

Support vector regression for high-resolution beach surface moisture estimation from terrestrial LiDAR intensity data

Junling Jin^{a,*}, Jeffrey Verbeurgt^a, Lars De Sloover^a, Cornelis Stal^{a,b}, Greet Deruyter^c, Anne-Lise Montreuil^d, Sander Vos^e, Philippe De Maeyer^a, Alain De Wulf^a

^a Department of Geography, Ghent University, Krijgslaan 281 S8, 9000 Ghent, Belgium

^b Department of Real-estate and Applied Geomatics, University College Ghent, Valentin Vaerwyckweg 1, 9000, Ghent, Belgium

^c Department of Civil Engineering, Ghent University, Technologiepark 904, 9052 Ghent, Belgium

^d Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Elsene, Belgium

^e Department of Hydraulic Engineering, Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Keywords:

Support vector machine regression
Machine learning
Terrestrial LiDAR
Surface moisture
Coastal monitoring

ABSTRACT

Beach Surface Moisture (BSM) is a key attribute in the coastal investigations of land-atmospheric water and energy fluxes, groundwater resource budgets and coastal beach/dune development. In this study, an attempt has been made for the first time to estimate BSM from terrestrial LiDAR intensity data based on the Support Vector Regression (SVR). A long-range static terrestrial LiDAR (Riegl VZ-2000) was adopted to collect point cloud data of high spatiotemporal resolution on the Ostend-Mariakerke beach, Belgium. Based on the field moisture samples, SVR models were developed to retrieve BSM, using the backscattered intensity, scanning ranges and incidence angles as input features. The impacts of the training samples' size and density on the predictive accuracy and generalization capability of the SVR models were fully investigated based on simulated BSM-intensity samples. Additionally, we compared the performance of the SVR models for BSM estimation with the traditional Stepwise Regression (SR) method and the Artificial Neural Network (ANN). Results show that SVR could accurately retrieve the BSM from the backscattered intensity with high reproducibility (average test RMSE of $0.71\% \pm 0.02\%$ and R^2 of $0.98\% \pm 0.002\%$). The Radial Basis Function (RBF) was the most suitable kernel for SVR model development in this study. The impacts of scanning geometry on the intensity could also be accurately corrected in the process of estimating BSM by the SVR models. However, compared to the SR method, the predictive accuracy and generalization performance of SVR models were significantly dependent on the training samples' coverage, size and distribution, suggesting the need for the training samples of uniform distribution and representativeness. The minimum size of training samples required for SVR model development was 54. Under this condition, SVR performed similarly to ANN with a test RMSE of 1.06%, but SVR still performed acceptably (with an RMSE of 1.83%) even using extremely few training samples (only 16 field samples of uniform distribution), far better than the ANN (with an RMSE of 4.02%).

1. Introduction

Beach Surface Moisture (BSM) plays an important role in the coastal studies which investigate the land-atmospheric water and energy fluxes (McLachlan, 1989), biogeochemical cycling (Legates et al., 2011), groundwater resource budgets (Chen and Hu, 2004; Horn, 2002; Smit et al., 2019) and coastal beach/dune development (Davidson-Arnott et al., 2008; Schmutz and Namikas, 2018). Particularly, it is often cited as a key supply limiting factor for the aeolian sediment transport on the

beach (Anthony et al., 2009; Cornelis and Gabriels, 2003; Darke and Neuman, 2008; Edwards and Namikas, 2015). However, BSM is highly variable in both time and space. Because it is not only affected by precipitation, evaporation, groundwater table, sediment size and porosity (similarly with the common surface soil moisture) but also by tidal and wave action (Atherton et al., 2001; Jin et al., 2020; Nield et al., 2011; Yang and Davidson-Arnott, 2005). This thus requires the BSM to be measured at a sufficient temporal (e.g., one-hour interval) and spatial (e.g., one meter) resolution in order to understand and quantify the aeolian

* Corresponding author.

E-mail address: Junling.Jin@ugent.be (J. Jin).

<https://doi.org/10.1016/j.jag.2021.102458>

Received 15 April 2021; Received in revised form 9 July 2021; Accepted 19 July 2021

Available online 29 July 2021

0303-2434/© 2021 The Author(s).

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

transport process in detail.

Traditionally, BSM can be acquired through in situ measurements, mainly including the sample gravimetric method (Davidson-Arnott, 2010; Nield et al., 2011; Wiggs et al., 2004), soil moisture probes (Bauer et al., 2009; Namikas et al., 2010; Schmutz and Namikas, 2011), handheld radiometer and spectroradiometer (Edwards et al., 2013a; Edwards et al., 2013b; Zhu et al., 2010), etc. However, these spot-based methods have limited meaning in measuring a substantial beach section due to the strong temporal and spatial variability of the BSM. The optical brightness method (with a ground-based video system) can estimate the large-area BSM at a high scanning frequency in a non-contact way (Darke et al., 2009; McKenna Neuman and Langston, 2006; Montreuil et al., 2018; Yang et al., 2019). However, it is subject to the weather and environment (only working in the daytime), especially the sunlight conditions, and thus hard to achieve a high measurement accuracy. Additionally, the common satellite remote sensing techniques for measuring the soil surface moisture on a large scale could not meet the requirements of high spatiotemporal resolutions to estimate the BSM (Ali et al., 2015; Anderson and Croft, 2009).

Recently, terrestrial laser scanning (TLS) has shown great promise for an accurate BSM estimation because of the advantages of active scanning (deployable at daytime and nighttime), high spatial and temporal resolution (through continuous measurements) (Jin et al., 2020; Jin et al., 2021; Nield et al., 2014; Ruessink et al., 2014; Smit et al., 2018; Tan et al., 2020), as well as a large spatial scale (long-range LiDAR systems can typically measure a large beach section, such as 400×800 m, with a single scan). According to the radar range equation, the backscattered intensity is closely related to the target surface reflectance, and the latter could theoretically be used to derive the BSM when other surface properties remain constant (Nolet et al., 2014). Additionally, the backscattered intensity of a LiDAR system operating in the shortwave infrared spectrum is quite sensitive to target surface moisture variation, which makes it an ideal tool to monitor BSM changes (Yang et al., 2019).

However, it is difficult to construct a specific model for the complex (nonlinear) and unknown relations between the LiDAR intensity and BSM, which vary with different LiDAR systems and beach types. Normally, the surface moisture is estimated by means of stepwise regression (SR) methods (Jin et al., 2020; Smit et al., 2018; Tan et al., 2020). Firstly, the original backscattered intensity is calibrated using data-driven or model-driven approaches based on the radar range equation (Höfle and Pfeifer, 2007; Kaasalainen et al., 2011), and then the relation between the calibrated intensity and the BSM is fitted based on a priori model (e.g., the exponential models) (Nolet et al., 2014; Philpot, 2010; Zhu et al., 2010). The procedure of parametric modeling is rather complicated and may cause an accumulation of errors. Besides, the radar range equation, as the prior knowledge of intensity calibration, may not be suitable for some specific terrestrial LiDAR systems in which the built-in software of the scanner could automatically reduce the backscattered intensity acquired at close ranges (Kaasalainen et al., 2011; Kaasalainen et al., 2009b; Tan and Cheng, 2015).

Another issue is the small number of available BSM samples for the stepwise regression modeling, given that the collection of field moisture samples is laborious, costly and time-consuming. In addition, for long-range LiDAR systems, it is difficult to obtain intensity correction data through indoor experiments (due to the length limitation of the laboratory) like the short- and middle-range LiDAR systems (Fang et al., 2015; Jin et al., 2020; Kaasalainen et al., 2009a; Tan and Cheng, 2015). A possible approach to deal with these limitations involves the use of non-parametric modeling to interpolate BSM (Ballabio, 2009), such as machine learning algorithms.

Machine learning algorithms provide an alternative method to solve the inversion problems of geophysical parameters with only little prior knowledge of the theoretical model, such as the Artificial Neural Networks (ANN), Decision Tree (DT) and Support Vector Machines (SVM). A detailed introduction and comparison about these common machine

learning algorithms for inversion problems can be found in the review literature (Carter and Liang, 2019; Lary et al., 2016; Maxwell et al., 2018; Mountrakis et al., 2011; Padarian et al., 2020). Recently, the Support Vector Regression (SVR), as a version of the SVM for regression, has been widely and successfully used in retrieving geophysical and biophysical parameters from remote sensing datasets (Ballabio, 2009; de Souza et al., 2019; Deiss et al., 2020; Deka, 2014; Hoa et al., 2019; Xiao et al., 2018). In these applications, the estimation performance of SVR is typically comparable to (or better than) the one using the well-known ANN algorithm, due to its advantages of structural risk minimization based on a strict statistical theory. Particularly, this method is praised for its ability to deal with small training datasets (de Souza et al., 2019). Considering that it is hard to collect a large number of moisture samples on the study beach due to the time and labor factor, the SVR seems to be a suitable method for developing the BSM estimation model. The high efficiency of SVR to retrieve large scale soil moisture from the remote sensing data (e.g., multi/hyperspectral images and synthetic aperture radar data) has been demonstrated (Ahmad et al., 2010; Ezzahar et al., 2020; Holtgrave et al., 2018; Pasolli et al., 2015; Pasolli et al., 2011). However, to our knowledge, no studies have tested SVR (or other machine learning algorithms) yet for estimating the soil moisture or beach surface moisture from the LiDAR intensity data.

This study sets out to investigate the feasibility of retrieving BSM from the LiDAR intensity data using the SVR algorithm, without correcting the impacts of scanning geometry on backscattered intensity in advance. Point cloud data of a high spatiotemporal resolution were collected using a long-range static terrestrial LiDAR (Riegl VZ-2000) (Vos et al., 2017), which was permanently deployed on top of a building close to the Ostend-Mariakerke beach, Belgium. By referring to the relevant studies on the application of SVR in the field of remote sensing (de Souza et al., 2019; Xiao et al., 2018), this study conducted investigations to answer the following: (i) how well (accuracy, reproducibility and generalization capability) SVR performs in estimating BSM from the LiDAR intensity data; (ii) whether SVR models can eliminate the impacts of the range and incidence angle on the backscattered intensity in the process of estimating BSM, considering the high dependence of original TLS intensity data on the scanning geometry; (iii) what are the minimum number and density of training samples required to develop an SVR model, investigating SVR's potential to deal with small sizes of BSM training samples, given that the collection of field BSM samples would be relative laborious and time-consuming; and (iv) how the SVR algorithm performs in comparison with the traditional SR method already tested in previous literature and the well-known ANN algorithm. Here, the ANN was used mainly for comparative analysis of the performance of SVR and ANN under conditions of limited training samples.

2. Materials and methods

2.1. Study site

The study site is a relatively flat sandy beach located at Ostend-Mariakerke, Belgium (51.213° N, 2.872° E), with a slope of approximately 2° (Fig. 1). It is a typical dissipative beach with a width exceeding 200 m at low tide. The neap tide measures about 3.5 m while the spring tide is up to 5 m (Deronde et al., 2008). The mean annual precipitation of the site is 748 mm and the mean annual temperature 10.6° C (Montreuil et al., 2018). The common wind speed ranges from 3 to 8 m/s, mainly coming from the southwest. In addition, the particle size of the sediments (quartz sand) slightly raises from $291 \mu\text{m}$ (D50) at the backshore to $337 \mu\text{m}$ (D50) at the intertidal zone, predominantly consisting of fine and medium sand (Montreuil et al., 2018). There is neither vegetation nor buildings on the beach.

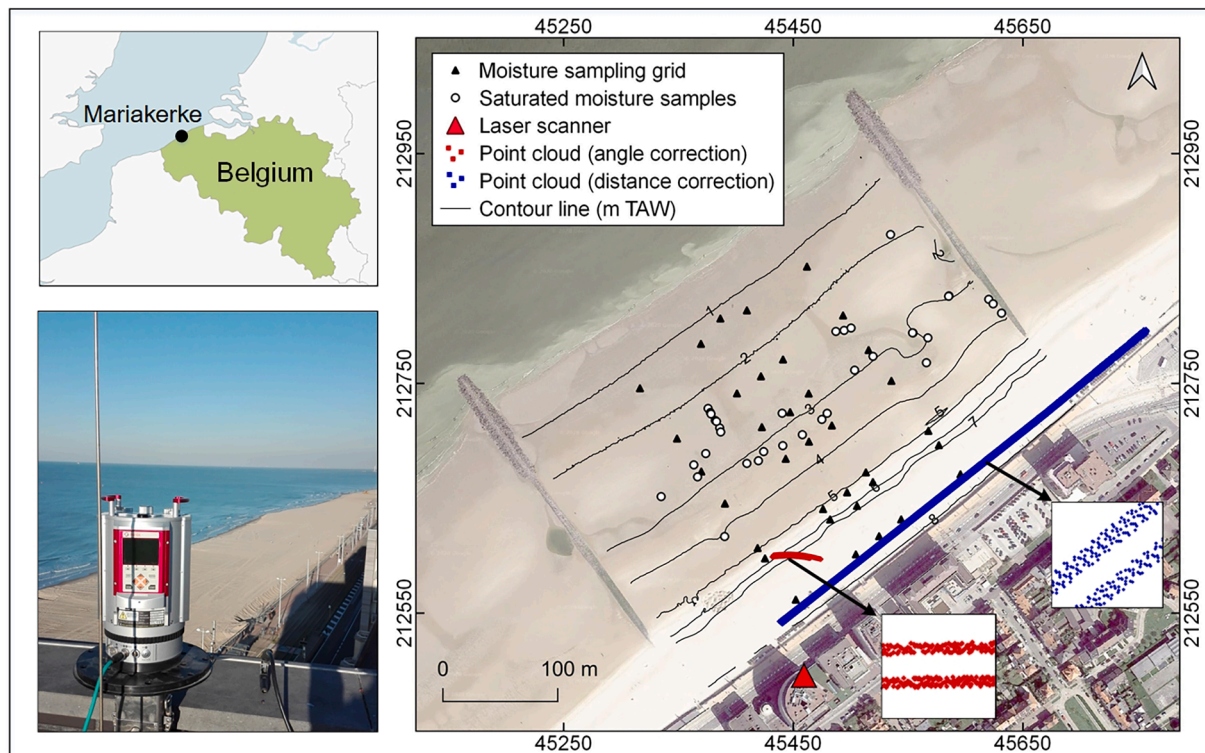


Fig. 1. Location of the study area at the Ostend-Mariakerke beach and the field deployment. The Riegl VZ-2000 scanner is deployed on top of a 42 m high building close to the target beach (CRS: EPSG 31 370). The red and blue points in the magnified images (the bottom right corner) represent the TLS point clouds used for correcting the impacts of scanning geometry on the intensity data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Data collection and pre-processing

