TETRA TECH

To:	Eric Hughes
From:	Brian Pickard, Aileen Molloy
Date:	April 2, 2025
Subject:	Evaluation of PA DEP Draft 2025 Remote Sensing BMP Verification Methodology

Overview

In 2025, the Pennsylvania Department of Environmental Protection (PA DEP) piloted a remote sensing methodology to verify agricultural Best Management Practices (BMPs)—specifically reduced tillage, conservation tillage, and high residue tillage management—in the Chesapeake Bay watershed. This approach uses satellite imagery and machine learning to identify conservation tillage practices across large areas, aiming to improve efficiency and coverage compared to traditional field-based verification methods. The Chesapeake Bay Program (CBP) established guidelines in 2017[1] for verifying BMP implementation using remote sensing. These 2017 Remote Sensing Verification guidelines set forth the required accuracy standards, validation procedures, and quality assurance steps that any remote-sensing based BMP verification method should follow.

The purpose of this document is to evaluate Pennsylvania's draft remote sensing verification methodology against the CBP's 2017 guidelines, to determine if the state's approach complies with each element of the guidance. The findings are based on Tetra Tech's review of the *Pennsylvania Department of Environmental Protection Remove Sensing Pilot Project: Draft Methodology Guidance for Remote Sensing Verification of Conservation Tillage BMPs*, submitted January 13, 2025, and prepared by Resolve Hydro LLC. The evaluation is organized in a threshold-based, criterion-by-criterion manner, checking each key requirement from the 2017 guidance and noting whether PA's implementation meets it, with explanations for any gaps. Based on the findings, recommendations are provided to address any shortcomings in PA's methodology and provide optional suggestions for how the methodology may be better aligned with current industry standards in remote sensing. An appendix is included with separate recommendations for improving the CBP's 2017 remote sensing verification guidelines themselves, for future consideration.

Method Summary

PA's method is model agnostic in that it does not prescribe a particular remote sensing model, but rather outlines the requirements that the selected model must meet, including distinguishing between the various conservation tillage BMPs, providing adequate spatial resolution (60 m or less), including transparent model documentation, meeting performance evaluation criteria, and separating training and evaluation data.

The methodology provides an approach for estimating a minimum sample size using a multinomial distribution and selecting survey areas, as described in the 2017 Remote Sensing Verification guidelines. The methodology also describes an approach for selecting overlapping ground truthing locations. The survey implementation includes a data collection period of March to the end of June to accurately capture the period when soil disturbance conditions can be detected immediately

[1] Norfleet et al. 2017. Draft Final Recommendation Report for the Establishment of Uniform Evaluation Standards for Application of Remote Sensing to Identify and Inventory Agricultural Practices for the Chesapeake Bay Program Partnership's Watershed Model. Prepared for Chesapeake Bay Program. Annapolis, MD.

after planting. The method suggests documenting the link between ancillary data layers and conservation tillage to support the model - adhering to CBP guidelines for ground truthing, avoiding double counting by ensuring each pixel is only assigned to one tillage regime and ensuring that those implementing the model and data collection have appropriate technical knowledge and skills.

For analysis and reporting the method uses a pixel-scale confusion matrix to compare remote sensing data with ground-truth observations. The method then prescribes the procedures in the Roadside Transect Surveys verification guidelines[2] for calculating the unbiased estimates of true marginal proportions of each tillage regime with confidence intervals, producer's accuracy and variance, user's accuracy and variance and the overall accuracy and variance, followed by scaling for the estimated proportions of estimated cropland to derive the total estimated acreage of conservation tillage in the evaluation area. Alternatively, the method suggests using the ratio of the user's accuracy to the producer's accuracy to correct for bias in the total conservation tillage BMP implementation areas, using the methodology in the 2017 Remote Sensing Verification guidelines. This second method is recommended if the overall accuracy is at least 71% and the lower confidence limit values for producer's accuracy and user's accuracy are at least 70%. The method concludes with the recommendation to use the proportions from the analysis if the lower confidence limit on the overall accuracy exceeds 50% but recommends trying to achieve a higher level.

