An end-user-oriented framework for the classification of multitemporal SAR images

ARTICLE *in* INTERNATIONAL JOURNAL OF REMOTE SENSING · JANUARY 2016

Impact Factor: 1.65

5 AUTHORS, INCLUDING:

Donato Amitrano

University of Naples Federico II

21 PUBLICATIONS **41** CITATIONS

SEE PROFILE

Gerardo Di Martino

University of Naples Federico II

87 PUBLICATIONS **216** CITATIONS

SEE PROFILE

Antonio Iodice

University of Naples Federico II

244 PUBLICATIONS **1,568** CITATIONS

SEE PROFILE

To appear in the *International Journal of Remote Sensing*
Vol. 00, No. 00, Month 20XX, 1–13

An end-user oriented framework for the classification of multitemporal SAR images

D. Amitrano[a]*, G. Di Martino[a], A. Iodice[a], D. Riccio[a] and G. Ruello[a]

[a] *University of Napoli Federico II, Department of Electrical Engineering and Information Technology*

(*Received 00 Month 20XX; accepted 00 Month 20XX*)

In this paper, we present an end-user-oriented framework for multitemporal synthetic aperture radar (SAR) data classification. It accepts as input the recently introduced Level-1α products, whose peculiarities are the high degree of interpretability and the increased class separability with respect to single grayscale images. These properties make the Level-1α products very attractive for the application of simple supervised classification algorithms. In fact: i) the high degree of interpretability of the maps makes the training phase extremely simplified; ii) the good separation between the classes allows for obtaining excellent results using simple discrimination rules. It results in a simple, fast, accurate and repeatable framework.

1. Motivations

Since the beginning of the synthetic aperture radar (SAR) history, the scientific community aimed to extract information from data with the purpose to classify objects on the imaged scenes. One of the first definitions for this activity was given in Harger (1973): "Classification consists in the discrimination and identification of randomly reflecting areas, characterized by a certain reflectivity spectral density". This definition highlights two aspects: i) *Discrimination*, i.e. the separation of objects which share similar characteristics of backscattering at the operating wavelength. This operation is carried out at a physical/signal level; ii) *Identification*, i.e. the process of the assignment of a label (or a class) to a group of scatterers identified as similar. This operation is carried out at a semantic level and is powered by the knowledge of the SAR backscattering mechanism.

A more articulate definition of classification is provided in Campbell and Wynne (2011): "Classification is the assignment of objects [...] to classes based on their appearance on the imagery. Often a distinction is made between three levels of confidence and precision. *Detection* is the determination of the presence or absence of a feature. *Recognition* implies [...] that the object can be assigned an identity in a general class or category. Finally, *Identification* means that the identity of an object or feature can be specified with enough confidence and detail to place it in a very specific class". This definition put the accent on the *appearance* of the objects, and thus on physical models which allows for the interpretation of the world representation, i.e. the image, as filtered by the sensor.

*Corresponding author. Email: donato.amitrano@unina.it

In recent years, the increased volume of available data and calculating power raised the user requests concerning classification procedures. Thus, while in '70s and '80s the rules were essentially dictated by the SAR community and by research needs, today the pressure on this activity comes from a wide variety of professionals and scientists belonging to different disciplines who posed new user-requirements to the SAR community.

In this context, several outstanding solutions were found. Some authors exploited multitemporal datasets for extracting time-dependent variables used to discriminate the behavior of the scene features (Bruzzone et al. 2004). This allowed to devise robust classification schemas, but the difficulties related to the implementation of the processing chain could limit their usage. Camps-Valls et al. (2008) couples multitemporal SAR and optical data to set up a framework based on composite kernels. Davidson and Ouchi (2003) improved a segmentation scheme with the use of multitemporal data. In Engdahl and Hyyppa (2003), the computation of some multitemporal quantities allows to reduce the dimension of the original dataset, selecting the best-suited group of images for unsupervised classification. In Quegan et al. (2000), backscattering characteristics of a time series have been analyzed for forest mapping. In Gaetano et al. (2014) a Markov random field framework for multitemporal data segmentation using an homomorphic transformation was applied.

All these works used multitemporal data in order to improve the performances of previously introduced schema or to establish new methods of data fusion. This allowed to better identify features which are assumed to be stable (i.e. the forest mapping presented in Quegan et al. (2000)), or which hold a specific behavior along time with respect to a given multitemporal quantity (see Davidson and Ouchi (2003)). Here, we want to focus on the simplification of the classification schema, in which multitemporal data are exploited for building a product suitable to be treated with standard algorithms.

In general, given an application, different approaches lead to different solutions and different answers to the following questions: 1) What is the best algorithm/technique to reach the goal? 2) Which parameters must be tuned for obtaining a more reliable output?

The answers to these questions dramatically change as a function of the expertise of the operator.

As for the first question, hundreds of algorithms/techniques with excellent results exist. An expert user could use his/her skills to select the best solutions for the given application. A non-expert end-user look toward solutions already implemented in some available software suite. Moreover, most of the existing methods suffer from the problem that both the end-user and the SAR expert operator have a limited control/vision of the operations that lead to the final products.

As for the second question, the great complexity of modern classification algorithms require non trivial expertise in the phase of tuning the required parameters.

However, the active participation of end-users in classification procedures represent an opportunity to improve the performances in terms of compliance to requirements of practical scenarios. This can be possible dulling the hurdles that they can encounter answering to the two aforementioned questions. Hence, it is necessary to include simple algorithms in a framework in which the knowledge/expertise required for classifying a scene is diluted and made accessible to the end-users. This is feasible via an extreme simplification of the classification schemas, starting from an adequate choice of the input data. This position is in countertrend with the past literature in which the development of new algorithms/techniques rather then the

reasoning on input data is privileged.

