

## Extracting the full value of the Landsat archive: Inter-sensor harmonization for the mapping of Minnesota forest canopy cover (1973–2015)



Jody C. Vogeler<sup>a,\*</sup>, Justin D. Braaten<sup>b</sup>, Robert A. Slesak<sup>a,c</sup>, Michael J. Falkowski<sup>d</sup>

<sup>a</sup> Department of Forest Resources, University of Minnesota, Saint Paul, MN 55108, USA

<sup>b</sup> College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331, USA

<sup>c</sup> Minnesota Forest Resources Council, Saint Paul, MN 55108, USA

<sup>d</sup> Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO 80523, USA

### ARTICLE INFO

#### Keywords:

Canopy cover  
Landsat time series  
LandsatLinkr  
LandTrendr  
Minnesota

### ABSTRACT

Remote sensing estimates of forest canopy cover have frequently been used to support a variety of applications including wildlife habitat modeling, monitoring of watershed health, change detection, and are also correlated to various aspects of forest structure and ecosystem function. Although data from the long running Landsat earth observation program (1972–present) have been previously utilized to characterize forest canopy cover, the variability in spatial and spectral resolutions between the Landsat sensors has generally limited analyses to readily comparable imagery from the TM and ETM+ sensors, which omits large portions of the full temporal record. In this study, we present an R package, LandsatLinkr, which automates the processes for harmonizing Landsat MSS and OLI imagery to the spatial and spectral qualities of TM and ETM+ imagery, allowing for the generation of annual cloud-free composites of tasseled cap spectral indices across the entire Landsat archive. We demonstrate the utility of LandsatLinkr products, further enhanced through the LandTrendr segmentation algorithm, for characterizing forest attributes through time by developing annual forest masks and maps of estimated canopy cover for the state of Minnesota from 1973 to 2015. The forest mask model had an overall accuracy of 87%, with omission and commission errors for the forest class of 17% and 10%, respectively, and 9% and 16% for non-forest classification. Our resulting maps depicted a significant positive trend in forest cover across all ecological provinces of Minnesota during the study period. A random forest model used to predict continuous canopy cover had a pseudo  $R^2$  of 0.75, with a cross validation RMSE of 5%. Our results are comparable to previous Landsat-based canopy cover mapping efforts, but expand the evaluation time period as we were able to utilize the entire Landsat archive for assessment.

### 1. Introduction

Remote sensing of forest attributes continues to advance the field of forest ecology and management by expanding our spatial and temporal records, ultimately leading to a deeper understanding of forest ecosystem pattern and processes. High (< 10 m) and medium spatial resolution remote sensing data (10–100 m) provide detailed depictions of within-stand forest characteristics, while also providing synoptic views of the complex dynamics and interactions of patches across large spatial extents (Cohen and Goward, 2004). Long running satellite programs, such as Landsat, are expanding opportunities to monitor forest trends through time (Huang et al., 2010; Kennedy et al., 2010), improving our understanding of forest disturbance and recovery patterns (Masek et al., 2013; Kennedy et al., 2015).

Remote sensing estimates of within-stand forest structural

attributes, such as canopy cover, have frequently been used to support a variety of applications related to research and management (Hansen et al., 2013; Koy et al., 2005). When viewing the forest from above, canopy cover is defined as the proportion of the forest floor in a given unit of space that is obscured by the vertical projection of tree canopies (Jennings et al., 1999). Canopy cover often correlates with additional forest structural attributes, such as stand basal area and volume (Jennings et al., 1999), and serves as an important input into fire behavior models (Pierce et al., 2012). Further, as an identified driver of wildlife habitat use, canopy cover may directly provide hiding cover (Schwab and Pitt, 1991) and nesting substrates (Swanson et al., 2008) for certain species, and also governs the amount of available light for understory growth and associated nesting and foraging resources for wildlife (Jennings et al., 1999). The health and functioning of watersheds may also be correlated with canopy cover through thermal

\* Corresponding author.

E-mail address: [jvogeler@umn.edu](mailto:jvogeler@umn.edu) (J.C. Vogeler).

regulation of streams (Moore et al., 2005), the introduction of woody debris (Crook and Robertson, 1999), buffering of nutrient loading (Jones et al., 2001), and erosion control (Hartanto et al., 2003).

Annual maps of historic canopy cover allow for the characterization of forest resources at a given point in time, as well as the monitoring of forest change and recovery trends which aid in the prediction of biotic and abiotic stressors on forest systems into the future. Indeed, the patterns of insects and pathogens are of great interest to many forest managers, which can be identified through slow declines in canopy cover (often represented by vegetation indices) that result from mortality or defoliation through time (Neigh et al., 2014). In addition to monitoring a variety of slow or abrupt forest disturbances, annual maps of canopy cover may aid in the tracking of recovery following disturbance events (Pickell et al., 2016). A variety of methods have been devised by foresters to measure canopy cover (Jennings et al., 1999), although variations in data collection efforts across space and time can make it difficult to assemble a contiguous data set. One useful alternative is leveraging Landsat data, which provide a consistent data source with a temporally rich archive of imagery at spatial resolutions appropriate for characterizing forest canopy cover (Ahmed et al., 2015; Pierce et al., 2012) and is available free to the public as of 2008 (Woodcock et al., 2008). Until recently, however, utilizing Landsat for estimates of canopy cover through time has been constrained by the different spatial, spectral, and/or radiometric properties of the varying Landsat sensors.

New approaches for the harmonization of multi-sensor imagery and creation of comparable vegetation indices are expanding the utility of the Landsat archive for historic forest mapping purposes (Braaten et al., 2017; Pflugmacher et al., 2012; Roy et al., 2016). Although many algorithms for analyzing Landsat time series image stacks have emerged in recent years (Brooks et al., 2014; Huang et al., 2010; Hughes et al., 2017; Jin et al., 2013; Kennedy et al., 2010; Vogelmann et al., 2012; Zhu et al., 2015; Zhu and Woodcock, 2014), most of the applications have been limited to leveraging data from the Thematic Mapper (TM) (1984–2012) and the Enhanced Thematic Mapper Plus (ETM+) (1999–present) sensors. Exclusion of data from the earliest sensor, the Multispectral Scanner (MSS) (1972–1999), and the latest sensor, the Operational Land Imager (OLI) (2013–present), are likely due to differences in the spatial, spectral, and/or radiometric resolutions of these sensors, which require much additional processing to incorporate them harmoniously into a time series with TM and ETM+ data. However, the additional twelve years (1972–1984) of imagery available through MSS sensors may improve the value of the Landsat record for characterizing forest ecosystems dynamics, as the cumulative time series approaches a more ecologically significant amount of time (Pflugmacher et al., 2012), and inclusion of OLI ensures continuation past ETM+.

In addition to harmonization between sensors, pixel level characterizations of forest attributes through time may benefit from the removal of year-to-year noise inherent to spectral imagery to better depict realistic patterns of forest recovery and change. Although Landsat time series change detection often utilizes such a segmentation or fitting procedure as an initial step in the identification of disturbance patches, such as that used in the LandTrendr algorithm (Kennedy et al., 2010), few studies have focused on the value of such fitted and smoothed annual products for the mapping of more specific forest attributes (Moisen et al., 2016).

In this study, we present an automated system for normalizing Landsat MSS and OLI imagery to the spatial and spectral qualities of TM and ETM+ imagery, allowing for the generation of annual cloud-free composites of spectral indices across the Landsat archive. This system, termed LandsatLinkr, is implemented as a code library for the R programming environment (R Development Core Team, 2016). We demonstrate the utility of LandsatLinkr and subsequent LandTrendr (Kennedy et al., 2010) fitted products for characterizing forest attributes through time by developing annual forest masks and maps of estimated canopy cover for the state of Minnesota from 1973 to 2015.

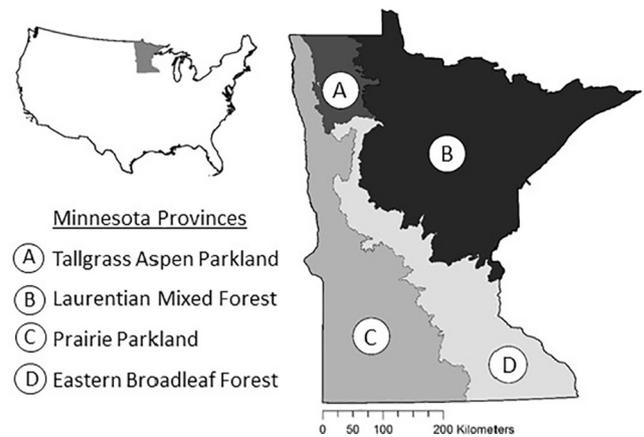


Fig. 1. The Minnesota, USA, study area divided by ecological provinces.

We focus on Minnesota where forests are not only vital for sustaining one of the largest state industries, timber, but are also important components of wetland systems that cover a large portion of the state, and as critical habitat for many wildlife species of conservation interest. Therefore, maps of canopy cover across the state and through time may provide valuable resources to a variety of Minnesota land management, monitoring, and research efforts.

## 2. Methods

### 2.1. Study area

The area of study is the entire state of Minnesota, USA, which encompasses the Laurentian Mixed Forest, Eastern Broadleaf Forest, Prairie Parkland, and Tallgrass Aspen Parklands ecological provinces (MN DNR, 1999; Fig. 1). Statewide, agricultural and forest land comprise approximately 50% and 30%, respectively, of total area. Surface waters cover approximately 10% of the total area, with the remaining 10% including managed grasslands, developed urban, and mining land uses (Rampi et al., 2016). There is a strong agricultural to forest land cover gradient extending from the southwest to northeast portions of the state. The regional climate is continental, with mean annual precipitation ranging from 500 to 800 mm and a mean growing season (May–Oct.) temperature of 11–16 °C. Annual precipitation is roughly comprised of about one-third snow and two-thirds rainfall. A variety of forest types occur in the region, but dominant forest types include the aspen-birch and spruce-fir types, and to a lesser extent oak-hickory and pine (Miles and VanderSchaaf, 2015). Wetlands, including forested bogs, peatlands, and swamps, are found extensively throughout the state (MPCA, 2015).

