THESIS of SHORT DOCTORAL (PhD) DISSERTATION

TAREK ALAHMAD

MOSONMAGYARÓVÁR
2026

THESIS OF SHORT DOCTORAL (PHD)

DISSERTATION SZÉCHENYI ISTVÁN UNIVERSITY

ALBERT KÁZMÉR FACULTY OF AGRICULTURAL AND FOOD SCIENCES IN MOSONMAGYARÓVÁR
DEPARTMENT OF BIOENGINEERING AND PRECISION TECHNOLOGY
WITTMANN ANTAL MULTIDISCIPLINARY DOCTORAL SCHOOL OF PLANT, ANIMAL AND FOOD SCIENCES

HABERLANDT GOTTLIEB DOCTORAL PROGRAM FOR PLANT SCIENCE

DOCTORAL SCHOOL LEADER: PROF. DR. LÁSZLÓ VARGA, DSC

PROGRAM LEADER:
PROF. DR. GYULA PINKE DSC

SUPERVISOR(S):
Dr. habil. Anikó Nyéki, Associate Professor

Prof. Dr. Miklós Neményi, Professor Emeritus, MHAS

Real-Time Soil Moisture Prediction and Analysis for Precision Crop Production

SUBMITTED BY
TAREK ALAHMAD

MOSONMAGYARÓVÁR

2026

1. INTRODUCTION

1.1. Background and Research Problem

Precision Agriculture (PA) has emerged as a critical paradigm for improving resource productivity and fostering environmental stewardship in modern agriculture through the management of spatial variability in the field. Soil moisture content (SMC) is a critical component of soil-crop-atmosphere interactions that influences the availability of water, energy balance, and crop growth. However, the conventional methods of SMC measurement have been found to be limited in their application for improving irrigation management. For instance, gravimetric measurement is labor-intensive and spatially limited, in-situ measurement requires dense networks of sensors that are costly to deploy, whereas remote sensing is limited to near-surface measurement. Further, many modeling efforts focus only on the shallow soil profile (0-20 cm), resulting in insufficient information for irrigation management that relies on the dynamics of root-zone soil moisture. For example, crops such as maize extract water from depths of more than 150 cm; however, most of the water is extracted from the upper 60 cm of the profile during the vegetative growth stage. Therefore, to achieve efficient irrigation management, there is a critical need for reliable measurement and prediction of SMC at different depths. Though machine learning (ML) algorithms have been found to achieve reliable prediction of SMC, their poor interpretability is a critical factor limiting their application for decision-making in modern agriculture. Development of decision-making frameworks using multiple sources of information such as field measurement, IoT-based meteorological measurement, and satellite-derived vegetation indices is a critical research gap that is yet to be fully explored.

1.2.Objectives

This dissertation addresses these gaps by developing depth specific, soil-aware ML models with explainable AI (XAI) capabilities. The main objectives are:

- To build an integrated dataset for depth-specific soil moisture modelling by combining gravimetric SMC, IoT meteorology, and Sentinel-derived vegetation indices (NDVI/NDMI).
- To characterize spatio-temporal and depth-dependent SMC dynamics across three different soil texture and two growing seasons of maize.
- To evaluate different predictor scenarios: (i) vegetation indices only (NDVI/NDMI), (ii) meteorological variables only, and (iii) combined predictors for depth-specific SMC prediction accuracy.
- To benchmark and optimize models by comparing machine learning approaches under a consistent train–validation–test protocol, assessing depth- and soil-specific strategies versus pooled modelling to improve generalizability.
- To apply explainable machine learning techniques (SHAP and permutation importance) to quantify global and depth-dependent feature contributions, and to translate these results into agronomically interpretable driver mechanisms for SMC variability.
- To synthesize predictive and explainability findings into actionable

implications for precision crop production and irrigation decision support align with Sustainable Development Goals (SDGs) 6 (Clean Water and Sanitation) and 13 (Climate Action).

