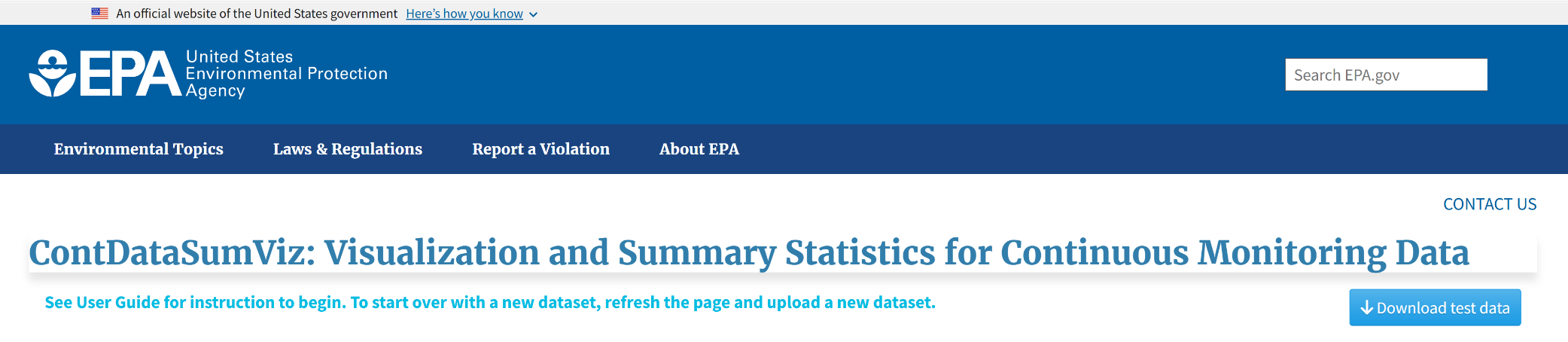
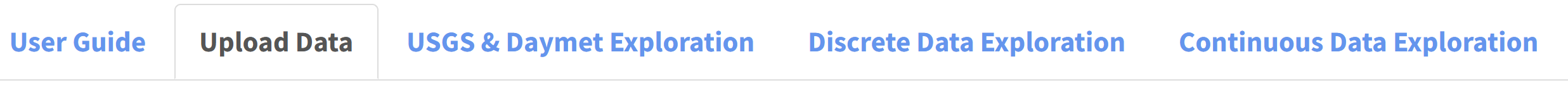
ContDataSumViz User Guide

Data Visualization & Reporting for Continuous Monitoring Data

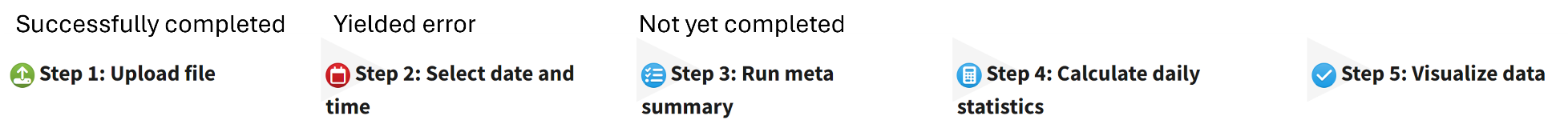
# Summary

This tool contains data summary and visualization modules for continuous monitoring data. Tool functionalities appear as the user provides required inputs. 

The landing tab—**Upload Data**—walks the user through the process of uploading their data. Test data are available from the Download test data button in the upper right corner of the application. As steps of the **Upload Data** tab are successfully completed, additional tabs will display at the top of the page. The **USGS & Daymet Exploration** tab allows the user to download USGS gage data at their gage of choice and Daymet daily weather and climatology data at their coordinates of choice, for the same time period as their uploaded data or a time period of their choosing. Users can then view this downloaded data alongside their uploaded data. The **Discrete Data Exploration** tab allows users to overlay discrete data points on a continuous time series of the same parameter. The **Continuous Exploration** tab provides daily summary statistics and visualizations for any uploaded parameters, as well as specific statistics for temperature and hydrology parameters.



Upload data progress is represented by the icons labeled with the upload data steps at the top of the page. Icons for steps that have not yet been completed are blue, successfully completed steps are green, and steps yielding errors are red.



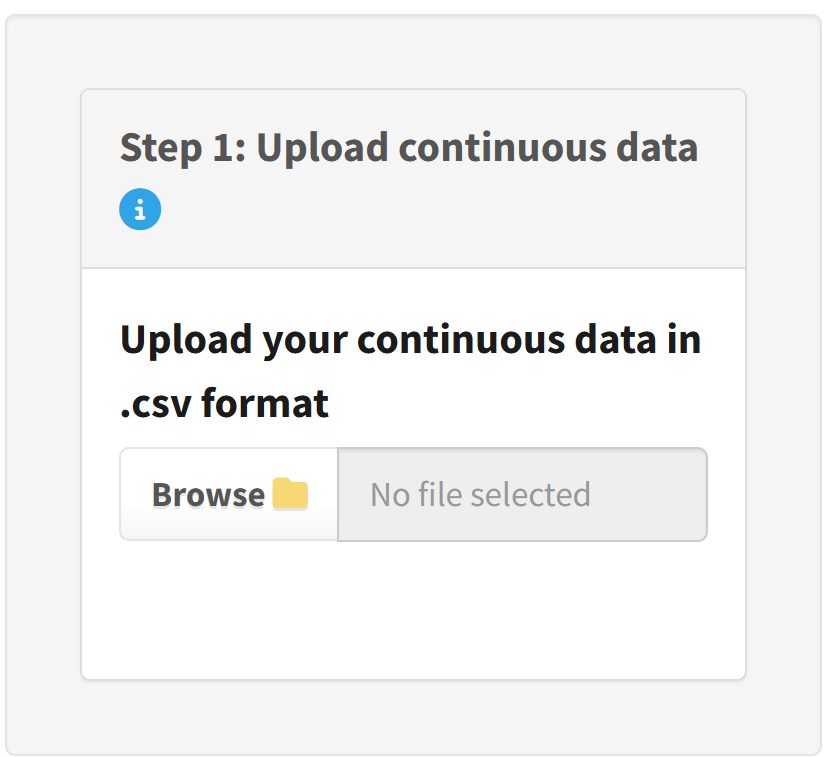
## Start over with a new dataset

To analyze a new dataset, refresh the page to reset the inputs and upload the new input data file in Step 1.

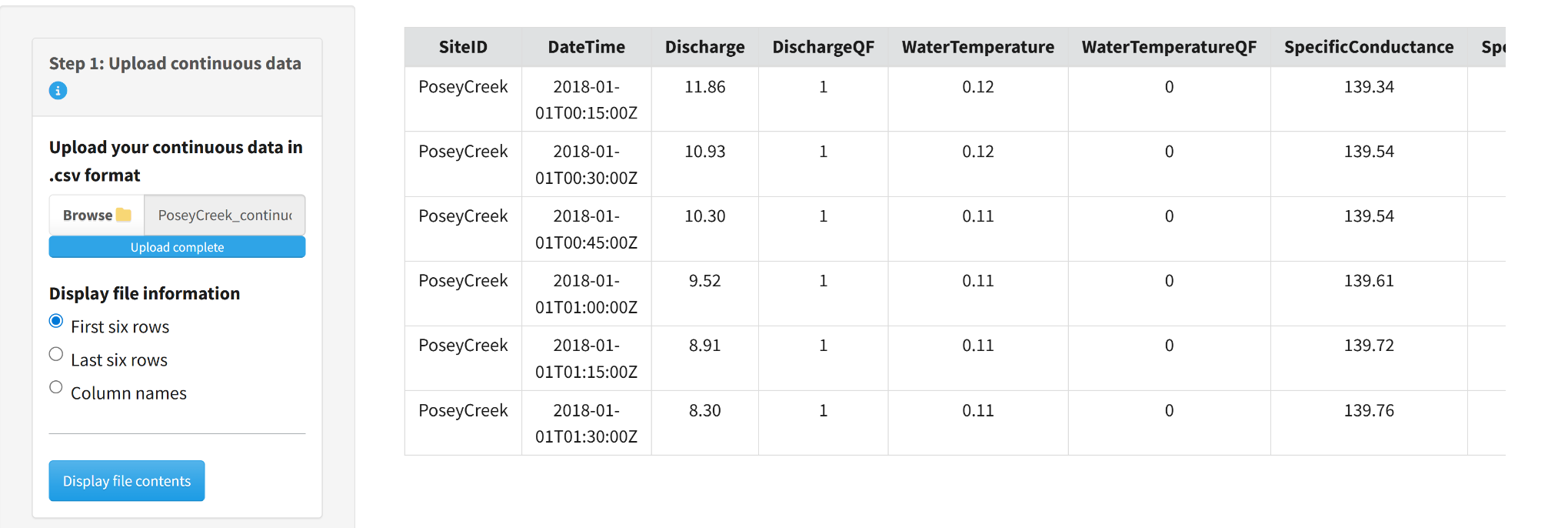
# Upload Data

## Step 1: Upload file

Ensure that your dataset is saved as a .csv file on your computer. To upload the dataset, select the browse button in the lefthand sidebar. Navigate to your file in the popup file explorer window and select open.



