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.

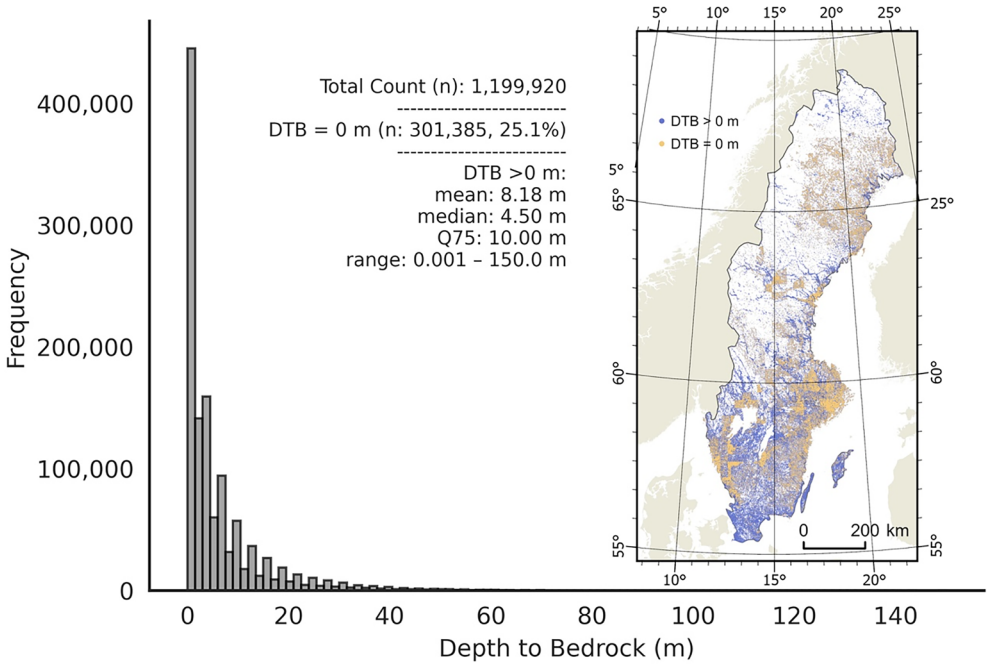

FIGURE 1 | Histogram and spatial distribution of depth-to-bedrock (DTB) observations (n = 1,199,920) over Sweden. The strong right skewness, zero inflation (25.1%), and spatial clustering might pose potential challenges for modeling.

