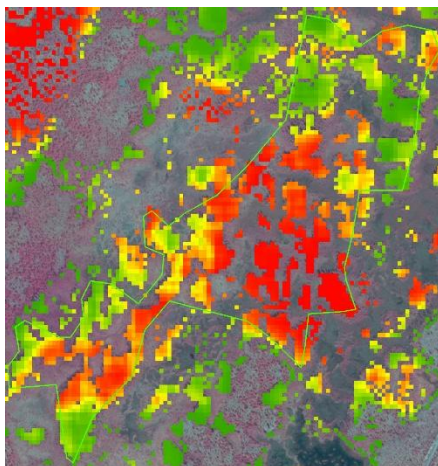




Dataproduktspecifikation

Indikationer på skog med naturvärden





Version	Datum	Beskrivning
0.1	2024-10-18	Första utkast
0.2	2025-08-26	Redigering
0.2_mall	2025-09-04	Redigering
0.3	2025-12-15	Redigering
0.4	2026-01-13	Redigering
0.5	2026-03-09	Förslag första version

Vad är Digitala värden i skog (DVIS)

Digitala kunskapsunderlag i form av kartor visar beräknade värden utifrån olika typer av fjärranalysdata och referensdata där ingen bedömning har gjorts på plats i fält. Kartorna ger en objektiv och skalbar grund för vidare bedömningar inom exempelvis skogsbruk, naturvård och samhällsplanering.

Digitalt framställda kartor är kraftfulla verktyg för analys, men bör inte användas direkt för rapportering, statistik eller beslut utan vidare bedömning. Detta beroende på att de innehåller osäkerheter kopplade till datainsamlingsmetoder, modellantaganden och upplösning, vilket kan påverka noggrannheten på lokal nivå. Kartorna kan innehålla fel eller generaliseringar som inte alltid fångar komplexiteten i verkliga förhållanden. För att säkerställa tillförlitlighet bör de kompletteras med expertbedömningar eller fältmätningar för att bekräfta resultaten och anpassa dem till specifika användningsområden. **Kartorna utgör inte beslut, ställningstagande eller bedömning av Skogsstyrelsen**, och är inte lämpliga att använda som enda underlag för till exempel myndighetsbeslut, ersättningar eller civilrättsliga tvister.

Inledning dataproduktspecifikation

Detta dokument och övriga dataproduktspecifikationer som beskriver dataprodukter inom DVIS ska ses som en definition av en dataprodukt. Det innebär en beskrivning, bland annat:

- .. egenskaper som finns
- .. hur den tagits fram samt
- .. de krav som ställs på dataprodukten.

Det här dokumentet riktar sig till användare av data primärt och inte för aktörer som ämnar att skapa/ta fram dataprodukten.

Denna dataproduktspecifikation är uppbyggd enligt ISO-standard vilket ställer krav på både innehåll och dokumentets struktur. Nedan följer en kort beskrivning av kapitelinnehåll, vilken också kan ses som en guide beroende på vilken information du som läsare eftersöker om dataprodukten:

- *Dataproduktens beskrivning, syfte och tillkomsthistorik*: Kapitel 1.2 och 7
- *Dataproduktens egenskaper och struktur*: Kapitel 4
- *Information rörande datakvalité*: Kapitel 6
- *Information om tillgängliggörande/nyttjande av dataprodukt*: Kapitel 8 och 10

Bakgrund

Detta dokument har tagits fram som ett stöd för datadelning och vidare användning av den beskrivna dataprodukten fjärranalysindikerade skogar med naturmiljövärden.

Syfte

Dokumentets syfte är att tillhandahålla dokument som beskriver dataproduktens informationsmodell, användningsområden, datakvalitetsinformation, tillkomsthistoria samt metoder för tillhandahållande.

Krav och rekommendationer

Krav på dataprodukten uttrycks i respektive tillämpligt avsnitt av denna dataproduktspecifikation (kapitel 4 och kapitel 6).

Innehållsförteckning

1	Översikt.....	7
1.1	Information om dataproduktspecifikationen	7
1.2	Termer och definitioner.....	8
1.3	Information om dataprodukten.....	9
2	Dataproduktens omfattningar.....	10
3	Identifiering av dataprodukten.....	11
	Datainnehåll och struktur	12
3.1	Beskrivning	12
3.1.1	Begreppsmodell.....	12
3.1.2	Informationsmodell	12
3.1.3	Informationsutbytesmodell.....	12
4	Referenssystem	12
5	Datakvalitet/Kvalitetskrav	13
5.1	Vad innebär datakvalitet i kontext av fjärranalys och AI-framtagna dataprodukter?	13
5.1.1	Utvärdering mot träningsdata.....	13
5.1.2	Visuell granskning.....	13
5.1.3	Utvärdering med data från riksskogstaxeringen.....	14
5.2	Utvärdering genom fältinventering	14
	Kartunderlagets prediktioner har utvärderats mot data från fältinventering av skogar med både höga och låga naturvärden (se bilaga 13.6).	14
5.3	Datakvalitetselement/Datakvalitet.....	14
5.3.1	Logisk konsistens – Format.....	14
5.3.2	Logisk konsistens – Konceptuell konsistens.....	15
5.3.3	Fullständighet - Övertalighet.....	15
5.3.4	Fullständighet – Brist	15
5.3.5	Användbarhet	16
6	Datafångst/Tillkomsthistorik	16
6.1	Datakällor	16
6.2	Produktion/metod	17
6.3	Träningsdata	18
7	Underhåll och uppdatering.....	27
7.1	Underhållsfrekvens.....	27
7.2	Uppdatering	27
8	Metadata datamängd	27
8.1	Metadataelement för rapportering av datakvalitet.....	28
9	Tillhandahållande och användningsexempel	29

9.1	Leveransmetoder och användningsexempel.....	29
9.1.1	Nedladdning	29
9.1.2	Format nedladdning.....	29
9.1.3	REST-tjänst	29
9.1.4	Wms-tjänst.....	29
9.2	Användningsexempel av dataprodukt	29
9.3	Presentationsregler	29
10	Övrig information.....	29
11	Referenser.....	29
12	Bilagor.....	30
12.1	Metadata, datakvalité exempel	30
12.2	AI och modeller.....	30
12.3	Kvalitetssäkring.....	31
12.4	Utvärdering mot Riksskogstaxeringens provytor.....	31
12.5	Utvärdering mot fältinventering	39
12.6	Begreppsmodell.....	42
12.7	Informationsmodell	42
12.8	Informationsutbytesmodell.....	42

1 Översikt

1.1 Information om dataproduktspecifikationen

Titel	Dataproduktspecifikation för Indikationer på skog med naturvärden
Datum	2026-02-16
Ansvarig part	Skogsstyrelsen
Språk	Svenska
Ämnesområde	Miljö
Syfte	Ett sammanhållet dokument med beskrivning av dataprodukten fjärranalysindikerade skogar med naturmiljövärden och dess kvalitetskrav samt egenskaper för datadelning
Identifierare	DPS:Skogsstyrelsen_naturmiljövärden_v1-00
Ajourhållning av dataproduktspecifikation	Förvaltning och ajourhållning av denna dataproduktspecifikation sker under Skogsstyrelsens försorg

1.2 Termer och definitioner

Generella och specifika termer som används i detta dokument.

Specifika termer	
Term	Beskrivning
Epoch	AI-modellen har tränat på alla exempel i träningsdatasetet en gång.
Recall	Mäter hur stor andel av de verkligt positiva fallen som AI-modellen lyckas identifiera korrekt
Precision	Mäter hur stor andel av de fall som AI-modellen klassificerar som positiva som faktiskt är korrekta.

Generella termer	
Term	Beskrivning
Begreppsmodell	En begreppsmodell beskriver verksamhetens centrala begrepp och deras relationer på en övergripande nivå. Den är oberoende av teknisk implementation och används för att skapa en gemensam förståelse av vad informationen betyder i verksamheten.
Informationsmodell	En informationsmodell beskriver hur informationen är strukturerad, till exempel informationsobjekt, attribut och relationer. Den konkretiserar begreppsmodellen och definierar vilken information som ska finnas och hur den hänger ihop.
Informationsutbytesmodell	En informationsutbytesmodell beskriver hur informationen ska utbytas mellan system och via tjänster. Den visar vilken del av informationsmodellen som exponeras och i vilket format eller struktur den levereras.

1.3 Information om dataprodukten

Namn	Indikationer på skog med naturvärden
Beskrivning	Dataprodukten <i>Indikationer på skog med naturvärden</i> består av kartunderlag som visar indikationer på att ett skogsområde har, eller saknar, naturskogsegenskaper. Kartunderlagens upplösning är 10x10meter (pixel), där värdet på respektive pixel är modellens prediktion av förekomst av naturskogsegenskaper. Dataprodukten har tagits fram med hjälp av maskininlärning, och består av tre delprodukter som tagits fram separat för olika geografiska regioner: • Fjärranalysindikerade skogar med naturskogsegenskaper i Sydost boreal region • Fjärranalysindikerade skogar med naturskogsegenskaper i Nordost boreal region • Fjärranalysindikerade skogar med naturskogsegenskaper i Alpin region I kartjänsten visas de olika delprodukterna som ett gemensamt kartunderlag.
Utsträckning (rumslig)	Norra Sverige alpin, nordväst boreal och sydostboreal region enligt EU:s indelning, se karta. Det innefattar länen Norrbotten, Västerbotten, Västernorrland, Jämtland, Gävleborg, och Dalarna, samt delar av Värmland, Västmanland, Örebro och Uppsala län.
Syfte	Modellernas prediktioner, i form av heltäckande kartor över skogsmarken, syftar till att visa var det finns indikationer på att skogen har naturskogsegenskaper ("höga naturvärden") eller saknar naturskogsegenskaper ("låga naturvärden"). Underlaget ska ses som ett komplement till fältinventering.

Tänkta användare	Skogsägare, myndigheter och övriga intressenter.
Datakällor	Se kapitel 7.
Tillkomsthistorik	Underlagen för de tre regionerna togs fram i april 2025. Underlagen klipptes mot Skogsstyrelsens skikt med avverkad areal från november 2025, och viss annan bearbetning gjordes i december 2025 (se kapitel 7 för mer information).
Ajourhållning	Uppdateringar där avverkade skogar klipps bort görs årligen. Vidareutveckling och uppdatering av den här versionen görs under 2026, och efter det uppdateras dataunderlagen när nytt data som väsentligt kan förbättra modellen finns tillgängligt (se kapitel 8).

2 Dataproduktens omfattningar

Detta dokument och dess beskrivningar gäller för hela dataprodukten som en sammanhållen datamängd. Dataprodukten är inte uppdelad i delmängder med avvikande innehåll, kvalitet eller leveranskrav och beskriver objektens aktuella tillstånd vid leveranstillfället.

