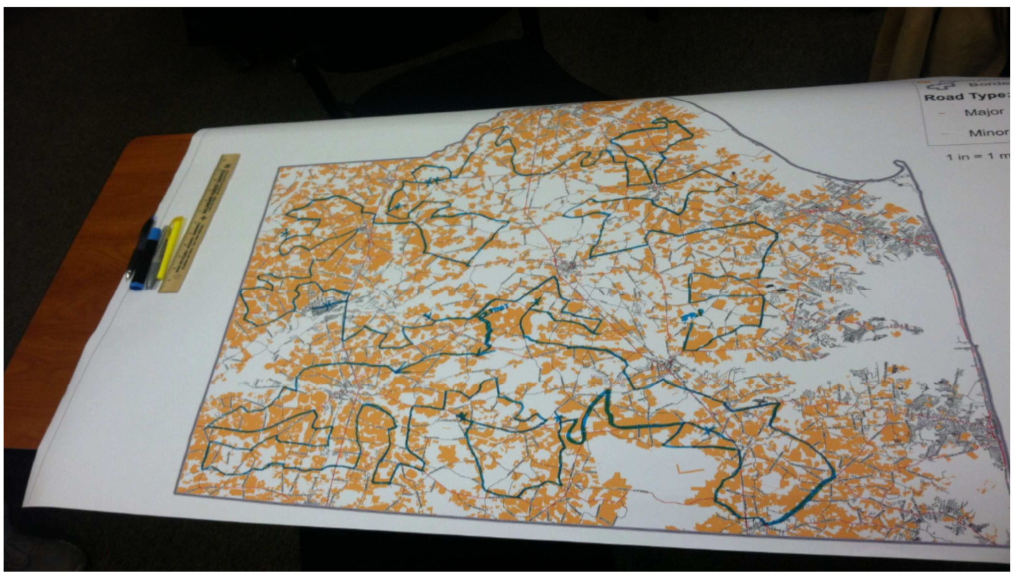
CBP/TRS-317-17

Recommendation Report for the Establishment of Uniform Evaluation Standards for Application of Roadside Transect Surveys to Identify and Inventory Agricultural Conservation Practices for the Chesapeake Bay Program Partnership's Watershed Model



Chesapeake Bay Program
Science. Restoration. Partnership.

March 16, 2017

Prepared for

Chesapeake Bay Program
410 Severn Avenue
Annapolis, MD 21403

Prepared by

Steven Dressing, Tetra Tech
Jon Harcum, Tetra Tech

With:

Mark Dubin
Agricultural Technical Coordinator
Chesapeake Bay Program Office
University of Maryland Extension

Review Comments Provided By:

Chad Watts, Conservation Technology Information Center
Susan Richards, Capital Resource Conservation and Development Area Council
Tim Sexton, Virginia Department of Conservation and Recreation

Contract Support Provided by

TETRA TECH

EPA Contract No. EP-C-12-055
Task Order No. 003

Cover Photo Credit
Marcia Fox, Delaware Department of Natural Resources and Environmental Control

Introduction

An assessment of roadside transect (i.e., windshield) survey accuracy was performed to determine the suitability of this method to identify agricultural conservation practices for credit in the Chesapeake Bay Program (CBP) partnership's watershed model. The assessment is based on a literature review, method documentation, and a detailed evaluation of the strengths and weaknesses of selected metrics.

Literature Review

Roberts and Coleman (1988) conducted a survey of crop residue in southern Ontario in 1987 using a combined roadside transect and on-site survey of over 10,000 fields. Sample size for the roadside transect survey was determined using a multinomial distribution method. They drove along predetermined random routes twice through the study area, with stops approximately every 0.8 km (0.5 mi) to record information on present crop, residue levels of the previous crop, and tillage type. The authors noted that field verification of residue categories by actual measurement is important in establishing the error associated with estimating residue cover. Reliability of the crop residue estimate was judged by a cross-tabulation of actual measured residue cover with estimated residue cover. Table 1 shows that overall accuracy for the 946 fields where actual residue measurements were taken was 96 percent. The greatest degree of accuracy occurred for the 75-100% residue category, but only six fields were included in that analysis, resulting in a 90% confidence interval of 61-100 percent accuracy. Accuracy for the 0-15% residue category was 97 percent, while the lowest level of accuracy was 91 percent for the 45-75% residue category.

Table 1. Accuracy of Residue Cover Estimation (Roberts and Coleman 1988)

Residue Category (%)	Total Fields	Fields Correct	Percent Correctly Classified[1]	90% Confidence Interval[2]
0-15	725	706	97	96%-98%
15-30	108	99	92	86%-96%
30-45	62	59	95	88%-99%
45-75	45	41	91	81%-97%
75-100	6	6	100	61%-100%
All	946	911	96	95%-97%

[1] This is equivalent to the Proportion Correct (PC) metric.
[2] 90% confidence interval using exact binomial distribution.

Thoma et al. (2004) tested the accuracy of the tillage roadside transect survey method (Hill 1998) by comparing ground measurements with results from a roadside transect survey performed on 161 randomly-selected fields in Minnesota. Fields were categorized into five residue categories (0-15, 16-30, 31-50, 51-75, and 76-100%) for comparison. Residue type was classified as corn (100 fields), bean (53 fields), or corn + bean (161 fields). Comparisons were performed on three sets of residue categories: (a) two categories of 0-30 and 31-100%; (b) three categories of 0-30, 31-75, and 76-100%; and (c) five categories of 0-15, 16-30, 31-50, 51-75, and 76-10%. Accuracy was assessed using the percentage of fields correctly classified which is equivalent to the Proportion Correct (PC) metric.

Roadside transect survey results ranged from 45 to 77 percent correct, depending on the type of residue and number of categories used for grouping (Table 2). The lowest overall accuracy occurred with five categories for beans (45 percent), corn + bean (49 percent), and corn (50 percent). All other overall accuracies were 70 percent or greater.

Table 2. Average Accuracy of the Tillage Roadside Transect Survey Visual Classifications of Crop Residue Cover Conducted by Agency Personnel in Three Minnesota Counties (Thoma et al. 2004)

Number of Categories	Corn + Bean n = 161[1]	Corn n = 100	Bean n = 53
	% Fields Correctly Classified		
5	49	50	45
3	74	70	77
2	74	70	77

[1]The Corn + Bean data groupings were not the sum of Corn and Bean groups because additional fields that had both corn and soybean residue mixed in the same field were included in the Corn + Bean group.

