





Article

AI-Driven Integration of Sentinel-1 SAR for High-Resolution Soil Water Content Estimation to Enhance Precision Irrigation in Smallholder Maize Systems, Vhembe District

Gift Siphwiwe Nxumalo ^{1,*}, Tondani Sanah Ramabulana ², Zibuyile Dlamini ³, Tamás János ¹,
Nikolett Éva Kiss ¹ and Attila Nagy ¹

¹ Institute of Water and Environmental Management, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, 146B Böszörményi Str., 4032 Debrecen, Hungary

² Institute of Geography and Earth Sciences, Faculty of Sciences, University of Pécs, Ifjúság útja, 7624 Pécs, Hungary

³ Doctoral School of Environmental Sciences, Institute of Environmental Sciences, University of Agriculture and Life Sciences, Páter Károly Str. 1., 2100 Gödöllő, Hungary

* Correspondence: nxumalo.gift.siphwiwe@agr.unideb.hu

Abstract

Climate variability threatens smallholder maize production in semi-arid Southern Africa, necessitating accurate irrigation management. We developed an Earth Observation–machine learning framework integrating Sentinel-1 SAR, TU Wien retrievals, and meteorological data to generate daily 10 m resolution root-zone soil moisture estimates (0–100 cm) for South Africa’s Vhembe District (2017–2022). Five algorithms—Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), and Multivariate Adaptive Regression Splines (MARS)—were calibrated using ~50,000 observations from two monitoring stations across six depths and five growing seasons. RF and XGBoost achieved highest accuracy ($R^2 = 0.96\text{--}0.97$, $RMSE < 0.025\text{ cm}^3/\text{cm}^3$), detecting critical irrigation thresholds (management allowable depletion = $0.23\text{ cm}^3/\text{cm}^3$, field capacity = $0.35\text{ cm}^3/\text{cm}^3$) with operational precision ($nRMSE < 0.05$). Depth-stratified validation revealed strong SAR surface correlations ($r = 0.84\text{--}0.85$ at 10 cm) declining systematically with depth ($r < 0.2$ below 40 cm), confirming ML models integrate satellite observations at shallow layers with meteorological gap-filling at depth. District mapping showed 79–94% of maize areas required irrigation during dry years (2017–2019, 2021–2022) versus 32% in wet 2020–2021. The framework provides a transferable pathway for precision irrigation in smallholder systems, pending vegetation-corrected retrievals and expanded validation.

Keywords: root-zone soil moisture; precision irrigation; TU Wien soil moisture retrieval; Sentinel-1 Synthetic Aperture Radar (SAR); machine learning; drought monitoring



Academic Editor: Nektarios N. Kourgiyalas

Received: 19 January 2026

Revised: 11 February 2026

Accepted: 14 February 2026

Published: 16 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Climate change and increasing hydroclimatic variability threaten agricultural sustainability in water-scarce regions such as the Vhembe District of Limpopo, South Africa. Smallholder farmers here rely heavily on maize production, which is increasingly constrained by recurrent droughts and erratic rainfall [1,2]. Soil moisture is central to this challenge, as it regulates evapotranspiration, crop water uptake, and yield performance. Imbalances between soil water availability and atmospheric demand can severely reduce productivity [3–6]. While mulching and straw incorporation can enhance retention [7], effective irrigation scheduling requires accurate, timely monitoring of soil water content [8].

Conventional methods, such as gravimetric sampling, tensiometers, and TDR probes, provide accurate point measurements but are labor-intensive, costly, and spatially limited [9]. Remote sensing offers a scalable alternative, with Sentinel-1 Synthetic Aperture Radar (SAR) being particularly valuable in cloudy subtropical regions. Its C-band radar operates independently of daylight and weather, enabling consistent observations [10,11]. However, vegetation, surface roughness, and topography often distort backscatter signals, reducing retrieval accuracy in heterogeneous agricultural landscapes [12].

The TU Wien change detection model has become a widely used retrieval approach, exploiting temporal backscatter variations to isolate soil moisture dynamics from static surface effects [13]. It underpins global products such as the Copernicus Global Land Service (CGLS) Soil Water Index and the European Space Agency Climate Change Initiative (ESA CCI) Soil Moisture dataset, providing consistent long-term monitoring [14,15]. Yet, its sensitivity is strongest in the top 5–10 cm of soil and declines under dense vegetation and complex terrain, limiting its application to crop-relevant root-zone water dynamics. Integrating TU Wien retrievals with meteorological data and machine learning (ML) offers a pathway to extend soil moisture estimates into deeper soil layers and improve robustness in fragmented landscapes.

Machine learning has demonstrated strong potential for capturing nonlinear relationships between SAR backscatter, meteorological drivers, and soil moisture. Regional applications include terrain-augmented ML for forest soil moisture in Sweden [16], Sentinel-based XGBoost models in China [17], and global-scale gradient-boosted regression trees [18]. Deep learning approaches, including convolutional neural networks, have further improved retrieval accuracy under crop cover [19,20]. However, ensemble ML models often outperform shallow neural networks [21,22], and hybrid machine learning–deep learning (MD–DL) strategies for root-zone estimation remain underexplored.

Despite these advances, most Earth observation (EO) studies remain limited to surface retrievals or coarse-scale global products, with minimal operational translation for smallholder irrigation [23]. In Vhembe, fragmented fields, mixed cropping, variable soils, and sparse validation data pose additional challenges [24].

While Sentinel-1 SAR data, the TU Wien change detection model, meteorological variables, and in situ soil moisture have each been applied separately or in partial combinations for soil moisture estimation, no study has integrated all of these with machine learning to produce high-resolution (~10 m), daily root-zone (0–1 m) soil moisture maps tailored to the near-field irrigation needs of smallholder farmers in southern Africa. The specific objectives of this study are to:

- Enhance retrieval accuracy by applying vegetation and topographic corrections to Sentinel-1 backscatter.
- Derive daily 10 m surface soil moisture at 10 cm depth (2017–2022) using the TU Wien change detection model, calibrated and validated with Agricultural Research Council (ARC) probe measurements.
- Integrate TU Wien retrievals with meteorological variables (rainfall, temperature, wind speed, and humidity) and train five machine learning models to estimate root-zone soil moisture at 20–100 cm depth, recognizing that while maize water uptake is concentrated in the upper 0–60 cm [25], deeper layers (60–100 cm) provide critical buffering capacity during prolonged drought [26].
- Generate high-resolution daily soil moisture maps for the Vhembe District and assess their reliability against in situ observations.
- Develop a framework to translate soil moisture estimates into prototype outputs, including per-field water-deficit alerts and irrigation scheduling recommendations for smallholder maize farmers, and evaluate their technical feasibility at two calibration sites.

By linking physically based retrievals with data-driven prediction, this study advances soil moisture monitoring from retrieval to application, transforming EO-based products into actionable irrigation guidance for smallholder systems in southern Africa.

2. Materials and Methods

2.1. Study Area

The study was conducted in the Vhembe District, Limpopo Province, South Africa (22.1–23.3° S, 29.5–31.3° E), which spans approximately 12,500 km² of semi-arid to subtropical savanna (Figure 1). The district is predominantly rural and characterized by mixed crop–livestock systems, moderate-fertility soils (mainly Ferralsols and Luvisols), and a climate transitioning from subtropical lowlands to tropical savanna at higher elevations. Land cover is dominated by smallholder croplands (primarily maize and sorghum), interspersed with grasslands and sparse forests.

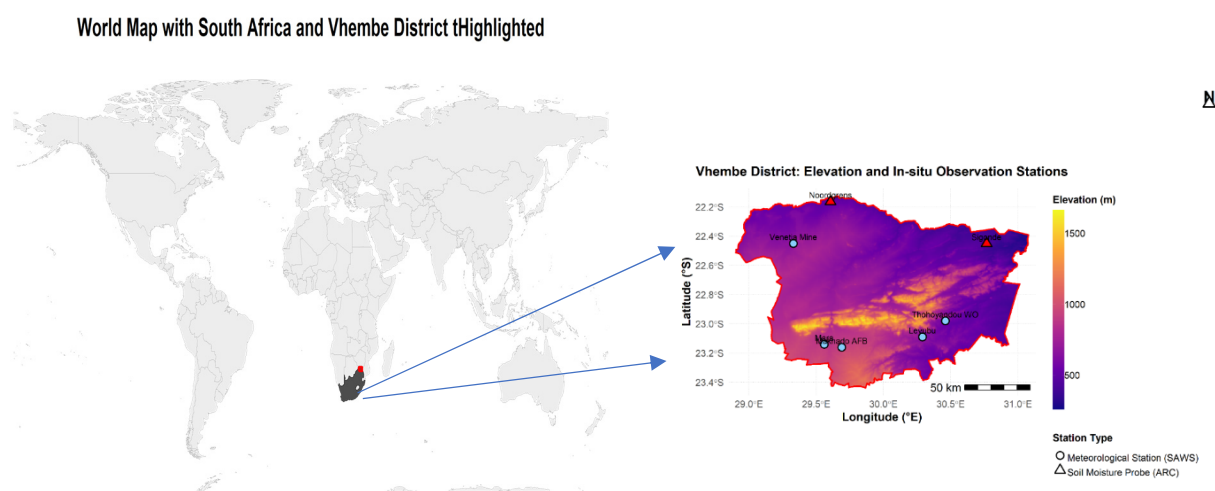


Figure 1. Study area overview showing the location of the Vhembe District in South Africa (left) and elevation map (right) with in situ observation stations for meteorological and soil moisture measurements. The map illustrates spatial distribution of measurement infrastructure supporting soil moisture and climate analysis in the study.

Topography ranges from ~200 m in river valleys to ~1800 m in the Soutpansberg Mountains, producing marked microclimatic variations that directly influence soil moisture dynamics. Rainfall is highly seasonal (≈ 400 – 800 mm annually, concentrated from October to March), with frequent droughts and heat stress events limiting agricultural productivity [27,28]. Irrigation is practiced mainly in valley systems, supported by rivers, dams, and boreholes.

Maize production illustrates the district’s vulnerability to water stress. Under rainfed conditions, yields typically range from 0.8 to 1.6 t ha⁻¹ [29], while irrigated schemes such as Dzindi can achieve 4 – 6 t ha⁻¹ with improved inputs [30], and intensive green-maize systems may reach $\sim 22,000$ cobs ha⁻¹ [31]. However, extreme drought years—including 2015–2016 [32] and 2018–2020/21—led to widespread crop failures across Limpopo [33,34]. These hydroclimatic challenges underscore the critical need for improved soil moisture monitoring and precision irrigation management in Vhembe.

2.2. Datasets

This study integrated Sentinel-1 SAR observations, in situ soil moisture profiles, meteorological records, and global satellite products to generate high-resolution soil moisture estimates for the Vhembe District (workflow shown in Table 1).

Table 1. Detailed description of the datasets used in this study.

Category	Product/Source	Variable(s)	Spatial Resolution	Temporal Resolution	Time Period
Remotely sensed data	Sentinel-1 SAR (VV, VH) (European Space Agency Sentinel Data Hub: https://www.sentinel-hub.com/)	Backscatter (σ^0), normalized indices	10 m	6 days (interpolated to daily)	1 November–31 March 2017–2022
	SRTM DEM: https://earthexplorer.usgs.gov/ (accessed on 2 July 2025)	Elevation, topographic indices	30 m	Static	—
Ground-based data	ARC probes: https://www.arc.agric.za/ (accessed on 25 May 2025)	Soil moisture (10, 20, 40, 60, 80, 100 cm)	Point (~cm depth)	Daily (aggregated from hourly)	1 November–31 March 2017–2022
Meteorological data	South African Weather Service (SAWS): https://www.weathersa.co.za/home/recentclimate (accessed on 19 April 2025)	Rainfall, temperature, humidity, wind speed	Station-based	Daily	1 November–31 March 2017–2022
Processed EO product	TU Wien Change Detection Model (from Sentinel-1)	Surface soil moisture (0–10 cm)	10 m	Daily (gap-filled)	1 November–31 March 2017–2022
Machine learning data	Derived features (RF, XGB, KNN, MARS, SVM inputs)	Root-zone soil moisture (20–100 cm)	10 m	Daily	1 November–31 March 2017–2022

Note: Sentinel-1 native temporal resolution is 6 days; daily products were generated through temporal gap-filling as described in Section 2.3.

2.2.1. Sentinel-1 SAR Data

Sentinel-1 Ground Range Detected (GRD) imagery in vertical–vertical (VV) and vertical–horizontal (VH) polarizations at 10 m spatial resolution formed the core dataset for soil moisture retrieval. The Sentinel-1 constellation provides a 6-day revisit frequency over the study area. Daily soil moisture estimates were generated through temporal interpolation and gap-filling procedures described in Section 2.3.

Preprocessing was performed in Google Earth Engine (GEE) following ESA-recommended protocols [35], including thermal noise removal, precise orbit correction, radiometric calibration to sigma nought (σ^0), terrain correction with the 30 m Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM), and speckle filtering (Lee sigma filter with 7×7 kernel). The resulting terrain-corrected σ^0_{VV} and σ^0_{VH} backscatter coefficients served as inputs to the TU Wien change detection model.

2.2.2. Topographic Data: SRTM DEM

The SRTM DEM at 30 m spatial resolution was used for terrain correction of Sentinel-1 backscatter and to derive topographic predictors (elevation, slope, aspect) for machine learning models. These topographic indices account for terrain-induced vari-

ations in backscatter and influence on soil moisture distribution patterns across the heterogeneous landscape.

2.2.3. Soil Moisture Reference Data (ARC Probes)

Soil moisture reference data were collected using Delta-T Devices DFM capacitance probes installed by the Agricultural Research Council (ARC) at two monitoring stations within representative maize fields in Vhembe District: Noordgrens and Sigonde. Each station is equipped with one multi-depth capacitance probe measuring volumetric soil moisture at six depths (10, 20, 40, 60, 80, and 100 cm) at hourly intervals, which were aggregated to daily mean values for this analysis. Data quality control included removal of outliers (values outside physically plausible ranges of 0.05–0.55 cm³/cm³) and flagging of periods with sensor malfunction.

The Sigonde ARC station, positioned in the southern sub-humid plains at moderate elevation (~600–700 m), features deep clay-loam soils and subtropical vegetation mosaics dominated by smallholder maize and vegetables. The Noordgrens ARC station, located in the northern lowlands (~400 m), encompasses sandy loam soils within a semi-arid savanna and bushveld setting, supporting irrigated maize cultivation. Calibration of surface soil moisture retrievals was conducted at 10 cm depth using ARC probe data because the TU Wien change detection model is most sensitive to near-surface (0–10 cm) soil moisture dynamics. Validation of root-zone predictions used deeper soil moisture observations at 20–100 cm to assess the machine learning models' ability to extrapolate from surface conditions and meteorological predictors to the full root zone. These ARC locations provide robust in situ observation points spanning diverse soil types (sandy loam vs. clay-loam), vegetation systems, and topographic gradients, thereby supporting reliable calibration and validation of soil moisture models throughout the Vhembe District.

While the two ARC monitoring sites represent contrasting soil types (sandy loam at Noordgrens vs. clay loam at Sigonde), the relatively small difference in empirically derived dry/wet thresholds (<0.03 cm³/cm³)—calculated as the difference between site-specific 5th and 95th percentiles (see Section 2.4, for detailed methodology)—suggests similar seasonal moisture extremes at both locations. This does not imply that true texture-specific hydraulic properties are uniform across the district. Rather, it indicates that our percentile-based calibration captures observed seasonal moisture variability common to both sites. Soil texture maps (where available) or localized calibration would be needed to define spatially explicit field capacity and permanent wilting point values for heterogeneous landscapes with greater textural contrasts.

2.2.4. Meteorological Data (South African Weather Service)

Daily meteorological data were obtained from the South African Weather Service (SAWS) network (Figure 2), including:

- Rainfall (mm)
- Maximum and minimum air temperature (°C)
- Relative humidity (%)
- Wind speed (m/s)

These variables, along with their 1–3 day lagged values, were incorporated as explanatory predictors in the machine learning models to capture temporal moisture dynamics and atmospheric controls on evapotranspiration. For spatial predictions at locations without weather stations, meteorological data were interpolated using inverse distance weighting from the three nearest SAWS stations (maximum interpolation distance: 50 km).

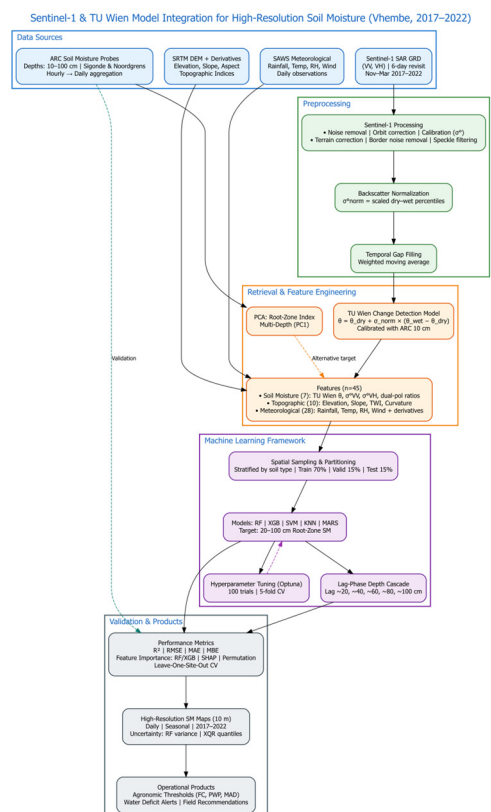


Figure 2. Workflow for soil water content (SWC, cm³/cm³) mapping in the Vhembe District using Sentinel-1 TU WIEN change detection, lagged meteorological and soil predictors (1–3-day lags), principal component analysis (PCA), and machine learning. Model output units correspond to management-relevant soil water thresholds; all validation is performed on independent test data.

