OBJECT-BASED SEGMENTATION AND MACHINE LEARNING CLASSIFICATION FOR LANDSLIDE DETECTION FROM MULTI-TEMPORAL WORLDVIEW-2 IMAGERY

A thesis submitted to the faculty of
San Francisco State University
in partial fulfillment of
The Requirements for
The Degree

Master of Science
In
Geographic Information Science

by

Owen Patrick Parker

San Francisco, California

December 2013

CERTIFICATE OF APPROVAL

I certify that I have read *Object-based Segmentation and Machine Learning Classification for Landslide Detection from Multi-Temporal WorldView-2 Imagery* by Owen Patrick Parker, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirements for the degree: Master of Science in Geographic Information Science at San Francisco State University.

Leonhard Blesius
Associate Professor of Geography

Jerry Davis
Professor of Geography

OBJECT-BASED SEGMENTATION AND MACHINE LEARNING CLASSIFICATION FOR LANDSLIDE DETECTION FROM MULTI-TEMPORAL WORLDVIEW-2 IMAGERY

Owen Patrick Parker
San Francisco, California
2013

Landslides are pervasive hazards that pose significant risk to human populations. Routine quantification of landslide occurrence is necessary for hazard mitigation, traditionally compiled from manual interpretation of aerial imagery. To increase precision, reduce costs, and expedite analysis, much effort is focused on landslide identification from satellite imagery, with object-based methods rapidly emerging as a viable approach. Recent work has also utilized machine learning classifiers to increase automation and transferability. This study built on previous work to apply object-based image analysis (OBIA) and machine learning classification to sub-meter and multi-temporal WorldView-2 imagery. The primary objective was to explore scenarios resulting in optimal classification, considering: (1) random forest (RF) versus support vector machine (SVM) classifiers, (2) multispectral versus fused image resolutions, (3) binary versus multi-class structures, and (4) variations in sample size. A study area was selected involving challenging image composition and an extended capture window to test the robustness of the method to non-ideal conditions. The *eCognition* software allowed for image segmentation. Following selection of training samples, the *R* software was then utilized for machine learning classification with both RFs and SVMs. Classification was performed for each parameter combination over 100 replications, with accuracy assessed against a manual reference inventory. Optimal results were observed for RF at the largest sample size using a binary class structure and fused resolution, with an average F-score of 60.2 ± 1.3%. RF classifications consistently reached ~3-5% higher accuracy versus SVM when compared between specific parameter combinations. RFs demonstrated higher run-to-run stability, both in terms of spatial results and lower variance by area, as well as lower processing cost by an order of magnitude. These findings aid future work in determining optimal classification frameworks. The need for future research is also highlighted, including automation of sample selection and further refinement of the image segmentation task.

I certify that the Abstract is a correct representation of the content of this thesis.

_____ _____

Chair, Thesis Committee Date

ACKNOWLEDGEMENT

The author would like to thank the thesis committee, Drs. Leonhard Blesius and Jerry Davis, for numerous reviews and thoughtful insight. Accolade is also due to Seth Hiatt of the Institute for Geographic Information Science at San Francisco State University for illuminating discussions regarding statistical methods and both the *R* and *eCognition* softwares.

This research was supported in part by the SFSU College of Science and Engineering Advisory Board Scholarship and the Department of Geography Steven Pease Memorial Award. Satellite imagery was provided courtesy of the Department of Geography.

The author also thanks his parents, Steven and Karen Parker, and his partner, Jamie Hawk, for tireless encouragement and sacrifice, without which this work would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Landslides are natural hazards of concern to administrative bodies due to impacts on infrastructure and degradation of the environment (Schuster, 1996). Mass wasting is synergistically exacerbated by anthropogenic disturbance (Guthrie, 2002; Schuster & Kockleman, 1996) and can pose significant risk to human populations (Van Westen et al., 2006). Considered over time, landslides are the most persistent of all geologic hazards and often pose the greatest risk to human populations, particularly in developing nations (Nadim et al., 2006; Varnes, 1985). Much effort has focused on routine quantification of landslide occurrence in order to identify hazards and mitigate risk (Mantovani et al., 1996; Metternicht et al., 2005).

Comprehensive, multi-temporal inventory maps provide a indispensable foundation for landslide susceptibility analyses (Carrara et al., 1992; Varnes, 1985). Landslide re-occurrence is more likely than initial mobilization (Ardizzone et al., 2007), such that routine inventories allow determination of localized process rates for hazard mitigation (Van Westen et al., 2006). Knowledge of landslide occurrence is also useful for sediment transport estimation, in turn used to prevent surface erosion and preserve stream quality (Brardinoni et al., 2003; Guthrie, 2002). Although quantitative assessments depend on catalogued landslide inventories, records are often unavailable or are substantially incomplete (Van Westen et al., 2006; Wills & McCrink, 2002).

Traditionally, landslide maps are assembled by manual delineation onto stereoscopic aerial photographs (Carrara et al., 1992; Galli et al., 2008; Varnes, 1985). Such expert analysis often requires years for completion and is subject to uncertain costs, such that scale or precision

are compromised to meet deadline pressures (Ardizzone et al., 2007; Galli et al., 2008; Guzzetti et al., 2000). In addition, manual landslide inventories are inherently subjective (Albrecht, 2010), and vary according to intended purpose, methods employed, and perception of natural phenomenon by technicians (Van Westen et al., 1999; Wills and McCrink 2002).

Attempts have been made at landslide identification using satellite imagery (Mantovani et al., 1996; Metternicht et al., 2005) in the interest of increasing precision, reducing costs, and expediting analysis. Additionally, satellite acquisitions are standardized over extensive portions of the planet, allowing for broad-scale analyses. Validated methods incorporating satellite imagery would also be of benefit to rapid emergency response efforts (Mondini et al., 2011b; Weirich & Blesius, 2007). Nonetheless, state-of-the-art digitization workflows continue to rely on aerial photography and labor-intensive human interpretation (Ardizzone et al., 2007; Galli et al., 2008).

Use of satellite imagery for detection and monitoring of mass wasting has seen important contributions in the past decade. Pixel-value change in optical imagery has been utilized to successfully track slow-moving landslides on the basis of both spectral difference (Hervás et al., 2003) and sub-pixel feature tracking algorithms (Debella-Gilo & Kääb, 2011). Active-sensor LiDAR data has driven advances in monitoring large, slow-moving landslides (Glenn et al., 2006; McKean & Roering, 2004), including dormant failures obscured under canopy (Booth et al., 2009; Van Den Eeckhaut et al., 2007, 2012). However, these methods are generally not transferable to more common and devastating forms of landslides that occur at high velocity. Single events, such as seismic shock or intense rainfall, slope failures *en masse*

(Malamud et al., 2004; Stark & Hovius, 2001; Wieczorek, 1984), resulting in broad impacts that demand immediate and accurate analysis. Yet, to date, no automated methodology has been proven to consistently replicate the accuracy of human image interpretation.

The recent introduction of very high resolution (VHR) sensors has increased the capability to distinguish landslides from spaceborne platforms. Following the launch of Ikonos in 1999 were additional sensors at meter or sub-meter resolutions, including SPOT, Quickbird, Worldview-1 and -2, GeoEye, and Pleiades-1A and -1B. The resolutions available now meet the demands of operational mapping, such that VHR products are gaining acceptance for landslide digitization (Ardizzone et al., 2007; Mondini et al., 2011a; Weirich & Blesius, 2007).

Although new sensors provide new opportunities, VHR imagery has introduced unique challenges related to heightened variability within scene features (Blaschke, 2010). Satellite imagery analysis has traditionally relied on pixel-based methods (e.g. unsupervised clustering algorithms) for tasks such as land cover classification (Cihlar, 2000). However, for certain tasks, pixel-based analyses perform poorly (Blaschke, 2010; Cleve et al., 2008). This follows a long-standing commentary (Aplin, 2006; Cracknell, 1998; Fisher, 1997; Woodcock & Strahler, 1987) highlighting the insufficiency of arbitrary pixelation to adequately represent landscape features. Specifically, landslides are discrete objects of inconsistent shape and spectral response, inadequately differentiated from the spectral background to be accurately detected by pixel-based clustering algorithms (Barlow et al., 2006; Metternicht et al., 2005). Landslide detection at high resolution is sensitive to intra-class overlap of feature spectra, resulting in a salt-and-pepper effect in classifications from VHR imagery (e.g. Hervás et al., 2003). With previous-

generation sensors of lower resolution, sensitivity to noise is reduced by generalization of spectral radiance over a broad instantaneous field of view (IFOV; Cracknell, 1998). Although sensors such as Landsat TM (30 m) or ASTER (15 m) are appropriate for land cover classification, these resolutions are beyond the extent of small and narrow mass wasting occurrences such as shallow translational slides or earth flows (Brardinoni et al., 2003; Nichol & Wong, 2005).

Several pixel-based studies (e.g. Mondini et al., 2011a, 2011b; Nichol & Wong, 2005) have achieved useful results by employing change detection transforms, such as principle component analysis (PCA), and machine learning classification such as discriminant analysis (DA) and artificial neural networks (ANN). In addition, object-based methods have rapidly emerged as a viable alternative for the analysis of VHR imagery (Blaschke, 2010). Object-based image analysis (OBIA) overcomes the limitations inherent to pixel-based methods by grouping pixels into coherent image regions that present similarity of associated attributes (Benz et al., 2004), allowing the user to reclaim control of image generalization (Burnett & Blaschke, 2003). By nesting pixels within the context of their discrete representations, OBIA mimics the human logic process (Benz et al., 2004; Blaschke, 2010).

Research into landslide mapping with OBIA has delivered steady advancements in accuracy and methodological sophistication (Anders et al., 2011; Barlow et al., 2006; Chang et al., 2011; Lahousse et al., 2011; Lu et al., 2011; Martha et al., 2010, 2011, 2012; Moine et al., 2008; Stumpf & Kerle, 2011). Initial work (e.g. Barlow et al., 2006; Martha et al., 2010; Moine et al., 2008) established a foundation of understanding into indicator variables that are strongly correlated with landslide occurrence across diverse environments. Much effort has been

devoted to implementing user-driven workflows within the *eCognition* software. In addition to image segmentation, *eCognition* incorporates a wide variety of object-refinement routines and a semi-programmatic graphical user interface (GUI) environment that allows development of "rulesets" to carry analysis from segmentation through classification (see Anders et al., 2011; Barlow et al., 2006; Martha et al., 2010). A persistent shortcoming of such an approach is reliance on the user-driven process of trial-and-error parameter refinement (Blaschke, 2010). The selection of meaningful image features and ruleset parameters retains the subjectivity, time commitment, and dependence on expert knowledge inherent to previous, manual methods of interpretation. There is a need for a transferable and highly automated means of classification that approaches the accuracy obtained by a skilled operator.

Recently, pioneering efforts have successfully utilized machine learning classifiers to increase the automation and transferability of landslide detection (Chang et al., 2011; Stumpf & Kerle, 2011). Machine learning algorithms are widely used tools for variable selection and classification, such as in bioinformatics (Díaz-Uriarte & Alvarez de Andrés, 2006; Statnikov et al., 2008). Several approaches gaining popularity in the field of remote sensing include support vector machines (SVMs; Vapnik, 1999) and random forests (RFs; Breiman, 2001). SVMs and RFs are both non-parametric and non-linear (i.e. requiring no assumptions of normality or separability in the underlying data distribution), making them well suited to satellite imagery applications. Both algorithms require supervised training, but are robust to statistically under-represented sets when properly applied (Díaz-Uriarte & Alvarez de Andrés, 2006; Statnikov et al., 2008). Although artificial neural networks (ANNs) have a tradition of satellite imagery

applications (Mas & Flores, 2008; Paola & Schowengerdt, 1995), comparison studies have highlighted the mathematical transparency, comparable accuracy, greater speed, and relative ease of implementation of SVMs (Dixon & Candade, 2008; Foody & Mathur, 2004a) and RFs (Chan & Paelinckx, 2008; Gislason et al., 2006; Lawrence et al., 2006).

SVMs are a family of boundary classifiers with a history of remote sensing applications (Mountrakis et al., 2011). The learning algorithm attempts to find an optimal decision boundary, or hyperplane, to separate the dataset into a pre-defined number of classes (Vapnik, 1999). In situations of class overlap, as is common with imagery datasets, the assumption of linear separability fails. Thus, SVM applications introduce slack-variable kernels (e.g. radial basis functions) to allow higher order and non-linear solutions. Assuming appropriate selection of features, kernels and parameters, SVMs provide robust classification under class-imbalance and variable cross-correlation (Statnikov et al., 2008). Although SVMs do not include internal measures of variable importance, they may be applied independently to variable selection (Archibald & Fann, 2007; Díaz-Uriarte & Alvarez de Andrés, 2006), with comparable results to RFs (Pal, 2005, 2006).

