Application of Local Binary Patterns in Digital Images to Estimate Botanical Composition in
Mixed Alfalfa – Grass Fields

Keenan C. McRoberts^a, Brent M. Benson^b, Erika L. Mudrak^c,
David Parsons^d, Debbie J. R. Cherney^e

^aDepartment of Animal Science, Cornell University, 149 Morrison Hall, Ithaca, NY 14853,
USA, kcm45@cornell.edu

^b203 Solutions LLC, 414 Water Street, Unit 1808, Baltimore, MD 21202, USA,
mbenson99@gmail.com

^cCornell Statistical Consulting Unit, Cornell University, B09 Savage Hall, Ithaca, NY 14853,
USA, mudrak@cornell.edu

^dSchool of Land and Food, University of Tasmania, Private Bag 98, Hobart, Tasmania, Australia,
David.Parsons@utas.edu.au

^eDepartment of Animal Science, Cornell University, 329 Morrison Hall, Ithaca, NY 14853,
USA, djc6@cornell.edu

Corresponding Author: Keenan C. McRoberts, kcm45@cornell.edu

Abstract

Botanical composition in mixed stands of alfalfa and grass is a critical parameter in equations estimating harvest fiber concentration for dairy rations. Composition is difficult to estimate by visual observation. Digital image analysis in mixed stands could reduce botanical composition uncertainty and improve spring harvest management decisions. Mixed stands were sampled ($n=168$) in farmers' fields in Tompkins County, New York in May 2011. A digital image was taken of standing samples at 5-Megapixels resolution using a Canon PowerShot A3100IS, and alfalfa and grass height relationships were recorded. After clipping representative samples at 10-cm above ground level, samples were manually separated into alfalfa (*Medicago sativa* L.) and timothy grass (*Phleum pratense* L.), and dried to calculate fractions on a dry matter basis. Uniform rotation invariant local binary patterns (LBP) were extracted from whole images and 64 x 64 pixel tiles, and were used to develop regression equations estimating grass fraction. Tiles were manually classified as alfalfa (0), grass (1) or unclassifiable. An iterative process selected most accurate local binary pattern operator settings. Grass fraction was estimated in three regression model development approaches: (1) using average tile LBP histogram bins from whole images and botanical height relationships, (2) developing a binary tile classification model from tile LBP histogram bins, and using tile model-predicted grass probability averaged for tiles in whole images (grass coverage estimate) and botanical height relationships as inputs in whole image models, and (3) using LBP histogram bins extracted directly from whole images (1024 by 1024 pixel square) and height relationships. Predictive accuracy in whole image models using tile LBP histogram averages was highest for models generated from LBP tile histogram bin means (R^2_{pred} up to 0.847), followed closely by combined tile models and whole image models (R^2_{pred} up to 0.807), with pairwise correlations between tile model-generated grass coverage

estimates and sample grass fraction up to 0.895. Local binary patterns are effective in differentiating alfalfa and grass under field conditions, because the method is robust to changes in color and illumination. Furthermore, key LBP histogram bins (e.g., symmetric edges) strongly differentiate alfalfa and grass in tiles. The LBP method is promising based on this study, but further evaluation under diverse field conditions, including different cameras and grass species, is necessary to assess usefulness.

Keywords: Alfalfa-grass forage; spring harvest; forage nutritive value; stand composition; pattern recognition

1. Introduction

Timing of spring forage harvests in the northeast U.S. is critical to ensure high quality forage for dairy cattle production throughout the growing season. Spring forage harvest timing can be predicted based on neutral detergent fiber (NDF) concentration (Parsons *et al.*, 2006b), and target NDF depends on the class of livestock being fed when forages are the principal source of fiber in rations. Target NDF at harvest for dairy cattle is approximately 50% of dry matter for pure grass stands for silage and 40% of dry matter for alfalfa (Cherney *et al.*, 2006). Other forage quality parameters, such as protein and fiber digestibility, are important for ration balancing, but they are not as useful for harvest date targets.

Forage sampling of farmers' fields in New York has produced simple equations for the prediction of nutritive value and harvest timing for pure stands of alfalfa (*Medicago sativa* L.), grass (e.g., *Phleum pratense* L., *Phalaris arundinacea* L., *Dactylis glomerata* L., and

Festucaarundinacea Schreb.), and mixed stands of alfalfa and grass (Parsons *et al.*, 2006a; Parsons *et al.*, 2009; Parsons *et al.*, 2012). Forage stands can also be tested for dry matter loss and changes in NDF at variable stubble height, another important management factor (Parsons *et al.*, 2009; Parsons *et al.*, 2012). These equations have proven useful over a range of conditions and years.

Alfalfa is sown with a perennial grass companion on over 80% of the alfalfa acreage in New York (Cherney *et al.*, 2006), and accurate mixed stand equations are important for effective harvest management. Required inputs for mixed stand equations include alfalfa maximum height, stand composition (grass fraction in the stand), and targeted harvest NDF concentration (Parsons *et al.*, 2006a). Botanical or stand composition is a critical parameter in the equation, and is difficult to accurately predict by visual observation (McRoberts *et al.*, 2012a). Difficulty in estimating stand composition was reported by extension educators at the Cornell Field Crop Extension Educators' Retreat in April 2011 as the principle problem limiting the utility of mixed stand equations for farmers in the northeast U.S. Overestimating grass by just 20% can result in late harvests by five or more days, potentially leading to NDF at harvest > 5% past target levels. Underestimating grass fraction results in early harvests, lower spring forage yields and higher unit cost per ton of forage dry matter produced. Thus, misestimating composition represents critical potential nutritive and economic losses for dairy farms. Reducing uncertainty in the stand composition estimate could improve the quality and timing of spring forage harvests.

