RFID-based Soil Moisture Sensor for Smart Agriculture: a Gaussian Mixture Model Approach

Nedal M. Benelmekki, Elvis Díaz Machado, Javier Del Rio Toledano,
Antoni Morell, Jose Lopez Vicario.
{nedal.martinez, elvis.diaz, javier.delrio, antoni.morell, jose.vicario}@uab.cat
Wireless Information Networking (WIN), Universitat Autònoma de Barcelona (UAB).
Edifici Q. Escola d'Enginyeria, Carrer de les Sitges, Cerdanyola del Vallès (08193 Barcelona).

Abstract—**In this work, we present an RFID-based indirect soil moisture sensor based on the application of Machine Learning. More specifically, we suggest an unsupervised approach that does not require information about the real height and moisture levels. This approach can be of great interest in practical agricultural deployments, where the careful deployment of tags at specific depths within the soil is challenging. It allows an estimation of the posterior probability of moisture, based on the available Received Signal Strength Indicator (RSSI) and phase. The suggested method enables the RFID system to operate as a sensor by probabilistically quantifying measurement uncertainty, which is a key distinction from existing methodologies. In this paper, we focus on two differentiated moisture cases to show the validity of our approach. Future research will extend the proposed methodology to a wider set of moisture levels.**

I. INTRODUCTION

The fourth industrial revolution, known as Industry 4.0, integrates emerging technologies to improve automation, intelligence, and efficiency in production processes and supply chains. However, its influence extends beyond manufacturing, pushing innovation across multiple fields, including agriculture. This evolution has led to Agriculture 4.0, where advanced technologies optimize efficiency while ensuring sustainability, a necessity to face current challenges such as droughts and climate change [1].

Among the innovations that fuel this new generation of agriculture, the Internet of Underground Things (IoUT) and Wireless Underground Sensor Networks (WUSNs) are key technologies. IoUT consists of sensors and communication devices, partially or completely buried underground for real-time soil sensing and monitoring [2], while WUSNs usually involve the use of sensor nodes completely buried in the soil, where each device contains all necessary sensors, memory, a processor, a radio, an antenna, and a power source [3]. These underground networks have shown promise for environmental monitoring, as they enable autonomous data collection from buried sensor nodes using various methods, such as unmanned vehicles, while minimizing interference with surface farming activities. Moreover, in addition to preserving arable land, underground deployment guards against common issues that affect aboveground nodes, such as theft or environmental degradation [4].

Despite their advantages, underground wireless networks face communication challenges, as electromagnetic wave propagation in soil faces higher attenuation [5]. Additionally, RF signals lose strength and exhibit phase variations when propagating through conductive materials [6, 7]. This phenomenon, while problematic for conventional communication, opens up an opportunity for sensing: Given that water is a well-known conductor and plays a key role in the electrical properties of the soil, variations in soil moisture directly influence the behavior of RF signals. Since soil moisture is critical in agriculture, affecting irrigation strategies and crop health, analyzing these variations of the RF signal could provide a promising method for indirect detection of soil moisture. Thus, this work presents a preliminary exploration of this concept using RFID technology and Machine Learning to develop an indirect soil moisture sensor. Although traditionally used for identification and tracking, RFID has recently found applications in agriculture, including soil moisture monitoring [8, 9]. Its low cost, energy efficiency, and non-invasive operation make it an ideal candidate for IoUT-based sensing. In particular, UHF RFID passive tags have been validated for agricultural sensing [10, 11], addressing several limitations of traditional sensors, such as battery dependence, calibration complexity, and high maintenance requirements.

To extract soil moisture information from RFID signal variations, this study adopts a probabilistic Machine Learning approach based on Gaussian Mixture Models (GMM). Unlike supervised learning methods that require labeled data for training, GMM operate in an unsupervised manner, which allows it to adapt to real-world, uncontrolled agricultural environments. Additionally, by quantifying uncertainty through Bayesian inference, the approach yields a clear estimation of soil moisture through the posterior probability conditioned on RSSI and phase variations, improving reliability in sensing.

