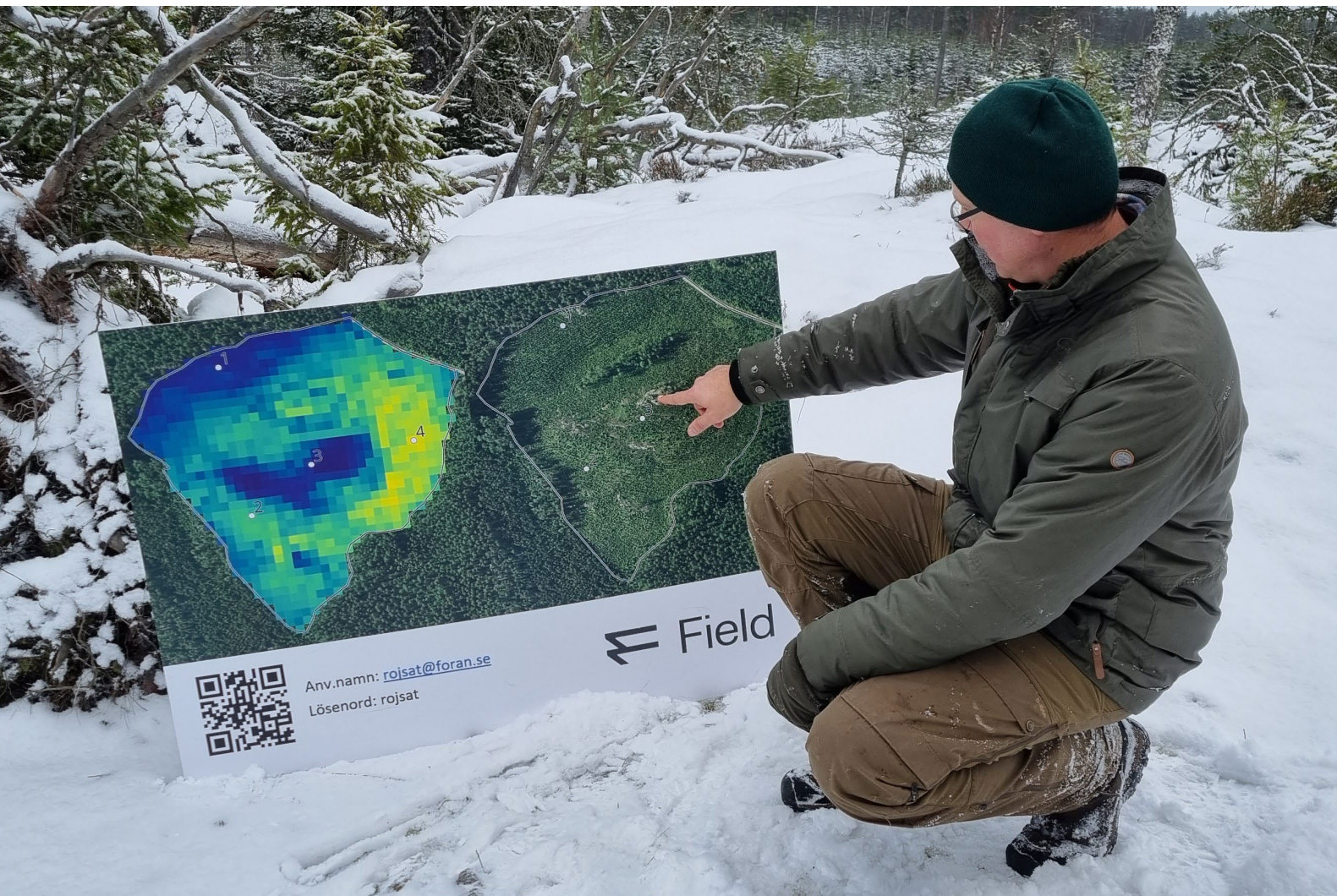


RöjSAT – en AI-modell för prediktion av behov och tid för röjning med fjärranalys

Alva Ringi, Vilma Hägg Edelönn, Kari Hyll, Maria Nordström



Demonstration av RöjSAT-modellen i fält. Foto: Maria Nordström, Skogforsk.

Förord	3
Summary	4
Sammanfattning.....	5
Bakgrund.....	6
Syfte och mål.....	7
Material	7
Referensdata från röjning insamlade i fält	7
Fjärranalysdata	8
Metod	9
Databearbetning.....	9
Modellering.....	10
Ramverk för analys av RøjSAT-modellens olika komponenter	12
Resultat och diskussion	12
Analys av insamlingsmetoden för referensdata från röjning	12
Analys av samband mellan referensdata och fjärranalysinformation	14
Analys av effekten av förbearbetning av data	17
Effekt av modellval	20
Effekten av dataunderlagets storlek på modellens prestanda	22
Sammanfattning av analysen.....	23
Slutsatser	24
Framtida arbete.....	24
Referenser.....	25



Uppsala Science Park, 751 83 Uppsala
skogforsk@skogforsk.se
skogforsk.se

Kvalitetsgranskning (Intern peer review) har genomförts i april 2026 av
Henrik Persson, seniorforskare. Därefter har Magnus Thor, Forskningschef,
granskat och godkänt publikationen för publicering 9 juli 2026.

Redaktör: Caroline Rothpfeffer, caroline.rothpfeffer@skogforsk.se
©Skogforsk 2026 ISSN 1404-305X

Förord

Den här rapporten beskriver en fortsättning på det arbete som inleddes år 2020 med projektet RøjSAT, finansierat av Vinnovas utlysning *Från AI-forskning till innovation våren 2019*. Projektet genomfördes i samarbete med Örebro universitet, Field Sweden, Södra, Mellanskog och Sveaskog och resulterade i en första modell för att skatta röjningsbehov utifrån satellitbilder, fältdata och AI. Arbetet som beskrivs i den här rapporten har genomförts inom ramarna för forskningsprogrammet Mistra Digital Forest. I fortsättningsprojektet har vi haft fokus på att fördjupa arbetet med dataflöden och modellering samt att öka förståelsen för modellens prestanda och användbarhet som underlag för vidare utveckling.

Vi vill rikta ett stort tack till Mistra Digital Forest, vars finansiella stöd har gjort det möjligt att arbeta vidare med projektet och fördjupa kunskapen kring hur metoden kan utvecklas och tillämpas. Vi vill även tacka alla er som bidragit med värdefulla inspel under arbetets gång.

Uppsala, augusti 2026

Författarna

Summary

This work is a continuation of the RøjSAT project, in which a model for predicting the time required to conduct pre-commercial thinning based on openly available remote sensing data has been developed. This phase of the project was focused on identifying limitations in the existing data and methodology, testing various preprocessing strategies, and evaluating the performance of different models given the available data. The results show that the models' performance is primarily limited by the data quality rather than by the choice of model type or complexity. Correlation analyses and learning curves suggest that simply adding more data of the same type will not solve the problem — the critical factor is the informational content of the data.

Through systematic evaluations of various preprocessing steps, such as outlier filtering and handling of edge effects, data quality has been improved, providing a more realistic assessment of the extent to which the models are generalized. Comparisons across different model types indicate that performance is similar, meaning that advanced models do not offer any clear advantage with the current dataset. These findings were also discussed during a workshop with industry stakeholders, where the importance of clearly defining the target variable and including data that reflects a range of practical situations—such as objects that do not require pre-commercial thinning — was highlighted as key areas for development.

Sammanfattning

Detta arbete utgör en fortsättning på RøjSAT-projektet, där en modell för att prediktera röjningstid från öppet tillgängliga fjärranalysdata har utvecklats. Den här delen av projektet har fokuserat på att identifiera begränsningar i befintliga data och metodik, testa olika strategier för förbearbetning och utvärdera hur olika modeller presterar givet tillgängliga data. Resultaten visar att modellernas prestanda i första hand begränsas av datamaterialets egenskaper snarare än av val av modelltyp eller komplexitet. Korrelationsanalyser och inlärningskurvor indikerar att mer data av samma typ som tidigare inte löser problemet – det är informationsinnehållet som är avgörande.

Genom systematiska utvärderingar av olika förbearbetningssteg, som filtrering av extremvärden och hantering av kanteffekter, har datakvaliteten förbättrats och gett en mer realistisk bild av hur generaliserbara modellerna är. Jämförelser mellan olika modelltyper visar att prestandan är likartad, vilket innebär att avancerade modeller inte ger någon tydlig fördel med nuvarande datagrund. Dessa insikter diskuterades också under en workshop med branschaktörer, där vikten av att tydligt definiera målvariabeln och inkludera data som speglar olika praktiska situationer – exempelvis objekt utan röjningsbehov – lyftes som centrala utvecklingsområden.

Bakgrund

Röjning är en skogsvårdsåtgärd som innebär avverkning av unga träd utan ekonomiskt uttag av gagnvirke. Syftet är att forma den växande skogen genom val av trädslag samt frigöra ljus, vatten och näring så att de kvarvarande träden får bättre förutsättningar att utvecklas, vilket ger ekonomisk vinst på längre sikt (Skogskunskap 2025). Det finns olika typer av röjning som genomförs vid olika tidpunkter under skogens omloppstid: planröjning, ungskogsröjning och underväxtröjning innan avverkning. Ungskogsröjning är den dominerande typen och utförs vanligtvis när bestånden är cirka 5–15 år gamla och 2–5 meter höga (Skogsstyrelsen 2025). Ungskogsröjning i ett bestånd genomförs en eller flera gånger under en omloppstid, beroende på exempelvis tillväxt och betestryck.

Behovet av röjning i Sverige är stort och eftersatt, vilket gör att effektiva metoder för att identifiera och prioritera behovet är av stor betydelse. Enligt Riksskogstaxeringen finns över en miljon hektar produktiv skogsmark med ett omedelbart behov av röjning i Sverige (Nilsson m.fl. 2025). Idag skattas röjningsbehov främst genom manuella bedömningar, till exempel visuellt från helikopter, vilket både är resurskrävande och inte särskilt exakt. Mer objektiva och skalbara metoder behövs därför för att stödja planering, prissättning och uppföljning av röjningsarbetet.