RFs are an ensemble approach to maximizing the potential of tree-based classifiers (e.g. decision trees) by growing a 'forest' of trees from random subsets and allowing these to vote for the most likely class in a manner similar to bagging (Breiman, 2001). RFs have been successfully applied to remote sensing classification tasks (Duro et al., 2012; Gislason et al., 2006; Lawrence et al., 2006; Pal, 2005, 2006). In addition to aforementioned benefits, RFs are appealing for OBIA due to embedded variable importance measures, ease of implementation, and algorithm

efficiency (Breiman, 2001). RFs are relatively robust to predictor interactions, but are nontheless sensitive to over-fitting under class-imbalance (Blagus & Lusa, 2010) and over-estimation of variable importance under cross-correlation (Nicodemus et al., 2010; Strobl et al., 2007).

Class-imbalance and class-overlap are a common problem for ensemble- or kernel-based classifiers, given a tendency to over-fit to the majority class in the interest of minimizing total error (He & Garcia, 2009; Mountrakis et al., 2011). Landslides represent a minority class of the overall landscape (Malamud et al., 2004; Stumpf & Kerle, 2011), such that their classification is susceptible to imbalance. As such, a training sample based on a natural class balance within the scene will tend to favor the majority class and under-fit to landslides, resulting in excessive misses. A popular solution involves oversampling the minority class by adding either random replicates or synthetic samples to bolster feature space representation (Chawla et al., 2002; He & Garcia, 2009). These methods are useful when training data are limited, although image analysis benefits from the ability to train additional samples as needed to reach a prescribed class ratio. Stumpf & Kerle (2011) proposed a routine by which the degree of landslide over-sampling is iteratively adjusted until a balance is reached between errors of omission and commission.

OBIA and machine learning algorithms have previously been proven as viable methods for the detection of landslides from VHR imagery. However, uncertainty remains as to both the limits of applicability of the method and the ideal framework for image pre-processing, classifier determination, sample selection, and parameter settings for specific test applications. This study explored two broad objective questions. First, how reasonable is application of the OBIA,

machine learning classification task to the following: (1) broad scales of analysis, (2) an extended imagery time step and non-ideal environmental conditions, and (3) an operational and practical test case? Second, which scenarios result in optimal classifier performance, considering: (1) random forests versus support vector machines, (2) multispectral versus fused image resolution, (3) binary versus multi-class class structure, and (4) variation in sample size? A study area was selected involving challenging image composition and an extended image capture window in order to test the robustness of the method to non-ideal conditions. The results build on previous applications of the machine learning landslide detection task that have involved both random forests (Stumpf & Kerle, 2011) and support vector machines (Chang et al., 2011). The insights gained into performance capabilities and limitations will assist in both guiding real-world applications and determining appropriate directions for future research.

STUDY AREA

2.1 Site and Situation

The study area encompassed Silverado and Modjeska Canyons, at the eastern edge of Orange County, California (Fig. 1). Rugged terrain includes both publicly and privately owned lands situated between the Santa Ana Mountains of Cleveland National Forest and the historic

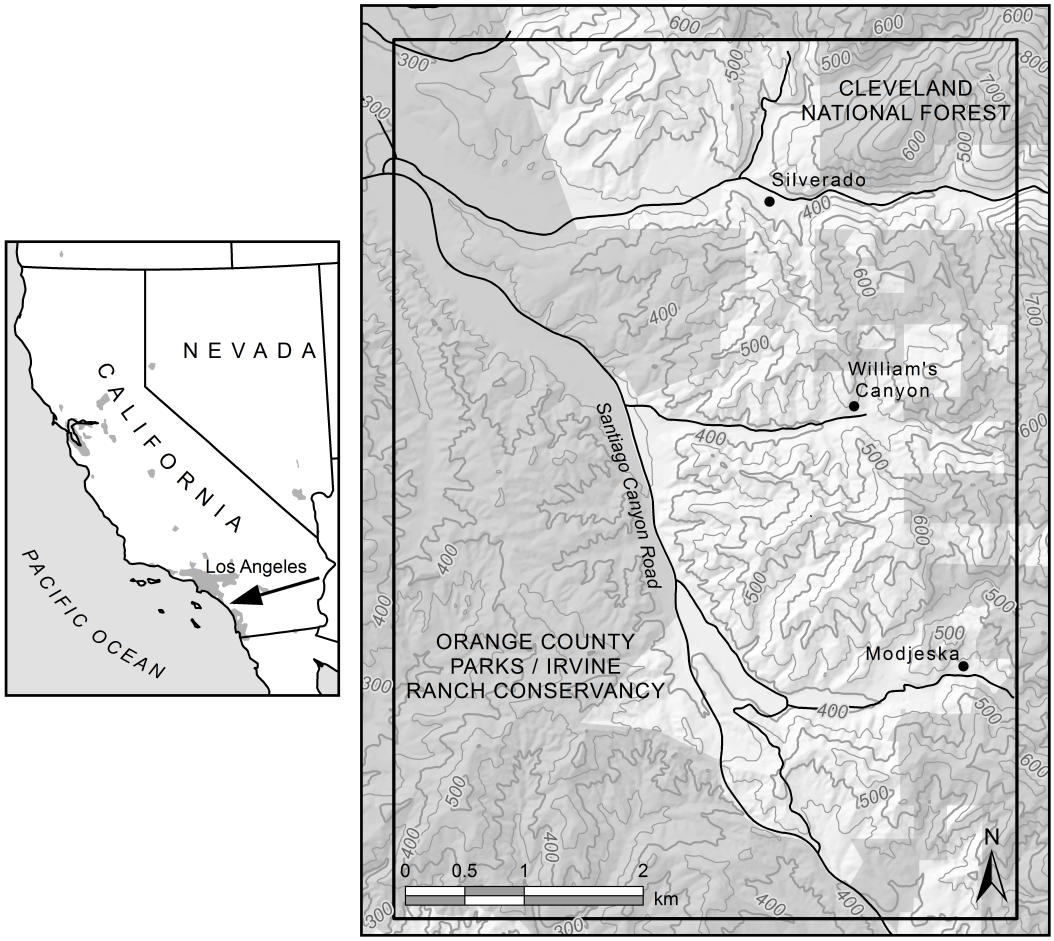

Fig. 1. Overview of study area, with location inset. *Inset:* Location within California; Urban regions highlighted in gray. *Overview map:* Imagery bounding box is outlined, with publicly administered lands in gray.

canyon rangelands of Irvine Ranch. An arid, Mediterranean climate results in dry summers and seasonal winter precipitation, averaging 88 mm yearly. Parent materials are predominantly Jurassic to Quaternary sedimentary, volcanic, and shallow intrusive formations, including complex layering of sandstone and shale variants (Fritsche & Behl, 2008). The region is susceptible to soil erosion and mass wasting due to a combination of unconsolidated, sedimentary geology, steep hillslopes, and inconsistent vegetation cover. Landscape formation is dominated by mass wasting and fluvial erosion, resulting in dissected canyons with 20° to 45° hillslopes and arroyos with considerable depth of accumulated alluvium. In addition to widespread landslide occurrence (Fig. 2), there are numerous cross-correlated features present in the image scene, including gullies, fire roads, streambeds, and urban interface. The study area remains predominantly rural despite urban density in the vicinity. The eastern half falls under jurisdiction of the Irvine Company and Orange County Parks, and is closed to public access, whereas the western half is mostly privately-owned in-holdings within the Cleveland National Forest administrative boundary.

From 21 October to 8 November 2007, the 115 km^2 Santiago Fire severely reduced already sparse vegetation cover within the southern two-thirds of the study area, prompting treatments to mitigate sediment yield (USFS, 2007). Later, from 18 December to 22 December 2010, a severe storm delivered 55 cm of rainfall to the area[1], resulting in flooding and emergency evacuation. The storm event was accompanied by widespread mass wasting along destabilized hillslopes (Fig. 2).

---

[1] http://www.intercanyon.org/winter-storm-2010

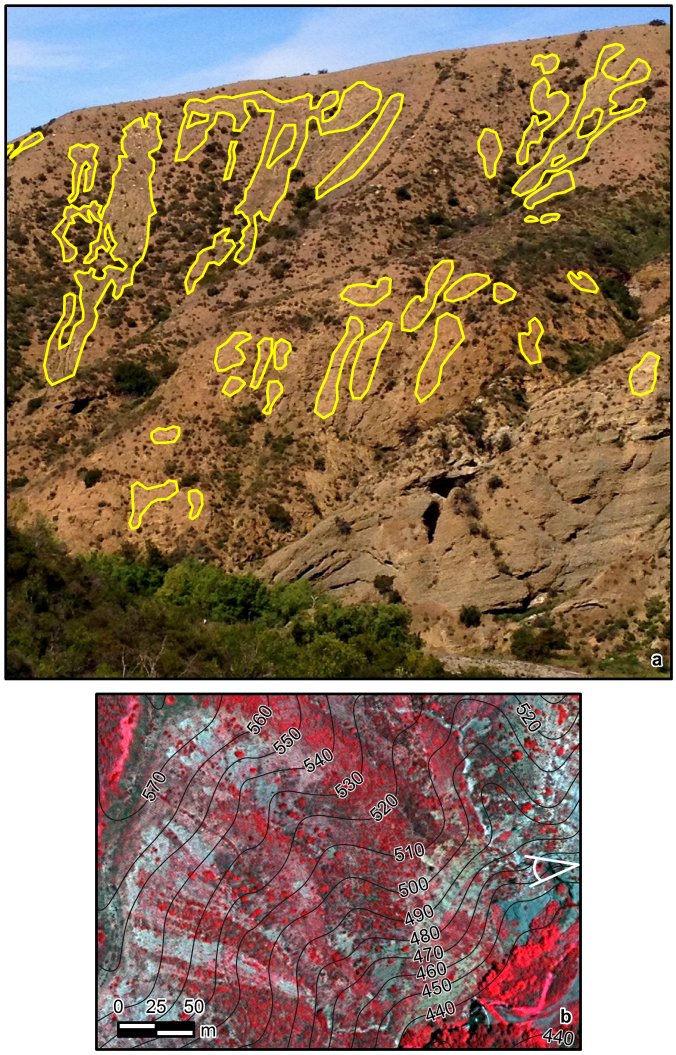

Fig. 2. Destabilized hillslopes (a) with delineated landslides (yellow outlines). False color imagery overview provided for reference (b), with contour lines in meters and angle of view indicated.

2.2 Imagery

Suitable WorldView-2 satellite imagery was acquired from swaths captured on 7 April 2010 and 28 April 2012 (Table 1). Potential imagery was available immediately after the winter storm (captured 24 December 2010), but did not reflect all landslides. Over the next month,

additional slope failures occurred as antecedent soil moisture accumulated and increased positive pore pressure. A single annum time step would have been ideal to minimize vegetation growth within failure tracks while maintaining consistent phenology, however no imagery was available during the same season in 2011. Owing to obscured signals of mass wasting and numerous cross-correlated features, the scene and situation posed a significant challenge to any classification method. Both scenes were received as 1-band panchromatic and 4-band multispectral, level 1-B GeoTIFF products, resampled by the vendor to the required subset using a modulation transfer function (MTF) kernel, with at-sensor geometric distortion corrections included.

Table 1

Overview of attributes for study area, imagery, and data.

Coordinate (UL)	33° 45' 34.9" lat.
	-117° 40' 21.0" lon.
Coordinate (LR)	33° 41' 28.7" lat.
	-117° 36' 52.7" lon.
Study Area (km^2)	40
Sensor	Worldview-II
Sensor Bands	Pan, B-G-R-NIR
Sensor Resolution (m)	0.5/2.0 (pan./multi.)
Imagery Date 1	07 April 2010
Off-Nadir Look Angle (°)	10.5
Sensor Zenith (°)	21.9
Sensor Azimuth (°)	238.0
Solar Zenith (°)	29.9
Solar Azimuth (°)	149.5
Imagery Date 2	28 April 2012
Off-Nadir Look Angle (°)	16.8
Sensor Zenith (°)	19.0
Sensor Azimuth (°)	213.0
Solar Zenith (°)	21.1
Solar Azimuth (°)	152.2
DEM Source (base data)	USGS NED (contour line)
DEM Resolution (m)	10

METHODS

3.1 Imagery Pre-processing

Imagery was first processed to approximate absolute surface reflectance values, with radiometric corrections applied to account for sensor gain and bias as well as atmospheric transmittance. The WorldView-2 imagery was delivered with detector variation adjustments included (Updike & Comp, 2010), and an at-sensor correction was calculated per pixel for each spectral band. At-satellite radiances were then converted to surface reflectance by the dark object subtraction (i.e. DOS3) method (Song et al., 2001), with additive effects removed by global subtraction of values sampled from water bodies and shadow (Chavez, 1996). Multiplicative effects were approximated by radiative transfer modeling, assuming simplified Lambertian surfaces and a cloudless Rayleigh atmosphere with zero aerosol optical depth. Estimates were made for optical thickness according to Kaufman (1989), and for atmospheric downwelling according to the 6S model (Vermote et al., 1997), generalized to the "US Standard 62" profile.