The authors believed that the "relatively poor classification accuracy" using five categories (Table 2) was due in part to observations being made at oblique angles which they reasoned would result in overestimation of crop residue for categories with <25% residue cover. Table 3 summarizes the accuracy of roadside transect survey observations by crop residue cover category. The lowest levels of success (as measured by the percentage of assessments that matched those made by field measurement) shown in the "Equal" column were found for the cover categories 20-24.9% (5 percent) and 45-54.9% (14 percent). In cases where residue level exceeded 34.9% there were no cases where roadside transect survey estimates exceeded measured residue levels. Over-estimation was most common for the cover categories between 10 and 24.9%. Under-estimation was most common for the cover categories 45-54.9% (86 percent), 55-69.9% (50 percent), and 25-34.9% (46 percent); it should be noted, however, that sample sizes for categories above 34.9% were small (n=1 to n=12).

Table 3. Percentages of Roadside Transect Survey Visual Classification of Crop Residue Cover that were Below, Above, and Equal to Field Measurement Values (Thoma et al. 2004)

Cover Category Based on Field Measurement	n	% Visual Estimates from Roadside Transect Survey that were Below, Above, or Equal to the Measured Crop Residue		
		Below	Above	Equal
0-9.9%	55	0	27	73
10-19.9%	38	8	55	37
20-24.9%	18	39	56	5
25-34.9%	24	46	12	42
35-44.9%	12	25	0	75
45-54.9%	7	86	0	14
55-69.9%	6	50	0	50
70-100%	1	0	0	100

Thoma et al. (2004) concluded that the tillage roadside transect survey method has numerous problems, including different personnel performing the survey in different counties, the inherent bias of observers, and the small representation of fields in the county. In addition, they noted that accuracy would improve if fewer residue cover categories were used. Particular problems were noted in discriminating small differences in residue cover near the 30% category boundary.

Daughtry et al. (2006) concurred with the observations of Thoma et al. (2004) stating that roadside surveys performed using the CTIC (Conservation Technology Information Center) method (Hill 1998) are subjective and that the techniques vary from county to county. They also claimed that no program exists for monitoring tillage objectively over broad areas.

Zheng et al. (2013) developed procedures to use Landsat 7 imagery to generate county maps with information on three tillage categories: non-conservation tillage (<30% crop residue), conservation tillage (30-70% residue), and conservation tillage-no till (>70% residue). Their remote sensing procedure (RS) resulted in tillage maps with overall classification accuracies ranging from 69 to 79 percent when compared versus field-level crop residue data measured using the line-point transect method. When they compared county-level results from RS with CTIC roadside transect tillage data, they found that the absolute differences between RS and CTIC data ranged from 0 to 45 percent for all tillage categories (Table 4). Figure 1 shows that the roadside transect survey method always under-estimated acreage in the 30-70% residue category, while generally over-estimating acreage in the <30% and >70% residue categories when compared with RS estimates. The red symbols in Figure 1 indicate questionable results for one county where the percentages for the CTIC roadside transect survey totaled 90% rather than 100%. This observation resulted in some of the greatest differences between RS and CTIC estimates. While neither method is considered “truth” in this comparison, the pattern of underestimation in the 30-70% residue category is somewhat consistent with the observations by Thoma et al. regarding Table 3 (2004).

Table 4. Differences between CTIC and Remote Sensing Estimates for Nine Counties (Zheng et al. 2013)

Residue Category	Difference Between CTIC and Remote Sensing (%) (negative values indicate CTIC less than Remote Sensing)								
<30%	0	26	5	3	5	41	3	-1	-18
30-70%	-9	-43	-8	-4	-18	-45	-34	-26	-9
>70%	9	17	3	2	12	-6	31	27	27

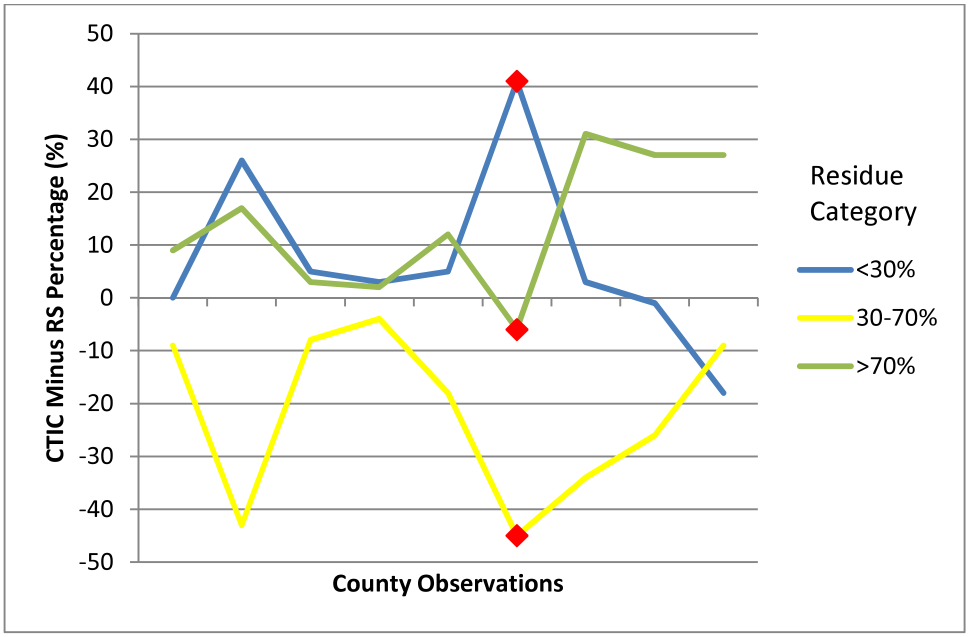


Figure 1. Difference Between CTIC and RS Estimates by Residue Category (Data from Zheng et al. 2013).