Forskning om skattning av röjningsbehov med fjärranalys har bedrivits i Sverige sedan 1990-talet. Tidiga studier använde vegetationsindex som NDVI och IPVI för att identifiera lövröjningsbehov, men precisionen begränsades av låg upplösning (20–30 m) och svårigheter att skilja lövträd från markvegetation (Persson & Bank 1998; Dahl 2001; Pettersson 2005). Senare studier kombinerade satellitbilder med laserdata eller drönbaserad fotogrammetri, vilket förbättrade noggrannheten men till höga kostnader som försvårar storskalig tillämpning (Hyvönen 2002; Watt m.fl. 2013; Fransson 2014). Internationellt har nyare studier visat potential för maskininlärning, men en genomgående utmaning är bristen på detaljerade fältdata för träning och validering (Arumäe m.fl. 2022).

Mot denna bakgrund har Skogforsk sedan år 2020 bedrivit projektet RøjSAT tillsammans med skogsföretag och andra forskningsutförare. Projektet syftar till att utveckla metoder för att skatta röjningsbehov med hjälp av öppna fjärranalysdata. I den första projektfasen som finansierades av Vinnova utvecklades en AI-modell som skattar röjningsbehov uttryckt som röjningstid per 10 x 10 m² pixel genom att kombinera GNSS-loggning från röjare med satellitdata från Sentinel-2.

Resultaten från tidigare arbete inom RøjSAT-projektet visade att satellitdata kan fånga mönster kopplade till röjningstid, men också att modellens prestanda begränsas av datakvalitet, arbetsmetodik i fält och hantering av spatialt beroende data (Hyll m.fl. 2024). I den här arbetsrapporten presenteras en fortsättning som finansierats av forskningsprogrammet Mistra Digital Forest med det övergripande målet att fördjupa analysen av det befintliga datamaterialet och metodiken, för att bättre förstå vilka faktorer som begränsar modellens generaliserbarhet och vilka delar av analyskedjan som har störst påverkan på modellens prestanda.

Syfte och mål

Syftet med arbetet var att undersöka hur olika faktorer kopplade till datainsamling, databearbetning och modellering påverkar RøjSAT-modellens prediktioner. Detta för att skapa förståelse för vilka insatser som krävs för att förbättra framtida prediktioner av röjningsbehov baserat på fjärranalysdata. För att uppnå syftet var målen med arbetet att:

- Identifiera begränsningar i befintliga data och metodik
- Testa olika strategier för förbearbetning av data och variabelkonstruktion inför modellering
- Utvärdera hur olika modeller presterar givet tillgängliga data

Material

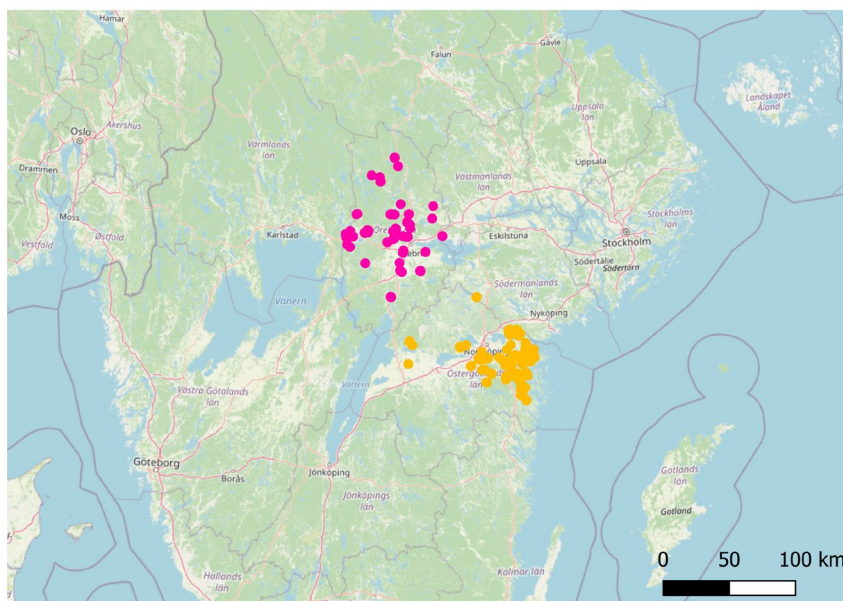
I detta avsnitt beskrivs det datamaterial som ligger till grund för arbetet. Till stor del har samma datamaterial som i projektets tidigare del använts, en mer utförlig beskrivning av materialet återfinns i den tidigare arbetsrapporten *Skattning av röjningsbehov med digitala verktyg* (Hyll m.fl. 2024).

Referensdata från röjning insamlade i fält

Under RøjSAT-projektets tidigare del genomfördes en insamling av referensdata i fält. Datainsamlingen bestod av positioner som samlades in genom att röjare utrustades med en mobiltelefon som registrerade GNSS-punkter med tidsstämplar under röjningsarbetet. Valet av metod för insamling av referensdata baserades på antagandet att röjningstid är en lämplig proxy för att fånga både behovet av röjning och svårighetsgraden hos arbetet.

GNSS-data samlades in från tre olika företag: Södra, Mellanskog och Sveaskog. Endast data från Södra nyttjades i modellutvecklingen under det tidigare arbetet. Att komplettera med data från Mellanskog lyftes som ett lämpligt steg för vidareutveckling, vilket har gjorts i den här uppföljande delen av projektet. I den tidigare delstudien saknades positionering över stora delar av objekten i Sveaskogs data vilket medförde osäkerhet i modelleringen, därför nyttjades inte de objekten (Hyll m.fl. 2024). Inte heller i den här delen av projektet gjordes någon ansats att nyttja dessa data.

Røjningsobjekten från vilka fältdata samlats in filtrerades baserat på om den utförda röjningen bedömdes vara ungskogsrøjning, planrøjning eller underväxtrøjning. Endast ungskogsrøjning var intressant för modellutvecklingen. Inga försök att förändra metodiken för filtreringen gjordes i den här projektdelen, utan de objekt från Södra och Mellanskog som bedömdes vara ungskogsrøjningsobjekt och ha komplett GNSS-data i projektets tidigare del nyttjades även i den här projektdelen. Det material som användes i analyserna i den här delen av projektet består av totalt 185 röjningsobjekt som röjdes under 2020 och 2021, se Figur 1 och Tabell 1.



Figur 1. Röningsobjekt från vilka data användes i studien. De rosa prickarna är objekt som röjts av Mellanskog och de gula prickarna är objekt som röjts av Södra.

Tabell 1. Dataunderlaget beskrivet som antal röningsobjekt respektive antalet pixlar per företag.

Företag	Antal objekt	Antal pixlar
Södra	109	29 094
Mellanskog	76	21 702

Fjärranalysdata

Satellitdata i form av Sentinel-2-bilder över röningsobjekten, tagna i juli 2020, hämtades via Skogsstyrelsens REST-API¹. För merparten av objekten innebar detta att satellitbilderna representerade den senaste sommaren innan rönning utfördes. För ett mindre antal objekt som röjdes under andra halvan av 2021 representerade bilderna istället föregående års sommar. Den uppsättning av våglängdsband som användes var B2 – Blå, B3 – Grön, B4 – Röd, B5 – Red Edge 1, B6 – Red Edge 2, B7 – Red Edge 3, B8A – Red Edge 4, B8 – Near Infrared, B11 – Short Wave Infrared 1, B12 – Short Wave Infrared 2 samt scenklassificeringsbandet SCL (Skogsstyrelsen 2025). Bland de tillgängliga våglängdsbanden från Sentinel-2 finns även B1 och B9, dessa exkluderades eftersom de avser kustnära aerosoler respektive vattenånga och inte bedömdes tillföra relevant information för prediktion av rönningstid.

Endast bilder som uppfyllde kvalitetskraven på molnfrihet och frånvaro av optiska artefakter inkluderades. Antalet bilder från juli 2020 som uppfyllde kvalitetskraven varierade mellan en och sju för de olika röningsobjekten, i snitt fanns cirka fyra användbara bilder över ett objekt. För varje objekt beräknades sedan ett medelvärde av de tillgängliga bilderna så att varje objekt representerades av en uppsättning pixlar med värden för de genomsnittliga förhållandena för perioden.

¹ REST - Skogsstyrelsen

Som komplement till satellitdata användes fyra öppet tillgängliga skogliga rasterprodukter: marktäcke, markfuktighet, trädhöjd och lutning. Dessa hämtades år 2024 i sin senaste tillgängliga version från Skogsstyrelsens REST-API.

Innan analysen omsamlades samtliga dataunderlag till en gemensam pixelstorlek på $10 \times 10 \text{ m}^2$.

Metod

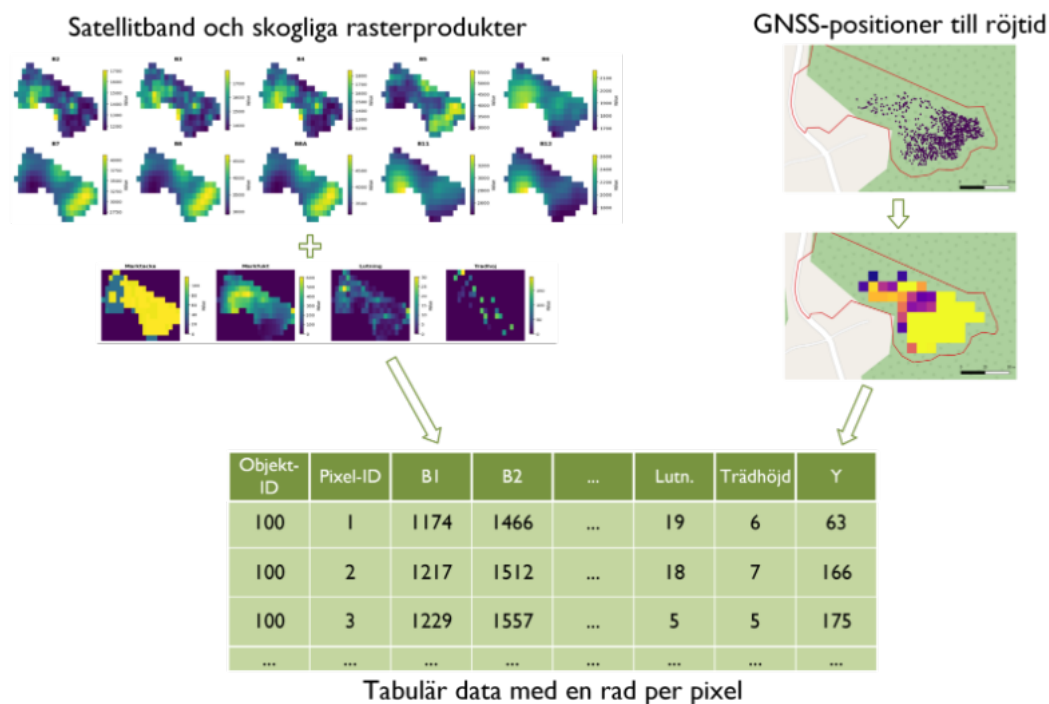
Det här avsnittet beskriver först metoden för bearbetning av datamaterialet och modelleringsprocessen som ligger till grund för studien. Därefter presenteras ett analysramverk som använts för att undersöka hur olika delar av flödet påverkar prediktionerna av röjningstid.

