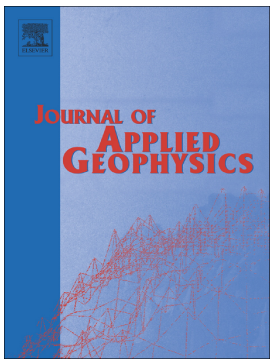
Toward the optimisation of the Kalman Filter approach in ground penetrating radar application for detection and locating buried utilities

Afrasiabi, Arasti; Faramarzi, Asaad; Chapman, David; Keshavarzi, Alireza; Stringfellow, Matthew

DOI:
10.1016/j.jappgeo.2023.105220



Toward the optimisation of the Kalman Filter approach in ground penetrating radar application for detection and locating buried utilities

Arasti Afrasiabi, Asaad Faramarzi, David Chapman, Alireza Keshavarzi, Mathew Stringfellow

PII: S0926-9851(23)00298-7

DOI: https://doi.org/10.1016/j.jappgeo.2023.105220

Reference: APPGEO 105220

To appear in: *Journal of Applied Geophysics*

Received date: 23 April 2022

Revised date: 3 September 2023

Accepted date: 26 October 2023

Please cite this article as: A. Afrasiabi, A. Faramarzi, D. Chapman, et al., Toward the optimisation of the Kalman Filter approach in ground penetrating radar application for detection and locating buried utilities, *Journal of Applied Geophysics* (2023), https://doi.org/10.1016/j.jappgeo.2023.105220

Toward the optimisation of the Kalman Filter approach in ground penetrating radar application for detection and locating buried utilities

Arasti Afrasiabi, Asaad Faramarzi, David Chapman, Alireza Keshavarzi, Mathew Stringfellow

Dr Arasti Afrasiabi

School of Engineering, Department of Civil Engineering

University of Birmingham, UK

Dr Asaad Faramarzi

School of Engineering, Department of Civil Engineering

University of Birmingham, UK

Dr David Chapman

School of Engineering, Department of Civil Engineering

University of Birmingham, UK

Mr. Alireza Keshavarzi

Independent Researcher

Yazd, Iran

Mr. Mathew Stringfellow

RSK Group, UK

Abstract

Ground Penetrating Radar (GPR) stands as a pivotal non-destructive tool for identifying and assessing buried utilities. However, the noisy GPR radargram data requires extensive expert interpretation, making it time-consuming and subjective. While various signal processing techniques, like time-zero correction and background subtraction, help mitigate noise, mathematical estimators like the Kalman Filter (KF) and particle filter present advanced solutions. Notably, KF, known for its efficiency and computational benefits, excels in denoising and decluttering GPR radargrams. This research introduces an innovative KF-based optimisation algorithm tailored to minimise user input and pinpoint buried utilities and anomalies. Leveraging the distinctive KF parameter, Normalised Innovation Squared (NIS), the algorithm aims for enhanced target detection. A genetic algorithm-driven multi-objective optimisation model evaluates the method's efficacy, focusing on the receiver operating characteristic (ROC) and mean of innovations. Cost functions considered encompass noise covariances and optimal Fourier analysis frequency. Preliminary results showed a 23.13% rise in the area under the ROC with optimised parameters (91.4%) compared to user-selected ones (68.27%). This method not only reduces GPR data noise but also augments the detection of buried utilities. Traditional chi-squared hypothesis testing was replaced with an NIS signal function analysis, facilitating more refined anomaly detection. Collectively, this study offers a transformative approach to GPR data post-processing, emphasising efficiency and reduced user dependency.

Keywords

Ground Penetrating Radar (GPR), Kalman Filter, Normalised Innovation Squared (NIS), Multi-objective optimisation, Receiver Operating Characteristic (ROC).

1 Introduction

1.1 Background and Motivation

Ground-penetrating radar (GPR) has been extensively used to assess buried utilities' location and condition in shallow underground spaces. Although GPR is a powerful tool for detecting subsurface objects, it faces challenges in terms of signal alterations, attenuation, noise, and clutter. These factors can create difficulties in interpreting GPR data and detecting target signals. Various techniques have been proposed to address these issues, including the Kalman Filter (KF), a well-established estimation method. However, traditional KF methods have limitations in terms of adaptability and optimisation, which can be further improved.

1.2 Literature Review and Related Work

Numerous methods have been proposed for GPR data processing and background removal. Recent approaches include an adaptive denoising method based on fast algorithm for independent component analysis (FastICA) and wavelet transform modulus maxima (WTMM) multifractal spectrum (Li et al., 2022), a deep learning approach to estimate the bulk relative permittivity of compound materials and filter out direct-coupling and surface reflections (Wickramanayake et al., 2022), and the conventional direct current (DC) Subtraction method (Cheng et al., 2023; Zhao et al., 2022). Earlier algorithms for background signal removal encompass particle filtering (Ng et al., 2008), spectral domain features (Ho et al., 2004), maximum likelihood (Ho and Gader, 2002; Smitha and Singh, 2016), eigenvalue-based methods (Khan and Al-Nuaimy, 2010), median filtering (Kim et al., 2007), parametric models (Van Merwe and Gupta, 2000), adaptive ground bounce (Wu et al., 2001), and singular value decomposition (Liu et al., 2017).

One of the well-established techniques for estimation problems, such as those associated with radar and navigation systems, is the Kalman Filter (Simon, 2006). The KF has been used to a limited extent for estimating the background signal of GPR radargrams. Carevic (1999) first proposed the KF-based algorithm for this purpose, which several researchers later extended and implemented (Luo and Fang, 2005; Ng et al., 2008; Zoubir et al., 2002a). Carevic (1999), conducted a stepwise analysis of backscattered signals to estimate the internal condition of the system at specific moments, referred to as the system state (Simon, 2006). Carevic's algorithm is based on calculating a parameter known as 'innovation,' which indicates the residual difference between the actual and estimated signal values. Target signal presence was determined using a chi-squared (χ^2) hypothesis test (Bar-Shalom et al., 2001). This statistical test measures how well a model corresponds to observed data (McHugh, 2012). However, the chi-squared method has limitations, including the lack of a definitive method for setting an optimal threshold value for hypothesis rejection and the unknown distribution of test statistics in GPR applications (Gurbuz, 2012; Ng et al., 2008). Moreover, Carevic's algorithm could not determine the width of the buried target, limiting its detection capacity to a single target per trace (Ng et al., 2008; Zoubir et al., 2002b). Subsequent research used Receiver Operating Characteristic (ROC) curves, defined as the area under the curve plotting false versus true detection rates, for validation of Carevic algorithm and achieved a target recognition rate over 90% (Kerekes, 2008; Luo and Fang, 2005; Zoubir et al., 2002a).

The recent studies on GPR showcases a variety of applications and methodologies, frequently utilising KF for positioning and data optimisation. Linna et al. (2020) focus on agricultural applications and suggest that deep learning techniques could be employed for data interpretation, although they provide no empirical validation. Zou et al. (2020) use a KF in conjunction with a low-cost GPS for 3D tree root mapping and offer a successful case study, but lack a discussion on limitations. Kaniewski and Kraszewski (2020) propose an ultrawideband radio-based positioning system for demining, introducing an extended KF with a novel pendulum motion-based dynamics model. While simulations indicate better accuracy, they do not provide real-world testing results. Patel and Ferguson (2021) employ GPR in Arctic conditions, using a Velodyne LiDAR and an extended KF to stabilise a drone-carrying payload, yet fail to discuss the method's limitations. Hou et al. (2022) introduce an automatic tunnel disease detection method that uses GPR and a KF alongside support vector machines (SVM), but neglect to address false positive or negative risks. Kaniewski and Kraszewski (2023) further refine their positioning algorithm, focusing on the KF's dynamics model but still lack empirical validation. Lastly, Patel (2023) discusses the challenge of stabilising GPR payloads in the Arctic and evaluates various controllers for performance, stating that none provided optimal control. Overall, these studies demonstrate the adaptability and promise of GPR technologies across multiple domains but often fall short in real-world testing and limitations discourse.

Despite the proven advantages of KF in GPR application, significant challenges remain. A key issue is the absence of a robust algorithm to accurately identify target width. Additionally, the manual determination of KF input values, such as noise covariance matrices, is time-consuming and prone to human error. Finally, existing studies have not explored the KF's potential for assessing the conditions of the ground that supports buried anomalies.

1.3 Research Objectives and Contributions

This paper presents an innovative approach that combines the KF with genetic algorithms to improve GPR data processing and buried target detection. The main contributions of our proposed method are:

- Addressing the limitations of previous KF-based approaches, particularly the inability to accurately identify the width of buried targets and the need for manual determination of some input values (e.g., noise covariance matrices).

- Providing a more accurate and efficient method for identifying the width of buried targets and estimating the supporting ground conditions.

- Minimising the need for manual input, reducing the risk of human error, and streamlining the GPR data processing workflow.

- Incorporating recent advancements in the field and addressing the current challenges, contributing to the ongoing development of GPR technology and its applications in various industries, such as infrastructure inspection, remote sensing, and environmental monitoring.

1.4 Paper Organisation

The remainder of this paper is organised as follows: Section 2 presents the proposed semi-autonomous algorithm that combines the KF with genetic algorithms, including its detailed description, mathematical formulation, and implementation steps. Section 3 presents the results and performance evaluation of the proposed algorithm. Section 4 concludes the paper, summarising the main contributions and findings.

By presenting an innovative approach that combines KF with genetic algorithms to improve GPR data processing and buried target detection, this paper aims to contribute to the advancement of GPR technology and its applications in various industries.

2 Methodology

This section elaborates on the research methods employed in this study, from theoretical foundations to specific algorithms and their implementation. For a concise overview of the complete research process, readers may refer to subsection 2.4.1 "Overview of Research Process."

2.1 Overview of the Kalman Filter

KF is a recursive, iterative algorithm designed to estimate the state of a linear dynamic system in the presence of noise. It was first introduced by Rudolf E. Kalman in 1960 and has since become a widely used tool in various fields, including control systems, signal processing, navigation, and computer vision (Simon, 2006). The main advantage of KF is its ability to provide optimal estimates of unknown variables in real time while considering uncertainties in both the system model and the observations (Grewal and Andrews, 2008; Särkkä, 2013). In its most basic form, KF consists of two main steps: prediction and update. In the prediction step, the filter uses the system's state-space model to predict the state at the next time step, incorporating the system's dynamic properties and known control inputs. In the update step, the filter incorporates the new observations or measurements to update the predicted state, producing an optimal estimate of the current state based on all available information (Simon, 2006). KF relies on the knowledge of the system's state-transition model, observation model, and the associated noise covariance matrices (Bar-Shalom et al., 2001).

In the context of GPR background removal, KF can be used to estimate the underlying background signal, which can then be subtracted from the observed GPR data to reveal the desired target signals. The effectiveness of KF in this application depends on the accurate representation of the system's dynamics and the noise properties, as well as the careful tuning of its parameters (Carevic, 1999; Kaniewski and Kraszewski, 2020; Zoubir et al., 2002a).

