A Classification Approach Based on SVM for Electromagnetic Subsurface Sensing

Andrea Massa, *Member, IEEE*, Andrea Boni, and Massimo Donelli, *Member, IEEE*

***Abstract*—In clearing terrains contaminated or potentially contaminated by landmines and/or unexploded ordnances (UXOs), a quick wide-area surveillance is often required. Nevertheless, the identification of dangerous areas (instead of the detection of each subsurface object) can be enough for some scenarios/applications, allowing a suitable level of security in a cost-saving way. In such a framework, this paper describes a probabilistic approach for the definition of risk maps. Starting from the measurement of the scattered electromagnetic field, the probability of occurrence of dangerous targets in an investigated subsurface area is determined through a suitably defined classifier based on a support vector machine. To assess the effectiveness of the proposed approach and to evaluate its robustness, selected numerical results related to a two-dimensional geometry are presented.**

***Index Terms*—Electromagnetic scattering inverse problems, pattern classification, subsurface sensing.**

I. Introduction

In the world, there are many areas contaminated (or potentially contaminated) with unexploded ordnance (UXO) and antitank/antipersonnel landmines. A report published in 1997 indicates that the approximate number of UXOs is of about 110 million in 70 countries [1], [14]. Moreover, recent missions in Bosnia, Afghanistan, and Iraq probably (and unfortunately) will further increase such an estimate. To return these zones to a civilian use, the ordnances should be obviously removed. However, in several cases, the former bombing ranges have been unused for many years and the UXO locations are partially known or completely unknown. Thus, a wide-area surveillance is needed in order to circumscribe those regions where the dangerous targets reside. Such a process is inevitably time-expensive and involves complex acquisition procedures. Consequently, high costs (from few hundred dollars for an acre in the case of surface or shallow targets up to a couple of million dollars for subsurface objects) occur. This is one of the main motivations of the growing research interest in developing unsupervised techniques able to effectively (in terms of time and resources) repair landmine/UXO-contaminated areas. Several solutions have been proposed based on various methodological approaches (e.g., see [2] and the references cited therein), which consider different sensor modalities such as ground-sensors (e.g., magnetometers, electromagnetic radars, sensors based on electromagnetic induction, etc.) or synthetic aperture radars. In general, these techniques are aimed at achieving the following goals: 1) correctly localizing a large number of dangerous targets, thus ensuring the future security of cleaned areas; 2) reducing the false-alarm rate, which strongly contributes to the costs of the clearing procedure; 3) reducing the time devoted to the detection process, thus realizing a *quick* area surveillance.

In such a framework, electromagnetic approaches based on learning-from-samples (LFS) techniques [3]–[5] have been recently proposed for the online [after the *learning process* (or *training phase*) performed once and off-line] detection of subsurface objects. The detection process is recast as a regression problem where the unknowns (i.e., the position as well as the geometric and dielectric characteristics of the target) are directly evaluated from the data (i.e., the values of the scattered field) by approximating the data-unknowns relation through an off-line data fitting process (*training phase*). LFS regression-based approaches demonstrated their effectiveness in dealing with detection processes where a limited number of unknowns (related to a single object) is considered. However, because of the complexity of the underlying architecture, some difficulties occur when a larger number of unknowns (related to multiple objects) is taken into account. Unfortunately, from a structural point of view, the regression technique does not permit one to simultaneously identify multiple positions. As a consequence, LFS regression-based approaches turn out to be very effective for the detection of a single (or few organized in a single cluster of objects) buried object, whereas they are not so suitable in dealing with the detection of multiple targets.

On the other hand, it should be pointed out that the identification of free-areas and an estimate of the concentration of subsurface objects (instead of the localization of each buried scatterer) might be enough in several situations. Then, the goal of a subsurface sensing technique could be moved from the "object detection" to the "definition of a risk map." Consequently, a classification approach, instead of a regression one, should be employed.

Within the framework of spatial statistics and remote sensing, relevant advances have been carried out in the last years for estimating a probability field. Let us consider the Markov random field Bayesian classifier (MRF-BC) or the autoregression models described in [6, ch. 7]. In this paper, a classification approach based on a LFS technique is proposed for an online subsurface sensing. Starting from the knowledge of the scattered field values collected above the surface, the method is aimed at defining a risk map of the domain under test. By considering a support vector machine (SVM)-based classifier, the proposed method estimates the *a posteriori* probability of the presence of subsurface dangerous objects.

The advantages, which can be also found better detailed in the related literature, of the use of such an instance-based classification method compared to more traditional optimization techniques can be summarized as follows: 1) no *a priori* knowledge about the system that generated the data is required, but

Manuscript received May 10, 2004; revised May 25, 2005.

The authors are with the Department of Information and Communication Technologies, University of Trento, 38050 Trento, Italy (e-mail: andrea.massa@ing.unitn.it).

Digital Object Identifier 10.1109/TGRS.2005.853186

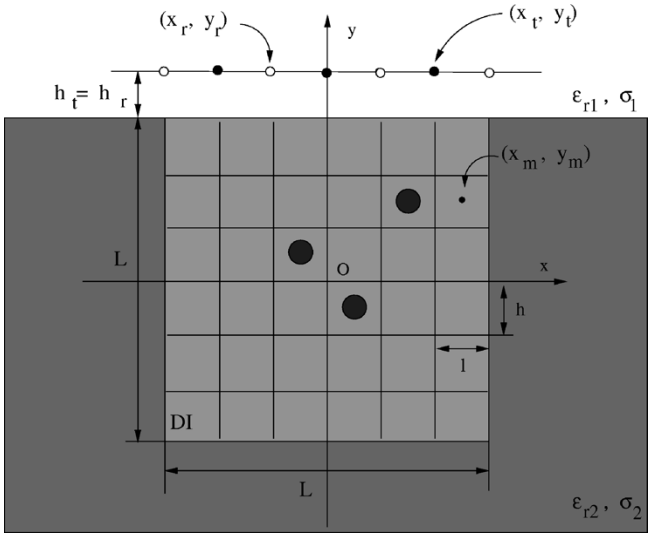


Fig. 1. Problem geometry.

only a set of input–output measurements; 2) the algorithm for the solution of the arising quadratic optimization problem with constraints is simple and reliable (no local minima/maxima occur); 3) SVMs are based on statistical learning theory by Vapnik and Chervonenkis, which permits one to design optimal classifiers with a solid theoretical framework; and 4) the estimation function is composed by a weighted sum of kernel functions, that can be easily implemented in hardware [7] for real-time applications.

The paper is organized as follows. The electromagnetic problem is formulated, and the SVM-based approach is outlined in Sections II and III, respectively. For the numerical validation, a selected set of results is presented to assess the effectiveness of the proposed approach (Section IV). Toward this end, a two-dimensional problem is dealt with. Both noiseless and corrupted measurement data are considered to check the robustness of the proposed approach. Finally, some conclusions and final remarks are presented in Section V.

II. Problem Formulation

Let us consider a typical two-dimensional buried-object scenario as shown in Fig. 1. The upper region presents the same characteristics of the vacuum ($\varepsilon_{r1} = 1.0$, $\sigma_1 = 0.0$). The lossy subsurface region, which models the soil, is characterized by a conductivity σ_2 and by a relative dielectric permittivity ε_{r2}. Moreover, let us assume that the investigation domain D_I lies entirely in the subsurface medium, $D_I = \{-L/2 \le x \le L/2, -L/2 \le y \le L/2\}$. A set of targets (either dielectric or lossy) are supposed to belong to D_I and illuminated by T transmitters located at known positions (x_t, y_t), $t = 1, \ldots, T$ above the air–ground interface. By considering underlining for vector notation, let $\underline{E}_{\text{inc}}$ be the so-called "incident field," i.e., the electric field distribution due to the transmitters in the absence of buried scatterers. The "scattered field" $\underline{E}_{\text{scat}}$ is collected by a set of sensors placed at given positions (x_r, y_r), $r = 1, \ldots, R$ close to the air–ground interface.

Under the hypotheses of an isotropic background medium and that the electromagnetic sources be z-directed electric line currents, both incident and scattered electric fields are also z-directed ($\underline{E}_{\text{inc}} = E_{\text{inc}}(x,y)\hat{z}$, $\underline{E}^{(t)}_{\text{scat}} = E^{(t)}_{\text{scat}}(x,y)\hat{z}$). In order to define a risk map, let us model the investigation domain through a two-dimensional lattice of M square cells of linear dimension l whose center coordinates are (x_m, y_m), $m = 1, \ldots, M$. The state χ_m of the mth cell can be either empty (if any scatterer belongs to the cell) $\chi_m = -1$ or occupied $\chi_m = +1$. Then, the problem can be formulated as follows: "*starting from the scattered field measurements* $\underline{\Gamma}_E = \left\{E^{(t)}_{\text{scat}}(x_r, y_r);\, r = 1, \ldots, R;\, t = 1, \ldots, T\right\}$, *find the probability* q_m *that the* m *– th cell is occupied* $(m = 1, \ldots, M)$." That is, *determine the probability array* $\underline{Q}$, which is a function $\Im$ of the scattering data $\underline{\Gamma}_E$

$$\underline{Q} = \Pr\left\{\underline{\chi} = \underline{1} \,|\underline{\Gamma}_E\right\} = \Im\left(\underline{\Gamma}_E\right) \tag{1}$$

where $\underline{Q} = \{q_m, m = 1, \ldots, M\}$ and $\underline{\chi} = \{\chi_m, m = 1, \ldots, M\}$. Such a statement defines a classification problem. A solution based on the SVM will be detailed in the following by assuming the knowledge of a set of known examples (i.e., input–output relations $\left\{\left(\underline{\Gamma}_E, m, \chi_m; m = 1, \ldots, M\right)^{(n)}; n = 1, \ldots, N\right\}$ called *training set*). For sake of clarity, let us denote by N the number of scattering configurations, $\Phi = M \times N$ being the number of patterns.

