Macrophyte classification in the Lower Paraná River floodplain: an object-based approach on multi-temporal COSMO-SkyMed X-band data

Natalia Soledad Morandeira[1,3]
Rafael Grimson[2,3]
Patricia Kandus[1]

[1] Instituto de Investigaciones en Ingeniería Ambiental (3iA)
Universidad Nacional de San Martín (UNSAM)
Campus Miguelete, General San Martín, Buenos Aires, Argentina
nmorandeira,pkandus@unsam.edu.ar

[2] Escuela de Ciencia y Tecnología (ECyT)
Universidad Nacional de San Martín (UNSAM)
Campus Miguelete, General San Martín, Buenos Aires, Argentina
rgrimson@unsam.edu.ar

[3] Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Argentina

Abstract. We propose a multi-temporal approach to discriminate macrophyte vegetation types in the Lower Paraná River floodplain. During a low intensity flood pulse, seven X-Band COSMO-SkyMed HImage images were acquired, covering a nine-month period. The region was first segmented with a mean-shift segmentation method using the information from the complete temporal series. Next, objects were classified with the expectation maximization algorithm into spectral classes. These spectral classes were assigned to six information classes defined through field sampling: Bulrush marshes, Short broad-leaf marshes, Tall broad-leaf marshes, Short grasslands and grass marshes, Tall grasslands and grass marshes, and Water. Class interpretation was based on the backscatter dynamics in relation to plant coverage and to hydrometric level. We related backscattering coefficient changes to interaction mechanisms: mirror reflectance (water-covered areas), volume dispersion (emergent vegetation with medium backscatter) and double-bounce (emergent flooded vegetation with high backscatter when water is present). The accuracy of the obtained product was assesed by comparing it with 55 field samplings. Global accuracy was 71.2%, whereas Kappa index was 63.4%. This work points out the usefulness of X-Band data for flood monitoring and macrophyte vegetation type discrimination. In a mosaic of herbaceous wetlands, the dynamics associated with flood pulse may change within patches in different geomorphological settings and topographical positions. The knowledge on the relation between vegetation, local settings and floods is essential for interpreting and predicting how backscattering coefficients and other SAR-derivated parameters vary with flooding.

Keywords: active microwave, flood pulse, wetlands
Palavras chave: áreas úmidas, pulso de inundação, microondas ativas

1. Introduction

Wetlands associated to floodplains are strongly influenced by the hydrological regime and the flood pulse (JUNK; BAYLEY; SPARKS, 1989). Flood height, duration, frequency, and energy (among other parameters) depend on the geomorphological settings and the local topographic positions. Thus, the effect of flood pulse on vegetation usually varies locally. The Lower Paraná River floodplain (also known as Paraná River Delta) is covered by a mosaic of wetlands dominated by herbaceous communities and dotted by shallow lakes. Vegetation zones dominated by one or a few macrophyte species are typical and are in close relation to flood

regime. Since flood pulses may lead to a distinct effect in each macrophyte community, a multi-temporal analysis can help to characterize wetland dynamics. The knowledge on how flood interacts with vegetation and how, in consequence, a remotely sensed variable is affected (e.g., reflectance, backscatter coefficient) is essential for multi-temporal classification procedures in floodplain wetlands.

The ability of Synthetic Aperture Radar (SAR) data to detect water below the vegetation has been promising for wetland vegetation discrimination and flood monitoring. Few SAR applications are focused specifically on herbaceous wetlands (e.g.,(NOVO et al., 2002; COSTA et al., 2002)), and most of them use C-band or L-band data (HENDERSON; LEWIS, 2008). Previous works with X-band data (9.6 GHz with a wavelength of 3.1 cm for high-resolution COSMO-SkyMed system (CSK from now on)) have reported high backscatter for marsh vegetation such as sedges, emergent reeds and grasses (HENDERSON; LEWIS, 2008). Due to its high frequency, X-band signal is expected to have low penetration and to be affected by water, soil and vegetation roughness. Besides, X-band backscatter values are expected to saturate with much lower biomass than in C- or L-band. RAMSEY III (1998) reported that X-band never reached the water surface in flooded marshes. In contrast, Pulvirenti et al. (2011) modelled the contribution of double-bounce scatter in flooded areas covered by wheat (with -10 to -5 dB).