2.2.5. Processed EO Product (TU Wien Soil Moisture Retrievals)

Together, these datasets enabled a two-stage workflow: (i) TU Wien model application to Sentinel-1 SAR for daily 10 cm soil moisture retrievals, and (ii) machine learning prediction of root-zone soil moisture (20–100 cm) using TU Wien retrievals and meteorological drivers strictly during the maize cropping season (1 November–31 March 2017–2022).

2.3. Sentinel-1 GRD Preprocessing

As illustrated in Figure 2, Sentinel-1 GRD imagery was preprocessed in GEE using standard ESA workflows [35]. The processing chain included:

1. Thermal noise removal to reduce radiometric artifacts
2. Precise orbit correction using restituted orbit files
3. Radiometric calibration to sigma nought (σ^0) backscatter coefficient
4. Terrain correction using the SRTM 30 m DEM with Range-Doppler orthorectification
5. Speckle filtering using Lee sigma filter (7×7 kernel) to reduce granular noise while preserving spatial features

Water masking was applied using the JRC Global Surface Water occurrence dataset to exclude pixels with >20% permanent water presence, preventing contamination of soil moisture retrievals by open water bodies.

The normalized backscatter index (Equation (1)) was computed as:

$$\sigma_{norm}^0 = \frac{\sigma_{dry}^0 - \sigma_{wet}^0}{\sigma_{wet}^0 - \sigma_{dry}^0} \tag{1}$$

where:

- σ^0 = terrain-corrected backscatter coefficient (VV or VH),
- σ_{dry}^0 = reference dry backscatter (5th percentile from 2017 to 2022 seasonal time series),
- σ_{wet}^0 = reference wet backscatter (95th percentile from 2017 to 2022 seasonal time series),
- σ_{norm}^0 = normalized backscatter index (0–1), representing relative soil moisture status.

Important note on vegetation influence: These dry and wet reference values are derived from multi-year time series spanning all crop growth stages (bare soil, canopy development, senescence). Therefore, σ_{dry}^0 and σ_{wet}^0 represent composite statistics integrating both soil moisture extremes and seasonal vegetation structure changes. Ideally, vegetation correction algorithms (e.g., Water Cloud Model) should be applied to isolate the soil backscatter component before normalization [15,36]. This study used uncorrected backscatter as proof-of-concept, with the ML framework implicitly compensating for vegetation effects through temporal and meteorological predictors. This limitation is discussed further in Section 4 (Lines 1189–1193).

To generate continuous daily soil moisture estimates from the 6-day Sentinel-1 revisit cycle, temporal gap-filling was performed using a weighted moving average approach. This interpolation method assumes a relatively smooth temporal evolution of soil moisture between observations and may underestimate abrupt changes caused by localized convective rainfall or rapid drainage events. For each missing date, σ_{norm}^0 values were interpolated from the closest available observations (± 3 days) using inverse temporal distance weighting (Equation (2)):

$$\sigma_{norm,t}^0 = \frac{\sum_i w_i \sigma_{norm,i}^0}{\sum_i w_i}, \quad w_i = \frac{1}{|t - t_i| + 1} \quad (2)$$

where t is the target date, t_i are available observation dates within ± 3 days, and w_i are temporal weights. This approach aims to preserve short-term moisture dynamics while providing continuous daily coverage required for operational irrigation scheduling, though temporal interpolation may smooth rapid moisture changes occurring between satellite overpasses (e.g., during intense rainfall events).

The normalized σ^0_{VV} and σ^0_{VH} were subsequently used in the TU Wien change detection model to derive daily surface (10 cm) soil moisture estimates.

2.4. TU Wien Change Detection Model

The TU Wien change detection model was applied to normalized Sentinel-1 backscatter (σ_{norm}) to estimate surface soil moisture at 0–10 cm depth [13,15]. Sentinel-1 C-band SAR (5.6 cm wavelength) is primarily sensitive to moisture in the top 2–5 cm of bare or sparsely vegetated soil, with penetration depth decreasing under dense canopy cover [37,38]. Therefore, TU Wien retrievals should be interpreted as near-surface moisture proxies rather than direct measurements at 10 cm depth. Our calibration using 10 cm ARC probe data establishes an empirical statistical relationship between satellite-observed surface backscatter dynamics and subsurface moisture at the sensor depth, but this does not imply that C-band radar directly measures moisture at 10 cm. The machine learning framework developed in Section 2.5 statistically propagates these surface moisture proxies into deeper soil layers (20–100 cm) by leveraging temporal persistence, meteorological drivers, and topographic controls—a data-driven approach that complements but does not replace physical understanding of infiltration and percolation processes. Importantly, TU Wien retrieval performance degrades substantially under moderate-to-dense vegetation cover, where C-band backscatter becomes dominated by canopy scattering rather than soil conditions [15,36] and is limited to shallow soil layers (<10 cm) due to microwave penetration constraints [37,39]. To address these limitations, our framework integrates ML algorithms with multiple data

sources—SAR, meteorological variables, and temporal features—allowing adaptive weighting of predictors based on local retrieval conditions.

This hybrid approach enables consistent root-zone moisture estimation across heterogeneous vegetation cover and soil depths. The model assumes that temporal fluctuations in radar backscatter are primarily driven by soil moisture variations, while static factors such as soil roughness and long-term average vegetation structure remain relatively stable. However, intra-seasonal vegetation phenology (crop emergence, canopy development, senescence) can introduce systematic backscatter variations unrelated to soil moisture, particularly under moderate-to-dense crop cover [15,36]. The dry and wet reference backscatter values (Equation (1)) are calculated across the entire 2017–2022 maize growing season dataset and therefore implicitly integrate these vegetation effects. While dedicated vegetation correction should ideally precede TU Wien retrieval [40,41], this study demonstrates a proof-of-concept workflow where the ML framework (Section 2.5) compensates for uncorrected vegetation influences through temporal lagged features, NDVI-derived indices (where available), and meteorological predictors. This limitation and its implications for operational deployment are discussed in Section 4. By anchoring observations between long-term dry and wet reference states, the method reduces reliance on ancillary inputs and ensures consistency across heterogeneous conditions (Equation (3)).

Surface soil moisture (θ) was retrieved as:

$$\theta = \theta_{\text{dry}} + \sigma_{\text{norm}} \times (\theta_{\text{wet}} - \theta_{\text{dry}}) \quad (3)$$

where:

- θ = estimated volumetric soil moisture [cm^3/cm^3]
- θ_{dry} = dry reference soil moisture, set to $0.10 \text{ cm}^3/\text{cm}^3$ (10% volumetric water content) based on the 5th percentile of ARC probe values across all depths and seasons (2017–2022)
- θ_{wet} = wet reference soil moisture, set to $0.50 \text{ cm}^3/\text{cm}^3$ (50% volumetric water content) reflecting the 95th percentile of ARC probe values
- σ_{norm} = normalized backscatter index (0–1, from Equation (1) in Section 2.3)

Important clarification on hydraulic thresholds: The choice of 5th and 95th percentiles as dry and wet references was based on analysis of the complete ARC probe dataset (2017–2022, all depths: 10–100 cm, $n = 8760$ observations per site). These percentiles represent site-specific extremes while excluding outliers that might reflect sensor errors or exceptional events. Critically, the resulting $\theta_{\text{dry}} = 0.10 \text{ cm}^3/\text{cm}^3$ and $\theta_{\text{wet}} = 0.50 \text{ cm}^3/\text{cm}^3$ are empirical management thresholds derived from multi-year field observations, not pedologically defined soil hydraulic constants. θ_{dry} ($0.10 \text{ cm}^3/\text{cm}^3$) approximates typical permanent wilting point values (-1500 kPa matric potential) reported for maize in similar semi-arid environments [29] and represents the lower bound of observed seasonal moisture. θ_{wet} ($0.50 \text{ cm}^3/\text{cm}^3$), however, does not represent field capacity in the classical sense (water content at -33 kPa after drainage). True field capacity for agricultural soils typically ranges from 0.20 to $0.40 \text{ cm}^3/\text{cm}^3$, depending on texture [42]. Our observed 95th percentile value of $0.50 \text{ cm}^3/\text{cm}^3$ likely approaches total porosity and represents near-saturated conditions immediately following heavy rainfall or irrigation events, before gravitational drainage to field capacity occurs. This high threshold is useful for irrigation management as it indicates when fields are “fully recharged” and require no additional water input, even though soil moisture will subsequently decline to true field capacity within 24–48 h. We acknowledge that true field capacity and permanent wilting point are soil-texture-dependent properties that vary spatially across the heterogeneous Vhembe landscape (clay-loam at Sigonde vs. sandy loam at Noordgrens). While our percentile-based approach provides practical decision thresholds, operational users should ideally calibrate these

values using texture-specific pedotransfer functions [42] or direct laboratory measurements of water retention curves.

Soil type differences were assessed by calculating separate percentiles for Noord-grens (sandy loam) and Sigonde (clay-loam) sites. The 5th/95th percentiles differed by $<0.03 \text{ cm}^3/\text{cm}^3$ between sites, falling within measurement uncertainty ($\pm 0.02 \text{ cm}^3/\text{cm}^3$ for DFM probes). This relatively small difference reflects the fact that our percentile-based thresholds capture seasonal moisture extremes observed at both sites rather than fundamental texture-specific hydraulic properties. Therefore, unified dry/wet references were applied across the district to maintain consistency and avoid introducing artificial spatial discontinuities in the soil moisture product. Future applications in areas with more pronounced textural contrasts should consider site-specific calibration or incorporation of soil texture maps to define spatially variable thresholds. This formulation scales normalized Sentinel-1 SAR backscatter between site-specific dry and wet soil moisture conditions, enabling the retrieval of daily surface soil moisture dynamics at high spatial resolution (10 m). These TU Wien surface soil moisture estimates were subsequently combined with meteorological predictors and in situ data to support machine learning-based root-zone soil moisture modeling described in Section 2.5.

2.5. Machine Learning Models

To extend surface soil moisture estimates into the crop-relevant root zone (20–100 cm depth), we developed a machine learning framework integrating TU Wien retrievals [43], antecedent soil moisture conditions, meteorological drivers, and topographic features. This section describes the conceptual framework, input features, spatial sampling strategy, and model architectures.

2.5.1. Conceptual Framework and Workflow

Root-zone soil moisture dynamics reflect the interplay of surface infiltration, vertical percolation, evapotranspiration, and lateral redistribution—processes that operate across multiple timescales. To capture these dynamics, we implemented a lag-phase prediction framework that leverages temporal persistence in soil water status. Specifically, soil moisture at each depth layer was predicted using:

1. TU Wien surface soil moisture (0–10 cm) from the current day
2. Antecedent soil moisture from shallower depths (1–3-day lags), representing vertical water movement
3. Meteorological variables from previous days (1–3-day lags), capturing atmospheric demand and recharge
4. Static topographic features, controlling lateral flow and local water accumulation

Figure 2 illustrates this workflow. The prediction proceeds sequentially through the soil profile: 10 cm TU Wien estimates inform 20 cm predictions, which in turn inform 40 cm predictions, continuing through 60, 80, and 100 cm layers. This physically informed cascade ensures deeper predictions integrate both recent surface dynamics and cumulative antecedent conditions. While this study estimates soil moisture across the full 0–100 cm profile, we acknowledge that the agronomic relevance of these depth layers differs for maize irrigation management. Active water uptake by maize is concentrated in the upper 30–60 cm, where root density and hydraulic conductivity are highest [25,44]. These shallow layers respond rapidly to rainfall and irrigation, making them critical for short-term scheduling decisions (1–3-day horizons). Conversely, deeper layers (60–100 cm) contribute to long-term drought buffering and plant water status during extended dry spells but have limited relevance for day-to-day irrigation control [26]. In our irrigation decision support framework (Section 3.4.3), we prioritize 20–60 cm predictions for operational scheduling

while using 80–100 cm estimates to assess seasonal drought risk and long-term water availability. This depth-weighted approach reflects the physiological reality that shallow soil moisture governs immediate crop water stress, while deep reserves determine overall resilience to multi-week rainfall deficits.

Limitation regarding temporal root zone development: This framework predicts soil moisture at fixed depth layers (10, 20, 40, 60, 80, 100 cm) throughout the growing season, implicitly treating the “agronomically relevant root zone” as static. In reality, maize rooting depth increases progressively during crop development: predominantly 0–30 cm during early vegetative stages (V4–V8), extending to 60–80 cm at peak growth (VT–R3), and potentially exceeding 100 cm under water deficit conditions [25,44]. Our approach does not explicitly account for these temporal changes in effective rooting depth. While the ML models adaptively weight different soil layers based on temporal context (as shown in feature importance analysis, Supplementary Figures S1–S3), this weighting is learned from data patterns rather than explicitly programmed phenological stages. Future refinements could incorporate crop growth models (e.g., DSSAT, AquaCrop) to define time-varying effective rooting depth zones, thereby improving the physiological relevance of irrigation recommendations during early vs. late season stages. This limitation is further discussed in Section 4.

2.5.2. Input Features and Temporal Structure

All machine learning models were trained using an identical, standardized feature set comprising 45 predictor variables (Tables 2 and 3):

A. Soil Moisture Features (7 variables):

- TU Wien surface soil moisture at 10 cm depth, current day (θ)
- Lagged soil moisture from shallower depth: SWC_{t-1}, SWC_{t-2}, SWC_{t-3} (3 lags)
- Depth-specific indices: vertical gradient ($\theta_{\text{surface}} - \theta_{\text{previous_depth}}$), cumulative 3-day mean

B. Meteorological Features (28 variables):

- Rainfall [mm]: current day and 1–3-day lags (4 variables)
- Air temperature [°C]: daily mean, maximum, minimum, and their 1–3-day lags (12 variables)
- Relative humidity [%]: current day and 1–3-day lags (4 variables)
- Wind speed [m/s]: current day and 1–3-day lags (4 variables)
- Derived variables: Cumulative 3-day rainfall, mean 3-day temperature, vapor pressure deficit (VPD) (4 variables)

C. Topographic Features (10 variables):

- Elevation [m], slope [degrees], aspect [degrees]
- Topographic wetness index (TWI), topographic position index (TPI)
- Flow accumulation, curvature (profile, plan, total)
- Distance to nearest stream [m]

The notation “*t*” refers to the prediction target date. For example, to predict soil moisture on 15 January 2020:

- SWC(*t*-1) = soil moisture from 14 January 2020
- Rainfall(*t*-2) = rainfall from 13 January 2020
- Temperature(*t*-3) = temperature from 12 January 2020

This lag structure captures the temporal memory inherent in soil water dynamics, where current moisture status reflects not only recent inputs but also cumulative antecedent conditions over preceding days.

2.5.3. Spatial Sampling Strategy

To balance computational efficiency with spatial representativeness, approximately 10,000 training points were randomly sampled per growing season within agricultural areas of the Vhembe District (total study area: $\sim 12,500 \text{ km}^2$). This sampling density of $\sim 0.8 \text{ points/km}^2$ ensured:

1. Adequate representation of spatial heterogeneity across soil types, topographic gradients, and land management practices
2. Computational feasibility for model training and hyperparameter tuning within reasonable time frames ($\sim 2\text{--}4 \text{ h}$ per model on standard workstations)
3. Statistical robustness with sufficient sample size ($n \approx 50,000$ total across five seasons) for reliable model generalization

Sampling was stratified by land cover type (cropland vs. non-cropland, derived from Sentinel-2 NDVI thresholds) to focus on maize production areas while maintaining geographic diversity. Points were extracted using the terra and sf R packages, with spatial alignment and masking ensuring all predictor layers shared a common coordinate reference system (WGS84/UTM Zone 35S) and 10 m resolution.

2.5.4. Machine Learning Algorithms

To first validate the TU Wien surface retrievals against ARC probe measurements at 10 cm depth (Section 3.2), ML models were trained using only TU Wien theta as the predictor variable. This baseline assessment establishes the accuracy of calibrated satellite retrievals before incorporating additional features for root-zone estimation (Section 3.3). The five representative algorithms spanning different statistical learning paradigms were selected to compare predictive performance and identify optimal approaches (see Table 3 for key hyperparameters tuned):

A. Random Forest (RF)

Random Forest constructs an ensemble of decision trees using bootstrapped samples and random feature subsets (typically \sqrt{p} features per split, where p = total features). Each tree votes on the final prediction, with averaging reducing overfitting and enhancing robustness to noisy, heterogeneous data [45]. RF excels at capturing nonlinear feature interactions and handling collinear predictors—common characteristics of soil moisture datasets.

B. Extreme Gradient Boosting (XGBoost)

XGBoost sequentially builds trees that iteratively correct residuals from previous iterations, optimizing a regularized objective function that balances bias and variance [46]. Its gradient-based optimization and built-in regularization (L1/L2 penalties) provide computational efficiency and high accuracy, particularly for large datasets with complex patterns [21,47].

C. Support Vector Machines (SVM)

SVM projects input features into a higher-dimensional space via a kernel function (radial basis function used here), seeking an optimal hyperplane that maximizes the margin between classes or minimizes prediction error in regression. The radial basis kernel captures nonlinear relationships and has demonstrated reliable performance in Sentinel-1-based soil moisture retrievals [48,49].

D. k-Nearest Neighbors (KNN)

KNN predicts soil moisture by averaging values from the k most similar observations in feature space, measured by Euclidean distance. This non-parametric, instance-based method adapts well to localized patterns and has effectively predicted site-specific soil moisture profiles in heterogeneous agricultural landscapes [50,51].

