Accuracy assessment of Pennsylvania streams mapped using LiDAR elevation data: Method development and results

David Saavedra, geospatial analyst

Louis Keddell, geospatial analyst

Michael Norton, geospatial program manager

Cassandra Pallai, geospatial program manager

February, 2018

1. Introduction

a. Purpose

The purpose of this assessment was to quantify the positional accuracy of streams mapped in the Pennsylvania portion of the Susquehanna River watershed using LiDAR-derived digital elevation data. This information, in conjunction with accuracy rates of the High-Resolution Land Cover Dataset (Claggett, et al., 2018), will be valuable in establishing a degree of confidence in the number of restorable acres reported through the Restoration Opportunity Areas Analysis conducted by the Chesapeake Conservancy (Saavedra, et al., 2018).

b. Overview of assessment

At its core, accuracy assessment is an evaluation of the positional/thematic accuracy of mapped spatial data against data which are regarded to be more accurate, known as reference data. Reference data often include field-collected data or aerial imagery. A review of the current literature on accuracy assessments revealed that many methodologies were aimed at assessing the accuracy of thematic maps such as land cover classifications. These assessments identify samples in reference data and assess the accuracy of the mapped dataset at those sample locations. The result is a table containing several accuracy metrics, including overall accuracy, producer's accuracy, user's accuracy, and rates of omission and commission errors (Congalton & Green, 2008).

Due to fundamental differences between a land cover map which classifies every pixel in a landscape into one of several classes and a stream map which only classifies streams and contains no data elsewhere, conventional accuracy assessment methodologies are not directly portable. For this reason, existing methods had to be adapted to assess the positional accuracy of mapped streams.

This assessment used a subset of mapped stream segments as samples and evaluated whether or not a corresponding stream existed in the reference imagery (described in section 2a). Streams that were "missed" (i.e. streams that were visible in the reference imagery but not included in the mapped stream dataset) were not assessed. Because of this, the assessment is only capable of reporting on user's accuracy, defined as "the probability that a sample from [a classified map] actually represents that category on the ground" (Story & Congalton, 1986), and its inverse, commission error rate. No statement can be made on producer's accuracy, defined as "the probability that a reference (ground) sample will be correctly classified [in the map]" or its inverse, omission error rate (Story & Congalton, 1986). See the confusion matrix below for a visual representation of the assessment:

	Actual positive *Stream visible in reference imagery*	**Actual negative** *Stream not visible in reference imagery*
Predicted positive *Stream mapped in dataset*	True positive *Stream accurately mapped*	False positive *Error of commission*
Predicted negative *Stream not mapped in dataset*	Not assessed	Not assessed

Table 1: Confusion matrix used in assessment.

2. Data preparation

a. Data used

The stream dataset that is the subject of this accuracy assessment was developed in 2016 using LiDAR-derived digital elevation (DEM) data. It was created by compiling the best-available LiDAR DEMs for the New York, Pennsylvania, and Maryland portions of the Susquehanna River watershed and merging them into one seamless DEM, a raster with a pixel resolution of 1 x 1 meters. This seamless DEM was hydrologically conditioned using a geoprocessing operation known as *breaching* (see section 4c for more information). Flow direction and flow accumulation layers were created from the hydro-conditioned DEM. A contributing area threshold of 60 acres was applied to the flow accumulation layer to delineate the stream network such that any pixels with greater than or equal to 60 acres of land area draining to them were included in the stream network and pixels with fewer than 60 acres of contributing area were excluded from the stream network. This raster version of the stream network was then converted to polyline format which was smoothed for aesthetic and cartographic purposes and ultimately used in this assessment (Saavedra, et al., 2018).