III. SVM-Based Classification Approach

The proposed SVM-based classification approach is formulated as the following two-step procedure:

- *Step 1:* determining a decision function $\hat{\Phi}$ that correctly classifies an input pattern $(\underline{\Gamma}_E, m)$ (not necessarily belonging to the training set);
- *Step 2:* mapping the decision function $\hat{\Phi}\left\{(\underline{\Gamma}_E, m)\right\}$ into an *a posteriori* probability $\Pr\left\{\underline{\chi} = \underline{1} \,|\underline{\Gamma}_E\right\}$.

A. Definition of the Decision Function

At this step, the status χ_m of each cell of the lattice has to be determined. Mathematically, such a problem formulates in the definition of a suitable discriminant function $\hat{\Phi}$ separating the two classes, which are labeled as $\chi = +1$ and $\chi = -1$. Since these classes are nonlinearly separable, the definition of a nonlinear (in terms of the original data $\underline{\Gamma}_E$) discriminant function is usually required as well as the solution of an optimization problem where multiple optima (also local optima) are present. As a matter of fact, such a solution is implemented when artificial neural networks (ANNs) are considered (see [8] and the references cited therein).

Unlike ANN, SVM defines a linear decision function corresponding to a hyperplane that maximizes the separating margin between the classes and it requires the solution of an optimization problem where only one minimum there exists. More in detail, the linear data-fitting is not carried out in the original input space $\Re\left\{\underline{\Gamma}_E\right\}$, but in a higher dimensional space $\aleph\left\{\underline{\varphi}\left(\underline{\Gamma}_E\right)\right\}$ (called *feature space*) where the original examples are mapped through a nonlinear operator,[1] $\underline{\varphi}(\bullet)$. The nonlinear SVM classifier so obtained is defined as

$$\hat{\Phi}\left(\underline{\varphi}\left(\underline{\Gamma}_E, m\right)\right) = \underline{w} \cdot \underline{\varphi}\left(\underline{\Gamma}_E, m\right) + b, \qquad m = 1, \ldots, M \tag{2}$$

[1]Because of the formulation of the problem at hand, it is easy to verify (10) that actually one does not need to know the $\underline{\varphi}(\bullet)$ function, but only its dot product in the feature space according to the so-called "*kernel trick*" [10].

where $\underline{w}$ and b are the parameters of $\hat{\Phi}$ to be determined during the training phase. The hyperplane so-defined causes the largest separation between the decision function values for the "margin" training examples from the two classes. Mathematically, such a hyperplane can be found by minimizing the following cost function

$$\Omega(\underline{w}) = \frac{1}{2}\|\underline{w}\|^2 \tag{3}$$

subject to the separability constraints

$$\begin{aligned} &\underline{w}\cdot\underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right)+b \geq +1 \text{ for } \chi_m^{(n)}=+1, \qquad m=1,\ldots,M \\ &\underline{w}\cdot\underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right)+b \leq -1 \text{ for } \chi_m^{(n)}=-1, \qquad n=1,\ldots,N. \end{aligned} \tag{4}$$

In this sense, SVM can be considered as a kind of regularized network, as indicated in [9].

However, since the training data in the feature space are generally noncompletely separable by a hyperplane, *slack variables* (denoted by $\xi_{(m)}^{(n)}$) are introduced to relax the separability constraints in (4) as follows:

$$\begin{aligned} &\underline{w}\cdot\underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right)+b \geq 1-\xi_{(m)+}^{(n)} \text{ for } \chi_m^{(n)}=1, \qquad m=1,\ldots,M \\ &\underline{w}\cdot\underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right)+b \leq \xi_{(m)-}^{(n)}-1 \text{ for } \chi_m^{(n)}=-1, \qquad n=1,\ldots,N. \end{aligned} \tag{5}$$

Such a procedure is justified by the Cover's theorem, a key point in the SVM methodology as indicated in [10, p. 200].

Thus, the cost function in (3) turns out to be

$$\Omega(\underline{w}) = \frac{\|\underline{w}\|^2}{2} + \frac{C}{\sum_{m=1}^{M}\left\{N_{(m)}^{-}+N_{(m)}^{+}\right\}} \times \sum_{m=1}^{M}\left\{\sum_{n=1}^{N_{(m)}^{+}}\xi_{(m)+}^{(n)} + \sum_{n=1}^{N_{(m)}^{-}}\xi_{(m)-}^{(n)}\right\} \tag{6}$$

where $N_{(m)}^{+}$ and $N_{(m)}^{-}$ indicate the number of training patterns for which $\chi_m^{(n)} = +1$ and $\chi_m^{(n)} = -1$, respectively. The user-defined hyperparameter C controls the tradeoff between the *empirical risk* (i.e., the training errors) and the model complexity [the first term in (7)] to avoid the *overfitting*. In that case, the decision boundary too precisely corresponds to the training data. Thereby, the method is unable to deal with data outside the training set [10, chs. 5 and 7].

Moreover, to include *a priori* knowledge about class distributions [11], two weighting constants can be defined $\lambda_+ = C/\sum_{m=1}^{M} N_{(m)}^{+}$ and $\lambda_- = C/\sum_{m=1}^{M} N_{(m)}^{-}$ [12], and (6) modifies as follows:

$$\Omega(\underline{w}) = \frac{\|\underline{w}\|^2}{2} + \lambda_+\sum_{m=1}^{M}\sum_{n=1}^{N_{(m)}^{+}}\xi_{(m)+}^{(n)} + \lambda_-\sum_{m=1}^{M}\sum_{n=1}^{N_{(m)}^{-}}\xi_{(m)-}^{(n)}. \tag{7}$$

In order to minimize (7), it can be observed that a necessary condition is that $\underline{w}$ is a linear combination of the mapped vectors $\underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right)$

$$\underline{w} = \sum_{m=1}^{M}\sum_{n=1}^{N}\left\{\alpha_m^{(n)}\chi_m^{(n)}\underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right)\right\} \tag{8}$$

where $\alpha_m^{(n)} \geq 0$, $n = 1,\ldots,N$, $m = 1,\ldots,M$ are Lagrange multipliers to be determined. Moreover, from the Karush–Khun–Tucker conditions at the optimality [13], b turns out to be expressed as follows:

$$b = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N_{sv}}\left\{\chi_m^{(n)} - \sum_{q=1}^{M}\sum_{p=1}^{N}\left\{\alpha_m^{(p)}\underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right)\cdot\underline{\varphi}\left(\underline{\Gamma}_E^{(p)}, q\right)\right\}\right\}}{N_{sv}} \tag{9}$$

N_{sv} being the number of patterns $\left(\underline{\Gamma}_E^{(n)}, m\right)$ for which $\alpha_m^{(n)} \neq 0$ (called *support vectors*). Since support vectors lie on the hyperplane for which (5) is satisfied with equality, they are taken into account for the classification while the others are neglected. Such an event reflects the "sparsity" property of the SVM classifier allowing the use of few input patterns.

Substituting (8) and (9) in (2) yields

$$\hat{\Phi}\left(\underline{\varphi}\left(\underline{\Gamma}_E\right), m\right) = \sum_{p=1}^{M}\sum_{n=1}^{N}\left\{\alpha_m^{(n)}\chi_m^{(n)}\Theta\left(\underline{\Gamma}_E^{(n)}, \underline{\Gamma}_E, p, m\right)\right\} + \frac{\sum_{m=1}^{M}\sum_{n=1}^{N_{sv}}\left\{\chi_m^{(n)} - \sum_{q=1}^{M}\sum_{p=1}^{N}\left\{\alpha_m^{(p)}\Theta\left(\underline{\Gamma}_E^{(n)}, \underline{\Gamma}_E^{(p)}, p, m\right)\right\}\right\}}{N_{sv}} \tag{10}$$

where $\Theta\left(\underline{\Gamma}_E^{(i)}, \underline{\Gamma}_E^{(j)}, p, m\right) = \underline{\varphi}\left(\underline{\Gamma}_E^{(i)}, p\right)\cdot\underline{\varphi}\left(\underline{\Gamma}_E^{(j)}, m\right)$ is a suitable *kernel function* [14]. Then, the decision function is completely determined when the Lagrange multipliers are computed. Toward this end, the constrained optimization problem formulated in (7) and (5) is reformulated in a more practical dual form. The solution of the dual problem, which is equivalent to the solution of the primal optimization problem (3)–(4), appears in (11), shown at the bottom of the page, subject to $\sum_{n=1}^{N}\sum_{m=1}^{M}\alpha_m^{(n)}\chi_m^{(n)} = 0$, $\alpha_m^{(n)} \in [0, \lambda_-]$ if $\chi_m^{(n)} = -1$ and $\alpha_m^{(n)} \in [0, \lambda_+]$ otherwise.