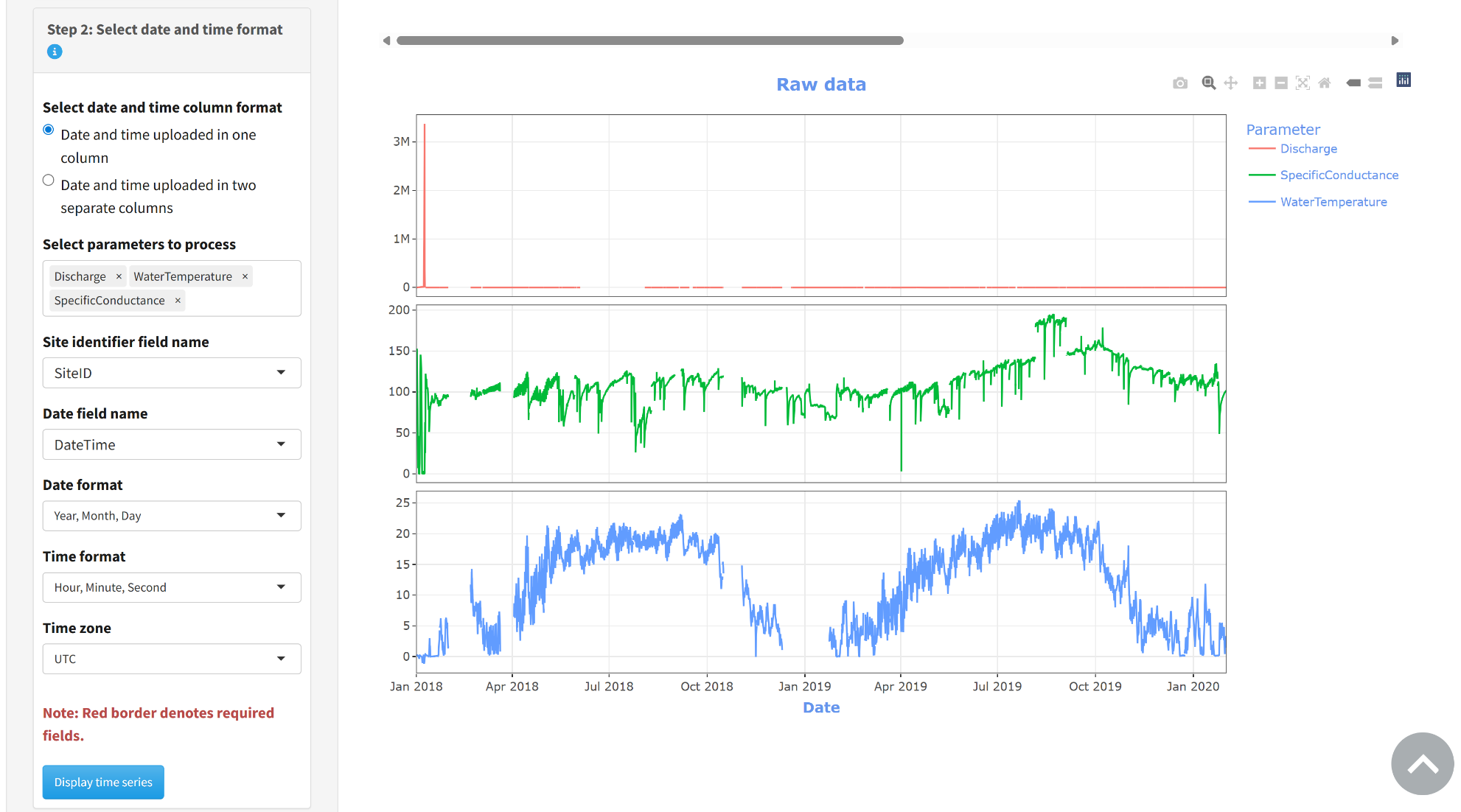
Select the file information you would like to display—the first six rows, the last six rows, or column names—then select display file contents. The date format displayed may be different from the display in Excel, because Excel often adds date formatting without alerting the user. If you open the file in a simple text editor like Notepad, the date format should match the application display.



## Step 2: Select date and time

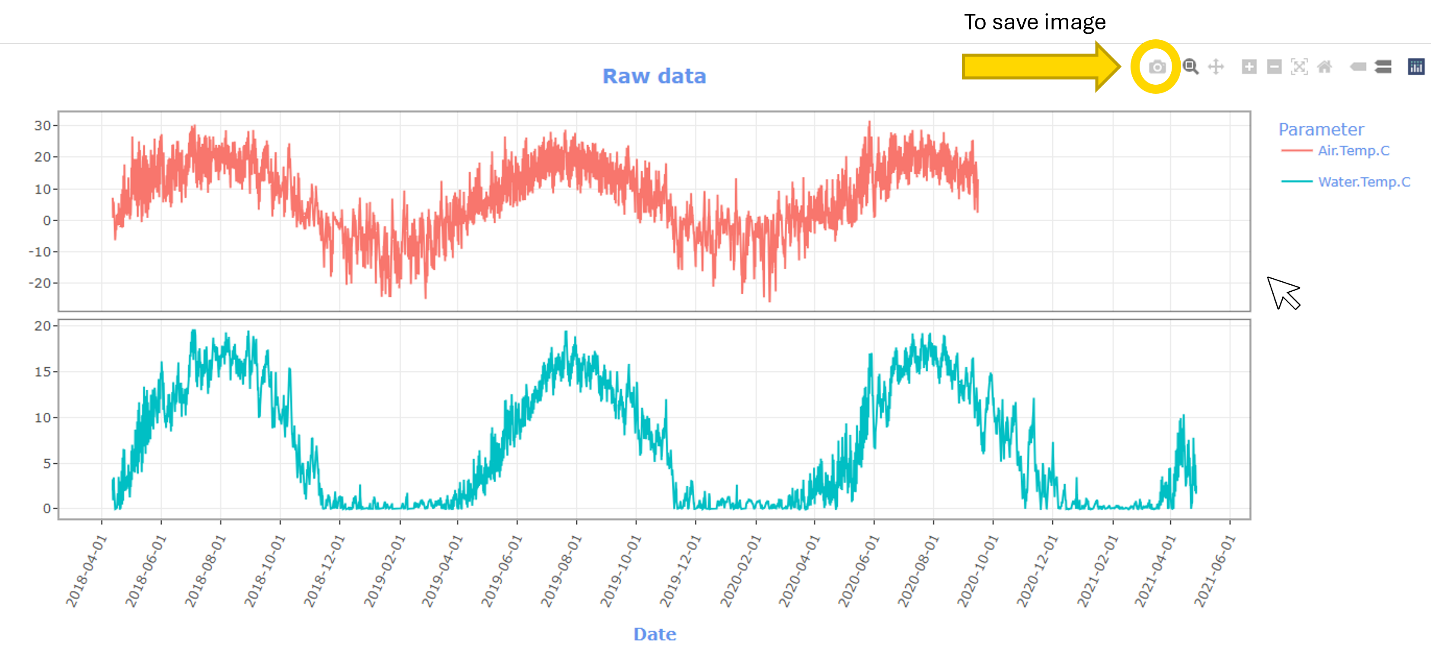
Selecting the display file contents button in Step 1 will display Step 2 in the lefthand sidebar. Select whether date and time are contained in a single column or two columns in the uploaded data file. Select the parameters to process, site identifier field name, date/time field name, date format, time format, and time zone using the dropdowns below. For more information about the time zone options, including their deviation from UTC in standard time and daylight savings time visit: <https://en.wikipedia.org/wiki/List_of_tz_database_time_zones>

If the date and time format parameters match the uploaded file, the time series will display after clicking the display time series button. If there is a mismatch, an error will return upon selecting the display time series button. Correct the error(s) in the parameter selections, then click Display time series, which will display the plot and Step 3.

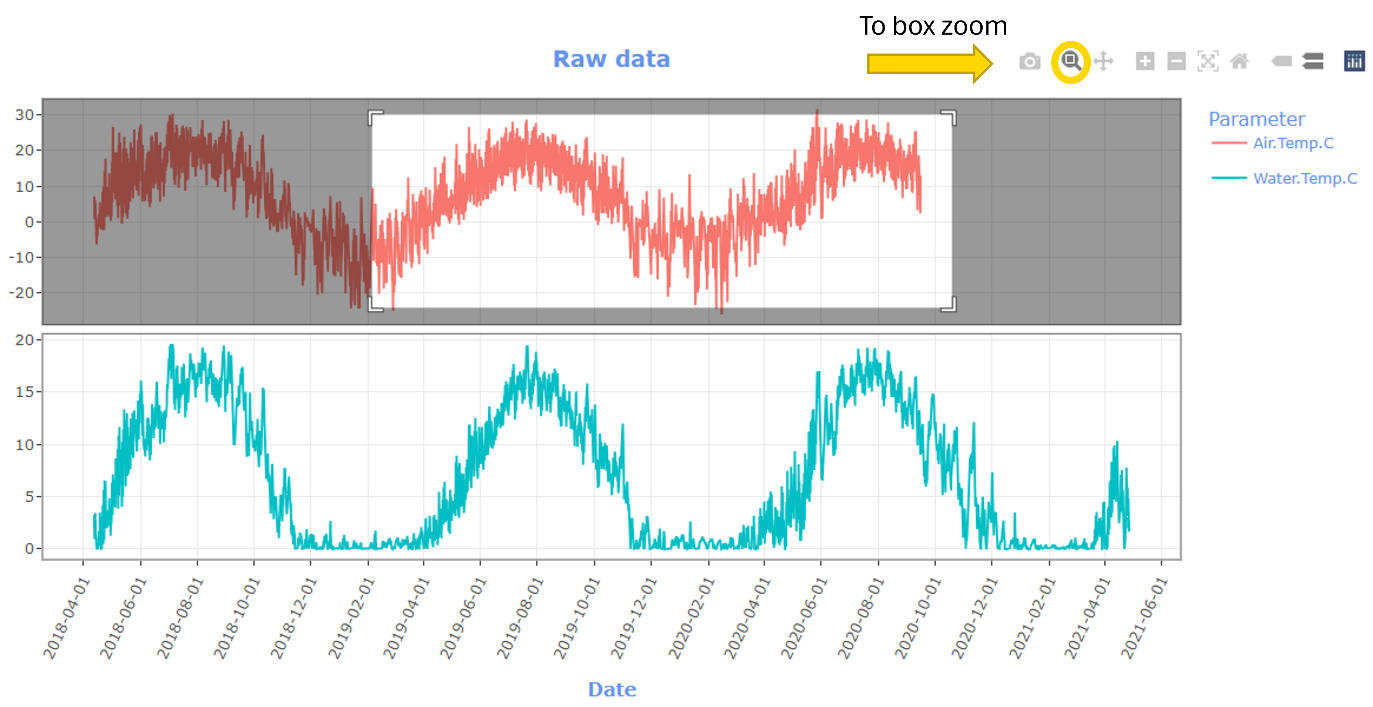


Single observation gaps (a missing value with non-missing data in the time point immediately preceding and following) are filled in for continuous data visualizations with the value preceding the missing one. This gap filling only occurs for the raw data time series, USGS gage/Daymet/base data time series, and discrete-continuous time series visualizations. Downloaded data will contain the same missing observations as the uploaded data. This feature is to account for minor differences in the frequency of parameter observations (e.g., one parameter has 15-minute observations, and another has 30-minute observations). Larger gaps (i.e., 2 or more consecutive NAs) are not filled. If parameter frequency differences result in larger gaps (e.g., one parameter has 15-minute observations and another has hourly observations), we recommend that users generate plots for these parameters separately or summarize the finer resolution parameter before uploading the data (e.g., for the 15-minute and hourly data, summarize the 15-minute data to 30-minute or hourly data).