Evaluation of PA Methodology Against CBP's 2017 Guidelines

The CBP's 2017 remote sensing verification guidance for BMPs outlined several key criteria to ensure that any remote-sensing based method produces reliable and acceptable results. Below is a crosswalk of each major guideline element versus the PA DEP 2025 pilot methodology and brief explanation:

2017 CBP Guideline Element	**PA 2025 Method Compliance**	**Notes**
Use of standard accuracy metrics – Utilize confusion-matrix statistics (e.g. Hit Rate (producer's accuracy), False Alarm Rate, overall accuracy) for performance evaluation.	Yes	The PA methodology uses a confusion matrix to derive Producer's Accuracy (Hit Rate), False Alarm Rate (User's accuracy), and overall classification accuracy, consistent with CBP's standard accuracy metrics.
Minimum accuracy threshold (71% overall) – Demonstrate that the remote sensing model achieves 71% or higher overall accuracy (as a minimum acceptable level for BMP identification) and a lower confidence limit of at least 70%	No	The PA methodology applies a lower confidence limit of 50%; however, this is derived from the Roadside Transect Surveys verification guidelines and is not consistent with the Remote Sensing verification guidelines. Using the verification guidelines from another method is not appropriate and the values from the 2017 Remote Sensing Guidelines need to be applied.
Confidence interval and statistical significance – Compute confidence intervals for accuracy and ensure the lower 90% confidence bound of the accuracy is above the required threshold (to demonstrate statistical confidence in the result).	Yes	Pennsylvania's method explicitly calculates accuracy confidence intervals and applies CBP's acceptance criteria that the lower bound of the 90% confidence interval for accuracy must exceed the selected threshold.

[2] Dressing and Harcum. 2017. 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. Prepared for Chesapeake Bay Program. Annapolis, MD.

2017 CBP Guideline Element	PA 2025 Method Compliance	Notes
Ground-truth validation sample – Perform field verification using a representative sample of sites (approximately 20 or more, strategically selected) to validate the remote sensing results, rather than relying on a fixed percentage of sites.	Yes	The PA DEP pilot includes a field sampling component that aligns with the 2017 guidance. Rather than attempting to visit 10% of all identified fields (which would be impractical), PA selects on the order of 20+ ground truth sites distributed across the region to verify the satellite classifications. This approach follows CBP recommendations that a well-chosen sample (around 20 or more points) can suffice for accuracy assessment, provided the sample represents the variability of conditions.
Bias correction of results – Adjust the reported BMP implementation results for any classification biases, using the error rates from the confusion matrix (i.e. apply bias correction if the remote sensing systematically over- or under-detects BMPs).	Yes	Pennsylvania's methodology plans for bias correction based on field verification findings. After constructing a confusion matrix from the ground-truth comparison, the PA team can adjust (bias-correct) the remote sensing results if systematic errors are detected. This means if the satellite model missed some BMPs or had false positives at a consistent rate, the final reported counts can be corrected using statistical adjustments, as per CBP guidance.
Independent validation data – Ensure the data used for accuracy assessment is independent of the data used to train or develop the model (to avoid overestimating accuracy).	Yes	The PA pilot methodology keeps training and validation separate. It uses independent ground truth data for the accuracy evaluation, which was not used in training the remote sensing model, thereby yielding an unbiased accuracy estimate. This practice is in line with the guidelines' intent to have a truly independent accuracy assessment.
Adherence to CBP BMP definitions and reporting requirements – Use CBP-approved BMP definitions/classifications and report all required attributes for verified practices (e.g. location, county or watershed, practice type, acreage, year of implementation).	Yes	The PA approach was developed in coordination with CBP partners and adheres to the standard definitions for conservation tillage BMPs. Each identified "conservation tillage" instance from the remote sensing is recorded with its location (e.g. county or watershed), the area (acres) under the practice, the year of implementation, and is classified according to the CBP's categories of tillage practice. This means the output of PA's verification method can be seamlessly integrated into CBP's reporting system, fulfilling documentation requirements.

