



Beståndsanpassad aptering: metodutveckling och utfallsprövning

- Effekter av varierande längdstyrning i
egenskapsskilda beståndsgrupper

*Stand specific bucking instructions: method development and outcome evaluation
- Effects of varying the bucking length control in stand groups with varying characteristics*

William Svederberg

Examensarbete • 30 hp

Sveriges lantbruksuniversitet, SLU

Skogssfakulteten, Institutionen för skogens biomaterial och teknologi

Jägmästarprogrammet

Rapport från Institutionen för skogens biomaterial och teknologi, 2023:6

Umeå 2023



Beståndsanpassad aptering: metodutveckling och utfallsprövning - Effekter av varierande längdstyrning i egenskapsskilda beståndsgrupper

*Stand specific bucking instructions: method development and outcome evaluation
- Effects of varying the bucking length control in stand groups with varying characteristics*

William Svederberg

Handledare: Dag Fjeld, Sveriges lantbruksuniversitet, Institutionen för skogens biomaterial och teknologi
Bitr. handledare: Oskar Gustavsson, Skogforsk
Bitr. handledare: Johan J Möller, Skogforsk
Examinator: Ola Lindroos, Sveriges lantbruksuniversitet, Institutionen för skogens biomaterial och teknologi

Omfattning: 30 hp
Nivå och fördjupning: Avancerad nivå, A2E
Kurstitel: Masterarbete i skogsvetenskap – Skogens biomaterial och teknologi
Kurskod: EX0956
Program/utbildning: Jägmästarprogrammet
Kursansvarig inst.: Institutionen för skogens biomaterial och teknologi
Utgivningsort: Umeå
Utgivningsår: 2023
Serietitel: Rapport från Institutionen för skogens biomaterial och teknologi,
Delnummer i serien: 2023:6

Nyckelord: beståndsanpassad aptering, maskininlärning, produktionsstyrning, skördardata

Sveriges lantbruksuniversitet

Fakulteten för Skogsvetenskap (S)

Institutionen för skogens biomaterial och teknologi (SBT)

Sammanfattning

Idag är tillgången till skogliga data mycket stor. En bidragande faktor till detta är insamlingen av data vid maskinell avverkning. Denna datamängd används i stor utsträckning för bland annat produktionsuppföljning men med viss bearbetning kan de användas för att digitalt återskapa den avverkade skogen. I denna studie presenteras och utvärderas en metod för att dela in den återskapade, digitala skogen, i apteringsgrupper med hjälp av maskininlärning baserat på genomsnittliga skogliga mätvärden som exempelvis diameter och höjd. Därefter utforskas effekterna av att variera längdstyrningen vid aptering av dessa egenskapsskilda apteringsgrupper. Detta gjordes för att fastställa hur resulterande timmerandel, fördelningsgrad, stockegenskaper och andel stamfelsved påverkas av att kort eller långt timmer produceras i de olika grupperna.

Utifrån skördardata från SCA återskapades över 4 miljoner stammar. Avverkningsobjekten segmenterades sedan i beräkningsytor utefter variation i övre höjd. De resulterande beräkningsytorna delades därefter in i så kallade apteringsgrupper med hjälp av klustringsalgoritmen K-means, en oövervakad maskininlärningsalgoritm som jämför beräkningsytorna och grupperar dem utefter deras egenskaper. Apteringsgrupperna användes därefter för att skapa stambanker som nyttjades för att testa apteringsutfall genom simulering på ca 70 000 stammar. För att validera stambankerna och delar av apteringsutfallet gjordes jämförelser mot apteringsgrupperna. Totalt gjordes 24 apteringsförsök vilka sammanställdes och analyserades avseende timmerandel, fördelningsgrad, stockegenskaper och andel stamfelsved.

Studiens resultat påvisar en komplex interaktion mellan olika variabler, vilket gör produktionsstyrning av denna typ komplicerad. Resultatet tyder bland annat på att timmerandelen alltid blir högre vid hård styrning mot långt timmer samtidigt som fördelningsgraden minskar med givet fördelningsönskemål. Medelstamvolymens påverkan på apteringsutfallet diskuteras då denna tycks ha en påverkan på utfallet i flera aspekter. Beståndsanpassad aptering som koncept diskuteras med koppling till studiens resultat och tidigare projekt för att synliggöra att en stor mängd faktorer påverkar apteringsutfallet i varje bestånd. Slutligen konstateras att vidare studier behövs för att identifiera och utvärdera alla effekter som variation av apteringsinstruktionerna mellan skog med olika egenskaper resulterar i.

Nyckelord: beståndsanpassad aptering, maskininlärning, produktionsstyrning, skördardata

Abstract

Today, the availability of forestry data is vast. One contributing factor to this is the collection of data during mechanized logging. This data is widely used for production monitoring, but with some processing, it can be used to digitally recreate the harvested forest. This study presents and evaluates a method for dividing the recreated, digital forest into bucking groups using machine learning based on average stand measurements such as diameter and height. The effects of varying length control when running simulated bucking on these groups with distinctive characteristics are then explored. This is done to determine how the resulting timber percentage, distribution, log properties, and proportion of logs with timber dimensions classed as pulpwood are affected by producing short or long timber in the different groups.

Based on harvesting data from SCA, over 4 million stems were recreated. The recreated harvest sites were then segmented into calculation areas based on variation in upper height. The resulting calculation areas were then divided into so-called bucking groups using the clustering algorithm K-means, an unsupervised machine learning algorithm that compares the calculation areas and groups according to their characteristics. The bucking groups were then used to create stem banks, which were subsequently used to test bucking outcomes through simulation on approximately 70,000 stems. To validate the stem banks and parts of the bucking outcome, comparisons were made against the bucking groups. In total, 24 bucking simulations were carried out, which were compiled and analysed with respect to timber percentage, distribution, log properties, and proportion of logs with timber dimensions classed as pulp wood.

The results of the study demonstrate a complex interaction between different variables, which makes production control of this type complicated. The results suggest, among other things, that the timber percentage always increases with strict control towards long timber, while the distribution decreases with a given desired distribution. The impact of the average stem on the bucking outcome is discussed as it appears to have an impact on the outcome in several aspects. Stand-adapted bucking as a concept is discussed in connection with the study's results and previous projects to demonstrate that many factors affect the bucking outcome in each stand. Finally, it is concluded that further studies are needed to clarify all the effects that can be identified by varying the bucking instructions between forests with distinctive characteristics.

Keywords: stand specific bucking instructions, machine learning, forest harvest, production control

Förord

Detta examensarbete har genomförts vid Institutionen för skogens biomaterial och teknologi, Sveriges lantbruksuniversitet. Arbetet omfattar 30 hp på avancerad nivå.

Att genomföra denna studie har varit en rolig och intressant utmaning med många nyanser, vilket gett mig värdefulla erfarenheter som jag kommer att bära med mig länge. Jag vill tacka Oskar Gustavsson och Johan J Möller, mina handledare vid Skogforsk, för en intressant idé, ett värdefullt stöd och stort engagemang under arbetets gång.

Ett ytterligare tack riktas också till John Arlinger, som hjälpt till med att skapa stambanker från skördardata och Ingemar Eriksson som hjälpt till med bearbetning av skördardata.

Studien hade inte kunnat genomföras utan ett omfattande datamaterial. Därför vill jag också rikta ett varmt tack till Anna Bylund, Andreas Eriksson, Edward Sjödin och Henrik Sjölander på SCA Skog som effektivt möjliggjort datainsamlingen och som bidragit med värdefulla insikter kopplade till materialet och studien som helhet.

Slutligen vill jag tacka min handledare Dag Fjeld, SLU som väglett mig genom arbetet och som hjälpt mig att utforma studien och dess omfattning.

Umeå, Mars 2023

William Svederberg

Innehållsförteckning

Förord	7
Tabellförteckning	10
Figurförteckning	12
1. Inledning	14
1.1 Svenskt skogsbruk - Kortvirkesmetoden.....	14
1.2 Aptering.....	14
1.3 Stambanker.....	15
1.4 Datainsamling i det praktiska skogsbruket	15
1.5 Maskininlärning och informationsutvinning	16
1.6 Utnyttjande av skördardata i tidigare apteringsstudier.....	16
1.7 Beståndsanpassad aptering och vision	17
1.8 Syfte och mål	19
2. Material och metod	20
2.1 Datainsamling	21
2.2 Datastruktur.....	22
2.3 Mjukvara och programmoduler	23
2.4 Förberedande bearbetning av datamaterialet.....	24
2.4.1 Datatvätt.....	24
2.4.2 Standardisering av data.....	25
2.5 Klustering av beräkningsytor.....	26
2.5.1 Klustering med K-means	27
2.6 Analys av stamegenskaper inom apteringsgrupper.....	29
2.7 Analys av timmerandelar i rådata	29
2.8 Analys av stamfelsved	29
2.9 Simulering av apteringsutfall.....	29
2.9.1 Apteringsinstruktioner	30
2.9.2 Uttag av beståndsvariabler för validering mot källmaterial.....	32
3. Resultat	33
3.1 Klustering av beräkningsytor.....	33
3.1.1 Kluster dominerade av gran.....	33
3.1.2 Kluster dominerade av tall	39

3.1.3	Kluster dominerade av övriga trädslag	45
3.1.4	Kluster med blandad trädslagsfördelning	51
3.2	Apteringsanalys.....	53
3.2.1	Timmerandel.....	54
3.2.2	Fördelningsgrad	55
3.2.3	Stockegenskaper	56
3.2.4	Stamfelsved	57
4.	Diskussion	58
4.1	Studiens resultat	58
4.2	Metoddiskussion	60
4.3	Datamaterialet.....	61
4.4	Tillämpning och vidare studier	62
4.5	Slutsatser	64
	Referenser.....	65
	Bilaga 1.....	67
	Bilaga 2.....	68
	Bilaga 3.....	69
	Bilaga 4.....	70
	Bilaga 5.....	71
	Bilaga 6.....	72

Tabellförteckning

Tabell 1. Studiens delar, metod för analys samt nyttjat indata per steg	20
Tabell 2. Definition av innehåll i de tre datagrupper som användes under studien.....	22
Tabell 3. Mjukvara nyttjad under studien med versionsnummer och användningsområde	23
Tabell 4. Programmoduler använda inom python i jupyter lab för hantering och manipulering av data.....	24
Tabell 5. Antal observationer före och efter datatvätt för respektive datagrupp. Värden inom parentes anger alternativ version av datamatriken med en mer försiktig tvätt	25
Tabell 6. Prismatriser för tall- respektive grantimmer. Längd anges i centimeter och diameter anges i millimeter. Färgkodat från rött till grönt för ökande pris (kr/m ³ to).....	30
Tabell 7. Pris per diameterklass (kr/m ³ fub) för massaved och stamfelsved	30
Tabell 8. Fördelningsönskemål (%) per längdklass för kort, neutral och lång apteringsinstruktion. Fördelningsönskemålet var detsamma för alla diameterklasser.....	31
Tabell 9. Genomsnittliga beståndsegenskaper för apteringsgrupper dominerade av gran. Värden inom parentes anger standardavvikelse för given variabel	37
Tabell 10. Genomsnittliga stamegenskaper för tall och gran i grandominerade kluster. Värden inom parentes anger standardavvikelse per variabel.....	38
Tabell 11. Genomsnittliga beståndsegenskaper för apteringsgrupper dominerade av tall. Värden inom parentes anger standardavvikelse för given variabel	43
Tabell 12. Genomsnittliga stamegenskaper för tall och gran i talldominerade apteringsgrupper. Värden inom parentes anger standardavvikelse per variabel	44
Tabell 13. Genomsnittliga beståndsegenskaper för apteringsgrupper dominerade av övriga trädslag. Värden inom parentes anger standardavvikelse för given variabel	49
Tabell 14. Genomsnittliga stamegenskaper för tall och gran i apteringsgrupper dominerade av övriga trädslag. Värden inom parentes anger standardavvikelse per variabel	50
Tabell 15. Genomsnittliga beståndsegenskaper för apteringsgrupper med blandad trädslagsfördelning. Värden inom parentes anger standardavvikelse för given variabel	51

Tabell 16. Genomsnittliga stamegenskaper för tall och gran i apteringsgrupper med blandad trädslagsfördelning. Värden inom parentes anger standardavvikelse per variabel	52
Tabell 17. Stamantal per stambank och jämförelse av stambanksdata (vänster) och källmaterialet (höger) för ett urval av beståndsvariabler	53
Tabell 18. Resulterande timmerandel (%) per apteringsgrupp och apteringsinstruktion. Fetmarkerat anger högsta timmerandel per grupp. Avser total timmerandel summerat över alla trädslag.....	54
Tabell 19. Resulterande fördelningsgrad för olika apteringsinstruktioner avrundat till närmsta heltal. Uppdelat för talltimmer (T) och grantimmer (G). Fetmarkerat anger bästa fördelningsgrad per apteringsgrupp och trädslag	55
Tabell 20. Genomsnittliga timmerlängder (cm) per apteringsgrupp och apteringsinstruktion. Kompletterat med egenskaper kopplade till respektive grupp	56
Tabell 21. Stamfelsesandelar (%) per apteringsinstruktion och grupp. Trädslag anges som T = Tall och G = Gran. Fetmarkerat anger högsta noterade stamfelsesandel per apteringsgrupp och trädslag. Referensvärden (ref.) hämtade per trädslag från källmaterialet från SCA	57
Tabell 22. Min- och max-värden för variabler i varje datagrupp för utsortering av oönskade värden	69

Figurförteckning

Figur 1. Exempel på ett avverkningsobjekt indelat i apteringsgrupper baserat på skogens eller beräkningsytornas variation som anges med olika färger. Exemplet har skapats på stamnivå utefter denna studies resulterande apteringsgrupper.....	18
Figur 2. Område för datainsamling.....	21
Figur 3. E-värde per antal kluster för beräkningsytor dominerade av gran. Optimalt antal kluster enligt analysen anges med streckad linje.....	33
Figur 4. Genomsnittlig siluettkoefficient per antal kluster för beräkningsytor dominerade av gran.	34
Figur 5. Siluettgrafer för k=3 kluster av grandominerade beräkningsytor.....	35
Figur 6. Jämförelse av kardinalitet och magnitud för k=3 kluster för grandominerade beräkningsytor.....	36
Figur 7. E-värde per antal kluster för beräkningsytor dominerade av tall. Optimalt antal kluster enligt analysen anges med streckad linje.....	39
Figur 8. Genomsnittlig siluettkoefficient per antal kluster för beräkningsytor dominerade av tall.	40
Figur 9. Siluettgrafer för k=2 kluster av talldominerade beräkningsytor.....	41
Figur 10. Jämförelse av kardinalitet och magnitud för k=3 kluster för talldominerade beräkningsytor.....	42
Figur 11. E-värde per antal kluster för beräkningsytor dominerade av övriga trädslag. Optimalt antal kluster enligt analysen anges med streckad linje.....	45
Figur 12. Genomsnittlig siluettkoefficient per antal kluster för beräkningsytor dominerade av övriga trädslag.....	46
Figur 13. Siluettgrafer för k=2 kluster av beräkningsytor dominerade av övriga trädslag.....	47
Figur 14. Jämförelse av kardinalitet och magnitud för k=2 kluster för beräkningsytor dominerade av övriga trädslag.....	48
Figur 15. Parvis jämförelse av vid klustring ingående variabler för grandominerade beräkningsytor.....	67
Figur 16. Parvis jämförelse av vid klustring ingående variabler för talldominerade beräkningsytor.	68

Figur 17. Siluettgrafer för k=2 till k=5 kluster för grandominerade beräkningsytor. Figure 17. Silhouette graphs for k=2 to k=5 clusters for spruce dominated areas.	70
Figur 18. Siluettgrafer för k=2 till k=4 kluster för talldominerade beräkningsytor.	71
Figur 19. Siluettgrafer för k=2 till k=5 kluster för beräkningsytor dominerade av övriga trädslag.	72

1. Inledning

1.1 Svenskt skogsbruk - Kortvirkesmetoden

Under 2021 uppgick nettoavverkningen på svensk skogsmark till närmare 77 miljoner m³fub. Denna volym bestod till ca 50 % av sågtimmer, 41 % av massaved och resterande del av bränslesortiment och annat virke (Skogsstyrelsen 2022). I svenskt praktiskt skogsbruk nyttjas kortvirkesmetoden till närmare 100 % avseende både manuell och maskinell avverkning av sågtimmer, massaved och bränsleved (Gellerstedt & Dahlin 1999). Detta innebär att trädstammen, redan i skogen, delas upp i en rad olika virkessortiment utefter dess fysiska egenskaper. Indelningen görs för att passa ett intrikat system där transport, lagring och bearbetning av virke är optimerat för denna typ av kortare virke, till skillnad från exempelvis helstamsmetoden som används på många andra platser i världen (Gellerstedt & Dahlin 1999).

1.2 Aptering

I grunden finns det två principer för aptering, fördelningsaptering och värdeaptering. Dessa två skiljer sig åt genom att värdeapteringen styr strikt mot att maximera värdet i varje stam. Vid fördelningsaptering tillåts istället ett beslutat maximalt avsteg från värdeoptimeringen av den enskilda stammen för att totalt över alla stammar möta en beställning från industrin. Fördelningsaptering har uppstått utifrån ett behov av att säkerställa att det virke som produceras vid avverkning är det som industrin efterfrågar. Vid förädling är detta viktigt för att öka utnyttjandegraden av varje enskild (von Essen & Möller 1997).

Vid fördelningsaptering är grunden, på samma sätt som vid värdeaptering, en prismatris som anger ett värde för varje längd och diameterkombination. Som komplement adderas en önskefördelning som uttrycker önskad andel av stockar i olika längder och diametrar. Under upparbetningen av en stam vid avverkning jämför apteringsdatorm tidigare utfall mot det totala behovet utifrån önskefördelningen och styr apteringen mot saknade segment i önskefördelningen

med givna restriktioner om att inte göra för stora avsteg mot optimalt stockvärde. Resultatet av apteringen utvärderas slutligen genom fördelningsgraden. Denna kan beräknas enligt:

$$\text{Fördelningsgrad} = 1 - \frac{\sum |m - n|}{2}$$

(1)

där m är önskemålet för en viss längd och diameterkombination och n är utfallet för en viss längd och diameterkombination och $\sum m = \sum n = 1,0$. En fördelningsgrad närmare 1 innebär således en bättre uppfyllnad av önskefördelningen (von Essen & Möller 1997).

1.3 Stambanker

För att kunna göra verklighetsförankrade analyser av exempelvis apteringsutfall med förändrade apteringsinstruktioner krävs ett underlag med stammar som mätts med hög noggrannhet. En stambank kan ses som en samling med digitaliserade stammar som skapats utifrån mätvärden från verkliga stammar. Underlaget till staminformationen kan hämtas från exempelvis skördardata eller från inventeringsdata. Från den inhämtade informationen återskapas sedan digitalt stammar med mål att representera den verkliga skogen. Detta med avseende på fysiska mått, kvalitet och stamfel (Möller & Moberg 2007).

1.4 Datainsamling i det praktiska skogsbruket

Under 2014 implementerades StanForD 2010 i stor skala i Sverige efter att ha genomgått tester sen en första version implementerades 2011. Idag registreras stora mängder data vid avverkning där skördarens aggregat registrerar information om varje avverkad stam. Denna information består av fysiska egenskapsdata i form av trädslag och sortiment samt diameter- och längdmått. Samtidigt innehåller datamängden också normalt ytterligare information om den avverkade stammens geografiska position, avverkningstidpunkt och en rad andra egenskaper som bland annat identifierar typ av kap. Datat kopplas också till enskilda objekt vilket möjliggör uppföljning och betalningsgrundande kalkyler per hanterat objekt (Arlinger & Möller 2021)

Den datamängd som samlas in av skördarna kan användas som underlag för uppföljning, betalning för maskinarbete eller som betalningsgrund för säljaren av avverkningen (Wilhelmsson et al. 2019). Alternativt kan datamaterialet betraktas

som en digital samling av högupplöst beståndsdata, som beskriver varje stam som avverkats med stor detaljrikedom avseende dess fysiska beskafter. Den digitala skogen.