The reference data used in this assessment were high-resolution digital orthoimages collected over the state of Pennsylvania from 2003 to 2006 and combined into a single, seamless, statewide layer. This layer is available as an ArcGIS map service, hosted by Penn State University[1]. This dataset was chosen as the reference dataset for several reasons. The first and primary reason was that these aerial images appear to have been collected during leaf-off conditions, allowing for better identification of features on the ground, including streams. The second reason was that the images were collected at the same time as, or slightly before, the collection of the majority of the LiDAR data used in the creation of the stream network dataset. This is important as it means that features reflected in the LiDAR data, and thus the stream dataset, are also likely to be present in the aerial imagery; if the imagery and LiDAR were collected many years apart, the possibility arises that human modification of the landscape may result in differences between the features reflected in the LiDAR and the imagery. The third reason for using this reference dataset was a matter of convenience and consistency; having the images available from one source in a seamless, statewide layer made the data very easy to work with and provided a degree of consistency that would not be present if the reference data were compiled from several sources.

b. Classification of streams

The details of sampling methods will be explained further in Section 3, but it is worth noting here that in order to ensure even distribution of random samples from streams of all sizes, the stream network dataset was broken up into five separate classes based on contributing area. While Strahler stream order (Strahler, 1957) may have been a more appropriate way to classify the streams, this metric was not computed during the initial creation of the stream dataset and it was not practical to retroactively calculate this information. Instead, a workable proxy was created by classifying the stream dataset into five classes based on contributing area using the geometric interval classification method in ArcGIS.

---

[1] Metadata for the map service is available at:
http://www.pasda.psu.edu/uci/FullMetadataDisplay.aspx?file=AerialPhotoColor_cached.xml

The resulting classification is below:

Class 1*: contributing area from 242,814 – 442,946 square meters*

Class 2*: contributing area from 442,947 – 2,426,346 square meters*

Class 3*: contributing area from 2,426,347 – 22,082,752 square meters*

Class 4*: contributing area from 22,082,753 – 216,886,745 square meters*

Class 5*: contributing area from 216,886,746 – 2,147,483,646 square meters*

In addition to the classification above, only streams greater than 20 meters in length were considered for assessment. This eliminated the assessment of small polyline fragments which are not truly representative of actual stream segments, but rather are a result of the conversion from raster to polyline during the creation of the stream network dataset.

3. Sampling

a. Adaptation of National Standard for Spatial Data Accuracy standards

In 1998 the Federal Geospatial Data Committee established the National Standards for Spatial Data Accuracy (NSSDA) to provide a standardized and statistically-based method for assessing the positional accuracy of geospatial data (FGDC, 1998). The NSSDA sets forth several requirements for positional accuracy sample selection, each of which are satisfied in this assessment. One of these requirements is that the reference data must be independent of the data that was used to create the geospatial data being assessed and must be of higher accuracy. This requirement is satisfied by using high-resolution aerial orthophotos as a reference dataset. The NSSDA also requires a minimum of 20 samples to be assessed. The assessment developed for this report uses 100 samples per stream class for a total of 500 samples.

Furthermore, the NSSDA requires that samples be well-defined and easily visible in the reference data in addition to being well-distributed throughout the geographic project area and reflective of the possible distribution of error in the dataset. The following stratified random sampling approach was used to satisfy these requirements:

1. Classify stream dataset into five classes as described in Section 2b
2. Assign a random number to every stream in each class
3. In ascending order based on random number, perform visual inspection of streams to identify 100 qualifying samples from each of the five classes

In order to qualify as a sample, the ground needed to be clearly visible in the reference imagery at the location of the mapped stream and the analyst performing the inspection needed to be able to determine the presence or absence of a stream in the reference imagery with a high degree of confidence. This sampling approach was done for each of the five stream classes for a total of 500 samples.

4. Assessment methods

a. Assessment overview

Once 100 sample streams from each of the five stream classes had been identified, the samples were exported into three separate but identical datasets so that three analysts could independently review the samples. Each of the three analysts reviewed all 100 samples from each of the five classes, for a combined total of 1500 individual assessments.

