Using Machine Learning to Automate Finding Impeded Surface Water Flow

Sam Strobel (ISDA) and Theodore Stumpf (MPH)

Presenters and Agencies

Sam Stroebel - The Indiana State Department of Agriculture (ISDA)

- Department of Soil Conservation
- Member of the Indiana Conservation Partnership

Non-regulatory agency that works with private landowners (mostly farmers) to help them implement soil and water conservation practices

Theodore Stumpf - The Management Performance Hub of Indiana (MPH)

- Senior Data Scientist

The Indiana Management Performance Hub combines a collaborative and innovative approach with industry-leading technical innovation to facilitate data-driven decision making and data-informed policy making.

INDIANA STATE DEPARTMENT OF AGRICULTURE

MANAGEMENT PERFORMANCEHUB

Background: The Problem

Land conservation professionals work with landowners to help them implement conservation practices on their land, generally agricultural fields.

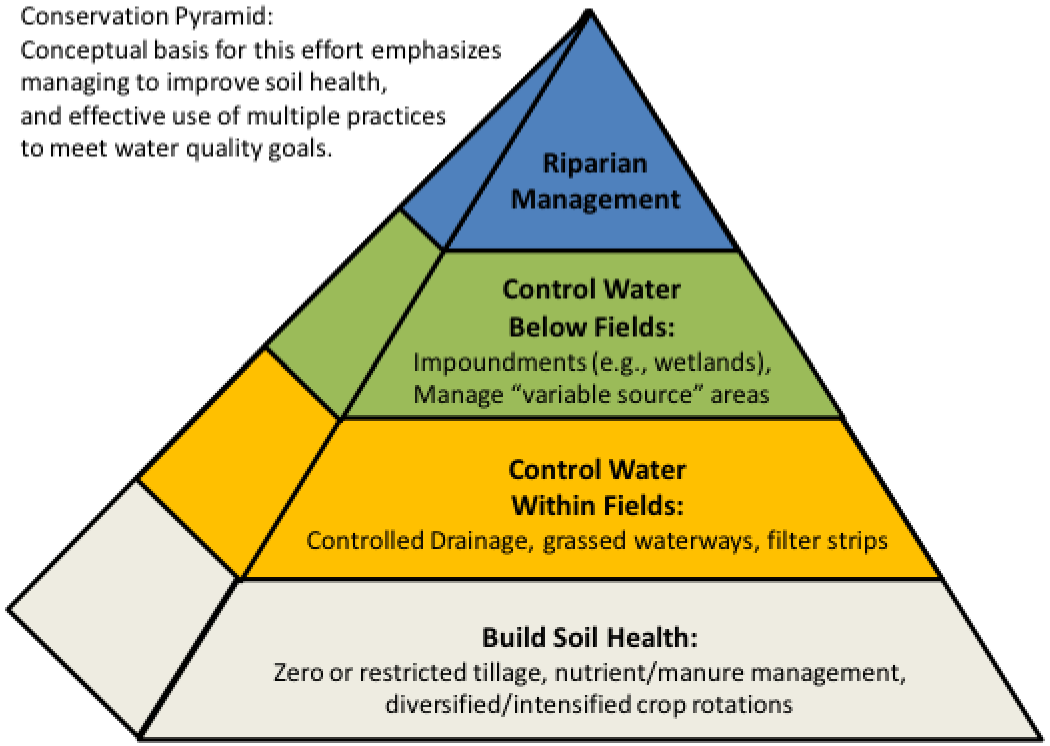
USDA and a group of partners created a GIS tool called the **Agricultural Conservation Planning Framework** (ACPF) to help conservation professionals determine where specific conservation practices may be needed based on hydrology.

INDIANA STATE DEPARTMENT OF AGRICULTURE

Background: ACPF

Three components of ACPF

- A GIS toolbox, and associated geodata to generate potential locations for a soil conservation practices
- A database of field boundaries, land use, and soil data
- A framework for conservation priorities

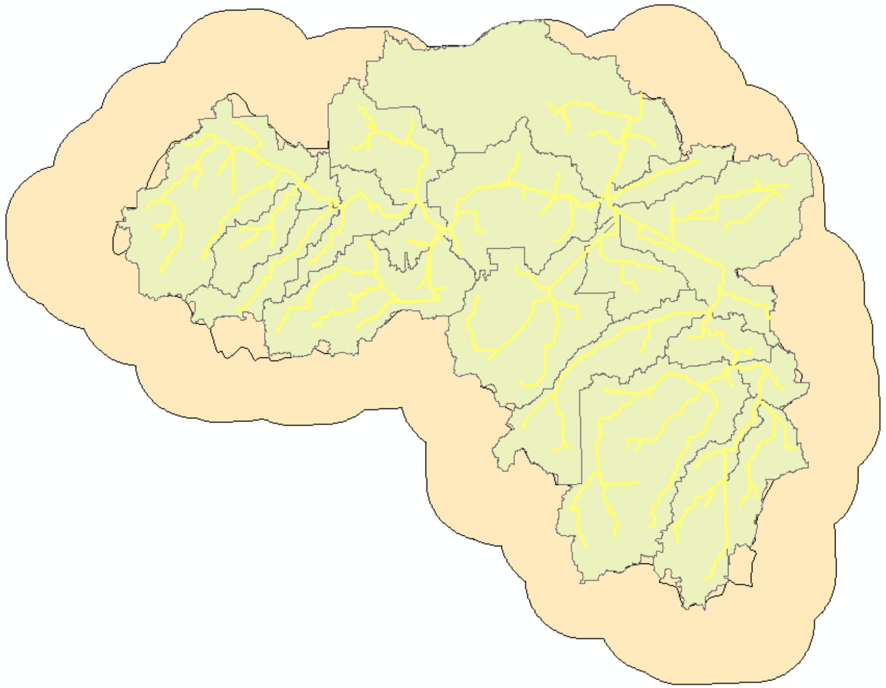


https://acpf4watersheds.org/the-framework/

INDIANA STATE DEPARTMENT OF AGRICULTURE

What Does ACPF Do?

