This is a “preproof" accepted article for Journal of Glaciology. This version may be subject to change during the production process. 10.1017/jog.2025.10071

A quantile regression forest estimate of Greenland's subglacial topography

Steven PALMER,[1] Charlie KIRKWOOD,[2] Chunbo LUO[3], Mathieu MORLIGHEM[4]

[1] *Geography, Faculty of Environment, Science and Economy, University of Exeter, UK* [2] *Institute for Data Science and Artificial Intelligence, University of Exeter, UK* [3] *Computer Science, Faculty of Environment, Science and Economy, University of Exeter, UK* [4]*Department of Earth Sciences, Dartmouth College, Hanover NH 03755, USA* <*s.j.palmer@exeter.ac.uk*>

ABSTRACT. Accurate knowledge of basal topography is required for numerical modelling efforts to predict how Earth's ice sheets will respond to continued warming. The widely used BedMachine v3 dataset has limitations with respect to its use in modelling studies, particularly in estimating uncertainties. Machine learning approaches offer promise in addressing this gap, with quantile regression forests (QRF) especially suited to geospatial data. Here, we apply a novel QRF approach to map the basal topography of Greenlands ice sheet using airborne radio echo sounding (RES) data. Compared to BedMachine, our model reduces the root-mean-squared-error of ice depth predictions by 18%, from 232 m to 190 m. It also significantly improves uncertainty calibration: 89.8% of new observations fall within our 90% prediction interval, versus 68% for BedMachine. The QRF model achieves a lower continuous ranked probability score (92 m vs. 130 m), indicating improved balance between accuracy and uncertainty. Our volume estimate for the Greenland ice sheet is 0.7% higher than BedMachines, though we emphasise differences in the predicted shape of subglacial features like outlet glacier troughs. This approach offers a computationally efficient, accessible method for deriving subglacial topography from RES data, while providing better-calibrated uncertainty estimates than existing models.

INTRODUCTION

Ice thickness and bed elevation are fundamental for modelling glacier flow dynamics and ice-sheetclimate interactions (Durand and others, 2011; Bamber and others, 2013). Measuring these parameters requires remote-sensing techniques that penetrate ice and can cover extensive areas. Radio-echo sounding (RES) has been used since the 1950s for subglacial topography detection in Antarctica and Greenland and remains the primary method for mapping ice-sheet basal topography from aircraft (Robin and others, 1969; Dowdeswell and Evans, 2004; Bingham and Siegert, 2007; Schroeder and others, 2020). RES survey accuracy is influenced by factors such as rough topography, subglacial and englacial water, and crevassed ice, which can scatter or attenuate radar signals (Jordan and others, 2017; Chu and others, 2016; Kendrick and others, 2018). Additional factors, like radar instrument specifications and survey platform stability, further affect accuracy (Lapazaran and others, 2016).

Due to the vast size of the ice sheets, RES data are generally collected along flight lines with gaps of several kilometers, which requires interpolation to generate continuous bed topography maps essential for modelling ice dynamics and subglacial hydrology (Studinger and others, 2010). Bed topography significantly influences ice and subglacial water flow, the estimation of total ice volume, and the rate at which ice is discharged to the surrounding ocean (Bamber and others, 2013). As stated in the most recent IPCC assessment report, Greenlands ice mass, which accounted for 24.5% of observed sea-level rise from 19012018, will continue to impact sea level due to ongoing ice mass loss (Fox-Kemper and others, 2021). Accurate bed topography data are needed to model the ice sheet's response to changes in meltwater input, which is increasing with atmospheric warming. Subglacial topography is crucial for controlling ice flow and discharge, grounding line position, and ice front dynamics, particularly for marine-terminating sectors (Durand and others, 2011; Cooper and others, 2019; Van den Broeke and others, 2016). Attributes of the ice-bed interface such as roughness and the distribution of free water, determine basal traction and consequently control the rate of movement of ice mass from the accumulation zone to the ablation zone. (Gudmundsson, 1997; Bingham and Siegert, 2009; Hoffman and others, 2016). Diurnal, seasonal and longer-term changes in subglacial meltwater storage and distribution further impact basal friction and ice dynamics (Palmer and others, 2011, 2013, 2015; Chu and others, 2016). Accurate bed topography is therefore critical to understanding ice sheet responses to meltwater variability and other climate-driven factors.

Since its release in 2017, the BedMachine Greenland v3 data have become the most widely-used es-

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

timates of sublacial topography in Greenland (Morlighem and others, 2017). Quantifying uncertainty in these data is important to assign confidence to the outputs of modelling studies that use these data, as well as to inform future RES data acquisition strategies, although this task has proven challenging (Morlighem and others, 2013, 2017). The BedMachine data contains uncertainty estimates which are derived via different methods in different regions (Morlighem et al., 2017 supplementary materials). Near the ice margin, where mass conservation is used, these estimates are obtained through propagation of errors and assume that the ice flow is in steady-state'. Further inland, where kriging is used, uncertainties are estimated through proximity to radar observations, reaching several hundreds of metres in many places.

Previous studies have shown that machine learning approaches such as random forests (Breiman, 2001) have the potential to improve estimates of geophysical parameters. Quantile regression forests (QRFs) (Meinshausen and Ridgeway, 2006) are an extension of random forests that provide information about the conditional distribution of the target variable (e.g., ice depth) for a given input (e.g., position in space), rather than just a prediction of the conditional mean value of the target variable. These decision tree-based learning algorithms have become widely used in spatial machine learning applications (Prasad and others, 2006; Hengl and others, 2015; Kirkwood and others, 2016), and various investigations have been carried out into their properties as spatial learning algorithms (Hengl and others, 2018; Møller and others, 2020; Sekulić and others, 2020). Unlike more traditional methods, QRF does not assume linearity or specific error distributions, allowing it to provide more flexible and spatially nuanced uncertainty estimates. This data-driven technique can yield prediction intervals that better reflect the true variability in bed elevation, especially in regions with sparse observations or complex subglacial terrain, thereby improving confidence in glaciological modelling.

In this study, we investigate the potential for a QRF approach to (i) improve ice sheet bed elevation datasets and (ii) to provide a more rigorous quantification of uncertainty. We present a quantile regression forest to map the basal topography of Greenland's ice sheet, with estimated uncertainty. Our new method uses the approach of Møller and others (2020) to provide oblique geographic coordinates i.e. a range of different compass orientations as input features to the QRF, so that spatial thresholds are not restricted to align north-south and east-west, leading to more realistic estimates of subglacial topography.

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

DATA

We used data assets published by the BedMachine v3 study (Morlighem and others, 2017), as well as ice velocity rasters from the MEaSUREs project Joughin and others (2015, 2018), and ice surface elevation data from the Greenland Ice Mapping Project (Howat and others, 2014). To train our quantile regression forest, we use the same airborne radar observations of ice thickness as were used to develop BedMachine v3; these primarily come from NASA's Operation IceBridge, with data processed by the Center for Remote Sensing of Ice Sheets (CReSIS; Paden and others, 2010 updated 2019) as well as other smaller scale survey projects (see section 2 of Morlighem and others, 2017). To test the accuracy and calibration of our quantile regression forest approach, and to provide fair comparison with BedMachine v3, we use circum-Greenland ice thickness measurements collected during the PROMICE airborne surveys Sørensen and others (2018) as a test dataset for both BedMachine and our quantile regression forest (Fig. 1).

METHODOLOGY

BedMachine inferred ice thickness using a combination of a mass conservation approach along the periphery of the ice sheet and ordinary kriging within the interior (Fig. 1; Morlighem and others, 2017). Mass conservation relies on modelling the transport of ice, and the directions and magnitude of the ice speed are critical inputs, as they control how the ice is distributed spatially. In regions of slow speed, flow directions become poorly defined, and the accuracy of mass-conservation inferred ice thickness drops. We find that mass conservation works well for regions where the ice velocity is greater than 30-40 m yr^{-1}, which is 30% of the ice sheet area. In this new approach, we use a quantile regression forest (Meinshausen and Ridgeway, 2006) across the entire ice sheet. We model the subglacial bed elevation directly as our target variable, rather than ice thickness. This has the effect of making predictions from the terminal nodes of each tree in our quantile regression forest physically horizontal, rather than physically perpendicular to the surface of the ice, as would be the case when predicting ice thickness. We found this resulted in better generalisation, accounting for $\sim$10 of the overall $\sim$40 metre reduction in root-mean-squared-prediction-error of our regression forest approach compared to BedMachine.

