Soil Texture and pH Mapping Using Remote Sensing and Support Sampling

Onur Yüzügüllü, Noura Fajraoui, and Frank Liebisch

***Abstract*—Soil pH and texture are valuable information for agriculture, supporting the achievement of high productivity and low environmental impact, which is the basis for sustainable agricultural production. In this study, we present novel soil mapping techniques that integrate high-spatial-resolution satellite and ground data, surpassing traditional methods in precision and reliability. By synergizing remote sensing data, including polarimetric synthetic aperture and multispectral imagery, with climate and terrain information, alongside coarse-resolution soil data, we achieved high accuracy, with an average error of less than 6%, in predicting soil pH and texture parameters. Notably, the approach allows for detailed mapping at the pixel level, revealing nuanced variability within 10×10 m field pixels. Considering the accuracy, the method establishes itself as a benchmark for field management guidelines integrating a precision sampling approach, offering actual and high spatial resolution information crucial for sustainable agricultural practices. This holistic approach allows new opportunities to revolutionize soil management practices, facilitating variable rate applications, soil moisture, and fertilization mapping and ultimately enhancing agri-environmental sustainability.**

***Index Terms*—Machine learning (ML), precision agriculture, remote sensing, soil health, soil mapping.**

I. Introduction

THE mapping of soil properties, encompassing vital parameters such as pH and soil texture, is fundamental for comprehending the complex dynamics within the agri-environment. It facilitates agricultural practices fostering soil fertility and health, leading to more efficient use of fertilizer and water resources. In addition, it enables the identification of regions with higher yield potential [1], [2], [3], [4]. The implications extend to plant growth dynamics, fostering sustainable agricultural practices, and preventing soil degradation—an essential step in mitigating climate change impacts [5], [6], [7].

In agriculture, the importance of soil pH and texture cannot be overstated as essential pillars for successful crop cultivation and effective management of input resources such as fertilizers. Soil pH, which denotes the soil's acidity or alkalinity, significantly affects the availability of nutrients to plants and soil stability [8], [9], [10], [11]. Deviations from the optimal pH range can lead to nutrient deficiencies or toxicities, adversely affecting crop growth and yield and fostering soil degradation [12]. Similarly, soil texture, characterized by the proportions of sand, silt, and clay particles, directly influences crucial factors such as water-holding capacity, drainage and aeration of soils, nutrient availability, and soil organic matter stability [13]. This information relies on knowing soil texture is pivotal as it directly impacts farmers' ability to implement efficient irrigation, manage water resources, and select appropriate crop varieties. The combined soil pH and texture assessment is the bedrock for precision agriculture practices, such as variable rate application. It allows farmers to customize their strategies, such as precision fertilization or seeding, to maximize crop productivity while minimizing environmental impact and costs [10], [11].

Traditionally, soil mapping in agricultural contexts heavily relied on conventional methods involving labor-intensive soil sampling and subsequent laboratory analyses. While these methods have proven effective, their time-consuming nature and significant financial requirements have been widely recognized. The difficulties in employing these techniques on a national, regional, or global scale have become increasingly apparent. Moreover, the substantial spatial variability of soil properties [14], [15], [16], [17] often leads to inaccuracies when extrapolating calculations to larger areas.

Innovative techniques have been continuously developed and improved to address this urgent need for more efficient and extensive soil data acquisition. Image-based soil mapping, in particular, has emerged as a promising solution, transcending the limitations of traditional methodologies [18], [19], [20], [21]. This approach seamlessly incorporates various secondary data sources into mapping, like the SCORPAN method [22], enhancing soil databases' precision and spatial coverage. In this context, remote sensing-based soil mapping offers a significant advantage in terms of cost-effectiveness, particularly when compared to the resource-intensive traditional soil survey methods [23].

Furthermore, modeling-based approaches provide an objective quantitative measure of prediction uncertainty, often absent in conventional methods [24], [25]. This inherent advantage refines the accuracy of soil property predictions, leading to more reliable soil maps. This advancement significantly benefits agricultural land management decision support thanks to three key factors as follows.

Manuscript received 22 February 2024; revised 22 April 2024; accepted 26 June 2024. Date of publication 3 July 2024; date of current version 24 July 2024. The work of Frank Liebisch was supported in part by the EJP Soil Project STEROPES under Grant 862695, Grant H2020, and Grant H2020-SFS-2019-1 and in part by EJP SOIL(2020))–Stimulating Novel Technologies from Earth Remote Observation to Predict European Soil Carbon within the Horizon 2020 program scheme. *(Corresponding author: Onur Yüzügüllü.)*

Onur Yüzügüllü and Noura Fajraoui are with the AgriCircle AG, Bahnhofstrasse 28b, Pfäffikon, 8808 Schwyz, Switzerland (e-mail: onur@agricircle.com).

Frank Liebisch is with the Agroscope, Water Protection, and Substance Flows, Reckenholzstrasse 191, 8045 Zurich, Switzerland.

Digital Object Identifier 10.1109/JSTARS.2024.3422494

1) *Efficient and Cost-Effective Monitoring:* Applicable across large areas cost-effectively.
2) *Overcoming Logistical Challenges:* Fills gaps where ground data collection is difficult or impossible.
3) *Noninvasive Approach:* Methods avoid direct soil interaction and disturbance.

Remote sensing captures soil pH and texture parameters by leveraging their temporal and electromagnetic characteristics. pH, gauged by hydrogen ion concentration, influences soil's physical, chemical, and biological properties. In the electromagnetic spectrum, chemical alterations impact the visible and near-infrared range [26], while physical changes affect electrical conductivity and dielectric properties [27]. Conversely, soil texture, reflecting particle size distribution, shapes how soil particles interact with electromagnetic waves. Simultaneously, variations in soil texture influence changes in soil moisture, detectable through shifts in dielectric properties [28].

Numerous research efforts have explored the correlation between satellite data and soil mapping. Notable studies conducted in the 1990s showcased the enhanced capability of satellite data in providing valuable insights for soil mapping [29], [30], [31]. These investigations highlighted how satellite technology substantially contributes to our understanding of soil composition, moisture levels, and other critical soil characteristics.

In the 2000s, significant progress was made in mapping soil properties, driven by the integration of remote sensing technology. This included increased satellite deployments and the usage of higher-resolution data, such as aerial photography, facilitating more accurate examination of spatial variation [18], [32], [33]. Researchers also incorporated optical and radar data to model soil property changes, with findings indicating the effective integration of radar imagery with optical data for extracting soil and plant information [34].

During the 2010s, there was a significant shift in soil mapping research, marked by the launch of Sentinel satellites with a spatial resolution of 10 to 20 m. This development greatly enhanced the capacity for detailed soil analysis [19], [35], [36]. During this decade, research also focused on data fusion techniques, integrating satellite imagery with ground-based measurements to improve the accuracy and applicability of soil mapping methods [19], [35], [36], [37], [38]. Notably, the release of SoilGrids250 m in 2017 marked a significant milestone, as it combined satellite data and machine learning (ML) to map various soil properties globally, both on the surface and in the soil profile down to 2 m [39].

Simultaneously with the satellite revolution, the first two decades of the 21st century saw a rapid development in advanced algorithms and ML techniques for processing and interpreting extensive remote sensing data. Nevertheless, directly mapping soil parameters with remote sensing remains a challenge due to complex interactions between various soil properties and their spectral response [21]. In this context, ground truth data, meticulously collected through strategically designed field sampling campaigns for soil texture and pH, plays a crucial role. By integrating Sentinel-2 data with these ground observations, researchers can leverage advanced statistical methods and ML algorithms like Random Forest or Support Vector Machines to develop increasingly accurate spatial predictions of soil texture and pH [40], [41], [42], [43]. This data-driven approach holds significant potential beyond traditional soil maps. Studies have shown the potential for applications in precision agriculture, enabling targeted interventions like optimized fertilizer application [44] and improved crop yield prediction [45]. In addition, these maps can contribute to environmental monitoring and sustainable land use management by providing insights into soil health and potential degradation risks.

This study presents an advanced approach that integrates high-spatial resolution satellite (10×10 m), topography (30×30 m), and a minimal number of ground data, whose locations are determined according to Precision Sampling [21], to calibrate the predictions spatially for comprehensive soil mapping, setting a new standard in precision agriculture and environmental management. The inclusion of a recent ground measurement provides an improvement in obtaining an accurate soil map, which makes the outputs of this method a reliable guideline for field management practices, enabling informed decision-making and optimized agricultural strategies. Furthermore, the inclusion of spectral information, climate data, topographical features, and coarse-resolution soil information in the soil mapping process represents a significant advancement in the field, offering a holistic understanding of soil properties and their spatial variability at a fine scale.

We introduce two frameworks: One relying solely on satellite data and ML and the other integrating supportive ground measurements for refinement. The ML components within both frameworks assess various algorithms, the light gradient boosting machine (LGBM), random forest (RF), and multilayer perception (MLP). To ensure objectivity, the frameworks undergo validation using an independent set of samples not used in the model development phase. The aim is to generate high-resolution (10 × 10 m) soil property maps utilizing an approach that combines satellite-based ML predictions with ground measurements, as was previously achieved for soil organic carbon [46]. The Precision Sampling technique [21], guided by spectral heterogeneity in Sentinel-1 and Sentinel-2 images, enhances predictions at the pixel level. The performance and robustness of the proposed approach are rigorously tested through three validation levels during model training, post-training refinement, and finally, validation using the precision sampling ground measurements.

The rest of this article is organized as follows. The initial section focuses on the data, delivering a comprehensive overview of the data used in this study. Subsequently, the methods section details the preprocessing steps, the development of models, and the various post-processing techniques implemented. Following this, the results and discussion section outlines and examines diverse validation cases with and without refinement via ground measurements to assess the model performances. The subsequent discussions explain the practical implications of the findings. Finally, Section V concludes this article.

II. Data

This section presents the datasets used in this study, incorporating information regarding climate, topography, and soil taxonomy, which significantly influence soil properties [47], [48], [49].

TABLE I
SUMMARY OF THE DATASETS USED FOR ANALYSIS, CONTAINING DATA PROPERTIES, INCLUDING ACCORDING TO SOURCES, RESOLUTION, SAMPLING DEPTH, SAMPLING PERIOD, AND SAMPLE COUNTS FOR EACH SOIL PARAMETER

Name	Reference	Resolution	Sampling depth	Sampling period	pH	Clay	Silt	Sand
ITACyL	[51]	Point	0–30	2000–2021	20919	23021	23019	23022
LUFA Nord-West	Personal contact	Field	0–30	1995–2015		107	107	
LUCAS	[52]	Point	0–20	2009–2018	19498	18620	18620	18620
NABODAT	[53]	Point	0–30	2000–2017	8217	5621	5559	3393
OpenAgrar	[54]	Point	0–10	2011–2018	3094	3028	3028	3028
ISRIC	[55]	Point	0–10	2000–	15927	21108	12781	9806
ISMN	[56]	Point	0–30	2017–		242	242	242
Total					67655	71747	63356	58111

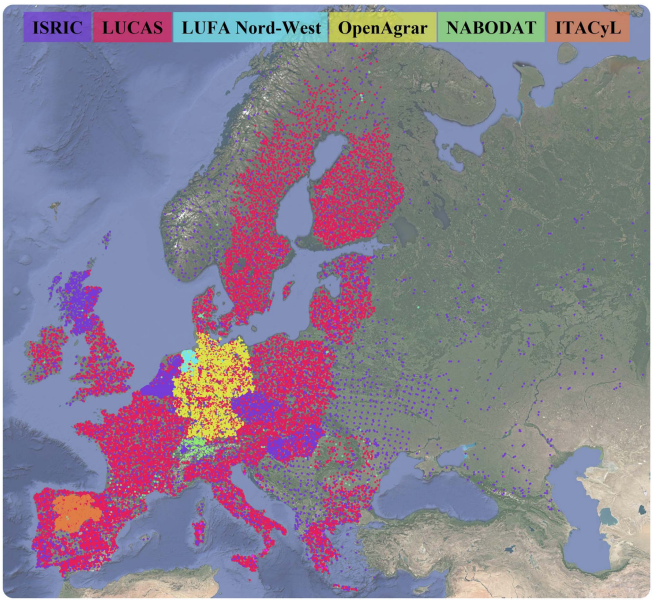

Fig. 1. This study used available agricultural soil property data (n = 100611), which are colored according to their sources, as in purple for ISRIC, pink for LUCAS, blue for LUFA Nord-West, yellow for OpenAgrar, green for NABODAT, and orange for ITACyL.

