

Opstellen van tarieven voor Inlandse eik en Beuk in Vlaanderen ten behoeve van het berekenen van houtvolumes

Statistische evaluatie van de regressiemodellen en overzicht van de resultaten (technisch rapport deel III)

Paul Quataert, Beatrijs Van der Aa en Pieter Verschelde

INBO.R.2011.18



BEUK
#0005
Dist: 00
IC1: 26

Auteurs:

Paul Quataert, Beatrijs Van der Aa en Pieter Verschelde

Instituut voor Natuur- en Bosonderzoek
Wetenschappelijke instelling van de Vlaamse overheid

Instituut voor Natuur- en Bosonderzoek

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

Vestiging:

INBO Geraardsbergen
Gaverstraat 4, 9500 Geraardsbergen
www.inbo.be

e-mail:

paul.quataert@inbo.be

Wijze van citeren:

Quataert, P., Van der Aa, B. Verschelde, P. (2011). Opstellen van tarieven voor Inlandse eik en Beuk in Vlaanderen ten behoeve van het berekenen van houtvolumes. Statistische evaluatie van de regressiemodellen en overzicht van de resultaten (technisch rapport deel III).. Rapporten van het Instituut voor Natuur- en Bosonderzoek 2011 (18). Instituut voor Natuur- en Bosonderzoek, Brussel.

D/2011/3241/169

INBO.R.2011.18

ISSN: 1782-9054

Foto cover:

Digitale meetklem

Verantwoordelijke uitgever:

Jurgen Tack

Druk:

Management ondersteunende diensten van de Vlaamse overheid

Dit onderzoek werd uitgevoerd in opdracht van:

Agentschap voor Natuur en Bos, Koning Albert-II laan 20, 1000 Brussel , TWOL: 0L200000505 B&G/2001/BOA1



Agentschap voor
Natuur en Bos

Modellen voor de houtvolumes voor Inlandse eik en Beuk in Vlaanderen

Statistische evaluatie van de regressiemodellen en overzicht van de resultaten (technisch rapport deel III).

Paul Quataert, Beatrijs Van der Aa en Pieter Verschelde

Inhoud

1	Inleiding	7
2	Materiaal en methoden	8
2.1	De variabelen: symbolen en notaties.....	8
2.1.1	De basisvariabelen.....	8
2.1.2	De achtergrondvariabelen	8
2.1.3	Het volume van de kegel als “proxy”	9
2.2	De modellen.....	9
2.2.1	De algemene vorm.....	9
2.2.2	Beknopte notatie voor de structurele component	10
2.2.3	Voorstellen en motivatie voor de structurele component	10
2.3	Algemene strategie van de modelbouw.....	11
2.3.1	Verkenning van de gegevens (EDA)	12
2.3.2	Modelanalyse	14
2.3.3	Vergelijking van de modellen	15
2.4	De tarieven	17
3	Beuk: verkenning van de gegevens (EDA)	18
3.1	Het meetbereik en de selectie van de gegevens.....	18
3.2	Opbouw van de dataset	19
3.3	De referentieboom en het referentiebereik per ecoregio	20
3.4	Relatie tussen omtrek en hoogte	21
3.5	Verkenning van het model met twee ingangen	21
3.6	Verkenning van het model met één ingang (en ecoregio).....	22
3.7	Verkenning van het relatief volume	22
3.8	Variantie van het volume in functie van de omtrek.....	23
4	Beuk: statistische analyse van de modellen	24
4.1	De veeltermmodellen	24
4.1.1	Veeltermmodel met twee ingangen	24
4.1.2	Veeltermmodel met één ingang	27
4.1.3	Veeltermmodel met één ingang en ecoregio.....	28
4.2	De logaritmische modellen	33
4.2.1	Logaritmisch model met twee ingangen	33
4.2.2	Logaritmische model met één ingang	35
4.2.3	Logaritmische model met één ingang en ecoregio	37
4.3	Relatief model met twee ingangen	41
5	Beuk: vergelijking van de modellen	44
5.1	Vergelijking van de voorspellingen van de modellen	44
5.1.1	De referentieboom	44
5.1.2	De tarieven met twee ingangen	44
5.1.3	De tarieven met één ingang per ecoregio.....	46
5.2	Vertekening en variabiliteit van de modellen	47
5.2.1	Distributie van de gestandaardiseerde residu's	47
5.2.2	Juistheid en precisie van de voorspellingen	47
5.2.3	Grafische voorstelling van de accuraatheid.....	48
5.3	De invloed van de meetfout op de hoogte	50
5.4	Algemeen besluit	51
5.4.1	De tarieven met twee ingangen	51
5.4.2	De tarieven met één ingang	51

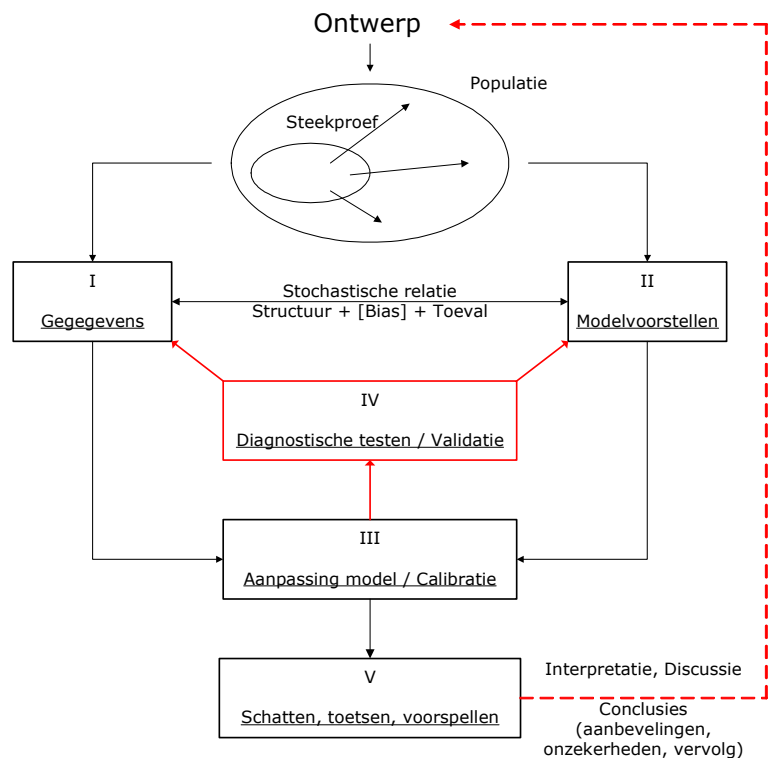
6	Beuk: tarieven (rudimentaire tabellen)	52
6.1	Inleiding.....	52
6.2	Tarief met twee ingangen (veeltermmodel)	53
6.2.1	De tarieven in tabelvorm	53
6.2.2	Controle per klasse.....	54
6.2.3	Grafische voorstelling van de tarieven.....	55
6.3	Tarieven met één ingang (veeltermmodel)	56
6.3.1	De tarieven in tabelvorm	56
6.3.2	Controle per klasse (Tabel 6-5 en Figuur 6-6)	57
6.3.3	Grafische voorstelling van de tarieven.....	58
6.4	Tarieven met één ingang (logaritmisch model).....	59
6.4.1	De tarieven in tabelvorm	59
6.4.2	Controle van de tarieven	60
6.4.3	Grafische voorstelling van de tarieven.....	61
6.5	Tarief voor grote omtrekken	61
7	Eik: verkenning van de gegevens (EDA)	63
7.1	Het meetbereik en selectie van de gegevens.....	63
7.2	Opbouw van de dataset.....	63
7.3	De referentieboom en het referentiebereik per ecoregio	65
7.4	Relatie tussen omtrek en hoogte	65
7.5	Verkenning van het model met twee ingangen	66
7.6	Verkenning met het model met één ingang (en ecoregio).....	66
7.7	Verkenning van het relatief model	67
7.8	Variantie van het volume in functie van de omtrek	67
8	Eik: statistische analyse van de modellen	69
8.1	Veeltermmodellen	69
8.1.1	Veeltermmodel met twee ingangen.....	69
8.1.2	Veeltermmodel met één ingang.....	72
8.1.3	Veeltermmodel met één ingang en ecoregio	74
8.2	Logaritmische modellen	78
8.2.1	Logaritmisch model met twee ingangen	78
8.2.2	Logaritmisch model met één ingang	81
8.2.3	Logaritmisch model met één ingang en ecoregio	83
8.3	Relatief model met twee ingangen.....	86
9	Eik: vergelijking van de modellen	90
9.1	Vergelijking van de voorspellingen van de modellen	90
9.1.1	De referentieboom	90
9.1.2	De tarieven met twee ingangen.....	90
9.1.3	De tarieven met één ingang per ecoregio	92
9.2	Vertekening en variabiliteit van de modellen.....	93
9.2.1	Distributie van de gestandaardiseerde residu's.....	93
9.2.2	Juistheid en precisie van de voorspellingen.....	94
9.2.3	Grafische voorstelling van de accuraatheid	94
9.3	De invloed van de meetfout op de hoogte.....	96
9.4	Algemeen besluit.....	97
9.4.1	De tarieven met twee ingangen.....	97
9.4.2	De tarieven met één ingang.....	97
10	Eik: tarieven (rudimentaire tabellen)	98
10.1	Inleiding.....	98
10.2	Tarief met twee ingangen	99
10.2.1	De tarieven in tabelvorm	99
10.2.2	Controle per klasse.....	100

10.2.3	Grafische voorstelling van de tarieven	101
10.3	Tarieven met één ingang (veeltermmodel)	102
10.3.1	De tarieven in tabelvorm	102
10.3.2	Controle per klasse	103
10.3.3	Grafische voorstelling van de tarieven	104
10.4	Tarieven met één ingang (logaritmisch model)	105
10.4.1	Voorstelling van de tarieven in tabelvorm	105
10.4.2	Controle van de tarieven.....	105
10.4.3	Grafische voorstelling van de tarieven	106
10.5	Tarief voor grote omtrekken.....	107
Lijst van figuren		108
Lijst van tabellen.....		113

1 Inleiding

Tarieven voorspellen het houtvolume van een boom op basis van enkele basisafmetingen van de boom (zoals omtrek en hoogte) en (eventueel) de kenmerken van de omgeving waarin de boom staat (bestandskwaliteit, bodemeigenschappen, hydrologische toestand, regio). In dit tweede deel van het technisch rapport beschrijven we hoe we tarieven ontwikkeld hebben op basis van statistische modelbouw. Als leidraad hebben we hierbij Figuur 1-1. Essentieel in het schema is stap IV. Voordat we de conclusies trekken uit een bekomen model, moeten we controleren of het model wel volledig consistent in elkaar zit. Zo niet is er een bijsturing nodig. Ofwel zijn de gegevens niet in orde ofwel moeten we de modellen aanpassen. Hoe we concreet met dit schema omgaan, staat beschreven in Materiaal en Methoden.

STATISTISCHE MODELBOUW



Figuur 1-1 De essentiële stappen bij statistische modelbouw gesitueerd binnen de fasen van het empirisch onderzoek.

2 Materiaal en methoden

2.1 De variabelen: symbolen en notaties

2.1.1 De basisvariabelen

Er zijn drie basisvariabelen:

- Het houtvolume V (m³).
- De omtrek C (m) (gemeten op 1.50 m).
- De hoogte H (m) (tot waar de stam een diameter heeft van 7 cm).

Bedoeling is dat we op basis van omtrek en hoogte (tarieven met twee ingangen) of omtrek alleen (tarieven met één ingang) het volume voorspellen. Conform de algemene notatie bij statistische modelbouw duiden we het voorspelde volume op basis van het tarief aan met $\hat{V}$ (V hoedje). De afwijking tussen het voorspelde volume en het opgemeten volume heet het residu:

$$e = V - \hat{V} \quad (2.1)$$

Heel dikwijls zullen we werken met een procentuele afwijking omdat naarmate de bomen groter worden, de afwijkingen ook groter worden:

$$e_p = \frac{V - \hat{V}}{\hat{V}} \quad (2.2)$$

Voor een vlotte notatie gebruiken we soms C^2 , C^3 , C^2H , ... voor C^2 , C^3 en C^2H . De functie $\text{poly}(\cdot, p)$ duidt een veelterm aan van orde p . Bijvoorbeeld:

$$\text{poly}(C, 3) = \beta_0 + \beta_1 C + \beta_2 C^2 + \beta_3 C^3 \quad (2.3)$$

Andere afkortingen zijn $\log V$, $\log C$ en $\log H$ voor $\log_{10}(V)$, $\log_{10}(C)$ en $\log_{10}(H)$.

2.1.2 De achtergrondvariabelen

De enige achtergrondvariabele die we gebruikt hebben bij de modelbouw is F , de ecoregio van het bestand waarin de boom staat. We onderscheiden vier ecoregio's:

- Z: Zoniën
- B: Brabants district
- V: Vlaams district
- K: Kempen district

Zoniën is strikt gezien geen apart ecoregio, maar het bos is op zich voldoende groot en heeft andere kenmerken, zodat een onderscheid verantwoord is.

Ecoregio gebruiken we in de modelbouw als een surrogaatvariabele die de standplaatskenmerken weergeeft omdat tal van kenmerken zoals bodemtype, bodemgesteldheid, drainageklasse en voedselrijkdom hiermee gerelateerd zijn. De kwaliteit van de standplaats bepaalt sterk hoe hoog de bomen (kunnen) worden. Dus bij tarieven met één ingang, die alleen gebaseerd zijn op de omtrek, is ecoregio een belangrijke kandidaatvariabele om het houtvolume te voorspellen.

2.1.3 Het volume van de kegel als "proxy"

Een simpel model om het houtvolume van een boom te voorspellen, is een boom voor te stellen als een (afgeknotte) kegel. Als $C = 2\pi R = \pi D$ de omtrek van het grondvlak en H de totale hoogte, dan is de corresponderende inhoud van de kegel:

$$V_{kegel} = \frac{1}{3} \pi R^2 H = \frac{1}{3} \pi \left(\frac{C}{2\pi} \right)^2 H = \frac{C^2 H}{12\pi} \cong \frac{C^2 H}{37.7} \quad (2.4)$$

Deze berekening is niet helemaal juist want we meten de omtrek niet aan de basis en de hoogte is niet de volledige hoogte. Maar aangezien deze variabele slechts een hulpvariabele is, volstaat deze benadering. Het **relatieve volume** interpreteren we als **een vormfactor** φ die aangeeft in hoeverre het werkelijke volume afwijkt van het volume van een kegel:

$$relV = \frac{V}{V_{kegel}} = \varphi \quad (2.5)$$

2.2 De modellen

2.2.1 De algemene vorm

We hebben ons beperkt tot modellen met volgende (toch nog vrij algemene) basisvorm:

$$g(V) = \mu_V + \varepsilon \quad \begin{cases} E\{g(V)\} = \mu_V = h(C, H, F) \\ E\{\varepsilon\} = 0; \quad Var\{\varepsilon\} = \sigma_\varepsilon^2; \quad \varepsilon \propto \mathcal{N}(0, \sigma_\varepsilon^2(\mu_V)) \end{cases} \quad (2.6)$$

Hierbij stellen we het volume V of $g(V)$, een transformatie ervan, voor als een som van:

- Een structurele of "voorspelbare" component μ_V die het gemiddelde voorstelt van het volume die we via de functie h kunnen berekenen uit de variabelen C , H en/of F . De functie h is in principe om het even welke functie, die we heel dikwijls benaderen door een veelterm met onbekende coëfficiënten die we kunnen bepalen door het model aan te passen aan de gegevens.
- Een stochastische of "onvoorspelbare" component ε , die de toevallige fluctuaties voorstellen ten opzichte van het gemiddelde. Deze toevalscomponent of ruisterm wordt veroorzaakt door meetfouten, de biologische variabiliteit van de bomen (elke boom is verschillend) en eventuele tekortkomingen in het model (in zoverre ze geen systematische invloed hebben en geen vertekening veroorzaken).

Zoals we nog zullen motiveren, hebben we drie types transformaties onderzocht:

- V : geen transformatie (de oorspronkelijke schaal),
- $\log V$: de logaritmische transformatie (de log-schaal), en
- $relV$: herschaling ten opzichte van het volume van een kegel (relatieve schaal).

Een belangrijk kenmerk is dat de ruis additief is ten opzichte van het getransformeerde volume. Er staat een somteken tussen de structurele en stochastische component. Maar als we terugtransformeren, is dat niet meer het geval. Bijvoorbeeld voor het logaritmische model krijgen we na terugtransformatie een multiplicatief model:

$$\log_{10}(V) = \mu_V + \varepsilon \Rightarrow V = 10^{\mu_V + \varepsilon} = 10^{\mu_V} 10^\varepsilon \quad (2.7)$$

Een ander punt van aandacht is dat we toelaten dat de variantie van de ruisterm niet constant is. Een klassieke veronderstelling is dat de ruis homoskedastisch is (d.w.z. overal gelijk), maar bij veel biologische groeiprocessen is dat niet het geval. Daarom bouwen we deze heteroskedasticiteit van meet af aan in het model in.

2.2.2 Beknopte notatie voor de structurele component

De functie h is onbekend en daarom benaderen we deze functie met een veelterm. Een voorbeeld met één ingang is:

$$\mu_v = f(C) = \beta_0 + \beta_1 C + \beta_2 C^2 + \beta_3 C^3 \quad (2.8)$$

en met twee ingangen:

$$\mu_v = f(C, H) = \beta_0 + \beta_1 C + \beta_2 C^2 + \beta_3 C^3 + \beta_4 H + \beta_5 CH + \beta_6 C^2 H + \beta_7 C^3 H \quad (2.9)$$

Om niet telkens alle componenten te moeten expliciteren, gebruiken we een beknoptere notatie waarbij we alle coëfficiënten, ook het intercept (β_0), laten vallen. De twee voorbeelden worden:

$$V \sim C + C^2 + C^3 \quad \text{of} \quad V \sim \text{poly}(C, 3) \quad (2.10)$$

en:

$$V \sim C + C^2 + C^3 + H + HC + HC^2 + HC^3 \quad \text{of} \quad V \sim \text{poly}(C, 3) * H \quad (2.11)$$

Het maalteken (*) wijst op een interactie tussen termen. Dat komt neer op een vermenigvuldiging tussen continue variabelen. Bij een interactie tussen een categorische variabele en continue variabele, is de betekenis dat de coëfficiënten van de continue variabelen afhankelijk zijn van de klasse van de variabele. Bijvoorbeeld:

$$\log V \sim \log C * F \quad (2.12)$$

wordt voluit:

$$E\{\log V\} = \beta_{0(f)} + \beta_{1(f)} \log C \quad (2.13)$$

m.a.w. we laten de coëfficiënten β_0 en β_1 afhangen van de ecoregio.

2.2.3 Voorstellen en motivatie voor de structurele component

Heel dikwijls is de motivatie voor een veeltermmodel zwak. In de wiskundige analyse wordt wel aangetoond dat we veel functies kunnen benaderen door veeltermen via een reeksontwikkeling. Als we ons beperken tot een klein interval van een bepaald punt kan een eerste orde model voldoende zijn: we benaderen de functie door een rechte (de raaklijn aan de figuur). Door meer en meer termen op te nemen, krijgen we een steeds betere benadering die we over een groter bereik kunnen toepassen.

Voor de opstelling van tarieven kunnen we veeltermmodellen iets beter motiveren. We kunnen immers in een eerste benadering ervan uitgaan dat het volume van een boom recht evenredig is met het volume van een kegel met omtrek van het grondvlak C en hoogte H :

$$V \propto V_{\text{kegel}} \propto C^2 H \quad (2.14)$$

Deze relatie nemen we als het uitgangspunt om **kandidaatmodellen** op te stellen waaruit we kunnen selecteren door na te gaan welke modellen het best aansluiten bij de gegevens.

De veeltermmodellen

Voor het veeltermmodel vullen we de term $C^2 H$ in (2.14) aan met combinaties van C en H :

$$P2 \leftrightarrow V \sim C + C^2 + C^3 + H + CH + C^2 H \quad (2.15)$$

De maximale orde van het model is drie (de termen $C^2 H$ en C^3).

Met deze functie kunnen we relaties modelleren die afwijken van het volume van de kegel. Model (2.15) heeft twee ingangen. Het model met één ingang bekomen we door de termen met H te laten vallen.

$$P1 \leftrightarrow V \sim C + C^2 + C^3 \quad (2.16)$$

Hoe hoog een boom wordt, hangt sterk samen met de eigenschappen van de standplaats. Deze standplaatskarakteristieken zijn sterk gecorreleerd de ecoregio. Vandaar het idee om per ecoregio een apart tarief met één ingang te bepalen:

$$P1f \leftrightarrow V \sim (C + C^2 + C^3) * F \quad (2.17)$$

De logaritmisch modellen

Door van (2.14) de logaritme te nemen, bekomen we:

$$\log_{10}(V) \propto 2 \log_{10}(C) + \log_{10}(H) \quad (2.18)$$

Hieruit leiden we volgend meer algemeen model af:

$$L2 \leftrightarrow \log_{10}(V) \propto \log_{10}(C) + \log_{10}(H) + \log_{10}(C) : \log_{10}(H) \quad (2.19)$$

In dit model zijn de coëfficiënten van $\log C$ en $\log H$ niet meer noodzakelijk 2 en 1 en hebben we ook een interactieterm toegevoegd. Het is interessant om na te gaan in hoeverre de interactieterm significant is en in hoeverre de geschatte regressiecoëfficiënten van $\log C$ en $\log H$ afwijken van de theoretische kegelwaarden (2 en 1).

Model (2.19) is een tarief met twee ingangen. We bekomen het model met één ingang door de termen met hoogte erin te laten vallen:

$$L1 \leftrightarrow \log_{10}(V) \propto \log_{10}(C) \quad (2.20)$$

Opnieuw is het zinvol het model op te splitsen per ecoregio:

$$L1f \leftrightarrow \log_{10}(V) \propto \log_{10}(C) * F \quad (2.21)$$

Het relatief model

In het relatieve model werken we met de verhouding van het werkelijke volume en het kegelvolume. Volgens (2.14) is deze verhouding gelijk aan een evenredigheidsconstante φ .

$$V_{rel} = \frac{V}{V_{kegel}} = \varphi \quad (2.22)$$

Deze φ kunnen we nu laten afhangen van de omtrek en de hoogte. We houden hierbij het model zo eenvoudig mogelijk:

$$R2 \leftrightarrow V_{rel} = \varphi(C, H) \sim C + H \quad (2.23)$$

Een model met één ingang is hier niet mogelijk want om het relatief volume te berekenen, moet de hoogte gekend zijn.

2.3 Algemene strategie van de modelbouw

We hebben drie types regressiemodellen onderzocht: een veeltermmodel (P), een logaritmisch model (L) en een relatief model (R). Voor het veeltermmodel en logaritmisch model hebben we drie varianten: een model met twee ingangen waarbij we het volume voorspellen op basis van omtrek en hoogte, een globaal model met één ingang met alleen omtrek als voorspellende variabele en een model met één ingang uitgesplitst per ecoregio. In deze laatste variant is ecoregio een surrogaat voor hoogte en hiermee kunnen we ook eventuele vormverschillen tussen de ecoregio's opvangen.

Samengevat hebben we dus naargelang de schaal drie types **kandidaatmodellen**:

- P : een eerste serie is geformuleerd in de originele schaal en werkt rechtstreeks met V uitgedrukt als een veelterm of polynoom van C en H.
- L : bij de tweede groep gebruiken we de variabelen in de log-schaal ($\log V$, $\log C$, $\log H$).

- R : het derde type modelleert het relatieve volume (2.5) in functie van van C en H. Deze modellen zijn uiteraard alleen bruikbaar bij tarieven met twee ingangen aangezien we hiervoor het volume van de kegel moeten kunnen bereken.

Behalve voor het relatief model, hebben we een tarief T met twee ingangen en met één ingang, deze laatste al dan niet opgesplitst per ecoregio. Tabel 2-1 geeft een overzicht.

Tabel 2-1 Overzicht van de kandidaatmodellen

Type model	Tarieftype		
	T2: twee ingangen	T1: één ingang	T1f: per ecoregio
P: veeltermmodel	P2	P1	P1f
L: logaritmisch model	L2	L1	L1f
R: relatief model	R2	-	-

Bedoeling is uit deze kandidaatmodellen het meest geschikte model te kiezen. Hiertoe volgden we zowel voor Beuk en Inlandse eik apart een driestappenplan gespreid over drie achtereenvolgende hoofdstukken per soort (in een vierde hoofdstuk komen dan de tarieven zelf aan bod):

- (1) Verkenning van de gegevens.
- (2) Statistische analyse van de modellen.
- (3) Vergelijking van de modellen.

Dus, na een eerste verkennende analyse onderzochten we in het hoofdstuk "statistische analyse van de modellen" de verschillende kandidaatmodellen één voor één en gaven we een aparte bespreking van de statistische kwaliteit. In het daarop volgende hoofdstuk vergeleken we de resultaten van deze bespreking om het beste model te selecteren.

2.3.1 Verkenning van de gegevens (EDA)

De drie voornaamste functies van de gegevensverkenning (EDA = exploratory data analysis) zijn:

- (1) **Kennismaken en zich vertrouwd maken** met de gegevens met als concreet doel ze grondig te beschrijven als documentatie voor de latere fases van het onderzoek. Daarenboven heeft de lezer graag een concreet beeld van de ruwe gegevens om de resultaten beter te kunnen situeren en beoordelen.
- (2) **Inschatten van de gegevenskwaliteit en het empirisch bereik van de gegevens** met hieraan gekoppeld de ontwikkeling van een analysedatabank en eventueel een verbetering van de gegevens (*data cleaning*). Kennis van het empirisch bereik is essentieel om de reikwijdte van de resultaten in te schatten. Een aparte analysedatabank is nodig omdat de ruwe gegevens zich meestal niet lenen tot een directe verwerking, maar eerst omgevormd moeten worden tot bruikbare analysevariabelen.
- (3) **Het maken van een eerste analyse** van de gegevens en een eerste interpretatie van de resultaten als referentiekader voor de resultaten van een meer doorgedreven statistische analyse. Gevorderde statistische technieken zijn wel krachtiger, maar kunnen mathematische artefacten introduceren. Daarom is het wenselijk een analyse te maken van de gegevens op basis van zo weinig mogelijk veronderstellingen.

Concreet zullen we hier verslag geven over een vijftal belangrijke aspecten.

(1) Meetbereik en selectie van de gegevens

Een belangrijke doelstelling van het project was tarieven op te stellen voor grotere diameters. Maar vanaf een bepaalde omtrek (4 m voor Beuk en 3 m voor Inlandse eik) waren er maar een beperkt aantal **geïsoleerde waarnemingen** beschikbaar. Knelpunt is dat de parameters van de regressiemodellen heel gevoelig zijn aan geïsoleerde punten die buiten de kern van het meetbereik liggen omdat ze werken als een soort hefboom op de ligging van de regressievergelijking. Als deze **hefboom punten** (*leverage points*) onvoldoende representatief zijn (of als we hieraan twijfelen), dan is het beter ze bij de modelbouw weg te laten en achteraf te controleren

in hoeverre deze waarnemingen in de lijn liggen van het model of niet. Dat kan als een verlies aan informatie begrepen worden. Maar dat is niet het geval. Het is niet door een paar grote bomen extra in de dataset op te nemen, dat de **empirische onderbouwing** van het model breder wordt. Integendeel, we lopen het risico dat slechts enkele niet representatieve bomen een heel grote invloed uitoefenen op de regressievergelijking. Daarom onderzochten we het omtrekbereik op **continuïteit**. Op basis hiervan beperkten we voor Beuk de omtrek tussen 0.35 m en 4 m en voor Inlandse eik tussen 0.35 m en 3 m. Hierdoor vielen een heel beperkt aantal waarnemingen weg voor de modelbouw. Alle resultaten van het rapport zijn gebaseerd op deze selectie.

(2) Distributie van de basisvariabelen in relatie tot de ecoregio

De **basisvariabelen** voor de tarieven zijn omtrek (C), hoogte (H) en volume (V). Hiervan onderzochten we de distributie zowel grafisch (boxplots en cumulatieve distributie) als numeriek (kengetallen zoals percentielen en gemiddelde). Belangrijk aandachtspunt was in hoeverre deze distributies verschillen naargelang de ecoregio om na te gaan of een apart tarief nodig is per ecoregio. Daarnaast bekeken we ook de distributie van **het relatieve volume** (re/V), een variabele die afgeleid is uit de drie basisvariabelen. In (2.5) is het relatieve volume gedefinieerd als de verhouding tussen het opgemeten volume V en het volume van de kegel met een grondvlak met omtrek C en een hoogte H . We kunnen het relatieve volume interpreteren als een **vormfactor**. Vraag was of de vorm de boom verschilt naargelang de ecoregio en in hoeverre met andere woorden de vorm van de boom afhangt van de standplaatseigenschappen.

(3) De referentieboom

Om de voorspelde volumes van de verschillende kandidaatmodellen vlot te kunnen vergelijken, hebben we uit de **kengetallen** van de afmetingen voor elke soort een **"gemiddelde" referentieboom** afgeleid. Voor Beuk is dat een boom met omtrek 1.8 m en een hoogte 32.5 m. Voor Inlandse eik is dat een boom met omtrek 1.5 m en een hoogte van 25 m. Hierbij merken we op dat één gemeenschappelijke "gemiddelde" boom voor alle ecoregio's niet mogelijk was want de ecoregio's verschillen te veel. In het bijzonder waren de bomen in het Kempens district gemiddeld veel kleiner. Toch bleek ook in dat geval de gemeenschappelijke referentieboom geen uitbijter te zijn, maar wel een boom aan de grote kant.

(4) Grafische verkenning van de relaties tussen de afmetingen van de bomen

Vervolgens maakten we een grafische verkenning van de relaties tussen omtrek, hoogte, volume en relatief volume. Het doel hiervan was de **plausibiliteit** van de verschillende kandidaatmodellen en **de rol van ecoregio** te beoordelen. Hierbij gingen we niet alleen na of de **vorm** van de modellen goed was, maar onderzochten we ook of de **modelassumpties** in orde waren. In het bijzonder weten we voor biologische groeiprocessen dat de voorwaarde voor **homoskedasticiteit** niet vervuld is. Naarmate de bomen groter worden, neemt de variabiliteit toe. Bij regressiemodellen kunnen we hiervoor corrigeren door van schaal te veranderen (heel dikwijls wordt daarom met logaritmische modellen gewerkt). Een andere oplossing is te werken met een **gewogen regressie** waarbij de gewichten omgekeerd evenredig zijn met de variantie (of met het kwadraat van de standaardafwijking). Daarom onderzochten we ook hoe de standaardafwijking gerelateerd was met de dimensie van de boom. Als indicator voor de dimensie namen we het volume van de kegel.

(5) Standaardafwijking van de ruis op de gegevens

Om de relatie tussen de dimensie van de boom en de variabiliteit te **kwantificeren**, zochten we naar het best passende regressiemodel tussen de standaardafwijking van het (opgemeten) volume en het volume van de corresponderende kegel (die een maat is voor de afmetingen van de boom). Hiertoe deelden we het kegelvolume op in vijftig klassen met ongeveer evenveel waarnemingen. Binnen elke klasse werd de standaardafwijking bepaald van de (opgemeten) volumes en uitgezet ten opzicht van het kegelvolume. Uit deze analyse bleek dat de standaardafwijking ongeveer toenam volgens een lineair verband. Aangezien bij de gewogen lineaire regressie de gewichten omgekeerd evenredig moeten zijn met de variantie, hebben we de gewichten gelijk gesteld met de inverse waarde van de kegelvolumes in het kwadraat.

2.3.2 Modelanalyse

Voor elk model (type en variant) uit Tabel 2-1 hebben we dezelfde vijf stappen doorlopen:

- (1) Voorstelling van het **startmodel**
- (2) Selectie van de significante termen in het startmodel via **stapsgewijze regressie**.
- (3) **Residuanalyse** om de kwaliteit van het model te evalueren (modeldiagnose).
- (4) Bespreking van de **geschatte coëfficiënten**.
- (5) **Grafische voorstelling** van het model.

(1) + (2) Selectie van het model

Het **startmodel** is het **maximale model** dat we in beschouwing nemen. Niet noodzakelijk zijn alle termen belangrijk en daarom gaan we met een **stapsgewijze regressie** na welke termen significant zijn en welke termen niet. De niet-significante termen laten we vallen en zo bekomen we het **finale model**. Omdat we geïnteresseerd zijn in een voorspellend model, hebben we geen aandacht besteed aan de hiërarchie van de termen. We hebben dus toegelaten dat een term van een lagere orde uit het model verdwijnt, ook als de hogere orde term nog aanwezig is. Na de modelselectie onderzochten we met een ANOVA-tabel de relatieve bijdrage van de termen. Voor een goede interpretatie moeten we er wel rekening mee houden dat de volgorde van de termen van belang is. De tabel schat in wat de bijdrage is van een term *nadat* de vorige termen in het model zijn opgenomen (technisch: de ANOVA-tabel geeft de **type I sum of squares**). Een andere volgorde van de termen kan andere resultaten geven. Toch krijgen we hieruit een goed beeld welke termen de grootste bijdrage geven aan het model.

(3) Modeldiagnose (residuanalyse)

Een residu is het verschil tussen een waarneming en het model. We kunnen een residu interpreteren als een schatting van de ruisterm. We mogen dus verwachten dat de residu's min of meer dezelfde kenmerken hebben als de ruis. Of dat het geval is, onderzoeken we met een **residu-analyse**. In essentie gaan we hierbij na of de distributie van de residu's dezelfde eigenschappen heeft als de theoretische distributie van de ruis. Belangrijkste kenmerken van de ruis zijn **statistische onafhankelijkheid** (de afwijkingen ten opzichte van het model zijn puur toeval en beïnvloeden elkaar niet), **homoskedasticiteit** (de variantie van de ruis is overal even groot) en **normaliteit** (de fluctuaties ten opzichte van het model zijn min of meer normaal verdeeld). Aan de hand van grafieken onderzoeken we in hoeverre de verdeling van de residu's aan deze voorwaarden beantwoordt.

Hier weten we a priori dat de ruis bij de veeltermmodellen **heteroskedastisch** is en toeneemt met de grootte van de boom. Tijdens de modelbouw hebben we hiervoor de gepaste maatregelen genomen door te werken met een gewogen regressie of te werken in een logaritmische schaal. Als we met gewone residu's werken (het verschil tussen voorspeld volume en opgemeten volume) zullen we uiteraard deze heterogeniteit opmerken. Maar de eigenlijke vraag die ons interesseert is of we erin geslaagd zijn de variantie te stabiliseren door te werken met de gewichten. Daarom **standaardiseren** of **normaliseren** we eerst de residu's door ze te delen ze delen door een maat die evenredig is met de toename van de spreiding. Aangezien de standaardafwijking recht evenredig is met de boomgrootte, delen we de residu's door de voorspelde waarde van het volume. Zoals ook hoger al aangegeven in (2.2), kunnen we deze residu's ook interpreteren als de **proportionele of procentuele** (als we vermenigvuldigen met 100) **afwijkingen**.

Deze procentuele afwijkingen gebruiken we daarom in de residu analyse en we gaan alle klasieke kenmerken na. Op basis van **QQ-plots** (globaal en opgesplitst) per ecoregio onderzoeken we normaliteit en gaan we na of er geen uitbijters (uitzonderlijke grote waarden) aanwezig zijn in de residu's. Om te controleren of de modelvorm correct is, zetten we deze residu's uit in **residugrafieken** ten opzichte van de omtrek en de hoogte. Als we hierin een patroon ontdekken, dan moeten we naar een betere modelvorm zoeken. Tenslotte verkennen we met **boxplots** van residu's in functie van de bestanden en de ecoregio's of er geen bijzondere bestanden zijn of effecten op schaal van de ecoregio. Vooral dat laatste is van belang omdat we op basis hiervan informatie krijgen of een apart tarief per ecoregio wenselijk is.

(4) De coëfficiënten van het regressiemodel

De tabel met coëfficiënten geeft de geschatte waarden van de parameters of coëfficiënten van de regressiemodellen. Aangezien de termen geselecteerd zijn vanuit hun voorspellende waarde en de modellen eerder een “ad hoc” karakter hebben, is een interpretatie van de termen minder aan de orde. Wel hebben we aandacht besteed aan de coëfficiënt die hoort bij de kegelterm (C2H) omdat we deze coëfficiënt kunnen interpreteren als een **vormfactor**. De andere termen geven aan hoe het kegelmodel verder gecorrigeerd moet worden en kunnen we het best grafisch interpreteren (volgende rubriek).

In het bijzonder heeft de **intercept** van deze modellen weinig betekenis. Het is de geschatte waarde bij een hoogte en omtrek gelijk aan nul. Een redenering zou kunnen zijn, dat we in dat geval, per definitie het intercept nul moet zijn. Maar dat klopt niet omdat het functionele verband tussen de dimensies van een boom en het houtvolume kan veranderen bij kleine bomen. We extrapoleren dus het model naar een bereik waarvoor het niet geschikt is en waarvoor we geen gegevens hebben. Een analyse en interpretatie van het intercept is dus weinig zinvol. Daarom hebben we het intercept aangevuld met **het voorspelde houtvolume van de referentieboom**, die min of meer typisch is voor de boomsoort in het Vlaamse grondgebied. Voor Inlandse eik was de referentieboom $C = 1.5$ m en $H = 25$ m en voor Beuk $C = 1.8$ m en $H = 32.5$ m.

(5) De voorstelling van de modellen (voorspellingen)

In de laatste rubriek geven we een grafische voorstelling van de geschatte modellen. Voor modellen met één ingang is dat heel eenvoudig een curve in functie van de omtrek al dan niet uitgesplitst per ecoregio. Voor tarieven met twee ingangen, moeten we ook rekening houden met de hoogte. Dat realiseren we op twee manieren. Enerzijds maken we grafieken waarbij we de omtrek laten variëren bij een constante hoogte. Maar omdat de hoogte sterk gerelateerd is met de omtrek (naarmate de omtrek stijgt, neemt de hoogte toe), zijn bepaalde combinaties van hoogte en omtrek niet mogelijk. Daarom geven we de curven per hoogte alleen binnen het omtrekbereik dat realistisch is.

De andere optie is een grafiek in functie van de omtrek met het houtvolume berekend met de gemiddelde hoogte die correspondeert met de omtrek. De gemiddelde hoogte hebben we afgeleid met een niet-parametrisch regressiemodel zoals geïmplementeerd in de functie *loess* in S-plus. In notatie (we gebruiken het “tilde” symbool, om aan te geven dat het om een specifieke voorspelling gaat van het volume):

$$\tilde{H} = \text{loess}(C) \rightarrow \tilde{V} = g(C, \tilde{H}) = g(C, \text{loess}(C)) \quad (2.24)$$

Om de waarnemingen aan deze curve toe te voegen, is een omweg nodig. Want de hoogte van de waarnemingen valt niet altijd samen met de verwachte hoogte bij een bepaalde omtrek. Als we de oorspronkelijke waarnemingen toevoegen aan deze figuur, dan krijgen we een verkeerd beeld van de variabiliteit. Maar aangezien het volume volgens de modellen ongeveer lineair toeneemt met de hoogte, kunnen we de waarnemingen corrigeren door volgende lineaire transformatie:

$$V_c \approx \frac{\tilde{H}}{H} V \quad (2.25)$$

Dat is een benadering, maar hiermee filteren we het hoogte-effect voor een groot gedeelte uit de waarnemingen zodat een goed beeld bekomen wordt van de variabiliteit ten opzichte van het model.

2.3.3 Vergelijking van de modellen

In het vorig hoofdstuk werden de modellen apart beoordeeld. In een volgende stap confronteerden we de modellen met elkaar om het meest geschikte model te vinden:

- (1) Ten eerste onderzochten we in hoeverre de voorspellingen van de verschillende modellen van elkaar verschillen.
- (2) Een tweede luik was een vergelijking van de voorspellingskwaliteit van de modellen.

- (3) Ten derde gingen we na wat de invloed was van de meetfout op de hoogte op de kwaliteit van het tarief met twee ingangen om de robuustheid ervan te testen.

(1) Vergelijking van de modelvoorspellingen

We vergeleken de modeltypes uit Tabel 2-1 grafisch twee aan twee, apart voor tarieven met twee ingangen en voor tarieven met één ingang. Als het verschil klein is, dan is de modelkeuze niet kritiek. Het patroon van de verschillen geeft informatie over hoe de vorm van de modellen verschilt. Hierbij kijken we niet meer in detail naar het globaal model met één ingang, omdat uit de vorige stap duidelijk was geworden dat we de tarieven met één ingang moeten opsplitsen per ecoregio.

(2) Vergelijking van de kwaliteit

De voorspellingskwaliteit van de modellen onderzochten we aan de hand van twee criteria. Het eerste criterium was in hoeverre de voorspellingen gemiddeld gezien juist zijn en of de modellen niet systematisch afwijken of vertekend zijn. Hiermee evalueren we **de juistheid van het model**. Het kengetal dat we hiervoor nemen, is de mediaan van de gestandaardiseerde residu's. We verkozen dit kengetal omdat bij een veeltermmodel de gemiddelde waarde van de residu's per definitie gemiddeld nul is en dus is deze maat niet bruikbaar om te evalueren. De mediaan is ook gevoelig voor een eventuele asymmetrie in de gestandaardiseerde residu's.

Het tweede criterium is de precisie van de voorspellingen als maat voor **de variabiliteit van het model**. Hiervoor keken we naar het interval waarbinnen 95% van de gestandaardiseerde residu's zich bevinden (d.w.z. tussen de 2.5% percentiel en het 97.5% percentiel). Dit interval kunnen we interpreteren als het 95% voorspellingsinterval relatief ten opzichte van de geschatte waarde. Als bijvoorbeeld het interval loopt van -20% tot 20%, dan verwachten we de werkelijke waarde in 95% van de gevallen tussen 0.8 en 1.2 keer de voorspelde waarde.

(3) Impact van de meetfout

Tot nu toe veronderstelden we dat we de hoogte zonder meetfout konden meten. Maar in de praktijk is dat niet het geval. Meetfouten tot 10% zijn niet uitzonderlijk om de hoogtemetingen notoir moeilijk zijn. Daarom is het belangrijk te weten hoe robuust de modellen zijn tegen eventuele meetfouten. Om het effect van de meetfout na te gaan op de voorspellingen van de tarieven met twee ingangen, voerden we een simulatie uit, waarbij we aan de hoogte ruis toevoegden volgens volgend model:

$$H_m = H \left(1 + \frac{\varepsilon}{100} \right) \quad \varepsilon \propto \mathcal{N}(0, \sigma_\varepsilon^2) \quad \sigma_\varepsilon = 5, 10, 15 \quad (2.26)$$

Hierbij is H de werkelijke hoogte en H_m de opgemeten hoogte.

De grootte van de ruis neemt bijgevolg proportioneel toe met de hoogte van de boom. Aangezien de ruis normaal verdeeld is, zal H_m in 95% van gevallen (ongeveer) in volgend interval liggen:

$$0 \leq \Delta_H = \left| \frac{H_m - H}{H} \right| \leq \frac{2\sigma_\varepsilon}{100} \quad (2.27)$$

Bijvoorbeeld voor $\sigma_\varepsilon = 10$ situeert de relatieve meetfout Δ_H zich in 95% van de gevallen binnen de 20% van de werkelijke waarde.

Omdat het gemiddelde soms iets meer zegt, berekenden we met een simulatie de gemiddelde waarde en mediaan van Δ_H :

$$E[\Delta_H (\%)] = 0.79\sigma_\varepsilon \quad \& \quad Med[\Delta_H (\%)] = 0.67\sigma_\varepsilon \quad (2.28)$$

Bijgevolg, als $\sigma_\varepsilon = 5, 10$ of 15 , dan is de meetfout op de hoogte ongeveer 4, 8 of 12%. Aangezien we een gemiddelde meetfout van 10% op de hoogte kunnen verwachten, onderzochten we het effect van de meetfout voor deze waarden van σ_ε .

2.4 De tarieven

De laatste stap was de voorstelling en bespreking van de tarieven voor de geselecteerde modellen. Hierbij hebben we volgende punten bekeken:

- (1) Het empirisch bereik van de modellen
- (2) Opstelling van de volumetabellen binnen het empirisch bereik
- (3) Inschatting van de modelfout
- (4) Grafische voorstelling van de tarieven

(1) Het empirisch bereik van de modellen

Modellen zijn nooit perfect en zijn maar geldig binnen een bepaald bereik. Daarom brachten we eerst het meetbereik van de gegevens in kaart door met contourplots duidelijk te maken binnen welke contouren de modellen met gegevens onderbouwd zijn en in hoeverre het meetbereik verschilt naargelang de ecoregio. Want over het algemeen leidt het gebruik van een model binnen het empirisch bereik tot relatief weinig problemen. De voorspellingen zijn dan interpolaties en zelfs wanneer het model niet volledig juist is, blijven de fouten over het algemeen klein, zeker als de waarnemingen gelijkmatig verspreid liggen over het volledig waarnemingsgebied. Maar naarmate we ons verder buiten het waarnemingsgebied wagen, hebben we veel minder garantie dat het model daar ook van toepassing is en neemt de modelonzekerheid toe. Gebruik van het modellen in dat gebied is niet verboden, maar we moeten ons wel van potentiële problemen bewust zijn.

(2) Berekening van de volumetabellen

Tot net over de grens van de hierboven afgeleide contouren hebben we de volumetabellen (tarieven) berekend. Voor de tarieven met twee ingangen bekeken we alleen plausibele combinaties van de omtrek en hoogte. De omtrek varieerde van 0.35 m tot 4 m (voor Beuk) of 3 m (voor Eik) in stappen van 0.1 m en de hoogte van 10 m tot 50 m (voor Beuk) of 40 (voor Eik) in stappen van 2.5 m. Deze punten beschouwen we als het centrum van het interval waarvoor de volumes geldig zijn. Dus voor de omtrek reikt elk interval van -0.05 m tot +0.05 m ten opzichte van het centrum en voor de hoogte is dat van -1.25 m tot 1.25 m. Deze rudimentaire tabellen zijn bedoeld als een eerste aanzet voor de meer definitieve tabellen. Dezelfde oefening maakten we voor de tarieven met één ingang, waarbij we het omtrekbereik lieten afhangen van de ecoregio. In het bijzonder in het Kempens district is het bereik heel smal. Uiteraard kunnen we altijd met de formules extrapoleren, maar om de redenen hierboven uiteengezet, is dat niet zonder risico.

(3) Modelonzekerheid

De tabellen geven voor een combinatie van omtrek en hoogte (tarieven met twee ingangen) of voor een combinatie van omtrek en ecoregio (tarieven met twee ingangen) het volume. Voor elke combinatie waarvoor we waarnemingen hebben, berekenden we het gemiddelde volume en berekenden we het procentuele verschil met het voorspelde volume:

$$A = \frac{V_{\text{gem}} - \hat{V}_{\text{tarief}}}{\hat{V}_{\text{tarief}}} \% \quad (2.29)$$

Deze procentuele afwijkingen gelijken op gestandaardiseerde residu's. Het verschil is dat we niet werken met individuele waarnemingen, maar met de gemiddelde waarde voor een welbepaalde combinatie van hoogte en omtrek.

Met deze procentuele afwijkingen onderzochten we of er zich geen patroon aftekende in functie van de omtrek of de hoogte. In het bijzonder waren we geïnteresseerd in hoeverre het model ook aan de randen van het empirisch bereik nog goed scoorde. Hiertoe maakten we een tweedimensionele contourgrafiek om de relatieve nauwkeurigheid in beeld te brengen.

(4) Grafische voorstelling van de tarieven

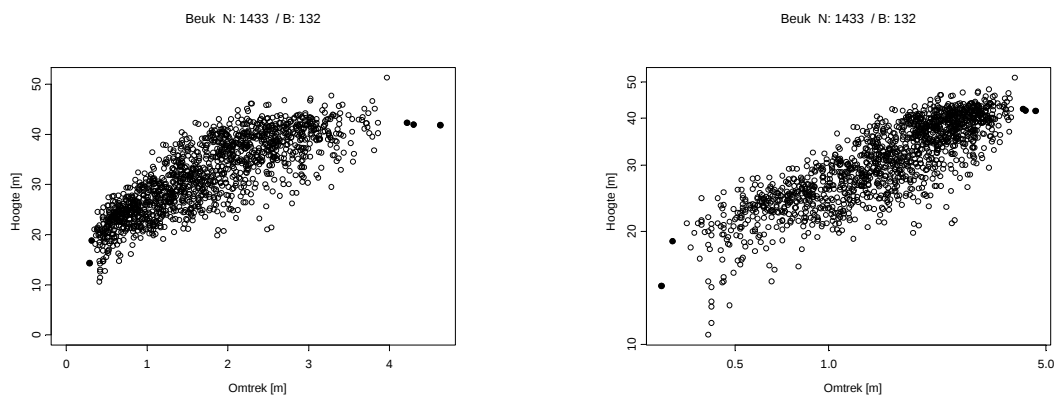
De grafieken namen we over uit het deel modelanalyse waar voor elk model een grafische voorstelling gemaakt werd.

3 Beuk: verkenning van de gegevens (EDA)

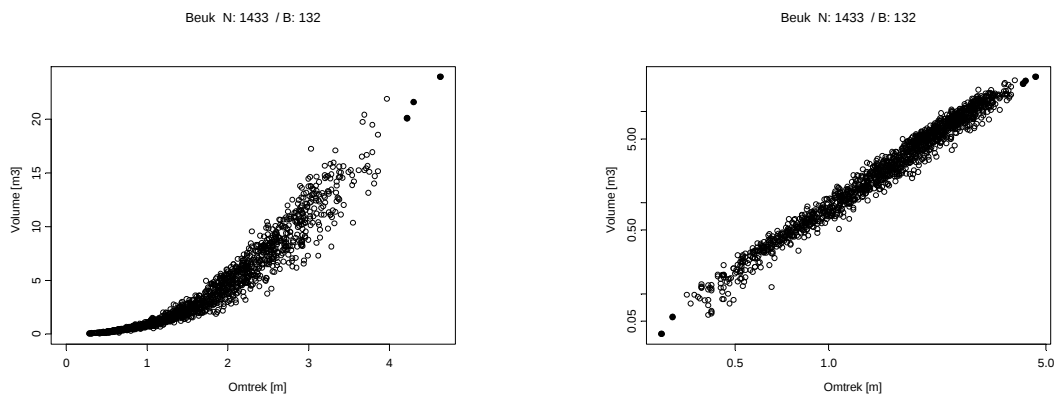
3.1 Het meetbereik en de selectie van de gegevens

Het bereik en de precisie van de statistische modellen is sterk bepaald door het meetbereik en het aantal gegevens. Figuur 3-1 en Figuur 3-2 geven hiervan een eerste beeld. In totaal gaat het om 1433 waarnemingen (N) gespreid over 132 bestanden (B). Vijf punten (in het zwart) liggen buiten het omtrekbereik 0.35m tot 4m. Deze punten hebben we niet meegenomen bij de modellering van de tarieven. Het zijn hefboom punten. Ze liggen een stuk afgezonderd van het normale bereik en hebben potentieel een grote invloed op de coëfficiënten.

Deze keuze verandert niets aan de reikwijdte van het model. Ongeacht we deze punten meenemen of niet, de tarieven zijn geldig voor omtrekken tussen 0.35 m en 4 m. Extrapolatie naar hogere waarden kan, maar we hebben geen gegevens om dat te controleren. Een gerichte meetcampagne is nodig om deze lacune op te vullen. Wel zullen we in de tabellen tarieven geven tot even buiten dat interval, maar de gebruiker moet rekening houden met een grotere (relatieve) fout omdat we buiten het interval de vorm van het model niet konden verifiëren. Wel zullen we op het einde controleren in hoeverre de modellen de weggelaten punten goed voorstellen.



Figuur 3-1 Beuk: verkenning van de relatie tussen omtrek en hoogte. Links: originele schaal, rechts: log-schaal. De punten in het zwart geven de waarnemingen aan buiten het bereik van 0.35 m tot 4 m.



Figuur 3-2 Beuk: verkenning van de relatie tussen omtrek en volume. Links: originele schaal, rechts: log-schaal. De punten in het zwart geven de waarnemingen aan buiten het bereik van 0.35 m tot 4 m.

