Supplement to Multi-scale trend analysis of water quality using error propagation of generalized additive models

Marcus W. Beck ([mbeck@tbep.org](mailto:mbeck@tbep.org)), Perry de Valpine ([pdevalpine@berkeley.edu](mailto:pdevalpine@berkeley.edu)), Rebecca Murphy ([rmurphy@chesapeakebay.net](mailto:rmurphy@chesapeakebay.net)), Ian Wren ([ianw@sfei.org](mailto:ianw@sfei.org)), Ariella Chelsky ([ariellac@sfei.org](mailto:ariellac@sfei.org)), Melissa Foley ([melissaf@sfei.org](mailto:melissaf@sfei.org)), David B. Senn ([davids@sfei.org](mailto:davids@sfei.org))

Last build 2021-05-05 18:04:13

# Comments on the select argument in `mgcv

When select = TRUE is included, the comparison between models S and SY changes. This option tells mgcv to penalize the coefficient of the linear terms in the spline. This would be appropriate if cont\_year was an explanatory variable subject to variable selection, but it is irrelevant if including both a linear and spline term for cont\_year. If select = TRUE is used, models S and SY would still be effectively equivalent, but AIC selection would suggest that one model is superior. This result would be an artifact of the choice in model SY to include a linear trend in cont\_year both as a separate term and as part of the spline, with the latter subject to penalization.

# Back-transformation of model results

Model results were back-transformed from log-space to aid in the interpretation of trends ([Bradu and Mundlak 1970](#ref-Bradu70), [Duan 1983](#ref-Duan83)). Back-transformation was accomplished using equation (1) for estimates of mean values and endpoints of confidence intervals from GAM results, such that:

where the back-transformed, expected value of the response variable (chl-a) is a function of the predicted value (mean or confidence interval endpoint) in log-space and a dispersion estimate from the model. The dispersion is the residual variance estimated from the GAM fit.

# Figures

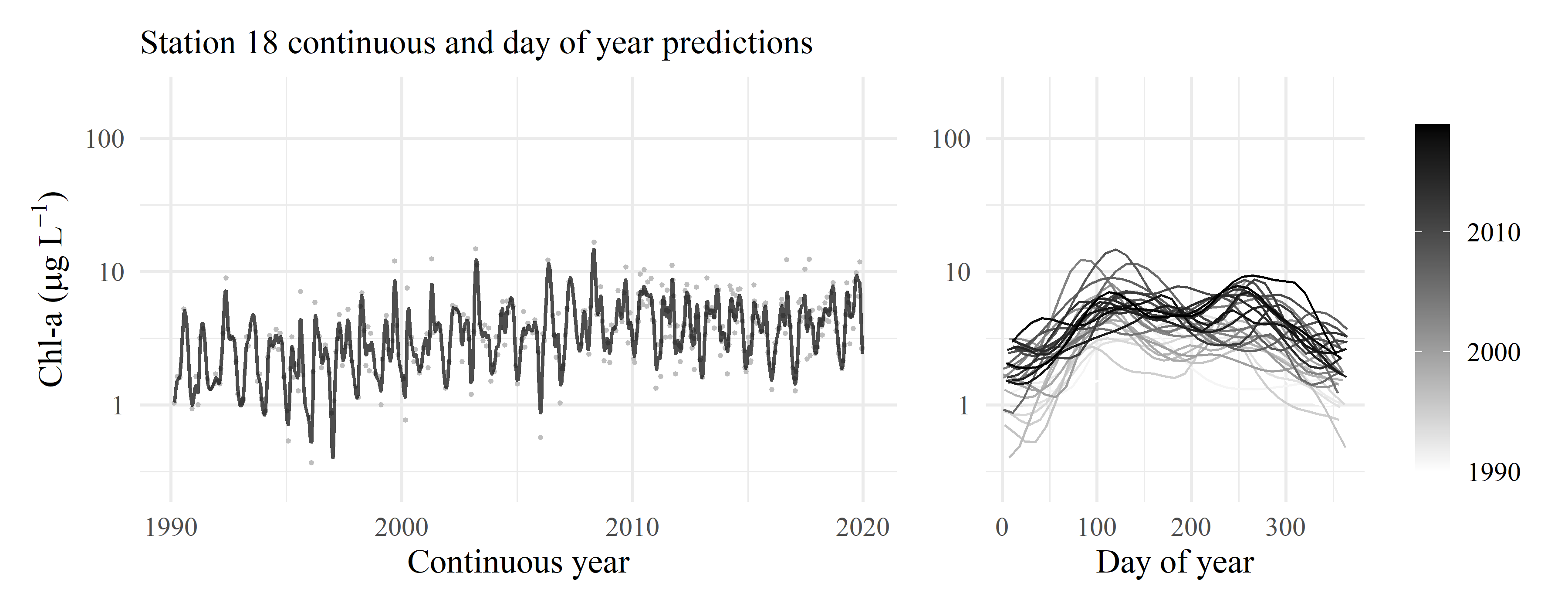


Figure 1: GAM predictions for station 18 for model S. The results show predictions across the time series and predictions by day of year. Observed data in are shown with the gray points.