E. Multivariate Adaptive Regression Splines (MARS)

MARS fits piecewise linear or cubic spline regressions, automatically detecting break-points and interaction terms in the predictor-response relationship [52]. Its flexibility in modeling nonlinear thresholds (e.g., rainfall-runoff transitions, temperature-evaporation responses) makes it suitable for hydrological applications, though it is less commonly applied than ensemble methods [53].

2.5.5. Training Procedure and Cross-Validation

All models were trained on 70% of the data ($n \approx 35,000$), with 15% reserved for validation (hyperparameter tuning, $n \approx 7500$) and 15% for independent testing (final performance evaluation, $n \approx 7500$). Data splitting was performed using stratified random sampling to ensure balanced representation of seasons, sites, and soil moisture ranges.

Hyperparameter optimization utilized Optuna [54], a Bayesian optimization framework that iteratively refines the search space based on prior trial performance. For each model, 100 optimization trials were conducted using 5-fold cross-validation on the training set, with the objective function defined as:

$$\text{Objective} = \text{minimize}(\text{RMSE}) + \text{maximize}(R^2)$$

Early stopping was enabled for gradient boosting models (XGBoost) to prevent overfitting, halting training when validation error failed to improve for 20 consecutive iterations.

All input features were standardized (zero mean, unit variance) prior to model fitting to ensure comparability across variables with different scales and units. Standardization is particularly critical for distance-based methods (KNN, SVM) and improves convergence for gradient-based algorithms (XGBoost).

2.6. Feature Importance Analysis and Model Diagnostics

To assess predictor contributions and ensure model interpretability, feature importance was quantified for all algorithms:

A. Tree-Based Models (RF, XGBoost):

- RF: Permutation importance, measuring the decrease in model accuracy when each feature is randomly shuffled
- XGBoost: Gain-based importance (total reduction in loss function attributable to each feature) and SHAP (SHapley Additive exPlanations) values for instance-level contributions

B. Non-Tree Models (SVM, KNN, MARS):

- Permutation importance: Applied uniformly across all models for consistent comparison
- Partial dependence plots: Visualizing marginal effects of key predictors on soil moisture

Feature importance results are presented in Supplementary Figures S1–S3 and discussed in Section 4. Correlation analysis (Supplementary Figure S10) confirmed that while lagged soil moisture variables exhibit moderate-to-strong autocorrelation ($r = 0.6\text{--}0.8$), the multi-model framework distributes predictive skill across meteorological and topographic inputs, demonstrating that accuracy arises from feature synergy rather than redundancy alone.

2.7. Spatial Prediction and Operational Mapping

After model training and validation, optimal models were applied to generate wall-to-wall spatial predictions of daily root-zone soil moisture across the Vhembe District. For each 10 m pixel, predictions utilized:

1. TU Wien surface soil moisture from the corresponding pixel

2. Meteorological data interpolated from the nearest 1–3 SAWS stations using inverse distance weighting (maximum interpolation radius: 50 km; typical weights: 60%/25%/15% for nearest/second/third stations)
3. Topographic features extracted from the SRTM DEM at the pixel location

This approach extends calibration from the two ARC probe sites (Noordgrens and Sigonde) to the entire district by leveraging:

- Spatial generalization: Trained models capture relationships between satellite-observed surface moisture, atmospheric drivers, and root-zone dynamics that are governed by universal physical processes (infiltration, percolation, evapotranspiration)
- Local adaptation: Topographic predictors and interpolated meteorology account for site-specific conditions even in unmonitored locations

Important caveat on spatial generalization: While the framework assumes that relationships learned at calibration sites are governed by transferable physical processes, this assumption remains untested at the ~12,498 km² of non-monitored areas. Spatial predictions at uncalibrated locations should be treated as model-based estimates with uncertain quantitative accuracy rather than validated observations. The district-wide maps presented in this study demonstrate the technical feasibility of the spatial prediction workflow but do not constitute operational soil moisture products ready for irrigation management without independent field verification.

Prediction uncertainty was assessed using Random Forest’s ensemble variance (standard deviation across individual trees) and XGBoost’s quantile regression (5th and 95th percentiles). Spatial predictions are presented in Section 3.4.

To validate spatial predictions beyond ARC locations, we conducted:

- Leave-one-site-out cross-validation: Training on one site (e.g., Noordgrens), predicting the other (Sigonde), iteratively
- Qualitative spatial consistency checks: Comparing predicted moisture patterns against independent indicators (rainfall maps, topographic wetness)

Results (Section 3.4.2) demonstrate that RF and KNN models successfully captured spatial gradients and local heterogeneity, while MARS, SVM, and XGBoost showed reduced spatial discrimination in non-calibrated areas.

Table 2. Multiple machine learning algorithms.

ML Method	Representative Equation (with Citation)	Implementation for 10 cm Surface Validation
k-Nearest Neighbors (KNN)	$\hat{y} = (1/k) \sum_i y_{(i)}$ [50,55]	TU Wien surface soil moisture retrieval [cm ³ /cm ³]. Preprocessing: Standardization (z-score normalization). Hyperparameters tuned: Number of neighbors (k = 3–25), distance metric (Euclidean, Manhattan), weighting scheme (uniform, distance). Validation: 10-fold cross-validation. Purpose: Non-parametric baseline to assess whether simple proximity-based averaging can reproduce observed soil moisture from TU Wien estimates.
Random Forest (RF)	$\hat{y} = (1/T) \sum_t h_t(x)$ [45]	Predictor: TU Wien surface soil moisture retrieval [cm ³ /cm ³]. Preprocessing: Standardization. Hyperparameters tuned: Number of trees (100–1000), max depth (5–50), min samples per split (2–10), max features per split (sqrt, log2). Validation: 10-fold cross-validation with out-of-bag error estimation. Purpose: Ensemble learning to capture nonlinear transformations between TU Wien retrievals and ground-truth measurements while reducing overfitting through bootstrap aggregation.

Table 2. Cont.

ML Method	Representative Equation (with Citation)	Implementation for 10 cm Surface Validation
MARS (Multivariate Adaptive Regression Splines)	$f(x) = \sum_m c_m B_m(x)$ [52,53]	Predictor: TU Wien surface soil moisture retrieval [cm ³ /cm ³]. Preprocessing: Standardization. Hyperparameters tuned: Maximum basis functions (10–100), degree of interactions (1–3), pruning penalty (2–5). Validation: 10-fold cross-validation. Purpose: Piecewise linear regression to identify potential breakpoints or thresholds in the TU Wien retrieval-to-observed relationship (e.g., saturation effects, sensor limitations)
Gradient Boosting Machine (GBM/XGBoost)	$f^m(x) = f_{m-1}(x) + v h_m(x)$ [21,47]	Predictor: TU Wien surface soil moisture retrieval [cm ³ /cm ³]. Preprocessing: Standardization. Hyperparameters tuned: Number of trees (200–1000), max depth (3–15), learning rate η (0.01–0.3), subsample ratio (0.5–1.0), L1/L2 regularization. Validation: 10-fold cross-validation with early stopping. Purpose: Sequential error correction through boosting to iteratively refine the mapping from TU Wien retrievals to observed values, emphasizing regions where initial predictions were poor.
Support Vector Machine (SVM)	$f(x) = \sum_i \alpha_i y_i K(x_i, x) + b, K(x_i, x) = \exp(-\ x_i - x\ ^2)$ [48,51]	Predictor: TU Wien surface soil moisture retrieval [cm ³ /cm ³]. Kernel: Radial Basis Function (RBF). Preprocessing: Standardization (critical for kernel methods). Hyperparameters tuned: Regularization parameter C (0.1–100), kernel coefficient γ (e^{-4} to 1). Validation: 10-fold cross-validation. Purpose: Nonlinear regression via kernel transformation to project TU Wien retrievals into higher-dimensional space where the relationship with observed soil moisture may be more linear

NOTE: All models in Section 3.3.1 use only TU Wien θ [cm³ cm⁻³] as the predictor to isolate retrieval performance and provide a baseline before adding other features. Model skill (R², RMSE, MAE, NRMSE, MSE, MBE) is obtained via 10-fold cross-validation using predictions from held-out folds only. For root-zone depths in Section 3.3.2, the feature set was expanded to include lagged soil moisture, meteorological variables (with 1–3-day lags), and topographic attributes (elevation, slope, aspect, TWI). Five algorithms (KNN, RF, MARS, GBM, SVM) were tested to represent diverse learning paradigms and assess which best models the TU Wien-to-observed mapping.

Table 3. Hyperparameter tuning ranges for the five machine learning models.

Model	Hyperparameters	Search Range
SVM	Kernel	{RBF}
	Regularization parameter (C)	[0.1, 100]
	Kernel coefficient (γ)	[e^{-4} , 1]
KNN	Number of neighbors (k)	[3, 25]
	Distance metric	{Euclidean, Manhattan}
	Weighting scheme	{uniform, distance}
RF	Number of trees (n_estimators)	[100, 1000]
	Maximum tree depth	[5, 50]
	Minimum samples per split	[2, 10]
	Maximum features per split	{sqrt, log2}
XGB	Number of trees (n_estimators)	[200, 1000]
	Maximum depth	[3, 15]
	Learning rate (η)	[0.01, 0.3]
	Subsample ratio	[0.5, 1.0]

Table 3. Cont.

Model	Hyperparameters	Search Range
MARS	Column sampling ratio	[0.5, 1.0]
	L1/L2 regularization (α, λ)	[0, 10]
	Maximum number of basis functions	[10, 100]
	Degree of interactions	[1, 3]
	Penalty for adding terms (pruning penalty)	[2, 5]

2.8. Statistical Evaluation Analysis

The performance of the five machine learning models (KNN, RF, SVM, MARS, and XGB) in predicting SWC from Sentinel-1 TU Wien retrievals was evaluated using multiple complementary error metrics. Each metric captures a distinct aspect of model performance

- Pearson correlation coefficient (r , Equation (4)): Measures the strength of linear association between predicted and observed SWC
- Coefficient of determination (R^2 , Equation (5)): Quantifies the proportion of variance explained by the model
- Root Mean Square Error (RMSE, Equation (6)): Captures the average magnitude of prediction errors, sensitive to large deviations
- Normalized RMSE (NRMSE, Equation (7)): Expresses RMSE relative to the mean observed value for dimensionless comparisons across depths and sites
- Mean Squared Error (MSE, Equation (8)): Provides the mean of squared errors
- Mean Absolute Error (MAE, Equation (9)): Gives a robust average error less influenced by outliers
- Mean Bias Error (MBE, Equation (10)): Reveals systematic over- or underestimation

Together, these metrics permit a comprehensive assessment of both the magnitude and direction of errors, supporting rigorous validation of Sentinel-1-based SWC predictions for irrigation management [56–59].

$$r = \frac{n \sum_{i=1}^n (y_i \hat{y}_i) - \sum_{i=1}^n y_i \sum_{i=1}^n \hat{y}_i}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (y_i - \hat{y}_i)^2}} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}}{(\bar{y})} \quad (7)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (10)$$

where: y_i is predicted soil moisture [cm^3/cm^3]; (y_i) is the measured soil moisture [cm^3/cm^3]; \bar{y} is the average of the measured soil moisture; n is the number of field samples used for validation.

All performance metrics reported in this study were calculated exclusively on an independent 15% test set ($n \approx 7500$ observations) that was held out from both training and hyperparameter tuning. This holdout set was stratified by season, site, and soil moisture range to ensure representative sampling. Additionally, leave-one-site-out cross-validation was performed by training on one ARC location (e.g., Noordgrens) and predicting the other (Sigonde), iteratively, to assess model transferability (results presented in Section 3.4.2) Model performance on training and validation sets is provided in Supplementary Table S2 for transparency.

2.9. Soil Moisture Dimensionality Reduction and Feature Importance Analysis

To reduce dimensionality and capture dominant variance in soil moisture across depths, Principal Component Analysis (PCA) was applied to the standardized daily volumetric soil moisture measurements from six depths (10, 20, 40, 60, 80, and 100 cm) [60]. The soil moisture data matrix X with n observations and $p = 6$ variables was centered and scaled prior to PCA to ensure equal weighting of each depth. PCA decomposes X into orthogonal principal components via eigen decomposition of the covariance matrix (Equation (11)):

$$C = (1/(n - 1)) X_T X = A \Lambda A_T \quad (11)$$

where C is the covariance matrix, $A = [a_1, \dots, a_p]$ contains the eigenvectors, and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ are the eigenvalues ordered such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$.

The first principal component (PC1) for each observation, representing the linear combination of soil moisture depths corresponding to the maximum variance direction, was calculated as (Equation (12)):

$$\text{PC1} = X a_1 \quad (12)$$

where a_1 is the eigenvector associated with the largest eigenvalue λ_1 . PC1 served as an integrated root-zone soil moisture index synthesizing the multi-depth profile, mitigating multicollinearity, and highlighting dominant variability. This PC1 index was employed as an alternative target variable for machine-learning model evaluation in Section 3.3.4, demonstrating models' capacity to reproduce integrated root-zone dynamics. PC1 typically explained 75–85% of total variance in the soil moisture profile, confirming its effectiveness as a summary metric.

2.10. Software & Tools

SAR data preprocessing was primarily conducted using GEE, leveraging its robust, scalable cloud computing platform for efficient batch processing and standardized Sentinel-1 workflows. ESA SNAP 8.0 was also utilized for complementary Sentinel-1 preprocessing steps such as calibration and terrain correction.

Machine learning model development, hyperparameter optimization, and evaluation were implemented entirely in R (version RStudio-2025.09.0-387), using key packages including caret, xgboost, earth, Random Forest, and e1071. The reticulate package facilitated interfacing with Python 3.11-based tools, when necessary, though core ML workflows relied on R's native environment due to its flexibility and broad environmental modeling community support.

Geospatial analyses and map visualizations were performed using a combination of ArcGIS Pro 3.2 and R's spatial ecosystem (terra, sf, ggplot2, patchwork), balancing ArcGIS's advanced spatial data handling capabilities with R's reproducible statistical mapping functions. This integrated software environment supported seamless data management, structured ML workflows, and high-quality visualization of soil moisture dynamics.

Detailed software versions, package dependencies, and environment setup scripts are provided in Supplementary Table S1 to facilitate replication. All code used in this study

is available upon request and will be deposited in a public repository (GitHub/Zenodo) upon manuscript acceptance.

3. Results

3.1. Meteorological Conditions

Interannual variability in rainfall, temperature, humidity, and wind speed across five maize growing seasons (2017/18–2021/22) is illustrated in Figure 3. Total growing-season rainfall was approximately 555 mm (2017/18), 403 mm (2018/19), 475 mm (2019/20), 507 mm (2020/21), and 398 mm (2021/22). Higher-rainfall seasons (2017/18, 2019/20, 2020/21) produced above-average soil moisture recharge and reduced irrigation demand, while 2018/19 and 2021/22 exhibited lower totals and higher temperature extremes, driving pronounced moisture deficits. Maximum daily temperatures routinely exceeded 35 °C during the drier seasons. Relative humidity tracked rainfall events closely, and wind speed anomalies during drier years likely intensified evapotranspiration. These meteorological patterns collectively shaped the seasonal soil moisture dynamics examined in subsequent sections.

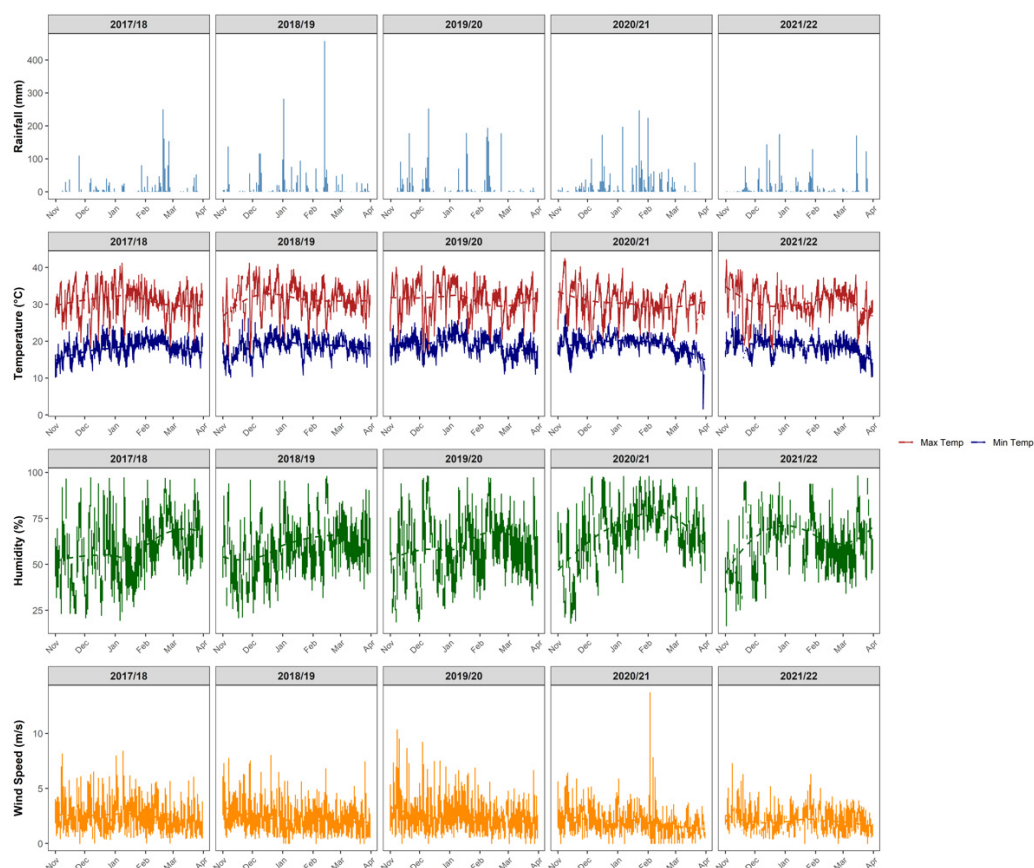


Figure 3. Meteorological trends during maize growing seasons in Vhembe (2017–2022): temporal patterns of rainfall, temperature extremes, relative humidity, and wind speed shown for each season. Panels illustrate interannual climate variability and its potential influence on soil moisture dynamics, crop water use, and agronomic management.

