



OPEN

DATA DESCRIPTOR

# A realistic 2D multi-offset, multi-frequency synthetic GPR data set as a benchmark for testing new algorithms

G. Roncoroni<sup>1</sup>, P. Koyan<sup>2</sup>✉, E. Forte<sup>1</sup>, J. Tronicke<sup>2</sup> & M. Pipan<sup>1</sup>

We present a 2D multi-offset, multi-frequency synthetic GPR data set specifically designed to evaluate and test processing, analysis and inversion techniques. The data set replicates realistic subsurface conditions at four sections separated by 2 m. We modeled four multi-offset GPR profiles at 50, 100 and 200 MHz frequencies using realistic wavelets. The data set provides a robust framework for validating advanced GPR algorithms and techniques such as pre-stack depth migration, amplitude versus offset analysis and full waveform inversion. Extensive technical validation ensures data reproducibility and affordability. The standardized, realistic synthetic data set can be used as a reliable benchmark for developing and testing new algorithms and methods, thereby advancing the understanding of subsurface imaging and real-world data interpretation.

## Background & Summary

Ground-penetrating radar (GPR) is a non-invasive geophysical method widely used for subsurface imaging in various fields such as geology, archaeology, environmental sciences, and civil engineering<sup>1</sup>. Multi-offset (MO) GPR, a technique that involves deploying multiple receivers along a transect to observe various wave types, has gained significant attention due to its ability to enhance subsurface imaging compared to traditional common-offset (CO) applications<sup>2–4</sup>. Specifically, MO GPR can assure that each subsurface point is imaged by multiple wavefronts, while in the commonly employed CO data collection each point in the subsurface is only sampled by a single wavefront. Such a technique is usually referred as to “Multi-fold” (MF) and was originally applied in reflection seismic, contributing to an exponential growth soon after the digital recording revolution. MO GPR offers benefits such as detailed estimates of subsurface electromagnetic wave velocity fields, improved signal-to-noise ratio in reflection sections, and the potential to adapt advanced seismic processing schemes for GPR data<sup>5</sup>. By combining MO data with advanced processing and analysis techniques like pre-stack depth migration<sup>6</sup> or amplitude-versus-offset (AVO) analysis<sup>7</sup>, it becomes possible to reduce noise, estimate 2D or 3D velocity models, and enhance the quality of subsurface images and estimates of physical parameters<sup>8</sup>.

Quantitative analysis and inversion of MO GPR surveys involve estimating subsurface properties such as moisture content, soil water content, or hydraulic conductivity. Recent advancements in MO GPR have focused on quantitative off-ground approaches, improved resolution through full-waveform inversion, and the utilization of time-lapse measurements to gain new insights into dynamic soil hydrologic processes<sup>9,10</sup>. The application of MO GPR extends beyond traditional GPR studies to areas like agriculture, where it has been used to link horizontal cross-hole variability with root image information for crop root system analysis<sup>11</sup>. Additionally, MO GPR has been employed in environmental studies to monitor dynamic unsaturated flow phenomena and to estimate saturated hydraulic conductivity in sandy soils<sup>12,13</sup>. The technique has also shown promise in detecting and monitoring contaminants like LNAPL (Light Non-Aqueous Phase Liquid) in aquifers<sup>14</sup>. Other successful applications focus on archaeological surveys demonstrating the possibility to obtain improved subsurface imaging, even in highly inhomogeneous subsurface conditions<sup>15</sup>. However, MO GPR is not a standard technique due to its inherent logistical constraints and the very demanding time required to acquire the data.

<sup>1</sup>University of Trieste, Department of Mathematics, Informatics and Geosciences, Trieste, Italy. <sup>2</sup>University of Potsdam, Institute of Geosciences, Potsdam, Germany. ✉e-mail: [koyan@uni-potsdam.de](mailto:koyan@uni-potsdam.de)

Full waveform inversion (FWI) is a state-of-the-art technique widely recognized for its effectiveness in imaging subsurface structures and physical parameters using seismic data<sup>16</sup>. Despite its advantages, the implementation of FWI comes with significant challenges that researchers are actively addressing. Recent advancements in the field have seen the application of Convolutional Neural Networks (CNNs) to solve inverse problems in imaging, showing improvements over traditional methods like compressed sensing<sup>17</sup>. Additionally, studies have highlighted the use of numerical simulations as the foundation for GPR inversion techniques, such as FWI, which has shown promise in enhancing resolution compared to conventional methods<sup>18,19</sup>. Researchers have also explored novel approaches like optimizing acquisition setups for cross-hole GPR FWI using checkerboard analysis, which leverages the full recorded signal to improve imaging accuracy<sup>20</sup>. Furthermore, innovative FWI approaches have been developed to estimate parameters like the radius of subsurface cylindrical objects, showcasing the potential of FWI in diverse applications beyond traditional geological studies<sup>21</sup>. Studies have also delved into applications of FWI in assessing concrete properties like chlorides and moisture content, demonstrating the versatility of FWI in material science investigations<sup>22</sup>. Moreover, FWI has been utilized for mapping soil moisture profiles at the field scale, emphasizing the capability of FWI to maximize information retrieval accuracy from GPR data<sup>23</sup>. The integration of FWI with other geophysical data, such as electrical resistivity tomography and reflection seismics<sup>24,25</sup>, presents opportunities for joint inversion approaches that are not commonly explored, but hold promise for future research directions<sup>26</sup>. Additionally, the use of generative adversarial networks in deterministic inversion approaches has been investigated, showcasing the potential of machine learning in enhancing inversion processes<sup>27</sup>.

In recent years, multi-channel GPR equipment has been developed with arrays of antennas, in which each of them can be used either as a transmitter or as a receiver. An array is a noticeable method of increasing the productivity rate by collecting several parallel profiles instead of just one. In this way it is also possible to improve the spatial resolution aiming to collect full-resolution 3D datasets<sup>28</sup> although not directly exploiting the above-described MF advantages. Examples of 3D GPR surveys with arrays of antennas are reported for archaeological<sup>29–31</sup>, engineering<sup>32</sup> and road inspection applications<sup>33</sup>.

To collect MF GPR data, several studies with either customary multichannel systems<sup>34</sup> or combining time-consuming single-channel measurements<sup>2,3</sup> have been used. Using only one transmitting and one receiving antenna and keeping the azimuth of the antennas constant there are three convenient acquisition geometries exploited to collect MF data sets<sup>4</sup> (Figure Supplementary 1). The easier, but more time-consuming way, is to symmetrically increase the antennas offset keeping their midpoint constant. This acquisition geometry is usually referred to as Common Midpoint Gather (CMP) as in reflection seismics. An alternative way is to move only one antenna away from the other (this geometry is often referred to as Wide Angle Reflection and Refraction – WARR, while in reflection seismic is reported to as Common Shot Gather – CSG). Collecting a series of separated records laterally shifted by a constant distance, it is then possible to combine and sort the data into several CMPs. Another way to obtain MF data, is acquiring several CO profiles along the same path with a different offset each and then sort the data as CMPs. This is the least time-consuming approach, but indeed it is time demanding since a maximum subsurface folding  $n$  can be obtained by subsequently collecting  $n$  separated CO profiles, each of them with a different offset. Further details about the different acquisition geometries of MO GPR data can be found in<sup>4</sup>.