2017 CBP Guideline Element	PA 2025 Method Compliance	Notes
Separate verification for each BMP type (no "composite" BMP credit) – Verify and report each BMP type independently. If multiple BMPs are assessed, each must meet accuracy and confidence criteria on its own; results for different practices should not be combined into a single aggregate credit or accuracy measure.	Unclear	It is unclear whether the three conservation tillage BMPs are being treated as separate BMPs and evaluated separately within the analysis. Each BMP needs to be treated independently (e.g., independent sampling data for validation, statistical confidence intervals, and overall accuracy assessment. *Applicable if the method is expanded beyond conservation tillage BMPs.* If Pennsylvania applies this remote sensing approach to multiple BMPs in the future, it will need to evaluate each practice separately. The CBP does not recognize composite or pooled credit for multiple BMP categories – for example, a combined accuracy for "tillage + cover crops" together would not be acceptable. Each BMP (e.g. tillage, cover crops, etc.) would require its own verification statistics and must individually meet the guidance criteria. PA's methodology should be prepared to maintain distinct accuracy assessments for each practice type to remain compliant with CBP requirements.

Overall, Pennsylvania's 2025 remote sensing verification methodology aligns well with the majority of the CBP's 2017 remote sensing verification guidelines. The approach uses the recommended statistical methods, incorporates required confidence checks, and follows the prescribed validation procedures and data reporting standards. These areas of compliance indicate that PA's method was largely designed with the CBP criteria in mind, which is appropriate for eventual acceptance of the data in the Bay Program accounting. The notable exceptions identified are: (1) an incorrect reference to a *"50% accuracy" threshold*, which is not part of the remote sensing guidance (this needs correction to the proper ~70% threshold), and (2) the consideration of separate verification for each BMP type. The second point is an important requirement to plan for as the methodology scales up – PA must ensure it does not combine multiple BMPs into one verification result. Apart from these issues, the PA DEP remote sensing methodology appears to be in line with CBP's 2017 expectations.

Recommendations for PA Methodology Improvements

Based on the evaluation above, the following **targeted recommendations** are offered to enhance Pennsylvania's remote sensing BMP verification methodology and address the identified shortcomings. These recommendations focus on improving PA's approach itself (the implementation of the verification method), ensuring it fully meets the CBP guidance and is robust and credible for long-term use:

- *Use the Correct Accuracy Threshold:* Remove any reference to a 50% accuracy threshold in the methodology, as it is not relevant to remote sensing verification. PA should explicitly adopt the CBP-recommended accuracy benchmark (approximately 70% minimum overall accuracy, with the corresponding confidence requirement) as the standard for success. In practice, this means

PA's remote sensing model should continue to be calibrated to achieve at least ~70% accuracy or higher, and the verification reports should document that this criterion is met with 90% confidence. Eliminating the 50% figure (which comes from a different framework) will avoid confusion and ensure PA's results are judged by the correct standard.

