Wetlands Ecol Manage (2020) 28:577–594
https://doi.org/10.1007/s11273-020-09731-2

ORIGINAL PAPER

A comparison of data mining techniques and multi-sensor analysis for inland marshes delineation

Joao P. D. Simioni · Laurindo A. Guasselli · Guilherme G. de Oliveira · Luis F. C. Ruiz · Gabriel de Oliveira

Received: 30 October 2018 / Accepted: 29 May 2020 / Published online: 8 June 2020

Abstract Inland Marsh (IM) is a type of wetland characterized by the presence of non-woody plants as grasses, reeds or sedges, with a water surface smaller than 25% of the area. Historically, these areas have been suffering impacts related to pollution by urban, industrial and agrochemical waste, as well as drainage for agriculture. The IM delineation allows to understand the vegetation and hydrodynamic dynamics and also to monitor the degradation caused by human-induced activities. This work aimed to compare four machine learning algorithms (classification and regression tree (CART), artificial neural network (ANN), random forest (RF), and k-nearest neighbors (k-NN)) using active and passive remote sensing data in order to address the following questions: (1) which of the four machine learning methods has the greatest potential for inland marshes delineation? (2) are SAR features more important for inland marshes delineation than optical features? and (3) what are the most accurate classification parameters for inland marshes delineation? To address these questions, we used data from Sentinel 1A and Alos Palsar I (SAR) and Sentinel 2A (optical) sensors, in a geographic object-based image analysis (GEOBIA) approach. In addition, we performed a vectorization of a 1975 Brazilian Army topographic chart (first official document presenting marsh boundaries) in order to quantify the marsh area losses between 1975 and 2018 by comparing it with a Sentinel 2A image. Our results showed that the method with the highest overall accuracy was k-NN, with 98.5%. The accuracies for the RF, ANN, and CART methods were 98.3%, 96.0% and 95.5%, respectively. The four classifiers presented accuracies exceeding 95%, showing that all methods have potential for inland marsh delineation. However, we note that the classification results have a great dependence on the input layers. Regarding the importance of the features, SAR images were more important in RF and ANN models, especially in the HV, HV + VH and VH channels of the Alos Palsar I L-band satellite, while spectral indices from optical images were more important in the marshes delineation with the CART method. In addition, we found that the CART and ANN methods presented the largest variations of the overall accuracy (OA) in relation to the different parameters tested. The multi-sensor approach was critical for the high OA values found in the IM delineation (> 95%). The four machine learning methods can be accurately applied for IM delineation, acting as an important low-cost tool for monitoring and managing these environments,

J. P. D. Simioni (✉) · L. A. Guasselli · G. G. de Oliveira · L. F. C. Ruiz
Center for Remote Sensing and Meteorology, Federal University of Rio Grande Do Sul, 9500 Bento Goncalves Av., Porto Alegre, RS 91501-970, Brazil
e-mail: joao.delapasse@ufrgs.br

G. de Oliveira
Department of Geography and Planning, 100 Saint George St, Toronto, ON M5S 3G3, Canada

in the face of advances in agriculture, soil degradation and pollution of water resources due to agrochemical dumping.

Keywords Machine learning · GEOBIA · Image classification · Wetlands

Introduction

Wetlands are ecosystems at the interface between aquatic and terrestrial environments; they may be continental or coastal, natural or artificial, permanently or periodically inundated by shallow water or consist of waterlogged soils (Junk et al. 2014). Wetlands have high biodiversity, playing an important role in maintaining/improving water quality, flood mitigation, aquifer recharge (Cowardin et al. 1979), microclimate regulation (Şimşek and Ödül 2018) and carbon sequestration (Mitsch et al. 2013). Wetlands can be classified as swamp, fen, bog, wet meadow and shallow water (Jahncke et al. 2018). Inland marshes are wetland types characterized by the presence of non-woody plants as grasses, reeds or sedges (Neiff et al. 2002; Keddy 2008). Due to the high accumulation of organic matter in these environments, it is possible to observe the formation of peatland layers or soils rich in organic matter (Sasser et al. 2017).

The delineation of inland marshes allows to understand the hidrodynamic patterns and also to monitor the development of agriculture over these areas (Junk et al. 2014). Although public awareness of wetland conservation has increased in the past few decades (Brock et al. 1999; Mui et al. 2015), the degradation of these areas has increased significantly. In the state of Rio Grande do Sul, southern Brazil, marsh areas are considered permanent preservation areas by legislation. However, there are no accurate inventories or mapping of the marsh areas located in the state. Due to conversion to agricultural areas, urbanization and waste disposal, important marsh areas have been extensively degraded in Rio Grande do Sul (Brenner 2016; Silva 2016).

The advance of geographic information systems has led to the improvement of methods used for the classification of remotely sensed images (França and Amaral 2013; Girolamo Neto 2014; Neves 2015). Remote sensing has an important role in the delineation of wetlands in regional scale (Karlson et al. 2019). The improved spatial, temporal, spectral and radiometric resolutions of optical sensors, especially Landsat 8 Operational Land Imager (OLI) and Sentinel 2 Multispectral Instrument (MSI) satellites, allows the development of strategies for continuous monitoring of vegetation patterns, water level oscillation and loss and degradation in wetlands (Sánchez-espinosa and Schröder 2019).

In the past few years, Synthetic Aperture Radar (SAR) sensors have been increasingly used for delineation and monitoring different wetland ecosystems in South America (Kandus et al. 2018), such as the Amazon river floodplain (Ferreira-Ferreira et al. 2015; Furtado et al. 2016; Pereira et al. 2018), the Brazilian Pantanal floodplain (Evans et al. 2010, 2014), the Paraná River floodplain (Morandeira et al. 2016; Gayol et al. 2019), and inland marshes areas (Grimson et al. 2019; Simioni et al. 2019).

SAR images have advantages in delineation of wetlands in comparison with optical images (White et al. 2015; Dabboor and Brisco 2018). SAR microwaves can collect data at day or night, and are able to penetrate through clouds and interact three-dimensionally with vegetation, detecting canopy structural characteristics and soil moisture (Morandeira et al. 2016; Pereira et al. 2018).

There are many SAR sensors, with different spatial resolutions, polarizations and wavelengths (Mahdianpari et al. 2017). The use of X (2.43–3.75 cm) and C (3.75–7.5 cm) bands are preferable for mapping herbaceous, less dense wetlands, and the L-band (15–30 cm) is indicated for mapping woody wetlands such as swamps and other high biomass wetlands (Dabboor and Brisco 2018).

With the wide range of satellite imagery available lately, different methods and techniques allow to delineate and classify wetlands with high accuracy. Geographic Object-Based Image Analysis (GEOBIA) offers a promising framework for segmenting landscapes into heterogeneous wetlands. The main benefits of GEOBIA in relation to pixel-based methods include: (1) the possibility of incorporating the shape, texture and any relevant contextual variables of the object into the classification, (2) to soften part of the local variation within the objects, which can reduce the speckle noise and increase the accuracy of the classification; and (3) to explain the landscape hierarchy, different land cover types and ecosystem

structure, by using multiple layers of objects at different spatial scales (Blaschke et al. 2014). Dronova (2015) states that there is an increase in precision using GEOBIA when compared to pixel-based approaches. Kamal and Phinn (2011) found a difference of $\sim$ 20% between GEOBIA and pixel-based aproaches.

Another method that has been widely used in recent years is the knowledge discovery in database (KDD), an exploratory automatic data analysis, designed to identify and to organize spatial patterns from large and complex remote sensing data sets (Maimon and Rokach 2005). One of KDD main steps is data mining (DM), a classification process used to identify patterns and to establish relationships within large data sets (Hand 2007). DM comprises a set of methods and algorithms with different usability and accuracy. Among them, we highlight the artificial neural networks (ANN), k-nearest neighbors (k-NN), classification and regression tree (CART), and random forest (RF) methods (Behrens and Scholten 2007). These algorithms have been successfully applied to delineate and classify different wetlands types. Bao and Ren (2011) applied ANN to classify the landscape heterogeneity in a wetland in DaLinor Lake and Van Beijma et al. (2014) applied RF to classify the vegetation in a salt marsh in Llanrhidian salt marshes, both with accuracy greater than 90%. RF is widely applied in studies to map wetlands with GEOBIA because of its ability to handle large datasets from different sources and the ability to evaluate the importance of features (Mahdianpari et al. 2017). The combination of multiple sources was a determining factor for the high accuracy ($> 93\%$ and $> 94\%$) observed by Silva et al. (2010), which used GEOBIA to map seasonal changes in aquatic macrophyte cover, and by Dubeau et al. (2017), which mapped wetlands using optical data, SAR and a Digital Elevation Model (DEM), respectively. By using DEM, SAR, and optical imaging data separately, Dubeau et al. (2017) found that the OA decreased to 89%. Wang et al. (2019) applied different DM methods to classify land cover in a coastal wetland in Linhong River Estuary Wetland with accuracy of 86.6% with RF and 77.2% with the k-NN method. Baker et al. (2006), using CART, obtained 73.1% accuracy in the mapping of wetlands and riparian areas in Gallatin River watershed, USA.

Although the integration between multi-sensor analysis, GEOBIA and data mining techniques for wetlands mapping has been well discussed in the literature (Belluco et al. 2006; Silva et al. 2010; Van Beijma et al. 2014; Walsh et al. 2014; Wester et al. 2018; Wang et al. 2019), this is the first study to apply these methods to map inland marshes. We hypothesize that the multi-sensor approach of the Sentinel-1 and Sentinel-2 satellites may have a high potential to discriminate the vegetation of wetlands.