Using the last approach<sup>15</sup>, compared CO and MO results for archaeological applications, while<sup>35</sup> used reflection tomography to estimate water content variations. A multi-channel system with four receiving antennas is reported by<sup>36,37</sup> to estimate soil porosity and water content using different offsets.

Even though these approaches are promising, all the described strategies have some disturbances because combining single-channel measurements, despite the used acquisition geometry is time consuming<sup>4,38</sup> and the accuracy depends on the actual location of the antennas on the ground, often requiring smoothing of scattered coordinates and data binning<sup>37</sup>.

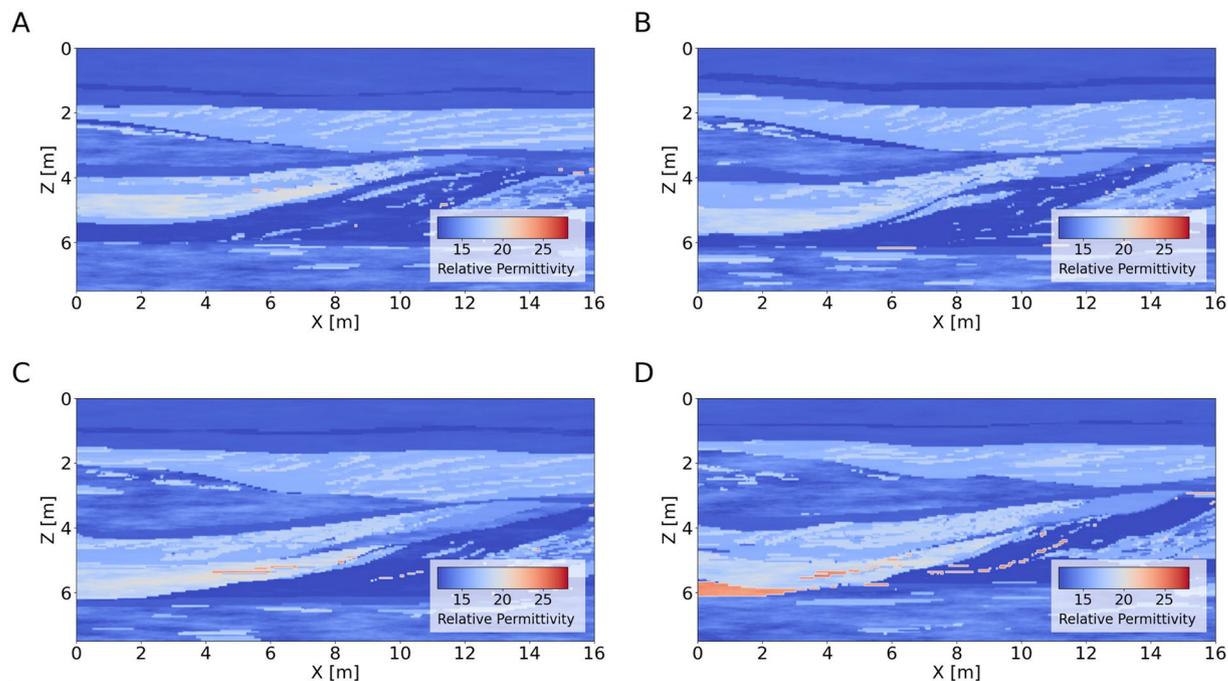
A few years ago, a multi-channel and MO equipment (WARR machine or SPIDAR) was tested and launched on the market by Sensors & Software (Mississauga, ON, Canada), allowing measuring with up to eight channels connected with different antennas offsets (almost) simultaneously<sup>39,40</sup>. This new system is specifically developed to make it possible to collect MO GPR data with up to seven different offsets (ranging from 0.25 m to 1.75 m) at the same speed as one single-channel (constant offset) data set. Such a new instrument has been used for different purposes including soil characterization<sup>41</sup> and improved subsurface velocity analysis<sup>42</sup>.

Despite the previously cited examples, nowadays GPR data sets are mainly acquired using CO geometries and antenna arrays are usually exploited in order to reduce the data acquisition time, increase the spatial resolution, and obtain 3D full-resolution data sets.

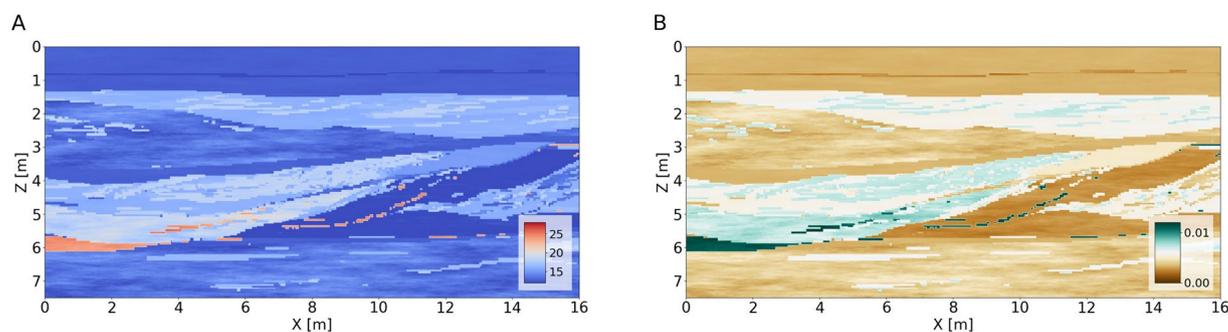
It is therefore crucial to have MO (and therefore MF) data with a large enough offset range (depending also on the frequency of the antenna used) and high spatial coverage to properly test the performance of GPR processing, analysis, and inversion algorithms.

In reflection seismics, there is a well-known synthetic data set called “Marmousi” which has been used for more than 35 years as an industry standard and benchmark data set to test, evaluate, compare, and implement advanced processing and inversion techniques. The Marmousi model<sup>43,44</sup> was created in 1988 mimicking the geometries and the physical parameters of real seismic data of the Kwanza (a.k.a. Quanza) Basin (Angola). Since its original implementation the model was geometrically extended and modified from acoustic to fully elastic, the latter being usually referred to as Marmousi2<sup>45</sup> as a consequence of the advancement in computer hardware capabilities and new algorithms implementation. More recently, a 3D version of the model was made available for the scientific community<sup>46</sup> allowing for simulating even complex 3D seismic data sets using various acquisition geometries.

