USER GUIDE

**Batch Analysis of Daily Temperature Values R Tool**

**Version 1.5 (March 2022)**

Jason Williams, Idaho DEQ Lewiston Regional Office


**Tool Description**

This document is a guide to using the Batch Analysis for Daily Temperature Values Tool R tool ('Batch Analysis DV Tool'). The Idaho Department of Environmental Quality (DEQ) developed this tool to enable batch analysis of many temperature logger datasets simultaneously. It was developed for projects where there are so many temperature logger datasets that manually processing each logger dataset and comparing temperatures to criteria is not feasible within time available. If you only have a handful of temperature logger datasets, this tool may not save you time relative to manually screening and analyzing data or using other tools.

The tool enables daily temperature statistics from multiple loggers to be screened, analyzed, & plotted simultaneously. The number of daily temperature records (site/day combinations) that can be analyzed at once is limited to the maximum number of rows possible in the Microsoft Excel template used as the tool input (1,048,576 rows). Users enter daily temperature data and information about when criteria apply into the Microsoft Excel template. The R tool then reads in these data, screens data based on user-specified criteria, outputs an excel file with results, and outputs a pdf file with a plot for each site/year combination. Prior experience with R is helpful, but not required.

This tool was designed to use daily temperature statistics as inputs rather than raw data for several reasons. First, some widely used public temperature data sources are available as daily temperature statistics. Second, using daily statistics as inputs makes it easier for users to place all project temperature data into only one input file and to manage project data. For projects with many raw data files, the number of raw data rows can quickly exceed Excel's row capacity. While R or other programming languages or tools could be used to format and combine all raw data into one massive csv file to use as input, most users of this tool likely do not have experience or resources needed for that type of data wrangling. However, if the project includes raw data, users will need to calculate daily temperature statistics by site/day (DailyMax, DailyMean, DailyMin, N) and format their data as required to populate the tool Excel template.

DEQ has conducted numerous tests to evaluate tool output accuracy, and those tests suggest output is reliable. However, we likely haven't encountered every possible situation. Therefore, users are responsible for reviewing and verifying accuracy of tool outputs.

**Terms of Use & Restriction of Liability**

Neither the State of Idaho nor the Idaho Department of Environmental Quality, nor any of their employees make any warranty, expressed or implied, or assume any legal liability or responsibility for the accuracy, completeness or usefulness of any information or data provided. This tool and associated information should not be used without first reading and understanding its limitations. The Department of Environmental Quality may update, modify, or revise the tool at any time, without notice.

**General Approach**

Detailed instructions for using the tool are provided below. The following general approach is recommended:

1. If you have raw temperature data, calculate daily temperature statistics by site/day (DailyMax, DailyMean, DailyMin, N) and format data as described in detailed instructions below.
2. Enter data into the Excel template 'data' and 'when criteria apply' tab as described in detailed instructions below.
3. Enter preliminary data screening values in 'user_input_for_batch_analysis_DV' script as described in detailed instructions below.
4. Run the tool and examine the excel and pdf plot outputs to evaluate whether data screening performed by the tool is adequate for user purposes.
5. If needed, edit data screening values in 'user_input_for_batch_analysis_DV' script and/or use the 'Manual Exclude' column in the template to flag/exclude additional data the user considers not suitable for comparison to criteria.
6. Re-run the tool and repeat steps 3-4 until data screening meets user requirements.
7. Examine tool outputs to verify they are correct and meet user needs.

Entering uncleaned data into the template and then flagging and excluding data using the tool's screening features is recommended because tool output documents what site/day combinations were excluded and associated reasons. Alternatively, users can also enter only cleaned data into the template, but users would then be responsible for documenting data excluded and exclusion reasons.

**Detailed Instructions**

1. If needed, install R and R studio (both are free).

2. Create a new folder and save all tool files within that directory. Note the tool will not work within a zipped folder.

3. Install the 'renv' R package. Open a new R script using R studio. Type

   install.packages("renv")