This paper is devoted to show that the recently introduced Level-1α products (Amitrano et al. 2015), based on the physical knowledge of electromagnetic scattering models and required parameters, allows for the building of such a framework. It gives the opportunity to use simple, well-known user-oriented classification tools which, thanks to the characteristics Level-1α imagery, provide excellent results on complex scenes. Thus, our framework satisfies the fundamental end-user requirements of simplicity, velocity, accuracy and repeatability.

2. Multitemporal Level-1α products

In this Section, we briefly recall the rationale of the recently introduced Level-1α products (Amitrano et al. 2015). These products are obtained from high resolution Single Look Complex (SLC) SAR data, by properly combining two opportunely processed intensity channels and the interferometric coherence of the complex data involved in the composition.

In order to show the characteristics of these products, we use a particularly illustrative example relevant to a RGB composition of a semi-arid environment.

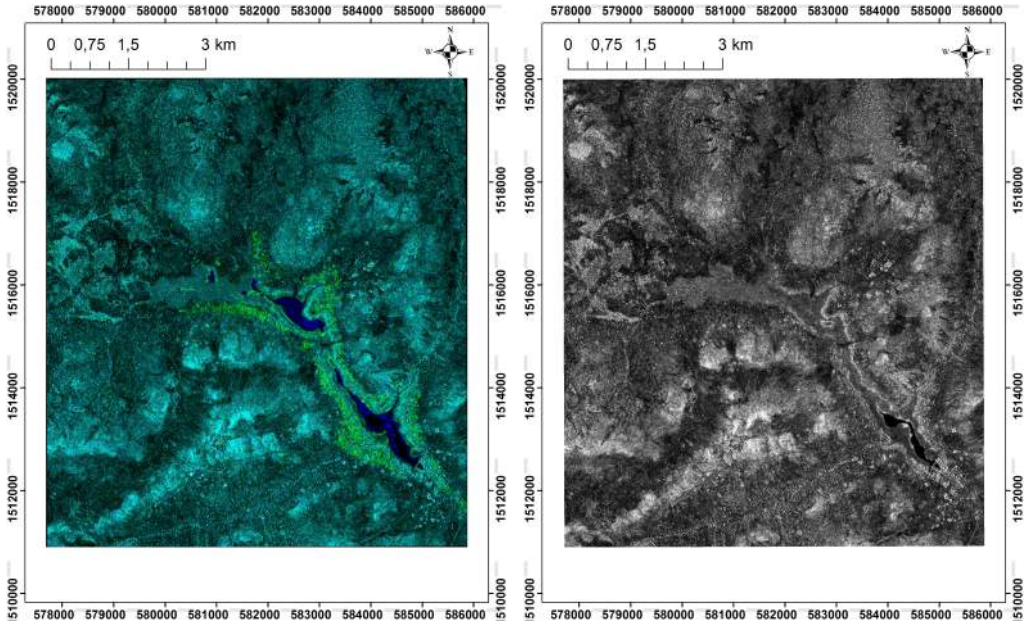
Figure 1(a) shows a Level-1α product of a rural area of Northern Burkina Faso. This area is characterized by a Sudanese-Sahelian climate, which consists of a brief rainy season and a severe dry season at the peak of which (towards the end of April) the environment is almost completely dry (Amitrano et al. 2014). This information is important, since it will allow to explain the colorimetric response of the Level-1α product.

The product depicted in Figure 1(a) is built loading on the green and blue band two images acquired on March 2011 and April 2011, respectively. The acquisition of March 2011 is also depicted in Figure 1(b). This comparison trivially highlights the advantages of a color composite representation instead of a grayscale one in terms of interpretability of the map. In fact, as an example, the reader can immediately associate the green color with vegetation or the blue with surface water.

For the ease of the reader, in the following we briefly summarize the results obtained in Amitrano et al. (2015) and describe how to interpret Level-1α products.

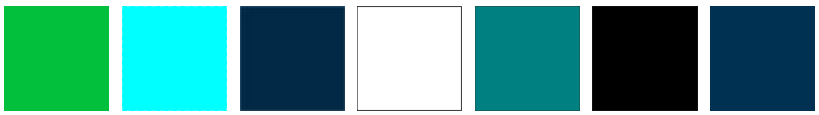
The interpretation of the map is guided by the physical principles that govern the interaction between the incident electromagnetic wave and the observed surface. For instance, with the aid of Figure 2 and Figure 3, the following qualitative classification can be made:

- Blue color indicates the presence of seasonal water (see Figure 3(a)). In fact, the blue color witnesses a strong dominance in the electromagnetic response of the image of April with respect to the March's one, which is negligible. Therefore, in April (blue band) the basin is almost empty and the electromagnetic response is due to bare soil, while in March (green band), because of the presence of surface water, that response is very low;
- Black color indicates that the response is low in both the intensity images (e.g., in the immediate nearby of the dam) and this testifies the presence of surface water which persists in both the acquisitions (see Figure 3(a));
- Balance between the two intensity channels means that no changes have occurred between the two acquisitions. Hence, this characteristic is proper of stable features. Of course, the intensity level at which the two channels are balanced identifies different objects. For example, cyan indicates trees (see Figure 3(a)). A balance at medium intensity level of the two primary



(a) (b)

Figure 1. Tougou basin, (a) COSMO-SkyMed Level 1α product -Blue band: 28 April 2011 (dry season); Green band: 27 March 2011 (dry season); Red band: interferometric coherence- and (b) 27 March 2011 intensity map. Spatial resolution 3 m $\times$ 3 m.



