Capability of high-resolution RGB imagery to accurately document residue in row-crop fields

J.A. Lory, P. Upadhyay, T.A.P. Lagaunne, C. Spinka, R. Miller, G. Davis, and G.N. DeSouza

Abstract:

Sustaining agriculture and civilizations requires protecting topsoil from erosion. Residue on the soil surface dissipates energy in raindrops, runoff, and wind, reducing detachment and transport of soil particles. Our objective was to contrast transect residue estimates based on: i. in-field evaluation of 100 evenly-spaced points along a 15-m tape; ii. evaluation of imagery of the same tapes obtained at two ground sampling distances: 0.014 and 0.06 cm pixel^{-1}; and iii. a 100-point grid analysis of 50 contiguous 0.3-m X 0.2-m images (0.014 cm pixel^{-1}) adjacent to the tape. Data was collected from May through early July 2018 in row-crop fields in four Missouri counties: Audrain, Boone, Callaway, and Cooper. We used data from 60 15.2-m transects in 21 fields. Residue estimates based on field readings of the transect tape ranged from 6% to 99%. Reading images of tapes resulted in a mean decrease of 2.5 percentage points of residue. The bullseye grid point method documented tremendous variability along the transect (mean transect sd=12.4) and a substantially lower estimate of residue cover compared to field tape readings (mean difference -10.3 percentage points). Consistent with a systematic bias to over-estimate residue with the tape transect method, the difference increased with increasing residue for residue <50%, was more pronounced with soybean (*Glycine max* [L.] Merr.) than corn (*Zea mays* L.), and the difference could be eliminated by rereading tapes with scrupulous attention to the details of transect methodology. Image-based methods have the potential to improve the accuracy of residue estimates by facilitating methods less prone to reader bias such as the bullseye grid approach. Additionally, to improve accuracy, image-based protocols

should be integrated into training and quality control assurance protocols of Natural Resources Conservation Service assessments of residue using transect methods.

Key Words:

line-transect method—bullseye grid-point method—ground truth—erosion

Sustaining agriculture and civilizations requires protecting topsoil from erosion (Montgomery 2007a; Pennock 2019; Weltz et al. 2020). The soil A horizon, compared to other diagnostic horizons, is best suited to support crop production due to factors such as highest soil organic matter, enhanced soil aggregate stability and soil porosity, and greatest nutrient-supplying power. In many soils, erosion of the A horizon reduces the depth of soil available to support crop production (Pennock 2019). Soil erosion rates continue to exceed soil formation rates in row-cropped fields in the US and globally (Montgomery 2007b; Cruse et al. 2013; Pennock 2019).

Surface residue in row-crop fields is defined as non-photosynthetic material on the soil surface and is primarily plant-derived material retained in the field from previous crops. Residue cover has a beneficial effect on a wide range of agro-ecological functions of the soil including reduced weed emergence, runoff, evaporation, soil sealing, and soil crusting; and increased soil carbon sequestration and soil fauna activity (Lal 2009; Ranaivoson et al. 2017; Cherubin et al. 2018). These benefits of residue provide protection to the soil from wind and water erosion in row-crop systems. Increasing the amount of residue on the soil surface increases many of the beneficial effects of residue including reducing erosion potential. Coverage of the soil surface with residue correlates with the degree of protection (Lal 2009; Ranaivoson et al. 2017; Cherubin et al. 2018) and serves as an important input for models used to estimate soil loss from agricultural fields such as WEPP, RUSLE and RUSLE2 (Flanagan et al. 2007; Dabney et al. 2011; Weltz et al. 2020).

The USDA Natural Resources Conservation Service (NRCS) evaluates residue cover on thousands of fields annually as part of documenting farmer compliance with Food Security Act regulations for highly erodible land (HEL; USC 1985). The standard approach used by NRCS personnel is a line-transect method, reading 100 points evenly spaced on 15.2-m or 30.4-m tape laid at a 45-degree angle to the direction of farming. The National Agronomy Manual (USDA NRCS 2011) details the protocols for the transect method emphasizing the need to view transect points from directly above and defining a hit as residue touching a pin-head sized transect point with a surface coverage diameter greater than or equal to 2.4 mm. The stated accuracy of residue estimates using the tape transect method is $\pm$15% of the mean field residue cover (95% confidence interval) when averaging results from five tapes and $\pm$32% of the field mean when averaging three tapes (USDA NRCS 2011).

There is limited published research on the precision and accuracy of the transect method for estimating residue cover. The most detailed assessment of field performance of the transect method was done by Morrison Jr. et al. (1995). The observed variation among trained personnel averaged 8.8 percentage points with a maximum range of 34% cover across all readings and could be as high as 14 percentage points with multiple readers reading the same device. They concluded reader variation was the dominant source of variation in transect methods tested and reader variation is likely common in field assessments. They also noted that variability along a transect was routinely at the same scale as variation among readers.

Multiple observers have documented a bias of the transect method to over-estimate residue cover compared to other methods (Laflen et al. 1981; Laamrani et al. 2017). This discrepancy typically is attributed to the tendency of readers to deviate from a 90-degree downward view angle of the transect point to include close calls. There is variability among descriptions of the transect method in applied and peer-reviewed publications (e.g. Shelton et al. 1990; Morrison Jr. et al. [1993, 1995]; Eck and Brown 2004; USDA NRCS 2011). All emphasized the importance of looking downwards at a constant angle. They are inconsistent in emphasizing the need to read a pinpoint transect point. Recommended minimum residue size ranges from 2.4 to $\leq$3.2 mm and some recommend carrying an appropriate-diameter dowel to improve judgement of minimum residue diameter.

Grid methods to estimate residue cover were first applied to images projected onto white screens with dots for identifying assessment points (Laflen et al. 1981, Morrison Jr. et al. 1993). Subsequently digital methods allowed superimposing a grid on a computer screen (e.g. Laamrani et al. 2017). These systems, while preserving or exceeding the number of points read with line-transect methods, sampled small areas relative to the transect methods they were compared to, typically assessing one to five images of less than 1 m^2.

In its purest form, residue sampling on a line transect or on an image-based grid should follow a binomial distribution reflecting the yes-no nature of the determination at each sampling point. However, evidence that readers can vary in their determinations suggest that variability will be greater than predicted by the binomial distribution.

Our objective was to determine the comparative precision and accuracy of three strategies for estimating residue cover based on the standard NRCS transect method. Specifically, our objective was to contrast residue estimates based on: i. infield evaluation of 100 evenly-spaced points along a 15-m tape; ii. evaluation of imagery of the same tapes obtained at two ground sampling distances (GSDs): 0.014 and 0.06 cm pixel^{-1}; and iii. a 100-point grid analysis of 50 contiguous images (0.014 pixel cm^{-1}) adjacent to the tape. Our goal is to use this information to inform the best ways to document field tape readings taken in the field by NRCS personnel and to obtain accurate estimates of residue cover for images used as ground truth in the development and testing of algorithms estimating residue coverage from imagery.

Methods and Materials:

Data acquisition. Data was collected from May through early July 2018 in fields planted to corn (*Zea mays* L.) or soybean (*Glycine max* [L.] Merr.) prior to corn reaching V4 and soybean reaching V3 stage. Personnel from NRCS identified farmer fields and obtained permission for the project team to access fields in four Missouri counties: Audrain, Boone, Callaway, and Cooper (table 1). We used data from 60 tapes in 21 fields. All fields were classified as HEL by

NRCS. Location details on residue type, date of sampling, planted crop, crop stage of growth on day of sampling, and number of tapes per site is provided in table 1. Additional sites were visited but not included due to advanced stage of a growing crop or poor weed control. Three tapes were eliminated from the initial analysis of reading field tapes because they were read by two instead of three readers in the field.

