DS421 SUMMER 2016 PROJECT SUMMARY

SARA STOUDT

1. Investigation of Chlorophyll

1.1. **Goals and Approaches.** The overarching goal of this portion of the project is to understand what variables help explain the variation of chlorophyll over time and at each station. We want to understand if these relationships differ over time and/or over space. My approach was to first see which other variables that we have information on at the same times and places as chlorophyll are correlated with chlorophyll. I included these variables in what I refer to as the "parsimonious" model. I then added to a "full" model additional variables that were identified by David as variables whose relationships with chlorophyll, in context, would be interesting to understand. I built models including these variables separately for each station. Then I experimented with truly spatial models by using the station ID as a factor in a model, so I could fit one model and be able to predict chlorophyll for each station at once. After this I went back to make more "bare bones" models per station that only incorporate the date (to incorporate trends over time) and the day of year (to incorporate seasonal trends) as well as the interaction between the two, allowing these trends to change depending on the value of the other one in order to have more fair comparisons between the various models. I then moved towards models that incorporate a sense of "flow" using chlorophyll values from neighboring stations and measures of true flow throughout the area.

  Throughout all of these models, I fitted the models in a nested way so that we could see how much each variable contributed to the overall fitted value. I created a Shiny application (a web application written in R) to be able to compare and contrast the fitted values from each model, the contributions of each variable to the fitted values in each model, the overall root-mean-squared-error (RMSE) per station of each model plotted on a map, and the contribution of volumetric flow averaged yearly and averaged monthly for each station.

1.2. **Details About the Setup.** I chose to use a log link function in each Generalized Additive Model (GAM) since the distribution of chlorophyll is highly skewed right. I was deciding between log transforming the response and fitting a linear model ($log(y) = X\beta + \epsilon$) versus using a Generalized Linear Model using a log link function ($log(y + \epsilon) = X\beta$) The residuals versus fitted values plots all look reasonable for constant variance under the GAM approach, so it did not seem necessary to use the log transform which would affect the variance as well as the linearity.

With this choice we also don't have to deal with the back-transform bias where the back transform of the mean of the response on a log scale is not equivalent to the mean of the response on the raw scale.

I also had to choose which stations to focus on. I decided to pick stations that had records throughout the widest range of dates (from 1975-2015). These stations 15 stations (out of 41) are: C10, C3, D10, D12, D19, D22, D26, D28A, D4, D41, D6, D7, D8, MD10, and P8. This narrowed down the number of stations to a reasonable number (since I was often fitting a model per station) while still representing a wide spatial range. For a few of these stations, there was sufficient missingness in some of the variables of interest that I just decided to remove the variable from that particular model for the station. For full spatial models, I only used variables that were available across all of the stations.

1.3. **Parsimonious Model.** The parsimonious model used temporal components and variables that visually looked correlated with chlorophyll overall. The day of year was included to account for any seasonality and was fit with a cyclic spline so that the ends match up from year to year. The date was included to account for any trends over time. The non-time covariates that looked correlated were pheo (pheophytin a), tn (total nitrogen), and do.per (dissolved oxygen per ?). Note that do.per was more highly correlated than do (dissolved oxygen), but I'm not sure what the "per" means. If in context, the "per" variable doesn't make sense, it can be replaced with ordinary dissolved oxygen. All but day of year were fit with thin plate regression splines. These splines have some optimality properties, are low rank (needing to fit many fewer coefficients), and isotropic (rotation of the covariate coordinate system will not change the smoothing). The same model form was fit for each station. The maximum basis dimensions per variable change per station to allow for extra flexibility if need be and are subject to the constraints of the data at each station.

Overall, pheophytin a is the non-temporal predictor variable that contributes the most variability in the fitted values of chlorophyl. Dissolved oxygen (per) is more erratic; for some stations it has similar variability in its contribution to the overall fitted value as pheophytin, and for others it matches less well. The day of year and date variables yield different contributions per station which we want to keep in mind when we are fitting full spatial models. For some stations the variability in the contribution of the temporal variables wanes as time goes on, and the matching of the contributions to the true values is better in some time periods than others (better early on).

