Received: 3 April 2024 | Revised: 27 August 2024 | Accepted: 23 September 2024

DOI: 10.1111/ejss.13589

RESEARCH ARTICLE

European Journal of Soil Science WILEY

Mapping soil thickness by accounting for right-censored data with survival probabilities and machine learning

Stephan van der Westhuizen[1,2,3] | **Gerard B. M. Heuvelink**[2,3] | **David P. Hofmeyr**[4] | **Laura Poggio**[3] | **Madlene Nussbaum**[5] | **Colby Brungard**[6]

[1]Department of Statistics and Actuarial Science, Stellenbosch University, Stellenbosch, South Africa

[2]Soil Geography and Landscape Group, Wageningen University, Wageningen, The Netherlands

[3]ISRIC-World Soil Information, Wageningen, The Netherlands

[4]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

[5]Department of Physical Geography, Utrecht University, Utrecht, The Netherlands

[6]Department of Plant and Environmental Sciences, New Mexico State University, Las Cruces, New Mexico, USA

Correspondence
Stephan van der Westhuizen, Department of Statistics and Actuarial Science, Stellenbosch University, Stellenbosch, South Africa.
Email: stephanvdw@sun.ac.za

Funding information
National Research Foundation (South Africa), Grant/Award Number: 129856

Abstract

In digital soil mapping, modelling soil thickness poses a challenge due to the prevalent issue of right-censored data. This means that the true soil thickness exceeds the depth of sampling, and neglecting to account for the censored nature of the data can lead to poor model performance and underestimation of the true soil thickness. Survival analysis is a well-established domain of statistical modelling that can deal with censored data. The random survival forest is a notable example of a survival-related machine learning approach used to address right-censored soil property data in digital soil mapping. Previous studies that employed this model either focused on mapping the probability of soil thickness exceeding certain depths, and thereby not mapping soil thickness itself, or dismissed it due to perceived poor performance. In this study, we propose an alternative survival model to map soil thickness that is based on the inverse probability of censoring weighting. In this approach, calibration data are weighted by the inverse of the probability that soil thickness exceeds a certain depth, that is, a survival probability. These weights can then be used with most machine learning models. We used the weights with a regular random forest, and compared it with a random survival forest, and other strategies for handling right-censored data, through a comprehensive synthetic simulation study and two real-world case studies. The results suggest that the weighted random forest model produces competitive predictions, establishing it as a viable option for mapping right-censored soil property data.

KEYWORDS
digital soil mapping, inverse probability of censoring weighting, random survival forest, soil depth, survival analysis

1 | INTRODUCTION

Over the past decade, the number of applications of machine learning in digital soil mapping (DSM) has increased dramatically, and it is now common to use machine learning to map key soil properties, such as soil organic carbon, nitrogen and pH (Minasny & McBratney, 2016). One reason for this is because 'off-the-shelf' machine learning models, in

particular random forest (RF) models, often produce soil maps with better prediction accuracy compared with other models such as multiple linear regression or geostatistical models (Wadoux & McBratney, 2020). Soil thickness, generally defined as the depth from the soil surface to the lithic or paralithic contact (U.S. Department of Agriculture, 2017), is another soil property that is important to map. It provides vital information for studies involving, for example, carbon storage (Olson & Al-Kaisi, 2015), crop suitability assessment (Fan et al., 2016) and plant habitat functions (Siemer et al., 2014). However, soil thickness remains of the poorest mapped soil properties in DSM (Chen et al., 2022).

Soil thickness may not be homogeneously defined as it may depend on the subsequent usage of this property, for example, for assessment of different soil functions (Greiner et al., 2017; Leenaars et al., 2018). For example, plant exploitable (effective) soil thickness and water storage capacity exceed beyond the weathered and structured soil layer into loose sediment parent materials formed from glacial till or fluvial deposits (Amelung et al., 2018). Consequently, achieving a homogeneous multi-purpose recording of soil thickness is not a simple task (BGS-SSP, 2010). Another difficulty concerning mapping soil thickness is that it is often right-censored which means that the true soil thickness is larger than the sampling depth (Chen et al., 2019; Malone & Searle, 2020). Right-censored soil thickness data occur for two main reasons: (1) determination of soil depth may not be part of the study for which samples are obtained, in which case surveyors may be instructed to auger no deeper than a given depth as that is the part of the soil which is of interest; (2) soil sampling to depth could be expensive and needs specialised equipment, and even with specialised equipment, there is a maximum feasible sampling depth (Malone & Searle, 2020). It is important to account for the censored nature of soil thickness data, because if right-censored data are treated as if they were true measurements, then predictions may severely underestimate the true soil thickness (Shangguan et al., 2017; Vaysse & Lagacherie, 2015).

Mapping soil thickness has been a subject of considerable effort, and there are two predominant approaches found in the scientific literature—mechanistic modelling and statistical modelling. In this paper, we will only focus on the latter approach, and for further information about mechanistic models, we refer the reader to Bonfatti et al. (2018), Minasny and McBratney (1999, 2006), and Schoorl et al. (2002). Examples of statistical methods that have been used to model soil thickness range from traditional methods such as principal component analysis, correlation analysis and linear regression (Chaplot et al., 2010; Florinsky et al., 2002; Moore et al., 1993; Odeh et al., 1991; Zhang et al., 2018), to geostatistical models (Kuriakose et al., 2009; Nussbaum et al., 2018; Odeh et al., 1995) and machine learning models, such as RF (Baltensweiler et al., 2021; Nussbaum et al., 2018; Tesfa et al., 2009), quantile regression forest (QRF) (Liu et al., 2022), support vector machines (Suleymanov et al., 2021), cubist and gradient boosting (Mulder et al., 2016; Nussbaum et al., 2018). None of the above mentioned studies used statistical models which accounted for the fact that some of the soil thickness observations are right-censored (Chen et al., 2019; Malone & Searle, 2020).

There are, however, studies that used statistical methods that incorporate strategies for dealing with right-censored data in mapping soil thickness. Kempen et al. (2015) predicted peat thickness in the Netherlands and corrected for censored observations by adding simulated values from a beta distribution. Lacoste et al. (2016) accounted for right-censored data by adding a fixed value of 30 cm to the soil depths that were censored. Brungard et al. (2021) converted soil thickness data to soil depth classes, that is, a categorical variable. Another way of dealing with right-censored soil thickness data was described in Chen et al. (2019), which used random survival forest (SRF) (Ishwaran et al., 2008) to map the probability that soil thickness exceeds certain depths (i.e. probabilistic prediction). While the study by Chen et al. (2019) did not generate maps of soil thickness itself, the authors did offer guidance on how to create these maps from the SRF model output. This included suggestions such as calculating the median from the predicted probability distribution or using the soil thickness values obtained at a predefined probability threshold.

In principle, knowing the exceedance probabilities at all depths can be used to obtain point and interval predictions of soil thickness because the exceedance probabilities define the cumulative probability distribution of soil thickness, but SRF and other survival models only produce exceedance probabilities for depths observed in the calibration data set. To the best of our knowledge, no study in DSM has produced soil thickness maps with SRF.

Highlights

- We propose a survival machine learning approach to model right-censored soil thickness data.
- Comparisons are made between our methodology and others for dealing with right-censored data.
- Survival machine learning is a viable option when the proportion of censoring is not too large.

Furthermore, it was mentioned in Malone and Searle (2020) that SRF yielded unsatisfactory results, which led to the authors using a different modelling approach to map soil thickness. This was mainly due to the complex structure of soil thickness data in Australia, which included many rock outcrop and very deep soil observations (e.g. soils surveyed deeper than 10 m), as well as a large proportion of censored data. The authors decided to use an alternative non-survival approach where three separate RF models were used to map soil thickness. The first RF model classified the occurrence of rock outcrops, the second model predicted soil thickness within a 0–2 m range, while accounting for right-censored data with an approach similar to what was used in Kempen et al. (2015), and the third model classified the occurrence of deep soils. One drawback of this approach is that three separate models need to be calibrated which is more complex.

Modelling results may also be less accurate if the proportion of censored data is large. In Willems et al. (2018), the authors conducted a simulation study and observed noteworthy bias in the estimated probability function of the response variable when the proportion of censored data was as large as 35% when using standard survival methods. Therefore, an additional survival method to SRF might also be needed to model right-censored soil thickness data. In Malone and Searle (2020), the proportion of censored data was close to 40%, which could also have contributed to the poor results of SRF.

In this paper, we propose an inverse probability of censoring weighting (IPCW) method, similar to Vock et al. (2016), to model soil thickness with a regular RF model, but with the calibration data weighted according to the exceedance probabilities of the response variable. Using a regular RF model has the advantage of producing soil thickness maps without the intermediate steps of obtaining point predictions from the exceedance probabilities as required with SRF. The method we propose involves two main stages. First, it estimates a survival function of censored soil thickness data from which it then calculates the inverse probability of censoring (IPC) weights for each observation in the calibration set. Second, it incorporates these weights with the RF model. It should be noted that the IPCW method is model-agnostic as long as the model can incorporate observation-based weights.

The main objective of this paper is to illustrate the use of the IPC weighted RF, from now on denoted as IPC weighted random forest (IPC-RF), and compare it with SRF and other strategies for dealing with right-censored data for mapping soil thickness. The models are compared by means of a comprehensive synthetic simulation study in which we investigate model performance, for example, with different proportions of censored data. The models are also compared in two real-world case studies, one from Zurich, Switzerland, and one from Maine, USA.

2 | MATERIALS AND METHODS

2.1 | Handling censored data

In the statistical literature, there are four main approaches for dealing with right-censored data. These are: (1) complete-data analysis; (2) imputation; (3) dichotomous data analysis and (4) survival analysis. Note that data can also be left-censored, interval-censored or truncated. In this paper, we only focus on right-censored soil property data and how it is accounted for with survival analysis. For a detailed discussion on the different types of censoring and truncation and how to handle them, we refer the reader to Leung et al. (1997), McDonald and Moffitt (1980), Tobin (1958), and Turkson et al. (2021).

In the complete-data analysis approach, censored data are ignored, and only the observations in the data set that are non-censored are considered (i.e. hard observations). This is the most simple approach for dealing with censored data, but it could also be very inefficient because there is a loss of information when the sample size is lowered by not using the censored data. Moreover, it can also lead to biased results if the censored observations were not censored at random, that is, if the censoring mechanism is possibly informative (refer to Web Appendix A).

Imputation refers to the process of filling in data for observations that were censored. This method is popular with missing data but may not be appropriate for censored data. For right-censored data, two extreme approaches to impute censored data are to either use left-point or right-point imputation. The former assumes that the true observation will be observed almost immediately after censoring, that is, the surveyed depth is then used, while the latter assumes that the true observation is very unlikely to be observed, and therefore, a very large value for the censored observation is imputed. If this approach is considered, stringent assumptions about the censoring mechanism are needed (Leung et al., 1997). This added complexity makes this approach not an ideal choice for handling censored data. In Baltensweiler et al. (2021), the authors used left-point imputation as only a small part of the data was censored.

