Preliminary Study on RFID Signal Characteristics Above Water using Probabilistic Modeling with Sparse Gaussian Processes

Elvis Díaz Machado, Nedal Martínez Benelmekki, Javier Del Rio Toledano,
Antoni Morell, Jose Lopez Vicario.
{elvis.diaz, nedal.martinez, 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—**This paper presents a preliminary investigation into the relationship between UHF RFID signal characteristics (Received Signal Strength Indicator - RSSI, and phase) and the vertical distance between the reader antenna and a water surface. Due to experimental constraints preventing direct soil moisture measurement, the study was adapted to use a 50-liter container filled with water, varying the reader height and transmission power. Initial manual data inspection revealed an apparent linear relationship between the signal phase and reader height, as well as between phase and the volume of water added (affecting water level). We outline the modified experimental methodology and propose the use of Feedforward Neural Networks (FFNN) for baseline modeling and Sparse Gaussian Process Regression (SGPR) to model the observed relationships and quantify measurement uncertainties efficiently. This work serves as a foundational step towards developing robust RFID-based moisture sensing techniques, providing a methodology for controlled experimentation and data analysis using scalable probabilistic machine learning models.**

I. Introduction

The drive to optimize resource management, particularly water usage in agriculture and industrial processes, has spurred research into innovative sensing technologies. Monitoring liquid levels or soil moisture is critical, but traditional methods can be invasive, costly, or labor-intensive. Radio Frequency Identification (RFID) technology, primarily known for tracking and logistics, presents a potential alternative for non-invasive, indirect sensing [9, 10, 5, 11]. This approach leverages the interaction between electromagnetic waves and the dielectric properties of materials, such as water, potentially enabling efficient monitoring of liquid levels or moisture content.

The underlying principle relies on the effect of the medium (e.g., water, moist soil) on the propagation of radio waves. Changes in the water content or the distance the signal travels through or reflects off water alter the received signal's properties. Previous research, often focusing on soil moisture, has explored using passive UHF RFID tags, demonstrating correlations between signal characteristics (like attenuation or phase shift) and water content [6, 7]. The potential for remote interrogation makes RFID suitable for large-scale or automated monitoring.

However, indirect measurements are susceptible to uncertainties arising from environmental factors, material variability, and instrument limitations. Accurately mapping RFID signal characteristics (RSSI and phase) to physical variables like distance or water level requires methodologies that can effectively manage these uncertainties. Probabilistic modeling techniques, particularly Sparse Gaussian Process Regression (SGPR), are well-suited for this task. They allow for the quantification of uncertainty in measurements, model parameters, and predictions, while remaining computationally tractable for larger datasets often encountered in real-world sensing applications. This leads to more reliable interpretations and data-driven decisions.

This paper details a preliminary experimental methodology adapted to investigate the relationship between RFID signals and reader height above a water surface within a controlled setup. We systematically varied key parameters and collected signal data. We propose using FFNNs for baseline modeling and SGPR to explicitly incorporate and quantify uncertainty in a scalable manner. Initial observations suggest linear trends in phase measurements, motivating the application of these modeling techniques.

II. Experimental Setup

This research aimed to examine the correlation between the vertical distance to a water surface and RFID signal characteristics (RSSI and phase), accounting for measurement uncertainties. Due to unforeseen limitations in measuring soil moisture accurately within our setup, the experiment was revised to use a container of water.

A. Overview

The experimental design focused on creating a controlled environment to study the interplay between reader height, transmission power, water level, and RFID signals. Key parameters were methodically varied:

- **Reader Height:** The vertical distance between the RFID reader's antenna and the water surface was adjusted across multiple levels.
- **Transceiver Power:** The reader's transmission power was automatically varied using its API across a predefined set of levels.
- **Water Volume/Level:** Water was added incrementally to the container, changing the water level and thus the reflection point relative to the fixed tag (or potentially altering signal path if the tag was submerged).

These variations allowed observation of how each parameter influences RSSI and phase readings.

Fig. 1: Final experimental setup (water container).

B. Water Container Setup

Instead of soil, a 50-liter container was used and filled incrementally with water. Initially, the container was empty (representing the baseline). Water was then added in known volumes (e.g., 8-liter increments), allowing the water level to rise systematically. A passive UHF RFID tag was placed at a fixed position below the container, using a styrofoam case to maintain position, and provide a similar to air environment in close proximity.

C. RFID System

A Zebra FX7500 reader operating in the UHF band (specifically fixed at 865.7 MHz for this study) was used.