### 2.2. Landsat imagery

The Landsat earth observation program has been collecting satellite image data from 1972 to present. This archive represents the longest global earth observation record from remote sensing. Landsat sensors include MSS, TM, ETM+, and OLI, which have been deployed on eight different satellites (only seven of which attained orbit). Satellites 1–3 only carried the MSS sensor, 4–5 carried both MSS and TM sensors (coincident image pairs), 6 carried the ETM sensor (was lost on launch), 7 carries the ETM+ sensor, and 8 carries the OLI and Thermal Infrared (TIRS) sensors. Satellites 1–3 had a higher altitude orbit as compared to satellites 4–8, so image data exists according to two different World Reference System (WRS) grids, the former being in WRS-1 and the latter being WRS-2. To fully include the area of the state of Minnesota in our analysis, we utilized Landsat imagery from 28 WRS-1 (MSS) and 28 WRS-2 (MSS, TM, ETM+, and OLI) scenes to create annual growing

season composites of tasseled cap spectral indices (Crist and Cicone, 1984). We viewed mean NDVI curves by day-of-year for all scenes to determine an ideal range of peak growing season dates from which to select our imagery (early July–early September). A total of 5945 images across all scenes and years were downloaded from the USGS Earth Explorer website (<http://earthexplorer.usgs.gov/>) and incorporated into our inter-sensor harmonization and creation of annual composites. After initial assessments, we found that there was insufficient image coverage of our study area during the first year of Landsat MSS, 1972, thus we removed this year from our study. We requested and downloaded the Surface Reflectance High Level Data Product available for TM and ETM+ (LEDAPS algorithm; Masek et al., 2006), and OLI (L8SR algorithm; Vermote et al., 2016) images, and Level 1 Product Generation System (LPGS) images for MSS. TM and ETM+ surface reflectance images were considered the spatial and spectral standard for the time series and remained unaltered. MSS and OLI data, however, went through a normalization process to harmonize them with the TM/ETM+ data.

The work flow for harmonizing the MSS and OLI data to TM/ETM+ and generating annual cloud-free composites was handled by the LandsatLinkr (LLR) system (Braaten et al., 2017). LLR is an R package designed to automate much of the pre-processing and harmonization required to create annual cloud-free composites of tasseled cap indices across a user-defined study area throughout the Landsat archive. Using previously established methodologies, LLR spectrally and spatially aligns images from multiple sensors through additional georegistration, resampling of coarser resolution MSS images, and the modeling and harmonization of tasseled cap indices. The six LLR steps include: 1) MSS unpacking; 2) TM/ETM+ unpacking; 3) OLI unpacking; 4) MSS to TM harmonization; 5) OLI to ETM+ harmonization; and 6) annual composite creation (Fig. 2). Steps 1–5 are completed at the scene level and step 6 incorporates all scenes required to cover a user-defined study area.

In the first step, MSS images are unpacked (decompressed, stacked, re-projected) and georegistration accuracy is assessed. For those images with an a total positional RMSE  $\geq 0.5$ , the georegistration is enhanced using an image-to-image tie point search and warp procedure (MSSwarper), which is based on the methods presented in Kennedy and Cohen (2003). Unfortunately, in our case there were a large number of MSS images for some of our scenes that were not available as desired USGS L1T products (i.e., orthorectified) for which LLR is set up to handle. They were instead only available as L1G products (systematic geocorrection based on spacecraft ephemeris data) which left gaps in our time series stacks. To improve the initial georegistration of L1G images, we did a shift and rotation procedure in ArcGIS by manually selecting three tie points from an additional 96 images. Following this step, we then continued with MSSwarper on the manually corrected L1G images, further fine-tuning the georegistration and filling in the gaps in our annual stacks. LLR then further radiometrically corrects the MSS images to top-of-atmosphere and surface reflectance. Surface reflectance is calculated using the COST method (Chavez, 1996), with the *Lhaze* parameter or “dark object value” estimated from an automated implementation of the histogram method (Chavez, 1988). Finally, cloud masks are created following the MSSsvm procedure (Braaten et al., 2015), which along with surface reflectance are resampled to 30 m to match the spatial resolution of the later Landsat sensors.

In the next LLR steps TM, ETM+, and OLI image data are automatically decompressed, stacked, re-projected, and a cloud mask is derived from the product-included Fmask (Zhu and Woodcock, 2012) cloud and cloud shadow mask. Fmask reclassification is performed to match the MSS cloud mask classes produced above to allow efficient mask application during the compositing process. For the TM and ETM+ data, tasseled cap spectral transformations brightness, greenness, and wetness (Crist, 1985) are produced, as well as a derivation of greenness and brightness called the tasseled cap angle (Powell et al., 2010).

LLR steps 4 and 5 involve the inter-sensor harmonization of MSS to TM and OLI to ETM+ indices. MSS images are normalized to the TM images through the modeling of TM tasseled cap indices from the 4 MSS bands using samples from offsets of coincident images and multiple linear regression; a similar approach presented by Roy et al. (2016) is utilized to transform OLI data to TM. This is done separately for each tasseled capped index to create index specific models. Since the models incorporate variance introduced by multiple individual MSS/TM relationships from the samples within the sets of coincident images (which occurred during the period of overlap between MSS and TM with Landsat 4 and 5), the final models represent mean or aggregate models which are then applied to all MSS images from a given scene. Since aggregate models were created using MSS Landsat 4 and 5 which are on the WRS-2 grid, and MSS Landsat 1–3 are WRS-1, the model from the WRS-2 scene that most overlaps each WRS-1 scene is used for harmonization. OLI images are harmonized to ETM+ images in the same way as MSS to TM, except that only 6 of the 7 OLI reflectance bands are included as predictors (band 1 – ultra blue is excluded), and near-date images are used since there are not coincident image pairs available for ETM+/OLI. We evaluated the performance of the LLR harmonization between sensors by calculating the difference between annual summer image composites produced for each sensor for the same year during sensor overlaps, which are intermediate products available from LLR.

The final step of LLR is to create annual composites of the tasseled cap indices. All of the previous LLR steps are completed on an individual scene basis, while the compositing step is applied to a user defined area which can encompass multiple WRS-1 and WRS-2 scenes. Images from a given year are compiled using one of several summary metric options within the compositing function, ignoring clouds, shadows, or no data pixels such as those from the ETM+ SLC-off lines. We chose to utilize the median summary statistic as we found this method best removed any remaining pixel noise from the composites and was the most suited for our purposes.

After creating annual composite stacks using LLR, we then applied the LandTrendr temporal segmentation algorithm presented in Kennedy et al. (2010). While LandTrendr is mostly used as a change detection algorithm, the first step involves a segmentation procedure that fits vertices to the spectral trends for each pixel. Among the outputs are revised annual pixel values interpolated from lines between the vertices. We used this approach to smooth the annual trends to minimize year to year noise, providing a better representation of true forest dynamics, which do not usually exhibit abnormal growth spikes or dips for a single year (Fig. 3). In addition to the annual spectral indices from Landsat, we also included topographic variables in our modeling efforts. We downloaded the Minnesota 30 m Digital Elevation Model (DEM; MDNR, 2004) to represent elevation and to create maps of slope, aspect, and several aspect transformations in ArcGIS 10.3 (Table 1).

### 2.3. Canopy cover estimates

To generate canopy cover estimates to use in model training and validation, we randomly identified 1340 sample locations across the state which we first evaluated to ensure that all forest and non-forest classes in the 2011 National Land Cover Dataset were represented (NLCD; Homer et al., 2015). Following similar methodologies as Coulston et al. (2012), we created  $3 \times 3$  Landsat pixel windows ( $90 \text{ m} \times 90 \text{ m}$ ) surrounding each sample location. Within each  $90 \text{ m} \times 90 \text{ m}$  sample plot, we generated a 100 point dot grid using the fishnet tool in ArcGIS 10.3. These dot grids were overlaid onto false color composites from 2008 4-band National Agriculture Imagery Program (NAIP) aerial imagery, and the proportion of dots within each sample window that intersected tree crowns was used to estimate canopy cover for that plot. The observer used surrounding landscape in addition to the plot view within the NAIP imagery to aid in differentiating between forest and non-forest vegetation prior to designating

