Development of Data Processing Tools for the Analysis of Radargrams in Utility Detection Using Ground Penetrating Radar

Florence Sagnard

Université Lille Nord de France, IFSTTAR, Villeneuve-d'Ascq, France

Abstract—The extraction of quantitative information from Ground Penetrating Radar (GPR) data sets (radargrams) to detect and map underground utility pipelines is a challenging task. This study proposes several algorithms included in the main stages of a data processing chain associated with radargrams. It comprises pre-processing, hyperbola enhancing, hyperbola detection and localization, and parameter extraction. Additional parameters related to the GPR system such as the frequency band and the polarization bring data sets additional information that need to be exploited. Presently, the algorithms have been applied step by step on synthetic and experimental data. The results help to guide future developments in signal processing for quantitative parameter estimation.

Keywords—buried pipes, dielectric characterization, GPR, ICA, PCA, polarization, template matching, ultra-wide band.

1. Introduction

The rapid growth of buried utility networks of different types (i.e., fiber optics, telecommunication lines, electrical cables, water and gas pipes, district heating network) to fulfill services in the urban landscape need mapping of the underground to update urban cadastral databases, to contribute to space saving and a wise use of land resources when planning for new networks [1], [2]. Among non-destructive techniques, Ground Penetrating Radar (GPR) appears the most suitable device for locating and identifying objects made of a dielectric or conductive solid or hollow material buried within 1.5 m of the ground surface. A few international organizations promote recommendations to properly use GPR in utility engineering, such as the ASCE (CI/ASCE 38-02) and the ASTM (ASTM D6432-99) international (2011) in North America [3], [4], EuroGPR in Europe, and the CEI (306-08) in Italy [5]. Moreover, European COST (Cooperation in Science and Technology) Action TU1208 "Civil engineering applications of Ground Penetrating Radar" [6] promotes applied research and emits recommendations on GPR applied to civil engineering. One of its working groups is devoted to utility detection and mapping by GPR. In this work, ground-coupled radar systems have been employed because, compared to air-coupled

systems, the energy transfer of electromagnetic (EM) waves in the sub-surface and the penetration depth is increased. The moving of the GPR system along a linear path gives a distance-time graph called radargram (also called B-scan, a set of traces or A-scans), contains diffraction hyperbolas associated with buried dielectric discontinuities due to the presence of circular-section objects that induce reflection with the emitted signal. It must be underlined that only small circular-section objects generate hyperbolas in radargrams. In the present work, the surveyed scenarios hosted long pipes compared to the GPR central wavelength, with a diameter shorter than 100 mm. The interpretation of hyperbolas, that represent the signatures of the objects, is a challenging task because several physical and geometrical parameters (i.e., soil heterogeneity and variations, soil absorption, antenna coupling, antenna-target coupling, target proximity) contribute to blur the information or mitigate contrasts. Consequently, several processing methods are developed in the literature because of the non-uniqueness of the solution, and they are generally adapted to a field situation because the EM problem is tedious.
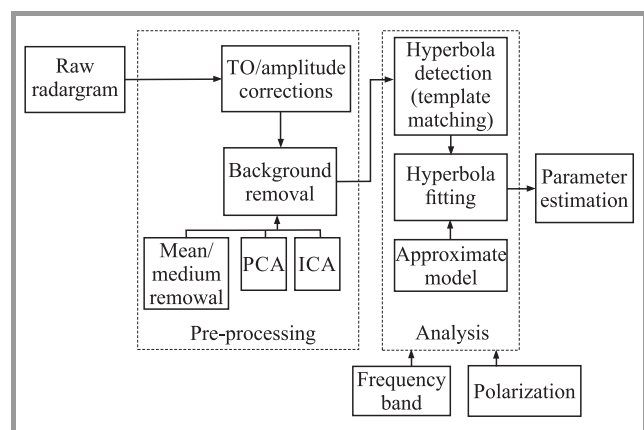
Fig. 1. The several stages of the processing link for the quantitative analysis of a radargram associated with buried targets.

Thus, the data processing chain synthesized in Fig. 1 has been developed to particularly analyze theoretical and experimental data sets. The processing is made of two main

stages associated with pre-processing and analysis algorithms to extract quantitative information from the hyperbola signatures associated with detectable targets. The information is relative to the evaluation of the target location, depth, size and dielectric nature. Attention has been paid to clutter removal in the case of a shallow target depth and a low permittivity contrast between the target and the soil, and on hyperbola detection in the presence of multiple targets and a heterogeneous soil. Parameters such as the frequency band and polarization have been considered as they can bring more detailed information from radargrams with additional processing.
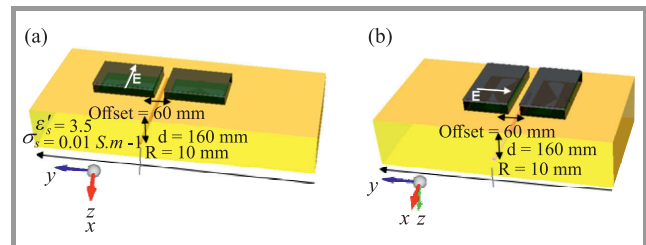
2. The Measurements

The acquisition of radargrams was achieved using several GPR systems operating at frequencies ranging from 300 MHz to 1.5 GHz either in the time domain (GSSI SIR 3000, UtilityScan DF) or in the frequency domain (a SFCW GPR conceived in our laboratory). The SFCW (Step Frequency Continuous Wave) GPR is a laboratory system and cannot make at present fast measurements such as commercial systems. The sampling distance step was around 5 mm using commercial GPRs, and defined to 40 mm using the laboratory system. The two types of GPR systems were used to make comparisons, and results are presented in this paper at 900 MHz.

The SIR 3000 was equipped with one of the three antenna pairs operating at the central frequencies 500, 900 MHz and 1600 MHz, and the UtilityScan DF worked with a double frequency antenna system at 300 MHz and 800 MHz that is supposed to collect data at the same sampling distance because the switching device is rapid enough during walking. Only the transverse magnetic (TM) polarization with respect to the pipe axes has been considered because these GPR systems only consider this polarization.

The SFCW GPR is made of a pair of shielded bowtie slot antennas (A4 sheet size designed on a single-sided FR4 substrate) has been used with a portable Anritsu MS 2026B Vector Network Analyzer (VNA) in the frequency band [0.05; 4 GHz] (1601 samples, intermediate frequency 500 Hz, 2 m coaxial N cables) [7]. A full two ports calibration was made that allows to measure the four $S_{ij}(f)$ coefficients.

Measurements in the frequency domain and in an Ultra-Wide Band (UWB) offer the advantages of selecting if needed a frequency band and shaping a Gaussian pulse associated with this frequency band, and using easily the polarization diversity. The complex transmission coefficient $\tilde{S}_{21}(f)$ measured at each scanning distance with a step of 40 mm is transformed in the time domain using an inverse Fourier transform to obtain a radargram. An apodization of each frequency curve has been made to smoothly extend the measured band from 4 to 9 GHz, and then perform the product with the spectrum of the excitation signal (the first derivative of the Gaussian function with duration 0.5 ns) used in Finite Difference Time Domain (FDTD) simula-

tions. Two symmetric antenna configurations have been defined such as presented in Fig. 2 to consider two perpendicular polarizations: the end-fire configuration (or TM mode) with both antennas aligned along their larger dimension, and the broadside configuration or transverse electric mode (TE) with both antennas facing each other symmetrically along their larger dimension. The laboratory made SFCW GPR has been easily modeled as all the antenna characteristics and geometry are known.

Fig. 2. The two polarization configurations of the SFCW GPR: (a) TM polarization (end-fire), and (b) TE polarization (broadside).

The measurements have been performed in two different test sites:

- a large sandy box belonging to the public square Perichaux in Paris 15th district [8]. The sand was wet on the surface, and its permittivity has been estimated to a real permittivity ε_s'. A couple of canonical objects (pipe or blade) dielectric or conductive with lateral dimension less than 25 mm have been buried at a depth close to 160 mm. They have been separated by a distance around 750 mm;

