



Konstruktion, test och underhåll av simuleringsfunktioner i Riksskogstaxeringen



Hans Åkesson & Bertil Westerlund

Arbetsrapport 419 2014

Sveriges lantbruksuniversitet
Institutionen för skoglig resurshushållning
901 83 UMEÅ
www.slu.se/srh
090/7868100



ISSN 1401-1204
ISRN SLU-SRG-AR-419-SE

Konstruktion, test och underhåll av simuleringsfunktioner i Riksskogstaxeringen

Hans Åkesson & Bertil Westerlund

Nyckelord; simulering, trädvolym, tillväxt, Riksskogstaxering

Arbetsrapport 419 2014

Sveriges lantbruksuniversitet
Institutionen för skoglig resurshushållning
Utgivningsort: Umeå
Utgivningsår: 2014

ISSN 1401-1204
ISRN SLU-SRG-AR-419-SE

Förord

År 2005 påbörjades utvecklingsprojektet ”Översyn av Riksskogstaxeringens databas”. De övergripande målsättningarna med projektet var att minimera följande identifierade problem med det gamla databssystemet:

- Systemet var alltför känsligt för personella förändringar.
- Underhåll av systemet var för komplicerat.
- Behovet att kunna möta kunders och användares krav och behov förväntades förändras och öka.

Förutom att ta fram en helt ny och mer ändamålsenlig databasmodell med tillhörande dokumentationshantering samt uppgradera till en enhetlig teknisk plattform så gjordes en fullständig revision av de beräkningsrutiner som används för att med de data som registrerats i fältarbetet skapa nödvändiga beräknade variabler. Några exempel på sådana är volymer och tillväxt för träd. I Riksskogstaxeringen så bestäms trädslag för och mäts diametern på alla träd på provytorna som är högre än 1,3 m, s.k. klavträd. Det är frågan om ungefär 100 000 träd årligen. Av dessa träd väljs ca 13 000 träd ut till provträd som mäts mer detaljerat, exempelvis med avseende på ålder, höjd och krongräns. För att t.ex. volym, tillväxt eller ålder ska kunna beräknas för alla klavträd tilldelas (simuleras) provträdsegenskaper till samtliga klavträd.

I denna arbetsrapport beskriver Hans Åkesson och Berti Westerlund hur simuleringsfunktionerna i Riksskogstaxeringens nuvarande databas togs fram. Rapporten syftar både till att dokumentera förfaringssättet och ge värdefulla råd inför framtida revisioner.

I egenskap av projektledare för detta utvecklingsprojekt vill jag rikta ett stort tack till författarna för ett väl utfört arbete!

Umeå i april 2014

Per Nilsson

Sammanfattning

I den här arbetsrapporten beskrivs hur en revision av Riksskogstaxeringens beräkningar av volym, tillväxt och ålder för de inventerade träden kan göras. Samt en del tips och idéer som kan vara värdefulla. Viktigt är att resultatet har tillräckligt bra noggrannhet, litet systematiskt fel samt att volym, tillväxt och ålder harmoniserar. Simuleringsystemet består i huvudsak av två delar, en prediktionsdel där exempelvis klavträdet volym skattas med funktioner samt en simuleringsdel där de predikterade volymerna korrigeras mot provträden och en ”störning” adderas så att klavträdens volymer speglar alla träd volymer med avseende på väntevärde och spridning. För att resultatet ska bli bra är det av betydelse att hantverket med att ta fram prediktionsfunktionerna görs så att rätt ingående variabler väljs och hur de bör omformas samt hur funktionsindelade kategorier väljs. Det här arbetet tar minst 6 manmånader för två kompetenta personer med stor erfarenhet av Riksskogstaxeringens design, variabler, databaser samt verktyg för funktionsbyggande..

Nyckelord; simulering, trädvolym, tillväxt, Riksskogstaxering

Innehåll

Syfte	5
Introduktion	6
Generell modell för klavträdssimulering.....	7
Val av tillämpad simuleringsmetod.....	9
Några ord om design av funktionmodeller.....	9
Metod för harmoni mellan simulerade egenskaper	10
Prediktionsdelen av simuleringen	10
Val av funktionsindelade kategorier	10
Val av variabler	13
Min- och max-värden för prediktionsfunktionerna.....	17
Test av prediktionsfunktionernas väntevärde.....	18
Några tips för modelleringen av prediktionsfunktionerna	19
Se upp med träd av klava dimensioner.....	20
Beståndsdata som avser det huggningsklassbestämmande skiktet kan inte användas för träd som är överståndare, beståndsrest eller skärm	20
Inklavad grundyta och medeldiameter kan inte användas obetingat.....	20
En ökning av brösthöjdsdiametern får inte generera en skattad volymminskning.....	20
Skatta alltid den väderkorrigerade tillväxten	22
Beräkna ytavvikelsen i simuleringen enligt provträdens uppmätta väder-korrigerade tillväxt och deras skattade väderkorrigerade tillväxt.....	22
Observera att tillväxten skattas på två (2) olika sätt.....	22
Använd inte trädålder på andra ägoslag än produktiv skogsmark vid tillväxt-skattning .	22
Skatta inte stamvolymen efter den klavade stubben med hjälp av klavträdens formhöjdsskattningsfunktioner.....	22
Begränsa värderummet för klavstubbarnas skattade formhöjd	22
Simuleringsdelen av simuleringen	23
Min- och max-simuleringsvärden	23
Test av simuleringen	24
Test av simuleringens väntevärdesriktighet	24
Test av simuleringens spridning och harmoni.....	25
Litteratur.....	27
Bilaga	29

Syfte

I detta dokument har vi samlat de erfarenheter vi fått när vi utvecklat och uppdaterat Riksskogstaxeringens simulerade egenskaper på klavträd och klavstubbar. I Riksskogstaxeringen och även i andra inventeringar där träd klavas dvs trädets diameter mäts i brösthöjd med en klave tas det ut ett subsampel av träd som mäts på fler sätt förutom diameter. För att kunna använda dessa egenskaper använder vi i första läget funktioner. Uppdateringar av simuleringsfunktionerna är något som måste ske regelbundet med ett visst intervall. De kan vara bra att vid behov testa hur bra simuleringen fungerar jämfört med de senaste sju till tio årens provträd. Om nödvändigt bygger man då om funktionerna med hjälp av det mest aktuella provträden. Antingen behåller man de oberoende variablerna och byter till nya värden på koefficienterna eller om detta inte är tillfyllest måste man ta fram en ny uppsättning oberoende variabler.

Det ovan beskrivna arbetet har hittills i huvudsak utförts av dokumentets författare. Vi har lärt oss detta hantverk till stor del genom "trial and error". Därför är den här rapporten viktig för att de som övertar detta arbete inte går i alla fällor som vi till en början gjorde och att man förstår att vara så noggranna i detta arbete som vi förstätt det. Mycket tid måste anslås för att utföra en uppdatering.

För detta arbete behövs åtminstone två personer med nedan beskriven kompetens som vid en fullständig revidering kan avsätta totalt minst sex(6) manmånader.

En person måste vara kompetent och erfaren i statistisk modell byggande (regressionsanalys, nonlinear, multivariate mm) kunna verktygen och välja ut signifikanta variabler.

Den andra personen måste vara numerisk kunnig, ha stor kännedom om Riksskogstaxeringens databaser och databashanteraren.

Båda personerna måste vara insatta i hur Riksskogstaxeringens (RT) data samlats in, deras definitioner och hur de förändrats över tiden. Brister i detta kompetensområde är det som troligast leder till att resultatet inte blir så bra som man kan förvänta sig.

Att detta arbete sker med metoder som används vid skogsforskning, betyder inte att det är forskning. Det teoretiska arbetet bakom simuleringsmodellen är redan gjord och de som arbetar med detta ska huvudsakligen inrikta sig på leverans i form av sproc:ar och udf:er i RT-databasen - en skriftlig rapport av resultaten är sekundärt.

Angreppssättet kan inte heller vara analytiskt utan empiriskt. Ibland går det inte att motivera funktionerna utifrån en teoretisk modell. Är det så att variabeln $dbh^{2,3}$ vid en regressionsanalys ger en bättre prediktion, så använder vi den utan kunna motivera den utifrån en *á priori*-modell. Fungerar det, så fungerar det.

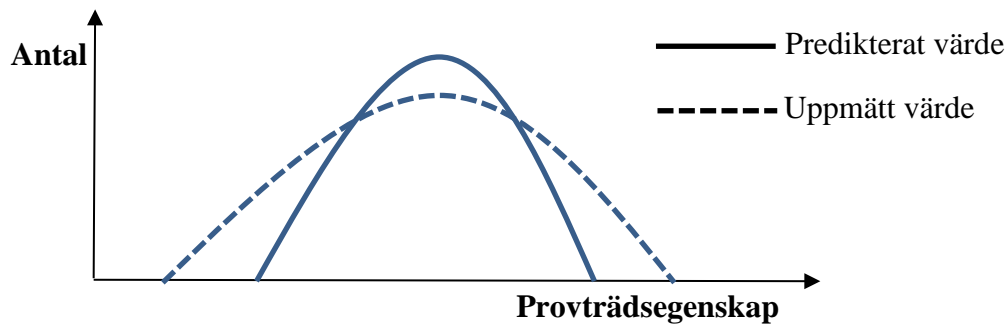
Mycket tid måste läggas ner på prediktioner utan bias, men man kan inte räkna med att alltid nå ett optimalt resultat. När det tar tid eller blir alltför komplext att nå vad man från början anser är tillräcklig noggrannhet, måste man analysera betydelsen av att leverera ett resultat med lägre precision. Man kanske finner att för t.ex. brösthöjdsålder för grova lönnar på myr, som är väldigt sällsynt så behöver man inte nå optimum.

Detta dokument kan också användas som underlag för implementering av simuleringen i form av programmeringskod. Av den anledning har vi valt att beskriva statistiska storheter i en mer explicit form än den som vanligen används. Se nedanstående exempel:

I stället för $Var(y)$ skriver vi $\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$

Introduktion

Från och med 1993 används simulering vid tilldelning av klavträdsegenskaper som är exklusivt uppmätta på provträd. Vid simulering nyttjas regressionsanalysens enkla komputationella förfarande vid tilldelning av värden till klavträden. Det är ett känt fenomen att regressionsfunktioner har en utjämnande effekt på sina utdata jämfört med de data använts för att skapa funktionerna. Prediktioner m.h.a. regressionsfunktioner har alltså en tendens att gå mot populationsmedelvärdet. Detta visas i figur 1.



