Remote Sensing of Environment 114 (2010) 2816–2832

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the Dry Chaco ecoregion of South America

Matthew L. Clark [a,*], T. Mitchell Aide [b], H. Ricardo Grau [c], George Riner [a]

[a] Center for Interdisciplinary Geospatial Analysis, Department of Geography and Global Studies, Sonoma State University, Rohnert Park, California, USA 94928
[b] Department of Biology, University of Puerto Rico, PO Box 23360, Río Piedras 00931-3360, Puerto Rico, USA
[c] CONICET, Instituto de Ecología Regional, Universidad Nacional de Tucumán, Casilla de Correo 34 (4107) Yerba Buena, Tucumán, Argentina

ARTICLE INFO

Article history:
Received 8 December 2009
Received in revised form 16 June 2010
Accepted 2 July 2010

Keywords:
Land cover and land use change
MODIS Enhanced Vegetation Index (EVI)
Time series analysis
Vegetation phenology
Random Forests
Google Earth interpretation
Dry Chaco ecoregion

ABSTRACT

Land use and land cover (LULC) maps from remote sensing are vital for monitoring, understanding and predicting the effects of complex human–nature interactions that span local, regional and global scales. We present a method to map annual LULC at a regional spatial scale with source data and processing techniques that permit scaling to broader spatial and temporal scales, while maintaining a consistent classification scheme and accuracy. Using the Dry Chaco ecoregion in Argentina, Bolivia and Paraguay as a test site, we derived a suite of predictor variables from 2001 to 2007 from the MODIS 250 m vegetation index product (MOD13Q1). These variables included: annual statistics of red, near infrared, and enhanced vegetation index (EVI), phenological metrics derived from EVI time series data, and slope and elevation. For reference data, we visually interpreted percent cover of eight classes at locations with high-resolution QuickBird imagery in Google Earth. An adjustable majority cover threshold was used to assign samples to a dominant class. When compared to field data, we found this imagery to have georeferencing error <5% the length of a MODIS pixel, while most class interpretation error was related to confusion between agriculture and herbaceous vegetation. We used the Random Forests classifier to identify the best sets of predictor variables and percent cover thresholds for discriminating our LULC classes. The best variable set included all predictor variables and a cover threshold of 80%. This optimal Random Forests was used to map LULC for each year between 2001 and 2007, followed by a per-pixel, 3-year temporal filter to remove disallowed LULC transitions. Our sequence of maps had an overall accuracy of 79.3%, producer accuracy from 51.4% (plantation) to 95.8% (woody vegetation), and user accuracy from 58.9% (herbaceous vegetation) to 100.0% (water). We attributed map class confusion to limited spectral information, sub-pixel spectral mixing, georeferencing error and human error in interpreting reference samples. We used our maps to assess woody vegetation change in the Dry Chaco from 2002 to 2006, which was characterized by rapid deforestation related to soybean and planted pasture expansion. This method can be easily applied to other regions or continents to produce spatially and temporally consistent information on annual LULC.

1. Introduction

Land use and land cover (LULC) maps are vital for monitoring, understanding and predicting the effects of complex human–nature interactions that span local, regional and global scales. For example, a spatial depiction of land conversion, such as deforestation for agriculture or pastures, or incremental changes, such as forest degradation and reforestation, are important for reducing uncertainty in carbon stocks and emissions, developing strategies for biodiversity protection, and understanding how globalization affects local and regional land use trends (Houghton, 2005; De Fries et al., 2007). Assessment of rapid land use changes, such as deforestation in the tropics (Archard et al., 2002), requires frequent measurements if it is to be incorporated into management and policy decisions. Furthermore, global issues span political and cultural boundaries, and so LULC maps need to be produced with spatially and temporally consistent information and accuracy. To meet these requirements, we need to develop cost-effective ways for automating the processing of satellite images and the production of LULC maps with high temporal resolution (Defries & Belward, 2000; Skole et al., 1997).

There is a strong tradition of using data from medium resolution sensors (10–60 m)–especially Landsat–for mapping LULC change at local to national scales (Alves & Skole, 1996; Steininger et al., 2001; Roberts et al., 2002; Zak et al., 2004; Boletta et al., 2006; Killeen et al., 2007; Gasparri & Grau 2009; Huang et al., 2009). This level of spatial resolution is generally sufficient for detecting fine-scale land use patterns. However, data costs, small image extent, cloud cover, haze,

* Corresponding author. Tel.: +1 707 664 2558.
E-mail address: mateolclark@gmail.com (M.L. Clark).

0034-4257/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2010.07.001

and infrequent measurements can make data from medium resolution sensors impractical for regional and global mapping (Asner, 2001; Hansen et al., 2008).

Satellites such as MODIS, SPOT-Vegetation, and MERIS offer multispectral measurements with lower spatial resolution (250 to 1000 m), relatively large scenes, and near-daily coverage that allow multiple observations in a year despite cloud coverage. Multi-temporal and multispectral analysis of these data can be used to produce LULC maps and other land cover descriptors, such as the timing, length and frequency of vegetation growing seasons. Several global land cover maps have been produced from low resolution satellites: 1.1-km AVHRR (IGBP DISCover , Loveland et al., 2000; UMD GLCC, Hansen et al., 2000), 1-km SPOT-Vegetation (GLC2000, Bartholomé & Belward, 2005), 500 m and 1000 m MODIS (MOD12Q1, Friedl et al., 2002; MCD12Q1, Friedl et al., 2010), and 300 m MERIS (Globcover, Bicheron et al., 2008). These map products generally focus on separating natural vegetation types for global carbon assessment and differ by source images, spatial scale, reference data, classification techniques and class rules, making comparison problematic (Herold et al., 2008). Evergreen broadleaf trees and areas without vegetation (snow, ice, barren) tend to be well classified, but accuracy is poor with large pixels that mix spectral and temporal signals from trees, shrubs and herbaceous vegetation (Herold et al., 2008). Most global maps provide "baseline" information from a single time period (mostly circa 2000), thus precluding analysis of LULC change using one product with consistent error. The 500 m MCD12Q1 MODIS product offers annual LULC maps from 2001 to 2007, with plans to continue into the future, but these products have just been released and have not been thoroughly assessed for class accuracy and change detection (Friedl et al., 2010).

In summary, there is a lack of LULC map products at regional to continental scales with spatially and temporally consistent information content and error, which prevents analyses of coupled natural and human systems across political boundaries and through time. To satisfy this need, we develop a scalable method for mapping annual LULC at these spatial scales. We use the Dry Chaco ecoregion in South America as a case study; however, the main impetus for this study is a larger project focused on recent LULC change in Latin America and the Caribbean. Our method is novel in that it integrates: 1) reference data interpreted from high-resolution imagery sampled in space and time within an Internet-based tool (Google Earth), 2) a flexible classification scheme based on percent cover thresholds, 3) predictor variables that respond to phenological variation in MODIS vegetation index time series data, 4) annual maps produced at 250 m scale, and 5) a Random Forests classifier that is robust in the face of heterogeneous classes and reference data error.

2. Study area: Dry Chaco ecoregion

To develop and test our method, we worked in the Dry Chaco ecoregion that spans Argentina, Bolivia and Paraguay (Olson et al., 2001). This is the second largest ecoregion in Latin America, covering 790,000 km^2 between 17°32′26″S and 33°52′7″S latitude and 67°43′12″W and 57°59′26″W longitude (Fig. 1), and includes the largest continuous neotropical dry forest (Eva et al., 2004). The ecoregion is characterized by a monsoonal climate with a strong seasonality (dry winters, rainy summers), but average temperature and in particular rainfall vary significantly across the area. Annual mean temperature varies from 18°C in the southern part of the ecoregion to 21°C in the north, and precipitation varies from 500 mm/year in the center to 1000 mm/year in the eastern and western extremes (Minetti, 1999). The vegetation is dominated by dry forest trees and shrubs, but natural grasslands occur in areas with sandy soils and frequent fires. Most of the ecoregion is flat, with elevation rising in the wetter, western side. Elevation above mean sea level for the ecoregion calculated from a digital elevation model (see Section 2.7) had a range

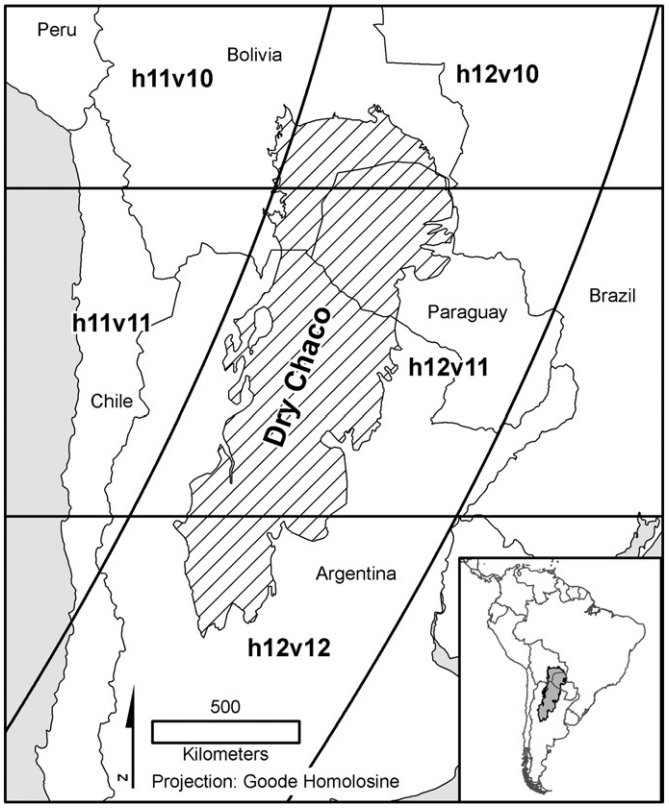


Fig. 1. The Dry Chaco ecoregion study site. MODIS tile extents are shown as the grid, and tile numbers are given with horizontal and vertical reference numbers (e.g., h11v10).

of 56 to 3577 m and average of 326 ± 289 m. Lowland areas, having elevation below 700 m, cover 91% of the ecoregion.

Historically, much of the ecoregion has been severely degraded by extensive cattle ranching and timber and charcoal extraction. In addition, agriculture has occurred in the foothills of the Andes and in irrigated valleys for more than a century. Citrus plantations (mostly lemon, some grapefruit and oranges) and sugar cane are important crops near the Andes. In the northern humid zones there are some banana plantations, while olives are cultivated in the drier southwest. The region also has some scattered tree plantations (pine, eucalyptus and poplar) and minor fruit orchards including blueberries, peaches, figs, and walnuts. During the last 30 years, the conversion of forest to agriculture has accelerated, mostly driven by growing global food demand (Grau et al., 2005). In Argentina, the majority of this new deforestation has occurred in the wetter eastern and western parts of the ecoregion, where millions of hectares of Chaco forest have been replaced with soybeans and pastures (Gasparri & Grau, 2009).

As a consequence of its large area and rapid land use change, the ecoregion has the largest carbon stock and the largest source of emissions from deforestation in Latin America outside the Amazon basin (Gasparri et al., 2008). The ecoregion is part of the Tropical/Subtropical Dry Broadleaf Forest biome, which globally has a high percentage of area converted and relatively low protection (Hoekstra et al., 2005), and it has the largest continuous habitat for large mammals (e.g., jaguars, peccaries) outside the Amazon basin, making it important for regional conservation (Altritcher and Boaglio, 2004; Altritcher et al., 2006; Redford et al., 1990). Given its large extent and rapid changes, developing remote sensing methods to monitor LULC is a priority for the Dry Chaco. Existing studies have used Landsat images, covering parts of the ecoregion in Bolivia (Killeen et al., 2007), Paraguay (Huang et al., 2009) and Argentina (Zak et al., 2004; Boletta et al., 2006; Gasparri & Grau 2009). Methods varied from automated per-pixel classification to visual interpretation; however, despite the lack of consistent products, all studies reported accelerated deforestation since the 1990s.

3. Methods

3.1. Google Earth reference data collection

We collected reference data with human interpretation of high-resolution imagery in Google Earth (GE, http://www.earth.google.com). In Latin America, most of this imagery is from Digital Globe's QuickBird satellite. Google superimposes high-resolution images over coarser images according to image availability, quality and date, and the final image mosaic is streamed to the GE application from Google's servers. The use of GE provides scalability to our method in that its high-resolution imagery is free to access, easy to navigate, is distributed across the region, and can be interpreted with a consistent set of rules.