- an utility zone built under the urban test bed Sense-City in Marne-La-Vallée (France) [9]. This zone is under the 10 m wide traffic circle with lawn at the center. The underground contains two series of 5 trenches parallel to each other with size 0.3×4 m and separated by an offset of 40 cm (Fig. 3). Each target has been positioned in the middle of a trench at a depth ranging from 14.5 to 64.5 cm from the 4.4 cm thick asphalt surface. The two series of trenches include dielectric or conductive pipes (diameter 63 mm) and dielectric or conductive blades (thickness between 1 and 3 cm). The pipes have been filled with air or water. The trenches were filled with a soil of fine elements (without gravels) commonly used in urban underground filling. The near sub-surface is made of four main layers corresponding to asphalt (layer 1), aggregate cement (layer 2), a natural soil (layer 3), and a quite wet natural soil (layer 4) located under a geotextile (a thin discontinuity). The real permittivities of the soil layers have been estimated as: asphalt (layer 1) $\varepsilon_1' = 4.5$ (corresponding to a time propagation of 0.64 ns), aggregate cement (layer 2) $\varepsilon_2' = 7.7$ (1.4 ns), natural soil $\varepsilon_3' = 34$ (8.5 ns).
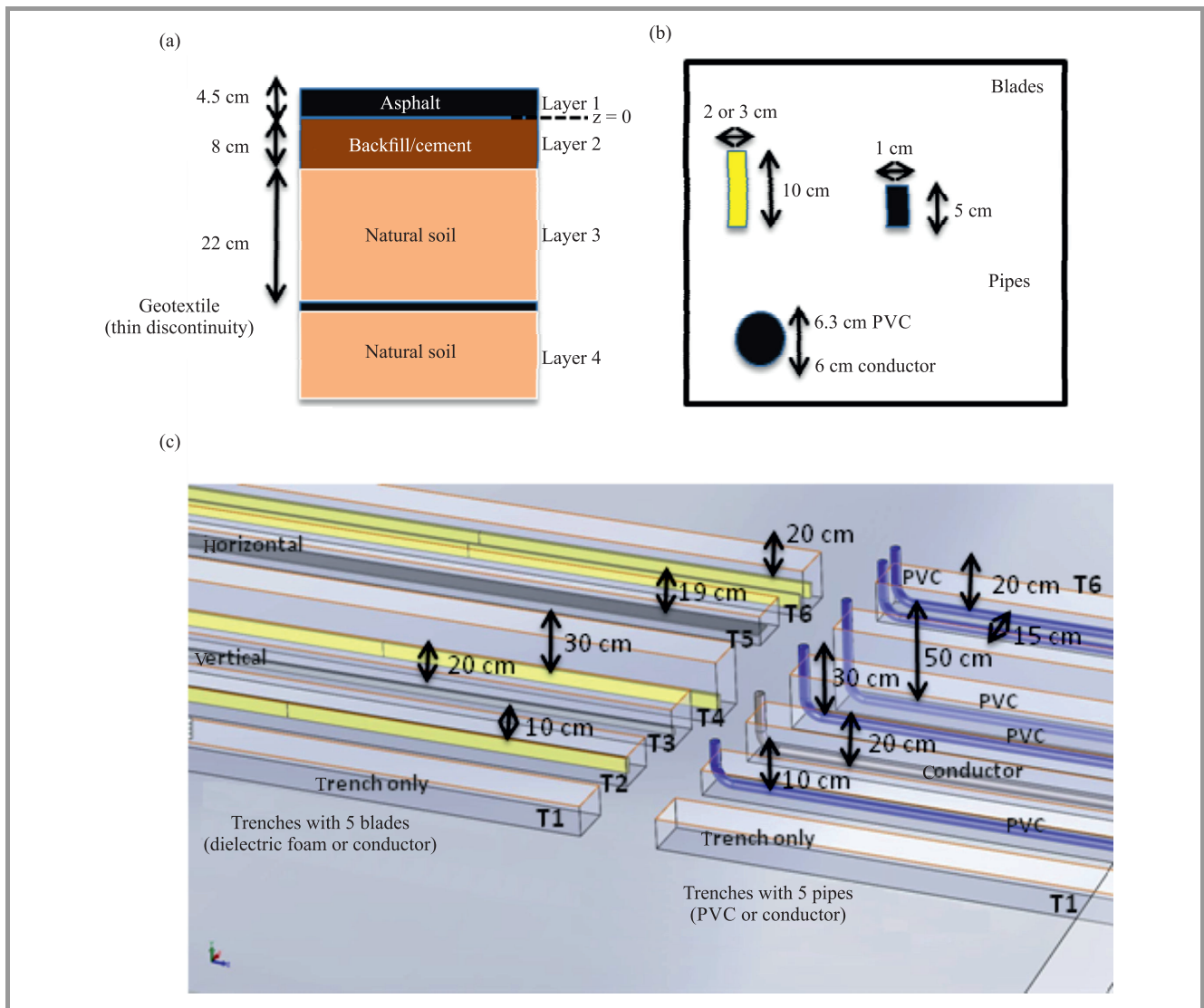
Fig. 3. Structure of the sub-surface in Sense-City site: (a) the main multilayers of the soil, (b) geometries of the buried objects, (c) distribution of the buried pipes and strips.

3. Modeling

A 3D full-wave FDTD modeling using the commercial software EMPIRE has enabled to analyze EM phenomena as a function of parameters such as soil and pipe permittivity, pipe depth, and antenna polarization and make comparison with experimental results [8], [9]. The complete SFCW GPR made of a pair of shielded bowtie slot antennas (frequency band [0.46 ; 4] GHz) and designed on a FR4 substrate (real relative permittivity $\varepsilon' = 4.4$ and thickness $e = 1.6$ mm) has been modeled in the presence of a semi-infinite soil [7]. Each antenna is enclosed in a shielded conductive rectangular box filled with a three-layered absorbing foam with total dimensions $362 \times 23 \times 67.5$ mm. The offset between antennas in simulations and experiments has been fixed to 60 mm, and the elevation h_s above the soil is $h_s = 40$ mm. The soil electrical parameters (ε'_s, σ_s) are assumed constant across the frequency range. The GPR

system is moved linearly on the soil surface with a step $\Delta y = 40$ mm (see Fig. 2) to acquire a radargram. The excitation signal generally used in the simulations is the first derivative of the Gaussian function with a duration equal to 0.5 ns (99% of the total energy) and begins at 0.33 ns.

4. Pre-Processing

The pre-processing step aims to prepare radargrams before analyzing the hyperbola signatures of sub-surface targets to extract quantitative information. After time scaling and amplitude scaling (if needed), the main step consists of clutter removal to reduce the reflection component of the ground surface, direct waves between antennas, coupling between the antenna system, and a shallow target, as well as scattering induced by soil heterogeneities. The clutter often hides target signatures because its amplitude is far stronger and the arrival time is shorter. Several clut-
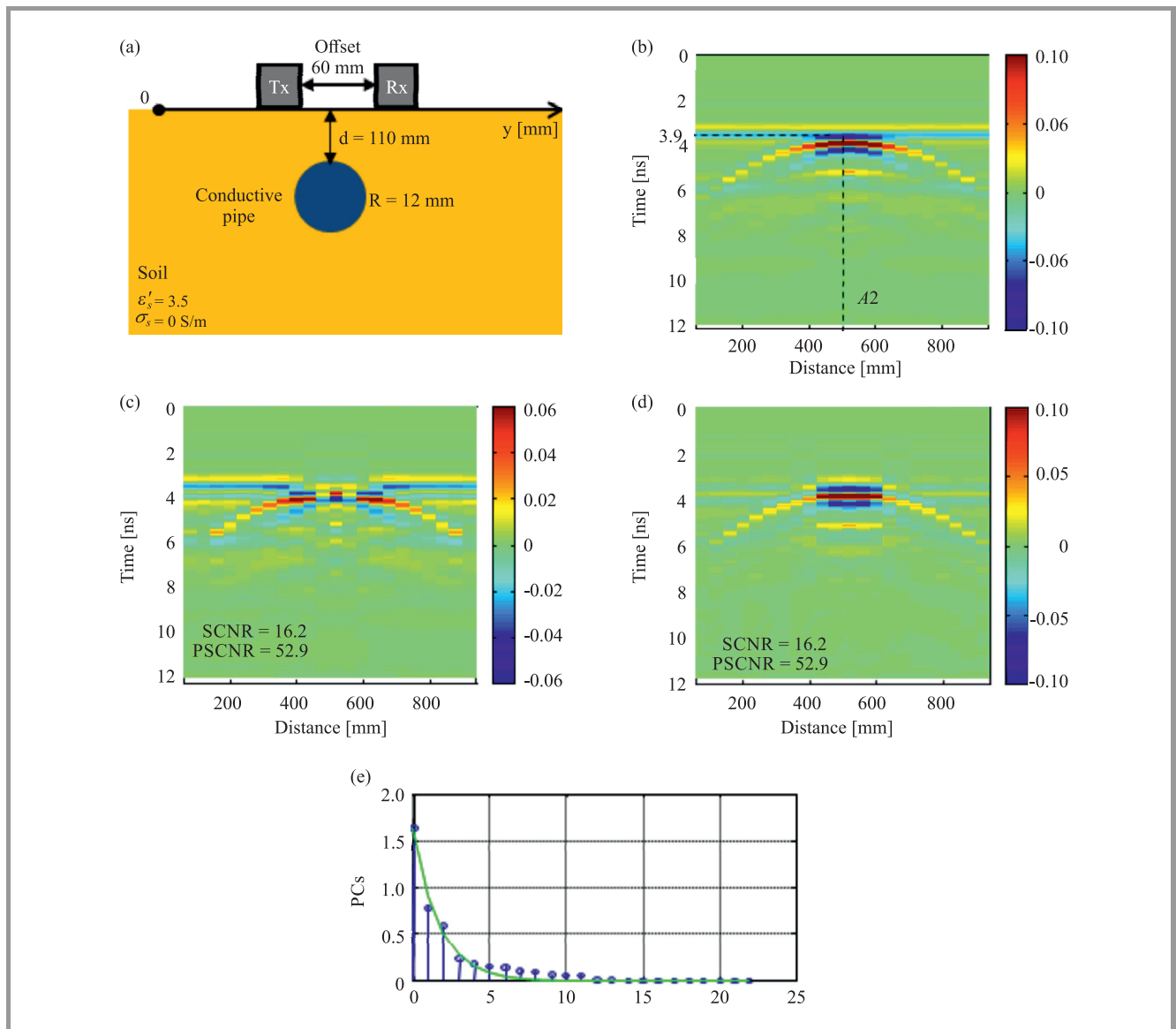
Fig. 4. Pre-processing applied on a synthetic radargram issued from FDTD simulations: (a) geometry of the soil structure including a buried conductive pipe probed by a pair of shielded bowtie slot antennas (TM polarization), (b) raw radargram, (c) radargram processed by PCA, (d) radargram processed by PCA modified, (e) distribution of the PCs. (See color pictures online at www.nit.eu/publications/journal-jtit)

ter reduction techniques were proposed in the literature and their development is an active research topic because the clutter is not uniform in a radargram, it is unsteady and it has different types [10].

