Letter

A Web Platform Development to Perform Thematic Accuracy Assessment of Sugarcane Mapping in South-Central Brazil

**Marcos Adami *, Marcio Pupin Mello, Daniel Alves Aguiar,
Bernardo Friedrich Theodor Rudorff and Arley Ferreira de Souza**

Remote Sensing Division (DSR), National Institute for Space Research (INPE),
Av. dos Astronautas, 1758, São José dos Campos-SP, 12227-010, Brazil;
E-Mails: mello@ieee.org (M.P.M.); daniel@dsr.inpe.br (D.A.A.);
bernardo@dsr.inpe.br (B.F.T.R.); arley@dpi.inpe.br (A.F.S.)

* Author to whom correspondence should be addressed; E-Mail: adami@dsr.inpe.br;
Tel.: +55-12-3208-6425; Fax: +55-12-3208-6488.

Abstract: The ability to monitor sugarcane expansion in Brazil, the world's largest producer and exporter of sugar and second largest producer of ethanol, is important due to its agricultural, economic, strategic and environmental relevance. With the advent of flex fuel cars in 2003 the sugarcane area almost doubled over the last decade in the South-Central region of Brazil. Using remote sensing images, the sugarcane cultivation area was annually monitored and mapped between 2003 and 2012, a period of major sugarcane expansion. The objective of this work was to assess the thematic mapping accuracy of sugarcane, in the crop year 2010/2011, with the novel approach of developing a web platform that integrates different spatial and temporal image resolutions to assist interpreters in classifying a large number of points selected by stratified random sampling. A field campaign confirmed the suitability of the web platform to generate the reference data set. An overall accuracy of 98% with an area estimation error of −0.5% was achieved for the sugarcane map of 2010/11. The accuracy assessment indicated that the map is of excellent quality, offering very accurate sugarcane area estimation for the purpose of agricultural statistics. Moreover, the web platform showed to be very effective in the construction of the reference dataset.

Keywords: Canasat Project; stratified random sampling; classification; remote sensing

1. Introduction

Due to its agricultural [1], economic [2], strategic [3,4] and environmental [5–7] relevance, sugarcane cultivation in the South-Central region of Brazil has been annually monitored and mapped using Landsat-like images and visual interpretation since 2003 through the Canasat Project (www.dsr.inpe.br/laf/canasat/en) [8]. The annual thematic maps have been used not only to estimate the cultivated sugarcane area but also as reference for monitoring sugarcane harvesting practices [9], for assessing land use change in response to sugarcane expansion [10,11], and for analyzing crop yield [12]. Although these maps were carefully created using images acquired during specific periods of the sugarcane crop calendar, they have not yet been evaluated with an objective method of quality assessment to determine their utility and applicability [13–16].

Foody [17] pointed out that the accuracy of land cover thematic maps should be assessed, not only to provide quality measurement, but also to determine a confidence level for decisions and analyses based on these maps. Indeed, accuracy assessments of thematic maps are essential for validation, acceptance and utilization of land cover maps [16,18]. However, the accuracy assessment process of thematic maps is not always a simple task [19,20]. Difficult access to extensive geographic regions and frequent land use changes can hinder the process of accuracy assessment but should not reduce the credibility of these assessments [21].

Positional and thematic errors are the two major types of errors that need to be evaluated in the accuracy assessment of thematic maps. Positional errors are associated with the misregistration between the thematic classification and the reference data [22,23]. Thematic errors are associated with erroneous labeling of either automatic and/or visual classification procedures and are the major error source of thematic maps [24].

Thematic maps of the Canasat Project estimated 8.35 million hectares of cultivated sugarcane in the South-Central region of Brazil for crop year 2010/11 [25]. According to the Brazilian Institute for Geography and Statistics (IBGE [2]), this cultivated sugarcane represents 87% of the national sugarcane area; the remaining 13% (1.23 million hectares) are cultivated in the Northeast region of Brazil. It is interesting to note that the sugarcane area has more than doubled from 2003 to 2010 in Brazil's South-Central region [25], highlighting its great potential for sugarcane expansion; while the northeast region has remained relatively stable over this same period [2] as there is less available land for expansion. Sugarcane crop in the South-Central region is largely mechanized and consequently cultivated on relatively flat terrain that is easy to access; however, the extensive cultivated area makes it difficult to carry out a field campaign for the validation of thematic sugarcane map. Thus, the objective of this work was to assess the accuracy of area estimation and thematic mapping of sugarcane by the Canasat Project in the 2010/2011 crop year using a novel web platform developed to combine different spatial and temporal image resolutions to classify a large number of points selected by a stratified random sampling procedure.

Considering the difficulties and restrictions inherent to the accuracy assessment process, Stehman [26] proposed the use of a regression estimator along with ancillary data gathered by specialists to reduce field work. Dorais and Cardille [27] integrated the high spatial resolution of images available on Google Earth with a time series of images from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor for monitoring deforestation and evaluating map quality. A similar

process was used by Cohen *et al.* [28] to detect forest disturbance and recovery using a Landsat time series integrated with Google Earth. Indeed, combining images of high spatial resolution with those of high temporal resolution for visual analyses of specific points by specialists seems to be a novel and valuable approach to be used in the accuracy assessment process. Thus, a web platform was developed to simultaneously analyze georeferenced high-spatial resolution (Landsat-like) images and high-temporal resolution (MODIS) images, to validate the maps generated by the Canasat Project and also introduce a novel method for determining the accuracy of the sugarcane map.

