Ecological Indicators 121 (2021) 106946

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Measuring nested frequency of plants from digital images with SampleFreq

Samuel E. Cox[a,*], D. Terrance Booth[b], Robert D. Berryman[c]

[a] USDI Bureau of Land Management, Wyoming State Office, 5353 Yellowstone Rd, Cheyenne, WY 82009, United States
[b] USDA Agricultural Research Service (retired), High Plains Grassland Research Station, 8408 Hildreth Rd, Cheyenne, WY 82009, United States
[c] Berryman Consulting, 3091 Mill Vista Rd #1506, Highlands Ranch, CO, 80129, United States

ARTICLE INFO

Keywords:
Vegetation
Density
Frequency
Photography
Monitoring

ABSTRACT

As our understanding of ecological systems grows, natural resource management becomes ever more dependent on timely, accurate, and inexpensively-collected monitoring data to support management decisions. Vegetation cover, density, and frequency are abundance metrics used in resource management; however, frequency data can be collected more quickly than density data and with more repeatability and less sensitivity to inter- and intra-seasonal variation in plant morphology. Moreover, frequency is perhaps the best method for monitoring invasive species across extensive areas. A limitation to the use of frequency data is that plot size affects frequency. The optimal plot size is one that yields measurements suitably removed from 0 or 100% to allow detection of both upward and downward frequency trends, yet the optimum plot size cannot be known before sampling. We addressed this conundrum by developing SampleFreq software that facilitates frequency measurements from digital nadir images of any scale with up to 10 nested plot sizes within the confines of the image dimensions. We conducted accuracy and agreement tests of the software using both artificial populations and field plots. Using artificial population plots, accuracy across all users was 93.4% with a repeatability coefficient of 1.4%, indicating high precision. In a field test, SampleFreq and standard field data averaged a 3.4% difference, and were within approximately 10.5% of each other 95% of the time. From the same field test, SampleFreq repeatability coefficient was 6.7%, while the field method was 4.3%, illustrating that both methods have relatively high precision. Because SampleFreq has high potential accuracy, high agreement with field data, and high precision across a range of users, we recommend using SampleFreq with nadir digital images as a suitable alternative method for monitoring plant frequency.

1. Introduction

"Would it not be presumptuous to think that all conceivable breakthroughs in sampling methodology have now been made, that all we must do is refine the known techniques and standardize their use?" –Shultz, Gibbens & Debano (1961)

Natural resource management decisions depend on timely, accurate, inexpensively-collected monitoring data. Land management agencies like the United States Bureau of Land Management or the United States Forest Service are tasked with approving or renewing permits for commercial land use (e.g. grazing, mineral extraction) that rely on monitoring data. Where those monitoring data are not produced in a timely manner, delays in permitting cost commercial enterprises and reduce or delay royalties to the federal government. Where those monitoring data are inaccurate, the resource condition may not actually support the management decision. Where those data are collected at great cost, either insufficient data are collected or the citizens who collectively own, and fund management of, public lands end up paying too much.

Vegetation abundance measurements regularly used in environmental surveys are cover, density, frequency, and mass (Mueller-Dombois and Ellenberg, 1974; Elzinga et al., 1998). Frequency is the percentage of sample plots in which a species of interest occurs (Hyder et al., 1963). Occurrence may be rooted frequency (positive count only if the plant is rooted in the plot/frame), or more rarely, canopy frequency (any part of the plant in the frame is a positive count), in which the sample contains most of the species of interest (Hyder et al., 1963; Barbour et al., 1987). Density is a count of units per area (e.g. stems/m^2), whereas frequency only measures presence or absence within the unit area. By avoiding determination and counting of separate units, frequency data can be gathered with greater speed than density data (Elzinga et al., 1998). Though often correlated with cover in homogenously-distributed plant communities, frequency can, and does, vary independently from cover (Barbour et al., 1987; Elzinga et al., 1998).

* Corresponding author.
E-mail addresses: secox@blm.gov (S.E. Cox), terry_booth@outlook.com (D.T. Booth), r.berryman@comcast.net (R.D. Berryman).

https://doi.org/10.1016/j.ecolind.2020.106946
Received 9 March 2020; Received in revised form 14 August 2020; Accepted 8 September 2020
Available online 05 October 2020

For example, frequency can remain constant for a single plant in a frame, while cover could increase severalfold during the growing season. This intra-season stability of frequency data is a method advantage. Frequency data are collected simply, quickly, and objectively with high repeatability, since decisions are limited to presence/absence. Frequency is less sensitive than cover or density to inter- and intra-seasonal variation in plant morphology (Elzinga et al., 1998; Coulloudon et al., 1999). Walker (1970) found that among 8 monitoring methods, frequency documented the most species, and he concluded, "*Frequency was the only method to provide acceptable estimates of the importance of all species without the expenditure of excessive amounts of time*". Of all methods, it is perhaps the best for detection of exotic species invasions across large areas, since the principle question is that of simple presence.

Frequency precision = 100%/n, and thus the method requires large sample sizes, on the order of 20–50 plots, to achieve useful detection precision. For example, 20 plots yields a detection precision of 100%/20 = 5%, but changes in frequency < 5% will be undetectable. For ± 1% precision, 100 plots are required. The frequency method's principal limitation is that plot size and shape dictate frequency, and temporal or spatial comparisons of frequency are only valid across same-sized plots (Elzinga et al., 1998), a limitation not shared by density or cover data. Optimal plot size is one that yields frequency measurements suitably far enough away from 0% or 100% to allow detection of both upward and downward frequency trends. For example, a plot size that returns a frequency measurement of 10% will only ever be able to record a frequency reduction of 10%; any frequency reduction beyond that will be beyond the limit of the plot size to measure (Coulloudon et al., 1999). A plot size yielding frequency between 20 and 80% imparts maximal change detection across a wide range of dispersion patterns (Coulloudon et al., 1999; Heywood and Debacker, 2007). Curtis and McIntosh (1950) concluded that the optimal plot size is ~2X the mean area of the most common species, and the smallest that detects all randomly-distributed species at < 86% frequency. Given those guidelines, Hyder et al. (1963) calculated that 15–25 cm square plots best measured frequency in Oregon sagebrush/Idaho fescue rangeland. Hyder et al. (1965) determined that 5–40 cm square plots best measured frequency in Colorado shortgrass prairie. Mosley et al. (1989) calculated that 5–50 cm square plots best measured frequency in Idaho mountain meadows. All of these studies determined optimal plot size *after data were collected*. Determining the optimal plot size prior to field work is largely guesswork, or as Hyder et al. (1963) put it, "*less objective than one might desire*." Given expectations of varying frequency by species, nested frequency plots maximize sampling efficiency across multiple species by improving the chance that at least one of the plot sizes within the nested set is near-optimal for every species of interest (Coulloudon et al., 1999). However, even nested plot sizes are determined *a priori*, with no guarantee that nested plots contain an optimal plot size for every species of interest, or that a given plot size will be optimal for a particular species over time. Thus, the disadvantage of nested frequency plots is that the first observations lock in the plot sizes that must be used forever after if the data are to yield trend information.

A disadvantage of all field methods is susceptibility to classification variation by users (Vittoz and Guisan, 2007). For example, Walker (1970) recorded 10% user variation with line point intercept, and this number was confirmed by Cagney et al. (2011) who reported 11% user variation with the same method. It is rare that a single technician records field data for a particular area over multiple decades, and so any temporal change that appears to occur in landscape vegetation data is potentially the product of different users and their varying classification definitions and styles as much as actual changes in vegetation. Even if a single technician did collect data over multiple decades, it is reasonable to assume changing skill level and visual acuity of that technician over time further biasing data interpretation.

