Supporting Materials: GAM Model Assessment*

Michael Schramm

Abstract

This document includes model summaries and diagnostics for each Nitrate-Nitrogen (NO_3-N) and Total Phosphorus (TP) Generalized Additive Model (GAM) developed in the Lavaca River Watershed.

1 Modeling Approach

Site-specific Generalized Additive Models (GAMs) were developed for Nitrate-Nitrogen (NO_3-N) and Total Phosphorus (TP). Previous papers suggest that other regression based load estimators, namely Weighted Regressions on Time Discharge and Season (WRTDS) produce results similar to GAMs where flow and season are primary drivers in nutrient loads (Beck and Murphy 2017). However, the mgcv package used to implement GAMs are easily extended to additional predictor variables and therefore chosen to be applied here (Beck and Murphy 2017; Robson and Dourdet 2015; Kuhnert et al. 2012).

A suite of potential flow-based predictor variables were developed based on (Zhang and Ball 2017).

$$Y = s(ddate) + s(yday) + s(log1p(Q)) + s(stfa) + s(ma)$$

where Y is the response variable that is a function of the sum of some smoothed predictor variables. For each of the stream sites the response variable is daily NO_3-N or TP load in kg/day.

- $ddate$ is date converted to decimal format;
- $yday$ is the day of the year as a numeral between 1 and 366;
- $log1p(Q)$ is mean daily streamflow plus one, log transformed;
- $stfa$ is a short-term flow anomaly, a unitless term that reflects the difference in the current discharge from flows in the previous month;

*This project was funded by a Texas Coastal Management Program grant approved by the Texas Land Commissioner, providing financial assistance under the Coastal Zone Management Act of 1972, as amended, awarded by the National Oceanic and Atmospheric Administration (NOAA), Office for Coastal Management, pursuant to NOAA Award No. NA21NOS4190136. The views expressed herein are those of the author(s) and do not necessarily reflect the views of NOAA, the U.S. Department of Commerce, or any of their subagencies.

- ma is the exponentially smoothed flow which is used to incorporate the influence of past flows on current load or concentration estimates;

The model was slightly altered for the Lake Texana site where daily load measurements are not a function of natural flow processes, but of operation decisions. Here, we expect concentration and load discharged from the dam to vary as a function of tributary inputs and some unknown lake metabolism processes. At this location, daily concentrations were modeled as a function of total inflow to the lake. Each of the flow based covariates ($log1p(Q)$, $stfa$, and ma) were calculated based on total inflow from gaged tributaries. Mean daily discharge was also included as a covariate under the assumption that hydraulic dynamics can substantially vary near the outlet of the dam based on the amount of water being discharged and influence nutrient concentration. Even under high inflow conditions, discharge might be low if it has been dry, conversely discharge could be higher than inflows if lake levels are high. Total loads below Palmetto Bend Dam were estimated using nutrient concentration in Lake Texana and reported mean daily dam discharges.

Thin plate regression splines were used for $ddate$, $log1p(Q)$, $ltfa$, and ma. A cyclic cubic regression spline was used for $yday$ to ensure ends of the spline match (day 1 and day 366 should match). First-order penalties were applied to the smooths of flow-based variables which penalize departures from a flat function to help constrain extrapolations for high flow measurements. GAMs were fit using the `gam()` function in the "mgcv" package in R version 4.2.1. Basis dimensions used for smooths were adjusted after using the `gam.check()` function to ensure models were not oversmoothed. Model residuals were inspected for distributional assumptions using the "gratia" package.

Left-censored nutrient concentrations were not uncommon in this dataset. Several methods are available to account for censored data. We decided to transform left-censored data to one-half the detection limit based on the fact that higher concentrations and loadings are typically associated with high-flow events and low-flow/low-concentration events will account for a small proportion of total loadings (McDowell et al. 2021). The "cenGAM" package in R provides the Tobit I family to accommodate censored data using the "gam" function in R. Censored Gamma models can be fit using a Bayesian framework with the "brms" package in R. Initial exploration using "cenGAM" and "brms" packages resulted in models that overestimated nutrient concentrations relative to "mgcv" (Bergbusch et al. 2021). All models were fit using the Gamma family and log link function.

Hold-out data is often used to validate predictive ability of a model. Given the small-sample size, we used all the available data to fit models at each site and implemented repeated 5-fold cross validation to assess model performance. Cross-validation predictions were assessed using Nash-Sutcliffe Efficiency (NSE), R^2, and Percent Bias across all folds.

2 Model Results

2.1 Lavaca River at Edna, 08164000

2.1.1 NO_3-N

Table 1: NO_3-N GAM summary - Lavaca River at Edna, USGS-08164000.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-1.950	0.146	-13.335				0.000 ***
B. smooth terms	s(ddate)				1.977	17	0.299	0.036 *
	s(yday)				0.001	4	0.000	0.442
	s(log1p(Flow))				2.320	9	0.724	0.019 *
	s(ma)				0.000	9	0.000	0.917
	s(ltfa)				0.000	9	0.000	0.445

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.107, Deviance explained 0.255
-REML : -58.587, Scale est: 1.582, N: 74

Table 2: Summary of goodness-of-fit metrics for 5-fold cross-validation of NO_3-N GAM at Lavaca River at Edna, USGS-08164000.

Goodness of Fit Metric	Median (IQR)
NSE	0.758 (0.714, 0.765)
R^2	0.761 (0.728, 0.771)
Percent Bias	-7.80 (-9.02, -4.15)

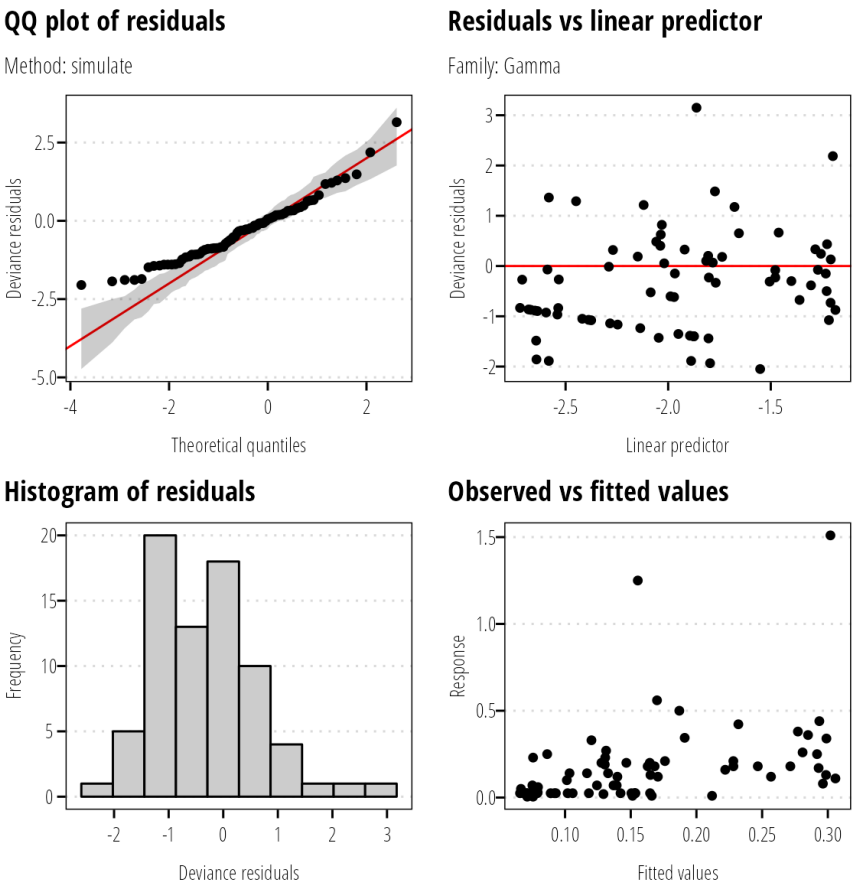

Figure 1: Diagnostic plot for NO_3-N model at USGS-08164000

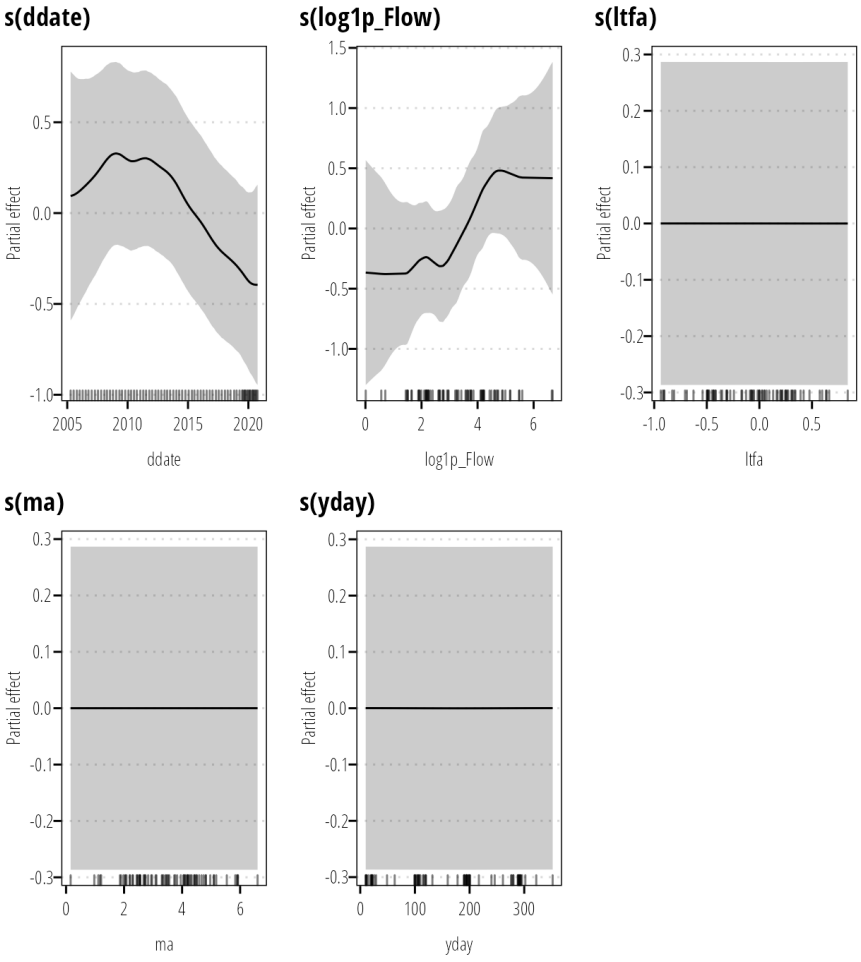

Figure 2: Partial effects of covariates in NO_3-N model at USGS-08164000

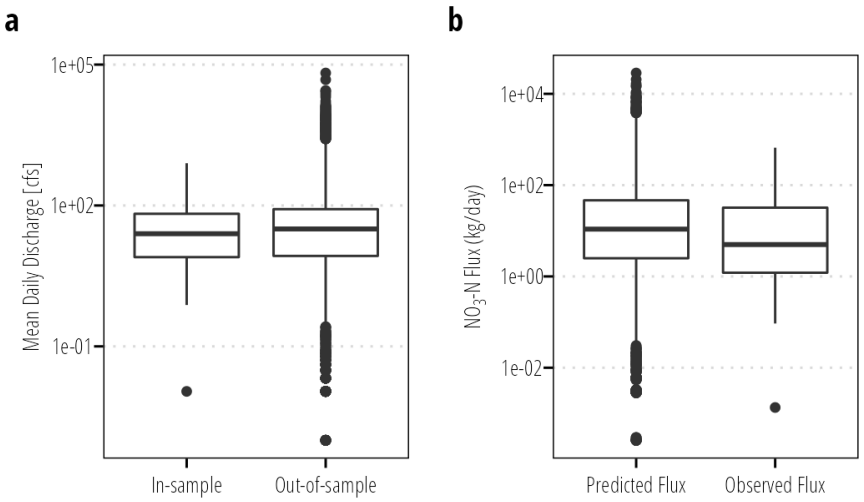

Figure 3: Comparisons of (a) in-sample and out-of-sample mean daily discharge and (b) predicted daily fluxes (for both sampled and non-sampled days) and measured daily fluxes.

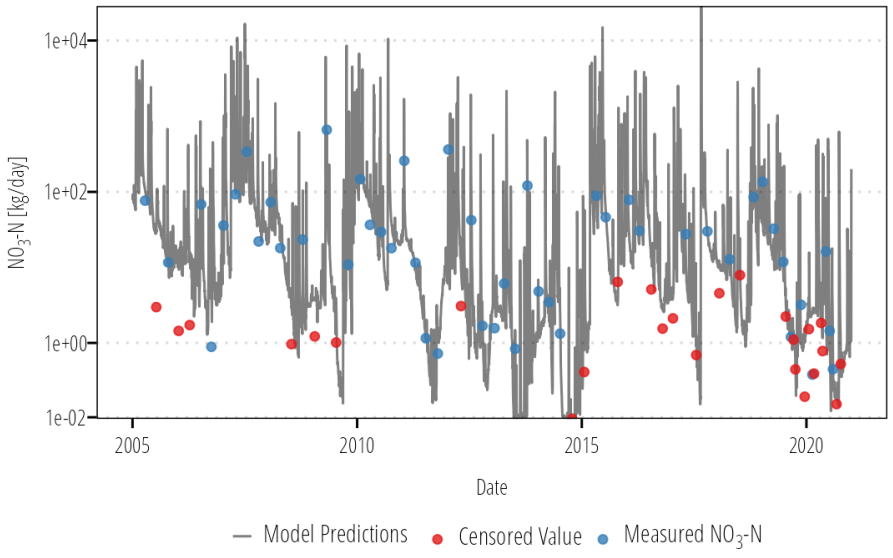

Figure 4: Time series plot of NO_3-N model predictions and observed values at USGS-08164000

2.1.2 TP

Table 3: TP GAM summary - Lavaca River at Edna, USGS-08164000.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-1.611	0.045	-35.811				0.000 ***
B. smooth terms	s(ddate)				3.262	17	0.408	0.045 *
	s(yday)				1.266	8	0.352	0.094 +
	s(log1p(Flow))				0.953	4	0.405	0.133
	s(ma)				0.000	5	0.000	0.510
	s(stfa)				2.585	4	2.857	0.003 **

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.274, Deviance explained 0.250
-REML : -70.944, Scale est: 0.162, N: 80

Table 4: Summary of goodness-of-fit metrics for 5-fold cross-validation of TP load GAM at Lavaca River at Edna, USGS-08164000.

