



Validation of the USGS Landsat Burned Area Essential Climate Variable (BAECV) across the conterminous United States



Melanie K. Vanderhoof*, Nicole Fairaux, Yen-Ju G. Beal, Todd J. Hawbaker

U.S. Geological Survey, Geosciences and Environmental Change Science Center, P.O. Box 25046, DFC, MS980, Denver, CO 80225, United States

ARTICLE INFO

Article history:

Received 2 November 2016
Received in revised form 6 June 2017
Accepted 25 June 2017
Available online 3 July 2017

Keywords:

Burned area
Essential climate variable
Fire
Landsat
Validation
Wildfire

ABSTRACT

The Landsat Burned Area Essential Climate Variable (BAECV), developed by the U.S. Geological Survey (USGS), capitalizes on the long temporal availability of Landsat imagery to identify burned areas across the conterminous United States (CONUS) (1984–2015). Adequate validation of such products is critical for their proper usage and interpretation. Validation of coarse-resolution products often relies on independent data derived from moderate-resolution sensors (e.g., Landsat). Validation of Landsat products, in turn, is challenging because there is no corresponding source of high-resolution, multispectral imagery that has been systematically collected in space and time over the entire temporal extent of the Landsat archive. Because of this, comparison between high-resolution images and Landsat science products can help increase user's confidence in the Landsat science products, but may not, alone, be adequate. In this paper, we demonstrate an approach to systematically validate the Landsat-derived BAECV product. Burned area extent was mapped for Landsat image pairs using a manually trained semi-automated algorithm that was manually edited across 28 path/rows and five different years (1988, 1993, 1998, 2003, 2008). Three datasets were independently developed by three analysts and the datasets were integrated on a pixel by pixel basis in which at least one to all three analysts were required to agree a pixel was burned. We found that errors within our Landsat reference dataset could be minimized by using the rendition of the dataset in which pixels were mapped as burned if at least two of the three analysts agreed. BAECV errors of omission and commission for the detection of burned pixels averaged 42% and 33%, respectively for CONUS across all five validation years. Errors of omission and commission were lowest across the western CONUS, for example in the shrub and scrublands of the Arid West (31% and 24%, respectively), and highest in the grasslands and agricultural lands of the Great Plains in central CONUS (62% and 57%, respectively). The BAECV product detected most (>65%) fire events >10 ha across the western CONUS (Arid and Mountain West ecoregions). Our approach and results demonstrate that a thorough validation of Landsat science products can be completed with independent Landsat-derived reference data, but could be strengthened by the use of complementary sources of high-resolution data.

Published by Elsevier Inc.

1. Introduction

Accurate mapping of the extent and timing of burned area is critical to quantifying and modeling greenhouse gas emissions (Crutzen and Andreae, 1990; Palacios-Orueta et al., 2005; Randerson et al., 2005), carbon and nutrient cycling (Conard et al., 2002; Bond-Lamberty et al., 2007), and changes to ecosystem structure (Thonicke et al., 2001; Goetz et al., 2005). Consequently, fire disturbance has been identified by the Global Climate Observing System (GCOS) program as one of the high priority Essential Climate Variables (ECV) (Global Climate Observing System, 2004) and major efforts have been undertaken to produce global burned area products (Mouillot et al., 2014). The

products developed to-date use coarse-scale satellite imagery (300 m to 1 km) (e.g., Moderate Resolution Imaging Spectrometer (MODIS) burned area product (MCD45, MCD64), Geoland2, fire_cci burned area (BA)). Such datasets provide information critical for climate modeling and are effective for capturing global fire patterns at a high temporal frequency, but may be limited in their ability to map fire heterogeneity, detect small fires (Stroppiana et al., 2012) or provide enough historical context, necessary to discern temporal trends (Mouillot et al., 2014) and relationships with climate and other drivers (Podur et al., 2002; Miller et al., 2009; Whitman et al., 2015). In addition, because of the tremendous amount of spectral diversity in the signal of burned areas across diverse vegetation types, fire combustion levels (e.g., ash, char, soot), and burn severities (e.g., ground vs crown fires), the accuracy of existing global burned area products is relatively low with documented errors of omission and commission for burned areas ranging from 51%

* Corresponding author.

E-mail address: mvanderhoof@usgs.gov (M.K. Vanderhoof).

to 93%, and 42% to 94%, respectively (Padilla et al., 2014a, 2015; Chuvieco et al., 2016).

This study presents a validation of the Landsat Burned Area Essential Climate Variable (BAECV) product across a sample set of locations and times using an independently derived reference dataset. The BAECV, developed by the U.S. Geological Survey (USGS), aims to capitalize on the long time period covered by Landsat imagery to provide wall-to-wall maps of burned areas across the conterminous United States (CONUS) (1984–2015), and could be extended to other regions with appropriate training data (Hawbaker et al., 2017). The product will be provided as a wall-to-wall raster of burned area across CONUS at 30 m resolution and an annual time-step, with a minimum fire size of 4.05 ha (45 pixels) (<https://www.sciencebase.gov/catalog/item/57867943e4b0e02680c14fec>). Landsat sensors can provide a longer temporal record of burned area relative to existing global burned area products and potentially with increased accuracy and detail (Stroppiana et al., 2012). Landsat has been used extensively to map burned areas, predominantly for local and regional studies (Mitri and Gitas, 2004; Bastarrika et al., 2011; Petropoulos et al., 2011; Mallinis and Koutsias, 2012). In recent years Landsat has been used to map fire and other disturbance types across portions of CONUS but for limited years (Masek et al., 2008; Boschetti et al., 2015) and/or Landsat path/rows (Cohen et al., 2010; Kennedy et al., 2010; Thomas et al., 2011). These efforts have been largely restricted to forest cover and fire is often not distinguished from other disturbance types (e.g., harvest, insect) (Goward et al., 2016). The most comparable effort to date is the Monitoring Trends in Burn Severity (MTBS) product which is also derived from Landsat (Eidenshink et al., 2007). The BAECV differs from the MTBS dataset in several important ways. Because the BAECV product generation is automated, the BAECV can potentially provide a more complete census of burned areas, relative to the MTBS dataset which relies on manually mapping reported large fires ($\geq 2 \text{ km}^2$ in the eastern U.S. and $\geq 4 \text{ km}^2$ in the western U.S.) (Eidenshink et al., 2007). The BAECV utilizes all available Landsat images, in contrast the MTBS effort began prior to the Landsat archive becoming freely available in 2008, which required them to be strategic in their image selection for the earlier years of the dataset. In addition, the MTBS made a conscious decision to provide limited mapping of prescribed fires, common in the southeastern United States due to the sheer number of such fires. Despite the advantages, automation can be expected to introduce errors in burned area extent (e.g., missing fires, over-mapping fires, or disagreeing on fire extent), necessitating an independent validation of the BAECV product.

Validation of burned area products and the provision of accuracy statistics to users is essential to allow users to decide when and how to utilize datasets, correctly interpret results, and provide feedback to improve products (Roy et al., 2005; Morisette et al., 2006). The Committee on Earth Observation Satellites (CEOS), Land Product Validation Subgroup (LPVS), formed in 2000, has specified that validation is a critical component in the generation of ECV products, and should follow internationally agreed upon validation best practices to measure accuracy, precision (standard error of accuracy estimates), and temporal stability at comprehensive spatial and temporal scales (Morisette et al., 2006). Validations typically produce pixel-level or point-level error matrices, derived by cross-tabulating ECV products with independent reference maps (Bastarrika et al., 2011; Stroppiana et al., 2012; Padilla et al., 2014a, 2015). Linear regression analysis has also been used to compare the proportion of burned area defined by the product, relative to reference maps (Roy et al., 2008; Roy and Boschetti, 2009). Comparisons between global burned area products have also examined differences in the spatial and temporal distribution of burned area and calculated patch indices to explore a product's ability to map small fires (Chuvieco et al., 2016).

The source of reference datasets varies by study and product. Fire perimeter datasets, such as the U.S. Geospatial Multi-Agency Coordination (GeoMAC) dataset, tend to either focus on large fires or are designed to

meet the needs of fire managers and do not provide a complete census of all fires (Eidenshink et al., 2007; Walters et al., 2011). This design makes perimeter datasets good references for large, single fire events (Mitri and Gitas, 2004; Henry, 2008; Bastarrika et al., 2011), but insufficient for a validation at a national or global scale. Because of the limitations of fire perimeter datasets, burned area maps derived from remotely sensed imagery are typically validated using a reference map derived from a finer resolution source of imagery (Roy and Boschetti, 2009; Mallinis and Koutsias, 2012; Padilla et al., 2014a). For coarse-resolution products this is non-problematic, as moderate-resolution sources of imagery, collected at regular intervals, are widely available (e.g., Landsat and ASTER) (Roy and Boschetti, 2009).

The utilization of high-resolution imagery (e.g., IKONOS, Quickbird-2, Geoeye-1, Worldview-2, 3) to validate a national or global burned area product, however, faces several challenges. High-resolution imagery has been successfully utilized to detect burned areas (Mitri and Gitas, 2006; Holden et al., 2010; Mallinis and Koutsias, 2012). Yet, as the satellites typically collect imagery on demand, the coincidence of images collected over burned patches, prior to vegetation recovery, is sporadic making it challenging to defend a sampling strategy and requiring classification of burned area from a single image instead of a pre- and post-fire image pair. In addition, these satellites have only been in orbit since late 1990s or early 2000s, meaning they can only be used to validate a portion of the temporal extent of a Landsat science product. These satellites also typically lack short-wave infrared (SWIR) bands, which have been found to be useful in detecting burned areas (Chuvieco, 1997). The spatial scale at which fire events occur should also be considered relative to the reference data. The small extent of high-resolution images (13 to 18 km across), relative to Landsat image extents (185 km across) means only portions of larger fires are often contained within high-resolution imagery, limiting the number of fire events being validated.

Instead of using high-resolution imagery, validation of Landsat disturbance products to date have typically relied on the derivation of independent datasets from Landsat images, complemented by high-resolution imagery, as available (Thomas et al., 2011; Masek et al., 2013). Burned patches are often visually distinct, but have high spectral diversity resulting from variability in soil type, pre-fire vegetation cover, fire severity and time since fire, and that can make it challenging to detect burned areas across diverse environments in an automated manner (Bastarrika et al., 2011). Therefore, forest disturbance events are identified through the visual examination of pre- and post- Landsat images by experienced image analysts (Masek et al., 2008; Huang et al., 2009; Stroppiana et al., 2012). Although including a manual component in imagery analysis is a common practice to improve the quality of reference datasets (e.g., Mitri and Gitas, 2004; Henry, 2008; Petropoulos et al., 2011), observer-dependent variability has also been documented, although not explicitly for mapping burned areas (Mazz, 1996; Baveye et al., 2010). Using multiple observers is one technique that has been used to reduce errors of omission in other areas of science and image analysis, but is not widely done (Mazz, 1996; Nichols et al., 2000).

Thorough validation of remote sensing products is essential prior to their acceptance by the scientific community, proper use, and integration into management and modeling activities. This study seeks to validate USGS's Landsat BAECV (1984–2015) using an independent dataset derived from Landsat across a sample of 28 Thiessen scene areas and five years, complemented by high-resolution imagery. Our research questions included:

- (1). How does the subjectivity of visual image interpretation affect the quality of the reference dataset and influence accuracy statistics?
- (2). What is the accuracy of USGS's BAECV across diverse land cover types and regions of the conterminous U.S. (CONUS)?
- (3). How stable are the accuracy statistics through time?
- (4). How does burn size influence the accuracy of the BAECV product?