3.2. In Situ Soil Moisture Observations from ARC Probes

Soil moisture profiles from the two ARC monitoring stations revealed distinct depth-specific and site-specific patterns strongly governed by contrasting soil textures (Figures 4 and 5). At Noordgrens (sandy loam), upper layers (10–40 cm) exhibited marked seasonal variability—rapid depletion and quick post-rainfall recharge—consistent with high hydraulic conductivity and low water-retention capacity. Deeper layers (60–100 cm)

showed dampened fluctuations reflecting slower vertical drainage. At Sigonde (clay loam), moisture retention was greater across all depths, with slower depletion rates and smoother seasonal cycles, reflecting the higher field capacity and lower hydraulic conductivity of clay-loam soils. Both sites showed maximum moisture values during peak rainfall months (December–February) and minima in dry-season months (July–September), with pronounced interannual variability aligned with the meteorological patterns described in Section 3.1.

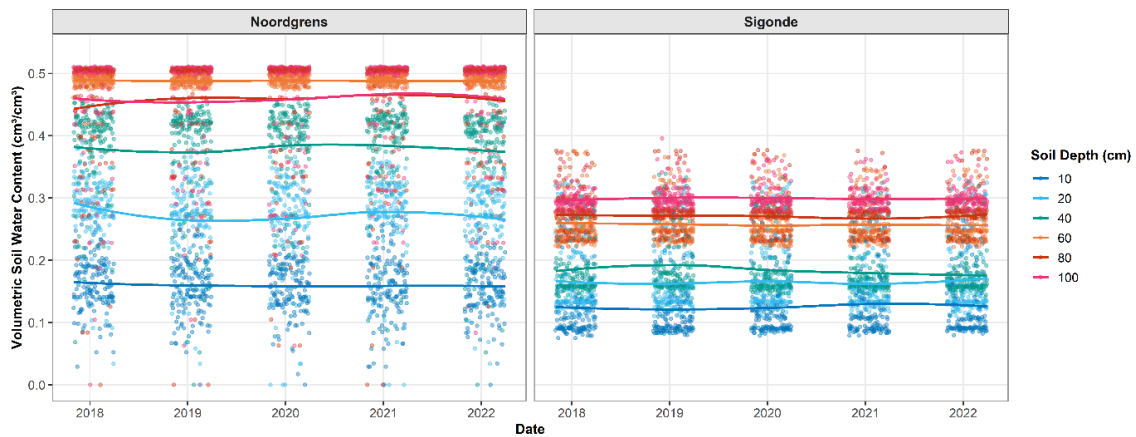


Figure 4. Temporal dynamics of in situ volumetric soil water content (cm^3/cm^3) at six depths (10, 20, 40, 60, 80, 100 cm) for Noordgrens and Sigonde monitoring stations during maize growing seasons (November–March, 2017–2022). Panels show daily measurements from Agricultural Research Council (ARC) capacitance probes, illustrating depth-specific moisture variability and seasonal recharge-depletion cycles.

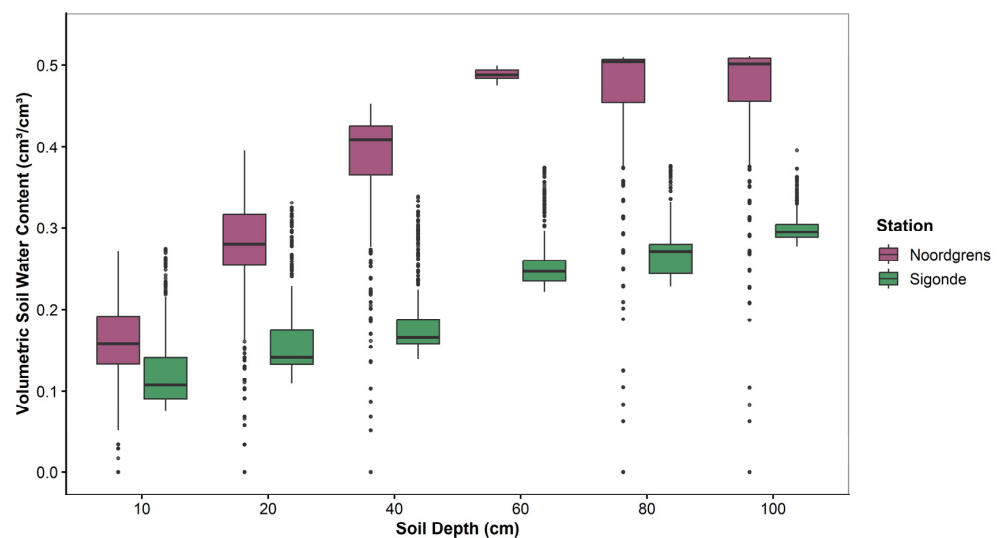


Figure 5. Statistical distribution of volumetric soil water content (cm^3/cm^3) by soil depth at Noordgrens and Sigonde stations during maize growing seasons (November–March, 2017–2022). Boxplots display median (central line), interquartile range (box), $1.5 \times$ interquartile range whiskers, and outliers (points) for each depth layer (10–100 cm). Data aggregated from daily ARC probe measurements ($n \approx 8760$ observations per site).

Satellite-to-In Situ Correlation Analysis

Comparison of TU Wien surface retrievals against ARC probe measurements revealed strong correlations at 10 cm depth at both stations (Noordgrens: $r = 0.85$, $\text{RMSE} = 0.036 \text{ cm}^3/\text{cm}^3$; Sigonde: $r = 0.84$, $\text{RMSE} = 0.067 \text{ cm}^3/\text{cm}^3$), but systematic signal decay with increasing depth (Figure 6A). At 20 cm, correlations declined to $r = 0.43$ – 0.47 ,

falling below $r = 0.2$ at 40 cm and becoming negligible or negative ($r = -0.08$ to 0.14) below 60 cm at both sites. Despite similar correlation profiles, absolute error magnitudes diverged substantially between stations: Sigonde maintained RMSE below $1.3 \text{ cm}^3/\text{cm}^3$ across the full profile (10–100 cm), whereas Noordgrens exhibited progressive error inflation from $0.36 \text{ cm}^3/\text{cm}^3$ at 10 cm to $2.89 \text{ cm}^3/\text{cm}^3$ at 100 cm.

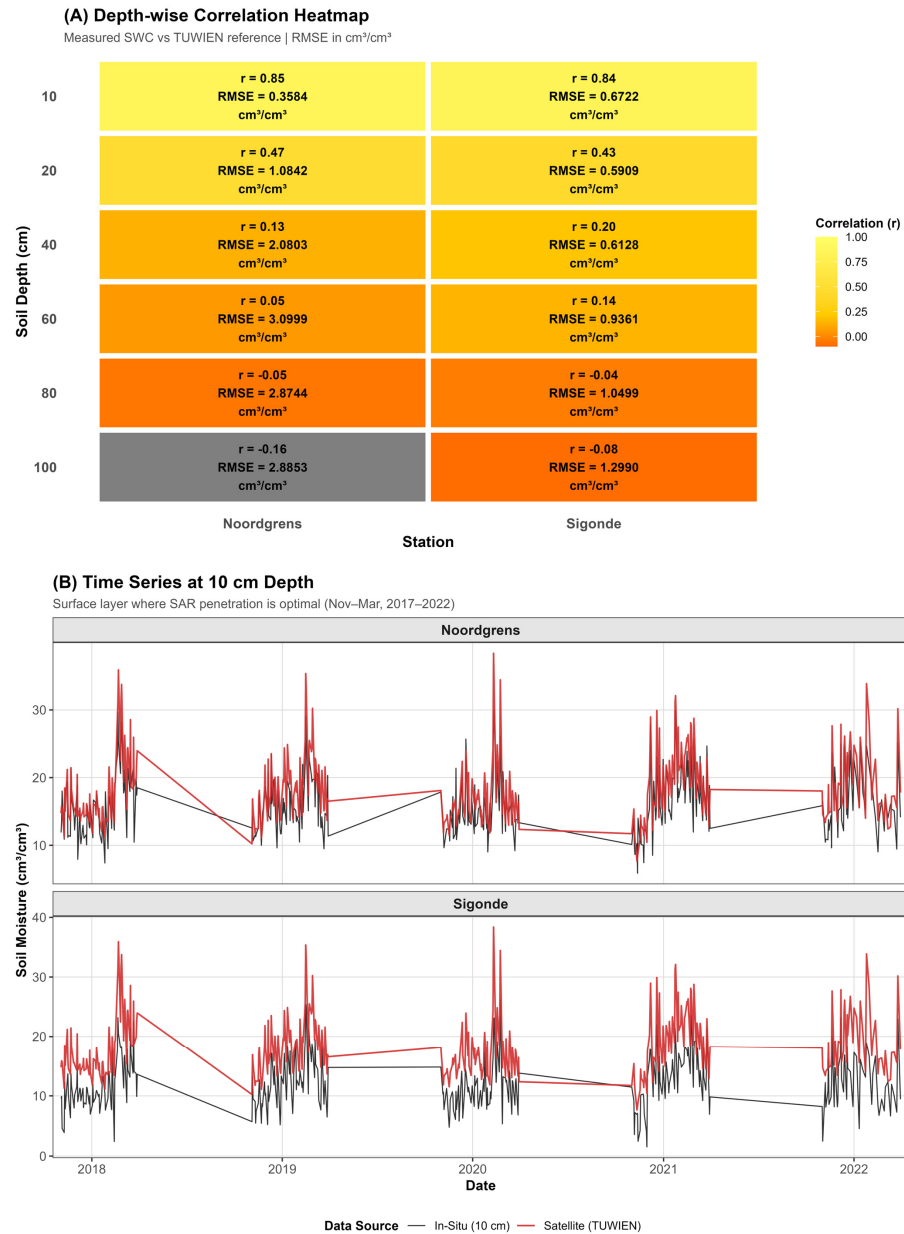


Figure 6. Depth-dependent SAR soil moisture retrieval performance. **(A)** Correlation heatmap showing systematic signal decay with depth at both stations, with RMSE values in cm^3/cm^3 . Strong correlations ($r > 0.8$, yellow-green shading) observed at 10 cm depth, declining to weak or negative correlations ($r < 0.2$, orange shading) at deeper layers (60–100 cm). **(B)** Time series validation at 10 cm depth where SAR penetration is optimal, comparing satellite-derived (TUWIEN, red) and in situ measurements (black) from November to March, 2017–2022. All RMSE values are expressed as volumetric soil moisture (cm^3/cm^3).

These contrasting error profiles—similar temporal dynamics but divergent absolute accuracy—reflect site-specific differences in vegetation density and soil surface conditions. They directly motivate the two-stage EO–ML approach employed in this study: machine learning compensates for depth-related SAR signal attenuation by incorporating lagged

moisture and meteorological variables at greater depths, while leveraging direct satellite information at surface layers (10–20 cm). The physical interpretation of these site differences and their implications for model transferability are discussed in Section 4.

3.3. Machine Learning Model Performance for Root-Zone Soil Moisture Prediction

3.3.1. Surface Soil Moisture Calibration (10 cm Depth)

All five ML models demonstrated excellent agreement between TU Wien surface retrievals and ARC probe observations at 10 cm depth (Figure 7), with $R^2 \geq 0.97$ across all algorithms. These results reflect statistical calibration performance at the two ARC sites, where near-surface SAR backscatter (sensitive to the top 2–5 cm) was empirically related to 10 cm probe measurements via the TU Wien change detection framework; the high R^2 values reflect strong temporal covariance between surface and shallow subsurface moisture rather than direct C-band measurement at 10 cm. Random Forest achieved $R^2 = 0.969$ (RMSE = $0.016 \text{ cm}^3/\text{cm}^3$, MAE = $0.013 \text{ cm}^3/\text{cm}^3$); SVM, XGBoost, MARS, and KNN showed similar accuracy ($R^2 = 0.976$; RMSE = $0.014\text{--}0.015 \text{ cm}^3/\text{cm}^3$; MAE = $0.011\text{--}0.012 \text{ cm}^3/\text{cm}^3$). Bias was negligible at both sites (MBE ≈ 0), and normalized RMSE values ($0.051\text{--}0.058$) fell well below the 0.10 threshold required for precision irrigation applications. Comparing ML-calibrated results ($R^2 \geq 0.97$, RMSE $\leq 0.016 \text{ cm}^3/\text{cm}^3$) against raw TU Wien performance ($r = 0.84\text{--}0.85$, RMSE = $0.036\text{--}0.067 \text{ cm}^3/\text{cm}^3$) demonstrates a 55–76% RMSE reduction attributable to ML calibration.

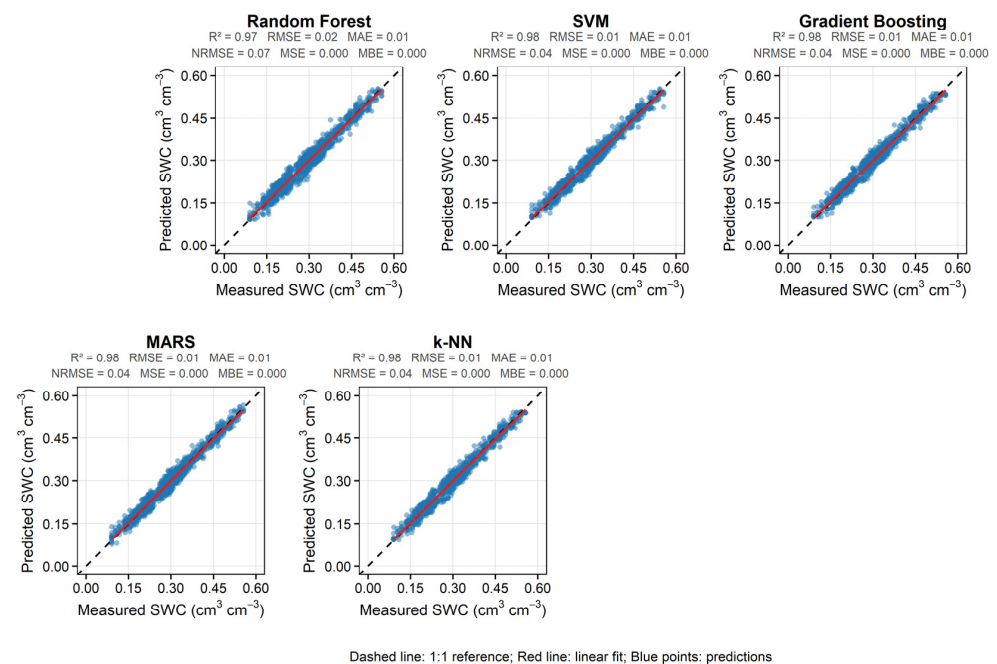


Figure 7. Comparison between Predicted and Observed Soil Water Content (SWC) at 10 cm Depth for the 2017–2022 Period Using TU Wien as Predictor. The scatterplots display 10-fold cross-validated daily interpolations for five machine learning models: Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), Multivariate Adaptive Regression Splines (MARS), and K-Nearest Neighbors (KNN). Model performance metrics are shown in each panel, including coefficient of determination (R^2), mean squared error (MSE), mean absolute error (MAE), mean bias error (MBE), root mean square error (RMSE), and normalized RMSE (NRMSE). Units are volumetric soil moisture (cm^3/cm^3).

3.3.2. Root-Zone Predictions (20–100 cm Depth)

The ML framework extends TU Wien surface retrievals to root-zone depths using antecedent moisture conditions (1–3-day lags), meteorological forcing, and topographic

controls; the physical assumptions and limitations of this statistical extrapolation are discussed in Section 4.

Model performance varied systematically with depth and algorithm (Figure 8). RF delivered the highest accuracy across the full profile, peaking at 60 cm ($R^2 = 0.989$, $nRMSE = 0.015$) with RMSE consistently below $0.013 \text{ cm}^3/\text{cm}^3$ and negligible bias ($MBE = -0.0002$ to $0.00004 \text{ cm}^3/\text{cm}^3$); all $nRMSE$ values remained <0.05 . XGBoost showed outstanding performance across the entire profile ($R^2 = 0.966$ – 0.988 , $nRMSE = 0.016$ – 0.047 , $RMSE = 0.006$ – $0.014 \text{ cm}^3/\text{cm}^3$). SVM-Radial achieved strong accuracy at all depths ($R^2 = 0.856$ – 0.958 , $RMSE = 0.012$ – $0.025 \text{ cm}^3/\text{cm}^3$), with optimum performance at 60 cm ($R^2 = 0.958$, $nRMSE = 0.03$). KNN reached peak skill at 60 cm ($R^2 = 0.946$, $nRMSE = 0.033$), with a modest decline at 100 cm ($R^2 = 0.842$, $nRMSE = 0.065$). MARS maintained acceptable accuracy at core depths (60 cm: $R^2 = 0.954$, $nRMSE = 0.031$) with zero bias across all layers ($MBE = 0$).

