaties

3.2.1 Stedelijk gebied

Het stedelijk testgebied bevindt zich in de stad Amersfoort en is ruim 21 ha groot (zie Figuur 3.1). Dit gebied is gekozen, omdat het representatief is voor verschillende soorten groen binnen een stedelijk gebied: het bevat laanbomen, kleine parken en privétuinen. Daarnaast heeft het bedrijf NEO B.V. in het verleden meerdere studies uitgevoerd in dit gebied, waardoor er verschillende aanvullende datasets beschikbaar zijn.



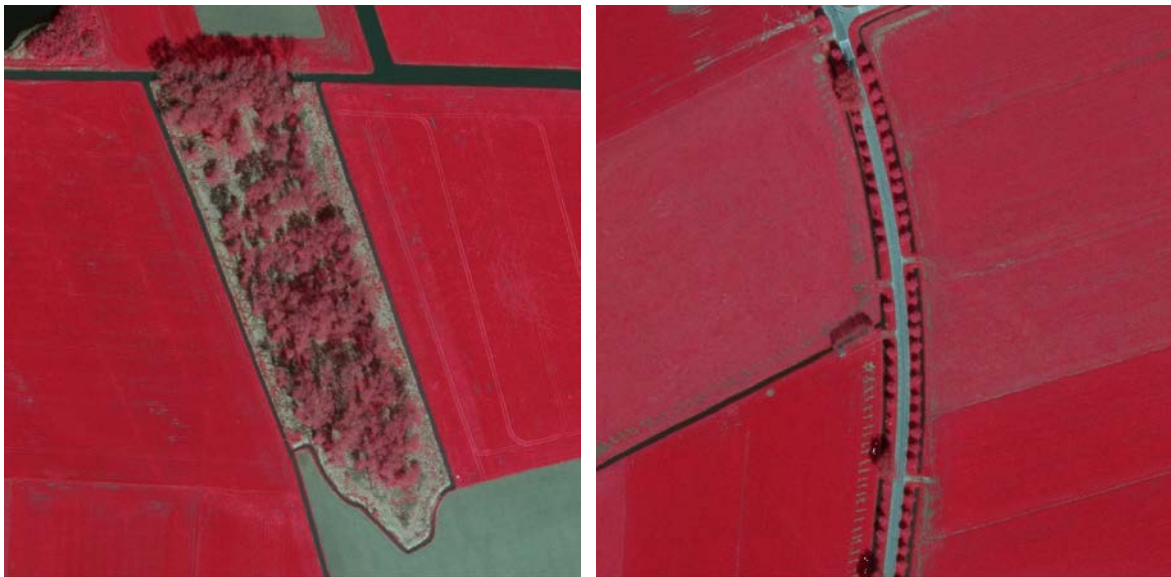
Figuur 3.1 Het stedelijk studiegebied in Amersfoort (ondergrond Google Earth Luchtfoto).

3.2.2 Landelijk gebied

Het landelijk testgebied is gelegen in het 'Land van Heusden en Altena' (zie Figuur 3.2), gelegen ruim 20 km ten noordwesten van de stad 's-Hertogenbosch. Dit testgebied, groot 529 ha, is gekozen omdat er binnen dit gebied bomen in verschillende vormen voorkomen (o.a. bomenrijen, laanbomen, productiebos, bomen gelegen op het erf). Figuur 3.3 laat een aantal verschillende elementen zien die aanwezig zijn in het testgebied.



Figuur 3.2 Landelijk studiegebied in het 'Land van Heusden en Altena' (ondergrond Formosat-2 image, 25-5-2012).



Figuur 3.3 Op de linkerfoto is een productiebos te zien; op de rechterfoto een bomenrij (ondergrond: infrarood luchtfoto uit 2008).

3.2.3 Parkgebied

3.2.3.1 Begrenzing van het validatiegebied

Het validatiegebied 'Campus' ligt direct ten noorden van de bebouwde kom in Wageningen. Het is een deel van het Born-complex van Wageningen UR. Het heeft een oppervlakte van 75,2 ha.



Figuur 3.4 Ligging Campus. Kaart: Kadaster Top10NL, versie sept. 2010.

Door diverse sloop- en bouwactiviteiten en opruimen van beplanting sinds voorjaar 2010 zijn delen van de campus letterlijk onherkenbaar veranderd. De uiteindelijke begrenzing van het validatiegebied is voornamelijk bepaald door de mogelijkheid om begin april 2013 nog te herkennen wat er ten tijde van de AHN-opname (februari 2010) was.



Figuur 3.5 Definitieve begrenzing van het validatiegebied. Oppervlakte: 54.2 ha, waarvan in februari 2010 bebouwd: 13%. Luchtfoto: Kadaster/Cyclomedia 2010 40 cm RGB.

3.2.3.2 Aanwezige beplanting

Het validatiegebied is rijk aan opgaande beplanting. Aanwezig zijn solitaire en laanbomen van diverse leeftijden, soorten en afmetingen. Er is een bosgebied en er zijn boomgroepen met en zonder ondergroei, solitaire bomen en windsingels. De onderstaande foto's geven een beeld van de situatie. De foto's dateren van april 2013.



Figuur 3.6 Links een houtwal in de noordoosthoek. Rechts: laanbeplanting, Droevendaalsesteeg.



Figuur 3.7 Links: bos met rododendrons, noordzijde campus. Rechts: hagen en laanbeplanting aan De Born, het oudste deel van de campus.

3.3 Beschrijving veldmetingen/referentiemodel

3.3.1 Stedelijk gebied

3.3.1.1 Metingen ter plaatse

De validatie van BEM wordt gedaan door de geëxtraheerde bomen te vergelijken met metingen die ter plaatse zijn uitgevoerd. De gemeentelijke bomendatabase van Amersfoort is in dit onderzoek gebruikt als de meting ter plaatse. Daarnaast is aanvullend beeldmateriaal gebruikt om de gemeentelijke bomendatabase van Amersfoort te verbeteren. Dit heeft er onder andere toe geleid dat er bomen zijn bevestigd, verwijderd, verplaatst of zijn toegevoegd aan de database. Naast het editen van de gemeentelijke bomendatabase zijn ook bomen die op privégrond staan toegevoegd aan de database. Het resultaat is een bomendatabase voor het testgebied, waarbij alle bomen zijn vastgelegd als puntlocatie. Een overzicht van de voor dat werk gebruikte datasets geeft Tabel 3.3.1.

Tabel 3.3.1

Overzicht van de gebruikte datasets voor de evaluatie van het stedelijk testgebied

Naam	Data type	Resolutie	Datum
Geëxtraheerde boomkronen (V5;V6;V7)	Polygoon	Niet bekend	Winter 2009
Luchtfoto Amersfoort	Raster	5 cm	Winter 2009
Luchtfoto Amersfoort (CIR)	Raster	25 cm	2011
AHN2 DSM	Raster	50 cm	Winter 2009
AHN2 DTM	Raster	50 cm	Winter 2009
Bomendatabase van de gemeente Amersfoort	Punten	Niet bekend	2009
Kadastrale percelen	Polygoon	Niet bekend	2012
Top10Smart (gebaseerd op TOP10NL)	Raster	2,5 m	2010

Naast het vastleggen van de puntlocaties zijn ook de boomkronen bij benadering vastgesteld en vastgelegd. Het laatste is nodig om de werkelijke boomkroonoppervlakte te kunnen vergelijken met de oppervlakte die met BEM is geëxtraheerd. Aangezien er geen externe bron beschikbaar is met de boomkroonoppervlakte, is er een referentiedataset gecreëerd op basis van een luchtfoto. Voor elke puntlocatie is handmatig een cirkel getrokken om de boomkroonoppervlakte te benaderen.



Figuur 3.8 Handmatig gecreëerde boomkronen gelegen in studiegebied van de gemeente Amersfoort (ondergrond: Luchtfoto).

Omdat V7 van het algoritme alleen bomen extraheert met een hoogte van > 4 m en een boomkroonoppervlakte van > 5 m², is de referentiedatabase voor deze versie hierop aangepast. Het aantal veldmetingen gelegen op zowel privé- als gemeentegrond wordt getoond in Tabel 3.3.2.

Tabel 3.3.2

Aantal bomen vastgelegd in de referentiedatabase en het aantal bomen in de referentiedatabase conform de voorwaarden van V7 van BEM.

Gemeente	Privégrond	Gemeentegrond	Totaal
Aantal bomen referentiedatabase totaal	283	420	703
Aantal bomen referentiedatabase o.b.v. V7	147	234	381

3.3.1.2 Gemeente- en privé bomen

Binnen het stedelijke studiegebied wordt onderscheid gemaakt tussen bomen die staan op gemeente grond en bomen die staan op privégrond. Voor beide type bomen wordt BEM getest. De reden hiervoor is tweeledig: in de eerste plaats zijn gemeenten voornamelijk geïnteresseerd in bomen die staan op gemeentegrond. Aangezien gemeenten gezien worden als een potentiële klant voor informatie over boomkronen, is het belangrijk om de nauwkeurigheid van het boomextractiealgoritme in beeld te hebben voor gemeentebomen.

In de tweede plaats is de verwachting dat de nauwkeurigheid van het boomextractiealgoritme verschilt voor bomen gelegen op gemeentegrond en bomen op privégrond. De reden hiervoor is dat over het algemeen bomen op privégrond lastiger te detecteren zijn dan bomen op gemeentegrond. Dit komt doordat privé bomen vaak onderdeel zijn van tuinen, en omringd worden door andere bomen/vegetatie. Gemeentebomen zijn over het algemeen groter en zijn vaker solitair, waardoor ze makkelijker herkend worden.



Figuur 3.9 Het testgebied wordt in deze foto afgebakend met een gele lijn; de oranje vlakken geven de gemeentegrond aan en groene vlakken geven de privégrond aan.

Om onderscheid te kunnen maken tussen bomen die staan op gemeentegrond en bomen die staan op privégrond, is gebruikgemaakt van kadastrale gegevens (www.kadaster.nl). Deze gegevens laten op basis van polygonen zien welke grond toebehoort aan de gemeente en welke grond gezien moet worden als privégrond (zie Figuur 3.9). Zowel de referentiedataset als de bomen die zijn geëxtraheerd met BEM zijn op basis van de kadastrale gegevens opgedeeld. Voor de referentiedataset is dit gedaan op basis van de stamlocatie, voor de geëxtraheerde boomkronen is dit gebeurd op basis van de BM_X en BM_Y attribuutwaarden.

3.3.1.3 Boomlocaties

Om te bepalen hoe goed het boomextractiealgoritme in staat is om individuele bomen te detecteren, worden alle boomlocaties die zijn bepaald door het algoritme vergeleken met de boomlocaties die in het veld zijn gemeten. Vier mogelijke situaties zijn gedefinieerd (zie ook Figuur 3.10):

1. **Werkelijk positieve kroonprojecties (WPK):** als een boomkroonprojectie een (of meerdere) veldmeting bevat.
2. **Werkelijk positieve veldmeting (WPv):** als een veldmeting zich binnen een kroonprojectie bevindt.
3. **Fout negatief (FN):** een in het veld gemeten stamlocatie bevindt zich buiten de geëxtraheerde boomkroonprojecties.
4. **Fout positief (FP):** een geëxtraheerde boomkroonprojectie bevat geen veldmeting.



Figuur 3.10 Op de linkerfoto wordt een werkelijk positieve kroonprojectie weergegeven (de rode punt representeert de veldmeting); op de middelste foto twee fout negatieve situaties en op de rechterfoto een fout positieve kroonprojectie (weergegeven in blauw).

Er moet bij deze methode rekening worden gehouden met het feit dat puntlocaties (de veldmetingen) worden vergeleken met oppervlaktes (de geëxtraheerde boomkronen). Als gevolg hiervan kan een enkele geëxtraheerde boomkroonprojectie meerdere veldmetingen bevatten. Hierdoor komt het aantal werkelijk positieve boomkroonprojecties niet overeen met het aantal werkelijk positieve veldmetingen. Hoe deze vier situaties zich verhouden tot elkaar is weergegeven in onderstaande formules:

1. Aantal veldmetingen in referentiebestand = WPv + FN
2. Aantal geëxtraheerde boomkroonprojecties = WPK + FP

De prestaties van het extractiealgoritme worden weergegeven aan de hand van de termen 'volledigheid' en 'correctheid'.

De volledigheid geeft aan in hoeverre het boomextractiealgoritme in staat is om alle individuele bomen die zich bevinden binnen het studiegebied ook daadwerkelijk te extraheren. De volledigheid wordt weergegeven als het percentage werkelijk positieve veldmetingen uit het referentiebestand (WPv), oftewel:

$$\text{Volledigheid} = \text{WPv} / \text{aantal veldmetingen in referentiebestand}$$

De correctheid geeft aan in hoeverre de door het algoritme geëxtraheerde bomen ook in werkelijkheid bomen zijn. Dit wordt weergegeven als het percentage werkelijk positieve kroonprojecties (WPK), oftewel:

$$\text{Correctheid} = \text{WPK} / \text{aantal geëxtraheerde boomkroonprojecties}$$

3.3.1.4 Boomkroonoppervlakte

Om te bepalen hoe goed het boomextractiealgoritme in staat is om de boomkroonoppervlakte te bepalen, worden twee evaluaties uitgevoerd. Als eerste wordt de overlappende oppervlakte tussen de geëxtraheerde boomkronen en de referentieboomkronen en het bijbehorende verschil bepaald. Evenals bij de boomlocaties worden weer drie situaties onderscheiden:

1. **Werkelijk positief:** de oppervlakte die overlapt tussen de geëxtraheerde boomkroonoppervlakte en de referentieoppervlakte;
2. **Fout negatief:** de oppervlakte van de referentiekronen die niet overlapt met de geëxtraheerde boomkronen;
3. **Fout negatief:** de oppervlakte van de geëxtraheerde boomkronen die niet overlapt met de referentiekronen.



Figuur 3.11 Op de linkerfoto wordt de geëxtraheerde boomkroonprojectie (groen) vergeleken met de referentiedata (rood); op de rechterfoto wordt dezelfde vergelijking getoond, alleen dan voor meerdere bomen.

Net zoals bij de boomlocatie wordt bij de boomkroonoppervlakte ook gekeken naar de volledigheid en correctheid van het boomextractiealgoritme.

In de tweede plaats wordt een meer specifieke evaluatie uitgevoerd van de boomkroonprojectie. In deze analyse wordt de boomkroonoppervlakte van de correct gedetecteerde bomen (d.w.z. als de locatie werkelijk positief is) vergeleken met de boomkroonoppervlakte van de referentiedataset (zie Figuur 3.11). Dit geeft een beeld van hoe goed het boomextractiealgoritme in staat is om voor een gedetecteerde boom de boomkroonprojectie te modelleren. Zowel de root-mean-square error (RMSE) als de determinatiecoëfficiënt (R^2)⁹ van het boomextractiealgoritme worden voor dit doel berekend. Het komt soms voor dat het boomextractiealgoritme een boomkroon genereert die overlapt met meerdere bomen uit de referentiedataset. Als dit zich voordoet, worden de boomkronen uit de referentiedataset samengevoegd tot één boomkroon (zie Figuur 3.5). De evaluatie is alleen uitgevoerd voor bomen gelegen op gemeentegrond.

⁹ R^2 is een maat die informatie geeft over de mate waarin een model de werkelijke data benadert. Indien alle voorspelde waarden overeenstemmen met de werkelijke waarden, dan is $R^2 = 1$. Merk op dat een (perfect) verband niets zegt over de causaliteit in de data.

3.3.1.5 Boomhoogte

Naast de kwaliteit van de locatie en de boomkroonoppervlakte is ook gekeken naar de kwaliteit van de geëxtraheerde boomhoogte. In het CP-bestand hebben alle boomkronen drie attributen die iets zeggen over de boomhoogte:

- **GEM_MV**: Gemiddelde hoogte van het maaiveld beneden de boomkroon (AHN2 DTM);
- **BOOM_HOOGT**: De maximale genormaliseerde hoogte (delta_h_raster) binnen de boomkroonprojectie;
- **GEM_HOOGTE**: De gemiddelde genormaliseerde hoogte (delta_h_raster) binnen de boomkroonprojectie.

De werkelijke hoogte van elke parameter wordt bepaald aan de hand van AHN2-rasterdata. Aangezien het boomextractiealgoritme dezelfde dataset gebruikt voor de hoogteberekening, wordt verwacht dat deze vergelijking dezelfde hoogtewaarden zal opleveren.

3.3.2 Landelijk gebied

3.3.2.1 Referentiemodel

In tegenstelling tot het stedelijk testgebied is er geen bestaande referentiedataset voorhanden die gebruikt kan worden om de kwaliteit van de geëxtraheerde bomen te bepalen voor het landelijk gebied. Aangezien het landelijk testgebied groter is en zich er in dit gebied significant meer bomen bevinden dan in het stedelijk testgebied, was het handmatig creëren van een referentiebestand geen optie. In plaats daarvan is ervoor gekozen om een semi-geautomatiseerde, object-gebaseerde benadering te gebruiken. De eerste stap van deze benadering is het segmenteren van het testgebied in objecten. Deze objecten zijn vervolgens handmatig geclassificeerd en geëdit middels een visuele controle op basis van luchtfoto's en AHN2-rasters (zie Tabel 3.3.3). Een nadeel van deze methode en de gebruikte databronnen is dat het erg lastig is om onderscheid te maken tussen individuele bomen. Als gevolg hiervan, en vergelijkbaar met de door het algoritme geëxtraheerde boomkronen, bestaat de referentiedataset alleen uit boomkroonprojecties.

Tabel 3.3.3

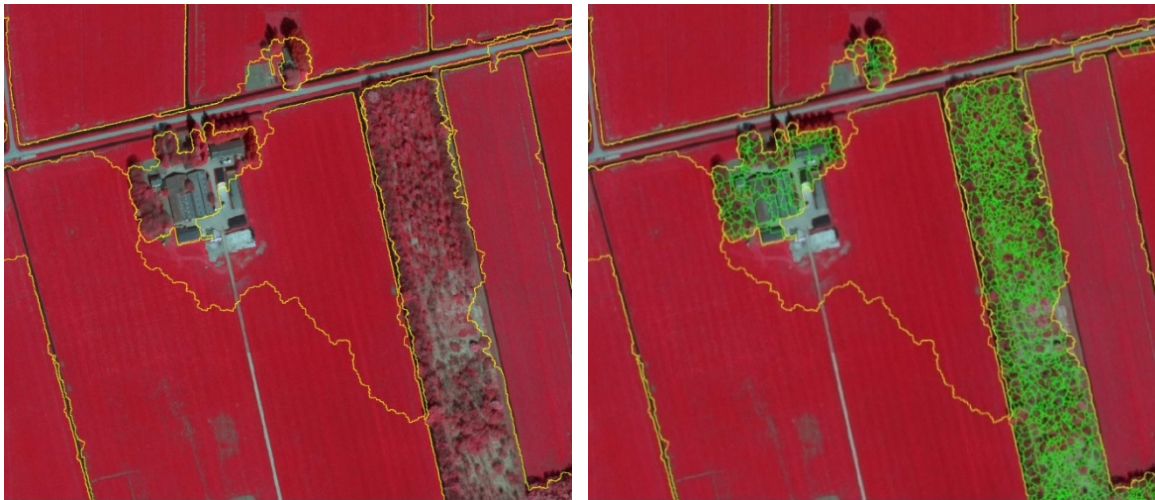
Overzicht van de gebruikte datasets voor de evaluatie van het landelijk testgebied

Naam	Data type	Resolutie	Datum
Luchtfoto (CIR)	Raster	25 cm	2011
AHN2 ongefilterd	Raster	50 cm	2011
Geëxtraheerde boomkronen (V5;V6)	Polygonen	Niet bekend	2011

De eerste fase van de classificatie is uitgevoerd met het softwarepakket eCognition 8.64.¹⁰ Met behulp van eCognition is de ongefilterde rasterdata van AHN2 opgedeeld in objecten met ongeveer hetzelfde oppervlak als een gemiddelde boomkroon. Om dit proces te versnellen wordt er gebruikgemaakt van twee verschillende schaalniveaus. De eerste segmentatie vindt plaats op perceelniveau, waarbij het doel is om zo groot mogelijke, homogene objecten te creëren. De segmentatie is gedaan op basis van hoogtedata, zodat alle objecten min of meer dezelfde gemiddelde hoogte hebben. Alle hoge elementen, zoals bomen, huizen, schuren, silo's etc. worden hierdoor gescheiden van lage elementen als weilanden, akkers, wegen etc. Vervolgens zijn de objecten die bomen bevatten handmatig geclassificeerd, en vervolgens verder gesegmenteerd op boomniveau (zie Figuur 3.12). Op basis van AHN2-data en luchtfoto's zijn de (potentiële) boomobjecten handmatig geclassificeerd als boom of geen-boom. Daarnaast zijn van de gecreëerde dataset de boomkronen daar waar nodig geëdit om de begrenzing van de boomkroon zo veel mogelijk overeen te laten komen met het gebruikte beeldmateriaal.