2.2 Current Practice and the Proposed Data Processing of GPR Radargrams

Noise and clutter in GPR radargrams can suppress useful target signals and essential information about buried anomalies, making the radargram interpretation a challenging process (Milsom and Eriksen, 2011). Typically, the interpretation of GPR radargrams includes several pre-processing and filtering steps (Huber and Hans, 2018), such as: i) dewow filtering (removes signal saturation caused

by the gap between the device antenna and ground surface); ii) DC-drift correction (shifts the deviated signal from zero with increasing depth); iii) bandpass filtering (retains the responses within the antenna frequency range); iv) signal/amplitude gain (ensures the visibility of the target signal increases with depth); and v) averaging filter (removes horizontal bands in radargrams). Of these, the last two steps have been the subject of several research papers in connection with the development and implementation of effective algorithms (Huber and Hans, 2018). Additionally, numerous works have been devoted to overcoming the DC-drift issue in GPR signals (Goodman and Piro, 2013). The proposed methodology in the current paper offers a series of new solutions to resolve these challenges.

2.3 Genetic Algorithm and Its Role in Parameter Optimisation

Genetic Algorithms (GAs) are a family of heuristic search algorithms inspired by the process of natural selection and genetics (Mirjalili, 2019). They have been extensively used to solve optimisation problems, particularly when the solution space is large, complex, and poorly understood. GAs work by iteratively generating a population of candidate solutions and evaluating their fitness based on a predefined objective function (Mirjalili, 2019). The fittest individuals are then selected for reproduction using genetic operators, such as crossover and mutation, to produce the next generation of solutions (Kramer, 2017). This process is repeated until a termination criterion, such as reaching a maximum number of generations or achieving a specified fitness level, is met.

In the context of the proposed methodology, a Genetic Algorithm is employed to search for an optimal set of parameters for the KF. The standard KF requires the determination of several parameters, such as the initial state estimate, the state transition matrix, the measurement matrix, and the noise covariance matrices (Hartikainen and Särkkä, 2011). Some of these parameters can be difficult to determine accurately, especially in complex environments with non-stationary noise and dynamic target properties (Simon, 2006). By utilising a Genetic Algorithm, the performance of KF can potentially be improved in terms of noise reduction, target detection, and overall adaptability. The GA is used to optimise the noise covariance matrices, which directly affect the filter's ability to distinguish between the signal and noise (Sen and Mallick, 2018). Moreover, the GA aids in finding an optimal combination of averaging and gain filters to enhance the signal quality further (Katoch et al., 2021).

In summary, integrating a Genetic Algorithm with KF in the proposed methodology is theorised to offer a robust optimisation framework for elevating the processing of GPR radargrams. This approach could potentially lead to more accurate and efficient detection of buried targets and the estimation of supporting ground conditions, while also diminishing the necessity for manual intervention and curtailing the potential for human errors.

2.4 Proposed Algorithm and Framework

The proposed algorithm in this study consists of three sections, as illustrated in Figure 1.

The first section adapts the averaging and gain filters into KF algorithm to enhance signal quality. This integration is a novel approach to GPR data processing that combines the strengths of both

techniques to improve the visibility of target signals. The second section involves adapting and utilising a multi-optimisation algorithm using evolutionary computing, chosen over conventional optimisation strategies due to the problem's complexity and the ability to generate Pareto solutions (PSs) associated with cost functions (area under the ROC curves and the mean of innovations). This innovative approach leverages the power of genetic algorithms to optimise the performance of the KF, ensuring more accurate and efficient target detection. The third section develops sequential KF algorithms to investigate the radargram and distinguish noise, target, and anomaly signals, generating improved quality radargrams for visual confirmations. This new framework is designed to address the limitations of traditional KF methods and enhance their adaptability and performance in complex GPR data processing scenarios.

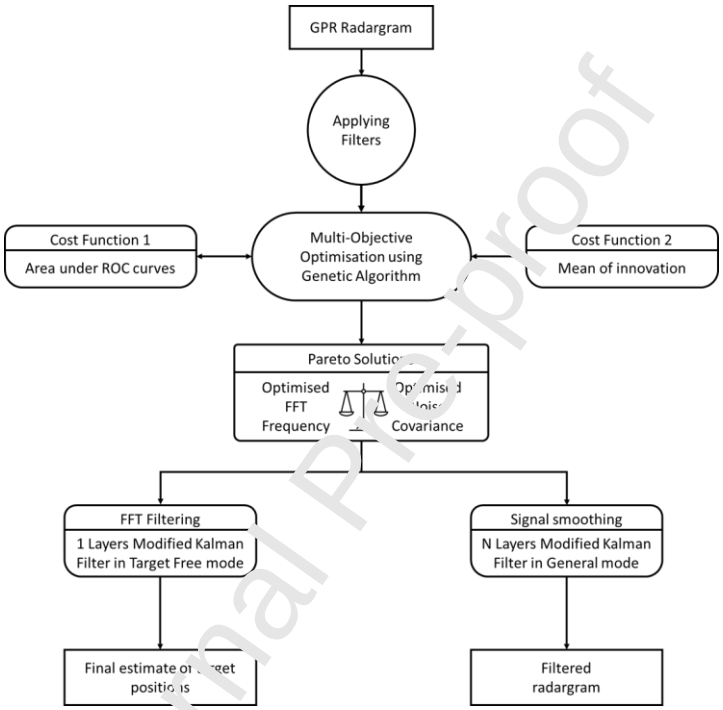


Figure 1. A flowchart of the proposed optimised Kalman Filter algorithm for detecting buried utilities and improved radargram quality

2.4.1 Overview of Research Process

To offer a comprehensive understanding of the research conducted, this subsection outlines the complete process, from the setup of the trial site to the data processing methods employed. The sequence of activities is as follows:

- **Trial Site Preparation**: An 11 m x 13 m trial site was constructed with 23 scan lines and five linear pipes at shallow depths (refer to Section 2.6.1 for details).
- **Data Collection**: GPR data were recorded in DZT GGSI format using a single-channel antenna with a centre frequency of 400MHz. The site's ground conditions were also thoroughly evaluated.
- **Initial Data Preprocessing**: Before applying the KF, the raw GPR data underwent averaging and gain filtering to enhance signal quality.

- **Parameter Optimisation**: Multi-objective Genetic Algorithm (MOGA) was employed for optimisation. Two cost functions were used to identify the optimal parameters for the KF and Fast Fourier Transform (FFT).
- **Kalman Filtering**: Based on the optimised parameters, the KF algorithm was applied to the pre-processed GPR data.
- **Target Detection**: The normalised innovation squared (NIS) signal from the KF was then used to detect targets, with results evaluated using ROC curves.
- **Evaluation and Validation**: Final processed radargrams were used for visual confirmation by experts, and the algorithm's performance was assessed using statistical metrics.

The framework mentioned above was developed by creating a modified version of KF algorithm. The design of the KF algorithm is a problem-specific task, and various types of KF algorithms have been proposed in the literature for specific problems (Simon, 2006). In this study, a discrete KF (Hartikainen and Särkkä, 2011; Simon, 2006) was developed to estimate the state-space model variables considering the physics of the system, such as the ground and the target, and the discontinuity of the GPR impulse signals. State-space models are mathematical models that describe the fundamental physics of a system by characterising its dynamics through several states (Van Drongelen, 2010). The state-space model for a GPR system in a KF application will be presented in the following paragraphs.

2.5 The Kalman Filter in GPR Applications

2.5.1 State-Space Model for GPR

This section presents the state-space model of GPR in the form of a series of mathematical equations (Eq. 1 to Eq. 10), followed by a brief description of the discrete KF. For further details on the process, readers are referred to Carevic (1999). The mathematical representation of the basic model for scattered GPR signal consists of the summation of the signal components (Eq. 1 and 2), namely the background signal $s^b\ (n,k)$, target signal $s^t\ (n,k)$, and the corresponding noise $\omega\ (n,k)$ (Carevic, 1999).

$$y\ (n,k) = s^t\ (n,k) + s^b\ (n,k) + \omega\ (n,k) \tag{1}$$

In vector format, this can be expressed as:

$$\begin{pmatrix} y(0,k) \\ y(1,k) \\ \vdots \\ y(N,k) \end{pmatrix} = \begin{pmatrix} s^t(0,k) \\ s^t(1,k) \\ \vdots \\ s^t(N,k) \end{pmatrix} + \begin{pmatrix} s^b(0,k) \\ s^b(1,k) \\ \vdots \\ s^b(N,k) \end{pmatrix} + \begin{pmatrix} \omega(0,k) \\ \omega(1,k) \\ \vdots \\ \omega(N,k) \end{pmatrix} \tag{2}$$

where indexes $n = \{1, \dots, N\}$ and $k = \{1, 2, 3, \dots\}$ refer to the sample points on the vertical axis and the distance from the origin (trace numbers) on the horizontal axis of a radargram, respectively, as illustrated in Figure 2.

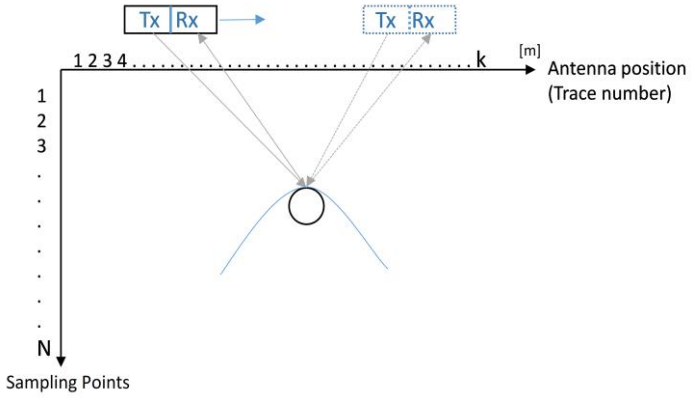


Figure 2. A typical GPR survey model (Tx refers to transmitter, Rx refers to receiver)

Taking into account the backscattered signal components s^t, s^l and ω, the KF algorithm can be utilised to distinguish and separate the noise from target signals under two general state assumptions (Carevic, 1999):

1. The quiescent (target-free) state, in which the return signal comprises only backscattered ground signals, known as background signals, as shown in Eq. 3 (Carevic, 1999):

$$y(k) = s^b(k) + \omega_1(k) \tag{3}$$

where:

$$s^b(k) = s^b(k-1) + \upsilon_1(k) \tag{4}$$

Here, υ_1 and ω_1 are the model and measured noises, respectively, assumed to be zero-mean Gaussian white noise with variances of $\sigma^2_{\upsilon_1}$ and $\sigma^2_{\omega_1}$.

2. The augmented (target-present) state, in which the scattered signal composition includes ground, target, and noise signals, as demonstrated in Eq. 6 (Carevic, 1999):

$$y(k) = s^b(k) + s^t(k) + \omega_2(k) \tag{6}$$

where:

$$s^b(k) = s^b(k-1) \tag{7}$$

$$s^t(k) = s^t(k-1) + b(k) \tag{8}$$

and the equation for added bias:

$$b(k) = b(k-1) + \upsilon_2(k) \tag{9}$$

Where υ_2 and ω_2 are the model and measured noises, respectively, assumed to be zero-mean Gaussian white noise with variances of $\sigma^2_{\upsilon_2}$ and $\sigma^2_{\omega_2}$.