II. PROBLEM FORMULATION

In this paper, we present an RFID-based indirect soil moisture sensor based on Machine Learning. By placing an RFID tag underground, our objective is to estimate the moisture level based on RSSI measurements. In a scenario with no obstacles between the RFID tag and the reader (only the soil) the RSSI can be expressed as [12]:

$$\begin{aligned} RSSI(dBm) =& g(P_{TX}(dBm), G_T(dB), G_R(dB)) \\ &- 10\beta \log(h_r, d_g) - \varphi(s_{moist}) \end{aligned} \tag{1}$$

where $g(P_{TX}(dBm), G_T(dB), G_R(dB))$ is the resulting combination of transmit power and transmit/receiver gains, β is the round-trip path loss exponent, h_r is the reader height (in meters), d_g is the distance from the ground surface to the buried RFID tag (expressed in meters) and $\varphi(s_{moist})$ is the loss (in dB) introduced by the soil in accordance to the moisture level (expressed by s_{moist}). By using the provided model, one can estimate s_{moist} by means of $RSSI(dBm)$. To do so, information about $g(P_{TX}(dBm), G_T(dB), G_R(dB))$, h_r and d_g is required to avoid any ambiguity regarding the nature of RSSI loss. This information is necessary to know whether a RSSI decrease comes from moisture or a variation of the reader's location. In a real scenario, however, accurate information is not always assured. Besides, equation (1) may be affected by modeling imperfections.

In order to consider a more realistic environment, in this work we consider the following assumptions:

- h_r and d_g are unknown. Instead, a distance range is considered.
- The soil moisture level is reduced to only two possible states, but the specific state is not known.
- The rest of parameters of the RFID system are unknown.

Since these assumptions limit the applicability of supervised learning (we cannot build a labeled training data-set), we propose an unsupervised learning approach based on GMM.

III. GMM FOR SOIL MOISTURE ESTIMATION

A Gaussian Mixture Model (GMM) is a probabilistic model that represents the overall data distribution as a weighted sum of several Gaussian component densities. Commonly used in Machine Learning as an unsupervised learning model, its parameters can be estimated using the Expectation-Maximization (EM) algorithm [13]. In the EM algorithm, the posterior probabilities are calculated in the Expectation Step and the parameters are updated in the Maximization Step. As flexible density estimators, GMMs can be used to construct more complex models in Machine Learning, as well as being able to cluster data. Below, we provide a summary of the key points on EM applied to GMM. The interested reader is referred to [13] for further details.

Consider a GMM with $K \in \mathbb{N}$ components (groups, or clusters) for N data points, as defined in [13]:

$$p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \pi) = \prod_{n=1}^{N} \prod_{k=1}^{K} [\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)]^{Z_{n,k}}$$

where $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ denotes the observed data, $\mathbf{Z} = (z_{n,k})_{n \in 0,1,\ldots,N-1 k \in 0,1,\ldots,K-1}$, is a binary ($z_{n,k} \in \{0, 1\}$, $\sum_{k=1}^{K} z_{n,k} = 1$) latent variable matrix (unobservable variable that influences observed data), $\pi_k \geq 0, \sum_{k=1}^{K} \pi_k = 1$ are the respective distribution weights (mixing coefficients), and $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the Gaussian probability density with mean $\boldsymbol{\mu}_k \in \mathbb{R}^d$ and covariance matrix $\boldsymbol{\Sigma}_k$. The objective of EM is to estimate the parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \pi)$ that maximize the log-likelihood ($\log p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \pi)$). A "step by step" recipe can be extrapolated in order to implement GMM with high-level programming languages, such as Python [14]:

1) **Initialization:** Randomly initialize $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and π.
2) **Expectation (E) Step:** Calculate the posterior probabilities (responsibilities), $\gamma_{n,k}$, by [13]:

$$\gamma_{n,k} = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

3) **Maximization (M) Step:** Update the parameters using the calculated posteriors by [13]:

$$N_k = \sum_{n=1}^{N} \gamma_{n,k} \quad \pi_k = \frac{N_k}{N} \quad \boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{n,k} \mathbf{x}_n$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{n,k} (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T$$

4) **Convergence:** Repeat steps 2 and 3 until the change in the log-likelihood falls below a predefined threshold.