Goodness of Fit Metric	Median (IQR)
NSE	0.77 (0.71, 0.81)
R^2	0.77 (0.72, 0.82)
Percent Bias	-7.45 (-9.10, -6.35)

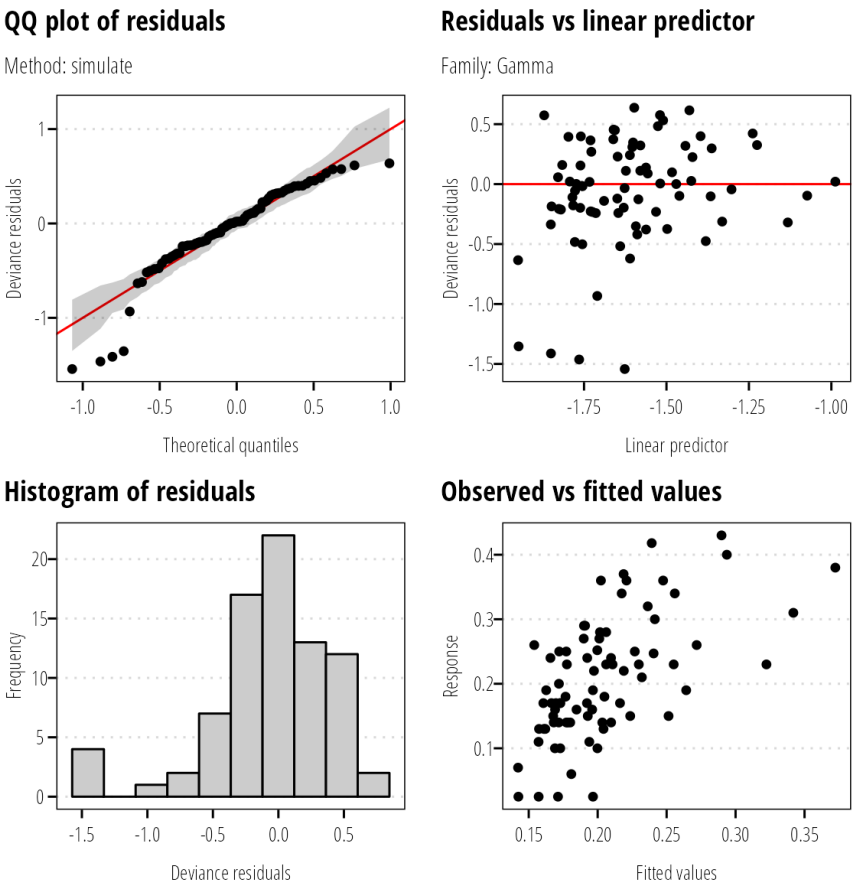

Figure 5: Diagnostic plot for TP model at USGS-08164000

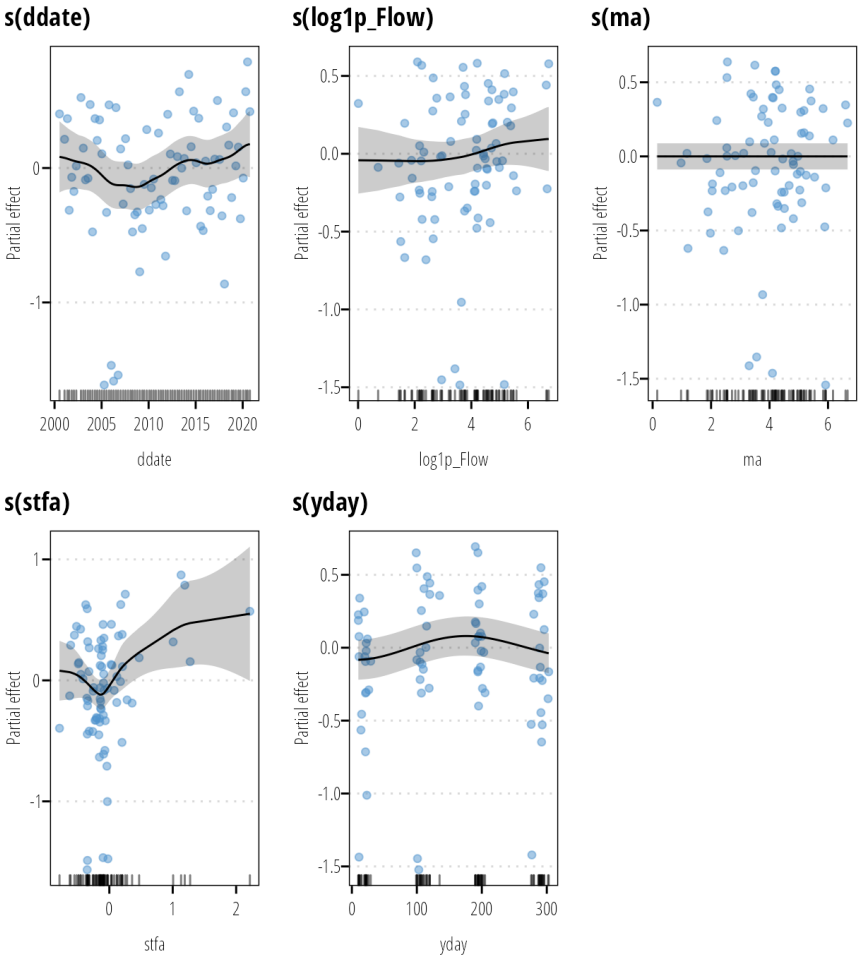

Figure 6: Partial effects of covariates in TP model at USGS-08164000

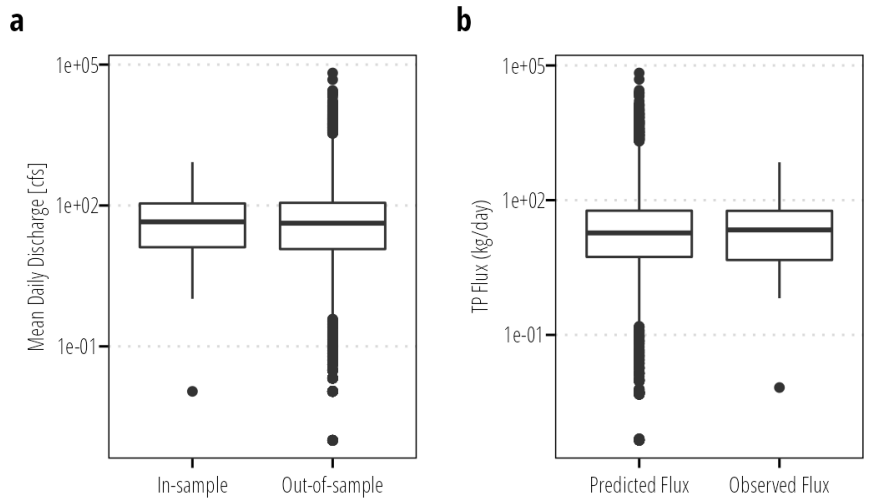

Figure 7: Comparisons of (a) in-sample and out-of-sample mean daily discharge and (b) predicted daily fluxes (for both sampled and non-sampled days) and measured daily fluxes.

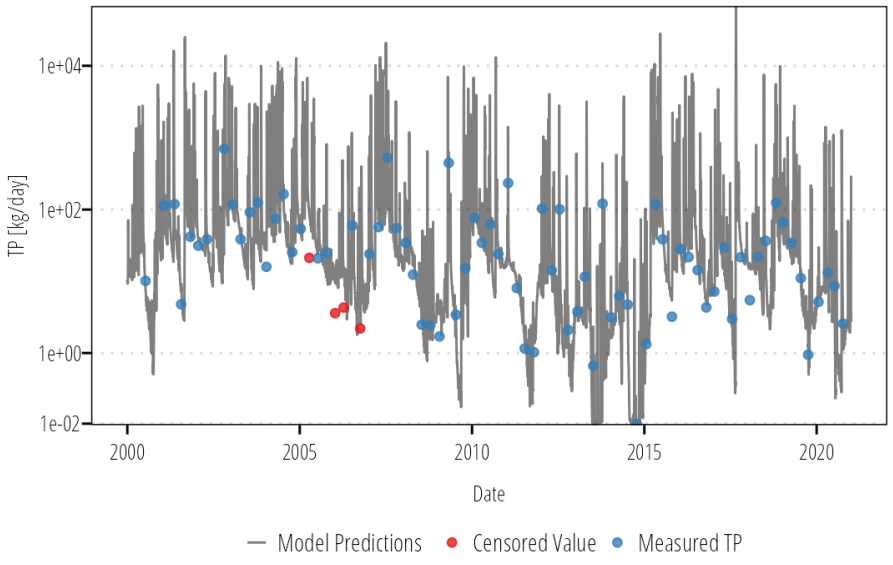

Figure 8: Time series plot of TP model predictions and observed values at USGS-08164000

2.2 Navidad River at Strane Pk nr Edna, 08164390

2.2.1 NO_3

Table 5: NO_3-N GAM summary - Navidad River at Strane Pk nr Edna, USGS-08164390.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-2.037	0.102	-20.057				0.000 ***
B. smooth terms	s(ddate)				1.685	17	0.781	0.001 ***
	s(yday)				2.486	4	5.143	0.000 ***
	s(log1p(Flow))				4.072	5	11.579	0.000 ***
	s(ma)				2.227	4	3.098	0.001 **
	s(ltfa)				0.001	9	0.000	0.387

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.717, Deviance explained 0.767
-REML : -46.034, Scale est: 0.00733, N: 59

Table 6: Summary of goodness-of-fit metrics for 5-fold cross-validation of NO_3-N concentration GAM at Navidad River at Strane Pk nr Edna„ USGS-NO308164390.

Goodness of Fit Metric	Median (IQR)
NSE	0.59 (0.53, 0.70)
R^2	0.69 (0.59, 0.78)
Percent Bias	-15.9 (-20.9, -13.0)

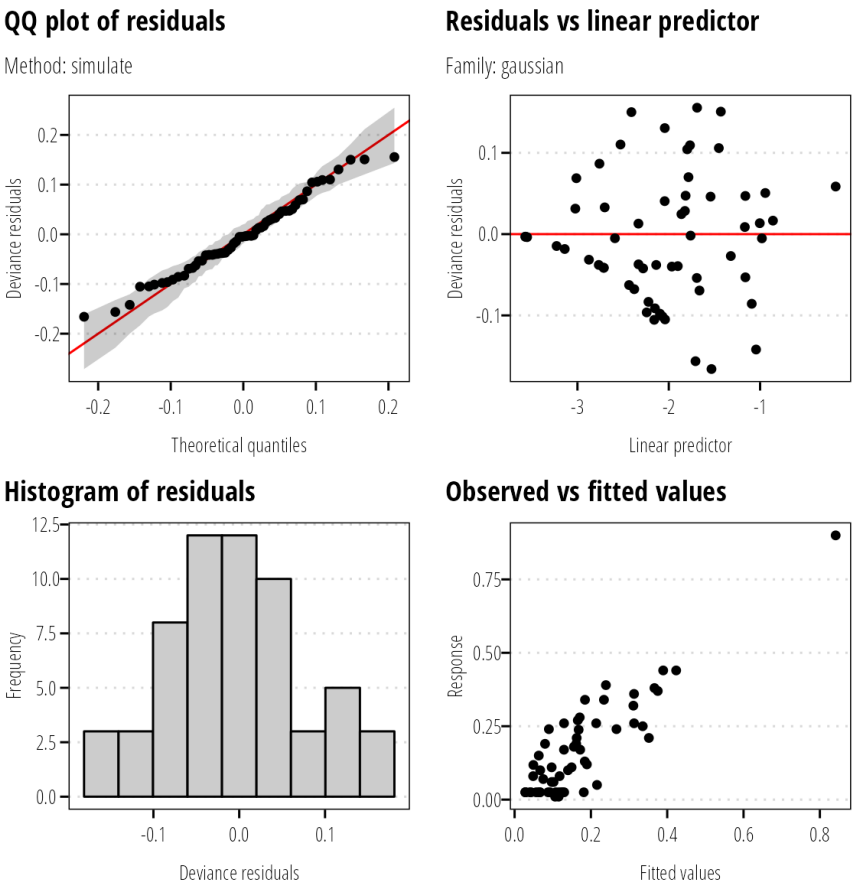

Figure 9: Diagnostic plot for NO_3 model at USGS-08164390.

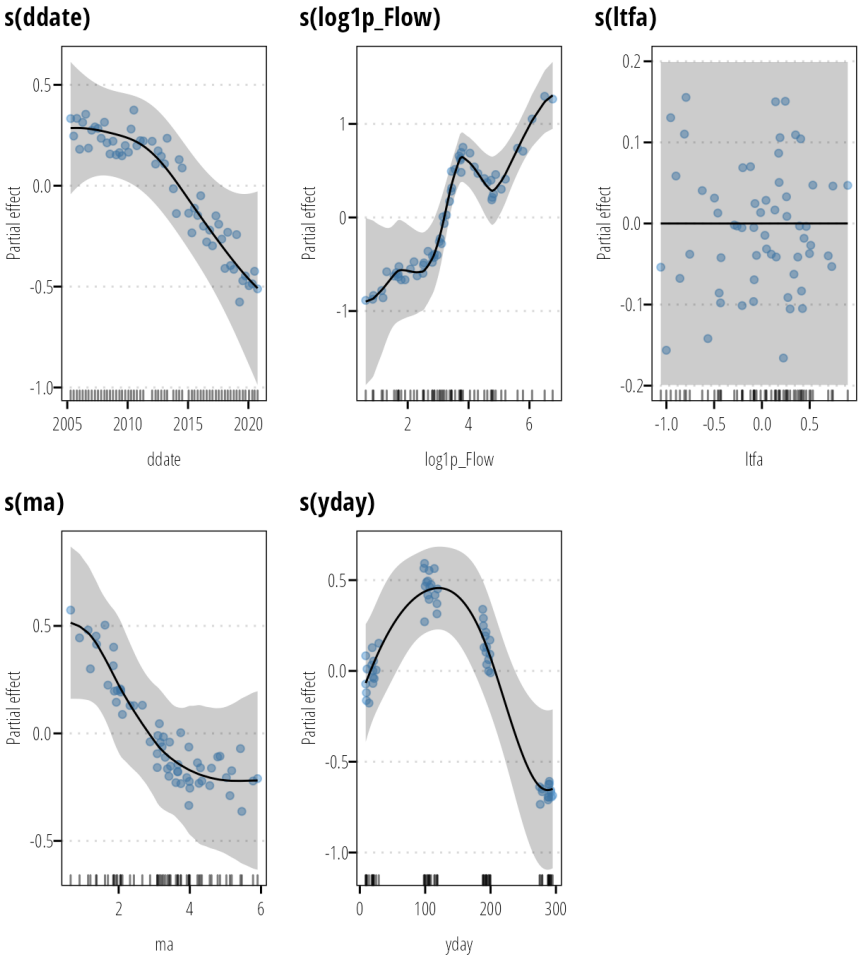

Figure 10: Partial effects of covariates in NO_3-N model at USGS-08164390.

