# Weekly Aggregation of CHLA, LSWT, and Turbidity Time Series for Global Lakes

## 1. Introduction

This study describes a standardized pipeline for converting daily satellite-derived lake observations (chlorophyll-a [CHLA], lake surface water temperature [LSWT], and turbidity) into a weekly aggregated format. The method is designed to preserve temporal continuity while minimizing noise due to outliers. It enables consistent time series analysis across variables and lakes and supports applications such as trend assessment, modeling, and hydro-ecological classification.

## 2. Datasets

Each daily file corresponds to a unique lake and contains:  
- date: Observation date  
- n: Number of observations used  
- ntot: Total number of input values  
- nNA: Number of missing values  
- mean: Mean value  
- median: Median value  
- stddev: Standard deviation  
  
The pipeline processes datasets stored in three folders:  
- CHLA: Chlorophyll-a concentration (mg/m³)  
- LSWT: Lake surface water temperature (K)  
- turbidity: Water turbidity (FNU or equivalent)

## 3. Methodology

### 3.1 Preprocessing and Filtering

- Files were filtered to include only data from 1 January 2002 onward.  
- Rows containing outliers were identified based on predefined thresholds:  
 - CHLA ≥ 50  
 - LSWT ≥ 305  
 - Turbidity ≥ 20  
- For rows classified as outliers, only the mean, median, and stddev values were masked (set to NaN). All rows were retained to preserve the weekly structure.

### 3.2 Temporal Aggregation

- For each lake, a weekly time index was computed where each week begins on January 1st and spans 7 consecutive days.  
- Data were grouped by Year and Week to compute:  
 - Cumulative sums of n, ntot, nNA  
 - Averages of mean, median, stddev (ignoring NaN values)  
- week\_start and week\_end were added to represent the actual time span of each week.

## 4. Output

Each lake’s file is saved in a weekly .csv format with the following fields:

|  |  |
| --- | --- |
| Column | Description |
| Lake\_ID | Unique identifier parsed from filename |
| Year | Calendar year |
| Week | Week index within the year (1–53) |
| week\_start | First date of the week |
| week\_end | Last date of the week |
| n | Sum of daily observation counts |
| ntot | Sum of total expected values |
| nNA | Sum of missing daily values |
| mean | Average of daily mean values (NaN ignored) |
| median | Average of daily medians |
| stddev | Average of daily standard deviations |

Output filenames follow this structure:  
ID[Lake\_ID]\_[Lake\_Name]\_[VARIABLE]\_weekly.csv  
  
For example:  
- ID3\_victoria\_CHLA\_weekly.csv  
- ID3\_victoria\_LSWT\_weekly.csv  
- ID3\_victoria\_TURB\_weekly.csv

## 5. Remarks

This pipeline ensures temporal consistency across all lakes and variables by preserving weekly intervals and handling outliers conservatively. It supports robust analysis of seasonal and inter-annual dynamics of lake properties globally.