1.4. **Full Model.** The full model included everything in the parsimonious model and then added some extra variables that are interesting in context but did not look particularly predictive of chlorophyll in pairwise scatterplots. The additional variables that were added are: sio2 (silicon), tp (total phosphorus), tss (total suspended solids), nh4 (ammonium), sal (salinity). Again, all of these were fit

with thin plate regression splines, the same model form was fit for each station, and the maximum basis dimensions per variable change per station. Salinity, ammonium, and total nitrogen are the variables that have the most missingness and therefore don't occur in every station's model.

Many of the added variables in the full model have flat contributions (or very small magnitude of variability) to the fitted values including total phosphorus, total suspended solids, ammonium, and salinity. The parsimonious model fitted values often visually match better than those of the full model, especially in magnitude for high peaks. This suggests that adding a bunch of other variables doesn't necessarily give us that much of a practical benefit, even though the RMSEs are often marginally better for the full model merely from having more access to additional information to tune predictions.

1.5. **Spatial Models.** The above models were fit per station which did not allow for a true spatio-temporal analysis since each station could not leverage information from the other stations. I tried a few different set-ups for the spatial models in increasing level of complexity.

(1) Spatial Model 1: Day of year and date with station as a factor. Same day of year and date trend, station just acts as an adjustment on the intercept.
(2) Spatial Model 2: Day of year and date with station as a factor. Date is allowed to differ by station.
(3) Spatial Model 3: Day of year and date with station as a factor. Date and day of year are both allowed to differ by station.
(4) Spatial Model 4: Day of year and date with station as a factor. Date is allowed to differ by station. The interaction between date and day of year is allowed to differ by station. Note that I didn't let day of year differ by station here to scale back the complexity.

I went back and did Spatial Model 3 [which turned out to be the highest performing] but adding one covariate (pheophytin a, then total nitrogen separately) to see if we could get a boost from a highly correlated covariate from the parsimonious model that we hadn't allowed in a spatial model yet.

It looks like there is some benefit of having a per station date term for some stations, and for others it doesn't really help that much. The middle stations benefit from a date term by station. Those on the extremes that are more isolated (left, top, bottom) do not. The day of year term by station doesn't seem to help any of the stations, and similarly for the interaction term, it is just overkill.

The intercepts for each station are very similar for spatial models 1, 2, and 4 while model 3 has larger intercepts in general. The intercepts do tend to be similar at nearby stations which suggests that there is some similarity based on proximity that is picked up on without explicitly including spatial coordinates in the models. The deviance explained goes down as the models get more complicated, and most of the stations have the lowest RMSE for Spatial Model 3 which allows for a

different day of year and date pattern by station. This makes sense since this allows for the most flexibility.

Looking at the RMSE per station, overall we do better on left most stations (and the north most station), and do worse on stations to the right (including the southern most station). There is the same relative ordering of the RMSE by station across Spatial Models 1 and 2 and the interaction and parsimonious models, meaning that the "hard" stations remain consistent across all of these models. However, the stations on the right do less poorly in the parsimonious model, suggesting that the extra covariates are helping there.

Now that we are incorporating more data and having many smooths being able to vary across each of 15 stations we have increasing complexity that carries over into the computational costs of fitting these models.

1.5.1. *Computational Considerations.* I experimented with a few different solutions to the computational problems that came with fitting the spatial models. This really comes down to patience when working on a personal computer; nothing here is extremely computationally intensive, but for my work flow I wanted to have reasonable wait times of a few minutes per model.

I refreshed my memory on how to take advantage of the Statistical Computing Facility's clusters and figured out how to move files back and forth and submit and monitor progress of various jobs. This is a good solution for fitting models that you are pretty sure of the size of the basis dimensions, but you don't want to wait around for the model to fit. For example, when I wanted to expand Spatial Model 3's basis dimensions to make sure I was making a fair comparison to another model later on in my process, I put it up on the cluster, and came back to get the results after a few hours.

However, in the trial and error phase of choosing basis dimensions, it is still kind of a pain to keep moving files back and forth to check diagnostic plots of various models. For this, I used mainly the "bam" function in the mgcv package. This is a multiple thread version of the "gam" function. There was a huge improvement in speed with only using 4 of the maximum 8 threads on my laptop.

I also recently learned how to switch my linear algebra package in R to be the vecLib version of BLAS which can give up to a linear time improvement.

Once I learned to take advantage of all of these options, the computational complexity of the spatial model fitting was not a major obstacle to progress.