The methods proposed appear initially suited to targets with small lateral dimension, overlapping hyperbola signatures, varying clutter along the scanning direction (a non-stationary signal), overlapped target and clutter signals in time and frequency [11]. The semi-automatic methods tested and compared are the mean or median subtraction [12], the Principal Component Analysis (PCA) that has been modified and the Independent Component Analysis (ICA) [13], [14]. Basically, PCA relies on the second order statistics to perform the correlation analysis, whereas ICA requires higher order statistics (4th moment). Both

methods have the advantage of requiring limited prior information on the data set.

4.1. PCA

The objective of PCA is to express the original data set in another domain by means of any appropriate linear transformation. PCA is a subspace projection method that is based on either on a Singular Value Decomposition (SVD) or an Eigenvalue Decomposition (EIG) to select relevant components in different subspace projections [13]. Each eigenvalue is related to a certain eigenvector (principal component), and the eigenvalues are ordered in the descending order. The main problem of this algorithm is how to choose the principal components where information is contained.

Every eigenvalue represents a specific amount of variance in measured data, and the objective is to obtain a new set of uncorrelated data.

From the initial data matrix $X(M \times N)$ (with M amplitude signals A-scans and N time samples; $M < N$) the first k_1 components associated with the clutter and the last $N - (k_2 + 1)$ components associated with noise are generally removed thus giving a new data set:

$$X_0 = \sum_{k_1+1}^{k_2} u_i s_i v_i^T, \tag{1}$$

where u_i and v_i are eigenvectors.

However, difficulties can be met when the clutter does not have the highest energy within the radargram in the case of a target that induces a strong Radar Cross Section (RCS) such as a conductive or a water filled target, or a high contrasted dielectric target with the surrounding soil [15]. Moreover, it appears difficult to select the index rank defined by k_1 and k_2 when there are multiple targets and reflections. Thus, we have proposed to modify PCA and apply the algorithm on a radargram with a reduced time scale (investigation depth) and a splitting of the scanning scale associated with a selected hyperbola signature into 3 parts. This splitting is based on an energy criterion that compares energy in the apex area and energy in the arms. This splitting allows to analyze separately the area close to the apex and relative to the two arms as they show distinct amplitude value and interaction with the clutter. It particularly concerns target response with high energy that is comparable to the clutter energy.

An illustration of the application of PCA and PCA modified is presented in the case of a conductive pipe with radius $R = 12$ mm (see Fig. 4a), buried at a depth d of 110 mm (that is, $\frac{3}{4}$ of the wavelength λ_{soil} at 1 GHz) in a soil with real permittivity $\varepsilon' = 3.5$($\frac{d}{\lambda_{soil}} \approx 1$ at 1 GHz), that has been probed by the pair of bowtie slot antennas used in the SFCW GPR (TM polarization). The data set have been obtained from FDTD simulations (see Fig. 4b) and, in this case, there is overlapping between clutter and hyperbola response. The arrival times of the clutter and the target are 3.4 and 3.9 ns, respectively. The PCs selected represents 90% of the energy, the first one is removed as it represents the clutter component (see Fig. 4e). The processing by conventional PCA depicts visually bad performance in reducing the coupling between antennas as part of the horizontal component of antenna coupling remains (see Fig. 4c). PCA modified preserves the target hyperbola shape (see Fig. 4d) but it does not eliminate completely the clutter over the apex, and a clutter remains in the time interval 3…4 ns.

4.2. ICA

ICA has been recently proposed as an alternative to eigenvector decomposition in radar images as it has proved itself to be a very promising tool to better interpret non-Gaussian heterogeneous clutter. ICA that is based on higher order statistical moments, can recover statistical independent sources and the mixing mechanism without having any physical background of the latter [14], [15]. The ICA model associated with a data matrix X that is assumed random is generated from a source matrix S through a linear process:

$$X = SA, \tag{2}$$

where X is the matrix used in the PCA algorithm, $S = (s_1, s_2, \ldots, s_{N2})$ is the $M \times N_2$ source matrix ($N_2 \leq N$), and A is the $N_2 \times N$ mixing matrix.

ICA assumes that the components of the source matrix are statistically independent with respect to each other, that makes ICA more efficient in the separation process. Before processing with the calculations, X is transposed ($X^T = A^T S^T$) and then ICA uses centering, whitening, and dimensionality reduction algorithms as pre-processing steps to simplify and reduce the complexity of the problem. The FastICA algorithm (fixed-point algorithm first proposed by Hyvärinen and Oja [15]) available in a Matlab package leads to the determination of two sub-solutions: 1) the estimation of the mixing matrix A, and 2) the estimation of the source signals S. To measure the non-Gaussianity (the independence) of the sources a non-linear and non-quadratic function, also called a contrast function, is used that helps to optimize the performance of the ICA algorithm.

In this work, negentropy has been used to measure the non-normality or the degree of unstructuredness of a ran-
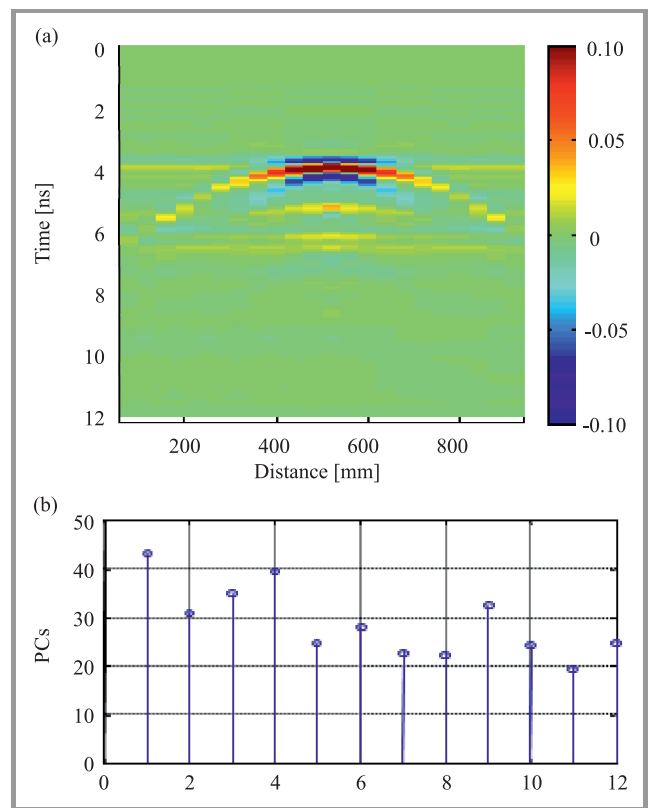
Fig. 5. Pre-processing applied on the synthetic radargram of Fig. 4a using ICA: (a) radargram processed by ICA, (b) Kurtosis criterion applied on ICs.

dom vector. Negentropy is non-negative and has a high value for non-Gaussian variables. In practice, negentropy is difficult to calculate and must be approximated, a solution proposed by Hyvärinen [15]. An iteration based on the fast-fixed-point technique that uses random vectors as starting values, allows to estimate step by step the several independent components (ICs). Afterwards, the main difficulty relies on the selection of ICs which contain information. The Kurtosis criterion applied to the ICs has served to define a threshold beyond which the ICs selected contain information.

As an illustration, the previous synthetic radargram associated with a conductive pipe buried at 110 mm depth has been analyzed by the ICA algorithm. The results in Fig. 5a show that the clutter is better reduced by ICA than by PCA modified, as there is no artifact above the apex, however a weak horizontal clutter remains visible. The Kurtosis criterion applied on ICs are plotted in Fig. 5b.

4.3. Comparison Criteria between Processing Algorithms

The visual comparison of the reconstructed images with reduced clutter provides a first qualitative assessment of the algorithms tested. Attention has to be paid to spatial coherency along the scanning direction in such a way that the visibility of a hyperbola must be strengthened as compared to that of the clutter. Thus, a few tools have been used to quantitatively compare the performance of the clutter reduction algorithms.

Signal to Clutter and Noise Ratio (SCNR)

The ratio is defined as the average energy of the reconstructed clutter-reduced image (represented by a matrix F of real values $f_{i,j}$ with $i = 1,\ldots,M$ and $j = 1,\ldots,N$) with the average energy contained in the "noise" image (named G with elements $g_{i,j}$ with $i = 1,\ldots,M$ and $j = 1,\ldots,N$), that is obtained by subtracting the raw image from the processed image such as [17]:

$$SCMR_{dB} = 10\log\frac{P_{\text{target}}}{P_{\text{clutter+noise}}} =$$

$$10\log\frac{\sum\limits_{i=1,N}\sum\limits_{j=1,M}|f_{i,j}|^2}{\sum\limits_{i=1,N}\sum\limits_{j=1,M}\left(|g_{i,j}-f_{i,j}|\right)^2}. \quad (3)$$

In practice, we don't have the opportunity to acquire an image without clutter and noise, that limits the interpretation of the SCNR.