In this work, two classes are considered to cover the two soil moisture cases. As for $\mathbf{x}_n$ and latent variable $\mathbf{z}_{n,k}$, the former refers to the experimental measurements obtained, whereas the latent information stands for the soil moisture level. Then, $K = 2$ and $\mathbf{x}_n$ becomes a 2D n-th measurement vector where each component is related to the RSSI and phase measurements, respectively. To generate the clusters, we depart from N measurements with different reader heights and soil moistures. As noted above, this information is not known in the algorithm.

For an initial set of measurements that can be carried out directly in situ (without any special deployment), the algorithm is iterated until convergence as previously described. Once the clusters are generated we have the following for $k = 1, 2$:

- $\boldsymbol{\mu}_k$: the centroid of each cluster being the two components the average RSSI and phase, respectively.
- $\boldsymbol{\Sigma}_k$: the covariance matrix of each cluster.
- π_k: the mixing coefficient for each cluster.

Then, we can compute the posterior probability of a new RFID measurement $\mathbf{x}_j$ belonging to each cluster conditioned on available data as follows [13]:

$$\gamma_{j,k} = \frac{\pi_k \mathcal{N}(\mathbf{x}_j|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{l=0}^{K-1} \pi_l \mathcal{N}(\mathbf{x}_j|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} \tag{2}$$

With that posterior probability the soil's moisture can be determined based on the class with the highest probability. The algorithm provides us with the soil moisture estimate (the obtained class) and the measurement uncertainty (the posterior probability).

IV. EXPERIMENTAL SETUP

The setup consists of a container filled with soil, where a commercial UHF RFID tag (Dogbone Monza 4D) is buried at a predetermined depth. Moisture sensors are placed on the soil surface and at three different depths to monitor the volumetric water content (VWC) gradient across the container. An Arduino-based microcontroller unit (MCU) manages the moisture sensors and displays real-time readings on an LCD screen. Fig. 1 presents a schematic representation of the setup, and Fig. 2 shows a picture of the assembled system.

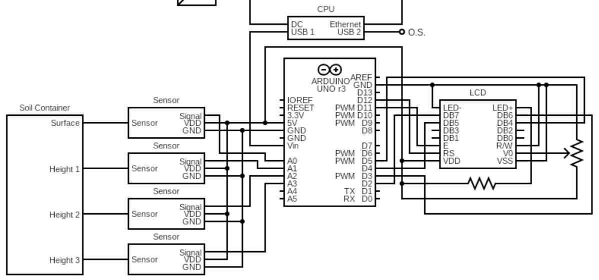


Fig. 1: Schematic of the proposed experimental setup.

A. Soil Preparation

1) **Soil deployment:** A 50L container was filled with soil of known composition.
2) **Moisture Levels:** To simulate a range of soil moisture conditions, water was added incrementally to the dry soil in a controlled manner, ensuring a uniform distribution throughout the container. Taking advantage that VWC is a purely volumetric measurement, we tested a wide range of moistures through $VWC = \frac{m_w - m_d}{m_d} \cdot \rho_b$ where m_d is the mass of the "dried" soil used as a baseline, m_w is the mass of the soil after adding the desired amount of water, and ρ_b is the water density, which can be assumed to be $1.0g/cm^3$ [15].
3) **Calibration:** The soil moisture content was increased in steps of approximately 6%, starting from dry conditions. Moisture was measured using a capacitive sensor.