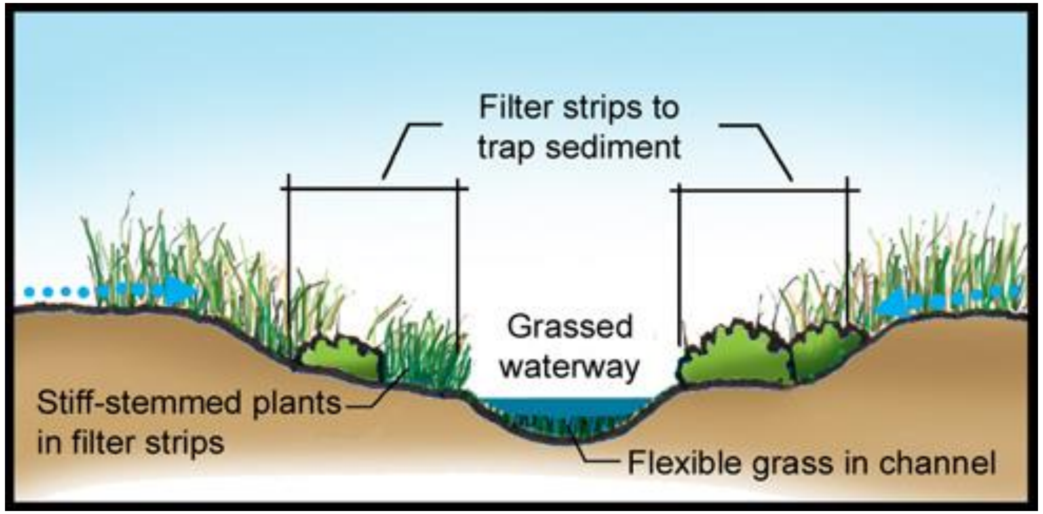
- Uses digital elevation model (DEM) to construct a hydrologic model of the overland flow of a HUC12 watershed to high level of detail
- Uses field boundaries to add information about the location of agricultural fields atop the hydrology
- Uses spatial algorithms to determine potential locations of specific conservation practices

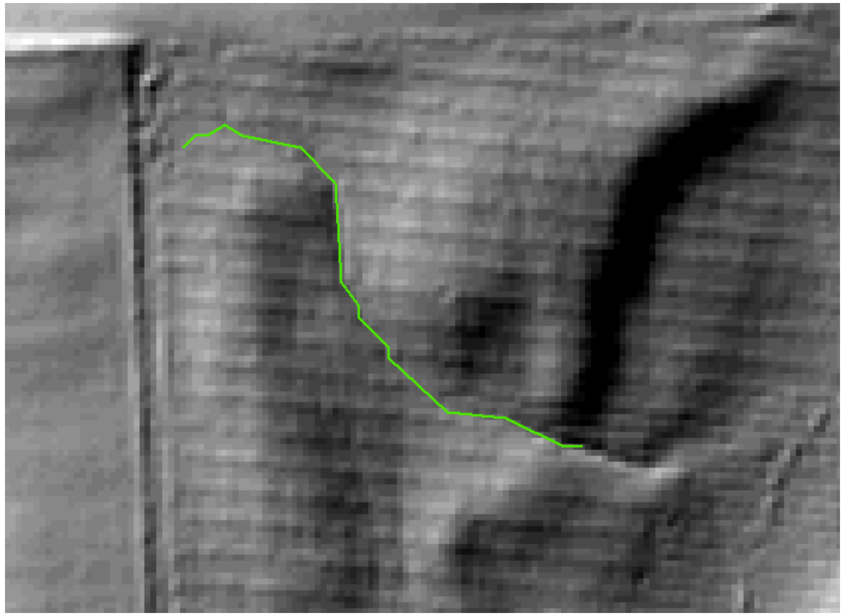


INDIANA
STATE DEPARTMENT OF
AGRICULTURE

ACPF Outputs Example

A grassed waterway is a vegetated channel that carries runoff at a nonerosive velocity to a stable outlet.





