

CROP AREA ESTIMATES WITH RADARSAT FEASIBILITY STUDY IN THE TOSCANA REGION - ITALY

G. Narciso^a*, B. Baruth^a, A. Klisch^a

^aJRC, IPSC – Agriculture Unit, 21020 Via Fermi 1, Ispra, Varese, Italy

Commission VI, WG VI/4

KEY WORDS: SAR, Crop Mapping, Data Integration, Rule Base Classification, Forward Chaining

ABSTRACT:

Crop area estimates are crucial for the Common Agricultural Policy of the European Union to manage future intervention on stocks and export policies. The requirement is for early and accurate information, provided in real time and at the lowest possible cost. On this base a study was started by the Joint Research Centre (JRC) – Ispra of the EU Commission, to assess the feasibility of satellite Synthetic Aperture Radar (SAR) imagery in providing such estimates. The study area was located in the Tuscany region of Italy, and the investigation concentrated on a classification procedure linked to SAR data properties. The study tested a rule-based classification capable of assigning class labels without employing training samples. These “rules” relied on crop-specific temporal sequences of cultivation practices identified by changes in backscattering. The proposed classifier provided a level of accuracy well below requirements though of some relevance for a set of 5 crop groups (winter cereals, spring crops, trees, pasture and forage crops). The project produced however a number of useful considerations on the operational use of SAR data for land use mapping.

1. INTRODUCTION

The Common Agricultural Policy of the European Union (EU) requires crop forecasts to manage future intervention on stocks and better reflect on export policies. Information on crop production is also particularly important for the cereal production areas in the Black Sea area (Ukraine, Russia) and Kazakhstan which are EU direct competitors.

Area estimates are considered more stable than yield estimates and, since trend extrapolations or the averages of recent years are often used, any alternative estimate would be adequate in case it was associated to a low error. Accuracy should be within the terms of the statistical law (Ref. Eurostat) and a reference error tolerance is $\pm 5\%$ for the early season estimates, improving to $\pm 1\%$ nearing the end of the agricultural season. In this context and responding to a broad request from the EU Commission, the Joint Research Centre (JRC)-Ispra undertook a study project to assess the suitability of Synthetic Aperture Radar (SAR) imagery for the early estimation of crop acreage.

1.1 Project objectives

The base requirement of the EU Commission is for area estimates supplied as early as possible and updated along the agricultural season. The project aims to assess the capability of SAR images to respond to this requirement and classify land use accurately starting early in the season at the lowest possible processing cost. The use of reference ground truth is a critical constraint to any operational scenario of a fast turn-over from data acquisition to the supply of the information. In view of this the research also aims to explore the possibility to rely exclusively on the SAR data, without using ground truth and devise a processing chain customized to the scope.

With respect to final estimates, the target error tolerance requested has to be considered more a target than a constraint.

1.2 Test area

The test site was the Tuscany region, in Central Italy. The topography of the area ranges from flat areas along the coastline to mountainous zones on the Apennine chain. Plains and valleys cover approximately 10 % of the territory. The land use is predominantly agricultural on flat areas, mixed agricultural and forestry in the hilly and mountainous areas. Typical average size of agricultural parcels is 1 ha and the altitude range for main agricultural areas is 0-400 m. The study concentrated on agricultural areas below the 400 m altitude.

1.3 Methodological Approach

The Synthetic Aperture Radar (SAR) is an active sensor which measures backscattering of its own signal. Images are obtained from a side-looking instrument and are used for a wide range of remote sensing applications. The active electro-magnetic waves are scarcely affected by atmospheric conditions giving this sensor its all-weather and day-and-night operativity. The backscattering of SAR bands X (7- 12.5 GHz) and C (5.3 GHz) respond differently to changes in surface geometry and dielectric properties. The instrument can transmit horizontal (H) or vertical (V) polarization and receive backscattering in either H or V or a combination of both and the response is in general linked to the geometry of the target. The signal is characterized by a low signal to noise ratio and is affected by air and surface moisture contents, temperature, free water and snow. SAR images are 5 to 10 times cheaper than optical images with a similar spatial resolution and this makes a multi-temporal approach with full coverage of large areas affordable.