Manual and semi-automated approaches have been attempted to estimate botanical composition in legume – grass stands. Visual estimation methods have been used historically as the principal

method. Rayburn and Green (2014) developed a visual reference guide for mixed stands of clover and grass to help calibrate the eye for human field estimation. Rayburn (2014) tested a manual point count method by iteratively superimposing a randomly placed virtual point count grid on mixed stand images, counting the number of points touching grass, legumes, forbs, bare ground, and dark shadows, and quantifying the points. Point counts were then regressed on actual botanical composition. The method was strong when using at least 100 points per image (R^2 ranging from 0.45 to 0.98). However, it is time consuming and would require manual image processing by users as well as equation calibration with different species combinations, sampling seasons, and cameras. Height relationships of species sampled were not used in their method development.

In a pot experiment, Himstedt *et al.* (2009) discovered high correlations between legume coverage and actual legume dry matter fraction in mixed stands ($R^2=0.89$ across three legume species for two sward ages; $R^2=0.96$ for alfalfa). Actual legume coverage was calculated by manually circling areas covered by legumes and dividing by total area. They also estimated coverage by processing grayscale images using morphological operators, including a multi-step erosion process, followed by a dilation process with the same number of steps. Erosion effectively removed small objects such as grass leaves, while dilation blew the remaining objects (inner portion of alfalfa leaves) back up to their approximate original size. Grayscale thresholding was used to separate legume leaves from everything else in the image, and to estimate coverage as *legume leaves/total area*. The relationship between actual coverage and estimated coverage was strong for samples with higher percent coverage ($R^2=0.88$ overall,

$R^2=0.84$ for alfalfa). However, sample size was small (64 images) and lighting and growth conditions were controlled.

Himstedt *et al.* (2010) furthered equation development using logit-transformed legume coverage in statistical model development, and selected a multivariate model predicting legume dry matter contribution with effects including logit-transformed legume coverage, total biomass in the sample, and their interaction. Equation testing on field samples yielded a strong relationship with legume dry matter contribution for clover – grass mixes ($R^2=0.98$). Practicality of such an equation for field use is questionable without further investment in technology such as field spectroscopy given the need for the total dry matter biomass variable. Single cameras were used in all tests. The technique would require further testing under variable field conditions to evaluate potential field use.

In a field study, Post *et al.* (2007) related plant canopy spectral reflectance (wavelengths 680 nm and 705 nm in the second derivative spectra) with alfalfa fraction in a mixed stand ($R^2=0.6$ to 0.7 , $n=95$). The approach is promising for further investigation, and potentially for post-calibration field use (Post *et al.*, 2007). Others have implemented variations on canopy spectral reflectance to predict stand composition with promising results (Kawamura *et al.*, 2011). However, spectral technology may not be accessible for end users.

More sophisticated image processing methods such as artificial intelligence (Aitkenhead *et al.*, 2003) and texture classification (Sabeenian and Palanisamy, 2010) have been tested to discriminate between vegetation types (e.g., crops and weeds). Methods that permit crop – weed

discrimination in real time, combined with robotic herbicide application and cultivation systems, play an important role in precision agriculture. Local binary patterns (LBP), commonly known for their use in facial recognition (Ahonen *et al.*, 2006), provide a powerful, robust, computationally efficient method for texture classification in image analysis (Ojala *et al.*, 2002). Under field conditions and with different image acquisition devices, illumination variability and color variability is high. The application of rotation invariant uniform LBPs to grayscale images could be useful for estimating alfalfa-grass stand composition under field conditions, because it is robust to changes in illumination and color (Ojala *et al.*, 2002).

The study objective was to develop a practical, farmer-accessible method that can be applied to estimate stand composition (i.e., grass and alfalfa dry matter fractions in binary mixes) under variable field conditions. The dataset was tested using multiple approaches including geometric pattern matching, color separation, blob detection, and tile extraction with fast Fourier transformation (combined with naïve Bayes classifier artificial intelligence and trained and untrained support vector machines) with unsatisfactory results (McRoberts *et al.*, 2012a). In this paper a method is developed that combines digital image analysis using local binary patterns with statistical modeling to estimate alfalfa-grass stand composition. The sampling process, local binary pattern method, and its implementation with several processing approaches to estimate stand composition are described.

2. Methods

2.1 Sampling

Mixed stands with different representative proportions of alfalfa and timothy grass (*Phleum pratense* L.) were identified in farmers' fields in Tompkins County, New York (42° 36' N, 76° 30' W) in spring 2011 ($n=168$, including 3 pure grass and 5 pure alfalfa samples). Representative samples were selected and delineated using a round hoop (66-cm diameter), which was rested on the vegetative canopy. A digital image (JPEG format) was taken at 5-Megapixels resolution using an affordable, farmer-accessible, point-and-shoot digital camera (Canon PowerShot A3100IS) with automatic settings. A small bubble level was mounted on the camera LCD screen to ensure consistent image orientation parallel to the ground. Hoop edges were located as close to the image edges as possible and photos were taken freehand at approximately 1 m above the grass canopy under natural field illumination conditions. Shadows were avoided in the image area. In each sample, alfalfa maximum height (maximum height of terminal bud), grass maximum height (maximum height of tallest extended grass blade), and grass canopy height were measured using a meter stick. Lighting conditions associated with each image were noted (e.g., full light or shade).

A 10-cm high quadrat was inserted under the hoop after photo acquisition to establish cutting height, and the hoop was lowered to ground level. Forage within each hoop (sample) was harvested at 10-cm above ground level, the typical forage cutting height by farmers in New York State, by cutting at the height of the quadrat using battery-operated grass clippers. Harvested forage from each sample was manually separated into alfalfa and grass fractions. Plant species other than alfalfa and timothy grass were either absent or present only in trace quantities relative to alfalfa and timothy grass, and were discarded. Samples were dried to stable weight at 60°C in

a forced-air oven. Samples were weighed directly after removal from the oven to determine alfalfa and grass dry fractions in each sample, calculated as $grass\ fraction = grass\ dry\ weight / (alfalfa\ dry\ weight + grass\ dry\ weight)$. Mean grass fraction for samples containing both alfalfa and grass was 0.413 with standard deviation of 0.148 (n=160).