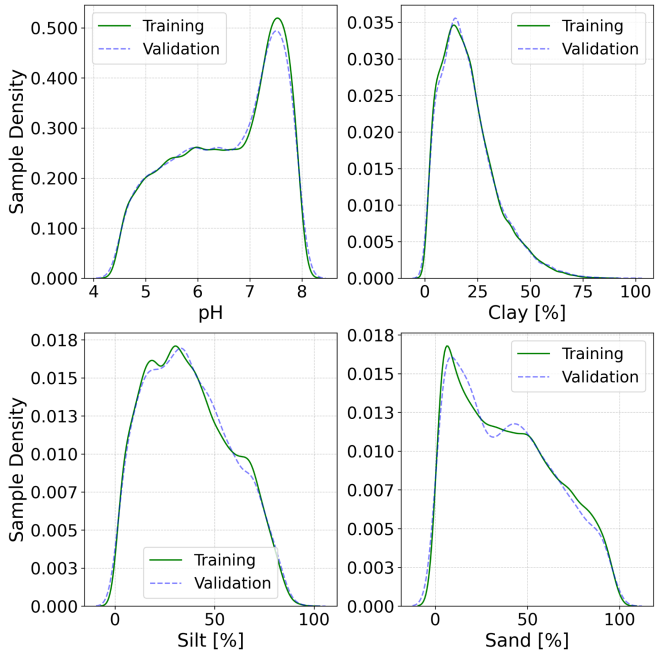

Fig. 2. Distributions of the soil parameters used to develop the prediction models shown for the training and validation dataset.

A. Ground Data

1) Available Soil Databases: To facilitate the spatial modeling of soil properties, we collected ground data from various openly available sources and put it into a consolidated database. Emphasizing uniformity, we standardized the units for each parameter individually. The data were collected from diverse European locations, as illustrated in Table I and Fig. 1. Table I offers details on resolution, sampling depth, data source reference, and sampling period for the topsoil's pH, clay, silt, and sand. For pH, we used the points with the $CaCl_2$ extraction method [50] for pH values. However, the measurement methods for the soil texture parameters were not given for all samples, so all available soil texture measurements were used in the development without applying a method filter.

In Fig. 1, spatial distributions of the sampling locations are given. Specific datasets such as ITACyL [51], LUFA Nord-West, NABODAT [53], and OpenAgrar [54] represent their respective countries, exhibiting concentrated sampling in regions like northern Spain, Germany, and Switzerland. In addition, the LUCAS [52] dataset covers a substantial expanse of Europe, encompassing countries such as France, Great Britain, Italy, Greece, Czechia, Poland, Slovakia, Austria, and Denmark, thereby contributing to a comprehensive repository of soil property measurements.

Fig. 2 shows the histograms of the target parameters, namely, pH, clay, silt, and sand, for training and validation datasets. The preliminary exploration of the data reveals that the pH distribution appears left-skewed, ranging from 3.3 to 8.1, with increased population density observed between 6.2 to 7.8, aligning with the prevalence of mildly acidic to neutral soils previously reported in Europe [57]. Besides, soil texture parameter histograms display a right-skewed distribution toward higher values for clay and silt, while sand demonstrates a relatively uniform distribution. These findings underscore most specific soil types within the studied regions and provide crucial insights into the overall soil composition.

When it comes to soil texture, using ternary plots is a common practice to understand the distribution of the soil class. In Fig. 3, it is evident that samples are scarce in the clay, silt, and sandy

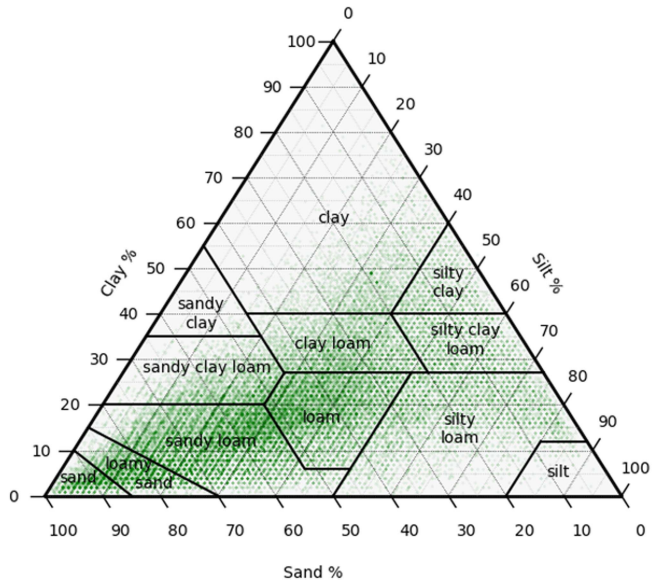

Fig. 3. Ternary distribution of soil texture ground measurements indicating the soil type classification according to the classes presented in unified soil classification system (USCS) [58].

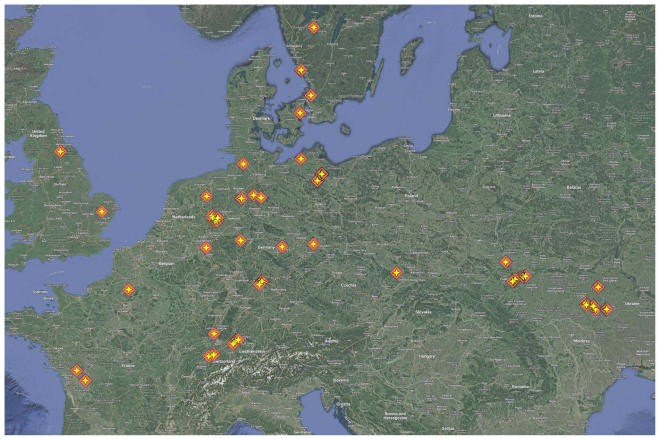

Fig. 4. Locations of 636 fields sampled with Precision Sampling for pH and soil texture. Points in the near vicinity are combined for simplicity in the visualization.

clay classes. This absence could lead to a reduced prediction capability in the corresponding soil classes, particularly in the lack of ground measurement support. This highlights the importance of ensuring a balanced representation of soil types for robust predictions across all classes.

2) Precision Sampling Data: We collected soil samples from 636 European fields using Precision Sampling [21] for parameter map refinement and independent validation purposes. In each field, validation samples are placed at least 30 m away from the refinement samples, making sure that they are not in close vicinity. These fields were chosen from diverse regions, as shown in Fig. 4, encompassing Great Britain, Germany, Switzerland, Poland, and Ukraine. Moreover, the diverse geographic distribution of these sampled fields ensures varied agricultural landscapes characterized by differing climates, soil textures, topography, and management practices.

B. Satellite Data

In our investigation, we extensively used data from two integral components of the Copernicus program: Sentinel-1 (S1) and Sentinel-2 (S2). These essential satellites are part of the collaborative efforts between the European Space Agency (ESA) and the European Union. The Copernicus program is a crucial Earth observation initiative dedicated to furnishing precise and up-to-date information about various environmental aspects of our planet. This mission is accomplished through a fleet of satellites collectively known as Sentinels, which diligently capture Earth's atmosphere, oceans, and land data.

Sentinel-1 (S1) plays a critical role in our work, leveraging its Synthetic Aperture Radar operating in the C-band with a center frequency of 5.405 GHz. This mission encompasses two satellites, S1a and S1b, collectively facilitating a temporal resolution of six days. Our soil property prediction models extensively relied on the Ground Range Detected data, which was processed by ESA and offered an original spatial resolution of 10 m.

Sentinel-1 stands out for its capacity to capture data using two polarizations: (i) vertical transmission-vertical received (VV); and (ii) vertical transmission-horizontal received (VH). Combined with the satellite's operating frequency, these polarizations render Sentinel-1 data uniquely sensitive to the Earth's surface's physical and dielectric properties. This unique sensitivity significantly enhances its suitability for precise soil property prediction.

In conjunction with Sentinel-1, we incorporated Sentinel-2 (S2), which is the Multispectral Instrument. This highly adaptable instrument captures data across a spectrum of 13 electromagnetic bands, spanning from visible to shortwave infrared (SWIR) wavelengths. Crucially, Sentinel-2 ensures a revisit period of six days, guaranteeing frequent and systematic observations of the Earth's surface. The spatial resolution of these bands varies, with some providing a high resolution of 10 m while others offer a slightly coarser resolution of 20 m. In addition, Sentinel-2 encompasses three bands dedicated to cloud screening and atmospheric corrections, each with a resolution of 60 m. Combining these bands equips Sentinel-2 with detailed and comprehensive soil property prediction capability.

The spectral properties of Sentinel-2 data align seamlessly with the expected spectral signature of soil, as highlighted in prior research [59]. This phenomenon can be attributed to the interplay between incoming radiation and soil particles, resulting in a saturation effect. Notably, near-infrared (NIR) and shortwave SWIR wavelengths are absorbed and scattered by soil particles. This unique interaction limits the additional insights derived about soil composition and properties using these wavelengths. As NIR and SWIR wavelengths are absorbed and scattered, the soil's response becomes less sensitive to further changes in the electromagnetic spectrum. Essentially, this saturation effect arises because soil particles efficiently absorb energy at NIR and SWIR wavelengths, and subsequent increases in energy do not yield significant changes or enhancements in the measurement of soil characteristics. The basic statistics of the satellite data-driven features are given in Appendix A.

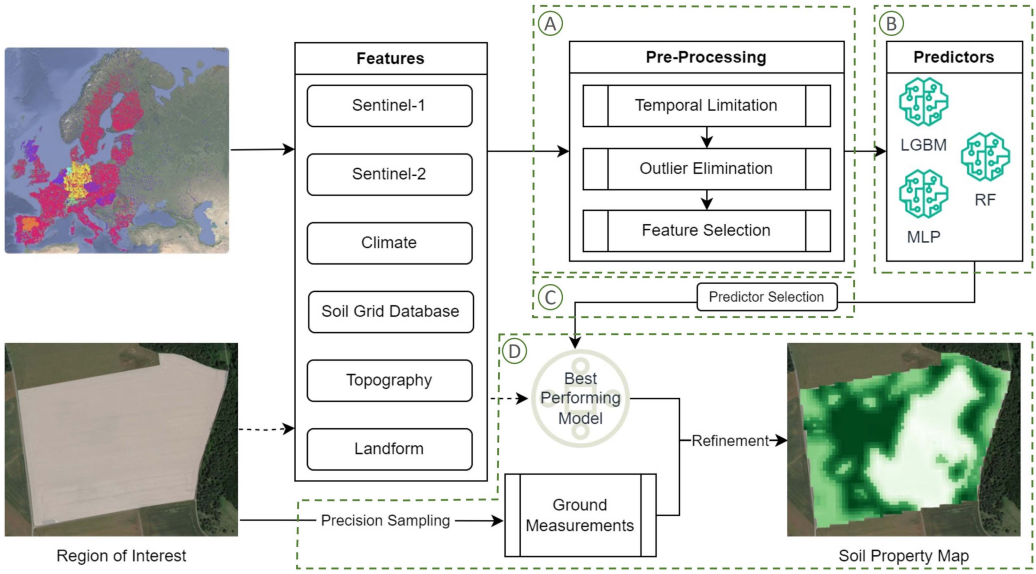

Fig. 5. Methodology followed in developing the presented soil property prediction models shows four interconnected steps in frames labeled accordingly.

C. Climate Data

The climate data used in this study was obtained through the Meteomatics API [60]. This process requires API requests, where we provide the precise latitude and longitude coordinates of soil sampling locations. The acquired data is particularly informative, encompassing daily temperature and precipitation metrics. Our focus centered on the mean and standard deviation of temperature and the mean and cumulative precipitation spanning 2017 to 2023. These parameters gave us a comprehensive overview of the climate conditions in the regions where our soil sampling efforts were concentrated. The basic statistics of the climate data-driven features are provided in Appendix A.

D. Soil Grid Database

To enhance the precision of our predictions, we utilized the 250 m soil property maps obtained through the Soilgrids REST API [61], which provided a low-resolution overview. These maps played a critical role in guiding our predictions. Our methodology involved supplying the API with latitude and longitude coordinates for various locations. Subsequently, we extracted the surface soil texture properties, including essential parameters such as clay, silt, and sand, from the API's response. By integrating this supplementary data, we enhance our predictive models' accuracy and reliability, strengthening our models' overall robustness. The basic statistics of the soil-grid data-driven features are given in Appendix A.

E. Topography Data

We obtained our topography data from the Copernicus Global 30 m digital elevation model (DEM), which we accessed through the Google Earth Engine using the COPERNICUS/DEM/GLO30 identifier. This dataset provides a spatial resolution of 30 m and offers detailed insights into the elevation variations across our sampling locations. In addition to fundamental elevation data, we extracted the topographical slope from this DEM. Integrating elevation and slope data helped us understand the terrain and hydrological characteristics. The basic statistics of the topography data-driven features are given in Appendix A.

III. Methods

This article focuses on soil property mapping, achieved through ML model predictions and ground measurements obtained from agricultural fields. To accomplish this, our soil property mapping approach encompasses four key components, as depicted in the four interconnected stages that constitute our comprehensive approach to mapping soil properties, ensuring our predictions are robust and reliable.

The proposed methodology shown in Fig. 5 consists of the following steps after the feature collection: Data preprocessing, training different model frameworks, selecting the best-performing model, and finally, refinement and validation of the produced maps.