Delmängder skulle i framtiden kunna definieras om det finns behov av att särskilja delar av dataprodukten med olika egenskaper, exempelvis baserat på:

- Tematisk indelning
- Kvalitetsnivå
- Aktualitet
- Leveransform

Sådana delmängder är för tillfället inte definierade i detta dokument.

3 Identifiering av dataprodukten

Titel	Indikation på skog med naturvärden
Alternativ titel	-
Produktnummer	01
Versionsnummer	2026:1.00
Syfte	-
Ämnesområde	Miljö
Metod för rumslig representation	Raster
Rumslig upplösning	10x10 (m)
Geografisk utsträckning	Norra Sverige (Alpin, Nordväst boreal och Sydostboreal region)
Kompletterande information	-

Datainnehåll och struktur

3.1 Beskrivning

Datamängden innehåller indikationer på skogar med naturvärden och innehåller följande information:

3.1.1 Begreppsmodell

Dataproduktspecifikationen kommer att uppdateras med en begreppsmodell som beskriver centrala begrepp för naturvärden och indikationer på naturvärden, se bilaga i kapitel 12.6.

3.1.2 Informationsmodell

Dataproduktspecifikationen kommer att uppdateras med en informationsmodell som visar datamängdens innehåll se bilaga i kapitel 12.7.

3.1.3 Informationsutbytesmodell

Dataproduktspecifikationen kommer att uppdateras med en informationsutbytesmodell för utbytesformat se bilaga i kapitel 12.8.

4 Referenssystem

Rumsligt referens-system	Nationella koordinatsystem SWEREF99 TM (EPSG:3006)
Temporalt referens-system	Gregorianska kalendern UTC +1 sommartid (CEST) UTC ± 0
Referenssystemets omfattning	-

5 Datakvalitet/Kvalitetskrav

I detta kapitel redogörs för de kvalitetskrav som finns på dataprodukten och således har bäring på användning av produkten. Resultat av dem i detta kapitel angivna kvalitetsparametrar finns i metadata för dataprodukten.

För en kortare beskrivning av AI i kontexten för denna produkt, se kapitel 13.3.

5.1 Vad innebär datakvalitet i kontext av fjärranalys och AI-framtagna dataprodukt?

För att utvärdera kvaliteten på produkterna inom projektet har en kvalitetssäkringsprocess inklusive olika former för utvärdering tagits fram (se bilaga 13.4). En nyckelfaktor är ett nära samarbete mellan datavetare och domänexperter inom naturvård. Ett noggrant urval av träningsytor ligger till grund för att skapa en så rättvis statistisk modell som möjligt. En rättvis modell innebär en modell utan bias.

Resultatet av iterationerna är ett kartunderlag som visar indikationer på skog naturvärden.

5.1.1 Utvärdering mot träningsdata

Modellerna i de olika regionerna har utvärderats med ett slumpmässigt urval av ca 20% av träningsdata, som sparades undan och inte användes i träningen av modellen (se bilaga 13.4,13.5)

Modellerna i de olika regionerna har utvärderats på åsidosatta data från träningen. Valideringsdata är 18% och testdata är 10%. Valideringsdata används under träningens gång för att mäta träffsäkerheten efter varje epoch. När modellen är färdigtränad testas modellen mot testdata. Men för att bestämma vilken modell som presterat bäst har vi främst tittat på träffsäkerheten mot valideringsdatat. Denna utvärdering skapar en bias till valideringsdatat och den säger inte hur bra modellen fungerar i alla skogar, utan bara hur bra den fungerar i det annoterade datasetet.

Mätmetoden under träningen är pixelbaserad. Det betyder att ett stort objekt har större inflytande på modellen än ett litet objekt.

Modellens träffsäkerhet inom kända ytor har beräknats som *Precision* och *Recall*, men det säger inget om träffsäkerheten utanför kända ytor. *Precision* beskriver andelen pixlar som modellen klassar som naturvärden som faktiskt var naturvärden. *Recall* beskriver vilken andel av alla faktiska naturvärden som modellen klassade som naturvärden.

5.1.2 Visuellt granskning

Personer med naturvårdskompetens och skoglig kompetens har visuellt granskat utvecklingsunderlag med AI-modellens prediktioner mot andra kartunderlag och/eller beståndsdata för bedömning av kvalitet och

användningsområden. Granskningen har utförts i två steg, först internt på Skogsstyrelsen där resultat från olika AI-modeller jämförts för att välja ut den som presterar bäst, som ett viktigt komplement till utvärdering av modellresultaten genom Precision och Recall. Därefter har det bästa utvecklingsunderlaget lämnats till en extern referensgrupp för utvärdering och bedömning av kvalitet, brister och användningsområden.

5.1.3 Utvärdering med data från riksskogstaxeringen

Kartunderlagets prediktioner har utvärderats mot riksskogstaxeringens provytor, där indikationsvärden i kartunderlaget jämförts med provytor med kända naturvärden samt provytor med skog i olika åldersklasser (se bilaga 13.5)

5.2 Utvärdering genom fältinventering

Kartunderlagets prediktioner har utvärderats mot data från fältinventering av skogar med både höga och låga naturvärden (se bilaga 13.6).

5.3 Datakvalitetselement/Datakvalitet

Kapitel	Kategori	Datakvalitetselement	Omfattning
6.3.1	Logisk konsistens	Format	datamängd
6.3.2	Logisk konsistens	Konceptuell konsistens	geodataobjekt
6.3.3	Fullständighet	Övertalighet	datamängd
6.3.4	Fullständighet	Brist	datamängd
6.3.5	Användbarhet		datamängd

5.3.1 Logisk konsistens – Format

Krav	Att fastställd lagringsstruktur överensstämmer med informationslagringsmodell
Datakvalitetskategori	Logisk konsistens
Datakvalitetselement	Format
Kvalitetsmått	
Acceptansnivå	
Definition	
Beskrivning	Lagringsstruktur för tillhandahållande stämmer med den struktur som anges i kapitel 4

5.3.2 Logisk konsistens – Konceptuell konsistens

Krav	Geodataobjekt skall överensstämma med informationsmodellen
Datakvalitetskategori	Logisk konsistens
Datakvalitetselement	Konceptuell konsistens
Kvalitetsmått	Antal objekt
Acceptansnivå	100%
Definition	-
Beskrivning	Geodataobjekt i dataproduct skall egenskapsmässigt (attribut) överensstämma med modellen som finns definierad i kapitel 4

5.3.3 Fullständighet - Övertalighet

Krav	Dataprodukten har en viss övertalighet av objekt
Datakvalitetskategori	Fullständighet
Datakvalitetselement	Övertalighet
Kvalitetsmått	Boolean
Acceptansnivå	Genomgått kvalitetssäkring och bedömd som acceptabel av Skogsstyrelsen
Definition	Felaktiga objekt som fångats upp på grund av brister i AI-modellen och kvalitet på indata
Beskrivning	Vissa objekt förekommer i datamängden trots att motsvarande objekt inte finns i verkligheten. Det kan förekomma så kallade falskt positiva objekt vilket kan skapa en övertalighet, krav är att 90% av objekten i datamängd skall vara korrekt identifierade.

5.3.4 Fullständighet – Brist

Krav	Dataprodukten saknar vissa objekt (skog med naturvärden)
Datakvalitetskategori	Fullständighet
Datakvalitetselement	Brist
Kvalitetsmått	Boolean

Acceptansnivå	Genomgått kvalitetssäkring och bedömd som acceptabel av Skogsstyrelsen
Definition	Saknade objekt (skog med naturvärden) som inte fångats upp på grund av brister i AI-modellen och kvalitet på indata
Beskrivning	AI-modellens prediktioner begränsas av olika faktorer

5.3.5 Användbarhet

Krav	Dataprodukten har utvecklats genom feedback från intern myndighetspersonal och från extern referensgrupp (se kap 13.4)
Datakvalitetskategori	Användbarhet
Datakvalitetselement	-
Kvalitetsmått	Boolean
Acceptansnivå	Dataprodukt granskad av referensgrupp och anses som användbar inom angiven kontext.
Definition	-
Beskrivning	Datakvalitetsparametern <i>Användbarhet</i> refererar till att/om feedback levererats av intressenter inom referensgruppen.

6 Datafångst/Tillkomsthistorik

6.1 Datakällor

Dataunderlagen bygger på senaste [Skogliga grunddata](#) (2018–2024, löpande uppdateringar efter skanning omdrev 2) samt [Nationella marktäckedata](#) (2015–2018). Dataunderlagen bygger också på flygbilder från 1950- och 1960-talen.

Tabell 1: Dataunderlag för i AI-modellerna för regionerna sydost boreal, nordväst boreal och alpin.

Datakällor	SO	NV	Alpin
Skogliga grunddata: volym (m ³ sk/ha, grundytbevåg medelhöjd (dm), grundyta (m ² /ha), grundytbevåg medeldiameter (cm) och Löv/Avlöv (Med löv/utan löv), dvs. band 1,2, 3, 4, 9	x	x	x
Tillväxt (beräknad utifrån skogliga grunddata; skillnad i trädhöjd mellan omdrev 1 och 2 av laserscanningen)	x	x	

Trädslag (kNN), SLU	x	x	x
Nationella höjdmodellen (lutning och lutningsriktning), Lantmäteriet.	x	x	x
Markfuktighetskartan, SLU	x	x	x
Skogsbeklädd mark utifrån ortofoton 1955-65 (AI-genererad), SKS intern	x	x	
Nationella marktäckedata (NMD) som filter för skogsmark	x	x	x

6.2 Produktion/metod

Separata AI-modeller har tagits fram för tre biogeografiska regioner; Alpin, Nordväst boreal och Sydost boreal region. Skogens egenskaper förväntas skilja sig åt mellan regionerna, och varje AI-modell tränas enbart med data från en specifik region.

AI-modellerna består av djupa konvolutionella neurala nätverk, så kallat Convolutional Neural Network (CNN). Modell arkitekturen är samma för alla modeller som kallas U-Net [\[1505.04597\] U-Net: Convolutional Networks for Biomedical Image Segmentation](#). För att ytterligare stärka modellens spatiala förmåga har det implementerats attention mekanismer [\[1804.03999\] Attention U-Net: Learning Where to Look for the Pancreas](#). Till skillnad från vanligt förekommande U-Net modeller så är vår arkitektur specifik anpassad för att läsa in många olika bildkanaler samtidigt. Jämfört med många andra domäner där det räcker med 3-kanaliga bilder (RGB) för att göra sin analys.

Modellerna analyserar de specifika valda fjärrdataunderlag för att göra sin prediktion. Det betyder att alla fjärrdataunderlag behöver vara spatialt tillgängliga för där analysen ska ske.

Träningsdatat är begränsat till skog och därmed tränas modellen endast mot områden med skog. Detta betyder att modellerna inte är tillförlitliga på områden där det saknas skog. Detta hanteras genom att maska bort prediktioner där skog saknas enligt Nationellt marktäckedata (NMD).