We first selected random points across the Dry Chaco ecoregion, with point centers at least 1000 m apart. Points were snapped to the closest MODIS pixel (Section 3.4). At each pixel center, a 250×250 m square (62,500 m^2) representing the pixel was generated along with a 4×5 internal grid of cells. Each cell covered 5% of the grid (each internal cell was 62.5×50 m = 3125 m^2). Reference sample points and associated grid layers were developed in the Interrupted Goode Homolosine (IGH) projection, WGS84 datum within ArcGIS 9.2 and then projected to the geographic coordinate system (GCS, i.e., simple cylindrical projection with latitude, longitude) within a KMZ file for viewing in GE (v4.3 to v5.0). Samples with no high-resolution QuickBird imagery in GE were removed from the sample set. Within each sample grid, two technician-level interpreters estimated the percent cover of seven classes with increments of 10%: woody vegetation (*Woody*), herbaceous vegetation (*Herb*), agriculture (*Ag*), plantations (*Plant*), built-up areas (*Built*), bare areas (*Bare*), and water (*Water*). Each interpreter estimated percent cover based on criteria presented in Table 1 (mixed woody is explained below and in Section 2.8.1). Fig. 2 shows typical examples of six classes (excluding *Water* and *MixWoody*) as seen in GE QuickBird imagery with the 250×250 m interpretation grid, which is slightly larger than a MODIS pixel. Each interpreter was allowed to consider the larger landscape context around the sample in assessing component classes, but all percent cover estimates were confined to the land cover in the sample grid. QuickBird image date is available in GE's status bar and the year was recorded for each sample. Samples were then labeled with the class that had the majority cover, one label from each technician. If the two class labels agreed, the estimates of class cover were averaged to provide one estimate for the sample. If the two class labels disagreed, then the technician interpretations were discarded and an "expert" interpreter (one of the authors) estimated the sample's final percent cover, and then the majority-class label came from the expert's estimation. A mixed woody/natural vegetation (*MixWoody*) class was created from samples with a blend of *Woody*, *Herb* and *Bare* cover (see Section 3.8.1).

Our initial interpretations with random sampling yielded 3147 samples, with an overabundance of non-riparian woody vegetation. In contrast, there were relatively few samples for bare, built, plant, water, and riparian area woody, as these areas covered a small fraction of the landscape; and thus, we implemented a stratified random sampling approach for these areas. We first digitized polygons in GE for Bare, Built, Plant, Water and riparian vegetation areas with QuickBird images. Random samples were generated within these polygons and then interpreted using the same protocol, yielding an additional 461 reference samples (13% of total).

3.2. Reference data accuracy assessment

Reference data collected from GE were expected to have spatial and interpretation error. Spatial error may be caused by terrain displacement in QuickBird images that have not been orthorectified. Interpretation error could be in percent cover estimates for the seven land cover classes or in recording QuickBird image dates. These errors are expected to vary among interpreters (Congalton & Mead, 1983; Powell et al., 2004), and can result from disagreement on class percent cover, level of training, fatigue, and data recording errors.

To test the spatial and interpretation accuracy of our GE reference data, we collected an independent set of points from field observations. Points were collected with a Trimble GeoExplorer 3 GPS receiver (PDOP < 6.0, > 4 satellites, and average of 70 positions per point) on January 2 and 3, 2009 along a highway route through the Santiago del Estero, Tucuman, and Salta provinces of NW Argentina, 160 to 950 m elevation. We applied differential corrections to 18 of 62 locations using a public base station (SOPAC, UNSA Salta, 24°43′38.84″ S, 65°24′27.52″W); the remaining 44 locations were not corrected due to missing base station data. At each location, we recorded the compass bearing and estimated distance to the center of patches dominated by one of our land cover classes. Google Earth "placemarks" were digitized with reference to GPS field locations and notes. A 250×250 m sample grid was then generated for each placemark and loaded into GE. Four technicians and one expert estimated percent cover and recorded image date for each field sample using the interpretation criteria in Table 1 (see Section 3.6). Some GPS points were located at road intersections visible in GE QuickBird imagery. The point location (placemark) of the GPS receiver was digitized in GE using detailed field notes and photographs as reference. The difference in IGH projected coordinates (see Section 3.4) between the placemark and GPS point was then used to determine GE QuickBird spatial error in meters.

Table 1
Visual criteria used for estimating percent cover of land use/land cover classes in Google Earth QuickBird image samples (Section 3.2). The class label of each sample, and subsequent classification scheme for a map, was determined by a variable majority cover threshold (Section 3.8.1).

Class	Abbreviation	Visual criteria
Built-up areas	*Built*	Urban and industrial buildings, infrastructure and associated roads
Water	*Water*	Lakes and large rivers
Bare areas	*Bare*	In addition to including areas of bare soil, which could be common in deserts, this class also includes ice, snow, sand dunes, rock, salt flats, and dry riverbeds. Open-pit mines with exposed soil/rock are included in this class.
Agriculture	*Ag*	Agricultural fields with annual crops (e.g., sugar cane, corn, wheat, soybean, rice). Perennial crops (e.g., citrus plantations) are included in the plantation class. Crops can usually be detected by plow lines, rectilinear shapes, and nearby roads and infrastructure. Bare soil in this context was classified as crops, but fallow agricultural land was classified as herbaceous or woody.
Plantations	*Plant*	The major characteristics of plantations are: perennial vegetation and the regular spacing of the plants. Common examples in the Chaco are pine and eucalyptus plantations, citrus and olive orchards, and vineyards. Roads, bare ground, or grass within the plantation were considered as part of the plantation.
Herbaceous vegetation	*Herb*	This class is usually dominated by native or planted grasses and herbs. The most common land use in this class is cattle pasture, which can be distinguished by trails and watering holes. This class can be confused with agriculture but is usually more heterogeneous in color (green, gray, brown) and texture.
Woody vegetation	*Woody*	Trees and shrubs are the major components of this class. Although most areas in this class are natural areas, woody vegetation can also occur within agricultural and urban regions.
Mixed woody vegetation	*MixWoody*	Not interpreted directly in Google Earth. *Woody*, *Herb* and *Bare* percent cover make up this class (see Section 3.8.1)

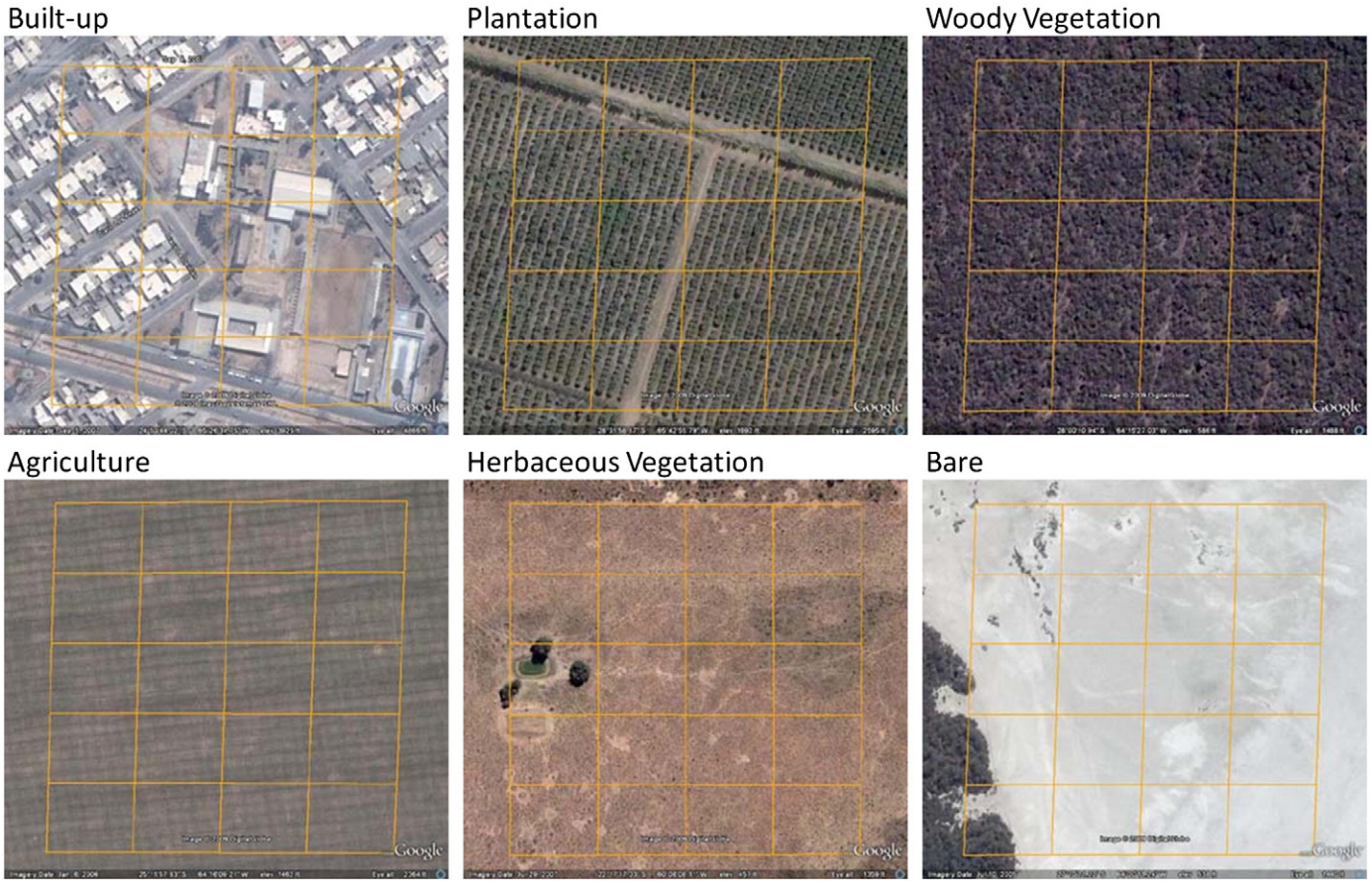


Fig. 2. Example reference samples in Google Earth for six land use/land cover classes. The grid is centered on a MODIS pixel and covers 250×250 m. Each grid cell is 5% of the total sample area.

3.3. MODIS imagery

We used the MODIS MOD13Q1 Vegetation Indices 250 m product (Collection 5) for LULC classification. The product is a 16-day composite of the highest-quality pixels from daily images and includes the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), blue, red, near infrared (NIR), mid-infrared reflectance and pixel reliability (Huete et al., 2002). We focused our study on MODIS tiles H12V10, H12V11, and H12V12, which cover most of the Dry Chaco (Fig. 1). Scenes from day 289, year 2000 to day 273, and year 2008 (184, 16-day scenes) were analyzed in this study. For each day, scenes were reprojected from their native Sinusoidal projection to the IGH projection (sphere radius of 6,378,137.0 m) using nearest-neighbor resampling. The original cell size of 231.7 m was maintained in the reprojection. The IGH projection is a composite, equal-area projection that is useful for calculating area over regional to global scales. A key component of this research was to have the GE sample grids and MODIS pixels cover the same area. Ideally, the MODIS scenes should be reprojected to IGH after classification to minimize geolocation mismatch with GE sample grids. However, GE requires GCS and sample grids projected from the sinusoidal projection to GCS were extremely skewed, especially at higher latitudes, which added error to our percent cover estimates. By projecting rasters to IGH first, we could generate GE sample grids centered on MODIS pixels that were relatively square when projected to GCS for overlay in GE.

3.4. Temporal profiles of classes

The NDVI and EVI vegetation indices (VIs) target the red and NIR spectral regions and respond to vegetation photosynthetic pigment concentrations (mostly chlorophyll) and structure. Detailed time series of these data follow the annual growth cycles–or phenology–of vegetation found in a pixel. Fig. 3 plots the temporal profiles of EVI and NDVI for example classes shown in Fig. 2, with vertical bars indicating the QuickBird image date. In native Chaco forest, trees and shrubs are mostly green (leaf-on) in the wet summers, producing higher VI values (Fig. 3—woody). Winters are relatively dry and many trees and shrubs lose their leaves (Fig. 2—woody), causing VI values to decrease. NDVI and EVI values follow the same temporal trend, but NDVI values are higher. At the other extreme are bare areas, such as dry lake beds (Fig. 2—bare). There is minimal to no vegetation in these areas, and VI values remain fairly flat through the year (Fig. 3—bare). Built-up areas have EVI and NDVI profiles that partly follow the seasonal cycle of precipitation due to urban vegetation, such as trees and herbaceous lots, yet vegetation index (VI) values are relatively low (Fig. 3—built-up) due to high cover of non-photosynthetic surfaces, such as roof-tops and streets (Fig. 2—built-up). Plantations have VI profiles that respond to seasonal precipitation cycles, yet since they are generally irrigated, values are relatively high in the dry winter months relative to native forest (Fig. 3—plantation vs. woody). In this example, the plantation's temporal profile shows a steady increase in VI values from 2001 to 2008, likely the result of increasing biomass and productivity. Herbaceous vegetation is highly variable in the Dry Chaco landscape and includes natural grasslands and pastures for livestock. The QuickBird image for the herbaceous example shows trails radiating from a watering hole, indicating that the area is a cattle pasture (Fig. 2—herbaceous). As with most classes, the herbaceous VI profiles follow seasonal precipitation, yet values are more erratic than woody vegetation. Agriculture in the Dry Chaco is predominantly rain-fed, and areas with mechanized agriculture (e.g., soybeans) have