1) *Data preprocessing:* In this initial step, we investigate our data. This involves setting temporal boundaries, removing outliers, and keeping pertinent features from our input datasets via correlation analysis. Doing so ensures that the data used for modeling is high quality and directly relevant to our objectives.
2) *Training different model frameworks:* We train three models using the preprocessed data. These models include the LGBM, RF, and MLP. The aim here is to explore various modeling approaches and measure their effectiveness in predicting soil properties at the pixel level.
3) *Selecting the best-performing model framework:* Following the training phase with different frameworks, we assess the performance of these models. We employ a

mean absolute percent error (MAPE) and coefficient of determination (R^2) evaluation metrics and performance criteria to identify the model framework that accurately and reliably predicts the soil property. The top-performing model framework, having the lowest MAPE and highest R^2, is chosen for the next step.

4) The selected model framework is then employed to generate initial soil property maps. These maps undergo further refinement and validation using the precision sampling approach and dataset. These ground measurements are used to fine-tune and validate the predicted soil property values, improving the accuracy and precision.

A. Preprocessing

1) Temporal Data Selection: Specific soil parameters, namely pH, clay, silt, and sand, are temporally stable and exhibit minimal temporal variation. We implemented a robust filtering process to maintain the relevance and consistency of the soil data used in our model development. Specifically, we included only measurements obtained after 2000, while data collected before that date were excluded.

This temporal filtering process ensured that our model used the most relevant and current pH and soil texture data available. By doing so, we aimed to strengthen our model's precision and practicality, relying on the latest and most pertinent information to make accurate predictions and informed assessments of soil properties.

2) Outlier Elimination: To ensure the quality and reliability of our input data, we executed an outlier elimination during the preprocessing phase. Our approach was methodical: We investigated the distribution of each feature concerning each target parameter independently. We singled out samples that fell within the top and bottom 1% percentile range for each feature as potential outliers. Furthermore, any sample flagged as an outlier in at least 80% of the features was removed.

This method effectively identified and eliminated outliers from our dataset. By excluding extreme values that could potentially adversely affect our analysis and modeling results, we ensured the integrity of our data. Following this outlier elimination process, we observed that the parameters' distributions began to tend to a Gaussian distribution with slight skewness. This characteristic is particularly advantageous for feature scaling, as it contributes to the smooth and effective convergence of ML algorithms, ultimately enhancing the reliability of our model.

3) Feature Selection: The feature selection was carried out using a linear correlation analysis, which enables us to assess the linear relationships between individual parameters and their interactions. Our goal was to comprehensively understand the influence of each parameter in a linear context. We conducted correlation analysis across all the features present in our dataset. We paid particular attention to highly correlated features during the feature elimination process. Specifically, we identified features with an absolute correlation coefficient exceeding 0.9. From this subset of highly correlated features, we retained the one that exhibited the strongest correlation with each target parameter. The selected features for each model are given in the tables of Appendix A. Our intention behind this approach was to identify the most pertinent and informative features for our modeling objectives. This process serves the dual purpose of reducing the dimensionality of our dataset while selecting the features that have the greatest influence in predicting soil properties accurately.

B. Model Development

In the evaluation phase, we systematically assessed the performance of different ML frameworks for each target parameter. We employed a methodology that involved training the models on a subset of the data and validating their performance on an independent set of samples. We analyzed the results of various ML algorithms to identify the most suitable framework for accurately predicting each soil property.

1) ML Frameworks: Assessing various ML frameworks is critical in effectively mapping soil properties using satellite data. We selected three frameworks from established research and implementations in our evaluation, considering each framework's distinctive features and advantages [62], [63]. The algorithms are implemented in the widely used *scikit-learn* package [64] and include:

1) *LGBM:* LGBM operates as a gradient boosting framework that relies on tree-based learning algorithms. Its efficiency in both training and prediction is noteworthy. LGBM uses histogram-based algorithms and supports parallel computing, rendering it adept at handling large datasets. With an array of customizable hyperparameters, we could fine-tune its performance according to our specific requirements [65].
2) *RF:* RF operates as an ensemble learning method that combines multiple decision trees to generate predictions. By aggregating results from various models, RF enhances accuracy and robustness. Furthermore, it offers insights into feature importance, facilitating the feature selection process [66].
3) *MLP:* MLP functions as an artificial neural network comprising multiple layers of interconnected neurons. It excels in discerning complex patterns within data and making predictions through feed-forward propagation. MLP models can be trained with or without backpropagation, allowing for optimization of their performance based on specific requirements [67].

These algorithms offer different and diverse modeling approaches, each with unique strengths in capturing complex relationships between features and soil parameters. We aim to evaluate these algorithms to determine which can most accurately account for the observed feature variations with the highest precision. Subsequently, we select the best-performing model to proceed with the subsequent steps. This evaluation process entails a comprehensive assessment and comparison of each algorithm's performance using suitable metrics and validation techniques.

2) Data Split: The data splitting step in ML model development is a pivotal stage for assessing the model's performance and capacity to generalize effectively to new, unseen data. This step involves strategically dividing the dataset into two different subsets: 1) the training set; and 2) the validation set.

The training set typically encompasses around 80% of the entire dataset. This substantial portion is the foundation upon which the model is built and fine-tuned. The model is exposed to this training set during training, aiming to learn the intricate relationships and patterns between the input features and the target soil property. The model iteratively adjusts its hyperparameters to minimize prediction errors, acquiring a deep understanding of how input features influence the target soil property. This understanding forms the basis for its subsequent predictions.

C. Selecting the Best Performing Model

In contrast, the validation set accounts for the remaining 20% of the data. Its significance lies in its independence from the training set; validation points should not overlap with the instances used for model training. This independence ensures that the model's performance on the validation set is an unbiased evaluation of its predictive capabilities. We gain valuable insights into its robustness and generalizability by subjecting the model to the validation set. We assess whether the model can make accurate predictions under unfamiliar conditions, a crucial aspect of its overall reliability.

1) Hyperparameter Optimization: Hyperparameter optimization is a crucial phase in the development of ML models. It thoroughly searches for the ideal combination of hyperparameters to optimize a model's performance with a given dataset. Hyperparameters are essential configuration settings that must be defined for each model. In this study, we employ the Hyperopt package [68], a powerful Python library that has gained widespread recognition for its flexibility and efficiency in hyperparameter tuning. It uses Bayesian optimization, designed to systematically explore the hyperparameter space, which essentially represents a multidimensional parameter space containing various hyperparameters, to pinpoint the most advantageous configuration.

The process of hyperparameter optimization starts with the definition of the objective function. This function quantifies the model's performance based on specific criteria. Our study aims to minimize the validation error between the measured and estimated values (MAPE) while maximizing the R^2. Then, we initiate the iterative exploration, where the Hyperopt package selects an initial configuration of hyperparameters. The Bayesian optimization algorithm uses this initial configuration to evaluate the model's performance on the validation data. This is followed by selection, which learns from previous evaluations and leverages this knowledge to decide which configurations to explore next. This iterative cycle continues until the algorithm converges to an optimal hyperparameter configuration. The optimization aims to find the combination that yields the best possible model performance according to our predefined objectives.

2) Model Performance Assessment Metrics: The model performance assessment was conducted based on the validation scores. To assess the accuracy of the models, three metrics were employed: Root Mean Squared Error (RMSE), the coefficient of determination (R^2), and MAPE.

RMSE measures the average magnitude of the errors between predicted and measured values, calculated by taking the square root of the average of the squared differences between predicted and measured values. The coefficient of determination R^2 measures how well the model explains the variance. It ranges from 0 to 1, with a higher value indicating a better model fit to the measured data. MAPE, given in (1), assesses the absolute percentage error between predicted and measured values, while lower values indicate better prediction.

$$\text{MAPE} = \left| \frac{\text{Prediction} - \text{Measurement}}{\text{Measurement}} \right| \times 100. \tag{1}$$

In the best model selection, the algorithm that achieved the highest validation R^2 and lowest RMSE was selected as the final model by comparing the validation scores of the different models. A lower RMSE indicates a better fit between the predicted and actual values, indicating higher predictive accuracy and better model performance in capturing the variations in target parameters.

For a more rigorous assessment of the model's accuracy, we calculated the MAPE in 50%, 80%, and 90% confidence intervals (CI). To do so, we divided the entire parameter range into 20 equidistant intervals. We systematically filter the data points for each interval, enabling us to calculate the percentage errors between the model's predictions and the actual measurements. Later, we compute the error distribution's 50th, 80th, and 90th percentiles for each parameter range.

D. Final Soil Property Mapping

This step employs a refinement process to enhance the accuracy and detail of the predicted soil property map. Support sampling is particularly crucial in geospatial analysis, where values at unmeasured locations need to be estimated based on the values observed at sampled locations. Considering the parameter's temporal changes, particularly for pH, the inclusion of ground measurements conducted at a close time to prediction solves the problem caused by variation resulting from the parameter's dynamic nature prior to planned precision management.

In our approach, we use the Precision Sampling [21] points taken for the refinement and left out the validation points in the field [46]. The refinement starts with thin-plate spline interpolation [69]. It is a sophisticated 2-D interpolation technique to fill in the gaps between sampled soil measurements. It is a popular choice in geostatistics [70] due to its ability to generate a continuous surface from scattered data points. It is like creating a smooth, continuous map of soil properties.

The next step involves smoothing for noise reduction, similar to applying a box-car filter to the soil property map. Imagine examining a pixel on the map and considering the values of

its neighboring pixels within a 3×3 grid, essentially covering a 30× 30 m area. By averaging or smoothing these values, we effectively reduce any noise or irregularities that might have arisen during the interpolation process. This smoothing operation ensures that our final map more coherently represents the spatial variation of soil properties.

After mapping the soil properties, we conducted internal and external model validations, during which we assessed the models' prediction performances by comparing ground measurements to predictions with and without precision sampling refinement.

IV. Results and Discussion

This section outlines the comprehensive evaluation of our ML models. It is divided into four key subsections as follows.

A) Data preprocessing that shows the number of samples that are used for model training and validation as well as remaining features after the selection.

B) Internal model assessment investigates the evaluation of the model's performance using the same data sources that were used for training (as detailed in Section II-A1). It tests how well the models perform when assessed against data they are already familiar with regarding the source. This is crucial in understanding how effectively the models capture the underlying patterns and relationships within the training data source.

C) Validation without precision sampling explores how the models generalize to entirely new and unseen data collected separately by AgriCircle. This provides valuable insights into the models' ability to make accurate predictions in real-world scenarios beyond their training data.

D) Validation with refinement presents the results of model performance on validation data (as explained in Section II-A2), but with the added benefit of Precision Sampling-based refinement using the samples collected by AgriCircle. It shows the models' performance when fine-tuned with additional ground measurements, showcasing their adaptability and refinement capabilities. Ultimately, this thorough assessment ensures a clear picture of how well the models perform across different scenarios and data sources, validating their effectiveness in predicting soil pH and texture.

A. Data Preprocessing

The original pH dataset has 67 655 samples. After the elimination of outliers, the pH prediction model was reduced to include 59 797 samples, of which 47 838 were used for training and 11 959 for validation. Through feature selection, the model retained 21 features. These features encompass VV mean and standard deviation, the mean value of B1, B4, B6, B9, B11, and B12, the standard deviation of B9, B11, and B12, means of temperature and precipitation, as well as the standard deviation of temperature, clay, silt and sand, and landform, taxonomy, and elevation.

In the clay model framework, the original dataset comprises 71 747 samples. Upon eliminating outliers, the clay prediction model incorporates 63 367 samples, with 50 694 allocated for training and 12 673 for validation. Through feature selection, the model retains 22 features. These encompass VV and VH mean, the standard deviation of VH, mean values of B1, B2, B4, B6, B9, B11, and B12, and standard deviations of B9, B11, and B12, mean temperature and precipitation, clay, silt, and sand, landform, taxonomy, elevation, and slope.

In the silt model framework, the original dataset contains 63 356 samples. Following the removal of outliers, the model integrates 53 535 samples, with 42 828 designated for training and 10707 for validation. After feature selection, the model preserves 23 features. These encompass VV and VH mean, along with the standard deviation of VH, mean values of B1, B6, B9, B11, and B12, as well as standard deviations of B9, B11, and B12, mean temperature and precipitation and temperature standard deviation, clay, silt, and sand, landform, taxonomy, elevation, and slope.

In the sand model framework, the original dataset consists of 53 111 samples. After the elimination of outliers, the model incorporates 44 390 samples, with 35 512 allocated for training and 8878 for validation purposes. The model maintains 21 features. These encompass VV and VH mean and standard deviation, mean values of B1, B2, B6, B11, and B12, as well as the standard deviations of B1, B9, B11, and B12. In addition, mean temperature and precipitation, temperature standard deviation, clay, silt, and sand, and elevation are retained.

The predictive models for pH, clay, silt, and sand content share similarities in the features used for refinement and prediction. Each model incorporates a diverse array of spectral indices derived from satellite imagery, including the mean of VV, B1, B11, and B12 and standard deviation values of various bands such as B1, B9, B11, and B12. In addition, climate variables, such as mean temperature and precipitation, are included. Low-resolution soil data for clay, silt, and sand content, as well as terrain characteristics like elevation, are also retained across all models. This consistent feature selection approach ensures the models capture the complex relationships between environmental predictors and soil properties, seemingly contributing to accurate predictions and enhancing our understanding of soil dynamics at a spatial scale.

