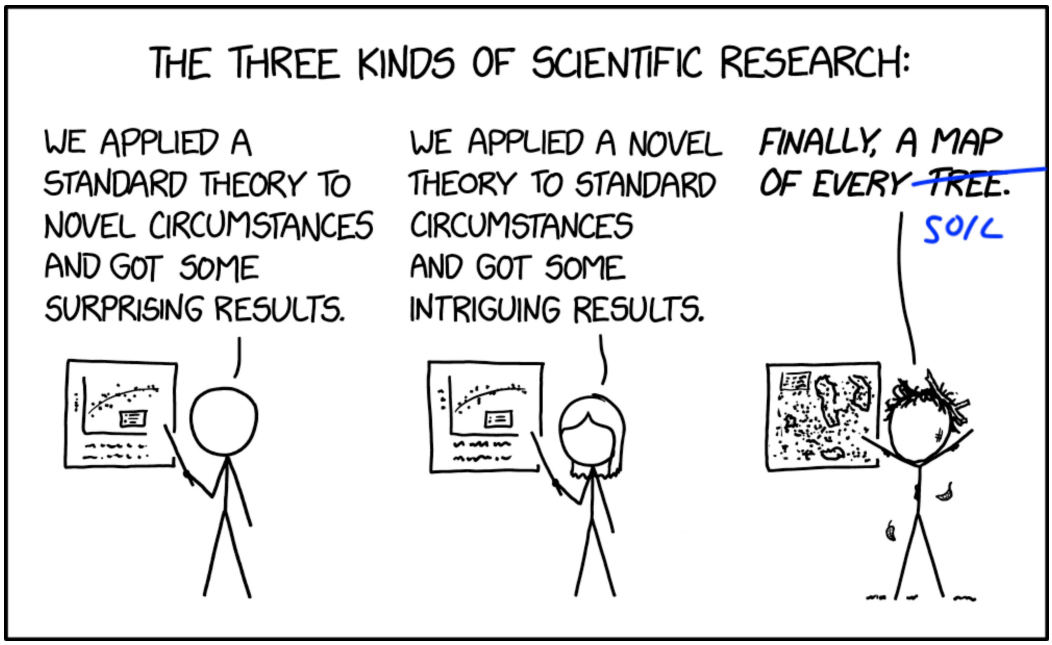
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2026:21
Faculty of Forest Sciences

Mapping surface deposits with machine learning

Yiqi Lin



Mapping surface deposits with machine learning

Yiqi Lin
Faculty of Forest Sciences
Department of Forest Ecology and Management
Umeå

SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS
Umeå 2026

Acta Universitatis Agriculturae Sueciae
2026:21

ISSN 1652-6880
ISBN (print version) 978-91-8124-238-6
ISBN (electronic version) 978-91-8124-268-3
https://doi.org/10.54612/a.5crgaqrdn7

Swedish University of Agricultural Sciences, Department of Forest Ecology and Management, Umeå, Sweden

Print: SLU Grafisk service, Uppsala 2026

Mapping surface deposits with machine learning

Abstract

In Sweden, detailed Quaternary deposit maps cover only about one-third of the country. This thesis examined whether machine learning and deep learning can accelerate surface deposit mapping and improve depth-to-bedrock prediction, while evaluating the opportunities and limitations of a data-driven approach. Surface deposit classification was tested using Extreme Gradient Boosting and multi-view evidential deep learning, while a two-part framework was developed to separate bedrock outcrop classification from continuous depth prediction, providing spatially explicit uncertainty estimates. In addition, historical land use was classified from scanned historical maps. Performance varied considerably by deposit type. Peat and bedrock outcrops were reliably identified and are promising candidates for automation. Till achieved high aggregate performance, but the most confident till predictions were frequently incorrect. Sorted sediments remained beyond the reach of current approaches. Depth-to-bedrock predictions were accurate within 10 m, though the model was less precise in this range. At greater depths, the model increasingly underestimated depth while becoming overconfident in its predictions. A recurring finding was that machine learning and existing map products are best viewed as complementary. Machine learning offers higher spatial resolution and explicit uncertainty estimates, but struggles with deposit classes and depth ranges where the link between surface data and the target is weak. It is therefore better positioned as a tool that supports, rather than replaces, expert mapping. While producing uncertainty estimates is technically feasible, bridging the gap between their production and practical use remains an open challenge.

Keywords: Quaternary deposits, depth-to-bedrock, machine learning, evidential deep learning, uncertainty quantification, LiDAR, Sweden

Kartläggning av jordarter med maskininlärning

Abstrakt

Bara en tredjedel av Sverige täcks idag av detaljerade jordartskartor. Den här avhandlingen undersökte om maskininlärning och djupinlärning kan påskynda jordartskartering och förbättra prediktioner av djupet till berggrunden, samt vilka möjligheter och begränsningar ett datadrivet tillvägagångssätt medför. Klassificering av jordarter testades med Extreme Gradient Boosting och evidentiell djupinlärning baserad på flervysdata. Ett tvådelat ramverk utvecklades för att separera identifiering av berghällar från kontinuerlig djupprediktion, med rumsligt explicita osäkerhetsestimat. Dessutom kartlades även historisk markanvändning från inskannade historiska papperskartor. Resultaten visade en betydande variation i prestanda mellan olika jordarter. Berghällar och torv var lovande kandidater för automatisering med tillförlitliga osäkerhetsestimat. Morän uppnådde hög prestanda, men de mest säkra prediktionerna var ofta felaktiga. Prediktioner av djup till berggrund var träffsäkra ner till 10 meter, även om modellen var mindre precis i detta intervall. Vid större djup underskattade modellen i ökande grad djupet, samtidigt som den blev alltför säker i sina prediktioner. Ett återkommande resultat var att maskininlärning och befintliga kartprodukter bör betraktas som komplement. Maskininlärning erbjuder högre rumslig upplösning och osäkerhetsestimat, men var mindre tillförlitlig där sambandet mellan målvariabeln och de tillgängliga data är svagt, exempelvis vid större djup eller för sorterade sediment. Maskininlärning erbjuder högre rumslig upplösning och explicita osäkerhetsestimat, men brister där sambandet mellan ytdata och målvariabel är svagt. Metoderna bör därför ses som ett verktyg snarare än en ersättning för expertkartering. Även om osäkerhetsestimat är tekniskt genomförbara, kvarstår utmaningen att överbrygga gapet mellan deras framtagning och praktisk användning.

Nyckelord: jordarter, jorddjup, maskininlärning, evidentiell djupinlärning, osäkerhetskvantifiering, LiDAR, Sverige

Contents

List of tables 9

List of figures 11

Abbreviations 13

Introduction 15

1. BACKGROUND 19
   - 1.1 The Quaternary 19
   - 1.2 Quaternary surface deposits vs. soils 21
   - 1.3 Distribution, classification, and mapping of QDs 22

2. RESEARCH AIMS 25

3. TECHNICAL FOUNDATIONS 27
   - 3.1 Data 27
     - 3.1.1 Point observations 27
     - 3.1.2 Environmental covariates 28
   - 3.2 Machine Learning methods 30
     - 3.2.1 Ensemble-based algorithms (Papers I, II, IV) 30
     - 3.2.2 Convolutional Neural Networks (Paper III) 30
   - 3.3 Evaluation and interpretability 32
     - 3.3.1 Validation strategies 32
     - 3.3.2 Feature importance (Papers I, II, IV) 32
   - 3.4 Quantifying and validating uncertainty 33
     - 3.4.1 Quantile Regression Forest (Paper II) 33
     - 3.4.2 Evidential Deep Learning (Paper III) 34

4. MAIN FINDINGS 37
   - 4.1 Surface deposit classification (Papers I & III) 37
   - 4.2 Depth-to-bedrock prediction (Paper II) 40
   - 4.3 Historical land use (Paper IV) 41

5. OVERALL REFLECTIONS 43

5.1 Model evaluation 43
5.1.1 Choice of validation approach 43
5.1.2 Limitations of the reported metrics 44
5.2 The interpretability challenge 46
5.2.1 Spatial features as predictors 46
5.2.2 Model training with legacy maps: benefits and risks 47
5.2.3 Uncovering association-based leakage 49
5.3 Understanding uncertainty 50
5.4 Future work 53

6. CONCLUSION 57

References 59

Popular science summary 69

Populärvetenskaplig sammanfattning 70

Acknowledgements 71

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Lin, Y., Lidberg, W., Karlsson, C., Sohlenius, G., Westphal, F., Larson, J., & Ågren, A. M. (2025). Mapping soil parent materials in a previously glaciated landscape: Potential for a machine learning approach for detailed nationwide mapping. *Geoderma Regional*, 40, e00905. https://doi.org/10.1016/j.geodrs.2024.e00905

II. Lin, Y., Peterson, G., Karlsson, C., Westphal, F., Lidberg, W., & Ågren, A. M. (2026). A two-part framework for depth to bedrock prediction and uncertainty assessment in Sweden. *European Journal of Soil Science*, 77(2), e70308. https://doi.org/10.1111/ejss.70308

III. Lin, Y., Westphal, F., Lidberg, W., Peterson, G., Karlsson, C., Sohlenius, G., & Ågren, A. M. Multi-View Evidential Deep Learning for Surface Deposit Mapping (manuscript)

IV. Ågren, A. M., & Lin, Y. (2024). A fully automated model for land use classification from historical maps using machine learning. *Remote Sensing Applications: Society and Environment*, 36, 101349. https://doi.org/10.1016/j.rsase.2024.101349

My contribution to the papers included in this thesis was as follows:

I. Methodology, Software, Formal Analysis, Writing – original draft, review & editing, Visualization

II. Methodology, Software, Formal Analysis, Writing – original draft, review & editing, Visualization

III. Methodology, Formal Analysis, Writing – original draft, review & editing, Visualization

IV. Software, Writing– review & editing

List of tables

Table 1. Typical glacial and postglacial deposits in Sweden. 22

Table 2. Summary of the inclusion of each spatial layer across the papers and their specific roles. 30

Table 3. Interpretation of evidential fusion outputs for the proposed map-auditing workflow. Combinations of image consensus, image-image conflict, and image-map conflict guide automatic acceptance, revision, or expert review..... 53

List of figures

Figure 1. Multidirectional hillshade (0.5-m LiDAR DEM) of a section of the highest coastline (HC) in northern Sweden. Above the HC, rough, irregular terrain is characteristic of bedrock outcrops and till. Below, the surface shows wave-reworked deposits with visible beach ridges. A lake and an incised channel are visible in the middle of the image. 16

Figure 2. Quaternary deposit (QD) maps cover the entire country, but with varying levels of detail. The central and northern parts of Sweden currently lack detailed (> 1:50,000) QD maps. (a) a QD map at 1:1,000,000, (b) a QD map at 1:50,000, and (c) DEM of the same map area. 17

Figure 3. Schematic illustration of a typical sediment profile in eastern Sweden, showing minimal sediment cover in higher areas and thick layers of fine-grained marine and lake sediments at lower elevations (adapted from Wahlgren et al., 2019). 20

Figure 4. Overview of thesis structure and methodological development.. 25

Figure 5. Overview of a standard Convolutional Neural Network. Input data undergoes feature extraction before being flattened and passed into a dense neural network for multi-class classification. 31

Figure 6. Comparison of standard softmax (left) and Evidential Deep Learning (EDL) (right) outputs. When the model is uncertain (top) or has conflicting evidence (bottom), softmax produces identical probabilities. EDL distinguishes the two: low belief masses indicate insufficient evidence, while high belief masses for competing classes indicate conflicting evidence.... 35

Figure 7. Left: in a tree ensemble, each tree is trained on a random subset of the data, and outputs are combined by majority vote or averaging. Disagreement between trees is smoothed away. Right: In multi-view Evidential Deep Learning, each view is trained independently on different data and produces its own belief distribution. Dempster-Schafer combination rules combine prediction while preserving disagreement as a measurable conflict term (*C*), yielding not only the prediction, but also uncertainty and conflict between views. 36

Figure 8. Schematic illustration of the challenge of point-scale evaluation caused by unit-size mismatch. 45

Figure 9. Legacy map labels can convey contextual information even when they differ from the field label. Left: the legacy QD map assigns a single label to the entire polygon. Right: In reality, the same area contains multiple classes. The map label is ‘wrong’ at that location, but still reflects the characteristics of the overall area, which may help narrow the set of plausible classes. 48

Figure 10. QuadMap representation of depth-to-bedrock (Paper II). Colour indicates predicted depth: green for shallow and brown for deep (0 -14 m). Pixel size encodes uncertainty, with larger pixels representing more uncertain predictions. 52

Figure 11. Agreement and conflict between the multi-view Evidential Deep Learning model’s predictions and field labels. Grey circles indicate consensus. Yellow circles indicate low-conflict errors (model uncertain but incorrect). Red circles indicate high-conflict errors (model confident but incorrect). In the proposed map auditing workflow, a similar mechanism would instead show conflict between image views and the legacy map, computed at every location without requiring field data. 55

Abbreviations

AUSE	Area under sparsification error
CNN	Convolutional neural network
CV	Cross validation
DEM	Digital elevation model
DL	Deep learning
DSM	Digital soil mapping
DST	Dempster-Shafer theory
DTB	Depth to bedrock
EDL	Evidential deep learning
GIS	Geographic information system
HC	Highest coastline
IDW	Inverse distance weighting
ka BP	kilo-annum (thousand years) Before Present
LiDAR	Light detection and ranging
MCC	Matthews correlation coefficient
MDHS	Multi-directional hillshade
ME	Mean error
ML	Machine learning
MMU	Minimum mapping unit
PICP	Prediction interval coverage probability
QD	Quaternary deposit
QCP	Quantile coverage probability
QRF	Quantile regression forest
RF	Random Forest

RMSE	Root mean square error
SASI	Swedish agricultural soil inventory
SFSI	Swedish forest soil inventory
SGU	Geological Survey of Sweden
SHAP	SHapley Additive exPlanations
XGBoost	Extreme Gradient Boosting

Introduction

Formed over the past 2.58 million years during the Quaternary period (Lukas et al. 2017), Quaternary surface deposits are the literal foundation for most human activity in mid-latitude regions. The knowledge of them is essential to how we plan, build, manage, and protect both infrastructure and the environment (Fookes 1997; Clarke 2017; Lukas et al. 2017). Quaternary deposit (QD) maps describe the distribution, stratigraphy, and characteristics of these materials and are widely used by government agencies, engineers, and researchers. Inadequate QD information has real consequences. Quick clay and transitional zones between deposit types are prone to rapid landslides and sinkholes, and increasing extreme weather events have exposed these vulnerabilities through embankment failures and slope instability (Salas-Romero et al. 2016; With et al. 2022; Ochsner et al. 2024). Beyond hazard mitigation, accurate QD information supports peatland restoration, groundwater protection, and optimizes drilling by providing estimates of deposit thickness (Griffiths et al. 2011; Hasan et al. 2025; Laudon et al. 2025). Reliable QD data delivers clear economic, societal, and scientific benefits (Bernknopf et al. 1983).

QD mapping methods have evolved considerably: in the pre-digital era, geologists traversed terrain with spades and probes, sketching deposit boundaries directly onto aerial photographs, which were later scanned into digital formats (Sohlenius et al. 2003). The introduction of Geographic Information Systems (GIS) shifted this work to the desktop, where experts interpreted Digital Elevation Models (DEMs), aerial photography, thematic maps, and satellite imagery to delineate QDs as polygons before field validation (Johnson et al. 2015; Chandler et al. 2018). The availability of airborne laser scanning using Light Detection and Ranging (LiDAR) fundamentally changed QD mapping by providing DEMs at resolutions previously unattainable (Fig. 1). While large-scale landforms such as end-moraine complexes and major eskers were already identifiable on coarser DEMs, LiDAR reveals meso-scale features[1] in unprecedented detail: those too small for most topographic maps, too subtle to detect in the field, and

[1] Glacial lineations, small end moraines, beach ridges, faultlines, etc. (Johnson et al., 2015)

with relief too low to distinguish on aerial photographs (Dowling et al. 2013; Putkinen et al. 2017; Greenwood & Hughes 2022). LiDAR DEMs have consequently become a routine starting point for QD mapping projects across Scandinavia and beyond (Webster et al. 2006; Roering et al. 2013; Johnson et al. 2015).

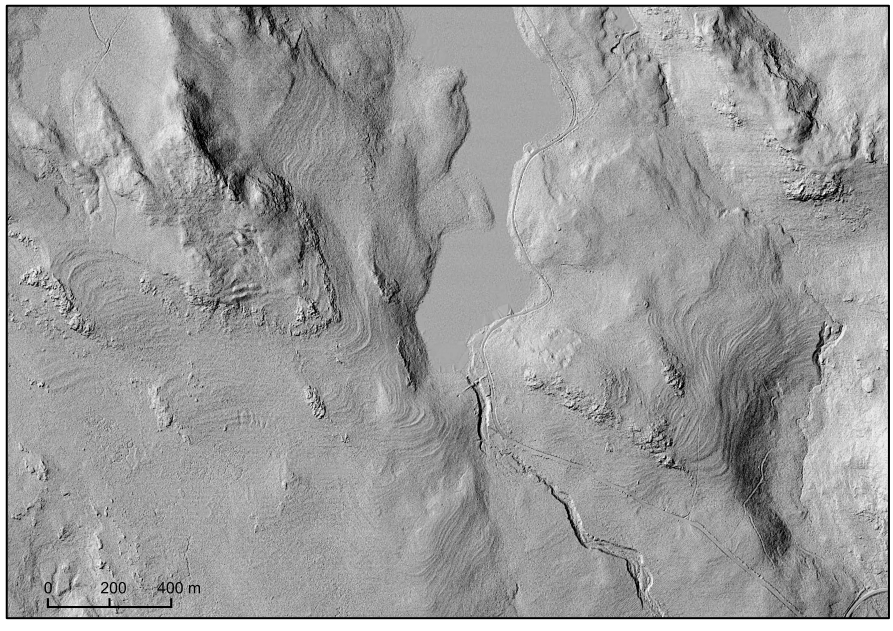


Figure 1. Multidirectional hillshade (0.5-m LiDAR DEM) of a section of the highest coastline (HC) in northern Sweden. Above the HC, rough, irregular terrain is characteristic of bedrock outcrops and till. Below, the surface shows wave-reworked deposits with visible beach ridges. A lake and an incised channel are visible in the middle of the image.

Despite these technological advances, manual delineation of QDs remains slow. National and regional geological surveys struggle to update their maps in a timely manner (Kohonen & Tarvainen 2021; McMartin et al. 2023). The Geological Survey of Sweden (SGU) has been producing QD maps for over 150 years; however, producing a 1:50,000 map for an area of 625 km^2 typically requires 250–350 h of work (Cecilia Karlsson, personal communication). Because the process is labour-intensive, updates are usually carried out as individual commissioned projects driven by localized needs. As a result, two-thirds of the country, especially the northern regions,

still lack detailed map coverage, relying instead on overview maps at scales of 1:100,000 or coarser (Karlsson et al. 2021) (Fig. 2).

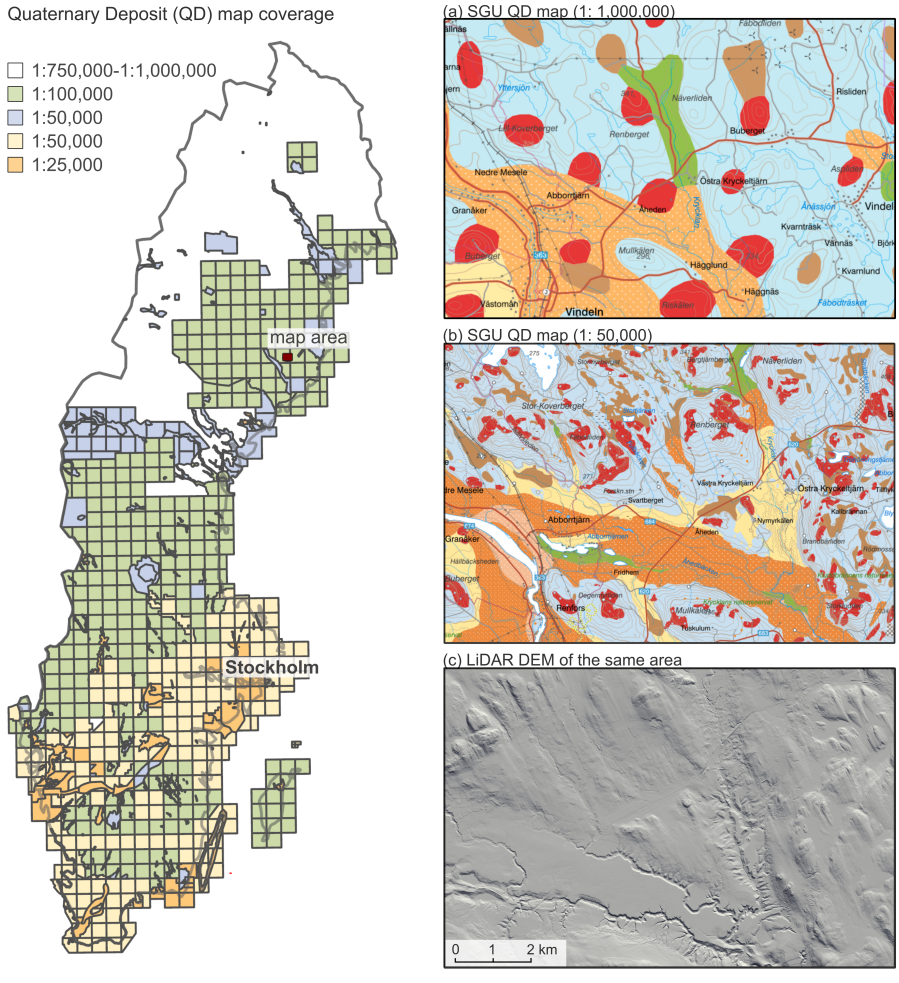


Figure 2. Quaternary deposit (QD) maps cover the entire country, but with varying levels of detail. The central and northern parts of Sweden currently lack detailed (> 1:50,000) QD maps. (a) a QD map at 1:1,000,000, (b) a QD map at 1:50,000, and (c) DEM of the same map area.

Rapid advances in computational capacity, data storage, and algorithm development, alongside the increasing availability of remotely sensed data, now offer a new opportunity to assess the potential of Machine Learning (ML) for surface deposit mapping. ML methods make minimal assumptions

about the relationships between predictors and the target variable, making them well-suited for capturing complex, non-linear patterns in high-dimensional environmental data (Breiman 2001b; Ryo & Rillig 2017). While ML has demonstrated considerable success in related fields such as geomorphology, land use/land cover classification, and Digital Soil Mapping (DSM) (Talukdar et al. 2020; Wadoux et al. 2020; Han et al. 2023), efforts to date in mapping surface deposits remain limited: pilot projects by engineering firms have explored depth-to-bedrock (DTB) mapping at local scales (Abbaszadeh Shahri et al. 2021), DSM studies have predicted specific soil properties in agricultural regions (Piikki & Söderström 2019; Adler et al. 2022), and individual deposit types such as peat have been mapped nationally (Rimondini et al. 2023). A comprehensive, multi-class ML framework for nationwide QD mapping has yet to be explored. This is largely because geological data are fundamental national infrastructure, with its characterization requiring immense resources, a high level of standardization, and long-term institutional commitment, a task that has traditionally been reserved for experts at geological surveys (McMillan 2002).

The convergence of nationwide LiDAR availability, advancements in algorithm development and computing power, Sweden's open data policies, and active institutional collaboration between government agencies and research institutions laid the foundation for this PhD project. Its main goal is to evaluate the potential of ML methods to accelerate QD mapping. Yet, as I quickly discovered, mapping surface deposits with ML is not just an exercise in data processing. Inferring what the ground is made of based on what the surface looks like demands a level of expertise in both geology and geospatial modelling that no Python package can offer. The extent to which this goal can be achieved, and what is gained or lost in the process, is what I will aim to outline in this thesis.

1. BACKGROUND

1.1 The Quaternary

Geologists divide Earth's 4.5-billion-year history into hierarchical units[2] of time, and the most recent of these is the Quaternary Period, the 'fourth age'. All Periods have ended, except the Quaternary. Spanning the last 2.58 million years, it is the only geological period still ongoing today (Gibbard et al. 2010; Lukas et al. 2017). Throughout this period, the Earth's climate has oscillated between glacial and interglacial intervals. The warm intervals often reached temperatures comparable to those of today, while the colder phases were dominated by glacial and periglacial processes, particularly in what are now temperate mid-latitude regions. The Quaternary began with the Pleistocene Epoch, commonly known as the 'Last Ice Age', which ended around 11,700 years ago (11.7 ka BP). In Fennoscandia[3] and northern Europe, the last glacial marking the final chapter of the Pleistocene was the Weichselian[4] glaciation (115-11.5 ka BP). The warmer phase we currently live in is called the Holocene (Cohen et al. 2013). Because each glacial cycle could preserve, rework, or erase deposits from previous ones, relatively little is known about earlier glaciations (Fredén 1990; Kleman et al. 2008; Wohlfarth et al. 2008). The Swedish landscape we see today, therefore reflects primarily the processes that occurred during and after the Late Weichselian, with only scattered remnants of older records surviving.

At their greatest extent (the Last Glacial Maximum), around 26.5-20 ka BP, the tremendous ice volume exerted a force that depressed the Earth’s crust downwards. Deglaciation across Fennoscandia was gradual, spanning roughly 22 to 9.7 ka BP (Patton et al. 2016; Stroeven et al. 2016), with the ice retreating from Sweden after 15.0 ka BP. As the ice load lifted, the crust

[2] Eon, Era, Period, Epoch, and Age.

[3] Fennoscandia, geologically, includes Scandinavia (Norway, Sweden, and Denmark), the Baltic Sea, Finland, and the Kola and Karelia regions of Northwestern Russia; in ice sheet glaciology, the term is used in a slightly more restrictive sense (Scandinavia, the Baltic Sea basins, and Finland) (Greenwood & Hughes 2022).

[4] Equivalent glacial periods have other regional names, e.g., Würm glaciation in the Alpine regions, Wisconsin in North America, Devensian in the British Isles etc. (Bowen 2009; Hughes et al. 2013).

began to rebound toward isostatic equilibrium (Björck 1995; Lambeck et al. 1998; Milne et al. 2001). It is a process that continues today, with uplift rates of around 10 mm/year in the Bothnian Bay, close to Umeå where this thesis was written (Steffen & Wu 2011; Vestøl et al. 2019).

A key geological marker linked to this rebound is the Highest Coastline[5] (*högsta kustlinje*, HC) (Fig. 1 and 3), the highest level reached by the postglacial sea before the land started to rise (Mörner 1979). The HC serves as a fundamental reference in geology, marking the boundary between areas once submerged by the sea and those that remained above it. It is central to understanding past sea levels and crustal rebound, and to predicting future landscape and sea level change (Mörner 1979; Peterson 2022). The HC formed at different times across the country as deglaciation progressed from south to north over several thousand years, with its elevation varying from less than 20 m above sea level in the south to 289 m in the north (Karlsson et al. 2021).

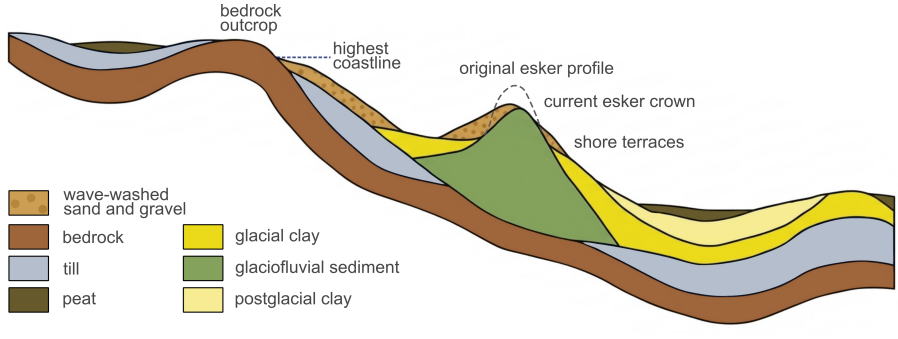


Figure 3. Schematic illustration of a typical sediment profile in eastern Sweden, showing minimal sediment cover in higher areas and thick layers of fine-grained marine and lake sediments at lower elevations (adapted from Wahlgren et al., 2019).

[5] Not to be confused with the High Coast (*Höga Kusten*), nearly identical both in Swedish and in English. Unlike the Highest Coastline, the High Coast is a region. Together with Finland's Kvarken Archipelago, it has been designated a UNESCO World Heritage site due to its uniqueness as an example of ongoing postglacial isostatic rebound (Poutanen & Steffen 2015).

1.2 Quaternary surface deposits vs. soils

One recurring challenge during this PhD is clarifying exactly what is being mapped, a task made more difficult by overlapping disciplinary boundaries and further complicated by a misleading translation[6]. This is reflected in the terminology used across the papers themselves: Paper I refers to 'parent material', Paper II predicts 'depth-to-bedrock (DTB)', and Papers III and this thesis adopt 'surface deposit' in the title, while the term 'QD' is dotted throughout. The term QD refers to the unconsolidated (loose) sedimentary material overlying bedrock, formed during the Quaternary period. Since most of them in mid-latitude regions were formed during the Quaternary, they are sometimes referred to as surface/surficial/superficial deposits to emphasize their immediate location at the land surface (Palmu et al. 2021).

It is perhaps easier to define what they are not: they are not 'soil' in the traditional pedological sense, which refers to the weathered surface layers that develop from parent materials through pedogenic (soil-forming) processes: the gradual formation of distinct horizons through biological, chemical, and physical weathering over time. However, even that definition shifts depending on the focus of the discipline (McBratney & Hartemink 2024). In Fennoscandia, QDs are the main 'parent materials' from which soils develop. 'Soil' development is limited compared to older landscapes elsewhere in the world, as simply not enough time has passed since deglaciation, and chemical weathering is limited due to the cold climate, leaving little opportunity for deep or complex soil profiles to form[7] (Mclaren & Turkington 2013; Spohn & Stendahl 2022). As a result, most of the 'soils' in Sweden are thin (20 – 50 cm) (Fredén 1990), incomplete weathered layer atop QDs, typically podzols[8] and histosols[9] (Jones et al. 2009).

---

[6] The Swedish word *jordart* (surface deposit/soil parent material) is literally translated as 'soil type' (*jord*-soil, *art*-type) and is sometimes loosely used as such. The actual pedological soil profile, known as *jordmån* is a terminology that is rarely maintained among the lay public.

[7] Most soils in Scandinavia are between 9,000 and 14,000 years old (Spohn and Stendahl, 2022).

[8] Acidic mineral soils typically found under coniferous forests, characterised by a leached grey horizon above a hard, reddish-brown layer. Podzols are common in boreal regions but not exclusive to them, and not all boreal soils are podzols (McLaren and Turkington, 2013).

[9] Organic, peaty soil.

1.3 Distribution, classification, and mapping of QDs

QDs are typically classified by two criteria: their genesis (the process by which they formed) and particle size (Fredén 1990; McMillan & Powell 1999). Terms such as gravel, sand, silt, and clay describe only the grain size, not how or when they were formed. Most sedimentary material is derived from bedrock that was eroded at one location and transported to another by glacial ice, meltwater, or wind before being deposited. A small proportion forms through in situ weathering and disintegration of the underlying bedrock, a process more characteristic of non-glaciated parts of the world (Karlsson et al. 2021). Peat, in contrast, is a postglacial deposit composed of organic material accumulated under waterlogged, anoxic conditions. Glacial (till) and glaciofluvial sediments are the most widespread deposits, covering over 75% of the country's land surface. In areas below the HC, wave and current action reworked the original deposits, washing out fine material from the upper layers. The landscape above the HC is unaffected by such coastal processes. Consequently, a location's elevation relative to the HC provides a practical and logical framework for interpreting and mapping these deposits (Fig. 3).

QDs and landforms are intimately linked: the shape of the land surface reflects how sediments were deposited (Benn & Evans 2014) (Table 1). Some deposits carry almost diagnostic geomorphological signatures: moraines and drumlins indicate glacial till; eskers suggest glaciofluvial sand and gravel; hummock-and-hollow microtopography is characteristic of peat. This is, of course, a gross simplification, as glacial deposits are often overlain by glaciofluvial and postglacial sediments, and the reality is considerably more complex (Fredén 1990; Putkinen et al. 2017).

Table 1. Typical glacial and postglacial deposits in Sweden.

Genesis	Deposit Type	Sorting Grade	Key Landforms	Occurrence
Glacial	Till (direct ice deposition)	very poorly sorted	moraines, drumlins, hummocky terrain	~75% of land surface
	Glaciofluvial (glacial meltwater deposition)	well- sorted stratified;	eskers, deltas, outwash plains	common but localized

Postglacial	Fluvial (modern river transport)	well-sorted, stratified	floodplains, terraces, alluvial fans	along rivers and lowlands
	Wave-washed (marine/lacustrine)	well- sorted	beaches, clay plains	below the HC
	Aeolian (wind-deposited)	very well sorted	dunes	rare, coastal
	Peat (organic accumulation)	n/a (organic)	bogs, fens, mires	extensive, especially in the north

The SGU is the central government agency responsible for geological information on bedrock, soil, and groundwater, and has been producing QD maps (*jordartskartor*) since its inception. Modern maps are produced in a GIS environment, with LiDAR-derived DEMs as the primary data source (Karlsson et al. 2021). Beyond maps based on direct observations, SGU also produces maps of subsurface properties using modelling. QD types and DTB are available as 3D models[10], but only for very localized areas. The existing DTB map is generated using inverse distance weighting (IDW) interpolation (Daniels & Thunholm 2014). ML has also been used to predict the likelihood of acid sulphate soil[11] occurrence.

[10] https://www.sgu.se/produkter-och-tjanster/kartor/geologi-i-3d/jordarter-och-jorddjup---3d-modeller/

[11] https://www.sgu.se/produkter-och-tjanster/kartor/kartvisaren/jordkartvisare/sur-sulfatjord/

2. RESEARCH AIMS

The overarching aim of this thesis is to investigate the potential of ML to accelerate surface deposit mapping in Sweden. The research began with two straightforward objectives: (1) identifying deposit type (classification) and (2) estimating deposit depth (regression) (Fig. 4). Together, they provide an opportunity to explore different facets of using ML for spatial prediction, centred around the following research questions:

- How well can ML predict surface deposit type and depth, and where and why does it fail?
- How reliable are the uncertainty estimates produced by the chosen frameworks, and what do they reveal about the limits of the model's predictions?
- What opportunities does a data-driven approach offer, and what are the limitations?

The thesis is organized as a collection of 4 papers, each representing a stage in the project's development.

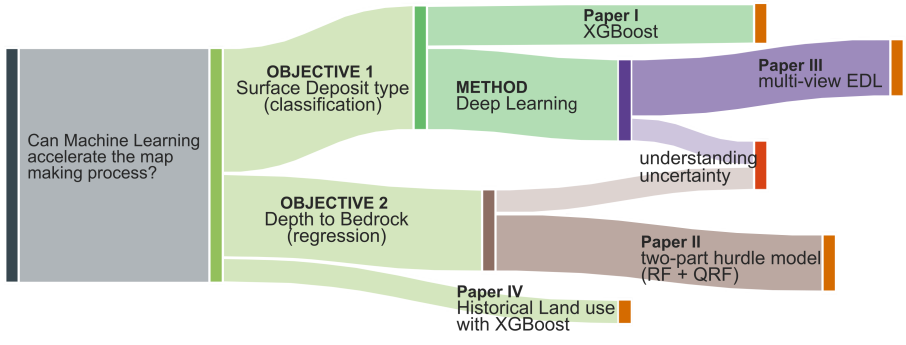


Figure 4. Overview of thesis structure and methodological development.

Paper I: Establishing the baseline. This study followed a standard supervised learning workflow: relating point observations to pre-processed terrain and hydrological indices to make predictions, ranking feature importance and interpreting the drivers of model predictions, followed by a spatial evaluation of model performance. The model was compared against legacy QD maps.

Establishing this baseline was a necessary first step to understanding the advantages and limitations of ML-based mapping relative to the manual approach, and the limitations it exposes.

Paper II: Regression with a zero-inflated dataset. DTB presents a modelling challenge: the dataset, aggregated from multiple existing sources, contains a substantial proportion of zero values (25%) representing true bedrock outcrops. These are governed by different processes than sediment-covered locations, and treating both in a single regression framework is inappropriate. This study addressed this through a two-part hurdle approach that combined a binary classifier and a regression model, while introducing uncertainty quantification via quantile regression forest (QRF) (Meinshausen 2006). It also explored methods for visualizing uncertainty.

Paper III: Shift toward Deep Learning (DL). This study reduced the reliance on handcrafted features in the previous ML studies by learning representations directly from LiDAR imagery using Convolutional Neural Networks (CNNs). The key methodological contribution was the combination of evidential deep learning (EDL) (Sensoy et al. 2018) with evidence fusion across multiple input views. Rather than producing a single class probability, the framework explicitly distinguishes between lack of evidence and conflicting evidence, a distinction that remains rare in geospatial mapping studies and has direct operational value.

Paper IV: Application to historical land use. This paper applied the Extreme Gradient Boosting (XGBoost) (Chen & Guestrin 2016) methodology from Paper I to classify historical land use from scanned mid-20th-century maps, producing reliable arable land classifications suitable for downstream analyses.

3. TECHNICAL FOUNDATIONS

The ML approaches used in this thesis fall into two categories with distinct data needs and workflows. Conventional tree-based ensemble ML methods (Papers I, II, and IV) operate on structured tabular data, where each labelled observation is associated with a set of predictor variables. These models learn relationships between covariates and the target variable, then apply the learned relationships to predict at unseen locations (Hastie et al. 2009). The DL approach (Paper III) automates much of the feature extraction and learns hierarchical representations directly from image data (LeCun et al. 2010; 2015). Despite these differences, both are supervised learning methods and require labelled training data.

No fieldwork was conducted as part of this thesis. All data was compiled from existing monitoring programs and databases. Relying on legacy datasets has well-known limitations: sampling designs reflect original survey objectives rather than modelling needs, creating methodological complications and interpretability challenges. For example, reporting protocols and classification schemes differ between surveys and over time, necessitating harmonization through aggregation and reclassification (Paper I). Combining DTB measurements from diverse sources can produce highly right-skewed, zero-inflated distributions that require specific modelling considerations (Paper II). Nonetheless, the spatial coverage and abundance of these legacy datasets could not have been achieved within the timeframe of a single PhD. The following sections describe the key datasets used across the papers (Table 2).

3.1 Data

3.1.1 Point observations

Swedish Forest Soil Inventory (SFSI) (Papers I and III)

Approximately 20,000 georeferenced soil observations distributed across Sweden on a systematic grid, with denser sampling in the south. Urban areas, arable land, water bodies, and high mountain regions are excluded. At each sampling location, parent material (five classes) and grain size (nine classes)

were determined from samples collected at least 20 cm below the humus layer using a probe (The Swedish National Forest Inventory 2021). We recategorized the samples into till, peat, coarse sediments, fine sediments, and bedrock outcrop classes.

Swedish Agricultural Soil Inventory (SASI) (Paper I)

Approximately 18,000 topsoil samples from agricultural land, aggregated from two national campaigns by the Swedish Board of Agriculture and the Swedish Environmental Protection Agency. Samples were classified as coarse or fine sediments based on the relative proportions of sand, clay, and silt.

Depth-to-bedrock (DTB) observations (Paper II)

Approximately 1.2 million measurements from eight SGU databases, the same data underlying the existing national IDW-interpolated DTB map (Daniels & Thunholm 2014). Sources include stratigraphic drilling data (well archives, peat archives, stratigraphic and hydrogeological databases), direct surface observations (bedrock outcrops, glacial striations, petrophysical measurements), and geophysical surveys. The dataset contains both confirmed measurements and estimated DTB values.

Historical land use from the economic map (Paper IV)

The economic map (*Ekonomiska kartan*[12]) (the Swedish Land Survey) is a historical map series covering most of Sweden (excluding the mountain range), produced between 1935 and 1988 and later digitized as GeoTIFFs. A total of 16,654 labelled samples were generated through manual digitization on 938 map tiles from 10 regions into four classes: arable land, forest, open land, and map graphics.

3.1.2 Environmental covariates

LiDAR DEM and derivatives

DEMs are fundamental to predictive mapping. Its derivatives (slope, curvature, topographic wetness index, geomorphons, surface roughness, etc.) transform raw elevation data into features that highlight different

[12] https://www.lantmateriet.se/sv/geodata/vara-produkter/produktlista/ekonomiska-kartan/#anchor-0

topographical characteristics and serve as inputs to ML models. Papers I and II used DEMs created from Sweden's first round of nationwide LiDAR scanning campaign (0.5 points m^{-2}), which produced elevation models at 2 m resolution. The subsequent update (1-2 points m^{-2}) produced DEMs at 0.5 m resolution (The Swedish Land Survey 2022). Paper III used multidirectional hillshade (MDHS) derived from the 0.5 m DEM as the sole raster input. Centered on each SFSI soil pit, images of three sizes (32×32, 256×256, and 1024×1024 pixels, covering 16 m, 128 m, and 512 m on the ground) were created, producing a total of 19,797×3 images. The images were labelled with the same five classes as in Paper I.

Other raster data

- **Quaternary Deposit (QD) Map**[13] (SGU): Compiled by aggregating the highest-quality mapping available as of 2020. It features variable scales ranging from 1:25,000 to 1:1,000,000 (Fig. 2).
- **Depth-to-Bedrock (DTB) Map**[14] (SGU): Provides existing DTB estimates generated through GIS analysis and IDW interpolation. Details about this map are provided in Paper II.
- **Distance to the HC** (SGU): Delineates the maximum post-glacial marine limit. It is a critical boundary between wave-washed zones and unmodified uplands.
- **National Land Cover Map**[15] (Swedish Environmental Protection Agency, 2019): A 10 m-resolution thematic map comprising 25 land cover and land use classes. It was created by integrating multisource data, including Sentinel-1 and -2 imagery, LiDAR, existing SGU geological maps, and various thematic maps.
- **Age Since Deglaciation** (SGU): Estimates the timing of ice-sheet retreat across Sweden (Hughes et al. 2016). Areas that deglaciated earlier have been exposed to postglacial processes, such as wave washing, sediment reworking, and peat accumulation, for longer.

[13] https://www.sgu.se/produkter-och-tjanster/geologiska-data/jordarter--geologiska-data/jordartsdata/

[14] https://www.sgu.se/produkter-och-tjanster/geologiska-data/jordarter--geologiska-data/jorddjupsdata/

[15] https://www.naturvardsverket.se/verktyg-och-tjanster/kartor-och-karttjanster/nationella-marktackedata/

Table 2. Summary of the inclusion of each spatial layer across the papers and their specific roles.

Data layer	Role	Paper I	Paper II	Paper III	Paper IV
QD map	Input	●	●	●	—
	Baseline	●	—	●	—
DTB map	Input	●	—	●	—
	Baseline	—	●	—	—
LiDAR DEM derivatives	—	●	●	●	—
Distance to HC	—	●	●	●	—
National land cover map	—	●	—	●	—
Age since deglaciation	—	●	●	●	—
Economic map	Input	—	—	—	●
	Training	—	—	—	●

● = Included in study

3.2 Machine Learning methods

3.2.1 Ensemble-based algorithms (Papers I, II, IV)

Ensemble methods combine predictions from multiple individual ML models, typically decision trees, to achieve stronger generalization than any single model (Dietterich 2000; Zhou 2012). The two methods used in this thesis differ in how trees are combined. Random Forest (RF) (Breiman 2001a) trains trees independently on random subsets of the data, with predictions averaged (regression) or decided by majority vote (classification) (Paper II). XGBoost (Chen & Guestrin 2016) trains trees sequentially, with each new tree correcting the errors of its predecessors, and the final prediction is the weighted sum of all trees (Papers I and IV).

3.2.2 Convolutional Neural Networks (Paper III)

CNNs are a DL approach designed for grid-structured data such as images (LeCun et al. 2015). Unlike traditional ML models that require hand-crafted features as input, CNNs learn to extract useful patterns directly from imagery during training. They do this using small learnable filters that slide across the input image, detecting spatial patterns at each location, conceptually similar

to how terrain indices are calculated using moving windows (van der Meij et al. 2022). A CNN is built from a sequence of layers, each performing a different role (Fig. 5). Convolutional layers apply these filters to the input, with early layers picking up simple patterns like edges and texture, and deeper layers combining them into more complex structures such as shapes. Pooling layers reduce the spatial size of the output by summarizing small regions into single values, keeping the important information while making computation manageable. Batch normalization stabilizes training by keeping values within a consistent range between layers. Dropout randomly switches off parts of the network during training to prevent it from memorizing the training data rather than learning generalizable patterns. Finally, fully connected layers combine everything the network has learned across the image into a class prediction. Through this layered structure, a CNN progressively transforms raw pixel values into increasingly abstract representations that capture what is relevant for classification.

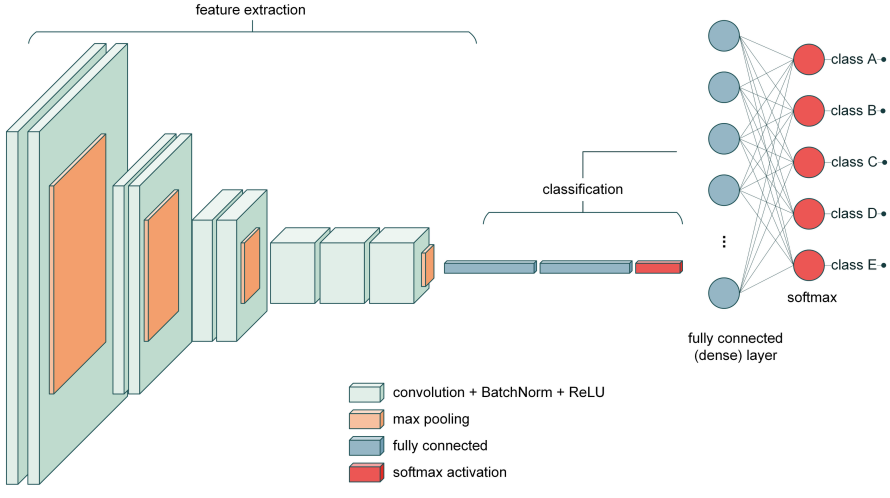


Figure 5. Overview of a standard Convolutional Neural Network. Input data undergoes feature extraction before being flattened and passed into a dense neural network for multi-class classification.

3.3 Evaluation and interpretability

3.3.1 Validation strategies

In Paper I, we constructed a held-out test set intended to approximate the Swedish landscape. SASI samples were first subsampled so that the ratio of SFSI to SASI samples (forest to agricultural land) roughly matched the national land cover proportions (approximately 68% forest to 7% agriculture). From this combined dataset, we then drew a stratified random sample, stratified by deposit type within each of the five SFSI sampling regions (Fig. 1d in Paper I), with weights inversely proportional to the region's sampling density to achieve a more or less uniform spatial coverage across the country. Hyperparameters were tuned using 5-fold cross-validation (CV) on the training set.

Papera II used stratified 5-fold CV with folds stratified by depth. For each outer fold, hyperparameters were tuned via an internal 5-fold CV on the 4 training folds. Performance metrics are reported as the mean across the 5 outer folds.

Paper III used stratified 10-fold CV. One fold served as the test set, the next as the validation set, and the remaining eight as the training set. A dedicated validation set was needed because CNNs are trained iteratively and require a separate partition to monitor overfitting and guide early stopping. Performance metrics are reported as the mean across the 10 folds.

In Paper IV, 20% of the digitized maps (181 tiles) were held out for testing, ensuring the model was evaluated on unseen sheets with varying colour characteristics and scanning artifacts.

3.3.2 Feature importance (Papers I, II, IV)

SHAP (SHapley Additive exPlanations) is an interpretability method that explains model predictions by assigning each feature a value representing its average marginal contribution across all possible feature combinations (Lundberg & Lee 2017). We specifically used the TreeSHAP (Lundberg et al. 2019) for XGBoost and RF. In classification tasks, SHAP values are computed per class, with the aggregated value indicating which features

matter most overall and which classes each feature is most influential for. In regression, they quantify how each feature pushes the predicted value above or below the baseline. SHAP supports both local explanation of individual predictions and global summaries of feature importance across the dataset and can be visualized through a variety of plots.