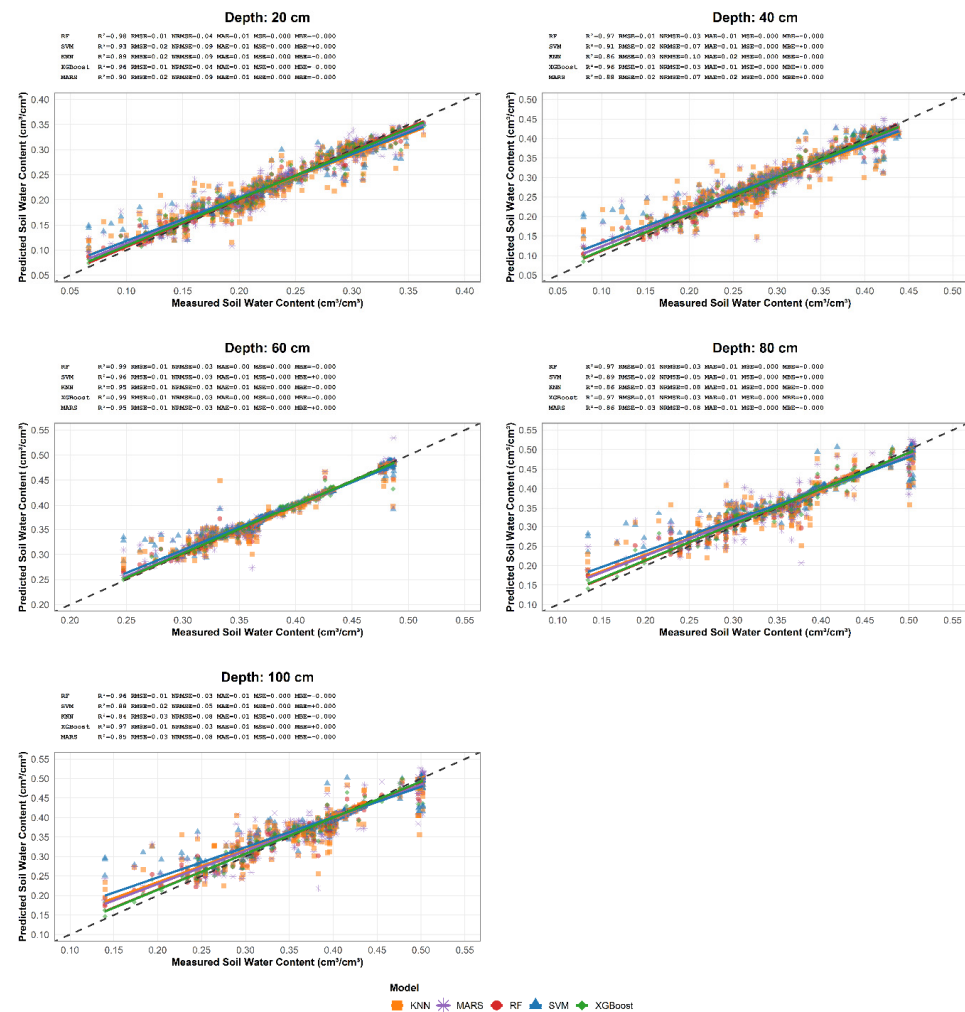


Figure 8. Predicted versus observed daily soil moisture content (cm^3/cm^3) at depths 20–100 cm at Noordgrens and Sigonde ARC monitoring stations (2017–2022) using 5 ML models. Model performance metrics include mean bias error (MBE), mean squared error (MSE), root mean squared error (RMSE), coefficient of determination (R^2), and normalized RMSE ($nRMSE$) for each soil depth. All metrics represent independent test set (15% holdout) validation performance at the two calibration sites and do not directly validate spatial predictions at non-monitored locations across the district (see Section 3.4.2 for spatial validation assessment).

All metrics reported above are based on an independent 15% test set held out from both training and hyperparameter optimization at the two ARC calibration sites; they

do not directly validate spatial predictions at non-monitored locations across the district (see Section 3.4.2).

3.3.3. Comparison of Daily Mean Predictions Across All Depths

When aggregating predictions across all depths (Figure 9), RF exhibited the strongest agreement with measured values ($R^2 = 0.96$, $RMSE = 0.023 \text{ cm}^3/\text{cm}^3$, $nRMSE = 0.04$), followed closely by other algorithms (MARS, XGBoost, KNN, SVM: $R^2 = 0.86\text{--}0.88$, $RMSE = 0.036\text{--}0.047 \text{ cm}^3/\text{cm}^3$, $nRMSE = 0.06$). All models exceeded operational suitability thresholds ($R^2 > 0.85$, $nRMSE < 0.06$) for irrigation scheduling applications.

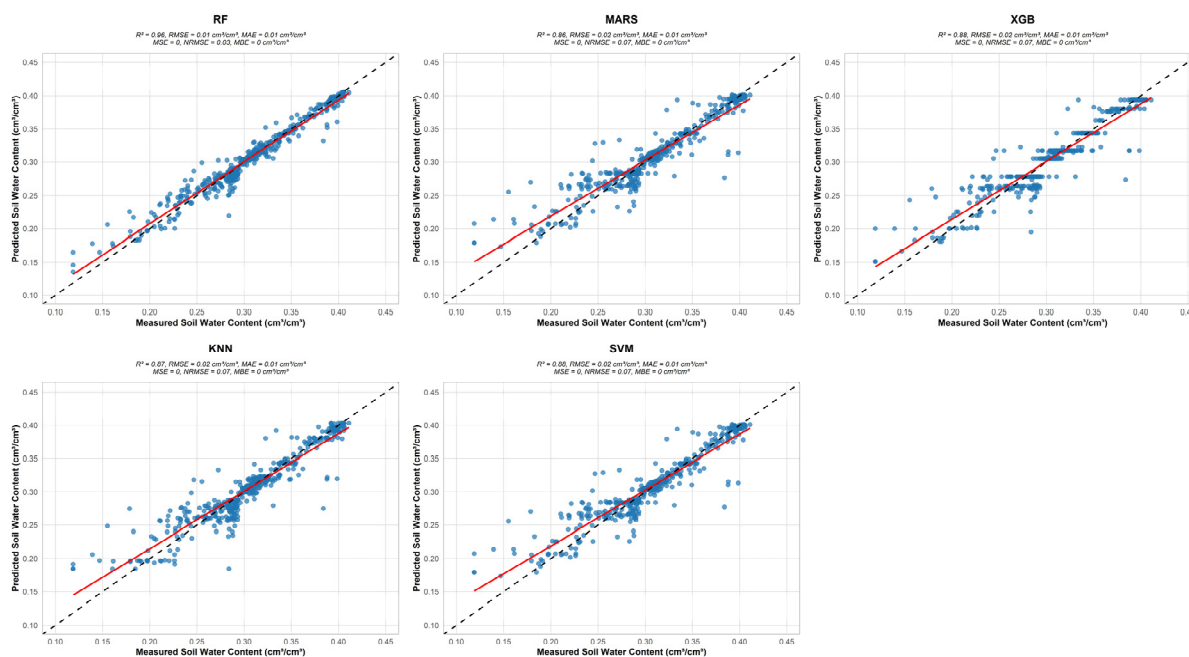


Figure 9. Daily mean predicted versus observed soil water content (cm^3/cm^3) aggregated across all depths in Vhembe (2017–2022), comparing five machine learning models: Random Forest (RF), Multivariate Adaptive Regression Splines (MARS), Extreme Gradient Boosting (XGB), k-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Panels show model fit statistics: R^2 , RMSE, MAE, MSE, $nRMSE$, and MBE.

3.3.4. Root-Zone Integration via Principal Component Analysis

PCA-based predictions (Figure 10) demonstrated each model’s ability to reproduce a single integrated root-zone moisture index (PC1, explaining 75–85% of total soil profile variance). Random Forest achieved outstanding accuracy ($R^2 = 0.957$, $RMSE = 0.005 \text{ cm}^3/\text{cm}^3$, $nRMSE = 0.219$); XGBoost and KNN showed strong consistency ($R^2 \approx 0.88\text{--}0.92$); MARS and SVM yielded satisfactory fits ($R^2 \approx 0.87\text{--}0.88$). The $nRMSE$ values below 0.10 achieved by RF, XGBoost, and KNN for depth-specific predictions indicate sufficient precision for automated irrigation scheduling. Leave-one-site-out cross-validation (Supplementary Table S2) confirmed partial model transferability: RF and KNN maintained acceptable accuracy ($R^2 = 0.82\text{--}0.87$) when predicting one site from training on the other, while MARS, SVM, and XGBoost showed greater sensitivity to site-specific conditions ($R^2 = 0.65\text{--}0.74$). Feature importance analyses (Supplementary Figures S1–S3) confirmed that predictive skill arises from feature synergy, lagged moisture variables, meteorological predictors, and topographic inputs all contribute, rather than temporal autocorrelation alone.

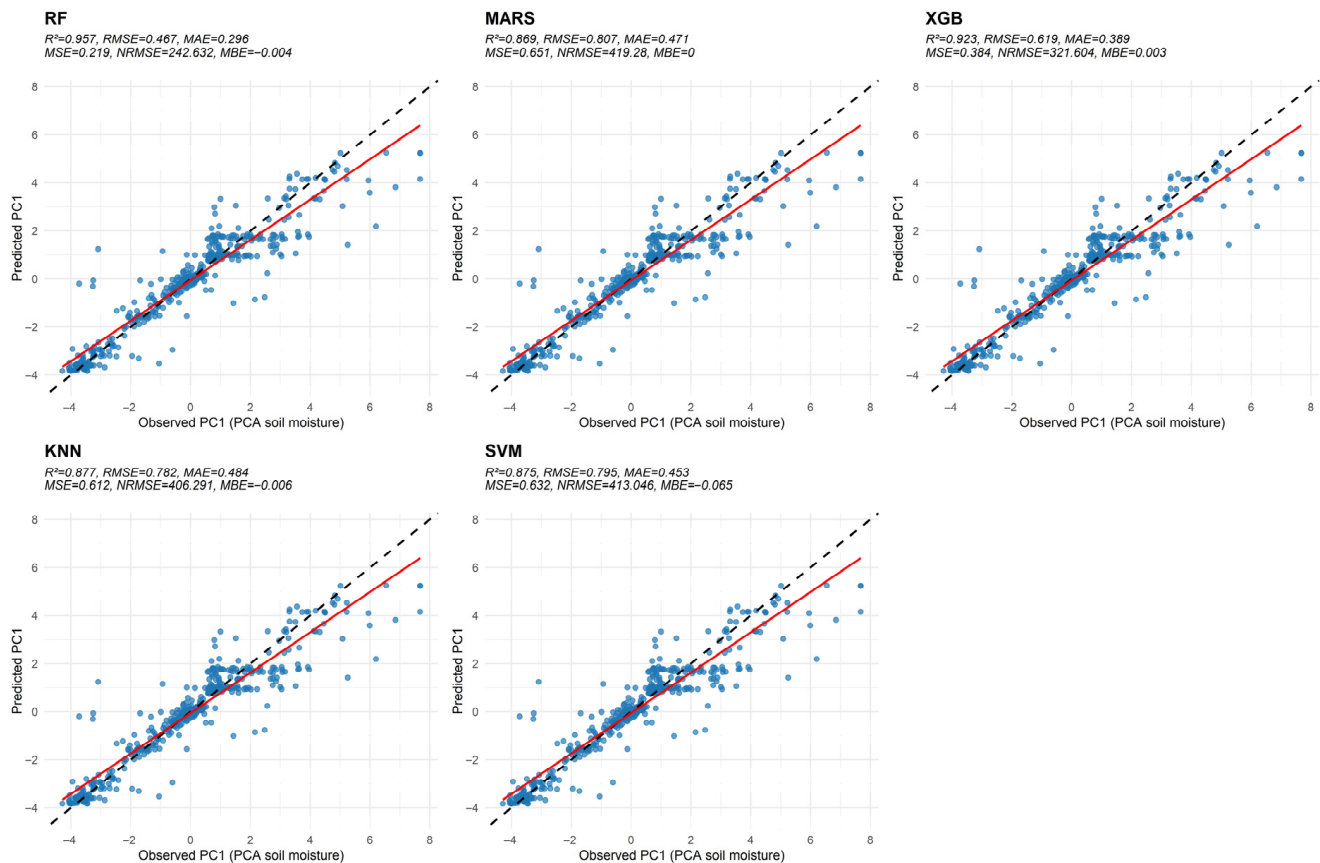


Figure 10. PCA-based soil moisture prediction (PC1) across five machine learning models (Random Forest [RF], Multivariate Adaptive Regression Splines [MARS], Extreme Gradient Boosting [XGB], k-Nearest Neighbors [KNN], and Support Vector Machine [SVM]) for the Vhembe district (2017–2022). Each panel compares predicted versus observed first principal component (PC1) scores, summarizing multi-depth soil moisture dynamics.

3.4. Spatiotemporal Analysis of Soil Moisture Dynamics

3.4.1. Seasonal Surface Soil Moisture Patterns (2017–2022)

High-resolution (10 m) surface soil moisture maps (Figure 11) revealed clear interannual and spatial gradients across the district. Southern and central Vhembe consistently showed elevated moisture (orange tones), while peripheral and northern areas (blue tones) exhibited persistently drier conditions and elevated drought risk. Wetter seasons (2019/20, 2020/21) produced visible expansion of well-watered zones; drier years (2018/19, 2021/22) were characterized by larger areas failing to meet optimal thresholds. Spatial patterns show strong qualitative coherence with topographic controls and drainage features, though quantitative accuracy at non-ARC pixels relies on the assumption that trained relationships between TU Wien surface moisture, meteorology, and topography remain consistent across the district.

3.4.2. Spatial Prediction Performance and Generalization

Leave-one-site-out cross-validation confirmed partial model transferability: RF and KNN maintained acceptable accuracy between sites ($R^2 = 0.82$ – 0.87 , $RMSE = 0.028$ – 0.035 cm^3/cm^3), while MARS, SVM, and XGBoost showed degraded performance ($R^2 = 0.65$ – 0.74 , $RMSE = 0.042$ – 0.058 cm^3/cm^3). Spatial moisture maps for 2017–2022 (Figure 12) demonstrate that RF and KNN captured distinct soil moisture gradients and local heterogeneity, accurately delineating wetter corridors in Makhado and Thulamela municipalities. The

superior performance of RF and KNN reflects their capacity to leverage generalizable hydrological relationships.

Vhembe District: Seasonal Soil Water Content (10m Resolution)

Sentinel-1 TU Wien Surface Soil Moisture (0–10 cm depth, 2017–2022)

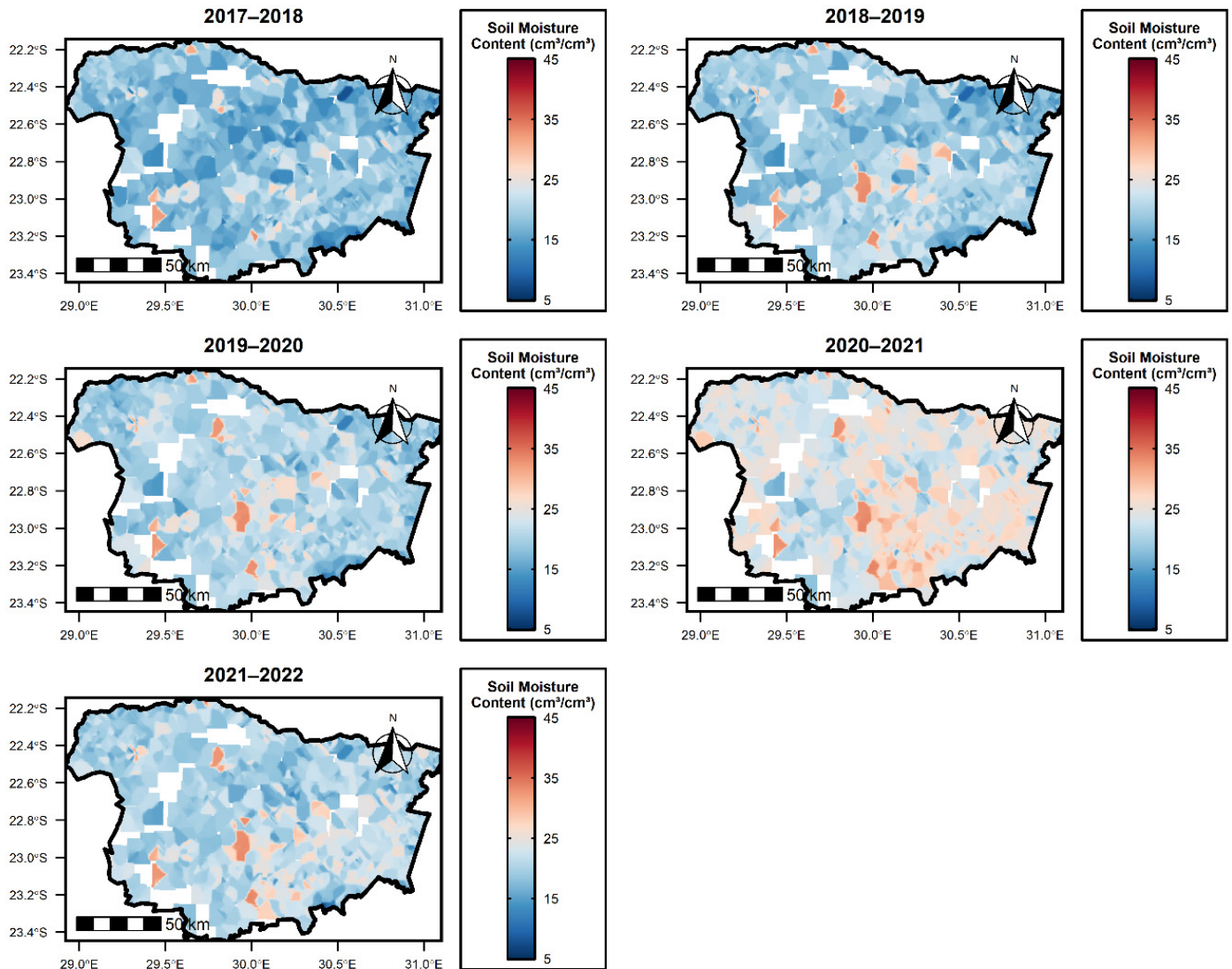


Figure 11. Seasonal maps of surface soil water content (SWC, 0–10 cm; cm^3/cm^3) for the Vhembe District, 2017–2022, at 10 m resolution. Each panel shows mean TU Wien Sentinel-1 SWC for one growing season, local neighborhood interpolated from satellite retrievals.