At each location, project staff and NRCS personnel placed three to five 15.2-m tapes at a 45 degree angle to the planted row direction in accordance with the line-transect method for estimating residue cover (USDA NRCS 2011). In contrast to standard NRCS assessment approach, tapes were placed to capture a range of conditions in the field instead of attempting to obtain three to five estimates of "representative" residue conditions. This was justified because our objectives did not include estimating residue cover at the field scale. A side of the tape was selected for reading ensuring an acute sun angle to minimize shadows from the tape interfering with tape reading. Images were then obtained from two GSDs. The first set of images was obtained 1.0-m above the ground on a tripod-mounted Canon EOS Rebel T6i Digital SLR camera (Canon USA, Melville, NY) with a 24-mm lens and 24.2 megapixel resolution which generated an image of 6,000 X 4,000 pixels. Estimated GSD for these images was 0.014 cm pixel^{-1}. Typically, 51 images were obtained per tape by moving the tripod 0.30 m between images. The second set of images was obtained by a pilot flying a DJI Phantom 4 PRO um-manned aerial vehicle (UAV; DJI, Shenzhen, China) maintaining a camera elevation of 2-m above the ground. The stock camera had a 24-mm lens and 20-megapixel resolution which generated an image of 5,472 X 3,648 pixels. Estimated GSD of the 2-m images was 0.06 cm pixel^{-1}. Typically, 17 images were taken along each tape. To minimize parallax effects, all images were taken perpendicular to the soil surface. All UAV-based imagery was done in duplicate with ISO set to 100.

After imagery was obtained, three trained personnel obtained 100 readings at the same marks on the tape at 0.152-m intervals. All tape readers recorded their assessment of the presence or absence of residue at each point. The percent residue tape estimate was then calculated as the number of points with residue. The definition of residue was based on the NRCS Agronomy Manual (USDA NRCS 2011) where a point was considered as positive for residue if, when looking directly down on the tape, residue touched a set point of the mark and had a soil surface coverage of $\geq$2.4 mm in all directions. If green plants interfered with reading the tape, they were moved aside to allow reading of residue below the plant leaves.

Image-based tape reading. The 0.014 and 0.06 cm pixel^{-1} images were assessed for residue cover following a protocol similar to reading the tape in the field and using, to the degree possible, the same points on the tape for assessment as were used in the field. Each tape was read by three trained technicians. From the images, 100 readings were obtained at marks on the tape at 0.152-m intervals and the reading at each point was recorded for their assessment of the presence or absence of residue. If the point on the tape was obscured by a plant, the reader shifted 0.076 m further along the tape to make the reading; if that location was covered, they shifted 0.076-m backwards along the tape. For each point, the reader reviewed overlapping images and selected the image that had the least parallax effect on the reading (image where reading point was most central to the image). As with the field readings of the tape, percent residue cover was then counted as the number of read points with residue.

Image-based residue estimate using the bullseye grid method. Using the software package Adobe Photoshop (Adobe, San Jose CA) and starting at the 0 point on the tape on the side of the tape used for reading, a 2,400 X 1,600 pixel area (approximately 0.305-m X 0.020-m surface area of the soil) was cut out and saved as a unique image from the 0.014 cm pixel^{-1} images. This process was repeated for each 0.305-m section of the 15.2-m tape resulting in 50 sequential images of residue cover contiguous to, but not including the tape, on the side of the tape used for reading. For each 0.305-m segment of the tape, the image was obtained from the photo where that section of the tape was most central.

Image-wise estimates of residue cover in each of these images was then obtained using a novel grid method. Images were imported into the software PowerPoint (Microsoft, Redmond WA) and a 50- (5X10 points) or 100-point grid (10 X 10 points) was superimposed on the image with the point represented by a period (Calibri font, 12 point; figure 1). Additionally, the point was centered in a circle with a diameter of the minimum size of residue (2.4 mm); actual circle size was adjusted to the correct size on the image based on the GSD of a pixel (figure 1). Residue was defined as when residue touched the point and the residue filled 50% or more of the circle. Contributing residue had to touch the center point directly or touch residue touching the center point. When live plant leaf covered the soil at a grid point, the reader randomly selected a point near grid point and assessed that point. Four readers contributed to grid reading of 50 images per tape for 60 tapes (3000 images). Two readers were randomly assigned each tape, and each read a different 50-point grid on each of the 50 images from a tape (100 points per image). From each tape, two images were randomly selected for quality control (n=120). On these images the assigned readers read a second 50 point grid and the other two readers read a 100-point grid resulting in four readings of the image each with n=100 and a total points read on the image of n=400.

Statistical analysis. Accuracy of readings was assessed based on expected distribution assuming a binomial distribution. One standard deviation estimated as:

$$1\ standard\ deviation = \sqrt{((p\ X\ (1-p))/n} \qquad \text{eq. 1}$$

where *p* equals the fraction of residue and n equals the number of points read. When assessing situations where reader values contribute directly to means for comparison, the standard deviation was calculated as:

$$1\ standard\ deviation = \sqrt{((p\ X\ (1-p))\ /\ n}\ \ X\ \sqrt{(r-1)/r} \qquad \text{eq. 2}$$

where *p* equals the fraction of residue, n equals the number of points read, and *r* equals the number of readers. An estimate of the expected percentage of the quality control residue data (n=120) within one standard deviation of the mean, based on a binomial distribution, was determined empirically by modeling each point individually based on n=100 readings and 1000 iterations.

All statistical analysis was done using the Statistical Analysis System (SAS ver. 9.4, SAS Institute Inc., Cary NC).

Results and Discussion:

Precision of in-field tape readings. Residue estimates based on field readings of the transect tape ranged from 6% to 99% for the 57 tapes. Although readers were reading the same points on the tape there was significant variability in residue estimates among readers (figure 2a). This indicated readers did not always agree on the determination of residue at a given point on the tape. When evaluated by tape, points where all three readers agreed on the presence or absence of residue ranged from 54% to 98% (average 80%) with highest agreement associated with low and high residue tapes. There was evidence of bias in the readings at points where there was disagreement among readers. For tapes with a mean reading below 25% residue (n=15), 39% of the time readers disagreed, two of the readers chose residue. In contrast, for tapes with a mean reading greater than 75% residue (n=16), 72% of the time readers disagreed, two of the readers chose residue. For tapes in the 25% to 75% range, there was no apparent bias; the odds of two readers agreeing on residue or no residue was 49%. This implied tape readers had a bias to reading no residue on low-residue tapes and a greater bias to read residue on high-residue tapes. We also compared readers. In the 2018 data set, an NRCS representative was one of the readers on all tapes. We compared the NRCS reader with the other trained readers who had at least 15 overlapping readings (n=4 readers). Two of these readers averaged significantly lower residue estimates than the NRCS representative (paired t-test, $P<0.01$).

These results document that field reading of tapes can include significant error from differences in interpretation of residue by the reader and that the errors in low and high residue tapes may not be random. These precision errors will result in greater than expected error compared to binomial distribution and should be considered when estimating accuracy of tape reading in the field.

Precision of image-based tape readings. Reading the tape from images decreased precision compared to field reading of the tapes (figure 2). Across all tapes, there was a reader bias towards lower estimates of residue for readings taken from an image, -2.2 percentage units for the 0.014 cm pixel^{-1} images and -2.9 percentage units for the 0.06 cm pixel^{-1}. Compared to field readings of the tape, reading the tape from the images resulted in more outlier readings and a trend towards a wider distribution of the data, as quantified by the box and whisker plots, particularly for reading with 25% residue cover or more (figure 2). There was little evidence that the bias of reading tapes from images correlated with residue cover (P>0.1).

To further understand how readings differed from images compared to the field we assessed all 5,700 point tape readings (57 tapes X 100 points) individually, across the three readers for each tape. There were clear similarities in reader patterns with both field- and image-based tape readings. As with field tape reading, there was a bias for readers to estimate no residue for tapes with less than 25% residue cover and a bias to estimate residue for tapes with greater than 75% residue cover. The number of readings where all readers agreed was similar for the field (79%) and 0.014 cm pixel^{-1} images (81%) declining to 75% for 0.06 cm pixel^{-1} images. While the number of points where all three readers agreed was similar across reading approaches, the specific readings where they agreed were not necessarily the same for the field tape and the image-based readings. For example, when comparing reading in the field and 0.014 cm $pixel^{-1}$ images, in the field all three readers agreed there was no residue for 2,119 readings; in the 0.014 cm $pixel^{-1}$ images, all three readers agreed there was no residue 87% of the time they previously agreed no residue in the field. Similarly, for the 2,435 times three readers agreed there was residue in the field, they agreed 85% of the time when reading these points from the 0.014 cm $pixel^{-1}$ images. For 0.06 cm $pixel^{-1}$ images, performance dropped to 83% for no-residue points in the field and to 76% for residue points in the field.