- *Ensure Separate Verification for Each BMP (No Composite Reporting):* As the remote sensing program expands, PA should verify each type of BMP independently and report accuracy and confidence levels for each one. The statistical analysis needs to be clearly articulated for each conservation tillage BMP. Separately, the current pilot is only for conservation tillage, but if in the future PA applies remote sensing to other practices (for example, cover crops or nutrient management), it must treat them as separate verification projects. Do not combine multiple BMPs into a single "composite" accuracy assessment or credit. Each practice should have its own ground-truth sample, confusion matrix, accuracy calculation, and bias correction (if needed). This will ensure that every BMP meets CBP's verification requirements on its own merits. Planning for this now will make the methodology scalable to other practices while staying compliant.
- *Refine Ground-Truth Sampling Design:* Continue to carefully design the field sampling strategy each year to ensure it captures the full range of farming conditions and practices in Pennsylvania. Because the verification uses a relatively small number of ground-truth sites (on the order of 20–30), each site must be strategically selected. PA should stratify or distribute the sampling to cover different geographies, crop types, and management conditions (e.g. various soil types or tillage equipment) so that the accuracy estimate is truly representative. This will minimize the risk of bias in the verification. If analysis shows high uncertainty (for example, if the confidence interval is close to dropping below the threshold), PA should consider increasing the sample size modestly or refining the sampling approach in those areas to strengthen the results. In summary, optimize the field verification sampling to maintain confidence in the accuracy findings.
- *Implement Bias Corrections in Practice:* When the field verification reveals any systematic bias (for instance, if the remote sensing consistently misses a certain fraction of conservation tillage fields or falsely labels some conventional tillage as no-till), PA should apply the necessary bias correction to its results. This involves adjusting the final reported count of verified BMP acres using the error rates from the confusion matrix. Doing so will improve the accuracy of the reported BMP implementation numbers and ensure that Pennsylvania's crediting in the Bay Program is fair and not skewed by any known biases in the remote sensing model. PA has included this step in the methodology design; the recommendation is to follow through with it diligently each year the program is run. All bias-adjustment calculations should be documented and reviewed to maintain transparency.

By implementing these recommendations, Pennsylvania can strengthen its remote sensing verification program. The goal is to ensure the methodology not only meets the 2017 CBP guidelines in letter and spirit, but also is reliable and defensible as a tool for tracking BMP implementation. Over time, a robust remote sensing verification approach will help Pennsylvania confidently report conservation tillage (and potentially other practices) with reduced field labor, while maintaining the credibility and accuracy required by the Chesapeake Bay Program.

Appendix: Recommendations for Improving CBP's 2017 Remote Sensing Guidelines

The following suggestions are outside the scope of evaluating Pennsylvania's current methodology, and instead address the CBP's 2017 remote sensing verification guidelines themselves. These are offered as potential improvements for the CBP to consider in future updates to the verification framework, especially in light of technological advances and lessons learned since 2017.

1) *Update Guidance to Incorporate Technological Advances:* The 2017 remote sensing verification guidelines should be revisited to reflect the significant advancements in remote sensing technology and data availability. Since 2017, new satellite platforms (e.g. high-resolution constellations like PlanetScope) and improved analytical techniques (e.g. machine learning and cloud computing) have become accessible and proven in pilot projects. The guidelines could be updated to encourage the use of higher-resolution imagery, multi-temporal data, and advanced classification algorithms that can greatly enhance BMP detection accuracy and coverage. In 2017, achieving ~70% accuracy might have been an ambitious target given the tools available; today, accuracies above 85% have been demonstrated with modern methods. The CBP should incorporate these developments, potentially setting more ambitious accuracy goals or at least acknowledging that better performance is now attainable. This update would make the guidance more relevant and effective in the current technological context.

2) *Improve Clarity on Accuracy Requirements and Framework:* The CBP should clarify the accuracy verification requirements to prevent any confusion between different verification frameworks. For instance, the current remote sensing guidance expects ~70% accuracy with certain confidence statistics, whereas other BMP verification pathways (like some field-based or inspection protocols) might use different metrics or thresholds (e.g. a 50% rule for something else). The updated guidelines should explicitly state the required accuracy threshold for remote sensing methods and that it is distinct from any other framework's criteria. Clearly distinguishing these will help jurisdictions avoid misapplying the wrong thresholds. Additionally, the guidance could spell out what happens if the accuracy criterion is not met (e.g. whether partial credit can be given or results must be discarded) – the 2017 document focused on achieving the goal, but providing direction for outcomes below the threshold would add clarity for edge cases.