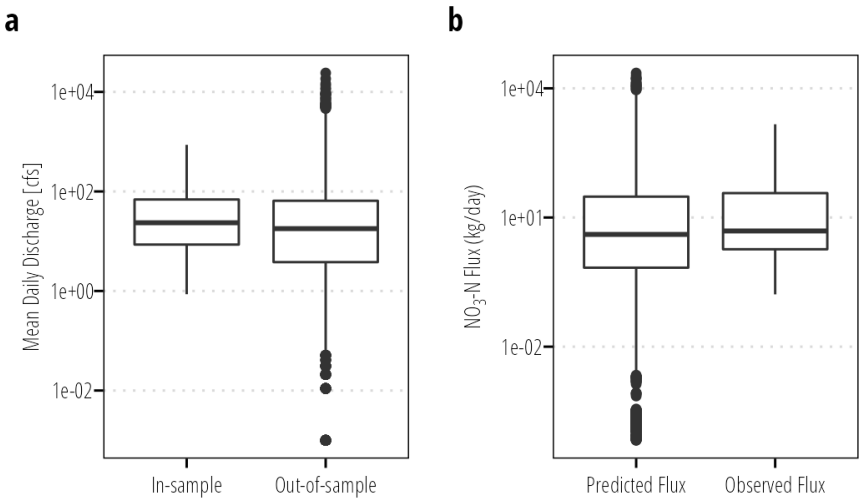

Figure 11: Comparisons of (a) in-sample and out-of-sample mean daily discharge and (b) predicted daily fluxes (for both sampled and non-sampled days) and measured daily fluxes.

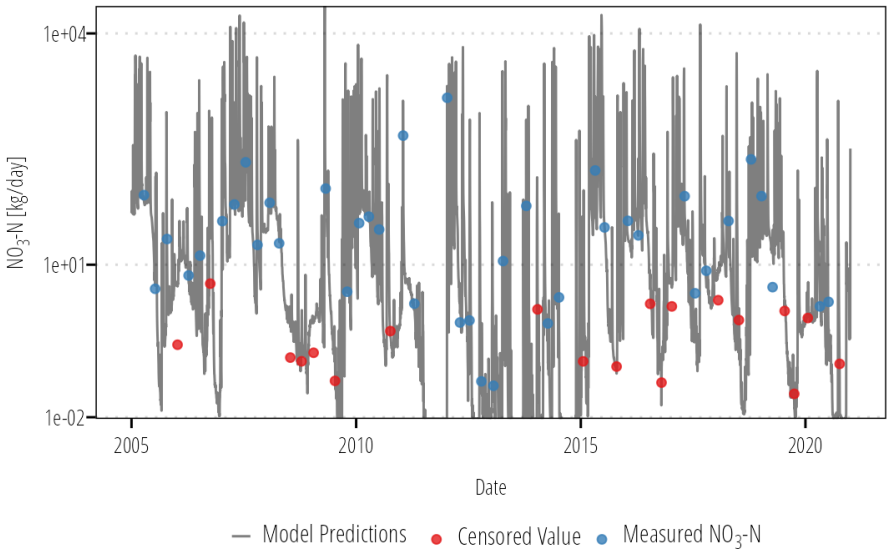

Figure 12: Time series plot of NO_3 model predictions and observed values at USGS-08164390.

2.2.2 TP

Table 7: TP GAM summary - Navidad River at Strane Pk nr Edna, USGS-08164390.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-1.597	0.038	-42.298				0.000 ***
B. smooth terms	s(ddate)				7.120	17	3.465	0.000 ***
	s(yday)				0.456	4	0.147	0.270
	s(log1p(Flow))				2.630	5	2.428	0.002 **
	s(stfa)				0.000	5	0.000	0.690
	s(ma)				0.000	5	0.000	0.759

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.550, Deviance explained 0.486
-REML : -76.491, Scale est: 0.110, N: 77

Table 8: Summary of goodness-of-fit metrics for 5-fold cross-validation of TP load GAM at Navidad River at Strane Pk nr Edna, USGS-08164390.

Goodness of Fit Metric	Median (IQR)
NSE	0.951 (0.944, 0.960)
R^2	0.982 (0.973, 0.984)
Percent Bias	-9.10 (-10.10, -8.27)

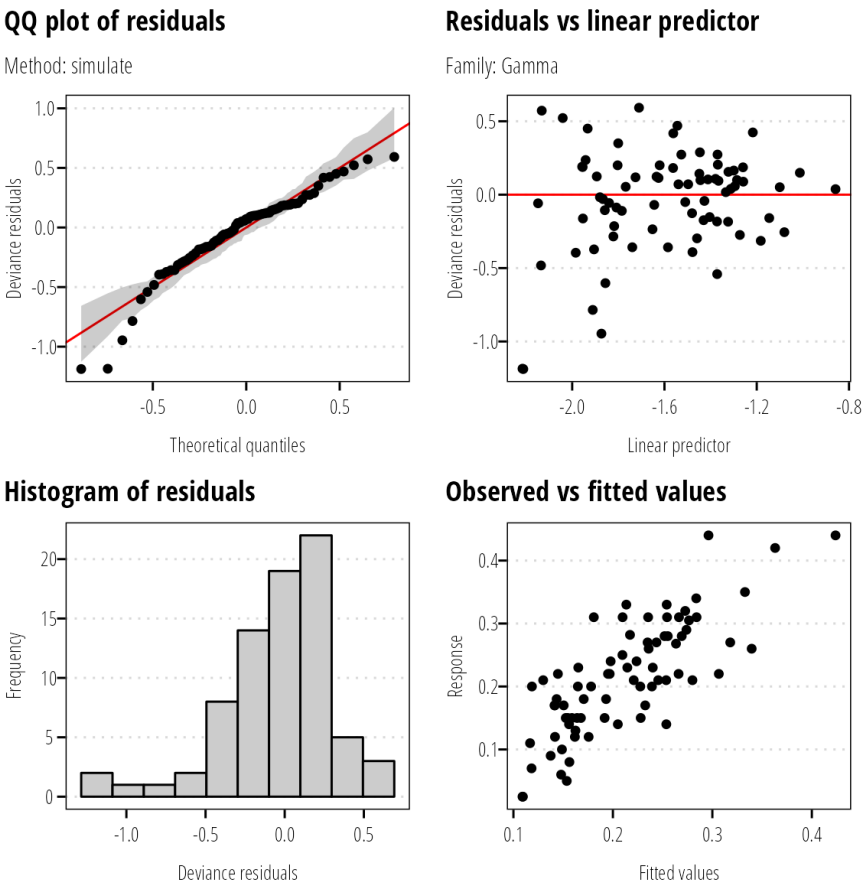

Figure 13: Diagnostic plot for TP model at USGS-08164390.

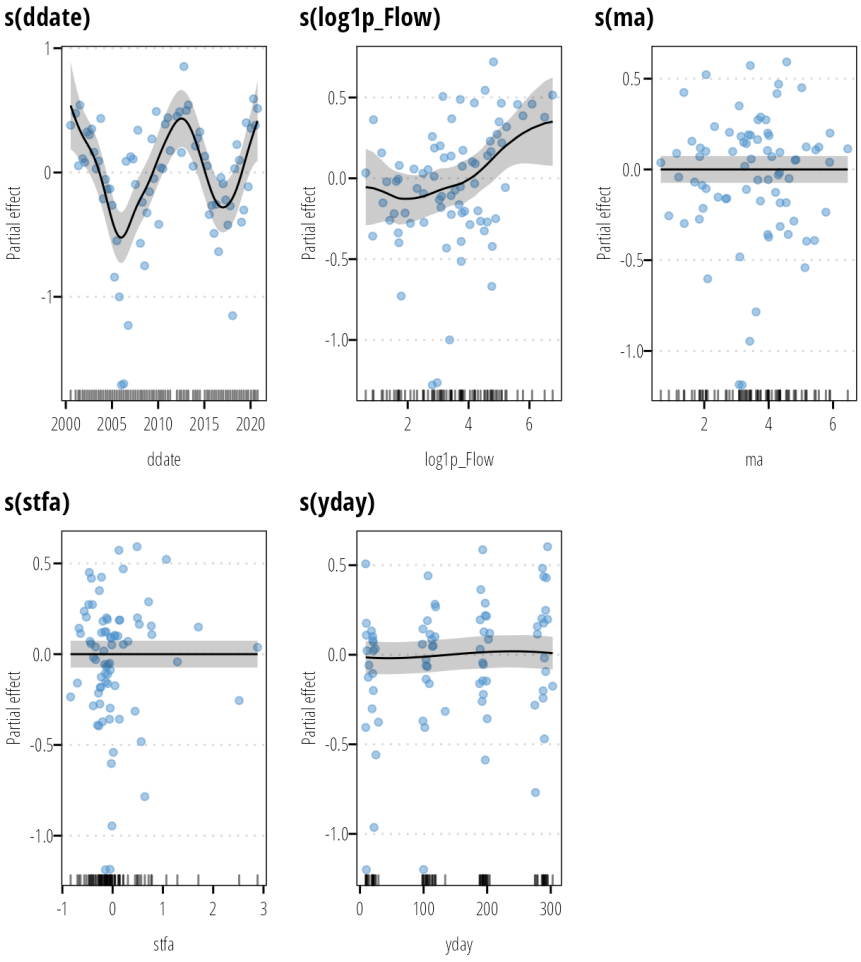

Figure 14: Partial effects of covariates in TP model at USGS-08164390.

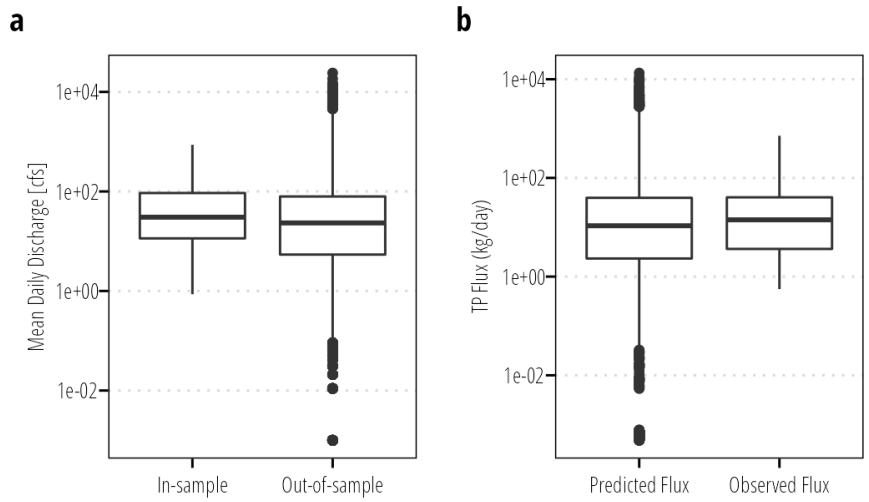

Figure 15: Comparisons of (a) in-sample and out-of-sample mean daily discharge and (b) predicted daily fluxes (for both sampled and non-sampled days) and measured daily fluxes.

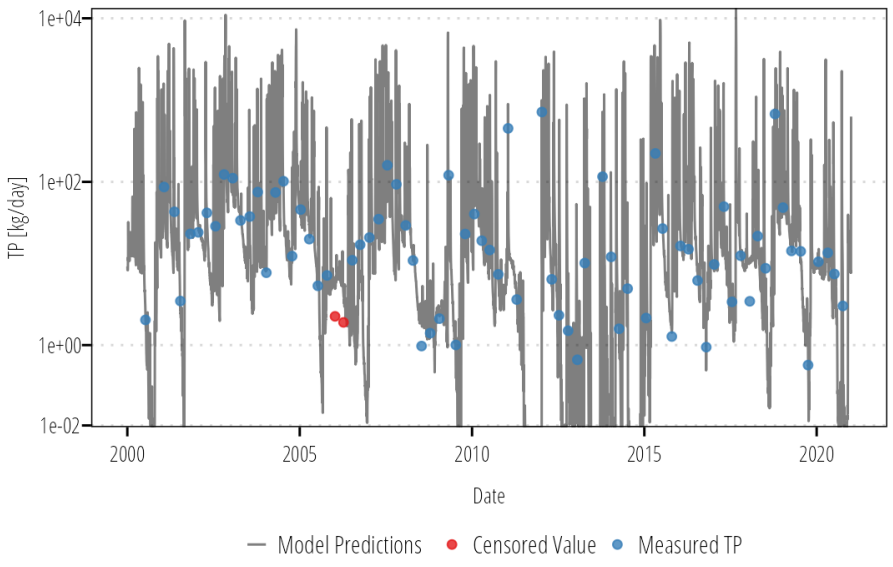

Figure 16: Time series plot of TP model predictions and observed values at USGS-08164390.

2.3 Sandy Creek nr Ganado, USGS-08164450

2.3.1 NO_3

Table 9: NO_3-N GAM summary - Sandy Creek nr Ganado, USGS-08164450.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-1.946	0.080	-24.212				0.000 ***
B. smooth terms	s(ddate)				0.000	17	0.000	0.916
	s(yday)				2.005	4	2.720	0.003 **
	s(log1p(Flow))				2.391	5	1.043	0.071 +
	s(ma)				4.107	5	4.440	0.000 ***
	s(ltfa)				0.605	5	0.138	0.302

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.274, Deviance explained 0.518
-REML : -51.543, Scale est: 0.362, N: 56

Table 10: Summary of goodness-of-fit metrics for 5-fold cross-validation of NO_3-N concentration GAM at Sandy Creek nr Ganado, USGS-08164450.