2.2.1. LiDAR data

The measurements were carried out using a shortwave near-infrared laser scanner Riegl VZ-2000 with a laser wavelength of 1,550 nm. The scanner's field of view was 360° (horizontal) \times 100° (vertical). The actual effective measurement distance is about 500 m under the circumstances of this study, with a beam divergence of 0.30 mrad. It is worth mentioning that a custom-made shield was used to protect the scanner from environmental impacts (e.g., wind, sand and rain). From November 2017 to December 2018, the scanner was permanently deployed on top of a 42 m high building close to the target beach (Fig. 1, (De Sloover et al., 2019)). The target beach was measured repeatedly at one-hour intervals to detect the beach topography and surface moisture changes. Each scan duration took about 4 min. A total of nine scans were used in this study, consisting of eight scans respectively associated with the BSM sample collections of eight times during April 18, 19 and 21–26, 2018 (BSM samples were collected once a day, at low tide), and one scan measured at 10:00 on April 17, 2018, used to produce the BSM map of the entire study area.

The scanner recorded both the 3D coordinates (relative to the scanner center) and the backscattered intensity of each reflected point. The raw coordinates of the point clouds were transformed into the Belgian national coordinate system employing twelve reflectors deployed on the beach and positioned through an RTK-GPS (Vos et al., 2017). As part of the Belgian CREST (Climate Resilient Coast) Project, this study adopted the Belgian Lambert 72 projected coordinate system (EPSG: 31,370) (VLIZ, 2020; Yang et al., 2020). The orthometric height was taken towards the TAW (Tweede Algemene Waterpassing) reference level (Ostend Height, EPSG: 5,110), which is an equipotential gravity surface of about 2.3 m under the conventional geoid EGM96. It is worth mentioning that the density of the point clouds gradually reduced with the increased scanning distance from $\sim 1,400$ points/m² at 60 m to ~ 3

points/m² at 440 m, with the horizontal and vertical angular sampling steps of 0.03° .

2.2.2. Surface moisture samples

The surface moisture samples were collected at an established sampling grid consisting of 5×7 points (Fig. 1), which extended from the relatively dry upper-beach to the dissipative intertidal area (Montreuil et al., 2018; Schmutz and Namikas, 2018). The location of each sampling point was obtained using an RTK-GPS device. A total of 114 moisture samples were taken after the TLS scan during April 18–26, 2018. The sample size was approximately 10 cm \times 10 cm with a thickness of about 0.5 cm. The sample moisture contents were measured in the laboratory using the gravimetric method (calculating the ratio between the water weight in the sample and the sample's total weight) (Davidson-Arnott, 2010; Nield and Wiggs, 2011).

The average intensity, range and incidence angle of each field moisture sample was extracted from the corresponding TLS point clouds in a 1×1 m grid cell, in which the incidence angle was derived based on the plane fitting using the nearby points (to obtain the corresponding normal vector of the fitted plane) (Jin et al., 2021). In this study, only the moisture samples with a scanning range < 250 m (a total of 55 effective samples) were adopted in the next moisture retrieval because the point cloud density at long ranges was too low (especially in the high moisture region) to obtain an accurate incidence angle (Jin et al., 2021).

It should be said that the moisture distribution of these field samples was relatively concentrated, mainly at the ranges of 0–1% and 15–20%, due to the collection from the established sampling grid at low tide each day (Jin et al., 2021). Only a few samples with saturated moisture were collected during the TLS measurements (even no high moisture samples at the ranges > 350 m). To address this issue, we manually extracted additional 35 samples with saturated moisture (the white circles in Fig. 1) from the TLS point clouds acquired at 10:00 on April 17, 2018. The samples were extracted based on the corresponding high resolution

(centimeter-level) orthophotos simultaneously acquired on the beach by means of a drone (DJI Phantom 4 pro). The orthophotos have been transformed into the same coordinate reference system as the TLS point clouds, by employing the black-white square targets placed on the beach surface and positioned through an RTK-GPS.

To ensure that the extracted samples were saturated, the samples should simultaneously satisfy the following three demands: (i) the sampling points should be located at the transition zone from the flat beach to the trough area (or shallow waters); (ii) the very thin water film is just starting to appear at the surface of the sampling area; (iii) there are at least three reflected points in each sampling area of 1×1 m (normally no reflected points from the water surface at long ranges due to the large incidence angles). Based on the laboratory test results (Jin et al., 2020), the gravimetric moisture contents of the 35 saturated samples were set to 26%. The results were also verified by the field moisture samples collected on the Ostend-Mariakerke and Koksijde-Groenendijk beach, Belgium (Jin et al., 2020; Jin et al., 2021).

The average intensity and range of each saturated sample were extracted from the corresponding TLS point clouds (1×1 m). The ranges of these samples changed from 139 m to 403 m. Because the TLS point clouds were too sparse at long ranges, the incidence angles of these saturated samples were calculated employing the point clouds derived from the corresponding high-resolution orthophotos. The image-derived point clouds were in the same coordinate reference system as the TLS point clouds and more details about the image-derived point clouds can be found in Jin et al. (2021). It is worth mentioning that these samples were only utilized as a reference to investigate the impact of sample coverage.

2.2.3. Intensity correction data

To investigate whether SVR models can eliminate the scanning geometry impacts on the backscattered intensity in the process of estimating BSM, the SVR training data should cover sufficiently large ranges of the scanning range and the incidence angle. Traditionally, these data can be obtained from the indoor correction experiments in which the target samples were measured at different ranges and incidence angles by the TLS (Fang et al., 2015; Jin et al., 2020; Kaasalainen et al., 2009a; Tan and Cheng, 2015). However, it is impractical to remove the laser scanner from the fixed deployment position (in this study). Moreover, it is difficult to find a laboratory with a length exceeding 500 m for indoor correction experiments.

In the present study, the correction data were extracted from the TLS point clouds acquired on the upper-beach, which could be considered as the natural homogenous target with similar surface properties (Jin et al., 2021). The surface moisture contents on the upper-beach were $0.05\% \pm 0.01\%$ at 10:00 on April 18, 2018, obtained by calculating the average of the 10 field moisture samples collected on the upper-beach. As shown in Fig. 1 (the red point clouds), we selected the point clouds within two very narrow arc-like strips (about 30 m length and 40 cm width) for the incidence angle correction. The incidence angle ranges of the two arc-like strips were $45\text{--}87^\circ$ (one sample per degree, a total of 41 samples extracted for SVR training) and $57\text{--}84^\circ$ (one sample per degree, a total of 28 samples for SVR testing), respectively. The former covers the majority of the possible incidence angles in this study (about 54° at 61 m and 87° at 440 m, regardless of the beach terrain relief). More details about the pre-processing of correction data are to be found in the previous literature of Jin et al. (2021).

Similarly, we selected the point clouds within two 1 m-width long strips (the blue point clouds in Fig. 1) for the range correction. The average intensity, range and incidence angle of the point clouds were calculated at intervals of 1 m. The two strips (for SVR training and testing respectively) covered all the required measurement ranges of 61–440 m, and a total of 39 samples were acquired for each strip (one sample per 10 m).

2.3. Support vector regression

Support vector regression is a non-parametric regression technique developed based on the hypothesis of structural risk minimization (Cortes and Vapnik, 1995). It can minimize the empirical risk and achieve a good generalization for the unseen data, even using few training data. Numerous studies have proven the robustness and high efficiency of the SVR algorithms for retrieving large-scale soil moisture from remote sensing data (Ahmad et al., 2010; Ezzahar et al., 2020; Holtgrave et al., 2018; Pasolli et al., 2015; Pasolli et al., 2011). This study focuses on retrieving the BSM from the high-resolution LiDAR intensity data using the common ε -insensitive SVR (ε -SVR). The detailed mathematical expressions of SVR can be found in Cristianini and Shawe-Taylor (2000) and Smola and Schölkopf (2004). Here, the basic concept of SVR is described below.

With the ε -SVR, we defined a set of training data $\{(x_1, y_1), \dots, (x_n, y_n)\}$, where x represents the feature vector ($x \in R^k$) and y denotes the target output (i.e., the BSM) ($y \in R$). k is the feature dimension and n the sample number. The SVR aims to find an appropriate function (hyperplane) $f(x)$ to describe the input–output mapping. The function $f(x)$ can be linearly expressed as $f(x) = w^T \phi(x) + b$, where w represents the weighting vector and b the bias. ϕ denotes a mapping function for projecting the original feature space to a higher-dimensional space. The standard form of ε -SVR is expressed as follows (Cristianini and Shawe-Taylor, 2000):

$$\begin{aligned} & \text{Min} : \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & \text{subject to} \begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i \\ f(x_i) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, n. \end{cases} \end{aligned} \quad (1)$$

Here, C is the penalty factor that tunes the trade-off between the model generalization and the tolerance for errors (de Souza et al., 2019; Pasolli et al., 2011). When using a larger C value, the generalization of the SVR models will decrease and it might cause overfitting ($C > 0$). ξ_i and ξ_i^* illustrate the slack variables. ε demonstrates the width of the insensitivity zone ($\varepsilon > 0$). A larger ε value will decrease the amount of the selected support vector and might cause training errors (Pan et al., 2015). For the non-linear cases, the SVR algorithm utilizes the kernel functions $K(x_i, x_j)$ to transform or linearize the input data, in which $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. In this study, we tested four common kernel functions as follows (de Souza et al., 2019):

$$K(x_i, x_j) \begin{cases} x_i \cdot x_j & \text{Linear} \\ (-\gamma x_i \cdot x_j + r)^d & \text{Polynomial} \\ \exp\left(-\gamma |x_i - x_j|^2\right) & \text{Radial Basis Function (RBF)} \\ \tanh(-\gamma x_i \cdot x_j + r) & \text{Sigmoid} \end{cases} \quad (2)$$

The polynomial degree d and the parameters γ and r can be tuned through the cross-validation method (Sunder et al., 2020). Fig. 2 describes the overall workflow to estimate the BSM from LiDAR intensity data with the SVR algorithm.

2.4. SVR training and evaluation

In this study, three input features/variables were selected to develop SVR models for estimating BSM, i.e., the backscattered intensity, scanning ranges and incidence angles. The selection of input features was done by referring to the available scientific literature, in which LiDAR intensity data were used to estimate BSM through the SR method (Jin et al., 2021; Nield et al., 2014; Philpot, 2010; Smit et al., 2018; Tan et al., 2020). Among the three input features, the backscattered intensity has a

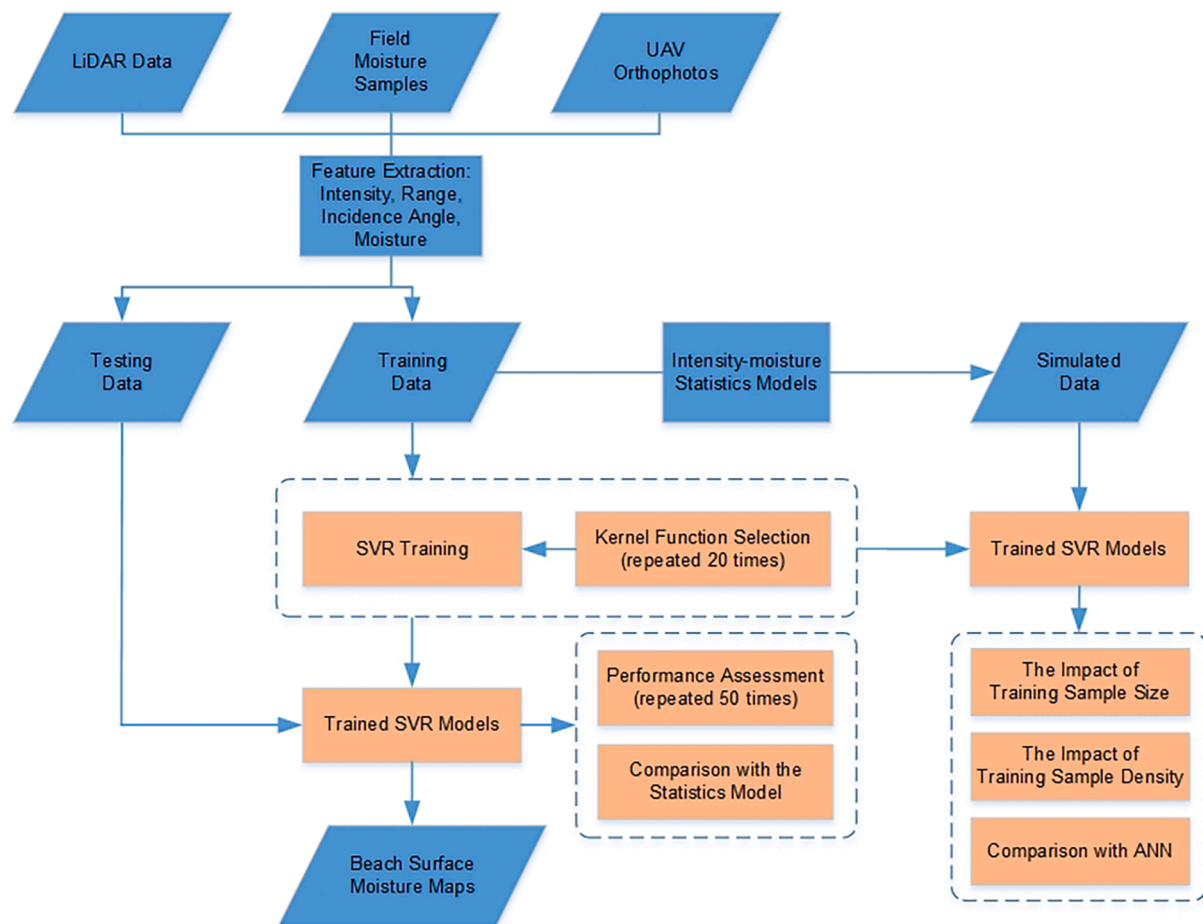


Fig. 2. Workflow for the estimation of beach surface moisture from the LiDAR intensity data using the SVR algorithm.