2.2 Image processing, tile extraction and classification

Images were cropped to remove the hoop and non-sampled area outside the hoop using the GNU Image Manipulation System (GIMP). The resulting ellipse from each original image was converted to grayscale in Python using the derivation of luminance signal equation (Radiocommunication Sector of ITU, 2002) (Figure 1). Grayscale conversion and subsequent image processing were undertaken using the Scikit-Image library for Python (Van der Walt *et al.*, 2014).

Figure 1 Inner hoop area in mixed stand (alfalfa – grass) images was manually extracted from original image (left) in GIMP. The resulting color ellipse (center) was converted to grayscale (right) in Python.

Tile extraction (Polder *et al.*, 2007) divided the grayscale image into 64 x 64 pixel tiles for local processing (Figure 2). Larger and smaller tile sizes meeting the LBP operator criteria (power of

two) were attempted, but were not practical for tile classification (e.g., too small to determine species in tile; both species simultaneously in tile due to large size). Images contained 462 tiles per image with standard deviation of 47.7 tiles per image. Tiles were manually classified as either mostly grass (1) or mostly alfalfa (0). Tiles with an equal amount of alfalfa and grass or material other than these species (e.g., bare ground, other species, unclear tile image areas) were considered unclassifiable, and were treated as missing data for classification purposes. Most tiles (78.4%) were successfully classified as alfalfa or grass. Out of 77,667 total tiles, 15,185 were classified as grass and 45,709 as alfalfa.

Figure 2 Tiles (64 x 64 pixels, center) were cropped from original images (left, with tile grid) for individual analysis. The 7 x 7 pixel inset in the tile (center) is extracted (right) to demonstrate a symmetric edge pattern in the focal pixel (boxed) for radius (R) = 2, number of sample points (P) = 8.

2.3 Local binary pattern determination in tiles

Local binary patterns (Ojala *et al.*, 2002) were then extracted using the Scikit-Image library for Python (Van der Walt *et al.*, 2014). Rotation invariant uniform LBPs effectively capture key information about stand composition in the image while minimizing sources of non-composition-related variation arising from lighting and color (Ojala *et al.*, 2002).

Local binary patterns were determined for each individual pixel in each tile by comparing pixel grayscale value with a number of points (P) located equidistantly from each other on an arc of radius R from the central pixel (Ojala *et al.*, 2002). For pixels on or near tile edges, points lying outside the tile area were assigned a grayscale value of 0. Each point was assigned a 1 if its grayscale value exceeded or was equal to the focal pixel and 0 if it was less than the central pixel (Figure 2). For example, with $P=8$, the resulting binary assignment for the 8 points generated 2^P possible patterns. Patterns were limited to uniform rotation invariant local binary patterns with ≤ 2 bitwise transitions between 0 and 1, which accounted for the majority of patterns (Ojala *et al.*, 2002). There were $P - 1$ uniform rotation invariant patterns with 2 transitions and two patterns with 0 transitions (all 1, all 0). Patterns with greater than 2 transitions were pooled into a final non-uniform pattern. In the case of $P=8$, there were 7 uniform rotation invariant patterns with 2 transitions, 2 patterns with 0 transitions, and 1 non-uniform pattern with > 2 transitions, for a total of 10 uniform rotation invariant pattern values represented (Figure 3). Each central pixel was assigned a single value between 0 and 9 that represented its unique pattern. Continuing with the $P=8$ example, pattern 0 (all 0s) detected light spots (focal pixel surrounded by darker pixels) while pattern 8 (all 1s) detected dark spots (focal pixel surrounded by lighter pixels). Pattern 4 was a symmetric edge detector. Patterns between 0 and 4 and 4 and 8 were edges of varying curvature. The same interpretation applied to LBP pixel values assigned for each distinct P . Additional detail on the implementation of uniform rotation invariant LBPs is available in Ojala *et al.* (2002).

Tiles contained 4096 pixels (64 x 64 pixels) and each pixel was assigned an LBP pattern value between 0 and 9 for the $P=8$ example. Thus, a histogram containing 10 frequencies described

each tile. Each frequency was calculated by counting the total number pixels containing each pattern value for each of j bins, and reporting $pattern\ count\ in\ bin / 4096$ as the frequency. Each frequency represented a fraction of total patterns expressed in the tile, and for each tile, $\sum LBP\ frequencies = 1$.

Figure 3 A) Example of uniform rotation invariant patterns for 8 points ($P=8$). Bin 0 represents lights spots. Bin 8 (P) represents dark spots. Bins 1 to 3 ($P/2 - 1$) and bins 5-8 ($P/2 + 1$ to $P-1$) are edges of varying curvature and contain 8 possible rotations; three are shown here for each bin. Bin 9 ($P+1$) represents all non-uniform patterns, three of which are shown here. B) LBP histogram for a hypothetical tile for an analysis with $P=8$. Each frequency is calculated by counting the total number of pixels containing each pattern type and dividing by the number of total pixels in the tile.