In this work, we propose a multi-temporal approach to discriminate macrophyte vegetation types in the Lower Paraná River floodplain. We performed an object-based unsupervised classification. During a low intensity flood pulse in a nine-month period, we characterized the dynamics of the vegetation types. We assumed that vegetation types (described with field-works) were unchanged during the study and, thus, assigned backscattering changes in each class to the presence and height of water column.

2. Methods

2.1. Study area

The study was conducted in the Lower Paraná River floodplain (also known as Paraná River Delta), next to Ramallo city (Figure 1), which is subject to the flood pulse of the Paraná River. The climate is temperate humid. The area is relatively flat: the mean difference between high and low positions at a local scale is 65 cm, as recorded in 17 transects by the authors (MORANDEIRA, 2014). Herbaceous communities distributed with a zonation pattern dominate the marshes, whereas forest patches cover ca. 1.5% of the study area (estimated from Enrique (2009)). The main productive activity is cattle, which depends on the natural herbaceous vegetation of the area. Fishery and apiculture are also important productive activities.

2.2. Field sampling and information classes

Vegetation sampling was conducted during summers 2010–2011 and 2011–2012 in 42 georreferenced sampling sites distributed along the study area. Based in three Braun-Blanquet abundance-cover censuses per site (MUELLER-DOMBOIS; ELLENBERG, 1974), we classified each local community into five main vegetation types. In each site, the vegetation type covered an homogeneous area of at least 1000 m^2. We also recorded the location of ten sites with open water (shallow lakes and rivers). The following description of the vegetation types is based in Morandeira and Kandus (In revision).

Bulrush marshes (A) are dominated by equisetoid herbs (*Schoenoplectus californicus, Cyperus giganteus*), usually growing in low flooded positions. Plants are characterized by vertical photosynthetic stems of 140 to 250 cm height. Short broadleaf marshes (B) are generally shorter than 80 cm and have few leaves with high leaf areas. Usually occur in low topographic positions, in generally flooded or soil-saturated sites. Typical species are

Sagittaria montevidensis, Eclipta prostrata and *Enydra anagallis*. Tall broadleaf marshes (C) are 150 to 250 cm height and usually have abundant leaves and ramified stems. Typical species are *Baccharis salicifolia, Conyza bonariensis, Polygonum acuminatum* and *Ludwigia* cf. *peruviana*. Short grasslands and grass marshes (D) are less than 50 cm in height and have few leaves. This vegetation type may occur in tall topographic positions (such as leeves) or in medium topographic positions. Typical species are *Cynodon dactylon, Paspalum vaginatum* and *Echinochloa helodes*. Tall grasslands and grass marshes (E) are 50 – 150 cm height and grow in tall or medium topographic positions and have abundant leaves. Typical species are *Panicum elephantipes, Hymenachne pernambucense, Echinochloa crus-gallis, Bolboschoenus robustus* and *Leptochloa fusca*.

Minimum aboveground green biomass is similar within vegetation types (100 to 290 $g.m^{-2}$). Maximum biomass ranges between 620 $g.m^{-2}$ and 3340 $g.m^{-2}$ and can be ordered as follows: D < B < C < A < E. We hypothetize that the interaction between the signal and the targest depends on the amount of biomass, on its distribution in leaves and stems (plant structure and architecture), on whether a water film or a water column is present below vegetation, and on soil wetness and roughness (specially for low biomass vegetation types).

2.3. COSMO images

The HImage CSK acquired images are listed in Table 1. They cover a region of about 42 km $\times$ 45 km (Figure 1) and have a pixel resolution of 2.5 m $\times$ 2.5 m. The original images were calibrated, geocoded and clipped using NEST Software ((European Space Agency, 2012)). Then, we applied a 7 $\times$ 7 pixel Lee filter in order to reduce the speckle noise on each image. Finally, we converted the filtered images from intensity to dB and created a stack with the seven 23948 $\times$ 20861 preprocessed images.