Goodness of Fit Metric	Median (IQR)
NSE	0.45 (0.24, 0.49)
R^2	0.46 (0.33, 0.56)
Percent Bias	-16 (-21, -8)

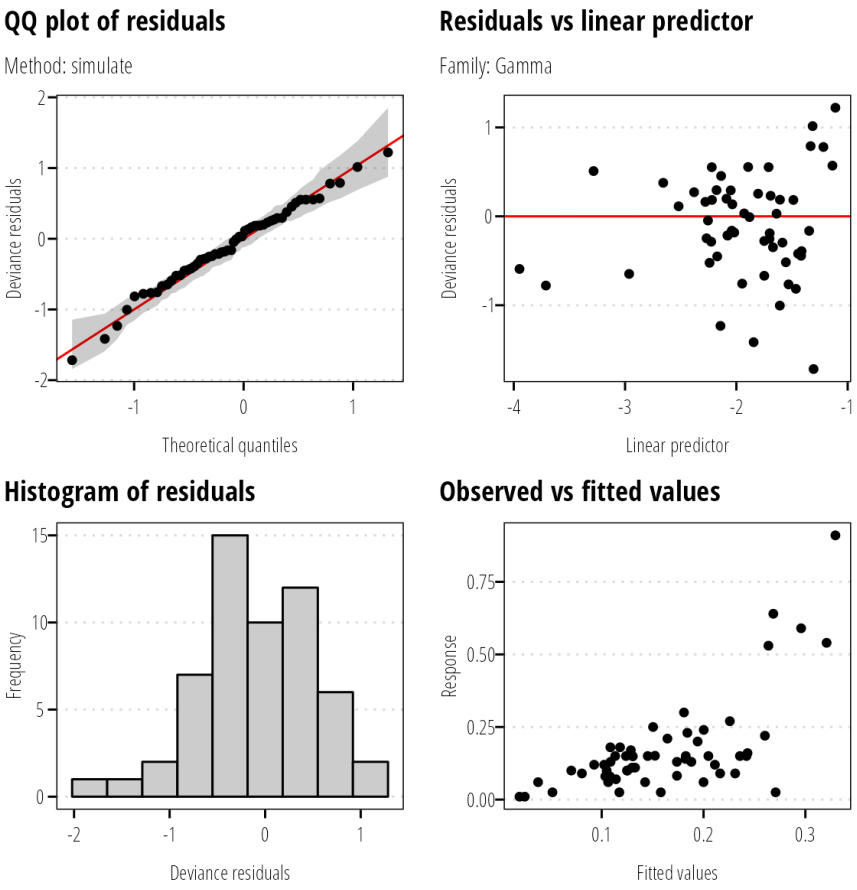

Figure 17: Diagnostic plot for NO_3 model at USGS-08164450.

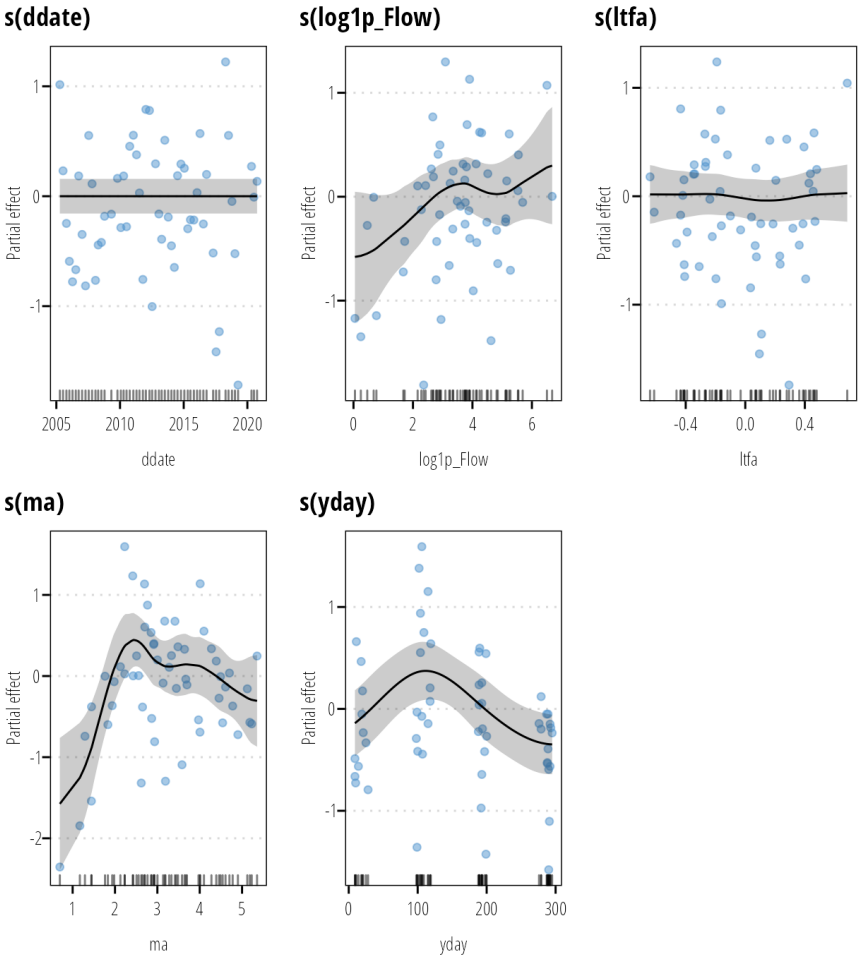

Figure 18: Partial effects of covariates in NO_3-N model at USGS-08164450.

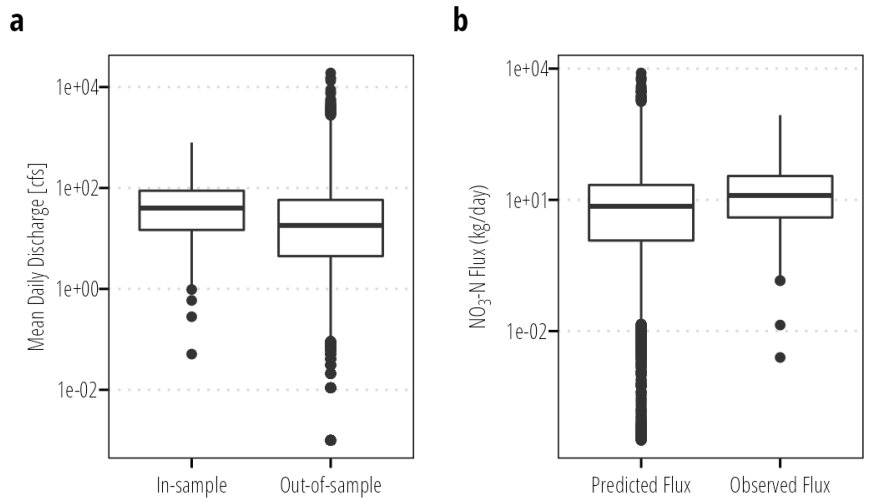

Figure 19: Comparisons of (a) in-sample and out-of-sample mean daily discharge and (b) predicted daily fluxes (for both sampled and non-sampled days) and measured daily fluxes.

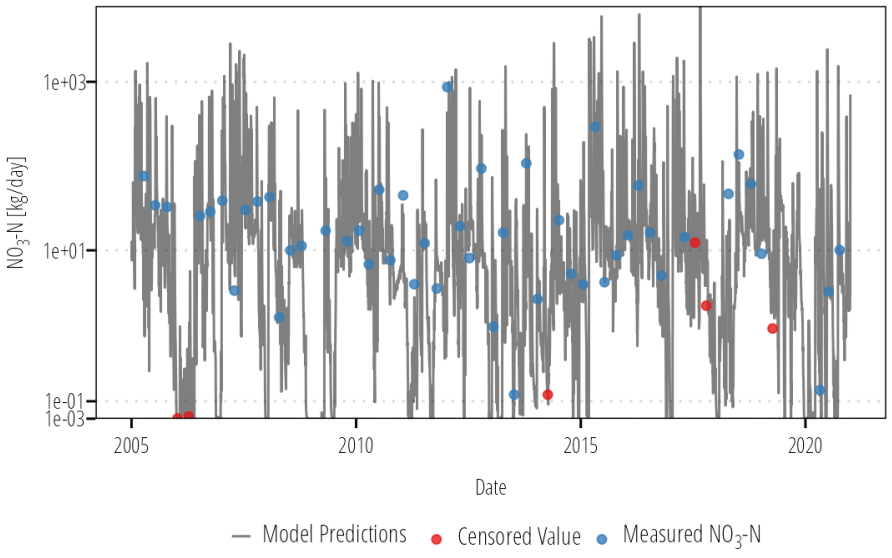

Figure 20: Time series plot of NO_3 model predictions and observed values at USGS-08164450.

2.3.2 TP

Table 11: TP GAM summary - Sandy Creek nr Ganado, USGS-08164450.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-1.729	0.067	-25.973				0.000 ***
B. smooth terms	s(ddate)				9.316	17	4.295	0.000 ***
	s(yday)				0.000	4	0.000	0.730
	s(log1p(Flow))				6.939	9	2.967	0.000 ***
	s(stfa)				2.097	5	0.757	0.090 +
	s(ma)				2.171	4	3.529	0.000 ***

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.757, Deviance explained 0.824
-REML : -34.024, Scale est: 0.00944, N: 75

Table 12: Summary of goodness-of-fit metrics for 5-fold cross-validation of TP load GAM at Sandy Creek nr Ganado, USGS-08164450.

Goodness of Fit Metric	Median (IQR)
NSE	0.78 (0.56, 0.81)
R^2	0.81 (0.67, 0.86)
Percent Bias	-6 (-9, -3)

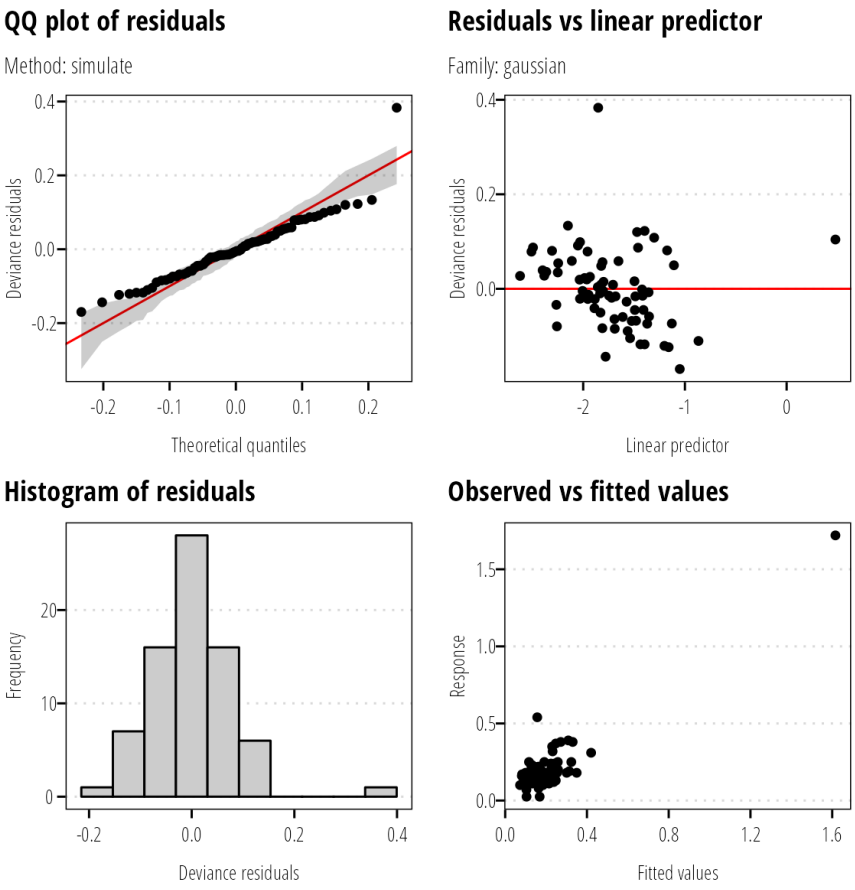

Figure 21: Diagnostic plot for TP model at USGS-08164450.

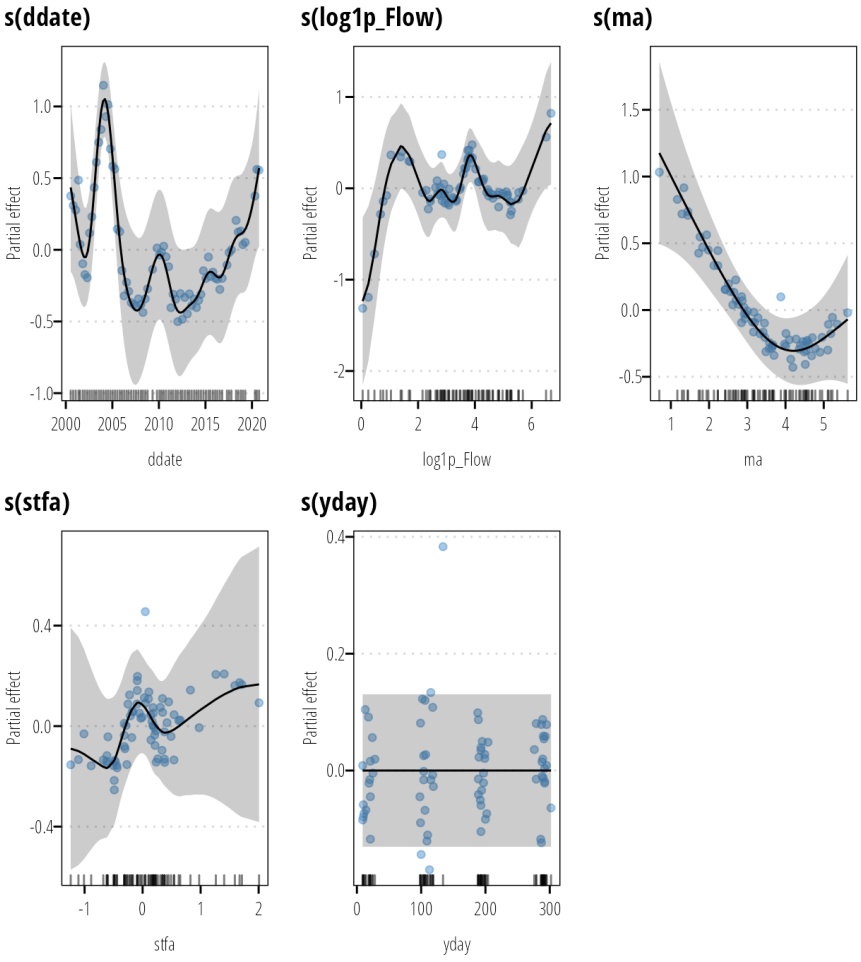

Figure 22: Partial effects of covariates in TP model at USGS-08164450.

