

# make prediction plots

David Nguyen

2022-12-28

## Prediction grid

First need to choose what values to predict mean log median of medians nitrate level. We want to make sure we aren't making prediction at DOC/P values that are very sparse. We will use a grid of values between 3 - 97% percentiles of DOC while P is held at the median, and vice versa. Note, it would be better to use more values than just the median, but it is hard to show in a plot.

Using these percentiles were determined subjectively by choosing what seemed to be a good trade-off between predicting over a large interval of values while keeping to concentrations that are frequently observed. See `joint_dist_cp.pdf` file in google drive for additional details.

The quantiles will be determined for each subset, e.g., lakes in north america, not quantiles for the entire data set. This means that model predictions for different subsets of the data will not overlap the same values of DOC/P.

## plots

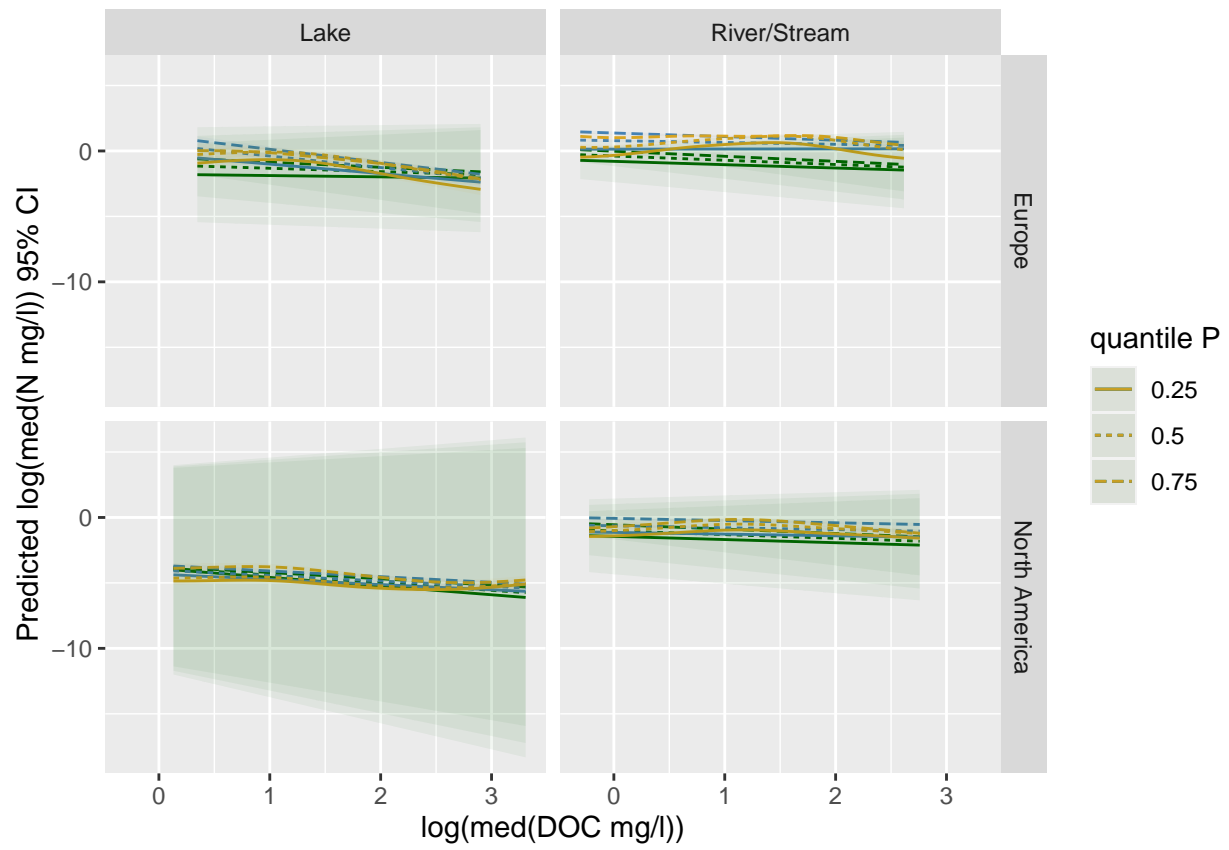
Color legend: SVC = darkgreen, OLS = steelblue, GAM = goldenrod

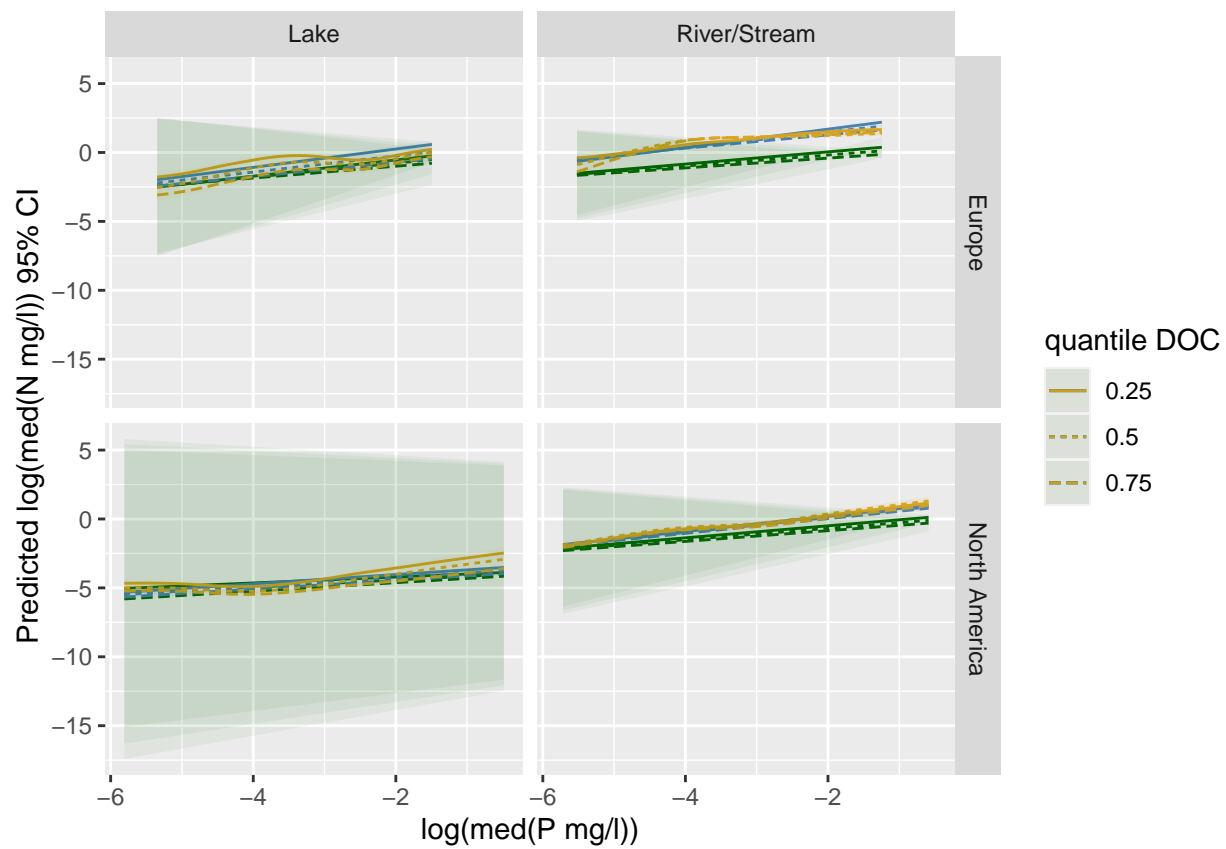
When DOC is varied, P is held at median and vice-versa.