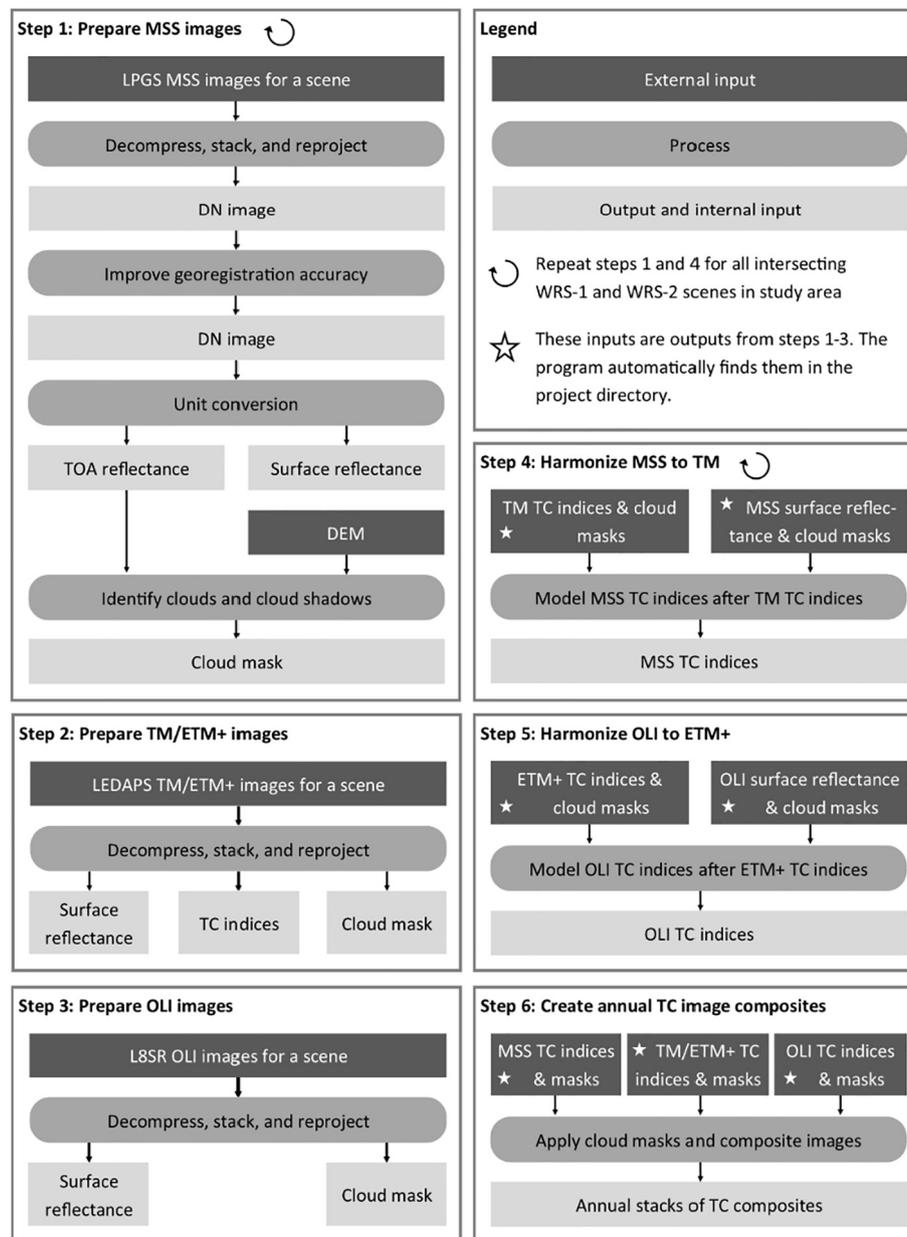


Fig. 2. Work flow for LandsatLinkr, an R package which automates the pre-processing, harmonization, and compositing of imagery from all sensors throughout the Landsat archive (1972–present).

canopy vs. non-canopy within the sample plots. When using aerial imagery for estimating canopy cover, observers must decide how to classify shadows within forest stands. In our study, we choose to classify dark shadows within a forest stand as non-canopy cover. A single observer conducted the photo interpretation to minimize any potential bias between different observers (Frescino and Moisen, 2012).

2.4. Forest cover models

Landsat tasseled cap and topography predictors corresponding to 2008, the year of NAIP imagery acquisition, were extracted for each 90 m square plot using zonal statistics in ArcGIS 10.3 (Table 1). We first created a statistical model to differentiate between forest and non-forest where plots with < 10% canopy cover were considered as non-forest. To create this forest mask model, we employed the random forest (RF) statistical modeling approach (Breiman, 2001). The RF algorithm is a classification and regression tree technique that has achieved excellent results in the generation of classifications or continuous predictions

from remotely sensed data (e.g., Falkowski et al., 2009; Hudak et al., 2008; Lawrence et al., 2006). Indeed, RF has rapidly been adopted by the ecological modeling community as an attractive alternative to traditional statistical approaches because of its flexibility in handling data with non-normal distributions, a large number of predictor variables, and the use of averaged bootstrap training samples for improved predictions (Cutler et al., 2007). In the case of classification, the RF algorithm develops classifications by growing numerous (100s to 1000s) classification trees from a random subset of training data (approximately 63% random subset), while randomly permuting predictor variables at each node. The RF algorithm then determines the final classification by selecting the most common classification outcome at each node within the group of multiple trees (Breiman, 2001; Lawrence et al., 2006; Prasad et al., 2006). Bootstrap error estimates are calculated for each tree in the forest by classifying the response variable(s) from each observation in the portion of training data not selected for model development (approximately 37% of the training data). After all the trees in the forest are grown, overall classification error is

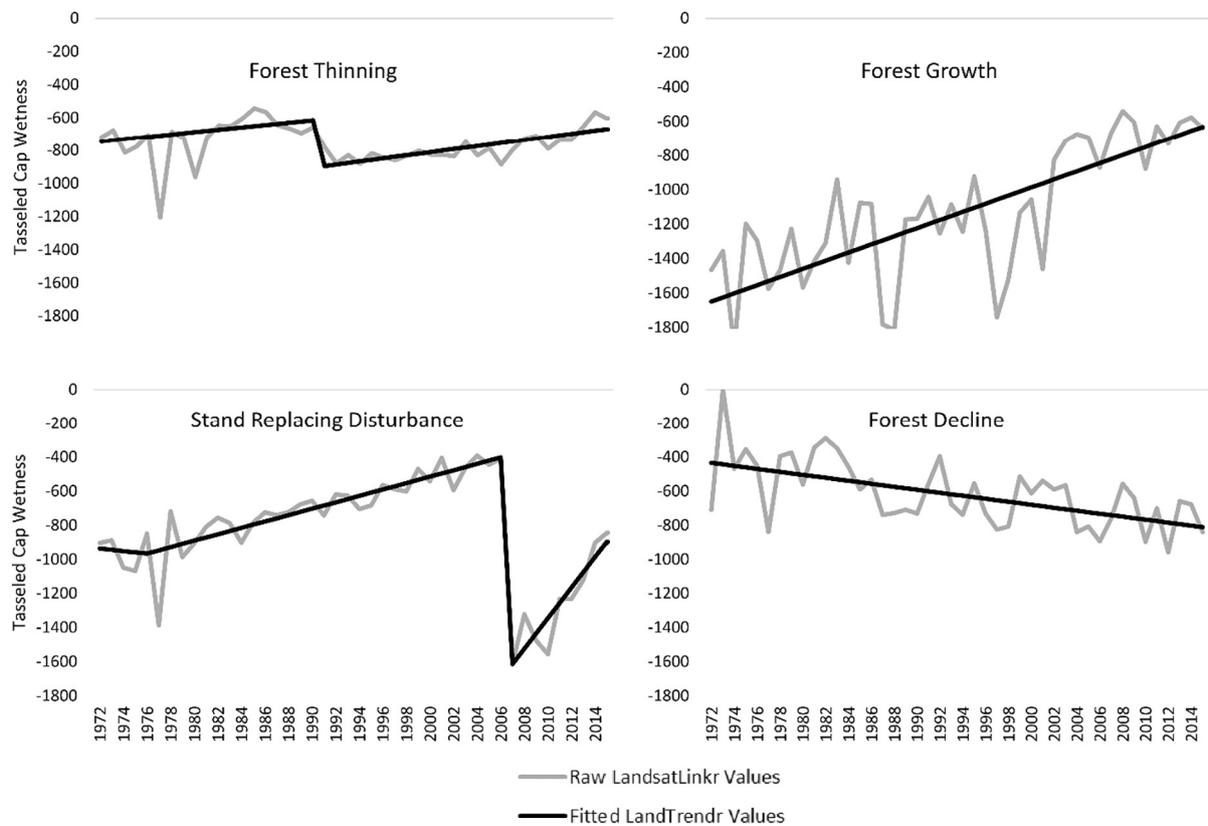


Fig. 3. Comparison of raw annual Landsat composite values from LandsatLinkr and fitted vertices from LandTrendr segmentation algorithm. Examples include tasseled cap wetness values for four pixels in Minnesota.

calculated by averaging errors across all trees in the forest; this is analogous to cross-validated estimate of error and accuracy (Cutler et al., 2007). The algorithm also calculates the influence each predictor variable has upon model accuracy based on the ratio of improvement in the mean squared error across bootstrap replicates, which can be used

to determine the relative importance of each variable used in the classification. For this study, the RF algorithm was implemented as the RandomForest package (Liaw and Wiener, 2002) in the R statistical program (R Development Core Team, 2016).

Although collinearity between predictors has been reported as less

Table 1

Descriptions and abbreviations for Landsat-derived tasseled cap predictors and DEM extracted topographic metrics included in the forest mask and canopy cover models. Model metrics were summarized for the 90 m × 90 m canopy cover training plots and the models were later mapped using 3 × 3 focal grids of the predictors. “X” indicates the inclusion of a predictor in the final random forest mask (RF Mask) and canopy cover (RF Cover) models.

Predictor	Abbrev.	Description	Included in final models	
<b>Landsat</b>				
Tasseled cap brightness		Soil reflectance <sup>a,b</sup>	RF Mask	RF Cover
Mean	TCB_mn		X	X
Standard deviation	TCB_sd		X	
Tasseled cap greenness		Variations in green vegetation <sup>a,b</sup>		
Mean	TCG_mn		X	X
Standard deviation	TCG_sd			
Tasseled cap wetness		Correlated with forest structure <sup>a,c</sup>		
Mean	TCW_mn		X	X
Standard deviation	TCW_sd		X	X
Tasseled cap angle		TCA = arctan(TCG/TCB) <sup>d</sup>		
Mean	TCA_mn		X	X
Standard deviation	TCA_sd		X	
<b>Topography</b>				
Elevation	ELEV	Elevation (meters)		
Slope	SLOPE	Degree topographic slope	X	
Aspect transformed	ASPECT	ASPECT = COS * [45 - (aspect in degrees)] + 1		
ssina	SSINA	SSINA = percent slope * SIN * (aspect in degrees)		
scosa	SCOSA	SCOSA = percent slope * COS * (aspect in degrees)		
<b>Prediction maps</b>				
Forest mask prediction	MASK	Forest vs. non-forest prediction included as a factor	NA	X

<sup>a</sup> Presented in Crist (1985).