To overcome these disadvantages, we developed SampleFreq (2020). This free software facilitates nested frequency classification from digital images and allows *ad hoc* selection of nested plot sizes within the confines of the image dimensions. This essentially allows users to "go back in time" and redefine plot sizes for image data sets at will, utilizing a plot size that is not just optimal at one point in time, but over all points in time (as opposed to a field-sampling plot used at a moment in time that can never be resized). SampleFreq utilizes digital, nadir images of any scale that are often already collected for other image analysis purposes, such as vegetation cover with SamplePoint (Booth and Cox, 2008; Curran et al., 2020), linear feature measurements with ImageMeasurement (Booth et al., 2006a), green cover measurements with eCognition or VegMeasure (Laliberte et al., 2007; Louhaichi et al., 2001) or species cover measurements with Imagine (Everitt et al., 2001). The US Forest Service, the Natural Resource Conservation Service and the Bureau of Land Management have recommended collection of plot photos for more than 20 years (Coulloudon et al., 1999), resulting in large collections of permanent photographic records that can be analyzed, or reanalyzed, with SampleFreq. In this study, we sought to validate the operation of this software in order to allow land managers to use it confidently for vegetation monitoring.

Accuracy is close agreement of sample means with actual parameter means. Precision is close agreement of sample means to each other, without reference to the true mean (Barbour et al., 1987). SampleFreq was designed to facilitate measurements with high accuracy and high precision—we tested the null hypothesis that SampleFreq measurements would be neither. Our goals in this study were to assess 1) the accuracy and precision of the software by using known, artificial population plots, 2) the agreement of data collected using the software with that of standard field-collected data, and 3) the precision of software measurements from field images.

2. Materials and methods

2.1. SampleFreq software

SampleFreq software facilitates manual nested frequency object classification of digital images of any known scale. We developed SampleFreq using the C# programming language as a tool to measure plant frequency, though it could be used to measure frequency of any visible object. SampleFreq's user interface allows a user to define up to 10 nested frame sizes, define up to 60 object classes to measure, and then facilitates nested frequency measurements from nadir-perspective images by requiring presence/absence determination by the user through the act of clicking on-screen buttons (Fig. 1). Classification data are automatically saved to an Excel spreadsheet (Microsoft, Redmond, WA, USA). A user can adjust image brightness, contrast, and magnification, to see detail. Upon analysis completion, users can create a summary spreadsheet of the nested frequency data that is comparable in format to standard field data results.

2.2. Creating artificial random populations for testing software accuracy and precision

Accuracy assessment requires a standard for comparison. Curtis and McIntosh (1950) evaluated accuracy of frequency frame sizes using large photographs of scattered, recognizable objects (nuts, safety pins) as artificial populations. We employed a similar strategy and created artificial populations of colored dots within digital images in order to precisely define the comparison standard, accepting the disadvantage that these colored dots do not replicate real world vegetation shapes or colors.

We used ArcMap 10.0 (ESRI, Redlands, CA, USA) to fill a virtual 4 × 5-m rectangle with 10 random point populations. Each population was represented by a unique color and symbolized as 2-cm diameter dots (3.14 cm^2), where dot density varied with dot color (Table 1).

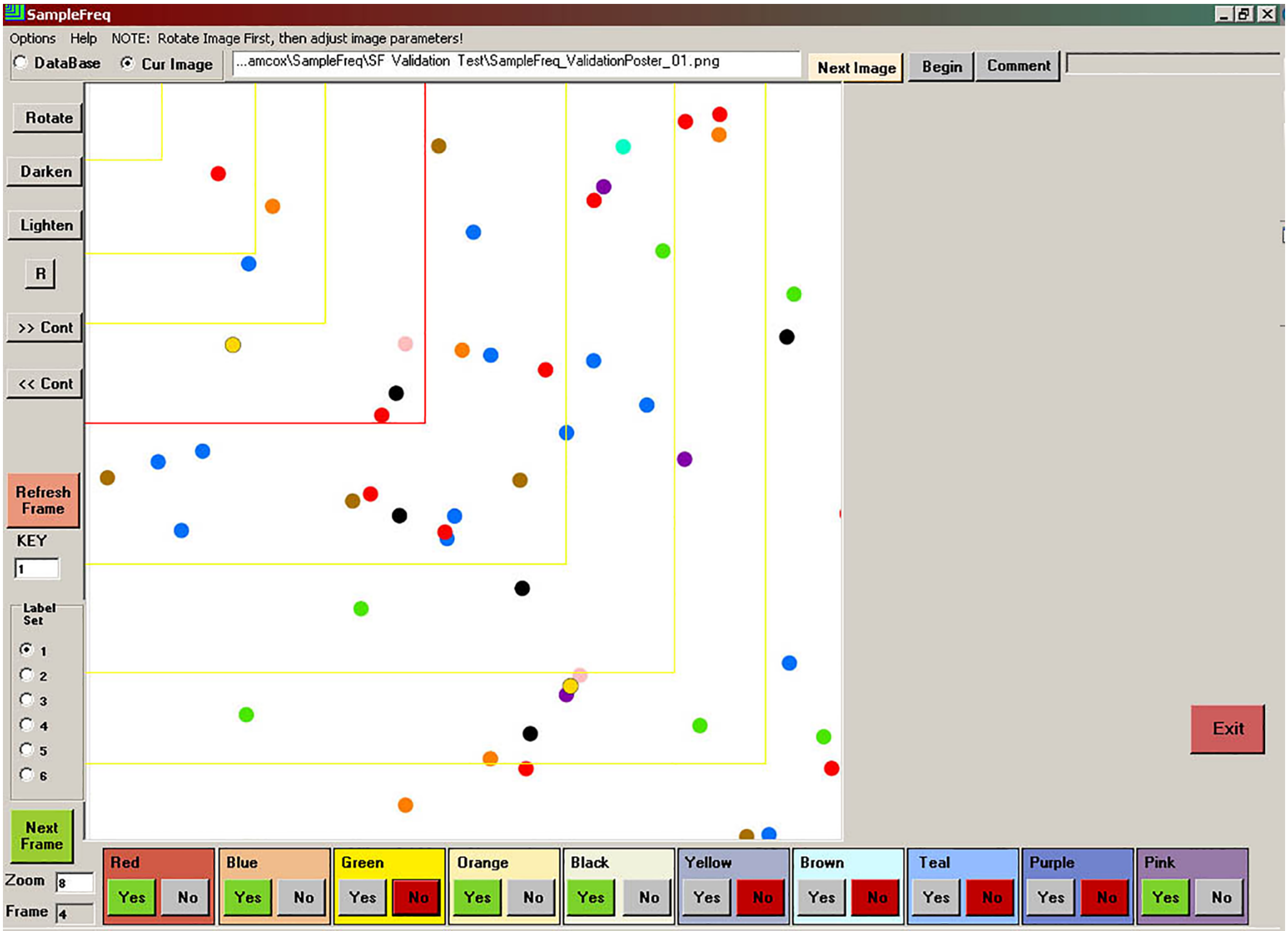

Fig. 1. *SampleFreq* screenshot showing series of 8 nested frames (yellow boxes with active frame in red) overlaid on one of 20 (A) artificial population images of randomly-distributed colored dots and (B) field-trial images acquired in mixed-grass prairie. User-defined classbuttons are shown on the bottom with presence (yes) or absence (no) classification options.

Some colors are naturally more eye-catching, and we anticipated that some bright colors, such as red, might be more accurately measured. Color as a variable is qualitative, categorical, and thus not readily subject to statistical comparison, but color luminance is a quantitative, discrete variable that *is* measurable. Color luminance describes emitted light, but in digital imagery, potential emitted light has limits, resulting in light emissions constrained by the gamut of the display hardware (Poynton, 1996). Simply put, a computer monitor isn't bright enough to display real-world color and brightness. As a result, image colors are described by luma, which is formulated from the gamma-compressed RGB values of digital images. If shown three equal intensities of color: red, green, and blue, humans perceive green as brighter than red, and red brighter than blue (Poynton, 1996). To compensate for this physiological feature of the human eye, coefficients are added to the luma equation to weight the primary colors based on an international standard (Rec. 601), thus giving digital images high-fidelity color (Poynton, 1996; Bezryadin et al., 2007). Color luma was calculated using a set of standardized correction coefficients, as follows (from Poynton, 1996):

Table 1

Color, density, luma and luma rank of random dot populations used in the artificial population, along with classification accuracy across all users (n = 15) by color, including the sum of all omission errors, commission errors and combined errors. For each color, there were a total of 120 classification values (8 frames × 15 users), so the total error counts are partitions of 120. Measurement standard deviation (SD) are shown for each color, and overall.