Interpretion: remember, we did log-log regression, so the slope of the curve is what corresponds to the exponential decay parameter  $k$  in Taylor and Townsend. So, if the slope of the curve is negative, there is a exponential decaying relationship and if the slope of the curve is positive, there is an exponential increasing relationship.

Notice that for the GAM, we can see that some values of DOC/P concentration are positive whereas others are negative, Scientifically, this suggests that the concentration of a chemical can affect whether it is associated with higher or lower nitrate.

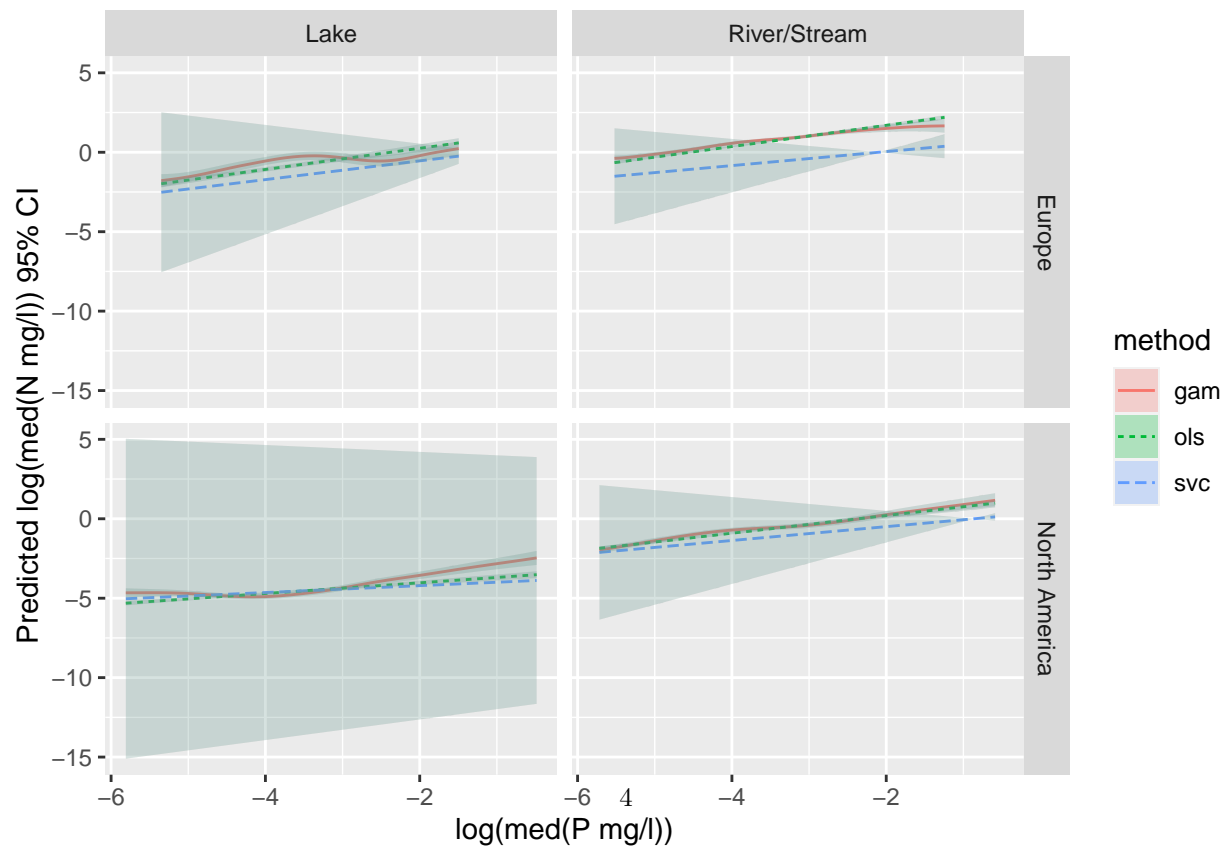
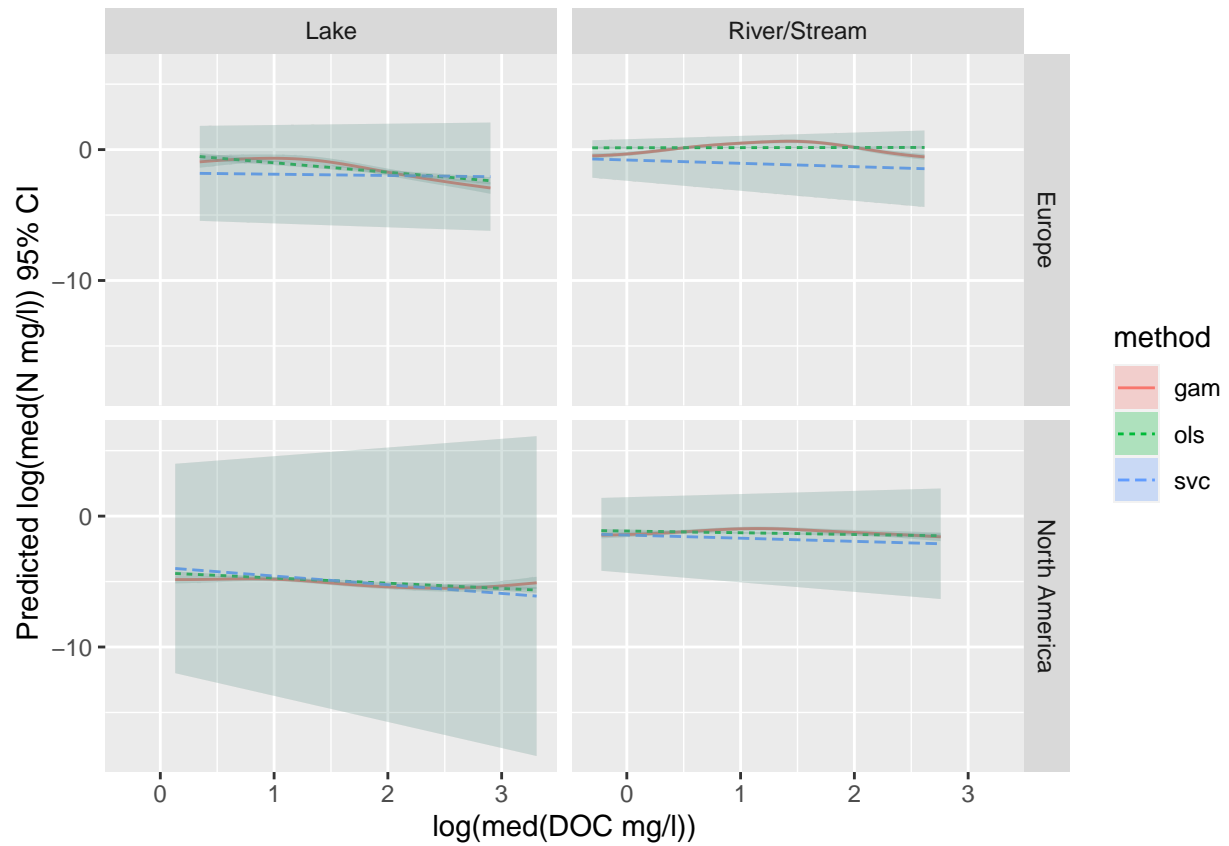
## All quantiles