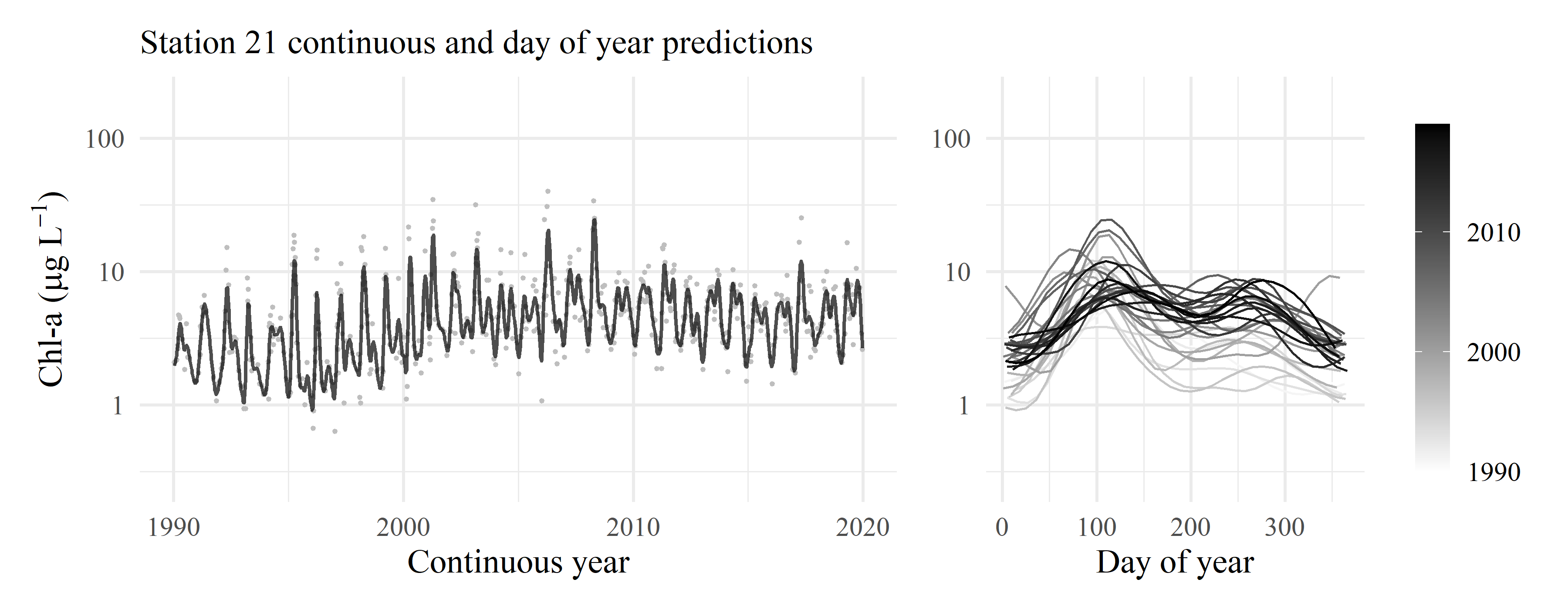


Figure 2: GAM predictions for station 21 for model S. The results show predictions across the time series and predictions by day of year. Observed data in are shown with the gray points.

