Remote Sensing of Environment 187 (2016) 156–168

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas

Charlotte Pelletier[a,*], Silvia Valero[a], Jordi Inglada[a], Nicolas Champion[b], Gérard Dedieu[a]

[a] CESBIO-UMR 5126, 18 avenue Edouard Belin, 31401 Toulouse CEDEX 9, France
[b] IGN Espace-MATIS/Université Paris-Est, 6 avenue de l'Europe, 31521 Ramonville CEDEX, France

ARTICLE INFO

Article history:
Received 22 February 2016
Received in revised form 13 September 2016
Accepted 3 October 2016
Available online xxxx

Keywords:
Land cover mapping
Satellite image time series
High resolution
Random forest
Temporal features
Sentinel-2

ABSTRACT

New remote sensing sensors will acquire High spectral, spatial and temporal Resolution Satellite Image Time Series (HR-SITS). These new data are of great interest to map land cover thanks to the combination of the three high resolutions that will allow a depiction of scene dynamics. However, their efficient exploitation involves new challenges, especially for adapting traditional classification schemes to data complexity. More specifically, it requires: (1) to determine which classifier algorithms can handle the amount and the variability of data; (2) to evaluate the stability of classifier parameters; (3) to select the best feature set used as input data in order to find the good trade-off between classification accuracy and computational time; and (4) to establish the classifier accuracy over large areas.

This work aims at studying these different issues, and more especially at demonstrating the ability of state-of-the-art classifiers, such as Random Forests (RF) or Support Vector Machines (SVM), to classify HR-SITS. For this purpose, several studies are carried out by using SPOT-4 and Landsat-8 HR-SITS in the south of France. Firstly, the choice of the classifier is discussed by comparing RF and SVM algorithms on HR-SITS. Both classifiers show their ability to tackle the classification problem with an Overall Accuracy (OA) of 83.3 % for RF and 77.1 % for SVM. But RF have some advantages such as a small training time, and an easy parameterization. Secondly, the stability of RF parameters is appraised. RF parameters appear to cause little influence on the classification accuracy, about 1% OA difference between the worst and the best parameter configuration. Thirdly, different input data – composed of spectral bands with or without spectral and/or temporal features – are proposed in order to enhance the characterization of land cover. The addition of features improves the classification accuracy, but the gain in OA is weak compared with the increase in the computational cost. Eventually, the classifier accuracy is assessed on a larger area where the landscape variabilities affect the classification performances.

1. Introduction

New satellite missions - such as Sentinel, Venμs, or Landsat Data Continuity Mission (LDCM) - will acquire High Resolution optical Satellite Image Time Series (HR-SITS). The large swath, the short revisit time, the high spatial resolution of about 10 m, and the spectral bands from visible to infra-red will become essential to monitor large territories. For instance, Sentinel-2 satellites will provide a global cover of continental surfaces every five days in 13 spectral bands from 10 to 60 m (Drusch et al., 2012).

The use of HR-SITS is of great interest for the development of high-level operational products, such as global land cover maps. For this specific purpose, the processing should be designed to operate with a robust classifier, and suitable input data. More precisely, standard classification processing chains need to be adapted in order to: (1) achieve a good trade-off between classification performances, the stability of the classifier and computational time; (2) provide the classifier with the best input data, which fully exploit the quantity of information given by HR-SITS; and (3) deal with the data variability arising from the landscape diversity over large areas. Note that HR-SITS will cover large areas where climate, human activities, and landscape soils and slopes may differ.

Multi-temporal classification issues have been tackled in different ways in remote sensing literature. Some approaches have proposed to select key dates representing discriminative phenological stages in order to classify multi-temporal data. The most common strategy consists in selecting images acquired on two different seasons (Rogan et al., 2002; Rodriguez-Galiano et al., 2012). These key dates generally correspond to a low cloud cover period,

* Corresponding author.

http://dx.doi.org/10.1016/j.rse.2016.10.010
0034-4257/

and to higher differences in spectral signatures of vegetation categories. Carrão et al. (2008) and Masse et al. (2011) also proposed automated date selection methods using sorting criteria and genetic algorithms respectively. They sought to extract the most condensed and pertinent information from high temporal resolution SITS. However, the selection of key dates is a complex step for operational land cover mapping because the acquisition of (cloud free) images is not ensured at the key dates, and climatic change or human activities may change these key dates from one year to another.

More general methods (without date selection) classified SITS with a high temporal resolution, but at a low spatial resolution (Alcantara et al., 2012; Gong et al., 2006). On the contrary, the classification of high spatial resolution images has been introduced with single or few images (Lu et al., 2004). Generally, a unique image cannot distinguish all the land cover categories. Likewise, global land cover initiatives suffer from the lack of sensors combining high temporal and spatial resolutions (Wardlow and Egbert, 2008; Gong et al., 2013; Chen et al., 2015; Wang et al., 2015).

Therefore, an operational scheme for HR-SITS should work with high spatial, spectral, and temporal resolutions. In addition of large area processing, this scheme should also respect user requirements such as a frequent update, a high number of land cover categories, or an automated process.

Above-mentioned classification techniques rely on supervised and unsupervised approaches. Many comparisons between both types have been performed in the literature, showing that supervised methods — Maximum Likelihood (ML), Neural Networks (NN) (multilayer perceptron of Atkinson and Tatnall (1997)), Support Vector Machines (SVM) (Vapnik, 1995, 1998), and Decision Trees (DT) (Breiman et al., 1984; Hansen et al., 1996; Friedl and Brodley, 1997) — outperform unsupervised methods (Szuster et al., 2011; Khatami et al., 2016).

More specifically, ensemble learning methods (bootstrap, boosting, *etc.*) have recently received a strong interest. They consist in learning several weak classifiers to generate a classifier with a strong decision rule. A well-known ensemble learning method is Random Forests (RF) of Breiman (2001), which has demonstrated its ability to yield accurate land cover maps (Belgiu and Drăguţ, 2016). It accomplished performances comparable to traditional classifiers such as DT or SVM, with a lower computational time (Inglada et al., 2015; Rodriguez-Galiano et al., 2012; Gislason et al., 2006).

To help the classifier to learn the decision rule, features (also named variables and attributes) are used as input data in the classification system. The number and quality of input features are related to resulting accuracies, but also to computational time. Hundreds of spectral features, such as the NDVI (Normalized Difference Vegetation Index) for vegetation depiction, NDWI (Normalized Difference Water Index) for water detection or NDBI (Normalized Difference Built-up Index) for building detection, have been proposed and compared in remote sensing domains (Mróz et al., 2004; Silleos et al., 2006; Yeom et al., 2013). In the same way, a great number of spatial features (geometric, texture, *etc.*) have been proposed (Haralick, 1979; Trias-Sanz, 2006; Lv et al., 2014). For land cover mapping, temporal features receive less interest because of the lack of high temporal SITS. However, they have proven their ability to improve the classification accuracies, especially on vegetation categories (Jia et al., 2014; Valero et al., 2016).

Features are mainly used to reduce the dimensionality of the data without discarding the main information. Before the introduction of accurate classifier methods that can handle complex and high quantity of data, only specific features were computed in line with the classification problem (Xiao et al., 2005). Currently, hundreds of features are computed (Dalla Mura et al., 2010; Huang and Zhang, 2013), and then the best subset is selected (Gressin et al., 2013; Paget et al., 2015). In addition, robust classification methods, such as SVM or RF, have shown that their performances are likely to remain unchanged even by adding insignificant features.

When working with SITS acquired at high temporal resolution, the contribution of several input features is uncertain (Fernández-Delgado et al., 2014). Indeed, the spectral signatures of temporal profiles can be enough to characterize land cover categories. However, features can help to deal with large variability of the landscape when working on large areas. Therefore, it becomes interesting to determine the best feature subset of smaller size among all available features in order to achieve equal accuracies and reduce computational cost.

This work aims at assessing the robustness of classification methods to provide accurate land cover maps over large areas with HR-SITS. Specifically, it addresses the evaluation of RF performances on large areas by using different feature sets as input data through several studies. Firstly, the choice of RF classifier is discussed by comparing it with the well-known SVM. Secondly, the RF parameter sensitivity is analyzed. Then, the use of different sets of features as input data in the classification system is studied. Finally, the classifier stability is tested on a larger area covering around 20,000 km.^{2}.

This paper is organized as follows: Section 2 describes the data; Section 3 details classification scheme, and more precisely the input features used, and RF classifier; Section 4 is devoted to results and discussions; and finally Section 5 draws the conclusion.

2. Data

2.1. Study area

Two study areas are selected in the south of France having a temperate climate and a mean annual precipitation of 650 mm (Fig. 1). The slope in the whole scene varies greatly with the presence of the Pyrénées mountains in the south, the Massif Central mountains in the northeast, and lowlands in the remaining area. Main land cover categories are agricultural fields (principally winter crops), roads, urban areas, and forests (broad-leaved and conifers).

2.2. Satellite images

Landsat-8 and SPOT-4 (Take-5 experiment, Hagolle et al. (2015)) images are used as a simulation of the upcoming ESA's Sentinel-2 data. The combination of both sensors provides images with a temporal average gap of 13 days, approaching the temporal resolution of Sentinel-2 of five days. Concerning the spatial resolution, Sentinel-2 has variable band resolutions from 10 to 60 m, but the main bands have a spatial resolution of 10 or 20 m close from the one of Landsat-8 and SPOT-4. The major difference with Sentinel-2 time series concerns the spectral resolution due to the absence of red-edge bands for Landsat-8 and SPOT-4 sensors. The satellite characteristics are summarized in Table 1.

USGS (United States Geological Survey) and THEIA Land Data Centre pre-processed Landsat-8 and SPOT-4 images respectively: they ortho-rectified images and converted digital number values to top-of-atmosphere reflectances. Then, top-of-atmosphere reflectances are converted to top-of-canopy reflectances by using MACCS processing chain (Multi-sensor Atmospheric Correction and Cloud Screening, Hagolle et al. (2015)) for both satellite images. Images with more than 80% of cloudy data, are omitted in the final time series.