1.5 Maskininlärning och informationsutvinning

Människans möjlighet att förstå och dra slutsatser från ett datamaterial är starkt begränsad av materialets omfattning. Ett datamaterial med ett stort antal observationer och tillhörande variabler är därför direkt omöjligt att analysera utan externa verktyg (Alpaydin 2014).

Maskininlärning innebär i grunden att en människa skapar ett beslutssystem för att göra det möjligt för en maskin (dator) att lösa en viss uppgift. Maskinen tränas sedan med en mängd data för att känna igen egenskaper eller mönster och ta beslut om en respons utifrån vad den lärt från att arbeta med träningsdata (Alpaydin 2014). Utvecklingen av maskininlärning har lett till oövervakad och övervakad maskininlärning. De två olika typerna av maskininlärning skiljer sig främst åt avseende dess metod för att lösa olika problem. Övervakad inlärning innebär att en given metod inte är beroende av en fastställd svarsvariabel. Istället tillåts metoder av denna typ finna egenskaper i ett datamaterial och producera resultat utifrån detta som sedan kan tolkas av användaren. Detta i motsats till övervakad inlärning som kräver att given metod lär sig att känna igen specifika egenskaper som leder till en specifik svarsvariabel (Alpaydin 2014).

När maskininlärning appliceras på ett datamaterial med avsikt att analysera och få insikt i materialet kallas detta för informationsutvinning (eng. data mining). Genom informationsutvinning kan en människa alltså dra slutsatser från egenskaper i ett datamaterial på kort tid genom att nyttja denna typ av tekniska lösningar (Alpaydin 2014).

1.6 Utnyttjande av skördardata i tidigare apteringsstudier

Skogforsk, det svenska skogsbrukets forskningsinstitut, arbetar aktivt med att utreda möjligheterna att nyttja skördardata för att förbättra utbytesprognoser som görs innan avverkningar för att skatta förväntade utfall av olika sortiment. Med start 2015 identifierades initialt ett behov av förbättrade utbytesprognoser med grund i en intervjustudie där tillfrågade skogsföretag tydligt meddelade en önskan om mer precisa utfallsprognoser. Den huvudsakliga och högst rankade anledningen till detta

är behovet av noggranna utbytesprognoser i den 1–3 månaders plan som löpande läggs för verksamheten (Möller et al. 2015b).

Senare studier med avsikt att undersöka möjligheten till ökade utbytesprognoser har gjorts med koppling mot laserdata i kombination med skördardata. Söderberg et. al (2021) presenterar att det är möjligt att, med relativt god precision, uppskatta flertalet skogliga variabler som grundyta, stamvolym och grundtevägd medeldiameter samt grundtevägd medelhöjd genom denna metod. I och med denna kunskap och en eventuell framtida utveckling av metoden för än högre precision är det möjligt att i framtiden kunna ha större tillit till skattade beståndsegenskaper.

1.7 Beståndsanpassad aptering och vision

Redan under tidigt 2000-tal påbörjades arbete med att undersöka möjligheten till beståndsanpassad aptering med målsättningen att öka timmerandelen och samtidigt tillfredsställa industrins beställningar (Kivinen 2007). Kivinen studie visade att fördelningsgraden kunde ökas genom att anpassa apteringsinstruktionerna efter beståndsegenskaper. Dock kunde detta enbart uppnås till kostnad av ett lägre timmerutfall.

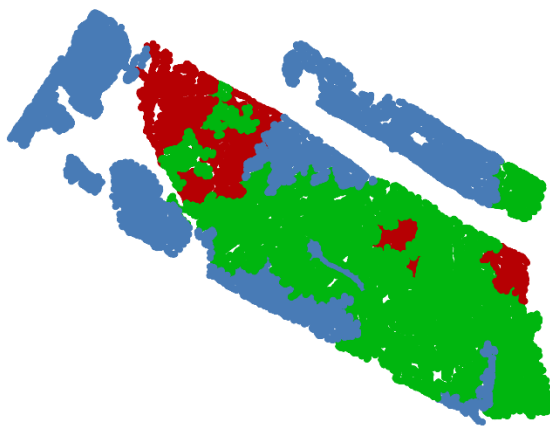
Sedan Kivinen studier utfördes har teknikutvecklingen fortsatt. Dagens utbredda datainsamling innebär att det nu finns en ständigt växande databas innehållande information om varje avverkad stam. Det finns därför nu möjlighet att med ett större material och mer utvecklad teknik, utvärdera beståndsanpassad aptering på nytt. Denna studie kommer att fokusera på hur apteringsutfallet påverkas av att nyttja olika längdstyrning för olika typer av skog. Syftet med denna avgränsning är att utröna hur apteringsutfallet förändras då olika apteringsinstruktioner med olika längdstyrning används i mer homogena grupper av bestånd. Utvärderingen kan ses som en förstudie med avsikt att vägleda framtida utveckling inom området för apteringsteknik.

Visionen om framtida praktisk implementering av anpassad aptering som tagits fram för denna studie baseras på nyttjande av skördardata från tidigare avverkningar enligt följande steg:

1. Genom att bearbeta det data som skördarna samlar in, återskapas skogen i digital form med data om varje stam, objekt och stock. Den återskapade skogen delas upp i så kallade beräkningsytor baserat på övre höjd med avsikt att skapa så homogena ytor som möjligt (Möller et al. 2015a). Varje beräkningsyta blir alltså en delarea av objektet med likvärdiga egenskaper. På detta vis har datamaterialet på objektsnivå komprimerats för att vara mer

hanterbart. Om dessa beräkningsytor sedan grupperas utefter egenskaper erhålls grupper med liknande skogstyper som nyttjas för att skapa anpassade apteringsinstruktioner för varje grupp.

2. När det i nästa steg är dags att avverka ett nytt objekt speglas detta mot de olika grupper som definieras och därav bestäms också på förhand hur apteringen ska utföras i beräkningsytor inom objektet. Detta gäller givetvis så länge inventering av objektet gjorts med hög upplösning och med pålitliga data. Objektet som ska avverkas har därmed delats in utefter den, ofta variationsrika, skogens egenskaper i vad som skulle kunna benämnas som apteringsgrupper. Denna information om tilldelade apteringsgrupper matas in i ett GIS och skickas till den uppkopplade skördaren som automatiskt anpassar apteringen när den går in i ett område med en annan apteringsgrupp. I Figur 1 ges exempel på hur en avverkningstrakt skulle kunna se ut då den delats in i apteringsgrupper. I denna rapport kommer benämningarna kluster och apteringsgrupp att användas likvärdigt beroende på vilket steg i analysen som avses. Då metod beskrivs och diskuteras kopplat till utförande av klustring används generellt ”kluster”. I annat fall används ”apteringsgrupp”.



Figur 1. Exempel på ett avverkningsobjekt indelat i apteringsgrupper baserat på skogens eller beräkningsytornas variation som anges med olika färger. Exemplet har skapats på stamnivå utefter denna studies resulterande apteringsgrupper.

Figure 1. Example of a harvesting object divided into “bucking groups” denoted by different colors. The example has been produced from stem level data in combination with the “bucking groups” that has resulted from this study.

1.8 Syfte och mål

Syftet med studien var att utforma och pröva en metod för att gruppera beräkningsytor från skördardata utefter beståndsegenskaper med hjälp av maskininlärning. Samtidigt ämnade studien att utreda effekterna av att variera längdstyrningen vid aptering av stammar i dessa egenskapsskilda kluster av beräkningsytor, även kallade apteringsgrupper.

Studien hanterade följande frågeställningar:

- Hur väl kan beräkningsytor skapade med skördardata grupperas med hjälp av maskininlärningsalgoritmen K-means?
- Hur skiljer sig beståndsegenskaperna mellan de olika grupperna av beräkningsytor?
- Hur påverkas andelen producerat timmer, fördelningsgraden och andelen stamfelsesved om längdstyrningen varieras i de olika grupperna av beräkningsytor?
- Hur påverkas egenskaperna hos producerade timmerstockar om längdstyrningen varieras i de olika grupperna av beräkningsytor?

Studiens frågeställningar stöddes av delmål enligt följande:

1. Att med hjälp av K-means algoritmen skapa apteringsgrupper bestående av beräkningsytor med liknande egenskaper samt att kvantifiera relationen mellan antal grupper, dess sammansättning och variation inom dem.
2. Att utifrån skördardata skapa underlag för validering av stambanker nyttjade vid apteringssimulering. Detta i form av genomsnittliga stamegenskaper som diameter, höjd och formkvot.
3. Att utifrån skördardata ta fram nyckeltal i form av andel stamfelsesved för jämförelse mot utfall vid apteringssimulering.
4. Att för varje apteringsgrupp kvantifiera utfall av timmerandelar, fördelningsgrad, stockegenskaper och andel stamfelsesved genom simulerad aptering med apteringsinstruktioner där längdstyrningen varieras.

2. Material och metod

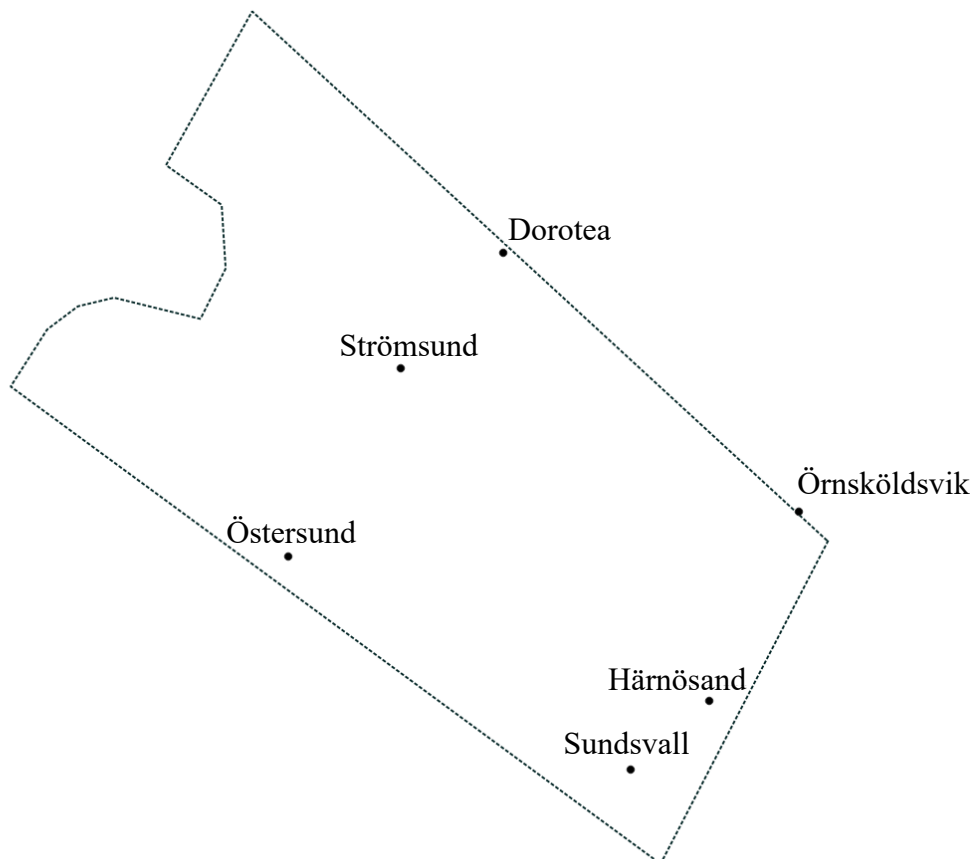
Studien delades in i fyra delar inom vilka de fyra delmålen hanterades enligt Tabell 1. Momenten beskrivs i vidare detalj under följande avsnitt.

Tabell 1. Studiens delar, metod för analys samt nyttjat indata per steg
Table 1. Parts of the study, method used for analysis and used data per part

Del	Metod	Indata
1) Skapa apteringsgrupper utifrån stamegenskaper och redovisa relationen mellan antal grupper och variation inom grupperna	Utför klusteranalys. Presentera variation inom kluster genom att utvärdera E-värde, siluettkoefficient, kardinalitet och magnitud	Objektsdata framtaget ur skördardata
2) Ta fram nyckeltal för validering av stambanker	Skapa relationer mellan det klustrade objektsdatat och ta fram tabeller och grafik för utvärdering av nyckeltal i Power BI	Objektsdata, stockdata och stamdata
3) Ta fram nyckeltal för andel stamfelsved i rådata	Skapa datamatrix med utsorterad stamfelsved. Skapa relationer mellan klustrat objektsdata, stockdata och stamfelsdata i Power BI. Ta fram tabeller och grafik för utvärdering av nyckeltal	Stamdata och stockdata framtaget ur skördardata
4) Kvantifiera utfall av apteringsmodellering med varierad längdstyrning i de olika apteringsgrupperna	Simulera apteringsutfall i stambanker byggda på de olika grupperna. Styr mot korta eller långa timmerlängder samt en neutral längdfördelning. Utvärdera utfall av timmerandelar, stockegenskaper, stamfelsved och fördelningsgrad. Validera mot nyckeltal från tidigare delmoment	Stambanker skapade med grund i klustrat objektsdata

2.1 Datainsamling

SCA Skog bidrog med skördardata till studien. Ett geografiskt område fastställdes för att samla in skördardata från slutavverkning för år 2021. För att skydda personuppgifter som kan vara kopplade till virkesordrar vid köp från privatpersoner bestämdes att datamaterialet endast skulle hämtas från avverkning på skog som ägs av SCA. Området valdes utifrån där SCA har högst densitet av egen skog. En shapefil skapades i Qgis för att definiera området för datainsamling (Figur 2). För att fånga upp variation i beståndsegenskaper beroende på geografisk plats ställdes krav på att maskinerna som informationen hämtades från skulle vara verksamma i olika delar av undersökningsområdet. När det gäller materialets utformning ställdes krav på att det skulle följa StanForD 2010-standarden för att möjliggöra enhetlig hantering av all data. Dessutom krävdes att koordinater skulle finnas med i materialet i form av kranspetsposition eller maskinposition för att möjliggöra inhämtning av information om stammarnas växtplats.



Figur 2. Område för datainsamling.
Figure 2. Area for data gathering.

2.2 Datastruktur

Datamaterialet i form av hpr-filer extraherades av SCA och sändes till Skogforsk som genomförde bearbetning av materialet. Bearbetningen gjordes med hjälp av mjukvarorna hprYield och hprCM med avsikt att återskapa stammar från stockar och segmentera avverkningsobjekten i beräkningsytor på ca 1 ha. Resultatet av bearbetningen grupperades i tre olika upplösningar enligt tabell 2.

Tabell 2. Definition av innehåll i de tre datagrupper som användes under studien
Table 2. Definition of contents in the three data groups used in the study

Upplösning	Innehåll
Objektsdata	Sammanställning av alla stammar inom ett avverkningsobjekt kopplat till en virkesorder. Innehåller beräkningsytor som definierar en beståndsindelning inom objektet baserat på trädhöjd.
Stamdata	Sammanfogade helstammar med parametrar som beskriver en given stams fysiska egenskaper. Innehåller identifieringsnycklar för att koppla till objektsdata.
Stockdata	Information om varje stock som apaterats från varje stam. Beskriver stockarnas egenskaper i form av bland annat sortiment och fysiska dimensioner. Innehåller identifieringsnycklar för att koppla till objektsdata samt stamdata.

Skogforsks bearbetning av materialet innebar beräkning av ett flertal variabler för att beskriva både ståndorten och enskilda stammar. Exempel på variabler som beräknades vid den förberedande bearbetningen är ståndortsindex, formkvot per stam, ålder, grundyta och volym per hektar. Om höjd över havet saknades inhämtades detta från allmänna karttjänster. Temperatursumma för beräkningsytorna beräknades från latitud och höjd över havet.

2.3 Mjukvara och programmoduler

Under studiens gång användes flera olika programvaror för att bearbeta och visualisera datamaterialet (tabell 3).

Tabell 3. Mjukvara nyttjad under studien, med versionsnummer och användningsområde
Table 3. Softwares used during the study, with version number and area of use

Mjukvara	Version	Användning
ForestPrognosis	-	Konstruktion av stambanker
hprCM	-	Återskapa stammar och beräkna biobränslekvantitet
hprYield	-	Segmentera avverkningsobjekt och återskapa stammar
Jupyter Lab	3.4.4	Beräkning och visualisering av data med programmeringsspråket Python
Microsoft Excel 2019	16.0.10391.20029	Enklare bearbetning och visualisering av data
Microsoft Power BI	3222.1031.11418.0	Visualisering av data och tabellkonstruktion
Timan	4.0.7.5	Apteringssimulering och analys av utfall
Qgis Desktop	3.16.16	Definiering av område för datainsamling

De delar av studien där data bearbetades och analyserades genomfördes med grund i programspråket Python. För att tillgängliggöra funktionalitet och förenkla analysprocessen nyttjades modulerna i tabell 4.

Tabell 4. Programmoduler använda inom python i Jupyter lab för hantering och manipulering av data

Table 4. Modules used within python in Jupyter lab for handling and manipulation of data

Modul	Användning
Glob	Förenklad inläsning av filer
Kneed	Hjälpmedel för val av antal kluster
Matplotlib	Visualisering av data
Numpy	Numeriska beräkningar
Os	Inläsning av filer
Pandas	Läsa, strukturera och manipulera data
Seaborn	Visualisering av data
Sklearn (Scikit-learn)	Klustring av datapunkter

2.4 Förberedande bearbetning av datamaterialet

För att kunna nyttja det erhållna datamaterialet vid analys fanns ett behov av att strukturera materialet, rensa rader som saknade värden och filtrera bort oönskade värden som skulle störa senare analyser. Samtidigt fanns ett behov av att göra inledande analyser av det råa datamaterialet för att fastställa avgränsningar, skapa en uppfattning av datamaterialets omfattning och strukturera arbetsgången för kommande analyser. Följande delkapitel beskriver hur datamaterialet förberetts inför vidare analyser.

2.4.1 Datatvätt

Materialets innehåll utforskades i ett första steg i Power BI för att finna datapunkter som innehöll värden så låga eller höga att de var utanför de rimliga ramarna för mätvärdet. Detta för att avgöra rimliga avgränsningar för tvätt av materialet.

Materialet i var och en av de inlästa filerna var organiserat i en datamatrix med rader och kolumner med enbart ett värde i varje cell. I och med detta uppfyllde datamaterialet kraven enligt principen för *tidy data* vilka beskriver kriterier och rekommendationer för förenklad datahantering (Neo 2020). För att underlätta kommande analyser gjordes en rensning av alla rader med celler innehållande NULL-värden eller som saknade värden.

Utifrån tidigare steg filtrerades materialet utefter valda, relevanta max- och min-värden. Filtreringen gjordes även utifrån önskad logik i materialets samband mellan

variabler, exempelvis att godtagbar följd mellan rot diameter, mittdiameter och topp diameter på en enskild stock bestämdes vara i minskande ordning från rot till topp. En utförlig beskrivning av alla filtreringsregler som användes presenteras i bilaga 3.

Mängden indata till studien presenteras i tabell 5.

Tabell 5. Antal observationer före och efter datatvätt för respektive datagrupp. Värden inom parentes anger alternativ version av datamatriken med en mer försiktig tvätt
Table 5. Number of observations prior, and post to data cleaning. Values within parentheses denotes an alternative version of the dataset after a less aggressive cleaning

Datagrupp	Ingående antal observationer	Dataförlust orsakad av datatvätt	Slutligt antal observationer
Objektsdata	5191	1004	4187
Stamdata	4 663 387	1 754 542 (678)	2 908 845 (4 662 709)
Stockdata	12 105 671	4 797 860	7 307 811

Störst andel av dataförlusten utgjordes här av NULL-värden. För stamdata stod denna typ av fel för 1 754 280 förlorade observationer och felet var saknade formkvoter. Då formkvoter inte var avgörande i vissa delar av studien användes en mindre aggressiv datatvätt för dessa delar vilket resulterade i endast 678 bortplockade observationer. Dessa dataförluster utgjordes av stammar som inte klarat begränsningar kopplade till max- och min-värden för olika variabler. Förluster i stockdatat utgjordes till ca 8 % av NULL-värden enbart från avsaknad av sortimentskod. Resterande förluster utgjordes av stockar utanför analysens fokus, exempelvis långa stockar avsedda för produktion av stolpar.

