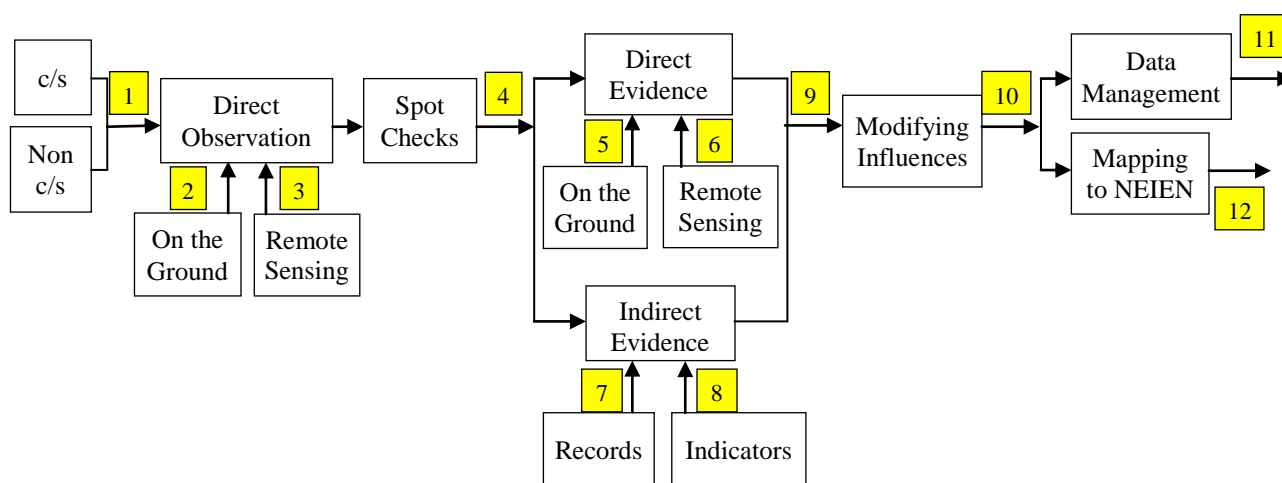


## 2.1.4 Verification Pathways and Potential Error Sources

Verification is a multidisciplinary process that can involve multiple individuals, groups, or agencies, as well as multiple technical and administrative approaches. As noted in section 2.1.2, verification needs to address both the presence or implementation of a practice and the degree to which the practice meets standards and specifications of performance. These two determinations could be made via direct observation, analysis of records and reporting (presumably based on direct observation made by one or more others), or a statistically-based approach that applies direct observation or analysis of records and reporting to a sampled subset of locations where the practice should be implemented.

The pathways associated with direct observation, analysis of records/reporting, and statistical sampling approaches are conceptually illustrated in Figures 1-3, respectively.



**Figure 1. Pathway and error sources for direct observation.** Numbered boxes denote steps where error or confidence can be evaluated.

Key to Figure 1.

