

Sluttrapport fra prosjektet:

## Videre utvikling av Geoskog som prognoseverktøy 2016-2017

### Om prosjektet

Viken Skog har vært prosjektansvarlig på vegne av partnere i Geoskog samarbeidet som består av Statskog, SB-Skog, Foran, Jordskifteretten og Viken med Geodata som forvalter av kildekode og programvare utvikler.

Basert på oppnådde resultater beskrevet i selve rapporten ansees prosjektet som særs vellykket og nye versjoner av programvaren er allerede tatt i bruk.

**Prosjektet ble støttet av Skogtiltaksfondet, kontrakt B-2016-08 «*Videre utvikling av skog Norges prognoseverktøy.*»**

Denne rapporten er distribuert til alle partnere i Geoskog samarbeidet, Skogtiltaksfondet og vil bli publisert på Viken Skogs hjemmeside.

Askim 18. august 2017, prosjektansvarlig Viken Skog Svein Dypsund

## Innholdsfortegnelse

Videreutvikling av Geoskog som prognoseverktøy .....	3
Innledning .....	3
Materiell og Metoder .....	3
Geoskog versjoner .....	3
Datasett med ulike antall bestand .....	3
Utvikler & fagligansvarlig .....	4
Testing .....	4
Beskrivelse av gjennomførte aktiviteter .....	4
Ytelsesforbedring ved kjøring av prognose basert på balansekvantum .....	4
Forbedre datavask-programmet .....	5
Bedre fleksibilitet .....	5
Bedre brukervennlighet .....	5
Feilretting .....	5
Resultat .....	5
Ytelsesforbedring ved kjøring av prognose basert på balansekvantum .....	5
Sammenligning mellom Geoskog for fase 2 og referanse .....	8
Forbedre datavask-programmet .....	9
Bedre fleksibilitet .....	10
Bedre brukervennlighet .....	10
Feilretting & nyere funksjonalitet .....	10
Konklusjoner .....	11
Vedlegg sluttregnskap .....	11

## Videreutvikling av Geoskog som prognoseverktøy

### Innledning

Geoskog er en løsning som ble påbegynt utviklet av Geodata AS i 2012 på oppdrag fra Statskog SF. Bakgrunnen var Statskogs kjøp av Borregård skoger hvor man så behovet for å gjøre funksjonaliteten i Avvirk2000 og Verdsett mer tilgjengelig. Geoskog er utviklet i scriptspråket Python noe som gjør kildekoden tilgjengelig for de som har tilgang til løsningen. Dette sikrer at det både i dag og i fremtiden vil være enkelt å videreutvikle Geoskog. Avvirk 2000 var skrevet i Fortran, noe som gjør videreutvikling og innsikt i koden vanskelig. Statskog besluttet derfor å utvikle en ny løsning (Geoskog), basert på de samme tilvekstfunksjonene som ble benyttet i Avvirk 2000. Løsningen er tilpasset Statskogs behov, og gir en prognose for stående volum for hvert 5. år de neste 100 år frem i tid. Geoskog støtter 3 forskjellige hogstmetoder (hogstmodenhet, balansekvantum og angitt årlig volum).

### Materiell og Metoder

#### Geoskog versjoner

Ytelsesforbedringen hadde som utgangspunktet den tidligere versjonen av Geoskog (2015.1) der det ble observert lang prosesseringstid ved kjøring av prognose basert på balansekvantum. Versjon 2015.1 er den versjonen som var i bruk ute hos brukerne før vi startet å teste. Versjon 2016.3 hører til fase 1 i søknaden. Vi lagde også en nyere versjon som heter 2017.1 som hører til fase 2 i søknaden.

#### Datsett med ulike antall bestand

Vi har testet med ulike størrelser for predefinert antall bestand som utgjorde en skogeiendom med følgende antall (se tabell 1); en må påpeke at siden den tidlige versjonen av Geoskog (2015.1) balansekvantum tok veldig lang tid, kunne vi ikke gjenskape tidsforbruket her for alle størrelser, men derimot har vi kunnet bruke disse bestand i de nyere versjoner.