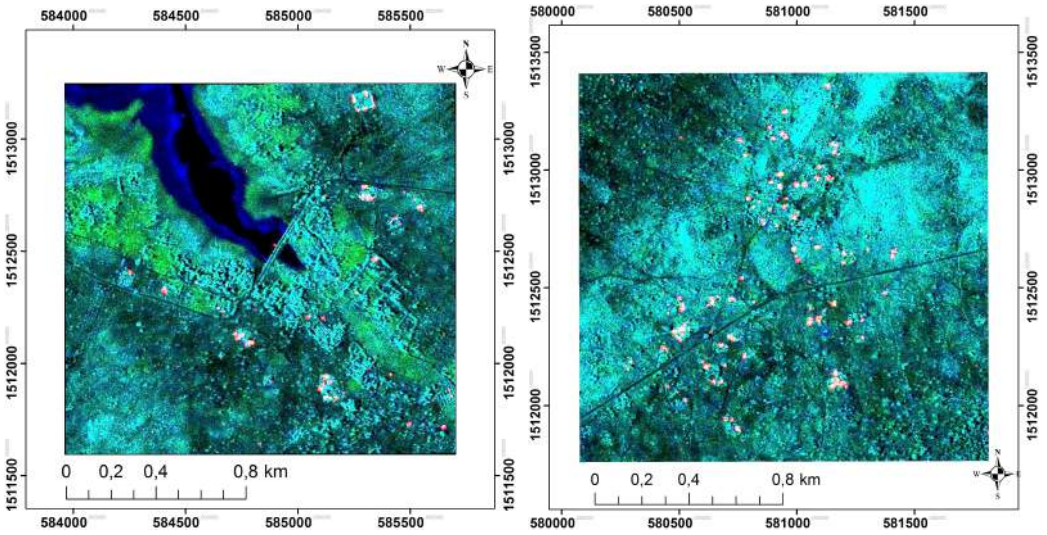
(a) (b) (c) (d) (e) (f) (g)

Figure 2. Table of colors associable with the classes identifiable on the RGB composition depicted in Figure 1(a). In particular: (a) Dark pastel green ("Grasses"), (b) Cyan ("Trees"), (c) Oxford blue ("Seasonal water"), (d) White ("Man-made scatterers"), (e) Tea leaf ("Rocks", "Humid soil"), (f) Black ("Weak scatterers"), (g) Prussian blue ("Bare dry soil").

colors, i.e. tea leaf color (see Figure 2), marks humid soils and rocks. Finally, the balance at low levels of green and blue (with tonalities which could be identified with Prussian blue or Dark green) indicates bare soil. Since in Figure 1(a) the two images have been both acquired in the dry season, the dominant color is cyan in its various gradation, i.e. soil with various percentage of moisture (see Figure 3(a) and Figure 3(b));

- Tonalities of green color indicate low vegetation and crops (see Figure 3(a)). These areas can be observed only in the immediate nearby of the dam and are identified by the zones in which the electromagnetic response of the March's image (green band) is dominant with respect to April's (blue band) thanks to the backscattering enhancement due to volumetric effects triggered by leafs (Fung 1979);
- Bright targets indicate man-made objects (see Figure 3(a) and Figure 3(b)), having a high contribution of all the bands which form the RGB composition. These features would be indistinguishable from trees or other highly reflective natural targets without taking into account of interferometric coherence.

As explained in Amitrano et al. (2015), the main characteristic of Level-1α imagery is that the association color-object (being physical-based) is stable for variation of climatic conditions and scene structure. Therefore, the above consideration



(a) (b)

Figure 3. Details of the Level-1α product depicted in Figure 1(a): (a) a rural area in the nearby of a dam and (b) a small human settlement.

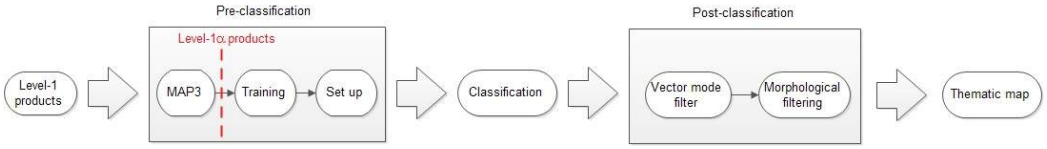


Figure 4. Block diagram of the proposed classification framework. The pre-classification block is detailed in Section 3.1. The post-classification phase is discussed in Section 3.3. In particular, the vector mode filter implementation is described in 3.3.1; a solution for morphological filtering is provided in 3.3.2. Results of the classification framework are discussed in 4.

hold for an image acquired on an urbanized scene in temperate environment as the one we will analyze in Section 4, given a proper selection of the reference image, which is guided by the application (Amitrano et al. 2015), such as the detection of seasonal crops.

3. Definition of the classification framework

In Figure 4 the block diagram of the proposed framework is shown. It is articulated in three steps, pointing to the redistribution of the complexities from the classification phase to those of pre- and post-classification. This simple architecture is possible thanks to the usage of the Level-1α products which are characterized by i) a high level of interpretability and ii) a good separability between the observable classes. The former characteristic makes these products very attractive for supervised classification procedures, since their interpretability facilitates the training phase. The latter characteristic allows to carry out the classification procedure with simple algorithms which require only few parameters to be set.

