**Water budget closure: extended work plan (FY18 work plan)**

**Personnel:**

*Scott Worland (lead) 50%*

*David Blodgett 25%*

*Andy Bock 25%*

*Tara Gross 25%*

*Sam Saxe (in kind)*

*Summer students? 50%*

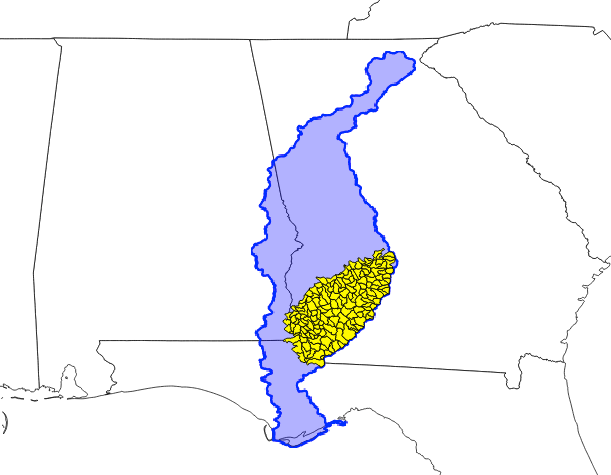
**Motivation and primary objective**

A simple water budget includes precipitation, streamflow, change in storage, evapotranspiration, and residuals:It is essential to include the human component–namely consumptive water use–to close the water budget and reduce the magnitude of the residuals. The largest withdraw volumes are for public supply, irrigation, self-supplied industrial, and thermoelectric. This project proposes to lay the groundwork for incorporating water-use data into national-scale water budgets.

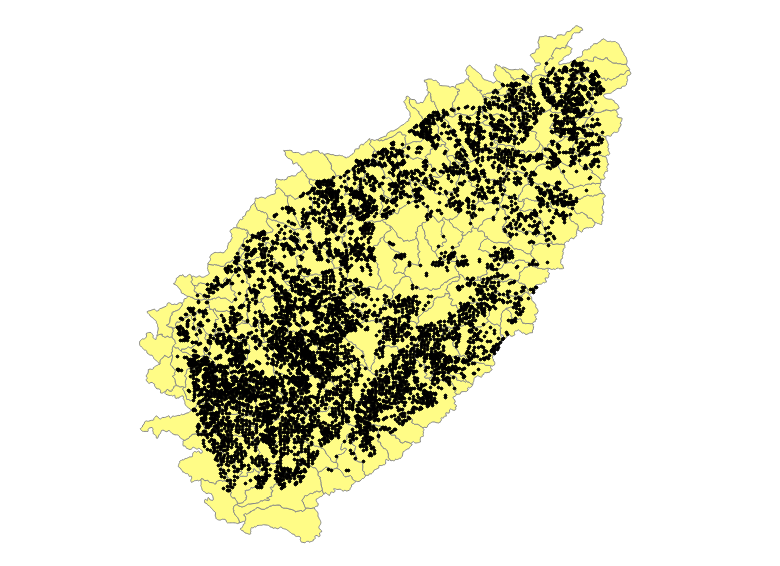
**Apalachicola-Chattahoochee-Flint (ACF) Basin case study**

**Phase 1: Site selection**

The first phase is an exploratory data analysis in a portion (164 HU12s) of the ACF basin in the state of Georgia.



The region was selected because it is part of a WaterSMART [focus area](https://www.sciencebase.gov/catalog/item/56abad4be4b0403299f4630e) and partial water-use data has been aggregated for the region for previous modeling efforts. The USGS South Atlantic WSC (Jamie Painter) has provided monthly values for ~8000 agricultural fields in the region for 2008-2012.



Spatially referenced data for thermoelectric, public supply, and self-supplied industrial consumptive water use are not currently available and compiling these data is a primary goal of phase 1.