Datsett	Geoskog versjon			Datsett	Geoskog versjon		
	2015	2016.3	2017.1		2015	2016.3	2017.1
1000	X	X	X	20000	-	X	X
2000	X	X	X	30000	-	X	X
3000	X	X	X	40000	-	X	X
4000	X	X	X	50000	-	X	X
5000	X	X	X	60000	-	X	X
6000	X	X	X	70000	-	X	X
7000	-	X	X	80000	-	X	X
8000	-	X	X	90000	-	X	X
9000	-	X	X	100000	-	X	X
10000	-	X	X				

Tabell 1. Antall bestand som ble brukt i testingen. X= har blitt kjørt

Videre utvikling av Geoskog som prognose verktøy

### Utvikler & fagligansvarlig

Rui Rua var hovedutviklere i prosjektet som gikk igjennom videreutviklingen av Geoskog. Felipe Verdú var fagligansvarlig person for Geoskog.

### Testing

Testingen ble gjennomført først fra Python skriptet direkte fra kildekoden. Det ble satt i koden en så kalt Python Profiler som gav tidsforbruket for hver kjøring så vi kunne registrere tiden på et systematisk sett for de ulike antall bestand i analysen. Vi kjørte kun prognose basert på balansekvantum.

Vi angrep de delene («issue») i kildekoden som gav det høyeste tidsforbruket i disse preliminare analysene. Settet å jobbe var dermed veldig avhengig av kunne forbedre en issue og ikke påvirke resultatet. For å kunne verifisere at resultatet ikke ble endret etter hvert inngrep, det nye resultatet ble testet systematisk med en fasit fra versjonen 2015.1 (referanse).

Når man nådde en optimal forbedring for denne issue der ingen forbedring kunne oppstå med tanke på tidsramme og prioriteringer testede vi også via Geoskog verktøykassen for ArcMap med samme datasett som beskrevet i tabell 1. Hvis dette testet i ArcMap gav liknende resultat som for skriptet kunne vi avslutte saken og vi gikk videre til neste issue.

### Beskrivelse av gjennomførte aktiviteter

#### Ytelsesforbedring ved kjøring av prognose basert på balansekvantum

Geoskog har hatt lang prosesseringstid ved kjøring av prognose basert på balansekvantum. Selv ved kjøring på relativt begrensede områder, er tidsbruken uforholdsmessig høy, og denne øker eksponentielt med antall bestand. Ved kjøring på en gjennomsnittlig stor skogeiendom, utgjør ikke prosesseringstiden noe stort problem, men i en del tilfeller av enkeltkjøring på svært store områder, har dagens løsning vist seg å ikke være brukbar i praksis.

Ytelsesforbedring i Geoskog er gjennomført i to faser.

Fase 1: Ved hjelp av tidtaking på enkeltoperasjoner under kjøring av programmet, har vi klart å isolere noen av de mest tidkrevende elementene, og optimalisere disse.

Fase 2: For å få ytterligere forbedring av ytelse, er det behov for en mer gjennomgripende gjennomgang av programmets kildekode for å optimalisere prosesseringsalgoritmer og minnebruk. Ut fra de analyser som er gjort, vil det være mulig å forbedre ytelsen betraktelig

## Videre utvikling av Geoskog som prognose verktøy

### Forbedre datavask-programmet

- a) Grenseverdier, som i dag er hardkodet, skal gjøres konfigurerbare.
- b) Funksjonalitet for beregning av flere attributter en i dag, blant annet tilvekst.

Datavask-programmet er en modul i Geoskog som kontroller verdier og sammenhengen mellom verdiene samt beregner manglende verdier (utelukkende basert på publiserte norske forskningsrapporter). Modulen rapporterer også bestand med feil verdier og feil i sammenhengen mellom verdier.

### Bedre fleksibilitet

- a) Rapportmodulen i Geoskog gjøres mer fleksibel ved å innføre mulighet for å splitte rapportene basert på en egenskap valgt av brukeren.
- b) Geoskog stiller i dag strenge krav til grunnlagsdataenes modell. Dette vil kunne gjøre mer fleksibelt ved å gi brukeren mulighet til styre tilordning av databasefelter.

Geoskogs rapportmodul leverer ut prognose resultater for avvirkning etter hogstmodenhet, årlig fast kvantum eller balansekvantum.

### Bedre brukervennlighet

- a) Forbedre feilmeldinger, slik at det er enklere for brukeren å lokalisere og rette feil i grunnlagsdataene.
- b) Gjøre funksjonaliteten tilgjengelig gjennom en egen verktøylinje for Geoskog i ArcMap.

### Feilretting

Det foreligger i dag en liste med kjente feil som skal rettes.