2. Methods

2.1. BAECV product algorithm

The BAECV algorithm is explained in detail in Hawbaker et al. (2017). The algorithm is a supervised machine learning approach and relied heavily on the MTBS data as the primary source of observed burned area to train and evaluate the algorithm. Of the 32 years of Landsat 4, 5, and 7 data (1984–2015), MTBS data were not available for 2014 or 2015 at the time of the product generation. Of the remaining 30 years for which MTBS data were available, 24 of these years were used to derive training data points, and the remaining 6 years (1988, 1993, 1998, 2003, 2008, and 2013) were retained for testing and validation. The BAECV algorithm uses a suite of predictor variables calculated from dense time series of Landsat data including both single-scene, pre-fire surface conditions (e.g., 3-year lagged means and standard deviations), and change from pre-fire surface conditions. These variables were used as the inputs to train a generalized boosted regression model (Hastie et al., 2009) which uses a sequence of classification and regression tree models (Breiman et al., 1984) to predict the probability that a pixel has burned in any given Landsat image. Burn classification images were generated by applying thresholds and a region-growing method to the burn probability images so that pixels with very high burn probability ($\geq 98\%$) and adjacent pixels with high burn probability ($\geq 95\%$) were also classified as burned. Burned area patches < 4.05 ha (45 pixels) were removed.

2.2. Validation sampling design

The fire_cci project, part of the European Space Agency's (ESA) Climate Change Initiative (CCI) generated a global reference dataset for a burned area product that met the CEOS LPVS stage 3 validation requirements. They used a stratified sampling scheme to select 102 global

Thiessen scene areas (TSAs), constructed by Cohen et al. (2010) from Landsat World Reference System II (WRS-II), across diverse biomes, fire regimes, and climate conditions (Padilla et al., 2014a). The TSAs provided non-overlapping partitioning of the study region. The fire_cci's global reference dataset was limited temporally and only included 2008 as the standard year for validation of all fire_cci products (Padilla et al., 2014a, 2015).

Our sample design to validate the BAECV was implemented to be complementary to the fire_cci project. We used stratified, random, one-stage cluster sampling, where each cluster was defined by a Thiessen scene area (Stehman, 2009). Using this sampling design, we augmented the existing fire_cci validation TSAs ($n = 9$) within CONUS with an additional 19 TSAs for a total of 28 TSAs (Fig. 1). The TSAs were stratified across the major Olson biomes (Olson et al., 2001) (Fig. 1). We used the same, simplified versions of Olson's biomes as used by the fire_cci project. Within CONUS, this included (1) temperate forest, (2) Mediterranean forest, (3) temperate grassland and savannah, (4) tropical and subtropical grasslands and savannah, and (5) xeric/desert shrub (Padilla et al., 2014a). TSAs with a burn area extent above the 80th percentile in 2008 within a given biome were defined as high burned area stratum. Burn area extent was defined using the Global Fire Emissions Database (GFED) version 3 (Giglio et al., 2009, 2010). We sampled disproportionately in the high burned area stratum to ensure that enough burned area was mapped within the reference dataset. We acknowledge that if a path/row was identified as a high burned area stratum in 2008 it may not have been classified similarly in the other sample years, however using consistent path/rows between years allowed for the reference dataset to contain variability in burned area extent between years and allowed us to test for temporal stability in accuracy statistics between years. The final number of TSAs included in the sample and within CONUS is shown in Table 1. We similarly used 2008, and additional years were included to represent 5-year increments and help assess the temporal stability of the BAECV products

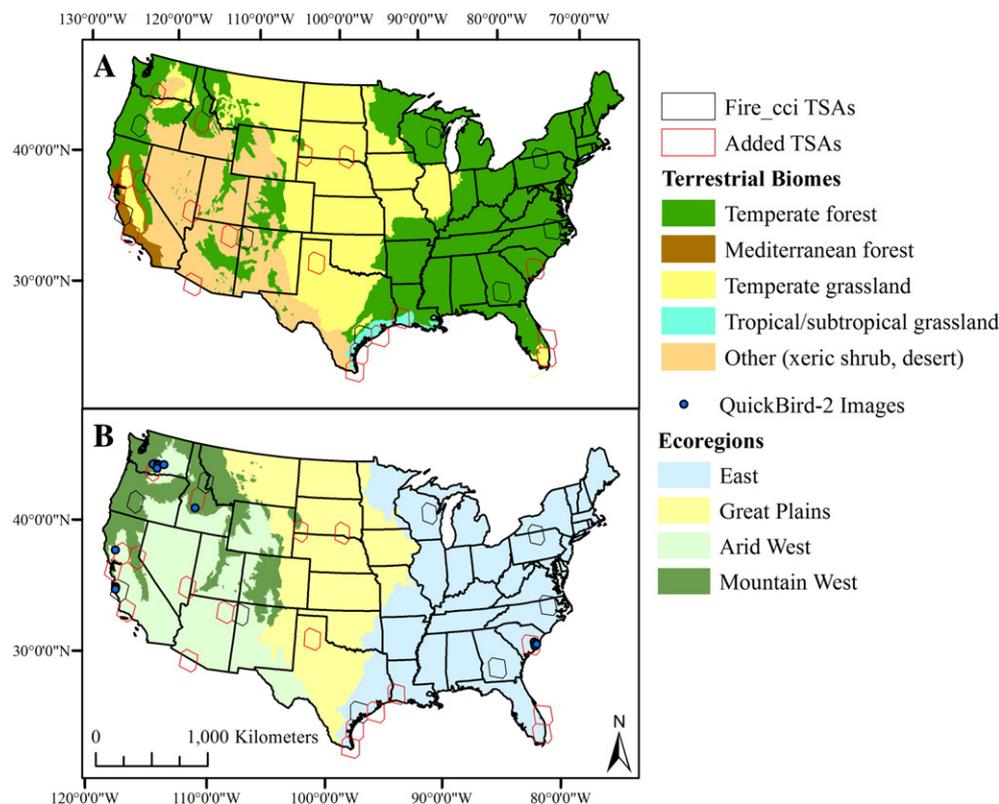


Fig. 1. A) The spatial distribution of the validation Thiessen scene areas (TSAs) based on Landsat path/rows in relation to Olson terrestrial biomes, and B) the validation TSAs and high-resolution image locations in relation to the U.S. Environmental Protection Agency's Level I ecoregions.

Table 1
Distribution of sampled and population Thiessen scene areas (TSAs) by biome and stratum. Each sampled TSA was then sampled for 5 separate years, however high/low BA stratum was determined from 2008, alone. Total number is calculated for the conterminous United States (CONUS). BA: burned area.

Biome	Total TSAs sampled	High BA stratum sampled	Low BA stratum sampled	Total number of TSAs	Total high BA stratum	Total low BA stratum
Temperate forest	11	6	5	224	45	179
Mediterranean forest	3	2	1	12	2	10
Temperate grassland and savanna	5	2	3	124	25	99
Tropical & subtropical savanna	4	2	2	7	2	5
Xeric/desert shrub	5	3	2	83	17	66
CONUS	28	15	13	450	91	359

(2008, 2003, 1998, 1993 and 1988). The specific Landsat path/rows and years used to validate were specifically avoided when training the BAECV, ensuring the independence of the pixels used in the accuracy assessment.

2.3. Landsat-resolution reference dataset

Reference data were generated from pre- and post-fire cloud-free (<20% cloud cover) Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) images for each of the 28 validation path/rows and 5 validation years. Images were limited to those with a RMSE <10 m and available as a L1T Surface Reflectance product. The dates of the 140 pre- and post-fire image pairs (a total of 280 Landsat images) were selected to maximize the amount of burned area mapped and are listed in the (Appendix Table A1). The pre- and post-image pairs do not specifically represent a probability sample within a year, but were designed to target changes incurred over the peak fire season. Peak fire season was determined using the distribution of total burned area by month as derived from the MODIS burned area product (MCD45, 2001–2015). The distribution of the image pairs in relation to peak fire season by U.S. ecoregion is shown in Table 2. The time gap between the pre- and post-fire images averaged 90 days, but ranged from 16 to 319 days. “New” burned area (post-fire – pre-fire) maps were generated using the Burned Area Mapping Software (BAMS), which is a semi-automated algorithm developed by the University of Alcalá, Madrid (Bastarrika et al., 2014; Padilla et al., 2014a). The algorithm was trained on manually selected polygons containing (1) clearly burned pixels and (2) spectrally similar but less distinct burned pixels (Bastarrika et al., 2014). Four vegetation indices were calculated for the pre- and post-fire images, and these were utilized in a supervised classification. Indices used included the Normalized burn ratio (NBR, García and Caselles, 1991; Key and Benson, 2006), the Mid-infrared Burned Index (MIRBI, Trigg and Flasse, 2001), the Global Environmental Monitoring Index (GEMI, Pinty and Verstraete, 1992) and the Normalized Difference Vegetation Index (NDVI, Tucker, 1979). The algorithm applied a region-growing function between the two types of training polygons, while cut-off values for each variable were extracted from the training polygons. Additional details are provided in Bastarrika et al. (2011, 2014).

The output maps were manually edited. When available, the analysts utilized ancillary datasets (e.g., MTBS, MODIS active fire points, MODIS burned area, aerial imagery) to improve confidence in their selection of training pixels and manual edits. Manual edits were completed for almost every image and primarily consisted of removing fire commission errors. Common sources of error that were corrected for included non-fire related change within agricultural and wetland cover types, drought related vegetation mortality, clear-cutting events, and occasional spectral confusion with bare rock and open water. The FMask from the Landsat surface reflectance product was applied to mask out clouds, cloud shadows, snow and open water (Zhu and Woodcock, 2014). For Landsat 7 ETM+ images SLC off pixels were masked out. The low-, medium- and high-intensity development classes (i.e., urban areas) were masked out using NLCD data (Homer et al., 2015) to reduce spectral confusion between burned areas and impervious surfaces. Fires with patch sizes of <4.05 ha (45 pixels) were also removed from the reference data to be comparable with the minimum mapping unit of the BAECV product.

Due to the substantial subjectivity involved in visual image interpretation, three different analysts individually generated a burned area map for each validation TSA and year. From these outputs, three renditions of the reference burned area maps were generated, in which burned area extent ranged from liberal (or inclusive) to conservative (exclusive). Burned area extent was defined as 1) at least one analyst agreed a given pixel was burned (1 agreed) (inclusive), 2) at least two of the three analysts were required to agree a given pixel was burned (2 agreed), 3) all three analysts were required to agree a pixel was burned (3 agreed) (exclusive).

2.4. High-resolution reference dataset

The reference burned area map used to validate the BAECV was selected by comparing all three renditions of the reference dataset to high resolution imagery. The criteria for the QuickBird-2 imagery (2 m resolution) acquired from Digital Globe was that it 1) contained newly burned areas and 2) was located within the validation TSAs and years. QuickBird-2 imagery is collected on demand; images with burned areas are largely incidental, limiting the number of images found. Sixteen QuickBird-2 images were used, representing four states (California,

Table 2
The distribution of the image pairs (pre- and post-Landsat images) in relation to the peak fire season for each ecoregion. BA: burned area.

Ecoregion	Peak fire season	BA occurring in peak season (%)	Pairs occurred in non-peak season (%)	Pairs overlapped peak season (%)	Pairs spanned entire peak season(s) (%)	Pairs overlapped or spanned entire peak season(s) (%)
Arid West	June–August	70	20	42.5	37.5	80
Mountain West	July–September	91	0	56	44	100
Great Plains	March–April and Sep–Oct	39 and 37 (76 total)	0	100	0	100
East	Sept–Oct and Nov–May	48 and 50 (98 total)	0	93	7	100

Table 3

The QuickBird-2 imagery used to compare the three renditions of the reference dataset. All Landsat images were from the Thematic mapper sensor. SC: South Carolina. ID: Idaho. CA: California. WA: Washington.