¹⁰ <http://www.ecognition.com/>

De kleur-infraroodluchtfoto is niet gebruikt voor het belijnen van de boomkroon in verband met het omvalseffect¹¹ van de beelden. Zoals Figuur 3.13 laat zien, maakt dit effect het erg lastig om de boomkronen te belijnen.



Figuur 3.12 Op de linkerfoto is er gesegmenteerd op perceelniveau(schaal 1); op de rechterfoto is een verdere segmentering te zien tot op boomniveau (schaal 2).



Figuur 3.13 Op beide beelden wordt in het groen de boomkroonprojectie getoond die gebaseerd is op AHN2-rasterdata. Op basis van het getoonde beeldmateriaal is het nagenoeg onmogelijk om de individuele bomen te detecteren.

3.3.2.2 Boomlocaties

Dit is niet getoetst voor het landelijk gebied.

¹¹ Loodrecht opgenomen luchtfoto's hebben altijd last van een omvalseffect. Dit wordt veroorzaakt door de beeldhoek van de camera. Van alle objecten die exact onder het midden van de lens staan, wordt de bovenkant op de juiste plaats ten opzichte van het maaiveld weergegeven. Hoe verder objecten naar de rand van het beeld staan, des te meer zal de bovenkant verschuiven ten opzichte van het maaiveld. Hoe groter de beeldhoek van de lens is, hoe groter het omvalseffect zal worden.

3.3.2.3 Boomkroonoppervlakte

Vergelijkbaar met de methode beschreven in de vorige paragraaf wordt de kwaliteit van het boomextractiealgoritme bepaald door de geëxtraheerde boomkroonoppervlakte te vergelijken met de boomkroonoppervlakte van de referentiedataset. Aangezien de referentiedataset geen individuele bomen of boomlocaties bevat, is dit de enige evaluatie die is uitgevoerd voor het landelijk testgebied. Het referentiebestand referentie en de geëxtraheerde boomkroonoppervlakte zijn met elkaar vergeleken om iets te kunnen zeggen over de werkelijk positieve, de foute negatieve en de foute positieve oppervlakte.

3.3.2.4 Boomhoogte

Dit is niet getoetst voor het landelijk gebied.

3.3.3 Parkgebied

3.3.3.1 Model van de situatie in 2010: SVO en PC

De werkelijke situatie ten tijde van de inwinning van de data (februari-april 2010) dient als referentie. Die situatie is in 2013 zo goed mogelijk vastgelegd in de vorm van polygonen die vegetatieobjecten voorstellen op basis van de Topografische Kaart 1:10.000, editie september 2010, en luchtfoto's uit 2010 en 2011. Voor deze handmatige modelconstructie is de ESRI ArcGIS-software gebruikt (versie 10.1).

Het uit de luchtfoto afgeleide model van de opgaande beplanting is vastgelegd in twee bestanden. Deze verdeling is conform de twee subklassen die door het CityGML informatiemodel worden onderscheiden binnen de klasse Vegetatie: 'Solitary Vegetation Object' (SVO, vrijstaande bomen) en 'PlantCover' (PC, meervoudig vegetatiedek). Dit onderscheid is gedigitaliseerd in de vorm van cirkels voor de vrijstaande bomen en in de vorm van grillige polygonen voor het vegetatiedek.

Tabel 3.3.4

Overzicht van de gebruikte datasets voor de evaluatie van het parkgebied

Naam	Data type	Resolutie	Datum
Luchtfoto RGB	Raster	40 cm	September 2010
Luchtfoto RGB	Raster	25 cm	Groei seizoen 2011
Terreinfo's	JPG	n.v.t.	Eind maart 2013
Geëxtraheerde Boomkroon datasets: V5, V9.2	SHP	n.v.t.	V5: 1e helft 2012 V9.2: 27 April 2013
Visual Tree Assessment van Campusterrein. VTA-velden:	Excel	Per boom	Februari 2012
Stamnummer, Boomsoort, Conditie, Stamdiameter, Hoogte, Levensverwachting, Gebreken, Staat van onderhoud, Beheermaatregelen, Inzet materieel, VTA-kwalificatie, Opmerkingen.			

3.3.3.2 Modelconstructie

SVO-model: aan de hand van de op de luchtfoto zichtbare kronen zijn handmatig cirkels geplaatst die de omvang benaderen. De terreinfo's en de luchtfoto van 2011 zijn daarbij als ondersteuning gebruikt.

Plaats en omvang van de individuele boomkronen zijn primair gebaseerd op de gebruikte luchtfoto's. Een handicap daarbij was de omvalling in de luchtfoto's, dus de vertekening door de zijdelingse opname, in combinatie met de geworpen zonneschaduw. Een andere handicap was dat de resolutie van de beschikbare luchtfoto 40 cm was, wat aan de grove kant bleek bij het herkennen van stamposities.

Opvallend was dat de VTA-stamposities in een aantal gevallen meters verwijderd waren van de stampositie zoals die op de luchtfoto's te zien is. De stamposities zoals aangegeven in de VTA-rapportage hebben daarom bij de modelconstructie verder geen rol gespeeld.

PC-model: waar individuele kronen op de luchtfoto niet te onderscheiden waren, maar wel opgaande beplanting, is handmatig met een polygoon de omtrek van het betreffende vegetatieobject weergegeven.

3.3.3.3 Correcties op het SVO-model

Bij de vergelijking van de referentiedata met de geëxtraheerde boomkroonprojecties bleken nog negentien objecten in het model te ontbreken, met een gezamenlijke oppervlakte van 155 m². Om het model zo compleet mogelijk te krijgen, is besloten om deze objecten na controle in de luchtfoto van 2010 alsnog toe te voegen aan het model.



Figuur 3.14 Het model van het validatiegebied in het voorjaar van 2010. SVO: paars, PC: groen.

Een van de objecten in V5 die het probleem van het digitaliseren op basis van luchtfoto's illustreert, is de boom die in het VTA-bestand is aangegeven als boom nummer 747. Dit is een fijnspar aan de zuidwestkant van het Orion-gebouw, dat er overigens in 2010 nog niet stond. De gedigitaliseerde cirkel lag noordelijk van de VTA-positie, de geëxtraheerde boomkroon lag zuidelijk van het VTA-punt. De hart-op-hart afstand tussen de cirkel en de boomkroon is ongeveer 6 m (Figuur 3.15).

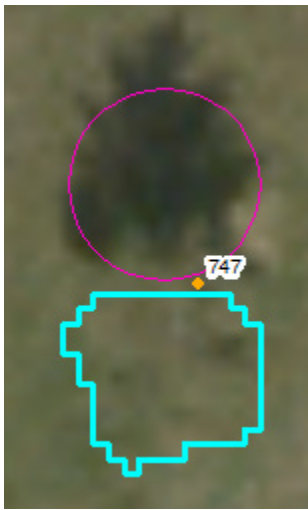
De boomkroon uit AHN2 heeft een omvang die ongeveer overeenkomt met het schaduwbeeld op de luchtfoto: de modelcirkel heeft een oppervlakte van 31,43 m², terwijl die van de AHN2-kroonprojectie 29,75 m² is. Om dubbeling in het modelbestand te voorkomen is de oorspronkelijke modelcirkel eruit verwijderd.

Dit illustreert de moeilijkheid om boomkronen te ontleen aan de luchtfoto's. Het AHN2-bestand is gebaseerd op berekende posities en heeft dus geen last van omvalling. Daarom is de op AHN2 gebaseerde positie betrouwbaarder.

De consequentie van deze werkwijze (aanvulling van referentiemodel met element uit te toetsen product) is wel dat het product van BEM en het referentiemodel niet meer helemaal onvermengd zijn.

Eindresultaat van de controlebewerking was de toevoeging van 19 vegetatieobjecten aan het SVO-bestand van het model. Slechts 1 van de modelcirkels is verwijderd. Het bestand bevat hierna 810 objecten.

Er zijn geen objecten toegevoegd aan het PlantCover-model. Dat bestand bevat 55 objecten met een gezamenlijke oppervlakte van 72.849 m².



Figuur 3.15 Boom 747 in het SVO-model (cirkel) en in het kronenbestand.

BEM produceert boomkroonprojecties, die aan elkaar kunnen grenzen of vrijstaand zijn. Geëxtraheerde kroonprojecties die aan elkaar grenzen, zouden in het ideale geval in het referentiemodel voorkomen als deel van één PC-object. Bij gebruik van de aanname dat elke boomkroon minstens één stam onder zich heeft, zou het aantal boomstammen in het veld (binnen het PC-object) ongeveer overeen moeten komen (gelijk of wat groter) met het aantal geëxtraheerde boomkroonprojecties. Als dat zo is, dan heeft de BEM-methode voor kroonbelijning binnen groepen van aaneengesloten vegetatie enige betrouwbaarheid. Voor drie gemodelleerde PC-objecten is dat getoetst.

Daarvoor is het aantal stammen geteld in 3 PlantCover-objecten van de 2 BEM-versies V5 en V9.2. De PlantCover-objecten 10 en 12 liggen aan de westkant van de Bornsesteeg, tussen Akkermaalsbos en het slootje langs de noordkant van park De Blauwe Bergen (zie ook Figuren 3.16 en 3.17). PlantCover-object 10 bevat (per begin april 2013) in het veld 24 stammen. Object 12 bevatte 54 stammen. Zo te zien, zijn er sinds de AHN-opname in februari 2010 weinig of geen stammen gerooid in de twee gebieden. Wel is sindsdien vrijwel alle ondergroei verwijderd. In het ideale geval worden van de extractiemethode voor de objecten 10 en 12 dus uitkomsten verwacht van 24 respectievelijk 54 kronen.



Figuur 3.16 Zuideinde van PlantCover-object nr.12.



Figuur 3.17 PlantCover-objecten 10, 12.

Het derde PlantCover-object is nr. 17, een groep van 8 hoge oude populieren ten zuiden van het Atlas- gebouw. Van BEM wordt voor dit object een score van 8 kroonprojecties verwacht.

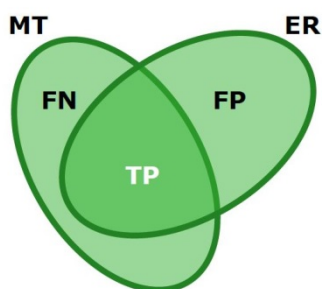


Figuur 3.18 PlantCover-object nr.17.



3.3.3.4 Validatie: volledigheid en correctheid

Bij deze validatie gaat het om het verschil tussen het model van de werkelijke situatie, gemaakt aan de hand van luchtfoto's, en de collectie boomkroonprojecties, geëxtraheerd uit het AHN2-bestand. Hierna zijn deze modellen aangeduid als respectievelijk 'Model Truth' (MT) en 'Extraction Result' (ER). Daarbij is MT dus de norm, ongeacht de eventuele tekortkomingen ervan.



Figuur 3.19 Onderdelen van de kwaliteitsbepaling.

In paragraaf 3.3.1.3 is een vergelijkbaar concept gehanteerd. Wat hier als MT is aangeduid komt overeen met wat daar AVM heet (Aantal veldmetingen in referentiebestand). ER hier correspondeert met AGB (Aantal geëxtraheerde boomkroonprojecties) daar.

Er zijn drie categorieën van overeenkomst tussen model en extractieresultaat te onderscheiden (zie figuur):

- False Negatives (FN, Fout negatief): wat er in ER ontbreekt t.o.v. MT (error of omission)
- False Positives (FP, Fout positief): wat er in ER meer zit dan in MT (error of commission)
- True Positives (TP): wat er zowel in ER als MT zit

In paragraaf 3.3.1.3 worden de concepten FN en FP op vergelijkbare wijze gehanteerd als op deze plaats. Een verschil is dat daar binnen het concept TP nog nader onderscheid wordt gemaakt tussen de concepten WPK (Werkelijk Positieve kroonprojectie, onderdeel van AGB) en WPv (Werkelijk Positieve Veldmeting, onderdeel van AVM).

Bij de benadering die in deze paragraaf is gebruikt (zie Figuur 3.19), zijn de verhoudingen tussen de gebruikte concepten weer te geven door de volgende formules:

$$MT=TP+FN$$

$$ER=TP+FP$$

Met behulp van het bovenstaande kunnen ook de waarden van de volgende parameters worden berekend (Xiao *et al.*, 2012):

- **Volledigheid = $TP/(TP+FN)$**

naarmate FN kleiner is, er dus in ER minder ontbreekt, nadert Volledigheid tot 1.

- **Correctheid = $TP/(TP+FP)$**

naarmate FP kleiner is, dus als ER minder 'valse' objecten bevat, nadert Correctheid tot 1.

- **Kwaliteit = $TP/(TP+FN+FP)$**

naarmate Volledigheid en Correctheid groter zijn, nadert Kwaliteit tot 1.

In de validatie voor het parkgebied zijn de parameterwaarden voor BEM-versies 5 en 9.2 bepaald. Daarbij zijn afzonderlijke berekeningen gemaakt voor het PC-model (aaneengesloten boomkronen) en het SVO-model (solitaire boomkronen). De methode om de getallen te bepalen is in de volgende paragrafen beschreven.

3.3.3.5 Validatie van de aaneengesloten boomkronen (PC)

Omdat het resultaat van boomkroondetectie geen uitspraken doet over stamlocaties of aantallen stammen, noch over aantal stammen per kroon, kan de validatie alleen plaatsvinden aan de hand van vergelijking tussen de gemodelleerde en de geëxtraheerde kroonprojectieoppervlakte. Daarbij gaat het om de ligging van de geëxtraheerde boomkroonprojecties ten opzichte van de PC-elementen in het model. De mate van overeenkomst is uitgedrukt in vierkante meters.

De vergelijking van de oppervlakte van aaneengesloten kronendek in het model (=MT) met de oppervlakte van aaneengesloten objecten in het extractieresultaat (=ER) is met gebruikmaking van ArcGIS als volgt uitgevoerd:

1. Bepaling van totaal aantal eenheden in MT en ER

De waarde van MT is gelijk aan het aantal PC-polygoon in het model.

De waarde van ER is verkregen door het uitvoeren van de ArcGIS Intersect-opdracht (methode: met zwaartepunt binnen source layer (=PlantCover)) om te bepalen hoeveel boomkronen geheel of grotendeels binnen de PC-polygoon liggen.

Daarna nog verwijderen van objecten waarvan de attribuutwaarde voor boomhoogte kleiner is dan 1.30 m¹². De gezamenlijke oppervlakte van de resulterende polygonen geeft de waarde van ER.

2. Bepaling van het aantal ten onrechte gedetecteerde eenheden (FP) en het aantal ontbrekende objecten (FN)
De omvang van FP is bepaald door ArcGIS het symmetrisch verschil te laten bepalen tussen de ER-PC-selectie en de MT-PC.
FP wordt gevormd door het deel van het Symmetric Difference-bestand dat buiten de PC-polygonen ligt. De FN's liggen erbinnen. Binnen het SymDiff-resultaat vormen de polygonen met de attribuutwaarde -1 de False Negatives. De overige objecten behoren tot FP.
3. Bepaling van hoeveelheid correct gedetecteerde eenheden ('TP')
Toepassing van de clipbewerking op de ER-kronen met de PlantCover-polygonen als mal resulteert in de door geëxtraheerde kronen gedekte oppervlakte binnen de PlantCover-polygonen.

De berekende waarden voor MT, ER, FN, FP en TP zijn daarna benut om tot de uitkomst van de validatie te komen via de onderstaande twee stappen.

1. Boekhoudkundige controle:
TP+FN moet MT zijn
TP+FP moet ER zijn
2. Berekening van de validatie uitkomsten voor het PC-deel van het referentiemodel:
Compleetheid = $TP/(TP+FN)$ = waarde
Juistheid = $TP/(TP+FP)$ = waarde
Kwaliteit = $TP/(TP+FN+FP)$ = waarde

3.3.3.6 Validatie van de oppervlakte van vrijstaande boomkronen (SVO)

De vergelijking van het aantal kronen van vrijstaande bomen in het model (=MT) met het aantal kroonprojecties in het extractieresultaat (=ER) is met gebruikmaking van ArcGIS als volgt uitgevoerd:

1. Bepaling van totaal aantal objecten in MT en ER
De waarde van MT is gelijk aan het aantal SVO-polygonen in het model.
De waarde van ER is gelijk aan het totaal van alle geëxtraheerde kronen voor het validatiegebied minus het aantal kronen waarvan het zwaartepunt binnen een PlantCover polygoon ligt.
2. Bepaling van hoeveelheid ten onrechte gedetecteerde objecten (FP)
FP-objecten voldoen niet aan a) het hoogtecriterium van 1.30 m, noch aan b) het liggingscriterium: overlap met een SVO-modelcirkel. Selectie met deze criteria geeft de waarde van FP.
3. Bepaling van hoeveelheid correct gedetecteerde objecten (TP)
De ten behoeve van stap 2 uitgevoerde selectie op geëxtraheerde objecten die géén overlap hebben met een modelcirkel geeft ook de waarde van TP.
4. Bepaling van hoeveelheid ontbrekende objecten (FN)
De waarde van FN wordt bepaald door de aftrekking $FN=MT-TP$.

¹² Model Bomenverordening voor Amsterdam 2013:

'boom: een houtachtig, opgaand gewas, zowel levend als afgestorven, met een omtrek van de stam van minimaal 31 centimeter op 130 centimeter hoogte boven het maaiveld; in geval van meerstammigheid geldt de omtrek van de dikste stam'.

De berekende waarden voor MT, ER, FN, FP en TP zijn daarna benut om tot de uitkomst van de validatie te komen via de onderstaande twee stappen.

1. Boekhoudkundige controle:

TP+FN moet MT zijn

TP+FP moet ER zijn

2. Berekening van de validatie uitkomsten voor het SVO deel van het referentiemodel:

Compleetheid = $TP/(TP+FN)$ = waarde

Juistheid = $TP/(TP+FP)$ = waarde

Kwaliteit = $TP/(TP+FN+FP)$ = waarde

3.3.3.7 Boomhoogte

Voor het parkgebied is niet getoetst op de juistheid van de berekende hoogtewaarden van de kroonprojecties.

4 Resultaten en discussie

4.1 Resultaten kwaliteitscontrole stedelijk gebied

Deze paragrafen geven een overzicht van de validatieresultaten voor de drie soorten gebieden: stedelijk gebied, landelijk gebied en parkgebied. In principe worden voor elk van de gebieden drie onderwerpen behandeld: locatie, oppervlakte en hoogte. Voor het landelijk gebied komt echter alleen oppervlakte aan de orde.

4.1.1 Veldmeting

Tabel 4.1 geeft een overzicht van de aanpassingen die zijn uitgevoerd op de referentiedataset van de gemeente Amersfoort. Naast het aanpassen van de bestaande dataset zijn ook de bomen die staan op privégrond toegevoegd aan de database. Tabel 4.2 geeft een beeld van het aantal bomen dat is verschoven.

Tabel 4.1

Overzicht van de aanpassingen van de referentiedataset van het stedelijk gebied.

Bomen in de referentiedataset	Aantal	Percentage
Totaal aantal bomen	346	
Aantal correcte bomen	129	37,2%
Aantal verplaatste bomen	164	47,4%
Aantal verwijderde bomen	53	15,3%
Aantal toegevoegde bomen (op gemeentegrond)	27	
Aantal toegevoegde bomen (op privégrond)	383	

Tabel 4.2

Overzicht van het aantal verplaatste bomen in meer detail.