The task of mapping Greenland's subglacial bed topography from point-sampled airborne radar observations of ice thickness (or equivalently of bed elevation, by subtracting ice thickness from surface elevation) is inherently a spatial interpolation problem. However, more information is available than just the airborne

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

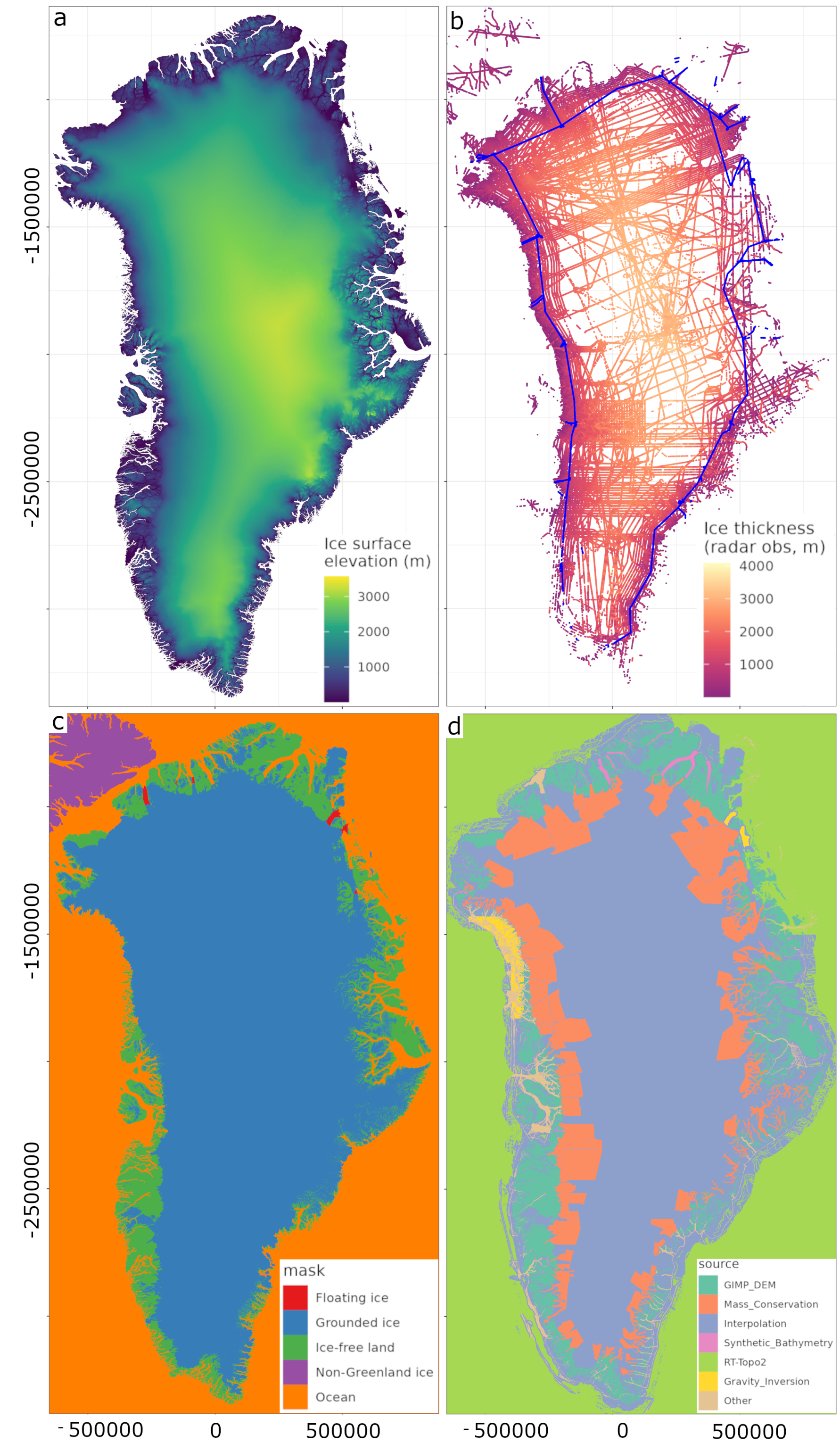

Fig. 1. Ice sheet surface elevation (a), RES observations of ice thickness(b), BedMachine mask (c), BedMachine sources (d). The PROMICE observations used to test both the QRF and BedMachine predictions are shown in blue on panel b.

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

radar observations and their spatial locations; we also have spatially-continuous knowledge of the ice sheet's surface in the form of gridded satellite data of surface elevation (Howat and others, 2014) and surface ice velocity (Joughin and others, 2015). Previous studies have shown that these covariate datasets contain information about subglacial topography (Smith and others, 2006; Gudmundsson, 2003), so we include them as input features to our QRF model (Fig. 2). We deduce that the degree to which surface features are informative of the subsurface will vary across the ice-sheet, and that there may be different interactions between different surface features at different locations. The desire to accommodate this flexibility, as well as to handle the scale of data involved (over 30 million airborne radar observations are available), motivates our investigation of machine learning for subglacial bed elevation prediction in this study.

To implement our quantile regression forest, we used the Ranger package Wright and Ziegler (2017) for the R language for statistical computing R Core Team (2023). The quantile regression forests algorithm Meinshausen and Ridgeway (2006) extends the popular 'random forest' machine learning algorithm to enable provision of conditional quantile predictions in addition to the random forest's typical prediction of the conditional mean. In a quantile regression forest, conditional quantiles are calculated from the empirical distribution of target variable observations that, for each tree, fall within the same leaf as the new prediction candidate. This enables the construction of prediction intervals to communicate prediction uncertainty; the calibration of which we assess as part of this study.

Data preparation

The collection of ice depth observations over Greenland by airborne radar survey has produced a spatially biased sample; all points on the map have not had equal probability of being observed, but rather observations are concentrated in flight lines, the densities of which vary significantly across the map, with some areas being 'hot spots' for observations while others are relatively unobserved. Any attempt to minimise prediction error equally across all observations is therefore not an attempt to minimise prediction error equally across the map, which should really be our aim.

To mitigate the effect of this sampling bias in our modelling, we first resampled the radar observations according to the following scheme: For n desired number of observations to use in modelling (we chose $n =$ 3.7 million in this study; equivalent to 1/10th of all the radar observations, and a good balance between computational cost and model quality), randomly generate easting and northing coordinates (NSIDC North pole stereographic, EPSG: 3413) of a location over Greenland, then take the nearest neighbouring radar

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

observation to be observation $i \in [1,\ n]$. Allow the same observation to be used whenever it is nearest to subsequent randomly generated coordinates (i.e., allow replacement). Following this scheme produces a resampled set of radar observations in which the probability of each observation being included is inversely proportional to the spatial sampling intensity of the original survey, as advocated for by De Bruin and others (2022). This effectively re-weights the observations, by resampling, to have equal representation across the map. For small n (up to about 10,000) the resultant resampled observation sets appear indistinguishable from a spatially random sample. For large n (over about 100,000) the influence of flight lines is unavoidable, but at least observations in sparsely sampled locations are up-weighted, meaning that metrics calculated against the resampled observation dataset are more representative of the map as a whole, rather than being biased to over-represent observation hotspots.

Since BedMachine utilised all radar observations that were available to it (i.e., no observations were held out as a test set for the final BedMachine product), we have a challenge in making a fair comparison between BedMachine and other approaches. Fortunately, an additional radar observations dataset from the PROMICE project (Sørensen and others, 2018, shown in blue in Fig. 1b) was not used to produce BedMachine, and so we utilise it here as our test set for comparing to BedMachine. This is also the reason we compare against BedMachine, since more recent versions of BedMachine have assimilated the PROMICE dataset. The PROMICE radar observation dataset is spatially biased in that it was acquired using flight missions oriented parallel to the ice sheet margin around the periphery of Greenland, with no observations in the interior. As such, the evaluations we make of our QRF, and of BedMachine, using this PROMICE dataset are biased towards the performance of the models around the ice periphery (although as we shall see, this tends to be a difficult area to predict). To provide some reassurance of the performance of our QRF in the ice interior, we also report metrics from a cross validation scheme, but it is not possible to compare this with BedMachine; to do so would require running the BedMachine procedure on the same folds.