The multispectral bands were pan-sharpened, or 'fused,' in order to maximize the precision of discernible features while retaining a majority of spectral response. Potential negative impacts from fusion were considered, such as confusion of spectral signature and introduction of noise or artifacts at feature edges (see Karathanassi et al., 2007). However, analysis by Mondini et al. (2011b) found that the magnitude of noise and artifacts introduced by fusion are an order of magnitude less than the difference in pixel values between landslide and non-landslide classes. Several fusion methods were tested, including PCA (Chavez et al., 1991),

Gram-Schmidt (Laben, 2000), wavelet (Núñez et al., 1999), and high-pass filter (HPF; Gangkofner et al., 2008). As also observed in previous studies (Gupta & Dey, 2012; Karathanassi et al., 2007; Yuhendra et al., 2012), objective measures of accuracy between methods did not agree with observation, such that HPF fusion was empirically superior despite statistical shortcomings. Under visual assessment, the HPF algorithm provided the most consistent radiometric quality and stable results from which fine-scale features could be discerned.

Precise orthographic registration to ground geometry is critical for landslide detection (Barlow et al., 2006; Chang et al., 2011; Lahousse et al., 2011; Lu et al., 2011; Martha et al., 2011), and was achieved by a rational function model transform, built from vendor-supplied coefficients. A 10 m digital elevation model (DEM) was acquired from the USGS National Elevation Dataset, derived as an interpolation of archive stereoscopic contours (Gesch et al., 2002). As per Toutin (2004), calibration to ground geometry was applied with 7 ground control points acquired from GPS field survey (Fig. 3), with less than 0.1 m horizontal and vertical accuracy on all readings. Ortho-rectification was achieved using *ENVI* 5.0 software and validated against independent test points, arriving at 1.1 m and 1.6 m for 95% circular error probable (CE95) for the 2010 and 2012 images, respectively (Table 2). Cubic convolution was applied for all image resampling, despite the traditional preference for nearest neighbor on the basis of radiometric matching of reference spectra. In this case of sub-meter imagery over steep terrain, radiometric concerns were outweighed by the potential for spatial error introduced from pixel displacement by the nearest neighbor model (Toutin, 2004). Furthermore, an OBIA approach

Table 2

Imagery ortho-rectification statistics.

Image	2010	2012
Adjustment GCPs	7	7
Independent GCPs	9	10
Independent Plan. RMSE (m)	0.6	1.0
Independent Plan. CE95 (m)	1.1	1.6

generalizes the radiometric response (Blaschke, 2010), mitigating the need for absolute pixel-to-pixel comparison.

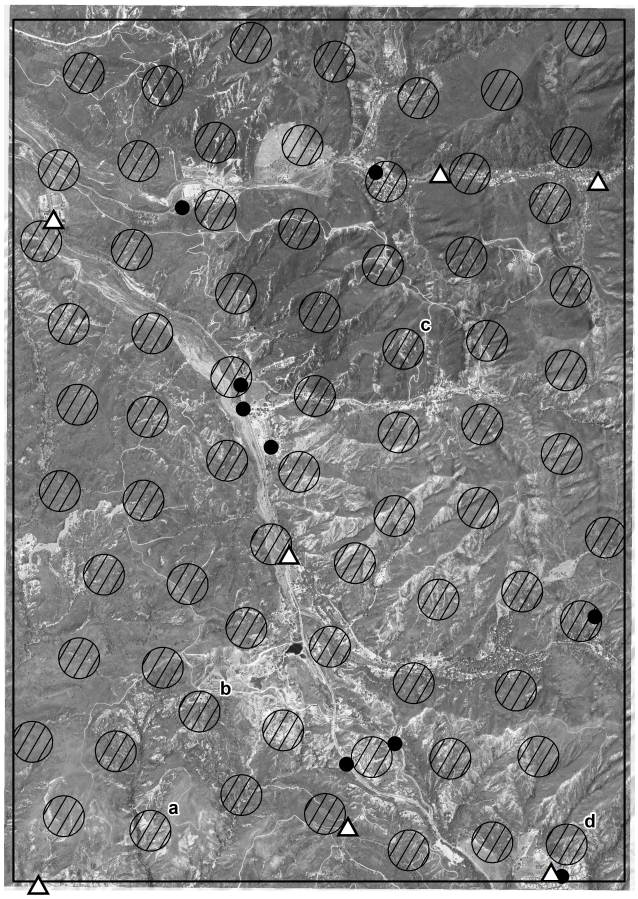

Fig. 3. Study area overview with 2012 imagery basemap. Distribution of adjustment (triangles) and validation (points) GCPs and classification test plots (hatched circles) are shown. The outline demarcates the final pre-processed imagery footprint. Letters (a-d) indicate test plot locations corresponding to Figures 4-6 & 9.

Due to multi-temporal inconsistencies in sun angle and phenology, the 2010 image was additionally adjusted to match the radiometric response of 2012 (Hall et al., 1991). Relative normalization was achieved by modeling a linear regression from a selection of pseudo-invariant features (PIF) between captures (Yang & Lo, 2000). PIF selection and normalization was fully automated by the multivariate alteration detection (MAD) method, a statistical ordination procedure based on canonical correlation (Canty et al., 2004; Nielsen et al., 1998; Schroeder et al., 2006). Because the MAD algorithm requires co-registration of image pixels, relative normalization was forestalled until after ortho-rectification.

Several image and topographic derivatives were calculated (Table 3). The first was normalized difference vegetation index ratio (NDVI), defined as $(NIR - R)/(NIR + R)$, which highlights greenness and indicates vegetation cover in regions of low to medium biomass (Huete et al., 2002). Measures of image texture were derived from gray level co-occurrence matrices (GLCM; Haralick et al., 1973). GLCMs offer improvement to VHR classification, particularly for anthropogenic or directional features (Puissant et al., 2005), and have been applied in previous object-based landslide analyses (Martha et al., 2010; Moine et al., 2008; Stumpf & Kerle, 2011; Wang & Niu 2010). However, GLCMs carry a heavy computational burden, thus only the filters *contrast* and *correlation* were computed (Stumpf & Kerle, 2011). Multi-temporal image differences and transforms also are a benefit to change detection (Lu et al., 2004) and landslide identification (Lu et al., 2011; Nichol & Wong, 2005). Principle component analysis is a linear, orthogonal transformation of the dataset that maximizes variance, and was applied to the full band stack from both image dates (Deng et al., 2008; Lu et al., 2011). Change in NDVI ($NDVI_{\Delta}$)

was also calculated, as was red difference (R_{Δ}). For the Southwestern United States, Chavez & MacKinnon (1994) emphasized the use of R_{Δ} rather than $NDVI_{\Delta}$. Under arid conditions lack of precipitation may result in differences in apparent greenness despite equivalent seasonality and vegetation cover. Alternately, high levels of red and near infrared (NIR) reflectance for sandy soils tend to obscure the signal of $NDVI_{\Delta}$.

Table 3

List of feature variables used as classifier input.

	Feature(s)
Spectral	B-G-R-NIR (2010, 2012)
	NDVI (2010, 2012)
	Band difference (G-R)
	NDVI difference
	PCA
	PCA (multispectral)*
Shape	Compactness
	Smoothness
	Length / Width
Context / Texture	GLCM (cor., con.)
	Main direction
	Direction minus net mean aspect
Ancillary, Topographic	Dlevation
	Slope (pct.)
	Curvature
	Aspect (N-S, E-W, net mean)
	Openness (pos., neg.; 30 m, 250 m)
Ancillary, Anthropogenic	Roads, euclidian distance
	Roads, kernel density (250 m)

Additional ancillary datasets were compiled, including topographic and road-network derivatives. The 10 m resolution of the original DEM was deemed too coarse to provide meaningful information for objects derived at or near 0.5 m. That is, considering uncertainty of one-half cell width for raster values abstracted over polygon boundaries, a 10 m resolution would result in erroneous statistics for image objects of less than roughly 20 m in any dimension. Thus, the DEM was generalized to match the panchromatic resolution by four

progressive applications of bilinear spline resampling to 5 m, 2.5 m, 1 m, and 0.5 m, followed by a low-pass filter. Although resampling introduced error, following subsequent segmentation this impact was mitigated by generalization over polygon area. From the result, slope was computed. Curvature calculation from the resampled DEM was particularly sensitive to artifacts, and instead was calculated from the original DEM and independently resampled to 0.5 m. Openness is a topographic derivative that measures enclosure from neighboring terrain (Yokoyama et al., 2002), providing a useful indication of landscape context for geomorphologic classification (Anders et al., 2011). Openness is scale-dependent, and thus was derived at 30 m and 250 m search radii both in positive and negative directions (i.e. above- and below-ground, respectively). Aspect is another terrain measure that provides context as a result of differing solar insolation and precipitation patterns on north- and south-facing slopes. Although aspect calculation is trivial per-pixel, when averaged over polygon boundaries uncertainty results from modulation between 0° and 360°. Therefore, mean aspect was computed as vector resultant decompositions of slope into north-south and east-west components (Davis, 2002). Transportation network data was also utilized (Barlow et al., 2006; Martha et al., 2010, 2011), as road cuts have a strong anthropogenic link to landslide risk resulting from hillslope destabilization and concentration of overland flow (Guthrie, 2002; Maharaj, 1993). In order to represent the abundance of emergency access roads and abandoned Jeep trails, roads were digitized from the imagery with additional reference to 1988 and 2004 aerial ortho-photos obtained from the USDA National Agriculture Imagery Program (NAIP). The *Kernel Density* tool in

ESRI *ArcGIS* 10.0 was then applied at a search radius of 250 m (Silverman, 1986), providing an aggregated measure of potential road impact.

3.2 Image Segmentation and Classification

3.2.1 Image Segmentation

The *eCognition* software was used to derive objects from pre-processed image pixels. The segmentation algorithm, titled multi-resolution image segmentation (MRIS; Benz et al., 2004), is a region-growing routine that merges objects upwards from the pixel level based on a user-specified balance of shape and spectral measures. Beginning with a pseudo-random set of pixels, the algorithm compares neighbors and merges if the result minimizes internal heterogeneity. Mean values are then computed for the newly created regions from selected input layers. The algorithm continues to merge pair-wise according to a linear weighting function (Benz et al*.,* 2004):

$$f = \omega_{spectral} h_{spectral} + \omega_{shape} h_{shape} \tag{1}$$

with $\omega_{shape} + \omega_{spectral} = 1$, where $\omega_{spectral}$ and ω_{shape} are user-defined weighting parameters . Respective heterogeneity criteria (h) measure change between the pre- and post-merge image objects. The spectral heterogeneity criterion measures change in deviation of pixel values, whereas the shape heterogeneity criterion measures change in object shape according to metrics referred to as “smoothness” and “compactness.” The scale parameter (f) governs the size to which objects may grow. Weighting parameters (ω) allow for balancing smoothness versus compactness and spectral versus spatial criteria.

Segmentation was conducted by trial-and-error parameter refinement of the MRIS algorithm, independently at the multispectral and fused resolutions. Default MRIS weights were used (i.e. $\omega_{compt} = 0.5, \omega_{shape} = 0.1$), partially due to the difficulty of objectively determining optimal values. Regardless, the default weight for shape is justifiable for landslide detection where strict spectral differentiation and loose shape control are desired (Martha et al., 2011). A typical implementation of input bands (Benz et al., 2004; Blaschke, 2010) was tested on the post-event image at both multispectral and fused resolutions, with all channels (blue, green, red, and NIR [*B-G-R-NIR*]) weighted equally (i.e. *1-1-1-1*). As landslide occurrence is correlated with image change, an additional scenario was tested, involving replacement of *NIR* with both G_{Δ} and R_{Δ}. The weighting scheme was also adjusted to favor the information provided by the change ratios, with B-G-R-G_{Δ}-R_{Δ} weights set as *1-1-1-3-3*. Thus, image change was intended to have twice the contribution as post-event spectral values. The results of the second scheme involving G_{Δ} and R_{Δ} were found to be superior and were used for final classification. At each respective resolution, the scale parameter was iteratively adjusted until a balance was reached between over- and under-segmentation of the landslide class (Figs. 4, 5, & 6), with final f-values of 10 for multispectral and 20 for fused resolutions. Mean zonal statistics were tabulated for all final objects from image layers, image derivatives, and ancillary data as listed in Table 3.