Peak Signal to Noise and Clutter Ratio (PSCNR)

Another parameter is the PSCNR that uses presently the peak signal instead of its average energy such as [18]:

$$PSCNR_{dB} = 10\log\frac{NM\max(|f|)^2}{\sum\limits_{i=1,N}\sum\limits_{j=1,M}\left(|g_{i,j}-f_{i,j}|\right)^2}. \quad (4)$$

Again, this ratio has some limits.

Receiver Operating Characteristics (ROC) curves

ROC curves appear more appropriate to compare quantitatively clutter reduction algorithms. The ROC curve is simply a graph of detection versus false alarm probabilities parameterized by a threshold S [19]. From a ROC curve (1×1 square region), fundamental information extracted from metrics (in general area-under-roc-curve AUC) is derived. A ROC curve is calculated by comparing pixel to pixel two binary images, a reference image containing only the desired information and the image processed by an algorithm using a varying threshold S in the amplitude range of the raw image. The reference image defined corresponds to a skeleton of the hyperbola response including the first positive and negative amplitudes. The threshold applied on the raw image serves to compute a binary image to make a comparison with the reference image.

As an illustration, these 3 parameters have been evaluated for the radargrams of Figs. 4c-d (PCA), and also Fig. 5a (ICA). The largest SCNR and PSCNR values belong to the median subtraction technique (14 and 72.9 respectively), and afterwards the better performances are attributed to PCA modified and ICA. The ROC curves visualized in Fig. 6 show that the curve associated with PCA is slightly below the one issued from the raw data. The curve associated with PCA modified appears higher than the one belonging to ICA.
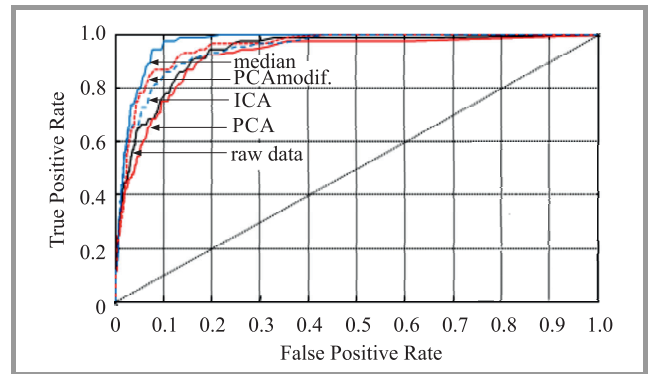

Fig. 6. ROC curves associated with the different pre-processing algorithms (median subtraction, PCA, ICA) applied on the raw radargram of Fig. 4a.

As a conclusion, the median subtraction technique appears a robust method for clutter reduction but it has its limits in the case of the overlapping of clutter and hyperbola response. As a conclusion, the median subtraction technique appears a robust method for clutter reduction that has however its limits, in the case of overlapping between clutter and hyperbola response. The PCA algorithm shows better performance in the case of the relative pipe diameter $\frac{d}{\lambda_{soil}} \leq 0.7$ and when there exists partial overlapping between clutter and hyperbola. When the hyperbola response has a high amplitude, PCA modified improves the removing of the clutter The ICA algorithm shows better performance in the case $\frac{d}{\lambda_{soil}} \geq 1$ and without overlapping.

5. Hyperbola Detection and Fitting

In practice, the difficulties encountered in hyperbola detection in a radargram rely on the heterogeneity of the background and a weak contrast between the hyperbola and the background. Thus, we have proposed a semi-automatic detection algorithm based on template matching that doesn't require a training period [20]. It is applied after a pre-processing stage and is made of a series of steps:

- A template containing a typical hyperbola signature is built, that is issued from synthetic or experimental data. The template does not need to be fully similar to the responses to be detected. When the hyperbolas in the radargram cannot be detected by a single pattern, several patterns can be used in sequence. Each defined template includes a small portion of a hyperbola signature near the apex. It has to be scaled in time, distance and amplitude to the radargram that has to be analyzed.

- An amplitude distance map is calculated by translating on the radargram the predefined template at every possible positions. A mean amplitude distance is evaluated according to the L_1 norm.

  Assuming an observed image G with elements $g_{i,j}$ or $g(i,j)$ and size $M \times N$, and a template image T with elements $t_{i,j}$ or $t(i,j)$, both images with scaled amplitudes, we define the L_1 norm distance map E between g and t by the following relation [18]:

$$E(m,n) = \sum_{i=1}^{M} \sum_{j=1}^{N} |t(i,j) - g(i-m,j-n)|. \quad (5)$$

  The summation is evaluated at all pixels (i,j) of the template t that is translated to all possible positions (m,n) along the observed image g. The position (m,n) at which the smallest value $E(m,n)$ is obtained corresponds to the best match between the template t and the corresponding sub-image in g.

- A threshold value increased progressively by the user allows to select a number of minima on the distance map that corresponds to a high level of similarly with the template.

- At the selected positions, the hyperbola points close to the apex that correspond to a maximum or a minimum amplitude have to be extracted. Because higher order reflections in a hyperbola pattern may produce a stronger amplitude as compared to the amplitude of the first reflection, an user interaction is necessary (semi-automatic) to select a hyperbola curve either on the upper or on the lower half zone of the central template position. Starting from the vertical symmetry axis of the template, close points belonging to the hyperbola curve on the left and on the right legs (usually 3 or 4 points) are extracted step-by-step. Because the number of curve points extracted is limited,

a polynomial fitting of the second order was made to refine the estimate of the abscissa y_0 at the apex.

- The estimation of the several parameters describing the hyperbola curve such as d, R, V, the target depth and radius and the celerity in the soil, respectively (see Fig. 7) is obtained using a hyperbola fitting of the curve points to the classical analytical relation according to the LS criterion. Parameters SR and y_0, representing respectively the center-to-center distance between antennas and the horizontal coordinate of the object, are a priori known. The position of the radar referenced at the center of the system is located using coordinate y_k.
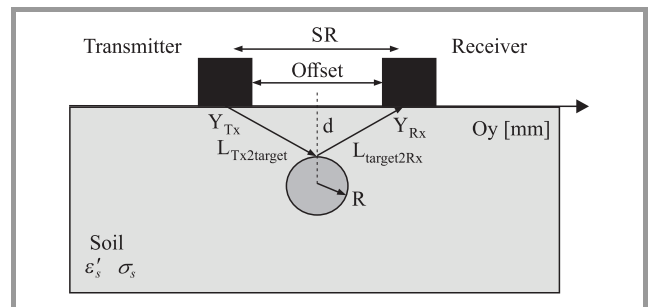
Fig. 7. Cross-view of a pipe buried in a soil and parameters associated with the ray-path theory.

The equations associated with the travel-time write as follows [21]:

$$\begin{cases} y_T = y_k - \frac{SR}{2} \\ y_R = y_k + \frac{SR}{2} \end{cases}, \quad (6)$$

$$\begin{cases} T_{Tx2target} = \left[(y_0 - y_T)^2 + (d+R)^2 \right]^{\frac{1}{2}} - R \\ T_{target2Rx} = \left[(y_0 - y_R)^2 + (d+R)^2 \right]^{\frac{1}{2}} - R \end{cases}, \quad (7)$$

where $T_{Tx2target}$ and $T_{target2Rx}$ represents the ray paths between the transmitter or the receiver to the buried object. The velocity v of the medium depends on the real relative permittivity ε_s' such as:

$$v = \frac{c}{\sqrt{\varepsilon_s'}}; \quad (\varepsilon_s'' \ll \varepsilon_s'). \quad (8)$$

The generalized hyperbola equation (y_k, t_k) including the antenna offset is expressed by:

$$t_k = \frac{T_{Tx2target} + T_{target2Rx}}{v}. \quad (9)$$

Thus, (y_0, t_0) represents the position in coordinate and time of the maximum of the hyperbola. It must be underlined that the hyperbola depends on five parameters (SR, y_0, d, R, v). In general, the pipe radius cannot be properly estimated, as it appears smaller than the radiating aperture of one antenna (on the order of the first Fresnel zone in the soil). The minimum of the sum of squares of the

distances between theoretical $y_{k,theo}$ and measured $y_{k,meas}$ data has been solved using the Gauss-Newton method as follows:

$$F = \sum_{k=1}^{K} \left(y_{k,theo} - y_{k,meas}\right)^2. \qquad (10)$$

Some constraints relative to the parameter ranges have been added in the algorithm to better drive the solution.
The Hessian matrix in 2D and particularly its eigenvalues can be used to characterize the uncertainties on the estimated parameters (d, v).
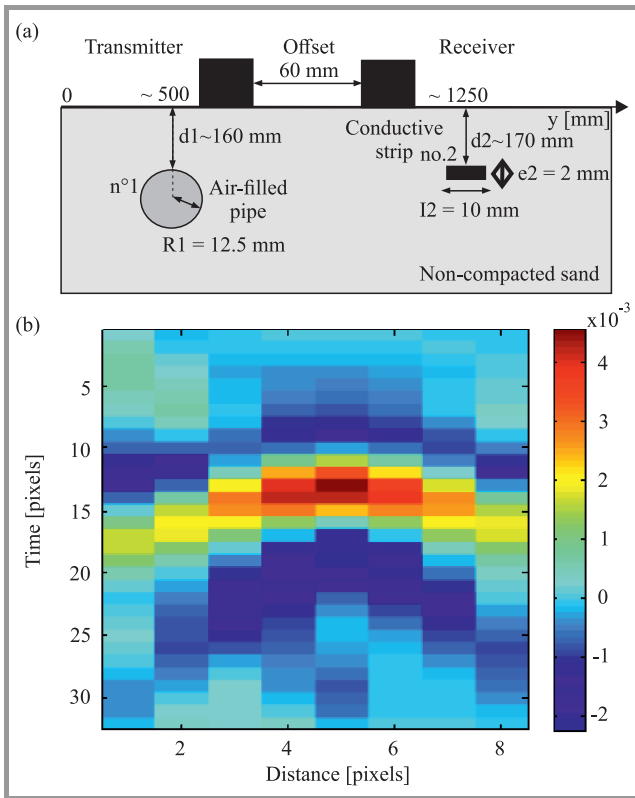
5.1. Application to Experimental Data