1.6. **Flow Models.** I started this project with a side motivation of thinking about how to incorporate flow information on data that is set up on a network. The network here is a spatial one, the set-up of the stations as compared to the flow of water in the Bay Delta area.

In the spatial models, the station is just a factor. There is no explicit information about which stations are close to one another. However, just considering

the Euclidean distance between two stations may not be the most indicative version of distance if the two stations are not flow connected. If we could somehow incorporate information about the flow network structure into the GAM, we may be able to leverage more spatial information than if we merely used longitude and latitude to determine "closeness".

I had two approaches for trying to incorporate flow that approach the problem from different angles. Using the chlorophyll series of a neighboring station (correlation-wise) as a covariate, incorporates some information about the network that doesn't explicitly use Euclidean distance between stations. A next step would be to do the same thing only using the station that is nearest in a Euclidean sense to see if there are any major differences.

I also used flow data from `http://www.water.ca.gov/dayflow/` as a covariate. Since I used this in a by station approach, this doesn't actually incorporate information about the spatial network, but it does help us think about flow in general.

1.6.1. *Neighboring (Correlation-Wise) Station Chlorophyll.* My first approach was to compare the chlorophyll series between each station and see which station's chlorophyll are most correlated.

Checking correlations in a preliminary exploration, there were several pairs that had high correlation in their chlorophyll values.

Absolute value of >0.8 Correlation:

- D7 to D8, D10
- D10 to D4
- D12 to D19
- D19 to D28A
- D26 to D28A

Since not every station was measured on the same days, merging the station series by day to get a true correlation left many missing value pairs between stations. Every station had about 600 measurements, so I decided that any correlation made with less than 100 pairwise comparisons that weren't NA were not to be used because they could be unduly high due to small numbers of comparisons. Then each station's "best match" was merged on a "closest date" criteria to be a covariate for a by station model. This means that even if the chlorophyll measurements were taken a day apart, they would be merged. This was to help with missingness issue, although I admit that the treatment of missigness in this whole workflow is not necessarily the best way of doing things. I also tried up to a lag of 5 to see if a lag would improve the correlation between different stations, but the best lag was always zero.

The RMSEs for the chlorophyll "flow" models are less than the RMSEs for both versions of "true" flow models described below. This correlation network structure rather than a mere spatial network may be a useful thing to pursue further.

1.6.2. *True Flow.* My second approach was inspired by Marcus Beck's work on adjusting for flow in his Weighted Regression on Time, Discharge, and Season models. The use of flow as a covariate can help account for the variance in the response that is attributed to flow instead of allowing it to be potentially confounded with a temporal trend. To make this approach more spatial, I should go back and fit a spatial model with flow as a covariate allowing as many of the temporal components to vary by station as possible. Then at least the flows from each station can be leveraged. A variant on the flow approach that I did try was to look at the volumetric flow data, determine which source made the largest contribution on average to each station, and then using flow data matching the largest contribution source from the water flow data. This was not always possible since for example, there was no obvious matching water flow data for the Carquinez Straight at Martinez even though a few stations had this as their highest volumetric component. To better understand flow over time, I made plots that show the composition of flow at each station aggregated by year to see changes over time, and by month to see changes over seasons. These can be found in the Shiny app.

For the flow models, I also built simple day of year, date, and interaction between day of year and date models per station to have a better baseline for comparison. The RMSEs for interaction models are smaller than for total flow model for 7/15 stations. The RMSEs for interaction models are smaller than for station specific flow model for 7/15 stations. The RMSEs for total flow are less than the RMSE for flow specific for 7/15 stations. However, this could be partially confounded with how much the interaction term helps at each station. I should either add the interaction term to the flow models or take out the interaction term from the baseline models to check this.