Based on the consideration above, this work aimed to compare four machine learning algorithms (classification and regression tree (CART), artificial neural network (ANN), random forest (RF), and k-nearest neighbors (k-NN)) using active and passive remote sensing data in order to address the following questions: (1) which of the four machine learning methods has the greatest potential for inland marshes delineation? (2) are SAR features more important for inland marshes delineation than optical features? and (3) what are the most accurate classification parameters for inland marshes delineation? To address these questions, we used data from Sentinel 1A and Alos Palsar I (SAR) and Sentinel 2A (optical) sensors, in a geographic object-based image analysis (GEOBIA) approach. In addition, we performed a vectorization of a 1975 Brazilian Army topographic chart (first official document presenting marsh boundaries) to quantify marsh area losses between 1975 and 2018.

Methodology

Study area

The Banhado Grande (BG) marsh (29°57′ S, 50°41′ W) is located within the Gravataí river basin, state of Rio Grande do Sul, Brazil. The BG is a marsh remnant in a predominantly agricultural landscape, with an original area of 5,591 ha (Geographic Service Directorate 1975; Ramos et al. 2014) and it is included in the Banhado Grande System (BGS) wetlands complex. The BGS is an extensive mosaic of wetlands, which is formed by marshes, wet meadow (WM) and rice crops (RC). The BGS is delimited by a 20-m altimeter quota (Fig. 1).

In large flood pulses, the area comprised by the BGS floods, connecting different small inner marshes (Leite and Guasselli 2013; Simioni et al. 2017). The connectivity established during flood pulses is

responsible for large amounts of nutrient and sediment exchanges between wetlands.

The mean annual precipitation varies between 1500 mm and 1700 mm. August has the highest mean precipitation (140 mm) and April has the lowest (86 mm). The average annual temperature varies between 17 and 20 °C. In the hottest month, January, the average temperature oscillates between 23 and 26 °C, and in the coldest month, July, it oscillates between 11 and 14 °C.

Most of the vegetation in the BG is basically composed of three large compartments, (i) *Cyperaceaes* of the genus *Scirpus*; (ii) *Pontederiaceaes* of the genus *Eichhornia*, and (iii) transition vegetation, characterized by diverse macrophytes, grasses and shrubby. Plants of the Scirpus genus are characterized by dense vegetation, ranging from 1.5 to 2.5 m above the water level and cover about 28% of the BG area, and do not show significant seasonal differences, given the high leaf turnover, with continuous growth and appearance of new green leaves even during the winter months (Leite and Guasselli 2013). *Eichhornia* plants have seasonal variability regulated by flood pulses, have an average height between 10 and 16 cm and cover approximately 15% of the BG area (Leite and Guasselli 2013). The other 57% of the BG area consists of open water and transitional vegetation, characterized by diverse macrophytes, grasses and shrubby (Leite and Guasselli 2013). The BG marsh presents rice crops and wet meadow areas within its limits (Ramos et al. 2014). In addition to the marsh samples, we also collected samples for rice crop and wet meadow. The dominant species in each of the classes are described in Table 1.

Image processing

The satellite images were obtained from three sensors: Alos Palsar (Phased Array L-band Synthetic Aperture Radar) 1 L-band, Sentinel 1a C-band (level Ground Range Detected—GRD) and Sentinel 2a Multispectral Instrument with preprocessing level 1C (Table 2). Both Alos Palsar 1 L-band and Sentinel 1a C-band images were obtaining from the Alaska Satellite Facility website (www.vertex.daac.asf.alaska.edu). The Also Palsar I satellite collected data from 2006 and 2011, the image from 12/05/2011 is the only full-polarimetric image collected in the study area. We understand that different dates used in the study may interfere in the results, however there have been studies showing the efficacy of Alos Palsar I L-band to map marshes (Simioni et al. 2019). To minimize this problem, we used images with same day and month to the Alos Palsar I image acquisition. These images have full and dual-polarization, with L and C bands multi-frequencies. The Sentinel 2a MSI images were obtained from the Copernicus Program website (www.scihub.copernicus.eu).

For the synthetic aperture radar (SAR) images, we first conducted a radiometric calibration (RC), which is required to compare SAR images obtained with different sensors or same sensor at different times. The radiometric calibration corrects the pixel values in order to truly represent the backscatter of the imaged surface, so the values were converted to intensity. The speckle noise causes the cancellation (destructive interference) or the sum of the signal (interference), generating images with a "salt and pepper" effect. To mitigate this effect, the Lee sigma filter was applied, with a 5 × 5 movable window (Furtado et al. 2016). Due to the lateral geometry of the SAR image, the generated image is mapped in the inclined plane (Slant Range Domain). For the Sentinel 1a C-band image, we applied a terrain correction in order to convert the image from the inclined plane to the ground level (Ground Range Geometry) and then defined a cartographic system. For the terrain correction, altimetry images from the Shuttle Radar Topography Mission (SRTM) were used, with spatial resolution of 30 m and the Mercator Transverse Universe (UTM) projection system, spindle 22, southern hemisphere and horizontal datum SIRGAS 2000.

For the Sentinel 2a MSI satellite images, we first performed the atmospheric correction and transformation of radiance to reflectance by using the *sen2cor* tool (Kaplan and Avdan 2017). Then, the bands 11 and 12 of Sentinel 2a were resampled from 20 to 10 m by the bilinear interpolation method (Zhou et al. 2003). To calculate the spectral indices we used the bands B3 (green), B4 (red), B8 (NIR), B11 (SWIR 1) and B12 (SWIR 2). Since we have extracted these spectral indices from the Sentinel 2a images, we did not use the bands 1, 5, 6, 7, 8A, 9 and 10.

RC and refined Lee filter processing of the SAR images and atmospheric correction of Sentinel 2a magnets were performed on the SNAP desktop software. After pre-processing the images, we obtained 29 features: 16 from vegetation, soil and

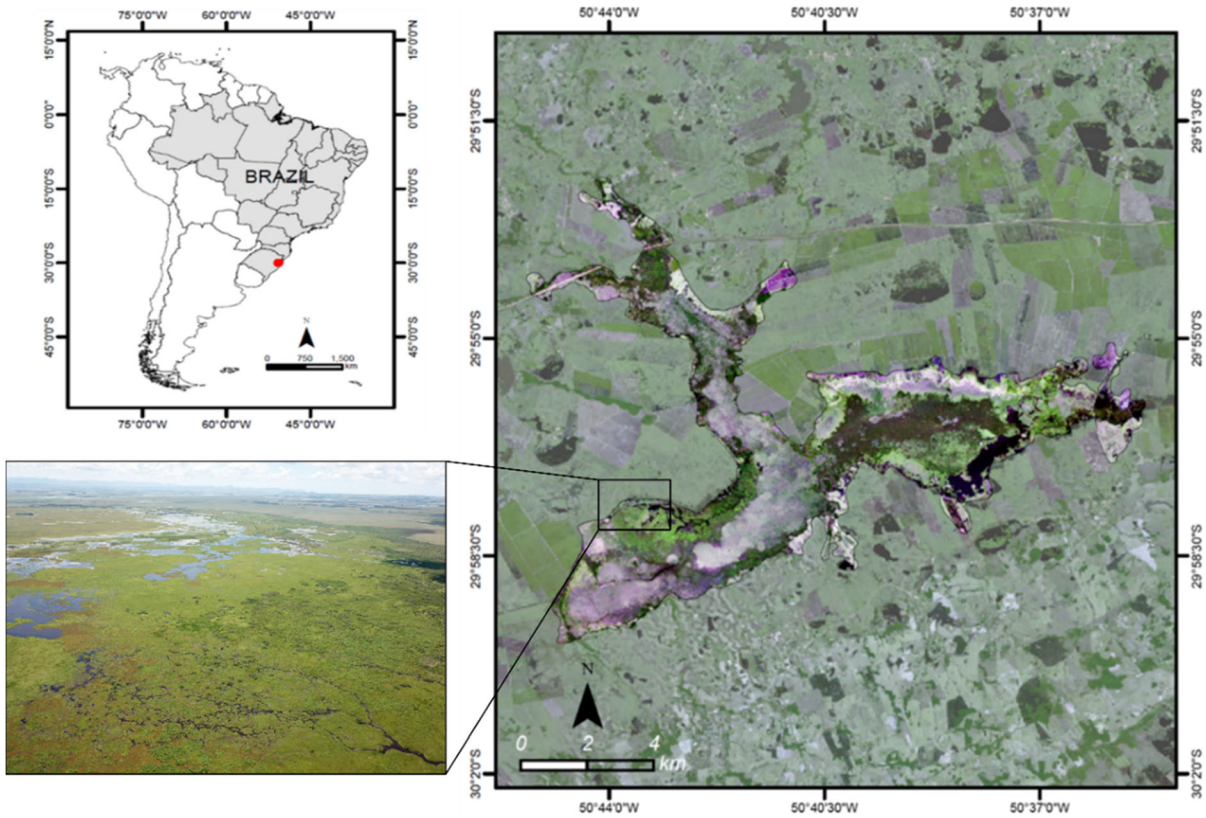


Fig. 1 Localization of the Banhado Grande (BG) marsh

Table 1 Orbital sensors used in the study

Date	Sensor	Polarization	Resolution (m)	Path/Frame	Frequency
12/05/2011	Alos Palsar 1	Quad-Pol	12.5	67/6590	L-band
12/04/2018	Sentinel 1a	VV + VH	10.0	66/78,211	C-band
12/07/2018	Sentinel 2a MSI	–	10.0	14,621/38	–

water indexes (Sentinel 2a image) and 13 from polarimetric channels (Sentinel 1a and Alos Palsar 1 images) (Table 3).