State	Landsat path/row	Number of quickbird-2 images	quickbird-2 date	Landsat pre-image date	Landsat post-image date
SC	p16r37	3	21-Apr-08	10-Feb-08	1-May-08
SC	p16r37	2	8-Jun-08	10-Feb-08	1-May-08
ID	p41r29	1	13-Aug-03	20-Jul-03	24-Oct-03
CA	p43r35	3	13-Nov-08	14-May-08	19-Sep-08
CA	p44r33	1	18-Nov-08	6-Jun-08	12-Oct-08
WA	p45r28	3	28-Sep-03	14-Jun-03	3-Sep-03
WA	p45r28	1	25-Sep-03	14-Jun-03	3-Sep-03
WA	p45r28	1	2-Sep-03	14-Jun-03	3-Sep-03
WA	p45r28	1	28-Aug-03	14-Jun-03	3-Sep-03
Total:	5	16 images			
	path/rows				

Washington, Idaho and South Carolina) and five of the 28 path/rows (p16r37, p41r29, p43r35, p44r33, p45r28) (Table 3, Fig. 1). The QuickBird-2 imagery was atmospherically corrected and converted to ground reflectance using ATCOR (the Atmospheric Correction module) in PCI Geomatica (Richter and Schlöpfer, 2016). Burned area was identified using maximum likelihood supervised classification with null values, in which each image was trained on manually selected “burned” and “unburned” polygons. A sieve filter was applied to reduce noise in the output images with the filter size varying based on the amount of noise produced by each classification. Burned area outputs were further manually edited in ArcGIS 10.3 using visual interpretation of the QuickBird-2 imagery as well as ancillary datasets (e.g., MTBS, MODIS fire points) as relevant. Image processing was completed using PCI Geomatica.

2.5. Comparison of the Landsat reference dataset versions

Because of the small number of high-resolution images used and the often larger than ideal date gap between the post-fire Landsat images and the high-resolution images, this analysis should not be interpreted as a validation of the reference dataset, but instead a means to compare the three version of the reference dataset. To determine which rendition of the reference dataset to use, the reference dataset (30 m resolution) was up-sampled to match the QuickBird-2 data (2 m resolution) for a pixel to pixel comparison (Padilla et al., 2014a). Metrics presented included overall accuracy, omission error, commission error, dice coefficient, and relative bias. Omission and commission errors were calculated for the category “burned” (Roy and Boschetti, 2009; Padilla et al., 2014a). The dice coefficient is the conditional probability that if one classifier (product or reference data) identifies a pixel as burned, the other one will as well, and therefore integrates omission and commission errors (Fleiss, 1981; Forbes, 1995). The relative bias provides the proportion that burned area is under (negative) or overestimated (positive) relative to the total burned area of the reference product (Padilla et al., 2014a). We note that the burned area within the Landsat reference dataset is calculated as change between post- and pre-fire images, while the high-resolution imagery is assumed to be post-fire, a potential inconsistency between the datasets. In addition, burned areas for which we observed disagreement between the Reference burned area and the high-resolution burned area (i.e., one mapped a fire and other did not, not disagreement regarding mapping within fire heterogeneity) we used ancillary datasets (MODIS fire points and MTBS) to determine the best known date of the fire. If the differences in image collection date in relation to the date of the fire was found to be the cause of the disagreement, then the burned area was masked out and not included in the comparison.

2.6. Validation of the BAECV

The reference data were designed to map all burned patches that occurred between the two selected Landsat scenes (pre-fire and post-fire). Even though the BAECV algorithm examines each individual Landsat image, its final products are annual composites (Hawbaker et al., 2017). To create comparable reference and BAECV datasets, the BAECV product was modified to apply the burn classification step to the pre-fire and post-fire scenes used to create the reference data and then making sure that only “new” burned area that occurred between the two scenes was assessed. The number of burned and unburned pixels included in the validation, as defined by the reference dataset, are provided in the (Appendix Table A2). Accuracy metrics reported included overall accuracy, omission error, commission error, dice coefficient, and relative bias and were reported from pixel-level summaries. Omission and commission errors were calculated for the category “burned” as calculated by Padilla et al. (2014a). The dice coefficient and relative bias metrics are described above in section 2.5. To account for the influence of stratification and clustering (Stehman, 1997), the pixel-level accuracy metrics were calculated for each TSA, individually. Standard errors, reported for accuracy metrics by ecoregion and CONUS, were then estimated accounting for the stratified sampling design (Stehman et al., 2007; Padilla et al., 2015). The general estimator for each accuracy metric was defined as the stratified combined ratio estimator (Cochran, 1977):

$$\hat{R} = \frac{\sum_{h=1}^H K_h \bar{y}_h}{\sum_{h=1}^H K_h \bar{x}_h} \quad (1)$$

where H is the number of strata, K_h is the size of stratum h , \bar{y}_h and \bar{x}_h are the sample means of y_t and x_t of stratum h , and y_t and x_t are the numerator and denominator of each accuracy metric equation, respectively (Padilla et al., 2014a; Appendix A). The estimated variance of \hat{R} was in turn defined as:

$$\hat{V}(\hat{R}) = \frac{1}{\hat{R}^2} \sum_{h=1}^H \frac{K_h^2}{k_h(k_h-1)} \sum_{t \in h} d_t^2 \quad (2)$$

where, k_h is the number of TSAs sampled in stratum h and \hat{X} and d_t are defined as:

$$\hat{X} = \sum_{h=1}^H K_h \bar{x}_h \quad (3)$$

$$d_t = (y_t - \bar{y}_h) - \hat{R}(x_t - \bar{x}_h) \quad (4)$$

Lastly, the standard error was calculated as:

$$SE = \sqrt{\hat{V}(\bar{y}_{st})} \quad (5)$$

Linear regressions were also performed, comparing the amount of burned area mapped by the reference dataset to the amount of burned area mapped by the BAECV product (ha per TSA and year). Accuracy metrics were presented by ecoregions (Fig. 1), land cover type, and burn size. Ecoregions were based on the U.S. Environmental Protection Agency’s Level I ecoregions and included the Mountain West, Arid West, Great Plains and East (Omernik and Griffith, 2014) (Fig. 1). Ecoregions were presented as they are more commonly used in the United States than Olson’s biomes and show similar spatial patterns in their distribution (Fig. 1). The National Land Cover Database (NLCD) (Homer et al., 2015) was used to stratify the validation results by land cover type. The relative abundance of NLCD land cover types within each of the ecoregions are shown in Table 4. The Arid West is dominated by shrub/scrub, the Mountain West is dominated by evergreen forest

Table 4

The relative abundance of National Land Cover Database (NLCD) land cover types across CONUS and within each of the four U.S. Environmental Protection Agency Level I ecoregions.

NLCD land cover types (2006)	CONUS (ha)	CONUS (%)	Arid West (%)	Mountain West (%)	Great Plains (%)	East (%)
Deciduous forest	87,625,705	12.1	0.4	3.9	3.0	26.7
Evergreen forest	93,412,343	12.9	9.6	50.8	1.7	9.9
Mixed forest	16,186,130	2.2	0.4	1.9	0.1	4.8
Shrub/scrub	174,633,568	24.2	64.8	23.2	12.4	4.0
Grasslands/herbaceous	117,627,598	16.3	9.3	10.7	36.9	2.7
Pasture/hay	53,751,191	7.4	1.6	1.8	6.7	11.9
Cultivated crops	125,299,764	17.3	5.4	1.3	30.4	16.0
Woody wetlands	31,243,109	4.3	0.4	0.8	1.0	9.5
Emergent herbaceous wetlands	10,501,421	1.5	0.4	0.5	1.3	2.0
Other (developed, barren, open water)	12,110,510	1.7	7.7	5.1	6.5	12.5

and shrub/scrub, the Great Plains is dominated by grasslands and cultivated crops and the East is the most diverse, dominated by deciduous forest, evergreen forest, pasture/hay, and woody wetlands (Table 4).

To evaluate if there was a relationship between accuracy and burned area size, the reference dataset was converted to shapefiles. Polygons from a single year, co-located within 1 km of each other were aggregated into a single burned area polygon. This step prevented heterogeneous fires, which map as multiple polygons, from being considered multiple small fires. Any overlap between the BAECV and aggregated reference data was considered detection of a burn event. Errors of omission were calculated for burned areas based on binned size classes.

Temporal stability, or the change in the accuracy of the product over time, is a validation component required by the CEOS LPVS. Cuzick's Test for Trend, a nonparametric extension of the Wilcoxon rank-sum test (Cuzick, 1985), was used to detect a temporal trend in the BAECV accuracy metrics (omission error and commission error) from 1988 to 2008. To test for significant differences in median accuracy (omission error and commission error) among years, a nonparametric test was used, the repeated measures analysis of variance (ANOVA) with a rank transformation. A rank transformed repeated measures ANOVA was selected over the Friedman rank sum test as used by Padilla et al. (2014b). Unlike the Friedman rank sum test, the rank transformed repeated measures ANOVA incorporates both the rank of accuracy scores, as well as the magnitude of differences in accuracy (Zimmerman and Zumbo, 1993). The error statistics calculated for each TSA and year were used as the inputs into each of the temporal stability statistical tests.

3. Results

3.1. Comparison of Landsat reference dataset renditions

The three renditions of the Landsat reference dataset were compared to 16 high-resolution images, each of which contained at least one burned patch. As the rendition of the burned area extent became

Table 5

Comparison of the three renditions of the Landsat reference dataset to burned area defined by the QuickBird-2 imagery. Errors of omission and commission were calculated for burned areas only.

Accuracy metrics	Reference - 1 agreed	Reference - 2 agreed	Reference - 3 agreed
Burned area omission error (%)	47	32	24
Burned area commission error (%)	30	33	40
Overall accuracy (%)	94	94	93
Dice coefficient (%)	61	68	67
Relative bias (%)	-24	1	27

more conservative (i.e., more analysts were required to identify a pixel was burned), errors of omission for burned areas decreased from 47% to 24%, while errors of commission for burned areas increased from 30% to 40% (from the most liberal to the most conservative rendition of the reference dataset) (Table 5). The rendition of the reference dataset in which burned area extent was defined by requiring at least two of the three analysts to identify a pixel as burned showed the best balance between errors of omission (32%) and commission (33%) (Table 5). The "2 agreed" rendition of the Landsat reference dataset was therefore selected as the primary reference dataset for validation of the BAECV. However, the "1 agreed" and "3 agreed" renditions are still helpful in bounding uncertainty of error estimates.

We can visually compare patterns between the high-resolution imagery, the Landsat "2 agreed" reference dataset and the BAECV. Some fire events showed high agreement among those mapped with the QuickBird-2 imagery, the Landsat reference dataset and the BAECV (Fig. 2). In such cases, error was primarily disagreement regarding within-fire heterogeneity (Fig. 2). In contrast, other fire events, such as those mapped in South Carolina woody wetlands showed adequate agreement between the QuickBird-2 output and the Landsat reference dataset, but poor agreement with the BAECV (Fig. 3). Differences in the detection or omission of fire events between the QuickBird-2 and Landsat reference map demonstrate the challenge of using high-resolution imagery. It is unclear how much of the difference between the high-resolution image and the reference dataset represents "true" error, or if some portion of the disagreement is due to QuickBird-2 mapping "old" fires not expected to be mapped by the Landsat reference dataset or the BAECV. This problem stems from the limited availability of same-year, pre-fire imagery from high-resolution data sources.

3.2. Pixel-level validation of the BAECV

Accuracy of the BAECV varied depending on the rendition of the reference dataset used. BAECV errors of omission were lowest using the "3 agreed" rendition of the reference dataset, while errors of commission were lowest using the "1 agreed" rendition of the reference dataset (Table 4). In Fig. 4 we show an example of variation between the three reference datasets, relative to the BAECV. Using the "3 agreed" rendition, for example, instead of the "2 agreed" rendition would have classified the BAECV's pixels within the northern burned area in the example provided, as errors of commission (Fig. 4). Differences in accuracy were substantial among the three levels, for instance in the Arid West errors of omission ranged from 27% to 52% from 3 to 1 analysts agreed, and errors of commission ranged from 16% to 39% from 1 to 3 analysts agreed (Table 6). The version of the reference dataset used influenced the accuracy results of the Great Plains the least and the East, the most. Changing what version of the reference dataset was used to validate the Great Plains changed omission error by 10% and commission error by 17% (Table 6). In the East, changing what version of the reference dataset was used to validate the East BAECV changed omission error by 26% and commission error by 33% (Table 4). The range in error rates across the three renditions of the reference dataset can help bound uncertainty levels.