### Resultat

#### Ytelsesforbedring ved kjøring av prognose basert på balansekvantum

##### *Fase 1*

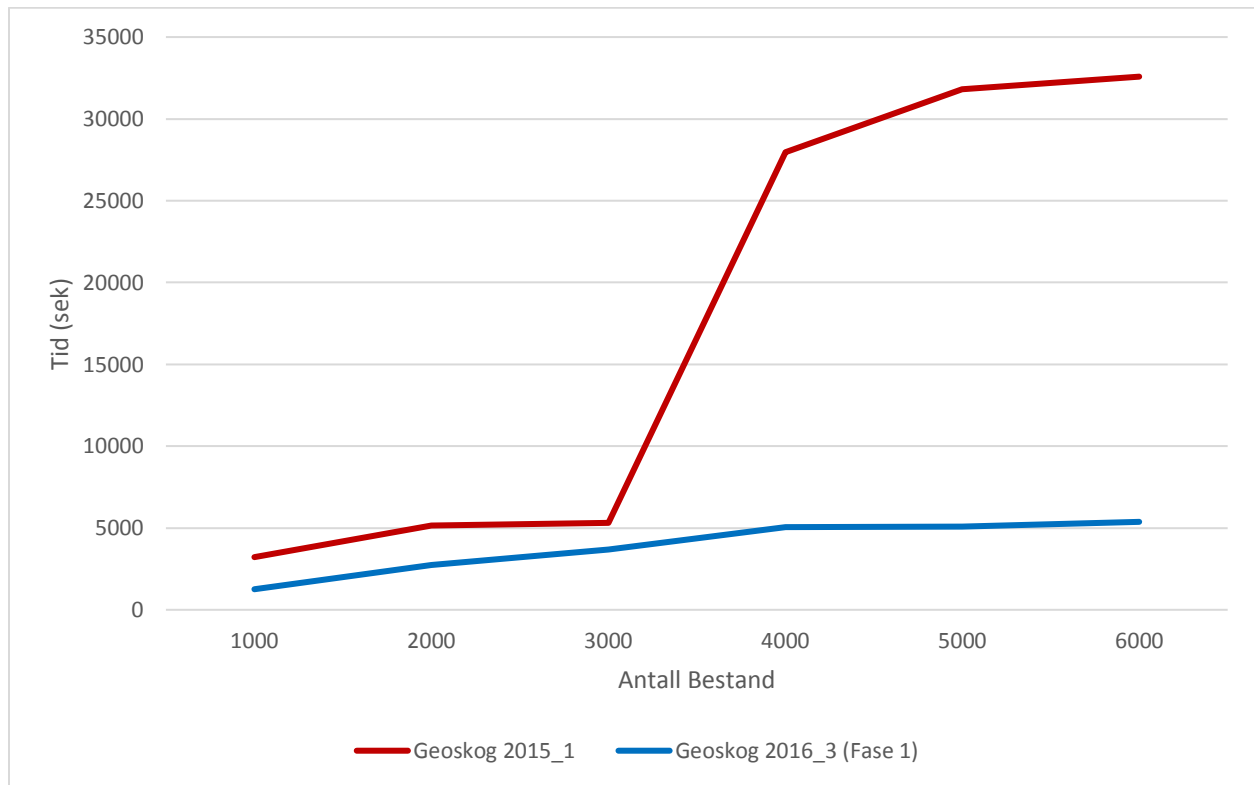
Resultatet av fase1 har kommet ut som versjon 2016\_3 for Geoskog. Det er blitt en ytelsesforbedring (i henhold til versjonen 2015\_1) som utmerker seg allerede fra første datasett (1000) med en forbedring som passerer 250% (256%) og der det virkelig slår til er med et større datasett på 4000 bestand der vi passerer 500% ytelsesforbedring for å komme over 600% for et datasett av 6000 bestand. Et datasett med 3000 bestand tok i vår referanse ca. 28000 sekunder mens i fase1 tok ca. 5045 sekunder.

## Videre utvikling av Geoskog som prognose verktøy

Antall bestand	Forbedring (%)	Balansekvantum kjørt med Geoskog 2015_1 (sek)	Balansekvantum kjørt med Geoskog 2016_3 (Fase 1) (Sek)
1000	256	3218	1257
2000	187.8	5143	2738
3000	144.9	5323	3673
4000	554.1	27956	5045
5000	624.4	31814	5095
6000	606.2	32581	5375

Tabell 2. Forbedring for fase 1

I diagram 1, ser man effekten av forbedringen som ble implementert i fase 1 i form av et mer linjer utvikling av tidsforbruket for fase 1 i motsetning til utviklingen for vår referanse som øker eksponentielt med antall bestand.



