Assessment of statistical-based clutter reduction techniques on ground-coupled GPR data for the detection of buried objects in soils

E. Tebchrany, F. Sagnard, V. Baltazart, J-P. Tarel, X. Dérobert
IFSTTAR, 14-20 bd Newton, 77420 Champs-sur-Marne, France
elias.tebchrany@ifsttar.fr, florence.sagnard@ifsttar.fr, vincent.baltazart@ifsttar.fr

***Abstract*— A bi-static Ground Penetrating Radar (GPR) has been developed for the detection of cracks and buried pipes in urban grounds. It is made of two shielded Ultra Wide Band (UWB) bowtie-slot antennas operating in the frequency band [0.3;4] GHz. GPR signals contain not only responses of targets, but also unwanted effects from antenna coupling in air and in the soil, system ringing, and soil reflections that can mask the proper detection of useful information. Thus, it appears necessary to propose and assess several clutter reduction techniques as pre-processing techniques to improve the signal-to-noise ratio, discriminate overlapping responses issued from the targets and the clutter, and ease the use of data processing algorithms for target detection, identification or reconstruction. In this work, we have evaluated on Bscan profiles three different statistical data analysis such as mean subtraction, Principal Component Analysis (PCA), and Independent Component Analysis (ICA) considering a shallow and a medium depth target. The receiver operating characteristics (ROC) graph has allowed to evaluate the performance of each data processing in simulations and measurements to further draw a comparison in order to select the technique most adapted to a given soil structure with its radar probing system.**

***Index Terms*— Clutter reduction, signal processing, object detection, polarimetry, ultra-wide band radar.**

I. INTRODUCTION

In the field of civil engineering, ground penetrating radar (GPR) is a well-known non destructive tool to probe natural or man-made materials or structures (road, bridges, or soils)([1]-[2] and the references herein). GPR is generally used to survey pavement thickness at traffic speed, detect and localize buried objects (pipes, cables, voids, cavities), zones of cracks and discontinuities in concrete or soils. Among both GPR configurations [2], the ground-coupled and the air-launched radar configurations, we have decided to work with a pair of transmitting and receiving bowtie-slot antennas in close proximity to the soil surface in order to minimize the significant reflection component issued from the air-soil interface which masks the signature of a shallow dielectric target (object or dielectric discontinuity), and also to improve the penetration of electromagnetic energy into the soil [3]. The GPR system operates in the frequency domain using a step-frequency continuous wave (SFCW) using a Vector Network Analyzer (VNA ANRITSU MS 2026B) in an ultra-wide band [0.3; 4] GHz to benefit from a wide dynamic range and a low figure noise. Its linear moving along the soil surface allows detecting the reflected waves induced by all discontinuities in the subsurface. In such a configuration, the direct and the ground-coupling in the antenna system which cannot easily be separated can overlap and obscure the target response, thus producing part of the clutter component. The clutter varies with soil dielectric characteristics and/or surface roughness and leads to uncertainty in the measurements [1]-[4]-[5].

The moving of the GPR system along a sampled linear path provides a 2D profile (Bscan), and thus an image that can be analyzed in both the frequency and the time domain. The detection of targets is usually focused on time imaging. Thus, the targets (limited in size) are usually shown by diffraction hyperbolas on a Bscan image that is an unfocused depiction of the scatterers. The contrast in permittivity and the ratio between the size of the target and the wavelength contribute to enhance the radar cross section (rcs) of the target, which means the magnitude of the target signature. However, Bscan images contain also the clutter signal composed of two overlapping signals, i.e., the direct coupling between the antennas and the reflected wave from the soil surface, the scattering on the heterogeneities due to the granular nature of the subsurface material, and finally some additive noise. In general, the clutter has a statistical nature because the ground surface is not perfectly flat and smooth, and shows heterogeneities. It appears as a group of overlapping signals that are uncorrelated to the target scattering characteristics but can occupy the same time and frequency domains as the target. Thus, clutter reduction appears as a first step, a pre-processing technique, in data processing algorithms aiming at target detection, identification or reconstruction. Clutter reduction techniques are classified in statistical signal processing, conventional filtering, wavelet packet decomposition, and non linear signal processing based on neural networks [4-7]. In this paper, the conventional mean subtraction technique (or background removal) and two statistical signal processing techniques, namely the Principal Component Analysis (PCA) and Independent