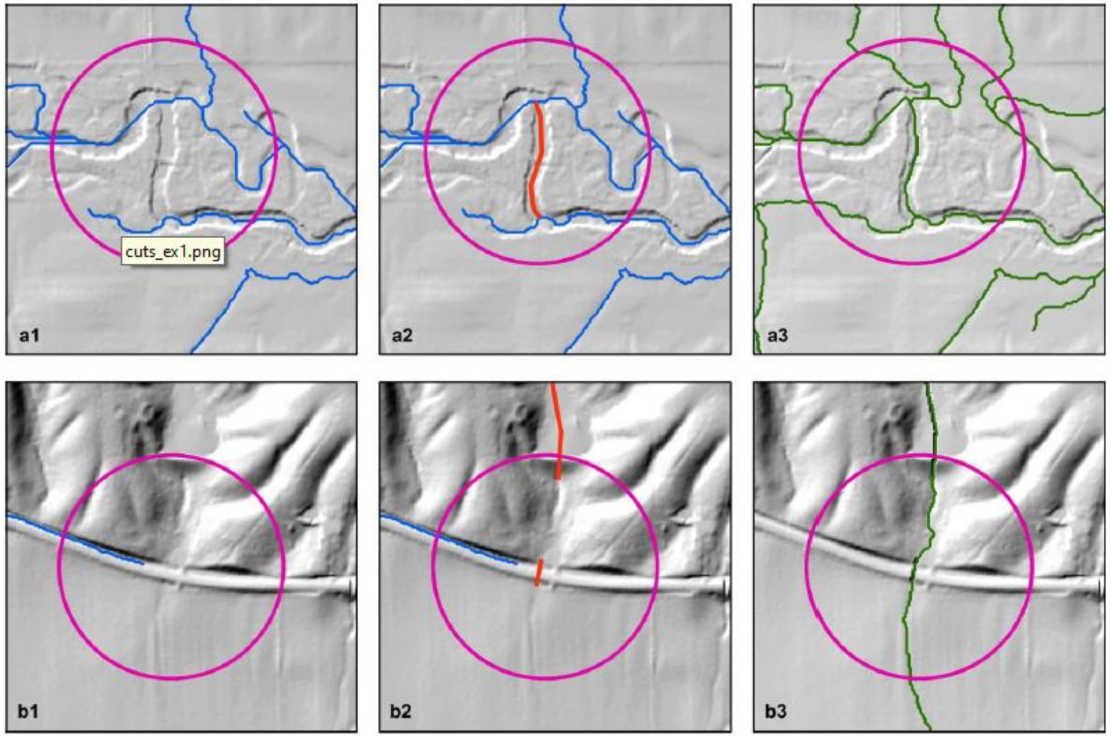
INDIANA STATE DEPARTMENT OF AGRICULTURE

ACPF Hydro: Enforcement

A core part of using the ACPF tool is creating an accurate local hydrologic model showing the surface water flow at the HUC12 level

To do this an analyst must complete a process called hydro-enforcement

- This involves finding flows that the algorithm thinks are impeded and modifying the DEM used in the flow characterization in order to correctly model the real-world flow



Original flowpaths — Digitized cut lines — Final flowpaths

Figure 3. Manually digitized cut lines and their impact on derived flow pathways

INDIANA STATE DEPARTMENT OF AGRICULTURE

ACPF Hydro: Enforcement

Examples of things that can cause a false impeded flow

- Culverts – man made structures for routing ditches and small streams under roads or other obstacles
- Bridges – Some bridges can be present even in bare-earth DEMs

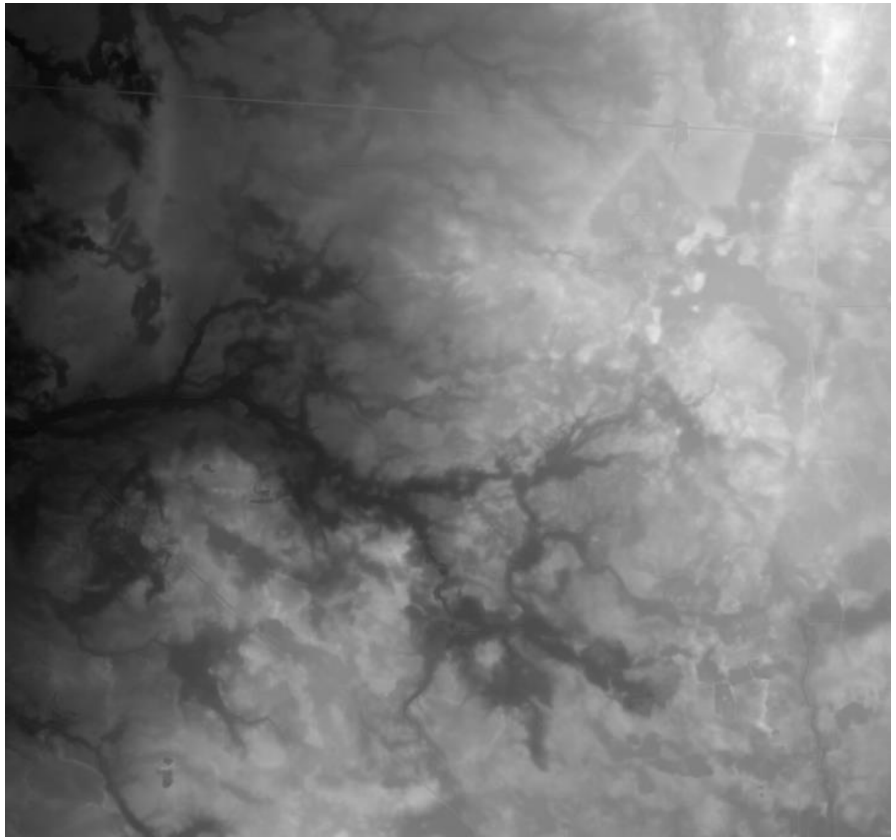
The hydro-enforcement process involves creating “cut” line features and “burning” them into the DEM file to fix false impeded flows