In the sandy box of the public square Perichaux, Paris 15th, the dielectric permittivity of the soil at this moment was estimated to 3.5 from common mid-point (CMP) measurements. Two targets, a 25 mm diameter dielectric pipe filled with air and a thin horizontal 10 mm width (2 mm thick) conductive strip were buried at the depths estimated to 160 and 170 mm respectively such as presented in Fig. 8a. Both objects are separated by a 750 mm distance. The scanning by the SFCW GPR has been performed in both end-fire and broadside polarizations (see Figs. 2a-b).
A synthetic template was computed from 3D FDTD simulations (software Empire) considering the detailed bowtie

Fig. 8. Geometry of the sandy box with two buried objects, an air-filled pipe and a horizontal conductive strip (a), synthetic template issued from FDTD simulations involving a pair of bowtie slot antennas (end-fire configuration) moved on a soil ($\varepsilon'_s = 3.5$, $\sigma_s = 0.01$ Sm^{-1}, $h_s = 10$ mm) with a buried conductive pipe ($R = 12.5$ mm) (b).

slot antenna geometry (Section 3) and 32 mm radius conductive pipe buried in a soil with a real relative permittivity $\varepsilon' = 3.5$ ($\sigma = 10^{-2}$ Sm^{-1}). The end-fire configuration has been considered. In this template visualized in Fig. 8b, it was wise to define a compact hyperbola signature with reduced multiple reflections to further detect different hyperbolas in a radargram in both polarizations. The template was scaled in time and amplitude to match the experimental time step $\Delta t = 5.56 \cdot 10^{-2}$ ns (step distance 40 mm used in the experiments), and the amplitude range of the experimental radargrams. Signed amplitudes have been used here to not deteriorate the image quality.

The template matching algorithm proposed has been performed on experimental radargrams considering the two orthogonal polarizations after application of the median subtraction technique. Results are presented for the geometry relative to Fig. 8a, and the time range has been defined to 5 ns. That leads to the L1 norm distance maps (also absolute distance maps) visualized in Figs. 9a and 9c. The position (m, n) at which the smallest value $E(m, n)$ is obtained corresponds to the best match between the template t and the corresponding sub-image in g. A threshold value allows to select a limited number of local minima corresponding to distances less than the threshold where the template is well matched with enough amplitude (visualized by "+" signs on the radargrams). The maximum threshold values leading to the detection of the most significant hyperbolas are 0.162 and 0.195 in the broadside and end-fire configurations, respectively (see Figs. 9b and 9d). Higher values give additional detections (false alarms) that don't correspond to buried objects but to background heterogeneities. In the end-fire polarization, the distance map cannot permit to detect the air-filled pipe.

The results of the parameter evaluation from the least-square (LS) fitting of each hyperbola detected are collected in Tables 1 and 2. In general, the positions y_0 of the objects appear correctly evaluated. Concerning the real permittivity value of the soil, the broadside configuration gives higher estimates as compared to the end-fire configuration, and consequently the target depths appear more important.
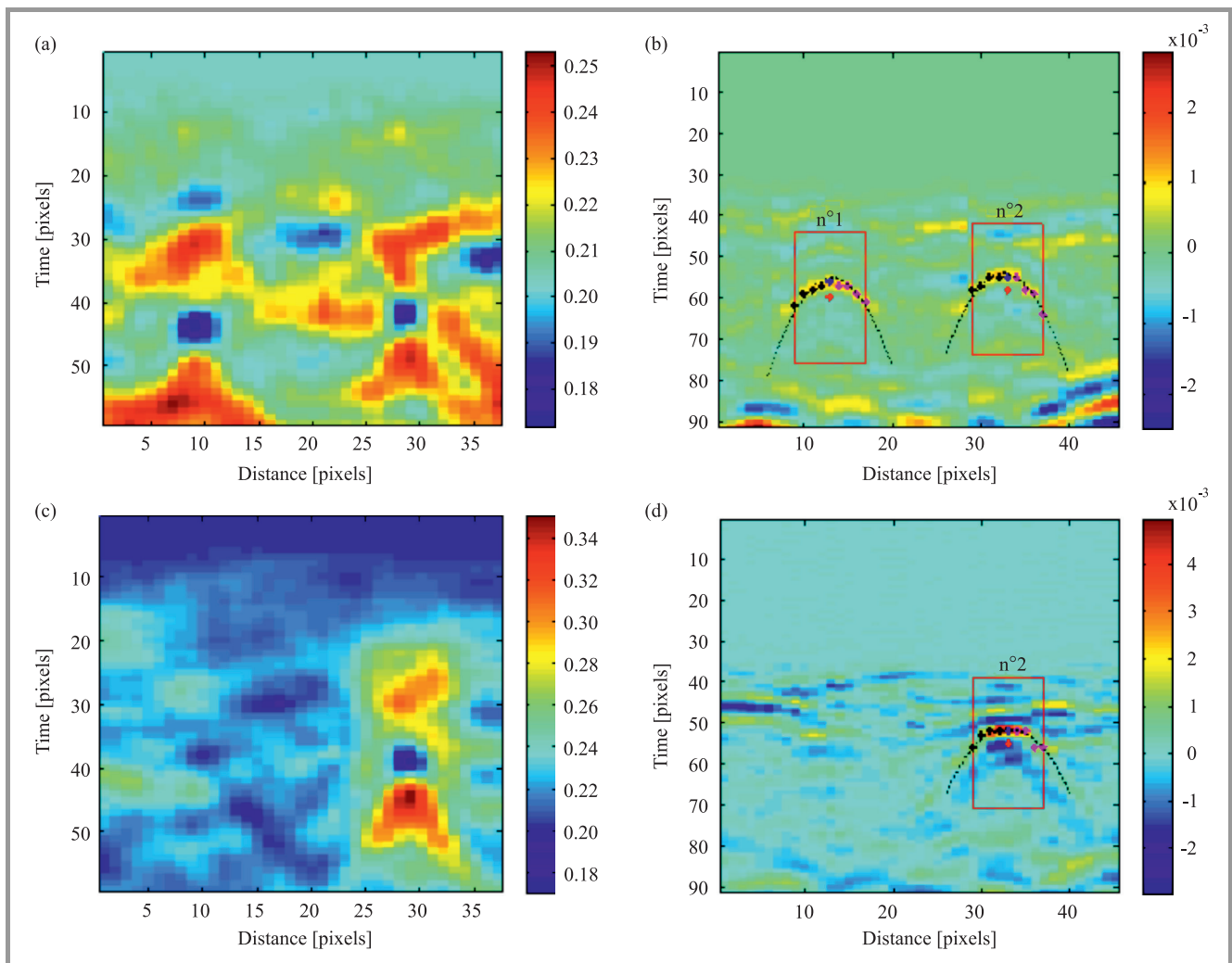
Table 1
Parameter estimation in the broadside configuration for the experimental B-scan of Fig. 9b (max. is associated with the detection of the positive amplitude)

Object	d [mm]	R [mm]	ε'	T_0 [ns]	y_0 [mm]	*fval*
Pipe no. 1 (max)	200	60	3.6	3.08	458.9	$2.68 \cdot 10^{-2}$
Pipe no. 1 (true values)	~160	12.5	3.5–4		~500	
Strip no. 2 (max.)	200	60	3.47	2.99	1232	$2.23 \cdot 10^{-2}$
Strip no. 2 (true values)	~170	5	3.5–4		~1200	

Fig. 9. Analysis of experimental radargrams (in pixels, time range 5 ns) associated with an air-filled pipe and a horizontal conductive strip using the template matching algorithm: (a, b) in the broadside and (c, d) end-fire configurations; (a, c) time-distance maps, (b, d) positions of the image template in the original B-scans and hyperbola fitting.

Table 2
Parameter estimation in the end-fire configuration
for the experimental B-scan of Fig. 9d

Object	d [mm]	R [mm]	ε'	T_0 [ns]	y_0 [mm]	$fval$
Strip no. 2 (max.)	165.6	25.5	2.65	2.82	1250	$1.85 \cdot 10^{-2}$

From Table 3 that collects eigenvalues of the Hessian matrix at the estimates of the depth and the velocity (d, v), we remark that the depths of the pipe and the strip have been both estimated to 200 mm, and more important than those a priori known (see Fig. 8a). In the case of a soil having weak permittivity variations, an additional step would be to find an optimum permittivity value issued from the several estimates. We remark that the objective function $fval$ associated with the LS fitting is presently higher for experimental data than for synthetic data of the order of 10^{-2}, because the image quality appears lower.