2. Materials and Methods

The thematic accuracy assessment of the sugarcane map from the Canasat Project for the South-Central region of Brazil was carried out for the 2010/2011 crop year (harvest from April 2010 to December 2010). The South-Central region of Brazil comprises the states of São Paulo, Minas Gerais, Paraná, Mato Grosso, Mato Grosso do Sul, Goiás, Rio de Janeiro, Espírito Santo, Santa Catarina and Rio Grande do Sul. However, the states of Rio de Janeiro and Espírito Santo have a relative small sugarcane area with low potential for expansion and the states of Santa Catarina and Rio Grande do Sul have an even smaller sugarcane area; therefore, these states were not considered in the present study. Although several subclasses of sugarcane were mapped (for details see Rudorff *et al.* [8]) they were aggregated as a single *sugarcane* class. Therefore, the thematic accuracy assessment accounted for a two-class thematic map, *i.e.*, *sugarcane* and *no sugarcane*. The following remote sensing images and ancillary data were used in the present work: (i) 396 images acquired by Landsat-5 and Landsat-7 from January 2009 through to September 2010; (ii) MODIS-EVI2 time series (February 2000–December 2011) of the MOD09 product for the entire South-Central region (tiles H12V10, H12V11, H13V10, H13V11, H14V10 and H14V11); (iii) a partial sugarcane map for São Paulo state provided by the sugarcane producers to the State Secretary of Environment (SMA-SP); and (iv) information on cultivated sugarcane in municipalities of the study area available at IBGE [2]. All Landsat images were registered based on the orthorectified images from the Enhanced Thematic Mapper Plus sensor (ETM+) of Landsat-7 [29] using a first order polynomial and the nearest neighbor interpolation method [22]. The root-mean-square (RMS) error of the georeferenced images was less than 0.5 pixels. The final preprocessing step applied a linear 2% contrast in all Landsat images. The remote sensing images were integrated in a web platform, using the Virtual Laboratory of Remote Sensing Time Series described by Freitas *et al.* [30].

2.1. Statistical Design

Unlike other crops, sugarcane must be cultivated near a sugar and/or ethanol processing plant to reduce transportation cost and minimize fast postharvest deterioration. Thus, sugarcane is only planted in municipalities that have a nearby processing unit. Because official statistics on cultivated area are a reliable source of information, we used the sugarcane area information from IBGE [2] as the initial step for stratification. Due to the large region covered by the mapping and the characteristics of sugarcane cultivation, municipalities with no sugarcane ($S = 0$) were excluded from the analyses. Stratified random sampling was conducted with the strata (h) chosen based on the proportion of the municipality covered by sugarcane (φ), given by

$$\varphi_i = \frac{S_i}{A_i} \tag{1}$$

where S_i represents the sugarcane area of the ith municipality estimated by IBGE [2]; and A_i represents the total area of the ith municipality. Once the municipalities were assigned to the strata, the municipality boundaries were erased leaving only the four strata (see Figure 1 for a display of the strata). Euclidean distances were computed considering the values of φ for each municipality in the grouping analysis, using the Ward clustering method [31], resulting in a dendogram (see Figure 1) to select the strata. This method minimizes the variance within each stratum. Although the variance of φ_i is not the key characteristic in the estimation of the sugarcane area or an estimation of accuracy, the variance depends on the proportion of the sugarcane area in each stratum, φ_h, since φ_h characterizes a feature of a pixel (point) which is the sampling unit. Thus, for each stratum h the proportion of area of sugarcane (based on the IBGE information) can be defined as φ_h, where φ_h is the ratio between the sum of S_i and the sum of A_i for all municipalities in stratum h.

For each stratum h the number of pixels (population—N_h) was obtained based on the spatial resolution of the Landsat images. We use the binomial function, which is a specific case of the multinomial function [20,32–34] recommended when the thematic map has only two mutually exclusive classes [33] (e.g., *sugarcane* and *no-sugarcane*), to estimate the sample size (n)

$$n = \frac{\left(Z_{\frac{\alpha}{2}}\right)^2 pq}{E^2} \tag{2}$$

where n is the sample size; $Z_{\alpha/2}$ is the two-tailed tabulated value for the standard normal distribution with 99% confidence level; p is the probability of occurrence of the *sugarcane* class, given by the mean of all values calculated in Equation (1) ($\bar{\varphi}$); q is the probability of occurrence of the *no sugarcane* class, given by the relation $q = 1 - p$. We adopted this value of p because it increases the sample size when compared with p values estimated using the expected map overall accuracy. E is the permitted sample error adopted as 2.5%. It is expected that stratified random sampling reduces the standard error relative to the simple random sampling. Indeed we verified that the standard deviation of the overall accuracy was reduced by 2.42 times when comparing the stratified random sampling with the simple random sampling. In fact, the binomial function and the adopted p value provided a larger number of sample points than would be required of stratified sampling to obtain the target sample error of 2.5% but not so large that sampling becomes unfeasible [35].

The standard deviation values in relation to φ_h were extracted along with the number of municipalities (M_h) and the number of pixels (N_h) of the Landsat images. Based on an adaptation of the optimal allocation described by Cochran [32], we used the standard deviation of φ_h instead of the proportion φ_h defined earlier. Thus, the sample size for each stratum (n_h) was calculated by

$$n_h = n \frac{N_h . sd(\varphi_h)}{\sum N_h . sd(\varphi_h)} \tag{3}$$

where n is the sample size for the entire study area (Equation (2)); N_h is the number of pixels of stratum h and sd(φ_h) is the standard deviation of φ in stratum h.

