FDTD Medium Dimension Selection Guidelines for GPR Synthetic Data Generation

Noushin Khosravi Largani, Seyed (Reza) Zekavat, and Himan Namdari

***Abstract*—Ground-penetrating radar (GPR) has been traditionally used for subsurface assessment. In many applications, such as precision agriculture via drone-borne radar, it is critical to use Machine Learning (ML) techniques to map GPR received signals into soil subsurface moisture and texture. Supervised ML methods need a large number of labeled data for their training process which is expensive and time-consuming to attain through actual field measurements. The gprMax software, which is created based on the Finite Difference Time Domain (FDTD) method, has been introduced as a reliable tool to emulate soil media and create synthetic labeled data. Proper selection of gprMax soil medium dimensions is critical to the generation of reliable synthetic data. The selection of large soil medium dimensions for gprMax emulations leads to synthetic data consistent with realistic scenarios. However, larger medium dimensions lead to higher computation complexity. This letter investigates and validates a proper selection of medium dimensions that maintains a trade-off across the accuracy and computational complexity of creating synthetic data. The results of this study are critical to researchers who adopt gprMax or any FDTD-oriented emulations for soil subsurface assessment.**

***Index Terms*—GPR, Impulse Response, FDTD, Emulation, Soil, Medium Dimension.**

I. Introduction

SOIL SUBSURFACE investigation is key to hydrological studies, environmental monitoring, and precision agriculture [1]–[3]. Ground-penetrating radar (GPR) is a conventional tool for soil subsurface assessment. GPR transmits electromagnetic waves towards the ground to characterize soil subsurface based on the received signal [2], [4]. Traditional GPR signal processing methods are complex and thus it is hard to leverage them for subsurface assessment of a large area such as a megafarm [5], [6]. Machine Learning (ML) can be considered as an effective approach for (real-time) subsurface assessment of large areas. However, supervised ML methods need a large number of labeled data that is very difficult to create using real field measurements for GPR applications. gprMax is open-source software that uses the Finite Difference Time Domain (FDTD) technique [7] capable of creating a large number of labeled GPR synthetic data.

The soil medium acts as a channel with an impulse response. The soil channel impulse response (CIR) is a key feature that can be used to extract soil subsurface information [4], [8]. Soil CIR contains information about the illuminated soil medium such as permittivity and number of layers at the frequency of interest. Soil CIR corresponds to:

$$h(t) = \sum_{l=1}^{L} a_l \delta(t - \tau_l), \tag{1}$$

in which a_l, τ_l, and L indicate channel gains, delays, and the number of multiple reflections. Parameters such as permittivity, conductivity, and depth of the medium determine a_l and τ_l [4].

The dimensions of a soil medium (i.e., x, y, and z) are shown in Fig. 1, where x and y represent the medium surface dimension, and z refers to the medium depth. If the dimensions of the illuminated medium vary, the soil CIR will change. Thus, dimension alterations impact the accuracy of subsurface feature analysis. As the medium dimensions increase, emulations become more accurate but expensive in terms of time and the required memory.

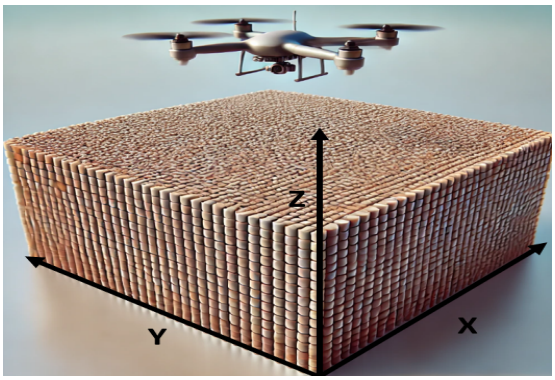

Fig. 1. The soil medium and its dimensions x, y, and z.

In this letter, we investigate the adoption of proper medium dimensions (x, y, and z) for gprMax emulations that support high precision and low complexity CIR extraction. To assess the impact of dimensions on gprMax synthetically generated data, we increase the medium dimensions gradually and obtain the corresponding CIR. The Euclidean distance between the extracted CIR channel gains for different consecutive medium dimensions is used as the precision measure. FDTD computation time is considered as the emulation Complexity. We select the dimension that facilitates a balance between precision and Complexity. The results of this letter can be used as a guideline for researchers who aim to use gprMax (or any FDTD-based synthetic data generation) to emulate soil medium illuminated by GPR signals. The results confirm that a surface dimension consistent with 1.5 times the maximum wavelength leads to acceptable precision and complexity.