2. MATERIALS AND METHODS

2.1. Study Area

Field experiments took place during the 2023 and 2024 maize growing seasons on a 23-hectare rainfed field owned by Széchenyi István University near Mosonmagyaróvár, Hungary (47°54'11.8"N, 17°15'08.9"E). The site is characterized by a temperate continental climate, and a mean annual precipitation of approximately 600 mm. Field terrain is nearly flat (elevation 133-138 m) with three soil textures according to USDA taxonomy: loam, sandy loam, and silt loam (Figure 1).

2.2.Data Collection

2.2.1. Gravimetric Soil Sampling

Soil samples were collected at fixed points near IoT sensor stations (within 1 m²) at five depths (5, 20, 40, 60, 80 cm) with three replicates per point. Sampling occurred biweekly in 2023 (9 dates) and on 23 dates across 2023-2024, totaling 705 samples. Samples were sealed, transported to the laboratory, and oven-dried at 105°C to constant mass.

Table 1: Summary of gravimetric samples collected across studies.

Study **Soil**

Loam,

Textures **Depths** **Replicates** **Total**
Sampling
Samples

Dates

Loam,

Article 1 **Article** 5 3 9 405

Sandy

loam, Silt loam

Silt loam 5 3 10 300 Loam,

2 Silt loam 5 3 23 705

Article 3

2.2.2. IoT-Based Meteorological Data

An on-site IoT sensor station (Boreas Ltd., EcoLogger Boreas datalogger unit BEL-06) continuously monitored: air temperature (°C),

relative humidity (%), precipitation (mm), wind speed (km/h), and solar radiation (W/m²). Data were transmitted via LoRaWAN at 10–15- minute intervals and synchronized with soil sampling timestamps. **2.2.3. Satellite-Derived Vegetation Indices**

Sentinel-2 Level-2A multispectral imagery closest to each sampling date (within ±3 days, cloud-free) was processed in QGIS. NDVI and NDMI were calculated:

NDVI = (NIR - Red) / (NIR + Red) using Sentinel-2 Band 8 (NIR, 842 nm) and Band 4 (Red, 665 nm)

NDMI = (NIR - SWIR1) / (NIR + SWIR1) using Sentinel-2 Band 8A (NIR narrow, 865 nm) and Band 11 (SWIR1, 1610 nm)

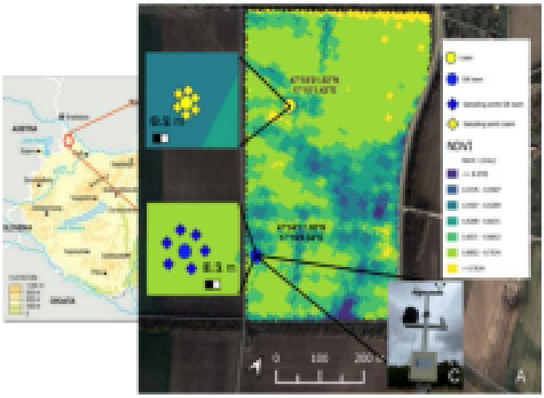

Figure 1: Study area location, soil texture classes, and sampling points with IoT sensor station and NDVI values.

2.3.Data Preprocessing and Feature Engineering

Data preprocessing was performed using Python (version 3.10.12). Data were synchronized, checked for outliers, and missing values imputed (forward-fill for short gaps). Soil texture was encoded as a categorical variable. The final dataset included predictors: meteorological variables (temperature, humidity, precipitation, wind

speed, solar radiation), depth, soil texture, and vegetation indices (NDVI, NDMI); target: SMC.

2.4.Statistical Characterization

To characterize SMC dynamics across depths and soil textures, descriptive statistics (mean, standard deviation, coefficient of variation) were calculated for each soil texture and depth using Python. Repeated measures ANOVA was conducted to test depth effects.

2.5.Modelling Frameworks

Three modelling frameworks were developed corresponding to the three articles:

2.5.1. Multi-Model Comparison (Article 1)

RFR, XGBoost, and LSTM were trained on meteorological predictors, depth and soil texture to predict SMC across three textures and five depths. Data split: 80% training, 20% testing. Hyperparameters tuned via grid search (RFR, XGBoost) and early stopping (LSTM).