One possible explanation for changed readings from the field to reading imagery could be the parallax effect. Due to the parallax effect, tape readings off the center of the image will not align with the same point on the ground as a field tape read from directly above. This could lead to a reader correctly assigning different results for the same point on a tape from the image compared to the field. This type of error is a particular class of precision error in that while it could lead to a different reading for the tape compared to the field reading, it is still a correct reading for the tape image. Comparing all 5,700 readings in the field, we only observed 55 times all readers agreed with one reading in the field and another reading from the tape with 0.014 cm-$pixel^{-1}$ images and 92 times for 0.06 cm $pixel^{-1}$ images. This implied the parallax effect altering the outcome of tape reading from images of tapes was not that important.

Another potential source of error is shadows obfuscating some readings on an image so the reader must guess based on what's visible. This problem is likely to be more of an issue with more residue. Shade from the tape also can be an issue. We minimized shade from the tape by reading the side of the tape facing the direction of the sun. It is difficult to quantify the role of shading, but it could explain some of the added variability with image reading of tapes. Consistent with this hypothesis is that variability in image-based readings was not evident in low residue tapes (figure 2).

We saw more outliers with image-based readings compared to field tape reading (figure 2). Some of these may be due to careless management of data entry, more likely in the monotony of reading tapes and entering data directly on the computer compared to filling in data sheets in the field. Reading an image can have benefits. It allows for more time and the ability to zoom in to see clearly. It also provides an opportunity to document readings and increase the number of readings without increasing labor time in the field.

In summary, these results suggested reading images of tapes will result in a small downward bias in residue estimates compared to field reading the tape. Otherwise, the same biases seen with field reading of tapes also existed in reading tapes from images. Reading a tape from an image did increase variability in reader outcomes, documented as up to an 80% increase in the standard deviation when reading 100 points. We concluded that reading tapes from images is defendable, but two steps are needed to ensure similar precision as with field reading of tapes. First, the number of readings should be increased to account for the larger error seen with image-read tapes (figure 2). Additionally, having two readers read 50% of the points, with unique points assigned to each reader, will increase precision by reducing the effect of reader bias on the outcome.

Image-based grid estimates of residue. Our first objective was to assess precision of reader interpretation of the bullseye target-style grid points (figure 1). For 10 images, four readers assessing the same 100-point grid agreed on 88% (SD=3.5) of the points (table 2). When they disagreed, almost 50% of the time it was one reader selecting no residue when the other three readers selected residue (table 2). In contrast, when the opposite was true (three readers selected no-reside and one selected residue), it represented a significantly lower percentage of the disagreements (27%, $P<0.01$). This shift is consistent with the hypothesis that the outcome of disagreements is not random. If the disagreements were random, 50%-50% disagreements would be most common, and the 75%-25% disagreements would also be evenly split. Our results imply there is a slight favoritism to select residue where it is a close call between residue and no residue.

These results were a 50% improvement over the precision of readers evaluating 100 points on a tape in the field. For tapes with the same range of residue (19% to 68%, n=32) the mean number of points with disagreement was 25% (SD=8.9) in the field or more than double what was found on gridded images. The combination of clarity provided by the high-resolution image plus the target-style point for evaluating residue cover led to improved reader precision. We concluded that the target-style grid system improved precision of readings among readers but there was still some evidence of a bias towards selecting residue in marginal situations.

Our second objective was to assess the precision of the quality control images. For quality control, we selected two images from each tape (n=120) to be read by all four readers, 100 points each using a unique target grid for each reader randomly applied to the image. Comparison of the four 100-point means (figure 3a) documented variation among readers. Specifically, in images with high residue cover (>80%) there was evidence that one reader consistently estimated lower values than the other readers. The most extreme outliers were from a series of tapes that had heavy small-grain straw residue. Reviewing the readings, one reader was more likely to select no-residue for partially shaded reading points than the other three readers. Paired t-tests among the four readers also indicated some small but systematic differences. Across all images, mean differences ranged from -0.7 to 2.9 percentage units and differences were significant in three of six comparisons among readers ($P \leq 0.01$). The bullseye method did not completely eliminate reader bias.

In figure 3a the dotted lines represent the theoretical expectation of one standard deviation for a binomial distribution based on 100 readings (eq. 2). For our distribution of residue, if the variability conformed with the binomial distribution, 82% of the data point should have fallen within that range; actual compliance was 55%. There is an expectation that the error rate will be higher than predicted by the binomial function because of precision errors documented above in reading images and tapes. To minimize reader bias, we tested a system of using two unique 50-point grids, one for each of two readers (n=100) for each image (figure 3b). Using this system, 61% of the data points fell within one standard deviation of the four-reader mean. This result is consistent with the expectation of precision error for 12% of the readings plus the expected precision of an n=100 binomial distribution. Based on these results, we concluded expected precision of our by-image estimate of residue using the bullseye method is likely to be more precise than field-based readings of a tape. It is important to note that these precision estimates do not account for any inherent bias across all readers in reading grids on images or tapes in the field.

Tape-wise estimates of residue cover based on image-based bullseye grids. Variability within a tape, based on image-wise residue estimates were highly variable (figure 4). The image-wise estimates documented residue along the transect in 0.3-m increments. For tapes with a field reading between 25% and 75% residue cover (n=23), the average standard deviation was 18%. Variability declined, as expected, near the extremes of high and low residue (figure 4). The high amount of variation along a tape reflected variations in cover associated with management practices such as systematic patterns of higher incorporation associated with the planter and random variations associated with patterns of residue dispersal from the previous crop (figure 5).

Compared to field estimates of residue cover, the image-based mean residue cover systematically estimated lower residue cover except for residue $\geq$85% (figure 6). Deviations from the one-to-one line could be large; between field tape readings of 25% and 75% the mean difference was 14% with a range of 6% to 23% (figure 6) with image-based means lower than field-based means. In that range, the mean deviation for corn (n=19) was 12.9% and the mean deviation for soybean (n=9) was 16.0%. When corn readings were restricted to the range of soybean residue levels (n=6) the mean corn difference was 11.9%. This documented a systematic difference between corn and soybean residue, but this difference was small compared to the systematic difference between field-based tape readings and the gridded image-based tape estimates.

As documented earlier, each method has precision biases. Greater precision error was associated with the field-tape reading resulting in over-estimating residue in high-residue situations and under-estimating residue in low-residue situations. These errors shift field-tape readings to the left along the *x*-axis for the lowest residue estimates and to the right along the *x*-axis for the highest residue estimates (figure 6). Correcting for these shifts could bring high-residue and low-residue estimates from images and tapes closer to the 1:1 line but will have little effect on points where differences are greatest between 25% and 75% residue cover. Precision errors were less for image-wise bullseye readings and slightly favored reading residue over no residue. Correcting for that bias would shift image-based errors slightly down on the *y*-axis increasing the distance of points from the one-to-one line. Consequently, biases documented in the previous sections do not account for the deviation of residue estimates by the two methods from the one-to-one line except, perhaps, at the highest and lowest residue levels.

With field tapes there are at least three additional possible sources of error that will consistently lead to a systematic increase in the residue estimate. The first mechanism with potential to create higher residue cover estimates from field tape readings is related to the ability of the tape reader to adjust their angle looking down on the field tape (e.g. Laflen et al. 1981). If a reader has any bias to adjust their visual angle they can increase the estimate of residue cover. Reading from an image eliminates the ability to adjust the angle of assessment eliminating this error. The previous section compared field-read tapes with image-based readings of tapes and suggested that this type of bias in field readings does occur. Image-based readings of tapes had a bias of more than two percentage points lower residue estimates compared to field readings. This error likely contributed to the differential between field-tape estimates and the image-based estimates but evidence from tape reading implies it does not fully explain the differences.

A second potential error is how readers interpret the tape mark in field readings. When the reader looks down on a measuring tape such as we used in the field, the marks are typically one third the functional diameter of residue. Readers are instructed in official NRCS methods to focus on an area the size of the end of a pin to determine if residue is touching the transect point (USDA NRCS 2011). To be done correctly, the reader must focus on the same pin-point location within the larger mark on the tape. Accepting readings from contacts anywhere else along the mark on the tape will lead to a systematic over-estimate of residue cover from the field transect method by expanding the size of the contact area considered. A third potential source of upwards bias are choices the reader makes when residue is close to the minimum dimensions for residue. Residue may touch the transect point but not meet the size criteria in all dimensions unless the reader looks past the reading point or peeks under the tape to see if the residue meets criteria in this obscured direction.