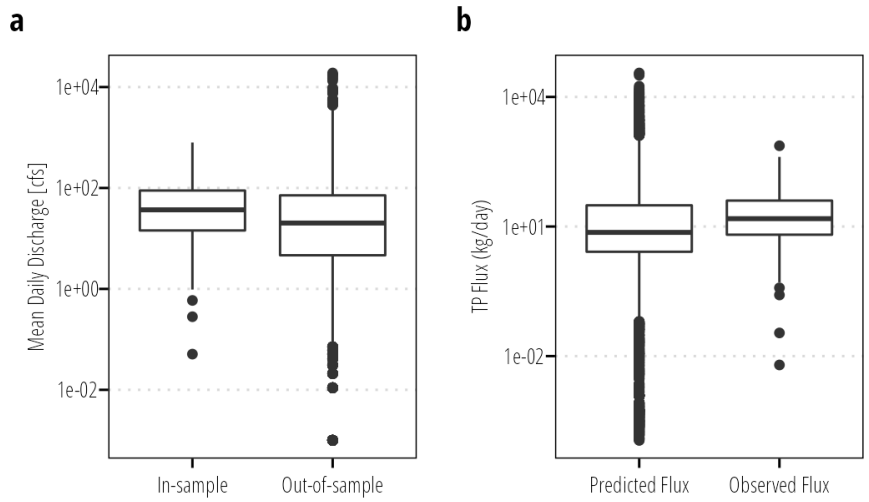

Figure 23: Comparisons of (a) in-sample and out-of-sample mean daily discharge and (b) predicted daily fluxes (for both sampled and non-sampled days) and measured daily fluxes.

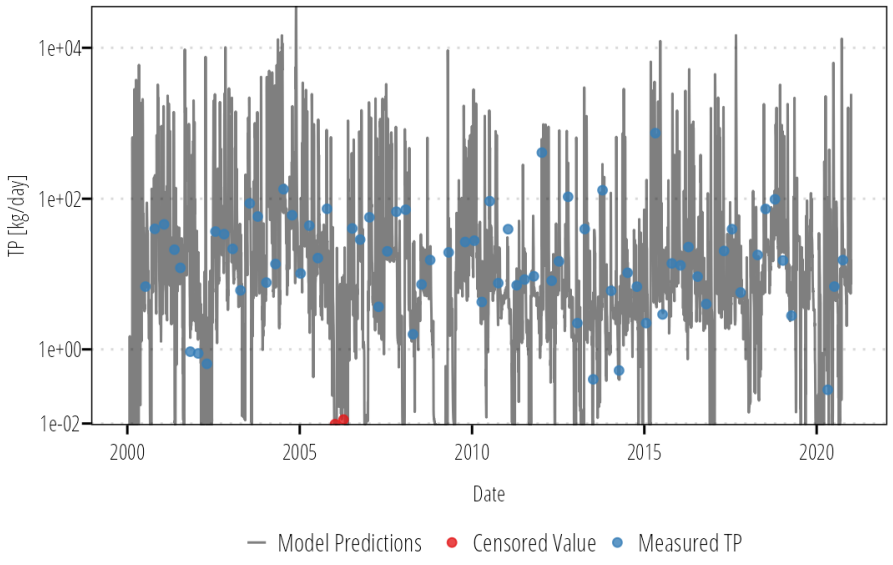

Figure 24: Time series plot of TP model predictions and observed values at USGS-08164450.

2.4 E Mustang Creek nr Louise, USGS-08164504

2.4.1 NO_3

Table 13: NO_3-N GAM summary - E Mustang Creek nr Louise, USGS-08164504.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-0.481	0.159	-3.028				0.004 **
B. smooth terms	s(ddate)				0.000	17	0.000	0.788
	s(yday)				2.557	4	6.590	0.000 ***
	s(log1p(Flow))				2.794	4	3.692	0.001 **
	s(ma)				0.000	5	0.000	0.684
	s(ltfa)				0.000	9	0.000	0.892

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.348, Deviance explained 0.498
-REML : 43.673, Scale est: 1.542, N: 61

Table 14: Summary of goodness-of-fit metrics for 5-fold cross-validation of NO_3-N load GAM at E Mustang Creek nr Louise, USGS-08164504.

Goodness of Fit Metric	Median (IQR)
NSE	0.38 (0.25, 0.41)
R^2	0.54 (0.45, 0.58)
Percent Bias	-46 (-54, -44)

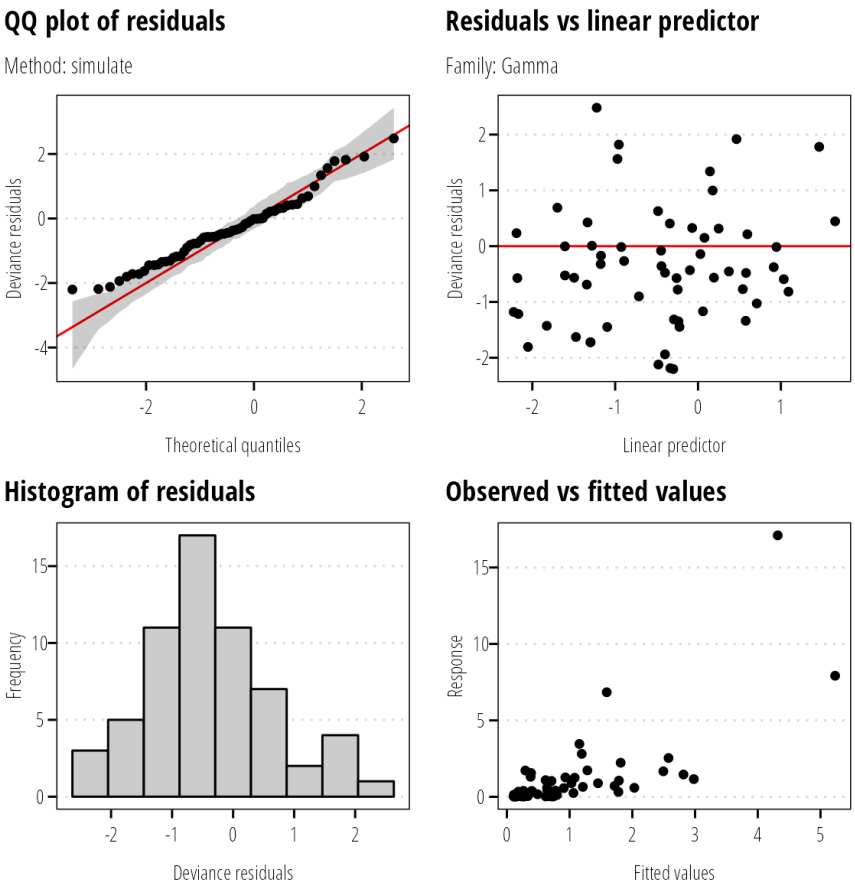

Figure 25: Diagnostic plot for NO_3 model at USGS-08164504

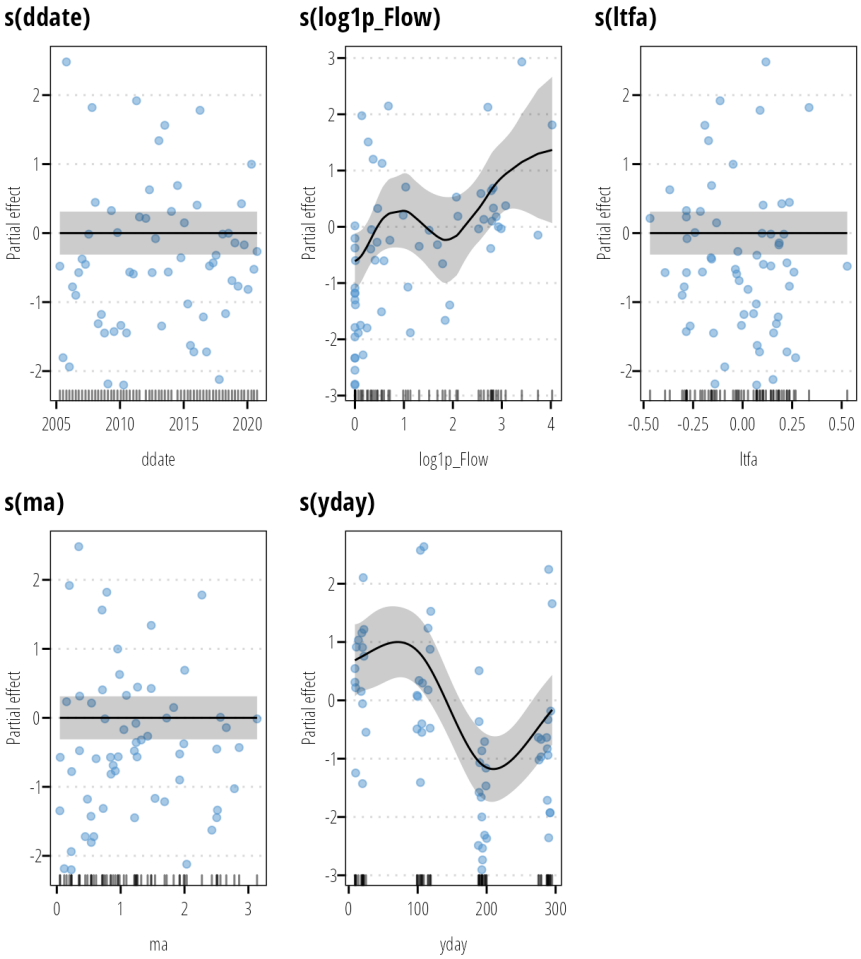

Figure 26: Partial effects of covariates in NO_3-N model at USGS-08164504

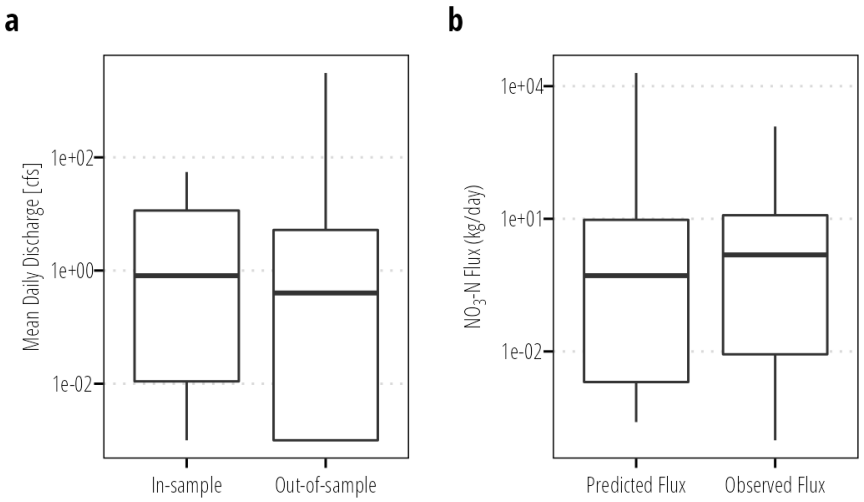

Figure 27: Comparisons of (a) in-sample and out-of-sample mean daily discharge and (b) predicted daily fluxes (for both sampled and non-sampled days) and measured daily fluxes.

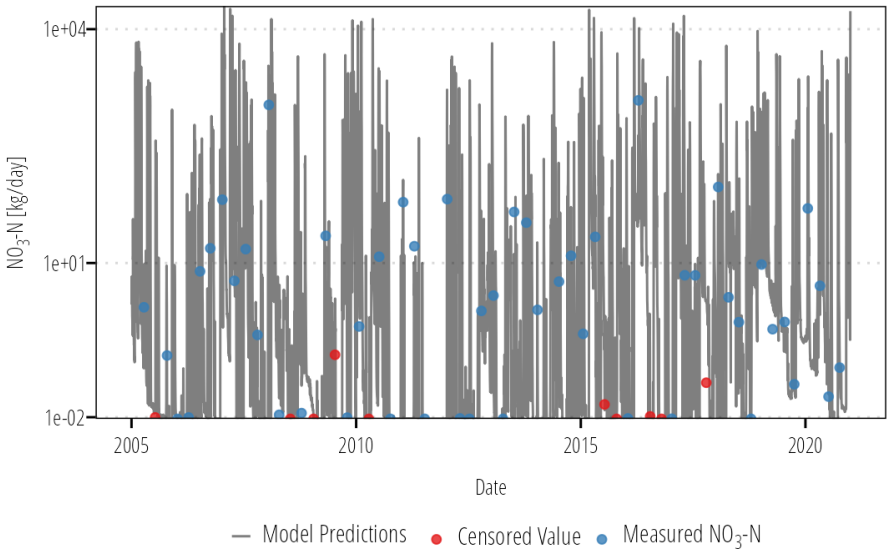

Figure 28: Time series plot of NO_3 model predictions and observed values at USGS-08164504

2.4.2 TP

Table 15: TP GAM summary - E Mustang Creek nr Louise, USGS-08164504.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-1.001	0.081	-12.331				0.000 ***
B. smooth terms	s(ddate)				0.044	17	0.003	0.343
	s(yday)				0.385	8	0.057	0.293
	s(log1p(Flow))				1.642	4	1.416	0.005 **
	s(ma)				1.184	5	0.447	0.086 +
	s(stfa)				1.015	4	0.415	0.117

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.263, Deviance explained 0.246
-REML : -2.438, Scale est: 0.521, N: 79

Table 16: Summary of goodness-of-fit metrics for 5-fold cross-validation of TP load GAM at E Mustang Creek nr Louise, USGS-08164504.