Figure 1. The four sugarcane strata and the *n* selected points in the study area; the trajectory of the field work; the visited points; and some illustrative photos from the field work.
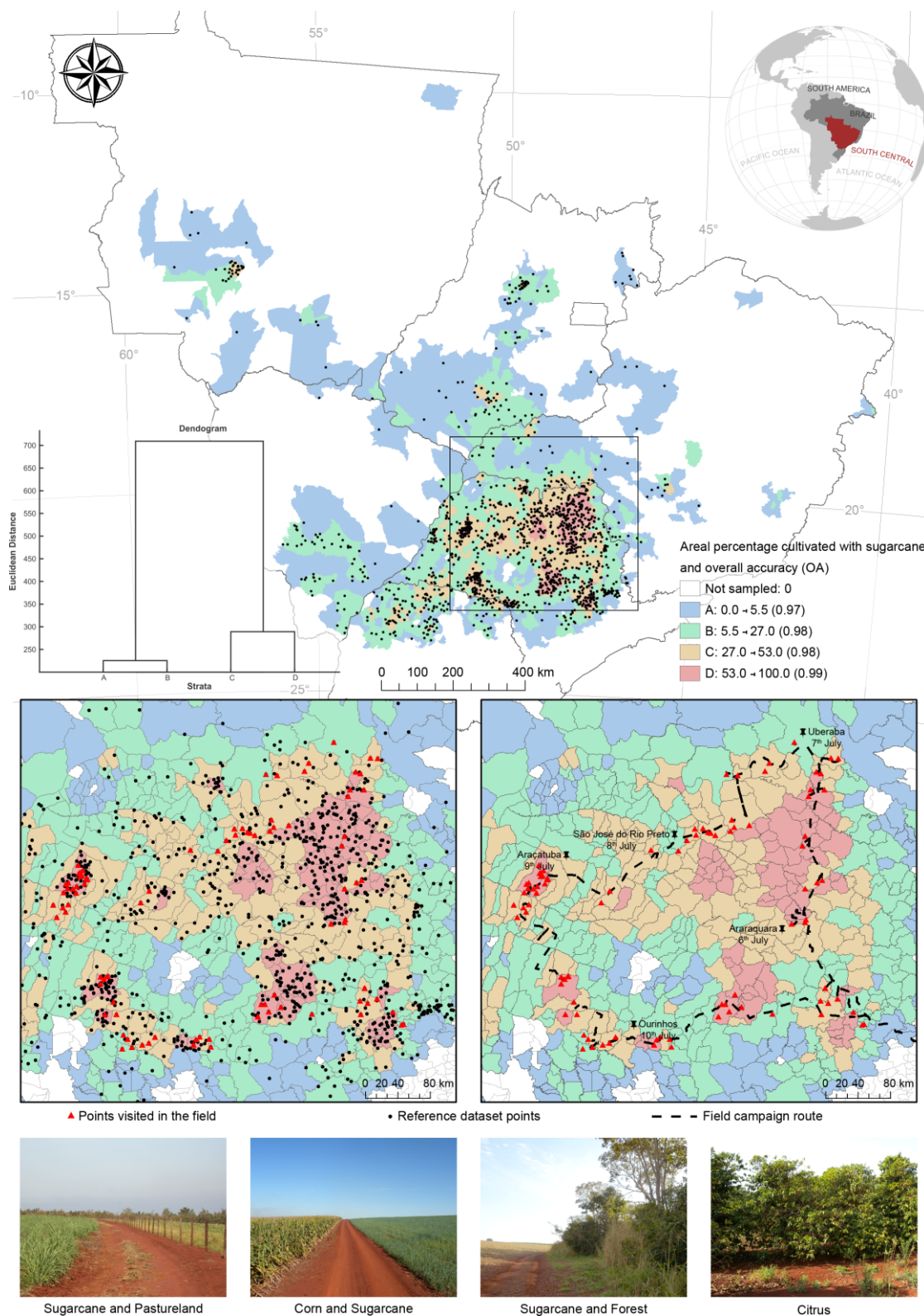
Thus the equations of user's accuracy and producer's accuracy for *sugarcane* (UA_{sh} and PA_{sh}) and *no-sugarcane* (UA_{nh} and PA_{nh}) classes and the overall accuracy (*OA*) are based on the error matrix [21,36–39] for each stratum (*h*), shown in Table 1.

Table 1. Error matrix for each stratum, with overall accuracy (*OA*), user's accuracy (*UA*) and producer's accuracy (*PA*) equations.

	Class	Reference Data		Row Total
		Sugarcane	*No-Sugarcane*	
Map	*Sugarcane*	$\boldsymbol{n_{11}}$	n_{12}	Tms $= n_{11} + n_{12}$
	No-sugarcane	n_{21}	$\boldsymbol{n_{22}}$	Tmn $= n_{21} + n_{22}$
	Column Total	Trs $= n_{11} + n_{21}$	Trn $= n_{12} + n_{22}$	$n_{\mathrm{h}} = n_{11} + n_{12} + n_{21} + n_{22}$
	OA_h		$(n_{11}+n_{22})/n_{\mathrm{h}}$	
	UA	UA$_{\mathrm{sh}} = n_{11}$/Tms	UA$_{\mathrm{nh}} = n_{22}$/Tmn	
	PA	PA$_{\mathrm{sh}} = n_{11}$/Trs	PA$_{\mathrm{nh}} = n_{22}$/Trn	

n_{ij} represents the number of pixels with map class *i* and reference class *j*.

The *OA*, *UA* and *PA* for the entire map was calculated based on the error matrix of each stratum, and considering weights (W_{h} is described further and presented in Table 2).