Figur 1.

Den naturliga variationen är dock viktig för att korrekt redovisa skogstillstånd- Okorrigerad volymprediktion t.ex. ger upphov till en felaktig volymfördelning av virkesförrådet med bl.a. en underskattning av andelen stora träd, vilket i sin tur ger en felaktig bild av gallringsbehov, möjlig avverkning osv. **Avsikten med klavträdssimulering är inte begränsat till att prediktera provträdsegenskaper hos klavträden. Den syftar också till att återställa spridningen hos de predikerade trädsegenskaperna.**

Både väntevärden och spridning skattas med stöd av provträden. Provträden används först som underlag för att skapa prediktionsfunktioner genom t.ex. regressionsanalys. Sedan används dessa prediktionsfunktioner på samma sätt på provträds materialet. De residualer som då uppstår mellan predikerade och uppmätta värden kan sedan användas till att skatta den spridning som måste simuleras till prediktionen för att återställa den ursprungliga spridningen. Denna simulering bör ske så att den bibehåller den naturliga variationen mellan träd inom ytor och mellan ytor. Prediktioner och spridningssimulering kan sedan tillämpas på klavträden.

Om y_i är provträds i :s (av totalt n provträd på ytan) värde för en godtycklig provträdsegenskap som man avser att simulera, så är \hat{y}_i motsvarande predikerade och \tilde{y}_i motsvarande simulerade provträdsvärde. Om prediktion och simulering är väntevärdesriktiga så gäller:

$$\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \approx 0 \quad \text{och} \quad \frac{1}{n} \sum_{i=1}^n (\tilde{y}_i - y_i) \approx 0$$

Formel 1 och 2

Om simuleringen återställer den ursprungliga spridningen kring provytans medelvärde \bar{y} så gäller:

$$\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \approx \frac{1}{n-1} \sum_{i=1}^n (\tilde{y}_i - \bar{y})^2 \quad \text{och} \quad \frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \leq \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

Formel 3 och 4

Metoden består således av två delar:

1 Prediktionsdelen (Vektorvärda) funktioner f som predikterar värdet på önskade variabler t.ex. stamvolym och tillväxt.

$[\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n] = f(x_1, x_2, \dots, x_m)$ där,

\hat{y}_i är en prediktion av provträdsegenskap y_i $i = 1, 2, \dots, n$

x_j är en inparameter. Trädslag och brösthöjdsdiameter bör ingå i \mathbf{x} .

2 Simuleringsdelen Slumptal från slumptalsfunktioner, som genererar tal efter samma fördelning som provträdsresidualerna (*predikterat – faktiskt*). Dessa slumptal adderas till de predikterade värdena. Se formel 5!

Westerlund (2013) beskriver simuleringen som en metod att utnyttja regressionsanalysens relativa enkelhet men ändå behålla den naturliga spridningen i materialet. Vidare anser de att en bra klavträdssimulering utmärks av:

- Inga systematiska avvikelser i tilldelningen för enskilda provytor
- Naturlig variation mellan provytor
- Inga systematiska avvikelser i tilldelningen för enskilda klavträd
- Naturlig variation mellan klavträd inom yta
- Harmoni mellan olika tilldelade egenskaper för varje klavträd
- Metoden är lätt att använda

Generell modell för klavträdssimulering

Provyta i med klavträden $1, 2, \dots, n$ har en för alla träd gemensam variabeluppsättning \mathbf{P}_i och en variabeluppsättning för varje enskilt klavträd \mathbf{T}_{ij} . Trädegenskapen y kan skattas väntevärdesriktigt m.h.a. en funktion f vilken bestäms m.h.a. regressionsanalys eller annan jämförbar metod:

$\hat{y}_{ij} = f(\mathbf{P}_i, \mathbf{T}_{ij})$ Högerledet kan förkortas till f_{ij}

Variansen av y_{ij} beror av två slumpkomponenter:

1. En på trädnivå – mellan träd inom provyta. Den betecknas γ_{ij} .
2. En på ytnivå – mellan provytorna. Den betecknas β_i .

Egentligen bör också en komponent som svarar mot variansen mellan provyteklustren (trakter) också inkluderas, men detta bidrag är så litet att man kan bortse från det (Westerlund 2013).

Om f_{ij} är väntevärdesriktigt gäller också att slumpkomponenterna har väntevärde 0 ($E(\beta_i)=0 \wedge E(\gamma_{ij})=0$). Variansen för bägge komponenterna kan beskrivas som $\text{Var}(\beta_i)=\sigma_\beta^2$ och $\text{Var}(\gamma_{ij})=\sigma_\gamma^2$. Dessa varianser måste skattas innan klavträdssimuleringen kan slutföras. I första fallet m.h.a. \mathbf{P}_i eller en delmängd därav och i det senare m.h.a. \mathbf{T}_{ij} . Simulering av y kan uttryckas som:

$$\tilde{y}_{ij} = f_{ij} + \beta_i + \gamma_{ij} \text{ eller } \tilde{y}_{ij} = \hat{y}_{ij} + \beta_i + \gamma_{ij}$$

Formel 5

β_i och γ_{ij} antas vara normalfördelade med väntevärde 0 och standardavvikelse som svarar mot σ_β och σ_γ .

$$\beta_i \sim N(0, \sigma_\beta) \text{ och } \gamma_{ij} \sim N(0, \sigma_\gamma)$$

Observera att ett slumptal för β_i dras en gång per yta och slumptal för γ_{ij} dras för varje enskilt träd.

Formel 5 utnyttjar inte det faktum att de flesta provytor har provträd som en delmängd av klavträden. Då provträden har uppmätta värden för y behöver dessa träd inte simuleras. Dessutom kan vi utnyttja den extra information vi får av provträdens residualvektor \mathbf{z} , där $z_k = (y_{ik} - f_{ik})$ där $k=1,2,\dots,m_i$ för provyta i med m provträd. Vi kan med hjälp av \mathbf{z} förbättra prediktionen av y för de klavträd som samtidigt inte är provträd genom att beräkna en predikerad provyteavvikelse δ_i som adderas till prediktionen. Denna provyteavvikelse medför att σ_β^2 reduceras. Vi betecknar den nya ytslumpskomponenten β'_i och vi kan uttrycka denna utvidgning av formel 5 som:

$$\tilde{y}_{ij} = f_{ij} + \delta_i + \beta'_i + \gamma_{ij} \text{ eller } \tilde{y}_{ij} = \hat{y}_{ij} + \delta_i + \beta'_i + \gamma_{ij}$$

Formel 6

Vidare blir simuleringen klart bättre om σ_γ^2 skattas klassvis med klasser av trädslag, diameter m.fl. träd och ytegenskaper.

Westerlunds tillämpade univariata modell

Den tillämpade univariata modellen har varit i bruk i RT sedan 1993. I denna modell upprätthåller man harmoni mellan simulerade egenskaper genom att värdet på en redan simulerad egenskap ingår i variabeluppsättningen som används för att simulera nästa egenskap osv. tills alla egenskaper simulerats.

Westerlund (2013) simulerar klavträdsegenskaper enligt följande modell:

$$\begin{cases} \tilde{y}_{ij} = f_{ij} + \omega_i \bar{z}_i + \gamma_{ij} & \text{för provytor med provträd} \\ \tilde{y}_{ij} = f_{ij} + \hat{\beta}_i + \gamma_{ij} & \text{för provytor utan provträd} \end{cases}$$

Formel 7

Förklaring följer nedan:

- Yta i har m provträd. Ett godtyckligt klavträd på yta i betecknas j och ett provträd k .
- $\hat{\beta}_i$ är ett normal fördelat slumpstal $\sim N(0, \hat{\sigma}_\beta)$.
- $\hat{\sigma}_\beta^2$ är i denna modell en konstant beräknad ur Westerlunds (2013) testmaterial om c:a 5 000 provträd.
- $\bar{z}_i = \frac{1}{m_i} \sum_{k=1}^{m_i} (y_{ik} - f_{ik})$ Höga och låga värden av $(y_{ik} - f_{ik})$ trunkeras så $|y_{ik} - f_{ik}| \leq 2\sigma_\beta$
- $\hat{\gamma}_{ij}$ är ett normalfördelat slumpstal $\sim N(0, \hat{\sigma}_\gamma)$. $\hat{\sigma}_\gamma^2$ beräknas av V_{ij} , som är en funktion för den skattade variansen för γ_{ij} , med trädslag och diameter som oberoende variabler. V_{ij} är ur testmaterialet beräknad med styckvis lineär regression.
- $\omega_i = \sqrt{\frac{\sigma_\beta^2}{\text{Var}(\bar{z}_i)}} = \sqrt{\frac{\sigma_\beta^2}{\sigma_\beta^2 + \frac{1}{m_i^2} \sum_{k=1}^{m_i} V_{ik}}}$ ty $\text{Var}(\bar{z}_i) = \sigma_\beta^2 + \frac{1}{m_i^2} \sum_{k=1}^{m_i} V_{ik}$

I denna modell finns inte någon slumpkomponent med reducerad ytvis varians β'_i enligt formel 6. Westerlund (2013) fann att β'_i blev i praktiken nära noll om antalet provträd var två eller fler på en yta. Felet som uppstod om denna term utelämnades även i fallet med ett (1) provträd var också litet, så man valde att bortse från denna term och förenkla modellen.

Val av tillämpad simuleringsmetod

I dagens läge tillämpas bara Westerlund (2013). Sören Holm utredde dock 1999 möjligheten av en multivariat modell, som var multivariat i både prediktions och simuleringsdelen. Den skulle säkra harmonin mellan simulerade egenskaper på ett bättre sätt än Westerlund (2013) och göra arbetet med att prediktera enklare. Modellen är å andra sidan mer komplex sin grundläggande struktur innefattande ett stort antal matrisoperationer. Det kan dock vara värt att undersöka detta angreppssätt i framtiden.

