Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Flooded area classification using pooled training samples: an example from the Chobe River Basin, Botswana

Mitchell P. Braget
Douglas G. Goodin
Jida Wang
James M. S. Hutchinson
Kathleen Alexander

SPIE.

Mitchell P. Braget, Douglas G. Goodin, Jida Wang, James M. S. Hutchinson, Kathleen Alexander, "Flooded area classification using pooled training samples: an example from the Chobe River Basin, Botswana," J. Appl. Remote Sens. 12(2), 026033 (2018), doi: 10.1117/1.JRS.12.026033.

Flooded area classification using pooled training samples: an example from the Chobe River Basin, Botswana

Mitchell P. Braget,[a] Douglas G. Goodin,[a,*] Jida Wang,[a] James M. S. Hutchinson,[a] and Kathleen Alexander[b]
[a]Kansas State University, Department of Geography, Manhattan, Kansas, United States
[b]Virginia Tech University, Department of Fish and Wildlife Conservation, Blacksburg, Virginia, United States

Abstract. In dryland systems, the flood pulse is the driving force in system dynamics but is highly variable in flow volume and landscape inundation features. Remote sensing can provide critical information fundamental to evaluating and forecasting flow behavior and population vulnerability; however, training and classifying an extensive time series of images is labor intensive, limiting the usefulness of these approaches in evaluating flood pulse dynamics and landscape interactions. Here, we provide an alternative approach that relies on only one set of "pooled" training samples for time series image classification and analysis. We test this approach by mapping the flood pulse in a time series of moderate resolution imaging spectroradiometer (MODIS) images from the Chobe River Basin of Botswana for the years 2014 to 2016. MODIS MOD09A1 images collected during the flooding season (February to July) were converted to Kauth–Thomas components, then sampled to form a training pool. Images were then classified using these pooled training samples. Results indicated that classification accuracies obtained using pooled training were statistically indistinguishable from classifications obtained from conventional training. Application of the method to another year's data (2013) also yielded comparably accurate results, suggesting that the training pool method remains robust when applied to image data other than that used to create the training pool. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JRS.12.026033]

Keywords: classification; flood mapping; times series; moderate resolution imaging spectroradiometer.

Paper 180140 received Feb. 14, 2018; accepted for publication May 30, 2018; published online Jun. 30, 2018.

1 Introduction

With the advent of numerical techniques for classifying digital remote sensing imagery, a wide variety of methodologies and approaches have been used to convert these images into usable spatial information.[1–4] These methods can be divided into two fundamental types; supervised and unsupervised classification. With supervised classification, labeled data are used to "train" an algorithm, which then assigns the remaining instances into its most probable category using features extracted from the data. With unsupervised classification, instances are assigned to unlabeled but similar clusters, which are then labeled as a class via posthoc assignment by a human analyst.[5] Throughout most classifications, the spatial units of classification have been pixels, however more recently, other spatial units, such as objects, have been used as the classifiable unit.[6,7] The features used for classification have also varied widely. Reflectance in discrete spectral bands distributed throughout the visible\near infrared (VNIR) and short-wave infrared (SWIR) regions of the spectrum are the most common inputs to classification algorithms;

*Address all correspondence to: Douglas G. Goodin, E-mail: dgoodin@ksu.edu

1931-3195/2018/$25.00 © 2018 SPIE

however, image texture, object spatial properties, derived spectral indices, and even temporal indices have all been used to form classification feature space.[8–11] Finally, hybrid techniques, which feature a combination of various approaches (e.g., supervised and unsupervised methods) have also been widely used for remote sensing classification.[8,9] In general, these approaches have been successful in allowing remote sensing scientists to make accurate and reliable thematic maps and information layers for the Earth′s surface. Although research into improving classification methods is on-going, the methods used are fundamentally effective.

Regardless of the methods used, one characteristic of image classification is that it is both time-consuming and labor intensive. If the supervised approach is used, considerable time and effort must be expended to adequately train the algorithm. For unsupervised methods, the time bottleneck tends to be in the labeling step, where an expert analyst manually assigns clusters to classes. If only a single image, or a small number of images are being classified, the time spent on training may not be a serious constraint, but what about circumstances in which a large num-ber of images need to be classified to a similar or identical set of land user or cover classes? Under such circumstances, the time needed to train and classify each image individually could become burdensome. Increasingly, investigations into ecosystem dynamics require that land-scape change data should be retrieved at relatively small time steps (e.g., monthly, weekly, or even daily) over considerable lengths of time, resulting in tens or even hundreds of images to classify. In these instances, current methodological approaches (supervised or unsupervised classification) limit the capacity for remote imagery to contribute to these investigations.

There are several potential solutions to the problem outlined above. For example, Li and Sheng[10] developed a self-adaptive unsupervised approach to mapping open water, which has been successfully used to monitor lake inundation dynamics in various regions.[11–15] This approach is highly adaptable, seasonally robust, and computationally efficient but only applicable in situations where a binary classification such as water versus nonwater is needed. Another approach, applicable when more classes are desired, is to classify based on spectral reflectance patterns (i.e., signatures) drawn from a pre-existing library. This approach is especially common in geological remote sensing, where spectral reflectance from rocks or minerals is algorithmically compared with known spectra (measured in situ or in the lab) and identified by matching the unknown spectral reflectance to the closest match in a library of potential identities.[16] This approach can be considered a type of supervised classification, but one in which the training data already exist, in the form of the spectral library. The advantage with the spectral library approach lies in the fact that a new training set does not have to be collected for every image to be classified, a clear savings of time and effort. The potential pitfalls of this method are also clear; however, whereas geological materials are relatively invariant in their spectral response, other Earth surface covers, such as vegetation and water are much more variable in both time and space, and thus harder to summarize by a single spectral library entry. For materials or surfaces that change over time, however, entities in the spectral library would need to encompass a wider range of potential spectral response. For these types of surface, a variation on the spectral library approach might be used, in which spectral library entries are replaced with groups of training samples, similar to a sample from a specific image but collected from multiple images across time. As this approach is similar to the idea of "pooling" data for analysis, these training samples might be termed pooled samples.

Use of preselected "pools" of training samples might substantially reduce the amount of time needed to classify many images, where the land cover to be classified has reflectance properties that stay relatively consistent across space and\or time. Flooding may be an example of this type of situation. Floods are often seasonal, widely distributed, and persistent over long-time periods, meaning that multiple images may be needed to fully characterize both its extent and duration. Floods are also notable for their impacts on human health, welfare, and safety, and there is a pressing need to be able to dynamically characterize these hydrological phenomena. This is especially true in dryland flood pulse dominated systems, where, the flood pulse is the driving force in system dynamics but is highly variable in flow volume and landscape inundation features, influenced by the character of the flood event and the geomorphological characteristics of the system.[17]