With the dichotomous data analysis approach, only the occurrence versus the nonoccurence of the true observation is analysed with a classification model (e.g. logistic regression). The problem with this approach is again a loss of information because the observation values themselves are ignored, and we only model the

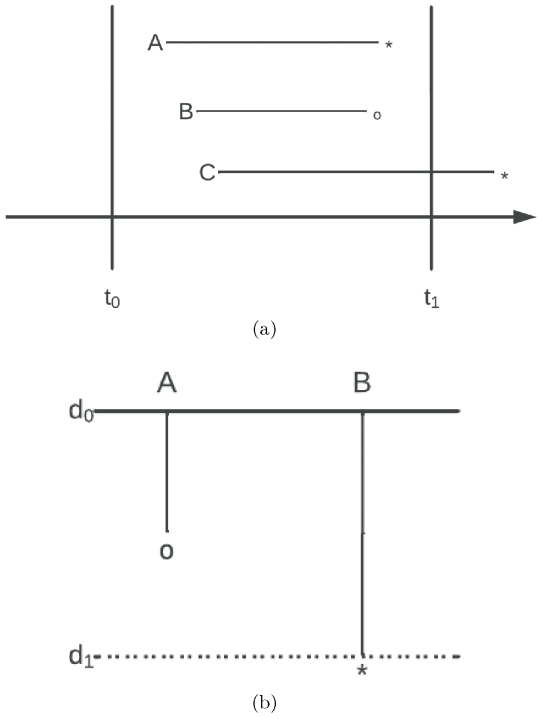

FIGURE 1 Graphical illustration of survival analysis and types of censored observations in (a) a typical epidemiological study and in (b) a digital soil mapping soil thickness study. * indicates an event of interest.

probability of observing the true observation. Recall that a part of the modelling approach in Malone and Searle (2020) analysed the occurrence of rock outcrops and deep soils as dichotomous variables.

Survival analysis refers to a collection of statistical methods with the goal of modelling a time-to-event response variable, also referred to as survival time. For example, in an epidemiological study involving cancer, the response variable could be the time until a patient goes out of remission. For some patients, the recorded data might be right-censored. This could happen when patients were lost to follow-up, or if they did not experience the event of interest while the study was conducted. In Figure 1a, an example of an epidemiological study is illustrated. As indicated on the horizontal axis, the duration of the study is from time t_0 to t_1, where '1' here refers to the end of the study. Subject A ends with an asterisk (*), which indicates an event of interest, and subject B ends with an open point (o), which indicates an event other than the event of interest. In the case of subject A, the time until the event of interest falls within the observation period, and therefore, the time of occurrence of the event of interest is known exactly, that is, not censored. Subject B is right-censored because an event other than the event of interest occurred within the observation period. Subject C is also right-censored because the event of interest occurred after time t_1, which is after the observation period ended.

In Figure 1b, an illustration of determining soil thickness, usually measured in distance (i.e. cm or m), is presented. Soil thickness at d_0 represents the soil surface and the line at d_1 represents the true soil thickness, assuming it is constant in this simple illustration. The observation at location A is censored because the true soil thickness was not reached. At location B, we reached the true soil thickness, and therefore, this observation is not censored.

2.2 | Survival analysis of right-censored soil thickness data

In the context of modelling soil thickness, the response variable is the depth at which the true soil thickness occurs, formally denoted as $D(s)$, with s a location in a region of interest, A. For simplicity, we omit the location s from the functions explained in this section. The survival function, $S(d)$, conveys information of the probability that the true depth exceeds a depth of d,

$$S(d) = P(D > d). \tag{1}$$

It is important to note that theoretically, the survival function is equal to one at the surface, that is, $S(0) = 1$, and then decreases monotonically as d increases so that $\lim_{d\to\infty} S(d) = 0$. It should also be noted that if there are locations with the lower soil boundary at the surface (e.g. rock outcrops), then $S(0)$ could be smaller than 1. The cumulative distribution function, that is, the reciprocal of Equation (1), is given by

$$\begin{aligned} F(d) &= P(D \le d), \\ &= 1 - S(d). \end{aligned} \tag{2}$$

In survival analysis, two other functions that are also often used include the hazard function and the cumulative hazard function (CHF). However, these functions are not used in this paper and hence not defined. We refer the reader to Kleinbaum et al. (2012) for detailed discussions about these functions.

Right-censoring occurs when the true soil thickness remains unknown when we augered to depth, d. Let us denote by $C(s)$, the maximum depth that we intend to sample at location $s \in \mathcal{A}$, even where $C(s) > D(s)$ (i.e. $C(s)$ can be considered as a 'censoring depth' if $C(s) < D(s)$). We then write an observation as $\{Y(s), \delta(s), x(s)\}$, where $Y(s) = \min(D(s), C(s))$, and $\delta(s) = I(D(s) \le C(s))$ and $x(s)$ is a p-vector of covariates. That is, at each sampling location, the depth reached, $Y(s)$, is recorded, as is an indicator $\delta(s)$, which tells us whether that depth was reached because the true soil thickness was obtained, that is, $Y(s) = D(s)$, or not, in which case $Y(s) = C(s)$.

Note also that it is assumed that $x(s)$ is known at all s in $\mathcal{A}$, including all sampling locations, regardless of whether an observation is censored or not.

Our aim is to predict $\{D(s), s \in \mathcal{A}\}$ from the covariates $x(s)$, using a model trained on calibration data $\{y(s_i), \delta(s_i), x(s_i)\}$ at sampling locations s_i for $i = 1, \ldots, n$. Here, $y(s_i)$, $\delta(s_i)$ and $x(s_i)$ are defined as before, and n is the sample size. Throughout this paper, it should be noted that we refer to a random variable with an uppercase letter, for example, D, C or Y, and a realisation of that variable by a lowercase letter, for example, d, c or y. Two types of prediction functions include point predictors and probabilistic predictors. If we denote s_0 to represent a prediction location, a point predictor will specify a value for $d(s_0)$, whereas a probabilistic predictor will provide an estimate of the conditional probability distribution for $D(s_0)$.

2.3 | Random survival forest for predicting soil thickness

For a detailed outline of SRF, we refer the reader to Ishwaran et al. (2008), and for details on how SRF has been implemented for modelling soil thickness, we refer to Chen et al. (2019). Here, we give a brief outline. SRF is an adaptation of the RF model and works on the same principles (Ishwaran et al., 2008). That is, trees are grown using bootstrap samples, a subset of covariates is randomly selected when tree nodes are split and the final ensemble is determined by a statistic computed on the observations in the terminal nodes (e.g. average). Two key differences between RF and SRF for regression include: (1) Instead of growing trees with the sum-of-squares splitting rule, SRF grows survival trees with the log-rank splitting rule (Ishwaran et al., 2008) that is based on the log-rank test which is a statistical test used to compare the survival functions of two groups (LeBlanc & Crowley, 1993; Segal, 1988); and (2) an ensemble statistic for SRF is the ensemble survival function. It should be noted that other splitting criteria are also available for SRF (Ishwaran et al., 2008).

Formally, consider a decision tree fit for an SRF model. Let h be a terminal node, and let $\{d_{1,h} < d_{2,h} < \ldots < d_{j,h} < \ldots < d_{m_h,h}\}$ be the m_h distinct (non-censored) depth observations in h. Furthermore, for $d_{j,h}$, $1 \le j \le m_h$, let $a_{j,h}$ be the number of (non-censored) observations with depth equal to $d_{j,h}$ and let $b_{j,h}$ be the number of observations whose depths (censored or not) exceed $d_{j,h}$ (note that $b_{j,h}$ is decreasing with j). Then, at a given node h, $1 - \frac{a_{j,h}}{b_{j,h}}$ may be interpreted as an estimate of the conditional probability of exceeding $d_{j,h}$, given previous depths have been reached. The survival function for h is then estimated from the Kaplan–Meier estimator (Ishwaran et al., 2008; Kaplan & Meier, 1958)

$$\widehat{S}_h(d) = \prod_{j: d_{j,h} \le d} \left(1 - \frac{a_{j,h}}{b_{j,h}}\right). \tag{3}$$

Given a covariate vector, $x(s_0)$, the survival function, $S(d|x(s_0))$, is estimated by dropping $x(s_0)$ down the tree, and then, the terminal node statistics are determined with

$$\widehat{S}(d|x(s_0)) = \widehat{S}_h(d), \text{ if } x(s_0) \in h.$$

Finally, the ensemble survival function is determined by averaging over all the trees. That is, if $\widehat{S}_b(d|x(s_0))$ is the estimate for the b-th tree, then the ensemble estimate is calculated with

$$\widehat{S}_e(d|x(s_0)) = \frac{1}{B}\sum_{b=1}^{B} \widehat{S}_b(d|x(s_0)), \tag{4}$$

where B is the total number of trees.

Note that although the SRF model does not provide point predictions, it is worth noting that for a non-negative random variable, say Z, with survival function S_Z, we have $E[Z] = \int_0^\infty P(Z > z)dz$. We can therefore obtain a prediction for the value of $d(s_0)$ using

$$\widehat{d}(s_0) = \int_0^\infty \widehat{S}_e(z|x(s_0))dz. \tag{5}$$

However, since $\widehat{S}_e(d|x(s_0))$ is only estimated up to $d = \max\{d(s_1), \ldots, d(s_n)\}$, attention needs to be paid to how we can extend this to all larger d, that is, it is necessary to complete the function by extrapolating into the tail. For simplicity, we assumed an exponential tail and estimated this by fitting a local-linear model to the pairs $\left\{\left(d(s_i), \log\left(\widehat{S}_e(d(s_i)|x(s_0))\right)\right)\right\}$, for $i = 1, \ldots, n$, using a kernel smoother as implemented in the package FKSUM (Hofmeyr, 2022). Then, if ℓ is the resulting local-linear estimate, we complete the tail of the estimated survival function by setting, for depth d^* larger than the greatest depth in the set of observations,

$$\widehat{S}_e(d^*|x(s_0)) = \exp(\ell(d^*)). \tag{6}$$

2.4 | Inverse probability of censoring weighting

In addition to SRF for point prediction in the presence of censoring, we propose an IPCW method which is well established in the statistical literature

(Braekers & Veraverbeke, 2005; Huang & Wolfe, 2002; Zheng & Klein, 1994) and has been applied in other domains such as the health sciences (Bandyopadhyay et al., 2014). IPCW may reduce the bias in predictions caused by censoring by correcting for locations where soil thickness was censored, by giving extra weight to the locations where soil thickness was not censored. Specifically, each sampling location is weighted by the inverse of an estimate of the probability of having remained uncensored until depth, d. Locations where soil thickness was censored receive zero weight, except if the censoring occurred beyond a predefined depth, τ (explained below). One of the advantages of the IPCW method is that it can readily be applied with most 'off-the-shelf' machine learning models as long as they can include observation weights. Below, we provide an outline of the IPCW method, and for more details and mathematical proofs, we refer the reader to Vock et al. (2016).