3.2 Opbouw van de dataset

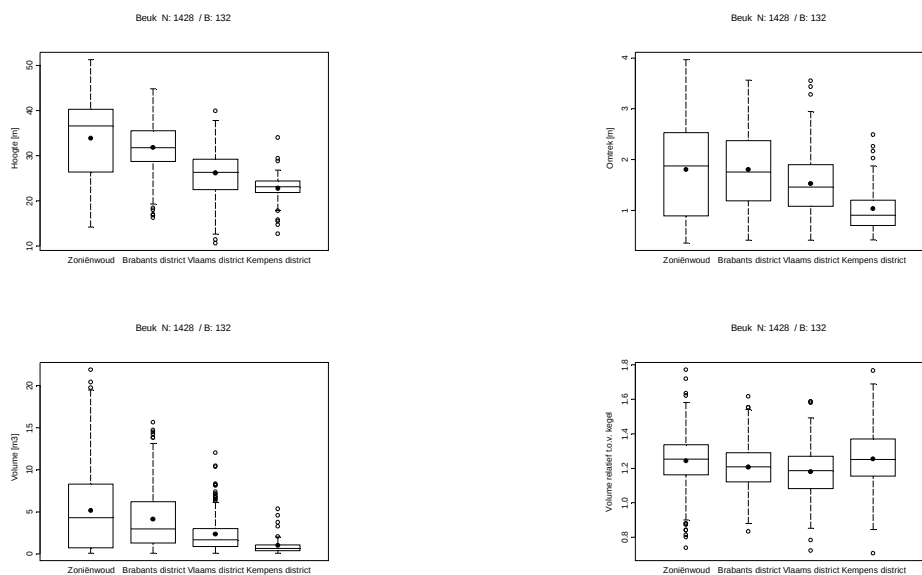
Na het weglaten van de vijf waarnemingen buiten het interval 0.35 m tot 4 m blijven 1428 metingen van bomen over verspreid over 132 bestanden (N: 1428 / B:132). Tabel 3-1 maakt een overzicht van enkele kenmerken van de dataset. De meeste bestanden (59) komen uit Zoniën. Over het Kempens district hebben we het minst informatie (7 bestanden en 4% van het aantal waarnemingen) en daarenboven is het omtrekbereik er heel beperkt (maar tot 2,50 m). Het model in het Kempens district zal weinig precies zijn met een beperkte geldigheid. Daarom zullen we bij de bespreking dit district dikwijls terzijde laten of apart houden. Ook van het Vlaamse district hebben we niet zoveel informatie (26 bestanden of 16 %), maar het bereik is wel een stuk groter (tot 3.55 m). Daarom zullen we hier toch een nauwkeurig model kunnen afleiden. Zoniën en het Brabants district bevatten respectievelijk 52 % en 28 % van de bomen.

We kunnen de ecoregio's duidelijk ordenen volgens grootte (Tabel 3-1, Figuur 3-3, Figuur 3-4). De grootste bomen komen voor in het Zoniënwoud, dan volgt het Brabants district en het Vlaamse district. Als we kijken naar de 95% percentiel van de hoogte (om eventuele outliers te elimineren), reikt de hoogte van de bomen in het Zoniënwoud tot 43.4 m, terwijl in het Brabantse district en het Vlaamse district dat respectievelijk 39.4 m en 34.7 m is. Alle andere maten wijzen in dezelfde richting.

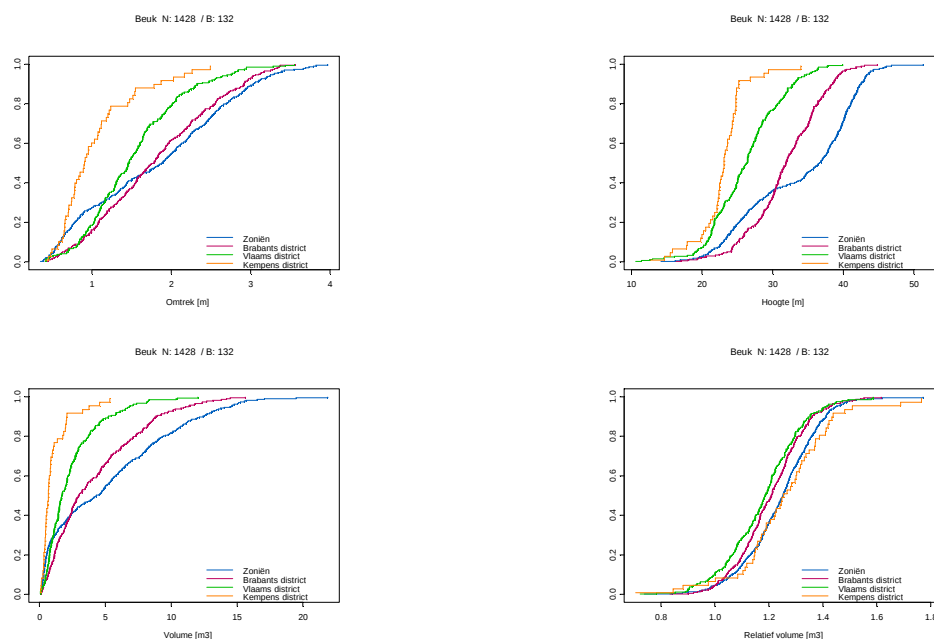
Een aparte bespreking is nodig voor het relatieve volume (het volume gedeeld door het kegel-volume). Deze maat is geen dimensie, maar een verhouding en kunnen we interpreteren als een vormfactor. Hier ligt de waarde voor de Kempen het hoogst, gevolgd door Zoniën, het Brabants district en het Vlaams district. Aangezien we in de Kempen alleen het kleinere segment van bomen vinden, is een mogelijke reden dat het relatieve volume daalt naarmate de boom groter wordt (zie Figuur 3-8). Als we het Kempens district buiten beschouwing laten, blijkt dat in het Zoniënwoud niet alleen de grootste bomen staan, maar voor vergelijkbare bomen verhoudingsgewijs het houtvolume groter is dan in de andere districten. Dus ook de vormfactor speelt een rol.

Tabel 3-1 Beuk: enkele richtgetallen, alle districten samen en opgesplitst per district.

	District				
	Alle	Zoniën	Brabants	Vlaams	Kempens
Aantallen					
+ metingen (bomen)	1428	746	399	229	54
+ percentage metingen	100%	52%	28%	16%	4%
+ bestanden	132	59	40	26	7
+ bomen/bestand (mediaan)	6	8	5	5	3
+ bomen/bestand (bereik)	1 – 88	1 – 71	1 – 88	1 – 37	1 – 31
Omtrek [m]					
+ bereik	0.35 – 3.97	0.35 – 3.97	0.41 – 3.56	0.41 – 3.55	0.42 – 2.49
+ 5% - 95% percentiel	0.53 – 3.14	0.52 – 3.27	0.64 – 3.09	0.69 – 2.73	0.49 – 2.07
+ mediaan / gemiddelde	1.65 / 1.73	1.88 / 1.80	1.76 / 1.80	1.46 / 1.53	0.90 / 1.03
Hoogte [m]					
+ bereik	10.6 – 51.3	14.2 – 51.3	16.3 – 44.8	10.6 – 39.9	12.7 – 34.0
+ 5% / 95% percentiel	20.6 – 42.5	21.0 – 43.4	23.6 – 39.4	19.0 – 34.7	15.7 – 27.5
+ mediaan / gemiddelde	31.4 / 31.6	36.6 / 33.8	31.8 / 31.8	26.3 / 26.2	23.1 / 22.8
Volume [m³]					
+ bereik	0.06 – 21.9	0.06 – 21.9	0.09 – 15.6	0.06 – 12.0	0.08 – 5.34
+ 5% / 95% percentiel	0.20 – 12.7	0.18 – 14.2	0.33 – 11.0	0.33 – 6.74	0.12 – 3.43
+ mediaan / gemiddelde	2.71 / 4.25	4.32 / 5.14	2.96 / 4.11	1.67 / 2.35	0.66 / 1.01
Relatief volume [-]					
+ bereik	0.707 – 1.772	0.739 – 1.772	0.833 – 1.616	0.723 – 1.587	0.707 – 1.766
+ 5% / 95% percentiel	0.990 – 1.436	1.012 – 1.445	1.005 – 1.421	0.932 – 1.402	0.943 – 1.491
+ mediaan / gemiddelde	1.227 / 1.223	1.252 / 1.242	1.208 / 1.207	1.186 / 1.180	1.251 / 1.254



Figuur 3-3 Beuk: boxplots van omtrek, hoogte, volume en relatief volume. De punten geven het gemiddelde aan.



Figuur 3-4 Beuk: cumulatieve distributies van omtrek, hoogte, volume en relatief volume.

3.3 De referentieboom en het referentiebereik per ecoregio

Bij de bespreking van de resultaten is het interessant om een referentieboom te definiëren waarvoor we het volume berekenen met de verschillende modellen. Op basis hiervan is een eerste elementaire bespreking mogelijk. Aangezien de standplaats heel sterk de dimensies van de boom beïnvloed, hebben we per ecoregio een apart bereik waar de mediaan van de omtrek en de hoogte liggen (Tabel 3-1). Om later eenvoudig te kunnen refereren, hebben we deze gegevens samengebracht in Tabel 3-2. Voor elke ecoregio apart geven we telkens een omtrek- en hoogtebereik en voor het volume gebruiken we de mediaan als een indicatief cijfer. De gezamenlijke referentieboom voor alle ecoregio's samen is gebaseerd op de karakteristieken voor het Zoniënwood en het Brabant district waarvan 75 % van de gegevens komen. **De dimensies van deze referentieboom zijn C = 1.8 m en H = 32.5 m.** Voor het Kempen district is de

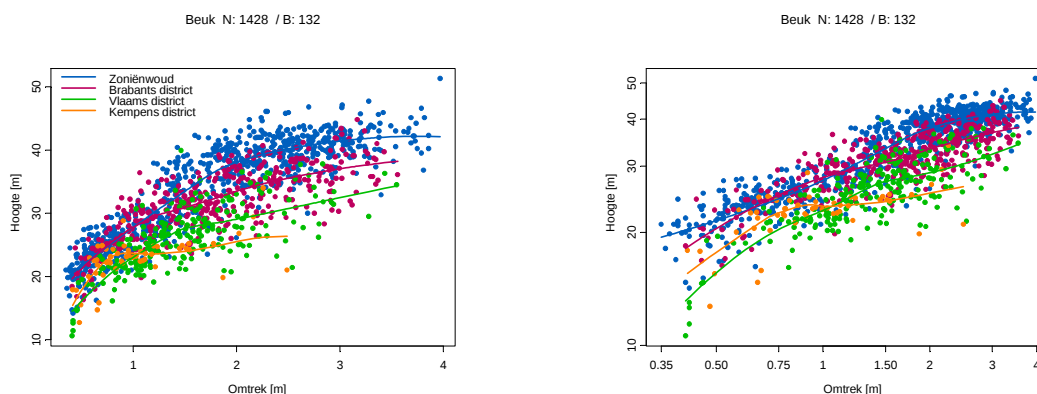
referentieboom niet alleen bijzonder groot, maar ook disproportioneel hoog. Dat geldt ook (zij het in mindere mate) voor het Vlaams district.

Tabel 3-2 Beuk: referentieboom en referentiebereik per ecoregio

	Ecoregio				
	Alle	Zoniën	Brabants	Vlaams	Kempens
Omtrek [m]	1.8	1.8 – 1.9	1.7 – 1.8	1.4 – 1.5	0.9 – 1.0
Hoogte [m]	32.5	35.0 – 37.5	30.0 – 32.5	25.0 – 27.5	22.5 – 25.0
Volume [m³]		4.32	2.96	1.67	0.66

3.4 Relatie tussen omtrek en hoogte

De hoogte vlakt af bij hoge waarden van de omtrek, ook in de log-schaal. Tussen de ecoregio's zijn er grote verschillen. Het grillig verloop in het Kempens district heeft te maken met de kleine steekproefgrootte. Als we dat district buiten beschouwing laten, is er een strikte ordening tussen de ecoregio's: Zoniën > Brabants district > Vlaams district. Bij een gelijke omtrek zijn de bomen het hoogst in het Zoniënwoud, dan komt het Brabants district, vervolgens het Vlaamse district en ten slotte het Kempens district (behalve voor lage waarden van de omtrek). Als we een model bouwen met alleen omtrek als variabele (tarief met één ingang), zullen we dus best opsplitsen per ecoregio.

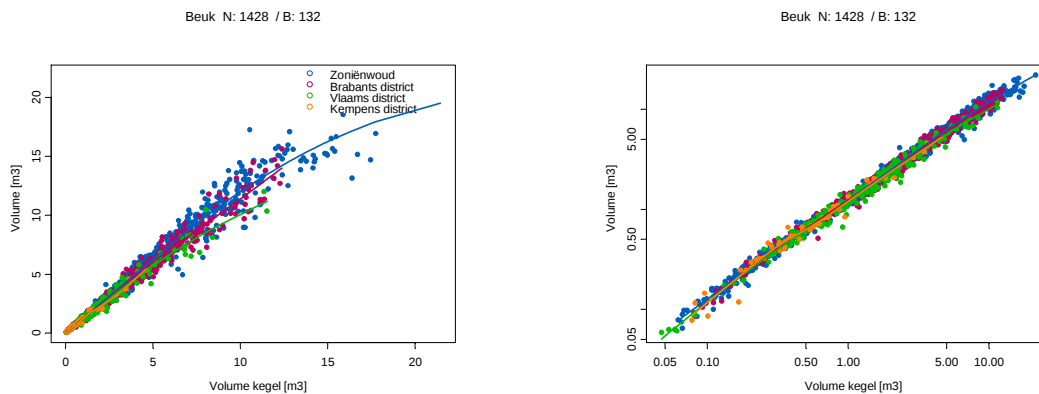


Figuur 3-5 Beuk: relatie tussen omtrek en hoogte (links: originele schaal, rechts: log-schaal).

3.5 Verkenning van het model met twee ingangen

De modellen met twee ingangen zijn geïnspireerd op het kegelvolume. Hier verkennen we in hoeverre deze modellen plausibel zijn door na te gaan hoe de relatie is tussen het werkelijke volume en het volume van een kegel is. In de originele schaal (Figuur 3-6, figuur links) zien we een haast lineair verband die afbuigt naar beneden bij grotere bomen. Een model met een term evenredig met het volume van een kegel zal dus al tot een heel goede voorspelling leiden en wellicht zal er weinig correctie nodig zijn.

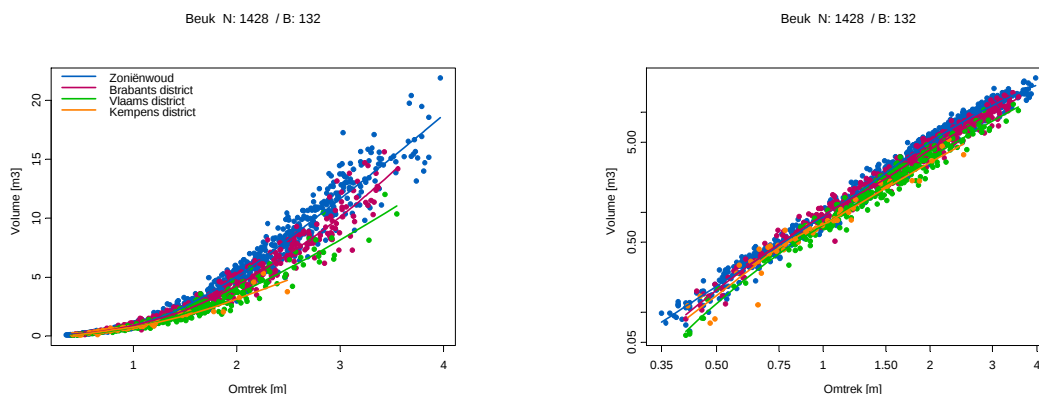
Er zijn wel verschillen tussen de ecoregio's. Het Zoniënwoud ligt hoger dan de rest en voor het Vlaams district buigt de curve plots af naar beneden. Maar het aantal gegevens daar is vrij beperkt en toeval kan een grote rol spelen. Van groter belang is dat de spreiding toeneemt met de grootte van de boom. Een gewogen lineaire regressie zal nodig zijn. In de log-schaal (Figuur 3-6, figuur rechts) is het verband nagenoeg lineair en is de ruis homogeen geworden. De log-transformatie heeft de variantie gestabiliseerd. Bijgevolg is in de log-schaal een gewone lineaire regressie zonder gewichten voldoende.



Figuur 3-6 Beuk: relatie tussen omtrek en het volume relatief ten opzichte van het volume van een kegel (links: originele schaal, rechts: log-schaal).

3.6 Verkenning van het model met één ingang (en ecoregio)

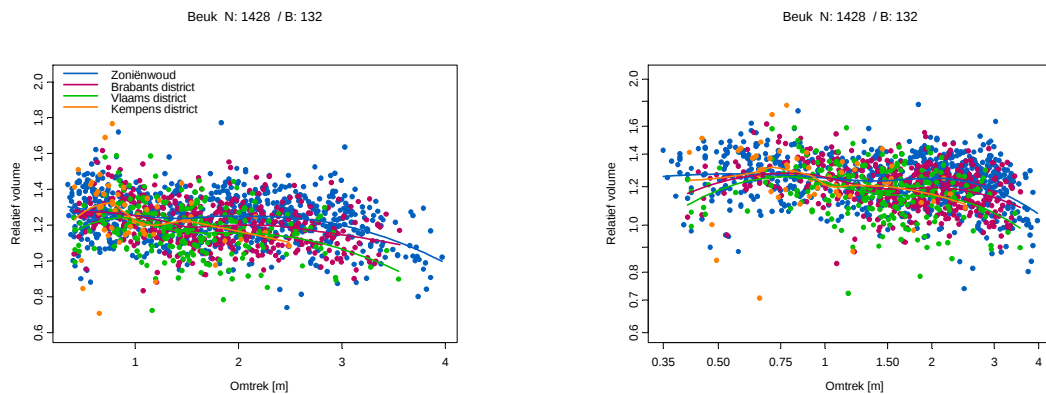
De relatie tussen omtrek en volume is nagenoeg lineair in de log-schaal. De variantie neemt toe met het volume, maar in de log-schaal is de variantie gestabiliseerd. De ordening van de curven is zoals in de vorige rubriek. Figuur 3-2 is een duidelijke indicatie dat de tarieven met één ingang zullen verbeteren door ze op te delen naargelang de ecoregio.



Figuur 3-7 Beuk: relatie tussen omtrek en volume (links: originele schaal, rechts: log-schaal).

3.7 Verkenning van het relatief volume

Het relatief volume ten opzichte van een kegel neemt af naarmate de omtrek toeneemt. Toch is de waarde in het interval van 1m tot 3m vrij stabiel in het Zoniënwoud en het Brabants district, met een waarde van ongeveer 1.2. Dat kunnen we ook afleiden uit de boxplots (Figuur 3-3). Opsplitsing per district geeft een duidelijke ordening: Zoniënwoud > Brabants district > Vlaams district. Het Kempen district past niet in het rijtje maar heeft een (te) beperkt aantal waarnemingen om hieruit definitieve conclusies te kunnen trekken.



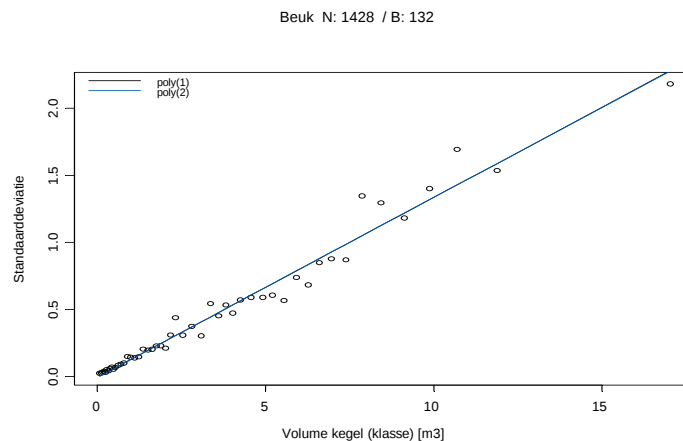
Figuur 3-8 Beuk: relatie tussen omtrek en het volume relatief ten opzichte van het volume van een kegel (links: originele schaal, rechts: log-schaal).

3.8 Variantie van het volume in functie van de omtrek

Een belangrijke voorwaarde voor lineaire regressie is dat de variantie van de ruis gelijk is (homoskedasticiteit) in functie van de waarde voorspeld door het model. Dat is hier duidelijk niet het geval. Een mogelijke oplossing is een gewogen regressie toe te passen waarbij de gewichten omgekeerd evenredig zijn met de variantie. Om de variantiefunctie te bepalen die de relatie beschrijft tussen de voorspelde waarde en de variantie, onderzoeken we hier de relatie tussen de standaardafwijking en het volume van de kegel (als benadering van de voorspelling die we hier nog niet kennen). Hiertoe deelden we het bereik van het volume van de kegel (tussen 0.05 m³ en 21.9 m³) op in 50 deelintervallen met ongeveer evenveel waarnemingen. In elk deelinterval bepalen we de standaardafwijking. Uit Figuur 3-9 en Tabel 3-3 blijkt dat de relatie tussen de standaardafwijking en het kegelvolume in goede benadering lineair is. Aangezien we de gewichten omgekeerd evenredig moeten kiezen met de variantie, kunnen we ze gelijk stellen één op het kwadraat van het kegelvolume. Hierdoor zal de impact van de variatie van de grotere bomen afgezwakt worden.

Tabel 3-3 Beuk: ANOVA-tabel voor een veeltermmodel van de standaardafwijking in functie van het volume van de kegel. De term poly(Vkegel,2) is een veelterm (polynoom) van de tweede graad. Aangezien de component van de tweede graad niet significant is en de component van de eerste graad wel, is de conclusie dat het verband tussen de standaardafwijking en de het volume van een kegel lineair is.

	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.477	0.011	42.529	0.000
poly(Vkegel, 2)1	3.469	0.079	43.718	0.000
poly(Vkegel, 2)2	-0.004	0.079	-0.054	0.957



Figuur 3-9 Beuk: relatie tussen de variantie van de waarnemingen en het volume van een kegel (in klassen van 0.25 m³).

4 Beuk: statistische analyse van de modellen

4.1 De veeltermmodellen

4.1.1 Veeltermmodel met twee ingangen

Startmodel

Het veeltermmodel met twee ingangen is in essentie een term evenredig met het volume van een kegel aangevuld met termen die functie zijn van de omtrek en de hoogte en de interactie tussen beide. We kunnen dit model neerschrijven in twee totaal equivalente varianten. Het enige verschil is dat de coëfficiënt van V_{kegel} een factor 12π groter zal zijn dan $C2H$.

$$P2 \Leftrightarrow V \sim \begin{cases} C2H + C + C2 + C3 + H + CH & (\text{var1}) \\ V_{kegel} + C + C2 + C3 + H + CH & (\text{var2}) \end{cases} \quad V_{kegel} = \frac{C2H}{12\pi} \quad (4.1)$$

Interessant is het "referentiemodel" waarbij we alleen rekening houden met V_{kegel} . Hieruit kunnen we de (vorm)factor afleiden waarmee we het volume van een kegel gemiddeld moeten vermenigvuldigen om het werkelijke houtvolume te bekomen.

$$P2r \Leftrightarrow V \sim V_{kegel} \quad (4.2)$$

Modelselectie (Tabel 4-1)

Uit een stapsgewijze regressie blijkt dat alleen H en C niet significant zijn ($p = 0.684$). Aangezien we op de eerste plaats een voorspellend model beogen, kunnen we deze lagere orde termen weglaten en het finaal model wordt:

$$V \sim C2H + C2 + C3 + CH \quad (4.3)$$

Uit de finale ANOVA-tabel (Tabel 4-1) blijkt dat een keer de term evenredig met het volume van de kegel is ingevoerd, de andere termen niet meer veel toevoegen (bekijk de sum of squares).

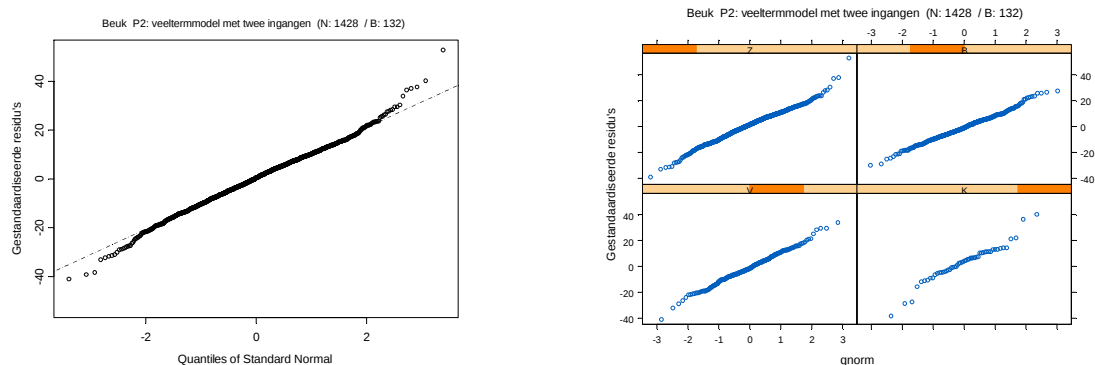
Tabel 4-1 Beuk: ANOVA-tabel van het finaal veeltermmodel met twee ingangen.

	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
C2H	1	1578	1578	91725	0.0000
C2	1	0	0	6	0.0152
C3	1	3	3	153	0.0000
CH	1	0	0	8	0.0052
Residuals	1423	24	0		

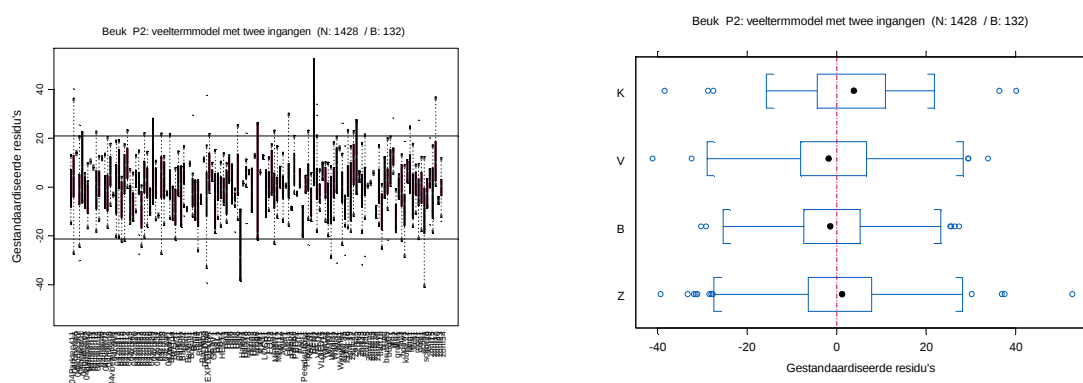
Residuanalyse (Figuur 4-1, Figuur 4-2 en Figuur 4-3).

De diagnostische grafieken op basis van de residu's vertonen geen structuur. De QQ-plot wijkt alleen in de extremen af van een rechte. Het model is niet perfect, maar sluit goed aan bij de gegevens. Ook zien we geen patronen in functie van het district. We hebben geen aanwijzingen dat een apart tarief met twee ingangen nodig is naargelang het district.

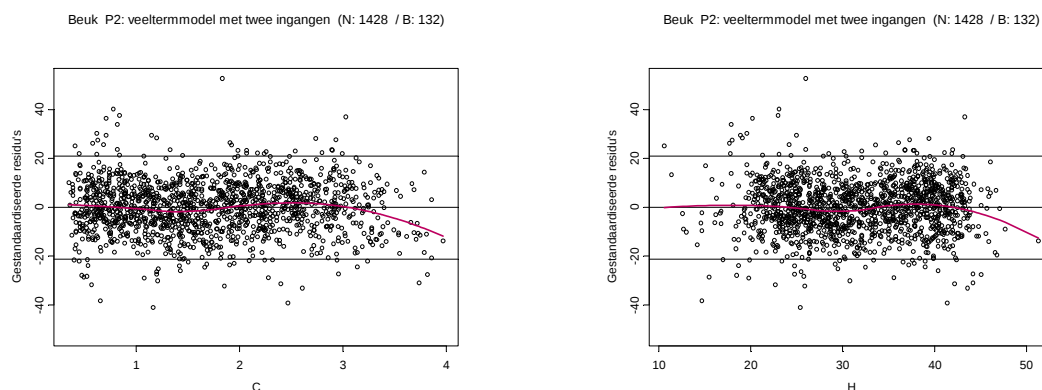
De gestandaardiseerde residu's zijn homoskedastisch. De relatieve fout is homogeen over het volledige bereik. Dit resultaat bevestigt dat spreiding recht evenredig is met het volume en dat de gewogen regressie met gewichten omgekeerd evenredig met het kwadraat van de kegelvolumes de variantie stabiliseert. Vijfennegentig procent van de gestandaardiseerde residu's valt tussen $\pm 20\%$. Dat impliceert een 95% voorspellingsinterval tussen 0.8 en 1.2 keer het geschatte volume.



Figuur 4-1 Beuk: residuanalyse voor het veeltermmodel met twee ingangen (P2): QQ-grafieken



Figuur 4-2 Beuk: residuanalyse voor het veeltermmodel met twee ingangen (P2): boxplots.



Figuur 4-3 Beuk: residuanalyse voor het veeltermmodel met twee ingangen (P2): Relatie met omtrek en hoogte.

Coëfficiënten van het model (Tabel 4-2, Tabel 4-3 en Tabel 4-4)

Eerst nemen we in het model alleen een term evenredig met het volume van kegel op. Het houtvolume is 1.21 keer het volume van een kegel (Tabel 4-2). Het houtvolume voor de referentieboom ($C = 1.8$ m en $H = 32.5$ m) is 3.40 m^3 .

Tabel 4-2 Eik: coëfficiënten van het veeltermmodel met alleen kegel

	Value	Std. Error	t value	Pr(> t)
Referentie:	3.40			
Vkegel	1.21	0.00423	286.86	0.00000

Door de andere termen in te voeren, verandert de coëfficiënt van Vkegel (Tabel 4-3). De reden is dat de toegevoegde termen met het volume van kegel gecorreleerd zijn. Een deel van de

voorspellende waarde van Vkegel wordt overgenomen door de andere termen. De impact op de referentieboom is echter klein. Het voorspelde houtvolume is afgerond opnieuw 3.40 m³.

Tabel 4-3 Beuk: coëfficiënten van het veeltermmodel met twee ingangen (variant 1: met Vkegel en referentie)

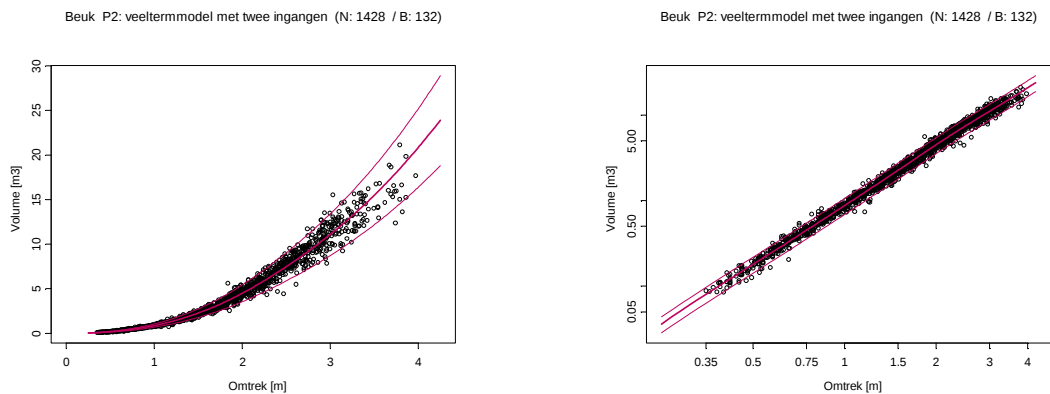
	Value	Std. Error	t value	Pr(> t)
Referentie:	3.39546			
Vkegel	1.36968	0.037915	36.13	0.00000
C2	-0.08560	0.019347	-4.42	0.00001
C3	-0.04996	0.011844	-4.22	0.00003
CH	0.00256	0.000914	2.80	0.00522

Tabel 4-4 Beuk: coëfficiënten van het veeltermmodel met twee ingangen (variant 2: met C2H en intercept)

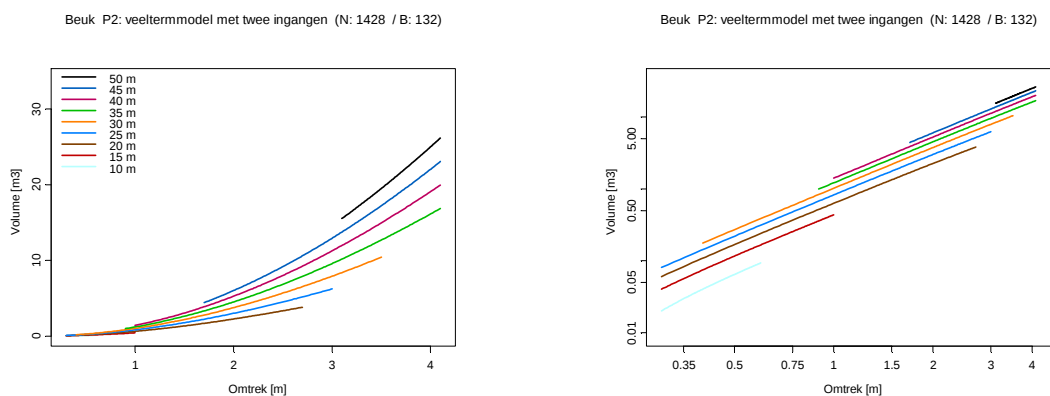
	Value	Std. Error	t value	Pr(> t)
(Intercept)	-0.01115	0.004438	-2.51	0.01212
C2H	0.03633	0.001006	36.13	0.00000
C2	-0.08560	0.019347	-4.42	0.00001
C3	-0.04996	0.011844	-4.22	0.00003
CH	0.00256	0.000914	2.80	0.00522

Voorstelling van het model (Figuur 4-4).

Over het volledige bereik blijkt het regressiemodel goed overeen te stemmen met de gegevens.



Figuur 4-4 Beuk: voorstelling van het veeltermmodel met twee ingangen (P2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de gestandaardiseerde residu's liggen.



Figuur 4-5 Beuk: grafische voorstelling van het veeltermmodel met twee ingangen (P2) per hoogteklasse (rechts in de log-schaal).

4.1.2 Veeltermmodel met één ingang

Startmodel

Het veeltermmodel met één ingang bekomen we door uit model (4.1) alle termen met hoogte erin te laten vallen:

$$P1 \leftrightarrow V \sim C + C2 + C3 = \text{poly}(C, 3) \quad (4.4)$$

Modelselectie (Tabel 4-5)

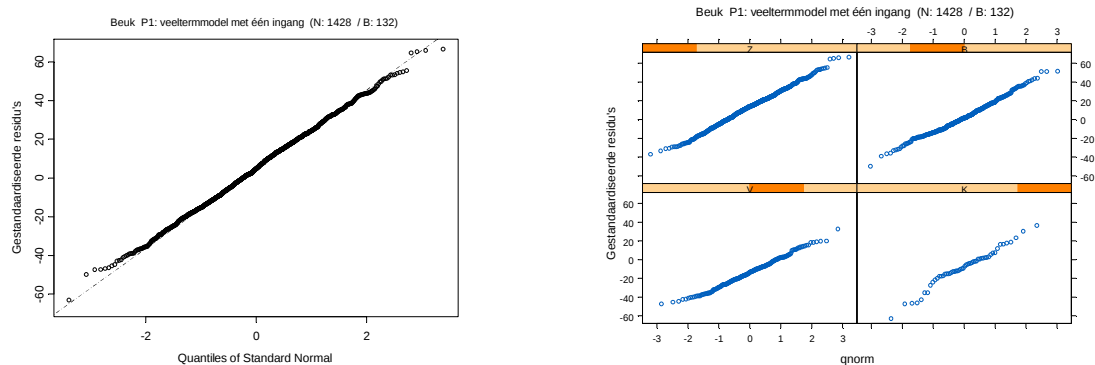
Alle termen van de derde orde polynoom zijn significant. De derdeorde term heeft verhoudingsgewijs maar een kleine inbreng.

Tabel 4-5 Beuk: ANOVA-tabel van het veeltermmodel met één ingang.

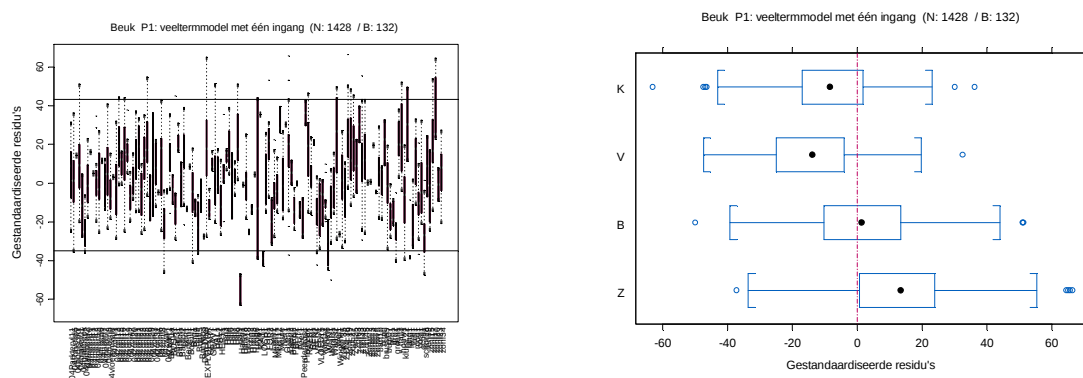
	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
C	1	1217	1217	20314	0.00e+000
C2	1	300	300	5003	0.00e+000
C3	1	3	3	47	1.29e-011
Residuals	1424	85	0		

Residuanalyse (Figuur 4-6, Figuur 4-7 en Figuur 4-8).

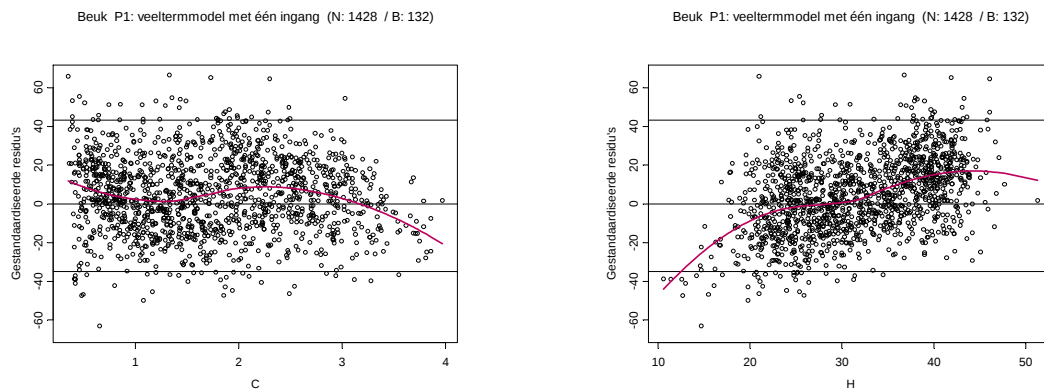
Ongeveer 95 % van de gestandaardiseerde residu's liggen nu in het interval ± 40 %. Het 95 % voorspellingsinterval is bijgevolg verdubbeld. Zoals verwacht is er een sterk correlatie van de residu's met de hoogte. Uit de relatie van de residu's ten opzichte van ecoregio blijkt dat deze variabele het model eventueel kan verbeteren.



Figuur 4-6 Beuk: residuanalyse voor het veeltermmodel met één ingang (P1). QQ-plots.



Figuur 4-7 Beuk: residuanalyse voor het veeltermmodel met één ingang (P1). Boxplots.



Figuur 4-8 Beuk: residuanalyse voor het veeltermmodel met één ingang (P1). Relatie met omtrek en hoogte.

Coëfficiënten van het model (Tabel 4-6)

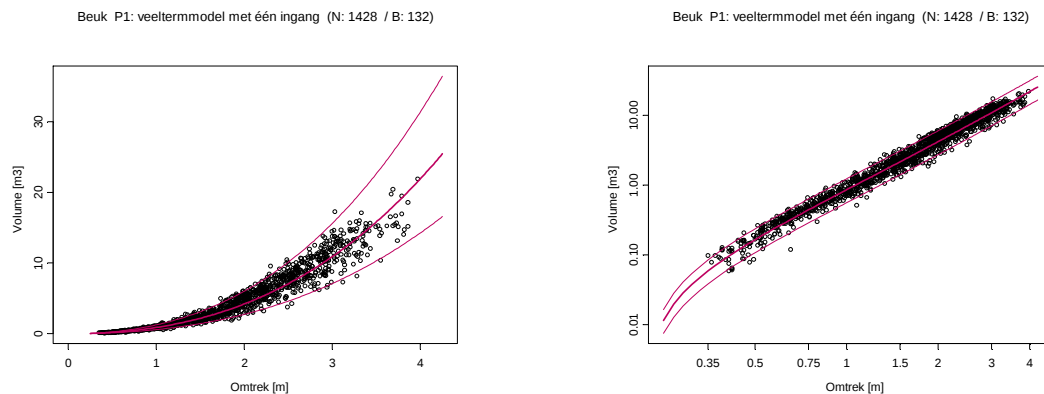
De derdeorde term is positief. Deze term is wellicht een substituuat voor $C2H$ in het veeltermmodel met twee ingangen. Want de hoogte H neemt toe met de omtrek C . Als we veronderstellen dat het verband in eerste benadering lineair is, dan wordt de term $C2H$ de term $C3$. Uiteraard klopt dat niet helemaal want de relatie tussen hoogte en omtrek vlakt af. Maar binnen het bereik van het model zal dat wellicht niet tot problemen leiden. Maar deze redenering toont wel aan dat we moeten oppassen bij een extrapolatie.

Tabel 4-6 Beuk: coëfficiënten van het veeltermmodel met één ingang.

	Value	Std. Error	t value	Pr(> t)
Referentie: (Intercept)	3.2820			
C	-0.0325	0.0250	-1.30	1.94e-001
C	-0.0228	0.0876	-0.26	7.95e-001
C2	0.7556	0.0856	8.83	0.00e+000
C3	0.1556	0.0228	6.83	1.29e-011

Voorstelling van het model (Figuur 4-9).

De globale overeenstemming tussen waarnemingen en model is goed.



Figuur 4-9 Beuk: grafiek van het veeltermmodel met één ingang (P1). Links = originele schaal & rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de gestandaardiseerde residu's liggen.

4.1.3 Veeltermmodel met één ingang en ecoregio

Startmodel

Vraag is in hoeverre we het model niet beter opsplitsen naargelang ecoregio. Hiertoe testen we volgend interactiemodel uit:

$$P1f \leftrightarrow V \sim (C + C2 + C3) * Fyto regio \quad (4.5)$$

Modelselectie (Tabel 4-7)

De ANOVA-tabel toont aan dat Ecoregio significant is en ook een hoge bijdrage aan het model levert. Het is een sterke voorspeller. We hebben er dus voordeel mee apart een model op te stellen per ecoregio (Tabel 4-8).

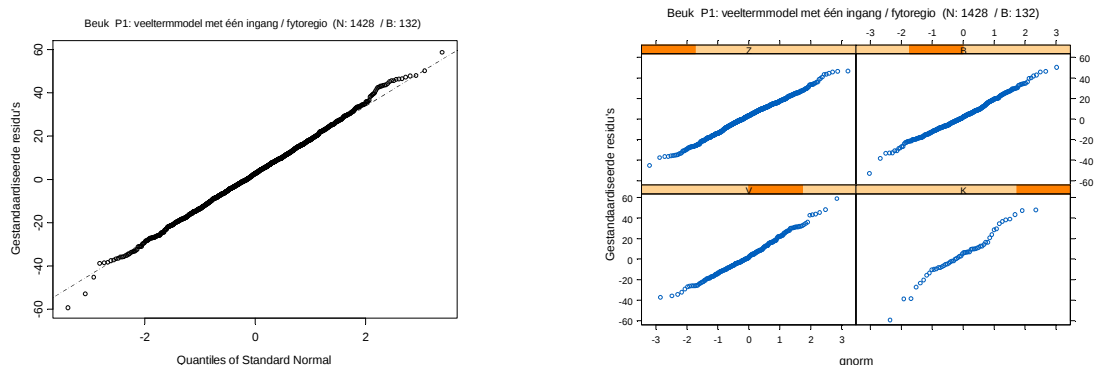
Tabel 4-7 Beuk: ANOVA-tabel van het veeltermmodel met één ingang en ecoregio.

	Terms	Resid.	Df	RSS	Df	Sum of Sq	F Value	Pr(>F)
1	C + C2 + C3		1424	85.3				
2	(C + C2 + C3) * Ecoregio		1412	56.0	12	29.3	61.6	0

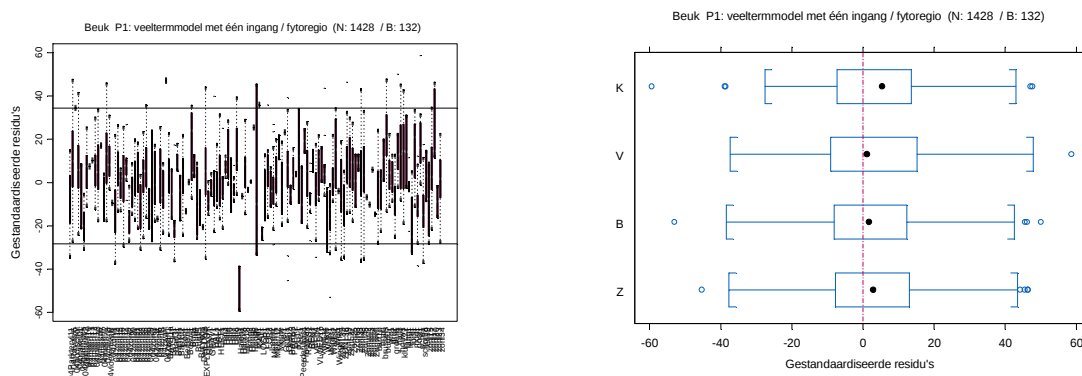
Residuanalyse (Figuur 4-10, Figuur 4-11 en Figuur 4-12)

Ten opzichte van het model met één ingang zonder ecoregio blijft de distributie van de gestandaardiseerde residu's min of meer gelijk. De correlatie van de residu's met de hoogte is iets afgezwakt zoals te verwachten was. Het voorspellingsinterval daalt tot ongeveer 35 % voor Zoniën, het Brabants en het Vlaamse district. In de Kempen blijft de voorspellingsmarge groter dan 40 %.

In het Brabants en het Vlaams district is de relatie tussen de gestandaardiseerde residu's en de omtrek heel egaal. Daar is het model dus goed bruikbaar. In Zoniën is er opnieuw een indicatie van overschatting bij grote waarde van de omtrek ($C > 3$ m). In het Kempens district is er een afgetekend patroon, maar dat betekent wellicht weinig, gezien het beperkt aantal gegevens en het beperkte omtrekbereik.

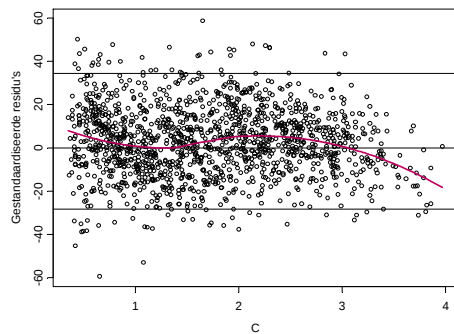


Figuur 4-10 Beuk: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). QQ-plots.

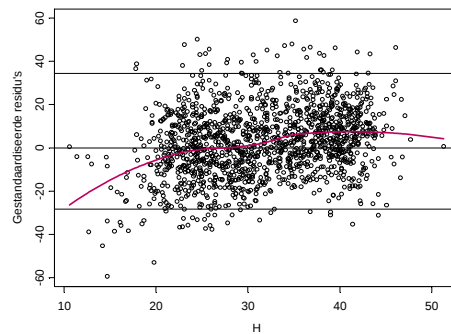


Figuur 4-11 Beuk: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). Boxplots

Beuk P1: veeltermmodel met één ingang / fytoecoregio (N: 1428 / B: 132)

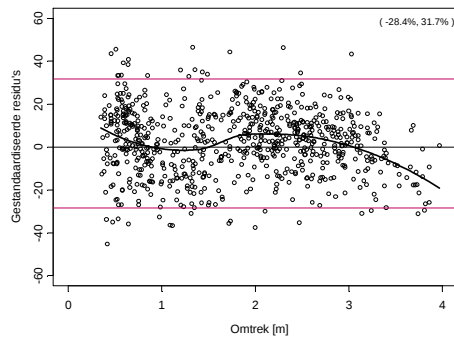


Beuk P1: veeltermmodel met één ingang / fytoecoregio (N: 1428 / B: 132)

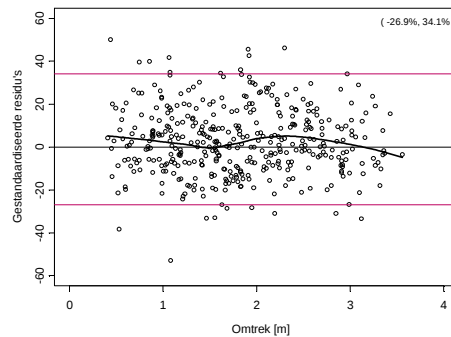


Figuur 4-12 Beuk: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). Relatie met omtrek en hoogte.

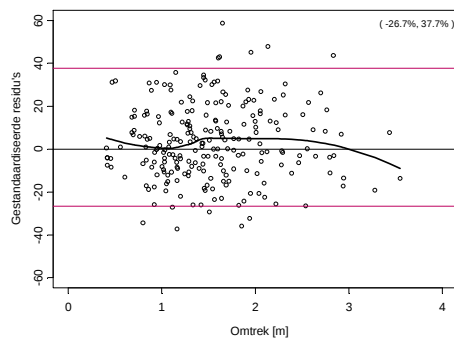
Beuk: Zoniënwoud [746] P1: veeltermmodel met één ingang / fytoecoregio



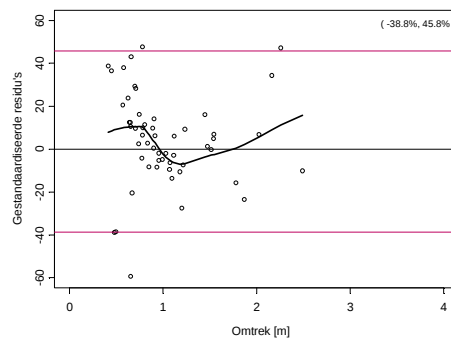
Beuk: Brabants district [399] P1: veeltermmodel met één ingang / fytoecoregio



Beuk: Vlaams district [229] P1: veeltermmodel met één ingang / fytoecoregio



Beuk: Kempens district [54] P1: veeltermmodel met één ingang / fytoecoregio



Figuur 4-13 Beuk: residuanalyse van het veeltermmodel met één ingang per ecoregio (P1f) in functie van omtrek (C).

Coëfficiënten van het model (Tabel 4-8)

Voor alle ecoregio's behalve het Kempens district is de coëfficiënt van de derde orde term (C3) positief. In de Kempen zal dus voor hogere waarden van de omtrek het model naar beneden afbuigen. Uiteraard is dat geen correct model, maar binnen het bereik waar de tarief geldig is, vormt dat geen probleem. Wel is dat (opnieuw) een indicatie dat we misschien beter steunen op de tarieven voor het Vlaamse district om de volumeschattingen voor de kempen uit te voeren.

Tabel 4-8 Beuk: coëfficiënten van het veeltermmodel met één ingang en ecoregio.