Table 2. Lower and upper limits of sugarcane % in each stratum and summary of the parameters used in the thematic accuracy assessment.

Stratum	A	B	C	D
Limits (in%)	(0; 5.5]	(5.5; 27]	(27; 53]	(53; 100]
φ_h	1.812%	13.623%	38.048%	64.794%
sd(φ_h)	0.007989	0.018522	0.034417	0.055521
M_{h}	286	343	199	74
N_{h}	12,495,627	28,040,236	24,634,031	25,620,349
W_{h}	0.1376	0.3088	0.2713	0.2822
n_{h}	104	396	504	500
n_{11}	49	191	246	249
n_{12}	3	7	6	1
n_{21}	0	2	6	6
n_{22}	52	196	246	244

n_{ij} is defined in Table 1.

2.2. Web Platform and Reference Database

The system architecture of the web platform, illustrated in Figure 2, was developed within the Virtual Laboratory of Remote Sensing Time-Series [30] and used to visually classify the randomly selected points (*n*) as *sugarcane* or *no-sugarcane* by the four independent interpreters to construct the reference dataset. The system is composed of a server and a client (browser/photo interpreter) side. The process begins after the photo interpreter logs in at https://www.dsr.inpe.br/laf/validamapacana/. After the user successfully logs in, the system obtains a list of all points, highlighting whether each point had or not been already interpreted by the logged user (Figure 2(2)), and build the webpage using HTML and Javascript (Figure 2(1)). As illustrated in Figure 3, once the browser's webpage is loaded (Figure 2(1)) it retrieves two images: a basemap using Google Maps (Figure 2(3)) and a partial

sugarcane map of São Paulo state obtained from the State Secretary of Environment (SMA-SP) (Figure 2(4)). Every map movement sends a new image request to Google Maps (Figure 2(3)) and also retrieves the appropriate shapefiles (Figure 2(4)). To view a data point, the photo interpreter must click on a specific numerical point ID (Figure 2(5)). Once the point was selected, both the ten-year MODIS-EVI2 time series data (Figure 3(6)) for that specific MODIS pixel and the list of available Landsat images around that point (Figure 3(3)) appear in the browser's window. Thus, the photo interpreter can choose the proper Landsat image (Figure 3(3)) that will be overlaid on the Google Maps image (Figure 2(7)) and used by him/her to classify the point as either *sugarcane* or *no-sugarcane* (Figure 3(5)). Once a point has been classified and saved (Figure 3(6)) the system highlights it as a classified one.

This web platform directly addresses the problem of how to go about monitoring and quantifying land-use land cover change over large areas with high accuracy without spending a lot of money on high-resolution data. This platform can be accessed at http://www.dsr.inpe.br/laf/class/validamapacana/en/ login: guest@guest.inpe.br password: 123456. The web platform consists of a Google Maps basemap, over which Landsat-5 images (bands 3, 4 and 5) taken during the 2009 and 2010 years. To the right of the basemap is a list of points (Figure 3). Each point is related to a specific MODIS-EVI2 pixel, which after being clicked, becomes highlighted on the basemap. Furthermore, clicking on a point brings up the corresponding 11-year MODIS-EVI2 [30,40] for that pixel. The user can roll over the MODIS-EVI2 time series bringing up the date on which each MODIS-EVI2 image was compiled, and use this information and the Landsat images to determine whether a point does or does not show evidence of sugarcane.