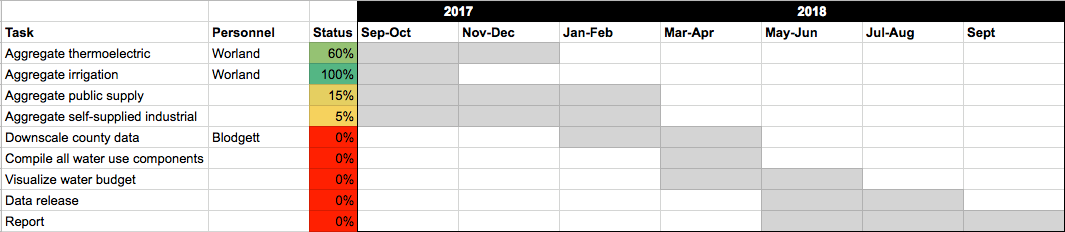
**Phase 1: Task examples**

A majority of the tasks for this project can be grouped into the following methodological categories:

1. **Data aggregation**: The project requires using data from a variety of sources (NWIS, local state data, etc) in a variety of formats (shapefiles, text files, databases). The specific tasks involve acquiring the data, documenting how it was generated, relating geometries (if applicable), and generated tables of time series consumption values for each water use category. Some individual tasks include:
   1. Using thermoelectric coefficients (Diehl and Harris, 2015) and EIA net generation data to calculate consumptive use at thermoelectric plants.
   2. Aggregate consumptive irrigation from coordinates to HUCs
   3. Associate NWIS permit numbers to individual well coordinates from the state of GA. Create polygons surrounding the cloud of wells (i.e., withdraw footprint) and aggregate the water-use data to HUCS.
   4. Do a similar task as described in C for self-supplied industrial
   5. Downscale county water-use data in AWUDS to HUCs
   6. Link water-use data to other components of the water budget
2. **Data visualization:** Generate plots of water budget components through time for HUCs. This will involve time series plots, residuals plots, spatial plots, and possibly and interactive combination of each.
3. **Data release:** The data from step one will be packaged as a single shapefile and made publicly available on Science Base.
4. **Report generation:** Lead the effort to write either an SIR or journal article of findings.

**Phase 1: Timeline**

Click image below to find link to up-to-date timeline information

[](https://docs.google.com/a/doi.gov/spreadsheets/d/1_vvjbhSh3_ZnG5Ri7w7HSZmce_dCnBJRr4ndd-uMw4k/edit?usp=sharing)

1/23/2018

* Need to do:
  + ~~Email Lan to setup a shared work area on the server~~
  + Concentrate on solutions #2 and #3 in Nancy’s emails
* Setup work directory:
  + D:/abock/WaterCensus18
  + Read through Nancy and Jamie’s email chain with Scott
    - On the WU/ga-permits.csv
      * CG is the permit\_tx (permit ID)
        + ≥ 7 characters, first three is the county
        + Find sites that share the same master permit number
        + Might have sites with no master permit record, or master permits with no source sites
      * CI column is the permit\_cd
        + NPDES (outfalls), EPWS (drinking water), and ALLC
    - Jamie’s email
      * CG column and site owner/alias name field can be used as an attribute for GW site types in NWIS
      * Kind of confusing, need to look for the NWIS fields
  + Think about attributes that we can use to delimit list to sites we don’t know.
    - Site w not lat long, but include county, huc12, or city
      * 5594 sites have lat/long
      * 5465 sites have huc 12 data
      * 4763 unique permit numbers
    - ~~Figure out GA county codes~~
      * ~~Tara is doing this~~
    - Need county shapefile
    - ~~Beginning and End dates?~~
      * ~~Doesn’t seem to be populated~~

1/24/2018

* ~~Lan set up a new workspace on studies~~
  + ~~Copied stuff over there~~
* Worked on the R-script
  + Read in NWIS, SWUDS
  + Create two new ID fields
    - 1 is the aggregate of SITE\_NO and AGECNY
      * Done for both the From/to in NWIS
    - The other field is the first 7 characters of PERMIT\_TX in SWUDS (Permit\_short)
  + Aggregated by master permit number
    - Calculated
      * Median Lat/Long
      * Percent of sites per permit no. that
        + Have NAs for lat long
        + Have 12-digit huc codes

1/25/2018

* Created tiers of sites
  + 100% of each permit\_short has huc 12 and lat/long (Tier One)
    - Can automatically assign huc12s?
      * Build Well Cloud from well x’s and y’s
  + 75 – 100% of each permit\_short has huc 12 and lat/long (Tier Two)
    - Make an educated guess for the Huc12
    - Are all huc12s the same for permit\_short
      * Assign HUC12 to row with missing Lat/Long
    - Else
      * Look at bringing in county/city data
      * Might have to look at where data plots
  + 50 – 75% of each permit\_short has huc 12 and lat/long
  + 0 – 50% of each permit short has huc 12 and lat/long
  + 0% of permit\_short have huc12 or lat/long
* Constructed initial code to convert points to polygons using the SF package
  + 2\_BuildGeom.R
* Tomorrow –
  + - Can we use county/city information for permits that have no HUC12 or Lat Long
    - Fields that might be useful (from GA\_permits.csv)
      * County Name/Fips code
      * Civil division codes