Figur 1. Tidsforbruket for balansekvantum for vår referanse og fase 1

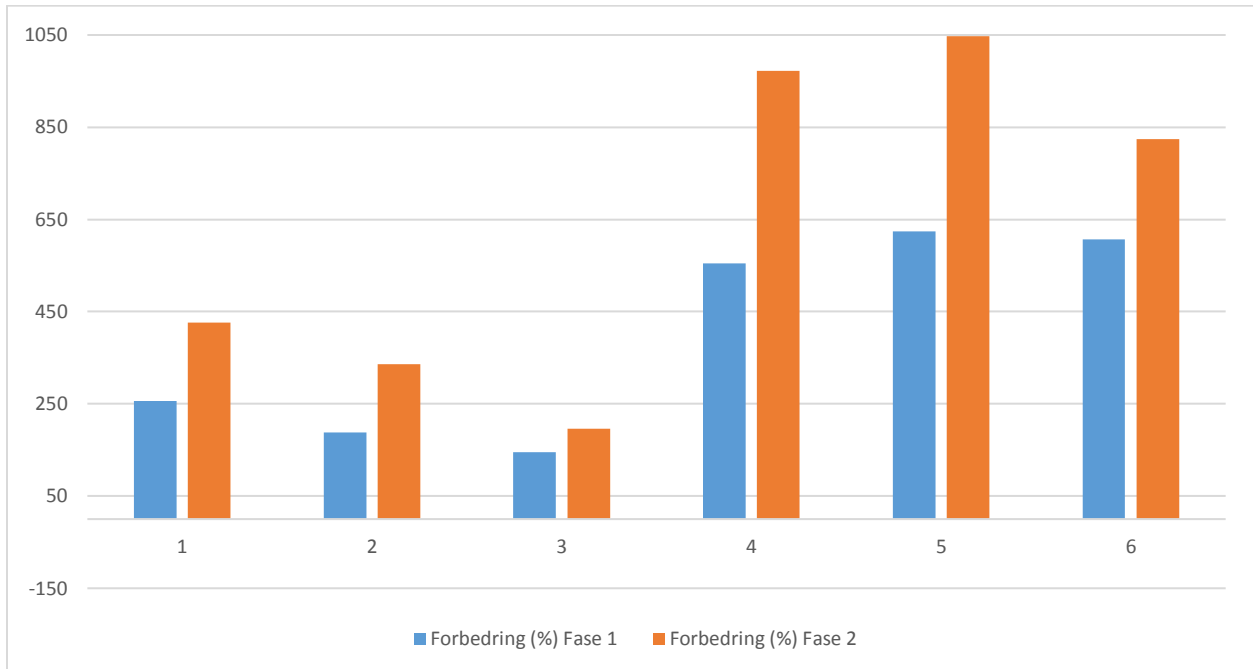
### Fase 2

Resultatet av fase 2 har kommet ut som versjon 2017.1 for Geoskog. Generelt sett, har ytelsesforbedringen doblet seg i henhold til vår referanse (Geoskog 2015.1). En kan påpeke at tidsforbruket for 1000 bestand for oppdaterte versjonen av Geoskog (fase 2) har forbedret med ca. 426% mot Geoskog 2015.1. For 6000 bestand, tidsforbruket har opplevd en økning på ca. 824% som forbedring (Tabell 3 og figur 2).

## Videre utvikling av Geoskog som prognose verktøy

Antall Bestand	Forbedring (%) Fase 1	Forbedring (%) Fase 2	Balansekvantum kjørt med Geoskog 2015.1	Balansekvantum kjørt med Fase 1 (Geoskog 2016.3)	Balansekvantum kjørt med Fase 2 (Geoskog 2017.1)
1000	256	427	3218	1257	754
2000	187.8	336	5143	2738	1530
3000	144.9	196	5323	3673	2711
4000	554.1	973	27956	5045	2874
5000	624.4	1048	31814	5095	3037
6000	606.2	824	32581	5375	3954