As satellites need several orbital cycles to cover a large area, HR-SITS are composed of multiple footprint images acquired at different dates. In order to work with a regular temporal sampling in the whole scene, images are temporally resampled (linear interpolation) as done in Inglada (2016a). Their work showed that although sample information is modified, classification accuracy does not change significantly (Inglada, 2016a).

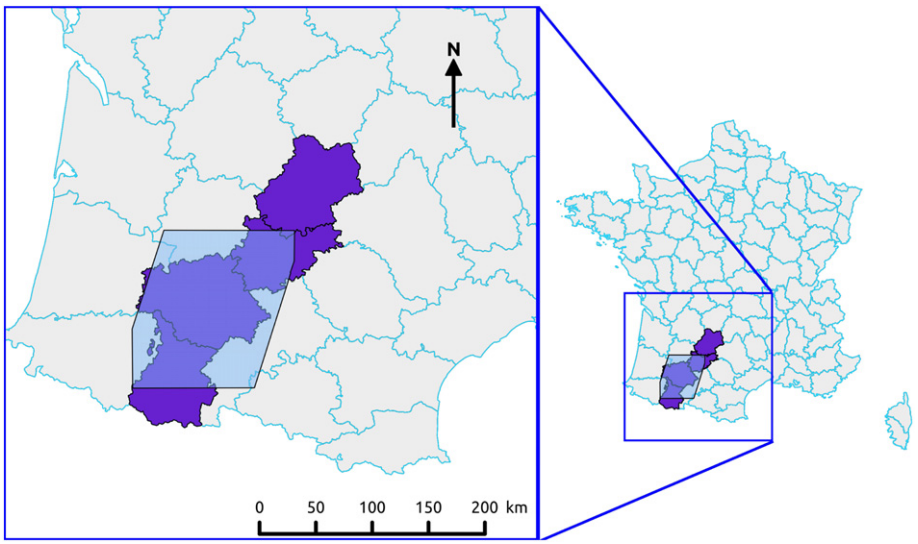

Fig. 1. Study areas. In blue, first study area. In purple, second study area. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Missing information, mainly due to the cloud coverage and the acquisition system, is detected by using Hagolle et al. (2010, 2015) method. Then, the missing values are filled in with a temporal linear interpolation (Inglada, 2016a; Inglada et al., 2015).

More precisely, the first data set covers 16,902 km^2 (in blue in Fig. 1). It is composed of 8 Landsat-8 and 15 SPOT-4 (Take-5 experiment) images. Landsat-8 images have been resampled at 20 m by using spatial cubic resampling method in order to take advantage of SPOT-4 spatial resolution, and to be closer to Sentinel-2 spatial resolution. Fig. 2 shows the temporal distribution of the images. The combination of both sensor images provides data that describe a complete vegetation cycle over time and that facilitates the recognition of land cover categories such as croplands.

The second data set covers 19,785 km^2 and crosses four French departments (in purple in Fig. 1). The scene cover requires eight 100×100 km^2 tiles of Landsat-8 images. Each tile contains from 9 to 21 acquired images depending on the satellite orbit and the cloud coverage. The images contained in each tile are resampled at a temporal resolution of 16-days (from the 19/04/2013 to 29/11/2013).

Table 1
Satellite characteristics.

Satellite	Sentinel-2		SPOT-4 Take-5		Landsat-8	
Temporal resolution	5 days		5 days		16 days	
Spectral (μm) and spatial (m) resolutions						
Coastal aerosol	0.43–0.45	60			0.43–0.45	30
Blue	0.46–0.52	10			0.45–0.51	30
Green	0.54–0.58	10	0.50–0.59	20	0.53–0.59	30
Red	0.65–0.68	10	0.61–0.68	20	0.64–0.67	30
Red-edge	0.70–0.71	20				
	0.73–0.74	20				
	0.77–0.79	20				
Near infra-red	0.78–0.90	10	0.78–0.89	20	0.85–0.88	30
	0.85–0.87	20				
	0.93–0.95	60				
Shortwave infra-red	1.57–1.66	20	1.58–1.75	20	1.57–1.65	30
	2.10–2.28	20			2.11–2.29	30
	1.37–1.39	60				

2.3. Reference data

The merging of three databases provides training and validation sets.

(1) The French Land Parcel Information System database (*Registre Parcellaire Graphique* in French) maps annually French crop fields. The position of the fields are determined at 1:15,000 scale. The field polygons can contain different land cover types which are declared by the farmers. Given the classification purpose, only fields with one land cover type are retained. The following land cover categories are selected from this database: wheat, corn, barley, rapeseed, sunflower, grassland, orchard, and vineyard.
(2) The French National Land Cover database, produced by the French mapping agency, maps the French territory with an update of three years. The geometric accuracy is about 1 m. Land cover types are determined by merging databases of the French mapping agency, and by using photo-interpretation when information is missing. From this database, the land cover categories used are impervious and pervious surfaces, bare soils, water, forests (broad-leaved, conifers, and mixed forests), and shrubs. Broad-leaved and conifers polygons are composed of the same tree species, and covering at least of 75% of trees. Mixed forest corresponds to a mixed of broad-leaved and conifers.
(3) The last database is composed of ground data collected during fieldwork campaigns. It represents about 1300 fields with homogeneous growing conditions and covering one land cover type. They have been visited three to four times during the main growing season of the year 2013. The location and the extent of the fields have been established by using a very high spatial resolution image. The following croplands have been listed: wheat, corn, barley, rapeseed, sunflower, soybean, and sorghum.

In order to avoid issues with samples belonging to polygon borders (which often mix different land cover categories), polygons from all the databases are eroded with a structuring element of 10 m (for urban surfaces) or 30 m (for other land cover categories).

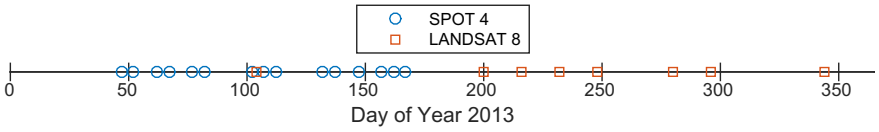

Fig. 2. Temporal distribution of Landsat-8 and SPOT-4 images for a first data set.

3. Classification scheme

3.1. Input features

Many features (also named variables) can be used as input data in a classification scheme. These features can be: (1) spectral features by combining spectral bands, (2) spatial features by using spatial information or by working at object-level instead of pixel-level, and (3) temporal features to characterize land cover categories which evolve over time such as croplands. In this paper, only spectral and temporal features will be analyzed.

Spectral features have been used successfully to improve the classification performances by better separating land cover categories with similar spectral signatures (Mróz et al., 2004; Silleos et al., 2006; Yeom et al., 2013). For instance, NDWI (McFeeters, 1996) for water detection, NDBI (Zha et al., 2003) for building detection, and the tasseled cap coefficients with Landsat images for the description of agricultural crops (Kauth and Thomas, 1976; Crist and Cicone, 1984; Huang et al., 2002; Baig et al., 2014). Yet, vegetation indexes, the most used indexes for land cover mapping, allow the depiction of vegetation stages.

For this study, twelve spectral features are selected from remote sensing literature for their performances on classification or their novelty. They are presented in Table 2. Due to the lower number of spectral bands of SPOT-4, only six of the twelve spatial features are computed for SPOT-4 images. These features are highlighted with asterisks in Table 2. As the NDVI (Normalized Difference Vegetation Index, Rouse et al. (1973)) is one of the most used vegetation indexes for land cover mapping, it is also analyzed separately from the other spectral features.

Temporal features, computed on SITS, can help the classifier to discriminate land cover categories which evolve in time. The extraction of temporal information is possible thanks to a high revisit time. In this work, temporal features correspond to either statistical values or phenological parameters.

Statistical values, extracted from the NDVI profile, have shown their ability to improve the characterization of some land cover categories (Valero et al., 2016). In this study, eight statistical values are computed from NDVI profile: mean, minimum, maximum, amplitude, Olympic mean (removing extreme values), median, and standard deviation.

Concerning the phenological parameters, Jönsson & Eklundh (2002, 2004,) developed the TIMESAT software, which extracts nine phenological parameters from NDVI profile, such as the beginning, the peak, or the length of the season. It requires at least three years of data with equally spaced date points in order (1) to estimate the number of seasons, and (2) to fit the NDVI profile to a model. It has been mainly used with low resolution image time series such as MODIS (Moderate Resolution Imaging Spectroradiometer) (Jia et al., 2014). Note that other curve fitting methods have also been proposed (Beck et al., 2006; Zhang et al., 2003).

This work relies on double logistic fitting as Inglada (2016b). Therefore, the NDVI profile is modeled by a double logistic function using six parameters (Eq. (1)): A the amplitude, B the minimum value, x_0 and x_2 the inflection points, and x_1 (or x_3) the rate of increase (or decrease) of the curve at the inflection point.

$$g(t) = A\left(\frac{1}{1+e^{\frac{x_0-t}{x_1}}} - \frac{1}{1+e^{\frac{x_2-t}{x_3}}}\right) + B \tag{1}$$

The algorithm performs a fit of NDVI profile by the double logistic function for the complete season. It uses a two-step estimation which is robust to the presence of several vegetation cycles in the season. A and B are estimated with the main NDVI cycle. Then, the four $x_i, 0 \leq i \leq 3$ remaining parameters are estimated by using the Levenberg-Marquardt iterative algorithm (a non-linear least-square algorithm) (Levenberg, 1994; Marquardt, 1963). Fig. 3 displays the fitting of an NDVI profile (red filled diamonds) by a double logistic function (blue curve). Both cyan lines represent the tangent at inflection points (x_0 and x_2).