Mousing over the time series will display the plotted values. To save the plot as an image, mouse over the plot to display the control panel in the upper righthand corner of the plot. Select the camera icon to open the file explorer window, then navigate to the location where you would like to save the plot and select save.

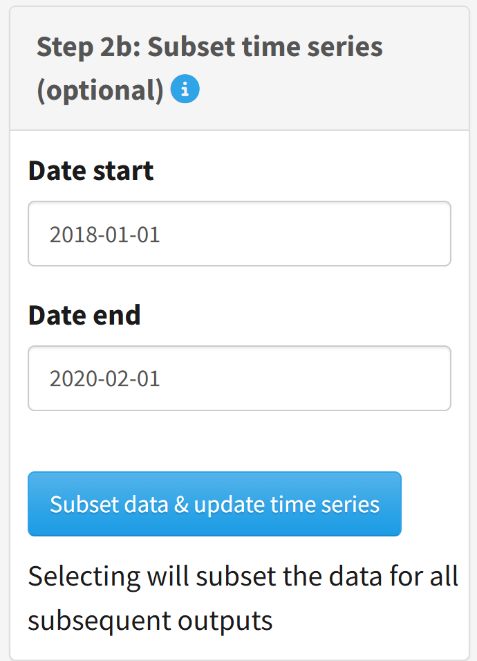


Zooming is available through the plus and minus buttons in the plot control panel, or by selecting the magnifying glass icon, which opens a box zoom that zooms to a user-drawn window. To reset the zoom, either double left click with the magnifying glass selected or select the home icon in the control panel.



## Step 2b: Subset Time Series (optional)

The start and end dates of this module will auto-populate with the start and end date of the uploaded data. To subset the time series for all subsequent analyses, input your desired start and end dates into the corresponding fields and select subset data and update time series. If you do not modify these fields, the entire uploaded time series will be analyzed.



If you would like to return to using the full time series after subsetting the data, please refresh the page and re-upload the data set.

## Step 3: Run meta summary

Use this module to generate a summary of missing and flagged data for each parameter, the period of record, and number of days in the period. First select whether the data contain quality flags. If you select yes, the module will display an input box for each parameter to indicate the quality flag column for that parameter. A single quality flag column can be selected for all parameters, or each parameter may have a unique quality flag column. Then, select the type of quality flags in the data. ContDataSumViz supports three quality flag types—fail, suspect, and not known—, which must be provided in the same column using unique flag codes. For each flag type selected, provide the corresponding code that indicates an observation has been flagged. Selecting the run meta summary button will generate a table summarizing the number of days with missing data, data flagged as fail, suspect, and not known for each parameter as well as the total period of record and number of days in the period. 

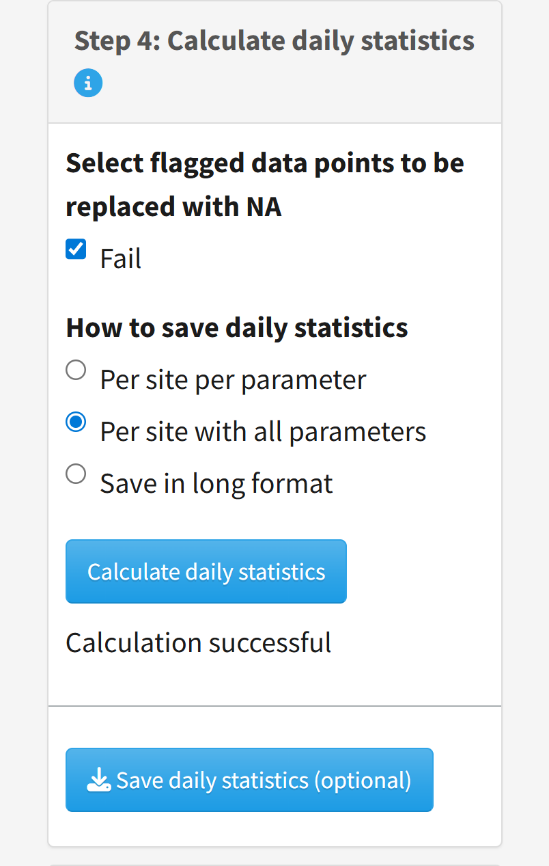
## Step 4: Calculate daily statistics

Check the boxes of quality flags previously identified in Step 3 to replace flagged observations with NA. Select the output format for saving daily statistics using the How to save daily statistics radio buttons. Select Calculate daily statistics to generate daily summaries and then optionally select Save daily statistics to download the calculated statistics. If you would like to change the daily statistics output format, you must select Calculate daily statistics again before selecting Save daily statistics.

*Per site Per parameter*: generates a zipped folder with different files for each parameter

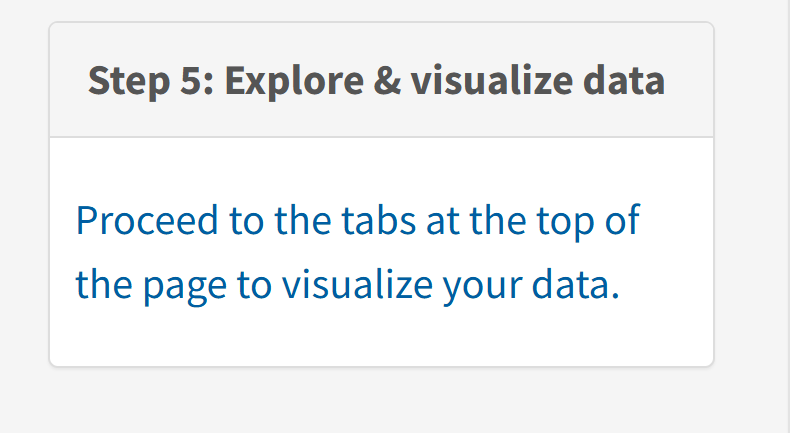
*Per site with all parameters*: generates a single file with summaries for all parameters

*Save in long format*: generates daily summaries with a single column for all parameters and a single column for their corresponding value for each date/time. Saving the data in long format is a step in formatting data in Water Quality eXchange (WQX) format for upload to the Water Quality Portal (WQP).



## Step 5: Visualize data

After successful calculation of daily statistics in Step 4, Step 5 will appear on the upload tab instructing the user to proceed to the other tabs at the top of the page to visualize their data. Select the text linked in the module to scroll to the tabs at the top of the page.



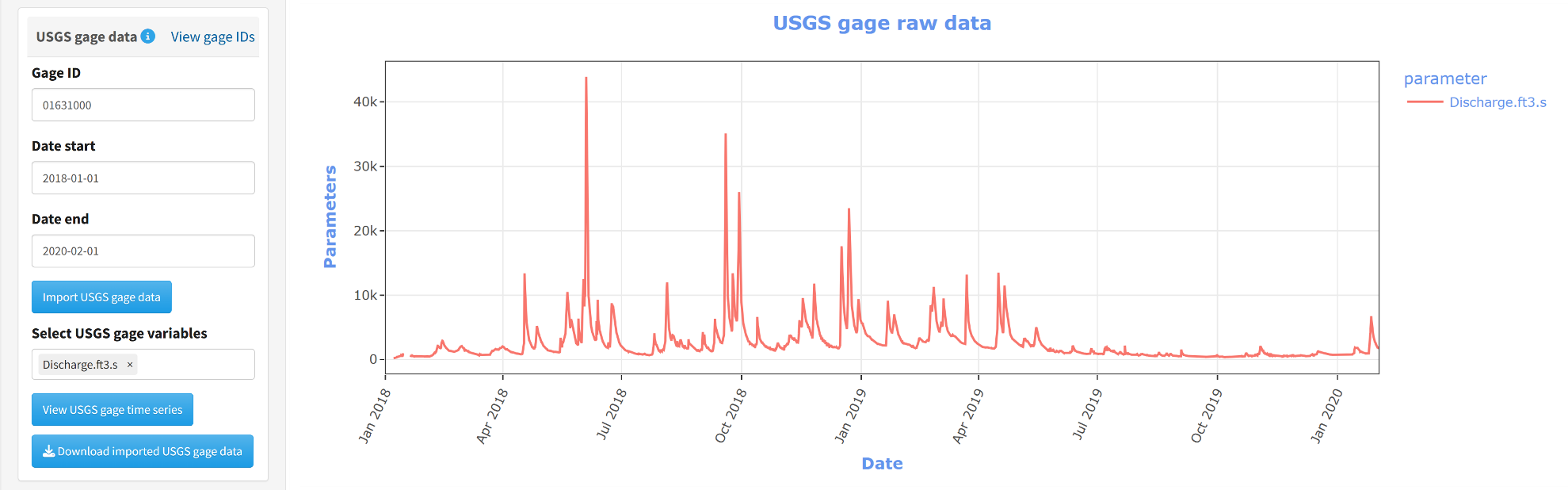
# USGS & Daymet Exploration

## USGS gage data

Input the id of the desired USGS gage in the Gage ID box. The Date start and Date end fields will auto-populate with the start and end date of the uploaded data but can be modified. The larger the requested date range, the longer it will take to download the data. After making these selections, click Download USGS gage data. To find a gage id, select the link in the upper right corner of the sidebar to visit the [USGS National Water Dashboard](https://dashboard.waterdata.usgs.gov/app/nwd/en/).

If you receive an error message, proceed to the USGS gage data error messages section of the user guide for additional guidance. When the USGS data have been successfully downloaded, the Select USGS gage variables box will appear. Select the desired variables from the dropdown and click View raw USGS data to generate a timeseries. When the data are successfully downloaded, the Download imported USGS gage data button will also appear, which you can select to save the downloaded gage data as a csv file.

The plot controls to zoom, pan, and save the USGS gage data plot are the same as for the raw data time series in the Upload Data tab.