Roadside Transect Survey Method Summary

The roadside transect survey method is described in “Cropland Roadside Survey Method: Procedures for Cropland Transect Surveys for Obtaining Reliable County- and Watershed-Level Tillage, Crop Residue, and Soil Loss Data (Hill 1998). This method is designed to gather information on various agricultural practices, primarily tillage and crop residue management systems. It can also be used to collect a variety of additional information. It is reported that “*when conducted properly*, this [method] provides a high degree of confidence in the data summaries.” The following discussion highlights the major features of a roadside transect survey; additional details can be found in Hill (1998).

Route

The first step in conducting a roadside transect survey (for residue or cover crops) is to establish a driving route that passes through all soil regions that are heavily used for crop production. Routes should be at least 110 miles long and avoid heavily urbanized areas and major highways wherever possible. Routes typically traverse east to west through a county five to eight times, but direction is not significant (Hill 1998). The same route is typically used if both residue and cover crop surveys are being performed (Chad Watts, CTIC, personal communication, February 21, 2017).

Survey Dates

Residue surveys should be performed after spring planting and emergence of the crop, but before canopy closure. Data typically collected for residue surveys include county, crop (e.g., corn, forage, soybeans), and acreage with various levels of residue (e.g., <15%, ≥15<30%, ≥30<60%, and ≥60%).

Some cover crop species can be identified easily at the early growth stages, but others are more difficult to distinguish (Chad Watts, CTIC, personal communication, February 21, 2017). It can be very difficult to distinguish between rye, wheat, and barley at the early growth states (Tim Sexton, Virginia Department of Conservation and Recreation, personal communication, February 17, 2017). For this reason, cover crop surveys may need to be completed in two parts, with cover crop species, estimated establishment date, establishment density, planting method and manure application recorded in the early fall. In late spring cover crop species and termination method may be confirmed. Alternatively, the Indiana Conservation Partnership performs its cover crop survey in the winter or early spring (Chad Watts, CTIC, personal communication, February 21, 2017).

Survey Team

Hill (1998) stated that the “ideal” survey team should consist of the county Extension agent, the NRCS district conservationist, the FSA county director, and a fourth person who could be a SWCD employee, supervisor, cooperator, or newspaper reporter [or recorder] who can assist in making observations. One person would drive, another would follow the route and track data collection points, the third would record data, and the fourth person would make desired field verifications (e.g., walk field to verify residue cover). More recently, however, Capital Resource Conservation and Development Area Council (Capital RC&D) in Pennsylvania has created county survey teams that are staffed by three individuals; two of whom work in multiple counties in order to achieve greater consistency of process between counties (PA DEP 2016). Each team includes one county agriculture agency staffer (from the county to be surveyed), one highly trained consulting technician, and one data entry technician. The consulting and data entry technicians staff multiple counties. The fourth person in this case is an independent quality control technician who would perform QC (quality control) on ten percent of observation points. Delaware has adopted an approach for roadside tillage surveys that is based on the method used in Pennsylvania (Fox and Monteith 2014).

Training

Training is typically provided for the survey team. For example, Capital RC&D developed training in collaboration with a technical consultant (PA DEP 2016). The one-day training required for the technician and GIS technician includes an overview of the entire survey process and review of multiple in-field examples of crop residue. The training is supported by multiple photo guides and written survey procedures. Pennsylvania is addressing the problem associated with observations being made at oblique angles by training technicians to walk around fields that do not have consistent or easily estimated residue (Susan Richards, Capital RC&D, personal communication, March 3, 2017). The county team member is trained via webinar to provide background regarding the procedures. In addition, in-field post-training testing of the consulting technicians is done during the first week of the survey by the technical consultant and documented for quality assurance. Evaluation of the data entry technicians is also conducted by the technical consultant and documented. Delaware has provided both inside and outside training in support of its residue survey (Fox and Monteith 2014).

In the past, CTIC had provided guidance on data collection methods and photos of various residue amounts so participants could calibrate their ocular estimates (Chad Watts, CTIC, personal communication, February 21, 2017). Where multiple people conducted the residue surveys, the quality of final results was subject to their ability to maintain this ocular calibration. Even with procedures that included routine verification of ocular estimates with field measurements, it was difficult to maintain acceptable quality assurance/quality control for surveys involving multiple people in multiple places.

Sample Size

Although not typically used by those performing residue surveys, a binomial distribution (Attachment A) can be assumed to determine sample size when there are two mutually exclusive choices. This would be the case, for example, where all BMP acres from the survey are tracked and submitted as either "cropland conservation tillage" acres or "not". Crop type and other associated information could also be tracked but the confidence level and error margin for the sampling design would not apply to such ancillary data.

The sample size (n) for typical roadside transect surveys is based on the following equation derived by Tortora (1978) assuming a multinomial distribution (Attachment B):

$$n = \chi^2_{\left(1,1-\left(\frac{a}{k}\right)\right)} \times q(1-q)/d^2$$

Where:

n = sample size

$X^2_{(1,1-(\frac{a}{k}))}$ = Chi-square value for one d.f. and the value (1-(a/k)) substituted for (1-α)

a = 1-p

p = confidence level we want in the proportions arrived at for each category (this is equivalent to (1-α) for the binomial calculation)

k = number of categories

q = *a priori* estimate of the proportion for each category (as a decimal). Use the q value for the category closest to 0.50 to ensure that sample size is sufficient for all categories. Use 0.50 when unknown. (q is the same as p for the binomial calculation).

d = allowable error in the proportions (e.g., +/- 10%), expressed as a decimal (same as d for binomial calculation)

Multinomial distributions can be assumed when there are three or more mutually exclusive choices. For example, three residue levels (e.g., <30%, ≥30 to <60%, and ≥60%) could be tracked, resulting in k=3. If these same three residue levels were tracked for four different crop or land use types (e.g., corn, soybeans, other crop, other land), the value of k would be 12 (3^x4=12). Tillage practices for use in Phase 6 of the Chesapeake Bay Program Watershed Model include the following four categories (i.e., k=4): conventional/high tillage (<15% residue cover OR 15-29% residue cover with full width tillage); low residue, strip till/no-till (15-29% residue cover, strip till or no till, and less than 40% soil disturbance); conservation tillage (30-59% residue cover); and high residue, minimum soil disturbance tillage (≥60% cover, minimum disturbance) (Phase 6.0 Conservation Tillage BMP Expert Panel 2016).