Tabell 3. Forbedring for fase 2

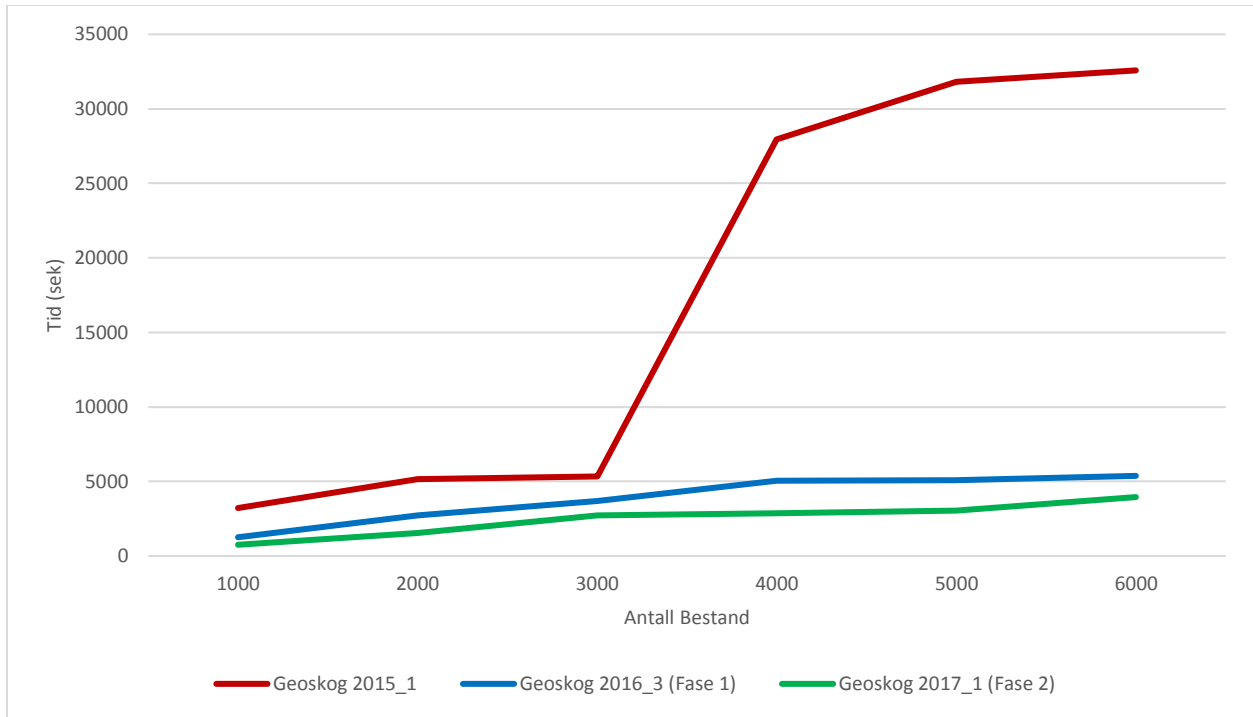


Figur 2. Forbedring for fase 1 og fase 2.

I figur 3, ser vi at tidsforbruket har blitt enda bedre for fase 2 i forhold mot fase 1. I alle fall, ser vi at tidsforbruket har blitt lineær i motsetning mot tidligere.

Generelt sett, ser vi at ytterligere økning av ytelsen for balansekvantum kan nås i den nyere versjonen av ArcGIS med 10.5 *numpy* som brukes av Geoskog, siden det har blitt mye raskere og det betyr i praksis at brukstiden har gått ned ytterligere ca. 25% mot ArcGIS 10.4.1.

