Campo Verde Database: Seeking to Improve Agricultural Remote Sensing of Tropical Areas

Ieda Del'Arco Sanches, Raul Queiroz Feitosa, *Senior Member, IEEE*, Pedro Marco Achanccaray Diaz, *Student Member, IEEE*, Marinalva Dias Soares, Alfredo José Barreto Luiz, Bruno Schultz, and Luis Eduardo Pinheiro Maurano

***Abstract*— In tropical/subtropical regions, the favorable climate associated with the use of agricultural technologies, such as no tillage, minimum cultivation, irrigation, early varieties, desiccants, flowering inducing, and crop rotation, makes agriculture highly dynamic. In this letter, we present the Campo Verde agricultural database. The purpose of creating and sharing these data is to foster advancement of remote sensing technology in areas of tropical agriculture, primarily the development and testing of methods for crop recognition and agricultural mapping. Campo Verde is a municipality of Mato Grosso state, localized in the Cerrado (Brazilian Savanna) biome, in central west Brazil. Soybean, maize, and cotton are the primary crops cultivated in this region. Double cropping systems are widely adopted in this area. There is also livestock and forestry production. Our database provides the land-use classes for 513 fields by month for one Brazilian crop year (between October 2015 and July 2016). This information was gathered during two field campaigns in Campo Verde (December 2015 and May 2016) and by visual interpretation of a time series of Landsat-8/Operational Land Imager (OLI) images using an experienced interpreter. A set of 14 preprocessed synthetic aperture radar Sentinel-1 and 15 Landsat-8/OLI mosaic images is also made available. It is important to promote the use of radar data for tropical agricultural applications, especially because the use of optical remote sensing in these regions is hindered by the high frequency of cloud cover. To demonstrate the utility of our database, results of an experiment conducted using the Sentinel-1 data set are presented.**

***Index Terms*— Agricultural mapping/monitoring, double cropping systems, free available database, remote sensing, synthetic aperture radar (SAR), tropical agriculture.**

I. Introduction

FOOD security is a major concern worldwide and faces the challenge of a continuously increasing global population and limited availability of natural resources. Consequently, agriculture is a key economic activity worldwide, primarily for food but also for fiber and energy (biofuel) production.

Tropical areas have an important position in global food production. Brazil, for instance, is one of the largest global producers and exporters of sugar, coffee, orange juice, soybean, maize, and beef. Brazil is also the lead producer of sugarcane ethanol, an alcohol-based biofuel. Much of this progress is the result of intense research in tropical agriculture. The Brazilian Cerrado biome, for example, was previously considered an area unsuitable for cultivation but has become an agricultural frontier in recent decades and is currently one of the top grain and beef-producing regions in the world [1].

To assure that food production meets the world demands and its environmental impacts are minimized, it is necessary to monitor agriculture activities regularly. Compared to temperate regions, this mission is considerably more challenging for tropical agricultural areas because of the favorable climate associated with the different cultivation systems adopted (e.g., no tillage, minimum cultivation, irrigation, crop rotation, and early varieties) cause intense dynamism and demand year-round monitoring. For this purpose, satellite remote sensing technology can contribute significantly, since it offers repetitive, timely, and accurate information regarding agricultural activity over large areas at relatively low cost [2].

Currently, a variety of high-quality remote sensing data are available free of charge that can be used to monitor agriculture, such as Moderate Resolution Imaging Spectroradiometer (MODIS) products and images from the Landsat series. Several studies have been conducted in this field using these data [2]–[4], but there is still considerable room for advancement, especially in tropical areas. For example, efficient methodologies to identify areas of double cropping (i.e., two consecutive crops cultivated on the same land within a single growing season) using multitemporal remote sensing data have been developed [5], but it remains difficult to identify which two crops are cultivated [4]. Moreover, most crop pattern recognition research has been conducted using a database of temperate regions [6], [7].