2.5.2. Soil-Specific GBR Models (Article 2)

Separate Gradient Boosting Regressor models were developed for loam and silt loam using meteorological predictors and depth. Hyperparameters optimized with GridSearchCV (5-fold). Data split: 70% training, 30% testing.

2.5.3. Explainable XGBoost with Multi-Source Fusion (Article 3)

XGBoost models were built for each soil texture (loam, silt loam) and depth (5-80 cm) under two scenarios: (A) vegetation indices only (NDVI, NDMI); (B) combined (vegetation + meteorological). SHAP analysis (TreeExplainer) applied to interpret feature contributions.

2.6.Model Evaluation Metrics

All models were evaluated using: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), and Nash-Sutcliffe Efficiency (NSE).

2.7 Feature Importance and Explainability

• For RFR: Gini-based importance (mean decrease in impurity). • For XGBoost: Built-in gain-based feature importance. • For

LSTM: Model-agnostic permutation importance. • For GBR: Built-in feature importance from scikit-learn. • For Article 3: SHAP analysis using Python shap library (v0.41). **3.**

RESULTS

3.1.Statistical Characterization of SMC Variability Silt loam retained the highest moisture values with mean SMC ranged between 15.0 and 22.5%, particularly at intermediate to deep layers. Loam showed fewer mean values (12.6–16.2%), while sandy loam showed the lowest SMC with maximum values at 80 cm depth Figure 2. Analysis of 705 gravimetric samples across two growing seasons revealed clear texture- and depth-dependent variability. In loam, mean SMC ranged from 12.6% (60 cm) to 16.2% (20 cm), with CV increasing from 24% (5 cm) to 39% (80 cm). In silt loam, mean SMC was higher (15.0-22.5%) and CV increased from 17.5% (5 cm) to 47.8% (80 cm). Repeated measures ANOVA confirmed significant depth effects ($p < 0.05$) for both loam and silt loam, justifying depth specific modelling.

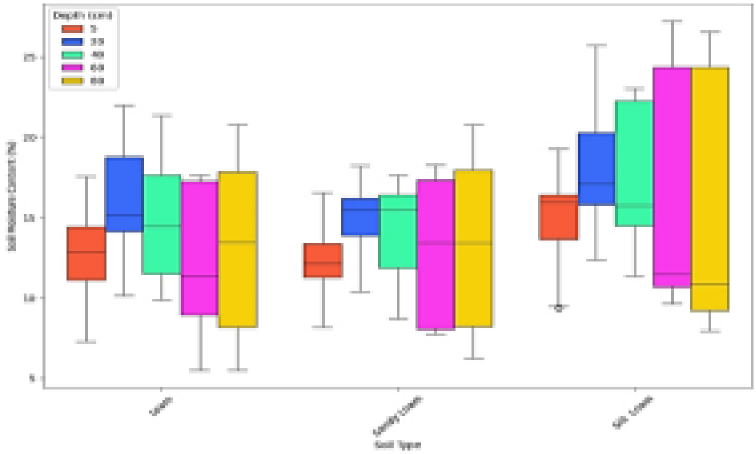

Figure 2. Soil moisture content variation in three soil textures at five different depths during 2023 season.

3.2.Predictive Performance Across Models and Scenarios 3.2.1. Multi-Model Comparison (Three Soil Textures) Across three textures and five depths, RFR provided the most consistent and highest performance with an R^2 ranging between 0.82 and 0.99, and RMSE < 1.5%. XGBoost demonstrated strong performance at shallow to moderate depths (5–40 cm), achieving a root mean square error (RMSE) as low as 0.10% in sandy loam at 40 cm. However, its accuracy decreased in deeper, fine-textured soils. The long short-term memory (LSTM) model effectively captured temporal dynamics at shallow to intermediate depths (5–60 cm), with a peak R^2 of 0.99 at 60 cm in loam. Nevertheless, its performance declined at greater depths.