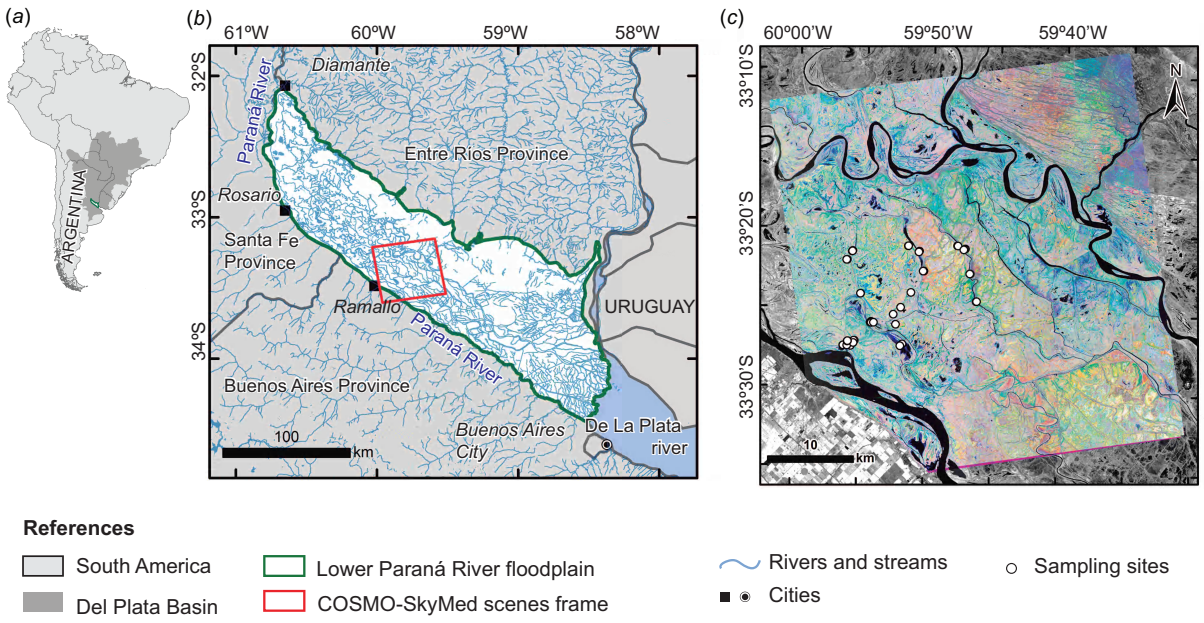

Figure 1: Study Area. (a) Location of the Del Plata basin and the Lower Paraná River floodplain in South America. (b) Lower Paraná River floodplain. (c) RGB composite of the second (R), fourth (G) and last (B) CSK images. The background image is a grayscale display of an infrared band from Landsat 5–TM scene.

Table 1: Features of the available HImage COSMO-SkyMed images and the corresponding hydrometric level in the Paraná River near Ramallo.

Date	CSK image		Hydrometric level
	Orbit	Incidence angle	at Ramallo (m)
08 – 20 – 2011	ascending	40°	2.57
09 – 21 – 2011	ascending	40°	2.74
10 – 23 – 2011	ascending	40°	2.48
11 – 16 – 2011	ascending	40°	2.77
02 – 13 – 2012	ascending	40°	1.78
03 – 15 – 2012	ascending	40°	1.64
04 – 16 – 2012	ascending	40°	1.42

2.4. Classification procedure

Since the spatial resolution of the CSK images is finer than the objects we are classifying, and due to the speckle noise, we considered that the most suitable approach is a object-based image analysis ((BLASCHKE et al., 2014)). This approach allowed us to take into account the contextual information to classify sets of pixels, considering the spatial patterns they create.

To produce the land-cover map from the stack of seven preprocessed CSK images described above, we first performed a segmentation of the clipped scenes. This step was accomplished by using the *mean shift* segmentation method from the OTB ((INGLADA; CHRISTOPHE, 2009)), with the range radius set to 2 dB. The result of this step was a segmentation of the space into almost 1.5 million segments. To each of these segments we associated its mean spectral signature (the vector consisting of the mean backscattering over the segment for each image), obtaining a distinctive temporal evolution in backscattering.

Next, we classified the mean signatures obtained in the segmentation step into spectral classes using the expectation-maximization (EM) algorithm ((MURPHY, 2012)) to find the optimal parameters of a mixture of 14 Gaussians model. To run this unsupervised clustering algorithm we used the implementation from the Python library PyPR. After obtaining the parameters of the proposed model, we assigned each segment to one of the 14 Gaussian clusters. The objects identified using this method are homogeneous in every CSK image, but the mean value of each spectral class varies for the seven different images.