Några ord om design av funktionmodeller

För att välja hur prediktionsfunktionerna ska se ut och vilka variabler som ska ingå måste man förstå vad som bestämmer volym och tillväxt. Ett trädets volym är en produkt av trädets grundytan i brösthöjd (g), höjd (h) och formtal (f).

$$volym = g \cdot f \cdot h$$

Grundytan (g) beräknas som

$$g = \pi \cdot dbh^2 / 4$$

Eftersom diametern i brösthöjd (dbh) mäts på alla ingående (klav)träd så är grundytet faktorn redan klar. Trädhöjd \times formtal är höjden på den cylinder som motsvarar trädets volym med grundytan vid brösthöjd benämnd formhöjd (fh). Därför är funktioner som skattar formhöjd effektiva vid beräkning av trädvolym. Här går det ofta bra att använda additiva modeller.

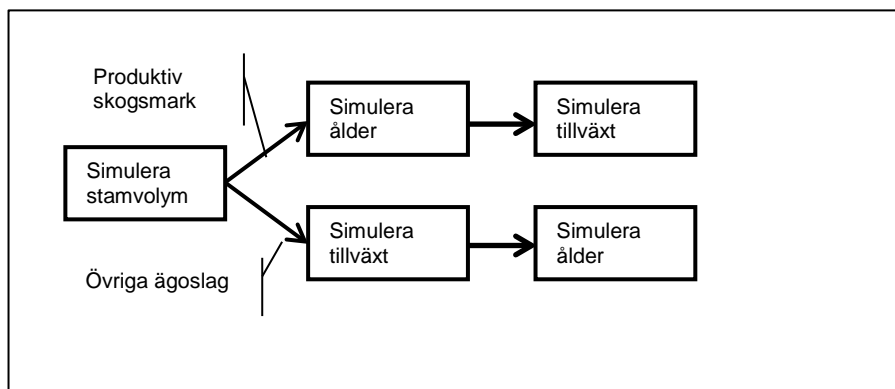
Vid prediktion av tillväxt är det viktigt att förutom trädets storlek få med tillväxtbestämmande ståndortsfaktorer. Här kan man gärna använda multiplikativa eller icke linjära modeller.

Viktiga ingående variabler i modellerna är diameter, medelhöjd, beståndsålder samt ståndortsbeskrivande variabler. Testa alla tillgängliga variabler i början med något stepwise verktyg. Studera residualplottarna för att se om några transformerade variabler, exv. Dbh^2 behövs. Testa alla signifikanta transformerade variabler också. Var försiktig med indikatorvariabler för trädslag, ägoslag och/eller huggningsklass etc. Om koefficienten för indikatorvariabeln är större än interceptet så bör man välja en annan modell.

Funktionerna bör vara trädslagsvisa för de vanligaste trädslagen samt i möjligaste mån ägoslagsvisa.

Metod för harmoni mellan simulerade egenskaper

Harmonisering mellan de simulerade egenskaperna stamvolym, brösthöjdsålder och femårig tillväxt säkerställs genom att låta en simulerad egenskap ingå som en oberoende variabel i skattningen av en annan egenskap. Den sekvens för klavträden i vilken vi simulerar egenskaper genom att låta det föregående simuleringsvärdet ingå i simuleringens skattningsdel framgår av nedanstående skiss.



Figur 2.

Prediktionsdelen av simuleringen

Stamvolymen på bark samt den avsatta och väderkorrigerade femåriga tillväxten ($m^3sk/5$ år) simuleras för tillfälliga klavträd som inte är provträd med borrprov eller permanent klavträd som inte är återinventerat provträd (s.k. återprovträd).

Den avsatta och väderkorrigerade femåriga tillväxten ($m^3sk/5$ år) simuleras för tillfälliga klavträd som inte är provträd med mätta borrprov eller permanent klavträd som inte är återprovträd.

GY är trädets grundyta dvs. $GY = \pi \cdot 1/4 \cdot 10^{-6} \cdot bhd^2$, om bhd är brösthöjdsdiameter i mm.

$$\text{Stamvolym} \quad y = GY \cdot (a + bx_1 + cx_2 + dx_3 + \dots)$$

$$\text{Tillväxt} \quad y = e^{a + bx_1 + cx_2 + dx_3 + \dots}$$

$$\text{Brhålder} \quad y = e^{a + bx_1 + cx_2 + dx_3 + \dots}$$

$$\text{Stubbvolym} \quad y = GY \cdot (a + bx_1 + cx_2 + dx_3 + \dots)$$

Val av funktionsindelade kategorier

Trädslag

Hur man väljer vilka trädslagskategorier som ska användas för att dela upp predikteringsfunktionerna måste man empiriskt testa fram. Om man väljer en alltför fin uppdelning riskerar man överanpassning på grund av för få träningsfall. I de fall man har få provträd att bygga funktionen med, är det bättre att ha breda trädslagskategorier och använda avvikande trädslag som indikatorvariabler.

Nedan redovisas den indelning som för närvarande tillämpas:

STAMVOLYM	Produktiv skogsmark:	Tall exkl. contorta, contorta, gran, björk, asp, al, ek, bok, övrigt ädellöv & övrigt löv
	Ej produktiv skogsmark:	Tall inkl. contorta, gran, björk, ädelt löv, al & övrigt löv
TILLVÄXT	Produktiv skogsmark:	Tall exkl. contorta,, contorta, gran, björk, ek bok, övrigt ädellöv & övrigt löv
	Ej produktiv skogsmark:	Tall inkl. contorta, gran och löv
BRHÅLDER	Produktiv skogsmark:	Tall inkl. contorta, björk & övrigt löv
	Ej produktiv skogsmark:	Barrträd & lövträd
STUBBVOLYM	Produktiv skogsmark:	Tall inkl. contorta, gran, björk & övrigt löv
	Ej produktiv skogsmark:	Tall inkl. contorta, gran, björk & övrigt löv

Diameterintervall

Uppdelning i diameterintervall är den mest tidsödande av alla indelade kategorier. För varje förslag till prediktering måste man studera residualplottar med diametern på x-axeln för att avgöra inom vilket diameterintervall som just den predikteringen fungerar. Nya indelningar av diameterintervall och regression inom föreslagna intervall måste tas fram och testas ända tills man når ett godtagbart resultat.

Man kan inte i förväg avgöra vilka diameterintervall som är optimala. Men några tumregler finns:

- **Det lägsta diameterintervallet måste vara smalt.** Provträdens stamvolym har en diskontinuitet i det att många trädsdrag beräknas genom S O Anderssons volymfunktioner för bhd 40 – 49 mm och byter till funktioner enligt Näslund, Eriksson, Matern & Hagberg m.fl. vid bhd \geq 50 mm. För tillväxt beräknad på provträd med maskinmätta borrhov tillkommer att beräkningar av barktjocklek och barktillväxt har en systematisk avvikelse vid små diametrar.
- **Det finns en punkt i de högre diameterintervallen där formhöjden börjar generera negativ volymtillväxt vid ökad diameter.** Detta fenomen beskrivs längre fram.

Ägoslag

Det är självklart att produktiv skogsmark alltid måste predikteras särskilt, eftersom det i RT är det viktigaste ägoslaget och alltid har mer information tillgänglig för prediktering. Impedimentägoslagen skiljer sig kraftigt från andra ägoslag med avseende på alla egenskaper man vill simulera, här är materialet mindre än för produktiv skogsmark. Om impedimenten ska uppdelas ännu finare är något som man måste testa sig fram till. Avverkning på icke produktiv skogsmark är så sällsynt att där finns ingen uppdelning för icke-skogsmarksägoslagen.

Nedan redovisas den indelning som för närvarande tillämpas:

Stamvolym	Produktiv skogsmark, naturbete, åker/väg/kraftledning/annan mark, myr & berg/fjällbarrskog.
Tillväxt	Produktiv skogsmark, naturbete/ åker/väg/kraftledning/annan mark, myr, berg & fjällbarrskog.
Brhålder	Produktiv skogsmark, naturbete/ åker/väg/kraftledning/annan mark, myr, berg & fjällbarrskog.
Stubbvolym	Produktiv skogsmark & naturbete/ åker/väg/kraftledning/annan mark/myr/berg/fjällbarrskog.

Medelhöjd över eller under 7 meter

Beståndsmedelhöjd gäller endast för det huggningsklassbestämmande skiktet. Om man inte tar hänsyn till detta vid prediktering för träd som inte tillhör det huggningsklassbestämmande skiktet, blir det en dålig skattning med systematisk avvikelse. Angreppssättet hittills har varit att för varje träslag dela upp skattningsfunktionerna på bestånd över och under sju meter. Man kan då på goda grunder anta att grova träd på bestånd med en medelhöjd < 7 meter troligen är överståndare och regressionen är här utförd på sådana provträd.

För närvarande används den ovanstående indelningen för predikering av stamvolym och ålder.

Huggningsart

Skattningsfunktionerna för klavstubbarnas volym skall uppdelas i olika huggningsarter och provträd till träningsmängden väljs så att passar så bra som möjligt till huggningsartens trädpopulation. Följer man inte denna princip blir skattningen systematisk avvikande. Till de formhöjdsfunktioner för klavstubbarna som utvecklades 2011 användes följande indelning i av klavstubbarnas formhöjds-skattning och urval av provträd att skatta på.

Skattningsfunktion av klavstubbbar	tränas på provträd
slutavverkade på produktiv skogsmark	i huggningsklass D
fröträdsavverkade på produktiv skogsmark	i huggningsklass A-B1 med bhd \geq 100 mm
övriga avverkade på produktiv skogsmark	i huggningsklass B2-C med bhd \geq 40 mm i huggningsklass A-B1 med bhd < 100 mm
andra ägoslag än produktiv skogsmark	andra ägoslag än produktiv skogsmark

Förrådsyta eller mellanliggande

När det gäller prediktering av klavstubbarnas volym finns relevant provyte- och bestånds-information på avverkade (tillfälliga) förrådsytor, men saknas på mellanliggande stubbprovytor som inte är förrådsprovytor. På avverkade förrådsprovytor har man t.ex. tillgång till både nuvarande bestånds grundyta och den avverkade grundytan, vilka adderade tillsammans ger grundytan vid avverkningstillfället.

Så följande uppdelning av skattningen utöver de vanliga i träslag, diameterintervall mm kan göras:

Ägoslag	Yttyp	Medelhöjd
Produktiv skogsmark	Förrådsyta	—
	Mellanliggande	\geq 7 m
< 7 m		
Andra ägoslag	—	—

Avverkad åldersintervall

Klavstubbbar efter tallar som slutavverkats i diameterintervallet 110 – 399 mm gav systematiska fel vid skattningar av volym om inte de delades upp i provytor med medelålder över och under 78 år. Detsamma gällde också tallar som slutavverkats i diameterintervall \geq 400 mm. De delades upp i provytor med medelålder över och under 70 år.