In Northern Botswana, the Chobe River is the only surface water in this dryland region. Flood pulse dynamics in this river system have important influences on both water quality and diarrheal

disease in children under five.[18,19] Regional scale remote sensing offers the only practical means for tracking and mapping flood dynamics in this system, but as in other systems, its use has been challenging, especially when the flooding creates temporary wetlands with emergent vegetation. Detecting the presence of standing water beneath flooded vegetation canopies using VNIR sensors has proven to be especially challenging.[20–23] Both thermal[24] and microwave [e.g., synthetic aperture radar (SAR)] images[25–28] have been used with success to attempt to overcome this limitation. As the flooding season in the Chobe River Basin (CRB) can be quite long (typically extending from February to July) and the inundated areas can change significantly over this time, it is also useful to be able to assess the temporal dynamics of flood coverage at frequent intervals during the flood season. High temporal-resolution optical images, such as acquired by the moderate resolution imaging spectroradiometer (MODIS) aboard the Terra and Aqua satellites, have been used to effectively capture inundation dynamics at 10 day or finer frequencies across cloud-prone regions.[14,29] All-weather conditions were further enabled by SAR imagery (e.g., ENVISAT, RADARSAT, ERS, ALOS PALSR, but their extraction of water extent can be compounded by surface roughening from wind and turbulence, which lead to an increase in the radar backscatter to the sensor.[30–32] Moreover, all remote-sensing data sources (optical, thermal, and microwave) experience the same fundamental challenge when used to map flooding over time, in which they require repetitive analysis of a number of images to extract similar land surface information from each of them. Addressing this need leads us to confront the problem discussed in Sec. 1, how to classify a series of images in a timely and efficient manner? In this paper, we evaluate the use of pooled classification methods as a method for more rapid mapping of floods and flood-inundated wetland using remote imagery of the CRB. Our goal here is not to provide maps of specific flood pulse events, but rather to test whether the pooled classification method is sufficiently accurate to warrant its further use for mapping dynamic pulse flooding events through time.

2 Study Area and Background

2.1 Chobe River Basin

The CRB is a seasonal marshland located on the border between Namibia and Botswana (Fig. 1). The CRB is a subbasin of the larger Upper Zambezi Basin with an area of roughly $4000\ \mathrm{km}^2$ and generally flat terrain with relief varying between 830 and 1050 m above the mean sea level.[24] The CRB is predominately a dryland system, with the Chobe River providing the only permanent surface water. River flow in the CRB is driven by the regional and local precipitation. Along with the Okavango Delta, the Chobe River is at the center of the Kavango–Zambezi Transfrontier conservation area (KAZA), the largest conservation area in the world, encompassing more than $170{,}000\ \mathrm{km}^2$ and contains the world's largest elephant population.[24] The KAZA provides a wildlife corridor, which contains water sources for agriculture and grazing.

Vegetation in the CRB is a heterogeneous mixed woody-herbaceous savanna ecosystem, with scrub woodlands in higher elevation zones, and bushland, shrub, and thicket mosaics in lower elevations.[33] Soils in the region are characterized as thin alluvial and volcanic in origin.[34] The CRB is situated in three ecoregions; the Zambezian-flooded grasslands, Zambezian and Mopane woodlands, and the Zambezian Balikiaea woodlands. The Zambezian-flooded grasslands ecoregion covers a large part of the CRB, especially the Zambezi River, and is a flooded grasslands and Savanna biome.

The hydrology of the CRB is characterized by flood pulses from the Zambezi and the Linyanti rivers.[35] A flood pulse occurs when there is a substantial amount of water going into a river system, water rises above the banks, and vegetation is inundated for periods of times ranging from a few weeks to months.[36] Flood pulses deliver valuable nutrients and water that are vital to plant development and are a principal driving force for the productivity and types of biota in the river-floodplain region.[37] These pulse floods happen during the rainy season, which occurs between November and April. The floods consistently start at the end of February when the Zambezi flood pulse enters the Zambezi wetlands.[24] Burke et al.[24] also found that the best predictor of flooding extent in the CRB is the discharge of the Zambezi River 64

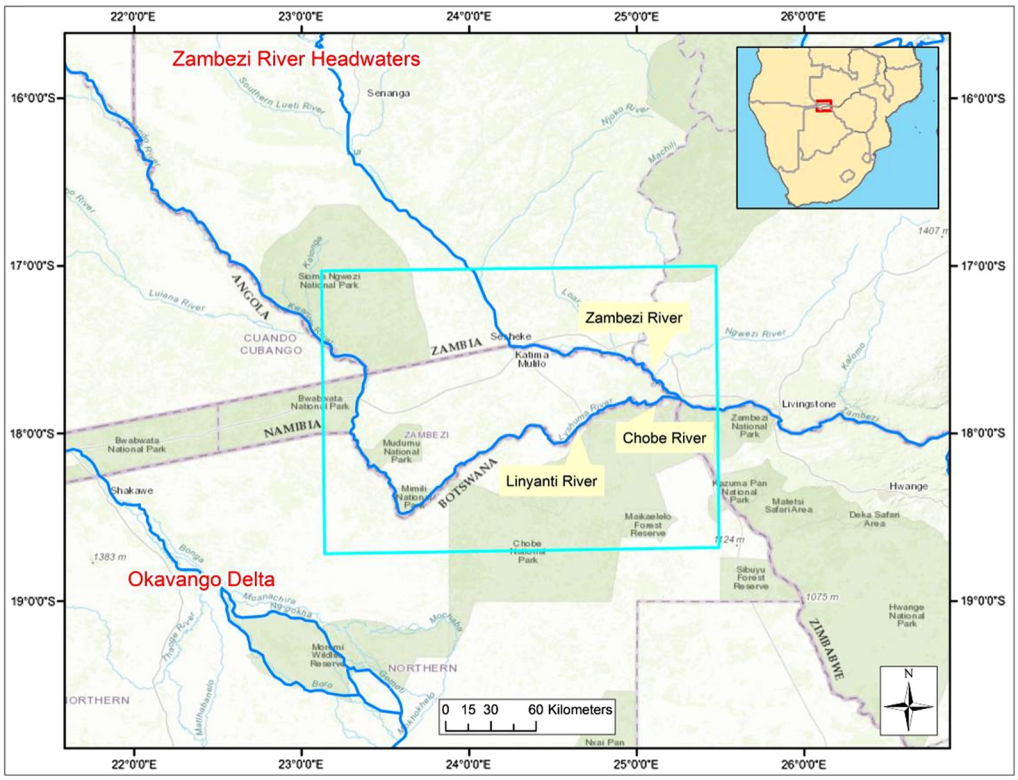

Fig. 1 Image map of the Chobe River region in central Africa. The study area is within the cyan rectangle.

days prior to flooding. Thus, the peak of these floods occur at the end of the rainy season and beginning of the dry season (April) when the peak discharge of the Zambezi River reaches the CRB. The CRB can experience a second flood pulse in the months of June and July when the Kwando River on the western side of the CRB has its peak discharge, which is outside of the traditional rainy season. If the flood pulse is high enough, water can move from the Kwando River, through the Linyanti channel, then through Lake Liambezi, and into the CRB.[38] However, this is not a frequent event, and Lake Liambezi often remains dry.

3 Classification Method: Pooled Training Approach

For this classification, we propose to use a form of supervised classification method in which a pooled training sample, analogous to the spectral library discussed earlier, provides the classification feature space and decision criteria for all images in a particular time series. The classification method is applied to a series of coarse resolution multispectral imagery.