3.2.2. GBR Model Performance (Two Soil Textures) Gradient Boosting Regression (GBR) model results indicated good predictive performance in terms of accuracy; the model was more accurate in simulating the behavior of the silt loam soil than the loam soil. This is evident by the high value of R^2 = 0.98, low value of

RMSE = 0.85%, and high value of NSE = 0.95 in the simulation results for the silt loam soil, compared to the loam soil. This is mainly due to low SMC.

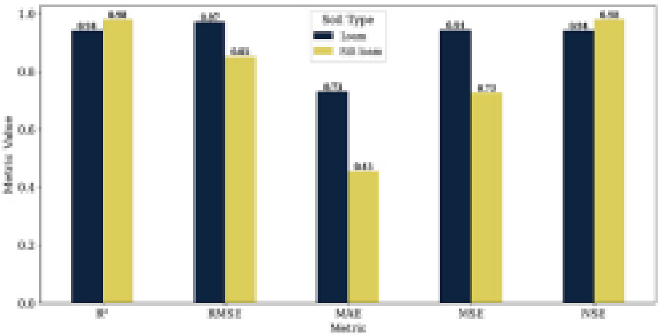

Figure 3: Evaluation metrics for soil moisture content prediction in loam and silt loam soils using the GBR model.

3.2.3. XGBoost Overall Model Performance with different inputs scenarios

Combining Sentinel-2 vegetation indices and IoT meteorological data consistently enhanced prediction accuracy in comparison with vegetation-only models across all depths and textures. The improvement was most pronounced in loam at shallow depths (5-20 cm) where RMSE reduction exceeded 60% dropped from 2.26% to 0.88% at 5 cm. In silt loam, vegetation-only models performed well with an R^2 up to 0.98 at 20 cm, but combined inputs still enhanced predictions to near-perfect accuracy ($R^2 = 0.99$). At deeper layers (40-80 cm), combined models maintained high accuracy with R^2 up to 0.92 in loam and around 0.98 in silt loam.

3.3.Feature Importance and Explainability Results

3.3.1. Multi-Model Feature Importance

Feature importance analysis showed that solar radiation was the most influential variable, particularly at deeper soil layers (40 to 80 cm),

with importance values ranging between 0.4 and 0.85 across all soil textures and depths. Whear at the surface layer (5 cm), precipitation emerged as a key predictor with importance value of 0.14 with RFR, 0.4 with XGBoost, and 0.25 with LSTM. Other variables such as humidity, temperature, wind speed varied in importance depending on model, depth, and soil texture.

3.3.2. GBR Feature Importance by Soil Texture

Results showed that in loam soils, precipitation and humidity were most influential factors (31% each), with soil depth also significant (27%). Solar radiation and wind speed were dominant in silt loam with values of 46% and 26% respectively, highlighting the role of radiation driven evapotranspiration in fine-textured soils.

3.3.3. SHAP Analysis for XGBoost Models

SHAP analysis emphasized depth-dependent variations in SMC drivers, where at upper layers (5-20 cm), NDVI was the most important predictor across both soil textures, confirming that denser, greener canopies indicate wetter soils. While, at intermediate depths (20-40 cm), both vegetation indices and solar radiation influenced moisture patterns. At deepest layers (60-80 cm), Solar radiation became the dominant factor, governing deep-soil moisture loss through root uptake and evaporation. Soil texture modulated these effects where NDVI retained importance at depth in silt loam due to stable moisture vegetation coupling, while meteorological factors remained influential throughout the profile in loam soils.

4. SCIENTIFIC THESES

1. Since the variation in soil moisture with depth is significantly affected by soil texture, the model should vary with depth and soil texture rather than using a one-size-fits-all approach.
2. Among all the algorithms, the best performing general model was found to be the Random Forest Regression model in predicting soil moisture levels at different depths and soil textures. This model was able to obtain high R^2 values ranging from 0.82 to 0.99 and low RMSE values less than 1.5% under all conditions. The best model to use in predicting soil moisture levels at different depths and soil textures was found to be the Random Forest model. However, the best model to use depends on the depth at which one wants to carry out the predictions. XGBoost was also seen to perform well in the shallower zones, and LSTM was seen to perform well in the upper and middle layers. However, these two algorithms were seen to lose their efficiency in deeper layers.
3. The primary environmental factors affecting soil moisture vary quantitatively in different soil textures. In loam soil, the two factors affecting soil moisture levels were found to be precipitation and humidity. Both these factors were seen to contribute 31% to the predictions. Similarly, in the case of silt loam soil, solar radiation was seen to dominate in affecting soil moisture levels by having the highest importance at 46%,

followed by wind speed at 26%.