A similar standard does not exist for GPR and, up to now, there are no MO synthetic but realistic GPR data sets made available to the scientific community to test new algorithms and procedures. It limits the



**Fig. 1** Relative permittivity [adimensional] along the 4 selected sections from the Herten 3D model. (A)  $y = 2$  m; (B)  $y = 4$  m; (C)  $y = 6$  m; (D)  $y = 8$  m.



**Fig. 2** Relative permittivity [adimensional] (A) and conductivity [S/m] (B) for the section at  $y = 8$  m as depicted in Fig. 1D.

reproducibility of the results and makes difficult to understand the subjectivity in choosing parameters and defining flows. To overcome these limitations, we simulate, present, and make publicly available a MO and multi-frequency data set across the model made available by<sup>47</sup> for which the same authors already provided a 3D single frequency CO data set<sup>48</sup>.

## Methods

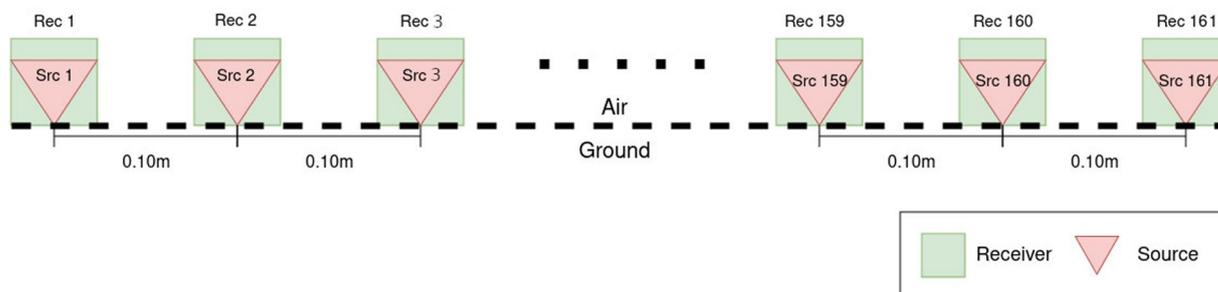
The sedimentary model used in the 3D modeling of GPR data bases on a high-resolution hydrofacies data set obtained from an aquifer-analog study within fluvio-glacial deposits<sup>49</sup>. The model represents a gravel quarry near the village of Herten in SW-Germany, where sand and gravel sequences formed in a braided-river regime characterize the near surface. The data set includes detailed hydrogeological properties and their spatial distributions within the quarry, covering an area of  $16 \times 10$  m ( $xx$ ) and a depth range ( $z$ ) of 7 m with a resolution of 0.05 m<sup>50</sup>. From this 3D model, we select 4 sections at 2 m, 4 m, 6 m, and 8 m along the  $y$  dimension, as depicted in Fig. 1 in terms of relative permittivity assuming a fresh-water saturated scenario.

The sedimentary model exhibits a variety of realistic features at different spatial scales, including thin interfaces and dipping layer sequences with varying electrical parameter contrasts (Fig. 2). These features make the model a challenging yet ideal target for testing and evaluating novel 2D GPR processing and inversion methods, for example FWI, migration or deconvolution algorithms.

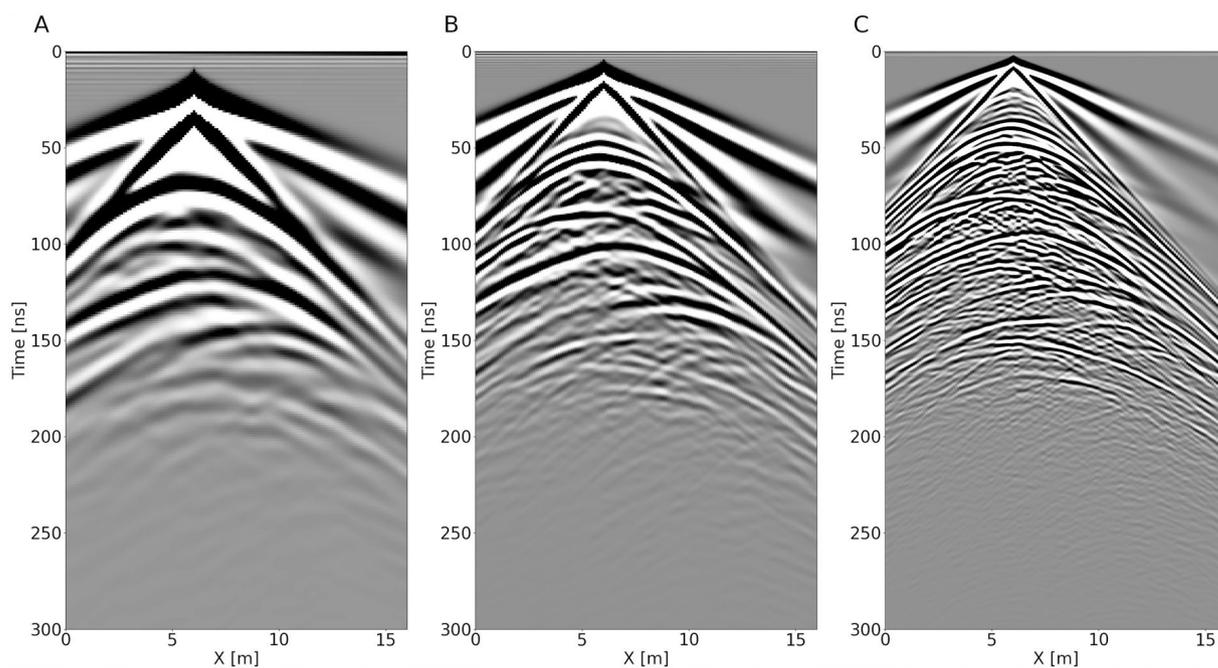
To perform forward modeling (i.e., to simulate synthetic GPR data), we use gprMax v.3.1.7<sup>51,52</sup>, an open-source electromagnetic modeling software (<https://github.com/gprmax/gprMax>) specifically developed for simulating GPR data using the Finite-Difference Time-Domain (FDTD) numerical method<sup>53</sup>. By utilizing the propagation physics of electromagnetic waves and the FDTD method, gprMax enables accurate simulation

Central frequency $f$ [MHz]	Timestep $dt$ [ns]	Recording time $t$ [ns]	Model discretization $d_{xz}$ [m]	Source/receiver position $x_{src}/x_{rec}$ [m]	Source/receiver position $y_{src}/y_{rec}$ [m]	Source/receiver position $z_{src}/z_{rec}$ [m]
50	0.116	500	0.05	0:0.1:16	2:2:8	0
100	0.058		0.025			
200	0.029		0.0125			

**Table 1.** Simulation parameters for the synthetic GPR data for the three different central frequencies equal to 50, 100 and 200 MHz, respectively.