The assessment process involved overlaying the mapped stream samples on the reference imagery and assessing whether the mapped stream represented a real stream on the ground (a true positive) or if the mapped stream did not represent a real stream on the ground (a false positive). The analysts followed a common set of guidelines regarding what constitutes true positive and false positive designations, but many cases required a discretionary decision to be made, which sometimes differed between analysts. Sections 4b, 4c, and 4d explain the assessment guidelines in further detail, as does the Appendix which contains images that accompany the guidelines written here. After three analysts independently reviewed all 500 samples, a final true positive or false positive designation was made for each of the 500 samples by siding with the majority rule of the analysts' decisions. To determine the user's accuracy for individual stream classes, the following equation was used:

$$\boldsymbol{UA_{class}} = \left(\frac{\boldsymbol{TP_{class}}}{\boldsymbol{TP_{class}} + \boldsymbol{FP_{class}}}\right) \times \mathbf{100}$$

Where $\boldsymbol{UA_{class}}$ is the user's accuracy for a given class of streams, $\boldsymbol{TP_{class}}$ is the total number of final true positive designations for the class, and $\boldsymbol{FP_{class}}$ is the number of final false positive designations for the class.

To calculate the user's accuracy across all stream classes, the following equation was used:

$$\boldsymbol{UA_{all}} = \left(\frac{\boldsymbol{TP_{all}}}{\boldsymbol{TP_{all}} + \boldsymbol{FP_{all}}}\right) \times \mathbf{100}$$

Where $\boldsymbol{UA_{all}}$ is the user's accuracy for all stream classes, $\boldsymbol{TP_{all}}$ is the total number of final true positive designations across all classes, and $\boldsymbol{FP_{all}}$ is the number of final false positive designations across all classes.

Commission error was calculated as:

$$\boldsymbol{CE} = \mathbf{100} - \boldsymbol{UA}$$

Where $\boldsymbol{CE}$ is the commission error, and $\boldsymbol{UA}$ is the user's accuracy for an individual class or across all classes.

b. Guidelines for true positive designations

In many cases, there was unambiguous agreement between mapped stream samples and streams in the reference imagery; these cases would naturally receive a true positive designation. However, there were several common cases where the true positive designation was less obvious but still warranted. The most common of these less-obvious cases was that in which the mapped stream segment fell completely within water in the reference imagery. This occurred commonly when stream segments were mapped within ponds, lakes, or wide rivers. This situation was anticipated in the creation of the stream dataset and was addressed by merging the stream dataset with the “water” class from the High-Resolution Land Cover Dataset before identifying restoration opportunity areas along the edges of streams and waterbodies.

Another common case where a true positive designation was appropriate but not obvious occurred when a mapped stream segment was in agreement with a stream in the reference imagery, but the headward extent of the reference stream went beyond that of the mapped stream. Because this assessment was not designed to evaluate omission error, the “missed” portion of the stream would not be considered and the mapped stream would receive a true positive designation.

A third scenario often encountered was that in which visibility was obstructed along portions of a stream in the reference imagery. In this case, if a mapped stream was in agreement with the visible portions of the reference stream and there was no reason to believe it was inaccurate along the obstructed portions of the reference stream, the mapped stream would receive a true positive designation.

c. Guidelines for false positive designations

As was the case with the true positive designations, many of the false positive designations were very clear and unambiguous. These were instances where the mapped stream may lie over a parking lot, athletic field, interstate highway, etc. in the reference imagery. In these cases, it was quite clear that there was no stream present in the reference imagery, nor anything that could potentially be interpreted as a stream, and the mapped stream would receive a false positive designation.

A common scenario encountered was mapped streams that appeared to follow ephemeral washes, roadside ditches, and other areas of concentrated flow in the reference imagery. In these cases, unless a *distinct channel* was visible in the reference imagery, the mapped stream segment would receive a false positive designation – evidence of concentrated flow alone was not considered sufficient to denote the mapped stream as a true positive.

d. Guidelines for discretionary cases

As mentioned previously in Section 4a, in many of the cases it was not entirely clear whether or not a mapped stream was accurate against the reference imagery and the analysts had to use their best judgement to evaluate the accuracy of the stream segment. It was known prior to developing the accuracy assessment that it would be rare for mapped streams to align *exactly* with streams in the reference imagery. This led to the question of how much error is acceptable for streams to still be considered accurate. This determination was at the discretion of the analyst, and was evaluated on a case-by-case basis. In general, mapped streams were considered accurate if they were not gross misrepresentations of streams in the reference imagery. Thus, if a mapped stream did not follow the meanders of a reference stream exactly, or it was misaligned such that it ran parallel to a reference stream instead of directly on it, the mapped stream would still be considered accurate. This guideline applies within reason, however, and if the misalignment or misplacement of a mapped stream was so severe that the analyst determined it was not representative of the stream in the reference imagery, it would be considered inaccurate and designated as a false positive.