We analysed and interpreted the multi-temporal signatures of the spectral classes, merging some of these classes together to obtain the five information classes that represent different land-cover classes in the map. The accuracy of the final product was evaluated by comparing the classes obtained with 41 homogeneous regions (ROI) generated around georeferenced field sampling sites. The global accuracy, the omission and commission errors and the Kappa index were computed ((CONGALTON, 1991)).

3. Results and discussion

From the classification procedure, we obtained 14 information classes, which we assigned to the six information classes (Figures 2 and 3). We based our interpretation on the spatial position of the field samples and on the analyses of the spectral signatures of each class through the studied period (Figure 3). Since field sampling was limited to two surveys, we assumed that plant coverage (vegetation type and biomass) remained the same in each site. A Kappa index of 63.4% was obtained (0.95 confidence interval: 63.1 – 63.6%), representing 71.5% of the maximum Kappa that can be obtained with our confusion matrix. Global accuracy was 71.2%, with a mean omission error of 40.1% and a mean commission error of 38.2%. The information

classes can be described as follows:

A. Bulrush marshes (10.6% of the total area). The three classes assigned to Bulrush marshes show a very high backscatter (> -6 dB) in at least half of the scenes. We suggest that backscatter was enhanced by double-bounce through all the studied period in class A1 and during August-November in classes A2 and A3. Double-bounce at targets dominated by *Schoenoplectus californicus* with the presence of standing water has been described and modeled for C-band ((GRINGS et al., 2005)) and at other plant targets for X-band ((PULVIRENTI et al., 2011)). Assuming a constant biomass of *S. californicus*, the backscatter decrease in A2 and A3 may be related to the hydrometric level decrease. The higher backscatter in A2 than in A3 may be due to higher *S. californicus* biomass. The A1 temporal pattern can be explained if we assume that A1 areas remained flooded all the period. Class A is mainly confused with class C (with both omission and commission errors).

B. Short broad-leaf marshes (13.9% of the total area). Two classes were assigned to broad-leaf marshes. B1 and B2 have medium backscatter and identical signature in the February-April period (-8.2 to -7.7 dB). During the high hydrometric level period (August-November), B1 has low backscatter whereas B2 has high backscatter. A possible explanation of this pattern is that B1 areas were at least partially covered by water (so that mirror reflection occurred), altogether with a little contribution of volume scatter from emergent vegetation. B2 may have partial contribution of double-bounce scattering during August-November. Note that B2 has low scatter in the September scene, probably due to the increase in water level in the floodplain which may cover part of the vegetation. Class B is mainly confused with class C.