The architecture presented in Figure 4 is rather standard, at least in the first two stages (named as “pre-classification” and “classification” in the picture). Therefore, it is poorly suited for being applied to SAR data. However, the introduced novelties, concerning the input products for the classifier and the post-processing phase (in which two morphological operators are used) allows for obtaining performances fully comparable with those of definitely more complex algorithms. It is worthwhile

to remark that Level-1α imagery is obtained by multitemporal processing of N images, even if each product involves only two of them, as stated in Amitrano et al. (2015). As all multitemporal processing, the quality of the products increases as the number of the available images increases.

The proposed approach looks towards temporal series of classification maps, introducing a novelty with respect to the past literature in which typically aimed at producing a single classification map by integrating the information acquired by N acquisitions considered as stand-alone events. Nevertheless, when N acquisitions of the same scene are available, a series of $N-1$ products and classification maps can be retrieved. This could be useful if a long-term environmental monitoring is required, since the changes between one acquisition and another can be recorded with a programmed revisit time. In the case of archive data the entire extent of the time series can be exploited to perform historic investigations and/or statistics. At semantic level, the detectable classes have a different meaning with respect to the classic SAR literature. In fact, our framework brings to a labeling which depends on the comparison of objects backscattering with respect to a pre-established reference situation, rather than on the temporal behavior of the scene with respect to some multitemporal quantities.

The block diagram presented in Figure 4 is quite generic. In the following subsections we delineate a schema which allows to reduce drastically the pre-classification activities necessary for supervised classification procedure and, in general, introduces significant simplification in all the other blocks of the chain.

3.1. *Pre-classification*

Level-1α products defined by MAP3 (Amitrano et al. 2015) are particularly well-suited for supervised classification procedures. In fact, these products have the peculiarity of introducing a good separation between the classes, which can be identified using simple algorithms. Moreover, the training phase is simplified thanks to their high degree of interpretability, which allows to identify objects by simple and fast visual inspection. Conversely, this operation could be very complex and tedious when considering a single channel image, especially if the operator has no a priori knowledge about the scene.

The supervised approach requires the user to select representative training data for each of a predefined number of classes. Classification performance is highly dependent on users ability in modeling the class distribution. With this respect, the users experience is crucial for identifying and locating the best training areas that, ideally, should be homogeneous and known, or, at least, recognizable (Townshend 1981). The introduction of the Level-1α products strongly mitigates these uncertainty factors since they can be easily managed and interpreted also by nonexpert users.

3.2. *Classification*

In our framework the complexity of the classification algorithm is redistributed toward the phases of pre- and post-classification, allowing to avoid the usage of complicated decision rules. Therefore, in this work we used the maximum likelihood (ML) classifier. This choice was guided by the simplicity of the set up and by their wide availability on commercial and open-source suites.

However, it is worthwhile to note that due to the strong inhomogeneity of the urban areas we exploit the high separability of this feature characteristic of the

Level-1α products using an interferometric coherence-derived mask. In fact, as explained in Richards (1999), the ML classification used in this work assumes that the statistics for each class in each band are normally distributed and calculates the probability that a given pixel belongs to a specific class. In Level-1α imagery, it is not possible to assume that the urban area has this type of distribution, being it very heterogeneous. Therefore, we used the interferometric coherence, which is represented in the red band of the Level-1α product, for identifying the features belonging to this class. In fact, it is well known that urban structures have typically a high value of interferometric coherence, therefore they can be easily detected by thresholding this quantity.

3.3. *Post-classification*

The output of the classification procedure could be mottled, especially if complex scenes are analyzed. Hence, an adequate post classification step is required in order to obtain a more homogeneous output map. In this work, we propose a simple post classification phase composed of two steps: i) application of a mode filter and ii) of a morphological filter that generates the final output. These activities are described in detail in the following paragraphs.

3.3.1. The vector mode filter

The mode filter is a reliable solution for filtering the output of the selected classifier, obtaining homogeneous classification maps with minimum alteration of its informative content.

The mode filter substitutes the central pixel of a sliding window with the mode computed within the window itself. However, the execution of that filter must take into account three requirements: simplicity, accuracy and velocity. In fact, the implementation of a sliding window running throughout the whole scene implies a high computational burden, which has to be carefully managed in order reduce the execution times. As a matter of fact, for a $m \times n$ image and a square sliding window of dimension l (with l odd), the number of windows to be evaluated is $(n - l + 1)(m - l + 1)$. Hence, for example, if the scene is 5000×5000 pixels and the sliding window is a three pixels square side, the number of windows to be processed is 24980004.

A solution for implementing the sliding windows particularly suitable for vectorial languages is to exploit the Hankel indexing (Partington 1989). To this end, assign to each element of the scene a linear index, for example proceeding along the rows of a $m \times n$ matrix, obtaining the matrix $\mathbf{I}$:$\mathbf{I}_{i,j} = j + (j - i)(i - 1) + n(i - 1), i = [1, \ldots, m], j = [1, \ldots, n]$. The purpose is to process the sliding window in vector form, i.e. to build an index matrix $\mathbf{W}$ in which each row collects the elements of a window.

In order to get this matrix, consider the Hankel-like matrix of dimensions $k \times l$, where $k = n - l + 1$. This matrix is such that its generic element a is such that $a_{i,j} = a_{i-1,j+1}$.