In optical remote sensing, cloud cover represents a major constraint, especially in tropical countries [8]–[10]. Alternatives to overcome or at least minimize this problem might be the combined use of data from different sensors, which is not an easy task. It requires cross-calibration

Manuscript received October 23, 2017; revised November 22, 2017; accepted December 29, 2017. Date of publication January 23, 2018; date of current version February 23, 2018. This work was supported by the "Science without Borders" CNPq/CAPES Program under Project 402.597/2012-5. *(Corresponding author: Raul Queiroz Feitosa.)*

I. Del'Arco Sanches and L. E. Pinheiro Maurano are with the National Institute for Space Research, São José dos Campos 12227-010, Brazil (e-mail: ieda.sanches@inpe.br; luis.maurano@inpe.br).

R. Q. Feitosa, P. M. Anchanccaray Diaz, and M. Dias Soares are with the Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro 22753801, Brazil (e-mail: raul@ele.puc-rio.br; pmad9589@ele.puc-rio.br; mdiasoares@gmail.com).

A. J. Barreto Luiz is with the Brazilian Agricultural Research Corporation (Embrapa), Jaguariuna 13820-000, Brazil (e-mail: alfredo.luiz@embrapa.br).

B. Schultz is with Geoambiente, São José dos Campos 12244-000, Brazil (e-mail: schultz.florestal@gmail.com).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2017.2789120

procedures to compensate for differences between sensors, and exploration of microwave synthetic aperture radar (SAR) data, which can be acquired in nearly all weather conditions (including cloudy days).

Even when sufficient remote sensing data are available, another great obstacle in the progress of earth-observation satellite-based agricultural monitoring is the lack of ground truth or reference data for the development and validation of methodologies. The combination of downloadable images with a catalog of supporting ground-based measurements is crucial in studies of regional processes [11]. Moreover, although satellite image-based estimates are the most cost-efficient way to obtain nationwide wall-to-wall statistics of cropland in a timely manner, satellite image alone does not provide satisfactory data, presuming the availability of field data representing all vegetation communities in a target area [12]. Collecting these data requires field campaigns that are time consuming and expensive and requires people with agricultural knowledge in the target region. Thus, reference data are essential for the development of automatic and accurate crop mapping from remote sensing data.

To the best of our knowledge, there is no public agricultural database for tropical regions. Motivated by this demand, we created a database for the municipality of Campo Verde, Mato Grosso state, Brazil, which is a great producer of soybean, maize, cotton, and beef in the Cerrado biome. This is a highly dynamic agricultural tropical area, where double cropping systems, crop-livestock rotation, and no-tillage systems are used. In this letter, we present this database, which contains the following:

1) reference data—shapes and land-use classes of 513 fields for the period between October 2015 and July 2016, which covers the Brazilian 2015/2016 crop year;
2) remote sensing images—a series of preprocessed Sentinel-1 SAR and Landsat-8/Operational Land Imager (OLI) data covering the same period.

We also present the results of an experiment using the Sentinel-1 sequence.

II. Campo Verde Reference Data

Campo Verde is a municipality of Mato Grosso (MT) state in the central west region of Brazil. This municipality is located in the Cerrado biome at a latitude of 15°32′48″ south and longitude of 55°10′08″ west (Fig. 1), and it has an area of 4782.118 km^2 with an altitude of 736 m. Campo Verde presents the Tropical Aw climate according to the Köppen–Geiger [13] classification. The average temperature is 22.3 °C. In winter, there is much less rainfall than in summer, and the average annual rainfall is 1726 mm. The predominant soils in this region are latosoils.

The base of the Campo Verde economy is agribusiness. The most cultivated crops are soybean (210 100 ha), maize (88 760 ha), cotton (81 996 ha), sorghum (3100 ha), and beans (2400 ha); numbers in parenthesis correspond to the planted area of each culture in 2015 [14]. There is also livestock (mainly cattle, chickens, and pigs) [15] and forestry (eucalyptus) production [16]. The agricultural calendar