It is hard to establish reader bias to select residue in marginal cases. However, the pattern of error in figure 6 suggests that this bias could be a significant contributor. This type of error should be smaller at low residue cover and increase as residue increases as more residue results in more opportunity for the error to occur. However, as we reach high levels of residue, the error should decrease or disappear as the heavy residue cover eliminates the potential for the error. This generally agrees with the pattern of differences in figure 6.

Another prediction is that the error may be more prevalent in soybean residue than in corn residue because it typically is thinner and less blocky creating more opportunities for marginal situations. We previously observed that differences between the image-based means and the field tape were greater with soybean than corn. Additionally, regression analysis documented a positive relationship between residue cover and the difference between field readings and image readings for residue less than 50% (figure 7). These results are consistent with a reader bias to select residue in marginal situations.

To test if a lack of rigor applied to the details of the tape reading method contributed to a systematic over-estimate of residue cover, we selected eight tapes and two readers. After careful training on the size of residue and focus on a specific pinpoint on the tape mark, we split the work of residue readings along the tape using images from the 0.014 pixel cm^{-1}, assessing residue every 2.5 cm (n=600; figure 8). Using the results of this assessment as the standard (*x*-axis), the estimated residue amounts dropped significantly from the field readings of the tape.

Furthermore, regression analysis indicated the gridded image tape means and the 100-point reread of the tape images were not statistically different than the 1:1 line. This result suggested that field readings were not accurate because of systematic bias leading to over-estimation of residue and the image grid method was providing accurate residue estimates.

In summary, there is definitive evidence that readings of tapes in the field were higher than the estimates derived from the mean of the 50 images along the tape for all but the highest residue situations. Our analysis strongly supports that tape reading based on the gridded image-based means are more accurate than the field readings. We suggest three mechanisms that led to reader bias resulting in higher estimates of residue in the field. While we cannot definitively prove which estimate is more accurate, the information we provide clearly documents methods that reduce potential bias, such as the bullseye grids on images and more rigorous application of the rules of while reading of images of a tape, similarly result in lower residue estimates. This strongly suggests the gridded-image based tape means were the more accurate estimate of residue cover.

Summary and Conclusions:

Results of this study have implications for the expected precision of field readings of tapes. We documented a precision error in field tape readings that will result in lower accuracy in readings than is expected with a binomial distribution. The binomial distribution assumes definitive yes/no results for predicting accuracy. For residue, there is an added error because readers can differ in their assessment of residue versus not residue at an average of 20% of points along a tape. Importantly, there is evidence that there are differences among readers and differences based on residue cover, favoring residue in high-residue situations and no residue in low-residue situations. We are not aware of any other paper that documents comparisons of multiple readers at each point along a 100-point transect in the field. However, differences among readers are frequently mentioned as a challenge for residue estimates derived by the transect tape method (e.g. Laflen et al. 1981; Morrison Jr. et al. 1995; Laamrani et al. 2017).

Our research also documented a high likelihood of systematic over-estimation of residue reading field tapes. We compared field residue estimates across 60 tapes from 20 fields using 50 images per tape and 100 points per images using the bullseye method (figure 6). Our delta values ranged from -5 % to 25% (mean 10%). Previous work concluded that estimating residue cover using grids on images decreased residue estimates compared to field readings of tapes (e.g. Laflen et al. 1981; Laamrani et al. 2017). Laflen et al. (1981) documented that field tape readings over-estimated residue cover by 6% to 10% compared to grid methods. In contrast to our comparison, they compared 10 field means of eight estimates of residue (eight grid-based images versus eight tape transects; residue range approximately 20 to 80% cover). The comparison of field means will mask differences among the methods compared to our comparison of by-tape values. Similarly, Laamrani et al. (2017) compared multiple estimates of residue using gridded images and line transects (five of each from 225 m^2 areas in 18 fields). Like Laamrani et al. (2017), we had high correlation values between the two methods (r^2= 0.89 compared to r^2=0.83) but our offset was greater (intercept = -10.9 versus -0.7). Their image-based values were based on one 0.75 m X 1.0 m quadrat set diagonally on the tape (1.25 m of the 15.2-m tape). The degree of variability we documented along a tape in figure 4 suggests the Laamrani et al. (2017) estimate of residue had low precision masking differences we observed.

Laflen et al (1981) attributed the bias of higher residue estimates in field readings of tapes to the tendency of readers to adjust their angle of sight if there was residue near the mark. Comparing field-based and image-based reading of transect tapes (figure 2) suggested this type of error only explained a two- to three-percentage point increase in residue estimates. The error of moving your angle of vision is not possible when reading an image. Our conclusion is that other forms of tape reader bias, possible in both image and field readings of tapes, can lead to upward bias. Possible contributors include expanding the size of the transect point and including residue of marginal size.

Our team of readers included two NRCS personnel plus members of our research team trained by NRCS personnel. There was no indication that NRCS readers systematically had lower residue estimates than other readers. Additionally, the close agreement of field-based and image-based readings of tapes (figure 2) suggested that project readers used similar criteria for field- and image-based readings of tapes. This implies that our field team learned the system, as trained by the NRCS personnel, and reliably applied it both in the field and when reading images of the tapes. However, when two readers re-read eight images of tapes after reviewing how to more rigorously apply the tape reading protocol, estimates of residue cover decreased and aligned with results of the grid methodology (figure 8). This result clearly suggests that the

NRCS and project team in this project had a bias when reading tapes that led to a systematic over estimation of residue. Our contention is that this bias is not unique to this project. The consistent result that using grids on images decreases residue estimates compared to field readings of tapes (e.g. Laflen et al. 1981; Laamrani et al. 2017) is consistent with this conclusion. The high degree of variability among readers consistently reported in assessments of residue estimates in the field (e.g. Laflen et al. 1981; Morrison Jr. et al. 1995; Laamrani et al. 2017) is also consistent with challenges in uniformly and accurately applying the tape reading protocol among readers.

Detailed protocols such as the Agronomy Handbook (USDA NRCS 2011) emphasize the importance of limiting readings to a pinhead sized point, consistently use the same angle looking down on the tape, and discard as "no-residue" questionable points. This documents that writers of the protocol understood the potential for these errors to bias readings, but the protocols do not emphasize how sensitive the methods are to failure to carefully follow these rules. The reality is that it is very difficult to document if readers are systematically adjusting their readings leading to bias. Personnel are largely trained by experienced readers and their success at learning the system is largely judged by matching results of an experienced reader. There is no internal standard that can be used to "ground truth" field readings of tapes. In this study we were unaware of the systematic error of our tape readings until the results of the grid-readings of images suggested there was a problem.

If our results represent what is common in field assessments of residue, these results have significant practical implications. Accounting for the upwards bias, particularly in soybean fields (figure 7), could have substantial impact on our assessment of these low residue systems. Accounting for the systematic overestimating residue documented in figure 7 could result in the reclassification of fields, particularly soybean fields, as having insufficient residue cover.

We concluded that transect tape measurements are not as precise as expected using binomial statistics and are subject to significant reader biases through small adjustments in reader technique. Such biases are likely leading to significant overestimates of residue cover particularly in soybean residue and are impossible to quantify using standard field methods used by NRCS.

The bullseye method applied to images limits the likelihood of some biases potentially affecting field reading of transect tapes. The small point size is consistent with the theoretical objective to assess residue at pinpoint. The surrounding circle with the radius of the minimum dimension of residue facilitates correctly determining if a point has sufficient residue coverage to qualify as residue. In our study, readers using the grid method reduced by nearly 50% the number of points where readers disagreed. The biggest drawback to image-based reading were shadows in the imagery that can make some determinations difficult. We limited the impact of differences among readers by assigning two readers to every image having each read a different 50-point grid on an image. Quality control documented that while these readings still had variability greater than predicted by binomial distribution, the results were substantial improved compared to single reader results (figure 3) and field-based readings. Additionally, our results indicated that tape means calculated from mean of the bullseye grids for 50 contiguous images along the tape were as accurate or more accurate than any other estimate of residue cover for a transect. This research also documented tremendous variability in residue cover at the 0.3-m X 0.2-m scale evaluated along each transect (Figs. 4 and 5). Across all 60 tapes the mean standard deviation of the 50 contributing images was 12.4 and for images between 25% and 75% residue cover (n=22), the mean standard deviation was 18.2.