This study is supported by funding from the United States Department of Agriculture (Grant Number: USDA NR223A750013G032).

Noushin Khosravi Largani, Seyed (Reza) Zekavat, and Himan Namdari are with the Data Science Department of Worcester Polytechnic Institute, USA (e-mail: nlargani@wpi.edu; rezaz@wpi.edu).

The method for impulse response extraction, the emulation parameter selection, and the verification of minimum surface dimension are explained in Sections II, III, and IV, respectively. The letter concludes in Section V.

II. Impulse Response Extraction

We use the transmitted and received signals that are created in the time domain by gprMax to extract the CIR of the soil medium that was presented in (1). Some initial components of the received signal correspond to the leakage between the antennas of the transmitter and receiver. To create a CIR that only includes the impact of the soil channel, we need to eliminate the antenna leakage. After removing the leakage effect, we obtain the FFT of the received signal that is expressed as:

$$Y = HX + N, \tag{2}$$

where Y, H, X, and N denote the Fourier transform of the received signal, CIR, transmitted signal, and noise, respectively. The CIR in the frequency domain is represented by $H = \frac{Y}{X}$ [4]. The CIR in the time domain ($h(t)$) is obtained using IFFT. The whole process has been detailed in Fig 2.

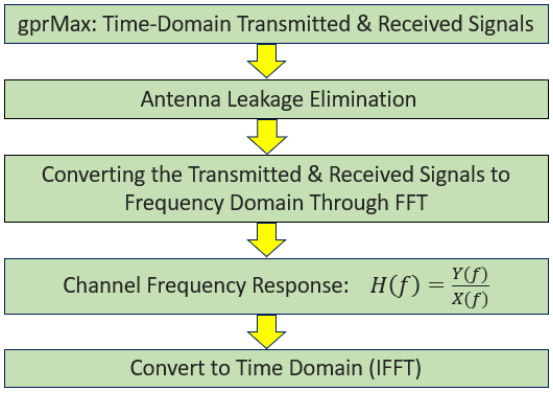

Fig. 2. The flowchart of extracting soil CIR.

III. Emulation Parameter Selection

We considered a single layer medium as shown in Fig. 1. for our emulations. Based on studies conducted in [9]–[11], we have selected the following pairs for soil medium permittivity (ϵ) and conductivity (σ):

$$(\epsilon, \sigma) = (4, 1), (8, 80), (38, 80), (38, 20). \tag{3}$$

We select Ricker for medium illumination as this waveform is widely used for GPR inspections and emulates pulses of real GPR systems. The center frequency (f_c) of Ricker is set to 825 MHz. The minimum (f_{min}) and maximum (f_{max}) frequencies of Ricker are derived based on the equations presented in [12], which have the values of 375.168 MHz and 1274.831 MHz, respectively. This frequency range is selected to maintain consistency with the soil Peplinski model, which is valid for frequencies from 300 MHz to 1300 MHz [13]. Therefore, the bandwidth of the transmitted waveform ($f_{max} - f_{min}$) is 899.663 MHz, and accordingly, the range resolution is approximately 17 cm. Thus, we set the depth of the illuminated medium to 20 cm (more than the range resolution). We adopt a Hertzian dipole antenna located at the center of the x-y plane with a height of 41 cm from the sdoil medium surface. The Hertzian dipole antenna is well-suited for qualitative observations, making it suitable for studying the impact of medium size. The selected height is more than $\frac{\lambda_{max}}{2}$, lying within the intermediate-field region of the antenna and providing appropriate output [14]. The dipole antenna's omnidirectional radiation pattern enables it to capture reflections from all directions. Among various types of dipole antennas, the Hertzian dipole exhibits the broadest pattern size [14], facilitating a more comprehensive medium-scale verification.

The minimum wavelength (23.5 cm) determines the FDTD pixel size in gprMax. The pixel size is set to 0.001m, based on the rule of thumb condition $\Delta x \leq \frac{\lambda_{min}}{10}$ [15]. Then, The vertical distance between the antenna and the medium's upper boundary can be calculated by considering a minimum of 20 pixels that is required for gprMax implementation. The surface (x-y plane) dimension of the medium is kept variable to study the proper dimension that maintains a trade-off between accuracy and time-memory efficiency. Table I summarizes the parameter selection.

TABLE I
Parameters Selected for Emulations.