Finally, since $\Omega_{\text{Dual}}(\underline{\alpha})$ is a convex and quadratic function of the unknown parameters $\alpha_m^{(n)}$, it is solved numerically by means of a standard quadratic programming technique (e.g., the

$$\max_{\underline{\alpha}}\left\{\Omega_{\text{Dual}}(\underline{\alpha})\right\} = \max_{\underline{\alpha}}\left\{\frac{\sum_{n=1}^{N}\sum_{p=1}^{N}\sum_{m=1}^{M}\sum_{q=1}^{M}\left[\alpha_m^{(n)}\alpha_q^{(p)}\chi_m^{(n)}\chi_q^{(p)}\Theta\left(\underline{\Gamma}_E^{(n)}, \underline{\Gamma}_E^{(p)}, p, m\right)\right]}{2} - \sum_{n=1}^{N}\sum_{m=1}^{M}\alpha_m^{(n)}\right\} \tag{11}$$

Platt's SMO algorithm for classification [15][2]). More in detail, the SMO algorithm breaks the large optimization problem at hand in a series of smaller ones characterized by only two variables and solved through an effective updating formula [15], thus inducing nonnegligible computational savings.

B. Mapping of the Decision Function Into the A Posteriori Probability

Concerning standard classification, the SVM classifier labels an input pattern according to the following rule [16]:

$$\chi_m = sign\left\{\hat{\Phi}\left(\underline{\varphi}\left(\underline{\Gamma}_E, m\right)\right)\right\}, \qquad m = 1, \ldots, M. \quad (12)$$

Unlike standard approaches, the proposed method is aimed at defining an *a posteriori* probability. Consequently, some modifications to the standard SVM-based classification approach are needed. Toward this aim, a set of efficient solutions has been proposed (e.g., see [14], [17]–[19]) either based on a direct training of the SVM with a logistic link function and a regularized maximum-likelihood score or based on *a posterior* fitting probability process.

The first class of approaches usually leads to nonsparse kernel machines and requires a significant modification of the SVM structure. In this paper, the *a posteriori* probability fitting method [19] is adopted since the use of a parametric model allows a direct fitting of the *a posteriori* probability $\Pr\left\{\underline{\chi} = \underline{1} \,|\underline{\Gamma}_E\right\}$. More in detail, such a model approximates the *a posteriori* probability through a sigmoid function

$$\Pr\left\{\chi_m = 1 \,|(\underline{\Gamma}_E, m)\right\} = \frac{1}{1 + \exp\left\{\gamma\hat{\Phi}\left(\underline{\varphi}\left(\underline{\Gamma}_E, m\right)\right) + \delta\right\}}, \qquad m = 1, \ldots, M \quad (13)$$

where γ and δ are unknown parameters to be determined.

To estimate the optimal values for the parameters of the sigmoid function, a fitting process is performed. A subset of the input patterns of the training set is chosen $\{(\underline{\Gamma}_E, m,\ \chi_m; \quad m = 1, \ldots, M)^{(s)};\ \ s = 1, \ldots, S\}$, where $\hat{\Phi}_m^{(s)} = \hat{\Phi}\left(\underline{\varphi}\left(\underline{\Gamma}_E^{(s)}, m\right)\right)$. Then, the following cost function is defined as in (14), shown at the bottom of the page, and successively minimized to define γ and δ according to the numerical procedure proposed by Lin *et al.* (see http://www.csie.ntu.edu.tw/~cjlin/)[3] to solve the problems (i.e., the use of a kind of Levenberg–Marquardt method for unconstrained optimization) of the implementation of Platt's probabilistic outputs method pointed out in [19].

Summarizing, the SVM optimization problem needs three successive steps: 1) determining the hyperparameters array (*model selection*), i.e., C and all the parameters that define the kernel function (e.g., the Gaussian width σ^2 when Gaussian kernels are used), by considering the *"training dataset"*; 2) determining the *functional parameters* $\underline{\alpha}$ and b starting from the *"training dataset"* and solving the dual problem (11); 3) determining the *a posteriori fitting parameters* γ and δ starting from a subset of the *"training dataset"* (*validation phase*); and 4) testing the SVM on a different dataset (*test phase*).

[2]An optimal implementation of the SMO algorithm is the "LibSVM" tool available at http://www.kernel-machines.org.

[3]Available at http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/.

IV. Numerical Validation

In this section, after a detailed description of the various steps carried out during the experimental validation, the results of a set of numerical examples are analyzed to assess the effectiveness, but also current limitations, of the proposed approach.

A. Reference Scattering Configuration

With reference to the geometry shown in Fig. 1, the following geometric and dielectric parameters are considered. The relative permittivity and the conductivity of the homogeneous subsurface region are $\varepsilon_{r2} = 4.0$ and $\sigma_2 = 1 \times 10^{-3}$ S/m, respectively. The investigation domain D_I is a 2.0 $\lambda\times$ 2.0 λ region, and it was discretized in a two-dimensional lattice of $M = 36$ square cells. The buried objects (modeling UXOs or landmines) are lossless circular cylinders of diameter $d_{\rm cil} = \lambda/6$ and characterized by a relative permittivity $\varepsilon_{\rm cil} = 5.0$. Concerning the measurement system, $R = 16$ ideal receivers are equally spaced along an observation line 2.0 λ-long and parallel to the air–ground interface at a distance $h_r = 0.6\ \lambda$ from the surface. The probing source ($t = 1$) is located at $x_t = 0.0$ and $y_t = 7/6\ \lambda$. The measurement data were synthetically computed by using a finite-element-based simulator and a perfectly matched layer truncation technique [20].

B. Training and Test Datasets

The choice of representative datasets is an open problem to be carefully addressed in order to limit its influence to the final result in terms of reliability as well as generality.

As pointed out in [21], the sources of variations in statistical validation tests can be identified in: 1) the randomness in the selection of test data; 2) the randomness in selecting training data; 3) the internal randomness of the training algorithm; and 4) the mislabeling of the dataset.

The former source of variation is of particular concern when the test set is a small fraction of the entire dataset. In this case, the variability due to randomly selecting the test set could be problematic. The source defined in 2) could influence the estimate of the classification. As a matter of fact, the training set is a particular realization of the input patterns. The third source of variation 3) can affect learning algorithms that depend on the random starting point (e.g., the backpropagation networks). However, the learning algorithm of a SVM is independent from its initialization, since the solution of the associated optimization problem is the global minimum of the error function regardless of the starting point. The mislabeling of the dataset is a source of errors for any learning algorithm, nevertheless SVMs have an intrinsic mechanism for dealing with misclassified data points [22], [23].

$$\Upsilon\left\{\gamma, \delta\right\} = -\sum_{s=1}^{S}\sum_{m=1}^{M}\left\{\frac{\chi_m^{(s)} + 1}{2}\log\left[\frac{1}{1 + \exp\left(\gamma\hat{\Phi}_m^{(s)} + \delta\right)}\right] + \left(\frac{1 - \chi_m^{(s)}}{2}\right)\log\left[\frac{\exp\left(\gamma\hat{\Phi}_m^{(s)} + \delta\right)}{1 + \exp\left(\gamma\hat{\Phi}_m^{(s)} + \delta\right)}\right]\right\} \quad (14)$$

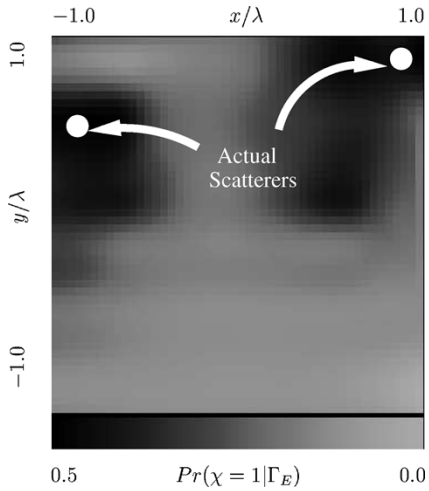


Fig. 2. Noiseless data. Experiment 1. Color level representation of the risk map for the two-target scenario [$(x_{\text{cil}}^{(1)} = -5/6\ \lambda,\ y_{\text{cil}}^{(1)} = \lambda/2)$ and $(x_{\text{cil}}^{(2)} = y_{\text{cil}}^{(2)} = 5/6\ \lambda)$].