Ståndortsindex för gran

Klavstubbar efter granar som slutavverkats i diameterintervallet ≥ 400 mm gav dåliga skattningar av volym annat än om de delades upp i fallen ståndortsindex för gran över och under 24 m.

Val av variabler

Höjd över havet och breddgrad

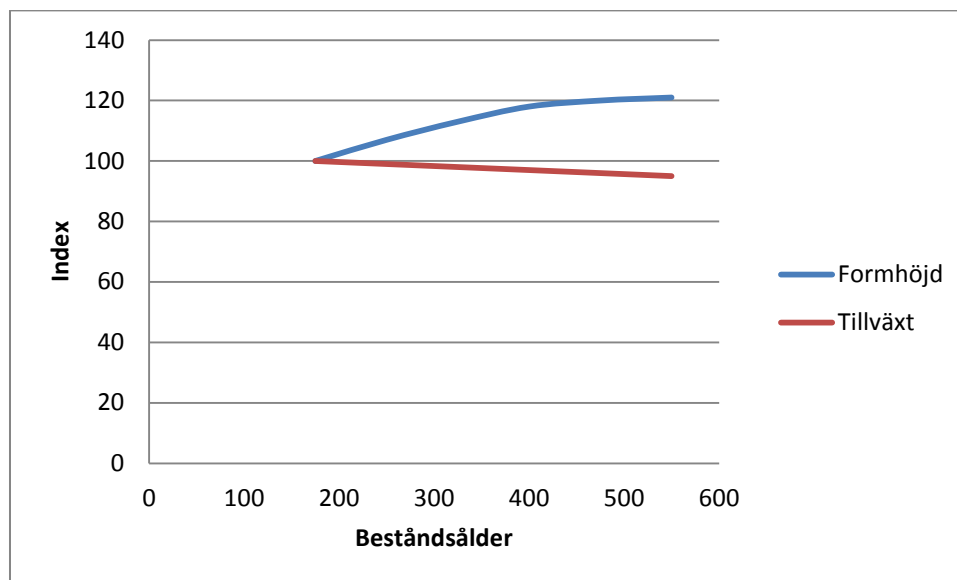
Höjd över havet i meter och nordlig breddgrad ingår i nästan alla predikteringar. De ingår i variabeluppsättningen både för sig eller som produkt (Altitud×Latitud) och är en mycket viktig variabel för provytans klimat.

Beståndsålder

För klavträdsvolym, tillväxt och trädålder på produktiv skogsmark är beståndsålder en viktig predikterande variabel. Den har dock varit föremål för ändringar i definition och datafångst. Därför ska provträd från provytor som där beståndsåldern är definierad på samma sätt som de klavträd som skall predikteras användas. Beståndsåldern avser också enbart det huggningsklassbestämmande skiktet och kan inte användas på överståndare, beståndsrester och träd i skärmar. Här måste funktioner utan beståndsålder användas.

Beståndsåldern har ställt till med problem när man ändrat definitionen av den, 1993. Tidigare hade vi ett takvärde för beståndsålder på 175 år. När det taket togs bort så skuttade prediktionen av stamvolym uppåt för träd på ytor med hög beståndsålder. Som man ser påverkas formhöjden/volymen kraftigt. När en inventeringsförändring som att ta bort ett maxvärde införs måste man hårdkoda ett takvärde, vänta i 5-6 år får att få ihop en tillräcklig mängd provträd och sedan utveckla nya predikteringsfunktioner.

Nedanstående diagram beskriver fenomenet:



Figur 3. Diagrammet visar hur en prediktering av ett klavträd i ett bestånd med en medelålder på 175 år förändras om medelåldern ändras till en högre ålder. Formhöjd och volymtillväxt är satt till ett index = 100 vid en beståndsålder på 175.

Vid prediktering av stamvolym har även inverterad beståndsålder använts för grova aspar på produktiv skogsmark för att prediktionen blev bättre.

Det finns ett mindre antal fall (<1 promille) där beståndsåldern saknas och då tillämpas följande empiriska regler:

- För prediktering av klavträdsvolym sätts beståndsåldern till 0 år om den är odefinierad.
- För prediktering av tillväxt sätts beståndsåldern till 4 år om den är odefinierad.
- För prediktering av ålder på produktiv skogsmark sätts beståndsåldern till 0 år om den är odefinierad.
- För prediktering av ålder på andra ägoslag än produktiv skogsmark sätts beståndsåldern till 12 år om den är odefinierad eller under 12 år.
- För prediktering av stubbvolym sätts odefinierad beståndsålder till 54 vid diverseavverkning, till 63 år vid avverkning av övrigt skikt och till 0 år i övriga fall.

Medelhöjd

Medelhöjden är en bra predikerande egenskap, men har sina fallor i användningen. På produktiv skogsmark avser den enbart huggningsklassbestämmande skiktet. Den mäts olika över och under sju meter. Dessutom har datafångst och definition ändrats sedan 1983, särskilt gäller detta för andra ägoslag än produktiv skogsmark.

Det finns ett antal fall där medelhöjden är odefinierad och då tillämpas följande empiriska regler:

- För prediktering av klavträdsvolym sätts medelhöjden till 0 meter på produktiv skogsmark och till 2.5 meter på andra ägoslag än produktiv skogsmark om medelhöjden är odefinierad.
- För prediktering av tillväxt sätts medelhöjden till 1.3 meter den är odefinierad.
- För prediktering av brösthöjdsålder sätts medelhöjden till 0 meter den är odefinierad.

Bonitet

Bonitet för bonitetsvisande trädslag används vid prediktering av väderkorrigerad femårig tillväxt. Tillväxten predikteras med en funktion av typen $\ln(y) = a + bx_1 + cx_2 + dx_3 + \dots$ och i den variabeluppsättning där förekommer boniteten alltid som $e \cdot \text{bonitet} + f \cdot \ln(\text{bonitet})$.

Grundyta

Grundyta är en viktig variabel vid prediktering av alla provträdsegenskaper. Men det är olika definitioner och fraktioner av grundyta som används:

Fältbedömd

Används vid prediktering av klavträdsvolym och klavträdsålder på produktiv skogsmark. Vid prediktering av klavstubbars volym används den tillsammans den avverkade grundytan som en skattning av grundytan före avverkning. På mellanliggande ytor skattas grundytan till 2 m^2 vid slutavverkning och fröträdsavverkning. Vid gallring till $2.146 \times \text{GrundytaAvverkad} + 4.6585$ och till $2.1178 \times \text{GrundytaAvverkad} + 4.6585$ för övriga avverkningar.

Inklavade klavträd alla

Används vid prediktering av volymtillväxt. Om inklavad grundyta $> 100 \text{ m}^2$ på produktiv skogsmark, naturbete, myr, berg och fjällbarrskog sätts grundytan till 100 m^2 . Om inklavad grundyta $> 50 \text{ m}^2$ på åker, väg, kraftledning och annan mark sätts grundytan till 50 m^2 .

Inklavade klavträd bhd $\geq 100 \text{ mm}$

Används vid prediktering av klavträdsvolym. Om inga träd med bhd $\geq 100 \text{ mm}$ klavats sätts denna grundyta till 0 m^2 .

Avverkad grundyta Används vid prediktering av klavstubbsvolym. Om inklavad grundyta $> 99 \text{ m}^2$ sätts den avverkade grundytan till 99 m^2 .

Medeldiameter

Medeldiameter avser alltid grundtytevägd medeldiameter beräknad på de inklavade träden. Det är olika definitioner och fraktioner av medeldiameter som används:

Inklavade klavträd bhd $\geq 100 \text{ mm}$ Används vid prediktering av klavträdsvolym, brösthöjdsålder och klavstubbsvolym. Om inga träd med $\text{bhd} \geq 100 \text{ mm}$ klavats sätts medeldiametern till 100 mm .

Avverkad grundyta Används vid prediktering av klavstubbsvolym.

Ståndortsindex för tall och för gran

Ståndortsindex för **tall** på produktiv skogsmark och naturbete används vid prediktering av:

Klavträdsvolym	Tall, björk, asp samt ädellöv exklusive ek och bok.
Brhålder	Tall och björk.
Stubbvolym	Tall

Ståndortsindex för **gran** på produktiv skogsmark och naturbete används vid prediktering av:

Klavträdsvolym	Gran, al, oädelt löv exklusive björk och asp, ek samt bok
Brhålder	Gran och löv.
Stubbvolym	Gran och löv.

Medelålder Avverkad

Medelålder (år) för avverkade träd i beståndet används för prediktering av klavstubbsvolym på produktiv skogsmark. Åldern för de avverkade träden sätts inte för alla huggningsarter, men kan ändå ingå i en prediktionsfunktion om den inte är alltför signifikant. Om ålder saknas sätts åldern till följande beroende på huggningsart, empiriska värden grundade på testmaterialet:

Mekanisk ungsogsgröjning	18 år
Underröjning i äldre skog	51 år
Diverseavverkning	57 år
Fröträdsavverkning	126 år
Avverkning av övrigt skikt	50 år
Mekanisk röjning och fröträdsavverkning	42 år
Mekanisk hyggesrensning och fröträdsavverkning	98 år

Diameter

Brösthöjdsdiameter förekommer nästan undantagslöst i alla predikterande funktioner, dock i olika fraktioner.

bhd	Enkelt diametervärde förekommer i nästan alla funktioner
bhd^2	Förekommer för klavträdsvolym, stubbvolym
bhd^3	Förekommer för klavträdsvolym
$\ln(\text{bhd})$	Förekommer för stubbvolym
$\text{bhd}^{2.5}$	Förekommer för stubbvolym
$\text{bhd}^{0.3}$	Förekommer för stubbvolym

Asp

Indikatorvariabel för trädslaget asp. Används när provträden som används i regressionen är få och trädslagskategorin är mycket vid (t.ex. kategorin "Lövträd"). Förekommer i funktioner för stubbvolym.

Al

Indikatorvariabel för trädslagen gråal och klibball. Används när provträden som används i regressionen är få och trädslagskategorin mycket vid (t.ex. kategorin "Lövträd"). Förekommer i funktioner för klavträdsvolym och stubbvolym

Ek

Indikatorvariabel för trädslaget ek. Används när provträden som används i regressionen är få och trädslagskategorin mycket vid (t.ex. kategorin "Lövträd"). Förekommer i funktioner för brhålder och stubbvolym.