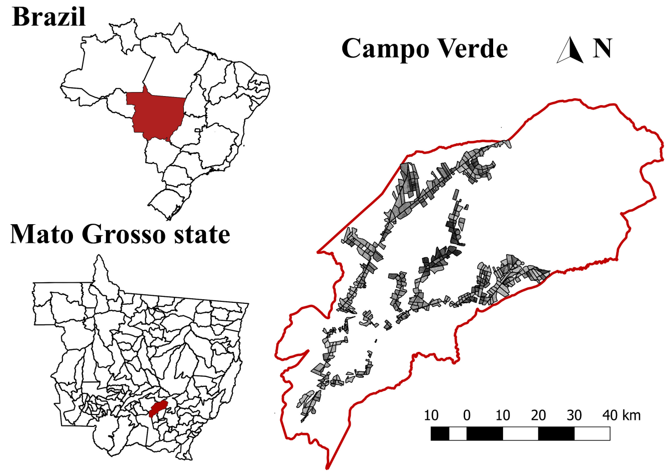

Fig. 1. Campo Verde municipality, Mato Grosso state, Brazil.

(Brazilian crop year) stretches from late August to the following July, with two planting periods during the rainy and dry seasons (first and second harvests, respectively). In Campo Verde, soybean is planted in September–December and harvested in January–April. Cotton can be planted in October to January and harvested in April–September. Maize is planted in January–March and harvested in June–September, and second-harvest maize is planted in January–March and harvested in June–September.

A. Field Campaigns and Land-Use Class Mapping Between October 2015 and July 2016

To build reference data for the Brazilian crop year 2015/2016, two field campaigns were conducted in Campo Verde between December 14, 2015, and December 18, 2015, and May 9, 2016, and May 13, 2016. The first campaign aimed to acquire information about the first harvest (summer crops and rainy season), and the last focused on the crops cultivated during the second harvest (dry season).

In each campaign, Google Earth images and mosaics of recent OLI/Landsat-8 images (WRS-2 path/rows: 225-70, 225-71, 226-70, and 226-71) were used to navigate online along the Campo Verde municipality using a Garmin Global Positioning System device connected to a laptop and Global Mapper software (Global Mapper Software LLC designs, Parker, CO, USA). The localization and land-use classes of several agricultural fields were recorded in each field work. A total of 513 fields were visited in both campaigns and are the areas selected to comprise the database.

The boundaries of the 513 agricultural fields selected were manually drawn based on RapidEye images (5-m spatial resolution). Along with the field data gathered in December 2015 and May 2016, a set of multitemporal Landsat-8/OLI images was used to visually classify the land use of all selected fields between October 2015 and July 2016. This procedure was conducted by an experienced interpreter.

B. Agricultural Characteristics and Practices

In the first campaign (December), most soybean fields were at close to maximum vegetative vigor and the cotton planting season was beginning. In the second campaign (May), most areas were cultivated with cotton or second-harvest maize. These are annual crops; their phenological cycle can last from approximately three to four months (soybeans and maize) and from four to six months (cotton), depending on the cultivar (early, semiearly, and late cycles).

The two field campaigns revealed important information, not only for the harvests themselves but also to map areas cultivated with double cropping systems (i.e., areas cultivated during both first and second harvests) and identify the types of crop rotation adopted in this region: soybean–maize, soybean–cotton, soybean–sorghum, soybean–pasture, soybean–beans, soybean–noncommercial crops (NCC), beans–cotton, maize–cotton, and NCC–cotton. In total, 14 land-use classes were detected: soybean, maize, cotton, beans, sorghum, NCC–millet, NCC–crotalaria, NCC–brachiaria, NCC–grasses (i.e., identified grasses), pasture, turf grass, eucalyptus, Cerrado, and uncultivated soil (i.e., bare soil, soil with crop residues from the previous harvest, and soil with weeds).

The areas cultivated with NCC are normally not harvested; this practice is adopted to improve soil conditions for the following harvest, for example: 1) millet—to provide soil ground cover and improve soil potassium; 2) crotalaria—to promote biological nitrogen fixation; and 3) brachiaria—to control soil nematode infestation.

Double cropping systems have been increasingly adopted in Brazil in recent years. This practice reduces the pressure to expand agricultural frontiers over natural vegetation areas.

In Campo Verde, we have observed areas that were cultivated with soybean in the first harvest and were later used as pasture for cattle (soybean–pasture rotation). In Brazil, it is estimated that 11.5 million ha (1.5 million in Mato Grosso state) are areas that use integrated agricultural production systems (crop–livestock–forest), mostly crop–livestock integration (83%) [17].

