**Workflow for Water Year Typing Analysis:**

*Modeled Predicted Unimpaired WYT Analysis*

*Alice Beittel*

**Data Parts:**

*Reference Gages:*

* *Characteristics*:
  + Gages that have real unimpaired flow rates (no naturalized or modeled data), period of record is > 15 years and within **calendar years** 1950-2015, no data issues found
  + Columns used: USGS\_Gage and NHDVI\_COMI
* *Source*:
  + 173 gages pulled from set of FFC reference gage set of 223 that meet the above requirements.
  + For an in depth README of the FFC reference gage set, see Noelle’s “FFC data README” word document.
* *Additional Comment(s):* 
  + Hat Creek (Gage ID: 11355500, COMID/NHDV1\_COMI: 18020003000206) was on this data set, but **not** on “COMID\_Monthly\_Natural\_Predict\_173” Alyssa obtained from TNC California Natural Flows Database. Could be explained by the following note from Noelle’s README: “One additional gage was added to the reference list that had been previously excluded. This was a gage in Class 9 (Hat Creek) that had originally been used in the UCD river classification analysis. This was considered important to add in because before its addition, class 9 only had 1 gage. Reference years were considered 1968-1988 based on the standards for the original UCD analysis.”

*TNC Monthly Mean Flows for all 173 reference gages (COMID/NHDV1\_COMI):*

* *Characteristics:* 
  + COMID, calendar year, calendar month, Estimated.Q (mean monthly flow rate), 90th percentile, 10th percentile
  + Period of Record for all COMIDs: **water years** 1950-2016
* *Source:* [*https://rivers.codefornature.org/*](https://rivers.codefornature.org/)

**Steps:**

Load “COMID\_Monthly\_Natural\_Predict\_173.csv” into R.

1. **Convert calendar years/months to water years/months:**
   1. Water year start: October; water year end: September of following year (example: year ending September 30, 1999 is called the "1999" water year)
   2. select columns needed (COMID, YEAR, MONTH, ESTIMATED.Q)
   3. 'rename' current year/month columns to "calendar" year/month and make functions to do the conversion:
      1. function for converting calendar year to water year (if month >= 10, add one year to the calendar year, if not return calendar year)
      2. function for converting calendar month to water month (if month == 10, 11, or 12 subtract 9, if not return calendar month + 3
   4. execute functions and create new columns for water year/month

**Result:** New data frame called ‘monthlymean’ containing COMID, Calendar Year, Calendar Month, Estimated.Q, Water Year, Water Month

1. **For each COMID, calculate quartiles bin cutoffs based on the *entire period of record* of monthly mean flows** => Each gage will have unique quartiles but the same quartiles will be used for each year within a gage's POR to determine the WYT.

**\*\*\*** Final data frame, “Modeled.Predicted.Unimpaired.15+.WYT\_09.21.2018.csv" has quartiles based on 15+ years of record within the range 1950**-**2016.

* 1. group all the monthly avg flows (Estimated.Q) by each COMID
  2. make quartiles for each COMID for entire POR => Q1=25%, Q2=50%, Q3=75%
  3. check...median should match Q2
  4. **‘POR\_avg\_annual\_flow’:** average of all monthly means for entire POR for each COMID(gage)
  5. **‘Count\_monthlyavg’:** number of average monthly values used to calculate the quartiles, should be 792

**Result:** New data frame called ‘quartiles’ containing COMID, Q1, Q2, Q3, POR median, POR average annual flow, POR month count

*9.26.2018 – Calculate quartile bin cutoffs based on rainy months (entire range: Oct-Mar, only Oct-Dec, only Jan-Mar)*

*Quartiles date frame names:*

* *OMquartiles => October-March quartiles*
* *ODquartiles => October-December quartiles*
* *JMquartiles => January-March quartiles*

*9.26.2018 – eliminated all flow values of zero in ‘monthlymean’ and renamed new data frame ‘monthlymean’*

* *There is now a data set with the quartiles based on entire POR, but with all the zero values eliminated.*

1. **Calculate the mean annual flow for each COMID (gage) based on each water year** (*example*: average of all monthly flows for water year 1968 for one reference gage).
   1. Grouped by COMID and then by water year
   2. Calculated **‘Mean\_Annual\_Flow’** for each COMID for each year
   3. **‘Count\_monthlyavg’:** Number of water months used to calculate the mean annual flow for ONE year.

\*\*\*Note: Because data set started with Jan, first water year average will only have 9 months of data instead of 12. Likewise the last water year average stops Dec 2016 and will only have 3 months of data instead of 12.

**Result**: ‘monthlymean’ now changed to ‘annualmean’ and contains COMID, water year, mean annual flow, water year month count

1. Join the 'annualmean' and 'quartiles' by COMID to get one table.

**Result**: New data frame called ‘all’ containing information in both ‘annualmean’ and ‘quartiles’.

1. **Categorize all mean annual flow values into WYT quartile bins**.

**IMPORTANT NOTE**: Quartile values are currently written as minimum values for a WYT bin (example: if mean annual flow is greater or EQUAL to the Q1 value and less than but NOT EQUAL to Q2 value, then that year is "below moderate. If mean annual flow is less than but NOT EQUAL to Q1 value, then that year is “dry").

1. **Join USGS gage ID to corresponding COMID**

**Result**: New data frame called ‘WYT\_USGS\_all’ containing USGS gage ID, COMID, water year, mean annual flow, count of water months per year, Q1, Q2, Q3 POR median, POR average annual flow, POR month count, water year type

1. **Select all desired columns for final data frame and export data as csv**

**Result:** Final data frame containing columns below:

|  |  |
| --- | --- |
| USGS\_Gage | USGS gage ID |
| COMID | Stream segment ID |
| STATION\_NA | Station name |
| Water\_year | Water year calculated from calendar year |
| Q1 | 25% percentile flow rate |
| Q2 | 50% percentile flow rate |
| Q3 | 75% percentile flow rate |
| Mean\_Annual\_Flow | Mean annual flow calculated from the water months for that water year |
| WYT | Water year type  **‘dry’** = mean annual flow is less than but NOT EQUAL to Q1 value  **'moderately dry'** = mean annual flow is greater or EQUAL to the Q1 value and less than but NOT EQUAL to Q2 value  **'moderately wet'** = mean annual flow is greater or EQUAL to the Q2 value,and less than but NOT EQUAL to Q3 value  **'wet'** = mean annual flow is greater or EQUAL to the Q3 value |
| POR\_avg\_annual\_flow | Average annual flow calculated from all water months for the entire period of record for one gage |
| Count\_water\_months | Number of water months used in calculating that water year’s ‘Mean\_Annual\_Flow’ |
| Count\_monthlyavg | Number of monthly flow averages used in calculating ‘POR\_avg\_annual\_flow’ |

**File address:**

X:\environmental\_flows\WaterYearType\Results\Modeled.Predicted.Unimpaired.15+.WYT\_09.21.2018.csv

Looking ahead:

1. Compare quartiles to observed data
2. Depending on comparison, we will then move on to calculating water year type for all stream segments in the state using this method.

Re-do quartiles

1. Mean of POR
2. Make plots of all monthly fows for POR