close relationship with beach surface reflectance based on the LiDAR equation (Kaasalainen et al., 2011; Kaasalainen et al., 2009a), and the reflectance can be used to derive the BSM when other beach surface properties (e.g., mineral composition, sand grain size and surface roughness) remain constant (Nield et al., 2014; Nolet et al., 2014; Philpot, 2010). The scanning ranges and incidence angles were adopted as the SVR input features to correct the scanning geometry impacts on the intensity in the process of estimating BSM. It is worth mentioning that this study used the incidence angle's cosine value as the SVR input feature rather than the incidence angle in radians. This is because the backscattered intensity is proportional to the incidence angle's cosine value based on the LiDAR equation (Fang et al., 2015; Xu et al., 2017). By using the cosine value, the complexity of the developed SVR models might be reduced to some extent. Besides, all input features were normalized to the 0–1 range to eliminate the effect of various scales on these features (Sunder et al., 2020).

Considering that SVR models' performance significantly depends on the selection of an appropriate kernel function, parameters tuning and sufficient and representative samples (Ballabio, 2009; de Souza et al., 2019; Deiss et al., 2020; Pasoli et al., 2015), this study tested four common kernel functions, i.e., linear, polynomial, Radial Basis Function (RBF) and sigmoid kernel, using the same training and test data. The kernel function with the best estimation performance on test data was selected for further SVR evaluation and application. In this study, the SVR models' performance was evaluated using two statistical criteria: RMSE (root mean squared error) and R^2 (the determination coefficient). The grid search method and 10-fold cross-validation were adopted to obtain the optimal SVR parameters, in which the parameter combination with the best cross-validation RMSE was chosen (Hsu et al., 2003;

Sunder et al., 2020). The trials for each kernel function were repeated 20 times by comprehensively considering the time cost and the robustness of training results.

The training data comprised 124 field samples, i.e., 44 samples collected from the sampling grid (80% of 55 effective samples) and 80 upper-beach samples for the intensity correction (41 for incidence angle correction and 39 for range correction). The test data consisted of 78 field samples, i.e., 11 samples collected from the sampling grid (20% of 55 effective samples) that were obtained at the different dates with those used in the training phase, and 67 upper-beach samples (28 for incidence angle and 39 for range). The upper-beach samples used for SVR training and testing were extracted from the different point clouds strips. An overview of the BSM estimation experiments, corresponding samples and related figures and tables in this study is given in Table 1.

Based on the selected kernel function, we retrained the SVR models repeatedly 50 times to further evaluate the reproducibility of the SVR algorithm. The prediction accuracy (RMSE) of the SVR model was obtained based on the test samples (i.e., 78 field samples). The intensity correction results were assessed by calculating the standard deviations of the predicted moisture contents respectively based on the incidence angle and range correction test samples (i.e., 28 and 39 upper-beach samples). The SVR model with the lowest test RMSE of 50 trials was selected to produce the beach surface moisture map of the entire study area (based on the scan acquired at 10:00 on April 17, 2018), in which each TLS point's incidence angle was calculated employing the high-density point clouds derived from the corresponding orthophotos, solving the low-density problem of the TLS point clouds at long ranges. A detailed introduction can be found in Jin et al. (2021).

Besides, the predicted moisture contents of the SVR model were

Table 1
Overview of the BSM estimation experiments, corresponding samples and related figures and tables.

Experiment		Training data	Test data	Fig./Tab.
Type	Description			
SVR	Kernel selection	44 BSM samples collected from the sampling grid; 80 upper-beach samples for intensity correction	11 BSM samples collected from the sampling grid; 67 upper-beach samples for intensity correction	Table 3
SVR/ SR	Method comparison	The same as above	The same as above	Figs. 3-5
SVR	The impact of sample coverage	44 BSM samples collected from the sampling grid; 80 upper-beach samples; 35 samples with a 26% moisture	11 BSM samples collected from the sampling grid; 67 upper-beach samples	Figs. 6-7
SVR	The impact of sample size	The simulated samples with the size of 20, 30, 50, 100, 200, 300, 400, 500, 700 or 1,000 (randomly subsampled from the 22,308 simulated training samples)	11 BSM samples collected from the sampling grid; 67 upper-beach samples; 2,520 simulated test samples; the field and simulated samples for evaluating intensity correction (Section 2.6)	Fig. 8
SVR	The impact of range interval	The simulated samples with the range interval of 10, 20, 40, 60, 80, 100, 130, 190 or 380 m	The same as above	Fig. 9a-c
SVR	The impact of angle interval	The simulated samples with the incidence angle interval of 2°, 4°, 6°, 8°, 10°, 12°, 14°, 22° or 42°	The same as above	Fig. 9d-f
SVR	The impact of moisture interval	The simulated samples with the moisture interval of 1%, 2%, 3%, 4%, 5%, 8%, 13% or 25%	The same as above	Fig. 9g-i
SVR/ ANN	Method comparison	54 simulated samples of uniform distribution	11 BSM samples collected from the sampling grid; 67 upper-beach samples	Figs. 10-11
SVR/ ANN	Method comparison	200 simulated samples of random selection	The same as above	–
SVR/ ANN	Method comparison	16 field samples of uniform distribution	The same as above	Figs. 12-13

compared with those estimated by the traditional SR method under the same training and test data. A moisture difference map of the two methods was produced (using a raster map of 1×1 m resolution) to further evaluate the SVR model's generalization ability and intensity correction performance. A detailed introduction on the SR method for retrieving BSM can be found in previous literature (Jin et al., 2020; Jin et al., 2021; Smit et al., 2018; Tan et al., 2020), and will be demonstrated briefly in Section 2.5. Also, we tried to validate the impact of the training samples' distribution (or coverage) on the generalization ability of the developed SVR models, by adding 35 samples of 26% moisture to the original training dataset and retraining (50 times) and testing the SVR models (see Table 1).

In this study, the grid search ranges of the SVR parameters were determined based on previous studies (Chang and Lin, 2011; Hsu et al., 2003) and the sample information as follows: $[2^{-12}, 2^{-2}]$ for the insensitive tube; $[2^{-10}, 2^{10}]$ for the parameter C ; $[3, 8]$ for the degree of the polynomial kernel function d ; $[2^{-10}, 1]$ for the coefficient r in the polynomial and sigmoid kernel functions. To reduce the computational time, we first conducted a preliminary grid search using the relatively large parameter interval. Then we obtained the final ranges of grid search for each parameter based on the cross-validation RMSE. The final grid search was carried out using a relatively small parameter interval and range. All SVR experiments were carried out utilizing the LIBSVM open-source library in Matlab (Chang and Lin, 2011).

2.5. The statistical model of BSM-intensity

For comparison purposes, the traditional statistical method (i.e., SR) was also adopted to estimate the BSM from the TLS backscattered intensity in this study (Jin et al., 2020; Jin et al., 2021; Smit et al., 2018; Tan et al., 2020). Its basic principles are that the original backscattered intensity was corrected firstly and then the relation between the BSM and the corrected intensity was fitted using least-squares regression. Previous studies had shown that the SR method can effectively eliminate the influence of the scanning geometry on the intensity data. Besides, compared to the machine learning approaches, there is no generalization problem for the statistical methods. In this study, the moisture contents predicted by the SR method were utilized as a comparative reference to evaluate the generalization performance and intensity correction results of SVR models, based on the moisture difference map of the two methods. According to the study of Jin et al. (2021), the original backscattered intensity I can be expressed as:

$$I = \delta e^{c_{BSM}} \cdot \sum_{i=0}^{N_1} [\beta_i (\cos\theta)^i] \cdot \sum_{i=0}^{N_2} (\gamma_i R^i) \quad (3)$$

Here, $\delta e^{c_{BSM}}$, $\sum_{i=0}^{N_1} [\beta_i (\cos\theta)^i]$ and $\sum_{i=0}^{N_2} (\gamma_i R^i)$ represent three independent functions which respectively illustrate the effect of the beach surface moisture BSM, the cosine incidence angle $\cos\theta$ and the range R on the backscattered intensity. δ , β_i and γ_i are the function coefficients. N_1 and N_2 denote the polynomial order. The three functions' parameters can be estimated stepwise by the least-square regression based on the intensity correction data and the moisture samples. It is worth mentioning that the procedure adopted the same data as those used for training SVR models (Table 1), and this allows us to compare the two methods straightforwardly. The detailed procedures for estimating the parameters of Eq. (3) can be found in Jin et al. (2021). After obtaining these parameters, the BSM can be derived using the reformulation of Eq. (3) as follows:

$$BSM = \frac{1}{c} \ln \left(\frac{I}{\delta \left(\sum_{i=0}^{N_1} [\beta_i (\cos\theta)^i] \right) \cdot \sum_{i=0}^{N_2} (\gamma_i R^i)} \right) \quad (4)$$

2.6. Investigation of the impact of the training sample size and density

Normally, SVR requires a training dataset consisting of sufficiently large and representative samples to achieve a good generalization (de Souza et al., 2019). However, in the actual field experiments, it is difficult to collect a large number of field moisture samples on the beach due to the time and labor factor. It is thus necessary to investigate what the required minimum sample size is to estimate the BSM using the SVR algorithm, and how the size and density (distribution) of the training samples affect the moisture retrieval accuracy and generalization ability of the SVR models. For this purpose, we utilized simulated datasets to train and test the SVR models (Table 1). Here, the simulated intensity data were generated based on Eq. (3) introduced in Section 2.5, which had been strictly derived and verified in previous literature (Jin et al., 2020; Jin et al., 2021; Tan et al., 2020). During generating the simulated intensity data, the ranges of the scanning range, incidence angle and BSM can be designed based on the experimental needs to generate sufficiently large (and dense) training samples. Similarly, in previous studies (Ahmad et al., 2010; Ait Hssaine et al., 2020; Mirsoleimani et al., 2019; Notarnicola et al., 2008; Paloscia et al., 2008; Paloscia et al., 2013; Pasolli et al., 2011), the simulated samples were extensively applied for training machine learning models to estimate the soil surface moisture from the remote sensing data.

In this study, the ranges of the input variables (for generating simulated intensity data) correspond to those acquired during the field measurements (as demonstrated in Table 2). Besides, the simulated data have a sufficiently large number of samples and a uniform distribution of the input variables, thus presenting more general BSM situations (0–25%) than the field samples. It should be noted that the simulated training dataset and the test datasets are completely independent of each other due to the use of the different values of input variables. Besides, we added Gaussian noise to the simulated intensity data for a good representation of reality. The noise variance is approximately 0.07, derived from the field point cloud data acquired on the upper-beach.

The SVR models were trained with the gradually increased training samples (e.g., 20, 50, 100...) which were subsampled randomly from the 22,308 simulated training samples (Table 2). The SVR trials were repeated 50 times for each sample size to obtain a robust result (with the re-sampled training samples for each trial). The minimum size of the training samples was determined when the prediction accuracy (RMSE) of the developed SVR models on the test datasets tended to be steady with the increased training sample sizes. Similarly, the influence of the sample density was also investigated by gradually increasing the range intervals, incidence angle intervals or moisture intervals of the training samples (also subsampled from the 22,308 simulated training samples) and then assessing the estimation accuracy (RMSE) of the developed SVR models on the test datasets. For example, we trained the SVR model using the samples with moisture contents of 0%, 13% and 25% (a moisture interval of about 13%) or using the samples with a range interval of 100 m. For each sample density case (or feature interval), the SVR training was repeated 20 times using the same training set (which was sufficient to obtain a robust result based on the findings of pre-trials).

After the training phase, the developed SVR models' estimation performance (RMSE and R^2) was assessed based on the simulated 2520

test samples. Considering that the simulated data might include the modeling errors of Eq. (3), the 78 field test samples were also exploited to assess the prediction performance of the developed SVR models under realistic field conditions. Besides, we also calculated the RMSE values of SVR predicted moisture based on the separate test datasets, consisting of three sets of simulated test datasets (simulated Range, Angle and Moisture test sets in Table 2) and three sets of field test datasets (28 upper-beach samples for incidence angle, 39 upper-beach samples for range and 11 samples from the sampling grid, see Section 2.4). Because the simulated Range and Angle test sets had a constant moisture content of 3.5% and the upper-beach range and angle test sets had a constant moisture content of 0.05%, their RMSE values could be used to assess whether the impact of the range and incidence angle on the back-scattered intensity has been accurately corrected. The RMSE values of the separate moisture test sets (field and simulated) were applied to assess the prediction accuracy of the trained SVR models under the different moisture intervals of training samples.