Databearbetningen och modelleringen i denna studie bygger vidare på den analyskedja som etablerades i första delen av RøjSAT-projektet. Där utvecklades ett flöde för att omvandla GNSS-positioner till röjningstid per pixel och kombinera med fjärranalysdata för modellering. Inom ramen för denna analyskedja utvecklades även en modell för prediktion av röjningstid, som i denna studie används som referensmodell.

Databearbetning

Utifrån de insamlade GNSS-positionerna beräknades hur lång tid som röjare uppehållit sig i ett mindre område. Detta gjordes genom att röjningsobjekten, alltså de geografiska indelningar av skog där röjningsarbete planerats att utföras, delades in i ett rutnät med $10 \times 10 \text{ m}^2$ stora pixlar. Genom att interpolera GNSS-positionerna för att återskapa röjarens rörelse över objektet, fördela positionen per sekund jämnt längs sträckan mellan två mätta positioner och beräkna uppehållstiden i sekunder som antalet positioner inom en pixel, erhöles ett mått på röjningstiden per pixel.

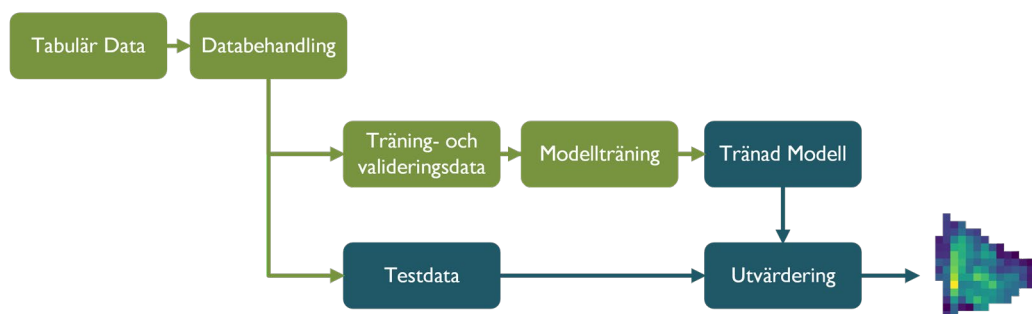
Inför modelleringen kombinerades referensdata i form av röjningstid per pixel och fjärranalysdata per pixel till tabulär form. Varje rad motsvarar en enskild pixel i ett objekt, med ett ID för både objekt och pixel. Kolumnerna utgörs av respektive våglängdsband i satellitdata och skogliga rasterprodukter, samt den härledda röjningstiden. Detta flöde illustreras av Figur 2.



Figur 2. Beskrivning av flödet av information från olika datakällor till tabellform.

Modellering

Modelleringsprocessen som användes beskrivs i Figur 3 och startar med tabulära data på den form som presenteras i Figur 2 ovan. Databehandling genomfördes sedan för att filtrera bort missvisande värden, vilket beskrivs närmare i kommande avsnitt. I detta steg genomfördes också en normalisering av indata för att omvandla variablerna till jämförbara skalor. Efter det gjordes en uppdelning i tränings- och valideringsdata som användes under modellträningen respektive testdata som hålls separat och används för att utvärdera den tränade modellen. Resultatet blir estimerade röjningstider per pixel och kan visualiseras som heatmaps över röjningsobjekten.



Figur 3. Beskrivning av modellerings- och analysflödet.

Den tidigare RøjsAT-modellen användes genomgående som en basmodell och utgör referens för de analyser och jämförelser som beskrivs i nästkommande kapitel. Basmodellen består av ett neuralt nätverk med tre dolda lager med ReLU som aktiveringsfunktion. Det första lagret innehåller 256 neuroner vilka halveras i

efterkommande lager till 128 och 64. Varje dolt lager följs av ett dropout-lager med en bortfallsandel på 0,3, 0,2 och 0,1 för respektive lager. Modellen tränades med Mean Absolute Error, MAE, som förlustfunktion och Adam-optimering.

Kodbasen för datahantering och modellutveckling från projektets tidigare del har successivt anpassats och byggts ut för att möjliggöra de olika typer av analyser och utvärderingar som beskrivs i nästkommande kapitel. I huvudsak har samma Pythonpaket som presenteras i den tidigare arbetsrapporten använts (Hyll m.fl. 2024), till exempel har Tensorflow-Keras fortsatt använts som ramverk för att implementera de olika modeller som har testats.

I den första projektfasen jämfördes flera typer av modeller i syfte att identifiera en fungerande modell för prediktion av rönjningstid, vilket slutligen resulterade i valet av ett neuralt nätverk. Resultatet indikerar att satellitdata i kombination med data över marktäckning, markfuktighet, lutning och trädhöjd via ett neuralt nätverk kan bidra till att skatta rönjningstid, samtidigt som flera begränsningar identifierades (Hyll m.fl. 2024). I denna vidareutvecklande projektdel har motsvarande modeller åter testats. Här har modelleringen i första hand använts som ett verktyg för att bedöma hur robusta olika modelltyper är, och för att avgöra om resultatets begränsningar beror på modellerna och deras utformning eller på dataunderlaget.

Olika modeller med olika komplexitet har testats. En linjär regressionsmodell användes som utgångspunkt för att beskriva den enklaste möjliga relationen mellan de förklarande variablerna och rönjningstid. Ett icke-linjärt alternativ som testades var en Random Forest-modell. Random Forest-modeller består av ett antal beslutsträd som tränas med slumpmässiga utgångsvärden, och vid regression ger varje träd en numerisk prediktion. Modellen tränas på oberoende observationer av träningsdata och resultatet fås genom att kombinera de enskilda trädens prediktioner (Breiman 2001). För modelleringen användes Python-biblioteket Scikit Learns RandomForestRegressor² med 100 beslutsträd, där varje träd som mest får ha tio nivåer och där varje lövnod som minst måste ha fem datapunkter.

Modelleringen har i huvudsak utgått från regression där rönjningstiden har predikterats som kontinuerliga värden. Försök att formulera ett klassificeringsproblem där rönjningstiderna delades in i tre olika klasser har också genomförts. Flera varianter av neurala nätverk utvärderades också med olika lagerdjup, antal neuroner och förlustfunktioner.

För att kunna jämföra olika modeller kvantitativt användes MAE, Mean Absolute Error, ett mått som beskriver en modells medelfel. I den här tillämpningen motsvarar MAE den genomsnittliga absoluta skillnaden, uttryckt i sekunder, mellan predikterad och uppmätt rönjningstid per pixel. MAE beräknas enligt

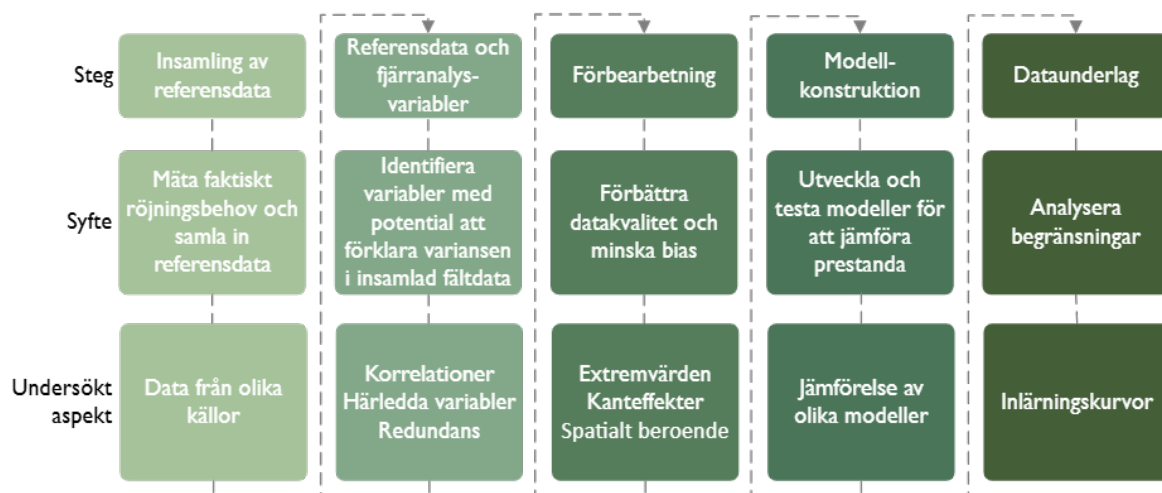
$$MAE = \left[n^{-1} \sum_{i=1}^n |e_i| \right]$$

där n motsvarar antalet observationer (pixlar) och e_i är avvikelsen för observation i , det vill säga skillnaden mellan predikterad och uppmätt rönjningstid för pixel i . (Willmott & Matsuura 2005).

² [RandomForestRegressor — scikit-learn 1.8.0 documentation](#)

Ramverk för analys av RøjSAT-modellens olika komponenter

För att undersöka hur olika faktorer kopplade till datainsamling, databearbetning och modellering påverkar RøjSAT-modellens prediktioner av röjningsbehov togs ett analysramverk fram, vilket beskriver olika områden som har analyserats vid olika steg i dataflödet. Ramverket illustreras av Figur 4 i fem steg, inklusive syftet med varje steg och vilka aspekter som har undersökts.



Figur 4. Analysramverk för vidareutvecklingen av RøjSAT-modellen.