High-resolution photography ($\leq$0.014 cm pixel^{-1}) can be used to document residue cover. Image-based methods have the potential to improve the accuracy of residue estimates by facilitating methods less prone to reader bias such as the bullseye grid approach. Image-based protocols should be integrated into the training and quality control assurance protocols of NRCS for field-readings using the transect method. Documenting test transects in the field with photography and then using grid methods outlined in this paper will provide an unbiased estimate of residue cover for comparison to field readings. Currently there is no such unbiased ground truth method available for quality control. Finally, the grid methods outlined in this paper can provide unbiased estimates of residue cover for images used in machine learning which requires accurate assignment of residue cover to imagery with a known precision as the basis for developing algorithms to accurately assess residue cover.

Acknowledgements:

We appreciate the work of research specialists David Kleinsorge, Krystal Burkett-Tysdale and Theresa Musket; and student workers Ahmed Krgo, Isaac Shee, Robert Phillips, and Dewi Kharismawati. We also thank Missouri USDA NRCS for financial and logistical support.

References:

Cherubin, M.R., D.M.S Oliveira, B.J. Feigl, L.G. Pimentel, I.P. Lisboa, M.R. Gmach, L.L. Varanda, M.C. Morais, L.S. Satiro, G.V. Popin, S.R. de Paiva, A.K.B. dos Santos, A.L.S. de Vasconcelos, P.L.A. de Melo, C.E.P. Cerri and C.C. Cerri. 2018. Crop residue harvest for bioenergy production and its implications on soil functioning and plant growth: A review. Scientia Agricola 75:255-272. doi 10.1590/1678-992X-2016-0459.

Cruse, R.M., S. Lee, T.E. Fenton, E. Wang, and J. Laflen. 2013. Soil renewal and sustainability. *In* Principles of Sustainable Soil Management in Agroecosystems, eds. R. Lal and B.A. Stewart, 477-500. CRC Press, Boca Raton, LA.

Dabney, S.M., D.C. Yoder, D.A.N. Viera and R.L. Bingner. 2011. Enhancing RUSLE to include runoff-driven phenomena. Hydrological Processes 25:1373-1390. doi: 10.1002/hyp.7897.

Eck, K.J., and D.E. Brown. 2004. Estimating corn and soybean residue cover. Agronomy Guide AY-269-W. Rev. June 2004. Purdue University Cooperative Extension Services, West Lafayette, IN.

Flanagan, D.C., J.E. Gilly and T.G. Franti. 2007. Water erosion prediction project (WEPP): Development history, model capabilities and future enhancements. Transactions of the

American Society of Agricultural and Biological Engineers 50:1603-1612. doi: 10.1303/2013.23968.

Laflen, J.M., M. Amemiya, and E.A. Hintz. 1981. Measuring crop residue cover. Journal of Soil and Water Conservation 36: 341-343.

Laamrani, A., P. Joosse, and N. Feisthauer. 2017. Determining the number of measurements required to estimate crop residue cover by different method. Journal of Soil and Water Conservation 72:471-479. doi: 10.2489/jswc75.5.471.

Lal, R. 2009. Soil quality impacts of residue removal for bioethanol production. Soil & Tillage Research 102:233-241. doi 10.1016/j.still.2008.07.003.

Montgomery, D. 2007a. Dirt: The Erosion of Civilizations. University of California Press, Oakland, CA.

Montgomery, D. 2007b. Soil erosion and agricultural sustainability. Proceedings of the National Academy of Sciences (PNAS) 104:13268-13272. doi 10.1073/pnas.0611508104.

Morrison Jr., J.E., C. Huang, D.T. Lightle and C.S.T. Daughtry. 1993. Residue measurement techniques. Journal of Soil and Water Conservation 48:478-483.

Morrison Jr., J.E., J. Lemunyon and H.C. Bogusch Jr. 1995. Sources of variation and performance of nine devices when measuring percent residue cover. Transactions of the American Society of Agricultural Engineers 38:521-529.

Pennock, D., 2019. Soil erosion: The greatest challenge to sustainable soil management. 100 p. FAO, Rome. License: CC BY-NC-SA 3.0 IGO.

Ranaivoson, L., K. Naudin, A. Ripoche, F. Affholder, L. Rabeharisoa, and M. Corbeels. 2017. Agro-ecological functions of crop residues under conservation agriculture. A review. Agronomy for Sustainable Development 37:26. doi.10.1007/s13593-017-0432-z.

Shelton, D.P., E.C. Dickey, P.J. Jasa, R. Kanable, R. Kanable, and S.S. Krotz. 1990. Using the Line-Transect Method to Estimate Percent Residue Cover. NebGuide G90-981, Publication 289, June 1990. Cooperative Extension, University of Nebraska, Lincoln, NE.

USC (United States Code). 1985. The food security act of 1985. United States Code 16:3801-3862.

USDA NRCS (USDA Natural Resources Conservation Service). 2011. Crop Residue. *In* National Agronomy Manual, 190-V-NAM, 4th Ed., Feb. 2011, Subpart 503E, 503-126 - 503-128. USDA, Washington, DC.

Weltz, M.A., C. Huang, B.A. Newingham, J. Tatarko, S.K. Nouwakpo, and T. Tsegaye. 2020. A strategic plan for future USDA Agricultural Research Service erosion research and model development. Journal of Soil and Water Conservation 75:137A-143A.

Tables:

Table 1

Location (latitude and longitude), sampling date and crop, and residue details for 2018 study sites.

ID	County	Approximate Location	Sampling Date	Crop (Stage)*	Dominant Residue*	Tapes at Location
2018-01	Audrain	39.260, -92.272	08 May 2018	C (V2)	SB	3
2018-02	Audrain	39.257, -92.267	20 May 2018	NE	C	2(3)†
2018-03	Audrain	39.264, -92.270	08 May 2018	C (V2)	SB	3(4)
2018-04	Audrain	39.269, -92.265	08 May 2018	NE	C	3
2018-06	Audrain	39.267, -92.266	10 May 2018	NE	C	3(4)
2018-07	Audrain	39.270, -92.212	14 May 2018	C (V2)	SB	3
2018-09	Audrain	39.257, -92.150	29 May 2018	SB	C	3
2018-11	Callaway	39.029, -92.078	30 May 2018	SB (V1)	WSG	3
2018-15	Boone	39.215, -92.204	23 May 2018	SB (V1)	C	3
2018-16	Boone	39.212, -92.203	30 May 2018	SB (V1)	C	3
2018-20	Cooper	38.778, -92.657	17 May 2018	C (V3)	SB	3
2018-21	Cooper	38.782, -92.654	17 May 2018	C (V3)	WSG	3
2018-22	Cooper	38.817, -92.623	18 May 2018	SB (VE)	C	3
2018-25	Cooper	38.815, -92.623	18 May 2018	NE	C	3
2018-26	Callaway	39.032, -92.079	04 Jun 2018	NE	WSG	3
2018-27	Cooper	38.822, -92.620	18 May 2018	C (V3.5)	SB	3
2018-40	Boone	38.901, -92.214	13 Jun 2018	NE	SB	3
2018-41	Boone	38.901, -92.210	14 Jun 2018	NE	C	3
2018-42	Boone	38.900, -92.206	14 Jun 2018	NE	SB	3
2018-43	Boone	38.901, -92.210	14 Jun 2018	SB (VE)	C	1

*C=corn; SB=soybean; WSG=winter small grain; NE=not emerged; VE=vegetative stage emerged; V#= vegetative stage.

†57 tapes were used for comparisons between field-based and image-based tape readings. At three locations an additional tape was included for the comparison of field readings and the image-based tape mean (n=60).

Table 2

Precision test of grid-image method for estimating residue. Residue means and standard deviation (sd) are for four readers.