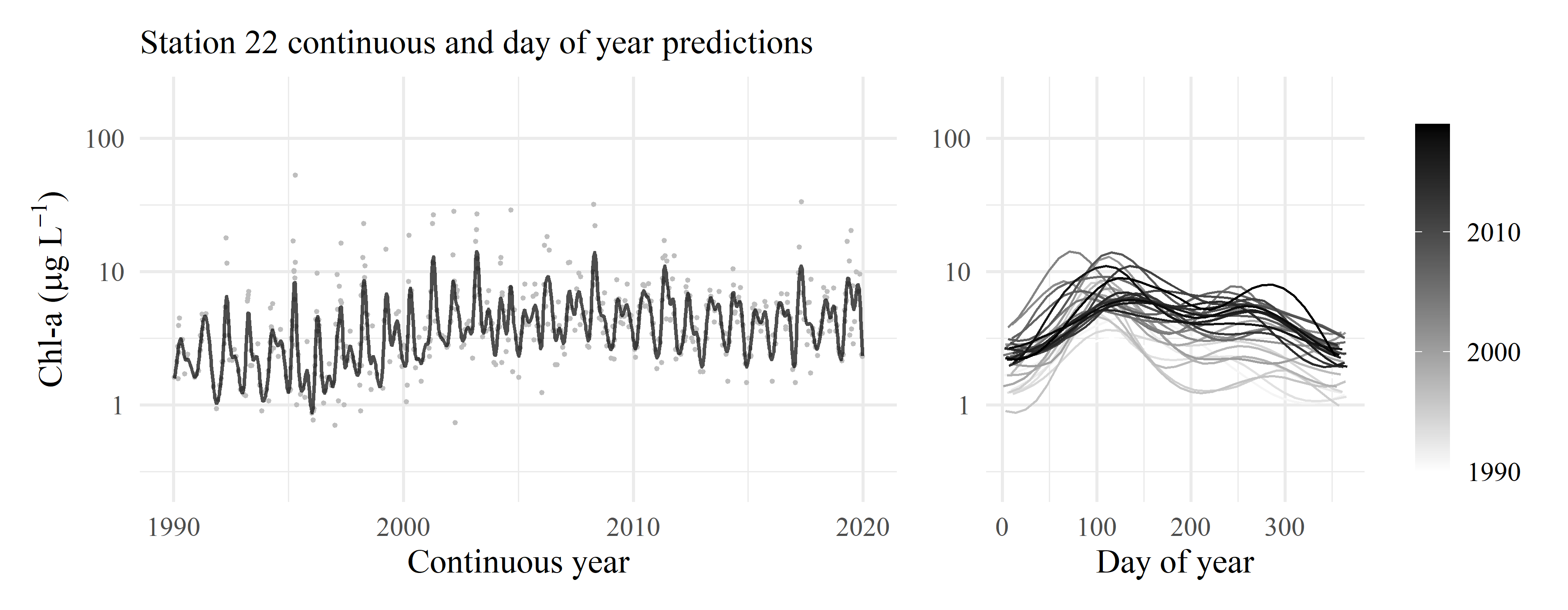


Figure 3: GAM predictions for station 22 for model S. The results show predictions across the time series and predictions by day of year. Observed data in are shown with the gray points.

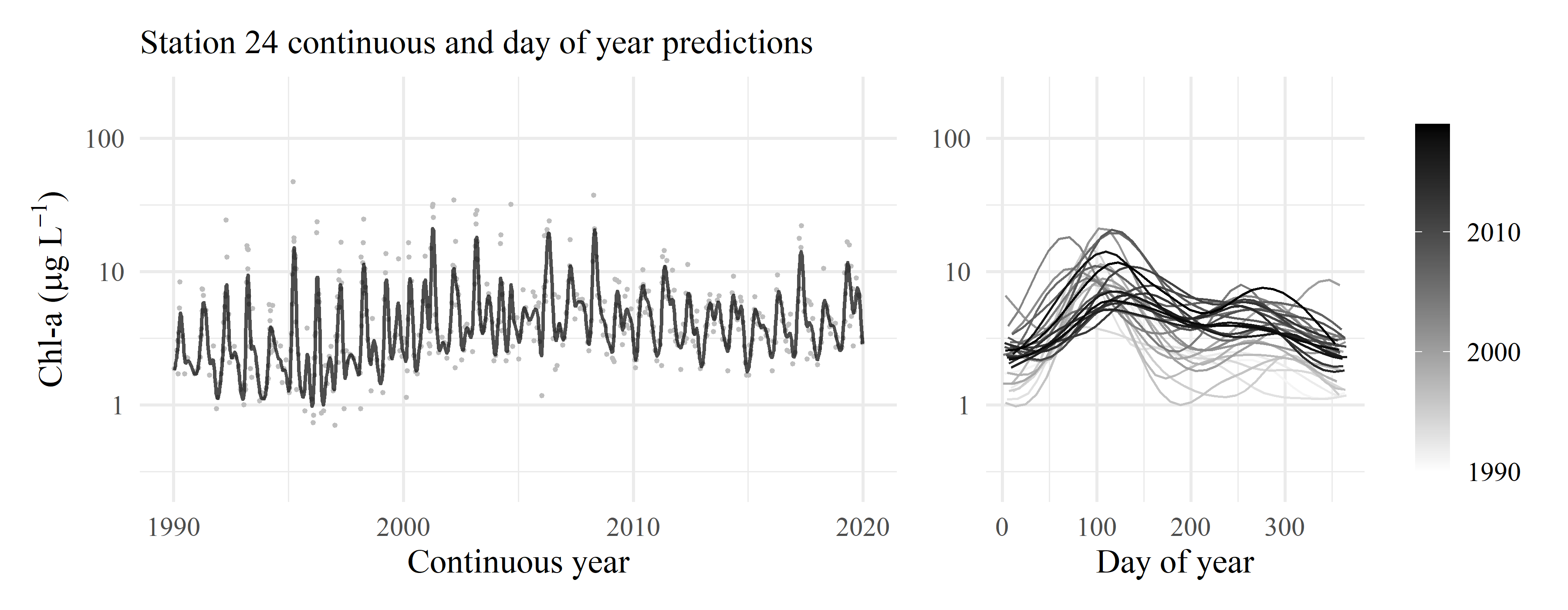


Figure 4: GAM predictions for station 24 for model S. The results show predictions across the time series and predictions by day of year. Observed data in are shown with the gray points.