Feature engineering

The target variable, y, for our QRF is subglacial bed elevation. We wish to make predictions of y as a function of inputs x. These inputs consist of the spatial coordinates of each observed location (i.e., easting and northing in EPSG:3413) along with their corresponding values of satellite-observed covariates, which in this study are surface elevation, ice flow speed and ice flow angle. Thus in raw form (without feature

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

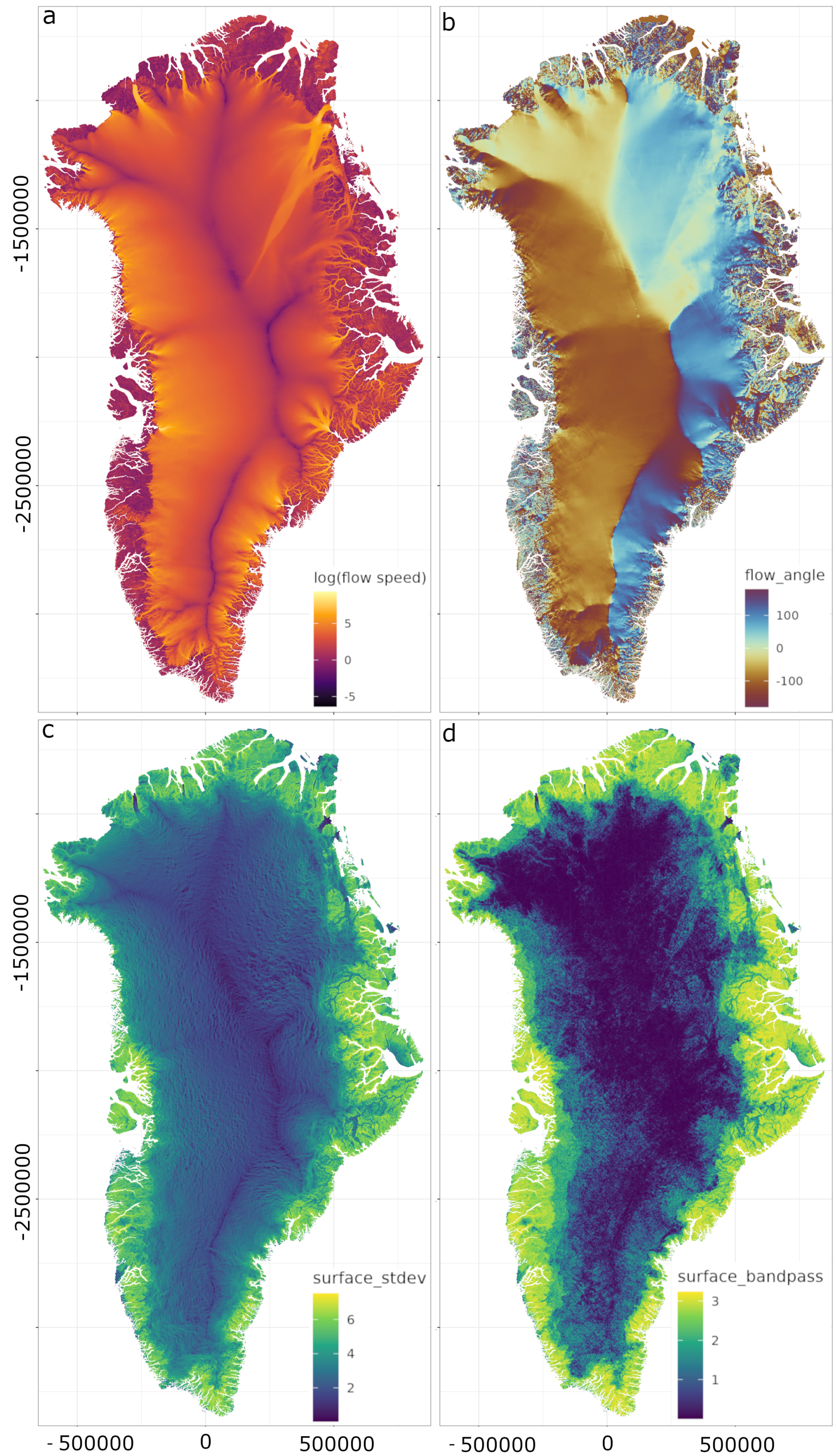

Fig. 2. Covariates used in the QRF: ice flow speed (a), ice flow angle (b) surface roughness (c) and roughness bandpass (d).

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

engineering) our input x_{raw} would be five dimensional: easting, northing, elevation, flow speed and flow angle would be our features.

However, in this study, we make three key feature engineering decisions to produce x_{eng} upon which our QRF is constructed:

1. We rotate our spatial axes to add new compass orientations in addition to easting and northing. In total, we provide 16 equal-angled rotations about the compass as our spatial input features. This allows our QRF to perform spatial interpolation that is more smooth and less 'blocky' than if the decision trees were forced to make partitions on easting and northing only.

2. We take the sine and cosine of the ice flow angle raster, so that the 'circular' nature of this variable is better represented (otherwise, 5 degrees and 355 degrees appear to be 350 degrees apart, when in reality they are only 10 degrees apart).

3. We filter the surface elevation raster to generate two derived features (although many more are possible); A) a simple surface roughness measure as the standard deviation of 5x5 window passed across the raster, and B) a simple bandpass of surface elevation for which we first subtract the mean of a 5x5 window (leaving only residual high frequency information) and then take the mean of this residual in a 5x5 window.

The benefits of feature engineering to tailor the inputs, x, towards the machine learning task at hand (to learn $f(x)$ such that $y = f(x)$) are well documented, (e.g. Kuhn and Johnson (2019)). Feature engineering helps guide a machine learning model by shaping its assumptions about how input variables relate to the target outcome. In Bayesian terms, it can be seen as helping to represent our prior belief about the possible relationships between the input variables x_{raw} and the predicted outcome y. The features we have engineered to include in this study are by no means exhaustive and there are many more that could be tried. We suggest that the engineering, or learning, of optimal features for this task would be a worthwhile focus for a future research project.

Hyperparameter tuning

With our observations resampled to compensate for the spatial sampling bias of flight lines (Bartlett and others, 2020), and with our features engineered to include additional compass orientations and derivatives of our satellite observed covariates, we proceeded to find suitable hyperparameters for the construction of

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

our quantile regression forest. Hyperparameters are *a priori* configuration values that influence how the QRF model learns from the training data. Our aim is to balance the trade-off between overfitting the data (i.e., spuriously fitting to noise) and underfitting the data (i.e, failing to fit to all of the signal). To assess our performance in this task, we use a k-fold cross-validation procedure (where k= 6), such that we split the training data (our radar observations excluding the PROMICE dataset) into six folds, and then train 6 QRFs, with the same hyperparameter values, each on a different set of 5 of the folds, so that their performance can be tested on the fold they did not see during training (see Appendix: Cross validation by field season). In this way, as we repeat the procedure using different hyperparameter values, we can assess the ability of the QRF algorithm with different hyperparameter values to make good predictions of subglacial bed elevation at locations it has not observed during training.

Typically, a k-fold cross validation procedure would split the data into k folds at random (Kohavi and others, 1995). However, to do so in our case would result in having observations from different folds side-by-side together in the same flight lines. Tuning hyperparameters on such folds would result in optimising for interpolation along flight lines, rather than between flight lines. Since large swathes of the map lie between flight lines (Fig. 1b), it makes sense to prioritise between-flight-line predictive performance, as this is more representative of the challenge of mapping Greenland's subglacial bed topography as a whole. In order to achieve this, we create our six folds for cross-validation by splitting the observations by the years in which they were collected; the aircraft in different years flew different routes, and so this results in six folds of distinct flight lines. The folds are assembled by years as follows: Fold 1 - years 1993 to 2001, Fold 2 - years 2002 to 2010, Fold 3 - years 2011 to 2012, Fold 4 - years 2013 to 2014, Fold 5 - years 2015 to 2017, Fold 6 - years 2018 to 2019. Splitting the data in this way produces 6 folds with close-to-equal number of observations.

By manually tuning hyperparameters to minimize the root-mean-squared error of predictions in cross-validation, we arrived at the following hyperparameters for our QRF: max.depth = 0 (unlimited), min.node.size = 9, splitrule = "variance", replace = TRUE, sample.fraction = 0.95, mtry = 5 (default for our number of inputs). We do not limit the maximum depth of the trees, but instead specify a minimum node size of 9 observations. Each tree is grown on a 95% random subsample of the data, sampling with replacement. Trees are grown using the typical 'variance' split rule, such that the intra-node variance of the resultant partitions is minimised. We do not change the 'mtry' hyperparameter, which controls how many variables are considered for each decision split in a tree. By default, it is set to the square root of the total number

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

of input features, rounded to the nearest integer. Since our model has 22 input features, this results in an 'mtry' value of 5. For the cross-validated hyperparameter tuning, we grew 20 trees per forest, this achieved a cross-validated R^2 of 0.97 and root-mean-squared-error of 172 metres — these cross-validation metrics are of interest mainly in the context of comparison with results for the PROMICE test set, which we report in the results section.