* giovanni.narciso@jrc.it.

In order to classify SAR images it is necessary to deal properly with the inherent fuzziness of the signal. Further to this, the surface geometric features considered at any particular moment in time, are not necessarily coupled to any specific land use class. The potential of this sensor to answer to the project requirements however relies in the possibility to identify crop-specific variations of backscattering along a times series of images. Their evolution however may be linked to specific sequences of agricultural practices and vegetation development phases, especially in the context of an agriculturally homogeneous territory. If such link can be assessed as consistent and stable it can be defined a 'rule' and it can then be used to classify land use along a sequence of SAR images as they are acquired. The classification can be updated the estimates in near-to real time and without the support of ground truth. A classification approach which can deploy this potential can be defined as "rule-based" and is to a large extent, comparable to an un-supervised classification. Running a "rule-based" classification is an associative task based on an 'if-then' inference. There are two main methods of reasoning when using these inferences: "backward-chaining" and "forward-chaining"; the one to use is determined by the nature of the problem.

(<http://www.amzi.com/manuals/amzi7/xsip/05forward.htm>)

A "backward-chaining" procedure starts with a list of possible goals and works backwards looking for the data or models that will allow achieving any of these goals. "Forward-chaining" on the contrary, using available data, searches which goal they may define (then-clause).

"Forward-chaining" is "data driven" and, in an "if-then" logic, the "if" clause is known and goes before the "then" clause. This approach suits the requirements as, in the case of repetitive decisions such as the classifications of images, it can provide consistent answers, overcoming the need of a training set. Specifically for the classification of SAR data a "forward-chaining" approach can be used to build a rule-matrix with the function of a look-up table, coding a stable transitivity between backscattering and land use. Such a link can potentially be used for classifications in different contexts without the need of further inputs.

2. SAR DATA

A total of 34 images RADARSAT-1, SLC (Single Look Complex), Standard mode images were acquired from July 2004 to August 2005. The acquisition was done along two parallel orbits with a total of 4 frames providing almost full coverage of Tuscany. The choice of the band-width and the polarization were guided by the constraints of having no pre-emptive knowledge of the characteristics of the target, giving more relevance to the homogeneity and comparability of data along the time series. The temporal resolution of 24 days may be considered adequate to capture surface changes in critical periods of the crop season (i.e. ploughing, sowing, crop emergence and development, harvesting, etc.). Other features of the images were an incidence angle of 30-37 (Beam 3 image mode), 100x100 Km coverage, a resolution of 15.7 m, with an azimuth of 8.9 m, with an 11.6 m and an azimuth of 5.1 m. pixel spacing range.



Figure 1. The Tuscany region overlapped with the selected four SAR frames (Radarsat-1 Standard 3 Beam Mode).

2.1 Pre-processing

The pre-processing chain was defined with the scope of guaranteeing an automatic co-registration so as to speed the data analysis. The pre-processing was done separately on the 4 frames with a number of different processing sequences repeated on the images as they were acquired. The first sequence is applied for a reference SLC. It includes the generation of a look-up table and the estimation of a topography dependent, pixel normalization factors. A specific geo-coding software module was used employing no ground control point but rather image correlation matching based on a large number of automatically selected image chips. Using the digital elevation model a SAR image was simulated in the slant range geometry of the image frame. The reference image was then correlated to the simulated SAR image.