Color	Density (dots/m^2)	Luma	Luma Rank	Accuracy	Omission Errors	Commission Errors	Total Errors	SD
Black	6	0	1	92.5	5	4	9	0.75
Blue	9	95	4	95.8	5	0	5	0.30
Brown	4	116	5	95.0	1	5	6	1.13
Green	8	158	7	100.0	0	0	0	0.00
Orange	7	151	6	92.5	9	0	9	0.78
Pink	1	209	10	95.8	5	0	5	0.45
Purple	2	59	2	85.0	15	3	18	1.54
Red	10	76	3	95.0	0	6	6	0.30
Teal	3	172	8	93.3	5	3	8	1.30
Yellow	5	202	9	89.2	3	10	13	0.54
All				93.4	48	31	79	0.71

$$Luma(Y') = 0.299R' + 0.587G' + 0.114B' \quad (1)$$

where R' = gamma-compressed red band value, G' = gamma-compressed green band value and B' = gamma-compressed blue band value. For all tests in this study, the luma of each color was used as a single-number quantitative surrogate for color that allowed us to conduct statistical tests on color effects. As each color has a unique luma, the terms color and luma are largely interchangeable in the context of our results.

We saved the resulting figure as a 4000 × 5000-pixel TIF file, with resulting resolution of 1-mm ground sample distance (GSD = the real-world distance depicted by one side of a square image pixel). We sliced this image into twenty 1000 × 1000-pixel squares (1 × 1 m) using PhotoShop 12.0.3 (Adobe, San Jose, CA, USA), saving each square as a PNG file named sequentially 1–20. This process created a group of simulated m^2 quadrat images representing a 10-class population with random distribution and known frequency for each color.

2.2.1. Measuring the artificial random populations for testing software accuracy and precision

We used ArcMap to create a virtual, square-shaped, 8-frame nested frequency feature class (Fig. 1) and overlaid this 8-frame feature on each artificial plot. We used the Identity tool to locate dot centers by frame, and sorted results in Excel to determine the smallest nested frame occupied by each color in each artificial plot. We entered these data into a SampleFreq summary file of the same format generated by the software.

2.3. Software testing

Fifteen rangeland management professionals used SampleFreq 1.0 to measure frequency of all 10 colors across all 20 images from Section 2.2, with the criteria that the dot center must fall within the active frame in order to be counted (Fig. 1). To ensure standardization, we provided each user with a SampleFreq database pre-formatted with the entire image set, the correct image GSD, and the same 8 nested frame sizes/shapes described above in Section Section 2.2.1. We assessed the influence of red-green colorblindness on SampleFreq performance by including a 16th user with this condition, but these data were not used in calculations of overall accuracy or precision of the software.

2.3.1. Accuracy assessment

We compared each of the 15 users' 80 measured frequency values (10 colors × 8 frames) with the standard. Since the artificial population frequency is known, we simply graded user responses against known values. The proportion of correct user responses is the user accuracy.

We tested for differences in user accuracy by color and by frame size (1 color × 8 frames/user; 1 frame × 10 colors/user) using a non-parametric Kruskal-Wallis test, followed by pairwise Wilcoxon Rank Sum tests with Bonferroni-corrected alpha to control Type I error (npar1way procedure, SAS 9.2; SAS Institute, Cary, NC). Since density varied independently by color (density and luma are not correlated; Table 1), we examined the pattern of accuracy by both color and dot density. We identified classification errors (among 80 classifications/user) into errors of omission (a user failed to count a dot in the frame it occurred) and commission (a user counted a dot in a frame where it didn't exist), and similarly tested these for differences by luma to assess whether some colors were more prone to errors of commission or omission (Kruskal-Wallis test, pairwise Wilcoxon Rank Sum tests with Bonferroni's alpha correction). As with accuracy, tests of differences by color also tested differences by density. All relationships were also assessed with correlation coefficients (Excel 2016; Microsoft, Redmond, WA), to determine the extent, if any, of linear relationships "accuracy × luma/density", and "errors of omission/commission × luma/density".

User responses were regressed on known standard values to generate coefficients of determination (r^2) to describe how well user measurements fit the artificial population values across the full range of potential frequency values. A coefficient of determination is not by itself a test of accuracy since data can be highly correlated, but offset; however, in combination with other tests, a regression can add to the understanding of how two data sets relate to each other (Bland and Altman, 1986). The regression line can indicate method disagreement and bias, if the slope and Y-intercept are appreciably different than 1 and 0, respectively.

We compared user accuracy by gender and the use of corrective lenses with Welch's T-tests to correct for unbalanced sample sizes. We used coefficients of determination to assess relationships of age (6 levels of 5-year interval groups) and screen resolution (5 ranked levels) with accuracy. Specifically, age was recorded by 5-year interval, and a person was classified as wearing corrective lenses if they wore eyeglasses or contact lenses on a regular basis.

2.3.2. Precision assessment

Measurement standard deviation is a simple measure of precision, and when multiplied by 1.96, becomes the repeatability coefficient, which is the value below which the absolute differences between two measurements would lie with 0.95 probability (Bland and Altman, 1986, 2003). We calculated measurement standard deviation by frame size, and by color. We then calculated the overall method repeatability coefficient.

2.4. Agreement with the nested frequency field method

Users of the software should be confident that data generated using SampleFreq will replicate data collected using a standard nested frequency frame. We tested a null hypothesis that the two methods will have significant and meaningful differences.

Twenty random plots were established within a 10-acre polygon of mixed grass prairie at the High Plains Grassland Research Station in Cheyenne, WY (41.183°, −104.900°). Each plot was 1 × 1 m square, marked with 4 orange nails and a numbered survey flag for easy location by recruited observers. One person acquired nadir images of all plots in 24 min using a Canon 1100D 12-megapixel digital camera mounted on a 2 m-AGL aluminum camera frame, as described by Booth et al. (2004). (We have since determined that a monopod-mounted camera is more efficient than the aluminum frame in most situations; see Curran et al., 2020). Images were acquired with the camera set to aperture priority mode (*f*/14), ISO = 200, with a resulting shutter speed range of 1/60–1/160 (We now know that 1/60 s is too slow a shutter speed and often results in blurred images when examined at the pixel level. We recommend using $< 1/250$ s). Images were cropped to the plot corners using Photoshop CS5 (Adobe, San Jose, CA), color-balanced to achieve natural color and illumination, and resized to exactly 2100-pixels wide to ensure all images were of equal resolution (0.48 mm GSD).

We measured rooted frequency of 5 forb species using the same 8 nested square frame sizes/shapes as the artificial population test above. To align field and software observations, two nested frequency frames were fabricated out of 1-cm steel rod and 2.5-cm steel angle stock, like the design illustrated by Coulloudon et al. (1999). Twelve rangeland professionals were recruited to measure nested frequency from all 20 plots both in the field using the metal nested frequency frame, and in the lab using SampleFreq 1.0. All observers were given the same data sheets and written instructions on data collection protocol. All observers used the same image set to reduce bias from varying photography techniques. Half the observers acquired their field data first, and the other half acquired their SampleFreq data first, to balance potential bias from performing one method prior to the other. Every observer was given a directory with a blank SampleFreq database and all 20 plot images for analysis. Observers classified all images and returned completed databases and summaries to the authors. One observer did not adhere to the instructions to record rooted frequency and recorded canopy frequency instead; that observer's data were discarded from analysis.