**Fig. 3** Sketch of the acquisition geometry used for all profiles and each frequency. The 161 Receivers have fixed locations, while the 161 source locations (shot points) are moved by steps of 0.1 m from position 1 to position 161.

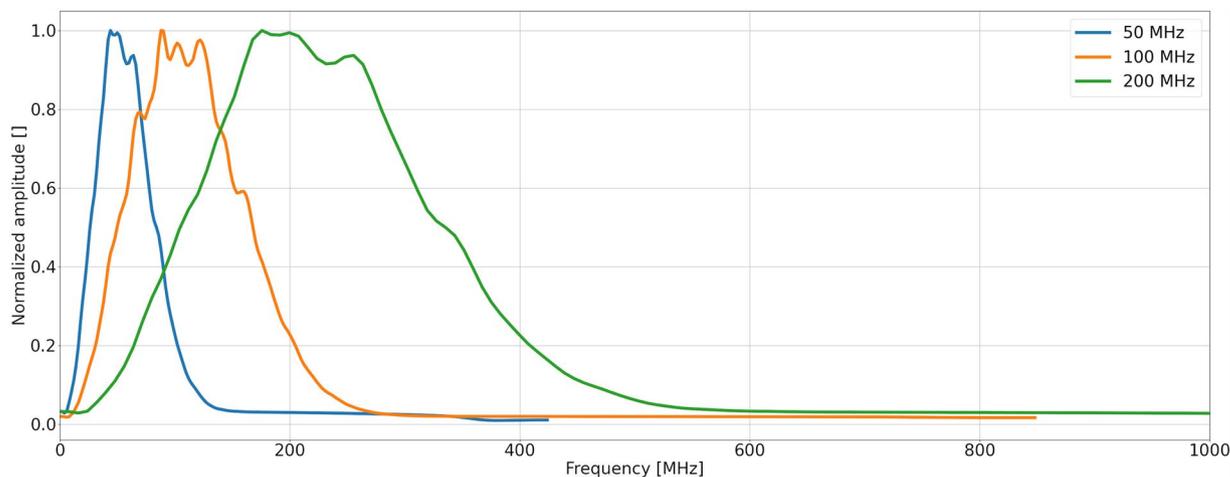


**Fig. 4** Comparison of three CSG simulated at  $x_{src} = 6$  m with 50 MHz (A), 100 MHz (B), and 200 MHz (C) central frequencies.

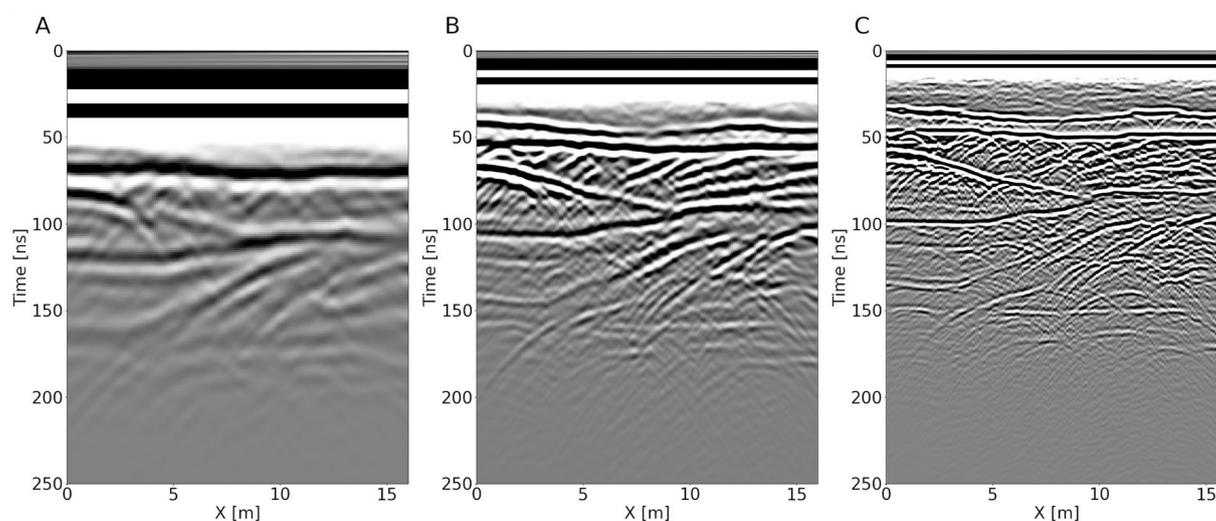
of GPR signals, making it valuable for investigating signal processing approaches and enhancing interpretation skills<sup>54,55</sup>. One of the notable strengths of *gprMax* is its capability to simulate real-world GPR scenarios, providing users with insights into expected outcomes during surveys and aiding in the enhancement of signal processing and interpretation capabilities<sup>56</sup>. Additionally, *gprMax* is fully parallelized, enabling it to leverage multiple CPUs and GPUs for efficient and high-performance simulations<sup>57</sup>. The simulation is performed using the parameters reported in Table 1.

The acquisition geometry is depicted in Fig. 3. For each of the four sections selected within the model, we simulate 161 shots (i.e., one source every 0.1 m from 0 m to 16 m) and, in each case, record the EM field using 161 receivers (i.e., one receiver every 0.1 m from 0 m to 16 m). A detailed sketch of the model used for the forward modeling including the air layer and PML region can be found in Figure Supplementary 2.

We generate 161 CSG for each of the four sections and for each of the three frequencies. We therefore provide a data set of 12 MO and multi-frequency GPR profiles: we simulate four 2 m separated GPR profiles for



**Fig. 5** Mean amplitude spectra among each frequency data set, each normalized to 1. (A) 50 MHz, (B) 100 MHz, (C) 200 MHz. The stability of the simulation is evident due to the absence of both high and low spurious frequency components.