Eventually, the four $t_i, 0 \leq i \leq 3$ parameters are computed as meaningful to describe the phenological cycle. These parameters are displayed on Fig. 3. Starting date t_0 (or ending data t_3) is the date for which the tangent at x_0 (or x_2) (inflection point) intercepts the minimum value. t_1 (or t_2) is the date for which the tangent at x_0 (or x_2) intercepts the maximum value. $t_2 - t_1$ defines the length of the plateau. Eqs. (2) to (5) denote how these parameters are computed.

$$t_0 = x_0 + \frac{B - g(x_0)}{g'(x_0)} \tag{2}$$

$$t_1 = \frac{A + B - (g(x_0) - x_0 g'(x_0))}{g'(x_0)} \tag{3}$$

$$t_2 = \frac{A + B - (g(x_2) - x_2 g'(x_2))}{g'(x_2)} \tag{4}$$

$$t_3 = x_2 + \frac{B - g(x_2)}{g'(x_2)} \tag{5}$$

A total of seventeen features are computed from the curve fitting: A, B, $A + B$, $x_i, 0 \leq i \leq 3$, $g'(x_0)$, $g'(x_2)$, $t_i, 0 \leq i \leq 3$, $t_2 - t_1$, $t_3 - t_0$, an approximation of the area under the curve $A(t_1 - t_0)/2.0 + A(t_3 - t_2)/2.0 + A(t_2 - t_1)$, and the error done by the least-square algorithm.

The spectral and temporal features are used to evaluate their influence on the classification performances. Spectral features are computed for each image of the time series. Temporal features are computed on NDVI profiles. These features are used in combination with the spectral bands of images at each date, *i.e.* four spectral bands for SPOT-4 and seven spectral bands for Landsat-8 times the number

Table 2
Formulas of spectral features.

Names	Formula	Commentaries
Normalized Difference Vegetation Index	$NDVI = \frac{NIR-R}{NIR+R}$*	Rouse et al. (1973)
Normalized Difference Water Index	$NDWI = \frac{Gr-NIR}{Gr+NIR}$*	McFeeters (1996)
Modified Normalized Difference Water Index	$MNDWI = \frac{Gr-SWIR}{Gr+SWIR}$*	Xu (2006) *MNDWI* with *SWIR*1, and *MNDWI*2 with *SWIR*2.
Normalized Difference Built-up Index	$NDBI = \frac{SWIR1-NIR}{SWIR1+NIR}$*	Zha et al. (2003)
Modified Normalized Difference Built-up Index	$MNDBI = \frac{SWIR2-NIR}{SWIR2+NIR}$	Shingare et al. (2014)
Index-based Built-up Index	$IBI = \frac{NDBI-\frac{NDVI+MNDWI}{2}}{NDBI+\frac{NDVI+MNDWI}{2}}$*	Xu (2008)
Brightness	$Brightness = 0.3561 \times Bl + 0.3972 \times Gr + 0.3904 \times Red + 0.6966 \times NIR + 0.2286 \times SWIR1 + 0.1596 \times SWIR2$	Tasseled cap transformation describes characteristics of agricultural crops and other scenes (Crist and Cicone, 1984; Kauth and Thomas, 1976) - only for Landsat images.
Greenness	$Greeness = -0.3344 \times Bl - 0.3544 \times Gr - 0.4556 \times Red + 0.6966 \times NIR - 0.0242 \times SWIR1 - 0.2630 \times SWIR2$	Although Baig et al. (2014) derived tasseled cap coefficients for Landsat-8 images, Landsat-7 tasseled cap coefficients are used (Huang et al., 2002).
Wetness	$Wetness = 0.2626 \times Bl + 0.2141 \times Gr + 0.0926 \times Red + 0.0656 \times NIR - 0.7629 \times SWIR1 - 0.5388 \times SWIR2$	
Brilliance	Spectral band norm*	*Brilliance* for common bands between SPOT-4 and Landsat-8, and *Brilliance*2 for Landsat-8 bands.

Ubl: ultra-blue. *Bl*: blue. *Gr*: green. *R*: red. *NIR*: nearest infra-red. *SWIR*: short-wave infra-red.
* Spatial features which can be computed for SPOT-4 images.

of images in the time series. In total, five feature sets are studied in the following:

1. spectral bands only (SB),
2. spectral bands and spectral features (SB-SF),
3. spectral bands and NDVI (SB-NDVI),
4. spectral bands and temporal features (SB-TF),
5. spectral bands, NDVI and temporal features (SB-NDVI-TF).

Table 3 gives the number of total features for both data sets described in Section 2.

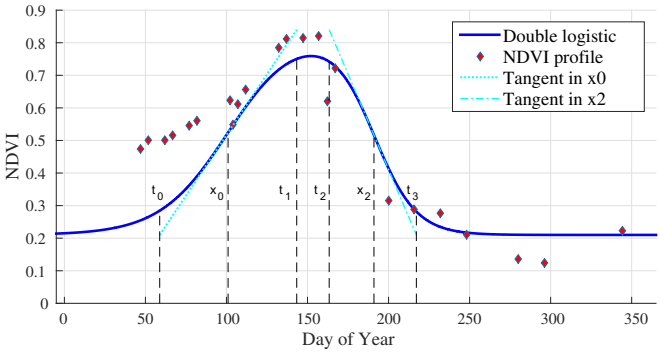

Fig. 3. Fitting of NDVI profile with a double logistic function. Starting date t_0 (or ending data t_3) is the date for which the tangent at x_0 (or x_2) (inflection point) intercepts the minimum value. t_1 (or t_2) is the date for which the tangent at x_0 (or x_2) intercepts the maximum value. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

3.2. Random Forests

Ensemble learning methods (bagging, boosting, *etc.*) have been widely used in remote sensing literature during the last decades. Random Forests (RF) of Breiman (2001) are a well-known ensemble learning method that combines K binary CART trees (Classification And Regression Trees) (Breiman et al., 1984). The input sample subsets used to build each tree are obtained from bootstrapping, an equiprobable selection technique with replacement (Breiman, 1996).

Each tree is built by performing an individual learning algorithm that splits the input variable set into subsets based on an attribute value test (for instance, the Gini coefficient). Contrary to classical Decision Trees (DT), the RF trees are built without pruning and by randomly selecting at each node a subset of input variables. Currently, this number of variables used to split a RF node (denoted by m) corresponds to the square root of the number of input variables (Liaw & Wiener, 2002). By limiting the number of variables used for a split, the computational complexity of the algorithm is reduced, and the correlation between the trees also decreases. The split process of RF decorrelates the different trees, thereby the classification result is less variable and more reliable.

Ideally, this process is repeated recursively on each derived subset until the node contains very similar samples, or when the splitting no longer adds value to the predictions. For implementation purposes, the tree building can be stopped when a maximum depth (*max_depth*) is reached, or when the number of samples at the node is below a *min_samples* threshold.

During the decision phase, each tree (individual learner) classifies the input feature data. Then, the RF output the class label that received the majority of votes.

RF are compared to another traditional classifier, the Support Vector Machines (SVM) (Vapnik, 1995, 1998) whose the principle is not detailed in this study.

Table 3
Number of total features for each data and feature sets. Data set 1 is composed of 15 SPOT-4 and 8 Landsat-8 images, whereas data set 2 contains 15 Landsat-8 images.

	SB	SB-SF	SB-NDVI	SB-TF	SB-NDVI-TF
Data set 1	116	302	139	141	164
Data set 2	105	285	120		

SB: spectral bands. SF: spectral features. TF: temporal features.

3.3. Sampling procedure

The sampling procedure is performed in three distinct steps that results in two independent sets of pixels: one for the training and one for the validation. The first step consists in splitting reference data into two disjoint training and validation sets of polygons. The split is made at a polygon level in order to ensure that there are no pixels from the same polygon in the training and the validation sets. The second step consists in randomly selecting the same number of training samples, *i.e.* pixel coordinates with their associated land cover category, per category from the training set of polygons. For under-represented land cover categories, the total available training samples are used. This random selection is independent from the polygons. Finally, all the pixels in the validation polygons are used for the validation of the produced land cover maps. The unit of assessment is therefore the pixel for the validation procedure.

In the first study area, training and validation samples are mainly selected from the fieldwork campaigns and the French National Land Cover database. Only vineyard and orchard are extracted from the French Land Parcel Information System database. It represents a total of 18 land cover categories. For the training set, the number of training samples is firstly fixed at 5000 samples per category when possible. However, some analyses on the influence of the number of training samples are also carried out by limiting the number of training samples to 750, 2000, and 10,000 samples per category. The numbers of training and validation samples used for the first study area are displayed in Table 4. Furthermore, the sampling procedure is repeated five times with different random draws from training and validation samples in order to statistically evaluate the results by computing confidence intervals.

In the second study area, reference data are composed of 12 classes, extracted from the French Land Parcel Information System and the French National Land Cover database. The training area is here spatially located within a circle of 15 km radius (red circle in Fig. 4). 15,000 training samples per category are randomly selected when possible in the red circle area. 35,000 validation samples per category are randomly selected outside the training area, but within the second study area remaining (in grey in Fig. 4). Nineteen sub-areas are defined for local validation (blue rectangles in Fig. 4). Each validation sub-area is a 20×15 km ^{2} rectangle, distant from 15 km of other validation sub-areas. Confusion matrices will be computed in each sub-area in order to appraise the influence of the training area location. The number of training (in red circle) and validation samples (total of samples in blue rectangles) for this second study area are also displayed in Table 4. Here, the sampling procedure is not repeated because almost all available training pixels in the red circle area are used. Thus, only broad-leaved, wheat, sunflower and grassland categories have more than 15,000 pixels in the red circle area.

3.4. Evaluation

The Overall Accuracy (OA) and the F-Score values are computed from the confusion matrices in order to evaluate the accuracy of the produced land cover maps (Congalton, 1991). The F-Score is a per category measure which corresponds to the harmonic mean of the user's accuracy and the producer's accuracy. It reaches its best value at 1 and worst at 0. User's and producer's accuracies will be also displayed in the result tables, but not commented.