2.4.2 Standardisering av data

När data analyseras med syfte att utvinna information och mönster är det viktigt att analyserat data har samma enheter för att undvika felaktiga resultat. Om olika variabler i datamaterialet har olika enheter, kan de algoritmer som används för att analysera data inte tolka och hantera data på ett korrekt sätt, utan att ge felaktiga viktade resultat till följd av att de olika variablernas storhet varierar. För att kunna jämföra data med olika enheter och få relevanta resultat måste materialet standardiseras eller normaliseras. I denna studie standardiserades datamaterialet med verktyget "StandardScaler" som är en del av sklearn-modulen. Verktyget skalar värden enligt Z-score metoden (scikit-learn 2022). Värden skalades således till att ha ett medelvärde om 0 och en standardavvikelse om 1. Standardiseringen av datamaterialet innebar en möjlighet att analysera data på ett korrekt sätt och få meningsfulla resultat (Mohamad & Usman 2013)

2.5 Klustring av beräkningsytor

För att skapa apteringsgrupper av bestånd med liknande egenskaper nyttjades klustringsmetoden K-means för att ta fram ett antal bestandsgrupper med särskilda egenskaper.

För att underlätta tolkningen av apteringsgrupperna delades objektsmaterialet även upp baserat på trädslagsandel och ståndortsindex enligt följande:

1. Beräkningsytor dominerade av gran
2. Beräkningsytor dominerade av tall
3. Beräkningsytor dominerade av övriga trädslag
4. Beräkningsytor med blandad trädslagsfördelning

Grupp 1–3 fastställdes utefter trädslagsandelar, och det dominerande trädslaget behövde utgöra mer än 50 % av trädslagsfördelningen (%volym) på varje beräkningsyta i gruppen. Grupp 4 definieras av att varken gran, tall eller övriga trädslag enskilt står för mer än 50 % av trädslagsfördelningen. Denna indelning möjliggjorde analys av eventuella effekter på stamegenskaper i blandskogar i motsats till om dessa beräkningsytor hade inkluderats i grupp 1–3.

Klustringen utfördes separat i grupp 1–4 på beräkningsytor med unika indikatorvärden. Från datamaterialet bestämdes följande egenskapsbeskrivande parametrar styra klustringen:

1. Grundytevägd medelhöjd
2. Grundytevägd medeldiameter
3. Medelstamsvolym
4. Formkvot
5. Andel gran (%Volym)
6. Andel tall (%Volym)
7. Andel övriga trädslag (%Volym)

För att stödja valet av antalet apteringsgrupper tog frågan upp med Skogforsk och SCA för att avgöra en hanterbar mängd grupper. Det slutliga valet av antal apteringsgrupper gjordes baserat på deras inspel i kombination med den föregående tekniska analysen som beskrivs i följande avsnitt.

2.5.1 Klustering med K-means

Algoritmen K-means är välkänd inom området för maskinlärning som en enkel metod för oövervakad maskinlärning. Algoritmen använder en iterativ process för att bilda kluster genom att tilldela varje objekt i datamängden till det närmaste medelvärdet för varje kluster. Detta medelvärde, eller centroiden, beräknas som medelvärdet för alla objekt i klustret. Algoritmen itererar över en process där den flyttar enskilda observationer mellan klustren och utvärderar det sammanlagda avståndet till vardera observations tilldelade centroid. Detta pågår till dess att det maximala antalet iterationer gjorts eller tills inga datapunkter längre byter kluster och att det totala avståndet mellan observationer och centroider således inte förändras (Han et al. 2011).

Algoritmen kräver ett bestämt antal slutliga kluster, varför en önskad mängd kluster behövde definieras. Valet av antal kluster är direkt avgörande för analysens resultat. Ett för lågt antal kluster innebär att spridningen inom varje kluster blir hög. Ett för högt antal kluster innebär i sin tur att analysen kan ta onödigt lång tid samtidigt som överlapp mellan kluster ofta ökar (Han et al. 2011). För att skapa ett beslutsunderlag för antalet kluster beräknades därför ett spridningsmått för $k=1$ till $k=20$ kluster. Spridningsmättet kan beskrivas som

$$E = \sum_{i=1}^k \sum_{p \in C_i} \text{dist}(p, c_i)^2 \quad (2)$$

där E är summan av det kvadrerade euklidiska avståndet mellan alla datapunkter p och deras tilldelade klustercentroid c_i för kluster C_i (Han et al. 2011). Det summerade värdet för varje iteration med k kluster plottades sedan i en figur med en markering för optimalt antal k som beräknats med hjälp av modulen "kneed". Denna metod för att avgöra antalet kluster kallas för "armbågsmetoden" (eng. Elbow method) och går ut på att hitta det antal kluster där det inte längre ger en stor vinst i minskad variation vid tillägg av ett ytterligare kluster. Detta kan beskrivas som punkten där avtagande utbyte nås.

Att finna ett optimalt antal kluster är dock inte alltid möjligt enbart genom att studera E -värdets utveckling med antal kluster. Detta då det bara ger en yttlig beskrivning av klustrens utformning. För att få mer insikt i detta gjordes även en siluettanalys där en genomsnittlig siluettkoefficient beräknades för olika antal kluster. Syftet med siluettanalys är att undersöka hur väl varje individuell punkt i en klustering passar in i sitt eget kluster och hur väl den skiljer sig från andra kluster.

Siluettkoefficienten kan enligt Rousseeuw (1987) beskrivas som

$$s(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & \text{om } a(i) < b(i), \\ 0, & \text{om } a(i) = b(i), \\ \frac{b(i)}{a(i)} - 1, & \text{om } a(i) > b(i). \end{cases} \quad (3)$$

där $s(i)$ är siluettkoefficienten s för objekt i . $a(i)$ representerar det genomsnittliga avståndet till alla andra objekt inom klustret för punkt i inom kluster a . $b(i)$ kan beskrivas som

$$b(i) = \min_{C \neq A} d(i, C) \quad (4)$$

där $d(i, C)$ representerar det genomsnittliga avståndet till alla objekt i kluster C utanför kluster a för objekt i . För att välja $b(i)$ identifieras den minsta genomsnittliga avvikelsen till närliggande kluster. Genom att $b(i)$ jämförs med $a(i)$ erhålls information om klustrets täthet (medelavståndet $a(i)$ inom klustret) och dess avstånd från närmsta närliggande kluster (minsta medelavståndet $b(i)$). Från dessa jämförelser erhålls ett genomsnittligt värde

$$-1 \leq s(i) \leq 1 \quad (5)$$

för hela uppsättningen av kluster där värden nära eller under 0 indikerar att det finns observationer som tilldelats fel kluster eller att dess klustertillhörighet är indifferent och att observationen skulle kunna ingå i flera kluster (Rousseeuw 1987).

Kvalitén på klustersammansättningen är också beroende på relationen mellan enskilda klusters kardinalitet (antal observationer i kluster a) och magnitud (det totala avståndet mellan alla observationer i kluster a och dess centroid). En bra sammansättning av kluster identifieras här som att förhållandet mellan kardinalitet och magnitud är detsamma för samtliga kluster. Avvikelser från detta samband innebär att givet kluster håller en onormal mängd observationer jämfört med dess spridning av observationer runt centroiden (Google Developers 2022a).

Slutligen beror valet av antal kluster också på applikationen. Beroende på vad grupperingen förväntas användas till är det viktigt att utvärdera valet av kluster mot dess funktionalitet. I fallet för denna studie förväntades en begränsad mängd kluster om 1–4 kluster per analyserad grupp.

2.6 Analys av stamegenskaper inom apteringsgrupper

För att ta fram stamegenskaper per apteringsgrupp skapades relationer mellan klustrat objektsdata och tvättat stamdata genom att länka beräkningsyte-ID mellan datamatriskerna i Power BI. Därefter skapades tabeller med genomsnittliga egenskaper per träslag och apteringsgrupp. Diameter i brösthöjd, trädhöjd, höjd över havet, stamantal per apteringsgrupp och medelstamsvolym togs fram för varje apteringsgrupp. Detta gjordes för stammar av gran och tall då detta var de träslag som ämnades analyseras i apteringssimuleringen.

2.7 Analys av timmerandelar i rådata

Timmerandelen beräknades per apteringsgrupp genom relaterade tabeller i Power BI. Beräkningsyte-ID kopplades mellan klusterdata och stockdata varpå en timmervolym per apteringsgrupp och träslag kunde tas fram genom DAX-uttryck. Timmerandelar per träslag beräknades därefter genom att dividera volymen timmer per träslag med den totala virkesvolymen per träslag.

2.8 Analys av stamfelsesved

Att analysera andelen stamfelsesved krävde en initial definition av vad som klassas som stamfelsesved. Alla stockar med minst 14 cm toppdiameter som klassats som massaved eller energived bestämdes därför vara stamfelsesved enligt rekommendation från Skogforsk. Förenklat är detta alltså alla stockar som håller timmerdimensioner men som av någon anledning inte klassats som timmer. Stockar klassade som stamfelsesved sorterades ut träslagsvis i två separata datamatrisker, en för gran och en för tall. Stamfelsesved filtrerades ej ut för övriga träslag då timmeruttag i de övriga träslagen är relativt ovanliga i dagsläget. Filtringen gjordes i det tvättade datamaterialet med stockdata då analysen krävde att varje stock hade en sortimentsbeteckning och att stockar med avvikande form ej inkluderades.

2.9 Simulering av apteringsutfall

Simuleringen utfördes med modulen Aptan, som används för att simulera aptering i programmet Timan 4.0.7.6. I simuleringen användes åtta unika stambanker som skapats för de olika klustren. Stambankerna konstruerades av Skogforsk utifrån ett slumpmässigt urval av beräkningsytorna i respektive apteringsgrupp. Därefter konstruerades tre olika apteringsinstruktioner för apteringssimuleringen.

2.9.1 Apteringsinstruktioner

Den längdneutrala apteringsinstruktionen som användes för simuleringarna skapades av Skogforsk och dess konfiguration låg till grund även för instruktionen med styrning mot korta, eller längre timmerstockar. Instruktionen konfigurerades för att möjliggöra aptering av timmer av tall- och granstammar. Övriga träslag tilläts endast apteras till massaved. I grundinstruktionens konfiguration ingick en rad styrande parametrar med anpassningar per träslag. I grunden för värdeberäkningarna fanns två prismatriser för timmer och nio matriser för massaved och stamfelsved (Tabell 6 & 7). Stamfelsved definierades som ett eget sortiment för att underlätta analyser av simuleringens utfall.

Tabell 6. Prismatriser för tall- respektive grantimmer. Längd anges i centimeter och diameter anges som toppdiameter i millimeter. Färgkodat från rött till grönt för ökande pris (kr/m³to)
Table 6. Price matrices for pine and spruce sawlogs. Length is given in centimeters and diameter is given as top diameter in millimeters. Color coded from red to green for increasing price (kr/m³to)

Talltimmer		Diameter											
Längd	120	140	160	180	200	220	240	260	280	300	320	340	360
340	280	348	383	490	540	580	604	630	649	663	663	670	575
370	296	369	405	512	564	598	624	650	669	684	684	691	593
400	312	389	422	539	588	624	644	663	683	698	698	698	599
430	325	405	435	550	600	630	650	670	690	705	705	705	605
460	335	417	444	561	612	636	656	677	697	712	712	712	611
490	344	429	452	572	624	643	663	683	704	719	719	719	617
520	351	437	457	578	630	649	670	690	711	726	726	726	623
550	354	441	461	583	636	655	676	690	711	726	726	726	623

Grantimmer		Diameter											
Längd	120	140	160	180	200	220	240	260	280	300	320	340	360
340	280	378	409	481	544	580	604	620	630	639	649	670	575
370	296	400	432	502	569	598	624	640	650	660	669	691	593
400	315	427	456	529	593	624	644	653	663	673	683	698	599
430	325	440	465	540	605	630	650	660	670	680	690	705	605
460	335	453	474	551	617	636	656	667	677	687	697	712	611
490	344	466	484	562	629	643	663	673	683	694	704	719	617
520	351	475	488	567	635	649	670	680	690	700	711	726	623
550	354	480	493	572	641	655	676	680	690	700	711	726	623

Tabell 7. Pris per diameterklass (kr/m³fub) för massaved och stamfelsved
Table 7. Price per diameter class (kr/m³sub) for pulpwood and wood of sawlog dimensions that has been classified as any other product than sawlogs

Produkt	Toppdiameterklass (mm)	
	50	140
Tall – Massaved	330	0
Tall – Massaved (Stamfelsved)	0	330
Gran – Massaved	330	0
Gran – Massaved (Stamfelsved)	0	330

Talltimmer begränsades till en maxlängd om 580 cm, en minsta längd om 310 cm och en maxdiameter om 500 mm. Grantimmer begränsades till en maxlängd om 580 cm, en minsta längd om 340 cm och en maxdiameter om 521 mm. Toppdiameter mäts alltid 10 cm från toppändan av given stock. För samtliga analyser bestämdes inga premier delas ut för några apteringsalternativ. Talltimmer och grantimmer bestämdes också generera fyra respektive två olika kvalitetsklasser enligt biometrias anvisningar om kvalitetsbestämning (Biometria 2022). Prismatriserna ändrades inte mellan olika kvalitetsklasser. De ändrades heller inte vid byte av stambank.

Vid fördelningsaptering uttrycktes fördelningsgraden som ett relativt antal stockar per diameterklass. Den längdneutrala apteringsinstruktionen tilläts dock inte tolerera prisförändringar mot värdeapteringen. Vid konstruktion av de andra två apteringsinstruktionerna ändrades den neutrala instruktionen enbart i fördelningsönskemålen. Fördelningsmatrisen ändrades för att önska en större andel långa, respektive korta längder (tabell 8). Pristoleransen ställdes till 50 promille för styrning mot kort och långt timmer. Den neutrala instruktionen kan alltså likställas med strikt värdeaptering till följd av i princip nolltolerans mot prisavsteg.

Tabell 8. Fördelningsönskemål (%) per längdklass för kort, neutral och lång apteringsinstruktion. Fördelningsönskemålet var detsamma för alla diameterklasser

Table 8. Desired distribution (%) per length class for short, neutral, and long price settings. The desired distribution was the same for all diameter classes

Längdklass (cm)	Apteringsinstruktion		
	Kort	Neutral	Lång
340	30	12	0
370	28	12	0
400	20	12	5
430	12	12	5
460	5	12	12
490	5	12	20
520	0	12	28
550	0	12	30

Alla de åtta stambankerna apterades tre gånger, en gång med varje apteringsinstruktion. Resultattabeller från Aptan sammanställdes i separata excel-dokument för enkla jämförelser mellan alternativen.

2.9.2 Uttag av beståndsvariabler för validering mot källmaterial

Genom att använda verktyget HPR-analys i Timan kunde uppgifter om de olika stambankerna tas ut för jämförelse mot källmaterialet från SCA. Beståndsvariabler togs ut från HPR-filer genererade vid körning med neutral lista. Variabler som analyserades var medelstamsvolym, grundtevägd diameter (DGV) och grundtevägd höjd (HGV).

3. Resultat

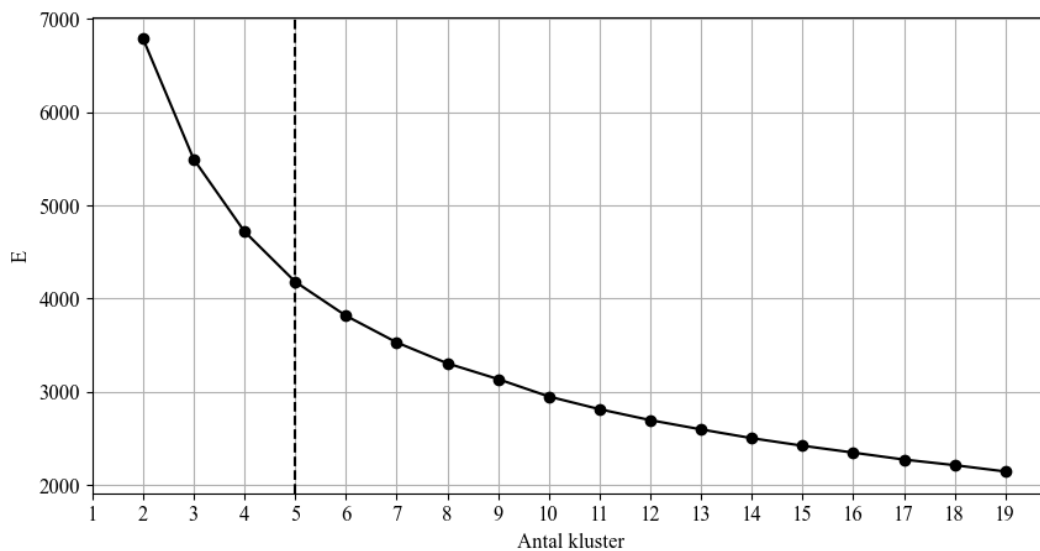
3.1 Klustring av beräkningsytor

3.1.1 Kluster dominerade av gran

Detta delkapitel presenterar utfallet för klusterindelning av beräkningsytor med andel gran >50 %.

Utvärdering med armbågsmetoden

Utvärderingen av antal kluster framtagna genom armbågsmetoden resulterade i ett rekommenderat klusterantal enligt Figur 3. Enligt denna analys vore det bäst att skapa 5 kluster för att beskriva olika typer av beräkningsytor.

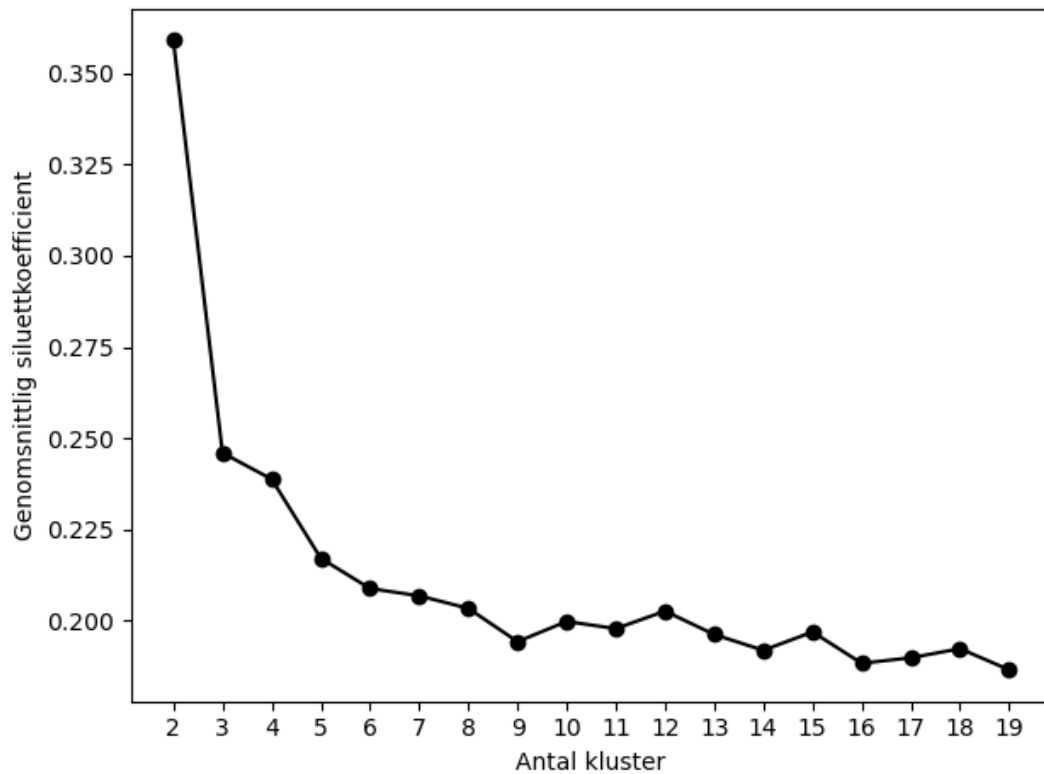


Figur 3. E-värde per antal kluster för beräkningsytor dominerade av gran. Optimalt antal kluster enligt analysen anges med streckad linje.

