Act 1: CWA Data Engineering

Data | Data Sources | Data Extraction

AGENDA, Feb 2

- Reviewed pdf to understand data required

- Reviewed data sources for match

- Identified potential data required from NHD

- Identified next steps

- EDA Presentation Slides

- Others

  - Meeting with previous team

  - Resources: Estimation of Compute & Min. Sample Size

Data Engineering Goals

A. To explore and understand possible data sources and specific data elements that can help construct a good feature set for our future model.

   1. Understand data elements that are taken into consideration for CWA determination & geographical distribution

   2. Identify availability and location of data sets that may be relevant

   3. Identify and distinguish which satellite dataset serves best for the problem at hand. Enumerate purpose & information it provides.

   4. Identify what other information needs to be combined from other datasets. List specific data elements and data sources.

   5. Identify what data needs to be combined and how, for training data preparation. Specifically understand what data needs to be/ can be presented in spatial Vs non-spatial formats.

B. To develop a data pipeline that generates a training data set.

1.0 Understanding CWA Determination

- CSV Elements

- CWA Determination from pdf

[Click to open worksheet](#)

CWA Determination Analysis

Tableau Interactive Visualizations (To help internal analysis; Could be useful for web deliverable)

1. Geographic Distribution of Projects with Details

2. Case Determinations - Timeline

3. CheckBox Analysis

4. Distribution of CWA determinations - TNW, RPW, non-RPW (requiring significant nexus determination) and others

    Click to access visualizations

Data Requirements

- Delineate water from land
- Identify TNWs
- Identify non-RPWs
- Identify wetlands
- Assess "significant nexus"

2.0 Data & Data Sources

Data Source	Data/Information
GEI: Landsat 8	RGB - water Vs land; hydrophytic vegetation - NDVI; water - NMDWI
Sentinel 2	RGB - water Vs land; hydrophytic vegetation - NDVI; water - NMDWI
GEI: Global Surface Water Explorer - JRC Global Surface	wetland hydrology, ephemeral water body
NHD - National Hydrography Dataset WBD − Watershed Boundary Dataset	headwaters indicator; ephemeral/intermittent indicator
SRTM - HydroSHEDs	30m resolution; more suited for regional analysis of hydrology
SRTM - Elevation	Explore NED instead
NED - National Elevation Dataset	More accurate than SRTM for elevation
SSURGO - Soil Survey Geographic db	to identify riparian areas
NWI - National Wetland Inventory	polygon data of wetlands

NHD – National Hydrology Dataset
WBD – Watershed Boundary Dataset

- <u>Source: USGS</u>

- The **NHD** and **WBD** are geospatial datasets that map the Nation's surface water network and hydrologic drainage areas.
  - The NHD represents the Nation's rivers, streams, canals, lakes, ponds, glaciers, coastlines, dams, and stream gauges, and related features.
  - The WBD represents drainage areas of the country as eight nested levels of hydrologic units.
  - They are the most up-to-date and geographically inclusive hydrography datasets for the Nation. (although EPA criticizes their "completeness")
  - Scale 1:24,000-scale or larger

NHDPlus HR – most refined data avaialble

- The **NHDPlus HR** is a nationwide, integrated **hydrography and elevation** geospatial dataset that includes the NHD stream network, WBD hydrologic units, local drainage areas or "catchments", flow direction and flow accumulation rasters, and <u>Value-Added Attributes (VAAs)</u>.

- <u>Value-added attributes (VAAs)</u>, include
    - An elevation-based catchment area for each flow line in the stream network
    - Catchment characteristics, including mean annual precipitation, mean annual temperature, and mean annual runoff, and mean latitude
    - Cumulative drainage area characteristics
    - Mean annual flow (1971-2000) and velocity estimates for each flow line in the stream network
    - Flow direction, flow accumulation, elevation, and hydro-enforced elevation rasters
    - Headwater node areas
    - Flow line min/max elevations and slopes

- NHDPlus HR consists of vector and raster data layers.

- As of 2020, a "**Beta" version** of NHDPlus HR has been developed for approximately **80% of the Nation**, including most of the conterminous U.S. and focused areas of Alaska.

- **Decision Point**: NHD + WBD + NED or NHD Plus HR ?
  - Would any of the "known" issues of NHDPlus HR Beta prohibit us from using the data set? Should we be looking at extracting and combining relevant data from NHD, WBD and NED explicitly for our purpose?

- **Questions:**
  - Can David's team help us identify a set of indicators that could proxy for the evaluation criteria related to hydrology of the sites?
  - Does NHD have indicators for water bodies that are TNWs?
  - Does NHD include wetlands? If no, how can we obtain similar characteristics?
  - Does NHD include any soil related indicators that can help identify hydric soil?
  - In what format do we require the data to be provided to us?

SRTM - HydroSHEDs

- [Source: WWF](#)
- HydroSHEDS is based on high-resolution elevation data obtained during a Space Shuttle flight for NASA's Shuttle Radar Topography Mission (SRTM).
- It offers a suite of geo-referenced data sets (vector & raster) at various scales of river networks, watershed boundaries, drainage directions, and flow accumulations.
- Available resolutions range from 90-500m

NHD offers us data at a much higher resolution, if only we can obtain data from David Keiser's team.

SSURGO

- Source: USDA

- Hydric soil is soil that is formed under conditions of saturation, flooding or ponding long enough during the growing season to develop anaerobic conditions in the upper part.

- Hydric soils support the growth and regeneration of hydrophytic vegetation.

- A combination of the hydric soil, hydrophytic vegetation, and hydrology properties define wetlands . Therefore, an area that meets the hydric soil definition must also meet the hydrophytic vegetation and wetland hydrology definitions in order for it to be correctly classified as a jurisdictional wetland.

SSURGO was used to identify riparian corridors or flood plains. Does the NHD have information that can help us assess the same?

NWI – National Wetlands Inventory

- [Source: U.S Fish & Wildlife Service](#)
- Available as polygon data within Google Earth as .kml file
- To find out if we can access this for free or if we need to purchase Google Earth Pro. Limitations may exist around use for analysis outside of GEI.

Next Steps

- Setup meeting with David Keiser and team to collaborate on NHD data for hydrology, watershed and elevation
- Understand how NWI data can be used within GEI (.klm) for identifying wetlands
- Confirm that use of NWI + NDVI is sufficient enough to identify wetlands
- Simultaneously scale up on GEE to extract images