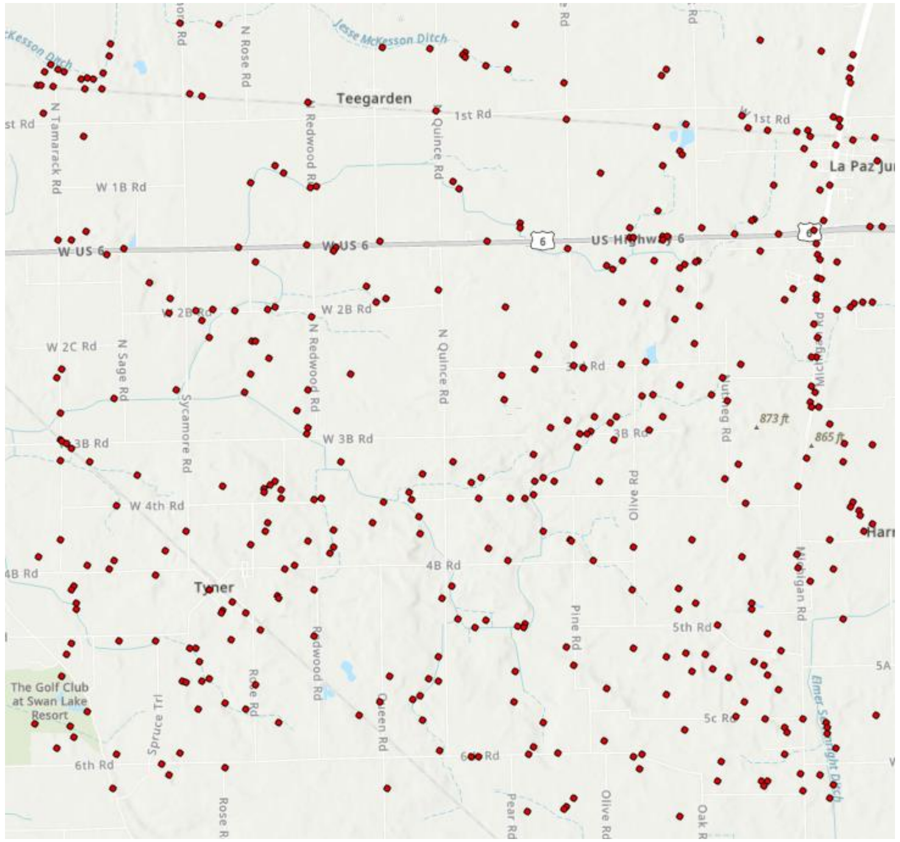
- This process can be very time consuming and is the major bottle neck for efficiently creating ACPF watershed models

The Project

ISDA utilized MPH to see if they could produce a machine learning / AI product to more quickly identify falsely impeded flows in DEM / Flow data.