2.5.2 Overview of the proposed discrete Kalman Filter algorithm

After determining the state-space model parameters based on the system's physics and taking into account the linearity of the equations as illustrated in the previous section, the system model variables can be estimated through the following equations (Eq. 10-14) in two steps (Carevic, 1999; Simon, 2006):

Step 1: Make an a priori estimate or prediction ($\hat{x}_k^-$), of the system state using measurement equations (Eq. 10-12) (Simon, 2006):

A priori estimate: $$\hat{x}_{k+1}^- = \mathbf{A}\hat{x}_{k-1}^+ + \mathbf{B}u \quad (10)$$

Error covariance matrix: $$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}^+\mathbf{A}^T + \mathbf{Q} \quad (11)$$

Kalman gain: $$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (12)$$

In these equations, u represents the system input. The variables **B**, **Q**, **P**, and A denote the 'input', 'system noise covariance', 'estimation-error covariance', and 'transition matrix of the model' matrices, respectively. The variables **R** and **H** represent 'the measurement noise covariance' and 'the measurement model matrix', respectively, in their matrix forms.

Step 2: Obtain a posteriori state estimate by updating the a priori state estimate (Simon, 2006):

Updating estimate $$\hat{x}_k^+ = \hat{x}_k^- + \mathbf{K}_k(y_k - \mathbf{H}_k\mathbf{A}\hat{x}_k^-) \quad (13)$$

Updating P: $$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k) + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \quad (14)$$

The input values $\mathbf{A}, \mathbf{H}, \mathbf{Q}$ and $\mathbf{R}$ specified for each state are presented in Table 1 (Carevic, 1999). The estimation-error covariance matrix provides information regarding the expected innovation, which serves as an indicator of the state estimate (x_k) accuracy. $\mathbf{P_0}$ is a diagonal matrix, and the initial value for the diagonal elements is set to 1E6 due to the high degree of uncertainty in the noise value (Simon, 2006).

Table 1. Input variables

Target-free	Target-present
$\mathbf{A} = \mathbf{I}_{N\times N}$	$\mathbf{A} = \begin{pmatrix} \mathbf{I} & 0 & 0 \\ 0 & \mathbf{I} & \mathbf{I} \\ 0 & 0 & \mathbf{I} \end{pmatrix}_{3N\times 3N}$
$\mathbf{H} = \mathbf{I}_{N\times N}$	$\mathbf{H} = (\mathbf{I} \quad \mathbf{I} \quad 0)_{N\times 3N}$

$$\mathbf{Q} = \sigma_{\upsilon_1}^2 \times \mathbf{I}_{N\times N} \quad \mathbf{Q} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_{\upsilon_2}^2 \mathbf{I} \end{pmatrix}_{3N\times 3N}$$

$$\mathbf{R} = \sigma_{\omega_1}^2 \times \mathbf{I}_{N\times N} \qquad \mathbf{R} = \sigma_{\omega_2}^2 \times \mathbf{I}_{N\times N}$$

To implement the proposed optional signal processing filters in this paper, Eq. 13 has been modified to develop a new parameter that incorporates signal gain, averaging filter, and signal drift removal into the KF. The following paragraphs provide further details.

2.5.3 Incorporation of signal pre-processing filters

The amplitude of GPR signals can be enhanced based on the quality of the initial signal by multiplying a time-varying gain function into the initial signal (Daniels, 2004). Commonly used gain functions, namely power and exponential gain functions (Eq. 15-16), have been selected and combined for this study (Eq. 17) (Huber and Hans, 2018).

$$G_{power}(n) = n^{\alpha} \tag{15}$$

$$G_{exponetial}(n) = e^{\alpha n} \tag{16}$$

$$G_{combi} = n^{\alpha} \times e^{\alpha n} \tag{17}$$

where, (G) represents the gain function, (n) is the sampling point, and (α) specifies the intensity of the signal amplification. To eliminate signal drift, a moving average subtraction function (Eq. 18-19) was multiplied by the gain function (Eq. 20) (Benedetto et al., 2013; Goodman and Piro, 2013):

$$f_{Sub} = y(n,k) - \bar{y}(n,k) \tag{18}$$

where:

$$\bar{y}(n,k) = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{19}$$

The parameter $\bar{y}(n,k)$ is the average value of all the traces of each radargram, which will be subtracted from the individual trace throughout N successive steps inside the algorithm.

$$y_G(n,k) = G_{combi} \times f_{sub} \tag{20}$$

Finally, to modify Eq.13, a new function $\xi(n,k)$ as expressed in Eq.21 was developed, which replaced the classical system input parameter (u_k) from Eq.13 with the new parameter (ξ_k) as expressed in Eq. 22 for when $\alpha = 1$:

$$\xi(n,k) = n \times e^{n} \times \big(y(n,k) - \bar{y}(n,k)\big) \tag{21}$$

$$\hat{x}_k^{+} = \boldsymbol{A}\hat{x}_k^{-} + \boldsymbol{K}_k(\xi_k - \boldsymbol{H}_k\boldsymbol{A}\hat{x}_k^{-}) \tag{22}$$

The general overview of the modified KF algorithm, after deploying the modified equation (Eq. 22), can be expressed as a flowchart (Figure 3). The procedure for target signal identification using the

proposed KF differs from the previously mentioned literature and will be discussed in the following sections.

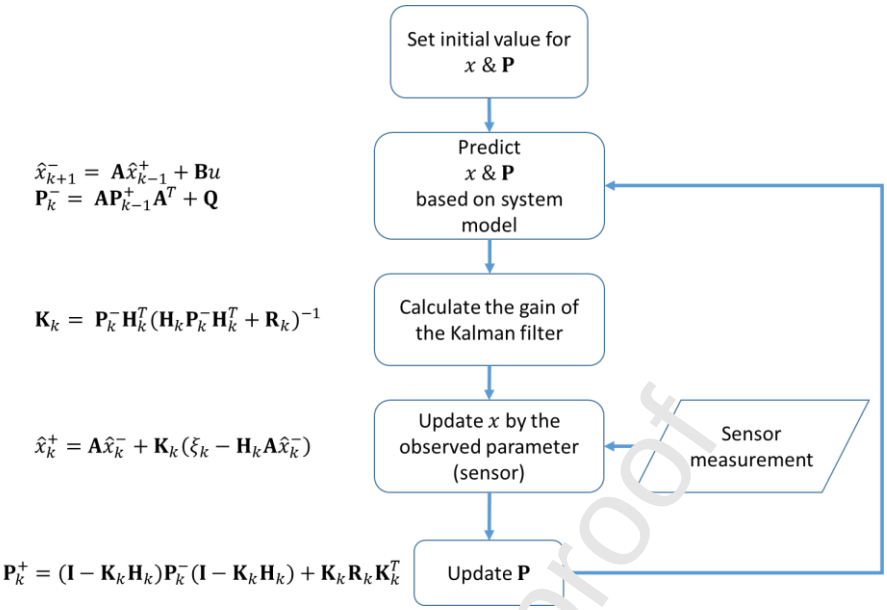


Figure 3. Discrete Kalman Filter flowchart

2.6 A novel target detection procedure using Normalised Innovation Squared (NIS)

A novel approach to target detection is introduced in this study, utilising the Normalised Innovation Squared (NIS) parameter, a unique feature of the KF. This innovative method paves the way for more accurate and efficient detection of buried targets, addressing the limitations of existing methods.

As previously highlighted, the accuracy of the KF can be determined through the parameter innovation $\boldsymbol{\nu}$, in its general form as presented in Eq. 23, which is the residual difference between the value of the estimated return signal by the KF and its actual measured value from the receiver antenna.

$$\boldsymbol{\nu} = y(k) - \mathbf{H}\mathbf{A}\hat{x}_k^- \tag{23}$$

However, as discussed earlier, detection of target signals using this approach in correlation with the chi-squared testing method has several limitations, including lack of accuracy, reliance on user inputs, and dependency on a predefined (often irrelevant) database. Therefore, to overcome these problems a new approach is suggested in this paper that uses the NIS in a different way for updating KF estimations, resulting in a more cohesive and innovative method.

Equation 23 is used to obtain the NIS, which can be expressed as (Eq.24):

$$NIS(k) = \nu(k)^T\,\mathbf{S}^{-1}\,\nu(k) \tag{24}$$

Similar to the innovation parameter, the NIS can be considered as an indicator of KF performance but is scalar, where $\mathbf{S}$ is the innovation covariance (Eq.25):

$$\mathbf{S} = \mathbf{H}(k)\mathbf{P}(k)^{-1}\mathbf{H}(k)^T + \mathbf{R}(k) \tag{25}$$

The value of NIS will be updated at each step of the estimation and illustrates the health status of the KF in terms of accuracy. A significant increase in the NIS value during the Kalman filtering process is an indicator of a growing difference between the KF estimate and the sensor output, which is a trigger for a faulty estimate. In other words, the higher the value of NIS, the more significant the difference between the KF estimate and the measured signal. In this study, all computed NIS values were recorded separately (as a chain), and a new signal called the NIS signal was generated, which was used to detect buried anomalies.

By characterising the NIS signal, the analysis can be performed only within the KF algorithm without the need for any external testing approach that might add biases to the outcome. Thereby, the decision on target positions or ground anomalies will be based on real measured data instead of a selected threshold by an expert.

To comply with the modification of the KF in the previous section (Eq. 22), the parameter $y(k)$ of the *innovation* (Eq. 23) was replaced with the parameter $\xi(k)$ (Eq.21) from which a modified NIS equation was specified (Eq.27).

$$\boldsymbol{\nu}' = \xi(k) - \mathbf{H}\mathbf{A}\hat{x}_k^- \tag{26}$$

$$NIS(k) = \boldsymbol{\nu}'(k)^T \, \mathbf{S}^{-1} \, \boldsymbol{\nu}'(k) \tag{27}$$

This approach not only addresses the limitations of the traditional target detection methods but also integrates the proposed improvements in the KF algorithm, resulting in a more innovative and unified solution. An accurate investigation of the NIS signal obtained from real data requires a good overview of its characteristics, which could be obtained from a noise-free radargram.

2.6.1 Generating noise-free NIS signal using numerical modelling

Numerical modelling is employed to generate a noise-free NIS signal. This step significantly enhances the accuracy of target detection. The innovative use of numerical modelling in conjunction with the NIS signal is a key aspect of the methodology. To generate a noise-free radargram, a survey configuration was numerically modelled with zero noise contribution using a trial site. The GPR data used in this study were obtained from a series of measurements taken from an accurately constructed trial site in conjunction with the FINDIT project (Metje et al., 2022), which covered 23 scan lines across an 11 m x 13 m area with five linear pipes installed at shallow depths. Pipe covering depths ranged from 223 mm to 814 mm (Figure 4). GPR data were recorded in DZT GGSI format using a single-channel antenna with a centre frequency of 400MHz. Scan line spacing was 0.5 m, with lines located perpendicular to the pipes. The ground condition was tested and confirmed as a 6.1 m thick glacial sand and gravel over red crag sand. Hyperbola matching in GPR indicated an approximate value of 6.5 for the dielectric constant of the sand.