Bok

Indikatorvariabel för trädslaget bok. Används när provträden som används i regressionen är få och trädslagskategorin mycket vid (t.ex. kategorin "Lövträd"). Förekommer i funktioner för stubbvolym.

Contorta

Indikatorvariabel för trädslaget contortatall. Används när provträden som används i regressionen är få och trädslagskategorin mycket vid (t.ex. kategorin "Barrträdslag"). Förekommer i funktioner för brhålder och stubbvolym.

Bete

Indikatorvariabel för ägoslaget naturbete. Används när provträden som används i regressionen är få och ägoslagskategorin mycket vid (t.ex. kategorin "andra ägoslag än produktiv skogsmark"). Förekommer i funktioner för klavträdsvolym.

Åker

Indikatorvariabel för ägoslaget åker. Används när provträden som används i regressionen är få och ägoslagskategorin mycket vid (t.ex. kategorin "andra ägoslag än produktiv skogsmark"). Förekommer i funktioner för klavträdsvolym.

Berg

Indikatorvariabel för ägoslaget berg. Används när provträden som används i regressionen är få och ägoslagskategorin mycket vid (t.ex. kategorin "andra ägoslag än produktiv skogsmark"). Förekommer i funktioner för klavstubbsvolym.

Kalmark

Indikatorvariabel för huggningsklass kalmark A. Används när provträden som används i regressionen är få och huggningsklasskategorin mycket vid (t.ex. kategorin "A-C"). Förekommer i funktioner för klavstubbsvolym.

Gallringsskog

Indikatorvariabel för huggningsklass C (gallringsskog). Används när provträden som används i regressionen är få och huggningsklasskategorin mycket vid (t.ex. kategorin "A-C"). Förekommer i funktioner för klavstubbsvolym.

Stamvolym

Simulerad stamvolymsvärde ingår i variabeluppsättningen för funktioner som predikterar tillväxt och är betydelsefull för harmonisering.

Formhöjd

Formhöjd baserad på simulerad stamvolym förekommer i funktioner som predikterar tillväxt och brhålder.

Stamvolym per år

Simulerad stamvolym dividerat med simulerat antal år i brösthöjd används för att prediktera tillväxt på produktiv skogsmark för klavträd med låg (< 9 år) brösthöjdsålder.

Tillväxtkvot

Kvoten mellan simulerad femårig tillväxt och simulerad stamvolym används för att prediktera brösthöjdsålder på andra ägoslag än produktiv skogsmark.

Kvoten mellan diameter och medeldiameter

Kvoten mellan brösthöjdsdiameter och inklavad medeldiameter för träd med bhd \geq 100 mm används för att prediktera klavträdsvolym och brösthöjdsålder på andra ägoslag än produktiv skogsmark.

Brösthöjdsålder

Simulerad brösthöjdsålder har använts för att prediktera tillväxt på produktiv skogsmark.

Min- och max-värden för prediktionsfunktionerna

Regression med många oberoende variabler med alltid en risk för "outliers" med onaturligt stora och små värden.

För prediktering av klavträdens och klavstubbarnas formhöjder har vi valt att empirisk bestämma ett värdeintervall som omfattar 99,95% av klavträdens formhöjder. Rent teoretisk kan formhöjden inte understiga 1.3 meter, men för träd med bhd \geq 40 mm ligger minivärdena något högre än så. Nedan listas de min och maxvärden som tillämpas.

Min och maxvärden för formhöjd (m)		
Trädslag	Min	Max
Tall (Pinus sylvestris)	2,35	19,00
Bergtall	1,66	19,00
Lärk	2,35	19,00
Övrig tall	2,35	19,00
Gran (Picea abies)	2,25	19,00
Övriga Picea-arter	2,25	19,00
Abies-arter	2,25	19,00
Övriga barrträd	2,25	19,00
Främmande granar	2,25	19,00
Björkar	1,70	19,00
Värtbjörk	1,90	19,00
Glasbjörk	1,70	19,00
Asp	1,88	19,00
Ek	1,61	19,00
Bok	2,37	19,00
Ask	1,67	19,00
Alm	2,29	19,00
Lind	2,49	19,00
Lönn	1,99	19,00
Avenbok	2,49	19,00
Fågelbär	2,28	19,00
Contortatall	2,14	19,00
Klibbal	2,12	19,00
Gråal	1,77	19,00
Sykomorlönn	1,84	19,00
Sälg	1,48	19,00
Rönn	1,74	19,00
Icke ädla löv exklusive björk, klibbal, gråal, sykomorlönn, sälg och rönn	1,67	19,00
Oxel	1,84	19,00

För prediktering av brösthöjdsålder tillämpas ett minvärde på 3.5 år.

För väderkorrigerad femårig volymtillväxt (I) tillämpas följande:

Tillfälliga klavträd och permanenta nya klavträd

Låt följande beteckningar gälla

I Väderkorrigerad femårig volymtillväxt ($m^3sk/5$ år)

V Stamvolym (m^3sk)

För ett predikerat värde \hat{I} tillämpas följande minvärden, dvs. om \hat{I} understiger minvärdet sätts I till minvärdet:

$$\min(I) = \begin{cases} 0.0 & \text{Om } \hat{I} \leq 0.0 \\ 0.3476V^{0.696} & \text{Om } \hat{I} > 0.0 \end{cases}$$

Permanent återträd

Tillväxten på återträden predikteras som volymskillnad enligt formhöjdsfunktionerna, men här gäller en begränsning av hur stor skillnaden mellan nuvarande diameter och förutvarande får vara.

Låt följande beteckningar gälla:

bhd_0 Nuvarande brösthöjdsdiameter (mm)

bhd_5 Förutvarande brösthöjdsdiameter (mm) dvs. för fem år sedan.

$FH(bhd, \dots)$ Formhöjdsfunktion (m)

Diameterdifferensen måste ligga i intervallet $[-14, 65]$, vilket innebär att

om $bhd_0 - bhd_5 < -14$ $bhd_5 = bhd_0 + 14$

om $bhd_0 - bhd_5 > 65$ $bhd_5 = bhd_0 - 65$

och volymtillväxten kan nu skattas till

$$\hat{I} = \frac{\pi}{4} 10^{-6} (FH(bhd_0^2, \dots) - FH(bhd_5^2, \dots))$$

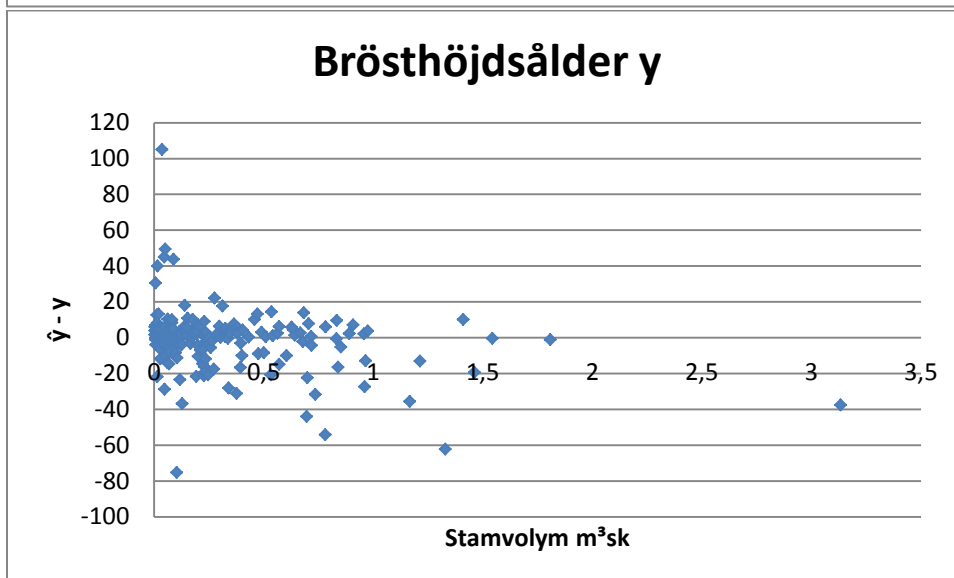
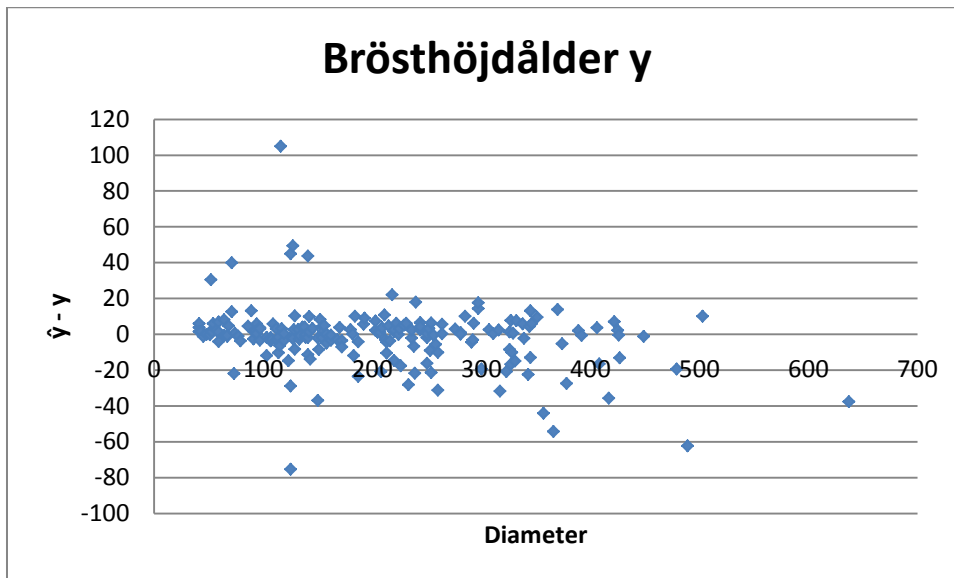
Test av prediktionsfunktionernas väntevärde

RMSE och förklaringsgrad för kombinationer av olika kategorivariabler används vid test av var man behöver förbättra predikteringen. Värdena hämtas från regressionsanalysen eller görs i efterhand på träningsmaterialet (provträden som ingår i analysen).