The accuracy of the BAECV when compared to the reference dataset ("2 agreed") showed high overall accuracy (>99%) across all ecoregions because of the prevalence of unburned areas and the relative ease of mapping unburned as unburned relative to accurately mapping burned areas (Table 6). When errors of omission and commission for burned areas were calculated, errors varied between ecoregions (Table 6). Both errors of omission and commission were lowest in the Arid West (31% and 24%, respectively) and highest in the Great Plains (62% and 57%, respectively) (Table 6). Over 30% of the Great Plains is mapped as cultivated crops or cropland by NLCD (Table 4) (Homer et al., 2015). Burned patches within and adjacent to cropland are not typically mapped by federal fire perimeter datasets and are often too small to be identified by coarse resolution sensors. They therefore potentially

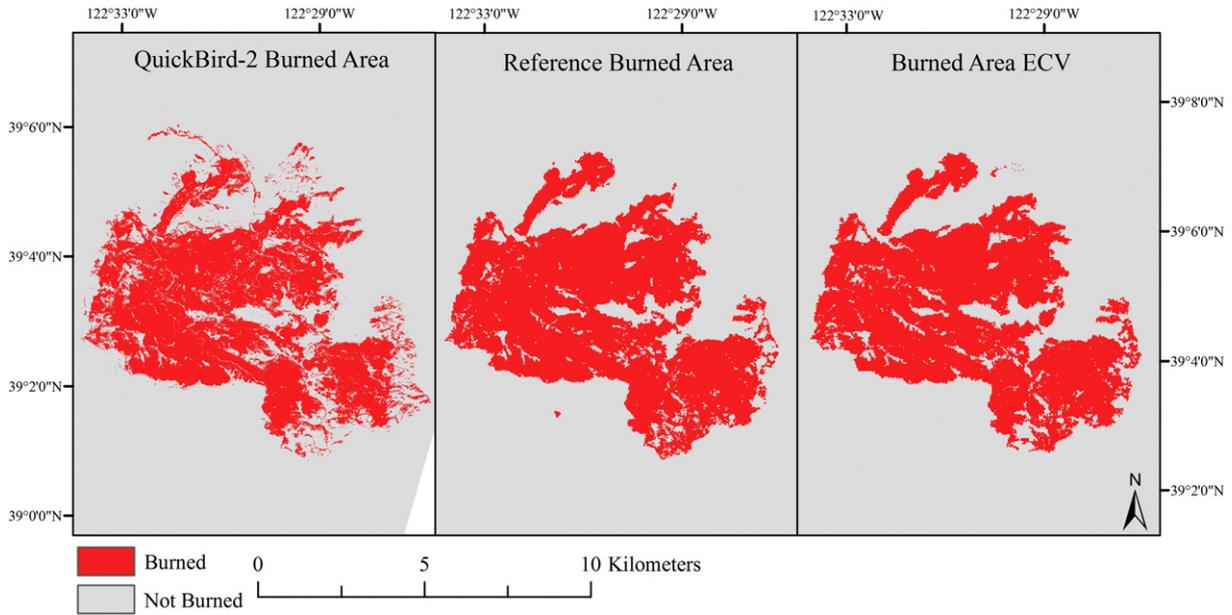


Fig. 2. An example comparison of a burned area within shrub/scrub cover in California as mapped by QuickBird-2 (image collected on November 18, 2008), the Landsat reference dataset (“2 agreed”) (post-image collected on October 12, 2008, path 44, row 33) and the Burned Area Essential Climate Variable (BAECV) (post-image, October 12, 2008, path 44, row 33).

represent a novel contribution of the BAECV. Lower accuracy within this cover type, however, influenced overall accuracy particularly in the Great Plains and East. Excluding the cropland cover type reduced commission error from 59% to 51% in the Great Plains and reduced omission error in the East from 63% to 50%. Cultivated crops also showed the weakest accuracy when burned area accuracy was assessed by NLCD cover type (Table 7).

Differences in accuracy among land cover types can also be contextualized by considering differences in the relative frequency or abundance of burned area among land cover types. Burned area was most often identified by the BAECV in shrub/scrub (32%), evergreen forest (24%) and grasslands (21%) land cover types (Table 7). Shrub/scrub

showed one of the lowest omission and commission errors for burned area (32% and 23%, respectively), while evergreen forest (40% and 33%, respectively) and grasslands (40% and 32%, respectively) showed errors of omission and commission similar or better than the average accuracy across CONUS (42% and 33%, respectively) (Table 8). Examples of agreement and disagreement between the reference dataset and the BAECV across multiple land cover types are shown in Fig. 5.

The amount of burned area by TSA and year were compared between that mapped by the reference dataset and that mapped by the BAECV. Significant correlations were observed in the Mountain West and Arid West, even after excluding a TSA with much higher burned area than other TSAs in the Arid West (Fig. 6). The correlation in the

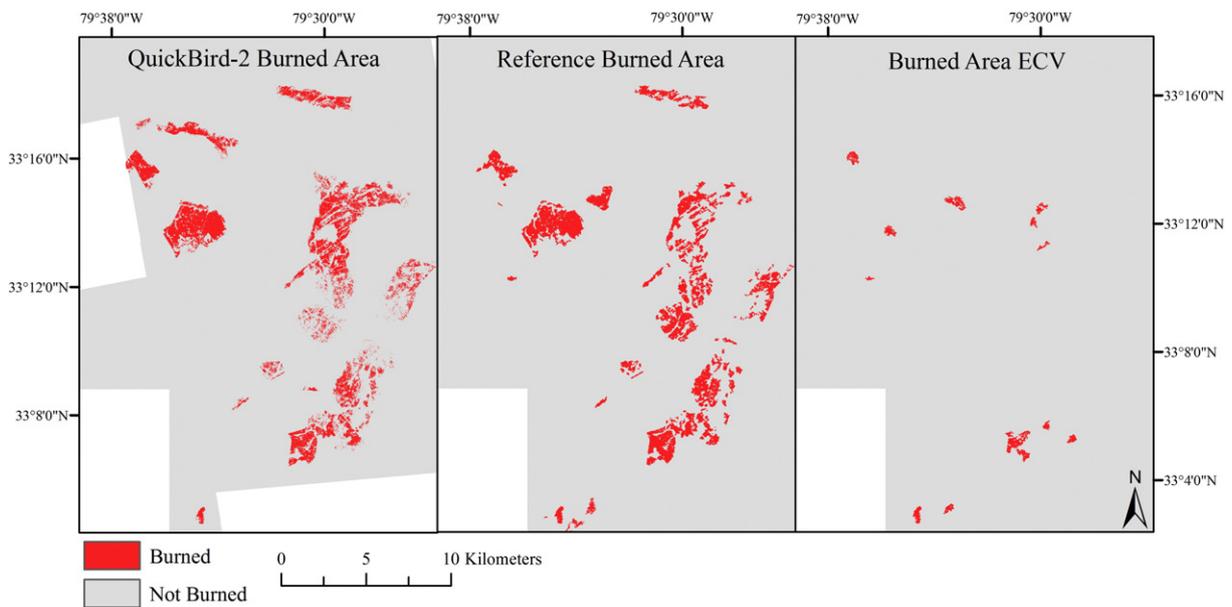


Fig. 3. An example comparison of a burned area within woody wetlands in South Carolina as mapped by QuickBird-2 (image collected on April 21, 2008), the Landsat reference dataset (“2 agreed”) (post-image collected on May 1, 2008, path 16, row 37) and the Burned Area Essential Climate Variable (BAECV) (post-image, May 1, 2008, path 16, row 37). White areas represent high-resolution image extent and areas masked due to fires occurring between image collection dates.

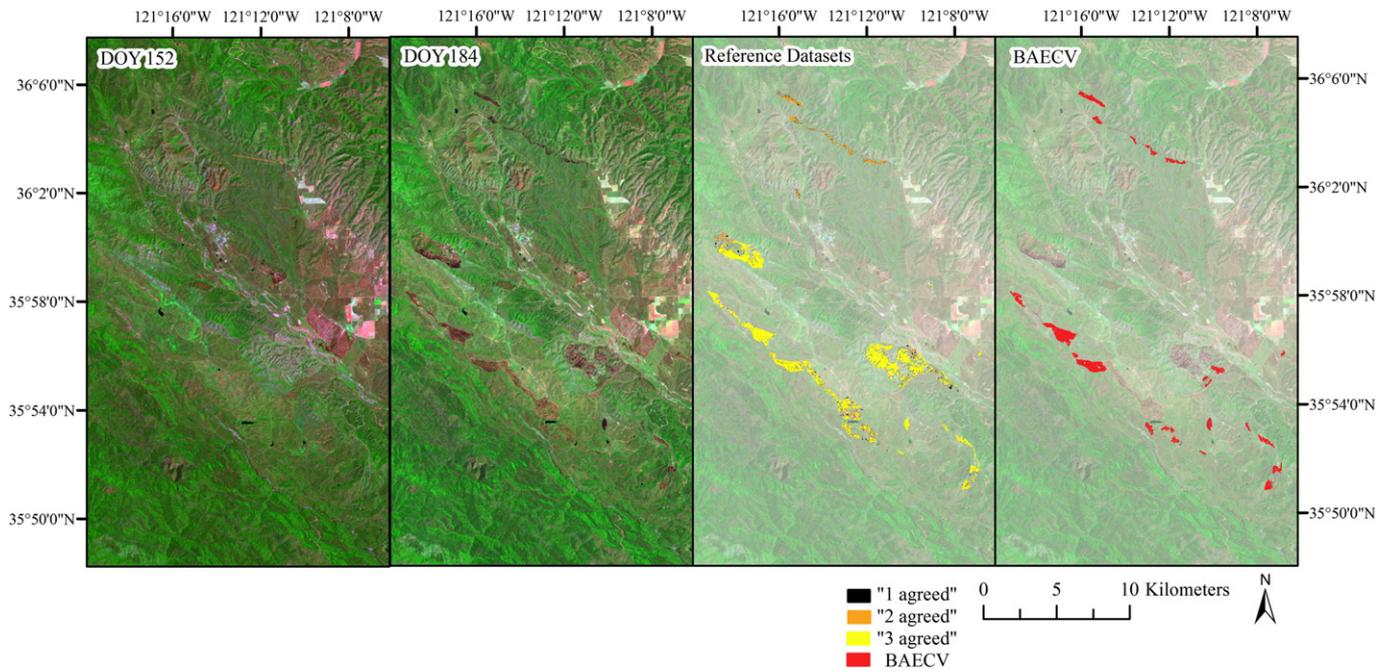


Fig. 4. A comparison from southern California (path 43, row 35, 2003) between the pre- and post-Landsat images used to derive the reference dataset, the three renditions of the reference dataset (1 to 3 analysts agreed a pixel was burned), and the Burned Area Essential Climate Variable (BAECV).

Great Plains was heavily influenced by an outlier TSA, in which the BAECV missed actively burning agricultural fires. In the East, the correlation was weak using all TSAs, but became significant when excluding the cultivated crops cover type from TSAs, suggesting that the BAECV's accuracy within active agricultural areas limits its use for monitoring trends in agricultural fires in this region (Fig. 6). When all TSAs and years were considered together the amount of burned area detected by the BAECV was significantly correlated with the amount of burned area detected by the reference dataset ($r^2 = 0.62$, $RMSE = 3057$ ha, $p < 0.01$) (Fig. 7).

At the scale of CONUS, the accuracy metrics including errors of omission, dice coefficient and relative bias varied substantially between the five sample years, suggesting the importance of sampling from multiple years (Table 8). Commission errors were relatively stable between the five sampled years. Omission errors, however, were particularly high in 1993 ($69 \pm 8\%$) and particularly low in 2008 ($19 \pm 6\%$) (Table 8). The year 1993 was a wet year across much of the CONUS, in particular the Mountain West, resulting in very little burned area in the reference dataset. The smaller burned area sample size in the reference dataset would have exaggerated the importance of errors and could have

potentially influenced the aggregated results. Standard errors decreased from individual years to aggregated years because of the increase in sample size, given a TSA as the sample unit (Table 8).