Finally, each acquired image is re-projected to the chosen projection of the reference image (UTM Zone 32 North, EPSG: 32632) and re-sampled to 20 x 20 m pixel size. This process leads to image-to-image precisions of better than < 0.15 SLC pixel in range and < 0.35 pixel in azimuth. Finally image the de-speckling is done after the pre-processing of the individual scenes. The de-speckling however was omitted from signature extraction

2.2 Transformation from backscattering to byte images

The pre-processed images contain normalized backscattering coefficients that still have to be converted to the log scale in dB values according to the following equation:

$$db \sigma_{i,j} = 10 \cdot ALOG10(\sigma_{i,j}); \sigma_{i,j} > 0 \quad (1)$$

Negative values have been considered as the result of the interpolation during geo-coding and thus, were excluded from further calculations. Afterwards, all images were scaled to a consistent range of dB values. It was done to get data that are similar to optical byte images, and that can be easily handled. The following equation has been applied:

$$BYTE db \sigma_{i,j} = \frac{255 \cdot (db \sigma_{i,j} - t_1)}{t_2 - t_1}; t_1 < db \sigma_{i,j} < t_2 \quad (2)$$

Where t_1 and t_2 are the lower and upper thresholds for the data range of $db\sigma_{ij}$ that are used for the further analysis.

3. ANCILLARY DATA

A number of ancillary data sets were used to improve the interpretation and classification of the images.

3.1 Land cover

The land cover information used as topographic and thematic reference for the analysis was derived from the Corine land cover data base. Given the nature of the project, focusing on agriculture, the exclusion of all those areas which are not properly agriculture helped to avoid possible sources of noise.

3.1.1 Meteorological data

A time series of meteorological data was acquired from the meteorological station of the interest area to support the interpretation and classification. Significant variations in the calendar of agricultural practices and crop growth cycle may occur depending on weather conditions. Meteorological data are also important to assess the soil moisture content (affecting surface scattering through dielectric properties). In winter, subzero temperatures can drastically change dielectric properties of soils, because water is bound in ice, leading to a steep drop in backscattering. A snow cover will “shield” the underlying soil surface from contributing to the overall backscattering signature.

3.1.2 Elevation data

The SRTM, digital elevation model as the backscattering of bare soil and vegetation is sensitive to terrain slopes as these modulate the local incidence angle.

3.1.3 Statistical information

Statistics on planted surfaces for the years of analysis (2004 - 2005) were the main references for the exploration of the rule base and for the evaluation of the accuracy of the classification. Main source of these statistics is the national crop estimates survey campaign, run by the Italian Ministry of Agriculture (AGRIT project). The survey is based on a “point frame approach” and relies on the direct reporting of land use, observed in the field over a large number of (a-dimensional) point samples. The survey was run in the years 2004 and 2005, on a sample of over 16.000 points, 8.067 of which are on agricultural land. The points are characterized by x; y coordinates and the survey legend is very detailed.

4. DATA ANALYSIS

The process of building a rule base for the classification of the time series of SAR images was structured into three main phases: a pre-analysis, the coding of the rules and the classification itself.

4.1 Pre-Analysis

The link between the progression of backscattering and the agronomic cycle of a specific crop is an assumption that needs an “experimental assessment”. For the scope, an empirical surface geometry evolution model was developed for a number

of crops. This model was then compared to a time series of backscattering measures. The recursive comparison of the 2 curves was run over a number of crops aggregation scenarios. This exploration assessed that backscattering discriminates at best an aggregation of five broad crop groupings: autumn and winter cereals (AWC), spring and summer crops (SSC), tree crops (T), permanent pastures (PP) and rotation forage (RF).

4.1.1 Surface Roughness Model

The evolution of surface roughness of cultivated fields is connected to a specific calendar of cultivation practices and crop development stages. In consideration of this, surface roughness was coded on an empirical scale of 5 levels, ranging from “high” to “low” and ranked from “most rough” to “least rough” for different stages of crops cycles. Soil background, expressed as percentage (%) cover, is a further characterizing element. It decreases along the crop cycle as the closing of the canopy proceeds. The three variables: level, ranking and percentage cover; were combined in an indicator describing the evolution of surface roughness and defined as “Surface Roughness Score (SRS)”. On this base it was possible to draft a model of the evolution for each reference crop class. The following graph shows the trend of curve for the final aggregation of crops.