We estimate the censored depth survival function, $S(c)$, with the Kaplan–Meier estimator (Kaplan & Meier, 1958) analogous to Equation (3), but here it is estimated from the censored soil thickness data and from the entire calibration set (instead of just the node in a tree)

$$\widehat{S}(c) = \prod_{j:c_j < c} \left(1 - \frac{a_j^*}{b_j}\right), \tag{7}$$

where a_j^* is the number of observations that were censored at depth c_j, and b_j is the number of observations whose depths exceed c_j. Note that the definition of b_j is the same as $b_{j,h}$ in Equation (3), but a_j^* here refers to the number of observations that were censored, as opposed to $a_{j,h}$ in Equation (3), which was the number of observations where the soil thickness was reached (in node h). In addition, it should be noted that here, the Kaplan–Meier method provides a marginal estimate of $S(c)$, that is, irrespective of the covariates. Next, for each observation, we define an IPC weight (Vock et al., 2016),

$$w(s_i) = \begin{cases} \dfrac{1}{\widehat{S}(\min(Y(s_i), \tau))}, & \text{if } \delta(s_i) = 1 \text{ or } C(s_i) \geq \tau, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

We select τ to be the largest value which does not result in extremely large weights given to some observations, which could detrimentally affect the performance of the model. For example, the value for τ could be such that $\widehat{S}(\tau) = 0.1$, which means that resulting weights from Equation (8) would not be larger than 10. The τ parameter can potentially be fine-tuned, for example, with cross-validation, but this will require additional steps as the censored data need to be accounted for. For this reason, the τ parameter was not fine-tuned in this paper. The final step in the IPCW method is to incorporate the weights in Equation (8) with a machine learning model by placing the weights on the observations in the calibration data.

2.5 | Synthetic simulation study

Synthetic data simulation experiments were performed in the R programming language (R Core Team, 2022) to compare the performance of SRF and IPC-RF, and other modelling approaches, to account for right-censored soil thickness data under two main censoring scenarios. The first scenario is when censoring occurred at a fixed depth mimicking a situation where a survey follows a protocol to always reach a predefined depth, and the second is when censoring occurred at different depths for surveys without this constraint. In both of these scenarios, we investigated model performance with various sample sizes, n; proportions of censored data, ρ; and different values for the depth at which censoring occurred, λ. Note that we also investigated model performance under different censoring mechanisms (see Web Appendix B for details).

The first modelling approach to which SRF and IPC-RF was compared was a random forest model with all data (ARF), that is, regardless if data were censored or not all data were treated as non-censored data (i.e. left-point imputation). The second approach was a random forest model with only hard data (HRF), that is, censored data were excluded when the RF model was fitted. This approach was introduced in Section 2.1, that is, the complete-data analysis approach. We compared SRF and IPC-RF with ARF and HRF to investigate how true soil thickness might be underestimated when survival data are not appropriately utilised (i.e. potential bias) and how prediction accuracy might be affected.

In each experiment, we simulated a synthetic data set with response values, $z(s_k)$, for $k = 1 \cdots N$, from a zero-mean Gaussian process with a covariance structure governed by an isotropic spherical variogram model similar to the simulation study in van der Westhuizen et al. (2022). The response values were transformed to a log-normal distribution, which would then represent soil thickness, $d(s_k)$. We also created three covariates, which were functions of $z(s_k)$. A sample was obtained by randomly selecting n of the $N = 10{,}000$ grid values. For more detailed information on the simulation study, readers are directed to Web Appendix B.

In the first scenario, for a certain censoring depth, λ, we randomly selected ρ proportion of the observations that were larger than λ and replaced the observation with λ. To illustrate, suppose $\rho = 0.2$ and $\lambda = 60$, then 20% of

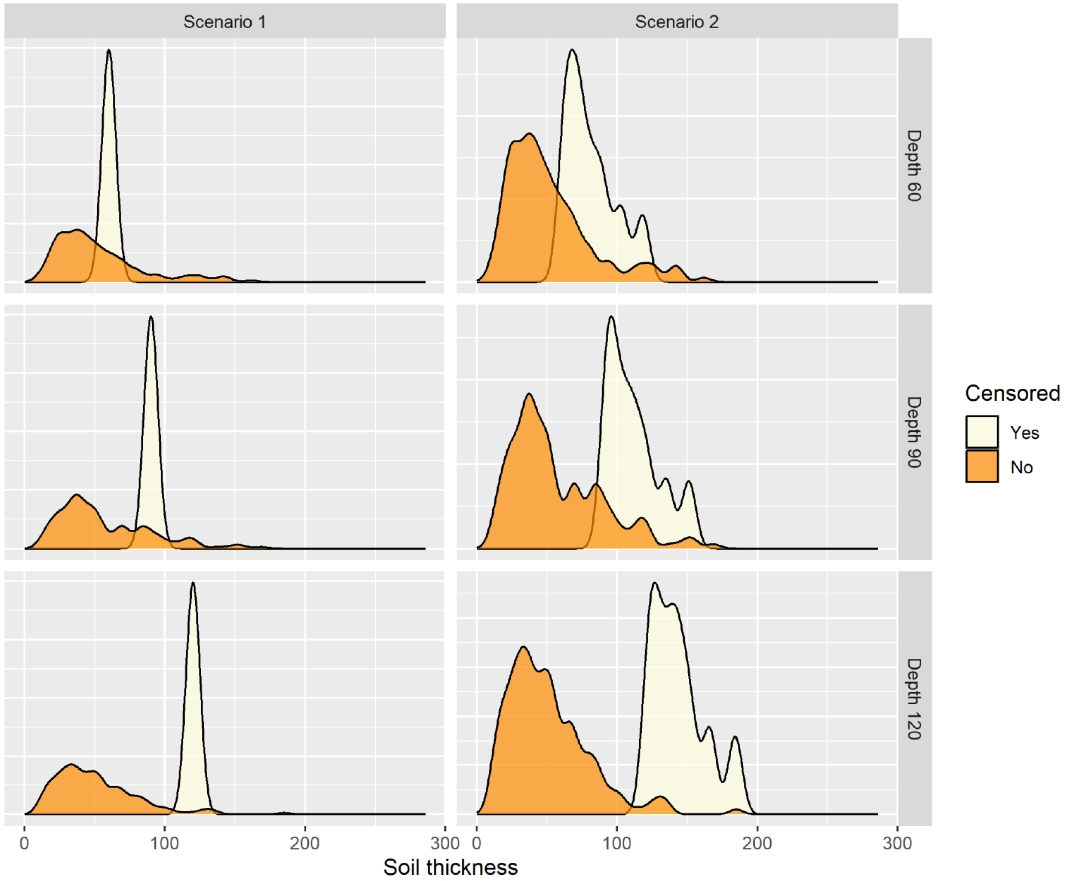

FIGURE 2 Scenarios with fixed censoring (left) and different censoring depths (right) are depicted with probability distributions from one of the simulations. Non-censored distributions are depicted with dark orange while the censored distributions are depicted by light yellow. For Scenario 1, a mixed discrete-continuous distribution is presented with the mass indicated by the light yellow (censored data). The two scenarios are shown for a censoring proportion of 0.6 and for different depths (60, 90, 120).

the observations that were larger than 60 were randomly selected and replaced with a value of 60. In the second scenario, observations to be censored were selected in the same way as in the first scenario, that is, we randomly selected ρ proportion of the observations that were larger than λ, but then, these were not set to a fixed censoring depth, but rather assigned various censoring depths. Specifically, a set of percentiles was determined from the actual depths before being censored. Each of the selected observations were then assigned one of the percentiles that was also larger than λ. Figure 2 provides an illustration of the difference between the two scenarios. For Scenario 1, a mixed discrete-continuous distribution is presented with the probability mass indicated by the light yellow (censored data at a fixed depth), while the non-censored data is presented by the dark orange distribution. The two scenarios are shown for a censoring proportion of 0.6 and for different depths (60, 90, 120).

The simulations were conducted using the following parameter values: $n = \{400, 800\}$, $\rho = \{0, 0.1, 0.3, 0.6, 0.9\}, \lambda = \{60, 90, 120\}$. It is important to note that when $\rho = 0$, it means that there were no censored observations, and that the censoring depths, λ, are multiples of $\{1.0, 1.5, 2.0\}$ times the mean of the simulated soil thickness values on the grid. Finally, each combination was repeated 200 times, resulting in a total of $60 \times 200 = 12,000$ simulations.

The response values in a selected sample are denoted by $\{y(s_i), \delta(s_i)\}$ for $i = 1, ..., n$, where $\delta(s_i)$ has the same definition as in Section 2.2. The modelling approaches (ARF, HRF, SRF, IPC-RF) were fitted on $\{y(s_i), \delta(s_i)\}$ along with $x(s_i)$ the generated covariate values. In the case of ARF and HRF, $\delta(s_i)$ was ignored, and the models were either fitted on all or only the non-censored observations. All the random forest models were implemented with the randomForestSRC package (Ishwaran et al., 2008) in R. In addition, the hyper-parameter, m_{try}, as well as the minimum node size and the number of trees, were kept at the default values, that is, 1 (when $p = 3$), 5 and 500, respectively. For IPC-RF, we used the case.wt functionality in randomForestSRC to incorporate the weights estimated with Equation (8). Note that this

argument puts the weights on the calibration observations in order to be selected for the bootstrap samples that are used to build the trees. The parameter, τ, was set such that $\left\{\widehat{S}(\tau)=0.1\right\}$. Finally, predictions were generated for the entire grid, and the models were evaluated with the mean error (ME) and root mean square error (RMSE)

$$\mathrm{ME}=\frac{1}{N}\sum_{k=1}^{N}\left(d(s_k)-\widehat{d}(s_k)\right), \tag{9}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(d(s_k)-\widehat{d}(s_k)\right)^2}, \tag{10}$$

where the $\widehat{d}(s_k), k=1...N$ are the predictions over the entire grid. Note that the $d(s_k)$ are the known true soil thickness values, which are known in this simulation study, and therefore, it was not required to account for censoring of the test data in the evaluation step as opposed to a real-world case study.

2.6 | Real-world applications

The first real-world application is from Maine, USA, which is located in the North-Eastern region of the country and has a surface area of approximately 91,646 km^2. The case study consisted of 5666 sampling locations with on average 1 sampling location per 16 km^2 (Figure 3). The locations were purposely chosen by soil surveyors during initial mapping and traditional soil survey update efforts. The dot sizes in Figure 3 are proportional to the soil thickness, and the dark orange dots represent locations where soil thickness was observed. There were 3856 locations where soil thickness was censored which meant that 68.1% of the observations were right-censored. Locations where the true soil thickness was recorded occurred mostly in the north-central and central regions of the study. Soil thickness was defined as the depth from the soil surface (including any organic horizons) to a lithic (i.e. bedrock contact).