Valet att använda CNN i stället för ett mer traditionellt Feedforward Neural Network (FNN) möjliggör att modellerna får inflytande från omkringliggande pixlar för en pixels prediktion. Ett CNN skalar ner en bilds upplösning i så kallat egenskapskartor för att få ut viktiga mönster i olika områden av en bild. Den gör det i flera steg, och det är definierat i modellarkitekturen hur många gånger den ska skalas ned. I vårt fall har vi valt att skala ner bilden till halva upplösningen 4 gånger. Det betyder att när bilden är som mest nedskalad i egenskapskartor så täcker en egenskapspixel 160 meter. Alltså 16 pixlar över de 10 meters pixlarna vårt fjärrdata har. Viktigt att veta är att det betyder inte nödvändigtvis att 160 meter har inflytande på prediktionen. Om det visar sig att det inte är statistiskt användbart vid träning av modellen, så kommer den funktionaliteten reduceras till väldigt liten påverkan.

Modellerna har tagits fram med djupa neurala nätverk, som implementerar en modifierad version av en Attention U-Net algoritmen ([Attention U-Net: Learning Where to Look for the Pancreas](#)). Närmare beskrivning av metoden för en vetenskaplig publikation (inte review-granskad) finns här [Deep learning segmentation model produces accurate biodiversity predictions at 10 m spatial resolution](#).

Modellerna bygger på att det finns heltäckande dataunderlag. Data för tillväxt har visat sig vara viktigt för att förklara resultaten i modellerna. Eftersom data för tillväxt bygger på att det finns laserdata från två laserskanningsperioder är datat inte rikstäckande. Vi har därför valt att inte använda tillväxt i modellen för alpin zon, och att producera två betaversioner med och utan tillväxt i Nordostboreal zon. I betaversionen med tillväxt Nordboreal zon saknas därför data i vissa områden där det inte finns två laserscanningar ("hål i kartan").

Ett filter från Nationellt marktäckedata (NMD) har tillämpats för att endast inkludera mark klassad som skogsmark. Även mark som klassats som temporärt ej skogsmark i NMD är bortmaskad.

6.3 Träningsdata

Träningsdata som har använts är binärt. Antingen har skogen naturvärden i träningsdatat eller så har den inte det. Modellen behandlar frågan den ska lösa som ett binärt problem utifrån denna träningsdata. I sina prediktioner strävar då modellen efter att klassa varje pixel i någon av dessa kategorier som värden mellan 0 och 1, och resultaten visar att prediktionerna är koncentrerade till pixelvärden $<0,1$ och $>0,9$.

En översikt av det träningsdata som använts i AI-modellerna för de tre olika regionerna visas i tabell 2. Viss bearbetning har gjorts för att förbättra kvaliteten på träningsdatat. För träningsdatat för höga naturvärden har objekt som har lite eller ung skog uteslutits (objekt som har en medianhöjd under 3 meter). För alla objekt har en negativ buffert på 20 meter använts för att minska kantzonseffekter, och objekt som är mindre än 0,5 hektar har exkluderats.

Träningsdatats egenskaper och representativitet avseende olika nyckelbiotopstyper, storleksklasser och geografisk placering visas i figurerna. Datat representativitet påverkar AI-modellens förväntade resultat, där träningsdata som är mer vanligt kan förväntas identifieras i större utsträckning av AI-modellen. Biotoper som är mer ovanliga i träningsdatat eller svåra att upptäcka i fjärranalysdata som exempelvis lövskogsbiotoper, brandfält, källor och lodytor, är svårare att upptäcka för AI-modellen.

Tabell 2: Total area träningsdata i hektar som använts i AI-modellerna för de olika regionerna, sydost boreal (SO), Nordväst boreal (NV) och Alpin. Arealen har avrundats till närmaste heltal. Antal objekt som använts för träning visas inom parentes.

Träningsdata höga naturvärden	SO	NV	Alpin
Nyckelbiotoper (Skogsstyrelsen och storskogsbruket)	63 492 (17 620)	133 487 (19 157)	46 855 (4 429)
Skogliga livsmiljötyper från NNK (Fullgod och icke-fullgod, fältbesökt eller granskad)	66 316 (10 050)	183 180 (7 753)	94 843 (1 601)
Träningsdata ej höga naturvärden			
Produktionsskogar 30-80 år	57 029 (8759)	52 659 (3 924)	1 052 (102)
Tidigare hyggen 1990-2000	163 374 (36 715)	531 033 (44 947)	19 642 (3040)
Tidigare hyggen 2000-2015	189 228 (61 267)	498 924 (47 307)	24 327 (3 984)
Skogsbeklädd mark utifrån flygfoton 1955-65			10 669 (402)

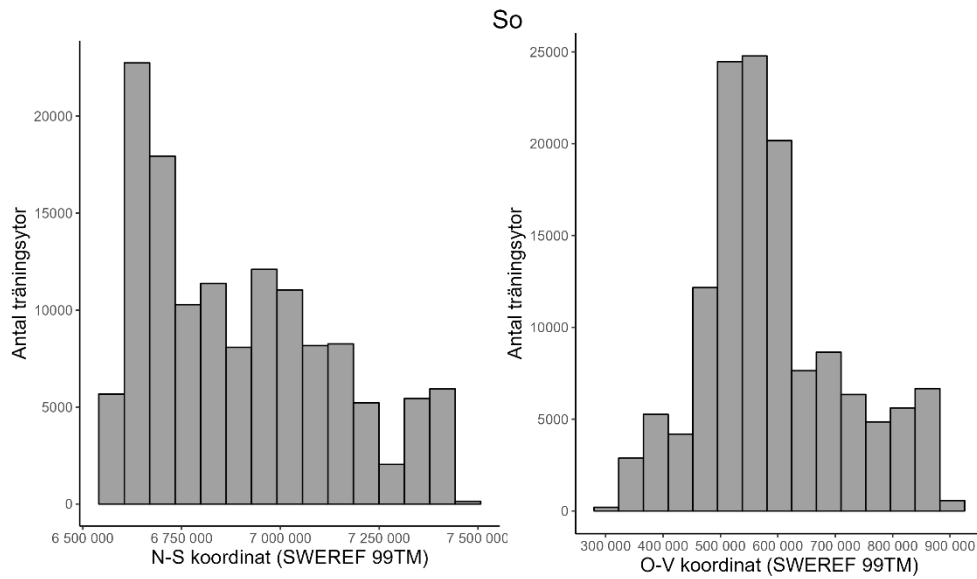
Tabell 3: Totalt antal träningsobjekt av de tio vanligaste livsmiljötyperna (sett över alla tre regioner) som använts inom de olika regionerna sydost boreal (SO), Nordväst boreal (NV) och Alpin, samt totalt antal objekt

Naturtyp	SO	NV	Alpin	Totalt antal objekt
9010	8235	5600	925	14760
9006	123	84	6	213
9050	103	62	41	206
9030	102	0	0	102
9009	61	17	0	78
9060	39	6	12	57
9040	0	4	47	51
9070	14	10	4	28
9080	23	3	0	26

9008	3	7	0	10
------	---	---	---	----

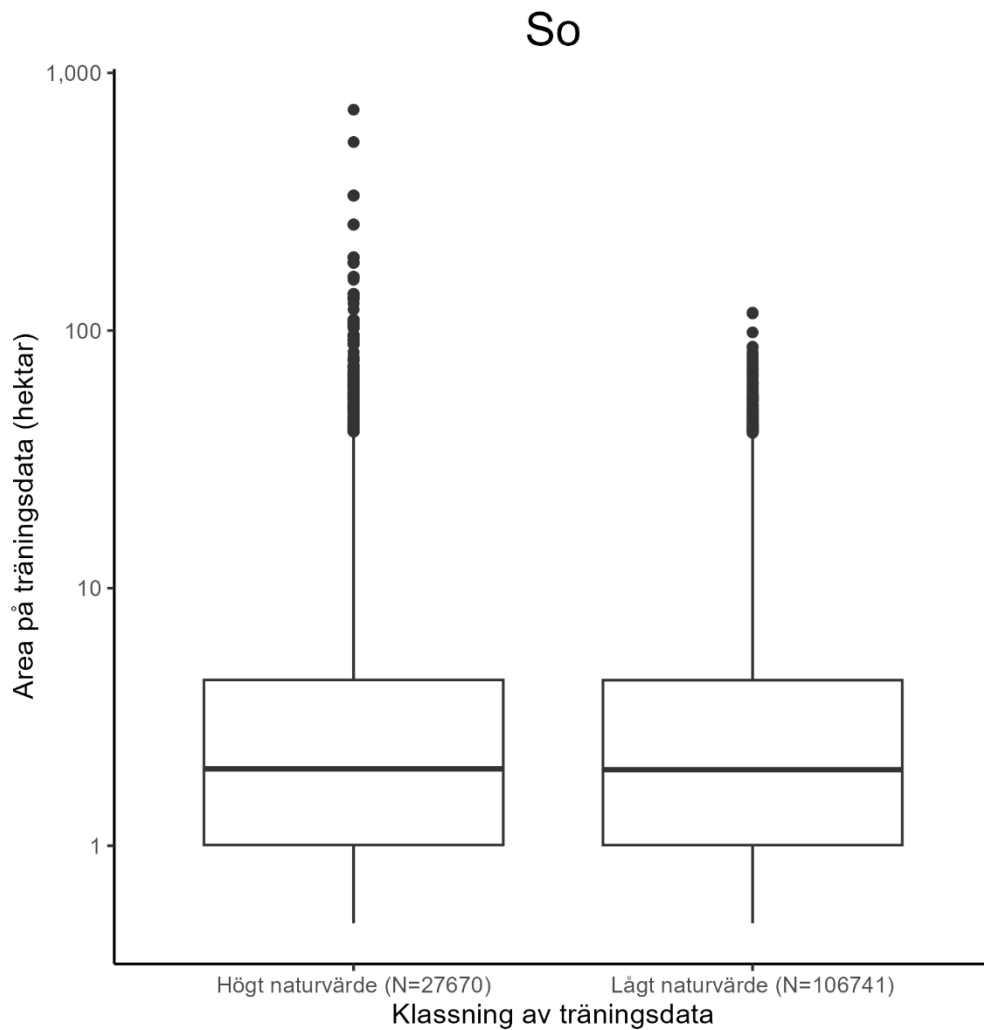
Tabell 4: De tio, till antalet, vanligast förekommande primära biototyperna inom nyckelbiotoper som använts som träningsdata i Sydostboreal region. Detta inkluderar såväl Skogsstyrelsens som storskogsbrukets egna nyckelbiotoper, i den mån information om biotyp finns tillgänglig.