			Percentage of readers selecting residue on disagreement points		
Residue Type	Residue (sd)	Points disagree	75% residue	50% residue	25% residue
	(%)	(%)			
Soybean	17.5 (4.1)	18	28	33	39
Soybean	18.5 (3.8)	11	36	28	36
Soybean	24.0 (2.9)	14	57	14	29
Corn	30.0 (2.8)	10	70	0	30
Soybean	37.0 (1.4)	15	60	20	20
Corn	38.8 (1.5)	14	36	50	14
Corn	39.8 (4.2)	14	57	7	36
Corn	43.5 (1.7)	5	40	2	40
Corn	56.0 (2.7)	11	36	46	18
Corn	68.3 (1.9)	11	73	18	9
Means		12.3	49.3	23.6	27.1

Figures:

Figure 1

Two examples of a 2,400- X 1,800-pixel image (ground sampling distance = 0.014 cm pixel^{-1}) with a 100-point target grid randomly assigned. Radius of the circle around the assessment point is scaled to be equal to the minimum dimension for residue cover (2.4 mm) to facilitate assessment of residue. To be classified as residue, residue must touch the center point or touch residue that touches the center point and cover $\geq$50% of the area of the circle. Image on the left has corn residue and image on the right has soybean residue.

Figure 2

Reader precision of field tapes read in the field, from images taken 1 m above the soil surface (0.014 cm $pixel^{-1}$), and from images taken 2 m above the soil surface (0.06 cm $pixel^{-1}$) compared to the three-reader mean of the field readings. Lines of the box indicate the 25th, 50th and 75th percentiles; the (x) is the mean and the whiskers are the minima and maxima excluding outliers (data points shown).

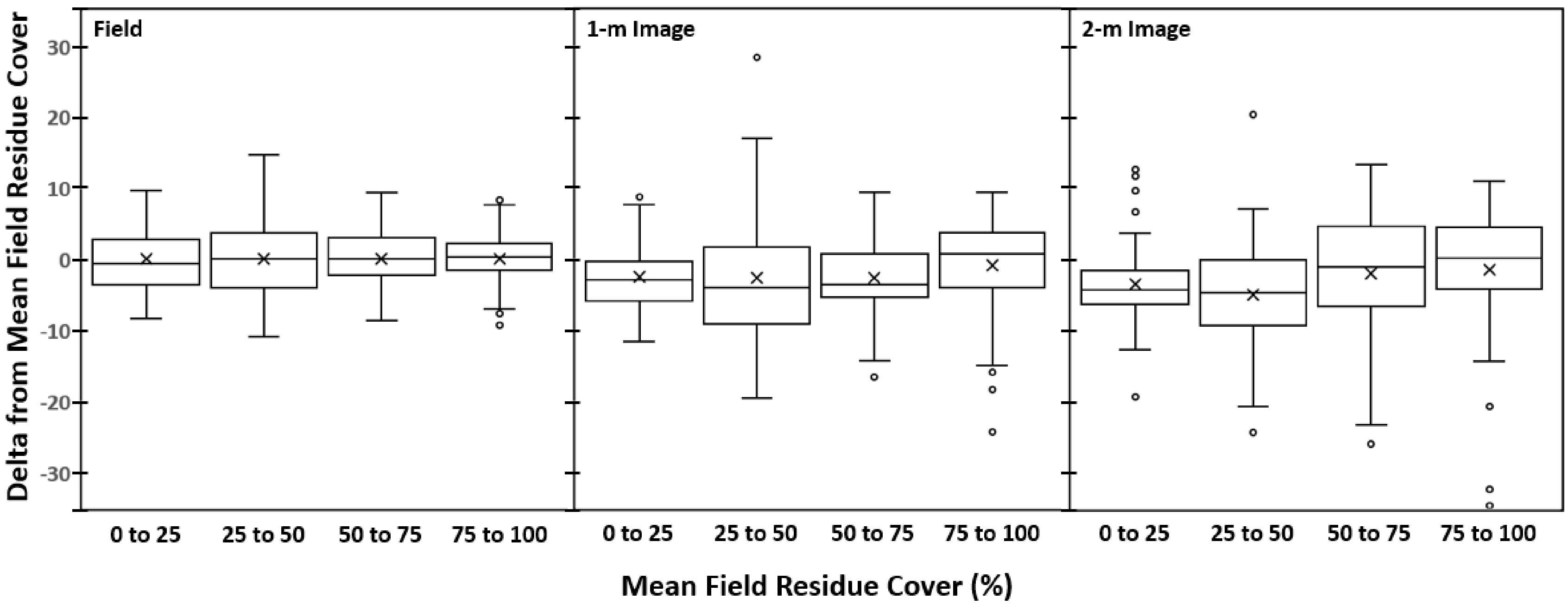

Figure 3

Deviation of reader means from the four-reader mean residue estimate for 120 images using the bullseye grid method. Points represent 100-point estimates of the 120 images by four different readers (a) or two-readers estimates (n=100; each reader reading 50 points) of the 120 images (b). Dashed line is one standard deviation based on the binomial distribution with n=100 (eq. 2).

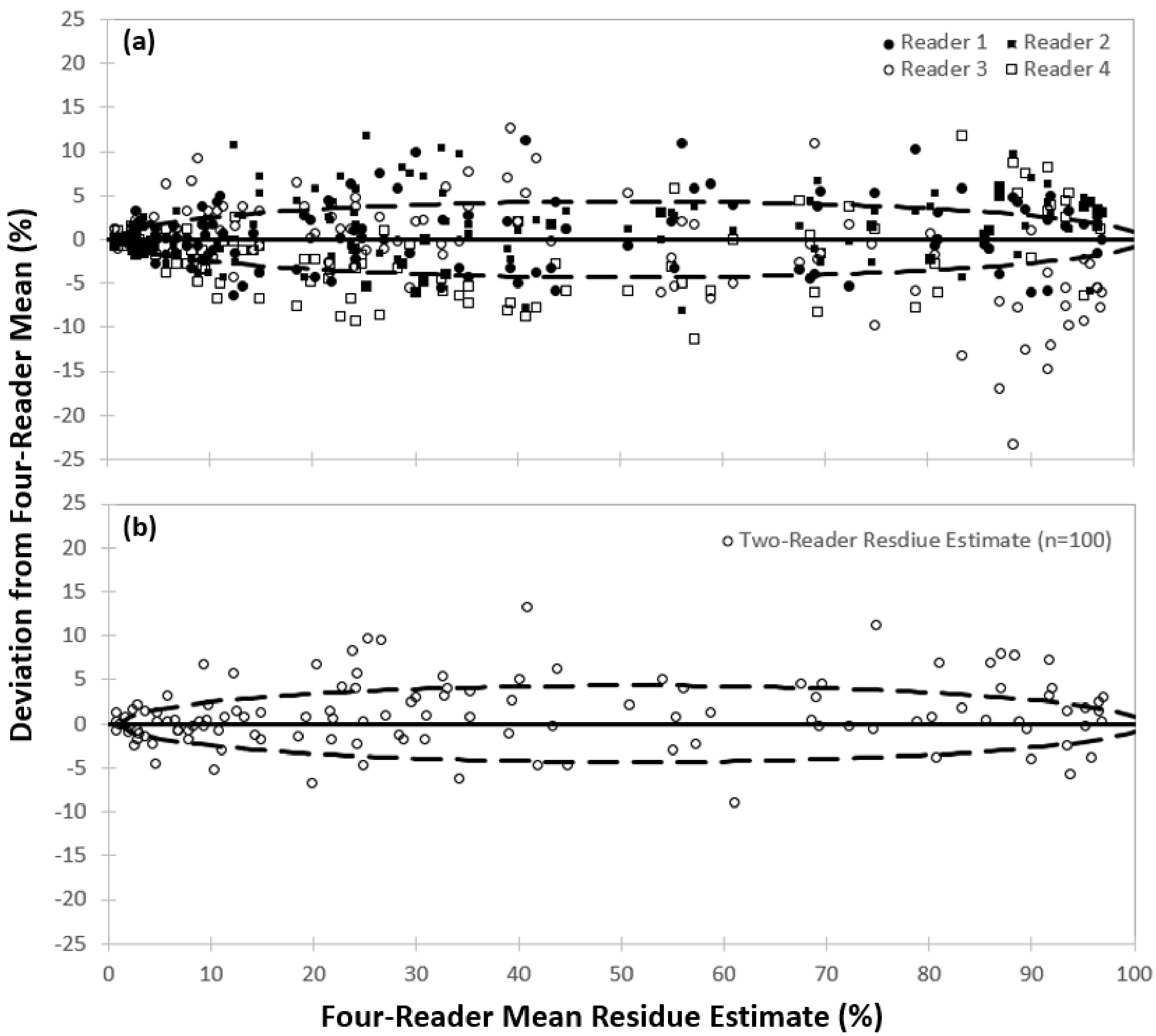

Figure 4

Field estimate of residue cover (percent) versus the image-wise estimate of residue cover (percent). Image-wise estimate of residue cover based on the sum of two 50-point grids using the bullseye system (n=100); each 50-point grid is read by a different reader.