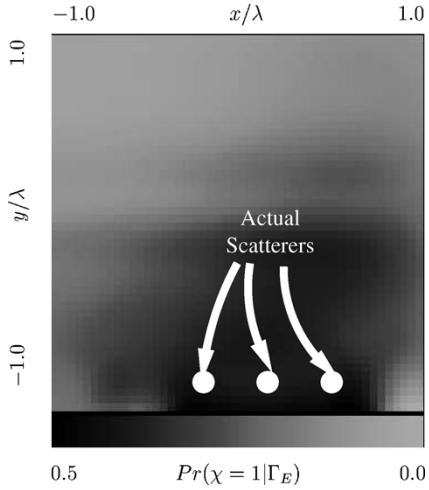


Fig. 3. Noiseless data. Experiment 1. Color level representation of the risk map for the three-target scenario [$(x_{\text{cil}}^{(1)} = -\lambda/6,\ y_{\text{cil}}^{(1)} = -5/6\ \lambda)$, $(x_{\text{cil}}^{(2)} = \lambda/6,\ y_{\text{cil}}^{(2)} = -5/6\ \lambda)$, and $(x_{\text{cil}}^{(3)} = \lambda/2,\ y_{\text{cil}}^{(3)} = -5/6\ \lambda)$].

TABLE I
NUMERICAL EXPERIMENTS 1, 2, AND 3. STATISTICS OF THE DANGEROUS-AREA-LOCALIZATION ERROR, ς

Experiment	*No. 1*	*No. 2*	*No. 3*		
SNR [dB]	∞	∞	20	10	5
$min_p\{\varsigma^{(p)}\}$	0.065	0.151	0.139	0.193	0.053
$max_p\{\varsigma^{(p)}\}$	1.024	1.046	0.995	0.992	0.936
$av_p\{\varsigma^{(p)}\}$	0.312	0.385	0.347	0.358	0.378
$var_p\{\varsigma^{(p)}\}$	0.077	0.098	0.074	0.075	0.073

According to such considerations and following some of the arising guidelines to avoid both incorrect and unreliable results, different datasets were generated. As far as the *training set* is concerned, various scattering configurations $N_{(\text{train})} = N_{(\text{train})}^{(2)} + N_{(\text{train})}^{(3)}$ ($N^{(2)}$ and $N^{(3)}$ being the numbers of scattering configurations with two and three-targets, respectively) were considered. The locations of the scatterers in D_I for the "*training*" dataset as well as for the "*test*" dataset were randomly chosen in a discrete grid of points imposing that

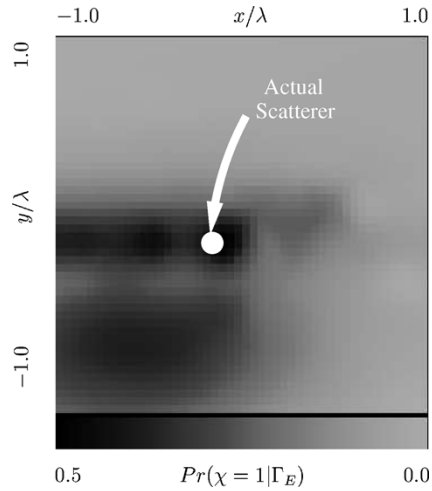


Fig. 4. Noiseless data. Experiment 2. Color level representation of the risk map for the single-target scenario $(x_{\text{cil}}^{(1)} = y_{\text{cil}}^{(1)} = -\lambda/6)$.

TABLE II
NUMERICAL EXPERIMENT 2. COMPARATIVE ASSESSMENT BETWEEN SVM-BASED REGRESSION APPROACH AND SVM-BASED CLASSIFICATION APPROACH. VALUES OF THE DANGEROUS-AREA-LOCALIZATION ERROR, ς, FOR DIFFERENT POSITIONS OF THE ACTUAL SCATTERER

	SVM − based Approach					
SNR [dB]	*Regr.*			*Clas.*		
∞	0.086	0.211	0.015	0.059	0.400	0.189
35	0.103	0.153	0.063	0.073	0.289	0.308
20	0.116	0.056	0.072	0.151	0.515	0.458
10	0.101	0.192	0.097	0.106	0.621	0.346
5	0.077	0.297	0.096	0.221	0.516	0.520
Test Case	*(a)*	*(b)*	*(c)*	*(a)*	*(b)*	*(c)*

TABLE III
NUMERICAL EXPERIMENT 3. ESTIMATE OF THE ERROR IN EVALUATING THE AREA OF THE DANGEROUS ZONE

SNR [dB]	∞	35	20	10
$\frac{\hat{\Delta}^{(1)}}{\Delta^{(1)}}$	2.023	3.230	3.249	4.456
$\frac{\hat{\Delta}^{(2)}}{\Delta^{(2)}}$	2.760	3.249	3.278	4.752

the objects, belonging to the same scattering example, cannot be located at the same position. More in detail, the coordinates of the grid points $\left[x_{(t)}^{pq},\ x_{(t)}^{pq}\right]$ were defined according to the following rule:

$$\begin{cases} x_{(t)}^{pq} = x_{(t)}^{0} + (p-1)\,\Delta x_{(t)}, & p = 1, \ldots, P_{(t)} \\ y_{(t)}^{pq} = y_{(t)}^{0} + (q-1)\,\Delta y_{(t)}, & q = 1, \ldots, Q_{(t)} \end{cases} \tag{15}$$

where the subscript t indicates the "training" (train) or the "test" (test), respectively; $x_{(\text{train})}^{0} = -y_{(\text{train})}^{0} = -5\lambda/6$, $\Delta x_{(\text{train})} = \Delta y_{(\text{train})} = \lambda/3$, $P_{(\text{train})} = Q_{(\text{train})} = 6$, and $x_{(\text{test})}^{0} = -y_{(\text{test})}^{0} = -11\lambda/12$, $\Delta x_{(\text{test})} = \Delta y_{(\text{test})} = \lambda/12$, $P_{(\text{test})} = Q_{(\text{test})} = 24$, respectively.

As far as the following experiments are concerned, if it is not specified, $N_{(\text{train})}^{(2)} = 35$ and $N_{(\text{train})}^{(3)} = 34$, thus $\Phi_{(\text{train})} = \Phi_{(\text{train})}^{(2)} + \Phi_{(\text{train})}^{(3)}$ patterns being $\Phi_{(\text{train})}^{(3)} = 1260$ and $\Phi_{(\text{train})}^{(3)} = 1224$, respectively. These patterns were also used during the *validation test* aimed at defining the *a posteriori* fitting model. The optimal values of the fitting parameters turned out to be $\gamma = -0.533$ and $\delta = 1.272$, which demonstrated to be appropriate for the whole numerical assessment.

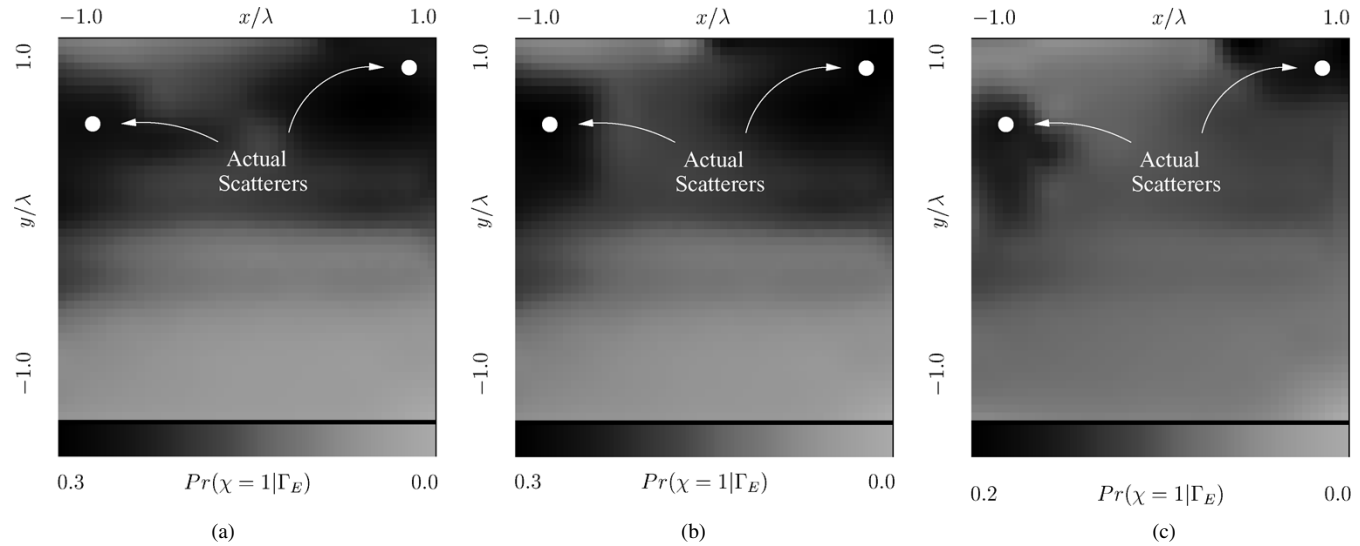


Fig. 5. Noisy data. Experiment 3. Risk maps for the two-target scenario [$(x_{\text{cil}}^{(1)} = -5/6\ \lambda,\ y_{\text{cil}}^{(1)} = \lambda/2)$ and $(x_{\text{cil}}^{(2)} = y_{\text{cil}}^{(2)} = 5/6\ \lambda)$] when (a) SNR $= 35$ dB, (b) SNR $= 20$ dB, and (c) SNR $= 10$ dB.