To detect the air-filled pipe in Fig. 9d, joint information from radargrams issued from the parallel and end-fire configurations (polarization diversity) could be used. Thus, the mean distance map was calculated from the individual distance maps in the broadside and end-fire configurations (see Figs. 9a and 9c) leading to the result visualized

Table 3
Eigenvalues of the Hessian matrix at the estimates
of the depth d and velocity v from hyperbola fitting
in both polarizations

Configuration	Eigenvalues for $(d; v)$
Parallel Pipe no. 1	$(4.1 \cdot 10^{-5}; 2.1 \cdot 10^{-2})$
Strip no. 2	$(4.2 \cdot 10^{-5}; 1.9 \cdot 10^{-2})$
End-fire Strip no. 2	$(4.0 \cdot 10^{-5}; 9.6 \cdot 10^{-3})$

in Fig. 10a. A weak threshold (0.18) situated between the previous ones (0.162 and 0.195) associated with both polarizations (see Fig. 10b), has allowed here to detect both hyperbolas without false alarms in the end-fire configuration, that was not possible when this polarization was only considered. Further insight into the solutions of the parameters estimated in the fitting has been gained by calculating the Hessian matrix H at the stationary point to evaluate its nature using its eigenvalues.
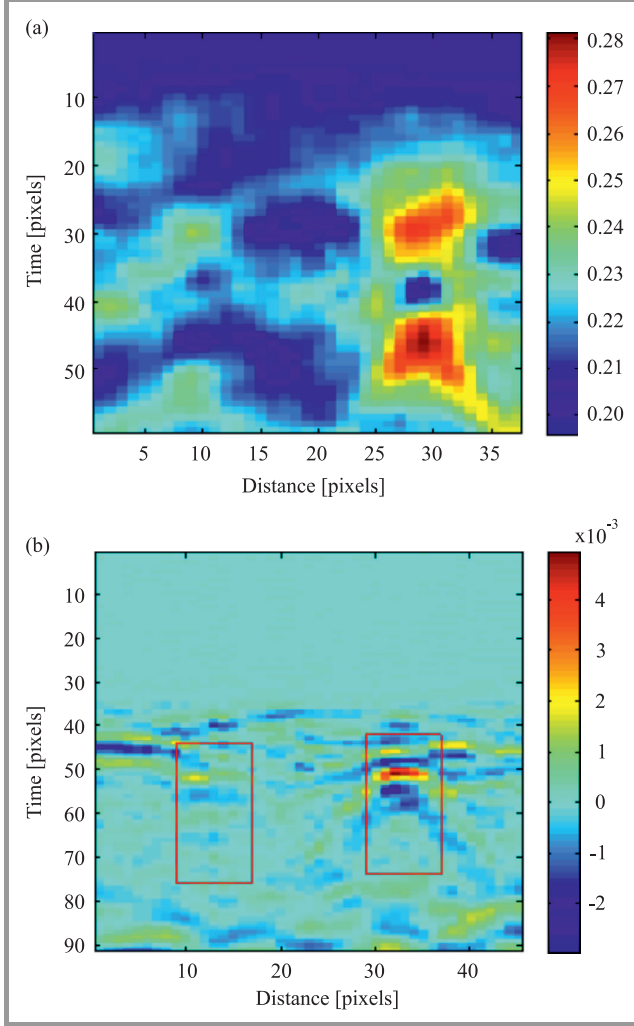
Fig. 10. Mean distance map in the parallel and end-fire configurations associated with an air-filled pipe and a horizontal conductive strip (Fig. 8a) (a) and positions of the image template in the experimental B-scan in the end-fire configuration (b).

The rate of convergence and sensitivity to round-off errors is given by the condition number of matrix H, that is the ratio of its largest to its smallest eigenvalues. In the present examples, fixing the pipe radius that cannot be evaluated properly has led to a decrease of the condition number. The eigenvalues associated with the several fittings in both polarizations are collected in Table 3. In general, the condition number is high, that means that correlation may exist between the two parameters and thus the convergence of the estimation algorithm appears slow.

6. Polarization Diversity

As a target can depolarize the incident field, thus giving a scattering response depending and the orientation of the antennas relative to the target, it is worth to exploit the polarization to characterize the target, and its orientation, particularly in the case of a long target such as a cylinder.

6.1. Analytical Modeling

The influence of the dielectric characteristics of a canonical target such as a cylinder (index d) illuminated by a electromagnetic (EM) polarized plane wave is described by the scattering theory [16], [23]. The backscattered fields depend strongly on the orientation of the cylinder relative to the antennas, its dielectric properties, and the radius of the cylinder compared to the wavelength λ_0 in the surrounding medium (index 0). The medium is usually air, but can be a dielectric material such as considered here. The scattering properties of a long cylinder has been described by the Bessel and Hankel functions. The scattering width (SW) is the equivalent area proportional to the apparent size of the target (in the far-field zone) that scatters the field relative to the incident power in each direction ϕ. Four expressions associated with SW consider both TE and TM polarizations and the case of a dielectric or a conductive cylinder as follows [23]:

TE polarization

Conductive cylinder:

$$SW = \frac{2\lambda_0}{\pi}\left|\sum_{n=0}^{+\infty}\varepsilon_n\frac{J_n'(\beta_0 R)}{H^{(2)'}(\beta_0 R)}\cos(n\phi)\right|^2 \quad (11)$$

Dielectric cylinder:

$$SW = \frac{2\lambda_0}{\pi}\left|\sum_{n=0}^{+\infty}\varepsilon_n \times\right.$$

$$\left.\times\frac{\eta_d J_n'(\beta_d R)J_n(\beta_0 R)-\eta_0 J_n'(\beta_0 R)J_n(\beta_d R)}{\eta_0 J_n(\beta_d R)H_n^{(2)'}(\beta_0 R)-\eta_d J_n'(\beta_d R)H_n^{(2)}(\beta_0 R)}\right| \quad (12)$$
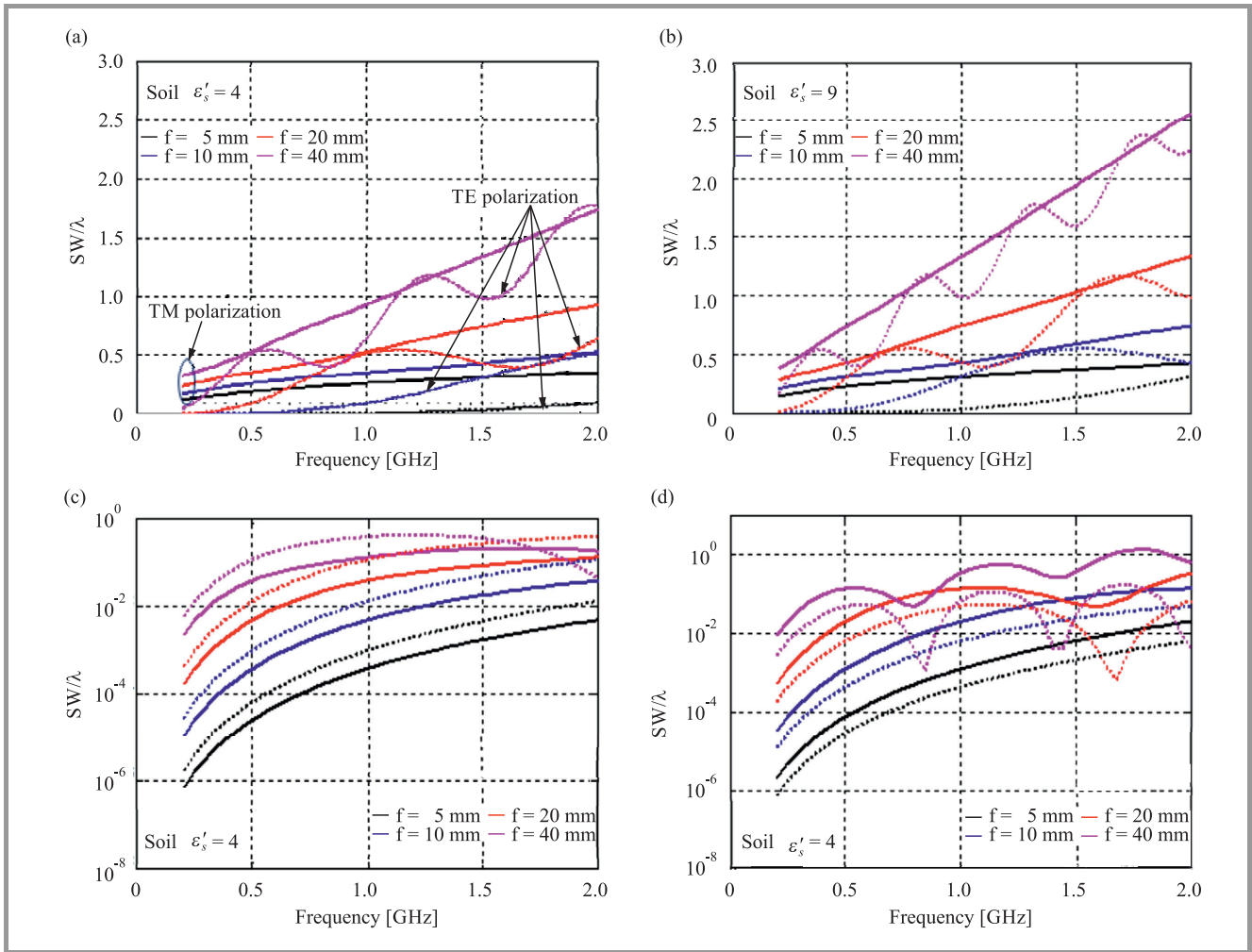
TM polarization