3.4 Quantifying and validating uncertainty

Uncertainty arises at many stages of ML-based mapping: measurement and reporting errors, location imprecision, limited or biased training data, weak explanatory power of the covariates, and model choice and misspecification (Nelson et al. 2011). Some sources are reducible with more or better data; others reflect variability in the phenomenon that cannot be easily explained away (Hüllermeier & Waegeman 2021; Fakour et al. 2024). Uncertainty is different from error: error is the actual discrepancy between a predicted and observed value at a given location, whereas uncertainty reflects the acknowledgment/awareness that predictions may deviate from the true value, though the exact magnitude remains unknown (Heuvelink 2018; Singh et al. 2024; Szatmári & Pásztor 2025).

Uncertainty quantification in spatial mapping has advanced considerably, with recent work emphasizing spatially explicit estimates, yet adoption remains uneven, and communication to end users is an unresolved challenge (Arrouays et al. 2020; Piikki et al. 2021). Producing uncertainty estimates is only part of the task: the reliability of these estimates, meaning whether the prediction is truly accurate when the model claims to be confident, is rarely validated with the same rigor as the way uncertainty is quantified (Schmidinger & Heuvelink 2023; Singh et al. 2024). In this thesis, uncertainty is quantified through prediction intervals (Paper II) and belief masses (Paper III), providing a measure of reliability at every predicted location. The reliability of these estimates is assessed using approaches suited to each framework.

3.4.1 Quantile Regression Forest (Paper II)

Quantile Regression Forest (QRF) (Meinshausen 2006) is an extension of the standard RF algorithm. Both RF and QRF build their ensemble of decision trees the same way: each training observation is passed through a

series of learned decision rules until it reaches a terminal node. What differs is what the leaf nodes store: rather than their mean, QRF stores all individual training values at each leaf node. When predicting for a new location, QRF pools these stored values across all trees to produce an empirical distribution of likely outcomes, from which prediction intervals can be constructed. Wide intervals indicate greater uncertainty; narrow intervals suggest confident predictions. Calibration of the prediction intervals is assessed using the Prediction Interval Coverage Probability (PICP), which measures whether the nominal coverage level matches the actual proportion of true values falling within the intervals. To complement it, we compute the Quantile Coverage Probability (QCP) (Schmidinger & Heuvelink 2023), which for each nominal quantile level gives the fraction of observations falling at or below the predicted quantile. Whereas PICP summarizes coverage at a single interval width, QCP diagnoses the shape of miscalibration across the predictive distribution, revealing whether it is biased (shifted), mis-scaled (too wide or narrow), or skewed (Schmidinger & Heuvelink 2023).

3.4.2 Evidential Deep Learning (Paper III)

Evidential Deep Learning (EDL) (Sensoy et al. 2018) is a classification framework that distinguishes between two types of uncertainty: lack of evidence (too little information is available) and conflicting evidence (information pointing in different directions). EDL modifies only the output layer of a deterministic neural network. Rather than outputting fixed class probabilities via softmax[16], the network outputs parameters of a higher-order evidential distribution, known as the Dirichlet distribution, which can be seen as a distribution over possible softmax outputs (Sensoy et al. 2018; Schreck et al. 2024). This is able to express not just the most likely class but also the confidence in that assignment (Fig. 6).

[16] Softmax is applied to the network's final layer to convert raw outputs into class probabilities that sum to 1. It does this by exponentiating each value and dividing by the sum. When trained with cross-entropy loss, the network is optimized to minimize the difference between these probabilities and the true class labels. Since the loss is smallest when the correct class probability is close to 1.0 and all others are close to 0, the model is incentivised to make its predictions as confident as possible. This leads to overconfident and unreliable outputs (Guo et al. 2017; Sensoy et al. 2018).

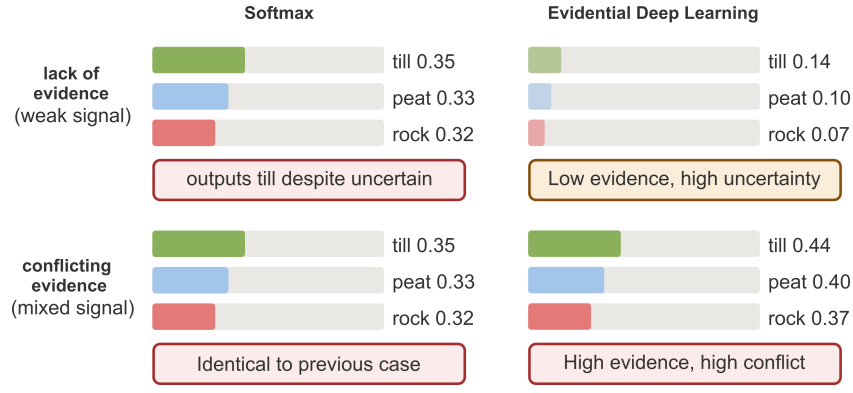


Figure 6. Comparison of standard softmax (left) and Evidential Deep Learning (EDL) (right) outputs. When the model is uncertain (top) or has conflicting evidence (bottom), softmax produces identical probabilities. EDL distinguishes the two: low belief masses indicate insufficient evidence, while high belief masses for competing classes indicate conflicting evidence.

Paper III used a multi-view architecture in which separate EDL networks were trained on different inputs: LiDAR MDHS imagery at three spatial extents and a tabular input incorporating additional spatial information. Each view produces independent predictions and uncertainty estimates. These are fused using a reduced version of Dempster-Shafer theory (DST) (Dempster, 1967; Shafer, 1976) proposed by Han et al. (2023). DST provides a mathematical framework for combining evidence that explicitly quantifies ignorance (lack of evidence) and treats conflict between sources as a measurable term (C) rather than averaging it away. In simpler terms, the DST fusion works like a panel of experts deliberating a case, each presented with different evidence. If all agree, the verdict is clear. If one is unsure but the others are confident, the uncertain opinion carries little weight. If some are confident yet disagree, the disagreement is recorded as conflict rather than cancelling each other out. This also differs from tree-based ensembles like RF, where individual trees are trained on a subset of the same type of input and their outputs are simply averaged (Fig. 7).

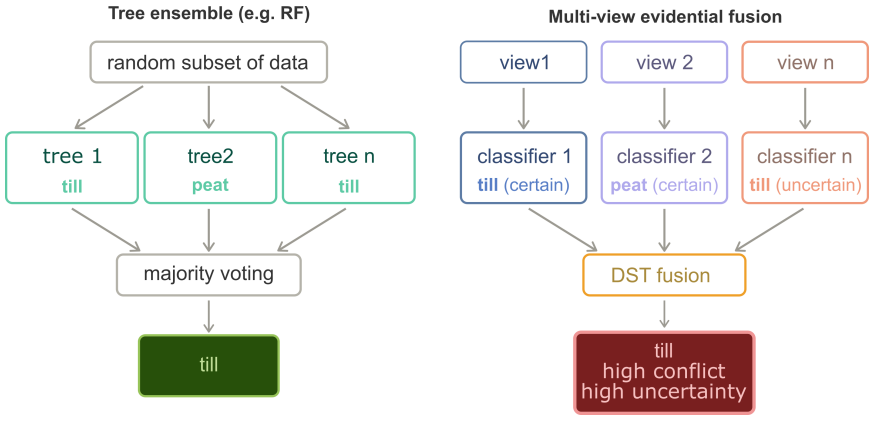


Figure 7. Left: in a tree ensemble, each tree is trained on a random subset of the data, and outputs are combined by majority vote or averaging. Disagreement between trees is smoothed away. Right: In multi-view Evidential Deep Learning, each view is trained independently on different data and produces its own belief distribution. Dempster-Schafer combination rules combine prediction while preserving disagreement as a measurable conflict term (C), yielding not only the prediction, but also uncertainty and conflict between views.

The quality of uncertainty estimates was assessed using sparsification curves, which progressively remove (sparsify) the most uncertain predictions and track how error changes. A model with reliable uncertainty estimates should show monotonic improvement in the evaluation metric as more uncertain samples are removed (Kristoffersson Lind et al. 2024). The sparsification curve is compared to an oracle curve, which represents the ideal case achieved by removing predictions in order of true error magnitude. The Area Under Sparsification Error (AUSE) (Ilg et al. 2018) is the area between the sparsification curve and the oracle curve, with lower values indicating better alignment between predicted uncertainty ranking and actual prediction errors. Sparsification evaluates whether uncertainty estimates correctly rank predictions from least to most reliable; it does not assess whether those scores are accurate in absolute terms.

4. MAIN FINDINGS

4.1 Surface deposit classification (Papers I & III)

Paper I demonstrated the potential for ML to automate surface deposit mapping at a national scale. A model trained solely on LiDAR-derived indices achieved an MCC of 0.45, outperforming the most detailed legacy QD maps (1:25,000-1:50,000; MCC = 0.36) and substantially outperforming overview maps (1:250,000-1:1,000,000; MCC = 0.18). Including legacy maps as covariates improved performance to an MCC of 0.56 (Table 2 in Paper I), though SHAP analysis revealed that legacy QD and land cover maps dominated all other predictors (Fig. 4 in Paper I). Paper III's multi-view EDL approach achieved a baseline MCC of 0.39, improving to 0.48 with legacy maps included (Table 3 in Paper III). Spatially stratified evaluation showed that both XGBoost (by physiographic region, Fig. 5 in Paper I) and multi-view EDL (by existing QD map quality, Table 4 in Paper III) performed better in the north than in the south, likely because the northern landscape is geologically simpler than that in southern regions below the HC. Paper III showed that different deposit types relied on evidence at different spatial scales: till benefited from a broader landscape context (~512 m), while bedrock outcrops and peat were best captured at a medium scale (~128 m). The small scale (~16 m) did not prove informative for any deposit type compared to its medium and large counterparts (Table 1 in Paper III). Given that the model receives both small and medium scales at the same 0.5 m resolution, this suggests that contextual information is necessary for deposit classification, at least with the current resolution.

The main argument for applying ML to surface deposit mapping is speed and scalability: two-thirds of Sweden cannot be mapped at a detailed level manually within any reasonable timeframe. However, performance varies substantially across deposit types, and not all classes are equally suited for automation.

Bedrock outcrops show clear automation potential. Although the class strictly indicates the absence of a deposit, it must be identified before deposit classes can be assigned elsewhere. The binary DTB classifier in Paper II

achieved a misclassification rate of 4.05%, and uncertainty estimates from Paper III's multi-view EDL framework are highly reliable for this class, meaning high-confidence predictions are generally correct. Performance would likely improve further with a bedrock outcrop-focused training set such as the outcrop observations from Paper II.

Peat mapping has already been demonstrated across several Swedish studies (Ågren et al. 2022 at 2 m resolution; Rimondini et al. 2023 at 10 m resolution, SGU's peat map[17]), and efforts have shifted toward mapping peatland types and microtopography (Karlson & Bastviken 2023; Noumonvi et al. 2025). Whether the models in these studies improve boundary delineation relative to existing products remains to be analyzed. The added value here lies less in detecting peat per se and more in providing spatially explicit uncertainty estimates that allow peat information to be used more safely in downstream analyses (Ballard et al. 2011; Beucher et al. 2023; Laudon et al. 2025).

Sorted sediments are unsuitable for automation under the current labelling scheme. **Coarse sediments** perform particularly poorly ($F1 < 0.1$ with ML, < 0.2 with DL), largely because the label groups genetically distinct deposits with very different terrain signatures: an esker has high slope and ridge-like curvature, while a delta or beach is flat with low slope and smooth curvature, yet they all share the same 'coarse' label. The model is presented with training samples that contradict each other. Rethinking the classification scheme to reflect landform would better align with what the data can distinguish. **Fine sediments** achieve higher performance ($F1 \sim 0.3$) but are overconfident (Fig. 6 in Paper III); fine sediments performed better than coarse likely because they are mostly flat glaciomarine and glaciolacustrine clay plains that form in low-energy environments, primarily beneath the HC, giving them a more consistent surface signal.

Till requires caution despite strong performance ($F1 \sim 0.8$). Confusion matrix and sparsification analysis (Fig. 5 and 6 in Paper III) show that most misclassifications default to till and that the model's most confident till predictions are frequently incorrect. This overconfidence is likely driven by

[17] https://www.sgu.se/produkter-och-tjanster/kartor/kartvisaren/jordkartvisare/torv/

the dominance of till in the training data. Defaulting ambiguous pixels to till resembles conventional mapping practice, where till acts as a background class for everything not explicitly delineated as something else. It may be worth reconsidering whether till should continue to function this way as it covers approximately 75% of the landscape, and direct effort toward the non-till classes instead.

Both the traditional ML and DL approaches used in this thesis are fundamentally data-driven: they learn associations between inputs and labels without knowledge of the processes that shaped the landscape. In traditional ML, the landscape context must be encoded manually. The surrounding landscape is summarized into terrain indices computed at different combinations of window sizes and resolutions, which in turn determines the scale of context captured, with little guidance on which configuration is optimal for a given geomorphic process (Padarian et al. 2019; Maxwell & Shobe 2022). In DL, the surrounding landscape is provided via input image patches. Paper III used three spatial extents, but the choice of which extents to provide remains a modelling decision.

Both methods attempted to encode landscape processes through the above/below HC variable and other spatial features. XGBoost treats process-relevant variables as covariates; in Paper III the features were incorporated as tabular input to provide additional context to image inputs. In both cases, the models were free to explore their effectiveness. There is growing interest in embedding domain knowledge directly into ML architectures (Wadoux et al. 2020; Von Rueden et al. 2023; Minasny et al. 2024). A more principled approach would encode process knowledge as a rule, since such constraints are known: wave-washed deposits cannot exist above the HC, and confirmed bedrock outcrops cannot be any other class. Encoding these directly would guarantee their presence at prediction time, rather than relying on the model to learn them in relation to other information. There are existing data worth considering for the future: many large glacial and glaciofluvial landforms have already been mapped by SGU and could serve as observational priors (Minasny et al. 2024). Other geological data, such as glacial striations and reconstructed ice-flow directions, carry information about past glacial processes, but translating them into model-ready constraints is less straightforward.

4.2 Depth-to-bedrock prediction (Paper II)

A defining characteristic of the DTB dataset is the high prevalence of bedrock outcrops (DTB = 0). Fitting a single regression model to a dataset dominated by zeros can create biased predictions and uncertainty estimates (Zuur et al. 2010). This motivated the two-part framework, where a binary RF classifier first separated bedrock outcrops from sediment-covered locations, and a QRF then predicted continuous DTB values. The two-part framework outperformed both the existing national IDW interpolation and a coarse global model in overall metrics.

The framework's clearest advantages over the existing IDW-based map are in bedrock outcrop delineation and DTB prediction in data-sparse regions. IDW assigns zero values to all mapped outcrop polygons, which typically encompass areas larger than the actual outcrops (Daniels & Thunholm 2014; Karlsson et al. 2021). In data-sparse areas, IDW degraded when the nearest observations were distant (ME = −2.72 m at <1 pt/km²), whereas the ML model remained stable (ME = 0.85 m) because it applies learned relationships between environmental covariates and depth rather than relying on proximity to nearby observations. This finding is particularly useful for mountainous and northern inland areas with limited field data.

For continuous depth prediction, the two-part model performed comparably to or better than IDW between 0 and 5 m. All models shifted from overestimating shallow depths to underestimating deeper depths. Beyond 5 m, IDW maintained smaller mean errors than the two-part model, which reached an ME of 12.44 m beyond 30 m compared to 5.68 m for IDW. This pattern is likely because the link between surface topography and bedrock depth weakens with increasing depth, so deep and shallow deposits can appear similar in topographic data. The training data are heavily skewed toward shallow observations (75% below 10 m, median 4.5 m), which biases the learned relationships toward this range.

The uncertainty estimates from QRF were slightly conservative overall ($PICP_{90}$ = 93.2%; Fig. 4a in Paper II). At shallow depths (2–5 m), the prediction intervals were wider than necessary ($PICP_{90}$ = 98.3%; Fig. 4b in Paper II), indicating conservative model behaviour and imprecise predictions. However, this can be beneficial for high-stakes applications

where capturing a broad range of plausible values is preferred. Beyond 30 m, the prediction intervals were too narrow while the model simultaneously underestimated depth. This combination of being wrong and overconfident is dangerous for practical applications. QCP diagnostics (Fig. 4c in Paper II) clarified the mechanism of miscalibration across depths, which were otherwise masked by the reliability plot (Fig. 4a in Paper II), underscoring the need to combine multiple metrics with stratified evaluation.

4.3 Historical land use (Paper IV)

Paper IV showed that ML can classify land use from historical orthophoto-based maps with high overall performance (MCC = 0.86). Arable land and graphics (map text and symbols) were classified with high accuracy (F1 = 0.95 and 0.98), while forest and open land were harder to distinguish (F1 = 0.84 and 0.87) due to colour and texture similarities in the source orthophotos (Table 1 in Paper IV). The model generalized well to unseen maps with varying colour characteristics arising from differences in production year and scanning conditions. SHAP analysis revealed that arable land was primarily identified through colour-based features. The arable land class is suitable for regional-scale analyses of agricultural change. Forest and open land classes, however, should be used with caution: their misclassification rates are comparable to the expected land use changes between time periods, making it difficult to distinguish real change from classification error.

5. OVERALL REFLECTIONS

5.1 Model evaluation

A fundamental distinction exists between model performance and map accuracy (Wadoux et al. 2021; Rolf 2023). Model performance, reported as metrics throughout these studies (MCC, F1, RMSE), measures how well the model predicts at held-out locations under specific evaluation conditions. Map accuracy is a property of the final mapped product across the prediction domain, and ideally requires a probability sample collected separately from the model training process to be estimated validly (Wadoux et al. 2021; Rolf 2023). The following subsections reflect on the validation approach used to produce the metrics in this thesis, and on the limits of what those metrics can tell us.

5.1.1 Choice of validation approach

Random CV is known to produce optimistic accuracy estimates when training data are spatially clustered (Wadoux et al. 2021; Milà et al. 2022). This is a concern for Paper II, where the DTB dataset combines multiple sources. In extreme cases, nearby observations may have nearly identical covariate values and targets, which means randomly assigned validation samples could effectively serve as pseudo-replicas of the training samples (Wang et al. 2023). Spatial CV methods have been proposed to address inflated metrics due to spatial autocorrelation between sampling points (Le Rest et al. 2014; Roberts et al. 2017; Valavi et al. 2019), but these methods withhold entire areas, forcing the model to predict into space it has not encountered during training. Since our prediction domain lies within the sampling domain, our main goal was not to test the model's spatial transferability (Wadoux et al. 2021; de Bruin et al. 2022). We attempted k-fold nearest-neighbour distance matching CV (Linnenbrink et al. 2024), which builds CV folds so that test-to-training distances resemble those the model will encounter at deployment. This gives a more honest estimate of error, but the approach did not scale to 1.2 million data points. A more concerning issue is that with 75% of observations shallower than 10 m, the model has limited support for deeper depths. We used stratified random CV,

which ensures the validation folds span all depths and revealed the depth-dependent over- and underprediction. However, it does not directly address pseudo-replicas or extrapolation in feature space, which are properties of the training data; a validation method can only reveal them (Jemeļjanova et al. 2024). With large datasets like this, it is worth considering whether sampling bias should be reduced at its source, for example, through thinning or weighting. Treating all available data as equally informative simply because it is available leaves these biases unaddressed. However, even a balanced dataset would not fully resolve deep DTB prediction, as this is a fundamental limitation of using surface-based covariates rather than a data availability issue.

Spatial clustering is less of a concern in Papers I and III, where the SFSI sampling follows a systematic grid, though with higher density in southern Sweden. In Paper I, we constructed a spatially uniform test set through stratification by dataset to reflect national land-use proportions and weighting by regional sampling density. Paper IV is not a spatial prediction problem. The classification task is based on spectral and filter-derived features of scanned map imagery, where variation lies in colour differences due to paper aging, printing techniques, and scanning artifacts across sheets. Splitting at the tile level ensures the model is evaluated on unseen sheets with potentially different visual properties, which is appropriate for this task.

5.1.2 Limitations of the reported metrics

In Papers I–III, we compared the performance of the ML models with the legacy maps using the same point-based evaluation. Point-support evaluation remains the standard in spatial mapping (Bishop et al. 2015), largely because field observations are collected at points, and true map accuracy across the full prediction area cannot be measured directly (Wadoux et al. 2021). However, this comparison has several limitations. First, the evaluation aligns with what the ML model was optimized to do: predicting at point locations using a subset of the dataset it was trained on. Legacy QD maps were produced to represent the dominant deposit type within polygons at a given cartographic scale, without point-level accuracy as their objective. Evaluating both at points favours the ML model. Second, legacy map ‘errors’ at point locations often reflect the spatial-unit mismatch rather than genuine

mistakes (Fig. 8). The SFSI field observations cover ~314 m², the ML model predicts at 4 m² pixels, and legacy maps use minimum mapping units (MMUs) (Saura 2002; García Álvarez 2017) ranging from 625 m² to 1,000,000 m² (Karlsson et al. 2021). A polygon label represents the dominant class within that unit (Warner et al. 2009), so a sampled point that happens to be on a secondary deposit will appear as an error. Small deposits that the ML model can resolve are invisible at map scale. The XGBoost's MCC of 0.56 is therefore a model performance metric calculated from point predictions, while the legacy map's 0.36 reflects how well a generalized polygon product matches point observations. These numbers are not directly comparable. Claims that ML outperforms the best available QD map likely overstate the improvement given these limitations. The comparison in Paper II is more straightforward: both the two-part model and the existing IDW interpolation use the same dataset and share the same objective of predicting depth at unsampled locations, making it a more direct comparison of methods.

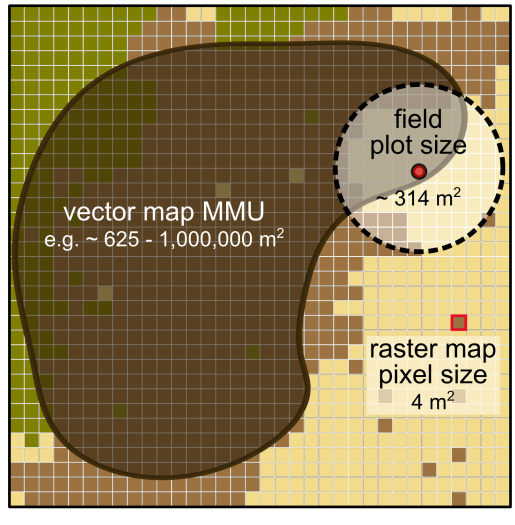


Figure 8. Schematic illustration of the challenge of point-scale evaluation caused by unit-size mismatch.

Another limitation of aggregate metrics is that they can conceal important patterns. We reported stratified results (Olofsson et al. 2014; Nelson et al. 2021), but even disaggregated metrics do not reveal whether the resulting map is spatially coherent. A model can achieve low RMSE or high MCC

while placing predictions in incorrect locations (Congalton 1991; Lyu 2025). The ML model's higher performance across the papers does not guarantee that the maps will be visually coherent or geologically meaningful. Detecting such spatial artifacts requires inspecting the predicted maps, which we did not do at scale within the constraints of this thesis.

5.2 The interpretability challenge

Since supervised ML models are designed to minimize prediction error, they have no mechanism to distinguish genuine patterns from spurious correlations arising from biases in the training data (Sasse et al. 2025; Ye et al. 2025). Post-hoc interpretation methods like SHAP can help expose such issues, but only when data characteristics and predictor-target relationships are scrutinized (Kaufman et al. 2012; Davis et al. 2023; Zhu et al. 2023). The broader lesson is that feature importance scores reflect what a model found useful in a specific dataset, but rankings obtained from that dataset may not generalize and should be validated on independent data. Interpreting them uncritically risks mistaking data biases for genuine insights (Jiang et al. 2024). The following subsections illustrate some challenges encountered in the papers.

5.2.1 Spatial features as predictors

Spatial features are simultaneously useful and a source of methodological complexity in environmental ML (Hengl et al. 2018; Koldasbayeva et al. 2024). While they are often top predictors, their high importance can stem from different underlying mechanisms. In Paper I, a classification task, class-specific SHAP values showed that Northing was highly important only for the two sorted sediment classes from the SASI dataset (Fig. 4b in Paper I). If Northing reflected a generalizable covariate–target relationship, it would also be important for predicting the coarse and fine sediment classes in the SFSI dataset. The class-specific pattern indicates that the model has instead learned a spatial shortcut: since SASI samples were collected exclusively from agricultural land, mainly in southern Sweden (Fig. 1 in Paper I), Northing acts as a proxy for SASI dataset membership rather than a predictor of sediment type. This likely contributed to the inflated performance on SASI classes (F1 > 0.72, Fig. 3 in Paper I).

In Paper II, Northing and Easting were important for predicting DTB (Fig. 5 in Paper II), but with reasons that are plausible yet difficult to verify, as many covariates are spatially structured. Sweden's geology and topography follow broad directional gradients: bedrock becomes harder and less erodible from south to north, sediment cover becomes thinner from southeast lowlands to mountainous northwest, and earlier deglaciation in the south allowed for longer sediment accumulation. The SHAP values for individual features aligned with these gradients and are largely interpretable. The challenge is that the contributions of Northing and Easting cannot be separated from the spatially structured covariates themselves (lithotectonic type, age since deglaciation, etc.), and the explicitly spatial predictors may provide redundant information already captured by other features. Ablation testing is recommended to identify such redundancies during model training (Meyer et al. 2019). Beyond interpretation, models that rely heavily on coordinates risk producing map artifacts tied to coordinate values rather than landscape processes (Meyer et al. 2019; Pawley et al. 2024), which again requires evaluating predicted maps at a broader scale.

5.2.2 Model training with legacy maps: benefits and risks

Legacy maps represent a form of expert knowledge (Minasny et al. 2024), but also carry limitations from outdated methods, lower resolution, and inconsistencies with modern data standards (Liu et al. 2022; Li et al. 2025; Yang et al. 2025). ML approaches have been used to integrate new data sources to improve the delineation and accuracy of legacy maps (Pahlavan-Rad et al. 2016). In Papers I and III, including the QD map consistently improved ML (MCC = 0.45 without vs. 0.56 with) and DL (MCC = 0.39 without vs. 0.48 with) model performance. In Paper III, the improvement was not sensitive to the original map generalization level, with performance gains consistent across areas covered by ‘generalized’ maps (<1:100,000; +0.10 MCC) and areas with ‘detailed’ maps (>1:50,000; +0.08 MCC) (Fig. 2).

Understanding how the models leverage information from legacy maps and other environmental covariates to achieve this performance gain matters. If the model predominantly copies legacy labels, it risks propagating legacy map errors and boundary imprecision into the final product (Behrens et al.

2008) and may create the illusion of improved map quality when predictions are simply the old map rendered at finer resolution (Malone et al. 2017; Wadoux et al. 2020). An alternative explanation is that legacy maps can provide contextual information even when they differ from the field label at the validation location. Because polygon maps assign a single label to the dominant class within an MMU of up to 1,000,000 m² (Section 5.1.2), a label that disagrees with the field observation at a specific point may still correctly describe the general area. For example, when a training point falls on a peatland within a till-dominated polygon, the model receives 'till' as the legacy map covariate (Fig. 9). The legacy map label is 'wrong' at that point, but the model still gains insights from its combination with other covariates. In a scenario where environmental covariates indicate either peat or fine sediment, a legacy label of 'till' tells the model that this area is till-dominated, making peat (a common inclusion within till landscapes) more likely than fine sediment. In this interpretation, the legacy map functions less as a direct answer and more as a contextual constraint that can help the model classify correctly.

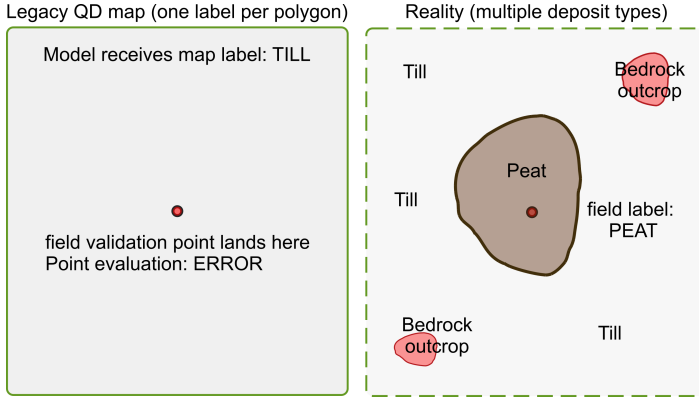


Figure 9. Legacy map labels can convey contextual information even when they differ from the field label. Left: the legacy QD map assigns a single label to the entire polygon. Right: In reality, the same area contains multiple classes. The map label is 'wrong' at that location, but still reflects the characteristics of the overall area, which may help narrow the set of plausible classes.

For XGBoost (Paper I), disentangling these mechanisms is challenging. The legacy map dominated the modelling process because it was the most efficient predictor, particularly for classes like till, bedrock outcrop, and peat, which have directly corresponding classes in the target (Fig. 4b in Paper I).

A three-way comparison between the legacy-informed model, an ablated model trained without the QD map, and the map label at each validation location would help clarify the mechanism. Where the ablated model is wrong but the legacy-informed model is correct, the legacy map contributed, and comparing against the map label can distinguish why. This would reveal the mechanism more clearly than SHAP values, which describe only association.

In XGBoost, the legacy map enters alongside all other covariates and shapes the model during training, so how the model uses it cannot be easily disentangled. In the evidential fusion framework, each view is trained independently and produces its own belief distribution before fusion, allowing direct comparison of what each source concluded and how fusion changed the outcome. If the fused predictions consistently match the legacy map labels regardless of the image views, this indicates reproduction. If the legacy map and image views are both wrong, yet the fused output is correct, this would suggest that disagreement between sources suppresses incorrect answers and tips the prediction toward the correct class. If such cases are common, it would suggest that the legacy map contributes beyond providing a direct answer and that the resulting product is a genuine improvement. The architecture does not in itself improve predictive performance, but it makes the distinction clearer.

5.2.3 Uncovering association-based leakage

A subtler and in some ways more problematic interpretability challenge arises from association-based leakage, where a feature is not the target variable itself but becomes a surrogate for it through the sampling context (Kaufman et al. 2012; Davis et al. 2023). In Paper I, sorted sediments on agricultural land (SASI samples) achieved strong performance (F1 > 0.72), with the land cover map being the most important feature (Fig. 4 in Paper I). Paper I acknowledged that the land cover map's importance was tied to the arable land class, which the SASI samples relied on heavily, but did not identify it as a leakage.

The high importance of the land cover map is more precisely described as association-based leakage, where a feature available at prediction time (land cover map) is deterministically related to the outcome (‘agricultural’

coarse or fine sediments) through the sampling context (arable land) (Davis et al. 2023). The model was trained with samples defined as coming from agricultural land and then provided with a feature that directly encodes whether a location is agricultural. What makes this case more insidious is that the leakage is not self-evident and could easily be misinterpreted as something more meaningful: coarse and fine sediments form fertile plains suited for agriculture, so agricultural land use serves as a plausible indicator of sediment type. The motivation for including the land cover map was sound, and a non-class-specific SHAP value for this feature would have attributed this high importance with a readily available, convincing explanation.

The class-disaggregated SHAP values exposed the mechanism. The land cover map contributed almost exclusively to predictions for SASI samples, not SFSI samples (Fig. 4 in Paper I). If the model had genuinely learned the sediment-land use relationship, the land cover map should contribute to predictions across all sorted sediment samples regardless of dataset origin. Removing the land cover map caused a sharp drop in SASI samples' F1 (~0.5, Fig. 3 in Paper I), confirming that the strong performance on SASI samples depended on the land cover map, and the final model MCC was inflated precisely because of that. This reflects a flaw in the modelling design in Paper I: by combining both dataset origin and deposit class within the target labels, the model was presented with two problems simultaneously: classifying where samples came from and classifying the deposit type, where it solved the former but was credited for the latter.

5.3 Understanding uncertainty

Performance metrics might provide a sense of model reliability. Nilsson et al. (2020), for example, called precision above 0.8 'very good' (Paper IV). As discussed in Section 5.1.2, aggregate metrics describe average behaviour and do not reveal how reliable any individual prediction is. Till achieved the highest F1 scores, yet its most confident predictions were frequently wrong (Paper III). The DTB model achieved low overall RMSE but severely underestimated depth beyond 30 m (Paper II). Without spatially explicit uncertainty estimates, users have limited means to judge where on the map

the model can be trusted. Current QD and DTB maps convey uncertainty only descriptively (Daniels & Thunholm 2014; Karlsson et al. 2021).

In Papers II and III, we used models that naturally produce uncertainty estimates and evaluated whether the quantified uncertainty was valid. Paper I did not provide uncertainty information. Paper IV used the term 'uncertainty', probably too loosely to describe misclassification. XGBoost can output class probabilities, but these reflect only how often the model sees each class given its training data. A model can assign high probability to a class simply because the training data is dominated by that class, as is the case with till.

Spatially explicit uncertainty could, in theory, prevent false confidence in map products and allow uncertainty to be factored into decision-making (MacEachren et al. 2005; Lark et al. 2022). In practice, however, a gap exists between producing and validating uncertainty estimates and translating them into something end users can act on (Kinkeldey et al. 2017). Even within the research community, terms like 'prediction interval,' 'calibrated,' 'conservative,' and 'over/undercovers' require familiarity with the underlying concepts (Lark et al. 2022); otherwise, outputs such as PICP and QCP values and associated plots can be more confusing than informative. Each step from model output to decision (producing metrics, quantifying uncertainty, validating, communicating, and ensuring it informs decisions) becomes increasingly abstract (Srivastava et al. 2026), and end users' decision processes rarely match the sophistication of the methodology (Lark et al. 2022). This disconnect was evident in our biannual meetings with the reference group, which includes stakeholders from government agencies and companies with a need for improved maps. Uncertainty information was acknowledged as important, but its practical use remained unclear, both in terms of the format users would prefer and how they would incorporate it into their workflows. The general sense is that everyone demands such information, but even when made available, it is hard to use and rarely influences the planning process.

We nevertheless tried to address the communication barrier by visualizing uncertainty directly on maps. In Paper II, we explored two approaches: displaying uncertainty alongside the prediction, and the coincident QuadMap method (Padarian & McBratney 2023), which

combines prediction and uncertainty on a single map using pixel size to indicate uncertainty (Fig. 10). Neither was tested with end users. Evidence on how users respond to uncertainty information is mixed: some studies suggest it increases trust through transparency, while others find it may overwhelm end users, reduce confidence, or create confusion (Korporaal et al. 2020; van der Bles et al. 2020; Courteille et al. 2025). This thesis demonstrated what is technically feasible. Future work should focus on testing these approaches with end users and tracking whether uncertainty information influences decisions in practice, rather than simply meeting a methodological expectation.

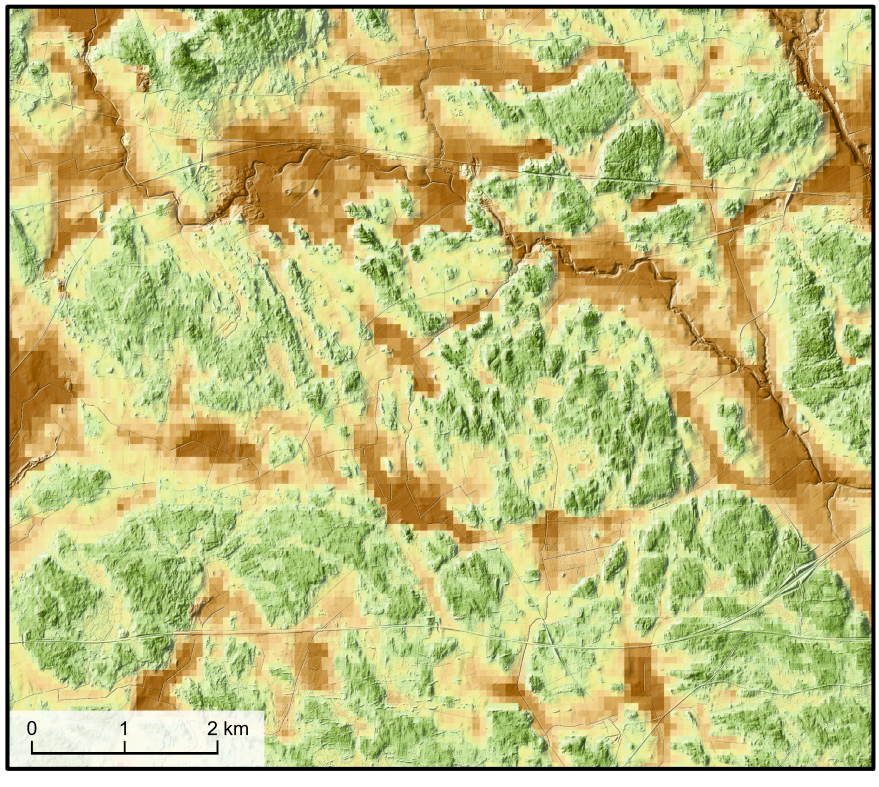


Figure 10. QuadMap representation of depth-to-bedrock (Paper II). Colour indicates predicted depth: green for shallow and brown for deep (0 -14 m). Pixel size encodes uncertainty, with larger pixels representing more uncertain predictions.

5.4 Future work

QD mapping is a century-old but ongoing task, and completing the remaining two-thirds of Sweden at sufficient detail through manual mapping is not feasible within a reasonable timeframe. Considerable work has focused on refining legacy maps, such as using spatial disaggregation methods to downscale coarse polygons to finer rasters (Bui & Moran 2001; Odgers et al. 2014). More recently, DL has been used to produce surficial geology predictions with the vision that experts would refine the initial machine predictions (Latifovic et al. 2018). Adding legacy maps as covariates in ML model training improved performance, but how models balance information from environmental covariates against legacy map labels was not investigated in our current work. The question is therefore not whether ML will play a role or whether to use legacy maps, but how they can be used together to achieve this goal.

The multi-view evidential fusion framework from Paper III can offer one path forward, where ML is used to audit the legacy map rather than disaggregate, or predict a new map from scratch. DST fusion produces per-view belief, per-view uncertainty, and a conflict term (C) quantifying disagreement between views. These outputs are well-suited for a map-auditing workflow. Because Dempster's rule is defined pairwise, C is computed at each fusion step and is therefore available without additional computation (Han et al. 2022). This allows C to be measured separately between image views and between the fused image evidence and the legacy map. Rather than treating the legacy map as another covariate, it can enter via a dedicated pathway, with its labels serving as fixed evidence that image-based evidence either confirms or contradicts. Depending on the image views' predicted class and confidence, high-C areas can either be automatically revised (e.g., if images confidently predict bedrock outcrop or peat) or flagged for expert review (Table 3). Full deployment would require modifying the input so that the legacy QD map, currently one-hot encoded and bundled with other covariates, enters independently.

Table 3. Interpretation of evidential fusion outputs for the proposed map-auditing workflow. Combinations of image consensus, image-image conflict, and image-map conflict guide automatic acceptance, revision, or expert review.

Image-image consensus	Image-image conflict (*C*)	Image-map conflict (*C*)	Interpretation	Action
Agree	Low	Low	Consensus across all sources	Accept classification
Agree	Low	High	Images agree but contradict map	Bedrock/peat: accept. Others: flag for review
Ignorant	Low	Low	Insufficient image evidence	Retain map label; prioritize for review
Disagree	High	High	Conflict across sources	Accept fusion result (caution) or flag for review

Paper III demonstrated this concept, though not yet in auditing form (Fig. 11). In its current implementation, we compared the final fused prediction against field validation labels: grey points indicate consensus, yellow points indicate locations where the model was uncertain and incorrect, and red points indicate locations where the model confidently predicted a different class. This visualization is limited to validation locations because it requires field labels. In the proposed workflow, *C* would instead measure disagreement between the fused LiDAR imagery views and the legacy map, without requiring field data. A further improvement would be to train the model on the original, unharmonized classes from the legacy map, preserving distinctions between deposit types that are collapsed into grain size-based categories. The model could then audit the legacy map against its own classification scheme rather than a simplified version.

Translating this into an operational system has implementation challenges. Calibrating conflict thresholds for each class to determine what constitutes high or low disagreement, and ensuring the volume of flagged areas is manageable, are two apparent issues. If too large a portion of the map requires human intervention, it offers little practical advantage over manual mapping. Despite these challenges, the workflow reframes EDL and evidential fusion from a prediction method into a support tool for map production itself, where uncertainty and conflict assume a different role that can lead to specific actions (accept, revise, or refer for expert review), directing expert attention to the locations where it matters most.

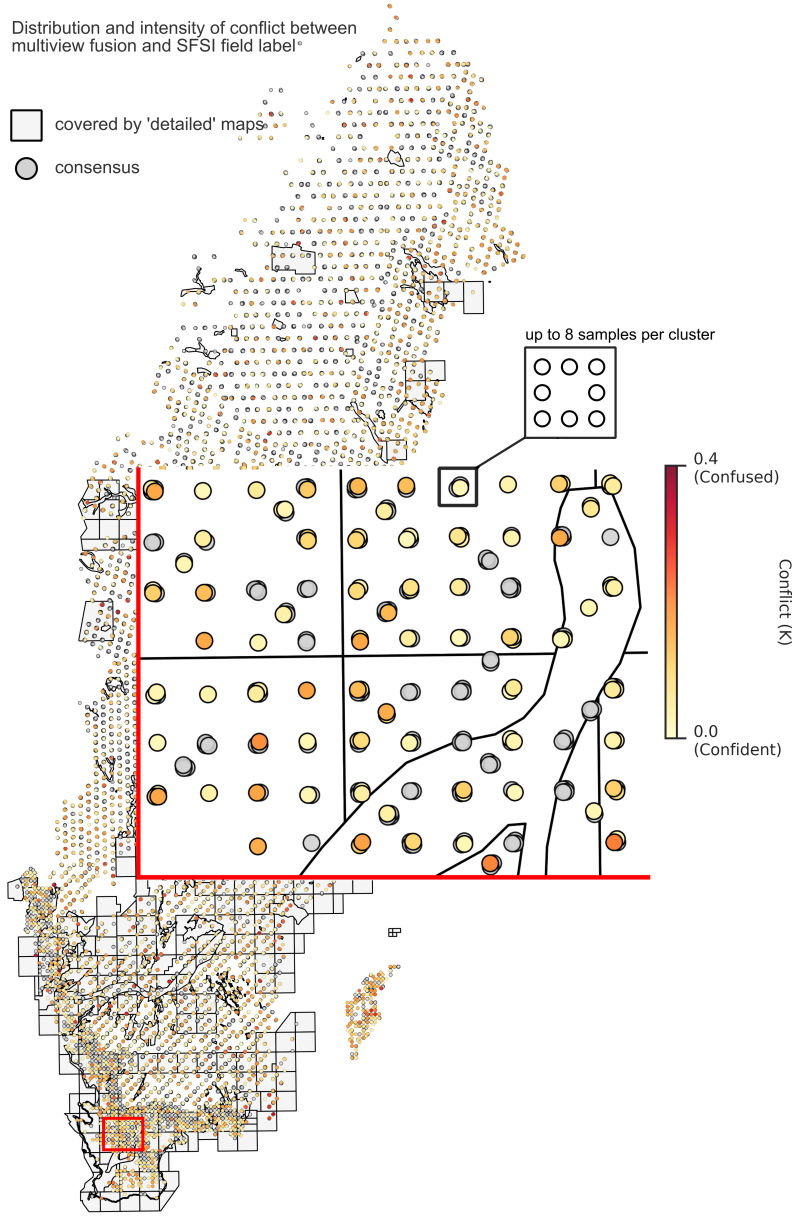


Figure 11. Agreement and conflict between the multi-view Evidential Deep Learning model's predictions and field labels. Grey circles indicate consensus. Yellow circles indicate low-conflict errors (model uncertain but incorrect). Red circles indicate high-conflict errors (model confident but incorrect). In the proposed map auditing workflow, a similar mechanism would instead show conflict between image views and the legacy map, computed at every location without requiring field data.

6. CONCLUSION

This thesis set out to address three questions: how well ML can predict surface deposit type and depth; how uncertainty in predictions can be quantified and validated; and what the opportunities and limitations are.

Regarding the first question, not everything can be reliably mapped using current approaches. Performance varies substantially by deposit class and depth range. Bedrock outcrops and peat can be mapped reliably and are realistic candidates for automation. Till achieves high aggregate performance, but models tend to default to till whenever the input is ambiguous. Sorted sediments remain beyond the reach of current approaches. For DTB, ML remains accurate within the operationally important shallow ranges (small ME under 10 m). Its main advantages over IDW are a substantially lower bedrock misclassification rate, spatially explicit uncertainty estimates, and stable performance in data-sparse areas.

On the second question, this thesis demonstrated that QRF (Paper II) and multi-view EDL (Paper III) produce spatially explicit uncertainty as a direct output of the model. This is an attractive feature of both frameworks. Validating these uncertainty estimates is equally important, and doing so revealed that DTB predictions at greater depths, till, and sorted sediment classifications may not be trustworthy. Despite initial exploration of visualization approaches (Paper II), how this information can be translated into something that informs decision-making remains largely open.

On the third question, the central insight is that ML and manual interpretation face different bottlenecks and are complementary rather than interchangeable. Manual mapping is constrained by time and resources. Given enough of both, an expert can map even the most geologically complex areas. ML is constrained by the data it is trained on: most fundamentally by whether enough labelled observations are available, but also by whether the covariates carry information about the target, and whether the training observations cover the full range of conditions. Ambiguous classes like sorted sediments and deep DTB predictions will remain difficult regardless of how sophisticated a model is. The disadvantage of ML becomes clear when data limitations force compromises, such as

grouping detailed QD classes into broader categories for model training. This common practice results in information loss in the final output. However, when ML does work, it offers high-resolution predictions with spatially explicit uncertainty.

ML is often described as data-driven, but applying it well involves far more than fitting a model to data. The structure of the data and the intended use of the model must be understood in order to make reliable choices about validation strategy, which in turn shapes the reported metrics and their interpretation. Each stage of the workflow, from data preparation to evaluation to interpretation, requires careful consideration to avoid misinterpreting or overstating the results. Rather than a replacement for expert interpretation, ML is better positioned as a tool, and like any tool, it works best when you understand how to use it.

References

Abbaszadeh Shahri, A., Shan, C., Zäll, E. & Larsson, S. (2021). Spatial distribution modeling of subsurface bedrock using a developed automated intelligence deep learning procedure: A case study in Sweden. *Journal of Rock Mechanics and Geotechnical Engineering*, 13 (6), 1300–1310. https://doi.org/10.1016/j.jrmge.2021.07.006

Adler, K., Piikki, K., Söderström, M. & Eriksson, J. (2022). Digital soil mapping of copper in Sweden: Using the prediction and uncertainty as decision support in crop micronutrient management. *Geoderma Regional*, 30, e00562. https://doi.org/10.1016/j.geodrs.2022.e00562

Ågren, A.M., Hasselquist, E.M., Stendahl, J., Nilsson, M.B. & Paul, S.S. (2022). Delineating the distribution of mineral and peat soils at the landscape scale in northern boreal regions. *SOIL*, 8 (2), 733–749. https://doi.org/10.5194/soil-8-733-2022

Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-de-Forges, A.C., Morgan, C.L.S., Roudier, P., Poggio, L. & Mulder, V.L. (2020). Impressions of digital soil maps: The good, the not so good, and making them ever better. *Geoderma Regional*, 20, e00255. https://doi.org/10.1016/j.geodrs.2020.e00255

Ballard, C.E., McIntyre, N., Wheater, H.S., Holden, J. & Wallage, Z.E. (2011). Hydrological modelling of drained blanket peatland. *Journal of Hydrology*, 407 (1), 81–93. https://doi.org/10.1016/j.jhydrol.2011.07.005

Behrens, T., Schmidt, K. & Scholten, T. (2008). An Approach to Removing Uncertainties in Nominal Environmental Covariates and Soil Class Maps. In: Hartemink, A.E., McBratney, A., & Mendonça-Santos, M. de L. (eds) *Digital Soil Mapping with Limited Data*. Springer Netherlands. 213–224. https://doi.org/10.1007/978-1-4020-8592-5_18

Benn, D. & Evans, D.J.A. (2014). *Glaciers and Glaciation, 2nd edition*. 2. ed. Routledge. https://doi.org/10.4324/9780203785010

Bernknopf, R., Brookshire, D., Sollar, D., Mckee, M., Sutter, J., Matti, J. & Campbell, R. (1983). *Societal Value of Geologic Maps*.