Quantitative validation was limited to the two ARC probe locations; spatial accuracy at the remaining $\sim 12,498 \text{ km}^2$ of Vhembe was assessed qualitatively against the topographic wetness index, seasonal rainfall accumulation, and Sentinel-2 NDVI. While predicted patterns showed strong qualitative coherence with these independent indicators, this does not substitute for quantitative ground-truth validation. District-wide soil moisture predictions should be interpreted as model-based spatial estimates with unknown quantitative accuracy at non-monitored pixels.

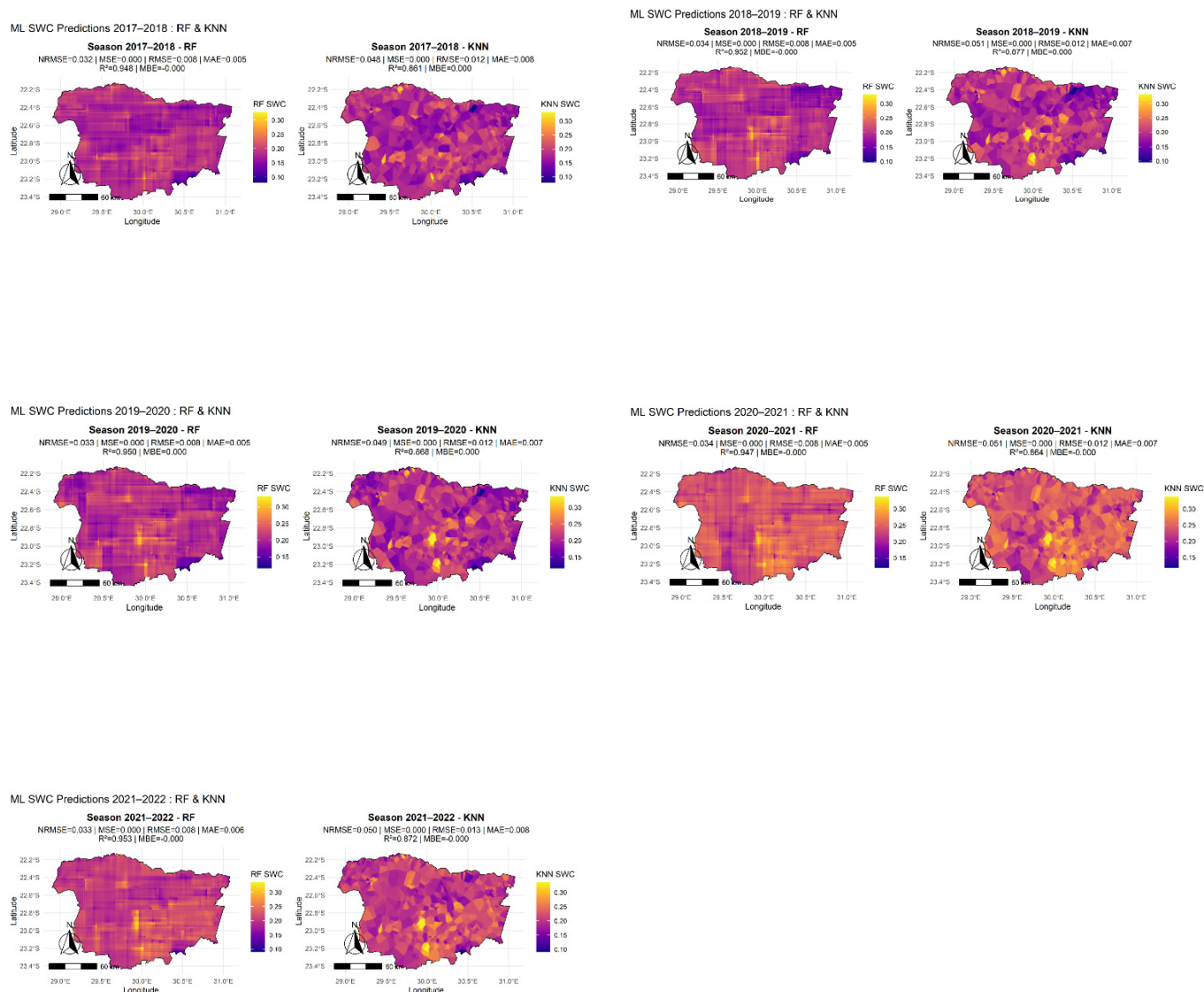


Figure 12. Seasonal mean volumetric soil moisture (SWC, cm³/cm³) predictions for Vhembe district (2017–2022) using Random Forest (RF) and k-Nearest Neighbors (KNN) models (Best selected performing models). Maps show spatial patterns of daily mean SWC across five summer growing seasons, illustrating temporal dynamics and interannual variation critical for irrigation planning.

3.4.3. Irrigation Decision Support Mapping

The irrigation decision framework (Figure 13) classifies RF-predicted soil moisture into five management categories based on empirically derived hydraulic thresholds (Critical Stress: <0.12 cm³/cm³; Irrigate Soon: 0.12–0.23 cm³/cm³; Monitor: 0.23–0.25 cm³/cm³; Optimal: 0.25–0.35 cm³/cm³; Above Optimal: >0.35 cm³/cm³). These thresholds were derived from the 5th and 95th percentiles of multi-year ARC probe observations and represent operational irrigation decision points rather than universal soil hydraulic constants; spatial application requires field verification.

During 2017–2018, 93.9% of the district required near-term irrigation (only 2.1% optimal). Comparable patterns persisted in 2018–2019 (88.5% requiring irrigation, 4.7% optimal) and 2021–2022 (79.1% requiring irrigation, 7.7% optimal). The 2020–2021 season was the most favorable, with only 31.5% requiring irrigation while 38.8% maintained optimal moisture. Critically stressed areas remained minimal throughout (<1%). For operational scheduling, shallow-layer moisture (20–40 cm) should be prioritized for short-term decisions, with deeper layers (60–100 cm) used as indicators of seasonal drought resilience.

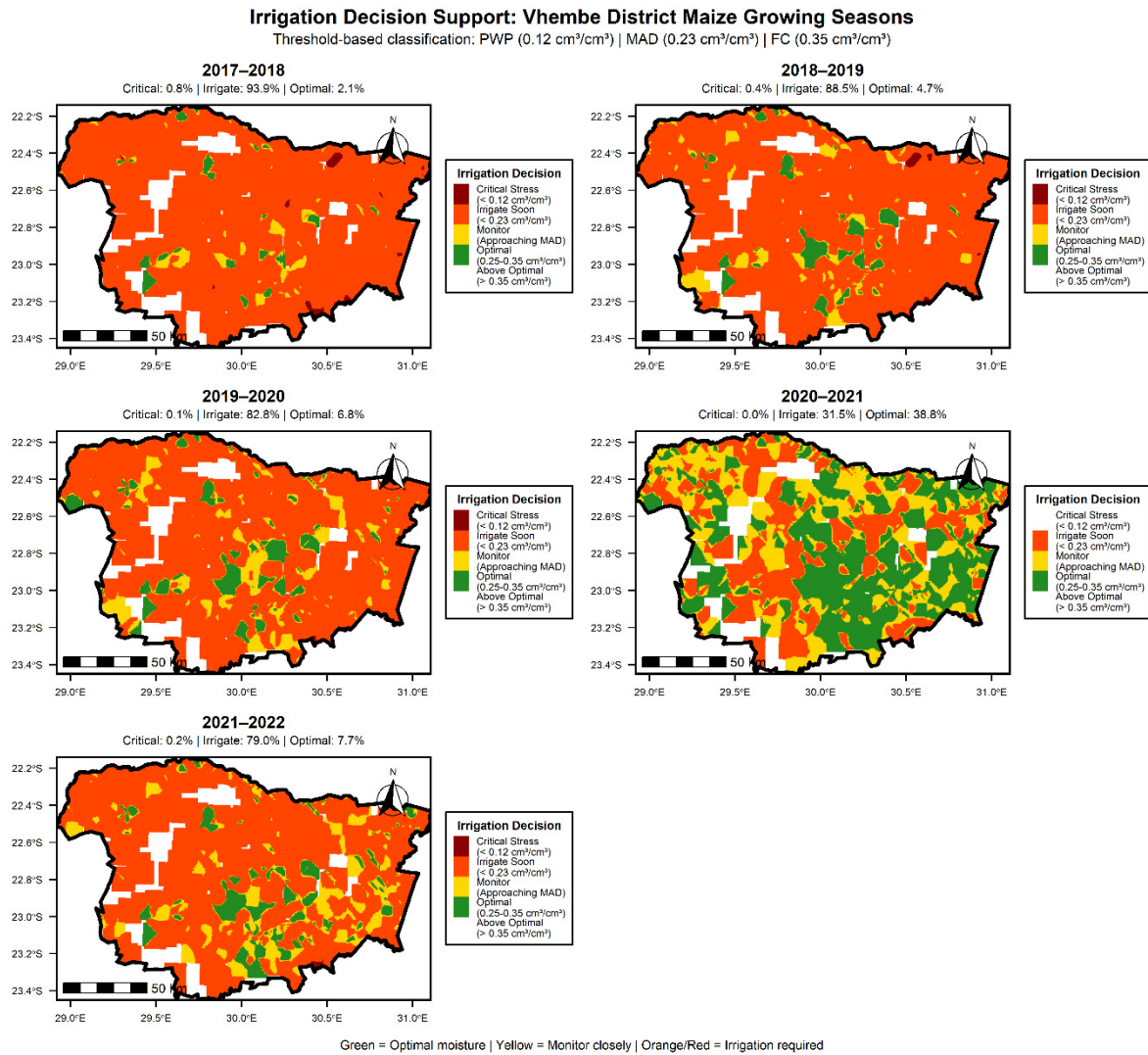


Figure 13. Prototype Irrigation Decision Support Framework for Vhembe District Maize Growing Seasons (2017–2022). Spatial distribution maps showing irrigation decision thresholds based on soil moisture classification derived from permanent wilting point ($PWP = 0.12 \text{ cm}^3/\text{cm}^3$), management allowable depletion ($MAD = 0.23 \text{ cm}^3/\text{cm}^3$), and field capacity ($FC = 0.35 \text{ cm}^3/\text{cm}^3$). Decision support categories include: Critical Stress ($<12\%$)—immediate irrigation required; Irrigate Soon ($12\text{--}23\%$)—schedule irrigation within 1–2 days; Monitor ($23\text{--}25\%$)—watch closely as irrigation may soon be needed; Optimal ($25\text{--}35\%$)—sufficient soil moisture, no irrigation needed; and Above Optimal ($>35\%$)—excess moisture, delay irrigation. Green areas represent optimal conditions, yellow indicates zones to monitor, and orange/red show areas requiring irrigation. White areas denote missing data. These classifications are based on models trained at two sites (Noordgrens and Sigonde) and require field verification before operational use in specific locations.

3.4.4. Water Deficit Quantification Across Growing Seasons

Water deficit analysis (Figure 14) estimated irrigation requirements relative to the MAD threshold ($0.23 \text{ cm}^3/\text{cm}^3$). Mean seasonal deficits ranged from 5.8 mm (2020–2021) to 50.5 mm (2017–2018), with maximum localized deficits exceeding 150 mm in northern and southeastern hotspots. Persistent deficit zones across multiple seasons indicate structural water availability constraints—likely related to soil texture, topographic position, and distance from water sources—that may require infrastructure investment beyond scheduling optimization alone. These deficit estimates inherit the spatial extrapolation uncertainties noted in Section 3.4.2 and require field verification before use in water allocation planning.

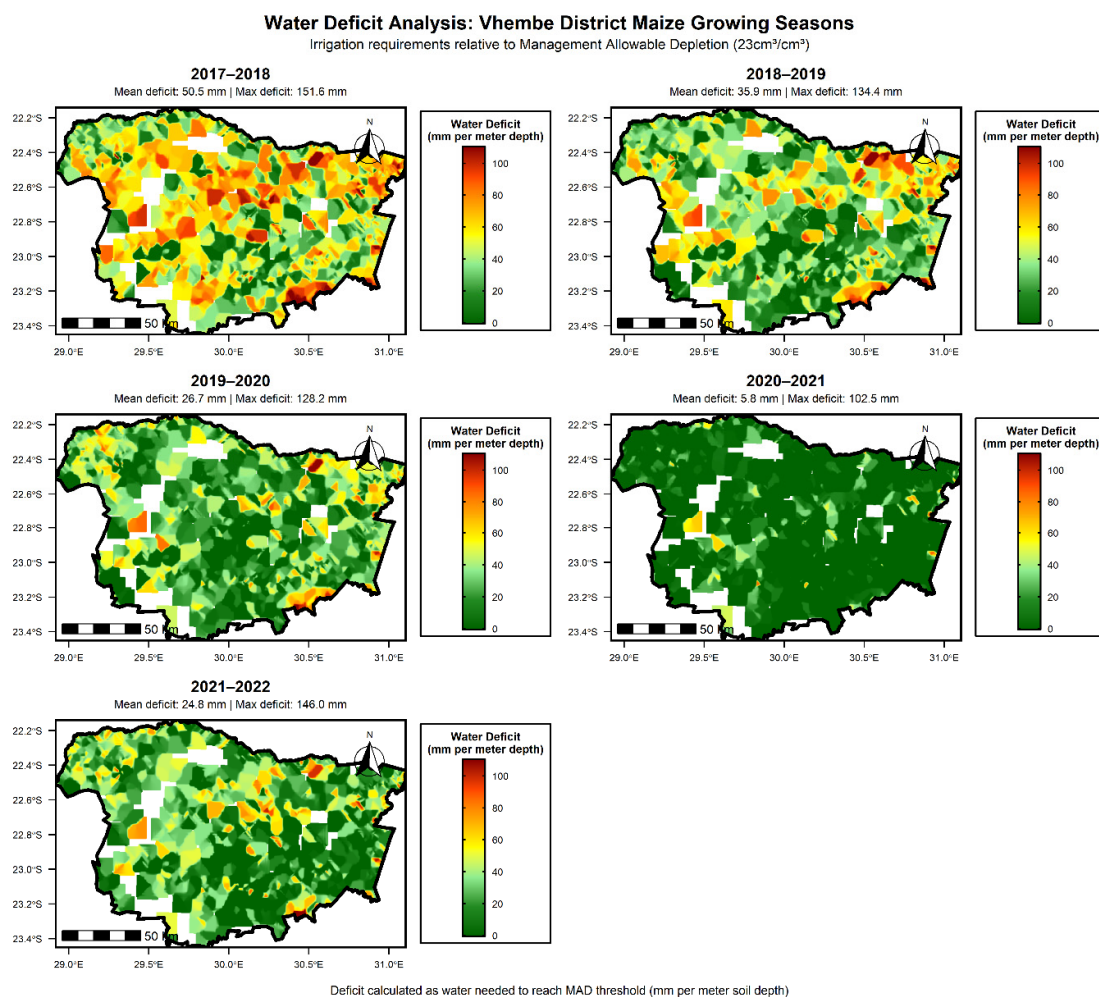


Figure 14. Model-Predicted Water Deficit Analysis for Vhembe District Maize Growing Seasons (2017–2022). Spatial representation of estimated irrigation water deficit (mm per meter soil depth) relative to the Management Allowable Depletion ($\text{MAD} = 0.23\text{ cm}^3/\text{cm}^3$). Deficit values are derived from Random Forest spatial predictions calibrated at two sites. Green areas indicate minimal or no deficit, while yellow to red gradients represent increasing water deficit levels. The mean and maximum deficits for each season are displayed above the respective maps, representing the estimated amount of water needed to restore soil moisture to the MAD threshold. Quantitative accuracy of deficit estimates at non-monitored locations requires field verification.

3.4.5. Feature Importance and Model Diagnostics

Feature importance analyses (Supplementary Figures S1–S3) and the correlation matrix (Supplementary Figure S10) revealed that while lagged soil moisture variables exhibited moderate-to-strong autocorrelation ($r = 0.6\text{--}0.8$), predictive skill was distributed across meteorological (rainfall, temperature, humidity) and topographic inputs. All performance metrics were calculated on the independent 15% test set ($n \approx 7500$ observations); Supplementary Figure S11 presents full test-set validation scatterplots for all models and depths, confirming robust out-of-sample generalization.

4. Discussion

This study demonstrated that integrating Sentinel-1 SAR, TU Wien retrievals, and meteorological data within a machine learning framework can generate temporally accurate, site-calibrated root-zone soil moisture estimates at two monitoring locations in smallholder maize systems—representing methodological progress over earlier surface-

limited or coarse-resolution approaches in terms of spatial resolution (10 m) and depth integration (0–100 cm), though validation remains limited to two sites [13,23].