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

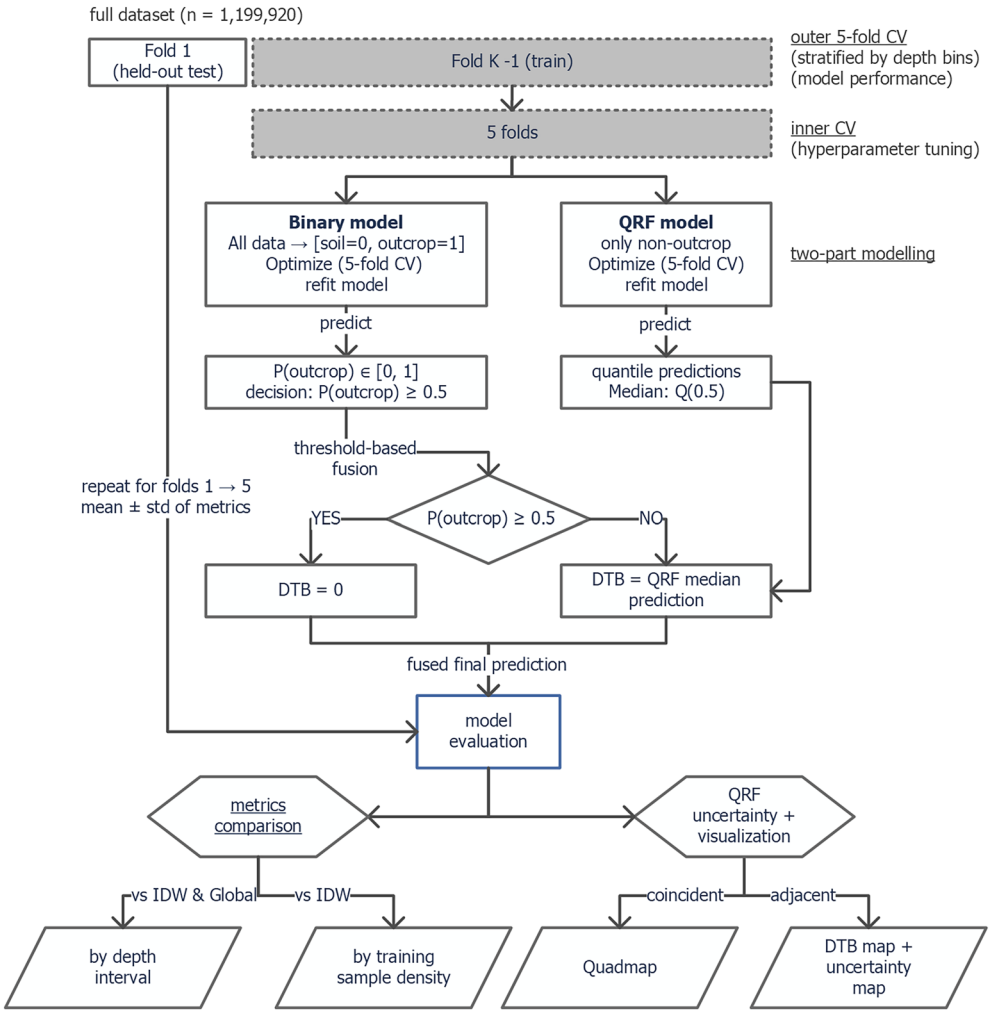

FIGURE 2 | Two-part modeling framework for Depth to Bedrock (DTB) prediction: Random Forest (RF) binary classification distinguishes bedrock outcrops from non-bedrock areas, and Quantile regression forest (QRF) predicts DTB. Models are fused at prediction time with a threshold (P (outcrop) $\geq 0.5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

$$\sigma = \frac{UL - LL}{2 \times 1.645} \quad (1)$$

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.

$$\text{normalized uncertainty} = \frac{\text{UL} - \text{LL}}{\text{median prediction}} \quad (2)$$

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

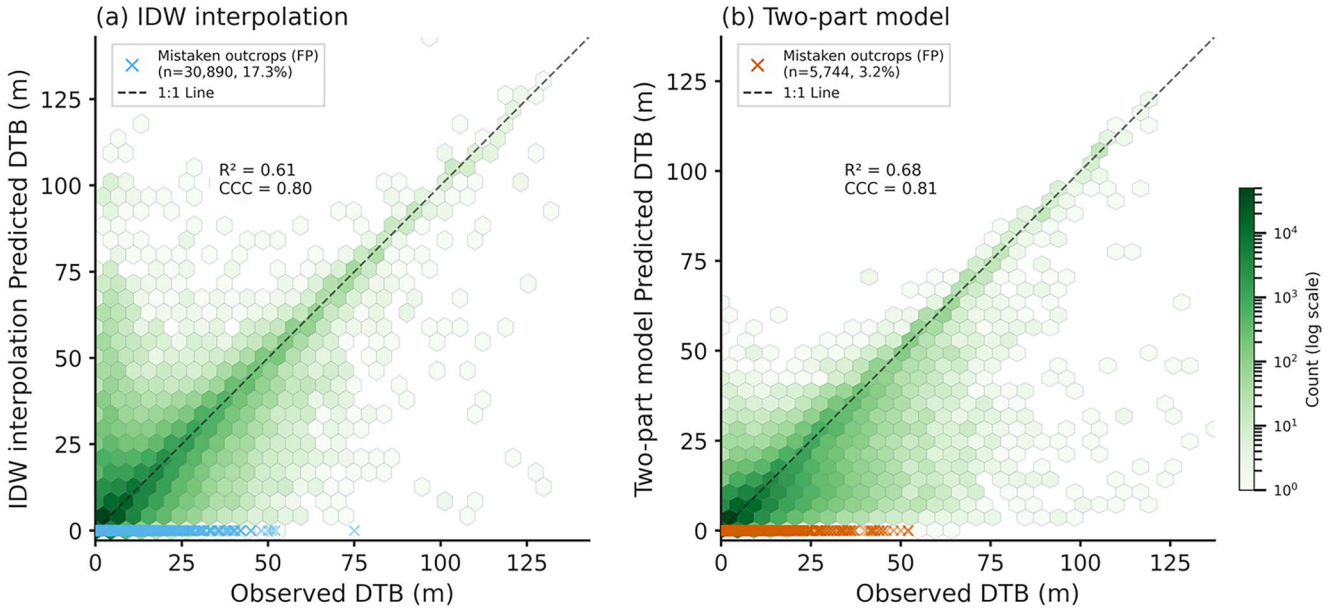

FIGURE 3 | Hexbin plots comparing observed versus predicted depth to bedrock (DTB) values for (a) Inverse Distance Weighting (IDW) interpolation and (b) the proposed two-part model. Hexagon color intensity represents point density. The diagonal dashed line represents perfect prediction (1:1 line).