Beucher, A.M., Weber, P.L., Hermansen, C., Pesch, C., Koganti, T., Møller, A.B., Gomes, L., Greve, M.B. & Greve, M.H. (2023). *Updating the Danish peatland maps with a combination of new data and modeling approaches*. DCA - Nationalt Center for Fødevarer og Jordbrug.

Bishop, T.F.A., Horta, A. & Karunaratne, S.B. (2015). Validation of digital soil maps at different spatial supports. *Geoderma*, 241–242, 238–249. https://doi.org/10.1016/j.geoderma.2014.11.026

Björck, S. (1995). A review of the history of the Baltic Sea, 13.0-8.0 ka BP. *Quaternary International*, 27, 19–40. https://doi.org/10.1016/1040-6182(94)00057-C

van der Bles, A.M., van der Linden, S., Freeman, A.L.J. & Spiegelhalter, D.J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 117 (14), 7672–7683. https://doi.org/10.1073/pnas.1913678117

Bowen, D.Q. (2009). Last Glacial Maximum. In: *Encyclopedia of Paleoclimatology and Ancient Environments*. Springer, Dordrecht. 493–495. https://doi.org/10.1007/978-1-4020-4411-3_122

Breiman, L. (2001a). Random Forests. *Machine Learning*, 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L. (2001b). Statistical Modeling: The Two Cultures. *Statistical Science*, 16 (3), 199–215. http://www.jstor.org/stable/2676681

de Bruin, S., Brus, D.J., Heuvelink, G.B.M., van Ebbenhorst Tengbergen, T. & Wadoux, A.M.J.-C. (2022). Dealing with clustered samples for assessing map accuracy by cross-validation. *Ecological Informatics*, 69, 101665. https://doi.org/10.1016/j.ecoinf.2022.101665

Bui, E.N. & Moran, C.J. (2001). Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma*, 103 (1), 79–94. https://doi.org/10.1016/S0016-7061(01)00070-2

Chandler, B.M.P., Lovell, H., Boston, C.M., Lukas, S., Barr, I.D., Benediktsson, Í.Ö., Benn, D.I., Clark, C.D., Darvill, C.M., Evans, D.J.A., Ewertowski, M.W., Loibl, D., Margold, M., Otto, J.-C., Roberts, D.H., Stokes, C.R., Storrar, R.D. & Stroeven, A.P. (2018). Glacial geomorphological mapping: A review of approaches and frameworks for best practice. *Earth-Science Reviews*, 185, 806–846. https://doi.org/10.1016/j.earscirev.2018.07.015

Chen, T. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA, August 13 2016. 785–794. ACM. https://doi.org/10.1145/2939672.2939785

Clarke, B.G. (2017). *Engineering of Glacial Deposits*. CRC Press. https://doi.org/10.1201/9781315149356

Cohen, K., Finney, S., Gibbard, P. & Fan, J. (2013). Cohen, K.M., Finney, S.C., Gibbard, P.L., Fan, J.X. 2013. The ICS International Chronostratigraphic Chart. Episodes 36, 199-204. *Episodes*, 36. https://doi.org/10.18814/epiiugs/2013/v36i3/002

Congalton, R.G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37 (1), 35–46. https://doi.org/10.1016/0034-4257(91)90048-B

Courteille, L., Tardieu, L., Boukhelifa, N., Lutton, E. & Lagacherie, P. (2025). What is the best way to communicate the uncertainty of a digital soil mapping product? Some lessons from an end-users survey. *Geoderma*, 459, 117302. https://doi.org/10.1016/j.geoderma.2025.117302

Daniels, J. & Thunholm, B. (2014). *Rikstäckande jorddjupsmodell*. (2014:14). https://resource.sgu.se/produkter/sgurapp/s1414-rapport.pdf

Davis, S.E., Matheny, M.E., Balu, S. & Sendak, M.P. (2023). A framework for understanding label leakage in machine learning for health care. *Journal of the American Medical Informatics Association : JAMIA*, 31 (1), 274–280. https://doi.org/10.1093/jamia/ocad178

Dietterich, T.G. (2000). Ensemble Methods in Machine Learning., Berlin, Heidelberg, 2000. 1–15. Springer. https://doi.org/10.1007/3-540-45014-9_1

Dowling, T.P.F., Alexanderson, H. & Möller, P. (2013). The new high-resolution LiDAR digital height model ('Ny Nationell Höjdmodell') and its application to Swedish Quaternary geomorphology. *GFF*, 135 (2), 145–151. https://doi.org/10.1080/11035897.2012.759269

Fakour, F., Mosleh, A. & Ramezani, R. (2024). A Structured Review of Literature on Uncertainty in Machine Learning & Deep Learning. arXiv. https://doi.org/10.48550/arXiv.2406.00332

Fookes, P.G. (1997). Geology for Engineers: the Geological Model, Prediction and Performance. *Quarterly Journal of Engineering Geology*, 30 (4), 293–424. https://doi.org/10.1144/GSL.QJEG.1997.030.P4.02

Fredén, C. (1990). *National Atlas of Sweden: Geology*. SNA Publishing.

García Álvarez, D. (2017). Cartographic Scale and Minimum Mapping Unit Influence on LULC Modelling: *Proceedings of International Workshop on Geomatic Approaches for Modelling Land Change Scenarios*, Porto, Portugal, 2017. 327–334. SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0006383003270334

Gibbard, P.L., Head, M.J., Walker, M.J.C. & Stratigraphy, the S. on Q. (2010). Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. *Journal of Quaternary Science*, 25 (2), 96–102. https://doi.org/10.1002/jqs.1338

Greenwood, S.L. & Hughes, A.L.C. (2022). Chapter 5 - Glacial landscapes of Fennoscandia. In: Palacios, D., Hughes, P.D., García-Ruiz, J.M., & Andrés, N. (eds) *European Glacial Landscapes*. Elsevier. 37–43. https://doi.org/10.1016/B978-0-12-823498-3.00051-0

Griffiths, K.J., MacDonald, A.M., Robins, N.S., Merritt, J., Booth, S.J., Johnson, D. & McConvey, P.J. (2011). Improving the characterization of Quaternary deposits for groundwater vulnerability assessments using maps of recharge and attenuation potential. *Quarterly Journal of Engineering Geology and Hydrogeology*, 44 (1), 49–61. https://doi.org/10.1144/1470-9236/09-009

Guo, C., Pleiss, G., Sun, Y. & Weinberger, K.Q. (2017). On Calibration of Modern Neural Networks. *Proceedings of International Conference on Machine Learning*, July 17 2017. 1321–1330. PMLR. https://proceedings.mlr.press/v70/guo17a.html [2026-04-15]

Han, W., Zhang, X., Wang, Y., Wang, L., Huang, X., Li, J., Wang, S., Chen, W., Li, X., Feng, R., Fan, R., Zhang, X. & Wang, Y. (2023). A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities. *ISPRS Journal of Photogrammetry and Remote Sensing*, 202, 87–113. https://doi.org/10.1016/j.isprsjprs.2023.05.032

Han, Z., Zhang, C., Fu, H. & Zhou, J.T. (2022). Trusted Multi-View Classification with Dynamic Evidential Fusion. arXiv. https://doi.org/10.48550/arXiv.2204.11423

Hasan, M., Su, L., Cui, P. & Shang, Y. (2025). Development of deep-underground engineering structures via 2D and 3D RQD prediction using non-invasive CSAMT. *Scientific Reports*, 15 (1), 1403. https://doi.org/10.1038/s41598-025-85626-7

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer. https://doi.org/10.1007/978-0-387-84858-7

Heuvelink, G.B.M. (2018). Uncertainty and Uncertainty Propagation in Soil Mapping and Modelling. In: McBratney, Alex.B., Minasny, B., & Stockmann, U. (eds) *Pedometrics*. Springer International Publishing. 439–461. https://doi.org/10.1007/978-3-319-63439-5_14

Hughes, A.L.C., Gyllencreutz, R., Lohne, Ø.S., Mangerud, J. & Svendsen, J.I. (2016). The last Eurasian ice sheets – a chronological database and time-slice reconstruction, DATED-1. *Boreas*, 45 (1), 1–45. https://doi.org/10.1111/bor.12142

Hughes, P.D., Gibbard, P.L. & Ehlers, J. (2013). Timing of glaciation during the last glacial cycle: evaluating the concept of a global 'Last Glacial Maximum' (LGM). *Earth-Science Reviews*, 125, 171–198. https://doi.org/10.1016/j.earscirev.2013.07.003

Hüllermeier, E. & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning*, 110 (3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

Ilg, E., Cicek, O., Galesso, S., Klein, A., Makansi, O., Hutter, F. & Brox, T. (2018). Uncertainty Estimates and Multi-Hypotheses Networks for Optical Flow. *Proceedings of Proceedings of the European Conference on Computer Vision*

(ECCV), 2018. 652–667. https://openaccess.thecvf.com/content_ECCV_2018/html/Eddy_Ilg_Uncertainty_Estimates_and_ECCV_2018_paper.html [2026-01-29]
Jemeljanova, M., Kmoch, A. & Uuemaa, E. (2024). Adapting machine learning for environmental spatial data - A review. *Ecological Informatics*, 81, 102634. https://doi.org/10.1016/j.ecoinf.2024.102634
Jiang, S., Sweet, L., Blougouras, G., Brenning, A., Li, W., Reichstein, M., Denzler, J., Shangguan, W., Yu, G., Huang, F. & Zscheischler, J. (2024). How Interpretable Machine Learning Can Benefit Process Understanding in the Geosciences. *Earth's Future*, 12 (7), e2024EF004540. https://doi.org/10.1029/2024EF004540
Johnson, M.D., Fredin, O., Ojala, A.E.K. & Peterson, G. (2015). Unraveling Scandinavian geomorphology: the LiDAR revolution. *GFF*, 137 (4), 245–251. https://doi.org/10.1080/11035897.2015.1111410
Jones, A., Stolbovoy, V., Tarnocai, C., Broll, G., Spaargaren, O. & Montanarella, L. (2009). *Soil Atlas of the Northern Circumpolar Region*. https://doi.org/10.2788/95795
Karlson, M. & Bastviken, D. (2023). Multi-Source Mapping of Peatland Types Using Sentinel-1, Sentinel-2, and Terrain Derivatives—A Comparison Between Five High-Latitude Landscapes. *Journal of Geophysical Research: Biogeosciences*, 128 (4), e2022JG007195. https://doi.org/10.1029/2022JG007195
Karlsson, C., Sohlenius, G. & Becher, G.P. (2021). *Handledning för jordartsgeologiska kartor och databaser över Sverige*. (2021:17). The Geological Survey of Sweden.
Kaufman, S., Rosset, S., Perlich, C. & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Trans. Knowl. Discov. Data*, 6 (4), 15:1-15:21. https://doi.org/10.1145/2382577.2382579
Kinkeldey, C., MacEachren, A.M., Riveiro, M. & Schiewe, J. (2017). Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science*, 44 (1), 1–21. https://doi.org/10.1080/15230406.2015.1089792
Kleman, J., Stroeven, A.P. & Lundqvist, J. (2008). Patterns of Quaternary ice sheet erosion and deposition in Fennoscandia and a theoretical framework for explanation. *Geomorphology*, 97 (1), 73–90. https://doi.org/10.1016/j.geomorph.2007.02.049
Kohonen, J. & Tarvainen, T. (2021). *Developments in map datamanagement and geological unit nomenclature in Finland*. GTK. https://doi.org/10.30440/bt412
Korporaal, M., Ruginski, I.T. & Fabrikant, S.I. (2020). Effects of Uncertainty Visualization on Map-Based Decision Making Under Time Pressure. *Frontiers in Computer Science*, 2. https://doi.org/10.3389/fcomp.2020.00032
Kristoffersson Lind, S., Xiong, Z., Forssén, P.-E. & Krüger, V. (2024). Uncertainty quantification metrics for deep regression. *Pattern Recognition Letters*, 186, 91–97. https://doi.org/10.1016/j.patrec.2024.09.011
Lambeck, K., Smither, C. & Johnston, P. (1998). Sea-level change, glacial rebound and mantle viscosity for northern Europe. *Geophysical Journal International*, 134 (1), 102–144. https://doi.org/10.1046/j.1365-246x.1998.00541.x
Lark, R.M., Chagumaira, C. & Milne, A.E. (2022). Decisions, uncertainty and spatial information. *Spatial Statistics*, 50, 100619. https://doi.org/10.1016/j.spasta.2022.100619
Latifovic, R., Pouliot, D. & Campbell, J. (2018). Assessment of Convolution Neural Networks for Surficial Geology Mapping in the South Rae Geological Region, Northwest Territories, Canada. *Remote Sensing*, 10 (2), 307. https://doi.org/10.3390/rs10020307

Laudon, H., Järveoja, J., Ågren, A., Peichl, M. & Lindgren, A. (2025). Rewetting drained forested peatlands: A cornerstone of Sweden's climate change mitigation strategy. *Ambio*, 54. https://doi.org/10.1007/s13280-025-02220-x

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J. & Bretagnolle, V. (2014). Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. *Global Ecology and Biogeography*, 23 (7), 811–820. https://doi.org/10.1111/geb.12161

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, 521 (7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Kavukcuoglu, K. & Farabet, C. (2010). Convolutional networks and applications in vision. *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, May 2010. 253–256. https://doi.org/10.1109/ISCAS.2010.5537907

Li, W., Chen, W., Zhang, J., Li, C. & Ma, X. (2025). Geological object recognition in legacy maps through data augmentation and transfer learning techniques. *Applied Computing and Geosciences*, 25, 100233. https://doi.org/10.1016/j.acags.2025.100233

Linnenbrink, J., Milà, C., Ludwig, M. & Meyer, H. (2024). kNNDM CV: *k*-fold nearest-neighbour distance matching cross-validation for map accuracy estimation. *Geoscientific Model Development*, 17 (15), 5897–5912. https://doi.org/10.5194/gmd-17-5897-2024

Liu, X., Zhu, A.-X., Yang, L., Pei, T., Qi, F., Liu, J., Wang, D., Zeng, C. & Ma, T. (2022). Influence of legacy soil map accuracy on soil map updating with data mining methods. *Geoderma*, 416, 115802. https://doi.org/10.1016/j.geoderma.2022.115802

Lukas, S., Preusser, F., Evans, D.J.A., Boston, C.M. & Lovell, H. (2017). Chapter 2 The Quaternary. *Geological Society, London, Engineering Geology Special Publications*, 28 (1), 31–57. https://doi.org/10.1144/EGSP28.2

Lundberg, S. & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. arXiv. https://doi.org/10.48550/arXiv.1705.07874

Lundberg, S.M., Erion, G.G. & Lee, S.-I. (2019). Consistent Individualized Feature Attribution for Tree Ensembles. arXiv. https://doi.org/10.48550/arXiv.1802.03888

Lyu, F. (2025). Evaluating the Evaluation Matrices: Integrating Spatial Assessment in Geospatial AI Model Training and Evaluation. https://hdl.handle.net/10919/136885 [2026-02-24]

MacEachren, A.M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M. & Hetzler, E. (2005). Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32 (3), 139–160. https://doi.org/10.1559/1523040054738936

Malone, B.P., Minasny, B. & McBratney, A.B. (2017). Using Digital Soil Mapping to Update, Harmonize and Disaggregate Legacy Soil Maps. In: Malone, B.P., Minasny, B., & McBratney, A.B. (eds) *Using R for Digital Soil Mapping*. Springer International Publishing. 221–230. https://doi.org/10.1007/978-3-319-44327-0_8

Maxwell, A.E. & Shobe, C.M. (2022). Land-surface parameters for spatial predictive mapping and modeling. *Earth-Science Reviews*, 226, 103944. https://doi.org/10.1016/j.earscirev.2022.103944

McBratney, Alex.B. & Hartemink, A.E. (2024). Define soil. *Soil Security*, 14, 100135. https://doi.org/10.1016/j.soisec.2024.100135

Mclaren, J. & Turkington, R. (2013). Boreal Forest Ecosystems. In: *Encyclopedia of Biodiversity*. 626–635. https://doi.org/10.1016/B978-0-12-384719-5.00260-4

McMartin, I., Kerr, D., McClenaghan, M.B., Duk-Rodkin, A., Tremblay, T., Parent, M. & Rice, J. (2023). Introduction and summary; Geological Survey of Canada Bulletin 611. 1–6. https://doi.org/10.4095/331419

McMillan, A.A. (2002). Onshore Quaternary geological surveys in the 21st century—a perspective from the British Geological Survey. *Quaternary Science Reviews*, 21 (8), 889–899. https://doi.org/10.1016/S0277-3791(01)00064-6

McMillan, A.A. & Powell, J.H. (1999). *BGS Rock Classification Sheme, Volume 4: Classification of artificial (man-made) ground and natural superficial deposits—applications to geological map and datasets in the UK. British Geological Survey Research Report, RR 99-04. Classification of artificial (man-made) ground and natural superficial deposits — applications to geological maps and datasets in the UK*. https://nora.nerc.ac.uk/id/eprint/3228/1/RR99004.pdf [2026-02-08]

van der Meij, W.M., Meijles, E.W., Marcos, D., Harkema, T.T.L., Candel, J.H.J. & Maas, G.J. (2022). Comparing geomorphological maps made manually and by deep learning. *Earth Surface Processes and Landforms*, 47 (4), 1089–1107. https://doi.org/10.1002/esp.5305

Meinshausen, N. (2006). Quantile Regression Forests.

Meyer, H., Reudenbach, C., Wöllauer, S. & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, 411, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

Milà, C., Mateu, J., Pebesma, E. & Meyer, H. (2022). Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation. *Methods in Ecology and Evolution*, 13 (6), 1304–1316. https://doi.org/10.1111/2041-210X.13851

Milne, G.A., Davis, J.L., Mitrovica, J.X., Scherneck, H.-G., Johansson, J.M., Vermeer, M. & Koivula, H. (2001). Space-Geodetic Constraints on Glacial Isostatic Adjustment in Fennoscandia. *Science*, 291 (5512), 2381–2385. https://doi.org/10.1126/science.1057022

Minasny, B., Bandai, T., Ghezzehei, T.A., Huang, Y.-C., Ma, Y., McBratney, A.B., Ng, W., Norouzi, S., Padarian, J., Rudiyanto, Sharififar, A., Styc, Q. & Widyastuti, M. (2024). Soil Science-Informed Machine Learning. *Geoderma*, 452, 117094. https://doi.org/10.1016/j.geoderma.2024.117094

Mörner, N.-A. (1979). The Fennoscandian uplift and late cenozoic geodynamics: geological evidence. *GeoJournal*, 3 (3), 287–318. https://doi.org/10.1007/BF00177634

Nelson, M.A., Bishop, T.F.A., Triantafilis, J. & Odeh, I.O.A. (2011). An error budget for different sources of error in digital soil mapping. *European Journal of Soil Science*, 62 (3), 417–430. https://doi.org/10.1111/j.1365-2389.2011.01365.x

Nelson, M.D., Garner, J.D., Tavernia, B.G., Stehman, S.V., Riemann, R.I., Lister, A.J. & Perry, C.H. (2021). Assessing map accuracy from a suite of site-specific, non-site specific, and spatial distribution approaches. *Remote Sensing of Environment*, 260, 112442. https://doi.org/10.1016/j.rse.2021.112442

Nilsson, M., Ahlkrona, E. & Jönsson, C. (2020). Regionala jämförelser mellan Nationella Marktäckedata och fältdata från Riksskogstaxeringen och NILS.

Noumonvi, K.D., Havertz, N.H., Bohlin, J., van der Linden, S., Nilsson, M.B. & Peichl, M. (2025). HuHoLa: A novel Hummock-Hollow-Lawn mire microtopography modelling approach. *Ecological Modelling*, 508, 111212. https://doi.org/10.1016/j.ecolmodel.2025.111212

Ochsner, M., Fisher, R. & Palmqvist, C.-W. (2024). The impacts of weather on railway infrastructure in Sweden. *Sustainable and Resilient Infrastructure*, 9 (6), 582–598. https://doi.org/10.1080/23789689.2024.2340371

Odgers, N.P., Sun, W., McBratney, A.B., Minasny, B. & Clifford, D. (2014). Disaggregating and harmonising soil map units through resampled classification trees. *Geoderma*, 214–215, 91–100. https://doi.org/10.1016/j.geoderma.2013.09.024

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E. & Wulder, M.A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42–57. https://doi.org/10.1016/j.rse.2014.02.015

Padarian, J. & McBratney, A.B. (2023). QuadMap: Variable resolution maps to better represent spatial uncertainty. *Computers and Geosciences*, 181, 105480. https://doi.org/10.1016/j.cageo.2023.105480

Padarian, J., Minasny, B. & McBratney, A.B. (2019). Using deep learning for digital soil mapping. *SOIL*, 5 (1), 79–89. https://doi.org/10.5194/soil-5-79-2019

Pahlavan-Rad, M.R., Khormali, F., Toomanian, N., Brungard, C.W., Kiani, F., Komaki, C.B. & Bogaert, P. (2016). Legacy soil maps as a covariate in digital soil mapping: A case study from Northern Iran. *Geoderma*, 279, 141–148. https://doi.org/10.1016/j.geoderma.2016.05.014

Palmu, J.-P., Ojala, A., Virtasalo, J., Putkinen, N., Kohonen, J. & Sarala, P. (2021). Classification system of superficial (Quaternary) geological units in Finland. *Bulletin of the Geological Survey of Finland*, 412, 115–169. https://doi.org/10.30440/bt412.4

Patton, H., Hubbard, A., Andreassen, K., Winsborrow, M. & Stroeven, A.P. (2016). The build-up, configuration, and dynamical sensitivity of the Eurasian ice-sheet complex to Late Weichselian climatic and oceanic forcing. *Quaternary Science Reviews*, 153, 97–121. https://doi.org/10.1016/j.quascirev.2016.10.009

Pawley, S.M., Atkinson, L., Utting, D.J., Hartman, G.M.D. & Atkinson, N. (2024). Evaluating spatially enabled machine learning approaches to depth to bedrock mapping, Alberta, Canada. *PLOS ONE*, 19 (3), e0296881. https://doi.org/10.1371/journal.pone.0296881

Peterson, G. (2022). Inlandsisen, landhöjningen och landskapet: Ett geologiskt perspektiv. 213.

Piikki, K. & Söderström, M. (2019). Digital soil mapping of arable land in Sweden – Validation of performance at multiple scales. *Geoderma*, 352, 342–350. https://doi.org/10.1016/j.geoderma.2017.10.049

Piikki, K., Wetterlind, J., Söderström, M. & Stenberg, B. (2021). Perspectives on validation in digital soil mapping of continuous attributes—A review. *Soil Use and Management*, 37 (1), 7–21. https://doi.org/10.1111/sum.12694

Poutanen, M. & Steffen, H. (2015). Land uplift at Kvarken Archipelago / high coast UNESCO World Heritage area. *Geophysica*, 50, 49–64

Putkinen, N., Eyles, N., Putkinen, S., Ojala, A., Palmu, J.-P., Sarala, P., Väänänen, T., Räisänen, J., Saarelainen, J., Ahtonen, N., Rönty, H., Kiiskinen, A., Rauhaniemi, T. & Tervo, T. (2017). High-resolution LiDAR mapping of glacial landforms and ice stream lobes in Finland. *Bulletin of the Geological Society of Finland*, 89. https://doi.org/10.17741/bgsf/89.2.001

Rimondini, L., Gumbricht, T., Ahlström, A. & Hugelius, G. (2023). *Mapping of peatlands in the forested landscape of Sweden using LiDAR-based terrain indices*. ESSD – Land/Pedology. https://doi.org/10.5194/essd-2023-77

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F. & Dormann, C.F. (2017). Cross-validation strategies for data with

temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40 (8), 913–929. https://doi.org/10.1111/ecog.02881
Roering, J.J., Mackey, B.H., Marshall, J.A., Sweeney, K.E., Deligne, N.I., Booth, A.M., Handwerger, A.L. & Cerovski-Darriau, C. (2013). 'You are HERE': Connecting the dots with airborne lidar for geomorphic fieldwork. *Geomorphology*, 200, 172–183. https://doi.org/10.1016/j.geomorph.2013.04.009
Rolf, E. (2023). Evaluation Challenges for Geospatial ML. arXiv. https://doi.org/10.48550/arXiv.2303.18087
Ryo, M. & Rillig, M.C. (2017). Statistically reinforced machine learning for nonlinear patterns and variable interactions. *Ecosphere*, 8 (11), e01976. https://doi.org/10.1002/ecs2.1976
Salas-Romero, S., Malehmir, A., Snowball, I., Lougheed, B.C. & Hellqvist, M. (2016). Identifying landslide preconditions in Swedish quick clays—insights from integration of surface geophysical, core sample- and downhole property measurements. *Landslides*, 13 (5), 905–923. https://doi.org/10.1007/s10346-015-0633-y
Sasse, L., Nicolaisen-Sobesky, E., Dukart, J., Eickhoff, S.B., Götz, M., Hamdan, S., Komeyer, V., Kulkarni, A., Lahnakoski, J.M., Love, B.C., Raimondo, F. & Patil, K.R. (2025). Overview of leakage scenarios in supervised machine learning. *Journal of Big Data*, 12 (1), 135. https://doi.org/10.1186/s40537-025-01193-8
Saura, S. (2002). Effects of minimum mapping unit on land cover data spatial configuration and composition. *International Journal of Remote Sensing*, 23 (22), 4853–4880. https://doi.org/10.1080/01431160110114493
Schmidinger, J. & Heuvelink, G.B.M. (2023). Validation of uncertainty predictions in digital soil mapping. *Geoderma*, 437, 116585. https://doi.org/10.1016/j.geoderma.2023.116585
Sensoy, M., Kaplan, L. & Kandemir, M. (2018). Evidential Deep Learning to Quantify Classification Uncertainty., 2018. Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2018/hash/a981f2b708044d6fb4a71a1463242520-Abstract.html [2026-01-12]
Singh, G., Moncrieff, G., Venter, Z., Cawse-Nicholson, K., Slingsby, J. & Robinson, T.B. (2024). Uncertainty quantification for probabilistic machine learning in earth observation using conformal prediction. *Scientific Reports*, 14 (1), 16166. https://doi.org/10.1038/s41598-024-65954-w
Sohlenius, G., Rudmark, L. & Hedenström, A. (2003). Mapping of unconsolidated Quaternary deposits.
Spohn, M. & Stendahl, J. (2022). Carbon, nitrogen, and phosphorus stoichiometry of organic matter in Swedish forest soils and its relationship with climate, tree species, and soil texture. *Biogeosciences*, 19 (8), 2171–2186. https://doi.org/10.5194/bg-19-2171-2022
Srivastava, V., Lei, F., MacEachren, A.M. & Maciejewski, R. (2026). *The Impact of Uncertainty Visualization on Trust in Thematic Maps*. *arXiv.org*. https://doi.org/10.1145/3772318.3790743
Steffen, H. & Wu, P. (2011). Glacial isostatic adjustment in Fennoscandia—A review of data and modeling. *Journal of Geodynamics*, 52 (3), 169–204. https://doi.org/10.1016/j.jog.2011.03.002
Stroeven, A.P., Hättestrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., Goodfellow, B.W., Harbor, J.M., Jansen, J.D., Olsen, L., Caffee, M.W., Fink, D., Lundqvist, J., Rosqvist, G.C., Strömberg, B. & Jansson, K.N. (2016). Deglaciation of Fennoscandia. *Quaternary Science Reviews*, 147, 91–121. https://doi.org/10.1016/j.quascirev.2015.09.016

Szatmári, G. & Pásztor, L. (2025). The impact of direct and indirect digital soil mapping approaches on spatial uncertainty. *Geoderma*, 460, 117448. https://doi.org/10.1016/j.geoderma.2025.117448

Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A. & Rahman, A. (2020). Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sensing*, 12 (7). https://doi.org/10.3390/rs12071135

The Swedish Land Survey (2022). Quality description laser data

The Swedish National Forest Inventory (2021). Fieldwork Instructions. Riksinventeringen av skog. https://www.slu.se/globalassets/ew/org/centrb/rt/dokument/faltinst/nfi_fieldwork_instructions_eng.pdf [2024-02-08]

Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10 (2), 225–232. https://doi.org/10.1111/2041-210X.13107

Vestøl, O., Ågren, J., Steffen, H., Kierulf, H. & Tarasov, L. (2019). NKG2016LU: a new land uplift model for Fennoscandia and the Baltic Region. *Journal of Geodesy*, 93 (9), 1759–1779. https://doi.org/10.1007/s00190-019-01280-8

Von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., Walczak, M., Garcke, J., Bauckhage, C. & Schuecker, J. (2023). Informed Machine Learning - A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems. *IEEE Transactions on Knowledge and Data Engineering*, 35 (1), 614–633. https://doi.org/10.1109/TKDE.2021.3079836

Wadoux, A.M.J.-C., Heuvelink, G.B.M., de Bruin, S. & Brus, D.J. (2021). Spatial cross-validation is not the right way to evaluate map accuracy. *Ecological Modelling*, 457, 109692. https://doi.org/10.1016/j.ecolmodel.2021.109692

Wadoux, A.M.J.-C., Minasny, B. & McBratney, A.B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, 210, 103359. https://doi.org/10.1016/j.earscirev.2020.103359

Wang, Y., Khodadadzadeh, M. & Zurita-Milla, R. (2023). Spatial+: A new cross-validation method to evaluate geospatial machine learning models. *International Journal of Applied Earth Observation and Geoinformation*, 121. https://doi.org/10.1016/j.jag.2023.103364

Warner, T.A., Nellis, M.D., Foody, G.M., Stehman, S.V. & Foody, G.M. (2009). Accuracy Assessment. In: *The SAGE Handbook of Remote Sensing*. SAGE Publications, Inc. 297–309. https://doi.org/10.4135/9780857021052

Webster, T.L., Murphy, J.B., Gosse, J.C. & Spooner, I. (2006). The application of lidar-derived digital elevation model analysis to geological mapping: an example from the Fundy Basin, Nova Scotia, Canada. *Canadian Journal of Remote Sensing*, 32 (2), 173–193. https://doi.org/10.5589/m06-017

With, C., Löfroth, H., Bastani, M., Persson, L., Rodhe, L., Hedfors, J. & Schoning, K. (2022). A methodology for mapping of quick clay in Sweden. *Natural Hazards*, 112 (3), 2549–2576. https://doi.org/10.1007/s11069-022-05278-y

Wohlfarth, B., Björck, S., Funder, S., Houmark-Nielsen, M., Ingólfsson, Ó., Lunkka, J., Mangerud, J., Saarnisto, M. & Vorren, T. (2008). Quaternary of Norden. *Episodes*, 31. https://doi.org/10.18814/epiiugs/2008/v31i1/011

Yang, J., Wang, T., Bi, Y. & Li, Z. (2025). Harnessing expert knowledge and legacy data for digital soil mapping with no new field surveys. *Geoderma Regional*, 42, e00998. https://doi.org/10.1016/j.geodrs.2025.e00998

Ye, W., Jiang, L., Xie, E., Zheng, G., Ma, Y., Cao, X., Guo, D., Qi, D., He, Z., Tian, Y., Coffee, M., Zeng, Z., Li, S., Huang, K., Wang, Z., Rehg, J., Kautz, H. & Zhang, A. (2025). *The Clever Hans Mirage: A Comprehensive Survey on Spurious Correlations in Machine Learning*. https://doi.org/10.48550/arXiv.2402.12715

Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman and Hall/CRC. https://doi.org/10.1201/b12207

Zhu, J.-J., Yang, M. & Ren, Z.J. (2023). Machine Learning in Environmental Research: Common Pitfalls and Best Practices. *Environmental Science & Technology*, 57 (46), 17671–17689. https://doi.org/10.1021/acs.est.3c00026

Zuur, A.F., Ieno, E.N. & Elphick, C.S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1 (1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

Popular science summary

Where do we find stable ground for railroads and highways? Where should we drill for water? Which hillsides might collapse after heavy rain? The answers lie beneath our feet, in the layers of till, sand, clay, gravel, and peat that cover the bedrock across Sweden. These surface deposits were left behind by glaciers, rivers, and seas over millions of years, and they determine everything from where we can safely build to where drinking water is found. Knowing what lies beneath the surface matters. The Geological Survey of Sweden has been mapping these deposits for over a century, but detailed maps suitable for modern use exist for only about a third of the country. Completing the rest by hand would take an unthinkably long time. This thesis explores whether machine learning, combined with high-resolution airborne laser scanning data, can speed up this process.

The short answer is: partly. Some deposit types, such as peat and exposed bedrock, have distinctive shapes in terrain data and can be automatically identified with high reliability. Others, such as different types of sand and gravel, are too complex for a computer. The thesis also tackled another related question: how deep are the deposits? Depth matters for any digging or drilling operation, and our model predicts it well in shallow areas but struggles with deeper deposits. We also repurposed a model to classify, from old maps, what land use once existed, information that is otherwise locked away but could be crucial for understanding where nutrient-rich former farmland once was and for guiding restoration efforts. We need to be honest about what the machine learning models can and cannot do. Rather than presenting model-generated maps as finished products, the thesis advocates for transparency about where predictions are uncertain. The goal is not to replace geologists but to help them work more efficiently.

Populärvetenskaplig sammanfattning

Var hittar vi stabil mark för järnvägar och motorvägar? Var bör vi borra efter vatten? Vilka sluttningar riskerar att kollapsa efter kraftigt regn? Svaren finns under våra fötter, i lagren av morän, sand, lera, grus och torv som täcker berggrunden i Sverige. Dessa jordarter avsattes av glaciärer, älvar och hav under miljontals år, och de avgör allt från var vi kan bygga säkert till var dricksvattnet finns. Att veta vad som finns under ytan har stor betydelse. Sveriges geologiska undersökning har kartlagt dessa avlagringar i över hundra år, men detaljerade kartor som passar moderna behov finns bara för ungefär en tredjedel av landet. Att färdigställa resten för hand skulle ta oöverskådligt lång tid. Denna avhandling undersöker om maskininlärning i kombination med högupplöst laserdata från flygburen skanning kan påskynda processen.

Det korta svaret är: delvis. Vissa jordarter, som torv och blottat berg, har tydliga former i terrängdata och kan identifieras automatiskt med hög tillförlitlighet. Andra, som olika typer av sand och grus, är för svåra för en dator att skilja åt. Avhandlingen tog sig också an en besläktad fråga: hur djupa är jordlagren? Djupet spelar roll vid alla schakt- och borrningsarbeten, och vår modell ger bra prediktioner på grunda djup men har svårare med djupare avlagringar. Vi använde också en modell för att klassificera historisk markanvändning från gamla kartor, information som annars är inlåst i arkiv men som kan vara avgörande för att identifiera var näringsrik före detta jordbruksmark en gång fanns och för att vägleda restaureringsinsatser. Vi behöver vara ärliga med vad maskininlärningsmodellerna kan och inte kan göra. Istället för att presentera modellgenererade kartor som färdiga produkter förespråkar avhandlingen för transparens kring var prediktionerna är osäkra. Målet är inte att ersätta geologer utan att hjälpa dem att arbeta mer effektivt.

Acknowledgements

In April 2022, as Europe descended into chaos, starting a PhD felt questionable at best and downright pointless at worst. Four years on, things have only spiralled further. The fact that I spent the past years living out life as a *doktorand* in this corner of the world is, above all, a privilege, a fleeting one, but a rare combination that affords a period of guaranteed funding and stability. Remote as it is, Umeå has offered a kind of respite while the outside world continues its fever dream. The most formidable part of this PhD wasn't the actual PhDing but grappling with a persistent sense of detachment: temporally, from the study subject itself, as billions of years of Quaternary history were reduced to LiDAR snapshots and performance metrics; and spatially, from the world at large. It remains a struggle to find the right distance: close enough to stay human, yet far enough to function. Looking back over the 'research' contained in these pages, one must imagine the *doktorand* happy. The writing of this thesis was completed in haste near the printing deadline, but the thinking behind these acknowledgments began much earlier, as I have many people to be thankful for. Corny as it may be, I will preserve them here in the only part of the book that stands a chance of being read, fully aware that the future me will likely cringe and wish for an 'unsend' button.

This thesis rests upon the efforts of many generations that came before (that actually traversed the land): the field surveyors who collected samples, the geologists who tirelessly mapped the country. Without them, this project wouldn't have been fathomable (or fundable). I also thank the funding bodies that paid for absolutely everything: my salary, my travels, shiny gadgets, compute and storage, and the desk I wish I had sat at more often. My heartfelt thanks go to my supervisors, who forged a coherent enough idea and pitched it with a straight face. To the opponent and committee members, if you've made it this far: it's a relief to have this work examined. I hope it will be a good discussion—or at the very least confirm it wasn't just me going in circles in my own head.

I feel lucky to have had three supervisors who care, not just about the project, but about us as humans. **Anneli**, who somehow always found time for me. I admire your willingness to speak up on issues, and I will not forget

the lengths you went to on my behalf. Thank you for showing me how fun research can be, for reframing my concerns when I was my own worst critic, for the necessary nudges, and the freedom to wander off—or, as they say in academic lingo, to 'develop as an independent researcher'. **William**, thank you for defending my weekends, for the 'doggos of encouragement', for stepping up over these past months, and for time and again pulling me out of the holes I've dug myself into. I admire how you cut through the noise to focus on what truly matters—something I aspire to. Finally, **Florian**, the mythical creature I have only glimpsed through the glow of a screen: thank you for your sharp eyes and sharper mind, your attention to detail, and your ability to explain complicated concepts with such clarity. You never dismissed our outsider (mis)understandings but instead guided us through the technical quagmire with patience and grace. Thank you also for graciously shouldering the blame for my confusion when in fact I had just confused myself. I have only immense gratitude for all of you.

To **Gustaf**, **Cecilia**, and **Gustav** at SGU: thank you for patiently explaining the Quaternary lexicon to a complete newbie, from glacial to postglacial, from *lerig morän* to *moränlera*, all of which I initially thought were the same. Thank you for guiding me in the field. Having you make sense of machine-spewed outputs likely saved me from the embarrassment of publishing total nonsense. Thanks also to **Matthias**, my half-time opponent, for pointing out gaps in my understanding; to **Eliza** and **Maja**, the current and former PhD study directors, for always having our backs; to **Ulf** and **Malin** for helping me navigate admin and money headaches; and to **Lucas**, whose tips and tricks made my small venture into the HPC world so smooth that I briefly felt like I knew what I was doing.

Life in a small town means people overlap in all sorts of (in)convenient ways, and I quickly discovered that classification isn't just a challenge for the models—it is also a challenge for me, as I try to sort the people who made these four years into neat groups, only to find that the hyperplane doesn't know where to land. I am profoundly grateful to all of you who blurred the line between colleagues and friends. To many of you, I could offer thanks many times over, but to save both paper and time, I will do it once. Please know that the gratitude is multiplied.

To all the *senpais* who did not spare their hard-earned wisdom about these turbid waters called academia, to the post-COVID generation of PhDs I looked to for camaraderie, laughter, and proof that I wasn't the only one barely keeping it together, to the trusty belayers who made sure I didn't fall to my demise, and to everyone in the department who showed me what an academic bubble looks like in the best possible way—thank you for all your help, and for sharing a piece of your world (and your fika) with me. In a climate that feels increasingly divided and less welcoming, this department has remained warm and open. I do not take that for granted, and it matters more than I can say.

To my officemates: **Mariana**, thank you for walking this path with me as fellow 'DEI hires', for staying tuned in to everything, everywhere, all at once, for your inspiring *Lebenslust,* and your astonishing enthusiasm for my failed pottery. **Olivia**, thank you for all the fun activities, for noticing when I had disappeared for too long, for your empathy and your nonjudgmental ear when I needed it. **Ulrika**, I thoroughly enjoyed all the conversations we had and the random animal fun facts. **Lena**, for your candidness and the steady calm you bring. And **Alejandro,** for being uncannily perceptive, and the afternoon philosophical deep dives that somehow left us with less clarity than before. This has been fun!

I have also met many wonderful people along the way: **Barry**, my unlicensed therapist, thank you for tolerating my chaotic texts and rants, for the reality checks, and for the times when we commiserated over life's absurdities. **Ash**, the living proof that human communication can be reduced to cat stickers, the years we 'motivated' each other to go to the gym were mostly fun, but I think I got more gains from the conversations about our common and not-so-common experiences. **Antonia**, you are a compass; thank you for maxing out the PhD experience so I don't have to. **Magda**, thank you for the *utmaning* to climb harder, and for listening to my overly dramatic reconstructions of life events as I *utmanar mig*. **Daniel**, for being my primary source of neighbourhood and mushroom intelligence, and for the lunches, coffees and clayey late nights, all of which we spent trying very hard to avoid other souls. **Koffi**, thank you for showing up and coming through. Messy as it might be, it's comforting to have you on the same (likely sinking) boat. **Vincent**, for lowering my blood pressure while relentlessly raising everyone else's. **Francoco** and **Eli**, for your support, your kindness,

and your generosity, for always having your door open, and for the epic trips you brought me on—and Francoco specifically, for picking up your phone (wink). **Vicky**, the only character who remains until season 5, I hope I can keep coming to you for advice and feel-good energy. **Haeun**, for the perfectly timed baked goods. **Alex**, for running the best pizzeria in town. **Lei**, **Sijia**, **Ruirui**, **Chuxian**, and **Yu**, for allowing me to pop into your office whenever I felt like petting fish. **Gillian**, the **Marcuses**, **Johannes**, **Arvid**, **Nicolas**, **Lorenzo**, **Sami**, **Kohsuke**, **Lea**, **Amanda**, **Monica**, **Corinne**, **Fabio**, and many more: thank you all for being part of this journey. I have learned a great deal from every one of you.

To my friends not in Umeå: **Chloe**, **Hyesoo**, **Anima**, **Andrea**, **Patrizia**, and **Pueh**—thank you for reminding me that life exists outside this town, for receiving my *Notes from the North* with such amusement, and for taking me in whenever I needed an escape. I must also thank my parents, who couldn't quite fathom my life choices but kept their mouths shut for the most part—the ultimate form of support from Asian parents. Thank you for maintaining good health and financial stability so I could roam the world in this quasi-irresponsible bliss. And **Cristian**, for your patience, optimism, endless positive affirmations (and, not least, your European passport)—all of which I sorely lack. I think your existence moors me in ways neither of us fully realizes. To the cattos: I almost forgot the collateral damage. Almost.

In some way, adjusting to life in this town felt like moving through the famous stages of grief: doubt, despair, bargaining, and eventual acceptance—and even occasional joy. As I was writing this, something suddenly clicked: Umeå helped me concretize the idea of 小確幸, the everyday bliss that always felt abstract before. Here I learned the quiet art of contentment—the kind that comes from noticing small but tangible moments of joy tucked within the mundane. It’s a lesson I didn’t know I needed.

April 17, 2026, at my desk, 5 hours to the deadline.

I

Geoderma Regional 40 (2025) e00905

Contents lists available at ScienceDirect

Geoderma Regional

journal homepage: www.elsevier.com/locate/geodrs

Mapping soil parent materials in a previously glaciated landscape: Potential for a machine learning approach for detailed nationwide mapping

Yiqi Lin [a,*], William Lidberg [a], Cecilia Karlsson [b], Gustav Sohlenius [b], Florian Westphal [a,c], Johannes Larson [a], Anneli M. Ågren [a]

[a] Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Skogsmarksgränd 17, 901 83 Umeå, Sweden
[b] Geological Survey of Sweden, Villavägen 18, 752 36 Uppsala, Sweden
[c] Department of Computing, Jönköping University, Gjuterigatan 5, 553 18 Jönköping, Sweden

ARTICLE INFO

Keywords:
Digital soil mapping
Soil parent materials
Airborne laser scanning
Machine learning
Extreme gradient boosting

ABSTRACT

Reliable information on soil-forming parent materials is crucial for informed decision-making in infrastructure planning, land-use management, environmental assessments, and geohazard mitigation. In the northern landscapes previously affected by glacial processes, these parent materials are predominantly Quaternary deposits. This study explored the potential of machine learning to expedite soil parent material mapping in Sweden. Two Extreme Gradient Boosting models were trained, one using terrain and hydrological indices derived from Light Detection and Ranging data, and the other incorporating additional ancillary map data. Both models were trained on 29,588 soil observations and evaluated against a separate hold-out set of 3500 observations. As a baseline, the existing most detailed maps achieved a Matthews Correlation Coefficient of 0.36. The Extreme Gradient Boosting models achieved higher MCC values of 0.45 and 0.56, respectively. To understand spatial variations in model performance, the second model was evaluated across 28 physiographic regions in Sweden. The results revealed that model performance varied across regions and deposit types, with till and peat exhibiting better performance than sorted sediments. These findings underscore the need for region-specific analyses to optimize the application of machine learning in digital soil mapping.

1. Introduction

Parent materials (PMs) are the initial state of the soil system (Jenny, 1994) and have a major influence on soil properties, which in turn affect nutrient availability, hydrology, and land stability (Anderson, 1988; Richter et al., 2019). Understanding the distribution of PMs is crucial for land-use planning, infrastructure construction, resource exploration, and geohazard recognition, as they carry significant environmental and societal impacts (Bernknopf et al., 1983; McMillan, 2002; Häggquist and Söderholm, 2015). PM can be broadly categorized as either primary in-situ or secondary transported material, such as alluvium, colluvium, aeolian, or glacial deposits (Gray and Murphy, 1999). In temperate regions, most soils have developed on soft rocks, or on unconsolidated sediments formed during the Quaternary period (the past 2.6 million years) through processes like erosion and deposition (Anderson, 1988). During glacial and postglacial times, these processes have created a mosaic of deposits at the surface that are different from the underlying bedrock geology (Lawley and Smith, 2008; Heung et al., 2014). The poor representation of the near-surface materials in existing geological maps often provide an erroneous view of soil PMs, further limiting their usefulness for soil modeling (Lawley and Smith, 2008; Lemercier et al., 2012). Producing PM maps (or often referred to as Quaternary Deposit (QD) maps in previously glaciated regions) is a labor-intensive, iterative process that often involves the collation and synthesis of diverse data sources. This often includes digitizing hand-drawn maps, compiling published and unpublished cartography or literature, and refining and updating maps by integrating data from new geophysical or geochemical investigations, with nationwide coverage typically achieved by amalgamating maps from different regions across the country (Weerts et al., 2005; Berg, 2009; Salazar Rincón et al., 2019).

The advent of Light Detection and Ranging (LiDAR) technology has revolutionized geomorphological and geological mapping across Scandinavia and beyond (Webster et al., 2006; Roering et al., 2013; Johnson et al., 2015; Ganerød et al., 2023), largely because it facilitated access to high-resolution Digital Elevation Models (DEMs), which have proven invaluable for identifying landscape features (Eilertsen et al., 2015;

* Corresponding author.
E-mail address: yiqi.lin@slu.se (Y. Lin).

https://doi.org/10.1016/j.geodrs.2024.e00905
Received 20 September 2024; Received in revised form 7 December 2024; Accepted 8 December 2024
Available online 10 December 2024

Peterson et al., 2017; Goodship and Alexanderson, 2020). Prior to the nationwide LiDAR scanning by the Swedish Land Survey (Lantmäteriet) in 2009, the methods for national scale QD mapping by the Geological Survey of Sweden (SGU) had remained unchanged for at least 15 years (Dowling et al., 2013). Such mapping relied heavily on ground investigations, supplemented by stereoscopic 3D viewing of aerial photographs and other thematic maps. Due to the time and effort required for manual mapping, priority has been given to areas of significant societal interest. As a result, while maps at a scale of 1:100,000 are available for large parts of Sweden, the highest-quality 1:25,000 scale maps are concentrated in the densely populated southern urban centers and agriculturally important regions (The Geological Survey of Sweden, 2024a). The vast northern inland areas, in contrast, rely on lower-resolution 1:1000,000 scale maps that depict generalized landforms (Fig. 1a) (The Geological Survey of Sweden, 2024b). Although QD maps at scales smaller than 1:100,000 are not intended for detailed planning and analysis, they are often the only resource available for national-scale mapping efforts (e.g., National Land Cover Map by The Swedish Environmental Protection Agency, 2020, Soil Moisture Map by Ågren et al., 2021). The lack of systematically available, detailed maps compromises the reliability of products derived from them - a problem that is not unique to Sweden (Lawley and Smith, 2008). Currently, at SGU, producing a 1:50,000 scale map for an area of 625 km^2 typically requires 250–350 h of work. This includes interpreting LiDAR and other remotely sensed data, conducting fieldwork, and compiling the maps. Given this approach, mapping the remaining areas of Sweden would take an estimated 50–100 years (personal communication, Cecilia Karlsson, Dec. 2024).