Biototyp	Antal objekt
Barrskog	2357
Barnaturskog	2316
Lövrik Barrskog	714
Gransumpskog	384
Bergbranter	372
Hällmarksskog	336
Kalkbarrskog	253
Skogsbäck	190
Lövbränna	181
Blandsumpskog	176



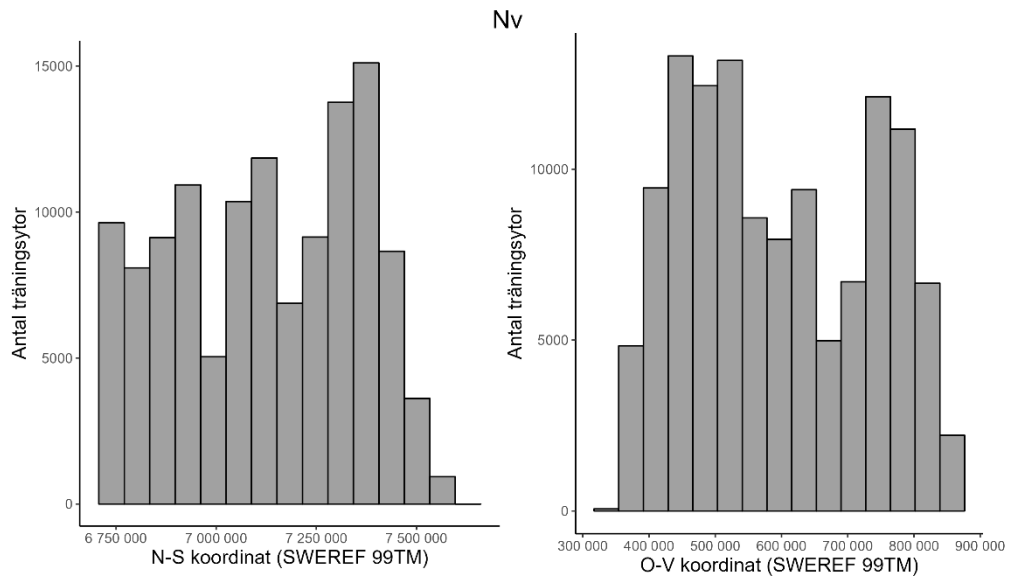
Figur 1: Histogram som visar spridningen av centrumkoordinaterna för alla träningsytor i Sydostboreal (SO)

Histogrammen ovan visar den nord-sydliga (vänster) och öst-västliga (höger) spridningen av centrumkoordinaterna för alla träningsytor i Sydostboreal (SO). Figuren visar att vi inom det område som utgör regionen finns en större representation av träningsdata från den södra delen jämfört med den norra delen. I öst-västlig spridning finns en något större representation av träningsdata från den västra delen av regionen.



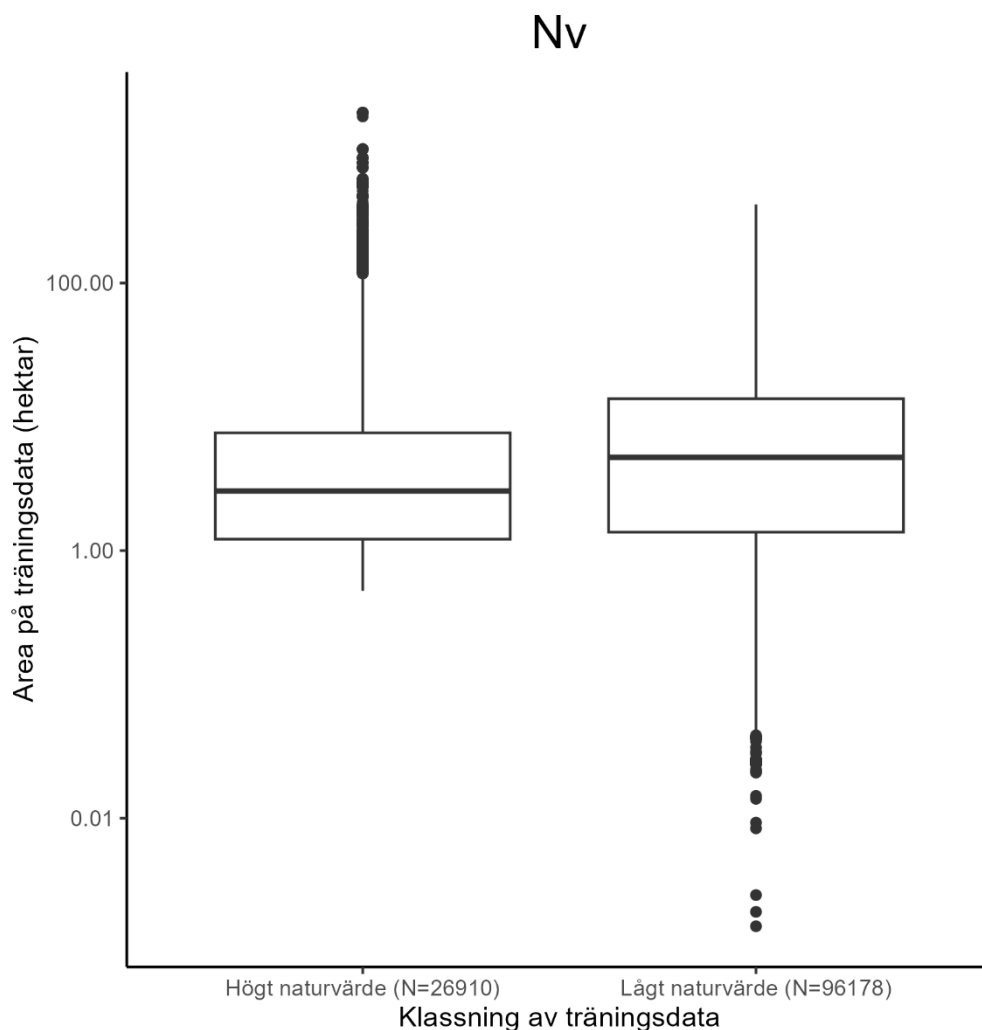
Figur 2: Boxplot som illustrerar storleksfördelningen Sydostboreal

Boxplot ovan illustrerar storleksfördelningen på de träningsobjekt som använts i region Sydostboreal, uppdelat på höga (vänster) och låga (höger) naturvärden. N refererar till mängden träningsobjekt i varje kategori. Medianvärde på arealen för höga naturvärden är 1,99 hektar, och 1,97 hektar för låga naturvärden. Notera att Y-axeln är logaritmisk.



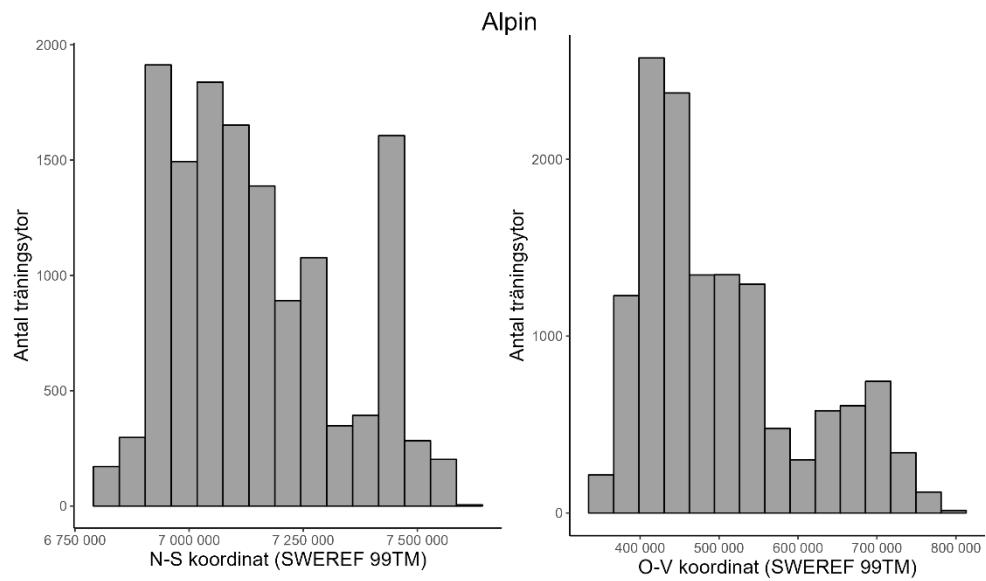
Figur 3: Histogram för spridningen av centrumkoordinaterna för alla träningsytor i Nordvästboreal (NV)

Histogrammen ovan visar den nord-sydliga (vänster) och öst-västliga (höger) spridningen av centrumkoordinaterna för alla träningsytor i Nordvästboreal (NV). Figuren visar att vi inom det område som utgör regionen finns en något större representation av träningsdata från den norra delen jämfört med södra delen. I öst-västlig spridning finns en viss underrepresentation av träningsdata från mitten av regionen



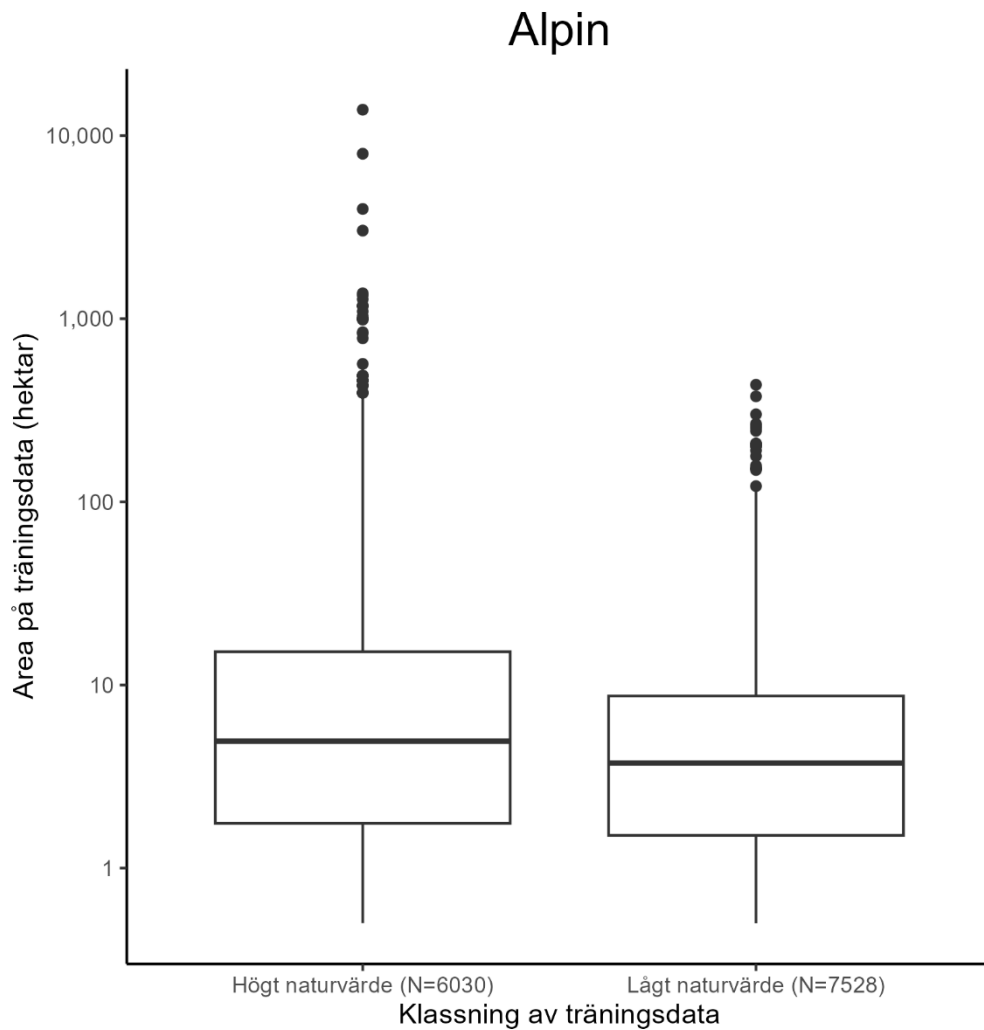
Figur 4: Boxplot som illustrerar storleksfördelningen på träningsobjekt Nordvästboreal