Conductive cylinder:

$$SW = \frac{2\lambda_0}{\pi}\left|\sum_{n=0}^{+\infty}\varepsilon_n\frac{J_n(\beta_0 R)}{H^{(2)}(\beta_0 R)}\cos(n\phi)\right|^2 \quad (13)$$

Dielectric cylinder

$$SW = \frac{2\lambda_0}{\pi}\left|\sum_{n=0}^{+\infty}\varepsilon_n \times\right.$$

$$\left.\times\frac{\eta_0 J_n'(\beta_d R)J_n(\beta_0 R)-\eta_d J_n'(\beta_0 R)J_n(\beta_d R)}{\eta_d J_n(\beta_d R)H_n^{(2)'}(\beta_0 R)-\eta_0 J_n'(\beta_d R)H_n^{(2)}(\beta_0 R)}\right| \quad (14)$$

Fig. 11. Influence of the pipe radius and the real permittivity on the scattering width (SW) as a function of the frequency in the case of (a, b) a conductive and (c, d) a dielectric air-filled cylinder ($\varepsilon' = 1$ in "c" and $\varepsilon' = 9$ in "d").

where β_0 and β_d are the wavenumbers, associated with the surrounding medium and the dielectric cylinder; they can be complex. $J_n(.)$ is the Bessel function of the first kind of order n, $J_n'(.)$ is its derivative, $H_n^{(2)}(.)$ is the Hankel function of the second kind of order n, and $H_n^{(2)'}(.)$ is its derivative.

The simplified modeling of the GPR system considers that the antennas, linearly polarized, are closely spaced. This results to a backscattering angle corresponding to $\phi = 180°$. Before analyzing theoretical and experimental B-scans, we have studied the scattered responses SW in a very large frequency band [0.2 ; 2] GHz considering several pipe radii and two dielectric contrasts between the pipe and the surrounding medium. The results collected in Figs. 11a-b, in the case of a conductive cylinder with a radius in the range 5...40 mm, highlight that the TM polarization (electric field parallel to pipe axis) generally leads to a higher backscattering response whatever the value of the real permittivity value (ε_0' or $\varepsilon_0'' = 0$) of the surrounding medium. We remark that the TE component oscillates under the TM component for a radius higher than 5 mm. In general, the

scattering width increases with the pipe radius and when the radius becomes large as compared to the wavelength of the medium; at 2 GHz, $\lambda_0 = 75$ mm for $\varepsilon_0' = 4$, and $\lambda_0 = 50$ mm for $\varepsilon_0' = 9$. Consequently, the TM polarization is preferred to detect a conductive pipe independently of its radius and the wavelength of the surrounding medium. Figure 11c-d associated with a dielectric cylinder show that the TE polarization gives a higher SW when the radius value is inferior to the wavelength and when the permittivity of the pipe (here filled with air) is less than the medium permittivity ($\varepsilon_0' = 4$). When the pipe is more dielectric ($\varepsilon_0' = 9$) than the medium, we remark that the TM polarization is in general higher than the TE polarization. We also observe that for a radius greater than 10 mm, TE and TM components oscillate and they can meet each other.

6.2. In-situ Measurements

The measurement results presented have been made in the test bed Sense-City (see Fig. 3). To facilitate the subsequent interpretation of vertical profiles with buried pipes, reference measurements were performed apart from the areas

with buried objects to obtain the dielectric characteristics of the multilayered soil. Using the commercial GPR GSSI SIR 3000, a scanning has been acquired to identify the several layers of the sub-surface. Consistently with the direct observations during excavation, the raw radargram at 900 MHz (without time zero correction and clutter removal) shown in Fig. 12 shows a soil made of four layers corresponding to asphalt (layer 1), aggregate cement (layer 2),
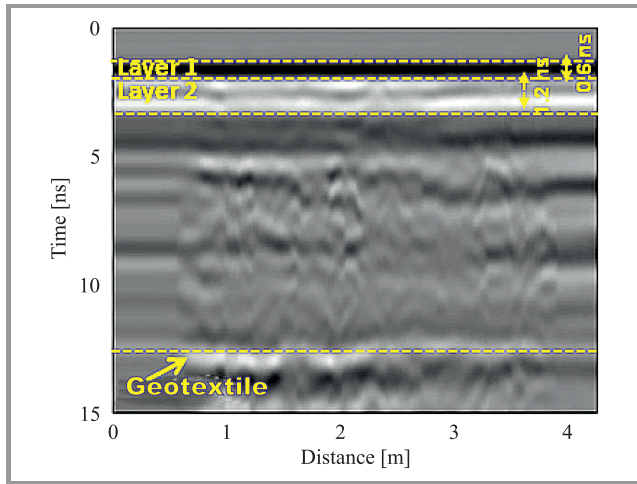
Fig. 12. Experimental radargrams (B-scans) of the subsurface without buried objects measured by the SIR 3000 system (linear gain and low and high pass filters applied) at 900 MHz in the TM polarization.
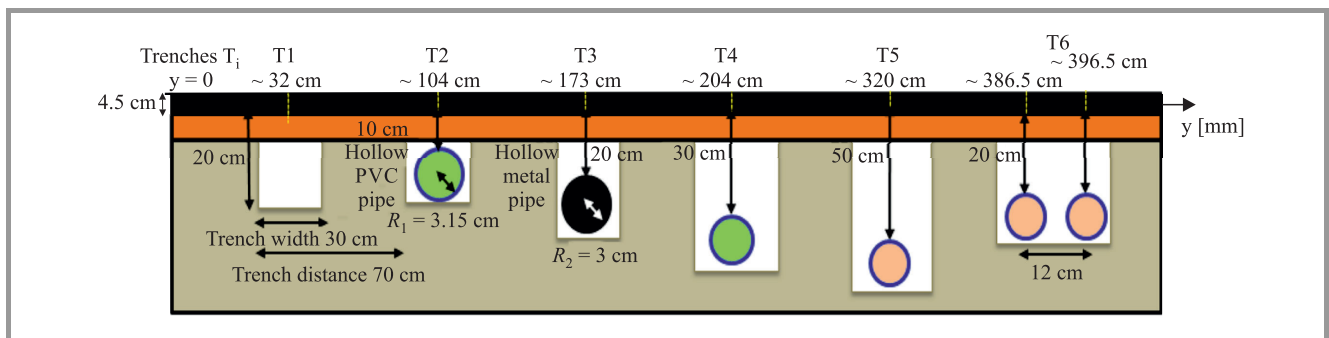
a natural soil (layer 3), and a wetter natural soil (layer 4) located under the geotextile (a thin discontinuity). A sublayer of layer 3 (layer 3b) of natural soil under the aggregate cement appears visible in this zone of the site and it is attributed to a compacted layer of natural soil. This layer will not be observed in the zone with the trenches because the soil has been excavated. From the radargram at high resolution, we have estimated the real permittivities ε' since we have an a priori knowledge of the thickness h of each layer from the construction phase. According to Fig. 3a, the real permittivity estimations of the soil layers are the following: asphalt (layer 1) $\varepsilon'_1 = 4.5$ (0.64 ns), aggregate cement (layer 2) $\varepsilon'_2 = 7.7$ (1.4 ns), natural soil (mean of layers 3b and 3) $\varepsilon'_3 = 34$ (8.5 ns). Based on these results,

a mean real permittivity of 12.7 is derived (decomposition in volumetric fractions).

Measurements have been made manually using the SFCW georadar system on the pipe zone of the test site whose cross section is presented in Fig. 13. The radargrams are presented in the time domain in both TE and TM polarizations (see Figs. 14a-b) considering a pulse with a peak frequency at 900 MHz. Neither low pass, nor high pass filters and nor time zero correction were applied here. The pipe in T5 was filled with air. A linear time gain [–10 ; 20] dB was applied to the radargram of Fig. 14a in the TE polarization (broadside). We remark in this figure that the air-filled pipe at depth 30 cm in T4 shows a detectable hyperbola signature (contrary to the SIR 3000 results). The signature of trench T1 appears weakly detectable, certainly because the distance step (4 cm) has not been defined here small enough (this will be improved in the future). In the TM polarization (end-fire), such as visualized in Fig. 14b, the measurements were made in two sequences that explains the vertical discontinuity in the radargram. In general, we remark that all the hyperbolas appear strongly attenuated that is consistent with the synthetic radargram obtained from FDTD simulations [22].
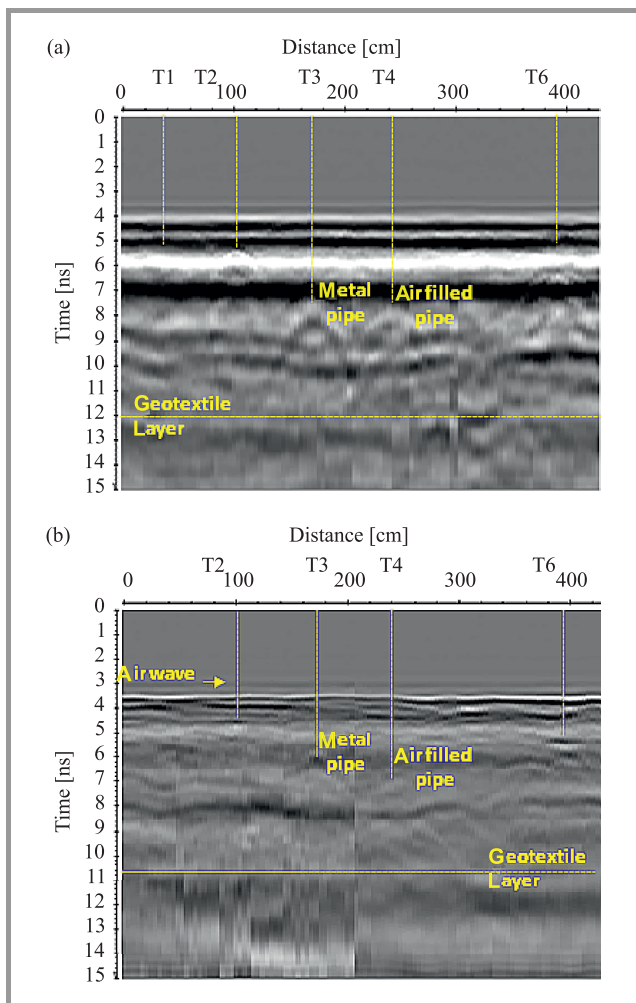
7. Conclusion