Similar to conventional soil mapping, in which soil surveyors or pedologists seek to establish conceptual relationships between soils and the landscape (Hudson, 1992), Digital Soil Mapping (DSM) techniques generalize and formalize soil classes or properties with spatially referenced environmental covariates for quantitative prediction (McBratney et al., 2003). Globally, DSM has transitioned from the academic domain to the operational stage (Arrouays et al., 2017; Zhang et al., 2017; Richer-de-forges et al., 2022). Advances in computing power and data storage, as well as the increasing availability of environmental datasets and regional soil databases, have fueled the rapid adoption of Machine Learning (ML) in DSM (Padarian et al., 2020; Wadoux et al., 2020). ML algorithms can handle complex relationships between large numbers of cross-correlated predictor variables and the target variable, a challenge for traditional geostatistical methods (Wadoux et al., 2020). Though DSM research covers a vast array of topics, its application to categorical mapping is limited (Wadoux et al., 2020). Recent developments have begun to expand this research area of multiclass mapping of soils over large areas (Ramcharan et al., 2018; Heung et al., 2022; Minarik et al., 2024, preprint). However, existing multiclass mapping studies of soil PMs are still often restricted to relatively small geographical extents (Lemercier et al., 2012; Latifovic et al., 2018), rely on limited field samples (Bonfatti et al., 2020; Prince et al., 2020; Krutskikh, 2022), or utilize synthetic training data (so-called 'pseudopoints') generated from polygon maps to improve spatial resolution through map disaggregation (Heung et al., 2014; Sorenson et al., 2023). In Sweden, DSM development has largely remained within the research realm, with studies typically conducted in specific regions (Söderström et al., 2016; Becher et al., 2019; Piikki and Söderström, 2019; Adler et al., 2022). Given the increasing demand for accurate and comprehensive QD maps in Sweden, especially for two-thirds of the country currently lacking detailed maps, there has been a growing interest in integrating DSM and ML techniques. In this context, our study is the first, to our knowledge, to explore the feasibility of generating wall-to-wall PM maps across a vast area (470,000 km^2) at a high resolution (4 m^2). Since most soil PMs in Sweden are QDs, we use the terms interchangeably to refer to the unconsolidated materials at the surface that are the foundation for eventual soil formation. However, we have retained the term 'QD maps' when directly referring to the official map products from the SGU.

In this study, we utilized data from 33,088 field observations obtained from extensive regional and national soil surveys for soil PM mapping in Sweden. Two separate Extreme Gradient Boosting (XGBoost) models were trained. The first model used exclusively terrain and hydrological indices derived from high-resolution airborne LiDAR data. A second model incorporating additional ancillary map data alongside the LiDAR-derived indices was trained and evaluated across Sweden's 28 physiographic regions. Our nationwide case study aimed to answer: (1) How do the ML-derived maps (using XGBoost) compare statistically to the existing QD maps? (2) Are certain PMs easier or more challenging to identify than others? (3) How do individual features impact the overall model performance, and the classification of individual PMs? and (4) Does the model performance exhibit spatial variation? Additionally, we provided a visual comparison between existing QD maps and the ML-derived maps, highlighting key differences and considerations for their application.

2. Material and methods

2.1. Study area

The study area spans most of Sweden, excluding the data-scarce mountain regions in the northwest (Fig. 1). Situated between 55° N and 70° N and 11° E and 25° E, Sweden covers an area of ca. 470,000 km^2. The country exhibits a diverse geomorphology, ranging from mountainous terrain in the northwest to lowland plains in the south. Located within the boreal and nemoboreal zone of northern Europe, Sweden's landscape is predominantly forested (68 %), followed by open mires and other open land uses (22 %) and arable land (7 %) (Statistics Sweden, 2019). Exposed bedrock constitutes a minor portion of the country's surface, while large parts of the bedrock are covered by Quaternary unconsolidated glacial deposits (glaciofluvium and till) and postglacial deposits (clay, silt, and peat), formed and shaped by glaciers, water, and to a lesser extent, wind (Donner and Donner, 2005). During the post-glacial times, the retreat of the glacier border triggered a period of rapid isostatic rebound and land uplift due to the release of pressure from the overlying ice cap (Karlsson et al., 2014). The central lowlands and areas around the coast that were submerged under the sea after glaciation are now partially covered by glacial and post-glacial clays (Agrell, 1979). With the land uplift, wave actions reworked till and glaciofluvial deposits in these areas to form littoral sand and gravel (Agrell, 1979) (Fig. 1b). The warmer humid climate, combined with a high groundwater table, created favorable conditions for peat formation (Franzén et al., 2012; Morris et al., 2018; Piilo et al., 2020). The country is divided into 29 distinct physiographic regions based on climatic, topographical, biological, and geomorphological factors (Helmfrid, 1996) (Fig. 1c).

2.2. Field data

For model training, we utilized field sampling data from (1) the Swedish Forest Soil Inventory (SFSI) (Fig. 1d), and (2) the Swedish Agricultural Soil Inventories (SASI) (Fig. 1e).

2.2.1. Swedish forest soil inventory (SFSI)

The Swedish Forest Soil Inventory (SFSI) consists of 19,650 georeferenced soil pits across 5 survey regions, excluding urban areas, cultivated land, water bodies, and high mountain regions (Fig. 1d). We utilized data collected during the sampling round from 2003 to 2012. PM and grain size were determined with a probe to collect soil samples from a minimum of 20 cm beneath the lower boundary of the humus layer. The pits were georeferenced with a GPS measurement error of up to 10 m. The SFSI classifies soil based on five PMs (well-sorted sediments, poorly sorted sediments, till, bedrock outcrops, and peat) and eight texture classes (cobble and stone (20–200 mm), gravel (2–20 mm), coarse-grained sand (0.6–2 mm), medium-grained sand (0.2–0.6 mm),

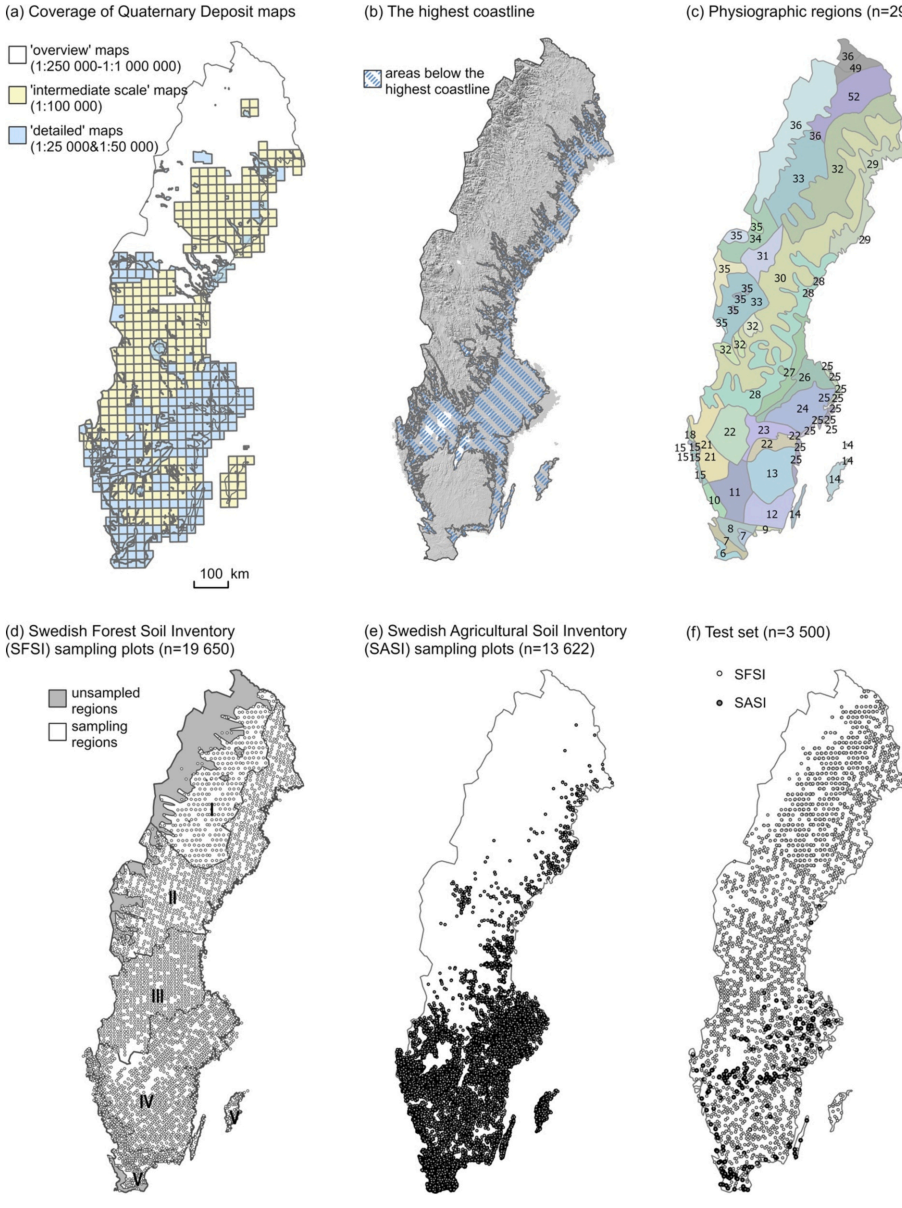


Fig. 1. Spatial data of Sweden. (a) Coverage of the Quaternary deposit (QD) maps by the Geological Survey of Sweden as of spring 2024, reclassified into three categories (source: the Geological Survey of Sweden). (b) The highest coastline is shown on a background map featuring a Digital Elevation Model (DEM). The blue hatched areas were submerged beneath the sea before land uplift following deglaciation (source: Geological Survey of Sweden). (c) Sweden is divided into 29 distinct physiographic regions based on geology, topography, and biology etc. See Fig. 5 for detailed legends (source: the National Atlas of Sweden). (d) Field sampling plots from the Swedish Forest Soil Inventory (SFSI). Sampling density increases from I to V. (e) Field sampling plots from the Swedish Agricultural Soil Inventory (SASI). Higher sampling density in the southern agricultural regions. (f) Distribution of the test set, constructed from (d) and (e).

fine-grained sand (0.06–0.2 mm), coarse silt (0.02–0.06 mm), fine silt (0.002–0.02 mm), clay (< 0.002 mm)) (The Swedish National Forest Inventory, 2021). Readers can refer to the SFSI sampling protocol for detailed instructions on PM and texture classification (The Swedish National Forest Inventory, 2021).

2.2.2. Swedish agricultural soil inventories (SASI)

Due to their similar characteristics, we combined topsoil samples (0–20 cm) from two national agricultural soil sampling campaigns: 12,599 topsoil samples (2011−2012) were collected by the Swedish Board of Agriculture (Jordbruksverket) from a more or less regular grid with a density of approximately 1 sample km^{-2} (Piikki and Söderström, 2019), and 5142 topsoil samples (1988–2017) collected by the Swedish Environmental Protection Agency (Naturvårdsverket) (Fig. 1e). Agricultural soil samples are concentrated in the south, reflecting the higher prevalence of agricultural land in that region. The determination of surface deposit types in our investigation was based on the percentage of sand, clay, and silt content. We classified fine sediments as those with a majority grain size ≤0.2 mm and coarse sediments as those with a majority grain size >0.2 mm. In total, the SASI dataset comprises 13,622 samples after data cleaning.

2.3. Data harmonization and partition

There are discrepancies in data collection methods and classification schemes between QD maps and field surveys. The SFSI and the original QD maps represent soil PMs at more comparable depths (SFSI at least 20 cm below the humus layer, QD maps 50 cm below the surface), while SASI represents topsoil up to 20 cm below the surface. To facilitate model training and comparison with the existing maps, a tailored harmonization process was devised collaboratively with the experts from the SGU where all 48 categories were reclassified into 7 more general soil PM classes: till, rock outcrops, peat, and two types of sorted sediments (coarse (fine sand-gravel) and fine (clay-silt)) from forest and agriculture land use, respectively. We defined sorted sediments as primarily transported and deposited by water, with a potential minor wind contribution. Fine sediments (clay-silt) were defined as having a grain size less than or equal to 0.2 mm, while coarse sediments (fine sand-gravel) were larger than 0.2 mm. Till, in contrast, was glacially deposited and generally unsorted. In Sweden, definitions of peat vary (at least 20 % organic material and minimum thickness of 30 cm, 40 cm, or 50 cm). In this study, we adopted the 50 cm threshold as defined by SGU (Ågren et al., 2022).

We created a test set with uniform sampling density across our study area (Fig. 1f), following simple random sampling principles (Brus et al., 2011) to generate a random subset from our original samples. This resulted in a test set totaling 3500 points (0.02 samples km^{-2}). The remaining 29,588 points were used for model training (0.06 samples km^{-2}). Further details regarding the harmonization procedure were provided in the Supplementary Material Table. A.1, and sample counts for all final classes in the training and evaluation sets were provided in the Supplementary Material Table. A.2.

2.4. Environmental covariates

2.4.1. The quaternary deposit (QD) maps from the Geological Survey of Sweden (SGU)

The QD maps from SGU provide a foundation for understanding Sweden's geology, surface deposit formation, and grain size composition (Karlsson et al., 2021). These maps depict the base layer (geology), the distribution of QDs 50 cm below the ground surface, and in some cases, thin/discontinuous surface layers between 0 and 50 cm at the ground surface. Specific landforms such as hummocky moraine and sand dunes are detailed with additional line and point layers (Karlsson et al., 2021). For this study, we classified the original QD map types into three categories based on their scale and cartographic methods. 'Detailed maps' (1:25,000 & 1:50,000) cover approximately 35.8 % of Sweden, primarily concentrated in southern urban centers and regions with agricultural importance (Fig. 1a). The 1:25,000 scale maps were created through a combination of manual interpretation of LiDAR DEMs and extensive field surveys. They are often an update from older 1: 50,000 maps. The 1: 50,000 scale maps were produced through extensive field surveys with manual interpretation of aerial photographs, topographic maps, and economic maps. 'Intermediate scaled maps' (1:100,000), covering 41.7 % of the country, were created by interpreting aerial photos and validated with less extensive field surveys. The remaining 22.5 % of the country, primarily in the northern inland, mostly rely on the 'overview maps' (between 1:250,000 and 1:1000,000) created partly based on aerial photograph interpretation with less intensive fieldwork, or digitizing analog, hard-copy maps(The Geological Survey of Sweden, 2018; 2024a; b; c; d). A nationwide QD map was provided in Supplementary Material Fig. A.1. This map was created by combining the highest-quality maps available for each region. For visualization on static media, we chose the version with simplified classes for readability without zooming possibilities. Readers can explore detailed, interactive maps on SGU's web map viewer at: https://apps.sgu.se/kartvisare/kartvisare-jordarter-25-100.html (for 'detailed' maps), or https://apps.sgu.se/kartvisare/kartvisare-jordarter-1-miljon.html (for 'overview' maps).

2.4.2. Environmental features for machine learning models

Two main sources of environmental features were used to train the ML models. Terrain and hydrological indices were derived from the original 2-m LiDAR DEM (Table 1). These features captured landforms and hydrological characteristics at various scales:

- 9 indices were derived directly from the high-resolution 2-m DEM for fine-scale details.
- To encompass larger landscape features, we resampled the 2 m DEM to 20 m and 50 m using the bilinear interpolation method to facilitate the calculation of an additional 13 and 4 indices, respectively. For detailed descriptions of these indices see the WhiteboxTools User Manual (Lindsay, 2024).

We further integrated ancillary data from existing maps and satellite imagery, including:

- Normalized Difference Vegetation Index (NDVI): Using cloud-free (<20 % cover) Sentinel-2 satellite imagery, we computed the median NDVI for the growing seasons (June 1 to September 30) from 2017 to 2023.
- Original Quaternary Deposit (QD) Map (The Geological Survey of Sweden, 2018; 2024a; b; c; d): This map was compiled from the best available QD maps in 2020 (scales vary from 1:25,000 to 1:1000, 000).
- Depth to Bedrock Map (Daniels and Thunholm, 2014): This map interpolates data from sources such as borehole drilling records, hydrogeological explorations, and seismic profiles. It uses the QD map as an input (assuming 0 m for rock outcrops).
- National Land Cover Map (The Swedish Environmental Protection Agency, 2020): The 2018 National Land Cover Map incorporates data from 41 sources, including the QD map, LiDAR DEM, Depth to Bedrock map, Sentinel-2 imagery, and other geospatial layers, resulting in a 10-m resolution map with 25 land cover classes (Minimum Mapping Unit (MMU): 0.01 ha).
- Distance to Highest Coastline (The Geological Survey of Sweden, 2015): This feature identifies the distance to a critical reference point – the highest coastline which indicates the maximum elevation reached by the sea after deglaciation and land rebound (Fig. 1b). Areas above this line are dominated by glacial deposits and peat, while areas below have more diverse water-deposited sediments. This feature is derived from data points in SGU and Geological Survey of Finland (GTK) databases.

Table 1
Feature layers used XGBoost model training. Each entry details the name, scale, specific settings in processing if applicable (such as search area or filter kernel size), data median, and range. All layers were resampled to 2 m resolution. Features indicated with an asterisk (*) were incorporated in the production of the National Land Cover Map. (a) LiDAR-derived indices. (b) Ancillary map data.

(a)

LiDAR derived digital terrain indices

Feature	Scale /Resolution	Specific Settings	Raster value median (range) /Number of classes	*Source*
Digital Elevation Model (DEM)*	2 m		110.94 (−2.07–863.65)	(The Swedish Land Survey, 2022)
Elevation above stream from a 1 ha stream network	2 m		2.0 (0.0–154.0)	(Lindsay, 2024)
Elevation above stream from a 10-ha stream network	2 m		4.0 (0.0–246.0)	Ibid.
Downslope index with 2 m drop	2 m		124.45 (2.0–31,250.0)	Ibid.
Downslope index with 2 m drop	20 m		0.07 (0.0–176.6)	Ibid.
Circular variance of aspect	2 m	3	0.05 (0.0–0.99)	Ibid.
Circular variance of aspect	20 m	5	0.21 (0.0–0.99)	Ibid.
Circular variance of aspect	50 m	11	0.28 (0.0–0.99)	Ibid.
Standard deviation of Slope	2 m	3	0.67 (0.02–21.93)	Ibid.
Deviation from mean elevation	2 m	7	−0.02 (−3.09–3.58)	Ibid.
Terrain ruggedness index	2 m		0.1 (0.0–4.14)	Ibid.
Average normal vector angular deviation	20 m	15	2.40 (0.11–56.0)	Ibid.
Multi-Scale roughness magnitude	2 m	Min_scale = 1, max_scale = 10, step = 2	1.98 (0.2–45.58)	Ibid.
Maximum Elevation deviation	2 m	Min_scale = 50, max_scale = 500, step = 50	−0.37 (−4.49–8.15)	Ibid.
Maximum Elevation deviation	20 m	Min_scale = 20, max_scale = 200, step = 11	0.0 (−5.55–3.42)	Ibid.
Maximum Elevation deviation	50 m	Min_scale = 50, max_scale = 500, step = 50	−0.49 (−4.92–7.8)	Ibid.
Max downslope elevation change	20 m		1.36 (0.21–3975.46)	Ibid.
Topographic Wetness Index	20 m		7.32 (−1.49–21.24)	Ibid.
Slope	20 m	20	2.58 (0.01–89.35)	Ibid.
Slope	50 m	50	1.53 (0.01–40.35)	Ibid.
Relative topographic positions	20 m	x = 11, y = 11	0.04 (−0.99–1.0)	Ibid.
Maximal curvature 20 m	20 m		0.0 (−0.02–0.19)	Ibid.
Maximal curvature 50 m	50 m		0.0 (−0.0–0.08)	Ibid.
Minimal curvature 20 m	20 m		−0.0 (−0.27–0.04)	Ibid.
Directional relief	20 m		278.43 (−44,738,810.1–1359.53)	Ibid.
Geomorphons	20 m	50	9 classes	Ibid.
Profile curvature	20 m		(−1.2–0.14)	Ibid.

(b)

Feature	Scale /Resolution	Specific Settings	Raster value median (range) /Number of classes	*Source*
Ancillary Map Data				
SGU Quaternary Deposit Map base layer up to 0.5 m below surface*	1:25,000 to 1: 1000,000		12 classes	(Karlsson et al., 2021)
National land cover map	10 m		23 classes	(The Swedish Environmental Protection Agency, 2020)
Age from deglaciation	?		11.51 (10.0–17.28)	(Hughes et al., 2016)
Distance to the highest coastline	1:100,000–1:200000		8.0 (0.0–255.0)	(The Geological Survey of Sweden, 2015)
Depth to bedrock*	10 m		7.43 (0–179.64)	(The Geological Survey of Sweden, 2023)
Normalized difference vegetation index (NDVI) (Sentinel 2) *	400 m		0.41 (−0.04–0.91)	
Coordinates				
Easting	2 m		568,679 (280795–913,007)	
Northing	2 m		6,920,018 (6141258–7,608,074)	

- Age from Deglaciation: Derived from DATED-1 time-slice reconstruction of the evolution of the extent of the Eurasian ice sheets 25–10 ka (Hughes et al., 2016).

All covariate data layers were resampled to a common resolution (2 m) and aligned to the national projection system (SWEREF99 TM). We utilized established open-source geospatial libraries for preprocessing data. Terrain and hydrological indices were calculated using WhiteboxTools (v2.3.0) (Lindsay, 2024). Other geospatial data manipulation, including resampling and raster value extraction, was facilitated by GDAL/OGR (v3.9.1) (GDAL/OGR contributors 2024) and Rasterio (v1.3.10) (Gillies, 2019) Python package. Sentinel-2 satellite imagery was accessed and processed through Google Earth Engine (GEE) (Gorelick et al., 2017) and its Python API geemap (v0.33.1) (Wu, 2020).

2.5. Extreme gradient boosting (XGBoost) model training

The Extreme Gradient Boosting (XGBoost) algorithm was employed

due to its efficiency in handling large datasets through optimized data structures and parallelization techniques (Chen and Guestrin, 2016). It is a tree ensemble model, but unlike Random Forest, which builds trees independently and derives the final prediction by majority voting (classification) or averaging (regression), XGBoost builds trees sequentially and corrects errors from the previous ones using gradient updates. It is less prone to overfitting due to built-in L1 and L2 regularization. For classification problems, the final prediction of XGBoost is assigned to the class with the highest probability. Its robustness has been demonstrated in national-scale applications in Sweden (Ågren et al., 2021; Ågren and Lin, 2024).

In this study, two XGBoost models were trained using the Python libraries scikit-learn (v1.5.1) (Pedregosa et al., 2011) and XGBoost (v2.1.1) (Chen and Guestrin, 2016). The first model utilized only LiDAR-derived indices as input features. The second model incorporated the same LiDAR features and additional maps (Table 1). To optimize hyperparameters, we employed the Tree-structured Parzen Estimator (TPE) sampler (Watanabe, 2023) implemented in Optuna (v3.6.1) (Akiba et al., 2019). The objective function was to maximize the mean Matthews Correlation Coefficient (MCC) (Matthews, 1975) across five-fold cross-validation on the training set ($n = 29{,}588$) over 100 trials. Each XGBoost model was trained with 100 fixed random states to ensure deterministic behavior and allow for a fair comparison of their performance. Throughout this study, all computations were performed on a Linux system (Ubuntu 22.04.2) with an AMD Ryzen ThreadRipper 3990× processor (64 cores, 2.9 GHz, 256 MB cache) and 256 GB of DDR4 RAM.

2.6. Model evaluation

2.6.1. Evaluation metrics

We evaluated model performance on the unseen hold-out set ($n = 3500$) using the following metrics: MCC (Matthews, 1975) for overall performance, and F1 score (harmonic mean of precision and recall) for class-specific performance. The precision, recall, and F1 scores of the best-performing XGBoost models (LiDAR-only and all features) based on MCC from 100 iterations were reported. Additionally, SHAP values were employed to identify the key features using the Python package SHAP (v 0.39.0) (Lundberg and Lee, 2017). A global measure was calculated by taking the mean of the absolute SHAP value for each feature across all instances in the hold-out test set. We computed MCC values across physiographic regions using the test set for the best-performing XGBoost model (all features). Only physiographic regions with a minimum of 80 field observations were selected to ensure sufficient data for reliable metric calculation. This approach provides a more nuanced understanding of the spatial variation in model performance and its applicability, particularly in regions lacking detailed QD map coverage.

2.6.2. Visual comparison of maps

To facilitate visual comparison, we focused on a peat-rich region in northern Sweden (64°23′N, 19°78′E). This area, well-known to the authors through decades of extensive research and established local knowledge (Laudon et al., 2013, 2021), provides a reliable basis for ground-truthing. Moreover, a high-quality QD map (1:25,000) produced by the SGU was available for this area. This map was created through extensive field investigations and the use of high-density LiDAR data (20 points m^{-2}), significantly exceeding the national average of 1–2 points m^{-2}. Four spatial representations of peat were presented: 1) an existing 'overview' QD map (1:1000,000), 2) a high-quality 'detailed' QD map (1:25,000), 3) an XGBoost (LiDAR only)-inferred map and 4) an XGBoost (all features)-inferred map. By displaying the 'overview' QD map, we aimed to exemplify the challenges faced by areas lacking high-resolution 'detailed' maps.

3. Results and discussion

3.1. Evaluation of existing QD maps and XGBoost model performance

To establish a baseline for assessing the performance of the XGBoost models, we evaluated the existing QD maps across scales using the hold-out test set ($n = 3500$) (Fig. 1f). On a country level, the 'overview' maps (1:250,000–1:1000,000) achieved a low level of agreement with the field data with an MCC of 0.18, followed by 'intermediate scale' maps (1:100,000) at 0.19. The highest agreement with the field plots was found with the 'detailed' maps (1:25,000–1:50,000) at an MCC of 0.36 (Table 2). Knight and Lunetta (2003) found a positive correlation between Minimum Mapping Units (MMU) size and accuracy metrics, up to the threshold of 6.4 ha, beyond which improvements plateaued. In contrast, our study revealed an opposite trend, with the evaluation metric MCC improving as MMU size decreased. However, a similar plateau effect was observed between 'intermediate-scale' maps (MMU: 2500–10,000 m^2) and 'overview' maps (MMU: 40000–1000,000 m^2) (Table 2) as the MMU sizes reached a comparable threshold to that identified in the aforementioned study (64,000 m^2). The trend in our study could be due to the fundamental differences in map-making methods and the classification details between different map types, as opposed to resampling raster maps to different MMUs in the study by Knight and Lunetta (2003). The 'detailed' maps were produced using rigorous methods, including interpreting LiDAR DEM and extensive field validation, enabled the identification of a more comprehensive range of surface deposit types. With smaller MMU of 625 m^2 for 1: 25,000 and 2500 m^2 for 1: 50,000 scales, the 'detailed' maps achieved average location errors (location of delineation between different QDs) between 25 and 100 m (Table 2). In contrast, 'overview' maps were created with a fixed display scale of 1:250,000 to 1:1000,000 where only large-scale QDs were identified. Their digitization was based on broader assumptions and generalizations, resulting in larger location errors (200 to 1000 m) (Table 2).

The XGBoost (LiDAR only) model achieved a mean MCC of 0.45 (Table 2), surpassing the MCC of 0.36 obtained by the 'detailed' maps. However, this result pertains to our 7-class ML model, which lacks the level of detail of the original 48-class QD maps (Supplementary Material Table. A.1). Although the relatively low MCC values of the QD maps raised concerns about potential error propagation into the ML process, the inclusion of them improved model performance, with XGBoost (all features) achieving an MCC of 0.56 (Table 2). This improvement can be attributed to the pedological expertise and insights captured in the original maps, in combination with the ability of ML techniques to filter out noise and capture underlying soil-environmental relationships (Teng et al., 2018; Liu et al., 2022). Another possible explanation could lie in the support (n), defined by volume, size, shape, and spatial orientation

Table 2
Evaluation of Quaternary Deposit Maps using Matthews Correlation Coefficient (MCC). Note that the number of data points within each map type may vary. A detailed breakdown of the precision, recall, F1 scores, and support (n) was provided in Supplementary Material Table. A.3.

Map type (scale)	MCC	n	Average location error (Karlsson et al., 2021)	Minimum mapping unit
'Overview' maps (1: 250,000–1: 1000,000)	0.18	934	200–1000 m	40,000–1000,000 m^2
'Intermediate scale' maps (1: 100,000)	0.19	1522	100–200 m	2500–10,000 m^2
'Detailed' maps (1: 25,000–1: 50,000)	0.36	987	25–100 m	625 & 2500 m^2
XGBoost (LiDAR only) (2 m)	0.45	3500	10 m	4 m^2
XGBoost (all features) (2 m)	0.56	3500	10–1000 m	4 m^2

(Gotway and Young, 2002). Point-based support usually presents the worst-case scenario, potentially leading to an underestimation of the true accuracy (Bishop et al., 2015). The actual quality of the QD maps might exceed the level indicated by this point-based evaluation. Nevertheless, spatially explicit comparisons between raster and vector datasets are challenging due to spatial and categorical mismatches (Wulder et al., 2006). Additionally, the GPS positioning errors of up to 10 m of the field samples can also undermine the reliability of the evaluation. Despite these uncertainties, the higher MCCs for the XGBoost models indicate the feasibility of using ML to automatically generate high-resolution future maps based on laser data and/or ancillary map data.

3.2. Visual comparison of maps

Waterbody polygons from the Swedish Property map (1:12,500) retrieved from the Swedish Land Survey were overlaid on all maps to aid visual interpretation. The limitation of using 'overview' maps for

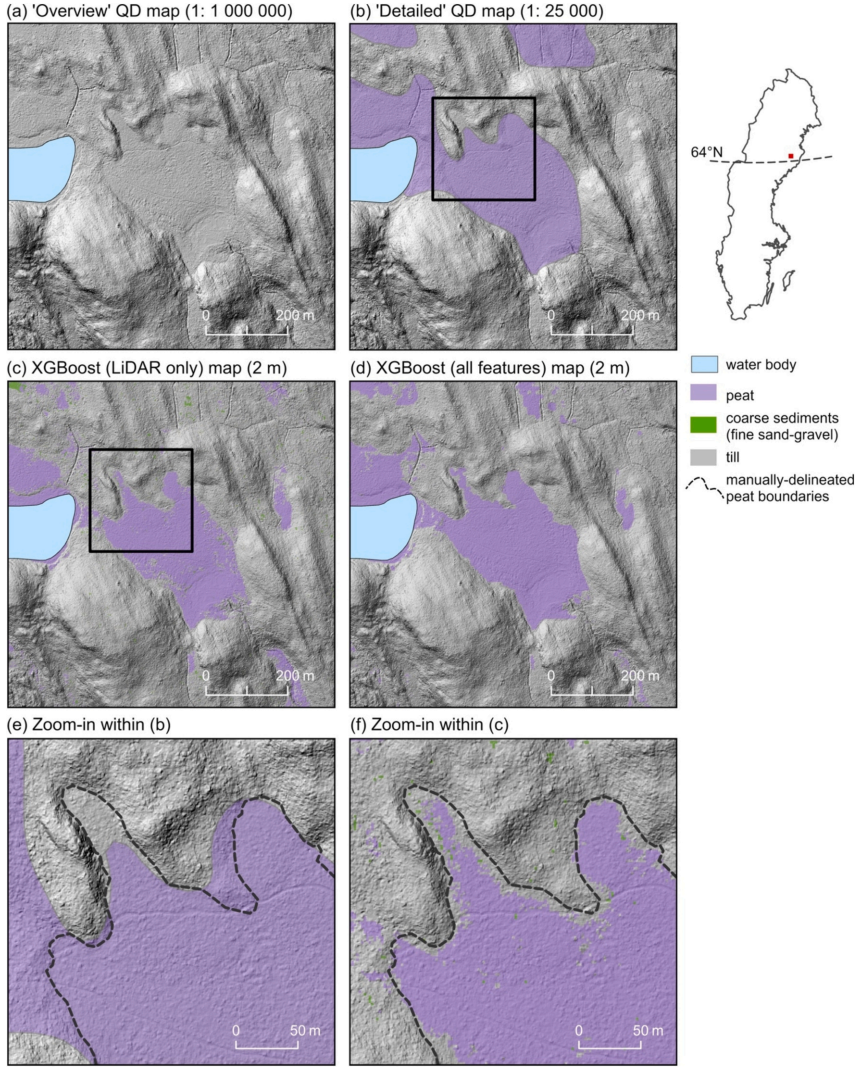


Fig. 2. Visual comparison of peat deposit distributions. (a) Quaternary Deposit (QD) map (1:1000,000) exemplifies the limitations of the 'overview' maps. (b) 'Detailed' QD map (1:25,000). (c) XGB (LiDAR only)-inferred map (2 m) and (d) XGB (all features)-inferred map (2 m). (e) and (f) zoom in to areas within panels (b) and (c), providing a closer look at the delineation of peat and till boundaries. The dashed lines in figures (e) and (f) indicate our interpreted border between till and peat based on the digital elevation model (DEM) and field visits.

detailed analysis is evident in Fig. 2a, where the peatland area was entirely omitted due to polygon map generalization. This process, commonly employed to enhance polygon map legibility at small scales, involves the elimination of certain small polygon features, smoothing of contours, and aggregation of closely spaced features (Smirnoff et al., 2012; Sayidov et al., 2020). While 'overview' QD maps (scales <1:100,000) are acceptable for national-level visualization, their use in other contexts such as GIS analyses, detailed planning, and decision-making requires caution (The Geological Survey of Sweden, 2024b). Conversely, the 'detailed' QD map (Fig. 2b), the XGBoost (LiDAR only) (Fig. 2c), and the XGBoost (all features) (Fig. 2d) exhibit similar spatial patterns for the peat extent. However, discrepancies were apparent when zooming in on specific areas within the 'detailed' QD map and the XGBoost (LiDAR only) map (Fig. 2e & f). The high-resolution raster maps also exhibit uncertainties, as demonstrated by the green pixels (coarse sorted sediments) in the XGBoost (LiDAR only) maps (Fig. 2c & f), known as the 'salt-and-pepper' effect (Blaschke et al., 2000, 2014; Whiteside et al., 2011). The isolated pixels are likely artifacts, as most landscape features need to be represented by areas larger than the size of an individual pixel (Whiteside et al., 2011). Various post-processing techniques to filter out noise could be applied to mitigate the visual clutter arising from undergeneralization in Fig. 2c-f. The harmonization process integrating materials from postglacial, glacial, and glaciofluvial processes (Supplementary Material Table A.1) further complicates the distinction between noise and actual landscape features based solely on model output. This also highlights the importance of expert knowledge when implementing DSM with ML during the model training process and for interpreting model outputs (Rossiter et al., 2022). The XGBoost (all features) map benefited from the high spatial resolution of a raster map, allowing for a detailed depiction of small landscape features, while incorporating existing QD maps during training prevented the generation of isolated pixels as observed in the XGBoost (LiDAR only) model. For practical purposes, the XGBoost (all features) might be the best model for application.

3.3. Predictive capability for different soil parent material classes

Our training data originates from two landscapes: forest (SFSI) and agricultural (SASI). To aid the visualization in Fig. 3, shapes were employed to represent 'detailed' QD maps (triangles), XGBoost (LiDAR-only) model (crosses), and XGBoost (all-features) model (circles). We used marker styles to differentiate PM types, with full markers representing PMs in forest land use, and half markers for those occurred in agricultural land use. Further color coding was applied (Fig. 3). Both XGBoost models excelled at classifying till, the dominant soil PM, achieving F1 scores of 0.82 (all features) and 0.80 (LiDAR only), while the 'detailed' QD maps achieved an F1 score of 0.60. This could be due to the slight discrepancies in the definition of till in the original QD maps and field sampling, where the QD maps can contain well-sorted sediments whereas field sampling defines till strictly as unsorted glacial deposits. XGBoost also demonstrated an improvement in peat classification, achieving F1 scores of 0.76 (all features) and 0.65 (LiDAR only) compared to 0.30 of 'detailed' QD maps (Fig. 3). This aligns with previous peat-specific DSM studies globally and in Sweden (Artz et al., 2019; DeLancey et al., 2019; Karlson et al., 2019; O'Leary et al., 2022; Rimondini et al., 2023). The multiclass nature of the XGBoost models offers more efficiency compared to managing multiple single-class models. The high-resolution models also have the potential to identify smaller or otherwise unmapped peatlands, which is valuable for peatland management and conservation efforts (Ågren et al., 2022).

Both XGBoost models struggled to differentiate between coarse and fine forest sorted sediments (Fig. 3), an issue noted in other multiclass DSM studies (Mulder et al., 2016; Ramcharan et al., 2018). Chen et al. (2022) highlighted the challenges in predicting coarse sediments due to their high spatial variability and recommended increasing field sampling as a potential strategy. McBratney et al. (2003) proposed

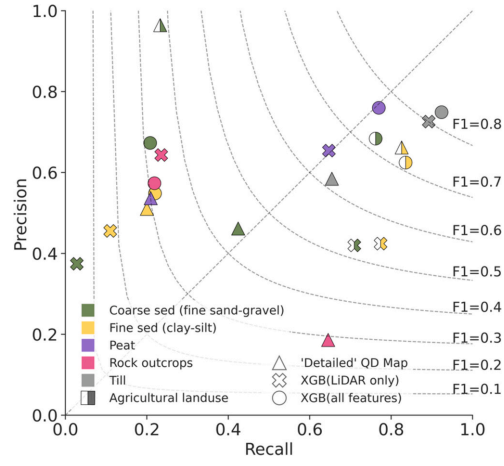


Fig. 3. The figure illustrates the performance of the 'detailed' Quaternary Deposit (QD) maps (triangles), XGBoost model trained with only LiDAR (crosses), and XGBoost model trained with all features (circles) for identifying different soil parent materials (PMs). Isolines represent F1 scores, the harmonic mean of precision (correctly identified) and recall (identification completeness). Half-filled markers denote PMs on agricultural land, while full markers represent forest land. Points under the 1:1 line are over-predicted and points above the line are under-predicted.

simplifying legends, reducing resolution, or conducting targeted sampling in regions with poor predictability. Similarly, Loiseau et al. (2021) recommended a threshold sampling density of about 1 sample 2 km^{-2} for clay, silt, and sand contents to avoid a substantial decrease in performance accuracy. Coarse and fine sorted sediments on agricultural lands, formed through simpler water deposition processes, had significantly more training samples (Supplementary Material Table. A.3). These soil PM types were effectively classified by XGBoost (all features), with F1 scores exceeding 0.72, which may suggest that increasing sampling density enhances model performance.

Rock outcrops on 'detailed' QD maps had high recall (0.63) but low precision (0.19) (Fig. 3, Supplementary Material Table. A.3), indicating frequent misclassification of other QDs as rock outcrops. This may be a result from the over-exaggeration of rock outcrop boundaries on polygon maps. In contrast, both XGBoost models demonstrated higher precision but lower recall, suggesting that while they often miss true rock outcrops, those identified as rock outcrops are typically accurate. This may be attributed to the limited availability of rock outcrop training samples (Supplementary Material Table A.2), as similarly observed with sorted sediments in forest landscapes. Additionally, the varying appearance of rock outcrops on LiDAR DEM could have presented further challenges. Several studies have highlighted the potential benefits of data resampling techniques, such as SMOTE (Synthetic Minority Oversampling Technique), to address imbalanced class distributions (Sharififar et al., 2019; Mirzaei et al., 2024). Thus, exploring class balancing techniques could be a promising strategy for enhancing model performance for minority classes.

3.4. Feature importance and the impact of LiDAR indices

SHAP values provided insights into the relative contributions of different features in the predictions of both XGBoost models (Fig. 4, showing only the top 15 most important features). The length of each bar in the SHAP plot represents the feature's overall importance as

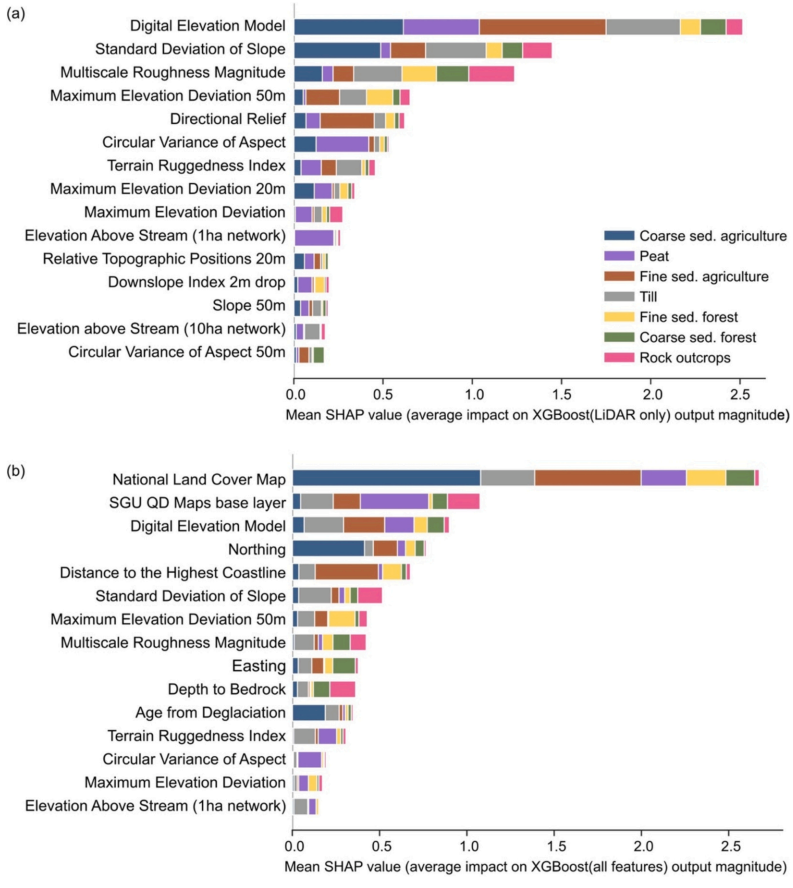


Fig. 4. SHAP (SHapley Additive exPlanations) values represent the relative importance of each feature in the XGBoost models' prediction of soil parent materials (PMs), with higher values indicating greater contribution. (a) SHAP summary plot for XGBoost (LiDAR only). (b) SHAP summary plot for XGBoost (all features).

determined by the model. In the XGBoost (LiDAR only) model, the original 2 m DEM was the most critical feature, particularly for identifying peat, till, and agricultural sediments (Fig. 4a). This finding underscores the capability of LiDAR in capturing high-resolution surface characteristics crucial for these classifications. The Circular Variation of Aspect (measures surface shape complexity, or texture) and Elevation above Stream (represents the relative vertical distance to drainage, measured along the downslope flow path) from a 1 ha stream network also contributed significantly to the prediction of peat (Fig. 4a). The characteristic flatness and poor drainage of peatlands were likely reflected in these indices, while a low flow accumulation threshold (1 ha) for Elevation above Stream captured areas with poor local drainage, which are favorable for peat development (Ehnvall et al., 2023). Similarly, Heung et al. (2014) found that aspect, distance to nearest stream, convergence index, and distance to nearest river to be the most important features for predicting parent material in the Canadian west coast, while slope, plan curvature, and profile curvature were the least important.

The most important feature for XGBoost (all features) to predict surface deposits was the National Land Cover Map (Fig. 4b). This map was crucial for identifying most surface deposit types, especially agricultural sorted sediments, likely due to the 'arable land' category within the map and the overrepresentation training data from the SASI dataset. For peat delineation, the original QD map was the most important (Fig. 4b), replacing Elevation above Stream as a key feature. The Distance to the Highest Coastline was crucial for identifying agricultural fine sediments, while Northing (latitude) and Age from Deglaciation significantly impacted the delineation of agricultural coarse sediments. These indices reflect the glacial history and post-glacial land rebound in the Bothnian Bay region (Stroeven et al., 2016; Nordman et al., 2020), which directly influenced the spatial distribution of sorted sediments. Rock outcrops were primarily identified based on Depth to Bedrock, which contains similar information as the original QD maps. The importance of Multiscale Roughness Magnitude diminished relative to other LiDAR-derived features (Fig. 4a) when additional ancillary map data were incorporated (Fig. 4b). This suggests that information from the QD Maps and Depth to Bedrock might have assumed the role previously played by Multiscale Roughness Magnitude in identifying rock outcrops for the XGBoost (LiDAR only) model.

Overall, conventional maps were more important than LiDAR-

derived indices in the XGBoost (all features) model. The inclusion of some conventional maps in the production of the National Land Cover map (marked with * in Table 1) may confound the interpretation of feature importance. Class imbalance within the training dataset can influence feature importance (Mirzaei et al., 2024). It is likely that applying resampling techniques to our training set, as discussed in Section 3.3, would have altered the feature importance rankings. Recent DSM studies found that though in general finer resolutions yield better performance in heterogeneous landscapes, and coarser resolution data is more effective in homogeneous landscapes; their effectiveness is highly dependent on the specific soil class or properties of interest, local environmental characteristics, and the geographical extent under consideration (Cavazzi et al., 2013; Larson et al., 2022; Piedallu et al., 2022). We calculated terrain and hydrological indices at different resolutions, examining the impact of scale on model performance, however, is beyond the scope of this study. A study conducted in eastern Canada using only four LiDAR-derived terrain indices achieved high overall performance in mapping parent materials, with Cohen's Kappa of 0.76 (Prince et al., 2020). Considering the recent trend towards utilizing high-dimensional and high-resolution data in DSM studies, it is imperative to conduct a meticulous feature selection process to ensure scale appropriateness and minimize model complexity.