Verplaatsing van de bomen (in meters)	Aantal	Percentage van het totaal aantal bomen
<0.5 m	49	14,2%
0.5 – 1 m	27	7,8%
> 1 m	88	25,4%

Gemiddelde verplaatsingsafstand is 1,28 m

4.1.2 Boomlocaties

In de onderstaande tabel wordt de nauwkeurigheid getoond van het boomextractiealgoritme voor de individuele boomlocaties. Zoals al eerder opgemerkt, is het referentiebestand aangepast voor V7 om ook een evaluatie mogelijk te maken van grote bomen.

Tabel 4.3

De nauwkeurigheid van het boomextractie algoritme voor het stedelijk gebied.

Omschrijving	V5	V6	V7	
Gemeente + privé	Referentiedataset	703	703	381
	Geëxtraheerde kroonprojecties	581	886	308
	Volledigheid	48,4%	68,6%	69,8%
	Correctheid	49,2%	43,1%	70,5%
Gemeente	Referentiedataset	283	283	147
	Geëxtraheerde kroonprojecties	234	249	113
	Volledigheid	51,6%	71,4%	61,9%
	Correctheid	59,4%	54,4%	73,5%
Privé	Referentiedataset	420	420	234
	Geëxtraheerde kroonprojecties	347	537	195
	Volledigheid	40,7%	58,3%	68,4%
	Correctheid	38,0%	33,0%	64,6%

4.1.3 Boomkroonoppervlakte

In Tabel 4.4 wordt een overzicht gegeven van de nauwkeurigheid van de geëxtraheerde boomkroonprojecties. De RMSE voor elk van de verschillende versies van het boomextractiealgoritme worden getoond in Tabel 4.5. In Figuur 4.1 wordt een scatterplot getoond met op de x-as de boomkroonoppervlakte uit de referentiedataset en op de y-as de boomkroonoppervlakte van de geëxtraheerde boomkronen. In dezelfde figuur worden voor alle drie versies van het algoritme ook de determinatiecoëfficiënten en de trendlijnen getoond. Deze laatste twee geven aan in hoeverre de geëxtraheerde boomkronen overeenkomen met de boomkronen uit de referentiedataset. Alleen de correct gedetecteerde (werkelijk positieve locatie) gemeentelijke bomen zijn hiervoor gebruikt.

Tabel 4.4

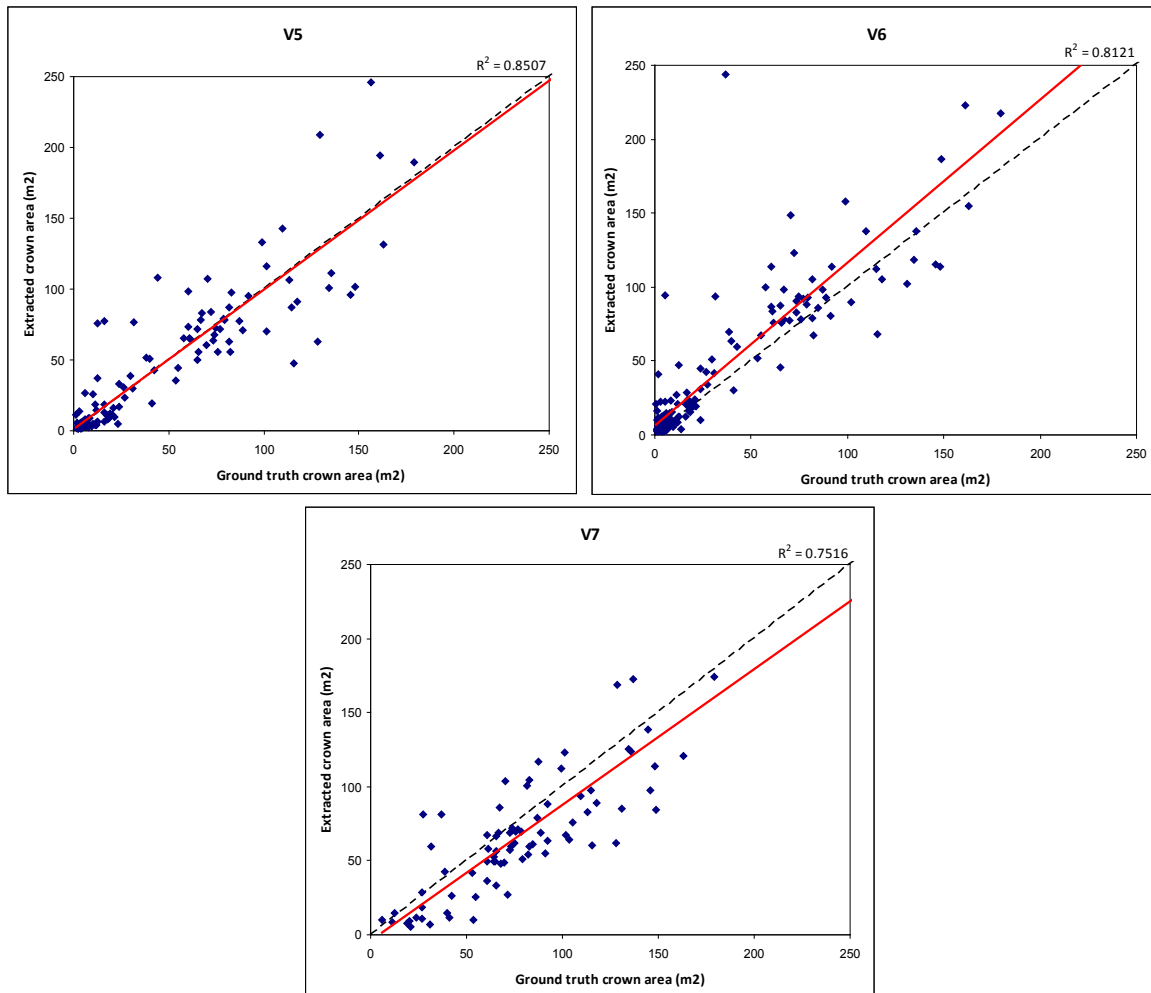
De nauwkeurigheid van het boomextractiealgoritme voor het stedelijk gebied m.b.t. oppervlakte.

Omschrijving	V5	V6	V7	
Gemeente + privé	Referentiedataset	20887,5 m ²	20887,5 m ²	20887,5 m ²
	Geëxtraheerde kroonprojecties	18219,8 m ²	28229,3 m ²	18157,3 m ²
	Volledigheid	61,5%	74,8%	66,5%
	Correctheid	70,5%	55,4%	76,5%
Gemeente	Referentiedataset	12472,9 m ²	12472,9 m ²	12472,9 m ²
	Geëxtraheerde kroonprojecties	9817,5 m ²	16942,0 m ²	11340,0 m ²
	Volledigheid	57,0%	72,5%	69,0%
	Correctheid	72,4%	53,4%	75,9%
Privé	Referentiedataset	8414,7 m ²	8414,7 m ²	8414,7 m ²
	Geëxtraheerde kroonprojecties	8402,3 m ²	11287,3 m ²	6817,3 m ²
	Volledigheid	68,3%	78,2%	62,8%
	Correctheid	68,4%	58,3%	77,6%

Tabel 4.5

Overzicht van de RMSE van zowel de referentiedataset als de geëxtraheerde bomen.

Omschrijving	V5	V6	V7	
Gemeente: werkelijk positieve bomen	Referentiedataset	6234,9 m ²	6950,4 m ²	6944,4 m ²
	Som van de kroonprojecties	6333,5 m ²	8859,0 m ²	6023,8 m ²
	Gemiddelde van de referentiedataset	44,9 m ²	36,6 m ²	83,7 m ²
	Gemiddelde van de kroonprojecties	45,6 m ²	46,6 m ²	72,7 m ²
	RMSE	22,6 m ²	31,5 m ²	28,9 m ²



Figuur 4.1 De drie scatterplots laten voor alle drie de versies van het boomextractiealgoritme zien hoe goed de boomkronen uit de referentiedataset (x-as) overeenkomen met de geëxtraheerde boomkronen (y-as). Tevens wordt in deze scatterplots de trendline (rood) getoond en de determinatiecoëfficiënt (R^2).

4.1.4 Hoogte

Om de hoogte te controleren die berekend is door het boomextractiealgoritme, is er een apart bestand gemaakt met hoogteparameters op basis van de AHN2-rasterdata. Aangezien het boomextractiealgoritme gebruikmaakt van dezelfde databron, zijn er geen grote verschillen gevonden. In Figuur 4.2 en de bijbehorende Tabel 4.7 wordt dit aan de hand van een zestal boomkroonprojecties dit nog eens geïllustreerd. Afgezien van enkele afrondingsverschillen zijn de hoogtewaarden nagenoeg gelijk aan elkaar.



Figuur 4.2 In deze figuur worden in totaal zes unieke bomen en hun BOOM_ID getoond met een luchtfoto als achtergrond. In Tabel 4.4 wordt een overzicht gegeven van de hoogtes die bij deze bomen horen.

Tabel 4.7

Overzicht van de hoogtewaarden van zowel de geëxtraheerde bomen als de referentiedataset.

BOOM_ID	Algoritme		Controle			
	GEM_MV	BOOM_HOOGTE	GEM_HOOGTE	GEM_MV	BOOM_HOOGTE	GEM_HOOGTE
1736	4,33	10,49	3,86	4,33	10,49	3,86
1743	4,23	8,63	2,83	4,22	8,63	2,83
1756	4,42	13,46	3,83	4,41	13,46	3,82
1762	4,64	8,95	2,56	4,64	8,95	2,55
1775	4,69	9,35	2,57	4,69	9,35	2,56
1778	4,74	9,27	2,64	4,74	9,27	2,63

4.2 Resultaten kwaliteitscontrole landelijk gebied

4.2.1 Boomkroonoppervlakte

In Tabel 4.8 wordt een overzicht gegeven van de nauwkeurigheid van de versies V5 en V6 van het boomkroon-extractiealgoritme voor het landelijk gebied.

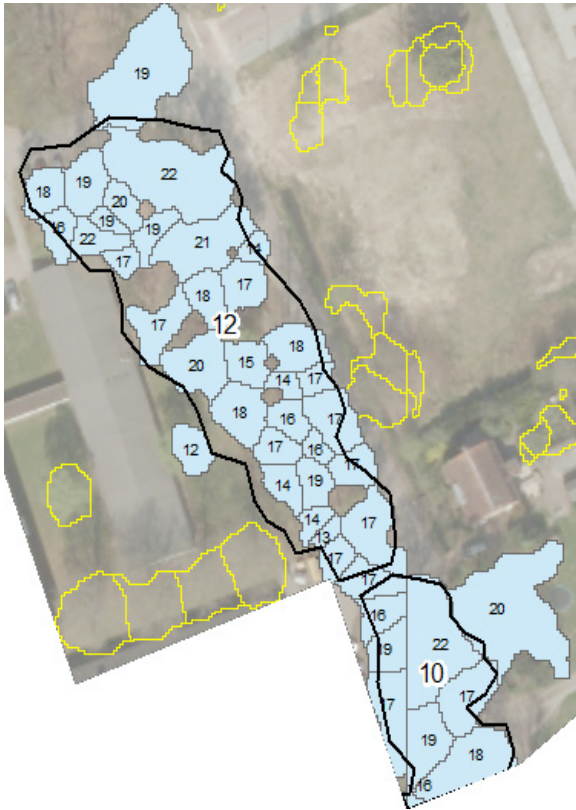
Tabel 4.8

Overzicht van de nauwkeurigheid van het boomkroonoppervlakte gegenereerd door versie V5 en V6 van het boomkroon-extractiealgoritme.

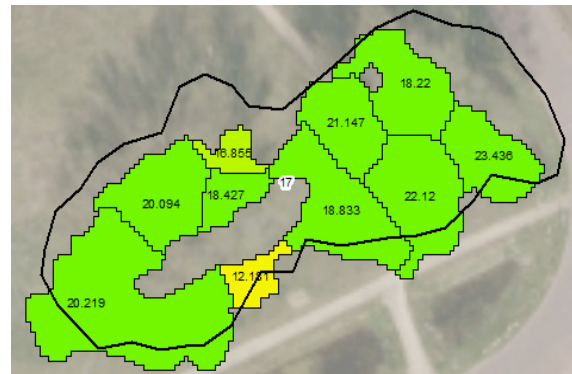
Omschrijving	V5	V6
Referentiedataset boomkroonoppervlakte	143791,6 m ²	143791,6 m ²
Som van de kroonprojecties	137555,5 m ²	172282,5 m ²
Volledigheid	85,2 %	96,5 %
Correctheid	89,0 %	80,5 %

4.3 Resultaten kwaliteitscontrole parkgebied

4.3.1 Boomlocaties



Figuur 4.3 Het ER van V5 voor PC's 10 en 12.



Figuur 4.4 Het ER van V5 voor PC 17.

4.3.1.1 Aantal kroonprojecties

BEM V5

Het extractieresultaat ER van V5 voor de drie proef-PlantCover-objecten was als volgt:

- Voor PC-object 10 en 12 (resp. 24 en 54 stammen op borsthoogte) produceerde V5 één aaneengesloten geheel van 44 kroonprojecties die overlappen met PC, terwijl het aantal ter plekke getelde stammen 78 bedraagt.
- Voor PC-object 17, dat 8 bomen omvat, produceerde V5 het aantal van 10 kronen.

BEM V9.2

Het extractieresultaat ER van V9.2 voor de drie proef-PlantCover-objecten was als volgt:

- Voor PC-object 10 en 12 (resp. 24 en 54 stammen op borsthoogte) produceerde V9.2 één aaneengesloten geheel van 37 kroonprojecties die overlappen met PC, terwijl het aantal stammen 78 bedraagt.
- Voor PC-object 17, dat 8 bomen omvat, produceerde V9.2 het aantal van 6 kronen.

Bij een gesloten kronendak (de PlantCover-objecten) extraheerden V5 én V9.2 dus zowel te weinig (PC objecten 10 en 12) als te veel (PC-object 17) boomkronen.

De uitkomsten van de vergelijking tussen onderscheiden boomkronen en getelde stammen in PlantCover-gebied 10, 12 en 17 zijn weergegeven in de onderstaande tabellen.

Tabel 4.9

De uitkomsten van de vergelijking tussen de aantallen onderscheiden boomkronen en het aantal getelde stammen in PlantCover-gebied 10, 12 en 17.

Omschrijving	PC10	PC12	PC17	MT	ER	FP	FN	TP
Getelde stammen	24	54	8	86				
Kroonprojecties (BEM v5)	PC10+PC12 samen: 44				54	2	32	52
Kroonprojecties (BEM v9.2)	PC10+PC12 samen: 37				43	0	43	43

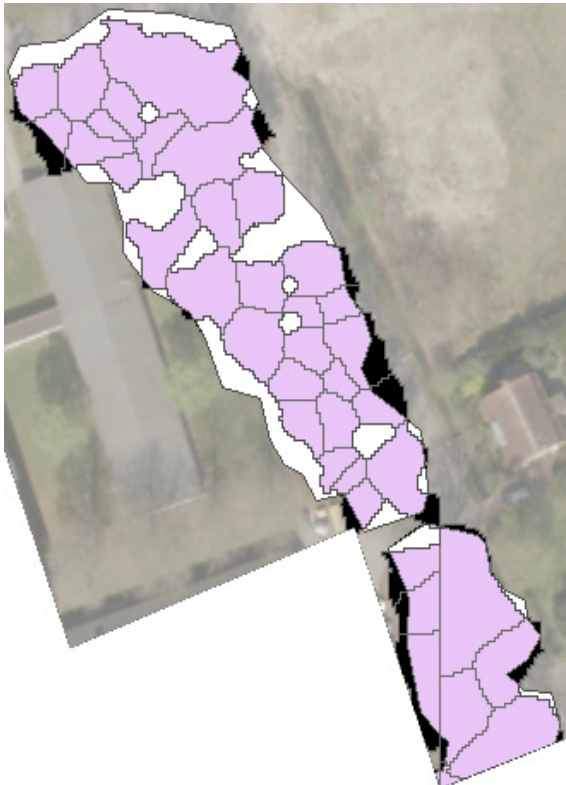
Tabel 4.10

Overzicht van de kwaliteit van versie 5 en 9.2 van BEM op het punt van de relatie tussen aantal stammen en aantal kronen.

Omschrijving	BEM v5	BEM v9.2
Compleetheid	0.62	0.50
Juistheid	0.96	0.95
Kwaliteit	0.60	0.48

4.3.1.2 Kwaliteitsparameters op basis van oppervlakte

Hoewel voor deze test-PC's de waarden voor MT en ER bekend zijn, zijn bepaling van FP en FN niet mogelijk zonder aanvullende informatie over de specifieke relatie tussen individuele stammen en kronen. Daarvoor zouden de stamlocaties bekend moeten zijn. Verder rekenen is daarom alleen mogelijk aan de hand van oppervlakte. De bepaling van de waarde van de kwaliteitsparameters op die basis is voor V5 als onderstaand uitgevoerd voor PC-gebieden 10 en 12.



Figuur 4.5 FN (wit), TP en FP (zwart) op PC 10+12 van BEM V5.

MT: Bepaal de oppervlakte van de PC-polygonen. De oppervlakte van PC's 10+12 is: 2588 m².

ER: Selecteer de kroonprojecties waarvan de zwaartepunten binnen PC 10+12 liggen. Er blijkt: ER= 39 stuks, met oppervlak 2291 m².

TP is het aantal m² van ER dat binnen PC's 10+12 ligt, FN is wat er binnen de PC-grens niet is toegekend. De waarden ervan zijn berekend met behulp van het symmetrisch verschil¹³ (SymDiff) van de kroonprojectiepolygoon binnen de PlantCover-polygoon.

Afleiding FP: Wat binnen en wat buiten de PC-polygoon ligt, onderscheidt zich door de attribuutwaarde van -1 in het SymDiff-bestand, die is toegekend aan de SymDiff-onderdelen binnen de PC1012-polygoon. Dat is FN. Het oppervlak daarvan is 522 m².

De overige SymDiff-onderdelen behoren tot FP.

Controle:

TP+FN moet MT zijn. TP+FN=2588. Correct

TP+FP moet ER zijn. TP+FP=2291. Correct

De resulterende waarden zijn, uitgedrukt in %:

Volledigheid = $TP/(TP+FN)$ = 80%

Correctheid = $TP/(TP+FP)$ = 90%

Kwaliteit = $TP/(TP+FN+FP)$ = 73%

Op dezelfde wijze zijn de parameterwaarden voor BEM V9.2 bepaald. De onderstaande tabel geeft een overzicht van de uitkomsten voor BEM versies 5 en 9.2.

Tabel 4.11

Overzicht van de kwaliteitsaspecten voor geëxtraheerde kroonprojecties ten opzichte van PC 10 en 12 in het referentiemodel op grond van oppervlakte.