3.2.2 Reference Inventory & Training Set Preparation

The study area was separated into train and test regions, with test zones as buffered sample points at 175 m radii (e.g. Möller et al., 2007), derived according to a systematically stratified unaligned sampling scheme (SSUS; Stehman 1992). This distribution method was

chosen to provide an even representation of variations in features and terrain. 70 such zones were defined, representing 6.7 km^2 or 16.8% of the study area (Fig. 3). Within test regions, landslide occurrence was manually digitized wall-to-wall by a trained geomorphologist, and the inventory validated by ground survey conducted April 2013. Manual mapping was conducted with reference to the pan-sharpened imagery and derivatives, in addition to visualization of the imagery overlaid on the terrain model. The inventory was interpreted conservatively; there was no inference of *de facto* occurrences or boundary limits (Wills & McCrink, 2002), such as accumulation zones or extended debris tracks within river washes.

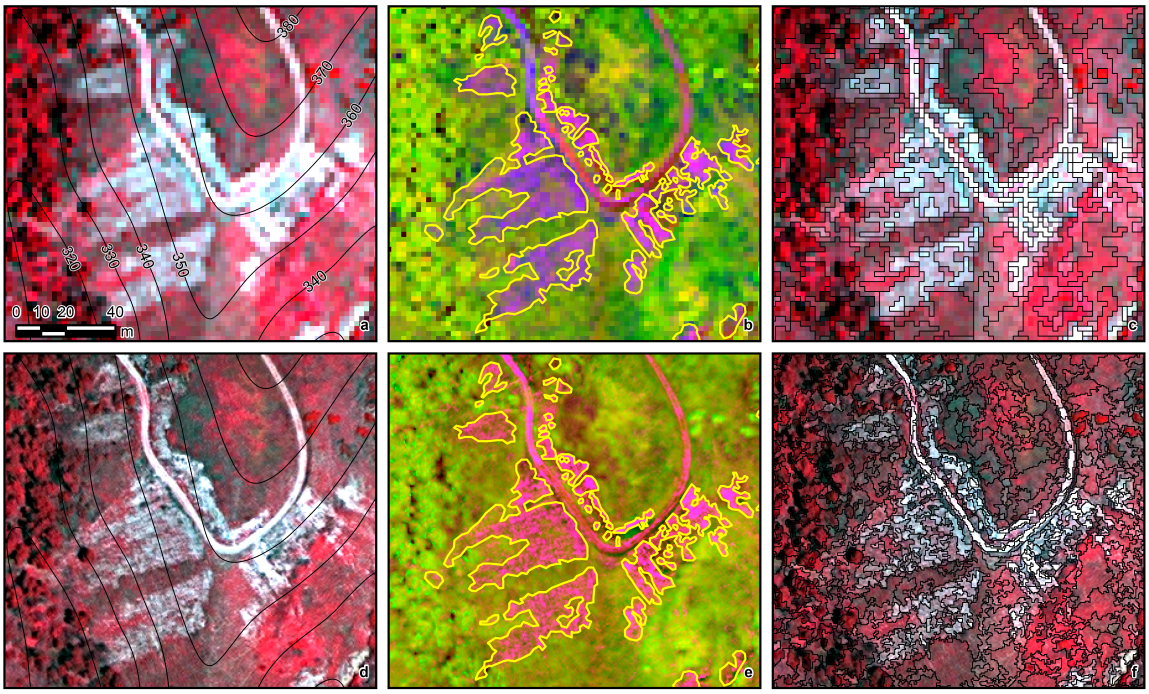

Fig. 4. *Sample location a*, including: a & d) False color images with contour line overlay and elevations in meters, b & e) Landslide reference inventory (yellow outline) over RGB composite image (R_{Δ}-$NDVI_{2012}$-PCA_4); c & f) Segmented image objects over false color images. Resolutions vary as multispectral (a-c) and fused (d-f).

Landslide classification is unique in that there is only one target class, thus either binary (i.e. single-class) or multi-class learning is possible. RFs are applicable to both approaches, whereas SVMs are designed for binary tasks but may be extended to multi-class scenarios (Hsu & Lin, 2002). An empirical review of, for example, Statnikov et al. (2008) indicates that the multi-class approach provides greater accuracy, and Stump & Kerle (2011) suggested that additional classes may allow the learner to better differentiate between cross-correlated predictive values. Thus, multiple non-landslide classes were discerned from the image scene, with particular attention to likelihood of confusion with landslides (Table 4). For several classes, there was sufficient variation of predictor variables within a set of image objects that it was necessary to further distinguish between changed and non-changed samples. For example, dirt roads were subdivided into examples that had either remained barren or had experienced vegetation re-colonization.

Table 4

Class definitions as sampled from segmented image objects.

Class	Sub-class
Landslides	
Barren (Bare Soil, Gully, Rock, etc.)	
Grasses	Phenology Consistent
Grasses	Phenology Changed
Shrubs & Trees	
Arroyos & River Plains	Stable
Arroyos & River Plains	New or Altered
Dirt Roads	Stable
Dirt Roads	Re-colonized by Vegetation
Paved Roads	
Buildings	
Shadow	

There is some discussion of pixel-based training for machine learning in the literature (e.g. Foody & Mathur, 2004b), yet little is known of robust methods for sampling image objects. Furthermore, although semi-automated sampling schemes involving active learning are in development (Stumpf et al., 2012), refinement and validation are still required. An approach was adopted involving manual selection of training samples from segmented image objects outside of the test zones. In an effort to minimize bias from spatial clustering and user preference, point locations were generated according to a pseudo-randomized SSUS grid within the training region, 200 each per target class. These points did not necessarily overlay on their intended objects. Within 175 m, each point was manually repositioned to the nearest polygon

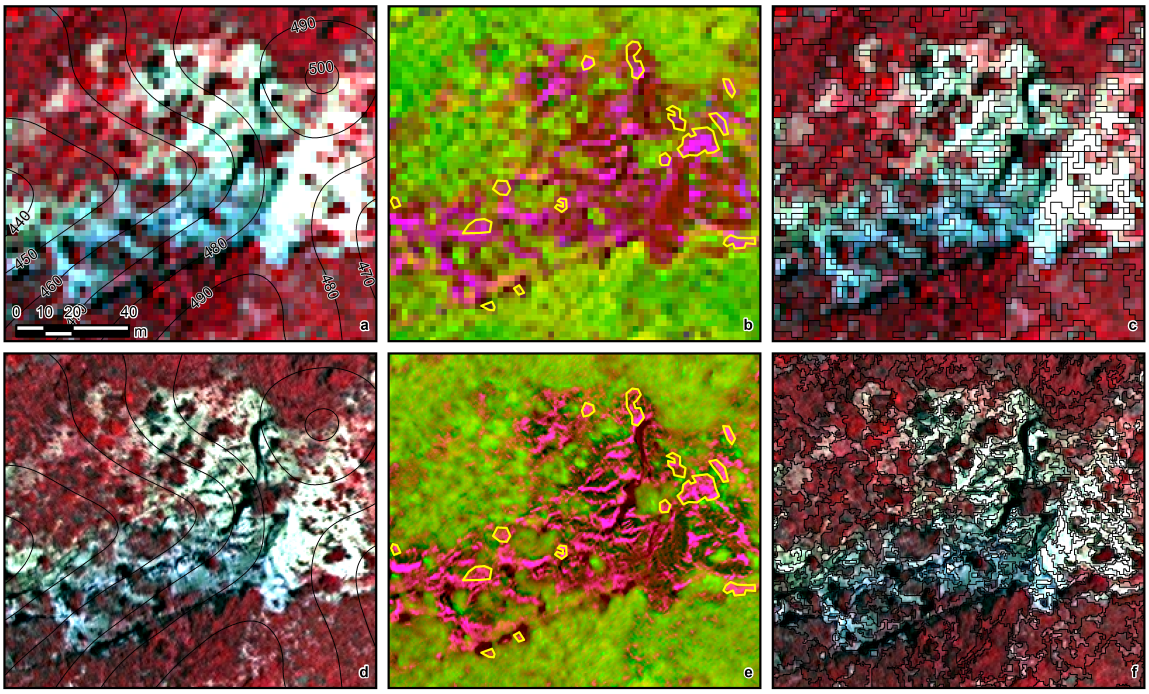

Fig. 5. *Sample location b*, including: a & d) False color images with contour line overlay and elevations in meters, b & e) Landslide reference inventory (yellow outline) over RGB composite image (R_{Δ}-$NDVI_{2012}$-PCA_4); c & f) Segmented image objects over false color images. Resolutions vary as multispectral (a-c) and fused (d-f).

corresponding wholly to the intended target at both the multispectral- and fused-resolution segmentations. If no appropriate location was found within the search window, the point was discarded. This approach yielded 40-120 samples per class. A final target of minimum 120 samples per class was fulfilled, at a compromise of reduced operator burden, by augmenting the initial sample pool with points chosen opportunistically from the image scene. For the landslide class only, 480 total samples were taken to allow for oversampling without exhausting representation from the sample pool.

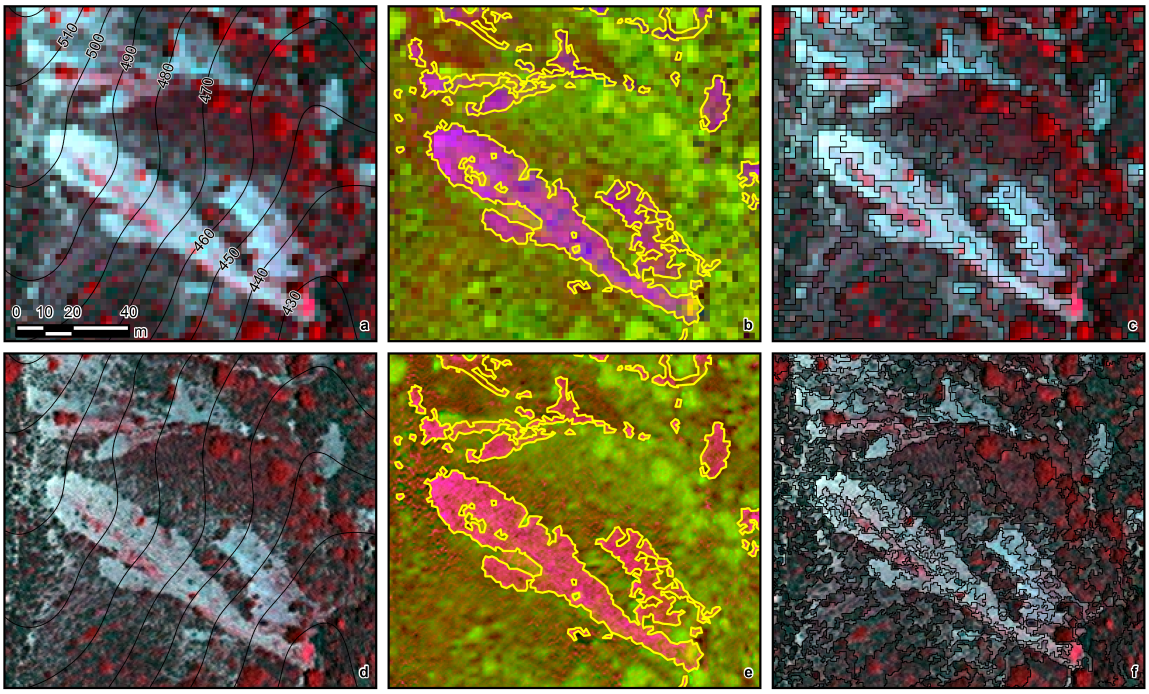

Fig. 6. *Sample location c*, including: a & d) False color images with contour line overlay and elevations in meters, b & e) Landslide reference inventory (yellow outline) over RGB composite image (R_{Δ}-$NDVI_{2012}$-PCA_4); c & f) Segmented image objects over false color images. Resolutions vary as multispectral (a-c) and fused (d-f).

3.2.3 Machine Learning Classification

Point samples were used to select and code training objects by class, and keyed tables were then exported to the *R* software (R Core Team, 2013). To test the sensitivity of classification to variation in sample size, each class sample was selected randomly from the full training set, by sets of 30 with replacement, to reach 30, 60, 90, and 120 observations total. For the binary case, all non-landslide classes were then combined. This approach was intended to explore the accuracy expected if a similar classification were initially trained at only a particular sample size.

Prior to classification, feature elimination was performed to reduce cross-correlation and improve the fit of the models, with top-ranked variables across all replications listed in Tables 5 & 6. SVMs require such a variable selection routine (Guyon et al., 2002; Statnikov et al., 2008), although the same is not consistently true for RFs. Although Díaz-Uriarte & Alvarez de Andrés (2006) recommend variable selection, RFs are robust to correlated variables with complex interactions such that an independent elimination step may limit the predictive power of the algorithm for imbalanced data (Blagus & Lusa, 2010; Cutler et al., 2007; Strobl et al., 2007). Nonetheless, feature elimination was performed for both models to allow for an equivalent and objective treatment (Statnikov et al., 2008). For the RF case, the *R* package *varSelRF* was applied (Díaz-Uriarte & Alvarez de Andrés, 2006), which builds random forests and iteratively removes 20% of the features based on importance measures calculated from initial out-of-bag (OOB) error. The ultimate set of selected features were those that produced the RF

Table 5

Multispectral resolution: Top-ranked features by classifier.