2.7. Comparing SVR and ANN for BSM estimation

Among the different machine learning methods, ANN has played a dominant role in soil moisture retrieval in the past three decades (Ali et al., 2015; Padarian et al., 2020). In this study, the ANN algorithm was also adopted to compare the estimation performance of ANN and SVR under conditions of limited training samples. The relevant theoretical background and detailed training methods can be found in previous literature (Behrens et al., 2005; Kolassa et al., 2018; McCulloch and Pitts, 1988; Paloscia et al., 2008). In this study, a three-layer structure (3:20:1) was used for ANN model development after several tests. To avoid overfitting, we adopted the Bayesian regularization back-propagation algorithm (Foresee and Hagan, 1997; MacKay, 1992) to train the ANN models, which can automatically optimize the number of the effective weights and biases used by the networks (Oliferenko et al., 2013; Zhang et al., 2010). The training was stopped when the ANN training achieved the maximum number of iterations (here set at 100 based on several pre-tests) or the mean square error (MSE) between the ANN outputs and the target data was <0.001 (corresponding to $\sim 0.8\%$ moisture). It should be mentioned that, in the pre-tests, the ANN models' estimation accuracy (MSE) on the calibration dataset tended to be steady after 10–30 times iterations.

The SVR and ANN models were developed using the same training and test dataset. The selection of the training dataset was done by referring to the experimental results in Section 2.6. In this study, we mainly compared the estimation accuracy and generalization ability of SVR and ANN algorithms under the same condition of the minimum training sample size or density (Table 1). The experiments were repeated 20 times for the SVR and ANN respectively and the model with the lowest test RMSE was used to produce the BSM map of the entire study area. Also, we produced the moisture difference maps between the predicted moisture by the developed SVR or ANN models and the statistical model, which were used to evaluate the generalization ability and intensity correction performance of the developed SVR or ANN models.

Table 2
Simulated data using the statistical model.

	Range (m)			Incidence Angle (°)			Moisture (%)			Count
	Min.	Max.	Int.	Min.	Max.	Int.	Min.	Max.	Int.	
Training	60	440	10	45	87	2	0	25	1	22,308
Test	62	438	30	46	84	2	0.5	24.5	3	2,520
Range test	62	438	10	Constant 63			Constant 3.5			39
Angle test	Constant 115			46	84	1	Constant 3.5			39
Moisture test	Constant 115			Constant 63			0.5	24.5	1	25

3. Results

3.1. Kernel function selection

At the initial stage of the SVR experiments, the SVR models (for estimating BSM) were developed using four different kernel function types (linear, polynomial, RBF and sigmoid) based on the field moisture samples (Table 1). As shown in Table 3, the average RMSE and R^2 of the 20-times SVR experiments (for each kernel function) were calculated to determine the optimal kernel function. The kernel function selection had a significant impact on the estimation accuracy of the developed SVR models. The SVR models utilizing the RBF kernel performed best on the test dataset with the lowest RMSE of 0.71% moisture and the highest R^2 of 0.98, followed by the polynomial kernel (0.86% and 0.96), linear kernel (1.91% and 0.84) and sigmoid kernel (2.01% and 0.83). The results were similar to those from previous studies (de Souza et al., 2019; Mountrakis et al., 2011). In addition, compared with the polynomial kernel, the RBF kernel had lower computational complexity (only three tuning parameters) and less computational time. Thus, the further SVR experiments adopted the RBF kernel function in this study.

3.2. Beach surface moisture estimation

Based on the methods introduced in Section 2.4, the scatter plot (Fig. 3a) shows the SVR model's training and testing results for the BSM estimation based on the field samples. The sample points were located closely around the bisector line (solid black line), indicating that the trained model performs well for both the training and test data. The RMSE and R^2 of the test data were 0.68% and 0.98, respectively. In addition, the SVR models provided consistent test results for 50 trials with an average RMSE of $0.71\% \pm 0.02\%$ and an average R^2 of $0.98\% \pm 0.002\%$, showing the high reproducibility of the SVR algorithm. In the moisture map (Fig. 3b), the SVR model seemed to perform well in the area covered by the training samples (i.e., the area of ranges less than 250 m and the upper-beach) but did not show satisfactory results at long ranges without the training samples covered (the upper-right of Fig. 3b).

According to the scatterplots in Fig. 4, the SVR predicted moisture was fully independent of the incidence angles and ranges. Besides, the standard deviation of SVR predicted moisture was quite small, with a moisture of 0.04% for the test data of incidence angle correction (the red solid triangle in Fig. 4a) and a moisture of 0.46% for the test data of range correction (the red solid triangle in Fig. 4b). This indicated that the impact of scanning geometry on the intensity had been accurately corrected. However, it should be mentioned that both test datasets were extracted from the upper-beach with quite low moisture.

In order to further evaluate the SVR model's generalization ability and intensity correction performance for the areas with the middle and high moisture, the SVR predicted moisture contents were compared with those retrieved using the SR method (see Section 2.5). As shown in Fig. 5a, the SR method performed satisfactorily for the training and test dataset, denoting a test RMSE of 0.79% and R^2 of 0.97. It should be mentioned that the RMSE was relatively lower than those from previous studies (Jin et al., 2021; Smit et al., 2018). To some extent, this might be caused due to using more low-moisture samples in this study (the lower-right of Fig. 5a). The standard deviations of SR predicted moisture based on the incidence angle or range correction test data were almost the

same as those of SVR, with values of 0.09% and 0.45% respectively. This indicated that the SR method had correctly eliminated the impact of scanning geometry on the intensity data.

Fig. 3c represents the difference map between the derived BSM using the trained SVR model and the SR model. Compared to the scatter diagram (which utilizes the limited sample points), the moisture difference map (over the entire study area) provided more intuitive and continuous information to assess the estimation accuracy and generalization ability of the trained SVR model. As shown in Fig. 3c, the predicted BSM contents of the SVR were almost equal to those predicted by the statistical model in the region covered by the training samples. The absolute moisture differences only amounted to less than 2% in this region.

However, it is noted that the SVR estimation performance significantly deteriorated for the ranges exceeding 250 m. Although the training data included 39 range correction samples (61–440 m) extracted from the upper-beach (with moisture of 0.05%), the results still seemed to be unsatisfactory for the high-moisture region at ranges greater than 250 m. One possible reason is the lack of high-moisture training samples at ranges exceeding 250 m. In order to address this issue, we tried to reduce the proportion of the low-moisture samples in the training dataset or tune the SVR parameters manually, but no significant improvement was observed for the estimation accuracy at long ranges. This indicates that the SVR needs a well-spread dataset of training samples to achieve a good generalization.

Furthermore, we added 35 samples of 26% moisture (Fig. 6c) to the training set and retrained SVR models using the new training set (repeated 50 times). The estimation performance of the retrained SVR models was quite consistent, with an average RMSE of $1.08\% \pm 0.05\%$ and an average R^2 of $0.96\% \pm 0.004\%$ on the test dataset. The prediction performance of the best (one) of the 50 retrained SVR models is shown in Fig. 6, with an RMSE of 0.98% and an R^2 of 0.97. Similar to Fig. 3, the predicted moisture of the retrained SVR model was also independent of the incidence angles and ranges (Fig. 7). The standard deviation of SVR predicted moisture was 0.01% for the test data of incidence angle correction and 0.47% for the test data of range correction. Besides, the retrained SVR model performed well in the entire study area, including the long ranges. Overall, almost all (absolute) differences between the BSM contents that were predicted by the retrained SVR model and the statistical model were less than 2%.

3.3. The impact of the training sample size

To investigate the SVR's potential to handle small sample datasets, we investigated and compared the estimation performance of the SVR models with different sizes or densities of the (simulated) training samples (based on the method introduced in Section 2.6). As shown in Fig. 8, the data points denote the average RMSE or R^2 of 50 trials for each sample size, and the error bars indicate the corresponding standard deviation. Overall, both the RMSE and R^2 of the developed SVR models gradually improved with the increased training sample sizes. The two statistics tended to be steady while using 200 (randomly subsampled) training samples with an RMSE of 1.05% and R^2 of 0.95 for the field test data. Regarding the simulated test data, the corresponding RMSE and R^2 were 1.29% and 0.97 respectively. The variability (standard deviation) of the estimation accuracy also gradually reduced as the increased size of the training sample, from 1.28% at 20 training samples to 0.16% at 200 training samples (for the field test dataset).

In Fig. 8c, the RMSE values of the separate test datasets were used to evaluate the intensity correction performance of the developed SVR models (based on the method in Section 2.6). The results illustrated that the SVR models' performance (RMSE) on the separate test datasets was similar to the one in Fig. 8a and also attained stabilization when using 200 (randomly subsampled) training samples, with the RMSE < 1% for the range and incidence angle test dataset. This indicated that the impacts of scanning geometry on the backscattered intensity had been corrected in the process of estimating BSM when using more than 200

Table 3
Performance of the SVR models with four kernel functions.

Kernel Function	Training		Testing	
	RMSE	R^2	RMSE	R^2
Linear	2.17%	0.88	1.91%	0.84
Polynomial	1.23%	0.96	0.86%	0.96
RBF	1.28%	0.96	0.71%	0.98
Sigmoid	2.26%	0.87	2.01%	0.83

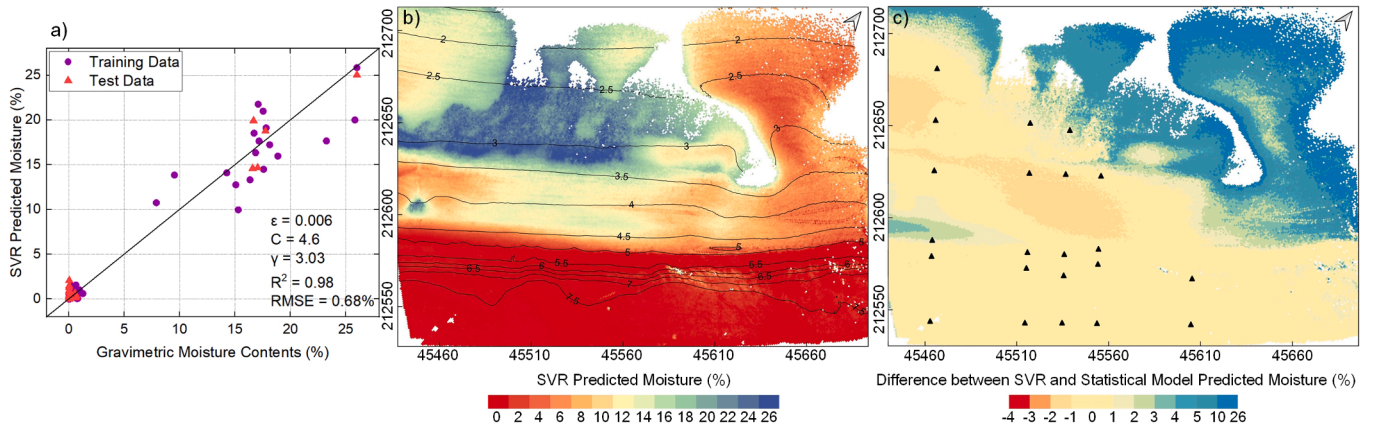


Fig. 3. (a) The scatter diagram of the relation between the measured and SVR predicted sample moisture based on the field samples. (b) The surface moisture map of the study area predicted by the developed SVR model. (c) The difference map between the predicted moisture respectively by the trained SVR model and the statistical model. The black triangles denote the locations of the 44 field moisture samples (ranges < 250 m) collected at the established sampling grid.

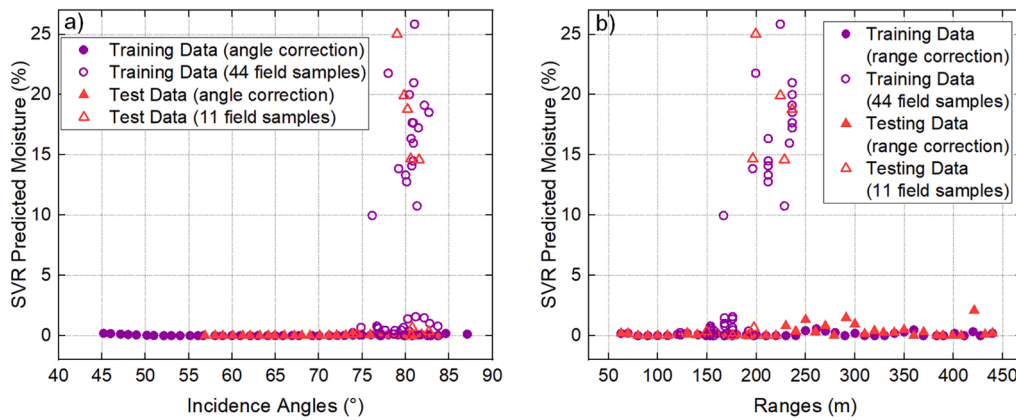


Fig. 4. (a) The scatter diagram of the relation between the SVR predicted sample moisture and incidence angles based on the field samples. (b) The scatter diagram of the relation between the SVR predicted sample moisture and the scanning range.