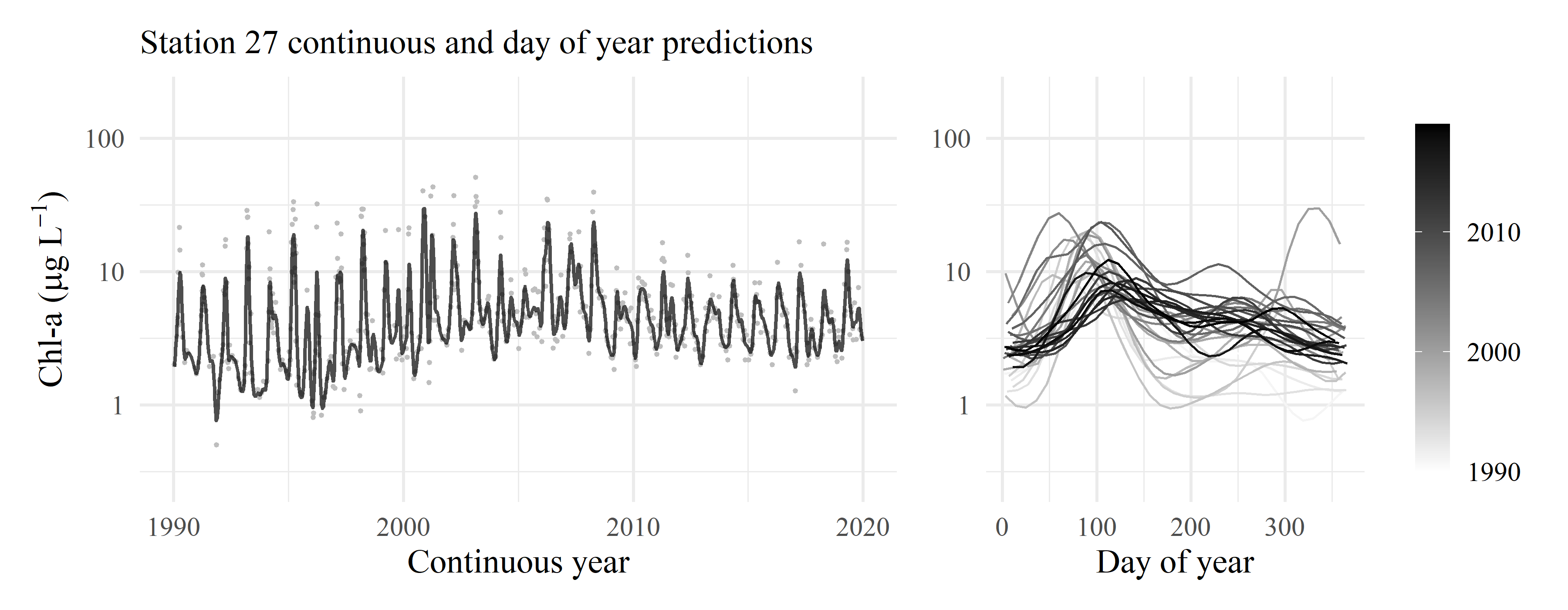


Figure 5: GAM predictions for station 27 for model S. The results show predictions across the time series and predictions by day of year. Observed data in are shown with the gray points.