3.3. Validation by fire size

According to GCOS requirements and the climate-modeling community there is a desire for a product that detects burned area patches smaller than 25 ha (Chuvieco et al., 2016). A primary advantage of mapping fires with Landsat relative to coarser resolution imagery is the ability to detect smaller fires. The BAECV aims to map fire events as small as 4.05 ha (10 acres). We evaluated the ability of the BAECV to detect fires by burned area size. Omission error related to the BAECV missing small fires may be difficult to interpret in a pixel-level analysis where large fires may contribute relatively more pixels for comparison. An analysis of detection/non-detection by fire size, however, can help highlight potential biases within pixel-based accuracy statistics. Across all ecoregions, the BAECV detected 50% of the 10 to 25 ha burned area patches, this percent increased as fire size increased up to 87% for large fires (>300 ha) (Table 9). Similar to the pixel-level accuracy

Table 6
Variation in the accuracy of the Burned Area Essential Climate Variable (BAECV) when using the three renditions of the reference dataset (1 to 3 analysts agreed a pixel was burned) (1988–2008). Standard errors, calculated as variation in accuracy statistics between Thiessen scene areas, are shown in parentheses. Shaded cells indicate the validation results by ecoregion using the selected “2 agreed” rendition of the Landsat reference dataset.

Ecoregion	Rendition	Overall accuracy	Omission error (%)	Commission error (%)	Dice coefficient (%)	Relative bias (%)
Arid West	1 agreed	99.9 (0.03)	52 (7)	16 (2)	61 (7)	−43 (13)
	2 agreed	99.9 (0.02)	31 (6)	24 (3)	72 (7)	−10 (16)
	3 agreed	99.7 (0.02)	27 (7)	39 (3)	67 (9)	19 (21)
Mountain West	1 agreed	99.8 (0.04)	48 (4)	18 (5)	63 (4)	−37 (9)
	2 agreed	99.8 (0.03)	41 (7)	32 (5)	63 (5)	−13 (16)
	3 agreed	99.8 (0.02)	37 (12)	46 (3)	58 (5)	18 (13)
Great Plains	1 agreed	99.9 (0.09)	68 (12)	51 (9)	39 (11)	−35 (14)
	2 agreed	99.9 (0.04)	62 (9)	57 (9)	40 (6)	−10 (12)
	3 agreed	99.8 (0.04)	58 (9)	68 (9)	36 (5)	31 (12)
East	1 agreed	99.9 (0.06)	78 (7)	35 (4)	33 (5)	−66 (11)
	2 agreed	99.9 (0.04)	67 (8)	47 (5)	41 (4)	−37 (15)
	3 agreed	99.8 (0.04)	53 (8)	69 (5)	38 (5)	53 (18)

Table 7

Variation in accuracy by the National Land Cover Database (NLCD) land cover type (using the “2 agreed” rendition of the reference dataset). Errors of omission and commission are presented for burned areas only. The relative distribution of burned area (BA) between cover types is also shown for the reference (28 Thiessen scene areas (TSAs)) and Burned Area Essential Climate Variable (BAECV) datasets (28 TSAs). Overall accuracy was 99.5% or higher across all land cover types and therefore is not shown.

Land cover type ^a	BA ha (%) (Reference)	BA ha (%) (BAECV)	Omission error (%)	Commission Error (%)	Dice coefficient (%)	Relative bias (%)
Deciduous forest	2.0	2.6	33	34	67	2
Evergreen forest	21.6	24.0	40	33	63	– 11
Mixed forest	5.0	6.2	20	20	80	0
Shrub/scrub	29.1	32.1	32	23	73	– 12
Grasslands/herbaceous	18.6	20.6	40	32	64	– 11
Pasture/hay	1.3	2.7	51	70	37	63
Woody wetlands	2.0	2.2	39	31	64	– 12
Emergent wetlands	8.8	4.0	76	33	36	– 63
Cultivated crops	11.5	5.6	90	88	11	– 15

^a Certain land cover classes were not included due to its exclusion from BAECV (e.g., developed) or minimal number of burned pixels (e.g., perennial ice/snow, bare rock, etc.).

statistics, the BAECV experienced more omission in the East relative to other ecoregions. In the Mountain West, for example, 76% of the 10 to 25 ha burned areas were detected and this only increased with larger fires (Table 9).

3.4. Temporal stability of the BAECV

Temporal stability in the accuracy of a product has been identified as essential for climate modelers and other users of burned area to be able to interpret trends in burned area over time. The temporal stability of the BAECV was evaluated across the five validation years (1988, 1993, 1998, 2003 and 2008). There was no statistical trend in the error of omission ($z = -1.41, p = 0.16$) or the error of commission ($z = -1.38, p = 0.17$) over the validation years (1988 to 2008) as determined by Cuzick’s Test for Trend, despite the potential for a trend due to a documented change in Landsat 5’s satellite orbit over its 27-year time series (Zhang and Roy, 2016). We also tested for significant differences in accuracy between years, not related to a temporal trend. There were no statistically significant differences between the errors of omission ($F = 2.32, p = 0.13$) and errors of commission ($F = 3.15, p = 0.08$) between the validation years as determined by a repeated measures ANOVA with a rank transformation. As the BAECV is expanded to include Landsat-8 data, it will be important to evaluate any potential change in accuracy.

4. Discussion

Since the opening of USGS’s Landsat archive in 2008, there has been great interest in generating relevant, remotely sensed products from Landsat imagery (Hansen and Loveland, 2012; Kovalsky and Roy, 2013; Masek et al., 2013). Validation of such products is widely accepted as necessary (Morissette et al., 2006; Hansen and Loveland, 2012). Given the long temporal range of Landsat imagery (1972–present) and moderate spatial resolution (30 m), however, it is challenging to generate and defend an independent dataset that can adequately validate Landsat-derived science products. Such a dataset must show a defensible sampling design and include a substantial number of observations

over both space and time. Disturbance-related products have taken advantage of spectrally diverse, but visually consistent changes in Landsat images (pre and post disturbance) that can be used to improve the identification of disturbances by incorporating a manual interpretation component (Henry, 2008; Petropoulos et al., 2011; Thomas et al., 2011). Relying on Landsat for both product generation and validation, however, limits our ability to assess inaccuracies imposed by the satellite sensor itself, such as spectral data quality, geolocation and mixed pixels (Strahler et al., 2006). It also prevents us from being able to quantify potential errors of omission within the BAECV due to the return interval of the Landsat satellites. This source of error can be expected to primarily occur in cover types that experience rapid recovery or manipulation following a fire event, for example fires within grassland and agricultural cover types. This approach, however, also has strengths in that it reduces problems of image registration and eliminates false error due to time gaps between the reference dataset and product dataset.

A novel component of our approach was that three independent datasets were derived and differences between the three datasets were used to bound uncertainty and improve confidence in our Landsat reference dataset. This approach showed that manually derived reference datasets that are typically considered “truth” can show considerable variability between analysts and still show disagreement when compared, for example, to burn patches mapped with high-resolution imagery. Defining burned area extent as requiring at least two of the three analysts to identify a pixel as burned, best balanced errors of omission and commission within the reference dataset. In the future, other alternatives could also be considered including having multiple analysts reach a consensus on pixels for which analysts disagreed. This finding could be used to improve future efforts to generate independent reference datasets.

The accuracy of the BAECV can be contextualized by comparing it to other efforts to map burned area with Landsat as well as global burned area datasets derived from coarser resolution imagery. When burned area algorithms are optimized for site level performance, the accuracy can exceed 95% (e.g., Mitri and Gitas, 2004; Petropoulos et al., 2011). As study areas grow in size, accuracy often starts to decrease (e.g., 15% to 30% error for burned area) due to variance imposed by local factors

Table 8

Variation in accuracy by year for the conterminous United States (CONUS) (1988–2008) using the “2 agreed” rendition of the Landsat reference dataset. Errors of omission and commission are presented for burned areas only. Standard errors, calculated as variation in accuracy statistics between Thiessen scene areas (28 per year), are shown in parentheses. BAECV: Burned Area Essential Climate Variable.

Year	Overall accuracy	Omission error (%)	Commission error (%)	Dice coefficient (%)	Relative bias (%)	Burned area (Reference, ha)	Burned area (BAECV, ha)
1988	99.8 (0.04)	37 (8)	38 (6)	62 (5)	3 (15)	85,485	87,288
1993	99.9 (0.04)	69 (8)	28 (6)	44 (8)	– 57 (12)	59,636	24,457
1998	99.9 (0.02)	45 (9)	32 (8)	61 (8)	– 20 (11)	50,964	40,613
2003	99.8 (0.05)	56 (12)	29 (5)	54 (10)	– 38 (12)	118,093	56,827
2008	99.8 (0.05)	19 (6)	32 (6)	74 (6)	19 (16)	101,607	120,628
88–08	99.9 (0.02)	42 (6)	33 (3)	62 (4)	– 14 (8)	415,785	329,814

(Bastarrika et al., 2011; Mallinis and Koutsias, 2012; Goodwin and Collett, 2014). Several Landsat-based efforts have mapped disturbance (fires, clearcut and insect mortality) across parts of CONUS but limited efforts to forest. The validations corresponding to these mapping efforts have relied on manual interpretation of Landsat images with errors of

omission and commission averaging 40–45% and 20–30%, respectively (Masek et al., 2008; Thomas et al., 2011). Validation for these efforts was supplemented using Forest Service Inventory and Analysis (FIA) plots and showed errors of omission and commission averaging 36% (0 to 100%) and 64% (26 to 100%), respectively. More recently,

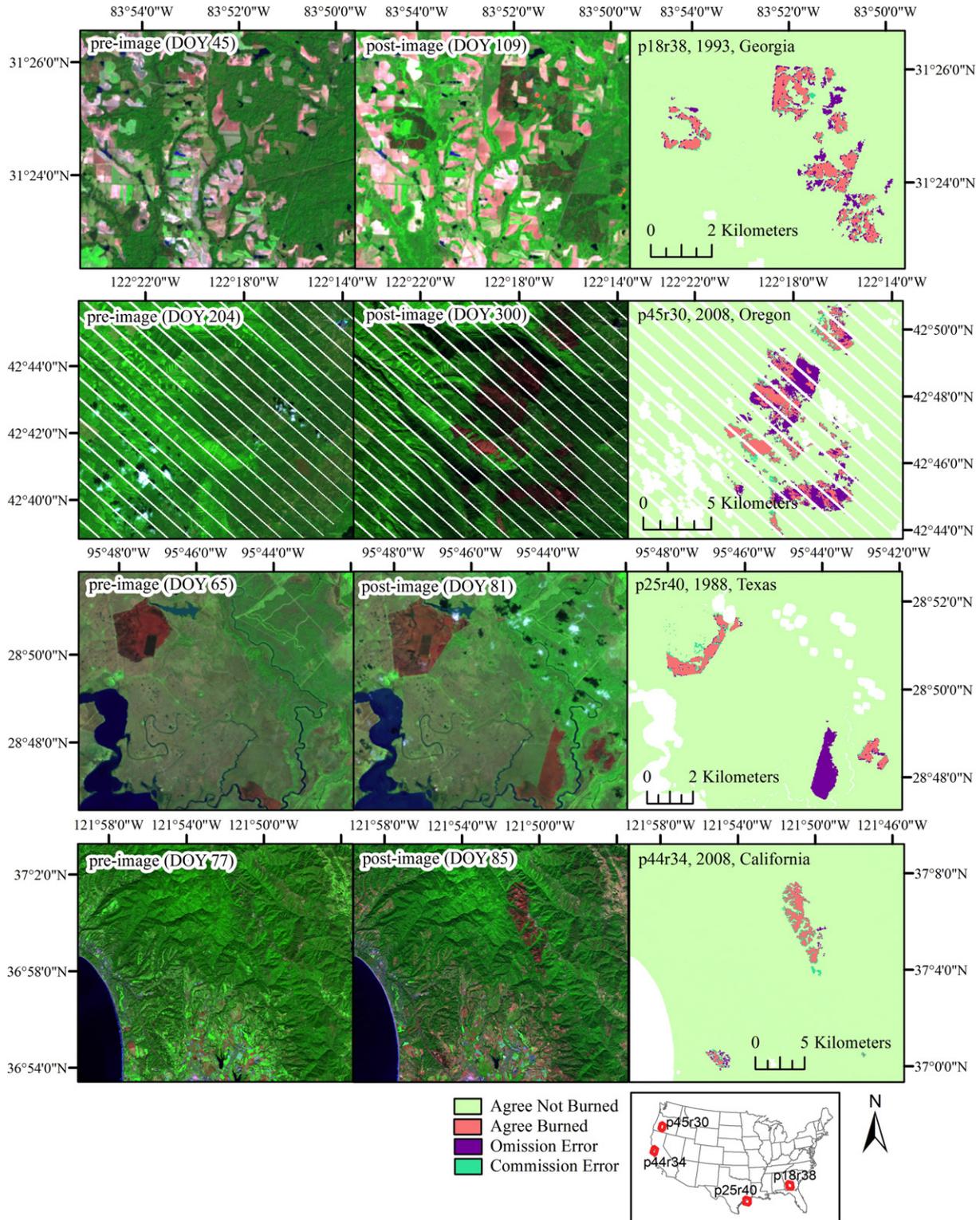


Fig. 5. Examples of the Landsat pre-fire and post-fire image pairs and corresponding agreement and disagreement between the reference dataset and BAECV from four Thiessen scene areas (TSAs) that showed errors of omission and commission, at the scale of a TSA, similar to the average observed across CONUS. The Georgia fires occurred within evergreen forest, the Oregon fires within grasslands, the Texas fires within emergent herbaceous wetlands, and the California fires within mixed forest and shrub/scrub.