In this work, signal and image processing techniques have been collected with the aim of extracting quantitative information from experimental radargrams containing hyperbola signatures from pipes and strips with lateral dimension less than around ten centimeters. The difficulties met in the analysis of GPR radargrams is the detection of a hyperbola pattern in a noisy background, in a weak image quality, with overlapping between signatures of the clutter and one or multiple targets (particularly buried in a shallow depth), and also the detection of a target with a small lateral dimension and a weak contrast with a perturbed surrounding medium. The main objective was to propose a data processing chain easily implementable to analyze by steps radargrams made of hyperbola signatures associated with relatively small pipes (less than 100 mm). Thus, a semi-automatic data processing has been proposed that is supposed to be adaptable to hyperbola signatures en-

Fig. 13. Cross-section of the pipe zone in the Sense-City test bed.

Fig. 14. Experimental radargrams measured by the SFCW GPR system in the: (a) TE (broadside), and (b) TM (end-fire) polarizations.

frequency time-reversal matrices [24], wavelet transform, or hyperbola detection that accounts for hyperbola miss-hapedness [25].

Acknowledgment

References

[1] Z. Liu and Y. Kleiner, "State of the art review of inspection technologies for condition assessment of water pipes", *Measurement*, vol. 46, no. 1, pp. 1–15, 2013.

[2] H. O. Henriques, M. Z. Fortes, L. Hudson, N. S. V. Silva, and F. O. Teixeira, "Use of radar illegal connections prospecting in buried or embedded cables", *Measurement*, vol. 47, pp. 221–227, 2014.

[3] CI/ASCE38-02, American Society of Civil Engineers, Standard guideline for the collection and depiction of existing subsurface utility data, 2002.

[4] ASTM Designation D 6432-11, Standard Guide for Using the Surface Ground Penetrating Radar Method for Subsurface Investigation, 2011.

[5] Italian Standard CEI-883, Regulations for performing preliminary surveys with ground-probing radar for before laying underground utilities and infrastructures, 2004.

[6] *Proceedings of the First General Meeting of COST Action TU1208*, L. Pajewski, and A. Benedetto, Eds. Rome, Italy, July 2013, Aracne, Rome, Italy, July 2013 [Online]. Available: http://www.cost.eu/domains_actions/tud/Actions/TU1208

[7] F. Sagnard, "Design of a compact ultra-wide band bow-tie slot antenna system for the evaluation of structural changes in civil engineering works", *Progress in Electromag. Res.*, PIER B, vol. 58, pp. 181–191, 2014.

[8] F. Sagnard and E. Tebchrany, "Using polarization diversity in the detection of small discontinuities by an ultra-wide band ground penetrating radar", *Measurement*, vol. 61, pp. 129–141, 2015.

[9] F. Sagnard, C. Norgeot, X. Derobert, V. Baltazart, E. Merliot, F. Derkx, and B. Lebental, "Utility detection and positioning on the urban site Sense-City using Ground Penetrating Radar Systems", *Measurement*, vol. 88, pp. 318–330, 2016.

[10] R. Solimene, A. Cuccaro, A. Dell'Aversano, I. Catapano, and F. Soldovieri, "Ground clutter removal in GPR surveys", *IEEE J. of Selec. Topics in Appl. Earth Observ. and Remote Sensing*, vol. 7, no. 3, pp. 792–798, 2014.

[11] E. Tebchrany, F. Sagnard, V. Baltazart, and J. P. Tarel, "Data processing of ground-penetrating radar signals for the detection of discontinuities using polarization diversity", European Geoscience Union General Assembly 2014, Vienna, Austria, 27 Apr. – 2 May 2014.

[12] G. T. Tesfamariam, D. Mali, and A. M. Zoubir, "Clutter reduction techniques for GPR based buried landmine detection", in *Proc. Int. Conf. on Signal Process., Commun., Comput. and Netw. Technol. ICSCCN 2011*, Thuckalay, India, 2011, pp. 182–186.

[13] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.

[14] A. Hyvärinen, "Independent component analysis: recent advance", *Philos Trans. A Math. Phys. Eng. Sci.*, vol. 371, no. 1984, 2013 (doi: 10.1098/rsta.2011.0534).

countered in a generally noisy and heterogeneous environment showing different dielectric contrasts with the targets. It has been observed that to remove the clutter, deep and shallow targets must be distinguished. In the case of a shallow target, which significantly influences the clutter and the interface reflection, the PCA algorithm appears in general more efficient than ICA. In the case of a deep target, ICA appears to be more efficient than PCA. The template matching algorithm has been adapted to the hyperbola detection by means of the definition of a adequate template that represents a hyperbola shape close to the apex. The template has been generated from FDTD simulations. The algorithm has been validated on experimental radargrams, and has been extended to take advantage of data sets that consider the polarization diversity. This additional parameter improves the algorithm robustness. Future researches aim at applying the template matching algorithm on radargrams with different target shapes (pipe and cracks) and dielectric contrasts with the surrounding soil. Improvement of the data processing chain could be made by adding complementary algorithms such as space-

[15] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications", *Neural Netw.*, vol. 13, no. 4, pp. 411–430, 2000.

[16] S. J. Radzevicius and J. J. Daniels, "Ground penetrating radar polarization and scattering from cylinders", *J. Applied Geophys.*, vol. 45, pp. 111–125, 2000.

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

[18] P. K. Verma, A. N. Gaikwad, D. Singh, and M. Nigam, "Analysis of clutter reduction techniques for through wall imaging in UWB range", *Progress in Electromag. Res. B*, vol. 17, pp. 29–48, 2009.

[19] T. Fawcett, "An introduction to ROC analysis", *Pattern Recogn. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

[20] F. Sagnard and J. P. Tarel "Template-matching based detection of hyperbolas in ground-penetrating radargrams for buried utilities", *J. of Geophys. and Engin.*, vol. 13, no. 4, pp. 491–504, 2016.

[21] S. Shihab and W. Al-Nuamy, "Radius estimation for cylindrical objects detected by ground penetrating radar", *Subsurface Sensing Technol. and Appl.*, vol. 6, no. 2, pp. 151–166, 2005.

[22] D. Vernon, *Machine Vision-Automated Visual Inspection and Robot Vision*. Prentice Hall, 1991.

[23] C. A. Balanis, *Advanced Engineering Electromagnetic*. Wiley, 1989.

[24] M. E. Yavuz, A. E. Fouda, and F. L. Teixeira, "GPR signal enhancement using sliding-window space-frequency matrices", *Progress in Electromag. Res.*, vol. 145, pp. 1–10, 2014.

[25] L. Mertens, R. Persico, L. Matera, and S. Lambot, "Automated detection of reflection hyperbolas in complex GPR images with no a priori knowledge on the medium", *IEEE Trans. Geosci. and Remote Sens.*, vol. 54, no. 1, pp. 580–596, 2016.

Florence Sagnard received the B.Sc. and the D.E.A. degree in Electronics from Université Pierre et Marie Curie, France, in 1990 and the Ph.D. degree from Université d'Orsay, France, in 1996. In 2005, she was qualified to conduct researches (HDR) from Université de Marne-La-Vallée, France. From 1997 to 2006, she was Assistant Professor at Université de Marne-La-Vallée and worked in the laboratory ESYCOM on the microwave characterization of materials. From 2002 to 2006 she worked as an invited researcher at INSA (Institut National des Sciences Appliquées), Rennes in UWB antennas, and transmission links. In 2007–2009, she joined CEREMA (Centre d'Etudes et d'expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement), Rouen, and since 2010 she is with IFSTTAR Marne-La-Vallée to develop SFCW GPR systems. Her research interests include data and image processing, FDTD simulations, UWB antennas, soil characterization, and civil engineering.
E-mail: florence.sagnard@ifsttar.fr
Université Lille Nord de France
IFSTTAR
20 rue Elisée Reclus
59650 Villeneuve-d'Ascq, France