B. RFID System

1) **Reader:** Zebra FX7500 RFID Reader operating in the UHF band (865.7MHz).
2) **Tags:** Passive UHF RFID tag (Dogbone Monza 4D).
3) **Antenna:** A Zebra AN480 Wide-Band RFID antenna.
4) **Tag Placement:** The RFID tag was placed at a fixed depth of 9.6cm inside a 12 x 5cm styrofoam case to ensure minimal soil compaction effects.

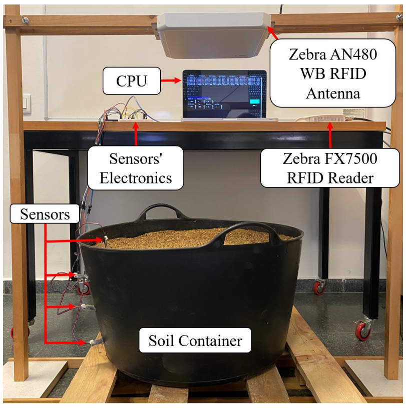


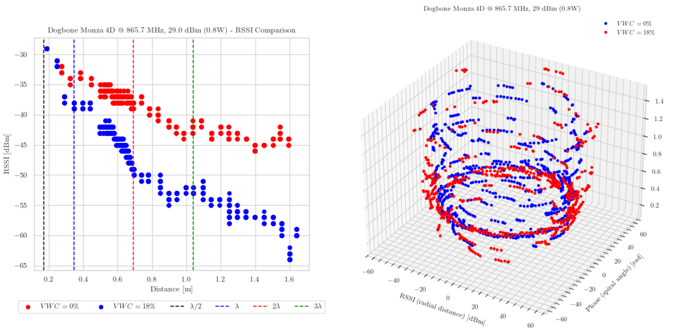
Fig. 2: Overview of the experimental setup.

C. Measurement Procedure

1) **Data Acquisition:** At each combination of moisture level, antenna height, and transceiver power, the following data was collected: *Height* - The vertical distance between the RFID antenna and the soil surface was systematically varied in 5cm steps and 1cm steps in regions of interest, *RSSI and Phase* - Values recorded continuously over a five second measurement window, *Moisture* - The soil's VWC was continuously recorded throughout the measurement session using capacitive sensors. Since each session lasted approximately 2h and was conducted on different days, evaporation effects were monitored by recording the volume of water added for calibration in 3L increments (data available at [16]).
2) **Environmental Control:** The experiment was conducted indoors to minimize external environmental influences such as wind and temperature variations. All tests were performed at room temperature.
3) **Data Synchronization:** Recorded data streams were timestamped to ensure accurate synchronization.

V. RESULTS AND DISCUSSION

In Fig. 3a we represent the relationship between RSSI and antenna height for the two VWC conditions considered in this work: 0% and 18%. As observed, there exists a clear dependence of signal attenuation on moisture level and propagation distance as described by (1). But the figure also shows the variability of real data with respect to the model. Under distance uncertainty, one can clearly observe that a bijective relation in terms of RSSI and moisture level does not exist. By incorporating the phase information (see Fig. 3b), ambiguity is reduced but data is still noisy.

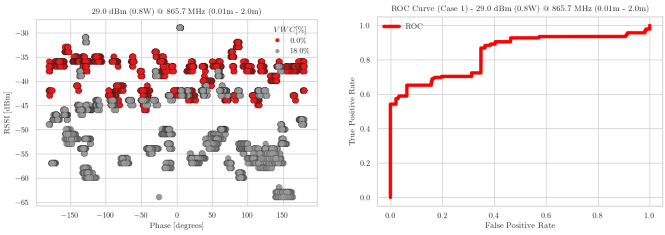


(a)

(b)

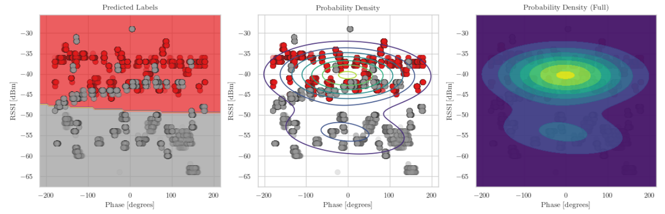
Fig. 3: Relationship between RSSI and antenna height for two VWC conditions (0% and 18%): a) RSSI vs. distance, b) RSSI and phase (x and y axis) vs. distance (z axis).