Component Analysis (ICA) techniques, have been considered and compared to reduce the clutter and to enhance the target signal. The conventional mean subtraction technique assumes a steady clutter signal, thus allowing the suppression of a horizontal coherent signal calculated over several traces. As a counterpart, the latter method is sensitive to any space variations of the clutter signal, i.e., the spatial variations of either the soil permittivity or the surface roughness. PCA and ICA are multivariable and statistical techniques aiming at data dimensionality reduction; as the goal in PCA is to find an orthogonal linear transformation that maximizes the variance of the variables, the goal of ICA is to express a set of random variables as linear combinations of statistically independent component variables.

In this paper, we propose to evaluate quantitatively, by means of the receiver operating characteristics (ROC) graph [9], the performance of PCA and ICA in Bscan profiles considering a shallow and a medium depth target. The study has been focused on a few scenarii for the soil structure for which the target is embedded at different depth within soil. Within this scope, the clutter signals can disturb the detection of the first arrival signals from the target. The level of the disturbance depends on the dielectric characteristics (conductive or dielectric, shape) of the target to be detected.

II. HARDWARE SETUP FOR SIMULATIONS AND MEASUREMENTS

The ultra-wide band (UWB) ground-coupled radar used and designed in our laboratory is composed of two bowtie-slot antennas with dimensions close to an A4 paper sheet, as shown in Fig. 1a. The antennas, designed on a single-sided FR4 substrate ($h = 1.5\,mm$; $\varepsilon_r' = 4.4$; $\tan\delta = 0.01$), operate in the frequency band [0.3; 4] GHz, and have a maximum gain of 5.5 dB [3]. To reduce the backward radiation in air and interactions with the surrounding environment, the antennas have been shielded with a metallic box filled with a 68 mm thick-layered absorbing material. According to previous studies [3], the antenna offset has been defined to 60 mm. To acquire Bscan profiles along a linear path on the soil surface with a 40 mm sampling step, two antenna configurations or polarizations (see Fig. 1b) have been considered: the parallel (end-fire or TE polarization), and the mirror configuration (broadside or TM polarization). FDTD simulations of all the antenna geometries and the soil have been made using the commercial software EMPIRE XCcel with an adaptative meshing resolution of lambda/15 in the different media. The excitation signal is the first derivative of the Gaussian function with the duration 0.5 ns and a time zero 0.3 ns; its spectrum has a maximum frequency at 1 GHz and a bandwidth of 3 GHz. A layered flat soil has been considered. The antennas are supposed to be positioned towards the soil surface at an elevation close to 10 mm. Measurements have been made using portable VNA in the step frequency (SF-GPR) acquisition mode in the band

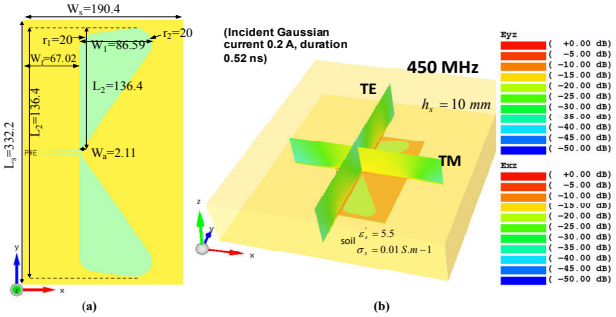

Fig. 1. The CPW-fed triangle bowtie slot antenna: (a) top view and (b) the radiated electric field in both planes TE and TM

[0.05; 4] GHz; 1601 frequency samples and an intermediate Frequency (IF) bandwidth of 500 Hz have been considered. A full two ports calibration has been made with two 2 m length radiofrequency cables. To obtain time data at finer resolution, zero padding has been performed with the bandwidth extended to 9 GHz. To compare synthetic and measured Bscans, the spectrum of the excitation signal used in the simulations has been multiplied to the frequency data and an inverse Fourier transform has been computed.
Both types of polarizations have been considered, as they influence the detection of either a conductive or a dielectric pipe in a given soil.