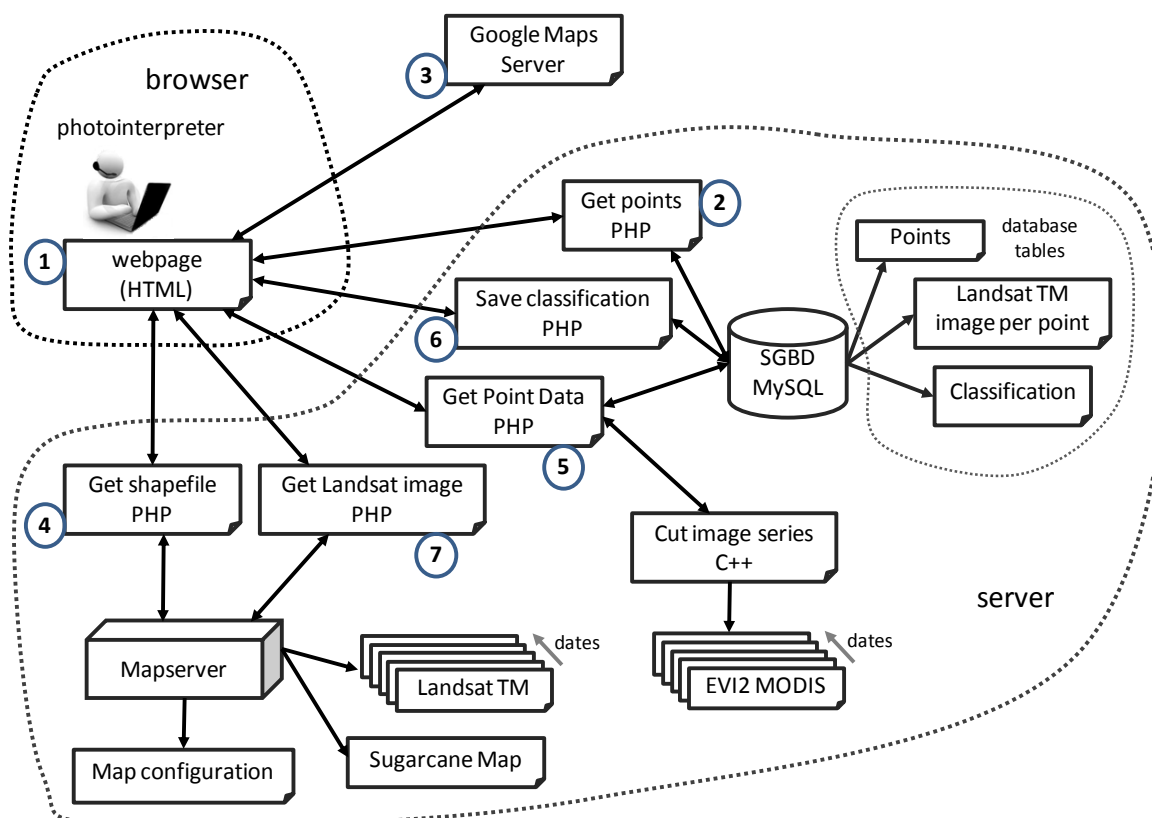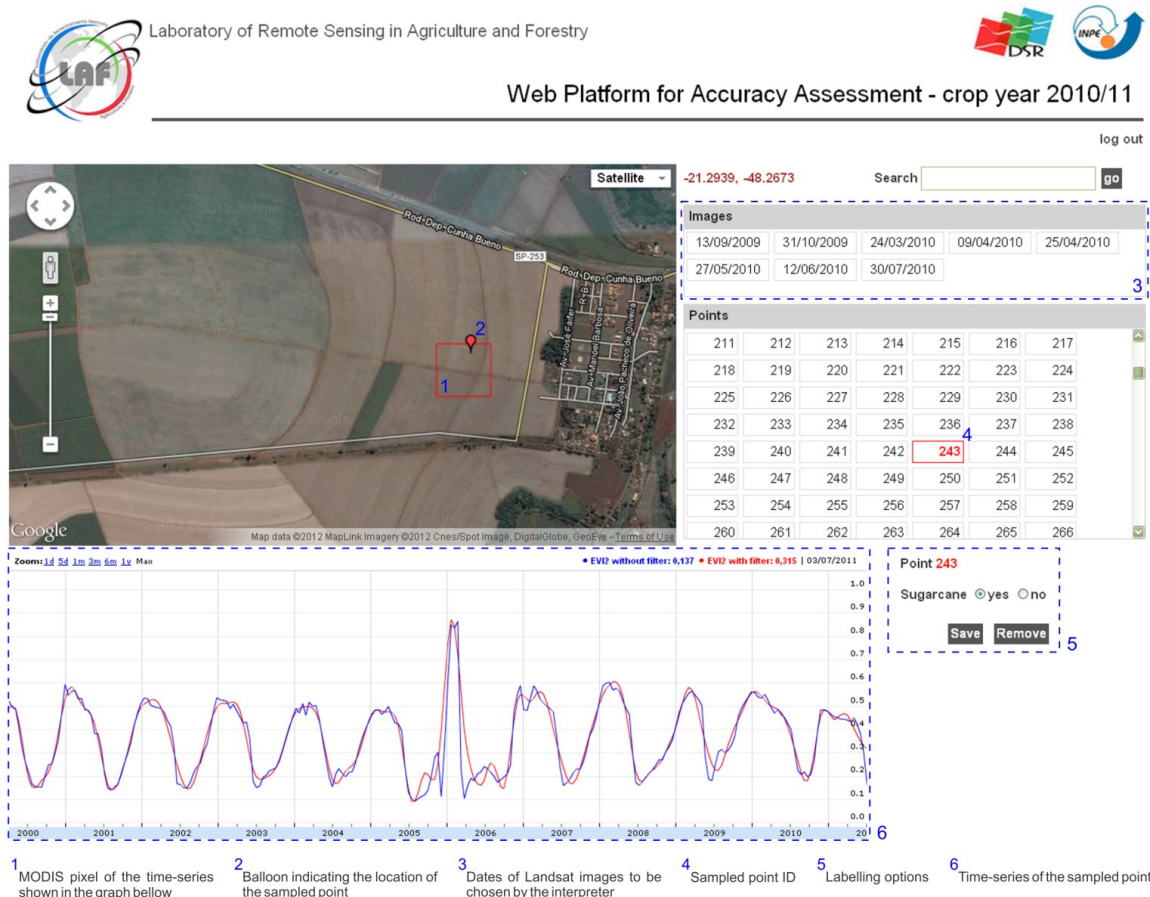**Figure 2.** System architecture.

Figure 3. The web platform developed within the Virtual Laboratory of Remote Sensing Time-Series [30] to classify the 1,504 selected points used to construct the reference dataset.

The classification of the *n* randomly sampled points was performed by four image interpreters following the sugarcane classification methods described by Rudorff *et al.* [8]. One of the four interpreters was specialized in sugarcane mapping and the classification of this interpreter prevailed over the other three in case of disagreement.

However, considering that the construction of the reference dataset based on the web platform is a relatively novel approach, a large field campaign was carried out to evaluate its actual effectiveness. The field campaign was performed from 5 to 10 July 2011 when 2,620 km across sugarcane areas were traversed in the states of São Paulo, Minas Gerais and Paraná. To access the sampled points of interest a Global Position System (GPS) device was integrated within the Global Mapper software. Photos were taken at each visited point and the current land use was briefly described.