2.4 Local binary pattern operator selection

An iterative optimization process (written in Python) selected R and P using between 500 and 5000 randomly selected classified tiles for training (50% alfalfa and 50% grass) and 2000 different randomly selected classified tiles (50% alfalfa and 50% grass) for testing. Testing tiles were static for each iteration, which tested multiple R and P combinations with different training sets. Different P and R combinations were evaluated for each training set, with R allowed to vary between 1 and 4 and P between 2 and $R \times 8$ (practical upper limit for P) in increments of 2. For each testing tile, KullbackLeibler (KL) divergence was computed between the testing tile LBP histogram and each training tile's LBP histogram. Testing tiles were estimated according to the lesser of $\sum(KL \text{ divergence for alfalfa tiles})$ and $\sum(KL \text{ divergence for grass tiles})$ as alfalfa or grass, respectively. The R , P selection process was repeated 59 times to assess stability of selected values and to select R , P candidates for use in statistical model development (below). The R , P combination that most accurately estimated human tile classifications (according to *test tiles correctly classified / total test tiles*) was selected from each iteration. Eight R , P sets emerged from this process (Appendix Table A.1), and were used in statistical model development (below).

2.5 Statistical methods to estimate stand composition from local binary pattern data

Three statistical modeling approaches were implemented to estimate stand composition (grass fraction response) from LBP data. Statistical modeling was undertaken in SAS for Windows 9.3 (SAS Institute Inc., 2011). Pure alfalfa and pure grass samples were excluded from whole image model development.

2.5.1 Whole image model using tile local binary pattern bin averages

The LBP tile data were reduced to a single histogram for each original hoop image by calculating the average value of each LBP frequency for each R, P combination. Candidate predictors in models estimating grass fraction (GFRAC) included average LBP histogram frequencies from each R, P set (bin_0 through bin_{p+1}), grass canopy height (GCPY), grass maximum height (GMAX) and alfalfa maximum height (AMAX). Regression model selection for GFRAC response was undertaken in PROC GLMSELECT using stepwise selection (SAS Institute Inc., 2011) (model selection based on *PRESS* statistic) with all candidate predictors, according to the model $y = X\beta + \varepsilon$. For each image i ,

$$GFRAC_{rpi} = \beta_0 + \beta_{h_0}bin_0 + \beta_{h_1}bin_1 + \cdots + \beta_{h_p}bin_p + \beta_{h_{p+1}}bin_{p+1} + \beta_2GCPY + \beta_3GMAX + \beta_4AMAX$$

Models selected by PROC GLMSELECT were tested for multicollinearity in PROC REG, and further reduced to ensure variance inflation factors near or below 10. Residuals and Q-Q plots were visually examined for normality. Model fit was evaluated based on lowest prediction sum of squares (*PRESS*) statistic and root mean squared error (RMSE), and highest R^2 and R^2_{pred} . *PRESS* is useful to assess model fit in cross-validation, and is calculated as $\sum_{i=1}^n (y_i - f_i)^2$, where n is number of observations, y_i is the actual value of the i th data point, and f_i is the predicted value for y_i from a model trained using all data except the i th observation (Meyers, 1990). The *PRESS* statistic is the sum of squares of prediction residuals for observations not used to fit the model. The R^2_{pred} is always lower than R^2 , because it accounts for model prediction capability according to: $R^2_{pred} = 1 - \frac{PRESS}{\sum_{i=1}^n (y_i - \bar{y})^2}$ (Meyers, 1990). The dataset was not

split for training and testing in this first-stage assessment of LBP methods for estimation of mixed stand composition. Top R , P sets were tested for performance differences in GFRAC estimation based on lighting conditions (full light or shade) by adding a binary covariate.

2.5.2 Tile model combined with whole image model

The two-phase model selection process first involved developing tile models to estimate tile probability of grass from local binary pattern bins. Second, average probability of grass, considered a grass coverage estimate for whole hoop images, was calculated and used as the key predictor in whole image model selection. Whole image models estimated grass fraction from the LBP-generated grass coverage estimate (from tile model), alfalfa maximum height and grass canopy height.

In the tile model, a binary response was defined by classified tile values (alfalfa = 0, grass = 1). Unclassified tiles were treated as missing data and ignored in analysis. Tile LBP models were selected using PROC GLIMMIX (SAS Institute Inc., 2011), with binomial distribution and logit link function, using LBP frequencies extracted from each R , P set (Appendix Table A.1), according to the model $\log\left(\frac{p_{grass}}{1-p_{grass}}\right) = X\beta + Zb + \varepsilon$, where p_{grass} is the probability a tile is grass, X is a matrix of fixed-effect predictors, β is a vector of regression coefficients for fixed effects, Z is a random effects matrix, b is a vector of random effects, and ε is a vector of residuals. Fixed effects included LBP frequencies for each tile (observation). Photo filename was included as a random effect to account for non-independence of tiles within photos. Tile model evaluation criteria included *maximum likelihood*, calculated as $\sum(classification * \log(\widehat{p_{grass}}) + (1 - classification) * \log(1 - \widehat{p_{grass}}))$, where *classification* is human

classified value for each tile (alfalfa (0) or grass (1)) and $\widehat{p_{grass}}$ is model-estimated probability of grass for each tile. This likelihood approach was chosen to assess model fit rather than classification error rate, because it does not require a specific cutoff (e.g., 0.5), which is important in this dataset due to mean grass fraction < 0.5 (and fewer tiles classified as grass than alfalfa), resulting in p_{grass} weighted toward alfalfa. This likelihood approach favors models predicting tile p_{grass} closer to 0 and 1 over less certain models (e.g., $p_{grass} = 0.5$). Model-predicted p_{grass} data were generated for all tiles in each R , P combination, and mean of tile-predicted p_{grass} for whole hoop images was used as a key predictor in whole image model selection. Performance of this parameter in predicting actual stand composition was also considered in selecting most effective tile models (and R , P combinations). Three approaches to tile model development were employed: (1) selection from bins representing light spots, symmetric edges, dark spots and non-uniform patterns, (2) selection from all significant histogram bins, and (3) selection from bins suggested by recursive partitioning in JMP 11.2.0 (SAS Institute Inc., 2013). Intraclass correlation was calculated to determine photo effect size in a null tile model (no fixed effects) according to Ene *et al.* (2015).