RF (binary)			RF (multi-class)			SV (binary)		
Name	Count	Importance	Name	Count	Importance	Name	Count	Rank
Red Diff.	100	0.078	Red (2012)	100	0.119	Blue (2012)	79	4.808
Green Diff.	100	0.050	Road Euc. Dist.	100	0.111	Slope	100	5.190
PCA 4	100	0.044	NDVI (2012)	100	0.100	PCA 3	78	6.577
Slope	100	0.037	NDVI (2010)	100	0.092	Red (2012)	71	7.076
NDVI Diff.	100	0.029	Blue (2012)	100	0.078	Red (2010)	38	7.463
Road Euc. Dist.	99	0.023	NDVI Diff.	100	0.078	NDVI Diff.	72	7.564
PCA 5	93	0.022	Red (2010)	94	0.055	Direct. - Aspect	78	7.882
Green (2010)	23	0.016	Green Diff.	97	0.049	Blue (2010)	44	8.377
PCA 3	78	0.013	Blue (2010)	100	0.049	PCA 6	25	8.456
Green (2012)	10	0.011	NIR (2012)	100	0.049	PCA 4	15	8.893
NDVI (2012)	24	0.010	PCA 4	100	0.048	PCA 5	14	9.286
PCA 2	35	0.010	Green (2012)	100	0.048	GLCM Corr. 0	13	9.462
PCA 6	43	0.010	PCA 1	96	0.048	Length / Width	25	9.624
NDVI (2010)	47	0.009	PCA 2	100	0.044	Open. (Neg 250)	12	9.983
Red (2012)	32	0.009	Red Diff.	100	0.042	Road Kern. Dens.	14	10.343
Red (2010)	5	0.008	NIR (2010)	96	0.042	Open. (Pos 250)	21	10.371
Blue (2010)	18	0.006	Open. (Neg 250)	100	0.041	GLCM Con. 0	22	10.618
Blue (2012)	11	0.005	Open. (Pos 250)	98	0.034	Compactness	9	11.022
NIR (2012)	7	0.005	PCA 5	99	0.033	Aspect (NS)	17	11.024
Smoothness	15	0.005	Slope	92	0.024	Aspect (EW)	12	11.333

Table 6

Fused resolution: Top-ranked features by classifier.

RF (binary)			RF (multi-class)			SV (binary)		
Name	Count	Importance	Name	Count	Importance	Name	Count	Rank
Red Diff.	100	0.054	Road Euc. Dist.	100	0.125	Red (2012)	84	5.426
Length / Width	74	0.052	Red (2012)	100	0.089	Slope	100	5.854
ELEV	15	0.044	Blue (2012)	100	0.085	Length / Width	96	6.604
PCA 3	100	0.039	NDVI (2010)	100	0.077	Blue (2012)	75	6.699
PCA 4 (Multi.)	100	0.034	NDVI (2012)	100	0.075	PCA 3	21	8.019
Slope	100	0.027	Length / Width	97	0.075	Red (2010)	46	8.165
Green Diff.	100	0.024	Green Diff.	99	0.066	PCA 1 (Multi.)	11	8.291
PCA 2 (Multi.)	20	0.022	PCA 4 (Multi.)	100	0.059	PCA 4 (Multi.)	26	8.308
Road Euc. Dist.	97	0.019	PCA 2 (Multi.)	100	0.049	PCA 3 (Multi.)	11	8.545
PCA 5	91	0.018	PCA 4	98	0.047	Direct. - Aspect	73	8.658
PCA 4	52	0.016	PCA 2	100	0.045	Blue (2010)	69	8.852
PCA 5 (Multi.)	69	0.014	Red Diff.	100	0.042	NDVI Diff.	26	9.008
NDVI Diff.	60	0.014	NIR (2012)	100	0.042	PCA 6	32	9.400
Red (2012)	27	0.012	Open. (Neg 250)	100	0.041	Green Diff.	15	9.600
Blue (2010)	23	0.010	Blue (2010)	100	0.041	PCA 5 (Multi.)	13	9.615
PCA 3 (Multi.)	27	0.009	NDVI Diff.	99	0.041	NDVI (2010)	11	9.618
NDVI (2012)	23	0.009	PCA 1	99	0.038	Green (2010)	20	10.070
NDVI (2010)	28	0.007	Green (2012)	100	0.037	GLCM Con. 135	10	11.240
PCA 6 (Multi.)	27	0.007	PCA 3	93	0.030	PCA 5	9	11.556
NIR (2012)	16	0.005	Open. (Pos 250)	90	0.029	Open. (Pos 250)	7	12.000

model with minimal OOB error. As per Svetnik et al. (2004), variable importance was not recomputed at each iteration. *varSelRF* makes direct use of the *randomForest* package (Liaw & Wiener, 2002), from which the only parameters requiring adjustment are *mtry*, which scales the number of variables tested at each node; *ntree*, the number of trees to grow; and *nodesize*, or the size of terminal nodes. For each factor, error decreases with increase in value, although there is generally a plateau beyond which significant performance costs are not justifiably compensated by gains in accuracy. Díaz-Uriarte & Alvarez de Andrés (2006) found error rates to stabilize at *mtry* above 1, *ntree* above 1000, and *nodesize* above 1. Here, for variable selection the initial forest was grown with *ntree* of 5000 and subsequent forests of 2000, and with *mtry* and *nodesize* of 2 and 1, respectively.

Variables for the SVM case were selected with the *mSVM-RFE* implementation[1] (based on Duan et al., 2005; Guyon et al., 2002), which models support vectors in the *e1071* package (Meyer et al., 2012) with the *LIBSVM* code (Chang & Lin, 2011). The most meaningful features were determined across 5-fold outer cross-validation (e.g. Foody & Mathur, 2004a), as follows. For each fold, top features were ranked according to minimized generalization error, and a grid search performed over a range of SVM cost and gamma parameters. Initially, value ranges were {0.01, 0.1, 0.5, 1, 5, 10, 50, 100} for cost and {0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1} for gamma, although after inspection of ranges utilized for equalizing class imbalance (see below) these were curtailed to 1 through 100 and 0.005 through 0.5, respectively. In all SVM models the radial basis kernel was applied. Optimal parameters from each fold were then used to

[1] http://www.colbyimaging.com/wiki/statistics/msvm-rfe

determine the ideal composition and number of features that minimized the generalization error across all folds. This required determining a local minimum of error values across the optimal grid search results for each number of features. Because multi-class variable selection is not supported for SVM, *mSVM-RFE* was applied to the binary training set and the resulting features used for the multi-class classification as well.

Following the selection of the optimal number and combination of features, class prediction was conducted for RFs using the *randomForest* package and for SVMs using the *e1071* package. In the RF case, the same static parameter values were used as those during feature elimination. For SVMs, only the selected variables were used to tune the model, again over a 5-fold outer cross-validated grid search of cost and gamma parameters. Trained models were then applied to prediction of object class membership for the independent test data.

A common challenge to machine learning is sensitivity to class-imbalance and class-overlap, leading to over- or under-prediction. Preliminary tests revealed that the dataset was indeed prone to under-prediction if a natural class balance was used. To correct this imbalance, a method similar to that demonstrated by Stumpf & Kerle (2011) was implemented. Screening runs were conducted to determine ideal values for β, with β representing the majority- to minority-class ratio. Beginning at a β of 1, i.e. severe over-sampling, 10 replicate runs were conducted and averaged, independently for each classifier and combination of resolution and sample size. Additional replicate sets were performed over β values incremented by 2 (i.e. 1, 3, 5, etc.). Accuracy was assessed independently for each run (see method below) and the results averaged for each β value. Second-order regressions were fit to the means of user's and

producer's accuracies (i.e. recall and precision). The intersection of the recall and precision trend lines was used to determine the optimal β value to balance accuracies for each respective sample size and image resolution (Figs. 7 & 8).

After determining β values, and according to the machine learning model as outlined, individual classifications were conducted. For each combination of classifier, sample size and image resolution (i.e. fused or multispectral), 100 replications were performed. Accuracy was assessed at each model run and averaged over all replications.

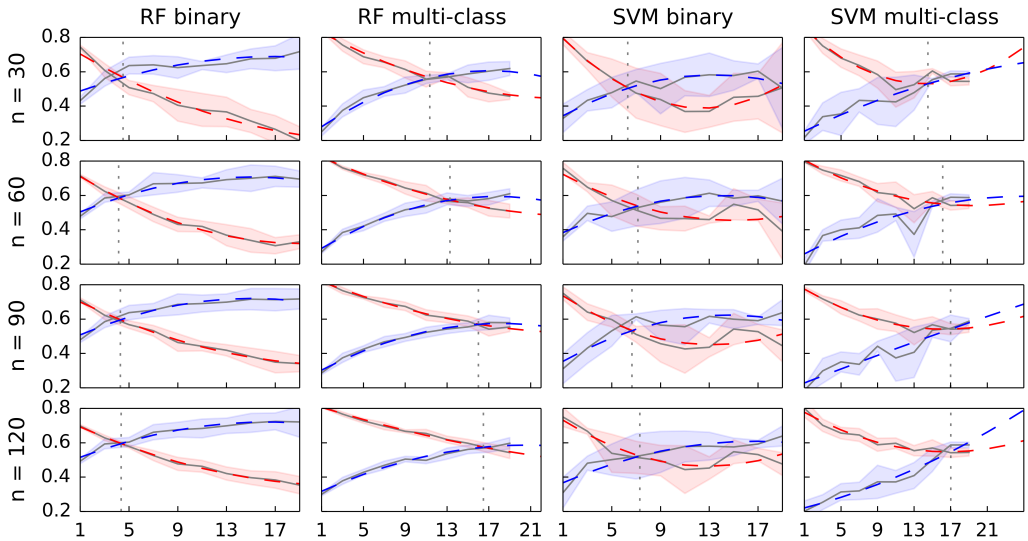

Fig. 7. *Fused resolution:* Determination of optimal class balance (β) to equalize user's accuracies (blue) and producer's accuracies (red). 2nd-order regressions are indicated as dashed trend lines. Each plotted data point (solid lines) represents the mean of 10 replications, with variances indicated as colored bounding plots.

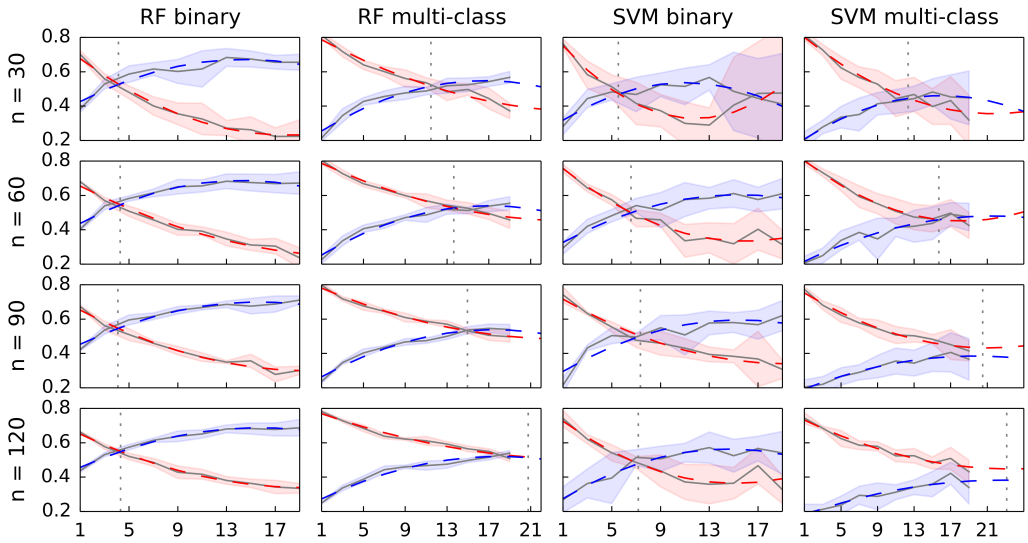

Fig. 8. *Multispectral resolution:* Determination of optimal class balance (β) to equalize user's accuracies (blue) and producer's accuracies (red). 2nd-order regressions are indicated as dashed trend lines. Each plotted data point (solid lines) represents the mean of 10 replications, with variances indicated as colored bounding plots.

RESULTS

Both the classified objects and the reference inventory landslides were mapped to the pixel level by converting to raster at 0.5 m resolution. Accuracy was then assessed pixel-by-pixel, allowing for an approximate comparison of classification results by area. Both user's and producer's accuracies were computed within the buffered test areas (Table 7), along with

Table 7

Classification accuracies, β -values, and processing time. Each row indicates the mean of 100 replications.