<sup>b</sup> Crist and Cicone (1984).

<sup>c</sup> Cohen et al. (1995).

<sup>d</sup> Presented in Powell et al. (2010).

**Table 2**

Estimated non-forest and forest classification error matrix with cells representative of the error-adjusted proportions of area. Observed classes are the columns, while predicted classes are the rows. Reported users', producer's, and overall accuracy measures include 95% confidence intervals.

Class	Non-forest	Forest	Total	User's	Producer's	Overall
Non-forest	0.46	0.09	0.55	0.85 ± 0.02	0.91 ± 0.02	0.87 ± 0.02
Forest	0.05	0.41	0.45	0.90 ± 0.02	0.83 ± 0.02	
Total	0.51	0.49	1			

of a concern with a RF approach, recent research has suggested otherwise (Murphy et al., 2010). Thus, we employ a model selection procedure to determine the optimal suite of predictor variables to use in the classification of forest presence and absence. The model selection procedure was implemented via the rUtilities package in R (Evans and Murphy, 2017) to develop the most parsimonious classification model, while retaining the highest possible classification accuracy. In order to reduce data redundancy and improve overall model interpretability, multi-collinear predictor variables were identified and removed via a multivariate variable screening process based upon Gram–Schmidt QR-Decomposition (Evans and Murphy, 2017; Falkowski et al., 2009). The final classification model is arrived at based on the criteria of smallest total and within class errors, and fewest numbers of predictor variables. In order to stabilize classification error, each RF model was run with 500 bootstrap replicates (i.e., individual classification trees), and then evaluated for the point of MSE stabilization which occurred at 400 trees for this model. After constructing our error matrix for the final forest mask model, we followed the methods presented in Olofsson et al. (2013) to calculate a poststratified estimator to translate the matrix into terms of unbiased proportions of area in the forest and non-forest classes. These sample-based estimates of area were incorporated into classification accuracy measures and associated confidence intervals (Olofsson et al., 2013).

We used a similar RF model selection procedure for canopy cover as with our forest mask models with the exception of designating a regression tree approach rather than a classification tree approach. In addition to the spectral and topographic variables, we included predictions from our forest mask classification as a factor in the models to aid in the continuous prediction of forest canopy cover across all land cover types within one model. To avoid overfitting the model, we chose the appropriate number of decision trees from our initial model run which incorporated all predictors by determining the number of trees at which the MSE had stabilized from the RF tree plot (300 trees). To assess bootstrap prediction errors, we averaged 100 iterations where 70% of the data was used to train each model and 30% was retained for validation. We calculated pseudo R<sup>2</sup> values using the model mean square error (MSE) and variance in the response variable (y) with the equation:

$$R^2 = 1 - \frac{MSE}{var(y)} \tag{1}$$

Relative importance of model predictors were assessed using variable importance values reported by RF which represent the percent increase in model MSE if a predictor was randomized.

### 2.5. Annual prediction mapping

To match the scale of the model variables, we created focal grids of the annual Landsat tasseled cap indices and static topographic predictors in ArcGIS at the 30 m Landsat resolution, where the center pixel within a moving focal window is assigned a summary statistic for the larger 3 × 3 pixel window. We applied the selected forest mask and canopy cover models to the Landsat time series stacks for the state of Minnesota using the focal grids within R to create annual maps of forest

masks and canopy cover from 1973 to 2015. While cover was modeled as a proportion (0–1), during predictive mapping we multiplied by 100 to produce percent canopy cover products.

We used the stacks of predictive maps for a general assessment of change in forest area and cover across the state through time. All trends presented for the forest/non-forest classification maps reflect error-adjusted estimations of area (Olofsson et al., 2013), although the unbiased estimator was calculated from the single year (2008) utilized in model creation and validation. While the ultimate objective of this study was to create maps of continuous cover through time, we derived five classes from our continuous cover products to aid in the visualization of temporal trends: (1) < 10% (from here on referred to as non-forest); (2) 10–24%; (3) 25–49%; (4) 50–74%; (5) ≥ 75%. We also divided the maps by ecological provinces (Fig. 1) to evaluate differences in trends across Minnesota ecoregions. The summaries of forest cover and cover classes through time were fit with linear coefficients and evaluated for significance of general trends across the full span of the time series using *t*-tests.

## 3. Results

### 3.1. Forest cover models

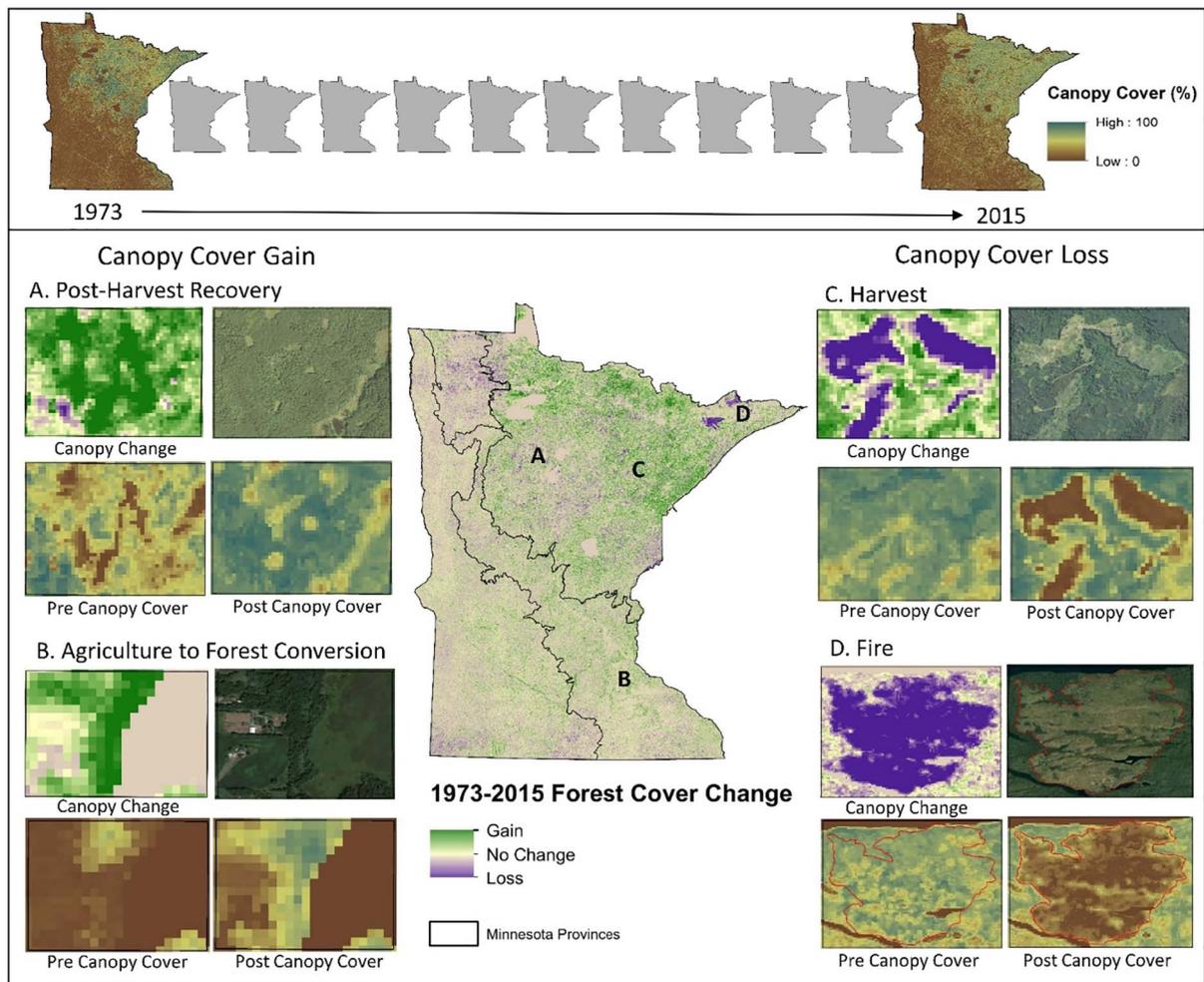
Through the use of Landsat-derived spectral indices and topographic information, we were able to create models of forest cover for the state of Minnesota. After the model selection procedure, our final random forest classification model of forest vs. non-forest included TCB\_mn, TCB\_sd, TCG\_mn, TCW\_mn, TCW\_sd, TCA\_mn, TCA\_sd and SLOPE predictor variables (Table 1). After incorporating the error-adjusted estimator of area, the forest mask model had an overall accuracy of 87% (Table 2), with omission and commission errors for the forest class of 17% and 10%, respectively, and 9% and 16% for non-forest classification. The selected random forest model for canopy cover had a pseudo R<sup>2</sup> of 0.75 while minimizing prediction errors with a cross validation RMSE of 5%. The importance of predictors was ranked as follows according to the RF algorithm (with values of relative importance): forest mask prediction (37.24); TCW\_mn (26.76); TCA\_mn (26.04); TCB\_mn (20.01); TCW\_sd (9.94); and TCG\_mn (9.61).