3.1 Data

MODIS MOD09A1 imagery for the study area was acquired from the LP DAAC website[39] and has previously been used to map flood regions.[40–42] It has the advantage of providing temporal and spatial resolution (500 m) that is superior to other available thermal or microwave data. The MOD09A1 product consists of seven reflectance bands, including coverage in the visible\near-infrared (VNIR) and SWIR. Each composite consists of compiled daily images over eight days, where pixels are selected based on quality, cloud, and viewing geometry, until the highest quality single value per pixel remains to be used in the composite. Imagery was collected from early February to early July for the years 2014, 2015, and 2016 with a total of 49 usable composites collected during the analysis time period. Ten images were excluded from the classification due to excessive cloudy conditions, with four in 2016, three in 2015, and three in 2014. These dates

Table 1 Acquisition dates for images used in this analysis. Time windows are defined and discussed in Sec. 3.4.

Time window	2014	2015	2016
1	14 March	14 March	10 February
2	17 May	25 May	24 May
3	18 June	26 June	25 June

were chosen to bound the yearly peak flood and extents. Each band in the MODIS dataset consists of 16-bit signed integer reflectance values that range from −100 to 16,000 with bands in each composite multiplied by 0.0001 to rescale to true reflectance. Each of the seven bands for a given date were stacked into one image. The study area was subset to the CRB based on previous studies of flood dynamics in this system.[43]

Although classifying all images within the flooding season from each of the years was the ultimate goal of our research, we did not use the entire data series to develop our methodological approach. Rather than evaluating the entire 15 image sequence for each of the 3 years, we selected images from three dates, spaced at the beginning, end, and middle of each flooding season, a total of nine images. Test images were selected at roughly the same time in each window (Table 1). The criteria for dividing the season into three time periods are explained in Sec. 4. These images were then classified using two variations of the pooled training method. In addition, one set of images was classified using conventional supervised classification to serve as a standard of comparison for the pooled methods.

3.2 Data Preparation

Following preprocessing, each image was converted into the Kauth–Thomas tasseled cap transformation using coefficients derived for the MODIS instrument (see Table 2).[44] The Kauth–Thomas (K–T) transformation is derived using the Gram–Schmidt orthogonalization method to "rotate" the features of a multivariable dataset into a new basis, whose components are linearly independent.[45] It is primarily used to enhance biophysical properties in the imagery by creating synthetic components corresponding to the brightness, greenness, and wetness properties of the data.[44] The K–T transform converts raw reflectance values into a few, more easily interpretable components, which can be used as input into classification algorithms. In our application, the K–T transformation was used to reduce the number of classification features from six (the full number of MODIS bands in the MOD09A1 data product) to three. It was also used because the criteria for defining the classes in our classification scheme was based largely biophysical surface properties, such as greenness and wetness, which can readily be extracted using features derived from the Kauth–Thomas transformation.

3.3 Land Cover Classes

Three cover classes were used for this classification; open water, emergent vegetation, and nonflooded areas. Direct observations in the study area were used to define the characteristics of the

Table 2 Coefficients for calculating the brightness, greenness, and wetness K–T components, derived for MODIS data.

	Red (B1)	NIR1 (B2)	Blue (B3)	Green (B4)	NIR2 (B5)	SWIR1 (B6)	SWIR2 (B7)
Brightness	0.4395	0.5945	0.246	0.3918	0.3506	0.2136	0.2678
Greenness	−0.4064	0.5129	−0.2744	−0.2893	0.4882	−0.0036	−0.4169
Wetness	0.1147	0.2489	0.2408	0.3132	−0.3122	−0.6416	−0.5087

three land cover categories and to identify areas representative of each type categorized as flooded and nonflooded. Flooded areas were of two types, visible surface water with little to no vegetation coverage (open water) and emergent vegetation defined as a mixture of water with aquatic vegetation protruding above the surface (emergent vegetation). Areas outside of these two types were categorized as nonflooded land cover. Open water and emergent veg-etation land covers were kept separate because the emergent vegetation class contains a mixture of both open water and vegetation overlying water or saturated soil. In effect, it resembles marsh-land. Open water and emergent vegetation also exhibit different tasseled cap component values in imagery, which, if combined, could result in misclassification.

3.4 Training and Classification Method

The most obvious challenge to implementing this methodology for rapid classification of a series of images is development of the pooled training dataset. This was done in two steps. First, it was necessary to determine if the spectral response of the surface remains roughly invariant through time, in which case only a single training pool is necessary, or if it varies enough to warrant splitting the training into two or more pools.

To assess this invariance, we plotted trajectories of the mean values and variance for each of the three KT components (brightness, greenness, and wetness) across the entire flooding season (February to December) for the years 2014 to 2016. Mean values for each cover type and year (Fig. 2) show some variation across the flooding season, as well as from year to year. This suggests that one single set of pooled training samples may not adequately be used to classify all images over an entire flooding season. Based on the shapes of the trajectories, we subjectively split the flooding season into three subperiods, indicated by the vertical lines in Fig. 2. In each of these time windows, the temporal pattern of variation in the three K–T components shows roughly coherent behavior. From day of year (DOY) 33 to approximately DOY 91, the land covers of emergent vegetation and nonwater exhibit decreasing values, whereas open water values remain consistent. Starting at DOY 91 to DOY 155, all three land covers exhibit relatively constant values or show very slight changes in values from composite to composite. After DOY 155, both emergent vegetation and the nonflooded class show increasing values of brightness and decreasing wetness, whereas open water values remain consistent. These subperiods are indicated by vertical lines in Fig. 2.

Once the number of time windows was established, the second step in the classification was to select training samples from within the time windows. Locations for extracting training samples were selected using a combination of field assessment and ancillary imagery. Obviously, field survey can only be used for imagery whose acquisition date overlaps the active period of the study. That is, field survey cannot be done retroactively. However, ancillary data sources are often available retroactively and can be used as a source of training data provided they can be correctly interpreted. In our study, field surveys conducted during the 2016 flood season were compared with visual interpretation of Landsat 8 OLI data from the same time frame. Using this approach, we were able to develop skill in visual identification of our three cover types from moderate resolution imagery. We then used other retrospective Landsat data, whose acquisition time corresponded to the MODIS composite data, to select training sites from the MODIS data by visual interpretation, and also to learn to recognize our three cover types on MODIS data. Once we were confident that we could visually recognize our three classes on the MODIS data, we selected further training sites by visualize inspection. Note that while training sites were selected using the procedure described above, the actual training data were from K–T transformed images.

Using the 3 years of data, each divided into three time windows, we tested two variations of the pooling classification technique, which are referred to as method I and method II, respectively. In method I, separate training pools were created for each of the three time windows. Training samples from all 3 years were then taken from several images within each time window, to ensure that the samples were as representative as possible for all years and dates within the window. Note that training samples consisted only of representative pixels from a single image. In other words, we did not select a training site, then use all pixels associated with that site across the entire time series. Instead, we used pixels only from the specific image from which that site