To assess method agreement, we compared the standard-method data to SampleFreq data using a variety of metrics. We plotted the data with a 1:1 equality line to examine the relationship of the SampleFreq measurements against the standard field measurements. We calculated the Pearson product moment correlation coefficient (r) for the relationship, which has some value on its own, but must be interpreted cautiously: data can be highly correlated without having high agreement (Bland and Altman, 1986). For example, 2 methods that produce identical results will show r equal to 1, a perfect correlation, but if method one produces results exactly twice as large as method two, r still equals 1, a perfect correlation despite low method agreement. For this reason, r (or r^2) alone doesn't measure agreement. Plotting the data with a 1:1 equality line protects against incorrect interpretation of the r statistic.

A low average difference between measurements would seem to be a good indicator of agreement, except that if some differences are large positive values, and others are large negative values, the mean of the values as a surrogate for agreement will be misleadingly low. In fact, the average difference between measurements is actually measuring bias: a high positive mean difference indicates that method 1 consistently overcounts, and vice versa. By squaring the differences, calculating the mean, then taking the square root, the effect of positive and negative differences is removed, so that the root mean square error (RMSE) measures absolute difference. As a single statistic, the RMSE is useful for describing overall agreement between data sets but may mask differences in agreement at various measurement levels and convey a sense of consistency that may not exist. For example, measurements of sagebrush frequency may show very low error when sagebrush frequency is low, but very high error when sagebrush frequency is high. For this reason, RMSE, though useful, must also be interpreted cautiously. We present a data plot, correlation and RMSE because they are familiar metrics to most; however, there is a better way to assess agreement.

Bland and Altman (1986) proposed a robust test of method agreement. A variable is measured by two methods, and the difference between the two values (A-B) is plotted on the Y axis against the mean of the two values (A + B)/2 on the X axis. This is repeated many times, at many levels of the variable. The goal is to examine method differences relative to a true value, but since the true value is almost always unknown, the average of the two methods is the closest approximation of the truth available. The resulting distribution illustrates how well two methods agree. If method differences are evenly distributed both above and below the mean difference across the entire X axis, and the differences are small, then the methods show good agreement. A measurement difference mean of zero indicates no bias, whereas if most errors are above the mean, then method A is measuring the variable higher than method B, and vice versa. The standard deviation of those differences can be used to quantify the agreement further. The empirical rule states that 95% of values will fall within 2 standard deviations of the mean in a normal distribution, so by doubling the standard deviation of the measurement differences (or more precisely, multiplying by 1.96) and adding or subtracting it from the mean of the method differences, an upper and lower value are determined that encompass 95% of anticipated method differences. Bland and Altman (1986) called these the limits of agreement. A small limits of agreement interval indicates high method agreement, with the exact determination of what constitutes "agreement" left to the particular requirements of the investigator.

Measurement standard deviation is a simple measure of precision, and when multiplied by 1.96, becomes the coefficient of repeatability, which is the value below which the absolute differences between two measurements would lie with 0.95 probability (Bland and Altman, 1986, 2003). We calculated measurement standard deviation by frame size, and by species, as we did for artificial population testing above. We then calculated a repeatability coefficient for both methods.

3. Results and discussion

3.1. Accuracy

SampleFreq accuracy of 10 populations of colored dots was 93.4 ± 4.2% and measurements were highly correlated to actual values ($r^2 = 0.99$, $p < 0.0001$, $n = 1200$). The best fit line equation for these data is $y = 1.0006x–0.0008$, which indicates virtually no method bias (~zero y-intercept), and a near-perfect relationship between the measured and known values (slope of 1). For comparison, Booth et al. (2006b), Booth et al. (2006c) reported cover-method coefficients of determination (r^2) to known cover values of 0.99 for steel point frame and SamplePoint, 0.98 for line point intercept, and 0.97 for laser point frame and ocular estimate. SampleFreq measurements show parity with these cover measurement methods. Individual user accuracies ranged from 85.0 to 97.5% ($n = 80$). A red-green colorblind user's accuracy was 92.5%, not significantly different than the mean accuracy for all users.

3.2. Accounting for the average 6.6% error

Though the error with SampleFreq is small, it is worth exploring the sources of error to determine if the method suffers from any systematic bias.

3.2.1. Neither color nor density explain accuracy variance

Accuracy varied by dot color ($p = 0.0038$, $n = 15$) in only 2 of 45 comparisons: yellow differed from brown ($p = 0.0004$, $n = 15$) and green ($p < 0.0001$, $n = 15$), all of which have mid-range luma and density (Table 1), thus neither variable explains the variance in accuracy. The lowest accuracies were with purple and yellow, the 2nd lowest and 2nd highest luma, respectively, with densities of 2 and 5 dots/m^2, neither of which support a trend of accuracy by luma or density (Table 1). The red-green colorblind user was the only one to misclassify any green dots (2), but did not misclassify any red dots, though 33% of observers did.

3.2.2. Accuracy varied by frame size

Accuracy across all nested frequency frame sizes is presumed to be equal (e.g., accuracy from a 2-m^2 frame should equal that from a 0.5-m^2 frame within the same nested series). Though careful classification by a user should affirm this assumption, we found that in practice this was not the case. Accuracy varied within the nested frame set by frame size ($p = 0.0002$, $n = 15$), with mid-size frames showing lower accuracy (Fig. 2). We speculate this is caused by two opposing trends. The smallest frame has the lowest probability of containing dots (Fig. 1), and thus presents a simpler classification scenario, e.g., if there are no dots in frame 1, it is easy to classify correctly. As frame size increases, users must scan more area with greater probability of encountering multiple dots, presenting increasing classification complexity and error risk. This trend is evident through the first four frames (Fig. 2). An opposite trend arises from the cumulative nature of nested frequency counts (each larger frame includes the entire area of all previous frames). Once a color is counted as present, it retains that status through subsequent frames. At some frame size, a color can achieve 100% presence across all plots, and from there on, it is always 100%, and the user does not have the option to classify it. This reduces the number of decisions a user must make at each subsequent frame. For example, Table 2 shows all ten colors < 100% frequency in frame 4 and a user must look for all ten; in frame five, two colors show 100% frequency, and are then "retired" from classification, so the user has only to look for eight colors, by frame 6, only six, and so on. This ever-decreasing classification complexity yields decreasing probability of

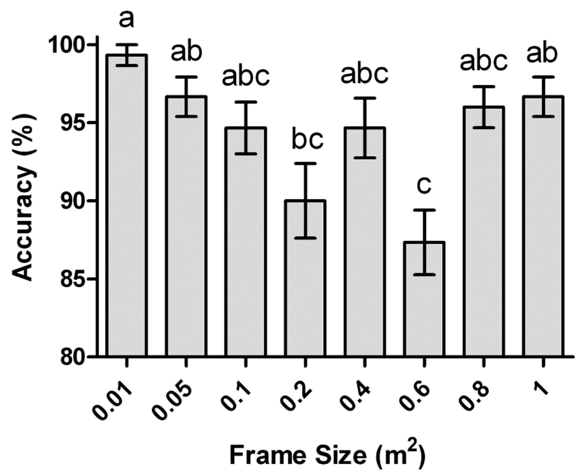

Fig. 2. Frequency measurement accuracy (n = 15 users) across all colors by nested frame size. Significant differences between frame sizes are noted with differing letters.

Table 2
Artificial population frequency by color and nested frame size. Each value in this table represents the known frequency of colored dots in the artificial population of 20 images against which observer data were graded.

Color	Frame Size (m^2)							
	0.01	0.05	0.1	0.2	0.4	0.6	0.8	1
Red	0	30	55	90	100	100	100	100
Blue	15	30	55	85	100	100	100	100
Green	10	45	65	85	90	100	100	100
Orange	0	45	60	85	95	100	100	100
Black	5	20	35	70	85	90	95	100
Yellow	5	40	45	65	70	90	100	100
Brown	5	25	35	55	70	90	95	100
Teal	0	10	25	50	75	85	90	100
Purple	0	0	10	20	30	60	80	95
Pink	0	5	10	20	30	45	55	65

making errors, demonstrated with a general accuracy increase from frames 5 to 8 (Fig. 2). We speculate these two trends explain the dip in accuracy among the mid-range frames in the nested plots. The implication for vegetation surveys is that not only does frame size have to be constant over time (Elzinga et al., 1998), but all frames within a nested frame design must also be constant: a 1 m^2 frame that is the smallest of six nested frames is not comparable to a 1 m^2 frame that is the mid-size of six nested frames. Given that complexity within frames seems to influence accuracy, subdividing larger frames to reduce frame complexity while conserving whole-frame tallies may mitigate this error risk.