For the final forest trained on all folds of radar observations (but excluding the PROMICE dataset which we reserved for testing only), we grew 100 trees using the hyperparameters described above.

RESULTS AND DISCUSSION

Test metrics

Our quantile regression forest (QRF) model outperforms BedMachine in both deterministic and probabilistic predictions when tested on the circum-Greenland PROMICE dataset, unseen by either method during training (Fig. 3). In terms of ice thickness (surface elevation minus predicted bed elevation), the QRF achieves a root-mean-squared error (RMSE) of 190 meters, representing a 18% reduction compared to BedMachine's 232 meters. This improvement, though modest, is significant in terms of increasing prediction accuracy, particularly when applied across large-scale ice-sheet models where even small gains can lead to more accurate estimates of ice-sheet volume and sea-level rise projections.

Additionally, the QRF model exhibits a higher R^2 value of 0.83 compared to BedMachine's 0.79 when compared with the PROMICE data, representing an increased correlation between observed and predicted ice thickness values. This modest improvement in deterministic performance translates into better fidelity in capturing subglacial features, which has implications for understanding Greenland's ice dynamics using numerical modelling (Fig. 4).

Predictive performance insights

Examining scatter plots of predicted versus observed ice thickness (Fig. 3), the QRF model reveals a more consistent performance across all observed depths compared to BedMachine, which demonstrates a tendency to under-predict ice thickness in shallow-ice regions. Notably, BedMachine often predicts zero or near-zero ice thickness in regions where the radar observations suggest depths of up to 500 meters (Fig. 4). This systematic underprediction, particularly along the shallow periphery, could have profound effects on estimates of ice-sheet dynamics and mass loss, which rely on accurate thickness measurements in these

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

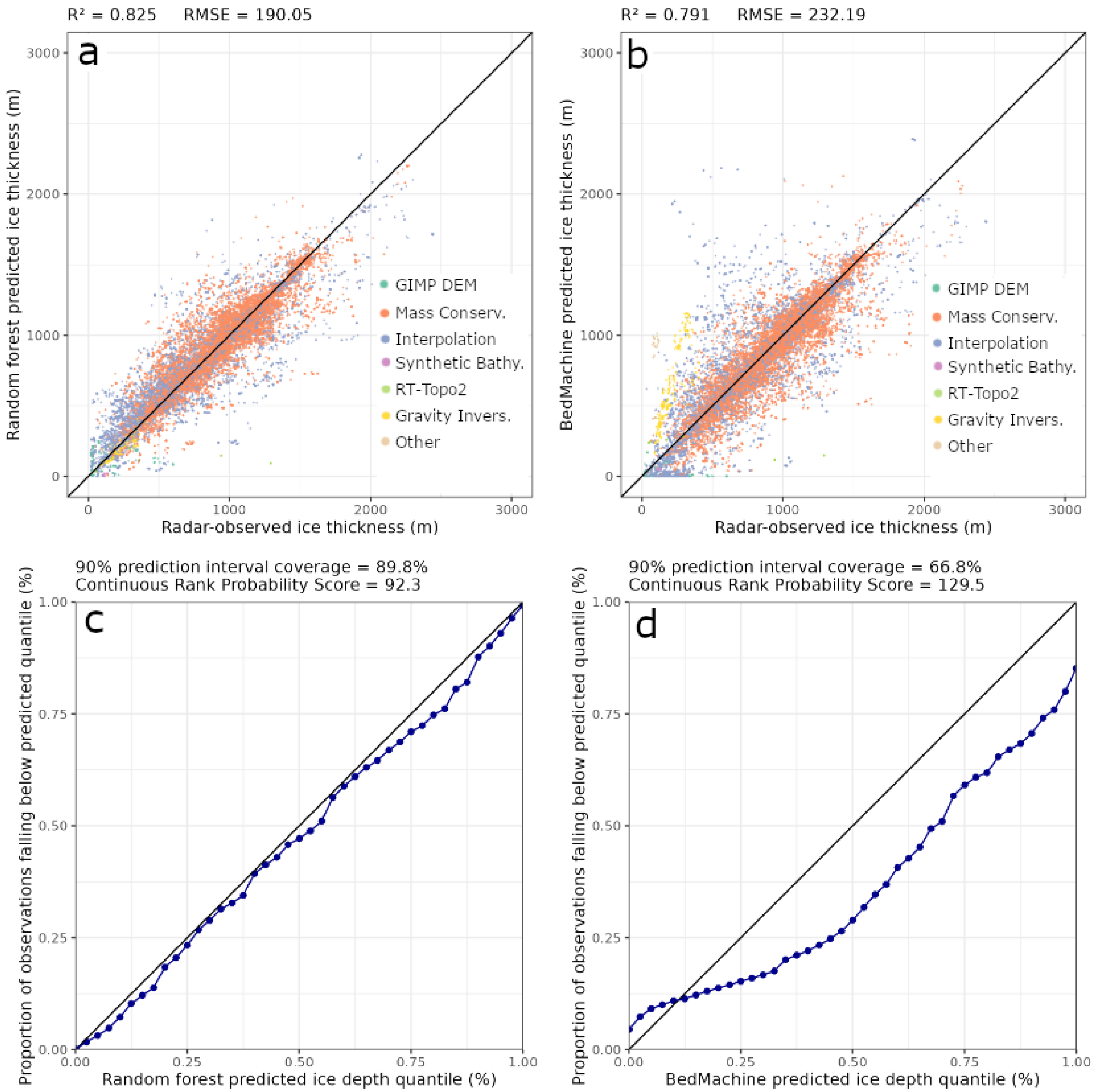

Fig. 3. QRF predictions versus observations (a), BedMachine predictions versus observations (b), QRF calibration (c), BedMachine calibration(d).

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

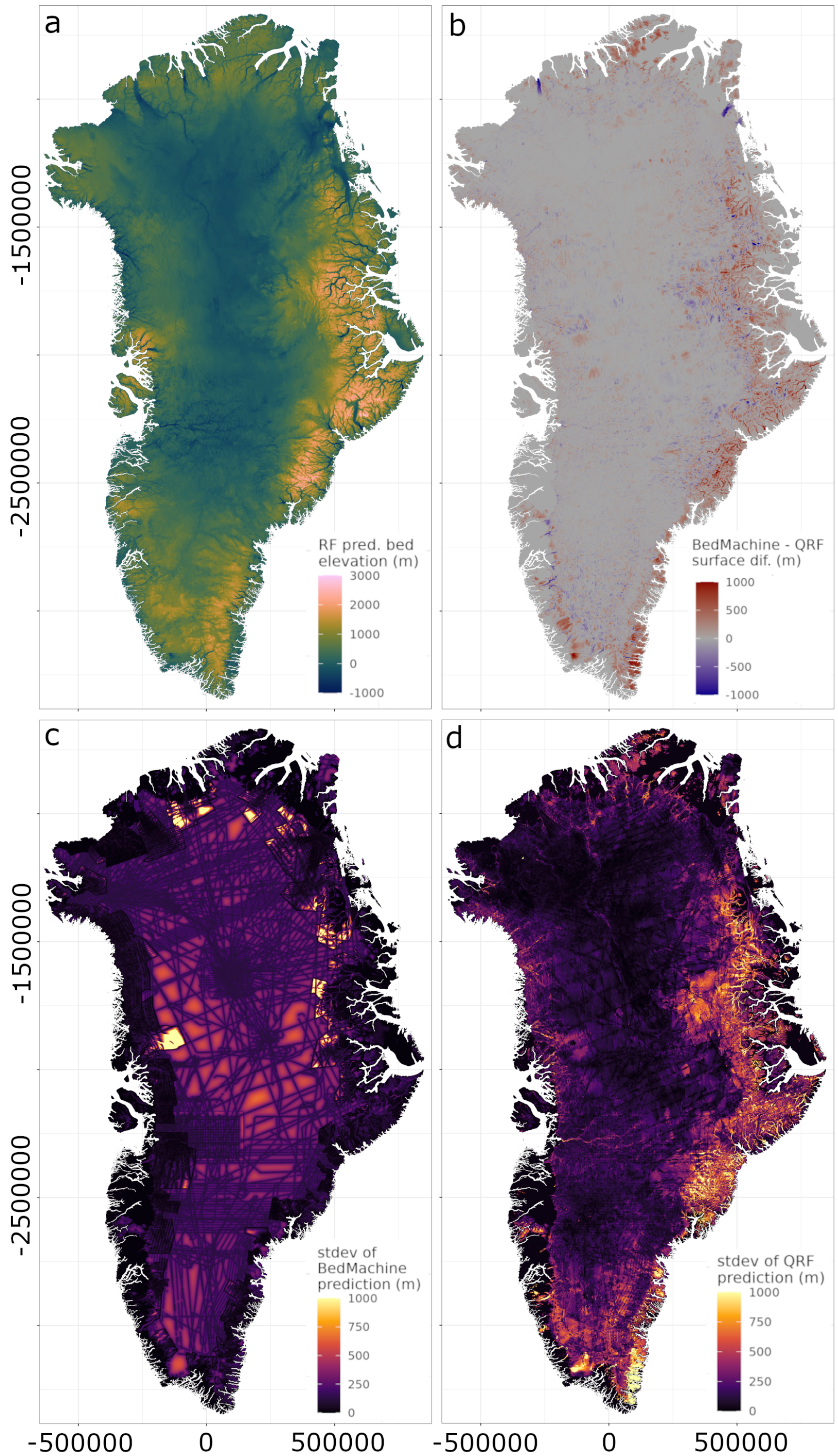

Fig. 4. QRF predicted bed elevation (a), BedMachine minus QRF (difference) (b), Bed Machine uncertainty (c), QRF uncertainty (d).