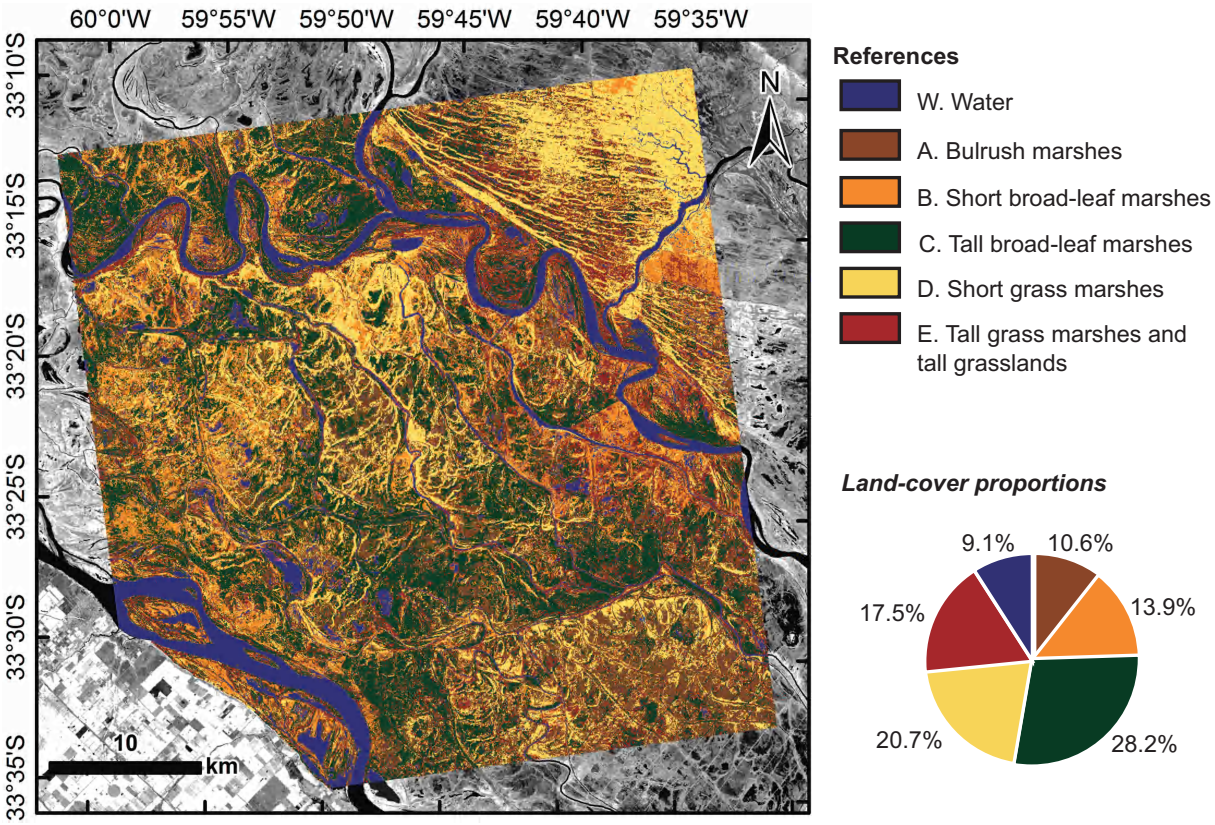

Figure 2: Macrophyte classification product on the COSMO-SkyMed multi-temporal series.