The transmission power of the reader was varied automatically using a script, incrementing by 4 dBm across a specified range [10-29.2] dBm (29.2 included) at each height setting, allowing assessment of power level impact on signal readings.

D. Measurement Procedure

Data acquisition occurred across various combinations of reader height, transceiver power, and water volume. The vertical distance ('height') between the reader antenna and the water surface was measured using a laser meter and adjusted systematically. Measurements were taken at 2 cm intervals in the [60-90] cm range identified as potentially interesting (roughly height of someone's hand at rest).

At each configuration (height, power, water volume), RSSI and phase values were recorded over 5-second intervals. The volume of water added was logged at each increment. Additional metadata from the reader (e.g., frequency, EPC, timestamps) were saved to $.csv$ files.

Experiments were conducted indoors to minimize environmental variability after initial outdoor tests confirmed comparable baseline readings. All data streams were timestamped for synchronization.

III. Data Analysis and Modeling

A. Initial Observations

Manual inspection of the collected data indicated observable trends. Specifically, the phase of the received RFID signal appeared to exhibit a roughly linear relationship with the reader height above the water surface. A similar linear trend was observed between the phase and the cumulative volume of water added to the container (which directly relates to the water level). RSSI values also varied and showed a logarithmic trend.

B. Modeling Approaches

The primary goal is to model the relationship between the controllable experimental parameters (reader height dist, transmission power power) and the observed RFID signal characteristics ('peakRssi', 'phase'). We employ two main machine learning models:

1) **Feedforward Neural Network (FFNN):** A standard FFNN is used as a baseline model for function approximation. It learns a deterministic mapping from inputs (height, power) to outputs (RSSI, phase). While often powerful predictors, FFNNs typically do not provide inherent uncertainty estimates with their predictions. The provided training recipe uses an Adam optimizer and standardizes inputs and outputs.
2) **Sparse Gaussian Process Regression (SGPR):** To address the need for uncertainty quantification and potentially scale to larger datasets, we employ SGPR. This probabilistic, non-parametric approach models the underlying function directly and provides principled uncertainty estimates for its predictions. It overcomes the computational limitations of standard Gaussian Processes for larger datasets through approximation techniques.

1) Sparse Gaussian Process Regression (SGPR): Gaussian Process Regression (GPR) is a powerful Bayesian method for regression that places a prior distribution over functions. Given training data, GPR computes a posterior distribution over functions, allowing for predictions at new input points along with associated uncertainty (variance) [3]. However, standard GPR requires computing and inverting the covariance matrix of the training data, which scales cubically with the number of data points N, i.e., $O(N^3)$ for training and $O(N^2)$ for prediction. This becomes computationally prohibitive for datasets beyond a few thousand points.

Sparse Gaussian Process Regression (SGPR) methods address this scalability issue by introducing a small set of M inducing points ($M \ll N$) [1, 2]. These inducing points act as a condensed summary of the full dataset. The core idea is that the predictions for the training data and test data are assumed to be conditionally independent given the function values at these inducing points.

Instead of working directly with the large $N \times N$ covariance matrix, SGPR methods typically optimize the locations of the M inducing points and the hyperparameters of the GP kernel (which defines the covariance between points) by maximizing a lower bound on the true marginal likelihood, often using variational inference techniques [2, 4].

The computational complexity of many SGPR methods scales approximately as $O(NM^2)$ for training and $O(M^2)$ for prediction, making them significantly more efficient than standard GPR when $M \ll N$. Despite the approximation, SGPR retains the key benefits of GPR:

- **Probabilistic Predictions:** Provides both a mean prediction and a variance (uncertainty estimate).
- **Non-parametric Flexibility:** Can model complex, non-linear relationships without predefined functional forms.
- **Kernel Engineering:** Allows incorporating prior knowledge about the function's properties (e.g., smoothness, periodicity) through kernel selection.

Common kernels include the Squared Exponential (RBF) kernel for smooth functions, Matérn kernels for controlling smoothness, and periodic kernels for cyclic patterns. The choice of kernel and the number of inducing points M are important modeling decisions.

2) Model Application in this Study: The FFNN and SGPR models will be trained on the collected dataset (using the provided data loader structure) to predict RSSI and phase based on height and power inputs. Their performance will be evaluated based on standard regression metrics (e.g., RMSE, MAE) and SGPR's ability to capture the observed trends and quantify prediction uncertainty. The scalability of SGPR makes it suitable should future experiments generate significantly larger datasets.