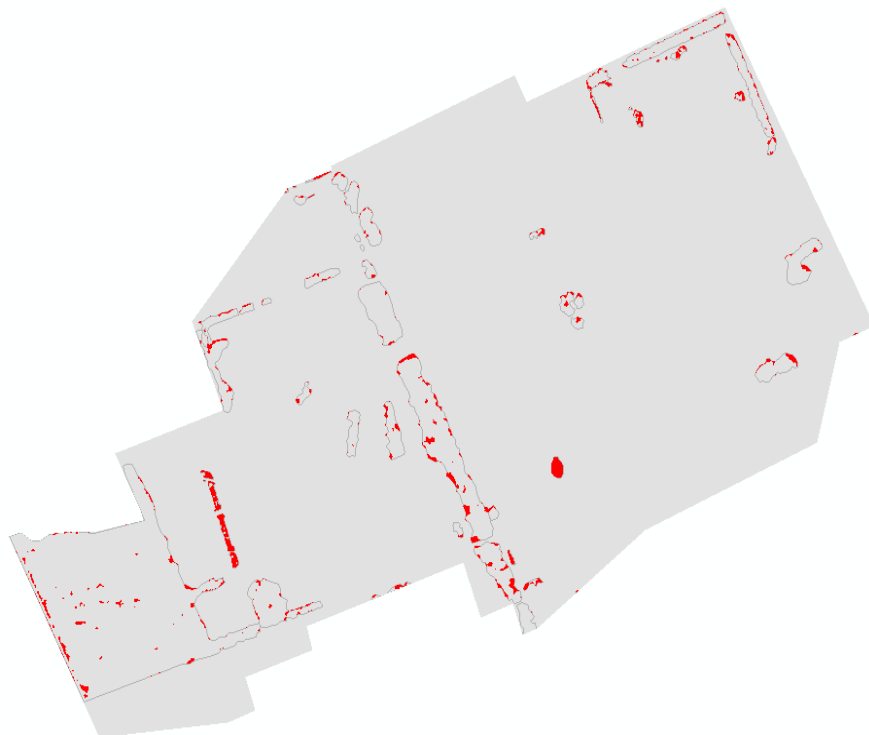
Omschrijving	V5	V9.2
MT Referentiedataset	2588 m ²	2588 m ²
ER Geëxtraheerde kroonprojecties	2291 m ²	2553 m ²
TP	2066 m ²	2256 m ²
FN	522 m ²	332 m ²
FP	225 m ²	296 m ²
Volledigheid	80%	87%
Correctheid	90%	88%
Kwaliteit	73%	78%

Te constateren valt dat voor deze PlantCover-objecten BEM versie 9.2 een 5% betere kwaliteit van het boomkronenbestand oplevert dan versie 5: hij is 7% vollediger, maar 2% minder correct.

4.3.2 Boomkronen

In deze paragraaf worden de resultaten van BEM versie 9.2 geïllustreerd voor zowel PlantCover-objecten als voor Solitaire Vegetatie Objecten. De tabellen daarna tonen de berekende waarden voor de kwaliteitsparameters voor versie 5 en versie 9.2. De onderstaande figuren illustreren waar BEM 9.2 FN en FP oplevert ten opzichte van het model met PlantCover-objecten.

¹³ Verzamelingenleer: symmetrisch verschil = Vereniging minus Doorsnede



Figuur 4.6 Het PC-model: de ligging van V9.2 Fout negatieve gebiedjes.

Tabel 4.12

Overzicht van de kwaliteitsaspecten voor geëxtraheerde kroonprojecties ten opzichte van de PlantCover-elementen in het referentiemodel op grond van oppervlakte.

Omschrijving	V5	V9.2
MT Referentiedataset	72849 m ²	72849 m ²
ER Geëxtraheerde kroonprojecties	64942 m ²	75321 m ²
TP	60583 m ²	67107 m ²
FN	12266 m ²	5742 m ²
FP	4360 m ²	8214 m ²
Volledigheid	83%	92%
Correctheid	93%	89%
Kwaliteit	78%	83%

De toename van de PC-Volledigheid in versie 9.2 ten opzichte van versie 5 is ten koste van de Correctheid gegaan. De Kwaliteit is toegenomen met slechts 5%, terwijl de Volledigheid met 9% toenam. De correctheid daalde echter met 4%.



Figuur 4.7 Het PC-model: de ligging van V9.2 Fout positieve gebiedjes.

Tabel 4.13

Overzicht van de kwaliteitsaspecten voor geëxtraheerde kroonprojecties ten opzichte van het Solitaire Vegetatie Objecten referentiemodel op grond van het deels overlappen met krooncircels in het model.

Omschrijving	V5	V9.2
MT Referentiedataset	810 obj.	810 obj.
ER Geëxtraheerde kroonprojecties	683 obj.	1132 obj.
TP	457 obj.	632 obj.
FN	353 obj.	178 obj.
FP	226 obj.	500 obj.
Volledigheid	56%	78%
Correctheid	67%	56%
Kwaliteit	44%	48%

Ook de toename van de SVO-Volledigheid in versie 9.2 ten opzichte van versie 5 is ten koste van de Correctheid gegaan. De Kwaliteit is toegenomen met slechts 4%, terwijl de Volledigheid met 22% toenam. De correctheid daalde echter met 11%.

De onderstaande figuur illustreert een aantal Fout positieve uitkomsten in het BEM 9.2 resultaat (ER) ten opzichte van het Solitaire Vegetatie Objecten model (MT).



Figuur 4.8 Een voorbeeld van Fout-positieve objecten, gegenereerd door BEM 9.2 ten opzichte van het SVO-model. In zwart vermoedelijk lantaarnpalen, in rood dak- en wandeffecten van gebouwen.

4.3.3 Hoogte

Voor het parkgebied is niet getoetst op de juistheid van de berekende hoogtewaarden van de kroonprojecties. Niettemin zijn er hier en daar vreemde waarden aangetroffen, zoals de variatie in het attribuut boomhoogte tussen 1.50 m en 32 m op plaatsen waar in het veld lantaarnpalen staan.

5 Conclusies en aanbevelingen

In dit hoofdstuk worden de resultaten van de validatieprocedure besproken. Eerst komen per gebiedstype de onvolkomenheden en fouten van BEM aan de orde. Steeds komen eerst de extractiefouten aan bod, gevolgd door de nauwkeurigheid.

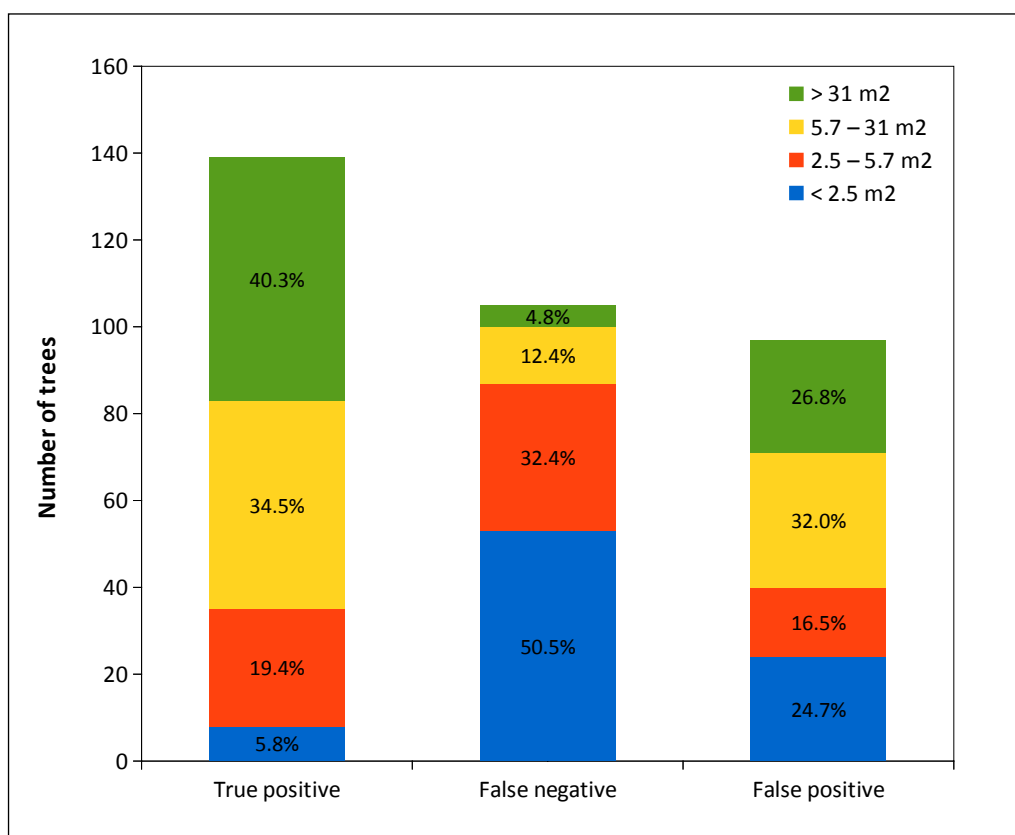
Daarna wordt, met name op grond van de bevindingen voor het stedelijk en het landelijk proefgebied, een aantal verbeteringsvoorstellen gedaan. Vervolgens worden de resultaten van deze voorstellen gepresenteerd.

5.1 Extractiefouten

5.1.1 Stedelijk gebied

5.1.1.1 Identificatie van individuele bomen

Zoals Tabel 4.1 en 4.2 in hoofdstuk 4 laten zien, mist BEM tussen de 30% en 50% van het totaal aantal bomen dat op gemeente- dan wel privégrond staat. Dit is het directe gevolg van de boomhoogte en de kroongrootte. Naarmate de boom kleiner is, wordt het lastiger om deze geautomatiseerd te detecteren. Met name de boomkroongrootte beïnvloedt het detecteren van bomen. Dit wordt geïllustreerd aan de hand van Figuur 5.1. De drie kolommen laten de validatieresultaten zien van BEM v5 voor bomen op gemeentegrond. Op basis van de boomkroonoppervlakte worden de geëxtraheerde bomen onderverdeeld in vier groepen.



Figuur 5.1 De staafdiagrammen laten de resultaten zien van versie V5 van het boomextractie-algoritme voor de bomen op gemeentegrond. De bomen zijn ingedeeld op basis van hun boomkroonoppervlakte.

Elke klasse representeert één kwartiel van de geëxtraheerde boomkroonoppervlakte (dat betekent dat elke klasse hetzelfde aantal geëxtraheerde bomen bevat). Door de data op deze manier te presenteren, kunnen we zien dat de werkelijk positieven voornamelijk bestaan uit grote bomen, terwijl de foute negatieven voornamelijk bestaan uit kleine bomen. Figuur 5.2 toont een voorbeeld van smalle bomen die zijn gemist door het boomextractiealgoritme. Wanneer we het AHN2-DSM bekijken (dit wordt getoond op de rechterafbeelding van Figuur 5.2), dan wordt al snel duidelijk dat dergelijke bomen gemist worden bij het classificeren van de data.



Figuur 5.2 Op de linker luchtfoto worden 4 foute negatieve boomkronen en hun bijbehorende oppervlakte (in m^2) getoond. Op de rechterfiguur worden dezelfde bomen getoond, alleen dan met AHN2-DSM als achtergrond.

Tabel 4.3 (zie hoofdstuk 4) laat zien dat het aantal foute positieven geproduceerd door versies V5 en V6 van het extractiealgoritme rond de 50% ligt. Versie V7 scoort aanzienlijk beter, maar dit is voornamelijk gelegen in het feit dat versie V7 zich focust op grote bomen. En zoals we zojuist hebben gezien, zijn deze makkelijker te classificeren dan kleine bomen. Het merendeel van de foute positieven is het gevolg van ruis in het stedelijk gebied. Straatmeubilair, zoals lantaarnpalen en verkeersborden, worden vaak ten onrechte geclassificeerd als bomen. In sommige gevallen worden ook de randen van gebouwen ten onrechte aangezien voor bomen (zie Figuur 5.3). Deze onterechte classificaties zijn alle het gevolg van het feit dat het boomextractiealgoritme alleen gebruikmaakt van hoogtedata.

Naast ruis worden sommige fout positieven veroorzaakt door AHN2-rastercellen zonder data. Deze AHN2-rastercellen zonder data worden gebruikt door versie V5 en V6 van het boomextractiealgoritme. In bepaalde gevallen leidt dit tot foute positieven. In Figuur 5.3 wordt hier een voorbeeld van gegeven. Versie V7 van het algoritme bewerkt de rastercellen zonder data eerst. Dit leidt tot een lager aantal fout positieven.



Figuur 5.3 Fout positieven: op de linker afbeelding wordt een verkeersbord getoond (groene belijning) die als boom is geclassificeerd, op de middelste afbeelding de rand van een gebouw. Op de rechter afbeelding wordt een door nodata-cellen (in blauw aangegeven) gesplitste boomkroon getoond.

5.1.1.2 Nauwkeurigheid van geëxtraheerde kroonprojecties

Verschillende resultaten voor de locatie en boomkroonoppervlakte

Op grond van de eerder beschreven resultaten kunnen we concluderen dat de nauwkeurigheid voor het in kaart brengen van de boomlocaties lager is dan de nauwkeurigheid voor de boomkroonoppervlakte. Dit verschil wordt voor een gedeelte verklaard door de variabele boomgrootte. Het missen of ten onrechte classificeren van kleine bomen heeft een minder groot effect op de boomkroonoppervlakte dan de nauwkeurigheid van de boomlocatie. Men kan zich voorstellen dat het missen van kleine bomen (met bijvoorbeeld een gemiddelde boomkroonoppervlakte van 5 m²) een minder grote impact heeft op de nauwkeurigheid van de boomkroonoppervlakte dan het missen van een enkele grote boom (met bijvoorbeeld een boomkroonoppervlakte van 100 m²). Echter, als we alleen de boomlocatie beoordelen, dan heeft elke gemiste of onterecht geclassificeerde boom een even groot effect op de nauwkeurigheid van de boomlocatie, ongeacht de grootte van de boom.

De RMSE (de wortel uit de gemiddelde kwadratische afwijking) (Tabel 4.5) en R² (Figuur 4.1) zijn berekend om te illustreren hoe goed de individuele boomkronen gekarteerd kunnen worden door het algoritme. Deze metingen zijn interessant, omdat zij uitgaan van werkelijk positieve boomextracties. Met andere woorden: deze metingen geven aan in hoeverre, wanneer een boom wordt gedetecteerd door het algoritme, de boomkroon overeenkomt met de werkelijkheid.

Hoe kleiner de RMSE-waarde, hoe kleiner het verschil tussen de geëxtraheerde boomkroon en de werkelijke boomkroon. Daar staat tegenover dat de RMSE als een op zichzelf staande waarde nagenoeg betekenisloos is. De RMSE wordt pas interessant wanneer verschillende modellen met elkaar vergeleken worden (in dit geval de verschillende versies van het algoritme). Op grond van de resultaten is de conclusie dat het resultaat van BEM v5 de laagste RMSE-waarde heeft. Dit is een interessant gegeven, want dit betekent feitelijk dat, hoewel versie V5 erg laag scoort op volledigheid, de geëxtraheerde boomkronen het best overeenkomen met de werkelijkheid. Daar staat tegenover dat als gevolg van de lage score op volledigheid de steekproefgrootte eveneens het kleinst is, wat een verklaring kan zijn voor de positieve score.

De R² is een meer algemene statistische waarde. Het kan onder meer gebruikt worden om verschillende modellen met elkaar te vergelijken, maar in tegenstelling tot de RMSE is de R² op zichzelf ook een betekenisvol getal. De R² geeft aan in hoeverre het model de werkelijkheid benadert. De waarde varieert van 0 tot 1, waarbij 0 aangeeft dat er geen overeenkomst is tussen model en werkelijkheid en 1 aangeeft dat er sprake is van een perfecte overeenkomst. Weer zien we dat BEM v5 goed scoort met een R²-score van 0.85. Dit bevestigt het positieve resultaat dat we eerder hebben gevonden op basis van de RMSE.

Het is belangrijk dat men zich realiseert dat er tot op heden relatief weinig energie is gestoken in het detecteren van en onderscheid maken tussen individuele boomkronen. Tot nu toe lag de focus voornamelijk op het correct extraheren van het totale kroonoppervlakte.

5.1.1.3 Hoogte

Er zijn geen grote verschillen gevonden tussen de resultaten van het boomextractiealgoritme en de waarden van het AHN2-bronbestand.

5.1.2 Landelijk gebied

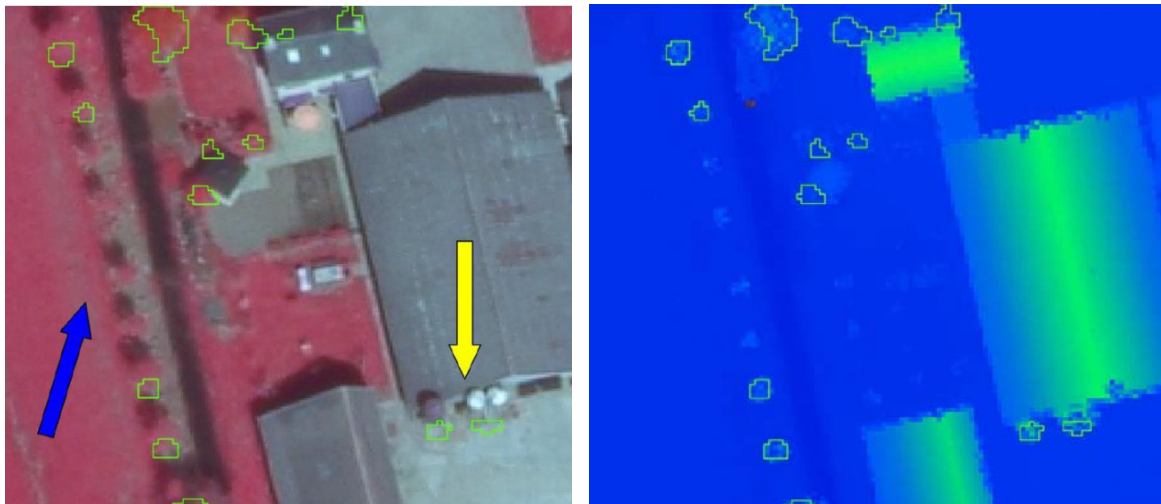
5.1.2.1 Identificatie van individuele bomen

Zoals verwacht presteert het extractiealgoritme aanzienlijk beter voor het landelijk gebied dan voor het stedelijk gebied. Hier is een aantal redenen voor aan te voeren. In de eerste plaats is de hoeveelheid ruis in het landelijk gebied stukken lager dan in het stedelijk gebied. Dit heeft onder meer als gevolg dat het aantal foute positieven voor het landelijk gebied lager is dan voor het stedelijk gebied. In de tweede plaats zien we in het landelijk gebied veel minder individuele bomen en zien we dat bomen veel vaker groeien in groepen. Bomen die dicht bij elkaar zijn gegroepeerd zijn beter te detecteren dan individuele bomen.

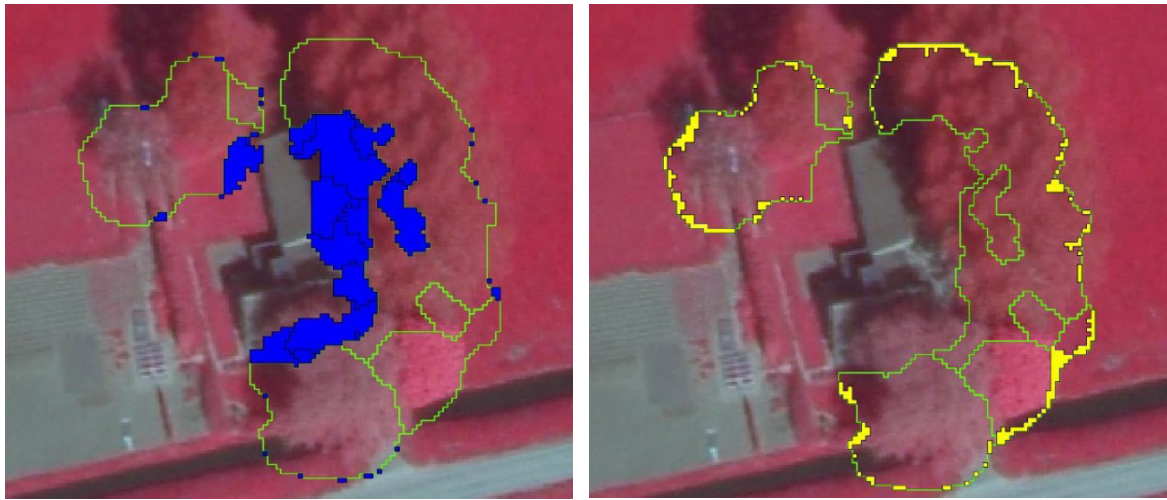
De fouten, ontstaan tijdens het extractieproces, zijn vergelijkbaar met de fouten die we zien voor het stedelijk gebied. Zo worden kleine bomen ook in het landelijk gebied gemist en wordt uiteenlopende bebouwing (graansilo's, schuren etc.) soms aangezien voor boomkronen (zie Figuur 5.4).

5.1.2.2 Nauwkeurigheid van geëxtraheerde kroonprojecties

Onnauwkeurigheden met betrekking tot de kartering van boomkronen hebben het grootste effect op de nauwkeurigheid van de resultaten. Dit betekent dat ondanks het feit dat bomen gevonden en geëxtraheerd worden, in sommige situaties delen van de boomkroon worden gemist dan wel dat bebouwing door het extractiealgoritme wordt aangezien voor boomkronen (zie ook Figuur 5.5).