3.5. Model performance across physiographic regions

This spatial analysis was only applied to XGBoost (all features) model due to its higher overall MCC. A spatial trend across Sweden's physiographic regions could be observed, where the northern inland regions (30–34), dominated by till and peat, generally achieved slightly higher MCCs compared to southern and coastal regions (11–29) (Fig. 5). As an exception, region 24 exhibited good model performance, which could be due to the high-quality QD map for this area. Till and fine-grained agricultural sediments were consistently overpredicted across the country (Fig. 5), likely owing to their overrepresentation in the training dataset (Supplementary Material Table A.2). In contrast, forest sediments were generally underpredicted. Peat and agricultural coarse sediments demonstrated more variable performance: peat was typically underpredicted in central Sweden (regions 24–26) but overpredicted in the north and south, while agricultural coarse sediments were overpredicted in the far south (regions 11–21) and underpredicted in south-central Sweden (regions 22–28). A closer examination of the class-wise training data distribution by region is necessary to determine if these patterns reflect data availability. Given that the model performs better in the northern inland areas, where only 'overview' and 'intermediate-scale' QD maps are available (Fig. 1a), there is great potential for using

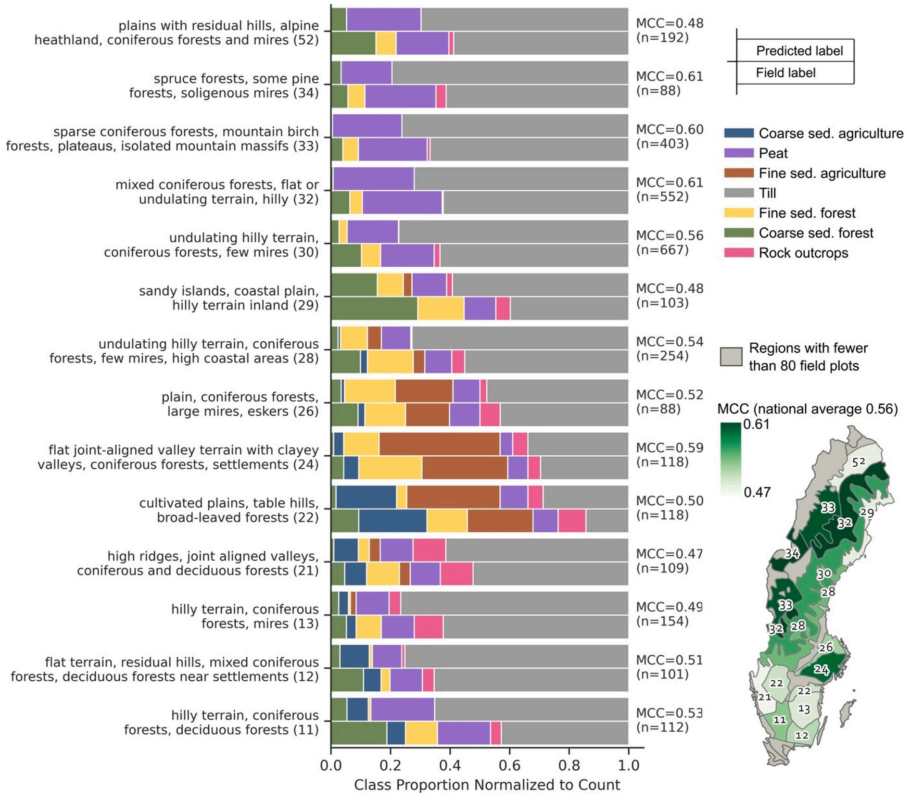


Fig. 5. Spatial variation in model performance across physiographic regions. Only regions with a minimum of 80 field observations were included in the analysis to ensure reliable Matthews Correlation Coefficient (MCC) calculation. The upper bars represent the predicted parent material (PM) class distribution for each region by the XGBoost (all features) model, while the lower bars represent the corresponding distribution of field-observed PM class. The proportion was scaled to sample counts to assist visual comparison.

the XGBoost (all features) model as a complementary tool to accelerate the mapping process in these northern regions.

With the increasing availability of legacy data and open soil databases, future studies are expected to rely more on these resources as a cost-efficient way of soil mapping (Hendriks et al., 2019). However, a key challenge in point-based DSM modeling remains: reconciling the pixel-based nature and the underlying pedogenetic and geomorphic processes of the actual soil landscapes (Rossiter et al., 2022). Although we incorporated features like Distance to the Highest Coastline to indirectly account for the glacial and postglacial processes, a more comprehensive approach may be warranted. Deep learning models, like Unets trained with raster image data may offer a promising avenue by incorporating the contextual information from neighboring cells and extracting multi-scale information automatically (Latifovic et al., 2018; Padarian et al., 2020; Taghizadeh-Mehrjardi et al., 2020), potentially leading to a more physically sound understanding of spatial prediction.

4. Conclusion

This study investigated the potential of Machine Learning (ML) for mapping soil parent materials (PMs) in Sweden, utilizing high-resolution airborne LiDAR data and ancillary maps. Our findings indicated that XGBoost models achieved higher Matthews Correlation Coefficients (MCCs) ranging from 0.45 to 0.56, in contrast to the existing polygon-based Quaternary Deposit (QD) maps, which showed MCCs between 0.18 and 0.36. The high-resolution maps generated by ML approach have the potential to identify previously unmapped smaller PM features. With adequate training data, the XGBoost algorithm successfully extracted valuable insights from the LiDAR data, effectively identifying till and peat. However, its performance decreased for PMs exhibiting high variability in grain size and composition. Regional analyses are crucial for capturing spatial variations in model performance and identifying areas where Digital Soil Mapping (DSM) can complement and enhance conventional mapping methods. Despite these advancements, expertise in soil-landscape relationships remains vital for interpreting model outputs and addressing potential discrepancies with landscape realities.

Funding sources

This work was supported by FORMAS Swedish Research Council for Sustainable Development [grant number 2021–00713, 2021–00115] and Knut and Alice Wallenberg Foundation [grant number 2018.0259 (Future Silviculture)] and partially supported by the Wallenberg AI, Autonomous Systems and Software Program—Humanities and Society (WASP-HS), funded by the Marianne and Marcus Wallenberg Foundation, the Marcus and Amalia Wallenberg Foundation, and Kempestiftelserna.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Christofer Engberg Hydén for providing the Swedish Forest Soil Inventory (SFSI) dataset and Johanna Wetterlind for the agricultural soil datasets (SASI).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geodrs.2024.e00905.

Data availability

Due to the sensitive nature of the Swedish Forest Soil Inventory plot geocoordinates, the exact data used in this study cannot be publicly shared. The Python codes used to generate the results will be available on GitHub:https://github.com/linyq017/Mapping-Soil-Parent-Materials-in-a-Previously-Glaciated-Landscape

References

Adler, K., Piikki, K., Söderström, M., Eriksson, J., 2022. Digital soil mapping of copper in Sweden: using the prediction and uncertainty as decision support in crop micronutrient management. Geoderma. Reg. 30, e00562. https://doi.org/10.1016/j.geodrs.2022.e00562.

Agrell, H., 1979. The Quaternary of Sweden. Univ. (Sveriges Geologiska Undersökning Serie C, Avhandlingar och uppsatser; 770).

Ågren, A.M., Lin, Y., 2024. A fully automated model for land use classification from historical maps using machine learning. Remote Sens. Appl. Soc. Environ. 36, 101349. https://doi.org/10.1016/j.rsase.2024.101349.

Ågren, A.M., Larson, J., Paul, S.S., Laudon, H., Lidberg, W., 2021. Use of multiple LIDAR-derived digital terrain indices and machine learning for high-resolution national-scale soil moisture mapping of the Swedish forest landscape. Geoderma 404, 115280. https://doi.org/10.1016/j.geoderma.2021.115280.

Ågren, A.M., Hasselquist, E.M., Stendahl, J., Nilsson, M.B., Paul, S.S., 2022. Delineating the distribution of mineral and peat soils at the landscape scale in northern boreal regions. SOIL 8 (2), 733–749. https://doi.org/10.5194/soil-8-733-2022.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. Association for Computing Machinery, New York, NY, USA, pp. 2623–2631. July 25 2019. https://doi.org/10.1145/3292500.3330701.

Anderson, D.W., 1988. The effect of parent material and soil development on nutrient cycling in temperate ecosystems. Biogeochemistry 5 (1), 71–97. https://doi.org/10.1007/BF02180318.

Arrouays, D., Lagacherie, P., Hartemink, A.E., 2017. Digital soil mapping across the globe. Geoderma. Reg. 9, 1–4. https://doi.org/10.1016/j.geodrs.2017.03.002.

Artz, R.R.E., Johnson, S., Bruneau, P., Britton, A.J., Mitchell, R.J., Ross, L., Donaldson-Selby, G., Donnelly, D., Aitkenhead, M.J., Gimona, A., Poggio, L., 2019. The potential for modelling peatland habitat condition in Scotland using long-term MODIS data. Sci. Total Environ. 660, 429–442. https://doi.org/10.1016/j.scitotenv.2018.12.327.

Becher, M., Sohlenius, G., Öhrling, C., 2019. Sur Sulfatjord - Egenskaper Och Utbredning.

Berg, R., 2009. Three-Dimensional Geological Mapping, Workshop Extended Abstracts. https://doi.org/10.13140/RG.2.1.1495.3041.

Bernknopf, R., Brookshire, D., Sollar, D., Mckee, M., Sutter, J., Matti, J., Campbell, R., 1983. Societal Value of Geologic Maps.

Bishop, T.F.A., Horta, A., Karunaratne, S.B., 2015. Validation of digital soil maps at different spatial supports. Geoderma 241–242, 238–249. https://doi.org/10.1016/j.geoderma.2014.11.026.

Blaschke, T., Lang, S., Lorup, E., Strobl, J., Zeil, P., 2000. Object-Oriented Image Processing in an Integrated GIS/Remote Sensing Environment and Perspectives for Environmental Applications.

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic object-based image analysis – towards a new paradigm. ISPRS J. Photogramm. Remote Sens. 87, 180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014.

Bonfatti, B.R., Demattê, J.A.M., Marques, K.P.P., Poppiel, R.R., Rizzo, R., de Mendes, W. S., Silvero, N.E.Q., Safanelli, J.L., 2020. Digital mapping of soil parent material in a heterogeneous tropical area. Geomorphology 367, 107305. https://doi.org/10.1016/j.geomorph.2020.107305.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62 (3), 394–407. https://doi.org/10.1111/j.1365-2389.2011.01364.x.

Cavazzi, S., Corstanje, R., Mayr, T., Hannam, J., Fealy, R., 2013. Are fine resolution digital elevation models always the best choice in digital soil mapping? Geoderma 195–196, 111–121. https://doi.org/10.1016/j.geoderma.2012.11.020.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. In: Proceedings of KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA, August 13 2016. ACM, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A. C., Walter, C., 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: a review. Geoderma 409, 115567. https://doi.org/10.1016/j.geoderma.2021.115567.

Daniels, J., Thunholm, B., 2014. Rikstäckande jorddjupsmodell. (2014:14). https://resource.sgu.se/produkter/sgurapp/s1414-rapport.pdf.

DeLancey, E.R., Kariyeva, J., Bried, J.T., Hird, J.N., 2019. Large-scale probabilistic identification of boreal peatlands using Google Earth Engine, open-access satellite data, and machine learning. PLoS One 14 (6), e0218165. https://doi.org/10.1371/journal.pone.0218165.

Donner, J., Donner, J.J., 2005. The Quaternary History of Scandinavia. Cambridge University Press.

Dowling, T.P.F., Alexanderson, H., Möller, P., 2013. The new high-resolution LiDAR digital height model ('Ny Nationell Höjdmodell') and its application to Swedish quaternary geomorphology. GFF 135 (2), 145–151. https://doi.org/10.1080/11035897.2012.759269.

Ehnvall, B., Ratcliffe, J.L., Bohlin, E., Nilsson, M.B., Öquist, M.G., Sponseller, R.A., Grabs, T., 2023. Landscape constraints on mire lateral expansion. Quat. Sci. Rev. 302, 107961. https://doi.org/10.1016/j.quascirev.2023.107961.

Eilertsen, R.S., Corner, G.D., Hansen, L., 2015. Using LiDAR data to characterize and distinguish among different types of raised terraces in a fjord-valley setting. GFF 137 (4), 353–361. https://doi.org/10.1080/11035897.2015.1111409.

Franzén, L., Lindberg, F., Viklander, V., Walther, A., 2012. The potential peatland extent and carbon sink in Sweden, as related to the Peatland / Ice Age Hypothesis. Mires Peat 10, 1–19.

Ganerød, A.J., Bakkestuen, V., Calovi, M., Fredin, O., Rød, J.K., 2023. Where are the outcrops? Automatic delineation of bedrock from sediments using deep-learning techniques. Appl. Comp. Geosci. 18, 100119. https://doi.org/10.1016/j.acags.2023.100119.

Gillies, S., et al.. Rasterio: Geospatial raster I/O for Python programmers. Mapbox. https://github.com/rasterio/rasterio.

Goodship, A., Alexanderson, H., 2020. Dynamics of a retreating ice sheet: a LiDAR study in Värmland, SW Sweden. GFF 142 (4), 325–345. https://doi.org/10.1080/11035897.2020.1822437.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Gotway, C.A., Young, L.J., 2002. Combining incompatible spatial data. J. Am. Stat. Assoc. 97 (458), 632–648. https://doi.org/10.1198/016214502760047140.

Gray, J., Murphy, B.W., 1999. Parent Material and Soils: A Guide to the Influence of Parent Material on Soil Distribution in Eastern Australia.

Häggquist, E., Söderholm, P., 2015. The economic value of geological information: synthesis and directions for future research. Res. Policy 43, 91–100. https://doi.org/10.1016/j.resourpol.2014.11.001.

Helmfrid, S., 1996. Geography of Sweden. SNA.

Hendriks, C.M.J., Stoorvogel, J.J., Lutz, F., Claessens, L., 2019. When can legacy soil data be used, and when should new data be collected instead? Geoderma 348, 181–188. https://doi.org/10.1016/j.geoderma.2019.04.026.

Heung, B., Bulmer, C.E., Schmidt, M.G., 2014. Predictive soil parent material mapping at a regional-scale: a random Forest approach. Geoderma 214–215, 141–154. https://doi.org/10.1016/j.geoderma.2013.09.016.

Heung, B., Bulmer, C.E., Schmidt, M.G., Zhang, J., 2022. Provincial-scale digital soil mapping using a random forest approach for British Columbia. Can. J. Soil Sci. 102 (3), 597–620. https://doi.org/10.1139/cjss-2021-0090.

Hudson, B.D., 1992. The soil survey as paradigm-based science. Soil Sci. Soc. Am. J. 56 (3), 836–841. https://doi.org/10.2136/sssaj1992.03615995005600030027x.

Hughes, A.L.C., Gyllencreutz, R., Lohne, Ø.S., Mangerud, J., Svendsen, J.I., 2016. The last Eurasian ice sheets – a chronological database and time-slice reconstruction, DATED-1. Boreas 45 (1), 1–45. https://doi.org/10.1111/bor.12142.

Jenny, H., 1994. Factors of Soil Formation: A System of Quantitative Pedology. Dover.

Johnson, M.D., Fredin, O., Ojala, A.E.K., Peterson, G., 2015. Unraveling Scandinavian geomorphology: the LiDAR revolution. GFF 137 (4), 245–251. https://doi.org/10.1080/11035897.2015.1111410.

Karlson, M., Gålfalk, M., Crill, P., Bousquet, P., Saunois, M., Bastviken, D., 2019. Delineating northern peatlands using Sentinel-1 time series and terrain indices from local and regional digital elevation models. Remote Sens. Environ. 231, 111252. https://doi.org/10.1016/j.rse.2019.111252.

Karlsson, C.S.J., Jamali, I.A., Earon, R., Olofsson, B., Mörtberg, U., 2014. Comparison of methods for predicting regolith thickness in previously glaciated terrain, Stockholm, Sweden. Geoderma 226–227, 116–129. https://doi.org/10.1016/j.geoderma.2014.03.003.

Karlsson, C., Sohlenius, G., Becher, G.P., 2021. Handledning för jordartsgeologiska kartor och databaser över Sverige. (2021:17). The Geological Survey of Sweden.

Knight, J.F., Lunetta, R.S., 2003. An experimental assessment of minimum mapping unit size. IEEE Trans. Geosci. Remote Sens. 41 (9), 2132–2134. https://doi.org/10.1109/TGRS.2003.816587.

Krutskikh, N., 2022. Mapping of the loose sediments of glacial and periglacial formations in areas with boreal vegetation using remote sensing. J. Appl. Remote. Sens. 16, 034528. https://doi.org/10.1117/1.JRS.16.034528].

Larson, J., Lidberg, W., Ågren, A.M., Laudon, H., 2022. Predicting soil moisture conditions across a heterogeneous boreal catchment using terrain indices. Hydrol. Earth Syst. Sci. 26 (19), 4837–4851. https://doi.org/10.5194/hess-26-4837-2022.

Latifovic, R., Pouliot, D., Campbell, J., 2018. Assessment of convolution neural networks for surficial geology mapping in the South Rae geological region, Northwest Territories, Canada. Remote Sens. 10 (2), 307. https://doi.org/10.3390/rs10020307.

Laudon, H., Taberman, I., Ågren, A., Futter, M., Ottosson-Löfvenius, M., Bishop, K., 2013. The Krycklan catchment study—a flagship infrastructure for hydrology, biogeochemistry, and climate research in the boreal landscape. Water Resour. Res. 49 (10), 7154–7158. https://doi.org/10.1002/wrcr.20520.

Laudon, H., Hasselquist, E.M., Peichl, M., Lindgren, K., Sponseller, R., Lidman, F., Kuglerová, L., Hasselquist, N.J., Bishop, K., Nilsson, M.B., Ågren, A.M., 2021. Northern landscapes in transition: evidence, approach and ways forward using the Krycklan catchment study. Hydrol. Process. 35 (4), e14170. https://doi.org/10.1002/hyp.14170.

Lawley, R., Smith, B., 2008. Digital soil mapping at a national scale: A knowledge and GIS based approach to improving parent material and property information. In: Hartemink, A.E., McBratney, A., de Mendonça-Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer, Netherlands, pp. 173–182. https://doi.org/10.1007/978-1-4020-8592-5_14.

Lemercier, B., Lacoste, M., Loum, M., Walter, C., 2012. Extrapolation at regional scale of local soil knowledge using boosted classification trees: a two-step approach. Geoderma 171–172, 75–84. https://doi.org/10.1016/j.geoderma.2011.03.010.

Lindsay, J., 2024. WhiteboxTools User Manual https://www.whiteboxgeo.com/manual/wbt_book/preface.html?search=2024-07-02.

Liu, X., Zhu, A.-X., Yang, L., Pei, T., Qi, F., Liu, J., Wang, D., Zeng, C., Ma, T., 2022. Influence of legacy soil map accuracy on soil map updating with data mining methods. Geoderma 416, 115802. https://doi.org/10.1016/j.geoderma.2022.115802.

Loiseau, T., Arrouays, D., Richer-de-Forges, A.C., Lagacherie, P., Ducommun, C., Minasny, B., 2021. Density of soil observations in digital soil mapping: a study in the Mayenne region, France. Geoderma Regional 24, e00358. https://doi.org/10.1016/j.geodrs.2021.e00358.

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. arXiv. https://doi.org/10.48550/arXiv.1705.07874.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta (BBA) - Protein Struct. 405 (2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

McMillan, A.A., 2002. Onshore quaternary geological surveys in the 21st century—a perspective from the British Geological Survey. Quat. Sci. Rev. 21 (8), 889–899. https://doi.org/10.1016/S0277-3791(01)00064-6.

Minarik, R., Hengl, T., Simoes, R., Isik, M.S., Ho, Y.-F., Tian, X., 2024. Soil type (world Reference Base) map of Europe based on ensemble machine learning and multiscale EO data. Res. Square. https://doi.org/10.21203/rs.3.rs-5244083/v1.

Mirzaei, F., Amirian-Chakan, A., Taghizadeh-Mehrjardi, R., Matinfar, H.R., Kerry, R., 2024. Soil textural class modeling using digital soil mapping approaches: effect of resampling strategies on imbalanced dataset predictions. Geoderma. Reg. 38, e00821. https://doi.org/10.1016/j.geodrs.2024.e00821.

Morris, P.J., Swindles, G.T., Valdes, P.J., Ivanovic, R.F., Gregoire, L.J., Smith, M.W., Tarasov, L., Haywood, A.M., Bacon, K.L., 2018. Global peatland initiation driven by regionally asynchronous warming. Proc. Natl. Acad. Sci. 115 (19), 4851–4856. https://doi.org/10.1073/pnas.1717838115.

Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Arrouays, D., 2016. GlobalSoilMap France: high-resolution spatial modelling the soils of France up to two meter depth. Sci. Total Environ. 573, 1352–1369. https://doi.org/10.1016/j.scitotenv.2016.07.066.

Nordman, M., Peltola, A., Bilker-Koivula, M., Lahtinen, S., 2020. Past and future Sea level changes and land uplift in the Baltic Sea seen by geodetic observations. In: Freymueller, J.T., Sánchez, L. (Eds.), Beyond 100: The Next Century in Geodesy. Springer International Publishing, pp. 161–167. https://doi.org/10.1007/1345_2020_124.

O'Leary, D., Brown, C., Daly, E., 2022. Digital soil mapping of peatland using airborne radiometric data and supervised machine learning – implication for the assessment of carbon stock. Geoderma 428, 116086. https://doi.org/10.1016/j.geoderma.2022.116086.

Padarian, J., Minasny, B., McBratney, A.B., 2020. Machine learning and soil sciences: a review aided by machine learning tools. SOIL 6 (1), 35–52. https://doi.org/10.5194/soil-6-35-2020.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Machine Learning Res. 12 (85), 2825–2830 [2024-08-15]. http://jmlr.org/papers/v12/pedregosa11a.html.

Peterson, G., Johnson, M.D., Smith, C.A., 2017. Glacial geomorphology of the south Swedish uplands – focus on the spatial distribution of hummock tracts. J. Maps 13 (2), 534–544. https://doi.org/10.1080/17445647.2017.1336121.

Piedallu, C., Pedersoli, E., Chaste, E., Morneau, F., Seynave, I., Gégout, J.-C., 2022. Optimal resolution of soil properties maps varies according to their geographical extent and location. Geoderma 412, 115723. https://doi.org/10.1016/j.geoderma.2022.115723.

Piikki, K., Söderström, M., 2019. Digital soil mapping of arable land in Sweden – validation of performance at multiple scales. Geoderma 352, 342–350. https://doi.org/10.1016/j.geoderma.2017.10.049.

Piilo, S.R., Korhola, A., Heiskanen, L., Tuovinen, J.-P., Aurela, M., Juutinen, S., Marttila, H., Saari, M., Tuittila, E.-S., Turunen, J., Väliranta, M.M., 2020. Spatially varying peatland initiation, Holocene development, carbon accumulation patterns and radiative forcing within a subarctic fen. Quat. Sci. Rev. 248, 106596. https://doi.org/10.1016/j.quascirev.2020.106596.

Prince, A., Franssen, J., Lapierre, J.-F., Maranger, R., 2020. High-resolution broad-scale mapping of soil parent material using object-based image analysis (OBIA) of LiDAR elevation data. CATENA 188, 104422. https://doi.org/10.1016/j.catena.2019.104422.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil property and class maps of the conterminous United States at 100-meter spatial resolution. Soil Sci. Soc. Am. J. 82 (1), 186–201. https://doi.org/10.2136/sssaj2017.04.0122.

Richer-de-forges, A.C., Arrouays, D., Poggio, L., Chen, S., Lacoste, M., Minasny, B., Libohova, Z., Roudier, P., Mulder, V.L., Nédélec, H., Martelet, G., Lemercier, B., Lagacherie, P., Bourennane, H., 2022. Hand-feel soil texture observations to evaluate the accuracy of digital soil maps for local prediction of particle size distribution. A case study in Central France. Pedosphere. https://doi.org/10.1016/j.pedsph.2022.07.009.

Richter, J., Owens, P., Libohova, Z., Adhikari, K., Fuentes, B., 2019. Mapping parent material as part of a nested approach to soil mapping in the Arkansas River Valley. Catena 178, 100–108. https://doi.org/10.1016/j.catena.2019.02.031.

Rimondini, L., Gumbricht, T., Ahlström, A., Hugelius, G., 2023. Mapping of peatlands in the forested landscape of Sweden using LiDAR-based terrain indices. ESSD – Land/Pedol. https://doi.org/10.5194/essd-2023-77.

Roering, J.J., Mackey, B.H., Marshall, J.A., Sweeney, K.E., Deligne, N.I., Booth, A.M., Handwerger, A.L., Cerovski-Darriau, C., 2013. 'You are HERE': connecting the dots with airborne lidar for geomorphic fieldwork. Geomorphology 200, 172–183. https://doi.org/10.1016/j.geomorph.2013.04.009.

Rossiter, D.G., Poggio, L., Beaudette, D., Libohova, Z., 2022. How well does digital soil mapping represent soil geography? An investigation from the USA. Soil 8 (2), 559–586. https://doi.org/10.5194/soil-8-559-2022.

Salazar Rincón, A.E., Rodríguez García, J.A., Cañas, V., Sevillano Matilla, A., Galindo Jiménez, I., Sánchez Jiménez, N., 2019. Quaternary geological mapping in Spain and contribution to the update of the International Quaternary Map of Europe https://digital.csic.es/handle/10261/2738132024-08-30.

Sayidov, A., Aliakbarian, M., Weibel, R., 2020. Geological map generalization driven by size constraints. ISPRS Int. J. Geo Inf. 9 (4), 284. https://doi.org/10.3390/ijgi9040284.

Sharififar, A., Sarmadian, F., Malone, B.P., Minasny, B., 2019. Addressing the issue of digital mapping of soil classes with imbalanced class observations. Geoderma 350, 84–92. https://doi.org/10.1016/j.geoderma.2019.05.016.

Smirnoff, A., Huot-Vézina, G., Paradis, S.J., Boivin, R., 2012. Generalizing geological maps with the GeoScaler software: the case study approach. Comput. Geosci. 40, 66–86. https://doi.org/10.1016/j.cageo.2011.07.013.

Söderström, M., Sohlenius, G., Rodhe, L., Piikki, K., 2016. Adaptation of regional digital soil mapping for precision agriculture. Precis. Agric. 17 (5), 588–607. https://doi.org/10.1007/s11119-016-9439-8.

Sorenson, P.T., Kiss, J., Bedard-Haughn, A.K., 2023. Improved parent material map disaggregation methods in the Saskatchewan prairies using historical bare soil composite imagery. Can. J. Soil Sci. 103 (1), 47–63. https://doi.org/10.1139/cjss-2021-0154.

Statistics Sweden, R. aand E.D, 2019. Markanvändningen i Sverige.

Stroeven, A.P., Hättestrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., Goodfellow, B.W., Harbor, J.M., Jansen, J.D., Olsen, L., Caffee, M.W., Fink, D., Lundqvist, J., Rosqvist, G.C., Strömberg, B., Jansson, K.N., 2016. Deglaciation of Fennoscandia. Quat. Sci. Rev. 147, 91–121. https://doi.org/10.1016/j.quascirev.2015.09.016.

Taghizadeh-Mehrjardi, R., Mahdianpari, M., Mohammadimanesh, F., Behrens, T., Toomanian, N., Scholten, T., Schmidt, K., 2020. Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in Central Iran. Geoderma 376, 114552. https://doi.org/10.1016/j.geoderma.2020.114552.

Teng, H., Viscarra Rossel, R.A., Shi, Z., Behrens, T., 2018. Updating a national soil classification with spectroscopic predictions and digital soil mapping. CATENA 164, 125–134. https://doi.org/10.1016/j.catena.2018.01.015.

The Geological Survey of Sweden, 2015. hogsta-kustlinjen-beskrivning.pdf. https://resource.sgu.se/dokument/produkter/hogsta-kustlinjen-beskrivning.pdf.

The Geological Survey of Sweden, 2018. jordarter-750000-mittnorden-beskrivning. https://resource.sgu.se/dokument/produkter/jordarter-750000-mittnorden-wms-beskrivning.pdf.

The Geological Survey of Sweden, 2023. jorddjupsmodell-beskrivning.pdf. https://resource.sgu.se/dokument/produkter/jorddjupsmodell-beskrivning.pdf [2023-11-30].

The Geological Survey of Sweden (2024a). jordarter-25-100000-beskrivning. https://resource.sgu.se/dokument/produkter/jordarter-25-100000-beskrivning.pdf.

The Geological Survey of Sweden (2024b). jordarter-1miljon-beskrivning. https://resource.sgu.se/dokument/produkter/jordarter-1miljon-beskrivning.pdf.

The Geological Survey of Sweden (2024c). jordarter-200000-vasternorrland-beskrivning. https://resource.sgu.se/dokument/produkter/jordarter-200000-vasternorrland-beskrivning.pdf.

The Geological Survey of Sweden (2024d). jordarter-250000-nordligaste-sverige-beskrivning. https://resource.sgu.se/dokument/produkter/jordarter-250000-nordligaste-sverige-beskrivning.pdf.

The Swedish Environmental Protection Agency, 2020. National Land Cover Database (NMD) - Product Description.

The Swedish Land Survey, 2022. Quality Description Laser Data.

The Swedish National Forest Inventory, 2021. Fieldwork Instructions. Riksinventeringen av skog. https://www.slu.se/globalassets/ew/org/centrb/rt/dokument/faltinst/nfi_fieldwork_instructions_eng.pdf.

Wadoux, A.M.J.-C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth Sci. Rev. 210, 103359. https://doi.org/10.1016/j.earscirev.2020.103359.

Watanabe, S., 2023. Tree-Structured Parzen Estimator: Understanding its Algorithm Components and Their Roles for Better Empirical Performance. arXiv. https://doi.org/10.48550/arXiv.2304.11127.

Webster, T.L., Murphy, J.B., Gosse, J.C., Spooner, I., 2006. The application of lidar-derived digital elevation model analysis to geological mapping: an example from the Fundy Basin, Nova Scotia, Canada. Can. J. Remote. Sens. 32 (2), 173–193. https://doi.org/10.5589/m06-017.

Weerts, H.J.T., Westerhoff, W.E., Cleveringa, P., Bierkens, M.F.P., Veldkamp, J.G., Rijsdijk, K.F., 2005. Quaternary geological mapping of the lowlands of the Netherlands, a 21st century perspective. Quat. Int. 133–134, 159–178. https://doi.org/10.1016/j.quaint.2004.10.011.

Whiteside, T.G., Boggs, G.S., Maier, S.W., 2011. Comparing object-based and pixel-based classifications for mapping savannas. Int. J. Appl. Earth Obs. Geoinf. 13 (6), 884–893. https://doi.org/10.1016/j.jag.2011.06.008.

Wu, Q., 2020. geemap: A Python package for interactive mapping with Google Earth Engine. J. Open Source Software 5 (51), 2305. https://doi.org/10.21105/joss.02305.

Wulder, M.A., White, J.C., Luther, J.E., Strickland, G., Remmel, T.K., Mitchell, S.W., 2006. Use of vector polygons for the accuracy assessment of pixel-based land cover maps. Can. J. Remote. Sens. 32 (3), 268–279. https://doi.org/10.5589/m06-023.

Zhang, G., Liu, F., Song, X., 2017. Recent progress and future prospect of digital soil mapping: a review. J. Integr. Agric. 16 (12), 2871–2885. https://doi.org/10.1016/S2095-3119(17)61762-3.

II

European Journal of Soil Science

WILEY

RESEARCH ARTICLE OPEN ACCESS

A Two-Part Framework for Depth to Bedrock Prediction and Uncertainty Assessment in Sweden

Yiqi Lin[1] | Gustaf Peterson[2] | Cecilia Karlsson[2] | Florian Westphal[1,3] | William Lidberg[1] | Anneli M. Ågren[1]

[1]Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Umeå, Sweden | [2]Geological Survey of Sweden, Uppsala, Sweden | [3]Department of Computing, Jönköping University, Jönköping, Sweden

Correspondence: Yiqi Lin (yiqi.lin@slu.se)

Received: 31 May 2025 | **Revised:** 28 January 2026 | **Accepted:** 16 March 2026

Keywords: depth to bedrock | digital soil mapping | quantile regression forest | two-part model | uncertainty visualization | zero-inflated data

ABSTRACT

Accurate mapping of depth to bedrock (DTB) in complex post-glacial landscapes is challenging due to high spatial variability and the prevalence of bedrock outcrops, which introduce "structural zeros" that violate standard regression modelling assumptions. To address this, we developed a two-part machine learning framework that separates bedrock outcrop classification from continuous depth prediction and applied it to a Swedish case study. The binary classifier effectively distinguished outcrops from sediment-covered areas (AUC = 0.96, F1-score = 0.83), whereas the regression component provided reliable DTB estimates in non-outcrop areas (R^2 = 0.68, RMSE = 5.74 m). The final fused model (R^2 = 0.67, RMSE = 5.80 m) outperformed both the existing national Inverse Distance Weighting interpolation model (R^2 = 0.61, RMSE = 6.61 m) and a global model evaluated over the study area (R^2 = 0.23, RMSE = 9.03 m). The two-part model remains robust in data-sparse regions. However, a depth-stratified uncertainty analysis revealed miscalibration in the uncertainty estimates of the regression component: in shallow ranges (2–15 m), the model overestimates uncertainty and produces overly wide prediction intervals. In deep ranges (> 30 m), it underestimates uncertainty while systematically underpredicts (mean error = 12.44 m). Our findings emphasize that zero-inflated datasets require special consideration in modeling approaches, and that depth-stratified evaluation is essential for understanding model reliability.

1 | Introduction

Unconsolidated sediments, also known as regolith or overburden, are integral to the functioning of the Earth's Critical Zone—the terrestrial surface layer from unaltered bedrock to the vegetation canopy (Brantley et al. 2017; Banwart et al. 2019). In mid-latitude regions, this regolith was primarily formed during the Quaternary period (2.58 Ma—present) (Gibbard and Head 2010), thus termed "Quaternary Deposits" (QDs). The landscape in these regions was shaped by glacial ice movement and associated erosional and depositional processes during repeated glacial–interglacial cycles, resulting in distinct geomorphological features and sedimentary deposits (Kleman et al. 2008). These deposits include glacial, post-glacial, and peat deposits (Donner and Donner 2005). Throughout this paper, for clarity, the term "soil" is used in a broad sense to refer to all types of unconsolidated sediments of glacial and post-glacial origin, which constitute the majority of near-surface materials in Fennoscandia, rather than the strict pedological definition, which typically considers soils as products of in situ weathering of geological substrata.

Understanding the characteristics of QDs, or soils (*sensu lato*), is fundamental to many disciplines (Lukas et al. 2017). Among these properties, Depth to Bedrock (DTB) is a critical parameter that influences a range of surface and subsurface processes. DTB is a result of complex interactions between geological, biological, geochemical, hydrologic, and climatic processes, as well as human activities (Anderson 1988; Jenny 1994; Dietrich et al. 2003; Lebedeva and Brantley 2013; Devkota et al. 2018).

Highlights

- Two-part framework explicitly models structural zeros in depth-to-bedrock mapping.
- Machine learning excels in shallow and sparsely sampled regions.
- Inverse distance weighting outperforms machine learning in deep ranges (> 10 m).
- Depth-stratified evaluation reveals hidden model performance limits.

It affects flow paths and groundwater storage capacity and regulates soil nutrient and water availability. These, in turn, affect land stability and various ecological and biogeochemical processes (Meyer et al. 2007; Ohnuki et al. 2008; Lanni et al. 2012; Kim et al. 2015; Silvestri et al. 2019). Accurate DTB mapping is essential for planning engineering and pursuing construction projects and various earth-surface modelling (Gomes et al. 2016; Lukas et al. 2017; Shangguan et al. 2017; Pawley et al. 2024).

DTB maps have been developed across various spatial scales, including watershed, regional, national, and global extents (Pelletier et al. 2016; Yan et al. 2020; Furze et al. 2021; Odom and Doctor 2023; Pawley et al. 2024; Oussou et al. 2025). Depending on the focus of the field, some related terms are used, such as soil depth or thickness (Tesfa et al. 2009; Patton et al. 2018; Chen et al. 2019; Malone and Searle 2020) and regolith thickness (Wilford and Thomas 2013; Karlsson et al. 2014; Wilford et al. 2016). Predominant methodologies include mechanistic modeling, which adopts a landscape evolution approach based on mass balance principles (Minasny and McBratney 1999; Dietrich et al. 2003; Pelletier and Rasmussen 2009; Bonfatti et al. 2018); interpolation from direct observations and/or secondary data (Daniels and Thunholm 2014; Karlsson et al. 2014); and data-driven modeling, such as geostatistics and machine learning (ML) (Hengl et al. 2014; Mulder et al. 2016; Shangguan et al. 2017; Chen et al. 2019; Yan et al. 2020; Furze et al. 2021; van der Westhuizen et al. 2024). Despite significant progress, accurately mapping DTB remains challenging because of its highly variable nature, further compounded by the high cost of direct measurements (Liu et al. 2022), inconsistencies in definitions (Lacoste et al. 2016), and the prevalence of right-censored data (where observed thickness underestimates actual thickness) (Chen et al. 2019, 2021; Malone and Searle 2020). One often overlooked challenge is the zero-inflated nature of DTB datasets, where an excessive number of zeros—a result of aggregating legacy bedrock outcrop datasets—violates the assumption of standard continuous regression models (Pennell et al. 2025). Zero-inflation can lead to poor model calibration, reduced interpretability, and biased uncertainty measures (Lambert 1992; Zuur et al. 2009). While Hengl et al. (2014) used zero-inflated models (Agarwal et al. 2002), they offered limited discussion on the modeling specifics and the impact of zeros. Yan et al. (2020) attempted to normalize data distribution via logarithmic transformation but noted that this approach failed to address excessive zero values, which likely led to the significant overestimation of shallow DTB.

Sweden presents a unique case for testing nationwide ML-based DTB mapping due to its extensive data availability. Two DTB maps with nationwide coverage are available for Sweden that offer valuable baselines for comparison. The Geological Survey of Sweden (SGU) developed an Inverse Distance Weighting (IDW) interpolation-based model in Geographic Information System (GIS) environment that utilizes approximately 1.2 million DTB samples (Daniels and Thunholm 2014). On a global scale, Shangguan et al. (2017) created a DTB map using Random Forest (RF) and Gradient Boosting, incorporating a compilation of soil profiles, borehole logs, and expert-generated pseudo-observations. A critical limitation of both models is the lack of spatially explicit uncertainty information. Maps produced by deterministic models like IDW can be adapted into categorical classes to imply uncertainty, and standard ML algorithms can approximate uncertainty by estimating the variance across ensemble predictions. Neither of these indirect measures offers a true representation of prediction uncertainty. Given uncertainty's impact on spatial analysis and decision-making, clear communication is crucial to prevent misinterpretation of the maps and ensure model limitations are acknowledged (Kinkeldey et al. 2014). Visualizing uncertainty and utilizing uncertainty information effectively remain persistent issues for both researchers and end-users (Arrouays et al. 2020).

To address these challenges, we adopted a two-part (hurdle) modeling framework (Heilbron 1994), which integrates a binary classifier with a truncated-at-zero regression model. The binary part acts as a "hurdle" to determine the possibility of bedrock outcrop presence (1, DTB = 0 m) or absence (0, DTB > 0 m), before the regression model is applied. We employed the Random Forest (RF) algorithm (Breiman 2001) for binary classification and Quantile Regression Forest (QRF) (Meinshausen 2006), a probabilistic extension of RF, for predicting DTB. Unlike traditional ML methods that provide point estimates, QRF retains all individual observations within each leaf node to construct conditional empirical distribution functions. This allows for extracting quantiles, computing prediction intervals (PIs), and characterizing distributional properties such as skewness and multimodality (Meinshausen 2006). QRF is increasingly used in Digital Soil Mapping due to its ability to quantify uncertainty via PIs at user-defined quantiles (Vaysse and Lagacherie 2017; Chen et al. 2021; Kasraei et al. 2021).

Leveraging the extensive geological data available at SGU, this study aims to understand whether ML methods utilizing environmental covariates are more effective than traditional interpolation techniques in capturing DTB variability. The primary objectives of this study are:

1. To assess the potential of a two-part framework for DTB mapping and uncertainty estimation:
2. To evaluate the performance and limitations of the existing national IDW model and a previously published global ML model for Sweden;
3. To investigate the key environmental covariates that contribute to identifying bedrock outcrops and predicting DTB; and

4. To present two approaches for visualizing prediction uncertainty

2 | Data and Methods

2.1 | Study Area

The study area encompasses the entirety of Sweden, situated between 55° N and 70° N and 11° E and 25° E, covering approximately 470,000 km^2. Sweden's bedrock consists primarily of Precambrian crystalline and metamorphic basement, with Phanerozoic sedimentary and Caledonian orogenic rocks occurring locally in the Scandinavian mountain belt and southernmost regions (Stephens 2020). Elevation ranges from sea level to slightly above 2000 m, with terrain transitioning from northwestern mountains to southern lowland plains. During the Quaternary, continental ice sheets repeatedly advanced and retreated across the Fennoscandian Shield (Wohlfarth et al. 2008; Stroeven et al. 2016). Glacial erosion stripped existing regolith and bedrock, creating striations and carving deep valleys, while glacial and glaciofluvial deposition left behind landforms such as drumlins, eskers, and moraines and associated sediments up to tens of meters thick (Wohlfarth et al. 2008). During the late-glacial period, central lowlands and coastal areas were submerged under the sea. As the ice sheet retreated, the land experienced isostatic rebound, a rapid uplift caused by the release of glacial pressure (Björck 1995). Post-glacial land uplift exposed older QDs, which were reworked by fluvial, marine, and aeolian processes that redistributed the sediments. The highest coastline represents the maximum post-deglaciation sea extent and is a crucial reference for understanding Swedish geomorphology and QD distribution (Peterson 2022). Areas above the highest coastline, never submerged, are predominantly covered by glacial (till, approximately 75% of the landscape) and glaciofluvial sediments, while areas below are characterized by post-glacial wave-washed and marine-reworked sediments (clay, silt, sand, gravel) (Fredén 1990). Additionally, the combination of a humid climate in the Holocene epoch, high groundwater tables, and flat terrain created favorable conditions for peat formation (Rundgren 2008). In this article, the term DTB refers to the vertical distance from the ground surface to the underlying bedrock or lithified parent material, encompassing the total thickness of QDs and weathered rock layers.

2.2 | DTB Point Observations

We compiled a total of 1,199,920 DTB observations from eight publicly available datasets at SGU, including stratigraphic sampling (well archive, peat archive, stratigraphic and hydrogeological database), surface observations (bedrock outcrops, glacial striations, and petrophysical measurements), and geophysical surveys (Table S1). The dataset exhibits a highly right-skewed and zero-inflated distribution (Figure 1), with values ranging from 0 m (exposed bedrock) to 150 m. The mean DTB is 8.18 m (±10.12 m), while the median is 4.5 m. There are two types of observation points:

- Observations with confirmed DTB (n = 1,098,771): cases where bedrock was encountered subsurface (n = 797,386) or at the surface (DTB = 0 m) (n = 301,385).
- Observations with assumed bedrock encounter (n = 101,149): cases where a hard surface was encountered, but validation via rock sounding was not explicitly performed due to the resource-intensive nature (e.g., additional deep drilling and/or sounding in multiple adjacent locations). These hard surfaces are assumed to be bedrocks, as lithified layers between bedrock and surface deposits are rare in our study area.

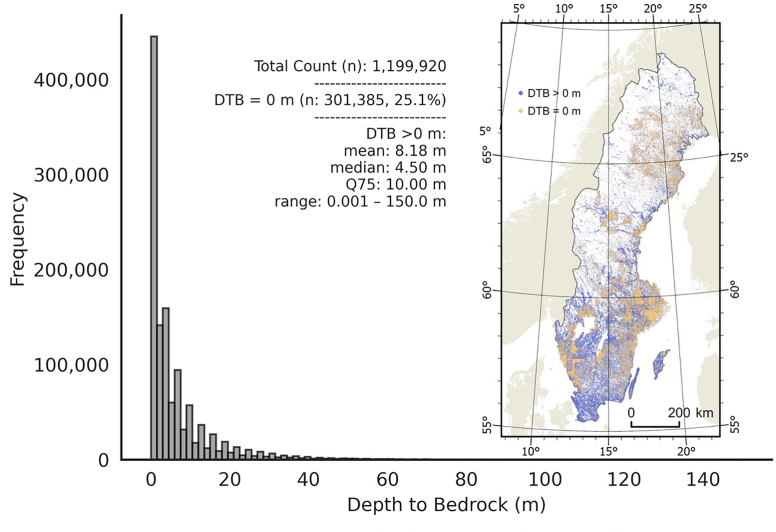


FIGURE 1 | Histogram and spatial distribution of depth-to-bedrock (DTB) observations (n = 1,199,920) over Sweden. The strong right skewness, zero inflation (25.1%), and spatial clustering might pose potential challenges for modeling.

2.3 | Existing Models

2.3.1 | National Inverse Distance Weighting (IDW)-Based Model

The Geological Survey of Sweden (SGU) produced a 10 m-resolution map using IDW interpolation method in combination with GIS analysis of QD map and the dataset described in Section 2.2 (Daniels and Thunholm 2014). As a deterministic method, IDW estimates values based on nearby observations (Burrough et al. 2015). Because different deposit types tend to have distinct depth ranges (e.g., clay is generally deeper than till), the modeling approach performed interpolation separately for each class, using a simplified QD map to assign a class to each point (Figure S1). This approach also allowed for adjustments; for example, bedrock outcrops were fixed at 0 m, and areas along bedrock faults received an additional 10 m (Daniels and Thunholm 2014). The resulting rasters were then masked to their respective QD extents and combined. Potential gaps (NoData values) were filled using IDW interpolation of all data, and focal statistics were applied to smooth the surface. We refer to this framework as the "IDW model," though it comprises additional GIS steps. The model can be considered highly reliable in areas with relatively flat bedrock and simple QD conditions, but not in geologically complex areas (Daniels and Thunholm 2014). The publicly available version of this map is a categorical map with predefined depth intervals, accessible via SGU's online map viewer at https://apps.sgu.se/kartvisare/kartvisare-jorddjup.html.

2.3.2 | Global Machine Learning-Based Model

Shangguan et al. (2017) created a series of ML-based global DTB maps using a compilation of soil profiles (ca. 130,000 locations), borehole logs (ca. 1.6 million locations), and expert-generated pseudo-observations to address sampling gaps. A stack of 155 covariates served as input data for the RF and Gradient Boosting Tree models, including DEM-derived hydrological and morphological features, lithologic units, MODIS surface reflectance bands, and vegetation indices. The training dataset incorporated 320,451 Swedish borehole records. Output products at 250-m resolution include: (a) absolute DTB, (b) censored DTB (up to 200 cm), and (c) occurrence of R horizon within 200 cm (%) (Shangguan et al. 2017). However, model evaluation focused on North America, leaving the reliability for Sweden uncertain. The maps used for our analysis were downloaded from http://globalchange.bnu.edu.cn/ in September 2024.

2.4 | Covariate Layers

Given that DTB in Sweden is largely shaped by glacial processes rather than in situ soil formation, we selected 21 features (18 continuous and 3 categorical) reflecting these processes influencing sediment thickness and bedrock exposure. Terrain indices were derived from LiDAR (Light Detection and Ranging)-based DEMs at 2, 20, and 50 m resolutions (Table 1a). Fine-scale indices (2 m resolution; e.g., roughness, ruggedness, downslope index) detect surface textures characteristic of bedrock outcrops and local gradients. Medium-scale indices (20 m resolution; e.g., slope, curvature, relative topographic position) were used as proxies for erosion and sediment transport potential. Large-scale indices (50 m resolution with a large window size) identify broad landforms such as valley bottoms and hilltops.

To represent landscape-scale controls, we calculated distances to key geological features: bedrock deformation zones (associated with deep deposits) and the highest coastline (delimiting post-glacial wave action and sediment redistribution). Legacy maps provided lithotectonic, surface geological, and geocoordinate information. Categorical variables were one-hot encoded to enable the assessment of the impact of individual classes on DTB prediction (Table 1b).

2.5 | Modelling Framework

2.5.1 | Two-Part Model

We implemented a two-part modeling framework (Figure 2).

- Binary classification: a RF model trained on all data to classify areas as either bedrock outcrop (1) or non-outcrop (0), with predictions exceeding a probability threshold classified as outcrops.
- DTB regression: a QRF model trained only on non-outcrop samples (truncated-at-zero). We constructed PIs from key conditional quantiles and used the median predictions (50th percentile) as the point estimate to compute evaluation metrics.

Models were fused using a 0.5 threshold: if the binary model predicted outcrop, DTB was set to 0; otherwise, the QRF median was used.

2.5.2 | Model Training

We employed a stratified 5-fold cross-validation (CV) to maintain consistent DTB distributions across depth ranges in each fold. Hyperparameters were optimized through 20 search trials using nested 5-fold CV. The binary RF was optimized to maximize Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) (Fawcett 2006) and minimize log-loss. The QRF model was optimized to minimize the Root Mean Squared Error (RMSE). The best hyperparameter combination was used to train the final model.

To interpret QRF behavior, we used a post hoc explainability method called SHapley Additive exPlanations (SHAP) (Lundberg and Lee 2017) with the TreeExplainer (Lundberg et al. 2020). We used a background dataset of 500 randomly sampled instances from the test set to simulate the reference distribution against which feature contributions are measured. SHAP values quantify the marginal contribution of each feature to a specific prediction across all possible feature combinations and therefore can be used as a measure of feature importance.

TABLE 1 | A list of (a) continuous features and (b) categorical features used to model Depth to Bedrock (DTB).