### 3.2. Landsat time series stacks

Landsat spectral indices harmonized across sensors and fitted to temporal trend lines facilitated the annual mapping of the forest mask and continuous canopy cover models across the state of Minnesota from 1973 to 2015 (Fig. 4). The R package presented here, LandsatLinkr (LLR), efficiently executed all necessary steps for the creation of the spatially and spectrally comparable annual tasseled cap composites (Fig. 5). These composites spanned the entire Landsat archive and successfully minimized seam lines between scenes, maximized data coverage on an annual basis, and adequately removed clouds and cloud shadows from the images. Evaluation of the distribution of differences between image data from two different sensors from the same annual composite period for a sample of points (n = 5964 for each year, distributed randomly throughout MN) show medians near zero and tight variance (Fig. 5), providing support for the successful inter-sensor harmonization by LLR.

### 3.3. Forest cover mapping and time series trends

Annual values of forest area (> 10% forest cover) from the mask maps depict a significant, although slight, positive trend across the state throughout the time period of the study (1973–2015; Fig. 5). Note that this estimate does not account for changes in forest types or stages of structural development, and may include some areas not commonly considered forests (e.g., shrub-scrub wetlands). Forest area accounted for 43.9% (± 2%) of Minnesota land area in 1973, expanding to 52.5%



**Fig. 4.** Representation of the annual stacks of forest canopy cover maps created for 1973–2015 including summary maps of canopy cover change within the Minnesota ecological provinces over the 43 year study period. Subsets are provided for example sources of canopy cover gain and loss observed during the study period (note that we make no claim that these examples represent dominant agents of change for Minnesota as this is beyond the scope of this study). Each example includes a local view of the change event from cover change (1973–2015) maps, a google earth photo corresponding to the post change condition, and pre- and post-change percent canopy cover for years directly prior and following each specific change event.

( $\pm 2\%$ ) by 2015 (water bodies were removed from this analysis). This significant positive trend was consistent across all Minnesota provinces (Fig. 6), where the Eastern Broadleaf and Laurentian Mixed Forest provinces experienced the greatest growth. While most of the provinces exhibited a constant positive trend across the 43-year study, the Tallgrass Aspen Parkland province began with 10 years of decline followed by an increase in forest for the remainder of the time series (Fig. 6). Examinations of the prediction maps revealed a slight overestimation of forest area within agricultural dominated areas on the western edge of Minnesota, primarily within the Prairie Parkland province.

The five cover classes including a non-forest class derived from the continuous canopy cover values facilitated further evaluation of the canopy cover mapping products and temporal trends for additional context and utility (Fig. 7). The histogram of cover values from the photo interpretation plots used in model creation were well represented by those of the corresponding 2008 cover map, with the exception of a slight under representation of the highest cover class in the map ( $\geq 75\%$  cover). Across the time-period of the study, all ecological provinces with the exception of the Tallgrass Aspen Parkland exhibited significant decreases in the non-forest cover class, corresponding to the significant increase in overall forest cover observed through our forest mask maps (Fig. 8). Trends within the moderate cover classes (10–24% and 25–49%) varied across the provinces (Fig. 8). All regions of the state experienced significant growth in the land area within the two highest

cover classes with the exception of a non-significant trend for the 50–74% class within the Tallgrass Aspen Parkland (Fig. 8). The greatest growth in these two high cover classes occurred in the Laurentian Mixed Forest province.

#### 4. Discussion

Through the use of the freely available Landsat archive, we created forest masks and maps of canopy cover for the state of Minnesota from 1973 to 2015, a valuable data set for a variety of research and management applications. This was in large part facilitated by the automation of processing steps by LandsatLinkr for the creation of spatially and spectrally harmonized image data from all Landsat sensors and multiple scenes. The harmonized spectral index predictors were further enhanced with the use of the LandTrendr segmentation algorithm for the smoothing of year-to-year spectral variations not associated with forest change dynamics. Many of the previous studies utilizing Landsat time series products for the mapping of forest attributes are temporally restricted to the post-MSS years ( $> 1984$ ; Potapov et al., 2015; Powell et al., 2010). The harmonization of MSS imagery to the later sensors allowed us to add an additional 11 years to our stacks of mask and cover maps, adding over a decade of identified patches of forest cover change and expanding the records of general forest area and cover class trends.

The accuracy of our canopy cover model is comparable to those

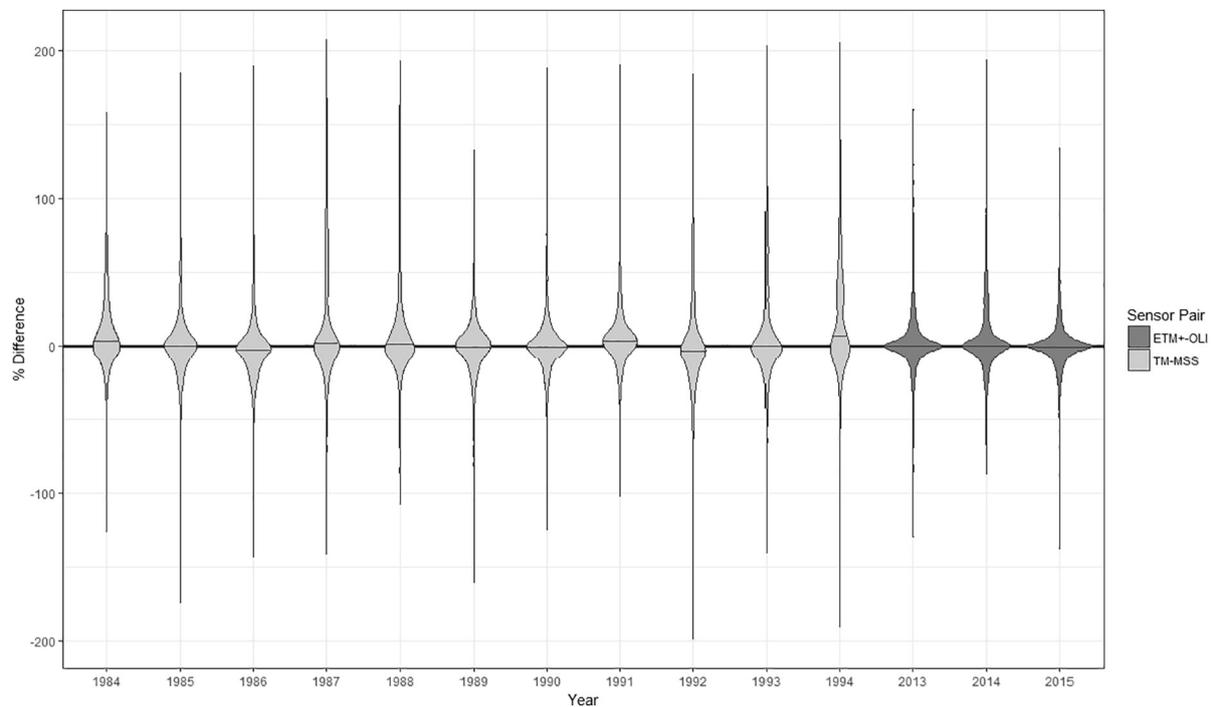


Fig. 5. Distributions of pair-wise difference between LandsatLinkr (LLR)-produced Tasseled Cap Wetness (TCW) annual image composites for years when two sensors concurrently acquired data. Differences between sensor data is cast as a percent of the range between two standard deviations surrounding the mean of TCW for Minnesota. Positive differences indicate that MSS and OLI TCW values are greater than TM and ETM+, respectively, and less, when negative.

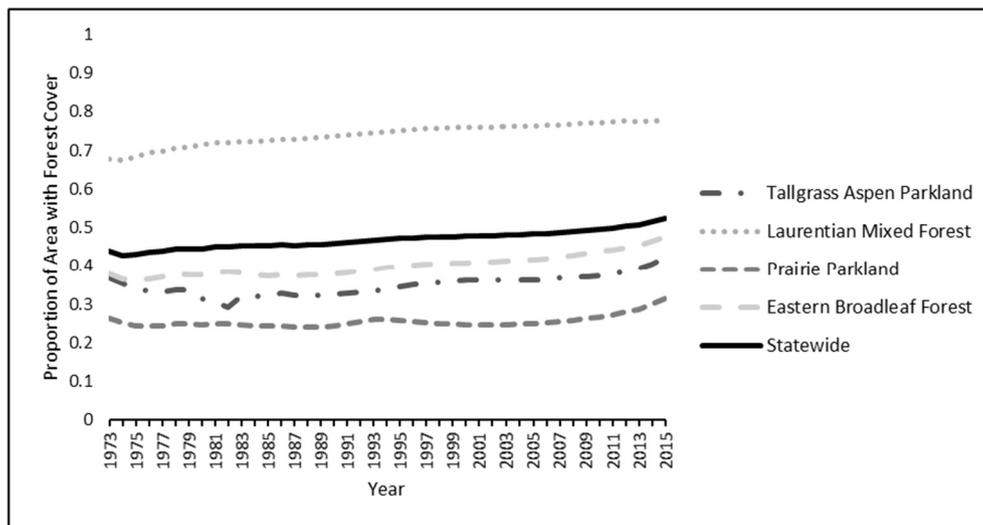


Fig. 6. Temporal trends (1973–2015) in the proportion of land statewide and within ecological provinces with forest cover derived from annual forest mask maps.