   Then, click 'Source' to run the script & install the package



   .
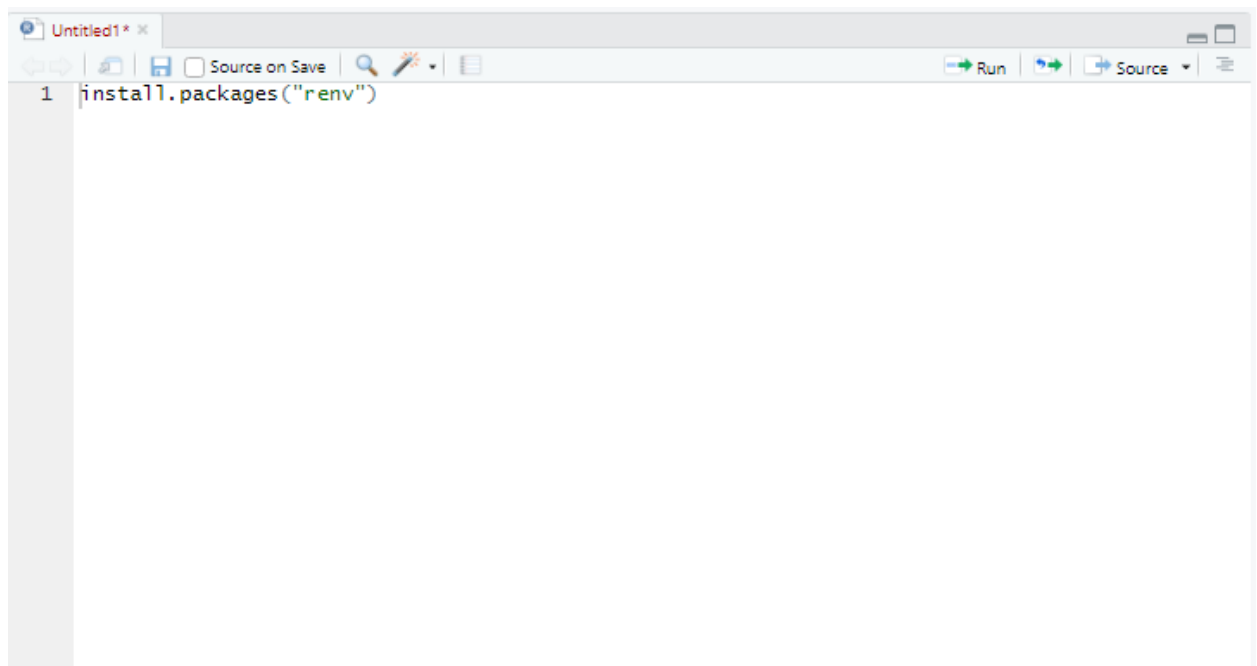
4. Open the template file 'template_for_batch_analysis.xlsx'. The template file is located within the 'inputs' subfolder. Enter or copy/paste daily temperature statistics data into the 'data' tab.

   a. Site (required) = unique identifier for the Site/Station that collected the temperature data. If you have multiple years of data for one site, give all data the same Site identifier
   b. Date (required) = format as mm/dd/yyyy
   c. DailyMax (required) = Daily Maximum temperature values (in degrees C)
   d. DailyMean (required) = Daily Mean temperature values (in degrees C)
   e. DailyMin (required) = Daily Minimum temperature values (in degrees C)
   f. N = (required) # of daily temperature observations

g. Source = user-provided data source name

h. Manual Exclude = this column enables the user to manually specify site/day combinations to exclude from temperature criteria analyses and plotting. The tool will flag and exclude any non-blank rows.

Use 'Save As' to give the populated template file a new name. Save this file in the same subfolder (i.e., the 'inputs' folder).

5. Enter information in the 'when criteria apply' tab.
   - Format start/end dates as mm/dd.
   - For salmonid spawning dates, use 'Salmonid Spawning-General' if spawning is one continuous period within a single calendar year; or enter both 'fall' and 'spring' dates if the spawning period is discontinuous. For example, if spawning spans all days 04/01 to 11/01 based on species present, enter 04/01 and 11/01 for 'Salmonid Spawning-General' and leave spring and fall dates blank. If the spawning period is discontinuous and there is a date gap when criteria do not apply, enter dates for 'spring' and 'fall', and leave 'general' dates blank.
   - If the spawning period spans the calendar new year, split this into two periods. For example, if spawning occurs over fall/winter between 10/01 and 04/01, do not enter 10/01 and 04/01 for 'general' dates. Instead, enter 10/01 to 12/31 for 'fall' and 01/01 and 04/01 for 'spring', and leave 'general' dates blank. End dates should not extend past 12/31.
   - If applicable, enter start/end dates as mm/dd for Idaho Bull Trout and EPA Bull Trout Criteria.
   - See 'Resources for Determining Where and When Idaho Numeric Temperature Criteria Apply' at the end of this document.

   Save the excel file.

6. Open the 'batch analysis DV tool' project file from the main folder. Once the file is open in R, proceed to the next step.