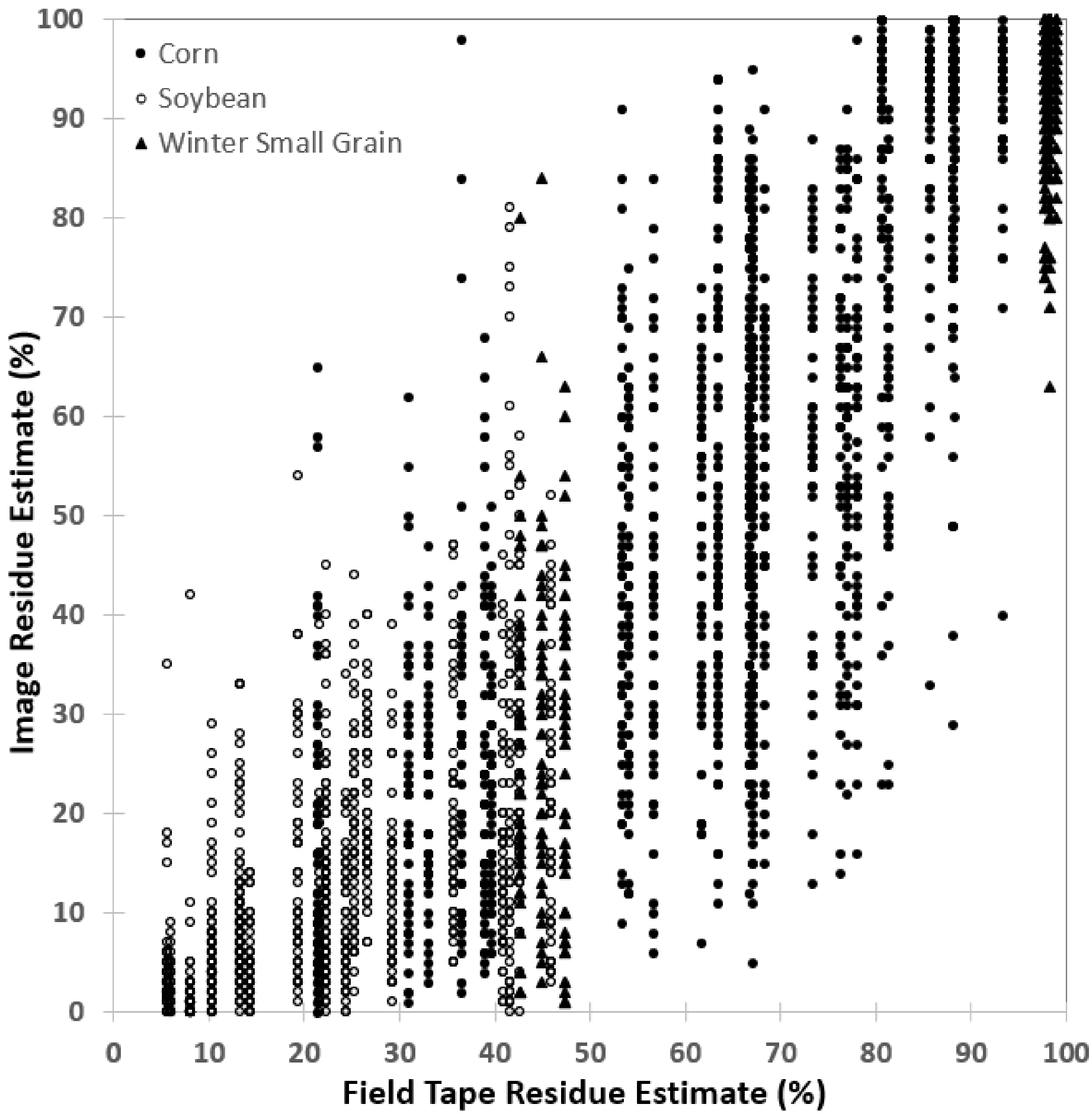

Figure 5

Examples of image-wise residue values for four tapes (field tape residue estimate): a. corn residue (67%); b. corn residue (37%); c. soybean residue (36%); and d. soybean residue (19%). All image-wise residue estimates based on two 50-point grids read by different readers (n=100) using the bullseye method.

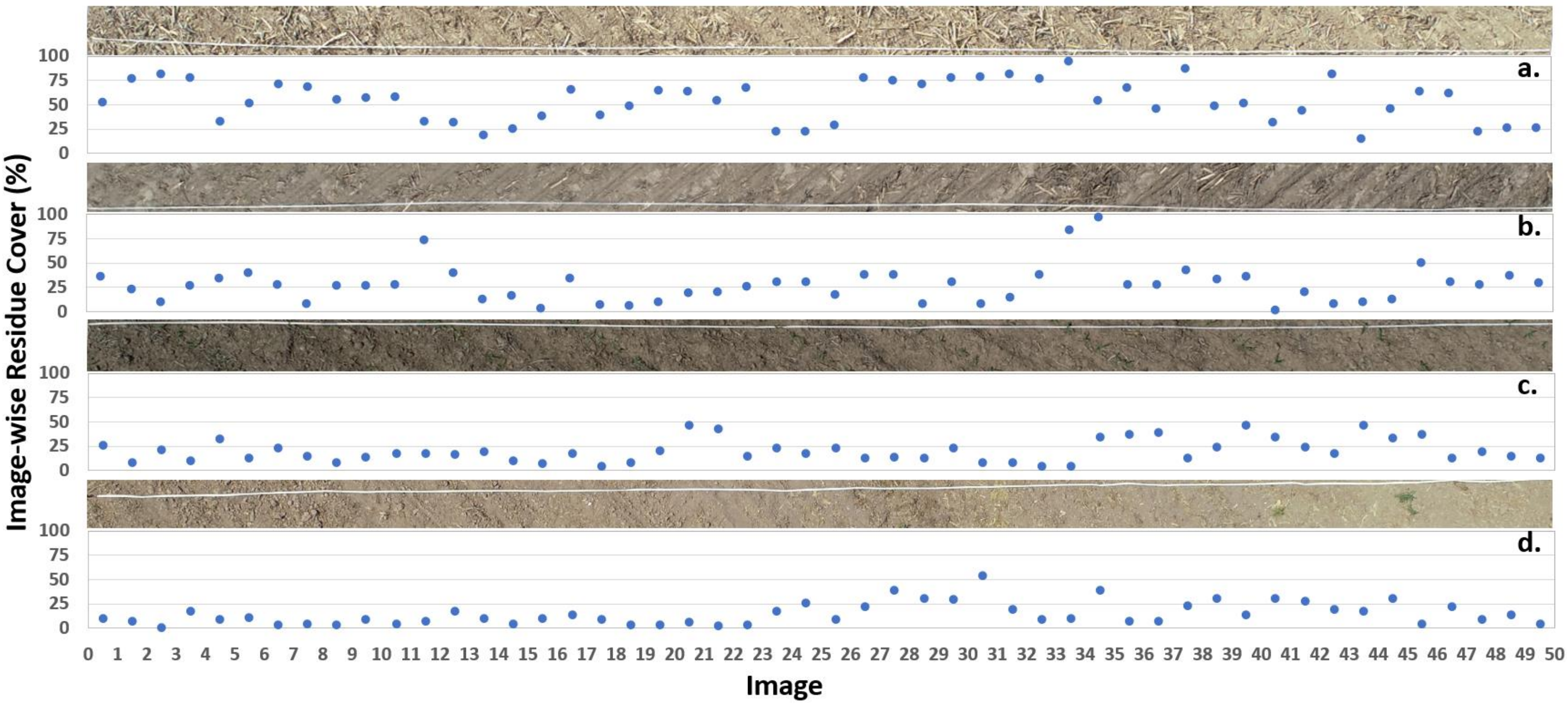

Figure 6

Field-based estimates of residue cover versus the mean image-wise readings. Field readings are based on n=100 transect points (three reader mean). Image estimate is the mean of 50 0.3-m by 0.2-m images contiguously representing the entire 15.2-m tape. Each image had n=100 readings. Solid line is the 1:1 line.