III. THE PROPOSED STATISTICAL CLUTTER REDUCTION TECHNIQUES

A. Data modeling

The GPR system with its transmitting and receiving antennas is moved linearly along the ground surface to detect the reflected waves from the soil subsurface. This displacement produces a set of time or frequency signals named traces (Ascans) at each spatial step; N is the number of spatial samples and M the number of time or frequency samples. Thus, the collected data can be expressed in a data matrix $X\,(M \times N)$, where $N < M$; considering a data sample X_{ij}, i denotes the frequency (time) index and j denotes the GPR position index. PCA and ICA are techniques of array processing and data analysis that will consider matrix X.

B. Principal Component Analysis (PCA)

PCA is a second-order statistical method, as only covariances between the observed variables are used in the estimation process. The variables are supposed Gaussianly distributed, linear and stationary. The purpose of PCA is to derive a relatively small number of decorrelated linear combination called principal components (PCs) of a set of random zero-mean variables ($K < M$) while retaining as much of the information from the original variables (frequency or time samples) as possible. The PCA algorithm can be performed using two approaches: the eigenvalue decomposition of the covariance matrix of the data and the singular value decomposition (SVD) of the data matrix. In this work, SVD has been performed on

matrix X from which the mean of each trace has been removed such as:

$$X = U \Sigma V^T \quad (1)$$

Where $\Sigma(N \times N)$ ($\Sigma = \Sigma^T$) is a diagonal matrix whose elements named the singular values σ_i, indicate the amount of information (variance) contained in each principal component, $U(M \times N)$ and $V(N \times N)$ are respectively the left and right singular vectors.

The amount of variance contained in each eigenvector is directly related to the corresponding eigenvalue from $\Sigma^2 = D$. Thus, the singular values can be sorted in descending order and the singular vectors in U,V reordered the same way. The clutter or non-target related signals are usually contained in the first few singular vectors. The criterion proposed in [4]-[5] has been used in order to select the really significant components containing the target information in the dataset and reduce its dimension such as:

- Eliminate some percentage below the sum of all singular vectors;
- Plotting singular values against the order number to look for breakpoints in the slope of the curve where the singular values of noise should slightly change and hence produce a flat slope.

The clutter-free dataset is then synthesized from the [*k1*, *k2*] indexed principal components as follows:

$$Y = \sum_{i=k_1}^{k_2} u_i \, \sigma_i \, v_i^T \quad (2)$$

C. Independent Component Analysis (ICA)

ICA analysis is a blind source separation (BSS) technique to extract statistically independent components (ICs called sources) generated simultaneously using the observed sequences of data [8]. The sources are supposed mixed by an unknown medium and the mixed signals are delivered by sensor measurements. The term 'blind' stresses the fact that the mixing structure and the sources are both unknown. As the PCA technique uses a data analysis based on second-order statistics adapted to signals with a Gaussian distribution (second-order are zero), the ICA method considers the majority of signals (natural signals) which do not have a Gaussian distribution and do have higher-order moments. The subspace formed by ICA is therefore not orthogonal as it is the case with components extracted by PCA. In addition, extracted ICs are truly statistical independent, thus the aim of ICA is to find new variables (ICs) that are both statistically independent and non-Gaussian.

The mathematical model for ICA, based on the same structure of data matrix X, expresses a linear transformation such as:

$$X = A S \quad (3)$$

Where $S(M \times N)$ represents the matrix containing N original sources as its columns and $A(N \times N)$ is a square mixing matrix. Thus, X represents a mixture of the N sources forming *S*.

The main computational issue in the ICA process is the estimation of the mixing matrix A, so that the ICs, i.e. the columns of matrix S can be obtained as follows:

$$S = A^{-1} X \quad (4)$$

And the estimate of *S* is thus defined $Y = W X$.