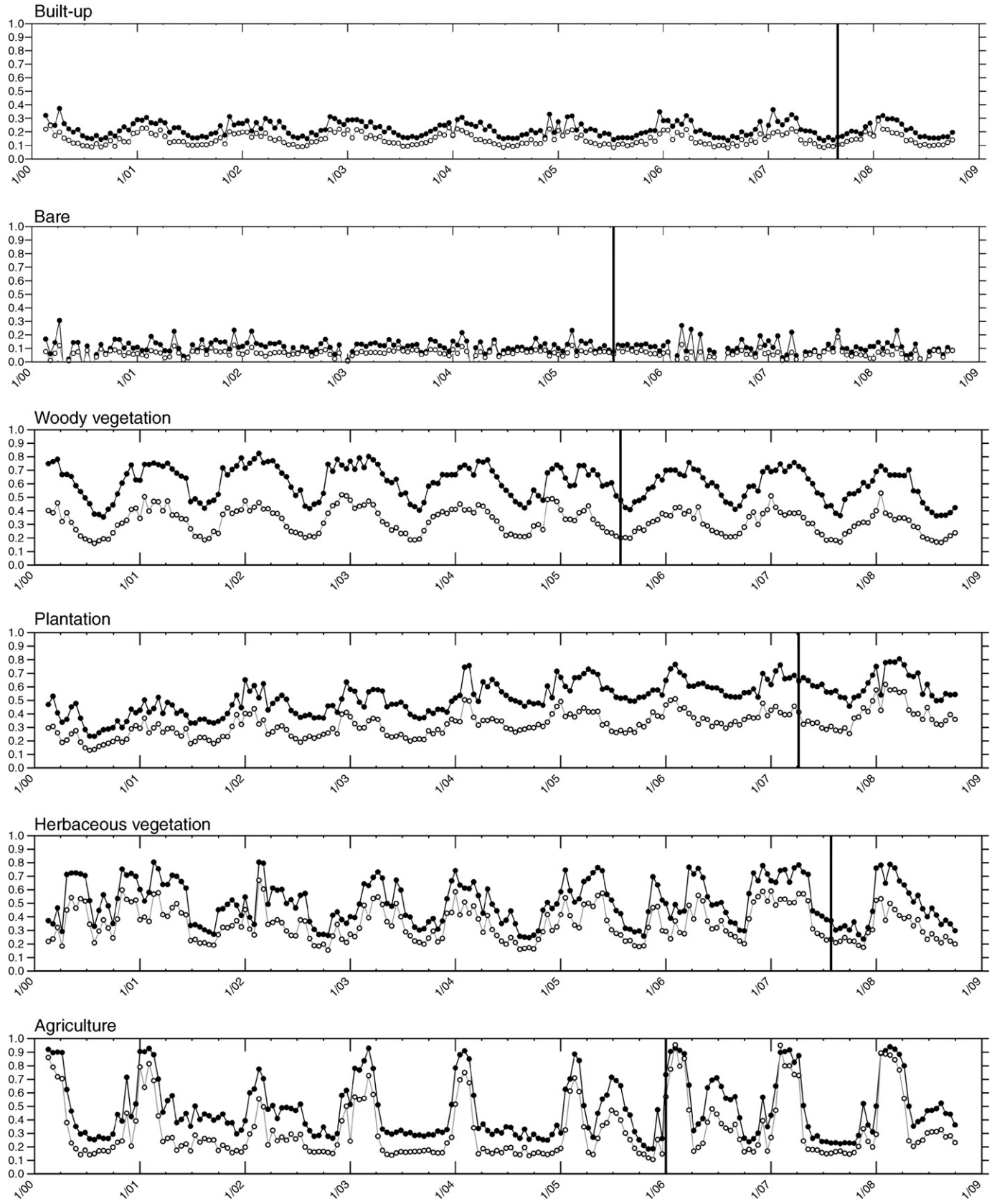


Fig. 3. Temporal profiles of NDVI (closed circles) and EVI (open circles) for the samples shown in Fig. 2. Vertical bars indicate the date when the sample's Google Earth high-resolution image was acquired.

steep and high amplitude VI profiles as crops are planted, mature and are harvested (Fig. 3—agriculture). In between crop cycles, when there are often fallow or tilled fields, VI values plummet to levels between bare soil or herbaceous vegetation in the dry season. In the agriculture example, likely a soybean field, the QuickBird image was taken at the beginning of the growing season when the VI values were

increasing rapidly (Figs. 2 and 3—agriculture). The VIs in this example track one growing season in 2003 and 2004, while there appear to be two crop cycles in 2005 and 2006.

3.5. Phenological variables

We started with the assumption that land cover classes have different phenological signals that can be used for automated classification. Our goal was to derive predictor variables from the VI time series that are sensitive to annual phenological change, yet reduce data volume and signal noise.

3.5.1. Annual statistics

We calculated the annual statistics mean, standard deviation, minimum, maximum and range for EVI, NDVI, red reflectance and NIR reflectance values from calendar years 2001 to 2007. The MOD13Q1 pixel reliability layer was used to remove all unreliable pixels (value = 3) prior to calculating statistics. We explored restricting our analysis to highly reliable pixels (value = 0), but this eliminated too many data points, and so signal error from marginally reliable pixels (value = 1) contributes to the total error in our process. Annual metrics are referenced by the input band *evi*, *ndvi*, *red* or *nir* and the statistic *_mean*, *_stddev*, *_min*, *_max*, and *_range*. For example, *evi_mean* is the mean EVI for 1 year.

3.5.2. TIMESAT processing of time series data

The TIMESAT program (Jönsson & Eklundh, 2004) was designed to analyze phenological signals found in time series data from satellite sensors. The program fits local functions to the time series data points, and then combines these functions into a global model. Phenological variables for each growing season are then extracted from the smooth model function, thereby reducing the influence of signal noise in the raw data. We used TIMESAT v.2.3 (Jönsson & Eklundh, 2004) to process EVI and NDVI data from Julian day 289, year 2000 to Julian day 273, year 2008 (23 points per years × 8 years = 184 points). This temporal window does not align with calendar years, but allowed the program to have ample data to fit a full function to the main Southern Hemisphere growing seasons 2001–2002 to 2007–2008. The MOD13 pixel reliability band was used to weight each point in the time series: value 0 (good data) had full weight (1.0), values 1–2 (marginal data, snow/ice) had half weight (0.5), and value 3 (cloudy) had minimal weight (0.1). Function-fitting parameters used in TIMESAT were: a Savitzky-Golay filter procedure, 3- and 4-point window over 2 fitting steps, adaptation strength of 2.0, no spike or amplitude cutoffs, season cutoff of 0.0, and begin and end of season threshold of 20%.

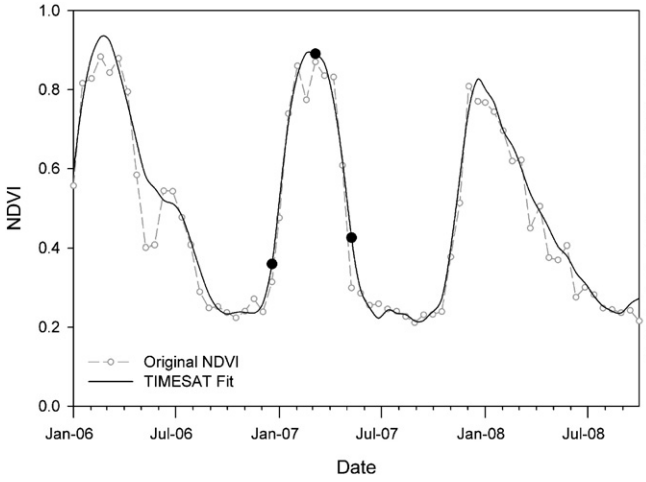
An example fit to NDVI data from an agriculture sample is shown in Fig. 4. Phenological variables extracted for each growing season included: 1. length of the season (*seas_length*); 2. base level, average of the left and right minimum values (*base_level*); 3. largest data value for the fitted function during the season (*maximum*); 4. seasonal amplitude (*amplitude*); 5. rate of increase at the beginning of the season (*left_der*); 6. rate of decrease at the end of the season (*right_der*); 7. small seasonal integral (*small_int*); 8. large seasonal integral (*large_int*); 9. number of seasons in a calendar year (*num_seas*); 10. time for the start of the season (*day_start*); 11. time for the end of the season (*day_end*); and, 12. time for the mid of the season (*day_mid*). A custom program was created to output these phenological variables as raster bands in a stack for each calendar year, 2001 to 2007. TIMESAT outputs fractions of band number for variables 10, 11, and 12. Julian day within a calendar year (0 to 366) was linearly interpolated from a lookup table of MODIS bands and Julian day. Each image pixel was processed independently and phenological variables were labeled with the year in which the season started. Only data for the first growing season in a calendar year were used in our analyses. If TIMESAT failed to retrieve a growing season for a year, the phenological variables for the year were set to a no data value.

3.6. Terrain variables

Additional auxiliary predictor variables used in the classification included elevation and slope derived from the Shuttle Radar Topography Mission (SRTM) 90 m Digital Elevation Model (DEM) that had missing data filled by the Consultative Group for International Agriculture Research (CGIAR, DEM Version 4, http://www.srtm.csi.cgiar.org). Degree slope was derived from the DEM at 90 m and then both elevation and slope were projected to the IGH projection using a bilinear interpolation method and a 231.7 m cell size, to match the MODIS-derived datasets.

3.7. Classification with Random Forests

We implemented per-pixel mapping of land cover with the Random Forests (RF) tree-based classifier (Breiman, 2001). Decision tree classifiers have an advantage over more traditional classifiers, like the maximum likelihood classifier, in that they make no assumptions about data distributions (e.g., normality), can handle data with different scales, and can adapt to noise and non-linear relationships inherent in remote sensing data (Friedl & Brodley, 1997). For example, a highly variable class such as agriculture, which includes different



TIMESAT variable	Value
1. seas_length	157
2. base_level	0.23
3. maximum	0.89
4. amplitude	0.67
5. left_der	0.17
6. right_der	0.17
7. small_int	7.19
8. large_int	4.69
9. num_seas	1
10. day_start	12/19/2006
11. day_end	5/9/2007
12. day_mid	3/6/2007

Fig. 4. Example TIMESAT function fit (black line) to 16-day MOD13 NDVI data (open circles with dashed line) for an agriculture pixel. TIMESAT phenological variables derived from the function are numbered on the figure for the 2006/2007 growing season, with variable names and actual values in the table. See Section 3.6.2 for more information.

crops, cropping patterns and fallow periods, may have multiple decision paths through a tree that all lead to the agriculture class.

The RF classifier builds multiple decision trees from bootstrapped sampling of the reference data. A final class is determined from the majority vote of the ensemble of trees—the forest. To decrease the time it takes to construct trees, RF tests subsets of the predictor variables for each decision branch (node). Roughly 2/3 of the reference data are sampled with replacement to build each tree, while 1/3 of the reference data are withheld from tree construction (called "out-of-bag", or OOB samples). The OOB samples are sent down trees for which they were not used, and the difference between the predicted and actual class is used to calculate an error matrix and unbiased estimate of accuracy (Breiman, 2001). RF also tracks predictor variable importance, which is calculated as a decrease in overall or class-level classification accuracy when the variable is permuted in the OOB samples.

Random Forests have been used in remote sensing to map land cover (Chan & Paelinckx, 2008; Gislason et al., 2006; Pal, 2005; Sesnie et al., 2008), forest biomass and structural parameters (Baccini et al., 2004; Hudak et al., 2008), forest successional stage (Falkowski et al., 2009), habitat (Korpela et al., 2009; Martinuzzi et al., 2009) and invasive species (Lawrence et al., 2006). In land cover studies, the RF classifier has been found to be stable and relatively fast, involve few user-defined parameters and yield overall accuracies that are either comparable to or better than other classifiers, such as maximum likelihood, spectral angle mapper and conventional decision trees (Lawrence et al., 2006), AdaBoost decision trees and neural networks (Chan & Paelinckx, 2008), and support vector machines (Pal, 2005).

3.7.1. Reference data preparation

Annual statistics and TIMESAT phenology were extracted for the year corresponding to the QuickBird image year (2002 to 2007) for each GE reference sample. Since samples came from different years, they included variability in space and time. Sample class labels were assigned according to the class with majority cover. Each GE sample thus included a majority class and its percent cover, and a suite of predictor variables: annual statistics, TIMESAT phenology and terrain variables. This dataset was filtered to remove samples with no TIMESAT values when using EVI or NDVI. This filtered version of samples is called the "reference dataset". There were 3309 total samples, spanning years 2002 to 2007 (Fig. 5). The spatial distribution of samples was clumped, reflecting the patchy distribution of QuickBird scenes in Google Earth.