1.7. **Overall Comparisons (RMSE per station/model combinations). By Station**

- C10: bad in general, need the extra covariates in full model to help
- C3: full and parsimonious models are best overall but the interaction and chlorophyll correlation models are on par
- D10: interaction and chlorophyll correlation models are best overall
- D12: interaction model best overall, on par with parsimonious, full, and chlorophyll correlation models
- D19: full model is best overall, on par with parsimonious model; interaction and chlorophyll correlation models are about the same order of magnitude
- D22: chlorophyll correlation model does best, on par with interaction and parsimonious models
- D26: parsimonious model does best, full model gets bad in certain categories, interaction and chlorophyll correlation models are about the same order of magnitude

- D28A: chlorophyll correlation model is best, parsimonious, full, and interaction models are next up and about the same order of magnitude
- D4: full model does best, parsimonious and interaction methods on par, chlorophyll correlation model not far behind
- D6: full model does best, parsimonious and interaction models on par, chlorophyl correlation model not far behind
- D7: full model does best, parsimonious, interaction, and chlorophyll correlation models are next up
- D8: full model does best, parsimonious model next interaction and chlorophyll correlation models a little bit behind
- MD10: full does best, parsimonious and interaction models next, chlorophyll correlation model decently behind
- P8: chlorophyll correlation does best, parsimonious and full models next

Overall it looks like chlorophyll correlation as another predictor variable is valuable. Trying a by station model with an interaction term, chlorophyll from a correlated station, and the effective predictors from the parsimonious model might be the best combination.

By Model (numbers out of 15)

- parsimonious model: annual 4 and 5 most often minimum RMSE (7, 4)
- full model: season 4, annual 5 and 6 most often minimum RMSE (4,3,3)
- interaction (baseline) model: annual 6, annual 4, 5, seasonal 1 most often minimum RMSE (5, 3,3,3)
- spatial model 1: annual 4, annual 5, seasonal 1 most often minimum RMSE (4, 3,3)
- spatial model 2: annual 5 and seasonal 4 most often minimum RMSE (6, 3)
- spatial model 3: annual 5, annual 4 and 6 most often minimum RMSE (5, 3,3)
- spatial model 4: annual 5 most often minimum RMSE (6)
- spatial model 3 with pheo: annual 5, 6 most often minimum RMSE (6, 3)
- spatial model 3 with total nitrogen: annual 5, 6, 3 most often minimum RMSE (6, 3,3)
- chlorophyll correlation: annual 3, 5, 6 most often minimum RMSE (4, 3,3)
- total flow : annual 5, 6, 4 most often minimum RMSE (6, 5, 4)
    Breakdown Key
- annual 1: [1975, 1982)
- annual 2: [1982, 1989)
- annual 3: [1989, 1996)
- annual 4: [1996, 2003)
- annual 5: [2003, 2010)
- annual 6: [2010, 2015] (fewer years than the others)
- seasonal 1: Jan, Feb, Mar

- seasonal 2: Apr, May, Jun
- seasonal 3: Jul, Aug, Sept
- seasonal 4: Oct, Nov, Dec
- after mid-90s better RMSE, we see this confirmed in the observations made from the shiny app

NOTE: More detailed results per model per station can be found in perStation-ModelObervations.tex and compareModelPerformanceByStation.tex

1.8. **Description of Shiny Functionality. Choices**

- station: C10, C3, D10, D12, D19, D22, D26, D28A, D4, D41, D6, D7, D8, MD10, and P8
- date range (1975-2015)
- upper limit for Y on fitted value plots (default calculated from data) [allows you to zoom]
- lower limit for Y on nested plots (default calculated from data) [allows you to zoom]
- upper limit for Y on nested plots (default calculated from data) [allows you to zoom]
- select up to two variables to display contributions towards fitted values for parsimonious model [declutters nested plots]
- select up to two variables to display contributions towards fitted values for full model [declutters nested plots]
- select up to two variables to display contributions towards fitted values for interaction model [declutters nested plots]
- select a spatial model (spatIntercept, spatDate.Dec, spatDOY, spatInteraction)
- select a model to look at RMSE breakdown (predPars, predFull, predInt, predSpat1, predSpat2, predSpat3, predSpat4, predSpat3Pheo, predSpat3Tn, chlPred, flowPred)
- remove C10 as outlier for RMSE breakdown plots (yes/no)

Plots

- location of chosen station
- fitted values parsimonious model
- fitted values full model
- fitted values interaction model
- fitted values chosen spatial model
- fitted values from chl from other station model
- fitted values from flow model
- component-wise predictions parsimonious model [log scale]
- component-wise predictions full model [log scale]
- component-wise predictions interaction model [log scale]
- component-wise predictions chosen spatial model [log scale]

- component-wise predictions chosen chl flow model [log scale]
- component-wise predictions chosen total flow model [log scale]
- component-wise predictions parsimonious model choose 2 variables [log scale]
- component-wise predictions full model choose 2 variables [log scale]
- component-wise predictions interaction model choose 2 variables [log scale]
- volumetric flow components (aggregated by year, aggregated by month)
- RMSE breakdown for chosen model by station
- commented out (app getting too unwieldly): show plots for all of spatial models at once for easier comparison

1.9. **In Progress.** I am in the process of trying to use the method described in O' Donnel, Rushworth, Bowman, and Scott's "Flexible regression models over river networks" in the Journal of the Royal Statistical Society, Applied Statistics, 2014.