As the first three studies aim at comparing different approaches (*e.g.* different classifier algorithms, input data sets or classifier parametrizations), the obtained results are averaged on five random trials from the training and validation samples. Thus, the results are not influenced by a specific split of training and validation samples. Furthermore, the confidence intervals computed from a T-distribution are displayed for the OA.

Concerning the classifier parameters, they have been optimized by a grid search. In the case of SVM, the chosen kernel – a Gaussian Radial Basis Function (SVM-RBF) – requires the optimization of two parameters: C the regularization parameter, and γ the width of the SVM-RBF kernel. The optimization is performed by a 10-fold cross-validation: $C = \{0.1, 0.5, 2.5, 12.5, 62.5, 312.5\}$, and $\gamma = \{1e^{-5}, 1.5e^{-4}, 2.25e^{-3}, 3.375e^{-2}, 5.5063e^{-1}\}$. The values of C (or γ) are obtained with a logarithmic grid range for five (or four) iterations with a minimum value of 0.1 (or 1.10^{-5}), and a logarithmic step of 5 (or 15).

Table 4
Number of training and validation pixels for the different studies.

	1st study area					2nd study area	
Number of pixels	T 750	T 2000	T 5000	T 10,000	V	T	V
Impervious	750	2000	5000	10,000	10,321	2968	13,280
Pervious	750	1045	1045	1045	1045	–	–
Bare soil	750	2000	2493	2493	2493	–	–
Water	750	2000	5000	7630	7630	899	16,383
Broad-leaved	750	2000	5000	10,000	202,829	15,000	14,851
Conifers	750	2000	5000	10,000	26,113	2179	22,801
Mixed	750	2000	5000	10,000	24,454	–	–
Shrub	750	2000	5000	10,000	26,519	906	24,709
Wheat	750	2000	5000	10,000	20,923	15,000	13,330
Rapeseed	750	2000	4350	4350	4350	8412	10,645
Corn	750	2000	5000	10,000	30,037	8065	12,762
Barley	750	1556	1556	1556	1556	10,720	12,577
Soy	750	2000	2268	2268	2268	–	–
Sorghum	257	257	257	257	257	–	–
Sunflower	750	2000	5000	10,000	10,730	15,000	12,840
Grassland	750	2000	5000	10,000	11,756	15,000	13,203
Orchard	750	2000	5000	10,000	18,120	–	–
Vineyard	750	2000	5000	10,000	52,188	639	2480
Total	13,007	32,858	71,969	129,599	453,589	94,788	157,831

T: training. V: validation.

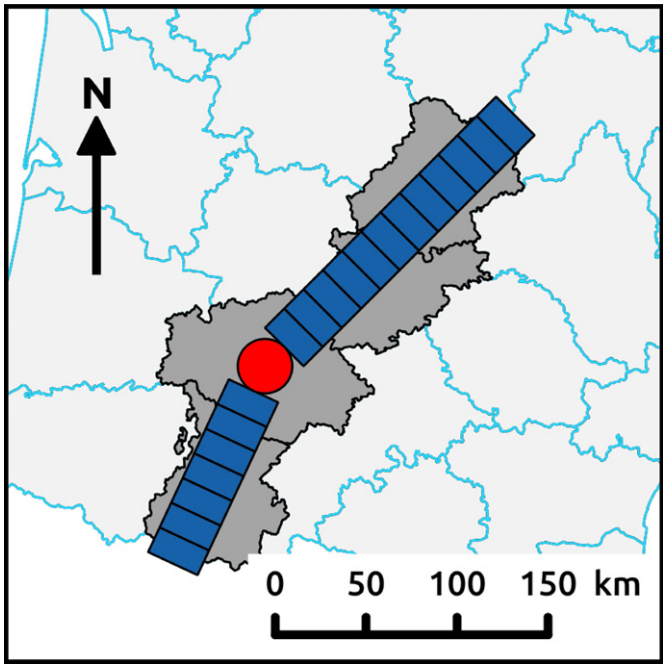

Fig. 4. Study area in grey with training area in red and validation areas in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The four RF parameters, described in Section 3.2, are optimized by selecting the combination of parameters achieving the best average OA on the five validation sets. The parameter ranges correspond to: $K = \{50, 100, 150, 200, 400\}$, $m = \{2, \sqrt{p}, p/3, p/2, p\}$ with p the total number of features, $max\_depth = \{10, 25, 50\}$, and $min\_samples = \{1, 10, 25, 50, 70\}$.

4. Results and discussion

4.1. Random Forests and Support Vector Machines comparison

The RF and SVM abilities to map land cover with HR-SITS are here compared. The evaluation is performed by using the first data set, presented in Section 2, composed of 18 land cover categories, shown in Table 4.

4.1.1. RF and SVM results

Tables 5 displays RF and SVM averaged F-Scores per land cover category, and OA with 95% confidence intervals for SB (the simplest case) and SB-SF (the case with the largest number of features).

A first analysis of OA values shows that RF outperform SVM for both input feature datasets, more exactly RF with SB-SF obtain the best results. However, the OA differences between both classifiers are not always observed by looking individual F-Score values. OA is a global measure, affected by the unbalanced number of validation samples. For instance, broad-leaved samples account for half of total validation samples. Therefore, the difference of 8% in broad-leaved F-Scores for SB-SF between RF and SVM gives rise to the large difference in OA.

Regarding the input feature sets, SVM have higher F-Score differences between SB and SB-SF than RF. SVM classifier is therefore more sensitive than RF to the input feature set.

It can be observed that variations per land cover category for "classifier - input features" combination can exceed 20%, because of the unbalanced problem that affects F-Score values. For instance, the unbalanced sample distribution between impervious and pervious surfaces causes high variations in pervious F-Scores for SB-SF between RF and SVM.

Regarding land cover categories with a number of training samples lower than 5000, Table 4, there is an in-between results in agreement with the literature on unbalanced training data problem: sometimes SVM perform better than RF (Lin & Chen, 2013), other time RF prevail (Khalilia et al., 2011). When the number of training samples is too low, the case of sorghum, results are dreadful for both classifiers (under 10%).

Besides the unbalanced training data set, the differences between both classifiers may also arise from the land cover thematics: SVM are better on buildings, grasslands, vineyard, and orchard; whereas RF are more accurate on forests, and globally on croplands. One can also notice that F-Score differences are higher when RF perform better than when SVM overcome.

Many studies have compared RF and SVM accuracies showing comparable performances between both classifiers using different remote sensing data (Waske & van der Linden, 2008; Duro et al., 2012; Hasan et al., 2012; Dalponte et al., 2013; Löw et al., 2013; Ghosh et al., 2014; Meyer et al., 2016).

Table 5 denotes a slight improvement with RF classifier probably due to RF ability to deal with the high-dimensionality of the feature space. Indeed, the RF algorithm splits each node with only the best feature among a random subset of features. This operation can be seen as an internal feature selection with no equivalent for the SVM.

4.1.2. Influence of the number of training samples

Table 5 displays RF and SVM accuracies achieved by the same number of training samples. However, it is well known that the number of training samples can affect the results, and its influence may differ for both classifiers. Table 6 (or 7) displays averaged F-Scores and OA obtained by using SVM (or RF) with a limited number of training samples per category of 750, 2000, and 5000 (or 2000, 5000, and 10,000). Table 4 displays the exact number of training and validation samples per category for each configuration. The different number of training samples, used to compare both classifiers, is mainly due to training times. For 5000 training samples per land cover category, the parameter optimization of SVM SB-SF takes almost 50 times longer than the parameter optimization of RF SB-SF. Furthermore, RF results show that its performances increase when the number of training samples increases. A test for 750 training samples per land cover category is therefore senseless for the RF. The input feature set used is SB-SF since it has previously obtained the best result.

Table 6 shows that SVM yield the best OA results with 2000 training samples per category. However, the analysis of F-Scores per category reveals that generally the best results are achieved with 5000 training samples per category. The difference of about 4% in broad-leaved F-Scores between both configurations may partially cause the difference in OA. Although results are lower for 750 training samples per category, the improvement by increasing the number of training samples is not so important.

According to Table 7, RF yield the best OA, and F-Score results with 10,000 training samples per category. However, as for SVM classifier, the accuracy gain is weak with increasing the sample size. Furthermore, using 2000 training samples can lead to better results for some land cover categories. For example, the best pervious F-Score is obtained with a limited number of training samples of 2000. This configuration reduces the gap between the number of pervious training samples (around 1000) and the one of other land cover categories, Table 4, leading to an increase of pervious F-Score. The same conclusion can be drawn for barley category which will be more clearly distinguished from wheat with a limited number of training samples of 2000 or 5000.

Table 5
Averaged F-Scores per land cover category and OA with 95% confidence intervals obtained on five random trials by using RF and SVM. The input feature sets are SB and SB-SF. Bold values correspond to the highest F-Scores and OA.