3.2.3. Accuracy by user gender, age, and eyesight—age matters

Across 15 users, accuracy did not vary by gender (p = 0.52, n = 5 men, 10 women) or by corrective lens use (p = 0.91, n = 6 with, 9 without). Accuracy increased with higher screen resolution (r = 0.63, p = 0.01, n = 6) and was 4.5% greater with resolutions of ≥1280 × 1024 (p = 0.04, unbalanced *t*-test n = 7 ≤ 1280 × 1024, n = 8 ≥ 1280 × 1024), though this doesn't imply causation.

Accuracy decreased as user age increased (r = −0.58, p = 0.03, n = 6 age classes [range = 20–54]). We have previously reported age-related classification bias in point intercept cover data collected both in the field and from a computer screen (Booth et al., 2005; Cagney et al., 2011). Diminished visual acuity is often associated with age-related changes in the eye. User visual acuity was not clinically rated, but we used corrective lens use as an admittedly imperfect indicator of visual acuity and found no relationship with user age (r = 0, p = 1.0, n = 6 age classes). Diminished visual acuity might also be expected to be associated with lower screen resolution preference, where text and objects appear larger and are thus easier to see, but age was not related to chosen screen resolution (r = 0.44, p = 0.10, n = 6), nor was corrective lens use (r = 0.20, p = 0.45, n = 6). Thus, we have no indication that visual acuity of users in our test group is related to age and have no other data to test or explain why classification accuracy decreased with user age.

3.2.4. Omission and commission errors

Cumulatively, users made 48 omission and 31 commission errors, often associated with dots on the boundary between two frames, where a user had to determine dot center location. This situation explains 60% of omissions and 74% of commissions, values that do not indicate systematic bias. In field applications, we expect a similar distribution of errors among users determining whether to include a plant within a frame.

Omission errors varied by color (p = 0.0181, n = 15), though we found no significant pairwise differences, and no correlation with luma (r = 0.22, p = 0.52, n = 10 luma classes; Table 1). Purple dots (luma = 59) were omitted most frequently (Table 1). Black has lower luma (0) and was only omitted a third as often. The next brighter color, red (luma = 76), was not omitted at all (Table 1). Thus, we have no evidence the luma systematically influences omission. Purple dot density was low (2 dots m^2), yet pink (1 dot/m^2), and teal (3 dots /m^2) were omitted only a third as often, while the second highest omission rate belonged to orange (7 dots/m^2). Density was not correlated with omission errors (r = 0.48, p = 0.16, n = 10 density classes; Table 1).

Commission errors also varied by color ($p < 0.0001$, n = 15), with yellow added more often than blue, green, orange, and pink ($p < 0.001$, Table 1). However, these results do not indicate any relationship between color and commission errors. Overall, there was no relationship between commission errors and luma (r = 0.13, p = 0.72, n = 10) or density (r = 0.06, p = 0.86, n = 10).

3.3. Precision from artificial populations

There are several ways to think about repeatability with SampleFreq. One indication of precision is that the slope of the measured-to-known frequency regression line had an extremely narrow 95% confidence interval of 0.997–1.005. Another indicator of high precision is that classification measurement standard deviation across all colors through all frames is 0.71% (Table 1), yielding a repeatability coefficient of 1.4%. In other words, two frequency measurements of the same variable will be less than 1.4% different, 95% of the time (Bland and Altman, 1986). However, examining standard deviations by frame size and color allows for a more nuanced interpretation (Table 3).

Table 3
Standard deviations of all measurements by frame size and color recorded during the artificial population trials. Lower values indicate higher repeatability.

Color	Frame Size (m^2)								*Mean*
	0.01	0.05	0.10	0.20	0.40	0.60	0.80	1.00	
Black	0.00	0.00	0.00	2.29	1.29	2.44	0.00	0.00	*0.75*
Blue	0.00	0.00	0.00	2.44	0.00	0.00	0.00	0.00	*0.30*
Brown	1.29	1.29	1.29	1.29	1.29	1.29	0.00	1.29	*1.13*
Green	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*0.00*
Orange	0.00	1.76	3.20	0.00	1.29	0.00	0.00	0.00	*0.78*
Pink	0.00	0.00	0.00	0.00	0.00	1.29	2.29	0.00	*0.45*
Purple	0.00	0.00	1.76	1.89	2.07	2.80	1.29	2.54	*1.54*
Red	0.00	0.00	0.00	2.44	0.00	0.00	0.00	0.00	*0.30*
Teal	0.00	1.29	1.29	1.89	2.07	2.58	1.29	0.00	*1.30*
Yellow	0.00	1.76	0.00	0.00	0.00	2.54	0.00	0.00	*0.54*
Mean	*0.13*	*0.61*	*0.75*	*1.22*	*0.80*	*1.29*	*0.49*	*0.38*	*0.71*

Repeatability, evidenced by smaller measurement standard deviation, is highest for green, blue and red, while it is lowest for teal, purple and brown, a pattern that does not indicate a relationship of repeatability with luma ($r = 0.15$, $p = 0.67$, $n = 10$) but does suggest a positive relationship with density, as red, green and blue are also the colors with the highest dot density (Table 1). Density seems to explain much of the variance in repeatability across colors ($r = 0.63$, $p = 0.049$, $n = 10$). Repeatability differed, barely, by frame size ($p = 0.04$, $n = 15$), though no pairwise differences were observed at the Bonferonni-corrected alpha. Repeatability was highest for the smallest and largest frame sizes (SD = 0.13%, 0.38%, respectively), and lowest for the mid-frame sizes (SD = 0.49–1.29%); the same pattern seen with accuracy (Table 3). We speculate that the same factors involved with a dip in accuracy across the mid frame sizes are at work here, though overall, a repeatability coefficient of 2.5% at a mid-frame size (0.6 m^2) still shows very high precision: there is 0.95 probability that any two measurements at this frame size should differ by < 2.5% (Bland and Altman, 1986).

3.4. SampleFreq agreement and precision compared well with the standard field method

Up to this point, we have only discussed data derived from artificial population tests. The frequency of the artificial population was precisely known, which allowed us to confidently assess method accuracy, and explore errors of omission and commission. Field plots do not share this quality. Their metrics are unknown and are measured by methods that may be subject to bias or error. Their frequency can never be determined with certainty. For this reason, we cannot assess accuracy of the software by examining field data, but we can assess agreement and precision. We approached this comparison with the assumption that neither method is inherently “correct”.

By all measures, SampleFreq data agreed very well with data from standard field methods. When plotted against each other, the data clustered along the 1:1 equality line, with no apparent bias (Fig. 3). Over 52% of all measurements were in perfect agreement between the two methods, and over 83% of measurements differed by ≤5%. Overall correlation of data between the two methods was 0.95 ($n = 440$), with a regression line equation of $y = 0.98x + 0.5$. The slope of the line is virtually 1, indicating consistent agreement, and the Y-intercept is 0.5 (Y axis units are % Frequency), indicating low bias (Fig. 3). Overall RMSE was 3.4% ($n = 440$). Given that a 20-plot frequency trial has a precision of only 5%, an error of 3.4% is less than one sampling unit.

Correlation of data between the two methods by species was always high ($r > 0.8$, $p < 0.0001$, $n = 88$) with RMSE always low (RMSE < 5%, $n = 88$, Table 4). Agreement of data by frame size was mixed: RMSE for all frame sizes was < 6%; data were well correlated for the 6 largest frame sizes (> 0.69), but not well correlated for the 2 smallest frame sizes (< 0.54, $n = 55$, Table 4). This lower correlation at the smaller frame sizes was due to the preponderance of “0” frequency values that were quickly skewed by even a few positive frequency values. Examination of the raw data reveals that at those small frame sizes, species presence was recorded 11 more times (out of 110 individual frame reads) using SampleFreq than the field method.