The existing research and physical reasoning also support these selected features. As they are basically the same information at a lower scale, the soil texture data from the soil grid database is an expected set of features. When it comes to climate data, [10] shows the importance of climate variations on the distribution of soil properties.

B. Internal Model Assessment

In this section, we perform a comparative analysis of the three distinct modeling frameworks explained in Section III-B1. The results of these frameworks are summarized in Table II. Each column group corresponds to a model framework, while each row presents the evaluation results for both the training and validation datasets for pH, clay, silt, and sand. In the Appendix, we provide further information regarding the model performances in Figs. 11–14.

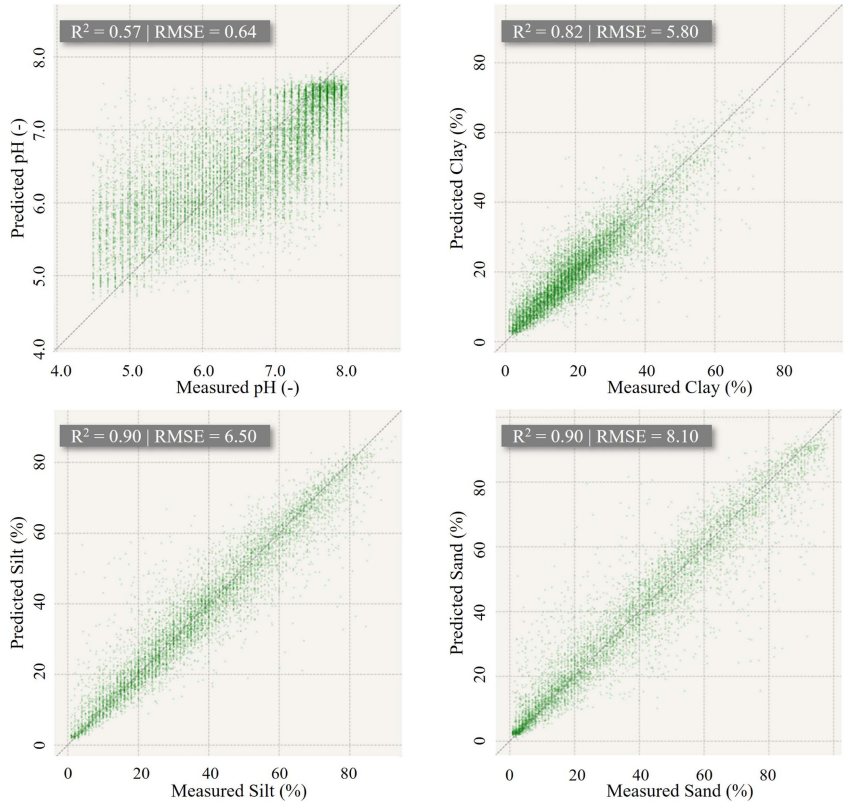

Fig. 6. Scatter plots between measured and predicted pH, clay, silt, and sand values using the best performing model framework for their validation data sets.

TABLE II
MODEL ASSESSMENT RESULTS FOR PH, CLAY, SILT, AND SAND FOR R^2 AND RMSE UNDER DIFFERENT FRAMEWORKS AFTER HYPERPARAMETER OPTIMIZATION

		LGBM		RF		MLP	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
pH (-)	Training	**76.97**	**0.47**	69.86	0.54	57.97	0.64
	Validation	**56.81**	**0.64**	54.77	0.66	50.50	0.69
Clay (%)	Training	**87.22**	**4.90**	86.73	4.90	81.48	5.80
	Validation	**81.74**	**5.80**	81.37	5.80	80.27	6.00
Silt (%)	Training	**96.55**	**3.90**	93.48	5.40	90.48	6.50
	Validation	**90.23**	**6.50**	89.94	6.60	89.25	6.80
Sand (%)	Training	91.92	7.50	**93.78**	**6.60**	90.24	8.20
	Validation	**90.29**	**8.10**	90.39	8.10	89.47	8.40

The best-performing frameworks are highlighted in bold.

We find the LGBM model consistently as the top-performing model for all four soil parameters II. The LGBM model demonstrates R^2 higher than 76% for all soil parameters in training, underscoring its capability to explain a substantial portion of the variability within them. As we investigate the validation, we see R^2 higher than 81% for soil texture-related parameters and 56% for the pH model. The high R^2 values in the validation set indicate the robust generalization ability of the LGBM model to new and unseen data, reaffirming its superior predictive power and reliability.

For instance, for pH prediction on the training set, the LGBM model achieves an R^2 value of 76.97% and an RMSE of 0.47. However, it is important to note that with this R^2 value, approximately 23% of the pH variability remains unexplained. This suggests that certain factors, such as unaccounted features, measurement errors, or inherent pH variability, persist beyond the scope of the selected features. The validation (see Fig. 6) reveals a specific trend: The models consistently overestimate pH values below six and underestimate them above 7.5. This can be attributed to the intricate relationships between the input features and pH and the tendencies of modeling frameworks to converge toward the mean. It's worth noting that the presence of data points with very similar values, often differing only by a small decimal point, emphasizes the challenges associated with precise soil property prediction at the pixel level.

We observe a minimum R^2 value higher than 80% for the tested clay, silt, and sand prediction frameworks, demonstrating higher robustness and generalization capabilities than the pH model. Here, we should mention that RF showed higher accuracy for sand than LGBM in the training dataset. However, LGBM showed a slightly higher R^2 and identical RMSE, which led us to proceed with the LGBM framework for consistency across the framework.

Fig. 6 presents scatter plots for the validation prediction using the LGBM model against the measured values. These scatter plots show the fit between predicted and measured values, providing insights into the model's performance. The achieved R^2 values of approximately 90% on the validation dataset, along with RMSE values of 0.64 and 5.8 for Silt and Sand, respectively, suggest good performance exhibited by the selected model. Specifically, these metrics imply that the model effectively captures the variance in Silt and Sand content, as evidenced by the high R^2 values. Moreover, the model demonstrates considerable

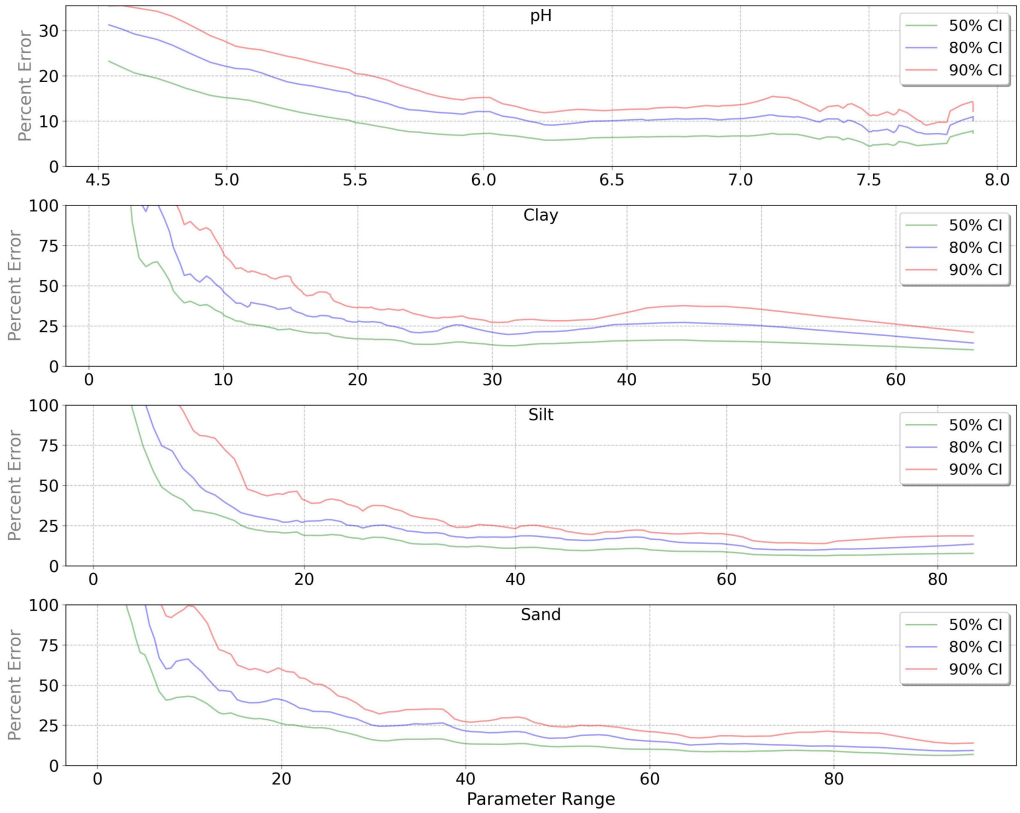

Fig. 7. Confidence intervals in MAPE for 50%, 80%, and 90%, calculated at different value ranges of pH, clay, silt, and sand.

accuracy for Clay, achieving an R^2 score of 81.74% alongside an RMSE of 5.8 were attained. This suggests the model can provide precise predictions for Clay content, indicating its utility in soil composition analysis. However, pH yields poor results, with an R^2 score of 56.1% and an RMSE of 0.64. Conversely, the model's performance is less favorable when assessing pH, as evidenced by an R^2 score of 56.1% and an RMSE of 0.64.

Fig. 7 illustrates the behavior of CI curves concerning prediction errors across different parameter values. A notable pattern emerges: As the parameter value increases, the MAPE, as defined in (1), tends to decrease. This pattern can be elucidated by considering the nature of the MAPE metric. At lower values of clay, silt, and sand content, the measurements approach zero, implying that any prediction higher than this point would result in a disproportionately high MAPE. This phenomenon gives rise to an asymptotic behavior in the confidence curve, characterized by a sharp decline in prediction error as parameter values increase. This observation highlights a critical aspect of the model's performance. It underscores that the model's predictions are less accurate than the ground truth measurements when dealing with low clay, silt, and sand content values. The model's prediction accuracy improves as these soil properties increase, leading to lower prediction errors and narrower confidence intervals.

Comparison of Figs. 2 and 7 provides a comprehensive view of the CI discussions. In Fig. 2, we observe that the 50% confidence intervals consistently maintain a relatively low prediction error, staying below 20% for all considered parameters. The validation results remain favorable when adopting a more stringent criterion, as represented by the 90% confidence intervals. For pH, the error remains around 15%, while for clay, it is less than 40%. Silt and sand parameters exhibit approximately 25% error within the 90% confidence interval. This stable pattern observed in the confidence curves across most parameter ranges reflects the robustness and reliability of the corresponding model. It suggests the model's predictions are generally consistent and accurate within the specified ranges, providing valuable insights into soil properties such as pH, clay, silt, and sand content.

C. External Model Assessment

After completing the internal validation assessment using the data points represented in Fig. 1, our analysis transitioned to evaluating the model's performance using a distinct set of data points collected from agricultural fields by AgriCircle, whose locations are presented in Fig. 4. The dataset consists of 554 fields with 2909 samples for pH, 234 with 1130 samples for clay, 234 with 1128 samples for silt, and 216 with 1038 samples for sand.

Using data collected by AgriCircle via Precision Sampling, we conducted a two-step validation. In the first step, we assessed the model's predictive capabilities on this new unseen set of ground measurements. In the second step, we split this set of data into two parts, with one part designated for the points to be

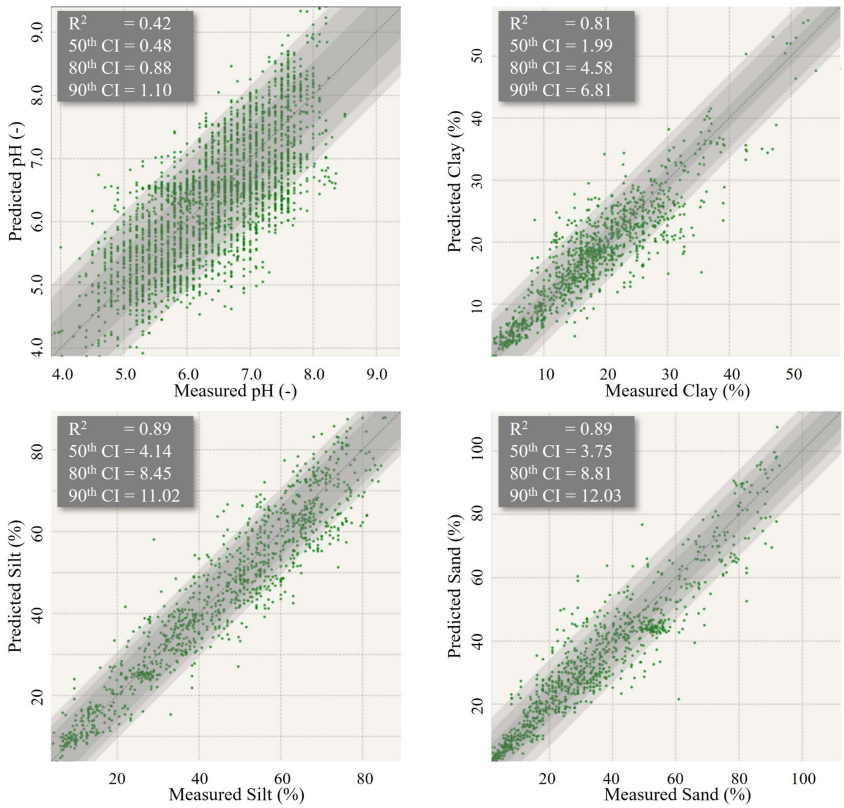

Fig. 8. Scatter plots between measured and predicted pH, clay, silt, and sand values on the Precision Sampling points before the ground measurement-based refinement. 50%, 80%, and 90% CIs are presented with bands in each subplot.

used for refinement, which were collected according to Precision Sampling, and the other part used for validation purposes, ensuring a distance of at least 30 m from the refinement samples.