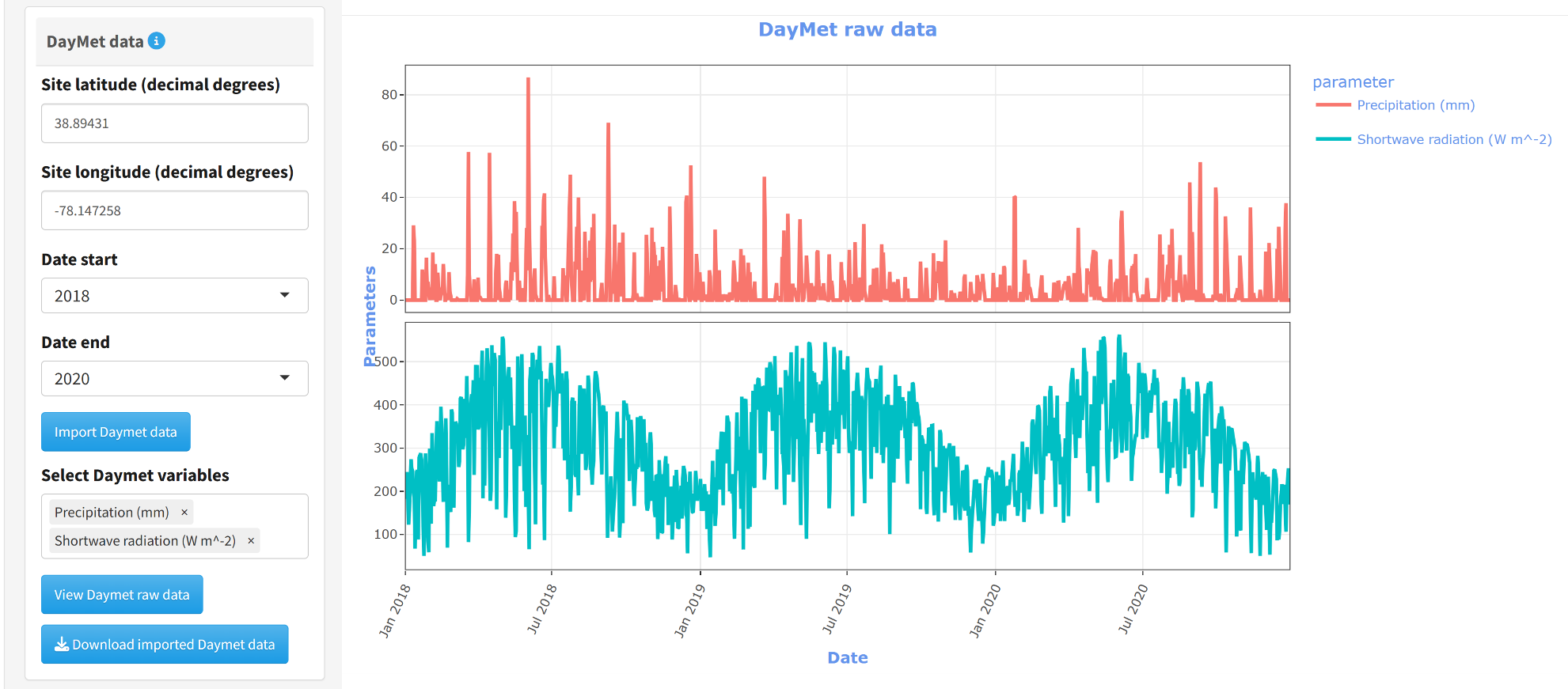


#### USGS gage data error messages

* “missing value where TRUE/FALSE needed”: this error generally displays when an invalid gage id is provided. Check that the id you are inputting is an id in the National Water Information System (NWIS) database.
* “arguments implying different number of rows: 1, 0”: this error can be displayed when the input date range is entirely outside of the available range for the gage. Check the site page for your selected gage to confirm that the available date range does not overlap that of the uploaded data. Select a new gage that has an overlapping time period with the uploaded data.
* Other errors: If the USGS web service times out, it is possible to see other errors. This may happen because too much data is being requested or the service itself is down. Try reducing the date range requested from the service. If that does not resolve the error, try downloading the data again in a few hours or days to rule out web service down time. See [USGS documentation](https://waterservices.usgs.gov/docs/instantaneous-values/instantaneous-values-details/) for more information on the USGS gage web service. If none of these solutions work, please submit an issue on the [ContDataSumViz repository](https://github.com/USEPA/ContDataSumViz/issues) with details about how this module failed (gage id and date range requested).

## DayMet data

To download modeled daily weather and climatology data from Daymet, input the latitude and longitude of your requested site in decimal degrees. The requested latitude and longitude must be in continental North America, Hawaii, or Puerto Rico. The Date start and Date end fields will auto-populate with the selected date range (in years) of the uploaded time series, provided they are within the range of available Daymet data. To modify the downloaded date range, change the dates in the corresponding fields. When Daymet data have been successfully downloaded, Select Daymet variables will appear. Select desired Daymet parameters (see table below) and select the View Daymet raw data button to display the time series. A button to download imported Daymet data will also display if the data have been successfully downloaded.



**Table 1**. Daymet parameters available for download in ContDataSumViz. Additional metadata concerning these variables are available on the [Daymet website](https://daac.ornl.gov/DAYMET/guides/Daymet_Daily_V4R1.html).

|  |  |
| --- | --- |
| **Parameter name** | **Meaning** |
| Precipitation (mm) | Total accumulated precipitation in water-equivalent depth in millimeters |
| Shortwave radiation (W m^-2) | Incident shortwave radiation in watts per square meter |
| Snow water equivalent (kg m^-2) | Snow water equivalent (estimate of the water contained in snowpack) in kilograms per square meter |
| Maximum air temperature (degrees C) | Maximum air temperature in a day in degrees C |
| Minimum air temperature (degrees C) | Minimum air temperature in a day in degrees C |
| Water vapor pressure (Pa) | Water vapor pressure in pascals |

## Merged base data, USGS data, and Daymet data

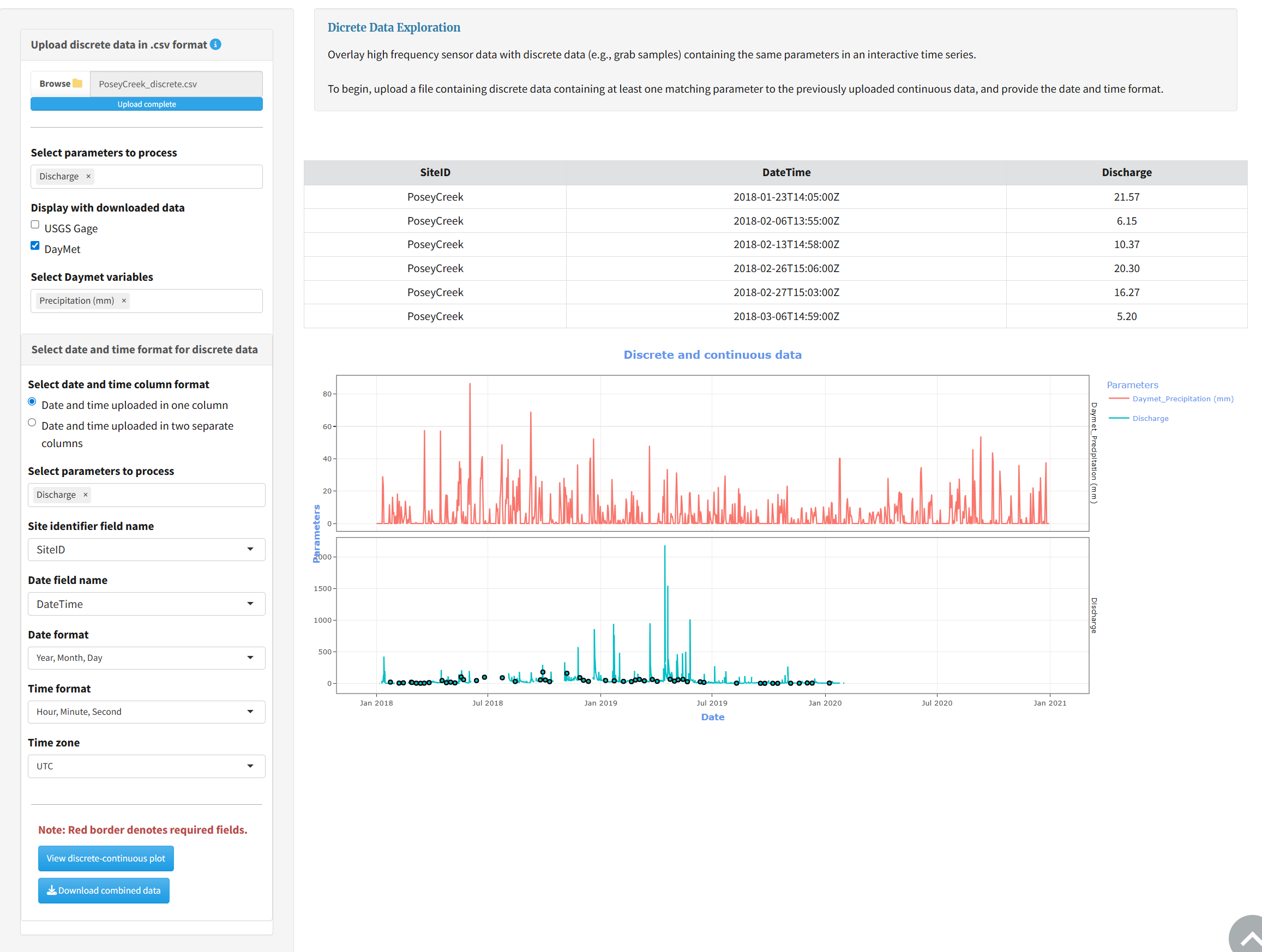
To display a plot with uploaded, USGS gage, and Daymet data, select the desired parameters from the Select base variable names dropdown and click View merged data. The plots will populate with the USGS gage and Daymet variables previously selected in the sidebar. A button will also appear in the side bar to Download merged data as a csv file.

# Discrete Data Exploration

This module allows users to plot discrete data on top of the uploaded continuous data. To begin, upload the .csv file with your discrete data. Select date and time parameters to match the uploaded data, as done in the Upload Data tab. The name of the parameter in the discrete file must match a parameter in the continuous data file, otherwise the application will display an error. Upon uploading the discrete data file, the application will display the first six rows in the main panel.

If you would like to view previously downloaded USGS gage or Daymet data along with the discrete-continuous plot, check the corresponding boxes and select the desired variables from the dropdowns.