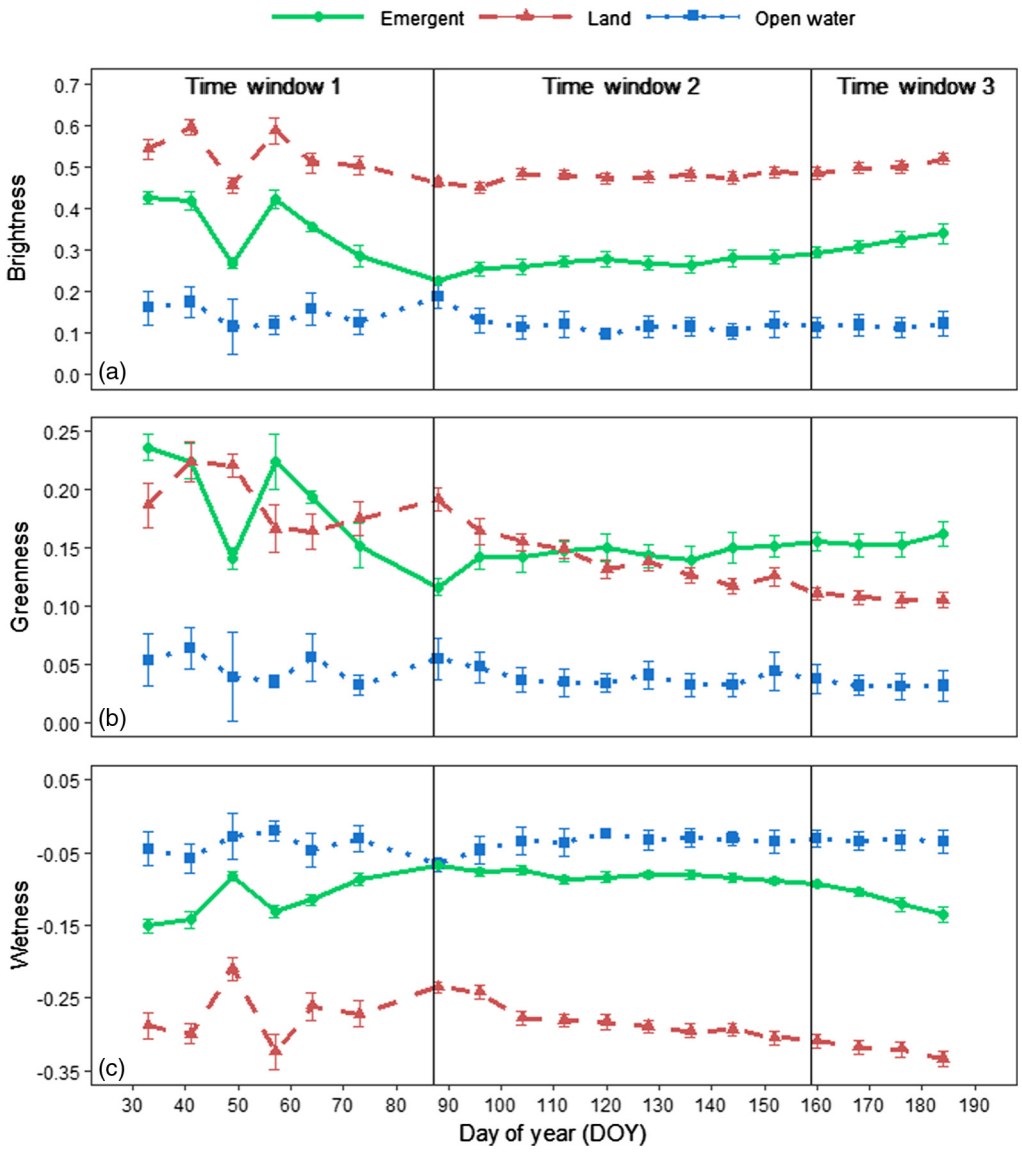

Fig 2 Mean seasonal tasseled cap component values for open water (blue), emergent vegetation (green), and nonflooded (red), from 2014 to 2016. Subfigures (a)-(c) display the brightness greenness, and wetness component, respectively, of the land cover values.

was selected. To reduce temporal bias of the training data, training samples were selected from images at the beginning, middle, and end of each time window. Approximately 25 to 30 training samples per class were selected from each individual image, yielding about 80 training samples per class. Method II was similar to method I, but only the second (middle) time window was used to collect samples for the training pool (see Fig. 2 for the time windows). We used the middle time window on the assumption that it would be most representative of the entire flooding sea-son. Collection of samples for the training pool for method II was also similar to method I; two-to-four samples were collected for each cover type, from images throughout the time series. The effectiveness of these two classification approaches was evaluated in two ways. First, we quantitatively compared classification results for selected dates from the three image years against the results from conventional training and classification (that is, use of training samples from each of the 3 years). Second, as a further evaluation for each method, we tested them by classifying a fourth year from the MODIS image archives, using the training sample pool developed for the 3 years from 2014 to 2016.

All classifications were done using classification trees. Classification trees are a nonparametric method for recursively partitioning a categorical dataset into successively more homogeneous

nodes using some splitting criteria, until an optimum classification is achieved. For this application of decision tree, we used a splitting criteria based on the Gini coefficient, which favors larger partitions in the data and was therefore preferable for a classification scheme with only a small number of classes.[46] To avoid overfitting, the decision tree was pruned using cross-validation error.[46] For this application, we chose the classification tree, because it has been shown to be effective for classification problems in remote sensing, and because it is less sensitive to the distributional structure of the input data.[47] The latter was of particular significance since the distributional characteristics of pooled, transformed image data may not conform to the assumptions of normality required for parametric classifiers. All calculations and classifications were done using the Rpart package within the R statistical programming languages.[48]

Classification accuracy was quantified by constructing a confusion matrix for each classification, from which categorical and overall accuracy (OA), along with the Kappa statistic and its variance were calculated. Each of these assessments was based on an independent sample drawn from the various classified images.[49] OA was used to assess the success of each method for placing pixels into their appropriate categories. The Kappa statistic (and its variance) were used to assess how well a classification performed relative to random chance. To test the relative merits of each method, it was also necessary to quantitatively compare classification accuracies of the various methods, to determine which was the most effect. This was done by pairwise z-tests, as suggested by Foody.[50] For these tests, the z-statistic is calculated as

$$z = \frac{\underline{x}_i - \underline{x}_j}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}, \tag{1}$$

where x_i and x_j are the numbers of correctly classified samples for two classifications i and j, n_i and n_j are numbers of samples in each classification, and

$$\bar{p} = (x_i + x_j)/(n_i + n_j). \tag{2}$$

4 Results

4.1 Method I

In method I, each of the three time windows defined based on the seasonal trajectory of the K–T components were evaluated separately for each of the three years of record (2014 to 2016). Comparison of the nine error matrices resulting from this classification shows some trends in the performance of the classifier relative to the three land classes in the classification scheme. In general, open water was the most accurately mapped class, based on the relatively small occurrence of confusion between it and the other two classes (Table 3). For many of the classifications, there was no confusion between classified and observed pixels from the validation sample, either as errors of omission (shown column-wise in the matrices) or commission (row-wise, on the matrices). Perhaps not surprisingly, the greatest confusion occurred between the open water class and the emergent class. In time window 2 of the 2014 classification, 10 pixels classified as open water were, in fact, members of the emergent class, the most confused instance among these classes for the method I classification. It should be noted that this highest level of interclass confusion still only represents a 17% commission error.

The emergent and nonflooded classes were the most confused classes using method I, a result that is consistent with other applications where flooding was mapped optically.[51] The largest amounts of confusion (and thus the highest errors) tended to occur in the first time window defined for the calculation. This was particularly true for the 2016 classification, 14 pixels from the emergent category were misclassified as land, a user error of 35% (where user error is determined by subtracting user accuracy from 1.0). Interestingly, pixels that were observed to be emergent vegetation in the validation sample were rarely confused with open water, as indicated by the higher producer accuracy (PA) values in the matrices.