	Tall	Gran	Björk	Ädellöv	Övrigt
Hukl	1.61	1.69	1.11	1.44	1.89
A-B1	85%	86%	81%	81%	79%
Hukl	0.87	0.93	0.99	1.22	1.47
B2-C	84%	88%	87%	87%	87%
Hukl	0.80	0.89	0.89	1.01	1.22
D	91%	88%	87%	87%	87%
Imped	1.70	1.77	2.01	2.10	1.83
	66%	67%	66%	66%	66%
Övriga	1.31	1.33	0.99	1.01	1.55
	72%	78%	75%	75%	75%

En tabell som ovan används för att analysera vilka kategorindelningar som prediktering bör delas upp i. Exemplet ovan är lite väl grovt, enbart uppdelad i kategorier av huggningsklass, ägoslag och trädslag, för en sådan analys men är belysande. Uppstyckningen av predikteringen i kategorier och intervall är viktig och måste ge väntevärdesriktiga resultat särskilt för de kategorier som användarna bedömer som viktiga. Generellt måste man lägga mer tid på ägoslaget produktiv skogsmark än för övriga ägoslag.

Plottar används flitigt för att studera överensstämmelse mellan provträdsvärden och predikteringen. Varje förslag till predikteringsfunktion och varje stycke i den funktionen bör testas mot medelfel och plottar. För stor avvikelse i medelfel eller konstigheter i plottarna medför att funktionerna måste revideras.



Exempel på värdefulla plottar. I detta fall provträdens brösthöjdsålder som en residual mellan predikerad ålder och mätt (fält eller maskin) ålder för lövträd på andra ägoslag än produktiv skogsmark. Plottar tillsammans med uppgifter som medelfel på -1.49 år och $s^2 = 1.41$ används för att godta eller förkasta predikteringsfunktioner. Om en prediktering förkastas får man prova med fler ingående variabler eller att stycka upp funktionen på t.ex. olika diameterintervall. I exemplet ovan har vi få provträd och egenskapen brösthöjdsålder är inte viktig på andra ägoslag än produktiv skogsmark och därför duger predikteringsfunktionen.

Några tips för modelleringen av prediktionsfunktionerna

Nedan presenteras några fällor som finns vid modellerandet och några goda råd om hur man undviker dem. De grundas på tjugo års erfarenhet av att arbeta med simuleringsmodeller i Riksskogstaxeringen.

Se upp med träd av klena dimensioner

När man skattar formhöjd/volym för klavträd skattar man egentligen formhöjd/volym enligt Näslund, Eriksson, Matern & Hagberg m fl. Det är dessa volymfunktioner som används för provträden, vilkas volym vi försöker optimera våra skattningsfunktioner mot.

Emellertid finns här en diskontinuitet då vissa trädslag använder S O Anderssons småträdsfunktioner för träd under 50 mm i diameter. Vid övergången från 49 mm till 50 mm i diameter kan skillnaden i volym-skattning bli relativt stort. Då vi bygger skattningsfunktionerna med provträden som träningsmängd måste noggranna analyser göras av skattning kring 50 mm. Antagligen måste man för de aktuella trädslagen alltid ha separata skattningsfunktioner för ett smalt lägsta diameterintervall.

Beståndsdata som avser det huggningsklassbestämmande skiktet kan inte användas för träd som är överståndare, beståndsrest eller skärm

Beståndsmedelhöjd, beståndsålder och trädslagsblandning är beståndsdata som avser endast det huggningsklassbestämda skiktet. Om man inte tar hänsyn till detta vid skattning av formhöjd/volym för träd som inte tillhör det huggningsklassbestämmande skiktet, kommer man att göra en dålig skattning med systematisk avvikelse.

Man skulle kunna utelämna beståndsmedelhöjd, beståndsålder och trädslagsblandning vid formhöjdsskattning, men särskilt beståndsmedelhöjden är ytterligt signifikant för träd som tillhör det huggningsklassbestämmande skiktet.

Angreppssättet hittills har varit att för varje trädslag dela upp skattningsfunktionerna på bestånd över och under sju meter och för olika diameterintervall. För bestånd under 7 meter använder man inte beståndsmedelhöjd, beståndsålder och trädslagsblandning. Diameterintervallen måste man empiriskt bestämma genom att testa sig fram. Om man gör denna uppdelning på bästa sätt så kommer ett grovt träd i ett bestånd under 7 meter inte beräknas med hjälp av beståndsmedelhöjd, beståndsålder och trädslagsblandning vid skattning av formhöjd. Ett sådant träd är med största säkerhet en överståndare och tillhör således inte det huggningsbestämmande skiktet.

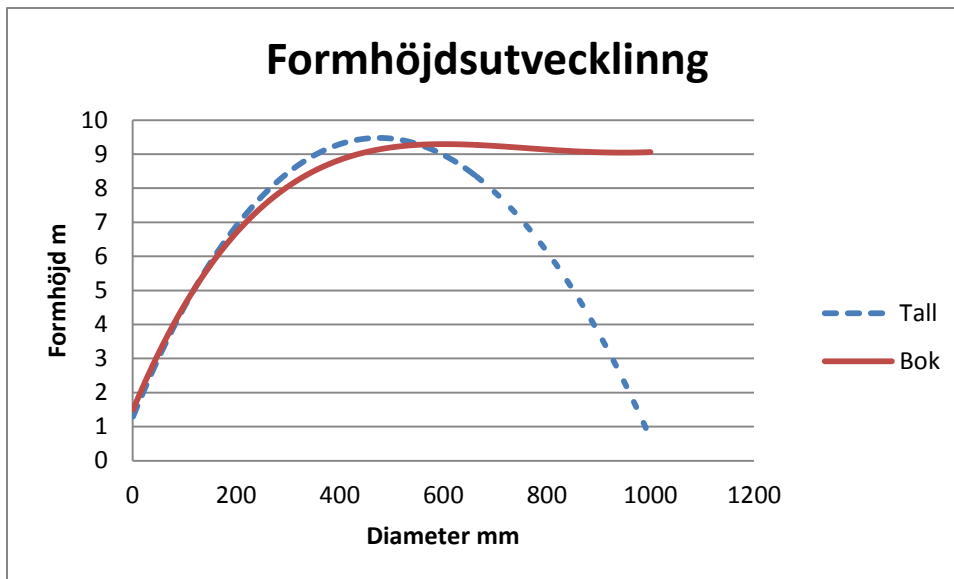
Om man delar upp skattningsfunktionerna på över och under 7 meters medelhöjd undviker man dessutom att använda värden som inte mäts eller skattas i fält vid medelhöjder under 7 meter. Den fältmätta grundytan t.ex. registreras inte om beståndsmedelhöjden understiger sju meter och därför är den inklavade grundytan ett bättre alternativ som oberoende variabel vid en formhöjdsskattning.

Inklavad grundyta och medeldiameter kan inte användas obetingat.

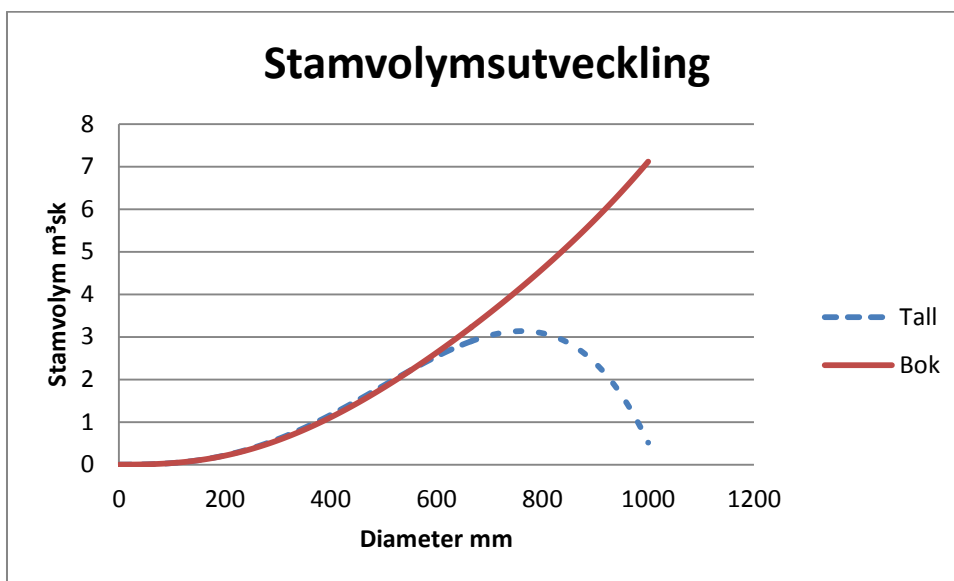
Inklavad grundyta och grundtyvägd medeldiameter är oberoende av beståndsmedelhöjd och signifikant vid skattning av formhöjd. Om grundytan och medeldiametern begränsas till klavträd med bhd ≥ 100 mm blir skattningen oftast bättre än om alla klavträd ingår. Detta har troligen att göra med hur delning av provytor fungerar och att antalet ”outliers” begränsas. Om inga klavträd med bhd ≥ 100 mm finns på ytan sätts grundytan till 0 m²/ha och medeldiametern till 100 mm.

En ökning av brösthöjdsdiametern får inte generera en skattad volymminskning

När träd blir äldre och grövre tenderar de att växa mer i diameter än på höjden. De får en tilltagande bulligare form och vid en viss grovlek minskar faktiskt formhöjden med avseende på diametern. När man med regression med hjälp av provträden skattar formhöjden kan formhöjdens avtagande tendens för grova träd bli alltför brant.



I diagrammet ovan syns en typisk regressionsskattning av formhöjd för tall och bok. Boken har en mycket svag avtagande formhöjd vid grövre diametrar. Tallen däremot har en drastiskt avtagande formhöjd. En svagt avtagande formhöjd innebär att volymen ändå blir större vid en diameterökning. Om formhöjden avtar som för tall i diagrammet ovan kommer en diameterökning ge en volymminskning, vilket inte verkar rimligt.



I det nämnda exemplet ger formhöjdsskattningen en maximal stamvolym på 3.136 m³sk vid en diameter på 760 mm för tall. Man måste alltså hitta denna punkt och t.ex. låsa fast volymutvecklingen på den maximala stamvolymen. I detta fall

Om tall och $bhd > 760$ mm beräknas formhöjd enligt $fh = 12.544 / [\pi \cdot (0.0005 \cdot bhd)^2]$

Allt detta kan tyckas onödigt då grova träd alltid blir provträd och därmed inte får en skattad stamvolym genom ett simulerat formhöjdsvärde. Men formhöjdsskattningarna används sedan 2011 även för att skatta volymtillväxten. Grova träd utan maskinmätta borrhov skulle då få en onaturligt låg tillväxt.