As commented in Section II, in this work we consider that h_r and d_g are not exactly known, but a distance range is available. Consequently, we consider two scenarios: Scenario 1 - Full Height Range (0.01m - 2.00m) and Scenario 2 - Limited Height Range (0.6m - 2.00m). As for the first scenario, in Fig. 4a, we show the phase and RSSI measurements for different reader heights and soil moisture levels. This information is unknown for the developed GMM but we use different colors to show how measurements related to different moisture levels are overlapping. In Fig. 4c, we are representing the clusters created by GMM. As observed, the algorithm is able to create two differentiated clusters, but a non-negligible number of samples belonging to the moist class (VWC of 18%) are grouped at the dry one. This is because the variation of RSSI levels observed when the soil moisture increases is higher. This could be because of multipath propagation induced by signal interactions within the wet soil. Once the clusters are created, we obtain a GMM-based soil moisture estimator. New RFID measurements can be used to estimate the moisture by computing the posterior probability given by (2). In Fig. 4b, we show the ROC of the obtained classifier. Although this scenario shows a non-negligible overlapping between classes, the proposed approach is able to show good results. The AUC is equal to 0.83 and, by interacting with the classification threshold, the accuracy can be increased to 76 %. The advantage of this method is that the threshold is directly related to the uncertainty of the measurement (the threshold is applied to the posterior probability). So, we can adjust to search for a better option in terms of measurement uncertainty and accuracy. A conservative selection (threshold equal to 0.5) provides us with a 70% of accuracy.



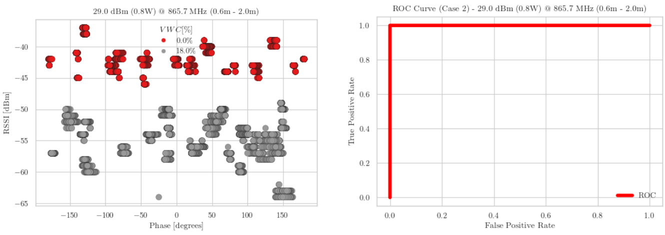
(a) Real phase-RSSI measurements for the two soil moisture cases (0% and 18% VWC).

(b) ROC Curve.



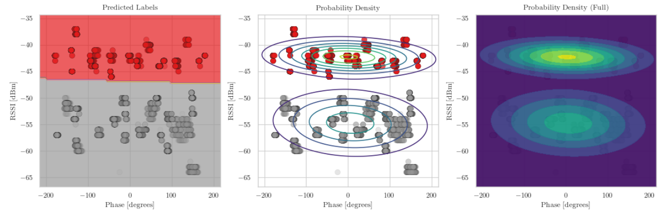
(c) GMM clusters generation and probability density results.

Fig. 4: Experimental results for Scenario 1: Full Height Range (0.01m - 2.00m).



(a) Real phase-RSSI measurements for the two soil moisture cases (0% and 18% VWC).

(b) ROC Curve.



(c) GMM clusters generation and probability density results.

Fig. 5: Experimental results for Scenario 2: Limited Height Range (0.6m - 2.00m).

Concerning the quality of obtained results, it is true that good accuracy results are usually set around 85-90%. But here it is worth mentioning that our method is an unsupervised approach with a set of assumptions that significantly limit the amount of required information (quite appropriate for real agricultural deployments). In order to show that the proposed method is able to attain good accuracy, we show results of Scenario 2 in Fig. 5. As expected, restricting the range to far-field distances ($h \geq 2\lambda$) enhances cluster separation. Regarding accuracy, the system collapses to 100%.