III. Sentinel-1 Data

Sentinel-1A is a polar-orbiting satellite that operates day and night and carries a 12-m-long advanced SAR working in the C-band.

To cover the Campo Verde municipality along the crop year 2015/2016, 27 Sentinel-1 images were acquired in the Interferometric Wide Swath Level-1 Mode with a swath of 250 km and a geometric resolution of 20 m. Two images per date at different times (see Table I) were necessary to cover the entire municipality area, except for January, since it covered the entire area of interest, resulting in a sequence of 14 images (see class distribution in Fig. 2). The images were acquired from the Sentinels Scientific Data Hub in level-1 ground range detected and were preprocessed using the Sentinel-1 tool.

First, a radiometric correction was performed according to the following equation:

$$\sigma^0 = \frac{DN^2}{A_\sigma^2} \tag{1}$$

where σ^0 is the radar return, DN is the pixel digital value in amplitude directly taken from the measurement file, and A_σ^2 is the calibration coefficient. The Sentinel products level-1 provide calibration reference tables for this purpose.

TABLE I
Sentinel-1 and Landsat-8/OLI Acquisition Dates Over Campo Verde

Satellite	Year	Month	Date
Sentinel-1	2015	October	29
		November	10, 22
		December	04, 16
	2016	January	21
		February	14
		March	09, 21
		May	08, 20
		June	13
		July	07, 31
Landsat 8 OLI	2015	October	26
		November	11, 27
		December	13
	2016	January	14
		February	15
		March	18
		April	03, 19
		May	05, 21
		June	06, 22
		July	08, 24

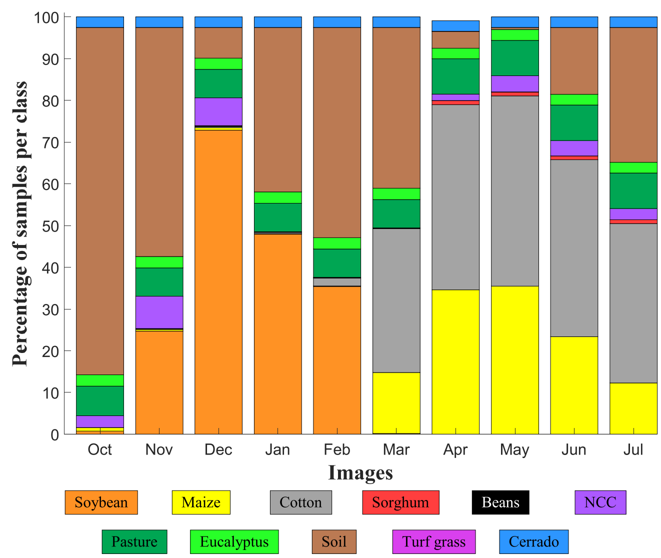

Fig. 2. Percentage of samples per class in each epoch.

Second, a range Doppler terrain correction was applied using a Shuttle Radar Topography Mission digital elevation model (DEM) to geocode the images. In this step, the images were georeferenced to the WGS84 system and the DEM and images were resampled for 10-m resolution using bilinear and nearest-neighbor interpolation methods, respectively. Third, the VV and VH bands in a linear scale were converted to dB.

Next, we performed co-registration of the Sentinel images using the RapidEye mosaic (5-m resolution) as reference. We used the nearest-neighbor resampling method, with root mean square set to 0.05 and the threshold (pixel accuracy) and polynomial order set to 1. Next, using ENVI 4.8 software (Exelis Visual Information Solutions), the bands were stacked to form a single image, which was georeferenced to the UTM projection Zone 21S and Datum WGS84 and resampled