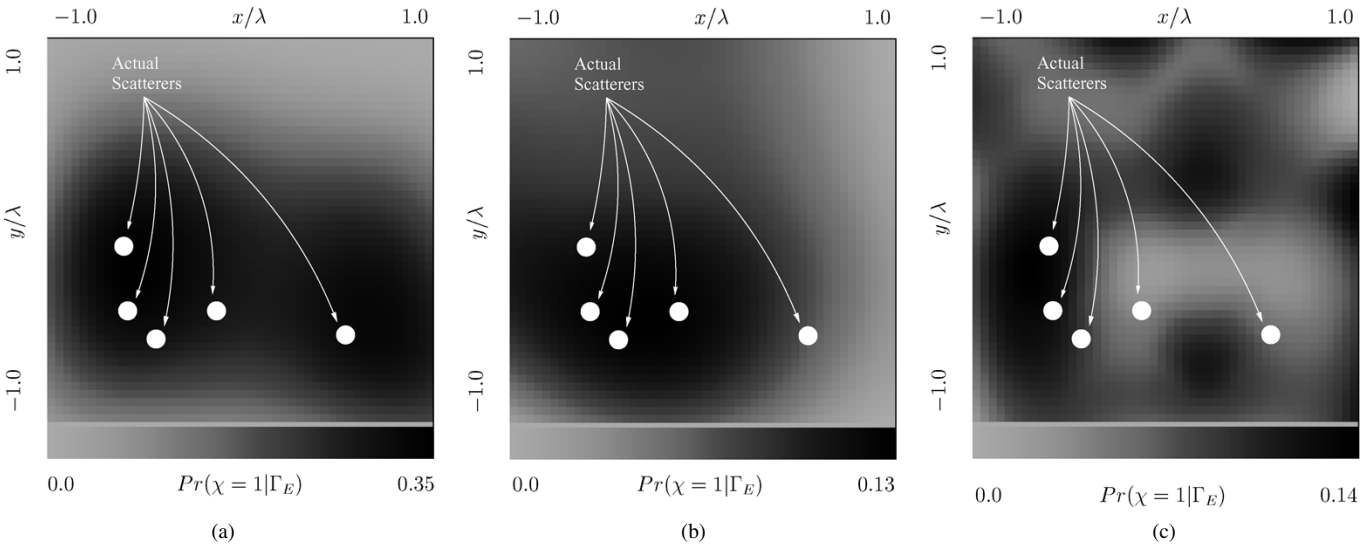


Fig. 6. Noiseless data. Experiment 4. Example (*i*). Risk maps obtained with different classification approaches: (a) SVM, (b) RBF, and (c) MLP. Buried objects are located at $(x_{\text{cil}}^{(1)} = -7/12\ \lambda,\ y_{\text{cil}}^{(1)} = -\lambda/12)$, $(x_{\text{cil}}^{(2)} = -7/12\ \lambda,\ y_{\text{cil}}^{(2)} = -5/12\ \lambda)$, $(x_{\text{cil}}^{(3)} = -\lambda/6,\ y_{\text{cil}}^{(3)} = -5/12\ \lambda)$, $(x_{\text{cil}}^{(4)} = -5/12\ \lambda,\ y_{\text{cil}}^{(4)} = -7/12\ \lambda)$, $(x_{\text{cil}}^{(5)} = 7/12\ \lambda,\ y_{\text{cil}}^{(5)} = -7/12\ \lambda)$.

C. Model Development

The selection of the appropriate SVM for solving a particular classification task is still an open problem. While the parameters of a SVM can be easily found by solving a quadratic programming problem, there are many proposals for identifying its hyperparameters (e.g., the kernel parameter or the regularization factor: C, σ^2, or p), but it is not clear yet which one is superior to the others. This task is usually denoted as the "*model selection problem*."

The model selection problem is strictly related to the evaluation of the generalization ability of the SVM. In fact, it is common use to select the optimal SVM (i.e., the optimal hyperparameters) by choosing the one with the lowest generalization error.

However, there has been some criticism on this approach, since the true generalization error obviously cannot be computed, thus it is needed to refer to an upper bound of its value. Minimizing an upper bound of the error rate can be misleading and the actual value can be quite different from the actual one. On the other hand, an upper bound of the generalization error, if correctly derived, is of paramount importance for estimating the applicability of the SVM to a particular classification task, especially on a real-world problem.

Several methods for model selection have been proposed (see [23] for a survey). They can be roughly classified in two distinct classes: bounds-based and data-based. The former are based on exact theorems derived from statistical learning theory [24], while the latter are based on estimations obtained by considering sets of measurements of the problem under investigation. They include the test set method or statistical intensive approaches, such as the k-fold cross validation, the leave one out and the bootstrap one. Despite the drawback of reducing the size of the training set for building an independent test set, in most cases this is the only way to avoid overly optimistic estimates. On the other hand, the most advanced

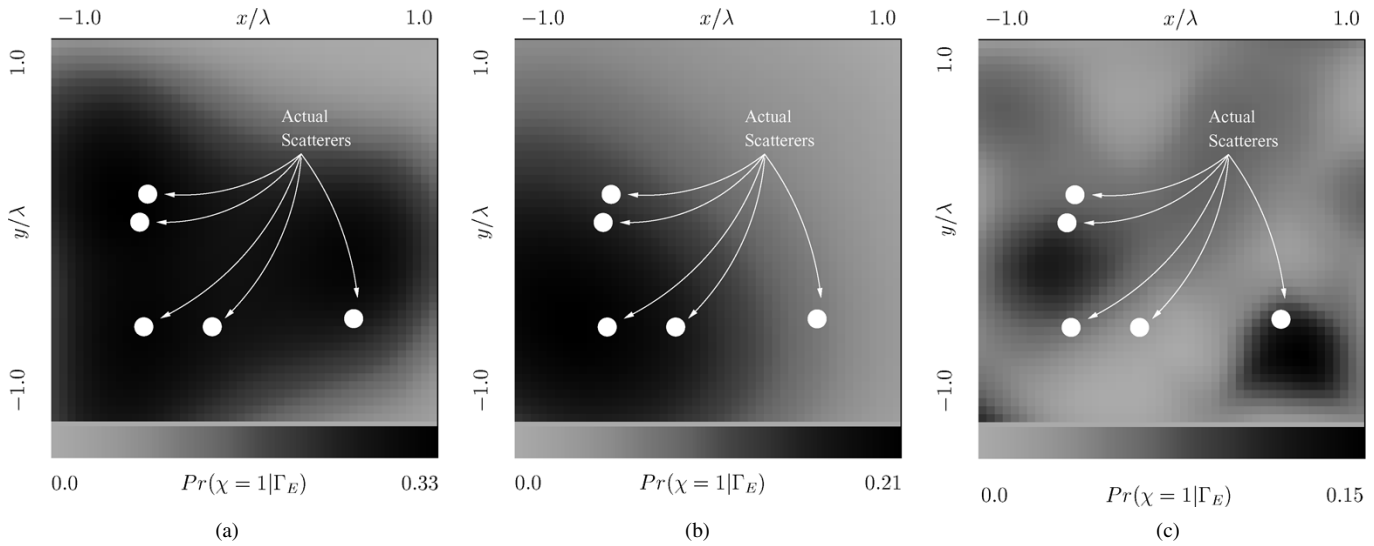


Fig. 7. Noiseless data. Experiment 4. Example (*ii*). Risk maps obtained with different classification approaches: (a) SVM, (b) RBF, and (c) MLP. Buried objects are located at $(x_{\text{cil}}^{(1)} = -\lambda/2, y_{\text{cil}}^{(1)} = \lambda/6)$, $(x_{\text{cil}}^{(2)} = -7/12\ \lambda, y_{\text{cil}}^{(2)} = 0.0)$, $(x_{\text{cil}}^{(3)} = -\lambda/2, y_{\text{cil}}^{(3)} = -\lambda/2)$, $(x_{\text{cil}}^{(4)} = -\lambda/6, y_{\text{cil}}^{(4)} = -\lambda/2)$, $(x_{\text{cil}}^{(5)} = 7/12\ \lambda, y_{\text{cil}}^{(5)} = -5/12\ \lambda)$.