1) Validation Without Precision Sampling: The validation analysis conducted in this phase serves as a critical step in understanding the performance and reliability of the trained models, shedding light on their effectiveness in accurately predicting soil properties across the designated geographical locations as shown in Fig. 8. Notably, the observed shifts in the R^2 values for the pH parameter require closer examination, revealing a decrease from 56.81% in the internal validation to 41.74% during the external validation process. This decline can be attributed to several factors, including the possibility of overfitting, inherent variability in the data sources employed, potential discrepancies in the spatial distribution of the training data, and the complex interplay of various environmental factors influencing pH levels at different geographical locations.

Furthermore, focusing on the evaluation of the soil texture parameters, we witnessed a reduction in the R^2 values, although within a relatively narrow range, from 81.74% to 80.73% for clay, from 90.23% to 89.41% for silt, and from 90.29% to 88.88% for sand. These marginal shifts underscore the complex nature of soil composition, emphasizing the inherent challenges associated with precisely predicting soil texture properties solely based on remote sensing data and associated feature sets.

While the initial validation accuracy is higher than 80%, particularly for the soil texture parameters, precision agriculture demands a higher level of accuracy, greater than 90%, to ensure effective and informed decision-making. Recognizing the critical need for improved precision, our approach underscores the statistical integration of ground measurements into the predicted soil property maps. This strategic integration process is a vital enhancement mechanism, facilitating the refinement of the predicted maps by applying the Precision Sampling methodology. By employing this approach, we aim to bridge the gap between the initial predictions and the required accuracy levels necessary for optimal precision agriculture practices and informed decisions in field management.

2) Validation With Precision Sampling: The preceding validation phase has emphasized the potential for enhancing the precision of pH, clay, silt, and sand predictions within agricultural fields. Fig. 9 provides valuable insights into the outcomes of validation points obtained from the dataset outlined in Section II-A2. This dataset leverages Precision Sampling pH, clay, silt, and sand measurements to fine-tune the predicted soil property maps through thin-plate spline interpolation. Notably, the validation points within agricultural fields maintain a minimum separation of 30 m from the refinement points, ensuring their independence. The scatter plots in Fig. 9 enable a comprehensive assessment of the adjusted soil property maps with their accuracy and reliability.

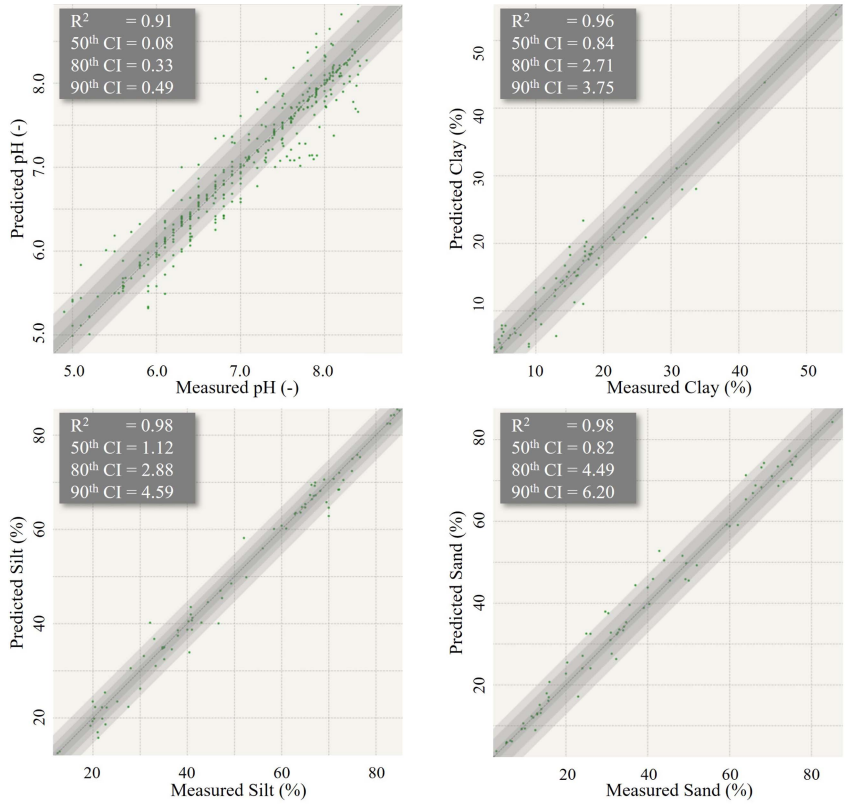

Fig. 9. Scatter plots between measured and predicted pH, clay, silt, and sand values on the validation points after the ground measurement-Precision Sampling based refinement. 50%, 80%, and 90% CIs are presented with the grey bands in each subplot.

Regarding pH predictions, the refined model achieves a R^2 value of 91.42% at the validation points, underscoring a robust correlation between the predicted pH values and the ground measurements. The low 50% CI value of 0.08 suggests relatively minor errors well within the confines of laboratory uncertainty. These results demonstrate the effectiveness of the refinement process in accurately estimating pH content facilitated by interpolation and smoothing techniques. For clay, silt, and sand, the R^2 values exceed 95%, accompanied by 50% CI values of 0.84, 1.12, and 0.82, respectively. When extending our analysis to the 90% CI, we observe an absolute error rate of 0.49 for pH and less than 6.2 for soil texture parameters.

Integrating ground measurements through the Precision Sampling approach is critical in improving the precision and reliability of soil property maps. By this approach, the refinement process effectively reduces the uncertainty associated with the predicted values, thereby enhancing the suitability of these maps for informing precise and targeted agronomic strategies. Even at a fine spatial resolution, this increased accuracy enables more informed decision-making and promotes optimized agricultural practices, contributing to improved crop yield and sustainable land management.

The high-resolution soil maps obtained through the refined methodology offer a detailed and comprehensive understanding of soil properties within agricultural fields, facilitating various precision agricultural practices such as variable rate application for sowing density, fertilization, and irrigation [71], [72], [73]. The practical examples depicted in Fig. 10 showcase four distinct agricultural fields across Europe. Each showcase represents a soil property map after postrefinement. It illustrates the enhanced and precise delineation of the respective soil property, empowering farmers and agronomists with valuable insights to optimize their land management strategies and improve overall agricultural efficiency and sustainability.

D. Practical Implications

In Fig. 10(a), the 46-hectare field in Ukraine highlights the importance of the presented pH map applicable for optimizing the lime application, likely improving crop productivity and yield [74] as well as soil structure and stability [75]. In Fig. 10(b), the 187-hectare field in Germany demonstrates the generated clay map with the aim of understanding the field's water holding capacity [76], [77], facilitating informed water management [78] and field drivability [79], [80] decisions. Similarly, the field depicted in subfigure 10-c, covering 32 hectares in Hungary, benefits from the calculated silt map, empowering the farmer in his aim to create a variable sowing rate map [81]. Lastly, Fig. 10(d) presents a 4-hectare field in England, where the sand, clay, and silt maps aid the landowner in efficiently managing irrigation activities [78], [82].

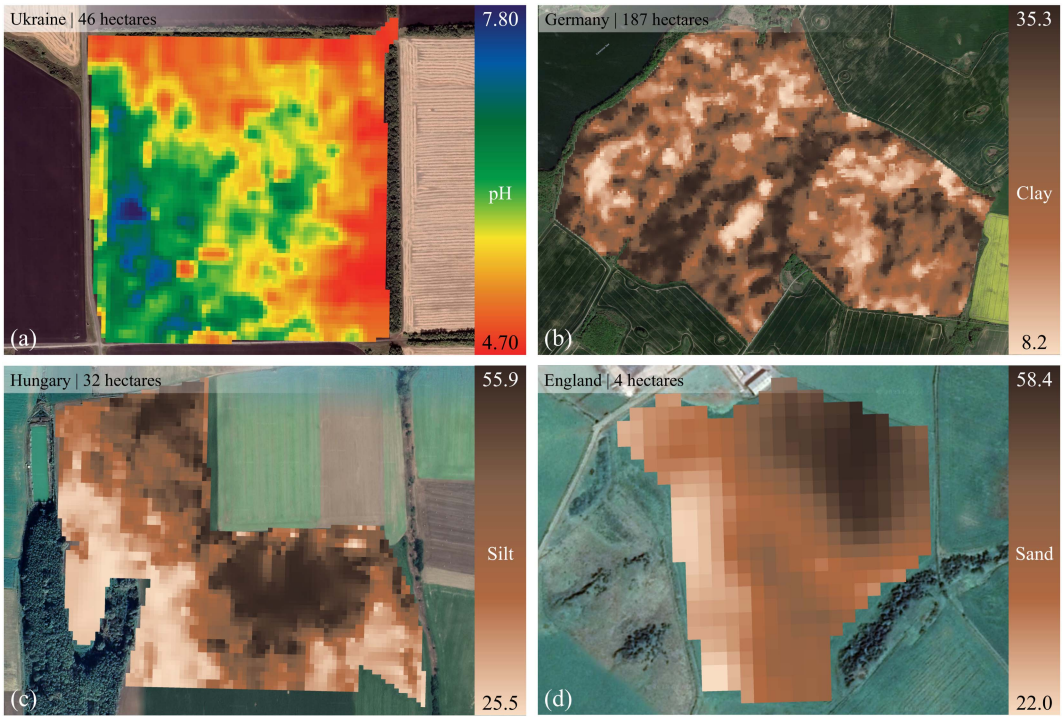

Fig. 10. Examples maps for predicted pH, clay, silt, and sand on various locations: (a) pH map of a field in Ukraine has an area of 46 ha, (b) Clay map of a field in Germany has an area of 187 ha, (c) Silt map of a field in Hungary has an area of 32 ha, and (d) Sand map of a field in England has an area of 4 ha.

As stated in this section, this research's practical implications cover a broad range of agricultural applications. Although soil sampling makes the presented approach more costly than methods that rely on remote-sensing data, the accuracy achieved makes the maps with support sampling superior.

V. Conclusion

This research is dedicated to developing and implementing an innovative algorithm that predicts soil texture and pH content in European soils using satellite, climate, and terrain data. Multispectral and radar imagery offer insights into soil spectral properties, complemented by climate data (temperature, precipitation) and terrain features (elevation, slope, aspect, hillshade). This integration allows the algorithm to navigate the intricate relationships between environmental features and target parameters. Training on a European dataset from various locations, the LGBM algorithm yields predictions with an average percentage error below 6% at the European scale when combined with precision-sampled ground measurements. Integration of precision sampling, also called support sampling, significantly enhances overall accuracy.

The framework is validated with independent datasets to ensure robustness, which demonstrates reliable performance across diverse soil types and geographic regions in Europe, both with and without refinement through ground measurements. The presented solution, with the support of ground measurements, carries profound implications for soil management, land use planning, and environmental research. Accurate predictions empower decision-making on soil health, sustainable land management, and optimization of fertilization and soil moisture practices. In the following research phase, our focus shifts to expanding predictive capabilities into soil moisture mapping. This extension holds promise for supporting irrigation and field drive-ability, enriching the toolkit for agriculture and environmental decision-makers.

Acknowledgment

The authors would like to thank Rodrigo Principe and Utku Berkalp Ünalan for setting up the Google Earth Engine framework to download the satellite data and Rene Dechow and Axel Don for their help in providing access and downloading model data for the OpenAgrar measurements.

Appendix A
Feature Statistics

In the appendix tables, VV and VH correspond to vertical–vertical and vertical–horizontal polarizations, B# corresponds to the respective band of Sentinel-2, and SGD corresponds to the Soil Grid Database.

Appendix B
Different Model Frameworks

This appendix presents the model training results using different frameworks, including the training and validation split.