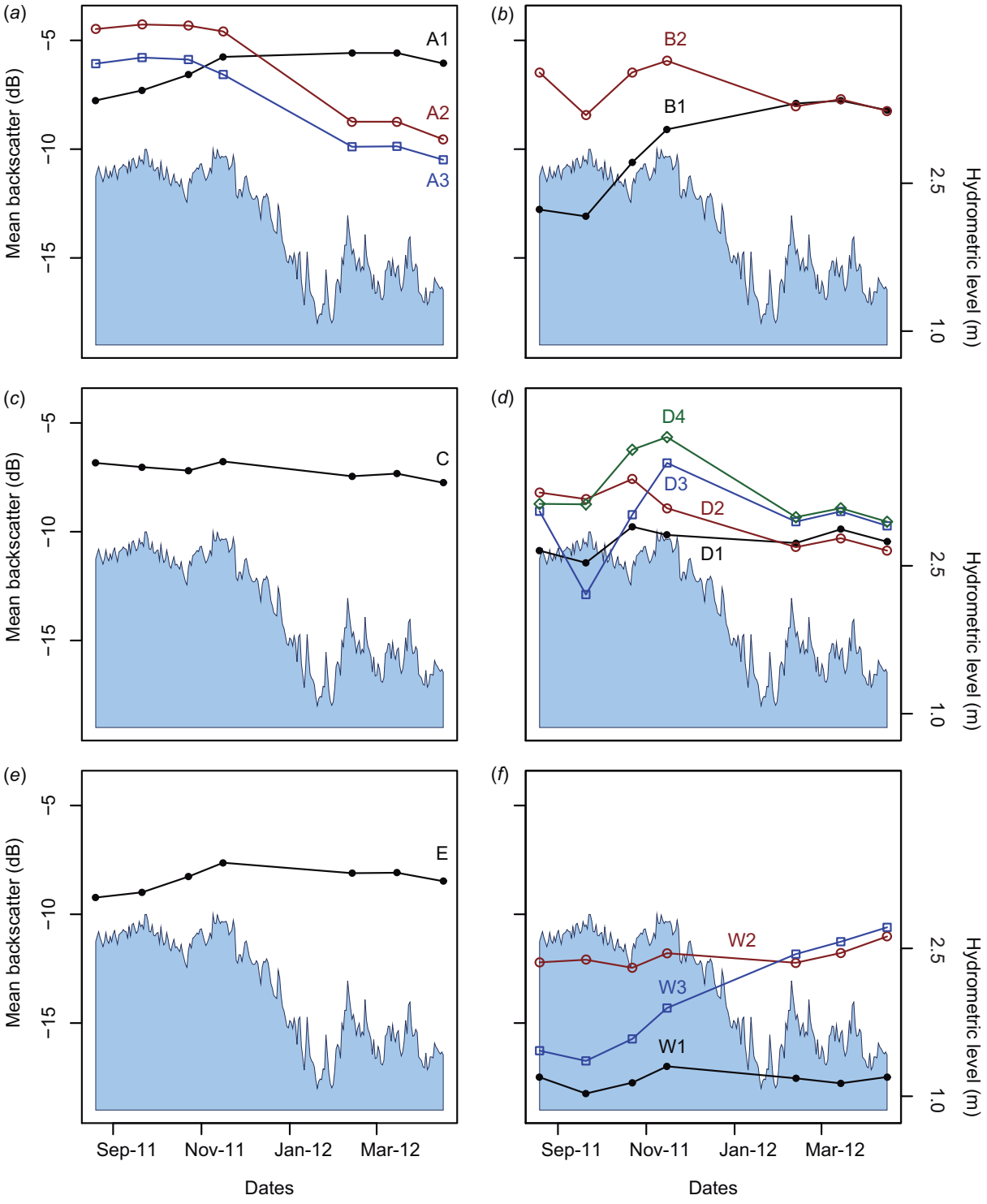

Figure 3: Temporal signatures of the information classes. (a) Bulrush marshes. (b) Short broad-leaf marshes. (c) Tall broad-leaf marshes. (d) Short grasslands and grass marshes. (e) Tall grasslands and grass marshes. One to four classes were assigned to each information class.

C. Tall broad-leaf marshes (28.2% of the total area). The class assigned to this vegetation type has a constant and relatively high backscatter (between -7.8 and -6.8 dB). This might be explained by high dispersion volume, which can be easily obtained in X-Band for this type of vegetation (tall and usually with high biomass and abundant leaves). This class was the most abundant in the floodplain. In the final product, class C is mainly confused with class A and B

(with both omission and comission error) and with class E (commision error).

D. Short grasslands and grass marshes (20.7% of the total area). Four classes were assigned to this vegetation type. The temporal pattern shows a maximum in October (B1 and B2) or in November (B3 and B4), a local minimum in September and low backscatter in February-April. We suggest that these short grasslands were flooded between August and November, probably almost covered by water in September (mirror reflection) and with only a water film in November (double-bounce scatter). In the February-April period, backscatter was relatively low due to the characteristic low biomass of this vegetation type. Class D is mainly confused with water (omission error).

E. Tall grasslans and grass marshes (17.5% of the total area). The temporal pattern of this class is similar to the one of class B but here the backscatter ranges between -9.2 and -7.6 dB. Dispersion volume may be lower than in class C due to biomass distribution: while here most of the biomass is in leaf blades, in class C most of the biomass is in leaves and stems (with more randomness in the orientation and position of the backscatterers). Class E has very low commission error, but omission error is high due to confusion with class C.

W. Water (9.1% of the total area). Three classes were assigned to water. W1 has a typical open water signature, with a constant low backscatter due to mirror reflection. W2 shows a constant median backscatter and was found in areas were double-bounce scatter contribution may occur: near ports, ships and vegetated coasts. W3 occurs in coastal areas that were probably completely covered by water between August and July, thus showing a signature similar to W1. With the hydrometric level decrease, W3 converged to W2 signature. Class W has low commission error (mainly due to misclassification of pixels belonging to class D) and very low omission error.

4. Conclusions

By using a multi-temporal approach, information about herbaceous land-cover classes can be extracted from X-Band HH SAR images from high-resolution COSMO-SkyMed data. The interaction mechanisms between the signal and the target (plant coverage and substrate) are highly affected by the hydrological status of the zone. Consequently, hydrological dynamics influenced the analysis of the SAR information and the classification obtained. Future works may include a sensitivity analysis on the minimum number of scenes needed to get the maximum accuracy and/or the comparison between object-based and per-pixel approaches. Electromagnetic models of the expected backscattering in each land-cover would improve the understanding on the underlying interaction mechanisms.