## Videre utvikling av Geoskog som prognose verktøy



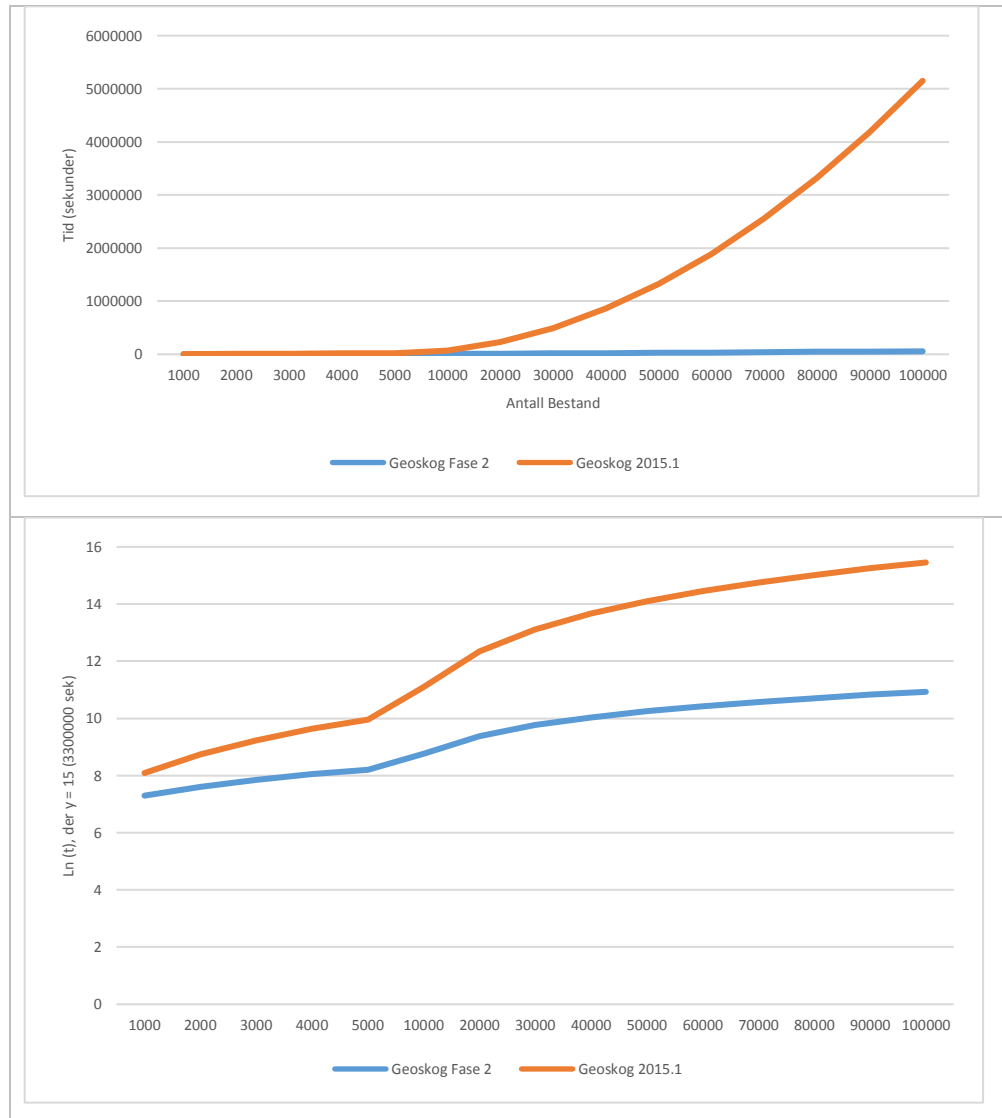
Figur 3. Tidsforbruket for balansekvantum for vår referanse, fase1 og fase 2.

### Sammenligning mellom Geoskog for fase 2 og referanse

Vi har estimert for Geoskog 2015.1 hvor mye tid skulle det ta med ulike datasett for ulike antall bestand opptil 100.000 bestand. Vi sammenlignet det med faktiske resultat fra oppdaterte Geoskog for fase 2. Som eksempel, kan vi se at for 10.000 bestand kjørt med fase 2, skal ta ca. 66000 sek (18 timer og 22 min), mens for vår referanse skulle ta 1 time og 47 min. For et datasett for 100.000 bestand, tidsforbruket for vår referanse er like med **60 dager**, mens for Geoskog for fase 2 tok det ca. 15 timer og 21 min. Dette tilsier at kjøre balansekvantum med vår referanse er i praksis umulig eller uaktuell. Å andre siden, med oppdaterte versjonen av Geoskog (fase 2), gjør det helt mulig å kjøre et stort datasett.



## Videre utvikling av Geoskog som prognose verktøy



Figur 4. Reelt tidsforbruk av Geoskog for fase 2 med opptil et datasett som består av 100.000 bestand. A) Overste figur, antall sekunder; B) nederst, logaritmen av antall sekunder.  $\ln(t) = 15$  er ca. 38,5 dager

### Forbedre datavask-programmet

Datavask-programmet er en modul i Geoskog som kontroller verdier og sammenhengen mellom verdiene samt beregner manglende verdier (utelukkende basert på publiserte norske forskningsrapporter). Modulen rapporterer også bestand med feil verdier og feil i sammenhengen mellom verdier.