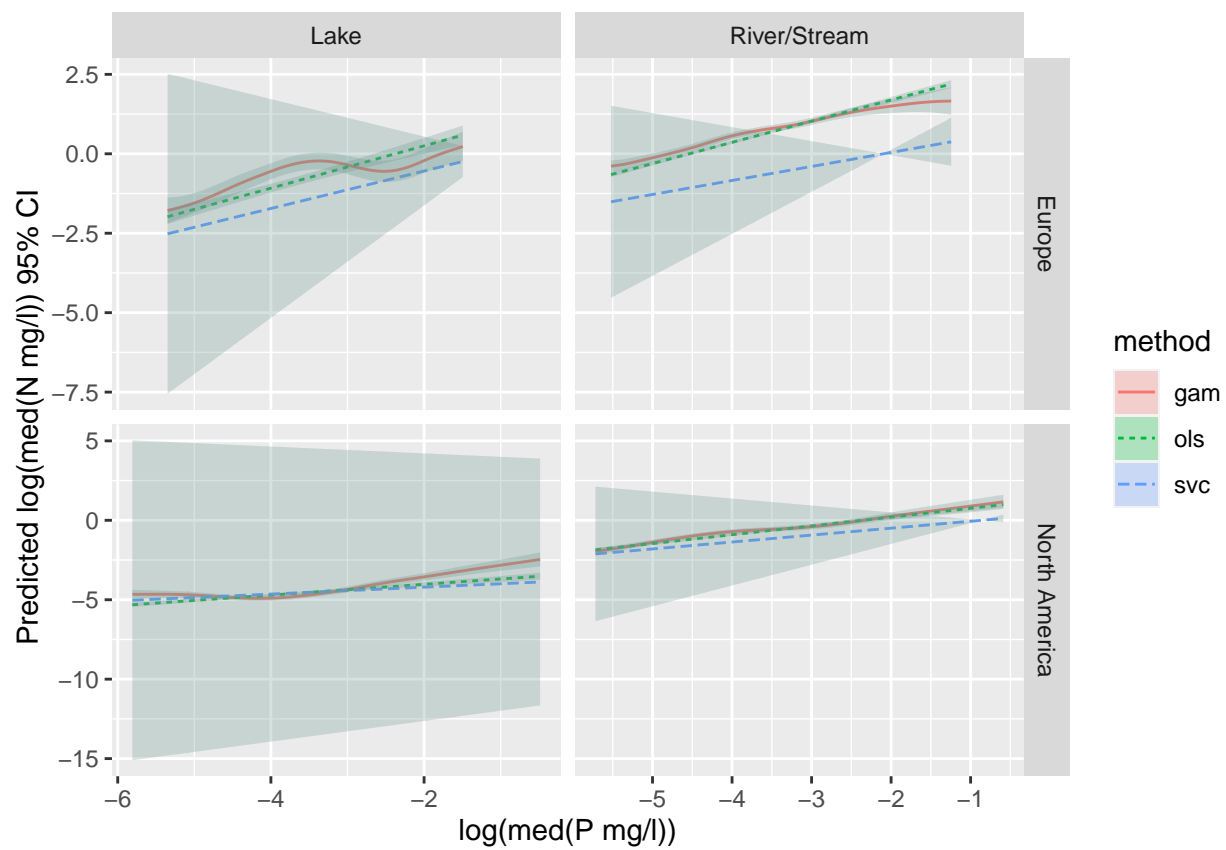
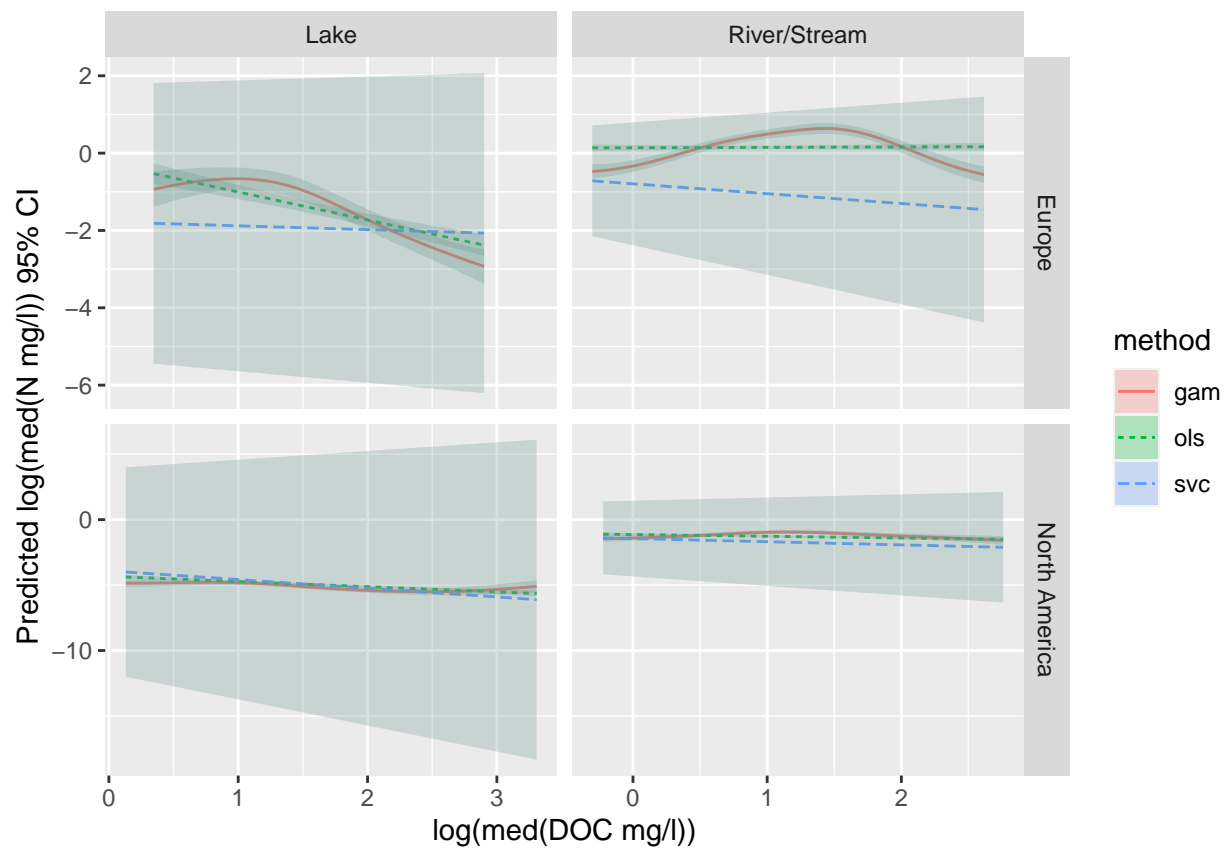