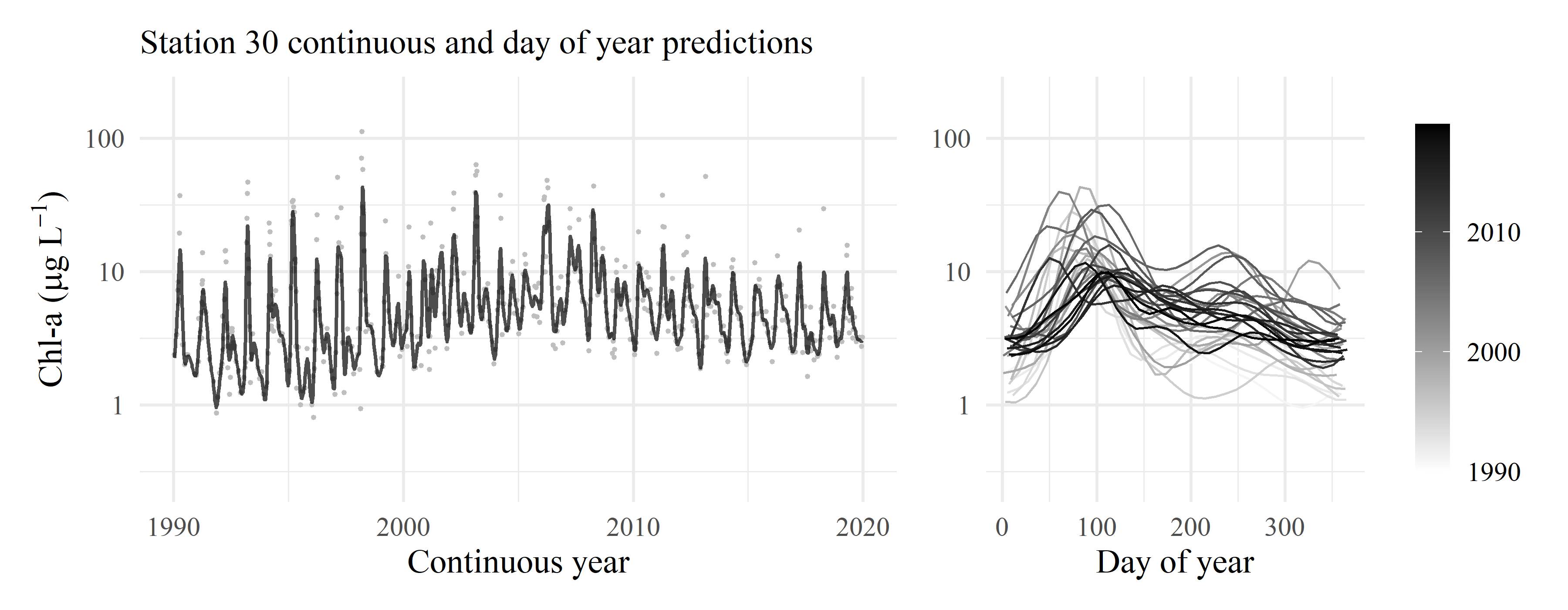


Figure 6: GAM predictions for station 30 for model S. The results show predictions across the time series and predictions by day of year. Observed data in are shown with the gray points.

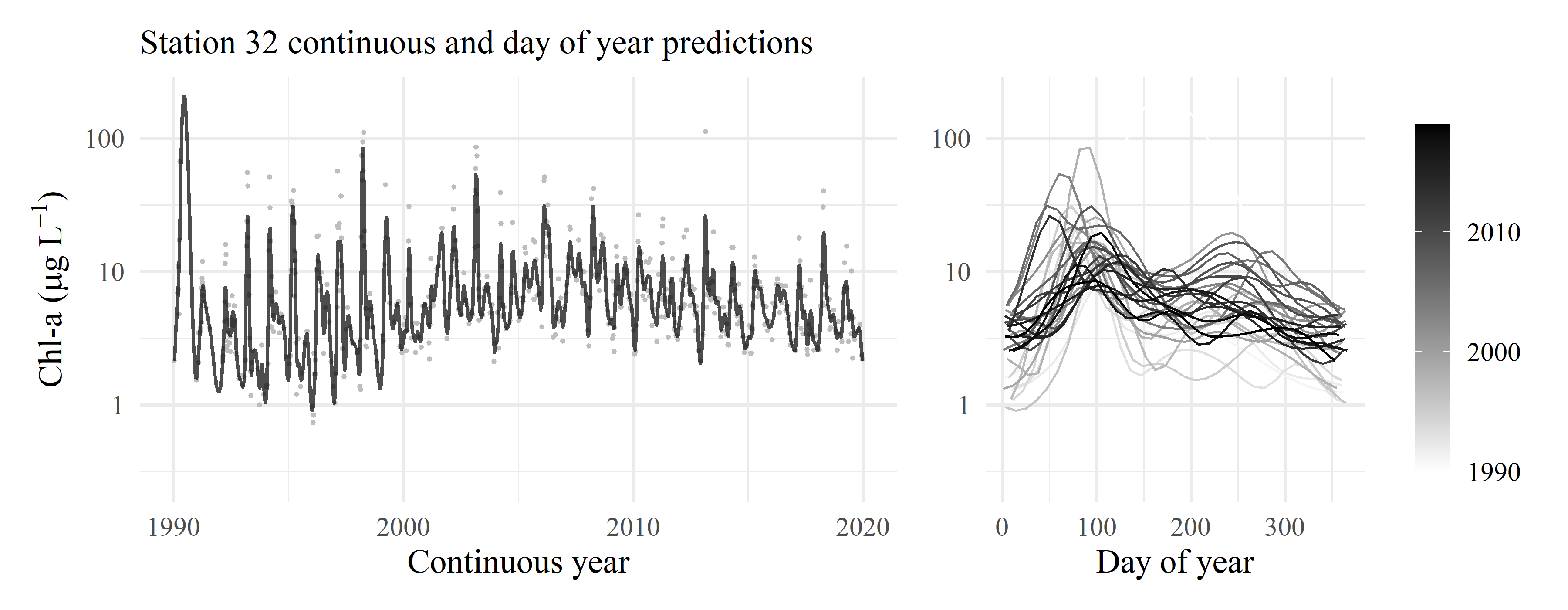


Figure 7: GAM predictions for station 32 for model S. The results show predictions across the time series and predictions by day of year. Observed data in are shown with the gray points.