Parameter	Value
Transmitted waveform	Ricker, $f_c = 825$ MHz
Depth of medium	20 cm
Antenna polarization	Y
Pixel size	0.001
Time window	10^{-8} s

IV. gprMax Minimum Surface Dimension Verification

One of the main disadvantages of the FDTD method is its brute-force calculation nature, which causes the memory requirements to scale with the third order of the emulation domain size $O((N)^3)$, and the computation time to scale with the fourth order $O((N)^4)$ [16]. N refers to the number of grid points in a single dimension. Hence, we consider the maximum order (i.e., N^4) as the bottleneck of the emulation Complexity. We set $x = y$ (because of the pattern of the dipole antenna) starting from 80 cm to 300 cm with a step size of 20 cm. The channel gains vector obtained for each pair of x,y is compared to the next x,y pair in terms of Euclidean distance to create an Error Percentage measure that corresponds to:

$$Error\ Percentage = \frac{\|d_i - d_{i-1}\|_2}{\|d_{i-1}\|_2} \times 100\%, \tag{4}$$

where d_i is the vector of a_l for $i = 100 : 300\ cm$ with the step of 20 cm, and d_{i-1} is the vector of a_l for $i = 80 : 280$ with the step of 20. $\|d_i - d_{i-1}\|_2$ represents the Euclidean distance between d_i and d_{i-1}. $\|d_i\|_2$ denotes ℓ_2-norm of vector d_i.

Fig 3 shows both the Error Percentage as well as the computational Complexity. As the Error Percentage is relatively calculated (between d_i and d_{i-1}), the horizontal axis shows the surface dimensions related to d_i. Thus, $x = 80\ cm$ is not represented in the Fig 3. It is observed that the channel

gains tend to get closer as the x, and y dimension increases. The dimensions x,y = 120 cm and more can be observed in the values where the Euclidean distance between channel gain vectors (d_i and d_{i-1}) is acceptably small, getting better from 120 cm to 300 cm. In addition, it is shown that the Complexity increases with the medium dimension. Based on these curves, we suggest that medium dimensions between 120-140 cm lead to low Error Percentage and Complexity.

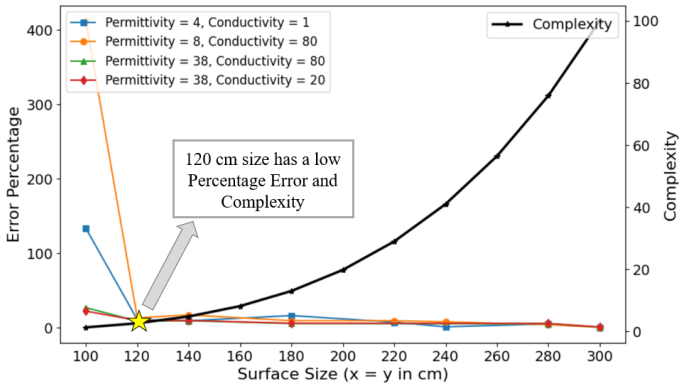

Fig. 3. Measure and Complexity for different x,y dimensions, $d_i = 100\ cm$.

Fig 4 is a zoom-in version of Fig 3 for d_i starting from $i = 120 : 300\ cm$. We have also removed the Complexity curve from this figure to allow a focus on values around 120 cm in dimension. It is observed that for different ranges of permittivity and conductivity, the difference between channel gains tends to zero when the dimension size for x and y reaches 300 cm. The observed fluctuations are caused by numerical dispersion in FDTD computations [17].

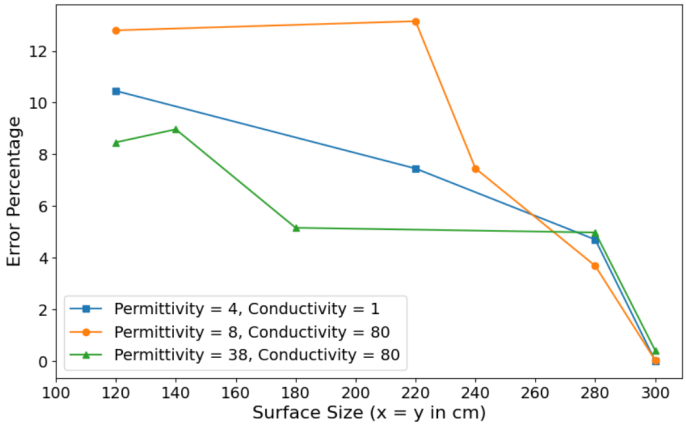

Fig. 4. Measure for different x,y dimensions, $d_i = 120\ cm$.