The matrix $\mathbf{H}$ collects on each row the first l elements of the sliding windows which involve the first row of the matrix $\mathbf{I}$. Therefore, we can place side by side to the $\mathbf{H}$ matrix l columns with an index jump of n with respect to the elements $H_{i,l}$. This operation has to be repeated $l - 1$ times, i.e. once for each remaining row of the sliding window. Hence, the intermediate matrix $\mathbf{T}$ has dimensions $k \times l^2$ and it is obtained as in the following pseudo-code:

FOR $i = 1, l$ DO
$\mathbf{T}(*, il\boldsymbol{u_1} + \boldsymbol{v_1}) = \mathbf{H} + ni\mathbf{h}$,
END

where $\boldsymbol{v_1}$ is a vector, $\boldsymbol{v_1} = [1, 2, \ldots, l]$, $\boldsymbol{u_1}$ is a vector of dimensions $[1 \times l]$ whose elements are all equal to 1, and $\mathbf{h}$ is a matrix whose elements are all equal to 1, i.e. $\mathbf{h}$:$\mathbf{h}_{i,j} = 1, \forall i \in [1, \ldots, k], \forall j \in [1, \ldots, l^2]$. We indicated with the asterisk $\mathbf{T}(*, \ldots)$ that the operation involves all the row of the matrix $\mathbf{T}$ using a notation typical of many programming languages.

The matrix $\mathbf{T}$ represents the sliding windows for the first l rows of the matrix $\mathbf{W}$. The last step is to replicate this matrix by rows k times inserting an indexes jump of l every k rows as follows:

FOR $j = 1, k$ DO
$\mathbf{W}(jk\boldsymbol{u_2} + \boldsymbol{v_2}, *) = \mathbf{T} + nj\mathbf{t}$,
END

where $\boldsymbol{v_2}$ is a vector $\boldsymbol{v_2} = [1, 2, \ldots, k]$, $\boldsymbol{u_2}$ is a vector of dimensions $[1 \times k]$ whose elements are all equal to 1, and $\mathbf{h}$ is a matrix whose elements are all equal to 1, i.e. $\mathbf{t}$:$\mathbf{t}_{i,j} = 1, \forall i \in [1, \ldots, nw], \forall j \in [1, \ldots, l^2]$. Each row of the matrix $\mathbf{W}$ is composed by the linear position index of the pixels belonging to a sliding window. The matrix dimensions are $nw \times l^2$, where $nw = n \times (l-1)/2 \times m \times (l-1)/2$, with l odd. Applying the indexes matrix obtained in such a way to the input image, we get the sliding windows in vector form in which the mode has to be computed. We implemented an optimized procedure to accomplish this task, which is structured as below:

- Let N_C be the number of classes present within the classification map and build a three dimensional matrix $\mathbf{C}$ of dimensions $nw \times l^2 \times N_C$ such that

$$\mathbf{C}(*, *, i) = \mathbf{Z} + i\mathbf{z}, \qquad i = 1, 2, , N_C, \tag{1}$$

  where $\mathbf{Z}$ is a matrix in which all the elements are equal to zero red and $\mathbf{z}$ is a matrix of the same dimensions of the matrix $\mathbf{Z}$ whose elements are all equal to 1. The matrix $\mathbf{C}$ is composed by a series of matrices in which all the elements are equals between them and equal to the position index of the element in the third dimension of $\mathbf{C}$. Hence, $\mathbf{C}(*, *, 1) = \mathbf{Z} + \mathbf{z}$, $\mathbf{C}(*, *, 2) = \mathbf{Z} + 2\mathbf{z}$ and so on;
- Let $\mathbf{S}$ the matrix of the sliding windows and replicate it in the third dimension N_C times. In such way, $\mathbf{S}$ and $\mathbf{C}$ are two matrices of dimensions $nw \times l^2 \times N_C$;
- Make the subtraction $\mathbf{A} = \mathbf{S} - \mathbf{C}$; in such way, along the third dimension of $\mathbf{A}$ we find a zero when the position index coincides with the class indicated in the matrix $\mathbf{S}$;
- Mark with one all the positions in which there is a zero in the matrix $\mathbf{A}$;
- Introduce the matrix $\mathbf{B}$ obtained by the summation by row of the matrix $\mathbf{A}$. This matrix have dimensions $nw \times N_C$. The values of $\mathbf{B}$ along the columns indicates how many times each class appears in each window;
- Compute the maximum of the matrix $\mathbf{B}$ by rows: the position index of the maximum of each rows indicates the mode of the window.

This vector approach allows to process a scene of size of the order of 4984×5831 in about one minute on a machine with 12 GB of RAM memory and 8 processors.

The same operation carried out with no vectorization has been completed in about 8 minutes.

3.3.2. Morphological filtering

Mathematical morphology is the most appropriate instrument for extracting image components that are useful in the representation and description of region shape, such as boundaries, skeletons and convex hull. Moreover, morphological techniques are also used for image preprocessing and/or postprocessing for filtering, thinning and pruning activities (Gonzalez and Woods 2007). In this section, in order to optimize the classification map, we use a "closing" morphological operator with square structuring element. It is obtained through the application of the two basic morphologic operators of dilation and erosion. This filter tends to smooth sections of contours fusing narrow breaks and long, thin gulfs and eliminating small holes and filling gaps in the contour (Gonzalez and Woods 2007).