Figure 3. E-value per number of clusters for spruce dominated areas. Optimal number of clusters are given by the dotted line.

Utvärdering med siluettmetoden

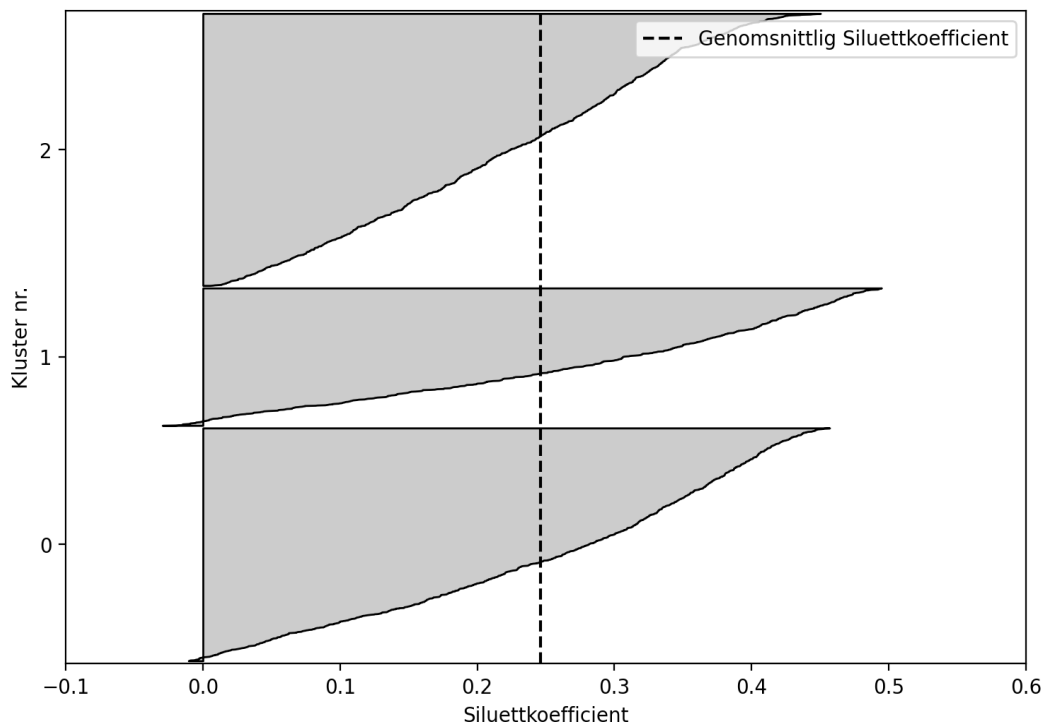
Analysen av den genomsnittliga siluettkoefficienten för olika antal kluster visade att det optimala antalet kluster avvek mot det resultat som erhöles från tidigare analys med armbågsmetoden. Högsta siluettkoefficient och således även rekommenderat antal kluster visade sig med denna metod vara $k=2$ kluster enligt Figur 4. Detta innebär att tätast och mest särskilda kluster skapas om antalet kluster bestäms vara 2.



Figur 4. Genomsnittlig siluettkoefficient per antal kluster för beräkningsytor dominerade av gran.
Figure 4. Mean silhouette coefficient per number of clusters for spruce dominated areas.

I Figur 5 visualiseras siluettkoefficienten för klusterantal $k=3$. Denna kompletteras av bilaga 4 (Figur 17) som visualiserar siluettkoefficienten för $k=2$ till $k=5$ kluster. Det är tydligt att två kluster ger en indelning som innebär en mindre osäker klassificering (Figur 17). Detta på grund av att endast ett fåtal observationer har en siluettkoefficient under 0. Samtidigt finns det fler observationer med högre siluettkoefficient än vid andra klusterantal vilket resulterar i ett högre genomsnittligt värde för siluettkoefficienten.

Storleken på de grå ytorna i graferna i Figur 5 & 17 vittnar även om förhållandet mellan mängden observationer i varje kluster och antal kluster. Yornas maximala höjd i Y-led samvarierar här direkt med antalet observationer i varje kluster. Det var här tydligt att tre kluster gav två större kluster och ett mindre kluster (Figur 5).

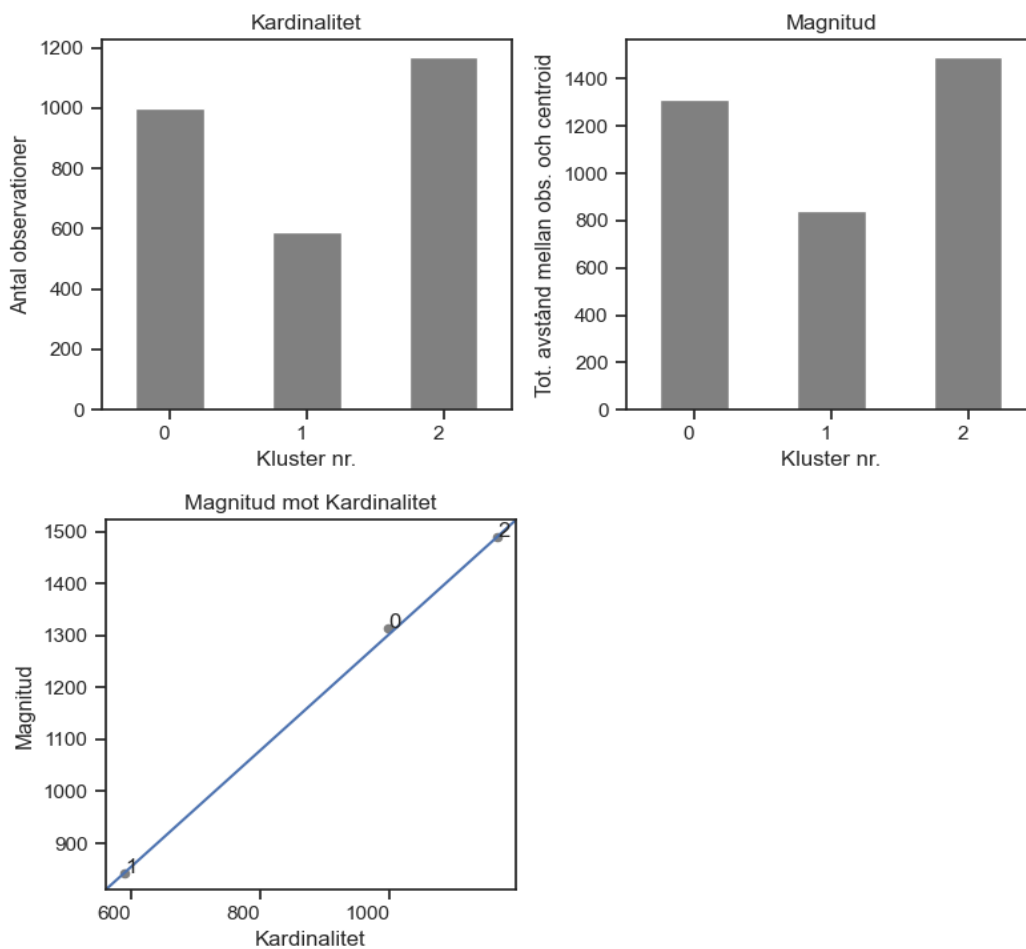


Figur 5. Siluettgrafer för $k=3$ kluster av grandominerade beräkningsytor.

Figure 5. Silhouette graphs for $k=3$ clusters of spruce dominated areas.

Utvärdering av kardinalitet och magnitud

Det slutliga valet blev att använda tre kluster för att beskriva de grandominerade bestånden. Detta efter att den praktiska användningen utvärderats mot resultatet från analys av E-värdet och siluettkoefficienten. Analysen av klustrens kardinalitet och magnitud påvisade ett godkänt samband mellan antal observationer per kluster och respektive klusters spridning av observationer. Detta samband gestaltas av att punkterna för kluster 0–2 alla ligger på eller nära linjen som beskriver önskat samband (Figur 6).



Figur 6. Jämförelse av kardinalitet och magnitud för $k=3$ kluster för grandominerade beräkningsytor.

Figure 6. Comparison of cardinality and magnitude for $k=3$ clusters for spruce dominated areas.

Slutlig klusterindelning

De tre apteringsgrupperna och dess egenskaper presenteras i tabell 9. Apteringsgrupperna karaktäriserades främst av differenser i medelstamsvolym. Gruppen ”Gran 0” innehöll beräkningsytor med genomsnittligt lägre medelstamsvolym medan gruppen ”Gran 1” innehöll beräkningsytor med genomsnittligt högre medelstamsvolym. Gruppen ”Gran 2” hade ett genomsnittligt värde mellan de andra två grupperna. Beräkningsytor med mindre träd visades förekomma på högre altitud och vid en lägre temperatursumma. Det genomsnittliga beräknade ståndortsindexet visade detta mönster och att boniteten följaktligen var lägre för beräkningsytor med genomsnittligt lägre medelstamsvolym.

Tabell 9. Genomsnittliga beståndsegenskaper för apteringsgrupper dominerade av gran. Värden inom parentes anger standardavvikelse för given variabel
Table 9. Average values for site properties for bucking groups dominated by spruce. Values within parentheses denote the standard deviation for each variable

Apteringsgrupp	Gran 0	Gran 1	Gran 2
Antal beräkningsytor	999	590	1167
Antal virkesordrar	432	247	516
Ståndortsindex*	G 19 (2)	G 25 (2)	G 22 (2)
Ålder*	110 (5)	108 (8)	108 (6)
DGV (cm)	21,5 (1,6)	29,5 (1,9)	25,1 (1,5)
HGV (m)	17,6 (1,3)	23,2 (1,4)	20,4 (1,1)
Formkvot	0,63 (0,02)	0,63 (0,02)	0,64 (0,02)
Beståndstäther (stammar/ha)	983 (283)	645 (172)	863 (239)
Volym per ytenhet (m³fub/ha)	172,2 (45,1)	282,2 (62,1)	235,0 (52,6)
Grundyta (m²)	25,5 (5,5)	32,3 (6,3)	30,0 (5,8)
Medelstamsvolym (m³fub)	0,17 (0,03)	0,43 (0,08)	0,27 (0,04)
Höjd över havet (m)	307,0 (100,9)	279,1 (125,3)	286,6 (111,8)
Temperatursumma (dygnsgrader)	844,5 (98,3)	873,2 (118,7)	868,8 (102,9)
Andel gran (%)	72,6 (12,9)	82,3 (12,8)	75,9 (13,6)
Andel tall (%)	15,9 (13,5)	12,2 (12,6)	15,1 (13,2)
Andel övrigt (%)	11,5 (8,5)	5,5 (6,2)	9,0 (7,6)
Medelarea per beräkningsyta (ha)	1,2 (0,4)	1,3 (0,4)	1,3 (0,4)

* Ålder och SI beräknad från skördardata.

Analysen av stamegenskaper i de tre klustren visade att stammar av tall konsekvent hade en genomsnittligt större diameter och höjd. Samtidigt hade tallarna även en högre formkvot. Standardavvikelse indikerade att det fanns en större variation av stamhöjder och diametrar i apteringsgrupp ”Gran 1” och ”Gran 2” avseende granstammar jämfört med de andra apteringsgrupperna och värden för tall (tabell 10).

Tabell 10. Genomsnittliga stamegenskaper för tall och gran i grandominerade kluster. Värden inom parentes anger standardavvikelse per variabel

Table 10. Average stem properties for pine and spruce in spruce dominated clusters. Values within parentheses denote the standard deviation per variable

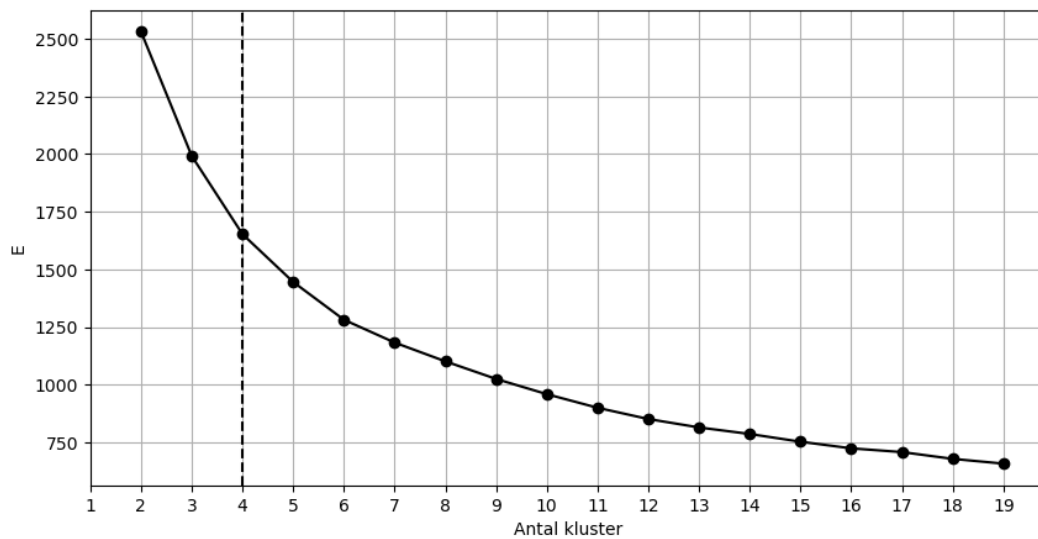
Trädslag	Stamegenskap	Apteringsgrupp		
		Gran 0	Gran 1	Gran 2
Tall	DBH (mm)	218 (75)	284 (81)	247 (81)
	Höjd (cm)	1704 (401)	2191 (463)	1963 (454)
	Formkvot	0,67 (0,05)	0,68 (0,05)	0,68 (0,05)
	Stamantal	134 507	29 213	114 776
Gran	DBH (mm)	163 (64)	232 (91)	191 (76)
	Höjd (cm)	1433 (365)	1925 (515)	1665 (440)
	Formkvot	0,63 (0,05)	0,63 (0,05)	0,63 (0,05)
	Stamantal	980 690	424 884	1 077 406

3.1.2 Kluster dominerade av tall

Detta delkapitel presenterar utfallet för klusterindelning av beräkningsytor med andel tall >50 %.

Utvärdering med armbågsmetoden

Utvärderingen av antal kluster genom armbågsmetoden resulterade i ett rekommenderat klusterantal enligt Figur 7 med tillhörande E-värde. Enligt denna analys var det bäst att skapa 4 kluster för att beskriva olika typer av beräkningsytor.

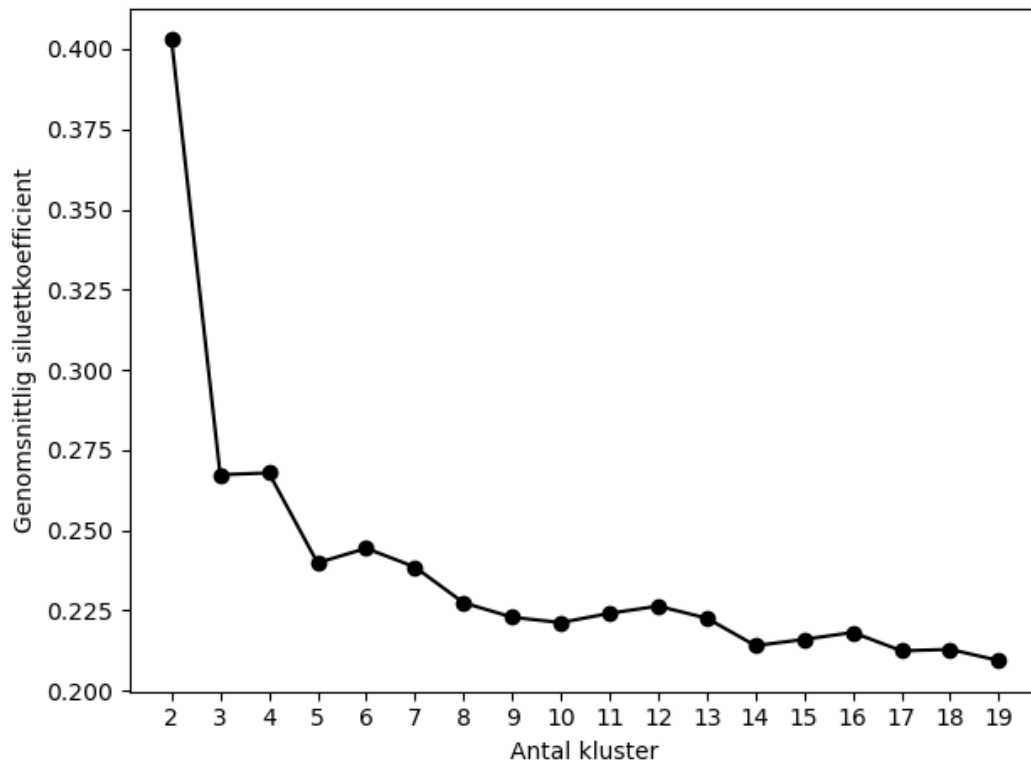


Figur 7. E-värde per antal kluster för beräkningsytor dominerade av tall. Optimalt antal kluster enligt analysen anges med streckad linje.

Figure 7. E-value per number of clusters. Optimal number of clusters are given by the dotted line.

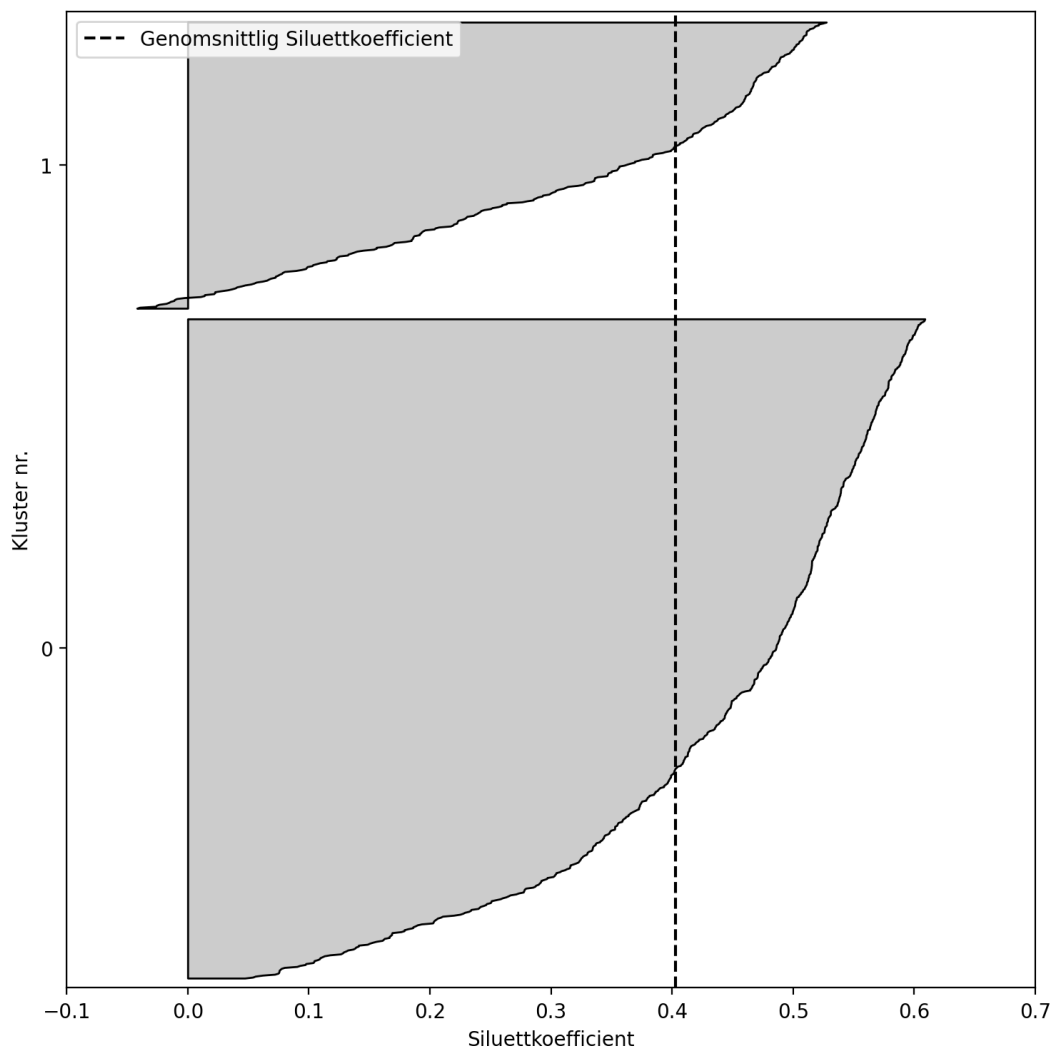
Utvärdering med siluettmetoden

Analysen av den genomsnittliga siluettkoefficienten för olika antal kluster visade att det optimala antalet kluster avvek mot det resultat som erhållits från tidigare analys med armbågsmetoden. Högsta siluettkoefficient och således även rekommenderat antal kluster visade sig med denna metod vara $k=2$ kluster enligt Figur 8. Detta innebär att tätast och mest särskilda kluster skapas om antalet kluster bestäms vara 2.