		Sample Size	β-value	AUC-ROC (%)	F-score (%)	User's accuracy (%)	Producer's accuracy (%)	Processing Time (s)
RF (binary)	Multi-spectral	30	4.1	74.7 ± 2.1	53.1 ± 1.7	50.2 ± 4.4	57.0 ± 3.7	17 ± 1
		60	4.3	75.5 ± 1.5	54.3 ± 1.2	51.7 ± 3.1	57.5 ± 3.1	36 ± 1
		90	4.1	76.4 ± 1.2	55.0 ± 0.9	53.6 ± 2.4	56.6 ± 2.2	58 ± 1
		120	4.3	76.4 ± 1.0	55.1 ± 0.8	53.7 ± 2.0	56.7 ± 2.0	80 ± 1
	Fused	30	4.5	77.1 ± 2.2	57.1 ± 2.2	54.9 ± 4.6	60.3 ± 5.2	20 ± 1
		60	4.2	78.3 ± 1.4	59.1 ± 1.8	57.4 ± 3.0	61.3 ± 4.5	41 ± 1
		90	4.3	78.6 ± 1.3	59.6 ± 1.5	58.0 ± 2.8	61.7 ± 4.1	63 ± 1
		120	4.3	79.2 ± 1.2	60.1 ± 1.3	59.1 ± 2.4	61.4 ± 3.9	88 ± 1
RF (multi-class)	Multi-spectral	30	11.5	75.6 ± 2.5	51.4 ± 2.1	52.3 ± 5.2	51.3 ± 4.7	30 ± 1
		60	13.6	75.9 ± 1.6	52.3 ± 1.1	52.8 ± 3.4	52.1 ± 3.2	58 ± 1
		90	15.0	76.4 ± 1.3	52.7 ± 1.0	53.8 ± 2.8	51.8 ± 2.6	88 ± 1
		120	20.8	74.0 ± 1.3	51.8 ± 1.4	48.9 ± 2.7	55.4 ± 2.7	116 ± 1
	Fused	30	11.4	77.8 ± 2.5	55.3 ± 2.5	56.5 ± 5.1	54.6 ± 4.2	40 ± 1
		60	13.3	78.2 ± 2.0	56.5 ± 2.0	57.3 ± 4.0	56.1 ± 3.9	71 ± 1
		90	16.0	77.8 ± 1.6	56.5 ± 1.7	56.5 ± 3.3	56.9 ± 3.4	103 ± 1
		120	16.5	78.9 ± 1.4	57.2 ± 1.5	58.7 ± 2.8	55.9 ± 2.9	136 ± 1
SVM (binary)	Multi-spectral	30	5.5	72.8 ± 4.4	45.5 ± 8.5	47.0 ± 8.8	46.7 ± 10.9	146 ± 8
		60	6.6	74.2 ± 3.3	49.5 ± 4.2	49.4 ± 6.8	50.9 ± 6.4	342 ± 19
		90	7.4	74.5 ± 2.2	49.6 ± 4.2	50.0 ± 4.0	49.9 ± 7.0	600 ± 32
		120	7.2	74.5 ± 2.0	48.4 ± 5.1	50.1 ± 4.0	47.8 ± 8.7	986 ± 49
	Fused	30	6.3	74.4 ± 5.0	49.9 ± 9.6	50.9 ± 9.9	51.6 ± 12.1	141 ± 7
		60	7.1	76.2 ± 3.2	54.2 ± 5.5	53.3 ± 6.6	56.5 ± 8.3	324 ± 23
		90	6.7	77.5 ± 2.8	54.6 ± 7.4	56.0 ± 5.4	54.3 ± 10.2	646 ± 42
		120	7.3	77.7 ± 2.0	54.9 ± 5.4	56.4 ± 4.0	54.2 ± 8.3	987 ± 53
SVM (multi-class)	Multi-spectral	30	12.3	70.1 ± 6.5	40.3 ± 11.8	41.7 ± 13.1	42.0 ± 13.6	112 ± 4
		60	15.7	71.3 ± 4.6	42.7 ± 9.0	43.8 ± 9.3	43.1 ± 11.0	224 ± 14
		90	20.5	69.3 ± 4.1	39.6 ± 8.6	40.0 ± 8.2	40.6 ± 10.8	353 ± 19
		120	23.1	69.0 ± 3.9	38.0 ± 9.1	39.6 ± 7.5	38.0 ± 12.0	518 ± 31
	Fused	30	14.5	69.6 ± 6.3	41.2 ± 12.2	41.5 ± 13.4	45.7 ± 15.8	113 ± 4
		60	16.1	73.0 ± 6.2	45.8 ± 12.4	47.5 ± 11.8	46.8 ± 15.0	230 ± 13
		90	17.0	73.8 ± 4.1	45.8 ± 10.1	49.2 ± 7.5	44.8 ± 13.0	391 ± 22
		120	17.0	74.6 ± 2.6	45.2 ± 7.4	50.9 ± 5.0	41.9 ± 10.0	620 ± 40

F-score and the area under the curve of the receiver operating characteristic (AUC-ROC). The F-score is a statistic that provides the harmonic mean of precision and recall (i.e. user's and producer's accuracies).

Best results were consistently achieved using the random forest classifier. The optimal parameter combination was ultimately observed at 120 samples using a binary class structure and fused resolution (e.g. Fig. 9), with an average F-score of 60.2 ± 1.3. Comparatively, the best SVM result was achieved with the same parameter set, at an F-score of 54.9 ± 5.4. As shown in Table 7, RF classifications consistently reached ~3-5% higher accuracy versus SVM when comparison was made between specific parameter combinations. RFs also demonstrated higher stability run-to-run, both in terms of consistency of spatial prediction and lower variance in accuracy. Finally, RF processing was faster than SVM by roughly an order of magnitude.

A comparative review of results between other parameter options also revealed systematic performance trends. For both classifiers and with all other settings held constant, increases from multispectral to fused resolution resulted in a ~2-4% increase in average

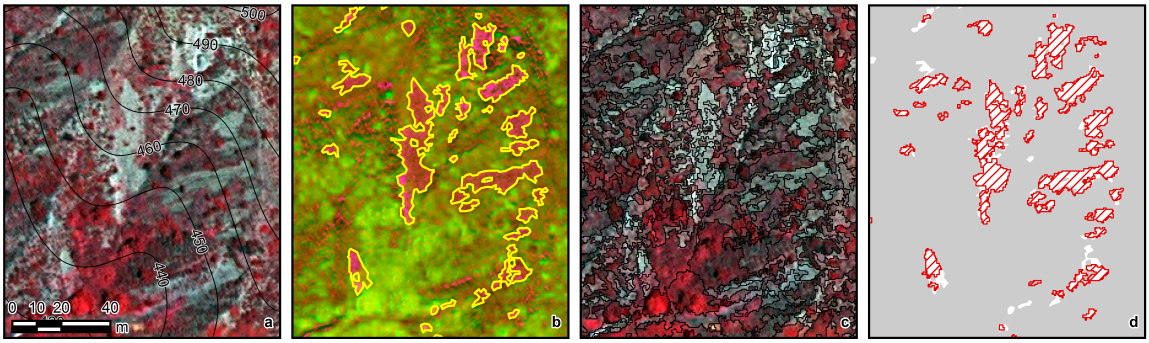

Fig. 9. *Sample location d*, including: a) False color image with contour line overlay and elevations in meters, b) Landslide reference inventory (yellow outline) over RGB composite image (R_{Δ}-$NDVI_{2012}$-PCA_4); c) Segmented image objects over false color images, d) . All resolutions are fused (i.e. 0.5 m).

accuracy, as well as an increase in variance. Classification run over a binary setup rather than against 12 target classes resulted in a ~2-3% increase in accuracy for RFs and a ~8-10% increase for SVMs. This observed performance offset was sufficient such that binary learning consistently exceeded the potential of the multi-class case despite increases in variance. The response of classifier performance to variation in sample size was less dependable. Although average accuracies fluctuated by ~2-4% across the range of 30 to 120 samples, the direction of change was inconsistent.

The feature elimination strategies used for RF and SVM resulted in a set of preferred features across runs (Tables 5 & 6). Although all learning strategies for RFs and SVMs tended to favor certain, similar features, results for the binary class structure were highly variable compared to the run-to-run stability of the multi-class scenario. Basic spectral values, ratios and transforms were chosen often by all routines, as were slope and measures of proximity to roads. Notably missing from the subset of meaningful features were both curvature and metrics derived from GLCMs.

DISCUSSION

5.1 Image Segmentation & Accuracy Assessment

Considering the preliminary step of the eCognition MRIS segmentation algorithm, the time delay from event to post-image capture posed a challenge to achieving optimal object delineation. Signal response within failure tracks was obscured by erosion, weathering of exposed materials, and vegetation regrowth. As a result, direct application of default MRIS parameters to the multispectral and fused products failed to consistently delineate landslide boundaries, even when strongly over-segmented. After multi-temporal change in both green and red bands was added to the algorithm, the extents of segmented objects, although over-segmented, were better constrained to landslide boundaries. Such an application of band-difference ratios may be well suited only to arid or semi-arid environments (Chavez & MacKinnon, 1994). It should be noted that the final scale parameter values as used cannot be reliably reapplied to diverse test environments.

Previous object-based landslide studies have not presented a standardized method of accuracy assessment. This stems from the fact that there is rarely absolute correspondence between segmentation boundaries and landslide occurrence as mapped in the reference inventory, and therefore direct comparison is not possible. The method used by Stumpf & Kerle (2011) involved generalization of the reference inventory to the object level based on a proportional cut-off value. However, the generalization error introduced was not appropriate for direct comparison of differing object boundary results at the multispectral and fused resolutions. Instead, the method of rasterizing the inventories and comparing at the pixel level

provided a precise gauge of performance, albeit at a greater processing cost. A point that deserves future study is development of an automated procedure for vectorized delineation of individual landslides from the resulting clusters of segmented image objects.

5.2 Class Balance & Sample Size

The preliminary model runs to equalize class-imbalance over β values provided useful insight into the relative performance of each classifier and combination of parameters (Figs. 7 & 8), particularly in regard to class structure. The highest degree of stability was observed for RFs in general, and specifically for multi-class runs. However, consistently higher β values were required in the multi-class case to equalize the user's and producer's accuracies. In effect, use of a multi-class approach reduced the importance of the landslide class in the classification model, and required additional constraint to avoid over-fitting. As a consequence, in the binary case the sample pool was consistently larger, providing the classifier with increased information content. Therefore, it is inconclusive if the higher accuracy observed for binary resulted from an inherent advantage over multi-class learning or from a sensitivity to sample size. For future research, more robust sampling should be targeted for multi-class prediction to prevent under-representation.

The manual sampling test sizes of between 30 to 120 samples per class represent a standard range for traditional applications of unsupervised, pixel-based classification. In the case of machine learning, however, commonly the true membership of the entire dataset is known *a priori*, and a full 20% is sampled as a training set (Blagus & Lusa, 2010; Mondini et al., 2011b; Statnikov et al., 2008; Stumpf & Kerle 2011). By limiting the sample size, this study induced

effects similar to those of so-called "small *n* large *p*" learning situations (Strobl et al., 2007), in which the number of representative samples (*n*) is insufficient in comparison to the number and complexity of interaction of predictor variables (*p*). A further limitation on learning potential resulted from unique replication sets being simulated by random selection from the original sample set, which may explain curtailed performances for the 90- and 120-sample cases (Table 7).

Although SVM results exhibited lower accuracy and higher variability, this may reflect a higher sensitivity to limited training knowledge. It remains to be seen if a similar study using extensive sample sets would also favor RFs. The need persists for an efficient and preferably automated approach for selection of sufficiently large sample sets before machine learning algorithms can be operationally applied to the landslide identification task over broad scales of imagery. For example, continued research into active learning approaches may provide a solution (e.g. Stumpf et al., 2012).

5.3 Feature Selection

From the results of feature elimination given in Tables 5 & 6, it is uncertain if the high run-to-run variability observed in the binary cases (i.e. for both RF and SVM) resulted from selection procedures or limited training knowledge. Comparatively, for multi-class models of the RF classifier, the variables selected were consistent across all replications. In theory, the SVM algorithm of recursive feature elimination is the more robust approach, in which the final predictors have been objectively cross-validated to achieve a minimum of error. In contrast, the RF strategy is naive in that features are only known to be meaningful for prediction of an

individual hold-out sample. The SVM routine demonstrated higher variability between replications, but it is not empirically apparent whether the approach arrived at a less correlated subset on average. In all cases, top-ranked features agreed with those meaningful to human interpretation, such as spectral bands, simple band ratios and transforms, and slope. Other than slope, the low importance accorded to topographic descriptors was expected, considering the coarse resolution of the elevation model relative to the imagery, and therefore the diminished information content of derivative data. Future research would benefit from highly accurate and precise topographic data such as from LiDAR scanning.