## 25th quantile

All axis on same scale

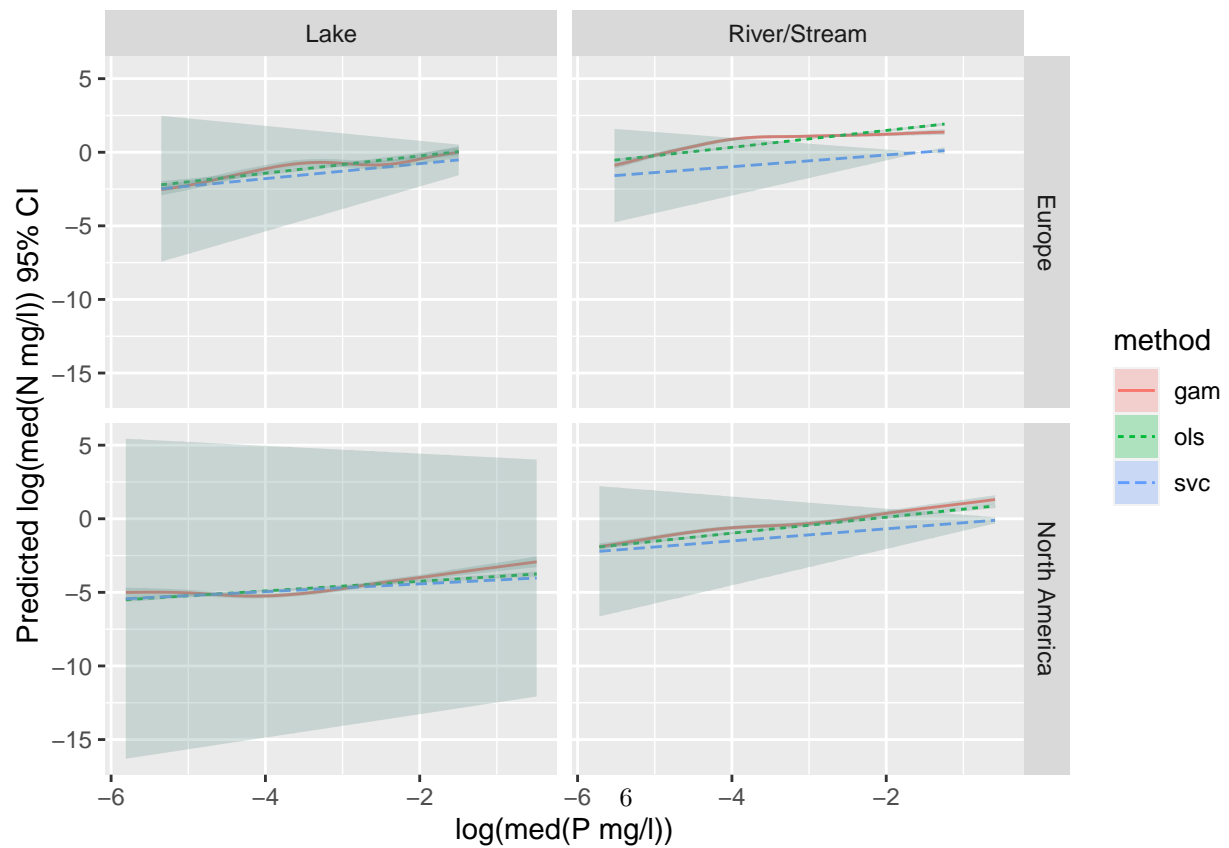
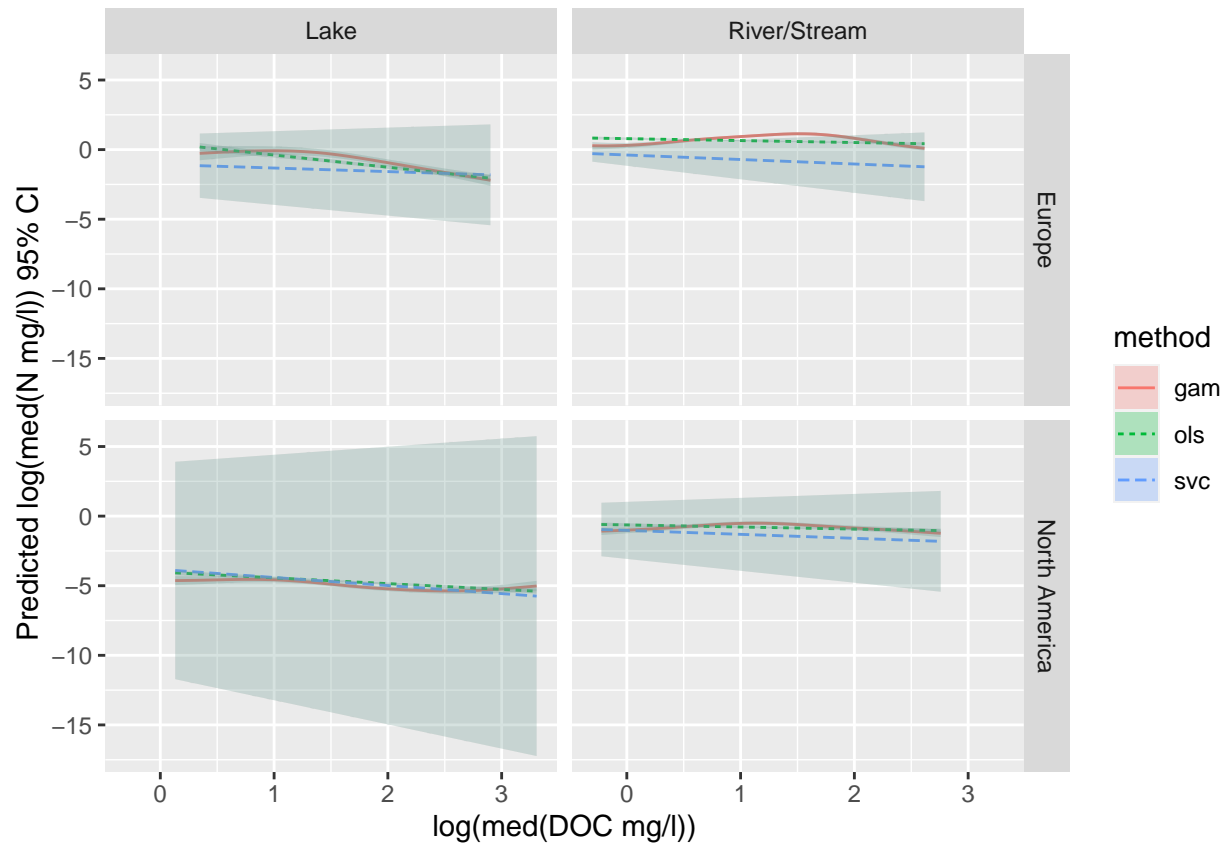