Resultat och diskussion

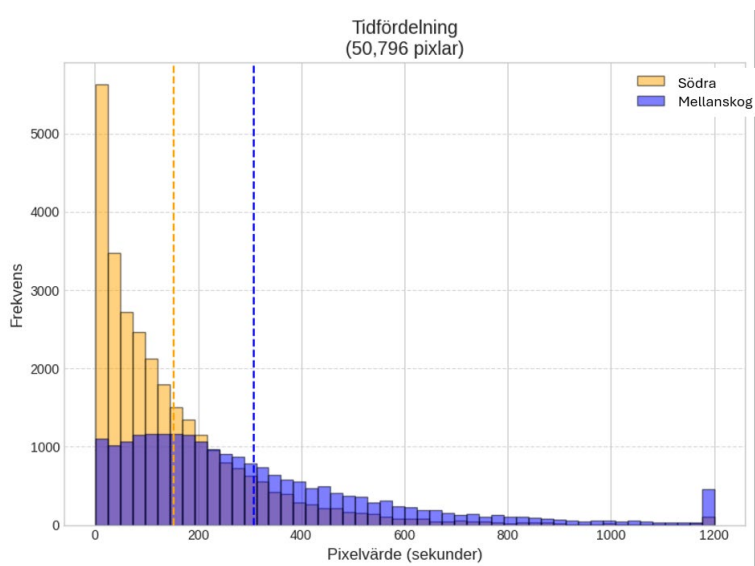
I avsnittet nedan presenteras och diskuteras de analyser som genomförts för vart och ett av de fem stegen i det analysramverk som introducerats ovan, samt de steg som tagits för att vidareutveckla RøjSAT-modellen.

Analys av insamlingsmetoden för referensdata från röjning

De fältdata som användes som referensdata för modelleringen samlades in av Södra och Mellanskog genom deras respektive anlitade röjarlag. Inledningsvis behandlades de två datamängderna separat, och kombinerades därefter i syfte att utöka underlaget för modellträning. Att slå samman de två datamängderna resulterade dock inte i bättre prediktioner. När basmodellen (beskriven i tidigare kapitel) tillämpades på data från respektive företag separat var MAE på testdata 95 sekunder för modellen baserad på data från Södra och 116 sekunder för modellen baserad på data från Mellanskog. När data från båda företagen kombinerades uppgick MAE till 102 sekunder. I alla dessa tre fall filtrerades extremvärden och den reviderade hanteringen av kanteffekter tillämpades, vilka båda beskrivs i efterkommande avsnitt.

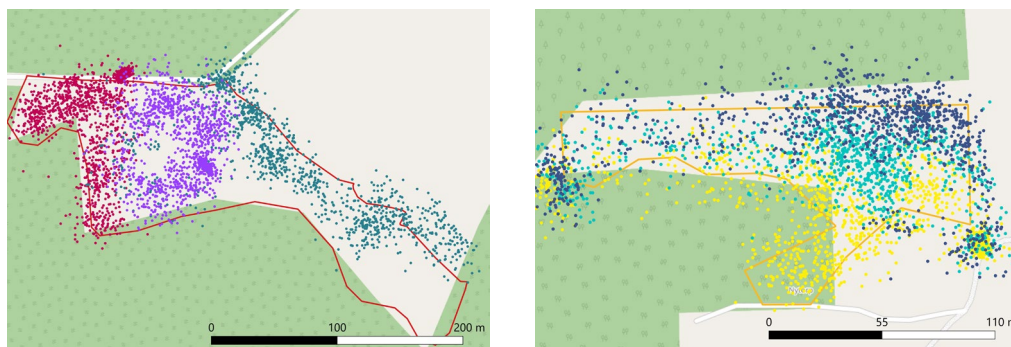
Att sammanslagningen av datamängderna inte resulterade i förbättrad prediktionsförmåga motiverade en närmare analys av skillnader mellan företagens data.

Analysen visade att fördelningen av uppmätta tider skiljde sig tydligt mellan företagen. Data från Mellanskog uppvisade en större andel höga röjningstider och även en högre medelröjningstid per $10 \times 10 \text{ m}^2$ -pixel jämfört med data från Södra, vilket illustreras i Figur 5. Dessa skillnader motiverade en fördjupad granskning av GNSS-data och röjarlagens arbetsmetodik.



Figur 5. Fördelning av uppmätt tid som röjare har tillbringat per $10 \times 10 \text{ m}^2$ -pixel. Gula och blå staplar motsvarar fördelningen för respektive företag, lila färg visar överlappet mellan fördelningarna. Tider över 1 200 sekunder/pixel har trösklats och visas som 1200 sekunder i figuren.

Genom kvalitativ analys av röjarnas rörelsemönster inom objekten kunde det konstateras att de två företagens arbetslag tillämpade olika arbetssätt, vilket illustreras i Figur 6. Analysen möjliggjordes genom att varje mobiltelefon och därmed röjares position gick att urskilja med ett unikt id. Röjarlagen som arbetade för Södra delade i större utsträckning upp sig mellan olika områden inom objektet. Röjarna som arbetade för Mellanskog rörde sig mer över hela objektet och därmed överlappade flera röjare samma områden. Diskussioner med en representant från Mellanskog bekräftade att de instruktioner de skickat med inför röjningsuppdrag motsvarade ett sådant rörelsemönster. Dessa olikheter resulterade i olika mönster av överlappning mellan individuella röjare, olika vistelsetid på olika delar av objektet och olika pixelvis tidfördelning, och påverkade därmed röjningstidens funktion som proxy för röjningsbehov. Försök till att hantera detta gjordes genom transformation av den underliggande fördelningen i respektive dataset för att se om det skulle göra dem mer jämförbara innan sammanslagning, men det resulterade i sämre MAE och användes därför inte.



Figur 6. Registrerade punkter under rökning med olika färg för enskilda rökare inom varje röklag. Till vänster visas ett objekt från Södra, till höger ett från Mellanskog.

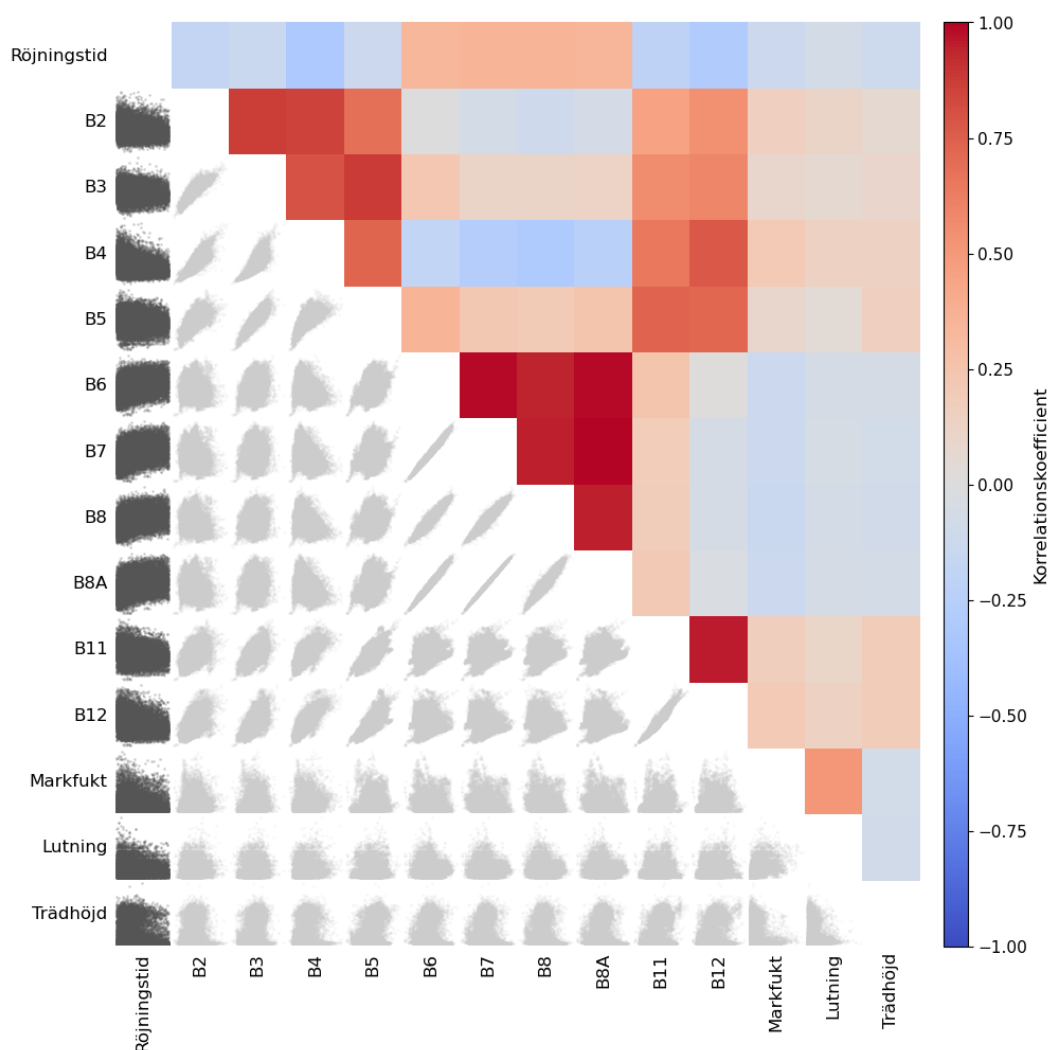
Resultaten belyser en central utmaning vid användning av rökningstid som responsvariabel: tidsåtgången påverkas inte enbart av rökningens behov utan även av hur arbetet organiseras och utförs. När datamängder från olika företag eller röklag kombineras utan kunskap om dessa organisatoriska skillnader riskerar modellen att lära in mönster som speglar olika arbetsmetodik snarare än rökningens behov. I detta projekt, där datamaterialet dessutom var begränsat i omfattning, innebär detta att kombinationen av datakällor sannolikt bidrog med mer brus än användbar variation, vilket även speglas i modellens prediktionsfel.