Boxplot ovan illustrerar storleksfördelningen på de träningsobjekt som använts i region Nordvästboreal, uppdelat på höga (vänster) och låga (höger) naturvärden. N refererar till mängden träningsobjekt i varje kategori. Medianvärde på arealen för höga naturvärden är 2,79 hektar, och 4,98 hektar för låga naturvärden. På grund av klippning i träningsdata i samband med avgränsning av regionen har det uppkommit ett mindre antal extremt små träningsobjekt bland de låga naturvärdena. Notera att Y-axeln är logaritmisk



Figur 5: Histogram för spridningen av centrumkoordinaterna för alla träningsytor i Alpin

Histogram ovan visar den nord-sydliga (vänster) och öst-västliga (höger) spridningen av centrumkoordinaterna för alla träningsytor i Alpin. Figuren visar att vi inom det område som utgör regionen finns en något större representation av träningsdata från den södra delen jämfört med den norra delen. I öst-västlig spridning finns en viss underrepresentation av träningsdata från mitten av regionen.



Figur 6: Boxplot som illustrerar storleksfördelningen på träningsobjekt Alpin

Boxplot ovan illustrerar storleksfördelningen på de träningsobjekt som använts i region Nordvästboreal, uppdelat på höga (vänster) och låga (höger) naturvärden. N refererar till mängden träningsobjekt i varje kategori. Medianvärde på arealen för höga naturvärden är 4,94 hektar, och 3,75 hektar för låga naturvärden. Notera att Y-axeln är logaritmisk.

7 Underhåll och uppdatering

7.1 Underhållsfrekvens

Produkten klipps mot utförda avverkningar årligen baserat på Skogsstyrelsens data.

7.2 Uppdatering

Kartunderlagen uppdateras när ny käll- och/eller träningsdata tillkommer som bedöms väsentligt kunna förbättra kartunderlaget.

För aktuell information rörande datamängdens aktualitet och uppdateringsfrekvenser hänvisas till datamängdens metadata i Skogsstyrelsens metadatakatalog.

8 Metadata datamängd

Metadata anges på datamängdsnivå enligt den Nationella metadataprofilen (NMDP 4.0). Tabell anger obligatoriska element för metadata på datamängdsnivå (se bilaga 13.1,13.2)

NMDP	Element	Multiplicitet
6.3.1	Titel	1
6.3.3	Sammanfattning	1
7.1.1.1	Resurstyp	1
7.1.2.1	Identifierare för resurs	1..*
7.1.2.5	Ämnesområde	1..*
6.3.6	Nyckelord	1..*
6.3.13	Omskrivande rektangel	1..*
6.5.1	Distributör	1..*
7.1.4.3	Tillkomsthistorik	1
6.3.11	Villkor för tillgång och nyttjande	1..*
6.3.4	Resurskontakt	1..*
6.2.3	Metadatakontakt	1..*
6.2.4	Datum för metadata	1
6.2.2	Språk i metadata	1
6.2.6	Metadatastandard	1
7.2.1.2	Rumsligt referenssystem	1..*
7.2.2.1	Metod för rumslig representation	1..*

8.1 Metadataelement för rapportering av datakvalitet

Metadataelement namn	Format
ISO 19157 namn	Logisk konsistens
Exempel	False/True
Exempel XML	Se Bilaga

Metadataelement namn	Konceptuell konsistens
ISO 19157 namn	Logisk konsistens
Exempel	False/true
Exempel XML	Se Bilaga

Metadataelement namn	Övertalighet
ISO 19157 namn	Fullständighet
Exempel	~ 90 %
Exempel XML	Se Bilaga

Metadataelement namn	Brist
ISO 19157 namn	Fullständighet
Exempel	~ 90 %
Exempel XML	Se Bilaga

Metadataelement namn	Användbarhet
ISO 19157 namn	Användbarhet
Exempel	Text

Exempel XML	Se Bilaga
--------------------	-----------

Metadataelement namn	Tematisk noggrannhet
ISO 19157 namn	Klassificeringsnoggrannhet
Exempel	Float
Exempel XML	Se Bilaga

9 Tillhandahållande och användningsexempel

9.1 Leveransmetoder och användningsexempel

9.1.1 Nedladdning

[Ladda ner geodata - Skogsstyrelsen](#)

9.1.2 Format nedladdning

GeoTIFF

9.1.3 REST-tjänst

[REST - Skogsstyrelsen](#)

9.1.4 Wms-tjänst

[wms - Skogsstyrelsen](#)

9.2 Användningsexempel av dataprodukt

[Digitala värden i skog - DVIS](#)

9.3 Presentationsregler

Denna dataprodukt har inga vedertagna presentationsregler.

10 Övrig information

11 Referenser

[Attention U-Net: Learning Where to Look for the Pancreas](#)

[Deep learning segmentation model produces accurate biodiversity predictions at 10 m spatial resolution \(Research Square\)](#)

[Skogliga grunddata](#)

[Nationella marktäckedata](#)

Tobias Andermann, Johan Häggmark, Patrik Olsson et al. Deep learning segmentation model produces accurate biodiversity predictions at 10 m spatial resolution, 05 September 2024, PREPRINT (Version 1) available at Research Square [<https://doi.org/10.21203/rs.3.rs-4734879/v1>]

12 Bilagor

12.1 Metadata, datakvalité exempel

Endast exempel, saknas idag.

```
<gmd:report>
  <gmd:DQ_CompletenessOmission>
    <gmd:measureDescription>
      <gco:CharacterString>
        Fullständighet (brist) av AI-detekterade objekt. Modellerna analyserar
        de specifika valda fjärrdataunderlag för att göra sin prediktion.
      </gco:CharacterString>
    </gmd:measureDescription>
    <gmd:result>
      <gmd:DQ_DescriptiveResult>
        <gmd:statement>
          <gco:CharacterString>
            Objekt saknas, dessa fångas inte p.g.a. brister i modell.
          </gco:CharacterString>
        </gmd:statement>
      </gmd:DQ_DescriptiveResult>
    </gmd:result>
  </gmd:DQ_CompletenessOmission>
</gmd:report>
```

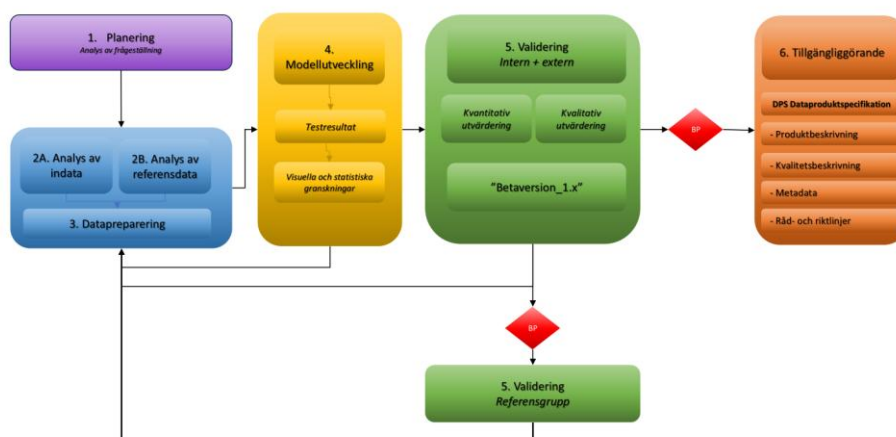
12.2 AI och modeller

Artificiell intelligens (AI) är teknik som gör det möjligt för datorer att lära sig saker genom att studera exempel, ungefär som människor lär sig av erfarenhet. I stället för att en utvecklare programmerar varje regel manuellt, får datorn se stora mängder data vilket leder till att den själv upptäcker mönster och samband.

En AI-modell kan ses som ett verktyg eller ett ”recept” som datorn använder för att fatta beslut. Genom att mata modellen med exempel – till exempel bilder, kartdata eller text – lär den sig hur olika situationer ser ut. Sedan kan den använda modellen för att göra bedömningar om nya områden den inte har sett förut.

I praktiken betyder det att en modell kan hjälpa till med sådant som att känna igen former i landskapet, förutsäga förändringar i naturen eller jämföra områden med varandra. Modellen ger inte alltid ett exakt svar och kräver kontroller men skapar ett systematiskt och konsekvent arbete och kan hantera mycket stora datamängder.

12.3 Kvalitetssäkring



Figur 7: Kvalitetssäkring görs under hela utvecklingsprocessen i olika steg.

Utvecklingen av kunskapsunderlagen är en iterativ process, där modellutveckling och utvärdering av resultat sker stegvis. En modell för kvalitetssäkring togs fram för alla underlag inom Digitala värden i skogen.

Innan utvecklingen startas planeras metoder och ingående datamängder. AI-modeller utvecklas och testas iterativt. Kvalitetssäkringen av resultaten består av olika moment, där en första visuell granskning/jämförelse och statistisk analys av olika resultat från AI-modeller görs. Då ett tillräckligt bra resultat erhållits, skapas en betaversion (utvecklingsunderlag) som granskas och bedöms internt på myndigheten med avseende på kvalitet och användbarhet. Godkänns den internt granskas och bedöms den sedan på liknande sätt av en extern referensgrupp från skogssektorn. Den externa referensgruppen kommer att delta i vidareutveckling av produkten under projektiden.

Då produkten utvärderats internt och externt och bedömts hålla tillräcklig kvalitet för användning, utvärderas dataunderlaget mot riksskogstaxeringens provytor och data från fältinventeringar. Kartunderlaget anpassas sen för tillgängliggörande i karttjänster och för nedladdning, och en dataproduktspecifikation tas fram (denna).

12.4 Utvärdering mot Riksskogstaxeringens provytor

Kartunderlagens prediktioner av naturvärden (indikationer mellan 0–1) utvärderades mot data från Riksskogstaxeringens provytor från åren 2020–2024. Endast odelade provytor på produktiv skogsmark användes. För utvärderingen tilldelades varje provyta en av fyra klasser A-D baserat på klassning av naturskogsegenskaper, samt beståndsålder och förväntade naturvärden.