Axis on different scales

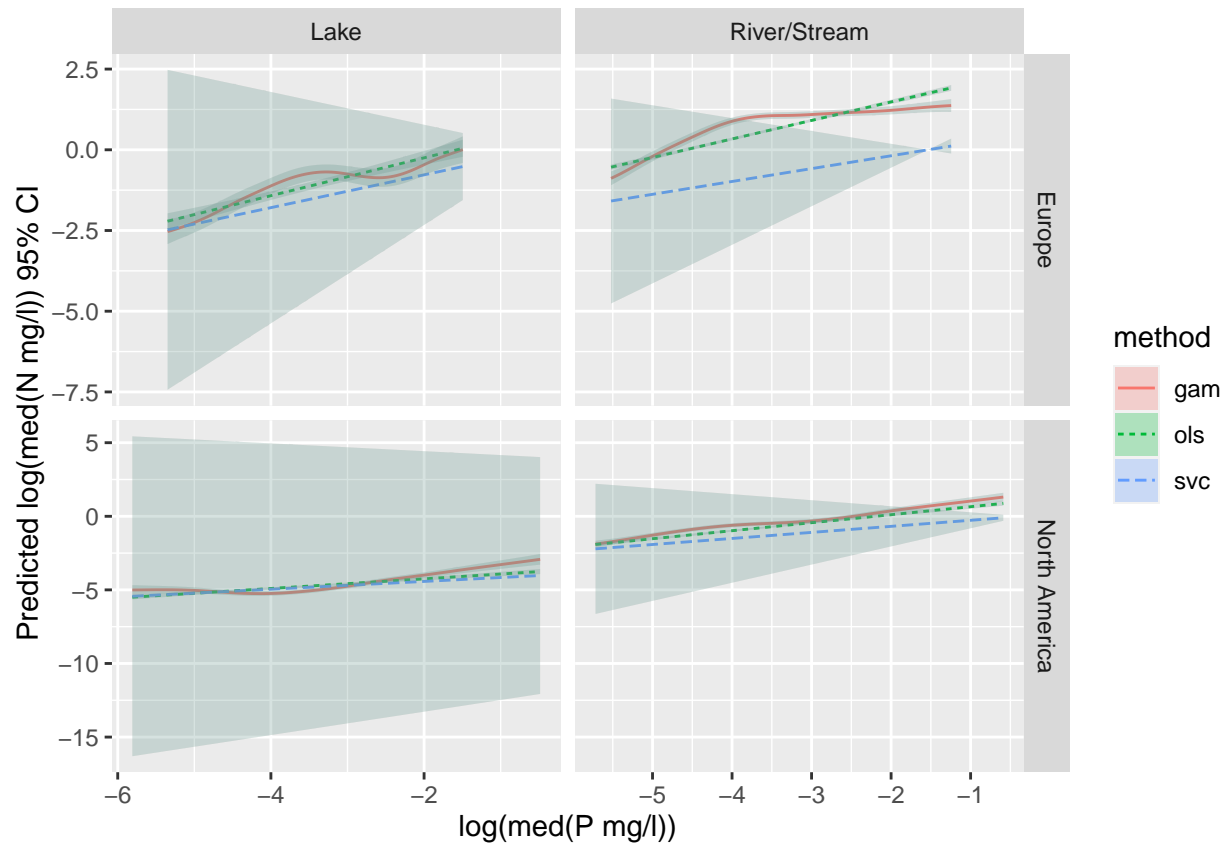
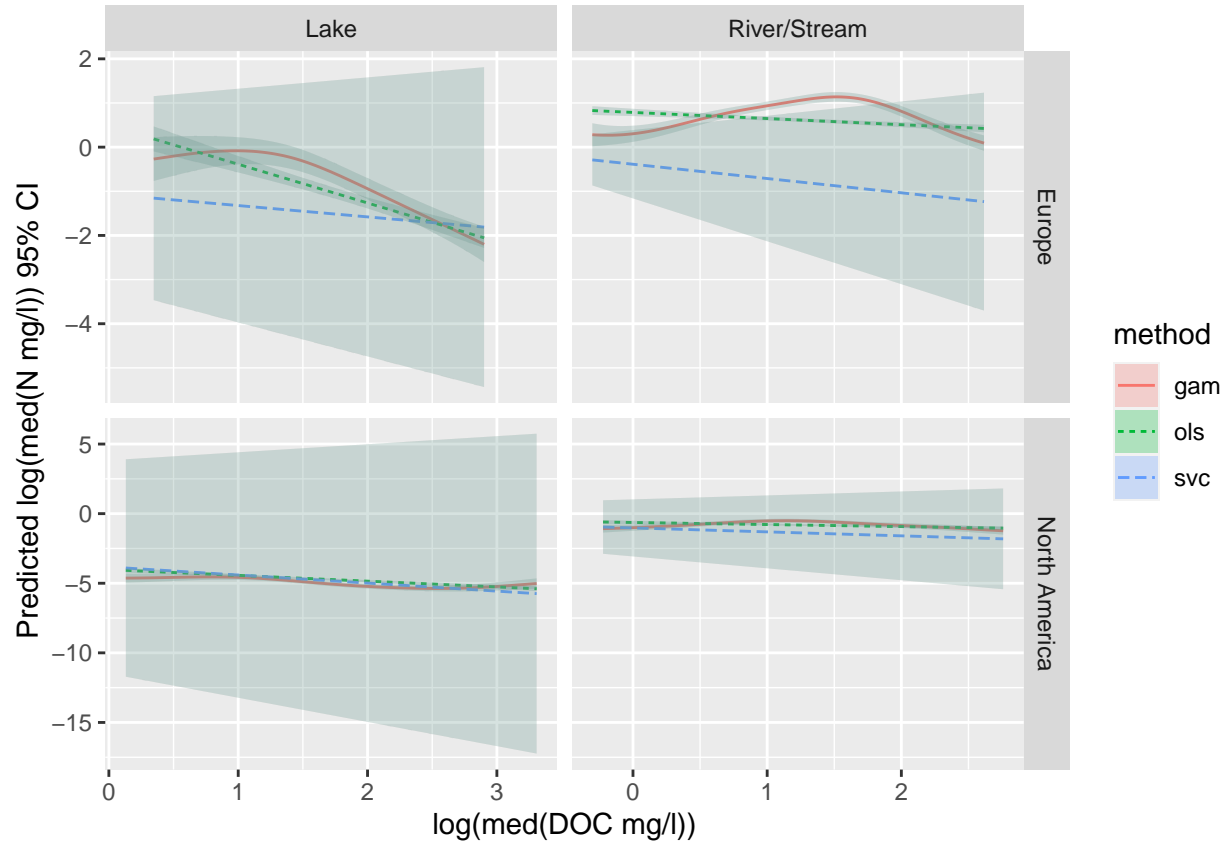