Skatta alltid den väderkorrigerade tillväxten

Skattningsfunktionerna för tillväxten är tränade på den väderkorrigerade tillväxten. Klavträd utan borrtäv som inte är återträd kan skattas direkt med hjälp av tillväxtskattningsfunktionerna. För återklavträd som tillväxtskattas genom skillnaden i stamvolym enligt formhöjdsfunktioner måste tillväxten först väderkorrigeras innan fortsatt beräkning sker.

Beräkna ytavvikelsen i simuleringen enligt provträdens uppmätta väder-korrigerade tillväxt och deras skattade väderkorrigerade tillväxt.

För provträd med borrtäv har tillväxten först beräknats för den avsatta tillväxten och sedan väderkorrigerats. Simuleringens ytcomponent beräknas då enkelt med hjälp av uppmätt väderkorrigerad tillväxt och skattad dito. För återprovträd har den avsatta tillväxten uppmätts genom volymskillnaden under fem år och sedan väderkorrigerats. För att beräkna simuleringens ytcomponent förfar man sedan på samma sätt som för ytor med borrtäv.

Först sedan den väderkorrigerade tillväxten simulerats beräknas den avsatta tillväxten.

Observera att tillväxten skattas på två (2) olika sätt

Vid simulering av tillväxt skattas provträd som inte är återprovträd och klavträd olika. Det är parallellt med provträdsberäknad tillväxt.

- Tillfälliga klavträd skattas med tillväxtskattningsfunktion
- Permanenta klavträd som inte är återträd skattas med tillväxtskattningsfunktion
- Återträd skattas genom volymskillnad enligt formhöjds-skattningsfunktion

Således finns det också två ytcomponenter vid simuleringen beroende på vilken skattningsmetod som används. Var uppmärksam på att slumpdelen i ytcomponenterna dock måste vara densamma oberoende av vilken ytcomponent som används!

Använd inte trädålder på andra ägoslag än produktiv skogsmark vid tillväxt-skattning

Var noga på att skilja på produktiv skogsmark och övriga ägoslag vid tillväxtsimulering. Vid fallet produktiv skogsmark har man tillgång till en simulerad brösthöjdsålder, men den saknas på andra ägoslag än produktiv skogsmark. Brösthöjdsåldern bör användas vid tillväxtskattning på produktiv skogsmark så den skattade tillväxten harmoniserar mot den simulerade stamvolymen.

Skatta inte stamvolymen efter den klavade stubben med hjälp av klavträdens formhöjds-skattningsfunktioner

De träd som avverkats är nästan alltid en avvikande population från de stammar som står kvar efter avverkning. Därför är det ingen bekväm genväg att på tillfälliga förrådsytor med stubbklavning använda de formhöjdsfunktioner som utvecklats för klavträden

Begränsa värderummet för klavstubbarnas skattade formhöjd

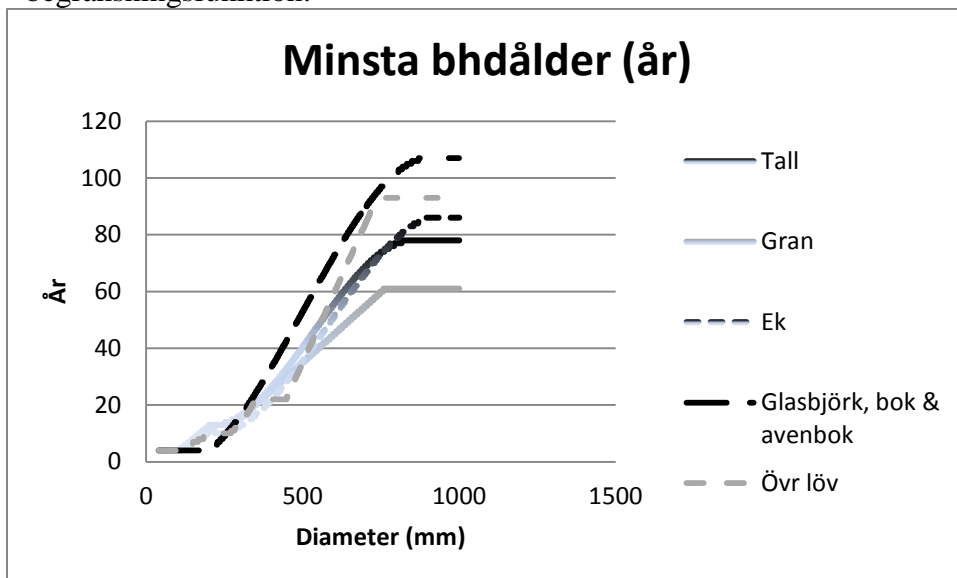
Regression med många oberoende variabler medför alltid en risk för "outliers" - onaturligt stora och små värden. Vi har valt att empiriskt bestämma ett värdeintervall som omfattar 99,95% av klavstubbarnas formhöjder. Rent teoretisk kan formhöjden inte understiga 1.3 meter, rent empiriskt ligger minivärdena något högre än så. Vilka värderum som tillämpas syns i källkod och i beräkningsdokumentet.

Simuleringsdelen av simuleringen

Min- och max-simuleringsvärden

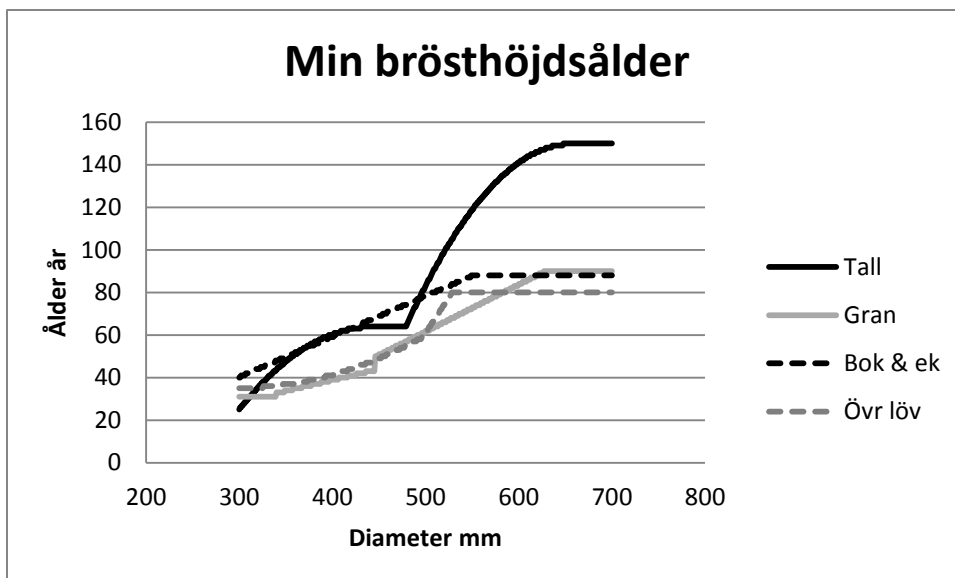
Även om den predikerande egenskapen har ett begränsat värderum i form av min- och maxvärde så kan simuleringens ytvisa och trädvisa komponenter ändå generera ett osannolikt extremvärde. Därför måste även det simulerade värdet begränsas till ett rimligt värdeintervall. Det bör påpekas att det är mycket sällsynt som ett simulerat värde hamnar utanför ”det rimliga värdeintervallet”, men har man c:a 100 000 klavträd om år så kommer det att hända några enstaka gånger.

- För simulerad formhöjd (m) för klavträd gäller ett minvärde på 1.3 m.
- För simulerad formhöjd (m) för klavstubbar gäller ett minvärde på 1.66 m.
- För simulerad volymtillväxt finns inga begränsningar av simulerat värde, eftersom begränsningen i prediktionsdelen är tillräcklig.
- För simulerad brösthöjdsålder gäller generellt minsta ålder 4 år.
- För simulerad brösthöjdsålder för träd utan trädhöjd gäller minsta ålder enligt en begränsningsfunktion:



Figuren visar klavträdens begränsningsfunktionens minvärden för trädslag och diameter.

- För simulerad brösthöjdsålder på icke-åldersmätta provträd på ytor med beståndsmedelhöjd gäller en ytterligare begränsning för minsta ålder om provträdet är minst 300 mm i brösthöjdsdiameter och trädhöjden minst 1,3333 ggr högre än medelhöjden samt att beståndsmedelhöjden är minst 7 meter, då gäller:



Figuren visar minvärden för trädslag och diameter för provträd $bhd \geq 300$ mm, trädhöjd $> 1.333 \times$ beståndsmedelhöjd samt att beståndsmedelhöjden ≥ 7 meter.

Test av simuleringen

Om det är frågan om revidering av simuleringsfunktionerna baserade på nytt provträdsmaterial så bör man i testinstans av RT-databasen genomföra den nya simuleringen för de år och egenskaper som de nya provträden representerar. Om man byggt en ny simulering på provträd 2007 – 2017 så bör den nya simuleringen testas gentemot den gamla. Om den inte uppvisar förbättringar för väntevärde, spridning och harmoni är det ingen idé att implementera den.

Test av simuleringens väntevärdesriktighet

Om vi har ett signifikant systematiskt fel i simuleringen så ligger medelfelet $bias = \frac{1}{n} \sum_{i=1}^n (\tilde{y}_i - y_i)$ utanför konfidensintervallet $\pm 1.96 \sqrt{n^{-2} \sum (\tilde{y} - y)^2}$.

Kategori	n	Volymtillväxt
Tall	40 068	0.000359 \pm 0.000230
Gran	46 686	-0.001299 \pm 0.00322
Björk	12 358	-0.000163 \pm 0.000513
...

Systematiska avvikelser kommer att finnas eftersom bl.a. antagandet om normalfördelning inte alltid gäller. Om det systematiska felet är litet kan man ändå acceptera simuleringen. Använd måttet relativt medelfel $\frac{1}{\bar{y}n} \sum_{i=1}^n (\tilde{y}_i - y_i)$.

Kategori	n	Volymtillväxt
Tall	40 068	0.798%
Gran	46 686	-2.208%
Björk	12 358	-0.609
...

Om felet är mindre än det som uppstått vid klavning och mätning så må det accepteras.