In addition, we performed the BG boundary vectorization using the 1:50:000 scale Brazilian Army topographic map (Geographic Service Directorate 1975) to quantify the loss of marsh area between 1975 and 2018. The topographic map was performed from surveys and fieldwork, and is the first official document presenting the BG limits.

Reference samples

The reference samples were collected by two methods. First, we collected 450 samples of the wet meadow and rice crop samples from fieldwork that took place between 12/01/2018 to 12/04/2018. To collect the samples, two Global Navigation Satellite System (GNSS) Ruide R90-X dual-frequency (L1/L2) receivers were used. The receiver's integrated GSM/GPRS modem enables the use of Networked Transport of RTCM via Internet Protocol (NTRIP) technology, which uses TOA-IP protocol to send GNSS receivers Real-time Kinematics (RTK) data. The reference base was the Brazilian Continuous Monitoring Network (BCMN) located in Porto Alegre, RS. Then, inland marshes, wet meadow and rice crops samples were collected by visual interpretation of World View 2 satellite images obtained on 12/08/2018. The multispectral resolution of the images is 2.0 m and the

Table 2 Indexes and polarizations used in the study

Index	Equation	Author(s)
Optical		
Weighted Difference Vegetation Index	WDVI = (B8 − g*B4)	Clevers et al. (1989)
Soil Adjusted Vegetation Index	SAVI = (((1 + L)*(B8-B4))/((B8 + B4 + L)))	Huete (1988)
Transformed Normalized Difference Vegetation Index	TNDVI = $\sqrt{}$((NDVI + 0.5))	Deering (1975)
Brightness Index	BI = $\sqrt{}$((((2*B4) + (2*B3))/2)	Escadafal (1989)
Brightness Index 2	BI_2 = $\sqrt{}$((((2*B4) + (2*B3) + (2*B8)))/3)	Escadafal (1989)
Ratio Vegetation Index	RVI = (B4/B8)	Pearson and Miller (1972)
Normalized Difference Water Index	NDWI = ((B8 − B11)/(B8 + B11))	Gao (1996)
Normalized Difference Water Index 2	NDWI_2 = ((B3 − B8)/(B3 + B8))	McFeeters (1996)
Normalized Difference Vegetation Index	NDVI = ((B8-B4)/(B8 + B4))	Rouse et al. (1973)
Normalized Difference Turbidity Index	NDTI = ((B4 − B3))/(B4 + B3))	Lacaux et al. (2007)
Normalized Difference Pond Index	NDPI = ((B3-B11*B12)/(B3 + B11*B12))	Lacaux et al. (2007)
Normalized Difference Index	NDI45 = ((B5 − B4)/(B5 + B4))	Delegido et al. (2011)
Modified Normalized Difference Water Index	MNDWI = ((B3-B11)/(B3 + B11))	Xu (2007)
Green Normalized Difference Vegetation Index	GNDVI = ((B8 − B3)/(B8 + B3))	Gitelson et al. (1996)
Difference Vegetation Index	DVI = (B8 − B4)	Richardson and Wiegand (1977)
Atmospherically Resistant Vegetation Index	ARVI = (((B8 − rb))/(B8 + rb))	Kaufman and Tanre (1992)

Polarization	Satellite	Band
SAR		
VV, VH and VV + VH	Sentinel 1a	C
VV, VH, HH and HV	Alos Palsar 1	L
VV + VH, VV + HH, and VV + HV	Alos Palsar 1	L
VH + HH and VH + HV	Alos Palsar 1	L
HH + HV	Alos Palsar 1	L

radiometric resolution is 11 bits. The visual interpretation of the World View 2 images was due to the impossibility of access to BG.

Segmentation

The segmentation was done through the object-based geographic analysis (GEOBIA), which segments the image in objects, which are groups of pixels representing homogeneous areas, entities or their (primitive) elements. These segments can then be classified into different categories by unsupervised, supervised or rule-based algorithms (Dronova 2015; Jones et al. 2018).

The segmentation was performed using the bands 2 (blue), 3 (green), 4 (red), 8 (near-infrared), 11 (Shortwave infrared SWIR) and 12 (SWIR) of Sentinel 2a MSI, and the channels VV and VH of Sentinel 1A, both with 10 m spatial resolution. We opted for the joint segmentation of SAR and optical images, because the joint analysis has shown better results compared to a sensor only (Macrì-Pellizzeri et al. 2002). For segmentation, we used the region-growth algorithm (Happ et al. 2013). The region-growth segmentation brings together adjacent pixels that meet a given heterogeneity criterion. Thus, the regions of the image are grouped or divided depending on whether the pixels have similar characteristics in terms of color, texture or shape. The region

Table 3 Dominant species for each class

Class	Dominant species/typology
Inland Marsh (IM)	Eichhornia crassipes, Eichhornia azurea, Salvinia auriculata, Hydrochleis nymphoides, Leersia sp., Cabomba australis, Leersia sp., Pontederia lanceolata, Nymphoides sp., Hygrophila sp., Polygonum spp., Myriophillum brasiliensis, Scirpus sp., Cyperus sp., Zizaniopsis sp., Eryngium pandanifolium sp.
Wet Meadow (WM)	Erianthus sp., Andropogon bicornis, Sida sp., Mimosa bimucronata, Panicum prionitis, grasses and shrubby
Rice Crops (RC)	Rice cultivation (Oryza sativa)

segmentation started with the launch of seed pixels from previously defined samples. It was obtaining 1,000 samples for each class analyzed (inland marshes, wet meadow and rice crop).

The region-growth algorithm is controlled by the similarity threshold (ST). Values from 0 to 1 were tested. High ST values mean higher freedom for growth of the regions (geographic objects) and vice versa. In addition, we evaluated the minimum number of cells (MNC), which determines the minimum size of the geographic objects. Low MNC values mean smaller geographic-objects generated by the segmentation. We tested MNC values between 200 and 600, varying by 200.

In order to incorporate the collected samples into the segmentation polygons, we performed the union of attributes by geographic location, through the spatial join plugin of QGis 3.4 Madeira. The segmentation was performed in the QGis 3.4 software, through the Geopatterns plugin (Ruiz 2019). Sample distribution ensured that no spatial object spanned more than one sampling point.

Classification and validation

In order to define the training and validation samples, 1000 points were generated on each of the three classes (inland marshes, wet meadow and rice crops), totaling 3000 samples. We used 80% of the samples for training and 20% for validation. The sampling and validation samples were randomly separated by the Geopatterns plugin, in the QGIS software 3.4. The accuracy of the classification was measured by the following indexes: (1) overall accuracy (OA) (Pontius and Millones 2011), (2) producer's accuracy (PA), and (3) user's accuracy (UA) (Story and Congalton 1986). The following non-parametric methods were used for classification: (1) classification and regression tree (CART), (2) artificial neural network (ANN), (3) random forest (RF), and (4) k-nearest neighbors (k-NN). The CART, RF and k-NN algorithms were applied based on the classification tool from the Geopatterns plugin (Ruiz 2019), and the ANN was applied in the MATLAB R2012b (MathWorks) software.

Classification and regression tree (CART)

The CART method uses the binary recursive partitioning analysis for class discrimination, in which each parent node is divided into two child nodes (Breiman et al. 1984; Lawrence and Wright 2001). The process is repeated by treating each child node as a parent node. When data from a node cannot be divided into additional nodes, it is called the terminal node. Once the first terminal node has been created, the algorithm repeats the procedure for each data set until all data are categorized as terminal nodes (Waheed et al. 2006).

The sample values 3, 5, 10, 15, 20, 25 and 30 were evaluated for tree depth and the values 8, 20, 40, 60, 80, 100 and 120 for the maximum number of samples at child nodes. The gini index (Rokach and Maimon 2005) was used as a measure of impurity of tree branches. The gini index measures the degree of heterogeneity of the data, searches the largest category in the data set (i.e., inland marsh) and tries to isolate it from other categories (Waheed et al. 2006).

The index of a node is given by Eq. 1:

$$Gini = 1 - \sum_{i=1}^{c} p_i^2 \quad (1)$$

where pi is the relative frequency of each class in each node, and c is the number of classes (in this case, 3).

Artificial neural network (ANN)

The ANN is a structured group of processing units, called artificial neurons, which comprises a mathematical model capable to approach complex relationships within data sets. Each neuron in the network can receive input signals, process them and send an output signal (Fantin-Cruz et al. 2011).

The ANN training was performed by the multilayered backpropagation method (Rumelhart et al. 1986) and the updating of the synaptic weights and internal connections of the network was done through the delta rule (Gurney 1997). The values of the input attributes were scaled from linear transformations according the Eq. 2:

$$\frac{(y_t - p_o)}{s_o} = ANN\left(\frac{(x_t - p_i)}{s_i}\right) \quad (2)$$

where x_t and y_t are the input and output variables, respectively, s_o and p_o are the scale and position parameters of the model outputs, and s_i and p_i are the scale and position parameters of the model inputs.