(a) Continuous features

Feature	Scale/resolution	Specific settings	Raster value median (range)	Source
Digital Elevation Model (DEM)	2 m		110.94 (−2.07–863.65 m)	The Swedish Land Survey (2022)
Elevation above stream from a 1 ha stream network	2 m		2.0 (0.0–154.0 m)	Lindsay (2024)
Downslope index with 2 m drop	2 m		124.45 (2.0–31250.0)	Lindsay (2024)
Circular variance of aspect	2 m	3 × 3 cells	0.05 (0.0–0.99)	Lindsay (2024)
Standard deviation of Slope	2 m	3 × 3 cells	0.67 (0.02–21.93)	Lindsay (2024)
Deviation from mean elevation	2 m	7 × 7 cells	−0.02 (−3.09–3.58)	Lindsay (2024)
Terrain ruggedness index	2 m	3 × 3 cells	0.1 (0.0–4.14)	Lindsay (2024)
Multi-Scale roughness magnitude	2 m	Min_scale = 1, max_scale = 10, step = 2	1.98 (0.2–45.58)	Lindsay (2024)
Maximum Elevation deviation	2 m	Min_scale = 50, max_scale = 500, step = 50	−0.37 (−4.49–8.15)	Lindsay (2024)
Slope	20 m	20	2.58 (0.00–89.84)	Lindsay (2024)
Profile curvature	20 m		0 (−0.02–0.03)	Lindsay (2024)
Relative topographic position	20 m	20 × 20 filter	0.05 (−1.0–1.0)	Lindsay (2024)
Relative topographic position	50 m	50 × 50 filter	0.00 (−1.0–1.0)	Lindsay (2024)
Easting	2 m		568,679 (280795–913,007)	
Northing	2 m		6,920,018 (6141258–7,608,074)	
Elevation under the highest coastline	1:100,000–1:200000		8.0 (0.0–255.0 m)	The Geological Survey of Sweden (2015)
Distance to fault lines	10 m		707.8 (0–65,535 m)	
Age since deglaciation	2 m		11.70 (10.0–17.29 ka)	Hughes et al. (2016)

(b) Categorical features

Feature	scale/resolution	Number of classes	Source
Quaternary deposit (QD) map with simplified classes	1:25,000 to 1: 1,000,000	8	Karlsson et al. (2021)
Geomorphons	20 m	10	Lindsay (2024)
Lithotectonic map	1:50,000 to 1: 250,000	23	The Geological Survey of Sweden (2023)

2.5.3 | Model Evaluation and Comparison With Existing DTB Maps

Binary classification performance was assessed using Precision, recall, F1 score, and ROC-AUC. The ROC curve plots True Positive Rate against False Positive Rate across probability thresholds; AUC near 1 indicates strong discriminatory power, whereas 0.5 indicates random performance (Fawcett 2006).

Regression performance was evaluated using RMSE, Mean Error (ME), also known as bias, Coefficient of Determination (R^2), and Concordance Correlation Coefficient (CCC) (Lin 1989). RMSE penalizes large errors by squaring

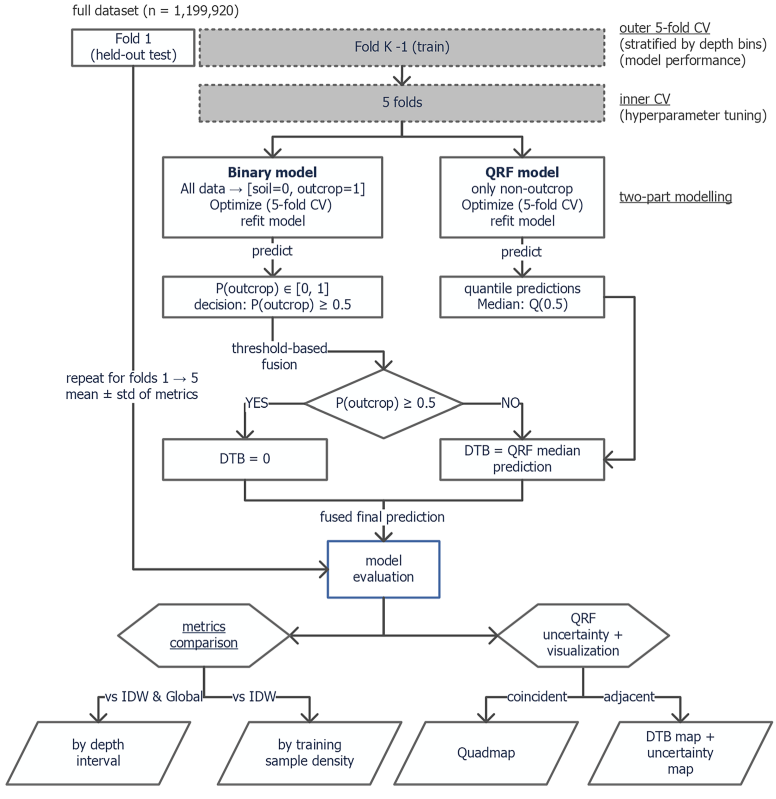


FIGURE 2 | Two-part modeling framework for Depth to Bedrock (DTB) prediction: Random Forest (RF) binary classification distinguishes bedrock outcrops from non-bedrock areas, and Quantile regression forest (QRF) predicts DTB. Models are fused at prediction time with a threshold (P(outcrop) $\geq 0.5 \rightarrow$ DTB = 0).

deviations and prevents the positive and negative deviations from offsetting each other. ME indicates clearly the direction of the error. R^2 measures the proportion of variance explained; CCC measures prediction-observation agreement (perfect concordance = 1).

Given the right-skewed distribution (90% of observations < 20 m), we binned depths into seven intervals (0–2, 2–5, 5–10, 10–15, 15–20, 20–30, > 30 m) to emphasize operationally important shallow ranges. We prioritized evaluating "false positives" (non-outcrops classified as outcrops), as these force predictions to zero and pose greater operational risk than false negatives (missed outcrops), which can be mitigated by the QRF's ability to predict shallow DTB values and are easier to manage in field operations.

For comparison, we ran the IDW-based model with four folds ($n = 959{,}936$) and evaluated on one fold ($n = 239{,}984$). We performed the interpolation only once instead of five times to reduce computational cost, as the IDW model was intended as a baseline for comparison rather than requiring a full stability assessment. From the global 250-m resolution absolute DTB map (Shangguan et al. 2017), we extracted values for all five test folds to calculate mean metrics, excluding locations where true DTB = 0 m, as the global map minimum in Sweden is 0.77 m.

To assess the robustness of the IDW and two-part model in data-sparse regions, we calculated error metrics stratified by local sampling density using the shared test fold. Sampling density is defined as the number of training points within a 5 km radius of each test location. The test set was stratified into six density classes, ranging from very sparse (< 1 pts/km^2) to very dense (> 50 pts/km^2), excluding points where IDW failed to produce predictions. This analysis was not performed for the global model due to its different training dataset.

2.5.4 | QRF Uncertainty Quantification and Visualization

To evaluate the calibration of the regression component, we analyzed only correctly classified sediment-covered areas, as false positives are assigned 0 m across all quantiles, making their predictions incompatible with interval-based evaluation. We computed the Prediction Interval Coverage Probability (PICP), which measures the proportion of observations that fall

TABLE 2 | Performance metrics for the binary bedrock outcrop classification, QRF DTB regression, and the final fused two-part model.

Metric	Binary (RF)	Regression (QRF)	Fused two-part	IDW	Global
Note	5-fold CV	5-fold CV	5-fold CV	Single fold	5-fold CV
RMSE (m)	—	5.74 ± 0.05	**5.80 ± 0.07**	6.61	9.03 ± 0.03
ME (m)	—	0.89 ± 0.03	1.01 ± 0.02	−0.70	**−0.21 ± 0.02**
R^2	—	0.68 ± 0.04	**0.67 ± 0.01**	0.61	0.22 ± 0.00
CCC	—	0.81 ± 0.00	**0.80 ± 0.00**	**0.80**	0.38 ± 0.00
ROC-AUC	0.96 ± 0.00	—	—	—	—
F1-score	0.83 ± 0.00	—	—	—	—
Precision	0.87 ± 0.02	—	—	—	—
Recall	0.80 ± 0.02	—	—	—	—
FP Rate (%)	—	—	4.05 ± 0.58	17.81	—

Note: Results are averaged across five cross-validation folds; IDW was evaluated on a single fold. The significant values are in bold.

within a specified PI (Papadopoulos et al. 2001; Shrestha and Solomatine 2006; Malone et al. 2011). Well-calibrated models show observed coverage matching nominal values, for example, PI90 should contain ~90% of observations. Deviations indicate miscalibration: PICP below nominal indicates undercoverage (intervals too narrow, model overconfident), whereas PICP above nominal indicates overcoverage (intervals unnecessarily wide, uncertainty overestimated) (Lilburne et al. 2024). Model calibration was assessed by plotting observed versus nominal coverage across multiple PIs; well-calibrated models align with the 1:1 line. These visualizations are termed "reliability plots" (Schmidinger and Heuvelink 2023; Lilburne et al. 2024).

We further diagnosed model behavior using the Quantile Coverage Probability (QRP). It has a similar logic to PICP, but instead of intervals, it evaluates single quantile predictions by measuring the proportion of observations falling below a specific predicted quantile (Schmidinger and Heuvelink 2023). Coverage exceeding nominal levels indicates overestimation (quantiles too high), while coverage below nominal indicates underestimation (quantiles too low).

Prediction uncertainty was visualized using coincident (single map) and adjacent (separate maps) approaches for a heterogeneous post-glacial area ~78 km northwest of Stockholm. The coincident approach adopts the QuadMap method, detailed by Padarian and McBratney (2023). QuadMap uses the Quadtree algorithm, which recursively divides the area into quadrants until meeting a predefined uncertainty threshold. The resulting visualization intuitively represents uncertainty through spatial resolution—areas with higher uncertainty appear as larger pixels, whereas areas with lower uncertainty are displayed at finer resolutions. As the Quadtree algorithm requires an uncertainty input to determine the splitting criteria whereas QRF generates quantile predictions, we converted PI_{90} into standard deviation (σ) using Equation (1), assuming normal error distribution (Padarian and McBratney 2023):

$$\sigma = \frac{UL - LL}{2 \times 1.645} \tag{1}$$

where σ is the standard deviation, UL and LL represent the upper (95th percentile) and lower (5th percentile) prediction limits, 1.645 is the z-score corresponding to the 95th percentile in a one-tailed standard normal distribution.

Fixed-resolution uncertainty maps were generated using two metrics: first, the ratio of PI_{90} over median prediction (Equation 2; Yan et al. 2020; Poggio et al. 2021), a relative metric that scales uncertainty by predicted depth to account for the large variation in DTB. Second, σ (Equation 1), which represents absolute uncertainty in meters and provides context for an intermediate step in the QuadMap method.

$$\text{normalized uncertainty} = \frac{\text{UL} - \text{LL}}{\text{median prediction}} \tag{2}$$

Spatial prediction used the best CV models at 2-m pixel resolution, with all covariates resampled and aligned to the national coordinate system (SWEREF99 TM). The final fused map combined binary and QRF predictions: outcrop pixels received 0 m depth, non-outcrop pixels received QRF median predictions. Final products were resampled to 10-m resolution for direct comparison with the existing IDW-based map.

3 | Results

3.1 | Two-Part Model Performance Metrics

The performance metrics for the binary bedrock classification and DTB regression models are averaged across five folds (Table 2). Precision (0.87 ± 0.01), recall (0.80 ± 0.02), F1-score (0.83 ± 0.00), and ROC-AUC (0.96 ± 0.00) values demonstrate that the RF model is robust at distinguishing bedrock outcrops from sediment-covered areas. QRF's performance, evaluated under ideal conditions when the binary classifier is correct, yielded an ME of 0.89 ± 0.03 m and an RMSE of 5.77 ± 0.07 m. The model explained a substantial portion of the depth variance ($R^2 = 0.68 \pm 0.04$) and exhibited strong agreement between predicted and observed values (CCC = 0.81 ± 0.00). The PICP for

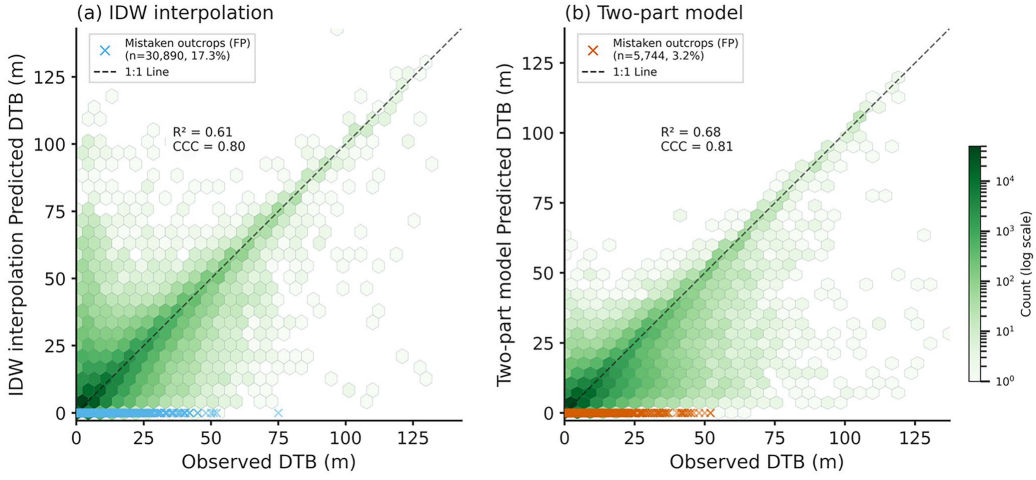


FIGURE 3 | Hexbin plots comparing observed versus predicted depth to bedrock (DTB) values for (a) Inverse Distance Weighting (IDW) interpolation and (b) the proposed two-part model. Hexagon color intensity represents point density. The diagonal dashed line represents perfect prediction (1:1 line).

PI_{90} ($PICP_{90}$) was 93.3% (Figure 4a), indicating that the model slightly overcovers and overestimates its uncertainty.

The integration of the binary and regression components into the final fused model resulted in minimal performance degradation ($\Delta RMSE = 0.06$ m, $\Delta ME = 0.12$ m) (Table 2). The integrated model achieved an overall ME of 1.01 ± 0.02 m, an RMSE of 5.80 ± 0.07 m, and an R^2 of 0.67 ± 0.01. The overall misclassification rate (non-outcrops mistaken as outcrops) was $4.05\% \pm 0.58\%$.

3.2 | Comparison With Existing DTB Maps

Overall, both the IDW and two-part models demonstrated strong predictive performance, with the two-part model slightly outperforming the IDW model (RMSE = 5.80 m vs. 6.61 m, $R^2 = 0.68$ vs. 0.61). The global model markedly underperformed (RMSE = 9.03 m, $R^2 = 0.22$). Both the IDW (CCC = 0.80) and the two-part model (CCC = 0.80) show substantial agreement between observed and predicted values. However, IDW has more dispersed predictions (Figure 3a) and a tendency to overpredict in the shallow regions (ME = −0.70 m), while the two-part model tends to underpredict the deeper ranges (ME = 1.01 m) (Figure 3b). One noticeable difference is that the IDW model (Figure 3a) suffered from a high false positive rate (17.18%), where true sediment covered areas were assigned zero depth.

Depth-stratified analysis revealed that the two-part model excelled in shallow depths (0–10 m), with a low misclassification rate across all depths (Table 3). The regression part demonstrated systematic bias, overestimating DTB (negative ME) in lower ranges while underestimating (positive ME) in higher ranges. This misclassification error exhibited a clear depth-specific trend, with the highest rate of $8.8\% \pm 1.3\%$ at 0–2 m. Between 2 and 10 m, the misclassification rate decreases rapidly from $3.7\% \pm 0.5\%$ (2–5 m) to $2.1\% \pm 0.2\%$ (5–10 m). Beyond 10 m, the misclassification rate remains low (< 2%). The IDW model, however, exhibits the lowest bias, particularly in deeper areas (> 30 m) (ME = 5.68 m). The global model exhibits overprediction in the shallow ranges and extreme underprediction in deeper ranges.

Stratifying model performance by training sample density reveals that the two-part model remains stable even as sampling density decreases, whereas the IDW model is highly sensitive to data scarcity (bottom part of Table 3). In data-sparse regions (< 1 pts/km^2), IDW performance degrades sharply (RMSE = 7.86 m) with a strong tendency to overpredict sediment depth (ME = −2.72 m). In contrast, the two-part ML model remains stable, even in the most data-sparse zones (RMSE = 5.90 m). The disparity in performance is wider at the extremes, whereas in moderately sampled areas (5–10 pts/km^2), the models performed similarly. Interestingly, IDW performance degrades in the most data-dense region (RMSE = 7.06 m). The complete depth-stratified analysis statistics are provided in Table S2.

3.3 | Reliability of QRF Uncertainty Estimates

We assessed QRF calibration using three complementary plots (Figure 4). Overall, the model produces conservative uncertainty estimates, with PICP values consistently above the 1:1 line (Figure 4a), indicating intervals wider than necessary (overall $PICP_{90} = 93.2\%$). However, this behavior varies with depth (Figure 4b): the model overestimates uncertainty at shallow-to-intermediate depths (2–15 m) but underestimates it at depth (> 30 m). Specifically, shallow-to-intermediate ranges show overcoverage ($PICP_{90}$: 98.3% at 2–5 m, 96.7% at 5–10 m, 94.5% at 10–15 m), while deep deposits (> 30 m) exhibit severe undercoverage (80.6%).

The QCP plot (Figure 4c) reveals the biases causing the deviation of $PICP_{90}$. In the 0–5 m range, median QCP exceeds 0.50,

TABLE 3 | Comparison of model performance metrics for Inverse Distance Weighting (IDW), the two-part model, and the global model by Shangguan et al. (2017).

		IDW		Two-part model		Global
	Class	**ME (m)**	**False positive rate (%)**	**ME (m)**	**False positive rate (%)**	**ME (m)**
By depth (m)	(0, 2]	**−1.18**	44	−1.35 ± 0.01	**8.85 ± 1.44**	−4.72 ± 0.02
	(2, 5]	−1.41	15	**−0.76 ± 0.04**	**3.71 ± 0.60**	−4.08 ± 0.01
	(5, 10]	**−0.2**	5	0.87 ± 0.04	**2.08 ± 0.25**	−1.25 ± 0.02
	(10, 15]	**0.89**	2	2.76 ± 0.05	**1.60 ± 0.22**	2.37 ± 0.02
	(15, 20]	**1.86**	1	4.63 ± 0.06	1.55 ± 0.31	6.0 ± 0.03
	(20, 30]	**2.89**	1	6.88 ± 0.09	1.22 ± 0.23	11.55 ± 0.08
	> 30	**5.68**	1	12.44 ± 0.31	**0.74 ± 0.12**	27.89 ± 0.14
	Class	ME (m)	RMSE (m)	ME (m)	RMSE (m)	Support (n)
By sampling density (pts/km^2)	< 1 (Very Sparse)	−2.72	7.86	**0.85**	**5.9**	17,167
	1–3	−1.59	6.68	**0.56**	**5.29**	50,319
	3–5	−0.6	5.57	0.6	**5.05**	35,596
	5–10	**−0.38**	5.29	0.47	**4.91**	55,521
	10–50	**−0.23**	5.61	0.58	**5.1**	63,941
	> 50 (Very Dense)	**−0.24**	7.06	0.4	**4.5**	15,676

Note: Top: Performance stratified by depth intervals, showing Mean Error (ME) and false positive rate (true sediment-covered areas misclassified as outcrops). Bottom: Performance stratified by training sample density classes, showing ME and Root Mean Square Error (RMSE). Density-stratified metrics were not computed for the global model. The significant values are in bold.

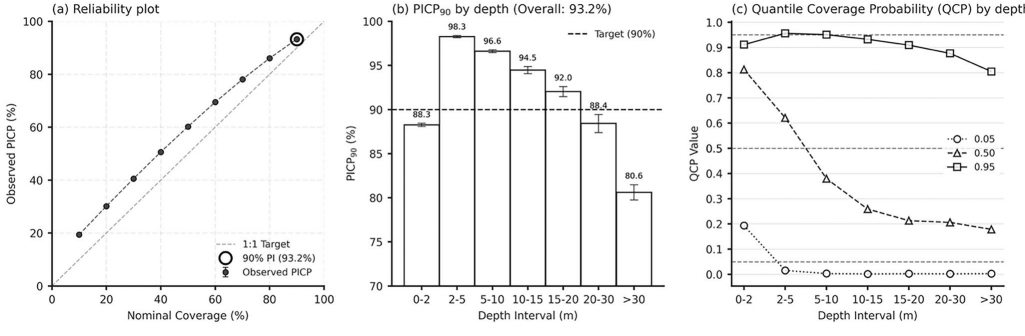


FIGURE 4 | Assessment of quantile reliability. (a) Reliability plot comparing nominal coverage probabilities against observed Prediction Interval Coverage Percentages (PICP). Points above the 1:1 diagonal indicate the model overcovers. (b) 90% Prediction Interval (PI_{90}) calibration by depth. The model exhibits overcoverage in the 2–15 m range but transitions to severe undercoverage in deeper areas (> 30 m). (c) Quantile Coverage Probability (QCP) plot reveals the biases causing miscalibration. Each line tracks the proportion of true observations falling below a predicted quantile (0.05, 0.50, 0.95) across the depth ranges.

indicating systematic overprediction of DTB. As depth increases, median QCP drops below 0.50, reflecting systematic underprediction. The 0.05 quantile (lower bound of PI_{90}) captures ~20% of observations at 0–2 m (above the nominal 5%), confirming the lower prediction bound is too high in shallow areas but drops to nearly 0% beyond 5 m. The 0.95 quantile (upper bound of PI_{90}) maintains near-target coverage (~95%) in shallow ranges but deteriorates to ~80% in deep ranges (> 30 m).

3.4 | Feature Importance for QRF Model

We generated a global summary (aggregated measure) of feature importance during the prediction stage for the QRF model (Figure 5). The QRF model relies heavily on direct indicators such as QD types (bedrock outcrops and glaciofluvial sediments); geology (lithotectonic information, distance to the highest coastline and fault lines); and topographic and surface

characteristics (Multiscale roughness magnitude, Relative topographic position, slope etc.). Bedrock outcrop class shows a strong negative correlation with DTB values when present. Glaciofluvial sediments show a strong positive correlation with DTB, whereas till exhibits a negative pattern with SHAP values tightly clustered near zero. DTB also shows a strong spatial trend across the study area, with Northing and Easting playing important roles in model prediction. DEM contributes minimally to the prediction (clustered near 0), but for certain areas, low DEM values strongly increase the predicted DTB (long positive tail).

3.5 | Visual Comparison of DTB and Uncertainty Maps

The visualization area lies below the highest coastline and is characterized by deep clay deposits (Figure 6a). The IDW map (Figure 6b) captures general depth patterns but assigns unrealistically large values to rugged terrains (e.g., lower right corner). The modeling decision sets bedrock outcrop polygons from the QD map to zero values (orange-red). The two-part model map (Figure 6c) shows sharper bedrock delineation, but a shallower maximum DTB (21.26 m) compared to IDW (27 m).

The uncertainty maps (Figure 6e,f) display the pixel-level predictions of the two uncertainty metrics (Equation 1 and 2), derived from the 95th and 5th percentiles of the QRF conditional distribution. They exhibit opposite spatial patterns: the relative uncertainty ratio (Figure 6e) shows inflated values in shallow, rugged terrain, while the absolute standard deviation (Figure 6f) indicates that error magnitude is highest in river valleys and flat terrains. The QuadMap (Figure 6d) encodes the uncertainty information in Figure 6f into variable spatial resolution, using a 0.5 m threshold to retain fine (10 m) resolution in high-certainty areas ($\sigma < 0.5$ m) while aggregating uncertain pixels into progressively coarser blocks up to 150 m. This produces an average resolution of 60 m with slightly reduced maximum DTB values.

4 | Discussion

4.1 | Considerations of the Two-Part Modeling Framework

DTB datasets contain "structural zeros"—zero that arise from processes that can only generate zeros (Zuur et al. 2009), that is outcrops have a DTB value of 0 m by definition. The two-part framework treats outcrops and sediment depth as distinct processes, alleviating the challenges that standard regression models often face with zero-inflated data (Lambert 1992; Zuur et al. 2009). A two-part framework enables both QRF-based uncertainty quantification and separate error diagnosis before model fusion. While other approaches exist for handling zero-inflated data (Lambert 1992), the two-part framework is better suited for structural zeros (Pennell et al. 2026).

Misclassification in shallow areas likely stems from two factors: model confusion in the ambiguous transition zone between outcrops and thin sediment cover, and the binary threshold setting, which directly controls outcrop identification. Analysis of false negatives (Figure S2) showed that when outcrops were missed, QRF still predicted shallow DTB values, explaining minimal performance degradation (ΔRMSE = 0.06 m, ΔME = 0.12 m) using a 0.5 threshold. False positives pose a greater risk, forcing predictions to zero regardless of actual depth (which can

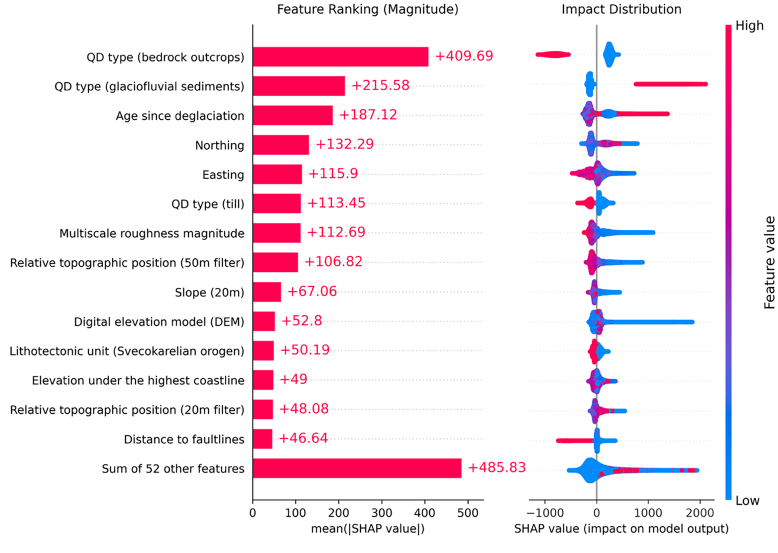


FIGURE 5 | SHAP summary plot showing the top 14 features impacting QRF prediction. The bar plot (left) represents the global magnitude of their influence on predicted DTB. The beeswarm plot (right) illustrates the distribution and the contribution of the feature to moving the prediction away from the baseline (mean) prediction.

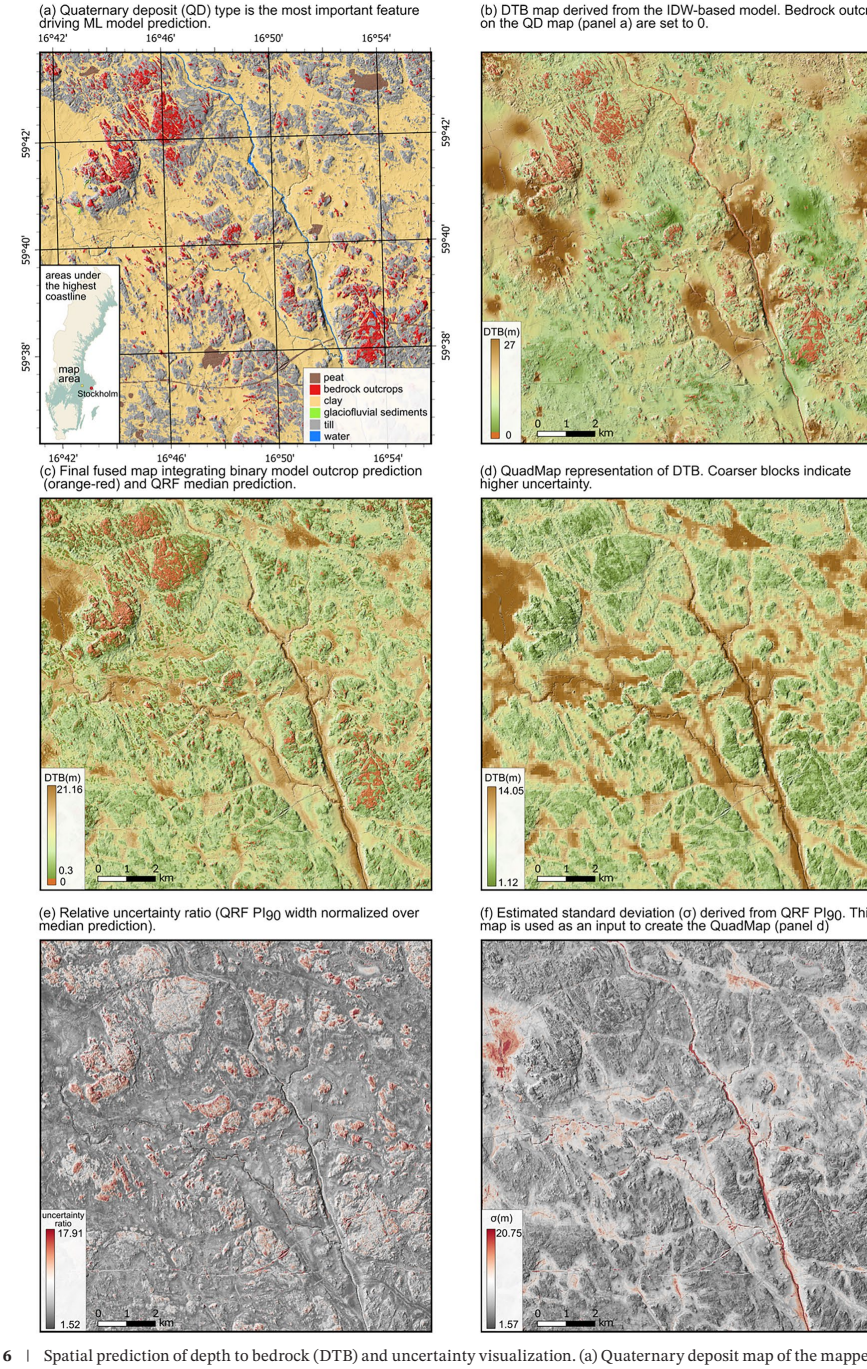


FIGURE 6 | Spatial prediction of depth to bedrock (DTB) and uncertainty visualization. (a) Quaternary deposit map of the mapped area, showing extensive clay plains. (b) DTB map derived from the IDW-based model. (c) Final fused map from the two-part model. (d) QuadMap representation of DTB with coarser blocks indicating higher uncertainty. (e) Relative uncertainty ratio (PI_{90} width normalized by median prediction). (f) Estimated standard deviation (σ) in meters.

exceed 50 m; Figure 3). Future work should analyze false negative probabilities to distinguish model confusion (probabilities near 0.5) from suboptimal threshold selection (probabilities far from 0.5) to ensure justified threshold choices. Additionally, this information would complement our current uncertainty analysis (Section 3.3), which focused exclusively on the regression component.

The QRF model exhibited systematic bias. The strong performance observed in shallow depths (0–10 m) is likely due to the high correlation between terrain attributes and DTB (Furze et al. 2021; Iwahashi 2025). This relationship is likely more distinct in areas above the highest coastline, where the landscape is characterized by glacial landforms sculpted by subglacial processes (Alley et al. 2019; Chandler and Evans 2021). These processes connected to the movement of the ice-sheet bed along the substrate seldom lead to landforms with relief higher than 10 meters, which explains the strong connection between surface topography and bedrock morphology.

Model accuracy deteriorates with increasing depth, a phenomenon presumably more pronounced in areas below the highest coastline where different processes have been active. These formerly submerged areas contain thicker sediments due to continuous glaciomarine and postglacial marine accumulation, which produced flat surfaces decoupled from bedrock morphology. As high-resolution LiDAR only captures surface conditions, deep sediments, which lack a unique surface signature, can appear visually similar to shallow sediments in topographical data, leading to a drop in ML performance (Furze et al. 2021). The weakened correlation is exacerbated by ML model's tendency to regress towards the mean. The "regression-to-mean" pattern has been observed across multiple DTB mapping studies (Hengl et al. 2014; Shangguan et al. 2017; Yan et al. 2020; Goodling et al. 2024), particularly when predictors only partially capture the underlying data variability (Shangguan et al. 2017).

The challenge of identifying deep deposits is not merely an algorithmic limitation of ML models, but an inherently difficult interpretive problem. Fraser et al. (2020) demonstrated that even among human experts interpreting LiDAR-derived hillshade images, the greatest disagreements occurred near transitions to deep soils or within isolated pockets of deep deposits surrounded by outcrops.

4.2 | Two-Part Model Compared to National and Global Models

IDW exhibits a 44% misclassification rate in the 0–2 m range, whereas the two-part ML model achieves 4.05% (±0.58%). This is due to the IDW modeling decision, which assigns zero values to all mapped outcrop polygons on the QD map. These polygons are generalized for visualization purposes and typically encompass areas larger than actual outcrops (Karlsson et al. 2021). The ML model uses high-resolution data for classification, enabling more precise delineation between outcrops and sediment-covered areas.

All methods show reduced accuracy as depth increases (Table 3). However, IDW maintains more reliable performance in deeper ranges. This confirms the suitability of IDW for data that is noisy and highly variable (Odom and Doctor 2023). Evaluation of model performance by sample density (Table 3) shows IDW performance deteriorates as sampling density decreases, reflecting the fundamental limitation of distance-based interpolation when nearest observations are distant. ML model's relatively stable metrics indicate that environmental covariates can provide predictive information even where observations are sparse, a critical advantage for mountainous and northern inland areas with limited field data. IDW's elevated error level in very densely sampled areas (> 50 pts/km^2) is likely because these areas correspond to regions with high spatial variability. The Precambrian bedrock surface in Sweden exhibits strong undulation, with linear valleys and depressions aligned with fracture zones—features shaped by pre-Quaternary chemical weathering and subsequent glacial erosion (Lidmar-Bergström 1996; Migoń and Lidmar-Bergström 2001). This complexity is difficult to capture through distance-based interpolation. Furthermore, some of the very dense sampling points are highly clustered observations from individual surveys along predefined transects. This might cause the interpolation to become dominated by these points.

Although the global model achieved $R^2 = 0.44$ for European validation, performance in Sweden was substantially lower ($R^2 = 0.22$), highlighting the challenge of transferring generalized relationships learned from a global dataset to complex glacial landscapes. The ME values by depth intervals and high RMSE (9.03 m) suggest that its overall low bias (ME = −0.21 m) is a result of the errors in shallow and deep areas canceling out each other. Our findings align with recent evaluations of the global DTB model in diverse geological contexts. Oussou et al. (2025) found inconsistent results across different water basins in Africa despite the overall maximum DTB values lying in similar ranges, while Yan et al. (2020) reported higher error metrics and more visual artefacts for China compared to a local model. Although global models provide useful baselines for data-scarce regions, this study demonstrates that models using local-specific data produce substantially more reliable predictions for operational applications (Bohn and Miller 2024).

As most of Sweden is covered by thin surface deposits (median DTB = 4.5 m), these findings suggest some methodological considerations for future iterations of DTB maps: ML methods excel in shallow DTB areas (< 10 m) and are more effective for sparsely sampled regions; IDW remains useful for deeper deposits where landform-depth relationships are less distinct, though performance depends on local sampling density.

4.3 | Uncertainty Quantification and Depth-Stratified Analysis

The QRF model exhibited significant overall overcoverage ($PICP_{10} = 19.4\%$, $PICP_{90} = 93.2\%$, Figure 4a), indicating it systematically overestimates its prediction uncertainty due to its inability to distinguish very deep sediments (discussed in more detail the Section 4.1). The problem is exacerbated by sparse training data at depth (> 75% of observations are less than 10 m) and multiple sources of input noise, including measurement errors, positional uncertainty from imprecise drill site coordinates, and

genuine spatial variability in subsurface conditions (Goodling et al. 2024). Widened PIs indicate that the model is imprecise, which may limit its usefulness. However, since the bias is small in these ranges, users can consider the map acceptable for most purposes (Malone et al. 2011). Moreover, a model that is cautious (overestimating uncertainty) could be beneficial for high-stakes applications, where capturing broader ranges of values allows for the preparation of worst-case scenarios.

The model shifts from overprediction in shallow ranges (< 5 m) to underprediction with increasing depth (Figure 4c). In a well-calibrated model, the 0.05, 0.50, and 0.95 quantile lines on the QCP plot (Figure 4c) should remain horizontal at their nominal values across depth ranges. The systematic overestimation in 0–2 m (ME = −1.35 m) shifts the PIs upward. Consequently, the lower prediction bound (0.05 quantile) is too high, capturing ~20% of observations instead of the nominal 5%. In deeper regions (> 20 m), model reliability decreases as it shifts towards undercoverage. The model's tendency to underpredict shifts PIs downwards, causing more samples to exceed the upper prediction bound, resulting in a lower-than-nominal $PICP_{90}$ value. The combination of large uncertainty and bias indicates not only poor prediction accuracy but also unaccounted error sources (Malone et al. 2011). In conclusion, the model is overly cautious in data-rich shallow areas but overconfident exactly where it performs worst. Predicted values in deep ranges should be treated with caution or avoided altogether. These findings underscore the necessity of stratifying evaluation by depth and incorporating multiple metrics, as global metrics alone would mask this uncertainty pattern.

4.4 | Features Driving DTB Prediction

Our analysis revealed that information from the QD map was particularly influential, with the bedrock outcrop class being the most dominant feature. This aligns with Kitterød (2017), who demonstrated that incorporating distance to bedrock outcrops reduced co-kriging estimation variance. Till, though exhibiting high overall importance, its values are clustered near zero. This pattern likely reflects till's prevalence as the most common QD type in the landscape rather than strong predictive power. Glaciofluvial sediments are typically associated with much deeper deposits. "Distance to fault lines" displayed a clear negative correlation with DTB, consistent with the logic that fractures in bedrocks create deep valleys where thick sediments accumulate. This aligns with Pawley et al. (2024) who similarly identified valley depth and proximity to channels as key DTB predictors in glacial landscapes characterized by deeply incised bedrocks.

Topographic variables exhibit a consistent trend: High values for "Multiscale roughness magnitude," "Relative topographic position," and "Slope" could suggest steep, rugged, or elevated terrain where sediment accumulation is expected to be limited. Conversely, low-gradient slopes, depressions, and flat areas, particularly those situated under the highest coastline, are accumulation zones and were positively correlated with deeper DTB. SHAP analysis also revealed multicollinearity in the environmental covariates, which can reduce model interpretability (Kasraei et al. 2024). The east–west gradient reflects the topographic transition from the mountainous west (high elevation, thin till cover, more exposed bedrock) to the eastern coastal lowlands (lower elevation, thick post-glacial sediment accumulation). The north–south gradient coincides with the ice sheet retreat: earlier deglaciation in the south ("Age since deglaciation") allowed longer exposure to postglacial processes and accumulation of organic and weathered materials. It also reflects the transition from the hard crystalline Precambrian shield in the north to the softer, more easily eroded Paleozoic calcareous sedimentary bedrock in the south ("Lithotectonic unit (Svecokarelian orogen)"). Easting and Northing essentially encapsulate the combined effects of topography, lithology, deglaciation history, and depositional environment and are not independent from the environmental features.

For SHAP computation, we used a background size of 500. Although small sample sizes can cause instability in mid-tier feature rankings, the identification of the most and least influential features remains robust (Yuan et al. 2022). We are therefore confident that the dominant drivers identified are trustworthy, despite potential variability in mid-tier feature rankings. Future research could investigate the specific feature contributions for shallow and deep predictions at the observation level. Such analysis might reveal whether surface attributes lose predictive power for deep sediments, potentially motivating separate shallow and deep modeling approaches.

4.5 | Uncertainty Visualization Remains a Challenge

The comparison between relative uncertainty ratio (Figure 6e) and absolute standard deviation (Figure 6f) reveals a critical limitation of normalizing prediction error in datasets with many zero or near-zero values (Poggio et al. 2021). Figure 6e shows the highest uncertainty in shallow areas and lowest in deep deposits, which appears to be an artifact of Equation (2) combined with the depth-dependent model behavior. In shallow areas (0–2 m), dividing the numerator by near-zero denominators artificially inflates the ratio. In the 2–10 m range, the model produces wide PIs ($PICP_{90} > 96\%$), which, when divided by small predicted values, yield disproportionately large ratios. Conversely, in deep areas where the model is overconfident ($PICP_{90} < 90\%$), the already too-narrow intervals are further suppressed when divided by large depth values. In contrast, the absolute σ map (Figure 6f) provides a more interpretable metric by highlighting areas where the actual magnitude of error (in meters) is significant. Since the estimated σ is proportional to the PI_{90} widths (Equation 1)—a sometimes preferred measure for communicating uncertainty (Chen et al. 2021; Kasraei et al. 2021)—it essentially conveys the same information.

The QuadMap approach (Figure 6d) embeds uncertainty directly into spatial resolution. The resulting "blocky" appearance in uncertain regions discourages the association of fine-resolution with high-confidence data. However, coincident display can increase visual complexity or clutter, potentially making it harder to discern both prediction values and their associated uncertainties (Kinkeldey et al. 2014). The spatial aggregation communicates data limitations, but the loss of detail can be perceived by users as less useful, potentially hindering

decision-making where precise spatial patterns are required (Courteille et al. 2025). The adjacent displays, however, require users to shift gaze between maps to compare spatial patterns (Kinkeldey et al. 2014). Thus, effective uncertainty visualization remains a challenge, requiring careful selection of the metric and the display method based on the intended use. Perhaps the uncertain map regions should be used as a warning for site investigation rather than being relied upon for decision-making.

5 | Conclusion

This study demonstrates that a two-part ML framework effectively addresses the challenge of "structural zeros" in DTB mapping. The approach outperformed (RMSE = 5.80 m, $R^2 = 0.67$) both national IDW interpolation and coarse-resolution global ML models. The model is robust in shallow areas (< 10 m) and sparsely-sampled regions, with a low bedrock outcrop misclassification rate (4.05%) and strong agreement with observed depths (CCC = 0.80). However, significant challenges remain in mapping deeper ranges (> 30 m), where the model underestimates both depth (ME = 12.44 m) and uncertainty ($PICP_{90} < 80.6\%$). In these data-scarce ranges, conventional IDW interpolation proved more reliable. Our results highlight the necessity of carefully selected uncertainty metrics for visualization, and model evaluation via depth-stratification, rather than relying on aggregate measures alone. Given that shallow deposits dominate the Swedish landscape, we recommend a hybrid approach that integrates the predictive capability of ML in shallow or sparsely-sampled areas with the reliability of IDW interpolation in deep zones.

Author Contributions

Yiqi Lin: methodology, software, investigation, formal analysis, visualization, writing – original draft, writing – review and editing, conceptualization, data curation, validation. **Gustaf Peterson:** writing – review and editing, conceptualization, investigation, formal analysis, resources, methodology, data curation. **Cecilia Karlsson:** writing – review and editing, resources, data curation. **Florian Westphal:** writing – review and editing, methodology, validation. **William Lidberg:** writing – review and editing, methodology, resources. **Anneli M. Ågren:** writing – review and editing, conceptualization, methodology, funding acquisition, supervision.

Acknowledgements

This work was supported by FORMAS [grant number 2021-00713, 2021-00115] and Knut and Alice Wallenberg Foundation [grant number 2018.0259 (Future Silviculture)] and was partially supported by Marianne and Marcus wallenberg foundation, and the Wallenberg AI, Autonomous Systems and Software Program—Humanities and Society (WASP-HS), funded by the Marianne and Marcus Wallenberg Foundation, the Marcus and Amalia Wallenberg Foundation, and Kempestiftelserna. Use of generative artificial intelligence tools: Claude (Anthropic, Claude Sonnet 4.5) was used to refine manuscript text for clarity and conciseness, and to assist with Python programming, including code debugging and optimization for data analysis and visualization. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This work was supported by FORMAS [grant number 2021-00713, 2021-00115] and Knut and Alice Wallenberg Foundation [grant number 2018.0259 (Future Silviculture)] and was partially supported by Marianne and Marcus wallenberg foundation, and the Wallenberg AI, Autonomous Systems and Software Program—Humanities and Society (WASP-HS), funded by the Marianne and Marcus Wallenberg Foundation, the Marcus and Amalia Wallenberg Foundation, and Kempestiftelserna.

Disclosure

Software implementation and computing environment: To ensure reproducibility, all computational workflows were executed within a Docker containerized environment. Data processing was conducted using Bash shell scripts (v.5.1.16), machine learning modeling and core analysis were conducted in Python (v.3.10.6). Calculation of terrain indices was performed using WhiteboxTools (v.2.3.5) (Lindsay 2024). Random Forest classification was implemented using scikit-learn (v.1.7.2) (Pedregosa et al. 2011), and Quantile Regression Forest using the quantile-forest package (v.1.4.0) (Johnson 2024). Model interpretability was assessed using SHAP (SHapley Additive exPlanations) via the shap package (v.0.48.0) (Lundberg and Lee 2017). Variable-resolution uncertainty visualizations were generated using Quadmap (v.1.0.0) (Padarian and McBratney 2023), a standalone command-line tool. Geospatial data handling utilized GDAL (v.3.8.4) (Rouault et al. 2024) and Rasterio (v.1.4.3) (Gillies 2024). Computations were executed on a workstation with dual Intel Xeon Platinum 8362 processors (2.8 GHz, 64 cores total) and 1 TB RAM.

Data Availability Statement

The Python codes used to generate the results are available on GitHub (https://github.com/linyq017/A-Two-Part-framework-for-Depth-to-Bedrock-Prediction-and-Uncertainty-Assessment-in-Sweden/).

References

Agarwal, D. K., A. E. Gelfand, and S. Citron-Pousty. 2002. "Zero-Inflated Models With Application to Spatial Count Data." *Environmental and Ecological Statistics* 9, no. 4: 341–355. https://doi.org/10.1023/A:1020910605990.

Alley, R. B., K. M. Cuffey, and L. K. Zoet. 2019. "Glacial Erosion: Status and Outlook." *Annals of Glaciology* 60, no. 80: 1–13. https://doi.org/10.1017/aog.2019.38.

Anderson, D. W. 1988. "The Effect of Parent Material and Soil Development on Nutrient Cycling in Temperate Ecosystems." *Biogeochemistry* 5, no. 1: 71–97. https://doi.org/10.1007/BF02180318.

Arrouays, D., A. McBratney, J. Bouma, et al. 2020. "Impressions of Digital Soil Maps: The Good, the Not So Good, and Making Them Ever Better." *Geoderma Regional* 20: e00255. https://doi.org/10.1016/j.geodrs.2020.e00255.

Banwart, S. A., N. P. Nikolaidis, Y.-G. Zhu, C. L. Peacock, and D. L. Sparks. 2019. "Soil Functions: Connecting Earth's Critical Zone." *Annual Review of Earth and Planetary Sciences* 47, no. 1: 333–359. https://doi.org/10.1146/annurev-earth-063016-020544.

Björck, S. 1995. "A Review of the History of the Baltic Sea, 13.0-8.0 Ka BP." *Quaternary International* 27: 19–40. https://doi.org/10.1016/1040-6182(94)00057-C.

Bohn, M. P., and B. A. Miller. 2024. "Locally Enhanced Digital Soil Mapping in Support of a Bottom-Up Approach Is More Accurate Than Conventional Soil Mapping and Top-Down Digital Soil Mapping." *Geoderma* 442: 116781. https://doi.org/10.1016/j.geoderma.2024.116781.

Bonfatti, B. R., A. E. Hartemink, T. Vanwalleghem, B. Minasny, and E. Giasson. 2018. "A Mechanistic Model to Predict Soil Thickness in a Valley Area of Rio Grande Do Sul, Brazil." *Geoderma* 309: 17–31. https://doi.org/10.1016/j.geoderma.2017.08.036.

Brantley, S. L., W. H. McDowell, W. E. Dietrich, et al. 2017. "Designing a Network of Critical Zone Observatories to Explore the Living Skin of the Terrestrial Earth." *Earth Surface Dynamics* 5: 841–860. https://doi.org/10.5194/esurf-5-841-2017.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45, no. 1: 5–32. https://doi.org/10.1023/A:1010933404324.