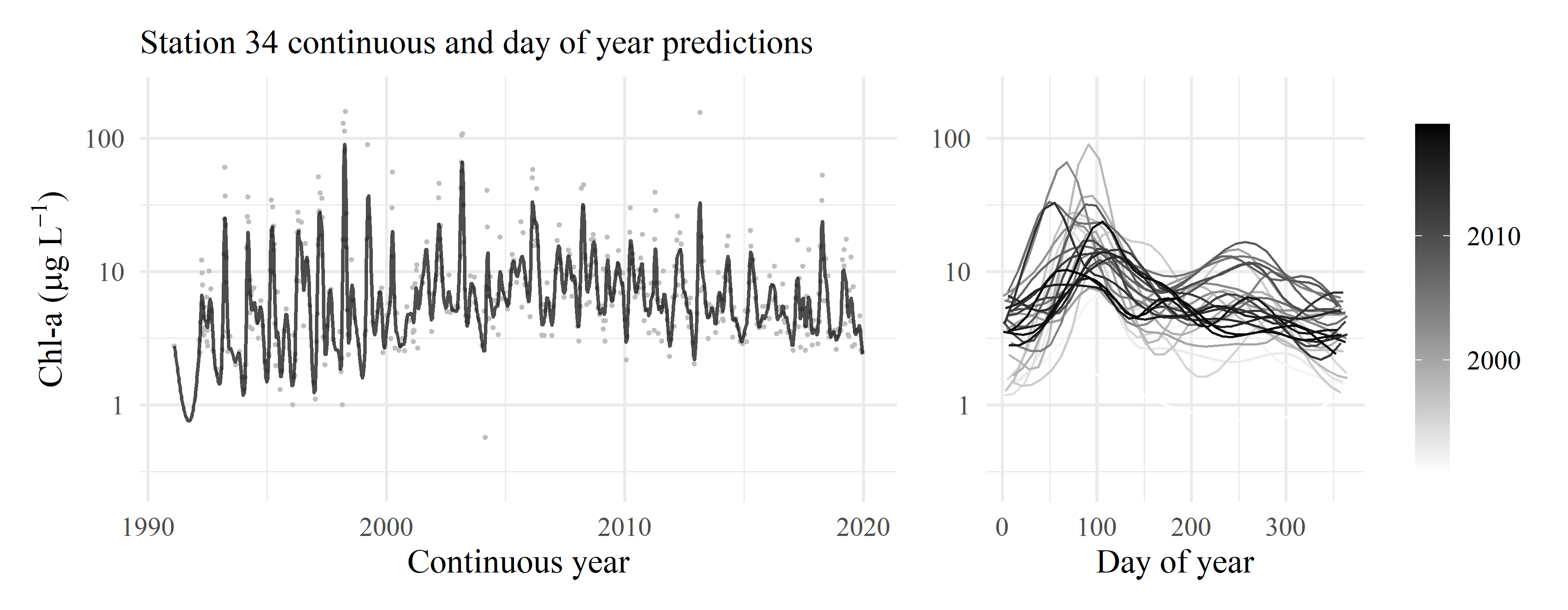


Figure 8: GAM predictions for station 34 for model S. The results show predictions across the time series and predictions by day of year. Observed data in are shown with the gray points.