Several model configurations were tested, varying the number of neurons in the hidden layer of the network. The activation function used in both layers of the model was the sigmoidal. The output of the model was calculated from the function presented in Eq. 3:

$$ANN\left(\frac{(x_t - p_i)}{s_i}\right) = f_o\left(\sum_h w_o f_h\left(\sum_i w_h\left(\frac{x_t - p_i}{s_i}\right) + b_h\right) + b_0\right) + e_o \quad (3)$$

where w_h, b_h, f_h, w_o, b_o and f_o are the synaptic weights (w), biases (b) and activation functions (f), respectively, of the hidden (h) and output (o) layers, while e_o is the expected error in the output layer.

Once the weights initialization is random, we executed 30 iterations for each model configuration. At each iteration, a maximum number of learning cycles for training interruption was defined if the model did not reach the convergence threshold. After several tests, the number of 15,000 cycles was defined, since it was sufficient to achieve an ideal fit for the synaptic weights. The training and calibration parts were separated for an internal cross-validation process in order to avoid the model overfitting. The relevance of the input attributes of the model was quantified by the relative contribution (RC) index (Oliveira et al. 2015).

Random forest (RF)

The RF is an ensemble classifier, which is comprised by a decision tree set {h (X, vk), k, 1, …}, in which *vk* are independently sampled random vectors equally distributed in all the trees of the forest (Breiman 2001). The RF method can produce multiple decision trees by using a subset of samples and training variables. The result of the classification process is the *X* class, with the highest number of votes among all trees (Han et al. 2012).

After the forest formation, there are many decision trees to be tested and all contribute for the classification of the object under study by choosing which class the target attribute should belong to. The RF defines its decision by counting the votes of the predictor components in each class and then selecting the winning class in terms of the number of votes accumulated (Han et al. 2012).

The number of trees and their maximum depth can be adjusted. We evaluated the values between 5 and 50 for both parameters. As a criterion of division it was used the gini index.

k-nearest neighbors (k-NN)

The k-NN is a supervised classification method based on the proximity of its neighbors in a sample space (Mazzillo JR and Anzanello 2015). The aim of k-NN is to form a generalization that is based on a training set, maximizing the accuracy of the classification (Han et al. 2012). The parameter k controls the number of neighbors to be analyzed. The k-NN values were defined from the following distances: Manhattan (power = 1), Euclidean (power = 2) and Chebyshev (power = 3).

Results

Model calibrations

The highest OA for the CART method (95.5%) was obtained considering a MNC of 400, a maximum depth of 10 and a number of child nodes of 20 (Fig. 2). As we increase the number of child nodes, the OA decreases. For the CART method, the smallest OA values (90%) were found considering a number of child nodes of 120 and a maximum depth of 10. For the

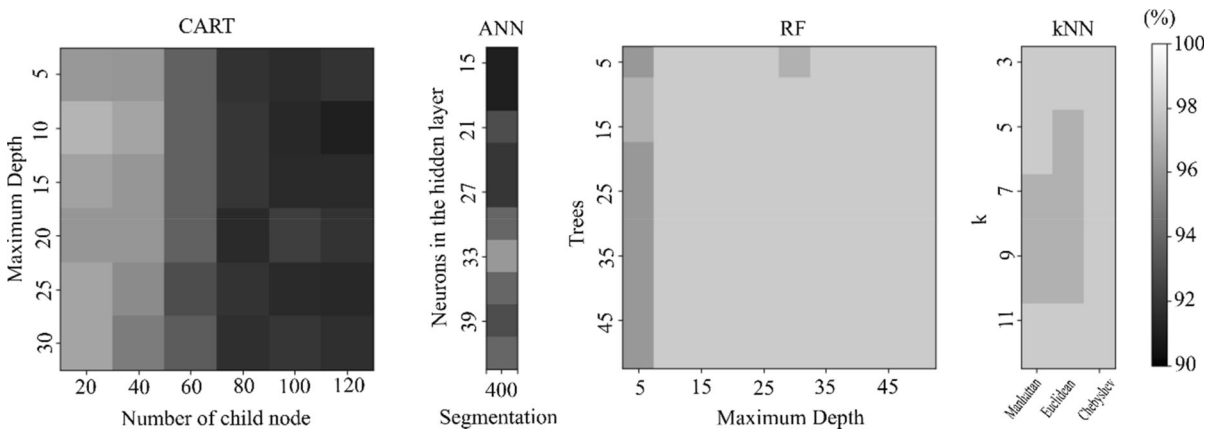


Fig. 2 Correct Proportion (CP) of each method: CART, ANN, RF, and k-NN

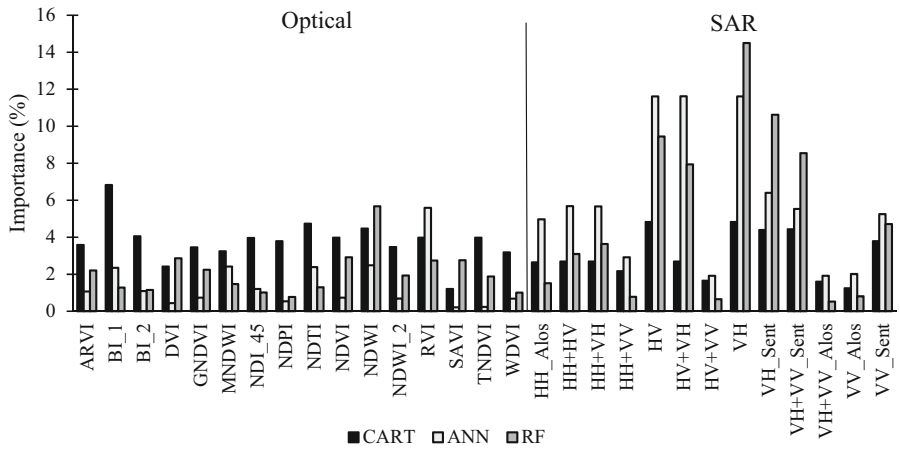


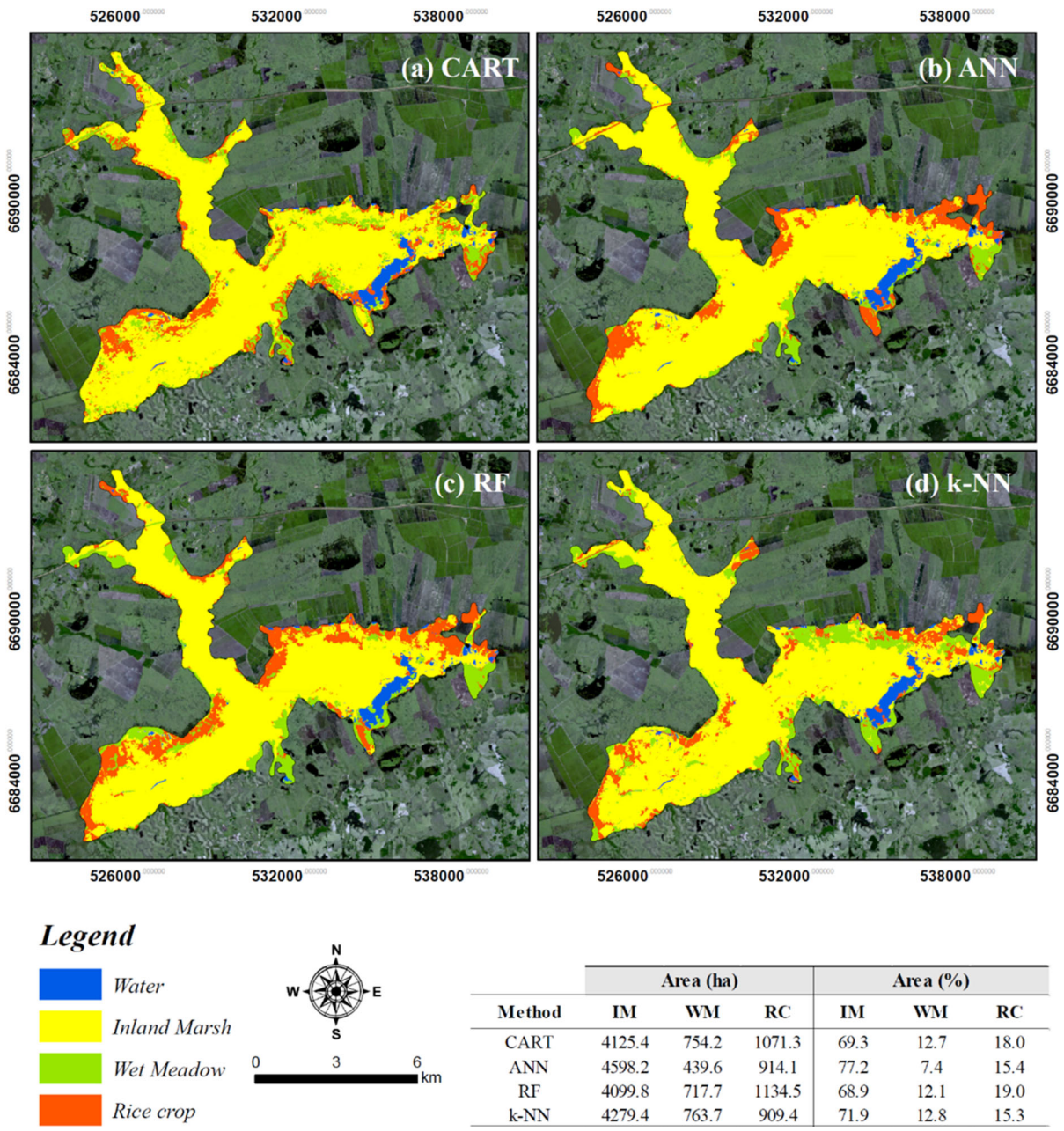
Fig. 3 Importance of the layers for CART, ANN and RF classifiers

ANN method, the highest OA (96%) was obtained with a MNC of 400 a number of neurons in the hidden layer equal to 33. As in the CART method, the lowest OA value for ANN was 90%. The RF presented the highest OA values (96%-98.2%) considering a MNC equal to 600. The k-NN presented the highest OA variations (96.5%-98.4%) by using the Chebyshev metric. In this metric, the distance between two vectors is the largest of their differences along any dimension.