		Value	Std. Error	t value	Pr(> t)
Zoniënwoud [N=746]	Referentie:	4.767			
	(Intercept)	0.1354	0.0256	5.29	1.6e-007
	C	-0.6449	0.0920	-7.01	0.000
	C2	1.4252	0.0931	15.31	0.000
	C3	0.0276	0.0246	1.12	0.261
Brabants district [N=399]	Referentie:	4.115			
	(Intercept)	-0.1140	0.0570	-2.001	4.6e-002
	C	0.1295	0.1812	0.715	0.470
	C2	0.8115	0.1611	5.038	0.000
	C3	0.0906	0.0403	2.246	0.025
Vlaams district [N=229]	Referentie:	3.401			
	(Intercept)	-0.0990	0.0575	-1.722	0.086516
	C	0.0865	0.2071	0.418	0.677
	C2	0.6974	0.2034	3.428	0.001
	C3	0.0672	0.0570	1.180	0.239
Kempens district [N=54]	Referentie:	2.991			
	(Intercept)	0.0501	0.178	0.281	0.7798
	C	-0.4841	0.664	-0.729	0.469
	C2	1.4592	0.735	1.986	0.053
	C3	-0.2409	0.236	-1.021	0.312

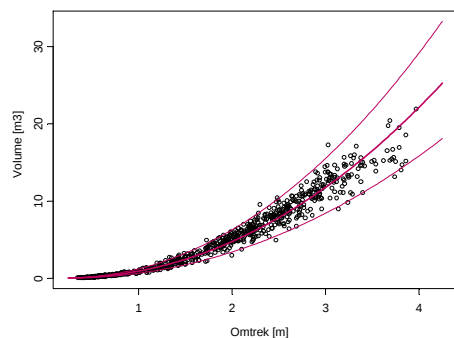
Tabel 4-9 Beuk: vergelijking van de coëfficiënten van het veeltermmodel met één ingang en ecoregio.

	District				
	Alle	Zoniën	Brabants	Vlaams	Kempens
Referentie:	3.282	3.753	3.276	2.708	2.501
(Intercept)	-0.0325	0.1354	-0.1140	-0.0990	0.0501
C	-0.0228	-0.6449	0.1295	0.0865	-0.4841
C2	0.7556	1.4252	0.8115	0.6974	1.4592
C3	0.1556	0.0276	0.0906	0.0672	-0.2409
Voorspellingsfout		-28.4% , 31.7%	-26.9% , 43.1%	-26.7% , 37.7%	-38.8% , 45.8%

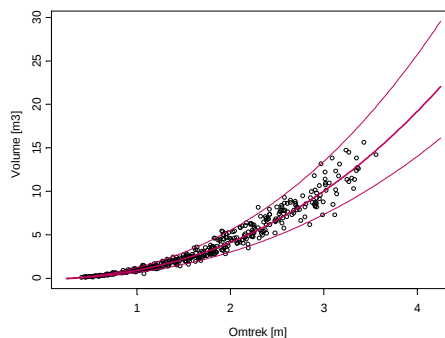
Voorstelling van het model (Figuur 4-14 en Figuur 4-15).

De modellen apart per ecoregio illustreren goed de ordening. Alleen het Kempens district doorbreekt het patroon een beetje, maar alleen voor lage waarden van de omtrek. Het verschil met het Vlaams district is niet groot behalve voor relatief hogere waarden van de omtrek. Maar in dat bereik hebben we heel weinig waarnemingen.

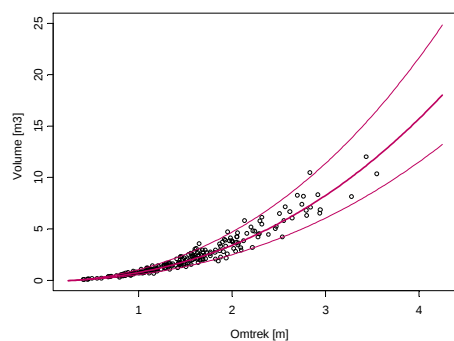
Beuk: Zoniënwoud [746] P1: veeltermmodel met één ingang / fyto regio



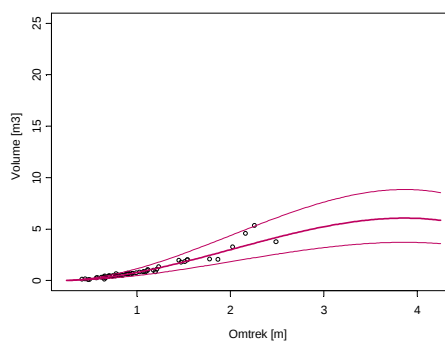
Beuk: Brabants district [399] P1: veeltermmodel met één ingang / fyto regio



Beuk: Vlaams district [229] P1: veeltermmodel met één ingang / fyto regio

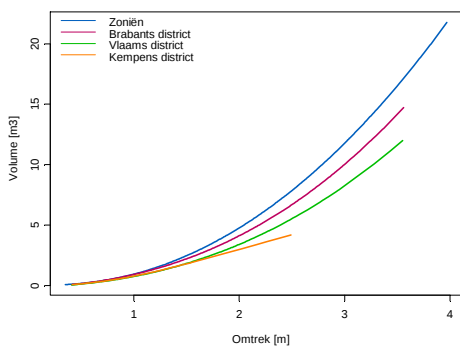


Beuk: Kempens district [54] P1: veeltermmodel met één ingang / fyto regio

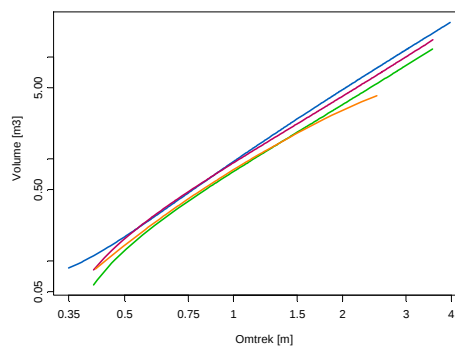


Figuur 4-14 Beuk: grafiek van het veeltermmodel met één ingang per ecoregio (P1f). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten onder en boven liggen 95% van de gestandaardiseerde residu's.

Beuk P1: veeltermmodel met één ingang / fyto regio (N: 1428 / B: 132)



Beuk P1: veeltermmodel met één ingang / fyto regio (N: 1428 / B: 132)



Figuur 4-15 Beuk: grafische vergelijking van de veeltermmodellen (P1f) in de originele schaal (links) en in de log-schaal (rechts).

4.2 De logaritmische modellen

4.2.1 Logaritmisch model met twee ingangen

Startmodel

Uit de verkennende analyse (o.a. Figuur 3-6) bleek dat een logaritmische transformatie de variatie stabiliseert. Vandaar dat we een lineair regressiemodel zonder weging kunnen uitproberen in de log-schaal. Hierbij proberen we het volume te voorspellen op basis van de log van de omtrek en de hoogte en een eventuele interactie tussen beide termen:

$$L2 \leftrightarrow \log_{10}(V) \approx \log_{10}(C) * \log_{10}(H) \quad (4.6)$$

Interessant is nat te gaan in hoeverre de interactie significant is en een groot effect heeft. Indien niet, dan kunnen we schrijven:

$$L2x \leftrightarrow \log_{10}(V) \approx \log_{10}(C) + \log_{10}(H) \quad (4.7)$$

Hieruit volgt na terugtransformatie (zie Materiaal en Methoden voor een bewijsvoering):

$$V \propto \varphi \frac{C^{\beta_1} H^{\beta_2}}{12\pi} \quad \text{met: } \varphi = 12\pi \cdot 10^{\beta_0} \quad (4.8)$$

Als $\beta_1 \approx 2$ en $\beta_2 \approx 1$ dan krijgen we ongeveer de formule van een kegel op een vormfactor φ na.

Modelselectie (Tabel 4-10)

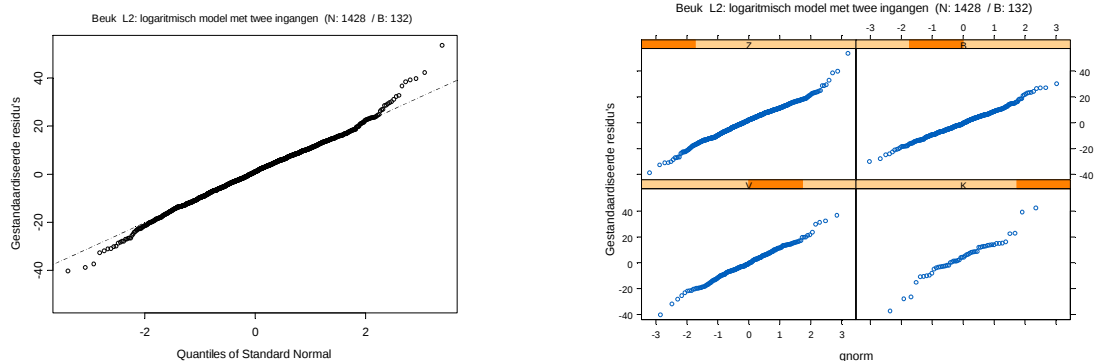
Uit de ANOVA-tabel blijkt dat de interactie (het product) tussen logC en logH nog weinig bijdraagt aan het model, maar de term is wel significant. Daarom bekijken we zowel het model met de interactie (Tabel 4-11) als zonder interactie (Tabel 4-12).

Tabel 4-10 Beuk: ANOVA-tabel van het logaritmisch model met twee ingangen.

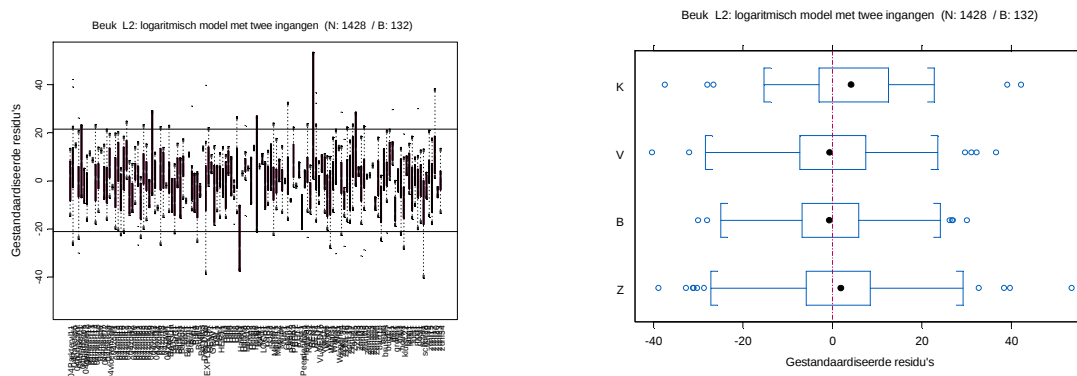
	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
logC	1	447	447	195842	0.00000
logH	1	7	7	3158	0.00000
logC:logH	1	0	0	8	0.00509
Residuals	1424	3	0		

Residuanalyse (Figuur 4-16, Figuur 4-17 en Figuur 4-18)

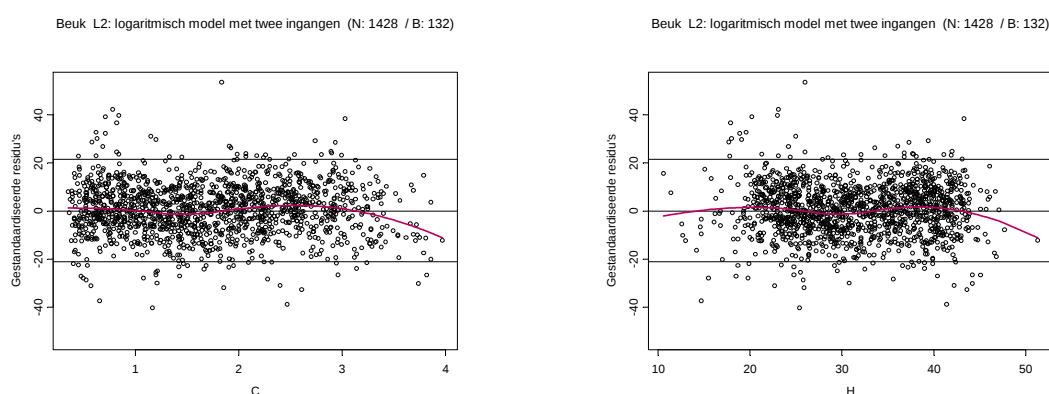
De residu's vertonen geen bijzondere patronen. In het bijzonder is het 95 % interval van de gestandaardiseerde residu's ± 20 % of de relatieve voorspellingsintervallen liggen opnieuw tussen 0.8 en 1.2 keer de voorspelde waarde. Alleen is er een beperkte afbuiging van de gestandaardiseerde residu's bij hoge waarden van de omtrek ($C > 3.5$ m).



Figuur 4-16 Beuk: residuanalyse voor het logaritmisch model met twee ingangen (L2): QQ-grafieken.



Figuur 4-17 Beuk: residuanalyse voor het logaritmisch model met twee ingangen (L2): boxplots.



Figuur 4-18 Beuk: residuanalyse voor het logaritmisch model met twee ingangen (L2): omtrek en hoogte.

Coëfficiënten van het model (Tabel 4-11 en Tabel 4-12)

Voor het model zonder interactieterm is het volume recht evenredig met een term die heel dicht zit bij de vorm van een kegel:

$$V \propto C^{1.9} H^{1.2} \quad (4.9)$$

Door het invoeren van de interactieterm (die significant is), veranderen de coëfficiënten omdat ze onderling gecorreleerd zijn, maar de invloed op het referentievolume is beperkt.

Tabel 4-11 Beuk: coëfficiënten van het logaritmisch model met twee ingangen (met interactie)

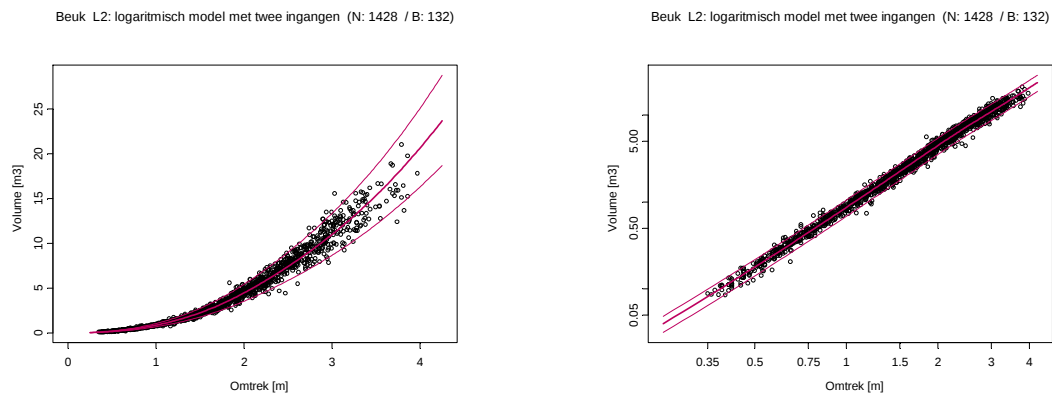
	Value	Std. Error	t value	Pr(> t)
Referentie:	3.374			
(Intercept)	-1.806	0.0319	-56.55	0.00000
logC	2.073	0.0708	29.27	0.00000
logH	1.228	0.0226	54.32	0.00000
logC:logH	-0.135	0.0483	-2.81	0.00509

Tabel 4-12 Beuk: coëfficiënten van het logaritmisch model met twee ingangen (zonder interactie)

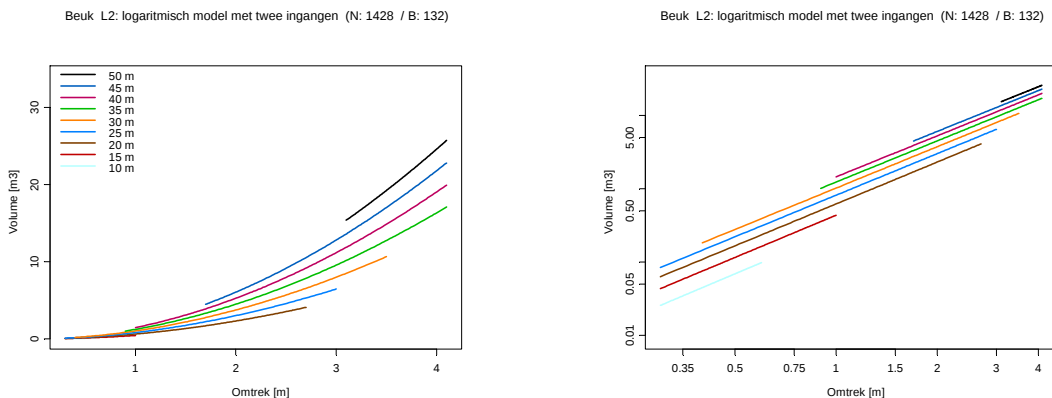
	Value	Std. Error	t value	Pr(> t)
Referentie:	3.36			
(Intercept)	-1.78	0.03069	-58.0	0
logC	1.88	0.00954	196.5	0
logH	1.21	0.02156	56.1	0

Voorstelling van het model (Figuur 4-19).

Zoals ook volgt uit bovenstaande residuanalyse is de aansluiting van de waarnemingen bij het model goed.



Figuur 4-19 Beuk: grafiek van het logaritmisch model met twee ingangen (L2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de gestandaardiseerde residu's liggen.



Figuur 4-20 Beuk: grafische voorstelling van het logaritmisch model met twee ingangen (L2) per hoogteklaas (rechts in de log-schaal).

4.2.2 Logaritmische model met één ingang

Startmodel

Het logaritmisch model met één ingang bekomen we door de hoogterm uit het model (4.6) weg te laten.

$$L1 \leftrightarrow \log_{10}(V) = \log_{10}(C) \quad (4.10)$$

Modelselectie (Tabel 4-13)

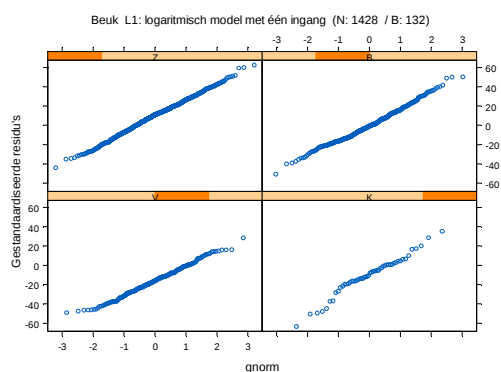
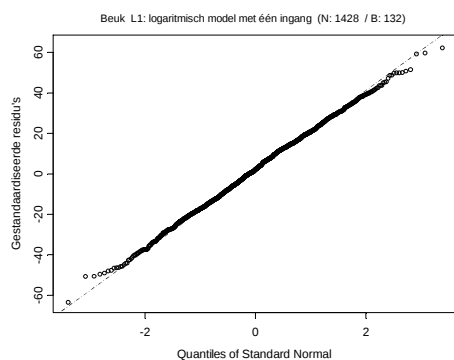
Voor de volledigheid geven we de ANOVA-tabel voor dit eenvoudige model.

Tabel 4-13 Beuk: ANOVA-tabel van het logaritmisch model met één ingang.

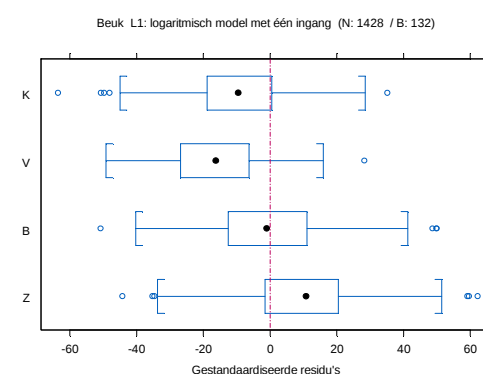
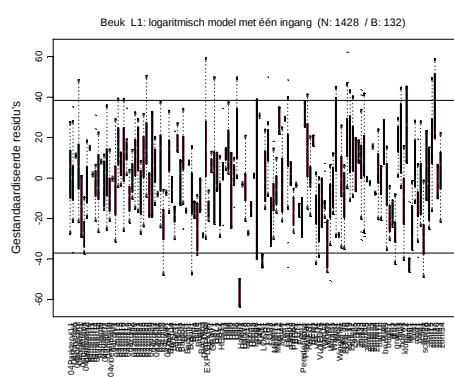
	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
logC	1	447	447	60843	0
Residuals	1426	10	0		

Residuanalyse (Figuur 4-21, Figuur 4-22 en Figuur 4-23)

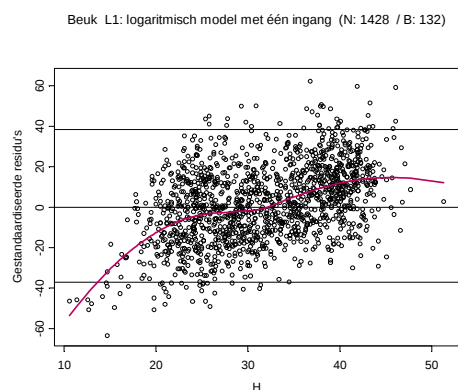
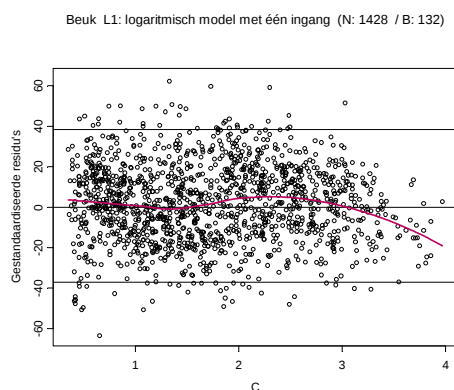
Uiteraard zorgt het ontbreken van de hoogte ervoor dat de schattingen minder precies zijn (ongeveer 40%). In functie van de hoogte is er een systematische (stijgende) trend. Ook is er een beperkte vertekening voor de omtrek. Voor een grote omtrek is er onderschatting. Omdat hoogte gecorreleerd is met de ecoregio, hebben we ook een verband met ecoregio. In het Zoniën-woud is er een systematische onderschatting, in het Brabants district is de schatting gemiddeld goed en in het Vlaams en het Kempens district is er een overschatting.



Figuur 4-21 Beuk: residuanalyse voor het logaritmisch model met één ingang (L1). QQ-plots.



Figuur 4-22 Beuk: residuanalyse voor het logaritmisch model met één ingang (L1). Boxplots.



Figuur 4-23 Beuk: residuanalyse voor het logaritmisch model met één ingang (L1). Relatie met omtrek en hoogte.

Coëfficiënten van het model (Tabel 4-14)

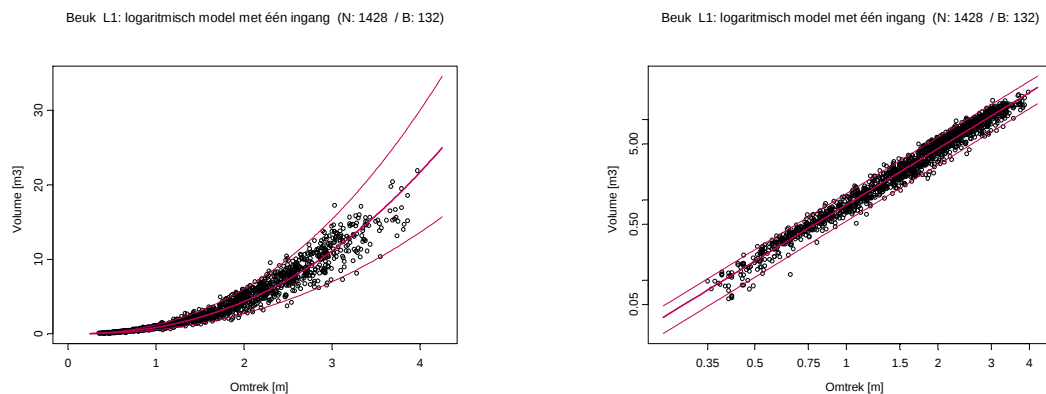
Ten opzichte van het logaritmisch model met twee ingangen verandert de waarde in de referentie nauwelijks (3.37 m^3).

Tabel 4-14 Beuk: coëfficiënten van het logaritmisch model met één ingang.

	Value	Std. Error	t value	Pr(> t)
Referentie:	3.40			
(Intercept)	-0.0618	0.00282	-21.9	0
logC	2.3222	0.00941	246.7	0

Voorstelling van het model (Figuur 4-24).

Onthank de grotere onnauwkeurigheid blijkt het model gemiddeld nog goed aan te sluiten bij de gegevens. Het model met één ingang is bijgevolg bruikbaar op een Vlaamse schaal om grove schattingen te maken.



Figuur 4-24 Beuk: grafiek van het logaritmisch model met één ingang (L1). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten liggen 95% van de gestandaardiseerde residu's.

4.2.3 Logaritmische model met één ingang en ecoregio

Startmodel

Het logaritmisch model met één ingang kunnen we opsplitsen per ecoregio. Op die manier krijgen we inzicht hoe het model afhangt van ecoregio:

$$L1f \leftrightarrow \log_{10}(V) \square \log_{10}(C) * Fytoregio \quad (4.11)$$

ANOVA (Tabel 4-15)

Correctie met ecoregio verbetert het model in belangrijke mate. De bijdrage is ook hoog-significant.

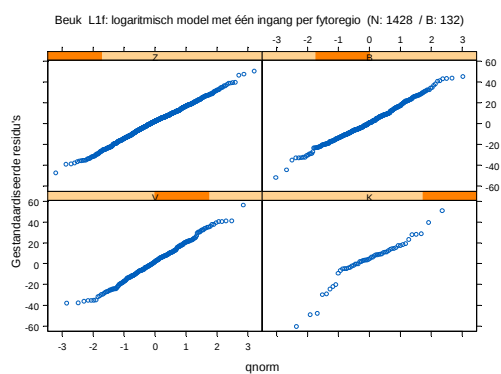
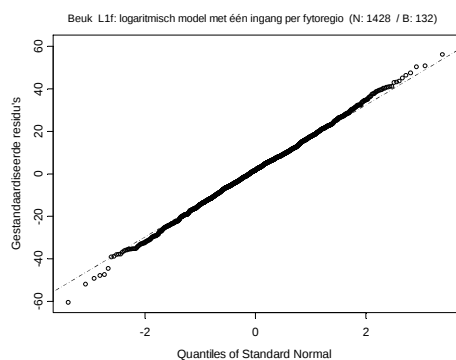
Tabel 4-15 Beuk: ANOVA-tabel van het logaritmisch model met één ingang en ecoregio.

	Terms	Resid.	Df	RSS	Sum of Sq	F	Value	Pr(F)
1	logC	1426	10.5					
2	logC * Ecoregio	1420	7.6	2.92	91.4	0		

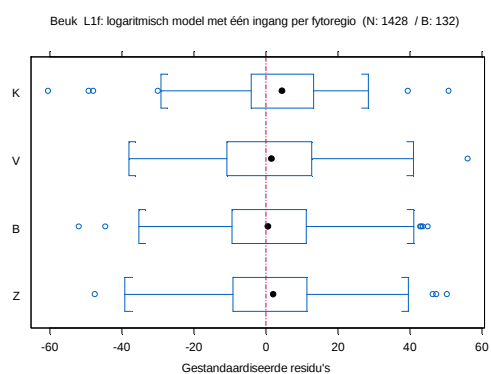
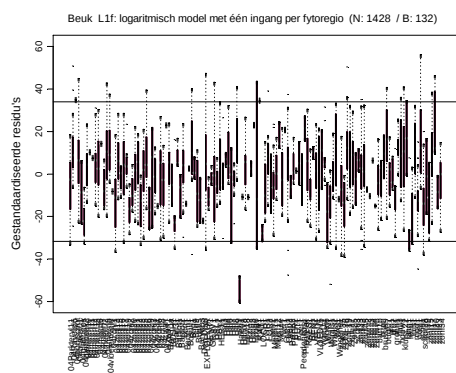
Residuanalyse (Figuur 4-25, Figuur 4-26, Figuur 4-27 en Figuur 4-28)

Correctie met ecoregio verkleint de gestandaardiseerde residu's. Voor Zoniën en Brabants district daalt de fout met 10 % tot ongeveer 30 %. Voor het Vlaams district en het Kempens district blijft de fout ongeveer 40 % (Figuur 4-28).

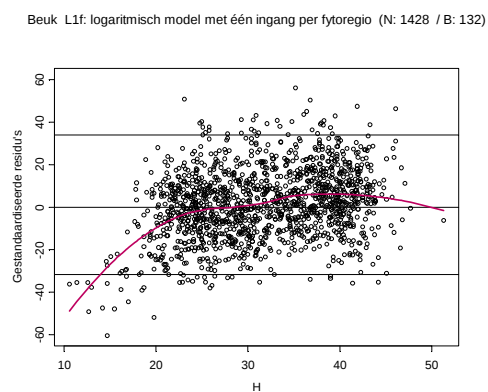
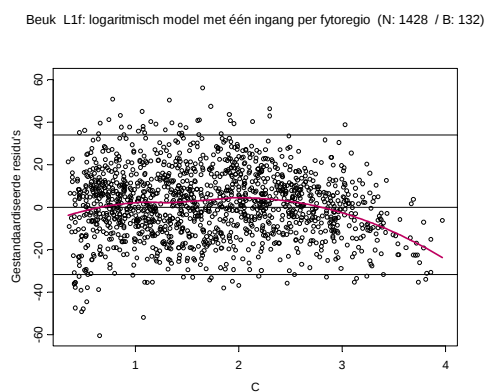
Zoals te verwachten is er voor één ingang een patroon van de gestandaardiseerde residu's ten opzichte van de omtrek. Maar door introductie van ecoregio is dat verband iets minder sterk. Behalve voor het Brabants district blijft er overal een parabolisch verband van de gestandaardiseerde residu's bestaan. Zeker in het Zoniënwoud neemt de onderschatting toe bij grotere waarden van de omtrek.



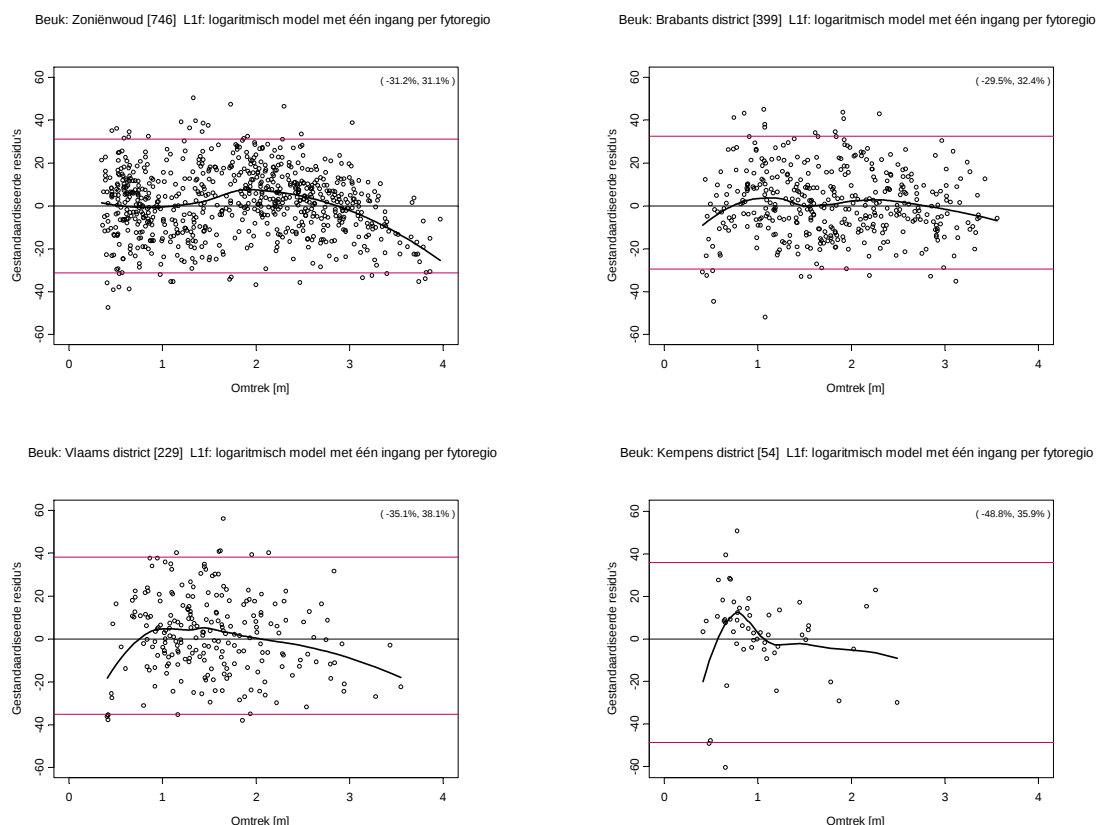
Figuur 4-25 Beuk: residuanalyse voor het logaritmisch model met één ingang per district (L1f). QQ-plots.



Figuur 4-26 Beuk: residuanalyse voor het logaritmisch model met één ingang per district (L1f). Boxplots.



Figuur 4-27 Beuk: residuanalyse voor het logaritmisch model met één ingang per district (L1f). Relatie met omtrek en hoogte.



Figuur 4-28 Beuk: residuanalyse van het logaritmisch model met één ingang per district (L1f) in functie van omtrek (C).

Coëfficiënten van het model (Tabel 4-16).

Het intercept betekent hier de schatting bij een omtrek van 1 m (dan is $\log C = 0$). Hieruit blijkt dat voor 1 m het Kempense district iets hoger ligt dan het Vlaamse district. Aangezien de helling (de coëfficiënt van $\log C$) in het Vlaamse district hoger is dan het Kempense district betekent dit dat in het Kempens district het volume hoger zal zijn dan in het Vlaamse district. Voor waarden groter dan 1 m wordt het houtvolume in het Vlaamse district al snel groter. Dat is bijvoorbeeld al het geval voor de referentieboom ($C = 1.8$ m). Verder blijkt opnieuw de klassieke volgorde op te gaan voor alle coëfficiënten: Zoniën > Brabants district > Vlaams district.

Tabel 4-16 Beuk: coëfficiënten van het logaritmisch model met één ingang en ecoregio.

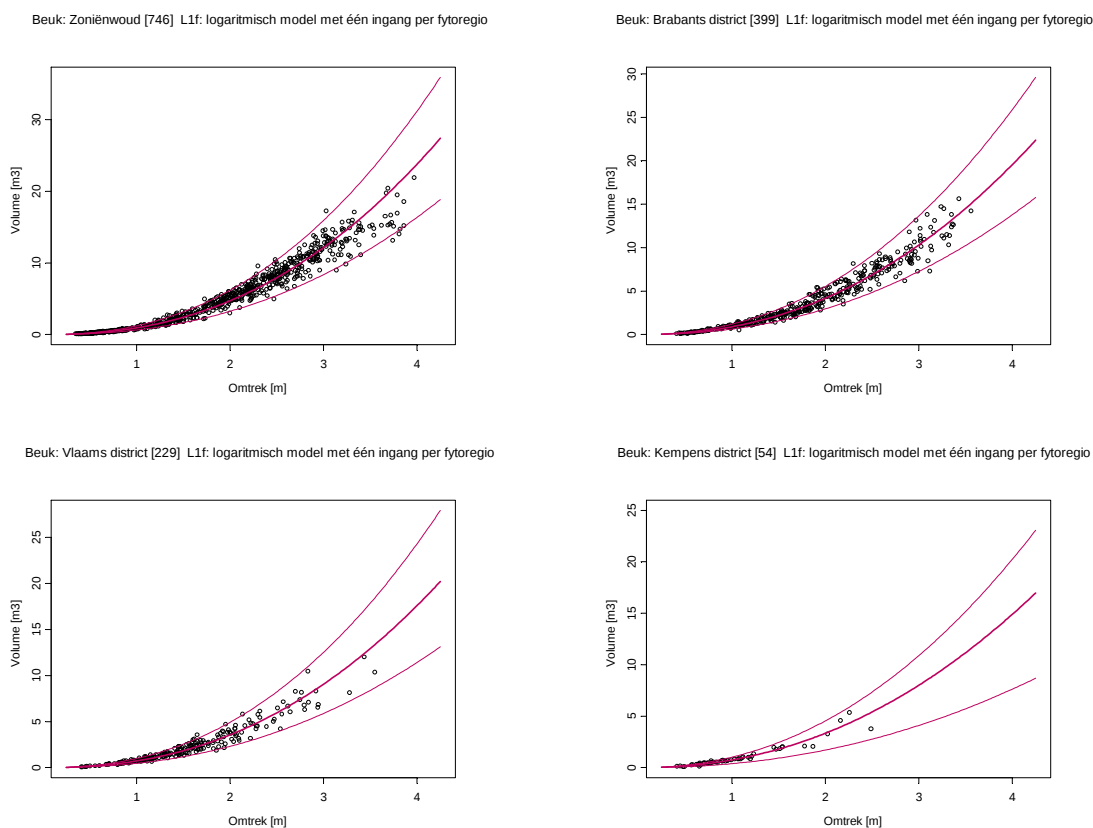
		Value	Std. Error	t value	Pr(> t)
Zoniënwood [N=746]	Referentie:	3.68			
	(Intercept)	0.932	0.0031	-9.85	0
	logC	2.336	0.0095	246	0
Brabants district [N=399]	Referentie:	3.31			
	(Intercept)	0.895	0.0051	-9.64	0
	logC	2.224	0.0169	132	0
Vlaams district [N=229]	Referentie:	2.79			
	(Intercept)	0.719	0.0068	-21.0	0
	logC	2.306	0.0289	80	0
Kempens district [N=54]	Referentie:	2.66			
	(Intercept)	0.747	0.0144	-8.84	0
	logC	2.159	0.0782	28	0

Tabel 4-17 Beuk: vergelijking van de coëfficiënten van het logaritmisch model met één ingang en ecoregio.

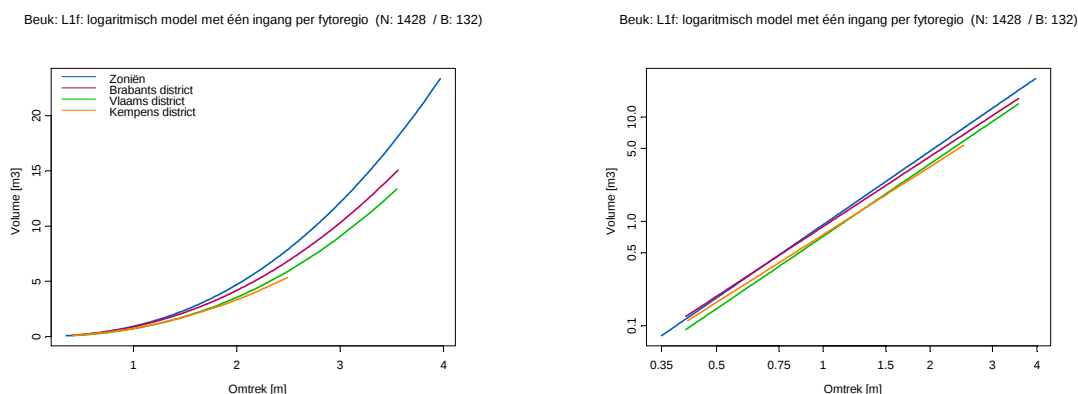
	District				
	Alle	Zoniën	Brabants	Vlaams	Kempens
Referentie:	3.40	3.68	3.31	2.79	2.66
(Intercept)	-0.062	0.932	0.895	0.719	0.747
log(C)	2.322	2.336	2.224	2.306	2.159
Voorspellingsfout		-31.2% , 31.1%	-25.9% , 32.4%	-35.1% , 38.1%	-48.6% , 35.9%

Voorstelling van het model (Figuur 4-29 en Figuur 4-30).

Zoals uit bovenstaande analyse te verwachten is, sluiten de gegevens goed aan bij de regressievergelijkingen. Zoals bleek uit de bespreking van de coëfficiënten heeft het Kempens district een groter houtvolume tot iets voorbij 1 m omtrek. Vanaf dan is het voorspelde houtvolume in het Vlaams district groter. Het Brabants district ligt uniform hoger dan het Kempens district, voor lage waarden van omtrek liggen de voorspellingen in Brabants district iets hoger dan in het Zoniënwoud. Maar voor de rest hebben de beuken in het Zoniënwoud gemiddeld het hoogste houtvolume.



Figuur 4-29 Beuk: grafiek van het logaritmisch model met één ingang per ecoregio (L1f). Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten onder en boven liggen 95% van de gestandaardiseerde residu's.



Figuur 4-30 Beuk: vergelijking van de logaritmische modellen (links in de originele schaal rechts en in de log-schaal).

4.3 Relatief model met twee ingangen

Startmodel

Het relatief model modelleert de verhouding van het volume tot het volume van een kegel als een lineaire functie van de omtrek en de hoogte. Op die manier krijgen we inzicht hoe de vormfactor afhangt van de grootte van een boom:

$$R2 \leftrightarrow relV = \frac{V}{V_{kegel}} = C + H \quad (4.12)$$

Modelselectie (Tabel 4-18)

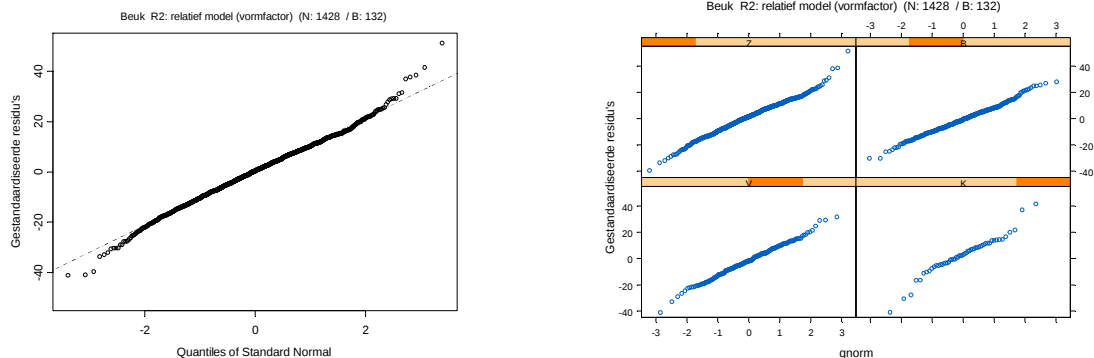
Opvallend is dat beide variabelen (hoogte en omtrek) maar een klein deel van de variatie (zie *sum of squares*) verklaren. Toch is dat in overeenstemming met de vorige modellen. We hebben hier al een groot deel van de variantie geëlimineerd door te delen door het kegelvolume. Na correctie hiervoor verklaren de andere variabelen in het model nog heel weinig.

Tabel 4-18 Beuk: ANOVA-tabel van het relatief model met twee ingangen.

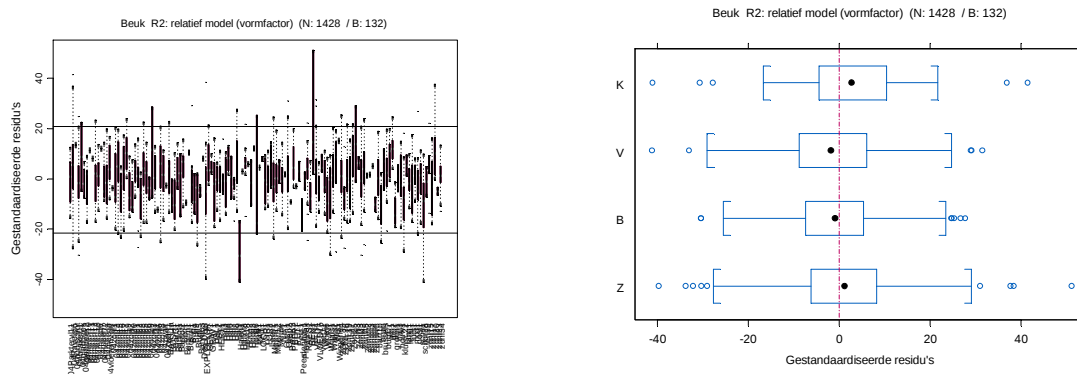
	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
C	1	1.6	1.59	91.5	0
H	1	1.3	1.33	76.9	0
Residuals	1425	24.8	0.02		

Residuanalyse (Figuur 4-31, Figuur 4-32 en Figuur 4-33).

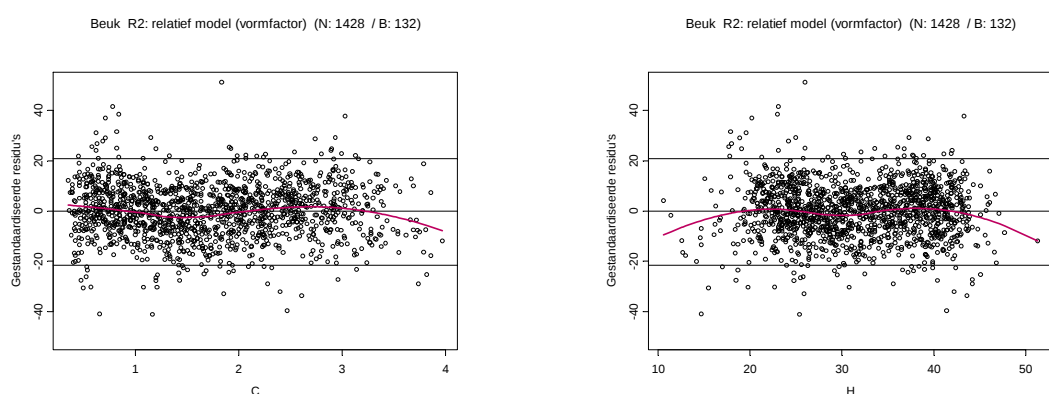
Opnieuw zijn er heel weinig anomalieën. Het relatief model met twee ingangen heeft dezelfde kwaliteit als het veeltermmodel en het logaritmisch model. De 95 % voorspellingsmarge is opnieuw ongeveer 20 % en deze fout is homogeen over het heel bereik. Voor grote waarden van de omtrek is er een tendens tot onderschatting, maar we hebben te weinig waarnemingen om dat formeel te toetsen.



Figuur 4-31 Beuk: residuanalyse voor het relatief model met twee ingangen (R2). QQ-grafieken.



Figuur 4-32 Beuk: residuanalyse voor het relatief model met twee ingangen (R2): boxplots



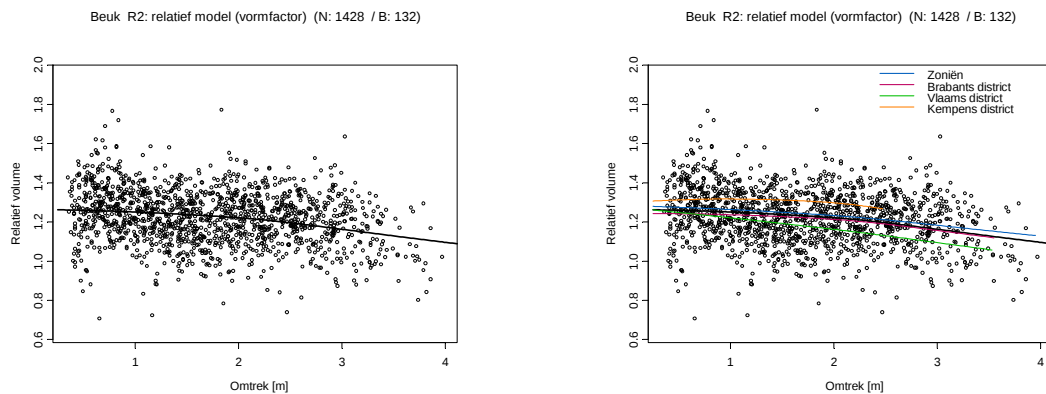
Figuur 4-33 Beuk: residuanalyse voor het relatief model met twee ingangen (R2). Relatie met omtrek en hoogte.

Coëfficiënten van het model (Tabel 4-19).

Zoals al duidelijk was uit de verkennende analyse (Figuur 3-8), daalt het relatieve volume in functie van de omtrek, maar er is wel een toename in functie van de hoogte. Beide trends compenseren elkaar gedeeltelijk. Als we het voorspelde relatief volume uitzetten in functie van de omtrek, waarbij de hoogte gelijk is aan de voorspelde hoogte bij de omtrek, dan tekent zich een zacht dalende trend af (Figuur 4-34). De figuur rechts toont aan dat het verband afhangt van de ecoregio. In het Kempens district ligt de curve iets hoger, in het Vlaams district iets lager.

Tabel 4-19 Beuk: coëfficiënten van het relatief model met twee ingangen.

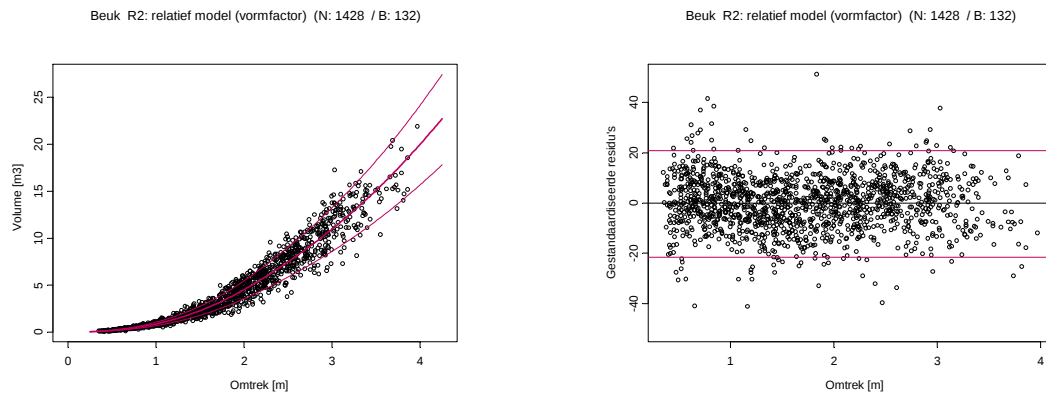
	Value	Std. Error	t value	Pr(> t)
Referentie:	3.41435	0.003513	347.96	0
(Intercept)	1.15239	0.017808	64.71	0
C	-0.09204	0.007262	-12.67	0
H	0.00725	0.000827	8.77	0



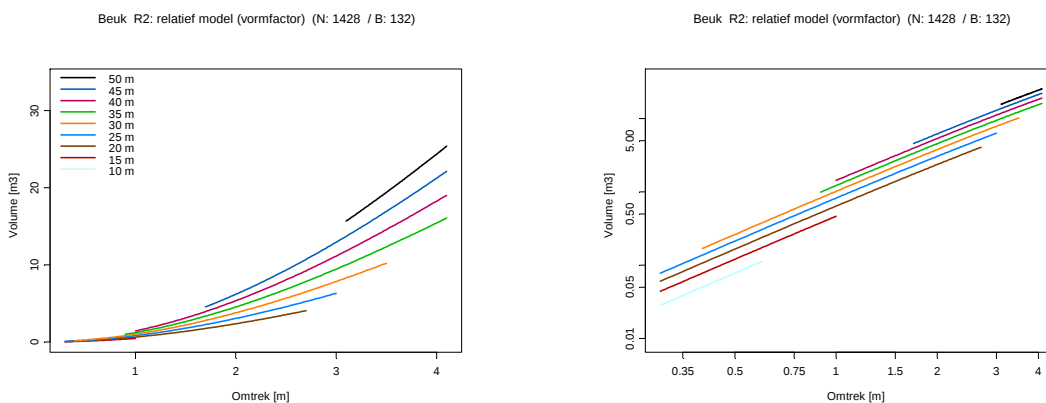
Figuur 4-34 Beuk: evolutie van het relatief volume in functie van de omtrek waarbij de hoogte gelijk is aan de voorspelde waarde bij de omtrek (rechts uitgesplitst per ecoregio)

Voorstelling van het model (Figuur 4-35).

Er is een goede overeenstemming tussen de waarnemingen en het model over het hele bereik.



Figuur 4-35 Beuk: grafiek van het relatief model met twee ingangen (R2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de gestandaardiseerde residu's liggen.



Figuur 4-36 Beuk: grafiek van het relatief model met twee ingangen (R2) per hoogteklaas (rechts in de log-schaal).

5 Beuk: vergelijking van de modellen

5.1 Vergelijking van de voorspellingen van de modellen

5.1.1 De referentieboom

De referentieboom heeft volgende afmetingen: $C = 1.8$ m en $H = 32.5$ m (Tabel 3-2). Tabel 5-1 geeft een overzicht van het voorspelde volume voor de referentieboom voor de verschillende modellen. De tabel is een compilatie van Tabel 4-3 en Tabel 4-9 voor de veeltermmodellen, Tabel 4-14 en Tabel 4-17 voor de logaritmische modellen en Tabel 4-19 voor het relatief model.

Voor de modellen met twee ingangen (T2) verschillen de (drie) modellen weinig. De voorspelde volumes van de modellen met één ingang (T1 en T1f) liggen wel iets verder uit elkaar. Maar de verschillen blijven klein met uitzondering voor het model voor het Kempens district, waar het verschil tussen het veelterm- en logaritmisch model 0.34 m³ is.

Het volume voor de referentieboom voor het model met twee ingangen (T2) is ongeveer 3.4 m³. Dat verandert weinig als we werken met een model met één ingang (T1). Maar wanneer we opsplitsen per ecoregio (T1f), komen uiteraard wel grote verschillen naar boven. De rangorde van het voorspelde volume is: $Z = 3.7$ m³ > $B = 3.3$ m³ > $V = 2.8$ m³ > $K = 2.6$ m³ (het gemiddelde van voorspellingen van model P1f en L1f in Tabel 5-1).