	RF SB			SVM SB			RF SB-SF			SVM SB-SF		
K,*m*, *max_depth*, *min_samples* | *C*, γ	400, 10, 50, 10			2,5, $2.25e^{-3}$			400,17, 50, 10			2,5, $1.5e^{-4}$		
	UA	PA	F	UA	PA	F	UA	PA	F	UA	PA	F
Impervious	87.4	76.7	90.4	89.8	88.9	89.2	86.6	93.7	89.9	92.0	93.5	**92.6**
Pervious	76.7	34.8	47.8	69.75	32.0	43.8	73.1	34.1	46.4	69.7	54.7	**61.2**
Bare soil	49.5	52.6	49.9	44.0	28.8	32.6	55.2	56.9	**55.3**	47.4	42.8	43.6
Water	98.5	97.1	97.7	99.8	77.5	87.2	98.8	99.6	**99.2**	99.8	86.3	92.4
Broad-leaved	93.3	82.1	87.2	92.8	75.8	83.3	93.9	83.2	**88.1**	90.7	72.2	80.2
Conifers	59.8	67.4	63.2	51.5	55.2	52.8	62.5	67.2	**64.6**	53.5	56.8	54.4
Mixed	21.2	31.6	24.9	19.2	31.1	23.6	24.3	35.8	**28.1**	13.1	26.8	17.5
Shrub	73.1	88.4	39.0	39.0	92.8	54.7	75.6	90.0	**82.0**	65.7	90.2	75.8
Wheat	91.0	92.8	91.9	90.1	83.6	86.7	91.6	92.9	**92.2**	90.3	92.5	91.4
Rapeseed	90.5	94.6	92.5	94.3	64.1	76.2	91.1	95.3	93.2	92.9	94.5	**93.7**
Corn	94.9	91.9	93.4	93.1	80.2	86.1	94.7	92.8	**93.8**	93.2	92.0	92.6
Barley	69.7	47.9	56.7	68.3	19.9	30.6	71.0	51.6	**59.7**	61.2	43.9	51.1
Soy	82.2	69.9	75.1	90.1	31.1	45.9	86.6	75.1	**80.1**	85.8	67.8	74.9
Sorghum	15.8	1.5	2.7	10.0	0.5	0.9	35.2	1.4	2.6	24.7	3.1	**5.3**
Sunflower	79.9	89.5	84.4	79.3	70.0	74.2	83.7	90.8	**87.1**	82.2	89.0	85.3
Grassland	68.8	83.7	75.4	74.2	73.7	73.9	69.2	84.6	76.0	75.9	83.8	**79.6**
Orchard	92.3	88.0	90.1	96.3	87.7	91.8	94.1	89.6	91.8	95.5	94.1	**94.8**
Vineyard	96.3	94.8	95.5	98.2	93.3	95.7	96.8	95.5	96.2	98.2	96.6	**97.4**
OA	82.1 ± 3.6			75.5 ± 2.4			**83.3 ± 3.9**			77.1 ± 5.2		

UA: user's accuracy. PA: producer's accuracy. F: F-Score. SB: spectral bands. SF: spectral features. *K*, the number of trees. *m*, the number of features randomly selected at each node. *max_depth*, the maximal depth of each tree. *min_samples*, the minimal number of samples per node. *C*, the regularization parameter. γ, the width of the SVM-RBF kernel.

4.1.3. Computational time

SVM and RF classifiers can lead to similar results with SB-SF, but another important comparison criterion is the computational time. SVM classifier is known for having a longer training time than RF (Pal, 2005). In order to compare training times of both classifiers, SB-SF input feature set is used with 5000 training samples per land cover category. In this context, Table 8 displays training times for both classifiers with SB-SF. A first case (**A**) gives the training times when the parameter configuration is known (the parameter configuration used here is the one displayed in Table 5). A second case (**B**) gives the required time to perform the optimization. For RF, training times are averaged for *m* variations. *K* is set to 100, *max_depth* to 50 and *min_samples* to 25. Section 4.3 will show that $K = 100$ gives similar results to $K = 400$, and that *max_depth* and *min_samples* influences are negligible.

Knowing the optimal parameter values, SVM is 12 times faster than RF algorithm. However, SVM parameter configuration is seldom straightforward and often requires an optimization phase, that blew

Table 6
Averaged F-Scores per land cover category and OA with 95% confidence intervals obtained on five random trials by using SVM with different numbers of training samples. The input feature set is SB-SF. Bold values correspond to the highest F-Scores and OA.

	SVM SB-SF 750			SVM SB-SF 2000			SVM SB-SF 5000		
C, γ	12.5, $1.5e^{-4}$			12.5, $1.5e^{-4}$			2.5, $1.5e^{-4}$		
	UA	PA	F	UA	PA	F	UA	PA	F
Impervious	92.8	87.2	89.8	92.6	91.2	91.8	92.0	93.5	**92.6**
Pervious	46.0	71.7	56.0	58.4	61.8	60.0	69.7	54.7	**61.2**
Bare soil	28.9	46.8	34.4	36.6	48.8	40.5	47.4	42.8	**43.6**
Water	99.8	82.3	90.0	99.8	84.5	91.3	99.8	86.3	**92.4**
Broad-leaved	93.2	74.1	82.4	93.3	77.0	**84.3**	90.7	72.2	80.2
Conifers	51.4	59.5	**54.8**	53.9	55.9	54.5	53.5	56.8	54.4
Mixed	18.6	35.1	24.2	19.8	36.6	**25.6**	13.1	26.8	17.5
Shrub	56.6	86.3	67.9	60.6	87.8	71.3	65.7	90.2	**75.8**
Wheat	90.6	89.8	90.2	90.1	91.4	90.7	90.3	92.5	**91.4**
Rapeseed	90.5	94.7	92.5	91.7	94.7	93.1	92.9	94.5	**93.7**
Corn	93.7	90.4	92.0	93.3	91.6	92.4	93.2	92.0	**92.6**
Barley	50.2	54.7	**52.2**	53.6	48.2	50.6	61.2	43.9	51.1
Soy	79.0	72.5	**75.1**	81.1	70.7	74.8	85.8	67.8	74.9
Sorghum	5.8	3.8	4.4	15.1	3.6	**5.6**	24.7	3.1	5.3
Sunflower	78.9	88.5	83.3	81.5	88.5	84.7	82.2	89.0	**85.3**
Grassland	72.4	79.5	75.7	75.0	81.5	78.0	75.9	83.8	**79.6**
Orchard	93.8	92.3	93.1	94.8	93.5	94.1	95.5	94.1	**94.8**
Vineyard	97.7	95.0	96.3	98.0	96.0	97.0	98.2	96.6	**97.4**
OA	77.7 ± 3.8			**79.4 ± 2.7**			77.1 ± 5.17		

UA: user's accuracy. PA: producer's accuracy. F: F-Score. SB: spectral bands. SF: spectral features. *C*, the regularization parameter. γ, the width of the SVM-RBF kernel.

Table 7
Averaged F-Scores per land cover category and OA with 95% confidence intervals obtained on five random trials by using RF with different numbers of training samples. The input feature set is SB-SF. Bold values correspond to the highest F-Scores and OA.

	RF SB-SF 2000			**RF SB-SF 5000**			**RF SB-SF 10,000**		
K, *m*, *max_depth*, *min_samples*	400, 2, 50, 10			400,17, 50, 10			400,17, 50, 25		
	UA	PA	F	UA	PA	F	UA	PA	F
Impervious	87.7	90.8	89.1	86.6	93.7	89.9	86.0	94.4	**89.9**
Pervious	55.3	46.6	**50.5**	73.1	34.1	46.4	81.5	23.6	36.3
Bare soil	40.2	63.2	48.3	55.2	56.9	**55.3**	64.0	49.0	54.6
Water	99.1	99.5	**99.3**	98.8	99.6	99.2	99.0	99.6	99.3
Broad-leaved	93.7	82.7	87.7	93.9	83.2	88.1	93.8	84.5	**88.8**
Conifers	63.1	71.2	**66.7**	62.5	67.2	64.6	62.0	66.6	64.1
Mixed	25.2	35.0	**28.6**	24.3	35.8	28.1	24.9	34.2	28.0
Shrub	71.9	86.3	78.3	75.6	90.0	82.0	75.7	91.2	**82.6**
Wheat	92.3	91.8	92.1	91.6	92.9	**92.2**	90.5	93.4	91.9
Rapeseed	90.2	94.4	92.2	91.1	95.3	93.2	93.0	94.0	**93.5**
Corn	95.1	91.9	93.5	94.7	92.8	**93.8**	94.3	93.1	93.7
Barley	58.5	58.6	58.5	71.0	51.6	**59.7**	77.1	43.8	55.6
Soy	82.0	73.9	77.3	86.6	75.1	**80.1**	88.9	69.7	77.7
Sorghum	15.0	1.4	2.6	35.2	1.4	**2.6**	12.5	0.4	0.8
Sunflower	80.6	89.2	84.7	83.7	90.8	**87.1**	83.5	90.7	86.9
Grassland	68.3	83.2	74.9	69.2	84.6	**76.0**	69.2	84.6	**76.0**
Orchard	91.8	87.7	89.7	94.1	89.6	91.8	94.4	89.7	**92.0**
Vineyard	96.1	94.7	95.4	96.8	95.5	96.2	96.9	95.8	**96.3**
OA	82.7 ± 3.8			83.3 ± 3.9			**83.8 ± 3.5**		

UA: user's accuracy. PA: producer's accuracy. F: F-Score. SB: spectral bands. SF: spectral features. *K*, the number of trees. *m*, the number of features randomly selected at each node. *max_depth*, the maximal depth of each tree. *min_samples*, the minimal number of samples per node.

up the training time from some minutes to several days. In contrast, RF are less sensitive to its parameter configuration resulting in a lower training time for optimization.

Pal (2005) and Meyer et al. (2016) reached the same conclusions about computational time for agricultural land cover mapping and rainfall retrievals respectively. As displayed in Table 8, both studies highlighted that time and experimentations required to RF parameter selection were quite small in comparison to SVM classifier.

4.2. Random Forest parameter sensitivity

The aim here is to evaluate the influence of RF parameters on classification performances by using the first data set described in Section 2. The five feature sets, presented in Section 3.1, are studied as input data in the classification scheme. As mentioned in Section 3.2, the implementation of RF requires the tuning of four parameters: (1) *K*, the number of trees; (2) *m*, the number of features randomly selected at each node; (3) *max_depth*, the maximal depth of each tree; and (4) *min_samples*, the minimal number of samples per node.

Table 9 displays the averaged OA values obtained for each input feature set for different *K* and *m* configurations on the five validation sets. *max_depth* and *min_samples* are set to 25. Specifically, the studied values are $K = \{50, 100, 150, 200, 400\}$, and $m = \{2, \sqrt{p}, p/3, p/2, p\}$ with p the total number of features.

The results allow to draw a general conclusion about the input feature sets: for all parameter settings, SB-SF input feature set performs better than the other input sets.