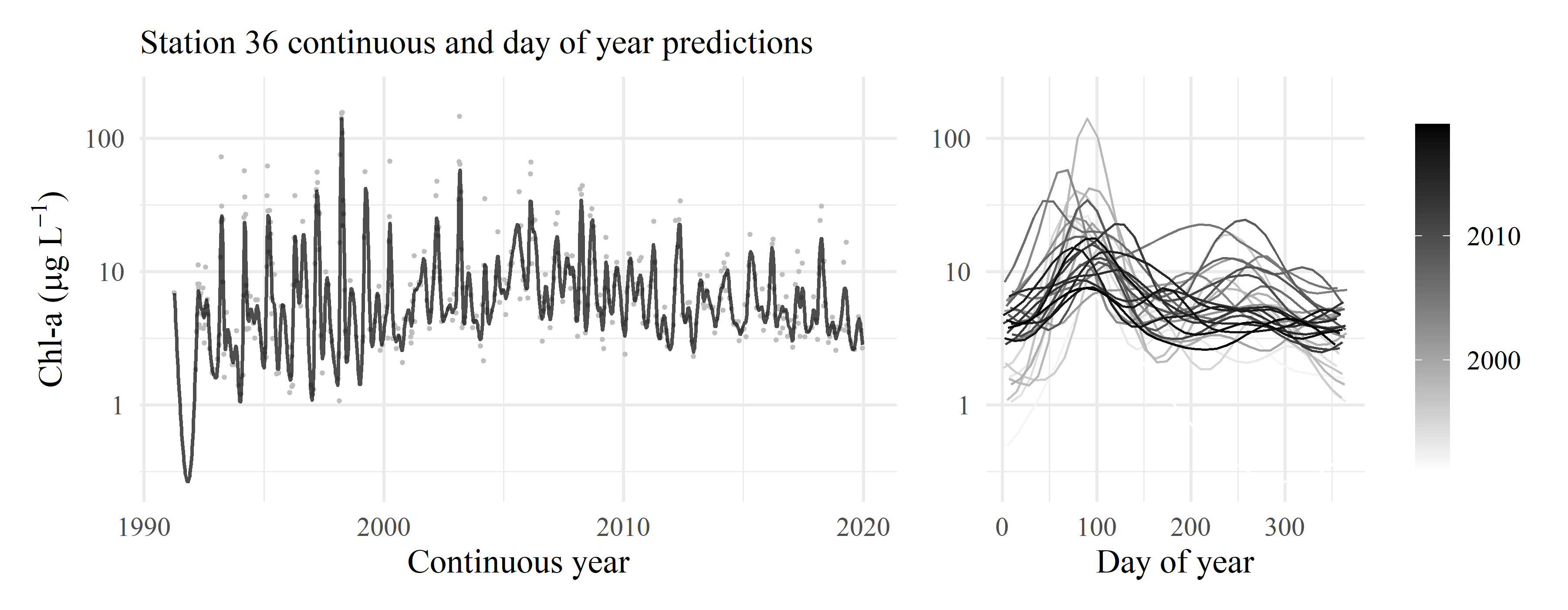


Figure 9: GAM predictions for station 36 for model S. The results show predictions across the time series and predictions by day of year. Observed data in are shown with the gray points.

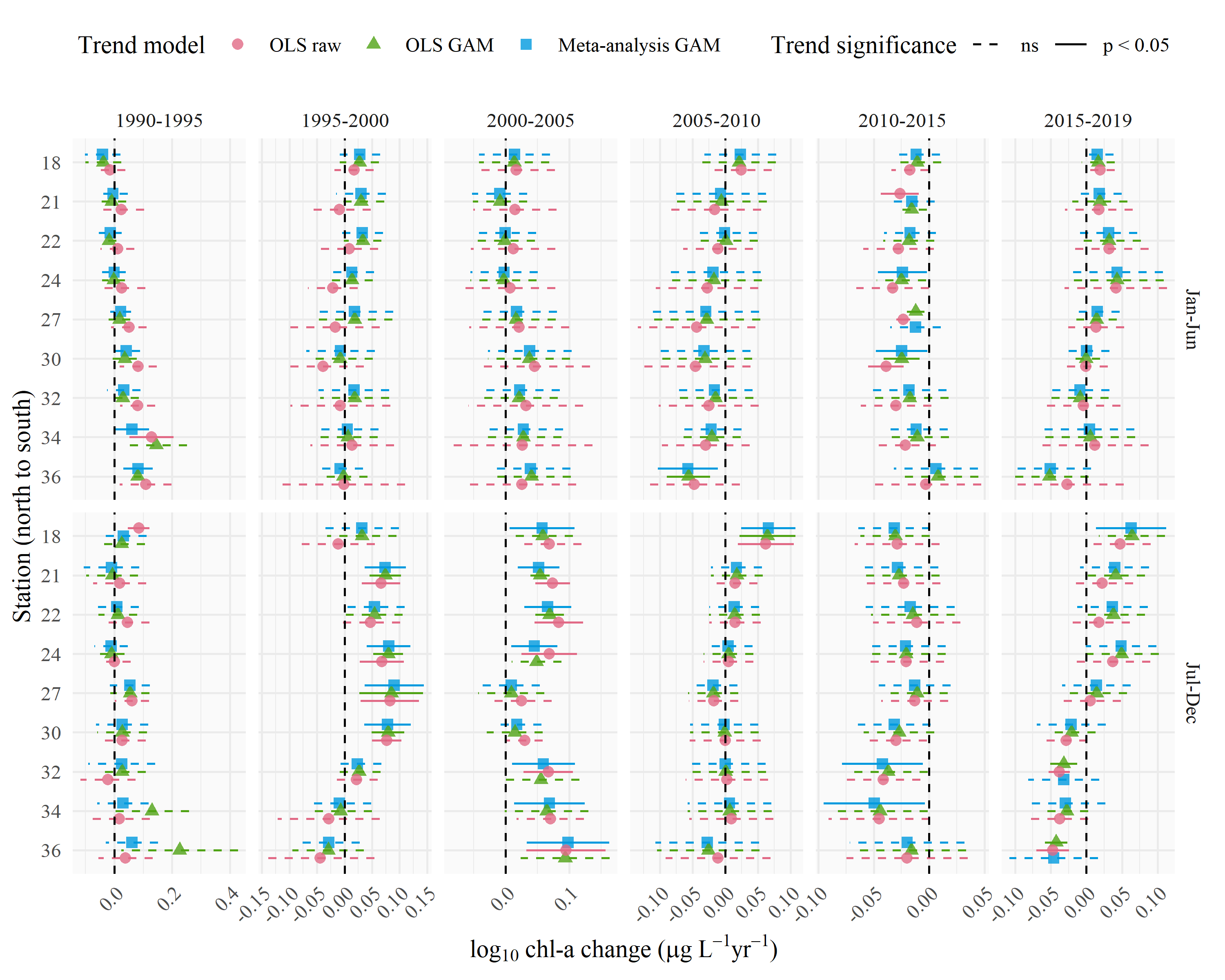


Figure 10: Trend estimate comparisons for three models applied to seasonal averages of chl-a in 5 year periods at each station. The “OLS raw” trend model is based on an ordinary least squares (OLS) regression fit to the seasonal averages of chl-a from the raw data, the “OLS GAM” trend model is based on an OLS regression fit to the seasonal averages of chl-a from the GAM model (without error propagation), and the “Meta-analysis GAM” trend model is based on a meta-analysis regression fit to the seasonal averages of chl-a from the GAM model. Values for each model are the log-slope estimates (+/- 95% confidence interval) as annual change per year within each season, with line style denoting trend significance.