Censoring occurred at a depth of 165 cm because this was the standardised excavation depth according to soil survey standards during the time of survey. There were seven observations that were non-censored that were deeper than 165 cm (these ranged from 170 to 213 cm). The covariates for this case study were prepared at 5 m pixel resolution and consisted of various morphological and hydrological derivatives of a state-wide LiDAR digital elevation model. Additional information about the covariates can be found in Appendix C. For computational

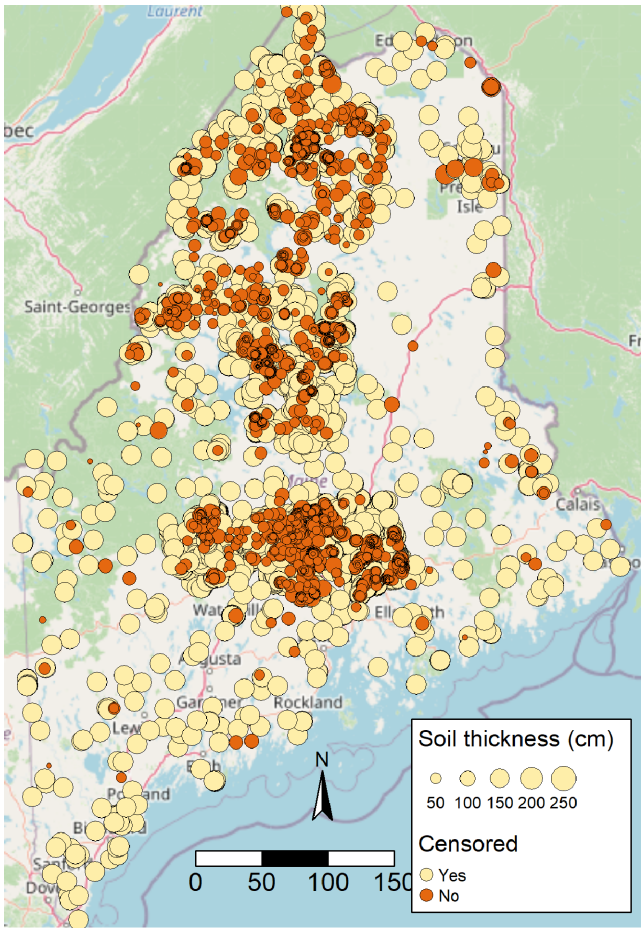

FIGURE 3 Locations of soil thickness recordings in Maine, USA.

purposes, the covariates were resampled to a resolution of 205 m. Additionally, six observations with missing values in the covariates were excluded, resulting in a total of 5660 observations available for model training.

For the second real-world application, we used soil thickness data from the arable land of the Canton of Zurich located in the North-East of Switzerland. The surface area of this region is approximately 1729 km^2. The majority (75%) of the data set originated from fully described and analysed soil profiles recorded for a conventional detailed soil mapping campaign in 1988–1997. We complemented these data with older surveyed locations, but not older than from 1975 what resulted in 3924 observations (Service center NABODAT, 2022), with thus on average 1 sampling location per 0.44 km^2 (Figure 4). Soil thickness was derived for this study from the in-situ recorded horizon qualifiers according to Swiss soil classification (Jäggli et al., 1998). Soil thickness was defined as the upper limit of the first occurring horizon considered unstructured parent material (i.e. horizons with only *C* or *R* qualifiers, excluding transition horizons). When no single *C* or *R* qualifier was recorded, then an observation

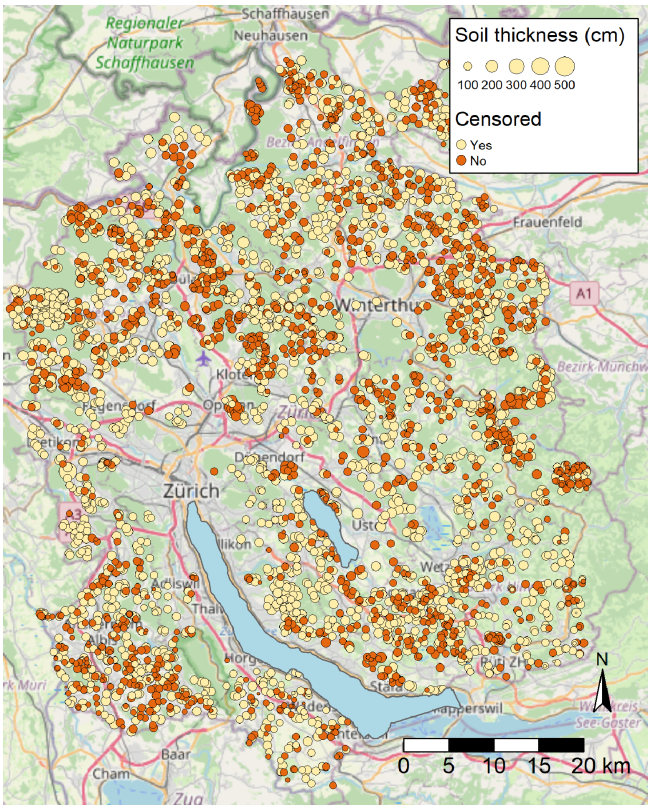

FIGURE 4 Locations of soil thickness recordings on the arable land of the Canton of Zurich, Switzerland.

was considered censored with the surveyed depth available only. In Figure 4, non-censored observations are shown as dark orange and censored observations as light yellow. Soil profile locations were purposely chosen by pedological experts to support conventional map polygon delineation (Jäggli et al., 1998). Both sets of observations are well-dispersed throughout the arable land of the Canton of Zurich. The proportion of censored observations for this case study was equal to 61.1%. The covariates used for this case study are presented in Web Appendix D.

2.7 | Model calibration and evaluation

As with the synthetic simulation study in Section 2.5, we employed the modelling approaches, ARF, HRF, SRF and IPC-RF to the two case studies introduced in Section 2.6. The models were calibrated and assessed with a k-fold nested cross-validation, which is the same approach used in van der Westhuizen et al. (2023). In this paper, we used $k = 10$ in the cross-validation and the RMSE was used for tuning. It should be noted that ARF, SRF and IPC-RF were calibrated on the non-censored and the censored data, while HRF was calibrated only on the non-censored data. For each model, we tuned the m_{try} $(10, \ldots, 30)$, and the minimum node size $(1, \ldots, 10)$ hyper-parameters. We used the default value for the number of trees as used in the RandomForestSRC package (i.e. 500 trees). In addition, for IPC-RF, we set τ, such that $\widehat{S}(\tau) = 0.1$.

To assess the models, we used standard performance metrics, that is, ME, RMSE and concordance correlation coefficient (CCC) (Lawrence & Lin, 1989). These metrics were determined on three variations of the test sets of the cross-validation. The first was to treat censored data as true observations by using left-point imputation, that is, using observation depth, thereby ignoring the censored nature of the data. This will not give a true reflection of model performance. However, to account for the fact that the test data might be censored, the second variation was to calculate the performance metrics only on non-censored data. Thirdly, we also calculated the metrics on test data that were first truncated to a certain threshold, d_{val}. The truncation is applied on non-censored observations in the test set, as well as on censored observations when the censored depth is at least as large as the threshold, d_{val}. For instance, suppose $d_{\text{val}} = 100\,\text{cm}$, then an observation (and the corresponding prediction) with soil thickness values larger than d_{val} will be truncated to 100 cm, otherwise the observation (and prediction) will not be changed. In both case studies, we performed testing with $d_{\text{val}} = \{100\,\text{cm}; 150\,\text{cm}\}$. We used this second approach as often map users are only interested in soil thickness down to a certain depth after which the knowledge of the exact soil thickness becomes less pertinent. For example, in wheat production, an agronomist might be interested in soil thickness only up to a value of 100 cm (Fan et al., 2016).

It should be noted that in survival analysis, model evaluation is usually done with Harrell's concordance index (C-index) as it accounts for censoring (Harrell et al., 1982). However, for the calculation of the C-index for a model, an estimate of $S(d)$ (or the CHF) is required. Therefore, we did not evaluate the models in this paper with this approach, as SRF is the only model that can directly produce this output. For an example of the C-index in soil thickness modelling, we refer the reader to Chen et al. (2019). For the case studies, we used the same software to implement the models as in Section 2.5.

2.8 | Evaluation by comparison of prediction maps

We also evaluated the models on the basis of prediction maps. The maps were produced by fitting the random forest models on all the training data (i.e. 5660 observations for

the Maine case study and 3924 for Zurich case study), except HRF which was fitted only on the non-censored data. Two sets of maps were produced by using different hyper-parameters. For the first set, the models were fitted with the hyper-parameters that resulted in the smallest test RMSE of the outer-loops in the cross-validation, calculated with the evaluation approach that only considered the non-censored data (refer to Section 2.7). The second set of maps was produced with the models that were fitted with hyper-parameters that resulted in the smallest test RMSE of the outer-loops in the cross-validation calculated with the truncated approach (refer to Section 2.7). We used $d_{val} = 100$ cm for the latter, and then, the map values were also truncated to $d_{val} = 100$ cm.

3 | RESULTS

3.1 | Synthetic simulation study

The RMSE results for the censoring scenarios (fixed vs. varying depths) are presented in Table 1. The table also shows the ratio (r) of censored to non-censored data for each simulation parameter combination. In Table 1, we also present the results for a RF model that was fitted with the true soil thickness which can be used as a baseline to compare with the other models. In addition, for each unique combination of simulation parameters, the best modelling approach is highlighted in boldface. The ME results are presented in Web Appendix B.

The RMSE results of the RF model indicated that regardless of the value of ρ and λ (and the censoring scenario), ARF, HRF, SRF and IPC-RF were compared against baselines of 24.0 and 22.3 when the sample size, n, was set to 400 and 800, respectively. When $\rho = 0.0$, the RMSE of ARF and HRF were comparable to that of the baseline. SRF performed worse than ARF and HRF when $\rho = 0.0$ for both censoring scenarios, while IPC-RF produced similar results to that of ARF and HRF.

A close look at Table 1 revealed the following highlights concerning the effect of simulation parameters. In general, it was observed that, for a specific censoring depth, sample size and censoring scenario, an increase in ρ led to a corresponding increase in the RMSE. As expected, this confirms a decrease in model performance as the information content of the data decreases due to fewer known soil thickness observations available to the models. For example, in the case of the first censoring scenario, when $\lambda = 60$, $\rho = 0.3$ and $n = 400$, the RMSE is equal to 26.5, 25.0, 26.3 and 24.4 for ARF, HRF, SRF and IPC-RF, respectively, and when ρ increased to 1 the RMSE increased to 39.0, 51.4, 38.9 and 39.0. For a larger sample size, at a given censoring proportion and depth, it was expected to note a decrease in RMSE. However, when $\rho = 1$, the RMSE values were similar between the two cases of $n = \{400,800\}$. Note that when $\rho = 1$, it means that all observations larger than or equal to λ were censored. Therefore, we do not expect an increase in model performance for a larger sample size. Finally, smaller RMSE values were noted in the case of higher λ values, especially for larger values of ρ. This is attributed to the increase in the number of censored observations when the censoring depth, λ, is smaller.

In terms of model performance in the first censoring scenario, IPC-RF consistently demonstrated superior results than SRF and the other modelling approaches (when $\rho \leq 0.6$, and regardless of censoring depth and sample size, with the only exception at $\rho = 0.6$ and $n = 800$ in which case SRF and IPC-RF were comparable). Then, when $\rho = 0.9$ (regardless of censoring depth and sample size), SRF exhibited the best performance, and when $\rho = 1$, the results of ARF, SRF and IPC-RF were comparable. Under the second censoring scenario, IPC-RF was consistently superior in comparison with SRF and the other modelling approaches (when $\rho \leq 0.3$, and regardless of censoring depth and sample size). Then, when $\rho \geq 0.6$, ARF was superior, surpassing IPC-RF. HRF and SRF performed poorly in the second censoring scenario, especially when λ was equal to 60. The only exception was when $\rho = 1$ wherein SRF produced comparable results with IPC-RF.