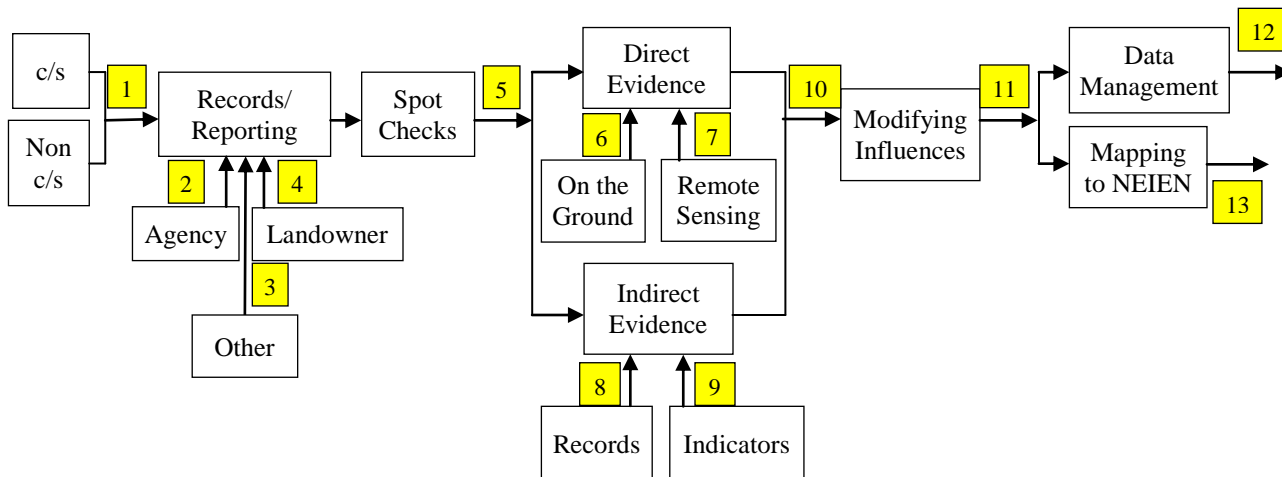
Error term	Description
1	Uncertainty whether both c/s and non c/s BMPs are detected.
2	Confidence in quality of on-the ground observation detecting BMP.
3	Confidence in quality of remote sensing detecting BMP.
4	Confidence in spot-checking approach.
5	Confidence that on-the-ground observation adequately determines BMP function.
6	Confidence that remote sensing data adequately characterize BMP functionality.
7	Confidence that records (e.g., NMP) capture BMP function.
8	Confidence that indicators support assumption of BMP function.
9	Overall confidence in scoring of BMP functionality.
10	Confidence that modifying influences (e.g., season, timing observations) are accounted for.
11	Uncertainty in effectiveness of data management to accurately record BMP data.
12	Uncertainty in mapping state/local BMP to NEIEN.

The pathways illustrated in Figure 1 include the option to use on-the-ground or remote sensing methods for detection, as well as a choice between direct or indirect evidence of function. Both the direct (on-the-ground or remote sensing) and indirect (records or indicators) evidence options for function also include options. Figure 1 identifies eleven unique (9 is the combined or overall error or confidence associated with assessment of functionality) error sources associated with using direct observation to verify BMP implementation. Clearly, the magnitude of these various error sources is likely to differ for each BMP (e.g., on-the-ground observation is usually better than remote sensing for detecting practices), and will vary for different BMPs (e.g., remote sensing is more reliable for structural and visible practices).

**Table 2. Direct observation scoring.** Ratings and scores are examples of highest level of confidence.

Rate confidence in results of operation on a scale of 1 (lowest) to 5 (highest)					
Where a choice is required ("OR" statement), select and rate only one operation					
Sector	Error term	Operation	Rating	Weighting	Component Score
<b>BMPs</b>	1	c/s & non c/s	5	4	20
<b>DETECTION</b>					
	2	on-the-ground	5	3	15
		OR			
	3	remote sensing		3	0
	4	Spot checks	5	3	15
<b>FUNCTION</b>					
	Direct Evidence				
	5	on-the-ground	5	2	10
		OR			
	6	remote sensing		2	0
		OR			
	Indirect Evidence				
	7	records		2	0
		OR			
	8	indicators		2	0
	9	Scoring of Function	5	2	10
	10	Modifying Influences	5	2	10
<b>REPORTING</b>					
	11	Data Management	5	2	10
	12	Mapping to NEIEN	5	2	10
				<b>TOTAL</b>	100

Table 2 illustrates a scoring system that could be used to compare the relative error of different direct observation approaches. This scoring system is scaled from 0 (worst) to 100 (best) for easy comparison and for potential use in adjusting crediting and modeling based on the approach that is used to verify each BMP. Tables 3 and 4 show parallel scoring for records/reporting and statistical sampling approaches, respectively. In each table, a maximum of 20 points is assigned to uncertainty whether both c/s and non-c/s practices are detected (a measure of program coverage), maximum 30 points each are assigned to detection and assessment of functionality and a maximum of 20 points are assigned to data management and reporting combined. The ratings and weightings incorporated in Tables 2-4 are intended as initial suggestions that can be changed to reflect the judgment of the AgWG membership. It should be noted that at present the assignment of the error term values and the scoring system will likely be categorical best-professional-judgment values (i.e., 1 - 5), but where actual numerical confidence/error terms are or will be available, the concept could be relatively easily retooled to use more absolute or quantitative expressions of confidence.



**Figure 2. Pathway and error sources for records/reporting.** Numbered boxes denote steps where error or confidence can be evaluated.

