1. Flow v. No Flow GAMs Fitted Values

- din: the different models have visually very similar predicted values for all stations
- nh: the different models have visually very similar predicted values for all stations except for C3 (no consistent difference between the two models' fitted values, one is not predicting more extreme values in general, the phase of the fitted values differ in places)
- no23: the different models have visually very similar predicted values for all stations except for C10, P8, D4, D6, D7 (magnitudes differ at peaks, less noticeable differences for this variable)
- station/response variable pairs that don't have similar fitted values between the two models are somewhat "irregular" series in that they have less reliable cyclical structure and/or the change in magnitude of cycles is not over time/predictable

2. Decomposition: Terms That Could Be Dropped

Each model includes all possible interactions between doy, date.dec and flow including the three way interaction. I suspect that some of these may be overkill. We can look at the nested components of the predicted values and see if they seem to have enough variability to be worth including. Note we do not need to do this to save on computation, each model is quick to fit, but it can help us interpret whether relationships (such as the doy trend or the flow trend) change over time or with one another.

- This is really station dependent. In general the three way interaction is probably overkill.
- Varying behavior in whether or not the two way interactions are important between flow and the time variables.
- C10: flo interactions could be dropped especially the 3 way interaction, flo has variability itself, seems to dampen the effect of doy, date.dec and their interaction, but this is not the norm if we look across the other stations.
- In general it seems that when date.dec and doy have a big impact on the fitted values, the interaction between them and flow is still important.
- In the cases where flow v. no flow fitted values look different, the impact of the time components seems to be dampened by flow.

3. Flow Normalization

I think I finally have this right.

(1) Take all possible combinations of [month, year, flow(in month)]
(2) Predict using GAM
(3) average to get a value per month/year

A spot check comparing my plots to the third plot on Marcus's app, the normalized values look reasonable. I also specifically checked all of the responses for C10 and P8 since they are used as examples in the draft.

For D6 din/no23 there still seems to be waves/oscillations in my flow normalized series. You can sort of see this in C10 and P8 if you stare at it. Could part of this be because I'm comparing the median of Marcus's work (somehow more robust to this kind of thing?) and the mean from mine.

I also think I figured out why my annual aggregated version looks less smooth than Marcus's. He uses annual.agg() on the normalized values, and I normalize then use annual.agg(). *Fix and make sure this is why.*

4. Dynagam Plots (C10 and P8)

Looking at my dynagam plots and finding mostly the same results as Marcus outlines in his draft.

Results: C10 din

- din: high flow, low concentration (negative association in dynaplots)
- changes in 2000s: old → new, red → blue, early 2000s yellow, late 2000s green, top to bottom goes red to blue, but yellow trumps red, and we see a boost in green in the summer and fall
- *highest in January, higher values at moderate to low flow rates depending on year* I see less peakedness in the curves except for the summer months, highest in February, March, January
- downward trend not so apparent in nh, in early years yes, but then flattens out, most variability in Dec, Jan, Feb.
- downward trend apparent in no23

Results P8

- seasonal variation more apparent in ammonium, big drop in variability in the summer
- highest concentrations in winter for both ammonium and no23
- pretty flat for May through August for nh, a little bit more going on for no23
- no23 flow/nutrient relationship strongest late in time series (blue-green curves above warm color curves)
- ammonium flow/nutrient relationship strongest in early 2000s, highest curves are green

5. Chl Data

In Marcus's draft he includes results on chl. In the original data that I was working with, chl is not included. I can easily build the same kind of models for chl and add to the Shiny app, but I want to use the same data to make it comparable.