from previous Landsat studies (Ahmed et al., 2015; Pierce et al., 2012), although utilizing predictor data available for the entire Landsat time series and encompassing a larger spatial extent and species composition gradient compared to many of the previous studies. With a few exceptions (Sexton et al., 2013; Hansen et al., 2011) including the NLCD tree cover products (Homer et al., 2015), many of the previous Landsat based canopy cover modeling efforts have been conducted within restricted forest types and/or small geographic areas (Ahmed et al., 2015; Carreiras et al., 2006). Pierce et al. (2012) utilized Landsat TM data for modeling aspects of canopy structure including cover for inclusion in fire behavior models within the coniferous dominated forests of Lassen Volcanic National Park, CA (42,900 ha). The study incorporated band reflectance and additional spectral indices to a similar set of tasseled cap and topographic metrics as our study with moderate model performance ( $R^2 = 0.66$ ; Pierce et al., 2012). Through the incorporation of

Landsat-derived disturbance information, Ahmed et al. (2015) were able to stratify the forests in their 2600 ha study area on Vancouver Island, BC Canada, into mature and young classes prior to canopy cover modeling efforts. The initial stratification improved model performance for mature forests ( $R^2 = 0.72$ ) over the combined data set ( $R^2 = 0.67$ ), although resulting in weaker model performance for the young forest class with an  $R^2$  of 0.59; RMSE remained constant at 7% for the divided and combined cover models (Ahmed et al., 2015). In our study, we were able to create a single canopy cover model for the entire state of Minnesota (> 22 million ha;  $R^2 = 0.75$ , RMSE = 5%) which includes 15 NLCD land cover classes, distinguishing forest from non-forest and estimating cover for forest types representing a range of composition from deciduous to coniferous and wetland to upland stands. Tasseled cap wetness was the strongest spectral predictor in our canopy cover models, consistent with findings of previous studies (Hadi et al., 2016;

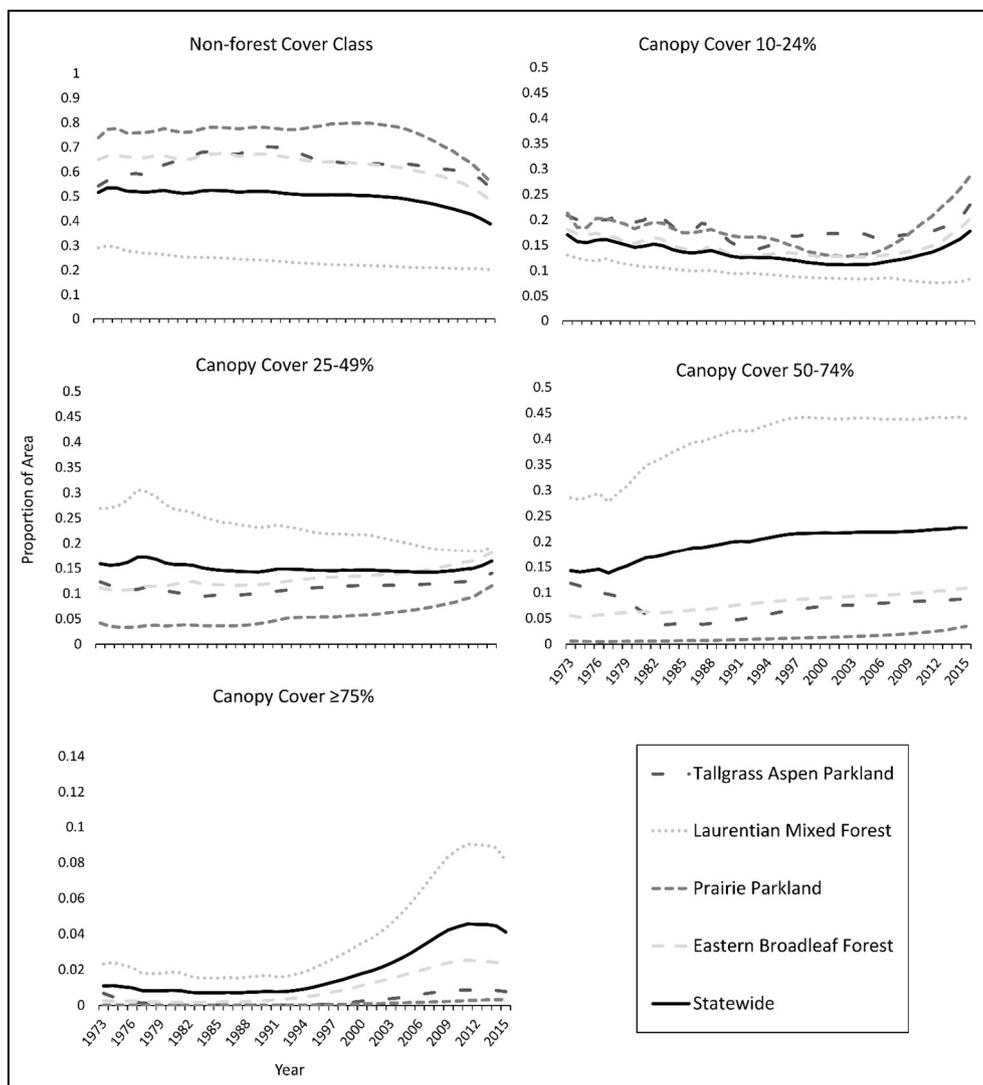


Fig. 7. Temporal trends (1973–2015) of proportion of land within forest canopy cover classes statewide and within Minnesota ecological provinces. Water bodies have been masked out and are not reflected in proportions. Note that the y-axis ranges vary to better depict the trends for each canopy cover class by province.

Healey et al., 2006).

Several continental to global scale tree cover products incorporating Landsat data do exist, although restricted in their ability to be applied to the Landsat archive due to the predictors utilized in their approach (Sexton et al., 2013; Coulston et al., 2012). The national tree cover models from the 2011 NLCD project achieved a range of predictive strengths when models were tested in the five focal areas across the country ( $R^2$  values from 0.65–0.87; Coulston et al., 2012); our results fall within this range. In addition to the NDVI and tasseled cap predictors, the 2011 NLCD cover models also incorporated the previously created 2001 tree cover maps and land cover data (Coulston et al., 2012). The NLCD tree cover product provides an extremely useful spatial data set for a large suite of applications across the continuous United States, although the need for land cover classifications as model inputs complicate the repeatability on an annual basis. Our approach may be an alternative for monitoring change at local or regional extents in the interim of more spatially continuous decadal projects such as NLCD. On a global scale, Sexton et al. (2013) utilized Landsat TM and ETM+ surface reflectance with additional axillary data at two time steps (2000 and 2005) to model the global MODerate-resolution Imaging Spectroradiometer (MODIS) Vegetation Continuous Fields (VCF) Tree Cover product, rescaling the 250 m tree product to the 30 m Landsat scale. When results were summarized across four validation

areas where the study utilized lidar canopy products as reference data, the Landsat-based model had an  $R^2$  of 0.811 and an RMSE of 14.637 (Sexton et al., 2013). Our utilization of spectral indices, which can be created for the entire Landsat archive, provides the opportunity to create annual historic records of forest cover for over four decades into the past and that are repeatable into the future. Through the harmonization of MSS and OLI imagery to TM/ETM+, our study assumes that the model created for a single time will be applicable to other years in the time series. We acknowledge that the accuracies presented here are for the single year model and the change products derived from the time series application of this model may have reduced accuracies.

Annual canopy cover maps may serve as model inputs for wildlife habitat relationships and additional forest attributes at single points in time, as well as providing the opportunity to monitor and interpret trends through time. Vogeler and Cohen (2016) reported that almost 80% of the US Fish and Wildlife Service habitat suitability models for forest inhabitants included some measure of canopy cover, highlighting the value of this metric for predicting and monitoring wildlife habitat for many species. In an effort to identify potential habitat for Myanmar's endangered Eld's deer (*Cervus eldi*), Koy et al. (2005) found significant relationships between their Landsat-derived canopy cover maps and deer distributions. There is increasing interest in leveraging historic datasets such as Landsat to map forest disturbance patches through