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.

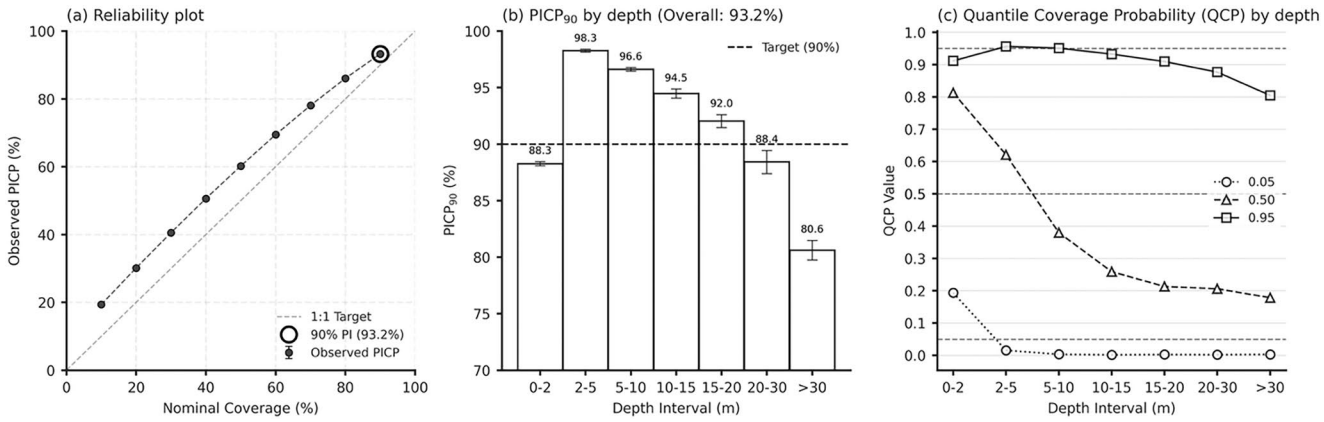

FIGURE 4 | Assessment of quantile reliability. (a) Reliability plot comparing nominal coverage probabilities against observed Prediction Interval Coverage Percentages (PICP). Points above the 1:1 diagonal indicate the model overcovers. (b) 90% Prediction Interval (PI_{90}) calibration by depth. The model 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

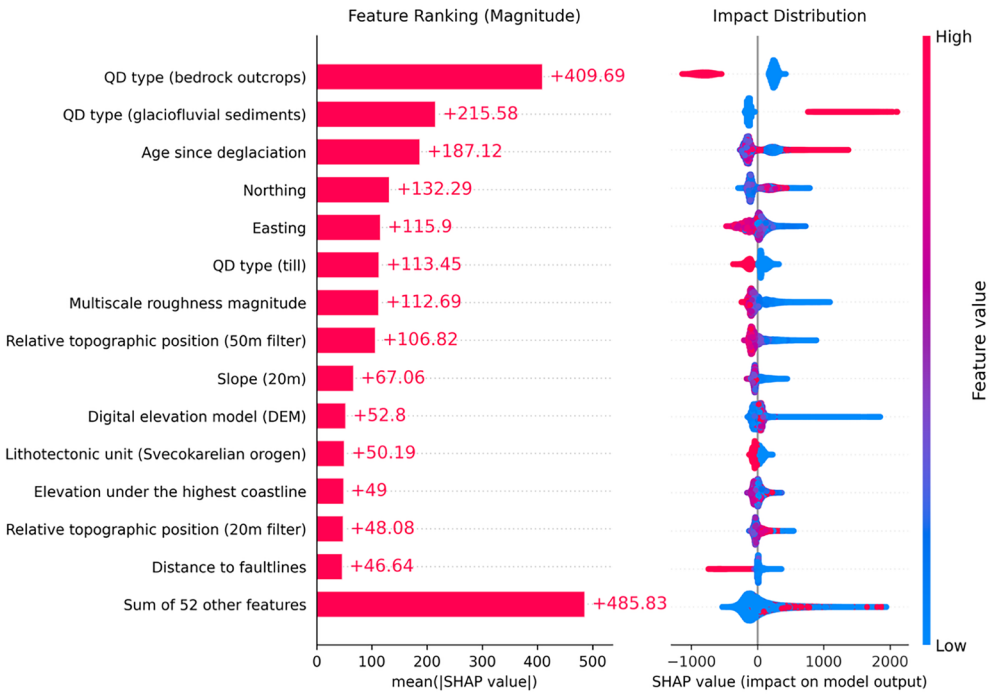

FIGURE 5 | SHAP summary plot showing the top 14 features impacting QRF prediction. The bar plot (left) represents the global magnitude of their influence on predicted DTB. The beeswarm plot (right) illustrates the distribution and the contribution of the feature to moving the prediction away from the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