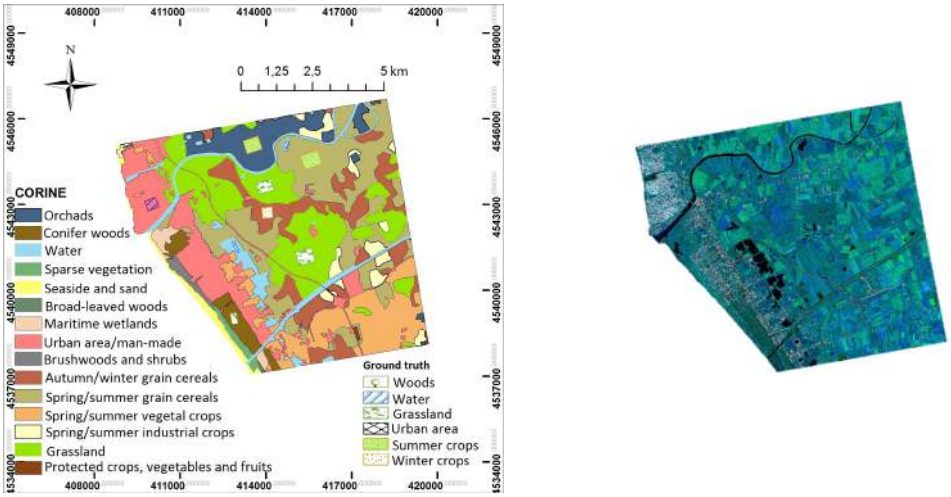
4. Experimental results

In this section we applied the proposed framework to a Level-1α product extracted from a set of COSMO-SkyMed images acquired over the city of Castel Volturno (Italy) in stripmap mode (3 m spatial resolution) and HH polarization between December 2009 and October 2011. In particular, the product used for classification involves two images acquired on December 2009 (reference image, blue band) and August 2010 (test image, green band). Therefore, according with the rationale of Level-1α imagery, this product is well-suited for detecting summer crops, since the enhancement of backscattering produced by the growth of plants and fruits (Fung 1979) causes for this feature a dominant response of the green band in the RGB composite. In general, since the colorimetric response of Level-1α imagery is invariant for scene and climatic conditions, all the considerations made in Section 2 about the association color-object hold for the scene here presented.

The validation of the classification procedure has been performed using an external ground truth derived from the CORINE land cover (Feranec et al. 2007). Seven classes ("Grassland", "Summer crops", "Urban areas", "Woods", "Winter crops", "Water" and "Temporary water") were extracted, as shown in Figure 5(a), for a total extension in the order of 200000 pixels. However, it is remarkable that the class "Temporary water" does not exist in the CORINE land cover product. For this reason, a ground truth for such class has been built exploiting the SAR product.

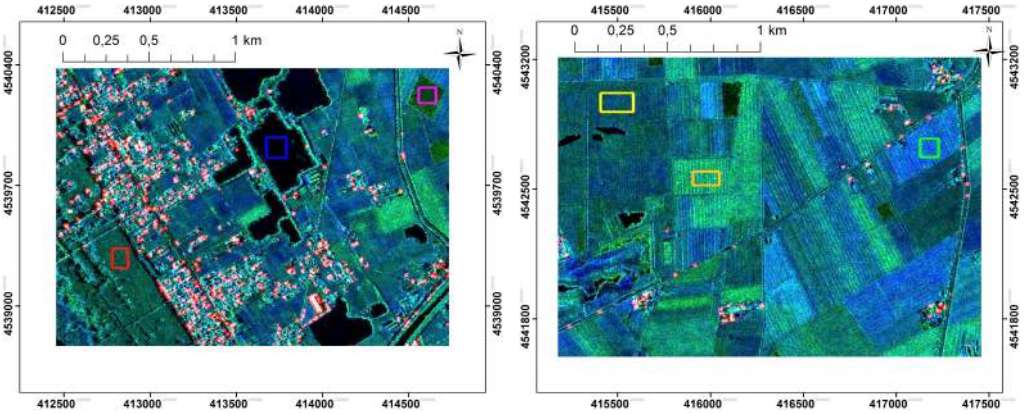
The Level-1α product of the study area is depicted in Figure 5(b). The used training sets are depicted in Figure 6. The first experiment was the solution of the classical SAR classification problem, i.e. a four classes classification (Bruzzone et al. 2004) for the features "Water", "Temporary water" (TW), "Urban areas" and "Woods/grasses/crops". Results of this experiment are shown in Figure 7(a) and Table 1. The registered kappa coefficient κ and overall accuracy are 0.78 and 93.65%, respectively.

As a general comment, in Table 1, a moderate interclass confusion between the classes "Urban areas" and "Woods/Grasses/Crops" (WGC) is registered. This is due to the high resolution of the input maps which contrasts with the ground truth extracted from a land cover whose objective is to identify large homogeneous areas, which, in the case of urban areas, include other features (such as trees and grassland or gardens) detectable on the SAR product.



(a) (b)

Figure 5. Castel Volturno, (a) CORINE land cover with the extracted ground truth and (b) the relevant Level-1α product. Reference image: December 2009; Test image: August 2010.



(a) (b)

Figure 6. Castel Volturno, training sets: (a) woods (red), water (blue), temporary water (magenta) and (b) winter crops (green), grasses (yellow), summer crops (orange).

Table 1. Castel Volturno, four classes classification confusion matrix. $\kappa = 0.78$, overall accuracy: 93.65%. PA: producer accuracy. UA: user accuracy.

	TW	WGC	Urban	Water	PA (%)	UA (%)
TW	89.75	1.30	0.52	0.47	89.75	59.84
WGC	10.25	96.61	33.46	0.19	96.61	95.97
Urban		1.13	65.44		65.44	86.97
Water		0.96	0.58	99.34	99.34	85.86

The model can be complicated separating "Woods" (prevalently located at south-west of our scene, as clearly observable in the Level-1α product of Figure 5(b)) from "Grassland and Crops" (GC), as shown in Figure 7(b) and Table 2. The separation is quite good and in fact the registered κ and overall accuracy (which are 0.741 and 90.44%, respectively) are only slightly lower with respect to the four classes problem.