Tabel 5-1 Beuk: schatting voor een referentieboom ($C = 1.8$ m, $H = 32.5$ m)

Type model	T2	T1	T1f (één ingang + ecoregio)			
District			Zoniën	Brabants	Vlaams	Kempens
Polynomiaal (P)	3.40	3.28	3.75	3.28	2.71	2.50
Logaritmisch (L)	3.37	3.40	3.68	3.31	2.79	2.66
Relatief (R)	3.41					

5.1.2 De tarieven met twee ingangen

Figuur 5-1 vergelijkt de modellen twee aan twee in functie van omtrek en hoogte. Tabel 5-2 geeft enkele kengetallen. Ongeveer 95 % van de verschillen ligt tussen -2.5 en 2.5 %. De uitschieters komen voor aan de randen, waar het model uiteraard minder goed is. Voor kleine waarden van de omtrek en de hoogte divergeren de modellen een beetje, maar globaal genomen zijn er geen grote verschillen tussen de modellen.

Tabel 5-2 Beuk: procentueel verschil tussen modellen voor tarieven met twee ingangen. Enkele kengetallen.

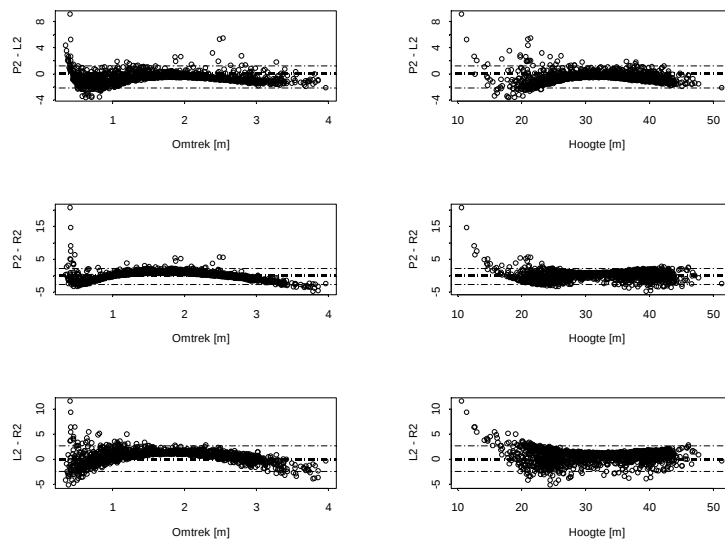
Modellen	Percentiel		
	2.5 %	Mediaan	97.5 %
P2 – L2	-2.18	-0.61	1.18
P2 – R2	-2.72	0.25	2.09
L2 – R2	-2.46	0.88	2.71

We geven een meer gedetailleerde bespreking per combinatie van de modellen op basis van Figuur 5-1:

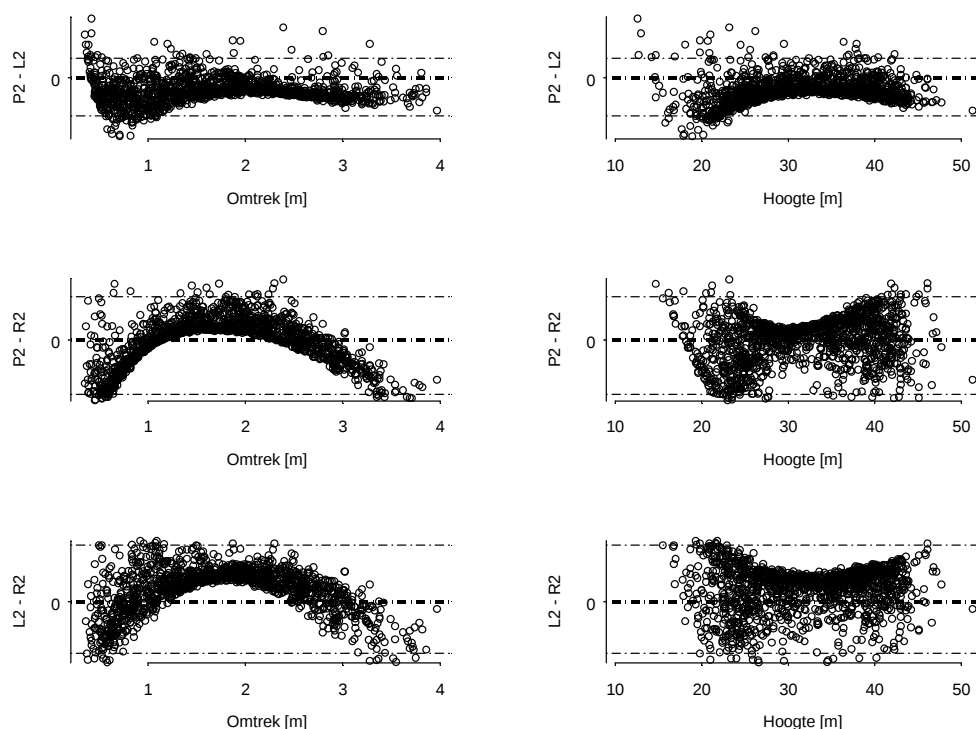
- P2 – L2:
 - o De verschillen zijn relatief klein op een paar uitschieters na: 95 % ervan ligt tussen -2.18 % en 1.18 %.
 - o L2 leidt gemiddeld tot een hogere schatting dan P2. Dat is vooral het geval aan de extrema van het bereik voor omtrek en hoogte.
 - o P2 is hoger voor een omtrek lager dan 1 m en hoger dan 3 m en voor een hoogte lager dan 20 m en hoger dan 40 m.

- P2 – R2:
 - o 95 % van de verschillen ligt tussen -2.72 % en 2.09 %. Bij kleine waarden van de omtrek en de hoogte zijn er verschillen van 5 % en meer.
 - o In het omtreksinterval 1 m tot ongeveer 3 m vallen de modellen nagenoeg samen.
 - o In functie van de hoogte is er geen duidelijk patroon. Wel zijn de verschillen het kleinst bij een hoogte van 30 m.
- L2 – R2:
 - o De verschillen zijn klein: 95 % ligt tussen -2.46 en 2.71. Dit interval is het meest symmetrisch. Bij de uiteinden lopen de verschillen op tot 5 % en meer.
 - o De reden is dat er zowel een omtreksinterval is waar L2 domineert (in het midden) en een interval waar R2 de grootste schattingen heeft.
 - o In functie van de hoogte is er veel minder een patroon, behalve bij heel kleine waarden waar het logaritmisch model het grootste is.

Beuk N: 1428 / B: 132



Figuur 5-1 Beuk: vergelijking van de modellen met twee ingangen. P2 = Veeltermmodel, L2 = logaritmisch model en R2 = relatief model. De punt-streep lijn in het vet is de nullijn. De punt-streep lijn onder en boven omvatten 95 % van de punten.



Figuur 5-2 Eik: vergelijking van de modellen met twee ingangen. P2 = Veeltermmodel, L2 = logaritmisch model en R2 = relatief model. De punt-streep lijn in het vet is de nullijn. De punt-streep lijn onder en boven omvatten 95 % van de punten. De figuur is dezelfde als Figuur 5-1, maar de uitbijters zijn weggelaten om meer details te zien.

5.1.3 De tarieven met één ingang per ecoregio

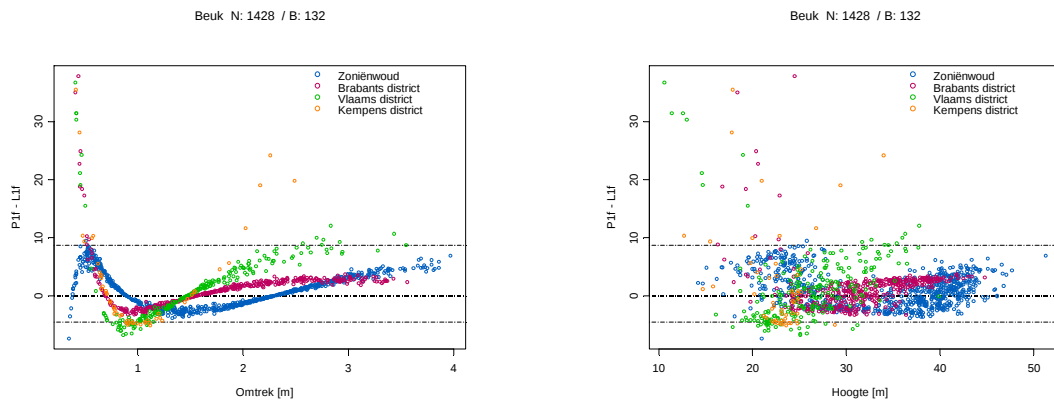
Figuur 5-3 vergelijkt P1f en L1f, de twee modellen met één ingang aangevuld met ecoregio. De patronen die naar boven komen zijn heel afgetekend. Het logaritmisch model heeft in het middenstuk de grootste waarden, maar naar de randen geeft het veeltermmodel de grootste waarden.

Voor het Brabants district blijken de modellen het best met elkaar in overeenstemming. In Zoniën is het verloop bijzonder. Over een breder omtrekbereik dan de rest, geeft het logaritmisch model de grootste waarde (van minder dan 1 m tot iets meer dan 2 m) en de curve gaat terug naar beneden voor heel kleine waarden van de omtrek.

Als we de verschillen uitzetten ten opzichte van de hoogte, verdwijnen de patronen grotendeels. Dat is ook te verwachten. Bij de tarieven met één ingang wordt geen rekening gehouden met hoogte en de modellen bepalen dus de relatie bij een uitgemiddelde waarde van de hoogte.

Tabel 5-3 Beuk: procentueel verschil tussen modellen voor tarieven met één ingang en ecoregio (P1f – L1f).

Ecoregio	Percentiel		
	2.5 %	Mediaan	97.5 %
Zoniën	-3.11	0.74	7.0
Brabants	-2.88	1.15	8.1
Vlaams	-5.79	0.40	19.7
Kempisch	-5.02	-1.04	26.8

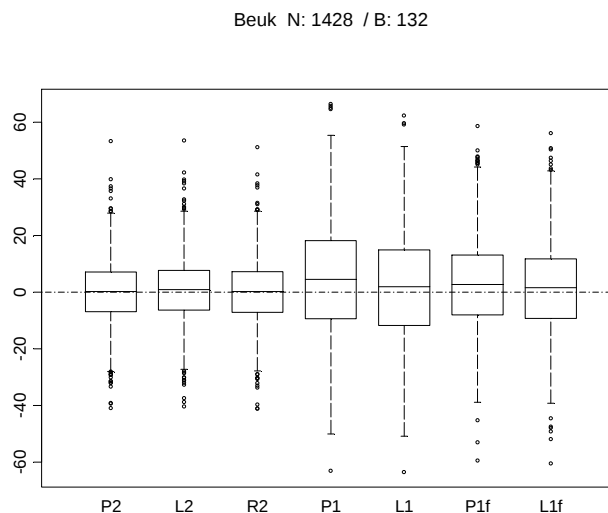


Figuur 5-3 Beuk: vergelijking van de modellen met één ingang per ecoregio in functie van de omtrek en de hoogte.

5.2 Vertekening en variabiliteit van de modellen

5.2.1 Distributie van de gestandaardiseerde residu's

Figuur 5-4 vergelijkt met boxplots de gestandaardiseerde residu's van de modellen. De modellen met twee ingangen (P2, L2 en R2) hebben de meest compacte distributie. De boxplots verschillen weinig. De mediaan van L2 ligt iets boven nul, wat wijst op een (kleine) positieve vertekening. Het veeltermmodel (P2) en het relatieve model (R2) zijn nagenoeg equivalent. Bij de modellen met één ingang worden de distributies compacter door toevoeging van ecoregio. Het logaritmisch model (L1f) vertoont globaal minder vertekening dan het veeltermmodel (P1f).



Figuur 5-4 Beuk: boxplots van de gestandaardiseerde residu's (%)

5.2.2 Juistheid en precisie van de voorspellingen

Als maat voor de juistheid nemen we de mediaan van gestandaardiseerde residu's (Tabel 5-4). Deze informatie kunnen we grafisch afleiden uit de boxplots (Figuur 5-4), maar de tabel geeft meer details voor de modellen met één ingang opgesplitst per ecoregio (L1f en P1f). In het ideale geval verwachten we een onvertekende symmetrische distributie.

Voor tarieven met twee ingangen, is het veeltermmodel nagenoeg onvertekend, maar ook bij het logaritmisch model en het relatief model is de vertekening klein (0.5 %). Voor de tarieven met één ingang stijgt de vertekening: 2 % voor het logaritmisch model en 4 % voor het veeltermmodel.

Door opsplitsing van de modellen met één ingang per ecoregio vermindert de vertekening, behalve voor het Kempens district, wellicht omdat de steekproef te klein is. In het Zoniënwoud en het Brabants district is de vertekening het kleinst voor het logaritmisch model, in Vlaanderen is het Veeltermmodel iets beter.

Tabel 5-4 Beuk: juistheid van de modellen met als maat de mediaan van de gestandaardiseerde residu's.

Ecoregio	P2	P1	P1f	L2	L1	L1f	R2
Zoniënwoud	1.3	13.4	2.9	1.9	10.8	2.0	1.2
Brabants district	-1.4	1.4	1.7	-0.7	-1.0	0.6	-0.9
Vlaams district	-1.7	-13.8	1.1	-0.6	-16.2	1.6	-1.8
Kempens district	3.9	-8.4	5.4	4.2	-9.6	4.5	2.7

Tabel 5-5 kwantificeert de voorspellingsfout. De kwaliteit van de voorspelling meten we af aan het 95 % interval van de gestandaardiseerde residu's. Voor tarieven met twee ingangen is de voorspellingsmarge ongeveer 21 % (het relatief model scoort net iets beter dan het veeltermmodel, gevolgd door het logaritmisch model). Voor de modellen met één ingang zonder ecoregio is de voorspellingsmarge bijna 40 %. Het logaritmisch model scoort iets beter (37.7 %). Opsplitsen naargelang de ecoregio verbetert de voorspellingsmarge tot 30 à 35 %, behalve voor de Kempen (bijna 50 %).

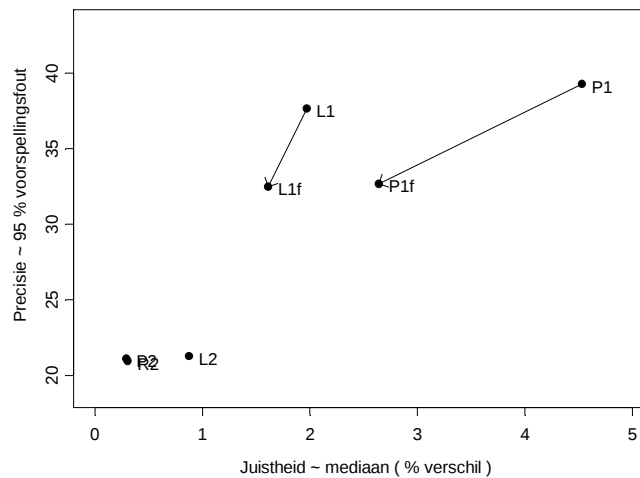
Tabel 5-5 Beuk: precisie van de modellen met als maat de relatieve voorspellingsfout van de gestandaardiseerde residu's.

Ecoregio	P2	P1	P1f	L2	L1	L1f	R2
Zoniënwoud	20.8	42.6	29.9	20.9	37.3	31.2	20.6
Brabants district	19.1	34.7	31.3	18.7	33.4	32.2	19.6
Vlaams district	21.6	39.0	32.9	21.6	41.1	35.7	21.8
Kempens district	31.4	46.6	44.4	31.3	48.7	48.4	32.8

5.2.3 Grafische voorstelling van de accuraatheid

De precisie en de juistheid definiëren samen de accuraatheid van het model. Figuur 5-5 geeft hiervan een grafische voorstelling door beide kwaliteitsmaten te combineren in een grafiek. In de X-as staat de juistheid van de voorspellingen, in de Y-as de precisie ervan. Hoe dichter de punten bij de oorsprong liggen, des te hoger de kwaliteit van de voorspelling.

Voor tarieven met twee ingangen, zijn P2 en R2 nagenoeg equivalent, en scoort L2 iets minder goed wegens een iets grotere vertekening. Voor tarieven met één ingang, zien we hoe de inbreng van ecoregio, zowel de vertekening als de voorspellingsfout vermindert. Globaal genomen scoort het logaritmisch model (L1f) iets beter, vooral wat de vertekening betreft.



Figuur 5-5 Beuk: vertekening en voorspellingsfout

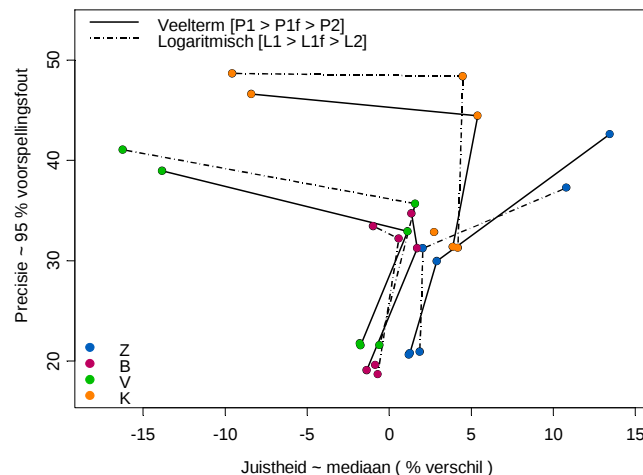
Figuur 5-6 gaat een stap verder en onderzoekt voor elke ecoregio de evolutie van de kwaliteit van de voorspellingen als we de modellen verbeteren uitgaande van een tarief met één ingang tot en met een tarief met twee ingangen. De punten die door lijnen verbonden zijn, zijn ofwel afkomstig van het veeltermmodel ofwel van het logaritmisch model. De vier punten niet verbonden door een lijn (soms half verborgen) stellen het relatief model voor aangezien hier maar één variant mogelijk was. De evolutie is in alle ecoregio's nagenoeg gelijk. We zien hoe introductie van ecoregio de systematische vertekening uit het model haalt en hoe introductie van hoogte (dus een tarief met twee ingangen) de voorspellingsfout verkleint. De verschillen tussen de verschillende modeltypes is klein.

Omdat deze figuur alle bevindingen samenbalt, geven we hier een gedetailleerde bespreking per ecoregio:

- Zoniën (Z):
 - o Voor de tarieven met één ingang zijn de gestandaardiseerde residu's gemiddeld positief wat betekent dat het model een onderschatting geeft. Dat is logisch omdat de bomen in Zoniën inderdaad veel hoger worden dan in de andere districten.
 - o Door introductie van fyto regio in het model verdwijnt de vertekening nagenoeg. De voorspellingsfout daalt tot 30 %.
 - o Toevoegen van hoogte (het model van twee ingangen) doet de voorspellingsfout verder dalen.
- Brabants district (B):
 - o In het Brabants district leidt geen enkel model tot een vertekening. Ook in termen van de voorspellingsfout is het model met één tarief lager dan de andere (ongeveer 35 %).
 - o De winst door fyto regio op te nemen in het model is klein. De voorspellingsfout daalt maar heel beperkt en behalve voor de Kempen is de kwaliteit van de modellen nu min of meer gelijkwaardig.
 - o Pas door toevoegen van de hoogte daalt de voorspellingsfout tot onder de 20 %. Hiermee scoort het Brabants district het best.

- Vlaams district (V):
 - o Voor de tarieven met één ingang zijn de gestandaardiseerde residu's gemiddeld negatief. Het model leidt tot een overschatting omdat de bomen in het Vlaams district kleiner blijven dan in Brabantse district en het Zoniënwoud.
 - o Door introductie van fyto regio in het model verdwijnt de vertekening nagenoeg. De voorspellingsfout daalt tot ongeveer 35 %.
 - o Toevoegen van hoogte (het model van twee ingangen) doet de voorspellingsfout verder dalen tot iets meer dan 20 %.
- Kempens district (K):
 - o Net zoals in het Vlaams district leidt het model met één tarief tot een overschatting. Daarenboven is de voorspellingsfout heel groot: bijna 50 %.
 - o Door introductie van fyto regio in het model wordt de vertekening lichtjes positief. De voorspellingsfout daalt nauwelijks.
 - o Toevoegen van hoogte (het model van twee ingangen) doet de voorspellingsfout slechts dalen tot 30 %.

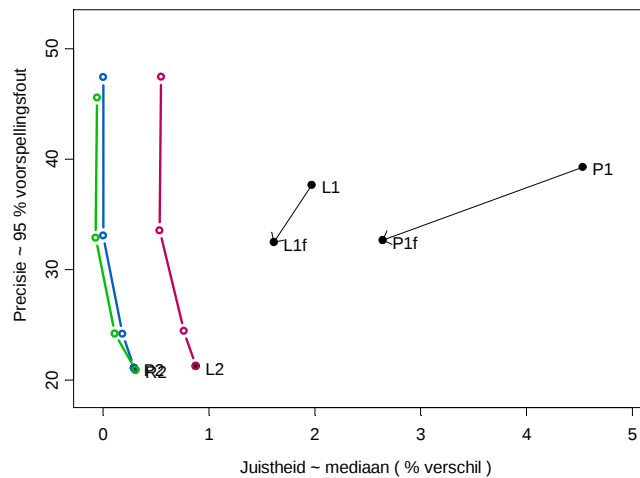
Beuk N: 1428 / B: 132



Figuur 5-6 Beuk: vertekening en voorspellingsfout opgesplitst naargelang fyto regio.

5.3 De invloed van de meetfout op de hoogte

Het is niet eenvoudig en tijdrovend de hoogte van een boom te bepalen. Ook is de meetfout soms groot omdat het zicht op de top van de boom gehinderd wordt door omringende bomen en/of eigen takken van de kroon. Om het effect van de meetfout op de kwaliteit van de tarieven in te schatten, voerden we een simulatie uit waarbij we meer en meer ruis toevoegen aan de hoogte. Figuur 5-7 geeft een grafische voorstelling hiervan. Hieruit blijkt dat de juistheid van de voorspellingen nauwelijks beïnvloed wordt door de meetfout op de hoogte. Maar de voorspellingsfout neemt gestaag toe. Deze evolutie is gelijk voor alle modellen met twee tarieven. Van zodra de relatieve meetfout gemiddeld 8 % is, wordt de voorspellingsfout van het model groter dan de tarieven met één ingang. Daarom is het misschien efficiënter om te werken met de modellen met één ingang uitgesplitst per fyto regio, als we de hoogte onvoldoende nauwkeurig kunnen bepalen.



Figuur 5-7 Beuk: invloed van meetfouten op de hoogte op de vertekening en voorspellingsfout. De gemiddelde relatieve fout in absolute waarde is respectievelijk 0%, 4%, 8% en 12%.

5.4 Algemeen besluit

5.4.1 De tarieven met twee ingangen

De verschillen tussen de drie modellen zijn heel klein. Het veeltermmodel (P2) en het relatief model (R2) zijn nagenoeg equivalent, het logaritmisch model (L2) is een klein beetje vertekend. Het 95% voorspellingsinterval is een relatieve afwijking van ongeveer 20 %. De keuze van de modellen is dus niet zo kritiek. Omdat we relatief model vooral vanuit een meer academische vraagstelling ingevoerd hebben, en het logaritmisch model iets minder goed scoort, gaat onze voorkeur uit naar het veeltermmodel.

5.4.2 De tarieven met één ingang

De 95 % voorspellingsmarge van modellen met één ingang verdubbelt ten opzichte van de modellen met twee ingangen, bijna 40 %. Maar het is mogelijk de modellen te verbeteren door een opsplitsing te maken in functie van ecoregio en de voorspellingsfout te verminderen tot 30 %. Alleen in het Kempens district blijven de voorspellingen slecht. De steekproef is er te klein en daarenboven hebben we maar over een heel beperkt omtrekbereik meetpunten. Het logaritmisch model en het veeltermmodel verschillen weinig van elkaar. We kunnen kiezen.

Vraag is dan nog wanneer we modellen met één ingang, dan wel met twee ingangen moeten toepassen. Uit bovenstaande analyse is duidelijk dat we nog veel kunnen winnen (in bepaalde districten tot 15 %) door te werken met twee ingangen op voorwaarde dat we de hoogte voldoende precies kunnen bepalen. Maar van zodra de gemiddelde meetfout in absolute waarde groter wordt dan 8 %, wordt de voorspellingsfout van de modellen met twee ingangen groter dan met één ingang.

6 Beuk: tarieven (rudimentaire tabellen)

6.1 Inleiding

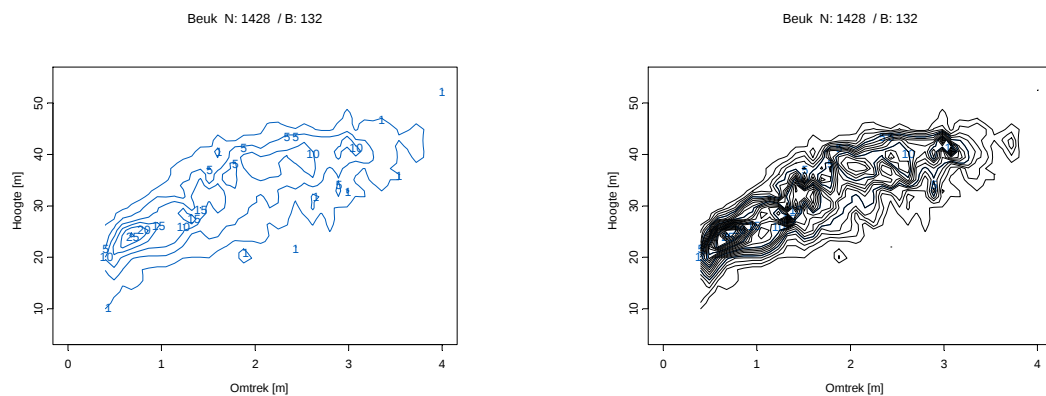
In deze rubriek geven we een eerste aanzet voor de tarieven. Hiertoe laten we omtrek variëren vanaf 0.35 m tot 4 m in stappen van 0.1 m en de hoogte van 10 m tot 50 m in stappen van 2.5 m. Deze punten zijn het centrum van een interval. Voor de omtrek reikt elk interval van -0.05 m tot +0.05 m ten opzichte van het centrum en voor de hoogte is dat van -1.25 m tot 1.25m.

De tabellen met de volumeschattingen hebben we alleen berekend voor plausibele combinaties van hoogte en omtrek. Tabel 6-1 geeft de combinaties in de databank. Opvullingen van hiaten in het bereik en uitbreidingen hebben we aangeduid met "-1". Als bij een ontbrekende combinatie van omtrek en hoogte de omringende combinatie wel bestaat, dan vulden we het hiaat op (voor interpolatie). Aan de randen hebben we het bereik één cel uitgebreid (voor extrapolatie).

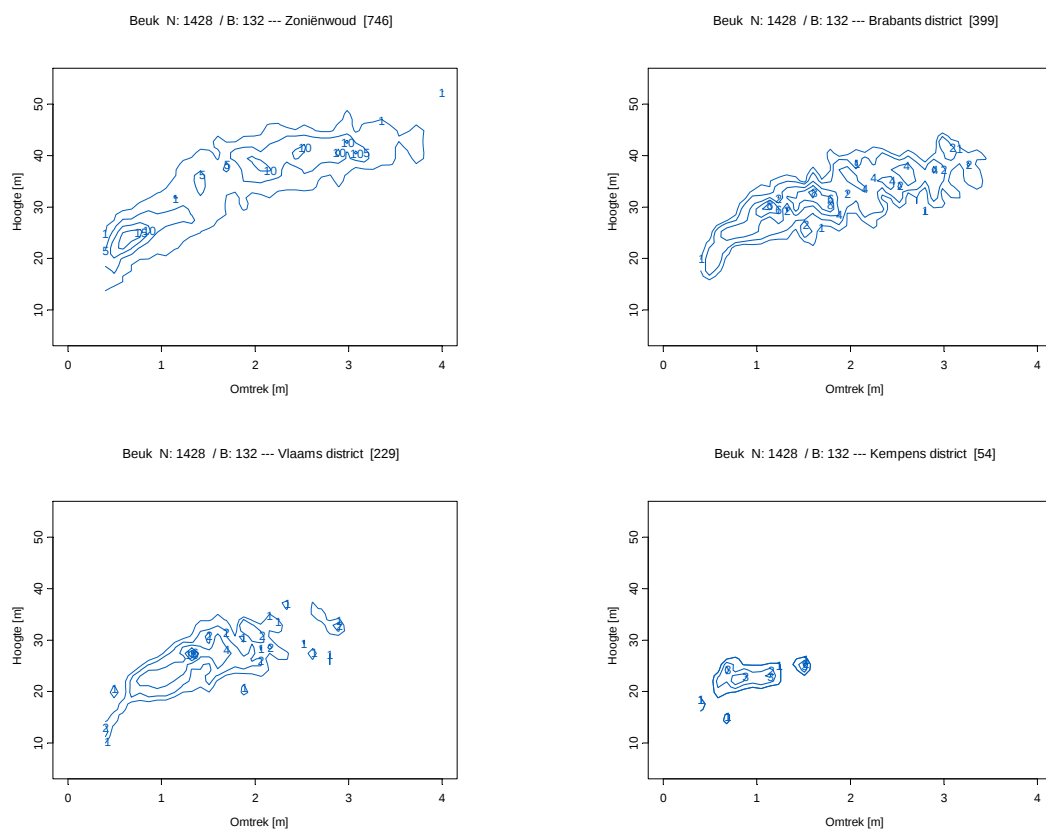
Van Tabel 6-1 maakten we een grafische voorstelling met contourgrafieken (Figuur 6-1 globaal en Figuur 6-2 per ecoregio). Voor het Kempens district zijn er bijna geen gegevens beschikbaar en ook het bereik in omtrek en hoogte is heel smal. In het Vlaamse district zijn de gegevens talrijker, maar het bereik gaat niet verder dan een omtrek van 3 m. In het Zoniënwoud hebben we de meeste gegevens en daar is ook het omtrekbereik het grootst (tot 4m). Ook tonen de figuren nog maar eens hoe de hoogte in functie van de omtrek sterk afhankelijk is van het district.

Tabel 6-1 Beuk: het meetbereik. Aantal bomen per combinatie van omtrek (vertikaal) en hoogteklaas (horizontaal) of ecoregio. De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd. De punten (•) zijn hiaten of uitbreidingen (Tabel 6-1).

	10	12.5	15	17.5	20	22.5	25	27.5	30	32.5	35	37.5	40	42.5	45	47.5	50	52.5	Z	B	V	K
0.4	1	3	2	5	11	3	2	•											18	3	4	2
0.5	•	1	4	9	15	17	7	1											40	8	4	2
0.6		•	1	4	8	23	21	6	•										46	10	2	5
0.7			2	1	5	19	28	6	1										36	11	6	9
0.8			1	1	4	15	25	11	3	•									34	8	11	7
0.9			•	•	4	12	18	9	9	1									21	15	12	5
1					4	12	10	15	10	4	•								19	18	14	4
1.1					4	11	10	13	14	2	1	1							15	20	16	5
1.2					1	6	13	12	15	2	5	2							21	18	13	4
1.3					1	2	5	23	6	9	3	3	•						19	15	18	•
1.4					1	2	4	10	16	13	13	5	2						33	18	14	1
1.5					1	3	8	10	7	14	1	5	3	•					17	17	14	4
1.6					•	2	7	7	14	15	8	4	•	2					17	22	20	•
1.7					•	1	3	8	9	9	2	7	3	1					15	16	12	•
1.8					•	•	4	5	10	6	4	3	11	1					16	21	6	1
1.9					3	•	4	4	5	10	10	12	11	2					29	19	12	1
2					•	•	3	4	4	3	9	14	12	2	•				28	13	9	1
2.1					•	1	2	•	5	6	11	14	7	2	1				28	12	9	•
2.2					•	•	•	5	5	7	8	15	13	6	2				35	20	5	1
2.3					•	•	•	2	4	2	8	6	10	5	2				20	12	6	1
2.4					•	1	•	•	•	1	10	6	14	9	•				26	14	1	•
2.5					1	1	1	•	2	5	5	11	10	11	2				33	11	4	1
2.6					•	•	•	2	1	•	10	13	11	8	1	•			27	15	4	•
2.7					•	•	•	•	2	1	6	6	9	8	•	1			21	10	2	
2.8						1	2	1	2	2	5	8	8	•	•				18	6	5	
2.9						•	•	1	7	5	6	13	8	2	•				25	14	3	
3								1	•	2	4	5	16	2	2				21	11	•	
3.1								•	2	•	1	14	4	•	•				14	7	•	
3.2								•	1	•	1	7	6	3	•				14	4	•	
3.3								1	•	3	4	4	3	2	1				10	7	1	
3.4								•	1	1	1	4	2	3	•				8	3	1	
3.5								•	1	1	1	1	1	1	•				4	•	1	
3.6									1	•	•	1	•	•	•				1	1	•	
3.7									•	2	5	2	•	•	•				9	•		
3.8										1	1	•	1	1	•				4			
3.9										•	1	1	•	•	•	•			2			
4											•	•	•	•	•	1	•		•			



Figuur 6-1 Beuk: het empirisch bereik van de modellen.



Figuur 6-2 Beuk: het empirisch bereik van de modellen per ecoregio

6.2 Tarief met twee ingangen (veeltermmodel)

6.2.1 De tarieven in tabelvorm

Als model voor het tarief met twee ingangen nemen we het veeltermmodel. De vergelijking is:

$$V = -0.01115 + 0.03633 C^2 H - 0.08560 C^2 - 0.04996 C^3 + 0.00256 CH \quad (6.1)$$

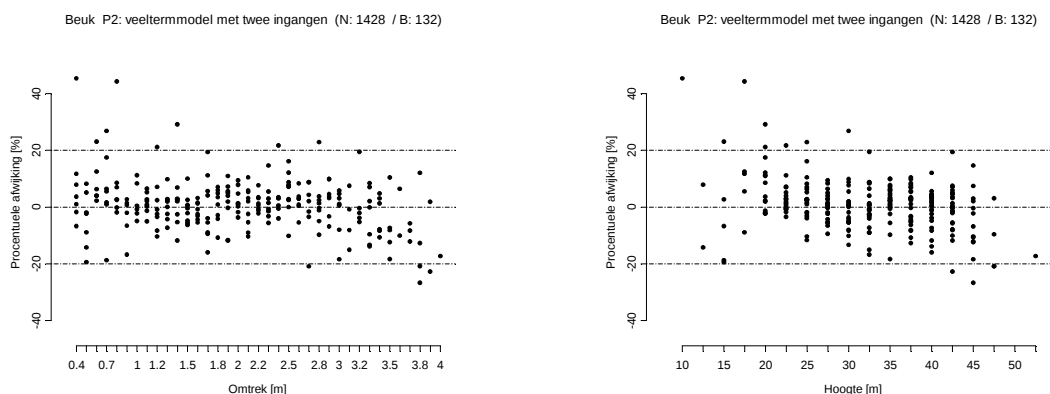
Tabel 6-2 geeft een rudimentaire tabel. Vanaf $H = 45 \text{ m}$ moesten we veel hiaten opvullen want daar zijn minder gegevens beschikbaar.

Tabel 6-2 Beuk: tarief met twee ingangen (veeltermmodel) in dm³. De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd. De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1).

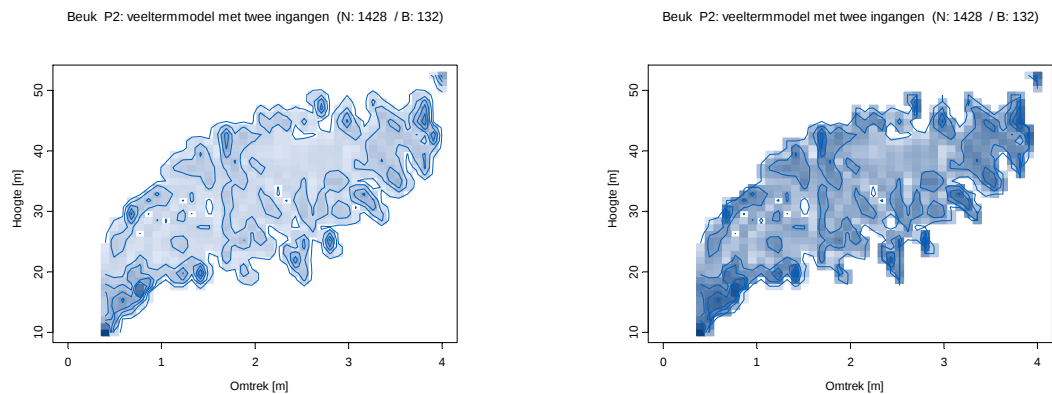
		Hoogte [m]																	
	10	12.5	15	17.5	20	22.5	25	27.5	30	32.5	35	37.5	40	42.5	45	47.5	50	52.5	
0.4	40	57	74	92	109	126	143	•160											
0.5	•65	91	117	143	168	194	220	246											
0.6		•130	166	203	240	276	313	349	•386										
0.7			224	273	322	371	420	469	518										
0.8			288	351	414	478	541	604	667	•731									
0.9			•359	•438	518	597	676	756	835	914									
1					631	728	826	923	1020	1117	•1214	•1312							
1.1					754	871	988	1105	1222	1339	1456	1573							
1.2					887	1025	1164	1302	1441	1579	1718	1856							
1.3					1029	1191	1353	1514	1676	1838	2000	2162	•2323						
1.4					1180	1367	1554	1741	1928	2115	2302	2489	2676						
1.5					1339	1553	1767	1981	2195	2409	2623	2837	3051	•3265					
1.6					•1507	1750	1993	2235	2478	2721	2964	3206	•3449	3692					
1.7					•1683	1956	2230	2503	2776	3050	3323	3596	3870	4143					
1.8					•1866	•2172	2478	2784	3090	3395	3701	4007	4313	4619					
1.9					2057	•2398	2738	3078	3418	3758	4098	4438	4778	5118					
2					•2256	•2632	3008	3384	3760	4136	4512	4888	5264	5641	•6017				
2.1					•2461	2875	3288	•3702	4116	4530	4944	5358	5772	6186	6600				
2.2					•2672	•3126	•3579	4033	4487	4940	5394	5848	6301	6755	7209				
2.3					•2890	•3385	•3880	4375	4870	5366	5861	6356	6851	7346	7842				
2.4					•3113	3652	•4190	•4729	•5267	5806	6344	6883	7421	7960	•8499				
2.5					3343	3926	4510	•5094	5677	6261	6845	7428	8012	8596	9179				
2.6					•3577	•4208	•4838	5469	6100	•6730	7361	7992	8622	9253	9883	•10514			
2.7							•5176	•5855	6534	7214	7893	8573	9252	9931	•10611	•11290			
2.8							5521	6251	6981	7711	8441	9171	9901	10631	•11361	•12091			
2.9							•5875	•6657	7439	8222	9004	9787	10569	11352	12134	•12916			
3									7909	•8746	9583	10419	11256	12092	12929	13766			
3.1									•8390	9283	•10176	11068	11961	12854	•13746	•14639			
3.2									•8882	9832	•10783	11733	12684	13635	14585	•15536			
3.3									9384	•10394	11404	12415	13425	14435	15445	16456			
3.4									•9896	10968	12040	13112	14183	15255	16327	•17398			
3.5										•11554	12689	13824	14959	16094	17229	•18364			
3.6											13351	•14551	•15751	16951	•18151	•19352			
3.7											•14026	•15293	16560	17827	19094	•20361			
3.8												16049	17385	•18721	20057	21393			
3.9													•16820	18226	19633	•21039	•22446	•23852	•25259
4														•19083	•20562	•22041	•23520	•24999	26478

6.2.2 Controle per klasse

Volgens Figuur 6-3 liggen de meeste afwijkingen tussen -20 % en 20 % (zie ook Tabel 6-3). De afwijkingen hebben geen patroon volgens omtrek, wel een beetje in functie van de hoogte. Een contourgrafiek van de absolute afwijking in twee dimensies (Figuur 6-4) toont dat de grootste afwijkingen zich aan de randen bevinden.



Figuur 6-3 Beuk: procentuele afwijking van het gemiddelde van de volumemetingen t. o. v. de tarieven.



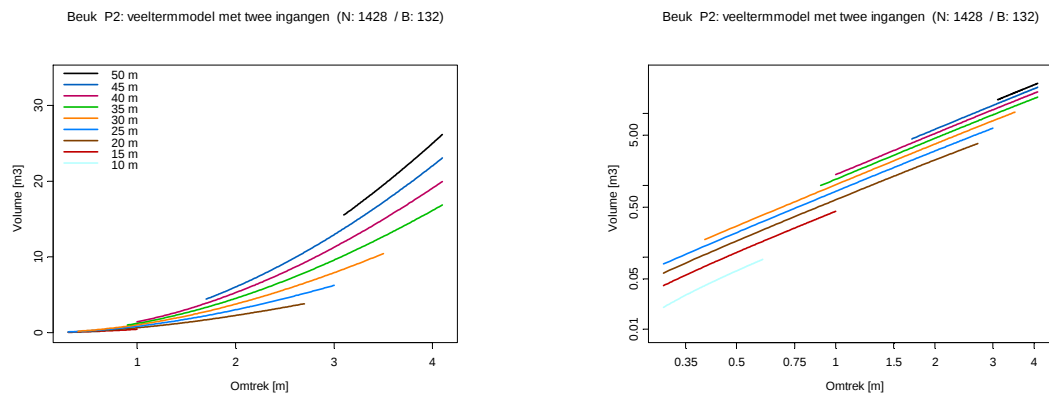
Figuur 6-4 Beuk: verkenning van het tweedimensioneel patroon van de absolute afwijkingen (links: in de originele schaal, rechts in de log-schaal).

Tabel 6-3 Beuk: vergelijking van het tarief met twee ingangen (veeltermmodel) ten opzichte van het gemiddelde per klasse. De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd.

	Hoogte [m]																		
	10	12.5	15	17.5	20	22.5	25	27.5	30	32.5	35	37.5	40	42.5	45	47.5	50	52.5	
0.4	45.3	7.8	-6.8	11.7	3.6	-1.8	1.0												
0.5		-14.3	-19.5	-9.0	-2.4	5.1	8.2	-2.0											
0.6			23.0	12.5	2.2	3.9	6.2	4.0											
0.7			-18.8	5.4	17.4	0.9	1.6	6.4	26.8										
0.8			2.7	44.3	8.5	6.9	2.5	-0.2	-1.9										
0.9					-2.1	1.4	2.7	-6.6	0.7	-16.8									
1					11.1	-0.8	-2.2	0.4	-5.0	8.3									
1.1					-1.3	0.2	5.2	2.7	0.8	6.4	2.1	-5.1							
1.2					21.1	-3.5	-10.4	-2.4	-8.4	-0.9	2.4	7.0							
1.3					1.9	2.8	2.4	-4.3	0.0	-7.3	-2.5	9.7							
1.4					29.1	2.9	2.7	-5.4	1.9	-2.6	2.4	6.8	-11.9						
1.5					-2.2	-0.5	2.5	-5.6	0.4	0.7	10.0	-6.4	-4.8						
1.6					-0.8	-3.2	1.5	1.2	-2.5	3.1	-4.4			-5.5					
1.7					11.1	-4.0	-9.5	-5.5	-9.0	4.1	5.6	-16.0	19.3						
1.8						7.0	0.8	-4.2	-3.0	5.9	-10.8	3.4	4.8						
1.9					10.9	-11.7	5.0	1.9	4.0	7.2	-0.4	5.6	-11.8						
2						-1.8	9.4	-3.7	1.0	1.3	7.8	0.2	4.7						
2.1					2.0	10.4		-5.4	-9.1	3.1	5.3	-2.5	5.8	-10.4					
2.2							1.2	7.6	0.6	-0.7	3.4	-1.2	3.1	-2.3					
2.3							2.9	5.4	-0.7	-5.6	-3.2	1.4	-1.5	14.6					
2.4					21.7			-4.1	0.5	2.9	3.4	-0.6							
2.5					12.1	7.1	16.1	-10.2	8.7	7.7	2.7	2.4	-0.1	7.4					
2.6							8.3	-5.5		0.8	2.9	3.7	0.4	2.9					
2.7								8.8	8.7	2.0	-3.5	4.3	-1.6			-21.0			
2.8						22.8	-0.4	-5.0	-1.2	-9.8	2.4	3.7	1.3						
2.9								9.8	-3.3	9.8	3.1	4.5	3.6	-6.9					
3								-8.0		5.7	1.0	0.6	4.6	-18.5	3.0				
3.1									-15.1		-8.2	-0.9	7.4						
3.2									19.4		-5.3	-3.9	-2.2	-0.5					
3.3								-13.4		-0.2	8.4	-13.9	7.0	2.0	-9.6				
3.4									1.3	4.7	-8.4	3.0	-7.9	-10.7					
3.5										-18.4	10.4	-7.4	-8.2	-12.4					
3.6										6.4			-10.1						
3.7												-8.4	-5.8	-12.2					
3.8											-12.7	12.0	-26.7	-20.9					
3.9												1.8	-22.8						
4																		-17.3	

6.2.3 Grafische voorstelling van de tarieven

Ten slotte is nog een interessante voorstelling de modellen uit te tekenen per hoogteklaas en wel zo dat ze alleen binnen het bereik liggen van de dataset. Voor de hoogste klasse ($H= 50\text{ m}$) zijn er maar twee waarnemingen.



Figuur 6-5 Beuk: grafische voorstelling van de tarieven per hoogteklasse (rechts in de log-schaal).

6.3 Tarieven met één ingang (veeltermmodel)

6.3.1 De tarieven in tabelvorm

De vergelijkingen zijn:

$$V = \begin{cases} Z \leftrightarrow 0.1354 + -0.6449 C + 1.4252 C^2 + 0.0276 C^3 \\ B \leftrightarrow -0.1140 + 0.1295 C + 0.8115 C^2 + 0.0906 C^3 \\ V \leftrightarrow -0.0990 + 0.0865 C + 0.6974 C^2 + 0.0672 C^3 \\ K \leftrightarrow 0.0501 + -0.4841 C + 1.4592 C^2 + -0.2409 C^3 \end{cases} \quad (6.2)$$

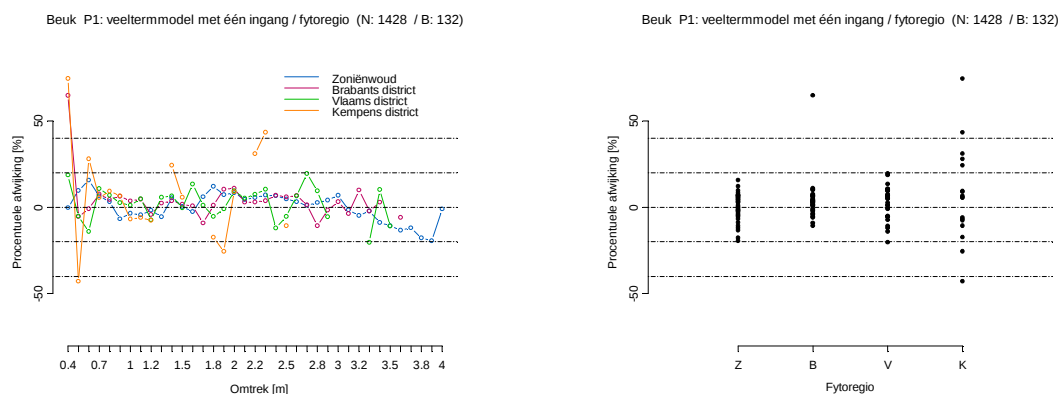
Tabel 6-4 geeft hiervan een eerste rudimentaire tabel waarbij we rekening houden met het bereik in elke ecoregio.

Tabel 6-4 Beuk: tarief met één ingang (veeltermmodel) en ecoregio in dm³. De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1). De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd.

	Zoniënwoud	Brabants district	Vlaams district	Kempens district
0.4	107	73	51	74
0.5	173	165	127	143
0.6	267	275	218	233
0.7	392	405	326	344
0.8	546	555	451	473
0.9	730	726	593	621
1	943	918	752	784
1.1	1187	1131	929	962
1.2	1461	1366	1125	1154
1.3	1766	1625	1340	•1357
1.4	2102	1906	1573	1571
1.5	2468	2212	1827	1794
1.6	2865	2541	2100	•2024
1.7	3293	2896	2394	•2260
1.8	3753	3276	2708	2501
1.9	4244	3683	3044	2745
2	4767	4115	3401	2991
2.1	5322	4575	3780	•3237
2.2	5908	5063	4182	3482
2.3	6527	5578	4607	3725
2.4	7178	6123	5054	•3963
2.5	7862	6696	5526	4195
2.6	8578	7300	6021	•4421
2.7	9327	7934	6541	
2.8	10109	8598	7086	
2.9	10924	9295	7656	
3	11772	10023	•8251	
3.1	12654	10784	•8873	
3.2	13570	11577	•9521	
3.3	14519	12405	10196	
3.4	15503	13266	10898	
3.5	16520	•14163	11628	
3.6	17572	15094	•12386	
3.7	18658	•16061		
3.8	19779			
3.9	20934			
4	•22125			

6.3.2 Controle per klasse (Tabel 6-5 en Figuur 6-6)

Bijna alle relatieve verschillen liggen tussen -20 % en 20 %. Er zijn een paar uitbijters voor kleine bomen waar het model minder precies is. Ook voor de Kempen komen een paar uitbijters voor. De afwijkingen vertonen geen patroon in functie van omtrek behalve tot ongeveer 3 m. Vanaf dan lijkt het model het volume te onderschatten. Bijna al de gegevens uit dat interval komen uit Zoniën.



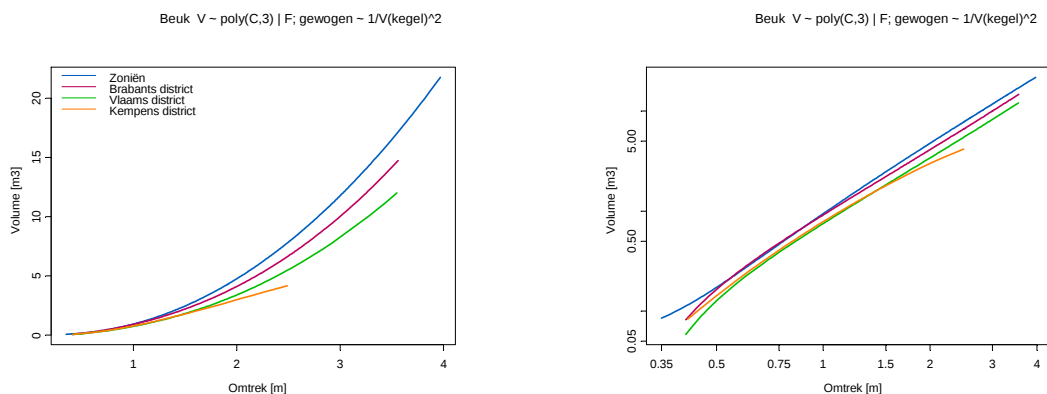
Figuur 6-6 Beuk: procentuele afwijking van het gemiddelde van de volumemetingen per combinatie van omtrek- en ecoregio ten opzichte van de waarde in de tarieventabel.

Tabel 6-5 Beuk: tarief met één ingang (veeltermmodel) en ecoregio. De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd.

	Zoniënwoud	Brabants district	Vlaams district	Kempens district
0.4	-0.3	64.7	18.6	74.4
0.5	9.7	-5.4	-5.2	-42.9
0.6	15.7	-0.9	-14.1	27.9
0.7	6.4	7.5	10.7	5.4
0.8	3.2	4.4	6.9	9.3
0.9	-6.7	6.4	2.5	6.5
1	-3.5	3.6	1.1	-6.8
1.1	-4.5	4.7	4.8	-6.0
1.2	-1.8	-4.2	-7.3	-7.6
1.3	-5.5	2.3	5.7	
1.4	5.3	3.8	6.5	24.2
1.5	0.4	1.7	-0.4	5.7
1.6	-2.6	0.7	13.4	
1.7	6.0	-9.1	1.1	
1.8	12.1	1.1	-5.4	-17.4
1.9	7.2	10.3	-1.0	-25.6
2	8.3	10.9	9.4	9.0
2.1	4.4	2.9	5.2	
2.2	5.6	3.0	7.4	31.0
2.3	7.0	3.7	10.4	43.3
2.4	6.6	6.9	-12.1	
2.5	4.7	5.9	-5.3	-10.7
2.6	3.2	6.8	6.5	
2.7	1.0	1.4	19.5	
2.8	2.6	-10.7	9.4	
2.9	4.1	-1.6	-5.6	
3	6.9	3.2		
3.1	-1.5	-3.7		
3.2	-4.8	9.9		
3.3	-2.2	-2.2	-20.3	
3.4	-8.8	2.8	10.2	
3.5	-10.7		-11.0	
3.6	-13.3	-5.9		
3.7	-11.9			
3.8	-17.7			
3.9	-19.5			
4	-1.1			

6.3.3 Grafische voorstelling van de tarieven

Aangezien we hier maar één variabele als ingang hebben, kunnen we rechtstreeks Figuur 4-15 overnemen.