Goodness of Fit Metric	Median (IQR)
NSE	0.851 (0.812, 0.855)
R^2	0.854 (0.839, 0.859)
Percent Bias	-9.2 (-12.7, -7.5)

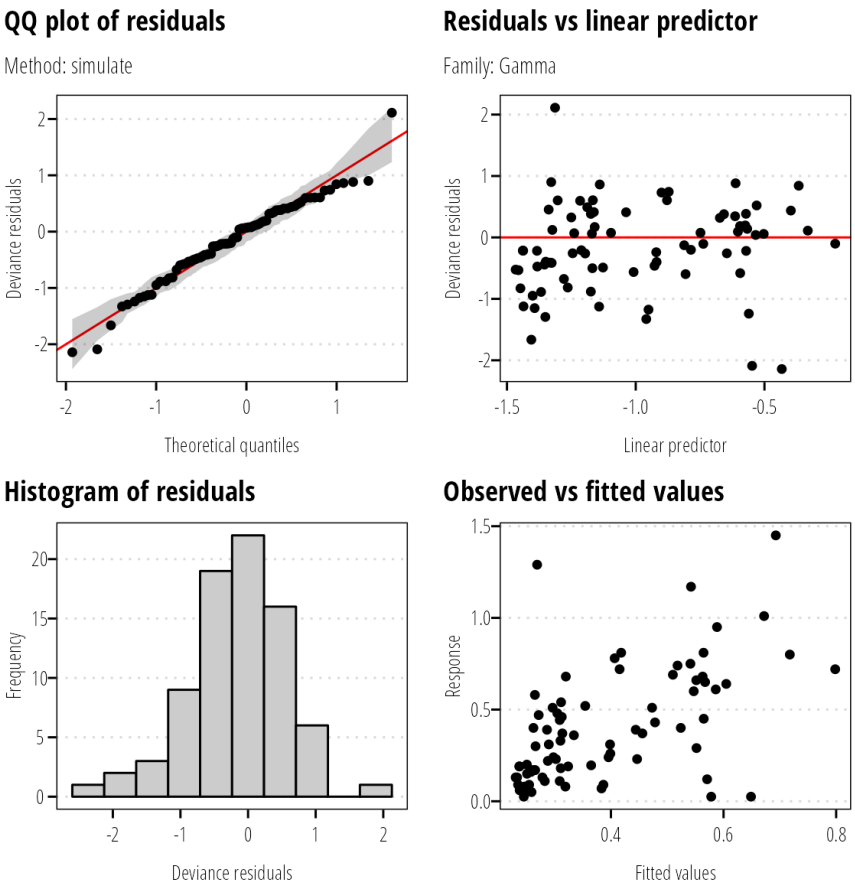

Figure 29: Diagnostic plot for TP model at USGS-08164504

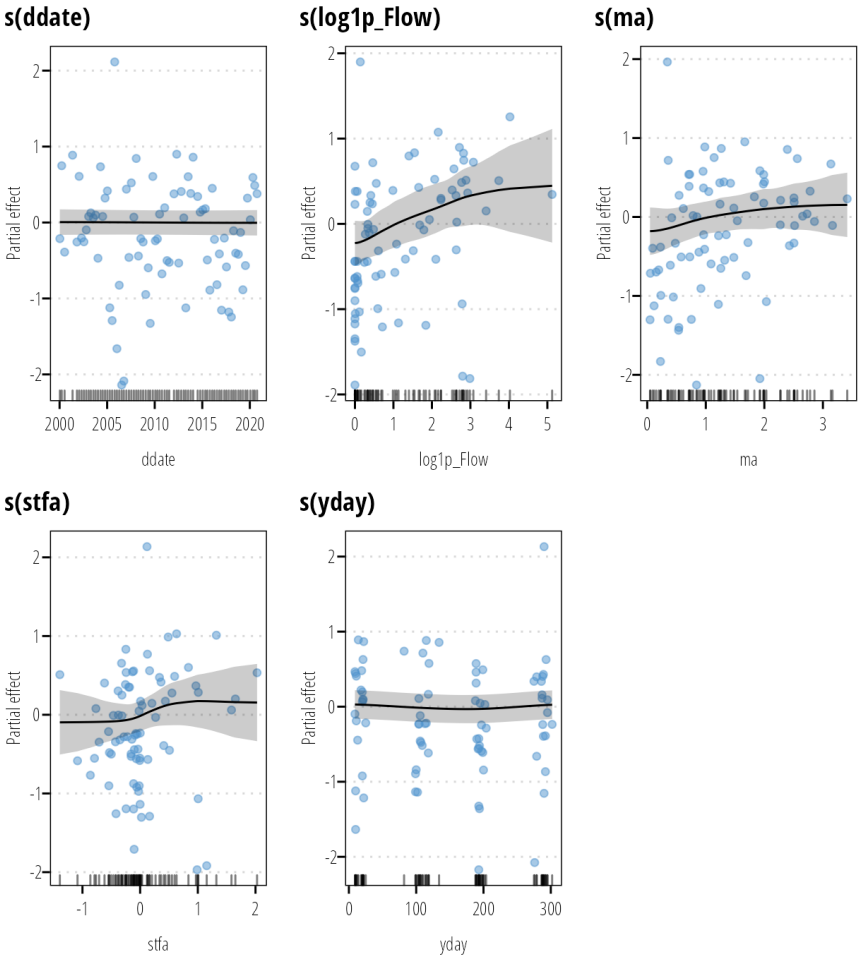

Figure 30: Partial effects of covariates in TP model at USGS-08164504

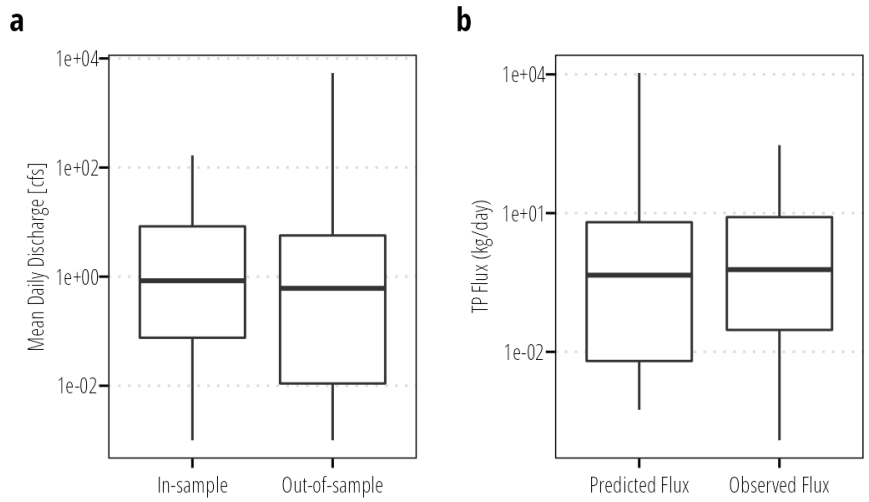

Figure 31: Comparisons of (a) in-sample and out-of-sample mean daily discharge and (b) predicted daily fluxes (for both sampled and non-sampled days) and measured daily fluxes.

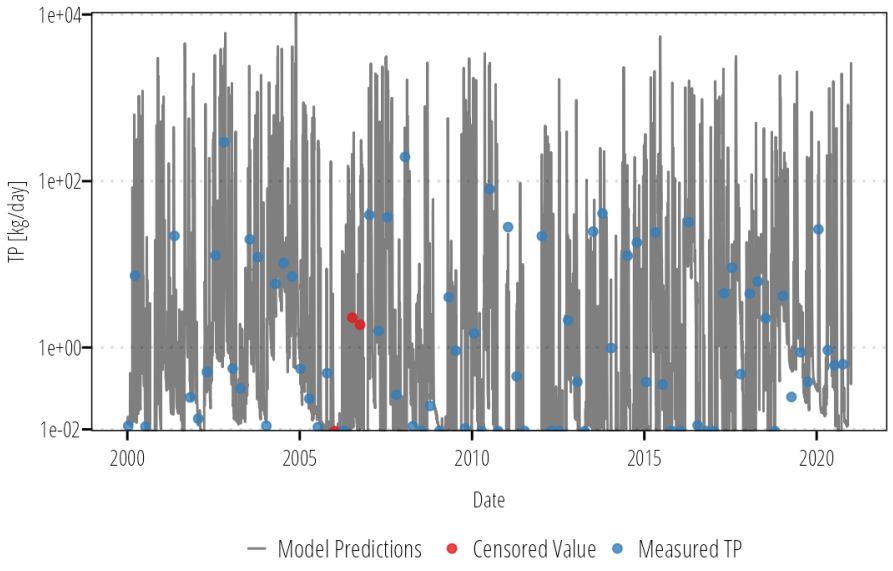

Figure 32: Time series plot of TP model predictions and observed values at USGS-08164504

2.5 W Mustang Creek nr Ganado, USGS-08164503

2.5.1 NO_3

Table 17: NO_3-N GAM summary - W Mustang Creek nr Ganado, USGS-08164503.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-1.244	0.085	-14.678				0.000 ***
B. smooth terms	s(ddate)				1.556	17	0.462	0.006 **
	s(yday)				2.672	4	13.475	0.000 ***
	s(log1p(Flow))				5.453	6	11.500	0.000 ***
	s(ma)				0.118	5	0.024	0.346
	s(ltfa)				2.270	9	0.928	0.007 **

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.510, Deviance explained 0.673
-REML : -3.704, Scale est: 0.453, N: 63

Table 18: Summary of goodness-of-fit metrics for 5-fold cross-validation of NO_3-N load GAM at W Mustang Creek nr Ganado, USGS-08164503.

Goodness of Fit Metric	Median (IQR)
NSE	0.41 (0.28, 0.67)
R^2	0.49 (0.41, 0.70)
Percent Bias	-13 (-18, -5)

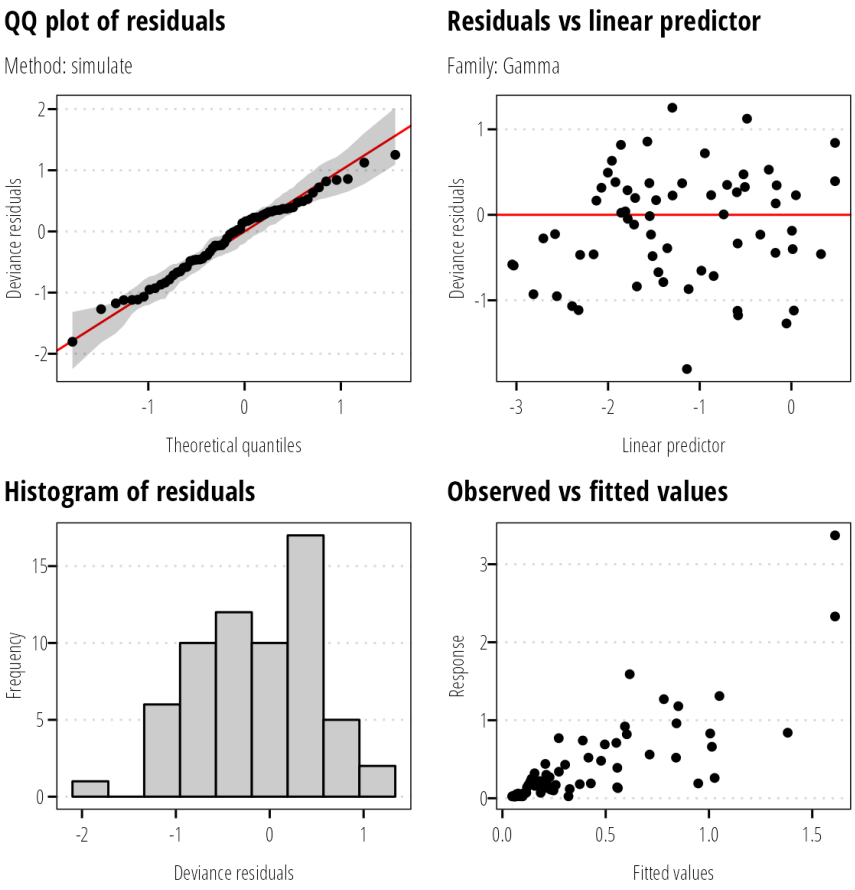

Figure 33: Diagnostic plot for NO_3 model at USGS-08164503

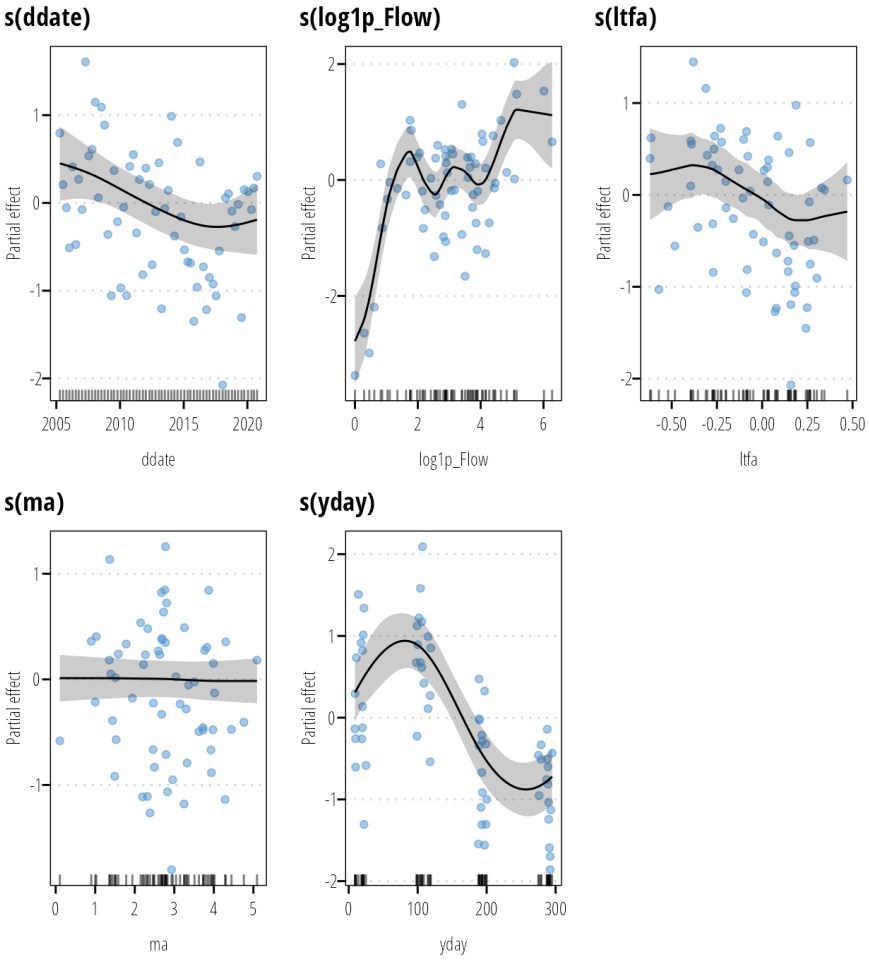

Figure 34: Partial effects of covariates in NO_3-N model at USGS-08164503

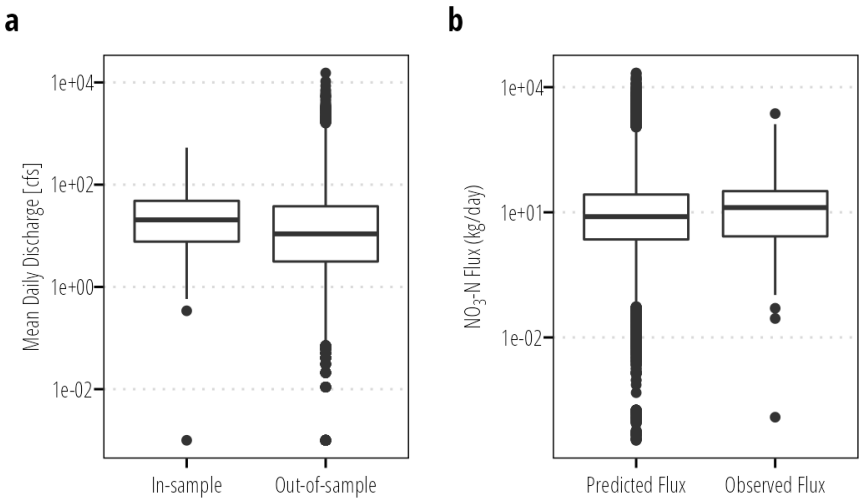

Figure 35: Comparisons of (a) in-sample and out-of-sample mean daily discharge and (b) predicted daily fluxes (for both sampled and non-sampled days) and measured daily fluxes.