Table 3 Error matrices for pooled classification Method I for all 3 years of analysis. Abbreviations are NF, not flooded; EV, emergent vegetation; OW, open water; UA, user accuracy; and PA, producer accuracy.

		2014					2015					2016				
		Observed														
		NF	EV	OW	Total	UA	NF	EV	OW	Total	UA	NF	EV	OW	Total	UA
Time window 1	NF	40	1	0	41	0.98	60	1	0	61	0.98	57	0	0	57	1.00
	EV	11	29	0	40	0.73	9	28	0	37	0.76	14	28	1	43	0.65
	OW	2	12	35	49	0.71	2	3	45	50	0.90	1	3	36	40	0.90
	Total	53	42	35	130		71	32	45	148		72	31	37	140	
	PA	0.75	0.69	1.00			0.85	0.88	1.00			0.79	0.90	0.97		
Time window 2	NF	36	0	0	36		33	0	0	33	1.00	36	0	0	36	1.00
	EV	3	45	2	50		9	36	0	45	0.80	4	38	1	43	0.88
Predicted	OW	0	10	50	60		3	5	36	44	0.82	0	8	42	50	0.84
	Total	39	55	52	146		45	41	36	122		40	46	43	129	
	PA	0.92	0.82	0.96			0.73	0.88	1.00			0.98	0.82	0.98		
Time window 3	NF	48	3	0	50		45	0	0	45	1.00	37	0	0	37	1.00
	EV	5	38	0	43		4	37	1	42	0.88	4	44	1	49	0.90
	OW	0	7	42	49		2	0	41	43	0.95	1	6	47	54	0.87
	Total	53	47	42	142		51	37	42	130		42	50	48	140	
	PA	0.91	0.81	1.00			0.87	1.00	0.98			0.88	0.88	0.89		

OA varies somewhat between the various time windows and years, from a high of 95% in the third window of 2015 to a low of 80% in the first time window of 2014 (see Table 4).

In general, classification accuracy was highest in 2015, which, based on visual inspection of the data, is most likely due to the presence of cloud cover in the three images selected for detailed analysis from the other 2 years (but we note that the selected images were chosen in part for their lack of clouds, hence results from other dates would have been similar, or perhaps even lower accuracy). The Kappa values and their variances show that all classifications performed significantly better than what would be expected from chance agreement. Based on the similarity of OA and Kappa values across all nine of the classifications, it is difficult to conclude that there is any systematic pattern in accuracy over the 3 years of the analysis. That is, none of the time windows appear to be significantly more or less accurate than any of the others, although this was not

Table 4 OA and Kappa values for classification method I for each of the 3 years analyzed. Exact image dates for the time windows in each of the 3 years are given in Table 1.

	2014		2015		2016	
Time window	OA	Kappa (Var)	OA	Kappa (Var)	OA	Kappa (Var)
1	0.80	0.70 (0.0027)	0.90	0.84 (0.0010)	0.86	0.81 (0.0017)
2	0.89	0.84 (0.0014)	0.86	0.79 (0.0022)	0.90	0.87 (0.0012)
3	0.90	0.84 (0.0015)	0.95	0.92 (0.0010)	0.91	0.87 (0.0013)

formally tested. However, it is noteworthy that in two of the three years, time window 1 was the least accurate of the three (only 2015 is an exception). This could possibly be explained by the fact that this time window shows the most variability in K–T component values (see Fig. 2), and is, therefore, most susceptible to unrepresentative training values.

4.2 Method II

Classification method II tested the efficacy of using training samples from only one time period to classify all of the images from a given flooding season. In order for this method to be effective, K–T component values must remain roughly similar for each cover type over the entire year, and be similar to those of the second time window, from which the training samples were collected. If these circumstances hold, the classification rules derived from one time period would be effective in either of the other two. We evaluated method II on only one of the three classification years, 2015, because it generally had less cloud contamination than did the other 2 years. For this classification, the pool of training samples was drawn only from the second time window, because this window generally had the most stable values, with least variance in the sample set (Fig. 2). Because the training samples came only from the second time window, they were not tested for that time window—this test would essentially be identical to the results from method I.

In general, the classification results using method II showed more interclass confusion (Table 5) and numerical accuracy metrics were lower than those from method I, although not drastically so (Table 6). As with method I, each of the classifications was significantly better than random, based on Kappa values. Method II had the lowest performance for the first time window, with an OA over 10% points less than method I. A visual comparison between the two regions on Fig. 2 indicates that time window 1 is the most variable of the three windows, and is

Table 5 Error matrices from the method II classifications. Abbreviations are NF, not flooded; EV, emergent vegetation; OW, open water; UA, user accuracy; and PA, producer accuracy.

			Observed				
			NF	EV	OW	Total	UA
Time window 1	Predicted	NF	60	2	0	62	0.97
		EV	17	40	2	59	0.68
		OW	0	20	55	75	0.73
		Total	77	62	57	196	
		PA	0.78	0.65	0.96		
Time window 3		NF	60	6	0	66	0.91
		EV	11	54	4	69	0.78
		OW	2	11	57	70	0.81
		Total	73	71	61	205	
		PA	0.82	0.76	0.93		

Table 6 OA and Kappa values for 2015 classifications using method II.

Time window	OA	Kappa (Var)
1	0.79	0.687 (0.0434)
3	0.83	0.751 (0.0309)

therefore the most susceptible to a less representative training sample. As with method I, the majority of the interclass confusion was between the open water and emergent vegetation classes, which in light of the error patterns in method I, is not surprising.

4.3 Comparison of Methods I and II with Conventional Supervised Classification

To evaluate whether use of pooled training samples yields accuracy comparable with conventional training, the three images from the 2015 dataset were also subjected to supervised classification using training samples collected only from those images. Visual comparison of the two methods shows similar results, which also correspond well to a visual interpretation of the imagery (Fig. 3). Quantitative comparison, however, shows that method I produces significantly more accurate classifications.

Somewhat surprisingly, the classification accuracy of the 2015 images obtained using conventional training was not as accurate as either of the pooled training methods. Comparing the

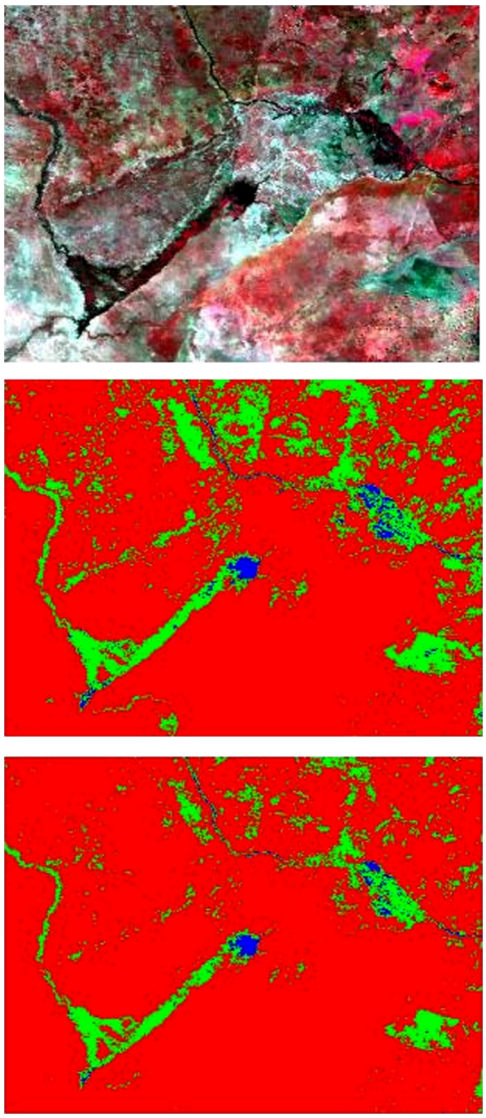

Fig. 3 Classification results for March 2015 for visual comparison. The original image (not converted to K–T components) is at top. The center classification was made using method I, the bottom classification is from method II. The nonwater class is shown in red, emergent vegetation is green, and open water is blue.