IV. Preliminary Results

This section outlines the format of results, incorporating initial visualizations derived from the collected data and the planned modeling approach. Detailed quantitative results from the full model training and evaluation are pending.

A. Model Performance Comparison

Visual comparisons provide further insight into model behavior. Figure 2 shows the absolute prediction error for both the FFNN (MLP) and SGPR models plotted against the measured RSSI and phase.

Figure 3 presents the Cumulative Distribution Function (CDF) of the absolute prediction errors for both models. This plot summarizes the overall error distribution; a curve shifted further to the left indicates that the model achieves lower errors more frequently. Based on this preliminary plot, the SparseGPR model appears to have a slightly better error profile, with a higher probability of producing very low errors compared to the MLP.

These results indicate the models' accuracy and the effectiveness of SGPR in quantifying prediction based on the experimental data, even with potentially large datasets.

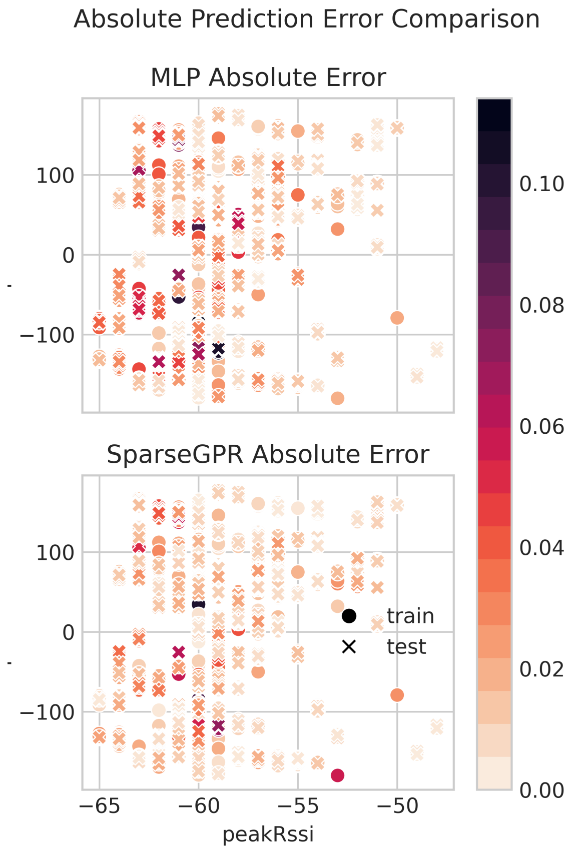

Fig. 2: Comparison of MAE for FFNN (MLP, top) and SparseGPR (bottom) across the observed RSSI/Phase.

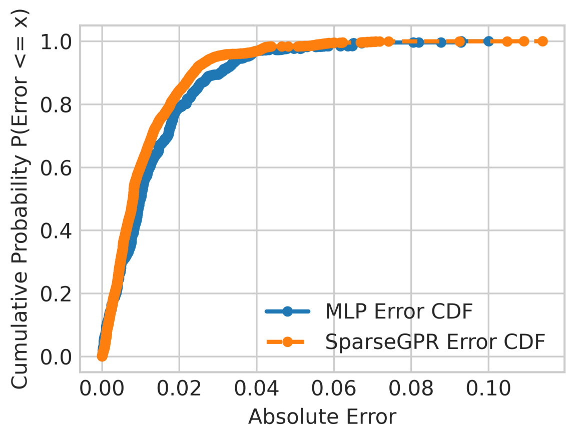

Fig. 3: Cumulative Distribution Function (CDF) of errors for the FFNN (MLP) and SparseGPR models.

V. Discussion

The preliminary findings suggest a predictable relationship, particularly linearity in phase, between RFID signal characteristics and the geometric setup (reader height, water level) in a controlled environment. This supports the potential use of phase information for distance or level sensing. The observed linearity simplifies modeling but also highlights the sensitivity of phase measurements.

The application of FFNN and SGPR provides a framework for not only predicting signal behavior but also understanding the associated uncertainties in a computationally efficient manner. SGPR is expected to offer valuable insights into the reliability of predictions, identifying regions in the parameter space (e.g., specific heights or power levels) where measurements are more or less certain, without the computational burden of full GPR. The use of inducing points in SGPR makes this approach scalable to the larger datasets anticipated in future, more extensive experiments or real-world deployments.

Limitations of this preliminary study include a single fixed RFID frequency, a simplified indoor environment. The relationship observed might change with different materials, water properties (e.g., salinity), tag types, frequencies, or in more complex multipath environments.