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

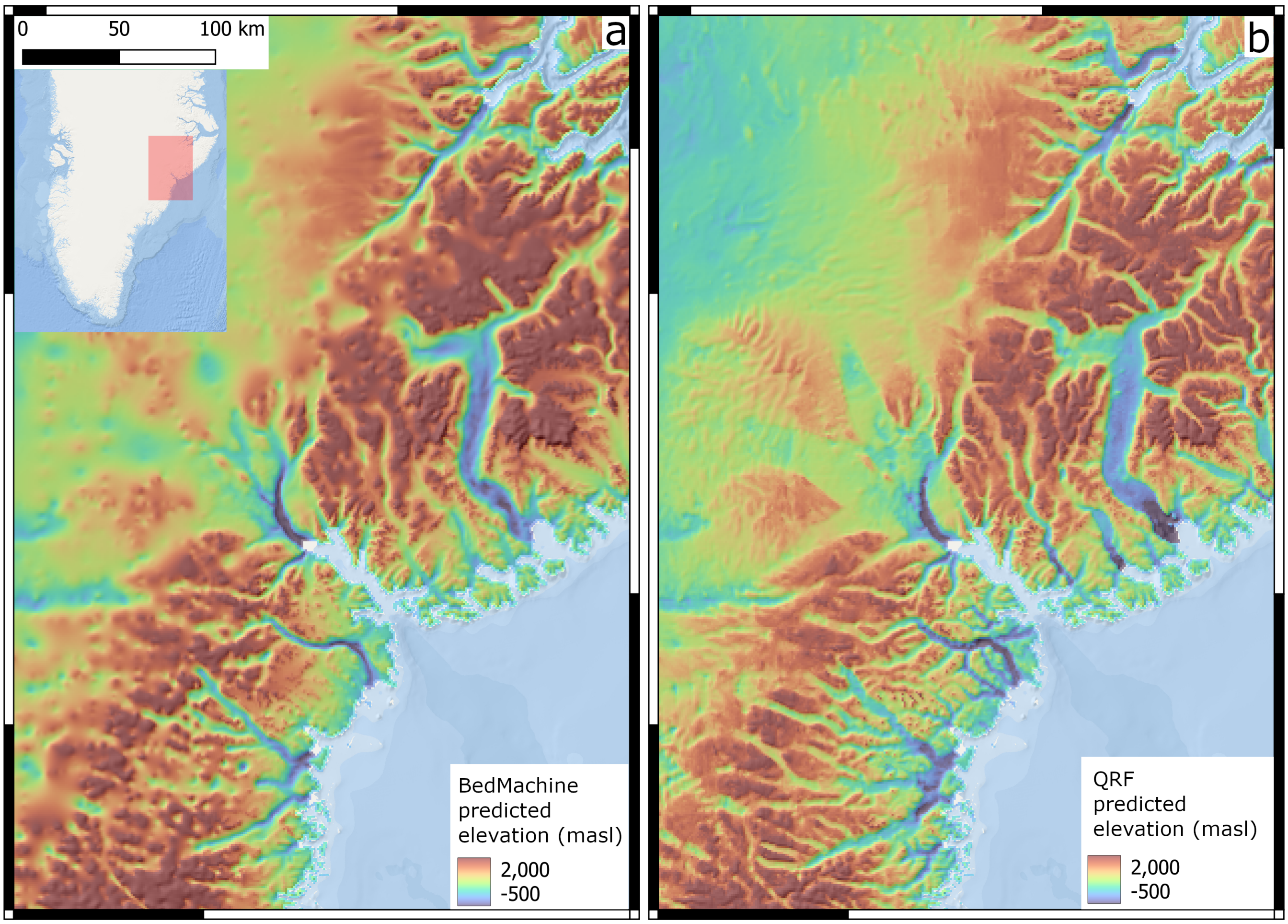

Fig. 5. Comparison of bed elevations predicated by BedMachine v3 (a) and our new QRF model (b) for the region near Kangerlussuaq Glacier in East Greenland.

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

crucial areas. In contrast, the QRF model provides a more balanced prediction distribution, with residuals scattered symmetrically around the diagonal for all observed ice depths.

These improvements in prediction accuracy are particularly relevant in regions of Greenland where the bedrock topography is complex, and ice dynamics are heavily influenced by subglacial features. Since BedMachine uses methods that have a tendency to smooth out important subglacial features, the QRF model's approach of using high-resolution satellite-derived surface and velocity data allows it to capture individual topographic features in finer detail. as can be seen in Fig. 5. This advantage is crucial for predicting the movement of subglacial water and ice flow, as accurate topographic representation can impact our understanding of ice discharge and basal hydrology, both of which are important for predicting sea-level rise.

Uncertainty quantification and practical relevance

A key strength of the QRF model lies in its ability to produce well-calibrated prediction intervals. BedMachine, on the other hand, exhibits underdispersed prediction intervals, with its 90% prediction interval capturing only 66.8% of test observations, indicating an apparent miscalibration in its predictions when compared with the PROMICE data (Fig 3d). This is problematic for practical applications, as models used to predict the response of ice sheets to climate change require not only accurate boundary conditions but also a robust representation of uncertainty. In contrast, the QRF model captures 89.8% of test observations within its 90% prediction interval, making it a more reliable tool for assessing uncertainties in ice-sheet behavior.

The continuous ranked probability score (CRPS) measures the difference between the predicted and observed cumulative distributions (Hersbach, 2000). The CRPS is widely used in probabilistic model evaluation owing to it being a proper scoring rule; one that rewards not only accuracy but also well-calibrated and sharp quantification of uncertainty (Gneiting and Raftery, 2007), with the CRPS' ultimate minimal value of zero only being achieved for a perfect deterministic prediction. The lower CRPS of our QRF model (92 metres compared to 130 metres for BedMachine) further highlights its advantage in balancing accuracy and uncertainty. This improved performance is especially valuable for quantifying the uncertainty in projections of sea level rise arising from the subglacial topography.

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

Mapping subglacial topography

The map generated by the QRF model (Fig. 4a) displays a more detailed and nuanced subglacial landscape compared to those produced by BedMachine (Fig. 5). The QRF model reveals a rich diversity of subglacial features, such as river channels, highlands, and lowland plains, particularly in Greenland's interior, where BedMachine uses kriging to interpolate ice thickness observations - sometimes over distances in excess of many kilometres. The use of satellite-derived ice surface and velocity data potentially enables the QRF to predict bed elevation with greater precision in regions where BedMachine relies solely on spatial interpolation of sparsely-sampled ice thickness measurements.

The differences between the two models' predictions are most pronounced at the ice-sheet periphery, where BedMachine relies on mass conservation for grounded ice and gravity inversion for fjord bathymetry. These methods, while useful, impose strict assumptions that may smooth out ravines and other important subglacial structures. In contrast, the QRF model offers more flexible predictions, which reveal deeper ice and lower bed elevations in subglacial ravines around Greenlands margins (Fig. 5). These features are important because they play a critical role in controlling ice flow, basal water distribution, and ice discharge into the ocean, factors that ultimately influence the rate of ice loss from Greenland.

In some regions, such as the southwest of the GrIS, BedMachine predicts deeper ice in subglacial ravines than the QRF model, but this appears to be the exception rather than the rule. Notably, large differences in estimated bed elevation exist between the QRF and BedMachine at floating ice shelves in northern Greenland due to the fact that the QRF is not able to represent the correct ice/sea/bed configuration. Such discrepancies highlight a limitation in the approach used here. Future work is needed to refine our approach in regions where the ice is floating or otherwise decoupled from the bedrock.