Another instance allowing for some acceptable degree of error was in the case of roads crossing streams. Because bridges, culverts, and other road crossings are higher in elevation than the streams which they cross, they act as dams in DEM data, blocking flow from the upstream side of the road to the downstream side. This was addressed through an automated pre-processing step called *breaching* (introduced in Section 2a) which carves an artificial channel through the road to connect the upstream and downstream portions of the stream channel. This allows water to flow freely “through” the road, where in reality it would flow under the road. In many cases this artificially carved channel, known as a *breach channel*, was carved straight through the road, allowing the stream to flow directly across the road. However, in some cases, the breach channel was carved in such a way that the stream was diverted along the road, eventually crossing it and reconnecting with the downstream channel in a location that was not immediately adjacent to the road. This type of error was known prior to the development of the accuracy assessment and, within reason, was not considered to negatively affect the accuracy of the mapped streams. In many cases the breach channel would only divert a mapped stream a short distance along a road and it would reconnect with the channel network shortly downstream of the road. In these cases, the analysts would consider the stream to be a true positive. However, there were occasions where a breach channel diverted the mapped stream a greater distance along a road, such that the stream segment reconnected to the channel network in a completely different location than the reference imagery would suggest. In these cases, analysts used their best discretion to decide if the breach error was severe enough that it misrepresented the stream in which case it would be designated as a false positive.

5. Results

The results of the accuracy assessment are presented in the table below:

Stream class	**Contributing area (sq. m.)**	**User’s Accuracy**	**Commission Error**
Class 1	242,814 – 442,946	57%	43%
Class 2	442,947 – 2,426,346	72%	28%
Class 3	2,426,347 – 22,082,752	86%	14%
Class 4	22,082,753 – 216,886,745	98%	2%
Class 5	216,886,746 – 2,147,483,646	98%	2%
All classes	242,814 – 2,147,483,646	82%	18%

Table 2: Accuracy assessment results. User’s accuracy is lowest in Class 1 and increases as contributing area increases.

6. Discussion

From the table in Section 5, it can be seen that the user’s accuracy is lowest (and consequentially, commission error is highest) for Class 1, the class consisting of streams with the smallest contributing area. As contributing area increases, user’s accuracy improves and commission error is reduced. This result was anticipated based on the methods used to create the stream dataset, as well as first-hand experience working with the dataset. The stream dataset was created using what is known as a constant contributing area threshold to delineate streams. In short, this method imposes a user-defined area threshold on flow accumulation layers to create a grid where cells with accumulation greater than the threshold are classified as part of the stream network, and cells with accumulation less than the threshold are not classified as part of the stream network. This method is commonly employed when delineating streams in a GIS environment because it is a straight-forward approach that benefits from working with raster layers such as DEMs, flow direction, and flow accumulation layers (O'Callagan & Mark, 1984).

Tarboton et al. (1991) suggest that extracted channel networks should closely resemble traditional cartographic representations of streams. Following this suggestion, a contributing area threshold of 60 acres was applied to the flow accumulation layer to extract a network with a channel density similar to that of the 1:24,000 scale National Hydrography Dataset (U.S. Geological Survey, 2007-2014). However, in the physical world, contributing area is only one of many factors affecting channel initiation; applying a constant contributing area threshold for channel initiation across a large geography will invariably result in over-inclusion/under-inclusion of headwater and lower-order streams in the extracted network, as reflected in the results of this assessment. As stream order and contributing area increase well beyond the initiation threshold, the impact of the threshold is reduced and the accuracy of these streams improves, a trend which can also be seen in Table 2.