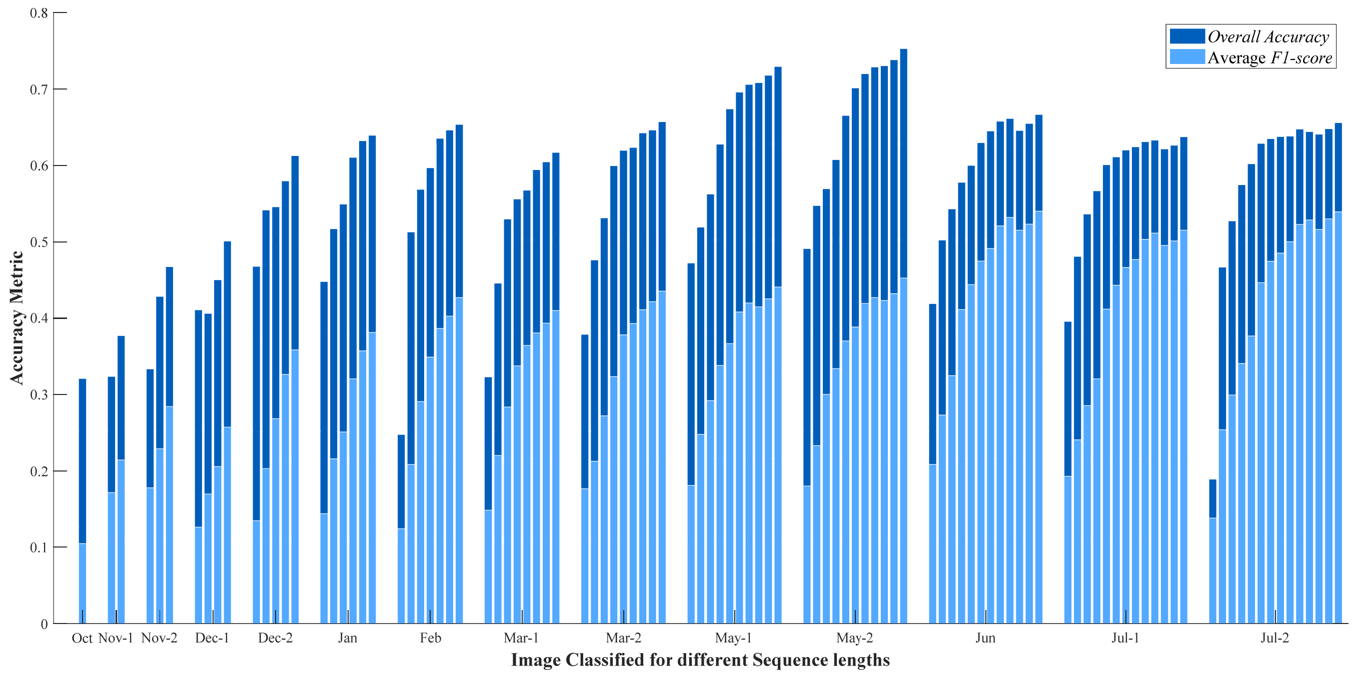

Fig. 3. OA (dark blue bars) and average F1-score (light blue bars) for different sequences (bar groups), taking each image in the database and adding former images to classify that image.

to 10 m. Finally, the images were clipped according to the shape of the Campo Verde municipality area.

IV. Landsat-8/OLI Data

Landsat-8 is a near-polar, sun-synchronous orbiting satellite with a 16-day repeat cycle, which carries two instruments; one instrument is the OLI. OLI is a push-broom sensor with 12-bit quantization. OLI has nine bands covering the spectrum from visible to shortwave infrared with 30-m spatial resolution (except the panchromatic band, which has 15-m resolution).

To cover the Campo Verde municipality along the crop year 2015/2016, 30 OLI images (Table I) corresponding to the Landsat-8 Level-1 data products (distributed as scales and calibrated digital numbers) were acquired from the United States Geological Survey Earth Resources Observation and Science Center. The images were mosaicked (WRS-2 path/row 226-70, 226-71) and clipped to the Campo Verde municipality area.

V. Experiment Using the Sentinel-1 Sequence

To illustrate the use of the Campo Verde database, we report in this section an experiment using the Sentinel-1 images in the data set.[1]

A. Multitemporal Classification by Image Stacking

We applied the conventional method for multitemporal crop mapping, known as image stacking. In this approach, pixels in all epochs are represented in a unique feature space such that spatially correspondent pixels share the same representation over all epochs.