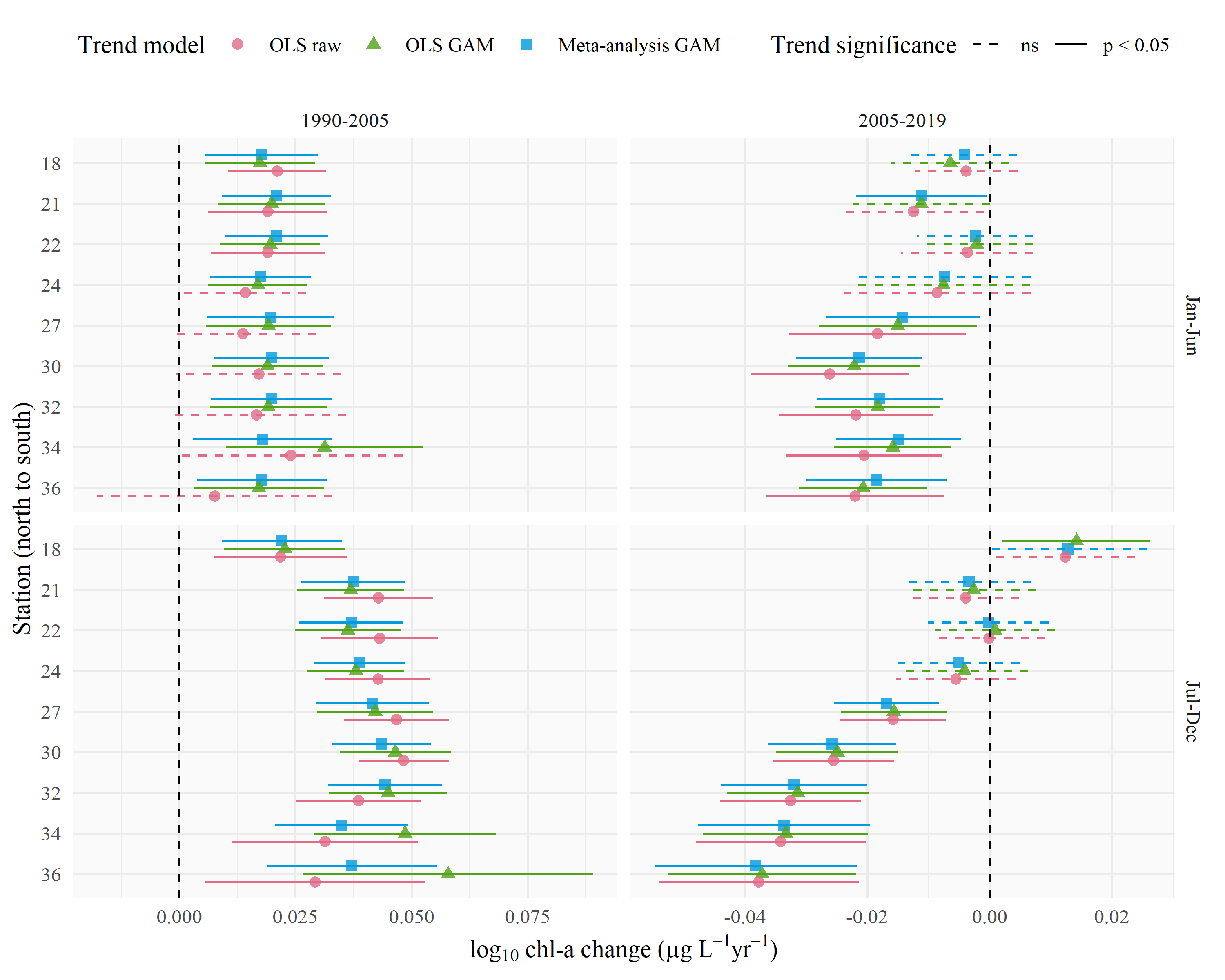


Figure 11: Trend estimate comparisons for three models applied to seasonal averages of chl-a in 15 year periods at each station. The “OLS raw” trend model is based on an ordinary least squares (OLS) regression fit to the seasonal averages of chl-a from the raw data, the “OLS GAM” trend model is based on an OLS regression fit to the seasonal averages of chl-a from the GAM model (without error propagation), and the “Meta-analysis GAM” trend model is based on a meta-analysis regression fit to the seasonal averages of chl-a from the GAM model. Values for each model are the log-slope estimates (+/- 95% confidence interval) as annual change per year within each season, with line style denoting trend significance.

# References

Bradu, D., and Y. Mundlak. 1970. Estimation in lognormal linear models. Journal of the American Statistical Association 65:198–211.

Duan, N. 1983. Smearing estimate: A nonparametric retransformation method. Journal of the American Statistical Association 78:605–610.