Ice volume estimates

One of the most significant applications of improved bed elevation estimates is in calculating total ice volume and, by extension, potential future sea-level rise contributions. While our QRF model does not explicitly model spatial dependence between locations, its structure inherently captures some spatial relationships through the arrangement of regression tree nodes. This allows for realistic estimates of total ice volume and uncertainty, a feature that is vital for large-scale models predicting Greenlands contribution to future sea-level rise. Our estimated ice volume is (3.011 ± 0.004) x 10^6 km^3, while the bedmachine estimate is (2.99 ± 0.02) x 10^6 km^3. This makes Our QRF estimate of the total Greenland ice sheet volume to be

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

0.7% greater than that estimated by BedMachine, though we place more importance in the differences the shape of the bed features, especially near the ice margin.

As we have seen, our QRF appears to be well calibrated when assessed against held-out test observations, i.e. a 90% prediction interval provided by the QRF will contain the true bed elevation in (close to) 90% of cases, and likewise for other prediction intervals. However, being able to reasonably quantify ice-thickness prediction uncertainty at individual locations does not necessarily mean that one can also reasonably quantify the uncertainty in an estimate of the total ice volume. If prediction errors are modelled entirely independently at each location, then when summed together to obtain a total, the errors will cancel out, resulting in a spuriously overconfident estimate of the total (Wadoux and Heuvelink, 2023). Conversely, if prediction errors have an unrealistically strong spatial dependence, they may not cancel out enough when summed together, potentially resulting in an unrealistically under-confident estimate of the total. While our QRF does not explicitly model spatial dependence, spatial dependence is implicitly captured in the structure of the quantile regression trees: Each tree has a number of terminal nodes, which each contain a set of training observations. The within-node spatial dependence is not modelled; we can 'simulate' observations from a node simply by sampling at random from the observed values it contains, this equates to an uncorrelated error distribution, like nugget variance in traditional geostatistics. The spatial dependence between nodes is however captured by the arrangement of their boundaries; regression trees will produce smaller terminal nodes where the 'length scale' of observations is shorter, i.e., where there is more high-frequency information to capture. Conversely, where the 'length scale' of observations is large, terminal nodes can also be large.

While a single decision tree can only crudely approximate the spatial dependence between observations, when combined together in the QRF, the quality of this approximation improves. We can think of the distribution of observations within each node of each tree as a nugget-effect-like independent noise (the data distribution), that each tree estimates. This is our aleatoric uncertainty (data uncertainty representing the inherent randomness in the observations, e.g. including measurement error). Meanwhile, the ensemble of mean values predicted by each tree throughout the feature space approximates a distribution over functions representing our epistemic uncertainty (reducible uncertainty representing our lack of knowledge). By treating our QRF as an approximate Bayesian model (with the ensemble of regression trees representing the 'poor man's posterior' as described on p.272 of Hastie and others, 2009), we can 'simulate' different maps of bed elevation from the ensemble, by first sampling one tree from our ensemble of trees (i.e, sampling

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

a configuration of parameters from the parameter distribution which our forest approximates) and then, at each location of the map, sampling one value from the relevant within-node empirical distribution of training observations (i.e, sampling from the data distribution).

Comparing the spatial autocorrelation properties of these 'simulated' maps to the spatial autocorrelation properties of held out test data (at the same locations) using variograms reveals them to be largely similar (Fig. 6), although there is a (somewhat rare) tendency for the regression trees to exhibit too much spatial variability at ranges of about 5 km. Despite this elevated 5 km spatial variability for some of the trees, the average of the forest as a whole (and therefore the overall spatial autocorrelation properties of our output bed map) matches quite well the spatial autocorrelation properties of the PROMICE test set (shown as the red line in Fig. 6h). The exaggerated 5 km spatial variability of some of the trees is unlikely to affect estimates of the total volume of ice, because for each tree this small-scale variability cancels out when summing over the full 250 km extent of Greenland. We therefore propose that it is reasonable to use the QRF to estimate total ice volume over Greenland.

Future work

There are several aspects of the work that warrant further investigation. Geophysical covariate observations such as gravitational field strength, could be used in the model. This may lead to improved predictions of bed elevations in areas of floating ice. Additionally, it would be worth attempting to optimise the input features, including exploring the sensitivity of the predicted bed elevation to the hyperparameters used e.g. the number of compass orientations. Future work is needed to explore the suitability of this approach for estimating bed elevations for the Antarctic ice sheets and Arctic ice caps. We suggest the next steps are to focus on quantifying the differences between BedMachine and the QRF at the outlet glaciers that drain the ice sheet interior, as estimates of ice sheet mass balance are sensitive to the geometry of these outlet glaciers.

CONCLUSIONS

The QRF model offers significant advancements over the widely-used BedMachine data in predicting Greenland's subglacial topography and ice thickness. The QRF model provides improvements to both deterministic accuracy and probabilistic predictions, achieving a 18% reduction in RMSE and a better representation of uncertainty, as demonstrated by its superior R^2 and continuous-ranked-probability score (CRPS). These

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

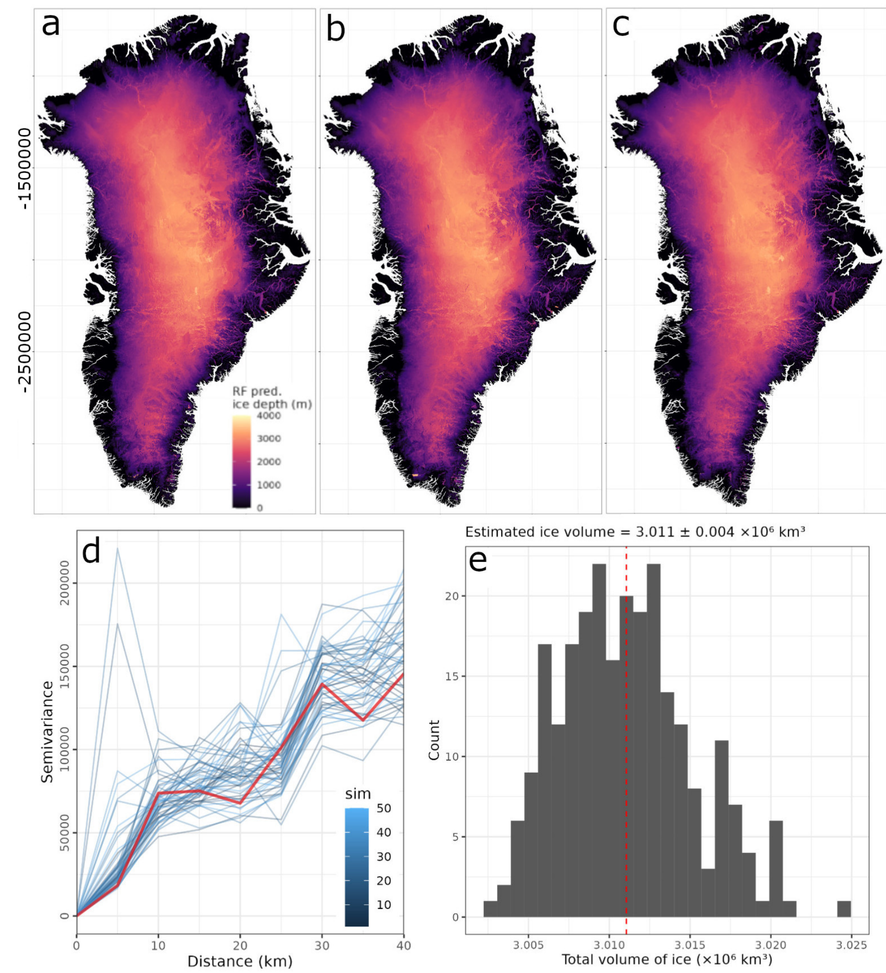

Fig. 6. Three example QRF 'simulated' ice thickness maps (a-c), Semivariograms of QRF 'simulations' (blue) vs held-out test observations (red)(d), and a histogram of QRF 'simulated' ice volumes (e). The vertical dashed red line shows the mean ice volume (3.011 ś 0.004) x 10^6 km^3. BedMachine's estimate is (2.99 ś 0.02) x 10^6 km^3.

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

improvements, while modest, hold substantial practical relevance for understanding ice dynamics in a warming climate and predicting future sea-level rise.