7. Summary

A method for assessing the positional accuracy of streams mapped in the Pennsylvania portion of the Susquehanna River watershed was developed and implemented. Following guidelines published by the NSSDA, a total of 500 streams were selected as samples to be independently assessed by three analysts. The analysts' assessments were then used to calculate the user's accuracy and commission error of the stream dataset. Results showed that accuracy was poorest for low-order streams and improved steadily as stream size increased. These results are consistent with expectations based on the methods used to delineate the stream channels. Research is currently being conducted with the goal of more reliably delineating headwater and low-order streams.

8. References

Claggett, P. et al. (2018). Chesapeake Bay 2013/2014 High-resolution Land Cover Dataset. *Working paper.* Chesapeake Bay Program, Annapolis, MD.

Congalton, R. G., & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press.

Federal Geographic Data Committee (FGDC). (1998). Geospatial Positioning Accuracy Standards, part 3: National standard for spatial data accuracy. *Subcommittee for Base Cartographic Data, 25p*.

O'Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer vision, graphics, and image processing*, *28*(3), 323-344.

Saavedra, D., Pallai, C., Dean C. (2018). Susquehanna River watershed enhanced flow path and restoration opportunity area analysis. *Working paper.* Chesapeake Conservancy, Annapolis, MD.

Story, M., & Congalton, R. G. (1986). Accuracy assessment: a user's perspective. *Photogrammetric Engineering and remote sensing*, 52(3), 397-399.

Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union*, *38*(6), 913-920.

Tarboton, D. G., Bras, R. L., & Rodriguez-Iturbe, I. (1991). On the extraction of channel networks from digital elevation data. *Hydrological processes*, *5*(1), 81-100.

U.S. Geological Survey. (2007-2014). National Hydrography Dataset. Available online at *https://nhd.usgs.gov*, accessed 02/26/2018.

Appendix

The purpose of this Appendix is to provide a visual illustration of the sample selection and assessment guidelines described in Section 3 and Section 4. In each of the images below, the stream segment in question is highlighted in a bright cyan color, while other nearby streams are shown in a darker transparent shade of blue. In some examples, the reference imagery is shown both with and without the stream dataset overlaid – this was done to better illustrate what is on the ground underneath the mapped stream.

1. Section 3a example – sample not well-defined or easily visible

Ground is not clearly visible in reference imagery due to evergreen tree canopy and dark shadows, making it difficult to determine the presence or absence of a stream in the imagery. This example would not be included as a sample in accuracy assessment as it does not meet the well-defined and clearly visible requirements of the NSSDA.

2. Section 4b example – stream correctly mapped

Mapped stream unambiguously follows stream visible in reference imagery. This is example would receive a true positive designation.

3. Section 4b example – stream mapped in waterbody

Mapped stream lies completely within water. This situation was anticipated in the creation of the stream dataset and was addressed by merging the stream dataset with the "water" class from the High-Resolution Land Cover Dataset before identifying restoration opportunity areas along the edges of streams and waterbodies. This example would receive a true positive designation.