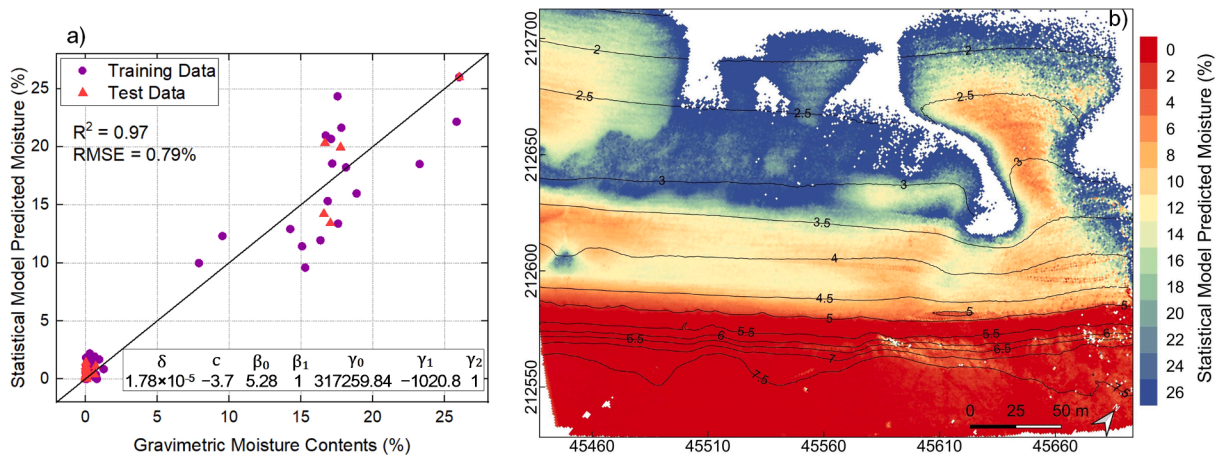


Fig. 5. (a) The scatter diagram of the relation between the measured moisture and the statistical model's predicted moisture, based on the field samples. The model parameters are shown at the bottom of the diagram. (b) The surface moisture map of the study area predicted by the statistical model.

(randomly subsampled) training samples.

3.4. The impact of the training sample density

Based on the method in Section 2.6, the RMSE and R^2 of the SVR

models varied as a function of the training samples' distance intervals (Fig. 9a–c), incidence angle intervals (Fig. 9d–f) and moisture intervals (Fig. 9g–i). With the increased intervals of the three features, the test RMSE of the SVR models remained steady ($RMSE < 1.5\%$) until reaching the thresholds and then deteriorated sharply (Fig. 9a, 9d and 9g). In this

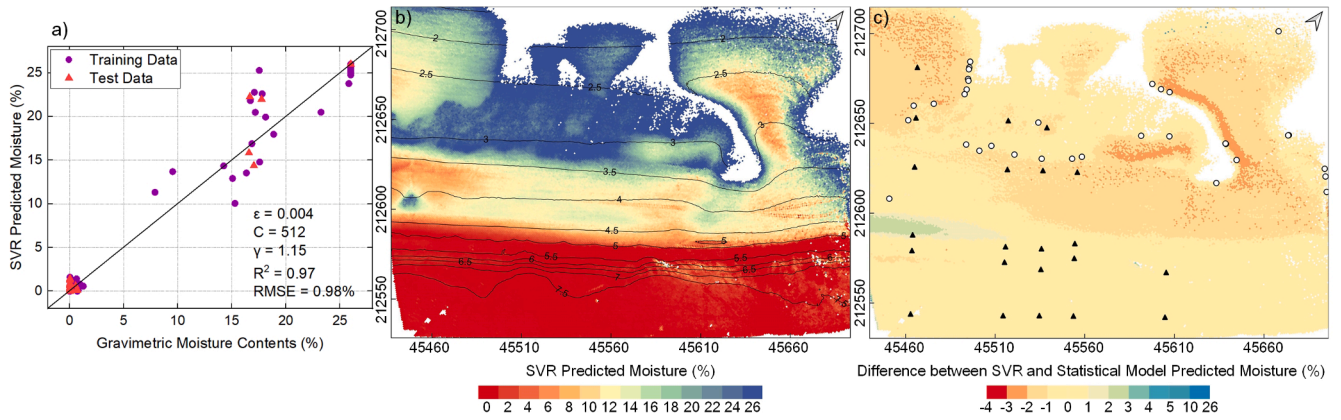


Fig. 6. (a) The scatter diagram of the relation between the measured and SVR predicted sample moisture, based on the field samples and additional 35 samples of 26% moisture (black circles). (b) The surface moisture map of the study area predicted by the trained SVR model. (c) The difference map between the predicted moisture by the trained SVR model and the statistical model, respectively. The black triangles denote the locations of the 44 field moisture samples (ranges < 250 m) collected at the established sampling grid.

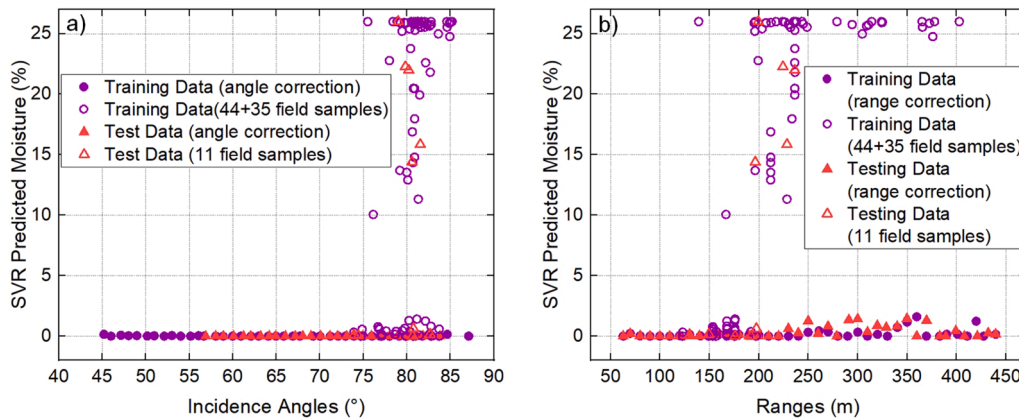


Fig. 7. (a) The scatter diagram of the relation between the SVR predicted sample moisture and incidence angles, based on the field samples and additional 35 samples of 26% moisture (b) The scatter diagram of the relation between the SVR predicted sample moisture and range.

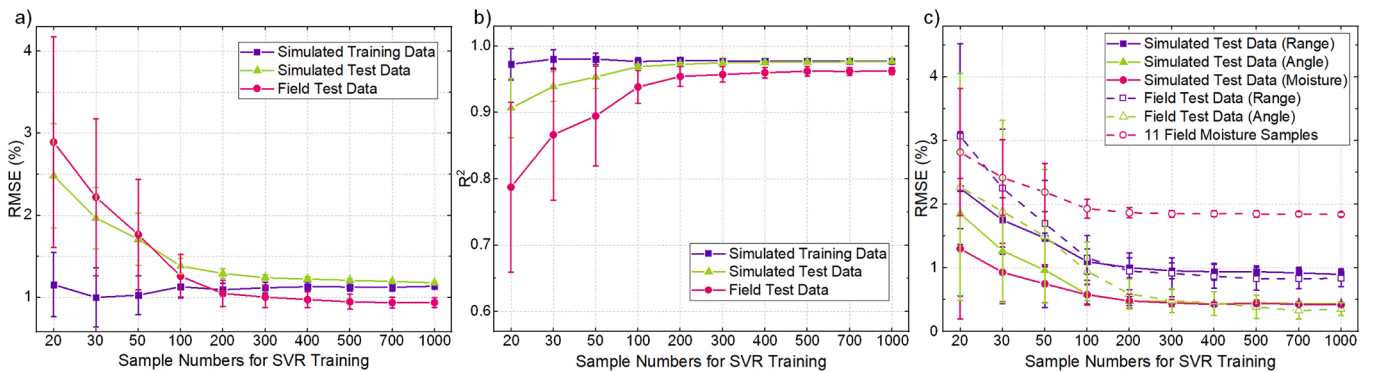


Fig. 8. (a) and (b) respectively show the RMSE and R^2 variation of the developed SVR models as a function of the sizes of the simulated training samples, based on the (simulated and field) comprehensive test dataset. (c) The estimation accuracy (RMSE) of the developed SVR models varied as a function of the simulated training samples' sizes based on the (simulated and field) separate test datasets for range, incidence angle and moisture.

study, the thresholds were 80 m for the range interval, 42° for the incidence angle interval, and 13% for the moisture interval. Similar trends were also observed for the determination coefficient R^2 in Fig. 9b, e and h.

The range intervals of training samples had a greater impact on the estimation accuracy of the SVR model than the incidence angle intervals and moisture intervals. Especially, the SVR models still performed well

even with the maximum angle interval of 42° (i.e., only including the minimum (45°) and maximum (87°) incidence angle), with an RMSE of $1.43\% \pm 0.74\%$ and R^2 of 0.95 ± 0.01 for the field test data (Fig. 9d–e). This is mainly because the TLS incidence angle had less impact on the backscattered intensity, having less impact on the predicted BSM as a result (Jin et al., 2021). Another possible reason is that we used the cosine values of incidence angle rather than the incidence angle in

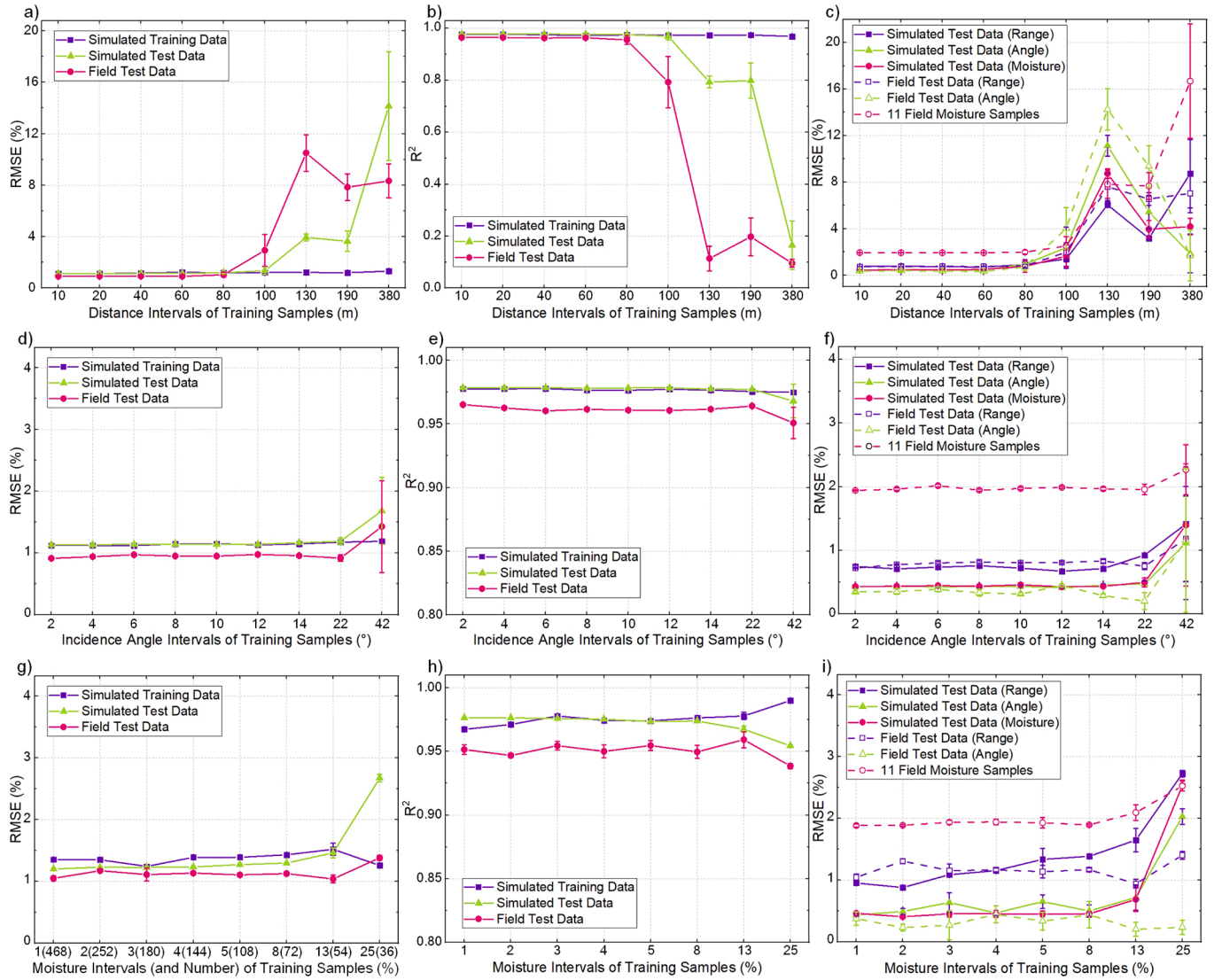


Fig. 9. The RMSE and R^2 variation of the developed SVR models as a function of the distance intervals (a–c), incidence angle intervals (d–f) and moisture intervals (g–i) of the simulated training samples. The developed SVR models were evaluated using both the simulated and field test datasets.

radians as the input feature of SVR. The former is linearly correlated with the backscattered intensity based on the LiDAR equation (Fang et al., 2015; Xu et al., 2017).

When we investigated the impacts of moisture interval on the SVR estimation performance (Fig. 9g–i), the range and incidence angle intervals of training samples were fixed at 80 m and 22° respectively (rather than 10 m and 2°). This could reduce the training sample size (close to the actual size of field moisture samples) and further investigate the potential of SVR models to deal with small sample datasets. As shown in Fig. 9g–i, the SVR models performed well when only using 54 training samples (i.e., based on the intervals of the 80 m range, 22° incidence angle and 13% moisture), with an RMSE of $1.04\% \pm 0.06\%$ and R^2 of 0.96 ± 0.01 for the field test data. The estimation accuracy was slightly better than the one using 200 randomly selected training samples in Fig. 8. Besides, when the range interval of training samples exceeded the threshold (80 m), the SVR predictive accuracies on the test datasets of the other two features also became bad (Fig. 9c,). Similar phenomena were also observed in Fig. 9f and 9i. This implies that all the three variables of the training samples should be sufficiently dense (at least reaching the threshold of each feature interval) to correct the effect of scanning geometry and accurately estimate the BSM from the LiDAR intensity data.