The reference dataset was then split by randomly selecting 70% and 30% of samples per class for training and testing, respectively, with a maximum limit of 700 samples in any given class (Table 2). The training and test datasets were filtered to provide 5 datasets with 60%, 80% and 100% majority cover thresholds. For example, the 60%

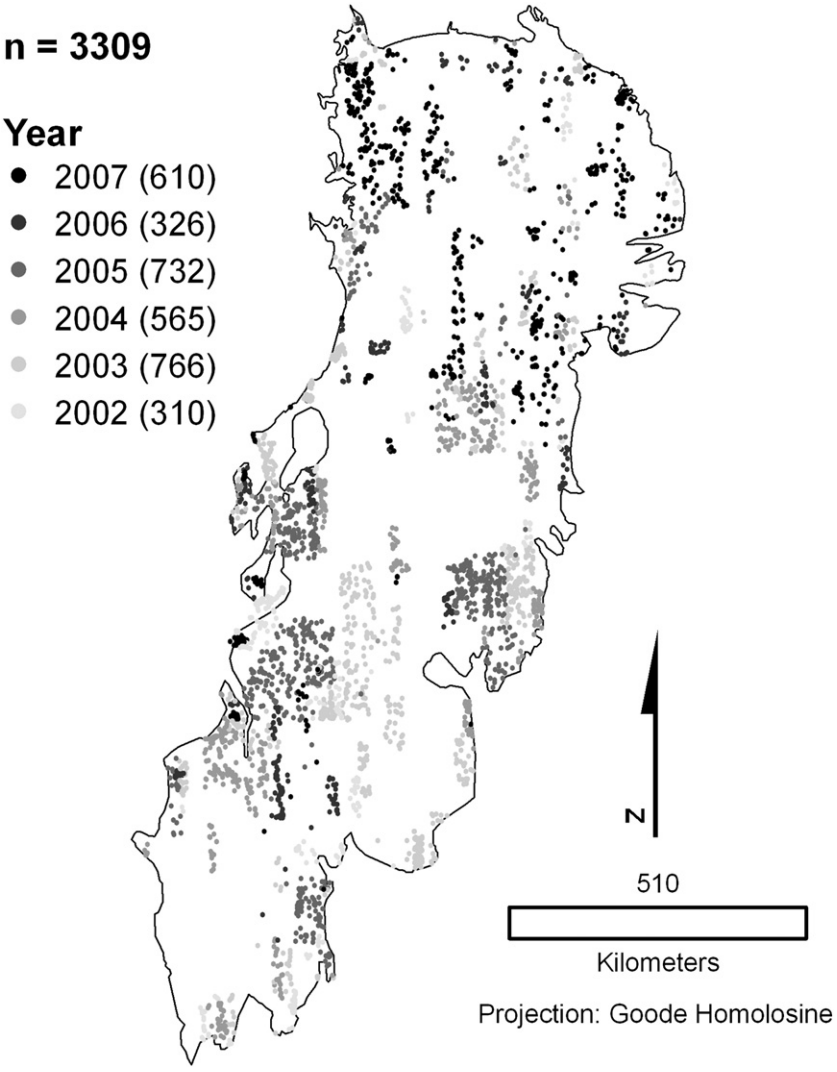


Fig. 5. Distribution of reference samples collected in Google Earth. Year corresponds to the image date of the high-resolution imagery under a sample. Total number of samples for a year is shown in parentheses.

Table 2
Reference samples derived from Google Earth interpretations and used in Random Forests classifier training and testing (accuracy assessment). Samples were labeled with their majority class with ≥60% to ≥100% cover thresholds. Mixed Woody samples had woody cover less than the threshold and remaining cover with herbaceous vegetation or bare less than the threshold.

Training dataset					
	No. of samples by majority-class percent cover				
Majority class	≥60%	≥70%	≥80%	≥90%	≥100%
Ag	304	294	280	252	213
Bare	88	78	65	50	34
Built	87	76	71	55	23
Herb	341	298	247	182	107
Plant	92	85	83	76	56
Water	76	76	76	73	72
MixWoody	54	145	249	401	484
Woody	700	660	620	568	485
Total	1742	1712	1691	1657	1474
Testing dataset					
Majority class	≥60%	≥70%	≥80%	≥90%	≥100%
Ag	131	128	124	110	85
Bare	38	32	21	19	15
Built	37	34	28	18	14
Herb	146	122	99	66	37
Plant	40	39	37	35	29
Water	33	32	32	31	30
MixWoody	24	63	99	172	206
Woody	300	286	273	249	215
Total	749	736	713	700	631
All Samples	2491	2448	2404	2357	2105

threshold training and test datasets included those reference samples whose majority class was ≥60% cover. Samples from a mixed woody (*MixWoody*) class were added to these datasets by using samples that had woody vegetation less than the threshold and with remaining area filled with herbaceous vegetation or bare areas with percent cover less than the threshold. So for the 60% threshold dataset, *MixWoody* samples had <60% woody cover with <60% herbaceous vegetation or bare areas (e.g., 50% *Woody*, 10% *Bare*, 40% *Herb*).

As an experiment to test the sensitivity of classification accuracy to class sample size, we randomly selected samples from the 80% threshold training dataset to create subsets with 20 to 100 maximum samples per class, in 20-sample increments. Each subset had the same samples from the smaller sample subset. For example, we randomly selected 20 samples per class and then added an additional 20 random samples per class to have a training subset with 40 samples per class.

The test datasets were not used in RF training and functioned to independently assess the accuracy of all classifications. Unless indicated, we report accuracy based on an independent test dataset rather than RF's OOB accuracy statistics. With traditional classifiers, including training data in the accuracy assessment of remotely sensed products can overestimate class accuracy (Conglaton, 1991). Furthermore, using an independent dataset allowed us to estimate the gain in accuracy from our post-classification, transition-rule filtering (Section 3.8.3). Breiman (2001) found that OOB statistics from RF are unbiased, and so here we sought to verify this result using our independent test dataset.

3.7.2. Building Random Forests (RF)

We implemented the RF classifier using R (v. 2.7.0; R Development Core Team, 2010) and the "randomForest" package (v. 4.5-27; Liaw & Wiener, 2002). Parameters used for all RFs were: 1000 trees (*ntree* = 1000), minimum of 5 samples in terminal nodes (*nodesize* = 5), and sqrt(p) as number of variables randomly sampled as candidates at each split, where p is number of variables (*mtry* = default). A large node size reduces the complexity of trees and runs faster. Initial experiments showed that a *nodesize* = 5 consistently produced RFs with slightly higher accuracy than RFs from *nodesize* = 10.

Random Forests were generated with the reference data at varying percent cover thresholds and with 6 different predictor variable sets for EVI and NDVI, respectively: 1. EVI/NDVI TIMESAT variables; 2. EVI/NDVI annual statistics; 3. set 2 + Red and NIR annual statistics; 4. sets 1 and 3; 5. set 4 + terrain variables (elevation, slope); and, 6. set 3 and terrain variables. All variable set and cover threshold tests were conducted using an R script that output the trained randomForest as an R object, a confusion matrix based on our independent test dataset, and predictor variable importance. We designed a program to apply a selected randomForest object to every pixel in the ecoregion for each year. The script is based on Python 2.5.1 and modules rpy-1.0.3, pywin32-212, numpy-1.1.0, and GDAL-1.5.2. It reads in predictor variable stacks from the appropriate year for each image line and for each pixel performs these steps: 1. apply the randomForest object (e.g., 1000 trees) to the predictor variables, 2. determine the pixel's class from the forest majority vote, and 3. store the class in a data matrix for storage in an output raster. The classification process resulted in land cover maps for each year from 2001 to 2007.

3.8. Temporal filtering with transition rules

We used a temporal filter to remove disallowed LULC transitions from the final classified maps (*sensu* Roberts et al., 2002). For example, it would be unlikely that a woody pixel could be deforested and turn into a built-up area, and then transition back to a woody area in the span of a few years. In contrast, pixels may transition from herbaceous vegetation to agriculture and back to herbaceous vegetation according to land management decisions. We implemented a temporal filter by passing a 3-year moving window through each map pixel, starting in 2001 and proceeding annually to 2007. The filter window thus had year n, $n+1$, and $n+2$. We first tested to see if the classes from year n and $n+2$ were equal; if the classes were equal, and class $n+1$ was a disallowed transition from n to $n+2$, then class $n+1$ was filled with the class from year n. The filter was then advanced 1 year in the temporal sequence, including fill values from previous years in the test for disallowed transitions. Disallowed transitions are defined in Table 3. An example of a disallowed 3-year transition is *Woody–Built–Woody*, while an allowed transition is *Ag–Herb–Ag*.

4. Results

4.1. Google Earth reference data accuracy

The mean horizontal distance between differentially corrected GPS and GE QuickBird control points was 4.3 ± 2.9 m (n = 5, min = 1.0 m,

Table 3
Disallowed class transitions used in the temporal filter (Section 3.9). The filter was a 3-year moving window through each map pixel, starting in 2001 and ending in 2007. The class of the start and end year (n and $n+2$) of the filter are in rows and the class of the middle year ($n+1$) is in columns. A permitted class transition in the middle year between the start and end year class is indicated with "yes", while a disallowed transition is indicated with "no".

		Year $n+1$							
	Class	*Bare*	*Built*	*Ag*	*Herb*	*Plant*	*Water*	*Woody*	*Mix Woody*
Year n and $n+2$	*Bare*	Yes	No	No	Yes	No	Yes	No	No
	Built	No	Yes	No	No	No	No	No	No
	Ag	Yes	No	Yes	Yes	No	Yes	No	No
	Herb	Yes	No	Yes	Yes	No	Yes	No	Yes
	Plant	No	No	Yes	Yes	Yes	No	No	No
	Water	Yes	No	Yes	Yes	No	Yes	No	No
	Woody	No	No	No	No	No	No	Yes	Yes
	MixWoody	No	No	No	Yes	No	No	Yes	Yes

max = 8.6 m). However, uncorrected points had a mean horizontal error of 14.9 ± 13.1 m (n = 16, min = 1.3 m, max = 43.0 m). This suggests that GPS horizontal position error could account for 10 m of the discrepancy between GPS and GE QuickBird control points. With corrected and uncorrected points combined, the mean horizontal error between GPS and GE QuickBird points was 12.4 ± 12.3 m, (n = 21, min = 1.0 m, max = 43.0 m), which is only 5.4% of the length of a MODIS pixel.

Technicians agreed in their majority class interpretation of GE images for 83% of the reference samples (Table 4, top panel). Consequently, 17% of the reference samples were reviewed by an expert, who then determined the final class label. Technicians disagreed more often on *Herb* samples or samples that had no clear majority cover (i.e., 50% cover by 2 classes). They had moderate disagreement on *Ag* and *Bare*, while they had >90% agreement for the other classes. Expert and technician interpretation of the majority class in GE imagery had an 88% and 86% overall agreement with observations from the field, respectively (Table 4, bottom panel). Most of the errors were in discriminating *Ag* and *Herb* samples.

4.2. Experiments with variable sets

We first added sets of predictor variables to the RF classifier (Section 3.8.2) while adjusting the percent cover threshold. Sets with variables based on EVI generally outperformed those based on NDVI, and so we present only results using EVI (Table 5). The overall accuracy calculated from the RF OOB process relative to the independent test dataset, across all variable sets and five percent cover thresholds, differed on average by −1.0 ± 1.1% without *MixWoody* and −0.3 ± 1.1% with *MixWoody* (OOB–independent); that is, the OOB and independent test accuracies were very similar, with OOB providing accuracy estimates that were on average, slightly pessimistic. In discussing results, we focus on accuracy estimates using the independent test dataset.

For all variable sets, overall accuracy was higher when the *MixWoody* class, which had spectrally mixed pixels, was excluded from the classification (Table 5). Overall accuracy steadily improved when mixed pixels were further filtered from the classification by raising the percent cover threshold, i.e., 100% had the highest pixel "purity" (Table 5—no *MixWoody*). In contrast, overall accuracy tended to decrease with a higher cover threshold when the *MixWoody* class was included in the classification (Table 5—with *MixWoody*).

For each cover threshold, there was a steady increase in overall accuracy with the addition of more predictor variables, with higher overall accuracies reached with sets 4 to 6. The lowest overall accuracies were with annual statistics (set 2), while TIMESAT phenology variables (set 1) had improved accuracy. When Red and NIR annual statistics were combined with EVI annual statistics (set 3), accuracy improved on average 8.5% and 7.5% without and with *MixWoody*, respectively. Adding TIMESAT variables to set 3 (set 4) increased overall accuracy on average 2.3% and 3.3% without and with *MixWoody*, respectively. Combining set 4 variables with elevation and slope terrain variables (set 5) increased accuracy on average 1.4% without *MixWoody* and 1.5% with *MixWoody*. Finally, adding terrain variables to annual statistics (set 4) variables, to create set 6, decreased accuracy slightly. The highest overall accuracy achieved was 86.1% without the *MixWoody* class and a 100% cover threshold. The highest overall accuracy with *MixWoody* was 77.7%, using a 60% threshold and variable set 5, yet the user and producer accuracies were 0.0% for *MixWoody* (Table 6). With the 80% cover threshold and set 5, overall accuracy was 74.2% and *MixWoody* user accuracy improved substantially to 54.9% (Table 6).

For operational mapping, we sought a classifier that had a relatively high overall and class accuracy, while accommodating mixed pixels—which are prevalent in the Dry Chaco landscape. We thus chose a classifier with *MixWoody* and a relatively low cover threshold (i.e., less pixel purity). We focused on RFs with variable set 5 since they had the highest overall accuracy (Table 5). The 60%

Table 4
Accuracy assessment of Google Earth interpretation. The top table shows percent agreement of 2 technicians in majority cover class for reference samples (training and testing ≥70% majority cover in Table 2). The bottom table shows agreement for samples visited in the field. There were 4 technician-level interpreters and one expert. Error matrices were calculated for each comparison: expert against field observations and technicians against field observations. Accuracies are percent of points in agreement for all points available (overall) or individual classes.