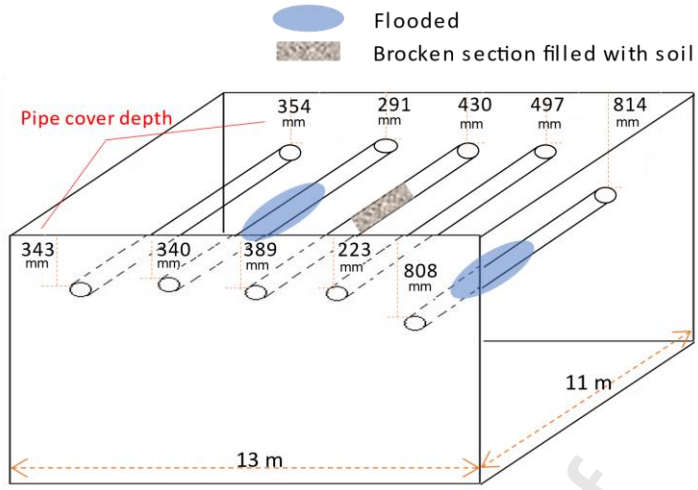


Figure 4. Detailed configuration of the buried pipes on the trial site

A numerical model was generated (Figure 5a) using the GPR modelling software matGPR (Tzanis and Kafetsis, 2004) from its numerical radargram (Figure 5b) to compare the characteristics of the numerically generated NIS signal with the NIS signal generated from real data.

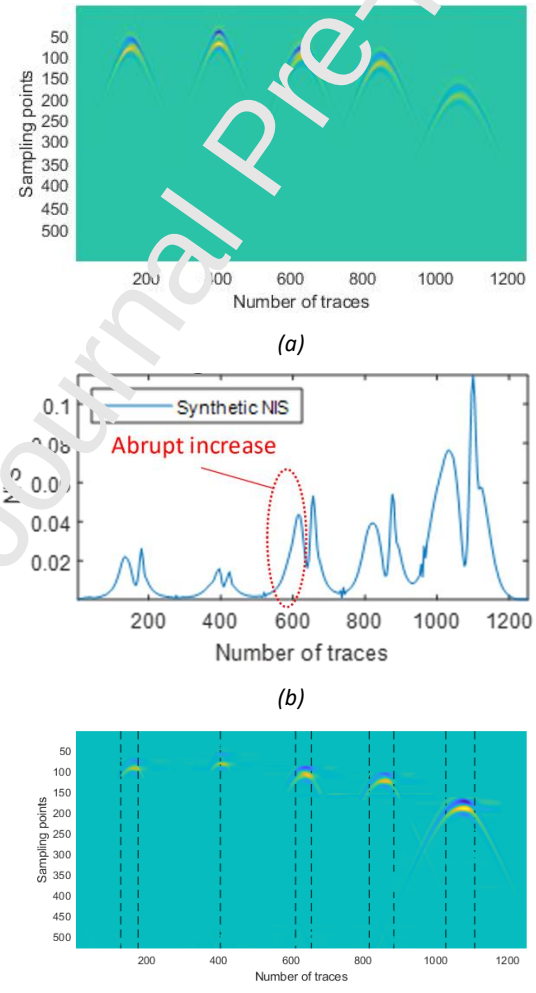


Figure 5. a) Numerically generated model, b) NIS *generated in the KF, c) identified peak location using the* NIS *signal*

An abrupt increase of NIS can be observed, for example, in the vicinity of trace number 600 (Figure 5b), where the KF enters the hyperbola zone, which is an indicator of a faulty KF estimate. A double signal-peak trend can be observed on the hyperbola tails (left and right of the vertex) instead of a single peak at the vertex. The logical explanation for these trends is the KF's attempt to correct its faulty estimate at these positions. After entering the KF into a target-present section, the corresponding trace containing target information is added into the system. Hence, the discrepancy between the predicted value and the sensor's output will be increased. Therefore, the KF attempts to reduce the discrepancy through subsequent updates. After an abrupt increase of the NIS signal, more weight will be put on the sensor data instead of the KF estimate. Consequently, the new estimate will be approximated concerning the sensor data, by which the amplitude of the NIS signal at the vertex will be reduced. The same trend occurs when the KF leaves the hyperbola zone and moves into a target-free section, where a second peak will be created.

As illustrated in Figure 5b, the NIS signal where a target (or anomaly) is present indicates a significantly higher amplitude compared to a target-free section. In other words, changes in NIS values where a target is present exhibit higher amplitude and lower frequency compared to background noise signals with higher frequencies and lower amplitudes. Therefore, sampling and evaluation of the NIS signal can be seen as a powerful tool to locate shallow pipes and ground anomalies. The position of the identified peaks can be indicated using dashed lines generated by the KF (Figure 5c).

Analysing the numerically generated NIS signal confirmed the accuracy of the proposed algorithm at this stage. However, real GPR data might contain severe ambient noises and other perturbations that result in a noisier and more ambiguous NIS signal, which can be processed and filtered by the KF.

2.6.2 Target detection algorithm using the Kalman Filter's NIS signal

The target detection algorithm leverages the KF's NIS signal, enabling a more precise identification of buried targets. This represents a significant improvement over existing methods, which often struggle with accurate target detection. Under the assumption of operating in the target-free state within the KF, the NIS values for each sampling point are calculated and recorded individually. The mean value of the generated NIS values is computed and used as a decision criterion to determine whether KF should continue operating in the target-free state or switch to the target-present state. Target and ground dielectric anomalies can be detected by evaluating the NIS signal at each point of the radargram and identifying the signal peaks. Similar to the noisy backscattered GPR signal, the NIS signal contains high-frequency noise that needs to be filtered. A commonly employed technique to determine frequency components (e.g., noise) involves utilising a suitable signal filtering tool, such as the Fast Fourier Transform (FFT). However, the performance of the FFT heavily relies on the selected range of frequencies that are allowed to pass through the filter.

The optimal range of Fourier frequencies was determined using a correlation between ROC curves (trade-off between the probability of true detections and false alarms) and FFT frequencies through an optimisation technique. Since different objectives are associated with these two cost functions, and no single solution satisfies both objectives simultaneously (non-dominated solution), the problem becomes nontrivial, necessitating a multi-objective optimisation approach. Due to the

problem's complexity, traditional optimisation methods are often insufficient for finding global solutions. Consequently, a Multi-Objective Genetic Algorithm (MOGA) was employed to address the optimisation problem.

It is important to note that determining the optimal FFT frequency range is a case-dependent task. As such, a general optimal frequency range cannot be established at this stage unless comprehensive testing is conducted for all soil and pipe type combinations.

2.7 Overview of the proposed optimisation algorithm

The proposed optimisation algorithm integrates the KF, NIS signal function analysis, and genetic algorithm-based multi-objective optimisation. This unique combination of techniques results in a highly effective and efficient approach to GPR data processing. The input variables of the KF, such as the model and sensor noise covariance matrices **Q** and **R**, and the FFT frequency, need to be determined appropriately before executing the KF algorithm. Due to the varying characteristics of these input parameters, two different cost functions—the mean of the innovation and the area under the ROC curves—were considered for the optimisation process.

2.7.1 Optimisation of noise covariances Q and R

An innovative approach to the optimisation of noise covariances **Q** and **R** is introduced. This optimisation significantly enhances the accuracy of target detection, further demonstrating the effectiveness of the methodology. Selecting inappropriate noise covariance parameters Q and R can significantly degrade the performance of the KF (e.g., manipulation of the NIS values). Noise is a random parameter that can have various origins, but it can be described mathematically as a probability density function (Van Drongelen, 2010). Deciding on an appropriate value is case-dependent and can be challenging, especially for the model noise covariance matrix (Q), where no information is available. According to the definition of the KF described by Simon, (2006), the innovation parameter must have a zero-mean value ($\mu = 0$).

These characteristics were harnessed to determine the optimised values of the noise covariance matrices Q and R by implementing MOGA-based optimisation, ensuring the mandatory zero-mean condition (Simon, 2006) is met. This aim was achieved using a standard method in the literature called the Weighted Sum Method (WSM), which computes a cost function (Eq. 3-28) as a linear weighted summation of the associated objectives (Alizadeh et al., 2019).

$$J = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{k}\sum_{j=1}^{k}\boldsymbol{\nu}_{ij}^{2}\right) = 0 \tag{28}$$

To avoid negative values and simplify the calculation, the values of the innovation were squared. Considering the dimensions of the radargrams of the associated dataset, a total of 1020 variables were used: 510 variables for Q and 510 variables for R.

2.7.2 Optimisation of the Fourier frequencies

The methodology also includes the optimisation of Fourier frequencies, a step that greatly improves the quality of GPR data processing. This represents another key innovation of the approach. Signal noise reduction must be conducted in a manner that retains useful information hidden within noisy signals. To determine the optimised value for the Fourier frequency, MOGA was employed to maximise the area underneath the ROC curves while minimising the means of innovation. The ROC analysis required a benchmark for classifying GPR signals to determine whether the signal is truly a target signal or a false alarm. This benchmark is defined by the expert as the approximate positions of potential targets after implementing pre-processing steps on one or more radargrams through the algorithm. This optimisation approach can effectively obtain the optimised Fourier frequency if implemented on a series of radargrams rather than a standalone one.

Configuration of MOGA parameters, including population size, number of generations, mutation probability, and crossover, were determined through trial and error. A population size of 30 and a generation size of 100, along with the GA's default mutation rate of 0.01, were selected for the model among standard parameter values defined in MATLAB. The computational time required to determine the optimal values using a personal computer (PC) with an Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz and 32 GB of memory was 06:31 hours. The processing time of the algorithm, after employing the optimal values of Q and R to post-process the radargrams, was 124 seconds on the same PC.

A typical result of the MOGA analysis is shown in Figure 6, with similar results obtained for all other radargrams of the trial site (Figure 4a). However, the analysis of Pareto solutions revealed that not all Pareto solutions guarantee an optimal result in terms of detection rate and false alarm ratio. This will be further discussed in the results and discussion section.