## 50th quantile

All axis on same scale



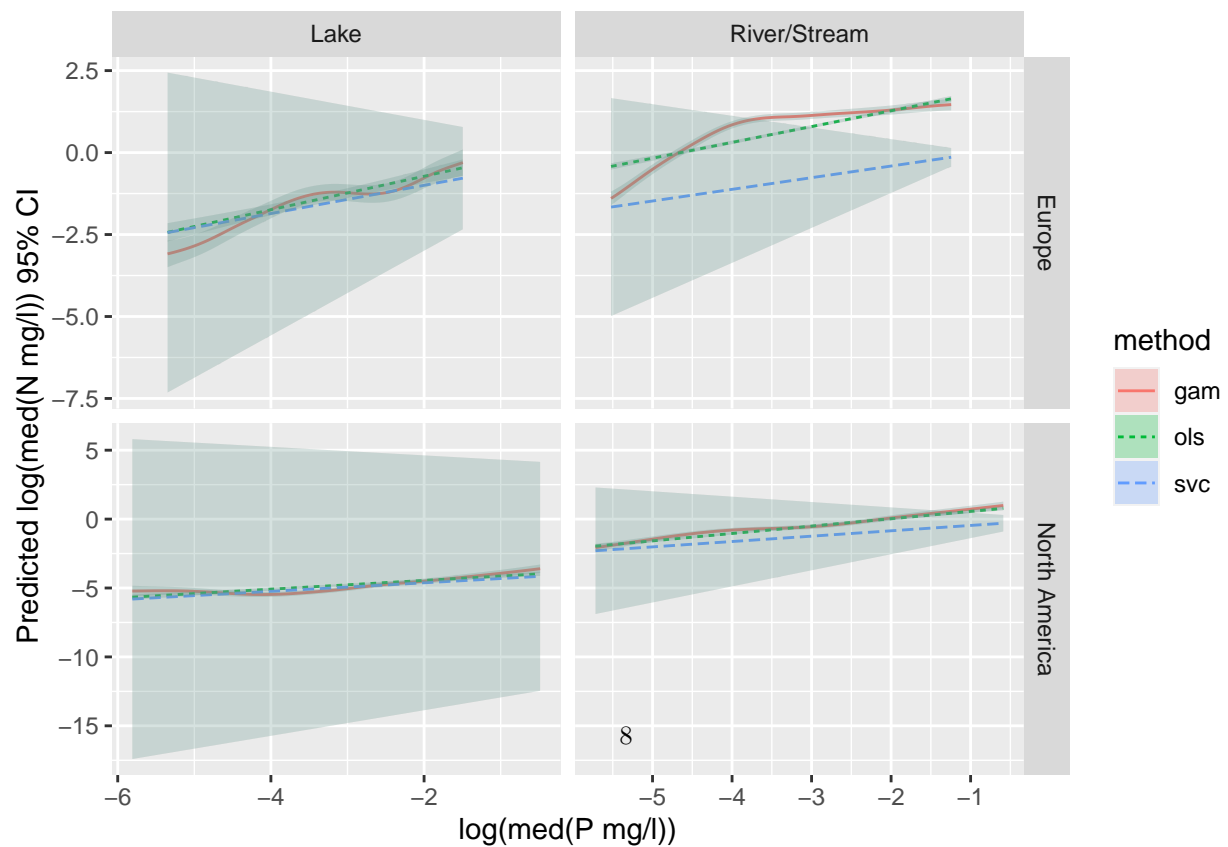
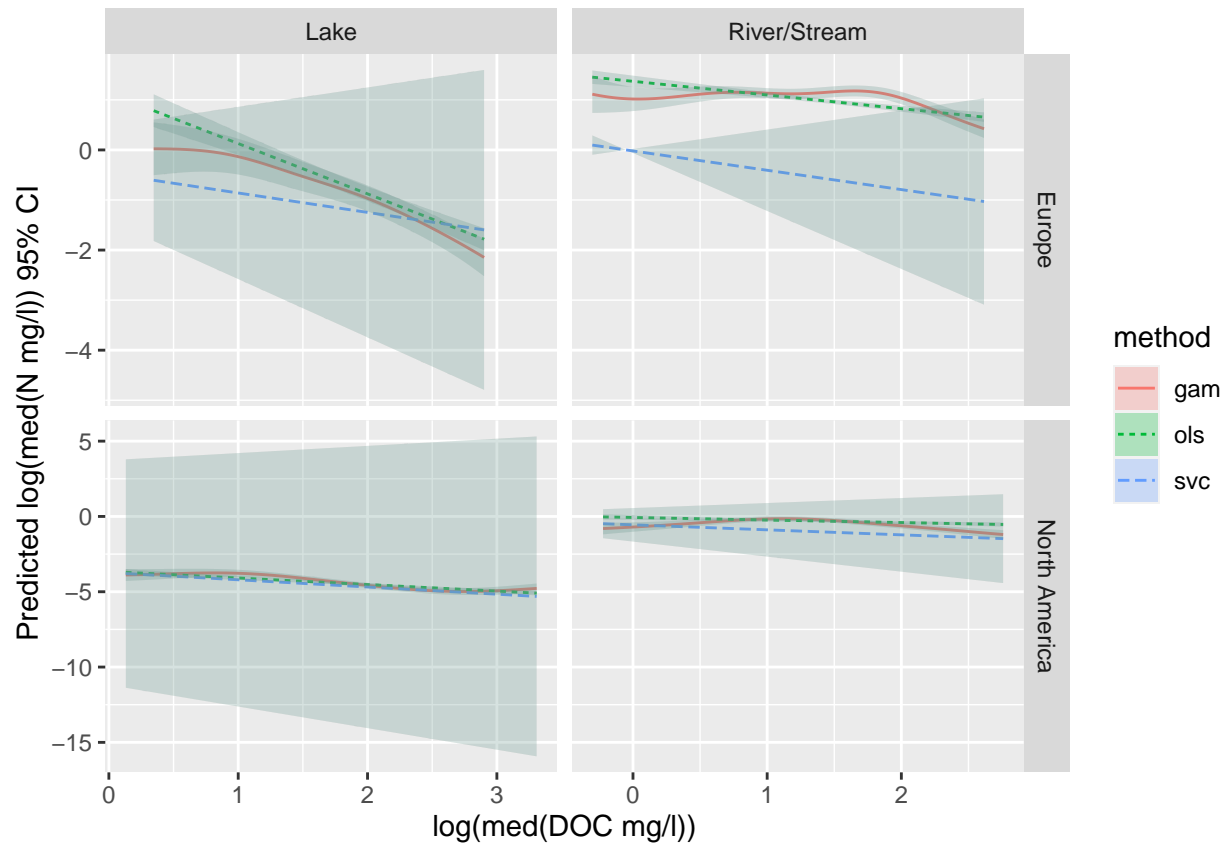
Axis on different scales