TABLE III
FEATURE STATISTICS FOR THE PH DATASET

Parameter	Description	Selected	Min	Max	Mean	Std
pH			4.50	8.00	6.54	0.99
VV	mean	x	−20.16	−0.01	−11.28	1.49
VH	mean		−22.83	−0.12	−18.38	1.44
VV	standard deviation	x	0.00	10.49	2.59	0.64
VH	standard deviation		0.00	20.39	2.53	0.95
B1	mean	x	0.01	0.54	0.10	0.04
B2	mean		0.03	0.55	0.13	0.04
B3	mean		0.04	0.55	0.16	0.04
B4	mean	x	0.06	0.56	0.20	0.05
B5	mean		0.07	0.58	0.23	0.05
B6	mean	x	0.08	0.58	0.26	0.05
B7	mean		0.10	0.58	0.28	0.05
B8	mean		0.09	0.60	0.30	0.05
B8A	mean		0.11	0.58	0.31	0.05
B9	mean	x	0.11	0.75	0.31	0.06
B11	mean	x	0.04	0.64	0.35	0.06
B12	mean	x	0.03	0.58	0.28	0.05
B1	standard deviation	x	0.00	0.40	0.08	0.07
B2	standard deviation		0.00	0.41	0.08	0.07
B3	standard deviation		0.00	0.39	0.08	0.06
B4	standard deviation		0.00	0.39	0.09	0.06
B5	standard deviation		0.00	0.39	0.09	0.06
B6	standard deviation		0.00	0.37	0.09	0.05
B7	standard deviation		0.00	0.35	0.10	0.05
B8	standard deviation		0.00	0.35	0.10	0.05
B8A	standard deviation		0.00	0.37	0.10	0.04
B9	standard deviation	x	0.00	0.55	0.10	0.05
B11	standard deviation	x	0.00	0.37	0.08	0.03
B12	standard deviation	x	0.00	0.40	0.07	0.02
Landform		x	0.00	42.00	30.36	6.42
Taxonomy		x	0.00	434.00	189.80	124.83
Elevation		x	0.00	2497.84	519.79	410.02
Slope			0.00	362.02	149.43	107.10
SGD Clay		x	0.80	76.38	23.54	9.12
SGD Silt		x	0.70	91.66	33.98	15.98
SGD Sand		x	0.30	98.83	42.46	20.58
Temperature	mean	x	−1.06	32.84	12.95	4.68
Temperature	standard deviation	x	0.81	17.09	7.36	2.49
Precipitation	mean	x	0.00	21.45	0.10	0.12
Precipitation	sum		0.51	43500.14	200.51	246.43

TABLE IV
FEATURE STATISTICS FOR THE CLAY DATASET

Parameter	Description	Selected	Min	Max	Mean	Std
Clay			1.00	97.00	21.65	13.56
VV	mean	x	−20.16	−0.01	−11.17	1.58
VH	mean	x	−22.84	−0.12	−18.23	1.47
VV	standard deviation		0.00	10.49	2.56	0.65
VH	standard deviation	x	0.00	20.39	2.51	0.97
B1	mean	x	0.02	0.54	0.11	0.05
B2	mean	x	0.04	0.55	0.13	0.05
B3	mean		0.04	0.55	0.16	0.05
B4	mean	x	0.06	0.56	0.19	0.05
B5	mean		0.07	0.58	0.23	0.05
B6	mean	x	0.08	0.58	0.26	0.05
B7	mean		0.10	0.58	0.28	0.06
B8	mean		0.10	0.60	0.29	0.06
B8A	mean		0.11	0.58	0.30	0.06
B9	mean	x	0.12	0.84	0.31	0.06
B11	mean	x	0.04	0.71	0.33	0.06
B12	mean	x	0.03	0.62	0.27	0.05
B1	standard deviation		0.00	0.40	0.10	0.08
B2	standard deviation		0.00	0.41	0.10	0.08
B3	standard deviation		0.00	0.39	0.10	0.08
B4	standard deviation		0.00	0.39	0.10	0.07
B5	standard deviation		0.00	0.39	0.10	0.07
B6	standard deviation		0.00	0.37	0.10	0.06
B7	standard deviation		0.00	0.35	0.10	0.06
B8	standard deviation		0.00	0.35	0.10	0.06
B8A	standard deviation		0.00	0.37	0.10	0.05
B9	standard deviation	x	0.00	0.55	0.11	0.06
B11	standard deviation	x	0.00	0.37	0.08	0.03
B12	standard deviation	x	0.00	0.40	0.07	0.02
Landform		x	0.00	42.00	30.18	6.45
Taxonomy		x	0.00	434.00	161.99	130.81
Elevation		x	0.00	2499.69	408.75	433.01
Slope		x	0.00	363.35	148.97	107.35
SGD Clay		x	1.36	77.98	23.68	9.01
SGD Silt		x	0.70	91.34	35.37	15.54
SGD Sand		x	0.29	99.74	40.24	19.98
Temperature	mean	x	−1.06	32.84	13.67	5.45
Temperature	standard deviation		0.78	17.09	7.68	2.90
Precipitation	mean	x	0.00	21.45	0.13	0.14
Precipitation	sum		0.51	43500.14	257.58	286.03

TABLE V
FEATURE STATISTICS FOR THE SILT DATASET

Parameter	Description	Selected	Min	Max	Mean	Std
Silt			1.00	95.10	37.23	21.01
VV	mean	x	−20.16	−0.01	−11.19	1.60
VH	mean	x	−22.83	−0.12	−18.23	1.51
VV	standard deviation		0.00	10.49	2.54	0.63
VH	standard deviation	x	0.00	20.39	2.51	0.95
B1	mean	x	0.02	0.54	0.11	0.05
B2	mean		0.04	0.55	0.13	0.05
B3	mean		0.05	0.55	0.16	0.05
B4	mean		0.06	0.56	0.20	0.05
B5	mean		0.07	0.58	0.23	0.05
B6	mean	x	0.08	0.58	0.26	0.05
B7	mean		0.10	0.58	0.28	0.06
B8	mean		0.10	0.60	0.29	0.06
B8A	mean		0.11	0.58	0.30	0.06
B9	mean	x	0.12	0.84	0.31	0.06
B11	mean	x	0.07	0.64	0.33	0.06
B12	mean	x	0.07	0.58	0.27	0.05
B1	standard deviation	x	0.00	0.40	0.10	0.08
B2	standard deviation		0.00	0.41	0.10	0.08
B3	standard deviation		0.00	0.39	0.10	0.08
B4	standard deviation		0.00	0.39	0.10	0.07
B5	standard deviation		0.00	0.39	0.10	0.07
B6	standard deviation		0.00	0.37	0.11	0.06
B7	standard deviation		0.00	0.35	0.10	0.06
B8	standard deviation		0.00	0.35	0.11	0.06
B8A	standard deviation		0.00	0.32	0.10	0.05
B9	standard deviation	x	0.00	0.55	0.11	0.06
B11	standard deviation	x	0.00	0.28	0.08	0.03
B12	standard deviation	x	0.00	0.34	0.07	0.02
Landform		x	0.00	42.00	30.05	6.48
Taxonomy		x	0.00	434.00	162.95	130.28
Elevation		x	0.00	2494.10	433.32	455.35
Slope		x	0.00	365.30	149.66	107.12
SGD Clay		x	0.80	75.82	23.64	9.18
SGD Silt		x	0.87	90.75	35.16	15.69
SGD Sand		x	0.29	99.04	41.19	20.25
Temperature	mean	x	−1.06	32.84	14.34	5.54
Temperature	standard deviation	x	0.80	17.09	7.95	3.05
Precipitation	mean	x	0.00	21.45	0.13	0.15
Precipitation	sum		0.51	43500.14	263.41	308.16

TABLE VI
FEATURE STATISTICS FOR THE SAND DATASET

Parameter	Description	Selected	Min	Max	Mean	Std
Sand			1.00	99.00	39.00	26.27
VV	mean	x	−20.16	−0.01	−11.13	1.59
VH	mean	x	−22.83	−0.12	−18.23	1.51
VV	standard deviation	x	0.00	10.49	2.54	0.63
VH	standard deviation	x	0.00	20.39	2.50	0.95
B1	mean	x	0.02	0.54	0.11	0.05
B2	mean	x	0.03	0.55	0.14	0.05
B3	mean		0.04	0.55	0.17	0.05
B4	mean		0.02	0.56	0.20	0.05
B5	mean		0.05	0.58	0.23	0.05
B6	mean	x	0.10	0.58	0.26	0.05
B7	mean		0.11	0.58	0.28	0.06
B8	mean		0.10	0.60	0.30	0.06
B8A	mean		0.12	0.58	0.31	0.06
B9	mean		0.12	0.84	0.31	0.06
B11	mean	x	0.04	0.71	0.33	0.06
B12	mean	x	0.03	0.62	0.26	0.05
B1	standard deviation	x	0.00	0.40	0.10	0.08
B2	standard deviation		0.00	0.41	0.11	0.09
B3	standard deviation		0.00	0.39	0.10	0.08
B4	standard deviation		0.00	0.39	0.11	0.08
B5	standard deviation		0.00	0.39	0.11	0.07
B6	standard deviation		0.00	0.37	0.11	0.07
B7	standard deviation		0.00	0.35	0.11	0.06
B8	standard deviation		0.00	0.35	0.11	0.06
B8A	standard deviation		0.00	0.32	0.10	0.05
B9	standard deviation	x	0.00	0.55	0.11	0.06
B11	standard deviation	x	0.00	0.28	0.08	0.03
B12	standard deviation	x	0.00	0.34	0.08	0.02
Landform			0.00	42.00	30.12	6.46
Taxonomy		x	0.00	434.00	161.15	128.93
Elevation		x	0.00	2497.53	458.37	460.81
Slope			0.00	364.53	149.22	107.15
SGD Clay		x	0.77	75.85	23.49	8.72
SGD Silt		x	0.90	92.17	36.21	15.44
SGD Sand		x	0.77	98.72	40.27	20.00
Temperature	mean	x	−3.08	32.84	13.85	5.47
Temperature	standard deviation	x	0.80	15.97	8.12	3.07
Precipitation	mean	x	0.00	21.45	0.14	0.15
Precipitation	sum		0.51	43500.14	278.25	314.49

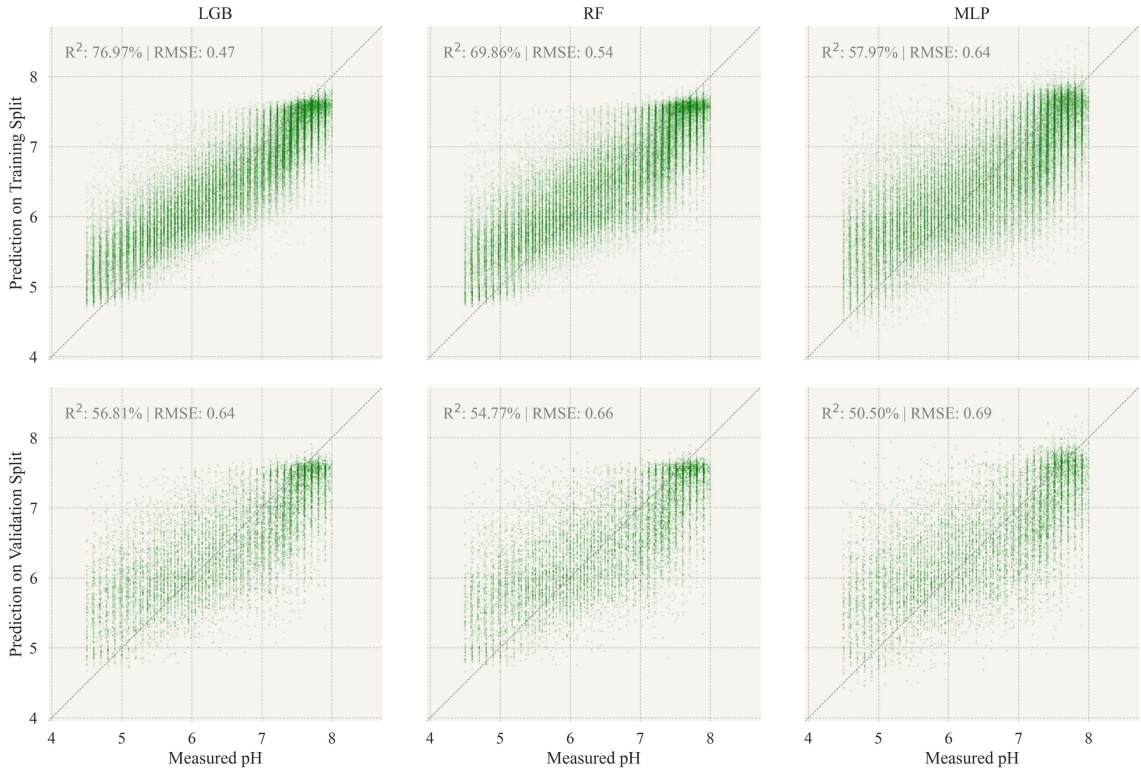

Fig. 11. Scatter plots between measured and predicted pH values using different frameworks for their training and validation datasets.