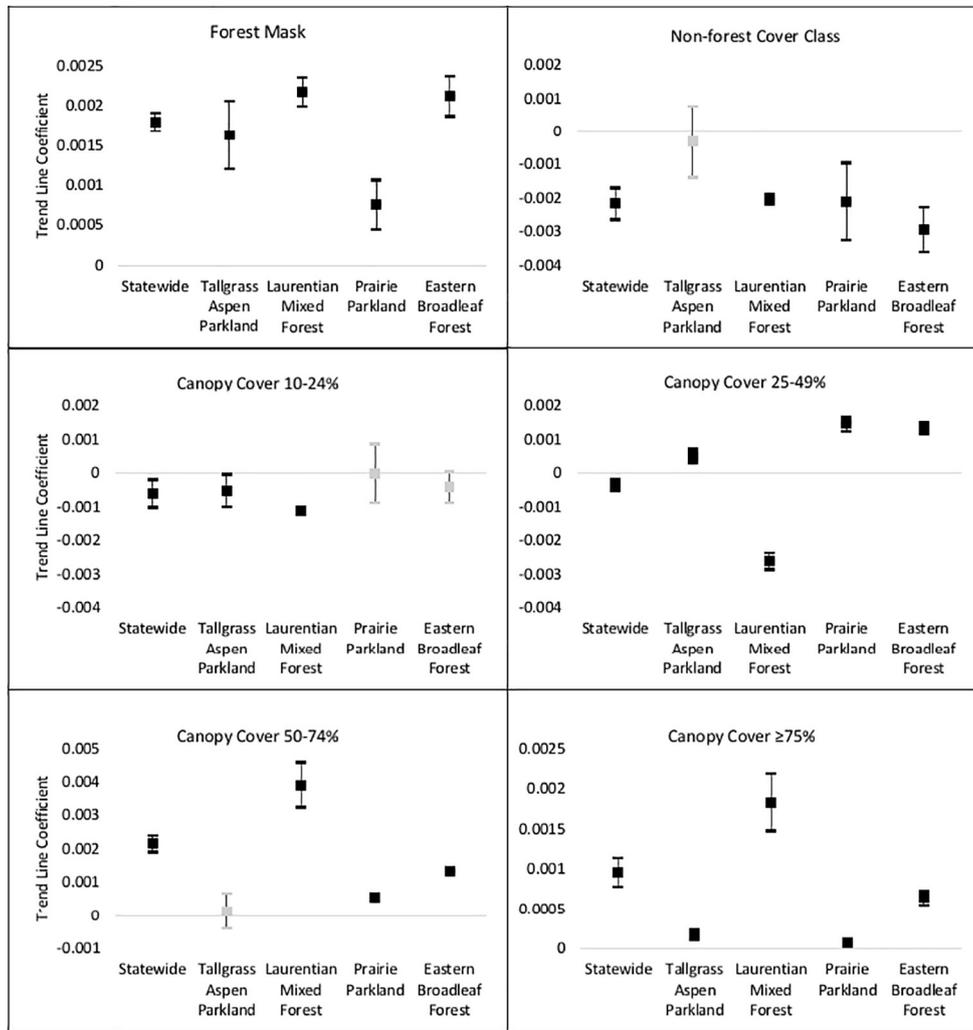


Fig. 8. Linear time series trend line coefficients with 95% confidence intervals for error-adjusted forest area from the forest mask maps, in addition to the proportion of the state represented by the 5 canopy cover classes derived from continuous canopy cover products (1973–2015). Non-significant coefficients shown in gray.

change algorithms such as LandTrendr (Kennedy et al., 2010). Information about changes in canopy cover within a disturbance patch may aid in classifying the agent of change which is of great interest to forest managers. The few studies that have incorporated imagery from the full time span of the archive have tended to focus on quantifying disturbance history to estimate current forest structure (Pflugmacher et al., 2012). Our stack of Landsat imagery was spatially and spectrally harmonized and cover models created for one time step should be applicable to additional years, but further research is needed to assess the accuracy of these models for application in future years.

During the four decades of our forest mask and cover products, we observed a significant increase in overall forest cover across the state of Minnesota, although this trend was not consistent across all cover classes. In a USDA Forest Service resource bulletin based on Forest Inventory and Analysis (FIA) data for Minnesota, live trees were reported to have increased during 1977–2003, as well as an increase of 5.6% in live tree biomass on timberlands within the same time-period (Miles et al., 2007), supporting the results of this study. Within the shorter time-period of 1990–2003, the study reported an overall decrease in Minnesota forest land of 4% (Miles et al., 2007). When we tested the forest area trend within our data for this truncated time-period, the significant increase in forest cover remained as found within the greater 1973–2015 study period. These contrasting results could be due to differences in the definitions of forest cover by FIA compared to this study, as well as a sampling-based vs. our spatially continuous

predictive mapping assessments. In our forest delineation, we take a snap-shot approach where there must be current forest cover to be classified as forest at a specific time period. FIA considers forest land as more of a land cover classification where even recently cleared areas with the absence of current cover will still be considered forest as long as the potential treed area is > 1 acre or at least 36.6 m (120 ft) wide for roadside, riparian, and shelterbelt strips. Updated Minnesota USDA forest resources reports for all years following the 1990–2003 time period found increases in overall forest area (e.g. Miles and Heinzen, 2008; Miles and VanderSchaaf, 2012, 2015), consistent with our findings. Miles et al. (2007) reported that during 1990–2003, there was greater timber growth than removal rates suggesting an increase in growing-stock volume. This may support our observation of the greatest increases occurring within the two higher forest cover classes.

While there were a variety of sources of forest cover loss and gain during our study, one example of forest gain was the encroachment of forest cover in land previously used for agriculture and pasture. We mostly observed this agriculture to forest conversion within the Eastern Broadleaf Forest province, which, along with the Laurentian Mixed Forest province, experienced the greatest gain in overall forest area. Marshlands reverting back to forest cover are an additional potential source of forest gain within Minnesota reported by Miles et al. (2007). This, along with forest regeneration within harvest patches may have contributed to the greatest increases we observed within the Laurentian Mixed Forest province for the two highest cover classes. That said, the

highest cover class ( $\geq 75\%$ ) also appeared to be under represented when histograms of data from the photo interpretation samples were compared to that of the corresponding 2008 predictive map (visual comparison, data not shown). The under representation of the highest cover class could potentially be a bias of photo interpretation methods for underestimating cover values (Greenfield et al., 2009) or an underestimation of higher cover values within the model.

## 5. Conclusions

In this study, we utilized the Landsat archive to create annual spatial records of canopy cover across the state of Minnesota from 1973 to 2015, a valuable resource for a multitude of research and management applications. We demonstrate the utility of the R package LandsatLinkr for expanding the temporal record of comparable Landsat spectral indices through the harmonization of Landsat MSS and OLI imagery to the spatial and/or spectral qualities of TM and ETM+. We found value in the further enhancement of the LandsatLinkr products with the use of the LandTrendr segmentation algorithm for the smoothing of year-to-year spectral variations not associated with forest change dynamics. While the accuracy of our canopy cover models were comparable to previous Landsat studies, and our temporal trends suggest that the forest area in Minnesota is expanding and moving towards more closed canopy conditions, all temporal trends presented here should be observed with caution as our study did not include a multi-year validation procedure. While the statewide and within province trends discussed here are a way to immediately assess our products, we believe the full value of our maps will be extracted when used in conjunction with additional datasets for applications such as evaluating habitat drivers for Minnesota wildlife species of conservation interest, as well as aiding in interpreting and classifying agents of change across the state, a project currently underway. Future studies should also assess the utility of these image stacks for characterizing local scale dynamics within specific disturbance events and associated recovery patterns. Within more specific applications of the mapping products for change, additional validation should be considered if reference data sets are available. Future studies should continue to assess the range of forest attributes able to be modeled and mapped using LandsatLinkr spectral products.

## Acknowledgements

We would like to acknowledge the Environment and Natural Resources Trust Fund, Minnesota Department of Natural Resources, and the Minnesota Forest Resources Council for project funding. We thank Kevin Buffington and Patrick Fekety for R assistance and additional insights which contributed to our study along with Erik Haunreiter and additional members of Warren Cohen's LARSE laboratory at Oregon State University. We appreciate access to the LandTrendr algorithm granted by Robert Kennedy.

## References

Ahmed, O.S., Franklin, S.E., Wulder, M.A., White, J.C., 2015. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the random forest algorithm. *ISPRS J. Photogramm. Remote Sens.* 101, 89–101.

Braaten, J.D., Cohen, W.B., Yang, Z., 2015. Automated cloud and cloud shadow identification in Landsat MSS imagery for temperate ecosystems. *Remote Sens. Environ.* 169, 128–138.

Braaten, J.D., Cohen, W.B., Yang, Z., 2017. LandsatLinkr. Zenodo: <http://dx.doi.org/10.5281/zenodo.807733>. (Last accessed 21 February 2017).

Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.

Brooks, E.B., Wynne, R.H., Thomas, V.A., Blinn, C.E., Coulston, J.W., 2014. On-the-fly massively multi-temporal change detection using statistical quality control charts and Landsat data. *IEEE Trans. Geosci. Remote Sens.* 52 (6), 3316–3332.

Carreiras, J.M., Pereira, J.M., Pereira, J.S., 2006. Estimation of tree canopy cover in evergreen oak woodlands using remote sensing. *For. Ecol. Manag.* 223 (1), 45–53.

Chavez, P.S., 1996. Image-based atmospheric corrections-revisited and improved. *Photogramm. Eng. Remote. Sens.* 62 (9), 1025–1035.

Chavez Jr, P.S., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sens. Environ.* 24 (3), 459–479.

Cohen, W.B., Goward, S.N., 2004. Landsat's role in ecological applications of remote sensing. *Bioscience* 54 (6), 535–545.

Cohen, W.B., Spies, T.A., Fiorella, M., 1995. Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, USA. *Int. J. Remote Sens.* 16 (4), 721–746.

Coulston, J.W., Moisen, G.G., Wilson, B.T., Finco, M.V., Cohen, W.B., Brewer, C.K., 2012. Modeling percent tree canopy cover: a pilot study. *Photogramm. Eng. Remote. Sens.* 78 (7), 715–727.

Crist, E.P., 1985. A TM tasseled cap equivalent transformation for reflectance factor data. *Remote Sens. Environ.* 17 (3), 301–306.

Crist, E.P., Cicone, R.C., 1984. A physically-based transformation of thematic mapper data—the TM tasseled cap. *IEEE Trans. Geosci. Remote Sens.* 22 (3), 256–263.

Crook, D.A., Robertson, A.I., 1999. Relationships between riverine fish and woody debris: implications for lowland rivers. *Mar. Freshw. Res.* 50 (8), 941–953.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J.C., Lawler, J.J., 2007. Random forests for classification in ecology. *Ecology* 88 (11), 2783–2792.

Evans, J.S., Murphy, M.A., 2017. *rfUtilities*. R Package Version 2.1-0. <https://cran.r-project.org/package=rfUtilities>.

Falkowski, M.J., Evans, J.S., Martinuzzi, S., Gessler, P.E., Hudak, A.T., 2009. Characterizing forest succession with lidar data: an evaluation for the Inland Northwest, USA. *Remote Sens. Environ.* 113, 946–956. <http://dx.doi.org/10.1016/j.rse.2009.01.003>.

Frescino, T.S., Moisen, G.G., 2012. In: Comparing alternative tree canopy cover estimates derived from digital aerial photography and field-based assessments. 2010 Joint Meeting of the Forest Inventory and Analysis (FIA) Symposium and the Southern Mensurationists. pp. 237–244.