4. Section 4b example – stream mapped correctly, differing headward extent in reference

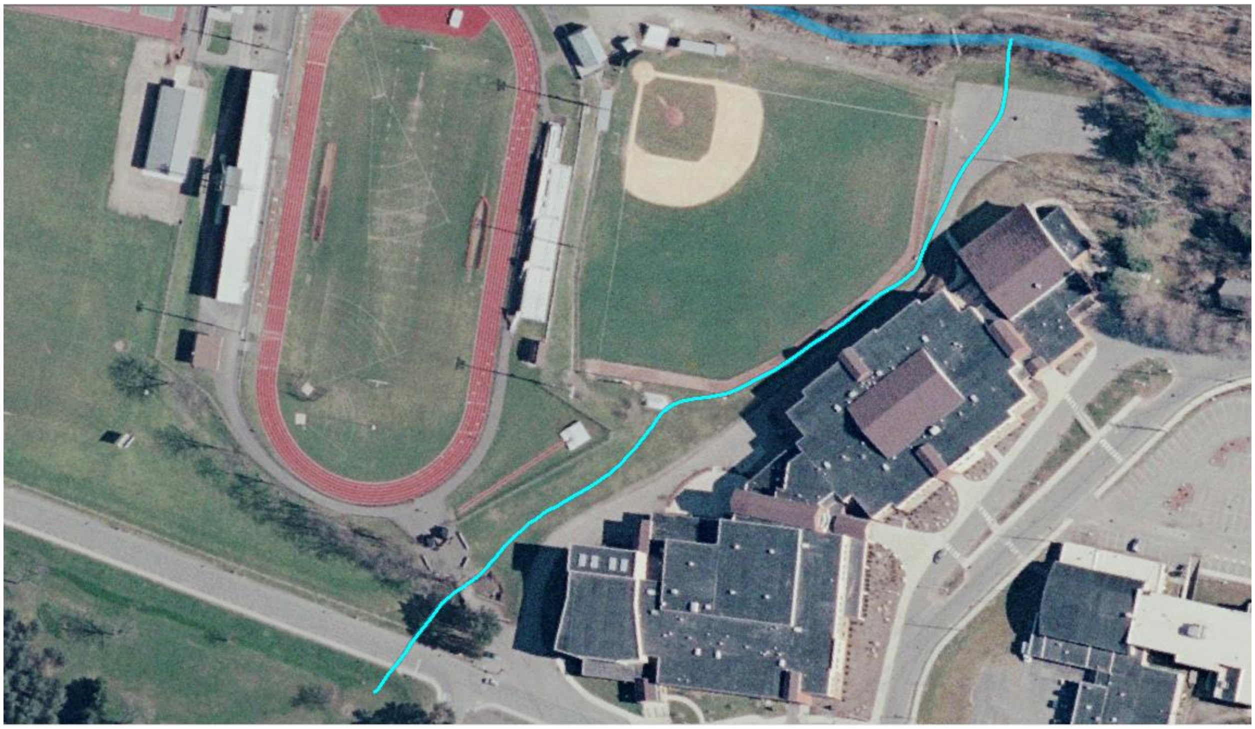
Mapped stream follows stream in reference imagery but the headward extent of stream in reference imagery exceeds that of the mapped stream. This example would receive a true positive designation as the mapped stream is accurate along its length.

5. Section 4b example – stream appears correct, visibility partially obstructed

Portions of a stream are visible in reference imagery along the length of the mapped stream, but other portions are obscured by tree canopy. In this instance, the mapped stream is accurate against the reference stream along the portions that are visible, and the surrounding context (headwaters to the west, confluence with larger stream to the east) suggest that the stream is mapped accurately. This example would receive a true positive designation.

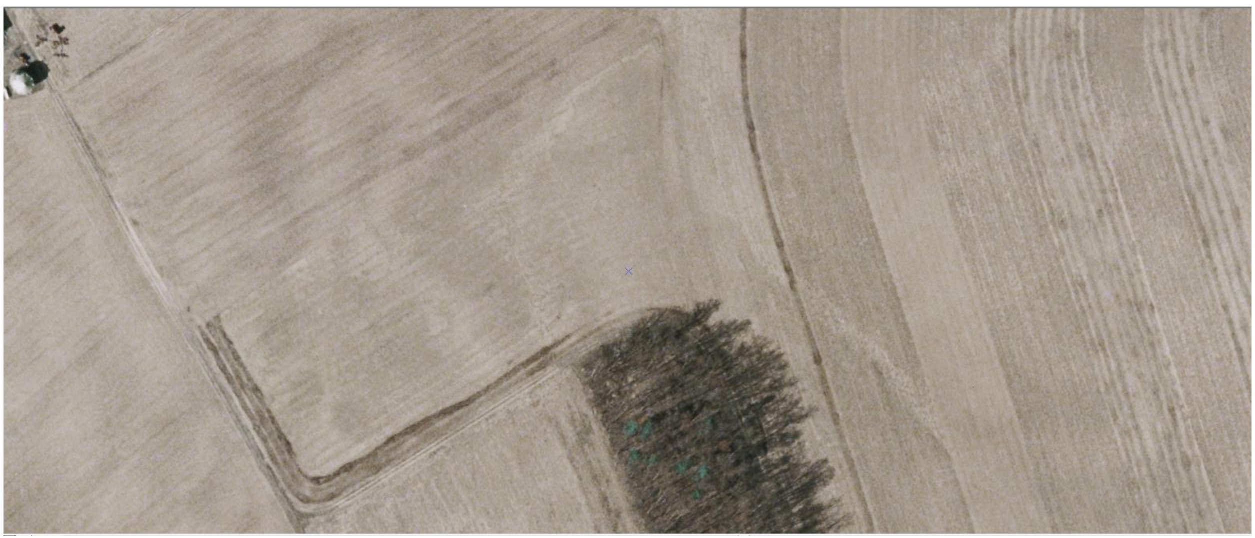
6. Section 4c example – stream mapped incorrectly