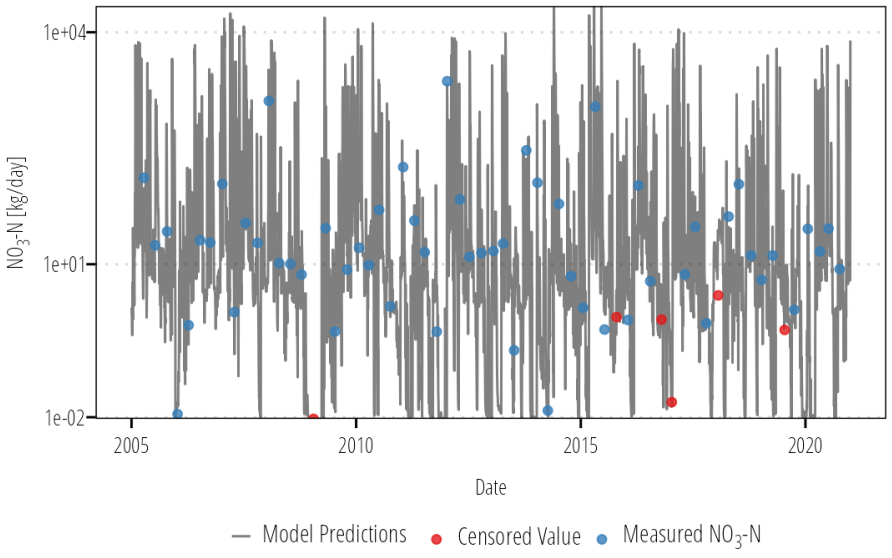

Figure 36: Time series plot of NO_3 model predictions and observed values at USGS-08164503

2.5.2 TP

Table 19: TP GAM summary - W Mustang Creek nr Ganado, USGS-08164503.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-1.150	0.063	-18.148				0.000 ***
B. smooth terms	s(ddate)				2.054	17	0.411	0.025 *
	s(yday)				0.000	4	0.000	0.573
	s(log1p(Flow))				0.000	9	0.000	0.648
	s(stfa)				0.235	5	0.050	0.342
	s(ma)				0.285	5	0.067	0.293

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.0843, Deviance explained 0.142
-REML : -33.755, Scale est: 0.325, N: 81

Table 20: Summary of goodness-of-fit metrics for 5-fold cross-validation of TP load GAM at W Mustang Creek nr Ganado, USGS-08164503.

Goodness of Fit Metric	Median (IQR)
NSE	0.864 (0.838, 0.882)
R^2	0.890 (0.887, 0.896)
Percent Bias	-6.50 (-9.12, -4.70)

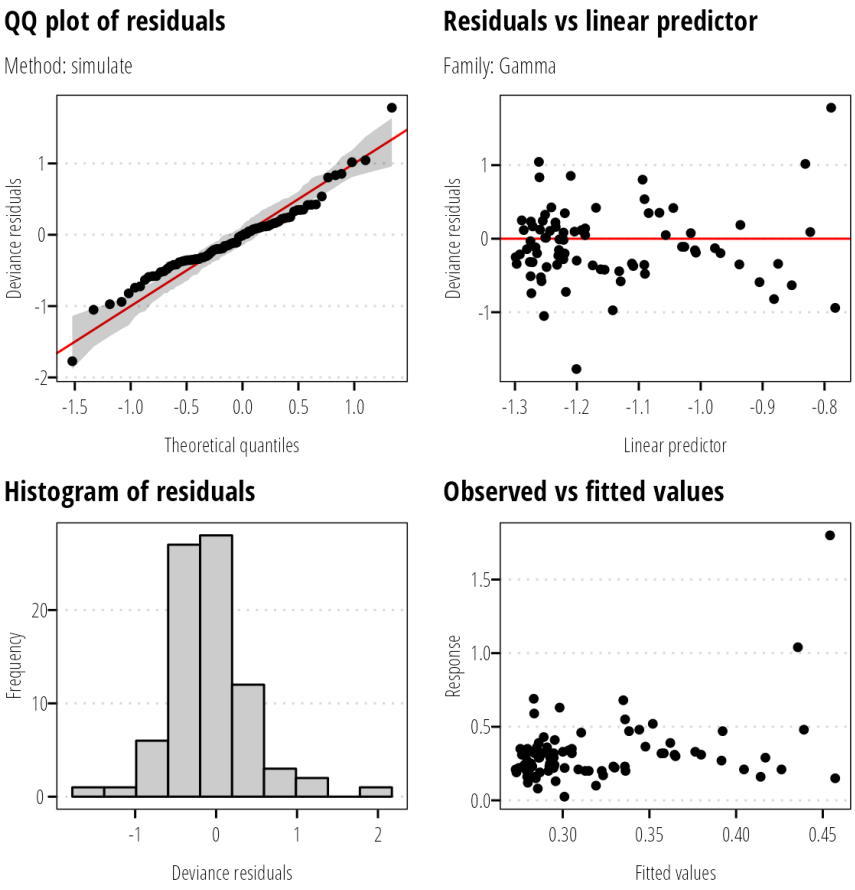

Figure 37: Diagnostic plot for TP model at USGS-08164503

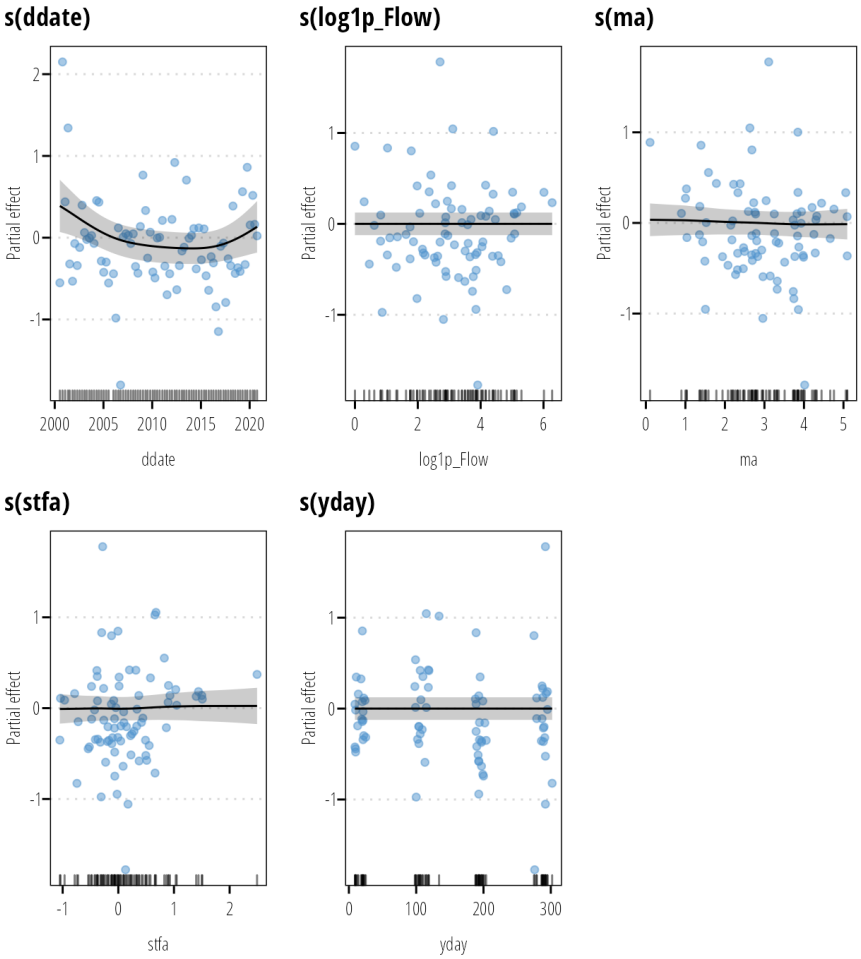

Figure 38: Partial effects of covariates in TP model at USGS-08164503

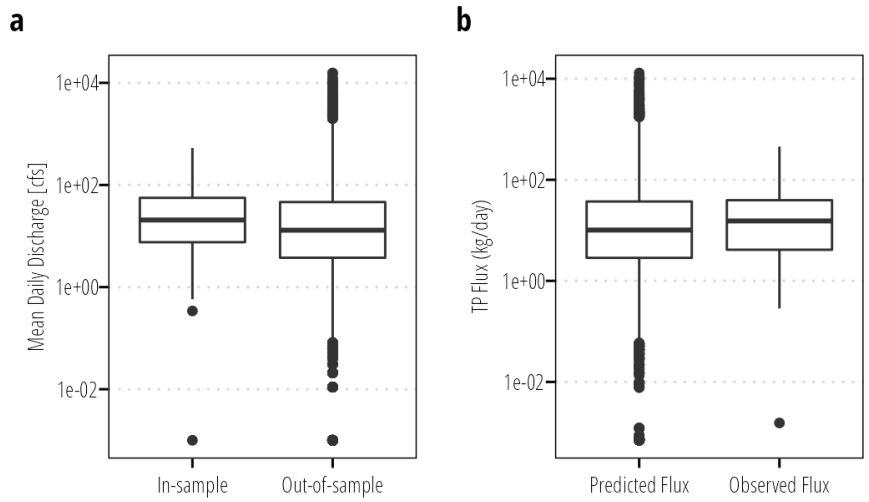

Figure 39: Comparisons of (a) in-sample and out-of-sample mean daily discharge and (b) predicted daily fluxes (for both sampled and non-sampled days) and measured daily fluxes.

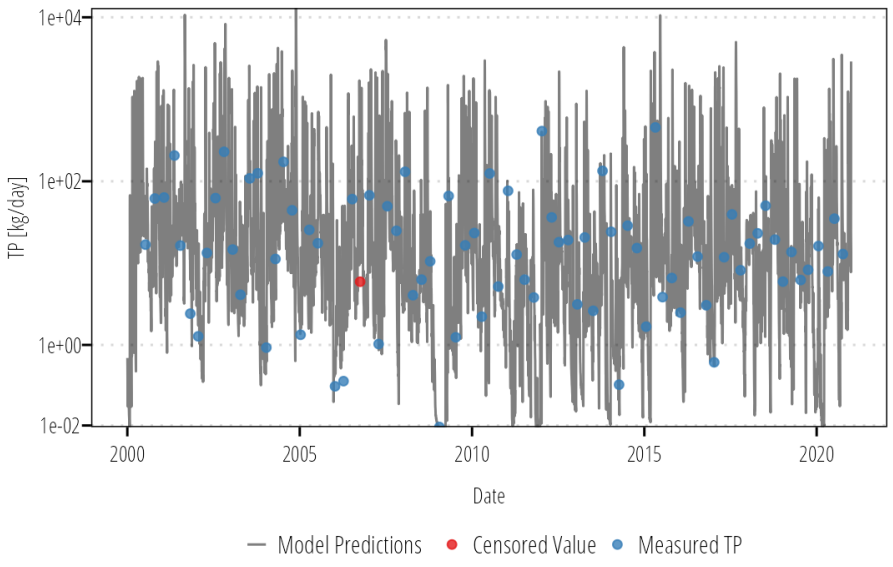

Figure 40: Time series plot of TP model predictions and observed values at USGS-08164503

2.6 Palmetto Bend Dam

2.6.1 NO_3

Table 21: NO_3-N GAM summary - Navidad River at Palmetto Bend Dam.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-1.450	0.087	-16.634				0.000 ***
B. smooth terms	s(ddate)				0.000	9	0.000	0.779
	s(yday)				2.836	8	5.179	0.000 ***
	s(log1p(Inflow))				0.000	4	0.000	0.467
	s(log1p(Flow))				6.058	9	2.712	0.000 ***
	s(ma)				2.665	5	2.101	0.002 **
	s(ltfa)				4.781	9	3.193	0.000 ***

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.746, Deviance explained 0.812
-REML : -15.004, Scale est: 0.017, N: 62

Table 22: Summary of goodness-of-fit metrics for 5-fold cross-validation of NO_3-N load GAM at Palmetto Bend Dam.