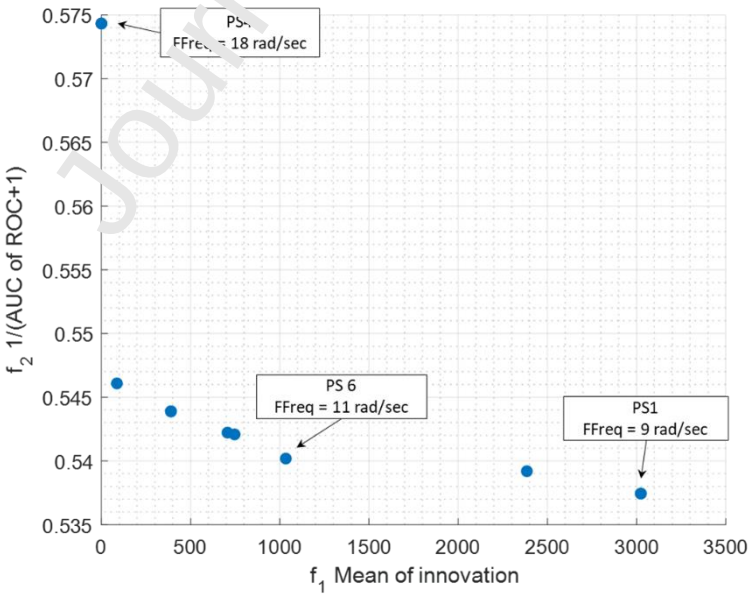


Figure 6. Pareto solutions of the 6th radargram at the trial site based on AUC of ROC curves and Fourier frequency (FFreq)

2.8 Solution to 'identify the target width'

The estimated radargram, obtained from a noisy backscattered signal after implementing the KF algorithm as presented in the previous section, can generally be considered the final radargram used by experts for visual confirmation. As discussed earlier, a significant issue with previous studies is their inability to identify the target width. To address this challenge, this study implements N independent KFs (where N is the number of sample points) on each row of the radargram, instead of processing an individual trace segment by segment. In other words, since each trace contains N sample points, N independent KFs can be executed simultaneously to estimate the information for each row individually. As a result, the average value of the NIS from the target-free state can be compared with the actual NIS value of each point on a row. This comparison allows for the evaluation of the information in each trace segment, enabling the detection of the endpoint of single or multiple buried targets. The pseudo-code for this procedure is illustrated in Figure 7.

```
1.  PROGRAM KF_GPR(Radargram, A, H, Q, R, FFreq)
2.  .    FOR Width = [ FullLength, 1 ] DO
3.  .        Calculate number of layers ;
4.  .        FOR p = 1 to final layer DO
5.  .            Select u ;
6.  .            Calculate u' ;
7.  .            FOR k = 1 to last trace DO
8.  .                KalmanFilter(A, H, Q, R) in target free mode;
9.  .                Saving NIS for each trace;
10. .                Go to next trace ;
11. .            ENDFOR
12. .            IF Width == FullLength THEN
13. .                FourierFilter(NIS, Ffreq);
14. .                Calculate the position of peaks in filtered NIS ;
15. .                Go to line 2 ;
16. .            ELSE
17. .                Calculate the average of NIS ;
18. .            ENDIF
19. .            FOR k = 1 to last trace DO
20. .                IF NIS > average of NIS THEN
21. .                    KalmanFilter(A, H, Q, R) in target present mode;
22. .                ELSE
23. .                    KalmanFilter(A, H, Q, R) in target free mode;
24. .                ENDIF
25. .                Go to next trace;
26. .            ENDFOR
27. .        ENDFOR
28. .    ENDFOR
29. END.
```

Figure 7. Pseudocode for the detection of target width

3 Results and Discussion

The results of the study demonstrate the effectiveness and innovation of the methodology. The approach to GPR data processing resulted in accurate target detection and significantly reduced noise, underscoring the benefits of the innovative combination of techniques

3.1 GPR survey data

As mentioned in Section 2.3.1, a typical GPR radargram, such as radargram 009 from the trial site dataset (Figure 8), illustrates the cross-section of the subsurface. The dataset used in this study consists of five buried pipes. Once converted into ASCII format, the radargram results in a 1350 x 510 matrix (traces x sampling points).

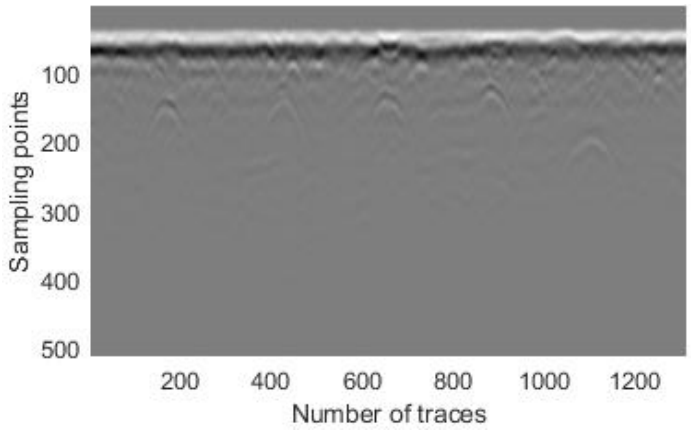


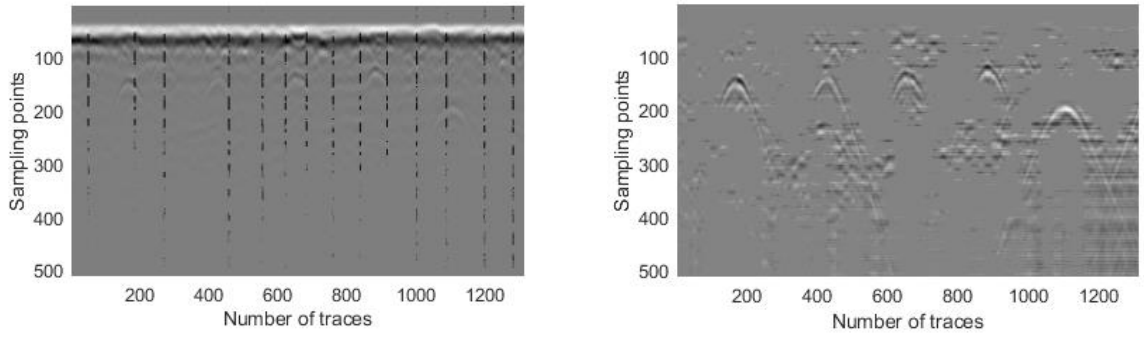
Figure 8. Typical GPR radargram

3.2 Optimisation of the GPR Outcome Using MOGA Analysis

The optimal Pareto solutions (PSs) were utilised to compare the performance of the algorithm by considering the position of the NIS signal peaks, which are visualised using dashed lines in the subsequent figures.

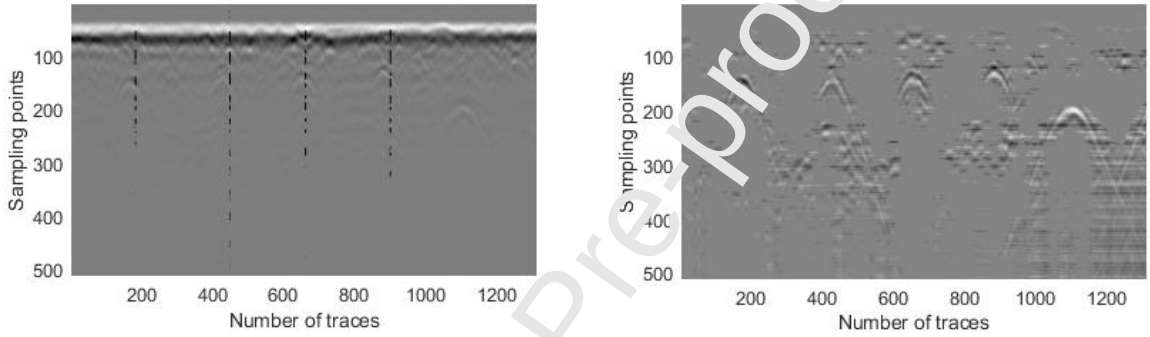
The detection outcomes of the algorithm using two PSs (PS1 and PS4) for two different radargrams (009 and 013) are presented in Figures 9 to 12. As discussed earlier, reflected GPR signals due to the contrast in dielectric permittivity of materials serve as indicators of existing targets or ground anomalies. The positions of such contrasts were detected by the proposed algorithm and are indicated automatically using dashed lines on the original radargrams (Figures 9a to 12a). The processed radargrams generated by the algorithm, with estimated and removed backgrounds, are presented in Figures 9b to 12b.

A comparison of PSs reveals that the solutions are optimised in different ways. In other words, using a PS with different optimisation criteria can deliver different results in terms of detected anomalies (dashed lines), which serve as indicators of the variation in the system characteristics in terms of the ROC and Fourier frequency. For instance, evaluation of the PS4 results showed that the corresponding mean value of the innovation tends to be minimised and produces a high number of dashed lines (Figure 9a), compared to the PS1 results, which tended to maximise the area under the ROC curves and produce a minimum number of dashed lines (Figure 11a). However, the minimum number of dashed lines in Figure 10a resulted in missing the fifth target hyperbola around the 1100th trace in Figure 10b. The dashed lines can also be considered as false alarms if an inappropriate PS is selected. The number of dashed lines has a direct relationship to the system's sensitivity, which can be set by the expert based on PSs.



(a) *(b)*

Figure 9. Comparison of radargrams. a) Original GPR radargram 009 with target positions indicated by dashed lines, determined using Pareto solution PS4. b) Radargram 009 processed using the KF algorithm and Pareto solution PS4, with noise and clutter removed.



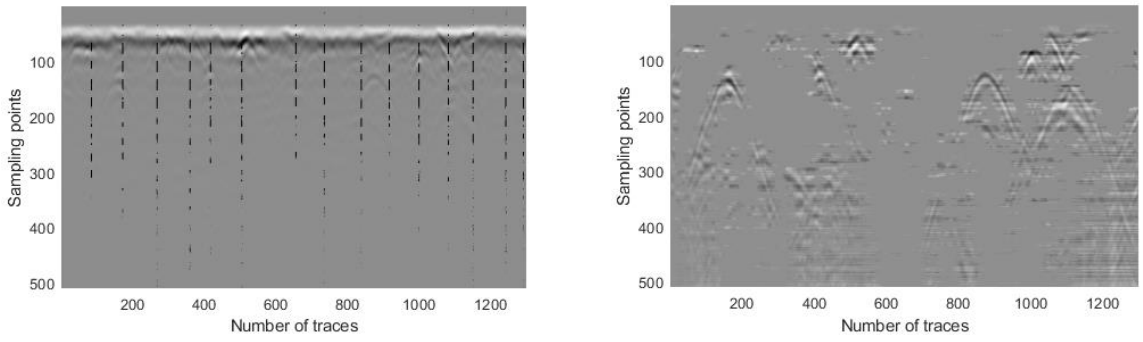
(a) *(b)*

Figure 10. Comparison of radargrams. a) Original GPR radargram 009 with target positions indicated by dashed lines, determined using Pareto solution PS1. b) Radargram 009 processed using the KF algorithm and Pareto solution PS1, with noise and clutter removed.

As illustrated in the site layout shown in Figure 4a, there is a pipe near the 700th trace in the radargram, which features a broken section with a blockage between the 12th and 15th traverses. The processed radargrams in Figure 11b and Figure 12b do not display any hyperbola signatures related to this pipe. However, the presence of the dashed lines generated by the algorithm clearly indicates a potential anomaly in this region. This suggests that the proposed algorithm is also capable of detecting defective target sections (e.g., broken pipe sections) and corresponding ground anomalies using dashed lines.