Acknowledgments: CSK imagery was acquired through the project "Multifrequency, multipolarization and multitemporal radar remote sensing of the Paraná River Wetland of Argentina: contribution of COSMO-Skymed data" (PI Patricia Kandus), of the Agenzia Spaziale Italiana. This work was funded by Comisión Nacional de Actividades Espaciales (CONAE, Argentina: AO SAOCOM N° 22) and by Agencia Nacional de Promoción Científica y Técnica (PICTO-CIN I–22; PICT 2012–2403). We thank Juan Nazar and Bello family for logistic help during field sampling. Several people collaborated during field work: M. Borro, M. Schoo, S. Varela, M. R. Derguy, M. S. Erario, A. Falthauser, M. Lanfiutti and V. Pacotti. We are also grateful to M. Salvia for her help during SAR data processing.

References

BLASCHKE, T. et al. Geographic object-based image analysis–towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 87, p. 180–191, 2014.

CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, v. 46, p. 35–46, 1991.

COSTA, M. P. F. et al. Biophysical properties and mapping of aquatic vegetation during the hydrological cycle of the Amazon floodplain using JERS-1 and Radarsat. *International Journal of Remote Sensing*, v. 23, n. 7, p. 1401–1426, 2002.

ENRIQUE, C. *Relevamiento y caracterización florística y espectral de los bosques de la Región del Delta del Paraná a partir de imágenes satelitales*. 93 p. Tese (Degree Thesis in Biological Sciences. Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.), Buenos Aires, 2009.

European Space Agency. *NEST v. 4C-1.1*. 2012. Available in: <http://nest.array.ca/web/nest>.

GRINGS, F. et al. Modeling temporal evolution of junco marshes radar signatures. *Geoscience and Remote Sensing, IEEE Transactions on*, v. 43, n. 10, p. 2238–2245, Oct 2005. ISSN 0196-2892.

HENDERSON, F. M.; LEWIS, A. J. Radar detection of wetland ecosystems: a review. *International Journal of Remote Sensing*, v. 29, n. 20, p. 5809–5835, out. 2008. ISSN 0143-1161.

INGLADA, J.; CHRISTOPHE, E. The Orfeo Toolbox remote sensing image processing software. In: *IGARSS (4)*. Cape Town: IEEE, 2009. p. 733–736.

JUNK, W. J.; BAYLEY, P. B.; SPARKS, R. E. The flood pulse concept in river-floodplain systems. *Canadian Special Publication of Fisheries and Aquatic Sciences*, Can. Spec. Publ. Fish. Aquat. Sci., v. 106, n. 1, p. 110–127, 1989.

MORANDEIRA, N. S. *Tipos funcionales de plantas en humedales de la planicie de inundación del Bajo Río Paraná (Entre Ríos , Argentina) y su observación con datos polarimétricos de radar*. 278 p. PhD Thesis in Biological Sciences. Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, 2014. Available in: <http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_5490_Morandeira.pdf>.

MORANDEIRA, N. S.; KANDUS, P. Distribution of herbaceous wetland plants in relation to hydrogeomorphic settings and soil features in the Lower Paraná River floodplain. *Aquatic Botany*, In revision.

MUELLER-DOMBOIS, D.; ELLENBERG, H. *Aims and methods of vegetation ecology*. New York: John Wiley, 1974. 547 p.

MURPHY, K. P. *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: The MIT Press, 2012. ISBN 9780262018029.

NOVO, E. M. L. a. d. M. et al. Relationship between macrophyte stand variables and radar backscatter at L and C band, Tucuruí reservoir, Brazil. *International Journal of Remote Sensing*, v. 23, n. 7, p. 1241–1260, 2002.

PULVIRENTI, L. et al. Flood monitoring using multi-temporal cosmo-skymed data: Image segmentation and signature interpretation. *Remote Sensing of Environment*, v. 115, n. 4, p. 990 – 1002, 2011. ISSN 0034-4257.

RAMSEY III, E. Radar remote sensing of wetlands. In: LUNETTA, R.; ELVIIDGE, C. (Ed.). *Remote sensing change detection: environmental monitoring methods and applications*. Chelsea, United States: Ann Harbor Press, 1998. p. 211–243.