Figuur 5.4 Op het linker CIR-beeld met boomextractie (in groen) wijst de blauwe pijl naar een aantal gemiste bomen. De gele pijl wijst naar een aantal ontdekt gedetecteerde bomen. Rechts het AHN2-raster.



Figuur 5.5 Op het linkerbeeld in blauw het gedeelte van de boom dat niet is gedetecteerd. Rechts in geel de oppervlakte die onterecht is geclassificeerd als boomkroon.

5.1.3 Parkgebied

In het vorige hoofdstuk is voor het parkgebied de kwaliteit van BEM-versies 5 en 9.2 bepaald. Daarbij werden drie aspecten onderscheiden:

- Kroonbepaling. De mate waarin het aantal onderscheiden boomkronen bij gebieden met aaneengesloten kronen in de buurt komt van het aantal in die gebieden getelde stammen.
- De mate waarin de oppervlakte van aaneengesloten boomkronen overeenkomt met de oppervlakte van die gebieden in het referentiemodel.
- De mate waarin het aantal geëxtraheerde kronen van vrijstaande bomen overeenkomt met het aantal kronen in een zelf vervaardigd referentiemodel.

De uitkomsten daarvan worden hierna samengevat.

5.1.3.1 Identificatie van individuele bomen

Aaneengesloten kronen: bij een gesloten kronendak (de PlantCover-objecten) extraheerden V5 én V9.2 zowel te weinig (PC-objecten 10 en 12) als te veel (PC object 17) boomkronen ten opzichte van het aantal getelde stammen. Duidelijk is dat de belijning van individuele kronen door BEM binnen een aaneengesloten kronendek nog onbevredigend is.

Solitaire kronen: de resultaten van de vergelijking tussen de gemodelleerde en de geëxtraheerde solitaire boomkronen in parkgebied zijn weergegeven in de onderstaande tabel.

Tabel 5.1

Overzicht van de kwaliteit van versie 5 en 9.2 van BEM.

Omschrijving	BEM v5	BEM v9.2
Volledigheid	56%	78%
Correctheid	67%	56%
Kwaliteit	44%	48%

Door BEM v9.2 werden meer kronen geëxtraheerd dan door BEM v5. Maar er werden ook meer niet-kronen geogst, zoals gebouwranden, vlaggenmasten en dergelijke.

De toename van de Volledigheid in versie 9.2 ten opzichte van versie 5 is ten koste van de Correctheid gegaan. Aan de hoge waarde van FP wordt bijgedragen door wand- en dakeffecten bij gebouwen (minstens 40 keer) en door lantaarnpalen (minstens 45 keer).

Onder de niet door BEM geëxtraheerde objecten – dus FN – zijn kleine en/of jonge boompjes en geknotte populieren, die daar een kwart van uitmaken.

5.1.3.2 Nauwkeurigheid van geëxtraheerde kroonprojecties

De resultaten van de vergelijking tussen de gemodelleerde en de geëxtraheerde oppervlakte van gebieden met aaneengesloten kronendek in parkgebied zijn weergegeven in de onderstaande tabel.

Tabel 5.2

Overzicht van de kwaliteit van versie 5 en 9.2 van BEM.

Omschrijving	BEM v5	BEM v9.2
Volledigheid	83%	92%
Correctheid	93%	89%
Kwaliteit	78%	83%

Te zien is een stijging van Volledigheid met 9%. Dat gaat gepaard met een daling van Correctheid van 4%. Het aandeel ruis in het extractieresultaat is dus toegenomen.

5.1.3.3 Hoogte

Zoals eerder aangegeven, is voor het parkgebied niet getoetst op de juistheid van de berekende hoogtewaarden van de kroonprojecties. Op de juistheid van het AHN2-uitgangsbestand werd vertrouwd.

Uit hier en daar aangetroffen vreemde waarden, zoals de variatie in het attribuut boomhoogte tussen 1.50 m en 32 m op plaatsen waar in werkelijkheid lantaarnpalen staan, moet achteraf geconstateerd worden dat er in de verwerking van de AHN-celwaarden tijdens de extractie waarschijnlijk toch fouten zijn geslopen.

5.1.4 Verschillen tussen de versies van BEM

Als we de resultaten van versie V5 en V6 met elkaar vergelijken, valt op dat voor zowel het stedelijk als het landelijk gebied V5 hoger scoort op correctheid dan op volledigheid. Versie V6 van het algoritme laat het omgekeerde resultaat zien. Het lijkt er dus op dat in dit geval het verbeteren van de compleetheid ten koste gaat van de correctheid wat betreft de boomextractie. In een stedelijke omgeving leidt dit tot een toename van de fout positieve boomkroonoppervlakte van 2.6 m² per m² toename van werkelijk positieve boomkroonoppervlakte. Voor het landelijk gebied is dit effect kleiner, met een toename van 1.1 m² van de fout positieve boomkroonoppervlakte per m² toename van de werkelijk positieve boomkroonoppervlakte.

Aangezien versie V7 zich alleen richt op grote bomen is het lastig om de resultaten van V7 met V5 en V6 te vergelijken. Versie V7 laat de beste classificatieresultaten zien. Ondanks het feit dat versie V6 beter scoort op compleetheid, is het gecombineerde resultaat voor versie V7 beter. Dit is waarschijnlijk het directe gevolg van het feit dat in deze versie kleine bomen buiten beschouwing worden gelaten.

Het verschil tussen versie 5 en versie 9.2 gaat mee in de trend die hiervoor al voor versies 6 en 7 is geschetst: toename van de volledigheid betekent afname van de correctheid. Tabel 5.3 vat al de uitkomsten overzichtelijk samen.

Tabel 5.3

Overzicht van de scores voor Volledigheid en Correctheid bij de verschillende versies van BEM voor de verschillende testgebieden.

	V5	V6	V7	V9.2
Stedelijk, Volledigheid kronen	48%	69%	70%	-
Parkgebied, Voll. Sol. kronen	56%			78%
Stedelijk, Correctheid kronen	49%	43%	71%	-
Parkgebied, Corr. Sol. Kronen	67%	-	-	56%
Stedelijk, Volledigheid oppervlakte	62%	75%	67%	-
Landelijk, Volledigheid oppervlakte	85%	97%	-	-
Parkgebied, Voll. Aaneengesloten kronen	83%	-	-	92%
Stedelijk, Correctheid oppervlakte	71%	55%	77%	-
Landelijk, Correctheid oppervlakte	98%	81%	-	-
Parkgebied, Corr. Aaneengesloten kronen	93%	-	-	89%

5.1.5 Experiment: verbeteren van BEM

De validatieresultaten waren aanleiding om enkele verbeteringsvoorstellen te onderzoeken. Deze verbeteringsvoorstellen zijn ook getest. Aangezien de broncode van het algoritme ten tijde van dit onderzoek niet beschikbaar was, was het niet mogelijk om het algoritme zelf te verbeteren. Dit betekende dat er met name gekeken is in hoeverre aanvullende bronnen de kwaliteit van het bestaande algoritme kan vergroten. Daarvoor zijn gebruikt: kleur-infraroodbeelden, topografische bestanden, satellietbeelden en bestaande boomdatabases.

Deze additionele bronnen zijn omgezet in een aantal binaire maskers. Op basis van deze maskers is het mogelijk om het zoekgebied van het extractiealgoritme van tevoren te definiëren. Deze werkwijze kan gecombineerd worden met het bestaande algoritme dat is opgezet door Alterra. Dat extractiealgoritme maakt alleen gebruik van hoogtedata. In aanvulling daarop is op basis van spectrale beelden een vegetatiemasker gecreëerd. Daarnaast is een apart zoekmasker gecreëerd met informatie over bebouwing. Het resultaat van deze aanvullende datasets is getest en gevalideerd voor het stedelijk testgebied in Amersfoort.

5.1.5.1 Vegetatiemasker

Het vegetatiemasker dient om onderscheid te maken tussen gebieden met en gebieden zonder vegetatie. De basis voor dit vegetatiemasker is een vegetatie-index. Daarvoor is gebruikgemaakt van de Normalized Difference Vegetation Index (NDVI). De NDVI maakt gebruik van rode en nabij-infrarode (NIR) golflengten die aangeven of er groene vegetatie aanwezig is of niet. Voor het stedelijk studiegebied, dat gelegen is in Amersfoort, zijn zowel luchtfoto's als satellietbeelden (Formosat-2) beschikbaar, die beide beschikken over rode en nabij-infrarode banden. Een overzicht van de specificaties van het gebruikte beeldmateriaal staat in Tabel 5.4.

Tabel 5.4

Overzicht van de specificaties van de luchtfoto en de Formosat-2 beelden.

Omschrijving	Banden	Resolutie	Cyclus	Datum
Formosat-2	Rood, groen, blauw en NIR	2 m	9 dagen	25-05-2012
Luchtfoto (CIR)	Rood, groen en NIR	0,25 m	jaarlijks	Lente 2011

Het grootste verschil tussen de luchtfoto's en de Formosat-2-beelden is de resolutie van het beeldmateriaal. De luchtfoto's zijn met hun celgrootte van 0.25 m veel gedetailleerder dan de Formosat-2-beelden die een resolutie hebben van 2 m. De luchtfoto's leveren daardoor een beter

vegetatiemasker op dan de Formosat-2-beelden. Daar staat tegenover dat de Formosat-2-beelden vrijelijk beschikbaar zijn, landsdekkend zijn en een hogere temporele resolutie hebben. Om een vergelijking te kunnen maken, is van beide type beelden een vegetatiemasker gecreëerd. Beide maskers zijn gecreëerd door visueel de NDVI-grenswaarde voor vegetatie/geen vegetatie te bepalen. Figuur 5.6 en 5.7 hieronder laten twee voorbeelden zien van het beeldmateriaal en het bijbehorende vegetatiemasker.



Figuur 5.6 Links een Formosat-2-beeld van het studiegebied in Amersfoort (d.d. 25-05-2012) in CIR; rechts het vegetatiemasker (groen is vegetatie).



Figuur 5.7 Links de luchtfoto, rechts het vegetatiemasker.

5.1.5.2 Gebouwenmasker

Het bestaande extractiealgoritme maakte reeds gebruik van een masker voor gebouwen. Dit masker is gebaseerd op Top10smart raster. Dit is een verrasterde versie van de Basisregistratie Topografie (BRT). Met een celgrootte van 2.5 m is de resolutie van Top10smart nogal grof.

Voor dit onderzoek is daarom gebruikgemaakt van gebouwen afkomstig van de Basisregistratie Grootchalige Topografie (BGT). De BGT is een landsdekkend topografisch bestand van Nederland. Alle objecten in de BGT zijn vastgelegd conform de IMGeo-standaarden (Geonovum, 2013). Het bestand is completer en nauwkeuriger dan de BRT. Daarnaast is de BGT een vectordataset. Dat betekent dat het verrasterd kan worden naar een celgrootte naar keuze. In Figuur 5.8 wordt aan de hand van een drietal afbeeldingen getoond hoe Top10smart zich verhoudt tot de BGT. Voor het masker is alleen gebruikgemaakt van bebouwing uit de BGT.



Figuur 5.8 *Verskil tussen Top10smart en de BGT: links de luchtfoto, in het midden Top10Smart en rechts de BGT (vector).*

5.1.5.3 Resultaten van de verbeteringen

Op basis van de twee verschillende vegetatiemaskers (luchtfoto en Formosat-2) is het extractieproces nogmaals twee keer uitgevoerd. BEM v10 maakt gebruik van het vegetatiemasker gebaseerd op de luchtfoto; BEM v11 maakt gebruik van het vegetatiemasker gebaseerd op de Formosat-2-beelden. Het gebouwenmasker is gebruikt voor beide versies. De resultaten staan in Tabel 5.5 en 5.6.

Uit het resultaat van versie V11 blijkt direct dat de Formosat-2-data ongeschikt zijn voor het creëren van een vegetatiemasker. Zowel wat betreft volledigheid als correctheid scoort versie V11 erg laag. Dus ondanks het feit dat de Formosat-2-data een aantal voordelen heeft boven luchtfotodata is het voor dit proces ongeschikt.

BEM v10 scoort aanzienlijk beter dan V11. Over het algemeen kunnen we stellen dat de volledigheid van V10 voor zowel boomlocatie als boomkroonoppervlakte vergelijkbaar is met de resultaten van V5 en V6. Ten aanzien van correctheid scoort versie V10 echter beter. Dit is onder meer het gevolg van de vegetatie- en gebouwenmaskers, die de hoeveelheid ruis (lantaarnpalen, randen van gebouwen etc.) reduceren.

Tabel 5.5

Overzicht van de nauwkeurigheid van boomlocatie, gegenereerd door BEM-versie 10 en 11.

Locatie bomen	Omschrijving	V10	V11
Gemeente + privé	Referentiedataset	703	703
	Geëxtraheerde kroonprojecties	518	330
	Volledigheid	66,0%	35,1%
	Correctheid	69,9%	54,2%
Gemeente	Referentiedataset	283	283
	Geëxtraheerde kroonprojecties	214	109
	Volledigheid	64,7%	21,6%
	Correctheid	79,4%	50,5%
Privé	Referentiedataset	420	420
	Geëxtraheerde kroonprojecties	304	221
	Volledigheid	58,8%	37,9%
	Correctheid	58,2%	52,0%

Tabel 5.6

Overzicht van de nauwkeurigheid van het boomkroonoppervlakte, gegenereerd door versie V10 en V11 van het boomkroon-extractiealgoritme.

Locatie bomen	Omschrijving	V10	V11
Gemeente + privé	Referentiedataset boomkroonoppervlakte	20887,5 m ²	20887,5 m ²
	Geëxtraheerde kroonprojecties	26828,0 m ²	17710,0 m ²
	Volledigheid	77,8%	52,8%
	Correctheid	60,6%	62,3%
Gemeente	Referentiedataset boomkroonoppervlakte	12472,9 m ²	12472,9 m ²
	Geëxtraheerde kroonprojecties	16655,8 m ²	12066,0 m ²
	Volledigheid	76,0%	58,7%
	Correctheid	56,9%	60,7%
Privé	Referentiedataset boomkroonoppervlakte	8414,7 m ²	8414,7 m ²
	Geëxtraheerde kroonprojecties	10172,3 m ²	5644,0 m ²
	Volledigheid	80,5%	44,0%
	Correctheid	66,6%	65,6%

5.2 Conclusies

In de voorgaande hoofdstukken van dit rapport zijn de resultaten besproken van verschillende versies van het boomextractiealgoritme. Elke versie heeft zo zijn eigen voor- en nadelen. In dit hoofdstuk worden de belangrijkste conclusies van de validatie van het landelijk, stedelijk en het parkgebied beschreven.

5.2.1 Conclusies validatie landelijk en stedelijk gebied

De verschillende versies van het boomextractiealgoritme zijn op meerdere manieren getest om een totaaloverzicht te verkrijgen van de kwaliteit ervan. Het is niet eenvoudig om aan te geven welke versie van het algoritme de beste is. Als die zou worden uitgedrukt in één enkele waarde – de berekende Kwaliteit – dan is direct de vraag wat de onderliggende waarden van Volledigheid en Correctheid waren voor de verschillende testgebieden, omdat die immers nogal verschillen. De prestaties van BEM worden uitgedrukt door een combinatie van de nauwkeurigheid van individuele boomlocaties, van boomkroonoppervlakte en van de afbakening van individuele bomen. Welke beoordeling het belangrijkste is, hangt af van wat de uiteindelijke gebruiker belangrijk vindt.

De eerder beschreven assessments van drie verschillende variabelen is alleen uitgevoerd voor het stedelijk gebied. Voor het landelijke gebied is alleen de nauwkeurigheid van de boomkroonoppervlakte getoetst. Om een beter beeld te krijgen van de prestaties van BEM voor het landelijk gebied is meer onderzoek nodig. Dit neemt niet weg dat de nauwkeurigheid van de boomkroonoppervlakte een goede indicator is voor de prestaties van het algoritme voor het landelijk gebied. Daarnaast maakt het een vergelijking tussen het landelijk en stedelijk gebied mogelijk.

Vanuit een commercieel oogpunt hanteren bedrijven zoals NEO vaak een kwaliteitseis van 95% nauwkeurigheid voor hun classificatieproducten. Aangezien deze waarde door geen enkele versie van het algoritme is gerealiseerd, kunnen we stellen dat het algoritme verder verbeterd dient te worden. De resultaten van de verschillende versies van het algoritme laten zien dat alleen het aanpassen van het algoritme weinig impact heeft op de gerealiseerde overalkwaliteit. Sterker nog, het lijkt erop dat wanneer de volledigheid van het resultaat toeneemt, dit ten koste gaat van de correctheid. Waarschijnlijk kunnen we dit toewijzen aan het feit dat de procedure alleen gebruikmaakt van AHN2-rasters als input.

Zoals BEM v10 laat zien, is het mogelijk om op basis van additionele informatie de kwaliteit van de classificering te verbeteren. Met name de volledigheid neemt toe. De correctheid daarentegen blijft nagenoeg gelijk.

5.2.2 Conclusies validatie parkgebied

5.2.2.1 Referentiemodel

Uit de kartering van de BEM-resultaten met het PlantCover-deel van het referentiemodel en de overlap-analyse ervan voor de verschillende versies blijkt dat de kwaliteit van het referentiemodel te wensen overlaat wat betreft het verloop en de ligging van de buitengrens van de PlantCover-polygonen. Er lijkt een systematische fout in te zitten, wat blijkt uit het feit dat de False Negatives vaak voorkomen aan de NO-zijde van objecten, en False Positives aan de ZW-zijde ervan. Mogelijk is dat te wijten aan de resolutie van de gebruikte luchtfoto-opnames en aan de omvalling van de daarop zichtbare objecten. Daarnaast valt op te merken dat er in de PlantCover-objecten van het referentiemodel geen open plekken zijn opgenomen. Het referentiemodel van het Campusterrein bood echter voldoende aanknopingspunten om te kunnen concluderen dat er verschillen zijn tussen de bekeken versies van BEM.

5.2.2.2 BEM

Uit het rekenwerk van de overlap-analyses blijkt dat versie 9.2 van BEM beter presteert dan versie 5. Dit komt op hoofdlijnen overeen met de conclusies van NEO over de prestaties van BEM voor het stukje stedelijk gebied in Amersfoort. Een duidelijk zwak punt van BEM is dat bij aaneengesloten boomkronen BEM geen betrouwbare grenslijn tussen de kronen weet te trekken. Ook kan met BEM geen betrouwbare stampositie per kroon worden berekend, omdat als uitgangsmateriaal alleen de hoogtes van de bovenkant van de boomkronen bekend zijn.

Ten slotte viel op dat in versie 9.2 in tientallen gevallen lantaarnpalen door BEM als boom waren aangemerkt. Zij lijken in het winterseizoen qua hoogte en diameter kennelijk op jonge boompjes, speciaal als ze in gras staan.

5.2.2.3 Uitgangsdata

De problematiek van de extractie van het juiste aantal van de juiste objecten wordt niet alleen veroorzaakt door een tekort aan data over waar vegetatie is, maar ook door de relatieve grofheid van het AHN2-raster. Jonge boompjes, zeker in het winterseizoen, zijn te klein en te dun om in het AHN2-grid terecht te komen.

Er valt te verwachten dat, als boomdetectie in plaats van op 50x50 cm gridcellen zou worden gebaseerd op de puntenwolk die aan AHN2 ten grondslag ligt, de detectie succesvoller (grotere correctheid, grotere volledigheid) zou kunnen zijn.

5.3 Conclusie met betrekking tot de kwaliteit

In de voorgaande hoofdstukken is de kwaliteit van het landsdekkende boomkronenbestand onderzocht en gerapporteerd, gebaseerd op verschillende versies van de Boom Extractie Methode (BEM). Aan de hand van drie verschillende kleine proefgebieden is de kwaliteit van het bestand beschreven. De gebruikte benaderingen zijn echter niet identiek, waardoor de resultaten niet helemaal vergelijkbaar zijn. Daarnaast zijn bij de beschreven validaties de kwaliteitseisen van mogelijke eindgebruikers onderbelicht gebleven. Voor een goede vergelijking van de kwaliteit is het van belang om te komen met een geïntegreerde benadering van de kwaliteit, waarbij de verschillende aspecten van het gebruik en het productieproces een plaats hebben.