The determination of the ICs begins by removing the mean values of the measurements (centering process) as in PCA. The next step is to whiten data X that means linearly transforming them by a matrix V such that $Z = V S$ is white, that is $E\{Z Z^T\} = I$ where I is the identity matrix. Matrix V can be easily found by PCA (eigenvalue decomposition) as normalizing the principal components to unit whitening data; the data are projected onto its principal component directions. As for Gaussian variables, uncorrelatedness implies independence, whitening eliminates all the dependence information in the data, this is not the case for non-Gaussian variables, and there is much more information in the data than what is used in whitening. The relation between non-gaussianity and independence could be explained easily using the central limit theorem (CLT) [8]. According to the CLT, the distribution of a sum of independent random variables tends to be "more" Gaussian than the original random variables. To obtain independent original sources we have to find W such that $y_i(t)$ is "least" Gaussian. The quantitative measure of non-Gaussianity of the signal (definition of a contrast function), which is generally based on kurtosis (a fourth order cumulant) and negentropy estimations, allows to obtain the mixing matrix. The optimization (minimization or maximization) of a contrast function enables the estimation of the independent components. Afterwards, the sources can be estimated. Our implementation uses the fixed point algorithm FastICA [8]. The selection of sources is based on the value of their Kurtosis fourth order cumulant [7].

IV. Evaluation on Simulated and Experimental Data

A. ROC curves as a method of assessment

The Receiver Operating Characteristic (ROC) methodology has been introduced in the early 80's and has become a conventional technique for performance assessment. It was firstly used to measure diagnostic performances of medical imaging systems. For a single target (a defective area in our case), the ROC analysis consists in measuring the binary response of the detection system (target present or not) to one stimulus, in our case an image, by calculating the true positive (TP) rate t_{pr} and the false positive (FP) rate f_{pr} [9]. A couple (t_{pr}; f_{pr}) corresponds to one point in the ROC plane for a given threshold in the range [0; max(amp)], where max(amp)

corresponds to the maximum magnitude of the absolute values in the Bscan. It is common to use the area under a ROC curve (AUC) that expresses the performance for the goodness of a ROC curve. In practice, the calculation of both rates TP and FP is based on the use of a reference binary image including two successive hyperbolas (see Fig. 3a) to draw a comparison with the raw binary image. The parameter used to plot the ROC curve is the threshold whose value has been used to convert the raw image into a binary image.

B. Simulated results

At first, synthetic data have been obtained from FDTD simulations where the antennas and its environment have been modeled as visualized in Fig. 2a and 2b. The polarizations TE and TM have been considered to probe a dielectric and a conductive pipe, respectively. The center-to-center distance between antennas are $SR = 291\ mm$ and $SR = 422\ mm$ respectively. A pipe with 12 mm radius buried at abscissa 500 mm in a two-layer soil ($\varepsilon_1^{'} = 5.5$, $\varepsilon_2^{'} = 9$, and $\sigma_{1,2} = 0.01\ S.m^{-1}$) at two depths 110 and 160 mm (0.86λ and 1.25λ at 1GHz) from the surface; the top layer is 70 mm thick. The radar is moved along axis Oy in the range [0; 1000] mm.

Considering a dielectric ($\varepsilon_{pvc} = 3.4$) pipe filled with air and probed in the mirror configuration (see Fig. 2a) at the depth 110 mm, the raw Bscan shown in Fig. 3a highlights the clutter responses associated with the direct wave in the soil and the reflection on the second soil layer at 3.2 ns and 3.7 ns, respectively. The first red diffraction hyperbola (positive waveform) occurs at 4.3 ns, and the second blue hyperbola (negative waveform) occurs at 4.8 ns. The two responses (clutter and target diffraction) overlap slightly. The time differences are similar to those obtained from the analytical ray-path model. We observe that the mean subtraction, PCA and ICA succeed in reducing significantly the clutter as shown on Fig. 3b, 3c and 3d respectively. The mean subtraction better enhances the first red hyperbola, while ICA better highlights the second blue hyperbola without introducing any distortion in the image; as a counterpart, PCA seems to introduce a signal artifact at 4.8 ns. However, it appears from ROC curves plotted in Fig. 4a (reference hyperbola in Fig. 3a) that ICA and the mean subtraction techniques afford the best performance in clutter removal over PCA, i.e., the corresponding ROC curve is the nearest to the high left corner.