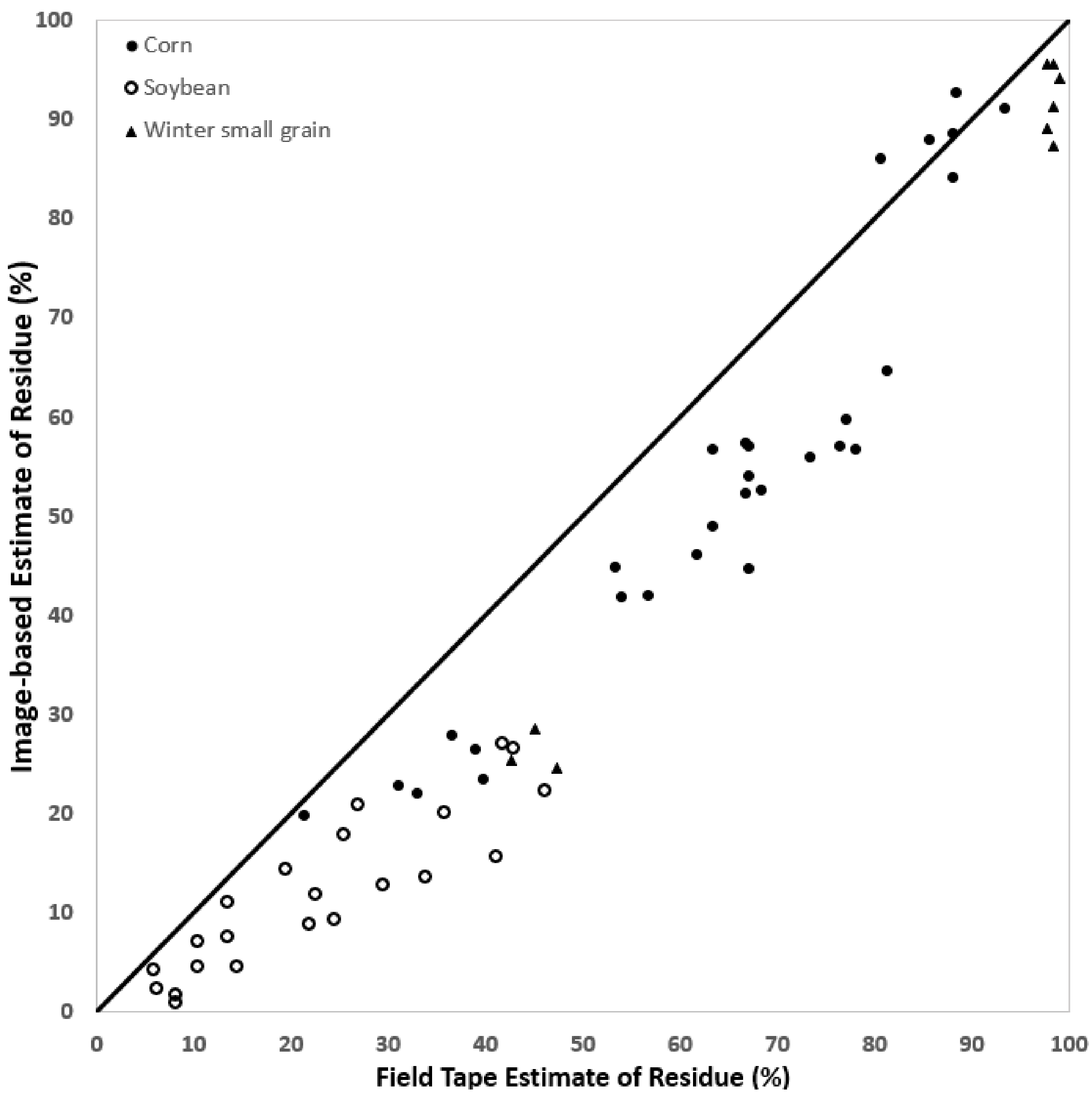

Figure 7

Effect of image-based estimate of residue on the delta between field tape reading values and image-based mean tape readings. Evaluated for residue cover less than 50%.

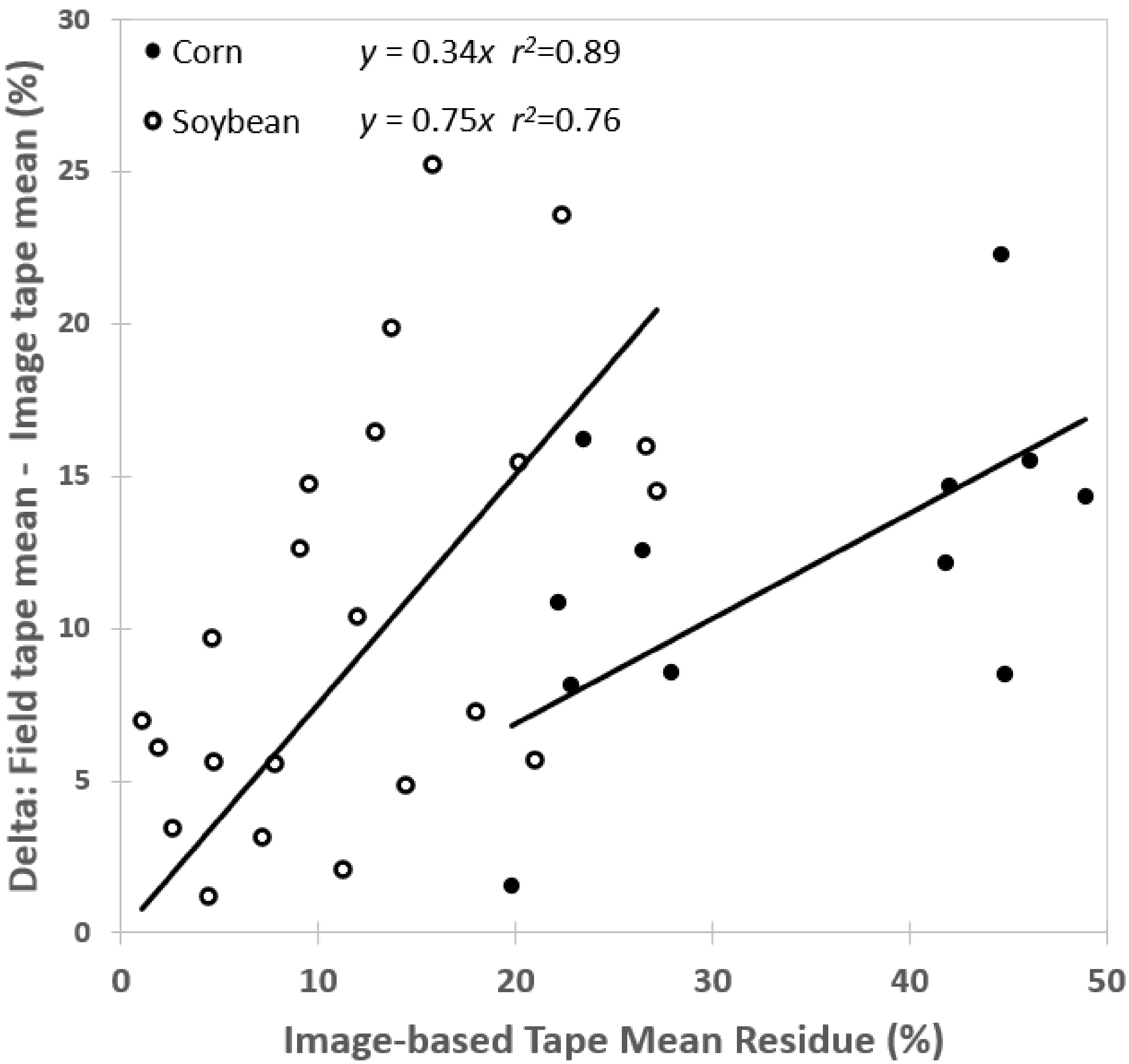

Figure 8

Subset of eight tapes comparing the 600-point reread of the tape image (0.014 cm pixel^{-1} ground sampling distance) with three estimates of residue cover: the 100-point reading from the field, the mean of the 50 gridded images, and the image-based reread of the same 100 points read in the field. Residue types included corn (squares), soybean (circles) and winter small grain (triangles). Solid line is the 1:1 line.