5.4 Classifier Comparison & General Trends

Although there was variability in individual classification results, systematic trends were observed across the run-specific spatial outputs (Fig. 10). Areas predicted by SVMs tended to fluctuate inconsistently between gross over- or under-estimation, whereas RF results were relatively stable over repetition. For both classifiers and at all parameter settings, a systematic source of error was poor prediction of landslide occurrence in cases of shallow translational complexes. Typically these locations involved pre-existing disturbance that served to obscure the signal of fresh scars. Another common false alarm was for regions of bare soil, particularly in the case of gully complexes that included areas in shadow. In these locales classification was challenging even for a trained interpreter, and only after fieldwork was the reference inventory validated. In addition, over steep topography such as gullies the learning algorithms were sensitive to false examples of image change that stemmed from image parallax during ortho-rectification. Other, less pervasive blunders included misappropriation of streambeds, buildings,

and managed fields, all of which were sporadic and resulted from particular objects that had changed between image dates (e.g. cut bank erosion, new construction, fire suppression management, etc.). Notably, roads were effectively differentiated in all learning scenarios. This was likely a result of effective learning from the road network input, and suggests the potential for other accurate semantic datasets to positively improve the landslide detection task (e.g. building footprint records, etc.) .

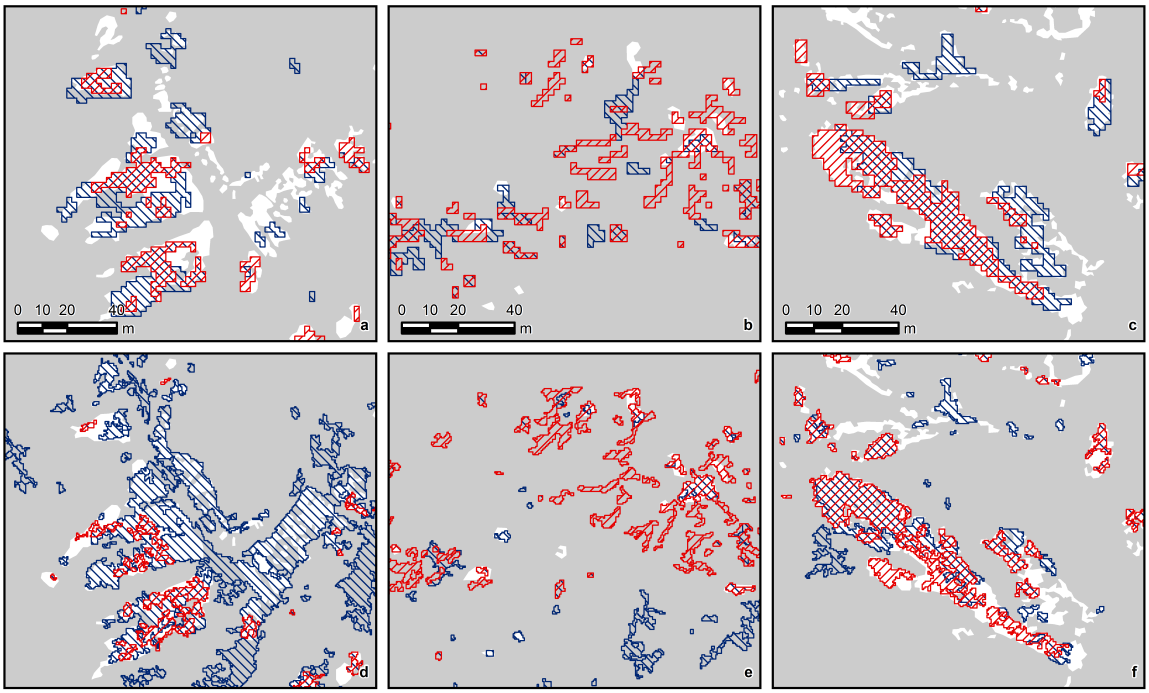

Fig. 10. Representative classification results for locations *a-c* (i.e. Figs. 4-6), showing best-case classifications at 120 samples for binary RF (red, 45° hatching) and binary SVM (blue, -45° hatching). The reference inventory is highlighted in white. Resolutions vary as multi-spectral (a-c) and fused (d-f).

CONCLUSION

In the past decade, methods for delineation of landslides from VHR satellite imagery have seen rapid advancement. Although object-based routines have been proven to provide viable and consistent results, such methods continue to rely on a user-driven approach to analysis that limits tractability across broad-scale application environments of diverse composition. Stumpf & Kerle (2011) demonstrated a method to increase the automation of landslide detection by implementing the random forest machine learning classifier. However, uncertainty remains as to an optimal framework for image pre-processing and for selection of both classifier and requisite parameter settings.

This study extended OBIA and machine learning landslide detection methods by comparing the use of classifiers, new image resolutions, class structures, and sample sizes. The random forest classifier was objectively compared to support vector machines and found to outperform under all scenarios. These findings free future analysts from uncertainty regarding optimal framework, while highlighting the need for future research, including automation of sample selection, as well as further refinement of the image segmentation task. The need also persists for optimization of image transform algorithms to enhance subtle landslide signals in challenging test environments.

REFERENCES

Albrecht, F., Lang, S. & Hölbling, D. (2010). Spatial accuracy assessment of object boundaries for object-based image analysis. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-4/C7, 29 June - 2 July. Ghent, Belgium*.

Anders, N. S., Seijmonsbergen, A. C., & Bouten, W. (2011). Segmentation optimization and stratified object-based analysis for semi-automated geomorphological mapping. *Remote Sensing of Environment, 115,* 2976–2985.

Aplin, P. (2006). On scales and dynamics in observing the environment. *International Journal of Remote Sensing, 27,* 2123–2140.

Archibald, R., & Fann, G. (2007). Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geoscience and Remote Sensing Letters, 4,* 674–677.

Ardizzone, F., Cardinali, M., Galli, M., Guzzetti, F., & Reichenbach, P. (2007). Identification and mapping of recent rainfall-induced landslides using elevation data collected by airborne Lidar. *Natural Hazards and Earth System Science, 7,* 637–650.

Barlow, J., Franklin, S., & Martin, Y. (2006). High spatial resolution satellite imagery, DEM derivatives, and image segmentation for the detection of mass wasting processes. *Photogrammetric Engineering & Remote Sensing, 72,* 687–692.

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing, 58,* 239–258.

Blagus, R., & Lusa, L. (2010). Class prediction for high-dimensional class-imbalanced data. *BMC Bioinformatics, 11,* 523.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing, 65,* 2–16 .

Blesius, L., & Weirich, F. (2010). Shallow landslide susceptibility mapping using stereo air photos and thematic maps. *Cartography and Geographic Information Science, 37,* 105–118.

Booth, A. M., Roering, J. J., & Perron, J. T. (2009). Automated landslide mapping using spectral analysis and high-resolution topographic data: Puget Sound lowlands, Washington, and Portland Hills, Oregon. *Geomorphology, 109,* 132–147.

Brardinoni, F., Slaymaker, O., & Hassan, M. (2003). A landslide inventory in a rugged forested watershed: A comparison between air-photo and field survey data. *Geomorphology, 54,* 179–196.

Breiman, L. (2001). Random forests. *Machine Learning, 45,* 5–32.

Burnett, C., & Blaschke, T. (2003). A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling, 168,* 233–249.

Canty, M. J., Nielsen, A., & Schmidt, M. (2004). Automatic radiometric normalization of multitemporal satellite imagery. *Remote Sensing of Environment, 91,* 441–451.

Carrara, A., Cardinali, M., & Guzzetti, F. (1992). Uncertainty in assessing landslide hazard and risk. *The ITC Journal, 2,* 172–183.

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

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2,* 1-27.

Chang, K.-T., Hwang, J.-T., Liu, J.-K., Wang, E.-H., & Wang, C.-I. (2011). Apply two hybrid methods on the rainfall-induced landslides interpretation. *19th International Conference on Geoinformatics, 24-26 June. Shanghai, China.*

Chavez, P. S. (1996). Image-based atmospheric corrections – revisited and improved. *Photogrammetric Engineering & Remote Sensing, 62,* 1025–1036.

Chavez, P. S., & MacKinnon, D. J. (1994). Automatic detection of vegetation changes in the Southwestern United States using remotely sensed images. *Photogrammetric Engineering & Remote Sensing, 60,* 571–583.

Chavez, P. S., Sides, S. C., & Anderson, J. A. (1991). Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogrammetric Engineering & Remote Sensing, 57,* 295–303.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16,* 321-357.

Cihlar, J. (2000). Land cover mapping of large areas from satellites: Status and research priorities. *International Journal of Remote Sensing, 21,* 1093–1114.

Cleve, C., Kelly, M., Kearns, F. R., & Moritz, M. (2008). Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography. *Computers, Environment and Urban Systems, 32,* 317–326.

Congalton, R. G., & Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, Florida: Lewis Publishers. 137 p.

Cracknell, A. P. (1998). Review article: Synergy in remote sensing--what's in a pixel? *International Journal of Remote Sensing, 19,* 2025–2047.

Crammer, K., & Singer, Y. (2001). On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research, 2,* 265–292.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., & Hess, K. T. (2007). Random forests for classification in ecology. *Ecology, 88,* 2783–2792.

Davis, J. C. (2002). *Statistics and Data Analysis in Geology*. New York: Wiley. 638 p.

Debella-Gilo, M., & Kääb, A. (2011). Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation. *Remote Sensing of Environment, 115,* 130–142.

Deng, J. S., Wang, K., Deng, Y. H., & Qi, G. J. (2008). PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing, 29,* 4823–4838.

Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics, 7*.

Dixon, B., & Candade, N. (2008). Multispectral landuse classification using neural networks and support vector machines: One or the other, or both? *International Journal of Remote Sensing, 29,* 1185–1206.

Drăguţ, L. (2010). ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science, 6,* 859–871.

Duro, D. C., Franklin, S. E., & Dube, M. G. (2012). Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. *International Journal of Remote Sensing, 33,* 4502–4526.

Esch, T., Thiel, M., Bock, M., Roth, A., & Dech, S. (2008). Improvement of image segmentation accuracy based on multiscale optimization procedure. *IEEE Geoscience and Remote Sensing Letters, 5,* 463–467.

Espindola, G. M., Camara, G., Reis, I. a., Bins, L. S., & Monteiro, A. M. (2006). Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing, 27,* 3035–3040.

Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters, 27,* 861–874 (2006).

Fisher, P. (1997). The pixel: A snare and a delusion. *International Journal of Remote Sensing, 18,* 679–685.

Foody, G. M., & Mathur, A. (2004a). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing, 42,* 1335–1343.

Foody, G. M., & Mathur, A. (2004b). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment, 93,* 107–118.

Fritsche, A. E., & Behl, R. J., Eds. (2008). *Geology of Orange County, California, and the Irvine Ranch National Natural Landmark*. Long Beach, CA: SEPM Society for Sedimentary Geology.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., & Reichenbach, P. (2008). Comparing landslide inventory maps. *Geomorphology, 94,* 268–289.

Gangkofner, U. G., Pradhan, P. S., & Holcomb, D. W. (2008). Optimizing the high-pass filter addition technique for image fusion. *Photogrammetric Engineering & Remote Sensing, 74,* 1107–1118.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., & Tyler, D. (2002). The national elevation dataset. *Photogrammetric Engineering and Remote Sensing, 68,* 5-11.

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

Glenn, N. F., Streutker, D. R., Chadwick, D. J., Thackray, G. D., & Dorsch, S. J. (2006). Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology, 73,* 131–148.

Gupta, A. & Dey, V. (2012). A comparative investigation of image fusion in the context of classification. *Journal of the Indian Society of Remote Sensing, 40,* 167–178.

Guthrie, R. H. (2002). The effects of logging on frequency and distribution of landslides in three watersheds on Vancouver Island, British Columbia. *Geomorphology, 43,* 273–292.

Guyon, I., Weston, J., & Barnhill, S. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46,* 389–422.

Guzzetti, F., Cardinali, M., Reichenbach, P., & Carrara, A. (2000). Comparing landslide maps: A case study in the Upper Tiber River Basin, Central Italy. *Environmental Management, 25,* 247–263.

Hall, F. G., Strebel, D. E., Nickeson, J. E., & Goetz, S. J. (1991). Radiometric rectification: Toward a common radiometric response among multidate, multisensor images. *Remote Sensing of Environment, 35,* 11–27.

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-3,* 610–621.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21,* 1263–1284.