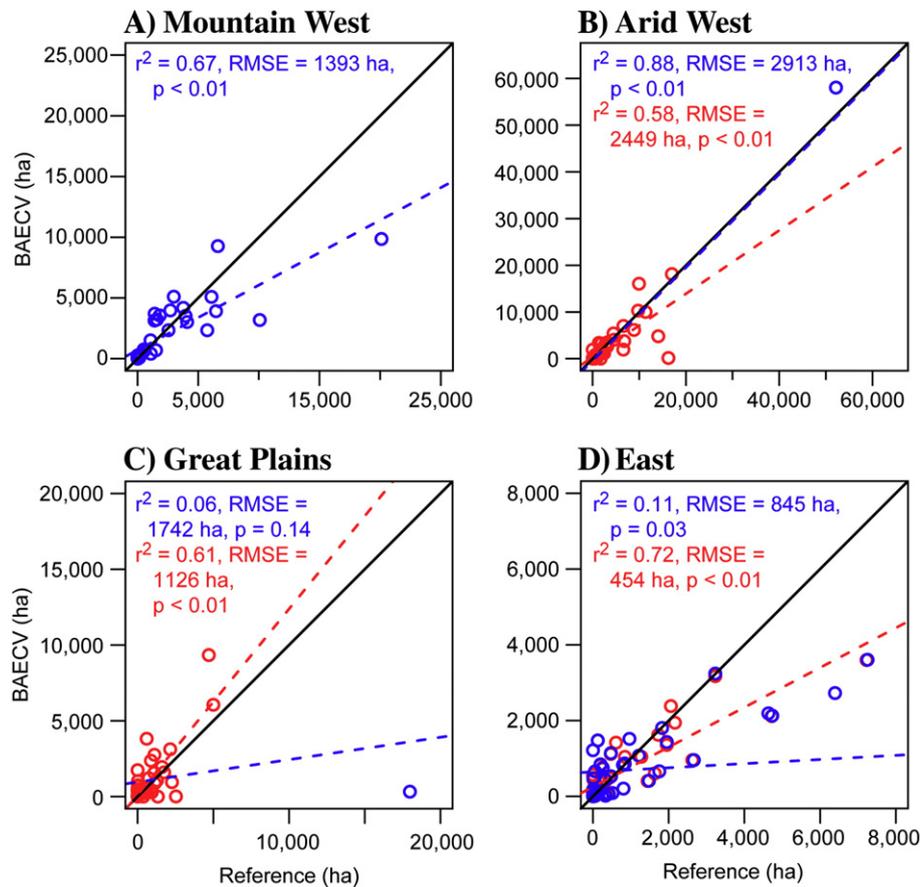


Fig. 6. The correlation between the amount of burned area identified by the reference dataset and the amount of burned area identified by the Burned Area Essential Climate Variable (BAECV) (blue) for ecoregions including A) the Mountain West, B) the Arid West, C) the Great Plains and D) the East. Each point represents the hectares burned by Thiessen scene area (TSA) and year. The correlation in the Arid West is shown with and without (red) an outlier TSA with much higher burned area than the other TSAs in the sample. The correlation in the Great Plains region is shown with and without (red) an outlier TSA with high rates of omission by the BAECV. The correlation in the East is shown with and without (red) pixels classified as cultivated crops within each TSA, a cover type that showed higher rates of omission by the BAECV. RMSE: root mean square error.

Boschetti et al. (2015) mapped burned areas from Landsat for a single year across much of the western United States. Relative to MTBS perimeters, the burned area extent showed 51% omission error and 14% commission error.

The spectral diversity within burned patches can be expected to increase as the vegetation and cover type diversity increases, as is the case in national and global scale efforts. This makes high levels of accuracy at these larger spatial extents considerably more challenging. We can use NLCD as a non-fire example of a national Landsat-based product. Although overall accuracy of NLCD cover classes averaged 78–79% for NLCD 2001 and 2006, the accuracy of cover class change was quite poor, with errors of omission ranging from 61 to 89% and errors of commission ranging from 18 to 73% (Wickham et al., 2013). At a global scale, accuracy of existing global burned area products, derived from coarser resolution imagery, also shows relatively weak accuracy. Errors of omission for such products have been documented to range from 51% to 93% for omission error and from 36% to 94% for commission error (Padilla et al., 2015; Chuvieco et al., 2016). This error may be higher for specific vegetation types. For example, although the MODIS burned area product (MCD45) performed well relative to other global burned area products (Padilla et al., 2015), it performed the worst in temperate forest (99% omission, 95% commission), which is a major biome across CONUS (Fig. 1).

Relative to other documented validation efforts, the BAECV's error rates (errors of omission and commission averaged 42% and 33%, respectively across CONUS) outperformed accuracy statistics reported

for coarser resolution global burned area products and showed rates of error slightly higher than efforts using Landsat imagery but constrained to forested cover types. Differences in the BAECV's error rates between ecoregions, in turn, can primarily be explained by issues specific to particular vegetation cover types. For instance, omission and commission errors were highest in agricultural cover types (i.e., cultivated crops and hay/pasture). In agricultural areas, frequent changes in site condition (green vegetation, non-photosynthetic vegetation, burned, tilled) make it challenging to use change-detection approaches to distinguish burn events. The limited training data within active agricultural areas also likely contributed to the poor performance within this cover type. Errors of omission were also high in emergent wetlands, in which a fire event often results in a transition from vegetation to open water. A portion of the errors may have also been related to variable severity. For example, prescribed fires, common in the southeastern U.S. are typically intended to clear understory vegetation (Waldrop et al., 1987, 1992), and therefore may be more difficult to detect if the overstory vegetation is largely undisturbed. Similarly, slightly higher error rates within evergreen forest, relative to other forest types, could be related to missing understory fire events, potential confusion with other sources of tree mortality, such as bark beetle damage, and misclassification within shaded pixels, common in high-relief environments, which are often dominated by evergreen forest.

Pixel-based accuracy metrics can help explain spatial variability in accuracy between biomes, ecoregions or land cover types; however, such metrics can confound error from disagreement in mapping within

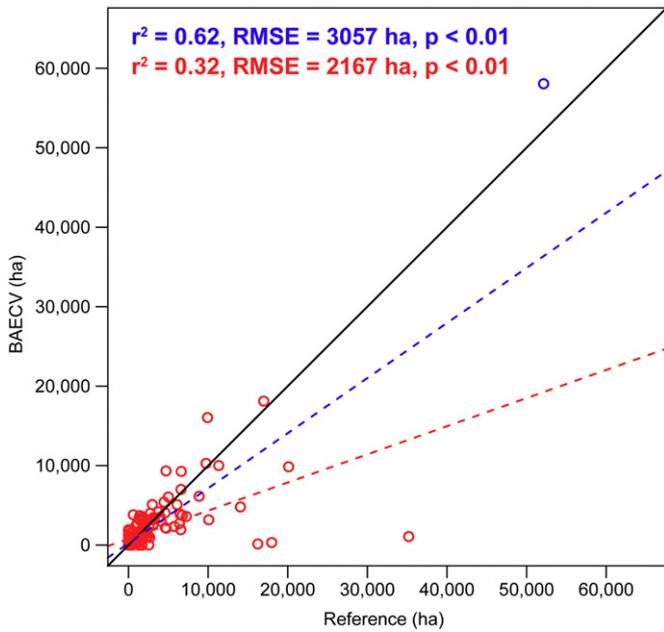


Fig. 7. The correlation between the amount of burned area identified by the reference dataset and the amount of burned area identified by the Burned Area Essential Climate Variable (BAECV) across the conterminous U.S. (CONUS) where each point represents a Thiessen scene area (TSA) and year. All cover types are included, but the correlation is shown with and without (red) an outlier TSA with much higher burned area than the other TSAs in the sample. RMSE: root mean square error.

Table 9

The percent of burned areas (or fire events) detected by the Burned Area Essential Climate Variable (BAECV) by fire size and region. Burned areas binned by size were derived from an aggregated version of the Landsat reference dataset (all polygons within 1 km of each other were considered as a single fire) to avoid fragmented or heterogeneous fires being considered as more than one fire. All years were combined (1988–2008).

Ecoregion	4.05 to 10 ha (%)	10 to 25 ha (%)	25 to 50 ha (%)	50 to 100 ha (%)	100 to 300 ha (%)	>300 ha (%)
Arid West	42	69	80	68	82	98
Mountain West	33	76	70	80	97	100
Great Plains	56	58	77	81	89	93
East	27	43	47	63	66	75
Conterminous U.S.	32	50	54	67	74	87
Total Fire Count	250	339	283	224	287	301

Appendix A

Table A1

The Landsat images used to generate the reference dataset. All images are from the Landsat TM sensor except for the starred dates, where images are from Landsat ETM+. Burned areas were identified by comparing a post-fire and pre-fire date, necessitating an image pair per path/row and year.

	1988		1993		1998		2003		2008	
Path/row	Pre-fire date	Post-fire date								
p15r35	6-Jul	26-Oct	9-Feb	15-May	27-Apr	21-Nov	14-Jul	19-Nov	26-May	27-Jun
p15r41	27-Jan	16-Mar	28-Mar	15-May	14-May	1-Jul	20-Jan	8-Mar	24-Apr	26-May
p15r42	16-Mar	4-Jun	28-Mar	15-May	11-Mar	12-Apr	20-Jan	8-Mar	24-Apr	11-Jun
p16r31	29-Jul	15-Sep	23-Jun	19-Oct	11-Oct	14-Dec	18-May	5-Jul	7-Apr*	16-Oct*
p16r37	7-Mar	8-Apr	7-Mar	6-May	22-May	27-Jan	31-Mar	27-Jan	10-Feb	1-May
p18r38	8-May	18-Dec	14-Feb	19-Apr	11-Jan	26-Nov	9-Jan	23-Oct	15-May	31-May
p24r29	31-Mar	2-May	5-Oct	21-Oct	26-Apr	17-Sep	8-Apr	26-May	1-May*	2-Jun*
p24r39	2-May	7-Sep	27-Mar	6-Nov	10-Apr	31-Jul	1-Oct	20-Dec	9-May	28-Jul
p25r40	6-Mar	22-Mar	2-Mar	26-Sep	13-Dec	29-Dec	30-Mar	9-Nov	30-Apr	19-Jul
p26r40	24-Jan	14-Apr	12-May	1-Sep	8-Apr	10-May	29-Sep	18-Dec	20-Mar	24-Jun
p26r41	24-Jan	14-Apr	31-Jul	20-Nov	8-Apr	11-Jun	29-Sep	18-Dec	20-Mar	24-Jun
p26r42	24-Jan	13-Mar	6-Feb	9-Mar	19-Jan	20-Feb	17-Jan	8-May	20-Mar	24-Jun

fire heterogeneity with error from mapping false fires or missing fires. The relative importance of each error type is likely to depend on the application of the dataset, but supplementing pixel-based accuracy metrics with additional assessments such as linear regressions and detection/non-detection by fire size can help provide a more complete understanding of the distribution of error within the BAECV. For example, the BAECV reliably mapped large fires (>25 ha) across most ecoregions, but was less reliable in its ability to consistently map small fires, particularly those under 10 ha. We saw significant correlations between the magnitudes identified as burned by the BAECV and reference dataset, but also observed that outlier TSAs could have a strong influence on the strength of this correlation.