Samtidigt är det önskvärt att en modell för prediktion av rökningens behov är generaliserbar över företag, regioner och arbetslag, vilket kräver att referensdata är representativa för de flesta förekommande förutsättningar för rökningens arbetet. För att rökningstid ska fungera som en stabil och generaliserbar responsvariabel behöver därför både datainsamling och modellutveckling integrera kunskap om hur arbetsorganisation och utförande påverkar tidsåtgången. Detta kan innebära behov av mer detaljerade metadata om arbetslagens instruktioner och arbetssätt, alternativt fler datakällor som minskar känsligheten för metodskillnader.

Analys av samband mellan referensdata och fjärranalysinformation

Korrelationsanalys

För att analysera sambanden mellan referensdata (skattning av rökningstid baserat på positioner) och tillgänglig fjärranalysinformation undersöktes relationen mellan rökningstid och varje enskild förklarande variabel. Detta gjordes dels visuellt med hjälp av spridningsdiagram, dels kvantitativt genom beräkning av Pearson-korrelation. Resultatet av båda dessa analyser presenteras i Figur 7. Variabeln *marktäcke* utelämnades från analysen då den är kategorisk och inte lämpar sig för spridningsdiagram eller Pearson-korrelation, eftersom värdena representerar klasser snarare än en kontinuerlig skala.



Figur 7. Samband mellan röjningstid och fjärranalysvariabler. I den nedre triangeln visas spridningsdiagram för varje variabelpar. I den övre triangeln visas Pearson-korrelationer, där färgskalan anger styrka och riktning på de linjära sambanden.

Resultaten visar att korrelationen mellan fjärranalysinformationen (våglängdsbanden från Sentinel-2 samt rasterprodukterna markfuktighet, lutning och trädhöjd) och den uppmätta röjningstiden mätt som Pearson-korrelation är låg till måttlig, som högst drygt 0,3 för våglängdsbanden B6-B8A, vilket indikerar att man bör förvänta sig en begränsad förklaringsförmåga vid modellering. Spridningsdiagrammen bekräftar denna bild genom att visa stor variation i röjningstid för givna värden av de enskilda variablerna, utan tydliga icke-linjära mönster. De låga korrelationsvärdena kan därmed inte förklaras av enbart icke-linjära relationer, utan speglar överlag svaga samband mellan fjärranalysdata och röjningstid.

Figuren visar också att flera av de förklarande variablerna är starkt korrelerade med varandra. Det gäller exempelvis för våglängdsbanden B6-B8A. Det innebär att datamaterialet innehåller redundans, modellen får många likartade signaler men lite ny information om variationen i röjningstid. Hur vi försökte hantera detta, genom att reducera antalet variabler, beskrivs i kommande avsnitt.

Konstruktion av härledda variabler (Feature Engineering)

För att renodla informationsinnehållet i satellitdata beräknades olika vegetationsindex från de olika våglängdsbanden. De index som användes var NDVI (Normalized Difference Vegetation Index) som beräknas enligt

$$NDVI = \frac{(NIR - Röd)}{(NIR + Röd)}$$

och SAVI (Soil-Adjusted Vegetation Index) som beräknas enligt

$$SAVI = \frac{(NIR - Röd)}{(NIR + Röd + L)} * (1 + L)$$

där *NIR* står för reflektansen i nära infrarött, *Röd* står för reflektansen i det röda bandet i det synliga spektrumet och *L* är en justeringsfaktor som används för att minska påverkan från markens reflektans (Vélez m.fl. 2023). I Sentinel-2-data valdes våglängdsbandet B8 för NIR då det har samma upplösning som det röda bandet B4. För justeringsfaktorn för markens reflektans, *L*, användes värdet 0,5, vilket är anpassat för medeltät vegetation (Huete 1988). De beräknade vegetationsindexen gav något högre korrelation med röjningstid, men fortfarande på en måttlig nivå (cirka 0,40–0,45). Att inkludera dessa index i modellträningen utöver de ursprungliga variablerna gav inte heller någon tydlig förbättring av modellprestanda. När basmodellen tillämpades på de ursprungliga våglängdsbanden samt de härledda banden NDVI och SAVI landade MAE för testdata på 104,9 att jämföra med 105,4 för basmodellen tränad med enbart de ursprungliga banden.

Hantering av redundans i fjärranalysdata

Den tidigare presenterade korrelationsanalysen visade på hög korrelation mellan olika våglängdsband från Sentinel-2. Detta indikerar att de fångar liknande information, vilket riskerar att försämra modellens effektivitet och generaliseringsförmåga (Jia m.fl. 2022). För att motverka detta gjordes en ansats att minska antalet ingående variabler. Detta gjordes genom två olika tillvägagångssätt som båda beskrivs av Jia m.fl. (2022). Det första följde principen av en algoritm kallad sekventiell bakåtsökning. Utgångspunkten för algoritmen är den fulla uppsättningen variabler som sedan reduceras med en variabel i taget så långt man kan utan att resultatet markant försämras. Valen av variabler att ta bort baserades på de som hade högst korrelation med andra variabler i korrelationsanalysen. Den andra ansatsen gjordes genom PCA – Principal Component Analysis – som reducerar korrelerade variabler till ett mindre antal icke korrelerade komponenter (Jia m.fl. 2022). Modellen tränades sedan med den mindre uppsättningen ingående variabler.

Resultatet blev att modellens prestanda inte blev mycket bättre än med den fulla variabeluppsättningen. Exempelvis kördes modellen med banden B7, B8, B8A och B12 borttagna, det vill säga banden med hög korrelation till andra kvarvarande band enligt korrelationsanalysen i Figur 10. Det resulterade i ett MAE på 104 sekunder. När basmodellen tränades på de tre PCA-komponenterna jämfört med den ursprungliga variabeluppsättningen minskade MAE för testdata bara aningen från 105 till 103 sekunder. Det är någorlunda väntat då de reducerade variablerna inte förändrar den totala informationsmängden, utan bara strukturerar om den befintliga signalen (Jia m.fl. 2022).

Att prestandan inte förändrades nämnvärt trots färre variabler tyder på att en stor del av den ursprungliga variabeluppsättningen innehöll redundans. Det talar för att, med utgångspunkt i de data som finns tillgängliga, använda en mer kompakt variabeluppsättning. Det kan vara fördelaktigt, både ur beräkningssynpunkt och i en framtida operativ kontext där enkelhet och tolkningsbarhet är önskvärda egenskaper (Jia m.fl. 2022).

Analys av effekten av förbearbetning av data

Filtrering av extremvärden

I den tidigare projektfasen noterades en osäkerhet kring pixlar med ovanligt höga röjningstider, och det framhölls att vissa tidsvärden sannolikt inte motsvarade faktisk röjning utan kunde motsvara perioder när röjaren befann sig i ett område utan att aktivt arbeta med röjning. Tanken var att hantera detta med en sensorbaserad indikation på när röjsågen var aktiv och inte. Insamlingen av dessa data begränsades dock av olika faktorer, till exempel en alltför grov medelvärdesbildning av vilka varvtal röjsågen körts på, och sensordata användes därför inte i slutändan, vilket beskrivs i den tidigare arbetsrapporten (Hyll m.fl. 2024). I stället utvecklades en tvåstegsstrategi för att hantera osäkerheter relaterat till höga röjningstider. Först identifierades pixlar där flera röjare hade vistats och där den sammanlagda tiden översteg 600 sekunder (10 minuter). I dessa fall behölls endast tiden för den röjare som hade vistats längst tid i pixeln. Därefter, om röjningstiden fortfarande översteg 600 sekunder efter denna filtrering, sattes värdet till 600 sekunder som en övre gräns.

Under vidareutvecklingen av modellerna aktualiserades problemet med extremvärden på nytt. Vid analys av fördelningen av röjningstider framkom att båda företagen hade en betydande andel pixlar med röjningstider som trösklades till 600 sekunder med ovanstående metod. För att minska osäkerheten togs därför alla pixlar med röjningstid över 600 sekunder bort ur datamaterialet, oavsett om det var en eller flera röjare som rört sig i pixeln. Tanken bakom denna åtgärd var att det inte med säkerhet gick att avgöra om den uppmätta tiden avsåg röjningsarbete eller exempelvis pauser.

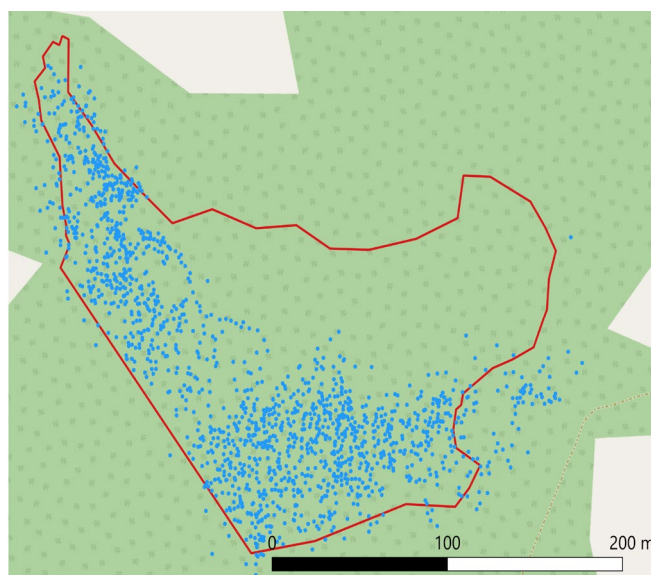
Liknande överväganden gjordes även för mycket låga röjningstider. En mycket låg tid kan indikera att området gått snabbt att röja, men det kan också bero på andra faktorer, till exempel att röjaren endast kortvarigt vistats i en del av pixeln, att pixeln passerats utan röjning eller avvikelser i positioneringen. I den tidigare delen av projektet gjordes ingen filtrering av låga röjningstider, i denna del togs beslutet att filtrera bort pixlar med röjningstider under 10 sekunder med motiveringen att det inte bör vara möjligt att röja ett område på 10 x 10 m² snabbare än så. Efter filtrering av extremvärden minskade det totala antalet pixlar från 50 796 till 43 945.

När basmodellen tillämpades på data där extremvärden inte hade filtrerats bort uppgick MAE för testdata till 119 sekunder. När pixlar med uppmätta röjningstider under 10 sekunder eller över 600 sekunder filtrerades bort blev MAE 98 sekunder.