3.5. Comparison of SVR with ANN

As shown in Fig. 10a and Fig. 11a, under the conditions of the minimum size and density of the training samples (i.e., the 54 samples in Section 3.4), both SVR and ANN achieved a high estimation accuracy, with an RMSE of 1.06% and 0.90%, respectively (for the field test dataset). The average RMSE of 20 experiments was $1.04\% \pm 0.06\%$ for SVR and $0.93\% \pm 0.09\%$ for ANN, indicating the high reproducibility and estimation accuracy. The difference map of the SVR and SR predicted moisture was almost monochromatic with the (absolute) moisture differences $< 1\%$ (Fig. 10c). A similar phenomenon was also observed for the ANN model in Fig. 11c. This suggested that both the developed SVR and ANN models achieved a good generalization and intensity correction performance. When training the SVR and ANN models using the 200 randomly selected training samples (see Section 3.3), the difference map of the SVR and SR predicted moisture (not shown) was similar to Fig. 10c, but a few scattered points with high moisture differences ($> 2\%$) were observed at long ranges. Besides, the SVR model performed slightly better than the ANN model with fewer points of high moisture difference.

We further compared the prediction performance (especially the generalization capability) of the SVR and ANN under extreme

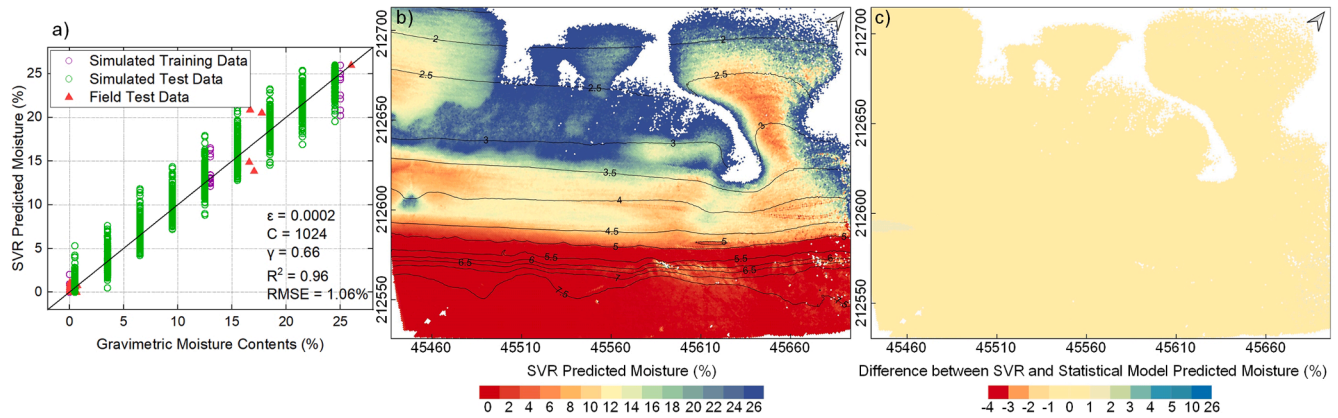


Fig. 10. (a) The scatter diagram of the relation between the actual and SVR predicted sample moisture, based on 54 simulated training samples of a uniform distribution. (b) The surface moisture map of the study area predicted by the developed SVR model. (c) The difference map between the predicted moisture contents by the trained SVR model and the statistical model, respectively.

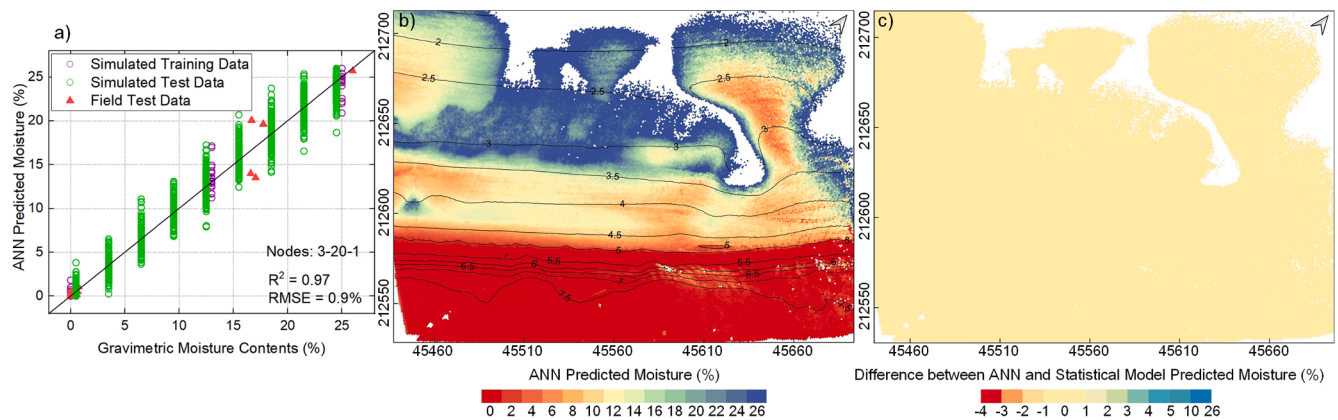


Fig. 11. (a) The scatter diagram of the relation between the actual and ANN predicted sample moisture, based on 54 simulated training samples of a uniform distribution (the same for those training the SVR). (b) The surface moisture map of the study area predicted by the trained ANN model. (c) The difference map between the predicted moisture contents by the trained ANN model and the statistical model, respectively.

conditions. Here, 16 field moisture samples (with a relatively uniform distribution in range and moisture) were selected from the field training dataset used in Fig. 6. The locations of these samples were shown in Fig. 12c (2 upper-beach samples with ranges > 360 m not shown). It should be noted that there were only samples with moisture of 0.05%

and 26% (no middle-moisture) at the ranges greater than 250 m. As shown in Fig. 12, even only using 16 training samples, the SVR model still performed acceptably with a test RMSE of 1.83% and an R^2 of 0.88. The average test RMSE and R^2 of 20 trials was $2.5\% \pm 0.24\%$ and 0.79 ± 0.03 , respectively. Compared with the statistical model's predicted

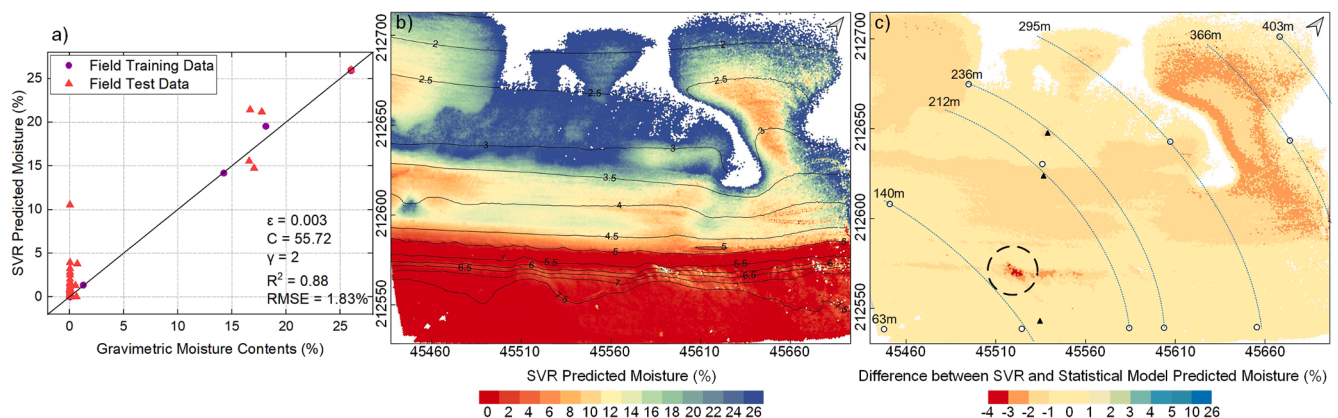


Fig. 12. (a) The scatter diagram of the relation between the measured and SVR predicted sample moisture, based on 16 field training samples with a uniform distribution. (b) The surface moisture map of the study area predicted by the developed SVR model. (c) The difference map between the predicted moisture contents by the trained SVR model and the statistical model. The blue lines denote the distribution of the 16 field samples (black triangles and circles). The dashed line denotes some points with a moisture difference of 3%–4% at the dune edge. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

moisture (Fig. 12c), the relatively large moisture difference (2%–3%) is mainly situated at ranges greater than 250 m. This is possible because of the lack of middle-moisture samples at long ranges. Also, there were some points with a moisture difference of 3%–4% at the dune edge (black dashed line) due to the relatively big topographic relief.

In contrast, based on the 16 field training samples, the ANN algorithm performed poorly with a test RMSE of 4.02% and R^2 of 0.58 (Fig. 13a). The average RMSE and R^2 of 20 trials amounted to $5.09\% \pm 0.98\%$ and 0.5 ± 0.08 , respectively. In particular, the predictions of the ANN model significantly deteriorated at long ranges, except in the sunken area of the upper-beach (black dashed line in Fig. 13c).

4. Discussion

Support vector regression showed good performances in estimating BSM from terrestrial LiDAR intensity data (Figs. 3–7), but some methodological decisions might affect its estimation performance. In this study, the SVR models utilizing the RBF kernel performed best on the test dataset, as has been observed in previous literature (de Souza et al., 2019; Mountrakis et al., 2011). The grid search method combined with 10-fold cross-validation was proven to be an effective method to identify optimum SVR parameters. However, in the process of tuning parameters, reducing the values of parameter ϵ will increase the number of support vectors and significantly increase the runtime of the model as a result. To reduce the computational time, one may first conduct a preliminary grid search using a coarse grid. After identifying a better and smaller region on the grid, the final grid search was carried out using finer parameter intervals on the small grid. This is similar to the method introduced by Hsu et al. (2003).

In this study, three input features (the backscattered intensity, scanning ranges and incidence angles) were selected for training SVR models to estimate the BSM. To our knowledge, this is the first use of SVR to retrieve BSM from LiDAR intensity data. Thus, the selection of input features was done by referring to some literature on traditional SR methods for BSM estimation (Jin et al., 2021; Nield et al., 2014; Philpot, 2010; Smit et al., 2018; Tan et al., 2020). The three input features selected were the same as the variables used to develop the BSM-intensity model based on the traditional SR method (Jin et al., 2021; Tan et al., 2020), which allows us to compare the two methods straightforwardly. Besides, based on the three input features, the generalization performance of the developed SVR models can also provide a reference for the study of the TLS backscattered intensity correction using SVR.

Theoretically, the TLS intensity data are also influenced by atmospheric conditions and instrumental mechanisms (Fang et al., 2015;

Höfle and Pfeifer, 2007). In this study, the TLS instrumental configurations were kept constant throughout. The atmospheric conditions near the study beach were also relatively stable, with an air temperature of about 14 ± 3 °C and a relative humidity of about $77 \pm 17\%$ (no precipitation) during the sampling period. Besides, since the atmospheric attenuation is closely related to the scanning range, the developed SVR models (using the range as the input feature) might correct the impact of atmospheric attenuation on TLS intensity data to some extent. Hence, the atmospheric attenuation of TLS intensity data was ignored in this study. Additionally, previous studies have shown the soil surface roughness plays a key role in retrieving surface soil moisture from the backscatter coefficient data of microwave remote sensing (Davidson et al., 2000; Ezzahar et al., 2020). However, for terrestrial laser scanning (normally based on visible or near-infrared wavelengths), only little literature is available and focusses on the impacts of surface roughness on TLS backscattered intensity at different incidence angles (Kaasalainen et al., 2009b; Pesci and Teza, 2008). A more detailed study is needed to fully investigate the relationship between soil/beach surface roughness and TLS intensity based on a series of controlled experiments (e.g., at different soil/sand roughness and considering the beam footprint sizes). In this study, the beach surface roughness was assumed to be constant, considering that most of the study area was relatively flat and a similar mineral composition (quartz sand) and grain size (D50: 291–337 μm) was visible on the beach.