Reference samples with ≥70% majority cover (training and testing)									
	Overall	*Ag*	*Built-up*	*Herb*	*Plant*	*Bare*	*Water*	*Woody*	*No majority*
Technician agreement	83%	77%	92%	59%	96%	74%	90%	92%	48%

Agreement of expert and technicians with field observations						
	Overall	*Ag*	*Built*	*Herb*	*Plant*	*Woody*
No. of samples	105	27	15	19	16	28
Expert	88%	78%	100%	63%	95%	100%
Tech mean	86%	63%	100%	73%	97%	100%

Table 5
Accuracy results for Random Forests with the 6 variable sets, excluding and including the *MixWoody* class, and a 60%, 80% and 100% cover threshold. Overall = independent test overall accuracy; OOB = out-of-bag overall accuracy.

	No *MixWoody*		With *MixWoody*	
Variable set	Overall	OOB	Overall	OOB
Set 1				
60	66.5%	65.6%	64.5%	63.7%
80	71.0%	68.9%	62.3%	59.8%
100	76.0%	76.0%	60.2%	62.3%
Set 2				
60	62.3%	60.1%	59.3%	57.9%
80	66.6%	63.5%	58.2%	53.6%
100	74.4%	72.1%	55.9%	57.5%
Set 3				
60	74.9%	75.2%	71.6%	72.8%
80	79.0%	78.8%	69.6%	69.1%
100	84.9%	83.8%	69.6%	72.2%
Set 4				
60	77.7%	76.9%	75.6%	74.0%
80	82.1%	80.4%	73.1%	71.1%
100	85.9%	84.9%	72.4%	74.6%
Set 5				
60	80.3%	78.6%	77.7%	75.7%
80	82.9%	82.0%	74.2%	71.7%
100	86.1%	86.7%	72.9%	74.9%
Set 6				
60	76.8%	78.1%	75.2%	76.3%
80	81.6%	81.4%	71.9%	71.6%
100	86.1%	84.9%	71.0%	73.7%

1. Timesat; 2. VI; 3. VI + Red + NIR; 4. VI + Red + NIR + Timesat; 5. VI + Red + NIR + Timesat + Terrain; 6. VI + Red + NIR + Terrain.

Table 6
Class producer and user accuracy results for the Random Forests classification with the *MixWoody* class, variable set 5 and a 60%, 80% and 100% cover threshold.

	Ag	*Bare*	*Built*	*Herb*	*Mix Woody*	*Plant*	*Water*	*Woody*
Class producer accuracy								
60	74.0%	68.4%	78.4%	67.8%	0.0%	47.5%	78.8%	95.3%
80	73.4%	81.0%	78.6%	53.5%	50.5%	51.4%	78.1%	92.3%
100	83.5%	80.0%	57.1%	29.7%	66.5%	27.6%	83.3%	87.4%
Class user accuracy								
60	82.9%	74.3%	76.3%	58.2%	0.0%	90.5%	96.3%	83.9%
80	82.7%	68.0%	81.5%	47.7%	54.9%	79.2%	100.0%	84.0%
100	74.0%	85.7%	80.0%	55.0%	65.6%	66.7%	100.0%	76.7%

Table 7
Class user accuracy for Random Forests classifications with a maximum limit on training samples per class (see Section 3.8.1). The label "All" refers to the classification using all available training data (Table 2—training dataset). All accuracies were assessed with the independent dataset (Table 2—testing dataset).

Samples per class	Overall	*Ag*	*Bare*	*Built*	*Herb*	*MixWoody*	*Plant*	*Water*	*Woody*
20	61.2%	71.4%	57.6%	53.2%	20.9%	39.1%	38.4%	93.1%	89.0%
40	68.2%	76.4%	61.3%	60.5%	46.7%	41.9%	39.7%	100.0%	89.2%
60	71.0%	80.4%	62.5%	65.9%	48.3%	45.5%	51.7%	100.0%	89.5%
80	71.8%	79.3%	62.5%	66.7%	51.1%	47.6%	51.7%	100.0%	90.4%
100	73.6%	84.1%	66.7%	77.1%	55.8%	47.7%	52.8%	96.2%	88.8%
All	74.2%	82.7%	68.0%	81.5%	47.7%	54.9%	79.2%	100.0%	84.0%

threshold was excluded because it could not classify *MixWoody*. The 100% threshold RF was also excluded because it was trained on samples with very high pixel purity. Producer and user accuracy for *MixWoody* was 33.0% and 10.9% higher for the 80% threshold over an RF with a 70% threshold, respectively (data not shown). In contrast, *Herb* producer and user accuracy was 12.9% and 4.8% lower with the 80% threshold relative to the 70% threshold, respectively (data not shown). Accuracy trends for other classes were less clear when comparing the 70% and 80% thresholds; however, we found that the 80% threshold produced more conservative maps depicting change in woody vegetation, and so this threshold was used for subsequent analyses (Section 4.3 to 4.5).

4.3. Sensitivity of Random Forests classifier to class sample size

Our final experiment involved adjusting the maximum sample size per class in the training dataset. We used an 80% cover threshold, variable set 5 (all variables) and included *MixWoody*. There was a general increase in overall and user accuracy with more samples per class (Table 7). Overall accuracy climbed steadily with the addition of training samples. In general, 60 to 100 samples per class produced only slightly lower overall map accuracy relative to using all training samples. *Water* and *Woody* were well classified with only 20 training samples. With 100 samples per class, the RF overall accuracy was 0.6% lower, and *Ag*, *Herb* and *Woody* had higher user accuracy relative to the RF trained with all samples. In contrast, the RF with all training samples had 7.2% and 26.4% higher user accuracy for *MixWoody* and *Plant*, respectively.

4.4. Final classification

Overall classification accuracy for our final map was 74.2% (Table 5—set 5, 80% threshold, *MixWoody*). In reviewing 182 misclassified samples, we found that some test samples were clearly misinterpreted. An expert re-estimated the percent cover for the 182 misclassified samples without prior knowledge of the original class label or RF class prediction. A new class label was then assigned to each sample and accuracy statistics were recalculated (Table 8). Nine test samples were excluded from the revised accuracy assessment because they no longer had >80% majority cover, nor fit into the *MixWoody* class. For the remaining results and discussion of the final map, we use accuracy statistics from Table 8. Class producer accuracy was lowest for *Plant* and highest for *Woody*, while user accuracy was lowest for *Herb* and highest for *Water*. Most *Woody* misclassified (omitted) samples went to *MixWoody*. *Woody* had lower user accuracy, due to misclassified (committed) *Herb*, *MixWoody* and *Plant* samples. *Herb* samples were confused with most classes, especially *Ag* and *MixWoody*, while most misclassified *Ag* samples were confused with *Herb*. *Plant* had low producer accuracy mainly due to confusion with *Woody*. *MixWoody* was also difficult to map. Most misclassified *MixWoody* samples were confused with *Herb* and *Woody*. The user accuracy of *MixWoody* was higher than its producer accuracy because samples were committed from *Herb* and *Woody*.

Fig. 6 shows the mean decrease in OOB overall and class producer accuracy when a given predictor variable was withheld from the RF. A variable that created a greater decrease in accuracy was thus considered to be relatively important in the RF. In terms of overall accuracy, the most important variables were the Red mean, minimum and maximum, NIR mean and base level. The TIMESAT variables were generally less important than Red and NIR annual statistics. Important variables for discriminating *Woody* were similar to those for overall accuracy. Red mean, EVI standard deviation and season length were important for classifying *Ag*. Plantations were best separated with elevation, EVI minimum and base level. Red mean was particularly important for *Built* accuracy, while NIR mean was important for *Water* accuracy.

There were some pixels in the EVI images that did not have clear seasonality (e.g., bare areas, water), which prevented TIMESAT from fitting a function and resulted in null values. Prior to transition-rule filtering, we filled pixels with null TIMESAT variables with the class from the RF using set 6 variables, 80% cover threshold (Table 5, does not include TIMESAT).

We extracted the class value from our transition-rule filtered maps (years 2001 to 2007) for the location and year of samples in the test dataset; filtered map and class values (test dataset with secondary review) were then compared as an error matrix (data not shown). The

Table 8
Error matrix for the final classification using all predictor variables (Set 5) with an 80% cover threshold and independent test data with a secondary expert review of misclassified samples (Section 4.4).

Classification	Reference									
	Ag	*Bare*	*Built*	*Herb*	*MixWoody*	*Plant*	*Water*	*Woody*	Total	User
Ag	96	–	–	8	4	1	1	–	110	87.3%
Bare	–	17	–	2	3	–	3	–	25	68.0%
Built	–	–	22	2	2	1	–	–	27	81.5%
Herb	18	–	3	63	13	5	4	1	107	58.9%
MixWoody	1	1	1	13	63	1	1	9	90	70.0%
Plant	1	–	–	1	1	19	–	1	23	82.6%
Water	–	–	–	–	–	–	25	–	25	100.0%
Woody	3	–	–	6	25	10	–	253	297	85.2%
Total	119	18	26	95	111	37	34	264	704	
Producer	80.7%	94.4%	84.6%	66.3%	56.8%	51.4%	73.5%	95.8%		79.3%

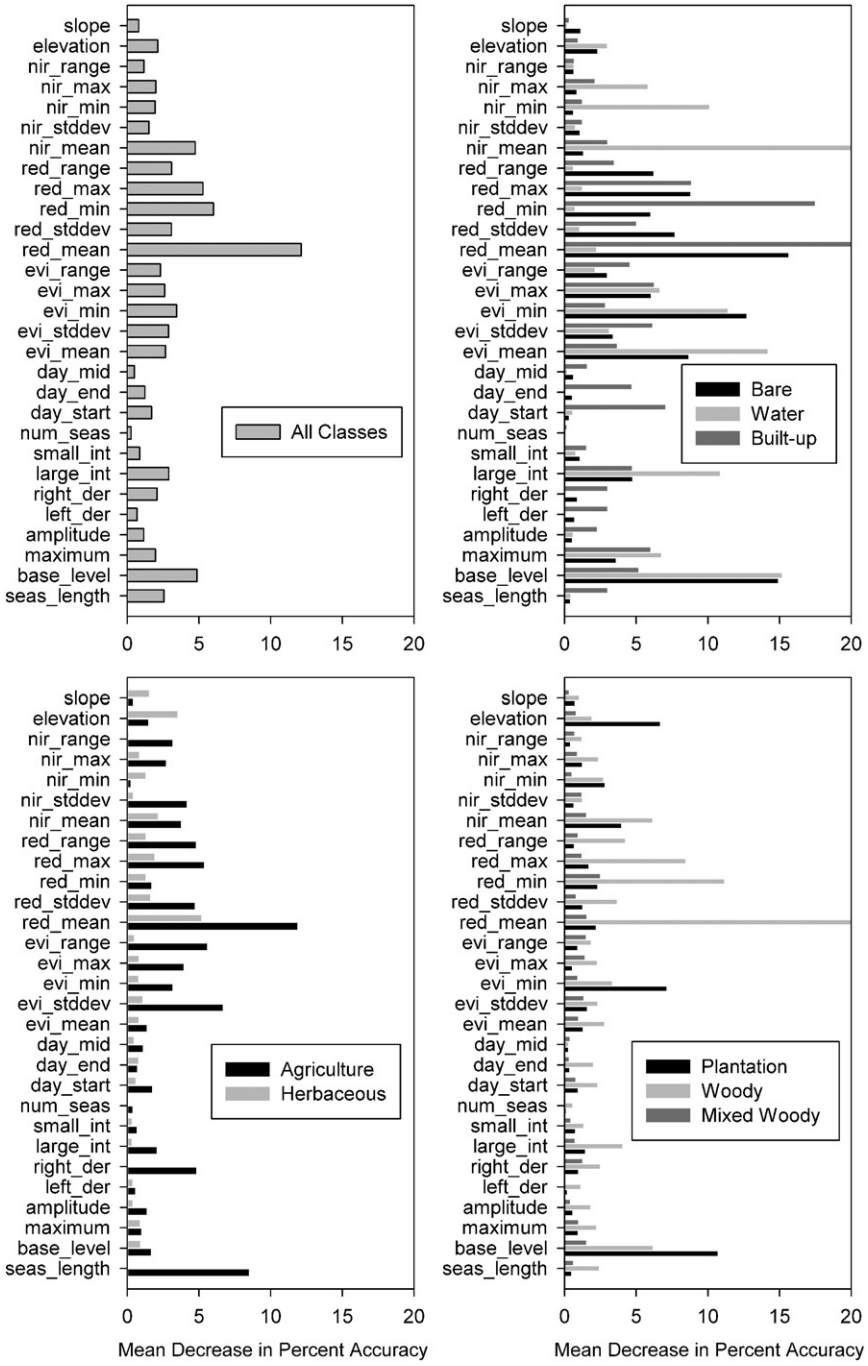


Fig. 6. Mean decrease in overall and class producer accuracy from 1000 trees in Random Forests. Input predictor variables were EVI set 5 (all variables). Those variables with higher mean decrease in accuracy are considered to be more important for overall or class-level classification.

transition-rule filter improved overall accuracy 0.1% relative to the unfiltered maps. User accuracy increased 6.5% for *Built*, 0.7% for *Plant*, and 0.5% for *Woody*. *Bare*, *Herb* and *Water* showed no improvement with transition-rule filtering, while *Ag* and *MixWoody* user accuracy decreased 0.8% and 1.2%, respectively.