1/29/2018

* Questions
  + site owner/alias name field – PPI? Only 114 sites in the SWUDs data
  + Just looking at ALLC/EPWS and AG/GW for drinking water
  + Unique IDs in the nwis file require a combination of
    - from\_agency + from\_site\_no + to\_agency + to\_site\_no +water\_cd+year+cn\_qnty\_cd+water\_use\_subtype\_cd+meas\_meth\_cd+data\_source\_cd
* The goals
  + 1. Derive site to aggregate relationships just using the tables
  + 2. Derive potential disaggregate relationships based on spatial data
    - HUC12
    - Lat/Long
    - Address
    - City/County Name
  + 3. Once the relationships are derived (which site may contribute to which Huc10/12)
    - Look at disaggregating the pumping data
* What we have for Georgia (not just subset for the ACF yet)
  + Looking at sites that are permit types “ALLC” and “EPWS” and are site type “AG” and “GW”
  + Ran some preliminary analysis by Permit\_TX (first seven digits)
    - Proportion of sites for each master permit that have
      * Huc12
      * Lat Long
      * Physical Addresses (street/city/zipcode – PO Box??)
    - Number of permits
      * With aggregate sites w/ no well sources
      * With well sources but no aggregate sites
    - Classified master\_permit based on the amount of information available for the individual well Sites:
      * Tiered based on master permit
        + Tier 1 – 100% populated for HUC12 and Lat Long
        + Tier 2 – 75-100%
        + Tier 3 – 50-100%

Take an educated guess for missing lat/long

Median Lat/long

Most common huc 10/12

* + - * + Tier 4 – 0 – 50%

Might require more manual work

Require multiple allocation/disaggregation scenarios

* + - * + Tier 5 –

No spatial information

* + - Before we go into any manual work, subset to just sites within our area of interest.

1/30/2018

* Met with Tara
  + Developed schematic powerpoint for project
  + She is going to add information on site\_no duplicates
  + With the info summarized we will shoot an email to Scott and Jamie
* Cleaned up the code and results considerably
  + Corrected the “Permit\_short” field calculation that was truncating EPWS GA permit IDs
  + Calculated the correct number of AG Sites with no GW children, and GW sites with no AG parents
  + Wrote out two files into tiers
    - 1 – children
    - 2 – parents
* NWIS File
  + Eliminate codes we don’t need
    - i.e. just “GW” in the “water\_cd”, not “SW” of “FN”
    - what is
      * cn\_qnty\_cd
      * water\_use\_subtype\_cd
      * meas\_meth\_cd
      * data\_source\_cd

2/6/2018

* Well Locations
  + Looked at well xy’s
  + Most that did not have identical tabular/spatial huc12s were because the huc12 were incorrectly attributed
  + Finished ALLC locations
  + Some linked by site\_no

2/7/2018

* Evaluated EPWS sites
  + 122 sites in the study area
  + Only two need to be double-checekd
  + Added attribute for how many EPWS locations were in AG
    - 63 out of 122 in AG list
  + No site\_no matched??

3/9/2018

* Work Wrap up
  + Andy
    - ~~Clean up 1\_csvLook.R to the final version~~
    - Clean up 2\_BuildGeom.R to the final version
      * ~~Get rid of well point clouds~~
      * ~~Write out files for~~ 
        + ~~GW~~
        + ~~EPWS~~
      * ~~Change overlay from Huc10 to Huc12~~
    - ~~Document Manual Changes to the Huc10 Field~~
    - ~~Perform overlay for EPWS Wells~~
    - Write new output file relating Huc10, Permit #, and Well ID
      * Fields we may have to consider in EPWS
        + Dups

Same lat/lon as ALLC locations

Do they have corresponding AG Sites

In\_AG

Sites with corresponding permits in AG File

IN Site\_no

Have sites in AG File

Only a few sites

What about the rest

* + Tara
    - Convert units (if necessary)
    - Re-aggregate volumes to the HUC10 level
  + Andy and Tara
    - Document procedure
      * Methods used to derive volumes
      * Statewide-statistics
      * characteristics of missing and/or mislabeled data
      * Emphasize which fields would be the most helpful if populated and/or properly populated