In het laatste hoofdstuk van dit rapport wordt een eerste aanzet gegeven voor een integraal kwaliteitsraamwerk. In dit kwaliteitsraamwerk zal zowel de productiekwaliteit als de door de gebruiker gewenste kwaliteit centraal staan.

6 Integraal kwaliteitsraamwerk

6.1 Inleiding

In de voorgaande hoofdstukken werd het boomkronenbestand van Alterra behandeld. De kwaliteit ervan kwam vooral aan de orde in termen van volledigheid en correctheid. Omdat de coöperatie Boomregister.nl van plan is om dat bestand in enige vorm als Open Data te ontsluiten, is de vraag relevant wat een gebruiker van dat bestand aan informatie over de kwaliteit ervan nodig zou hebben. In dit hoofdstuk wordt een voorstel gedaan voor een aanpak om die vraag voor een willekeurig bestand met ruimtelijke data te beantwoorden.

6.1.1 Kwaliteit van Open Data

Met de opkomst van Open Data zien we een toename van het aantal beschikbare gegevenssets. Het merendeel van al deze datasets heeft een ruimtelijke component (Algemene Rekenkamer, 2014). Een voorbeeld van zo'n collectie bestanden is het Nationaal Georegister¹⁴. In dergelijke bestanden is de informatiewaarde gelegen in de koppeling tussen gegevens en locatie. Steeds meer niet-geoschoolde gebruikers maken gebruik van deze data en beschouwen deze data vaak als perfect. Dat maakt de noodzaak voor de explicitering van de kwaliteit van deze open data alleen maar groter (Zrobek & Dawidowicz, 2014).

Het gebruik door een beperkt aantal medewerkers in dezelfde organisatie van een niet-open gegevensbestand dat niet volledig juist is (en dus niet volkomen betrouwbaar), hoeft geen probleem te zijn. Men weet hiervan en zal in voorkomende gevallen bij elkaar te rade gaan of in een bepaalde situatie de gegevens correct zijn en zo niet, wat de feitelijke situatie is. Als diezelfde gegevens ook met gebruikers buiten die organisatie worden gedeeld, dus als het bestand Open Data wordt, is dit gezamenlijke kwaliteitsbewustzijn over de tekortkomingen er niet, net zo min als de mogelijkheid de data te verifiëren bij collega's. Van Oort (2006) wijst erop dat de toename van Open Data gepaard gaat met een grotere afstand tussen makers en gebruikers van datasets, zodat de behoefte aan heldere en adequate metadata groter wordt.

Het moge duidelijk zijn dat bij een grotere afstand tussen makers en gebruikers hogere eisen aan de data moeten worden gesteld. Zo zal deze volledig moeten zijn en de werkelijkheid correct en actueel moeten weergeven. Begrijpelijk voor de gebruiker moeten zijn en niet voor tweeërlei uitleg vatbaar. De gegevens moeten ook consistent zijn in de zin dat een begrip overal hetzelfde betekent. Kortom, de data moet betrouwbaar zijn.

Hoewel men het erover eens is dat kwaliteit van data belangrijk is, zien we tegelijkertijd dat de beschrijving van datakwaliteit vaak slecht vindbaar en moeilijk te doorgronden is voor de (eind)gebruikers. Kwaliteitskenmerken worden in technische termen beschreven in metadatasets die vaak slecht leesbaar zijn of alleen voor ingewijden begrijpelijk. Met andere woorden: er bestaat een gat tussen de (informatie)behoefte van de gebruiker en de producent van gegevenssets (Pôças *et al.*, 2014). De vraag is op welke manier dit gat tussen de producent en de gebruiker overbrugd kan worden.

Deze vraag is niet nieuw. In de afgelopen decennia zijn tal van raamwerken en concepten ontwikkeld die proberen om dit gat te overbruggen (Van Oort, 2006). Vaak zijn deze raamwerken en concepten nogal technisch van aard, wat ze minder geschikt maakt om een vraag/eis/wens van de gebruiker te vertalen naar praktische kwaliteitskenmerken en eisen, die op hun beurt weer bruikbaar zijn voor de

¹⁴ Zie ook <https://www.nationaalgeoregister.nl>

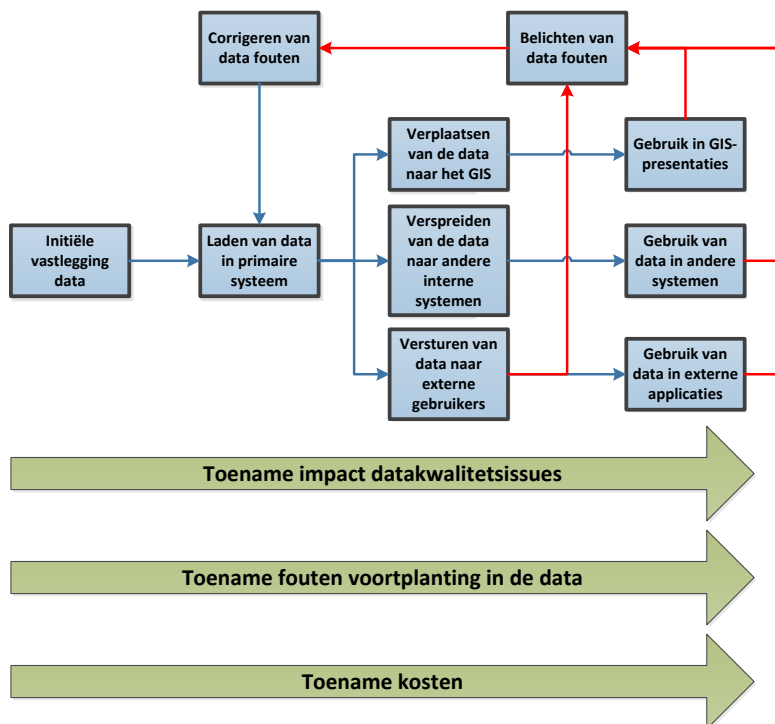
producent. Daarnaast zien we, dat als producenten hun data in termen van bestaande kwaliteitsraamwerken beschrijven, het voor de gebruiker vaak nog niet duidelijk is wat de kwaliteit is van de dataset die hij in een proces wil gebruiken.

In de komende paragrafen wordt voor de data afkomstig van het boomkronenbestand een raamwerk geschetst dat probeert het gat tussen de producent en de gebruiker zo veel mogelijk te dichten. Om een beeld te geven van de kwaliteit die je als producent moet leveren, wordt hier uitgegaan van drie soorten (potentiële) gebruikers van het boomkronenbestand. Vervolgens wordt een overzicht gegeven van de verschillende kwaliteitscriteria die beschreven moeten worden wil de gebruiker een oordeel kunnen vellen over de vraag of de data geschikt zijn voor zijn gebruiksdoel: wat is de 'fitness for use' (Bevelander *et al.*, 2014) van deze data? Idealiter wordt dit laatste natuurlijk in overleg met de gebruiker bepaald. Maar juist omdat dit zo lastig is (de eindgebruiker is bij open data vaak onbekend), wordt hier geprobeerd om een methodische handreiking te doen. Om een beter beeld te krijgen hoe die bij het boomkronenbestand is georganiseerd, zal daar in de volgende paragrafen ook aandacht aan besteed worden. Belangrijk om op te merken is dat hier alleen de open data-variant van het boomkronenbestand besproken wordt. De uiteindelijke eisen rondom de beschrijving van de kenmerken van de commerciële data zullen hetzelfde zijn, maar de wijze van communicatie, het gebruik, gebruikers en inhoudelijke kwaliteit zullen verschillen.

6.2 Kwaliteit in het proces

Zoals in de inleiding al werd betoogd, wordt kwaliteit vaak onderbelicht en wordt het door velen zelfs als lastig ervaren (Devillers *et al.*, 2007). Dit wordt in belangrijke mate veroorzaakt doordat eventuele problemen rondom kwaliteit vaak in een te laat stadium aan het licht komen. Kenmerkend voor kwaliteitsissues die pas laat aan het licht komen, is dat het vaak kostbaar is om deze op te lossen. Het onderstaande diagram illustreert dit nog eens.

Kosten van kwaliteit



Figuur 6.1 Impact van kwaliteit op het proces en de kosten.

In hetzelfde diagram zien we ook dat gaandeweg het proces niet alleen de kosten toenemen, maar ook de impact van data-kwaliteitsissues. Om dit te voorkomen is het dus zaak om vooraf een goed beeld te hebben van de gebruikers en het gebruik van de data. De volgende paragrafen proberen een overzicht te creëren van de gebruikers, het gebruik, maar ook de wijze waarop de interactie met de gebruiker plaatsvindt.

6.2.1 Gebruikers van het boomkronenbestand

De eerste vraag die je je als producent bij de productie van een bestand/gegevensset moet stellen, is: Wie zijn mijn gebruikers? In het geval van het boomkronenbestand moet je, met andere woorden, weten welke partijen en/of organisaties zitten te wachten op een landsdekkend bestand met alle bomen in Nederland. De belangrijkste reden om de gebruikers in beeld te hebben, is omdat deze voor een belangrijk deel bepalen wat de inhoud is of wordt van het boomkronenbestand. Als niemand bijvoorbeeld zit te wachten op de soortnaam van een boom, dan heeft het weinig zin om deze op te nemen. Zeker als daar heel veel moeite voor gedaan moet worden of grote investeringen mee gepaard gaan.

Op dit moment zijn er grofweg drie groepen waar het boomkronenbestand zich op richt. Misschien wel de belangrijkste groep vanuit het perspectief van mogelijke revenuen is de gemeentemarkt. Gemeenten in Nederland geven jaarlijks veel geld uit aan hun boombeheer (Jong *et al.*, 2011). Een andere groep is de rijksoverheid. Zij is bijvoorbeeld geïnteresseerd in het landelijk groen langs openbare wegen of in kleine landschapselementen (Krause *et al.*, 2011). De derde en laatste groep is de onderzoekswereld. Onderzoekers willen het boomkronenbestand bijvoorbeeld gebruiken om te bepalen of er ook sprake is van ontbossing in Nederland. Of om de impact van bomen op het stadsklimaat te bepalen (Klemm *et al.*, 2015; Georgi & Zafiriadis, 2006). In principe zou je ook nog een vierde groep kunnen aanwijzen: particuliere gebruikers. Deze laatste groep wordt in dit document echter buiten beschouwing gelaten.

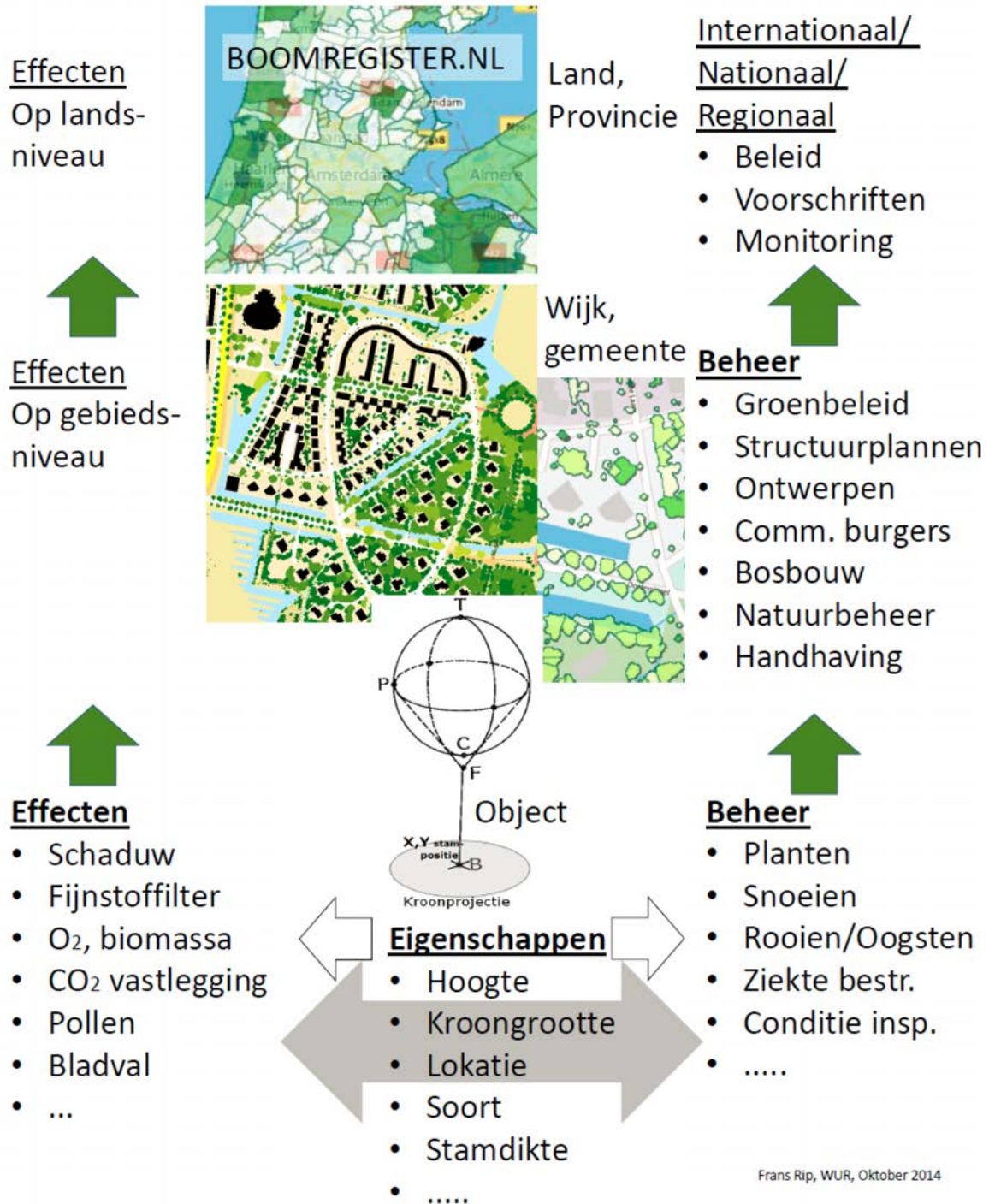
6.2.2 Gebruik van het boomkronenbestand

Naast het onderkennen van de verschillende gebruikers is het ook belangrijk om te weten wat de gebruikers willen doen met de informatie uit het boomkronenbestand. Eerder in dit rapport en in de vorige paragraaf zijn al wat voorbeelden gegeven. In grote lijnen is het gebruik op te splitsen in de volgende onderdelen:

- Inventarisatie: inventarisatie van het aantal bomen binnen een gemeente, de houtopstand in een bepaald gebied of de hoeveelheid groen binnen een gemeente.
- (Risico)analyse: bijvoorbeeld een analyse van de effecten van bomen op het stadsklimaat of het groeipatroon van bomen in relatie tot kabels in de grond.
- Vergelijking: vergelijking met bestaande datasets, bijvoorbeeld de gemeentelijke bomendatabase.
- Combineren: combineren met andere databronnen. Bijvoorbeeld het gebruik van bomen voor visualisatie doeleinden en 3D-weergave.
- Communicatie: het communiceren over specifieke issues of problemen waarbij bomen een rol spelen. Bijvoorbeeld bomen of gebieden waar een bepaalde ziekte of plaag zich voordoet.

Naast de bovenstaande toepassingsmogelijkheden speelt natuurlijk ook de schaal een belangrijke rol bij het gebruik van het boomregister. In onderstaande figuur wordt een overzicht geschetst van zowel de toepassingsmogelijkheden als de schaal.

Gegevens over bomen en hun effecten



Figuur 6.2 Mogelijke toepassingen van het boomregister.

6.2.3 Communiceren over het boomkronenbestand

Naast het gebruik en de gebruikers speelt communicatie ook een belangrijke rol bij het uiteindelijke gebruik en de toepassingsmogelijkheden. Vaak is het gebruik van sommige datasets gekoppeld aan de bekendheid van het bestand en de inhoud. Op dit moment geldt voor het boomkronenbestand dat de communicatie zich voornamelijk beperkt tot het zenden van informatie en het creëren van ruchtbaarheid met betrekking tot het bestaan van het bestand. Hieronder volgt een overzicht van de tot nu toe gebruikte communicatiemiddelen:

- Beurzen (Dag van Openbare Ruimte, Boominfodag)
- Internet (www.boomregister.nl)

-
- Onderzoek
 - Artikelen wetenschappelijk en vakbladen
 - Rapporten

De wijze van communiceren is kenmerkend voor producenten van open data. Een goed voorbeeld hiervan is het Nationaal Georegister. In principe is dit platform er alleen op gericht om informatie te ontsluiten en om bezoekers van de website te wijzen op nieuwe datasets. Er vindt verder eigenlijk geen interactie plaats tussen gebruiker en aanbieder.

6.2.4 Organisatie

Voordat we overgaan tot het beschrijven van de producteisen kijken we eerst naar de wijze waarop het beheer rondom het boomkronenbestand is georganiseerd. De reden dat we dit doen heeft te maken met de eerdere opmerking dat idealiter de gebruiker samen met de producent de 'fitness for use' bepaalt. Net als de communicatie rondom open data, is het opzetten van een organisatie rondom open data eveneens lastig. Wie neem je op in de organisatie? Hoe ga je om met wensen van de gebruikers zonder dat daar een financiële bijdrage tegenover staat? Op dit moment is er geen organisatie of afdeling die in contact staat met de open data-gebruiker van het boomkronenbestand.

6.3 Product- en kwaliteitskenmerken

Naast de organisatorische aspecten is de (potentiële) gebruiker uiteraard ook geïnteresseerd in de kwaliteit van de data zelf. Het eerste aanknopingspunt daarvoor is de zogenaamde metadata van de dataset. Metadata is informatie die karakteristieken van bepaalde informatie omschrijft. Denk daarbij bijvoorbeeld aan de auteur van het bestand, de aanmaakdatum, de bron en de updatefrequentie. Voor geo-informatie maken de meeste producenten gebruik van het Nederlands metadataprofiel versie 1.3.1 dat voor een belangrijk gedeelte gebaseerd is op ISO 19115. Ook voor het boomkronenbestand kan metadata worden opgesteld en beschikbaar worden gesteld aan potentiële gebruikers.

Om als gebruiker dus meer te weten te komen over de kwaliteit van het boomkronenbestand is een logische eerste bron de metadata (zie Bijlage 1). Wat direct opvalt is dat een groot aantal velden van de metadata gaat over technische en procesmatige aspecten van de data. Bijvoorbeeld over de productiedatum, de auteur, het uitleveringsformat, de updatefrequentie etc. Allemaal velden die weliswaar relevant zijn en de gebruiker kunnen helpen bij het bepalen van de fitness for use, maar die verder weinig vertellen over de inhoudelijke kwaliteit van de dataset (Devillers *et al.*, 2005). Dus stel een gemeente wil alle particuliere boombezitters een brief sturen op basis van de data uit het boomkronenbestand. Dan is het de vraag of het boomkronenbestand goed genoeg is om te voorkomen dat de particulieren onterecht aangeschreven worden (commissie) dan wel particulieren niet worden aangeschreven die wel aangeschreven moeten worden (omissie)¹⁵.

Om als gebruiker een oordeel te kunnen vormen over de inhoudelijke kwaliteit van een dataset heb je dus meer informatie nodig dan alleen de metadata. In de afgelopen jaren zijn er tal van artikelen en boeken gepubliceerd over het thema kwaliteit van ruimtelijke data (Devillers *et al.*, 2010). In de literatuur wordt vaak onderscheid gemaakt tussen interne kwaliteit, dat wil zeggen de kwaliteit die de producent nastreeft, en de externe kwaliteit, die ook wel 'fitness for use' wordt genoemd (Pôças *et al.*, 2014). Fitness voor use gaat veel meer in op de kwaliteitseisen die de gebruiker stelt. In de praktijk zijn beide vormen van kwaliteit eigenlijk onlosmakelijk met elkaar verbonden (Devillers *et al.*, 2005) en probeer je als producent altijd in samenspraak met de afnemer tot de gewenste eindkwaliteit te komen. Voor open data ligt dit wat lastiger. Vaak ken je als producent je eindgebruiker niet, weet je

¹⁵ Wat hier uiteraard ook een rol speelt, maar hier buiten beschouwing blijft, is de kwaliteit van andere bestanden. Bij combinatie van databestanden wordt de uiteindelijke kwaliteit sterk bepaald door het bestand met de slechtste kwaliteit. Dus in dit voorbeeld speelt ook de kwaliteit van het adressenbestand en de kwaliteit van de kadastrale gegevens een belangrijke rol.

niet hoe de door jou geproduceerde data gebruikt wordt en welke kwaliteit eigenlijk gewenst is vanuit het perspectief van de eindgebruiker.