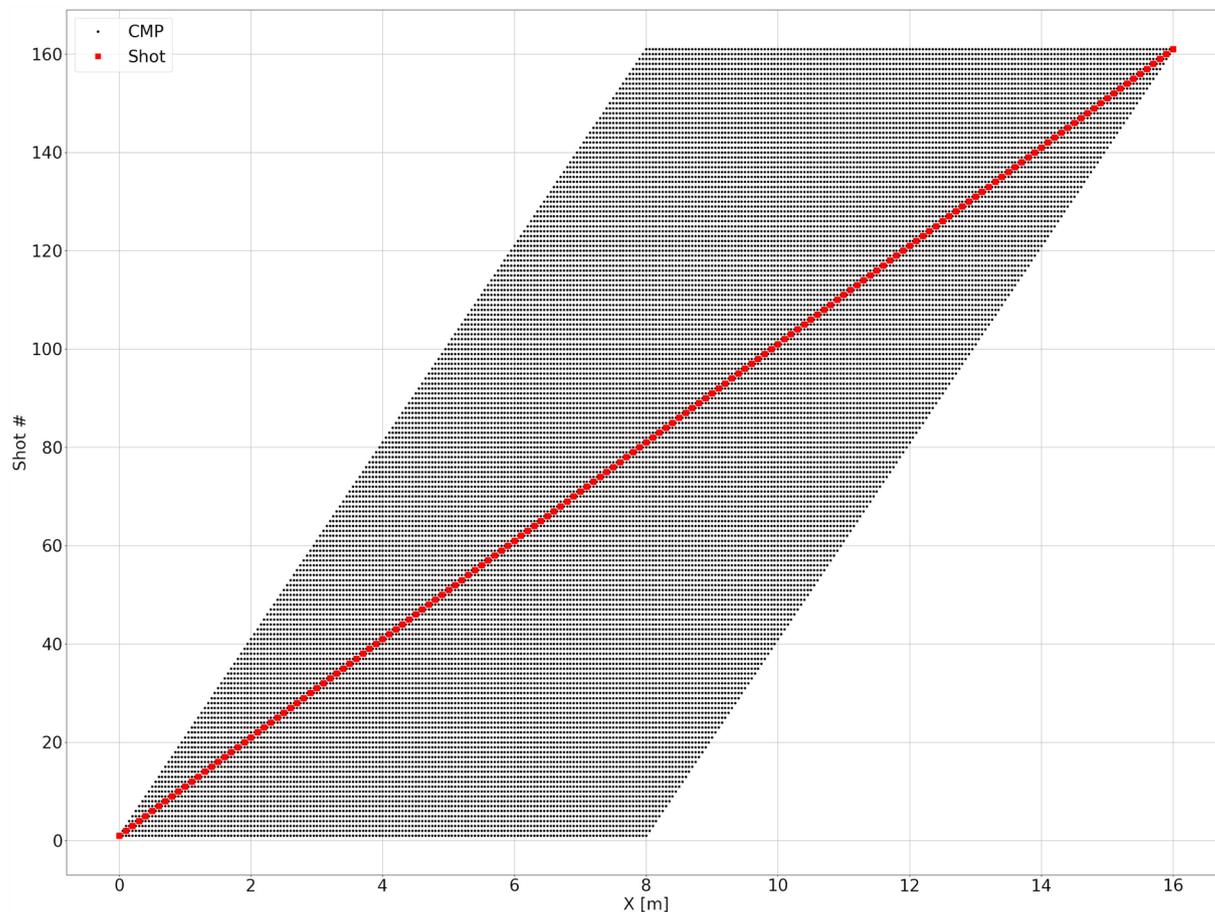
**Fig. 6** Example zero-offset profile with central frequencies of 50 MHz (A), 100 MHz (B), and 200 MHz (C). Due to the different shapes of the used source wavelets (see Figure Supplementary 3) the first arrivals (air waves) have slightly different time shifts.

Header	Start byte	End byte
Trace number	5	9
Trace ID	9	13
Recorder ID	13	17
Shot ID	17	21
Source X [cm]	73	77
Receiver X [cm]	81	85
CMP [cm]	181	185

**Table 2.** Location [byte] of the header information stored in the SEG-Y files.

50, 100 and 200 MHz central frequency Ricker wavelets. These wavelets and their spectra are shown in Figure Supplementary 3. An example CSG for each frequency is shown in Fig. 4. A more detailed example of four CSG for each frequency is shown in Figures Supplementary 4–6, for the 50, 100 and 200 MHz, respectively.

In order to demonstrate the differences between each profile within the model, the CSG at  $x_{src} = 6$  m is shown for each of the four sections in Figures Supplementary 7–9, for the 50, 100 and 200 MHz, respectively.



**Fig. 7** Stacking chart of the acquisition geometry. CMP are plotted as black dots, sources as red dots.

Comparing the CSG at different frequencies (Fig. 4), we observe a substantial increase in resolution with increasing frequency, making the data set an ideal test case for frequency-based inversion, such as frequency FWI, frequency merging<sup>58</sup>, or data deconvolution.

### Data Records

The dataset supporting this study is available on Figshare<sup>59,60</sup>. It includes synthetic multi-offset and multi-frequency ground-penetrating radar (GPR) data, stored in SEG-Y format, alongside the scripts and resources necessary for processing and validation.

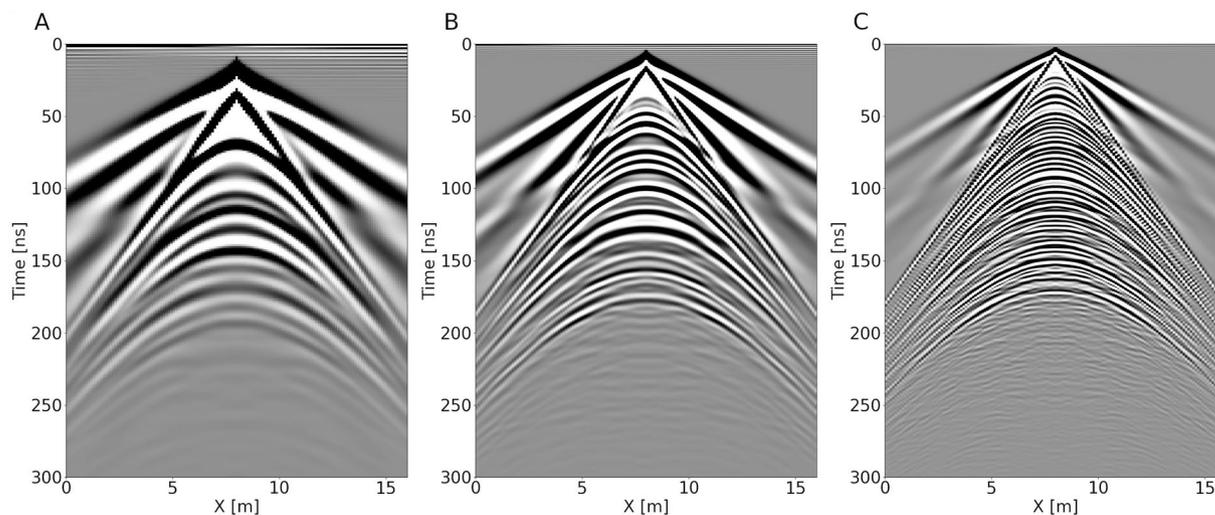
The GPR data is organized into subdirectories by frequency (50 MHz, 100 MHz, and 200 MHz), with each directory containing four profiles corresponding to the model sections at  $y = 2$  m,  $y = 4$  m,  $y = 6$  m, and  $y = 8$  m, *line0* to *line3*, respectively. Further details about the organization of the files and the structure of the dataset can be found on the data repository<sup>59,60</sup>. Each SEG-Y file is accompanied by header information specifying acquisition parameters, as outlined in the Usage Notes section.