This feature space is formed by stacking the monotemporal features of pixels at the same spatial coordinate to cause an $n \times d$ dimensional feature space for d features per epoch of a sequence comprising n images.

A classifier is then designed to map points of this feature space to a crop type in a given epoch.

B. Experimental Protocol

In our experiment, we extracted texture features of each pixel in each epoch. The features were correlation, homogeneity, mean, and variance, as in [18], computed from gray-level co-occurrence matrices in four directions (0°, 45°, 90°, and 135°) using 3×3 windows per polarization (VV and VH in this case). This approach yielded 32-D feature vectors for each pixel in each epoch. Next, feature vectors of spatially corresponding pixels were concatenated, forming a $32n$ dimensional feature vector, where n is the number of images in the sequence.

A random forest (RF) classifier was trained upon pixels of randomly selected fields. The RF consisted of 250 random trees with maximum depth equal to 25. As our database is unbalanced (see Fig. 2), samples of less abundant classes were replicated to obtain approximately 50 000 samples per class in each epoch. Finally, the classifier was tested on sites not used for training. Each field was used either for training or for testing over the whole sequence, using stratified random sampling from Quantum GIS to take approximately 20% for training and 80% for testing.

[1]The Campo Verde database is available in IEEE Dataport at https://ieee-dataport.org/documents/campo-verde-database.

The aforementioned protocol was applied for different sequences consisting of consecutive images from our database. In all cases, only the most recent image was classified.

C. Results

Fig. 3 summarizes the results in terms of overall accuracy (OA) (dark blue bars) and average F1-score (light blue bars) for all sequences considered in the experimental protocol. Each group of bars presents performance values corresponding to the same image: the one with an acquisition date that is indicated in the horizontal axis. The bars within a group correspond to different sequence lengths. The leftmost bar of a group refers to a sequence consisting only of the image being classified. The next bars to the right indicate the classification performance of the same image but upon sequences of increasing length, which were formed by adding earlier images consecutively. Therefore, the leftmost group has only one bar, since it corresponds to the October image, the earliest in the data set. The rightmost group refers to the latest image. It has 14 bars corresponding to sequences that can be formed by adding earlier images, reaching the earliest image.

Fig. 3 shows that the accuracy tended to increase as prior images were added to the sequence. This improvement was generally significant for up to five images and later declined. In certain cases, for longer sequences (see three leftmost groups), the inclusion of one more image to the sequence was even deleterious. The plot also shows that for a fixed sequence length, performance varied depending on the epoch, reaching the highest values in May. In general, in spite of comparatively higher values for OA, the average F1-score was low due to low F1 accuracies for classes with few samples.

The explanation for these observations requires a more elaborate discussion that goes beyond the scope of this letter.

VI. Conclusion

We argued in this letter that the available public multitemporal databases for crop mapping do not represent the crop dynamics of tropical regions. To fill this gap in the data, we created the Campo Verde agricultural database, which consists of reference data for 513 fields of Cerrado biome in Brazil over ten months. Unlike existing data sets from temperate regions, our database embodies a considerably more complex dynamic due to the favorable climate that allows multiple crops per year and more flexibility to plan the use of land.

Experiments were carried out considering the sequence of Sentinel-1 images performing a multitemporal analysis. Although our database contains a set of SAR and Landsat-8 images, the provided reference data can be used for remote sensing studies in general using data from other sensors (e.g., MODIS and Terra).

As future work, we will elaborate another database for other important Brazilian agricultural region (MATOPIBA). The two databases will provide valuable data for developing and testing methods for crop recognition and agricultural mapping.

References

[1] P. A. A. Pereira, G. B. Martha, Jr., C. A. Santana, and E. Alves, "The development of Brazilian agriculture: future technological challenges and opportunities," *Agricult. Food Secur.*, vol. 1, no. 4, Apr. 2012.

[2] C. Atzberger *et al.*, "Agriculture," in *Land Resources Monitoring, Modeling, and Mapping With Remote Sensing*, vol. 2. Boca Raton, FL, USA: CRC Press, 2015, pp. 71–112.