Pairwise correlations were evaluated between whole image GFRAC and LBP coverage estimates predicted by selected tile models for 8 R , P combinations (Appendix Table A.1). Whole image model selection was undertaken in PROC GLM (SAS Institute Inc., 2011) with GFRAC continuous response. Fixed effects included LBP coverage estimate (mean p_{grass}), AMAX and GCPY. Whole image models selected using the tile model-generated p_{grass} estimate did not include interactions or quadratic effects for ease of cross-model comparison. Model fit was

evaluated using the criteria outlined in Section 2.5.1. Final tile model selection was based on performance of whole image models containing tile overall p_{grass} estimate.

2.5.3 Whole image model using local binary pattern bins in whole images

Further image extraction was required to process whole images (in lieu of tiles), because processing image units must be a power of 2 pixels. The closest match was 2^{10} for whole images, thus a 1024 by 1024 pixel square was extracted from inside hoop images (approximately 60% of total hoop area) (Figure 4). The LBP histograms were extracted from whole images (rather than individual tiles) as described in section 2.3 by manually testing R, P combinations between 1, 4 and 7, 56. The GFRAC models were selected using automated model selection procedures for more than 50 R, P sets from LBP bins, AMAX and GCPY, with final models adjusted to eliminate multicollinearity. Statistical model selection was undertaken as in Section 2.5.1.

Figure 4 A 1024 x 1024 pixel square was extracted from original images for application of the local binary pattern operator on whole hoop images.

3. Results

3.1 Relationship between classified tiles and grass fraction

Grass coverage estimates generated by human tile classification underestimated actual GFRAC (as measured by dry biomass) in a direct relationship, particularly for low- and mid-range grass fractions (Figure 5A). Samples near GFRAC values of 0.5 contained higher proportion of unclassified tiles in original hoop images (Figure 5B).

Figure 5 Quadratic relationship between actual grass fraction and fraction estimated by tile classification (A, $n=160$), described by $y = -0.7266x^2 + 1.302x + 0.159$ ($r^2 = 0.872$). Dashed line is 1:1 relationship. Relationship between actual grass fraction and fraction unclassified tiles in each original image (B, $n=160$). Dashed line indicates actual grass fraction of 0.5.

3.2 Local binary patterns for alfalfa and grass differentiation

Symmetric edge detection bins were most significant and most effective at partitioning tiles into alfalfa and grass (Figure 6, Appendix Table A.2). Average symmetric edge proportion of total patterns in each tile was significantly higher for grass tiles than alfalfa tiles. Other patterns also exhibited mean differences, but average proportion of total patterns differed by a smaller margin for other bins. For example, average proportion of total patterns expressed for light spots, dark

spots, and non-uniform patterns was higher for alfalfa tiles than grass tiles. Consequently, uniform patterns were expressed in greater average proportion for grass tiles, which was driven by the symmetric edge bin and others near it (Appendix Table A.2). Bins neighboring the symmetric edge detector generally had higher grass fractions than alfalfa fractions, although more bins adhered to this trend above $P/2$ than below it. Conversely, bins neighboring light spots ($P=0$) and dark spots ($P=P$) tended to be expressed in higher proportion for alfalfa tiles than grass tiles.

Figure 6 Average percentage of total local binary patterns in dataset for grass tiles (gray line, $n=15,185$) and alfalfa tiles (black line, $n=45,709$) for $R=3$, $P=20$. Bin 0 represents lights spots, bins 1 to 9 and 11 to 19 edges of varying curvature, bin 10 symmetric edges, and bin 20 dark spots. Bin 21 (not shown) represents non-uniform patterns, and contained 42.1% of alfalfa tile patterns and 40.4% of grass tile patterns. Relationship around the symmetric edge detector bin and interpretation are similar for other R , P sets.

Bin 2 exhibited greater separation of alfalfa and grass than neighboring bins. Bins representing edges of varying curvature generally had low numerical mean differences between bin 0 and bin $P/2-2$, and bin $P-1$ and bin $P/2+3$ (alfalfa > grass), and higher differences near the symmetric edge detector (grass > alfalfa). The lowest R, P combination (1,8) differed slightly, due to smaller number of possible patterns, and was less effective in separating alfalfa and grass than other R, P sets tested.

3.3 Field measurements

Grass fraction was more highly correlated with grass canopy height ($r=0.38$) than grass maximum height ($r=0.21$) and alfalfa maximum height ($r=0.06$). Grass maximum height was not included as a candidate covariate in grass fraction models, because it usually dropped from models that included grass canopy height.

3.4 Model development on tile bin averages for whole images

Grass fraction models were selected individually from tile LBP means for R, P combinations, as well as GCPY and AMAX for original hoop images (Appendix Table A.3). Coefficient of determination (R^2) ranged from 0.50 to over 0.86, R^2_{pred} from 0.47 to 0.85 and calibration RMSE from 0.10 to 0.057. Top R, P combinations based on minimum PRESS were 4,32; 3,24; and 3,20. Top models (Table 1) generally had higher R, P combinations and were those selected from automated model selection with post-selection manual adjustments to prevent multicollinearity (rather than just from bins representing light spots, dark spots, symmetric edges, and non-

uniform patterns). All models contained the symmetric edge detector bin. Bin 0 (light spots) was frequently selected, except in top tier automated selection models.

Table 1 Five top models estimating grass fraction developed using tile local binary pattern bin means, grass canopy height and alfalfa maximum height for whole images. Each row represents a single model ($n=160$). Models were evaluated based on PRESS statistic, R^2 , R^2_{pred} , and $RMSE$, and sorted in ascending order by PRESS.