Our results indicate that the QRF model provides a more balanced and consistent prediction of ice thickness than BedMachine, which has a tendency to underpredict the width and depth of subglacial troughs at the ice sheet margin. This is crucial for studies of ice sheet mass balance because the cross-sectional area of these outlet glacier troughs determines the estimated rate of mass loss to the surrounding ocean.

One of the key contributions of the QRF model is its ability to produce well-calibrated uncertainty estimates, which are more reliable than those from BedMachine. The QRF model's prediction intervals captured 89.8% of test observations within a 90% confidence interval, a stark contrast to BedMachine's underdispersed intervals, which only captured 66.8%. This is crucial for decision-makers who rely on accurate risk assessments in scenarios where ice-sheet behavior is difficult to predict. For example, policymakers focused on sea-level rise adaptation would benefit from models that can account for the inherent uncertainty in ice-thickness predictions while minimizing errors in those predictions (Siegert and Pearson, 2021). These more robust uncertainty estimates are also useful for identifying priority areas for future RES surveys to target.

In terms of mapping subglacial topography, the QRF model unveils a richer and more detailed landscape than BedMachine, especially in regions where traditional methods like mass conservation and gravity inversion tend to over-smooth high-relief subglacial features. The model's ability to preserve the linearity of crucial features of the sublglacial landscape, such as ravines and channels, enhances our understanding of how these features influence ice flow and the routing of basal water.

Despite its strengths, the QRF model does have limitations, particularly in accounting for spatial dependence across locations. However, its structured ensemble approach partially captures these spatial relationships, leading to realistic estimates of total ice volume. This feature holds promise for improving future ice-sheet models that require robust volume predictions and well-calibrated uncertainty estimates. We suggest that the optimization of input features would be a worthwhile focus for a future research project.

Overall, we have demonstrated that a QRF approach represents a computationally low-cost option for estimating ice sheet basal topography, offering both improved accuracy and more robust uncertainty quantification when compared with BedMachine in grounded-ice areas.

https://doi.org/10.1017/jog.2025.10071 Published online by Cambridge University Press

REFERENCES

Bamber J, Griggs J, Hurkmans R, Dowdeswell J, Gogineni S, Howat I, Mouginot J, Paden J, Palmer S, Rignot E and Steinhage D (2013) A new bed elevation dataset for Greenland. *The Cryosphere*, **7**(2), 499–510 (doi: 10.5194/tc-7-499-2013)

Bartlett OT, Palmer SJ, Schroeder DM, MacKie EJ, Barrows TT and Graham AG (2020) Geospatial simulations of airborne ice-penetrating radar surveying reveal elevation under-measurement bias for ice-sheet bed topography. *Annals of Glaciology*, **61**(81), 46–57 (doi: 10.1017/aog.2020.35)

Bingham RG and Siegert MJ (2007) Radar-derived bed roughness characterization of Institute and Möller ice streams, West Antarctica, and comparison with Siple Coast ice streams. *Geophysical Research Letters*, **34**(21) (doi: 10.1029/2007GL031483)

Bingham RG and Siegert MJ (2009) Quantifying subglacial bed roughness in Antarctica: implications for ice-sheet dynamics and history. *Quaternary Science Reviews*, **28**(3-4), 223–236 (doi: 10.1016/j.quascirev.2008.10.014)

Breiman L (2001) Random forests. *Machine learning*, **45**, 5–32 (doi: 10.1023/a:1010933404324)

Chu W, Schroeder DM, Seroussi H, Creyts TT, Palmer SJ and Bell RE (2016) Extensive winter subglacial water storage beneath the Greenland Ice Sheet. *Geophysical Research Letters*, **43**(24), 12484–12492 (doi: 10.1002/2016GL071538)

Cooper MA, Jordan TM, Schroeder DM, Siegert MJ, Williams CN and Bamber JL (2019) Subglacial roughness of the Greenland Ice Sheet: relationship with contemporary ice velocity and geology. *The Cryosphere*, **13**(11), 3093–3115 (doi: 10.5194/tc-13-3093-2019)

De Bruin S, Brus DJ, Heuvelink GB, van Ebbenhorst Tengbergen T and Wadoux AMC (2022) Dealing with clustered samples for assessing map accuracy by cross-validation. *Ecological Informatics*, **69**, 101665 (doi: 10.1016/j.ecoinf.2022.101665)

Dowdeswell JA and Evans S (2004) Investigations of the form and flow of ice sheets and glaciers using radio-echo sounding. *Reports on Progress in Physics*, **67**(10), 1821 (doi: 10.1088/0034-4885/67/10/R03)

Durand G, Gagliardini O, Favier L, Zwinger T and Le Meur E (2011) Impact of bedrock description on modeling ice sheet dynamics. *Geophysical Research Letters*, **38**(20) (doi: 10.1029/2011GL048892)

Fox-Kemper B, Hewitt H, Xiao C, Aðalgeirsdóttir G, Drijfhout S, Edwards T, Golledge N, Hemer M, Kopp R, Krinner G, Mix A, Notz D, Nowicki S, Nurhati I, Ruiz L, Sallée JB, Slangen A and Yu Y (2021) *Ocean, Cryosphere and Sea Level Change*, 12111362. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (doi: 10.1017/9781009157896.011)

Gneiting T and Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, **102**(477), 359–378 (doi: 10.1198/016214506000001437)

Gudmundsson GH (1997) Ice deformation at the confluence of two glaciers investigated with conceptual map-plane and flowline models. *Journal of Glaciology*, **43**(145), 537–547 (doi: 10.3189/S0022143000035140)

Gudmundsson GH (2003) Transmission of basal variability to a glacier surface. *Journal of Geophysical Research: Solid Earth*, **108**(B5) (doi: 10.1029/2002JB002107)

Hastie T, Tibshirani R and Friedman J (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2nd edition, ISBN 9780387848570

Hengl T, Heuvelink GB, Kempen B, Leenaars JG, Walsh MG, Shepherd KD, Sila A, MacMillan RA, Mendes de Jesus J, Tamene L and others (2015) Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PloS One*, **10**(6), e0125814 (doi: 10.1371/journal.pone.0125814)

Hengl T, Nussbaum M, Wright MN, Heuvelink GB and Gräler B (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, **6**, e5518 (doi: 10.7717/peerj.5518)

Hersbach H (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, **15**(5), 559–570 (doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2)

Hoffman MJ, Andrews LC, Price SF, Catania GA, Neumann TA, Lüthi MP, Gulley J, Ryser C, Hawley RL and Morriss B (2016) Greenland subglacial drainage evolution regulated by weakly connected regions of the bed. *Nature Communications*, **7**(1), 13903 (doi: 10.1038/ncomms13903)

Howat IM, Negrete A and Smith BE (2014) The Greenland Ice Mapping Project (GIMP) land classification and surface elevation data sets. *The Cryosphere*, **8**(4), 1509–1518 (doi: 10.5194/tc-8-1509-2014)

Jordan TM, Cooper MA, Schroeder DM, Williams CN, Paden JD, Siegert MJ and Bamber JL (2017) Self-affine subglacial roughness: consequences for radar scattering and basal water discrimination in northern Greenland. *The Cryosphere*, **11**(3), 1247–1264 (doi: 10.5194/tc-11-1247-2017)

Joughin I, Smith BE, Howat I and Scambos T (2015) MEaSUREs Greenland Ice Sheet Velocity Map from InSAR Data, Version 2 (doi: 10.5067/OC7B04ZM9G6Q)

Joughin I, Smith BE and Howat IM (2018) A complete map of Greenland ice velocity derived from satellite data collected over 20 years. *Journal of Glaciology*, **64**(243), 1–11 (doi: 10.1017/jog.2017.73)

Kendrick A, Schroeder D, Chu W, Young TJ, Christoffersen P, Todd J, Doyle S, Box J, Hubbard A, Hubbard B and others (2018) Surface meltwater impounded by seasonal englacial storage in West Greenland. *Geophysical Research Letters*, **45**(19), 10–474 (doi: 10.1029/2018GL079787)

Kirkwood C, Cave M, Beamish D, Grebby S and Ferreira A (2016) A machine learning approach to geochemical mapping. *Journal of Geochemical Exploration*, **167**, 49–61 (doi: 10.1016/j.gexplo.2016.05.003)

Kohavi R and others (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence*, volume 14, 1137–1145, Montreal, Canada

Kuhn M and Johnson K (2019) *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Chapman & Hall/CRC Data Science Series, CRC Press, ISBN 9781351609463 (doi: 10.1201/9781315108230)

Lapazaran J, Otero J, Martín-Español A and Navarro F (2016) On the errors involved in ice-thickness estimates I: ground-penetrating radar measurement errors. *Journal of Glaciology*, **62**(236), 1008–1020 (doi: 10.1017/jog.2016.93)