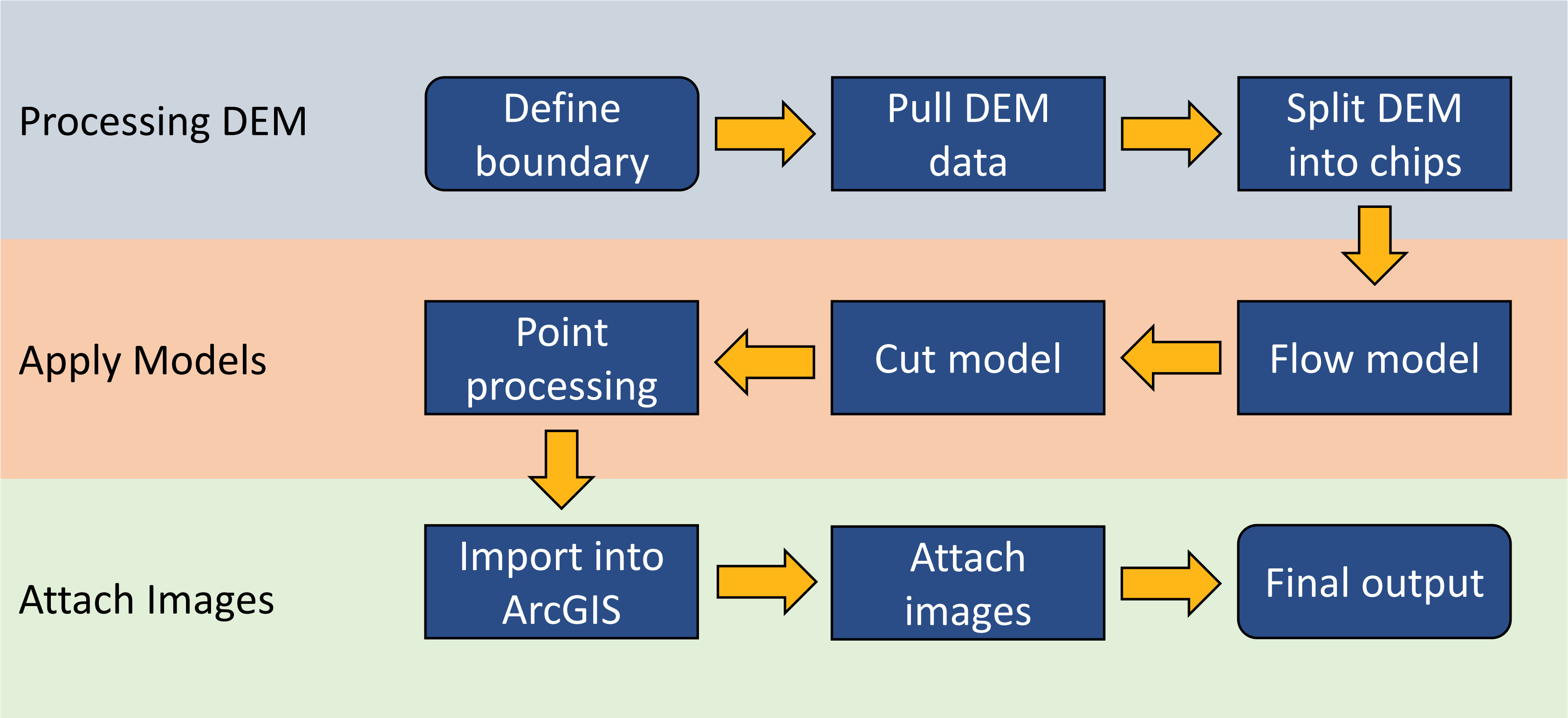
Converting a DEM into Data





MANAGEMENT
PERFORMANCEHUB

Data Processing Pipeline



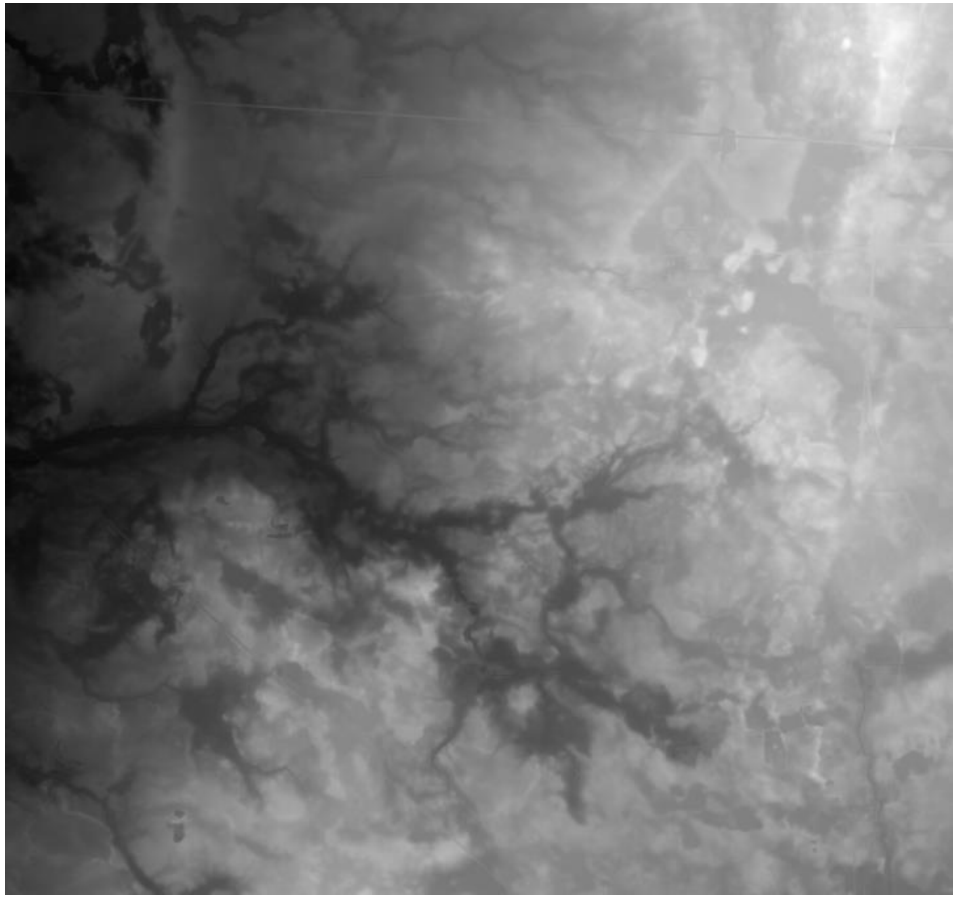
MANAGEMENT PERFORMANCEHUB

Source Data Characteristics

For this project, we are using the DEM provided by the Indiana Geographic Information Office. The elevation data was collected from 2016-2020 and it has a resolution of one point every 2.5 feet.

In our process, we downsample the resolution to one point every 2 meters.

This sample area is nearly 40 square miles, the image is 5100px x 5300px.