A plot of the method differences over the mean between the two methods (Fig. 4) shows an even distribution of error at all sampled frequency values. We expect 95% of differences between the two methods will lie within 2 standard deviations of the mean (or, more

Table 4
This table shows the agreement between data sets collected using SampleFreq and the traditional field nested frequency frame method. Agreement is examined by species, and by plot size. Overall agreement between methods is high ($R = 0.95$, RMSE = 3.4%, $n = 440$).

Species	r	RMSE (%)
fringe sage	0.99	0.4
globemallow	0.85	5.3
gumweed	0.98	2.8
thistle	0.80	4.3
toadflax	0.97	4.0
Plot (m^2)		
0.01	0.14	1.1
0.05	0.54	2.2
0.1	0.69	4.4
0.2	0.92	2.1
0.4	0.81	5.5
0.6	0.89	4.8
0.8	0.93	3.5
1	0.94	3.7
All	0.95	3.4

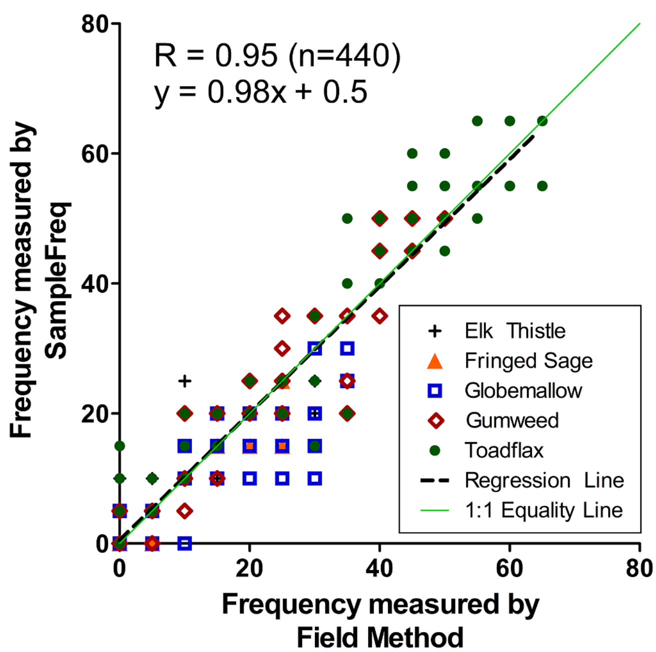

Fig. 3. Frequency measured by SampleFreq (Y) plotted against the standard field method (X). A line showing a perfect 1:1 equality is shown for reference. All 440 measurement comparisons are incorporated in the regression, but duplicate combinations (e.g. 5% SampleFreq × 5% Field), mask each other.

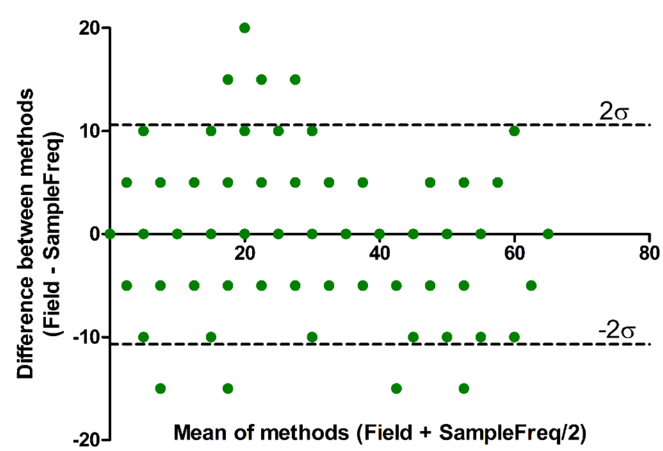

Fig. 4. Differences between all pairs of measurements from SampleFreq and the standard field method are plotted against the mean of those paired measurements. The mean difference between methods is −0.07, which is visually indistinct from zero; 95% of differences are expected within 2 standard deviations (2 s) — the Limits of Agreement — which are denoted as dashed lines. All 440 measurement comparisons are incorporated in the limits of agreement test, but duplicate combinations (e.g. 0% difference at mean 15% frequency), mask each other.

Table 5
Frequency measurement standard deviation values are used as indicators of repeatability. This table shows the standard deviation of all measurements by frame size and species for measurements recorded during the field trials by 12 observers.

Method	Species	Frame size (m^2)								*Mean*
		0.01	0.05	0.10	0.20	0.40	0.60	0.80	1.00	
SampleFreq	elk thistle	2.02	2.61	2.61	2.61	2.61	4.72	4.72	5.00	*3.36*
	fringed sage	0.00	0.00	0.00	1.51	3.44	0.00	0.00	0.00	*0.62*
	globemallow	0.00	0.00	2.61	2.52	4.72	6.36	6.32	6.11	*3.58*
	gumweed	1.51	3.44	4.91	2.34	6.88	3.37	3.23	3.02	*3.59*
	toadflax	2.02	4.37	7.54	6.88	5.48	5.84	6.07	10.09	*6.03*
	Mean	1.11	2.08	3.53	3.17	4.62	4.06	4.07	4.84	*3.44*
Field	elk thistle	0.00	3.75	0.00	0.00	3.50	1.51	1.51	1.51	*1.47*
	fringed sage	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*0.00*
	globemallow	0.00	2.02	2.02	2.02	2.02	4.05	4.91	4.05	*2.64*
	gumweed	0.00	2.61	5.22	2.61	6.74	2.61	2.61	2.61	*3.13*
	toadflax	2.24	3.87	4.67	4.67	3.44	3.37	2.34	5.13	*3.72*
	Mean	*0.45*	*2.45*	*2.38*	*1.86*	*3.14*	*2.31*	*2.27*	*2.66*	*2.19*

precisely, $\mu - 1.96$ s and $\mu + 1.96$ s). These are the limits of agreement between the two methods (Bland and Altman, 1986). Frequency measurements of the two methods differed by -0.07 ± 5.4, so we conclude that SampleFreq measurements will be no more than 10.7% below or 10.5% above a comparable field measurement, 95% of the time. Given that a 20-plot frequency trial has a precision of 5%, an error of 10.7% is approximately two sampling units. This does not imply that SampleFreq is in error, as it is just as likely that the SampleFreq data are more accurate than the field. This simply means that the methods rarely disagree by more than 10%, and the RMSE value indicates that they disagree by 3.4%, on average.

Using measurement standard deviation as an indicator of precision, the standard field method's precision across all frame sizes and species was 2.2%, compared to 3.4% for SampleFreq (Table 5). This is 3x and 5x greater than measurements from artificial populations, respectively, but given the additional variables of texture, height, and pattern presented by real plants, the lower precision for these measurement data is expected. Repeatability coefficients were 4.3% and 6.7% for the field and SampleFreq methods, respectively. This means that any two field measurements of the same variable should be < 4.3% different, 95% of the time, and < 6.7% different using SampleFreq (Bland and Altman, 1986). Since imprecise methods will rarely agree, high precision is required for high method agreement. We regard the difference in precision between these two methods to be minor and conclude that SampleFreq data can be regarded as having similar high precision as the traditional field method.

Contrasting repeatability coefficients of the SampleFreq method using field images (6.7%) with the artificial population (1.4%) illustrates the inherent uncertainties that arise when users are asked to classify vegetation. Compared to the classification of clear, obvious, equally-sized colored circles, users lost > 5% repeatability when classifying more complicated forms and colors of real vegetation. This intuitively makes sense, and we expect that examination of vegetation in the field similarly reduces classification precision. And while that precision is still relatively high, the difference between the potential precision of the method and the actual field precision is one reason why a permanent image record is so valuable: any doubts of classification accuracy can be easily examined in SampleFreq at any time. Field methods do not have a mechanism to doublecheck accuracy, so any precision shortcomings are not correctable (Cagney et al., 2011).