Layer importance and mapping

The importance of the layers in the classifications is presented in Fig. 3. The optical images were more relevant only in the CART method. The ANN and RF methods presented a similarity in the importance of the layers, with greater relevance for SAR images. In the ANN, the highest importance was for the HV, VH and HV + VH channels, of about 11.6%. The highest importance for ANN using optical images were observed in the RVI index. For the SAR images, the greatest importance was obtained by using the VH channel of Alos Palsar I through the RF method (14.5%). For the optical images, the greatest importance was obtained byusing the NDWI index through the RF method (5.7%).

In Fig. 4, we observe that the CART, ANN, RF and k-NN methods classified, respectively, 69.3%, 77.2%, 68.9%, and 71.9% of BG as IM. The ANN presented more homogeneous IM areas in the central portion of BG in comparison with the other methods. The RF method presented the smallest IM area and the largest RC area among the analyzed methods. The k-NN classified more RC fragmented areas inside IM zones.



	Area (ha)			Area (%)		
Method	IM	WM	RC	IM	WM	RC
CART	4125.4	754.2	1071.3	69.3	12.7	18.0
ANN	4598.2	439.6	914.1	77.2	7.4	15.4
RF	4099.8	717.7	1134.5	68.9	12.1	19.0
k-NN	4279.4	763.7	909.4	71.9	12.8	15.3

Fig. 4 Classification results for the diferent methods: (1) CART, (2) ANN, (3) RF, and (4) k-NN

The confusion matrix shows the values of PA, UA and Kappa for the different classifiers (Table 4). For the CART method, there was confusion between WM and RC classes within the IM areas. For the ANN method, the IM areas presented confusion only with RC, reaching 97.7% for UA. Both PA (98.6%) and kappa (0.93) were one tenth higher than the CART method. Both RF and k-NN presented 100% on UA for IM. The PA for IM using RF and k-NN was lower when compared to CART and ANN (97.2%). Nevertheless, RF and k-NN showed the highest overall kappa values (0.97 and 0.98, respectively).

Table 4 Confusion matrix for the different methods: (1) CART, (2) ANN, (3) RF, and (4) k-NN

	CART				ANN				RF				k-NN			
	IM	WM	RC	UA (%)	IM	WM	RC	UA (%)	IM	WM	RC	UA (%)	IM	WM	RC	UA (%)
IM	186	5	7	93.9	210	0	5	97.7	209	0	0	100	209	0	0	100
WM	0	199	6	97.1	1	177	15	91.7	0	193	4	98.0	0	192	2	99.0
RC	3	6	187	95.4	2	1	188	98.4	6	0	187	96.9	6	1	189	96.4
PA (%)	98.4	94.8	93.5	95.5	98.6	99.4	90.4	96.0	97.2	100	97.9	98.3	97.2	99.5	99.0	98.5
Kappa	0.92				0.93				0.97				0.98			

IM inland Marsh, *WM* wet meadow, *RC* rice crop, *UA* User's accuracy, *PA* producer's accuracy

Discussion

In this study, we analized several methods to delineate inland marshes in southern Brazil using machine learning algorithms and a multi-sensor approach. Optical remote sensing data provides spectral information of molecular and structural features related to leaf area index, biomass and coverage of the canopy. On the other hand, SAR data provides structural (roughness and geometry) and dielectric (water presence) features (Van Beijma et al. 2014). Within this context, recent work have shown that GEOBIA is a promising method in the mapping of wetlands, since it allows to separate the heterogeneity of plant communities into objects, reducing the noise in the spectral data (Dronova 2015; Mui et al. 2015).

Simioni et al. (2018) mapped inland marshes using a pixel-based approach, finding a 95.9% accuracy. In this study, using GEOBIA, we obtained a 98.5% accuracy. Despite the fact that the present study shows a better accuracy in relation to the work of Simioni et al. (2018), it is understood that we cannot say that GEOBIA was more reliable, since the difference in accuracy values is small (< 3%) and may be associated with the date of obtaining the image or even the method of classification used.

We also highlight the potential of data mining techniques for inland marshes delineation. The four methods analyzed showed accuracy greater than 90% for all parameters tested. CART and ANN presented 6% differences in the OA values, demonstrating the importance to evaluate different parameters in the models.

Our results also showed that the RF and k-NN classifiers presented the highest OA for inland marshes delineation (98.3% and 98.5%, respectively). The k-NN was the method that presented the highest OA. This method is considered a lazy learning (Guo et al. 2003), because it simply stores the entire training set and postpones all effort towards inductive generalization until classification time (Wettschereck et al. 1997). Although studies have stated that k-NN is highly sensitive to the definition of its distance function (Wettschereck et al. 1997; Guo et al. 2003), we found that the variation of k between 3 and 13 showed the same OA in Chebyshev distance and decreased only 1% in some values of k in the Manhattan and Euclidean distances. One explanation for the success of k-NN is that lazy learnings perform better for forecasts that use a single training set and few classes (Webb et al. 2011). In this study, we chose to use only three classes (IM, WM and RC). We also found that the UA (forecasting ability to represent reality) achieved 100% in the inland marshes areas for the RF and k-NN methods. However, both methods had the lowest PAs (the quality of the classification of training set pixels) for IM, of the order of 97.2%. The best PA verified for IM areas was found in the ANN method.

The L-band VH channel had the greatest importance in mapping wetlands for RF and ANN methods. The SAR images show differences in the backscatter response between flooded and non-flooded vegetation, playing an important role in the hydrological monitoring of wetlands (Baghdadi et al. 2001; Ferreira-Ferreira et al. 2015). For the Brazilian Pantanal, Evans et al. (2014) found a OA of 80% by mapping the vegetation cover with the integration of SAR C and L-bands images. According to the authors, the main errors found in the classification were due to the

similarity of wetland type classes in terms of vegetation structure and flood dynamics, and, therefore, similar backscatter characteristics.

When mapping wetlands using SAR C and X-bands data, Mleczko and Mróz (2018) obtained a OA of ~ 65% using Sentinel 1 C-band. Franklin and Ahmed (2017) used SAR C- band and Landsat 8 OLI images together with Lidar-derived geomorphometric variables, and found a OA of 91%. The authors highlight the importance of the multisensor approach, especially with optical and SAR images, to improve wetland mapping accuracy.

We highlight that among the four classifiers analyzed in this study, RF is the most common method found in literature to map wetlands (Millard and Richardson 2013; Van Beijma et al. 2014; Mahdianpari et al. 2017), showing overall accuracies higher than 90%. The k-NN method, which presented the highest accuracy in this study (98.5%), did not show the same performance on other studies over similar environments. Na et al. (2015), i.e., assessing the k-NN and RF methods to map flooded forests in China observed accuracies of 41.6% and 83.6 using k-NN and RF, respectively.

In relation to the CART and ANN methods, the accuracies obtained in this study (OA of 95.5% and 96%, respectively) are still more reliable than previous studies aiming to map inland marshes. The CART method, for example, was evaluated by Pantaleoni et al. (2009), where the authors observed a precision of 76.1% by using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images. The ANN method has been used for wetlands mapping since the last two decades (Augusteijn and Warrender 1998; Ghedira et al. 2000; Fantin-Cruz et al., 2011; Chatziantoniou et al., 2017), with accuracies ranging between 70 and 90%. Using machine learning methods to map land use and land cover with emphasis on wetlands, Chatziantoniou et al. (2017) point out that the marsh class presented the lowest accuracy in the optical classification, but when using SAR data the results were improved due to texture and backscatter. In addition, the authors highlighted the GEOBIA high ability to discriminate rice crop and wetland areas by using the shape characteristic.