Results of the six classes experiment are shown in Figure 7(c). In this case we separated "Summer crops" (Sc) from "Grassland and winter crops" (GWc). This separation is possible thanks to the characteristics of the Level-1α products. In fact,

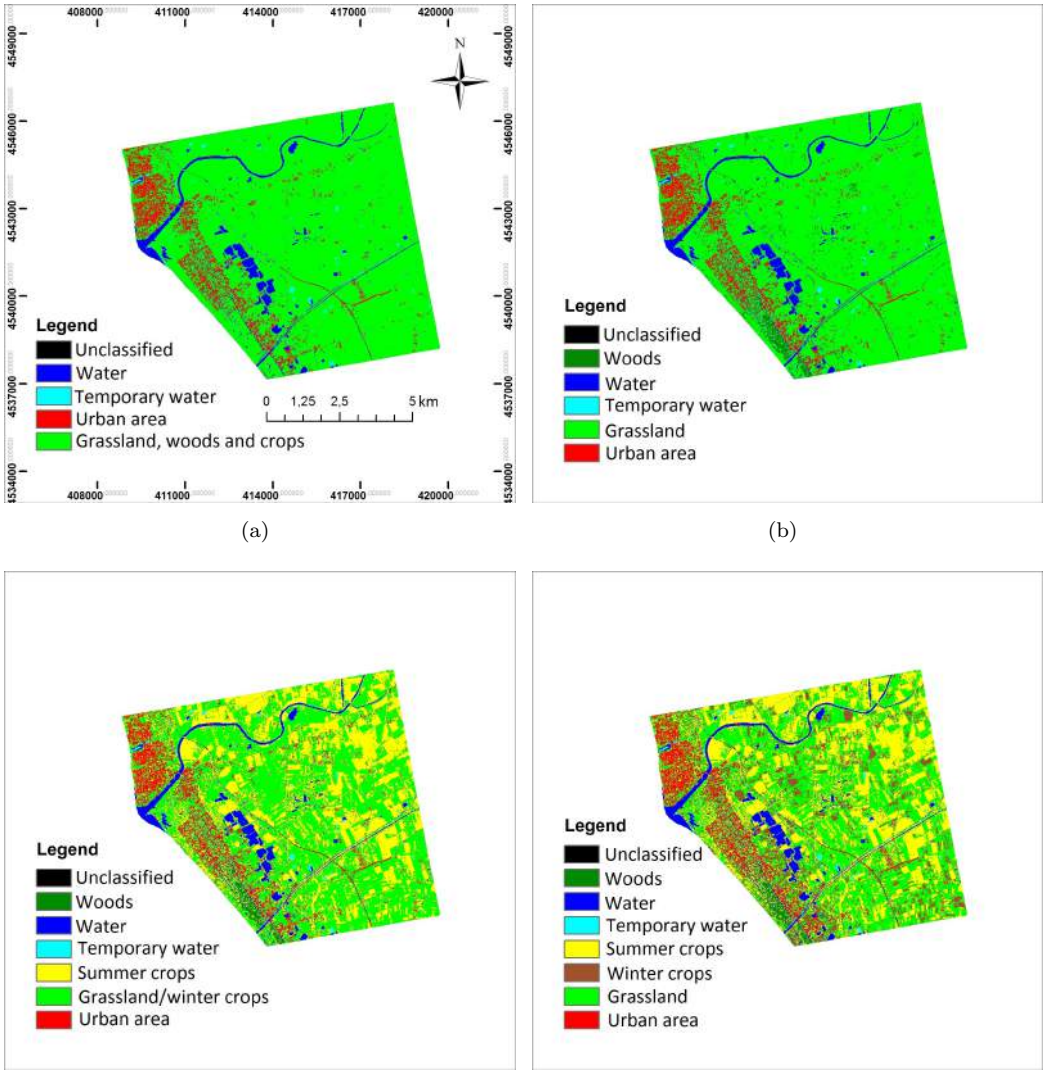


Figure 7. Castel Volturno, ML classification with (a) four, (b) five, (c) six and (d) seven classes.

Table 2. Castel Volturno, five classes classification confusion matrix. $\kappa = 0.741$, overall accuracy: 90.44%.

	TW	GC	Urban	Woods	Water	PA (%)	UA (%)
TW	89.75	1.26	0.52	1.95	0.47	89.75	59.84
GC	10.25	94.72	33.18	30.85	0.19	94.72	93.64
Urban		1.20	65.44	0.01		65.44	86.87
Woods		1.99	0.28	64.31		64.31	68.67
Water		0.83	0.58	2.89	99.34	99.34	85.56

summer cultivations have a response which turns into green since the dominance of the test (summer) band with respect to the reference (winter) one. However, this splitting caused a fall of the κ and of the overall accuracy to 0.563 and 70.46%, respectively. This can be explained by the fact that if the harvesting has been already performed on some fields (in fact, the test band has been acquired at the end of August), and their response looses the volumetric contribution given by the abundant presence of leafs and plants and turns toward those of a pasture. The complete confusion matrix for this experiment is reported in Table 3.

Table 3. Caserta dataset, six classes classification confusion matrix. $\kappa = 0.563$, overall accuracy: 70.46%. Class abbreviations: TW: Temporary water, GWc: Grassland and winter crops, Sc: Summer crops. PA: producer accuracy, UA: user accuracy.