Figuur 6-7 Beuk: grafische vergelijking van de veeltermmodellen (P1f) in de originele schaal (links) en in de log-schaal (rechts).

6.4 Tarieven met één ingang (logaritmisch model)

6.4.1 De tarieven in tabelvorm

De vergelijkingen per ecoregio zijn:

$$\log_{10}(V) = \begin{cases} Z \leftrightarrow 0.932 + 2.336 \log_{10}(C) \\ B \leftrightarrow 0.895 + 2.224 \log_{10}(C) \\ V \leftrightarrow 0.719 + 2.306 \log_{10}(C) \\ K \leftrightarrow 0.747 + 2.159 \log_{10}(C) \end{cases} \quad (6.3)$$

Uit (6.3) kunnen we onmiddellijk het volume afleiden voor een omtrek van 1 m. De logaritme hiervan is 0 en bijgevolg is de interpretatie van het intercept gelijk aan de waarde bij 1 m. Hieruit volgt dat bij 1 m Zoniën het hoogst ligt, kort gevolgd door het Brabants district (37 dm³ verschil). De volumes in het Vlaams district en het Kempens district liggen dan al tot 200 dm³ lager. Bij 1 m is het voorspelde volume in het Kempens district hoger dan in het Vlaams district. Maar aangezien de helling in de log-schaal groter is in het Vlaamse district, zal vanaf ongeveer 1,30 m het voorspelde volume in het Vlaamse district groter zijn.

Markant is dat de hellingscoëfficiënt in het Brabants district lager ligt dan in het Kempens district. In hoeverre dat een artefact is, is niet te achterhalen, want in het bereik dat het volume in het Vlaamse district groter wordt dan het Brabants district, hebben we geen gegevens. Door de tweede en de derde vergelijking in (6.3) aan elkaar gelijk te stellen, bekomen we dat hiervoor de omtrek ongeveer 140 m moet zijn. In de praktijk stelt zich dus totaal geen probleem en liggen de waarden voor het Brabants district uniform hoger dan het Kempens district.

Tabel 6-6 vertaalt de formule in een tabel met omtrek in stappen van 0.1 m waarbij we het bereik laten afhangen per ecoregio. De conclusies op basis van de formules worden hier bevestigd. De rangorde is nagenoeg monotoon over het volledige omtrekbereik. De schattingen voor Zoniën liggen systematisch hoger dan voor het Brabants district en de schattingen voor het Brabants district liggen systematisch hoger dan het Vlaams district. Alleen het Kempens district doorbreekt het patroon: tot 1.30 m zijn de houtvolumes hoger in het Kempens district; vanaf dan zijn de schattingen hoger in het Vlaams district.

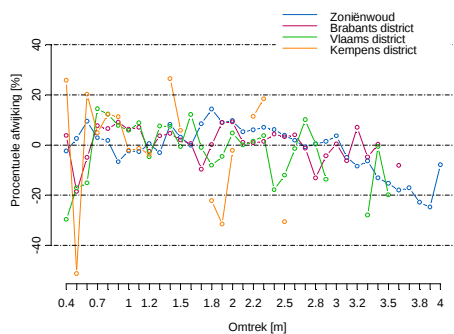
Tabel 6-6 Beuk: tarief met één ingang op basis van het logaritmisch model in dm³. De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1). De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd.

	Zoniënwoud	Brabants district	Vlaams district	Kempens district
0.4	110	117	87	103
0.5	185	191	145	167
0.6	283	287	221	248
0.7	405	405	316	346
0.8	554	545	430	461
0.9	729	708	564	595
1	932	895	719	747
1.1	1165	1106	895	917
1.2	1428	1342	1094	1107
1.3	1721	1604	1316	•1315
1.4	2046	1892	1561	1544
1.5	2404	2205	1831	1792
1.6	2796	2546	2125	•2059
1.7	3221	2913	2443	•2347
1.8	3681	3308	2788	2656
1.9	4177	3731	3158	2984
2	4708	4182	3554	3334
2.1	5277	4661	3978	•3704
2.2	5883	5169	4428	4096
2.3	6526	5707	4906	4508
2.4	7208	6273	5412	•4942
2.5	7930	6870	5947	5397
2.6	8691	7496	6510	•5874
2.7	9492	8152	7102	
2.8	10333	8839	7723	
2.9	11216	9557	8374	
3	12140	10305	•9055	
3.1	13107	11085	•9767	
3.2	14116	11896	•10509	
3.3	15168	12739	11281	
3.4	16264	13613	12086	
3.5	17403	•14520	12921	
3.6	18587	15459	•13789	
3.7	19816	•16431		
3.8	21090			
3.9	22409			
4	•23774			

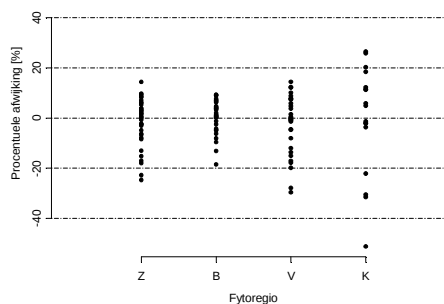
6.4.2 Controle van de tarieven

Tabel 6-7 en Figuur 6-8 tonen dat er bijna geen patroon is in de afwijkingen en er zijn ook heel weinig uitschieters. Opnieuw zijn de resultaten in het Kempens district het minst nauwkeurig. In het Zoniënwoud worden de volumes steeds meer onderschat in het bereik boven de 3 m. Meten van de hoogte lijkt aangewezen (maar ook daar neigen de formules naar een onderschatting).

Beuk L1f: logaritmisch model met één ingang per fyto regio (N: 1428 / B: 132)



Beuk L1f: logaritmisch model met één ingang per fyto regio (N: 1428 / B: 132)



Figuur 6-8 Beuk: procentuele afwijking van het gemiddelde van de volumemetingen per combinatie van omtrek- en ecoregio ten opzichte van de waarden uit de tabel.

Tabel 6-7 Beuk: vergelijking van het tarief met één ingang op basis van het logaritmisch model ten opzichte van het gemiddelde in elke klasse. De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).

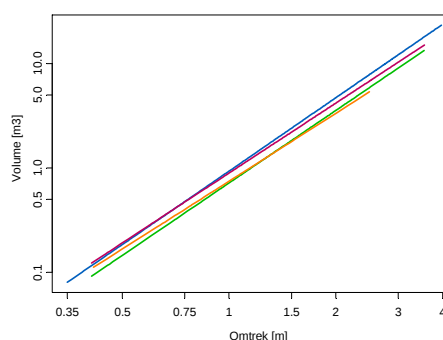
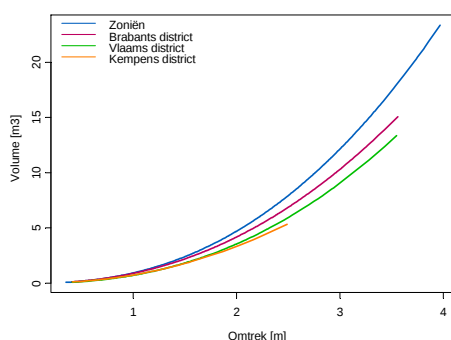
	Zoniënwood	Brabants district	Vlaams district	Kempens district
0.4	-2.4	3.8	-29.7	25.8
0.5	2.6	-18.6	-17.1	-51.3
0.6	9.4	-5.0	-15.1	20.2
0.7	2.9	7.7	14.4	4.8
0.8	1.8	6.4	12.3	12.2
0.9	-6.7	9.1	7.8	11.2
1	-2.4	6.2	5.8	-2.1
1.1	-2.6	7.0	8.8	-1.4
1.2	0.5	-2.5	-4.7	-3.7
1.3	-3.1	3.6	7.5	
1.4	8.2	4.6	7.3	26.4
1.5	3.0	2.0	-0.6	5.8
1.6	-0.1	0.6	12.1	
1.7	8.4	-9.7	-1.0	
1.8	14.3	0.1	-8.1	-22.2
1.9	9.0	8.9	-4.6	-31.6
2	9.7	9.2	4.7	-2.2
2.1	5.3	1.0	0.0	
2.2	6.0	0.9	1.5	11.4
2.3	7.0	1.4	3.7	18.4
2.4	6.1	4.4	-17.9	
2.5	3.9	3.2	-12.0	-30.6
2.6	1.8	4.0	-1.5	
2.7	-0.8	-1.3	10.0	
2.8	0.4	-13.2	0.4	
2.9	1.4	-4.3	-13.7	
3	3.7	0.4		
3.1	-4.9	-6.3		
3.2	-8.5	7.0		
3.3	-6.4	-4.8	-28.0	
3.4	-13.1	0.2	-0.7	
3.5	-15.3		-19.9	
3.6	-18.1	-8.1		
3.7	-17.1			
3.8	-22.8			
3.9	-24.8			
4	-7.9			

6.4.3 Grafische voorstelling van de tarieven

Voor de volledigheid nemen we hier Figuur 4-30 nog eens over. De grafieken zijn geordend: $Z > B > V$ (de curve voor het Kempens district laten we buiten beschouwing omdat we te weinig waarnemingen hebben over een te smal bereik).

Beuk: L1f: logaritmisch model met één ingang per fyto regio (N: 1428 / B: 132)

Beuk: L1f: logaritmisch model met één ingang per fyto regio (N: 1428 / B: 132)



Figuur 6-9 Beuk: vergelijking van de logaritmische modellen (links in de originele schaal rechts en in de log-schaal).

6.5 Tarief voor grote omtrekken

Tot besluit vergelijken we de voorspellingen van de modellen voor de waarnemingen die we aan de kant hebben gezet bij de modelbouw (Tabel 6-8). Alle bomen komen van het Zoniënwood. Op basis van deze kleine steekproef kunnen we uiteraard geen besluiten trekken. Voor tarieven met twee ingangen, scoort het relatief model (R2) heel goed; het veeltermmodel P2 en het lo-

garitmisch model L2 iets minder. Voor tarieven met één ingang zijn de voorspellingen systematisch te laag. Bij het veeltermmodel is de onderschatting iets minder groot.

Tabel 6-8 Eik: vergelijking van de modellen voor de waarnemingen met een omtrek groter dan 3m. F: = voorspelde waarde (*fitted value*) en R: = gestandaardiseerd verschil (*standardized residual*).

	Ecoregio	Omtrek	Hoogte	Volume	Vkegel	relV
1118	Zoniënwoud	4.63	41.8	24.0	23.8	1.01
1402	Zoniënwoud	4.22	42.3	20.1	20.0	1.01
1429	Zoniënwoud	4.30	41.9	21.6	20.6	1.05

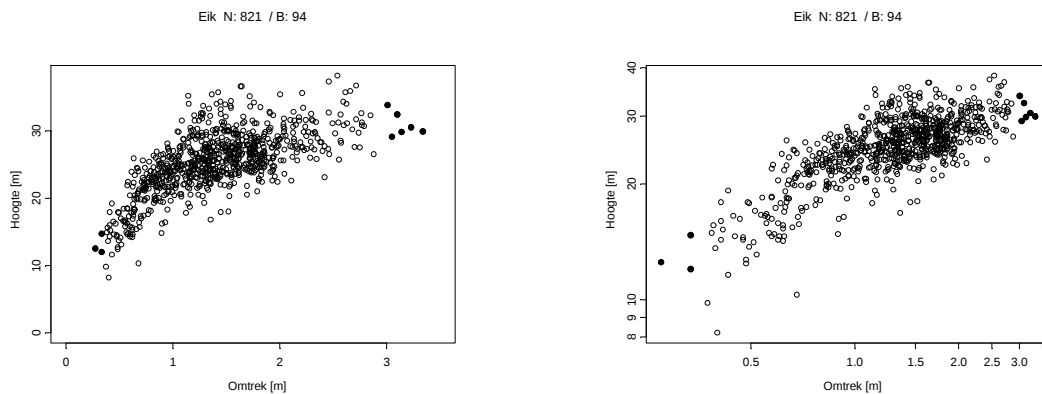
	F:P2	R:P2	F:L2	R:L2	F:R2	R:R2
1118	26.2	-8.7	26.2	-8.7	24.5	-2.1
1402	22.5	-10.9	22.4	-10.3	21.4	-6.1
1429	23.0	-6.3	22.9	-5.9	21.8	-1.0

	F:P1f	R:P1f	F:L1f	R:L1f
1118	30.4	-21.3	33.5	-28.4
1402	24.9	-19.2	26.9	-25.5
1429	25.9	-16.7	28.2	-23.3

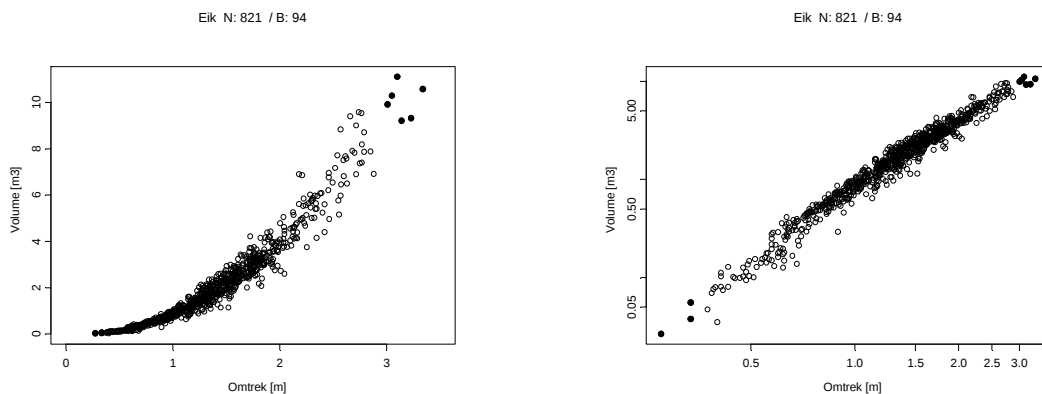
7 Eik: verkenning van de gegevens (EDA)

7.1 Het meetbereik en selectie van de gegevens

Het bereik van de statistische modellen is sterk bepaald door het meetbereik van de gegevens. Figuur 7-1 en Figuur 7-2 geven hiervan een eerste beeld. In totaal gaat het om 821 waarnemingen (N) gespreid over 94 bestanden (B). Negen punten (aangeduid) in het zwart liggen buiten het omtrekbereik 0.35 m tot 3 m. Deze punten hebben we niet meegenomen in de modellering want het zijn hefboom punten: ze liggen afgezonderd van het normale bereik en hebben potentieel een grote invloed op de coëfficiënten. Deze selectie verandert weinig of niets aan de reikwijdte van het model. Ongeacht we deze punten zouden meenemen of niet, de tarieven zijn in essentie geldig voor omtrekken tussen 0.35 m en 3 m. Extrapolatie is slechts in beperkte mate mogelijk. Om dat verder te motiveren, zullen we een keer de modellen geselecteerd en gecalibreerd zijn, de reikwijdte van de modellen onderzoeken. Ook zullen we dan nagaan hoe goed de modellen de weggelaten punten voorspellen.



Figuur 7-1 Eik: verkenning van de relatie tussen omtrek en hoogte. Links: originele schaal, rechts: log-schaal. De punten in het zwart geven de waarnemingen aan buiten het bereik van 0.35 m tot 3 m.



Figuur 7-2 Eik: verkenning van de relatie tussen omtrek en volume. Links: originele schaal, rechts: log-schaal. De punten in het zwart geven de waarnemingen aan buiten het bereik van 0.35 m tot 3 m.

7.2 Opbouw van de dataset

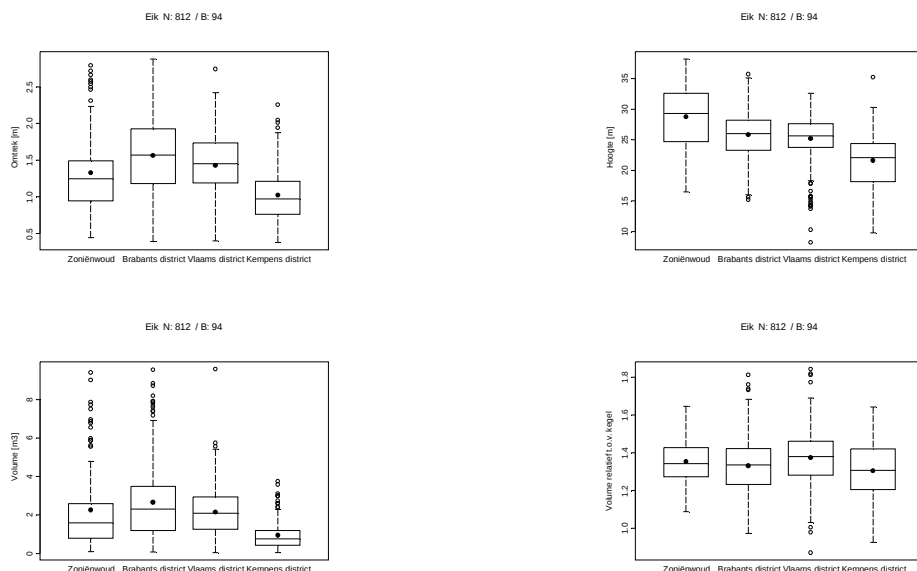
Na het weglaten van de negen waarnemingen blijven 812 metingen verspreid over 94 bestanden (N: 821 / B: 94). De meeste gegevens komen uit het Brabants en Vlaams district (samen ongeveer 70 %). De representativiteit van de steekproef komt aan bod in deel 1 van het technisch rapport. De waarnemingen zijn geen aselechte steekproef uit de populatie, maar toch bekomen we wellicht een eerste beeld van de verschillen tussen de ecoregio's. Tabel 7-1 en de

enkele grafieken (boxplots in Figuur 7-3 en cumulatieve distributies Figuur 7-4) maken een verkennende analyse.

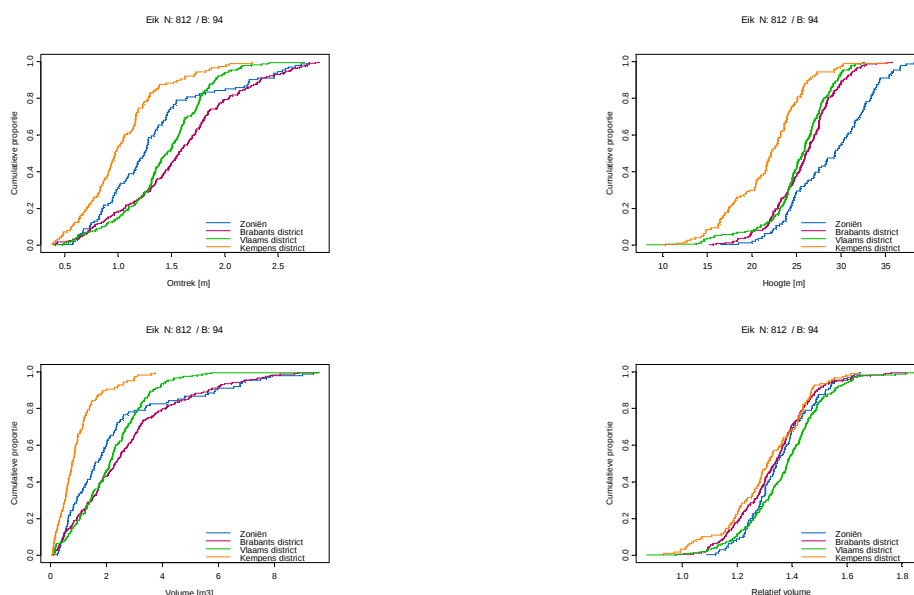
Volgens de boxplots zijn de bomen met de grootste omtrek afkomstig uit het Brabantse en Vlaamse district. Maar de bomen uit het Zoniënwoud zijn wel hoger. **Onder meer als gevolg hiervan, is de distributie van het volume in het Zoniënwoud heel asymmetrisch.** De mediaan van het volume ligt lager dan in het Brabants en Vlaams district, maar de gemiddelde waarde is bijna even groot. Over de hele lijn zijn de bomen afkomstig van het Kempens district het kleinst.

Tabel 7-1 Eik: enkele richtgetallen, alle districten samen en opgesplitst per district.

	District				
	Alle	Zoniën	Brabants	Vlaams	Kempens
Aantallen					
+ metingen (bomen)	812	117	279	284	132
+ percentage metingen	100%	14.4 %	34.4 %	35.0 %	16.3 %
+ bestanden	94	22	33	22	17
+ bomen/bestand (mediaan)		3.5	5	6	7
+ bomen/bestand (bereik)	–	1 – 19	1 – 36	1 – 58	1 – 26
Omtrek [m]					
+ bereik	0.38 – 2.88	0.45 – 2.79	0.39 – 2.88	0.40 – 2.74	0.38 – 2.26
+ 5% - 95% percentiel	0.60 – 2.34	0.64 – 2.50	0.65 – 2.58	0.62 – 2.04	0.47 – 1.77
+ mediaan / gemiddelde	1.36 / 1.39	1.25 / 1.33	1.57 / 1.56	1.45 / 1.43	0.97 / 1.02
Hoogte [m]					
+ bereik	8.2 – 38.2	16.5 – 38.2	15.2 – 35.7	8.2 – 32.6	9.8 – 35.2
+ 5% / 95% percentiel	16.7 – 32.3	21.3 – 35.7	19.6 – 31.3	15.8 – 30.1	14.4 – 28.2
+ mediaan / gemiddelde	25.5 / 25.3	29.3 / 28.7	26.0 / 25.8	25.6 / 25.2	22.1 / 21.6
Volume [m³]					
+ bereik	0.035 – 9.579	0.101 – 9.399	0.074 – 9.541	0.035 – 9.579	0.047 – 3.751
+ 5% / 95% percentiel	0.182 – 5.744	0.347 – 6.874	0.301 – 6.895	0.178 – 4.145	0.099 – 2.657
+ mediaan / gemiddelde	1.791 / 2.144	2.585 / 2.259	2.300 / 2.660	2.097 / 2.146	0.764 / 0.945
Relatief volume [-]					
+ bereik	0.87 – 1.84	1.09 – 1.65	0.97 – 1.81	0.87 – 1.84	0.93 – 1.64
+ 5% / 95% percentiel	1.10 – 1.58	1.17 – 1.55	1.10 – 1.56	1.13 – 1.61	1.02 – 1.53
+ mediaan / gemiddelde	1.35 / 1.35	1.34 / 1.35	1.34 / 1.33	1.38 / 1.38	1.31 / 1.30



Figuur 7-3 Eik: boxplots van omtrek, hoogte, volume en relatief volume. De punten geven het gemiddelde aan.



Figuur 7-4 Eik: cumulatieve distributie van omtrek, hoogte, volume en relatief volume.

7.3 De referentieboom en het referentiebereik per ecoregio

Ondanks de regionale verschillen zoeken we hier toch naar een globale referentieboom om de modellen op een eenvoudige manier met elkaar te kunnen vergelijken. Het is immers niet eenvoudig om de invloed van kleine verschillen in de coëfficiënten snel in te schatten en daarom zullen we bij elk model de waarde berekenen voor een referentieboom. In een later stadium zullen we de modellen over het volledige bereik vergelijken, maar voor een eerste elementaire bespreking een referentiepunt handig.

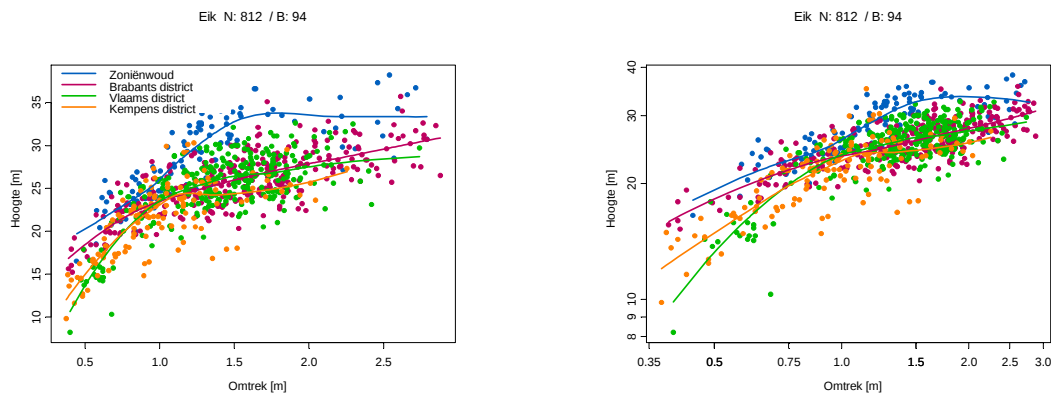
Hiertoe gaan we op zoek naar de boom met gemiddelde karakteristieken op basis van de mediaan van hoogte en omtrek. Hierbij keken we op de eerste plaats naar het Brabantse en Vlaamse district met 70 % van de waarnemingen. Volgens Tabel 7-1 is voor deze districten de mediaan van de omtrek respectievelijk 1.57 m en 1.45 m en de mediaan van de hoogte 26.0 en 25.6 m. Om af te ronden kozen we op basis hiervan als **referentieboom C = 1.5 m en H = 25 m**. Voor het Kempens district is de referentieboom groot, maar toch niet echt uitzonderlijk. Voor Zoniën is de referentieboom een eerder lage boom volgens de omtrek die nogal aan de hoge kant is.

Tabel 7-2 Eik: referentieboom en referentiebereik per ecoregio

	Ecoregio				
	Alle	Zoniën	Brabants	Vlaams	Kempens
Omtrek [m]	1.5	1.2 – 1.3	1.5 – 1.6	1.4 – 1.5	0.9 – 1.0
Hoogte [m]	25	27.5 – 30.0	25 – 27.5	25 – 27.5	20.0 – 22.5
Volume [m³]	2.1	2.6	2.3	2.1	0.75

7.4 Relatie tussen omtrek en hoogte

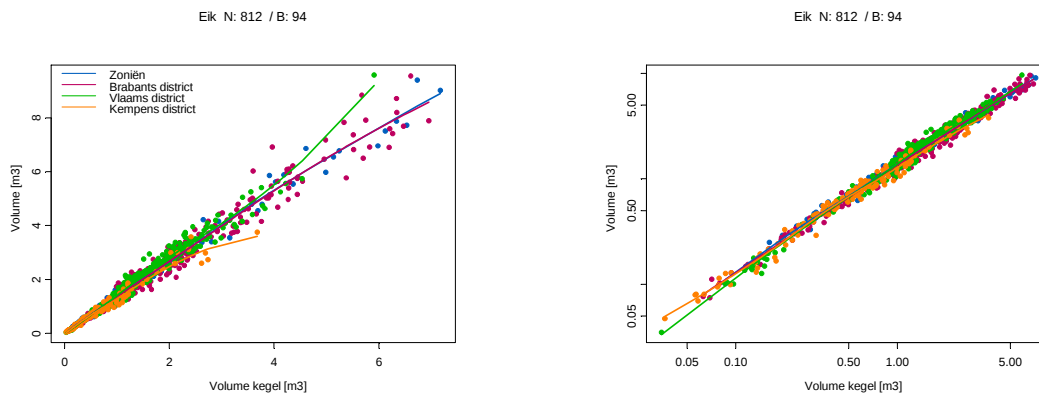
De hoogte vlakt sterk af bij hoge waarden van de omtrek (Figuur 7-5), ook in de log-schaal. Bij een gelijke omtrek zijn de bomen in Zoniën veel hoger dan in de andere districten. In het Brabants district zijn de bomen ook hoger, maar alleen bij kleinere bomen tot 1 m omtrek. Aangezien de kwaliteit van de standplaats de hoogte sterk bepaalt en de gemiddelde standplaats eigenschappen sterk gecorreleerd is met ecoregio, verwachten we een sterke correlatie tussen hoogte en ecoregio. Bij een gelijke omtrek zal het volume dus verschillen naargelang ecoregio. Om hiervoor te corrigeren moeten we de modellen met één ingang opsplitsen per ecoregio.



Figuur 7-5 Eik: relatie tussen omtrek en hoogte (links: originele schaal, rechts: log-schaal).

7.5 Verkenning van het model met twee ingangen

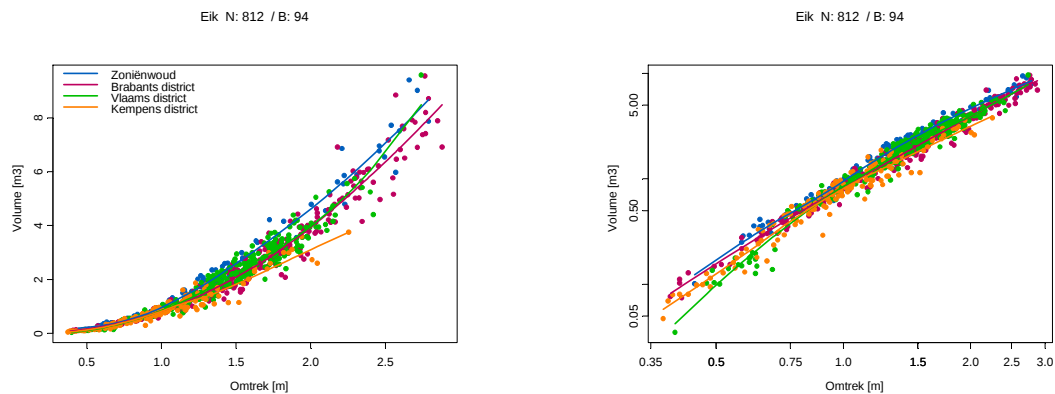
De modellen met twee ingangen zijn geïnspireerd op het kegelvolume. Hier verkennen we in hoeverre deze modellen plausibel zijn. In de originele schaal (Figuur 7-6, figuur links) zien we een haast lineair verband die wat afbuigt naar beneden bij grotere bomen. De verschillen tussen de ecoregio's zijn klein. Alleen voor grote bomen buigen het Vlaamse district en het Kempen district af (in tegengestelde richtingen). Maar het aantal gegevens daar is vrij beperkt en toeval kan een grote rol spelen. Van groter belang is dat de spreiding toeneemt met de grootte van de boom. Een gewogen lineaire regressie zal nodig zijn. In de log-schaal (Figuur 7-6, figuur rechts) is het verband nagenoeg lineair. Ook blijkt de variantie gestabiliseerd door log-transformatie. De variatie rond de rechte is gelijk over het hele bereik. Bijgevolg is een gewone lineaire regressie in de log-schaal zonder gewichten voldoende.



Figuur 7-6 Eik: relatie tussen het volume van de kegel en volume (links: originele schaal, rechts: log-schaal).

7.6 Verkenning met het model met één ingang (en ecoregio)

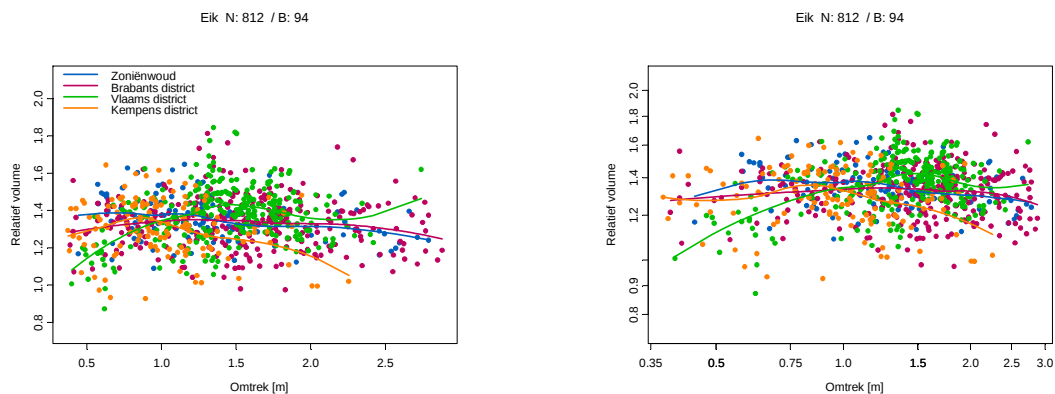
We verkennen hier de relatie tussen omtrek en volume om de basisvorm van de modellen met één ingang te onderzoeken (Figuur 7-7). In tegenstelling met de vorige figuur zijn er nu wel duidelijke verschillen naargelang de ecoregio. Zoniën ligt hoger dan het Vlaamse en Brabantse district die goed op elkaar gelijken. Het Kempen district ligt lager, behalve voor kleine waarde van de omtrek (zie log-schaal). Een apart model per ecoregio zal wellicht nodig zijn. In de originele schaal neemt de variantie toe met het volume. Een gewogen regressie is nodig. In de log-schaal is de relatie tussen omtrek en volume nagenoeg lineair. De variantie is er gestabiliseerd.



Figuur 7-7 Eik: relatie tussen omtrek en volume (links: originele schaal, rechts: log-schaal).

7.7 Verkenning van het relatief model

Het relatief model gaat uit van de veronderstelling dat de verhouding tussen het volume van de boom en het volume van een kegel berekend op basis van de hoogte en de omtrek nagenoeg een constante is of op zijn minst eenvoudig te modeleren is in functie van hoogte en omtrek. Hier onderzoeken we deze relatie (Figuur 3-6). In Zoniën en het Brabants district is het verloop gelijkaardig (zacht dalend), in het Vlaamse district hebben we eerder een stijging en in de Kempen is er een relatief sterke daling.



Figuur 7-8 Eik: relatie tussen omtrek en het volume relatief ten opzichte van het volume van een kegel (links: originele schaal, rechts: log-schaal).

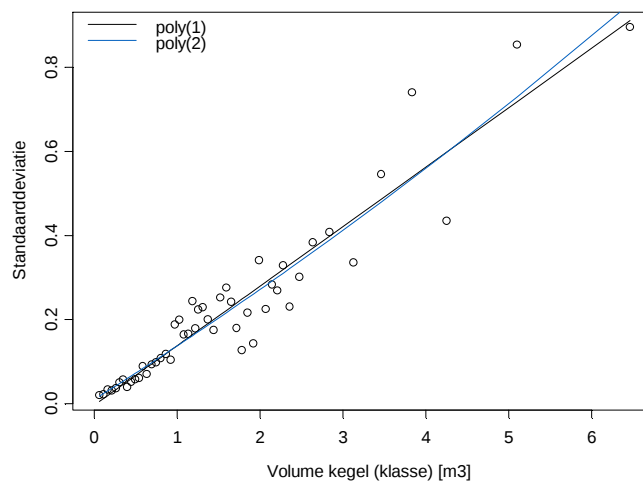
7.8 Variantie van het volume in functie van de omtrek

Een belangrijke voorwaarde voor lineaire regressie is dat de variantie van de punten om het model gelijk blijft als we deze uitzetten ten opzichte van de predictoren en de voorspellingen van het model (homoskedasticiteit). Dat is hier niet het geval: de variantie op het volume stijgt naarmate de bomen groter worden. Dat is heel dikwijls het geval met natuurlijke groeiprocessen. Transformeren naar de log-schaal helpt heel dikwijls om de variantie te stabiliseren. Een andere oplossing is te werken met een gewogen regressie. De gewichten moeten omgekeerd evenredig zijn met de variantie van het geschatte volume. Om de gewichten te bepalen, onderzoeken we de relatie tussen de standaardafwijking en het volume van de kegel (als benadering van het voorspelde volume dat we nog niet kennen). We werken met de standaardafwijking en niet met het kwadraat ervan (de variantie) om een stabielere figuur te krijgen. Figuur 7-9 toont de relatie tussen de standaardafwijking berekend in functie van het volume van de kegel. Deze relatie is nagenoeg lineair (Tabel 7-3 geeft een formele test). Bijgevolg moeten we de gewichten omgekeerd evenredig nemen met het kwadraat van het kegelvolume.

Tabel 7-3 Eik: ANOVA-tabel voor de veeltermmodellen van de standaardafwijking in functie van het volume van de kegel. De term $\text{poly}(\text{Vkegel}, 2)$ is een veelterm (polynoom) van de tweede graad. Aangezien de component van de tweede graad niet significant is en de component van de eerste graad wel, is de conclusie dat het verband tussen de standaardafwijking en de het volume van een kegel lineair is.

	Value	Std. Error	t_value	Pr(> t)
(Intercept)	0.223	0.008	26.209	0.000
$\text{poly}(\text{Vkegel}, 2)1$	1.305	0.060	21.741	0.000
$\text{poly}(\text{Vkegel}, 2)2$	0.062	0.060	1.032	0.307

Eik N: 812 / B: 94



Figuur 7-9 Eik: relatie tussen de variantie van de waarnemingen en het volume van een kegel (in klassen van 0.25 m³).

8 Eik: statistische analyse van de modellen

8.1 Veeltermmodellen

8.1.1 Veeltermmodel met twee ingangen

Startmodel

Het veeltermmodel met twee ingangen is in essentie een term evenredig met het volume van een kegel aangevuld met termen die functie zijn van de omtrek en de hoogte en de interactie tussen beide. We kunnen dit model neerschrijven in twee totaal equivalente varianten. Het enige verschil is dat de coëfficiënt van V_{kegel} een factor 12π groter zal zijn dan $C2H$.

$$P2 \leftrightarrow V \sim \begin{cases} C2H + C + C2 + C3 + H + CH & (\text{var1}) \\ V_{kegel} + C + C2 + C3 + H + CH & (\text{var2}) \end{cases} \quad V_{kegel} = \frac{C2H}{12\pi} \quad (8.1)$$

Interessant is het "referentiemodel" waarbij we alleen rekening houden met V_{kegel} . Hieruit kunnen we de (vorm)factor afleiden waarmee we het volume van een kegel gemiddeld moeten vermenigvuldigen om het werkelijke houtvolume te bekomen.

$$P2r \leftrightarrow V \sim V_{kegel} \quad (8.2)$$

Modelselectie (Tabel 8-1)

Uit een stapsgewijze regressie blijkt dat $C3$ niet significant is ($p = 0.396$). We kunnen deze term weglaten uit (4.1) en we krijgen als finaal model een interactie tussen H en een veelterm van de tweede orde in C :

$$V \sim C2H + C + C2 + H + CH = \text{poly}(C, 2) * H \quad (8.3)$$

Uit de ANOVA-tabel van het finaal veeltermmodel (Tabel 8-1) blijkt dat nadat de term evenredig met het volume van een kegel is ingevoerd, de andere termen niet meer veel toevoegen. Toch kunnen we geen ervan weglaten uit het model (zelfs al lijkt dat zo voor C en H , maar een keer alle termen in het model zijn, blijken zowel C als H significant, zie de tabel met coëfficiënten: Tabel 8-4).

Tabel 8-1 Eik: ANOVA-tabel van het finaal veeltermmodel met twee ingangen (stapsgewijs invoegen)

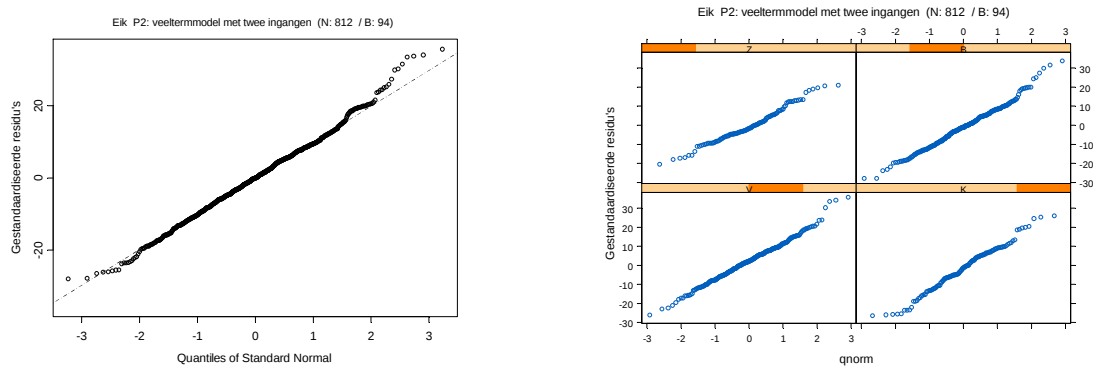
	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
C2H	1	1095	1095	56866	0.000
C	1	0	0	1	0.408
C2	1	0	0	7	0.008
H	1	0	0	1	0.224
CH	1	1	1	29	0.000
Residuals	806	16	0		

Residuanalyse (Figuur 8-1, Figuur 8-2 en Figuur 8-3)

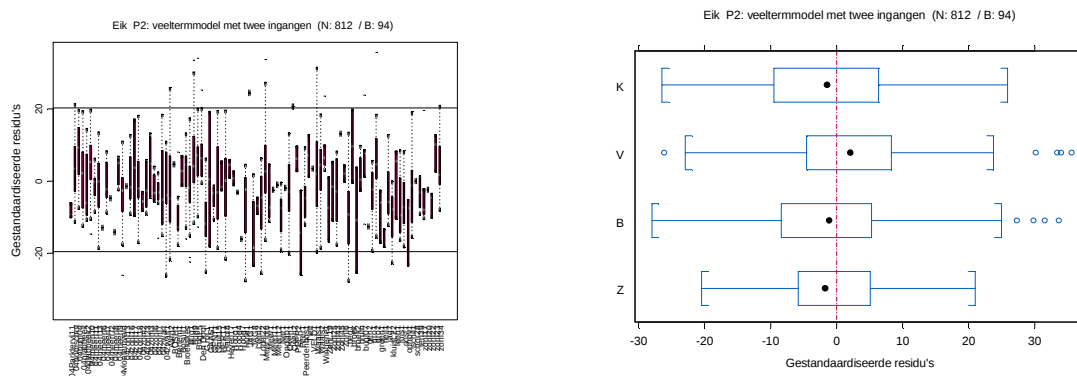
Het model sluit goed aan bij de gegevens. De diagnostische grafieken op basis van de gestandaardiseerde residu's vertonen weinig structuur.

- De QQ-plot wijkt alleen in de extremen af van een rechte. Ook per ecoregio vertonen deze grafieken geen grote anomalieën.
- De boxplots per ecoregio verschillen heel weinig. Bijgevolg hebben we geen aanwijzingen dat een apart model per district nodig is. Ook de boxplots per bestand zijn vrij homogeen. Uiteraard is er wat meer variatie omdat de boxplots gebaseerd zijn op een beperkt aantal waarnemingen.

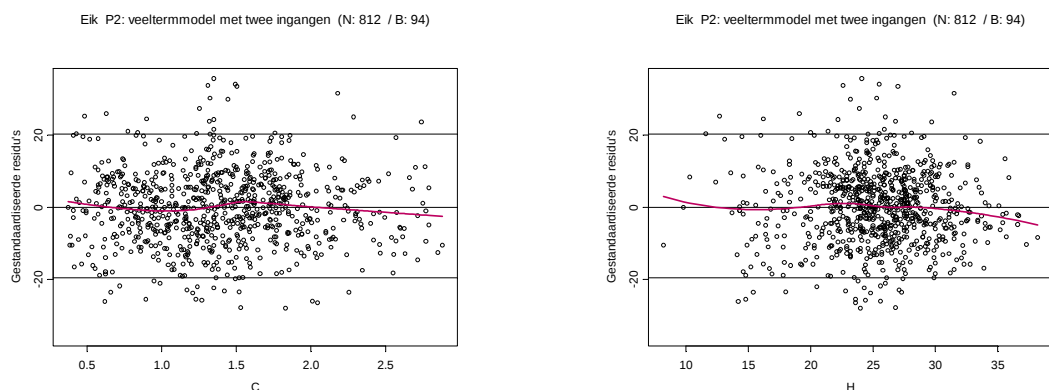
- We zien weinig structuur als we de residu's uitzetten in functie van de omtrek of de hoogte. De fout is na standaardisatie homoskedastisch over het volledige bereik van de omtrek en de hoogte. Vijfennegentig procent van de gestandaardiseerde residu's valt tussen $\pm 20\%$ en dat is meteen ook het 95 % voorspellingsinterval. Ten opzichte van de voorspelde waarde moeten we een marge nemen van 20 % opdat de werkelijke waarde in 95 % van de gevallen de werkelijke waarde zou bevatten.



Figuur 8-1 Eik: residuanalyse voor het veeltermmodel met twee ingangen (P2): QQ-grafieken



Figuur 8-2 Eik: residuanalyse voor het veeltermmodel met twee ingangen (P2): boxplots.



Figuur 8-3 Eik: residuanalyse voor het veeltermmodel met twee ingangen (P2): Relatie met omtrek en hoogte.

De coëfficiënten van het model (Tabel 4-2, Tabel 8-3 en Tabel 8-4)

Het veeltermmodel kunnen we interpretern als een term die het volume van een kegel voorstelt (C^2H) gecorrigeerd met de termen C , C^2 , H en CH . Daarom schatten we eerst het model met alleen de kegel term (Tabel 4-2). Het houtvolume is 1.36 keer het volume van een kegel.

Het houtvolume voor de referentieboom ($C = 1.5$ m en $H = 25$ m) is met dit model ongeveer 2 m^3 .

Tabel 8-2 Eik: coëfficiënten van het veeltermmodel met alleen kegel

	Value	Std. Error	t value	Pr(> t)
Referentie:	2.019			
(Intercept)	-0.00859	0.00185	-4.65	3.86e-006
Vkegel	1.35920	0.00582	233.59	0.00e+000

Door nu de andere termen in te voeren, verandert de coëfficiënt van Vkegel (Tabel 8-3). De reden is dat de toegevoegde termen hiermee gecorreleerd zijn. Deze termen nemen een deel van de voorspellende waarde van Vkegel over. De impact op de referentieboom is echter klein. Het voorspelde houtvolume is opnieuw ongeveer 2 m^3 .

Tabel 8-3 Eik: coëfficiënten van het veeltermmodel met twee ingangen (variant 1: met Vkegel en referentie)

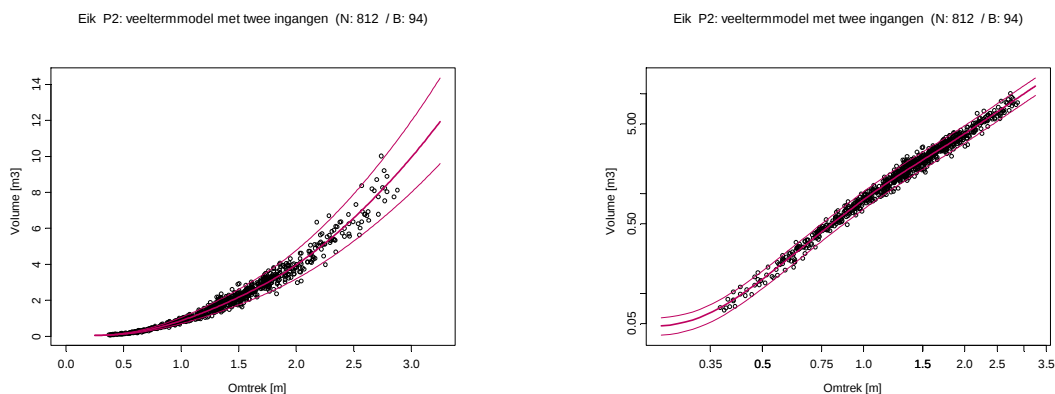
	Value	Std. Error	t value	Pr(> t)
Referentie:	2.025			
Vkegel	0.86644	0.11562	7.49	1.76e-013
C	-0.56118	0.11248	-4.99	7.44e-007
C2	0.29096	0.08266	3.52	4.56e-004
H	-0.00725	0.00167	-4.36	1.50e-005
CH	0.02496	0.00463	5.39	9.28e-008

Tabel 8-4 Eik: coëfficiënten van het veeltermmodel met twee ingangen (variant 2: met C2H en intercept)

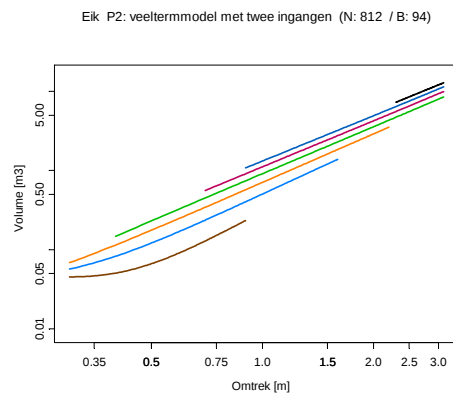
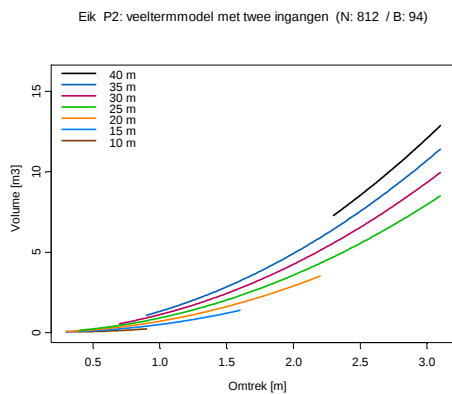
	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.16449	0.03421	4.81	1.82e-006
C2H	0.02298	0.00307	7.49	1.76e-013
C	-0.56118	0.11248	-4.99	7.44e-007
C2	0.29096	0.08266	3.52	4.56e-004
H	-0.00725	0.00167	-4.36	1.50e-005
CH	0.02496	0.00463	5.39	9.28e-008

Voorstelling van het model (Figuur 8-4).

De figuren stellen het volume als functie van omtrek voor in de originele schaal en in de log-schaal, waarbij we de hoogte gelijk nemen aan de verwachte waarde bij een gegeven omtrek. Impliciet neemt de hoogte hier dus toe in functie van de omtrek. We zien hoe de absolute voorspellingsmarge toeneemt, maar de relatieve voorspellingsmarge is een constante (ongeveer 20 %).



Figuur 8-4 Eik: grafiek van het veeltermmodel met twee ingangen (P2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de 95% voorspellingsintervallen.



Figuur 8-5 Eik: grafiek van het veeltermmodel met twee ingangen (P2) per hoogteklaas (rechts in de log-schaal).

8.1.2 Veeltermmodel met één ingang

Startmodel

Het veelterm met één ingang bekomen we door uit model (4.1) alle termen met hoogte erin te laten vallen:

$$P1 \leftrightarrow V \sim C + C2 + C3 = \text{poly}(C, 3) \quad (8.4)$$

Modelselectie (Tabel 8-5).

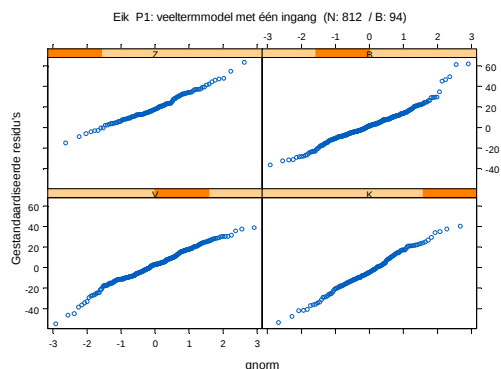
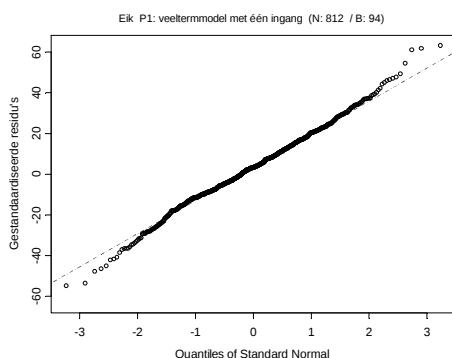
Alle termen van de derde orde veelterm zijn significant. De inbreng van de derde orde term is marginaal.

Tabel 8-5 Eik: ANOVA-tabel van het veeltermmodel met één ingang.