Table 7 Classification error matrices for conventional supervised classification, 2015 data set. Abbreviations are NF, not flooded; EV, emergent vegetation; OW, open water; UA, user accuracy; and PA, producer accuracy.

			Observed				
			NF	EV	OW	Total	UA
Time window 1	Predicted	NF	66	0	0	66	1.00
		EV	14	43	2	59	0.73
		OW	4	6	65	75	0.87
		Total	84	49	67	200	
		PA	0.79	0.88	0.97		
Time window 2		NF	56	0	0	56	1.00
		EV	24	26	0	50	0.52
		OW	3	26	55	74	0.74
		Total	83	42	55	180	
		PA	0.67	0.62	1.00		
Time window 3		NF	64	2	0	66	0.97
		EV	10	49	1	60	0.82
		OW	3	5	52	60	0.87
		Total	77	56	53	186	
		PA	0.8	0.88	0.98		

2015 OA and Kappa results for conventional classification (Tables 7 and 8) with both method I (see the 2015 column in Table 4) and method II (Table 6) shows that in some cases, the conventional classification method yielded notably lower values than did the pooled methods. Although not necessarily anticipated at the start of this analysis, this result is consistent with previous work where spectral signatures were generalized for supervised classification.[52] This result might arise from the fact that the number of training samples from the pooled samples (≈100 per class, per image, or about 300 for each pool) is larger than the ≈100 samples per class collected from each of the images classified using the conventional training methods. Generally, a larger number of training samples will provide a more reliable set of decision rules, especially when using a decision tree classification approach.[53] This result also reinforces the earlier point that the classification feature values (i.e., the K–T components) stayed roughly uniform across time, which further supports the use of pooled training samples when this condition holds.

Quantitative comparison of the classifications used the methods outlined in Sec. 3.4 further indicates that use of pooled training provided equal or better results to the conventional methods. A comparison of the pooled method I with method II (Table 9) shows that method I was

Table 8 OA and Kappa values for 2015 classifications using conventional supervised classification.

Time window	OA	Kappa (Var)
1	0.87	0.751 (0.039)
2	0.76	0.641 (0.048)
3	0.89	0.830 (0.035)

Table 9 Pairwise comparison analysis for method I (pooled training with three time windows) versus method II (pooled training with one time window) for 2015 analysis. Number correct and n are derived from the error matrix for each classification (tables K and M) and are the source of the X and N terms in Eq. (1).

Time window	Method I			Method II			Z	P
	Number correct (X)	N	OA	Number correct (X)	N	OA		
1	133	148	0.899	155	196	0.798	2.53	0.0014[a]
3	123	130	0.946	171	205	0.834	2.88	0.0039[b]

[a]Significant at α ¼ 0.05.
[b]Significant at α ¼ 0.01.

Table 10 Pairwise comparison analysis for method I (pooled training with three time windows) versus conventional supervised for 2015 analysis. Number correct and n are derived from the error matrix for each classification (tables K and L) and are the source of the X and N terms in Eq. (1). None of the comparisons reach statistical significance at α ¼ 0.05.

Time window	Method I			Conventional			Z	P
	Number correct (X)	N	OA	Number correct (X)	N	OA		
1	133	148	0.898	174	200	0.870	0.65	0.516
2	105	122	0.861	137	180	0.761	1.98	0.048
3	123	130	0.946	165	186	0.887	1.62	0.050

significantly better than method II for the third time window (i.e., significant at α ¼ 0.01) and marginally significant (at α ¼ 0.05) for the first time window. Comparison of method I with conventional classification (Table 10) showed no significant difference between the two, despite the higher OA values from method I.

4.4 Application of Pooled Training to 2013 Imagery

As a final evaluation of the effectiveness of the pooled training samples, we classified image data from a year (2013) that was not included in the original analysis, using the method I approach. These images were classified using the pooled training data collected from the 2014 to 2016 image series; no additional training samples were collected. Accuracy results for this classification were similar to the ones for the 2014 to 2016 data (see Tables 11 and 12). As with the previous classification, the best results, in terms of OA, were obtained from the second time window, and the lowest accuracy was associated with the first window. For each classification, the Kappa values and variances show that the accuracy achieved was significantly better than random assignment.

5 Discussion and Conclusions

In this study, pooled training methods were effective for classifying open water, water with emergent vegetation, and nonflooded land cover in the CRB study area from selected MODIS images drawn from each of three years from 2014 to 2016. The pooled training method was also successful for classifying images from 2013, a year from which images were not used to develop the training data pools. Results from the application of the training pool method to images collected during another flood year (2013) are encouraging as they suggest that it may be feasible to classify individual flooding images over the course of a season using only pooled samples from other years, without the need for additional training.

Table 11 Error matrices for classification of 2013 images, done using pooled training samples collected from the 2014–2016 imagery. Abbreviations are NF, not flooded; EV, emergent vegetation; OW, open water; UA, user accuracy; and PA, producer accuracy.

			Observed				
			NF	EV	OW	Total	UA
Time window 1		NF	30	0	0	30	1.00
		EV	4	51	6	61	0.84
		OW	0	17	61	78	0.78
		Total	34	68	67	169	
		PA	0.88	0.75	0.91		
Time window 2		NF	30	1	0	31	0.97
		EV	3	61	2	66	0.92
	Predicted	OW	0	5	52	57	0.91
		Total	33	67	54	154	
		PA	0.91	0.91	0.97		
Time window 3		NF	52	8	0	60	0.87
		EV	2	28	0	30	0.93
		OW	1	6	53	60	0.88
		Total	55	42	53	150	
		PA	0.95	0.67	1.00		

Table 12 OA and Kappa values for 2015 classifications using conventional supervised classification.

Time window	OA	Kappa (Var)
1	0.84	0.75 (0.044)
2	0.93	0.89 (0.032)
3	0.89	0.83 (0.040)

Although both methods were effective, they were not equally so. Method I, in which separate pools of training samples from each of three time windows were used to classify images from within that time window, was the best performer of the two pooled training methods we tested. Classification accuracies from method I were statistically superior to those of method II, and were statistically not distinguishable from results obtained using the traditional supervised classification approach of collecting training samples from each image to be classified. Based on this, if the pooled training approach is to be used for classifying a series of images, we recommend dividing the pools into regions, in which the training features to be used are more uniform through time.