R	P	R^2	$PRESS$	R^2_{pred}	$RMSE$	Model effects ^b
4	32	0.859	0.534	0.847	0.0566	bin2 bin16 bin19 bin27 AMAX
3	24	0.855	0.560	0.840	0.0577	bin10 bin12 bin19 bin22 GCPY AMAX
3	20	0.851	0.563	0.839	0.0582	^a bin0 bin10 bin21 GCPY AMAX
3	20	0.852	0.571	0.837	0.0582	bin3 bin7 bin10 bin18 GCPY AMAX
2	12	0.849	0.575	0.836	0.0585	bin5 bin6 bin11 GCPY AMAX

^a Model selected only from bins representing symmetric edges, light spots, dark spots, and non-uniform patterns.

^b Model effect representing symmetric edges is in bold for each model.

3.5 Combined tile model and whole image model development

Maximum likelihood for tile models ranged from -31,000 to -19,000, and tended to favor less parsimonious models within each R,P , and higher R,P combinations (Appendix Table A.4). The top tile models based on maximum likelihood were 4,32; 4,28; and 4,24 (Table 2). Maximum likelihood-selected tile models differed from models selected based on performance (low $PRESS$) in whole image models (Table 2). Coefficient of determination in whole image models ranged from 0.2 to over 0.82, R^2_{pred} from 0.2 to 0.81 and calibration $RMSE$ from 0.13 to 0.064 (Appendix Table A.4). Top R, P based on $PRESS$ were 3,24; 3,20; 4,32; and 4,28 (Table 2). Top models were generally selected from higher R , and P in the 20s. Models were selected three ways for each R,P , and models fueled by bins suggested in recursive partitioning regression generally performed better than those constructed using all significant bins or built using symmetric edge,

light spot, dark spot, and non-uniform pattern bins. All tile models contained the symmetric edge detector bin. Models selected from bins identified with recursive partitioning also frequently included bins neighboring symmetric edge detector bin and bin 2. The binary lighting effect was only significant in combined whole image model selection in models with all significant bins, and was not included in models. Variation among photos accounted for 40% of the variance in a null tile model (no fixed effects). The remaining 60% was due to variation among tiles.

Table 2 Six top models in combined tile model and whole image model selection.

<i>R</i>	<i>P</i>	Tile model (<i>n</i> =60,894 classified tiles) ^d					GFRAC corr (<i>r</i>)	Whole image model ^d (<i>n</i> =160 images)				
		<i>ML</i>	Model effects ^e					<i>R</i> ²	<i>PRESS</i>	<i>R</i> ² _{<i>pred</i>}	<i>RMSE</i>	
3	24	-20893	^c bin12	bin25	bin11	bin10	0.895	0.817	0.676	0.807	0.0641	
3	20	-20374	^c bin10	bin9	bin21	bin2	0.889	0.810	0.700	0.800	0.0652	
3	20	-19467	^b bin0	bin1	bin2	bin7	bin8	bin9	bin10	bin11		
			bin12	bin13	bin14	bin15	bin17	bin18	bin20			
4	32	-19457	^c bin16	bin15	bin14	bin2	bin3	0.884	0.800	0.737	0.789	0.0669
4	28	-19566	^c bin14	bin13	bin12	bin2	0.885	0.800	0.738	0.789	0.0669	
4	24	-18800	^b bin0	bin1	bin2	bin7	bin8	bin9	bin10	bin11		
			bin12	bin13	bin14	bin15	bin17	bin18	bin19			
			bin20	bin22	bin23	bin24	0.882	0.794	0.760	0.783	0.0679	

^a Model selected from symmetric edge, light spot, dark spot, and non-uniform pattern bins.

^b All significant LBP bins were selected.

^c Model selected from bins identified using recursive partitioning (SAS Institute Inc., 2013).

^d Tile model selection criteria included maximum likelihood (ML), performance of tile model-generated grass coverage estimates in whole image models with alfalfa maximum height and grass canopy height covariates (based on *R*², *PRESS*, *R*²_{pred}, and *RMSE*), and bivariate correlation of tile model-generated grass coverage estimates with grass fraction in dry biomass in whole hoop images (GFRAC corr (*r*)).

^e Symmetric edge effect is bold for each model.

3.6 Model development on whole hoop image local binary pattern bins

Whole image models selected from LBP bins extracted from whole hoop images did not perform as well as other approaches tested above. Fourteen top models were selected based on $PRESS < 1.10$, $RMSE < 0.08$, and $R^2_{pred} > 0.69$ (Appendix Table A.5). Additional models selected using only key bins (e.g., light spots, dark spots, symmetric edges, and non-uniform patterns) did not outperform those in Appendix Table A.5.

4. Discussion

Clear differentiation of local binary patterns in alfalfa and grass tiles is the principal reason that this method is effective in estimating grass coverage and grass fraction of dry matter from digital images representing mixed stand samples. Curved alfalfa leaves in tiles classified as alfalfa contain fewer symmetric edges and more non-uniform patterns relative to tiles with predominantly grass (more long, straight or slightly curving leaf edges). Thus, the symmetric edge detector bin is most effective in differentiating grass and alfalfa in the dataset used in this study. Similarly, lightspots, darkspots, and edges of higher curvature are present in higher fractions for alfalfa tiles, while edges closer to symmetric tend to be more prevalent in grass tiles regardless of R , P combination. Those farther from the symmetric edge detector are expressed more in alfalfa tiles, especially above the symmetric edge detection bin. Bin 2 also emerged frequently as a significant effect in tile models, and exhibited greater alfalfa – grass differentiation capacity than neighboring bins. Bin 2 represents slight curves close to light spots, which are more prevalent in alfalfa than grass.