The pattern of precision by species was identical between the two methods: highest for fringed sage and lowest for toadflax (Table 5). The pattern of precision by frame size was different between the two methods, except that measurements taken on the smallest frame showed the highest precision for both, which is consistent with the pattern seen from artificial population measurements.

Regardless of method, accuracy and precision are at least in part predicated on a user's positive identification of where a plant is rooted. Elzinga et al. (1998) discuss the importance of establishing how plants will be counted near boundaries, and whether the entire stem must be inside the frame. In some cases, a plant's canopy may obfuscate that determination from a nadir perspective. This is a disadvantage of the SampleFreq method, as well as all remote sensing methods that utilize nadir imagery (Elzinga et al., 1998; Chase and Chase, 2017). A field user can push aside vegetation or alter their look angle to make this determination, but a SampleFreq user cannot. In ecosystems with thick and overlapping canopy, nadir image analysis of frequency, multi-hit cover or species richness can't be expected to yield accurate or precise results. In arid settings, vegetation is typically low-growing and sparse, so canopy obfuscation is less a concern. Nevertheless, though we cannot explicitly test for it, some of the error observed in the SampleFreq method relative to the field method could be due to the canopy-obfuscation effect, especially when the target plant canopy is adjacent to the nested frame boundary.

4. Application

Field work utilizing methods like nested frequency involves judgment, and the use of precision software like SampleFreq cannot eliminate those judgments. In the field, judgments can never be reexamined; once the technician leaves the plot, it cannot be precisely recreated. The distinct advantage of image-based methods is that all judgments made from imagery can be reexamined. Non-image field data are forever limited to what the technician recorded. In contrast, the use of digital images with image analysis software such as SampleFreq, and including many other programs, allow numerous options for data collection, reexamination, or recollection, and thus represent more than refinement of a method (Shultz et al., 1961), but a fundamentally new approach to vegetation monitoring.

In most temperate regions of the earth, vegetation is most recognizable from imagery during a relatively brief mature phenological stage where flowers, seeds, mature leaf shapes or colors are observed. A user can photograph plots in summer, and delay analysis until fall/winter, allowing more plots to be photographed within a brief phenological stage of interest, thereby expanding the reach of the monitoring effort, and increasing statistical power through increased sampling. Even in the tropics, flowers or fruits that are produced during certain parts of the year can be important visual clues for plant identification from imagery. Thus, though traditional field methods are often ensconced in natural resource professions, image-based monitoring offers ample reward to those who refuse to simply "refine the known techniques and standardize their use" (Shultz et al., 1961).

Because image resolution is an important factor in accuracy and precision of image-based point sampling methods (Booth and Cox, 2006; Booth and Cox, 2009; Duniway et al., 2012; Weber et al., 2013), it is likely to be an important consideration for image-based frequency sampling as well. Field plots in this study were imaged at 0.48 mm GSD. We presume that since this resolution allowed quick and confident recognition of all cover types that it is sufficient for general field use, but it is probable that even higher resolution imagery, which is certainly attainable given improved digital image sensors, will yield frequency measurements that are more precise and accurate than we have described here.

In this study we examined imagery acquired in a semi-arid mixed-grass prairie of low-growing vegetation (< 40 cm in height); however, there is no practical reason this technique cannot be used with taller vegetation, provided the imagery captures the vegetation from a nadir perspective. In many cases this can be accomplished with a handheld-camera (Cagney et al., 2011), a terrestrial-based monopod or camera frame (Booth et al., 2004), or a long boom (Roshier et al., 1997). Taller vegetation may require the use of unmanned aerial systems (Curran et al., 2020) or fixed-wing aircraft (Duniway et al., 2012) to capture nadir perspectives. Regardless of the platform used, or the ecosystem of interest, SampleFreq can be used to measure plant frequency. As an example, vegetation point classification from nadir images using the closely-related program SamplePoint has been completed across many ecosystems with varying plant communities, including grasslands, temperate and tropical forests, alpine tundra, salt marshes and deserts (Guo et al., 2016, Parrish et al., 2017, Skipper et al., 2013, Goonan et al., 2009, Bacopoulos et al., 2018,Tabeni et al., 2014) as well as agricultural settings (Nielsen et al., 2015). We therefore think it reasonable to anticipate successful use of SampleFreq in these ecosystems, with the caveats that image resolution must be adequate for identification of the species of interest, and that the vegetation canopy allows determination of where a plant is rooted.

In any application, image quality is paramount to achieving desired results. Images must be well-focused with sharp image details present throughout the image. A small lens aperture (e.g. *f*/22) is desired because it achieves a greater depth of field within the image such that image elements in the canopy and on the ground are in sharp focus (a large image aperture will result in only portions of the image in sharp focus); however, in low light a larger aperture may be required for proper exposure. Fast shutter speeds (≤ 1/250 sec) are recommended to avoid image motion blur due to camera shake or vegetation movement due to wind. Maintaining optimal shutter speed and aperture for a proper exposure can usually be achieved through adjustment of the digital camera sensor's gain (denoted by ISO, a carryover from when photographic film's numerical light sensitivity was determined by the International Standards Organization). Proper exposure of both shadows and highlights is critical for image analysis, and techniques to mediate shadows include shading the plot or utilizing high dynamic range imagery (Booth et al., 2004; Cox and Booth, 2008). In all cases, capturing non-lossy 12–14 bit raw image files allows users to recover details in highlights and shadows through post-processing that are otherwise lost when capturing image files using 8-bit lossy JPG format. While capturing an image, the operator of the camera should position themselves to cast their shadow behind them, and not in the plot.

5. Conclusions

- SampleFreq can assist users in collecting accurate and precise data that has close agreement with data collected using standard field methods. We recommend using SampleFreq with nadir digital images as an advantageous alternative to standard methods for monitoring plant frequency.
- SampleFreq accuracy varied by frame size and we conclude that our speculated reasons may apply equally to standard field methods. We recommend users of standard field methods be aware of a possible effect of frame size on data-collection accuracy.
- SampleFreq classification accuracy decreased with user age within the range of 20–54 years. The reasons for this are not known.

CRediT authorship contribution statement

Samuel E. Cox: Conceptualization, Methodology, Investigation, Writing - original draft, Visualization, Project administration, Funding acquisition. **D. Terrance Booth:** Conceptualization, Methodology, Funding acquisition. **Robert D. Berryman:** Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank those who participated in the software testing: Spencer Allred, Travis Bargsten, Shayla Burnett, Jessica Crowder, Travis Decker, Amanda Gearhart, Mark Goertel, Larry Griffith, Ken Henke, Amanda Jones, Chris Keefe, Carmen Kennedy, Keith Klement, Marion Mahaffey, Brian Mealor, TJ Murry, Cheryl Newberry, Sherri Rowatt, Derek Stevenson, Sydney Thielke, and Mary Wilson. We also thank Mitch McClaran, Mort Kothman, and John Heywood for providing helpful reviews of early drafts. This project was supported by the United States Department of the Interior, Bureau of Land Management and the United States Department of Agriculture, Agricultural Research Service.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2020.106946. These data include Google maps of the most important areas described in this article.

References

Bacopoulos, P., Tritinger, A., Dix, N., 2018. Sea-level rise impact on salt marsh sustainability and migration for a subtropical estuary: GTMNERR (Guana Tolomato Matanzas National Estuarine Research Reserve). Environ. Model. Assess. https://doi.org/10.1007/s10666-018-9622-6.

Barbour, M.G., Burk, J.H., Pitts, W.D., 1987. Terrestrial Plant Ecology. Benjamin/Cummings Publishing Company, Menlo Park, CA, pp. 634p.

Bezryadin, S., Bourov, P., Ilinih, D., 2007. Brightness calculation in digital image processing. In: First International Symposium on Technologies for Digital Photo Fulfillment, pp. 10–15.

Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet i, 307–310.