Select the discrete data site identifier, date, and time format. Then select View Discrete-Continuous Plot to view the time series. The plot controls for zoom, pan, and save the continuous-discrete data plot are the same as for the raw data time series in the Upload Data tab. Merged discrete and continuous data can be downloaded as a csv file by selecting the Download combined data button.



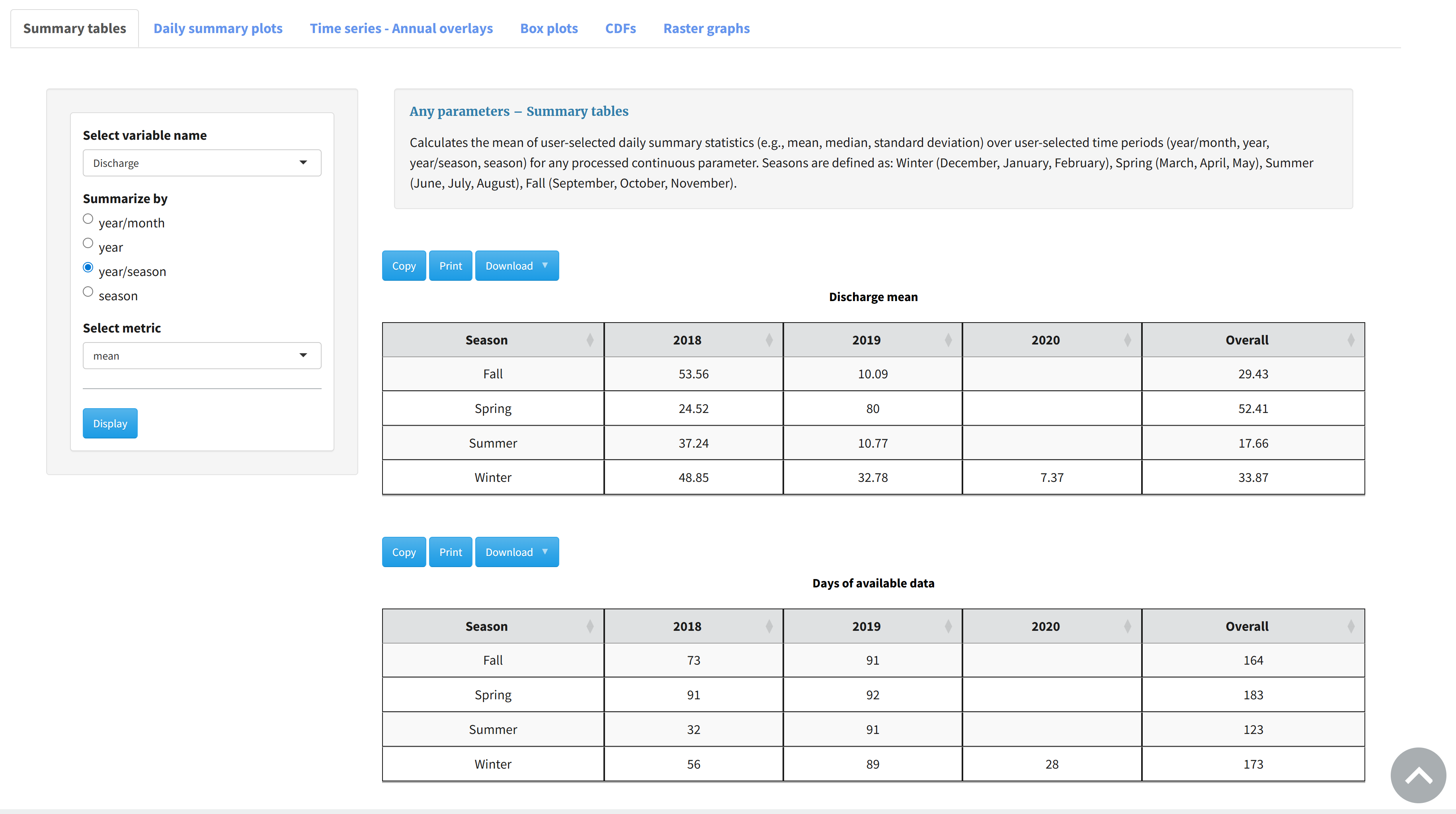
# Continuous Data Exploration

## Any parameters

Modules under this tab can calculate summaries and develop visualizations for any uploaded continuous parameter.

### Summary tables

This module allows users to calculate summary statistics for different variables, time periods, and metrics. Users can select any of the variables in their uploaded file from the Select variable name dropdown. Radio buttons display the time period summary options: year/month, year, year/season, and season†. The metrics options available in the dropdown are mean, median, minimum, maximum, range, standard deviation, variance, coefficient of variation, and number of observations. Options to copy, print, and download the summary table are displayed above the table. Downloads are available as csv, Excel, and PDF files. Below the results table is a table summarizing the number of days of available data used in the summary calculation. This table may be a useful reference for other summary plots like cumulative distribution functions, which are dependent on the amount of data provided.



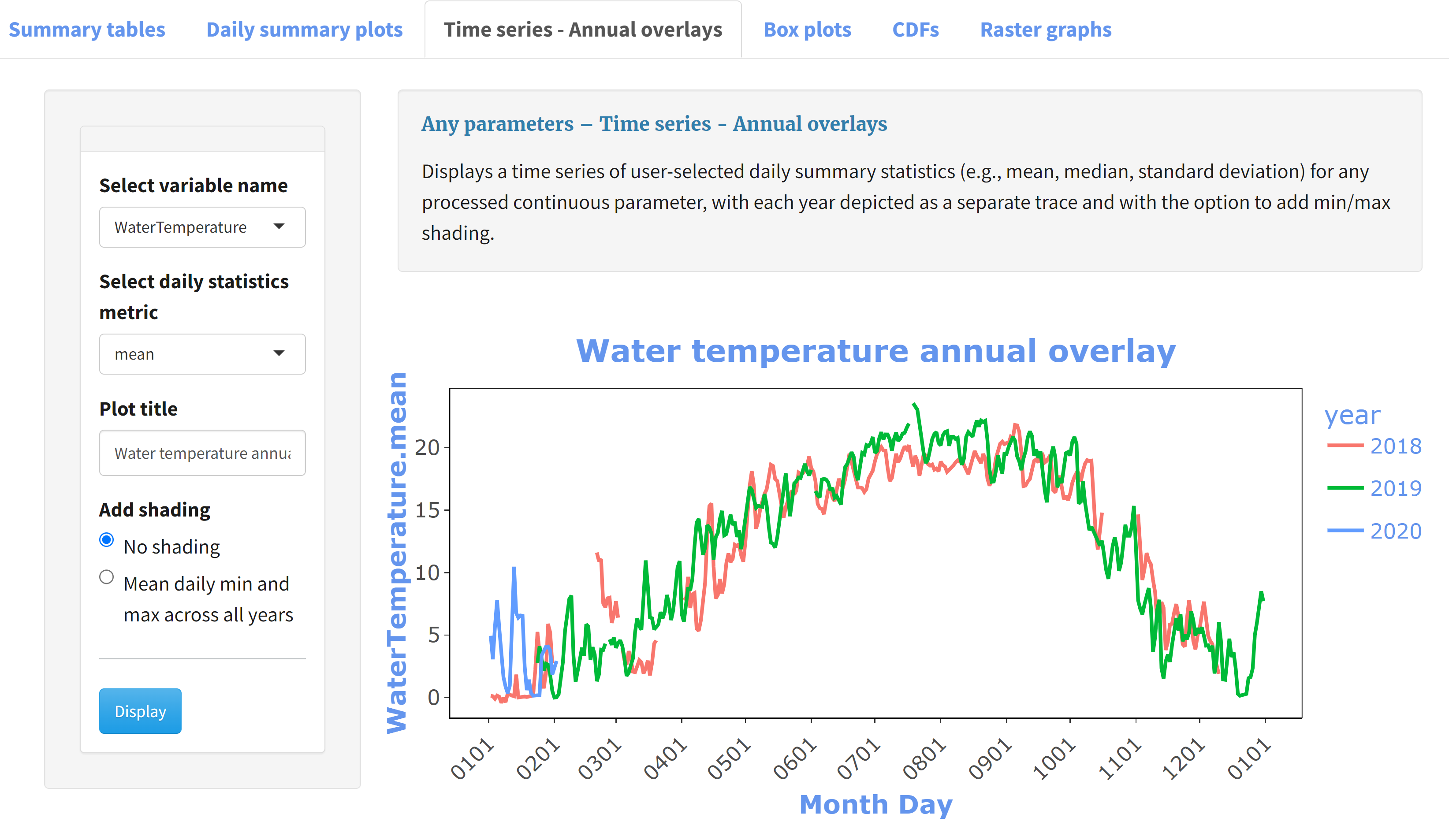
### Daily summary plots

This module displays time series of daily summaries for selected variables and offers the option of adding shading between the 25th and 75th percentiles or the minimum and maximum values. The Plot title text box allows users to alter the time series title. Available summary statistics are the same options as for the Summary tables tab above.



### Time series – Annual overlays

This module overlays time series of daily summary metrics for each year in the uploaded data. The available summary metrics are the same as the previous Any parameters tabs. Shading for mean daily minimum and maximum across all years can be added to the plot.



### Box plots

This module prepares boxplots of daily summary statistics grouped by different time periods. The daily statistics metrics options are the same as the above Any parameters tabs. The available summary time periods are: 1) month, 2) month (years side by side), 3) year, 4) season†, 5) season† (years side by side).



### CDFs

This module calculates and displays the empirical cumulative distribution function of the selected variable for each year subset by the selected season†. If no shading is selected, the plot is interactive and can be downloaded by mousing over the plot and selecting the camera icon. If shading is selected, the plot is not interactive. To save the plot, mouse over the figure, right click, and select Save image as.



### Raster graphs

This module displays raster graphs depicting daily summary statistics with rows representing years, columns representing days, and colors representing summary values. The Adjust plot aspect ratio parameter in the sidebar of this tab allows the user to adjust the plot aspect ratio. To save the plot, mouse over it, right click, and select Save image as.