Burrough, P. A., R. A. McDonnell, and C. D. Lloyd. 2015. *Principles of Geographical Information Systems*. OUP Oxford.

Chandler, B. M. P., and D. J. A. Evans. 2021. *Glacial Processes and Sediments*. Wiley-Blackwell. https://doi.org/10.1016/b978-0-12-409548-9.11902-5.

Chen, S., V. L. Mulder, M. P. Martin, et al. 2019. "Probability Mapping of Soil Thickness by Random Survival Forest at a National Scale." *Geoderma* 344: 184–194. https://doi.org/10.1016/j.geoderma.2019.03.016.

Chen, S., A. C. Richer-de-Forges, V. Leatitia Mulder, et al. 2021. "Digital Mapping of the Soil Thickness of Loess Deposits Over a Calcareous Bedrock in Central France." *Catena* 198: 105062. https://doi.org/10.1016/j.catena.2020.105062.

Courteille, L., L. Tardieu, N. Boukhelifa, E. Lutton, and P. Lagacherie. 2025. "What Is the Best Way to Communicate the Uncertainty of a Digital Soil Mapping Product? Some Lessons From an End-Users Survey." *Geoderma* 459: 117302. https://doi.org/10.1016/j.geoderma.2025.117302.

Daniels, J., and B. Thunholm. 2014. "Rikstäckande Jorddjupsmodell. (2014:14)." https://resource.sgu.se/produkter/sgurapp/s1414-rapport.pdf.

Devkota, S., N. M. Shakya, K. Sudmeier-Rieux, B. G. McAdoo, and M. Jaboyedoff. 2018. "Predicting Soil Depth to Bedrock in an Anthropogenic Landscape: A Case Study of Phewa Watershed in Panchase Region of Central-Western Hills, Nepal." *Journal of Nepal Geological Society* 55: 173–182. https://doi.org/10.3126/jngs.v55i1.22809.

Dietrich, W. E., D. G. Bellugi, L. S. Sklar, J. D. Stock, A. M. Heimsath, and J. J. Roering. 2003. "Geomorphic Transport Laws for Predicting Landscape Form and Dynamics." In *Prediction in Geomorphology*, 103–132. American Geophysical Union (AGU). https://doi.org/10.1029/135GM09.

Donner, J., and J. J. Donner. 2005. *The Quaternary History of Scandinavia*. Cambridge University Press.

Fawcett, T. 2006. "An Introduction to ROC Analysis." *Pattern Recognition Letters* 27, no. 8: 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Fraser, O. L., S. W. Bailey, M. J. Ducey, and K. J. McGuire. 2020. "Predictive Modeling of Bedrock Outcrops and Associated Shallow Soil in Upland Glaciated Landscapes." *Geoderma* 376: 114495. https://doi.org/10.1016/j.geoderma.2020.114495.

Fredén, C. 1990. *National Atlas of Sweden: Geology*. SNA Publishing.

Furze, S., A. M. O'Sullivan, S. Allard, T. Pronk, and R. A. Curry. 2021. "A High-Resolution, Random Forest Approach to Mapping Depth-To-Bedrock Across Shallow Overburden and Post-Glacial Terrain." *Remote Sensing* 13, no. 21: 4210. https://doi.org/10.3390/rs13214210.

Gibbard, P. L., and M. J. Head. 2010. "The Newly-Ratified Definition of the Quaternary System/Period and Redefinition of the Pleistocene Series/Epoch, and Comparison of Proposals Advanced Prior to Formal Ratification." *Episodes* 33, no. 3: 152–158. https://doi.org/10.18814/epiiugs/2010/v33i3/002.

Gillies, S. 2024. "Rasterio Documentation."

Gomes, G. J. C., J. A. Vrugt, and E. A. Vargas. 2016. "Toward Improved Prediction of the Bedrock Depth Underneath Hillslopes: Bayesian Inference of the Bottom-Up Control Hypothesis Using High-Resolution Topographic Data." *Water Resources Research* 52, no. 4: 3085–3112. https://doi.org/10.1002/2015WR018147.

Goodling, P., K. Belitz, P. Stackelberg, and B. Fleming. 2024. "A Spatial Machine Learning Model Developed From Noisy Data Requires Multiscale Performance Evaluation: Predicting Depth to Bedrock in the Delaware River Basin, USA." *Environmental Modelling & Software* 179: 106124. https://doi.org/10.1016/j.envsoft.2024.106124.

Heilbron, D. C. 1994. "Zero-Altered and Other Regression Models for Count Data With Added Zeros." *Biometrical Journal* 36, no. 5: 531–547. https://doi.org/10.1002/bimj.4710360505.

Hengl, T., J. M. de Jesus, R. A. MacMillan, et al. 2014. "SoilGrids1km—Global Soil Information Based on Automated Mapping." *PLoS One* 9, no. 8: e105992. https://doi.org/10.1371/journal.pone.0105992.

Hughes, A. L. C., R. Gyllencreutz, Ø. S. Lohne, J. Mangerud, and J. I. Svendsen. 2016. "The Last Eurasian Ice Sheets—A Chronological Database and Time-Slice Reconstruction, DATED-1." *Boreas* 45, no. 1: 1–45. https://doi.org/10.1111/bor.12142.

Iwahashi, J. 2025. "Depth to Bedrock in Japan: Insights From Borehole Data and Terrain Analysis Using Digital Elevation Models." *Progress in Earth and Planetary Science* 12, no. 1: 23. https://doi.org/10.1186/s40645-025-00698-3.

Jenny, H. 1994. *Factors of Soil Formation: A System of Quantitative Pedology*. Dover.

Johnson, R. A. 2024. "Quantile-Forest: A Python Package for QuantileRegression Forests." *Journal of Open Source Software* 9, no. 93: 5976. https://doi.org/10.21105/joss.05976.

Karlsson, C., G. Sohlenius, and G. P. Becher. 2021. "Handledning För Jordartsgeologiska Kartor Och Databaser Över Sverige. (2021:17)." The Geological Survey of Sweden.

Karlsson, C. S. J., I. A. Jamali, R. Earon, B. Olofsson, and U. Mörtberg. 2014. "Comparison of Methods for Predicting Regolith Thickness in Previously Glaciated Terrain, Stockholm, Sweden." *Geoderma* 226: 116–129. https://doi.org/10.1016/j.geoderma.2014.03.003.

Kasraei, B., B. Heung, D. D. Saurette, M. G. Schmidt, C. E. Bulmer, and W. Bethel. 2021. "Quantile Regression as a Generic Approach for Estimating Uncertainty of Digital Soil Maps Produced From Machine-Learning." *Environmental Modelling & Software* 144: 105139. https://doi.org/10.1016/j.envsoft.2021.105139.

Kasraei, B., M. G. Schmidt, J. Zhang, et al. 2024. "A Framework for Optimizing Environmental Covariates to Support Model Interpretability in Digital Soil Mapping." *Geoderma* 445: 116873. https://doi.org/10.1016/j.geoderma.2024.116873.

Kim, M. S., Y. Onda, J. K. Kim, and S. W. Kim. 2015. "Effect of Topography and Soil Parameterisation Representing Soil Thicknesses on Shallow Landslide Modelling." *Quaternary International* 384: 91–106. https://doi.org/10.1016/j.quaint.2015.03.057.

Kinkeldey, C., A. M. MacEachren, and J. Schiewe. 2014. "How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies." *Cartographic Journal* 51, no. 4: 372–386. https://doi.org/10.1179/1743277414Y.0000000099.

Kitterød, N.-O. 2017. "Estimating Unconsolidated Sediment Cover Thickness by Using the Horizontal Distance to a Bedrock Outcrop as Secondary Information." *Hydrology and Earth System Sciences* 21, no. 8: 4195–4211. https://doi.org/10.5194/hess-21-4195-2017.

Kleman, J., A. P. Stroeven, and J. Lundqvist. 2008. "Patterns of Quaternary Ice Sheet Erosion and Deposition in Fennoscandia and a Theoretical Framework for Explanation." *Geomorphology* 97, no. 1: 73–90. https://doi.org/10.1016/j.geomorph.2007.02.049.

Lacoste, M., V. L. Mulder, A. C. Richer-de-Forges, M. P. Martin, and D. Arrouays. 2016. "Evaluating Large-Extent Spatial Modeling Approaches: A Case Study for Soil Depth for France." *Geoderma Regional* 7, no. 2: 137–152. https://doi.org/10.1016/j.geodrs.2016.02.006.

Lambert, D. 1992. "Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing." *Technometrics* 34, no. 1: 1–14. https://doi.org/10.2307/1269547.

Lanni, C., M. Borga, R. Rigon, and P. Tarolli. 2012. "Modelling Shallow Landslide Susceptibility by Means of a Subsurface Flow Path Connectivity Index and Estimates of Soil Depth Spatial Distribution." *Hydrology and Earth System Sciences* 16, no. 11: 3959–3971. https://doi.org/10.5194/hess-16-3959-2012.

Lebedeva, M. I., and S. L. Brantley. 2013. "Exploring Geochemical Controls on Weathering and Erosion of Convex Hillslopes: Beyond the Empirical Regolith Production Function." *Earth Surface Processes and Landforms* 38, no. 15: 1793–1807. https://doi.org/10.1002/esp.3424.

Lidmar-Bergström, K. 1996. "Long Term Morphotectonic Evolution in Sweden." *Geomorphology* 16, no. 1: 33–59. https://doi.org/10.1016/0169-555X(95)00083-H.

Lilburne, L., A. Helfenstein, G. B. M. Heuvelink, and A. Eger. 2024. "Interpreting and Evaluating Digital Soil Mapping Prediction Uncertainty: A Case Study Using Texture From SoilGrids." *Geoderma* 450: 117052. https://doi.org/10.1016/j.geoderma.2024.117052.

Lin, L. I.-K. 1989. "A Concordance Correlation Coefficient to Evaluate Reproducibility." *Biometrics* 45, no. 1: 255–268. https://doi.org/10.2307/2532051.

Lindsay, J. 2024. "WhiteboxTools User Manual." https://www.whiteboxgeo.com/manual/wbt_book/preface.html?search=.

Liu, F., F. Yang, Y. Zhao, G. Zhang, and D. Li. 2022. "Predicting Soil Depth in a Large and Complex Area Using Machine Learning and Environmental Correlations." *Journal of Integrative Agriculture* 21, no. 8: 2422–2434. https://doi.org/10.1016/S2095-3119(21)63692-4.

Lukas, S., F. Preusser, D. J. A. Evans, C. M. Boston, and H. Lovell. 2017. "Chapter 2 the Quaternary." *Geological Society, London, Engineering Geology Special Publications* 28, no. 1: 31–57. https://doi.org/10.1144/EGSP28.2.

Lundberg, S., and S.-I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." https://doi.org/10.48550/arXiv.1705.07874.

Lundberg, S. M., G. Erion, H. Chen, et al. 2020. "From Local Explanations to Global Understanding With Explainable AI for Trees." *Nature Machine Intelligence* 2, no. 1: 56–67. https://doi.org/10.1038/s42256-019-0138-9.

Malone, B., and R. Searle. 2020. "Improvements to the Australian National Soil Thickness Map Using an Integrated Data Mining Approach." *Geoderma* 377: 114579. https://doi.org/10.1016/j.geoderma.2020.114579.

Malone, B. P., J. J. de Gruijter, A. B. McBratney, B. Minasny, and D. J. Brus. 2011. "Using Additional Criteria for Measuring the Quality of Predictions and Their Uncertainties in a Digital Soil Mapping Framework." *Soil Science Society of America Journal* 75, no. 3: 1032–1043. https://doi.org/10.2136/sssaj2010.0280.

Meinshausen, N. 2006. "Quantile Regression Forests."

Meyer, M. D., M. P. North, A. N. Gray, and H. S. J. Zald. 2007. "Influence of Soil Thickness on Stand Characteristics in a Sierra Nevada Mixed-Conifer Forest." *Plant and Soil* 294, no. 1: 113–123. https://doi.org/10.1007/s11104-007-9235-3.

Migoń, P., and K. Lidmar-Bergström. 2001. "Weathering Mantles and Their Significance for Geomorphological Evolution of Central and Northern Europe Since the Mesozoic." *Earth-Science Reviews* 56, no. 1: 285–324. https://doi.org/10.1016/S0012-8252(01)00068-X.

Minasny, B., and A. B. McBratney. 1999. "A Rudimentary Mechanistic Model for Soil Production and Landscape Development." *Geoderma* 90, no. 1: 3–21. https://doi.org/10.1016/S0016-7061(98)00115-3.

Mulder, V. L., M. Lacoste, A. C. Richer-de-Forges, and D. Arrouays. 2016. "GlobalSoilMap France: High-Resolution Spatial Modelling the Soils of France up to Two Meter Depth." *Science of the Total Environment* 573: 1352–1369. https://doi.org/10.1016/j.scitotenv.2016.07.066.

Odom, W., and D. Doctor. 2023. "Rapid Estimation of Minimum Depth-To-Bedrock From Lidar Leveraging Deep-Learning-Derived Surficial Material Maps." *Applied Computing and Geosciences* 18: 100116. https://doi.org/10.1016/j.acags.2023.100116.

Ohnuki, Y., C. Kimhean, Y. Shinomiya, and J. Toriyama. 2008. "Distribution and Characteristics of Soil Thickness and Effects Upon Water Storage in Forested Areas of Cambodia." *Hydrological Processes* 22, no. 9: 1272–1280. https://doi.org/10.1002/hyp.6937.

Oussou, F. E., C. E. Ndehedehe, N. Yalo, et al. 2025. "Improving the Depth-To-Bedrock Maps for Groundwater-To-Atmosphere Modeling in Africa." *Journal of Hydrology* 656: 132964. https://doi.org/10.1016/j.jhydrol.2025.132964.

Padarian, J., and A. B. McBratney. 2023. "QuadMap: Variable Resolution Maps to Better Represent Spatial Uncertainty." *Computers and Geosciences* 181: 105480. https://doi.org/10.1016/j.cageo.2023.105480.

Papadopoulos, G., P. J. Edwards, and A. F. Murray. 2001. "Confidence Estimation Methods for Neural Networks: A Practical Comparison." *IEEE Transactions on Neural Networks* 12, no. 6: 1278–1287. https://doi.org/10.1109/72.963764.

Patton, N. R., K. A. Lohse, S. E. Godsey, B. T. Crosby, and M. S. Seyfried. 2018. "Predicting Soil Thickness on Soil Mantled Hillslopes." *Nature Communications* 9, no. 1: 3329. https://doi.org/10.1038/s41467-018-05743-y.

Pawley, S. M., L. Atkinson, D. J. Utting, G. M. D. Hartman, and N. Atkinson. 2024. "Evaluating Spatially Enabled Machine Learning Approaches to Depth to Bedrock Mapping, Alberta, Canada." *PLoS One* 19, no. 3: e0296881. https://doi.org/10.1371/journal.pone.0296881.

Pedregosa, F., G. Varoquaux, A. Gramfort, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12, no. 85: 2825–2830.

Pelletier, J. D., P. D. Broxton, P. Hazenberg, et al. 2016. "A Gridded Global Data Set of Soil, Intact Regolith, and Sedimentary Deposit Thicknesses for Regional and Global Land Surface Modeling." *Journal of Advances in Modeling Earth Systems* 8, no. 1: 41–65. https://doi.org/10.1002/2015MS000526.

Pelletier, J. D., and C. Rasmussen. 2009. "Geomorphically Based Predictive Mapping of Soil Thickness in Upland Watersheds." *Water Resources Research* 45, no. 9: 319. https://doi.org/10.1029/2008WR007319.

Pennell, T., L.-P. Comeau, C. Feng, and B. Heung. 2025. "The Importance of Zeros in Digital Soil Mapping I: A Review." *Geoderma* 464: 117636. https://doi.org/10.1016/j.geoderma.2025.117636.

Pennell, T., L.-P. Comeau, J. Zhang, C. Feng, and B. Heung. 2026. "The Importance of Zeros in Digital Soil Mapping II: A Case Study of Depth-To-Bedrock Mapping in New Brunswick, Canada." *Geoderma* 465: 117651. https://doi.org/10.1016/j.geoderma.2025.117651.

Peterson, G. 2022. "Inlandsisen, Landhöjningen Och Landskapet: Ett Geologiskt Perspektiv. 213."

Poggio, L., L. M. de Sousa, N. H. Batjes, et al. 2021. "SoilGrids 2.0: Producing Soil Information for the Globe With Quantified Spatial Uncertainty." *Soil* 7, no. 1: 217–240. https://doi.org/10.5194/soil-7-217-2021.

Rouault, E., F. Warmerdam, K. Schwehr, et al. 2024. "GDAL (v3.9.1). Zenodo." https://doi.org/10.5281/ZENODO.5884351.

Rundgren, M. 2008. "Stratigraphy of Peatlands in Central and Northern Sweden: Evidence of Holocene Climatic Change and Peat

Accumulation." *GFF* 130, no. 2: 95–107. https://doi.org/10.1080/11035890801302095.

Schmidinger, J., and G. B. M. Heuvelink. 2023. "Validation of Uncertainty Predictions in Digital Soil Mapping." *Geoderma* 437: 116585. https://doi.org/10.1016/j.geoderma.2023.116585.

Shangguan, W., T. Hengl, J. Mendes de Jesus, H. Yuan, and Y. Dai. 2017. "Mapping the Global Depth to Bedrock for Land Surface Modeling." *Journal of Advances in Modeling Earth Systems* 9, no. 1: 65–88. https://doi.org/10.1002/2016MS000686.

Shrestha, D. L., and D. P. Solomatine. 2006. "Machine Learning Approaches for Estimation of Prediction Interval for the Model Output." *Neural Networks* 19, no. 2: 225–235. https://doi.org/10.1016/j.neunet.2006.01.012.

Silvestri, S., R. Knight, A. Viezzoli, et al. 2019. "Quantification of Peat Thickness and Stored Carbon at the Landscape Scale in Tropical Peatlands: A Comparison of Airborne Geophysics and an Empirical Topographic Method." *Journal of Geophysical Research: Earth Surface* 124, no. 12: 3107–3123. https://doi.org/10.1029/2019JF005273.

Stephens, M. 2020. "Chapter 1 Introduction to the Lithotectonic Framework of Sweden and Organization of This Memoir." *Geological Society, London, Memoirs* 50: 1–15. https://doi.org/10.1144/M50-2019-21.

Stroeven, A. P., C. Hättestrand, J. Kleman, et al. 2016. "Deglaciation of Fennoscandia." *Quaternary Science Reviews* 147: 91–121. https://doi.org/10.1016/j.quascirev.2015.09.016.

Tesfa, T. K., D. G. Tarboton, D. G. Chandler, and J. P. McNamara. 2009. "Modeling Soil Depth From Topographic and Land Cover Attributes." *Water Resources Research* 45, no. 10: 7474. https://doi.org/10.1029/2008WR007474.

The Geological Survey of Sweden. 2015. "hogsta-kustlinjen-beskrivning.pdf." https://resource.sgu.se/dokument/produkter/hogsta-kustlinjen-beskrivning.pdf.

The Geological Survey of Sweden. 2023. "Produktbeskrivning Berggrund 1:50 000 -1:250 000." https://resource.sgu.se/dokument/produkter/berggrund-50-250000-beskrivning.pdf.

The Swedish Land Survey. 2022. "Quality Description Laser Data." https://www.lantmateriet.se/globalassets/geodata/geodataprodukter/hojddata/quality_description_lidar.pdf.

van der Westhuizen, S., G. B. M. Heuvelink, D. P. Hofmeyr, L. Poggio, M. Nussbaum, and C. Brungard. 2024. "Mapping Soil Thickness by Accounting for Right-Censored Data With Survival Probabilities and Machine Learning." *European Journal of Soil Science* 75, no. 5: e13589. https://doi.org/10.1111/ejss.13589.

Vaysse, K., and P. Lagacherie. 2017. "Using Quantile Regression Forest to Estimate Uncertainty of Digital Soil Mapping Products." *Geoderma* 291: 55–64. https://doi.org/10.1016/j.geoderma.2016.12.017.

Wilford, J., and M. Thomas. 2013. "Predicting Regolith Thickness in the Complex Weathering Setting of the Central mt Lofty Ranges, South Australia." *Geoderma* 206: 1–13. https://doi.org/10.1016/j.geoderma.2013.04.002.

Wilford, J. R., R. Searle, M. Thomas, D. Pagendam, and M. J. Grundy. 2016. "A Regolith Depth Map of the Australian Continent." *Geoderma* 266: 1–13. https://doi.org/10.1016/j.geoderma.2015.11.033.

Wohlfarth, B., S. Björck, S. Funder, et al. 2008. "Quaternary of Norden." *Episodes* 31: 11. https://doi.org/10.18814/epiiugs/2008/v31i1/011.

Yan, F., W. Shangguan, J. Zhang, and B. Hu. 2020. "Depth-To-Bedrock Map of China at a Spatial Resolution of 100 Meters." *Scientific Data* 7, no. 1: 2. https://doi.org/10.1038/s41597-019-0345-6.

Yuan, H., M. Liu, M. Krauthammer, L. Kang, C. Miao, and Y. Wu. 2022. "An Empirical Study of the Effect of Background Data Size on the Stability of SHapley Additive ExPlanations (SHAP) for Deep Learning Models." https://doi.org/10.48550/arXiv.2204.11351.

Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith. 2009. "Zero-Truncated and Zero-Inflated Models for Count Data." In *Mixed Effects Models and Extensions in Ecology With R*, edited by A. F. Zuur, E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith, 261–293. Springer. https://doi.org/10.1007/978-0-387-87458-6_11.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** The DTB samples are sourced from eight databases. Sources marked with an asterisk (*) have a DTB of 0 m. **Table S2:** Performance metric comparison of IDW, the two-part model and the global model by Shangguan et al. (2017) by depth. **Figure S1:** Quaternary deposit (QD) map of Sweden with simplified classification. ML model prediction relies heavily on the bedrock outcrop class. **Figure S2:** Evaluation of RF classification thresholds ($T = 0.1, 0.2, ...0.9$) impact on fusion model performance. Left: (error distribution): The Kernel Density Estimate (KDE) plot shows the "cost" of missing an outcrop (false negative). It displays the predicted DTB values for samples that the RF misclassified as non-outcrops. As T increases, more outcrops are "missed," but the mean prediction error for these samples remains relatively low. Right (performance trade-offs): The MAE (blue line) reaches its minimum at a threshold of 0.4, which best balances two types of classification errors: false negatives (red) and false positives (yellow). However, as the threshold increases, MAE stays roughly the same, which suggests the misclassified outcrops did not have a huge impact on the overall performance. It could be that QRF is relatively good at predicting shallow DTB for outcrops (according to the left plot). This might suggest that a higher threshold (0.7 or 0.8) can be explored to minimize false positives.

IV

Remote Sensing Applications: Society and Environment 36 (2024) 101349

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

A fully automated model for land use classification from historical maps using machine learning

Anneli M. Ågren *, Yiqi Lin

Department of Forest Ecology and Management, Swedish University of Agricultural Science, Skogsmarksgränd, SE-907 36, Umeå, Sweden

ARTICLE INFO

Keywords:
Automation
Economic map
Extreme gradient boosting
Historical maps
Land use
Machine learning

ABSTRACT

Digital land use data before the age of satellites is scarce. Here, we build a machine learning model, using Extreme Gradient Boosting, that can automatically detect land use classes from an orthophoto map of Sweden (economic maps, 1:10 000 and 1:20 000) constructed from 1942 to 1988. Overall, the machine learning model demonstrated robust performance, with Cohen's Kappa and Matthews Correlation Coefficient of 0.86. The F1 values of the individual classes were 0.98, 0.95, 0.84, and 0.87 for graphics, arable land, forest, and open land, respectively. While the model can be used to detect land use changes in arable land, higher uncertainties associated with forest and open land necessitate further investigation at regional scales or exploration of improved mapping techniques. The code is publicly available to enable easy adaptation for classifying other historical maps.

1. Introduction

Understanding the land use and land cover change (LULUC) is vital to tackling large-scale challenges such as biodiversity loss, climate change, and food production (Winkler et al., 2021). Habitat loss caused by land use changes is recognized as one of the most pressing threats to biodiversity worldwide (Newbold et al., 2015, 2016; Loucks et al., 2008). Studies showed the lasting effect of historical land use on plant diversity. For example, grassland plant diversity in agricultural landscapes was shown to be controlled by land use more than 200 years ago (Gustavsson et al.,2007). Similarly, persistent differences between the vegetation of primary forests and post-agricultural secondary forests indicate that distribution patterns for many plant species still reflect the open agricultural environment of the nineteenth century (Bellemare et al.,2002). This study focuses on Sweden, which has undergone significant land use changes in the last century. The amount of arable land peaked between 1910 and 1950 at over 3.7 million hectares but decreased to 2.5 million hectares by 2007 (Swedish Board of Agriculture 2011) southern Sweden's total open land cover declined by 17% during this period (Auffret et al.,2018), with afforestation being the major driver. However, some arable land has also been converted into open land uses (Auffret et al.,2018). Historical landscape predicted present-day species richness better than current landscape, suggesting the long-term effects of historical land use (Auffret et al.,2018). Furthermore, climate change can compound plant responses to habitat conversion, as both shape the re-distribution of plant species in space and over time (Auffret and Svenning 2022). While studies on land-use changes have been performed in southern Sweden (Auffret et al.,2018), research in northern Sweden is still lacking. Understanding the intricate interplay between climate change and land use change necessitates studies in these areas, as the northern latitudes are experiencing even more rapid climate change (Previdi et al.,2021; Laudon et al., 2021).

Land use and land-cover change contributed to 12.5% of anthropogenic global carbon emissions from 1990 to 2010 (Houghton

* Corresponding author.
E-mail address: anneli.agren@slu.se (A.M. Ågren).

https://doi.org/10.1016/j.rsase.2024.101349
Received 13 June 2024; Received in revised form 27 August 2024; Accepted 8 September 2024
Available online 12 September 2024

et al., 2012). Drainage of wet organic soils for agricultural purposes increases greenhouse gas (GHGs: CO_2, N_2O, and CH_4) emissions by roughly 1 t CO_2 equivalents ha^{-1} $year^{-1}$, compared to undrained peat (Kasimir-Klemedtsson et al.,1997). Globally, GHG emissions from drained organic soils make up 32% of the total emissions from the agricultural sector but for Sweden the corresponding number is 44%. Rewetting drained peatlands is therefore an important climate mitigation option in Sweden. It has been suggested that forested peatlands previously used for agriculture are more nutrient-rich and emit higher levels of N_2O (Klemedtsson et al.,2005). Hence, former arable soils, now often afforested, might be good objects for rewetting. While the nutrient status of mineral forest soils and peatlands generally increases towards the south (Högberg et al., 2021) peatlands display the opposite patterns (Laudon et al., unpublished manuscript) with more eutrophic peatlands in northern Sweden compared to the south. Whether these nutrient-rich peatlands in the north are naturally nutrient-rich sites or former arable land, is still unknown.

Sweden's dependence on food imports from countries facing a broad range of potential climate risks (Horn et al., 2022) highlights the importance of domestic food production. Since the end of the 2nd World War, there has been a decrease in the number of farms and the total area of arable land due to farm closure and consolidation, particularly in northern Sweden, where 55–60% of the arable land was lost from 1944 to 1988 (Riksantikvarieämbetet 1996). However, the 2017 Swedish new food strategy (2016/17:104) aiming to increase domestic food production suggests that parts of the previously abandoned arable land might be recultivated.

Understanding historical land use across large areas, especially in northern Sweden, is crucial for addressing ecological, climate change, and food security challenges. Though present-day geographical databases are available on both national and EU level (Abdi 2020; Naturvårdsverket 2023), land use data pre-dating the launch of satellite imagery programs like Landsat in the 1970s is scarce (Mäyrä et al., 2023). The semi-automated approach proposed by Auffret et al. (2017) that digitized land use from historical maps was

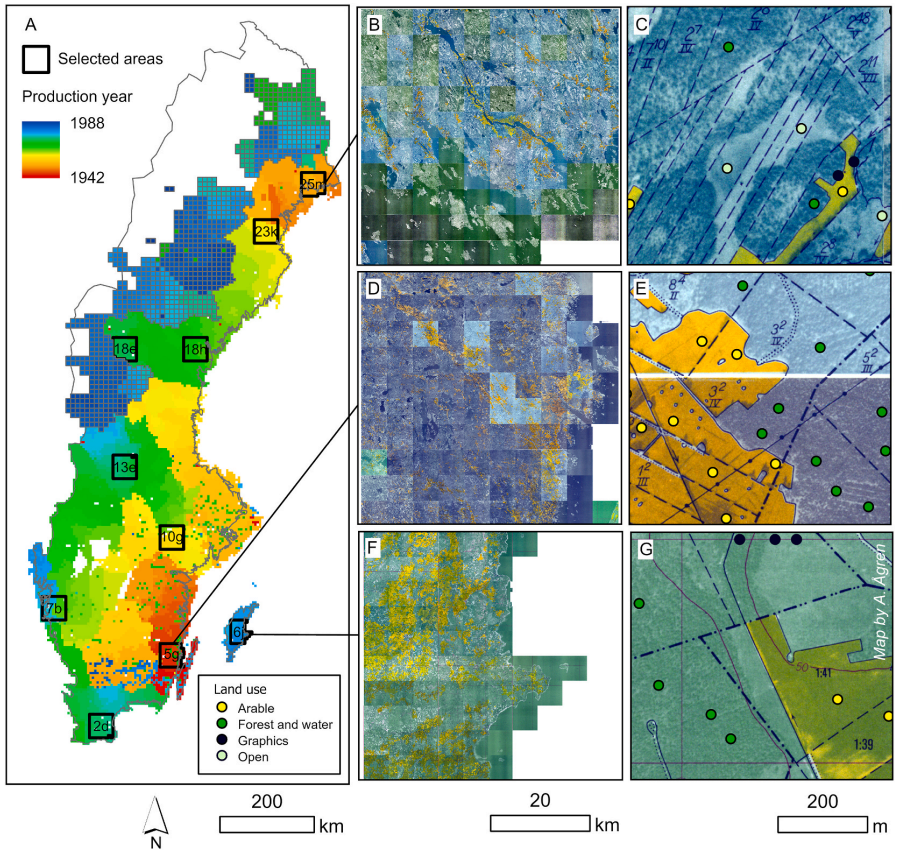


Fig. 1. A) Ten map regions for digitizing training and testing data were selected (A) to cover a range of production years and landscape types. B, D, F) Overview of three of the regions. Note the variability in color among the scanned maps. C, E, G) Example of digitized points, classified according to four classes, these are shown in the correct map scale (1:10 000). Note that the orthophotos are in the background of the map, with interpreted symbols drawn on top. The map tiles with a grey outline, in the northern inland of Sweden (A), are at the scale of 1:20 000, compared to 1:10 000 for the rest of Sweden. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

geographically limited to southern Sweden. Advancements in Machine Learning and Deep Learning algorithms, as well as the capability of supercomputers to handle high-resolution data for entire countries, are revolutionizing how maps are developed (Song et al., 2023). Recent successful applications in Sweden include automatic detection of ditch systems with high accuracy by deep learning (Laudon et al., 2022; Lidberg et al., 2023), machine learning generated soil moisture map at 2-m resolution (Ågren et al.,2021), which was further leveraged to create a novel peat map surpassing the accuracy of traditional methods (Ågren et al.,2022). These high-resolution maps provide invaluable new datasets for researchers and land-use planners working at large scales.

In this study, we developed a fully automated machine learning model for land use classification from scanned economic maps

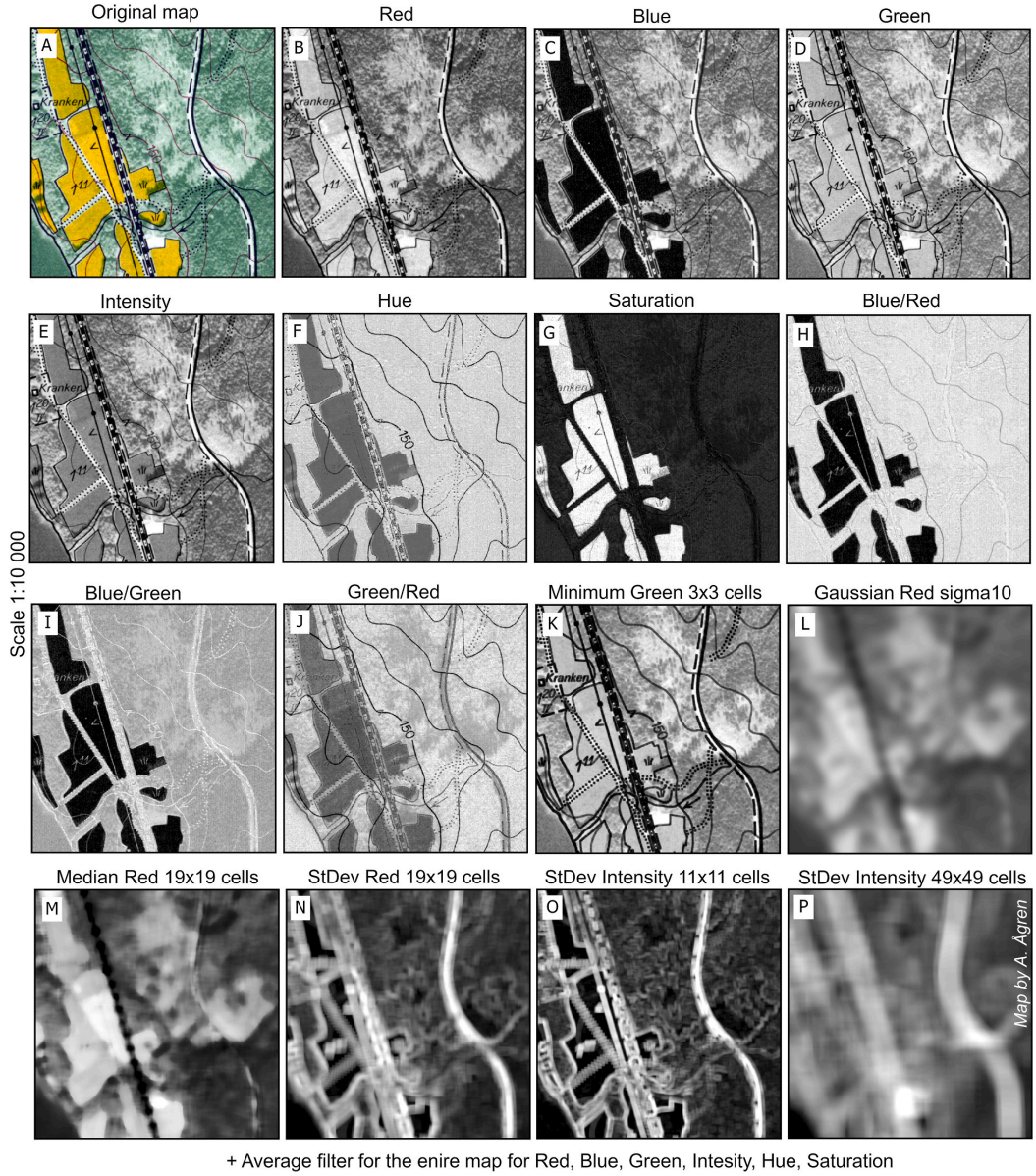


Fig. 2. The computed features from the scanned economic maps. 21 different features were derived. All maps show the same area, displayed at the scale of 1:10 000.

covering the entire country and evaluated the quality of the generated maps to determine their suitability for detecting land use changes over time.

2. Methods

The historical map series economic map (*Ekonomiska kartan in Swedish*) provides a detailed record of historical land use across most of the country (Fig. 1A), except the mountain range. In southern Sweden, each map sheet covers 25 km^2 (1:10 000). In northern inland Sweden, each sheet covers 100 km^2 (1:20 000) (Fig. 1A). The maps feature a monochrome aerial orthophoto in the background, with arable land colored in yellow and additional information such as roads, buildings, ownership boundaries, and place names in black (Fig. 1B–G). Maps from 1935 to 1947 have a black-and-white background photo, while for those produced after 1947, the photo was printed in green and white. More modern maps have topographical contour lines (Fig. 1G) and wetlands denoted with hatched areas. During the years 1992–1997, the maps were scanned as images by the Swedish Land Survey and released digitally as GeoTIFFs. The maps that were published online were produced from 1942 to 1988 (Fig. 1A).

2.1. Digitizing training/testing data for the machine learning model

The scanned economic maps from various production years and locations across Sweden were first explored visually. Due to variations in mapping methodologies, aging of the paper, and artifacts from printing and later scanning of the maps, land use categories can be expressed in a wide range of colors (Fig. 1B, D, F). To address this variability, 10 map regions were selected to cover a wide range of production years and landscape types (Fig. 1A). Within each region, up to 100 economic maps (Fig. 1B, D, F) were visually inspected to generate as diverse training/testing data as possible (Fig. 1C, E, G). Graphics were digitized at a scale of approximately 1:200–1:600 (i.e., zoomed in on pixel level), depending on the size of the object, while land-use classes were digitized at a scale of ca 1:2000–1:6000 (i.e., zoomed in on "landscape level"). Identifying land use classes relied primarily on color and texture within the orthophotos. Arable land was easily identifiable by its yellow color (though, in reality, it varied in shades from yellow to brown, orange, lime, and dark green with a hint of yellow). Generally speaking, 'forest and water' was darker (dark grey, blue, or green) than open land, with coniferous forest being darker and "spikier" than deciduous forest, which was lighter in color and "rounder" in appearance. Open land typically had the lightest colors. In the south, most open areas were fields, while in the north, most open areas were mostly peatlands or clear-cuts. However, land use was not always determined by color. The texture of the orthophoto on the map was another key determinant. Open areas often had a "smooth surface," i.e., little local variation in colors (Fig. 1C–G), while forests exhibited less uniform texture due to sunlit parts and shadows (Figs. 1C and 2A). Wet parts of peatlands, though often dark, could be visually distinguished from forests by their smoother homogeneous surface (Fig. 1C). Additionally, peatlands could also have bands with light and dark pixels, typical of string bogs. The minimum forest size, according to the definition from the Global Forest Resources Assessment (FRA), is 0.5 ha (ca 70 × 70 m) (Keenan et al., 2015). Open areas, however, can be smaller, such as a backyard with a lawn or the circular end of a cul-de-sac. Clear-cuts with a few seed trees were classified as open land (Fig. 1A), while a sparse natural forest was classified as forest land. Some of the classes were not easily identified even by human experts. We digitized numerous examples of such "tricky sites" with the aim of improving learning on them. Examples included dark, "patchy" green areas resembling forests but were situated in the mountains above the tree line where similar patterns were caused by field vegetation, or dark green areas on islands that were actually field vegetation and stoniness that had a forest-like appearance. A dense stand of even-aged forest could also be mistaken for an open field due to its homogeneous green impression. Based on map inspection, we expected arable land to be the easiest to classify and open land to be the most difficult. Therefore, we provided the model with additional examples of open land and forest. The final digitized points comprised 4312 for arable land, 2366 for 'forest and water', 4753 for open land, and 5223 for graphics. Local names provided additional information helpful for labelling, with descriptive names often ending with "-bog", "-moss", "-field", "-farm", "-forest", etc. In total we digitized 16 654 classified points from 938 maps, across the 10 study regions.

2.2. Extracting features to enhance the separation between land use classes

To enhance the separation between land use classes, several features were extracted from the original maps (Fig. 2A). The features were extracted with Whitebox tools v. 2.3.1 (Lindsay 2023), NumPy v. 1.24.4 (NumPy et al.,2023) and Rasterio v. 1.3.7 (Rasterio 2024). First, the composite images were separated into red, green, and blue bands (Fig. 2B–D), and converted to intensity, hue, and saturation (Fig. 2E–G). The ratios of the blue/red band, blue/green band, and green/red band were calculated (Fig. 2H–J). The graphics were most highlighted in the green band (Fig. 2D). To further enhance the graphics, a minimum filter on the green band with a kernel of 3 cells was applied (Fig. 2K). The arable areas appeared prominent in the blue band, saturation, blue/red, and blue/green features (Fig. 2C, G, I). 5 filters were calculated to improve the separation between forests and open land. In the clear-cut area in Fig. 2, individual trees/stones appeared as dark areas on the map. To prevent misclassification of these features as forest, we applied the Gaussian and median filters to the red band to smooth out color variations, as the background orthophoto was most visible in the red band (Fig. 2B). We used the 'fast almost Gaussian filter' (Kovesi 2010) algorithm with a Sigma of 10 cells (Fig. 2L), which provided a blurred effect visually similar to a Gaussian filter but was approximately 10 times faster. A median filter with a kernel of 19 cells was also applied to reduce color variations (Fig. 2M). 3 standard deviations were calculated to highlight texture. The standard deviations were calculated on the red band with a kernel of 19 cells and on the intensity band with kernels of 11 and 49 cells (Fig. 2N–P). To mitigate the impacts of color variations between maps (Fig. 1B, D, F) from different production years, the average red, green, blue, intensity, hue, and saturation for each individual map were derived. In summary, 21 different features were calculated for all 938 economic maps (Fig. 1). The values of these features were then extracted to the digitized point layer with classified land use.

2.3. Tuning, training, feature reduction, and application of the final XGBoost model

The dataset of 16 654 classified land use sites with extracted feature values was first divided into a training dataset (80%) and a separate test data set (20%) based on the digitized map tiles. All points from 750 randomly selected map tiles were used to train the model, and all points from the remaining 181 map tiles were set aside for evaluation. This approach allowed us to evaluate the model performance on "unseen maps" with different colors due to different production years. The machine learning algorithm Extreme Gradient Boosting (XGBoost) was chosen because it is designed to optimize both computational efficiency and model performance (Chen and Guestrin 2016). It incorporates parallelization and tree-pruning techniques and utilizes a more optimized algorithm that reduces memory usage and computational overhead.

The hyperparameter tuning of the model was performed with Bayesian optimization using the Optuna framework version 3.5.0 (Akiba et al.,2019). Specifically, the Tree-structured Parzen Estimator (TPE) sampler was applied on the training set (Watanabe 2023). The objective was to maximize the mean Matthews Correlation Coefficient (MCC) obtained from all 5 cross-validation folds over 100 trials. MCC gives the best measure of overall model performance for imbalanced multi-class datasets (Delgado and Tibau 2019). The overall quality of the model performance was high for our initial model, which included all 21 features and explored a wide hyperparameter search space using TPE. However, clear indications of overfitting were observed as the loss curves of training and testing datasets diverged (Fig. 3A). Various techniques exist to reduce model overfitting and improve the model's generalizability on unseen data. The model complexity was reduced by adjusting the bounding values of the hyperparameter search space. We decreased the number of features, maximum depth, and learning rate, and increased L1 and L2 regularization parameters. We implemented subsampling and early stopping during model fitting. Training features were iteratively removed based on their importance and collinearity. Hyperparameters were fine-tuned iteratively, with adjustments made in steps of 20%, while monitoring improvements in overfitting without compromising overall model performance. The total decrease in MCC after all iterations was <5‰. The process was stopped when an additional iteration resulted in a decrease of MCC by 1.5%. Finally, the best model hyperparameters from the tuning process were used to train the XGBoost model with 100 boosting rounds (Fig. 3B). The trained model with the 12 remaining features was applied to predict land use on unseen maps. For more details, we refer to the Python scripts on GitHub (https://github.com/anneliagren/Land-Use-Mapping-from-Historical-Maps). All processing was performed on two local Linux servers running Ubuntu 22.04.4. Server 1 is equipped with two Intel Xeon Platinum 8362 CPUs (2.8 GHz, together 64 cores/128 threads) and 1024 GB RAM. Server 2 is equipped with one AMD Ryzen ThreadRipper 3990X CPU (2.9 GHz, 64 cores/128 threads) and 256 GB RAM.

2.4. Evaluation metrics and SHAP-values

The model performance was evaluated on the separate test dataset of 181 maps (20%). We evaluated model performance based on these measures: Recall (user's accuracy), Precision (producer's accuracy), F1-value (Powers 2011), Matthews Correlation Coefficient (MCC) (Matthews 1975), and Cohen's Kappa (κ) (Cohen 1960). Recall answers the question; out of all the points for a certain class, how many were retrieved? Precision answers the question: out of all the points that the model retrieved for a certain class, how many were classified correctly? F1 is the harmonic mean between precision and recall (Powers 2011). Cohen's Kappa and Matthews Correlation Coefficient can often be used interchangeably, except for certain conditions, where Matthews Correlation Coefficient behaves more robustly (Delgado and Tibau 2019).

To gain insights into how the XGB model classifies land use, we calculated the Shapley Additive exPlanations (SHAP) values. Derived from game theory, SHAP values provide a unified approach to explaining the output of any machine learning model (Winter, 2002; Li 2022). Originally it was developed to calculate how much an individual player contributed to a game (Lundberg and Lee 2017). We calculated a global measure of feature importance, defined as the mean absolute value of each feature over all instances of the dataset using the Python library "Shap" version 0.44.1, on model predictions during testing. The absolute SHAP-values were calculated to compute the contribution of each feature (Fig. 2) to the detection of each land-use class and not if the feature contribute positively or negatively to the prediction (Li and Managi 2024).

2.5. Adding water features to the classified raster maps

We adopted the method used by Auffret et al. (2017) by adding water features using a modern vector layer, as water and forest could not be separated effectively. We opted for the more detailed water polygons from the Swedish Property map (1:12 500)

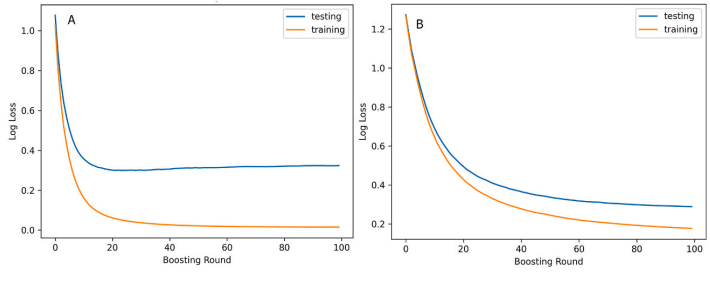


Fig. 3. Loss curves for A) Initial model with all 21 features. B) Final model after feature reduction and further tuning of the XGBoost model to reduce overfitting.

(downloaded 2021-08-09). The package gdalUtils was used to burn the modern water vector layer onto the final classified raster map (GDAL /OGR 2024). Given this approach, users of the maps should not use this map for detecting changes in water feature boundaries.

2.6. Post-processing

In some of the predicted maps, we observed the occasional irregularities in land use classification affecting one or a few cells (Fig. 4B). A post-processing step using the 'generalized classified raster' algorithm in Whitebox Tools v 2.3.0 was undertaken to remove the noise. This algorithm was preferred over a general smoothing filter (for example, Gaussian or majority), due to its capability to selectively modify without affecting feature boundaries within the classified raster (Lindsay 2023). The algorithm first identifies each contiguous group of cells in the input (i.e. a clumping operation) and then defines the subset of features that are smaller than the user-specified minimum feature size. In our case, the threshold was set to 9 cells to avoid removing small features representing footpaths. We used the method "longest" for class feature generalization, which reassigns small clumps with the same class value as the neighboring land use category with the longest shared border (Fig. 4C).

2.7. Mapping the northern inland of Sweden

While the 1:20 000 maps for the northern inland had been scanned, they had not yet been georeferenced and the borders around the maps remained (Fig. 5A). To address this issue, we implemented the following steps: 1) converted the original image to grayscale, 2) applied Otsu's thresholding (Otsu 1979) to create a binary image (Figs. 5B), 3) moved a bounding box of 5000 by 5000 pixels (red dashed box, Fig. 5) iteratively over the binary image, progressively decreasing the step size from 200, 100, 20, to 5 pixels. This process aimed to find the region of the target size that contains the most white pixels, corresponding to the map content area in the original map (green area), 4) cropped the original map with the location of the bounding box, and 5) the cropped image (Fig. 5C) was georeferenced using the shape file with the location of each file, in Fig. 1A. Apart from these extra steps, the maps were processed identically (calculating features and applying the classification model) as the rest of Sweden.