Hervás, J., Barredo, J. I., Rosin, P. L., Pasuto, A., Mantovani, F., & Silvano, S. (2003). Monitoring landslides from optical remotely sensed imagery: The case history of Tessina landslide, Italy. *Geomorphology, 54,* 63–75.

Hsu, C.-W., & Lin, C.-J. (2002). A comparison on methods for multiclass support vector machines. *IEEE Transactions on Neural Networks, 13,* 415–425.

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

Johnson, B. & Xie, Z. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS Journal of Photogrammetry and Remote Sensing, 66,* 473–483 (2011).

Karathanassi, V., Kolokousis, P., & Ioannidou, S. (2007). A comparison study on fusion methods using evaluation indicators. *International Journal of Remote Sensing, 28,* 2309–2341.

Kaufman, Y. J. (1989). The atmospheric effect on remote sensing and its correction. In: Asrar, G. (Ed.). *Theory and Applications of Optical Remote Sensing*. New York: Wiley. 752 p.

Laben, C. A., & Brower, B. V. (2000). *Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening (Gram-Schmidt)*. US Patent 6,011,875.

Lahousse, T., Chang, K. T., & Lin, Y. H. (2011). Landslide mapping with multi-scale object-based image analysis – a case study in the Baichi watershed, Taiwan. *Natural Hazards and Earth System Science, 11,* 2715–2726.

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

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

Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing, 25,* 2365–2401.

Lu, P., Stumpf, A., Kerle, N., & Casagli, N. (2011). Object-oriented change detection for landslide rapid mapping. *IEEE Geoscience and Remote Sensing Letters, 8,* 701–705.

Maharaj, R. J. (1993). Landslide processes and landslide susceptibility analysis from an upland watershed: A case study from St. Andrew, Jamaica, West Indies. *Engineering Geology, 34,* 53–79.

Malamud, B. D., Turcotte, D. L., Guzzetti, F., & Reichenbach, P. (2004). Landslide inventories and their statistical properties. *Earth Surface Processes and Landforms, 29,* 687–711.

Mantovani, F., Soeters, R., & Van Westen, C. J. (1996). Remote sensing techniques for landslide studies and hazard zonation in Europe. *Geomorphology, 15,* 3-4.

Mantovani, F., Gracia, F. J., Domenico de Cosmo, P., & Suma, A. (2009). A new approach to landslide geomorphological mapping using the open source software in the Olvera area (Cadiz, Spain). *Landslides, 7,* 69–74.

Martha, T. R., Kerle, N., Jetten, V., van Westen, C. J., & Kumar, K. V. (2010). Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology, 116,* 24–36.

Martha, T. R., Kerle, N., van Westen, C. J., Jetten, V., & Kumar, K. V. (2011). Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. *IEEE Transactions on Geoscience and Remote Sensing, 49,* 4928–4943.

Martha, T. R., Kerle, N., Van Westen, C. J., Jetten, V., & Kumar, K. V. (2012). Object-oriented analysis of multi-temporal panchromatic images for creation of historical landslide inventories. *ISPRS Journal of Photogrammetry and Remote Sensing, 67,* 105–119.

Mas, J. F., & Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing, 29,* 617–663.

McKean, J., & Roering, J. (2004). Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. *Geomorphology, 57,* 331–351.

McKeown, D. M., Cochran, S. D., Ford, S. J, McGlone, J. C., Shufelt, J. A., & Yocum, D. A. (1999). Fusion of HYDICE hyperspectral data with panchromatic imagery for cartographic feature extraction. *IEEE Transactions on Geoscience and Remote Sensing, 37,* 1261–1277.

Metternicht, G., Hurni, L., & Gogu, R. (2005). Remote sensing of landslides: an analysis of the potential contribution to geo-spatial systems for hazard assessment in mountainous environments. *Remote Sensing of Environment, 98,* 284–303.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2012). e1071: Misc functions of the Department of Statistics (e1071), TU Wien. R package version 1.6-1.

Moine, M., Puissant, A., & Malet, J.-P. (2009). Detection of landslides from aerial and satellite images with a semi-automatic method. Application to the Barcelonnette basin (Alpes-de-Haute-Provence, France). In: Malet, J.-P., Remaître, A., & Bogard, T. (Eds.). *Landslide Processes: From*

Geomorphologic Mapping to Dynamic Modelling, *Proceedings of the Landslide Processes Conference, a Tribute to Dr. Theo Van Asch, 6-7 February, Strasbourg, France.*

Möller, M., Lymburner, L., & Volk, M. (2007). The comparison index: A tool for assessing the accuracy of image segmentation. *International Journal of Applied Earth Observation and Geoinformation, 9,* 311–321.

Mondini, A. C., Chang, K.-T., & Yin, H.-Y. (2011a). Combining multiple change-detection indices for mapping landslides triggered by typhoons. *Geomorphology, 134,* 440-451.

Mondini, A. C., Guzzetti, F., Reichenbach, P., Rossi, M., Cardinalli, M., & Ardizzone, F. (2011b). Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sensing of Environment, 115,* 1743–1757.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 66,* 247–259.

Nadim, F., Kjekstad, O., Peduzzi, P., Herold, C., & Jaedicke, C. (2006). Global landslide and avalanche hotspots. *Landslides, 3,* 159–173.

Nichol, J., & Wong, M. S. (2005). Satellite remote sensing for detailed landslide inventories using change detection and image fusion. *International Journal of Remote Sensing, 26,* 1913–1926.

Nicodemus, K. K., Malley, J. D., Strobl, C., & Ziegler, A. (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinformatics, 11*.

Nielsen, A. A., Conradsen, K., & Simpson, J. J. (1998). Multivariate Alteration Detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *Remote Sensing of Environment, 64,* 1–19.

Núñez, J., Otazu, X., Fors, O., Prades, A., Palà, V., & Arbiol, R. (1999). Multiresolution-based image fusion with additive wavelet decomposition. *IEEE Transactions on Geoscience and Remote Sensing, 37,* 1204-1211.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and Planning A, 16,* 17–31.

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

Pal, M. (2006). Support Vector Machine-based feature selection for land cover classification: a case study with DAIS hyperspectral data. *International Journal of Remote Sensing, 27,* 2877–2894.

Paola, J. D., & Schowengerdt, R. A. (1995). A review and analysis of backpropagation Neural Networks for classification of remotely-sensed multispectral imagery. *International Journal of Remote Sensing, 16,* 3033–3058.

Parker, O., & Blesius, L. (2012). Object-based segmentation of multi-temporal Quickbird imagery for landslide detection. *XII International Symposium on Environmental Geotechnology, Energy and Global Sustainable Development, 27-29 June.* Los Angeles, CA..

Prasad, A. M., Iverson, L. R. & Liaw, A. Newer classification and regression tree techniques: Bagging and random foresets for ecological prediction. *Ecosystems, 9,* 181–199 (2006).

Puissant, A., Hirsch, J., & Weber, C. (2005). The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. *International Journal of Remote Sensing, 26,* 733-745.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Schroeder, T. A., Cohen, W. B., Song, C., Canty, M. J., & Yang, Z. (2006). Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon. *Remote Sensing of Environment, 103,* 16–26.

Schuster, R. L. (1996). Socioeconomic Significant of Landslides. In: Turner & Schuster. pp. 11-35.

Schuster, R. L., & Kockelman, W. J. (1996). Principles of Landslide Hazard Reduction. In: Turner & Schuster. pp. 91-105.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall. 175 p.

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and change detection using Landsat TM data: When and how to correct atmospheric effects? *Remote Sensing of Environment, 75,* 230–244.

Statnikov, A., Wang, L., & Aliferis, C. F. (2008). A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics, 9*.

Stehman, S. V. (1992). Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering & Remote Sensing, 58,* 1343–1350.

Stein, A. & de Beurs, K. Complexity metrics to quantify semantic accuracy in segmented Landsat images. *International Journal of Remote Sensing, 26,* 2937–2951 (2005).

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics, 8*.

Stumpf, A., & Kerle, N. (2011). Object-oriented mapping of landslides using random foresets. *Remote Sensing of Environment, 115,* 2564–2577.

Stumpf, A., Lachiche, N., Kerle, N., Malet, J.-P., & Puissant, A. (2012). Adaptive spatial sampling with active random forest for object-oriented landslide mapping. *IEEE International Geoscience and Remote Sensing Symposium, "Remote Sensing for a Dynamic Earth," 22-27 July, Munich, Germany*.

Svetnik, V., Liaw, A., Tong, C., & Wang, T. (2004). Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. Multiple Classier Systems, *Fifth International Workshop, MCS 2004, Proceedings, 9–11 June*. Cagliari, Italy.

Thomas, C., Ranchin, T., Wald, L., & Chanussot, J. (2008). Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics. *IEEE Transactions on Geoscience and Remote Sensing, 46,* 1301–1312.

Tong, X., Liu, S., & Weng, Q. (2010). ISPRS Journal of Photogrammetry and Remote Sensing Bias-corrected rational polynomial coefficients for high accuracy geo-positioning of QuickBird stereo imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 65,* 218–226 (2010).

Torgo, L. (2010). *Data Mining with R: Learning with Case Studies*. Chapman and Hall/CRC Press. 305 p.

Toutin, T. (2004). Review article: Geometric processing of remote sensing images: Models, algorithms and methods. *International Journal of Remote Sensing, 25,* 1893–1924.

Turner, A. K. & Schuster, R. L. (1996). *Landslides—Investigation and Mitigation*. National Research Council, Transportation Research Board Special Report 247. Washington, D.C.: National Academy Press. 673 p.

United States Department of Agriculture, Forest Service [USFS]. (2007). Burned-area Report. FSH 2509.13. Unpublished Report on File. Corona, CA. 9 p.

Updike, T. & Comp, C. (2010). Radiometric Use of WorldView-2 Imagery. Technical Note. DigitalGlobe, Inc. 17 p.

Van Den Eeckhaut, M., Poesen, J., Verstraeten, G., Vanacker, V., Nyssen, J., Moeyersons, J., Van Beek, L. P. H., & Vandekerckhove, L. (2007). Use of LIDAR-derived images for mapping old landslides under forest. *Earth Surface Processes and Landforms, 32,* 754–769.

Van Den Eeckhaut, M., Kerle, N., Poesen, J., & Hervás, J. (2012). Object-oriented identification of forested landslides with derivatives of single-pulse LiDAR data. *Geomorphology, 173-174,* 30-42.

Van Westen, C. J., Seijmonsbergen, A. C., & Mantovani, F. (1999). Comparing landslide hazard maps. *Natural Hazards, 20,* 137–158.

Van Westen, C. J., Van Asch, T. W. J., & Soeters, R. (2006). Landslide hazard and risk zonation – Why is it still so difficult? *Bulletin of Engineering Geology and the Environment, 65,* 167–184.

Vapnik, V. (1999). *The Nature of Statistical Learning Theory*. Springer-Verlag. 344 p.

Varnes, D. J. (1985). *Landslide Hazard Zonation: A Review of Principles and Practices*, UNESCO Natural Hazards Series No. 3. Paris: UNESCO. 63 p.

Vermote, E. F., Tanré, D., Deuzé, J. L., Herman, M., & Morcrette, J.-J. (1997). Second simulation of the satellite signal in the solar spectrum, 6S: An overview. *IEEE Transactions on Geoscience and Remote Sensing, 35,* 675–686 (1997).

Wang, X., & Niu, R. (2010). Landslide intelligent prediction using object-oriented method. *Soil Dynamics and Earthquake Engineering, 30,* 1478–1486.

Weirich, F., & Blesius, L. (2007). Comparison of satellite and air photo based landslide susceptibility maps. *Geomorphology, 87*, 352–364.

Wieczorek, G. F. (1984). Preparing a detailed landslide inventory map for hazard evaluation and reduction. *Bulletin of Association of Engineering Geologists, 21,* 337–342.

Wills, C. J., & McCrink, T. P. (2002). Comparing landslide inventories: The map depends on the method. *Environmental & Engineering Geoscience, 8,* 279–293.

Woodcock, C. E., & Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment, 21,* 311–332.

Yang, X., & Lo, C. P. (2000). Relative radiometric normalization performance for change detection from multi-date satellite images. *Photogrammetric Engineering & Remote Sensing, 66,* 967–980.

Yokoyama, R., Shlrasawa, M., & Pike, R. J. (2002). Visualizing topography by Openness: A new application of image processing to digital elevation models. *Photogrammetric Engineering & Remote Sensing, 68,* 257–265.

Yuhendra, I. A., Sumantyo, J. T. S., & Kuze, H. (2012). Assessment of pan-sharpening methods applied to image fusion of remotely sensed multi-band data. *International Journal of Applied Earth Observation and Geoinformation, 18,* 165–175.

Zhan, Q., Molenaar, M., Tempfli, K., & Shi, W. (2005). Quality assessment for geospatial objects derived from remotely sensed data. *International Journal of Remote Sensing, 26,* 2953–2974.