Jupyter notebooks provided in<sup>59,60</sup> demonstrate how to read SEG-Y files, visualize the profiles, and verify the acquisition geometry using python. The required computational environment is defined in conda environment file: *env.yml*.

### Technical Validation

**Data stability.** In order to illustrate the effectiveness of the forward modelling, we show the mean frequency spectrum for each frequency (Fig. 5). It is apparent that the spectra are centered around the expected central frequencies, with no spurious effects at either low or high frequencies edges (as they would be expected in case of data instability due to wrong model parametrization).

Furthermore, we compare the zero-offset GPR profiles (i.e.,  $x_{src} = x_{rec}$ ) for each frequency for the section at  $y = 2$  m (Fig. 6). We observe near-field effects, which are larger for lower frequencies, as expected. These near-field effects occur within the close range of the GPR antenna, where emitted electromagnetic wavefronts have not yet fully developed into far-field radiation patterns. For lower frequencies, such as 50 MHz, the near-field zone is wider, reaching a time length of about 70 ns (Fig. 6A). This extended near-field zone is due to the longer wavelengths associated with lower frequencies, which increase the distance over which the electromagnetic field transitions from the near-field to the far-field. Consequently, this makes not possible to extract the signal from the noise in the shallower portion of the data.



**Fig. 8** CMP gathers for the central transmitter location ( $x = 8$  m) along the  $y = 8$  m section for central frequencies of 50 MHz (A), 100 MHz (B), and 200 MHz (C).

In contrast, higher frequencies, such as 200 MHz, exhibit a much shorter near-field region, not exceeding 20 ns (Fig. 6C). The reduced near-field zone at higher frequencies allows improved separation between the direct wave and the shallowest reflections.

The increasing times of the reflectors with decreasing antenna frequencies in Fig. 6 can be attributed to the time lag of the Ricker wavelet used in the forward modelling, which decreases as the frequency increases. Specifically, the 50 MHz wavelet has a longer time lag compared to the 100 MHz and 200 MHz wavelets, as shown in Figure S9. This results in progressively shallower apparent depths for higher-frequency antennas. A zero-time correction is typically applied to real GPR data in the first steps of the processing flow<sup>1</sup> and can also be applied to this dataset.

**Data compatibility and geometry validation.** SEG-Y data provided in this paper (see Table 2 for headers) are tested to properly work on different commercial and open-source programs originally developed for both reflection seismic: ProMAX (Halliburton), Petrel 17 (Schlumberger), Seissee 2.22 (Dalmorneftegeofizika Geophysical Company), and GPR: Prism 2.70.04 (Radar Systems), ReflexW 9.5.7 (Sandmaier geophysical research).

In order to validate the data geometry, we analyze the stacking chart for the entire simulated data set. Figure 7, shows the shot numbers vs their  $x$  coordinate. The maximum folding (i.e., 161) is correctly reached for the CMP at 8 m.

In order to further validate the acquisition geometry, we compare three CMP at  $x = 8$  m, that is, the full-folding case (Fig. 8).

### Usage Notes

An example python code to read the header file can be found at<sup>59</sup> and in [https://github.com/Giacomo-Roncoroni/MO-GPR\\_data](https://github.com/Giacomo-Roncoroni/MO-GPR_data).

### Code availability

Codes for data validation and exemplary gprMax input files are reported in<sup>59</sup> and in [https://github.com/Giacomo-Roncoroni/MO-GPR\\_data](https://github.com/Giacomo-Roncoroni/MO-GPR_data).

Received: 12 August 2024; Accepted: 13 December 2024;

Published online: 06 February 2025

### References

- Jol, H. M. *Ground Penetrating Radar: Theory and Applications*. Elsevier, Amsterdam, 509 pp (2009).
- Berard, B. & Maillol, J.-M. Common- and multi-offset ground-penetrating radar study of a roman villa, Tourega, Portugal. *Archaeological Prospection* **15**(1), 32–46, <https://doi.org/10.1002/arp.319> (2007).
- Berard, B. A. & Maillol, J.-M. Multi-offset ground penetrating radar data for improved imaging in areas of lateral complexity: Application at a Native American site. *Journal of Applied Geophysics* **62**, 167–177, <https://doi.org/10.1016/j.jappgeo.2006.10.002> (2007).
- Forte, E. & Pipan, M. Review of multi offset GPR applications: data acquisition, processing and analysis. *Signal Processing* **132C**, 210–220, <https://doi.org/10.1016/j.sigpro.2016.04.011> (2017).
- Angelis, D., Warren, C., Diamanti, N., Martin, J. & Annan, A. P. Challenges and opportunities from large volume, multi-offset Ground Penetrating Radar data. <https://doi.org/10.5194/egusphere-egu21-13138> (2021).
- Bradford, J. H. Reverse-time prestack depth migration of GPR data from topography for amplitude reconstruction in complex environments. *J. Earth Sci.* **26**, 791–798, <https://doi.org/10.1007/s12583-015-0596-x> (2015).
- Zeng, X., McMechan, G. A. & Xu, T. Synthesis of amplitude versus offset variations in ground-penetrating radar data. *Geophysics* **65**, 113–125, <https://doi.org/10.1190/1.1444702> (2000).