Although the pooled classification approach was effective in this instance, it is appropriate to consider a few caveats related to its application. The classifications considered here are notable for the spectral contrast between classes, the relative temporal homogeneity of those classes, and the fact that only a relatively small number of classes were used. The latter condition makes the probability of placing a pixel in the correct class inherently higher regardless of the method used

as the number of ways an error can be made is smaller with fewer classes. Addressing this concern is one of the reasons that the Kappa coefficient was calculated for each classification. Kappa provides a way to quantify accuracy that takes into account the number of classes and hence the "degree of difficulty" of a particular classification. Each of our classifications yielded Kappa values indicating that the results were significantly higher than expectation based on random chance.

The first two conditions above (spectral contrast and temporal homogeneity) are significantly more limiting potential constraints on the use of training pools. Clearly, for training pools to be effective, the training signatures in the pool must be applicable across a wide range of images. For images drawn from a time series, this means that the reflectance patterns that characterize one type of surface must either remain similar over time, or be divisible into a relatively small number of subsets that remain similar across time. For surfaces dominated either by nonliving materials (e.g., inorganic materials, soils, water, etc.), or by vegetated surfaces of low temporal variation (such as nondeciduous forest, although this could be questioned), this situation may hold. However, even with the relatively time-invariant cover classes we used (water, water with emergent vegetation), there was enough temporal variability to warrant the use of more than one training pool per year. For vegetated surfaces that undergo distinct seasonal\phenological variation, development of a meaningful training pool might require a larger number of time windows and larger number of classes, which could potentially negate the time savings that are a prime benefit of the pooled classification method.

Following from the need to define time windows to develop adequate training pools is the problem of how to define these time windows. The underlying purpose of the time windows is to divide the image series into roughly homogeneous regions based on trajectories of classification features. In our analysis, this was done somewhat subjectively, based on "by eye" comparisons. Results from the classifications consistently showed that the first time window was the least accurately classified, and the trajectories of its classification features (Fig. 2) clearly show the most change and the most variability through time. The other two time windows show much more uniform behavior suggesting that a different division scheme might have been used, or that combining these two windows may have been appropriate. Alternatively, a more systematic method might have been used to divide the trajectories of classification features into time windows.

Despite the potential limitations in applying the pooled classification method with multiple time windows, it is clearly an effective alternative for classification tasks in which a large number of potentially similar images are to be classified. Recall from Sec. 2 that the purpose of our analysis was to produce maps of flooded areas to assess their vulnerability to water-borne diseases. Typically, flooding in the CRB is persistent over several months, so even if composited data are used to map it, it is still likely that about 15 images per year would require classification. For the 3-year duration of our study, this would mean classifying ?45 images. Ultimately, this analysis may be extended to encompass the entire duration of the MODIS mission, meaning that several hundred images would need to be classified. Clearly, training a classifier for each of several hundred images separately would require significant amounts of time and effort. Although we have applied our method only to flood mapping in one location, it is likely usable for other applications, as well. The method could be used to map transient events such as floods in other areas, provided a time series of images with sufficient temporal resolution to capture the event were available. It is also worth mentioning that, although our method was demonstrated using MODIS imagery, the conceptual framework of pooled training and classification may apply to other multispectral imageries and potentially SAR images with multiple radar frequencies and polarizations. Using the training pool method, the effort needed to accomplish this classification task would be reduced, making such analyses much more tractable.

Acknowledgments

This material was based on work supported by the U.S. National Science Foundation under grant number AGS 1518486, K. Alexander, PI. The article was improved by the comments of two anonymous referees.

References

1. D. Lu, "The potential and challenge of remote sensing-based biomass estimation," Int. J. Remote Sens. 27(7), 1297–1328 (2006).
2. M. Li et al., "A review of remote sensing image classification techniques: the role of spatio-contextual information," Eur. J. Remote Sens. 47(1), 389–411 (2014).
3. P. Du et al., "Multiple classifier system for remote sensing image classification: a review," Sensors 12(4), 4764–4792 (2012).
4. R. A. Schowengerdt, Techniques for Image Processing and Classifications in Remote Sensing, Academic Press (2012).
5. J. A. Dechka et al., "Classification of wetland habitat and vegetation communities using multi-temporal Ikonos imagery in southern Saskatchewan," Can. J. Remote Sens. 28(5), 679–685 (2002).
6. T. Blaschke, "Object based image analysis for remote sensing," ISPRS J. Photogramm. Remote Sens. 65(1), 2–16 (2010).
7. D. Liu and F. Xia, "Assessing object-based classification: advantages and limitations," Remote Sens. Lett. 1(4), 187–194 (2010).
8. G. S. Ruppert et al., "A hybrid classifier for remote sensing applications," Int. J. Neural Syst. 8(01), 63–68 (1997).
9. B. W. Heumann, "An object-based classification of mangroves using a hybrid decision tree—Support vector machine approach," Remote Sens. 3(11), 2440–2460 (2011).
10. J. Li and Y. Sheng, "An automated scheme for glacial lake dynamics mapping using Landsat imagery and digital elevation models: a case study in the Himalayas," Int. J. Remote Sens. 33(16), 5194–5213 (2012).
11. E. A. Lyons et al., "Quantifying sources of error in multitemporal multisensor lake mapping," Int. J. Remote Sens. 34(22), 7887–7905 (2013).
12. Y. Sheng et al., "Representative lake water extent mapping at continental scales using multi-temporal Landsat-8 imagery," Remote Sens. Environ. 185, 129–141 (2016).
13. C. Song et al., "Heterogeneous glacial lake changes and links of lake expansions to the rapid thinning of adjacent glacier termini in the Himalayas," Geomorphology 280, 30–38 (2017).
14. J. Wang, Y. Sheng, and T. S. D. Tong, "Monitoring decadal lake dynamics across the Yangtze Basin downstream of three Gorges Dam," Remote Sens. Environ. 152, 251–269 (2014).
15. K. Yang et al., "Recent dynamics of alpine lakes on the endorheic Changtang Plateau from multi-mission satellite data," J. Hydrol. 552, 633–645 (2017).
16. E. A. Cloutis, "Review article hyperspectral geological remote sensing: evaluation of analytical techniques," Int. J. Remote Sens. 17(12), 2215–2242 (1996).
17. W. J. Junk and K. M. Wantzen, "The flood pulse concept: new aspects, approaches and applications-an update," in Second Int. Symp. on the Management of Large Rivers for Fisheries, pp. 117–149 (2004).
18. K. Alexander and J. Blackburn, "Overcoming barriers in evaluating outbreaks of diarrheal disease in resource poor settings: assessment of recurrent outbreaks in Chobe District, Botswana," BMC Public Health 13(1), 775 (2013).
19. J. T. Fox and K. A. Alexander, "Spatiotemporal variation and the role of wildlife in seasonal water quality declines in the Chobe River, Botswana," PLoS One 10(10), e0139936 (2015).
20. V. Klemas, "Remote sensing of floods and flood-prone areas: an overview," J. Coast. Res. 31(4), 1005–1013 (2014).
21. L. C. Smith, "Satellite remote sensing of river inundation area, stage, and discharge: a review," Hydrol. Process. 11(10), 1427–1439 (1997).
22. M. Boschetti et al., "Comparative analysis of normalised difference spectral indices derived from MODIS for detecting surface water in flooded rice cropping systems," PLoS One 9(2), e88741 (2014).
23. A. Ogilvie et al., "Decadal monitoring of the Niger Inner Delta flood dynamics using MODIS optical data," J. Hydrol. 523(Suppl. C), 368–383 (2015).
24. J. J. Burke, N. G. Pricope, and J. Blum, "Thermal imagery-derived surface inundation modeling to assess flood risk in a flood-pulsed Savannah Watershed in Botswana and Namibia," Remote Sens. 8(8), 676 (2016).