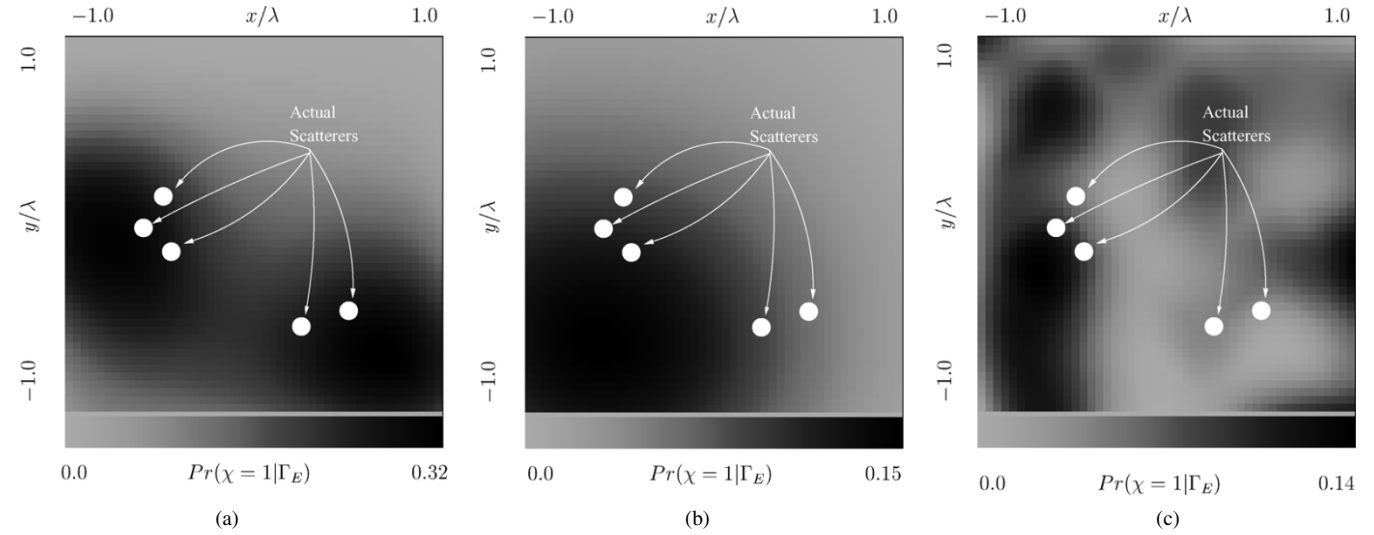


Fig. 8. Noiseless data. Experiment 4. Example (*iii*). Risk maps obtained with different classification approaches: (a) SVM, (b) RBF, and (c) MLP. Buried objects are located at $(x_{\text{cil}}^{(1)} = -\lambda/2, y_{\text{cil}}^{(1)} = \lambda/6)$, $(x_{\text{cil}}^{(2)} = -7/12\ \lambda, y_{\text{cil}}^{(2)} = 0.0)$, $(x_{\text{cil}}^{(3)} = -5/12\ \lambda, y_{\text{cil}}^{(3)} = -\lambda/6)$, $(x_{\text{cil}}^{(4)} = \lambda/4, y_{\text{cil}}^{(4)} = -7/12\ \lambda)$, $(x_{\text{cil}}^{(5)} = \lambda/2, y_{\text{cil}}^{(5)} = -\lambda/2)$.

methods try to estimate the generalization error directly from the empirical error. They represent the state of the art of the research in machine learning, even though their practical effectiveness is yet to be verified [25]. The bounds-based methods theoretically are the best because one can execute the model selection and at the same time estimate the generalization error. However, in many cases the bounds can be very loose and report a generalization error much greater than 1 (actually, any value greater than 0.5 is completely useless). As experimentally shown, the minimum of the upper bound often corresponds to the best hyperparameter choice, so its value can be of use for the model selection, if not for estimating the true generalization error.

On the other hand, since Gaussian kernels have shown good properties in solving classification problems [26] and their use instead of sigmoidal, linear, or polynomial kernels is justified by reduced numerical difficulties as well as to the fact that they are described with only one characteristic parameter (i.e., the Gaussian width), Gaussian kernel functions were adopted in this work

$$\Theta\left(\underline{\Gamma}_E^{(i)}, \underline{\Gamma}_E^{(j)}, p, m\right) = \exp\left\{-\left[\frac{\left\|\left(\underline{\Gamma}_E^{(i)}, p\right) - \left(\underline{\Gamma}_E^{(j)}, m\right)\right\|^2}{2\sigma^2}\right]\right\}. \tag{16}$$

Thus only two hyperparameters were tuned (i.e., C and σ^2) following the method of the test set proposed in [23]. More in detail, the optimal hyperparameters in a set of discrete samples are those values for which the generalization error is minimum. Accordingly, C was varied in the range $[1, 10^6]$ and σ^2 from 0.01 up to 10.0. After such a process, the optimal values of the hyperparameters turned out to be $C = 10^4$ and $\sigma^2 = 1.0$.

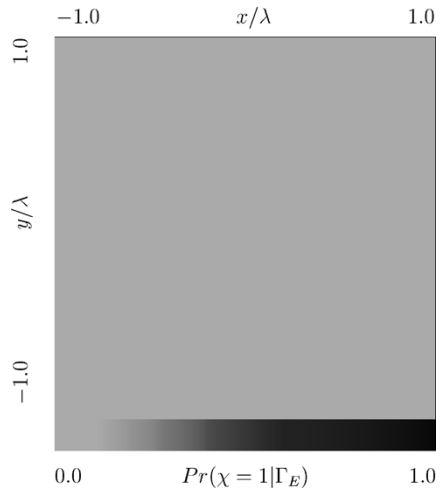


Fig. 9. Noisy data. Experiment 5. Risk map when no objects are hidden under the surface and $\Phi^{(0)}_{(\mathrm{train})} \in \Phi_{(\mathrm{train})}$.

D. Numerical Assessment

In order to analytically evaluate the effectiveness of the classification method in correctly locating the dangerous areas, a *dangerous-area-location error* ς is then defined

$$\varsigma = \frac{\sqrt{(x_{\mathrm{cil}} - \widetilde{x}_{\mathrm{cil}})^2 - (y_{\mathrm{cil}} - \widetilde{y}_{\mathrm{cil}})^2}}{\lambda} \tag{17}$$

where $(x_{\mathrm{cil}},\, y_{\mathrm{cil}})$ and $(\widetilde{x}_{\mathrm{cil}},\, \widetilde{y}_{\mathrm{cil}})$ are the actual and estimated coordinates of the center of a dangerous zone, respectively, being

$$\widetilde{x}_{\mathrm{cil}} = \frac{\sum_{m=1}^{M_{(j)}} \{x_m \Pr(\chi_m = 1 \,|\underline{\Gamma}_E)\}}{\sum_{m=1}^{M_{(j)}} \{\Pr(\chi_m = 1 \,|\underline{\Gamma}_E)\}}$$
$$\widetilde{y}_{\mathrm{cil}} = \frac{\sum_{m=1}^{M_{(j)}} \{y_m \Pr(\chi_m = 1 \,|\underline{\Gamma}_E)\}}{\sum_{m=1}^{M_{(j)}} \{\Pr(\chi_m = 1 \,|\underline{\Gamma}_E)\}} \tag{18}$$

where $M_{(j)}$ indicates the number of connected cells where $\Pr(\chi_m = 1 \,|\underline{\Gamma}_E) \neq 0$. Moreover, the extension of the estimated dangerous zone $\widetilde{\Delta}$ is defined as follows:

$$\widetilde{\Delta} = \pi \left\{ \sum_{m=1}^{M_{(j)}} \left\{ \frac{\rho_m \Pr(\chi_m = 1 | \underline{\Gamma}_E)}{\max_m \{\Pr(\chi_m = 1 | \underline{\Gamma}_E)\}} \right\} \right\}^2 \tag{19}$$

where $\rho_m = \sqrt{(x_m - \widetilde{x}_{\mathrm{cil}})^2 - (y_m - \widetilde{y}_{\mathrm{cil}})^2}$.

Within the numerical validation, the first experiment deals with a *test set* of $\Phi_{(\mathrm{test})} = 2484$ patterns (related to scattering configurations different from those of the training phase and concerned with two- and three-scatterers configurations, $N^{(2)}_{(\mathrm{test})} = 35$ and $N^{(3)}_{(\mathrm{test})} = 34$) and noiseless conditions. Figs. 2 and 3 show the risk maps obtained for two examples of the test set. The first example (Fig. 2) refers to a two-targets configuration where the scatterers are located at ($x^{(1)}_{\mathrm{cil}} = -5/6\ \lambda$, $y^{(1)}_{\mathrm{cil}} = \lambda/2$) and ($x^{(2)}_{\mathrm{cil}} = y^{(2)}_{\mathrm{cil}} = 5/6\ \lambda$). In such a case, the values of the error figures turn out to be equal to $\varsigma^{(1)} = 0.291$ and $\varsigma^{(2)} = 0.389$. Moreover, the highest values of the occurrence probability are very close to the actual positions of the scatterers. As far as the dimensions of the two dangerous zones (the objects being not-adjacent) are concerned, they are slightly over-estimated ($\widetilde{\Delta}^{(1)}/\Delta^{(1)} = 2.023$ and $\widetilde{\Delta}^{(2)}/\Delta^{(2)} = 2.760$, $\Delta^{(1)} = \Delta^{(2)} = (1/9)\lambda^2$).

The second example (Fig. 3) is related to a three-scatterers configuration. The objects are adjacent and lie at the bottom of the investigation domain. As expected, when the targets are buried far from the surface, the localization of the "dangerous zones" is more difficult. In spite of this, the approach is still able to localize these areas with an acceptable degree of accuracy ($\varsigma^{(1)} = 0.359$, $\varsigma^{(2)} = 0.385$, and $\varsigma^{(3)} = 0.502$).