Test av simuleringens spridning och harmoni

För att utröna om simulering är spridningsriktig kan man studera kovariansmatrisen för provträdens faktiska värden för formhöjd, femårig volymtillväxt och brösthöjdsålder. Jämför denna kovariansmatris med kovariansmatriserna för samma provträds predikterade värden och simulerade värden. Om diagonalvärdena är ungefär lika mellan de faktiska och de simulerade värdena så är simuleringen spridningsriktig. Om övriga element också är ungefär lika så kan simuleringen sägas vara harmonisk, dvs den upprätthåller korrelationen mellan de simulerade egenskaperna.

Nedanstående exempel från 1999 gäller kovariansmatriser (formhöjd, volymtillväxt och brösthöjdsålder) skattade för 82 452 provträd på 29 552 provvytor. Man ser att spridningen inte återställs till 100 % genom simuleringen men att den duger.

Provträdsvärden	Predikterade provträdsvärden	Simulerade provträdsvärden
$\begin{pmatrix} 6.04 & 0.08 & 34.90 \\ 0.08 & 0.00 & 0.07 \\ 34.90 & 0.07 & 1492.96 \end{pmatrix}$	$\begin{pmatrix} 4.85 & 0.07 & 29.92 \\ 0.07 & 0.00 & 0.18 \\ 29.92 & 0.18 & 981.30 \end{pmatrix}$	$\begin{pmatrix} 5.65 & 0.08 & 34.19 \\ 0.08 & 0.00 & 0.15 \\ 34.19 & 0.15 & 1281.65 \end{pmatrix}$

Spridningen mellan träden på en provyta och mellan provvytor kan testas på samma sätt genom "faktoriserade" kovariansmatriser där man analytiskt delar upp residualernas spridning ytviss och trädvis:

Låt följande beteckningar gälla:

m_i	Antal provträd på yta i
I	Antalet provvytor där $m_i \geq 2$
M	$\sum_{i=1}^I m_i$
z_{ikl}	Provträdsresidualer (faktiskt-predikterat) för yta i , träd k och simulerad egenskap l .
$\bar{z}_{i\bullet l}$	Medelvärdesvektor för ytor med m_i provträd och där $m_i \geq 2$.
$\Sigma_{träd}$	Den trädvisa kovariansmatrisen
$\hat{\sigma}_{träd_{rs}}^2$	(r,s):te elementet i $\Sigma_{träd}$
Σ_{yta}	Den ytvisa kovariansmatrisen
$\hat{\sigma}_{yta_{rs}}^2$	(r,s):te elementet i Σ_{yta}
SS_{rs}	Produktsumma mellan variabel r och s .

Konstruera $\Sigma_{träd}$ enligt:

$$\hat{\sigma}_{träd_{rs}}^2 = \frac{\sum_{i=1}^I \sum_{k=1}^{m_i} (z_{ikr} - \bar{z}_{i\bullet r})(z_{iks} - \bar{z}_{i\bullet s})}{\sum_{i=1}^I (m_i - 1)}$$

Bilda medelvärdesvektorer enligt:

$$\bar{z}_{\bullet\bullet l} = \frac{\sum m_i \cdot \bar{z}_{i\bullet l}}{\sum m_i}$$

Produktsumman mellan variabel(egenskap) r och s kan beskrivas som:

$$SS_{rs} = \sum_i m_i \cdot (\bar{z}_{i\bullet r} - \bar{z}_{\bullet\bullet r})(\bar{z}_{i\bullet s} - \bar{z}_{\bullet\bullet s})$$

$$E(SS_{rs}) = \left(M - 2 + \frac{\sum_{i=1}^I m_i^2}{M} \right) \cdot \sigma_{yta_{rs}}^2 + (I - 1) \cdot \sigma_{träd_{rs}}^2$$

Då är $\hat{\sigma}_{yta_{rs}}^2$ en unbiased skattning av det (:s):te elementet för Σ_{yta} enligt:

$$\hat{\sigma}_{yta_{rs}}^2 = \frac{\mathbf{SS}_{rs} - (I - 1)\hat{\sigma}_{träd_{rs}}^2}{M - 2 + \frac{\sum m_i^2}{M}}$$

Nedanstående exempel från 1999 gäller faktoriserade kovariansmatriser (formhöjd, volymtillväxt och brösthöjdsålder) mellan träden på provyta och mellan provytor.

82 452 provträd 29 552 provytor	Provträdsvärden	Predikterade provträdsvärden	Simulerade provträdsvärden
Mellan provträd	$\begin{pmatrix} 2.04 & 0.03 & 14.13 \\ 0.03 & 0.00 & 0.16 \\ 14.13 & 0.16 & 457.46 \end{pmatrix}$	$\begin{pmatrix} 1.29 & 0.03 & 12.40 \\ 0.03 & 0.00 & 0.25 \\ 12.40 & 0.25 & 150.00 \end{pmatrix}$	$\begin{pmatrix} 1.75 & 0.03 & 12.64 \\ 0.03 & 0.00 & 0.20 \\ 12.64 & 0.20 & 386.82 \end{pmatrix}$
Mellan provytor	$\begin{pmatrix} 3.84 & 0.05 & 16.26 \\ 0.05 & 0.00 & -0.14 \\ 16.26 & -0.14 & 975.42 \end{pmatrix}$	$\begin{pmatrix} 3.45 & 0.05 & 14.71 \\ 0.05 & 0.00 & -0.12 \\ 14.71 & -0.12 & 80.44 \end{pmatrix}$	$\begin{pmatrix} 3.75 & 0.05 & 17.79 \\ 0.05 & 0.00 & -0.09 \\ 17.79 & -0.09 & 951.61 \end{pmatrix}$

Samma som förut - inte perfekt men det får duga, det går inte att väsentligen förbättra simuleringen.

Litteratur

- Andersson, S.-O. 1954. Funktioner och tabeller för kubering av småträd. Meddelanden från Statens Skogsforskningsinstitut 44:12, 29 s
- Brandel, G. 1990. Volymfunktioner för enskilda träd. Tall, gran och björk. SLU, Inst f skogsproduktion, Rapport 26, 72 s., Garpenberg.
- Carbonnier, C. u.å. Funktioner för kubering av europeisk, sibirisk och japansk lärk. Manuskript, 12 s.
- Eerikäinen, Kalle, A Multivariate Linear Mixed-Effects Model for the Generalization of Sample Tree Heights and Crown Ratios in the Finnish National Forest Inventory, Society of American Foresters, 2009, pp. 480-493
- Eriksson, B (1985). The Swedish National Forest Inventory: A brief description. In Forstliche Nationalinventuren in Europa (ed Dr Peltz & T Cunia) Universität Freiburg, Mitteilungen der Abteilung für Forstliche Biometrie pp 85-3.
- Eriksson, H. 1973. Volymfunktioner för stående träd av ask, asp, klibbal och contorta-tall. Skogshögskolan, Inst. f. skogsproduktion, Rapporter och Uppsatser 26, 26 s., Stockholm.
- Feller, W (1971). An Introduction to Probability Theory and its Application. Second Edition. Wiley, New York.
- Hagberg, E. & Matérn, B. 1975. Tabeller för kubering av ek och bok. Skogshögskolan, Inst. f. skoglig matematisk statistik, Rapporter och Uppsatser 14, 118 s., Stockholm.
- Holm, S, Hägglund, B& Mårtensson, A (1979). A method for generalization of sample tree data from the Swedish National Forest Survey. (in Swedish) Swedish University of Agricultural Sciences, Dept of Forest Survey, Report nr 25. Umeå.
- Kari T. Korhonen& Annika Kangas, Application of nearest-neighbour regression for generalizing sample tree information. Scandinavian Journal of Forest Research.
- Kangas& Korhonen Generalizing sample tree information with semiparametric and parametric models. Silva Fennica pp 151-158. 1994
- Lappi, J. Calibration of height and volume equations with random parameters. Forest Science Volume: 37 Issue 3 pp 781-801 1991
- Loetsch, F, Zörher, F & Haller, K e (1973). Forest inventory. Volume 2. München
- Näslund, M (1940). Funktionen Und Tabellen zur Kubierung stehender Mäume. Kiefter, Fichte un Birke in Nordschweden. (in Swedish) Medd Stat Skogsförsöksanstalt Stockholm
- Näslund, M. & Hagberg, E. 1950. Skogsforskningsinstitutets större tabeller för kubering av stående träd. Tall, gran och björk i södra Sverige. Statens Skogsforskningsinstitut, Experimentalfältet, 200 s., Stockholm.
- Fridman, Jonas et.al. 2014. Adapting National Forest Inventories to changing requirements – the case of the Swedish National Forest Inventory at the turn of the 20th century. Silva Fennica

Westerlund, Bertil (2014) Generalizing sub-sample tree observations in forest inventories: the mixed model approach used in the Swedish National Forest Inventory. Unpublished manuscript, Swedish University of Agricultural Sciences, Dept of Forest Survey

Åkesson, Hans (1999). Neurala modeller och simulering av klavträäd i Riksskogstaxeringen. Examensarbetet i datavetenskap, Umeå Universitet, UMNAD 265.99

Bilaga

Exempel på predikteringsfunktioner och uppdelning av användningsområde

Funktioner för prediktering av formhöjd för björk på produktiv skogsmark

medelhöjd under 7 meter, diameter 100 mm och större

$$fh = -5.70007 + 0.029311 * \text{inklavad grundyta} - 0.000028097 * \text{latitud} \times \text{altitud} + 2.01926 * \text{diameter}^{0.3} + 0.10086 * \text{ståndortsindex}$$

medelhöjd 7 meter eller högre, diameter 100 mm och större

$$fh = -1.36350 + 0.035566 * \text{fältmätt grundyta} - 0.016040 * \text{beståndsålder} + 0.000038563 * \text{beståndsålder}^2 + 0.21855 * \text{medelhöjd} - 0.000012287 * \text{latitud} \times \text{altitud} + 0.94707 * \text{diameter}^{0.3} + 0.014544 * \text{ståndortsindex}$$

diameter 40-99 mm

$$fh = 21.0547 - 2889.46 * (1/\text{diameter} + 50) + 121164.96 * (1/\text{diameter} + 50)^2 + 0.033319 * \text{fältmätt grundyta} - 0.022100 * \text{beståndsålder} + 0.000060859 * \text{beståndsålder}^2 + 0.080835 * \text{medelhöjd} - 0.028157 * \text{diameter}$$