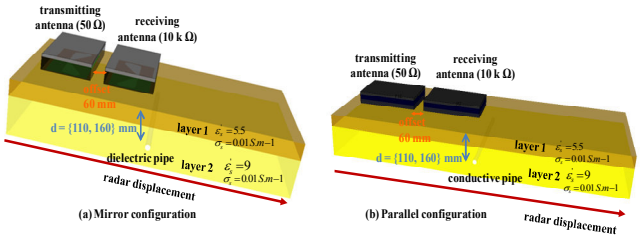

Fig. 2. Both GPR configurations modeled using FDTD simulations: (a) the mirror (TE) configuration, and (b) the parallel (TM) configurations

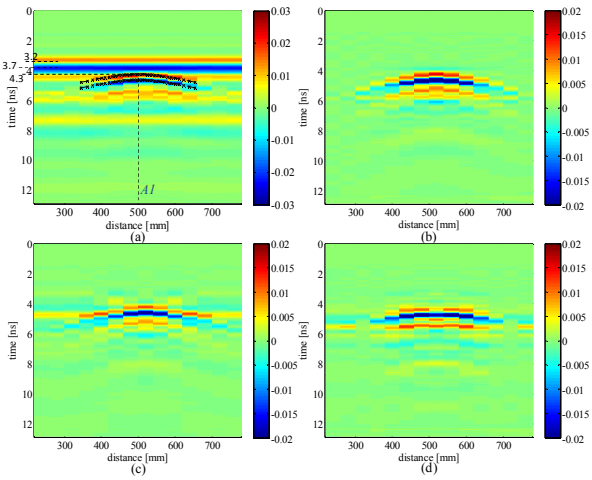

Fig. 3. Clutter reduction techniques applied on numerical data in the case of a dielectric air-filled pipe (r=12 mm, depth 110 mm) in the mirror configuration: (a) raw data with the reference hyperbolas, (b) mean subtraction, (c) PCA, and (d) ICA

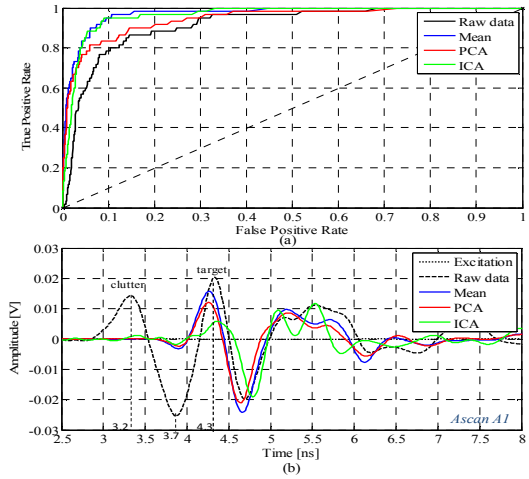

Fig. 4. Case of a dielectric air-filled pipe (r=12 mm) buried at a depth 110 mm; (a) ROC curves, and (b) Ascans A1 issued from the three clutter reduction algorithms

Fig. 4b shows the different Ascan signals at the apex of the hyperbola on Fig. 3a to Fig. 3d i.e., y = 500. It shows that the mean technique lowers more significantly the clutter component in the range [2;4] ns, while the three techniques correctly enhance the target signal ; ICA changes the target apex time position from 4.3 to 4.5 ns.

Considering a conductive pipe probed in the parallel configuration (see Fig. 2b) at depth 110 mm, the raw Ascan visualized in Fig. 5a shows that the arrival times of the clutter (4.1 ns) and the second soil layer (4.8 ns) overlap the arrival time of the diffraction hyperbola whose apex is estimated at 4.5 ns. According to Fig. 5b, 5c, and 5d, PCA appears more efficient in removing the clutter and the second soil layer signals than ICA. Because the target response influences the clutter and the second layer responses, the signals appear dependent as opposed to the assumption for ICA (independent sources) is not respected. The ROC curves in Fig. 6a and Ascans A2 at the hyperbola