Considering the difficulty of collecting a large number of field BSM samples, this study focused on model simplicity in terms of the minimum training sample size and density for SVR model development. Compared to the range interval of the training samples, the incidence angle interval only has a slight influence on the estimation accuracy of the SVR model. Even with the maximum angle interval of 42° (i.e., only including the minimum (45°) and maximum (87°) incidence angle), the SVR models still performed well. We also demonstrated that 200 randomly selected samples or 54 uniformly distributed samples (both from the simulated 22,308 training samples) were sufficient for training SVR models to accurately estimate BSM. Besides, the SVR models using 54 training samples of uniform distribution (Fig. 9g–i) performed slightly better than those using 200 randomly selected training samples (Fig. 8). This suggests that the SVR algorithm can achieve high estimation accuracies even with few training samples, provided that the training samples are representative and distributed evenly. This is similar to those observed in previous literature (Ballabio, 2009; Deiss et al., 2020; Pasolli et al., 2015). It should be noted that the experiments did not cover the ranges < 60 m, in which the built-in software of the scanner could automatically reduce the backscattered intensity (not monotonic reduction with the increasing range) and more training samples might be required for

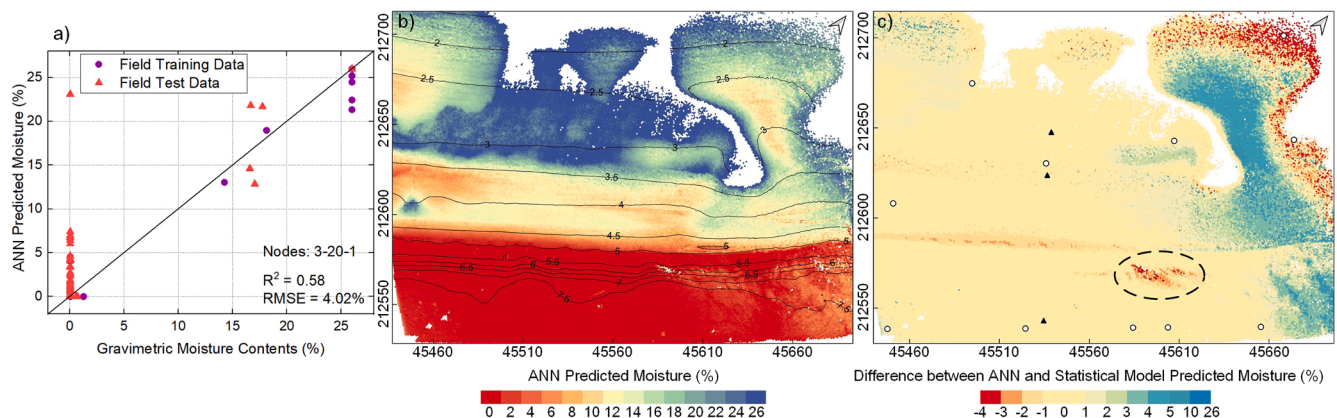


Fig. 13. (a) The scatter diagram of the relation between the measured and ANN predicted sample moisture, based on 16 field training samples with a uniform distribution. (b) The surface moisture map of the study area predicted by the trained ANN model. (c) The difference map between the predicted moisture contents by the trained ANN model and the statistical model. The black triangles and circles denote the locations of 16 field samples. The dashed line demonstrates some points with a moisture difference of 3%–4% in the upper-beach sunken area.

the range correction.

In this study, the SR method was used as a comparative reference to evaluate the generalization performance and intensity correction results of SVR models (Fig. 5). Based on the difference maps between the derived BSM using the SVR and SR models (Fig. 3c, Fig. 6c, Fig. 10c and Fig. 12c), we found that the SVR algorithm could accurately correct the effect of scanning geometry on the intensity in the process of estimating BSM, but the coverage and distribution of the training samples may significantly affect the generalization performance of the SVR model. In order to achieve good generalization of the SVR model, the training samples should cover a wide range of input features (e.g., at least covering the possible maximum, minimum and middle moisture, range and incidence angle). This finding could provide a reference for collecting field moisture samples or conducting indoor experiments for intensity correction in future studies.

We compared the SVR algorithm with ANN under minimal conditions in terms of the sample size and density (i.e., 54 uniformly distributed samples or 200 randomly selected samples), and both methods achieved a high estimation accuracy and generalization performance (Figs. 10–11). In actual field measurements, the incidence angle rises with the increase of scanning ranges, especially on relatively flat beaches. Theoretically, if the training samples cover the required scanning ranges (here 60–440 m), the corresponding incidence angles (54° – 87°) will also be covered. Moreover, the incidence angle only slightly affected the backscattered intensity. Thus, 18 field samples with a uniform distribution over the moisture and range value may be sufficient to develop an SVR model of high estimation performance. This finding was demonstrated in Fig. 12, in which the SVR performance was still acceptable even using only 16 field training samples of uniform distribution. Moreover, the SVR model performed much better than the ANN models under this condition, suggesting the better ability of SVR to deal with small training samples.

It should be mentioned that there are some limitations on the adopted long-range LiDAR, mainly including the sparse point density at long ranges and the beam divergence. In this study, the beach surface was assumed to be the extended Lambertian target (Höfle and Pfeifer, 2007) and the influence of beam divergence was ignored, considering that most of the study area was relatively flat. Because it was time-consuming to collect surface moisture samples using a large sampling size (e.g., 1×1 m) and the moisture contents in the later samples might vary somewhat during the sampling interval, this study collected surface moisture samples with a relatively small sampling size (i.e., 10×10 cm). However, there were only 1 or 2 TLS scanning points visible (even no points at ranges > 200 m) within the small sampling area of 10×10 cm. Thus, in the pre-processing of the moisture samples, the center locations of the moisture samples were connected to a corresponding 1×1 m grid cell in the TLS scan (rather than 10×10 cm grid cell). The average intensity, range and incidence angle of each surface moisture sample were extracted from the corresponding TLS point clouds of 1×1 m grid cell (> 9 reflected points). To be exact, the average moisture contents obtained from the two different sampling sizes might be slightly different. In this study, this difference was ignored because the surface moistures in the 1×1 m grid cell varied little, with an average standard deviation of 1.44% moisture (Jin et al., 2021).

Considering that the SVR algorithm is data-driven, the developed SVR models can be used directly in other study sites only if certain conditions are fulfilled, namely very similar sand types, surface roughness, atmospheric conditions and the same LiDAR instrument, etc. To improve the transferability of the developed SVR models, future studies will need to incorporate more input features (e.g., surface roughness and atmospheric factors) into the model training. However, it is hard to transfer the developed SVR models (based on a long-range LiDAR) to other TLS scanning data (e.g., those acquired by the common short- and middle-range TLS or the mobile TLS system) without re-training and parameter tuning. Compared to the long-range TLS adopted in this study, there is normally no problem of low point density for the short-

and middle-range TLS scanning using a geodetic tripod. However, the short- and middle-range TLS requires measurements at multiple scanning positions to cover a substantial beach section. This is quite time-consuming and the beach surface moisture in the later scan might change a little during the TLS scan interval.

5. Conclusions

This study provides, to our knowledge, the first assessment of the SVR algorithm to estimate high-resolution BSM from the LiDAR intensity data. A static long-range terrestrial LiDAR (Riegl VZ-2000) was applied to collect the point cloud data on the beach with a very high spatial (centimeters to decimeters) and temporal (one hour) resolution. Since the backscattered intensity depended on the scanning geometry of TLS, the scanning range and the incidence angle were employed together with the backscattered intensity as the input features to develop the SVR models. The grid search method (in combination with the 10-fold cross-validation) was used to tune the SVR parameters. Based on the results in this study, we conclude that:

- (i) Support Vector Regression (e-SVR) can accurately estimate beach surface moisture contents from the LiDAR intensity data (with an average test RMSE of $0.71\% \pm 0.02\%$ and R^2 of $0.98\% \pm 0.002\%$ for 50 times trials), which was slightly better than the traditional SR method. The RBF was the most suitable kernel for SVR model development in this study.
- (ii) The developed SVR models can eliminate the impacts of scanning geometry on the backscattered intensity in the process of estimating BSM, without correcting the backscattered intensity in advance.
- (iii) The coverage and distribution of the training samples will significantly affect the prediction performance of SVR models. For example, SVR performed poorly at long ranges without the training samples covered. The minimum size of training samples required for SVR model development was 54, provided that the samples were distributed widely and evenly (i.e., under the feature intervals of 80 m range & 22° incidence angle & 13% moisture).
- (iv) With sufficient training samples (at least 54 uniformly distributed samples or 200 randomly selected samples), the predictive performance of SVR was similar to that of ANN. It is worth noting that SVR still performed acceptably (with a test RMSE of 1.83%) even using extremely few training samples (only 16 field samples of uniform distribution), far better than the ANN (with a test RMSE of 4.02%).

CRedit authorship contribution statement

Junling Jin: Conceptualization, Methodology, Validation, Writing – original draft. **Jeffrey Verbeurg:** Investigation, Software, Formal analysis, Writing – review & editing. **Lars De Sloover:** Investigation, Software, Data curation, Writing – review & editing. **Cornelis Stal:** Conceptualization, Resources, Writing – review & editing. **Greet Deruyter:** Supervision, Writing – review & editing. **Anne-Lise Montreuil:** Resources, Writing – review & editing. **Sander Vos:** Resources, Writing – review & editing. **Philippe De Maeyer:** Supervision, Project administration, Funding acquisition. **Alain De Wulf:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgements

We would like to thank Pieter Rauwoens (KU Leuven), Annelies Vandenbulcke (Ghent University) and Sanders Vos (the Delft University of Technology and Baars-CIPRO [www.baars-cipro.nl]) for setting up and maintaining the laser scanner during the field campaign. Vayamundo Oostende is acknowledged for providing access to the rooftop installation location. Jeroen Saelens from the 3DDA Research Group (Ghent University) and Steven Muylaert from the Innovation Team of the Flemish Department of Public Works are recognized for acquiring, processing and providing the UAV data. Koen Vanstaen, Johan Verstraeten, Johan Vercruysse, Lennert Dobbelaere and Tom Seuryneck of Meetnet Vlaamse Banken are acknowledged for providing the meteorological data of the Belgian coast. Sabine Cnudde (Ghent University) is recognized for her guidance (thorough proofreading of the manuscript). This study was supported by the European Research Council [grant number 291206]; the Flemish Agency for Innovation by Science and Technology (VLAIO/IWT) [grant number 150028]; the Dutch NWO CoastScan Project [grant number 2018/STW/00505023]; and the Chinese Scholarship Council [grant number 201406410069].