When roadside transect surveys are used for cover crops, the value of k can be much larger. Cover crop practices for use in Phase 6 of the Chesapeake Bay Program Watershed Model include 3 types (traditional, traditional with fall nutrients, and commodity), 14 species, three planting dates (early, normal, late), and three planting methods (aerial, drilled, other) (Phase 6.0 Cover Crops BMP Expert Panel 2016). Overall, only 104 of the possible 378 (3x14x3x3=378) different combinations of cover crop type, species, plant date, and planting method are possible in practice. If total cropland acreage is to be used for calculations of acreages of various cover crops, then "no cover crop" should be added, increasing the number to 105 combinations. In addition, however, there are unique efficiency values for common low-till land uses and common high-till land uses in each of two regions: Coastal Plain/Piedmont crystalline/karst and Mesozoic lowlands/valley and ridge siliciclastic. Therefore, k=420 (4 x 105) for states including all combinations in their survey. The impact of increasing the number of categories, k, is discussed later in this section.

It should be noted, however, that roadside transect surveys were approved only for traditional cover crops in 2015 (AgWG 2015, Fox and Richards 2015). Neither commodity cover crops nor traditional cover crops with fall manure applications were approved for verification with roadside transect surveys, therefore reducing the maximum value of k.

The *a priori* estimates of the proportions for each category (q) are to be based on knowledge participants have regarding adoption of reduced tillage practices in the area. Absent any knowledge of expected proportions, an assumption of 0.50 should be used to provide the most conservative estimate of sample size. As surveys are performed over time, better estimates of q values can be obtained from empirical evidence.

Figure 2 illustrates how sample size varies as a function of confidence level (α=0.05 and 0.10) and *a priori* estimates of the proportion within a category (q=0 to 1). The number of categories is fixed at k=3 and the error margin is set to 0.05 for this illustration. As a reminder, the sample size is determined using the category q value that is closest to 0.5 to ensure that under-sampling does not occur. Where q=0.5, the sample sizes would be 573 and 453 for α=0.05 and 0.10, respectively.

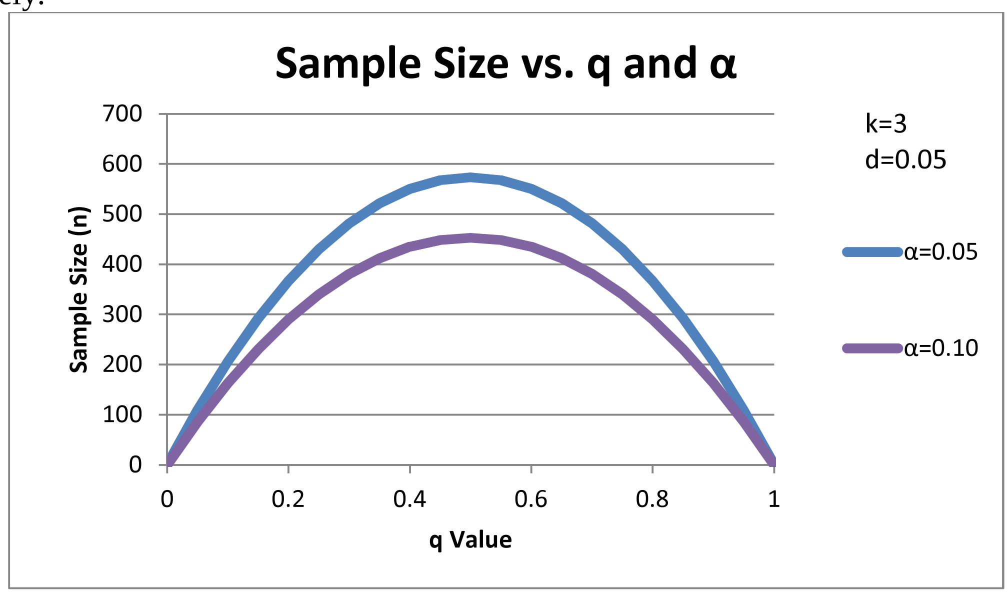


Figure 2. Sample Size as a Function of Confidence Level (1-α) and Proportion (q).

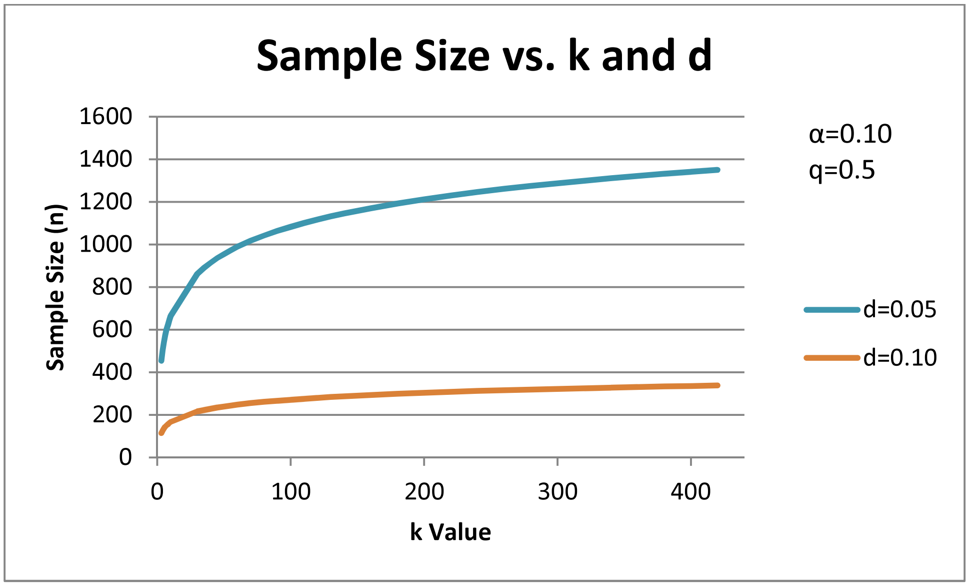


Figure 3. Sample Size as a Function of Number of Categories (k) and Error Margin (d).

Figure 3 illustrates how sample size varies as a function of the number of categories (k) and the error margin (d) at a fixed confidence level (α=0.10) and q value (0.5). In the case of cover crops with k=420, the sample sizes would be 1,350 and 338 for d=0.05 and 0.10, respectively.