25. D. E. Alsdorf, E. Rodríguez, and D. P. Lettenmaier, "Measuring surface water from space," Rev. Geophys. 45(2), 1–24 (2007).
26. G. Mallinis et al., "An object-based approach for flood area delineation in a transboundary area using ENVISAT ASAR and LANDSAT TM data," Int. J. Digit. Earth 6(suppl. 2), 124–136 (2013).
27. S. Martinis and C. Rieke, "Backscatter analysis using multi-temporal and multi-frequency SAR data in the context of flood mapping at River Saale, Germany," Remote Sens. 7(6) (2015).
28. S. Plank et al., "Mapping of flooded vegetation by means of polarimetric Sentinel-1 and ALOS-2/PALSAR-2 imagery," Int. J. Remote Sens. 38(13), 3831–3850 (2017).
29. I. Klein et al., "Global WaterPack—a 250 m resolution dataset revealing the daily dynamics of global inland water bodies," Remote Sens. Environ. 198, 345–362 (2017).
30. D. E. Alsdorf, "GEOPHYSICS: tracking fresh water from space," Science 301(5639), 1491–1494 (2003).
31. X. Ding and X. Li, "Monitoring of the water-area variations of Lake Dongting in China with ENVISAT ASAR images," Int. J. Appl. Earth Obs. Geoinformation 13(6), 894–901 (2011).
32. S. Martinis et al., "Comparing four operational SAR-based water and flood detection approaches," Int. J. Remote Sens. 36(13), 3519–3543 (2015).
33. N. G. Pricope et al., "Spatio-temporal analysis of vegetation dynamics in relation to shifting inundation and fire regimes: disentangling environmental variability from land management decisions in a Southern African Transboundary Watershed," Land 4(3), 627–655 (2015).
34. K. Nyamapfene, "A geographical overview of the soils of Zimbabwe and their agricultural potential," Geogr. Educ. Mag. 15, 63–73 (1992).
35. K. A. Alexander, A. Heaney, and J. Shaman, "Distant climate controls and dryland flood pulse dynamics influence diarrheal disease and population vulnerability to climate change," Nat. Cimate Change (2018), in press.
36. R. Hernandez Guzman et al., "Analysis of flood pulse dynamics in the lower basin of the San Pedro River (northwestern Mexico) using remote sensing," Lat. Am. J. Aquat. Res. 44(2), 293–304 (2016).
37. W. Junk, P. B. Bayley, and R. E. Sparks, "The flood pulse concept in river-floodplain systems," in Proc. of the Int. Large River Symp. (LARS), D. P. Dodge, Ed., Vol. 106, pp. 110–127, Canadian Special Publication of Fisheries and Aquatic Sciences (1989).
38. S. Long, T. E. Fatoyinbo, and F. Policelli, "Flood extent mapping for Namibia using change detection and thresholding with SAR," Environ. Res. Lett. 9(3), 035002 (2014).
39. https://lpdaac.usgs.gov/
40. Y. Chen et al., "An evaluation of MODIS daily and 8-day composite products for floodplain and wetland inundation mapping," Wetlands 33(5), 823–835 (2013).
41. A. S. Islam, S. K. Bala, and M. A. Haque, "Flood inundation map of Bangladesh using MODIS time-series images," J. Flood Risk Manag. 3(3), 210–222 (2010).
42. C. Huang, Y. Chen, and J. Wu, "Mapping spatio-temporal flood inundation dynamics at large river basin scale using time-series flow data and MODIS imagery," Int. J. Appl. Earth Obs. Geoinformation 26, 350–362 (2014).
43. N. G. Pricope and M. W. Binford, "A spatio-temporal analysis of fire recurrence and extent for semi-arid savanna ecosystems in southern Africa using moderate-resolution satellite imagery," J. Environ. Manage. 100(Suppl. C), 72–85 (2012).
44. S. E. Lobser and W. B. Cohen, "MODIS tasselled cap: land cover characteristics expressed through transformed MODIS data," Int. J. Remote Sens. 28(22), 5079–5101 (2007).
45. J. Leon Steven, Å. Björck, and G. Walter, "Gram–Schmidt orthogonalization: 100 years and more," Numer. Linear Algebra Appl. 20(3), 492–532 (2012).
46. L. Breiman, Classification and Regression Trees, Routledge, New York (2017).
47. T. Funkenberg et al., "The Ha Tien Plain—wetland monitoring using remote-sensing techniques," Int. J. Remote Sens. 35(8), 2893–2909 (2014).
48. T. Theneau and B. Ripley, "Rpart: recursive partitioning and regression trees. R package version 4.1-10," https://CRAN.R-project.org/package=rpart (11 June 2018).

49. P. Olofsson et al., "Good practices for estimating area and assessing accuracy of land change," Remote Sens. Environ. 148, 42–57 (2014).
50. G. Foody, "Thematic map comparison: evaluating the statistical differences in classification accuracy," Photogramm. Eng. Remote Sens. 70, 627–633 (2004).
51. P. A. Townsend and S. J. Walsh, "Modeling floodplain inundation using an integrated GIS with radar and optical remote sensing," Appl. Remote Sens. GIS Geomorphol. 21(3), 295–312 (1998).
52. A. G. Laborte, A. A. Maunahan, and R. J. Hijmans, "Spectral signature generalization and expansion can improve the accuracy of satellite image classification," PLoS One 5(5), e10516 (2010).
53. T. Kavzoglu and I. Colkesen, "A kernel functions analysis for support vector machines for land cover classification," Int. J. Appl. Earth Obs. Geoinformation 11, 352–359 (2009).

Mitchell P. Braget received his MA degree in geography from Kansas State University and is currently a GIS technician for the International Water Institute in Fargo, ND. His work is focused on water resources in the Red River of the North watershed.

Douglas G. Goodin is a professor of geography and director of the Remote Sensing Laboratory at Kansas State University. His research interests include biophysical remote sensing, thematic remote sensing, computational methods in spatial analysis, and application of remote sensing to environmental problems, particularly to analysis of infectious disease.

Jida Wang received his PhD in geography from University of California, Los Angeles (UCLA), and is currently an assistant professor in the Department of Geography, Kansas State University. His research focuses on terrestrial water resources (such as in lakes, reservoirs, and glaciers) and how their abundance and dynamics can be better understood using remote sensing and hydrological modeling.

James M. S. Hutchinson is a professor of geography and director of Kansas State University's Geographic Information Systems Spatial Analysis Laboratory. He studies issues related to grassland sustainability, surface water resources, agricultural biosecurity, and geographic visualization with an emphasis on real- and near-real time environmental monitoring. He has extensive international research and teaching experience, including work in China, Paraguay, Senegal, and France. He also currently serves as director of the university's secondary major in natural resources and environmental sciences.

Kathleen Alexander is a professor in the Department of Fish and Wildlife Biology at Virginia Tech University. Her research interests are exploring and understanding the factors that influence the emergence and persistence of emerging and re-emerging diseases at the human—wildlife—environmental interface, using a systems biology approach to ecosystem health integrated with public health. Her approach integrates critical crosscutting elements that can influence infectious disease dynamics such as culture and behavior, gender dimensions, and climate change.