Figur 8. Genomsnittlig siluettkoefficient per antal kluster för beräkningsytor dominerade av tall.
Figure 8. Mean silhouette coefficient per number of clusters for pine dominated areas.

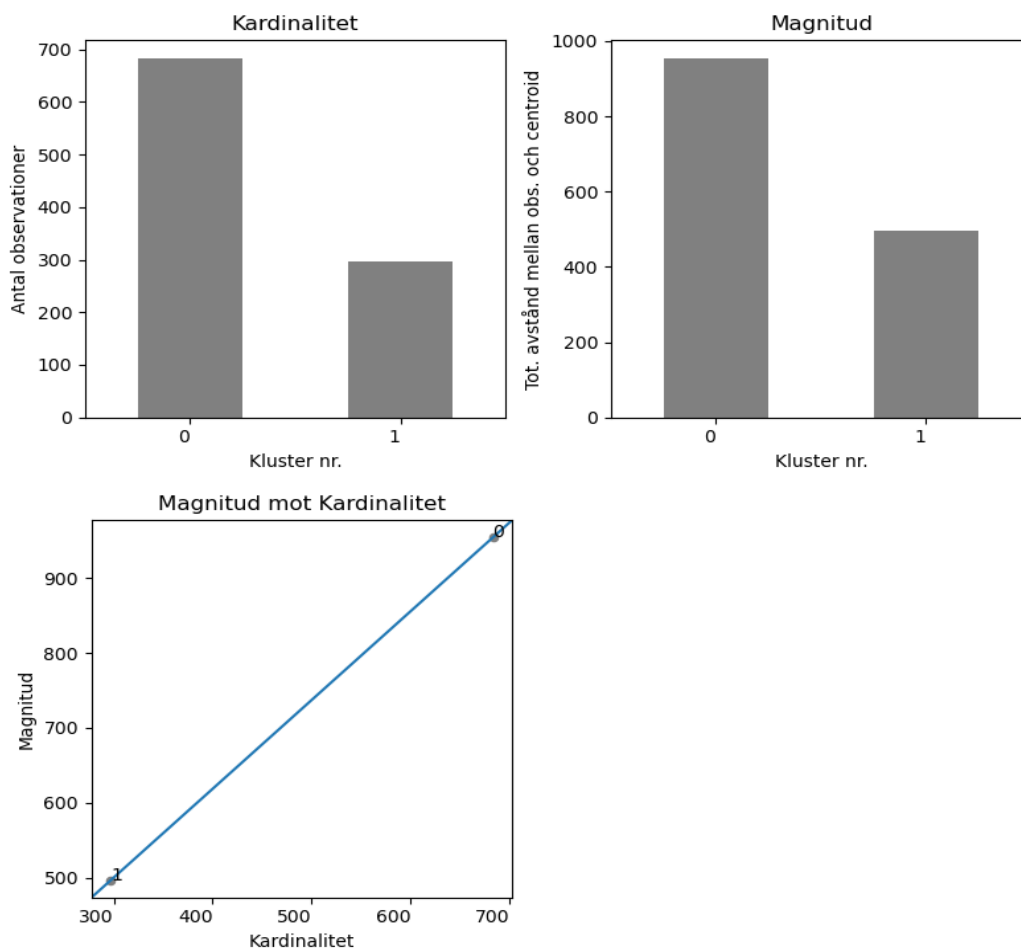
I Figur 9 visualiseras siluettkoefficienten för alla observationer för $k=2$ kluster. Figuren kompletteras med bilaga 4 (Figur 18) som visualiserar $k=2$ till $k=4$ kluster. Likt för grandominerade beräkningsytor gav två kluster den bästa indelningen sett till den genomsnittliga siluettkoefficienten. Två kluster resulterade även i ett stort och ett mindre kluster. Tre respektive fyra kluster resulterade i mer fraktionerade kluster och ett större klustringsfel (Figur 18). Två kluster gav ett litet klustringsfel för ett litet antal observationer men en genomsnittligt högre siluettkoefficient (Figur 9).



Figur 9. Siluettgrafer för $k=2$ kluster av talldominerade beräkningsytor.
Figure 9. Silhouette graphs for $k=2$ clusters of pine dominated areas.

Utvärdering av kardinalitet och magnitud

Det slutliga valet blev att använda två apertingsgrupper för att beskriva de talldominerade bestånden. Detta efter att den praktiska användningen utvärderats mot resultatet från analys av E-värdet och siluettkoefficienten. Analysen av klustrens kardinalitet och magnitud påvisade ett godkänt samband mellan antal observationer per kluster och respektive klusters spridning av observationer. Vid användning av tre eller fyra kluster kunde detta krav på samband inte tillfredsställas. Sambandet gestaltas i Figur 10.



Figur 10. Jämförelse av kardinalitet och magnitud för $k=3$ kluster för talldominerade beräkningsytor.

Figure 10. Comparison of cardinality and magnitude for $k=3$ clusters for pine dominated areas.

Slutlig klusterindelning

Likt vid klustringen av grandominerade beräkningsytor visade sig trädslagsandelarna ha liten påverkan på resultatet. Till följd av att dem ändå påverkade klustringen genom att skapa en ytterligare särindelning av de olika klustren inkluderades ändå dessa beslutsvariabler. En visualisering av resultatet för klustringen av talldominerade beräkningsytor presenteras i Bilaga 2. De två apteringsgrupperna och dess egenskaper presenteras i tabell 11.

Tabell 11. Genomsnittliga beståndsegenskaper för apteringsgrupper dominerade av tall. Värden inom parentes anger standardavvikelse för given variabel

Table 11. Average values for site properties for bucking groups dominated by pine. Values within parentheses denote the standard deviation for each variable

Apteringsgrupp	Tall 3	Tall 4
Antal beräkningsytor	683	297
Antal virkesordrar	276	124
Ståndortsindex*	T20 (2)	T24 (2)
Ålder*	111 (2)	111 (4)
DGV (cm)	23,0 (2,0)	28,3 (2,4)
HGV (m)	18,4 (1,6)	22,4 (1,7)
Formkvot	0,65 (0,02)	0,67 (0,02)
Stammar per hektar	775 (262)	512 (193)
Volym per ytenhet (m³fub/ha)	172,3 (52)	222,2 (62)
Grundyta (m²)	23,4 (5,8)	24,7 (6,3)
Medelstamsvolym (m³fub)	0,22 (0,05)	0,45 (0,11)
Höjd över havet (m)	282,1 (69,8)	273,0 (75,6)
Tempsumma	866,2 (69,1)	876,9 (67,7)
Andel gran (%)	24,4 (13,6)	19,8 (14,2)
Andel tall (%)	71,8 (14,7)	77,9 (14,9)
Andel övriga trädslag (%)	3,9 (4,1)	2,3 (3,7)
Area medel (ha)	1,3 (0,4)	1,3 (0,4)

* Ålder och SI beräknad från skördardata.

Likt tidigare observerat för grandominerade beräkningsytor visade tallarna i denna grupp i genomsnitt en större diameter, högre stamhöjd och en högre formkvot (tabell 12).

Tabell 12. Genomsnittliga stamegenskaper för tall och gran i talldominerade apteringsgrupper. Värden inom parentes anger standardavvikelse per variabel

Table 12. Average stem properties for pine and spruce in pine dominated bucking groups. Values within parentheses denote the standard deviation per variable

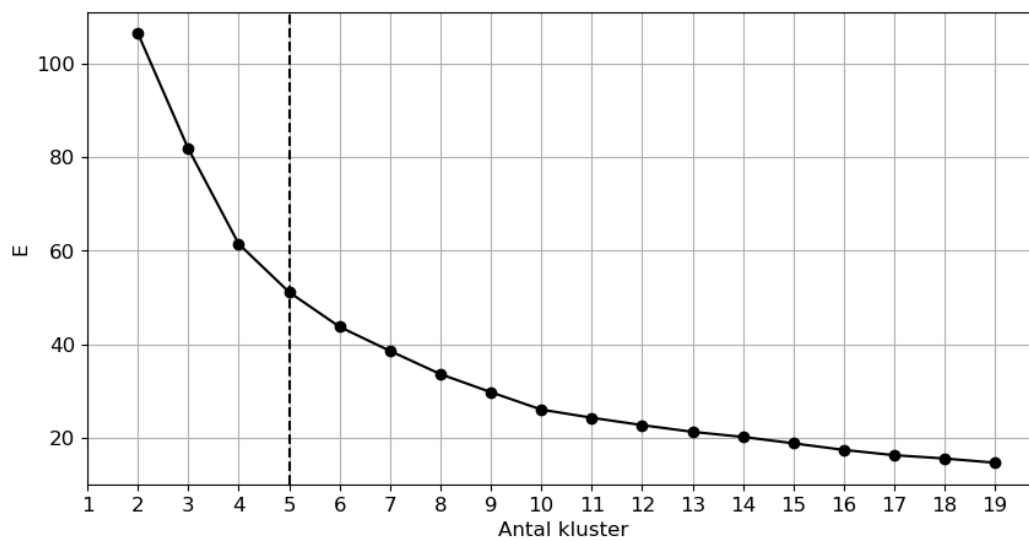
Trädslag	Stamegenskap	Apteringsgrupp	
		Tall 3	Tall 4
Tall	DBH (mm)	206 (69)	264 (72)
	Höjd (cm)	1728 (407)	2154 (408)
	Formkvot	0,67 (0,05)	0,69 (0,04)
	Stamantal	377 754	120 442
Gran	DBH (mm)	153 (59)	187 (75)
	Höjd (cm)	1373 (435)	1634 (577)
	Formkvot	0,63 (0,05)	0,63 (0,05)
	Stamantal	281 130	71 169

3.1.3 Kluster dominerade av övriga trädslag

Detta delkapitel presenterar utfallet för klusterindelning av beräkningsytor med andel övriga trädslag >50 %.

Utvärdering med armbågsmetoden

Utvärderingen av antal kluster genom armbågsmetoden resulterade i ett rekommenderat klusterantal enligt Figur 11 med tillhörande E-värde. Enligt denna analys vore det bäst att skapa 5 kluster för att beskriva olika typer av beräkningsytor.

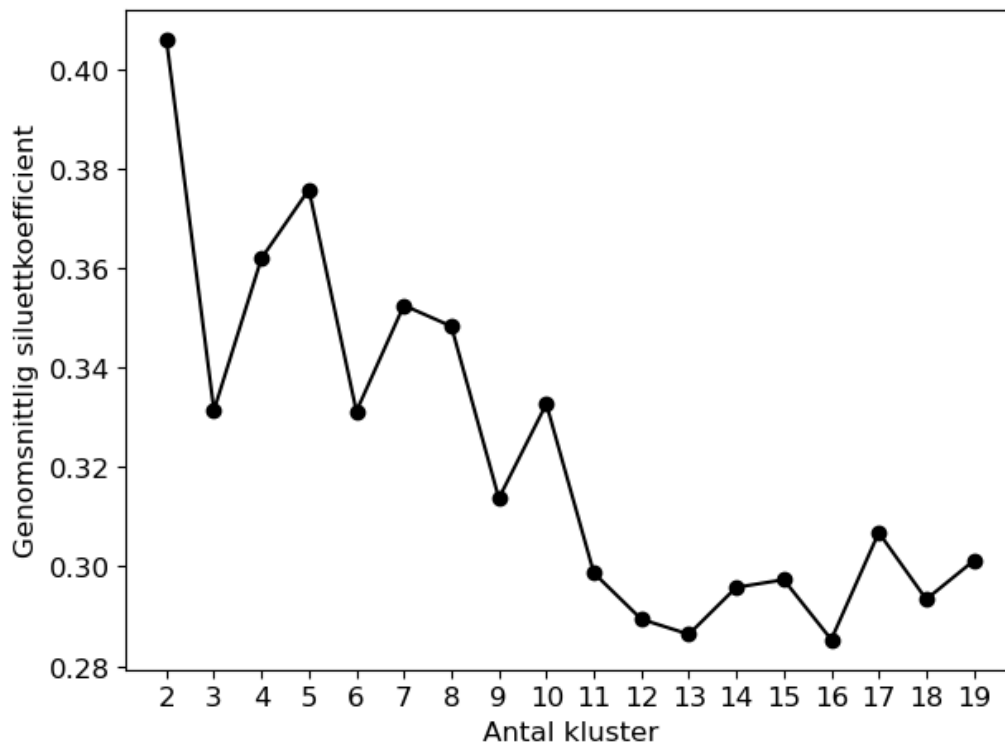


Figur 11. E-värde per antal kluster för beräkningsytor dominerade av övriga trädslag. Optimalt antal kluster enligt analysen anges med streckad linje.

Figure 11. E-value per number of clusters for areas dominated by other tree species. Optimal number of clusters are given by the dotted line.

Utvärdering med siluettmetoden

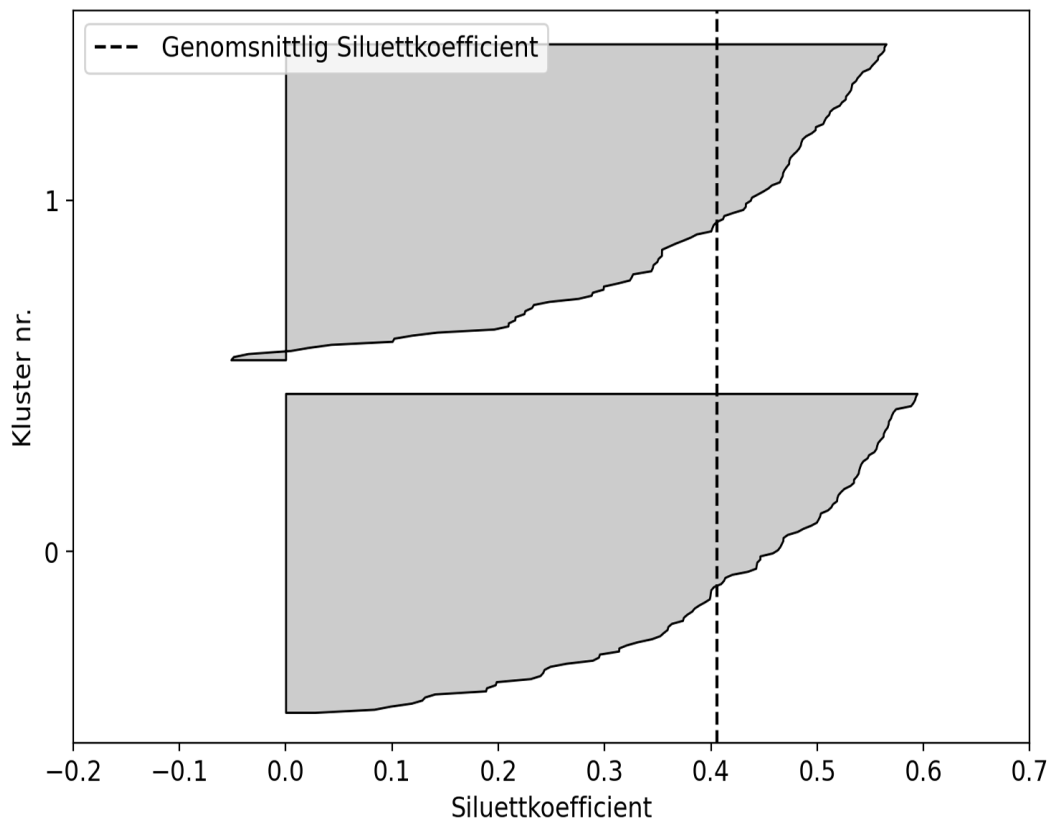
Analysen av den genomsnittliga siluettkoefficienten för olika antal kluster visade att det optimala antalet kluster avvek mot det resultat som erhöles från tidigare analys med armbågsmetoden. Högsta siluettkoefficient och således även rekommenderat antal kluster visade sig med denna metod vara $k=2$ kluster enligt Figur 12. Detta innebär att tätast och mest särskilda kluster skapas om antalet kluster bestäms vara 2.



Figur 12. Genomsnittlig siluettkoefficient per antal kluster för beräkningsytor dominerade av övriga trädslag.

Figure 12. Mean silhouette coefficient per number of clusters for areas dominated by other tree species.

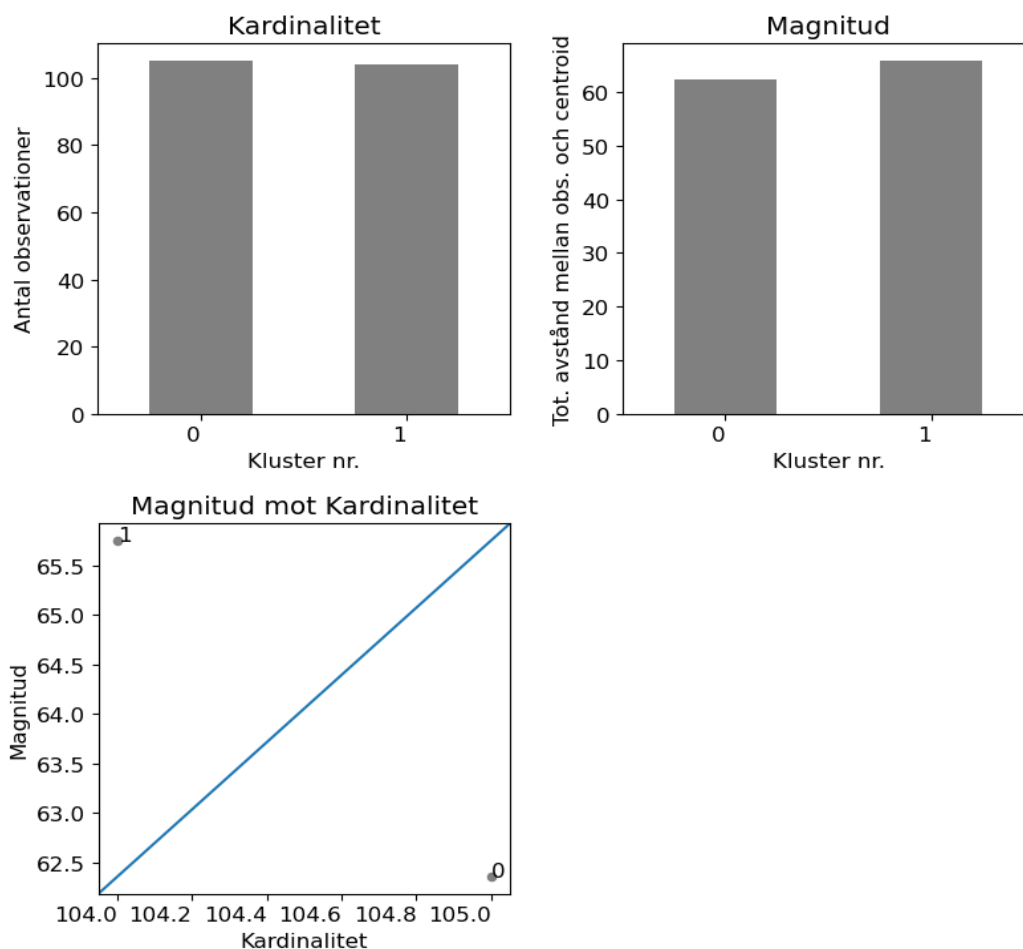
I Figur 13 visualiseras siluettkoefficienten för alla observationer för $k=2$ kluster. Figuren kompletteras med bilaga 6 (Figur 19) som visualiserar siluettkoefficienten för $k=2$ till $k=5$ kluster. Två kluster resulterade för denna gruppen i två jämstora kluster (Figur 13). Tre till fem kluster resulterar i mer varierande klusterstorlekar och ett större klustringsfel. För denna grupp visade sig siluettkoefficienten fluktuerar mycket vid val av olika mängd kluster. Detta syns tydligt i Figur 19 där en tydlig minskning av siluettkoefficienten syns från två till tre kluster men att denna sen stiger igen vid fyra och fem kluster.