Greenfield, E.J., Nowak, D.J., Walton, J.T., 2009. Assessment of 2001 NLCD percent tree and impervious cover estimates. *Photogramm. Eng. Remote. Sens.* 75 (11), 1279–1286.

Hadi, Korhonen, L., Hovi, A., Rönholm, P., Rautiainen, M., 2016. The accuracy of large-area forest canopy cover estimation using Landsat in boreal region. *Int. J. Appl. Earth Obs. Geoinf.* 53, 118–127.

Hansen, M.C., Egorov, A., Roy, D.P., Potapov, P., Ju, J., Turubanova, S., Kommareddy, I., Loveland, T.R., 2011. Continuous fields of land cover for the conterminous United States using Landsat data: first results from the Web-Enabled Landsat Data (WELD) project. *Remote Sens. Lett.* 2 (4), 279–288.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342 (6160), 850–853.

Hartanto, H., Prabhu, R., Widayat, A.S., Asdak, C., 2003. Factors affecting runoff and soil erosion: plot-level soil loss monitoring for assessing sustainability of forest management. *For. Ecol. Manag.* 180 (1), 361–374.

Healey, S.P., Yang, Z., Cohen, W.B., Pierce, D.J., 2006. Application of two regression-based methods to estimate the effects of partial harvest on forest structure using Landsat data. *Remote Sens. Environ.* 101 (1), 115–126.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States—representing a decade of land cover change information. *Photogramm. Eng. Remote. Sens.* 81 (5), 345–354.

Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z., Vogelmann, J.E., 2010. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sens. Environ.* 114 (1), 183–198.

Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., Falkowski, M.J., 2008. Nearest neighbor imputation modeling of species-level, plot-scale structural attributes from lidar data. *Remote Sens. Environ.* 112, 2232–2245.

Hughes, M.J., Kaylor, S.D., Hayes, D.J., 2017. Patch-based forest change detection from Landsat time series. *Forests* 8 (5), 166.

Jennings, S.B., Brown, N.D., Sheil, D., 1999. Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. *Forestry* 72 (1), 59–74.

Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., Xian, G., 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sens. Environ.* 132, 159–175.

Jones, K.B., Neale, A.C., Nash, M.S., Van Remortel, R.D., Wickham, J.D., Riitters, K.H., O'Neill, R.V., 2001. Predicting nutrient and sediment loadings to streams from landscape metrics: a multiple watershed study from the United States Mid-Atlantic Region. *Landscape Ecol.* 16 (4), 301–312.

Kennedy, R.E., Cohen, W.B., 2003. Automated designation of tie-points for image-to-image coregistration. *Int. J. Remote Sens.* 24 (17), 3467–3490.

Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—temporal segmentation algorithms. *Remote Sens. Environ.* 114 (12), 2897–2910.

Kennedy, R.E., Yang, Z., Braaten, J., Copass, C., Antonova, N., Jordan, C., Nelson, P., 2015. Attribution of disturbance change agent from Landsat time-series in support of habitat monitoring in the Puget Sound region, USA. *Remote Sens. Environ.* 166, 271–285.

Koy, K., McShea, W.J., Leimgruber, P., Haack, B.N., Aung, M., 2005. Percentage canopy cover—using Landsat imagery to delineate habitat for Myanmar's endangered Eld's deer (*Cervus eldi*). *Anim. Conserv.* 8 (3), 289–296.

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

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

- 18–22.
- Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.K., 2006. A Landsat surface reflectance dataset for North America, 1990–2000. *IEEE Geosci. Remote Sens. Lett.* 3 (1), 68–72.
- Masek, J.G., Goward, S.N., Kennedy, R.E., Cohen, W.B., Moisen, G.G., Schleeweis, K., Huang, C., 2013. United States forest disturbance trends observed using Landsat time series. *Ecosystems* 16 (6), 1087–1104.
- Miles, P.D., Heinzen, D., 2008. Minnesota's forest resources, 2007. In: Res. Note. NRS-24. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA (4 pp.).
- Miles, P.D., VanderSchaaf, C.L., 2012. Minnesota's forest resources, 2012. In: Res. Note NRS-175. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA (4 pp.).
- Miles, P.D., VanderSchaaf, C.L., 2015. Forests of Minnesota, 2014. In: Resource Update FS-44. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA (4 pp.).
- Miles, P.D., Jacobson, K., Brand, G.J., Jepsen, E., Meneguzzo, D., Mielke, M.E., Olson, C., Perry, C.H.H., Piva, R.J., Wilson, B.T., Woodall, C., 2007. Minnesota's forests 1999–2003 (part A). In: USDA Forest Service Resource Bulletin, NRS 12-A.
- Minnesota Department of Natural Resources, 1999. Ecological Classification System. <http://www.dnr.state.mn.us/ecs/index.html> (Last accessed 23 March 2017).
- Minnesota Department of Natural Resources, 2004. Minnesota Digital Elevation Model - 30 Meter Resolution. <https://gisdata.mn.gov/dataset/elev-30m-digital-elevation-model> (Last accessed 5 January 2017).
- Minnesota Pollution Control Agency, 2015. Status and Trends of Wetlands in Minnesota: Vegetation Quality Baseline. Report # wq-bwm-1-09. Minnesota Pollution Control Agency, St. Paul, MN. <http://www.pca.state.mn.us/index.php/view-document.html?gid=23346> (Last accessed 9 February 2017).
- Moisen, G.G., Meyer, M.C., Schroeder, T.A., Liao, X., Schleeweis, K.G., Freeman, E.A., Toney, C., 2016. Shape selection in Landsat time series: a tool for monitoring forest dynamics. *Glob. Chang. Biol.* 22 (10), 3518–3528.
- Moore, R., Spittlehouse, D.L., Story, A., 2005. Riparian microclimate and stream temperature response to forest harvesting: a review. *JAWRA J. Am. Water Resour. Assoc.* 41 (4), 813–834.
- Murphy, M., Evans, J.S., Storfer, A., 2010. Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* 91, 252–261. <http://dx.doi.org/10.1890/08-0879.1>.
- Neigh, C.S., Bolton, D.K., Diabate, M., Williams, J.J., Carvalhais, N., 2014. An automated approach to map the history of forest disturbance from insect mortality and harvest with Landsat time-series data. *Remote Sens.* 6 (4), 2782–2808.
- Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* 129, 122–131.
- Pflugmacher, D., Cohen, W.B., Kennedy, R.E., 2012. Using Landsat-derived disturbance history (1972–2010) to predict current forest structure. *Remote Sens. Environ.* 122, 146–165.
- Pickell, P.D., Hermosilla, T., Frazier, R.J., Coops, N.C., Wulder, M.A., 2016. Forest recovery trends derived from Landsat time series for North American boreal forests. *Int. J. Remote Sens.* 37 (1), 138–149.
- Pierce, A.D., Farris, C.A., Taylor, A.H., 2012. Use of random forests for modeling and mapping forest canopy fuels for fire behavior analysis in Lassen Volcanic National Park, California, USA. *For. Ecol. Manag.* 279, 77–89.
- Potapov, P.V., Turubanova, S.A., Tyukavina, A., Krylov, A.M., McCarty, J.L., Radeloff, V.C., Hansen, M.C., 2015. Eastern Europe's forest cover dynamics from 1985 to 2012 quantified from the full Landsat archive. *Remote Sens. Environ.* 159, 28–43.
- Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., Ohmann, J.L., 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: a comparison of empirical modeling approaches. *Remote Sens. Environ.* 114 (5), 1053–1068.
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199.
- R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria URL. <https://www.R-project.org/>.
- Rampi, L.P., Knight, J.F., Bauer, M., 2016. Minnesota Land Cover Classification and Impervious Surface Area by Landsat and Lidar: 2013 Update. Retrieved from the Data Repository for the University of Minnesota. <http://dx.doi.org/10.13020/D6JP4S>.
- Roy, D.P., Kovalsky, V., Zhang, H.K., Vermote, E.F., Yan, L., Kumar, S.S., Egorov, A., 2016. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sens. Environ.* 185, 57–70.
- Schwab, F.E., Pitt, M.D., 1991. Moose selection of canopy cover types related to operative temperature, forage, and snow depth. *Can. J. Zool.* 69 (12), 3071–3077.
- Sexton, J.O., Song, X.P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.H., Collins, K.M., Channan, S., DiMiceli, C., Townshend, J.R., 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *Int. J. Digital Earth* 6 (5), 427–448.
- Swanson, D.L., Ingold, J.L., Wallace, G.E., 2008. Ruby-crowned Kinglet (*Regulus calendula*). In: Rodewald, P.G. (Ed.), *The Birds of North America*. Cornell Lab of Ornithology, Ithaca Retrieved from the Birds of North America: <https://birdsna.org/Species-Account/bna/species/ruckin> (Last accessed: 2 February, 2017).
- Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* 185, 46–56.
- Vogeler, J.C., Cohen, W.B., 2016. A review of the role of active remote sensing and data fusion for characterizing forest in wildlife habitat models. *Revista de Teledetección* (45), 1–14.
- Vogelmann, J.E., Xian, G., Homer, C., Tolck, B., 2012. Monitoring gradual ecosystem change using Landsat time series analyses: case studies in selected forest and rangeland ecosystems. *Remote Sens. Environ.* 122, 92–105.
- Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., 2008. Free access to Landsat imagery. *Science* 320, 1011.
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* 118, 83–94.
- Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* 144, 152–171.
- Zhu, Z., Woodcock, C.E., Holden, C., Yang, Z., 2015. Generating synthetic Landsat images based on all available Landsat data: predicting Landsat surface reflectance at any given time. *Remote Sens. Environ.* 162, 67–83.