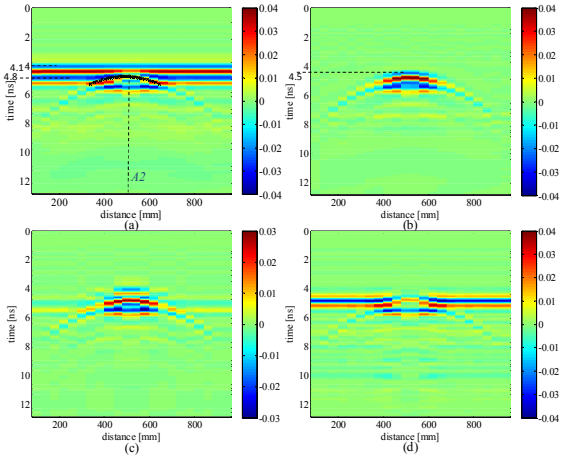

Fig. 5. Clutter reduction techniques applied on numerical data in the case of a conductive pipe (r=12 mm, depth 110 mm) in the parallel configuration: (a) raw data, (b) mean subtraction, (c) PCA, and (d) ICA

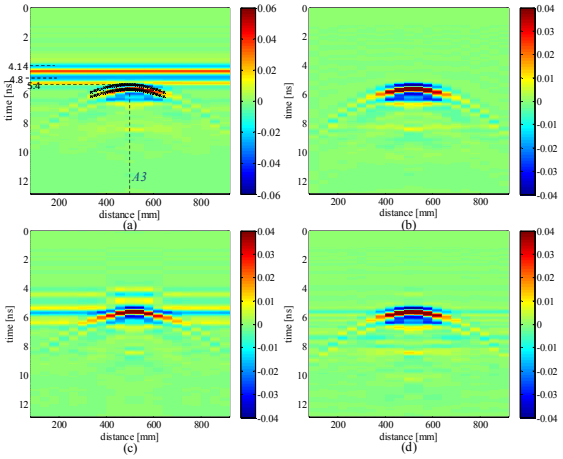

Fig. 7. Clutter reduction techniques applied on numerical data in the case of a conductive pipe (r=12 mm, depth 160 mm) in the parallel configuration: (a) raw data, (b) mean subtraction, (c) PCA, and (d) ICA

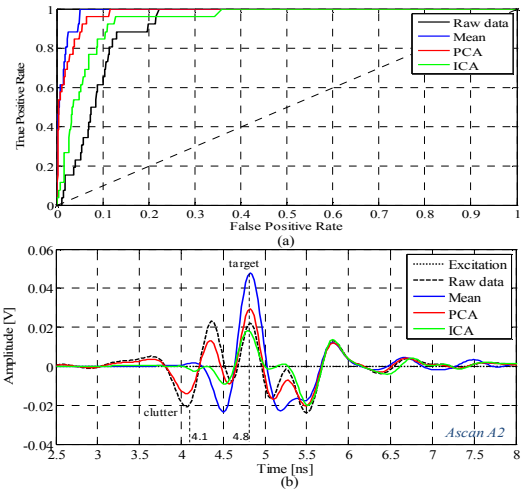

Fig. 6. Case of a conductive pipe (r=12 mm) buried at a depth 110 mm; (a) ROC curves, and (b) Ascans A2 issued from the three clutter reduction algorithms

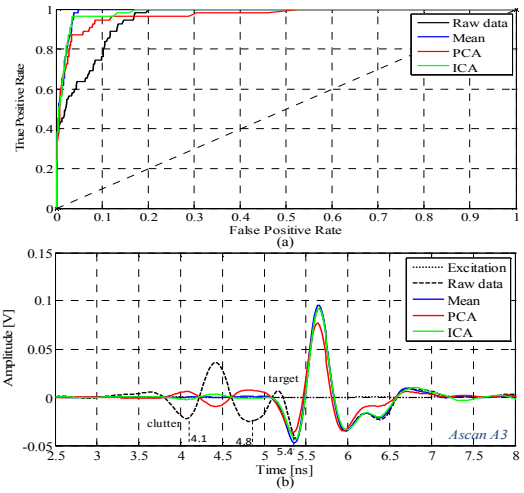

Fig. 8. Case of a conductive pipe (r=12 mm) buried at a depth 160 mm; (a) ROC curves, and (b) Ascans A3 issued from the three clutter reduction algorithms