Regarding the *K* parameter, it can be observed that accuracy rises slightly while the number of trees increases. Therefore, as the computational cost increases linearly with the number of trees, *K* can be set to 100 without a major loss in accuracy. These outcomes about *K* are consistent with other studies (Rodriguez-Galiano et al., 2012).

About the *m* parameter, Rodriguez-Galiano et al. (2012) worked with small values of *m* ($m = 1$) to yield the best results. The use of small values of *m* reduces the correlation among the individual trees resulting in better generalization. However, the default value recommended by the literature is $m = \sqrt{p}$ (Liaw & Wiener, 2002). In this study, Table 9 shows that the best results are obtained for this recommended *m* value (second column). However, the differences in OA are weak when comparing the results of $m = \sqrt{p}$ with $m = 2$ and $m = p/3$ (the first three columns) for a set number of trees. These results are also consistent with Cutler et al. (2007) and Boulesteix et al. (2012): they asserted that it is preferable to tune *m* for each case even if accuracy gain is weak.

Furthermore, Tatsumi et al. (2015) reported that while the number of training samples increases, the *m* parameter is less sensitive, and the accuracy improvement is mainly due to the increasing of good training samples. For this study, the weak gain in accuracy by tuning the *m* parameter may therefore be due to the huge number of training samples used (5000 per category).

In addition, it can be observed that the differences in OA between $m = \sqrt{p}$ (second column) and $m = p$ (last column) can reach more than 2% for all the feature sets for a set number of trees *K*. Therefore, contrary to Oliveira et al. (2012), a too high value of *m* causes a decline in OA. The case $m = p$ implies that all the features are used at each node to determine the best split which likely causes overfitting.

Table 8
Training times (with standard deviation) in seconds for SVM and RF classifiers. Row **A** shows training times when only the optimized parameter configuration is used. Row **B** shows training times when the optimization is performed.

	SVM SB-SF	RF SB-SF
A	265 ± 9 s	3,246 ± 375 s
B	209,856 ± 51,497 s (≃ 2 days 9 h)	4,207 ± 80 s (≃ 1 h)

SB: spectral bands. SF: spectral features.

Table 9
Averaged OA values obtained for five random trials by using different RF parameter configurations. m value is such as $m = 2$, $m = \sqrt{p}$, $m = p/3$, $m = p/2$, and $m = p$ with p the total number of features. Bold values correspond to the highest OA.

m\K	50	100	150	200	400
SB $p = 116$					
2	80.9	81.3	81.5	81.5	81.6
10	81.3	81.7	81.7	81.9	**82.0**
38	80.9	81.2	81.4	81.5	81.5
58	80.6	80.9	81.1	81.1	81.3
116	79.2	79.5	79.6	79.6	79.8
SB-SF $p = 302$					
2	82.2	82.6	82.8	82.7	82.9
17	82.7	82.3	83.0	83.1	**83.2**
100	82.0	82.9	82.6	82.5	82.7
151	81.8	82.2	82.2	82.3	82.4
302	80.4	80.8	80.9	80.9	81.0
SB-NDVI $p = 139$					
2	81.0	81.4	81.5	81.6	81.7
11	81.4	81.8	81.9	81.9	**82.0**
46	81.1	81.5	81.5	81.6	81.6
69	80.8	81.3	81.2	81.3	81.5
139	79.5	79.8	79.8	80.0	80.0
SB-TF $p = 141$					
2	80.6	81.1	81.3	81.4	81.4
11	81.3	81.6	81.7	81.8	**81.8**
47	80.8	81.4	81.4	81.5	81.6
70	80.6	81.2	81.1	81.3	81.3
141	79.6	79.9	80.0	80.1	80.1
SB-TF-NDVI $p = 164$					
2	80.9	81.2	81.3	81.3	81.4
12	81.3	81.7	81.8	81.9	**82.0**
54	81.1	81.3	81.5	81.5	81.6
82	80.8	81.1	81.3	81.3	81.3
164	79.4	79.7	79.7	79.8	79.9

SB: spectral bands. SF: spectral features. NDVI: Normalized Difference Vegetation Index. TF: temporal features. K, the number of trees. m, the number of features randomly selected at each node.

The impact of *max_depth* and *min_samples* parameters on classification is also studied. Both parameters have been less examined in the literature. From the previous results, m is set to $\sqrt{p}$, and K to 100. Table 10 displays averaged OA for SB (the simplest case) and SB-SF (the set achieving the best score) with different parameter configurations: $min\_samples = \{1, 10, 25, 50, 70\}$, and $max\_depth = \{10, 25, 50\}$.

Table 10 shows that an early stop of the tree building ($max\_depth = 10$) gives meaningless splits of samples, and so leads to poorer results. In addition, the accuracy differences between $max\_depth = 25$ or $max\_depth = 50$ are low. Therefore, *max_depth* will be set to 25.

Regarding the value of *min_samples*, the best results are obtained by either $min\_samples = 10$ or $min\_samples = 25$. Generally, the value of *min_samples* has a low impact on OA values.

Tables 9 and 10 also show that OA varies according to the different input data sets. The OA variations between the best parameter configuration and the other configurations are bigger for SB than for SB-NDVI or SB-SF. Therefore, increasing the number of features makes RF classifier less sensitive to its parameter configuration.

4.3. Feature set comparison

The purpose here is to evaluate the impact of different input feature sets on classification accuracies. The first data set, presented in Section 2, is used to build the five input feature sets. For each feature set, the RF parameters are optimized, as described in Section 3.4, in order to compare the best possible performances. Table 11 displays

Table 10
Averaged OA values obtained for five random trials by using different RF parameter configurations. Bold values correspond to the highest OA.

max_depth \ min_samples	1	10	25	50	70
SB $p = 116$					
10	78.1	78.3	78.1	78.1	78.0
25	81.3	**81.8**	81.7	81.3	80.8
50	81.2	**81.8**	81.6	81.2	80.9
SB-SF $p = 302$					
10	80.0	79.8	80.0	80.1	79.9
25	82.1	82.3	82.3	82.2	81.8
50	82.1	82.4	**82.5**	82.2	81.9

SB: spectral bands. SF: spectral features. *max_depth*, the maximal depth of each tree. *min_samples*, the minimal number of samples per node.

averaged OA with 95% confidence intervals, and averaged F-Scores per land cover category.

Comparing all the global results, SB-SF input feature set obtains the best scores. However, the resulting OA is only higher by about 1%. In addition, confidence intervals obtained by the four other feature sets overlap SB-SF confidence interval. The analysis per land cover category gives a similar conclusion. Adding spectral features to spectral bands improves F-Score values for almost all land cover categories. However, these F-Score differences between SB-SF and the other input feature sets are often small. Therefore, the results of the simplest cases SB or SB-NDVI are close to those of SB-SF. The lack of significant gain by adding some spectral features is probably due to the information redundancy between the different features.

Comparing SB-NDVI to SB-TF and SB-NDVI-TF, the results are extremely similar for all land cover categories. The proposed temporal features, computed from NDVI profiles, fail to bring additional relevant information to NDVI. Indeed, the related temporal information is already contained in the NDVI profile. As temporal features increase the global computational time without improving the classification accuracy, they will not be further analyzed.

The F-Score analysis also reveals that pervious, bare soil, mixed forest, barley, and sorghum categories obtain F-Scores under 60%. As denoted by Table 5, the number of training samples for these categories, except mixed forests, are below 5000. As bootstrap samples (Breiman, 1996) are selected from the training set to build each tree of the random forest, categories with less training samples will be disadvantaged. Mixed forests – which correspond to the presence of conifers and broad-leaved – obtain F-Scores under 30% for all the feature sets. At a 20 m spatial resolution, mixed forests represent a mixing either in pixels or at macroscopic scale. This category is therefore harder to classify.

The results suggest that SB-SF input feature set gives more information, but the overall classification results are slightly improved with regard to the other input feature sets. The feature contribution may differ over large areas where landscape changes. In this particular context, the variability of the data could be better addressed by adding some spectral features.

4.4. Random Forest stability over large areas

The goal here is to assess the RF accuracy and the contribution of the different input feature sets over large areas. More precisely, the classifier performances are evaluated when validation areas are moving away from the training area. For this purpose, the second data set, described in Section 2, is used. Besides, the circular training area and the nineteen validation sub-areas of Fig. 4 are considered. The evaluation is performed by computing a confusion matrix in each validation sub-area.

To confirm earlier results, the two feature sets that previously obtained the best results, and the simplest case are tested as input

Table 11
Averaged F-Scores per land cover category and OA with 95% confidence intervals obtained on five random trials by using optimized RF with different feature sets. Bold values correspond to the highest F-Scores and OA.