Figur 13. Siluettgrafer för $k=2$ kluster av beräkningsytor dominerade av övriga trädslag.
Figure 13. Silhouette graphs for $k=2$ clusters of areas dominated by other tree species.

Utvärdering av kardinalitet och magnitud

Det slutliga valet blev att använda en apertingsgrupp för att beskriva beräkningsytor dominerade av övriga trädslag. Alltså att inte dela upp denna grupp i kluster. Detta efter att den praktiska användningen utvärderats mot resultatet från analys av E-värdet och siluettkoefficienten. Analysen av klustrens kardinalitet och magnitud påvisade inte ett godkänt samband mellan antal observationer per kluster och respektive klusters spridning av observationer vid användning av mer än ett kluster. Sambandet vid användning av två kluster presenteras i Figur 14 där det är tydligt hur klustren inte visar önskat samband då magnitud och kardinalitet jämförs.



Figur 14. Jämförelse av kardinalitet och magnitud för k=2 kluster för beräkningsytor dominerade av övriga trädslag.

Figure 14. Comparison of cardinality and magnitude for k=2 clusters for areas dominated by other tree species.

Slutlig klusterindelning

Likt vid klustringen av grupperna med tall- respektive grandominerade beräkningsytor visade sig trädslagsandelarna ha liten påverkan på resultatet. Till följd av att de ändå påverkade klustringen genom att skapa en ytterligare särindelning av de olika klustren inkluderade klustringen ändå dessa beslutsvariabler. Värden för gruppens egenskaper presenteras i tabell 13.

Tabell 13. Genomsnittliga beståndsegenskaper för apteringsgrupper dominerade av övriga trädslag. Värden inom parentes anger standardavvikelse för given variabel
Table 13. Average values for site properties for bucking groups dominated by other tree species. Values within parentheses denote the standard deviation for each variable

Apteringsgrupp	Övrigt 5
Antal beräkningsytor	209
Antal virkesordrar	44
Ståndortsindex*	G19 (2)
Ålder*	110 (5)
DGV (cm)	22,1 (2,1)
HGV (m)	17,2 (1,5)
Formkvot	0,65 (0,02)
Stammar per hektar	790 (273)
Volym per ytenhet (m³fub/ha)	165,0 (42,5)
Grundyta (m²)	24,0 (5,5)
Medelstamsvolym (m³fub)	0,21 (0,05)
Höjd över havet (m)	320,2 (105,1)
Temperatursumma (Dygnsgrader)	849,5 (94,9)
Andel gran (%)	13,5 (13,0)
Andel tall (%)	1,6 (4,0)
Andel övrigt (%)	84,9 (14,3)
Area medel (ha)	1,2 (0,3)

* Ålder och SI beräknad från skördardata.

I denna grupp visade stamegenskaperna att granar och tallar på denna typ av beräkningsytor är mer lika varandra sett till DBH och stamhöjd (tabell 14). Tidigare identifierat mönster gäller dock även här, de stammar av tall som förekommer är i genomsnitt större än granarna. Stamantalet för övriga trädslag i denna grupp var 76 898 st. utan angiven art.

Tabell 14. Genomsnittliga stamegenskaper för tall och gran i apteringsgrupper dominerade av övriga trädslag. Värden inom parentes anger standardavvikelse per variabel

Table 14. Average stem properties for pine and spruce in bucking groups dominated by other tree species. Values within parentheses denote the standard deviation per variable

Trädslag	Stamegenskap	Apteringsgrupp
		Övrigt 5
Tall	DBH (mm)	184 (62)
	Höjd (cm)	1550 (357)
	Formkvot	0,66 (0,06)
	Stamantal	2 104
Gran	DBH (mm)	181,96 (59)
	Höjd (cm)	1509 (343)
	Formkvot	0,62 (0,05)
	Stamantal	20 181

3.1.4 Kluster med blandad trädslagsfördelning

Detta delkapitel presenterar utfallet för klusterindelning av beräkningsytor med blandad trädslagsfördelning. Denna grupp delades upp utefter angivet trädslag i ståndortsindex. Indelning i tall- respektive grandominerade beräkningsytor med blandad trädslagsfördelning blev den slutliga klusterindelningen för blandade skogar. Detta efter att E-värde och siluettkoefficient utvärderats och visat att vidare klustring av de två grupperna inte kunde göras med goda resultat. Detta till följd av att denna typ av skogar visade sig vara relativt ovanliga. Det totala antalet observationer i gruppen var 100 talldominerade beräkningsytor och 142 grandominerade ytor vilket är relativt få jämfört med andra grupper. Till följd av att klustring inte utförts efter den manuella grupperingen blir presentation av delar kopplade till val av antal kluster och detaljer om klusteringdelningen redundant, varför detta utelämnas ur rapporten.

De två apteringsgrupperna av blandskogar presenteras med nyckeltal i tabell 15.

Tabell 15. Genomsnittliga beståndsegenskaper för apteringsgrupper med blandad trädslagsfördelning. Värden inom parentes anger standardavvikelse för given variabel
Table 15. Average values for site properties for bucking groups with mixed tree species distribution. Values within parentheses denote the standard deviation for each variable

Apteringsgrupp	Blandat G 6	Blandat T 7
Antal beräkningsytor	142	100
Antal virkesordrar	112	73
Ståndortsindex*	G 20 (3,0)	T 20 (2,7)
Ålder*	108 (5,0)	112 (3,4)
DGV (cm)	23,3 (2,9)	23,2 (3,0)
HGV (m)	18,8 (2,3)	18,6 (2,1)
Formkvot	0,64 (0,02)	0,64 (0,01)
Stammar per hektar	895 (299)	899 (368)
Volym per ytenhet (m³fub/ha)	200,3 (64,3)	192,3 (64,1)
Grundyta (m²)	27,3 (6,4)	26,5 (6,9)
Medelstamsvolym (m³fub)	0,22 (0,08)	0,22 (0,08)
Höjd över havet	282,5 (95,7)	305,0 (69,4)
Temperatursumma	865,7 (92,1)	842,3 (68,9)
Andel gran (%)	46,4 (3,0)	41,1 (5,6)
Andel tall (%)	36,9 (9,9)	46,6 (2,9)
Andel övrigt (%)	16,7 (11,3)	12,3 (7,6)
Medelarea per beräkningsyta (ha)	1,2 (0,4)	1,2 (0,4)

* Ålder och SI beräknad från skördardata.

I denna grupp visade stamegenskaperna att granar och tallar på denna typ av beräkningsytor följer en tidigare identifierad trend där tallarna i genomsnitt är större än granarna. I denna typ av skogar är gran mer vanligt förekommande sett till stamantal. En detaljerad beskrivning av stamegenskaperna presenteras i tabell 16.

Tabell 16. Genomsnittliga stamegenskaper för tall och gran i apteringsgrupper med blandad trädslagsfördelning. Värden inom parentes anger standardavvikelse per variabel

Table 16. Average stem properties for pine and spruce in bucking groups with mixed tree species distribution. Values within parentheses denote the standard deviation per variable

Trädslag	Stamegenskap	Apteringsgrupp	
		Blandat G 6	Blandat T 7
Tall	DBH (mm)	235 (72)	229 (69)
	Höjd (cm)	1900 (388)	1866 (379)
	Formkvot	0,68 (0,05)	0,68 (0,05)
	Stamantal	25 555	23 169
Gran	DBH (mm)	185 (62)	182 (60)
	Höjd (cm)	1614 (423)	1587 (411)
	Formkvot	0,63 (0,06)	0,63 (0,06)
	Stamantal	57 867	36 588

3.2 Apteringsanalys

Totalt apterades 70 201 stammar under apteringssimuleringen. Stamantalet är ett resultat av det slumpmässiga urvalet av beräkningsytor för konstruktion av stambankerna. Vid jämförelse med referensmaterialet från SCA som legat till grund för stambankerna visade sig medelstamsvolym och HGV i genomsnitt vara högre i referensmaterialet. DGV var däremot i genomsnitt lägre i referensmaterialet. Generellt sätt ligger värden för stambankerna nära de värden som hämtats från källmaterialet avseende givna beståndsvariabler. En detaljerad jämförelse presenteras i tabell 17.

Tabell 17. Stamantal per stambank och jämförelse av stambanksdata (SBD) och källmaterialet (KM) för ett urval av beståndsvariabler

Table 17. Number of stems per stem bank and comparison of stem bank data (SBD) and source data (KM) for a selection of stand variables

Stambank	Stamantal (n)	Medelstamsvolym (m ³ fub)		DGV (mm)		HGV (m)	
		SBD	KM	SBD	KM	SBD	KM
Gran 0	10 745	0,16	0,17	218	215	17,0	17,5
Gran 1	7 122	0,42	0,43	298	295	23,2	23,1
Gran 2	9 325	0,26	0,27	253	250	20,1	20,4
Tall 3	8 317	0,21	0,22	232	230	17,8	18,4
Tall 4	5 193	0,42	0,45	286	284	22,2	22,4
Övrigt 5	9 576	0,18	0,21	221	221	18,2	17,2
Blandat G 6	10 151	0,2	0,22	236	233	18,5	18,8
Blandat T 7	9 772	0,2	0,22	235	232	18,1	18,6
Total	70 201	0,24	0,26	247	245	19,4	19,6

3.2.1 Timmerandel

Apteringsanalyserna visade genomgående att styrning mot längre timmer presterar bättre sett till att maximera timmerutfallet (tabell 18). Apteringsgrupp 5–7 genererade genomsnittligt lägre timmerandelar än alla andra grupper bortsett från grupp 0. För grupp 5 förklaras detta av att det inte har apterats något timmer i övriga trädslag. Istället gav exempelvis contorta enbart utfall i form av massaved.

Resultatet indikerade också att timmerandelarna varierar mindre ju högre medelstamsvolymen är vid jämförelse av kort och lång instruktion. För grupper dominerade av gran varierade denna skillnad mellan 1 % för den grupp som hade lägst medelstamsvolym och 0,8% för den grupp som hade högst medelstamsvolym. Avseende grupper dominerade av tall var skillnaden större. Här var skillnaden 1,8 % för den grupp som hade lägst medelstamsvolym och 0,8 % för den med högst medelstamsvolym.

Mellan apteringsgrupper var skillnaden i timmerandel också lägre vid aptering mot långt timmer. För grupper dominerade av gran var skillnaden mellan grupper ca 1,5 % lägre vid styrning mot långt timmer. Motsvarande värde för grupper dominerade av tall var 1 %. Andelen timmer varierade alltså mindre mellan apteringsgrupperna vid styrning mot långt timmer.

Tabell 18. Resultaterande timmerandel och egenskaper per apteringsgrupp och apteringsinstruktion. Fetmarkerat anger högsta timmerandel per grupp. Avser total timmerandel summerat över alla trädslag

Table 18. Resulting sawlog rate (%) per bucking group and bucking instruction. Bold denotes highest sawlog rate per cluster. Rates are given for the total production of timber irrespective of tree species

Apterings- grupp	Timmerandel (%) per apteringsinstruktion			Egenskaper		
	Kort	Neutral	Lång	DGV (cm)	HGV (m)	Medelstams -volym (m ³ fub)
Gran 0	49,5	50,1	50,5	21,51	17,55	0,17
Gran 1	71,4	71,8	72,2	29,53	23,18	0,43
Gran 2	61,1	61,6	62,0	25,06	20,43	0,27
Tall 3	63,0	64,0	64,8	22,98	18,36	0,22
Tall 4	79,0	79,3	79,8	28,35	22,35	0,45
Övrigt 5	11,6	11,6	11,8	22,08	17,16	0,21
Blandat G	54,3	54,8	55,3	23,33	18,85	0,22
6						
Blandat T	57,3	57,9	58,4	23,22	18,64	0,22
7						

3.2.2 Fördelningsgrad

När fördelningsapatering användes i instruktionen för korta, respektive långa längder gav den långa instruktionen aldrig den bästa fördelningsgraden för varken gran eller tall (tabell 19). Den neutrala instruktionen visade sig här vara den bästa för att uppnå högsta uppfyllnad av fördelningsönskemålet. Skillnaden i fördelningsgrad var i regel liten mellan kort och neutral instruktion med en genomsnittligt större minskning för lång lista. Längre och grövre skog identifierades ofta prestera bättre med styrning mot korta längder. Den genomsnittliga fördelningsgraden över alla apateringsgrupper var högst för styrning mot korta längder avseende tall. För gran gav neutral längdstyrning den genomsnittligt högsta fördelningsgraden.

Grupper med högre medelstamsvolym gav en lägre maximal skillnad i fördelningsgrad mellan apateringarna avseende huvudträdslaget. För grupper dominerade av gran var de maximala skillnaderna 13 %, 7 % och 4 % (Gran 0, Gran 2 & Gran 1) ordnat från låg till hög medelstamsvolym. För grupper dominerade av tall var de maximala skillnaderna 3 % och 5 % (Tall 3 och Tall 4) ordnat från låg till hög medelstamsvolym.

Tabell 19. Resultande fördelningsgrad för olika apateringsinstruktioner avrundat till närmsta heltal. Uppdelat för talltimmer (T) och grantimmer (G). Fetmarkerat anger bästa fördelningsgrad per apateringsgrupp och trädslag

Table 19. Resulting distribution rate for different bucking instructions rounded to nearest whole number. Divided for pine timber (T) and spruce timber (G). Bold denotes highest distribution rate per bucking group and tree species

Apteringsgrupp	Kort T	Neutral T	Lång T	Kort G	Neutral G	Lång G
Gran 0	80	81	72	80	84	71
Gran 1	88	82	78	85	83	81
Gran 2	85	84	79	82	84	77
Tall 3	79	84	80	78	83	73
Tall 4	84	84	81	80	83	80
Övrigt 5	83	71	67	86	78	68
Blandat G 6	83	84	78	81	83	74
Blandat T 7	82	85	79	82	85	74
Genomsnittlig fördelningsgrad	83	82	77	82	83	75

3.2.3 Stockegenskaper

Stockegenskaper analyserades för timmer med fokus på stocklängd. Apteringsmodelleringen visade att klen, kort skog producerar det längsta timret vid styrning mot korta längder. Samtidigt visades att samma typ av skog producerar de kortaste stockarna vid styrning mot neutral längdfördelning eller långa längder (Tabell 20).

Tabell 20. Genomsnittliga timmerlängder (cm) per apteringsgrupp och apteringsinstruktion. Kompletterat med egenskaper kopplade till respektive grupp

Table 20. Mean sawlog length per bucking group and bucking instruction. Completed with characteristics linked to each group

Apteringsgrupp	Apteringsinstruktion			Egenskaper			
	Kort	Neutral	Lång	DGV	HGV	Medelstamsvol.	Formkvot
Gran 0	412	441	466	21,51	17,55	0,17	0,63
Gran 1	405	447	479	29,53	23,18	0,43	0,63
Gran 2	409	447	475	25,06	20,43	0,27	0,64
Tall 3	414	454	475	22,98	18,36	0,22	0,65
Tall 4	409	454	479	28,35	22,35	0,45	0,67
Övrigt 5	403	432	461	22,08	17,16	0,21	0,65
Blandat G 6	408	448	472	23,33	18,85	0,22	0,64
Blandat T 7	409	449	473	23,22	18,64	0,22	0,64

3.2.4 Stamfelsesved

Utfallen av stamfelsesved för gran och tall i de olika klustren jämfördes med nyckeltal från källmaterialet. Stamfelsandelen för det simulerade utfallet varierade mellan att vara högre och lägre än andelarna i källmaterialet per trädslag. De största andelarna stamfelsesved återfanns vid styrning mot korta längder avseende tall och neutral längdstyrning avseende gran. Styrning mot långt timmer gav i regel den minsta andelen stamfelsesved. Denna tendens var tydligare för gran än tall. En utförlig beskrivning av andelen stamfelsesved med referensvärden presenteras i tabell 21.

Tabell 21. Stamfelsesvedsandelar (%) per apteringsinstruktion och grupp. Trädslag anges som T = Tall och G = Gran. Fetmarkerat anger högsta noterade stamfelsandel per apteringsgrupp och trädslag. Referensvärden (ref.) hämtade per trädslag från källmaterialet från SCA

Table 21. Share of logs with sawlog dimensions classified as any other product than sawlog timber (%). Bold denotes highest share per group and tree species. Reference values (ref.) fetched from the source material from SCA

Aptering sgrupp	Kort T	Neutral T	Lång T	Kort G	Neutral G	Lång G	Ref. T	Ref. G
Gran 0	7,9	7,1	6,7	10,7	10,7	10,8	5,3	12,9
Gran 1	6,9	6,9	6,5	12,2	12,3	12,0	4,8	11,8
Gran 2	9,2	8,2	7,4	11,4	11,4	11,2	4,6	12,2
Tall 3	7,3	6,2	5,7	7,0	7,5	7,0	2,8	8,7
Tall 4	5,6	5,9	5,3	8,6	9,0	8,6	2,0	8,0
Övr. 5	18,4	20,9	18,2	11,7	12,0	11,8	16,2	11,6
Blandat G 6	8,2	7,8	7,5	8,8	8,8	8,5	4,8	9,8
Blandat T 7	8,7	7,9	7,2	9,8	9,8	9,7	3,3	9,1

4. Diskussion

4.1 Studiens resultat

Grupperingen av beräkningsytor i kluster med liknande egenskaper visar att antalet möjliga kombinationer av egenskaper är stort. De styrvariabler som användes vid klustringen visade sig ha olika stor påverkan på resultatet. Genomsnittliga stamegenskaper som HGV, DGV och medelstamsvolym visade sig alltid ge de tydligaste indelningarna. Andra nyttjade egenskaper påverkade mindre. Detta beror sannolikt på den stora variation som finns för de andra variablerna. Vid parvisa jämförelser av variablerna är det tydligt att formkvot och trädslagsandelar varierar ungefär lika mycket oavsett värde på HGV, DGV och medelstamsvolym (Bilaga 1 & 2). För alla de 4 huvudgrupperna (Avsnitt 3.5) gav klustringen också ett överlapp avseende de mest styrande variablerna. Detta innebär att ett visst värde för HGV, DGV eller medelstamsvolym som identifieras i ett visst kluster också kan återfinnas i ett annat kluster. Naturligtvis gäller detta om variablerna studeras enskilt. Siluettgraferna kan användas för att avgöra hur unika klustren är då alla styrvariabler inkluderas. I de flesta fall finns observationer som klustrats fel. Antalet felaktiga klustertilldelningar tycks dock vara mycket lågt för de valda klusterantalen per grupp. Siluettanalysen visar också att den genomsnittliga siluettkoefficienten generellt är låg och närmare 0 än 1. Detta indikerar att klustren inte är helt tydligt åtskilda i ett sju-dimensionellt rum. Innebörden av detta är att algoritmen inte alltid kunnat placera varje beräkningsyta i ett enskilt kluster, i vissa fall finns det helt enkelt beräkningsytor som kan tilldelas två olika grupper med oförändrad avvikelse från gruppens medelvärde. De observationer som kan dela grupper eller som placerats i fel grupp är följaktligen långt ifrån alla gruppmedelvärden.