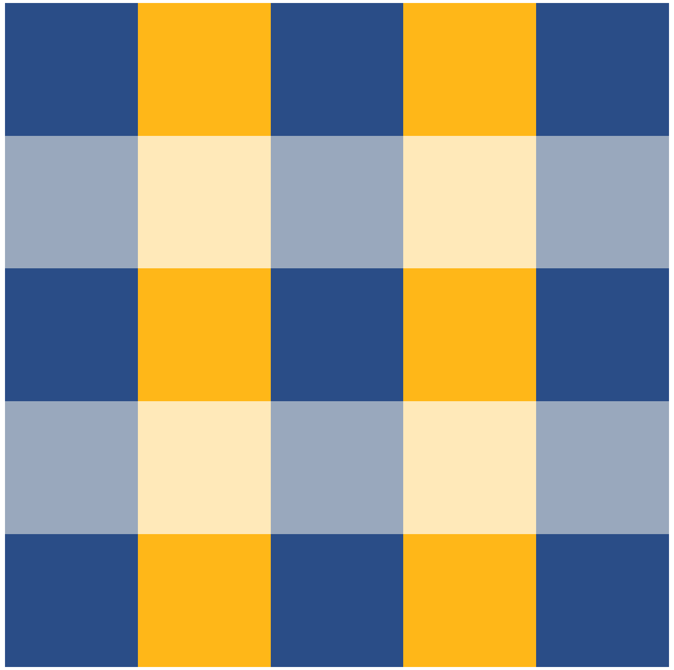


MANAGEMENT PERFORMANCEHUB

GIS INDIANA GEOGRAPHIC INFORMATION OFFICE INDIANAMAP

How Do You Climb a Mountain?

- One step at a time.
- The large DEM image is split into smaller “chips”, sized 24px x 24px.
- The larger image is then sampled in half-chip increments.
- Because the models assign one of three classes to each chip, this size balances output resolution and input detail.



MANAGEMENT PERFORMANCEHUB

Data Normalization

All the chips are normalized, with elevation data converted from meters to values between 0 and 1. A cap is placed on this compression, ensuring chips have a minimum range of 2 meters.

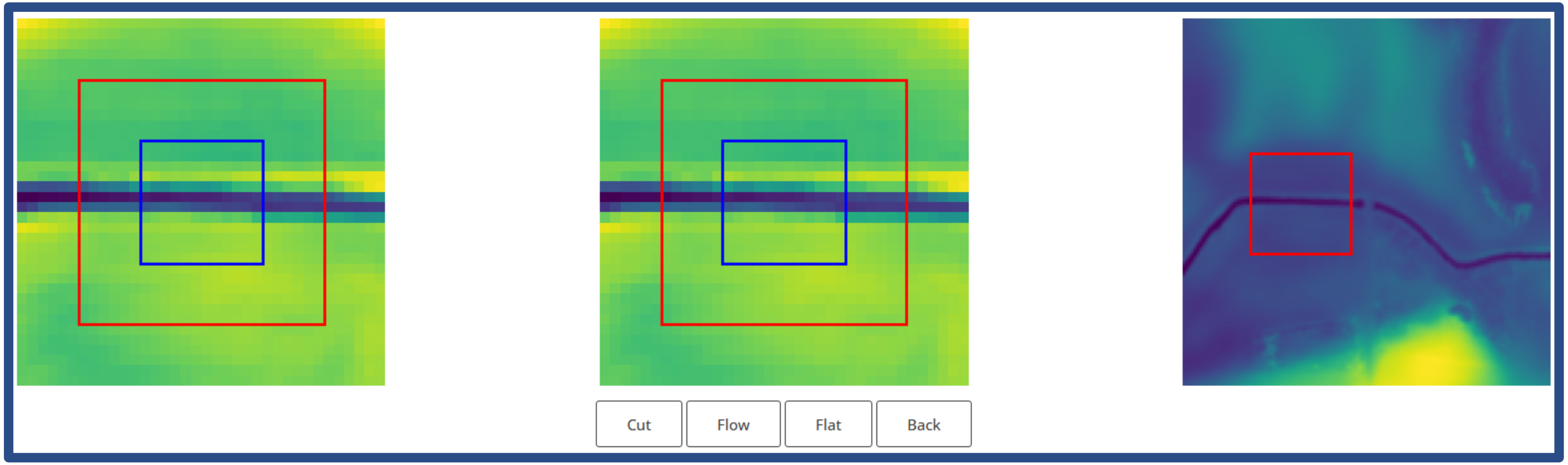
Some additional metadata is preserved from the original chip, such as the total height difference.

Additionally, the x and y slopes are calculated for the chip to provide additional information to the models to aid in detecting cuts / flows.

Training Data Collection

Training data was collected using a tool I created. It allowed me to quickly label a set of input data.

The training images were sampled from areas surrounding cut lines provided by ISDA.



MANAGEMENT PERFORMANCEHUB

Model Overview

Both models use the same overall structure.