	SB			SB-SF			SB-NDVI			SB-TF			SB-NDVI-TF		
K, *m*, *max_depth*, *min_samples*	400, 10 50, 10			400,17, 50, 10,			400,11, 50, 10			400,11, 25, 10			400,12, 50, 10,		
	UA	PA	F	UA	PA	F	UA	PA	F	UA	PA	F	UA	PA	F
Impervious	87.4	93.7	90.4	86.6	93.7	89.9	87.0	93.5	90.1	88.0	93.6	**90.7**	87.5	93.4	90.3
Pervious	76.7	34.8	**47.8**	73.1	34.1	46.4	73.6	32.9	45.3	76.4	33.8	46.8	72.7	32.3	44.6
Bare soil	49.5	52.6	49.9	55.2	56.9	**55.3**	50.4	55.3	51.6	51.3	55.7	52.5	51.5	57.4	53.3
Water	98.5	97.1	97.7	98.8	99.6	99.2	99.2	99.4	99.3	99.3	99.3	99.3	99.4	99.4	**99.4**
Broad-leaved	93.3	82.1	87.2	93.9	83.2	**88.1**	93.6	81.9	87.2	93.3	81.9	87.1	93.5	81.7	87.1
Conifers	59.8	67.4	63.2	62.5	67.2	**64.6**	60.3	67.7	63.7	60.5	68.0	63.9	60.5	68.3	64.0
Mixed	21.2	31.6	24.9	24.3	35.8	**28.1**	22.4	33.5	26.1	21.6	32.0	25.1	22.1	33.0	25.7
Shrub	73.1	88.4	79.9	75.6	90.0	**82.0**	73.6	88.5	80.2	73.4	88.5	80.1	73.8	88.3	80.2
Wheat	91.0	92.8	91.9	91.6	92.9	**92.2**	91.1	92.9	92.0	90.9	92.5	91.7	91.1	92.6	91.9
Rapeseed	90.5	94.6	92.5	91.1	95.3	**93.1**	90.9	94.4	92.6	90.0	94.0	92.0	90.7	94.2	92.4
Corn	94.9	91.9	93.4	94.7	92.8	**93.8**	94.7	92.1	93.4	94.6	92.1	93.3	94.5	92.1	93.3
Barley	69.7	47.9	56.7	71.0	51.6	**59.7**	68.5	50.3	57.8	69.3	48.5	56.9	66.8	49.5	56.7
Soy	82.2	69.9	75.1	86.6	75.1	**80.1**	82.7	71.0	76.0	83.2	70.3	75.7	82.6	70.6	75.6
Sorghum	15.8	1.5	2.7	35.2	1.4	2.6	16.2	1.6	**3.0**	15.0	1.4	2.6	15.6	1.6	**3.0**
Sunflower	79.9	89.5	84.4	83.7	90.8	**87.1**	80.3	89.9	84.8	79.3	88.7	83.7	80.1	89.4	84.4
Grassland	68.8	83.7	75.4	69.2	84.6	**76.0**	68.5	83.4	75.1	67.8	83.7	74.8	67.8	83.4	74.7
Orchard	92.3	88.0	90.1	94.1	89.6	**91.8**	92.2	88.2	90.1	91.6	87.6	89.6	91.7	88.1	89.9
Vineyard	96.3	94.8	95.5	96.8	95.5	**96.2**	96.4	94.7	95.6	96.2	94.5	95.4	96.4	94.6	95.5
OA	82.1 ± 3.6			**83.3 ± 3.9**			82.2 ± 4.0			82.0 ± 4.0			82.1 ± 4.2		

UA: user's accuracy. PA: producer's accuracy. F: F-Score. SB: spectral bands. SF: spectral features. NDVI: Normalized Difference Vegetation Index. TF: temporal features. *K*, the number of trees. *m*, the number of features randomly selected at each node. *max_depth*, the maximal depth of each tree. *min_samples*, the minimal number of samples per node.

data for the classifier. This study aims at investigating whether SB-SF or SB-NDVI are more robust and stable than only SB for classification over large areas. RF parameters are set as follows: $K = 100$, $m = \sqrt{p}$, $max\_depth = 25$, and $min\_samples = 25$.

Fig. 5 represents validation sample distributions for each land cover category contained in all the sub-areas. The horizontal axis represents the distance (in kilometers) between the evaluation area and the training area. Positive (or negative) distances represent sub-areas located at the northeast (or southwest) of the training area. The vertical dashed red line localizes training area position.

North-south gradient introduces landscape variations: from −67.5 to 97.5 km cropland categories are mainly represented, whereas from 112.5 to 187.5 km forest categories hold the majority. In the southeast of the area, −112.5 to −82.5 km, (or northeast at 187.5 km) land cover distribution strongly changes due to Pyrénées mountains (or Massif Central mountains) with mainly forests, shrubs, and grasslands. Note that the number of validation samples is very low at −112.5 km.

Fig. 6 displays the OA values computed in all the sub-areas for the three feature sets (SB, SB-SF, and SB-NDVI).

Fig. 6 shows that the behaviours of the input data sets are quite similar. However, for the areas close to the training area, adding spectral features to spectral bands, most of the time, slightly increases the OA values. For the areas far from the training area, SB or

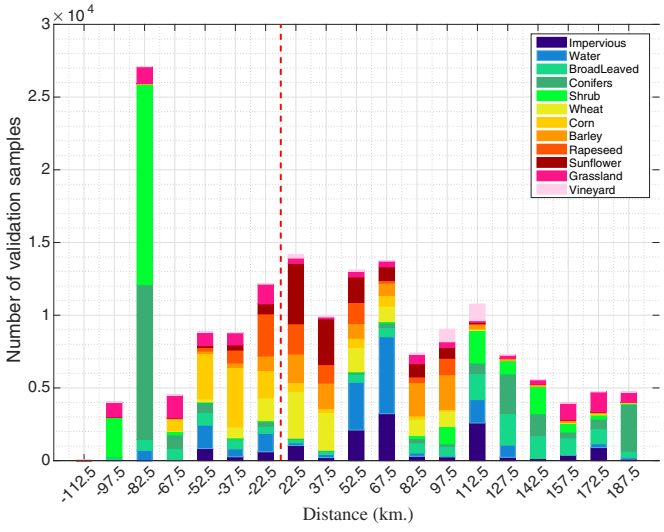

Fig. 5. Distribution of validation samples for each land cover category with respect to the training area distance. The vertical dashed red line shows training area location. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

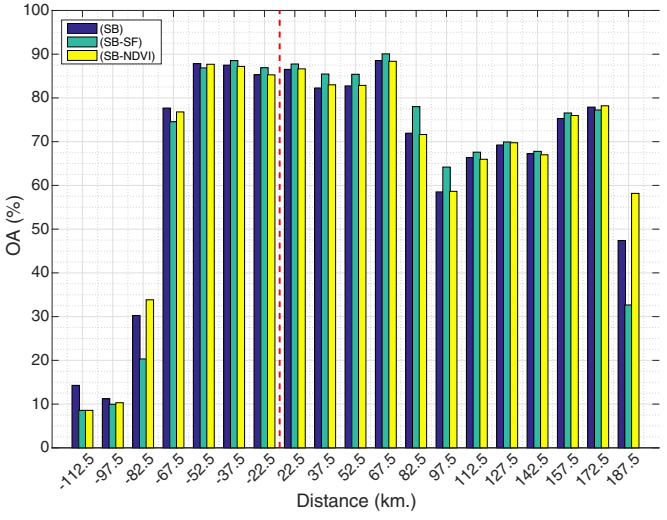

Fig. 6. Values obtained for OA for three feature sets with respect to the training area distance. The vertical dashed red line shows training area location. (sb: spectral bands, SF: spectral features). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

SB-NDVI obtain globally better OA values than SB-SF, but OA values are often low, under 0.5. Therefore, the addition of spectral features does not assure a classification accuracy improvement, even far from the training area.

More precisely, OA values remain constant from −52.5 to 67.5 km, where landscape is mainly composed of croplands and urban surfaces. These distances represent areas close to the training area. Hence, the RF classifier can achieve good performances as long as the landscape remains similar to that of the training area. But, results plunge dramatically, OA below 0.4, under −82.5 km due to the presence of the Pyrénées mountains. In this area, landscape is significantly different from that of the training area. Between 67.5 and 97.5 km, OA values fall of around 0.25: landscape varies with a greater presence of shrubs and vineyards (whose soils differ from those of the training area). OA values increase slightly from 112.5 to 172.5 km corresponding to areas with few croplands.

To conclude, the obtained results confirm that adding spectral features does not result in a huge accuracy gain, but it increases computational time. In addition, the results may give insights to implement a sampling strategy. Indeed, classification performances appeared to remain constant on a particular landscape. As a consequence, the stratification of large input datasets, *i.e.* the split of the study area into smaller areas according to landscape types, could be used to improve the land cover map quality.

5. Conclusion

Some of the main challenges of land cover mapping over large areas using HR-SITS have been discussed. In this context, the potential of RF has been shown.

RF have yielded comparable results to the traditional SVM method with a better trade-off between the classification performances and the computational times. They have also shown fewer variations in accuracy when input feature sets changed. RF are therefore a suitable tool to handle the amount of data provided by the HR-SITS. The results of SVM and RF also show some complementarity mainly for the categories with low accuracies. The merging of both classifiers may lead to more accurate results for these categories than a single classifier.

Spectral and temporal features have been also proposed in order to characterize scene dynamics. Although they improve the performances of classification, the gain in accuracy is weak compared with the increase in the computational cost. More precisely, it was expected that the proposed temporal features would add some relevant information to better discriminate the different land cover categories. However, RF already manage temporal information by exploiting the spectral signatures of HR-SITS, and therefore reduce confusion among land cover categories which evolve over time such as croplands. Furthermore, it has been shown that the setting of RF parameters causes little influence on the classification accuracy. To conclude, the use of only spectral bands is a good trade-off between accuracy and computational time.

The proposed classification scheme has also been tested in a larger area. More precisely, the impact on classification accuracy of a spatially localized training area has been investigated through different input feature sets. The study reveals that spectral features do not help the classifier even far from the training area. Furthermore, the classification performances decrease when the landscapes are different from those of the training area. Consequently, it may be interesting to stratify training samples according to the types of landscapes.

Future works will be carried out with Venμs and Sentinel-2 images (10 m). They will especially analyze the contribution of spatial features.

Acknowledgments

This work is funded by the French spatial agency (CNES) and the French mapping agency (IGN), and it has been carried out at CESBIO laboratory – supported by the University Paul Sabatier, CNRS (*Centre National de la Recherche Scientifique*), CNES, IRD (*Institut de Recherche pour le Développement*) – and MATIS laboratory supported by IGN.

The SPOT-4 imagery was obtained under the SPOT-4/Take-5 program. The SPOT-4/Take-5 imagery is copyrighted to CNES under "CNES 2013, all rights reserved. Commercial use of the product prohibited". Observation data were collected at the Regional spatial Observatory (OSR). OSR facilities and staff are funded and supported by the Observatory Midi-Pyrenean, the University Paul Sabatier, CNRS, CNES, IRD. The authors would like to thank all of the people involved in collecting field data for validation.