4.5. Land use/Land cover change from 2002 to 2006

The area of LULC for the Dry Chaco ecoregion from 2002 to 2006 is presented in Fig. 7. Note that only years 2002 through 2006 had class values changed by the transition-rule filter, and so we describe change within those years. *Woody* decreased each year while *Ag, Herb* and *MixWoody* expanded (Fig. 7). From 2002 to 2006, our maps showed a 6,969,951 ha decrease in *Woody*, or 11.9% of the 2002 area. In contrast, *Ag* increased by 1,650,564 ha, *Herb* increased by 984,634 ha and *MixWoody* increased 4,392,469 ha, which were 44.9%, 17.5% and 55.0% increases from their 2002 area, respectively. Of the pixels that were *Woody* in 2002 and another class in 2006, 68.9% went to *MixWoody* (5,695,721 ha), 13.3% went to *Ag* (1,100,303 ha) and 16.7% went to *Herb* (1,382,515 ha). There were also 659,233 ha of *MixWoody* that transitioned into *Woody* between 2002 and 2006. Given the classifier confusion of *Woody* and *MixWoody*, some of the *Woody* loss or gain is certainly related to misclassification with *MixWoody*; however, part of this change is likely related to forest degradation (e.g., *Woody* to *MixWoody*) or woody growth (e.g., *MixWoody* to *Woody*). There was much less confusion between *Woody* and *Ag* or *Herb* samples, and

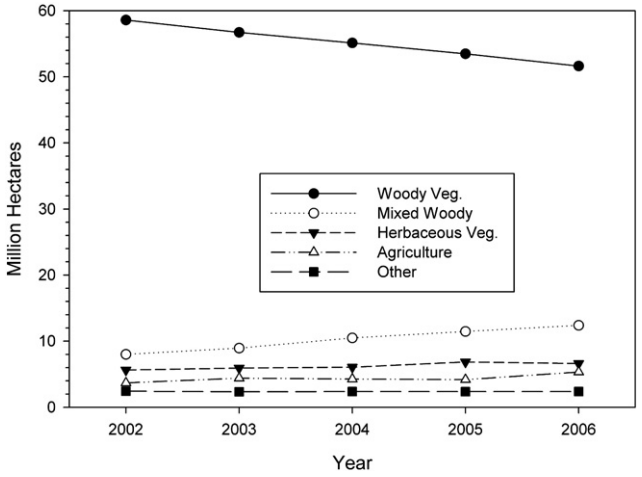


Fig. 7. Area of the Dry Chaco ecoregion covered by woody vegetation, agriculture, herbaceous vegetation, mixed woody and all other classes (plantations, water, bare, and built-up) for years 2002 to 2006.

transitions between these classes are more confident and represent deforestation (e.g., *Woody* to *Ag*, *Herb)* or reforestation (*Ag*, *Herb* to *Woody*).

The LULC maps for 2002 and 2006 show zones of deforestation in southwest Paraguay near Mariscal Estigarribia and northwest Argentina in the provinces of Salta, Tucuman and Santago del Estero (Fig. 8). Deforestation tended to be either on the periphery or in-fill of existing agricultural zones. Forests were converted mainly to agriculture in Argentina, while they were converted to pastures (*Herb*) in Paraguay. In both countries, converted areas typically featured large, regularly cut blocks, which were often established by first thinning forests (e.g., with tractors and chains) resulting in *Woody* transitioning to *MixWoody*, followed by *Herb* or *Ag* in subsequent years (Fig. 9—inset). There was a large increase in *MixWoody* from 2002 to 2006 in the southern part of the ecoregion between La Rioja and San Luis, Argentina (Fig. 8). It is not clear what created this consistent trend—*MixWoody* increased in every year during the study period. Given that this area includes comparatively more open-canopy forests (i.e., with spectral characteristics similar to *MixWoody*) due to drier conditions than the rest of the ecoregion, this may reflect subtle changes in canopy cover due to changes in grazing pressure or decreasing rainfall.

5. Discussion

An objective of this study was to establish a method for mapping annual LULC that can be easily applied at broader spatial scales with internally consistent information content and error. We were interested in producing annual maps of woody vegetation, herbaceous vegetation and agriculture, as these are the dominant classes responding to recent economic activity and other social factors. The final classification scheme used all predictor variables—TIMESAT phenological metrics, annual statistics from EVI, Red and NIR, elevation and slope. Our reference dataset included mixed pixels, with 80% or greater class cover. The exception was mixed woody vegetation, which was a heterogeneous class with 20–80% woody cover with bare or herbaceous understory. We performed a statistically rigorous accuracy assessment with an independent test dataset that had the same spatial scale and class evaluation protocol as the training dataset (*sensu* Stehman & Czaplewski, 1998). Samples were randomly selected from the entire region and through time and included mixed pixels, which lower accuracy statistics but provide a realistic assessment of a map's utility (Powell et al., 2004).

The final overall accuracy for our 2001–2007 maps was 79.3%. Woody vegetation had an 85.2% user's accuracy—the probability that a mapped pixel would be woody vegetation if we were to visit the pixel in the field, or virtually in Google Earth. Mixed woody was an intermediate class between herbaceous and woody vegetation, and so it was confused with both of these classes. Agriculture had similar user accuracy as woody vegetation (87.3%), while herbaceous vegetation had a low user accuracy of 58.9%. There was considerable confusion among herbaceous and agriculture samples, and when we combined the two classes, the resulting class had 85.3% user accuracy and overall map accuracy was 83.0%. If we considered just woody vegetation and combined all other classes, overall map accuracy was 92.2% and woody vegetation and the other class had 85.2% and 97.3% user accuracy, respectively. This *Woody*/*Other* classification has sufficient accuracy for detecting changes in woody vegetation, such as deforestation and reforestation zones.

5.1. Using Google Earth for reference data

The Dry Chaco ecoregion covers 79 million hectares and spans three countries with relatively low road density. Ground-based reference data collection was thus impractical for this study area. In addition, we sought a method that could scale to map LULC over larger regions. We thus explored collecting reference data from "virtual" field visits within Google Earth (GE). This relatively new source of data has recently been combined with other information (e.g., expert opinion, additional imagery) to provide reference data for classification and accuracy assessment of regional and global scale land cover maps (Bicheron et al., 2008; Helmer et al., 2009; Friedl et al., 2010). We found the main advantages of using Google Earth for reference data to be: 1) access is free, includes an intuitive globe interface, and permits quick viewing of a large archive of high-resolution images that are streamed over the Internet; 2) QuickBird high-resolution images were available across the region, particularly for areas with human activity, allowing sampling of spatial variation in predictor variables for LULC classes; 3) QuickBird images were acquired during the same years as the MODIS imagery, allowing sampling of inter-annual variation in MODIS-derived variables—each location was linked to a specific year for extracting variables; 4) relative to a physical field campaign, our time and monetary costs were extremely low, and there were no access constraints (e.g., property ownership, distance to roads, etc.); and, 5) we had a synoptic view of the entire 250×250 m sample plot, which is difficult to achieve on the ground. We found the georeferencing of GE QuickBird scenes to be excellent, with an average error on <5.4% the length of a MODIS pixel.

There are some obvious disadvantages in using GE for reference data. For one, we had no control over the spatial and temporal coverage of high-resolution imagery. Google attempts to provide the most recent, cloud-free images, and most scenes in Latin America are from the QuickBird satellite, launched in late 2001. In our random and stratified random sampling of available scenes, there were over 310 samples for each year from 2002 to 2007, yet no samples from 2001. Although not explored in this study, Google Earth does allow a user to pan through an archive of high-resolution images, providing greater opportunities for sampling reference data through time and verifying mapped LULC change. QuickBird scenes in Dry Chaco tended to cover areas of intense human activity (e.g., agriculture, pastures and urban areas) and nearby natural vegetation, and so there is bias towards these areas in random sampling. However, this bias is advantageous in that it allows greater sampling of anthropogenic classes, which are a relatively small fraction of Dry Chaco's area and spatially clustered.

Visual interpretation of high-resolution imagery has several limitations. Vegetation phenology, atmospheric conditions, illumination, and view angle varied among GE QuickBird scenes, which can increase variation in image interpretation. We used very general LULC classes that could be reliably identified using generic features by interpreters that had little *a priori* understanding of the landscape. For this reason, our woody classes did not distinguish between shrubs and

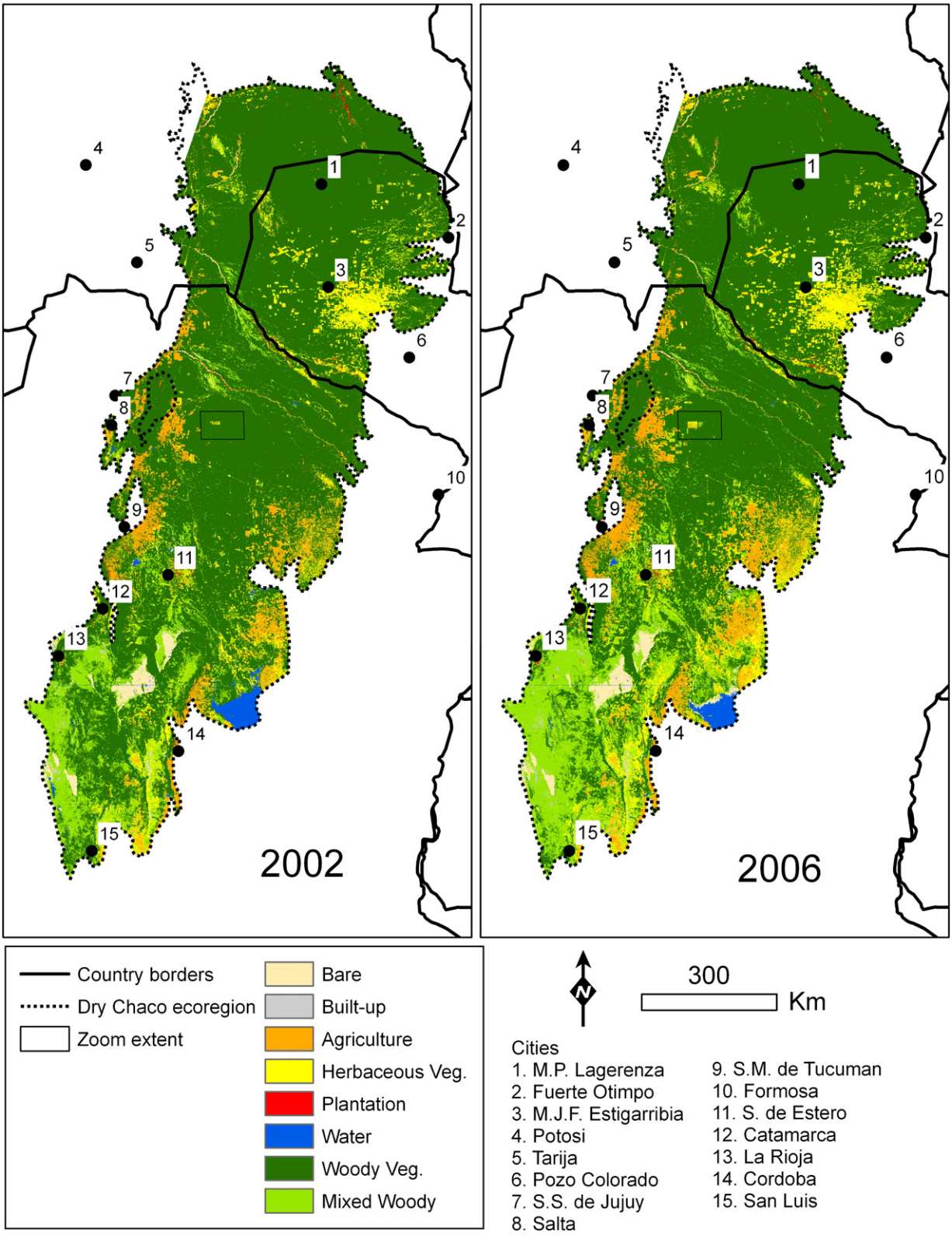


Fig. 8. Change in land use/land cover from 2002 to 2006 for the Dry Chaco ecoregion. The zoom extent is the area shown with more detail in Fig. 9.

trees, and pastures and low vegetation where combined in the herbaceous vegetation class. Even though natural vegetation degradation, successional stage (e.g., secondary forests) and crop type are important dimensions of LULC change, these properties were difficult to accurately distinguish in the high-resolution imagery. Technician and expert disagreements with field observations were mostly with herbaceous vegetation and agriculture. These classes were often difficult to distinguish in the imagery as the interpreter had to rely on visual cues such as plow lines for agriculture and cattle trails and watering holes for herbaceous pastures. However, some of the confusion between these two classes resulted from the time difference between interpreted GE images (years 2002 to 2007) and