The ME results, presented in Web Appendix B, indicated that for larger values of ρ ($0.6 \leq \rho \leq 1$) and smaller values for λ ($60 \leq \lambda \leq 90$), ARF, HRF and IPC-RF increasingly underestimated the true soil thickness, while SRF increasingly overestimated it. The reason for the overestimation of SRF for larger ρ is because more observations were sooner 'reached' in Equation (3) because they were censored at λ. This then led to larger survival probabilities for data with true depths larger than λ. However, in case of the first censoring scenario, SRF underestimated the true soil thickness. This is because in Equation (3), no distinct values larger than depth, λ, were available.

3.2 | Results for the real-world applications

3.2.1 | Maine case study

The distribution of soil thickness is illustrated in Figure 5 indicating that the distribution was skewed to the right. It is important to point out that almost 100% of the data with a depth of at least 165 cm were censored (only seven observations above this threshold were non-censored).

TABLE 1 Root mean square error results computed for all grid cells of the simulated grid for both censoring scenarios.

Parameters			$n = 400$						$n = 800$					
Scenario	λ	ρ	r	**RF**	**ARF**	**HRF**	**SRF**	**IPC-RF**	r	**RF**	**ARF**	**HRF**	**SRF**	**IPC-RF**
1	60	0.0	0.00	24.0	24.0	24.0	26.4	**23.7**	0.00	22.3	22.4	22.4	24.1	**22.2**
		0.1	0.04	24.0	24.5	24.2	26.3	**23.7**	0.04	22.3	23.0	22.5	23.9	**22.3**
		0.3	0.13	24.0	26.5	25.0	26.3	**24.4**	0.12	22.3	25.2	23.3	24.1	**22.9**
		0.6	0.29	24.0	30.5	27.4	26.6	**26.3**	0.29	22.3	30.0	25.4	**24.7**	**24.7**
		0.9	0.52	24.0	36.4	36.3	**32.5**	35.4	0.51	22.3	36.0	34.1	**32.2**	33.3
		1.0	0.58	24.0	39.0	51.4	**38.9**	39.0	0.60	22.3	**38.4**	51.9	39.2	**38.4**
	90	0.0	0.00	24.0	24.0	23.9	26.3	**23.6**	0.00	22.3	22.3	22.3	23.8	**22.1**
		0.1	0.02	24.0	24.3	24.1	26.4	**23.7**	0.02	22.3	22.6	22.4	23.9	**22.2**
		0.3	0.05	24.0	25.2	24.8	26.2	**24.2**	0.05	22.3	23.6	22.9	23.9	**22.5**
		0.6	0.11	24.0	27.7	26.7	26.7	**25.4**	0.11	22.3	26.9	24.9	24.6	**24.1**
		0.9	0.17	24.0	30.4	32.6	**28.1**	31.7	0.18	22.3	30.2	30.5	**26.5**	30.0
		1.0	0.20	24.0	**31.5**	41.7	31.6	31.6	0.20	22.3	**31.4**	44.3	33.5	31.5
	120	0.0	0.00	24.0	24.0	24.0	26.4	**23.7**	0.00	22.3	22.3	22.3	23.9	**22.1**
		0.1	0.01	24.0	24.1	24.1	26.1	**23.6**	0.01	22.3	22.6	22.5	24.0	**22.2**
		0.3	0.02	24.0	24.5	24.5	26.0	**23.9**	0.02	22.3	23.3	23.1	24.2	**22.6**
		0.6	0.04	24.0	26.0	26.0	26.5	**25.1**	0.04	22.3	24.1	23.6	23.9	**23.0**
		0.9	0.07	24.0	27.4	29.1	**26.8**	28.0	0.07	22.3	26.5	27.3	**24.9**	26.4
		1.0	0.08	24.0	**28.2**	32.7	28.8	**28.2**	0.08	22.3	**27.2**	31.7	27.4	**27.2**
2	60	0.0	0.00	24.0	24.0	24.0	26.4	**23.7**	0.00	22.3	22.4	22.4	24.1	**22.2**
		0.1	0.04	24.0	24.1	24.2	26.0	**23.7**	0.04	22.3	22.5	22.5	23.7	**22.2**
		0.3	0.12	24.0	24.6	25.0	26.4	**24.2**	0.12	22.3	23.1	23.2	24.2	**22.7**
		0.6	0.27	24.0	**25.3**	27.1	28.2	25.8	0.27	22.3	**24.2**	25.2	26.5	24.4
		0.9	0.48	24.0	**26.9**	34.3	35.5	29.2	0.47	22.3	**25.8**	31.9	38.4	28.5
		1.0	0.53	24.0	**27.5**	41.1	31.0	31.0	0.55	22.3	**26.5**	39.7	31.0	30.5
	90	0.0	0.00	24.0	24.0	23.9	26.3	**23.6**	0.00	22.3	22.2	22.3	23.8	**22.1**
		0.1	0.02	24.0	24.0	24.2	26.0	**23.6**	0.02	22.3	22.4	22.4	23.7	**22.2**
		0.3	0.05	24.0	24.3	24.8	25.9	**24.1**	0.05	22.3	22.7	22.9	23.7	**22.6**
		0.6	0.11	24.0	**24.9**	26.6	26.9	25.4	0.11	22.3	**23.6**	24.6	25.3	23.8
		0.9	0.16	24.0	**25.4**	31.5	28.5	27.0	0.17	22.3	**24.4**	29.3	29.8	26.0
		1.0	0.19	24.0	**25.6**	35.9	27.6	28.4	0.18	22.3	**24.5**	34.5	26.6	27.3
	120	0.0	0.00	24.0	24.1	24.0	26.4	**23.7**	0.00	22.3	22.3	22.3	23.9	**22.1**
		0.1	0.01	24.0	24.0	24.2	25.9	**23.7**	0.01	22.3	22.4	22.5	23.9	**22.2**
		0.3	0.02	24.0	24.2	24.7	25.9	**24.1**	0.02	22.3	22.9	23.1	24.0	**22.7**
		0.6	0.04	24.0	**24.6**	25.8	26.2	25.0	0.04	22.3	**22.8**	23.6	23.8	23.1
		0.9	0.07	24.0	**24.8**	28.9	26.5	25.5	0.07	22.3	**23.5**	26.8	24.9	24.2
		1.0	0.08	24.0	**25.1**	30.8	26.8	26.2	0.07	22.3	**23.6**	29.1	24.7	24.6

Note: Results are shown for censoring proportion $\rho = \{0,0.3,0.6,0.9,1\}$, censoring depth $\lambda = \{60,90,120\}$ and sample size $n = \{400,800\}$. Model results include RF fitted on the true soil thickness, and ARF, HRF, SRF, IPC-RF, fitted on censored soil thickness. For each unique combination of simulation parameters, the best approach is presented in boldface.

Abbreviations: ARF, random forest model with all data; HRF, random forest model with only hard data; IPC-RF, inverse probability of censoring weighted random forest; RF, random forest; SRF, random survival forest.

Table 2 displays the cross-validation (outer-loop) results for the Maine case study. The table showcases the ME, RMSE and CCC values for the ARF, HRF, SRF and IPC-RF models, assessed as discussed in Section 2.7. It should be noted that ME was calculated such that a negative value corresponds to an underestimated prediction. When model testing was performed with the first approach (treating censored data as true observations or

left-point imputation), both ARF and IPC-RF outperformed HRF and SRF. This is evident from the RMSE values, which were 44.7 and 44.4 for ARF and IPC-RF, respectively, while HRF and SRF had RMSE values of 88.1 and 83.8, respectively. In case of testing on only the non-censored data, it is apparent that HRF and SRF were superior. Lastly, when testing with the truncated data, we once again observed that ARF and IPC-RF exhibited superior performance in both cases of $d_{\text{val}} = \{100\text{cm}; 150\text{cm}\}$.

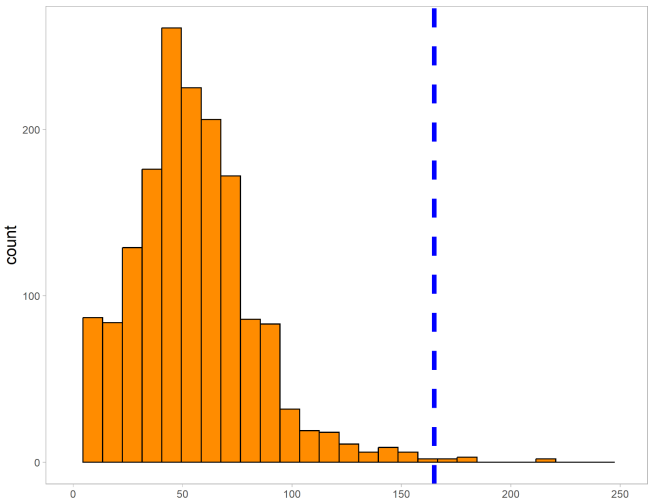

FIGURE 5 Distribution of soil thickness in Maine, USA. A histogram only for the non-censored data (dark orange) is shown based on 1804 observations. The blue dashed line represents the fixed censoring depth of 165cm.

The soil thickness maps for the Maine case study are displayed in Figure 6. For each of ARF, HRF, SRF and IPC-RF, two maps are shown. Figure 6a–d were produced with ARF, HRF, SRF and IPC-RF, respectively, fitted with the hyper-parameters which resulted in the smallest test RMSE calculated with the second evaluation approach discussed in Section 2.7 (non-censored data only). Figure 6e–h were produced with ARF, HRF, SRF and IPC-RF, fitted with the smallest test RMSE calculated with the third evaluation approach (i.e. truncated data). The latter four maps were then also truncated to 100 cm.

The maps produced by ARF and IPC-RF were very similar and showed no noteworthy differences. This is because with the IPC-RF model, τ was set to 160cm which meant that all censored data of 165 cm received a weight of one, and thereby were included in the calibration of the model. For ARF and IPC-RF, the deepest soils were observed in the central, north-eastern and south-western regions of Maine. HRF and SRF produced maps with much smaller soil thickness values, especially HRF. Most predictions produced by these two models were also less than 100cm. The 90th percentile of the predictions of HRF was 78.5cm while for the SRF model, it was 93.7cm. It is therefore clear that HRF and SRF severely underestimated (deeper) soil thickness in this study.

3.2.2 | Zurich case study

Figure 7 presents the distribution of the soil thickness data. It can be seen that the shape of the distributions of

TABLE 2 Cross-validation results for the Maine case study.