- *Klass A* Skog med bekräftade naturskogsegenskaper; naturskog eller naturtyp. Vi förväntade oss också att medianen för indikationsvärdena skulle vara statistiskt högst i A jämfört med övriga klasser.

- *Klass B* Skog med en beståndsålder äldre än 140 år men som ej klassats som naturskog eller naturtyp. Vi förväntade oss att medianen för indikationsvärdena i provytorna i B skulle vara statistiskt högre än i C och D, och lägre än i A.
- *Klass C* Skog med en beståndsålder mellan 80–140 år (barrdominerad skog) eller 60–140 år (lövskog). Vi förväntade oss att medianen för indikationsvärdena i provytorna i C skulle vara statistisk lägre än i A och B, och högre än i D.
- *Klass D* Skog med en beståndsålder under 80 år (barrdominerad skog) eller under 60 år (lövskog). Vi förväntade oss också att medianen för indikationsvärdena skulle vara statistiskt lägst jämfört med övriga klasser.

Indikationsvärden är beräknade för varje 10 x 10 meters pixel och varierar mellan 0 och 1. Värden över 0,5 bedöms som höga naturvärden, och värden under 0,5 bedöms som låga naturvärden. Ett högre värde ger en starkare indikation på att pixeln innehåller skog med naturvärden.

För beräkningarna användes det högsta indikationsvärdet bland pixlar som överlappade provytan, för att inte kanteffekter skulle påverka resultatet. Kanteffekter uppstår eftersom AI-modellen tar hänsyn till omkringliggande bestånd i sin analys. Inom vissa provytor som ligger nära andra ägoslag eller bestånd uppstod därför en stor spridning i modellens klassning av indikerade naturvärden.

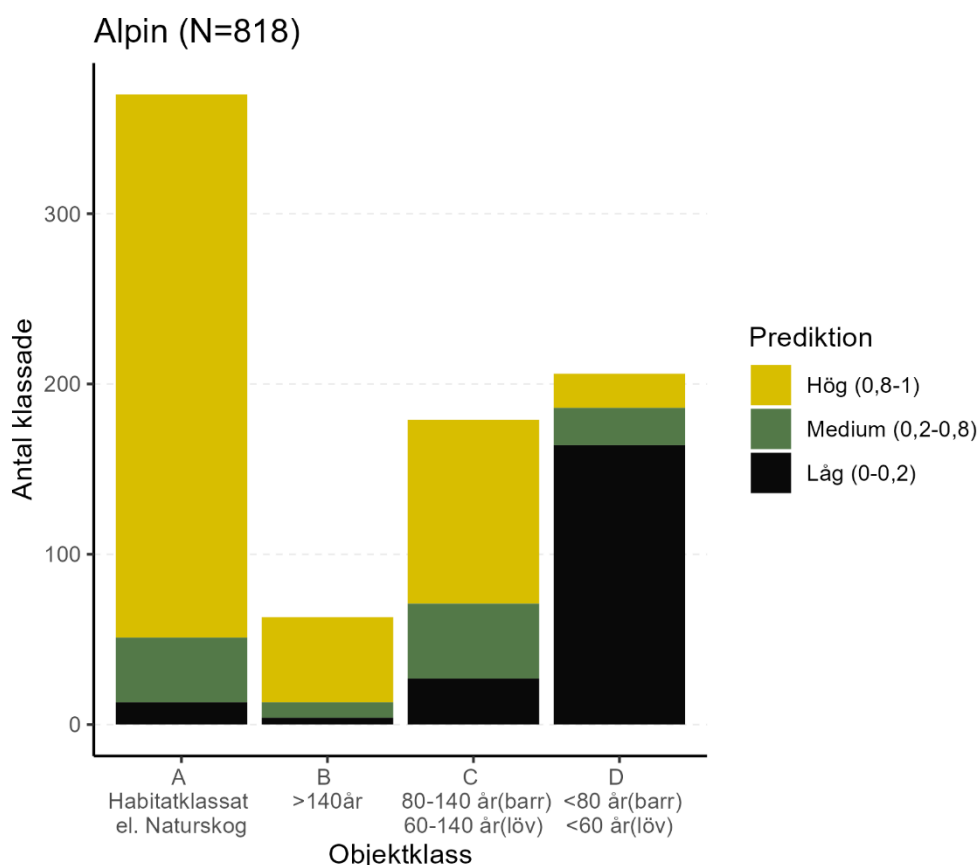
Den statistiska utvärderingen mot Riksskogstaxeringens provytor gjordes i R (version 4.1.2). Vi testade antagandet att indikationsvärdena skiljer sig åt mellan de olika klasserna (A högst och D lägst indikationsvärden) på det högsta pixelvärdet inom varje provyta med en generaliserad linjär modell (GLM) med en quasibinomial fördelning (nära binära data; antingen nära 0 eller nära 1), där klass A användes som jämförelse. Vi utförde även ett Tukey-test för parvisa jämförelser för alla nivåer. Separata analyser gjordes för tre geografiska områdena (Alpin, NV-Boreal, SO-Boreal).

En grafisk analys av fördelningen av indikationsvärden i olika klasser gjordes genom boxplotar (för spridning av data) samt stapeldiagram som visar fördelningen av indikationsvärden uppdelad på höga värden (0,8–1), mer osäkra prediktioner/medium värden (0,2–0,8), eller låga värden (0–0,2). Vi förväntade oss att provytorna i A skulle ha högst andel provytor med höga indikationsvärden (över 0,8). Vi förväntade oss att provytorna i B skulle ha en hög andel provytor med höga indikationsvärden (över 0,8). Vi förväntade oss att provytor i C skulle ha en blandning av höga och låga indikationsvärden (den mest osäkra klassen). Vi förväntade oss att provytorna i D skulle ha högst andel låga indikationsvärden (under 0,2).

I enlighet med våra prediktioner så visar resultaten från regressions- och parvisa analysen att indikationsvärdena är signifikant högre i provytor som habitatklassats eller klassats som naturskog (klass A) än i provytor som var yngre än 140 år och inte habitatklassats eller klassats som naturskog (klasserna C och D) för alla regioner ($p < 0,05$). I en av regionerna, sydost boreal, är

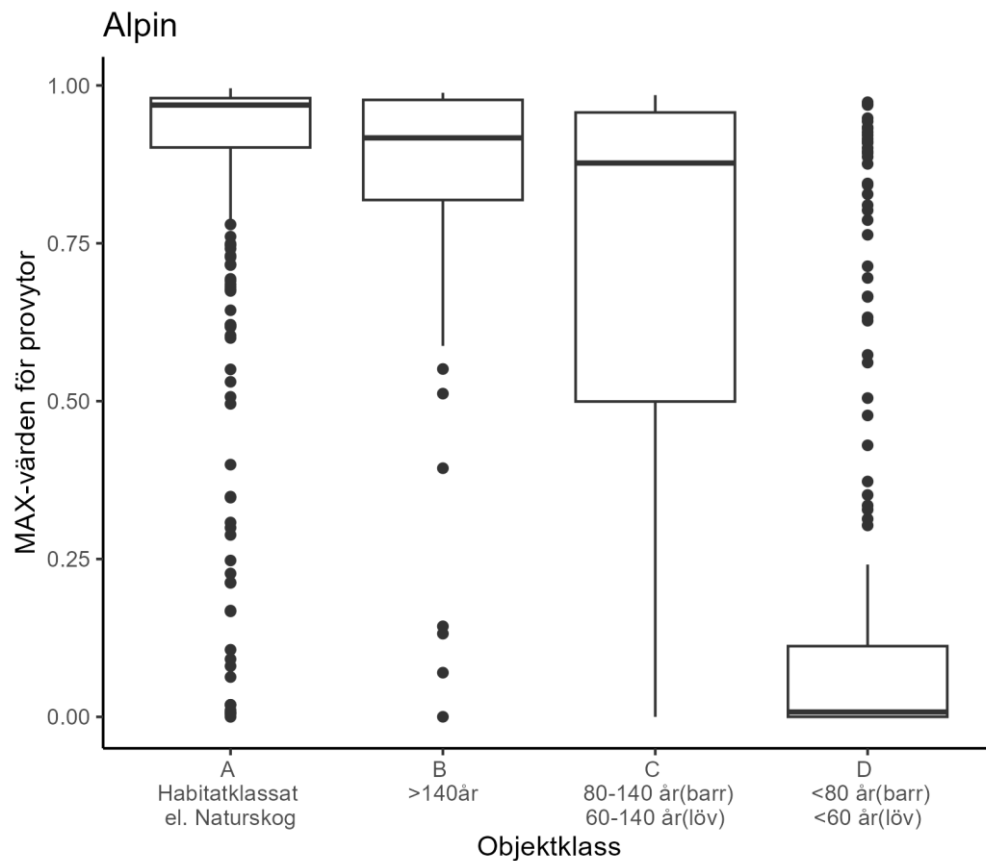
indikationsvärdena, i enlighet med våra prediktioner, statistisk högre i provytor med registrerade naturvärden (A) än i provytor med skog >140 år utan registrerade naturvärden (B) ($p < 0,05$). I motsats till våra prediktioner så skiljer sig indikationsvärdena mellan A och B i regionerna Alpin och Nordvästboreal. I linje med våra prediktioner så visar vår parvisa jämförelse efter justering av p-värdet för upprepade jämförelser ett signifikant högre indikatorvärde för klass A och B jämfört med klass C ($p < 0,05$), och provytor av klass C hade ett signifikant högre högsta pixelvärde än de av klass D i samtliga regioner ($p < 0,05$). I motsats till våra prediktioner så kunde vi inte detektera en signifikant skillnad mellan de högsta pixelvärdena i klass A och B i regionerna NV-Boreal eller Alpin.

Resultaten av den grafiska analysen visar att den största andelen provytor med höga indikationsvärden finns i A (Figurer 1,3,5) och lägst andel i D. Fördelningen av indikationsvärden i de olika klasserna i de tre regionerna visas i figurerna 2,4 och 6. I samtliga tre utvärderade regioner så ligger medianen för det högsta pixelvärdet i klasserna väldigt nära antingen 1 (högsta möjliga) eller 0 (lägsta möjliga), vilket visar att även om vi detekterar en signifikant skillnad i medianvärden mellan klasser, så är modellen nära binär.