Future work should involve extending experiments to different materials, frequencies, and potentially outdoor or more realistic settings, generating larger datasets where the benefits of SGPR become more pronounced. Also, validating the models' predictive capabilities and uncertainty quantification in these varied conditions. Exploring the simultaneous prediction of RSSI and phase using multi-output SGPR models [8] could also yield more robust models by capturing potential correlations between the signal characteristics. Further investigation into optimal kernel selection and the number/placement of inducing points for SGPR in this specific application context is also warranted.

VI. Conclusion

This paper outlined a modified experimental methodology for investigating UHF RFID signal characteristics (RSSI, phase) in response to varying reader height and water level within a 50L container. Initial data exploration revealed promising linear trends, particularly for signal phase relative to reader height and water volume. We proposed a modeling approach using FFNNs for baseline performance, and Sparse Gaussian Process Regression (SGPR) for scalable probabilistic modeling capable of quantifying prediction uncertainty. Preliminary results illustrate the expected trends and the comparison of model performance. This preliminary work establishes a foundation for further research into RFID-based sensing, emphasizing the importance of controlled experimentation and scalable probabilistic modeling like SGPR to handle inherent measurement uncertainties and larger datasets effectively.

VII. Data Availability

The dataset collected during this study, along with the code used for analysis and modeling (including the data loader and training scripts presented), are intended to be made available upon completion of the analysis in a public GitHub repository:

https://github.com/Wireless-Information-Networking/water_measures

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] Edward Snelson and Zoubin Ghahramani. "Sparse Gaussian Processes using Pseudo-inputs". In: *Advances in Neural Information Processing Systems*. Ed. by Y. Weiss, B. Schölkopf, and J. Platt. Vol. 18. MIT Press, 2005. URL: https://proceedings.neurips.cc/paper_files/paper/2005/file/4491777b1aa8b5b32c2e8666dbe1a495-Paper.pdf.

[2] Michalis Titsias. "Variational Learning of Inducing Variables in Sparse Gaussian Processes". In: *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*. Ed. by David van Dyk and Max Welling. Vol. 5. Proceedings of Machine Learning Research. Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA: PMLR, 16–18 Apr 2009, pp. 567–574. URL: https://proceedings.mlr.press/v5/titsias09a.html.

[3] Carl Edward Rasmussen and Hannes Nickisch. "Gaussian Processes for Machine Learning (GPML) Toolbox". In: *J. Mach. Learn. Res.* 11 (Dec. 2010), pp. 3011–3015. ISSN: 1532-4435.

[4] James Hensman, Nicolò Fusi, and Neil D. Lawrence. "Gaussian processes for Big data". In: *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*. UAI'13. Bellevue, WA: AUAI Press, 2013, pp. 282–290.

[5] C. Bauer-Reich et al. "An investigation of the viability of UHF RFID for subsurface soil sensors". In: *IEEE International Conference on Electro/Information Technology*. 2014, pp. 577–580. DOI: 10.1109/EIT.2014.6871828.

[6] Rafael V. Aroca et al. "Calibration of Passive UHF RFID Tags Using Neural Networks to Measure Soil Moisture". In: *Journal of Sensors* 2018 (2018), pp. 1–12. ISSN: 1687-7268. DOI: 10.1155/2018/3436503. URL: http://dx.doi.org/10.1155/2018/3436503.

[7] Sérgio Pichorim, Nathan Gomes, and John Batchelor. "Two Solutions of Soil Moisture Sensing with RFID for Landslide Monitoring". In: *Sensors* 18.2 (Feb. 2018), p. 452. ISSN: 1424-8220. DOI: 10.3390/s18020452. URL: http://dx.doi.org/10.3390/s18020452.

[8] Wessel P. Bruinsma et al. "Scalable exact inference in multi-output Gaussian processes". In: *Proceedings of the 37th International Conference on Machine Learning*. ICML'20. JMLR.org, 2020.

[9] Colleen Josephson et al. "Time-of-Flight Soil Moisture Estimation Using RF Backscatter Tags". In: *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*. 2020, pp. 5049–5052. DOI: 10.1109/IGARSS39084.2020.9323308.

[10] Roberta Avanzato et al. "Soil Moisture Estimation based on the RSSI of RFID Modules". In: *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. Vol. 2. 2021, pp. 647–652. DOI: 10.1109/IDAACS53288.2021.9661031.

[11] Irene Cappelli et al. "Battery-Less HF RFID Sensor Tag for Soil Moisture Measurements". In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021), pp. 1–13. DOI: 10.1109/TIM.2020.3036061.