Evaluation	Metric	ARF	HRF	SRF	IPC-RF
All (censored and non-censored)	ME	−1.3	−70.8	−65.2	−0.9
	RMSE	44.7	88.1	83.8	44.4
	CCC	0.527	0.079	0.087	0.534
Non-censored	ME	49.9	0.6	5.6	49.7
	RMSE	61.3	27.4	29.3	61.4
	CCC	0.162	0.354	0.287	0.163
Val. depth (100 cm)	ME	11.6	−26.1	−20.9	11.6
	RMSE	27.2	36.9	33.5	27.2
	CCC	0.343	0.207	0.233	0.351
Val. depth (150 cm)	ME	6.1	−60.5	−55.0	6.2
	RMSE	39.8	76.1	71.7	39.7
	CCC	0.533	0.092	0.106	0.538

Note: Evaluation metrics were calculated for the outer-loops of the nested cross-validation. The results are shown when using all the data as well as for two alternative strategies, that is, evaluating on non-censored data, and with truncated observations and predictions ($d_{\text{val}} = \{100\text{cm}; 150\text{cm}\}$).
Abbreviations: ARF, random forest model with all data; CCC, concordance correlation coefficient; HRF, random forest model with only hard data; IPC-RF, inverse probability of censoring weighted random forest; ME, mean error; RMSE, root mean square error; SRF, random survival forest.

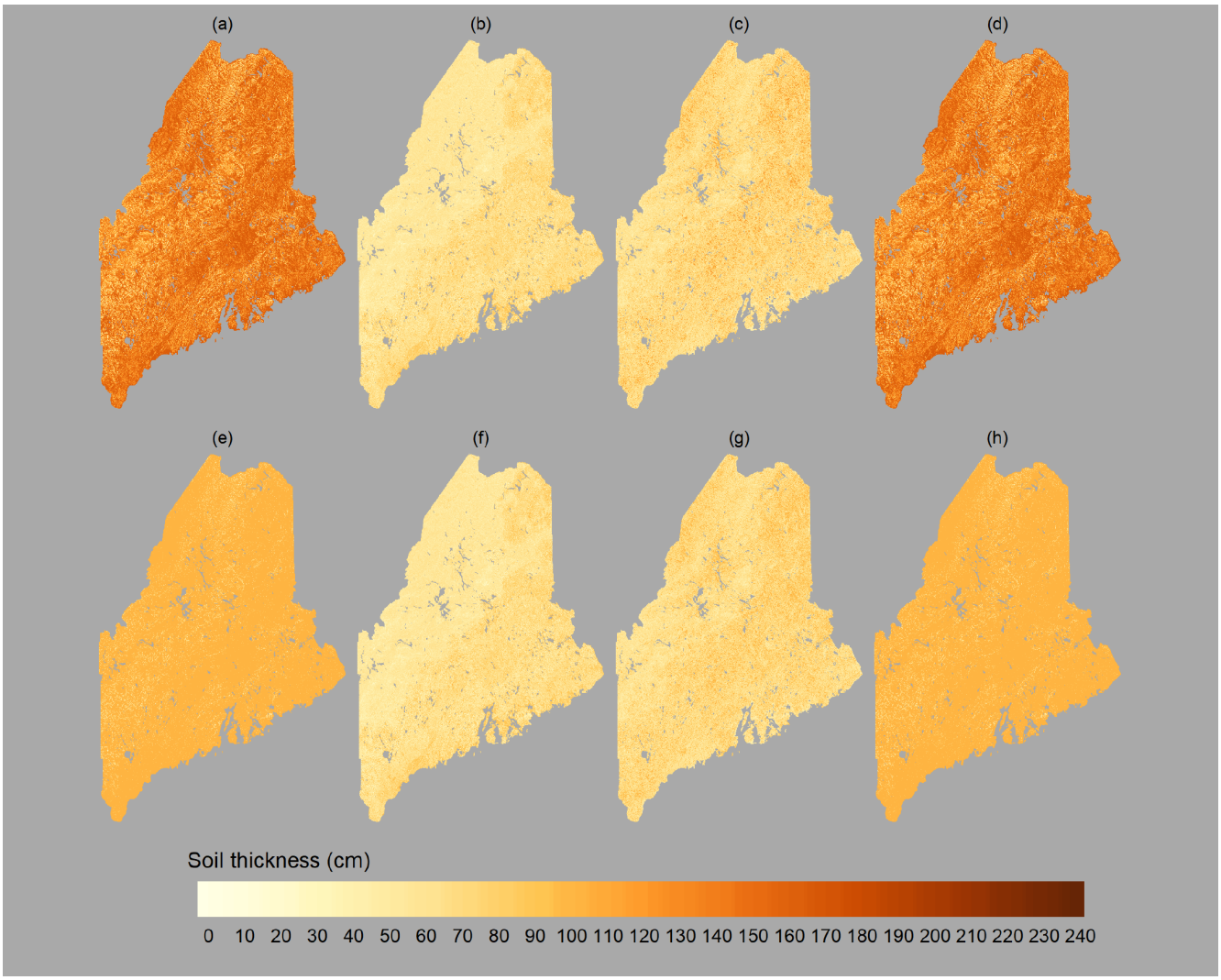

FIGURE 6 Predicted soil thickness maps (in cm) for the Maine case study for (a) random forest model with all data (ARF); (b) random forest model with only hard data (HRF); (c) random survival forest (SRF); (d) inverse probability of censoring weighted random forest (IPC-RF), fitted with the hyper-parameters which resulted in the smallest test root mean square error (RMSE) calculated with the non-censored data, and (e) ARF; (f) HRF; (g) SRF; (h) IPC-RF, fitted with the smallest test RMSE calculated with the truncated data. The latter four were also truncated to 100 cm.

the non-censored and censored data are similar (slightly right-skewed), but the distribution of the censored data (light yellow) has a larger mean. It is important to point out that about 75% of the data that were larger than the overall mean of 75 were censored.

Table 3 presents the cross-validation results for the Zurich case study. When the evaluation was done with the first evaluation approach (treating censored data as true observations), ARF demonstrated superior performance as expected. This is evident from the RMSE value which was lower than that of other models. In contrast to the Maine case study, we observed that HRF and SRF performed relatively well when testing with all the data, achieving RMSE values of 32.3 and 28.0, respectively. In terms of evaluation with only the non-censored data, HRF exhibited the best performance with an RMSE of 24.1, followed by IPC-RF with an RMSE of 25.4. When considering evaluation with the truncated data, ARF, SRF and IPC-RF were comparable in the 100 cm case. We also noted that SRF and IPC-RF produced the least biased results in the 100 cm case. In the 150 cm case IPC-RF performed best with a RMSE of 26.0 and a CCC of 0.427. However, in comparison with HRF, the difference between the RMSE values was negligible.

The maps produced by ARF, HRF, SRF and IPC-RF for the Zurich case study are presented in Figure 8. As with the Maine case study, HRF and SRF produced lower soil thickness values compared to that of ARF and IPC-RF. The truncated maps of HRF and SRF in Figure 8f,g are therefore very similar to Figure 8b,c, respectively.

ARF and IPC-RF produced larger values of soil thickness, especially IPC-RF as seen in the Western and East-central parts of the Canton of Zurich.

4 | DISCUSSION

Handling right-censored data poses a great challenge as it relates to a situation with data that have reduced information. While there has been some exploration in the literature of DSM on how to deal with censored data, as evidenced by studies such as Chen et al. (2019), Kempen et al. (2015), and Lacoste et al. (2016), users should be mindful that results will often fall short of optimal when compared to results that would have been obtained with data that were not censored (data for which the true observations are known), particularly when the proportion of censored data is large, and if censoring occurs at shallower depths. This is because less information of the true soil thickness is known, and if the censored nature of the data is then not accounted for, predictions will not reflect the true soil thickness process. The synthetic simulation study confirmed this. In Table 1, when the proportion of censored data was larger ($\rho \geq 0.3$) and when censoring occurred at shallower depths (i.e. $\lambda = 60$), all of the models produced much poorer results compared with the baseline (when the true soil thickness was known and used). A different simulation study that aimed to assess the estimation of several survival curves across different proportions of censored data (Willems et al., 2018) also revealed that results were notably worse when the proportion of censored data was larger than 0.35.

The novel application of using an IPCW random forest model to map soil thickness has to the best of our knowledge not been used in DSM. The IPC-RF model performed well in the simulation study which investigated model performance for different censoring scenarios. Specifically, IPC-RF consistently demonstrated superior performance when the proportion of censored data (that exceeded a specified threshold) was at most 0.6 for a fixed censoring

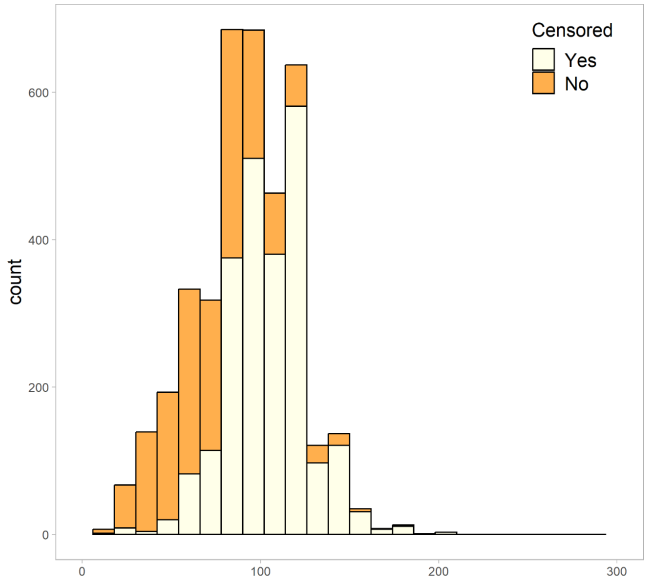

FIGURE 7 Distribution of censored and non-censored soil thickness in the Zurich case study. Two histograms are presented, one for the censored data (2350) and one for the non-censored data (1496).

TABLE 3 Cross-validation results for the Zurich case study.

Evaluation	Metric	ARF	HRF	SRF	IPC-RF
All	ME	−0.1	−17.0	−5.9	−8.5
	RMSE	26.3	32.3	28.0	29.1
	CCC	0.333	0.174	0.294	0.326
Non-censored	ME	15.0	0.2	8.5	6.5
	RMSE	28.3	24.1	26.3	25.4
	CCC	0.269	0.288	0.267	0.365
Val. depth (100 cm)	ME	6.9	−8.0	1.4	−1.6
	RMSE	20.0	21.4	19.4	19.8
	CCC	0.356	0.275	0.376	0.389
Val. depth (150 cm)	ME	12.9	−2.0	6.4	4.8
	RMSE	29.0	26.4	27.6	26.0
	CCC	0.306	0.293	0.296	0.427

Note: Evaluation metrics were calculated for the outer-loops of the nested cross-validation. The results are shown when using all the data as well as for two alternative testing strategies, that is, evaluation on non-censored data, and evaluation with truncated observations and predictions ($d_{val} = \{100\text{cm}; 150\text{cm}\}$).
Abbreviations: ARF, random forest model with all data; CCC, concordance correlation coefficient; HRF, random forest model with only hard data; IPC-RF, inverse probability of censoring weighted random forest; ME, mean error; RMSE, root mean square error; SRF, random survival forest.