4. The integration of Sentinel-2 vegetation indices with IoT meteorological data enhanced the accuracy of predictions significantly compared to vegetation index predictions alone. The improvement was more pronounced in loam soils at shallow

depth, with an improvement in model accuracy by more than 60%, reducing the RMSE by 2.26% to 0.88% at 5 cm depth. 5. Depth-dependent factors influencing soil moisture content were quantified using explainable AI. Results indicated texture dependent hierarchies in influencing factors. For loam soils, precipitation and humidity were the dominant factors at 31%, while in Silt Loam soils, solar radiation was the dominant factor at 46%, followed by wind speed at 26%. SHAP analysis indicated that in both soils, NDVI was the most important predictor near the surface, while solar radiation was the dominant predictor at depth (60-80 cm).

6. By integrating gravimetric SMC data with high-frequency meteorological data and satellite vegetation data, a strong foundation exists to develop depth-dependent ML models that can capture both high-frequency surface dynamics and deeper soil layer responses.
7. The integration of explainable machine learning with field data, meteorological data, and satellite data creates a strong scientific foundation to develop depth-dependent irrigation recommendations in precision crop production.

5. PRACTICAL THESES

1. Irrigation scheduling for loam soils should incorporate real-time meteorology in addition to canopy and deep layer prediction to differentiate between whole-profile depletion and surface drying. For silt loam soils, the close relationship between vegetation indices and soil moisture allows for scheduling of irrigation mainly

using satellite-derived NDVI trends, with some deep layer verification under extreme weather events.

2. Solar radiation is another important factor that influences the loss of deep soil moisture. During times of increased evaporative demand, the scheduling of irrigation should take into account the total amount of solar radiation exposure rather than relying solely on rainfall or surface layer information. This is especially true for loam soils because of the rapid propagation of atmospheric information through the profile.
3. Soil-type-specific approaches to monitoring will enable maximum efficiency in terms of cost and accuracy. For loam soils, there is a need to use integrated sensor-satellite systems with real-time meteorology to capture rapid changes in soil moisture. For silt loam soils, there is a possibility of using satellite-derived monitoring with some validation under extreme weather events.
4. The outputs of Explainable AI can be converted into simple decision rules. The SHAP tool can be used to develop heuristics, e.g., (i) when surface NDVI is high, and solar radiation is low, do not irrigate even if surface soil is dry; (ii) when there have been 3-

5 consecutive days of high solar radiation without appreciable rain, anticipate soil dehydration in the deep soil profile and irrigate regardless of surface soil status. These heuristics can be used by farmers.

5. Precision irrigation systems require not only technology but also expertise on the part of farmers. High-level expertise in agronomy is always needed to interpret the outputs of precision systems, understand when expert advice can be overridden, and develop

general rules based on specific situations. Precision technology can only be valuable when farmers have the expertise needed to use it critically.

6. A scalable architecture can be built on existing infrastructure. The proposed framework can be implemented using (i) existing free Sentinel-2 satellite images, (ii) existing low-cost IoT weather stations using LoRaWAN technology, (iii) open-source Python libraries (scikit-learn, XGBoost, SHAP) for machine learning, and (iv) cloud or edge computing technology for prediction delivery. This architecture can be used to provide expert decision support automatically, as well as being interpretable and understandable by farmers.
7. The modeling framework provides a proactive approach, not a reactive approach, in managing irrigation schedules. The system can predict moisture stress in the soil using real-time IoT sensors and weather forecasts, thus allowing farmers to schedule irrigation at times when it is most beneficial.