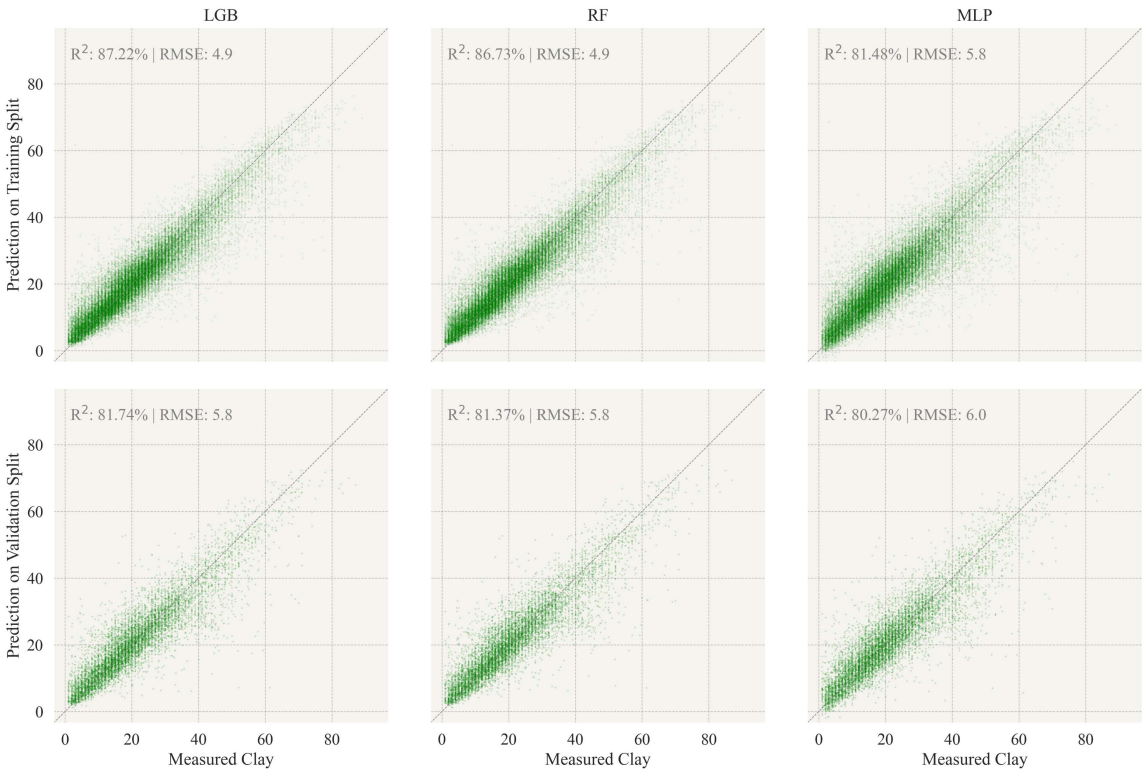

Fig. 12. Scatter plots between measured and predicted clay values using different frameworks for their training and validation datasets.

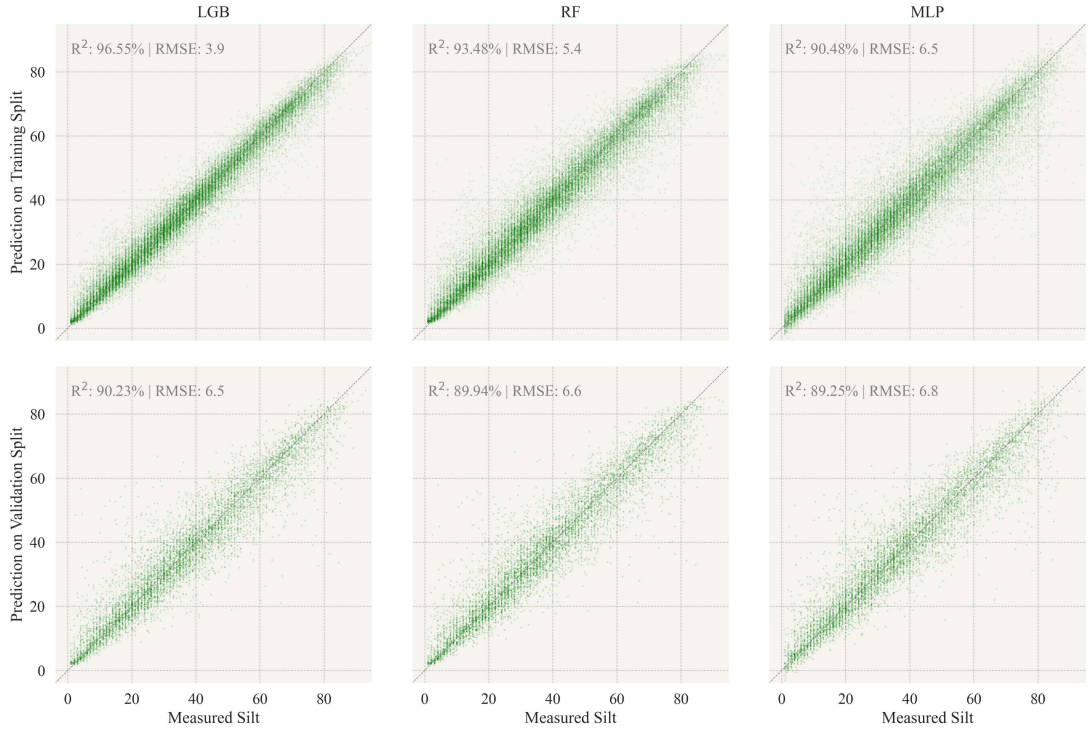

Fig. 13. Scatter plots between measured and predicted silt values using different frameworks for their training and validation datasets.

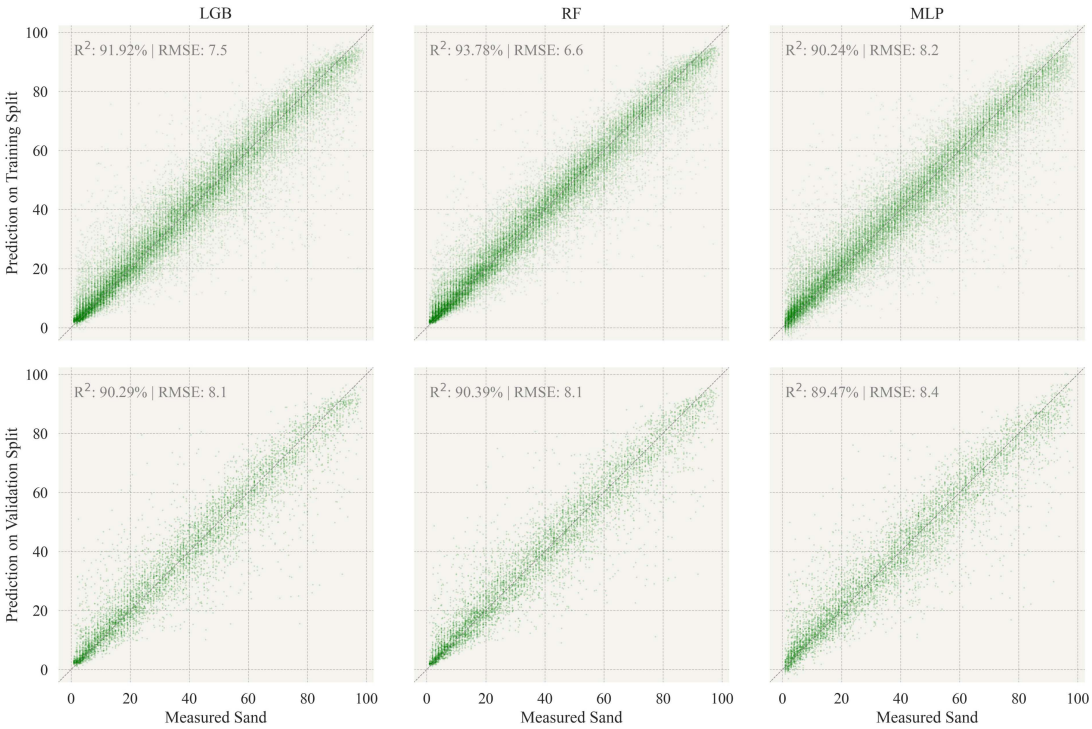

Fig. 14. Scatter plots between measured and predicted sand values using different frameworks for their training and validation datasets.

REFERENCES

[1] J. F. Parr, R. I. Papendick, S. B. Hornick, and R. E. Meyer, "Soil quality: Attributes and relationship to alternative and sustainable agriculture," *Amer. J. Altern. Agriculture*, vol. 7, no. 1/2, pp. 5–11, Jun. 1992.

[2] K. G. Cassman, "Ecological intensification of cereal production systems: Yield," *Proc. Nat. Acad. Sci.*, vol. 96, no. 11, pp. 5952–5959, May 1999.

[3] Z. L. Frogbrook, M. A. Oliver, M. Salahi, and R. H. Ellis, "Exploring the spatial relations between cereal yield and soil chemical properties and the implications for sampling," *Soil Use Manage.*, vol. 18, no. 1, pp. 1–9, Mar. 2002.

[4] W. Guo, S. J. Maas, and K. F. Bronson, "Relationship between cotton yield and soil electrical conductivity, topography, and landsat imagery," *Precis. Agriculture*, vol. 13, no. 6, pp. 678–692, Dec. 2012.

[5] R. M. Adams, B. H. Hurd, S. Lenhart, and N. Leary, "Effects of global climate change on agriculture: An interpretative review," *Climate Res.*, vol. 11, no. 1, pp. 19–30, 1998.

[6] P. Smith and J. E. Olesen, "Synergies between the mitigation of, and adaptation to, climate change in agriculture," *J. Agricultural Sci.*, vol. 148, no. 5, pp. 543–552, Oct. 2010.

[7] H. Yohannes, "A review on relationship between climate change and agriculture," *J. Earth Sci. Climatic Change*, vol. 7, no. 2, 2016, Art. no. 335.

[8] D. Neina, "The role of soil pH in plant nutrition and soil remediation," *Appl. Environ. Soil Sci.*, vol. 2019, Nov. 2019, Art. no. 5794869.

[9] S. O. Oshunsanya and S. O. Oshunsanya, "Introductory chapter: Relevance of soil pH to agriculture," in *Soil pH for Nutrient Availability and Crop Perform.*, IntechOpen, Dec. 2018.

[10] U. W. A. Vitharana, M. Van Meirvenne, D. Simpson, L. Cockx, and J. De Baerdemaeker, "Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area," *Geoderma*, vol. 143, no. 1, pp. 206–215, Jan. 2008.

[11] K. W. T. Goulding, "Soil acidification and the importance of liming agricultural soils with particular reference to the United Kingdom," *Soil Use Manage.*, vol. 32, no. 3, pp. 390–399, Sep. 2016.

[12] I. Nweke and L. Nsoanya, "Soil pH an indices for effective management of soils for crop production," *Int. J. Sci. Technol. Res.*, vol. 2, pp. 132–134, 2013.

[13] N. G. Juma, "Interrelationships between soil structure/texture, soil biota/soil organic matter and crop production," in *Soil Structure/Soil Biota Interrelationships*. Amsterdam, The Netherlands: Elsevier, 1993, pp. 3–30.

[14] N. C. Wollenhaupt, D. J. Mulla, and C. A. G. Crawford, "Soil sampling and interpolation techniques for mapping spatial variability of soil properties," in *The State of Site Specific Management for Agriculture*. Chichester, England, U.K.: Wiley, Oct. 1997, pp. 19–53.

[15] A. Gallardo, "Spatial variability of soil properties in a floodplain forest in Northwest Spain," *Ecosystems*, vol. 6, no. 6, pp. 564–576, Oct. 2003.

[16] F. Rosemary, U. W. A. Vitharana, S. P. Indraratne, R. Weerasooriya, and U. Mishra, "Exploring the spatial variability of soil properties in an Alfisol soil catena," *CATENA*, vol. 150, pp. 53–61, Mar. 2017.

[17] B. Usowicz and J. Lipiec, "Spatial variability of soil properties and cereal yield in a cultivated field on sandy soil," *Soil Tillage Res.*, vol. 174, pp. 241–250, Dec. 2017.

[18] E. Ben-Dor, "Quantitative remote sensing of soil properties," in *Advances in Agronomy*. Cambridge, MA, USA: Academic, Jan. 2002, vol. 75, pp. 173–243.

[19] Y. Ge, J. A. Thomasson, and R. Sui, "Remote sensing of soil properties in precision agriculture: A review," *Front. Earth Sci.*, vol. 5, no. 3, pp. 229–238, Sep. 2011.

[20] S. Grunwald, G. M. Vasques, and R. G. Rivero, "Fusion of soil and remote sensing data to model soil properties," in *Advances in Agronomy*. Cambridge, MA, USA: Academic, Jan. 2015, vol. 131, pp. 1–109.

[21] O. Yuzugullu, F. Lorenz, P. Fröhlich, and F. Liebisch, "Understanding fields by remote sensing: Soil zoning and property mapping," *Remote Sens.*, vol. 12, no. 7, 2020, Art. no. 1116.

[22] F. Ungaro, F. Staffilani, and P. Tarocco, "Assessing and mapping topsoil organic carbon stock at regional scale: A scorpan kriging approach conditional on soil map delineations and land use," *Land Degradation Develop.*, vol. 21, no. 6, pp. 565–581, 2010.

[23] A. C. Richer-de Forges et al., "Remote sensing data for digital soil mapping in french research–A review," *Remote Sens.*, vol. 15, no. 12, 2023, Art. no. 3070.