apex in Fig. 6b confirms that PCA is the best algorithm in this case to remove signals not containing target information, while the mean introduces an artifact by increasing the amplitude at the apex. The conductive pipe has now been buried more deeply at the depth 160 mm to obtain resolved responses. From the raw Bscan visualized in Fig. 7a, we observe that the arrival times of the clutter, the second soil layer and the hyperbola apex are 4.1, 4.8 and 5.4 ns, respectively. Presently, ICA affords better results than PCA in removing the clutter as mentioned by the ROC curves and the Ascans A3 in Fig. 8a and 8b.

C. Experimental results

The clutter removal techniques have been also applied on field data. The measurements with the UWB radar have been made in the large sandy box of the public square Perichaux, Paris 15th district. The sand was wet and not compacted with a depth estimated to 48 cm; its dielectric permittivity has been estimated to 3.5 according to previous work in [3].

In a first measurement campaign, a PVC air-filled pipe with a 12 mm radius buried in the sand box at 160 mm at position 500 mm has been probed with the GPR on a linear path [0;1000] mm (step 40 mm) using the mirror configuration. The raw Bscan in Fig. 9a shows components of clutter with significant roughness (unlike the smoothness in simulations) overlapping the hyperbola signature. As shown previously in numerical results where signals overlap, the best performance in clutter removal is obtained by PCA as visualized on the Bscans of Fig. 9b, 9c and 9d and on the ROC curves of Fig. 10a. Concerning the ROC curves, we remark that the FP rate does not reach zero because the hyperbola amplitude appears lower than components not containing target information.

In a second measurement campaign, a conductive pipe with a 12 mm radius at abscissa 450 mm has been probed in the parallel configuration. The experimental conditions remain the same as in the previous measurements. From Fig. 9a, we remark that the clutter does not overlaps the

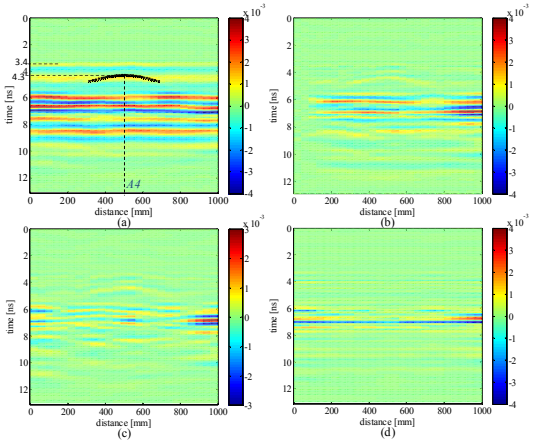

Fig. 9. Clutter reduction techniques applied on experimental data in the case of a dielectric air-filled pipe (r=12 mm, depth 160 mm) in the mirror configuration: (a) raw data, (b) mean subtraction, (c) PCA, and (d) ICA

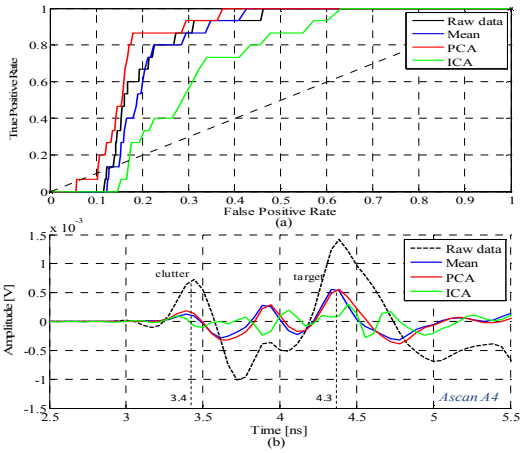

Fig. 10. Case of a dielectric air-filled pipe (r=12 mm) buried at a depth 160 mm; (a) ROC curves, and (b) Ascans A4 issued from the three clutter reduction algorithms

hyperbola response. In this case, the hyperbola signature is characterized by higher amplitude as compared to other signals. According to the processed Bscan in Fig. 11b and ROC curves in Fig. 11c, ICA shows the best performance in clutter removal because the hyperbola does not appear fully masked by the clutter. Moreover, we notice on Fig. 11b that ICA removes most of the multiple reflections occurring at late time delays, unlike PCA and the mean.