Naast de onbekendheid, bij open data, van de eindgebruiker, is er ook de vraag of het mogelijk is om de kwaliteit van de dataset te beschrijven. In welke termen kan dat? Zoals al eerder beschreven, zijn er de afgelopen decennia tal van publicaties verschenen die iets zeggen over kwaliteit. Elke publicatie legt echter weer net even een andere nadruk of gebruikt net weer andere termen (Van Oort, 2006; Meijer & Vullings, 2011; Laufer, 2011). In dit hoofdstuk beperken we ons tot een zestal publicaties die iets zeggen over kwaliteit van ruimtelijke data. Geordend naar jaar van publicatie:

- Aronoff (1989) presenteerde een interpretatie van het concept document USA-SDTS (Chrisman, 1987; Moellering, 1988) vanuit management perspectief.
- USA-SDTS (1992). *The United States of America spatial data transfer standard* bevat een sectie die kwaliteitselementen voor ruimtelijke data beschrijft.
- ICA (1995). Onder auspiciën van de International Cartographic Association, publiceerden Guptill en Morrison (1995) een boek getiteld 'Elements of Spatial Quality'. Dit boek bevat bijdragen van verscheidene auteurs.
- CEN/TC287 (1998). Technisch comité 287 van de Comité Européen de Normalisation (CEN) ontwikkelde een Europese pre-standaard ENV 12656. ISO werkte CEN/TC287 verder uit tot de standaard ISO/TC211.
- ISO/TC211 (1992). Technisch comité 211 van de International Standards Organisation (ISO) heeft een aantal internationale standaarden ontwikkeld voor geografische informatie:
 - ISO 19113:2002(E) – Quality Principles
 - ISO 19114:2002(E) – Quality Evaluation Principles
 - 19115-1:2014(EN) – Metadata
 - ISO 19105:2002(E) – Geographic Information – Conformance Testing
 - ISO 19157:2013(EN) – Geographic information – Data quality
 - Etc.
- Pôças *et al.*, (2014) presenteerden een manier om de fitness for use van ruimtelijke datasets te bepalen in de context van ecologische assessments en monitoring.

Bovenstaande publicaties hebben met elkaar gemeen dat ze kwaliteit van ruimtelijke data in vijf à zeven kenmerken beschrijven: lineage (totstandkoming), positionele nauwkeurigheid, thematische (attribuut) nauwkeurigheid, semantische nauwkeurigheid, compleetheid/volledigheid, logische consistentie en temporele nauwkeurigheid. Een overzicht van welke kenmerken in welke standaard voorkomen staat in het volgende overzicht:

Tabel 6.1

Voorbeelden van kwaliteitselementen afkomstig uit standaarden en van cartografische organisaties (naar Van Oort, 2006).

Kenmerk	CEN ¹	ICA ²	IGN ³	ISO ⁴	SDTS ⁵
Lineage		X	X	X	X
Positionele nauwkeurigheid	X	X	X	X	X
Thematische (attribuut) nauwkeurigheid				X	X
Semantische nauwkeurigheid		X	X	X	X
Volledigheid/compleetheid		X	X	X	X
Logische consistentie		X	X	X	X
Temporele informatie	X	X		X	

¹ (CEN/TC-287, 1994 and 1995), ²(Guptill and Morrison, 1995), ³(IGN, 1997), ⁴(ISO-TC/211, 2003), ⁵(FGDC, 2000)

De bovenstaande tabel laat zien dat standaarden en dataproducenten met name gericht zijn op interne kwaliteit (bijvoorbeeld compleetheid en logische consistentie) en daarnaast dat men vergelijkbare criteria belangrijk vindt. In de praktijk zien we dat de meeste standaarden langzamerhand richting ISO-standaard convergeren (Devillers *et al.*, 2005).

In tegenstelling tot de producent zal de gebruiker veel meer redeneren vanuit het gegeven dat een product een aantal eigenschappen moet hebben om er een bepaalde taak mee te kunnen uitvoeren. Een gemeente die alle particuliere bomenbezitters wil aanschrijven, heeft een bestand nodig waarin individuele bomen zijn vastgelegd en waarbij onderscheid gemaakt wordt tussen particuliere en gemeentelijke bomen (producteigenschappen). Daarnaast moet het voldoen aan een bepaalde kwaliteit om te voorkomen dat mensen ten onrechte níet worden aangeschreven of ten onrechte wél worden aangeschreven (kwaliteitseisen). Voor de gebruiker zijn dus zowel de producteigenschappen als de kwaliteitseisen belangrijke selectiecriteria.

In het onderstaande overzicht wordt de link gelegd tussen producteisen en kwaliteitselementen. De producteisen die in de praktijk het meest voorkomen en waar kwaliteitselementen aan gekoppeld kunnen worden, zijn genoemd. Voor de voorkomende producteisen staan hieronder de kwaliteitselementen die van belang kunnen zijn. Voor meer informatie over de totstandkoming van de onderstaande lijst en de inhoudelijke achtergrond wordt verwezen naar Vullings *et al.* (2015).

0. Objectdefinitie

1. Ieder object moet een bepaald kenmerk hebben* (per attribuut invullen, bv. als in Tabel 6.3)
 - a. Thematische nauwkeurigheid (in %)
 - b. Volledigheid (in %)
 - c. Actualiteit (datum)
 - d. Lineage (vrije tekst)
 - e. Logische consistentie (in %)
2. Publicatie, inwinningsdatum, verversingsfrequentie
 - a. Actualiteit (datum)
 - b. Temporele nauwkeurigheid, geldigheid (afwijking in dagen)
3. Ieder object heeft een geometrie (punt, vlak, lijn en/of raster)
 - a. Logische consistentie (% (aantal objecten juist))
 - b. Omissie/commissie (% (aantal objecten te veel of te weinig))
 - c. Positionele nauwkeurigheid (meters)
4. Brontype (dataset, service)
 - a. Beschikbaarheid (uur/week)
5. Resolutie
 - a. Positionele nauwkeurigheid (meters)
6. Meta-informatie van de dataset moet bepaalde kenmerken hebben
 - a. Volledigheid
7. Compatibiliteit eerdere versies
 - a. Lineage (vrije tekst)
8. Objecten moeten binnen een bepaalde gebied liggen

Om de lezer een idee te geven hoe dit raamwerk in de praktijk kan werken, wordt hieronder een fictieve use-case uitgewerkt voor de open data-variant van het boomkronenbestand. Daarbij wordt uitgegaan van de volgende acht attributen:

Tabel 6.2

De attributen van de geëxtraheerde boomkronen.

Attribuut naam	Beschrijving
BLAD_ID	Nummer van het corresponderende AHN2-blad
BOOM_ID	Uniek boomkroonnummer
BM_X	X-coördinaat van het kroon zwaartepunt (pseudostamlocatie)
BM_Y	Y-coördinaat van het kroon zwaartepunt (pseudostamlocatie)
GEM_MV	Gemiddelde maaiveldhoogte onder de kroon (AHN-2 DTM)
BOOM_HOOGT	Maximale genormaliseerde hoogte (delta_h_raster) binnen kroonprojectie
GEM_HOOGTE	Gemiddelde genormaliseerde hoogte (delta_h_raster) binnen kroonprojectie
SHAPE_AREA	Oppervlak van de kroonprojectie in m ²

Uitwerking van de kwaliteitselementen voor de bovengenoemde kenmerken

Producteis	Kwaliteitselement	Eis + Eenheid	Opmerkingen kwaliteitselement	Opmerkingen producteis
0	Voor elk object in het bestand moet een objectdefinitie aanwezig zijn	Bijvoorbeeld: 'boom: een houtachtig, opgaand gewas zowel levend als afgestorven, met een omtrek van de stam van minimaal 31 cm op 130 cm hoogte boven het maaiveld; in geval van meerstammigheid geldt de omtrek van de dikste stam.'	(Model Bomenverordening Amsterdam 2013)	
1	Ieder object moet een bepaald kenmerk hebben* (per attribuut invullen)	a) Thematische nauwkeurigheid	Ja-%	Criteria: verwijst de aanwezige waarde naar de objecteigenschap? Beantwoorden met % van objecten in bestand waarvoor dat zo is.
		BLAD_ID		NVT
		BOOM_ID	100%	Object polygoon moet boomkroon weergeven, geen andere objecten
		BM_X		NVT
		BM_Y		NVT
		GEM_MV		NVT
		BOOM_HOOGTE		NVT
		GEM_HOOGTE		NVT
		SHAPE_AREA		NVT
		b) Volledigheid	%	
		BLAD_ID	100%	Elk object moet binnen een AHN2 subunit liggen
		BOOM_ID	100%	Elk object moet een ID hebben
		BM_X	100%	X-waarde van kroon-zwaartepunt op maaiveld moet er zijn
		BM_Y	100%	Y-waarde van kroon-zwaartepunt op maaiveld moet er zijn
		GEM_MV	100%	Gemiddelde mv-hoogte onder de kroon moet er zijn
		BOOM_HOOGTE	100%	Tophoogte van de kroon moet er zijn
		GEM_HOOGTE	100%	Gemiddelde hoogte van de kroon moet er zijn
		SHAPE_AREA	100%	Oppervlakte van de kroonprojectie op maaiveld moet er zijn
		c) Actualiteit	Productie -datum	Wanneer is de waarde van het veld bepaald en klopt die waarde?
		BLAD_ID	april 2013	
		BOOM_ID	april 2013	
		BM_X	april 2013	
		BM_Y	april 2013	
		GEM_MV	april 2013	
		BOOM_HOOGTE	april 2013	
		GEM_HOOGTE	april 2013	
		SHAPE_AREA	april 2013	
				De waarde van deze velden is bepaald tijdens de productie van deze versie van het boomkronenbestand (CP)
				* per kenmerk kwaliteitselementen vastleggen. Kan per objectklasse verschillen.
				Afhankelijk van gebruikte criteria voor boom
				Ter bevestiging van herkomst
				Representatief punt per kroon
				Representatief punt per kroon
				Nodig voor berekening objecthoogte
				Hangt samen met maaiveld hoogte onder kroon
				Relevant voor kroonvorm en volume
				Relevant voor kroonvorm en -volume
				Varieert per kaartblad
				Prod. datum CP
				Prod. datum CP
				Prod. datum CP
				Prod. datum CP
				Prod. datum CP
				Prod. datum CP
				Prod. datum CP

Producteis	Kwaliteitselement	Eis + Eenheid	Opmerkingen kwaliteitselement	Opmerkingen producteis
	d) Lineage (totstand-koming)	'vrije tekst'	Moet duidelijk zijn hoe het veld gevuld is (procedure)	
	BLAD_ID			
	BOOM_ID		Tijdens het aanmaken van het bestand worden verschillende bestanden gebruikt (o.a. AHN2 en de BAG). Uit de huidige procesbeschrijvingen is niet af te leiden welke bestanden met welke datum zijn gebruikt.	
	BM_X			
	BM_Y			
	GEM_MV			
	BOOM_HOOGTE			
	GEM_HOOGTE			
	SHAPE_AREA			
	e) Logische consistentie	%	Attribuut- en formaat- validiteit	
	BLAD_ID			
	BOOM_ID			
	BM_X			
	BM_Y			
	GEM_MV			
	BOOM_HOOGTE		Op dit punt is overeenkomst met een informatiemodel voor de thematiek het ijkpunt	
	GEM_HOOGTE			
	SHAPE_AREA			
	Attribuut-definities			
	BLAD_ID		Genummerde onderdelen van een overzichtskaart. In dit geval het blad_id van een AHN2 subunit.	
	BOOM_ID		Gegenereerd administratief uniek nummer, bestaande uit 9 posities.	
	BM_X		Getal dat de geografische positie in Oost-West richting aangeeft in RDNEW.	
	BM_Y		Getal dat de geografische positie in Noord-Zuid richting aangeeft in RDNEW.	
	GEM_MV		Gemiddelde terreinhoogte ten opzichte van het NAP voor het gebied van de kroonprojectie.	
	BOOM_HOOGTE		Boomhoogte is de afstand tussen het maaiveld en de boomtop in meters.	
	GEM_HOOGTE		Gemiddelde afstand tussen het maaiveld en de hoogtes die voorkomen in de kroon in meters.	
	SHAPE_AREA		De oppervlakte van de boomkroonprojectie op het maaiveld in m ² .	
2 Data: - publicatie, - inwinning, Verversings- frequentie	Actualiteit	Datum	Eenmalig	
	temporele nauwkeurigheid geldigheid (van.. tot...)	afwijking (in dagen)	In principe beperkt de geldigheid zich tot opnamedata van het AHN2-bestand.	Opnamedata AHN2 verschillen per regio in de periode 2008-2012

Producteis	Kwaliteitselement	Eis + Eenheid	Opmerkingen kwaliteitselement	Opmerkingen producteis
3 Ieder object heeft een geometrie (punt, vlak, lijn, raster)*	logische consistentie	% (aantal objecten juist)	overlap / slivers komen niet voor in het bestand	* indien kwaliteitselementen voor verschillende gebieden verschillen, dan dat beschrijven
	logische consistentie	% (aantal objecten juist)	relatie tussen geometrietype en attribuut. Deze is 100%. als elk object een boom is.	
	omissie / commissie	% (aantal objecten te veel of te weinig)	Uitkomsten variëren met gebruikte wijze van validatie.	
	positionele nauwkeurigheid	meters	Voor kleine objecten (solitaire bomen) vermoedelijk vergelijkbaar met de nauwkeurigheid van AHN2 (0.5 m). Voor grote objecten: onduidelijk.	
4 brontype (dataset, service)	Beschikbaarheid (bv 24*7)	uur/week	Kaartbeeld is als webservice beschikbaar.	
5 Resolutie	positionele nauwkeurigheid	meters	NVT	
6 Meta-informatie van de dataset moet bepaalde kenmerken hebben*	volledigheid		aanwezigheid van bepaalde meta-informatiekenmerken, zoals verwijzing naar een objectencatalogus of conformiteitsbeschrijving (zie bijlage).	* Lijst van meta-informatievelen die voor de gebruiker van belang zijn.
7 Compatibiliteit eerdere versies *	Lineage	'vrije tekst'	Er is slechts één versie.	* welke elementen van compatibiliteit zijn van belang voor de gebruiker.
8 Objecten moeten binnen bepaald gebied liggen			In principe alleen Nederland, maar dit is niet gecontroleerd.	

6.4 Toetsen van de bruikbaarheid

Een overzicht van de verschillende relevante kwaliteitselementen geeft meer inzicht in de aspecten waar je op moet letten bij het beoordelen van de kwaliteit van een dataset. Maar daarmee is het beoordelen ervan an sich nog niet echt veel eenvoudiger geworden. Om het beoordelingsproces zo veel mogelijk te faciliteren, wordt voorgesteld om uit te gaan van een aantal generieke use-cases die, gecombineerd met een aantal vragen, de gebruiker kunnen helpen bij het beoordelen van de geschiktheid van een dataset. In de onderstaande tabel worden zes verschillende use-cases gepresenteerd en kort toegelicht. Deze use-cases overlappen in belangrijke mate met de verschillende vormen van gebruik van het boomkronenbestand zoals dat in paragraaf 6.2.2 is gepresenteerd.

Tabel 6.3

Verschillende use-cases die elk aparte eisen stellen aan de kwaliteit van een dataset.

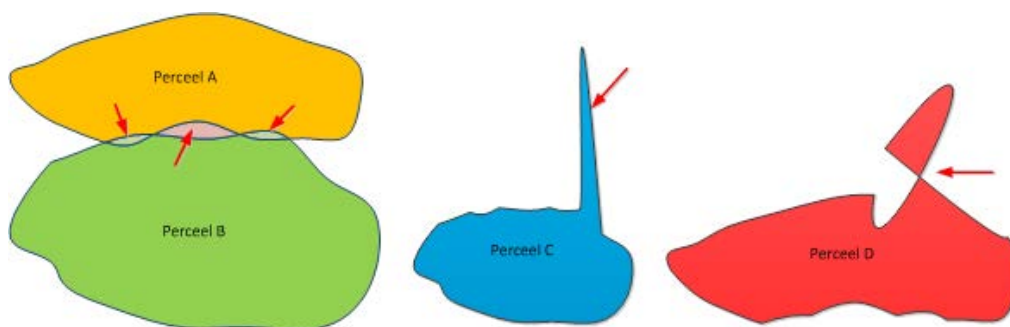
Generieke use-case	Beschrijving	Voorbeelden
Visualisatie - kleinschalig	U wilt een dataset gebruiken voor een visualisatie op een schaal van 1:10.000 of minder gedetailleerd.	Overzichtskaarten op regionale schaal
Visualisatie - gedetailleerd	U wilt een dataset gebruiken voor een visualisatie op een schaal van 1:10.000 of meer gedetailleerd.	Begeleidende informatie bij aanbestedingen en bestekken
Analyse	U wilt een dataset gebruiken voor het doen van een ruimtelijke analyse.	Berekenen van afstanden en routes; modelmatige toepassingen; genereren van statistieken; combineren en verrijken van datasets
Registratie van objecten	U wilt een dataset gebruiken voor het vastleggen van gegevens waar rechtszekerheid belangrijk bij is.	Kadaster; vastleggen van eigendom of gebruik
Communicatie	U wilt een dataset gebruiken om betrokkenen te informeren.	Infographics bij nieuwsitems; kaarten voor burgers bij gebiedsontwikkelingen
Beleid	U wilt een dataset gebruiken in beleidsmatige toepassingen.	Ontwikkelen, evalueren en monitoren van beleid; impact assessment
Ruimtelijke ingrepen	U wilt de dataset gebruiken als referentie voor een ruimtelijke ingreep.	Graafwerkzaamheden; bodemsanering; precisie landbouw

Voor elke use-case geldt dat verschillende aspecten van belang zijn. Voor een analyse is de lineage bijvoorbeeld belangrijker dan voor het maken van een kleinschalige visualisatie. Zo stelt elke use-case zijn eigen specifieke eisen (zie Tabel 6.5).

Om de gebruiker te helpen bij het vaststellen wat voor hem belangrijk is of niet, wordt een aantal vragen gesteld. Hierbij wordt geprobeerd om zo veel mogelijk onduidelijke termen te vermijden. Een voorbeeld: om te bepalen in hoeverre logische consistentie van belang is worden de volgende twee vragen gesteld:

- Is overlap relevant? Ja, Nee
Zo ja: geef aan hoe veel overlap toegestaan is (% aantal objecten, % vierkante meters)
- Is de aanwezigheid van slivers relevant? Ja, Nee
Zo ja: geef aan hoe veel slivers toegestaan zijn (% aantal objecten, % vierkante meters)

Beide vragen worden ondersteund door een illustratie die laat zien wat bedoeld wordt met overlap en slivers (zie Figuur 6.3). Vervolgens wordt het antwoord van de gebruiker vergeleken met de informatie die beschikbaar is over een of meerdere datasets. De use-case bepaalt in principe welke vragen gesteld worden, maar de gebruiker heeft altijd de mogelijkheid om zelf aan te geven welke vragen voor hem relevant zijn.



Figuur 6.3 Links wordt een voorbeeld van een kleine overlap (sliver) getoond, in het midden een spike en rechts een selfintersect (kruisende lijn).

Tabel 6.4

Welke kwaliteitskenmerken zijn van belang voor een generieke use-case?