Mapped stream does not follow any stream in reference imagery. This is unambiguous and this example would receive a false positive designation.



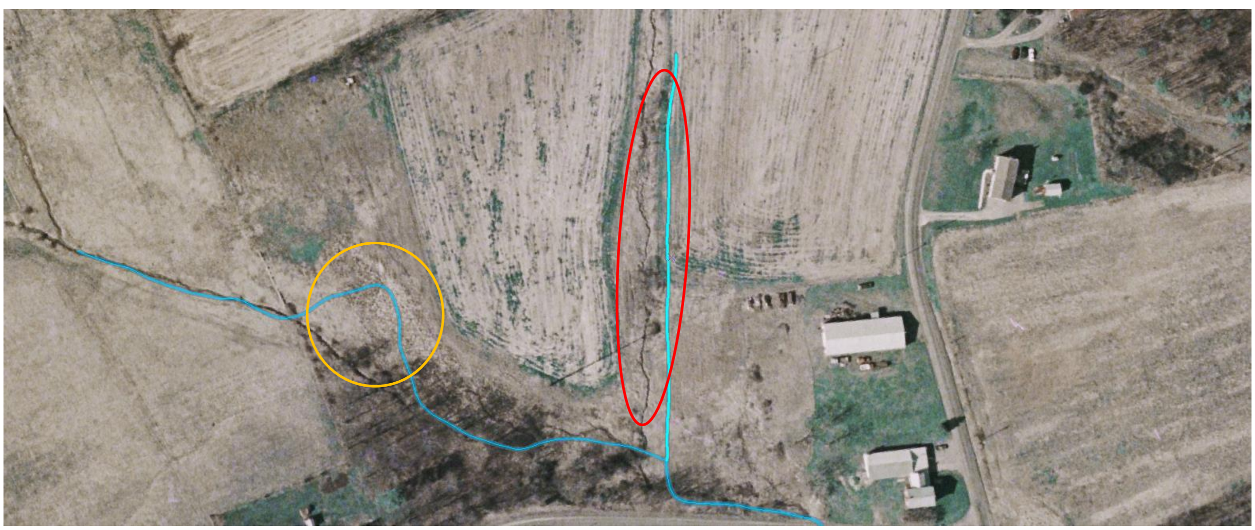
7. Section 4c – stream mapped on concentrated flow path, no channel visible

Mapped stream follows a path of concentrated flow (as evidenced by the faint ephemeral wash visible in imagery), but a distinct stream channel is not visible in reference imagery. This example would receive a false positive designation.



8. Section 4d example – stream mapped parallel to or deviates from reference stream

Mapped stream generally follows a stream visible in the reference imagery, however, mapped stream may run parallel to the reference stream (red ovals) or deviate slightly from the reference stream (orange circles). The accuracy of these examples would be at the discretion of the analyst based on their judgement of the severity of the misalignments and the degree to which the mapped stream represents or misrepresents the reference stream.



9. Section 4d example – breaching error re-routes mapped stream

Mapped stream follows a stream in the reference imagery for a portion of its length but deviates from the reference stream for another portion of its length as a result of a breaching error. In the bottom image, the reference stream can be seen flowing from the northwest to the southeast, crossing perpendicularly under the road and flowing into the main river immediately downstream of the road crossing (red oval). However, in the upper image, the mapped stream can be seen following the reference stream until it encounters the road, at which point it flows parallel to the road for some distance before crossing and reconnecting with the main river further downstream. The accuracy of this example would be at the discretion of the analyst based on their judgement of the severity of the breaching error and the degree to which the mapped stream represents or misrepresents the reference stream.