Manual tile classification tended to underestimate actual grass fraction in a direct quadratic relationship, particularly for low- and mid-range grass fractions (Figure 5). This could be attributed to broader alfalfa leaves obscuring grass leaves in the area viewed, and the lack of correction for alfalfa and grass height relationships. This is consistent with the linear relationship between actual grass fraction and researcher visual estimation of grass fraction for nearly 600 samples in 2004 ($y = 0.22 + 0.69x$, $R^2=0.43$, $RMSE= 0.147$) (McRoberts *et al.*, 2012a), which underestimated grass fraction for low- and mid-range fractions. Furthermore, the positive relationship in the present dataset between actual grass fraction and proportion unclassified tiles up to $GFRAC = 0.5$ suggests that human classification difficulty increases when stands are highly mixed. Tiles would be more likely to contain both alfalfa and grass, which precludes binary classification in these cases.

Grass fraction models selected from LBP tile bin means for whole hoop images are most promising for $GFRAC$ estimation, followed closely by combined tile model selection and whole hoop model selection. Consistent with most accurate R,P sets in the automated R, P selection procedure (Appendix Table A.1), leading R,P candidates in both model selection approaches are 3,24; 3,20; and 4,32, although other models from high R (3 or 4) and high P (20 to 24 for $R=3$ and 28 to 32 for $R=4$) performed nearly as well based on low $PRESS$ and other statistical selection factors considered. Leading models contain the symmetric edge detector, at least one neighboring bin, one bin near light spots, and often a bin between symmetric edges and dark spots.

Local binary patterns extracted directly from whole images (without tile extraction) were the least effective in estimating grass fraction in mixed stands in this dataset. This can be partly attributed to the image subsample required by the LBP operator (representing about 60% of whole hoop image area). Similar bins emerged for whole images as in tile models, and often included bin 0 (light spots), bin 2, symmetric edge detector, and bin P (dark spots). However, it is unlikely that whole image models would perform as well as tile models due to the scale of data analyzed. Furthermore, R , P selection is arbitrary and manual using this method due to lack of binary classified response feasibility for whole hoop images, which precludes automated R , P selection similar to the approach used with classified tiles.

Similar to facial recognition, smaller image sections in mixed stands may contain relationships that permit more effective pattern recognition upon aggregation. Tile size (64 x 64 pixels) was chosen so that tiles could be effectively classified by humans. Larger and smaller regions were tested, but presented complications due to areas too small to classify or large areas that usually contained both species. Analysis of local areas (tiles) in this study was more effective than whole image analysis with LBPs. This is consistent with Ahonen *et al.* (2004) facial recognition tests, which indicated that subdivision of images into smaller rectangular regions improved performance.

Images represented a range in illumination and color, but lighting did not contribute to most GFRAC estimation models. Local binary patterns are more robust than other methods to changes in illumination and color, such as those experienced under field conditions, due to their gray scale invariance (Ojala *et al.*, 2002). This is an important reason that the LBP approach is

favorable relative to other potential automated approaches for estimating mixed stand composition under field conditions. Promising alternative approaches requiring further field testing include erosion and dilation procedures (Himstedt *et al.*, 2009, 2010, 2012) and canopy reflectance (Post *et al.*, 2007). Aside from local binary patterns, erosion and dilation procedures may have the most potential for field use, although effectiveness would decrease for grass species with wider leaves in mixed stands, because they would be less likely to drop out completely during the erosion process. Preliminary testing was undertaken with other less effective approaches including tile extraction and fast Fourier transform (Polder *et al.*, 2007). These approaches had little alfalfa – grass differentiation capacity in the absence of species height relationships (McRoberts *et al.*, 2012a; McRoberts *et al.*, 2012b). More complex LBP processing approaches such as support vector machines could be considered.

Image resolution would impact the LBP operator. Five-megapixel images used in this study contain approximately 23.5 pixels per cm. Higher-resolution images taken with the same sampling protocol would contain more pixels per cm and would have to be reduced to the resolution used in this study for compatibility. This is beyond the scope of the current dataset, but will be important to test for future use with higher resolution devices.

5. Conclusions

Accurate grass fraction estimation with LBPs in digital images is promising based on this study. The current dataset contained images from a single grass species, single sampling year, and single camera. Development of GFRAC prediction equations will require training and testing on a more diverse dataset. If successful, resulting equations will contribute to improved alfalfa-grass

harvest management and economic outcomes for dairy farms in the northeast U.S. by reducing uncertainty in existing NDF prediction equations. Equations could be built into a web program and smartphone app requiring users to upload representative digital images from mixed alfalfa-grass fields and to input species height measurements. The program would return estimates of botanical composition (GFRAC), NDF concentration, and optimal harvest date (based on target NDF and estimated daily change in NDF given weather conditions until harvest). The GFRAC estimation approach developed here uses technology commonly available to farmers and consultants (digital or smartphone camera), permitting widespread use without investment in additional infrastructure. Such a tool would allow farmers and consultants to prioritize the order of harvest of alfalfa-grass fields to increase the odds of obtaining optimal forage NDF for lactating dairy cow diets.

6. Acknowledgments

The authors would like to thank Dr. J.H. Cherney for assistance with study design, planning and implementation and Brian Boerman at Agricultural Consulting Services for assistance in identification of farmer's fields for sampling. Thank you to Dr. Jacob Bien for assistance with the development and implementation of statistical methods. We thank Caitlin Niederhofer, Alyssa Cornelius, and Frances Melgar for assistance with manual image processing. We would like to acknowledge the following individuals for support with sample collection and processing: Ken Paddock, Caitlin Niederhofer, Eric Cerretani, Jeremy Schwartzbord, Mohammed Al Abri, Dan Miller, and Jacqueline Benson. This project was supported in part by the Northeast Sustainable Agriculture Research and Education (SARE) program (Graduate Student

583 GrantNumber GNE11-024). SARE is a program of the National Institute of Food and
584 Agriculture, U.S. Department of Agriculture.

585

586 **References**

587 Ahonen, T., Hadid, A., Pietikainen, M., 2006. Face description with local binary patterns:
588 Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine*
589 *Intelligence* 28, 2037-2041.