8. Becht, A., Appel, E. & Dietrich, P. Analysis of multi-offset gpr data: a case study in a coarse-grained gravel aquifer. *Near Surface Geophysics* **4**(4), 227–240, <https://doi.org/10.3997/1873-0604.2005047> (2005).
9. Mangel, A., Moysey, S., Ryan, J. & Tarbuton, J. Multi-offset ground-penetrating radar imaging of a lab-scale infiltration test. *Hydrol. Earth Syst. Sci.* **16**, 4009–4022 (2011).
10. Klotzsche, A., Jonard, F., Looms, M., van der Kruk, J. & Huisman, J. Measuring soil water content with ground penetrating radar: a decade of progress. *Vadose Zone Journal* **17**(1), 1–9, <https://doi.org/10.2136/vzj2018.03.0052> (2018).
11. Lärm, L. Linking horizontal crosshole gpr variability with root image information for maize crops. *Vadose Zone Journal*, **23**(1), <https://doi.org/10.1002/vzj2.20293> (2023).
12. Mangel, A., Moysey, S. & Bradford, J. H. Reflection tomography of time-lapse gpr data for studying dynamic unsaturated flow phenomena. *Hydrology and Earth System Sciences* **24**(1), 159–167, <https://doi.org/10.5194/hess-24-159-2020> (2020).
13. Léger, E., Sainteny, A., & Coquet, Y. Estimating saturated hydraulic conductivity from ground-based gpr monitoring Porchet infiltration in sandy soil. *Proceedings of the 15th International Conference on Ground Penetrating Radar*. <https://doi.org/10.1109/icgpr.2014.6970399> (2014).
14. Deeds, J. & Bradford, J. H. (2002). Characterization of an aquitard and direct detection of lnapl at hill air force base using GPR AVO and migration velocity analyses. *Proceedings Volume 4758, Ninth International Conference on Ground Penetrating Radar*, <https://doi.org/10.1117/12.462236> (2002).
15. Pipan, M., Baradello, L., Forte, E., Prizzon, A. & Finetti, I. 2-D and 3-D processing and interpretation of multi-fold ground penetrating radar data: a case history from an archaeological site. *Journal of Applied Geophysics* **41**(2–3), 271–292, [https://doi.org/10.1016/S0926-9851\(98\)00047-0](https://doi.org/10.1016/S0926-9851(98)00047-0) (1999).
16. Sun, J., Inananen, K., Zhang, T., & Trad, D. Implicit seismic full waveform inversion with deep neural representation. *Journal of Geophysical Research Solid Earth*, **128**(3), <https://doi.org/10.1029/2022jb025964> (2023).
17. McCann, M., Jin, K. & Unser, M. Convolutional neural networks for inverse problems in imaging: a review. *Ieee Signal Processing Magazine* **34**(6), 85–95, <https://doi.org/10.1109/msp.2017.2739299> (2017).
18. Stadler, S. & Igel, J. Developing realistic ftdt gpr antenna surrogates by means of particle swarm optimization. *Ieee Transactions on Antennas and Propagation* **70**(6), 4259–4272, <https://doi.org/10.1109/tap.2022.3142335> (2022).
19. Van der Kruk, J. *et al.* Quantitative multi-layer electromagnetic induction inversion and full-waveform inversion of crosshole ground penetrating radar data. *Journal of Earth Science* **26**(6), 844–850, <https://doi.org/10.1007/s12583-015-0610-3> (2015).
20. Oberröhrmann, M., Klotzsche, A., Vereecken, H. & van der Kruk, J. Optimization of acquisition setup for cross-hole: gpr full-waveform inversion using checkerboard analysis. *Near Surface Geophysics* **11**(2), 197–209, <https://doi.org/10.3997/1873-0604.2012045> (2012).
21. Liu, T. *et al.* Radius estimation of subsurface cylindrical objects from ground-penetrating-radar data using full-waveform inversion. *Geophysics* **83**(6), H43–H54, <https://doi.org/10.1190/geo2017-0815.1> (2018).
22. Kalogeropoulos, A., van der Kruk, J., Hugenschmidt, J., Büsch, S. & Merz, K. Chlorides and moisture assessment in concrete by GPR full waveform inversion. *Near Surface Geophysics* **9**(3), 277–286, <https://doi.org/10.3997/1873-0604.2010064> (2010).
23. Minet, J., Wahyudi, A., Bogaert, P., Vanclooster, M. & Lambot, S. Mapping shallow soil moisture profiles at the field scale using full-waveform inversion of ground penetrating radar data. *Geoderma* **161**(3–4), 225–237, <https://doi.org/10.1016/j.geoderma.2010.12.023> (2011).
24. Qinn, T., Bohlen, T. & Pan, Y. Indirect joint petrophysical inversion of synthetic shallow-seismic and multi-offset ground-penetrating radar data. *Geophysical Journal International* **229**(3), 1770–1784, <https://doi.org/10.1093/gji/ggac021> (2022).
25. Qinn, T., Bohlen, T. & Allroggen, N. Full-waveform inversion of ground-penetrating radar data in frequency-dependent media involving permittivity attenuation. *Geophysical Journal International* **232**(1), 504–522, <https://doi.org/10.1093/gji/ggac319> (2023).
26. Yan, P., Kalscheuer, T., Hedin, P. & Juanatey, M. Two-dimensional magnetotelluric inversion using reflection seismic data as constraints and application in the cosc project. *Geophysical Research Letters* **44**(8), 3554–3563, <https://doi.org/10.1002/2017gl072953> (2017).
27. Laloy, E. *et al.* Gradient-based deterministic inversion of geophysical data with generative adversarial networks: is it feasible? *Computers & Geosciences* **133**, 104333, <https://doi.org/10.1016/j.cageo.2019.104333> (2019).
28. Grasmueck, M., Weger, R. & Horstmeyer, H. Full-resolution 3D GPR imaging. *Geophysics* **70**, K12–K19 (2005).
29. Francese, R. G., Finzi, E. & Morelli, G. 3-D high-resolution multi-channel radar investigation of a Roman village in Northern Italy. *Journal of Applied Geophysics* **67**, 44–51 (2009).
30. Trinks, I. *et al.* Large-area high-resolution ground-penetrating radar measurements for archaeological prospection. *Archaeological Prospection* **25**, 171–195, <https://doi.org/10.1002/arp.1599> (2018).
31. Forte, E. *et al.* Optimised extraction of archaeological features from full 3-D GPR data. *Applied Sciences* **11**, 8517, <https://doi.org/10.3390/app11188517> (2021).
32. Muller, W. B. Semi-automatic determination of layer depth, permittivity and moisture content for unbound granular pavements using multi-offset 3-D GPR. *International Journal of Pavement Engineering* **21**(10), 1281–1296, <https://doi.org/10.1080/10298436.2018.1539485> (2020).
33. Fang, L., Yang, F., Xu, M. & Liu, F. Research on Development 3D Ground Penetrating Radar Acquisition and Control Technology for Road Underground Diseases with Dual-Band Antenna Arrays. *Sensors (Basel)* **23**(19), 8301, <https://doi.org/10.3390/s23198301> (2023).
34. Wollschläger, U., Gerhards, H., Yu, Q. & Roth, K. Multichannel ground-penetrating radar to explore spatial variations in thaw depth and moisture content in the active layer of a permafrost site. *The Cryosphere* **4**, 269–283, <https://doi.org/10.5194/tc-4-269-2010> (2010).
35. Bradford, J. H. Measuring water content heterogeneity using multifold GPR with reflection tomography. *Vadose Zone Journal* **7**, 184–193, <https://doi.org/10.2136/vzj2006.0160> (2008).
36. Bradford, J. H., Clement, W. P., & Barrash, W. Estimating porosity with ground-penetrating radar reflection tomography: A controlled 3-D experiment at the Boise Hydrogeophysical Research Site. *Water Resources Research*, **45**(4), <https://doi.org/10.1029/2008WR006960> (2009).
37. Bradford, J. H., Nichols, J., Mikesell, T. D. & Harper, J. T. Continuous profiles of electromagnetic wave velocity and water content in glaciers: An example from Bench Glacier, Alaska, USA. *Annals of Glaciology* **50**, 1–9, <https://doi.org/10.3189/172756409789097540> (2017).
38. Fisher, E., McMechan, G. A. & Annan, A. P. Acquisition and processing of wide-aperture ground-penetrating radar data. *Geophysics* **57**, 495–504, <https://doi.org/10.1190/1.1443265> (1992).
39. Annan, A. P., & Jackson, S. R. The WARR machine. In A. Giannopoulos & C. Warren (Eds.), 9th International Workshop on Advanced Ground Penetrating Radar (IWAGPR). Piscataway, NJ: Institute of Electrical and Electronic Engineers, <https://doi.org/10.1109/IWAGPR.2017.7996106> (2017).
40. Diamanti, N., Elliott, E. J., Jackson, S. R. & Annan, A. P. The WARR machine: System design, implementation and data. *Journal of Environmental and Engineering Geophysics* **23**, 469–487 (2018).
41. Kaufmann, M. S., Klotzsche, A., Vereecken, H. & van der Kruk, J. Simultaneous multichannel multi-offset ground-penetrating radar measurements for soil characterization. *Vadose Zone Journal* **19**(1), e20017, <https://doi.org/10.1002/vzj2.20017> (2020).
42. Angelis, D., Warren, C., Diamanti, N., Martin, J. & Annan, A. P. The effects of receiver arrangement on velocity analysis with multi-concurrent receiver GPR data. *Near Surface Geophysics* **20**(5), 519–530 (2022). ISSN 1569-4445.