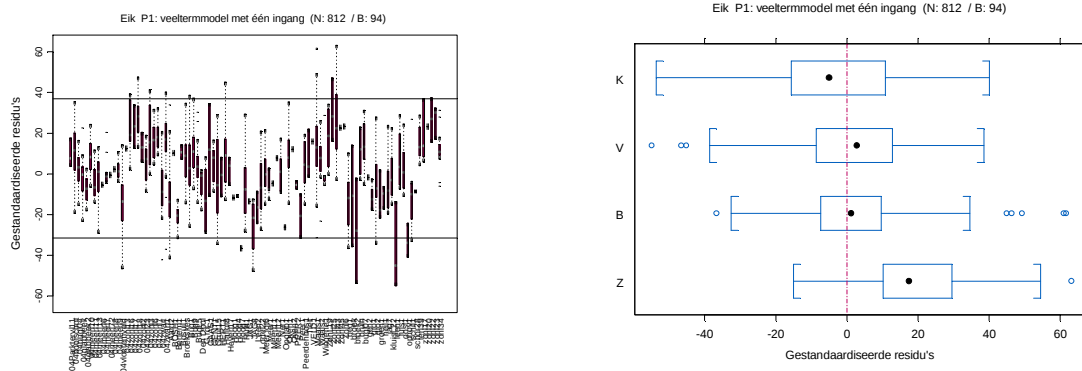
	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
C	1	946	946	17760	0.0000
C2	1	121	121	2277	0.0000
C3	1	0	0	6	0.0135
Residuals	808	43	0		

Residuanalyse (Figuur 8-6, Figuur 8-7 en Figuur 8-8).

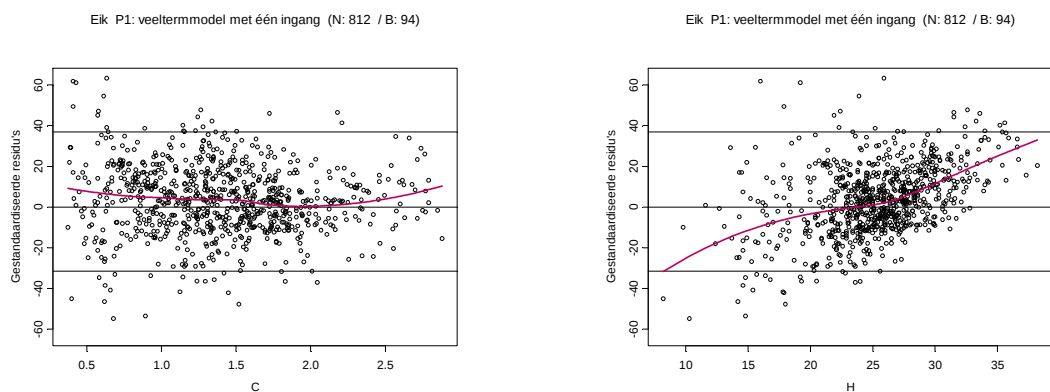
De QQ-plots zijn vrij goed. Omdat hoogte een belangrijke predictor is, die we hebben laten vallen, stijgen de residu's in functie van de hoogte. Omdat hoogte gecorreleerd is met ecoregio, zijn de gestandaardiseerde residu's gecorreleerd met ecoregio. Ten opzichte van omtrek is er geen patroon wat erop wijst dat het model op een juiste manier met omtrek rekening houdt. Ongeveer 95 % van de gestandaardiseerde residu's ligt tussen ± 40 %. Dit is een maat voor de voorspellingskwaliteit van het model.



Figuur 8-6 Eik: residuanalyse voor het veeltermmodel met één ingang (P1). QQ-plots.



Figuur 8-7 Eik: residuanalyse voor het veeltermmodel met één ingang (P1). Boxplots.



Figuur 8-8 Eik: residuanalyse voor het veeltermmodel met één ingang (P1). Relatie met omtrek en hoogte.

De coëfficiënten van het model (Tabel 8-6)

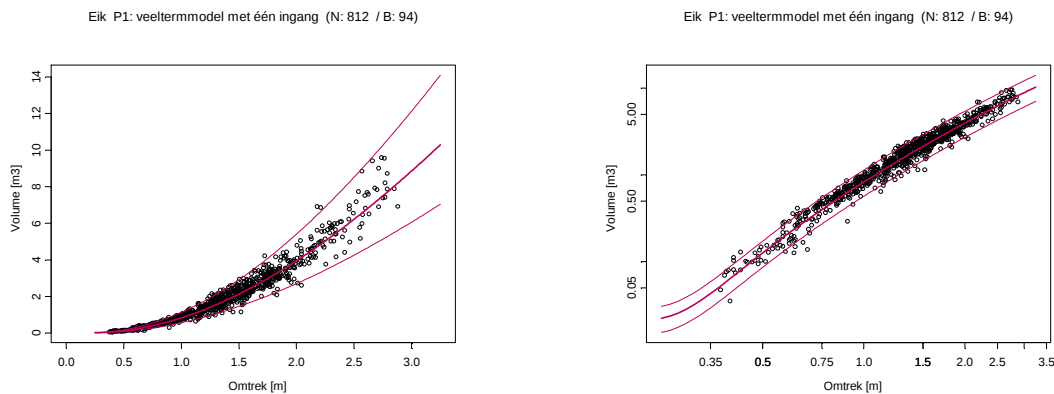
Aangezien het model met één tarief niet bruikbaar is, maar wel de versie opgesplitst per ecoregio, zijn de schattingen van weinig betekenis. Voor de volledigheid geven we toch de tabel. De referentieboom heeft een iets groter volume dan bij het model met twee ingangen.

Tabel 8-6 Eik: coëfficiënten van het veeltermmodel met één ingang.

	Value	Std. Error	t value	Pr(> t)
Referentie:	2.127	0.0162		
(Intercept)	0.096	0.0317	3.04	2.43e-003
C	-0.663	0.1185	-5.60	2.96e-008
C2	1.496	0.1283	11.66	0.00e+000
C3	-0.100	0.0406	-2.48	1.35e-002

Voorstelling van het model (Figuur 8-9).

Zoals uit de residuanalyse bleek, geeft het model wel een onvertkende schatting in functie van de omtrek. Maar de onzekerheid op de voorspelling is dubbel zo groot als bij het model met twee ingangen.



Figuur 8-9 Eik: grafiek van het veeltermmodel met één ingang (P1). Links = originele schaal & rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de relatieve residu's liggen.

8.1.3 Veeltermmodel met één ingang en ecoregio

Startmodel

Vraag is in hoeverre we het model niet beter opsplitsen naargelang ecoregio. Hiertoe testen we volgend interactiemodel uit:

$$P1f \leftrightarrow V \sim (C + C2 + C3) * Fytoregio \quad (8.5)$$

Modelselectie (Tabel 8-7)

De ANOVA-tabel toont aan dat Ecoregio significant is en een hoge bijdrage aan het model levert. Het is dus aanbevolen apart een model op te stellen per ecoregio:

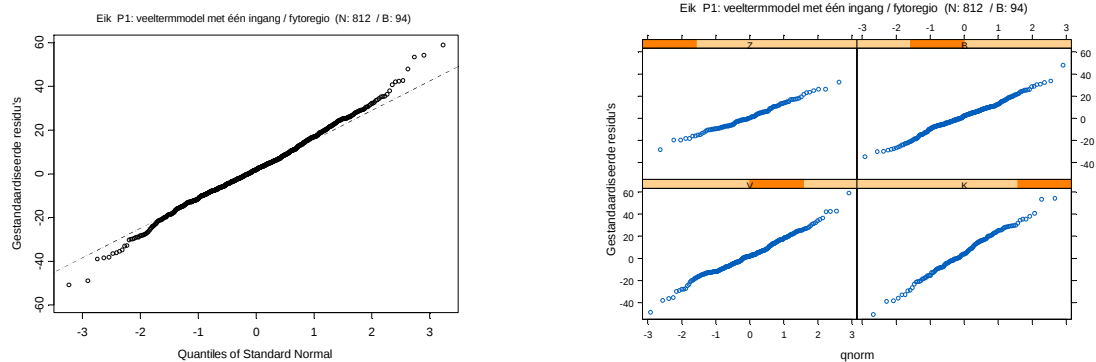
$$P1f \leftrightarrow V \sim (C + C2 + C3) \mid Fytoregio \quad (8.6)$$

Tabel 8-7 Eik: ANOVA-tabel van het veeltermmodel met één ingang en ecoregio.

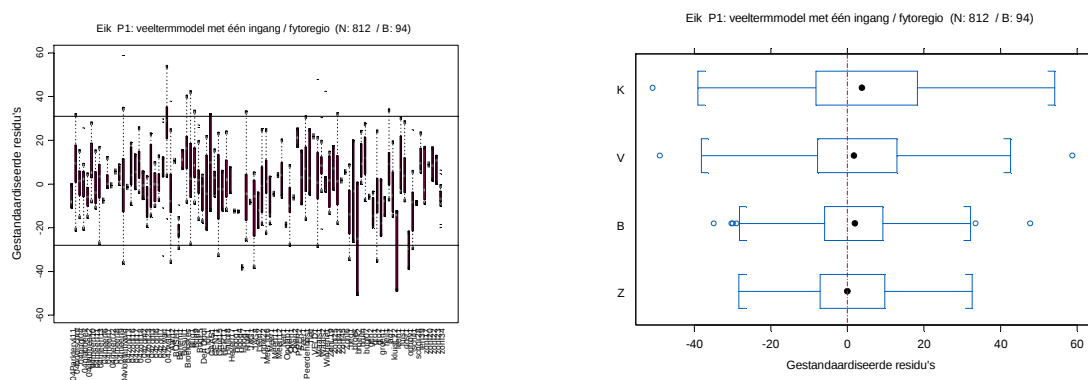
	Terms	Resid. Df	RSS	Df	Sum of Sq	F Value	Pr(F)
1	C + C2 + C3	808	43.0				
2	(C + C2 + C3) * Ecoregio	796	31.9	12	11.1	23.2	0

Residuanalyse (Figuur 8-10, Figuur 8-11, Figuur 8-12 en Figuur 8-13).

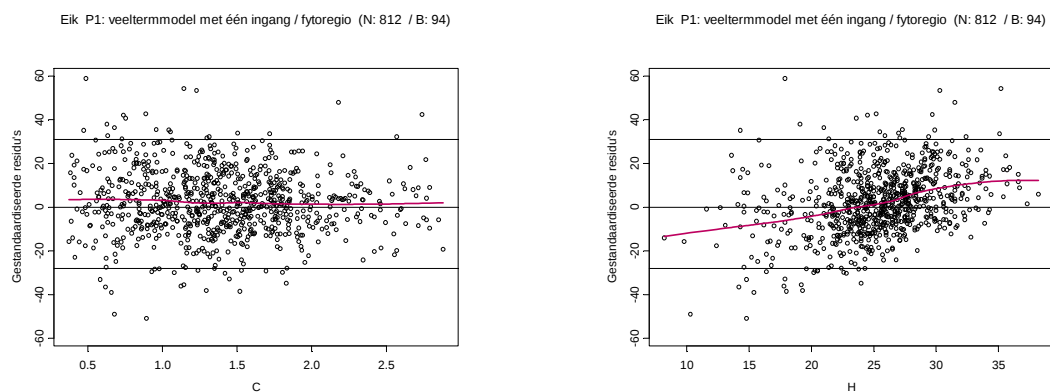
De QQ-plots wijken af aan de staarten. Het verband met ecoregio is afgevlakt, maar de gestandaardiseerde residu's blijven gecorreleerd met de hoogte. Dat is logisch. Door ecoregio in te brengen in het model, compenseren we voor de gemiddelde hoogte in een district, maar het model houdt uiteraard geen rekening met de hoogteverschillen binnen een regio. Opvallend is dat de breedte in de boxplots afhangt van ecoregio. De onzekerheid in het Vlaams en Kempens district is groter. Dat blijkt ook uit Figuur 8-13 waar we de residu-plots in functie van omtrek uitsplitsen per ecoregio. In het Zoniënwoud en het Brabants district, is de grootteorde van de fout 25 %, in het Vlaamse district ongeveer 30 % en in het Kempens district meer dan 35 %.



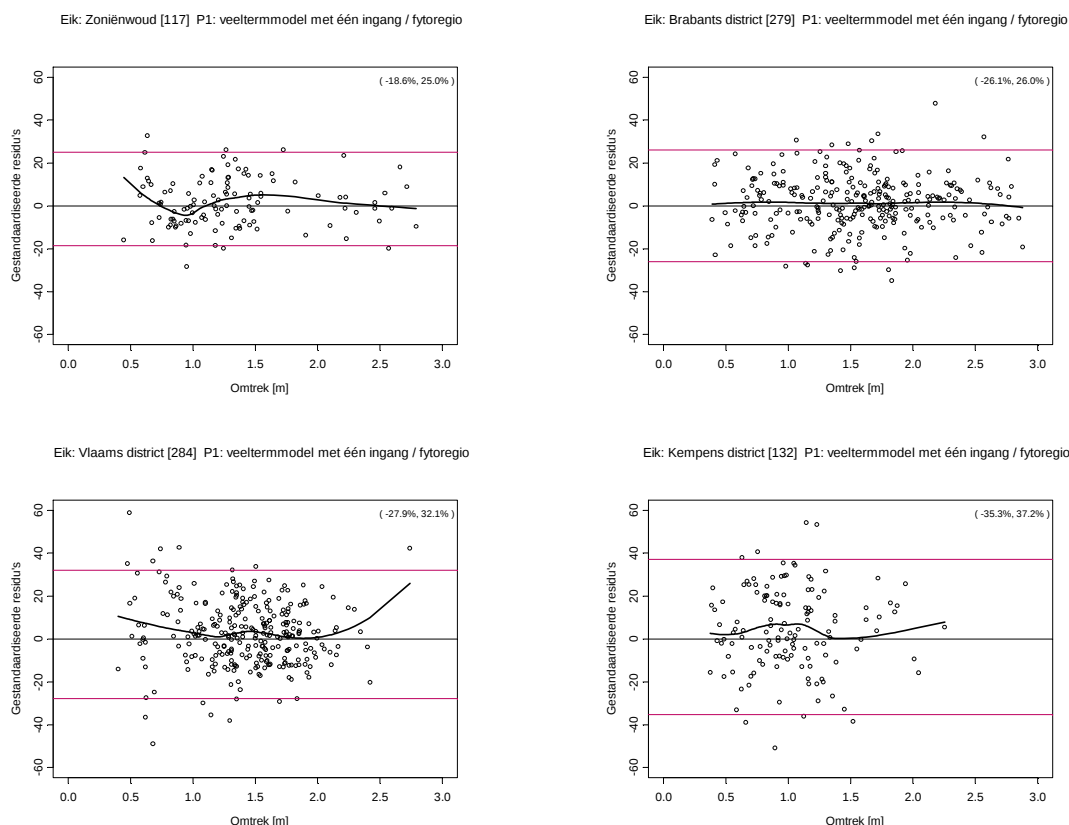
Figuur 8-10 Eik: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). QQ-plots.



Figuur 8-11 Eik: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). Boxplots



Figuur 8-12 Eik: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). Relatie met omtrek en hoogte.



Figuur 8-13 Eik: residuanalyse van het veeltermmodel met één ingang per ecoregio (P1f) in functie van omtrek (C).

De coëfficiënten van het model (Tabel 8-8 en Tabel 8-9)

Volgens de referentiewaarden (voor een boom van 1,5 m omtrek en een hoogte van 25 m) zijn de ecoregio's geordend: Zoniën (2.476 m³) > Brabants district (2.056 m³) > Vlaams district (2.182 m³) > Kempens district (1.811 m³). De voornaamste reden is dat de hoogte sterk afhangt van de regio. Uit de verkenning van de gegevens bleek al dat een boom in Zoniën relatief hoger wordt (op voorwaarde dat de steekproef representatief is) en dat vertaalt zich in een hogere schatting van het gemiddeld volume.

Tabel 8-8 Eik: coëfficiënten van het veeltermmodel met één ingang en ecoregio.

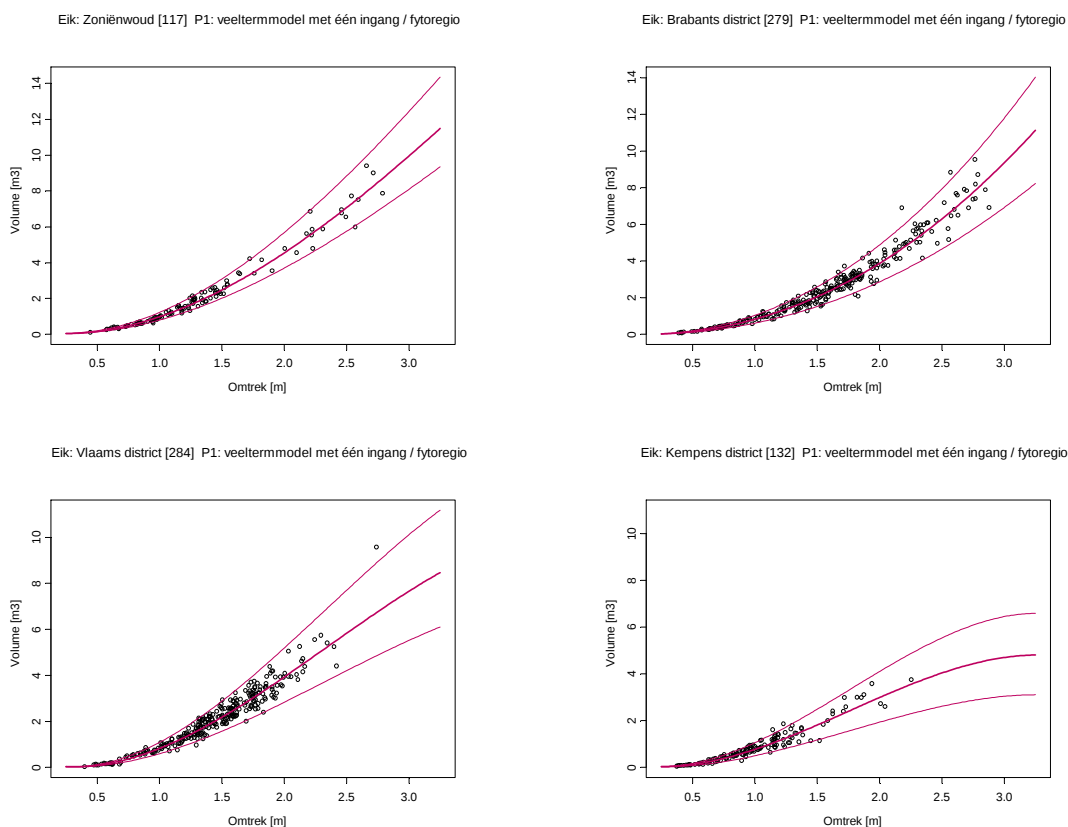
		Value	Std. Error	t value	Pr(> t)
Zoniënwood [N=746]	Referentie:	2.476			
	(Intercept)	0.138	0.0966	1.43	1.55e-001
	C	-0.813	0.3144	-2.59	1.10e-002
	C2	1.800	0.3036	5.93	3.37e-008
	C3	-0.146	0.0858	-1.71	9.07e-002
Brabants district [N=399]	Referentie:	2.056			
	(Intercept)	0.021	0.0457	0.466	6.42e-001
	C	-0.247	0.1637	-1.511	1.32e-001
	C2	1.018	0.1659	6.136	2.94e-009
	C3	0.034	0.0487	0.697	4.86e-001
Vlaams district [N=229]	Referentie:	2.182			
	(Intercept)	0.216	0.0592	3.64	3.22e-004
	C	-1.260	0.2201	-5.72	2.67e-008
	C2	2.181	0.2423	9.00	0.00e+000
	C3	-0.311	0.0799	-3.90	1.22e-004
Kempens district [N=54]	Referentie:	1.811			
	(Intercept)	0.138	0.0805	1.71	0.0898342
	C	-0.858	0.3385	-2.53	0.0124731
	C2	1.838	0.4240	4.34	0.0000291
	C3	-0.349	0.1599	-2.18	0.0311466

Tabel 8-9 Eik: vergelijking van de veeltermmodellen met één ingang en ecoregio.

	District				
	Alle	Zoniën	Brabants	Vlaams	Kempens
Referentie:	2.127	2.476	2.056	2.182	1.811
Intercept					
C	-0.663	-0.813	-0.247	-1.260	-0.858
C2	1.496	1.800	1.018	2.181	1.838
C3	-0.100	-0.146	0.034	-0.311	-0.349
95 % interval		-18.6% , 25.0%	-26.1% , 26.0%	-27.9% , 32.1%	-35.3% , 37.2%

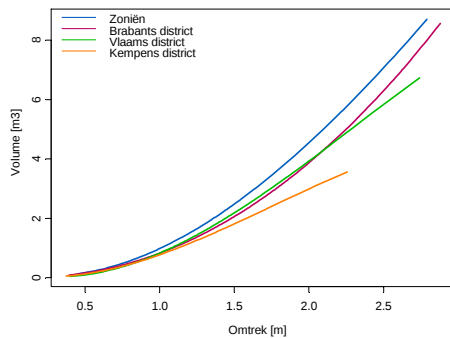
Voorstelling van het model (Figuur 8-14 en Figuur 8-15).

Figuur 8-14 geeft per ecoregio het resultaat. Overall is de overeenstemming met de gegevens heel goed. Uit Figuur 8-15 blijkt dat de curve in het Zoniënwoud voor het hele omtrekgebied hoger ligt dan alle andere. De curven voor het Vlaams district en het Brabants district liggen heel dicht bij elkaar behalve op het einde. Probleem is wel dat we in het Vlaams district slechts één boom hebben met een diameter groter dan 2.5 m en daarenboven is het een uitbijter. Het volume in het Kempens district ligt hoger bij kleinere bomen. Maar in absolute waarde betekent dat niet veel. Vanaf 1 m omtrek ligt het volume daar lager dan in de andere ecoregio's.

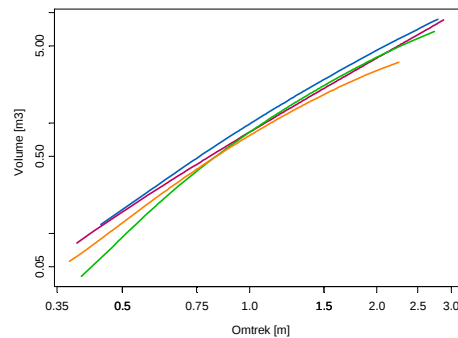


Figuur 8-14 Eik: grafiek van het veeltermmodel met één ingang per ecoregio (P1f). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten onder en boven liggen 95% van de relatieve residu's.

Eik P1: veeltermmodel met één ingang / fyto regio (N: 812 / B: 94)



Eik P1: veeltermmodel met één ingang / fyto regio (N: 812 / B: 94)



Figuur 8-15 Eik: grafische vergelijking van de veeltermmodellen (P1f) in de originele schaal (links) en in de log-schaal (rechts).

8.2 Logaritmische modellen

8.2.1 Logaritmisch model met twee ingangen

Startmodel

Uit de verkennende analyse (o.a. Figuur 7-2) bleek dat een logaritmische transformatie de variatie stabiliseert. Vandaar dat we een lineair regressiemodel zonder weging kunnen uitproberen in de log-schaal. Hierbij proberen we het volume te voorspellen op basis van de log van de omtrek en de hoogte en een eventuele interactie tussen beide termen:

$$L2 \leftrightarrow \log_{10}(V) \approx \log_{10}(C) * \log_{10}(H) \quad (8.7)$$

Interessant is nat te gaan in hoeverre de interactie significant is en een groot effect heeft. Indien niet, dan kunnen we schrijven:

$$L2x \leftrightarrow \log_{10}(V) \approx \log_{10}(C) + \log_{10}(H) \quad (8.8)$$

Hieruit volgt na terugtransformatie (zie Materiaal en Methoden voor een bewijsvoering):

$$V \propto \varphi \frac{C^{\beta_1} H^{\beta_2}}{12\pi} \quad \text{met: } \varphi = 12\pi \cdot 10^{\beta_0} \quad (8.9)$$

Als $\beta_1 \approx 2$ en $\beta_2 \approx 1$ dan krijgen we ongeveer de formule van een kegel op een vormfactor φ na.

Modelselectie (Tabel 8-10)

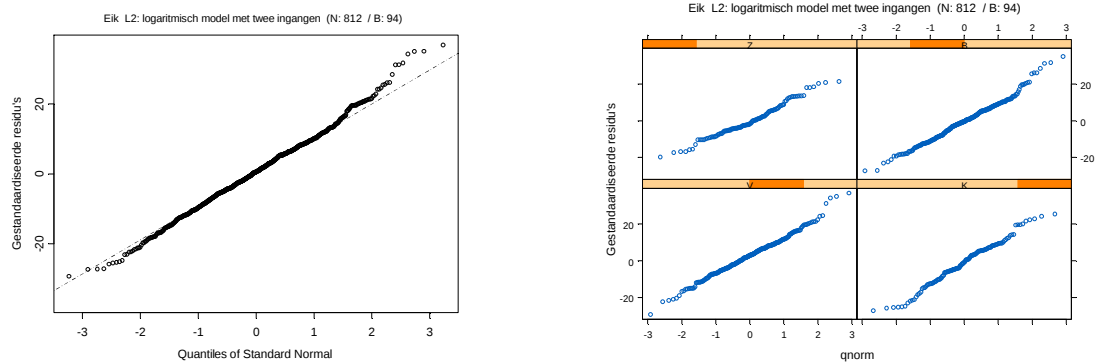
Uit de ANOVA-tabel blijkt dat de interactie (het product) tussen $\log C$ en $\log H$ nog weinig bijdraagt aan het model, maar de term is wel hoogsignificant. Bij "Coëfficiënten" onderzoeken we de grootte van de interactieterm.

Tabel 8-10 Eik: ANOVA-tabel van het logaritmisch model met twee ingangen.

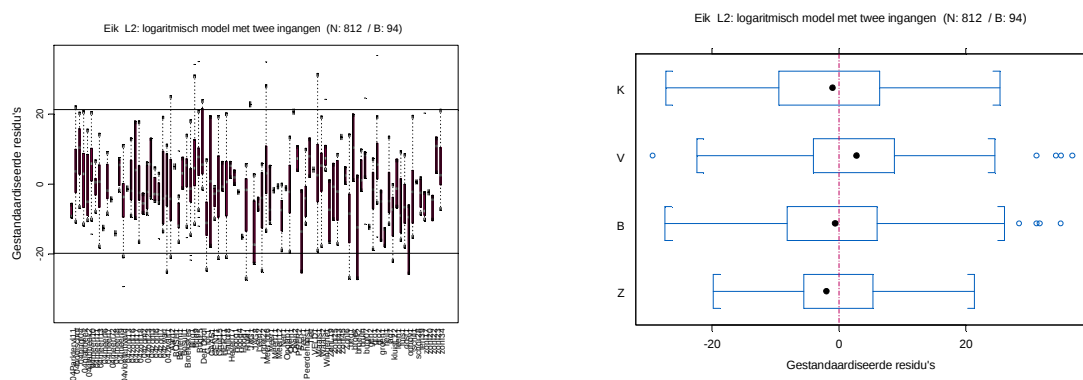
	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
logC	1	144	144	69225	0.00e+000
logH	1	3	3	1629	0.00e+000
logC:logH	1	0	0	27	2.48e-007
Residuals	808	2	0		

Residuanalyse (Figuur 8-16, Figuur 8-17 en Figuur 8-18)

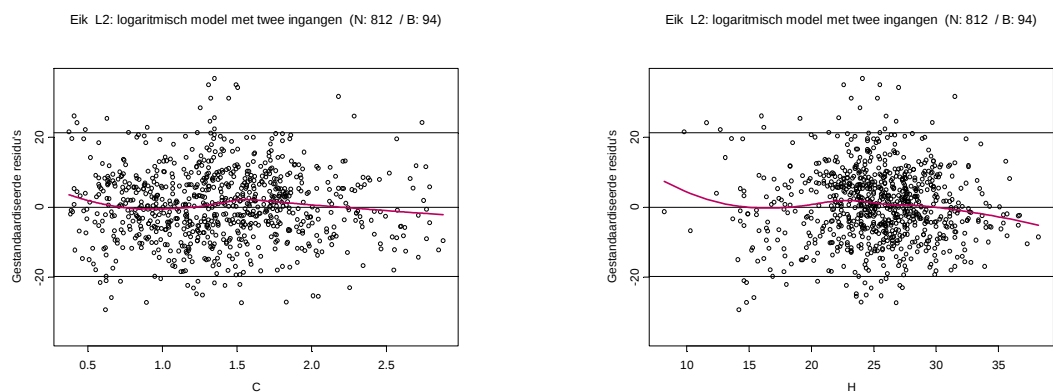
De residu's vertonen geen bijzondere patronen. In het bijzonder is het 95 % interval van de relatieve residu's ± 20 % of de relatieve voorspellingsintervallen liggen opnieuw tussen 0.8 en 1.2 keer de voorspelde waarde.



Figuur 8-16 Eik: residuanalyse voor het logaritmisch model met twee ingangen (L2): QQ-grafieken.



Figuur 8-17 Eik: residuanalyse voor het logaritmisch model met twee ingangen (L2): boxplots.



Figuur 8-18 Eik: residuanalyse voor het logaritmisch model met twee ingangen (L2): omtrek en hoogte.

Coëfficiënten van het model (Tabel 8-11 en Tabel 8-12)

Aangezien de interactieterm hoogsignificant is laten we term in het model. Tabel 8-11 geeft de parameterwaarden voor het finaal model. Voor de referentieboom is het volume net iets meer dan 2 m³. De interactieterm is negatief. Dat betekent dat de coëfficiënten van $\log C$ en $\log H$ dalen naarmate de boom groter wordt.

Tabel 8-11 Eik: coëfficiënten van het logaritmisch model met twee ingangen, met interactie

	Value	Std. Error	t value	Pr(> t)
Referentie:	2.008			
(Intercept)	-1.605	0.0388	-41.4	0.00e+000
logC	2.570	0.1158	22.2	0.00e+000
logH	1.117	0.0284	39.3	0.00e+000
logC:logH	-0.431	0.0828	-5.2	2.48e-007

Als we de interactieterm weglaten, bekomen een model die heel dicht de basisvorm van een kegel benadert:

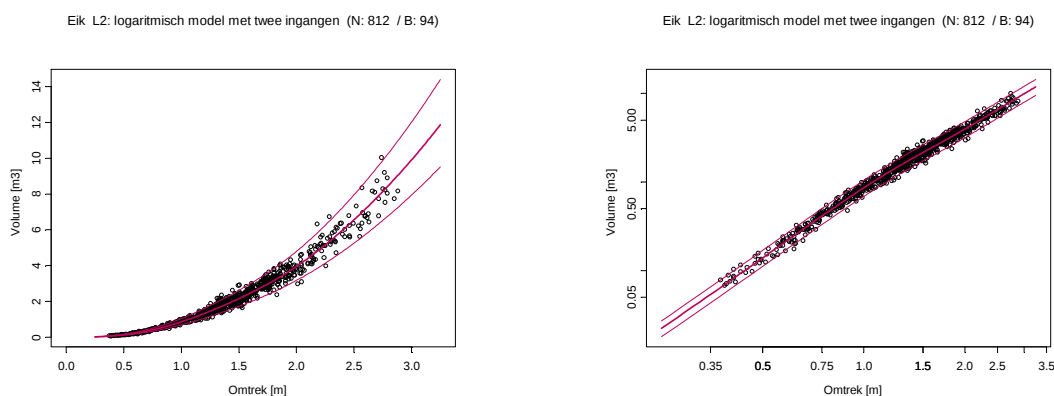
$$V \propto \frac{C^{1.97} H^{1.14}}{12\pi} \quad (8.10)$$

Tabel 8-12 Eik: coëfficiënten van het logaritmisch model met twee ingangen, zonder interactie

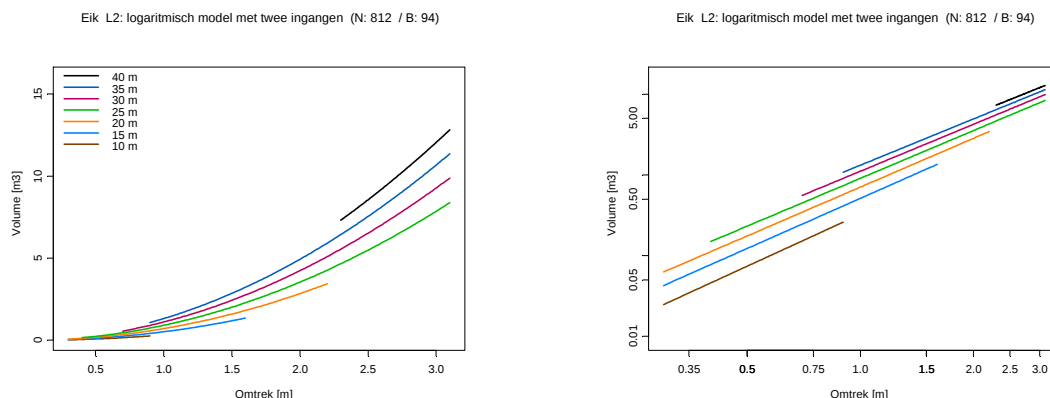
	Value	Std. Error	t value	Pr(> t)
(Intercept)	-1.64	0.0389	-42.1	0
logC	1.97	0.0136	144.8	0
logH	1.14	0.0286	39.7	0

Voorstelling van het model (Figuur 8-19).

De waarnemingen blijken goed aan te sluiten bij de voorspelde waarden. We krijgen een anaaloog verband als bij het veeltermmodel met twee ingangen.



Figuur 8-19 Eik: grafiek van het logaritmisch model met twee ingangen (L2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinen 95% van de relatieve residu's liggen.



Figuur 8-20 Eik: grafiek van het logaritmisch model met twee ingangen (L2) per hoogteklasse (rechts in de log-schaal).

8.2.2 Logaritmisch model met één ingang

Startmodel

Het logaritmisch model met één ingang bekomen we door de hoogterm uit het model (4.6) weg te laten.

$$L1 \leftrightarrow \log_{10}(V) \square \log_{10}(C) \quad (8.11)$$

Modelselectie (Tabel 8-13)

Uit de verkennende analyse bleek dat dit model kans heeft op slagen (o.a. Figuur 7-2). Zoals verwacht is de enige term in het model hoog significant (Tabel 8-13).

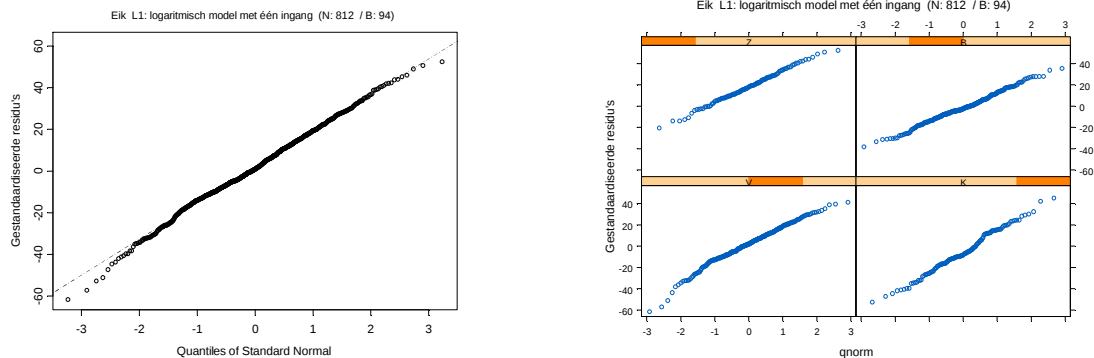
Tabel 8-13 Eik: ANOVA-tabel van het logaritmisch model met één ingang.

	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
logC	1	144	144	22754	0
Residuals	810	5	0		

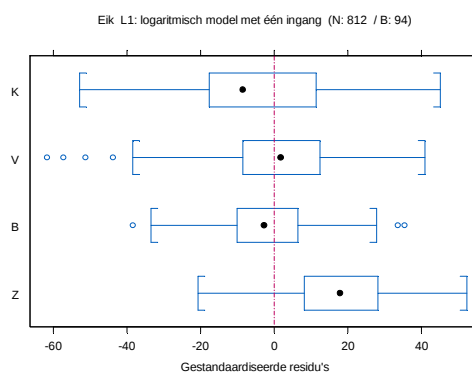
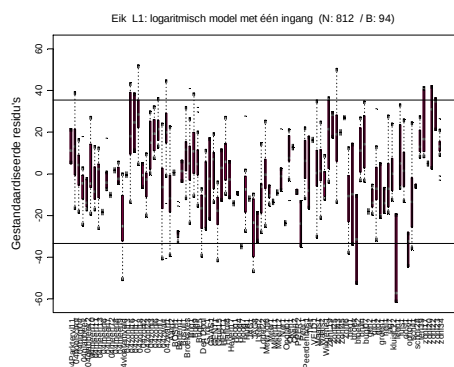
Residuanalyse (Figuur 8-21, Figuur 8-22 en Figuur 8-23)

De residuplots heeft een patroon ten opzichte van de omtrek. Voor hoge en lage waarden van de omtrek leidt het model gemiddeld tot een overschatting, voor tussenliggende waarden tot een onderschatting. Een complexer verband met de omtrek is nodig. Zoals verwacht (omdat hoogte niet in het model opgenomen is) is er een duidelijke relatie van de residu's met hoogte.

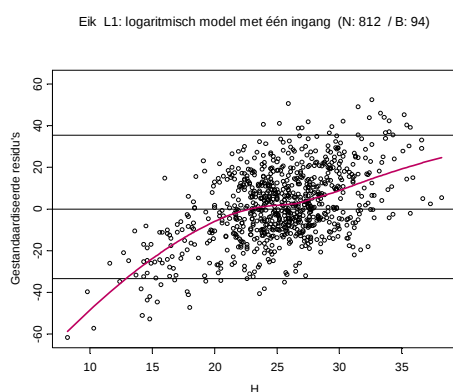
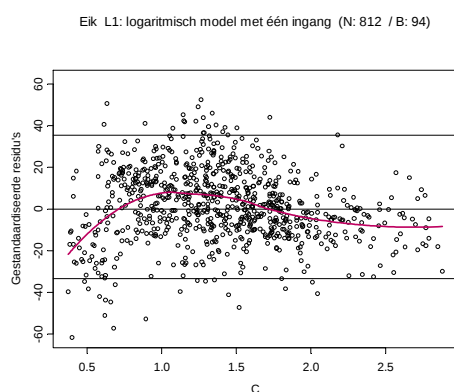
Omdat hoogte gecorreleerd is met de ecoregio, is er ook een relatie tussen ecoregio en de gestandaardiseerde residu's. In het Zoniënwoud is er gemiddeld een onderschatting, in het Brabantse district en in het Vlaamse district is de schatting gemiddeld. In het Kempen district is er een overschatting, maar aangezien de boxplot breder zijn dan de andere, blijkt ook het model iets minder goed te zijn.



Figuur 8-21 Eik: residuanalyse voor het logaritmisch model met één ingang (L1). Q-Q-plots.



Figuur 8-22 Eik: residuanalyse voor het logaritisch model met één ingang (L1). Boxplots.



Figuur 8-23 Eik: residuanalyse voor het logaritisch model met één ingang (L1). Relatie met omtrek en hoogte.

Coëfficiënten van het model (Tabel 8-14).

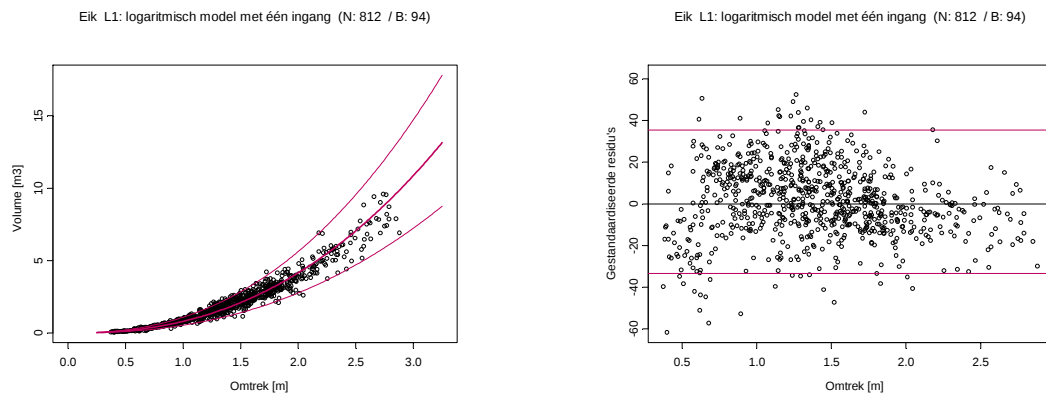
Het model is onvoldoende accuraat. Ter informatie geven we toch de tabel met de geschatte coëfficiënten.

Tabel 8-14 Eik: coëfficiënten van het logaritisch model met één ingang.

	Value	Std. Error	t value	Pr(> t)
Referentie:	2.10	0.00297	108	0
(Intercept)	-0.0952	0.00329	-28.9	0
logC	2.372	0.01572	150.8	0

Voorstelling van het model (Figuur 8-24).

In de figuur zien we hoe inderdaad voor grote en kleine (zie log-schaal) omtrekken, het model systematisch tot een te hoge schatting van het volume leidt. In het midden is er eerder een onderschatting (ook het beste zichtbaar in de log-schaal).



Figuur 8-24 Eik: grafiek van het logaritmisch model met één ingang (L1). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten liggen 95% van de relatieve residu's.

8.2.3 Logaritmisch model met één ingang en ecoregio

Startmodel

Het logaritmisch model met één ingang kunnen we opsplitsen per ecoregio. Op die manier krijgen we inzicht hoe het model afhangt van ecoregio:

$$L1f \leftrightarrow \log_{10}(V) \square \log_{10}(C) * Fytoregio \quad (8.12)$$

Modelselectie (Tabel 8-15)

Toevoeging van ecoregio verbetert het model in belangrijke mate. Zowel ecoregio en de interactie tussen ecoregio en de (log van de) omtrek zijn hoog significant (Tabel 8-15). Het is dus aan te bevelen een aparte regressie uit te voeren per ecoregio:

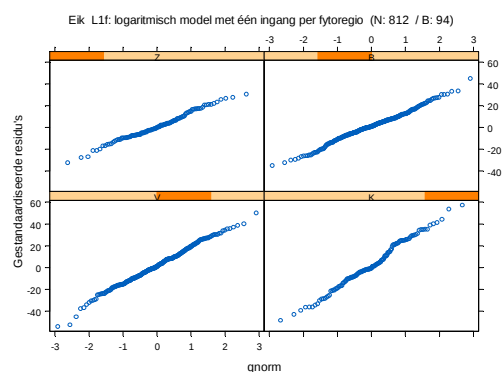
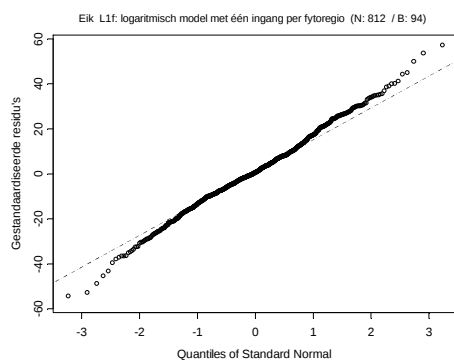
$$L1f \leftrightarrow \log_{10}(V) \square \log_{10}(C) \mid Fytoregio \quad (8.13)$$

Tabel 8-15 Eik: ANOVA-tabel van het logaritmisch model met één ingang en ecoregio.

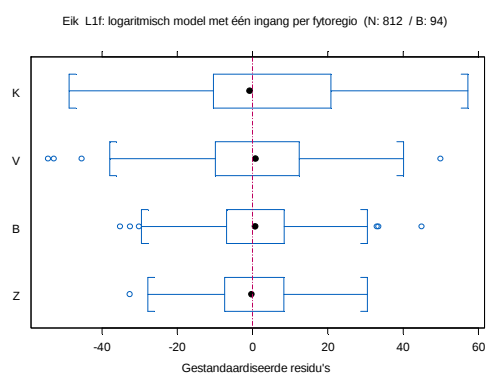
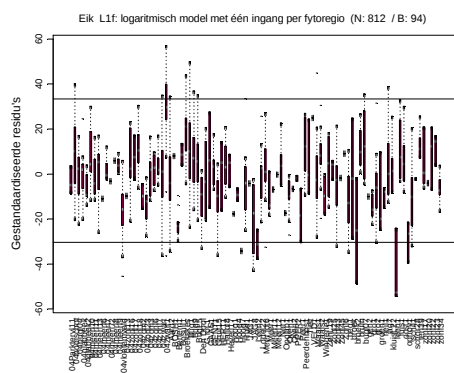
	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
logC	1	144	144	28042	0.00e+000
Ecoregio	3	1	0	49	0.00e+000
logC:Ecoregio	3	0	0	16	6.98e-010
Residuals	804	4	0		

Residuanalyse (Figuur 8-25, Figuur 8-26, Figuur 8-27 en Figuur 8-28)

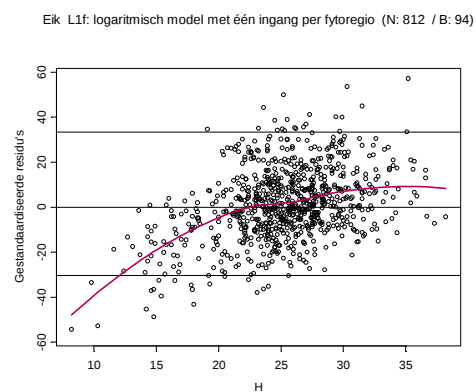
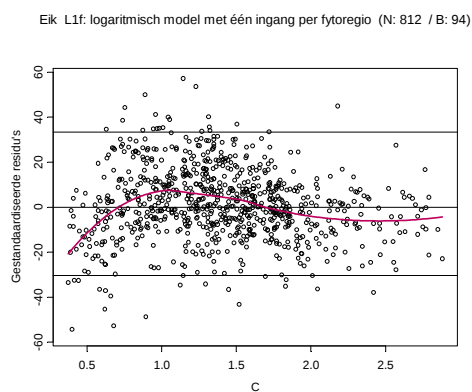
Het patroon ten opzichte van de ecoregio's is verdwenen, behalve dat de nauwkeurigheid blijft verschillen naargelang het district. Vooral in de Kempen en in mindere mate in het Vlaamse district is het model minder precies. Erger is dat het patroon van de gestandaardiseerde residu's in functie van de omtrek blijft. Een apart logaritmisch regressiemodel per ecoregio leidt niet tot een echte verbetering. Alleen in het Brabants district is er een goede overeenstemming met de gegevens. In het Vlaams district en het Kempens district totaal niet.



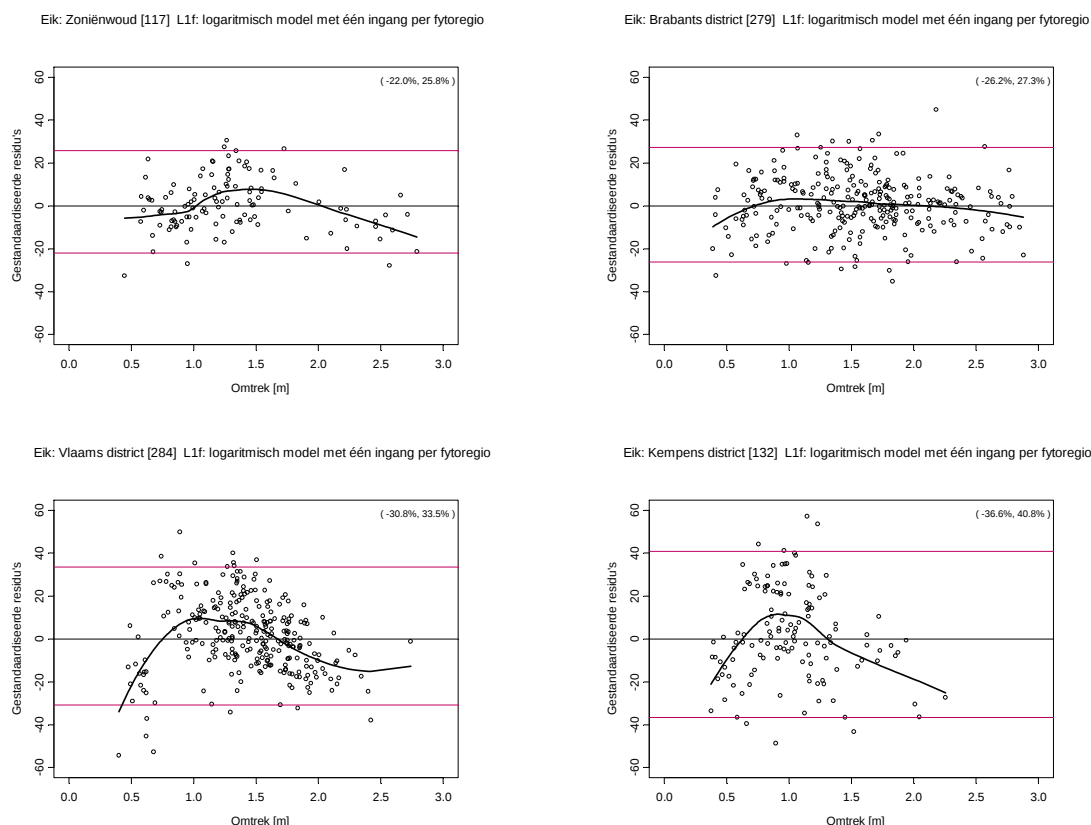
Figuur 8-25 Eik: residuanalyse voor het logaritmisch model met één ingang per district (L1f). Q-Q-plots.



Figuur 8-26 Eik: residuanalyse voor het logaritmisch model met één ingang per district (L1f). Boxplots.



Figuur 8-27 Eik: residuanalyse voor het logaritmisch model met één ingang per district (L1f). Relatie met omtrek en hoogte.



Figuur 8-28 Eik: residuanalyse van het logaritmisch model met één ingang per district (L1f) in functie van omtrek (C).

De coëfficiënten van het model (Tabel 8-16 en Tabel 8-17)

Voor de referentieboom krijgen we opnieuw volgende ordening: Zoniën > Vlaams district > Brabants district en Kempens district. Het is echter weinig zinvol de modelparameters verder te vergelijken, want het logaritmisch model heeft niet de goede vorm.

Tabel 8-16 Eik: coëfficiënten van het logaritmisch model met één ingang en ecoregio.

		Value	Std. Error	t value	Pr(> t)
Zoniënwood [N=746]	Referentie:	2.417			
	(Intercept)	0.956	0.00565	-3.45	0.0008
	logC	2.287	0.02903	78.79	0
Brabants district [N=399]	Referentie:	2.039			
	(Intercept)	0.812	0.00464	-19.5	0
	logC	2.271	0.01916	118.5	0
Vlaams district [N=229]	Referentie:	2.130			
	(Intercept)	0.768	0.00616	-18.6	0
	logC	2.515	0.03112	80.8	0
Kempens district [N=54]	Referentie:	1.949			
	(Intercept)	0.740	0.00836	-15.6	0
	logC	2.388	0.05026	47.5	0

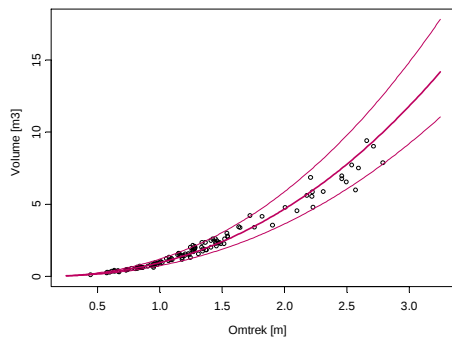
Tabel 8-17 Eik: vergelijking van de coëfficiënten van het logaritmisch model met één ingang en ecoregio.

	District				
	Alle	Zoniën	Brabants	Vlaams	Kempens
Referentie:	2.10	2.417	2.039	2.130	1.949
Intercept	-0.095	0.956	0.812	0.768	0.740
logC	2.372	2.287	2.271	2.515	2.388
95 % interval		-22.0% , 25.3%	-26.2% - 27.3%	-30.8% - 35.5%	-36.6% - 40.8%

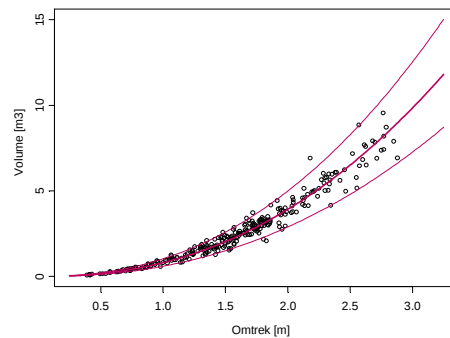
Voorstelling van het model (Figuur 8-29 en Figuur 8-30).