Hantering av kanteffekter

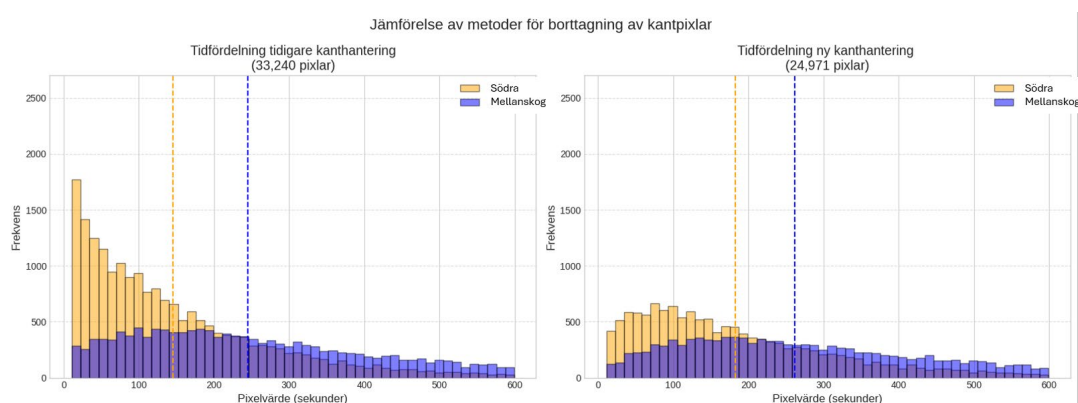
Problematiken med osäkra extremvärden identifierades särskilt i data från Södra, som uppvisade en stor mängd pixlar med mycket låga röjningstider. Ett möjligt fall där pixlar får väldigt låga uppmätta röjningstider är när röjare endast rört sig inom en liten del av pixeln. Det är naturligt att detta sker längs kanterna av röjningsobjekten, då det inte är

säkert att kanten på det område som har röjts linjerar med indelningen av objekten i $10 \times 10 \text{ m}^2$ -pixlar. För att hantera den uppkomna kanteffekten beskars polygonerna som beskriver utbredningen på de planerade röjningsobjekteten med en buffert från objektets yttre kant på tio meter under projektets tidigare del. Vid närmare inspektion av företagens registrerade positioner, framgick det att flera objekt inte hade röjts i sin helhet, utan att det område som faktiskt röjts utgjorde en mindre del av den objektpolygon som låg till grund för analysen. Figur 8 visar ett exempel på ett sådant objekt. Det här mönstret var vanligare för Södra med större andel korta röjningstider än för Mellanskog. Upptäckten om delvis röjda objekt gav insikt i att den tidigare metoden för hantering av kanteffekter inte tar hänsyn till att den yta som objektpolygonen beskriver inte nödvändigtvis motsvarar den yta som faktiskt röjdes.



Figur 8. Exempel på ett röjningsobjekt där hela objektet som beskrivits med objektpolygonen inte röjts.

Utifrån dessa observationer formulerades en hypotes att många av pixlarna med korta röjningstider representerade interiöra kanteffekter, det vill säga pixlar inom objektpolygonen som angränsade till områden där ingen röjning alls hade skett. I stället för att använda en tio meters bred buffert längs objektskanten prövade vi att filtrera bort alla pixlar som angränsar till en pixel utan uppmätt röjningstid, oavsett om dessa pixlar var belägna längs med kanten på en objektpolygon eller inte. Figur 9 visar skillnaden i fördelningen av röjtider med de olika varianterna av kanthantering.



Figur 9. Fördelning av uppmätt tid som röjare har tillbringat per 10 x 10 m²-pixel med olika hantering av kanteffekter. Gula och blå staplar motsvarar fördelningen för respektive företag, lila färg visar överlappet mellan fördelningarna. Till vänster har data inom en 10 meter bred buffert mot objektskanten filtrerats bort från dataunderlaget, till höger har pixlar som ligger intill pixlar utan registrerade positioner med röjare filtrerats bort. I båda fallen har även pixlar med röjningstider under 10 sekunder respektive över 600 sekunder tagits bort.

Den nya hanteringen av kanteffekter minskade antalet pixlar att använda för modellutveckling från 33 240 till 24 971 jämfört med den tidigare hanteringen, men bedömdes nödvändig för att uppnå tillräcklig datakvalitet, då pixlar med osäkra röjningstider annars riskerar att försämra modellens robusthet och prediktionsförmåga. När basmodellen tillämpades på materialet med tidigare kanthantering blev MAE på testdata 99 sekunder, och med den nya kanthanteringen 105 sekunder.

Förbearbetningens effekt på datamaterialets storlek

Filtreringen av pixlar med orimligt korta och långa röjningstider tillsammans med den reviderade hanteringen av kanteffekter minskade antalet pixlar tillgängliga för modellutveckling till ungefär hälften av den ursprungliga mängden. Åtgärderna bedömdes dock vara nödvändiga steg för att reducera osäkerheter i röjningstiden som responsvariabel. Figur 10 sammanfattar hur datamängden successivt reducerades genom de olika stegen. I slutändan användes 24 971 pixlar i de vidare analyserna av samband mellan referensdata och fjärranalysinformation samt i modellutvecklingen.

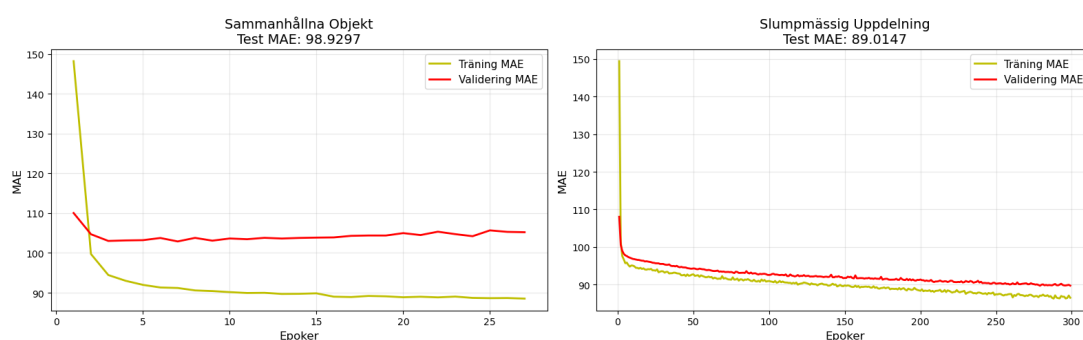


Figur 10. Successiv reduktion av antalet pixlar i datamaterialet med uppmätt röjningstid till följd av filtrering av extremvärden och hantering av kanteffekter.

Sammanhållning av pixlar från samma röjningsobjekt vid uppdelning för träning, validering och test

Basmodellens generella utformning användes, som tidigare beskrivits, som utgångspunkt för den vidare utvecklingen. Den tränades på pixelnivå och predikterar röjningstid för enskilda pixlar, utan hänsyn till data från omgivande pixlar. Pixlar som ligger nära varandra geografiskt har dock ett spatialt beroende, och därmed större sannolikhet att ha liknande värden, både i fjärranalysdata och uppmätt röjningstid, på grund av att de beskriver skog i samma område. Detta är en känd utmaning vid arbete med fjärranalysdata (Karasiak m.fl. 2024). För att minska risken för sådant läckage togs beslutet att pixlar som tillhör samma röjningsobjekt skulle hållas samman vid uppdelning i tränings-, validerings- och testdata, så att samtliga pixlar från samma röjningsobjekt hamnade i ett och samma av de tre dataseten. Slumpen användes fortsatt för att bestämma vilket röjningsobjekt som hamnade i vilket av de tre dataseten.

Figur 11 visar hur MAE för tränings- och valideringsdata utvecklades under träningen av basmodellen, samt MAE på testdata, dels när pixlar från samma röjningsobjekt hålls ihop, dels när de blandas slumpmässigt. När pixlar blandas slumpmässigt (till höger i figuren), blir MAE snarlikt för de tre olika dataseten, i det här fallet ca 90 sekunder. När pixlar från samma objekt istället hålls ihop blir MAE för träningsdata lägre än för validerings- och testdata, vilket är väntat eftersom modellen utvärderas på pixlar som i mindre utsträckning liknar de pixlar modellen har tränats på. Detta ger en mer rättvis bild av hur modellen kan förväntas prestera på nya, osedda data, vid praktisk användning.



Figur 11. Utveckling av MAE för tränings-, validerings- och testdata när pixlar från samma röjningsobjekt hålls ihop (till vänster) och när pixlar blandas slumpmässigt (till höger). Skillnaden i antal epoker beror på att träningen avbryts när validerings-MAE inte längre förbättras, vilket indikerar att modellen har nått sin kapacitet.

Det är dock inte bara pixlar *inom* ett röjningsobjekt som är spatialt beroende. Pixlar från röjningsobjekt som ligger nära varandra har också ett beroende. Det är därför sannolikt att åtgärden att hålla ihop pixlar från samma objekt inte är tillräcklig för att helt undvika läckage av information mellan tränings-, validerings- och testdata. En framtida utvecklingsmöjlighet är att undersöka mer avancerade metoder för uppdelning av data. Det kan också vara relevant att undersöka hur den totala geografiska spridningen på insamlade fältdata påverkar vart i landet, och för vilken typ av skog, en utvecklad modell är lämplig att använda.

Effekt av modellval

Ett antal olika varianter av modeller testades med samstämmigt utfall, varav ett urval presenteras i Tabell 2. Den neurala nätverksmodellen är den basmodell som beskrivits tidigare i rapporten. Alla modeller som presenteras i tabellen tränades på båda företagens data. Enligt tidigare motiveringar filtrerades pixlar med röjningstider under 10 sekunder eller över 600 sekunder bort och den omarbetade metoden för hantering av kanteffekter användes. Vidare hölls pixlar från samma röjningsobjekt ihop vid uppdelning i olika dataset för träning, validering och test.

Tabell 2: Jämförelse mellan olika modelltyper.