5. Conclusion

The BAECV, by largely automating the detection of burned areas and using the entire Landsat archive, potentially provides a more complete census of fires across all regions of the conterminous United States, improving our knowledge of fires in traditionally underrepresented land cover types such as grasslands and agricultural areas. Error within the BAECV was most concentrated in several land cover types, specifically the BAECV tended to underestimate burned area within emergent wetlands, overestimate burned area in pasture/hay and show high rates of both omission and commission in areas dominated by cultivated crops. Errors of omission were also higher for small fires (<10 ha). In contrast, errors were lowest in forest, shrub/scrub, and grassland, land cover types dominant across much of western and eastern CONUS. No temporal trend in accuracy was observed over the validation years selected. The BAECV is only one of several ongoing USGS efforts to create Landsat-based science products. Validation of disturbance-related Landsat science products may depend heavily on creating independent datasets from Landsat to validate systematically across space and over the entire temporal extent of the Landsat archive. Adding a manual component can potentially improve the accuracy of the validation dataset, however, variability in analyst interpretation can be substantial and should be considered.

Acknowledgements

We thank the U.S. Geological Survey, Climate and Land Use Mission Area, Land Remote Sensing Program for providing funding to support this research. We are grateful to Carol Mladinich and Susan Stitt for their role in initiating the validation effort and assessing preliminary results of the BAECV algorithm. Hayley Distler, Matthew Galbraith and Di Ana Mendiola were essential in generating the Landsat reference dataset. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Table A1 (continued)

	1988		1993		1998		2003		2008	
p30r30	16-Aug	15-Sep	24-May	29-Sep	9-Jul	11-Sep	1-Mar	20-May	17-Apr	19-May
p30r36	21-Feb	25-Mar	2-Feb	6-Apr	19-Mar	11-Sep	23-Jul	25-Sep	12-Feb	3-May
p33r30	4-Jul	6-Sep	2-Sep	4-Oct	31-Aug	5-Dec	12-Jul	14-Sep	9-Jun	29-Sep
p35r35	28-Mar	13-Apr	12-Jun	14-Jul	9-May	26-Jun	12-Sep	14-Oct	7-Jun	9-Jul
p36r35	4-Apr	22-May	22-Aug	9-Oct	16-May	3-Jul	1-Jul	21-Oct	21-Nov	7-Dec
p37r38	16-Jul	4-Oct	9-May	26-Jun	21-Apr	8-Jun	12-Oct	15-Dec	29-Feb	20-May
p39r34	11-May	16-Sep	23-May	5-Dec	19-Apr	24-Jul	4-Jun	6-Jul	19-Jun	7-Sep
p41r28	26-Jun	30-Sep	8-Jul	13-Nov	7-Aug	23-Aug	4-Jul	24-Oct	27-Jul*	29-Sep*
p41r29	10-Jun	30-Sep	8-Jul	28-Oct	7-Aug	26-Oct	20-Jul	24-Oct	11-Jul*	13-Sep*
p42r36	21-Mar	5-Sep	1-Sep	6-Dec	30-Aug	4-Dec	1-Jan	15-Oct	4-Mar	15-Nov
p43r33	10-Jul	12-Sep	13-Feb	24-Sep	21-Aug	8-Oct	18-Jul	22-Oct	15-Jun	2-Aug
p43r35	20-Mar	28-Sep	1-Apr	8-Sep	15-Apr	5-Aug	31-May	2-Jul	14-May	19-Sep
p44r33	27-Mar	21-Oct	11-Jun	15-Sep	25-Jun	13-Sep	26-Aug	11-Sep	6-Jun	12-Oct
p44r34	17-Jul	11-Dec	11-Jun	17-Oct	5-Mar	15-Oct	26-Aug	11-Sep	18-Mar	25-Aug
p45r28	8-Jul	10-Sep	18-Jun	9-Nov	31-May	22-Oct	14-Jun	16-Jul	29-Jun	17-Sep
p45r30	24-Jul	10-Sep	18-Jun	24-Oct	3-Aug	22-Oct	14-Jun	16-Jul	23-Jul*	26-Oct*

Table A2

The number of burned and unburned pixels within the reference dataset (30 m resolution) included in the validation, using the “2 agreed” rendition of the reference dataset. NA pixels were defined by CFMask provided with Landsat imagery and include pixels classified as cloud, cloud shadow and open water. Pixels classified as developed by the national Land Cover Database are also included in NA pixels as they were masked out. Total pixels differ between years due to the inclusion of Landsat ETM+ pairs with the scan-line error.

Ecoregion	Year	Burned pixels	Burned count %	Unburned pixels	Unburned count %	NA Pixels	NA count %	Total pixels
Arid West	1988	331,685	0.2	166,853,592	98.7	1,904,551	1.1	169,089,828
	1993	500,630	0.3	161,023,414	95.2	7,564,736	4.5	169,088,780
	1998	359,876	0.2	156,711,278	92.7	12,016,850	7.1	169,088,004
	2003	225,537	0.1	156,353,322	92.5	12,500,433	7.4	169,079,292
	2008	825,963	0.5	165,463,974	98.0	2,550,023	1.5	168,839,960
Mountain West	1988	459,910	0.5	87,418,460	89.0	10,330,854	10.5	98,209,224
	1993	1613	0.0	58,352,910	59.4	39,861,155	40.6	98,215,678
	1998	89,706	0.1	88,027,119	89.6	10,094,751	10.3	98,211,576
	2003	301,934	0.3	88,876,309	90.5	9,033,066	9.2	98,211,309
	2008	151,364	0.2	74,672,380	92.3	6,103,041	7.5	80,926,785
Great Plains	1988	111,506	0.1	107,521,042	91.8	9,532,519	8.1	117,165,067
	1993	57,758	0.0	106,281,189	90.8	10,723,999	9.2	117,062,946
	1998	78,606	0.1	91,735,746	78.3	25,309,490	21.6	117,123,842
	2003	257,764	0.2	109,864,416	93.8	6,963,709	5.9	117,085,889
	2008	54,262	0.0	76,326,944	67.8	36,125,470	32.1	112,506,676
East	1988	82,182	0.1	132,383,546	83.6	25,946,461	16.4	158,412,189
	1993	118,247	0.1	143,539,259	90.6	14,749,169	9.3	158,406,675
	1998	47,898	0.0	147,714,222	93.3	10,599,431	6.7	158,361,551
	2003	530,594	0.3	136,604,865	86.3	21,111,702	13.3	158,247,161
	2008	116,347	0.1	96,383,818	66.6	48,259,945	33.3	144,760,110

References

- Bastarrika, A., Chuvieco, E., Martin, M.P., 2011. Mapping burned areas from Landsat TM/ETM+ data with a two-phase algorithm: balancing omission and commission errors. *Remote Sens. Environ.* 115, 1003–1012.
- Bastarrika, A., Alvarado, M., Artano, K., Martinez, M.P., Mesanza, A., Torre, L., Ramo, R., Chuvieco, E., 2014. BAMS: a tool for supervised burned area mapping using Landsat data. *Remote Sens.* 6 (12), 12360–12380.
- Baveye, P.C., Laba, M., Otten, W., Bouckaert, L., Sterpaio, P.D., Goswami, R.R., Grinev, D., Huston, A., Hu, Y., Liu, J., Mooney, S., Pajor, R., Sleutel, S., Tarquis, A., Wang, W., Wei, Q., Sezgin, M., 2010. Observer-dependent variability of the thresholding step in the quantitative analysis of soil images and X-ray microtomography data. *Geoderma* 157 (1–2), 51–63.
- Bond-Lamberty, B., Peckham, S.D., Ahl, D.E., Gower, S.T., 2007. Fire as the dominant driver of central Canadian boreal forest carbon balance. *Nature* 450, 89–92.
- Boschetti, L., Roy, D.P., Justice, C.O., Humber, M.L., 2015. MODIS-Landsat fusion for large area 30 m burned area mapping. *Remote Sens. Environ.* 161, 27–42.
- Breiman, L., Friedeman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. 1st ed. Chapman and Hall, New York, NY, USA.
- Chuvieco, E. (Ed.), 1997. *Remote Sensing of Large Wildfires in the European Mediterranean Basin*. Springer-Verlag, Berlin.
- Chuvieco, E., Yue, C., Heil, A., Mouillot, F., Alonso-Canas, I., Padilla, M., Pereira, J.M., Oom, D., Tansey, K., 2016. A new global burned area product for climate assessment of fire impacts. *Glob. Ecol. Biogeogr.* 25 (5), 619–629.
- Cochran, W.G., 1977. *Sampling Techniques*. 3rd ed. John Wiley & Sons, New York, USA.
- Cohen, W.B., Yang, Z., Kennedy, R., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync-Tools for calibration and validation. *Remote Sens. Environ.* 114, 2911–2924.
- Conard, S.G., Sukhinin, A.I., Stocks, B.J., Cahoon, D.R., Davidenko, E.P., Ivanova, G.A., 2002. Determining effects of area burned and fire severity on carbon cycling and emissions in Siberia. *Clim. Chang.* 55 (1/2), 197–211.
- Crutzen, P.J., Andreae, M.O., 1990. Biomass burning in the tropics: impact on atmospheric chemistry and biogeochemical cycles. *Science* 250 (4988), 1669–1678.
- Cuzick, J., 1985. A Wilcoxon-type test for trend. *Stat. Med.* 4 (1), 87–90.
- Eidenshink, J., Schwind, B., Brewer, K., Zhu, Z., Quayle, B., Howard, S., 2007. A project for monitoring trends in burn severity. *Fire Ecol.* 3 (1), 3–21.
- Fleiss, J.L., 1981. *Statistical Methods for Rates and Proportions*. 2nd ed. John Wiley & Sons, New York, NY.
- Forbes, A.D., 1995. Classification-algorithm evaluation: five performance measures based on confusion matrices. *J. Clin. Monit.* 11 (3), 189–206.
- García, M.J.L., Caselles, V., 1991. Mapping burns and natural reforestation using thematic mapper data. *Geocarto Int.* 6, 31–37.
- Giglio, L., Lobada, T., Roy, D.P., Quayle, B., Justice, C.O., 2009. An active-fire based burned area mapping algorithm for the MODIS sensor. *Remote Sens. Environ.* 113, 408–420.
- Giglio, L., Randerson, J.T., van der Werf, G.R., Kasibhatla, P., Collatz, G.J., Morton, D.C., et al., 2010. Assessing variability and long-term trends in burned area by merging multiple satellite fire products. *Biogeosci. Discuss.* 7, 1171.
- Global Climate Observing System, 2004. *Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (GCOS-92 (ES))*, WMO/ID No. 1244. Retrieved from: https://www.wmo.int/pages/prog/gcos/Publications/gcos-92_GIP_ES.pdf.
- Goetz, S.J., Bunn, A.G., Fiske, G.J., Houghton, R.A., 2005. Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance. *Proc. Natl. Acad. Sci. U. S. A.* 102 (38), 13521–13525.
- Goodwin, N.R., Collett, L.J., 2014. Development of an automated method for mapping fire history captured in Landsat TM and ETM+ time series across Queensland, Australia. *Remote Sens. Environ.* 148, 206–221.