The maximum and minimum practical values for soil permittivity (38 and 4) and conductivity (80 and 1) adopted in (3) create a broader perspective for selecting an appropriate dimension. Permittivity directly affects the wavelength and conductivity affects the wave propagation and subsequently the channel gains. To adequately represent the wave's characteristics and to minimize boundary effects, the medium dimensions should be selected much higher (e.g., $5\lambda_{max}$) than the maximum wavelength [16]. However, the implementation of gprMax using $5\lambda_{max}$ dimension is computationally expensive in terms of time and memory. Thus, based on Fig. 3 and Fig. 4, we adopt a dimension of $1.5\lambda_{max}$, that is a compromise across precision and computation cost.

V. Conclusion

This letter studies the minimum acceptable medium dimensions to maintain proper precision *and* Complexity using gprMax. Emulations are conducted for a wide range of soil electric features, i.e., permittivity and conductivity values. To maintain a trade-off between accuracy and complexity for different soil features, the letter confirms that the minimum medium surface dimension should be in the order of $1.5\lambda_{max}$.

References

[1] V. Filardi, A. Cheung, R. Khan, O. Mangoubi, M. Moradikia, S. R. Zekavat, B. Wilson, R. Askari, and D. Petkie, "Data-driven soil water content estimation at multiple depths using SFCW GPR," in *2023 IEEE International Opportunity Research Scholars Symposium (ORSS)*, pp. 86-90, IEEE, 2023.

[2] H. Namdari, M. Moradikia, R. Askari, O. Mangoubi, D. Petkie, and S. Zekavat, "Advancing precision agriculture: Machine learning-enhanced GPR analysis for root-zone soil moisture assessment in mega farms," *IEEE Trans. Agrifood Electron.*, accepted for publication, 2024.

[3] H. Namdari, M. Moradikia, D. T. Petkie, R. Askari, and S. Zekavat, "Comprehensive GPR signal analysis via descriptive statistics and machine learning," in *2023 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)*, pp. 127-132, IEEE, 2023.

[4] S. Zheng, X. Pan, A. Zhang, Y. Jiang, and W. Wang, "Estimation of echo amplitude and time delay for OFDM-based ground-penetrating radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2384-2388, 2015.

[5] H. M. Jol, Ed., *Ground Penetrating Radar Theory and Applications*. Elsevier, 2008.

[6] L. B. Conyers, *Ground-Penetrating Radar for Archaeology*. Rowman & Littlefield, 2023.

[7] C. Warren, A. Giannopoulos, and I. Giannakis, "gprMax: Open source software to simulate electromagnetic wave propagation for ground penetrating radar," *Computer Phys. Commun.*, vol. 209, pp. 163-170, 2016.

[8] H. Huh, H. Goh, J. W. Kang, S. François, and L. F. Kallivokas, "Using the impulse–response pile data for soil characterization," *J. Eng. Mech.*, vol. 149, no. 10, pp. 04023078, 2023.

[9] S. Friedman, "Electrical properties of soils," in *Encyclopedia of Agrophysics*. Dordrecht: Springer Netherlands, 2011, pp. 242-255.

[10] J. Cihlar and F. T. Ulaby, *Dielectric Properties of Soils as a Function of Moisture Content*, NASA-CR-141868, 1974.

[11] B. Allred, J. J. Daniels, and M. R. Ehsani, *Handbook of Agricultural Geophysics*. CRC Press, 2008.

[12] Y. Wang, "Frequencies of the Ricker wavelet," *Geophysics*, vol. 80, no. 2, pp. A31-A37, 2015.

[13] N. R. Peplinski, F. T. Ulaby, and M. C. Dobson, "Dielectric properties of soils in the 0.3-1.3-GHz range," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 3, pp. 803-807, 1995.

[14] C. A. Balanis, *Antenna Theory: Analysis and Design*. John Wiley & Sons, 2016.

[15] K. S. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media," *IEEE Transactions on Antennas and Propagation*, vol. 14, no. 3, pp. 302–307, May 1966.

[16] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed. Artech House, 2005.

[17] B. Finkelstein and R. Kastner, "A comprehensive new methodology for formulating FDTD schemes with controlled order of accuracy and dispersion," *IEEE Trans. Antennas Propag.*, vol. 56, no. 11, pp. 3516-3525, 2008.