For completeness, by considering the whole test-set, the statistics of the *dangerous-area-location error* ς are given in the second column of Table I.

The second numerical experiment considers a more critical test scenario where a single target is supposed to be located in the investigation domain [$N^{(1)}_{(\mathrm{test})} = 36$, thus $\Phi^{(1)}_{(\mathrm{test})} = 1296$]. It should be pointed out that such a configuration does not belong to the training set. Concerning the effectiveness of the approach in dealing with this kind of test set, the localization statistics are reported in Table I (second column). As expected, the error figure increases as compared to the first experiment and the average value changes from $av_p\left\{\varsigma^{(p)}\right\}\big|_{\mathrm{Exp.1}} = 0.312$ to $av_p\left\{\varsigma^{(p)}\right\}\big|_{\mathrm{Exp.2}} = 0.385$.

As an example, the risk map for a sample of the test set ($x^{(1)}_{\mathrm{cil}} = y^{(1)}_{\mathrm{cil}} = -\lambda/6$) is shown in Fig. 4 ($\varsigma^{(1)} = 0.364$ and $\widetilde{\Delta}^{(1)}/\Delta^{(1)} = 4.718$).

For comparison purposes, the results obtained with the SVM-based classification approach were compared with authors prior approach based on regression [5] in terms of localization accuracy of the actual scatterer. Toward this end, three different positions of the actual object were considered: 1) ($x^{(1)}_{\mathrm{cil}} = \lambda/4$, $y^{(1)}_{\mathrm{cil}} = -\lambda/3$); 2) ($x^{(1)}_{\mathrm{cil}} = (3/4)\lambda$, $y^{(1)}_{\mathrm{cil}} = -(3/4)\lambda$); and 3) ($x^{(1)}_{\mathrm{cil}} = \lambda/4, y^{(1)}_{\mathrm{cil}} = -(3/4)\lambda$). As expected, the SVM-based regression approach overcomes the classification one in locating the scatterer as confirmed by the values of ς are reported in Table II for the noiseless case as well as for a noisy conditions.

As a matter of fact, since the SVM, as formulated by Vapnik and based on statistical learning theory, is a LFS technique used when no information (such as noise model and level, probability density function, etc.) on the problem at hand there exist, the SVM cost function is designed to work in different operative cases and various working conditions. Thus, to further evaluate the robustness of the proposed approach, also a noisy environment might be considered. Therefore, in the third experiment, corrupted measurement data were simulated by adding a uniform Gaussian noise (characterized by different values of the signal-to-noise ratio (SNR) in the range from 5 dB up to 35 dB in order to satisfactorily simulate realistic conditions for the electric field measurement) to synthetic data [5] of the test set used in the first experiment. Such a choice, instead of others forms of noise, should be considered as a first/reference attempt in a more general framework for giving a simple, although representative, indication on the behavior of the SVM-based approach in real situations.

To further confirm the effectiveness of the proposed approach in dealing with a noisy environment, the risk maps

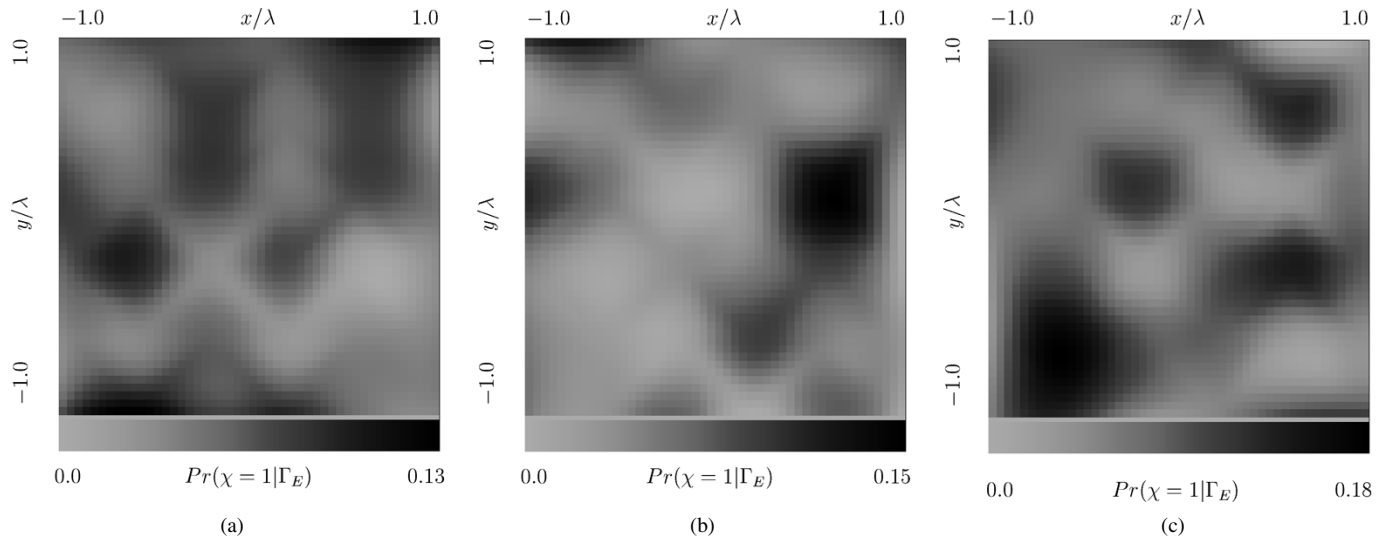


Fig. 10. Experiment 5. Risk maps when no objects are hidden under the surface and $\Phi^{(0)}_{(\text{train})} \notin \Phi_{(\text{train})}$. (a) Noiseless conditions. (b) SNR $=$ 10 dB. (c) SNR $=$ 5 dB.

for the two-targets configuration (where the actual scatterers are located as in Fig. 2) and for different SNRs are shown in Fig. 5 as representative examples. It can be noticed that, when SNR $\geq$ 20 dB the "contaminated zones" are quite correctly detected and located ($\varsigma^{(1)}\big|_{\text{SNR=35 dB}} = 0.383$, $\varsigma^{(2)}\big|_{\text{SNR=35 dB}} = 0.462$ and $\varsigma^{(1)}\big|_{\text{SNR=20 dB}} = 0.389$, $\varsigma^{(2)}\big|_{\text{SNR=20 dB}} = 0.464$). Otherwise, the performance of the approach reduces ($\varsigma^{(1)}\big|_{\text{SNR=10 dB}} = 0.526$, $\varsigma^{(1)}\big|_{\text{SNR=10 dB}} = 0.580$) even though the highest values of the occurrence probability are situated in correspondence with and close to the target positions.

As far as the prediction of the extension of the dangerous zones is concerned, Table III shows that the effectiveness of the approach reduces with the increasing of the SNR as well as, on average, for the localization accuracy (Table I, last three columns). This causes an overestimate of the dangerous areas since, due to the presence of the noise, the accuracy of the classification process reduces as well as the range of the estimated probability.

The fourth experiment (Experiment 4) is aimed at confirming the validity and the advantages of the SVM-based classification approach over some other standard nonlinear methods as regularized form of the familiar multilayer perceptron (MLP) or radial basis functions (RBF) neural networks. Toward this purpose and to face a more complex multiple-scatterers scenario, a data test set characterized by $\Phi_{(\text{test})} = 3600$ patterns related to different scattering configurations of five buried objects randomly distributed in D_I was considered. The classifiers were trained with $\Phi_{(\text{train})} = \Phi^{(2)}_{(\text{train})} + \Phi^{(3)}_{(\text{train})} + \Phi^{(5)}_{(\text{train})}$ patterns being $\Phi^{(2)}_{(\text{train})} = 1260$, $\Phi^{(3)}_{(\text{train})} = 1224$, and $\Phi^{(5)}_{(\text{train})} = 1008$, respectively. Figs. 6–8 show, in a comparative fashion, the results obtained with different classifiers and in correspondence with a set of representative examples. More in detail, the first example (*i*) refers to a configuration where two different clusters of objects can be recognized (i.e., the one with four scatterers and the other with one scatterer; Fig. 6). In the second example [Fig. 7, (*ii*)], three clusters of objects are present, while example (*iii*) deals with a two-clusters setup where the objects are equally distributed. Whatever the example, the SVM-based approach is able to identify each cluster with a satisfactory probability value ($\Pr\{\chi_m = 1 \,|\underline{\Gamma}_E\} \cong 0.35$ in the neighboring of the actual locations of the buried scatterers). On the other hand, the performance of the RBF-based classifier turn out to be inferior in terms of cluster-localization as well as estimated probability value [$\max_m [\Pr\{\chi_m = 1 \,|\underline{\Gamma}_E\}]]_{\text{RBF}} \cong 0.13$—example (*i*); $\max_m [\Pr\{\chi_m = 1 \,|\underline{\Gamma}_E\}]]_{\text{RBF}} \cong 0.21$—example (*ii*); $\max_m [\Pr\{\chi_m = 1 \,|\underline{\Gamma}_E\}]]_{\text{RBF}} \cong 0.15$—example (*iii*)]. As far as the MLP-based approach is concerned, the classification results are not accurate and several difficulties occur in the location of the actual scatterers [Figs. 6(c), 7(c), and 8(c)].