References

- Ahmad, S., Kalra, A., Stephen, H., 2010. Estimating soil moisture using remote sensing data: A machine learning approach. *Adv. Water Resour.* 33 (1), 69–80.
- Ait Hssaine, B., Merlin, O., Ezzahar, J., Ojha, N., Er-Raki, S., Khabba, S., 2020. An evapotranspiration model self-calibrated from remotely sensed surface soil moisture, land surface temperature and vegetation cover fraction: application to disaggregated SMOS and MODIS data. *Hydrol. Earth Syst. Sci.* 24 (4), 1781–1803.
- Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., Notarnicola, C., 2015. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sens.* 7 (12), 16398–16421.
- Anderson, K., Croft, H., 2009. Remote sensing of soil surface properties. *Prog. Phys. Geogr.* 33 (4), 457–473.
- Anthony, E.J., Ruz, M.-H., Vanh  e, S., 2009. Aeolian sand transport over complex intertidal bar-trough beach topography. *Geomorphology* 105 (1–2), 95–105.
- Atherton, R.J., Baird, A.J., Wiggs, G.F., 2001. Inter-tidal dynamics of surface moisture content on a meso-tidal beach. *J. Coastal Res.* 482–489.
- Ballabio, C., 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. *Geoderma* 151 (3–4), 338–350.
- Bauer, B.O., Davidson-Arnott, R.G.D., Hesp, P.A., Namikas, S.L., Ollerhead, J., Walker, I. J., 2009. Aeolian sediment transport on a beach: Surface moisture, wind fetch, and mean transport. *Geomorphology* 105 (1–2), 106–116.
- Behrens, T., F  rster, H., Scholten, T., Steinr  cken, U., Spies, E.-D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. *J. Plant Nutr. Soil Sci.* 168 (1), 21–33.
- Carter, C., Liang, S., 2019. Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* 78, 86–92.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* 2 (3), 1–27.
- Chen, X.I., Hu, Q.I., 2004. Groundwater influences on soil moisture and surface evaporation. *J. Hydrol.* 297 (1–4), 285–300.
- Cornelis, W., Gabri  ls, D., 2003. The effect of surface moisture on the entrainment of dune sand by wind: an evaluation of selected models. *Sedimentology* 50 (4), 771–790.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine Learning* 20 (3), 273–297.
- Cristianini, N., Shawe-Taylor, J. (Eds.), 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.
- Darke, I., Davidson-Arnott, R., Ollerhead, J., 2009. Measurement of beach surface moisture using surface brightness. *J. Coastal Res.* 251, 248–256.
- Darke, I., Neuman, C.M., 2008. Field study of beach water content as a guide to wind erosion potential. *J. Coastal Res.* 24 (5 (245)), 1200–1208.
- Davidson-Arnott, R., 2010. *Introduction to coastal processes and geomorphology*. Cambridge University Press.
- Davidson-Arnott, R.G.D., Yang, Y., Ollerhead, J., Hesp, P.A., Walker, I.J., 2008. The effects of surface moisture on aeolian sediment transport threshold and mass flux on a beach. *Earth Surface Processes and Landforms: J. British Geomorphol. Res. Group* 33 (1), 55–74.
- Davidson, M.W.J., Toan, T.L., Mattia, F., Satalino, G., Manninen, T., Borgeaud, M., 2000. On the characterization of agricultural soil roughness for radar remote sensing studies. *IEEE Trans. Geosci. Remote Sens.* 38 (2), 630–640.
- De Sloover, L., De Wulf, A., Stal, C., Verbeurgt, J., Vos, S., 2019. Case Study of a Hypertemporal Terrestrial LiDAR to Monitor a Macrotidal Beach: Assessment of Different Calibration Procedures. *International Multidisciplinary Scientific GeoConference: SGEM* 19 (2.2), 57–64.
- Souza, G.S.A.d., Soares, V.P., Leite, H.G., Gleriani, J.M., do Amaral, C.H., Ferraz, A.S., Silveira, M.V.d.F., Santos, J.F.C.D., Velloso, S.G.S., Domingues, G.F., Silva, S., 2019. Multi-sensor prediction of Eucalyptus stand volume: A support vector approach. *ISPRS J. Photogramm. Remote Sens.* 156, 135–146.
- Deiss, L., Margenot, A.J., Culman, S.W., Demyan, M.S., 2020. Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma* 365, 114227. <https://doi.org/10.1016/j.geoderma.2020.114227>.
- Deka, P.C., 2014. Support vector machine applications in the field of hydrology: a review. *Appl. Soft Comput.* 19, 372–386.
- Deronde, B., Houthuys, R., Henriet, J.P., Lancker, V.V., 2008. Monitoring of the sediment dynamics along a sandy shoreline by means of airborne hyperspectral remote sensing and LIDAR: a case study in Belgium. *Earth Surf. Processes Landforms* 33 (2), 280–294. <https://doi.org/10.1002/esp.1545>.
- Edwards, B.L., Namikas, S.L., 2015. Characterizing the sediment bed in terms of resistance to motion: Toward an improved model of saltation thresholds for aeolian transport. *Aeolian Res.* 19, 123–128.
- Edwards, B.L., Namikas, S.L., D’Sa, E.J., 2013a. Simple infrared techniques for measuring beach surface moisture. *Earth Surf. Proc. Land.* 38 (2), 192–197.
- Edwards, B.L., Schmutz, P.P., Namikas, S.L., 2013b. Comparison of surface moisture measurements with depth-integrated moisture measurements on a fine-grained beach. *J. Coastal Res.* 29 (6), 1284–1291.
- Ezzahar, J., Ouadi, N., Zribi, M., Elfakh, J., Aouade, G., Khabba, S., Er-Raki, S., Chehbouni, A., Jarlan, L., 2020. Evaluation of Backscattering Models and Support Vector Machine for the Retrieval of Bare Soil Moisture from Sentinel-1 Data. *Remote Sens.* 12 (1), 72.
- Fang, W., Huang, X., Zhang, F., Li, D., 2015. Intensity correction of terrestrial laser scanning data by estimating laser transmission function. *IEEE Trans. Geosci. Remote Sens.* 53 (2), 942–951.
- Foresee, F.D., Hagan, M.T., 1997. Gauss-Newton approximation to Bayesian learning. In: *Proceedings of International Conference on Neural Networks (ICNN’97)*. IEEE, pp. 1930–1935.
- Ho  a, P.V., Giang, N.V., Binh, N.A., Hai, L.V.H., Pham, T.-D., Hasanlou, M., Tien Bui, D., 2019. Soil salinity mapping using SAR Sentinel-1 data and advanced machine learning algorithms: a case study at Ben Tre province of the Mekong River Delta (Vietnam). *Remote Sens.* 11 (2), 128.
- H  fle, B., Pfeifer, N., 2007. Correction of laser scanning intensity data: Data and model-driven approaches. *ISPRS J. Photogramm. Remote Sens.* 62 (6), 415–433.
- Holtgrave, A.-K., F  rster, M., Greifeneder, F., Notarnicola, C., Kleinschmit, B., 2018. Estimation of Soil Moisture in Vegetation-Covered Floodplains with Sentinel-1 SAR Data Using Support Vector Regression. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation. Science* 86 (2), 85–101.
- Horn, D.P., 2002. Beach groundwater dynamics. *Geomorphology* 48 (1–3), 121–146.
- Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2003. *A practical guide to support vector classification*. Taipei.
- Jin, J., De Sloover, L., Verbeurgt, J., Stal, C., Deruyter, G., Montreuil, A.-L., De Maeyer, P., De Wulf, A., 2020. Measuring Surface Moisture on a Sandy Beach based on Corrected Intensity Data of a Mobile Terrestrial LiDAR. *Remote Sens.* 12 (2), 209.
- Jin, J., Verbeurgt, J., De Sloover, L., Stal, C., Deruyter, G., Montreuil, A.-L., Vos, S., De Maeyer, P., De Wulf, A., 2021. Monitoring spatiotemporal variation in beach surface moisture using a long-range terrestrial laser scanner. *ISPRS J. Photogramm. Remote Sens.* 173, 195–208.
- Kaasalainen, S., Jaakkola, A., Kaasalainen, M., Krooks, A., Kukko, A., 2011. Analysis of incidence angle and distance effects on terrestrial laser scanner intensity: Search for correction methods. *Remote Sens.* 3 (10), 2207–2221.
- Kaasalainen, S., Krooks, A., Kukko, A., Kaartinen, H., 2009a. Radiometric calibration of terrestrial laser scanners with external reference targets. *Remote Sens.* 1 (3), 144–158.
- Kaasalainen, S., Vain, A., Krooks, A., Kukko, A., 2009b. Topographic and distance effects in laser scanner intensity correction. *ISPRS Workshop Laser Scanning* 219–223.
- Kolassa, J., Reichle, R.H., Liu, Q., Alemohammad, S.H., Gentile, P., Aida, K., Asanuma, J., Bircher, S., Caldwell, T., Colliander, A., Cosh, M., Holfield Collins, C., Jackson, T.J., Mart  nez-Fern  ndez, J., McNairn, H., Pacheco, A., Thibeault, M., Walker, J.P., 2018. Estimating surface soil moisture from SMAP observations using a Neural Network technique. *Remote Sens. Environ.* 204, 43–59.
- Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. *Geosci. Front.* 7 (1), 3–10.
- Legates, D.R., Mahmood, R., Levya, D.F., DeLiberty, T.L., Quiring, S.M., Houser, C., Nelson, F.E., 2011. Soil moisture: A central and unifying theme in physical geography. *Prog. Phys. Geogr.* 35 (1), 65–86.
- MacKay, D.J.C., 1992. Bayesian interpolation. *Neural Comput.* 4 (3), 415–447.
- Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* 39 (9), 2784–2817.
- McCulloch, W.S., Pitts, W., 1988. A logical calculus of the ideas immanent in nervous activity. *S*, 15–27.
- McKenna Neuman, C., Langston, G., 2006. Measurement of water content as a control of particle entrainment by wind. *Earth Surface Processes and Landforms: J. Br. Geomorphol. Res. Group* 31 (3), 303–317.
- McLachlan, A., 1989. Water filtration by dissipative beaches. *Limnol. Oceanogr.* 34 (4), 774–780.
- Mirsoleimani, H.R., Sahebi, M.R., Baghdadi, N., El Hajj, M., 2019. Bare soil surface moisture retrieval from sentinel-1 SAR data based on the calibrated IEM and dubois models using neural networks. *Sensors* 19 (14), 3209.
- Montreuil, A.-L., Chen, M., Brand, E., Strypsteen, G., Rauwoens, P., Vandenbulcke, A., De Wulf, A., Dan, S., Verwaest, T., 2018. Dynamics of Surface Moisture Content on a Macro-tidal Beach. *J. Coastal Res.* 85 (sp1), 206–210.

- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* 66 (3), 247–259.
- Namikas, S.L., Edwards, B.L., Bitton, M.C.A., Booth, J.L., Zhu, Y., 2010. Temporal and spatial variabilities in the surface moisture content of a fine-grained beach. *Geomorphology* 114 (3), 303–310.
- Nield, J.M., King, J., Jacobs, B., 2014. Detecting surface moisture in aeolian environments using terrestrial laser scanning. *Aeolian Res.* 12, 9–17.
- Nield, J.M., Wiggs, G.F.S., 2011. The application of terrestrial laser scanning to aeolian saltation cloud measurement and its response to changing surface moisture. *Earth Surf. Proc. Land.* 36 (2), 273–278.
- Nield, J.M., Wiggs, G.F.S., Squirrell, R.S., 2011. Aeolian sand strip mobility and protodune development on a drying beach: examining surface moisture and surface roughness patterns measured by terrestrial laser scanning. *Earth Surf. Proc. Land.* 36 (4), 513–522.
- Nolet, C., Poortinga, A., Roosjen, P., Bartholomeus, H., Ruessink, G., Magar, V., 2014. Measuring and modeling the effect of surface moisture on the spectral reflectance of coastal beach sand. *PLoS ONE* 9 (11), e112151.
- Notarnicola, C., Angiulli, M., Posa, F., 2008. Soil moisture retrieval from remotely sensed data: Neural network approach versus Bayesian method. *IEEE Trans. Geosci. Remote Sens.* 46 (2), 547–557.
- Oliferenko, A.A., Oliferenko, P.V., Torrecilla, J.S., Katritzky, A.R., 2013. Rebuttal to “Comments on ‘Boiling Points of Ternary Azeotropic Mixtures Modeled with the Use of the Universal Solvation Equation and Neural Networks’”. *Ind. Eng. Chem. Res.* 52 (1), 545–546.
- Padarian, J., Minasny, B., McBratney, A.B., 2020. Machine learning and soil sciences: A review aided by machine learning tools. *Soil* 6 (1), 35–52.
- Paloscia, S., Pampaloni, P., Pettinato, S., Santi, E., 2008. A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR images. *IEEE Trans. Geosci. Remote Sens.* 46 (10), 3274–3284.
- Paloscia, S., Pettinato, S., Santi, E., Notarnicola, C., Pasolli, L., Reppucci, A., 2013. Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sens. Environ.* 134, 234–248.
- Pan, Z., Glennie, C., Legleiter, C., Overstreet, B., 2015. Estimation of water depths and turbidity from hyperspectral imagery using support vector regression. *IEEE Geosci. Remote Sens. Lett.* 12 (10), 2165–2169.
- Pasolli, L., Notarnicola, C., Bertoldi, G., Bruzzone, L., Remelgado, R., Greifeneder, F., Niedrist, G., Della Chiesa, S., Tappeiner, U., Zebisch, M., 2015. Estimation of soil moisture in mountain areas using SVR technique applied to multiscale active radar images at C-band. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8 (1), 262–283.
- Pasolli, L., Notarnicola, C., Bruzzone, L., 2011. Estimating soil moisture with the support vector regression technique. *IEEE Geosci. Remote Sens. Lett.* 8 (6), 1080–1084.
- Pesci, A., Teza, G., 2008. Effects of surface irregularities on intensity data from laser scanning: an experimental approach. *Ann. Geophys.* 51 (5–6), 839–848.
- Philpot, W., 2010. Spectral reflectance of wetted soils. *Proc. ASD IEEE GRS 2*, 1–12.
- Ruessink, G., Brakenhoff, L., van Maarseveen, M., 2014. Measurement of surface moisture using infra-red terrestrial laser scanning, EGU General Assembly Conference Abstracts.
- Schmutz, P.P., Namikas, S.L., 2011. Utility of the Delta-T Theta Probe for obtaining surface moisture measurements from beaches. *J. Coastal Res.* 27 (3), 478–484.
- Schmutz, P.P., Namikas, S.L., 2018. Measurement and modeling of the spatiotemporal dynamics of beach surface moisture content. *Aeolian Res.* 34, 35–48.
- Smit, Y., Donker, J.J., Ruessink, G., 2019. Spatiotemporal surface moisture variations on a barred beach and their relationship with groundwater fluctuations. *Hydrology* 6 (1), 8.
- Smit, Y., Ruessink, G., Brakenhoff, L.B., Donker, J.J.A., 2018. Measuring spatial and temporal variation in surface moisture on a coastal beach with a near-infrared terrestrial laser scanner. *Aeolian Res.* 31, 19–27.
- Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. *Statistics Comput.* 14 (3), 199–222.
- Sunder, S., Ramsankaran, R., Ramakrishnan, B., 2020. Machine learning techniques for regional scale estimation of high-resolution cloud-free daily sea surface temperatures from MODIS data. *ISPRS J. Photogramm. Remote Sens.* 166, 228–240.
- Tan, K., Chen, J., Zhang, W., Liu, K., Tao, P., Cheng, X., 2020. Estimation of soil surface water contents for intertidal mudflats using a near-infrared long-range terrestrial laser scanner. *ISPRS J. Photogramm. Remote Sens.* 159, 129–139.
- Tan, K., Cheng, X., 2015. Intensity data correction based on incidence angle and distance for terrestrial laser scanner. *J. Appl. Remote Sens.* 9 (1), 094094. <https://doi.org/10.1117/1.JRS.9.094094>.
- VLIZ, F.M.I., 2020. Crest project, <http://www.crestproject.be/en>.
- Vos, S., Lindenberg, R., de Vries, S., 2017. Coastscan: Continuous monitoring of coastal change using terrestrial laser scanning. *Proceedings of the Coastal Dynamics*.
- Wiggs, G.F.S., Baird, A.J., Atherton, R.J., 2004. The dynamic effects of moisture on the entrainment and transport of sand by wind. *Geomorphology* 59 (1–4), 13–30.
- Xiao, X., Zhang, T., Zhong, X., Shao, W., Li, X., 2018. Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data. *Remote Sens. Environ.* 210, 48–64.
- Xu, T., Xu, L., Yang, B., Li, X., Yao, J., 2017. Terrestrial laser scanning intensity correction by piecewise fitting and overlap-driven adjustment. *Remote Sens.* 9 (11), 1090.
- Yang, Q., Snyder, J., Tobler, W., 2020. Map projection transformation: principles and applications. CRC Press 28–32.
- Yang, X., Yu, Y., Li, M., 2019. Estimating soil moisture content using laboratory spectral data. *J. For. Res.* 30 (3), 1073–1080.
- Yang, Y., Davidson-Arnott, R.G.D., 2005. Rapid measurement of surface moisture content on a beach. *J. Coastal Res.* 213, 447–452.
- Zhang, H.-T., Xu, F.-Y., Zhou, L., 2010. Artificial neural network for load forecasting in smart grid. In: 2010 International Conference on Machine Learning and Cybernetics. IEEE, pp. 3200–3205.
- Zhu, Y., Weindorf, D.C., Chakraborty, S., Haggard, B., Johnson, S., Bakr, N., 2010. Characterizing surface soil water with field portable diffuse reflectance spectroscopy. *J. Hydrol.* 391 (1–2), 133–140.