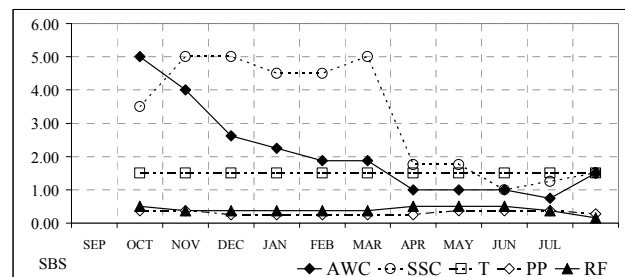


Figure. 2. Trend of the Surface Roughness Score (SRS)

4.1.2 Reference Backscattering

The time series of backscattering values was extracted from 3*3 pixel chips centred on a sample of 600 points of the AGRIT point survey on the SE frame. These values were processed to define a specific indicator (Surface Backscattering Score, SBS) comparable to the SRS. The measured “level” and “ranking” were coded as for the SRS on a scale of 5 and plotted on a graph, which, the previous one shows the trend for the 5 crop groupings.

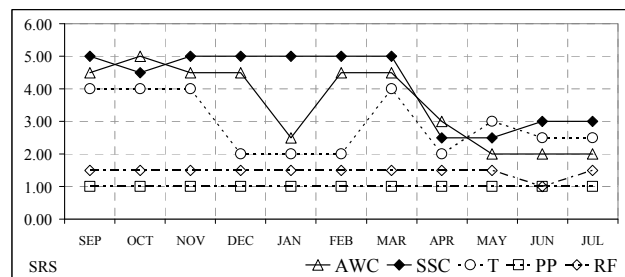


Figure 4 Trend of the Surface Backscattering Score (SBS)

4.1.3 Comparison of the Surface Roughness model and Backscattering

This exploration assessed that SBS and SRS discriminate land use in a similar way. For winter crops, the two indicators appear to be fairly coherent along most of the development cycle. Spring crops show a diverging trend at the beginning of the cycle as backscattering increases while the modelled surface roughness decreases. The two trends are more coherent in their average stability during winter. The drop of SBS in February (as for the winter crops) is probably caused by meteorological factors. In early spring the evolution of the two trend lines proceeds in parallel but diverges again starting from May onward.

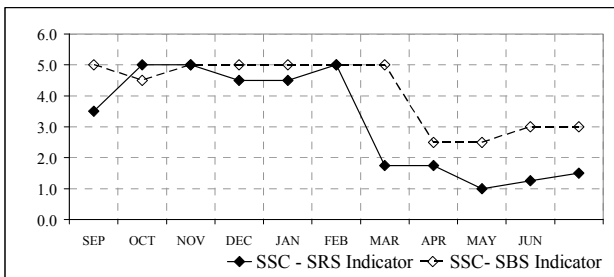


Figure 5 Trend comparisons for winter crops

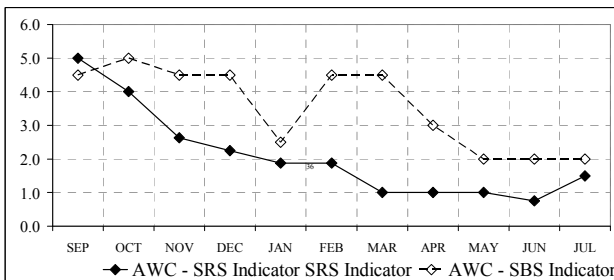


Figure 5 Trend comparisons for spring and summer crops

4.2 Rule coding

The pre-analysis allows the formalization of “rules” which are such because they are expected to be independent from contingency.

RULE 1: Surface roughness and backscattering are characterized by coherent levels of magnitude for specific crops
 RULE 2: Backscattering is characterized by different levels of magnitude during the agricultural cycle of different crops and these levels may rank differently between one land use class and another at different times.