Resultatet av apteringsanalysen påvisar en komplex interaktion mellan olika mätvärden och att det är komplicerat att finna en optimal lösning för de olika värden som undersöks. Ett optimalt scenario vore att finna en lösning där timmerandelen ökar ihop med fördelningsgraden, samtidigt som andelen stamfelsved minskar. Resultatet visar istället att en ökad timmerandel ofta innebär en lägre andel stamfelsved men på bekostnad av en minskad fördelningsgrad. Viktigt att beakta är

att fördelningsönskemålen ändrades utan att beakta beståndsegenskaperna per kluster. Istället ställdes samma höga krav på långa respektive korta längder i alla typer av skog. Apteringsmodelleringen skulle alltså kunna göras än mer beståndsanpassad vid vidare studier, till skillnad från denna studie som fokuserat på generella mönster avseende längdstyrningens påverkan. Att beakta beståndsegenskaper vid bestämmandet av önskefördelningen försvårar också analysen då det inte går att bestämma skillnader mellan olika grupper. Utförd analys ger en första inblick i interaktionen mellan vald längdstyrning och beståndsegenskaper.

Studiens resultat är i linje med vad som redan visats av Kivinen (2007) som fann att fördelningsönskemålen uppnås i högre grad om timmerandelen tillåts minska. Resultaten kan dock inte ses som helt jämförbara. Detta till följd av att denna studie inte nyttjat en konventionell matris för fördelningsönskemål. De starka styrningarna mot långt eller kort timmer skapar en önskad differentiering mellan instruktionerna men är inte anpassade till beståndsegenskaper. Detta i skillnad mot Kivins studie som sökte optimera apteringsinstruktionerna till en begränsad mängd stammar. Denna studie har istället nyttjat en alternativ metod där ett stort datamaterial nyttjats för att finna tendenser till att vissa bestånd gynnas av aptering av långt respektive kort timmer. Den har också belyst andra viktiga variabler så som stamfelsved.

En intressant del i studiens resultat är att beräkningsytor med klana stammar producerade de längsta stockarna vid aptering mot kort timmer och de kortaste stockarna vid aptering mot långt timmer (Tabell 20). Detta i motsats till att beräkningsytor med genomsnittligt grövre stammar producerade de kortaste stockarna vid aptering mot kort timmer och de längsta stockarna vid aptering mot långt timmer. Varför detta är resultatet är svårt att svara på. Men det handlar eventuellt om att det i klana bestånd inte produceras lika många timmerstockar per stam. Detta kan exemplifieras med en vidare inblick i apteringsutfallet:

- För apteringsgrupp ”Gran 0” med medelstamsvolym 0,16 m³fub producerades 0,53 st. timmerstockar per stam.
- För apteringsgrupp ”Gran 1” med medelstamsvolym 0,42 m³fub producerades 1,5 st. timmerstockar per stam.

Detta indikerar att det i de flesta fall apteras endast en timmerstock i klana grupper medan det i grövre grupper oftare produceras fler timmerstockar. På en grov stam blir det således eventuellt möjligt att skapa två kortare stockar istället för en. Det skulle vara en möjlig förklaring till att grövre grupper ger kortast timmer vid användning av kort instruktion. Generellt kan sägas att små träd ger mindre

valfrihet vid aptering och i kombination med hård längdstyrning anpassas längdutfallet för att inte överskrida maximal värdeavvikelse.

4.2 Metoddiskussion

Indelningen av beräkningsytorna i fyra grupper baserat på dominerande trädslag kan ses som en initial manuell klustring. Valet att göra detta motiveras av att det gav möjligheten att avgöra hur bestånds- och stamegenskaper skiljer sig beroende av det dominerande trädslaget. Främst visade sig denna indelning vara fördelaktig då den kortade ner tiden för iteration av alternativ kod vid klustringsanalysen och gav bättre prestanda vid hantering av materialet i exempelvis Power BI. Denna indelning antas inte ha haft några negativa effekter på studiens resultat eller applicerbarhet. Detta då just trädslagsandelar i det använda materialet kan ses som en av de mer säkra variablerna.

Vid tvätt av datamaterialet exkluderades en mängd värden som ansågs extrema eller ej trovärdiga. Acceptabla värden bestämdes utifrån kännedom om materialet efter visualisering av variablerna och avgränsningar sattes vid värden som indikerade tydliga extremvärden. Detta innebär att korrekta, men väldigt ovanliga stamegenskaper inte representeras i materialet. Samtidigt innebär avgränsningarna som fastslagits vid okulär bedömning att eventuella extremvärden kan ha kvarlämnats till följd av missbedömningar av vad som är ett rimligt maximalt eller minimalt värde för en given variabel. Resultatet har dock inte påvisat en uppenbar påverkan av extremvärden och resulterande klustring har genererat grupper som representerar alla typer av stammar som är normala att finna i praktiken.

Användandet av olika typer av klustringsmetoder kräver en god kännedom om materialet och i ett optimalt fall att hypotetiska kluster kan identifieras på förhand. På det viset kan användaren tolka resultatet enklare och bekräfta eller förkasta resultat vid jämförelse mot ett förväntat utfall. I denna studie har klustringen utförts på variabler i sju dimensioner. Detta ihop med materialets omfattning innebär att förväntade resultat inte kan identifieras på förhand. Således är det inte möjligt att förkasta eller anta att vald klustringsalgoritm som den bästa metoden. Validering skulle kräva jämförelse mot andra metoder vilket inte har kunnat göras inom tidsramen för denna studie. K-means är känd för att fungera bäst i sfäriska kluster och bäst resultat kan uppnås när föredragen form kombineras med att klustren är tydligt åtskilda (Google Developers 2022b). I detta fall är det tydligt att klustren inte är helt separerade, istället ligger observationerna för de klusterbestämmande variablerna spridda mellan min- och maxvärdet för given variabel. Interaktionen mellan variabler i det flerdimensionella rummet kan visserligen innebära att separationen är bättre än vad som visas vid visualisering i två dimensioner.

Siluettkoefficienterna indikerar dock att separationen är liten. Detta kan ses som motivation till att förkasta den valda klustringsmetoden. Att ett datamaterial ska vara helt anpassat till en specifik metod är dock osannolikt. Resultatet visar att vald klustringsmetod kunnat prestera godtyckliga kluster med hänsyn till den stora variationen och antalet möjliga variabelkombinationer i nyttjat datamaterial. Ett alternativ hade varit att ompröva klustringen med en alternativ metod som exempelvis hierarkisk klustring. Denna metod skulle kunna ge bättre eller likvärdiga resultat men processen är ofta mer långsam. Fördelen med att använda hierarkisk klustring skulle vara att antalet kluster inte måste bestämmas på förhand utan kan väljas efteråt (Alpaydin 2014).

Det stora antalet observationer och variabler gör även beräkningsarbetet långsamt under vissa analysprocesser. Att utföra en principalkomponentanalys för att minska dimensionaliteten skulle eventuellt ha kunnat göra arbetet något mer effektivt. Detta tros inte påverka det faktiska resultatet men kan eventuellt underlätta vid framtida liknande studier.

Simulering av apteringsutfallet gjordes utifrån tre apteringsinstruktioner med varierande längdstyrning. Genom att tvinga apteringen mot kortare eller längre längder kunde utfallen sedan jämföras mot utfallet från den längdneutrala listan. Simuleringen hade kunnat inkludera mer avancerade uppsättningar av exempelvis sortiment och fördelningsönskemål. Att undvika detta innebar dock en fördel i att den enskilt viktigaste frågan för studien kunde isoleras. Att minska antalet variabler i simuleringen innebar att resultaten kunde tolkas enklare. Med fler variabler hade det varit svårare att avgöra vilken faktor som föranlett givet resultat. Detta var positivt för studien. Samtidigt finns en utvecklingspotential då simulering med mer konventionella apteringsinstruktioner kan ge nya insikter bortom de som denna studie gett med förändrad längdstyrning.

4.3 Datamaterialet

Det data som använts för studien är hämtat från en mängd olika maskiner över en lång period. Maskinernas mätnoggrannhet är inte känd varför det också kan finnas okända avvikelser i de mätvärden som bestämmer stammarnas egenskaper. Mätvärdenas kvalité kan variera över tid och beror bland annat på kalibreringsintervall, förslitning av mätinstrument och förhållanden vid avverkning (Wilhelmsson et al. 2019). Den stora mängd stammar som inkluderats i denna studie tros dock bidra till att över- och underskattade värden jämnar ut varandra och att medelvärden för hela materialet kan antas vara trovärdigt. Det är också känt efter samtal med SCA att det regelbundet görs uppföljningar på mätinstrumenten i skördarnas aggregat. Detta bidrar också till datamaterialets tillförlitlighet. Enligt

Wilhelmsson et al. (2019) ger skördarmätningen en avvikelse i mätningen om några enstaka procent avseende längd och diameter så länge uppföljning och kalibrering görs regelbundet och med hög kvalitet.

4.4 Tillämpning och vidare studier

Denna studie har resulterat i insikter om möjligheten att hantera stora mängder skogliga data från maskinell avverkning. Samtidigt har studien också gett insikt i möjligheten att nyttja maskininlärning och andra tekniska verktyg för att analysera denna typ av data. Studien har visat att bearbetning av skördardata och uppdelning i beräkningsytor för vidare gruppering i kluster kan utföras. I vidare studier som utgår från stora mängder skördardata kan denna metod nyttjas för att få vidare insikter i datamaterialet. Samtidigt innebär klustringen att materialet kondenseras i en mindre mängd observationer vilket kan vara en fördel då en praktiskt hanterbar mängd klasser eller grupper efterfrågas.

Vidare studier avseende beståndsanpassad aptering är nödvändiga för att kunna dra en slutsats om huruvida detta är en metod att satsa på för att förändra dagens apteringsmetoder. Detta kan göras genom ytterligare apteringsanalyser utifrån denna studies klustringsresultat eller med grund i ny datainsamling. För att fördjupa denna studie skulle det vara relevant att satsa ytterligare på att bearbeta apteringsinstruktionerna vidare och på ett mer konventionellt sätt för att efterlikna applicering i den faktiska verksamheten. Ett alternativt sätt att angripa klustringen skulle vara att identifiera vilka stammar som lämpar sig bäst för olika produkter. Med den informationen styrs sedan klustringen mot att gruppera enskilda stammar eller beräkningsytor. Från den resulterande grupperingen görs sedan vidare apteringsanalyser med anpassade apteringsinstruktioner.

En annan aspekt som kan vara viktig för efterföljande studier är prismatrisens påverkan på apteringsutfallet. I denna studie nyttjades en generell matris som skulle motsvara ett konventionellt exempel. Denna ger i sig en styrning mot att producera det som totalt resulterar i det högsta värdet. Det är här viktigt att ej förväxla detta värde med exempelvis vad som sedan betalas till den som säljer virket. Vilka längd- och diameterkombinationer som ger högst värdeutfall presenteras i tabell 6. Det är tydligt att grövre och längre stockar ger det högsta värdeutfallet. Detta begränsar troligtvis möjliga utfall vid fördelningsaptering till följd av att värdeavvikelsen förhindrar vissa avsteg som skulle kunna förändra utfallet. Ett alternativt tillvägagångssätt skulle vara att tillåta högre värdeavvikelse vid fördelningsaptering eller att göra alla möjliga längd och diameterkombinationer lika mycket värda. Detta kan eventuellt ge andra insikter i interaktionen av att variera längdstyrningen i apteringsgrupperna.

Resultatet av utförd apteringsanalys indikerar att fördelningsaptering med styrning mot längre timmer kan vara fördelaktigt för att öka timmerandelen vid avverkning. Denna studie har dock använt en schablonmässig, orealistisk önskefördelning för att skapa en spridning mellan apteringarna. Detta kan förklara att samma indikation inte ses avseende fördelningsgraden. Anledningen till avvikelsen är att apteringsdatorn inte kan hitta alternativ för att uppnå önskefördelningen utan att bryta kravet på tillåten värdeavvikelse. Vidare studier kopplat till denna fråga skulle vara nödvändiga för att ge vidare insikt i effekten av att använda fördelningsaptering och att producera längre timmer. Genom nyttjande av mer realistiska matriser skulle slutsats kunna dras avseende den indikerade fördelen. Kan den ökade timmerandelen inte kombineras med en hög fördelningsgrad existerar heller inte incitamentet för att nyttja fördelningsapteringen: att öka värdekedjans totala värde.

4.5 Slutsatser

De slutsatser som kan dras från studiens resultat är att:

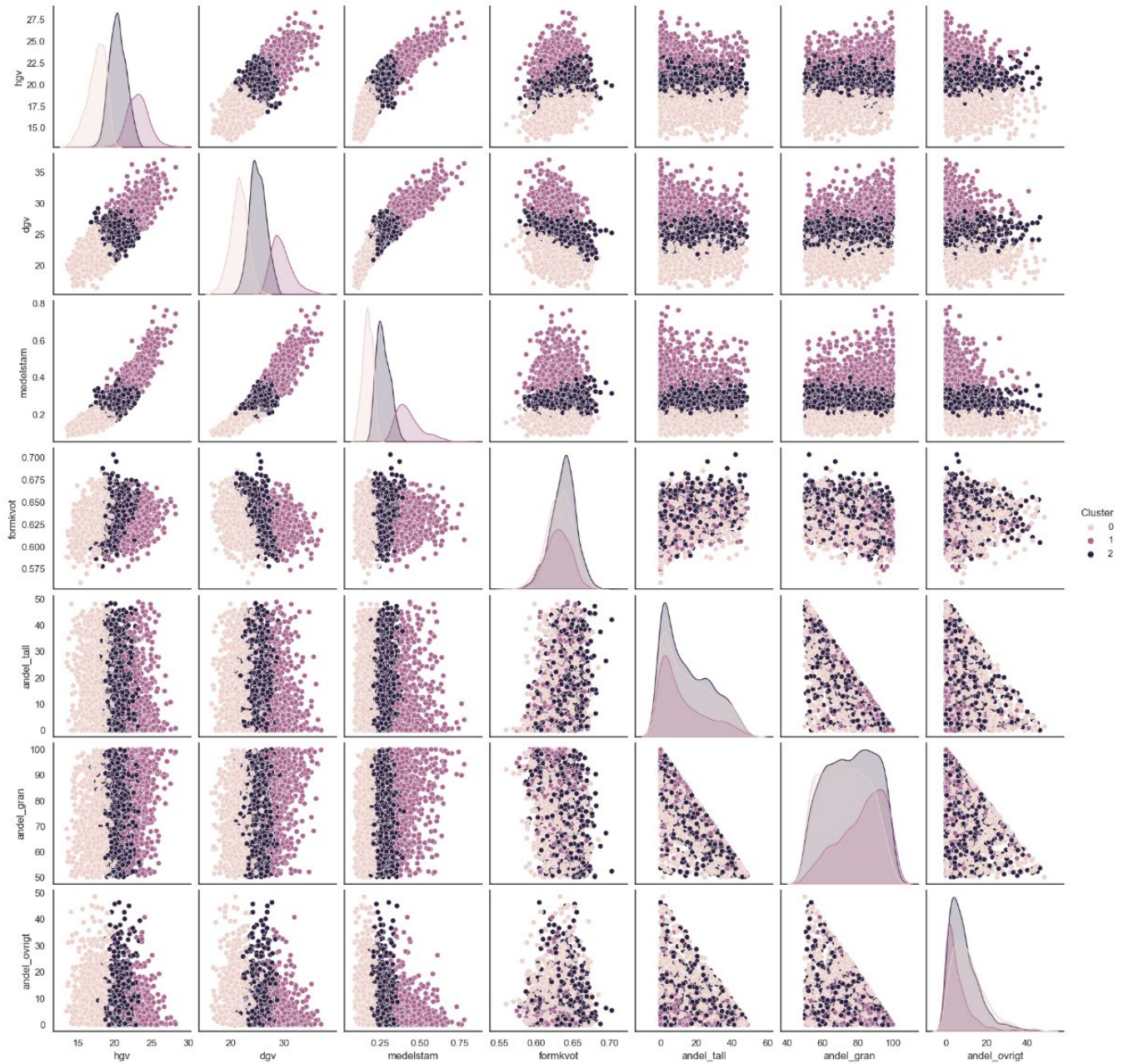
- Det är möjligt att med hjälp av maskininlärning skapa grupper av beståndstyper utifrån skördardata. Variationen av beståndsegenskaper var dock stor. Detta ledde till tvivelaktig klustring för enskilda observationer men ett generellt godtagbart resultat i form av kluster med separerade medelvärden. Analys av siluettkoefficientens värde för olika kluster visade att klustren inte var helt separerade.
- Stamegenskaper inom de skapade klustren varierade relativt mycket. Klustringen producerade alltså grupper med lika beståndsmedelvärden men när egenskaperna studerades på stamnivå tydliggjordes en större variation.
- Fördelningsapatering mot längre timmer enligt studiens metodik gav i det studerade området en högre timmerandel och en lägre andel stamfelsved. Detta skedde dock på bekostnad av en lägre fördelningsgrad enligt given önskefördelning. Timmerandelen ökade i relation till trädens genomsnittliga storlek.
- Det finns indikationer för att olika typer av skog presterade olika bra sett till fördelningsgrad. Längre och grövre skog presterade bäst med kortare timmerlängder eller neutral längdfördelning.
- Medelstamsvolymen tycks ha stor påverkan på skillnader i apteringsutfall. Högre medelstamsvolym innebar en mindre skillnad mellan apteringarna avseende timmerandel och fördelningsgrad.
- Styrning mot långa längder resulterade i en mindre skillnad i timmerandel mellan apteringsgrupperna. För gran var denna skillnad ca 0,15 % och för tall ca 1 %.