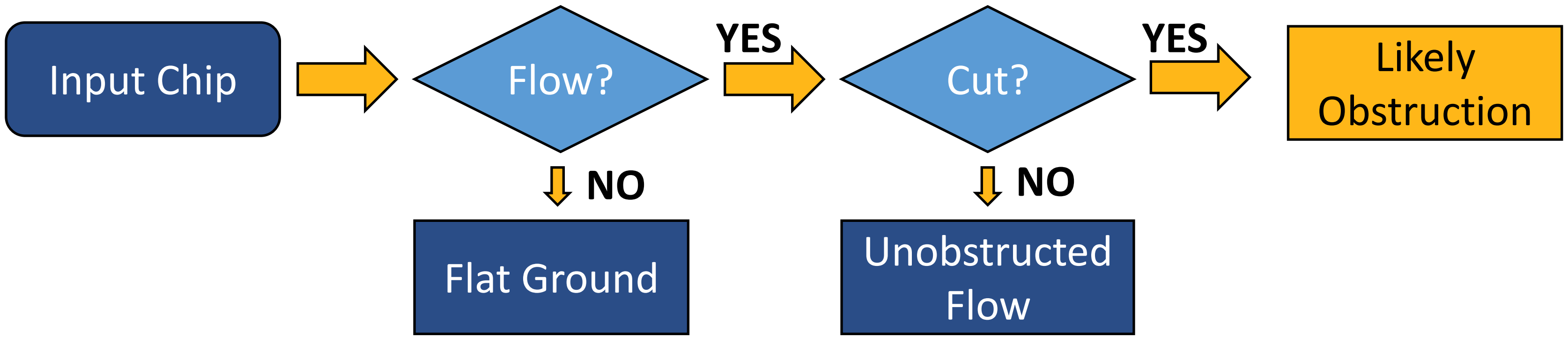
- At the start there are several convolution layers that extract data from the chip.
- The data is then “flattened” and the metadata is appended.
- Finally, the data flows through a few dense layers before being classified as one of two classes.

Two Models, Three Classes

Trying to create a multi-class model had its own challenges, so our solution was to use two unique models in sequence each with their own specialization.

The first model identifies a flow, obstructed or not as compared to flat or sloped ground.

The second model determines if the flow present is obstructed or not. Obstructed flows are marked as potential cut locations.

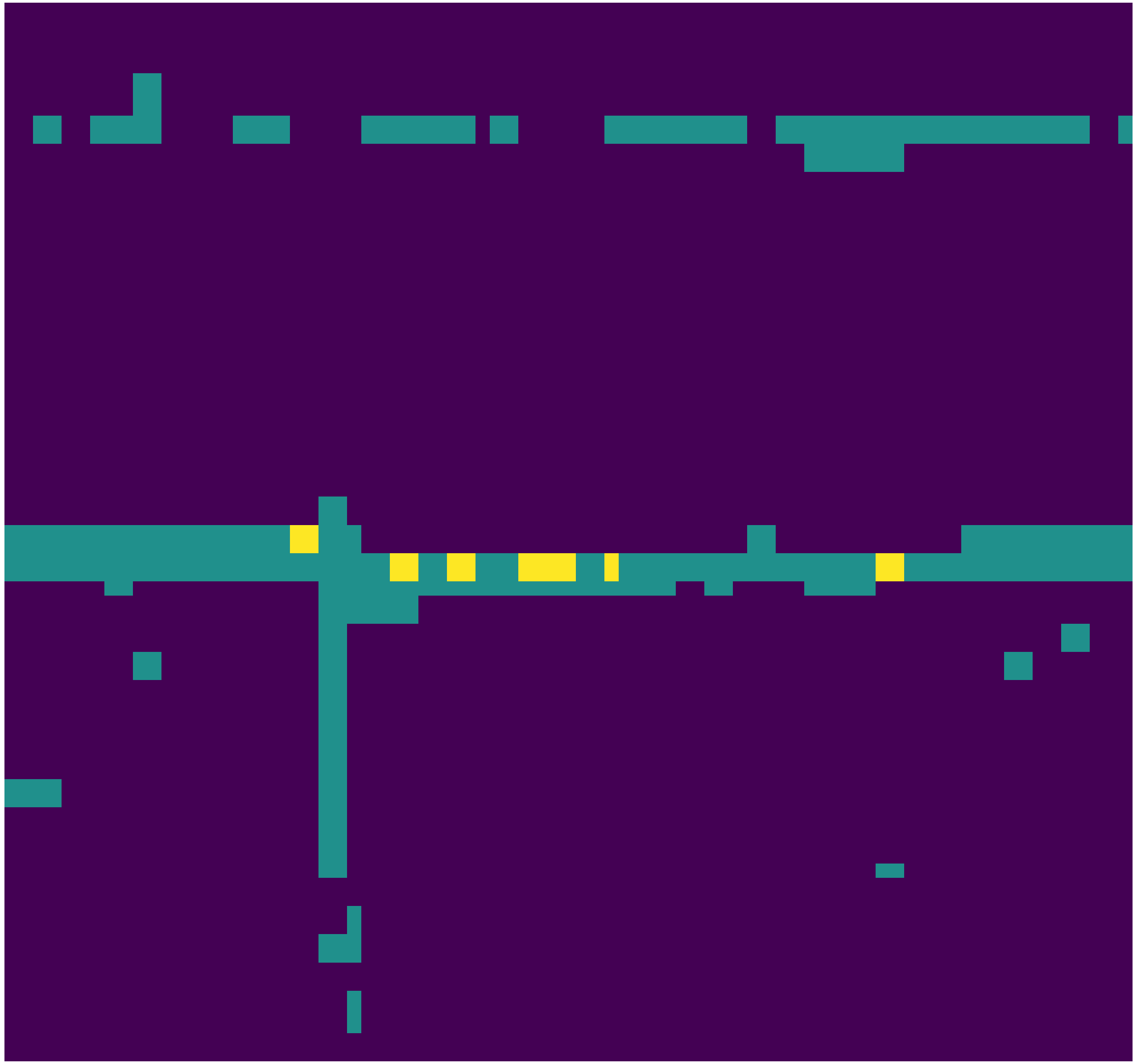


MANAGEMENT PERFORMANCEHUB

Sequential Model Output

Every chip is fed into the first model. Significant flows are marked, showing as cyan in this image.