A critical finding was the depth- and site-dependent SAR retrieval performance, revealing both the potential and limitations of C-band soil moisture estimation in semi-arid agricultural landscapes. While both stations demonstrated strong surface-layer correlations with TU Wien estimates ((Noordgrens: $r = 0.85$, $RMSE = 0.036 \text{ cm}^3/\text{cm}^3$; Sigonde: $r = 0.84$, $RMSE = 0.067 \text{ cm}^3/\text{cm}^3$ at 10 cm depth; Section 3.3.1), retrieval accuracy diverged substantially with depth and differed markedly in absolute error magnitudes between sites. At Sigonde, RMSE remained below $1.3 \text{ cm}^3/\text{cm}^3$ across the full profile (10–100 cm), whereas Noordgrens exhibited progressive error inflation ($0.36 \text{ cm}^3/\text{cm}^3$ at 10 cm increasing to $2.89 \text{ cm}^3/\text{cm}^3$ at 100 cm). This disparity persisted despite similar correlation decay patterns ($r < 0.2$ below 40 cm at both stations), indicating that temporal dynamics are captured equivalently, but absolute retrieval precision varies substantially with site characteristics.

The RMSE heterogeneity reflects well-documented physical constraints of C-band SAR soil moisture retrieval. Noordgrens' denser vegetation ($NDVI = 0.52 \pm 0.18$) causes volume scattering and attenuation effects that reduce microwave penetration and increase retrieval uncertainty [15,61], consistent with findings that woody cover $>30\%$ significantly degrades SAR soil moisture accuracy [40,41]. Sigonde's sparser canopy ($NDVI = 0.38 \pm 0.14$) and smoother clay-loam surface provide more favorable backscatter conditions [38,62]. These site differences also explain why the ML framework achieves comparable final performance ($R^2 = 0.96\text{--}0.97$) at both stations through fundamentally different predictor mechanisms: at Sigonde, the algorithm leverages genuine satellite-observed moisture signals, whereas at Noordgrens, greater reliance on meteorological predictors compensates for higher retrieval uncertainty at depth.

The systematic correlation decay below 20 cm (Figure 6A) aligns with theoretical C-band penetration limits of 5–10 cm in moist agricultural soils [63,64], confirming that SAR provides direct observations only of near-surface moisture. This is why the ML models predict root-zone moisture (20–100 cm) by statistically extrapolating from surface retrievals using antecedent moisture, meteorological forcing, and topographic controls—the framework does not explicitly model physical processes such as unsaturated hydraulic conductivity or root water uptake. Predictions, therefore, represent statistical inference from surface conditions and meteorological proxies rather than mechanistic simulation of vadose zone hydrology [65]. Consequently, leave-one-site-out cross-validation performance ($R^2 = 0.82\text{--}0.87$ for RF/KNN; Section 3.4.2) likely reflects the transferability of meteorological-hydrological relationships rather than independent SAR-based retrieval. District-wide predictions for root-zone moisture (>40 cm depth) depend primarily on interpolated weather data and moisture redistribution physics rather than direct Earth observation—a distinction that must be explicitly acknowledged in operational applications [66,67].

This heterogeneity in predictor contributions underscores a fundamental challenge in scaling EO–ML frameworks across heterogeneous agricultural landscapes. While our approach achieves uniformly high performance ($R^2 > 0.96$), the underlying mechanisms differ sufficiently that expanded validation across diverse soil-vegetation complexes is essential before claiming true spatial transferability [68,69]. Future work should explicitly quantify predictor importance variations across environmental gradients and develop ensemble approaches that weight satellite versus meteorological information adaptively based on real-time retrieval confidence metrics [70]. Ensemble models, particularly RF and XGBoost, surpassed the accuracies reported in related regional SAR–ML studies [17,21]. However, the R^2 values reflect strong temporal autocorrelation in soil moisture dynamics, amplified by 1–3 day lagged predictors. District-wide maps (Figures 10–13) should there-

fore be interpreted as model-based spatial estimates exhibiting strong qualitative coherence with landscape features but with uncertain quantitative accuracy at non-monitored pixels without additional ground validation. However, the current framework treats the agronomically relevant root zone as static across the growing season, whereas maize rooting depth increases progressively from 0 to 30 cm during early vegetative stages to 60–100 cm at maturity [25,44]. Future refinements incorporating dynamic rooting depth models would enable phenology-responsive scheduling, weighting shallow-layer predictions more heavily during early growth stages and progressively integrating deeper layers as roots develop.

The demonstrated ability to detect management-critical thresholds ($MAD = 0.23 \text{ cm}^3/\text{cm}^3$, $FC = 0.35 \text{ cm}^3/\text{cm}^3$) with high temporal accuracy ($R^2 > 0.96$, $nRMSE < 0.05$) at calibration sites suggests several operational pathways. The 10 m resolution enables field-scale monitoring tailored to individual smallholder plots (typically 0.5–2 ha in Vhembe) [31]. Daily temporal resolution supports responsive scheduling aligned with maize's peak evapotranspiration demand of 5–8 mm/day under semi-arid conditions [44,71]. Depth-stratified predictions allow differentiation between short-term irrigation needs (governed by 20–40 cm status) and longer-term drought resilience (reflected in 60–100 cm reserves). However, translating these technical capabilities into adopted practice requires addressing key implementation barriers: farmer access to forecast products through mobile platforms or extension services; alignment with existing water allocation schedules in communal schemes; and integration with local knowledge of soil variability and crop phenology not fully captured by district-scale models.

The irrigation deficit analysis (Figure 14) revealed mean seasonal deficits ranging from 5.8 mm (2020–2021) to 50.5 mm (2017–2018), with maximum localized deficits exceeding 150 mm. For context, maize grain yield typically declines by 5–7% per 10 mm of unmet water demand during critical growth stages, suggesting that unaddressed 2017–2018 deficits could have reduced yields by 25–35% in severely affected areas [72]. Conversely, precision irrigation targeting deficit hotspots could stabilize yields closer to 4–6 t ha⁻¹ achieved in well-managed schemes versus 0.8–1.6 t ha⁻¹ under rainfed conditions [29,73]. The spatial persistence of deficit hotspots in northern and southeastern Vhembe across multiple seasons indicates structural water availability constraints—likely related to soil texture, topographic position, and distance from water sources—that require infrastructure investment (boreholes, conveyance systems) alongside scheduling optimization.

The resulting irrigation decision maps demonstrated that most maize-growing areas frequently required irrigation, except in 2020–2021 when optimal conditions exceeded 35% of the district. The scalability and computational efficiency of RF and XGBoost (district-wide predictions generated in 2–4 h on standard workstations) make them viable for regional deployment across semi-arid Africa [42,74]. However, operational deployment faces several socio-technical challenges: the two-site calibration limits district-wide accuracy claims; smallholder farmers typically lack direct EO infrastructure access; and effective irrigation management requires not just information but also reliable water access, appropriate governance, and economic incentives [75,76]. Our study addresses the information component—demonstrating that high-resolution root-zone estimates are technically achievable—but cannot resolve the infrastructure, governance, and market constraints that often limit smallholder irrigation performance [74,77]. Future work should therefore pilot the framework within existing irrigation schemes where institutional capacity already exists, using participatory trials to co-develop decision rules that balance model recommendations with local knowledge and operational constraints.

A key limitation is the application of uniform irrigation thresholds ($FC \approx 0.35 \text{ cm}^3/\text{cm}^3$, $PWP \approx 0.12 \text{ cm}^3/\text{cm}^3$, $MAD \approx 0.23 \text{ cm}^3/\text{cm}^3$) across contrasting soil textures. True

field capacity and permanent wilting point are texture-dependent properties that vary by $0.05\text{--}0.15\text{ cm}^3/\text{cm}^3$ between sandy and clayey soils [41]. Our thresholds were derived empirically from the 5th and 95th percentiles of multi-year ARC observations and represent site-averaged management decision points. The small difference in percentile-based thresholds between sites ($<0.03\text{ cm}^3/\text{cm}^3$) reflects that both capture similar seasonal moisture extremes despite textural differences. Future refinements should incorporate pedotransfer functions to define spatially explicit FC and PWP values, particularly in regions with greater textural heterogeneity [78]. Farmers should treat our decision maps as general guidance and refine thresholds using soil-specific calibration or local soil moisture sensors.

5. Conclusions

This study developed and validated—at two calibration sites in the Vhembe District, South Africa—an EO–ML framework that integrates Sentinel-1 SAR backscatter, TU Wien soil moisture retrievals, and meteorological data to produce daily, 10 m resolution root-zone soil moisture estimates (0–100 cm) for smallholder maize systems.

Ensemble ML algorithms, particularly Random Forest and XGBoost, achieved high predictive accuracy ($R^2 > 0.96$; $\text{RMSE} < 0.025\text{ cm}^3/\text{cm}^3$) at the calibration sites, reliably reproducing empirically derived field-capacity ($\text{FC} \approx 0.35\text{ cm}^3/\text{cm}^3$) and management-allowable-depletion ($\text{MAD} \approx 0.23\text{ cm}^3/\text{cm}^3$) thresholds critical for irrigation scheduling. Depth-stratified validation confirmed strong SAR surface correlations ($r = 0.84\text{--}0.85$ at 10 cm), declining systematically below 20 cm ($r < 0.2$ at 40 cm), demonstrating that root-zone predictions rely primarily on meteorological gap-filling rather than direct satellite observation beyond the near-surface layer. District-wide spatial predictions showed that 79–94% of maize-growing areas required irrigation during dry years (2017–2019, 2021–2022) versus 32% in the wetter 2020–2021 season, with persistent deficit hotspots in northern and southeastern Vhembe indicating structural water availability constraints.

Key limitations include: (i) two-site calibration restricting spatial generalization; (ii) uniform hydraulic thresholds applied across heterogeneous soils (sandy loam to clay loam), requiring local calibration for precision use; and (iii) C-band SAR sensing depth confined to the top 2–5 cm, meaning subsurface predictions depend on statistical extrapolation from meteorological proxies. These limitations are fully discussed in Section 4.

Future work should expand calibration networks to 8–12 stratified sites, implement vegetation correction algorithms (e.g., Water Cloud Model) before ML integration, and couple EO–ML outputs with farmer-accessible decision-support platforms to translate this technical capability into improved water-use efficiency and smallholder resilience across sub-Saharan Africa.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/w18040499/s1>, Figure S1: Feature importance analysis for Random Forest model showing predictor contributions across soil depths (10–100 cm). Figure S2: Feature importance analysis for XGBoost model showing predictor contributions across soil depths (10–100 cm). Figure S3: Feature importance analysis for SVM, KNN, and MARS models showing predictor contributions across soil depths (10–100 cm). Figure S4: Random Forest ensemble variance (prediction uncertainty) maps for the Vhembe District showing spatial distribution of model confidence across five growing seasons (2017–2022). Figure S6: Spatial soil moisture predictions using MARS model across five growing seasons (2017–2022). Figure S7: Spatial soil moisture predictions using SVM model across five growing seasons (2017–2022). Figure S8: Spatial soil moisture predictions using XGBoost model across five growing seasons (2017–2022). Figure S9: Comparison of spatial prediction performance across all five machine learning models, highlighting differences in spatial discrimination capabilities. Figure S10: Correlation matrix showing relationships between lagged soil moisture variables, meteorological predictors, and topographic features used in machine

learning models. Figure S11: Independent test set validation scatterplots for all machine learning models across all soil depths (10–100 cm) demonstrating out-of-sample generalization performance. Figure S 12: Complete time series comparison of satellite vs. in-situ soil moisture across all depths (2017–2022). Table S1: Detailed software versions, package dependencies, and computational environment specifications used for data processing, machine learning model development, and spatial analysis. Table S2: Complete model performance metrics (R^2 , RMSE, MAE, MBE, MSE, NRMSE) for training, validation, and test datasets across all soil depths and machine learning algorithms.

Author Contributions: Conceptualization: G.S.N., T.S.R., A.N. and T.J.; Methodology: G.S.N., A.N., T.J. and N.É.K.; Software: G.S.N. and N.É.K.; Validation: G.S.N., T.S.R. and N.É.K.; Formal Analysis: G.S.N., T.S.R. and T.J.; Investigation: G.S.N., T.S.R., and Z.D.; Resources: G.S.N., A.N. and T.J.; Data Curation: G.S.N., T.S.R. and N.É.K.; Writing—Original Draft Preparation: G.S.N., T.S.R. and Z.D.; Writing—Review & Editing: All authors; Visualization: G.S.N., T.S.R. and N.É.K.; Supervision: G.S.N., A.N. and T.J.; Project Administration: G.S.N. and A.N.; Funding Acquisition: A.N. and T.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the TKP2021-NKTA-32 project. Project no. TKP2021-NKTA-32 has been implemented with the support provided by the National Research, Development and Innovation Fund of Hungary, financed under the TKP2021-NKTA funding scheme.

Data Availability Statement: The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request. Sentinel-1 SAR imagery is publicly accessible through the European Space Agency Copernicus Open Access Hub: <https://scihub.copernicus.eu/>). SRTM Digital Elevation Model data can be obtained from the USGS Earth Explorer portal (<https://earthexplorer.usgs.gov/>) (accessed on 2 July 2025). In situ soil moisture measurements from Agricultural Research Council (ARC) monitoring stations are subject to data sharing agreements and can be requested from the ARC: <https://www.arc.agric.za/> (accessed on 25 May 2025). Meteorological data from the South African Weather Service (SAWS) are available through SAWS data services (<https://www.weathersa.co.za/home/recentclimate>) (accessed on 19 April 2025) subject to their terms of use. Processing scripts and machine learning model code developed for this study will be made available in a public repository (GitHub/Zenodo) upon manuscript acceptance or can be requested from the corresponding author.

Acknowledgments: The authors gratefully acknowledge the Széchenyi Plan Plus program for financial support through grant RRF-2.3.1-21-2022-00008. We thank the field technicians and staff at the Noordgrens and Sigonde monitoring stations for their assistance with soil moisture data collection and maintenance of instrumentation, the Agricultural Research Council (ARC) of South Africa for providing soil water content measurements, and the South African Weather Service (SAWS) for providing meteorological data used in this study.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Atiah, W.A.; Amekudzi, L.K.; Akum, R.A.; Quansah, E.; Antwi-Agyei, P.; Danuor, S.K. Climate Variability and Impacts on Maize (*Zea Mays*) Yield in Ghana, West Africa. *Quart. J. Royal Meteor. Soc.* **2022**, *148*, 185–198. [[CrossRef](#)]
2. Suriadi, A.; Syarifinnur; Mulyati; Sumarsono, J.; Hadiawati, L.; Khaerana; Putra, G. Maize Production at Phenological Stages Affected by Water Irrigation Stress in Dryland Conditions. *IOP Conf. Ser. Earth Environ. Sci.* **2024**, *1377*, 012016. [[CrossRef](#)]
3. Datta, S.; Taghvaeian, S.; Ochsner, T.; Moriasi, D.; Gowda, P.; Steiner, J. Performance Assessment of Five Different Soil Moisture Sensors under Irrigated Field Conditions in Oklahoma. *Sensors* **2018**, *18*, 3786. [[CrossRef](#)]
4. Ford, T.W.; Quiring, S.M. Comparison of Contemporary In Situ, Model, and Satellite Remote Sensing Soil Moisture with a Focus on Drought Monitoring. *Water Resour. Res.* **2019**, *55*, 1565–1582. [[CrossRef](#)]
5. Sah, R.P.; Chakraborty, M.; Prasad, K.; Pandit, M.; Tudu, V.K.; Chakravarty, M.K.; Narayan, S.C.; Rana, M.; Moharana, D. Impact of Water Deficit Stress in Maize: Phenology and Yield Components. *Sci. Rep.* **2020**, *10*, 2944. [[CrossRef](#)]