Uit de figuren blijkt dat het logaritmisch model tot een systematische overschatting leidt voor grote omtrekken. Alleen voor het Brabants district is het model in goede overeenstemming met de gegevens.

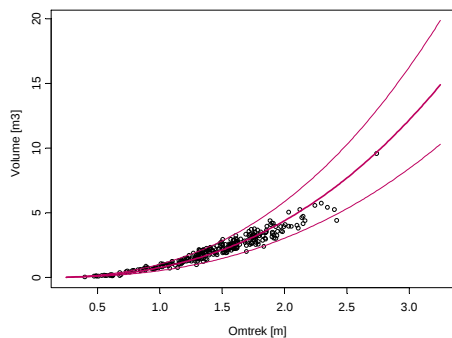
Eik: Zoniënwood [117] L1f: logaritmisch model met één ingang per fyto regio



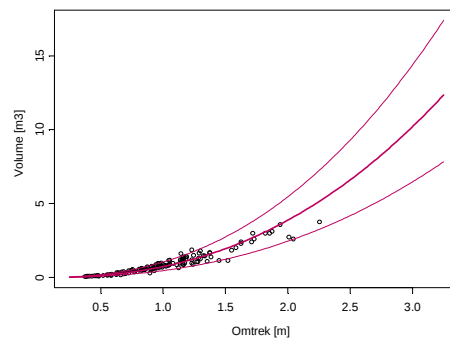
Eik: Brabants district [279] L1f: logaritmisch model met één ingang per fyto regio



Eik: Vlaams district [284] L1f: logaritmisch model met één ingang per fyto regio

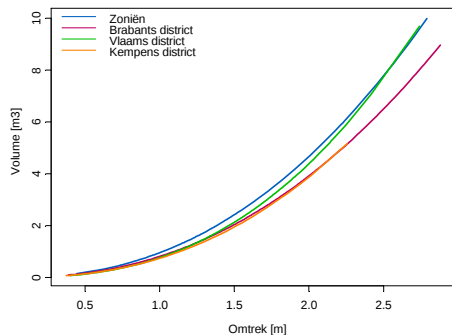


Eik: Kempens district [132] L1f: logaritmisch model met één ingang per fyto regio

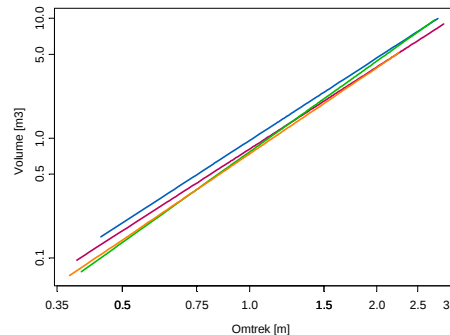


Figuur 8-29 Eik: grafiek van het logaritmisch model met één ingang per ecoregio (L1f). Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten onder en boven liggen 95% van de relatieve residu's.

Eik: L1f: logaritmisch model met één ingang per fyto regio (N: 812 / B: 94)



Eik: L1f: logaritmisch model met één ingang per fyto regio (N: 812 / B: 94)



Figuur 8-30 Eik: vergelijking van de logaritmische modellen (links in de originele schaal rechts en in de log-schaal).

8.3 Relatief model met twee ingangen

Startmodel

Het relatief model modelleert de verhouding van het volume tot het volume van een kegel als een lineaire functie van de omtrek en de hoogte. Op die manier krijgen we inzicht hoe de vormfactor afhangt van de grootte van een boom:

$$R2 \leftrightarrow relV = \frac{V}{V_{kegel}} = C + H \quad (8.14)$$

Modelselectie (Tabel 8-18)

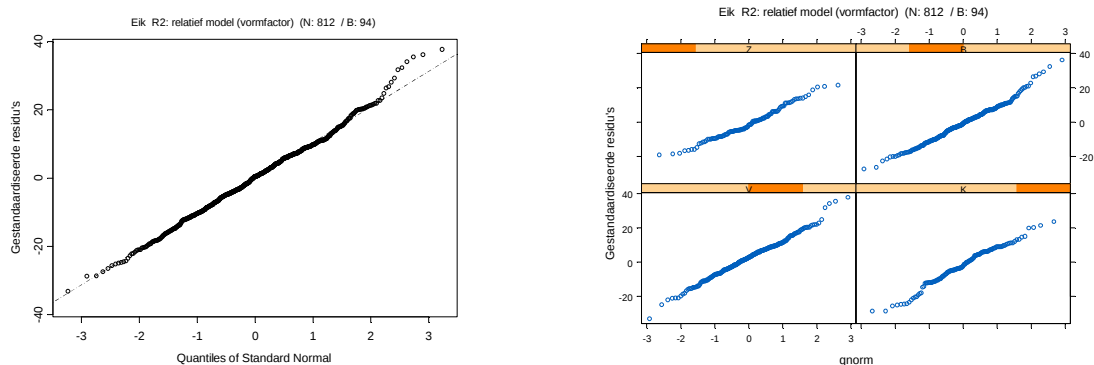
Zowel C als H zijn significant. De verhouding tussen het volume is geen constante, maar hangt af van de grootte van de boom.

Tabel 8-18 Eik: ANOVA-tabel van het relatief model met twee ingangen.

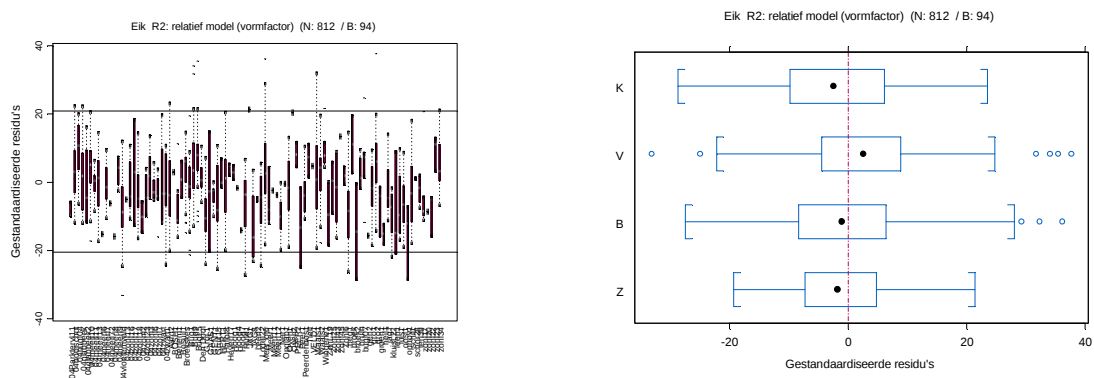
	Df	Sum of Sq	Mean Sq	F Value	Pr(>F)
C	1	1.6	1.59	91.5	0
H	1	1.3	1.33	76.9	0
Residuals	1425	24.8	0.02		

Residuanalyse (Figuur 8-31, Figuur 8-32 en Figuur 8-33)

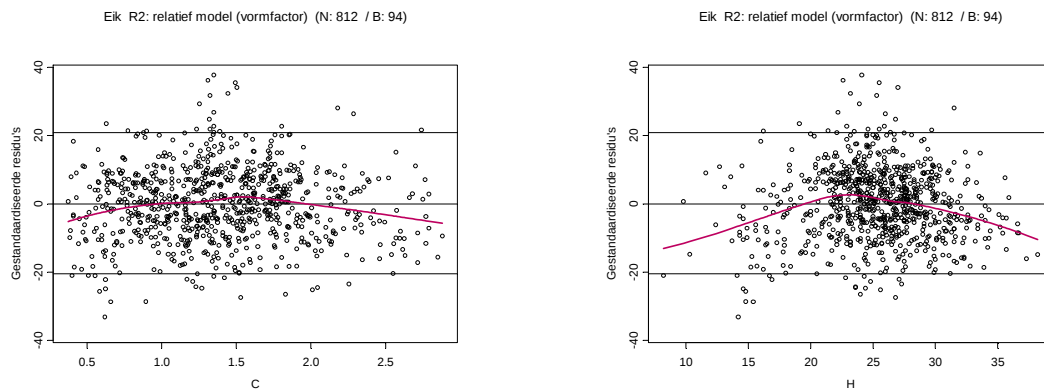
De QQ-plots zijn iets minder goed aan rechterstaart dan de andere modellen met twee ingangen. Er is misschien ook een iets grotere invloed van ecoregio. De plots ten opzichte van omtrek en hoogte zijn niet volledig vlak. Als geheel lijkt het model iets minder geslaagd.



Figuur 8-31 Eik: residuanalyse voor het relatief model met twee ingangen (R2). QQ-grafieken.



Figuur 8-32 Eik: residuanalyse voor het relatief model met twee ingangen (R2): boxplots



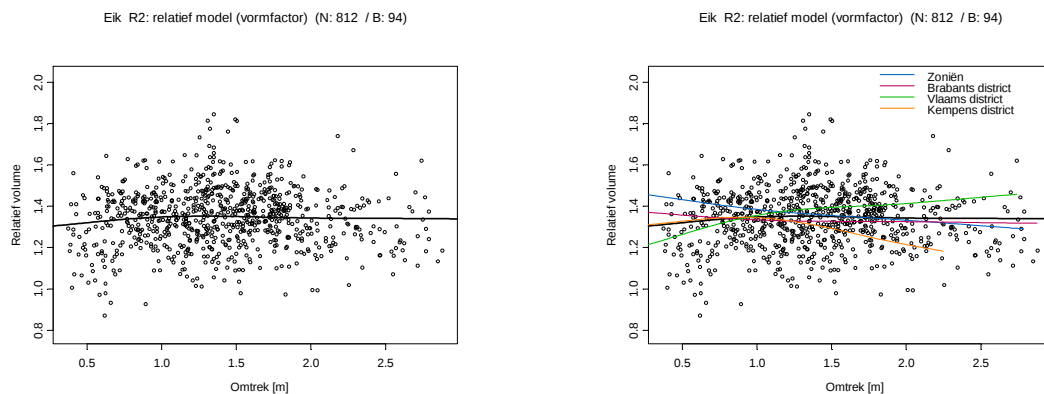
Figuur 8-33 Eik: residuanalyse voor het relatief model met twee ingangen (R2). Relatie met omtrek en hoogte.

De coëfficiënten van het model (Tabel 8-19 en Figuur 4-34)

Uit de verkennende analyse (Figuur 3-6) bleek het relatieve volume in functie van de omtrek nagenoeg constant te zijn. De coëfficiënten (Tabel 8-19) zijn niet alleen heel klein, maar compenseren elkaar: in functie van de omtrek daalt het relatieve volume, maar in functie van de hoogte (die positief gecorreleerd is met omtrek) is er een stijging. Als we het voorspelde relatief volume uitzetten in functie van de omtrek (Figuur 4-34), waarbij we de hoogte gelijkstellen aan de voorspelde hoogte bij de omtrek, dan is er globaal nagenoeg geen trend (linkerfiguur). Opgesplitst per ecoregio (rechterfiguur) zijn er wel trends. Maar alles bij elkaar zijn de verschillen klein. In Zoniën daalt de curve zacht, in het Brabants district is het relatief volume nagenoeg constant, in het Vlaams district is er een stijging en in de Kempen buigt de curve naar beneden vanaf 1.5 m.

Tabel 8-19 Eik: coëfficiënten van het relatief model met twee ingangen.

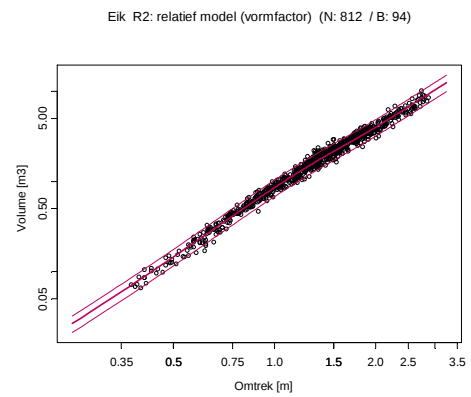
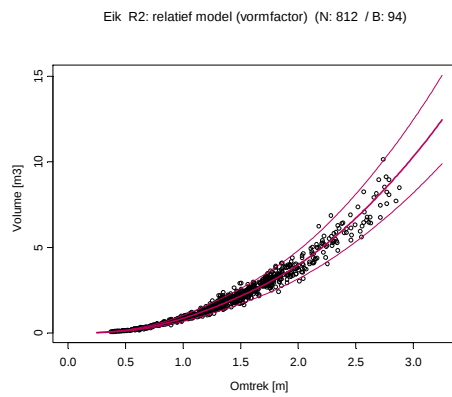
	Value	Std. Error	t value	Pr(> t)
Referentie:	1.99955			
(Intercept)	1.23767	0.02971	41.66	0.0000
C	-0.02896	0.01300	-2.23	0.0261
H	0.00584	0.00151	3.87	0.0001



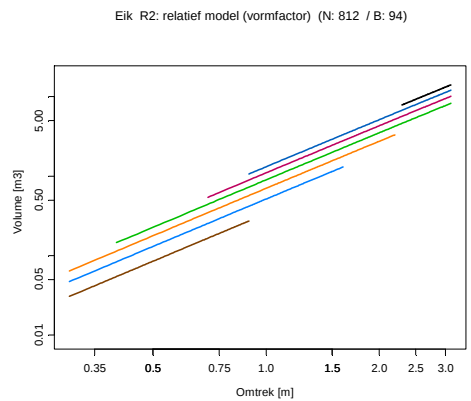
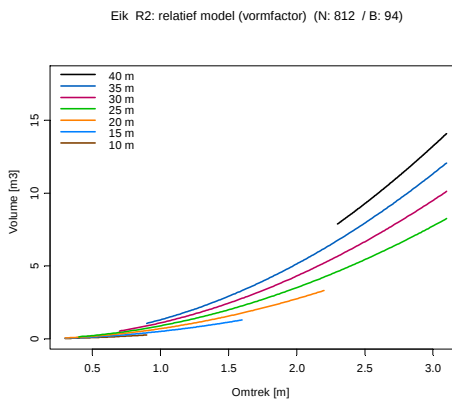
Figuur 8-34 Eik: evolutie van het relatief volume in functie van de omtrek waarbij de hoogte gelijk is aan de voorspelde waarde bij de omtrek (rechts uitgesplitst per ecoregio)

Voorstelling van het model (Figuur 8-35).

Globaal genomen blijkt het model vrij goed aan te sluiten bij de gegevens.



Figuur 8-35 Eik: grafiek van het relatief model met twee ingangen (R2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de relatieve residu's liggen.



Figuur 8-36 Eik: grafiek van het relatief model met twee ingangen (R2) per hoogteklaase (rechts in de log-schaal).

9 Eik: vergelijking van de modellen

9.1 Vergelijking van de voorspellingen van de modellen

9.1.1 De referentieboom

De referentieboom heeft volgende afmetingen: $C = 1.5$ m en $H = 25$ m. Tabel 9-1 geeft een overzicht van het voorspelde houtvolume voor de verschillende modellen. De tabel is een compilatie van Tabel 8-3 en Tabel 8-9 voor de veeltermmodellen, Tabel 8-14 en Tabel 8-17 voor de logaritmische modellen en Tabel 8-19 voor het relatief model.

Bij de tarieven met twee ingangen (T2) geven alle modellen nagenoeg dezelfde waarde: 2 m^3 . Dat deze waarden zo dicht bij elkaar liggen is niet zo verwonderlijk. De referentieboom ligt in het midden van de dataset en de modellen verschillen vooral aan de staarten. Bij de globale tarieven met één ingang (T1) ligt het voorspelde volume iets hoger: 2.1 m^3 . Dit is een klein verschil als we bedenken dat de referentieboom hier gedefinieerd is als de boom met een omtrek van 1.5 m zonder te refereren naar de hoogte.

De tarieven met één ingang maken geen gebruik van de hoogte en de modellen met één ingang uitgesplitst per ecoregio (T1f) zullen impliciet met eventuele verschillen in de hoogte rekening houden. Dat blijkt ook uit de overzichtstabel (Tabel 9-1).

- In het Zoniënwoud ligt het houtvolume een halve kubieke meter hoger bij eenzelfde omtrekwaarde. De reden is dat de hoogte van de bomen daar gemiddeld hoger is (Tabel 7-1).
- Om dezelfde reden liggen het volume in de Kempen lager omdat de bomen gemiddeld lager zijn (Tabel 7-1).
- Het volume in het Vlaamse district ligt (iets) hoger dan in het Brabants district, ondanks het feit de gemiddelde hoogte een halve meter lager ligt (Tabel 7-1). Blijkbaar hangt de vorm van de boom af van ecoregio.

Tabel 9-1 Eik: volumeschatting voor een referentieboom ($C = 2$ m, $H = 30$ m) voor de verschillende tarieven

Type model	T2	T1	T1f (één ingang + ecoregio)			
			Zoniën	Brabants	Vlaams	Kempens
Veelterm (P)	2.03	2.13	2.48	2.06	2.18	1.81
Logaritmisch (L)	2.01	2.10	2.42	2.04	2.13	1.95
Relatief (R)	2.00					

9.1.2 De tarieven met twee ingangen

Figuur 9-1 vergelijkt de modellen over het bereik van de omtrek en de hoogte. Omdat de uitbijters het zicht op het geheel maskeren, zoomt Figuur 5-2 in op het deel waarbinnen 95 % van de verschillen tussen de modellen zich situeert (Tabel 9-2).

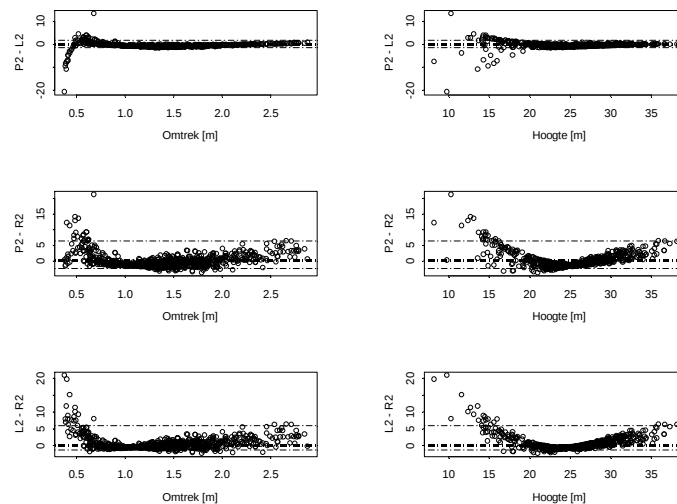
Tabel 9-2 Eik: procentueel verschil tussen modellen voor tarieven met twee ingangen.

Modellen	Percentiel		
	2.5 %	Mediaan	97.5 %
P2 – L2	-1.40 %	-0.62 %	1.72 %
P2 – R2	-2.41 %	-0.69 %	6.47 %
L2 – R2	-1.23 %	-0.08 %	5.88 %

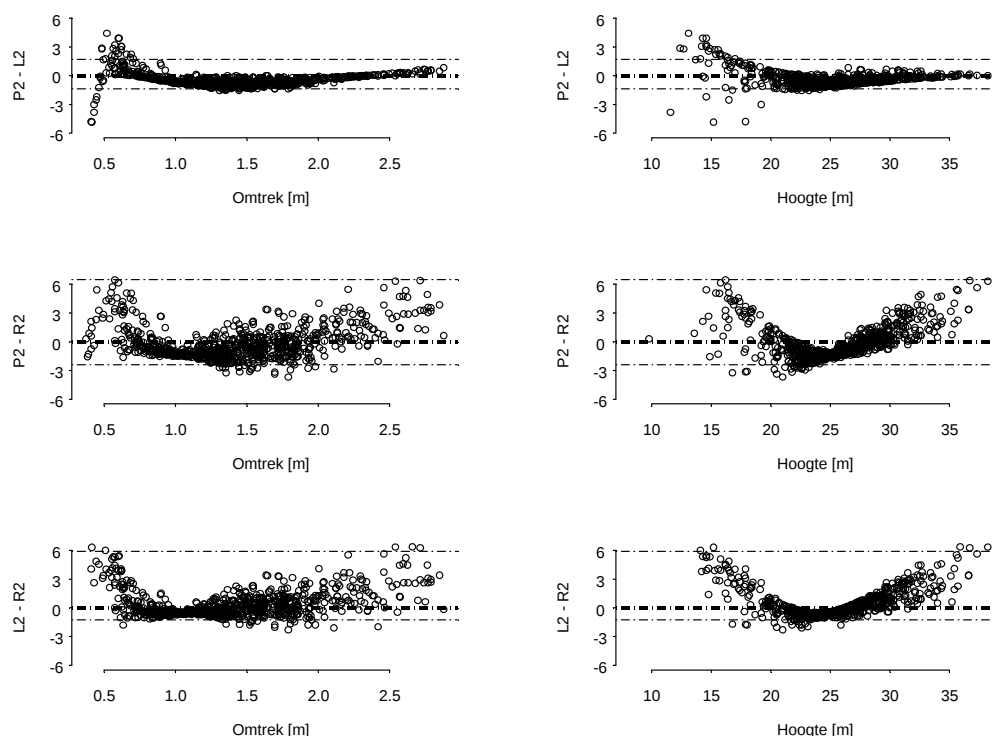
We geven een bespreking per model.

- P2 – L2:
 - o De fout is klein, een 2 % (95 % van de verschillen ligt tussen -1.40 % en 1.72 %). Maar er zijn wel een paar serieuze uitschieters.
 - o Aangezien de mediaan negatief is (-0.62 %) zijn de voorspellingen van het logaritmisch model in meer dan 50 % van de gevallen hoger dan het veeltermmodel. L2 leidt dus in meer gevallen tot een hogere schatting dan P2.
 - o Maar dat verschil is niet uniform. Voor kleine omtrekwaarden ($C < 0.75$ m) of kleine hoogtewaarden ($H < 15$ à 20 m), liggen de voorspellingen van het veeltermmodel hoger. Voor nog kleinere waarden van de omtrek ($C < 0.5$ m) keert het patroon nogmaals om.
- P2 – R2:
 - o Het verschil tussen deze twee modellen is groter en niet symmetrisch: 95 % ervan ligt tussen -2.41 % en 6.47 %.
 - o In het omtrekinterval van 1 m tot 2 m is er weinig verschil (1 à 2 %), maar aan de uiteinden divergeren de modellen steeds verder. Heel markant is het patroon in functie van de hoogte.
 - o De implicatie is dat bij een eventuele extrapolatie de voorspellingen van de modellen ver uit elkaar liggen.
- L2 – R2:
 - o Aangezien L2 en P2 vrij goed met elkaar in overeenstemming zijn (behalve voor kleinere bomen), geeft de vergelijking met L2 vergelijkbare resultaten als voor de vergelijking met P2.
 - o Dus opnieuw is er een asymmetrisch patroon in de verschillen en divergeren de modellen aan de grenzen van het bereik van de hoogte en de omtrek.

Eik N: 812 / B: 94



Figuur 9-1 Eik: vergelijking van de modellen met twee ingangen. P2 = Veeltermmodel, L2 = logaritmisch model en R2 = relatief model. De punt-streep lijn in het vet is de nullijn. De punt-streep lijn onder en boven omvatten 95 % van de punten.



Figuur 9-2 Eik: vergelijking van de modellen met twee ingangen. P2 = Veeltermmodel, L2 = logaritmisch model en R2 = relatief model. De punt-streep lijn in het vet is de nullijn. De punt-streep lijn onder en boven omvatten 95 % van de punten. De figuur is dezelfde als Figuur 9-1, maar de uitbijters zijn weggelaten om meer details te zien.

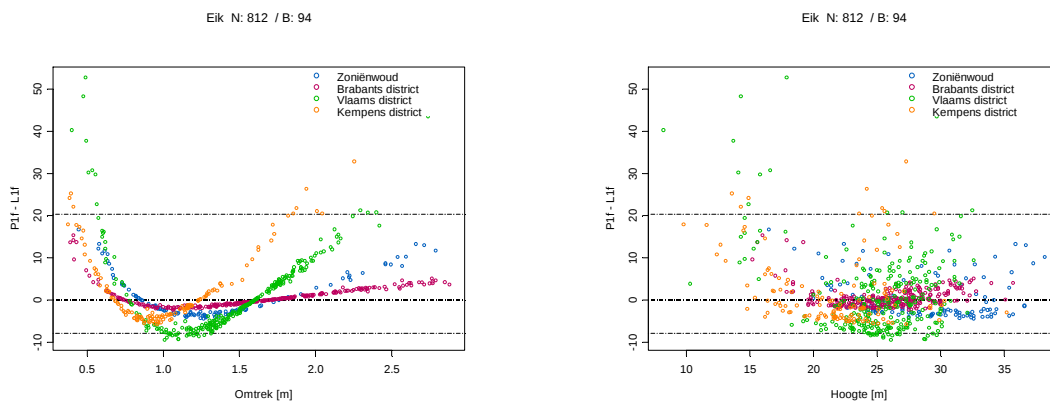
9.1.3 De tarieven met één ingang per ecoregio

Figuur 9-3 vergelijkt de modellen met één ingang aangevuld met ecoregio (P1f en L1f). Tabel 9-3 geeft enkele kengetallen. Uit de residuanalyse weten we dat het veeltermmodel (P1f) beter aansluit bij de gegevens. We onderzoeken hier hoe de twee modellen zich ten opzichte van elkaar situeren.

- Het algemene patroon is telkens gelijk. Voor de “gemiddelde” boom, zijn de voorspellingen van het logaritmisch model hoger, aan de uiteinden is het net andersom.
- Voor het Brabants district is er een goede overeenstemming van de modellen. Alleen voor $C < 0.5$ m is er een groot procentuele verschil. Dat komt overeen met de residuanalyse, waarbij we vaststelden dat daar zowel het logaritmische model als het veeltermmodel goed bij de gegevens aansloten.
- De resultaten voor het Zoniënwoud zijn vergelijkbaar met het Brabants district, maar de verschillen zijn wel groter.
- Voor het Kempens en Vlaams district, is de overeenstemming niet goed. Voor de gemiddelde boom leidt het logaritmisch model tot 10 % grotere schattingen, voor de kleine en grote bomen, leidt het veeltermmodel tot grotere waarden. Verschillen tot meer dan 20 % komen voor.

Tabel 9-3 Eik: verschil tussen de twee modellen voor tarieven met één ingang in functie van de ecoregio (P1f – L1f).

Ecoregio	Percentiel		
	2.5 %	Mediaan	97.5 %
Zoniënwoud	-4.06	-2.06	13.0
Brabants district	-1.93	-0.04	5.1
Vlaams district	-8.44	-1.14	29.2
Kempens district	-5.43	-1.39	23.6

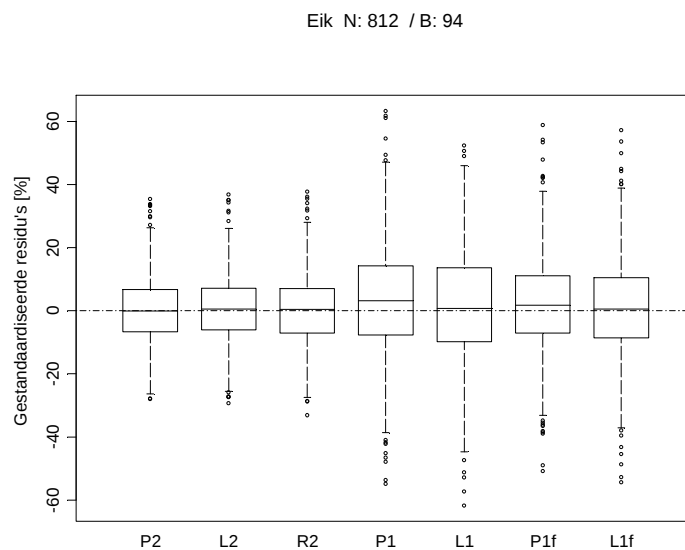


Figuur 9-3 Eik: vergelijking van de modellen met één ingang per ecoregio in functie van de omtrek en de hoogte.

9.2 Vertekening en variabiliteit van de modellen.

9.2.1 Distributie van de gestandaardiseerde residu's

Figuur 9-4 geeft de boxplots van de gestandaardiseerde residu's. De modellen met twee ingangen (P2, L2 en R2) hebben de meest compacte distributie. De boxplots ervan verschillen weinig. De medianen van L2 en R2 liggen iets boven nul, wat wijst op een (kleine) positieve vertekening. Bij de modellen met één ingang worden de distributies compacter en minder vertekend door toevoeging van ecoregio.



Figuur 9-4 Eik: boxplots van de gestandaardiseerde residu's (%)

9.2.2 Juistheid en precisie van de voorspellingen

Als maat voor de juistheid nemen we de mediaan van gestandaardiseerde residu's (Tabel 9-4). Deze informatie kunnen we grafisch afleiden uit de boxplots (Figuur 9-4), maar voor de modellen met één ingang en ecoregio splitst Tabel 9-4 de cijfers op per ecoregio. In het ideale geval verwachten we een onvertekende symmetrische distributie en bijgevolg een mediaan gelijk aan nul.

Voor de tarieven met twee ingangen heeft het veeltermmodel (P2) geen vertekening. Bij het logaritmisches (L2) en relatief model (R2) is er een kleine positieve vertekening. Voor de tarieven met één ingang blijft het logaritmisches model (L1) vrij goed scoren, maar bij het veeltermmodel (P1) neemt de vertekening sterk toe.

Door het model met één ingang aan te vullen met ecoregio, is er weinig verbetering voor het logaritmisches model (L1f) dat al goed scoorde, maar het veeltermmodel (P1f) verbetert wel, behalve in de Kempen.

Tabel 9-4 Eik: juistheid van de modellen met als maat de mediaan van de gestandaardiseerde residu's.

Ecoregio	P2	P1	P1f	L2	L1	L1f	R2
Zoniënwood	-1.7	17.5	0.04	-2.0	17.9	-0.3	-1.8
Brabants district	-1.1	1.2	2.0	-0.6	-2.7	0.7	-1.1
Vlaams district	2.1	2.8	1.8	2.8	1.8	0.8	2.6
Kempens district	-1.4	-4.9	3.8	-1.0	-8.5	-0.7	-2.5

Tabel 9-5 kwantificeert de precisie van de voorspellingen met als maat het 95 % voorspellingsinterval. Deze maat is procentueel. Ze geeft aan hoe groot we de afwijking verwachten procentueel ten opzichte van de voorspelde waarde. Dat is conform het statistisch model waarbij we vaststelden dat de standaardafwijking van de waarnemingen recht evenredig is met de gemiddelde waarde.

Voor tarieven met twee ingangen is de voorspellingsfout van het veeltermmodel (P2) het kleinst (20 %), kort gevolgd door het relatief (R2) en het log-model (L2) met ongeveer 21 %.

Door slechts met één ingang te werken verslechtert de voorspellingsfout tot 35 %. Door ecoregio toe te voegen, winnen we tot 10 % op de voorspellingsfout. In het bijzonder krijgen we een heel goede voorspelling voor Zoniën en Brabants district, waar de fout daalt tot ongeveer 25 %. Dat is maar 5 % slechter dan het model met twee ingangen. In het Vlaams district is de voorspellingsfout 30 % en in de Kempen ligt de fout boven de 35 %. Telkens scoort het veeltermmodel iets beter (ongeveer een tweetal procent winst).

Tabel 9-5 Eik: precisie van de modellen met als maat het 95 % relatieve voorspellingsinterval

Ecoregio	P2	P1	P1f	L2	L1	L1f	R2
Zoniënwood	18.0	42.5	23.1	18.0	42.5	25.9	18.4
Brabants district	19.7	28.9	26.1	20.4	27.6	26.8	20.3
Vlaams district	19.8	29.8	29.8	20.6	32.5	32.1	21.1
Kempens district	24.8	37.3	36.9	24.5	40.9	39.7	29.9

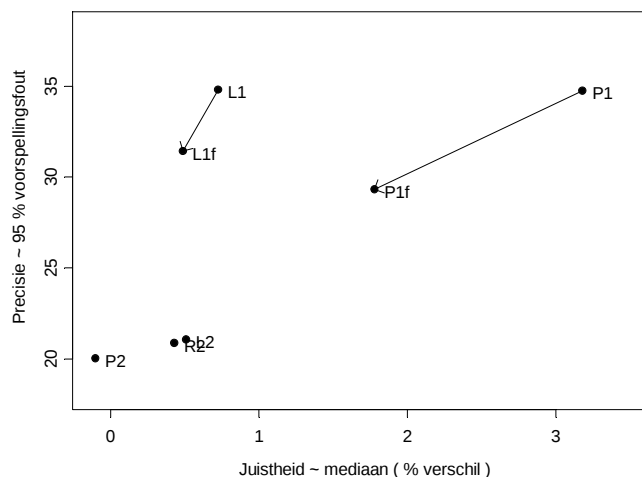
9.2.3 Grafische voorstelling van de accuraatheid

De precisie en de juistheid definiëren samen de accuraatheid van het model. Figuur 9-5 geeft hiervan een grafische voorstelling door beide kwaliteitsmaten te combineren in een grafiek. In de X-as staat de juistheid van de voorspellingen, in de Y-as de precisie ervan. Hoe dichter de punten bij de oorsprong liggen, des te hoger de kwaliteit van de voorspelling.

De drie modellen met twee ingangen hebben een heel vergelijkbare kwaliteit en zijn duidelijk beter dan de modellen met één ingang. De voorspellingsfout van P2, R2 en L2 is nagenoeg gelijk, maar er is iets minder vertekening bij P2.

Voor tarieven met één ingang, zien we mooi hoe de inbreng van ecoregio, zowel de vertekening als de voorspellingsfout vermindert. Het logaritmisch model (L1f) heeft globaal minder vertekening, het veeltermmodel (P1f) heeft een kleinere voorspellingsfout.

Eik N: 812 / B: 94

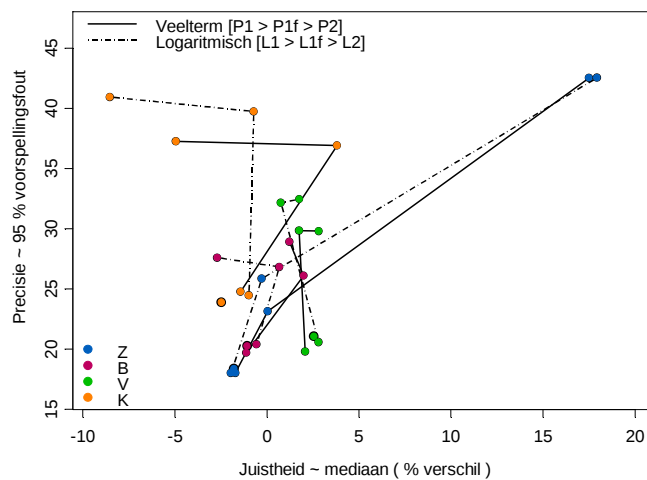


Figuur 9-5 Eik: vertekening en 95 % voorspellingsfout (procentueel).

Het globale beeld kunnen we verder verfijnen door op te splitsen per ecoregio. Figuur 9-6 toont per ecoregio hoe de kwaliteit van de modeltypes (veelterm, logaritmisch of relatief) evolueert door meer termen in rekening te brengen. Omdat het relatief type maar één variant heeft, is dat punt niet verbonden door lijnen. Ook hebben we het punt iets meer benadrukt door de rand dikker te maken. Globaal komt uit de figuur naar voor dat door introductie van ecoregio in het model de vertekening nagenoeg verdwijnt, maar dat hoogte nodig is om de 95 % voorspellingsmarge te verminderen tot 20 %. Ook blijkt het veeltermmodel een iets betere kwaliteit te hebben. Maar verschillen voor het volledig model (met twee ingangen) zijn echt heel klein.

Wel zijn er markante verschillen tussen de ecoregio's. Daarom geven we per ecoregio apart een bespreking.

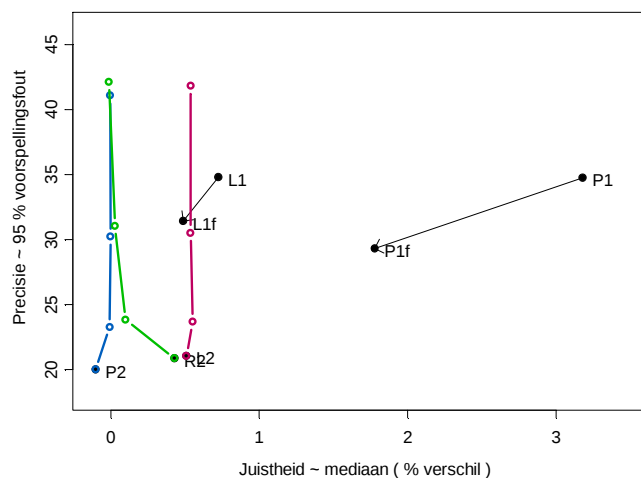
- In het Kempens district starten we met een voorspellingsfout van ongeveer 40 %. Opsplitsing per ecoregio verkleint wel de vertekening, maar niet de voorspellingsfout. Pas wanneer we hoogte toevoegen aan het model, daalt de voorspellingsfout tot 25 %. Deze waarde is de minst goede van alle districten.
- Voor Zoniën is zowel de vertekening (meer dan 15 %) als de voorspellingsfout (meer dan 40 %) heel groot als we met het model met één tarief werken. Het model verbetert sterk als we een apart model met één ingang ontwikkelen voor Zoniën. De vertekening verdwijnt nagenoeg en de voorspellingsfout zakt van 40 % naar ongeveer 25 % (nog iets lager voor het veeltermmodel). Door hoogte in het model op te nemen, zakt de schattingsfout onder de 20 %. Wel is er een beetje vertekening. Alle modellen met twee ingangen scoren nagenoeg even goed.
- In het Brabants district zijn de modellen met één ingang meteen vrij goed. Er is bijna geen vertekening en de voorspellingsfout ligt tussen de 25 à 30 %. Introductie van de hoogte verkleint de voorspellingsfout tot 20 %, iets hoger dan in het Zoniën. De kwaliteit van de drie types met twee ingangen verschilt nauwelijks.
- In Vlaams district is de evolutie van de kwaliteit vrij analoog als in het Brabants district. De voorspellingsfout is wel iets groter (tussen de 30 à 35 % om te starten). Introductie van de hoogte doet de fout dalen tot ongeveer 20 %. Alle modellen hebben dan een vergelijkbare kwaliteit.



Figuur 9-6 Eik: vertekening en 95 % voorspellingsfout (procentueel) uitgesplitst per ecoregio. De punten voor het relatief model (waarvan er maar één variant van is) zijn iets sterker omcirkeld en niet verbonden door lijnen.

9.3 De invloed van de meetfout op de hoogte

Het is niet eenvoudig en tijdrovend de hoogte van een boom te bepalen. De meetfout is soms groot omdat het zicht op de top van de boom gehinderd wordt door omringende bomen en/of eigen takken van de kroon. Om het effect van de meetfout op de kwaliteit van de tarieven in te schatten, voerden we een simulatie uit waarbij we meer en meer ruis toevoegen aan de hoogte. Figuur 9-7 geeft een grafische voorstelling hiervan. Hieruit blijkt dat de juistheid van de voorspellingen nauwelijks beïnvloed wordt door de meetfout op de hoogte (alleen bij R2 lijkt de vertekening te verdwijnen). Maar de voorspellingsfout neemt gestaag toe. Van zodra de relatieve meetfout gemiddeld 8 % is, wordt de voorspellingsfout van het model groter dan de tarieven met één ingang. Daarom is het misschien efficiënter om te werken met de modellen met één ingang uitgesplitst per ecoregio, als we de hoogte onvoldoende nauwkeurig kunnen bepalen.



Figuur 9-7 Eik: de invloed van gemiddelde meetfout op de hoogte (0 %, 4 %, 8 % en 12 %) op de vertekening van de het voorspelde volume en de voorspellingsfout.

9.4 Algemeen besluit

9.4.1 De tarieven met twee ingangen

Globaal genomen ziet de voorspelling het veeltermmodel (P2) er iets beter uit dan de twee andere modellen (Figuur 9-6). Er is nagenoeg geen vertekening en de 95 % voorspellingsmarge is 20 % behalve in het Kempens district (25 %). Ook blijkt uit de residuanalyse dat het veeltermmodel homogeen aansluit bij de gegevens over het volledige meetbereik van de omtrek en de hoogte.

Maar de keuze tussen L2 en P2 is niet kritiek. Alle modellen sluiten goed bij elkaar aan, behalve aan de randen. Maar per definitie hebben we hier geen informatie om uit te maken welk model het best is. Het relatief model (R2) gedraagt zich wel totaal anders. R2 hebben we ingevoerd vanuit een academisch perspectief. De achterliggende vraag was eerder in hoeverre de verhouding tussen het werkelijke volume en het volume van de kegel een constante was, dan een alternatief model te vinden.

9.4.2 De tarieven met één ingang

Zowel de vertekening als de voorspellingsmarge is heel groot bij alle modellen met één ingang en het is nodig de modellen op te splitsen per ecoregio. Overal verdwijnt dan de vertekening en behalve in het Kempens district, wordt ook de voorspellingsmarge veel beter. Zeker in het Zoniengewoud en het Brabants district, is de verbetering groot. De 95 % voorspellingsmarge daalt er tot ongeveer 25 %. In het Vlaams district is dat nog 30 % en in het Kempens district 35 %. De veeltermmodellen blijken iets beter te scoren dan de logaritmische modellen. Daarom bevelen we aan de veeltermmodellen te gebruiken. Vraag is dan nog wanneer we modellen met één ingang, dan wel met twee ingangen moeten toepassen. Uit bovenstaande analyse is duidelijk dat we nog veel kunnen winnen (in bepaalde districten tot 15 %) door te werken met twee ingangen op voorwaarde dat we de hoogte voldoende precies kunnen bepalen. Maar van zodra de gemiddelde meetfout in absolute waarde groter wordt dan 8 %, wordt de voorspellingsfout van de modellen met twee ingangen groter dan met één ingang.

10 Eik: tarieven (rudimentaire tabellen)

10.1 Inleiding

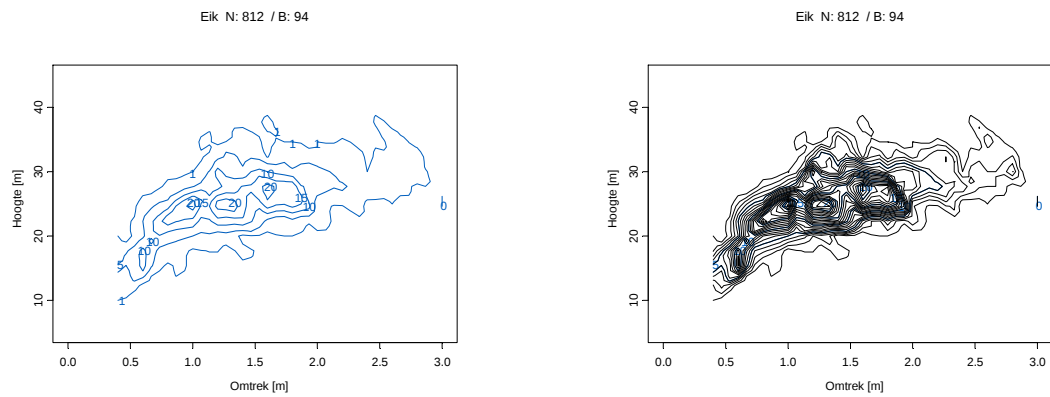
In deze rubriek geven we een eerste aanzet voor de tarieven. Hiertoe laten we omtrek variëren vanaf 0.35 m tot 3 m in stappen van 0.1 m en de hoogte van 10 m tot 40 m in stappen van 2.5 m. Deze punten zijn het centrum van een interval. Voor de omtrek reikt elk interval van -0.05 m tot +0.05 m ten opzichte van het centrum en voor de hoogte is dat van -1.25 m tot 1.25 m.

De tabellen met de volumeschattingen hebben we alleen berekend voor plausibele combinaties van hoogte en omtrek. Tabel 6-1 geeft de combinaties in de databank. Opvullingen van hiaten in het bereik en uitbreidingen hebben we aangeduid met "-1". Als bij een ontbrekende combinatie van omtrek en hoogte de omringende combinatie wel bestaat, dan vulden we het hiaat op (voor interpolatie). Aan de randen hebben we het bereik één cel uitgebreid (voor extrapolatie).

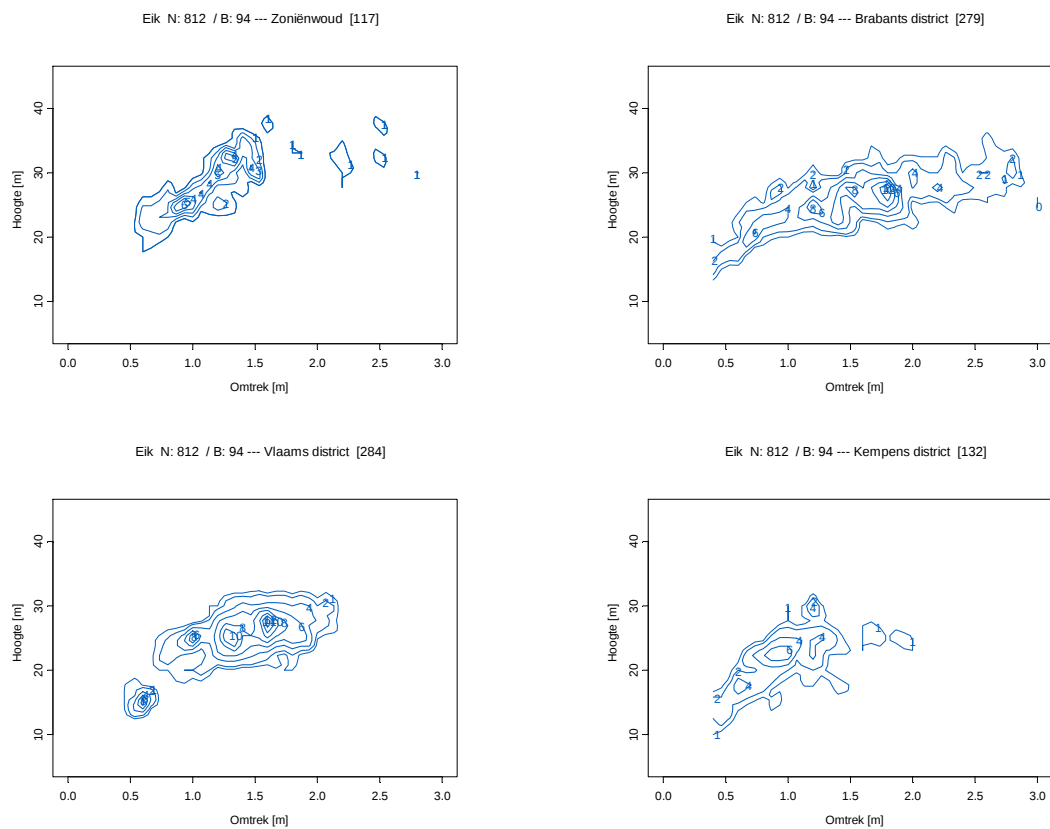
Van Tabel 6-1 maakten we een grafische voorstelling met contourgrafieken (Figuur 10-1 globaal en Figuur 10-2 per ecoregio). De meeste gegevens komen uit het Brabantse en het Vlaamse district. De gegevens van het Brabants district liggen verspreid over het volledige omtrekbereik van 0.35 m tot 3 m. In het Vlaams district is het omtrekinterval veel compacter, tot iets meer dan 2 m. In het Kempens district liggen de meeste waarnemingen onder 1.5 m. Dat is ook het geval in Zoniën. Zoals al vroeger geconstateerd is een opvallende eigenschap voor het Zoniën-woud dat de hoogte in functie van de omtrek gemiddeld hoger is dan de andere ecoregio's.

Tabel 10-1 Eik: meetbereik. Aantal bomen per combinatie van omtrek (vertikaal) en hoogteklaas (horizontaal) of ecoregio. Hiaten en/of uitbreidingen worden aangeduid met een punt (.). De gebieden in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).

	10	12.5	15	17.5	20	22.5	25	27.5	30	32.5	35	37.5	40	Z	B	V	K
0.4	1	2	6	2	1									1	5	1	6
0.5	.	4	4	6	.	.	.							.	3	5	6
0.6	.	.	12	12	4	5	2							7	7	11	10
0.7	1	.	2	4	11	11	5	.						7	12	5	10
0.8	.	.	.	3	7	19	6	2						9	9	7	12
0.9		2	2	6	15	17	3	.						7	11	9	18
1.0		.	.	5	14	24	5	1	.	.				9	10	14	16
1.1			1	4	7	12	8	2	1	2				8	10	10	9
1.2			1	2	14	23	8	12	3	.				12	14	20	17
1.3			.	3	13	25	13	6	7	.				13	13	33	8
1.4			2	.	15	17	16	6	3	4				10	21	27	5
1.5			1	1	9	11	17	12	4	2	.			10	21	24	2
1.6			.	.	4	19	23	10	.	.	2			2	19	34	3
1.7				1	1	15	19	8	4	1	.			1	23	22	3
1.8				1	4	18	20	6	1	1	.			2	26	22	1
1.9				.	1	15	8	8	3	.	.			1	12	19	3
2.0					3	2	8	7	1	1	.			1	12	7	2
2.1					1	1	5	5	3	.	.			1	7	7	.
2.2					.	1	6	3	4	1	.			5	8	2	.
2.3					.	3	4	4	4	.	.			1	11	2	1
2.4					1	.	4	1	.	.	.			.	4	2	.
2.5					.	.	2	3	2	.	2			5	4	.	.
2.6						1	2	2	1	3	.			2	7	.	.
2.7						.	2	2	1	1	1			2	4	1	.
2.8							.	4	2	.	.			1	5	.	.
2.9							1	.	.	.	.	.		.	1	.	.
3.0							.	.	.	.	.	.	.	.	.	.	.



Figuur 10-1 Eik: het empirisch bereik van de modellen (links met aantalcontouren in stappen van 5 en rechts in stappen van 1).



Figuur 10-2 Eik: het empirisch bereik van de modellen per ecoregio

10.2 Tarief met twee ingangen

10.2.1 De tarieven in tabelvorm

Als model voor het tarief met twee ingangen nemen we het veeltermmodel. De vergelijking is:

$$V = 0.1645 + 0.0230 C^2 H - 0.5612 C + 0.2910 C^2 - 0.00725 H + 0.0250 CH \quad (10.1)$$

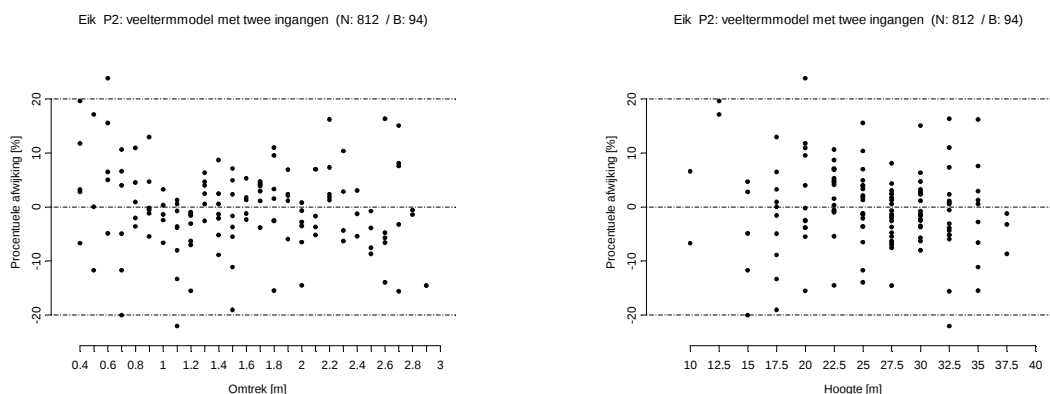
Tabel 6-2 geeft een rudimentaire tabel. Vanaf $H = 35 \text{ m}$ moesten veel hiaten opgevuld worden want daar zijn minder gegevens beschikbaar.

Tabel 10-2 Eik: tarief met twee ingangen (veeltermmodel) in dm³. De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1). De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).