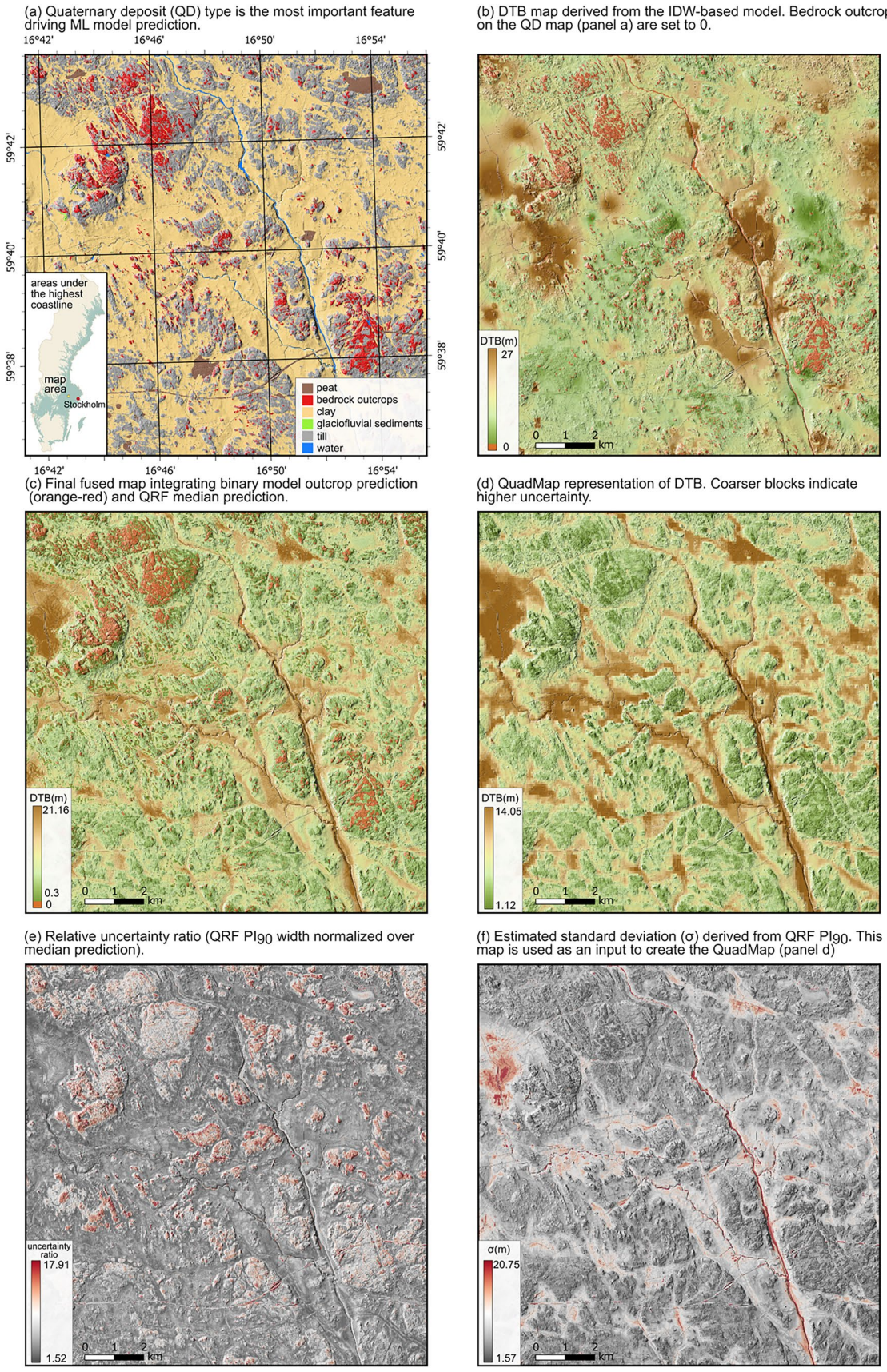

FIGURE 6 | Spatial prediction of depth to bedrock (DTB) and uncertainty visualization. (a) Quaternary deposit map of the mapped area, showing extensive clay plains. (b) DTB map derived from the IDW-based model. (c) Final fused map from the two-part model. (d) QuadMap representation of DTB with coarser 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.