	TW	GWc	Sc	Urban	Woods	Water	PA (%)	UA (%)
TW	89.75	0.03	2.59	0.52	1.95	0.47	89.75	59.84
GWc		80.07	34.55	16.97	8.72		80.07	68.5
Sc	10.25	17.36	57.22	16.21	22.12	0.19	57.22	68.12
Urban		0.86	1.58	65.44	0.01		65.44	86.87
Woods		1.62	2.40	0.28	64.31		64.31	68.67
Water		0.06	1.67	0.58	2.89	99.34	99.34	85.56

A further splitting is possible. In fact, winter crops should exhibit a higher backscattering in winter and, as a consequence, a response which turns into blue on the Level-1α product. The results of this seven classes procedure are shown in Figure 7(d). However this strongly depends on the type of cultivations and on the sowing time. In fact, in souther Italy, autumn/winter cereals are seeded usually in November due to the mild climate. The relevant fields not always reach a backscattering such as to be separated from grassland at the end of December. In fact, in the study area, this cultivations experienced their maximum growing stadium usually in May/June since the harvesting is performed at the beginning of summer. For this reason, the class "Winter crops" is highly absorbed by the "Grassland" one. The registered κ for this classification is 0.462 while the overall accuracy is 60.58%. The confusion matrix is omitted for brevity.

5. Conclusions

In this work we introduced an end-user-oriented framework for the classification of multitemporal SAR data which exploits the characteristics of the Level-1α products. The proposed framework makes use of simple tools well-known in literature and, fusing the concepts of classification and change detection, contextualizes its products in the field of the time series of classification maps. The extracted classes have a different semantic content with respect to the past literature since the detection is made by comparison of objects backscattering with respect to a reference situation rather by the analysis of the temporal behavior along the entire series.

The proposed approach exploits the peculiarity of the Level-1α products: interpretability and class separability. The former property makes our framework particularly well-suited for supervised classifications, since the training step is simplified and fast. The latter property allows for using of very simple classification algorithms. This allows for reduce the complexity of the whole activity, making this schema particularly attractive for nonexpert SAR users. In fact, from the end-users' standpoint, the proposed framework fulfills the requirements of simplicity, repeatability, velocity and accuracy. In particular: i) the use of the Level-1α optimize the the selection of the training sets; ii) the simplicity of the training step and of the decision rule ensures the classification procedure to be completed in a very short time. The post-classification phase also has a small computational burden, thanks to the vector form of the mode filter and to the mathematical morphology robustness; iii) the procedure is simply replicable either for different elements of the same time series or for other scenes with characteristics completely different; iv) the obtained results testify the potentiality of the proposed framework in terms of accuracy and reliability of the output maps.

Due to the simplicity of the schema, in many cases high level classes can be obtained. This is particularly useful in operative contexts when quick preliminary analysis are needed. In this optic, our framework acts as a pre-classifier whose results can be exploited for more detailed investigations.

References

Amitrano, D., G. Di Martino, A. Iodice, D. Riccio, and G. Ruello. 2015. "A New Framework for SAR Multitemporal Data RGB Representation: Rationale and Products." *IEEE Transactions on Geoscience and Remote Sensing* 53 (1): 117–133.

Amitrano, D., G. Di Martino, A. Iodice, D. Riccio, G. Ruello, M. N. Papa, F. Ciervo, and Y. Koussoube. 2014. "Effectiveness of high-resolution SAR for water resource management in low-income semi-arid countries." *International Journal of Remote Sensing* 35 (1): 70–88.

Bruzzone, L., M. Marconcini, U. Wegmuller, and A. Wiesmann. 2004. "An Advanced System for the Automatic Classification of Multitemporal SAR Images." *IEEE Transactions on Geoscience and Remote Sensing* 42 (6): 1321–1334.

Campbell, J. B., and R. H. Wynne. 2011. *Introduction to Remote Sensing.* New York, NY 10012: The Guilford Press.

Camps-Valls, G., L. Gomez-Chova, J. Munoz-Mari, J.L. Rojo-Alvarez, and M. Martinez-Ramon. 2008. "Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 46 (6): 1822–1835.

Davidson, G., and K. Ouchi. 2003. "Segmentation of SAR images using multitemporal information." *IEE Proceedings Radar Sonar and Navigation* 150 (5): 367–374.

Engdahl, M. E., and J. M. Hyyppa. 2003. "Land-Cover Classification Using Multitemporal ERS-1/2 Insar Data." *IEEE Transactions on Geoscience and Remote Sensing* 41 (7): 1620–1628.

Feranec, Jan, Gerard Hazeu, Susan Christensen, and Gabriel Jaffrain. 2007. "CORINE land cover change detection in Europe (case studies of the Netherlands and Slovakia)." *Land Use Policy* 24 (1): 234–247.

Fung, Adrian K. 1979. "Scattering from a Vegetation Layer." *IEEE Transactions on Geoscience Electronics* 17 (1): 1–6.

Gaetano, R., D. Amitrano, G. Masi, G. Poggi, A. Verdoliva, G. Ruello, and G. Scarpa. 2014. "Exploration of Multitemporal COSMO-SkyMed Data Via Tree-Structured MRF Segmentation." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (7): 2763–2775.

Gonzalez, R. C., and R. E. Woods. 2007. *Digital Image Processing.* Upper Saddle River, NJ: Prentice Hall.

Harger, R. O. 1973. "Synthetic Aperture System Design for Random Field Classification." *IEEE Transactions on Aerospace and Electronic Systems* 1 (5): 732–740.

Partington, J. R. 1989. *An introduction to Hankel operators.* Cambridge, UK: Cambridge University Press.

Quegan, S., T. Le Toan, J. J. Yu, F. Ribbes, and N. Floury. 2000. "Multitemporal ERS SAR Analysis Applied to Forest Mapping." *IEEE Transactions on Geoscience and Remote Sensing* 38 (2): 741–753.

Richards, J. A. 1999. *Remote Sensing Digital Image Analysis.* Berlin: Springer-Verlag.

Townshend, J. R. G. 1981. *Terrain Analysis and Remote Sensing.* London: George Allen and Unwin.