3. Results and Discussion

During the 2010/2011 crop year, 902 of the 2,362, municipalities considered in this study (those of São Paulo, Minas Gerais, Paraná, Mato Grosso, Mato Grosso do Sul and Goiás) cultivated sugarcane [2]. Figure 1 shows the dendogram and the spatial distribution of the four sugarcane strata that were defined based on the percentage of sugarcane in each municipality (φ). The lower and upper limits of the sugarcane percentage for each stratum were adjusted as follows: stratum A (0; 5.5]; stratum B (5.5; 27]; stratum C (27; 53]; and stratum D (53; 100] (Table 2). The sample size (n_h) for the entire

study area was 1,504. Table 2 summarizes the following parameters for each stratum h: proportion of area of sugarcane (φ_h) and standard deviation of φ_h (sd(φ_h)); number of municipalities (M_h); number of pixels (N_h) of the Landsat image; weight (W_h), given by $N_h/\Sigma N_h$; number of samples obtained by Equation (3) (n_h) for each stratum; and number of pixels with map class i and reference class j (n_{ij}), as described in Table 1.

All municipalities with more than 53% of sugarcane (stratum D; Table 1) were in São Paulo State (Figure 1) which was responsible for approximately 63% of the sugarcane area in the studied region in crop year 2010/2011 [2]. However, São Paulo state also has the smallest municipalities with an average size of 384 km^2 followed by Paraná (499 km^2), Minas Gerais (687 km^2), Goiás (1,382 km^2), Mato Grosso do Sul (4,578 km^2) and Mato Grosso (6,407 km^2); therefore, it was expected that the most densely cultivated sugarcane municipalities were located in those states with a smaller average for municipality size. Nevertheless, sugarcane has been planted for centuries in São Paulo state as a consequence of favorable soil and climatic conditions [41]. Moreover, there are also other factors that favor sugarcane production in São Paulo and its vicinity: positive socioeconomic aspects; agroindustry infrastructure; a large road network; close proximity to consumer markets; and significant local investment in plant breeding [42].

During the field campaign, 362 of the 1,504 points from the reference dataset were visited. They were distributed in the strata as follows: no points in stratum A; 28 points in stratum B; 114 points in stratum C; and 220 points in stratum D. All 362 points visited in the field were correctly classified by the interpreters indicating that the web platform was very useful in the construction of the reference dataset. Thus, it was possible to calculate the overall and by stratum accuracy indices presented in Table 3 for each stratum.

Table 3. Descriptive statistics of the following accuracy figures: overall accuracy (***OA***); producer's accuracy related to the *sugarcane* class (***PAs***); producer's accuracy related to the *no-sugarcane* class (***PAn***); user's accuracy related to the *sugarcane* class (***UAs***); and user's accuracy related to the *no-sugarcane* class (***UAn***).

Stratum	Statistic	*OA*	*PAs*	*PAn*	*UAs*	*UAn*
A	Estimated	0.97	1.00	0.95	0.94	1.00
	sd	0.0084	0.0000	0.0309	0.0326	0.0000
B	Estimated	0.98	0.99	0.97	0.96	0.99
	sd	0.0075	0.0073	0.0128	0.0132	0.0071
C	Estimated	0.98	0.98	0.98	0.98	0.98
	sd	0.0150	0.0096	0.0096	0.0096	0.0096
D	Estimated	0.99	0.98	1.00	1.00	0.98
	sd	0.0053	0.0095	0.0041	0.0040	0.0097
Overall	Estimated	0.98	0.98	0.97	0.97	0.98
	sd	0.0039	0.0027	0.0048	0.0049	0.0027

Table 3 shows that the accuracy values for all strata were above 96%, but stratum A with PAn_A and UAs_A of 95% and 94%, respectively. The smallest number of samples ($n_h = 104$), together with the lowest sugarcane percentage ($\leq 5.5\%$), contributed to the fact that no omission errors were observed for

the *sugarcane* class in stratum A. Therefore, the omission errors observed in the *no-sugarcane* class were responsible for the lowest accuracy performance of stratum A. In short, the Canasat sugarcane map overestimated in about 6% the sugarcane area in stratum A. Sugarcane overestimation was also observed for stratum B where the mean errors of inclusion and omission were 1%. Inclusion error in stratum B might be associated with cattle raising activity in the vicinity of sugarcane cultivated area which can cause interpretation errors, especially with well-cultivated pasture land [10,43]. In stratum C, the mean inclusion error of 2% (*UAs* = 98%) was compensated by the mean omission error of 2% (*Pas* = 98%) providing accurate area estimation. In stratum D, Canasat sugarcane map underestimated in about 2% the sugarcane area. Although stratum D presents the densest sugarcane cultivated area, other crops are also being cultivated that might cause minor interpretation confusion [8]. However, it is difficult to find a plausible technical explanation for such a low interpretation error which is likely to be at the quality limit of what can be achieved by visual interpretation of Landsat images for sugarcane mapping in this region.

Although the overall mean error of the sugarcane map was 2% (*OA* = 98%) the mean inclusion error of 2% (*UAs* = 98%) was compensated for by the mean omission error of 2% (*Pas* = 98%) providing a mean error associated with the estimate of the sugarcane area close to −0.5% that was calculated using a weighted mean of the strata, where the individual weights were computed by multiplying the area of the stratum by the average sugarcane proportion within the stratum (Tables 2 and 4). The mean area estimation error of −0.5% corresponds to an underestimation of less than 42 thousand hectares of sugarcane in the crop year 2010/2011 based on the sugarcane thematic map of the Canasat Project. It is worth mentioning that the visual Landsat based mapping include the within sugarcane-field road network that is estimated to be around 5% of the total sugarcane area [44]. Sugarcane for the beverage industry to produce "cachaça" or for cattle raising to produce silage is also included in this thematic sugarcane map. However, this sugarcane area is not very significant and remains quite stable from year to year with almost no influence on the relative annual sugarcane area estimation.