## Temperature

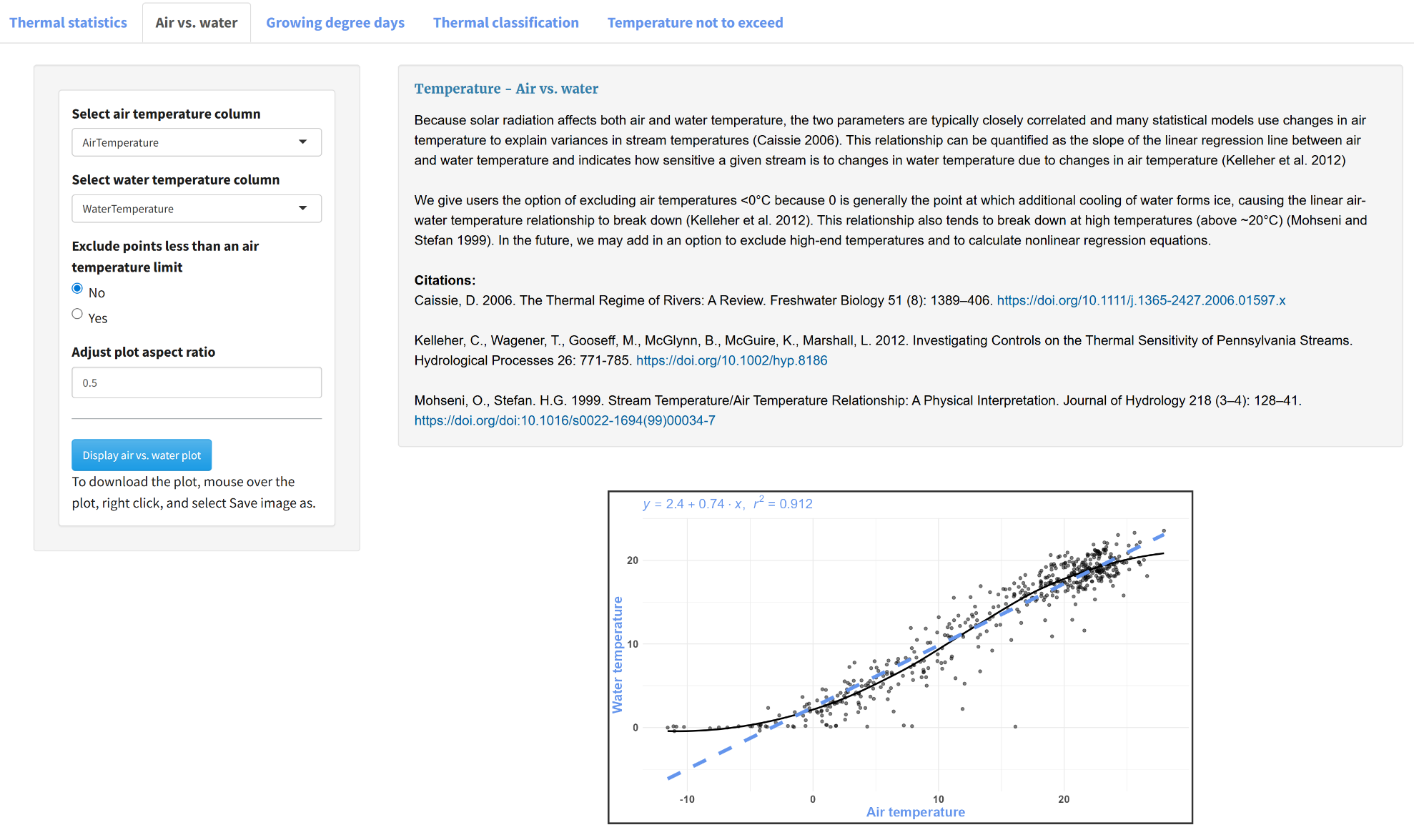
### Thermal statistics

This module calculates stream thermal regime statistics from [Tsang et al. 2016](https://doi.org/10.1080/03632415.2016.1210517). These metrics cover five categories of stream thermal regime components: frequency, magnitude, rate of change, timing, and variability. To run this module, select the columns representing SiteID and water temperature and input the critical temperature: the cutoff temperature used in the frequency calculations that sum the number of observations greater than the cutoff. To display the metrics tables, select Display stream thermal metrics. When the statistics have successfully calculated, the summary tables will appear in the main panel on the right and the Save thermal statistics in excel button will appear in the sidebar on the left, which generates an Excel workbook containing all the thermal statistics tables. Individual tables can be downloaded using the download button above the corresponding table. These tables can be downloaded as csv, Excel, and PDF files.



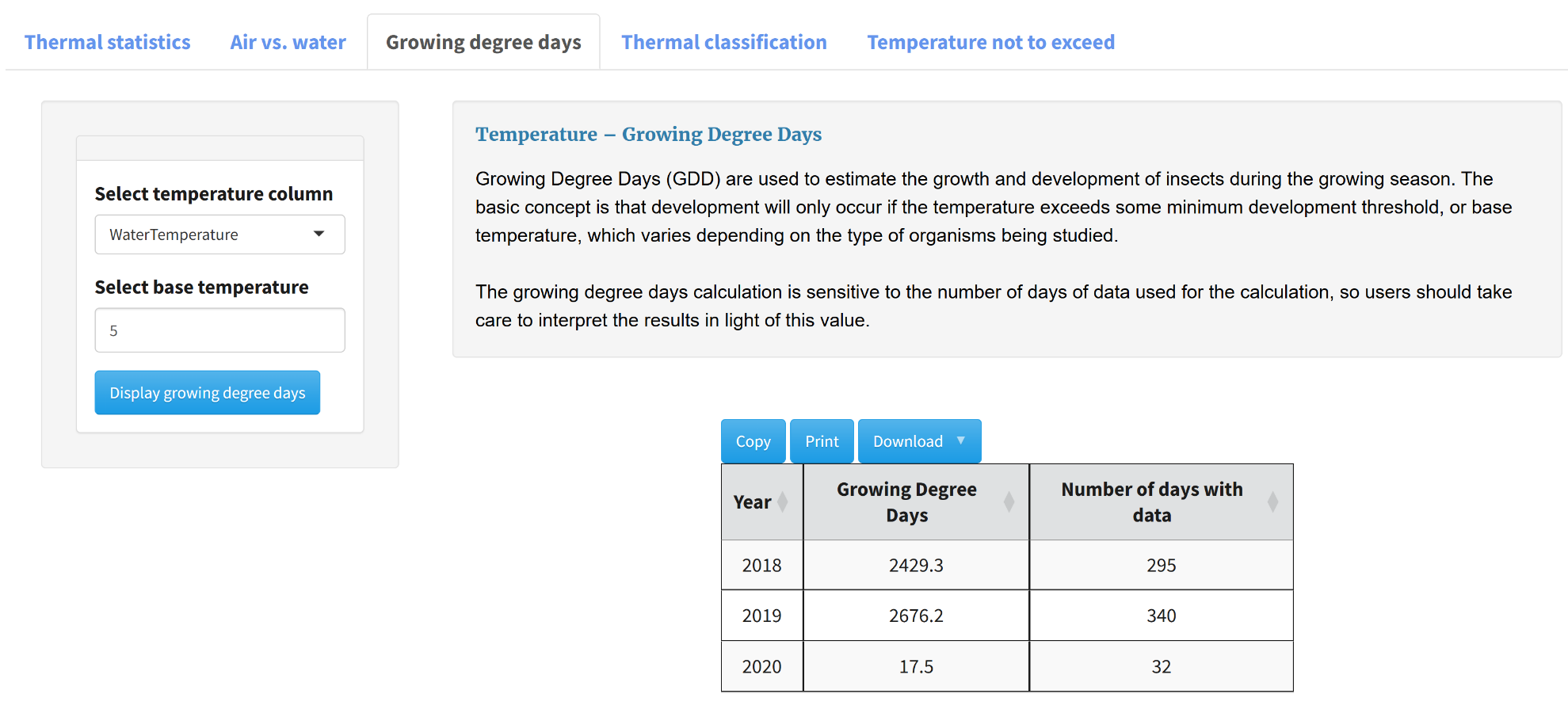
### Air vs Water

This module generates a plot of air temperature vs. water temperature. To save the plot, mouse over it, right click, and select Save image as. The blue dashed line indicates the linear best fit which corresponds to the regression parameters in blue in the upper lefthand corner of the plot. The black line represents a local polynomial regression fit of the data points in black. The sidebar contains an option to exclude values with air temperatures less than a specified value. The default value provided (when Limit the data points with air temperature = Yes) is 0˚ but can be adjusted. At 0 ˚C ice formation generally breaks down the relationship between air and water temperature.