The main idea here is to use penalized B-splines that respect the spatial network of the flow. Because it boils down to a spline approach, I thought it might be a nice comparison. However, the current obstacle is to get the data into the format expected by the smnet package that implements this approach. The data goes through extensive GIS preprocessing before being in the right format. I am trying to circumvent this process, drawing out the network by hand and trying to massage everything into the correct format so that we plug into the heart of main functionality of the smnet package. I am going through the source code, and making some progress, but it is not ready yet.

2. Generalized Additive Model (GAM) and Weighted Regression on Time, Discharge, and Season (WRTDS) Comparison

2.1. **Goals and Approaches.** The goal of this portion of the project is to mirror the comparison of GAMs and WRTDS done by Marcus Beck and Rachel Murphy (`https://github.com/fawda123/patux_manu/blob/jawra/patux_manu.pdf`) but provide the same framework in a new environment (Bay Delta). I have mainly gone through Marcus's work here (`https://github.com/fawda123/sf_trends`) and made the GAM models using the data that he pre-processed. I then built the comparison framework following the manuscript and a Shiny application mirroring Marcus's (`https://beckmw.shinyapps.io/sf_trends/`). Throughout, I learned more about the inner workings of the WRTDStidal package and the context of the data from Marcus.

The context of the project is to understand the variability in dissolved inorganic nitrogen, ammonium, and nitrate at many of the same stations considered in the chlorophyll part of the project. Flow is represented by flow records from the San Joaquin River, Sacramento River, or salinity as a tracer of flow depending on the station.

2.2. **Building the GAM Models.** To build these GAM models, I took a less individualistic approach per station because I didn't want the comparison of GAM and WRTDS to be unfairly impacted by fine tuning per station. The response variable was on the log scale to match what was used in the WRTDS models. Comparisons are made on the log scale to avoid any biases with backtransforming.

I started with the default parameters for the predictor variables day of year, date, and flow for each response, allowing for all possible 2 way interactions as well as a full 3-way interaction. The day of year portions were fit with cyclic splines to ensure they matched up at the ends. All others were fit with thin plate regression splines. I also fit the same model but without flow to see how much flow was helping in explaining the variability in each response.

From checking the diagnostics of these, I got a sense for how much the maximum basis dimensions needed to be expanded. I picked reasonably large basis dimensions per parameter and then fit a consistent model to each station/response variable pair. A few of the models yielded warnings, and then I adjusted the basis dimensions a bit to get the warnings to go away, but keeping the adjustments as consistent with one another as possible. I also fit the models in a nested way so that we could isolate the impact of each interaction above and beyond the individual components.

2.3. **Analyzing Flow.** Flow normalization attempts to correct for differences in flow that may obscure the overall trend in the response. To flow normalize, we take the averages of the predictions at a point given the expected flow that occurs in the same month across all years. For example, for a value at January 1990 and a value at January 2000, all of the January observations would go into creating the expected flow for both, but the spline will be different even given the same flow input, because it occurs at two different times.

We can also assess how the relationship between the response and flow by month (seasonal) changes over different years through plots. These are shown in the Shiny app using Marcus's dynagam function.

2.4. **Metrics of Comparison.** The metrics of comparison are taken from Beck and Murphy's manuscript. They all are calculated on the log scale. All of the metrics of comparison are calculated overall, in annual blocks, seasonal blocks, and flow levels.