Table 4. Overall error matrix weighted by stratum.

| | Class | Reference Data | | Row Total |
		Sugarcane	*No-Sugarcane*	
Map	*Sugarcane*	**732.11**	*19.89*	752.00
	No-sugarcane	*12.30*	**739.70**	752.00
	Column Total	744.41	759.59	*1,504.00*
	OA		98%	
	UA	97%	98%	
	PA	98%	97%	
	Area error	$\frac{(19.89 - 12.30)}{1,504.00}$	0.504%	42,077 ha

4. Summary and Final Considerations

In this work, we assessed the thematic mapping accuracy of the sugarcane map for the South-Central region of Brazil produced by the Canasat Project (www.dsr.inpe.br/laf/canasat/en/)

relative to crop year 2010/2011. To do this, we developed a web platform that integrated different types of remote sensing images and ancillary data to assist the visual interpretation and classification of 1,504 randomly sampled points. We also visited 362 points by traveling 2,620 km in the states of São Paulo, Minas Gerais and Paraná to check the effectiveness of the classification procedure of the web platform, which showed to be very effective in the construction of the reference dataset. The overall accuracy (*OA*) index was 98% varying from 97% for the stratum with less sugarcane (0 to 5.5%) to 99% for the stratum with most sugarcane (53 to 100%). Since part of the omission errors were compensated by the inclusion errors, the mean thematic error associated with the sugarcane area estimation was −0.5%, meaning an omission of less than 42 thousand ha out of a total of 8.3 million ha [25].

The thematic accuracy assessment indicated that the sugarcane map of the crop year 2010/11 from the Canasat Project has an excellent thematic accuracy providing sugarcane agricultural statistics of high confidence. However, it should be noted that this error refers only to the thematic accuracy assessment, since positional accuracy assessment was not evaluated in this work.

Acknowledgments

References

1. FAO (Food and Agriculture Organization of the United Nation). *FAOSTAT: FAO Statistical Database*. Available online: http://faostat.fao.org (accessed on 9 April 2012).
2. IBGE (Instituto Brasileiro de Geografia e Estatística). *Sistema IBGE de Recuperação Automática (SIDRA)*. Available online: http://www.sidra.ibge.gov.br (accessed on 20 January 2012).
3. Goldemberg, J. Ethanol for a sustainable energy future. *Science* **2007**, *315*, 808–810.
4. Leite, R.C.C.; Leal, M.R.L.V.; Cortez, L.A.B.; Griffin, W.M.; Scandiffio, M.I.G. Can Brazil replace 5% of the 2025 gasoline world demand with ethanol? *Energy* **2009**, *34*, 655–661.
5. Macedo, I.C.; Seabra, J.E.A.; Silva, J.E.A.R. Green house gases emissions in the production and use of ethanol from sugarcane in Brazil: The 2005/2006 averages and a prediction for 2020. *Biomass Bioener.* **2008**, *32*, 582–595.
6. Kim, H.; Kim, S.; Dale, B.E. Biofuels, land use change, and greenhouse gas emissions: Some unexplored variables. *Environ. Sci. Technol.* **2009**, *43*, 961–967.
7. Figueiredo, E.B.; La Scala, N., Jr. Greenhouse gas balance due to the conversion of sugarcane areas from burned to green harvest in Brazil. *Agr. Ecosyst. Environ.* **2011**, *141*, 77–85.
8. Rudorff, B.F.T.; Aguiar, D.A.; Silva, W.F.; Sugawara, L.M.; Adami, M.; Moreira, M.A. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. *Remote Sens.* **2010**, *2*, 1057–1076.

9. Aguiar, D.A.; Rudorff, B.F.T.; Silva, W.F.; Adami, M.; Mello, M.P. Remote sensing images in support of environmental protocol: Monitoring the sugarcane harvest in São Paulo State, Brazil. *Remote Sens.* **2011**, *3*, 2682–2703.

10. Adami, M.; Rudorff, B.F.T.; Freitas, R.M.; Aguiar, D.A.; Sugawara, L.M.; Mello, M.P. Remote sensing time series to evaluate direct land use change of recent expanded sugarcane crop in Brazil. *Sustainability* **2012**, *4*, 574–585.

11. Nassar, A.M.; Rudorff, B.F.T.; Antoniazzi, L.B.; Aguiar, D.A.; Bacchi, M.R.P.; Adami, M. Prospects of the Sugarcane Expansion in Brazil: Impacts on Direct and Indirect Land Use Changes. In *Sugarcane Ethanol: Contributions to Climate Change Mitigation and the Environment*, 1st ed.; Zuurbier, P.; van De Vooren, J., Eds.; Wageningen Academic Publishers: Wageningen, Gelderland, The Netherlands, 2008; pp. 63–92.

12. Sugawara, L.M. Variação Interanual da Produtividade Agrícola da Cana-De-Açucar por meio de um Modelo Agronômico. Ph.D. Dissertation, INPE, São José dos Campos, SP, Brazil, 2010.

13. Stehman, S.V. Selecting and interpreting measures of thematic classification accuracy. *Remote Sens. Environ.* **1997**, *62*, 77–89.

14. Smits, P.C.; Dellepiane, S.G.; Schowengerdt, R.A. Quality assessment of image classification algorithms for land-cover mapping: A review and a proposal for a cost-based approach. *Int. J. Remote Sens.* **1999**, *20*, 1461–1486.