[3] B. F. T. Rudorff, D. A. Aguiar, W. F. D. Silva, L. M. S. Sugawara, M. Adami, and M. A. Moreira, "Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using landsat data," *Remote Sens.*, vol. 2, no. 4, pp. 1057–1076, 2010.

[4] A. C. Coutinho, J. C. D. M. Esquerdo, L. S. de Oliveira, and D. A. Lanza, "Methodology for systematical mapping of annual crops in Mato Grosso do Sul state (Brazil)," *Geografia*, vol. 38, no. 1, pp. 45–54, 2013.

[5] B. Qiu, M. Zhong, Z. Tang, and C. Wang, "A new methodology to map double-cropping croplands based on continuous wavelet transform," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 26, pp. 97–104, Feb. 2014.

[6] B. K. Kenduiywo, D. Bargiel, and U. Soergel, "Crop type mapping from a sequence of TerraSAR-X images with dynamic conditional random fields," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. III-7, pp. 59–66, Jun. 2016.

[7] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.

[8] E. E. Sano, L. G. Ferreira, G. P. Asner, and E. T. Steinke, "Spatial and temporal probabilities of obtaining cloud-free Landsat images over the Brazilian tropical savanna," *Int. J. Remote Sens.*, vol. 28, no. 12, pp. 2739–2752, 2007.

[9] A. K. Whitcraft, E. F. Vermote, I. Becker-Reshef, and C. O. Justice, "Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations," *Remote Sens. Environ.*, vol. 156, pp. 438–447, Jan. 2015.

[10] I. D. R. Eberhardt *et al.*, "Cloud cover assessment for operational crop monitoring systems in tropical areas," *Remote Sens.*, vol. 8, no. 3, pp. 219–232, 2016.

[11] M. S. Moran, C. D. H. Collins, D. C. Goodrich, J. Qi, D. T. Shannon, and A. Olsson, "Long-term remote sensing database, Walnut Gulch Experimental Watershed, Arizona, United States," *Water Resour. Res.*, vol. 44, no. 5, pp. W05S10-1–W05S10-8, 2008.

[12] E. Tomppo, H. Olsson, G. Ståhl, M. Nilsson, O. Hagner, and M. Katila, "Combining national forest inventory field plots and remote sensing data for forest databases," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 1982–1999, 2008.

[13] M. C. Peel, B. L. Finlayson, and T. A. Mcmahon, "Updated world map of the Köppen–Geiger climate classification," *Hydrol. Earth Syst. Sci.*, vol. 11, no. 5, pp. 1633–1644, 2007.

[14] IBGE—Brazilian Institute of Geography and Statistics. (2015). *Municipal Agricultural Production*. Accessed: Apr. 25, 2017. [Online]. Available: http://www2.sidra.ibge.gov.br/bda/acervo/acervo9.asp?e=c&p=PA&z=t&o=11

[15] IBGE—Brazilian Institute of Geography and Statistics. (2015). *Municipal Livestock Research*. Accessed: Apr. 25, 2017. [Online]. Available: http://www2.sidra.ibge.gov.br/bda/tabela/protabl.asp?c=3939&z=t&o=24&i=P

[16] IBGE—Brazilian Institute of Geography and Statistics. (2015). *Production of Vegetal Extraction and Forestry*. Accessed Apr. 25, 2017. [Online]. Available: http://www2.sidra.ibge.gov.br/bda/tabela/protabl.asp?c=291&z=t&o=29&i=P

[17] ILPF—Integração Lavoura-Pecuária-Floresta. (2017). *ILPF em Números: Região 02—MT, GO e DF. 16 p*. Accessed Apr. 26, 2017. [Online]. Available: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/157581/1/2017-cpamt-ilpf-em-numeros-reginal-2.pdf

[18] J. D. B. Castro, R. Q. Feitoza, L. C. La Rosa, P. M. A. Diaz, and I. Del Arco Sanches, "A Comparative analysis of deep learning techniques for sub-tropical crop types recognition from multitemporal optical/SAR image sequences," in *Proc. 30th SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Niteroi, Brazil, Oct. 2017, pp. 382–389.