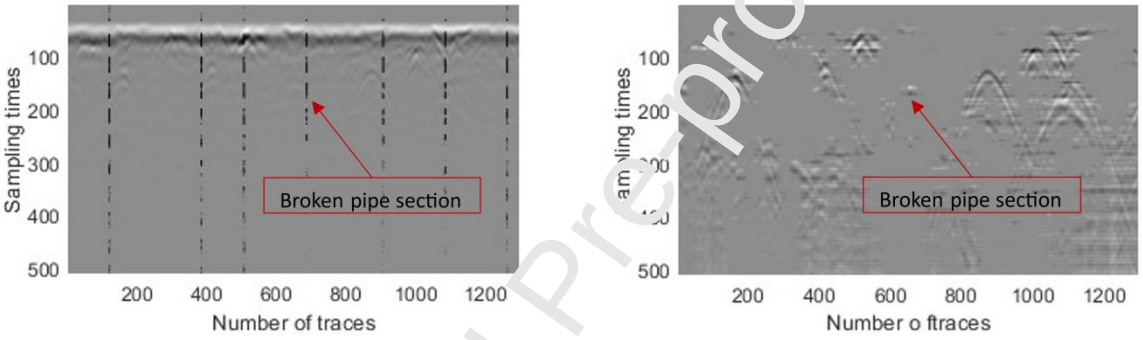
The benefits of the proposed algorithm become evident when comparing the processed radargrams on the right-hand side (e.g., Figure 11b) with those on the left (e.g., Figure 11a). The likely positions of anomalies are highlighted on the original radargrams by a series of dashed lines that were not previously obvious.

Alongside the actual five hyperbola signatures, additional near-surface hyperbolas were also detected. These near-surface anomalies could be due to factors such as pebbles, tree roots, or rabbit and mole holes, which contribute to the appearance of these extra hyperbola features.



(a) *(b)*

Figure 11. Comparison of radargrams. a) Original GPR radargram 013 with target positions indicated by dashed lines, determined using Pareto solution PS4. b) Radargram 013 processed using the KF algorithm and Pareto solution PS4, with noise and clutter removed.



(a) *(b)*

Figure 12. Comparison of radargrams. a) Original GPR radargram 013 with target positions indicated by dashed lines and the detected ground anomaly, determined using Pareto solution PS1. b) Radargram 013 processed using the KF algorithm and Pareto solution PS1 with noise and clutter removed and the detected ground anomaly.

3.3 Evaluating the Performance of the MOGA Optimisation of KF

To establish a benchmark for comparison, the results (the values of covariances Q and R) of the MOGA optimisations were compared against those obtained through trial and error. The significant differences between the MOGA-based generated NIS signals and the experimentally-derived NIS signals are illustrated in Figures 13 and 14. As shown on the vertical axis of each plot (Figure 13a and 13b), the experimentally-derived NIS signals indicated values with maximum orders of $x10^{11}$, and $x10^{12}$ for radargrams 009 and 013, respectively. In contrast, the MOGA-based NIS signals, utilising PS4, yielded values with maximum orders of $x10^{-23}$, and $x10^{-22}$, respectively (Figures 14a and 14b).

The high performance and accuracy of the proposed KF algorithm in this study are evident from the near-zero order of the MOGA-based NIS signals, which fully comply with the theory of the KF. As indicated by Simon (2006), smaller values of NIS indicate higher performance of the algorithm.

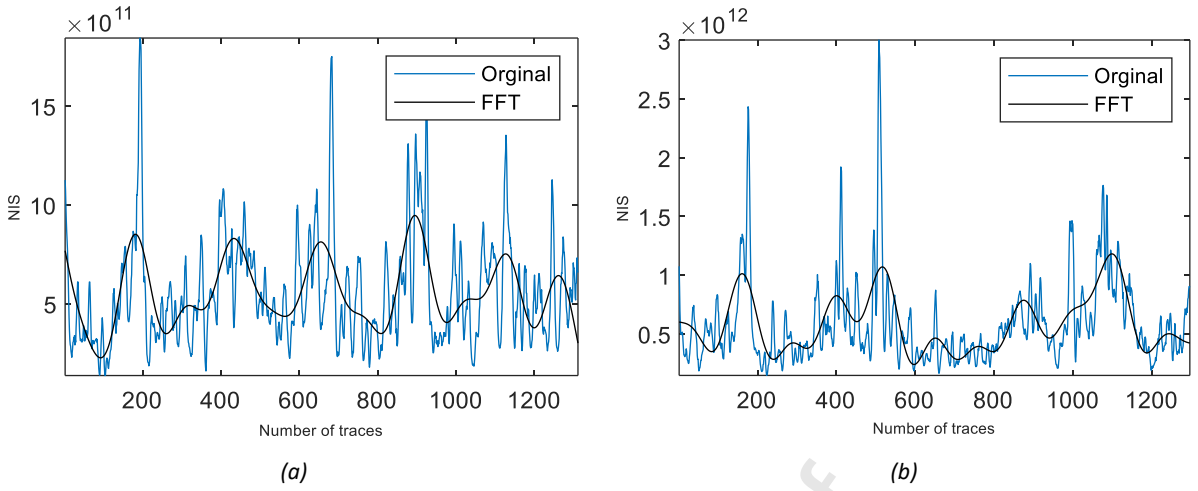


Figure 13. a) NIS *signal for experimentally obtained* **Q**-**R** *covariances from radargram 009, b)* NIS *signal for experimentally obtained* **Q**-**R** *covariances from radargram 013*

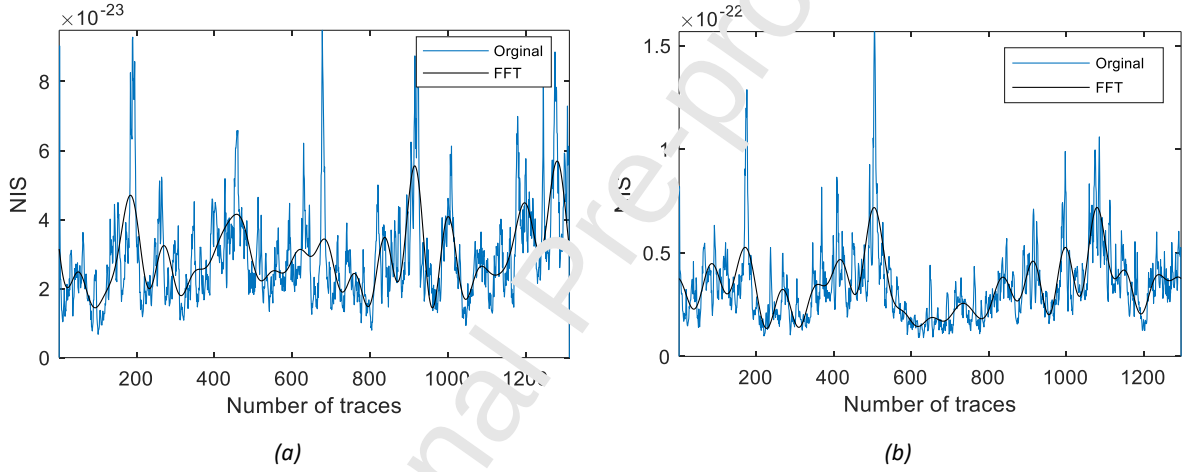


Figure 14. a) NIS *signal for optimised* **Q**-**R** *covariances from radargram 009 using PS4, b)* NIS *signal for optimised* **Q**-**R** *covariances from radargram 013 using PS4*

3.4 Generated ROC Curves from Pareto Solutions (PSs)

The area under each ROC curve for different Fourier frequencies in two PSs (PS4 and PS1) was calculated, and the associated covariance matrices R and Q were determined. A few examples of corresponding ROC curves associated with these two PSs are illustrated in Figures 15a and 15b. A higher general average value for the ROC curves was observed for PS1 (Figure 15b) compared to the ROC curves for PS4 (Figure 15a). As mentioned earlier, higher ROC values on the Y-axis indicate a higher probability of target detection against the probability of false alarms on the X-axis.

In general, the mean value of the areas under the ROC curves changed significantly if Fourier frequencies (FFreq) were selected randomly outside the PS. Lower mean values were observed after using randomly selected FFreq compared to optimised values associated with the PS. For instance, randomly selected FFreqs such as 20, 35, and 40 rad/sec, instead of those provided by PS4, yielded ROC mean values of 68.27%, 65.35%, and 65.45% respectively. In contrast, the PS4 optimal FFreqs

such as 9, 10, and 18 rad/sec yielded higher ROC mean values of 71.29%, 73.05%, and 68.89%, respectively. A similar pattern was observed for the PS1 results compared to PS4, but they generally yielded higher mean values in the range of 69.78% to 74.03%.

It should be noted that PS4 was the point on the Pareto solution curve where the weighted parameter of the innovation was optimised in terms of its zero-mean characteristic. In other words, the optimisation algorithm generated PS4 in a way that guaranteed the required zero-mean condition. On the other hand, PS1 was the point on the curve where the weighted parameter was the optimised area under the ROC curves, ensuring the maximum area under the curves

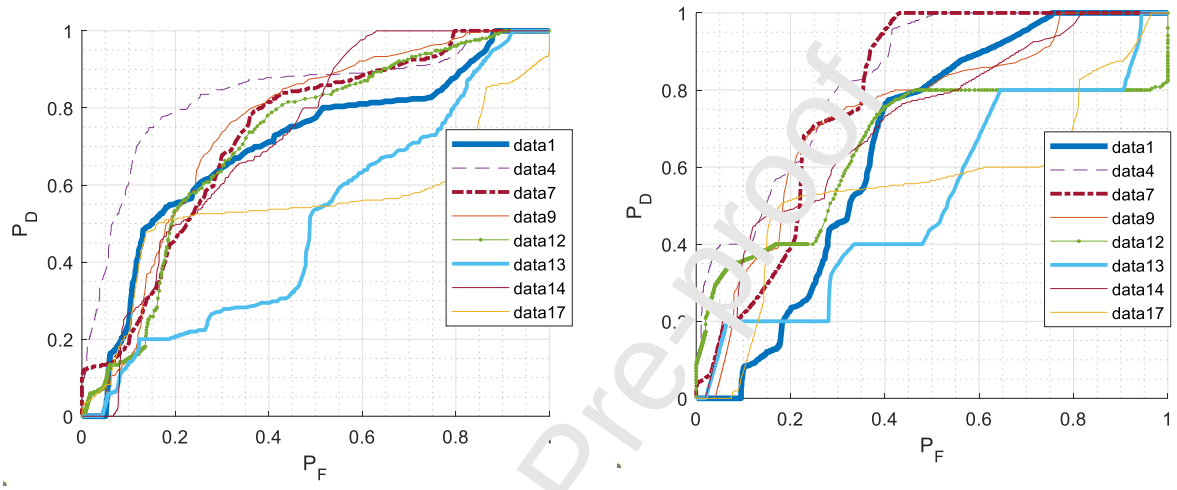


Figure 15. a) Generated ROC curves for all radargrams in dataset using PS4, b) Generated ROC curves for all radargrams in dataset using PS1

To further verify the performance of the algorithm in terms of functionality and detection biases, another independent dataset was analysed using the proposed algorithm in this study.

3.5 Verification of the Algorithm's Performance

The dataset was obtained from the northern section of the trial site discussed in Section 2.3.1, featuring identical ground conditions but different buried utilities. This included seven buried features, such as pipes and telecommunication cables, as well as an induction chamber. The associated Pareto solutions (Figure 16) demonstrated favourable values for the ROC curves, exceeding 91.4% (Y-axis), while maintaining an almost zero mean of innovation values on the order of E10-4 (X-axis).