15. Powell, R.L.; Matzke, N.; Souza, C., Jr.; Clark, M.; Numata, I.; Hess, L.L.; Roberts, D.A. Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. *Remote Sens. Environ.* **2004**, *90*, 221–234.

16. McRoberts, R.E. Satellite image-based maps: Scientific inference or pretty pictures? *Remote Sens. Environ.* **2011**, *115*, 715–724.

17. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201.

18. Stehman, S.V. Sampling designs for accuracy assessment of land cover. *Int. J. Remote Sens.* **2009**, *30*, 5243–5272.

19. Foody, G.M. Sample size determination for image classification accuracy assessment and comparison. *Int. J. Remote Sens.* **2009**, *30*, 5273–5291.

20. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed.; Taylor & Francis Group: New York, NY, USA, 2009.

21. Stehman, S.V.; Czaplewski, R.L. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sens. Environ.* **1998**, *64*, 331–344.

22. Xiaolong, D.; Khorram, S. The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Trans. Geosi. Remote Sens.* **1998**, *36*, 1566–1577.

23. Dicks, S.; Lo, T. Evaluation of thematic map accuracy in a land-use and land-cover mapping program. *Photogramm. Eng. Remote Sensing* **1990**, *56*, 1247–1252.

24. Zhu, Z.; Yang, L.; Stehman, S.V.; Czaplewski, R.L. Accuracy assessment for the US Geological Survey Regional Land-Cover Mapping Program: New York and New Jersey Region. *Photogramm. Eng. Remote Sensing* **2000**, *56*, 1247–1438.

25. CANASAT. *Sugarcane Crop Mapping in Brazil by Earth Observing Satellite Images*. Available online: http://www.dsr.inpe.br/laf/canasat/en/crop.html (accessed on 9 April 2012).

26. Stehman, S.V. Use of auxiliary data to improve the precision of estimators of thematic map accuracy. *Remote Sens. Environ.* **1996**, *58*, 169–176.

27. Dorais, A.; Cardille, J. Strategies for incorporating high-resolution google earth databases to guide and validate classifications: Understanding deforestation in Borneo. *Remote Sens.* **2011**, *3*, 1157–1176.

28. Cohen, W.B.; Yang, Z.; Kennedy, R. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync—Tools for calibration and validation. *Remote Sens. Environ.* **2010**, *114*, 2911–2924.

29. Tucker, C.J.; Grant, D.M.; Dykstra, J.D. NASA's global orthorectified Landsat data set. *Photogramm. Eng. Remote Sensing* **2004**, *70*, 313–322.

30. Freitas, R.M.; Arai, E.; Adami, M.; Ferreira, A.S.; Sato, F.Y.; Shimabukuro, Y.E.; Rosa, R.R.; Anderson, L.O.; Rudorff, B.F.T. Virtual laboratory of remote sensing time series: Visualization of MODIS EVI2 data set over South America. *J. Comp. Int. Sci.* **2011**, *2*, 57–68.

31. Ward, J.H., Jr. Hierarchical grouping to optimize an objective function. *J. Amer. Statis. Assn.* **1963**, *58*, 236–244.

32. Cochran, W.G. *Sampling Techniques*; John Wiley: New York, NY, USA, 1977; p. 428.

33. Stehman, S.V. Impact of sample size allocation when using stratified random sampling to estimate accuracy and area of land-cover change. *Remote Sens. Lett.* **2012**, *3*, 111–120.

34. Stehman, S.V.; Selkowitz, D.J. A spatially stratified, multi-stage cluster sampling design for assessing accuracy of the Alaska (USA) National Land Cover Database (NLCD). *Int. J. Remote Sens.* **2010**, *31*, 1877–1896.

35. Chen, P.Y.; Luzio, M.D.; Arnold, J.G. Spatial agreement between two land-cover data sets stratified by agricultural eco-regions. *Int. J. Remote Sens.* **2006**, *27*, 3223–3238.

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

37. Stehman, S.V.; Foody, G.M. Accuracy Assessment. In *The Sage Handbook of Remote Sensing*; Warner, T.A., Nellis, M.D., Foody, G.M., Eds.; SAGE: London, UK, 2009; pp. 297–309.

38. Czaplewski, R.L.; Patterson, P.L. Classification accuracy for stratification with remotely sensed data. *Forest Sci.* **2003**, *49*, 402–408.

39. Card, D. Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogramm. Eng. Remote Sensing* **1982**, *48*, 431–439.

40. Jiang, Z.; Huete, A.R.; Didan, K.; Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845.

41. Manzatto, C.V.; Assad, E.D.; Bacca, J.F.M.; Zaroni, M.J.; Pereira, S.E.M. *Zoneamento Agroecológico da Cana-de-açúcar: Expandir a Produção, Preservar a Vida, Garantir o Futuro*; Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisa de Solos, Ministério da Agricultura, Pecuária e Abastecimento: Rio de Janeiro, RJ, Brazil, 2009; p. 55.

42. Lucon, O.; Goldemberg, J. São Paulo—The "other" Brazil: Different pathways on climate change for state and federal governments. *J. Environ. Dev.* **2010**, *19*, 335–357.

43. Xavier, A.C.; Rudorff, B.F.T.; Shimabukuro, Y.E.; Berka, L.M.S.; Moreira, M.A. Multi-temporal analysis of MODIS data to classify sugarcane crop. *Int. J. Remote Sens.* **2006**, *27*, 755–768.
44. Paes, L.A.D. Personal Communication, 2012.