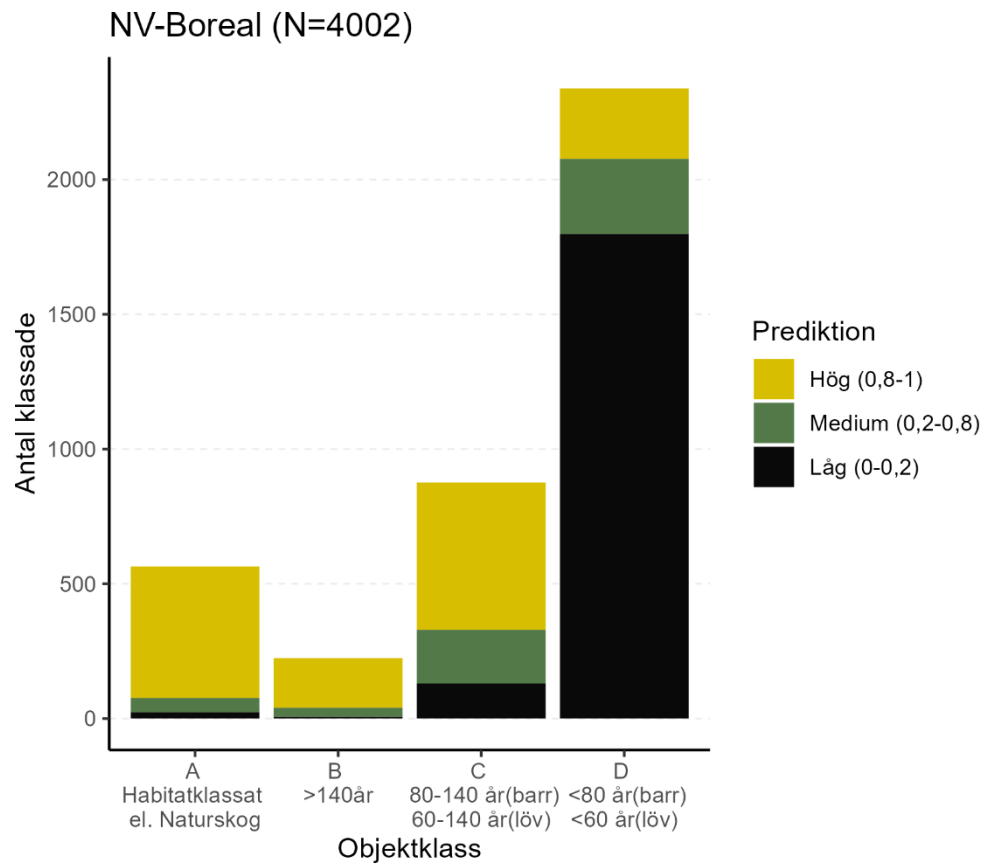
Figur 8: Fördelning av provytor

Bilden ovan visar fördelning av provytor med ett indikationsvärde som klassats som antingen högt (0,8–1), medium (0,2–0,8), eller lågt (0–0,2) inom Alpin region. N=antal provytor.

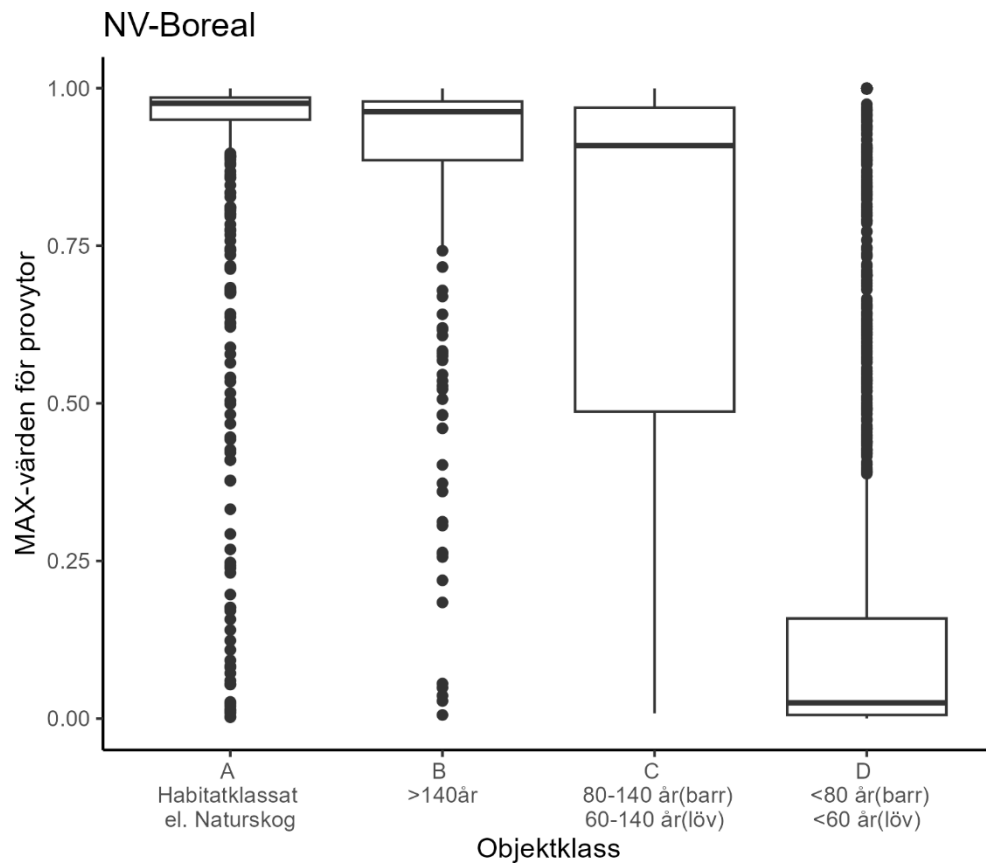


Figur 9: En boxplot som illustrerar fördelningen av högsta indikationsvärde Alpin region

I boxplot ovan illustreras fördelningen av högsta indikationsvärde inom provytor för alla objektklasser för alla provytor inom Alpin region. Grov horisontell mittenlinje illustrerar medianvärde, medan boxens övre och undre kant markerar 75% och 25% kvartilen, respektive.

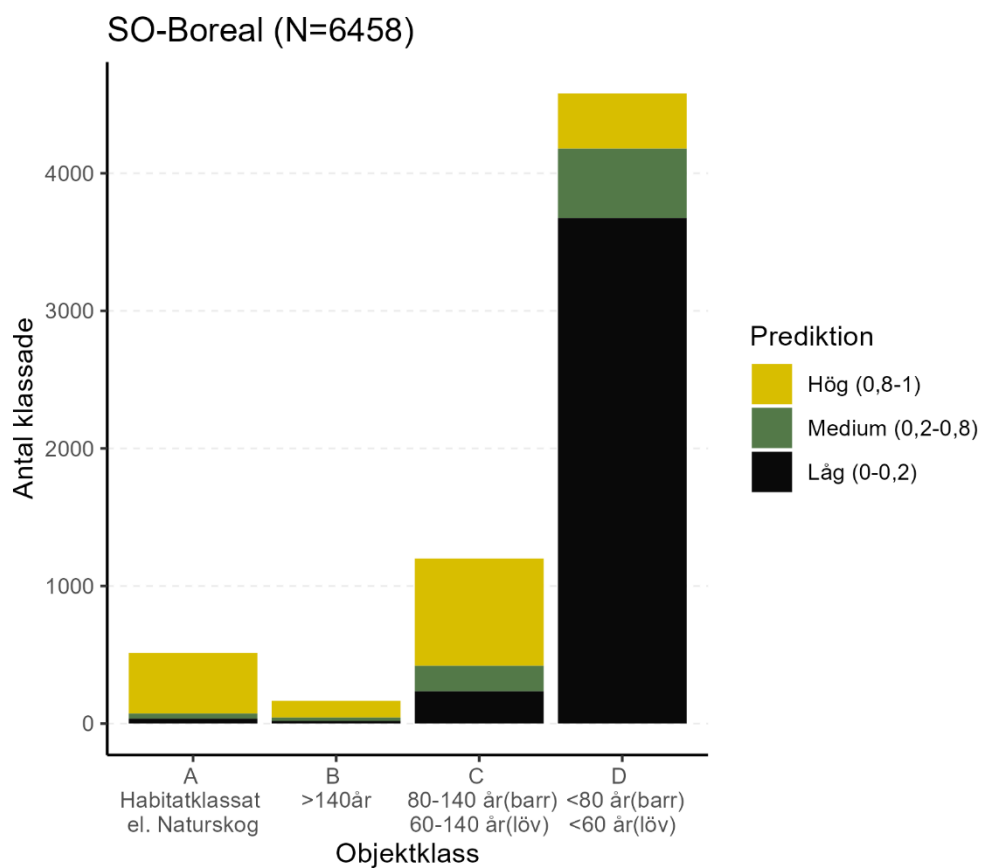


Figur 10: Fördelning av provytor med ett indikationsvärde som klassats som antingen högt (0,8–1), medium (0,2–0,8), eller lågt (0–0,2) inom NV-Boreal region. N=antal provytor

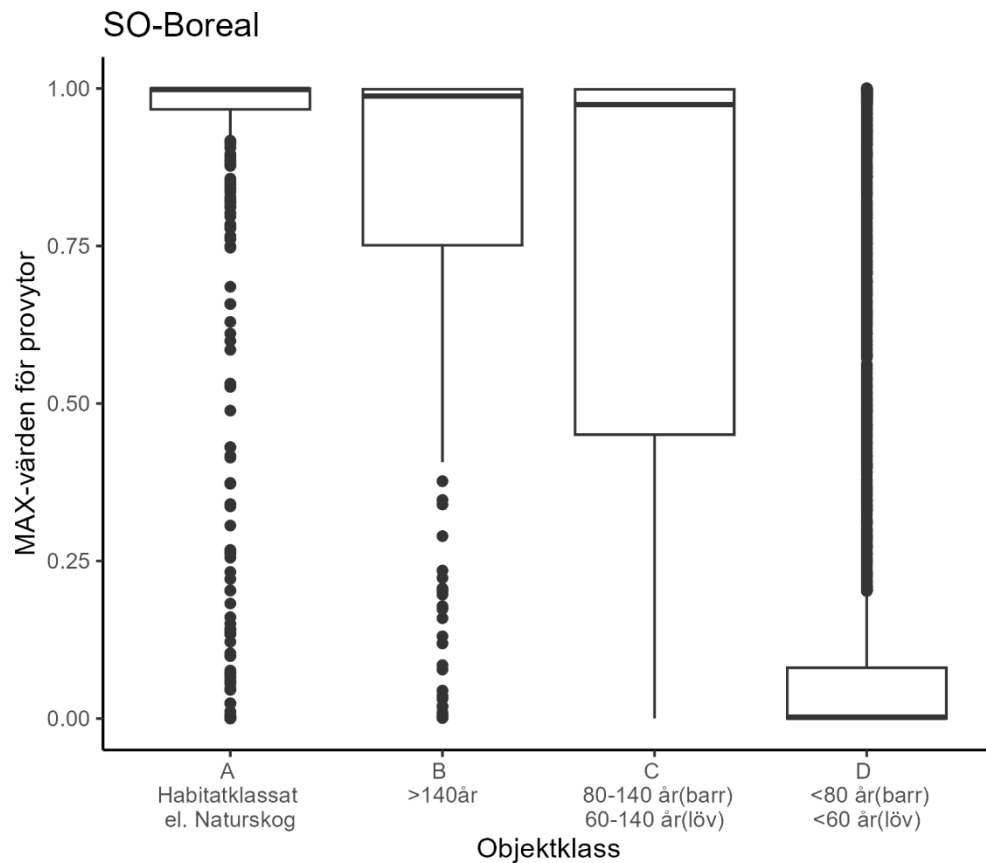


Figur 11: En boxplot som illustrerar fördelningen av högsta indikationsvärde inom NV-Boreal region

I boxplot ovan illustreras fördelningen av högsta indikationsvärde inom provytor för alla objektklasser för alla provytor inom NO-Boreal region. Grov horisontell mittenlinje illustrerar medianvärde, medan boxens övre och undre kant markerar 75% och 25% kvartilen, respektive.