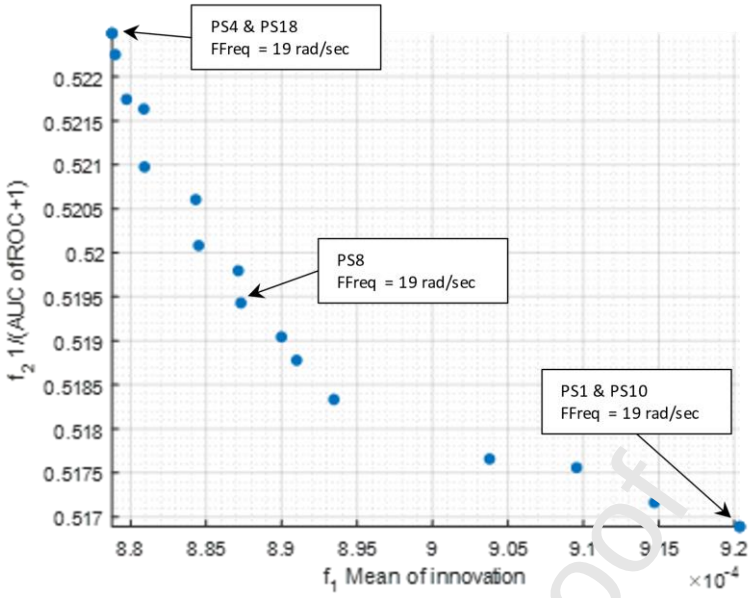
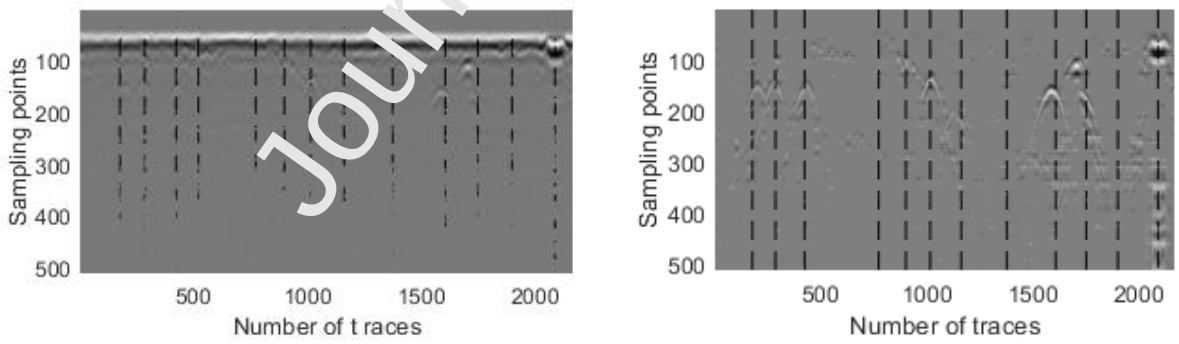


Figure 16. Pareto solutions for the north section of the first survey site but different utilities

The results of the algorithm for a selected radargram from the dataset (Figure 17a) using PS4 (or PS18) from the associated Pareto solution curve are shown in Figure 17b. The algorithm detects the estimated positions of possible anomalies, including target positions, and displays them in Figure 17a using dashed lines on both the original and processed radargrams. The findings indicate that all utilities (i.e., all seven buried utilities and the inspection chamber) were successfully identified. Some additional warning lines may be considered false alarms; however, as mentioned earlier, such warning lines will be generated if any irregularity is detected in the ground.



(a) *(b)*

Figure 17. Radargrams from the north section of the first survey site but different utilities a) Original radargram 10, b) processed radargram 10

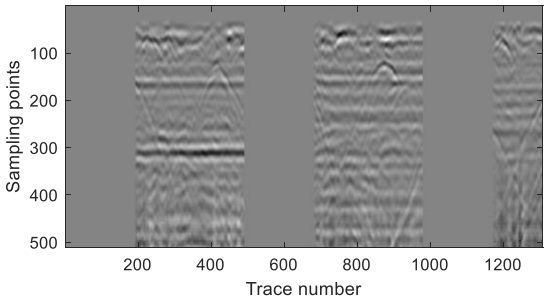
3.6 Evaluation of the Proposed Algorithm

In this study, the zero-mean property of KF's performance benchmark parameters (i.e. the innovation and NIS) was used to determine the optimal values for **Q** and **R** covariances. As described by Grewal

and Andrews (2008), assuming that sources of errors have a Gaussian distribution, the NIS parameter will have a chi-square distribution with m degrees of freedom (m is the length of innovation vectors). Although this property was used in Carevic (1999)'s proposed algorithm (discussed in Section 1 of this paper) to detect buried targets, it has considerable drawbacks and limitations. Specifically, there are five parameters (i.e. threshold, null hypothesis rejection parameters (K_0 and K_1), width of targets, and number of layers) that need to be specified by the user in advance. Due to variations in quality and specifications of different GPR datasets, such parameters are not unique and might vary for different datasets. Moreover, the performance of Carevic's algorithm is highly dependent on the correct choice of these parameters.

To address these limitations, a new approach was proposed that eliminates the need to specify parameter values in advance. Figure 18 illustrates a comparison between example results obtained with Carevic's algorithm (Figure 18a, b, c) and the result obtained from our proposed algorithm (Figure 18d), for which no parameter values were needed. Furthermore, the dependency of Carevic's algorithm on the correct choice of parameters can be observed in Figures 18a, b, c, as the width of targets cannot be identified by their algorithm. This can lead to problems such as missing targets (Figure 18a) or a high level of false alarms (Figure 18c). In contrast, our proposed algorithm uses the mean of NIS values as a benchmark for the detection of targets and ground anomalies, instead of the chi-squared testing hypothesis. This approach eliminates the need to determine such parameters through the user and has proven to be computationally cost-efficient and accurate, due to its semi-autonomous nature and minimisation of expert interaction for determining required parameters.

Furthermore, conventional pre-and post-processing of GPR data using software such as REFLEX can be a complex and time-consuming task, often requiring manual expert intervention (Ernenwein and Kvamme, 2008). The proposed algorithm in this study is computationally cost-efficient and accurate, as it minimises the need for expert intervention and is semi-autonomous in nature.



(a)

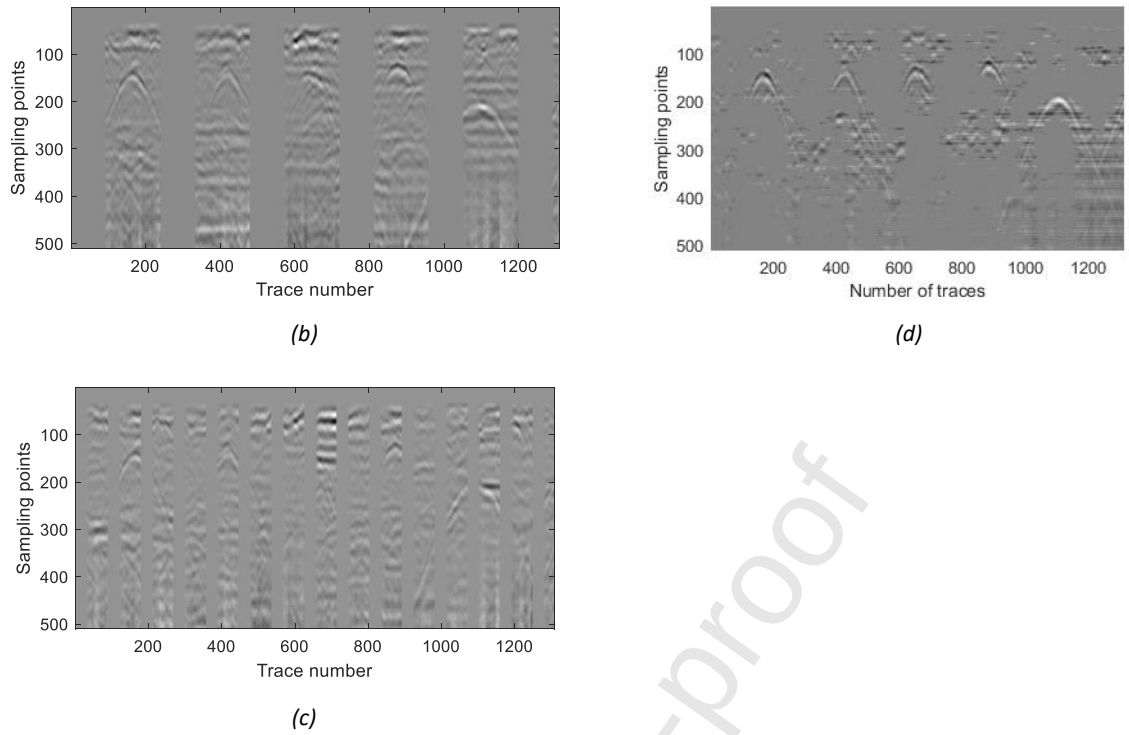


Figure 18. Comparison of the Carevic (1999) algorithm based on chi-squared hypothesis testing with the proposed algorithm in this research. (a-c) Results obtained with the Carevic algorithm using different user-defined window sizes, (d) Result obtained with the proposed semi-autonomous algorithm, and (e) Result obtained with the proposed autonomous algorithm

4 Summary and Conclusions

The efficiency and accuracy of processing and interpreting GPR data for detecting buried targets depend on the level of expert knowledge. The backscattered GPR signal comprises target signals, background noise, and clutter, making interpretation particularly challenging. In this paper, a novel semi-autonomous algorithm was presented that combines the KF and machine learning approaches, specifically, a genetic algorithm, to optimise the process of estimating the backscattered signal in a noisy environment. Potential buried targets were detected in this algorithm by generating, utilising, and processing the Normalised Innovation Squared (NIS) signal. The NIS is an internal parameter of the KF, and its updated value at each step illustrates the health status of the KF in terms of accuracy. Receiver Operating Characteristics (ROC) curves were utilised to evaluate the algorithm's performance. Furthermore, two commonly used pre-processing filters, namely the averaging filter and signal gain filters, were adapted into the proposed algorithm. This adaptation yielded improved quality of the final output radargram by cancelling background noise and appropriately amplifying the backscattered signals. Additionally, the proposed technique reduced the need for expert intervention. Expert involvement was required only for the visual confirmation of the approximate positions of the buried targets, as demonstrated by the KF for one or more processed radargrams before executing the optimisation procedure on all radargrams in the dataset.