Generieke use-case		Visualisatie kleinschalig	Visualisatie gedetailleerd	Analyse	Registratie van objecten	Communicatie	Beleid	Ruimtelijke ingrepen
Ruimtelijk	Logische consistentie (overlap / slivers)		X	X	X			X
	Logische consistentie (geometrietype)			X	X			X
	Omissie / commissie		X	X	X		X	X
	Positionele nauwkeurigheid	X	X	X	X		X	X
	Gebiedsdekking	X	X	X	X	X	X	X
	Resolutie (pos. nauwkeurigheid)	X	X	X				X
Thematisch	Thematische nauwkeurigheid		X	X	X	X	X	X
	Volledigheid		X	X	X		X	X
	Actualiteit	X	X	X	X	X	X	X
	Lineage			X	X			X
	Logische consistentie			X				X
Temporeel	Actualiteit			X	X			X
	Temporele nauwkeurigheid			X	X			X
Overig	Beschikbaarheid (bv 24*7)					X		X
	Volledigheid metainformatie	X	X	X	X	X	X	X
	Lineage (compatibiliteit)			X				X

De in dit hoofdstuk gedane suggesties zijn mede gebaseerd op de inhoud van Vullings *et al.* (2015) en Meijer *et al.* (2015).

6.5 Conclusies en aanbevelingen

In de voorgaande paragraaf werd een raamwerk geschetst dat de producent kan helpen bij het beschrijven van de dataset die hij heeft geproduceerd. Het raamwerk kan de gebruiker ook helpen bij het beoordelen van de fitness for use, de geschiktheid voor zijn gebruiksdoel. Binnen de scope van dit onderzoek is het niet mogelijk om een volledig kwaliteitsoordeel te geven over het boomkronenbestand. Toch is er een aantal conclusies te trekken en aanbevelingen te geven.

Op dit moment is er geen platform beschikbaar waar gebruikers met vragen over de open data-versie van het boomkronenbestand terecht kunnen. Een mogelijkheid zou kunnen zijn om hier een apart platform voor te creëren, met name gericht op open data-gebruikers, vergelijkbaar met bijvoorbeeld Open Street Map (OSM). Recentelijk is er een coöperatie opgezet rondom het vermarkten van het boomkronenbestand. Dus een andere optie zou kunnen zijn om dit onder te brengen bij deze nieuwe coöperatie.

Een andere manier om de interactie tussen de open data-gebruiker en de producent van het boomkronenbestand te verbeteren, is door alle gebruikers van dataset zich eerst te laten registreren. Voor de producent wordt dan iets meer inzichtelijk wie de gebruikers zijn. Daarnaast wordt het dan ook makkelijker om met deze groep in contact te komen.

De metadataset is in zijn huidige vorm onvoldoende om voor de gebruiker de fitness for use te bepalen. Los van de inhoud van de data speelt ook mee dat voor een groot aantal gebruikers de terminologie vaak te technisch is en te weinig houvast biedt om te bepalen of de dataset voor hen ook bruikbaar is (zie ook Devillers *et al.*, 2010).

Tot slot: zoals al eerder opgemerkt, is het niet mogelijk om alle vragen te beantwoorden die gebruikers rondom het boomkronenbestand kunnen hebben. Daar is eerst meer onderzoek voor nodig en moeten meer gebieden worden onderzocht. Voorgesteld wordt om in een vervolgonderzoek het in dit hoofdstuk gepresenteerde kwaliteitsraamwerk als uitgangspunt te kiezen voor verdere controles.

Literatuur

Algemene Rekenkamer, 2014. Trendrapport open data.

Ardila Lopez, J.P., 2012. Object - based methods for mapping and monitoring of urban trees with multitemporal image analysis. PhD thesis University of Twente. ITC Dissertation 209.

Aronoff, S., 1989. Geographic Information Systems: A management perspective. Ottawa, Canada: WDC Pub.

Bevelander, M., Bulens, J., Dijkstra, S., 2014. NCG-symposium 'Hoezo datakwaliteit?'. Geo-Info 2014-5 p44-45.

Bowler, D.E., Buyung-Ali, L., Knight, T.M., & Pullin, A.S., 2010. Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landscape and urban planning*, 97(3), 147-155.

Clement, J., 2012. Silvistar 3D Tree model and Tree extraction using LiDAR (ESRI User Conference paper).

CityGML: <http://www.citygml.org/>

CROW, 2012. Combineren van onder- en bovengrondse infrastructuur met bomen. CROW Publicatie 280, Ede, maart 2012

Devillers, R., Y. Bédard, M. Gervais, R. Jeansoulin, F. Pinet, M. Schneider, L. Bejaoui, M.A. Levesque, M. Salehi en A. Zargar, 2007. How to improve geospatial data usability: From metadata to Quality-Aware GIS Community. Spatial Data Usability Workshop. AGILE 2007 Conference, Aalborg, Denmark.

Devillers, R., Y. Bédard, R. Jeansoulin en B. Moulin, 2007. Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Sciences (IJGIS)*, vol. 21, n.3, p. 261-282.

Devillers, R., Bédard, Y., Jeansoulin, R., 2005. Multidimensional Management of Geospatial Data Quality Information for its Dynamic Use Within GIS. *Photogrammetric engineering and remote sensing* 2005 vol:71 iss:2 pg:205 -215

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., S. Wenzhong, 2010. Thirty years of research on spatial data quality: achievements, failures, and opportunities. *Transactions in GIS* 14-4, p.387-400.

Dienst Ruimtelijke Ordening Gemeente Amsterdam, 2013: Concept bomenverordening 2014. Amsterdam.
http://www.amsterdam.nl/gemeente/organisatie-diensten/dienst-ruimtelijke/bomenverordening#h3_1

Geonovum, 2013. IMGeo gegevenscatalogus, Deel 2.1, bron:
<http://www.geonovum.nl/dossiers/bgtimgeo/destandaard>, accessed: 10-6-2013.

Georgi, N.J. & Zafiriadis, K., 2006. The impact of park trees on microclimate in urban areas. *Urban Ecosystems*, 9(3), 195-209.

-
- ISO (International Standardisation Organisation), 2002. ISO 19113:2003 Geographic information – Quality principles, 29p.
- ISO (International Standardisation Organisation), 2003a. ISO 19114:2003 Geographic information – Quality evaluation procedures, 63 p.
- ISO (International Standardisation Organisation), 2003b. ISO 19115:2003 Geographic information – Metadata, 140 p.
- Jong, J.J. de, R.A. Smidt, J.H. Spijker en C.M. Niemeijer, 2011. Databank Gemeentelijk Groenbeheer, Rapportage boekjaar 2010, aangepaste versie. Wageningen, Alterra 2011, 83 blz, 45 fig., 17 tab.
- Klemm, W., Heusinkveld, B.G., Lenzholzer, S., Jacobs, M.H., & Van Hove, B., 2015. Psychological and physical impact of urban green spaces on outdoor thermal comfort during summertime in The Netherlands. *Building and Environment*, 83, 120-128.
- Klijn, E.H., G.R. Teisman, 2000. Governing Public-Private Partnerships; analysing and managing the processes and institutional characteristics of public-private partnerships in: Osborne (ed.) (2000).
- Koop, H., 1989. Forest dynamics. SILVI-STAR: A comprehensive monitoring system. Springer-Verlag.
- OGC (Open Geospatial Consortium), 2012. OGC CityGML Encoding Standard document version 2.0.0, <http://www.citygml.org>, accessed: 10-6-2013.
- Krause, A.U.M., Danes, M.H.G.I., Janssen, H., Kramer, H., Rip, F.I., 2011 Recording small landscape features with object recognition; Possibilities and limitations of automated procedures to support monitoring in the frame of the GeoCAP. Wageningen UR, Alterra rapport 2137.
- Kumar, S., Singh, N., & Prasad, R., 2010. Anhydrous ethanol: A renewable source of energy. *Renewable and Sustainable Energy Reviews*, 14(7), 1830-1844.
- Laufer, R., 2011. Prozedurale Qualitätsmodellierung und -management für Daten. Institut für Ingenieurgeodäsie der Universität Stuttgart.
- Meijer, M., Vullings, L.A.E., Bulens, J.D., Rip, F.I., Boss, M., Hazeu, G., Storm, M., 2015. Spatial Data Quality and Tooling. 9th International Symposium on Spatial Data Quality, 29-30 september, Montpellier.
- OGC, 2012. CityGML Encoding Standard. Gerhard Gröger, Thomas H. Kolbe, Claus Nagel, Karl-Heinz Häfele (Eds.). <http://www.opengis.net/spec/citygml/2.0>
- Oort, P.V., 2006. Spatial data quality: from description to application. Wageningen Universiteit.
- Oosterbaan, A., Hochbichler, E., Nicolescu, V.N., & Spiecker, H., 2009. Silvicultural principles, goals and measures in growing valuable broadleaved tree species. *Die Bodenkultur*, 45(60), 3.
- Rafiee, A., Dias, E., Koomen, E., 2013. Between Green and Grey: towards a new green volume indicator, in: Geertman S., Stillwell J., Toppen F. (eds.), *Proceedings of CUPUM 2013*, The 13th International Conference on Computers in Urban Planning and Urban Management. Planning Support Systems for Sustainable Urban Development, Utrecht, the Netherlands, 2013.
- Randrup, T.B., McPherson, E.G., & Costello, L.R., 2001. A review of tree root conflicts with sidewalks, curbs, and roads. *Urban Ecosystems*, 5(3), 209-225.
- Schouten, L., J. Clement, M. Flanagan, 2012. Bomen modelleren met laseraltimetrie. *Geo-Info 2012-7*, p.4-7.

-
- Smit, N., & van Thiel, S., 2002. De zakelijke overheid: publieke en bedrijfsmatige waarden in publiek-private samenwerking. Retrieved from <http://hdl.handle.net/1765/1468>
- Velarde, M.D., Fry, G., & Tveit, M., 2007. Health effects of viewing landscapes–Landscape types in environmental psychology. *Urban Forestry & Urban Greening*, 6(4), 199-212.
- Viana, H., Cohen, W.B., Lopes, D., & Aranha, J., 2010. Assessment of forest biomass for use as energy. GIS-based analysis of geographical availability and locations of wood-fired power plants in Portugal. *Applied Energy*, 87(8), 2551-2560.
- Vullings, L.A.E., Bulens, J.D., Rip, F.I., Boss, M., Meijer, M., Hazeu, G., Storm, M., 2015. Spatial Data Quality: What do you mean? AGILE Conference paper, Lissabon, 9-12 juni 2015.
- Xiao, W., Xu, S., Oude Elberink, S., Vosselman, G., 2012. Change detection of trees in urban areas using multi-temporal airborne lidar point clouds. <http://proceedings.spiedigitallibrary.org/>
- Żróbek, R., Dawidowicz, A. (Eds.), 2014. SPATIAL DATA IN WIDE GEOSPACE. Croatian Information Technology Society, University of Warmia and Mazury, University of Silesia.

Bijlage 1 Metadata Boomkronenbestand

Identificatie Contacten Metametadata Dekking Kwaliteit Inhoud Distributie

Identificatie

Boomkronen_OKT2012

Versie:2012, Silvi-Star

Unieke Identifier: 35decbd2-56c5-4af9-8b5b-265168370b04

Creatiedatum: 2012-10-01

Status: In ontwikkeling

Samenvatting: Bestand met projecties van boomkroonontrekken en inclusief de maximale hoogte van die kronen boven het maaiveld

Doel van vervaardiging: Technisch experiment: is een landsdekkend bestand van boomkronen + boomhoogte mogelijk op basis van AHN2, en kan dat aangevuld worden met boommodelleringsattributen volgens Silvi-Star?

Toepassingsschaal: 1:100 1:10000000

Resolutie: 0.5 m x 0.5 m

Ruimtelijk schema: Vector

Herzieningsfrequentie: Niet gepland

Onderwerpen: Flora en fauna, Hoogte, Natuur en milieu, Gezondheid, Locatie

Trefwoorden:

Hoogte, Bodemgebruik, Landgebruik, Menselijke gezondheid en veiligheid, Gebieden met natuurrisico's, Atmosferische omstandigheden

Thesaurus trefwoorden: GEMET - INSPIRE themes, version 1.0

Publicatiedatum thesaurus: 2008-06-01

Aanvullende informatie:

Alterra-rapport nr.2671, 2015: Meijer *et al.*, 2015: Boomkronen afleiden uit het Actueel Hoogtebestand Nederland. Kwaliteitsaspecten rondom het geautomatiseerd in kaart brengen van bomen op basis van het AHN2-bestand

download URL Alterra rapporten: <http://www.wageningenur.nl/nl/Expertises-Dienstverlening/Onderzoeksinstituten/Alterra/Publicaties/Alterra-rapporten.htm>

Gebruiksbeperkingen: Geen gebruiksbeperkingen

(Juridische) gebruiksrestricties: Niet toegankelijk

(Juridische) toegangsrestricties: Niet toegankelijk

Taal van de bron: Nederlands

Karakterset van de bron: utf8

Identificatie	Contacten	Metametadata	Dekking	Kwaliteit	Inhoud	Distributie
---------------	-----------	--------------	---------	-----------	--------	-------------

Contact

Metadata auteur: Frans Rip
Naam organisatie: GeoDesk Wageningen UR
Rol organisatie: Verstrekker
Website organisatie: <http://help.geodesk.nl>
Naam contactpersoon: Frans Rip
Rol contactpersoon: verstrekker
E-mail: geodesk.cqi@wur.nl

Identificatie	Contacten	Metametadata	Dekking	Kwaliteit	Inhoud	Distributie
---------------	-----------	--------------	---------	-----------	--------	-------------

Metametadata

Metadata taal: Nederlands
Metadata hiërarchieniveau: Dataset

Metadata unieke identifier: b3072f24-a692-4a92-a567-0eb35b97dd2c
Metadata wijzigingsdatum: 2015-01-21
Metadata standaard naam: ISO 19115
Metadata standaard versie: Nederlandse metadataprofiel op ISO 19115 voor geografie 1.3

Dekking

Code referentiesysteem: 28992

Verantwoordelijke organisatie voor namespace referentiesysteem: EPSG

Code verticaal referentiesysteem: 5709

Verantwoordelijke organisatie voor namespace verticaal referentiesysteem: Nederlandse Commissie voor Geodesie

Beschrijving geografisch gebied: Nederland

Omgrenzende rechthoek in meters:

Minimum x-coördinaat: 13555.4178

Maximum x-coördinaat: 278018.7452

Minimum y-coördinaat: 306833.1646

Maximum y-coördinaat: 614414.2939

Temporele dekking:

Van datum: 2008-01-01

Tot datum: 2012-03-06

Beschrijving temporele dekking: AHN2 is opgenomen en geproduceerd in de periode 2008-2012 (verschillende gebieden in verschillende jaren). Dat wil zeggen dat de afgeleide boomkronen klopten voor de opnametijd van het betreffende deelgebied. Er is geen sprake van één landsdekkende geldigheidsdatum.

Kwaliteit

Algemene beschrijving herkomst: De omtrekken van boomkronen (polygonen) zijn afgeleid uit de AHN2-gridcellen (0.5x0.5 m raster-resolutie). Daarna zijn enkele Silvi-Star-parameters berekend en toegevoegd.

Geometrische nauwkeurigheid:

Qua door polygonen bestreken oppervlakte: merendeel van de vegetatie wordt afgedekt

Qua individualiteit: is slecht voor aaneengesloten boomkronen en vrij goed voor solitaire bomen

Qua objecthoogte: ongeveer even goed als het AHN-grid (+/- 5 cm)

Topologische samenhang: Niet van toepassing

Volledigheid: Landsdekkend

Specificaties:

Titel: Boomkronen_Okt2012

Datum: oktober 2012

Conformiteitsindicatie: shapefile

Uitgevoerde bewerkingen:

Beschrijving:

- Aftrekken van de gridcelwaarde met de terreinhoogte van de maximale gridcelwaarde (de bovenkant), t.b.v. de hoogte in de gridcel boven maaiveld.
- Verkleinen zoekgebied door maken van uitsluitingsmaskers:
 - Het nodata-top10 masker: alles wat geen boomkroon kan zijn op basis van Top10NL en BAG.
 - Het STD_notrees masker. Masker bevat alle min of meer vlakke bovenkanten, bepaald aan de hand van standaardafwijkingpatroon in venster van 3x3 gridcellen. Cel met waarde die weinig afwijkt van omgeving komt in masker.
- Boomkroon bepaling door iteratieve celselectie rond piekwaarden in de niet gemaskeerde gebieden en clustering daarvan.
- Bepaling van de polygoon rond celcluster.
- Pseudostamlocatie door berekening van de centroïde van de polygoon.

Bewerkende organisatie: Jan Clement, Alterra: kroonprojecties.

Beschrijving: Silvi-Star-parameters toevoegen: van de Silvi-Star parameters zijn enkel de 4 perifere waarden bepaald van de kroonpolygoon (P1-P4). De kroontop (T) hoogte was al vastgelegd als kroonattribuut.

Datum bewerking: 2012-10-01

Bewerkende organisatie: Geodan (Azhar Rafiee)

Inhoud

Geen inhoudelijke informatie beschikbaar (applicatieschema).

Object naam: crownWithPeriferyPoints

Attributen**Shape****OBJECTID_1****OBJECTID****BLAD_ID****BOOM_ID****SOLITAIR****GRIDS****BM_X****BM_Y****GEM_MV****BOOM_HOOGT****GEM_HOOGTE****Shape_Leng****FID_1****Id****minX_x****minX_y****maxX_x****maxX_y****minY_x****minY_y****maxY_x****maxY_y****deltaX****deltaY****gemB****indexB****ShpIndex****Shape_Length****Shape_Area**

Distributie

Naam en versie van distributieformaten:

File Geodatabase Feature Class

Locatie dataset: \\...\...\Bomen_OKT2012\Bomen_sylvistar.gdb

Distributeur:

Naam organisatie: Wageningen UR GeoDesk

Rol organisatie: Distributeur

Website organisatie: _

Naam contactpersoon: Frans Rip

Rol contactpersoon: Verstrekken selecties op verzoek

E-mail: geodesk.cgi@wur.nl

Alterra Wageningen UR
Postbus 47
6700 AA Wageningen
T 0317 48 07 00
www.wageningenUR.nl/alterra

Alterra-rapport 2671
ISSN 1566-7197



Alterra Wageningen UR is hét kennisinstituut voor de groene leefomgeving en bundelt een grote hoeveelheid expertise op het gebied van de groene ruimte en het duurzaam maatschappelijk gebruik ervan: kennis van water, natuur, bos, milieu, bodem, landschap, klimaat, landgebruik, recreatie etc.

De missie van Wageningen UR (University & Research centre) is 'To explore the potential of nature to improve the quality of life'. Binnen Wageningen UR bundelen 9 gespecialiseerde onderzoeksinstituten van stichting DLO en Wageningen University hun krachten om bij te dragen aan de oplossing van belangrijke vragen in het domein van gezonde voeding en leefomgeving. Met ongeveer 30 vestigingen, 6.000 medewerkers en 9.000 studenten behoort Wageningen UR wereldwijd tot de aansprekende kennisinstellingen binnen haar domein. De integrale benadering van de vraagstukken en de samenwerking tussen verschillende disciplines vormen het hart van de unieke Wageningen aanpak.

To explore
the potential
of nature to
improve the
quality of life



Alterra Wageningen UR
Postbus 47
6700 AA Wageningen
T 317 48 07 00
www.wageningenUR.nl/alterra

Alterra-rapport 2671
ISSN 1566-7197

Alterra Wageningen UR is hét kennisinstituut voor de groene leefomgeving en bundelt een grote hoeveelheid expertise op het gebied van de groene ruimte en het duurzaam maatschappelijk gebruik ervan: kennis van water, natuur, bos, milieu, bodem, landschap, klimaat, landgebruik, recreatie etc.

De missie van Wageningen UR (University & Research centre) is 'To explore the potential of nature to improve the quality of life'. Binnen Wageningen UR bundelen 9 gespecialiseerde onderzoeksinstituten van stichting DLO en Wageningen University hun krachten om bij te dragen aan de oplossing van belangrijke vragen in het domein van gezonde voeding en leefomgeving. Met ongeveer 30 vestigingen, 6.000 medewerkers en 9.000 studenten behoort Wageningen UR wereldwijd tot de aansprekende kennisinstellingen binnen haar domein. De integrale benadering van de vraagstukken en de samenwerking tussen verschillende disciplines vormen het hart van de unieke Wageningen aanpak.