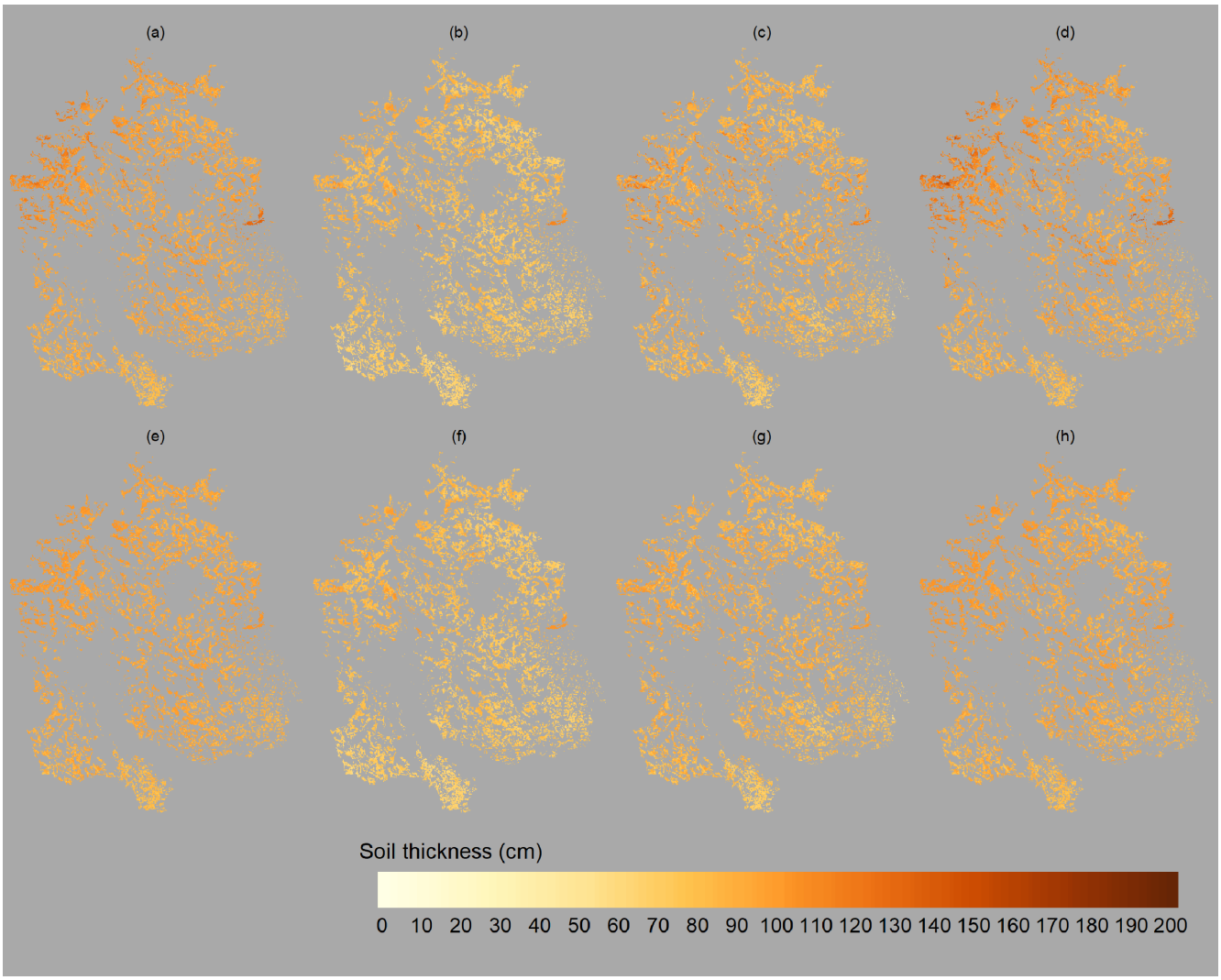

FIGURE 8 Predicted soil thickness maps (in cm) of arable land for the Zurich case study for (a) random forest model with all data (ARF); (b) random forest model with only hard data (HRF); (c) random survival forest (SRF); (d) inverse probability of censoring weighted random forest (IPC-RF), fitted with the hyper-parameters that resulted in the smallest test root mean square error (RMSE) calculated on the non-censored data, and (e) ARF; (f) HRF; (g) SRF; (h) IPC-RF, fitted with the smallest test RMSE calculated on the truncated data. The latter four were also truncated to 100 cm.

depth and no more than 0.3 for varying censoring depths. The superior performance of IPC-RF is attributed to the model's ability to directly account for censored data by assigning greater weight to the non-censored data. This results in an overall improved fitted random forest model and, consequently, more accurate predictions.

In the simulation study, in the case where all observations that exceeded a certain depth were censored ($\rho = 1$), the results revealed that the best option is to either use ARF or IPC-RF in case of Scenario 1, and to use ARF in case of Scenario 2. The reason for the comparable performance between ARF and IPC-RF in the first scenario is because the τ parameter allowed all censored observations to be included in the calibration step by giving equal weight to all. These findings suggest that due to the limited information available about the true soil thickness when $\rho = 1$, using a survival-related model like SRF may not necessarily confer an advantage. In such cases, a user might as well opt for an approach like ARF or a modelling approach in which the occurrence of the true (deep) soil thickness is modelled with an dichotomous data analysis approach as used in Malone and Searle (2020). Note that we did not investigate the performance of such an approach in this study as the comparison with a classification model was outside the scope of our study.

In the simulation study, model evaluation could be carried out using the true soil thickness data, but this will not be feasible in real-world applications. While Harrell's concordance index (*C*-index) is a common choice for model evaluation in survival analysis (Harrell et al., 1982),

it requires an estimate of the survival function or the CHF. Within the context of DSM, this method was for example used in Chen et al. (2019). Another method for evaluating models with censored data involves using an IPCW approach, as used in Graf et al. (1999). Although we explored this method in our study, its preference for IPC-RF due to a similar formulation in the model, providing an unfair advantage, led to notably superior results compared with other models. Consequently, we opted to exclude it from our study. Instead, we adopted two alternative and meaningful evaluation strategies in the DSM domain. The first involved evaluating only with non-censored data, while the second involved evaluation with data truncated to specific depths—100 and 150 cm. This decision was influenced by the potential interest of a map user in soil thickness down to a certain depth, beyond which the exact thickness may be less important.

The Maine real-world case study, similar to the first scenario in the simulation study where $\rho = 1$ and $\lambda = 120$, indicated that ARF and IPC-RF were comparable and outperformed HRF and SRF. This was evident from Table 2, specifically when the models were evaluated with all of the data in test set (using non-censored data and treating censored data as true), as well as when the models were evaluated with the truncated data. As noted before, this is not unexpected as limited information available about the true soil thickness is known. Therefore, a modelling approach like the one used in Malone and Searle (2020) could also be more appropriate, but such an approach could possibly be further improved by modelling the smaller depths with IPC-RF instead of using a regular random forest with imputations from a beta distribution (Kempen et al., 2015).

In the Zurich case study, both IPC-RF and SRF performed relatively well in terms of the RMSE metric, especially when the evaluation was conducted using truncated data. However, both survival models did overestimate the true soil thickness (more so in the case of SRF) when the data were truncated to 150 cm. In Malone and Searle (2020) the authors also noted overestimated results of SRF. It is noteworthy that conclusions for the Zurich case study are not as straightforward because several aspects are occurring simultaneously. It can be assumed that, in this case, censoring is informative (refer to Web Appendix A) and likely exhibits a spatial correlation structure with the observation depth. The same surveyors covered a pre-specified area (Jäggli et al., 1998) and were supervising the excavation of the profile pits being satisfied with the operator earlier or later. In addition, a relationship with gravel content of the underlying non-recorded soil horizon is to be expected. Moreover, censoring is not at a fixed depth as for the Maine case study and occurs at low soil thickness values already.

Further research is needed to also derive uncertainty maps for the models employed in the two case studies. For the random forest models, this can be accomplished using methods outlined in Hengl et al. (2018). Regarding SRF, one straightforward approach involves extracting soil thickness from the survival function, such as obtaining the 5% and 95% percentiles in case of a 90% prediction interval.

5 | CONCLUSION

Soil thickness data are often right-censored, indicating that the sampling depth is smaller than the true soil thickness. In this paper, we used an IPC weighted random forest model to address this issue, by assigning extra weight to non-censored data and zero weight to censored data, unless censoring occurred after a predefined depth. We compared the proposed model with a SRF and two other strategies for dealing with right-censored data (i.e. left-point imputation and using only non-censored data). The models were evaluated in a synthetic simulation study under various censored scenarios. The results of the simulation study showed that IPC-RF demonstrated superior performance when the proportion of censored data (that exceeded a certain depth) was equal to or less than 0.6. For larger proportions, left-point imputation (ARF) was superior. The models were also assessed in two case studies using metrics (ME, RMSE, CCC) that were computed on test data sets that were adjusted for censoring. In the Maine case study, IPC-RF and ARF outperformed SRF and HRF when the truncated evaluation approach was performed. In the Zurich case study, IPC-RF produced comparable results with the truncated evaluation approach as well as the approach with using only non-censored data. The simulation study and the case studies demonstrated that IPC-RF is a viable option for modelling right-censored soil thickness data.

AUTHOR CONTRIBUTIONS

Stephan van der Westhuizen: Conceptualization; investigation; funding acquisition; writing – original draft; methodology; validation; visualization; writing – review and editing; project administration; formal analysis; software. **Gerard B. M. Heuvelink:** Supervision; project administration; conceptualization; methodology; validation; writing – review and editing. **David P. Hofmeyr:** Conceptualization; methodology; validation; writing – review and editing. **Laura Poggio:** Conceptualization. **Madlene Nussbaum:** Data curation; writing – review and editing. **Colby Brungard:** Data curation; writing – review and editing.

ACKNOWLEDGEMENTS

This work is based on research supported in by the National Research Foundation (South Africa) (Grant number: 129856). We would like to thank the Canton of Zurich for sharing their soil survey data through the open Swiss national soil dataset. We would like to thank Nicholas Butler, Alaina Kresovic and Joshua Dera from the USDA-NRCS Dover-Foxcroft field office in Maine for digitising and sharing soil survey data.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Stephan van der Westhuizen https://orcid.org/0000-0001-9469-8427
Gerard B. M. Heuvelink https://orcid.org/0000-0003-0959-9358
Madlene Nussbaum https://orcid.org/0000-0002-6808-8956

REFERENCES

Amelung, W., Blume, H., Fleige, H., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., Stahr, K., & Wilke, B. (2018). *Scheffer/Schachtschabel Lehrbuch Der Bodenkunde*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-55871-3

Baltensweiler, A., Walthert, L., Hanewinkel, M., Zimmermann, S., & Nussbaum, M. (2021). Machine learning based soil maps for a wide range of soil properties for the forested area of Switzerland. *Geoderma Regional*, *27*, e00437. https://doi.org/10.1016/j.geodrs.2021.e00437

Bandyopadhyay, S., Wolfson, J., Vock, D., Vazquez-Benitez, G., Adomavicius, G., Elidrisi, M., Johnson, P., & O'Connor, P. (2014). Data mining for censored time-to-event data: A bayesian network model for predicting cardiovascular risk from electronic health record data. *Data Mining and Knowledge Discovery*, *29*, 1033–1069. https://doi.org/10.1007/s10618-014-0386-6

BGS-SSP. (2010). Klassifikation der Böden der Schweiz (KLABS): Bodenprofiluntersuchung, Klassifikationssystem, Definitionen Der Begriffe, Anwendungsbeispiele. Technical Report Bodenkundliche Gesellschaft der Schweiz.