2.8. Comparison with Auffret et al. (2017) for the south of Sweden

Five of the digitized study areas overlapped with the area previously classified by Auffret et al. (2017) in their work on semi-automated land use digitization in south Sweden using R (https://github.com/AGAuffret/HistMapR/). We compared our manually digitized points with the predictions by Auffret et al. (2017), focusing only on points classified as arable land, open land, and forest. All points could be used for evaluation purposes since our digitized points were independent of the training data used in Auffret et al. (2017). Some key differences existed between the two methods. While both studies incorporated water features from modern-day maps after the final predictions, Auffret et al. (2017) included a separate class for pixels that could not be classified, labeled as "borders". Due to these class discrepancies, direct comparisons of the total maps were impossible. However, we could perform comparison for the individual classes of arable land, forest, and open land, which were present in both studies.

2.9. Optional removal of graphics

Finally, we provided the option to remove graphic elements (e.g. contour lines, ownership borders, place names) from the maps. Classifying graphics as a separate class provided us the option to use it as a mask to remove them while preserving as much of the original land use data as possible. First, all graphics were converted to a value indicating missing data, commonly referred to as 'Nodata'. Then we employed the GDAL (GDAL /OGR 2024) function 'gdal.FillNodata' to fill the 'NoData' values. It started by searching within a neighborhood of 3 x 3 pixels and exponentially increasing neighborhood sizes (3, 6, 12, 24, 48, etc. pixels) until all 'NoData' cells were filled. Progressively increasing the search area allowed for more efficient computations while preserving local variations in the data.

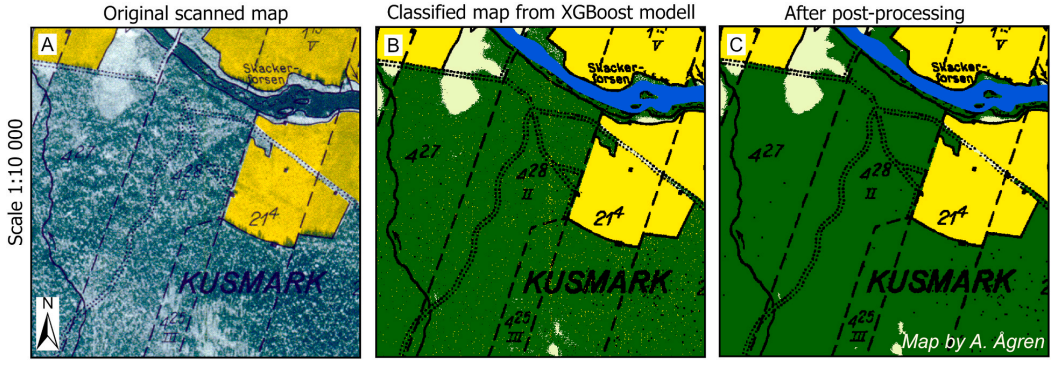


Fig. 4. A) The figure shows an example of the original scanned map, B) the classified land use map from the XGBoost model, and C) the map after the post-processing step after the removal of individual misclassified small "clumps".

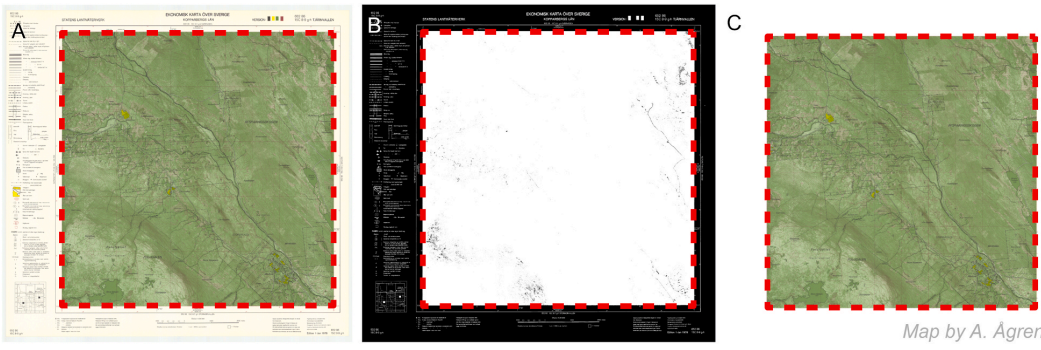


Fig. 5. Example of automated cropping of images. A) Original map. B) Binary image from Otsu's thresholding. C) Cropped map.

3. Results

3.1. Quality measures of the maps

The overall quality of the classified land use maps was high, as indicated by the statistical evaluation using the test dataset consisting of 181 maps (Table 1). While we presented accuracy for comparisons with other studies, we considered κ and MCC to be more robust metrics for the overall quality of the map, given the slightly unbalanced nature of our dataset. The accuracy for our final model was 0.90, κ and MCC both achieved a score of 0.86. Graphics and arable land had the highest performance for the individual classes, with F1 scores above 0.90. The most difficult land use to distinguish/identify was the forest and open land, with F1-scores of 0.84 and 0.87, respectively.

3.2. Feature importance with SHAP values

The SHAP values (Fig. 6) quantify the contribution of each feature to XGBoost's ability to accurately classify each land-use class. The length of each bar within a feature represents the feature's relative importance for individual classes. The cumulative length of all bars for a given feature reflects its overall impact on the model across all classes. For graphics, the minimum filter on the green band, green/red, and average green were the most important features. Blue/red and green/red contributed the most to classifying arable

Table 1
Precision, Recall, F1-score on the testing dataset for each class, and the overall performance measures of the model; Accuracy, Cohen's kappa (κ), Matthews Correlation Coefficient (MCC), and number of observations (n). The table presents the measures from the final XGBoost model after feature reduction and further tuning of the model to reduce overfitting (see section 2.3).

	Precision	Recall	F1-score	Accuracy	κ	MCC	n
Graphics	0.97	0.98	0.98				716
Arable	0.97	0.92	0.95				468
Forest	0.83	0.86	0.84				930
Open	0.88	0.86	0.87				976
Total			**0.91**	**0.90**	**0.86**	**0.86**	

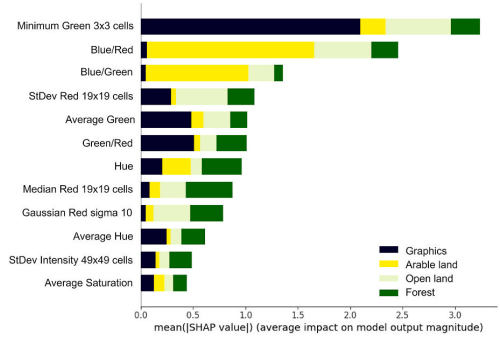


Fig. 6. The absolute SHAP values for the XGBoost model.

land. Classifying open land relied on a wider range of features, with the most important ones being the minimum filter on the green band, the standard deviation of the red band with a kernel of 19 cells, blue/red, and a Gaussian filter with a sigma of 10. Forest classification primarily involved the median filter on the red band with a 19 cells kernel, hue, and a Gaussian filter with a sigma of 10.

3.3. Visualization of the predicted maps

Visual inspection of the predicted maps and the original maps revealed a high degree of accuracy (Fig. 7A and B). There was also a high level of agreement between our new fully automated land use classification with the classification by Auffret et al. (Fig. 7B and C). The primary difference between our methods lay in the handling of graphics, which we maintained as a separate class, while Auffret et al. masked them out during preprocessing. This was especially evident in urban areas with high density of graphics (Fig. 7D–F). Both models delineated open and forest land based on the brightness or darkness of the orthophoto, which led to some limitations. Light-colored deciduous forests were misclassified as open areas (Fig. 7G–I), and some dark open areas (often wet soils) were misclassified as forests (Fig. 7J–L). The percentage of graphics ranged 0–27% on the evaluation maps. We observed an increase in graphics over time, from an average of ca 5% on the maps produced in the early 40s to ca 12% in the late 70s (Fig. 8).

3.4. Statistical comparison with Auffret et al., (2017) for the south of Sweden

In addition to the visual comparison with Auffret et al. (2017), we also compared Auffret et al.'s maps statistically with our manually digitized points. We compared the performance of the individual classes' arable land, forests, and open land, excluding classes that were not represented in both maps. Both models achieved high recall, or user's accuracy, for arable land, specifically, 0.92 in our study and 0.93 in Auffret et al. (2017). This indicates that 92% and 93% of the actual arable land were correctly identified by the models. Our model demonstrated higher precision, or producer's accuracy, for arable land 0.97 (Table 1) compared to Auffret et al.'s 0.58 (Table 2). This metrics measures the proportion of correctly identified arable land out of all pixels identified as arable land. For forest and water, our model had a lower recall 0.86 but higher precision 0.83 in comparison to Auffret et al.'s higher recall 0.91, and lower precision 0.58. For open areas, our model outperformed that of Auffret et al.'s by ca 10%, with recall and precision values of 0.86 and 0.88, respectively (Table 1), compared to 0.70 and 0.75, for Auffret et al. (Table 2). Based on F1 scores, which is the harmonic mean for precision and recall, our model generally performed better than the previous model by Auffret et al., with F1 scores 0.24 higher on arable land, 0.19 higher for forest, and 0.14 higher for open areas (Table 2).

3.5. Masking away graphics

By using the classified graphic as a mask (Fig. 9B), we could remove the graphics without changing any other cells in the maps and, therefore, maintain the integrity of the land use classes (Fig. 9C). For the majority of the map tiles three iterations (filter size 3x3, 6x6, and 12x12) was enough to fill all NoData cells, and mask away all graphics. Another common way to remove graphics is to use a general smoothing filter (Fig. 9D) as applied by Auffret et al., (2017). Which is similar to our post-processing step (Section 2.6, Fig. 4), but, with a larger kernel. However, this affects more pixels across the landscape and introduce more uncertainty in the maps (Fig. 9D).

4. Discussion

4.1. Land use classification model

4.1.1. Objective and significance of land use classification

The purpose of classifying the land use in these historical maps is to generate a dataset that can be used to detect changes in land use (Auffret et al.,2018) and learn more about the interactions on species diversity in a changing world where both climate and land use are changing (Auffret and Svenning 2022). We also aim for these maps to serve as planning tools for rewetting nutrient-rich peatlands, as historical land use may have long-lasting effects on soil carbon (Schulp and Verburg 2009), microbial diversity (Li et al., 2021), and N_2O emissions (Klemedtsson et al.,2005). This article focuses on the method development of detecting land use in historical maps and assessing the machine learning model classified maps for their intended use.

4.1.2. Modern land use data quality and XGBoost model assessment

As background to our classification, we examined the quality of modern-day land use data for Sweden (NMD), at 10 m resolution, derived from a combination of 41 different data layers of satellite imagery, laser data, digital terrain indices, current vector maps, etc. (Naturvårdsverket 2023). The NMD maps identify forests, wetlands, arable land, other open land, urban areas, and water. The quality of the NMD maps was assessed based on regional precision for Sweden, divided into 4 regions based on field data (Nilsson et al., 2020). The precision was (the range among the 4 regions is given in the in brackets): forest (0.93–0.97), arable land (0.96–0.97), wetland (0.91–0.98), other open land (0.38–0.75), urban land (0.46–0.58), water (0.99–100). Nilsson et al. (2020) further concluded that a precision above 0.8 is very good, 0.7–0.8 is good, 0.6–0.7 is acceptable, but below 0.6 there is a poor agreement between the map and reality.

Our machine learning model could not be directly assessed with field data, due to the lack of accurate positioning of the available field data during the map production years. Hence, we partitioned our digitized point layer into a separate evaluation dataset. Our XGBoost model had a precision of 0.97, 0.83, and 0.88 for arable land, forest and open areas, respectively. The inherent uncertainty of the original paper maps (García-Alvarez et al.,2019) is acknowledged but remains nonquantifiable, as the only information given by the National Land Survey on this, is that newer maps are more accurate than old maps.

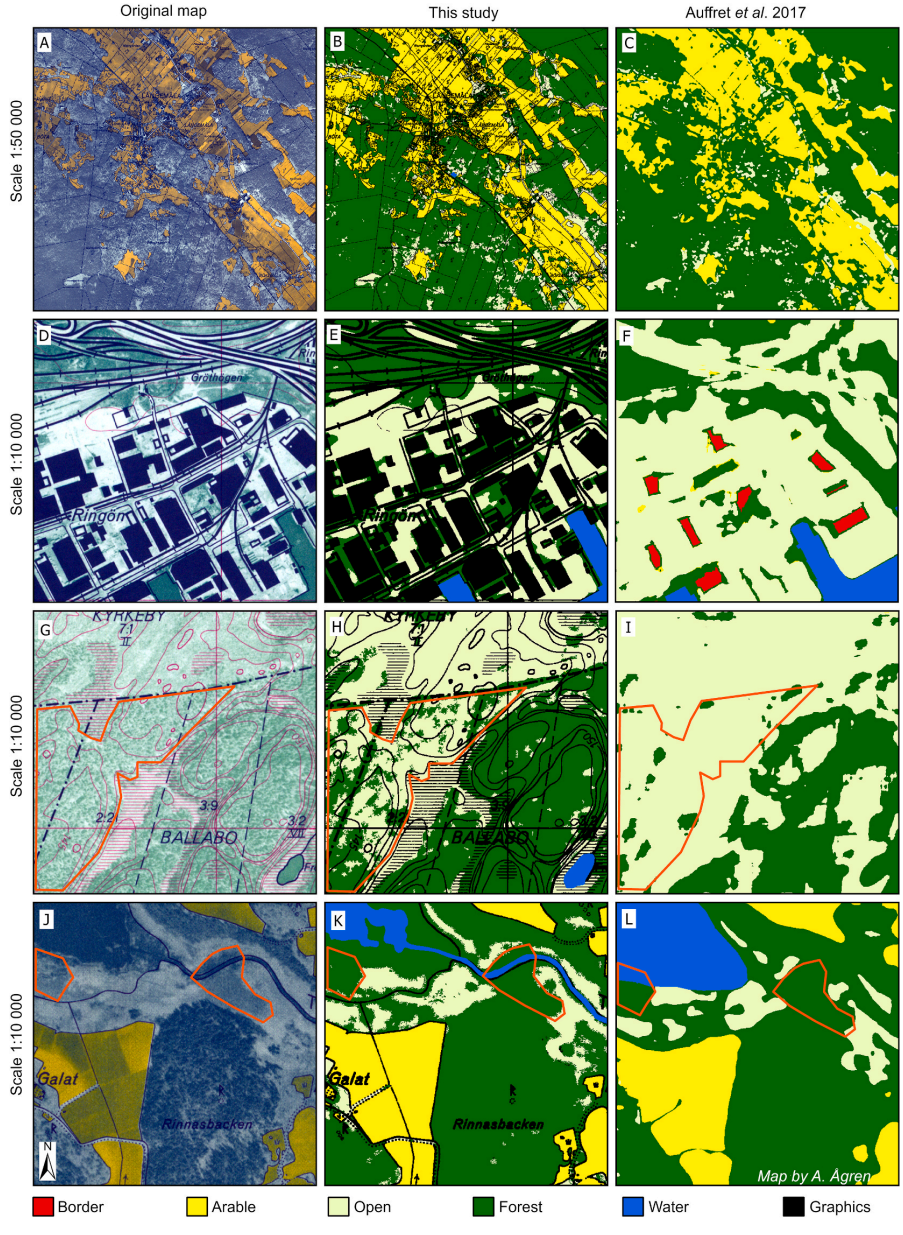


Fig. 7. Comparison of land use classification for four different areas in southern Sweden. The left panels show the original scanned maps, the center panels our land use classification, and the right panels display the classification by Auffret et al., (2017). Panels A–C show an entire map tile compressed to 1:50 000 while, panels D–L show zoomed-in sections of maps originally produced at a scale of 1:10 000. Orange polygons exemplify misclassified areas. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4.1.3. Challenges and solutions in historical orthomap classification

Both our XGBoost method and Auffret's model encountered difficulties in correctly mapping dark open areas (Fig. 7K and L), and light forests (Fig. 7H and I). In a separate model run (not shown) we digitized a further 5500 points representing dark open areas and

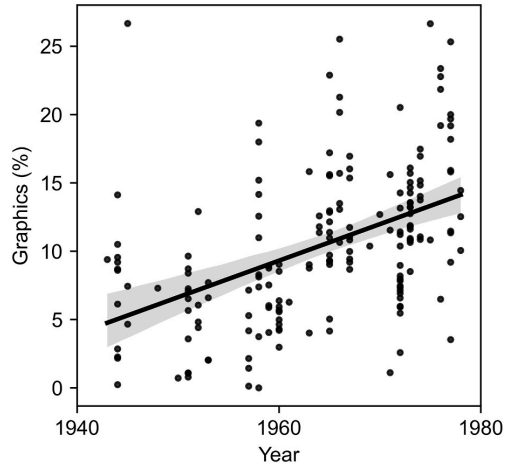


Fig. 8. Linear regression reveals a significant increase in graphics content of historical maps over time ($p < 0.001$, $R^2 = 0.22$). The grey shaded area depicts the 95% confidence interval for the regression line.

Table 2
Precision, Recall, F1-score, and number of observations (n) for the map by Auffret et al. (2017). The map was statistically evaluated against all of our training and testing points where the study areas overlapped. Furthermore, we show the difference between our map and Auffret et al.'s. A positive number indicates that our map has a higher quality, while a negative number indicates that Auffret's map has a higher quality.

	Precision	Recall	F1 score	n	Difference in precision	Difference in recall	Difference in F1
Arable	0.58	0.93	0.71	1740	0.39	−0.01	0.24
Forest	0.51	0.91	0.65	3022	0.32	−0.05	0.19
Open	0.75	0.70	0.73	3246	0.13	0.16	0.14

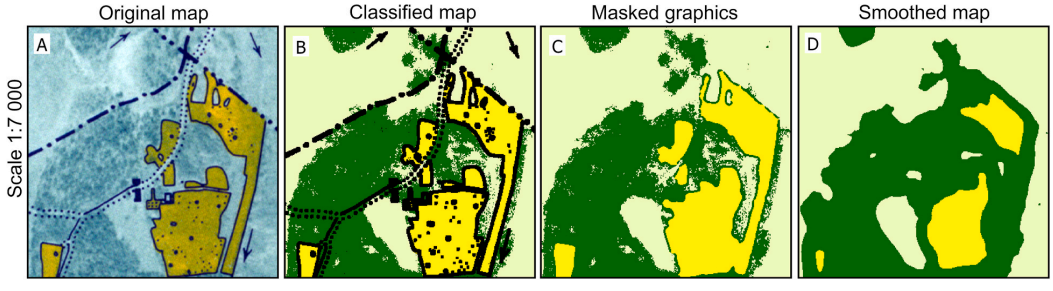


Fig. 9. All four maps show the same area. A) Original map. B) Our classified map. C) The map after masking graphics in our study. D) Classified map by Auffret et al., (2017), who applied a general smoothing filter in the first step of land use classification.

light forests. The intention was to improve the classification with additional training data, allowing it to learn to differentiate land use based on both color (light vs. dark) and texture (smooth vs. "textured"). However, this attempt to improve performance instead introduced noise in the model, leading to the deterioration of predictive performance. Hence, we opted to revert back to the original, more well-balanced dataset. Our analysis suggested that classification based on light and dark areas of the orthophoto would yield more accurate predictions overall. We observed illumination non-uniformity in some of the scanned maps (visible in Fig. 1B), characterized by lighter and darker stripes across the maps (Baltsavias 1994). These artifacts affected the land use classification in some of the images. Underexposed dark bands were over-classified as forest, while overexposed light bands were over-classified as open land. This banding was most pronounced for the northern inland maps, but, the same patterns were also present in the maps by Auffret et al., (2017) (not shown). To potentially improve the classification of forest and open land, we recommend evaluated a novel deep learning model Correcting Uneven Illumination Network (CUI-Net) for correcting light and dark stripes in the original images (Chao et al., 2023).

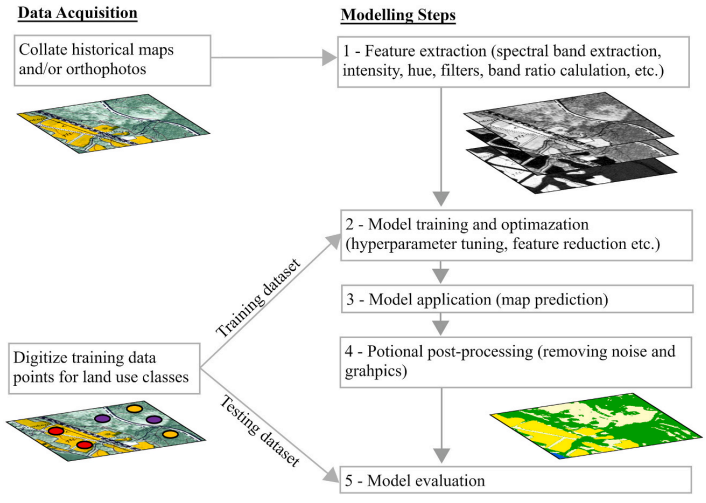


Fig. 10. Modeling Framework for Classifying Land Use from Pre-Satellite Historical Data. Python scripts for all calculations and instructions are found on https://github.com/anneliagren/Land-Use-Mapping-from-Historical-Maps.

The extent of graphic coverage on the maps depended on the production year, with more graphics added over time (Fig. 8). Among our 181 evaluation maps, graphics covered between 0 and 26%. Since graphics obscured the underlying photo, it was challenging to determine the land use hidden underneath. However, by assigning graphics to a separate class, we could quantify the uncertainty introduced by the overall coverage of the graphics (Fig. 8). Having a separate class allows users to easily mask the graphics away and interpolate with the surrounding land use if needed (Fig. 9). While the masking and interpolating method works well for rural areas where most of the graphics are wetlands, topographic lines, and ownership boundaries, it will introduce a larger bias for urban areas (Fig. 7D–F). Hence, we opted to retain the graphic as a separate class, allowing potential future map users the flexibility depending on their research questions. Additionally, we have included a Python script to facilitate efficient removal of graphics when desired. We observed a further complication in the analysis. Recent maps, characterized by a higher graphic content (Fig. 8), exhibited a reduction in background orthophoto contrast to enhance graphic visibility. This contrast reduction appeared to have negatively impacted land-use classification performance, particularly for forest and open land categories, as evidenced by the test map evaluations.

The automated clipping of the maps of northern inland generally performed well. However, some maps experienced distortion from scanning (often a minor rotation). The misalignment with the 5000x5000 cell window has, in some cases, resulted in partial map loss or map border inclusion in the clipping area. The location uncertainty introduced by scanning distortion amounted to ca 1%. To make the clipping more efficient, we enforced early stopping when the bounding box was within 5 m of the optimal placement. Except for the distorted map areas, we argue that this method still surpasses modern land use classification, which has a 10 m resolution.

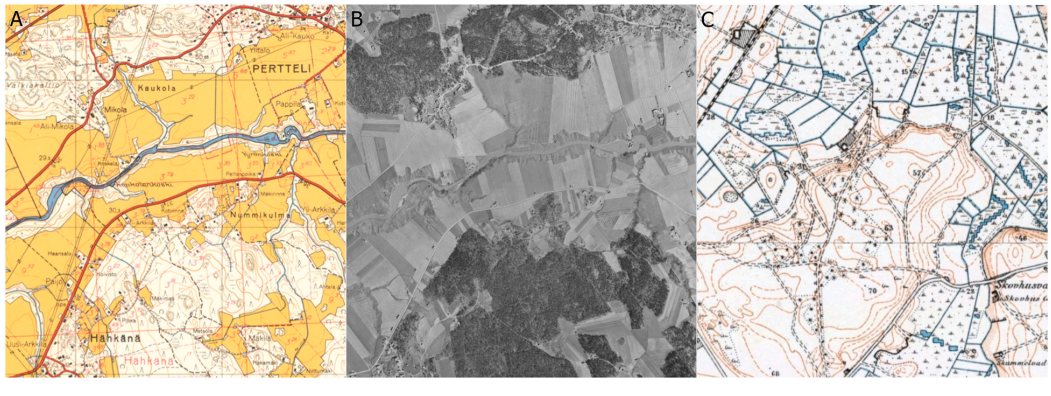


Fig. 11. Example of a Finnish basic map 1:20 000 (1964) (A) and a historical orthophoto (1959) (B) of the same area, from National Land Survey of Finland. The Danish topographical map Lave målebordsblade (C) 1:20 000, 1901–1971, Generalstaben.

4.2. Can the maps be used for detecting land-use changes over time?

4.2.1. Arable land

Ensuring the trustworthiness of land-use maps is important as they are crucial tools for understanding some of the world's most pressing challenges like biodiversity loss and climate change (Griffin 2020). Accurate assessment of land-use change from maps necessitates the estimation of map uncertainty (Prestby 2023). To properly evaluate the uncertainty, we calculated several measures of uncertainty both for the entire map and for the individual classes (Table 1). The uncertainties of the maps also have to be considered in the context of land-use change over time. Statistical records indicate a 21% decrease in arable land in Sweden from the period 1910–1950 to the year 2023 (Swedish Board of Agriculture 2011; Swedish Official Statistics 2023). While this data comes from the farmers' self-reporting to the board, which lacked georeferencing (before 1998), most of the individual fields can be identified on our model-produced maps. Our model had an F1-value of 0.95 for arable land (Table 1), and modern-day land use maps had a precision of 0.96–0.97 (Nilsson et al., 2020). These metrics suggest that a few local arable sites might be mislabeled by the model, and one should not trust the maps implicitly on the local level. However, when examining arable land at regional scales, the low uncertainty associated with both our model-produced maps and modern land use maps suggests minimal impact on the overall results.

4.2.2. Forest and open land

For forest land and open land, which were harder to detect, we can estimate that our map has an uncertainty of around 13–16% based on the F1 values (Table 1). In comparison, the modern-day land use maps had 3–7% uncertainty based on the producer's accuracy for forest land, while the uncertainty for wetland was 2–9% and other open land 25–68% (Nilsson et al., 2020). The higher uncertainties for forest and open land in both our and existing maps, limit the effective detection of small-scale land-use changes. Based on field measurements from the National Forest Inventory the productive forest land has increased by 5% from 1926 to 2020, while mires decreased by 13% over the same period (Swedish Official Statistics). While Auffret et al. (2018) reported a 17% decline in total open land cover in southern Sweden, our evaluation indicated an uncertainty of 27% for open land in their maps (Table 2). Furthermore, the broad range of uncertainty (25–68%) associated with open land in modern maps (Nilsson et al., 2020) undermines the certainty of this reported land-use change. Despite having a higher overall quality (14–24% more accurate (Table 2)), our map still exhibits an uncertainty of 13% for open land (Table 2). We therefore advise using the map with caution. Given the higher uncertainty of the classes forest and open land for both modern and historical maps, the suitability of the maps for detecting land use changes for those land use classes is questionable. Therefore, we recommend users conduct further investigations at more regional and local scales to assess the uncertainty further.

4.3. Applying the model to other countries

Long-term records of land use change require alternative data sources beyond satellites, as satellites dedicated to land-cover observation were not launched until 1972 (Belward and Skoien 2015). Historical maps and orthophotos provide valuable tools for detecting land use changes extending back more than 250 years. This approach is demonstrated by a Czech case study where historical land use was manually interpreted and digitized as vector data (Skalos et al.,2011). Manual digitization remains a viable option for small study areas (Skalos et al.,2011) and smaller countries, such as Latvia (65 300 km^2) (Piskinaite and Veteikis 2023), or even for larger countries when using low resolutions, as shown in a study of India at 5 arc-minute resolution (Tian et al., 2014). However, detailed manual interpretation of historical maps and orthophotos at scales of 1:10 000 to 1:20 000 is prohibitively time-consuming and expensive to perform across entire countries. Semi-automated methods using R and GRASS have been employed to detect land use in southern Sweden and Italy (Auffret et al.,2017; Gobbi et al., 2019). In this study, we developed a fully automated method for land use classification that could be applied across an entire country (470 000 km^2) at high resolution (1 m) by leveraging efficient calculations with parallel processing using Python scripts running on Linux servers. This approach allowed us to map the entire country within ca. two months of full time work. The same methodology can be adapted to other countries with similar datasets (Fig. 10).

Monochrome orthophotos became widespread in the 1930s–1960s and changed how maps were constructed. The production of photomaps began after the First World War, initially intended for military purposes but eventually expanding to civil applications as well (Baltsavias 1996). Maps based on orthophotos gained widespread popularity because orthophotomaps provided a wealth of ground details and information that could not be captured by standard vector maps (McKenzie 1973). The Swedish economic map is an orthophotomap with the orthophoto in the map background and interpreted vectorgraphics on top. Most European countries has similar maps that are in the scale of ca 1:10 000–1: 50 000, and were drawn based on aerial photography. However, in some European countries, these datasets can sometimes exist separately. In the Finnish basic map (1:20 000, 1949–1998, National Land Survey of Finland) (Fig. 11A), the land uses water, arable land, and graphics can be classified from the maps, while open land and forest land can be classified from the orthophotos (Fig. 11B) and then used to update the white background areas in the map (Fig. 11A). Unlike many modern maps, early geographical information was not standardized across countries. As a result, our tuned model cannot be directly applied to maps from other countries. However, the methods (https://github.com/anneliagren/Land-Use-Mapping-from-Historical-Maps) can be easily transferred to any country with historical maps and/or orthophotos from the same period (Fig. 10). For maps that use a greater number of symbols to indicate different land uses within polygons, such as the Danish topographical map (Lave målebordsblade, 1:20 000, 1901–1971, Generalstaben) (Fig. 11C), we suggest that a deep learning approach might be more effective. The complexity and diversity of symbols in such maps could be better interpreted by advanced deep learning techniques, which classify land use based on image segmentation techniques, possibly with a U-Net model (Mäyrä et al.,2023).

5. Conclusions

Our method demonstrates the potential for automated, high-quality arable land mapping. The maps can thus facilitate the identification of afforested former arable sites, which are likely to be nutrient-rich sites. Such nutrient rich sites often displays a high growth and yield, high biodiversity, and can be good candidates for rewetting, which can lead to tradeoffs for best management practice. Very little is known about nutrient-rich drained peatlands in northern regions; this map is, therefore, an important contribution to further research. However, with the developed methods we could not get a good estimation of open and forest land use change, despite F1 values above the threshold for "very good" suggested by Nilsson (>80%), as the magnitude of change for these land uses fall within the range of uncertainty estimates. To promote further research and broader application, the code for this method is publicly available on GitHub (https://github.com/anneliagren/Land-Use-Mapping-from-Historical-Maps).

Code availability

All code is openly available and published on GitHub under The Creative Commons CC0 Public Domain Dedication (https://github.com/anneliagren/Land-Use-Mapping-from-Historical-Maps)

Data sources

The original maps can be downloaded from: https://www.lantmateriet.se/sv/geodata/vara-produkter/produktlista/ekonomiska-kartan/

Declaration of generative AI

During the preparation of this work the authors used GitHub Copilot to assist in writing the Python scripts. Following this, the authors thoroughly reviewed and edited the content and take full responsibility for the content of the publication.

Ethical statement

No other related manuscripts are under consideration or in press elsewhere by the authors. We agree and comply with the general terms and general obligations of the authors.

None of the authors of this article have coauthored an article with any of the suggested reviewers.

CRediT authorship contribution statement

Anneli M. Ågren: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yiqi Lin:** Writing – review & editing, Software.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Anneli Agren reports financial support was provided by Swedish Reserch Council Formas. Anneli Agren reports financial support was provided by Knut and Alice Wallenberg Foundation. Anneli Agren reports a relationship with Knut and Alice Wallenberg Foundation that includes: funding grants. Anneli Agren reports a relationship with Swedish Research Council Formas that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All code is openly available and published on GitHub under The Creative Commons CC0 Public Domain Dedication (https://github.com/anneliagren/Land-Use-Mapping-from-Historical-Maps)

Acknowledgements

This work was funded by The Swedish Research Council Formas (proj no. 2021-00713, 2021-00115) and Knut and Alice Wallenberg Foundation (2018.0259 Future Silviculture). The funding sources had no involvement in study design, collection, analysis and interpretation of data, nor in the writing of the report.

References

2016/17:104 A National Food Strategy for Sweden – more jobs and sustainable growth throughout the country. Short Version of Government Bill 2016/17:104. Government Offices of Sweden, Ministry of Enterprise and Innovation. Elanders. 24 p. Stockholm.

Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience Remote Sens. 57, 1–20. https://doi.org/10.1080/15481603.2019.1650447.

Ågren, A.M., Larson, J., Paul, S.S., Laudon, H., Lidberg, W., 2021. Use of multiple LIDAR-derived digital terrain indices and machine learning for high-resolution national-scale soil moisture mapping of the Swedish forest landscape. Geoderma 404, 115280. https://doi.org/10.1016/j.geoderma.2021.115280.

Ågren, A.M., Hasselquist, E.M., Stendahl, J., Nilsson, M.B., Paul, S.S., 2022. Delineating the distribution of mineral and peat soils a the landscape scale in northern boreal regions. Soil 8, 733–749. https://doi.org/10.5194/soll-8-733-2022.
Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: a next-generation hyperparameter optimization framework. Kdd'19: Proceedings of the 25th Acm Sigkdd International Conferencce on Knowledge Discovery and Data Mining, pp. 2623–2631. https://doi.org/10.1145/3292500.3330701.
Auffret, A.G., Svenning, J.C., 2022. Climate warming has compounded plant responses to habitat conversion in northern Europe. Nat. Commun. 13 https://doi.org/10.1038/s41467-022-35516-7. ARTN 7818.
Auffret, A.G., Kimberley, A., Plue, J., Skånes, H., Jakobsson, S., Waldén, E., Wennbom, M., Wood, H., Bullock, J.M., Cousins, S.A.O., Gartz, M., Hooftman, D.A.P., Tränk, L., 2017. HistMapR: rapid digitization of historical land-use maps in R. Methods Ecol. Evol. 8, 1453–1457. https://doi.org/10.1111/2041-210x.12788.
Auffret, A.G., Kimberley, A., Plue, J., Waldén, E., 2018. Super-regional land-use change and effects on the grassland specialist flora. Nat. Commun. 9 https://doi.org/10.1038/s41467-018-05991-y. ARTN 3464.
Baltsavias, E.P., 1994. Test and calibration procedures for image scanners. In: ISPRS Commission I Symposium, pp. 163–170. https://doi.org/10.3929/ethz-a-004334530. Como, Italy, September 12-16: IAPRS.
Baltsavias, E.P., 1996. Digital ortho-images - a powerful tool for the extraction of spatial- and geo-information. ISPRS J. Photogrammetry Remote Sens. 51, 63–77. https://doi.org/10.1016/0924-2716(95)00014-3.
Bellemare, J., Motzkin, G., Foster, D.R., 2002. Legacies of the agricultural past in the forested present: an assessment of historical land-use effects on rich mesic forests. J. Biogeogr. 29, 1401–1420. https://doi.org/10.1046/j.1365-2699.2002.00762.x.
Belward, A.S., Skoien, J.O., 2015. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. ISPRS J. Photogrammetry Remote Sens. 103, 115–128. https://doi.org/10.1016/j.isprsjprs.2014.03.009.
Chao, K., Song, W., Shao, S., Liu, D., Liu, X.C., Zhao, X.B., 2023. CUI-Net: a correcting uneven illumination net for low-light image enhancement. Sci. Rep. 13 https://doi.org/10.1038/s41598-023-39524-5. ARTN 12894.
Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, San Francisco, California, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychologial Measurment 20, 37–46. https://doi.org/10.1177/001316446002000104.
Delgado, R., Tibau, X.A., 2019. Why Cohen's Kappa should be avoided as performance measure in classification. PLoS One 14. https://doi.org/10.1371/journal.pone.0222916.
García-Alvarez, D., Van Delden, H., Olmedo, M.T.C., Paegelow, M., 2019. Uncertainty challenge in geospatial analysis: an approximation from the land use cover change modelling perspective. Geospatial Challenges in the 21st Century 289–314. https://doi.org/10.1007/978-3-030-04750-4_15.
GDAL/OGR, 2024. Geospatial Data Abstraction software. Library, Open Source Geospatial Foundation. https://doi.org/10.5281/zenodo.5884351.
Gobbi, S., Ciolli, M., La Porta, N., Rocchini, D., Tattoni, C., Zatelli, P., 2019. New tools for the classification and filtering of historical maps. ISPRS Int. J. Geo-Inf. 8 https://doi.org/10.3390/ijgi8100455. ARTN 455.
Griffin, A.L., 2020. Trustworthy maps. Journal of Spatial Information Science 5–19. https://doi.org/10.5311/Josis.2020.20.654.
Gustavsson, E., Lennartsson, T., Emanuelsson, M., 2007. Land use more than 200 years ago explains current grassland plant diversity in a Swedish agricultural landscape. Biol. Conserv. 138, 47–59. https://doi.org/10.1016/j.biocon.2007.04.004.
Högberg, P., Wellbrock, N., Högberg, M.N., Mikaelsson, H., Stendahl, J., 2021. Large differences in plant nitrogen supply in German and Swedish forests - implications for management. For. Ecol. Manag. 482 https://doi.org/10.1016/j.foreco.2020.118899. ARTN 118899.
Horn, B., Ferreira, C., Kalantari, Z., 2022. Links between food trade, climate change and food security in developed countries: a case study of Sweden. Ambio 51, 943–954. https://doi.org/10.1007/s13280-021-01623-w.
Houghton, R.A., House, J.I., Pongratz, J., van der Werf, G.R., DeFries, R.S., Hansen, M.C., Le Quéré, C., Ramankutty, N., 2012. Carbon emissions from land use and land-cover change. Biogeosciences 9, 5125–5142. https://doi.org/10.5194/bg-9-5125-2012.
Kasimir-Klemedtsson, A., Klemedtsson, L., Berglund, K., Martikainen, P., Silvola, J., Oenema, O., 1997. Greenhouse gas emissions from farmed organic soils: a review. Soil Use Manag. 13, 245–250. https://doi.org/10.1111/j.1475-2743.1997.tb00595.x.
Keenan, R.J., Reams, G.A., Achard, F., de Freitas, J.V., Grainger, A., Lindquist, E., 2015. Dynamics of global forest area: results from the FAO global forest Resources assessment 2015. For. Ecol. Manag. 352, 9–20. https://doi.org/10.1016/j.foreco.2015.06.014.
Klemedtsson, L., von Arnold, K., Weslien, P., Gundersen, P., 2005. Soil CN ratio as a scalar parameter to predict nitrous oxide emissions. Global Change Biol. 11, 1142–1147. https://doi.org/10.1111/j.1365-2486.2005.00973.x.
Kovesi, P., 2010. Fast almost-Gaussian filtering. In: 2010 International Conference on Digital Image Computing: Techniques and Applications, pp. 121–125. https://doi.org/10.1109/DICTA.2010.30.
Laudon, H., Hasselquist, E.M., Peichl, M., Lindgren, K., Sponseller, R., Lidman, F., Kuglerová, L., Hasselquist, N.J., Bishop, K., Nilsson, M.B., Ågren, A.M., 2021. Northern landscapes in transition: evidence, approach and ways forward using the Krycklan Catchment Study. Hydrol. Process. 35, e14170 https://doi.org/10.1002/hyp.14170.
Laudon, H., Lidberg, W., Sponseller, R.A., Hasselquist, E.M., Westphal, F., Ostlund, L., Sandstrom, C., Jarveoja, J., Peichl, M., Agren, A.M., 2022. Emerging technology can guide ecosystem restoration for future water security. Hydrol. Process. 36 https://doi.org/10.1002/hyp.14729. ARTN e14729.
Li, Z.Q., 2022. Extracting spatial effects from machine learning model using local interpretation method: an example of SHAP and XGBoost. Comput. Environ. Urban Syst. 96 https://doi.org/10.1016/j.compenvurbsys.2022.101845. ARTN 101845.
Li, C., Managi, S., 2024. Mental health and natural land cover: a global analysis based on random forest with geographical consideration. Sci. Rep. 14 https://doi.org/10.1038/s41598-024-53279-7. ARTN 2894.
Li, J.Y., Hua, Z.S., Liu, T., Wang, C.W., Li, J., Bai, G., Lucker, S., Jetten, M.S.M., Zheng, M., Guo, J.H., 2021. Historical land use has long-term effects on microbial community assembly processes in forest soils. Isme Communications 1. https://doi.org/10.1038/s43705-021-00051-x. ARTN 48.
Lidberg, W., Paul, S.S., Westphal, F., Richter, K.F., Lavesson, N., Melniks, R., Ivanovs, J., Ciesielski, M., Leinonen, A., Agren, A.M., 2023. Mapping drainage ditches in forested landscapes using deep learning and aerial laser scanning. J. Irrigat. Drain. Eng. 149 https://doi.org/10.1061/Jidedh.Ireng-9796. ARTN 04022051.
Lindsay, J.B., 2023. WhiteboxTools User Manual. Geomorphometry and Hydrogeomatics Research Group, University of Guelph, Guelph, Canada, version 2.3.0.
Loucks, C., Ricketts, T.H., Naidoo, R., Lamoreux, J., Hoekstra, J., 2008. Explaining the global pattern of protected area coverage: relative importance of vertebrate biodiversity, human activities and agricultural suitability. J. Biogeogr. 35, 1337–1348. https://doi.org/10.1111/j.1365-2699.2008.01899.x.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc, Long Beach, California, USA, pp. 4768–4777. https://doi.org/10.48550/arXiv.1705.0787.
Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta Protein Struct. 405, 442–451. https://doi.org/10.1016/0005-2795(75)90109-9.
Mäyrä, J., Kivinen, S., Keski-Saari, S., Poikolainen, L., Kumpula, T., 2023. Utilizing historical maps in identification of long-term land use and land cover changes. Ambio 52, 1777–1792. https://doi.org/10.1007/s13280-023-01838-z.
McKenzie, M.L., 1973. Photoimages for map bases. J. Res. U. S. Geol. Surv. 1, 327–339.
Naturvårdsverket, 2023. Nationella Marktäckedata 2018 Basskikt, Produktbeskrivning, p. 59. Stockholm, Version 2.3.
Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., Börger, L., Bennett, D.J., Choimes, A., Collen, B., Day, J., De Palma, A., Díaz, S., Echeverria-Londoño, S., Edgar, M.J., Feldman, A., Garon, M., Harrison, M.L.K., Alhusseini, T., Ingram, D.J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Correia, D.L.P., Martin, C.D., Meiri, S., Novosolov, M., Pan, Y., Phillips, H.R.P., Purves, D.W., Robinson, A., Simpson, J., Tuck, S.L., Weiher, E., White, H.J., Ewers, R.M., Mace, G.M., Scharlemann, J.P.W., Purvis, A., 2015. Global effects of land use on local terrestrial biodiversity. Nature 520, 45–50. https://doi.org/10.1038/nature14324.

Newbold, T., Hudson, L.N., Arnell, A.P., Contu, S., De Palma, A., Ferrier, S., Hill, S.L.L., Hoskins, A.J., Lysenko, I., Phillips, H.R.P., Burton, V.J., Chng, C.W.T., Emerson, S., Gao, D., Pask-Hale, G., Hutton, J., Jung, M., Sanchez-Ortiz, K., Simmons, B.I., Whitmee, S., Zhang, H.B., Scharlemann, J.P.W., Purvis, A., 2016. Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. Science 353, 288–291. https://doi.org/10.1126/science.aaf2201.

Nilsson, M., Ahlkrona, E., Jönsson, C., Allard, A., 2020. Regionala Jämförelser Mellan Nationella Marktäckedata Och Fältdata Från Riksskogstaxeringen Och NILS. Sveriges lantbruksuniversitet & Naturvårdsverket, Umeå & Stockholm, p. 34.

NumPy, D., Steppi, A., van Beek, B., Harris, C., Laíns, F., Vanderplas, J., Yan, L., van Kerkwijk, M., Picus, M., Weber Mendonça, M., Shanker, N., Goldbaum, N., Gommers, R., Goswami, R., Adel, S., Berg, S., van der Walt, S., Reddy, T., Weckesser, W., 2023. NumPy user guide, version 1.26. https://numpy.org/doc/stable/index.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9, 62–66. https://doi.org/10.1109/TSMC.1979.4310076.

Piskinaite, E., Veteikis, D., 2023. The results of digitizing historical maps: comparison of Lithuanian land-use structure in the 19th and 21st centuries. Land 12. https://doi.org/10.3390/land12050946. ARTN 946.

Powers, D.M.W., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. J. Mach. Learn. Technol. 2, 37–63. https://doi.org/10.48550/arXiv.2010.16061.

Prestby, T.J., 2023. Trust in maps: what we know and what we need to know. Cartography and Geographic Information Science 1–18. https://doi.org/10.1080/15230406.2023.2281306.

Previdi, M., Smith, K.L., Polvani, L.M., 2021. Arctic amplification of climate change: a review of underlying mechanisms. Environ. Res. Lett. 16, 093003 https://doi.org/10.1088/1748-9326/ac1c29.

Rasterio, 2024. https://rasterio.readthedocs.io/en/stable/index.html.

Riksantikvarieämbetet, 1996. ODLINGSLANDSKAPET - en lång markanvändnings historia *Informationsavdelningen*. Kulturmiljöavdelningen 58. Stockholm.

Schulp, C.J.E., Verburg, P.H., 2009. Effect of land use history and site factors on spatial variation of soil organic carbon across a physiographic region. Agric. Ecosyst. Environ. 133, 86–97. https://doi.org/10.1016/j.agee.2009.05.005.

Skalos, J., Weber, M., Lipsky, Z., Trpáková, I., Santrucková, M., Uhlírová, L., Kukla, P., 2011. Using old military survey maps and orthophotograph maps to analyse long-term land cover changes - case study (Czech Republic). Appl. Geogr. 31, 426–438. https://doi.org/10.1016/j.apgeog.2010.10.004.

Song, Y.Z., Kalacska, M., Gasparovic, M., Yao, J., Najibi, N., 2023. Advances in geocomputation and geospatial artificial intelligence (GeoAI) for mapping. Int. J. Appl. Earth Obs. Geoinf. 120 https://doi.org/10.1016/j.jag.2023.103300. ARTN 103300.

Swedish Board of Agriculture, 2011. Agriculture in Figures Years 1866-2007. SCB-tryck, p. 209. Örebro.

Swedish Official Statistics, 2023. Jordbruksmarkens användning 2023. Slutlig Statistik. https://jordbruksverket.se/om-jordbruksverket/jordbruksverkets-officiella-statistik/jordbruksverkets-statistikrapporter/statistik/2024-02-07-jordbruksmarkens-anvandning-2023.-slutlig-statistik.

Tian, H.Q., Banger, K., Bo, T., Dadhwal, V.K., 2014. History of land use in India during 1880-2010: large-scale land transformations reconstructed from satellite data and historical archives. Global Planet. Change 121, 78–88. https://doi.org/10.1016/j.gloplacha.2014.07.005.

Watanabe, S., 2023. Tree-structured parzen estimator: understanding its algorithm components and their roles for better empirical performance. https://doi.org/10.48550/arXiv.2304.11127.

Winkler, K., Fuchs, R., Rounsevell, M., Herold, M., 2021. Global land use changes are four times greater than previously estimated. Nat. Commun. 12 https://doi.org/10.1038/s41467-021-22702-2. ARTN 2501.

Winter, E., 2002. Chapter 53 the shapley value. In: *Handbook Of Game Theory With Economic Applications*, 2025-2054. Elsevier. https://doi.org/10.1016/S1574-0005(02)03016-3.

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2026:21

This thesis evaluated machine learning and deep learning for mapping Sweden's Quaternary surface deposits and predicting depth-to-bedrock. Peat and bedrock were reliably classified, predictions for till were often confidently wrong, and sorted sediments remained challenging. Depth-to-bedrock predictions were accurate up to 10 metres but degraded thereafter. Machine learning is best positioned as a tool that supports expert mapping by providing higher-resolution predictions and explicit uncertainty estimates, rather than as a replacement.

Yiqi Lin received her PhD education at the Department of Forest Ecology and Management, SLU, Umeå. She holds M.Sc. degrees from Stockholm University and the KTH Royal Institute of Technology, and a B.A. from the University of British Columbia.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8124-238-6
ISBN (electronic version) 978-91-8124-268-3