Finally, the last experiment (Experiments 5) deals with the scattering scenario where there is no landmine at all hidden under the surface, but some noise is added to the data. Firstly, the SVM-based classifier was trained with a training set containing some patterns related to the free-object scenario ($\Phi_{(\text{train})} = \Phi^{(2)}_{(\text{train})} + \Phi^{(3)}_{(\text{train})} + \Phi^{(0)}_{(\text{train})}$, $\Phi^{(2)}_{(\text{train})} = 2412$, $\Phi^{(3)}_{(\text{train})} = 2412$, and $\Phi^{(0)}_{(\text{train})} = 10$).[4] As expected, whatever the noisy conditions, the classifier did not detected a buried scatterer (Fig. 9). Otherwise, when the free-scenario patterns were neglected and in presence of a nonnegligible level of noise (SNR $\leq$ 10 dB), the classification was not adequate as shown in Fig. 10(b) and in Fig. 10(c), respectively. For completeness, also the risk map for the noiseless condition is reported in Fig. 10(a).

V. Conclusion

In this paper, a classification approach for subsurface sensing of multiple buried targets has been proposed. A suitable SVM-based strategy has been developed for determining the probability of occurrence of buried targets and to define a "risk map" of the investigation domain.

The effectiveness of the approach has been preliminarily assessed by considering a two-dimensional geometry and noiseless as well as noisy conditions. The obtained results confirmed the ability of the method in detecting and locating multiple targets as well as in estimating the extension of the dangerous zones.

[4]Let us consider that the complete set of patterns related to the free-objects scattering scenario should be of 36 different elements since $N^{(0)}_{(\text{train})} = 1$.

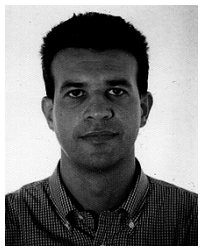
Future works, current under development, will be devoted to fully exploit the key-features of the approach as well as to assess its reliability in dealing with real experiments and three-dimensional scenarios. Moreover, as far as the hardware implementation [7] is concerned, several advances would be needed to allow an on-site testing.

ACKNOWLEDGMENT

The authors would like to thank S. Piffer and M. Conci for constant technical support as well as constructive comments. The first author would also like to thank E. Vico for her support. Moreover, the authors are indebted to the anonymous reviewers for their constructive comments, which significantly improved this paper

REFERENCES

[1] "Mine facts," in *CD ROM-Based Encyclopedia of Current World-Wide Mine Technology*. Washington, DC: OASD (SO/LIC) Acquisition, 2500 Defence Pentagon, 1997.
[2] "Special Issue on: "New Advances in Subsurface Sensing: Systems, Modeling and Signal Processing,"," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, Jun. 2001.
[3] S. Caorsi, D. Anguita, E. Bermani, A. Boni, M. Donelli, and A. Massa, "A comparative study of NN and SVM-based electromagnetic inverse scattering approaches to on-line detection of buried objects," *J. Appl. Comput. Electromagn. Soc.*, vol. 18, no. 2, pp. 1–11, Jul. 2003.
[4] I. T. Rekanos, "Inverse scattering of dielectric cylinders by using radial basis function neural networks," *Radio Sci.*, vol. 36, no. 5, pp. 841–849, 2001.
[5] E. Bermani, A. Boni, S. Caorsi, and A. Massa, "An innovative real-time technique for buried object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 927–931, Apr. 2003.
[6] S. Shekhar and S. Chawla, *Spatial Databases: A Tour*. Englewood Cliffs, NJ: Prentice-Hall, 2003.
[7] D. Anguita, A. Boni, and S. Ridella, "A digital architecture for support vector machines: Theory, algorithm and FPGA implementation," *IEEE Trans. Neural Netw.*, vol. 14, no. 5, pp. 993–1009, Sep. 2003.
[8] C. Christodoulou and M. Georgiopoulos, *Application of Neural Networks in Electromagnetics*. Norwood, MA: Artech House, 2001.
[9] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines," in *Advances in Large Margin Classifiers*, A. J. Smola, P. L. Bartlett, B. Scholkopf, and D. Schuurmans, Eds, MA: MIT Press, 2000.
[10] B. Schölkopf and A. Smola, *Learning with Kernels*. Cambridge, MA: MIT Press, 2002.
[11] Y. Lin, Y. Lee, and G. Wahba, "Support vector machines for classification in nonstandard situations," Dept. Statistics, Univ. Wisconsin, Madison, Tech. Rep. 1016, 2000.
[12] K. Morik, P. Brockhausen, and T. Joachims, "Combining statistical learning with a knowledge-based approach: A case study in intensive care monitoring," presented at the *16th Int. Conf. Machine Learning*, 1999.
[13] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[14] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Wiley, 1999.
[15] J. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods—Support Vector Learning*, B. Scholkopf, C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA: MIT Press, 1999.
[16] K. -R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, Mar. 2001.
[17] G. Wahba, "Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV," in *Advances in Kernel Methods—Support Vector Learning*, B. Scholkopf, C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA: MIT Press, 1999.
[18] T. Hastie and R. Tibshirani. (1996) Classification by pairwise coupling. Stanford Univ., Stanford, CA. [Online]Tech. Rep.
[19] J. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods," in *Advances in Large Marging Classifiers*, A. J. Smola, P. Bartlett, B. Scholkopf, and D. Schuurmans, Eds. Cambridge, MA: MIT Press, 1999.
[20] S. Caorsi and M. Raffetto, "Perfectly matched layer for the truncation of finite element meshes in layered half-space geometries and applications to electromagnetic scattering by buried objects," *Microw. Opt. Technol. Lett.*, vol. 19, pp. 427–434, Dec. 1998.
[21] T. G. Dietterich, "Approximate statistical test for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, pp. 1895–1923, Oct. 1998.
[22] C. Cortes and V. N. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[23] D. Anguita, A. Boni, S. Ridella, F. Rivieccio, and F. Sterpi, "Theoretical and practical model selection methods for support vector classifiers," in *Support Vector Machines: Theory and Applications*, L. Wang, Ed. Berlin, Germany: Springer-Verlag, 2005, vol. 177, Studies in Fuzziness and Soft Computing Series, ch. 7.
[24] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
[25] D. Anguita, S. Ridella, F. Rivieccio, and R. Zunino, "Hyperparameter design criteria for support vector classifiers," *Neurocomputing*, vol. 55, pp. 109–134, Oct. 2003.
[26] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.*, vol. 15, pp. 1667–1689, Jul. 2003.



Andrea Massa (M'03) received the laurea degree in electronic engineering and the Ph.D. degree in electronics and computer science from the University of Genoa, Genoa, Italy, in 1992 and 1996, respectively.

From 1997 to 1999, he was an Assistant Professor of electromagnetic fields at the Department of Biophysical and Electronic Engineering, University of Genoa, teaching the university course of Electromagnetic Fields 1. Since 2000, he has been an Associate Professor at the University of Trento, Trento, Italy, where he teaches the courses "Electromagnetic Fields 1," "Electromagnetic Techniques for Biomedical and Industrial Diagnostics," and "Wireless Communications 1." At the University of Trento, he has constituted a new research group in electromagnetic fields (currently made up over ten people working in different fields of electromagnetics) and has cooperated in designing a complete educational program in telecommunications engineering (B.S. and M.S. degrees, International Ph.D. School). He is currently the Director of the ELEDIALab, University of Trento and the Deputy Dean of the International Ph.D. School in Information and Communications Technologies. His research work since 1992 has been principally on electromagnetic direct and inverse scattering, microwave imaging, optimization techniques, wave propagation in presence of nonlinear media, wireless communications, and applications of electromagnetic fields to telecommunications, medicine, and biology.

Dr. Massa is a member of the PIERS Technical Committee, the European Microwave Association (EuMA), and the Inter-University Research Center for Interactions Between Electromagnetic Fields and Biological Systems (ICEmB).

Andrea Boni was born in Genoa, Italy, in 1969. He received the laurea degree in electronic engineering and the Ph.D. degree in electronic and computer science from the University of Genoa, Genoa, in 1996 and 2000, respectively.

After working as Research Consultant at the Department of Biophysical and Electronic Engineering, University of Genoa, he joined the Department of Information and Communication Technologies, University of Trento, Trento, Italy, where he teaches digital electronics. His main scientific interests are the study and development of digital circuits for advanced information processing, with particular attention to programmable logic devices, digital signal theory and analysis, statistical signal processing, statistical learning theory, and support vector machines. The application of such interests focuses on identification and control of nonlinear systems, pattern recognition, time series forecasting, and signal processing.

Massimo Donelli received the laurea degree in electronic engineering and Ph.D. degree in space science and engineering from the University of Genoa, Genoa, Italy, in 1998 and 2003, respectively.

He is currently with the Department of Information and Communication Technology, University of Trento, Trento, Italy. His main interests are electromagnetic inverse scattering, adaptive antennas synthesis, optimization techniques for microwave imaging, wave propagation in superconducting materials, and urban environments.