43. Versteeg, R. J. & Grau, G. (eds.) The Marmousi experience. Proc. EAGE workshop on Practical Aspects of Seismic Data Inversion (Copenhagen, 1990), Eur. Assoc. Explor. Geophysicists, Zeist (1991).
44. Versteeg, R. J. The Marmousi experience: Velocity model determination on a synthetic complex data set. *The Leading Edge* **13**, 927–936, <https://doi.org/10.1190/1.1437051> (1994).
45. Martin, G. S., Wiley, R. & Marfurt, J. Marmousi2: An elastic upgrade for Marmousi. *The Leading Edge* **25**(2), 156–166 (2006).
46. Pan, G., Liang, L., & Habashy, T. Three-dimensional frequency-domain elastic full-waveform inversion of Marmousi model. *SEG Technical Program Expanded Abstracts*, 1125–1130, <https://doi.org/10.1190/segam2016-13958342.1> (2016).
47. Koyan, P. & Tronicke, J. 3D modeling of ground-penetrating radar data across a realistic sedimentary model. *Computers & Geosciences* **137**, 104422, <https://doi.org/10.1016/j.cageo.2020.104422> (2020).
48. Koyan, P., & Tronicke, J. A synthetic 3D ground-penetrating radar (GPR) data set across a realistic sedimentary model. *Mendeley Data*, **V1**, <https://doi.org/10.17632/by3yh79hx4.1> (2019).
49. Bayer, P., Huggenberger, P., Renard, P. & Comunian, A. Three-dimensional high resolution fluvio-glacial aquifer analog: Part 1: Field study. *Journal of Hydrology* **405**(1–2), 1–9 (2011).
50. Comunian, A., Renard, P., Straubhaar, J. & Bayer, P. Three-dimensional high resolution fluvio-glacial aquifer analog - Part 2: Geostatistical modeling. *Journal of Hydrology* **405**(1–2), 10–23, <https://doi.org/10.1016/j.jhydrol.2011.03.037> (2011).
51. Warren, C., Giannopoulos, A. & Giannakis, I. gprMax: Open source software to simulate electromagnetic wave propagation for Ground Penetrating Radar. *Comput. Phys. Comm.* **209**, 163–170, <https://doi.org/10.1016/j.cpc.2016.08.020> (2016).
52. Warren, C. *et al.* A CUDA-based GPU engine for gprMax: Open source FDTD electromagnetic simulation software. *Comput. Phys. Comm.* **237**(2018), 208–218, <https://doi.org/10.1016/j.cpc.2018.11.007> (2018).
53. Zhang, D., Tang, R., Wen, H. & Fan, S. Propagation of ground penetrating radar waves in Chinese coals. *Plos One* **15**(5), e0233434, <https://doi.org/10.1371/journal.pone.0233434> (2020).
54. Yang, J. & Duan, Y. Identification of unstable subsurface rock structure using ground penetrating radar: an eemd-based processing method. *Applied Sciences* **10**(23), 8499, <https://doi.org/10.3390/app10238499> (2020).
55. Warren, C. & Giannopoulos, A. Influence of lossless and lossy, heterogeneous environments on ground penetrating radar antenna behavior. IEEE 15th Mediterranean Microwave Symposium (MMS), Lecce, Italy, 2015, pp. 1–5, <https://doi.org/10.1109/MMS.2015.7375397> (2015).
56. Giannakis, I., Giannopoulos, A. & Warren, C. Special Issue: Ground Penetrating Radar (GPR) Numerical Modelling Research and Practice. *Near Surface Geophysics* **22**(2), 1569–4445, <https://doi.org/10.1002/nsg.12301> (2024).
57. Warren, C., Giannakis, I., & Giannopoulos, A. GprMax: an open source electromagnetic simulator for generating big data for ground penetrating radar applications. *EGU General Assembly 2021*, <https://doi.org/10.5194/egusphere-egu21-10347> (2021).
58. Roncoroni, G., Forte, E., Santin, I. & Pipan, M. Deep learning-based multifrequency ground penetrating radar data merging. *Geophysics* **89**, F1–F9, <https://doi.org/10.1190/geo2023-0215.1> (2024).
59. Roncoroni, G. Multi-Offset Synthetic GPR Data. *figshare. Dataset*. <https://doi.org/10.6084/m9.figshare.26508976.v1> (2024).
60. Philipp, K. Giacomo, R. Jens, T. A synthetic 3D ground-penetrating radar (GPR) dataset across a realistic sedimentary model, V2. *Mendeley Data* <https://doi.org/10.17632/by3yh79hx4.2> (2025).

## Acknowledgements

This work has been partially funded by the Deutsche Forschungsgemeinschaft (DFG), Germany, under project no. 374920008. G. R. acknowledges the IG of the PNRR spoke 2, National Centre for HPC, Big Data and Quantum Computing (J93C22000540006).

## Author contributions

Roncoroni G.: Conceptualization, Code Writing, Technical Validation and Manuscript Draft. Koyan P.: Conceptualization, Technical Validation and Manuscript Draft. Forte E.: Technical Validation, Manuscript Draft and Review. J. Tronicke: Coordination and Final Manuscript Review. Pipan M.: Coordination and Final Manuscript Review.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-04300-1>.

**Correspondence** and requests for materials should be addressed to P.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025