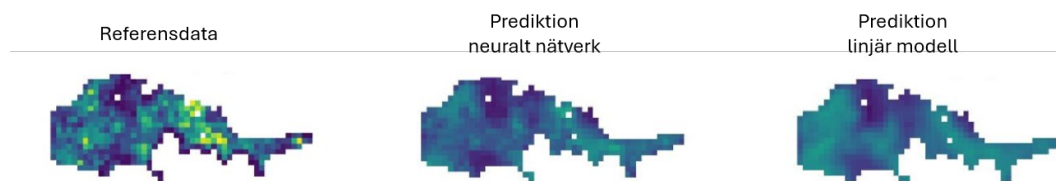
Modell	Träning MAE (s)	Validering MAE (s)	Test MAE (s)
Linjär Regression	-	-	98
Random Forest Originalband	-	-	105

Neuralt nätverk Originalband	94	102	105
---------------------------------	----	-----	-----

Jämförelse av MAE-värdena på testdata för de olika modellerna visar på små skillnader, där Random Forest och neurala nätverk till och med innebar en marginell försämring av MAE jämfört med den linjära modellen. För alla de olika modellerna var MAE på testdata omkring 100 sekunder, vilket innebär att prediktionen i genomsnitt avviker med 100 sekunder från den uppmätta röjningstiden för en enskild pixel. Det kan jämföras med den genomsnittliga uppmätta röjningstiden som var 226 sekunder per pixel.

En generell utmaning oavsett vilken modell som användes var att modellerna gärna predikterar värden nära medelvärdet på den uppmätta röjningstiden i träningsdata. Detta är ett välkänt mönster inom maskininlärningsbaserad regression, små värden tenderar att överskattas och stora värden att underskattas (Lee & Chen 2025). Försök att motverka detta gjordes genom att testa olika förlustfunktioner och genom att skala utdata, men effekten blev begränsad.

För att visualisera resultatet av modellerna, konstruerades heatmaps över röjningsobjekten för såväl referensdata som de olika modellernas predikterade röjningstider. Figur 12 visar ett exempel på heatmaps för ett röjningsobjekt. Figuren tydliggör problematiken med att modellerna, både det neurala nätverket och den linjära modellen, har svårt att fånga den fulla variationen i röjningstider, prediktionerna tenderar att ligga nära medelvärdet. Samtidigt tydliggör figuren att modellen fångar mönster av lägre och högre röjningstider väl. Detta stämmer överens med slutsatsen från tidigare projektdel där modellen också testades i fält – basmodellen fångar variationer inom objekt väl. Detta visar på metodens potential, samtidigt tydliggjordes ett behov av att definiera vilka krav som ställs på en modell för att den ska vara praktiskt användbar, vilket diskuteras närmare under avsnittet *Framtida arbete*.



Figur 12. Jämförelse mellan uppmätt röjningstid (referensdata) och predikterad röjningstid för ett exempelobjekt, där prediktioner från ett neuralt nätverk och en linjär modell visas. Röjtiden illustreras i en skala för tiden från mörkblått (lägst) till ljus (högst).

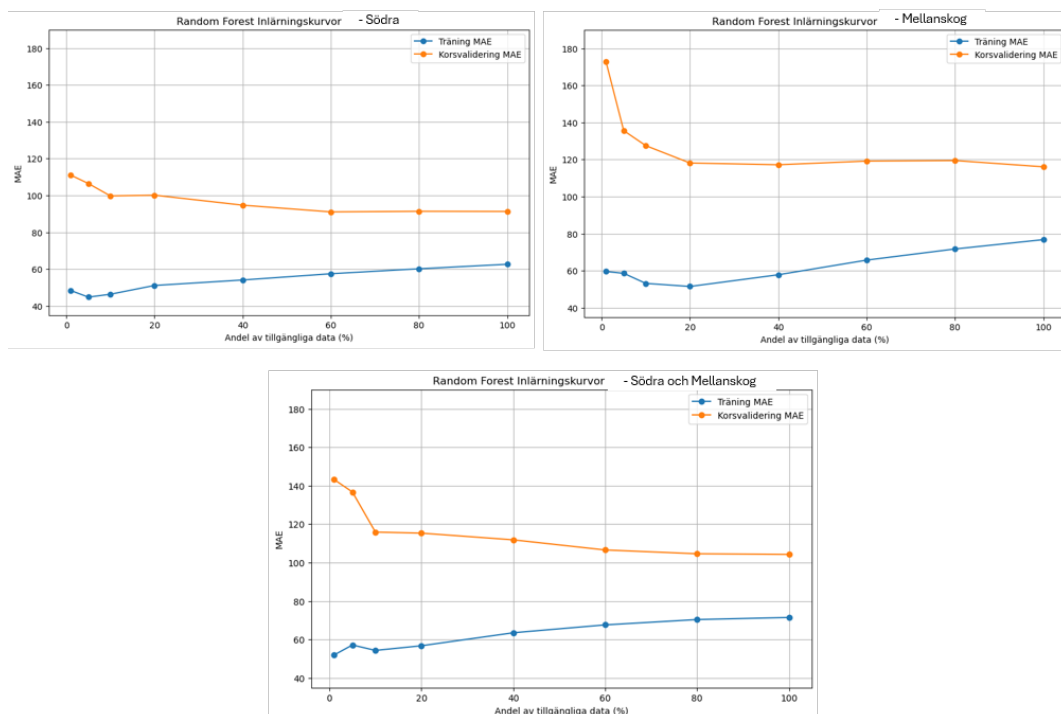
Avslutningsvis bör det noteras att flera av de förbearbetningssteg som införts i denna projektfas – såsom en striktare hantering av kanteffekter, borttagning av pixlar med osäkra extremvärden och en mer konsekvent avgränsning av de faktiska arbetsytorna ledde till ett mindre men mer tillförlitligt datamaterial. I den tidigare projektfasen bidrog vissa av de pixlar som nu har filtrerats bort till en skenbart högre modellprestanda, eftersom de innehöll mönster som modellerna kunde anpassa sig till men som inte speglade faktiska skillnader i röjningsbehov. När detta brus tas bort kvarstår ett renare underlag, men också ett där modellen har färre felaktiga eller överdrivna strukturer att lära sig från. De något sämre MAE-värdena i denna fas bör därför inte nödvändigtvis ses som en försämring, utan som en konsekvens av att prediktionerna nu utvärderas mot ett mer korrekt datamaterial.

Effekten av dataunderlagets storlek på modellens prestanda

För att bedöma om en större mängd referensdata i form av röjningstider skattade från röjarens position skulle kunna förbättra modellens prestanda analyserades inlärningskurvor som visar hur en modells fel, uttryckt som MAE, förändras när modellen tränas på successivt större delar av tillgängliga data. Analysen baserades på en Random Forest-modell då dessa modeller är snabbare att träna än neurala nätverk och tidigare analyser visat att olika modeller presterar ungefär lika bra.

Korsvalidering med fem grupper användes, där alla pixlar från samma röjningsobjekt alltid hamnar i samma grupp för att undvika informationsläckage. Modellen tränades upprepade gånger på ökande andelar, från 10 till 100% av tillgängliga data, och för varje andel beräknades felet för både tränings- och korsvalideringsdata. Andelarna avser här den proportion av den maximalt användbara träningsmängden inom respektive uppdelning för korsvalideringen. På grund av korsvalideringens konstruktion användes alltid en av de fem grupperna som validering, och träningsmängden omfattar således en delmängd av det totala materialet.

Inlärningskurvorna, som visas i Figur 13, tydliggör ett mönster som är typiskt för denna typ av analys och är väl dokumenterat (Mohr & van Rijn 2024). Felet för träningsdata ökar i början samtidigt som felet för valideringsdata minskar. När modellen tränas på ett litet dataset kan den i princip memorera detaljerna i träningsdata, vilket resulterar i ett lågt fel på träningsdata men högt fel på valideringsdata på grund av att modellen är överanpassad. När fler datapunkter tillkommer måste modellen hantera större variation och generalisera bättre. Detta leder till ett högre fel på träningsdata men ett lägre fel på valideringsdata, tills båda kurvorna gradvis stabiliseras.



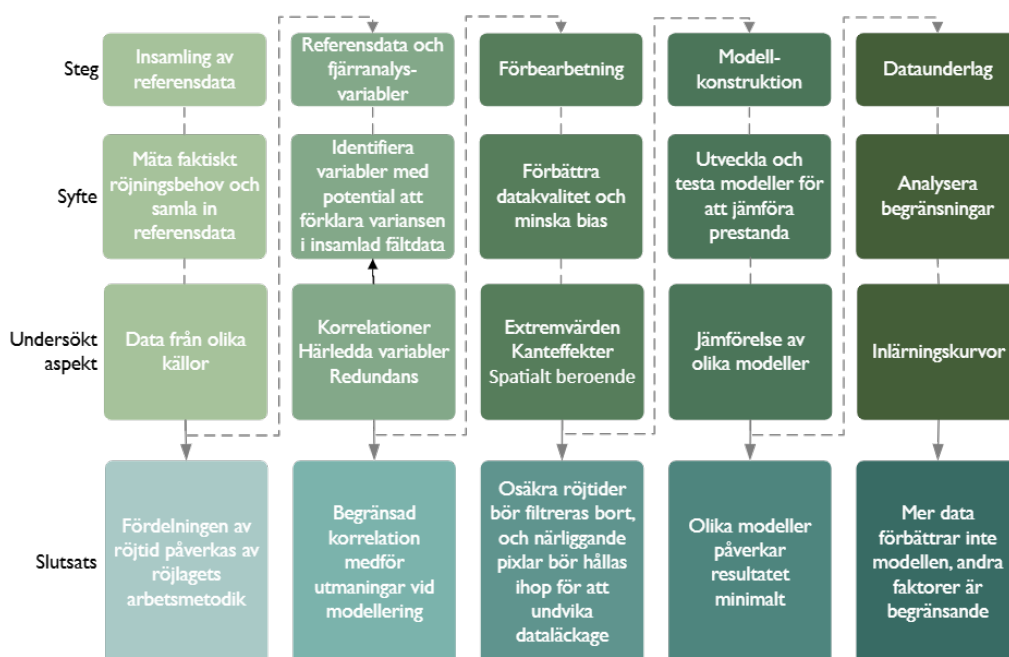
Figur 13. Inlärningskurvor för Random Forest-modellen tränad på datasetet från Södra (överst till vänster), Mellanskog (överst till höger) respektive för en kombination av de båda dataseten (nederst) som visar prediktionsfelet (MAE) för tränings- (blå linje) och korsvalideringsdata (röd linje) när olika andel av tillgängliga data använts för modellträning.