Additional Considerations Regarding Sample Size

Example calculations provided by Hill (1998) specify a 90 percent confidence interval (α=0.10) and an allowable error of five percent (d=0.05). Figure 4 shows sample size for α=0.10 and d=0.05 for three values of q (0.5, 0.7, 0.9) and k values ranging from 3 to 420. The differences between the lines for different q values illustrate the importance of adhering to the requirement that the sample size is determined using the category q value that is closest to 0.5. For example, at a k value of 40, the required sample sizes would be 914, 768, and 329 for q values of 0.5, 0.7, and 0.9, respectively.

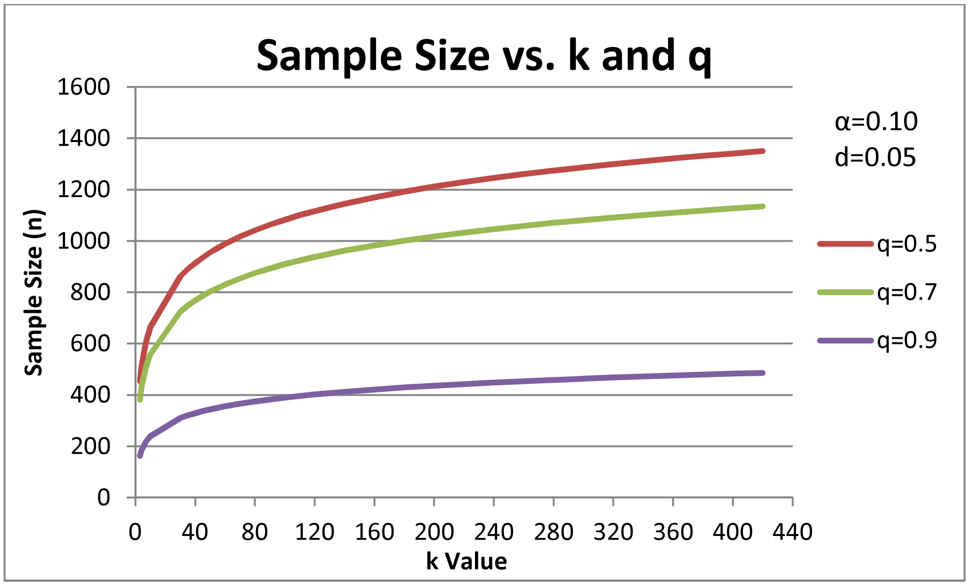


Figure 4. Sample Size as a Function of Number of Categories (k) and Category Proportion (q).

Careful consideration of categories is essential to developing a cost-efficient roadside transect survey. As visualized in Figure 4, required sample size doubles from k=3 (n=163) to k=38 (n=326) for q=0.9. For cover crops in particular, it may be prudent to strategize on the cover crop species to be tracked to maximize the benefit (e.g., model credit) of information gained for the minimum cost (i.e., lowest feasible k value). By considering the efficiencies for nitrogen (N), phosphorus (P), and total suspended solids (TSS) assigned to the cover crop species most likely to be grown in a jurisdiction, for example, it may be possible to create groupings that both simplify tracking and reduce the number of sample points required. As an example, traditional rye that is drilled early has the greatest N efficiency credit (0.45) for the Coastal Plain and Piedmont regions (Phase 6.0 Cover Crops BMP Expert Panel 2016). If this species is broadly applied and can be easily identified via roadside transect survey, it may be prudent to simply track it and a small set of other species (perhaps in groups[1]) depending on the expected prevalence and efficiency credits for other species grown in the jurisdiction. Alternatively, if q values are 0.15 or lower for all categories, tracking a large number of categories may yield a manageable sample size as indicated by the lower sample sizes for such q values in Figure 2.

Because urbanization typically reduces cropland acreage, the number of observation points that capture cropland is likely to decrease over time in many jurisdictions. For this reason it is advisable to over-sample initially to reduce the likelihood that the number of fixed sampling points (n) decreases below the required level over time. For example, between 1945 and 1997 ERS reported that the Northeast lost about 12% of cropland. This is a rate of about 0.25 percent per year. Using this rate and a 10-year timeframe, this would indicate a need to oversample by about 2.5 percent to ensure that fixed sample sizes meet statistical requirements over the 10-year period. Assuming a 20-year timeframe, over-sampling would need to be 5 percent.

[1] Efficiencies assigned to any grouped species may need to be the lowest values for any species in the group in accordance with CBP policy.

Data Collection

Data are collected by driving along the route and making observations on each side of the road. Drivers stop at even-interval distances from the starting point. Intervals are often set at a half mile, but interval length will depend on the route length and number of observations required. If no cropland field is encountered at the half-mile (or other pre-set) interval on either side of the road, participants are to continue driving until cropland is observed on at least one side of the road; this becomes the starting point. No data are collected at an observation point for any side of the road where no cropland field is encountered (Hill 1998). If no cropland field is encountered at a stopping point on one side of the road, participants are to record data only for the side with cropland.

For quality control, typically ten percent of the crop observations of each technician are visited by an independent quality control technician and documented. Verified observations are corrected, and in Pennsylvania the results of QC are used to determine the overall accuracy and consistency of each technician (Susan Richards, Capital RC&D, personal communication, March 3, 2017). Technicians in Pennsylvania are contacted and retrained if QC reviews show significant inconsistency between the reviewer and the technician; in some cases data may be discarded. However, it would also be beneficial if the results of QC are then used to determine overall accuracy of the method application (see *Metrics for Roadside Transect Survey Accuracy and Completeness* for a discussion).

Calculation of Acreage in Each Category

Survey results are used to calculate acreages of land in the various categories of interest. The simple approach is to multiply the total acreage of the specific crop (the most accurate available data should be used for this purpose) by the fraction of observations of that crop found in each category. For example, consider the acreage of crop A in tillage category I.