Meinshausen N and Ridgeway G (2006) Quantile regression forests. *Journal of Machine Learning Research*, **7**(6), 983–999 (doi: 10.5555/1248547.1248582)

Møller AB, Beucher AM, Pouladi N and Greve MH (2020) Oblique geographic coordinates as covariates for digital soil mapping. *Soil*, **6**(2), 269–289 (doi: 10.5194/soil-6-269-2020)

Morlighem M, Seroussi H, Larour E and Rignot E (2013) Inversion of basal friction in Antartica using exact and incomplete adjoints of a higher-order model. *Journalof Geophysical Research*, **118**, 1746–1753 (doi: 10.1002/jgrf.20125)

Morlighem M, Williams CN, Rignot E, An L, Arndt JE, Bamber JL, Catania G, Chauché N, Dowdeswell JA, Dorschel B and others (2017) BedMachine v3: Complete bed topography and ocean bathymetry mapping of Greenland from multibeam echo sounding combined with mass conservation. *Geophysical Research Letters*, **44**(21), 11–051 (doi: 10.1002/2017GL074954)

Paden J, Li J, Leuschen C, Rodriguez-Morales F and Hale R (2010 updated 2019) IceBridge MCoRDS L2 Ice Thickness, Version 1 Greenland. *Boulder, CO: NASA National Snow and Ice Data Center Distributed Active Archive Center* (doi: 10.5067/GDQ0CUCVTE2Q)

Palmer S, Shepherd A, Nienow P and Joughin I (2011) Seasonal speedup of the Greenland Ice Sheet linked to routing of surface water. *Earth and Planetary Science Letters*, **302**(3-4), 423–428 (doi: 10.1016/j.epsl.2010.12.037)

Palmer S, McMillan M and Morlighem M (2015) Subglacial lake drainage detected beneath the Greenland Ice Sheet. *Nature Communications*, **6**(1), 1–7 (doi: 10.1038/ncomms9408)

Palmer SJ, Dowdeswell JA, Christoffersen P, Young DA, Blankenship DD, Greenbaum JS, Benham T, Bamber J and Siegert MJ (2013) Greenland subglacial lakes detected by radar. *Geophysical Research Letters*, **40**(23), 6154–6159 (doi: 10.1002/2013GL058383)

Prasad AM, Iverson LR and Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, **9**, 181–199 (doi: 10.1007/s10021-005-0054-1)

R Core Team (2023) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria

Robin GdQ, Evans S and Bailey JT (1969) Interpretation of radio echo sounding in polar ice sheets. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **265**(1166), 437–505 (doi: 10.1098/rsta.1969.0063)

Schroeder DM, Bingham RG, Blankenship DD, Christianson K, Eisen O, Flowers GE, Karlsson NB, Koutnik MR, Paden JD and Siegert MJ (2020) Five decades of radioglaciology. *Annals of Glaciology*, **61**(81), 1–13 (doi: 10.1017/aog.2020.11)

Sekulić A, Kilibarda M, Heuvelink G, Nikolić M and Bajat B (2020) Random forest spatial interpolation. *Remote Sensing*, **12**(10), 1687 (doi: 10.3390/rs12101687)

Siegert M and Pearson P (2021) Reducing uncertainty in 21st century sea-level predictions and beyond. *Frontiers in Environmental Science*, **9**, 751978 (doi: 10.3389/fenvs.2021.751978)

Smith BE, Raymond CF and Scambos T (2006) Anisotropic texture of ice sheet surfaces. *Journal of Geophysical Research: Earth Surface*, **111**(F1) (doi: 10.1029/2005JF000393)

Sørensen LS, Simonsen SB, Forsberg R, Stenseng L, Skourup H, Kristensen SS and Colgan W (2018) Circum-Greenland, ice-thickness measurements collected during PROMICE airborne surveys in 2007, 2011 and 2015. *GEUS Bulletin*, **41**, 79–82 (doi: 10.34194/geusb.v41.4348)

Studinger M, Koenig L, Martin S and Sonntag J (2010) Operation Icebridge: using instrumented aircraft to bridge the observational gap between Icesat and Icesat-2. In *2010 IEEE International Geoscience and Remote Sensing Symposium*, 1918–1919, IEEE (doi: 10.1109/IGARSS.2010.5650555)

Van den Broeke MR, Enderlin EM, Howat IM, Kuipers Munneke P, Noël BP, Van De Berg WJ, Van Meijgaard E and Wouters B (2016) On the recent contribution of the Greenland Ice Sheet to sea level change. *The Cryosphere*, **10**(5), 1933–1946 (doi: 10.5194/tc-10-1933-2016)

Wadoux AMJC and Heuvelink GBM (2023) Uncertainty of spatial averages and totals of natural resource maps. *Methods in Ecology and Evolution*, **14**(5), 1320–1332 (doi: 10.1111/2041-210X.14106)

Wright MN and Ziegler A (2017) ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, **77**(1), 117 (doi: 10.18637/jss.v077.i01)

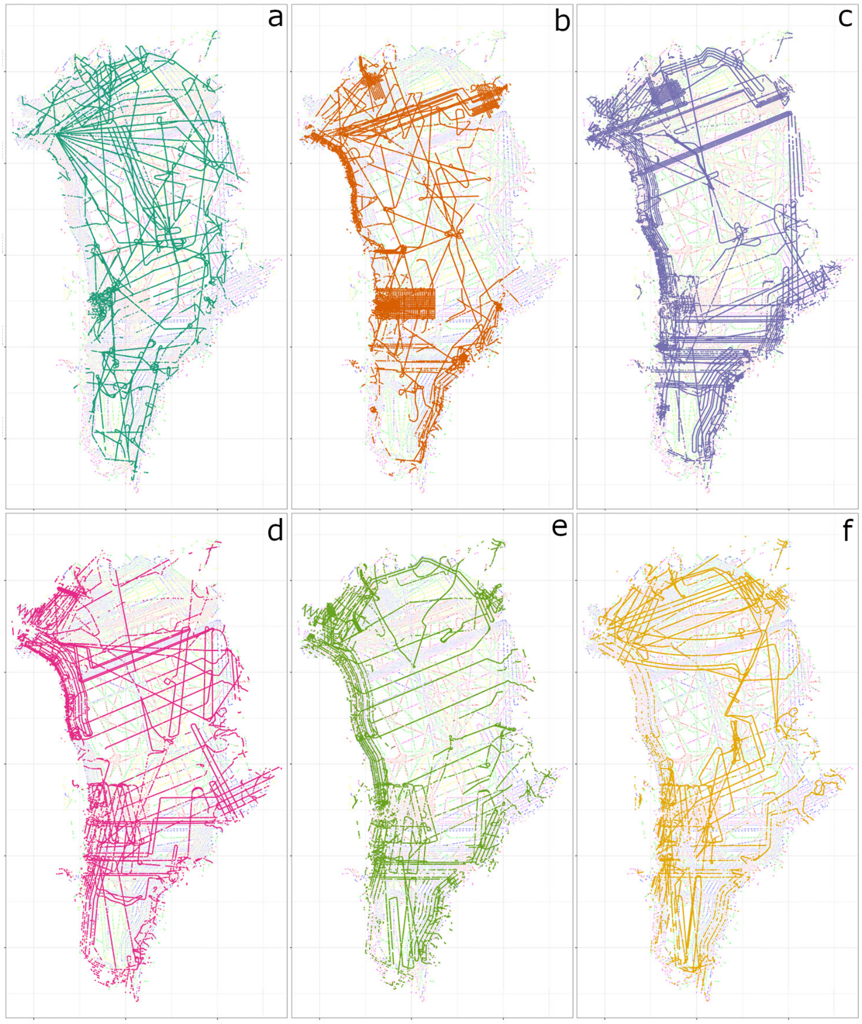

Fig. 7. Our six cross-validation folds (a-f). The folds were created by grouping observations by years of collection in order to produce six folds of approximately equal size with minimal overlap in flight lines, so that cross-validation reveals our QRF's ability to make reasonable predictions in the spaces between flight lines.

APPENDIX: CROSS-VALIDATION BY FIELD SEASON

To evaluate the ability of our QRF to make reasonable predictions in the (sometimes large) spaces between radar observation flight lines, and in order to tune its hyperparameters towards this purpose, we split our training data into six different folds according to their years of collection: 1993-2001 fold one, 2002-2010 fold two, 2011:2012 fold three, 2013:2014 fold four, 2015:2017 fold five, 2018:2019 fold six. These groupings produce folds of approximately equal size, and with little overlap between their respective flight lines (Fig. 7).