   R batch analysis DV tool.Rproj

7. In R, open the 'user_input_for_batch_analysis_DV.R' file. Specify user inputs following comment instructions in script and as listed below. The tool will flag days where user-specified thresholds are exceeded. Flags applied for each day will be documented in the 'dailystats' output tab, but flagged days are not included in criteria exceedance evaluations. Flagged and excluded days are included in output temperature plots as grey lines.

```
############################################################################################
# PART 1: USER INPUT
############################################################################################

# specify project name; the excel output files will have this in the filename; NO SPACES IN NAME!
projectname <-"Lochsa"

# specify name of excel file with data
filename <-"for_lochsa_batch_analysis.xlsx"

# specify names of tabs in template excel file
data.tab = "Data"
criteria.tab = "when criteria apply"

# flag & exclude days where daily mean differs from previous or subsequent daily mean above this value
# If you don't want to include this threshold in the analysis, enter 100

consecutive.mean.threshold <- 3

# flag & exclude days where daily max differs from previous or subsequent daily max above this value
# If you don't want to include this threshold in the analysis, enter 100

consecutive.max.threshold <- 10

# flag & exclude days with daily max above this threshold
# If you don't want to include this threshold in the analysis, enter 100

daily.max.threshold <-28

# flag & exclude days during October-May with daily mean above this threshold
# If you don't want to include this threshold in the analysis, enter 100
oct.to.may.mean.threshold <-15

# flag & exclude days with daily range above this threshold
# if you don't want to include this threshold in the analysis, enter 100
daily.range.threshold <-20

# flag & exclude days where # observations does not exceed this threshold
# if you don't want to include this threshold in the analysis, enter 1000
minimum.daily.obs.threshold <-24

# require that each site/day combo have # of daily observations reported
# enter "Y" to exclude cases where # obs not reported, or "N"
require.Nobs <-"N"
```

**projectname =** name for the project/analyses you are doing. The output excel file will be names using the project name.

**filename** = name of the excel file with your data from step 5

**data.tab** = name of the tab containing raw data in the excel file

**criteria.tab** = name of tab containing criteria application info in the excel file

**consecutive.mean.threshold** = within a site, if a daily mean value differs from the previous or subsequent mean by more than this threshold, data for the associated site/day will be flagged and excluded from criteria exceedance analyses.

**consecutive.max.threshold** = within a site, if a daily max value differs from the previous or subsequent max by more than this threshold, data for the associated site/day will be flagged and excluded from criteria exceedance analyses.

**daily.max.threshold** = if a daily max value exceeds this threshold, data for the associated site/day will be flagged and excluded from criteria exceedance analyses

**oct.to.may.threshold=** if a daily mean value from October- May exceeds this threshold, data for the associated site/day will be flagged and excluded from criteria exceedance analyses & plotting.

**daily.range.threshold=** if the daily range exceeds this threshold, data for the associated site/day will be flagged and excluded from criteria exceedance analyses

**minimum.daily.obs.threshold** = if the number of daily observations (N) is less than this threshold, data for the associated site/day will be flagged and excluded from criteria exceedance analysis and plotting

**require.Nobs**= enter "Y" to flag and exclude cases where # of observations is not avaialable, or "N" to retain these in criteria exceedance analyses

After you specified inputs, save the 'user_input_for_batch_analysis_DV.R' in R.

8. Next, open the 'batch_temp_analysis_DV_v1.4.R' file. Click the 'Source' button to run the script.

   The first time you run the script in a folder, R will download R libraries needed to run the script, which may take 30 minutes or more, depending on your internet connection and R packages you have installed locally. R downloads the exact version of the libraries used when the tool was developed. This ensures that if a newer version of the library becomes available, the tool use the version applied when the tool was created, ensuring consistent results. The next time you run the script within the folder, it will run faster because the libraries will already have downloaded.

9. When the tool runs, view output in the R Studio console window. Some messages in blue or white text will likely appear. For example, message '`summarise()` has grouped output by 'Site'. You can override using the `.groups` argument' may appear, but this is just an informational message and does not represent an error. Some additional messages may also appear in the R studio console window, depending on the particulars of your dataset. If the tool generates an excel and pdf output files within the 'output' subfolder, then the tool ran successfully. It is then your responsibility to examine the output to verify it is accurate and meets your needs.

10. If it runs with no errors, the script will create several files and save them in the 'outputs' subfolder. The tool will output one excel file and a temperature plots pdf file with one page per site. These files will be saved the 'output' subfolder of the directory you created. In R, you can use getwd() to confirm your current working directory.