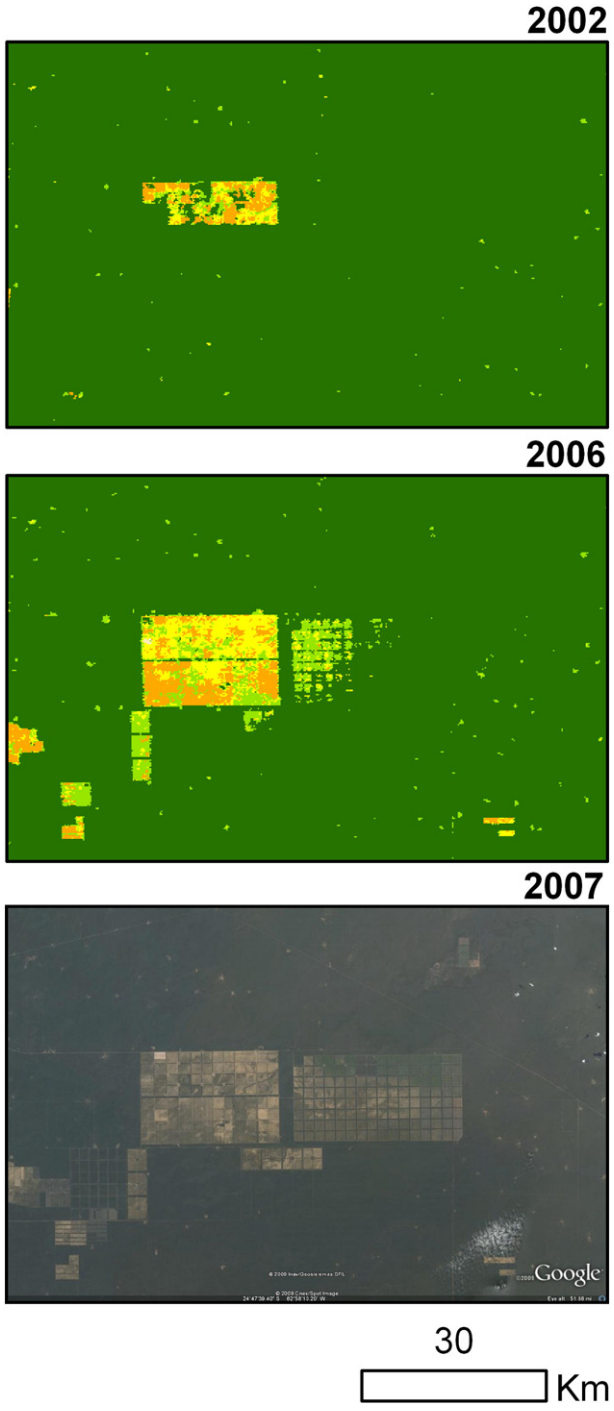


Fig. 9. A large block of agriculture expansion to the east of Salta, Argentina. The zone includes partially deforested blocks in 2006, which are detected as mixed woody vegetation.

the field visit in January, 2009. Pastures could pre-date agriculture fields visited in 2009 and some parcels can alternate annually between food crops (*Ag*) and forage for cattle (*Herb*), depending on factors such as climate and market prices; and thus, there could be mismatch between LULC in field observations and imagery that registers as an interpretation error. We tried to minimize this error by selecting field sites thought to be under regular cultivation at least 7 years. We can further reduce this error from temporal mismatch by combining *Herb* and *Ag* into a single class. With the *Ag*/*Herb* class, expert and average technician accuracy was 95.3% and 88.4%, respectively, and overall accuracy across all classes was 97.1% for the expert and averaged 94.8% for technicians. Therefore, for all classes except *Herb* and *Ag*, GE interpretation accuracy is quite high. We sought to block interpretation errors from entering the RF by comparing the majority cover class from 2 technicians for each sample; the expert then determined the final class only for those samples with technician disagreements. A disadvantage of this approach is that it reduced the total number of samples we could collect in a set time, as there were two technicians, and sometimes one expert, that had to view each sample.

High-resolution imagery over large areas has historically been expensive to acquire, and many past studies involving LULC mapping with low resolution imagery have used Landsat-scale map products as reference data (DeFries et al., 1998; Hansen et al., 2000; Carreiras et al., 2006). These data, which are limited in spatial extent and time periods, in turn have their own map generalization and error that propagates into the analysis of coarser resolution pixels. Freely available, GE high-resolution imagery removes some of these barriers, permitting spatial and temporal sampling of land cover variability that strengthens algorithm development and assessment of product accuracy (Bicheron et al., 2008; Helmer et al., 2009). The percent cover reference data collected in our study are flexible for use in other applications. For discrete LULC mapping, different classes can be defined with percent cover rules applied to "end members" (e.g., woody, agriculture, herbaceous). For example, an agriculture–woodland mosaic class could be created by selecting samples that are a mix of agriculture and woody vegetation, with neither cover greater than 80%. The thresholds could also be set to approximate a standardized system, such as the UN Land Cover Classification System (LCCS; Di Gregorio, 2005). Google Earth reference data can be used for calibration and validation of continuous field land cover products over broad areas, such as Vegetation Continuous Fields (VCF, Hansen et al., 2003) or fractional cover from spectral mixture analysis (Roberts et al., 2002; Carreiras et al., 2006).

5.2. LULC classification with the MODIS vegetation index product

The MOD13Q1 product has several advantages for LULC mapping. First, the product includes EVI and NDVI, as well as blue, red, NIR and mid-infrared bands, with a 16-day compositing scheme that helps eliminate cloudy and other unreliable pixels. Although there are alternative compositing techniques for MODIS daily imagery (reviewed in Dennison et al., 2007), MOD13 is favorable because image compositing is done prior to download, greatly reducing data download time, storage costs and processing time. Second, MOD13Q1 VI, red and NIR bands have 250 m resolution, whereas other MODIS products, such as the Nadir BRDF-Adjusted Reflectance (NBAR) based on Aqua and Terra (MCD43A4) have ≥500 m resolution. Although 500 m pixels are generally sufficient for the scale of land use in the Dry Chaco ecoregion, there are other regions where land use has a finer grain and requires 250 m or finer pixels (Hansen et al., 2008).

It is important to consider MODIS pixel georeferencing error, which in combination with GE georeferencing error (Section 5.1), can contribute to spatial mismatch between reference and MODIS data, which in turn, can translate into error in both RF decision rules and map accuracy assessment. Pixel georeferencing error results from the gridding of sensor observations that have inherent geolocation error and tend to overlap and include photons from larger surface areas with increasing view zenith angle (Wolfe et al., 2002; Tan et al., 2006). The current MOD13 C5 compositing algorithm attempts to reduce view-angle effects by selecting the best quality pixel (within highest 10% of NDVI over 16 days) with the smallest view angle (Didan & Huete, 2006). Geolocation of MODIS data has improved since launch, and empirical results estimate error to be on average 18 ± 38 m and 4 ± 40 m in the across-track and along-scan directions, respectively (Wolfe et al., 2002). Since GE samples were 250×250 m, or 18.3 m larger on each side than the gridded MOD13 data (231.7 m), they accommodated some spatial mismatch between the two datasets. However, georeferencing error is

certainly part of the total error observed in our RF decision rules and map accuracy statistics. We expect GE samples with mixed cover to contribute the most error to our analysis, as a spatial shift between the visually interpreted area to the actual area sensed by MODIS could change the relative proportion of classes mixing in the integrated spectral response through time. In contrast, GE samples with 100% cover by a single class, which excludes the *MixWoody* class, were generally from homogeneous patches, and so spatial mismatch with MODIS pixels caused minimal error in RF training and accuracy assessment. In our analysis, this best-case scenario of minimal georeferencing error was with the RF based on all predictor variables, which had an overall accuracy of 86.1% (Table 5—set 5, 100%, no *MixWoody*), with user accuracy ranging from 69.0% (*Herb*) to 100% (*Water*); and thus, the RF still has difficulty separating some classes, likely from spectro-temporal similarity and mislabeled classes in reference samples.

We used two approaches to extract predictive information from the MOD13 time series—TIMESAT and annual statistics. Variables from TIMESAT produced higher accuracies than did EVI annual statistics (Table 5); however, combined EVI, Red and NIR annual statistics yielded higher classification accuracies than TIMESAT. One advantage of TIMESAT is that it produces phenological variables that are straightforward to interpret. These variables could be used for monitoring functional and structural changes in natural vegetation due to forest degradation, climate change, or invasive species. However, a major disadvantage with TIMESAT was its multiple function-fitting parameters (see Section 3.6.2). We subjectively evaluated these parameters for fitting all LULC types, but a rigorous optimization analysis would be time consuming. Another disadvantage with TIMESAT is that samples without clear EVI seasonality (e.g., bare areas, water) are not fit, resulting in null values. This included 3.5% of reference samples in our TIMESAT EVI analysis. Chaco and other dry forests have pronounced seasonality that provides clear trends for TIMESAT function-fitting. In Amazonian seasonally dry forests, EVI and water indices are sensitive to canopy phenological cycles (Huete et al., 2006; Xiao et al., 2005), yet trends are noisy and temporal profiles may have missing data due to cloud cover; and thus, we expect TIMESAT to be less useful in more humid regions. Future research should explore alternative filtering techniques, such as wavelets (Galford et al., 2008), and predictive metrics based on MODIS time series data.

5.3. LULC classification with Random Forests

The Random Forests classifier is a relatively new technique for remote sensing (Chan & Paelinckx, 2008; Gislason et al., 2006; Lawrence et al., 2006, Pal 2005). Some advantages of RF include: few parameters to adjust—terminal node size, number of trees, and number of variables tested at each node; only the most important variables will tend to be used, and so having many correlated or weak predictor variables is not a problem; insight on the classification can be gained by analyzing variable importance; and, classification accuracy estimated by OOB is unbiased (Breiman, 2001; Lawrence et al., 2006). This last property of RF OOB accuracy is a benefit for remote sensing in particular, since reference data are time consuming to collect and are best used in training the classifier rather than being withheld for testing. Our results using an independent dataset indicate that RF OOB accuracy is generally more conservative, i.e., negatively biased (Table 5). For example, the final map overall accuracy was 71.7% and 74.5% with OOB and the original (non-reviewed) test dataset, respectively. However, when an expert reviewed the misclassified samples of our test dataset (Section 3.4), overall accuracy increased to 79.3%, while OOB remained 71.7% because it was based on training data. This result indicates that OOB accuracy statistics will be closer to those from an independent dataset as long as errors are similar, as they were before the secondary expert review of test data. One of the most important features of the RF classifier is that it can handle heterogeneous classes with non-parametric distributions; that is, a tree can arrive at a class through multiple pathways, provided that there are sufficient samples. In our study, an RF tree terminal node had a minimum of 5 samples, and so a class would need ≥ 10 total samples to have two separate terminal nodes—more samples allow for more terminal nodes. This property is particularly important for classes such as agriculture, which includes fallow fields and various crops with varying growing cycles. How many samples are needed per class to achieve a reasonable accuracy? Our experiments show that overall accuracy was only 0.6% less with up to 100 samples per class versus using all 1691 samples. However, there were differences in class user accuracy that should be taken into consideration. Water had >90% user accuracy with as few as 20 samples since it has a distinct low reflectance signal. Most classes had user accuracy that stabilized when the RF had between 60 to 100 samples per class, and only mixed woody and plantations benefited with the addition of all available samples. Plantations benefited the most from adding more than 100 samples per class; however, there were only 83 total plantation samples available, and so improvement in class accuracy was related to the addition of samples from other classes, which reduced commission error.

5.4. Land change in the Dry Chaco

From 2002 to 2006, our maps showed that a net 6.9 million hectares of closed-canopy ($\geq$80% cover), woody vegetation was lost from the Dry Chaco ecoregion. About 5.7 million hectares entered the mixed woody class, which had <80% woody cover mixed with herbaceous vegetation or bare soil, while about 0.7 million hectares transition from mixed woody into woody vegetation. Some of the loss of woody vegetation can be attributed to forest degradation, where forests have trees and shrubs removed as an intermediate step to agriculture or pastures. However, the classifier confused these two woody classes, and so we cannot attribute the increase of mixed woody only to forest degradation, and a large proportion of the *MixWoody–Woody* transition (i.e., the southwest of the study area) should be attributed to subtle and more reversible changes such as increases in grazing pressure and decreases in canopy cover due to changes in rainfall. There was less confusion of closed-canopy woody vegetation with herbaceous vegetation or agriculture; and thus, we considered pixels that transitioned between these classes from 2002 to 2006 to be reliable for mapping deforestation and reforestation. With this logic, there were 2.5 million hectares of woody vegetation (4% of 2002 area) that was deforested from 2002 to 2006. In Argentina, these trends reflect the clearing of Chaco forest for soybean production to meet demand from global markets during this time frame (Gasparri & Grau, 2009; Grau et al., 2005). The maps from this study will form the basis for a future analysis of socio-economic factors that are linked to land use change in this ecoregion.

6. Conclusion

We developed a method to map annual land use/land cover (LULC) at regional scales using source data and processing techniques that permit scaling to broader spatial and temporal scales, while maintaining a consistent classification scheme and accuracy. Our maps have a 250 m pixel size, include 8 general classes, and can be produced annually from year 2001 onward. Using the Dry Chaco ecoregion as a test site, our 2001–2007 maps had an overall accuracy of 79.3%. Herbaceous vegetation was difficult to distinguish from agriculture and mixed woody vegetation with our technique, due to sub-pixel mixing, limited spectral information, georeferencing error and mislabeled reference data. A generalized map of areas with and without closed-canopy, woody vegetation had an overall accuracy of 92.2%. When compared as a multi-temporal sequence, this level of classification provides a land cover change "alarm", similar to Vegetation Cover Conversion (Zhan et al., 2002), that maps recent human disturbance such as deforestation and large disturbances such as fire. Conversely, the annual sequence of LULC maps can track the

year pixels regenerate to woody vegetation, thus providing an estimate of time since recovery, or forest age (Helmer et al., 2009).