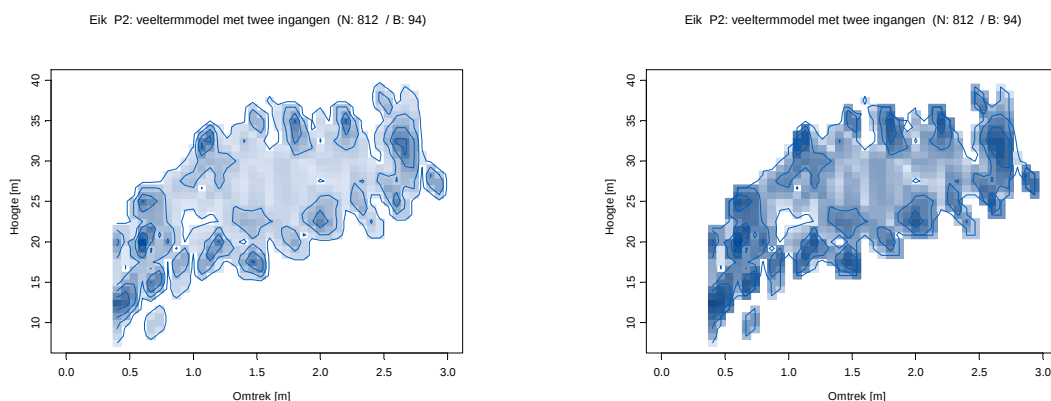
	Hoogte [m]												
	10	12.5	15	17.5	20	22.5	25	27.5	30	32.5	35	37.5	40
0.4	51	67	83	99	115								
0.5	•66	94	121	149	•176	•204	•231						
0.6	•93	•133	172	212	252	292	332						
0.7	129	•183	236	290	344	398	451	•505					
0.8	•176	•245	•313	382	450	519	587	656					
0.9		403	487	572	656	741	825	•910					
1		•505	•606	708	810	912	1013	1115	•1217	•1319			
1.1			740	860	980	1100	1220	1340	1460	1580			
1.2			887	1026	1166	1305	1445	1584	1724	•1863			
1.3			•1047	1207	1368	1528	1688	1848	2008	•2168			
1.4			1222	•1404	1586	1768	1950	2131	2313	2495			
1.5			1411	1615	1820	2025	2230	2435	2639	2844	•3049		
1.6			•1613	•1842	2071	2300	2528	2757	•2986	•3215	3444		
1.7				2083	2338	2592	2846	3100	3354	3608	•3862		
1.8				2340	2620	2901	3181	3461	3742	4022	•4303		
1.9				•2612	2919	3227	3535	3843	4151	•4459	•4767		
2					3235	3571	3908	4244	4581	4917	•5254		
2.1					3566	3932	4299	4665	5031	•5398	•5764		
2.2					•3914	4311	4708	5105	5503	5900	•6297		
2.3					•4277	4707	5136	5565	5995	•6424	•6853		
2.4					4657	•5120	5582	6045	•6507	•6970	•7433		
2.5					•5053	•5550	6047	6544	7041	•7538	8035		
2.6						5998	6530	7063	7595	8128	•8661		
2.7						•6463	7032	7601	8171	8740	9309		
2.8							•7552	8159	8766	•9374	•9981		
2.9							8091	•8737	•9383	•10029	•10675	•11321	
3							•8648	•9334	•10020	•10707	•11393	•12079	

10.2.2 Controle per klasse

Tabel 6-3 vergelijkt het tarief en het gemiddelde van de volumemetingen per klasse. De meeste afwijkingen liggen tussen -20 % en 20 %. De afwijkingen vertonen geen patroon in functie van omtrek en hoogte (Figuur 10-3). De kwaliteit van het model is vrij uniform over het volledige bereik. Over de geëxtrapoleerde waarden geven noch de tabel noch de figuur informatie. De voorstelling in twee dimensies (Figuur 6-4) toont hoe aan de randen de modelonzekerheid groter is.



Figuur 10-3 Eik: procentuele afwijking van het gemiddelde van de volumemetingen t. o. v. de tarieven.



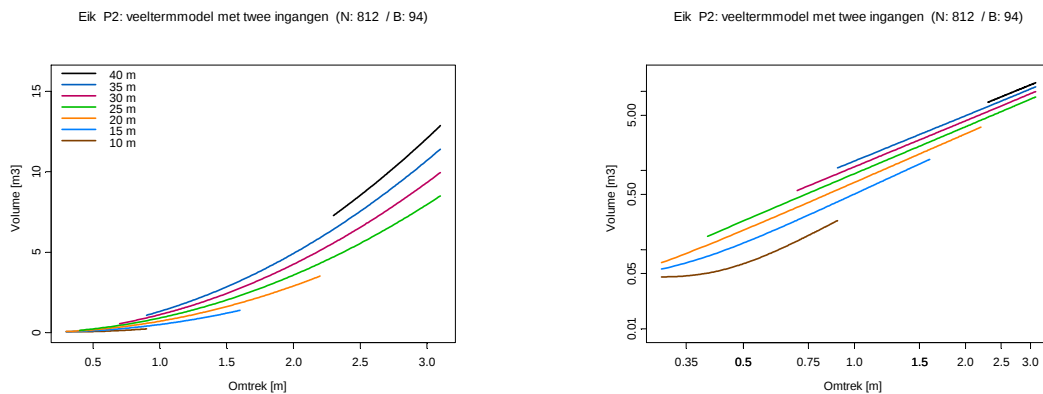
Figuur 10-4 Eik: verkenning van het tweedimensioneel patroon van de absolute afwijkingen (links: in de originele schaal, rechts in de log-schaal).

Tabel 10-3 Eik: vergelijking van het tarief met twee ingangen (veeltermmodel) ten opzichte van het gemiddelde per klasse. De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).

	Hoogte [m]													
	10	12.5	15	17.5	20	22.5	25	27.5	30	32.5	35	37.5	40	
0.4	-6.7	19.6	2.8	3.2	11.7									
0.5		17.1	-11.8	0.0										
0.6			-4.9	6.5	23.8	5.0	15.5							
0.7	6.6		-20.1	-5.0	4.0	10.6	-11.8							
0.8				0.9	10.9	4.5	-3.6	-2.1						
0.9			4.7	12.9	-0.2	-0.6	-1.2	-5.5						
1.0					-2.5	0.3	-1.4	-6.7	3.2					
1.1				-13.4	-3.9	-0.8	-3.7	0.5	-8.0	-22.1	1.2			
1.2				-1.6	-15.6	-1.0	-1.3	-7.0	-6.3	-3.1				
1.3					-2.6	4.6	4.0	2.4	6.3	0.5				
1.4				-8.9		8.7	-2.1	-1.4	2.4	-5.2	0.5			
1.5				-19.1	-5.5	7.1	4.9	-3.7	-1.7	2.3	-11.2			
1.6						5.3	1.3	1.7	-2.3			-1.3		
1.7					-3.8	4.1	3.9	4.3	4.7	1.1	2.9			
1.8					9.5	1.5	3.3	-2.6	-2.6	11.0	-15.5			
1.9						6.9	2.1	2.3	1.1	-6.0				
2.0						-14.6	-6.6	-0.7	-3.5	0.8	-2.8			
2.1						7.0	7.0	-1.7	-3.7	-5.2				
2.2							1.8	1.3	2.3	7.3	16.2			
2.3							10.3	-6.3	2.8	-4.4				
2.4						-5.4		3.0	-1.3					
2.5								-7.6	-0.8	-3.9		-8.7		
2.6							-14.0	-4.8	-5.8	16.3	-6.6			
2.7								8.0	15.0	-15.6	7.5	-3.3		
2.8									-1.4	-0.6				
2.9								-14.6						
3.0														

10.2.3 Grafische voorstelling van de tarieven

Figuur 10-5 vergelijkt de tarieven grafisch. Voor elke hoogte is het corresponderende omtrekke-reik gekozen, zodat we een realistisch beeld krijgen. De afwijkende curve voor een hoogte van 10 m (zie log-schaal) is niet noodzakelijk reëel aangezien daar de modelonzekerheid heel hoog is.



Figuur 10-5 Eik: grafische voorstelling van de tarieven per hoogteklasse (rechts in de log-schaal).

10.3 Tarieven met één ingang (veeltermmodel)

10.3.1 De tarieven in tabelvorm

Bij de validatie van het logaritmisch model tekende zich een patroon af. Daarom verkiezen we ook voor de tarieven met één ingang het veeltermmodel. De vergelijkingen zijn:

$$V = \begin{cases} Z \leftrightarrow 0.138 - 0.813C + 1.800C^2 - 0.146C^3 \\ B \leftrightarrow 0.021 - 0.247C + 1.018C^2 + 0.034C^3 \\ V \leftrightarrow 0.216 - 1.260C + 2.181C^2 - 0.311C^3 \\ K \leftrightarrow 0.138 - 0.858C + 1.838C^2 - 0.349C^3 \end{cases} \quad (10.2)$$

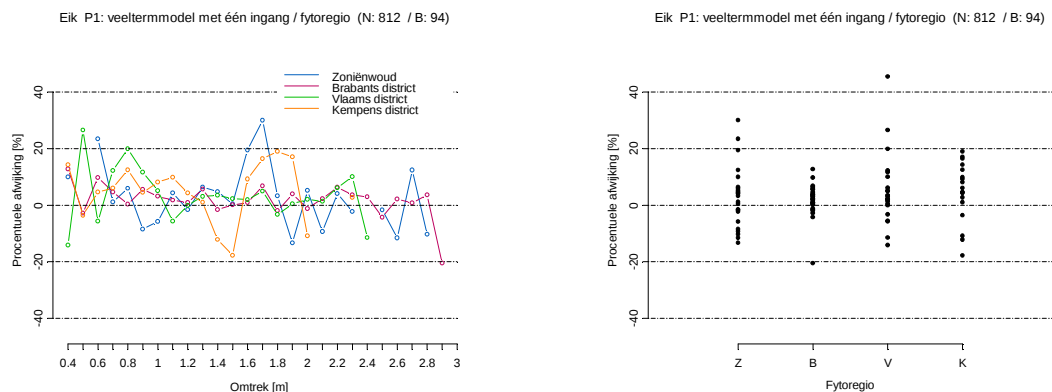
Net zoals bij het tarief met twee ingangen geven we alleen waarden binnen het bereik van de databank. In het Kempens district hebben de gegevens een smaller bereik. Voor grotere bomen raden we aan de tabel voor het Vlaamse district te gebruiken. Uiteraard is het altijd mogelijk met de formules zelf te rekenen.

Tabel 10-4 Eik: tarief met één ingang (veeltermmodel) en ecoregio in dm³. De waarden voor de referentieboom zijn onderlijnd. De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1). De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).

	Zoniënwoud	Brabants district	Vlaams district	Kempens district
0.4	92	87	41	66
0.5	•164	156	92	125
0.6	267	247	178	209
0.7	401	359	296	318
0.8	565	492	444	450
0.9	758	648	621	<u>601</u>
1.0	979	826	825	<u>770</u>
1.1	1228	1026	1054	955
1.2	<u>1502</u>	1249	1306	1153
1.3	<u>1803</u>	1495	1579	1364
1.4	2127	1764	<u>1872</u>	1584
1.5	<u>2476</u>	<u>2056</u>	<u>2182</u>	<u>1811</u>
1.6	2847	<u>2371</u>	2508	2044
1.7	3240	2710	2847	2280
1.8	3654	3073	3198	2518
1.9	4089	3460	3559	2754
2.0	4542	3871	3929	2988
2.1	5015	4306	4304	•3216
2.2	5504	4766	4684	•3438
2.3	6011	5251	5066	3650
2.4	•6533	5761	5450	•3851
2.5	7070	6297	•5832	
2.6	7621	6857	•6210	
2.7	8185	7444	6584	
2.8	8762	8056	•6951	
2.9	•9350	8695		
3.0	•9949	•9359		

10.3.2 Controle per klasse

Om deze tabel te valideren hebben we per klasse het gemiddelde volume van de gemeten bomen vergeleken met de waarde voorspeld door het tarief. Uit Tabel 6-5 blijkt dat bijna alle relatieve verschillen tussen -20 % en 20 % liggen. De afwijkingen vertonen ook geen patroon meer in functie van omtrek (Figuur 6-6). Het model is uniform geldig over het volledige bereik. Over de geëxtrapoleerde waarden kunnen we uiteraard niets zeggen.



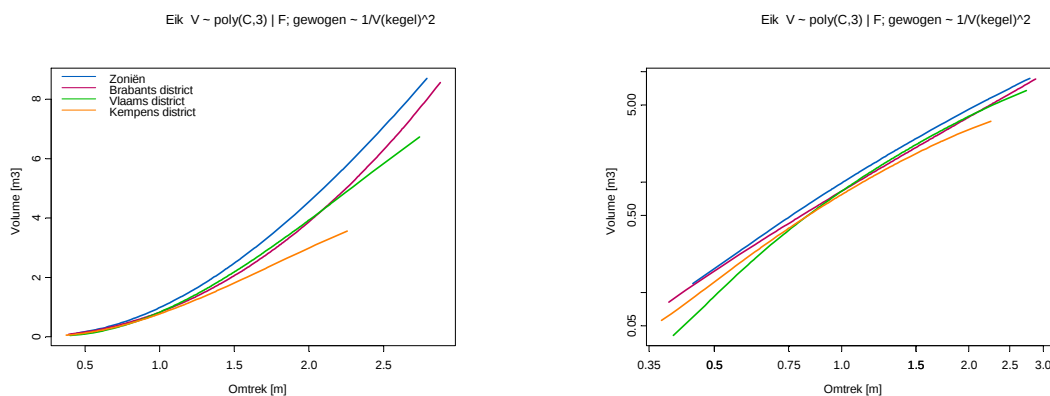
Figuur 10-6 Eik: procentuele afwijking van het gemiddelde van de volumemetingen per combinatie van omtrek- en ecoregio ten opzichte van de waarde in de tarieventabel.

Tabel 10-5 Eik: relatieve procentuele afwijking met één ingang (veeltermmodel) en ecoregio. De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).

	Zoniënwoud	Brabants district	Vlaams district	Kempens district
0.4	10.0	12.8	-14.1	14.3
0.5		-2.8	26.5	-3.6
0.6	23.4	9.7	-5.7	4.6
0.7	1.1	4.6	12.2	6.1
0.8	6.0	0.4	19.9	12.5
0.9	-8.5	5.6	11.7	4.6
1.0	-5.8	3.1	5.1	8.1
1.1	4.3	1.8	-5.7	9.8
1.2	-1.6	0.9	-0.1	4.3
1.3	6.4	5.7	3.1	1.0
1.4	4.8	-1.6	3.5	-12.2
1.5	0.5	0.1	2.3	-17.8
1.6	19.4	0.9	1.9	9.2
1.7	30.0	6.8	4.9	16.5
1.8	3.3	-2.0	-3.3	19.0
1.9	-13.3	3.9	0.4	17.0
2.0	5.2	-1.2	2.0	-10.9
2.1	-9.3	2.3	1.4	
2.2	4.1	6.3	6.2	
2.3	-2.3	3.6	10.0	2.8
2.4		3.0	-11.5	
2.5	-1.7	-4.3		
2.6	-11.6	2.1		
2.7	12.4	0.8	45.5	
2.8	-10.3	3.6		
2.9		-20.5		
3.0				

10.3.3 Grafische voorstelling van de tarieven

Figuur 6-8 is een kopie van Figuur 8-15. Gemiddeld ligt het voorspelde volume in Zoniën het hoogst. Het Brabants en Vlaams district hebben heel vergelijkbare waarden behalve aan de grenzen van het bereik. Vraag is of dat geen artefact is en of we niet beter deze tarieven groeperen. Ook het Kempens district wijkt sterk af vanaf ongeveer 1.5 m. Maar aangezien het omtrekgebied van de gegevens beperkt is, is niet duidelijk of het om een reëel verschil gaat.



Figuur 10-7 Eik: grafische vergelijking van de veeltermmodellen (P1f) in de originele schaal (links) en in de log-schaal (rechts).

10.4 Tarieven met één ingang (logaritmisch model)

10.4.1 Voorstelling van de tarieven in tabelvorm

Om het probleem met het logaritmisch model verder uit te diepen, maken we hier dezelfde bespreking als voor het veeltermmodel. De vergelijkingen per ecoregio zijn:

$$\log_{10}(V) = \begin{cases} Z \leftrightarrow 0.956 + 2.287 \log_{10}(C) \\ B \leftrightarrow 0.812 + 2.271 \log_{10}(C) \\ V \leftrightarrow 0.768 + 2.515 \log_{10}(C) \\ K \leftrightarrow 0.740 + 2.388 \log_{10}(C) \end{cases} \quad (10.3)$$

Aangezien het Vlaams district de grootste helling heeft, maar een kleiner intercept, liggen de waarden daar aanvankelijk lager, maar vanaf 1.30 m is het voorspelde houtvolume groter dan in Brabant. Uiteindelijk, voor heel grote waarden van de omtrek, zullen de voorspellingen in Vlaanderen zelfs hoger liggen dan in Zoniën. Maar het is wel belangrijk om te beseffen dat er voor deze waarden van de omtrek geen gegevens in de databank ter beschikking zijn. Eenzelfde opmerking geldt voor het Kempens district. Vanaf 2.30 m omtrek ligt het voorspelde volume hoger dan in het Vlaamse district, maar vanaf dan zijn er nog weinig gegevens ter beschikking.

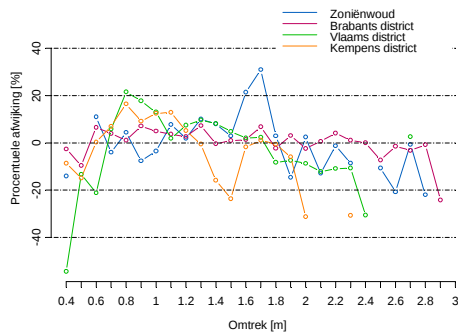
Tabel 10-6 Eik: tarief met één ingang op basis van het logaritmisch model in dm³. De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1). De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).

	Zoniën	woud	Brabants district	Vlaams district	Kempens district
0.4	118		101	77	83
0.5	•196		168	134	141
0.6	297		254	213	219
0.7	423		361	313	316
0.8	574		489	438	434
0.9	751		639	589	575
1	956		812	768	740
1.1	1189		1008	976	929
1.2	1451		1228	1215	1144
1.3	1742		1473	1486	1385
1.4	2064		1743	1791	1653
1.5	2417		2039	2130	1949
1.6	2802		2361	2506	2274
1.7	3218		2709	2918	2628
1.8	3668		3085	3370	3013
1.9	4151		3488	3860	3428
2	4667		3919	4392	3875
2.1	5218		4378	4965	•4354
2.2	5804		4866	5582	•4865
2.3	6426		5383	6242	5410
2.4	•7083		5930	6947	•5989
2.5	7776		6506	•7698	NA
2.6	8506		7112	•8496	NA
2.7	9272		7749	9342	NA
2.8	10077		8416	•10237	NA
2.9	•10919		9114	NA	NA
3	•11799		•9844	NA	NA

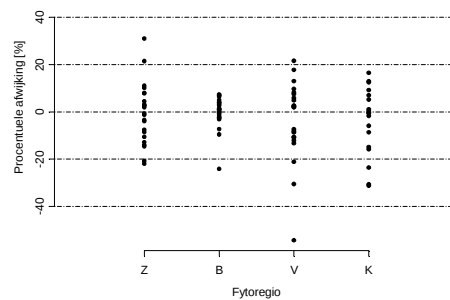
10.4.2 Controle van de tarieven

Tabel 6-7 en Figuur 6-8 tonen dat er een patroon is in de afwijkingen. Voor lage waarden van de omtrek is liggende de schattingen eerder hoger, voor grote waarden is er eerder een onderschatting.

Eik L1f: logaritmisch model met één ingang per fyto regio (N: 812 / B: 94)



Eik L1f: logaritmisch model met één ingang per fyto regio (N: 812 / B: 94)



Figuur 10-8 Eik: procentuele afwijking van het gemiddelde van de volumemetingen per combinatie van omtrek- en ecoregio ten opzichte van de waarden uit de tabel.

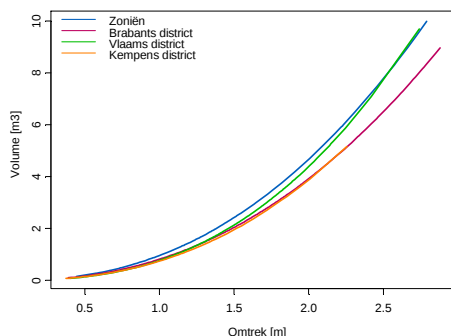
Tabel 10-7 Eik: vergelijking van het tarief met één ingang gebaseerd op het logaritmisch model ten opzichte van het gemiddelde in elke klasse. De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).

	<u>Zoniën</u> woud	<u>Brabants</u> district	<u>Vlaams</u> district	<u>Kempens</u> district	
0.4	-14.0	-2.7	-54.4	-8.7	
0.5		-9.6	-13.4	-15.0	
0.6	10.9	6.4	-21.2	0.3	
0.7	-4.0	3.9	5.9	7.0	
0.8	4.4	1.1	21.5	16.4	
0.9	-7.6	7.1	17.7	9.1	
1	-3.5	4.9	12.9	12.4	
1.1	7.7	3.6	1.8	12.8	
1.2	1.9	2.7	7.4	5.1	
1.3	10.1	7.2	9.6	-0.5	
1.4	7.9	-0.4	8.1	-15.9	
1.5	2.9	0.9	4.8	-23.6	
1.6	21.4	1.3	2.0	-1.8	
1.7	30.9	6.8	2.4	1.0	
1.8	2.9	-2.4	-8.2	-0.6	
1.9	-14.6	3.1	-7.4	-6.0	
2	2.4	-2.4	-8.7	-31.3	
2.1	-12.9	0.6	-12.1		
2.2	-1.3	4.1	-10.9		
2.3	-8.6	1.1	-10.7	-30.7	
2.4		0.0	-30.6		
2.5	-10.6	-7.4			
2.6	-20.8	-1.5			
2.7	-0.8	-3.2	2.5		
2.8	-22.0	-0.8			
2.9		-24.2			
3					

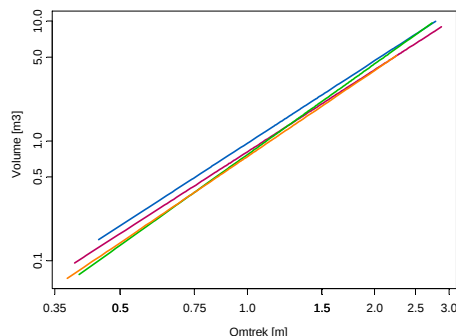
10.4.3 Grafische voorstelling van de tarieven

Figuur 10-9 is een kopie van Figuur 8-30. In het logaritmisch model verschillen Zoniën en het Brabants district voor grote waarden van de omtrek nog nauwelijks van elkaar. Tussenin is er wel een verschil. Wellicht omdat precies in dat bereik de hoogte van de bomen in Zoniën en het Brabants district sterk verschillend is. Het Vlaamse district en het Kempens district verschillen nauwelijks. Opnieuw is dat een indicatie dat we het tarief voor het Kempens en het Vlaamse district misschien wel samen moeten nemen.

Eik: L1f: logaritmisch model met één ingang per fyto regio (N: 812 / B: 94)



Eik: L1f: logaritmisch model met één ingang per fyto regio (N: 812 / B: 94)



Figuur 10-9 Eik: vergelijking van de logaritmische modellen (links in de originele schaal rechts en in de log-schaal).

10.5 Tarief voor grote omtrekken

Tot besluit vergelijken we de voorspellingen van de modellen voor de waarnemingen die we aan de kant hebben gezet bij de modelbouw (Tabel 6-8). Op basis van deze kleine steekproef kunnen we uiteraard geen besluiten trekken. Maar de cijfers bevestigen wel de algemene trend. Het veeltermmodel P2 en het logaritmisch model L2 sluiten het best aan bij de waarnemingen. Voor het logaritmisch model met één ingang zijn de voorspellingen systematisch te hoog. Bij het veeltermmodel is dat iets minder het geval.

Tabel 10-8 Eik: vergelijking van de modellen voor de waarnemingen met een omtrek groter dan 3m. F: = voorspelde waarde (*fitted value*) en R: = gestandaardiseerde residu's (*standardized residual*). Negatieve waarden wijzen op een overschatting want het voorspelde volume is dan groter dan de waargenomen volume.

	Ecoregio	Omtrek	Hoogte	Volume
304	B	3.05	29.1	10.3
340	B	3.23	30.5	9.3
448	B	3.01	33.8	9.9
751	B	3.14	29.8	9.2
794	B	3.34	29.9	10.6
800	Z	3.10	32.4	11.1

	F:P2	R:P2	F:L2	R:L2	F:R2	RvR2
304	9.39	9.7	9.31	10.6	9.47	8.7
340	10.94	-14.8	10.86	-14.2	11.16	-16.5
448	10.44	-5.0	10.40	-4.7	10.95	-9.4
751	10.14	-9.2	10.07	-8.5	10.29	-10.5
794	11.48	-7.9	11.38	-7.1	11.64	-9.2
800	10.65	4.4	10.59	4.9	11.04	0.6

	F:P1f	R:P1f	F:L1f	R:L1f
304	9.70	6.1	10.22	0.8
340	10.99	-15.2	11.64	-19.9
448	9.43	5.2	9.92	0.0
751	10.33	-10.9	10.92	-15.6
794	11.82	-10.5	12.56	-15.8
800	10.56	5.3	12.72	-12.6

Lijst van figuren

Figuur 1-1 De essentiële stappen bij statistische modelbouw gesitueerd binnen de fasen van het empirisch onderzoek.....	7
Figuur 3-1 Beuk: verkenning van de relatie tussen omtrek en hoogte. Links: originele schaal, rechts: log-schaal. De punten in het zwart geven de waarnemingen aan buiten het bereik van 0.35 m tot 4 m.....	18
Figuur 3-2 Beuk: verkenning van de relatie tussen omtrek en volume. Links: originele schaal, rechts: log-schaal. De punten in het zwart geven de waarnemingen aan buiten het bereik van 0.35 m tot 4 m.....	18
Figuur 3-3 Beuk: boxplots van omtrek, hoogte, volume en relatief volume. De punten geven het gemiddelde aan.	20
Figuur 3-4 Beuk: cumulatieve distributies van omtrek, hoogte, volume en relatief volume.	20
Figuur 3-5 Beuk: relatie tussen omtrek en hoogte (links: originele schaal, rechts: log-schaal).	21
Figuur 3-6 Beuk: relatie tussen omtrek en het volume relatief ten opzichte van het volume van een kegel (links: originele schaal, rechts: log-schaal).	22
Figuur 3-7 Beuk: relatie tussen omtrek en volume (links: originele schaal, rechts: log-schaal).	22
Figuur 3-8 Beuk: relatie tussen omtrek en het volume relatief ten opzichte van het volume van een kegel (links: originele schaal, rechts: log-schaal).	23
Figuur 3-9 Beuk: relatie tussen de variantie van de waarnemingen en het volume van een kegel (in klassen van 0.25 m ³).....	23
Figuur 4-1 Beuk: residuanalyse voor het veeltermmodel met twee ingangen (P2): QQ-grafieken.....	25
Figuur 4-2 Beuk: residuanalyse voor het veeltermmodel met twee ingangen (P2): boxplots. ..	25
Figuur 4-3 Beuk: residuanalyse voor het veeltermmodel met twee ingangen (P2): Relatie met omtrek en hoogte.	25
Figuur 4-4 Beuk: voorstelling van het veeltermmodel met twee ingangen (P2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de gestandaardiseerde residu's liggen.	26
Figuur 4-5 Beuk: grafische voorstelling van het veeltermmodel met twee ingangen (P2) per hoogteklasse (rechts in de log-schaal).	26
Figuur 4-6 Beuk: residuanalyse voor het veeltermmodel met één ingang (P1). QQ-plots.	27
Figuur 4-7 Beuk: residuanalyse voor het veeltermmodel met één ingang (P1). Boxplots.	27
Figuur 4-8 Beuk: residuanalyse voor het veeltermmodel met één ingang (P1). Relatie met omtrek en hoogte.	28
Figuur 4-9 Beuk: grafiek van het veeltermmodel met één ingang (P1). Links = originele schaal & rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de gestandaardiseerde residu's liggen.	28
Figuur 4-10 Beuk: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). QQ-plots.	29
Figuur 4-11 Beuk: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). Boxplots	29
Figuur 4-12 Beuk: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). Relatie met omtrek en hoogte.	30
Figuur 4-13 Beuk: residuanalyse van het veeltermmodel met één ingang per ecoregio (P1f) in functie van omtrek (C).....	30
Figuur 4-14 Beuk: grafiek van het veeltermmodel met één ingang per ecoregio (P1f). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten onder en boven liggen 95% van de gestandaardiseerde residu's.	32
Figuur 4-15 Beuk: grafische vergelijking van de veeltermmodellen (P1f) in de originele schaal (links) en in de log-schaal (rechts).	32
Figuur 4-16 Beuk: residuanalyse voor het logaritmisch model met twee ingangen (L2): QQ-grafieken.	33

Figuur 4-17 Beuk: residuanalyse voor het logaritmisch model met twee ingangen (L2): boxplots.	34
Figuur 4-18 Beuk: residuanalyse voor het logaritmisch model met twee ingangen (L2): omtrek en hoogte.	34
Figuur 4-19 Beuk: grafiek van het logaritmisch model met twee ingangen (L2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de gestandaardiseerde residu's liggen.	35
Figuur 4-20 Beuk: grafische voorstelling van het logaritmisch model met twee ingangen (L2) per hoogteklaas (rechts in de log-schaal).	35
Figuur 4-21 Beuk: residuanalyse voor het logaritmisch model met één ingang (L1). QQ-plots.	36
Figuur 4-22 Beuk: residuanalyse voor het logaritmisch model met één ingang (L1). Boxplots.	36
Figuur 4-23 Beuk: residuanalyse voor het logaritmisch model met één ingang (L1). Relatie met omtrek en hoogte.	36
Figuur 4-24 Beuk: grafiek van het logaritmisch model met één ingang (L1). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten liggen 95% van de gestandaardiseerde residu's.	37
Figuur 4-25 Beuk: residuanalyse voor het logaritmisch model met één ingang per district (L1f). QQ-plots.	38
Figuur 4-26 Beuk: residuanalyse voor het logaritmisch model met één ingang per district (L1f). Boxplots.	38
Figuur 4-27 Beuk: residuanalyse voor het logaritmisch model met één ingang per district (L1f). Relatie met omtrek en hoogte.	38
Figuur 4-28 Beuk: residuanalyse van het logaritmisch model met één ingang per district (L1f) in functie van omtrek (C).	39
Figuur 4-29 Beuk: grafiek van het logaritmisch model met één ingang per ecoregio (L1f). Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten onder en boven liggen 95% van de gestandaardiseerde residu's.	40
Figuur 4-30 Beuk: vergelijking van de logaritmische modellen (links in de originele schaal rechts en in de log-schaal).	40
Figuur 4-31 Beuk: residuanalyse voor het relatief model met twee ingangen (R2). QQ-grafieken.	41
Figuur 4-32 Beuk: residuanalyse voor het relatief model met twee ingangen (R2): boxplots.	42
Figuur 4-33 Beuk: residuanalyse voor het relatief model met twee ingangen (R2). Relatie met omtrek en hoogte.	42
Figuur 4-34 Beuk: evolutie van het relatief volume in functie van de omtrek waarbij de hoogte gelijk is aan de voorspelde waarde bij de omtrek (rechts uitgesplitst per ecoregio)...	43
Figuur 4-35 Beuk: grafiek van het relatief model met twee ingangen (R2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de gestandaardiseerde residu's liggen.	43
Figuur 4-36 Beuk: grafiek van het relatief model met twee ingangen (R2) per hoogteklaas (rechts in de log-schaal).	43
Figuur 5-1 Beuk: vergelijking van de modellen met twee ingangen. P2 = Veeltermmodel, L2 = logaritmisch model en R2 = relatief model. De punt-streep lijn in het vet is de nullijn. De punt-streep lijn onder en boven omvatten 95 % van de punten.	45
Figuur 5-2 Eik: vergelijking van de modellen met twee ingangen. P2 = Veeltermmodel, L2 = logaritmisch model en R2 = relatief model. De punt-streep lijn in het vet is de nullijn. De punt-streep lijn onder en boven omvatten 95 % van de punten. De figuur is dezelfde als Figuur 5-1, maar de uitbijters zijn weggelaten om meer details te zien.	46
Figuur 5-3 Beuk: vergelijking van de modellen met één ingang per ecoregio in functie van de omtrek en de hoogte.	47
Figuur 5-4 Beuk: boxplots van de gestandaardiseerde residu's (%).	47
Figuur 5-5 Beuk: vertekening en voorspellingsfout	49
Figuur 5-6 Beuk: vertekening en voorspellingsfout opgesplitst naargelang fytoregio.	50
Figuur 5-7 Beuk: invloed van meetfouten op de hoogte op de vertekening en voorspellingsfout. De gemiddelde relatieve fout in absolute waarde is respectievelijk 0%, 4%, 8% en 12%.	51

Figuur 6-1 Beuk: het empirisch bereik van de modellen.	53
Figuur 6-2 Beuk: het empirisch bereik van de modellen per ecoregio	53
Figuur 6-3 Beuk: procentuele afwijking van het gemiddelde van de volumemetingen t. o. v. de tarieven.	54
Figuur 6-4 Beuk: verkenning van het tweedimensioneel patroon van de absolute afwijkingen (links: in de originele schaal, rechts in de log-schaal).	55
Figuur 6-5 Beuk: grafische voorstelling van de tarieven per hoogteklaas (rechts in de log-schaal).	56
Figuur 6-6 Beuk: procentuele afwijking van het gemiddelde van de volumemetingen per combinatie van omtrek- en ecoregio ten opzichte van de waarde in de tarieventabel.	57
Figuur 6-7 Beuk: grafische vergelijking van de veeltermmodellen (P1f) in de originele schaal (links) en in de log-schaal (rechts).	58
Figuur 6-8 Beuk: procentuele afwijking van het gemiddelde van de volumemetingen per combinatie van omtrek- en ecoregio ten opzichte van de waarden uit de tabel.	60
Figuur 6-9 Beuk: vergelijking van de logaritmische modellen (links in de originele schaal rechts en in de log-schaal).	61
Figuur 7-1 Eik: verkenning van de relatie tussen omtrek en hoogte. Links: originele schaal, rechts: log-schaal. De punten in het zwart geven de waarnemingen aan buiten het bereik van 0.35 m tot 3 m.	63
Figuur 7-2 Eik: verkenning van de relatie tussen omtrek en volume. Links: originele schaal, rechts: log-schaal. De punten in het zwart geven de waarnemingen aan buiten het bereik van 0.35 m tot 3 m.	63
Figuur 7-3 Eik: boxplots van omtrek, hoogte, volume en relatief volume. De punten geven het gemiddelde aan.	64
Figuur 7-4 Eik: cumulatieve distributie van omtrek, hoogte, volume en relatief volume.	65
Figuur 7-5 Eik: relatie tussen omtrek en hoogte (links: originele schaal, rechts: log-schaal). ..	66
Figuur 7-6 Eik: relatie tussen het volume van de kegel en volume (links: originele schaal, rechts: log-schaal).	66
Figuur 7-7 Eik: relatie tussen omtrek en volume (links: originele schaal, rechts: log-schaal). .	67
Figuur 7-8 Eik: relatie tussen omtrek en het volume relatief ten opzichte van het volume van een kegel (links: originele schaal, rechts: log-schaal).	67
Figuur 7-9 Eik: relatie tussen de variantie van de waarnemingen en het volume van een kegel (in klassen van 0.25 m ³).	68
Figuur 8-1 Eik: residuanalyse voor het veeltermmodel met twee ingangen (P2): QQ-grafieken	70
Figuur 8-2 Eik: residuanalyse voor het veeltermmodel met twee ingangen (P2): boxplots.	70
Figuur 8-3 Eik: residuanalyse voor het veeltermmodel met twee ingangen (P2): Relatie met omtrek en hoogte.	70
Figuur 8-4 Eik: grafiek van het veeltermmodel met twee ingangen (P2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de 95% voorspellingsintervallen.	71
Figuur 8-5 Eik: grafiek van het veeltermmodel met twee ingangen (P2) per hoogteklaas (rechts in de log-schaal).	72
Figuur 8-6 Eik: residuanalyse voor het veeltermmodel met één ingang (P1). QQ-plots.	72
Figuur 8-7 Eik: residuanalyse voor het veeltermmodel met één ingang (P1). Boxplots.	73
Figuur 8-8 Eik: residuanalyse voor het veeltermmodel met één ingang (P1). Relatie met omtrek en hoogte.	73
Figuur 8-9 Eik: grafiek van het veeltermmodel met één ingang (P1). Links = originele schaal & rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de relatieve residu's liggen.	74
Figuur 8-10 Eik: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). QQ-plots.	75
Figuur 8-11 Eik: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). Boxplots	75
Figuur 8-12 Eik: residuanalyse voor het veeltermmodel met één ingang en ecoregio (P1f). Relatie met omtrek en hoogte.	75

Figuur 8-13 Eik: residuanalyse van het veeltermmodel met één ingang per ecoregio (P1f) in functie van omtrek (C).	76
Figuur 8-14 Eik: grafiek van het veeltermmodel met één ingang per ecoregio (P1f). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten onder en boven liggen 95% van de relatieve residu's.	77
Figuur 8-15 Eik: grafische vergelijking van de veeltermmodellen (P1f) in de originele schaal (links) en in de log-schaal (rechts).	78
Figuur 8-16 Eik: residuanalyse voor het logaritmisch model met twee ingangen (L2): QQ-grafieken.	79
Figuur 8-17 Eik: residuanalyse voor het logaritmisch model met twee ingangen (L2): boxplots.	79
Figuur 8-18 Eik: residuanalyse voor het logaritmisch model met twee ingangen (L2): omtrek en hoogte.	79
Figuur 8-19 Eik: grafiek van het logaritmisch model met twee ingangen (L2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de relatieve residu's liggen.	80
Figuur 8-20 Eik: grafiek van het logaritmisch model met twee ingangen (L2) per hoogteklaas (rechts in de log-schaal).	80
Figuur 8-21 Eik: residuanalyse voor het logaritmisch model met één ingang (L1). QQ-plots.	81
Figuur 8-22 Eik: residuanalyse voor het logaritmisch model met één ingang (L1). Boxplots.	82
Figuur 8-23 Eik: residuanalyse voor het logaritmisch model met één ingang (L1). Relatie met omtrek en hoogte.	82
Figuur 8-24 Eik: grafiek van het logaritmisch model met één ingang (L1). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten liggen 95% van de relatieve residu's.	83
Figuur 8-25 Eik: residuanalyse voor het logaritmisch model met één ingang per district (L1f). QQ-plots.	84
Figuur 8-26 Eik: residuanalyse voor het logaritmisch model met één ingang per district (L1f). Boxplots.	84
Figuur 8-27 Eik: residuanalyse voor het logaritmisch model met één ingang per district (L1f). Relatie met omtrek en hoogte.	84
Figuur 8-28 Eik: residuanalyse van het logaritmisch model met één ingang per district (L1f) in functie van omtrek (C).	85
Figuur 8-29 Eik: grafiek van het logaritmisch model met één ingang per ecoregio (L1f). Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. Binnen limieten onder en boven liggen 95% van de relatieve residu's.	86
Figuur 8-30 Eik: vergelijking van de logaritmische modellen (links in de originele schaal rechts en in de log-schaal).	86
Figuur 8-31 Eik: residuanalyse voor het relatief model met twee ingangen (R2). QQ-grafieken.	87
Figuur 8-32 Eik: residuanalyse voor het relatief model met twee ingangen (R2): boxplots.	87
Figuur 8-33 Eik: residuanalyse voor het relatief model met twee ingangen (R2). Relatie met omtrek en hoogte.	88
Figuur 8-34 Eik: evolutie van het relatief volume in functie van de omtrek waarbij de hoogte gelijk is aan de voorspelde waarde bij de omtrek (rechts uitgesplitst per ecoregio).	88
Figuur 8-35 Eik: grafiek van het relatief model met twee ingangen (R2). Links = originele schaal + rechts = log-schaal. Het volume is berekend bij een hoogte voorspeld op basis van de omtrek. De lijnen onder en boven zijn de limieten waarbinnen 95% van de relatieve residu's liggen.	89
Figuur 8-36 Eik: grafiek van het relatief model met twee ingangen (R2) per hoogteklaas (rechts in de log-schaal).	89
Figuur 9-1 Eik: vergelijking van de modellen met twee ingangen. P2 = Veeltermmodel, L2 = logaritmisch model en R2 = relatief model. De punt-streep lijn in het vet is de nullijn. De punt-streep lijn onder en boven omvatten 95 % van de punten.	91
Figuur 9-2 Eik: vergelijking van de modellen met twee ingangen. P2 = Veeltermmodel, L2 = logaritmisch model en R2 = relatief model. De punt-streep lijn in het vet is de nullijn.	

De punt-streep lijn onder en boven omvatten 95 % van de punten. De figuur is dezelfde als Figuur 9-1, maar de uitbijters zijn weggelaten om meer details te zien.	92
Figuur 9-3 Eik: vergelijking van de modellen met één ingang per ecoregio in functie van de omtrek en de hoogte.	93
Figuur 9-4 Eik: boxplots van de gestandaardiseerde residu's (%).....	93
Figuur 9-5 Eik: vertekening en 95 % voorspellingsfout (procentueel).....	95
Figuur 9-6 Eik: vertekening en 95 % voorspellingsfout (procentueel) uitgesplitst per ecoregio. De punten voor het relatief model (waarvan er maar één variant van is) zijn iets sterker omcirkeld en niet verbonden door lijnen.....	96
Figuur 9-7 Eik: de invloed van gemiddelde meetfout op de hoogte (0 %, 4 %, 8 % en 12 %) op de vertekening van de het voorspelde volume en de voorspellingsfout.....	96
Figuur 10-1 Eik: het empirisch bereik van de modellen (links met aantalcontouren in stappen van 5 en rechts in stappen van 1).	99
Figuur 10-2 Eik: het empirisch bereik van de modellen per ecoregio.....	99
Figuur 10-3 Eik: procentuele afwijking van het gemiddelde van de volumemetingen t. o. v. de tarieven.	100
Figuur 10-4 Eik: verkenning van het tweedimensioneel patroon van de absolute afwijkingen (links: in de originele schaal, rechts in de log-schaal).	101
Figuur 10-5 Eik: grafische voorstelling van de tarieven per hoogteklaas (rechts in de log-schaal).	102
Figuur 10-6 Eik: procentuele afwijking van het gemiddelde van de volumemetingen per combinatie van omtrek- en ecoregio ten opzichte van de waarde in de tarieventabel.	103
Figuur 10-7 Eik: grafische vergelijking van de veeltermmodellen (P1f) in de originele schaal (links) en in de log-schaal (rechts).	104
Figuur 10-8 Eik: procentuele afwijking van het gemiddelde van de volumemetingen per combinatie van omtrek- en ecoregio ten opzichte van de waarden uit de tabel.	106
Figuur 10-9 Eik: vergelijking van de logaritmische modellen (links in de originele schaal rechts en in de log-schaal).	107

Lijst van tabellen

Tabel 2-1	Overzicht van de kandidaatmodellen	12
Tabel 3-1	Beuk: enkele richtgetallen, alle districten samen en opgesplitst per district.....	19
Tabel 3-2	Beuk: referentieboom en referentiebereik per ecoregio	21
Tabel 3-3	Beuk: ANOVA-tabel voor een veeltermmodel van de standaardafwijking in functie van het volume van de kegel. De term $\text{poly}(V_{\text{kegel}}, 2)$ is een veelterm (polynoom) van de tweede graad. Aangezien de component van de tweede graad niet significant is en de component van de eerste graad wel, is de conclusie dat het verband tussen de standaardafwijking en de het volume van een kegel lineair is.	23
Tabel 4-1	Beuk: ANOVA-tabel van het finaal veeltermmodel met twee ingangen.....	24
Tabel 4-2	Eik: coëfficiënten van het veeltermmodel met alleen kegel	25
Tabel 4-3	Beuk: coëfficiënten van het veeltermmodel met twee ingangen (variant 1: met V_{kegel} en referentie).....	26
Tabel 4-4	Beuk: coëfficiënten van het veeltermmodel met twee ingangen (variant 2: met $C2H$ en intercept)	26
Tabel 4-5	Beuk: ANOVA-tabel van het veeltermmodel met één ingang.	27
Tabel 4-6	Beuk: coëfficiënten van het veeltermmodel met één ingang.....	28
Tabel 4-7	Beuk: ANOVA-tabel van het veeltermmodel met één ingang en ecoregio.	29
Tabel 4-8	Beuk: coëfficiënten van het veeltermmodel met één ingang en ecoregio.....	31
Tabel 4-9	Beuk: vergelijking van de coëfficiënten van het veeltermmodel met één ingang en ecoregio.	31
Tabel 4-10	Beuk: ANOVA-tabel van het logaritmisch model met twee ingangen.....	33
Tabel 4-11	Beuk: coëfficiënten van het logaritmisch model met twee ingangen (met interactie)34	
Tabel 4-12	Beuk: coëfficiënten van het logaritmisch model met twee ingangen (zonder interactie).....	34
Tabel 4-13	Beuk: ANOVA-tabel van het logaritmisch model met één ingang.	35
Tabel 4-14	Beuk: coëfficiënten van het logaritmisch model met één ingang.	36
Tabel 4-15	Beuk: ANOVA-tabel van het logaritmisch model met één ingang en ecoregio.....	37
Tabel 4-16	Beuk: coëfficiënten van het logaritmisch model met één ingang en ecoregio.	39
Tabel 4-17	Beuk: vergelijking van de coëfficiënten van het logaritmisch model met één ingang en ecoregio.	39
Tabel 4-18	Beuk: ANOVA-tabel van het relatief model met twee ingangen.	41
Tabel 4-19	Beuk: coëfficiënten van het relatief model met twee ingangen.....	42
Tabel 5-1	Beuk: schatting voor een referentieboom ($C = 1.8 \text{ m}$, $H = 32.5 \text{ m}$)	44
Tabel 5-2	Beuk: procentueel verschil tussen modellen voor tarieven met twee ingangen. Enkele kengetallen.	44
Tabel 5-3	Beuk: procentueel verschil tussen modellen voor tarieven met één ingang en ecoregio ($P1f - L1f$).	46
Tabel 5-4	Beuk: juistheid van de modellen met als maat de mediaan van de gestandaardiseerde residu's.	48
Tabel 5-5	Beuk: precisie van de modellen met als maat de relatieve voorspellingsfout van de gestandaardiseerde residu's.	48
Tabel 6-1	Beuk: het meetbereik. Aantal bomen per combinatie van omtrek (vertikaal) en hoogteklasse (horizontaal) of ecoregio. De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd. De punten (•) zijn hiaten of uitbreidingen (Tabel 6-1).	52
Tabel 6-2	Beuk: tarief met twee ingangen (veeltermmodel) in dm^3 . De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd. De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1).	54
Tabel 6-3	Beuk: vergelijking van het tarief met twee ingangen (veeltermmodel) ten opzichte van het gemiddelde per klasse. De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd.	55

Tabel 6-4 Beuk: tarief met één ingang (veeltermmodel) en ecoregio in dm ³ . De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1). De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd.....	57
Tabel 6-5 Beuk: tarief met één ingang (veeltermmodel) en ecoregio. De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd.....	58
Tabel 6-6 Beuk: tarief met één ingang op basis van het logaritmisch model in dm ³ . De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1). De gekleurde stroken geven per ecoregio het interval waarbinnen de mediaan van omtrek en hoogte liggen (Tabel 3-2). De waarde voor de algemene referentieboom is onderlijnd.....	60
Tabel 6-7 Beuk: vergelijking van het tarief met één ingang op basis van het logaritmisch model ten opzichte van het gemiddelde in elke klasse. De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).	61
Tabel 6-8 Eik: vergelijking van de modellen voor de waarnemingen met een omtrek groter dan 3m. F: = voorspelde waarde (<i>fitted value</i>) en R: = gestandaardiseerd verschil (<i>standardized residual</i>).	62
Tabel 7-1 Eik: enkele richtgetallen, alle districten samen en opgesplitst per district.....	64
Tabel 7-2 Eik: referentieboom en referentiebereik per ecoregio	65
Tabel 7-3 Eik: ANOVA-tabel voor de veeltermmodellen van de standaardafwijking in functie van het volume van de kegel. De term poly(Vkegel,2) is een veelterm (polynoom) van de tweede graad. Aangezien de component van de tweede graad niet significant is en de component van de eerste graad wel, is de conclusie dat het verband tussen de standaardafwijking en de het volume van een kegel lineair is.	68
Tabel 8-1 Eik: ANOVA-tabel van het finaal veeltermmodel met twee ingangen (stapsgewijs invoegen).....	69
Tabel 8-2 Eik: coëfficiënten van het veeltermmodel met alleen kegel.....	71
Tabel 8-3 Eik: coëfficiënten van het veeltermmodel met twee ingangen (variant 1: met Vkegel en referentie)	71
Tabel 8-4 Eik: coëfficiënten van het veeltermmodel met twee ingangen (variant 2: met C2H en intercept).....	71
Tabel 8-5 Eik: ANOVA-tabel van het veeltermmodel met één ingang.....	72
Tabel 8-6 Eik: coëfficiënten van het veeltermmodel met één ingang.	73
Tabel 8-7 Eik: ANOVA-tabel van het veeltermmodel met één ingang en ecoregio.	74
Tabel 8-8 Eik: coëfficiënten van het veeltermmodel met één ingang en ecoregio.	76
Tabel 8-9 Eik: vergelijking van de veeltermmodellen met één ingang en ecoregio.	77
Tabel 8-10 Eik: ANOVA-tabel van het logaritmisch model met twee ingangen.	78
Tabel 8-11 Eik: coëfficiënten van het logaritmisch model met twee ingangen, met interactie ..	80
Tabel 8-12 Eik: coëfficiënten van het logaritmisch model met twee ingangen, zonder interactie	80
Tabel 8-13 Eik: ANOVA-tabel van het logaritmisch model met één ingang.	81
Tabel 8-14 Eik: coëfficiënten van het logaritmisch model met één ingang.....	82
Tabel 8-15 Eik: ANOVA-tabel van het logaritmisch model met één ingang en ecoregio.	83
Tabel 8-16 Eik: coëfficiënten van het logaritmisch model met één ingang en ecoregio.	85
Tabel 8-17 Eik: vergelijking van de coëfficiënten van het logaritmisch model met één ingang en ecoregio.....	85
Tabel 8-18 Eik: ANOVA-tabel van het relatief model met twee ingangen.	87
Tabel 8-19 Eik: coëfficiënten van het relatief model met twee ingangen.....	88
Tabel 9-1 Eik: volumeschatting voor een referentieboom (C = 2 m, H = 30 m) voor de verschillende tarieven	90
Tabel 9-2 Eik: procentueel verschil tussen modellen voor tarieven met twee ingangen.....	90
Tabel 9-3 Eik: verschil tussen de twee modellen voor tarieven met één ingang in functie van de ecoregio (P1f – L1f).....	93
Tabel 9-4 Eik: juistheid van de modellen met als maat de mediaan van de gestandaardiseerde residu's.....	94
Tabel 9-5 Eik: precisie van de modellen met als maat het 95 % relatieve voorspellingsinterval	94

Tabel 10-1 Eik: meetbereik. Aantal bomen per combinatie van omtrek (vertikaal) en hoogteklaas (horizontaal) of ecoregio. Hiaten en/of uitbreidingen worden aangeduid met een punt (•). De gebieden in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).....	98
Tabel 10-2 Eik: tarief met twee ingangen (veeltermmodel) in dm ³ . De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1). De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).....	100
Tabel 10-3 Eik: vergelijking van het tarief met twee ingangen (veeltermmodel) ten opzichte van het gemiddelde per klasse. De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).....	101
Tabel 10-4 Eik: tarief met één ingang (veeltermmodel) en ecoregio in dm ³ . De waarden voor de referentieboom zijn onderlijnd. De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1). De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).	103
Tabel 10-5 Eik: relatieve procentuele afwijking met één ingang (veeltermmodel) en ecoregio. De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).	104
Tabel 10-6 Eik: tarief met één ingang op basis van het logaritmisch model in dm ³ . De waarden met een • zijn hiaten of uitbreidingen (Tabel 6-1). De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).	105
Tabel 10-7 Eik: vergelijking van het tarief met één ingang gebaseerd op het logaritmisch model ten opzichte van het gemiddelde in elke klasse. De referentieboom is onderlijnd. De getallen in het geel geven het centrum van de waarnemingen per ecoregio (Tabel 3-2).	106
Tabel 10-8 Eik: vergelijking van de modellen voor de waarnemingen met een omtrek groter dan 3m. F: = voorspelde waarde (<i>fitted value</i>) en R: = gestandaardiseerde residu's (<i>standardized residual</i>). Negatieve waarden wijzen op een overschatting want het voorspelde volume is dan groter dan de waargenomen volume.	107