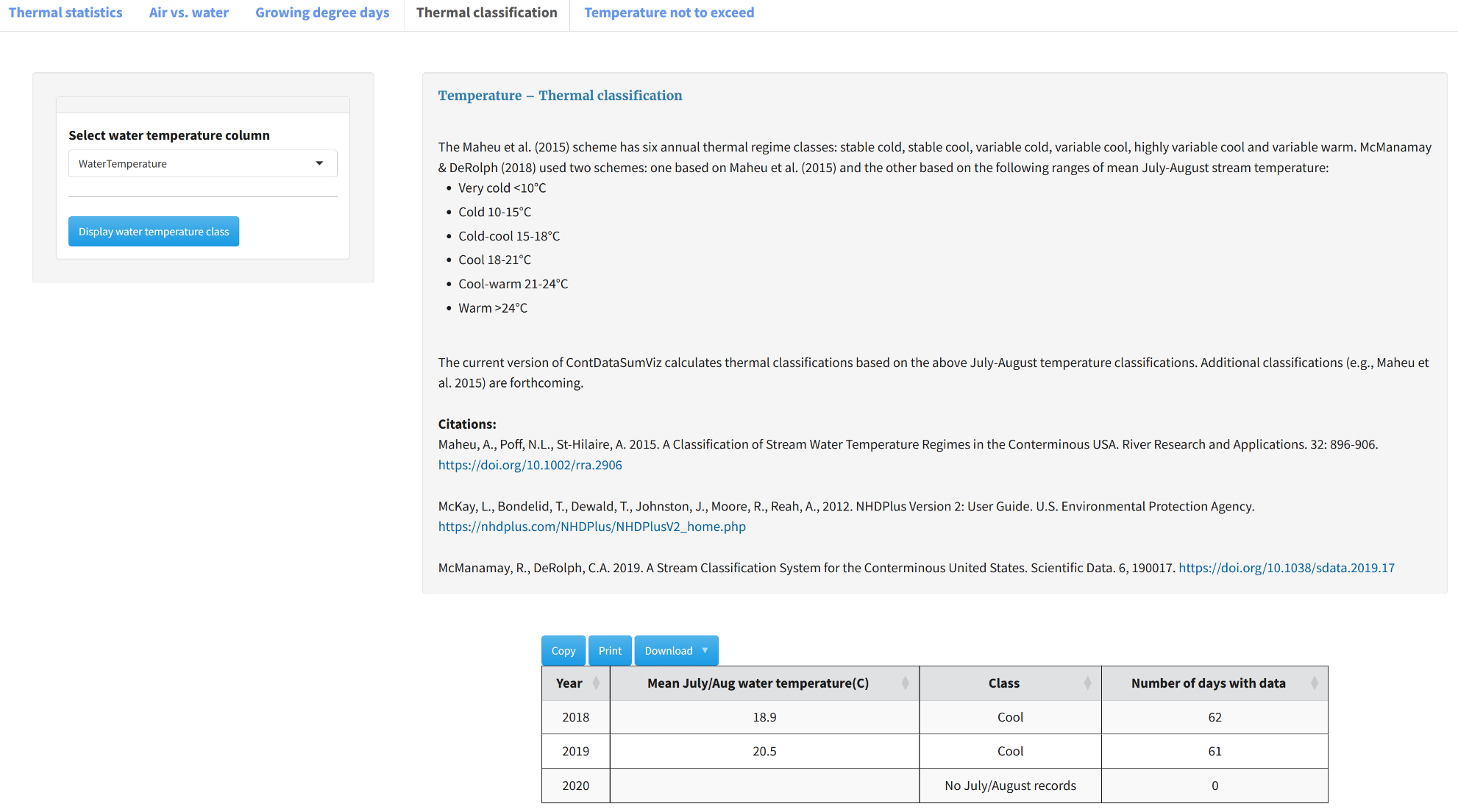
### Growing degree days

Select the temperature column and the base temperature for the growing degree day calculation. Left click “Display growing degree days” to run the calculation and display the results. Because this calculation is sensitive to the amount of data used, the resulting table displays the number of days with data as well as the growing degree days calculated.



### Thermal classification

This module classifies thermal regime by year based on a July-August temperature classification proposed by [McManamay & DeRolph (2018)](https://doi.org/10.1038/sdata.2019.17). Additional classification schemes like [Maheu et al. 2015](https://doi.org/10.1002/rra.2906) are forthcoming. To display water temperature class, select the water temperature column and left click “Display water temperature class.” If the uploaded data do not have July or August water temperature measurements for a given year, the class will be listed as “No July/August records.”



### Temperature not to exceed

This module was developed to support 4T3 and 6T3 calculations used in New Mexico water quality criteria.

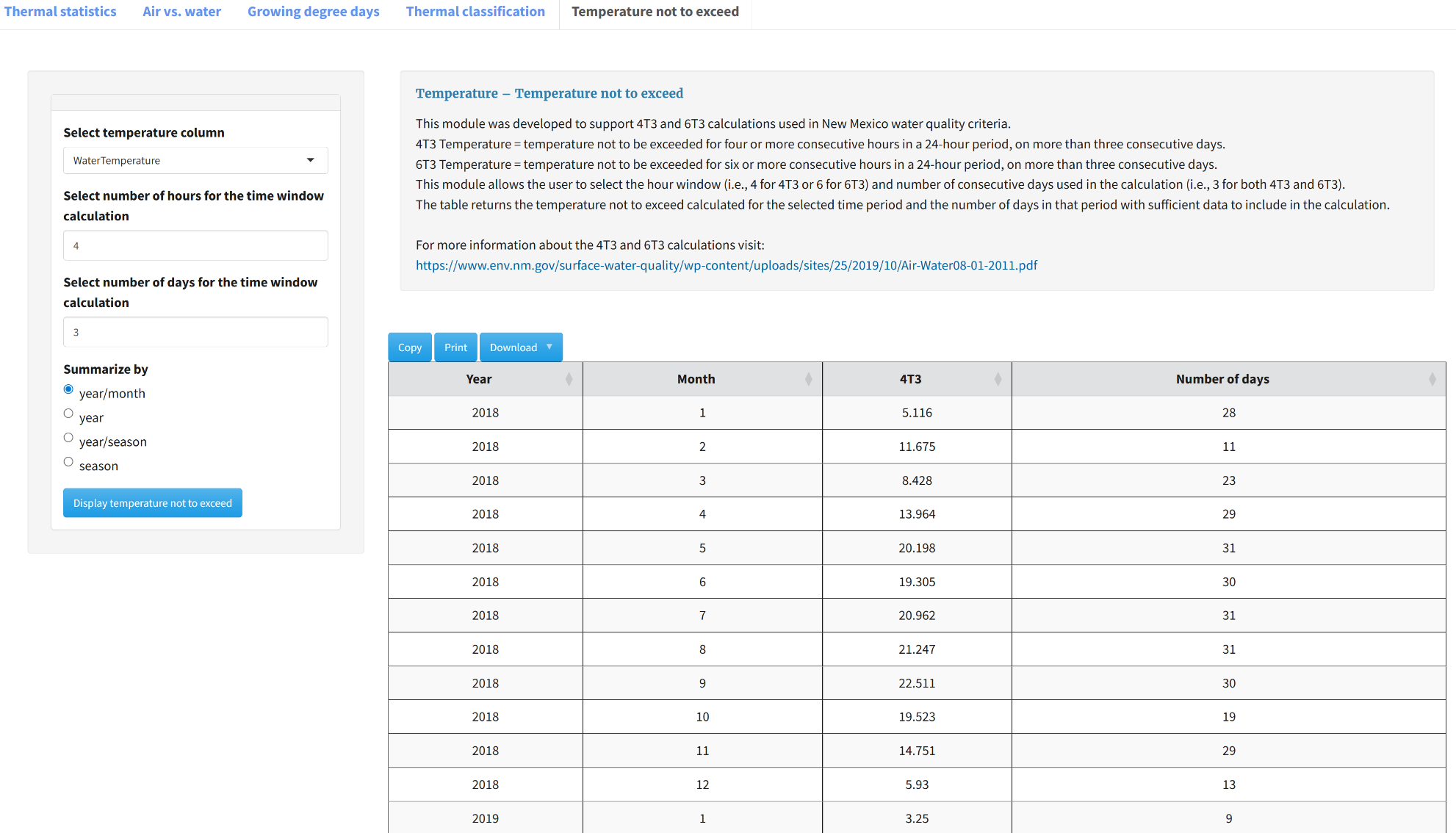
4T3 = temperature not to be exceeded for four or more consecutive hours in a 24-hour period, on more than three consecutive days.

6T3 = temperature not to be exceeded for six or more consecutive hours in a 24-hour period, on more than three consecutive days

This module allows the user to select the hour window (e.g., 4 hours in 4T3) and number of consecutive days used in the calculation (e.g., 3 days in 4T3). Results are returned by year and include the number of days included in the calculation.

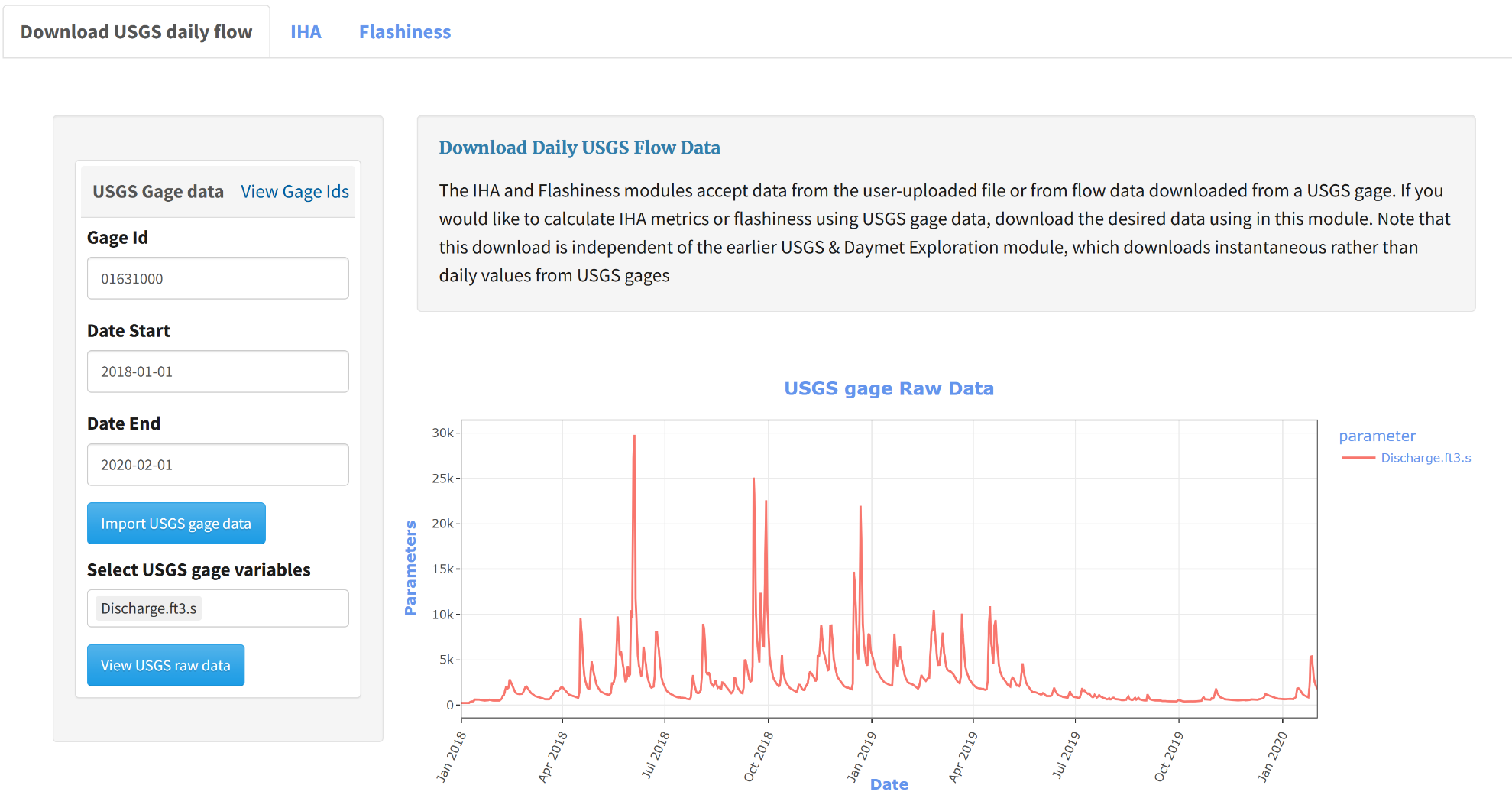
Note that times are included in the hour summary if times ≥ start time & times ≤ start time + # hours, while days are included if days ≥ start day and days ≤ start day + # days + 1.