For å ha en bedre kontroll på grenseverdier, som har vært hardkodet, har gjorts konfigurerbare. Disse kan styres med hjelp av konfigurasjonsverktøyet. Følgende variabler har blitt konfigurerbare:

- TREANTALLDAA for hogstklasse II: Ovre/Nedre grense for treantall (HKL2)
- TREANTALLDAA for hogstklasse > II: Ovre Nedre grense for treantall (HKL>2)

Videre utvikling av Geoskog som prognose verktøy

- c. Maks alder for hogstklasse V: Ovre aldersgrense for HKL5
- d. Ovre/Nedre grense for middeldiameter
- e. Ovre/Nedre grense for middelhøyde

Dessuten, det har blitt implementert nye funksjonalitet for beregning av flere attributter en i dag, blant annet tilvekst og total volum.

#### Bedre fleksibilitet

Det ble ikke gjort noe med denne siden det ble nedprioritert til et senere tidspunkt.

#### Bedre brukervennlighet

Brukergrensesnittet har gjennomgått mindre justeringer i all hovedsak *som en del av* feilretting og implementering av ny funksjonalitet slik det er beskrevet i neste avsnitt.

#### Feilretting & nyere funksjonalitet

Følgende rettelser er gjorda:

1. AREALPROD flyttes starten av tabellen.
2. Støtte for ArcGIS 10.4.1. og 10.5
3. Legg til massevirkeprosent samt skurprosent i resultattabellen.
4. Endring standardverdier i «default config fil». Det varsles likevel at det er veldig viktig at bruker tar en sikt på innholdet i «config filen» gjennom å bruke Konfigurasjonsverktøyet å lage en egen config fil til eget bruk.
5. Tilvekstkurve i rapporten I form av diagram og I tabell
6. I datavasken når en re kalkulerer volum, oppdateres det nå også VOLUM TOTAL (VOLUMTOT)
7. Når en fremskriver en bestand, oppdateres nå også VOLUM TOTAL (VOLUMTOT)
8. Ytterligere økning av ytelsen for balansekvantum.
  - a. I den nyere versjonen av ArcGIS, 10.5 *numpy* som brukes av Geoskog, har blitt mye raskere og det betyr i praksis att brukstiden har gått ned ytterligere ca. 25% mot ArcGIS 10.4.1.
  - b. Balansekvantum kjøres nå i «background modus», hvilket det hjelper å bruke mer minne enn CPU.
  - c. I 10.4.1 tok det 22t, 14 min og 31 sek
9. Ytelsesforbedring i økonomimodulen

Videre utvikling av Geoskog som prognose verktøy

## Konklusjoner

En kan konkludere at alle inngrep gjennomførte i både fase 1 og fase 2 har blitt en suksess som vist under resultat med en forbedring som i noe tilfelle er oppe i 1000% ytelsesforbedring.

Resultatet fra fase 2 gjør det mulig for alle brukere at bruke balansekvantum i en rimelig tidsperiode hvilket var ikke mulig tidligere. Det vil si, at kjøre enkelte prognoser for små eiendommer uten at det påvirke brukeren i sine daglige arbeidsvaner. Den nye oppdaterte versjonen gjør det også mulig å kjøre balansekvantum for store eiendommer med flere tusen bestand. I vårt eksempel, kunne vi kjøre balansekvantum for Geoskog for 100.000 bestand på overkant av 15 timer.

Det kan påpekes at bruket av numpy i ArcGIS for desktop i sin versjon 10.5 gir ytterligere en økning av ytelsen for balansekvantum kan nås siden det har blitt mye raskere og det betyr i praksis att brukstiden har gått ned ytterligere ca. 25% mot ArcGIS 10.4.1.

Metodene som ble brukte under denne videreutvikling har gitt vad vi forventet og det anbefales at utføre ytterligere analyser som kan tenkes for å forbedre enda mer balansekvantum modulen samt andre deler av Geoskog.

## Vedlegg sluttregnskap

Sluttregnskapet er satt opp med medgåtte timer fra utvikler miljøet og finansieringsplanen beskrevet i søknaden er fulgt. Selve sluttregnskapet er satt opp i vedlegget «*Geoskog utvikling finansiering og sluttregnskap - vedlegg til sluttrapporten*».