Breakdown

- annual: [1976, 1983), [1983, 1990), [1990, 1997), [1997, 2004), [2004, 2015] {has 2 extra years}
- seasonal: (Jan, Feb, Mar), (Apr, May, Jun), (Jul, Aug, Sept), (Oct, Nov, Dev)
- flow: quartiles of the flow for that station

Metrics

- root-mean-squared-error (RMSE): $\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}-y)^2}{n}}$
- deviance: sum of squared residuals
- average difference: $\left(\frac{\sum_{i=1}^{n}\hat{y}_{WRTDS,i}-\sum_{i=1}^{n}\hat{y}_{GAM,i}}{\sum_{i=1}^{n}\hat{y}_{GAM,i}}\right)*100$
- percent change: annual mean estimates for the first and last three years of flow-normalized estimates, excluding the annual aggregations that had limited annual mean data (i.e., seven years per period) double check, still might have done this wrong
- RMSE between: $\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_{WRTDS,i}-\hat{y}_{GAM,i})^2}{n}}$
- regression of wrtdsPred gamPred: is intercept significant? (>0 WRTDS estimates were on average larger than GAMs) is slope significant? (<1 GAMs fit a wider range of values compared to WRTDS)
- visually compare fitted values on Shiny app

Both methods have predictions that are at least visually very close to one another. The RMSE and deviance for GAM fitted values are less than that of WRTDS. In an experiment, D7 and D19 (in the center of the station network) were chosen to be fairly representative of the other stations. I expanded the maximum basis dimensions for the GAM models as far as I could under identifiability constraints, and visually the prediction look similar to those of the GAM model with the original choices basis dimension maximums. For the WRTDS models, the original window sizes were (0.5, 10, 0.5). I tried smaller window sizes (0.25, 5, 0.25), bigger window sizes (0.75, 15, 0.75) and an even bigger widows size (1,20,1). Still overall the expanded GAM does better than all of the WRTDS models. Of the WRTDS models, the smaller window model generally does better, with a few breakdown categories having the original WRTDS model being the best. From the Shiny app, it looks like WRTDS predictions are less extreme than the GAM predictions. So if GAMs are always "winning" it means these models get predicted values high/low enough in the right places, and the models are not losing big by predicting values large in magnitude when they shouldn't be.

The overall intercept is often significantly negative meaning GAM estimates were on average larger than those of WRTDS, and the slope is often significantly less than one, meaning that GAMs fit a wider range of values compared to WRTDS. These results are consistent with what we see when examining the comparison plots on the Shiny app.

NOTE: For more detailed results, see compareModels.tex.

2.5. Description of Shiny Functionality. Choices

- station (C10, C3, P8, D19, D26, D28, D4, D6, D7)
- response variable (din, nh, no23)
- plot type (observed, annual [avg])
- scale type (linear, natural log)
- date range (1975-2013)

Plots
- fitted values for GAM with and without flow
- fitted values for flow normalized GAM with and without flow
- with flow model contributions of each component on the fitted values
- no flow model contributions of each component on the fitted values
- GAM with flow and WRTDS fitted value comparison
- dynagam plots (changes in relationship between response and flow across time) for GAM with flow
- RMSE breakdown spatial plots

2.6. **In Progress.** I am currently investigating what is going on at station P8. In context, P8 is downstream of a wastewater treatment plant. Changes happened in the mid-2000s to reduce ammonium outputs from the treatment plant. Water quality changes in ammonium and to a lesser extent dissolved inorganic nitrogen have been shown in other work.

The biggest differences seen in the metrics of comparison between GAMs and WRTDS happens for P8, so I am looking into where the methods differ, and if these differences correspond to the times where we know something new is happening. Some of the differences in the metrics are huge, so first I want to make sure this isn't a missingness issue or some other artifact.

There is also a major discrepancy in the annual aggregated GAM with flow v. GAM without flow. The without flow model does much better, which is not consistent with the rest of the results. I refit the models to give higher maximum basis dimensions in case the model fit was off. This didn't help. It looks like there is some asymmetry around $y = 0$ in the residuals v. predicted values plot, but there are so few points where this is true, and on the annual aggregation level, this shouldn't be a big issue. I took those points out and refit the model, but this still didn't help. For dissolved inorganic nitrogen and nitrate, this station has an increased mean, median, and spread than all of the other stations for that response, but C10 is also up there and seems to be fine on the Shiny app. Ammonium is still a mystery; the flow model consistently predicts larger values than the no flow predictions, but both the distributions of flow and the response are consistent with those of other stations.

The annual, flow-normalized predictions for GAMs at P8 look weird compared to the WRTDS version, so I need to make sure the flow-normalization is not off or see why the model is behaving so poorly.