Then, once all potential flows have been identified, the second model checks for obstructions. Likely cuts are marked, showing as yellow in this image.



MANAGEMENT PERFORMANCEHUB

Converting to Geospatial Format

We now have a set of points for likely obstructions, but they are identified by which chip number they are. We need to convert them to a shapefile to get usable output.

To get the final shapefile, each point in the output is paired with the model's confidence values and some other data.

Additionally, the chips that were detected as obstructions are all saved in a folder as .png images.

MANAGEMENT PERFORMANCEHUB

ArcGIS Integration

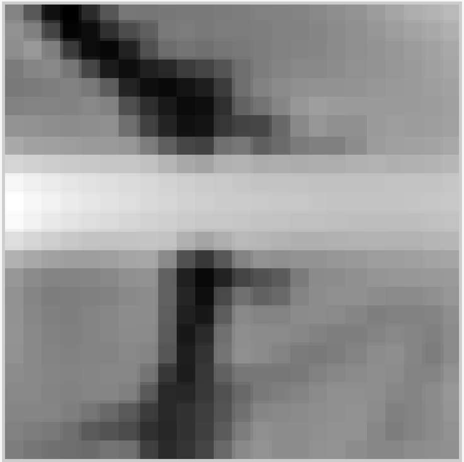
The final step involves integration with ArcGIS Pro to create the geodatabase file.

The shapefile export is formatted so that just a simple manual step is required to complete the workflow.

This provides a view of the data that is very user friendly and easy to validate.

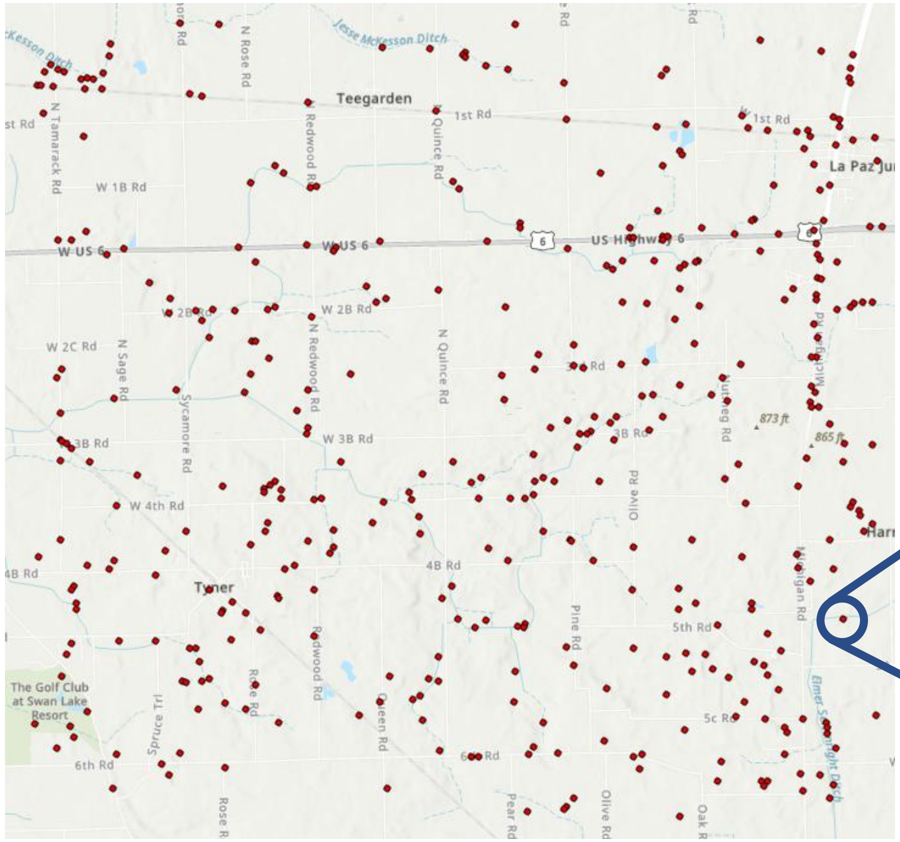
id	275
image_path	image_275.png
flow_con	100
cut_con	88.771546

image_275.png



MANAGEMENT PERFORMANCEHUB

Delivered Solution

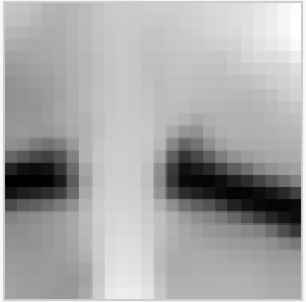


- A set of points for likely flow obstructions.
- The models' confidence for each of the points.
- An attached image to make it very easy to judge if the model correctly identified an obstruction so a cut can be applied if needed.

ModelOutput - image_100.png

OBJECTID	101
id	100
image_path	image_100.png
flow_con	100
cut_con	98.309231

image_100.png



MANAGEMENT PERFORMANCEHUB

THANK YOU!

Questions?

INDIANA STATE DEPARTMENT OF AGRICULTURE

MANAGEMENT PERFORMANCE HUB