The pre-analysis also showed that level, ranking and variation of backscattering may be described using an indicator defined as score

$$SCORE = f(aL; bR) * C \quad (3)$$

L= the absolute level of surface roughness;

R= is the “ranking” order

C= the % of soil cover

a, b are specific weights to account for the fact that the two variables may balance each other at different moments

This indicator needs however specific weights and coefficients to account for the fact that the two variables may balance each other at different moments

4.2.1 Forward Chaining Process

Forward chaining is automatic process, aimed at establishing the most stable link possible between backscattering and land use classes and therefore builds a rule-matrix working as a LUT for the classification of SAR images in an unsupervised mode. In the current analysis, the forward chaining process concentrated exclusively on the time series of the SE frame. The relative rule matrix was then applied for the classification of the neighbouring frames in order to test the validity of a localized knowledge expanded to homogeneous areas with no preliminary “training information”.

The process is incepted from what can be defined as a “trigger rule-matrix” based on a number on hypothesis. It then runs a number of cycles to adapt this trigger-matrix to the specific conditions of the area studied. The scope is to implement a procedure which increases its own knowledge with a recursive cycle of classification and testing. The final rule-matrix is the one in which the most accurate correspondence between the indicator and the land use classes of the training set can be achieved.

Once confirmed, these “rules” were used to define indicators specific for land use and period. A matrix of these indicators was built where the columns are the five reference land use classes and the rows are defined by the time sequence of backscattering.

5. CLASSIFICATION

The time series of SAR images was classified using a “Minimum Euclidean Distance” approach of pixel values to the reference values in the rule matrix. Each image was classified independently and all values which exceeded a specific threshold distance from the reference value were labelled as non-classified. A time series of twelve classified mosaics was assembled by merging the frames along the same orbit per acquisition date and between neighbouring orbits with the nearest acquisition dates. The classified surfaces of each mosaic were compared with the 2005 official statistics of Tuscany, where all reported crops were grouped into the five referenced groupings. Those pixels classified as “tree crops”, given the inherent lack of homogeneity of the class, were labelled as non-classified and were not compared.

5.1 Accuracy Assessment

The accuracy of the classification was assessed using a confusion matrix (Figure .and). All the samples mapped as urban areas, orchards, forested area and non agricultural uses on the Corine land cover map were excluded from the analysis. We are aware that this procedure is not formally correct. These confusion matrices are any case an assessment of the poor level of the classification.

	AWC	SSC	T	PP	RF
AWC	0	0	0	0	0
SSC	0	69	42	17	19
T	0	42	34	15	14
Pp	0	64	22	16	14
RF	0	37	13	16	15

Table1. Confusion matrix for a five crops grouping

			AWC	
			PP	
			RF	SSC
AWC	PP	RF	267	77
		SSC	71	34

Table2. Confusion matrix for a two crops grouping

The overall estimates were also compared with the official statistics of the "Regione Toscana". The classified surfaces of the first mosaic of the time series (September, 2004) differ significantly from the 2005 official statistics. The estimated area of winter cereal (Figure.6) evolves irregularly from September to December and differences from the official figures oscillate in a range of +/- 150.000 ha. In January the classification and the estimates coincide but the area planted is systematically overestimated all through the winter. Estimates and official figures eventually converge at the end of the season. A similar was observed for spring crops which starts with an over estimation ranging between 100.000 and 150.000 ha. It converges to the actual figures in January and is then underestimated throughout winter and spring. The estimates and the official figures converge again from April onward, in connection with the actual start of the crop cycle.