6. LIST OF PUBLICATION

1. Alahmad, T., Nyéki, A., Neményi, M. (2025*). Soil Moisture Content Prediction Using Gradient Boosting Regressor (GBR) Model: Soil-Specific Modeling with Five Depths*. Applied Sciences, 15(11), 5889. https://doi.org/10.3390/app15115889 (Q2, Impact Factor: 3.7)
2. Tarek Alahmad Miklós Neményi, Adrienn Széles, Nour Ali, Omar Hijazi, Anikó Nyéki (2025). *Spatiotemporal prediction of soil moisture content at various depths in three soil types using machine learning algorithms.* Frontiers in Soil Science

https://doi.org/10.3389/fsoil.2025.1612908 (Q2, Impact Factor: 3.7)
3. Alahmad, T., Nyéki, A., Neményi, M. (2025*). Explainable XGBoost-based models of root-zone soil moisture profiling using coupled Sentinel-2 and IoT data*. Discover Applied Sciences https://doi.org/10.1007/s42452-026-08673-3 (Q2, Impact Factor: 2.8)
4. Alahmad, T., Neményi, M., Nyéki, A. (2023). *Applying IoT Sensors and Big Data to Improve Precision Crop Production: A Review.* Agronomy, 13(10), 2603. https://doi.org/10.3390/agronomy13102603
5. Alahmad, T., Neményi, M., Nyéki, A. (2024). *Soil Moisture Content Prediction in Loam Soil with RFR Model*. Acta Agronomica Óváriensis, 65(2), 43-56. https://doi.org/10.17108/ActAgrOvar.2024.65.2.43
6. Alahmad, T., Neményi, M., Nyéki, A. (2025*). Soil moisture*

prediction with machine learning: combining vegetation indices and meteorological data. Precision agriculture '25, Brill. https://doi.org/10.1163/9789004725232_172

7. Alahmad, T., Neményi, M., Nyéki, A. (2025). *Enhancing Agricultural Sustainability using AI-Driven Soil Moisture Modeling: A Soil-Type and Depth Approach with SHAP Interpretability*. Acta Agronomica Óváriensis, https://doi.org/10.17108/ActAgrOvar.2025.66.2.5
8. Nyéki Anikó, Alahmad Tarek, Sana Arshad, Gombkötő Nóra, Neményi Miklós, Morad Mirzaei, Szabó Szilárd, Endre Harsanyi, Main Al-Dalahmeh, Safwan Mohammed. (2025). *Real-time*

monitoring of ammonia emissions from cereal crops using LoRaWAN-based sensing technology. Nature, Scientific Reports https://doi.org/10.1038/s41598-025-31661-3
9. B Ambrus, G Teschner, AJ Kovács, M Neményi, L Helyes, Z Pék, S Takács, T Alahmad, A Nyéki (2024). *Field-grown tomato yield estimation using point cloud segmentation with 3D shaping and RGB pictures from a field robot and digital single lens reflex cameras*. Heliyon 10.20 (2024). DOI: https://doi.org/10.1016/j.heliyon.2024.e37997
10. Neményi Miklós, Ambrus Bálint, Teschner Gergely, Alahmad Tarek, Nyéki Anikó, Kovács Attila J. (2023). *Challenges of ecocentric sustainable development in agriculture with special regard to the internet of things (IoT), an ICT perspective*. Agricultural Engineering Sciences Journal, online pp 1-10 (2023). https://doi.org/10.1556/446.2023.00099

11. Anikó Nyéki, Bálint Daróczy, Miklós Neményi, Bálint Ambrus, Gergely Teschner, and Tarek Alahmad. (2025). *Predicting maize growth and biomass: Integrating gradient boosted trees with sentinel images and IoT*. Agricultural Engineering Sciences Journal DOI: https://doi.org/10.1556/446.2025.00202
12. Nour Ali, Anikó Nyéki, Aïda Jalloul, Tarek Alahmad. (2024). *The Effect of Ascorbic Acid on Salt Tolerance and Seedling Performance in Triticum durum Defs. 'Douma 3' Under Salinity Stress in Syria*. Agronomy, 14(12), 2982; DOI: https://doi.org/10.3390/agronomy14122982