3) *Emphasize Separate Verification for Different BMPs:* To address potential misunderstandings, the CBP should explicitly note in the guidance that each BMP type must be verified independently when using remote sensing. The 2017 guidelines were written for conservation tillage as a case study, but as remote sensing is expanded to other practices, it's important to state that you cannot combine multiple BMPs into one verification exercise or one accuracy calculation. For example, if a state uses one satellite analysis to identify both cover crops and tillage practices, the verification of each practice must be reported separately with its own accuracy stats; there is no concept of a "composite BMP" accuracy in the CBP accounting. Making this clear in the guidelines will prevent any attempts to pool results and will maintain the integrity of each BMP's reported outcome.

4) *Incorporate Efficient Sampling Strategies:* The updated guidelines should incorporate lessons learned about field verification sampling efficiency. The original guidance mentioned a 10% field re-check or a minimum of ~20 points for accuracy assessment. Now, with more experience, CBP can provide refined advice on sampling. For instance, it could endorse a stratified random sampling approach with a minimum number of validation points (e.g. 20–30) that must capture the variability of the landscape. It could also note that simply using a flat percentage (like 10%) might not be necessary if a smaller well-planned sample achieves the desired confidence interval. Offering a formula or table relating sample size to confidence levels could help jurisdictions plan their field work more efficiently (so they collect neither too few nor unnecessarily many points). This would make verification efforts more efficient without compromising statistical rigor.

5) *Establish a Formal Remote Sensing Verification Protocol and Support System:* One major improvement on the CBP process side would be to formally adopt remote sensing verification as an approved method across the Chesapeake Bay jurisdictions. CBP could develop a standardized protocol or toolkit based on successful pilots (such as Pennsylvania's) and provide it to all states in the watershed. This protocol would outline the required steps, metrics, and QA/QC checks (essentially a refined version of the 2017 guidance, updated for current technology). Additionally, CBP could consider setting up a central support system – for example, a regional data hub for satellite imagery or a technical assistance team – to help states implement the protocol. This would ensure consistency in how remote sensing is used for BMP verification and help states with fewer resources to participate. By institutionalizing the approach, the CBP partnership can achieve basin-wide improvements in BMP reporting. In short, making remote sensing verification a formal part of the CBP guidance and providing shared support would enhance adoption and consistency.

6) *Address Outdated Assumptions and Data Limitations:* The revised guidelines should acknowledge and provide solutions for some limitations that were more acute in 2017. For example, cloud cover or image availability was a concern – now, with more satellites and even radar options (e.g. Sentinel-1), the guidance could mention using multi-source data (optical and radar) to mitigate weather-related data gaps. The guidelines could encourage approaches like data fusion or multi-date analysis to improve reliability. Another outdated assumption might be the cost: high-resolution imagery was expensive in 2017, but costs have come down or free sources (Sentinel-2, Landsat) are available; the guidance can update advice on how to obtain imagery cost-effectively. By updating these sections, the CBP can remove barriers that were noted in 2017 and empower jurisdictions to use the best available data.

7) *Provide Regular Updates and Foster Continuous Improvement:* Finally, the CBP should commit to periodically reviewing and updating its remote sensing verification guidelines. Given the rapid pace of innovation in remote sensing, what is state-of-the-art today may become outdated in a few years. Setting a schedule or trigger for guideline review (for example, every 3–5 years, or when major advancements occur) would help keep the verification framework current. In these updates, CBP can incorporate feedback from pilot projects, address any persistent areas of confusion, and raise or adjust standards as appropriate. The 2017 guidance was an excellent

foundation; a 2025 or 2026 update could build on that by incorporating everything learned in the intervening years. Regular updates ensure the guidelines remain relevant, effective, and aligned with best practices, ultimately strengthening the overall BMP verification and reporting process for the Chesapeake Bay Program.

By implementing these improvements, the Chesapeake Bay Program's remote sensing verification framework can become more clear, up-to-date, and efficient. This will support broader adoption of remote sensing techniques for BMP verification across the watershed, leading to more consistent and credible BMP reporting. As technology and methods continue to evolve, a proactive approach to guideline maintenance will help the CBP partnership stay at the forefront of verification science and ensure that the data underpinning water quality efforts are as accurate and reliable as possible.