**Data Preparation & Analysis Progress Report (prediction parameters)**

**1. Extraction of FireCCI Data from Google Earth Engine (GEE)**

* **Objective:**  
  Retrieve burned area information for all lake catchments in the study region using the ESA FireCCI51 MODIS Burned Area Product.
* **Approach:**
  + Lake catchment polygons were prepared as a shapefile and uploaded to GEE as an asset.
  + Using Python and the Earth Engine API, weekly aggregated summaries of burned pixels were extracted for each catchment, **grouped by land cover class** for each week and year (2002–2020).
  + The output is a set of CSV files (one per year), with each row representing a lake-week, and columns for burned pixel counts per land cover type.

### 2. ****Identification of the Most Fire-Affected Land Covers****

* **Objective:**  
  Determine which land cover types are most frequently or heavily burned in the lake catchment areas throughout the study period.
* **Approach:**
  + All annual FireCCI extraction CSVs were aggregated into a single DataFrame.
  + For each land cover class (e.g., cropland, shrubland, forest types), the total number of burned pixels was summed across all lakes and weeks.
  + Land cover codes were mapped to descriptive names for interpretability.
  + The results were sorted to rank land cover types by total burned area.
  + The process can also be performed per year or per lake if more granularity is needed.
* **Outcome:**
  + Identified the land cover types most frequently impacted by fire in the catchment dataset (e.g., shrublands, grasslands, croplands).
  + These land covers will be prioritized for inclusion in further analysis and modeling as potential predictors of post-fire lake ecosystem response.

| **Rank** | **Land Cover Name** | **Total Burned Pixels** |
| --- | --- | --- |
| 1 | Grassland | 28,250 |
| 2 | Tree broad-leaved deciduous | 23,559 |
| 3 | Shrub or herbaceous flooded | 17,874 |
| 4 | Shrubland | 16,408 |
| 5 | Tree needle-leaved evergreen | 13,591 |

**3. Extraction of ERA5-Land Climate Data from GEE**

* **Objective:**  
  Gather key climate variables for each lake catchment on a weekly basis.
* **Variables Extracted:**
  + 2-meter air temperature
  + Lake mix-layer temperature
  + Total precipitation
  + Surface runoff
  + Total runoff
* **Approach:**
  + ERA5-Land Daily Aggregated dataset was accessed via GEE.
  + For each week of each year, the **mean or sum** of each variable was calculated within each catchment using spatial reduction (zonal statistics).
  + The output is a set of CSV files (one per year), with rows per lake-week and columns for each variable.

### . ****Correlation Analysis between Chl-a and ERA5 Features****

* **Objective:**  
  Investigate which climate variables are most strongly associated with chlorophyll-a concentration (chla\_mean) in the lakes.
* **Approach:**
  + For each lake, chla\_mean values were paired with weekly ERA5 variables.
  + Both contemporaneous (same week) and **lagged** (previous weeks) correlations were computed using Pearson correlation.
  + Pairplot scatter matrices and lagged correlation analysis were used to visualize and quantify the strength and timing of relationships.
* **Outcome:**
  + Identified climate features and lag times with the strongest correlation to chla\_mean, which will be used for feature selection in subsequent modeling (e.g., LSTM neural network).

### 4. ****Correlation Analysis between Chl-a and ERA5 Features****

* **Objective:**  
  Investigate which climate variables are most strongly associated with chlorophyll-a concentration (chla\_mean) in the lakes.
* **Approach:**
  + For each lake, chla\_mean values were paired with weekly ERA5 variables.
  + Both contemporaneous (same week) and **lagged** (previous weeks) correlations were computed using Pearson correlation.
  + Pairplot scatter matrices and lagged correlation analysis were used to visualize and quantify the strength and timing of relationships.
* **Outcome:**
  + Identified climate features and lag times with the strongest correlation to chla\_mean, which will be used for feature selection in subsequent modeling (e.g., LSTM neural network).

The results for Lake 27 are shown below:

| **Feature** | **NoLag\_corr** | **BestLag** | **BestLag\_corr** |
| --- | --- | --- | --- |
| lake\_mix\_layer\_temperature\_C | -0.39 | 1 | -0.38 |
| runoff\_sum | -0.20 | 3 | -0.35 |
| surface\_runoff\_sum | 0.17 | 4 | -0.12 |
| temperature\_2m\_C | -0.26 | 4 | -0.42 |
| total\_precipitation\_sum | -0.19 | 3 | -0.23 |