Our procedure is transferable to other regions and could provide a relatively inexpensive means to monitor annual LULC at regional to continental scales. First, we implemented our method with free and open-source tools available on many operating systems, which also allows flexibility and scalability in processing data on workstations or servers. Second, our input satellite imagery and reference data are available across political boundaries. We used the free MOD13 MODIS product, which includes the most reliable pixels every 16 days and greatly reduced pixels affected by clouds. Since compositing of hundreds of reflectance scenes is done before download, our data storage and processing costs were minimized. Although MOD13 data do not have many spectral bands, they are temporally rich. This allowed us to calculate predictor variables that describe vegetation phenology and other temporal variation within pixels. The RF classifier was successful at discriminating LULC classes using these temporal variables, without the operator having to select optimal predictor variables or classifier parameters. Random Forests can handle classes with multi-modal distributions, which is particularly important when mapping at broad scales since class variance increases across environmental and anthropogenic gradients.

We relied on visual interpretation of high-resolution imagery in Google Earth for collecting reference data. This approach allows statistically rigorous sampling across broad spatial scales, with limits set more by time and labor costs than data availability. We found that available images covered all but one year in our study, and the ability to view historical images in Google Earth increases sampling possibilities in the spatial and temporal domains. Google Earth allows the remote sensing community to sample more area, with arguably more accuracy, than relying on medium resolution images or classified maps (e.g., from Landsat) for training and validation data. There is great potential in deploying Google Earth as part of a Web-based LULC observation system, whereby expert collaborators or facilitated volunteers (e.g., "crowd-sourcing") interpret imagery remotely via the Internet (Fritz et al., 2009). However, there are caveats to using these virtual reference data in other regions. Our results based on field data indicate that samples are well georeferenced, but there is class interpretation error, especially in distinguishing herbaceous vegetation and agriculture. Given the wide range of global LULC types and differences in human perception of these classes, it is important to control class error by establishing interpretation rules and a training protocol, with understanding and accuracy assessments based on field observations. In the future, we expect that new Web-based geospatial technology, such as Google Earth used in this study, will promote greater international participation in developing and validating remote sensing products that span spatial and temporal scales.

Acknowledgements

This research was funded by a Dynamics of Coupled Natural and Human Systems grant from the U.S. National Science Foundation (NSF 0709598) and from Fondo paa la Investigación Científica y Tecnológica, Argentina (FONCyT, PICT 2006-1693). We thank student workers at the Sonoma State University and University of Puerto Rico who helped interpret imagery in Google Earth. This work was greatly improved by four anonymous reviewers.

References

Achard, F., Eva, H., Stibig, H. J., Mayaux, P., Gallego, J., Richards, T., et al. (2002). Determination of deforestation rates of the world's humid tropical forests. *Science*, *297*, 999−1002.

Altritcher, M., & Boagio, G. (2004). Distribution and relative abundance of peccaries in the Argentine Chaco: associations with human factors. *Biological Conservation*, *116*, 217−225.

Altritcher, M., Boaglio, G., & Perovic, P. (2006). The decline of jaguars, *Panthera onca*, in the Argentine Chaco. *Oryx*, *40*, 302−309.

Alves, D. S., & Skole, D. L. (1996). Characterizing land cover dynamics using multi-temporal imagery. *International Journal of Remote Sensing*, *17*, 835−839.

Asner, G. P. (2001). Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*, *22*, 3855−3862.

Baccini, A., Friedl, M. A., Woodcock, C. E., & Warbington, R. (2004). Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters*, *31*, L10501.

Bartholomé, E., & Belward, A. S. (2005). GLC2000: A new approach to global land cover mapping from Earth Observation data. *International Journal of Remote Sensing*, *26*, 1959−1977.

Bicheron, P., Defourny, P., et al. (2008). GlobCover: products description and validation report, ESA GlobCover project. http://www.ionia1.esrin.esa.int/images/GLOBCOV-ER_Product_Specification_v2.pdf

Boletta, P. E., Ravello, A. C., Planchuela, A. M., & Grillo, M. (2006). Assessing deforestation in Argentine Chaco. *Forest Ecology and Management*, *228*, 108−114.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5−32.

Carreiras, J. M. B., Pereira, J. M. C., & Shimabukuro, Y. E. (2006). Land-cover mapping in the Brazilian Amazon using SPOT-4 vegetation data and machine learning classification methods. *Photogrammetric Engineering & Remote Sensing*, *72*, 897−910.

Chan, J. C. -W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, *112*, 2999−3011.

Congalton, R. G., & Mead, R. A. (1983). A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*, *49*, 69−74.

DeFries, R. S., & Belward, A. S. (2000). Global and regional land cover characterization from satellite data: an introduction to the Special Issue. *International Journal of Remote Sensing*, *21*, 1083−1092.

DeFries, R. S., Hansen, M. C., Townshend, J. R. G., & Sohlberg, R. S. (1998). Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing.*, *19*, 3141−3168.

DeFries, R., Achard, F., Brown, S., Herold, M., Murdiyarso, D., Schlamadinger, B., et al. (2007). Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. *Environmental Science and Policy*, *10*, 385−394.

Dennison, P. E., Roberts, D. A., & Peterson, S. H. (2007). Spectral shape-based temporal compositing algorithms for MODIS surface reflectance data. *Remote Sensing of Environment*, *109*, 510−522.

Di Gregorio, A. (2005). *Land Cover Classification System−−classification concepts and user manual for Software version 2.* FAO Environment and Natural Resources Service Series, No. 8, Rome 208 pp.

Didan, K., & Huete, A. (2006). *MODIS Vegetation Index Product Series Collection 5 Change Summary.*Terrestrial Biophysics and Remote Sensing (TBRS) laboratory, The University of Arizona June 29, 2006.

Eva, H. D., Belward, A. S., De Miranda, E. E., Di Bella, C. M., Gond, V., Huber, O., et al. (2004). A land cover map of South America. *Global Change Biology*, *10*, 731−744.

Falkowski, M. J., Evans, J. S., Martinuzzi, S., Gessler, P. E., & Hudak, A. T. (2009). Characterizing forest succession with Lidar data: An evaluation for the inland Northwest USA. *Remote Sensing of Environment*, *113*, 946−956.

Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, *61*, 399−409.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, *83*, 287−302.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., et al. (2010). MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, *114*, 168−182.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., et al. (2009). Geo-Wiki. Org: The use of crowd sourcing to improve global land cover. *Remote Sensing*, *1*, 345−354.

Galford, G. L., Mustard, J. F., Melillo, J., Gendrin, A., Cerri, C. C., & Cerri, C. E. P. (2008). Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sensing of Environment*, *112*, 576−587.

Gasparri, N. I., & Grau, H. R. (2009). Deforestation and fragmentation of Chaco dry forests in NW Argentina (1972–2007). *Forest Ecology and Management*, *258*, 913−921.

Gasparri, N. I., Grau, H. R., & Menghi, E. (2008). Carbon pools and emissions from deforestation in extra-tropical forests of northern Argentina from 1900 to 2005. *Ecosystems*, *11*, 1247-121.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, *27*, 294−300.

Grau, H. R., Aide, T. M., & Gasparri, N. I. (2005). Globalization and soybean expansion into semiarid ecosystems of Argentina. *Ambio*, *34*, 267−268.

Hansen, M. C., Defries, R. S., et al. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, *216–7*, 1331−1364.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., & Sohlberg, R. A. (2003). Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm.*Earth Interactions*, *7*(10) 15 pp.

Hansen, M. C., Stehman, S. V., Potapov, P. V., Loveland, T. R., Townshend, J. R. G., DeFries, R. S., et al. (2008). Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. *Proceedings of the National Academy of Science*, *105*, 9439−9444.

Helmer, E. H., Lefsky, M. A., & Roberts, D. A. (2009). Biomass accumulation rates of Amazonian secondary forest and biomass of old-growth forests from Landsat time series and the Geoscience Laser Altimeter System. *Journal of Applied Remote Sensing*, *3*, 033505.

Herold, M., Mayaux, P., Woodcock, C. E., Baccini, A., & Schmullius, C. (2008). Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, *112*, 2538−2556.

Hoekstra, J. M., Boucher, T. M., Ricketts, T. H., & Roberts, C. (2005). Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology Letters*, *8*, 23−29.

Houghton, R. A. (2005). Aboveground forest biomass and the global carbon balance. *Global Change Biology*, *11*, 945−958.

Huang, C., Kim, S., Song, K., Townshend, J. R. G., Davis, P., Altstatt, A., et al. (2009). Assessment of Paraguay's forest cover change using Landsat observations. *Global and Planetary Change*, *67*, 1−12.

Hudak, A. T., Crookston, N. L., Evans, J. S., Hall, D. E., & Falkowski, M. J. (2008). Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment*, *112*, 2232−2245.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195−213.

Huete, A. R., Didan, K., Shimabukuro, Y. E., Ratana, P., Saleska, S. R., Hutyra, L. R., et al. (2006). Amazon rainforests green-up with sunlight in dry season. *Geophysical Research Letters*, *33*, L06405.

Jönsson, P., & Eklundh, L. (2004). TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, *30*, 833−845.

Killeen, T. J., Calderon, V., Soria, L., Quezada, B., Steininger, M. K., Harper, G., et al. (2007). Thirty years of land-cover change in Bolivia. *Ambio*, *36*(7), 600−606.

Korpela, I., Koskinen, M., Vasander, H., Holopainen, M., & Minkkinen, K. (2009). Airborne small-footprint discrete-return LiDAR data in the assessment of boreal mire surface patterns, vegetation, and habitats. *Forest Ecology and Management*, *258*, 1549−1566.

Lawrence, R. L., Wood, S. D., & Sheley, R. L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment*, *100*, 356−362.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2* (3), 18−22.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., et al. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, *216–7*, 1303−1330.

Martinuzzi, S., Vierling, L. A., Gould, W. A., Falkowski, M. J., Evans, J. S., Hudak, A. T., et al. (2009). Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sensing of Environment*, *113*, 2533−2546.

Minetti, J. L. (1999). *Atlas climático del noroeste argentino.* Fundación Carl C: Zon Caldenius, Sede NOAA. Tucumán: Universidad Nacional de Tucumán.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., et al. (2001). Terrestrial ecoregions of the world: a new map of life on earth. *Bioscience*, *51*, 933−938.

Pal, M. (2005). Random Forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*, 217−222.

Powell, R. L., Matzke, N., de Souza, C., Jr., Clark, M. L., Numata, I., Hess, L. L., et al. (2004). Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. *Remote Sensing of Environment*, *9*, 221−234.

R Development Core Team (2010). *R: A language and environment for statistical computing.*Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0 http://www.R-project.org

Redford, K., Taber, A., & Simonetti, J. (1990). There is more to biodiversity than the tropical rain forest. *Conservation Biology*, *4*, 328−330.

Roberts, D. A., Numata, I., Holmes, K., Batista, G., Krug, T., Moteiro, A., et al. (2002). Large area mapping of land-cover change in Rondônia using multitemporal spectral mixture analysis and decision tree classifiers.*Journal of Geophysical Research-Atmospheres*, *107*, 8073 (LBA 40-1-40-18).

Sesnie, S. E., Gessler, P. E., Finegan, B., & Thessler, S. (2008). Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sensing of Environment*, *112*, 2145−2159.

Skole, D., Justice, C. O., Janetos, A. C., & Townshend, J. (1997). A land cover change monitoring program: a strategy for international effort. *Mitigation and Adaptation Stategies for Global Change*, *2*, 157−175.

Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment*, *64*, 331−344.

Steininger, M. K., Tucker, C. J., Townshend, J. R. G., Killeen, T. J., Desch, A., Bell, V., et al. (2001). Tropical deforestation in the Bolivian Amazon. *Environmental Conservation*, *28*, 127−134.

Tan, B., Woodcock, C. E., Hu, J., Zhang, P., Ozdogan, M., Huang, D., et al. (2006). The impact of gridding artifacts on the local spatial properties of MODIS data: implications for validation, compositing, and band-to-band registration across resolutions. *Remote Sensing of Environment*, *105*, 98−114.

Wolfe, R. E., Nishihama, M., Fleig, A. J., Kuyper, J. A., Roy, D. P., Storey, J. C., et al. (2002). Achieving sub-pixel geolocation accuracy in support of MODIS land science. *Remote Sensing of Environment*, *83*, 31−49.

Xiao, X., Zhang, Q., Saleska, S., Hutyra, L., de Camargo, P., Wofsy, S., et al. (2005). Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest. *Remote Sensing of Environment*, *94*, 105−122.

Zak, M., Cabido, M., & Hodgson, J. G. (2004). Do subtropical seasonal forests in the Gran Chaco, Argentina, have a future? *Biological Conservation*, *120*, 589−598.

Zhan, X., Sohlberg, R. A., Townshend, J. R. G., DiMiceli, C., Carroll, M. L., Eastman, J. C., Hansen, M. C., & DeFries, R. C. (2002). Detection of land cover changes using MODIS 250 m data. *Remote Sensing of Environment*, *83*, 336−350.