Total acreage of crop A: $\mathrm{Acreage}_{\mathrm{A}}$
Total number of sample points in crop A: $\mathrm{Points}_{\mathrm{A}}$
Total number of crop A sample points in category I: $\mathrm{Points}_{\mathrm{AI}}$

$$Acreage_{AI} = Acreage_{A} \times \frac{Points_{AI}}{Points_{A}}$$

For example, if there are 80,000 acres of soybeans, 200 observation points with soybeans as the crop, and 50 soybean observation points with <30% residue, the total acreage of soybeans with <30% residue is calculated as:

$$Acreage = 80{,}000\ x\ \frac{50}{200} = 20{,}000\ acres$$

Roadside Transect Survey Accuracy Metrics

For the purposes of this assessment, roadside transect surveys are assumed to include both data collection from the vehicle and field verification of a percentage of sampling points for which data were recorded. It is also presumed that more than two categories are used for classification. The procedures described in CBP (2017) for binary classifications would be appropriate if there are only two categories.

The accuracy of the roadside transect survey approach is determined by comparing vehicle-based observations with observations from in-field measurements. As described earlier, data are collected from a roadside transect survey and uniquely classified. A random sample of the vehicle-based observations is ground truthed for quality control. The possible outcomes from the comparison of roadside transect surveys and in-field measurements (ground truth results) can be summarized in a sample count error matrix as shown in Table 5—in this example there are four categories (k=4), but the approach can be extended to larger k values if necessary. The row totals represent the number of observations falling into each class based on the roadside transect results. The column totals represent the number of observations falling into each class based on the in-field measurements. If the ground truthing agreed perfectly with the roadside transect data, then the diagonal (gray) cells would be non-zero and equal to the corresponding row and column totals. The off-diagonal (non-gray) cells represent the number of observations miss-classified by the roadside transect method in comparison to the in-field measurements. If the ground truthing agreed perfectly with the roadside transect, then the off-diagonal cells would be zero. The marginal proportions, w_i, are the proportion of data classified into each category using all data from the roadside transect (and would not likely equal $\boldsymbol{n_{i+}/n}$).

Table 5. Sample Count Error Matrix Notation

Class		Reference/Ground Truth 1	2	3	4	**Row Total**	Marginal Proportions
Roadside Transect Results	1	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	$n_{1,4}$	$\boldsymbol{n_{1+}}$	w_1
	2	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$	$n_{2,4}$	$\boldsymbol{n_{2+}}$	w_2
	3	$n_{3,1}$	$n_{3,2}$	$n_{3,3}$	$n_{3,4}$	$\boldsymbol{n_{3+}}$	w_3
	4	$n_{4,1}$	$n_{4,2}$	$n_{4,3}$	$n_{4,4}$	$\boldsymbol{n_{4+}}$	w_4
Column Total		$\boldsymbol{n_{+1}}$	$\boldsymbol{n_{+2}}$	$\boldsymbol{n_{+3}}$	$\boldsymbol{n_{+4}}$	$\boldsymbol{n}$	

True (adjusted) proportions and confidence intervals can be computed using procedures provided by Congalton and Green (2008) and Olofsson et al. (2013). The approach to compute confidence limits is demonstrated with the example data shown in the top section of Table 6 (*Sample Count Error Matrix*). For example, there were 20 roadside transect observations classified as Category 1 and confirmed as Category 1 during the follow-up ground truthing, while there were 4 roadside transect observations classified as Category 1 but found to be in Category 3 during ground truthing. Unbiased estimators of individual cell probabilities can be calculated by adjusting for the marginal proportions using the below equation

$$\hat{p}_{i,j} = w_i \left(\frac{n_{i,j}}{n_{i+}}\right)$$

Table 6. Sample Error Matrix and Related Calculations

Sample Count Error Matrix	Class		Reference/Ground Truth 1	2	3	4	Row Total	Marginal Proportions
	Classified Data	1	20	10	4	1	**35**	0.3
		2	3	25	5	2	**35**	0.4
		3	1	3	20	6	**30**	0.1
		4	1	2	3	15	**21**	0.2
	Column Total		**25**	**40**	**32**	**24**	**121**	

Area Proportion Error Matrix			Reference/Ground Truth 1	2	3	4	Row Total
	Classified Data	1	0.171	0.086	0.034	0.009	**0.3**
		2	0.034	0.286	0.057	0.023	**0.4**
		3	0.003	0.010	0.067	0.020	**0.1**
		4	0.010	0.019	0.029	0.143	**0.2**
	Column Total (true or adjusted proportions w/ 90% confidence int.)		**0.219** ±0.055	**0.400** ±0.068	**0.187** ±0.056	**0.194** ±0.046	

Producer's Accuracy	0.784	0.713	0.357	0.735	
User's Accuracy	0.571	0.714	0.667	0.714	
Overall Accuracy					0.667

Var(Producer's Acc.)	5.24×10^{-3}	2.64×10^{-3}	5.32×10^{-3}	6.06×10^{-3}	
Var(User's Accuracy)	2.02×10^{-3}	1.69×10^{-3}	1.84×10^{-3}	1.69×10^{-3}	
Var(Overall Accuracy)					1.803×10^{-3}

This can be illustrated using the example data in Table 6. Selecting $n_{1,3}$ yields $\hat{p}_{1,3}$ = 0.3 x 4 / 35 = 0.034 (see cells with red borders in Table 6). In essence, these unbiased estimators are the proportions we would expect to observe after normalizing the data for sample size.

The unbiased estimators of the true (adjusted) marginal proportions, $\hat{p}_{+j}$, are the column totals in the *Area Proportion Area Matrix* section of Table 6. For example, 0.219 or 21.9% of the total area is Category 1. That is, although the roadside transect results indicate that 30% of the total area is in Category 1 (i.e., w_1=0.3), only 21.9% of the total area is expected to be in Category 1 after accounting for bias computed based on a comparison to ground truth data.