To run this module, select the hour and day window and left click “Display temperature not to exceed.” The defaults are set to calculate 4T3. Note that this module may take several minutes to run, as the moving window analysis is computationally intensive.



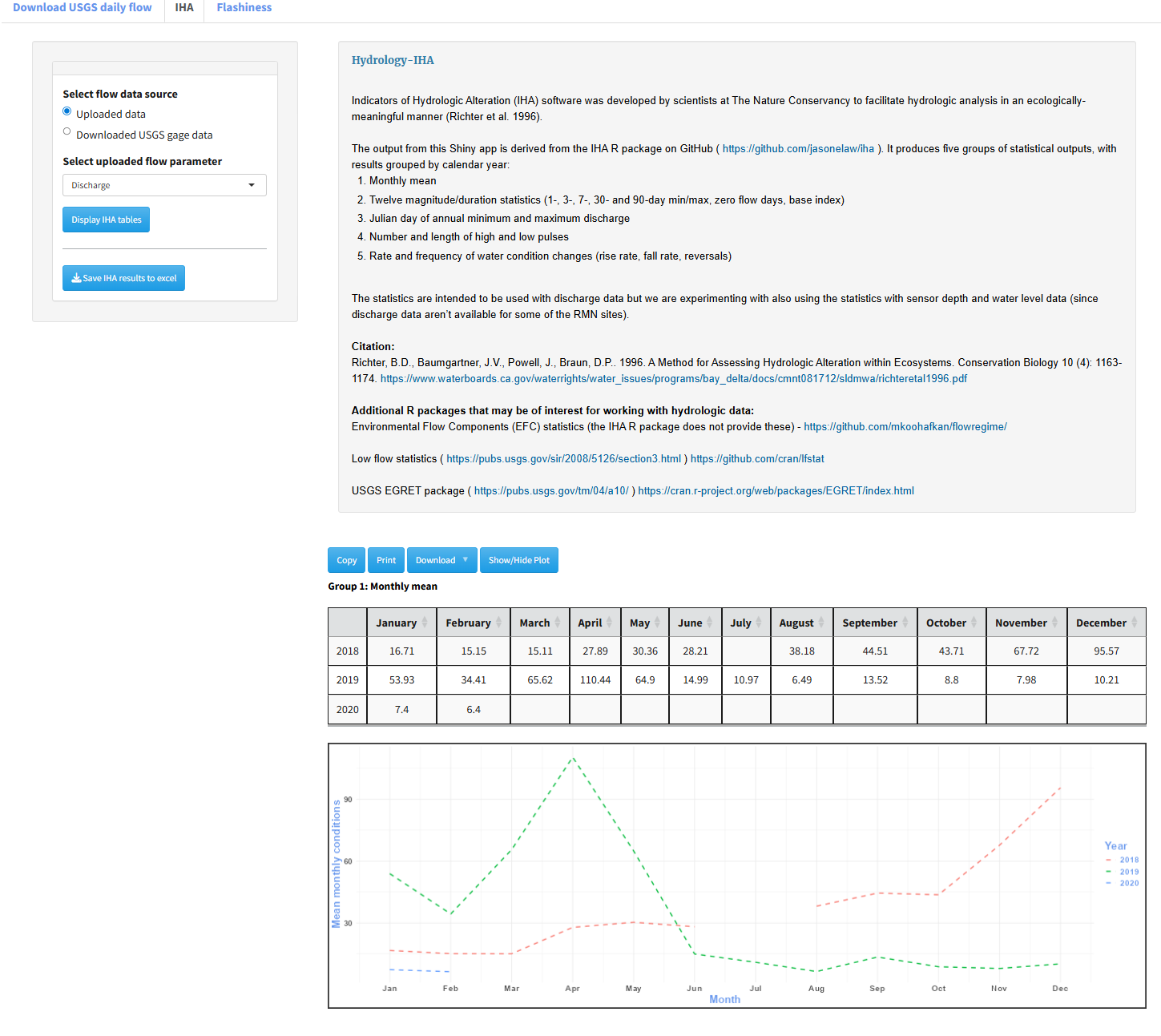
## Hydrology

### Download USGS daily flow

This module downloads mean daily values from USGS gages. Discharge data downloaded from this module can be used to populate the subsequent hydrology modules. This module is independent of the earlier USGS & Daymet Exploration module, which downloads instantaneous rather than daily summary values. To obtain data, input the gage ID, start and end dates, and select Import USGS gage data. To display a time series of the data, select the View USGS gage variables button. 

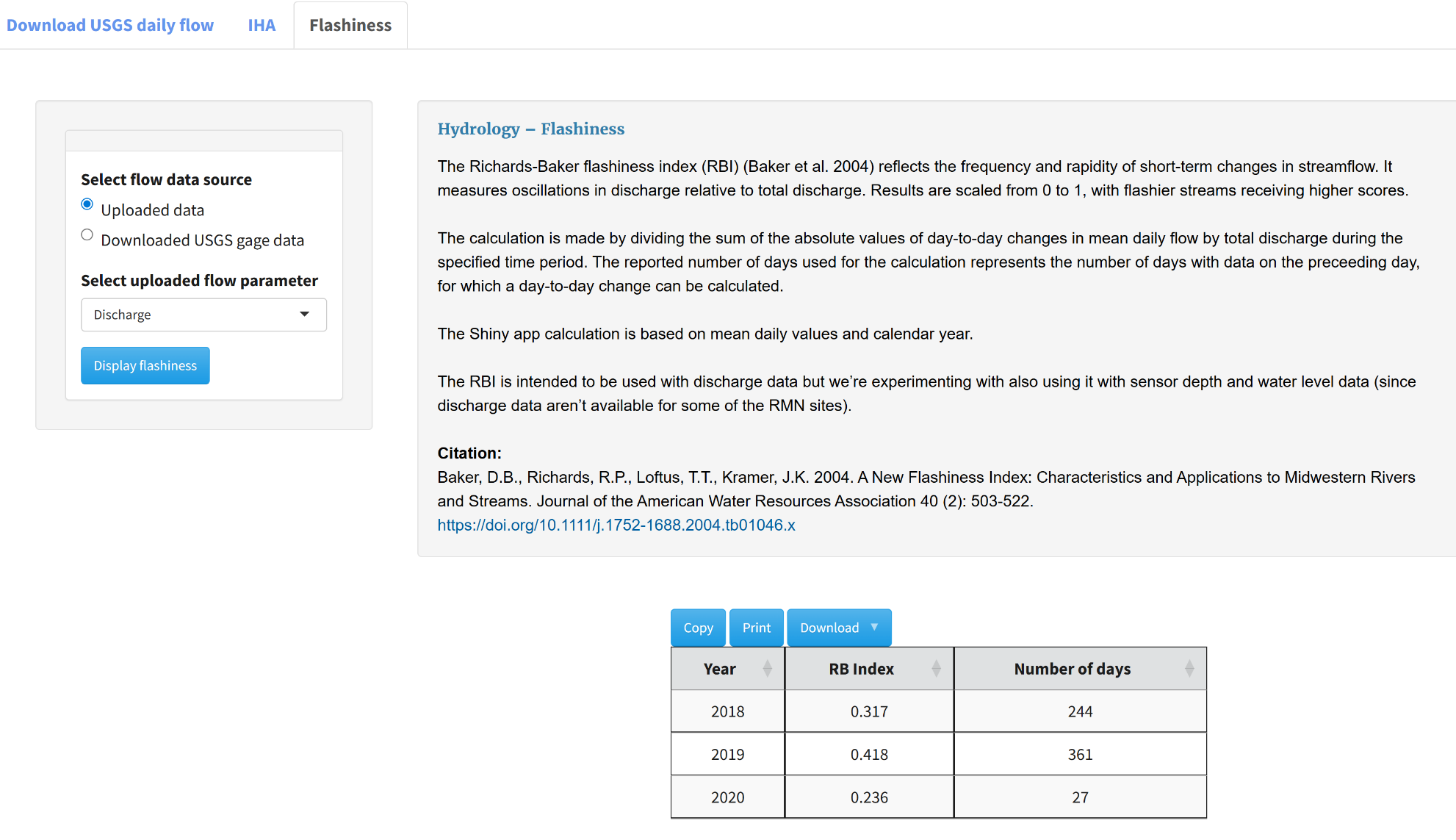
### IHA

This module calculates tables and figures for five groups of Indicators of Hydrologic Alteration (IHA): Group 1 Monthly mean, Group 2 Twelve magnitude/duration statistics (1-, 3-, 7-, 30- and 90-day min/max, zero flow days, base index), Group 3 Julian day of annual minimum and maximum discharge, Group 4 Number and length of high and low pulses, Group 5 Rate and frequency of water condition changes (rise rate, fall rate, reversals). Uploaded discharge data or downloaded daily discharge data from USGS gages can be used in this module. To download all statistics groups in a single Excel workbook, select Save IHA results to excel in the lefthand sidebar. To download individual tables, select the download button above the corresponding table. To display the corresponding plot, select Show/Hide Plot. To download a plot, mouse over the plot, right click, and select Save image as.



### Flashiness

This module calculates the Richards-Baker flashiness index, which represents the frequency and rapidity of short-term changes in streamflow. To calculate the Richards-Baker index, select the flow data source, flow variable, and left click “Display RB Index.” The resulting table summarizes the index for each year of data and displays the number of days used for the calculation, which includes days with flow data on the preceding day, for which a day-to-day change in flow can be calculated.



## Definition of season in ContDataSumViz

† Seasons in ContDataSumViz are defined as follows:

|  |  |
| --- | --- |
| **Season** | **Months** |
| Winter | December, January, February |
| Spring | March, April, May |
| Summer | June, July, August |
| Fall | September, October, November |