References

Alcantara, C., Kuemmerle, T., Prishchepov, A.V., Radeloff, V.C., 2012. Mapping abandoned agriculture with multi-temporal MODIS satellite data. Remote Sens. Environ. 124, 334–347.

Atkinson, P.M., Tatnall, A., 1997. Introduction neural networks in remote sensing. Int. J. remote sens. 18, 699–709.

Baig, M.H.A., Zang, L., Shuai, T., Tong, Q., 2014. Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. Remote Sens. Lett. 5, 423–431.

Beck, P.S., Atzberger, C., Høgda, K.A., Johansen, B., Skidmore, A.K., 2006. Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. Remote Sens. Environ. 100, 321–334.

Belgiu, M., Drăguţ, L., 2016. Random Forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.

Boulesteix, A.L., Janitza, S., Kruppa, J., König, I.R., 2012. Overview of Random Forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Min. Knowl. Disc. 2, 493–507.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R., 1984. Classification and Regression Trees. Taylor & Francis.

Carrão, H., Gonçalves, P., Caetano, M., 2008. Contribution of multispectral and multitemporal information from MODIS images to land cover classification. Remote Sens. of Environ. 112, 986–997.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., et al. 2015. Global land cover mapping at 30 m resolution: a POK-based operational approach. ISPRS J. Photogrammetry Remote Sens. 103, 7–27.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. of Environ. 37, 35–46.

Crist, E.P., Cicone, R.C., 1984. A physically-based transformation of Thematic Mapper data–the TM Tasseled Cap. IEEE Trans. Geosci. Remote Sens. 256–263.

Cutler, D.R., Edwards, T.C., Jr, Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.

Dalla Mura, M., Benediktsson, J., Waske, B., Bruzzone, L., 2010. Morphological attribute profiles for the analysis of very high resolution images. IEEE Trans. Geosci. Remote Sens. 48, 3747–3762.

Dalponte, M., Orka, H.O., Gobakken, T., Gianelle, D., Gianelle, D., 2013. Tree species classification in boreal forests with hyperspectral data. IEEE Trans. Geosci. Remote Sens. 51, 2632–2645.

Drusch, M., Bello, U.D., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for {GMES} operational services. Remote Sens. Environ. 120, 25–36.

Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 118, 259–272.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15, 3133–3181.

Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sens. Environ. 61, 399–409.

Ghosh, A., Fassnacht, F.E., Joshi, P., Koch, B., 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensor across three spatial scales. Int. J. Appl. Earth Obs. Geoinf. 26, 49–63.

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. Pattern Recogn. Lett. 27, 294–300.

Gong, P., Chen, Z., Tang, H., Zhang, F., 2006. Land cover classification based on multi-temporal MODIS NDVI and LST in Northeastern China. IEEE International Conference on Geoscience and Remote Sensing Symposium 2006 (IGARSS). pp. 1149–1152.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., et al. 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. Int. J. Remote Sens. 34, 2607–2654.
Gressin, A., Mallet, C., Vincent, N., Paparoditis, N., 2013. Updating land cover databases using a single very high resolution satellite image. The ISPRS Workshop on Image Seq. Anal. 2, 13–18.
Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images. Remote Sens. Environ. 114, 1747–1755.
Hagolle, O., Sylvander, S., Huc, M., Claverie, M., Clesse, D., Dechoz, C., Lonjou, V., Poulain, V., 2015. SPOT-4 (Take 5): simulation of Sentinel-2 time series on 45 large sites. Remote Sens. 7, 12242–12264.
Hansen, M., Dubayah, R., DeFries, R., 1996. Classification trees: an alternative to traditional land cover classifiers. Int. J. Remote Sens. 17, 1075–1081.
Haralick, R., 1979. Statistical and structural approaches to texture. Proceedings of the IEEE 67, 786–804.
Hasan, R.C., Ierodiaconou, D., Monk, J., 2012. Evaluation of four supervised learning methods for benthic habitat mapping using backscatter from multi-beam sonar. Remote Sens. 4, 3427–3443.
Huang, C., Wylie, B., Yang, L., Homer, C., Zylstra, G., 2002. Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. Int. J. Remote Sens. 23, 1741–1748.
Huang, X., Zhang, L., 2013. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. IEEE Trans. Geosci. Remote Sens. 51, 257–272.
Inglada, J., 2016a. OTB Gapfilling, A Temporal Gapfilling for Image Time Series Library. http://dx.doi.org/10.5281/zenodo.45572.
Inglada, J., 2016b. PhenOTB, Phenological Analysis for Image Time Series. http://dx.doi.org/10.5281/zenodo.45573.
Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., et al. 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. Remote Sensing 7, 12356–12379.
Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B., Wang, X., 2014. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. Remote Sens. 6, 11518–11532.
Jönsson,P., Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans. Geosci. Remote Sens. 40, 1824–1832.
Jönsson,P., Eklundh, L., 2004. Seasonality extraction by function fitting to time-series of satellite sensor data. Comput. Geosci. 30, 833–845.
Kauth, R.J., Thomas, G.S., 1976. The tasselled cap - a graphic description of the spectral-temporal development of agricultural crops as seen by landsat. LARS Symposia. pp. 159.
Khalilia, M., Chakraborty, S., Popescu, M., 2011. Predicting disease risks from highly imbalanced data using Random Forest. BMC Med. Inform. Decis. Mak. 11, 51.
Khatami, R., Mountrakis, G., Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. Remote Sens. Environ. 177, 89–100.
Levenberg, K., 1994. A method for the solution of certain nonlinear problems in least squares. Q. Appl. Math.
Liaw, A., Wiener, M., 2002. Classification and regression by Random Forest. R News 2, 18–22.
Lin, W.J., Chen, J.J., 2013. Class-imbalanced classifiers for high-dimensional data. Brief. Bioinform. 14, 13–26.
Löw, F., Dech, S., Conrad, C., 2013. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. ISPRS J. Photogramm. Remote Sens. 85, 102–119.
Lu, D., Mausel, P., Batistella, M., Moran, E., 2004. Comparison of land-cover classification methods in the Brazilian Amazon Basin. Photogramm. Eng. Remote Sens. 70, 723–731.
Lv, Z., Zhang, P., Benediktsson, J., Shi, W., 2014. Morphological profiles based on differently shaped structuring elements for classification of images with very high spatial resolution. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 4644–4652.
Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11, 431–441.
Masse, A., Ducrot, D., Marthon, P., 2011. Tools for multitemporal analysis and classification of multisource satellite imagery. 6th International Workshop on the Analysis of Multi-temporal Remote Sensing Images (Multi-Temp 2011). pp. 209–212.
McFeeters, S., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.
Meyer, H., Kühnlein, M., Appelhans, T., Nauss, T., 2016. Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. Atmos. Res. 169, 424–433.
others, Mróz, M., Sobieraj, A., 2004. Comparison of several vegetation indices calculated on the basis of a seasonal SPOT XS time series, and their suitability for land cover and agricultural crop identification. Tech. Sci. 7, 39–66.
Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., Pereira, J.M., 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and Random Forest. For. Ecol. Manag. 275, 117–129.
Paget, M., Gressin, A., Mallet, C., 2015. Multi-temporal optical vhr image fusion for land-cover mapping. IEEE International Geoscience and Remote Sensing Symposium 2015 (IGARSS). pp. 1913–1916.
Pal, M., 2005. Random Forest classifier for remote sensing classification. Int. J. Remote Sens. 26, 217–222.
Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J., 2012. An assessment of the effectiveness of a Random Forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104.
Rogan, J., Franklin, J., Roberts, D.A., 2002. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. Remote Sens. Environ. 80, 143–156.
Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symp., Washington, NASA 309–317.
Shingare, P.P., Hemane, P.M., Dandekar, D.S., 2014. Fusion classification of multispectral and panchromatic image using improved decision tree algorithm. International Conference on Signal Propagation and Computer Technology (ICSPCT) 2014. pp. 598–603.
Silleos, N.G., Alexandridis, T.K., Gitas, I.Z., Perakis, K., 2006. Vegetation indices: advances made in biomass estimation and vegetation monitoring in the last 30 years. Geocarto Int. 21, 21–28.
Szuster, B.W., Chen, Q., Borger, M., 2011. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. Applied Geography 31, 525–532.
Tatsumi, K., Yamashiki, Y., Torres, M.A.C., Taipe, C.L.R., 2015. Crop classification of upland fields using random forest of time-series landsat 7 etm+ data. Comput. Electron. Agric. 115, 171–179.
Trias-Sanz, R., 2006. Texture orientation and period estimator for discriminating between forests, orchards, vineyards, and tilled fields. IEEE Trans. Geosci. Remote Sens. 44, 2755–2760.
Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., Koetz, B., 2016. Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions. Remote Sens. 8, 55.
Vapnik, V.N., 1995. The nature of statistical learning theory. Springer Science & Business Media.
Vapnik, V.N., 1998. Statistical learning theory. Wiley New York.
Wang, J., Zhao, Y., Li, C., Yu, L., Liu, D., Gong, P., 2015. Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250 m resolution. ISPRS J. Photogramm. Remote Sens. 103, 38–47.
Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the US Central Great Plains. Remote Sens. Environ. 112, 1096–1116.
Waske, B., van der Linden, S., 2008. Classifying multilevel imagery from SAR and optical sensors by decision fusion. IEEE Trans. Geosci. Remote Sens. 46, 1457–1466.
Xiao, X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., Salas, W., Moore, B., 2005. Mapping paddy rice agriculture in southern China using multi-temporal modis images. Remote Sens. Environ. 95, 480–492.
Xu, H., 2006, jul. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033.
Xu, H., 2008, jul. A new index for delineating built-up land features in satellite imagery. Int. J. Remote Sens. 29, 4269–4276.
Yeom, J., Han, Y., Kim, Y., 2013. Separability analysis and classification of rice fields using KOMPSAT-2 High Resolution Satellite Imagery. Res. J. Chem. Environ 17, 12.
Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. Int. J. Remote Sens. 24, 583–594.
Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sensing of Environment 24, 583–594.