The variance of $\hat{p}_{+j}$ can be estimated as

$$Var(\hat{p}_{+j}) = \sum_{i=1}^{k} w_i^2 \frac{\frac{n_{i,j}}{n_{i,+}}\left(1 - \frac{n_{i,j}}{n_{i,+}}\right)}{n_{i,+} - 1}$$

Using the data in Table 6, the variance for $\hat{p}_{+1}$ can be estimated as

$$Var(\hat{p}_{+1}) = (0.3^2)\frac{\frac{20}{35}\left(1-\frac{20}{35}\right)}{35-1} + (0.4^2)\frac{\frac{3}{35}\left(1-\frac{3}{35}\right)}{35-1} + (0.1^2)\frac{\frac{1}{30}\left(1-\frac{1}{30}\right)}{30-1} + (0.2^2)\frac{\frac{1}{21}\left(1-\frac{1}{21}\right)}{21-1} = 0.00112$$

and the half-width confidence interval may then be computed as the square root of the variance times the corresponding normal statistic (i.e., $Z_{1-\alpha/2}Var^{0.5}$). So, the 90% confidence interval for the unbiased estimate of the proportion of observations that fall into Category 1 would be 21.9% +/-5.5% where 5.5% is computed as 100 x 1.645 x $0.00112^{0.5}$.

The producer's accuracy, or the probability of a correct determination given that the true class=i is computed as

$$\hat{\theta}_i = \frac{\hat{p}_{i,i}}{\hat{p}_i}$$

For example, the producer's accuracy for Category 2 is $\hat{\theta}_2$ = 0.286 / 0.400 = 0.713 (see Table 6). Note that $\hat{p}_{i,i}$ in this case is found in the *Area Proportion Error Matrix* of Table 6 where the Classified Data values is 2 and the Reference/Ground Truth value is also 2 (see green box). The value of $\hat{p}_i$ is also found in the *Area Proportion Error Matrix* of Table 6, but as the Column Total for Reference/Ground Truth value of 2 (see green box).

The user's accuracy is computed as

$$\hat{I}_i = {n_{i,i}}/{n_{i+}}$$

For example, the user's accuracy for Category 2 is $\hat{I}_2$ = 25 / 35 = 0.714 (see blue boxes in Table 6).

The overall accuracy, $\hat{P}$, is sum of the individual cell probabilities along the diagonal of the *Area Proportion Error Matrix* portion of Table 6, e.g., 0.171 + 0.286 + 0.067 + 0.143 = 0.667.

The variance for the overall accuracy can be calculated by

$$Var(\hat{P}) = \sum_{i=1}^{k}\left({p_{i,i}(w_i - p_{i,i})}/{(w_i n)} \right)$$

$$Var(\hat{P}) = \begin{matrix} (0.171)(0.3-0.171)/(0.3 \times 121) \\ +(0.286)(0.4-0.286)/(0.4 \times 121) \\ +(0.067)(0.1-0.067)/(0.1 \times 121) \\ +(0.143)(0.2-0.143)/(0.2 \times 121) \end{matrix} = 0.001803$$

The variance for the producer's accuracy can be calculated by

$$Var(\hat{\theta}_i) = \frac{p_{i,i}}{p_i^4}\left[p_{i,i}\left(\sum_{j\neq i}^{k}\frac{p_{i,j}(w_j - p_{i,j})}{w_j n}\right) + \frac{(w_i - p_{i,i})(p_i - p_{i,i})^2}{w_i n}\right]$$

For example,

$$Var(\hat{\theta}_1) = \frac{0.171}{0.219^4}\left[0.171(\dots) + \frac{(0.3 - 0.171)(0.219 - 0.171)^2}{(0.3)(121)}\right] = 0.0052352$$

$$where\ (\dots) = \left(\frac{0.034(0.4 - 0.034)}{(0.4)(121)}\right) + \left(\frac{0.003(0.1 - 0.003)}{(0.1)(121)}\right) + \left(\frac{0.010(0.2 - 0.010)}{(0.2)(121)}\right)$$

The variance for the user's accuracy can be calculated by

$$Var(\hat{I}_i) = \frac{p_{i,i}(w_i - p_{i,i})}{w_i^2 n}$$

For example,

$$Var(\hat{I}_1) = \frac{0.171(0.3 - 0.171)}{(0.3^2)(121)} = 0.002024$$

The half-width confidence interval for the overall, producer's, and user's accuracy may then be computed as the square root of the variance times the corresponding normal statistic (i.e., $Z_{1-\alpha/2}Var^{0.5}$). For example, using the variance of 0.001803 computed for the $Var(\hat{P})$, the half-width confidence interval is +/-7.0% where 7.0% is computed as 100 x 1.645 x $0.001803^{0.5}$. So the overall accuracy is 66.7% +/-7.0%.

Results and Discussion

Roadside transect surveys have a long history of practical use in quantifying the extent of conservation tillage practices at the county level. Standard methods exist for sampling protocols and computing sample sizes. Increasing the number of classification categories, k, results in increased sample sizes in order to maintain a constant allowable error, d. However as the number of categories increases, it is less likely that the *a priori* estimate of the proportion for any one category, q, would approach the conservative assumption of 0.5 and somewhat mitigate the sample size requirements associated with the increased number of categories.

Past studies have indicated biases exist with roadside transect surveys. There is a tendency to overestimate crop residues for categories with lower levels of in-field measured crop residue and underestimate crop residues for categories with higher levels of in-field measured crop residue.

Additional challenges with roadside transect surveys include different personnel performing the survey in different counties, the inherent bias of observers, and the small representation of fields in the county. For example, the ability to capture information at the species level for observed cover crops often varies county-by-county depending on the knowledge of the person in each county who conducts the survey (Chad Watts, CTIC, personal communication, February 21, 2017). The routes have been a challenge for residue surveys, and the lowest resolution at which CTIC is willing to assert the statistical validity of the data is at the county level.

Past studies also indicate that accuracy would improve if fewer residue cover categories were used. Simplifying the survey by redefining tillage definitions to encompass fewer categories makes a lot of sense to those who have been involved with residue surveys over the years (Chad Watts, CTIC, personal communication, February 21, 2017). Some argue that fewer distinctions between various residue amounts, especially those in the middle of the spectrum, may be appropriate because for most purposes it may be sufficient to know whether or not the field was tilled, and how intensely. The most important point here is that the categories and thresholds should be governed by appropriate category definitions and the level of specificity required for achieving objectives.

Nevertheless, procedures exist that can be used to estimate survey accuracy (overall, producer's and user's accuracies) and adjust the results of roadside transect surveys based on in-field measured ground truthing.

Recommendations

The above results and discussion section points to following a survey design that includes methodology documentation, sample size calculations, training, ground truthing, and survey accuracy assessments.