Figur 12: Fördelning av provytor med ett högsta pixelvärde som klassats som antingen högt (0,8–1), medium (0,2–0,8), eller lågt (0–0,2) inom SO-Boreal region.



Figur 13: En boxplot som illustrerar fördelningen av högsta indikationsvärde inom SO-Boreal region

I boxplot ovan illustreras fördelningen av högsta indikationsvärde inom provytor för alla objektklasser för alla provytor inom SO-Boreal region. Grov horisontell mittenlinje illustrerar medianvärde, medan boxens övre och undre kant markerar 75% och 25% kvartilen, respektive.

12.5 Utvärdering mot fältinventering

En första fältvalidering av kartunderlagens prediktioner av naturvärden gjordes i kommunerna Umeå, Vännäs, Robertsfors, Nordmaling, Bjurholm och Lycksele i Västerbotten under 2025. Syftet med fältvalideringen var att testa en ny metodik i mindre skala för statistisk utvärdering av överensstämmelsen mellan indikationsvärden för naturvärden i det digitala kunskapsunderlaget med observerade naturvärden i fält. Vi ville fånga upp både skogar som klassats korrekt som höga och låga naturvärden, samt skogar som klassats felaktigt som höga naturvärden (falska positiva), och skogar med höga naturvärden som missats i kartunderlaget (falska negativa) vilket kräver att inventeringen täcker in hela områden.

Inventeringsmetodiken baserades på Skogsstyrelsen metodik för bedömning av hänsynskrävande biotoper, som förenklades och anpassades till en linjeinventering i stället för provyteinventering. Inventeringen genomfördes i 1x1 km rutor längs fyra inventeringslinjer, som var utlagda proportionellt mot det aggregerade indikationsvärdet i respektive ruta i inventeringsområdet. Detta medför att en viktning gjordes mot skogar som hade höga naturvärden i kartunderlaget. Totalt inventerades 60 rutor för att få ett tillräckligt underlag för statistisk utvärdering. En fältapplikation togs fram i Fieldmaps för insamling av data under inventeringen. Inventerarna hade inte tillgång till kartunderlagen de skulle utvärdera eller data om naturvärden på plats

Vid inventeringen följde inventeraren transekterna och registrerade sträckor (antal meter längs linjen) som uppfyllde kriterierna för hänsynskrävande biotoper, dvs förekomst av naturvärde och vilken eller vilka hänsynskrävande biotoper som observerades. Bedömningen av naturvärden gjordes i huvudsak inom en fem meters zon från inventeringslinjen.

Tabell 5: Total sträcka inventerad under fältvalideringen år 2025.

Huvudgrupp	Antal transekter	Total längd(m)	Viktad medellängd(m)
Jordbruk och bete	16	665	55,1
Naturskogskaraktär	364	29 243	79,9
Terrängformationer	163	15 391	83,3
Vattenpräglade	208	11 045	61,4
Värdefulla strukturer	141	11 343	85,5

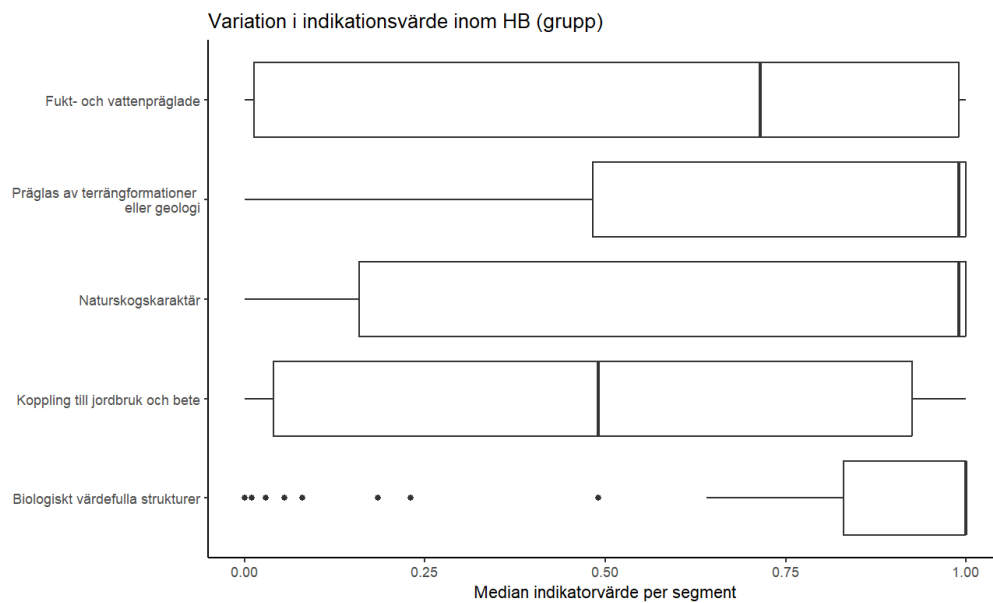
Resultaten från 2025 års fältvalidering sammanfattas i tabell 5. Tabellen redovisar antal observationer, total uppmätt linjelängd samt längdvägt medelvärde av DVIS per huvudgrupp.

Resultaten indikerar en god överensstämmelse mellan Indikationer på skog med naturvärden och fältinventeringen för huvudgrupperna *Värdefulla strukturer*, *Terrängformationer* och *Naturskogskaraktär*, vilka uppvisar höga längdvägda medelvärden ($\geq 0,8$). För *Vattenpräglade* miljöer är de längdvägda medelvärdena lägre, vilket kan tyda på en mer varierande eller svagare överensstämmelse mellan modell och fältdata. Eftersom fördelningen av kartunderlagets indikationsvärden är nära binär, d.v.s. de flesta värden finns mellan 0–0,20 och 0,80–1, så finns mycket få värden i mellanskiktet. Vi kan

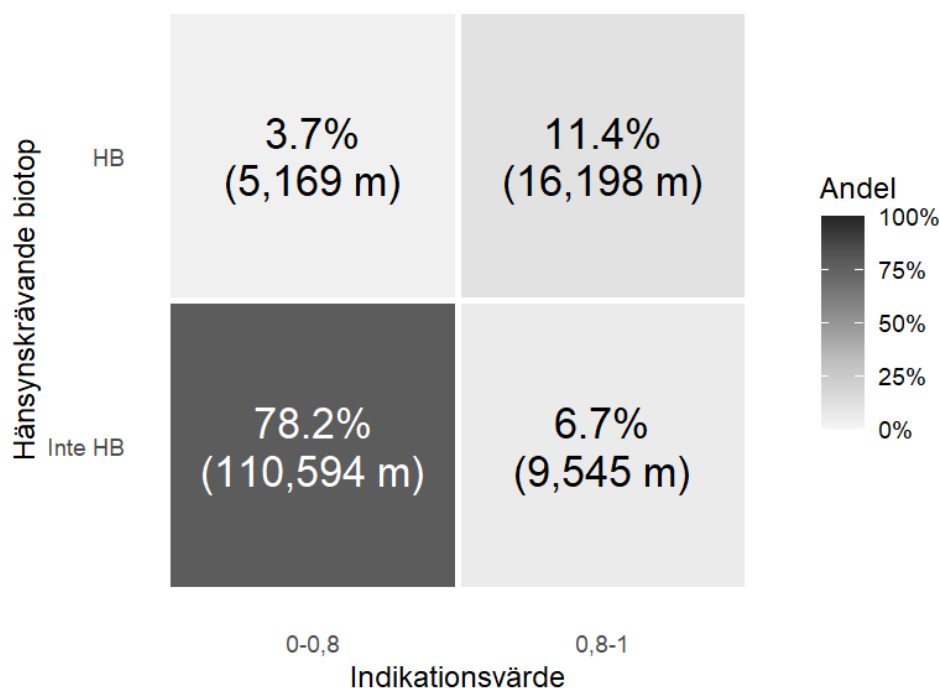
därför anta att indikationsvärden $<0,80$ representerar indikationer på ej höga naturvärden.

För huvudgruppen *Jordbruk och bete* är antalet observationer begränsat, vilket innebär att resultaten bör tolkas med försiktighet och inte möjliggör några säkra slutsatser.

Skillnader i spridning och central tendens mellan huvudgrupperna illustreras även i figuren nedan.



Figur 14: Boxplot som visar variationen i indikationsvärden för olika grupper av hänsynskrävande biotoper



Figur 15: Illustration av hur den uppmätta linjelängden i urvalet fördelar sig mellan kombinationer av hänsynskrävande biotoper

Bild ovan är illustration av hur den uppmätta linjelängden i urvalet fördelar sig mellan kombinationer av hänsynskrävande biotoper (HB) och ej hänsynskrävande biotoper, höga indikationsvärden (0,8–1,0) och låga indikationsvärden (0–0,80). Både färgens intensitet och procenttalet anger andelen av den totala uppmätta linjelängden som faller i respektive ruta, där mörkare nyans innebär större andel. I varje ruta redovisas även den faktiska linjelängden i meter inom parentes.

Varje ruta i Figur 15 motsvarar en specifik kombination av segment som klassats som HB eller inte och indikationsvärde i Indikation på skog med naturvärden. Figuren ger därmed en samlad bild av både relativ fördelning (andel) och absolut omfattning (meter) av de observerade linjesegmenten i urvalet, uppdelat efter huvudgrupp och klassning av Indikation på skog med naturvärden.

Av den totala inventerade sträckan så klassade Indikationer på skog med naturvärden 15,1% som höga naturvärden (indikationsvärde >0,8). Av dessa utgjordes 11,4 procentenheter av hänsynskrävande biotoper med ett högt indikationsvärde, och återstående 3,7 procentenheter utgjordes av hänsynskrävande biotoper som klassats med ett lågt indikationsvärde (falskt negativ). Av de 84,9% av den inventerade sträckan som bedömts som ej hänsynskrävande biotop så klassade Indikationer på skog med naturvärden 78,8 procentenheter med ett lågt indikationsvärde (<0,8), och 6,1 procentenheter som indikation på höga naturvärden (falskt positiv).

Sammantaget innebär detta att 89,6% av den inventerade sträckan klassats korrekt av Indikationer på skog med naturvärden.

12.6 Begreppsmodell

Kommer i version 1.0.

12.7 Informationsmodell

Kommer i version 1.0.

12.8 Informationsutbytesmodell

Kommer i version 1.0.