Bland, J.M., Altman, D.G., 2003. Applying the right statistics: Analyses of measurement studies. Ultrasound Obsetet. Gynecol. 22, 85–93.

Booth, D.T., Cox, S.E., Louhaichi, M., Johnson, D.E., 2004. Lightweight camera stand for close-to-earth remote sensing. J. Range Manag. 57, 675–678.

Booth, D.T., Cox, S.E., Johnson, D.E., 2005. Influencing digital measurements of ground cover. Rangeland Ecol. Manage. 58, 598–604.

Booth, D.T., Cox, S.E., Berryman, R.D., 2006a. Precision measurements from very large scale aerial digital imagery. Environ. Monit. Assess. 112, 293–307. https://doi.org/10.1007/s10661-006-1070-0.

Booth, D.T., Cox, S.E., Berryman, R.D., 2006b. Point sampling digital imagery with 'SamplePoint'. Environ. Monit. Assess. 123, 97–108.

Booth, D.T., Cox, S.E., Meikle, T.W., Fitzgerald, C., 2006c. The accuracy of ground cover measurements. Rangeland Ecol. Manage. 59, 179–188. https://doi.org/10.1007/s00267-008-9110-x.

Booth, D.T., Cox, S.E., 2008. Image-based monitoring to measure ecological change on rangelands. Front. Ecol. Environ. 6 (4), 185–190.

Booth, D.T., Cox, S.E., 2009. Dual-camera, high-resolution aerial assessment of pipeline revegetation. Environ. Monit. Assess. https://doi.org/10.1007/s10661-008-0562-5.

Cagney, J., Cox, S.E., Booth, D.T., 2011. Comparison of point intercept and image analysis for monitoring rangeland transects. Rangeland Ecol. Manage. 64, 309–315. https://doi.org/10.2111/REM-D-10-00090.1.

Chase, A.F., Chase, D.Z., 2017. Detection of Maya Ruins by LiDAR: applications, case study, and issues. In: Masini, N., Soldovieri, F. (Eds.), Sensing the Past. Geotechnologies and the Environment. Springer DOI1007/978-3-319-50518-3_22.

Coulloudon, B., Eshelman, K., Gianola, J., Habich, N., Hughes, L., Johnson, C., Pellant, M., Podborny, P., Rasmussen, A., Robles, B., Shaver, P., Spehar, J., Willoughby, J., 1999. Sampling Vegetation Attributes. Interagency Technical Reference 1734-4. BLM National Business Center, Denver, CO, pp. 163.

Cox, S.E., Booth, D.T., 2008. Shadow attenuation with high dynamic range images. Environ. Monit. Assess. https://doi.org/10.1007/s10661-008-0577-y.

Curran, M.F., Cox, S.E., Robinson, T.J., Robertson, B.L., Strom, C.F., Stahl, P.D., 2020. Combining spatially balanced sampling, route optimization and remote sensing to assess biodiversity response to reclamation practices on semi-arid wellpads. Biodiversity. https://doi.org/10.1080/14888386.2020.1733085.
Curtis, J.T., McIntosh, R.P., 1950. The interrelations of certain analytic and synthetic phytosociological characters. Ecology 31 (3), 434–455.
Duniway, M.C., Karl, J.W., Scharder, S., Baquera, N., 2012. Rangeland and pasture monitoring: an approach to interpretation of high-resolution imagery focused on observer calibration for repeatability. Environ. Monit. Assess. 184, 3789–3804. https://doi.org/10.1007/s10661-011-2224-2.
Elzinga, C.L., Salzer, D.W., Willoughby, J.W., 1998. Measuring and Monitoring Plant Populations. BLM Technical Reference 1730-1. BLM National Business Center, Denver, CO, pp. 477.
Everitt, J.H., Yang, C., Racher, B.J., Britton, C.M., Davis, M.R., 2001. Remote sensing of redberry juniper in the Texas rolling plains. J. Range Manage. 54, 254–259.
Guo, Q., Hu, Z., Li, S., Yu, G., Sun, X., Li, L., Liang, N., Bai, W., 2016. Exogneous N addition enhances the responses of gross primary productivity to individual precipitation events in a temperate grassland. Sci. Rep. 6, 26901.
Goonan, K., Manning, R., van Riper, C.J., Monz, C., 2009. Managing recreation on mountain summits in northern forest region of Main, New Hampshire, New York and Vermont. Proceedings of the 2009 Northeastern Recreation Research Symposium. GTR-NRS-P-66.
Heywood, J.S., Debacker, M.D., 2007. Optimal sampling designs for monitoring plant frequency. Rangeland Ecol. Manage. 60, 426–434.
Hyder, D.N., Conrad, C.E., Tueller, P.T., Calvin, L.D., Poulton, C.E., Sneva, F.A., 1963. Frequency sampling in sagebrush-bunchgrass vegetation. Ecology 44 (4), 740–746.
Hyder, D.N., Bement, R.E., Remmenga, E.E., Terwilliger Jr., C., 1965. Frequency sampling of blue grama range. J. Range Manage. 18, 90–93.
Laliberte, A.S., Rango, A., Herrick, J.E., Fredrickson, E.L., Burkett, L., 2007. An object-based image analysis approach for determining fractional cover of senescent and green vegetation with digital plot photography. J. Arid Environ. 69, 1–14. https://doi.org/10.1016/j.jaridenv.2006.08.016.
Louhaichi, M., Borman, M.M., Johnson, D.E., 2001. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. Geocarto Int. 16 (1), 65–70. https://doi.org/10.1080/10106040108542184.
Mueller-Dombois, D., Ellenberg, H., 1974. Aims and Methods of Vegetation Ecology. John Wiley and Sons, New York, NY, USA, pp. 547.
Mosley, J.C., Bunting, S.C., Hironka, M., 1989. Quadra and sample sizes for frequency sampling mountain meadow vegetation. Great Basin Natural. 49 (2), 241–248.
Nielsen, D.C., Lyon, D.J., Hergert, G.W., Higgins, R.K., Holman, J.D., 2015. Cover crop biomass production and water use in the central Great Plains. Agron. J. 107, 2047–2058.
Parrish, M., Demarais, S.I., Ezell, A.W., Wigley, T.B., Jones, P.D., Riffell, S.K., 2017. Retained vegetation density of streamside management zones and stringers in southern intensively managed pine forests. For. Ecol. Manage. https://doi.org/10.1016/j.foreco.2017.04.024.
Poynton, C.A., 1996. A Technical Introduction to Digital Video. Wiley, New York, NY, pp. 320.
Roshier, D., Lee, S., Boreland, F., 1997. A digital technique for recording of plant population data in permanent plots. J. Range Manag. 50, 106–109.
SampleFreq. 2020. www.SamplePoint.org. Accessed 09 March 2020.
Schultz, A.M., Gibbens, R.P., Debano, L., 1961. Artificial populations for teaching and testing range techniques. J. Range Manage. 14, 236–242.
Skipper, B., Grisham, B., Kalyvaki, M., McGaughey, K., Mougey, K., Navarrete, L., Rondeau, R., Boal, C., Perry, G., 2013. Non-overlapping distributions of feral sheep (Ovis aries) and stout iguanas (Cyclura pinguis) on Guana Island, British Virgin Islands. IRC Rept. Amphib. 20 (1), 7–15.
Tabeni, S., Garibotti, I.A., Pissolito, C., Aranibar, J.N., 2014. Grazing effects on biological soil crusts and their interaction with shrubs and grasses in an arid rangeland. J. Veg. Sci. https://doi.org/10.1111/jvs.12204.
Vittoz, P., Guisan, A., 2007. How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. J. Veg. Sci. 18, 413–422.
Walker, B.H., 1970. An evaluation of eight methods of botanical analysis on grasslands in Rhodesia. J. Appl. Ecol. 7 (403–416), p415.
Weber, K.T., Chen, F., Booth, D.T., Raza, M., Serr, K., Gokhale, B., 2013. Comparing two ground-cover measurement methodologies for semiarid rangelands. Rangeland Ecol. Manage. 66, 82–87.