590 Ahonen, T., Pietikainen, M., Hadid, A., Maenpaa, T., 2004. Face recognition based on the
591 appearance of local regions. *Proceedings of the 17th International Conference on Pattern*
592 *Recognition. ICPR, Cambridge, UK*, pp. 153-156 Vol.153.

593 Aitkenhead, M., Dalgetty, I., Mullins, C., McDonald, A., Strachan, N., 2003. Weed and crop
594 discrimination using image analysis and artificial intelligence methods. *Computers and*
595 *Electronics in Agriculture* 39, 157-171.

596 Cherney, J.H., Cherney, D.J.R., Parsons, D., 2006. Grass silage management issues. *Proceedings*
597 *from Silage for Dairy Farms: Growing, Harvesting, Storing, and Feeding. Natural*
598 *Resource, Agriculture, and Engineering Service, Harrisburg, PA*, pp. 37-49.

599 Ene, J., Leighton, E., Blue, G., Bell, B., 2015. Multilevel Models for Categorical Data Using
600 SAS® PROC GLIMMIX: The Basics. *SAS Global Forum 2015, Paper 3430-2015*,
601 *Dallas, TX*.

602 Himstedt, M., Fricke, T., Wachendorf, M., 2009. Determining the contribution of legumes in
603 legume–grass mixtures using digital image analysis. *Crop Science* 49, 1910-1916.

604 Himstedt, M., Fricke, T., Wachendorf, M., 2010. The relationship between coverage and dry
605 matter contribution of forage legumes in binary legume–grass mixtures. *Crop Science* 50,
606 2186-2193.

607 Himstedt, M., Fricke, T., Wachendorf, M., 2012. The benefit of color information in digital
608 image analysis for the estimation of legume contribution in legume–grass mixtures. *Crop*
609 *Science* 52, 943-950.

610 Kawamura, K., Watanabe, N., Sakanoue, S., Lee, H.J., Inoue, Y., 2011. Waveband selection
611 using a phased regression with a bootstrap procedure for estimating legume content in a
612 mixed sown pasture. *Grassland science* 57, 81-93.

613 McRoberts, K., Cherney, J., Benson, B., Cherney, D., 2012a. Reducing farmer uncertainty in
614 spring forage harvests: Digital image analysis and artificial intelligence to predict alfalfa-
615 grass stand composition. *ASA, CSSA, and SSSA International Annual Meetings*,
616 *Cincinnati, OH*.

- 617 McRoberts, K.C., Cherney, J.H., Benson, B., Cherney, D.J.R., 2012b. Image recognition for
618 alfalfa-grass mixtures. 2012 Joint Conference NAAIC, Trifolium, & Grass Breeders,
619 Ithaca, NY.
- 620 Meyers, R.H., 1990. Classic and modern regression with applications, 2nd edition. Duxbury
621 Press.
- 622 Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant
623 texture classification with local binary patterns. IEEE Transactions on Pattern Analysis
624 and Machine Intelligence 24, 971-987.
- 625 Parsons, D., Cherney, J., Gauch, H., 2006a. Estimation of preharvest fiber content of mixed
626 alfalfa–grass stands in New York. Agronomy Journal 98, 1081-1089.
- 627 Parsons, D., Cherney, J.H., Gauch, H.G., 2006b. Estimation of spring forage quality for alfalfa in
628 New York State. Forage and Grazinglands 4, DOI: 10.1094/FG-2006-0323-1001-RS.
- 629 Parsons, D., Cherney, J.H., Peterson, P.R., 2009. Preharvest neutral detergent fiber concentration
630 of alfalfa as influenced by stubble height. Agronomy Journal 101, 769-774.
- 631 Parsons, D., McRoberts, K.C., Cherney, J.H., Cherney, D.J.R., Bosworth, S.C., Jimenez-Serrano,
632 F.R., 2012. Preharvest Neutral Detergent Fiber Concentration of Temperate Perennial
633 Grasses as Influenced by Stubble Height. Crop Sci. 52, 923-931.
- 634 Polder, G., van Evert, F.K., Lamaker, A., Jong, A.d., Van der Heijden, G., Lotz, L., van der
635 Zalm, A., Kempenaar, C., 2007. Weed detection using textural image analysis. 6th
636 Biennial Conference of the European Federation of IT in Agriculture (EFITA), Glasgow,
637 UK. Available online at <http://edepot.wur.nl/28203>. .
- 638 Post, C., DeGloria, S., Cherney, J., Mikhailova, E., 2007. Spectral measurements of alfalfa/grass
639 fields related to forage properties and species composition. Journal of plant nutrition 30,
640 1779-1789.
- 641 Radiocommunication Sector of ITU, 2002. Parameter values for the HDTV standards for
642 production and international programme exchange. Recommendation ITU-R BT.709-5
643 (04/2002). International Telecommunication Union, Geneva.
- 644 Rayburn, E.B., 2014. Measuring legume content in pastures using digital photographs. Forage
645 and Grazinglands 12, DOI: 10.2134/FG-2011-0143-MG.
- 646 Rayburn, E.B., Green, J.T., 2014. Visual reference guide for estimating legume content in
647 pastures. Forage and Grazinglands 12, DOI: 10.2134/FG-2011-0176-DG.
- 648 Sabeenian, R., Palanisamy, V., 2010. Crop and weed discrimination in agricultural field using
649 MRCSF. International Journal of Signal and Imaging Systems Engineering 3, 61-69.

650 SAS Institute Inc., 2011. SAS for Windows. v. 9.3. SAS Inst., Cary, NC.

651 SAS Institute Inc., 2013. JMP PRO 11.2.0 for Windows. SAS Inst., Cary, NC.

652 Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N.,
653 Gouillart, E., Yu, T., the scikit-image contributors, 2014. scikit-image: image processing
654 in Python. PeerJ 2, DOI: 10.7717/peerj.7453.
655