The complete dataset and implementation code used in this work are available at [16, 17].

ACKNOWLEDGEMENTS

This publication is part of the project PID2022-139929NB-I00, funded by MCIN/AEI/10.13039/501100011033/FEDER, UE, and the Project 2021 SGR 00197. The authors thank Mr. Ernesto Díaz Prieto for his technical support in the laboratory.

REFERENCES

[1] Mohd Javaid et al. “Enhancing smart farming through the applications of Agriculture 4.0 technologies”. In: *International Journal of Intelligent Networks* 3 (2022).
[2] Mehmet C. Vuran et al. “Internet of underground things: Sensing and communications on the field for precision agriculture”. In: *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), pp. 586–591.
[3] Ian F. Akyildiz and Erich P. Stuntebeck. “Wireless underground sensor networks: Research challenges”. In: *Ad Hoc Networks* 4.6 (2006), pp. 669–686.
[4] Christophe Cariou et al. “Internet of Underground Things in Agriculture 4.0: Challenges, Applications and Perspectives”. In: *Sensors* 23.8 (2023).
[5] M. Can Vuran and Agnelo R. Silva. “Communication Through Soil in Wireless Underground Sensor Networks – Theory and Practice”. In: *Sensor Networks*. Ed. by Gianluigi Ferrari. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 309–347.
[6] David M. Pozar. *Microwave Engineering*. Fourth edition. Hoboken, NJ: John Wiley & Sons, Inc, 2012. 1 p. ISBN: 978-0-470-63155-3 978-1-118-21363-6.
[7] Staley Arya and Girish Kumar Tiwari. “Characterizing Radio Frequency Transmission and Attenuation in Underwater Wireless Communication”. In: *Proc. 2023 SMART GENCON*. 2023.
[8] Roberta Avanzato et al. “Soil Moisture Estimation based on the RSSI of RFID Modules”. In: *Proc. IDAACS*. Vol. 2. 2021, pp. 647–652.
[9] Irene Cappelli et al. “Battery-Less HF RFID Sensor Tag for Soil Moisture Measurements”. In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021).
[10] Rafael V. Aroca et al. “Calibration of Passive UHF RFID Tags Using Neural Networks to Measure Soil Moisture”. In: *Journal of Sensors* 2018 (2018).
[11] Sérgio Pichorim, Nathan Gomes, and John Batchelor. “Two Solutions of Soil Moisture Sensing with RFID for Landslide Monitoring”. In: *Sensors* 18.2 (2018).
[12] Mehmet C. Vuran and Ian F. Akyildiz. “Channel model and analysis for wireless underground sensor networks in soil medium”. In: *Physical Communication* 3.4 (2010), pp. 245–254.
[13] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information science and statistics. New York: Springer, 2006. ISBN: 978-0-387-31073-2.
[14] amber-kshz. *PRML*. Version master. May 24, 2020. URL: https://github.com/amber-kshz/PRML.
[15] Svatopluk Matula, Kamila Báťková, and Wossenu Legese. “Laboratory Performance of Five Selected Soil Moisture Sensors Applying Factory and Own Calibration Equations for Two Soil Media of Different Bulk Density and Salinity Levels”. In: *Sensors* 16.11 (2016).
[16] Nedal Martínez Benelmekki. *Laboratory Measurments of Passive UHF RFID Tags*. Version DRAFT VERSION. 2025. DOI: 10.34810/data2325. URL: https://doi.org/10.34810/data2325.
[17] Nedal Martínez Benelmekki and Elvis Díaz Machado. *RFID Reader FX7500: A Toolkit for RFID Performance Measurements and Analysis*. Version 4.0.0. 2025. URL: https://github.com/Wireless-Information-Networking/RFID-Reader-FX7500.