    The Excel output file includes several tabs:
    - **User inputs:** a record of values entered into 'user_input_for_batch_analysis.R'
    - **Dailystats:** for each site, daily temperature statistics and a record of flags applied and exclusions based on QAQC screening
    - **'WMT':** weekly maximum temperature values

- **'COLD Summary':** Idaho Cold Water Aquatic Life numeric temperature criteria exceedance results for sites where COLD criteria dates were specified. If there are multiple calendar years of data for a site, data are summarized by site/year combination.
- `**SS summary'**: Idaho Salmonid Spawning numeric temperature criteria exceedance information for sites where SS dates were specified. If there are multiple calendar years of data for a site, data are summarized by site/year combination.
- **'ID BT rearing'**: Idaho Bull trout rearing criteria exceedance information for sites where Bull Trout rearing dates were specified. The Maximum Weekly Maximum Temperature (MWMT), calculated as the single highest WMT during the specified rearing period, is also provided. If there are multiple calendar years of data for a site, data are summarized by site/calendar year combination.
- **'ID BT spawning'** provides Idaho Bull Trout spawning criteria exceedance information for sites where Bull Trout spawning dates were specified. If there are  multiple calendary years of data for a site, data are summarized by site/calendar year combination.
- '**EPA BT'** provides EPA Bull Trout spawning criteria exceedance information. This is calculated for all sites, regardless of inputs.  Disregard if not applicable.

11. Inspect output. If needed, remove or edit data in the excel template or user input R file, and then re-run the script.

**Resources for Determining Where and When Idaho Numeric Temperature Criteria Apply**

Table 1 summarizes the Idaho numeric temperature criteria the R tool was designed to apply. To use the tool, users must determine where and when criteria are applicable, and enter this information into the excel template (see instructions above). This is not always simple. Links to several relevant resources are provided below.

**Table 1.** Numeric temperature criteria applied by the Batch Analysis DV Tool.

| Criteria | Criteria Value(s) | Where Criteria Value(s) Apply | When Criteria Value(s) Apply |
|---|---|---|---|
| Cold Water Aquatic Life (IDAPA 58.01.02.250.02.b) | 22 °C daily maximum | all Idaho waters [a] | Year-round |
| | 19 °C daily average | all Idaho waters [a] | Year-round |
| Salmonid Spawning (IDAPA 58.01.02.250.02.f.ii) | 13 °C daily maximum | waters where SS is a designated or existing use [b] | during spawning and incubation periods for salmonids present |
| | 9 °C daily average | waters where SS is a designated or existing use [b] | during spawning and incubation periods for salmonids present |
| ID Bull Trout [c, d] (IDAPA 58.01.02.250.02g and g.i) | 13 °C 7-day average of daily maxima (MWMT) | waters in Idaho's 1996 bull trout conservation plan [e] | June-August (rearing period) |
| | 9 °C daily average | waters in Idaho's 1996 bull trout conservation plan [e] | September-October (spawning period) |
| Federal Bull Trout [d] 40 CFR § 131.33 | 10 °C 7-day average of daily maxima (MWMT) | waters specified in 40 CFR § 131.33(a)(2) | June-September |

[a]waters as defined in Idaho Code 39-3602(34), except those where CWAL is not designated but seasonal cold, warm water, or modified aquatic life uses are designated in IDAPA 58.01.02.110-160; all Idaho waters that lack aquatic life use designations receive CWAL presumed use protection (IDAPA 58.01.02.101.01). In the Lochsa, CWAL criteria apply to all streams.
[b]IDAPA 58.01.02.110-160 defines designated waters; existing uses are those attained on or after November 28, 1975, per IDAPA 58.01.02.10.38.
[c]ID Bull Trout and Federal Bull Trout criteria apply to CWAL beneficial use
[d]for waters where both ID and federal bull trout criteria apply, DEQ tests for exceedances using both criteria, and DEQ considers exceedance of either criteria to be a water quality standards violation
[e] waters > 600 m elevation north of the Salmon River Basin/Clearwater River Basin Divide, and waters > 1400 meters elevation south of the Salmon River Basin/Clearwater River Basin Divide in key watersheds listed in Table 6, Appendix F of Governor Batt's Idaho Bull Trout Conservation Plan: https://species.idaho.gov/wp-content/uploads/2016/05/bulltroutconservationplan-96.pdf

Relevant Resources:

- Idaho Water Quality Standards: https://adminrules.idaho.gov/rules/current/58/580102.pdf
- Idaho Water Body Assessment Guidance, 3rd edition: https://www2.deq.idaho.gov/admin/LEIA/api/document/download/14844
- Geographic Information Systems (GIS) datasets that document where various criteria apply may be available. Contact Idaho's Integrated Report Coordinator.
- Geography and Timing of Salmonid Spawning in Idaho: https://www.arcgis.com/home/item.html?id=2bda6efd734041debce3191bf5870f16