**Key to Figure 2.**

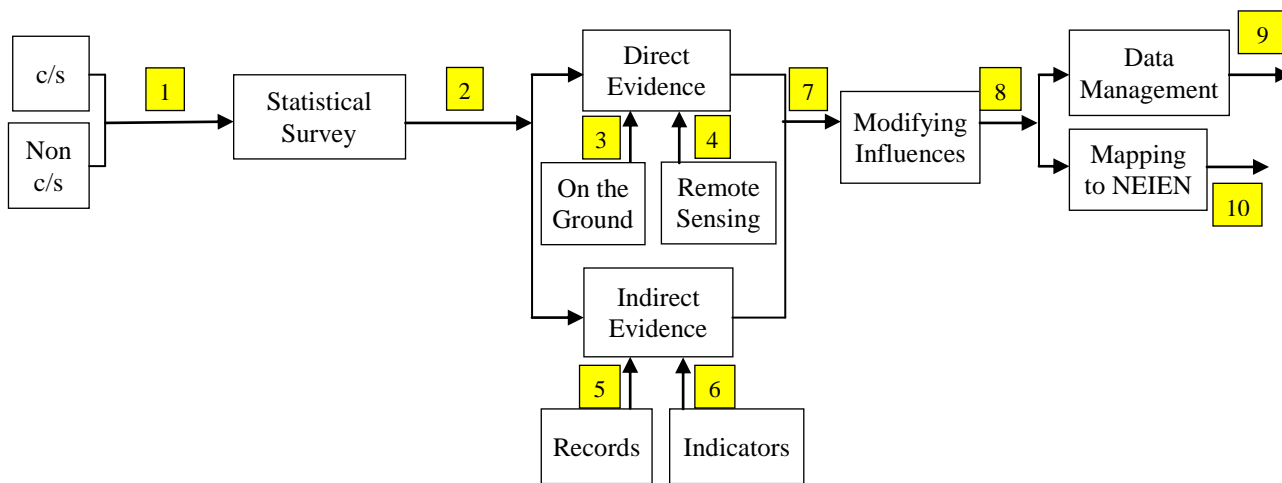
Error term	Description
1	Uncertainty whether both c/s and non c/s BMPs are detected.
2	Confidence in quality of agency records documenting BMPs.
3	Confidence in quality of other organization documenting BMPs.
4	Confidence in quality of landowner self-reporting of BMPs.
5	Confidence in spot-checking approach.
6	Confidence that on-the-ground observation adequately determines BMP function.
7	Confidence that remote sensing data adequately characterize BMP functionality.
8	Confidence that records (e.g., NMP) capture BMP function.
9	Confidence that indicators support assumption of BMP function.
10	Overall confidence in scoring of BMP functionality.
11	Confidence that modifying influences (e.g., season, timing observations) are accounted for.
12	Uncertainty in effectiveness of data management to accurately record BMP data.
13	Uncertainty in mapping state/local BMP to NEIEN.

The pathways illustrated in Figure 2 include the option to use agency, landowner or other records/reporting for detection, as well as a choice between direct or indirect evidence of function. Both the direct (on-the-ground or remote sensing) and indirect (records or indicators) evidence options for function also include options. Figure 2 identifies twelve unique (10 is the combined or overall error or confidence associated with assessment of functionality) error sources associated with using records and reporting to verify BMP implementation. As for direct observation, the magnitude of these various error sources is likely to differ for each BMP (e.g., agency records and reporting may be more consistent than landowner records and reporting, but may be less detailed in some cases), and will vary for different BMPs (e.g., records are likely to be a better source of information for nutrient management than is remote sensing).

Figure 3 for statistical surveys does not include spot checking but well-designed statistically-based sampling approaches will include QA/QC procedures that serve essentially the same purpose. Only ten unique sources of error are shown in Figure 3, but it should be noted that multiple factors determine the confidence or error (error term 2) associated with a statistically-based sampling approach, including sampling design, sample size, and methods used for detection of BMPs at sampled sites.