Bonfatti, B. R., Hartemink, A. E., Vanwalleghem, T., Minasny, B., & Giasson, E. (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

Braekers, R., & Veraverbeke, N. (2005). Cox's regression model under partially informative censoring. *Communications in Statistics□Theory and Methods*, *34*, 1793–1811. https://doi.org/10.1081/STA-200066346

Brungard, C., Nauman, T., Duniway, M., Veblen, K., Nehring, K., White, D., Salley, S., & Anchang, J. (2021). Regional ensemble modeling reduces uncertainty for digital soil mapping. *Geoderma*, *397*, 114998. https://doi.org/10.1016/j.geoderma.2021.114998

Chaplot, V., Lorentz, S., Podwojewski, P., & Jewitt, G. (2010). Digital mapping of a-horizon thickness using the correlation between various soil properties and soil apparent electrical resistivity. *Geoderma*, *157*, 154–164. https://doi.org/10.1016/j.geoderma.2010.04.006

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

Chen, S., Mulder, V. L., Martin, M. P., Walter, C., Lacoste, M., Richer-de Forges, A. C., Saby, N. A. P., Loiseau, T., Hu, B., & Arrouays, D. (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

Fan, J., McConkey, B., Wang, H., & Janzen, H. (2016). Root distribution by depth for temperate agricultural crops. *Field Crops Research*, *189*, 68–74. https://doi.org/10.1016/j.fcr.2016.02.013

Florinsky, I. V., Eilers, R. G., Manning, G. R., & Fuller, L. G. (2002). Prediction of soil properties by digital terrain modelling. *Environmental Modelling and Software*, *17*, 295–311. https://doi.org/10.1016/S1364-8152(01)00067-6

Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, *18*, 2529–2545. https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2529::AID-SIM274>3.0.CO;2-5

Greiner, L., Keller, A., Grêt-Regamey, A., & Papritz, A. (2017). Soil function assessment: Review of methods for quantifying the contributions of soils to ecosystem services. *Land Use Policy*, *69*, 224–237. https://doi.org/10.1016/j.landusepol.2017.06.025

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *JAMA*, *247*, 2543–2546.

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

Hofmeyr, D. P. (2022). Fast kernel smoothing in R with applications to projection pursuit. *Journal of Statistical Software*, *101*, 1–33. https://doi.org/10.18637/jss.v101.i03

Huang, X., & Wolfe, R. A. (2002). A frailty model for informative censoring. *Biometrics*, *58*, 510–550. https://doi.org/10.1111/j.0006-341x.2002.00510.x

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, *2*, 841–860. https://arXiv.org/abs/0811.1645v1

Jäggli, F., Peyer, K., Pazeller, A., & Schwab, P. (1998). Grundlagenbericht Zur Bodenkartierung Des Kantons Zürich. Technical Report.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457–481.

Kempen, B., Brus, D. J., & de Vries, F. (2015). Operationalizing digital soil mapping for nationwide updating of the 1:50,000 soil map of The Netherlands. *Geoderma*, *241–242*, 313–329. https://doi.org/10.1016/j.geoderma.2014.11.030

Kleinbaum, D. G., Klein, M., Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text volume 3*. Springer.

Kuriakose, S. L., Devkota, S., Rossiter, D. G., & Jetten, V. G. (2009). Prediction of soil depth using environmental variables in an anthropogenic landscape, a case study in the western ghats of kerala, India. *Catena*, *79*, 27–38.

Lacoste, M., Mulder, V. L., Richer-De-Forges, A. C., Martin, M. P., & Arrouays, D. (2016). Evaluating large-extent spatial modeling approaches: A case study for soil depth for France. *Geoderma Regional*, *7*, 137–152.

Lawrence, I., & Lin, K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*, 255–268.

LeBlanc, M., & Crowley, J. (1993). Survival trees by goodness of split. *Journal of the American Statistical Association*, *88*, 457–467.

Leenaars, J. G. B., Claessens, L., Heuvelink, G. B. M., Hengl, T., Ruiperez González, M., van Bussel, L. G. J., Guilpart, N., Yang, H., & Cassman, K. G. (2018). Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-saharan africa. *Geoderma*, *324*, 18–36. https://doi.org/10.1016/j.geoderma.2018.02.046

Leung, K. M., Elashoff, R. M., & Afifi, A. A. (1997). Censoring issues in survival analysis. *Annual Review of Public Health*, *18*, 83–104.

Liu, F., Yang, F., Zhao, Y., Zhang, G., & Li, D. (2022). Predicting soil depth in a large and complex area using machine learning and environmental correlations. *Journal of Integrative Agriculture*, *21*, 2422–2434. https://doi.org/10.1016/S2095-3119(21)63692-4

Malone, B., & Searle, R. (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

McDonald, J. F., & Moffitt, R. A. (1980). The uses of tobit analysis. *The Review of Economics and Statistics*, *62*, 318–321.

Minasny, B., & McBratney, A. B. (1999). A rudimentary mechanistic model for soil production and landscape development. *Geoderma*, *90*, 3–21. https://doi.org/10.1016/S0016-7061(98)00115-3

Minasny, B., & McBratney, A. B. (2006). Mechanistic soil–landscape modelling as an approach to developing pedogenetic classifications. *Geoderma*, *133*, 138–149. https://doi.org/10.1016/j.geoderma.2006.03.042 Advances in landscape-scale soil research.

Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma*, *264*, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017 Soil mapping, classification, and modelling: history and future directions.

Moore, I. D., Gessler, P. E., Nielsen, G. A. E., & Peterson, G. A. (1993). Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal*, *57*, 443–452.

Mulder, V. L., Lacoste, M., Richer-de Forges, A. C., & Arrouays, D. (2016). Globalsoilmap France: High-resolution spatial modelling the soils of France up to two meter depth. *Science of the Total Environment*, *573*, 1352–1369. https://doi.org/10.1016/j.scitotenv.2016.07.066

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M. E., & Papritz, A. (2018). Evaluation of digital soil mapping approaches with large sets of environmental covariates. *The Soil*, *4*, 1–22. https://soil.copernicus.org/articles/4/1/2018/. https://doi.org/10.5194/soil-4-1-2018

Odeh, I. O. A., Chittleborough, D. J., & McBratney, A. B. (1991). Elucidation of soil-landform interrelationships by canonical ordination analysis. *Geoderma*, *49*, 1–32.

Odeh, I. O. A., McBratney, A. B., & Chittleborough, D. J. (1995). Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. *Geoderma*, *67*, 215–226.

Olson, K. R., & Al-Kaisi, M. M. (2015). The importance of soil sampling depth for accurate account of soil organic carbon sequestration, storage, retention and loss. *Catena*, *125*, 33–37. https://doi.org/10.1016/j.catena.2014.10.004

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Schoorl, J. M., Veldkamp, A., & Bouma, J. (2002). Modeling water and soil redistribution in a dynamic landscape context. *Soil Science Society of America Journal*, *66*, 1610–1619. https://doi.org/10.2136/sssaj2002.1610

Segal, M. R. (1988). Regression trees for censored data. *Biometrics*, *44*, 35–47.

Service center NABODAT. (2022). Swiss soil dataset□Documentation version 6 (April 2022). Technical Report. https://nabodat.ch/index.php/de/service/bodendatensatz

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., & Dai, Y. (2017). Mapping the global depth to bedrock for land surface modeling. *Journal of Advances in Modeling Earth Systems*, *9*, 65–88. https://doi.org/10.1002/2016MS000686

Siemer, B., Obmann, L., Hinrichs, U., Penndorf, O., Pohl, M., Schürer, S., Schulze, P., & Seiffert, S. (2014). Bodenbewertungsinstrument Sachsen. Technical Report Tech. Rep., Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie....

Suleymanov, A., Abakumov, E., Suleymanov, R., Gabbasova, I., & Komissarov, M. (2021). The soil nutrient digital mapping for precision agriculture cases in the trans-ural steppe zone of Russia using topographic attributes. *ISPRS International Journal of Geo-Information*, *10*, 243. https://doi.org/10.3390/ijgi10040243

Tesfa, T. K., Tarboton, D. G., Chandler, D. G., & McNamara, J. P. (2009). Modeling soil depth from topographic and land cover attributes. *Water Resources Research*, *45*, W10438.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society*, *26*, 24–36.

Turkson, A. J., Ayiah-Mensah, F., & Nimoh, V. (2021). Handling censoring and censored data in survival analysis: A standalone systematic literature review. *International Journal of Mathematics and Mathematical Sciences*, *2021*, 9307475.

U.S. Department of Agriculture, N. R. C. S. (2017). National soil survey hand-book, title 430-VI. https://www.nrcs.usda.gov/resources/guides-and-instructions/national-soil-survey-handbook

van der Westhuizen, S., Heuvelink, G. B. M., & Hofmeyr, D. P. (2023). Multivariate random forest for digital soil mapping. *Geoderma*, *431*, 116365. https://doi.org/10.1016/j.geoderma.2023.116365

van der Westhuizen, S., Heuvelink, G. B. M., Hofmeyr, D. P., & Poggio, L. (2022). Measurement error-filtered machine learning in digital soil mapping. *Spatial Statistics*, *47*, 100572. https://doi.org/10.1016/j.spasta.2021.100572

Vaysse, K., & Lagacherie, P. (2015). Evaluating digital soil mapping approaches for mapping globalsoilmap soil properties from

legacy data in languedocroussillon (France). *Geoderma Regional*, *4*, 20–30. https://doi.org/10.1016/j.geodrs.2014.11.003

Vock, D. M., Wolfson, J., Bandyopadhyay, S., Adomavicius, G., Johnson, P. E., Vazquez-Benitez, G., & O'Connor, P. J. (2016). Adapting machine learning techniques to censored time-to-event health record data: A general-purpose approach using inverse probability of censoring weighting. *Journal of Biomedical Informatics*, *61*, 119–131. https://doi.org/10.1016/j.jbi.2016.03.009

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

Willems, S., Schat, A., van Noorden, M., & Fiocco, M. (2018). Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator. *Statistical Methods in Medical Research*, *27*, 323–335. https://doi.org/10.1177/0962280216628900

Zhang, W., Hu, G., Sheng, J., Weindorf, D. C., Wu, H., Xuan, J., Yan, A., & Gu, Z. (2018). Estimating effective soil depth at regional scales: Legacy maps versus environmental covariates. *Journal of Plant Nutrition and Soil Science*, *181*, 167–176. https://doi.org/10.1002/jpln.201700081

Zheng, M., & Klein, J. P. (1994). A self-consistent estimator of marginal survival functions based on dependent competing risk data and an assumed copula. *Communications in Statistics□Theory and Methods*, *23*, 2299–2311. https://doi.org/10.1080/03610929408831387

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: van der Westhuizen, S., Heuvelink, G. B. M., Hofmeyr, D. P., Poggio, L., Nussbaum, M., & Brungard, C. (2024). Mapping soil thickness by accounting for right-censored data with survival probabilities and machine learning. *European Journal of Soil Science*, *75*(5), e13589. https://doi.org/10.1111/ejss.13589