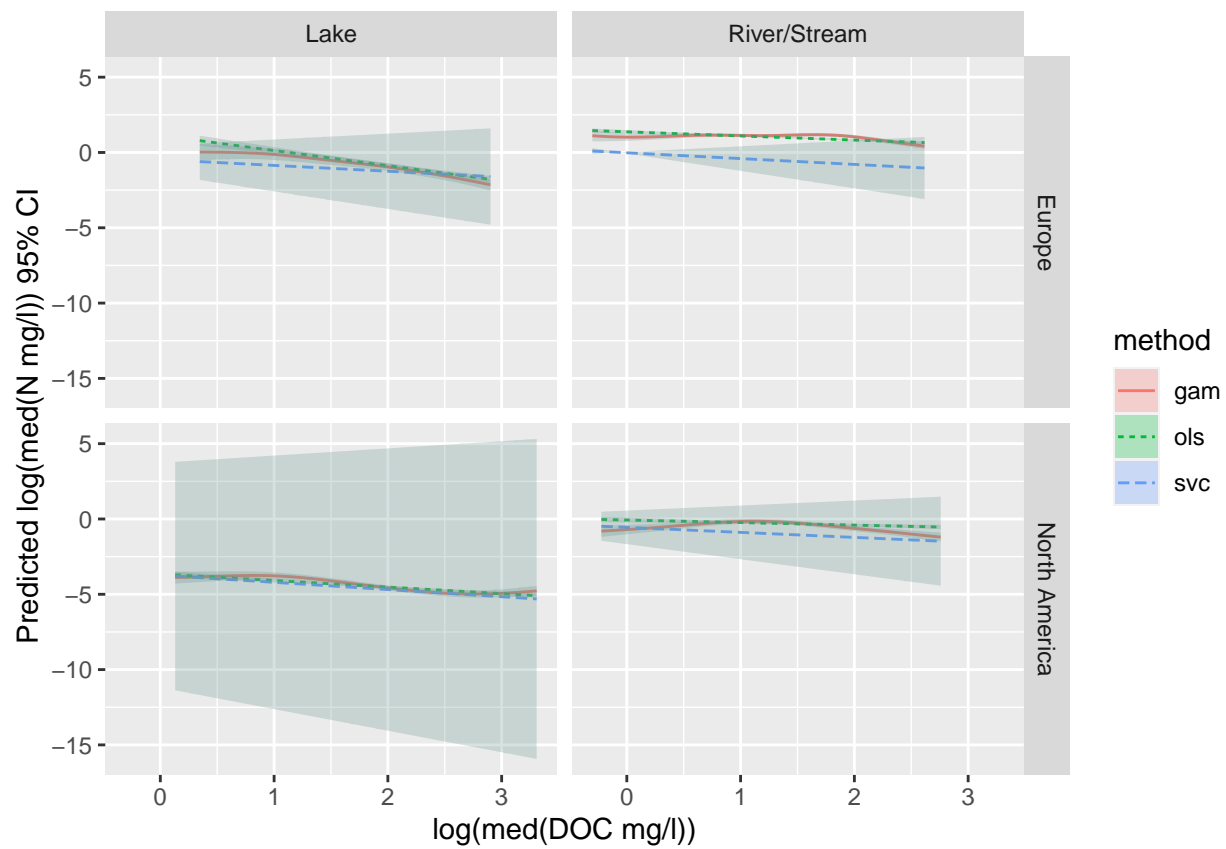
## 75th quantile

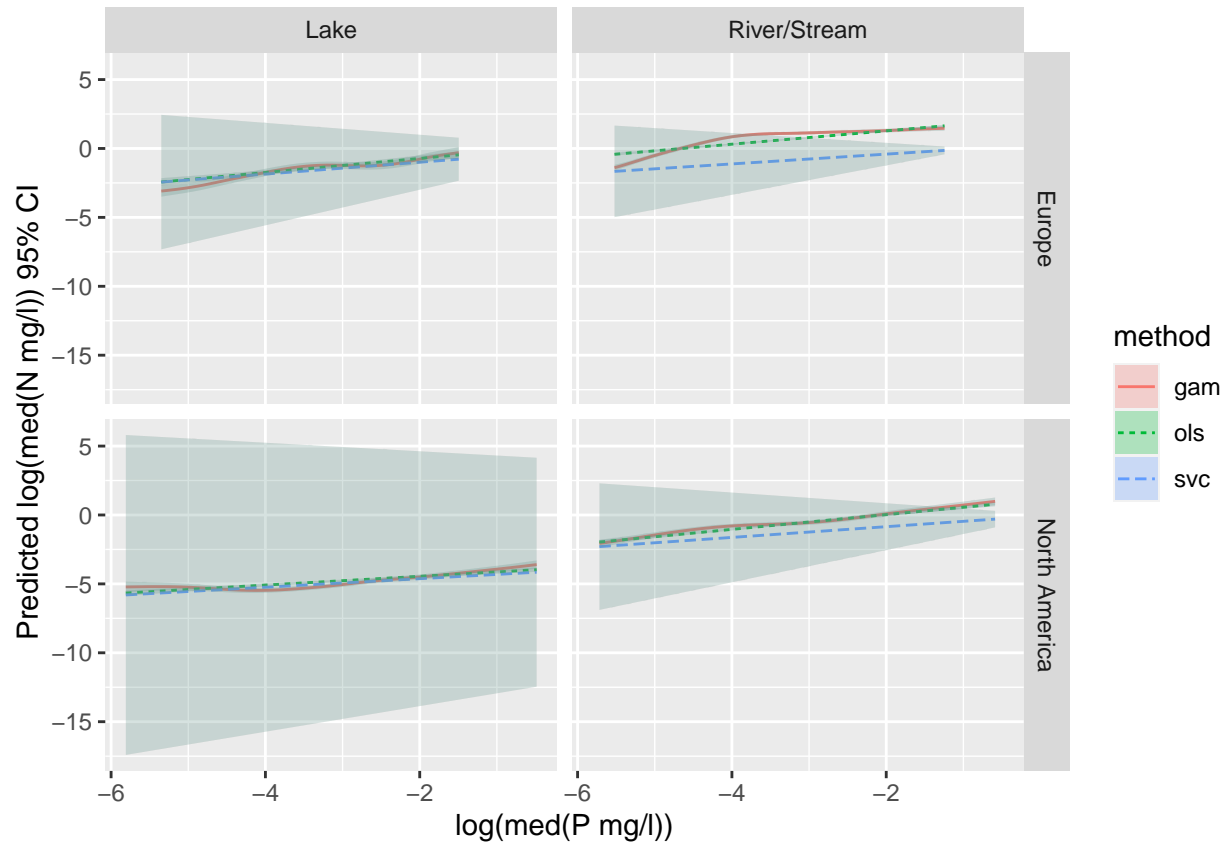
All axis on same scale

Axis on different scales









## Caveats

Not sure what's going on with the SVC confidence intervals. They seem weirdly narrow at some values.

The SVC spatial average prediction doesn't seem to match well with the GAM and OLS. I think a possible reason for this, is that when I get this prediction, I set all random effects equal to zero instead of marginalizing over them.

The GAM model I'm plotting does not include a spatial smooth. This is because there is no built-in predict method for predicting the response when the spatial smooth is set to zero. So it will have the same issues wrt to underestimation of uncertainty as the OLS model.