V. Conclusion

In this paper, we have presented the application of two different statistical techniques, namely PCA and ICA, for reducing the clutter and unwanted signals in ground-coupled GPR images in the case of a single target buried in a two-layered medium. Our objective was to study the performance and the shortcomings of each technique using ROC curves. Considering numerical and field data, we have noticed that in the case of a shallow target, which significantly influences the clutter and the interface reflection, PCA appears to be more efficient than ICA. While in the case of a deep target, ICA appears to be more efficient than PCA. Future studies aim at considering

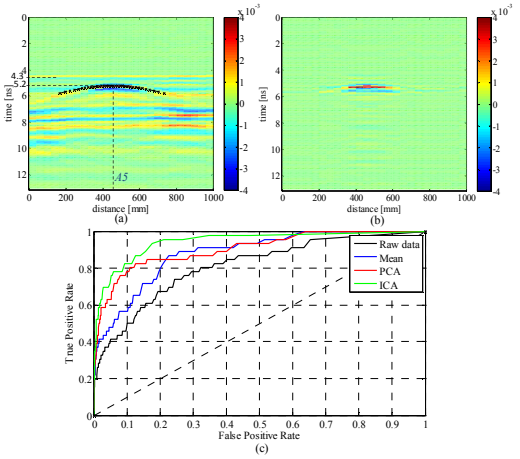

Fig. 11. Clutter reduction techniques applied on experimental data in the case of a conductive pipe (r=12 mm, depth 160 mm) in the parallel configuration: (a) raw data, (b) ICA, and (c) ROC curves

different dielectric contrast between the target and the soil and also several targets in a single Bscan. Moreover, the influence of the polarization could be studied in order to obtain the response of the clutter as a function of the polarization. All this pre-processing will help us to further detect small defaults in a civil engineering structure.

References

[1] X. Dérobert, C. Fauchard, P. Côte, E. L. Brusq, E. Guillanton, J. Dauvignac, and C. Pichot, “Step-frequency radar applied on thin road layers,” Journal of Applied Geophysics, vol. 47, pp. 317–325, July 2001.

[2] Proceedings of the General Meeting of COST Action TU1208, Rome (IT), 22-24 July 2013. http://www.cost.eu/domains_actions/tud/Actions/TU1208

[3] F. Sagnard, E. Tebchrany, and V. Baltazart, "Evaluation of an UWB ground-coupled radar in the detection of discontinuities using polarization diversity: FDTD modeling and experiments," 7th International Workshop on Advanced Ground Penetrating Radar (IWAGPR), IEEE, pp. 1-6, 2013.

[4] F. Abujarad, “Ground penetrating signal processing for landmine detection,” PhD Thesis, Otto-von-Guericke-UniversitÄat Magdeburg, March 2007.

[5] N.J.F. Bostanudin, “Computational methods for processing ground penetrating radar data,” PhD Thesis, University of Portsmouth, May 2013.

[6] P. K. Verma, , A. N. Gaikwad, D. Singh, and M. J. Nigam, "Analysis of clutter reduction techniques for through wall imaging in UWB range," Progress In Electromagnetics Research B, vol. 17, pp. 29-48, 2009.

[7] B. Karlsen, H. B. Sorensen, J. Larsen, and K. B. Jakobsen, "GPR detection of buried symmetrically shaped minelike objects using selective independent component analysis," AeroSense 2003, International Society for Optics and Photonics, pp. 375-386, 2003.

[8] A. Hyvärinen, and E. Oja, "Independent component analysis: algorithms and applications," Neural Networks, vol. 13, n. 4, pp. 411-430, 2000.

[9] T. Fawcett, "An introduction to ROC analysis,” Pattern Recognition Letters, vol. 27, n. 8, pp. 861-874, 2006.