Figure 5 shows the different classes in the southern flank of BG. The potential of GEOBIA on recognizing shapes in the segmentation process was decisive for the good performance in the class discrimination, especially for RC. We highlight that this image was collected in December (12/11/2018), which is the final period of tillering in the rice areas. All shoots that grow after the initial parent shoot grows from a seed. Tillers are segmented, and each segment phase its own two-part leaf. From planting to harvest, rice passes by three stages (vegetative phase, reproduction phase and ripening phase). In this sense, we note that that date of the image may affect the reflectance values, given the structure of the crops and also physiological stress,

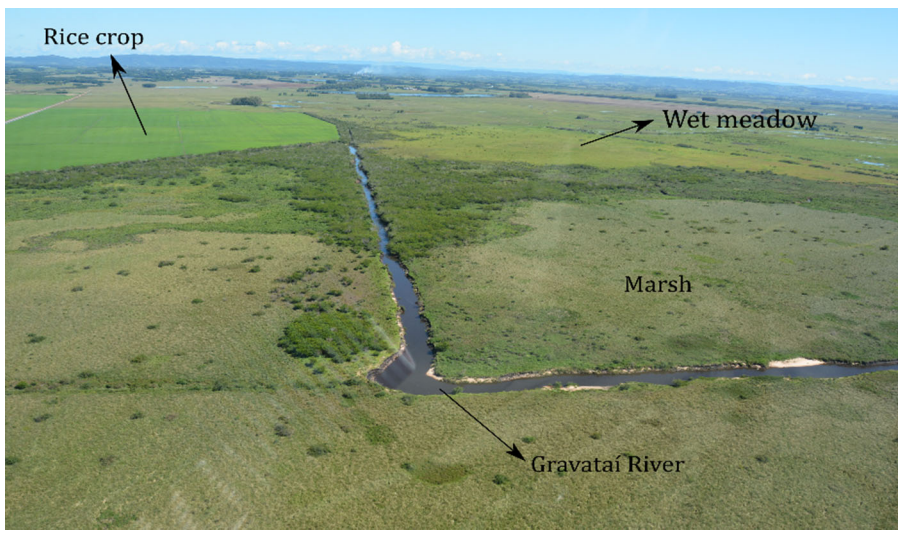


Fig. 5 Different types of vegetation within the Banhado Grande marsh

interfering in the classification results (Kuenzer and Knauer 2013).

In SAR images, Kuenzer and Knauer (2013) suggest that the final tillering period is ideal for separating rice from other classes, since the volume dispersion within the rice canopy and the interactions between plants and the surface of water result in an increase in the backscatter. Regarding polarization, we note that VV channels have higher backscatter coefficients than HH channels in the early stages of rice growth due to the physical structure of the plant, which consists mainly of short vertical leaves and stems during the early stages of rice growing (Kim et al. 2008). We found that the rice crop area within the BG ranges from 15.3% (k-NN) to 19% (RF), and is found mainly at the southern and northern limits of BG. However, the highest OA for RC areas (98.4%) were verified in the ANN method. This method estimated that 15.4% of the BG area in 1975 was currently occupied by RC.

Approximately 28% of the BG area is composed of *Cyperaceaes* of the genus *Scirpus* sp. (Fig. 6). This vegetation type is common in freshwater marshes of southern Brazil and Argentina (Pratolongo et al. 2005). This vegetation is dense, ranging from 1.5 to 2.5 m above the surface water. In a study on vegetation dynamics in a marsh, Guasselli (2005) found that there are no significant differences in reflectance of *Scirpus giganteus* during the year, given the high leaf turnover, with continuous growth and appearance of new green leaves even during the winter months (Pratolongo et al. 2005).

Regarding SAR images, Pope et al. (1994) point out that HH polarization allows to separate *Scirpus giganteus* from other aquatic vegetation types. The same is discussed by Bourgeau-Chavez et al. (2009), whichfound that HH/HV showed higher backscatter coefficients compared to wet meadow, for example. The models estimated that there are between 68.9% (CART) to 77.2% (ANN) of remaining natural vegetation within BG. The largest OA's found for IM areas were obtained by the RF and k-NN methods, with $\sim$ 100% accuracy. Both methods estimated similar values of remnants of IM areas in BG, 68.9% (RF) and 71.9% (k-NN).The models also estimated that between 7.4% (ANN) and 12.8% (k-NN) of the BG area correspond to wet meadow areas. The k-NN was the method that presented the highest overall accuracy in the WM classification, of $\sim$ 99% (Fig. 7). Ozesmi and Bauer (2002) highlight the high separability of WM from other vegetation types common in marshes areas, especially in the near infrared band. Chimner et al. (2019) used multi-sensor data (SAR and optical) to map Mountain Peatlands

Fig. 6 *Scirpus giganteus* in the Banhado Grande marsh

Fig. 7 Wet meadow in the Banhado Grande marsh

and wet meadows. The results showed that the wet meadow areas had the lowest overall accuracy (∼ 79%), being confused mainly with grassland or shrubland. It is important to note that we aimed to map the total area of inland marshes; however, in order to reduce the classes of analysis we incorporated grassland or shrubland areas as WM areas. This fact may justify the higher overall accuracy found in our study, since the largest WM class errors in the study by Chimner et al. (2019) are associated with the grassland or shrubland areas.

Finally, our results showed a decrease of the BG natural area since its first mapping, in 1975 (Geographic Service Directorate 1975; Ramos et al. 2014). It should be noted that the 1975 mapping was performed by visual interpretation of an aerophotogrammetric image, that is, a different method than the one used in this study. We understand that it is interesting to bring this information so that we can quantify the area of marsh lost since the first accurate mapping carried out. By using the k-NN method, which showed the highest OA, we found that ∼ 15.3% of the BG original area has been converted to agriculture. These results are important in terms of the development of more sustainable activities inn the area surrounding the BG, which has a key role in the hydrological patterns in the region.

Conclusion

We conclude that the four models analyzed can be applied with high precision (> 95.5%) to delineate inland marshes. The method that presented the highest overall accuracy to delineate inland marsh was k-NN (98.5%) accuracy, followed by RF (98.3%) accuracy. ANN presented a overall accuracy of 96%, and CART was the method that presented the lowest overall accuracy, of 95.5%.

For RF, the MNC 600 showed the highest accuracy and, for k-NN, the highest accuracy was found at MNC 200. It can be concluded that MNC values less than 600 allow high precision in marsh areas classification.

Regarding the importance of features, we conclude that SAR images are the most important in ANN and RF models. The greatest overall importance was found in the Alos Palsar I L-band VH channel, with 14.5% in the RF method, demonstrating the potential of SAR L-band images to delineate inland marshes. For the ANN method, the SAR channels HV, HV + VH and

VH presented the highest importance, both with 11.6%. Regarding the spectral indices, the feature BI_1 was more important in the CART model, 6.8%.The CART and ANN methods presented the largest variations of the accuracy in relation to the different parameters tested. For CART, we found that increasing the number of child nodes caused the production of trees with smaller overall accuracy, with a variation of $\sim$ 6% in the overall accuracy, and the parameter with the best OA was found in Neurons in the hidden layer 33. Thus, we verified that different parameters can generate differences of $\sim$ 6% in the final accuracy to delineate marshes in the CART and ANN methods.

For the RF, we did not obtain significant differences between the OA values by testing different parameters. We note that the OA values were similar between Maximum Depth 10 to 50 for all tree numbers except trees 5 and maximum depth 30. As in the RF method, the k-NN method showed little variation in OA values when testing different parameters. The values of k did not influence the OA results in the Chebyshev metric distance.

The method with the highest OA (k-NN) estimated that the BG area has decreased 28.1% since 1975, with 15.3% being converted to agricultural areas and 12.8% to wet meadow. This fact demonstrates the need for conservation strategies not only in BG, but in all wetlands of the state of Rio Grande do Sul.

We suggest, for further studies, the use of other data mining methods, as well as the use of both geomorphological attributes and land use temporal classification in the algorithms, as well as the evaluation of the results attributed to the use of different sets of images for the same classifier. We also recommend the application of these algorithms in other IM areas, including salt marshes, to verify the potential of the methods in different types of wetlands.

Acknowledgements This study was financed by the Coordination for the Improvement of Higher Education Personnel (CAPES) agency (Grant No: 88882.438941/2019-01).

References

Augusteijn MF, Warrender CE (1998) Wetland classification using optical and radar data and neural network classification. Int J Remote Sens 19:1545–1560. https://doi.org/10.1080/014311698215342

Baghdadi N, Bernier M, Gauthier R, Neeson I (2001) Evaluation of C-band SAR data for wetlands mapping. Int J Remote Sens 22:71–88. https://doi.org/10.1080/014311601750038857

Baker C, Lawrence R, Montagne C, Patten D (2006) Mapping wetlands and riparian areas using Landsat ETM+ imagery and decision-tree-based models. Wetlands 26:465. https://doi.org/10.1672/0277-5212(2006)26[465:mwarau]2.0.co;2

Bao Y, Ren J (2011) Wetland Landscape classification based on the BP neural network in DaLinor Lake Area. Procedia Environ Sci 10:2360–2366. https://doi.org/10.1016/J.PROENV.2011.09.368

Behrens T, Scholten T (2007) A comparison of data mining techniques in predictive soil mapping. Dev Soil Sci 31:353–617. https://doi.org/10.1016/S0166-2481(06)31025-2

Belluco E, Camuffo M, Ferrari S et al (2006) Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing. Remote Sens Environ 105:54–67. https://doi.org/10.1016/j.rse.2006.06.006

Blaschke T, Hay GJ, Kelly M et al (2014) Geographic object-based image analysis: towards a new paradigm. ISPRS J Photogramm Remote Sens 87:180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014

Bourgeau-Chavez LL, Riordan K, Powell RB et al (2009) Improving wetland characterization with multi-sensor, multi-temporal SAR and optical/infrared data fusion. In: Jedlovec G (ed) Advances in geoscience and remote sensing, 1st edn. InTech, London

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees Regression trees. Taylor & Francis, Oxfordshire

Brenner V (2016) Methodological proposal for renaturation of rectified section of the Gravataí River. Federal University of Rio Grande do Sul

Brock MA, Smith RGB, Jarman PJ (1999) Drain it, dam it: Alteration of water regime in shallow wetlands on the New England Tableland of New South Wales, Australia. Wetl Ecol Manag 7:37–46. https://doi.org/10.1023/A:1008416925403