6. Zhao, F.; Wang, G.; Li, S.; Hagan, D.F.T.; Ullah, W. The Combined Effects of VPD and Soil Moisture on Historical Maize Yield and Prediction in China. *Front. Environ. Sci.* **2023**, *11*, 1117184. [[CrossRef](#)]
7. Carvalho, A.A.D.; Montenegro, A.A.D.A.; Assis, F.M.V.D.; Tabosa, J.N.; Cavalcanti, R.Q.; Almeida, T.A.B. Spatial Dependence of Attributes of Rainfed Maize under Distinct Soil Cover Conditions. *Rev. Bras. Eng. Agric. Ambient.* **2019**, *23*, 33–39. [[CrossRef](#)]
8. Vennam, R.R.; Poudel, S.; Ramamoorthy, P.; Samiappan, S.; Reddy, K.R.; Bheemanahalli, R. Impact of Soil Moisture Stress during the Silk Emergence and Grain-filling in Maize. *Physiol. Plant.* **2023**, *175*, e14029. [[CrossRef](#)]
9. Zhou, Z.; Diverres, G.; Kang, C.; Thapa, S.; Karkee, M.; Zhang, Q.; Keller, M. Ground-Based Thermal Imaging for Assessing Crop Water Status in Grapevines over a Growing Season. *Agronomy* **2022**, *12*, 322. [[CrossRef](#)]
10. He, W.; Yokoya, N. Multi-Temporal Sentinel-1 and -2 Data Fusion for Optical Image Simulation. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 389. [[CrossRef](#)]
11. Lozac'h, L.; Bazzi, H.; Baghdadi, N.; Hajj, M.E.; Zribi, M.; Cresson, R. Sentinel-1/Sentinel-2-Derived Soil Moisture Product At Plot Scale (S² MP). In *Proceedings of the 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*; IEEE: Tunis, Tunisia, 2020; pp. 168–171.
12. Padrón, R.A.R.; Gula, M.O.B.; Ben, L.H.B.; Mezzomo, W. Calibration of the Capacitance Probe for Soil Moisture Monitoring. *Rev. Bras. De Agric. Irrig.-RBAI* **2022**, *16*, 131–137. [[CrossRef](#)]
13. Wagner, W.; Lemoine, G.; Rott, H. A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data. *Remote Sens. Environ.* **1999**, *70*, 191–207. [[CrossRef](#)]
14. Dorigo, W.; Wagner, W.; Albergel, C.; Albrecht, F.; Balsamo, G.; Brocca, L.; Chung, D.; Ertl, M.; Forkel, M.; Gruber, A.; et al. ESA CCI Soil Moisture for Improved Earth System Understanding: State-of-the Art and Future Directions. *Remote Sens. Environ.* **2017**, *203*, 185–215. [[CrossRef](#)]
15. Bauer-Marschallinger, B.; Freeman, V.; Cao, S.; Paulik, C.; Schaufler, S.; Stachl, T.; Modanesi, S.; Massari, C.; Ciabatta, L.; Brocca, L.; et al. Toward Global Soil Moisture Monitoring with Sentinel-1: Harnessing Assets and Overcoming Obstacles. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 520–539. [[CrossRef](#)]
16. Ågren, A.M.; Lidberg, W.; Strömberg, M.; Ogilvie, J.; Arp, P.A. Evaluating Digital Terrain Indices for Soil Wetness Mapping—A Swedish Case Study. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 3623–3634. [[CrossRef](#)]
17. Li, M.; Yan, Y. Comparative Analysis of Machine-Learning Models for Soil Moisture Estimation Using High-Resolution Remote-Sensing Data. *Land* **2024**, *13*, 1331. [[CrossRef](#)]
18. Sungmin, O.; Orth, R. Global Soil Moisture Data Derived through Machine Learning Trained with In-Situ Measurements. *Sci. Data* **2021**, *8*, 170. [[CrossRef](#)]
19. Guo, Y.; Chen, S.; Li, X.; Cunha, M.; Jayavelu, S.; Cammarano, D.; Fu, Y. Machine Learning-Based Approaches for Predicting SPAD Values of Maize Using Multi-Spectral Images. *Remote Sens.* **2022**, *14*, 1337. [[CrossRef](#)]
20. Hegazi, E.H.; Samak, A.A.; Yang, L.; Huang, R.; Huang, J. Prediction of Soil Moisture Content from Sentinel-2 Images Using Convolutional Neural Network (CNN). *Agronomy* **2023**, *13*, 656. [[CrossRef](#)]
21. Song, W.; Song, W.; Gu, H.; Li, F. Progress in the Remote Sensing Monitoring of the Ecological Environment in Mining Areas. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1846. [[CrossRef](#)]
22. Alaboz, P. Model Ensemble Techniques of Machine Learning Algorithms for Soil Moisture Constants in the Semi-arid Climate Conditions. *Irrig. Drain.* **2025**, *74*, 529–540. [[CrossRef](#)]
23. Massari, C.; Modanesi, S.; Dari, J.; Gruber, A.; De Lannoy, G.J.M.; Giroto, M.; Quintana-Seguí, P.; Le Page, M.; Jarlan, L.; Zribi, M.; et al. A Review of Irrigation Information Retrievals from Space and Their Utility for Users. *Remote Sens.* **2021**, *13*, 4112. [[CrossRef](#)]
24. Kahinda, J.-M.M.; Kapangaziwiri, E.; Hughes, D.; Khakhu, K. *Towards the Quantification of the Historical and Future Water Resources of the Limpopo River*; Water Research Commission: Pretoria, South Africa, 2022.
25. Kage, H.; Kochler, M.; Stützel, H. Root Growth and Dry Matter Partitioning of Cauliflower under Drought Stress Conditions: Measurement and Simulation. *Eur. J. Agron.* **2004**, *20*, 379–394. [[CrossRef](#)]
26. Lilley, J.M.; Fukai, S. Effect of Timing and Severity of Water Deficit on Four Diverse Rice Cultivars II. Physiological Responses to Soil Water Deficit. *Field Crops Res.* **1994**, *37*, 215–223. [[CrossRef](#)]
27. Materechera, F.; Scholes, M.C. Understanding the Drivers of Production in South African Farming Systems: A Case Study of the Vhembe District, Limpopo South Africa. *Front. Sustain. Food Syst.* **2022**, *6*, 722344. [[CrossRef](#)]
28. Shoko Kori, D.; Musakwa, W.; Kelso, C. Understanding the Local Implications of Climate Change: Unpacking the Experiences of Smallholder Farmers in Thulamela Municipality, Vhembe District, Limpopo Province, South Africa. *PLoS Clim.* **2024**, *3*, e0000500. [[CrossRef](#)]
29. Haarhoff, S.J.; Kotzé, T.N.; Swanepoel, P.A. A Prospectus for Sustainability of Rainfed Maize Production Systems in South Africa. *Crop Sci.* **2020**, *60*, 14–28. [[CrossRef](#)]
30. Lam, Q.D.; Rötter, R.P.; Rapholo, E.; Ayisi, K.; Nelson, W.C.D.; Odhiambo, J.; Foord, S. Modelling Maize Yield Impacts of Improved Water and Fertilizer Management in Southern Africa Using Cropping System Model Coupled to an Agro-Hydrological Model at Field and Catchment Scale. *J. Agric. Sci.* **2023**, *161*, 356–372. [[CrossRef](#)]

31. Denison, J.; Manona, S. *Principles, Approaches and Guidelines for the Participatory Revitalisation of Smallholder Irrigation Schemes*; Water Research Commission: Pretoria, South Africa, 2007; ISBN 978-1-77005-568-1.
32. Simanjuntak, C.; Gaiser, T.; Ahrends, H.E.; Ceglar, A.; Singh, M.; Ewert, F.; Srivastava, A.K. Impact of Climate Extreme Events and Their Causality on Maize Yield in South Africa. *Sci. Rep.* **2023**, *13*, 12462. [[CrossRef](#)]
33. Nxumalo, G.; Bashir, B.; Alsafadi, K.; Bachir, H.; Harsányi, E.; Arshad, S.; Mohammed, S. Meteorological Drought Variability and Its Impact on Wheat Yields across South Africa. *Int. J. Environ. Res. Public Health* **2022**, *19*, 16469. [[CrossRef](#)]
34. Ferreira, N.C.R.; Rötter, R.P.; Bracho-Mujica, G.; Nelson, W.C.D.; Lam, Q.D.; Recktenwald, C.; Abdulai, I.; Odhiambo, J.; Foord, S. Drought Patterns: Their Spatiotemporal Variability and Impacts on Maize Production in Limpopo Province, South Africa. *Int. J. Biometeorol.* **2023**, *67*, 133–148. [[CrossRef](#)]
35. Filippini, F. Sentinel-1 GRD Preprocessing Workflow. *Proceedings* **2019**, *18*, 11. [[CrossRef](#)]
36. Tamás, J.; Lénárt, C. Analysis of a Small Agricultural Watershed Using Remote Sensing Techniques. *Int. J. Remote Sens.* **2006**, *27*, 3727–3738. [[CrossRef](#)]
37. Ulaby, F.T.; Moore, R.K.; Fung, A.K. *Microwave Remote Sensing: Active and Passive*; Remote Sensing; Artech House: Norwood, MA, USA, 1981; ISBN 978-0-89006-190-9.
38. Baghdadi, N.; Bazzi, H.; El Hajj, M.; Zribi, M. Detection of Frozen Soil Using Sentinel-1 SAR Data. *Remote Sens.* **2018**, *10*, 1182. [[CrossRef](#)]
39. Balenzano, A.; Mattia, F.; Satalino, G.; Davidson, M.W.J. Dense Temporal Series of C- and L-Band SAR Data for Soil Moisture Retrieval Over Agricultural Crops. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 439–450. [[CrossRef](#)]
40. Bai, X.; He, B.; Li, X.; Zeng, J.; Wang, X.; Wang, Z.; Zeng, Y.; Su, Z. First Assessment of Sentinel-1A Data for Surface Soil Moisture Estimations Using a Coupled Water Cloud Model and Advanced Integral Equation Model over the Tibetan Plateau. *Remote Sens.* **2017**, *9*, 714. [[CrossRef](#)]
41. Paloscia, S.; Pettinato, S.; Santi, E.; Notarnicola, C.; Pasolli, L.; Reppucci, A. Soil Moisture Mapping Using Sentinel-1 Images: Algorithm and Preliminary Validation. *Remote Sens. Environ.* **2013**, *134*, 234–248. [[CrossRef](#)]
42. Saxton, K.E.; Rawls, W.J. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Sci. Soc. Amer. J.* **2006**, *70*, 1569–1578. [[CrossRef](#)]
43. Gala, T.S.; Aldred, D.A.; Carlyle, S.; Creed, I.F. Topographically Based Spatially Averaging of SAR Data Improves Performance of Soil Moisture Models. *Remote Sens. Environ.* **2011**, *115*, 3507–3516. [[CrossRef](#)]
44. Stirzaker, R.; Mbakwe, I.; Mziray, N.R. A Soil Water and Solute Learning System for Small-Scale Irrigators in Africa. *Int. J. Water Resour. Dev.* **2017**, *33*, 788–803. [[CrossRef](#)]
45. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
46. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: San Francisco, CA, USA, 2016; pp. 785–794.
47. Li, S.; Han, Y.; Li, C.; Wang, J. A Novel Framework for Multi-Layer Soil Moisture Estimation with High Spatio-Temporal Resolution Based on Data Fusion and Automated Machine Learning. *Agric. Water Manag.* **2024**, *306*, 109173. [[CrossRef](#)]
48. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
49. Ettalbi, M.; Baghdadi, N.; Garambois, P.-A.; Bazzi, H.; Ferreira, E.; Zribi, M. Soil Moisture Retrieval in Bare Agricultural Areas Using Sentinel-1 Images. *Remote Sens.* **2023**, *15*, 3502. [[CrossRef](#)]
50. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inform. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
51. Li, W.; Xiao, C.; Liang, X.; Yang, W.; Zhang, J.; Dai, R.; La, Y.; Kang, L.; Zhao, D. Precision Identification of Irrigated Areas in Semi-Arid Regions Using Optical-Radar Time-Series Features and Ensemble Machine Learning. *Hydrology* **2025**, *12*, 214. [[CrossRef](#)]
52. Dinesh, D.; Kumar, S.; Saran, S. Machine Learning Modelling for Soil Moisture Retrieval from Simulated NASA-ISRO SAR (NISAR) L-Band Data. *Remote Sens.* **2024**, *16*, 3539. [[CrossRef](#)]
53. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*; ACM: Anchorage, AK, USA, 2019; pp. 2623–2631.
54. Lamichhane, M.; Mehan, S.; Mankin, K.R. Soil Moisture Prediction Using Remote Sensing and Machine Learning Algorithms: A Review on Progress, Challenges, and Opportunities. *Remote Sens.* **2025**, *17*, 2397. [[CrossRef](#)]
55. Friedman, J.H. Multivariate Adaptive Regression Splines. *Ann. Statist.* **1991**, *19*, 1–67. [[CrossRef](#)]
56. Willmott, C.; Matsuura, K. Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Clim. Res.* **2005**, *30*, 79–82. [[CrossRef](#)]
57. Filgueiras, R.; Mantovani, E.C.; Fernandes-Filho, E.I.; Cunha, F.F.D.; Althoff, D.; Dias, S.H.B. Fusion of MODIS and Landsat-Like Images for Daily High Spatial Resolution NDVI. *Remote Sens.* **2020**, *12*, 1297. [[CrossRef](#)]
58. Hodson, T.O. Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [[CrossRef](#)]

59. Robeson, S.M.; Willmott, C.J. Decomposition of the Mean Absolute Error (MAE) into Systematic and Unsystematic Components. *PLoS ONE* **2023**, *18*, e0279774. [[CrossRef](#)] [[PubMed](#)]
60. Palazzolo, N.; Peres, D.J.; Creaco, E.; Cancelliere, A. Using Principal Component Analysis to Incorporate Multi-Layer Soil Moisture Information in Hydrometeorological Thresholds for Landslide Prediction: An Investigation Based on ERA5-Land Reanalysis Data. *Nat. Hazards Earth Syst. Sci.* **2023**, *23*, 279–291. [[CrossRef](#)]
61. Zribi, M.; Gorrab, A.; Baghdadi, N. A New Soil Roughness Parameter for the Modelling of Radar Backscattering over Bare Soil. *Remote Sens. Environ.* **2014**, *152*, 62–73. [[CrossRef](#)]
62. Lewis, B.L.; Kretschmer, F.F.; Shelton, W.W. *Aspects of Radar Signal Processing*; Artech House: Norwood, MA, USA, 1986; ISBN 978-0-89006-191-6.
63. Escorihuela, M.J.; Chanzy, A.; Wigneron, J.P.; Kerr, Y.H. Effective Soil Moisture Sampling Depth of L-Band Radiometry: A Case Study. *Remote Sens. Environ.* **2010**, *114*, 995–1001. [[CrossRef](#)]
64. Western, A.W.; Grayson, R.B.; Blöschl, G. Scaling of Soil Moisture: A Hydrologic Perspective. *Annu. Rev. Earth Planet. Sci.* **2002**, *30*, 149–180. [[CrossRef](#)]
65. Ford, T.W.; Harris, E.; Quiring, S.M. Estimating Root Zone Soil Moisture Using Near-Surface Observations from SMOS. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 139–154. [[CrossRef](#)]
66. Gruber, A.; De Lannoy, G.; Albergel, C.; Al-Yaari, A.; Brocca, L.; Calvet, J.-C.; Colliander, A.; Cosh, M.; Crow, W.; Dorigo, W.; et al. Validation Practices for Satellite Soil Moisture Retrievals: What Are (the) Errors? *Remote Sens. Environ.* **2020**, *244*, 111806. [[CrossRef](#)]
67. Tong, C.; Wang, H.; Magagi, R.; Goita, K.; Zhu, L.; Yang, M.; Deng, J. Soil Moisture Retrievals by Combining Passive Microwave and Optical Data. *Remote Sens.* **2020**, *12*, 3173. [[CrossRef](#)]
68. Adab, H.; Morbidelli, R.; Saltalippi, C.; Moradian, M.; Ghalhari, G.A.F. Machine Learning to Estimate Surface Soil Moisture from Remote Sensing Data. *Water* **2020**, *12*, 3223. [[CrossRef](#)]
69. Carranza, C.; Nolet, C.; Peziz, M.; Van Der Ploeg, M. Root Zone Soil Moisture Estimation with Random Forest. *J. Hydrol.* **2021**, *593*, 125840. [[CrossRef](#)]
70. Onyango, C.M.; Nyaga, J.M.; Wetterlind, J.; Söderström, M.; Piikki, K. Precision Agriculture for Resource Use Efficiency in Smallholder Farming Systems in Sub-Saharan Africa: A Systematic Review. *Sustainability* **2021**, *13*, 1158. [[CrossRef](#)]
71. Nxumalo, G.S.; Ramabulana, T.S.; Dlamini, Z.; Louis, A.; Nagy, A. Integrating OPTRAM and Machine Learning with Multimodal EO Proxies for Optimized Irrigation Scheduling in Smallholder Systems: A Vhembe District Case Study. *Front. Agron.* **2026**, *7*, 1697188. [[CrossRef](#)]
72. Bwambale, E.; Abagale, F.K.; Anornu, G.K. Model-Based Smart Irrigation Control Strategy and Its Effect on Water Use Efficiency in Tomato Production. *Cogent Eng.* **2023**, *10*, 2259217. [[CrossRef](#)]
73. Ndhleve, S.; Nakin, M.D.V.; Longo-Mbenza, B. Impacts of Supplemental Irrigation as a Climate Change Adaptation Strategy for Maize Production: A Case of the Eastern Cape Province of South Africa. *Water SA* **2017**, *43*, 222. [[CrossRef](#)]
74. Tesfay, M.G. Impact of Irrigated Agriculture on Welfare of Farm Households in Northern Ethiopia: Panel Data Evidence. *Irrig. Drain.* **2021**, *70*, 306–320. [[CrossRef](#)]
75. Bjornlund, H.; Van Rooyen, A.; Stirzaker, R. Profitability and Productivity Barriers and Opportunities in Small-Scale Irrigation Schemes. *Int. J. Water Resour. Dev.* **2017**, *33*, 690–704. [[CrossRef](#)]
76. Nxumalo, G.S.; Chauke, H. Challenges and Opportunities in Smallholder Agriculture Digitization in South Africa. *Front. Sustain. Food Syst.* **2025**, *9*, 1583224. [[CrossRef](#)]
77. Franke, A.C.; Machakaire, A.T.B.; Mukiibi, A.; Kayes, M.J.; Swanepoel, P.A.; Steyn, J.M. In-Field Assessment of the Variability in Water and Nutrient Use Efficiency among Potato Farmers in a Semi-Arid Climate. *Front. Sustain. Food Syst.* **2023**, *7*, 1222870. [[CrossRef](#)]
78. Vereecken, H.; Huisman, J.A.; Pachepsky, Y.; Montzka, C.; Van Der Kruk, J.; Bogen, H.; Weihermüller, L.; Herbst, M.; Martinez, G.; Vanderborght, J. On the Spatio-Temporal Dynamics of Soil Moisture at the Field Scale. *J. Hydrol.* **2014**, *516*, 76–96. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.