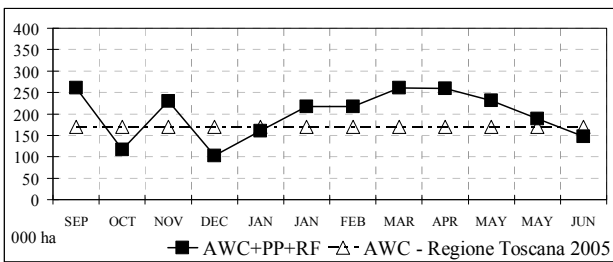


Figure 6. Classified surface Vs. Official statistics (AWC)

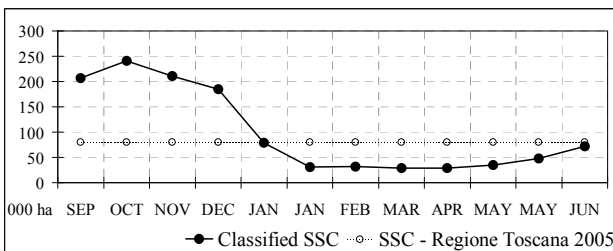


Figure 7. Classified surface Vs. Official statistics (SSC)

Permanent pastures and rotation forage are systematically over estimated. A further aggregation of winter crops, pastures and

forage crops was also tested and compared to spring crops. During winter, especially the winter of 2005, precipitations (rain and snow) masked the surface characteristics. During this period the relative differences between the estimates and the official statistics are more marked.

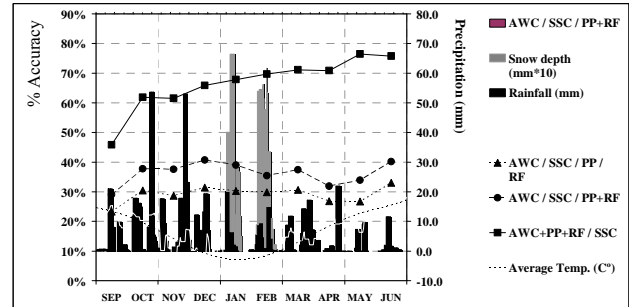


Figure 8. Evolution of Accuracy % along the season

The overall accuracy of the classification was also tested separately on two series of mosaics one for the eastern frames and the other for the western frames, Overall it resulted quite poor except for the case where winter crops, permanent pastures and rotation forage are considered as a single class. The level of accuracy of this configuration also shows an improvement along the times series which is constant for both the West and East frames (Figure 8).

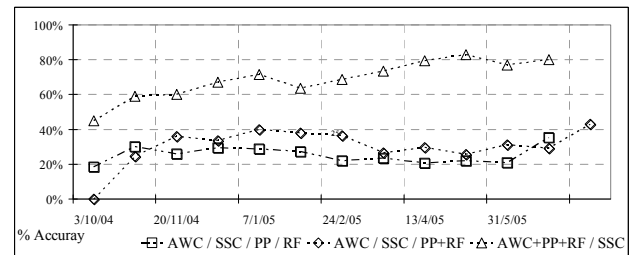


Figure 8. Evaluation of % accuracy (East frame)

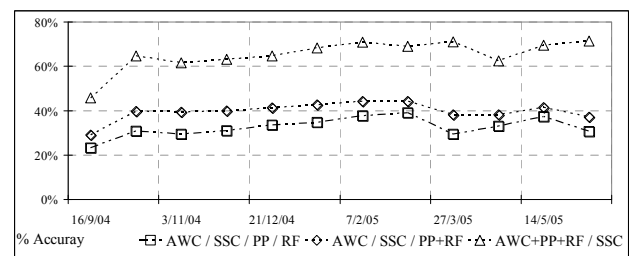


Figure 8. Evaluation of % accuracy (East frame)

6. CONCLUSIONS

The main conclusion is that whatever the constrains and the limitations, the classification attempt failed. However the applied methodology eventually brought to a classification of the images and this was not a given at the beginning of the study. The official statistics and the classifications are more or less comparable, at least in the order of magnitude of the overall

area planted surface and also this was not a given at the beginning of the study.

The methodological approach was chosen to respond to the project requirements and there are definitely some limitations at the origin. In the first instance there is no preliminary guarantee that the set accuracy targets can be met. A further limit is that that some reference ground-truth must in any case be used, at least in an initial phase, to stabilise the rule base itself (forward-chaining). In order to adapt to other agricultural environments the rules need to be revised with the consequent requirement for further ground truth.