The algorithm comprises two major sections. The focus of the first section is to generate an optimised quality radargram for visual confirmation by an expert. The second section of the

algorithm focuses on detecting buried targets and ground anomalies using a multi-objective optimisation algorithm. Three real GPR datasets were processed, leading to the conclusion that the reduction in the area under the ROC curves is proportional to the rate of false alarms concerning detected targets and potential ground anomalies. Based on the lowest optimal area under the ROC curve (>74%), the best Pareto solution was selected, successfully detecting all five targets. However, it should be noted that the false alarm rate in this study was influenced by multiple factors such as FFT frequency and ground homogeneity. Given that the study aimed to identify all ground anomalies, a general false alarm rate could not be feasibly provided. For a more quantifiable measure of the false alarm rate, comprehensive research with a broader dataset would be required. Within the optimised Pareto solutions, the lower the optimal area under the ROC curves, the higher the optimal detection rate of the algorithm, with no missing targets. Moreover, the algorithm detected broken or blocked pipe sections that were invisible in a normal radargram by generating warning lines. The effectiveness of the algorithm was assessed in terms of biased estimation using two different additional GPR datasets, for which no initial information was available for one concerning the number, position, type of buried targets, and ground conditions.

From the results of this paper, the following conclusions can be drawn:

- Adapting pre-processing filters into the proposed algorithm effectively reduced the pre-processing effort and enhanced the quality of the results.
- One of the main problems associated with employing the KF for GPR, namely the issue of not being able to achieve "width of target detection," can be solved by utilising a series of KF algorithms.
- The statistical chi-squared hypothesis method can efficiently be replaced by an NIS signal analysis for identifying buried utilities using the KF approach.
- An NIS signal analysis identifies target positions based on the real sampled data and does not require selecting a threshold and determining other coefficients by the expert during chi-squared testing.
- The complexity of GPR data post-processing and, consequently, the associated computational cost is significantly reduced, compared to existing conventional techniques, owing to the scalar format of NIS values and the semi-autonomous nature of the proposed algorithm.
- The multi-objective optimisation algorithm can optimise the detection rate by determining the optimised Fourier frequency and values of noise the covariances Q and R.

References

Alizadeh, V., Fehri, H., Kessentini, M., 2019. Less is more: From multi-objective to mono-objective refactoring via developer's knowledge extraction. In: Proceedings - 19th IEEE International Working Conference on Source Code Analysis and Manipulation, SCAM 2019. Institute of Electrical and Electronics Engineers Inc., pp. 181–192.

Bar-Shalom, Y., Li, X.-R., Kirubarajan, T., 2001. Estimation with Applications to Tracking and Navigation, Estimation with Applications to Tracking and Navigation.

Benedetto, A., Tosti, F., Luca Bianchini Ciampoli, F.D., 2013. An overview of ground-penetrating radar signal processing techniques for road inspections. J. Chem. Inf. Model. 53, 1689–1699.

Carevic, D., 1999. Kalman filter-based approach to target detection and target-background separtion in ground-penetrating radar data. In: Detection and Remediation Technologies for Mines and Minelike Targets IV. p. 1284.

Cheng, Q., Zhang, S., Chen, X., Cui, H., Xu, Y., Xia, S., Xia, K., Zhou, T., Zhou, X., 2023. Inversion of reclaimed soil water content based on a combination of multi-attributes of ground penetrating radar signals. J. Appl. Geophys. 213, 105019.

Daniels, D.J., 2004. Ground Penetrating Radar-2nd edition.

Ernenwein, E.G., Kvamme, K.L., 2008. Data processing issues in large-area GPR surveys: Correcting trace misalignments, edge discontinuities and striping. Archaeol. Prospect. 15, 133–149.

Goodman, D., Piro, S., 2013. GPR Remote Sensing in Archaeology. Geotechnol. Environ. Springer-Verlag Berlin Heidelberg.

Grewal, M., Andrews, A., 2008. Kalman filtering : Theory and Practice Using MATLAB.

Gurbuz, A.C., 2012. Determination of background distribution for ground-penetrating radar data. IEEE Geosci. Remote Sens. Lett. 9, 544–548.

Hartikainen, J., Särkkä, S., 2011. Optimal filtering with Kalman filters and smoothers – a Manual for Matlab toolbox EKF / UKF. Dep. Biomed. Eng. Comput. Sci. Aalto Univ. Sch. Sci.

Ho, K.C., Gader, P.D., 2002. A linear prediction land mine detection algorithm for hand held ground penetrating radar. IEEE Trans. Geosci. Remote Sens. 40, 1374–1384.

Ho, K.C., Gader, P.D., Wilson, J.N., 2004. Improving landmine detection using frequency domain features from ground penetrating radar. In: International Geoscience and Remote Sensing Symposium (IGARSS). pp. 1617–1620.

Hou, L., Zhang, Q., Zhang, R., 2022. Automatic Detection of Diseases in Tunnel Linings Based on a Convolution Neural Network and a Support Vector Machine. Electron. 11, 3290.

Huber, E., Hans, G., 2018. RGPR - An open-source package to process and visualize GPR data. In: 2018 17th International Conference on Ground Penetrating Radar, GPR 2018. Institute of Electrical and Electronics Engineers Inc.

Kaniewski, P., Kraszewski, T., 2020. Novel algorithm for position estimation of handheld ground-penetrating radar antenna. In: Proceedings International Radar Symposium. IEEE Computer

Society, pp. 100–102.

Kaniewski, P., Kraszewski, T., 2023. Estimation of Handheld Ground-Penetrating Radar Antenna Position with Pendulum-Model-Based Extended Kalman Filter. Remote Sens. 15, 741.

Katoch, S., Chauhan, S.S., Kumar, V., 2021. A review on genetic algorithm: past, present, and future. Multimed. Tools Appl. 80, 8091–8126.

Kerekes, J., 2008. Receiver operating characteristic curve confidence intervals and regions. IEEE Geosci. Remote Sens. Lett. 5, 251–255.

Khan, U.S., Al-Nuaimy, W., 2010. Background removal from GPR data using eigenvalues. In: Proceedings of the 13th Internarional Conference on Ground Penetrating Radar, GPR 2010.

Kim, J.H., Cho, S.J., Yi, M.J., 2007. Removal of ringing noise in GPR data by signal processing. Geosci. J. 11, 75–81.

Kramer, O., 2017. Genetic Algorithm Essentials. Studies in Computational Intelligence 679.

Li, R., Zhang, H., Chen, Z., Yu, N., Kong, W., Li, T., Wang, E., Wu, X., Liu, Y., 2022. Denoising method of ground-penetrating radar signal based on independent component analysis with multifractal spectrum. Meas. J. Int. Meas. Confed. 192, 110886.

Linna, P., Aaltonen, T., Halla, A., Gronman, J., Narra, N., 2020. Conceptual design of an autonomous rover with ground penetrating radar: Application in characterizing soils using deep learning. In: 2020 43rd International Convention on Information, Communication and Electronic Technology, MIPRO 2020 - Proceedings. Institute of Electrical and Electronics Engineers Inc., pp. 1174–1179.

Liu, C., Song, C., Lu, Q., 2017. Random noise de-noising and direct wave eliminating based on SVD method for ground penetrating radar signals. J. Appl. Geophys. 144, 125–133.

Luo, Y., Fang, G.Y., 2005. GPR clutter reduction and buried target detection by improved Kalman filter technique. In: 2005 International Conference on Machine Learning and Cybernetics, ICMLC 2005. IEEE, pp. 5432–5436.

McHugh, M.L., 2012. The Chi-square test of independence. Biochem. Medica 23, 143–149.

Metje, N., Chapman, D., Stringfellow, M., Rebegea, S., Tuckwell, G., Guy, M., Parkin, N., Roberts, D., Leech, C., Lord, A., 2022. Locating blockages in buried (telecoms) ducts: a new approach. Proc. Inst. Civ. Eng. - Smart Infrastruct. Constr. 175, 59–72.

Mirjalili, S., 2019. Genetic algorithm. Stud. Comput. Intell. 780, 43–55.

Ng, W., Chan, T.C.T., So, H.C., Ho, K.C., 2008. Particle filtering based approach for landmine detection using ground penetrating radar. IEEE Trans. Geosci. Remote Sens. 46, 3739–3755.

Patel, M., 2023. Payload position tracking and fractional control evaluation for a drone-based ground penetrating radar system.

Patel, M., Ferguson, P., 2021. Tracking and Estimation of a Swaying Payload Using a LiDAR and an Extended Kalman Filter. In: IEEE International Symposium on Robotic and Sensors Environments, ROSE 2021 - Proceedings. Institute of Electrical and Electronics Engineers Inc.

Särkkä, S., 2013. Bayesian filtering and smoothing, 1st ed, Cambridge University Press. Cambridge University Press.

Sen, M.K., Mallick, S., 2018. Genetic Algorithm with Applications in Geophysics. Springer, Cham, pp. 487–533.

Simon, D., 2006. Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches, Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches.

Smitha, N., Singh, V., 2016. Clutter reduction techniques of ground penetrating radar for detecting subsurface explosive objects. 2016 Int. Conf. Inf. Commun. Embed. Syst. ICICES 2016 1–8.

Tzanis, A., Kafetsis, G., 2004. A freeware package for the analysis and interpretation of common-offset ground probing radar data, based on general purpose computing engines. Bull. Geol. Soc. Greece 36, 1347.

Van Drongelen, W., 2010. Signal Processing for Neuroscientists, A Companion Volume, Signal Processing for Neuroscientists, A Companion Volume.

Van Merwe, A. Der, Gupta, I.J., 2000. A Novel signal processing technique for clutter reduction in gpr measurements of small, shallow land mines. IEEE Trans. Geosci. Remote Sens. 38, 2627–2637.

Wickramanayake, S., Thiyagarajan, K., Kodagoda, S., 2022. Deep Learning for Estimating Low-Range Concrete Sub-Surface Boundary Depths Using Ground Penetrating Radar Signals. IEEE Sensors Lett. 6.

Wu, R., Clement, A., Li, J., Larsson, E.G., Bradley, M., Habersat, J., Maksymonko, G., 2001. Adaptive ground bounce removal. Electron. Lett. 37, 1250–1252.

Zhao, L., Zhang, J., Jiao, S., Zheng, T., Li, J., Zhao, T., 2022. Ground surface detection method using ground penetrating radar signal for sugarcane harvester base-cutter control. Biosyst. Eng. 219, 103–123.

Zou, L., Wang, Y., Giannakis, I., Tosti, F., Alani, A.M., Sato, M., 2020. Mapping and assessment of tree roots using ground penetrating radar with low-cost GPS. Remote Sens. 12, 1300.

Zoubir, A.M., Chant, I.J., Brown, C.L., Barkat, B., Abeynayake, C., 2002a. Signal processing techniques for landmine detection using impulse ground penetrating radar. IEEE Sens. J. 2, 41–51.

Zoubir, A.M., Chant, I.J., Brown, C.L., Barkat, B., Abeynayake, C., 2002b. Signal processing techniques for landmine detection using impulse ground penetrating radar. IEEE Sens. J. 2, 41–51.

Declaration of interests

☐The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Arasti Afrasiabi reports financial support was provided by UK Research and Innovation. Arasti Afrasiabi reports a relationship with UK Research and Innovation that includes: funding grants.

Highlights

A semi-autonomous algorithm based on Kalman Filter (KF) for processing GPR data

Semi-autonomous detection and location of buried anomalies in the subsurface

Degerming optimised KF parameters using a multi-objective genetic algorithm

Producing a new signal (the NIS signal) for locating anomalies in the subsurface

Replacing chi-squared testing hypothesis by NIS signal analysis for anomaly detection