Chatziantoniou A, Psomiadis E, Petropoulos G (2017) Co-Orbital Sentinel 1 and 2 for LULC mapping with emphasis on wetlands in a mediterranean setting based on machine learning. Remote Sens 9:1259. https://doi.org/10.3390/rs9121259

Chimner RA, Bourgeau-Chavez L, Grelik S et al (2019) Mapping mountain peatlands and wet meadows using multi-date, multi-sensor remote sensing in the Cordillera Blanca. Peru Wetlands. https://doi.org/10.1007/s13157-019-01134-1

Clevers JGPW, Leeuwen HJC Van, Sensing R, Verhoef W (1989) Estimanting apar by means of vegetation indeces: a sensitivity analysis. In: XXIX ISPRS congr techcomm VII interpret photogr remote sens data, pp 691–698

Cowardin LM, Carter V, Golet FC, LaRoe ET (1979) Classification of wetlands and deepwater habitats of the United States. Washigton

Craft C, Craft C (2016) Inland marshes. In: Creat restoring wetl, pp 95–127. https://doi.org/10.1016/B978-0-12-407232-9.00005-1

Dabboor M, Brisco B (2018) Wetland monitoring and mapping using synthetic aperture radar. In: Wetlands management: assessing risk and sustainable solutions. Intechopen, London, p 13

Deering DW (1975) Measuring forage production of grazing units from landsat MSS data. In: Proc 10th int symp remote sens environ, pp 1169–1178

Delegido J, Verrelst J, Alonso L, Moreno J (2011) Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors 11:7063–7081

de França FRS, Amaral HCDJ (2013) Aplicação de Técnicas de Mineração de Dados para o Mapeamento do Conhecimento na Aprendizagem de Programação: Uma Estratégia Baseada na Taxonomia de Bloom. In: Congresso Brasileiro de Informática na Educação. UNICAMP, Campinas, pp 759–768

Dronova I (2015) Object-based image analysis in wetland research: a review. Remote Sens.

Dubeau P, King DJ, Unbushe DG, Rebelo LM (2017) Mapping the Dabus Wetlands, Ethiopia, using random forest classification of Landsat PALSAR and topographic data. Remote Sens. https://doi.org/10.3390/rs9101056

Escadafal R (1989) Remote sensing of arid soil surface color with Landsat thematic mapper. Adv Space Res 9:159–163

Evans TL, Costa M, Telmer K, Silva TSFF (2010) Using ALOS/PALSAR and RADARSAT-2 to map land cover and seasonal inundation in the Brazilian Pantanal. IEEE J Sel Top Appl Earth Obs Remote Sens 3:560–575. https://doi.org/10.1109/JSTARS.2010.2089042

Evans TL, Costa M, Tomas WM, Camilo AR (2014) Large-scale habitat mapping of the Brazilian Pantanal wetland: a synthetic aperture radar approach. Remote Sens Environ 155:89–108. https://doi.org/10.1016/j.rse.2013.08.051

Fantin-Cruz I, Pedrollo O, Castro NMR et al (2011) Historical reconstruction of floodplain inundation in the Pantanal (Brazil) using neural networks. J Hydrol 399:376–384. https://doi.org/10.1016/j.jhydrol.2011.01.014

Ferreira-Ferreira J, Silva TSF, Streher AS et al (2015) Combining ALOS/PALSAR derived vegetation structure and inundation patterns to characterize major vegetation types in the Mamirauá Sustainable Development Reserve, Central Amazon floodplain, Brazil. Wetl Ecol Manag 23:41–59. https://doi.org/10.1007/s11273-014-9359-1

Franklin S, Ahmed O (2017) Object-based Wetland Characterization Using Radarsat-2 Quad-Polarimetric SAR Data, Landsat-8 OLI Imagery, and Airborne Lidar-Derived Geomorphometric Variables. Photogramm Eng Remote Sens 83:27–36. https://doi.org/10.14358/PERS.83.1.27

Furtado LF, Silva TSF, Novo EML (2016) Dual-season and full-polarimetric C band SAR assessment for vegetation mapping in the Amazon várzea wetlands. Remote Sens Environ 174:212–222. https://doi.org/10.1016/J.RSE.2015.12.013

Gao B (1996) NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens Environ 58:257–266

Gayol MP, Morandeira NS, Kandus P (2019) Dynamics of shallow lake cover types in relation to Paraná River flood pulses: assessment with multitemporal Landsat data. Hydrobiologia 833:9–24. https://doi.org/10.1007/s10750-018-3847-x

Geographic Service Directorate (1975) Topographic chart. In: Brazilian Army. https://www.quoos.com.br//carta1dl/1dl_ASX139santo_antonio_da_patrulha.jpg. Accessed 26 Aug 2019

Ghedira H, Bernier M, Ouarda TBMJ (2000) Application of neural networks for wetland classification in RADARSAT SAR imagery. In: IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. IEEE, pp 675–677

Girolamo Neto CN (2014) Potencial de técnicas de mineração de dados para o mapeamento de áreas cafeeiras. Instituto Nacional de Pesquisas Espaciais

Gitelson AA, Kaufman YJ, Merzlyak MN (1996) Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens Environ 58:289–298

Grimson R, Gayol MP, Grimson R, et al (2019) Freshwater marsh classification in the Lower Paraná River floodplain: an object-based approach on multitemporal X-band COSMO- SkyMed data. https://doi.org/10.1117/1.JRS.13.014531

Guasselli L a. (2005) Dinâmica da vegetação no banhado do Taim, RS

Guo G, Wang H, Bell D et al (2003) KNN model-based approach in classification. Springer, Berlin, Heidelberg, pp 986–996

Gurney K (1997) An introduction to neural networks, 1st edn. Routledge, New York

Han J, Kamber M, Pei J (2012) Data mining: concepts and techniques, 3rd edn. Elsevier, Waltham

Hand DJ (2007) Principles of data mining. Drug Saf 30:621–622. https://doi.org/10.2165/00002018-200730070-00010

Happ PN, Feitosa RQ, Bentes C, Farias R (2013) Um algoritmo de segmentação por crescimento de regiões para GPUs. Bol Ciencias Geod 19:208–226. https://doi.org/10.1590/S1982-21702013000200004

Huete AR (1988) A soil-adjusted vegetation index (SAVI). Remote Sens Environ 25:295–309

Jahncke R, Leblon B, Bush P, LaRocque A (2018) Mapping wetlands in Nova Scotia with multi-beam RADARSAT-2 Polarimetric SAR, optical satellite imagery, and Lidar data. Int J Appl Earth Obs Geoinf 68:139–156. https://doi.org/10.1016/J.JAG.2018.01.012

Jones T, Marzen L, Mitra C, Barbour M (2018) Identification and classification of geographically isolated wetlands in North Alabama using geographic object based image analysis (GeOBIA). Geocarto Int 6049:1–16. https://doi.org/10.1080/10106049.2018.1438527

Junk WJ, Piedade MTF, Lourival R et al (2014) Brazilian wetlands: Their definition, delineation, and classification for research, sustainable management, and protection. Aquat Conserv Mar Freshw Ecosyst 24:5–22. https://doi.org/10.1002/aqc.2386

Kamal M, Phinn S (2011) Hyperspectral data for mangrove species mapping: A comparison of pixel-based and object-

based approach. Remote Sens 3:2222–2242. https://doi.org/10.3390/rs3102222

Kandus P, Minotti PG, Morandeira NS et al (2018) Remote sensing of wetlands in South America: Status and challenges. Int J Remote Sens 39:993–1016. https://doi.org/10.1080/01431161.2017.1395971

Kaplan G, Avdan U (2017) Mapping and Monitoring Wetlands Using SENTINEL-2 Satellite Imagery. ISPRS Ann Photogramm Remote Sens Spat Inf Sci 4:271–277

Karlson M, Gålfalk M, Crill P et al (2019) Delineating northern peatlands using Sentinel-1 time series and terrain indices from local and regional digital elevation models. Remote Sens Environ 231:111252. https://doi.org/10.1016/J.RSE.2019.111252

Kaufman YJ, Tanre D (1992) Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE Trans Geosci Remote Sens 30:261–270. https://doi.org/10.1109/36.134076

Keddy P (2008) Freshwater marshes. Encycl Ecol. https://doi.org/10.1016/B978-008045405-4.00338-4

Kim YH, Hong SY, Lee H (2008) Radar backscattering measurement of a paddy rice field using multi-frequency(l, c and x) and full-polarization. Int Geosci Remote Sens Symp 4:553–556. https://doi.org/10.1109/IGARSS.2008.4779781

Kuenzer C, Knauer K (2013) Remote sensing of rice crop areas. Int J Remote Sens 34:2101–2139. https://doi.org/10.1080/01431161.2012.738946

Lacaux JP, Tourre YM, Vignolles C, Ndione JA, Lafaye M (2007) Classification of ponds from high-spatial resolution remote sensing: application to Rift valley fever epidemics in Senegal. Remote Sens Environ 106:66–74

Lawrence RL, Wright A (2001) Rule-Based Classification Systems Using Classification and Regression Tree (CART) Analysis. Photogramm Eng Remote Sens 67:1137–1142. https://doi.org/10.1016/S0034-4257(01)00247-4

Leite MG, Guasselli LA (2013) Spatio-temporal dynamics of aquatic macrophytes in Banhado Grande, Gravataí River basin. Para Onde 7:17–24

Macrì-Pellizzeri T, Oliver CJ, Lombardo P (2002) Segmentation-based joint classification of SAR and optical images. IEEE Proc Radar Sonar Navig 149:281–296. https://doi.org/10.1049/ip-rsn:20020714