**Table 3. Records and reporting scoring.** Ratings and scores are examples of highest level of confidence.

Rate confidence in results of operation on a scale of 1 (lowest) to 5 (highest)					
Where a choice is required ("OR" statement), select and rate only one operation					
<b>Sector</b>	<b>Error term</b>	<b>Operation</b>	<b>Rating</b>	<b>Weighting</b>	<b>Component Score</b>
<b>BMPs</b>	1	c/s & non c/s	5	4	20
<b>DETECTION</b>					
	2	agency	5	3	15
		OR			
	3	other		3	0
		OR			
	4	landowner		3	0
	5	Spot checks	5	3	15
<b>FUNCTION</b>					
	Direct Evidence				
	6	on-the-ground	5	2	10
		OR			
	7	remote sensing		2	0
		OR			
	Indirect Evidence				
	8	records		2	0
		OR			
	9	indicators		2	0
	10	Scoring of Function	5	2	10
	11	Modifying Influences	5	2	10
<b>REPORTING</b>					
	12	Data Management	5	2	10
	13	Mapping to NEIEN	5	2	10
				<b>TOTAL</b>	100



**Figure 3. Pathway and error sources for statistical surveys.** Numbered boxes denote steps where error or confidence can be evaluated.

**Key to Figure 3.**

Error term	Description
1	Uncertainty whether both c/s and non c/s BMPs are detected
2	Confidence/error in statistical sampling (e.g., NASS)
3	Confidence that on-the-ground observation adequately determines BMP function
4	Confidence that remote sensing data adequately characterize BMP functionality
5	Confidence that records (e.g., NMP) capture BMP function
6	Confidence that indicators support assumption of BMP function
7	Overall confidence in scoring of BMP functionality
8	Confidence that modifying influences (e.g., season, timing observations) are accounted for
9	Uncertainty in effectiveness of data management to accurately record BMP data
10	Uncertainty in mapping state/local BMP to NEIEN

The framework described here is intended to illustrate an approach that could be used to objectively and consistently rate the various approaches that are and could be used by CB states to verify implementation of agricultural practices. This framework includes a broad accounting of potential error sources to help establish proper context when determining the relative importance and magnitude of errors associated with the detection and functional assessment of BMPs. Although this framework was developed for consideration by the AgWG, the general nature of the terminology used should facilitate application to other source sectors in the CBW.

Several issues for the application of this framework should be noted. First and foremost, the selection of the proper scoring scheme (direct observation, records/reporting, or statistical survey) is critical; the best option may not always be obvious, especially in the case of hybrid approaches. In the case of the Black Creek, IN retrospective BMP assessment (Bracmort et al. 2004 and 2006, Section 4.1.1), for example, assessment of BMP status was based on direct observation, while the sites to be assessed were selected at random from a population of all BMPs implemented. Although it is possible to interpret this as a case of direct observation, the statistical survey scoring scheme would be the preferred choice because all the BMPs assessed were drawn as a sample and the results are intended to be extrapolated to the population. It is suggested that all situations in which the BMPs verified are selected as a sample of a larger population be evaluated as a statistical survey, while situations where all BMPs in a watershed or region are verified (i.e., a census) should be evaluated as either direct observation or records/reporting.

Secondly, it is strongly recommended that each assessment using this framework be done by more than a single rater, as ratings are strongly based on best professional judgment, which will naturally vary among individuals. A panel of 3 to 6 raters may be a good choice. At the outset, some procedure for arriving at a consensus rating

**Table 4. Statistical survey scoring.** Ratings and scores are examples of highest level of confidence.

Sector	Error term	Operation	Rating	Weighting	Component Score
BMPs	1	c/s & non c/s	5	4	20
<b>DETECTION</b>					
	2	Statistical sampling	5	6	30
<b>FUNCTION</b>					
	Direct Evidence				
	3	on-the-ground	5	2	10
		OR			
	4	remote sensing		2	0
		OR			
	Indirect Evidence				
	5	records		2	0
		OR			
	6	indicators		2	0
	7	Scoring of Function	5	2	10
	8	Modifying Influences	5	2	10
<b>REPORTING</b>					
	9	Data Management	5	2	10
	10	Mapping to NEIEN	5	2	10
				<b>TOTAL</b>	100

should be worked out. This may range from a simple averaging of all scores to an interactive process where individuals adjust their ratings to achieve consensus.

Thirdly, it is critical to document assumptions and individual interpretations supporting individual ratings. Such documentation will help justify individual scores and help assure repeatability of results.

Finally, if this framework is applied across multiple jurisdictions over the long term, it would be important to perform periodic cross-checks among different groups of raters (e.g., among states) to ensure comparability and consistency of ratings across the Bay watershed.