An inherent constraint is also that sometimes human common sense is ignored. In practice once the inferences are established there is no space for "interpretation" which is the creative responses of experts to unusual circumstances.

The project itself was subject to a number of constraints:

- A single year of observation is too short a period for anything more than a hypothesis test.
- Tuscany is characterized by a complex topography and a wide variety of agricultural scenarios which are not particularly suited for the use of radar images.
- The evaluation of the accuracy relied used a set of available data which was not dedicated to the scope.

There are however some observations that can be made. The evolution of the accuracy levels especially from the second image onward is an indication that the methodology applied for the classification is sensitive to the evolution of surface characteristics. During the pre-analysis phase, the surface roughness model does not clearly discriminate between winter cereals and spring crops at the beginning of November and from December to January while the measure of backscattering does. The same situation is observed in late spring for winter crops, forage crops and permanent pastures. A possible interpretation is that the "modelled" trend of surface roughness misses a component which is otherwise caught by the SAR information.

The maximum discrepancies between classification and official statistics take place in winter. Noticeable is the effect of precipitation and especially snow cover in January and February 2005. Meteorological events may be local and episodic but are certainly frequent and should not affect the classification from SAR considering its all-weather capabilities. In this study however such interference was assessed. This specific advantage of SAR appears questionable by the decreased of sensitivity to the wet ground that follows cloudy weather.

The results concerning the accuracy in the identification of single land use classes are not encouraging. However it appears possible to discriminate between broad groups of crops.

As a final comment the forward chaining approach has the potential to characterize a management model affecting land use.

REFERENCES

Bardossy, A.; Samaniego, L.,(2002) Fuzzy rule-based classification of remotely sensed imagery, *Geoscience and Remote Sensing*, IEEE Transactions on Volume 40, Issue 2, Feb 2002 Page(s):362 – 374

Blasé, X. (2005), *Crop Monitoring by advanced SAR remote sensing techniques: modelling and experimental analysis*. PhD Thesis, Faculté des Sciences Agronomiques, Université Catholique de Louvain, Belgique

Buzzone, M. Marroncini, U. Wegmuller, and A. Wiesmann, (2004) "An advanced system for the automatic classification of multitemporal SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, no. 6, pp. 1321-1334, 2004.

Hagg, W.; Sties, M.,(1998); Knowledge and context based SAR classification for map-updating using fuzzy logic and GIS-methods *Geosciences and Remote Sensing Symposium Proceedings*, 1998. IGARSS apos; 98. 1998 IEEE International Volume 3, Issue, 6-10 Jul 1998 Page(s):1307 - 1309 vol.3

Hellmann, M., and Jager G.,(2002); Fuzzy rule based classification of polarimetric SAR data , *Aerospace Science and Technology*, Volume 6, Number 3, March 2002, pp. 217-232(16).

Hellmann, M. Jager, G. Kratzschmar, E. Habermeyer, M. Classification of full polarimetric SAR-data using artificial neural networks and fuzzy algorithms, *Deutsches Zentrum fur Luft- und Raumfahrt, Oberpfaffenhofen*;

Newell, A. (1982). The knowledge level. *Artificial Intelligence*. 18(1), 87-127.

Quegan S. and J. J. Yu (2001), "Filtering of multichannel SAR images," *IEEE Trans Geosci. Remote Sensing*, vol. 39, no. 11, 2001.

Sowter, A.; Lemoine, G.; van Leeuwen, H (1998).; SAR techniques for rapid area estimation of agricultural crops during winter and spring ; *Integrated Systems for Commercial Remote Sensing Applications* (Ref. No. 1998/207), IEE Colloquium on Volume , Issue , 28 Apr 1998 Page(s):6/1 - 6/5

ACKNOWLEDGEMENTS

Thanks to Dr J. Gallego (JRC-Ispra) for his precious suggestions for the devise of the methodology and the accuracy assessment.

Tanks to Mr J. Cerrani (formerly JRC-Ispra) for the support in the implementation of the forward chaining procedure.