- Goward, S.N., Huang, C., Zhao, F., Schleeweis, K., Rishmawi, K., Lindsey, M., Dungan, J.L., Michaelis, A., 2016. NACP NAFLD Project: Forest Disturbance History from Landsat, 1986–2010. ORNL DAAC, Oak Ridge, Tennessee, USA <http://dx.doi.org/10.3334/ORNLDAAC/1290>.
- Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. *Remote Sens. Environ.* 122, 66–74.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer, New York, NY, USA.
- Hawbaker, T.J., Vanderhoof, M.K., Beal, Y.-J., Takacs, J.D., Schmidt, G.L., Falgout, J.T., Williams, B., Fairaux, N.M., Caldwell, M.K., Picotte, J.J., Howard, S.M., Stitt, S., Dwyer, J.L., 2017. Mapping burned areas using dense time-series of Landsat data. *Remote Sens. Environ.* 198, 504–522 (this issue).
- Henry, M.C., 2008. Comparison of single- and multi-date Landsat data for mapping wild-fire scars in Ocala National Forest, Florida. *Photogramm. Eng. Remote. Sens.* 74 (7), 881–891.
- Holden, Z.A., Morgan, P., Smith, A.M.S., Vierling, L., 2010. Beyond Landsat: a comparison of four satellite sensors for detecting burn severity in ponderosa pine forests of the Gila Wilderness, NM, USA. *Int. J. Wildland Fire* 19, 449–458.
- Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States – representing a decade of land cover change information. *Photogramm. Eng. Remote. Sens.* 81, 345–354.
- Huang, C., Goward, S.N., Masek, J.G., Gao, F., Vermote, E.F., Thomas, N., Schleeweis, K., Kennedy, R.E., Zhu, Z., Eidenshink, J.C., Townshend, J.R.G., 2009. Development of time series stacks of Landsat images for reconstructing forest disturbance history. *Int. J. Digital Earth* 2 (3), 195–218.
- Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr – temporal segmentation algorithms. *Remote Sens. Environ.* 114 (12), 2897–2910.
- Key, C.H., Benson, N.C., 2006. Landscape assessment: sampling and analysis methods. In: Lutes, D.C., Keane, R.E., Caratti, J.F., Key, C.H., Benson, N.C., Gangi, L.J. (Eds.), FIREMON: Fire Effects Monitoring and Inventory System (General Technical Report RMRS-GTR-164-CD). Rocky Mountain Research Station, Fort Collins, CO.
- Kovalskyy, V., Roy, D.P., 2013. The global availability of Landsat 5 TM and Landsat ETM+ land surface observations and implications for global 30 m Landsat data product generation. *Remote Sens. Environ.* 130, 280–293.
- Mallinis, G., Koutsias, N., 2012. Comparing ten classification methods for burned area mapping in a Mediterranean environment using Landsat TM satellite data. *Int. J. Remote Sens.* 33 (14), 4408–4433.
- Masek, J.G., Huang, C., Wolfe, R.E., Cohen, W., Hall, F., Kutler, J., Nelson, P., 2008. North American forest disturbance from a decadal Landsat record. *Remote Sens. Environ.* 112, 2914–2926.
- Masek, J.G., Goward, S.N., Kennedy, R.E., Cohen, W.B., Moisen, G.G., Schleeweis, K., Haug, C., 2013. United States forest disturbance trends observed using Landsat time series. *Ecosystems* 16 (6), 1087–1104.
- Mazz, J.P., 1996. Analysis of observer variability in the assessment of FLIR performance. *Proceedings of the International Society for Optical Engineering*. 2743. <http://dx.doi.org/10.1117/12.241975>.
- Miller, J.D., Safford, H.D., Crimmins, M., Thode, A.E., 2009. Quantitative evidence for increasing forest fire severity in the Sierra Nevada and Southern Cascade Mountains, California and Nevada, USA. *Ecosystems* 12 (1), 16–32.
- Mitri, G.H., Gitas, I.Z., 2004. A semi-automated object-oriented model for burned area mapping in the Mediterranean region using Landsat-TM imagery. *Int. J. Wildland Fire* 13 (3), 367–376.
- Mitri, G.H., Gitas, I.Z., 2006. Fire type mapping using object-based classification of Ikonos imagery. *Int. J. Wildland Fire* 15, 457–462.
- Morisette, J.T., Baret, F., Liang, S., 2006. Special issue on global land product validation. *IEEE Trans. Geosci. Remote Sens.* 44 (7), 1695–1697.
- Mouillot, F., Schultz, M.G., Yue, C., Cadule, P., Tansey, K., Ciais, P., Chuvieco, E., 2014. Ten years of global burned area products from spaceborne remote sensing – a review: analysis of user needs and recommendations for future developments. *Int. J. Appl. Earth Obs. Geoinf.* 26, 64–79.
- Nichols, J.D., Hines, J.E., Sauer, J.R., Fallon, F.W., Fallon, J.E., Heglund, P.J., 2000. A double-observer approach for estimating detection probability and abundance from point counts. *Auk* 117 (2), 393–408.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R., 2001. Terrestrial ecoregions of the world: a new map of life on earth. *Bioscience* 51, 933–938.
- Omernik, J.M., Griffith, G.E., 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. *Environ. Manag.* 54 (6), 1249–1266.
- Padilla, M., Stehman, S.V., Chuvieco, E., 2014a. Validation of the 2008 MODIS-MCD45 global burned area product using stratified random sampling. *Remote Sens. Environ.* 144, 187–196.
- Padilla, M., Stehman, S.V., Litago, J., Chuvieco, E., 2014b. Assessing the temporal stability of the accuracy of a time series of burned area products. *Remote Sens.* 6 (3), 2050–2068.
- Padilla, M., Stehman, S.V., Hantson, S., Oliva, P., AlonsoCanas, I., Bradley, A., Tansey, K., Mota, B., Pereira, J.M., Chuvieco, E., 2015. Comparing the accuracies of remote sensing global burned area products using stratified random sampling and estimation. *Remote Sens. Environ.* 160, 114–121.
- Palacios-Orueta, A., Chuvieco, E., Parra, A., Carmona-Moreno, C., 2005. Biomass burning emissions: a review of models using remote-sensing data. *Environ. Monit. Assess.* 104 (1–3), 189–209.
- Petropoulos, G.P., Kontoes, C., Keramitsoglou, I., 2011. Burnt area delineation from a uni-temporal perspective based on Landsat TM imagery classification using support vector machines. *Int. J. Appl. Earth Obs. Geoinf.* 13 (1), 70–80.
- Pinty, B., Verstraete, M.M., 1992. GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetatio* 101 (1), 15–20.
- Podur, J., Martell, D.L., Knight, K., 2002. Statistical quality control analysis of forest fire activity in Canada. *Can. J. For. Res.* 32 (2), 195–205.
- Randerson, J.T., van der Werf, G.R., Collatz, G.J., Giglio, L., Still, C.J., Kasibhatla, P., Miller, J.B., White, J.W.C., DeFries, R.S., Kasischke, E.S., 2005. Fire emissions from C3 and C4 vegetation and their influence on interannual variability of atmospheric CO₂ and δ¹³C₂. *Glob. Biogeochem. Cycles* 19, 1–13.
- Richter, R., Schläpfer, D., 2016. Atmospheric/Topographic correction for satellite imagery: ATCOR-2/3 User Guide. (Version 9.0.2. Available online). http://atcor.com/pdf/atcor3_manual.pdf.
- Roy, D.P., Boschetti, L., 2009. Southern Africa validation of the MODIS, L3JRC, and GlobCarbon burned-area products. *IEEE Trans. Geosci. Remote Sens.* 47 (4), 1032–1044.
- Roy, D.P., Frost, P.G.H., Justice, C.O., Landmann, T., Le Roux, J.L., Gumbo, K., Makungwa, S., Dunham, K., Du Toit, R., Mhwandagara, K., Zacarias, A., Tacheba, B., Dube, O.P., Pereira, J.M.C., Mushove, P., Morisette, J.T., Santhana Vannan, S.K., Davies, D., 2005. The Southern Africa fire network (SAFNet) regional burned-area product-validation protocol. *Int. J. Remote Sens.* 26 (19), 4265–4292.
- Roy, D.P., Boschetti, L., Justice, C.O., Ju, J., 2008. The collection 5 MODIS burned area product – global evaluation by comparison with the MODIS active fire product. *Remote Sens. Environ.* 112, 3690–3707.
- Stehman, S.V., 1997. Estimating standard errors of accuracy assessment statistics under cluster sampling. *Remote Sens. Environ.* 60, 258–269.
- Stehman, S.V., 2009. Sampling designs for accuracy assessment of land cover. *Int. J. Remote Sens.* 30 (20), 5243–5272.
- Stehman, S.V., Arora, M., Kasetkasem, T., Varshney, P., 2007. Estimation of fuzzy error matrix accuracy measures under stratified random sampling. *Photogramm. Eng. Remote. Sens.* 73, 165–174.
- Strahler, A.H., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M.C., Herold, M., Mayaux, P., Morisette, J.T., Stehman, S.V., Woodcock, C.E., 2006. Global Land Cover Validation: Recommendations for evaluation and accuracy assessment of global land cover maps (GOFC-GOLD Report No. 25). Joint Research Centre, European Commission, Ispra, Italy.
- Stroppiana, D., Bordogna, G., Carrara, P., Boschetti, M., Boschetti, L., Brivio, P.A., 2012. A method for extracting burned areas from Landsat TM/ETM+ images by soft aggregation of multiple spectral indices and a region growing algorithm. *ISPRS J. Photogramm. Remote Sens.* 69, 88–102.
- Thomas, N.E., Huang, C., Goward, S.N., Powell, S., Rishmawi, K., Schleeweis, K., Hinds, A., 2011. Validation of North American forest disturbance dynamics derived from Landsat time series stacks. *Remote Sens. Environ.* 115, 19–32.
- Thonicke, K., Venevsky, S., Sitch, S., Cramer, W., 2001. The role of fire disturbance for global vegetation dynamics: coupling fire into a dynamic global vegetation model. *Glob. Ecol. Biogeogr.* 10 (6), 661–677.
- Trigg, S., Flasse, S., 2001. An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah. *Int. J. Remote Sens.* 22 (13), 2641–2647.
- Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150.
- Waldrop, T.A., van Lear, D.H., Lloyd, F.T., Harms, W.R., 1987. Long-Term Studies of Prescribed Burning in Loblolly Pine Forests of the Southeastern Coastal Plain. (General Technical Preprint SE-45). Department of Agriculture, Forest Service, Southeastern Forest Experiment Station, Asheville, NC: U.S.
- Waldrop, T.A., White, D.L., Jones, S.M., 1992. Fire regimes for pine-grassland communities in the southeastern United States. *For. Ecol. Manag.* 47 (1–4), 195–210.
- Walters, S.P., Schneider, N.J., Guthrie, J.D., 2011. Geospatial Multi-agency Coordination (GeoMAC) Wildland Fire Perimeters, 2008 (Data Series 612). U.S. Geological Survey, Reston, Virginia.
- Whitman, E., Battlori, E., Parisien, M.A., Miller, C., Coop, J.D., Krawchuk, M.A., Chong, G.W., Haire, S.L., 2015. The climate space of fire regimes in north-western North America. *J. Biogeogr.* 42 (9), 1736–1749.
- Wickham, J.D., Stehman, S.V., Gass, L., Dewitz, J., Fry, J.A., Wade, T.G., 2013. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sens. Environ.* 130, 294–304.
- Zhang, H.K., Roy, D.P., 2016. Landsat 5 thematic mapper reflectance and NDVI 27-year time series inconsistencies due to satellite orbit change. *Remote Sens. Environ.* 186, 217–233.
- Zhu, Z., Woodcock, C.E., 2014. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: an algorithm designed specifically for monitoring land cover change. *Remote Sens. Environ.* 152, 217–234.
- Zimmerman, D.W., Zumbo, B.D., 1993. Relative Power of the Wilcoxon Test, the Friedman Test, and Repeated-measures ANOVA on Ranks.