A two-step process may be appropriate for determining whether and how to accept data from roadside transect surveys. The first step would be to assess the methodology used, with an emphasis placed on conformance with the method described by Hill (1998) or a documented alternative. Hill's method is the only widely known method for roadside transect surveys applicable to cropland residue and, more recently, cover crops. Route delineation, survey dates, survey team qualifications, training, and data collection should follow Hill's recommendations unless alternative approaches can be justified. For example, routes and sampling points should be used consistently over time and training should be provided by qualified staff, preferably on an annual basis. The most accurate data available on crop acreages should be used to calculate acreages of land in the various categories. Sample size should be calculated for each effort using the multinomial equation and best available information regarding q values and anticipated loss of cropland over time. The basic recommendations for setting an acceptable sampling level include the following:

- The sample size for each county should be based on a minimum confidence interval of 90 percent and an allowable error, d, not to exceed 0.10.
- The *a priori* value of q should be set to 0.5 unless a previous survey or a pilot study can be used to justify an alternative value.

- The final sample size should include an additional 5 percent to allow for site loss due to urban encroachment unless an alternative value can be justified.

In the studies reviewed for this report, it was typical that approximately ten percent of sites per county (as calculated according to the above) be field-verified using an appropriate method. For example, the line-point transect method would be an appropriate method for field-verifying crop residue. The results of these quality assurance samples should be used to determine the accuracy of the roadside transect survey method. It should be noted that for larger numbers of classes, a ten percent sampling of sites for quality assurance might result in relatively few sites per class which would adversely affect the analysis outcome.

Because of differences in personnel from county-to-county and year-to-year, the results from the quality assurance are staffing specific. The basic recommendations for acceptability include the following:

- Each quality assurance analysis will report the sample count error matrix; the area proportion error matrix (including confidence intervals of the true or adjusted proportions); the producer's, user's, and overall accuracy estimates; and the confidence interval of the overall accuracy estimate.
- To use the true (adjusted) proportions resulting from the analysis, the lower confidence limit on the overall accuracy must exceed 50 percent. (A value of 50 percent was selected based on the lower range of survey accuracies discussed in the literature review section of this report.)
- A minimum confidence level of 90 percent should be used on all statistics.

References

AgWG. 2015. *Draft Call Summary: Agriculture Workgroup September 16-17, 2015*. U.S. Environmental Protection Agency, Chesapeake Bay Program, Agriculture Workgroup. Accessed February 24, 2017.
http://www.chesapeakebay.net/channel_files/22598/draft_call_summary_agwg_sept_16.v5.pdf

CBP (Chesapeake Bay Program). 2017. *Recommendation Report for the Establishment of Uniform Evaluation Standards for Application of Remote Sensing to Identify and Inventory Agricultural Conservation Practices for the Chesapeake Bay Program Partnership's Watershed Model*. CBP/TRS-316-17. Developed for Chesapeake Bay Agricultural Remote Sensing Technical Team of U.S. Environmental Protection Agency Chesapeake Bay Program, by Tetra Tech, Inc., Fairfax, VA, 12 p.

Congalton RG, and K. Green. 2008. *Sample Design Considerations. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 2nd edition. CRC Press, Boca Raton, Florida.

Daughtry, C.S.T., P.C. Doraiswamy, E.R. Hunt, Jr., A.J. Stern, J.E. McMurtrey, III, and J.H. Prueger. 2006. Remote sensing of crop residue cover and soil tillage intensity. *Soil Tillage Res*. 91:101–108.

Fox, M. and T. Monteith. 2014. *Delaware's Cropland Transect Survey*. Delaware Department of Natural Resources and Environmental Control, Delaware Watershed Assessment and Management Section. Presented to Agriculture Workgroup December 11, 2014. Accessed February 24, 2017. http://www.chesapeakebay.net/channel_files/21229/delawares_cropland_transect_survey_agwg.pdf

Fox, M. and S. Richards. 2015. Delaware and Pennsylvania's Pilot Cover Crop Transect Surveys. Presented to Agriculture Workgroup September 17, 2015. Accessed February 24, 2017. http://www.chesapeakebay.net/channel_files/22592/foxcover_crop_pilot_survey_agwg09172015_final.pdf

Hill, P. 1998. *Cropland Roadside Survey Method: Procedures for Cropland Transect Surveys for Obtaining Reliable County- and Watershed-Level Tillage, Crop Residue, and Soil Loss Data*. Publication developed as an educational resource by Indiana's T-by-2000 Soil Conservation Education Program. Accessed February 21, 2017 (NOTE: Document will be added to this site). http://www.ctic.org/crm

Olofsson, P., G.M. Foody, S.V. Stehman, and C.E. Woodcock. 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*. 129(203):122-131.

PA DEP. 2016. *Pennsylvania Department of Environmental Protection QA00 Addendum Verification Program Draft March 2016*. Pennsylvania Department of Environmental Protection. Accessed February 3, 2017. http://www.chesapeakebay.net/documents/PA_BMP_Verification_QAPP_Final.pdf

Phase 6.0 Cover Crops BMP Expert Panel. 2016. *Phase 6 Cover Crops Efficiency Values (excel) – Revised. Submitted to: Agriculture Workgroup Chesapeake Bay Program December 17, 2016*. Filename: phase6covercropefficienciesrevised_01232017.xlsx. Accessed February 6, 2017. http://www.chesapeakebay.net/calendar/event/24774/ phase6covercropefficienciesrevised_01232017.xlsx

Roberts, P. and D. Coleman. 1988. *A Survey of Crop Residue in Southwestern Ontario*. Report prepared for the Management Committee of the Soil and Water Environmental Enhancement Program. 64 pp. Accessed February 14, 2017. http://agrienvarchive.ca/download/sweep_06.pdf

Thoma, D.P., Gupta, S.C., Bauer, M.E., 2004. Evaluation of optical remote sensing models for crop residue cover assessment. *J. Soil Water Conserv*. 59, 224–233.

Tortora, R.D. 1978. A note on sample size estimation for multinomial populations. *The American Statistician* 32(3):100-102.

Zheng, B., J.B. Campbell, Y. Shao, and R.H. Wynne. 2013. Broad-scale monitoring of tillage practices using sequential Landsat imagery. *Soil Sci. Soc. Am. J.* 77:1755–1764.