Referenser

- Alpaydin, E. (2014). *Introduction to machine learning*. Third edition. Cambridge, Massachusetts: MIT Press. (Adaptive computation and machine learning series)
- Arlinger, J. & Möller, J.J. (2021). *StanForD 2010*. [Broschyr]. Uppsala: Skogforsk. https://www.Skogforsk.se/cd_20190114161847/contentassets/f6ab99f3707a4670baeb987798c6e0b6/stanford-2010-svensk.pdf [2022-09-14]
- Biometria (2022). Kvalitetsbestämning av sågtimmer av tall och gran. <https://www.biometria.se/publikationer/maetningsinstruktioner/maetningsbestaemmelser/nationella-maetningsbestaemmelser/>
- von Essen, I. & Möller, J.J. (1997). *Fördelningsapatering på mindre trakter*. (371). Uppsala: Skogforsk. <https://www.Skogforsk.se/contentassets/0e57b79dfbb944ddb08dc6dfac394af7/arbetsrapport-371-1997.pdf> [2022-09-16]
- Gellerstedt, S. & Dahlin, B. (1999). Cut-To-Length: The Next Decade. *Journal of Forest Engineering*, 10 (2), 17–24. <https://doi.org/10.1080/08435243.1999.10702731>
- Google Developers (2022a). *Interpret Results and Adjust Clustering | Machine Learning*. Google Developers. <https://developers.google.com/machine-learning/clustering/interpret> [2023-01-09]
- Google Developers (2022b). *k-Means Advantages and Disadvantages | Machine Learning*. Google Developers. <https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages> [2023-02-21]
- Han, J., Pei, J. & Kamber, M. (2011). *Data Mining: Concepts and Techniques: Concepts and Techniques*. Saint Louis, USA: Elsevier Science & Technology. <http://ebookcentral.proquest.com/lib/slub-ebooks/detail.action?docID=729031> [2022-11-16]
- Mohamad, I.B. & Usman, D. (2013). Standardization and Its Effects on K-Means Clustering Algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, 6 (17), 3299–3303. <https://doi.org/10.19026/rjaset.6.3638>
- Möller, J.J., Bhuiyan, N. & Hannrup, B. (2015a). *Utveckling och test av beslutsstod vid automatiserad gallringsuppföljning* (862). Uppsala: Skogforsk https://www.Skogforsk.se/cd_20190114161533/contentassets/1b9d70029980454596832b78317b9c51/utveckling-och-test-av-beslutsstod-vid-automatiserad-gallringsuppfoljning-arbetsrapport-862-2015.pdf [2023-03-09]
- Möller, J.J. & Moberg, L. (2007). *Stambank VMF Qbera*. (641). Uppsala: Skogforsk. https://www.Skogforsk.se/cd_20190114161857/contentassets/d11af9e2a791457abaf1fd65e035d6ed/641-stambank-vmf-qbera-hela.pdf [2023-01-16]
- Möller, J.J., Nordström, M. & Arlinger, J. (2015b). *Förbättrade utbytesprognoser*. (880). Uppsala: Skogforsk. [2022-09-05]
- Neo, B. (2020). *What is Tidy Data?* <https://towardsdatascience.com/what-is-tidy-data-d58bb9ad2458> [2022-11-10]

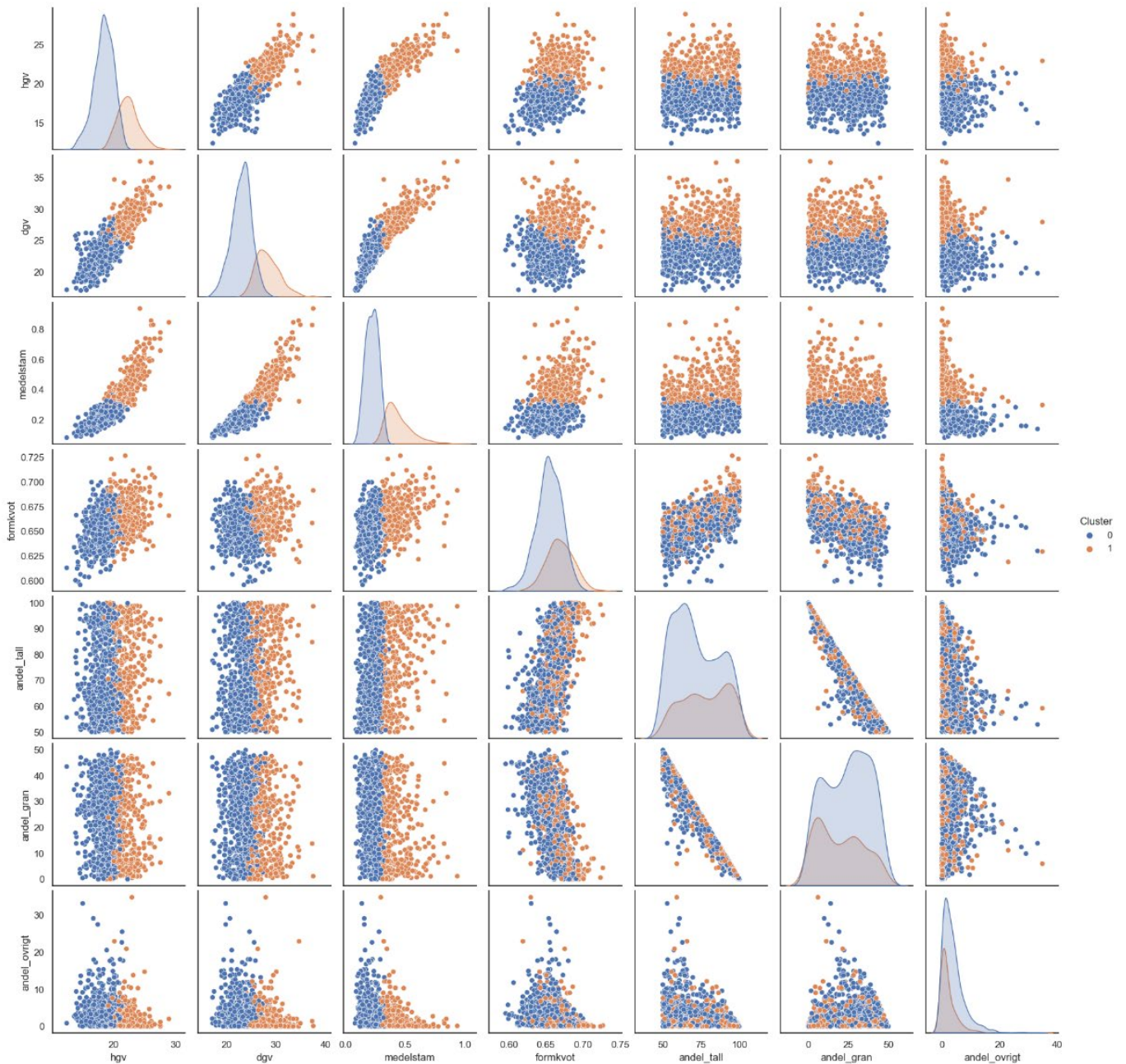
- Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- scikit-learn (2022). *sklearn.preprocessing.StandardScaler*. *scikit-learn*. <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.StandardScaler.html> [2023-02-21]
- Skogsstyrelsen (2022). *Avverkningen på rekordnivå 2021*. <https://www.skogsstyrelsen.se/nyhetslista/avverkningen-pa-rekordniva-2021/> [2022-09-14]
- Wilhelmsson, L., Möller, J.J. & Arlinger, J. (2019). *Betalningsgrundande eller betalningsstödjande virkesmätning med skördare*. (1032). Uppsala: Skogforsk https://www.Skogforsk.se/cd_20200203142819/contentassets/3fece182fb8f458dbdf3e59a81d8625a/arbetsrapport-1032-2019.pdf [2022-09-14]

Bilaga 1



Figur 15. Parvis jämförelse av vid klustring ingående variabler för grandominerade beräkningsytor.
Figure 15. Pair plot for comparing variables used for clustering spruce dominated areas.

Bilaga 2



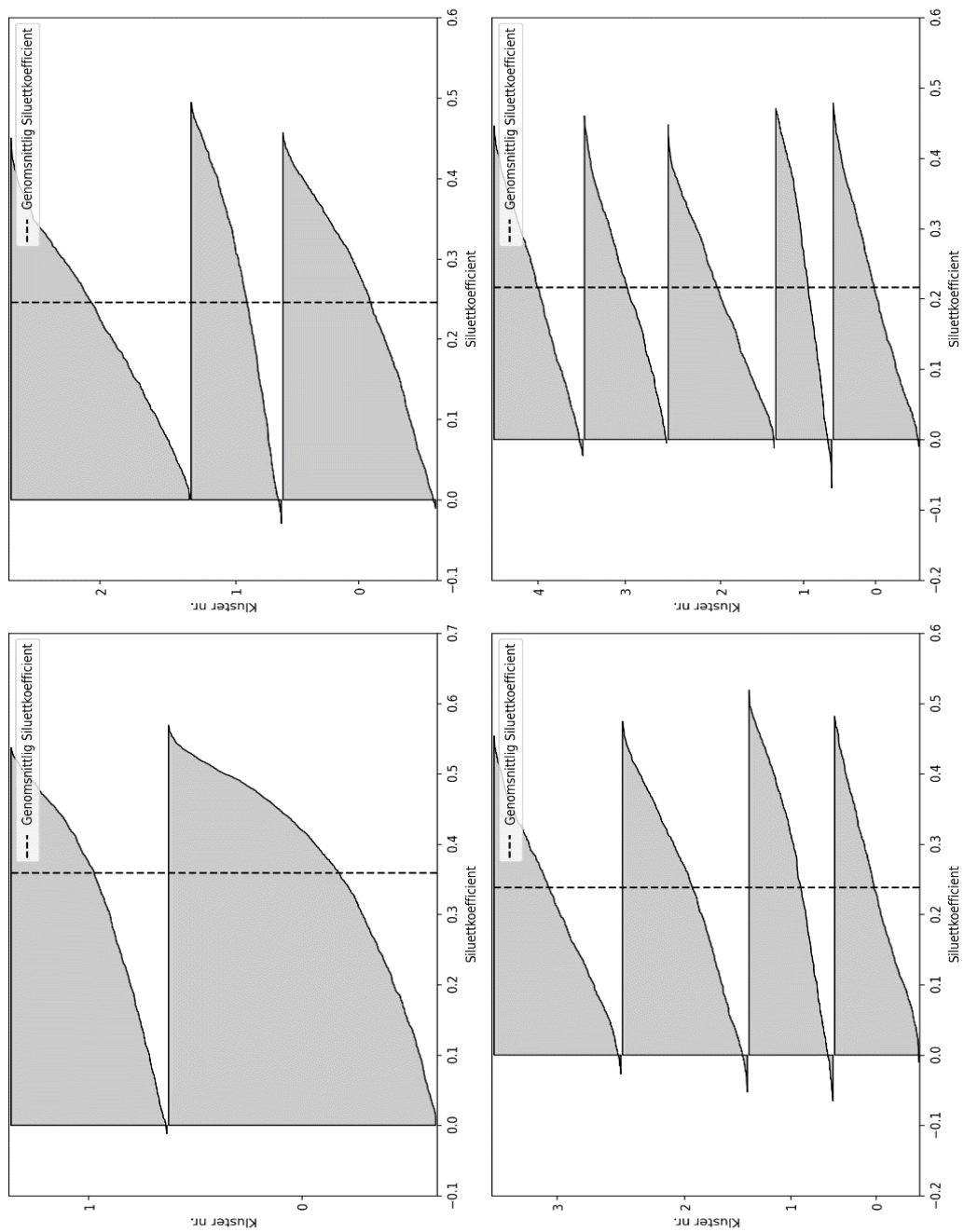
Figur 16. Parvis jämförelse av vid klustring ingående variabler för talldominerade beräkningsytor.
Figure 16. Pair plot for comparing variables used for clustering pine dominated areas.

Bilaga 3

Tabell 22. Min- och max-värden för variabler i varje datagrupp för utsortering av oönskade värden
Table 22. Min- and max-values for variables in different data groups used to exclude unwanted values

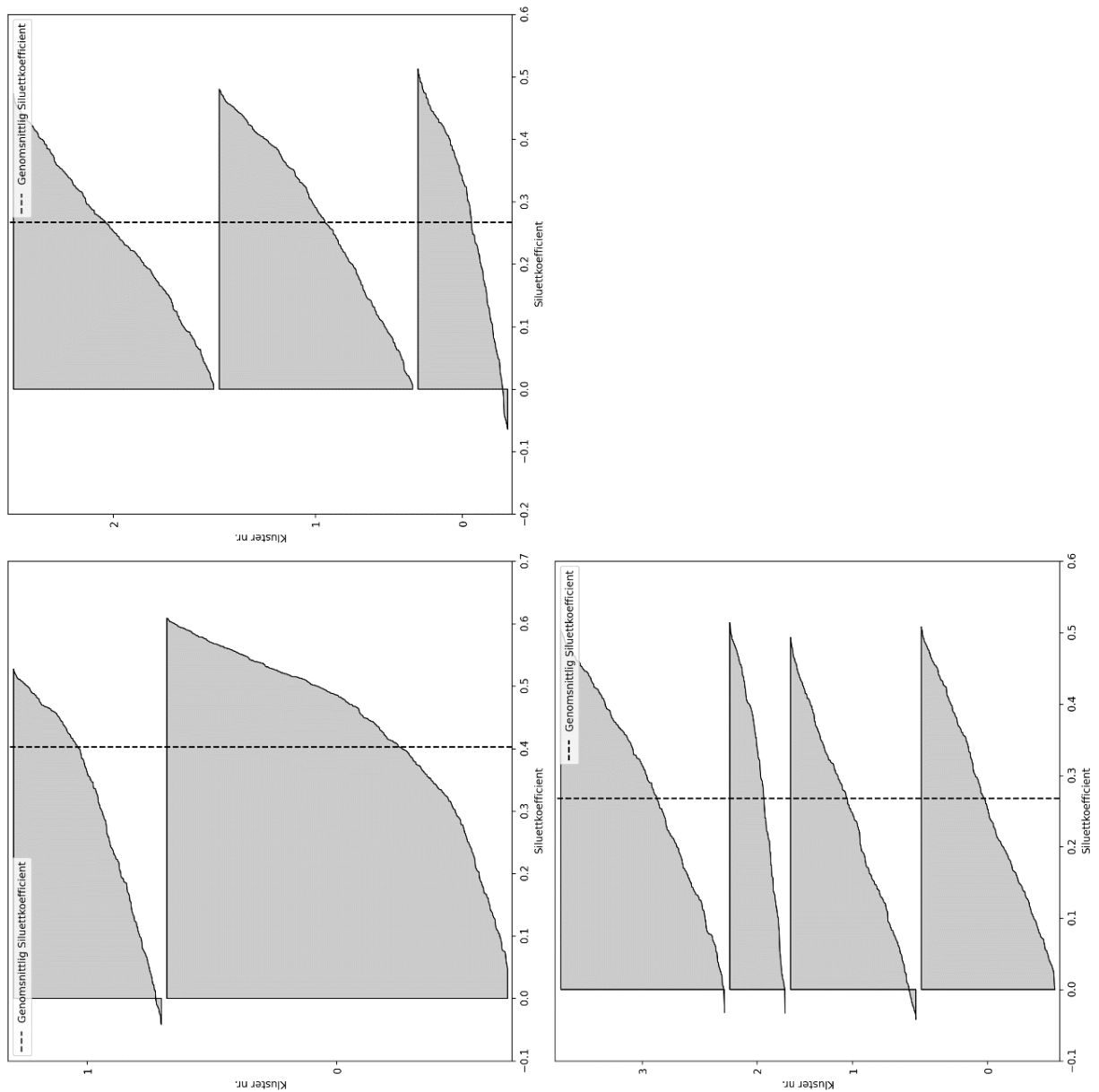
Variabel	Min	Max	Datagrupp
Andel per trädslag (%volym)	0	100	Objektsdata
Area (ha)	0,5	-	Objektsdata
DGV (cm)	0	60	Objektsdata
Grundyta (m ² /ha)	0	60	Objektsdata
HGV (m)	0	40	Objektsdata
Höjd över havet (m)	-	700	Objektsdata
Medelstamsvolym (m ³ fub)	0	1	Objektsdata
Stamantal (n)	0	3000	Objektsdata
Volym per hektar (m ³ fub/ha)	1	3000	Objektsdata
Ålder (år)	45	-	Objektsdata
Volym (m ³ fub)	0,01	1	Stockdata
Toppdiameter (mm)	0,01	600	Stockdata
Mittdiameter (mm)	0,01	500	Stockdata
Rotdiameter (mm)	50	700	Stockdata
Längd (cm)	-	700	Stockdata
Formkvot	0,3	1	Stamdata
Höjd (cm)	-	4000	Stamdata
DBH (mm)	-	800	Stamdata
Höjd över havet (m)	-	700	Stamdata

Bilaga 4



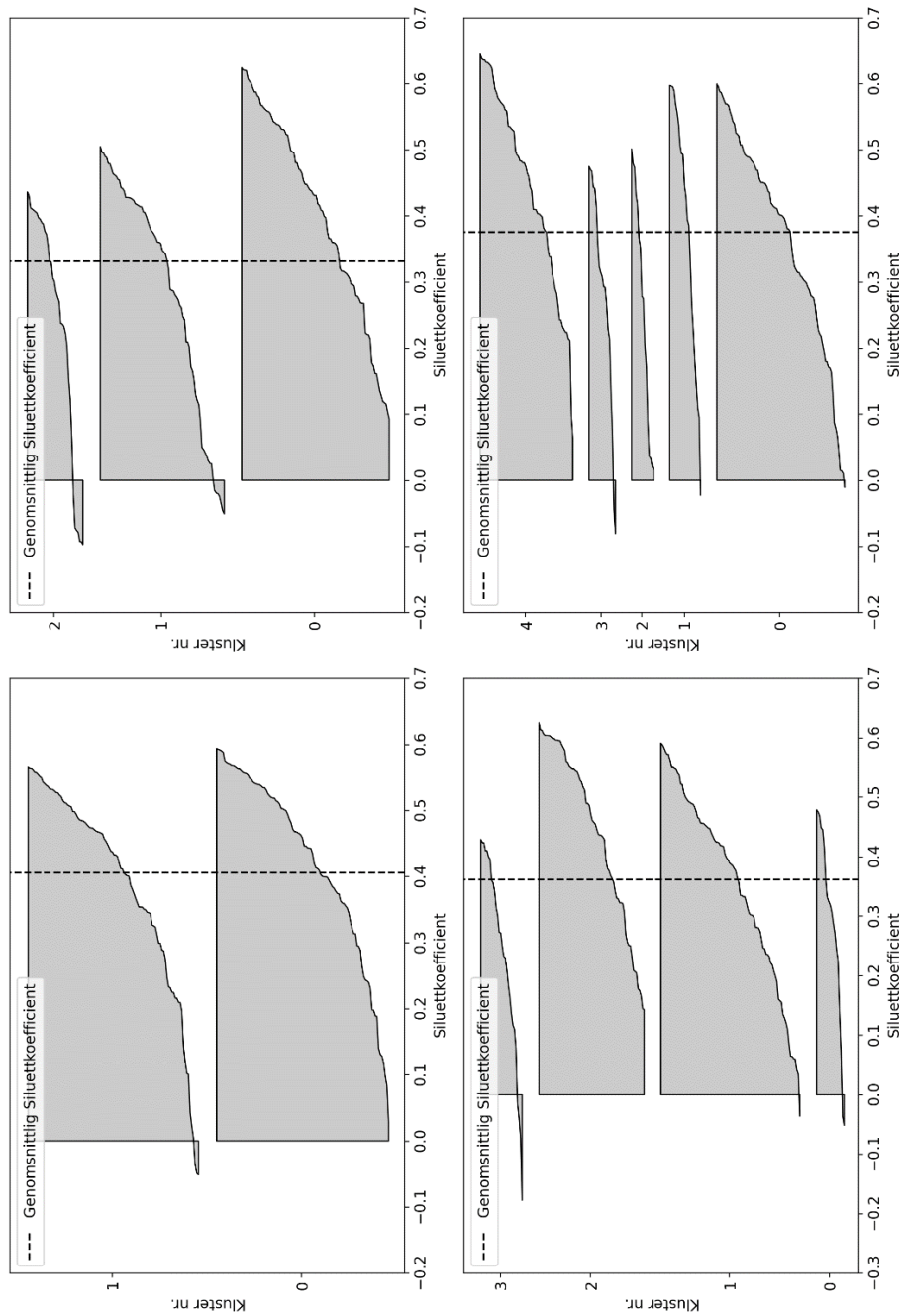
Figur 17. Siluettgrafer för k=2 till k=5 kluster för grandominerade beräkningsytor.
Figure 17. Silhouette graphs for k=2 to k=5 clusters for spruce dominated areas.

Bilaga 5



Figur 18. Siluettgrafer för k=2 till k=4 kluster för talldominerade beräkningsytor.
Figure 18. Silhouette graphs for k=2 to k=4 clusters for pine dominated areas.

Bilaga 6



Figur 19. Siluettgrafer för k=2 till k=5 kluster för beräkningsytor dominerade av övriga trädslag.
Figure 19. Silhouette graphs for k=2 to k=5 clusters for areas dominated by other tree species.