I detta fall planar kurvorna ut redan när cirka 60% av de tillgängliga tränings- och valideringsdata används, vilket indikerar att modellen relativt snabbt når sin kapacitet givet informationsinnehållet i datamaterialet. Denna punkt kallas mättnadspunkt, och efter denna ger inte ytterligare data någon meningsfull förbättring (Mohr & van Rijn, 2024). Resultatet tyder på att modellens begränsningar främst är kopplade till bristande informationsinnehåll i data, snarare än till datamängden i sig. För att förbättra prediktionernas noggrannhet krävs därför sannolikt insatser som stärker det insamlade materialets relevans. Exempel på sådana åtgärder är insamling av ytterligare fältvariabler, såsom stamtäthet eller fältmätt trädhöjd, eller kompletterande information om när röjsågar faktiskt används, vilket skulle kunna förtydliga skillnaden mellan aktivt röjningsarbete och övrig vistelse på objekten.

Som beskrivet under avsnittet om korrelationsanalys är sambanden mellan de förklarande variablerna (både satellitband, data från kompletterande rasterprodukter och härledda vegetationsindex) och responsvariabeln röjningstid begränsad, samtidigt som korrelationen mellan de förklarande variablerna var hög. Det stöder också samma slutsats, modellens prestanda begränsas av informationsinnehållet i data snarare än av mängden data.

Sammanfattning av analysen

Figur 14 sammanfattar den analys som genomförts baserat på basmodellen för det första RøjSAT-projektet och utifrån det ramverk som presenterades inledningsvis i kapitlet, kompletterat med de huvudsakliga slutsatserna från respektive analyssteg. Genomlysningen visar att modellernas prestanda i första hand begränsas av referensdata och dess informationsinnehåll, snarare än val av modelltyp eller modellkomplexitet. Förbättringar i datakvalitet genom förbearbetning har gett en mer realistisk bild av modellernas generaliseringsförmåga, medan vidare prestandavinster sannolikt kräver mer informativa eller kompletterande variabler.



Figur 14. Analysramverket för vidareutvecklingen av RøjSAT-modellen, kompletterat med sammanfattande slutsatser för respektive analyssteg.

Slutsatser

Från det arbete som redovisas i rapporten drar vi följande övergripande slutsatser:

- Modellernas prestanda begränsas i hög grad av tillgången på lämpliga referensdata för modellering. Olika modelltyper, som linjär regression, Random Forest och neurala nätverk, ger likartade resultat i termer av MAE baserat på befintliga data så mer data av samma typ kommer sannolikt inte förbättra prestandan nämnvärt. Förbättringar i datakvalitet och tillgång till fler variabler som beskriver exempelvis stamtäthet, trädhöjd eller sensordata som skiljer aktivt arbete från pauser, är därför viktigare än att utveckla mer komplexa modeller.
- Røjningstid påverkas av arbetsmetodik och individuella skillnader mellan røjare, vilket gör den osäker som entydigt mått på røjningsbehov. Skillnader i arbetsmetodik gör det dessutom utmanande att kombinera den här typen av manuellt insamlade data från olika källor.
- Spatial korrelation bör hanteras vid uppdelning av data för träning, validering och test för att undvika informationsläckage och överskattad modellprestanda.
- Även om utvärderingsmättet MAE tyder på en väsentlig avvikelse från referensdata när det gäller prediktion av faktiskt røjningstid så lyckas RøjSAT fånga relevanta skillnader inom ett røjningsobjekt, vilket kan vara värt att bygga vidare på.

Framtida arbete

Uti från de möjligheter och begränsningar som identifierades under utvecklingsarbetet initierades en workshop med aktörer i branschen för att diskutera hur arbetet med att skatta røjningsbehov med fjärranalys kan utvecklas vidare. En central slutsats från workshopen var behovet av att tydligare definiera *vad* som ska skattas med modellen. Det konstaterades att røjningstid inte nödvändigtvis är en entydig proxy för røjningsbehov, eftersom tidsåtgången även påverkas av exempelvis arbetsmetodik och terrängförhållanden. För framtida arbete betonades därför vikten av att målvariabeln definieras tydligt och att referensdata sedan samlas in med utgångspunkt i det fenomen som modellen avser att beskriva.

Vidare framkom att vilken typ av information om røjningsbehov som är mest relevant varierar mellan olika aktörer. För större skogsägande bolag lyftes möjligheten att använda fjärranalysbaserade metoder för att prioritera mellan objekt som särskilt värdefullt, exempelvis för att sortera bort bestånd utan røjningsbehov och därmed minska risken att skicka røjlag till objekt för att på plats i fält upptäcka att røjningsbehov saknas. I detta

sammanhang diskuterades behovet av att inkludera referensdata från objekt utan röjningsbehov, eller tydligare information om varför hela objekten i vissa fall inte har röjts. För mindre markägare och skogsägarföreningar bedömdes denna typ av storskalig prioritering mindre central, medan mer objektsnära information, som skattning av röjningskostnad, ansågs vara mer relevant.

En annan diskussionspunkt rörde möjligheten att utveckla en gemensam modell för att prediktera röjningsbehov eller röjningsarbete i olika skogstyper samt för både första- och andraröjningar. I diskussionen framhölls att första- och andraröjning delvis styrs av olika faktorer, där behovet av förstaröjning i större utsträckning påverkas av stamtäthet och lövuppslag, medan andraröjning är mer kopplad till trädhöjd då det ofta är önskvärt att vänta med andraröjning tills beståndet nått en höjd där risken för betesskador från älg minskat. Sammantaget talar detta för att olika typer av ungsogsröjning kan behöva hanteras separat, både vid datainsamling och modellering.

Vidare diskuterades möjligheten att skilja aktivt röjningsarbete från annan vistelse på röjningsobjekten, till exempel genom att samla in data från röjsågar eller smarta klockor. Samtidigt poängterades att framtida datainsamling inte får innebära alltför många extra moment för röjarna, eftersom det riskerar att påverka både datakvalitet och genomförbarhet negativt.

Utöver de frågor som diskuterades under workshopen har även andra utvecklingsmöjligheter identifierats under projektets gång. Även om arbetet i det här projektet pekar på att datakvalitet är den främsta begränsningen, kan nya datakällor öppna för andra typer av modeller. En möjlighet skulle kunna vara att arbeta mer direkt med bildanalys, exempelvis genom att använda konvolutionella neurala nätverk istället för att transformera satellitdata till tabulär form. Vidare lyftes möjligheten att använda satellitbilder från flera tidpunkter före röjning, exempelvis genom förändringsanalys över flera år, i syfte att bättre fånga tillväxtmönster som kan vara relevanta för att bedöma framtida röjningsbehov.

Referenser

- Arumäe, T., Lang, M., Sims, A. & Laarmann, D. 2022. Planning of commercial thinnings using machine learning and airborne lidar data. *Forests*, 13(2), 206 s.
- Breiman, L. 2001. Random forests. *Machine learning*, 45(1), 5-32.
- Dahl, M. 2001. Satellitbildsbaserade skattningar av skogsområden med röjningsbehov. Master thesis, SLU Sveriges Lantbruksuniversitet, 35 s.
- Fransson, T. 2014. Användning av drönare vid skogsvårdsplanering. BSc thesis, Linnéuniversitetet, 1-49 s.
- Huete, A. R. 1988. A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, 25(3), 295-309.
- Hyll, K., Ene, L., Hedberg, R., Sidén, F., Längkvist, M., Rahaman, A., Willén, E. & Nordström, M. 2024. Skattning av röjningsbehov med digitala verktyg. Skogforsk arbetsrapport [1201-2024-skattning-av-rojningsbehov-med-digitala-verktyg.pdf](#).

- Hyvönen, P. 2002. Kuvioittaisten puustotunnusten ja toimenpide-ehdotusten estimointi klähhimmän naapurin menetelmällä Landsat TM-satelliittikuvan, vanhaninventointitiedon ja kuviotason tukiaineiston avulla. *Metsätieteen aikakauskirja*, 3, 363–379.
- Jia, W., Sun, M., Lian, J., & Hou, S. 2022. Feature dimensionality reduction: a review. *Complex & Intelligent Systems*, 8(3), 2663–2693.
- Karasiak, N., Dejoux, J. F., Monteil, C., & Sheeren, D. 2022. Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*, 111(7), 2715–2740.
- Lee, H. & Chen, S. 2025. Systematic bias of machine learning regression models and correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Mohr, F. & van Rijn, J. N. 2024. Learning curves for decision making in supervised machine learning: a survey. *Machine Learning*, 113(11), 8371–8425.
- Persson, A. & Bank, H. 1998. Användning av satellitdata - hitta avverkad skog och uppskatta lövröjningbehov. Skogsstyrelsen. 42 s.
- Pettersson, M. 2005. Användning av satellitdata för lokalisering av skogsområden där lövröjning bedöms angelägen. Master thesis, SLU Sveriges Lantbruksuniversitet, 33 s.
- Nilsson, P., Markström, M. & Fridman, J. 2025. *Skogsdata 2025: Aktuella uppgifter om de svenska skogarna från SLU Riksskogstaxeringen: Tema: Tillväxt, avverkning och naturlig avgång*. SLU Institutionen för skoglig resurshushållning.
<https://res.slu.se/id/publ/143434>
- Skogskunskap. 2025. Varför röja? <https://www.skogskunskap.se/skota-skog/roja/rojningens-grunder/varfor-roja/> [2025-12-11].
- Skogsstyrelsen. 2025. Röjning. <https://www.skogsstyrelsen.se/bruka-skog/rojning/> [2025-12-10].
- Skogsstyrelsen. 2020. Sentinel2 – teknisk beskrivning. Dokumentversion 1.0. Skogsstyrelsen. Publicerad 2020-02-27. 1–7 s.
- Vélez, S., Martínez-Peña, R. & Castrillo, D. 2023. Beyond vegetation: A review unveiling additional insights into agriculture and forestry through the application of vegetation indices. *J*, 6(3), 421–436.
- Watt, M. S., Meredith, A., Watt, P. & Gunn, A. 2013. Use of LiDAR to estimate stand characteristics for thinning operations in young Douglas-fir plantations. *New Zealand Journal of Forestry*, 43, 1–10.
- Willmott, C. J. & Matsuura, K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82.