[24] Z. Libohova et al., "The anatomy of uncertainty for soil pH measurements and predictions: Implications for modellers and practitioners," *Eur. J. Soil Sci.*, vol. 70, no. 1, pp. 185–199, Jan. 2019.

[25] S. M. Brouder, B. S. Hofmann, and D. K. Morris, "Mapping soil pH," *Soil Sci. Soc. Amer. J.*, vol. 69, no. 2, pp. 427–442, Mar. 2005.

[26] R. V. Rossel et al., "A global spectral library to characterize the world's soil," *Earth- Sci. Rev.*, vol. 155, pp. 198–230, 2016.

[27] R. Shrestha, "Relating soil electrical conductivity to remote sensing and other soil properties for assessing soil salinity in northeast Thailand," *Land Degradation Develop.*, vol. 17, no. 6, pp. 677–689, 2006.

[28] M. F. Celik, M. S. Isik, O. Yuzugullu, N. Fajraoui, and E. Erten, "Soil moisture prediction from remote sensing images coupled with climate, soil texture and topography via deep learning," *Remote Sens.*, vol. 14, no. 21, 2022, Art. no. 5584.

[29] P. A. Agbu, D. J. Fehrenbacher, and I. J. Jansen, "Soil property relationships with spot satellite digital data in east central illinois," *Soil Sci. Soc. Amer. J.*, vol. 54, no. 3, pp. 807–812, 1990.

[30] M. Liengsakul, S. Mekpaiboonwatana, P. Pramojanee, K. Bronsveld, and H. Huizing, "Use of GIS and remote sensing for soil mapping and for locating new sites for permanent cropland–A case study in the "highlands" of Northern Thailand," *Geoderma*, vol. 60, no. 1–4, pp. 293–307, 1993.

[31] A. Leone, G. Wright, and C. Corves, "The application of satellite remote sensing for soil studies in upland areas of southern Italy," *Remote Sens.*, vol. 16, no. 6, pp. 1087–1105, 1995.

[32] A. B. McBratney, M. M. Santos, and B. Minasny, "On digital soil mapping," *Geoderma*, vol. 117, no. 1/2, pp. 3–52, 2003.

[33] J. Rogan and D. Chen, "Remote sensing technology for mapping and monitoring land-cover and land-use change," *Prog. Plan.*, vol. 61, no. 4, pp. 301–325, 2004.

[34] J. Qi, C. Wang, Y. Inoue, R. Zhang, and W. Gao, "Synergy of optical and radar remote sensing in agricultural applications," in *Ecosystems' Dynamics, Agricultural Remote Sensing and Modeling, and Site-Specific Agriculture*, vol. 5153. Bellingham, WA, USA: SPIE, 2003, pp. 153–158.

[35] V. Mulder, S. De Bruin, M. E. Schaepman, and T. Mayr, "The use of remote sensing in soil and terrain mapping–A review," *Geoderma*, vol. 162, no. 1/2, pp. 1–19, 2011.

[36] D. J. Mulla, "Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps," *Biosyst. Eng.*, vol. 114, no. 4, pp. 358–371, 2013.

[37] C. Ballabio, P. Panagos, and L. Monatanarella, "Mapping topsoil physical properties at European scale using the Lucas database," *Geoderma*, vol. 261, pp. 110–123, 2016.

[38] S. R. Swain et al., "Estimation of soil texture using sentinel-2 multispectral imaging data: An ensemble modeling approach," *Soil Tillage Res.*, vol. 213, 2021, Art. no. 105134.

[39] T. Hengl et al., "SoilGrids250 m: Global gridded soil information based on machine learning," *PLoS One*, vol. 12, no. 2, Feb. 2017, Art. no. e0169748.

[40] E. Vaudour, C. Gomez, Y. Fouad, and P. Lagacherie, "Sentinel-2 image capacities to predict common topsoil properties of temperate and mediterranean agroecosystems," *Remote Sens. Environ.*, vol. 223, pp. 21–33, 2019.

[41] G. Ma, J. Ding, L. Han, Z. Zhang, and S. Ran, "Digital mapping of soil salinization based on sentinel-1 and sentinel-2 data combined with machine learning algorithms," *Regional Sustainability*, vol. 2, no. 2, pp. 177–188, 2021.

[42] A. Crema, M. Boschetti, F. Nutini, D. Cillis, and R. Casa, "Influence of soil properties on maize and wheat nitrogen status assessment from sentinel-2 data," *Remote Sens.*, vol. 12, no. 14, 2020, Art. no. 2175.

[43] A. Gholizadeh, Z. Daniel, M. Saberioon, and L. Boruvka, "Soil organic carbon and texture retrieving and mapping using proximal, airborne and sentinel-2 spectral imaging," *Remote Sens. Environ.*, vol. 218, pp. 89–103, 2018.

[44] C. Hedley, "The role of precision agriculture for improved nutrient management on farms," *J. Sci. Food Agriculture*, vol. 95, no. 1, pp. 12–19, Jan. 2015.

[45] P. Muruganantham, S. Wibowo, S. Grandhi, N. H. Samrat, and N. Islam, "A systematic literature review on crop yield prediction with deep learning and remote sensing," *Remote Sens.*, vol. 14, no. 9, 2022, Art. no. 1990.

[46] O. Yuzugullu, N. Fajraoui, A. Don, and F. Liebisch, "Satellite-based soil organic carbon mapping on European soils using available datasets and support sampling," *Sci. Remote Sens.*, vol. 9, 2024, Art. no. 100118.

[47] E. C. Brevik, "The potential impact of climate change on soil properties and processes and corresponding influence on food security," *Agriculture*, vol. 3, no. 3, pp. 398–417, 2013.

[48] J. Seibert, J. Stendahl, and R. Sørensen, "Topographical influences on soil properties in boreal forests," *Geoderma*, vol. 141, no. 1/2, pp. 139–148, 2007.

[49] J. D. Phillips and D. A. Marion, "Soil geomorphic classification, soil taxonomy, and effects on soil richness assessments," *Geoderma*, vol. 141, no. 1/2, pp. 89–97, 2007.

[50] V. Houba, E. Temminghoff, G. Gaikhorst, and W. Van Vark, "Soil analysis procedures using 0.01 m calcium chloride as extraction reagent," *Commun. Soil Sci. Plant Anal.*, vol. 31, no. 9/10, pp. 1299–1396, 2000.

[51] "Inicio - ITACyL portal web," Accessed on: Dec. 22, 2022. [Online]. Available: https://www.itacyl.es

[52] "LUCAS - ESDAC - European commission," Accessed on: Dec. 12, 2022. [Online]. Available: https://esdac.jrc.ec.europa.eu/projects/lucas

[53] N. Marianne Stokar, "Startseite - NABODAT," Accessed on: Dec. 12, 2022. [Online]. Available: https://www.nabodat.ch/index.php/de

[54] C. Poeplau, A. Don, H. Flessa, A. Heidkamp, A. Jacobs, and R. Prietz, "Erste Bodenzustandserhebung Landwirtschaft—Kerndatensatz," OpenAgrar, Jan. 12, 2022, doi: 10.3220/DATA20200203151139.
[55] "ISRIC soil data hub," Accessed on: Dec. 12, 2022. [Online]. Available: https://www.isric.org/explore/isric-soil-data-hub
[56] "Home," Accessed on: Dec. 22, 2022. [Online]. Available: https://ismn.earth/en
[57] H. I. Reuter, L. R. Lado, T. Hengl, and L. Montanarella, "Continental-scale digital soil mapping using European soil profile data: Soil pH," *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie*, vol. 19, no. 1, pp. 91–102, 2008.
[58] A. Casagrande, "Classification and identification of soils," *Trans. Amer. Soc. Civil Eng.*, vol. 113, no. 1, pp. 901–930, Jan. 1948.
[59] E. Mohamed, A. Saleh, A. Belal, and A. Gad, "Application of near-infrared reflectance for quantitative assessment of soil properties," *Egyptian J. Remote Sens. Space Sci.*, vol. 21, no. 1, pp. 1–14, 2018.
[60] "Weather API | meteomatics," Accessed on: Dec. 13, 2022. [Online]. Available: https://www.meteomatics.com/en/weather-api
[61] "REST soilgrids API-swagger UI," Accessed on: Oct. 17, 2023. [Online]. Available: https://rest.isric.org/soilgrids/v2.0/docs
[62] S. Khanal, J. Fulton, A. Klopfenstein, N. Douridas, and S. Shearer, "Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield," *Comput. Electron. Agriculture*, vol. 153, pp. 213–225, 2018.
[63] T. Hengl et al., "Soilgrids1 km–global soil information based on automated mapping," *PLoS One*, vol. 9, no. 8, 2014, Art. no. e105992.
[64] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[65] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 3149–3157.
[66] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[67] M. W. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)–A review of applications in the atmospheric sciences," *Atmospheric Environ.*, vol. 32, no. 14/15, pp. 2627–2636, 1998.
[68] J. Bergstra, D. Yamins, and D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *Proc. 30th Int. Conf. Mach. Learn.*, Atlanta, GA, USA, Jun. 2013, vol. 28, pp. 115–123.
[69] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.
[70] C. Chen and Y. Li, "A robust method of thin plate spline and its application to DEM construction," *Comput. Geosci.*, vol. 48, pp. 9–16, Nov. 2012.
[71] B. Basso, C. Fiorentino, D. Cammarano, and U. Schulthess, "Variable rate nitrogen fertilizer response in wheat using remote sensing," *Precis. Agriculture*, vol. 17, pp. 168–182, 2016.
[72] R. P. Sishodia, R. L. Ray, and S. K. Singh, "Applications of remote sensing in precision agriculture: A review," *Remote Sens.*, vol. 12, no. 19, 2020, Art. no. 3136.
[73] L. Karthikeyan, I. Chawla, and A. K. Mishra, "A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses," *J. Hydrol.*, vol. 586, 2020, Art. no. 124905.
[74] J. Holland, P. White, M. Glendining, K. Goulding, and S. McGrath, "Yield responses of arable crops to liming–An evaluation of relationships between yields and soil pH from a long-term liming experiment," *Eur. J. Agronomy*, vol. 105, pp. 176–188, 2019.
[75] J. Blomquist, J.-E. Englund, and K. Berglund, "Soil characteristics and tillage can predict the effect of 'structure lime' on soil aggregate stability," *Soil Res.*, vol. 60, no. 4, pp. 373–384, 2022.
[76] A. Botha and B. Eisenberg, "Estimation of soil water retention from clay content and cation exchange capacity values of soils," *South Afr. J. Plant Soil*, vol. 10, no. 3, pp. 141–143, 1993.
[77] J. Jabro, R. Evans, Y. Kim, and W. Iversen, "Estimating in situ soil–water retention and field water capacity in two contrasting soil textures," *Irrigation Sci.*, vol. 27, pp. 223–229, 2009.
[78] W. Zhao, J. Li, R. Yang, and Y. Li, "Crop yield and water productivity responses in management zones for variable-rate irrigation based on available soil water holding capacity," *Trans. ASABE*, vol. 60, no. 5, pp. 1659–1667, 2017.
[79] P. Schjønning, L. Matieu, and M. H. Thorsøe, "Soil compaction—Drivers, pressures, state, impacts and responses," Cabi Digital Library, DCA - Danish Centre for Food and Agriculture, 2019. [Online]. Available: https://www.cabidigitallibrary.org/doi/full/10.5555/20203423303
[80] M. Stettler et al., "Terranimo–A web-based tool for evaluating soil compaction," *Landtechnik*, vol. 69, no. 3, pp. 132–138, 2014.
[81] E. E. da Silva et al., "Variable-rate in corn sowing for maximizing grain yield," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 12711.
[82] T. Kelly, T. Foster, D. M. Schultz, and T. Mieno, "The effect of soil-moisture uncertainty on irrigation water use and farm profits," *Adv. Water Resour.*, vol. 154, 2021, Art. no. 103982.

Onur Yüzügüllü received the Ph.D. degree in environmental engineering from ETH Zurich, Zurich, Switzerland, in 2017.

He completed the industrial postdoc with the Crop Science Laboratory, ETH Zurich. He is currently the Lead Data Scientist with AgriCircle AG, Schwyz, Switzerland. His research interests include remote sensing of soils and vegetation using machine learning methods for precision agriculture.

Noura Fajraoui received the Ph.D. degree in earth sciences from the University of Strasbourg, Strasbourg, France, in 2014.

She completed the Postdoc with the Risk Safety and Uncertainty Quantification group, ETH Zurich, Zurich, Switzerland. She is currently the Senior Data Scientist with AgriCircle AG, Schwyz, Switzerland. Her research interests include applied machine learning methods for geostatistical data.

Frank Liebisch received the Ph.D. degree in agricultural sciences from the ETH Zurich, Switzerland, in 2011.

He held a postdoctoral and researcher position with the group of Crop Science and coordinated the plant research station at the ETH Zürich. Currently, he leads the research group Water Protection and Substance Flows at Agroscope, Switzerland. His research interests include remote sensing and precision farming in the context of nutrient and resource management in various cropping systems.