Mahdianpari M, Salehi B, Mohammadimanesh F, Motagh M (2017) Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. ISPRS J Photogramm Remote Sens 130:13–31. https://doi.org/10.1016/j.isprsjprs.2017.05.010

Maimon O, Rokach L (2005) Introduction to knowledge discovery in databases. In: Maimon O, Rokach L (eds) data mining and knowledge discovery handbook. Springer, New York, pp 1–15

Mazzillo JRCA, Anzanello MJ (2015) Sistemática de seleção de variáveis para classificação de produtos em categorias de modelos de reposição. Gest Prod 22:201–212. https://doi.org/10.1590/0104-530X1052-13

McFeeters SK (2007) The use of the normalized difference water index (NDWI) in the delineation of open water features. Int J Remote Sens 17:1425–1432

Millard K, Richardson M (2013) Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR-SAR fusion using a random forest classifier. Can J Remote Sens 39:290–307. https://doi.org/10.5589/m13-038

Mitsch WJ, Bernal B, Nahlik AM et al (2013) Wetlands, carbon, and climate change. Landsc Ecol 28:583–597. https://doi.org/10.1007/s10980-012-9758-8

Mleczko M, Mróz M (2018) Wetland mapping using SAR Data from the Sentinel-1A and TanDEM-X missions: a comparative study in the Biebrza Floodplain (Poland). Remote Sens 10:78. https://doi.org/10.3390/rs10010078

Morandeira N, Grings F, Facchinetti C et al (2016) Mapping plant functional types in floodplain wetlands: an analysis of C-band polarimetric SAR data from RADARSAT-2. Remote Sens 8:174. https://doi.org/10.3390/rs8030174

Mui A, He Y, Weng Q (2015) An object-based approach to delineate wetlands across landscapes of varied disturbance with high spatial resolution satellite imagery. ISPRS J Photogramm Remote Sens 109:30–46. https://doi.org/10.1016/j.isprsjprs.2015.08.005

Na XD, Zang SY, Wu CS, Li WL (2015) Mapping forested wetlands in the Great Zhan River Basin through integrating optical, radar, and topographical data classification techniques. Environ Monit Assess 187:696. https://doi.org/10.1007/s10661-015-4914-7

Neiff JJ, Patino CA, Neiff ASP et al (2002) Response of a natural marsh to chemical and biological inputs of eutrophic waters (Saladas, Corrientes, Argentina). Rev Bras Recur Hídricos 7:53–62. https://doi.org/10.21168/rbrh.v7n3.p53-62

Neves AK (2015) O Uso de Séries Temporais e Mineração de Dados no Mapeamento de Cobertura do Solo e seus Padrões São José dos Campos. Instituto Nacional de Pesquisas Espaciais

Oliveira GG, Pedrollo OC, Castro NMR (2015) Simplifying artificial neural network models of river basin behaviour by an automated procedure for input variable selection. Eng Appl Artif Intell 40:47–61. https://doi.org/10.1016/j.engappai.2015.01.001

Ozesmi SL, Bauer ME (2002) Satellite remote sensing of wetlands. Wetl Ecol Manag 10:381–402. https://doi.org/10.1023/A:1020908432489

Pantaleoni E, Wynne RH, Galbraith JM, Campbell JB (2009) Mapping wetlands using ASTER data: a comparison between classification trees and logistic regression. Int J Remote Sens 30:3423–3440. https://doi.org/10.1080/01431160802562214

Pearson RL, Miller LD (1972) Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass Prairie. Remote Sens Environ 8:1355

Pereira L, Furtado L, Novo E et al (2018) Multifrequency and Full-Polarimetric SAR assessment for estimating above ground biomass and leaf area index in the Amazon Várzea Wetlands. Remote Sens 10:1355. https://doi.org/10.3390/rs10091355

Pontius RG, Millones M (2011) Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int J Remote Sens 32:4407–4429. https://doi.org/10.1080/01431161.2011.552923

Pope KO, Rey-Benayas JM, Paris JF (1994) Radar remote sensing of forest and wetland ecosystems in the Central

American tropics. Remote Sens Environ 48:205–219. https://doi.org/10.1016/0034-4257(94)90142-2

Pratolongo P, Vicari R, Kandus P, Malvárez I (2005) A new method for evaluating Net Aboveground Primary Production (NAPP) of Scirpus giganteus (Kunth). Wetlands 25:228–232. https://doi.org/10.1672/0277-5212(2005)025[0228:ANMFEN]2.0.CO;2

Ramos RA, Pasqualetto AI, Balbueno RA, et al (2014) Mapeamento e diagnóstico de áreas úmidas no Rio Grande do Sul, com o uso de ferramentas de geoprocessamento. In: Anais do Simposio de Áreas Protegidas. Viçosa, pp 17–21

Richardson AJ, Wiegand CL (1977) Distinguishing vegetation from soil background information. Photogramm Eng Remote Sens 43:1541–1552

Rokach L, Maimon O (2005) Decision trees. Data mining and knowledge discovery handbook, 1st edn. Springer, Berlin, pp 165–192

Rouse JW, Hass RH, Schell JA, Deering DW (1973) Monitoring vegetation systems in the great plains with ERTS. In: Third earth resour technol satell symp, vol 1, pp 309–317. https://doi.org/citeulike-article-id:12009708

Ruiz LF (2019) Proposed object-based classification and post-classification of land cover and use by unmanned aerial vehicle images. Federal University of Rio Grande do Sul

Rumelhart DE, Hinton GE, Williams RJ (1986) © Learning representations by back-propagating errors. Nature 323:264–265. https://doi.org/10.1038/324227a0

Sánchez-espinosa A, Schröder C (2019) Land use and land cover mapping in wetlands one step closer to the ground : Sentinel-2 versus landsat 8. J Environ Manage 247:484–498. https://doi.org/10.1016/j.jenvman.2019.06.084

Sasser CE, Evers-Hebert E, Holm GO et al (2017) Relationships of marsh soil strength to belowground vegetation biomass in louisiana coastal marshes. Wetlands. https://doi.org/10.1007/s13157-017-0977-2

Silva RC da (2016) Estudo da dinâmica da fragilidade ambiental na Bacia Hidrográfica do Rio Gravataí, RS. Universidade Federal da Bahia

Silva TSF, Costa MPF, Melack JM (2010) Spatial and temporal variability of macrophyte cover and productivity in the eastern Amazon floodplain: a remote sensing approach. Remote Sens Environ 114:1998–2010. https://doi.org/10.1016/j.rse.2010.04.007

Simioni J, Guasselli L, Etchelar C (2017) Connectivity among Wetlands of EPA of Banhado Grande, RS Conetividade entre Áreas Úmidas, APA do Banhado Grande, RS. Braz J Water Resour. https://doi.org/10.1590/2318-0331.011716096

Simioni JPD, Guasselli LA, Nascimento VF et al (2019) Integration of multi-sensor analysis and decision tree for evaluation of dual and quad: Pol SAR in L- and C- bands applied for marsh delineation. Environ Dev Sustain. https://doi.org/10.1007/s10668-019-00442-0

Simioni JPD, Nascimento VF, Guasselli LA et al (2018) Small inner marsh area delimitation using remote sensing spectral indexes and decision tree method in southern Brazil. Rev Teledetección. https://doi.org/10.4995/raet.2018.10366

Şimşek ÇK, Ödül H (2018) Investigation of the effects of wetlands on micro-climate. Appl Geogr. https://doi.org/10.1016/j.apgeog.2018.05.018

Story M, Congalton RG (1986) Remote sensing brief accuracy assessment: a user's perspective. Photogramm Eng Remote Sens 52:397–399

Van Beijma S, Comber A, Lamb A (2014) Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. Remote Sens Environ. https://doi.org/10.1016/j.rse.2014.04.010

Waheed T, Bonnell RB, Prasher SO, Paulet E (2006) Measuring performance in precision agriculture: CART-A decision tree approach. Agric Water Manag 84:173–185. https://doi.org/10.1016/j.agwat.2005.12.003

Walsh N, Bhattasali N, Chay F (2014) Mapping Tidal Salt Marshes

Wang X, Gao X, Zhang Y et al (2019) Land-cover classification of coastal wetlands using the RF algorithm for worldview-2 and Landsat 8 images. Remote Sens 11:1927. https://doi.org/10.3390/rs11161927

Webb GI, Sammut C, Perlich C et al (2011) Lazy Learning. Encyclopedia of Machine Learning. Springer, US, Boston, MA, pp 571–572

Wester SJ, Grimson R, Minotti PG et al (2018) Hydrodynamic modelling of a tidal delta wetland using an enhanced quasi-2D model. J Hydrol 559:315–326. https://doi.org/10.1016/j.jhydrol.2018.02.014

Wettschereck D, Aha DW, Mohri T (1997) A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. Artif Intell Rev 11:273–314. https://doi.org/10.1023/A:1006593614256

White L, Brisco B, Dabboor M et al (2015) A collection of SAR methodologies for monitoring wetlands. Remote Sens 7:7615–7645. https://doi.org/10.3390/rs70607615

Xu H (2007) Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int J Remote Sens 27:3025–3033

Zhou Q, Jing Z, Jiang S (2003) Remote sensing image fusion for different spectral and spatial resolutions with bilinear resampling wavelet transform. In: Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems. IEEE, Shanghai, pp 1206–1213