Goodness of Fit Metric	Median (IQR)
NSE	0.42 (0.34, 0.46)
R^2	0.60 (0.52, 0.66)
Percent Bias	-43 (-47, -38)

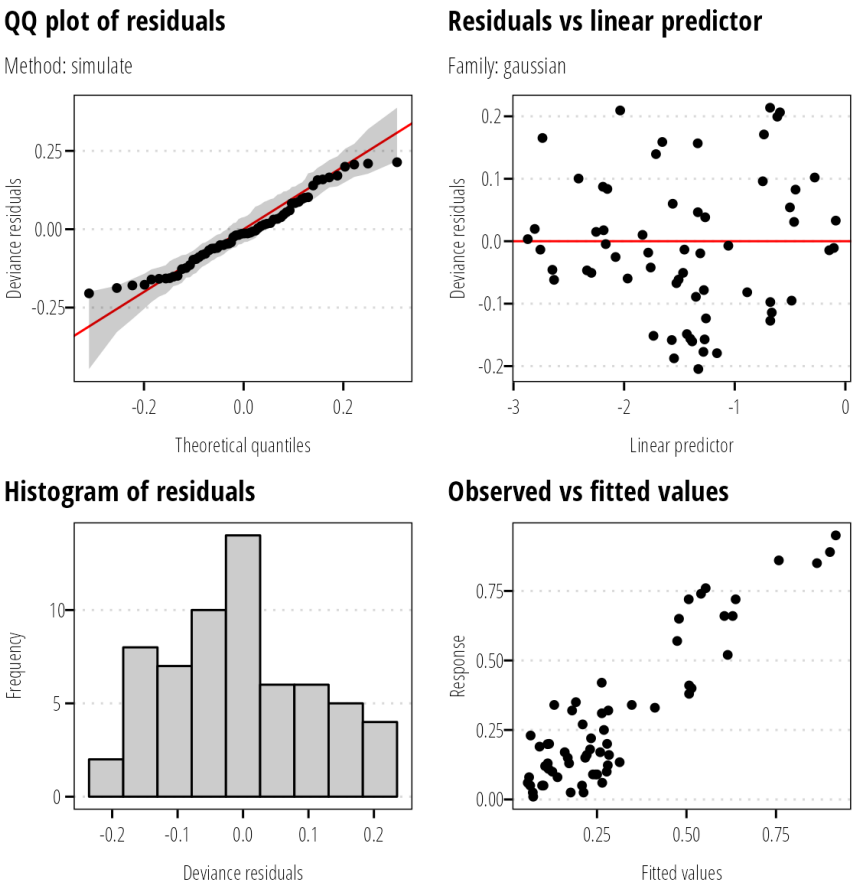

Figure 41: Diagnostic plot for NO_3 model below Palmetto Bend Dam.

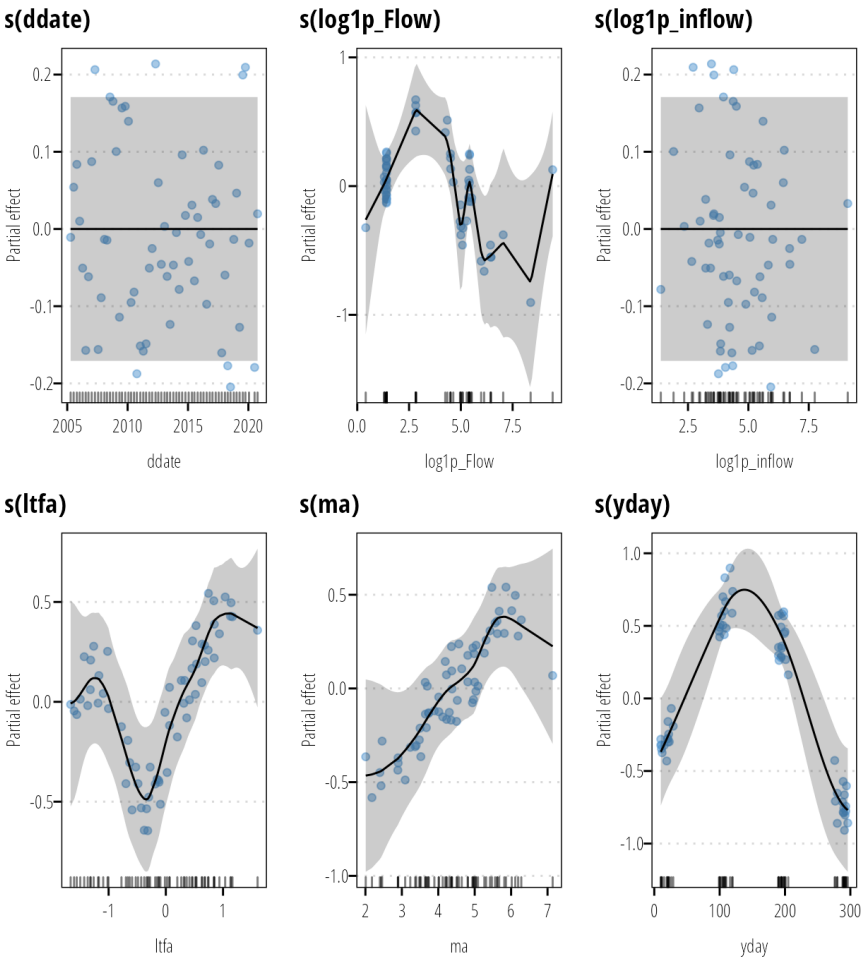

Figure 42: Partial effects of covariates in NO_3-N model below Palmetto Bend Dam.

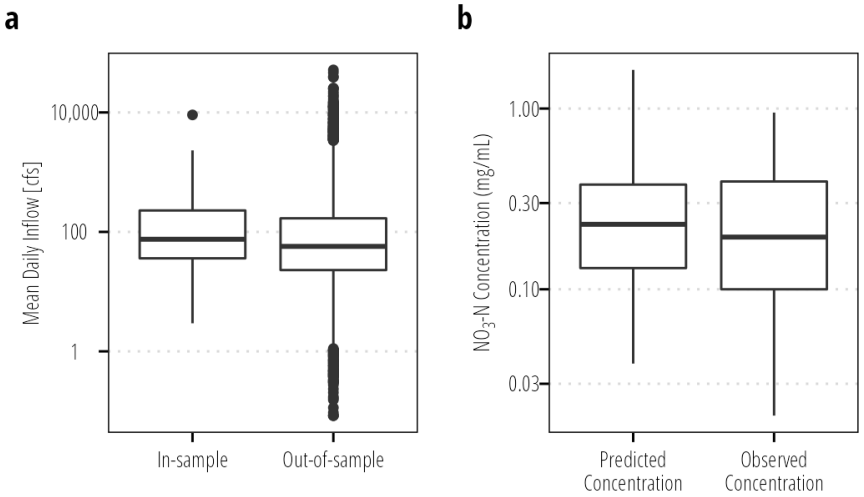

Figure 43: Comparisons of (a) in-sample and out-of-sample mean daily inflows and (b) predicted daily concentration (for both sampled and non-sampled days) and measured daily concentration.

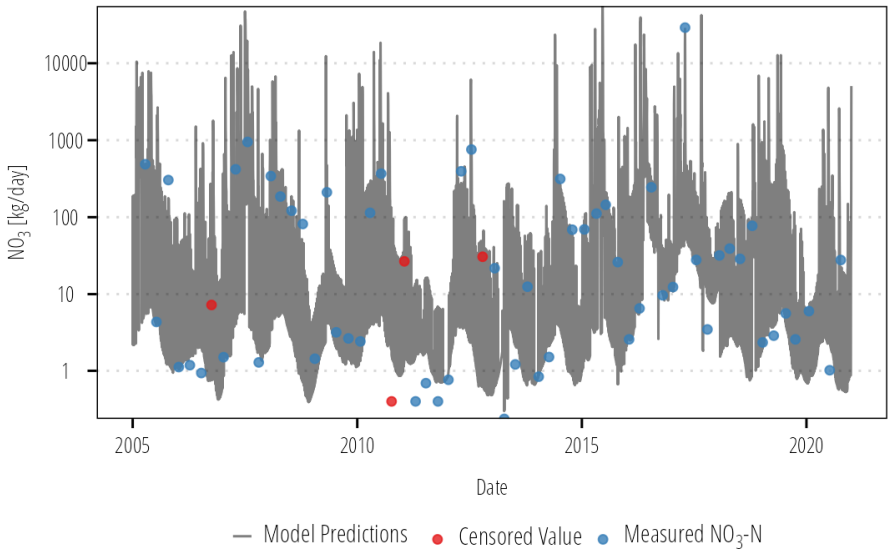

Figure 44: Time series plot of NO_3 model predictions and observed values below Palmetto Bend Dam.

2.6.2 TP

Table 23: TP GAM summary - Navidad River at Palmetto Bend Dam.

Component	Term	Estimate	Std.Error	t-value	edf	ref.df	F-value	p-value[1]
A. parametric coefficients	(Intercept)	-1.624	0.037	-44.377				0.000 ***
B. smooth terms	s(ddate)				3.214	8	1.862	0.001 ***
	s(yday)				1.309	8	0.374	0.088 +
	s(log1p(Inflow))				0.003	9	0.000	0.360
	s(log1p(Flow))				1.104	4	0.561	0.098 +
	s(stfa)				0.000	5	0.000	0.470
	s(ma)				2.262	5	1.669	0.006 **

[1] Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '+' < 0.1
Adjusted R-squared: 0.321, Deviance explained 0.388
-REML : -99.963, Scale est: 0.00403, N: 81

Table 24: Summary of goodness-of-fit metrics for 5-fold cross-validation of TP GAM at Palmetto Bend Dam.

Goodness of Fit Metric	Median (IQR)
NSE	0.877 (0.862, 0.911)
R^2	0.961 (0.956, 0.975)
Percent Bias	-17.6 (-21.1, -12.7)

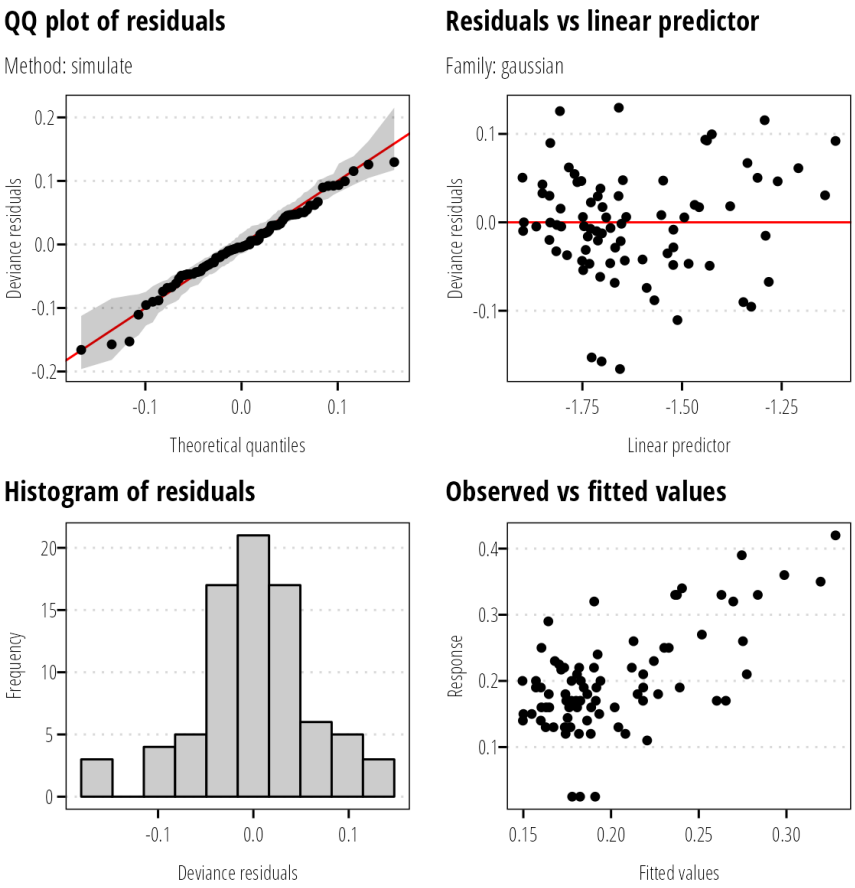

Figure 45: Diagnostic plot for TP model below Palmetto Bend Dam.

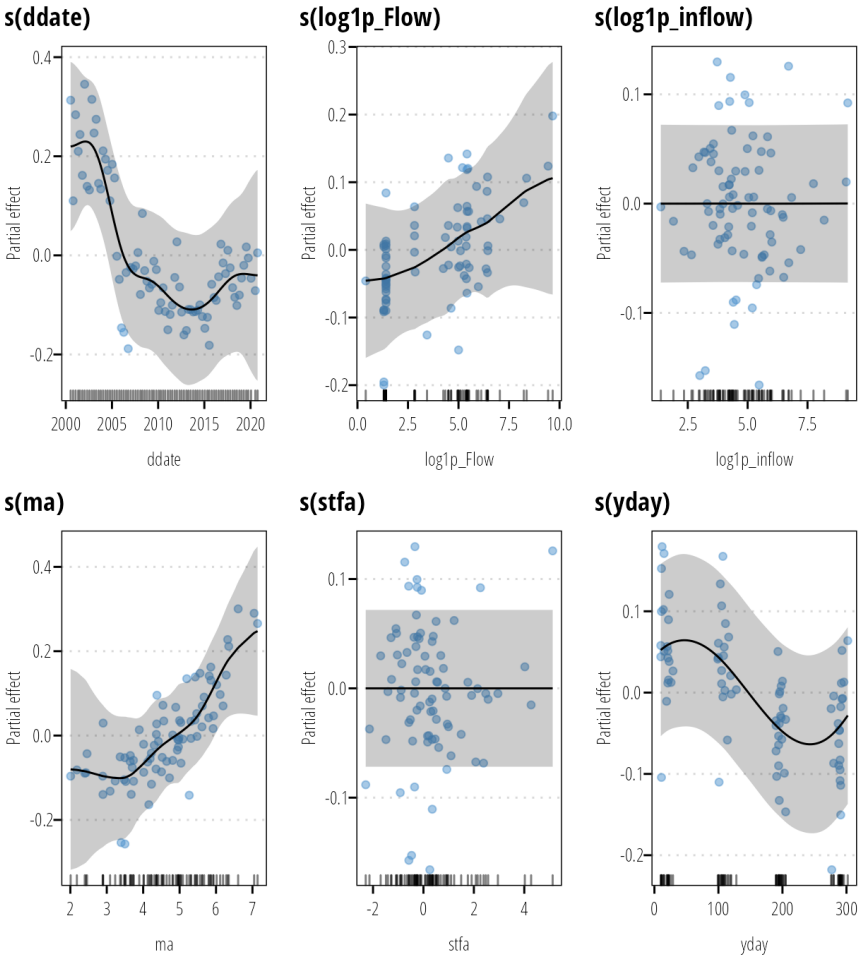

Figure 46: Partial effects of covariates in TP model below Palmetto Bend Dam.

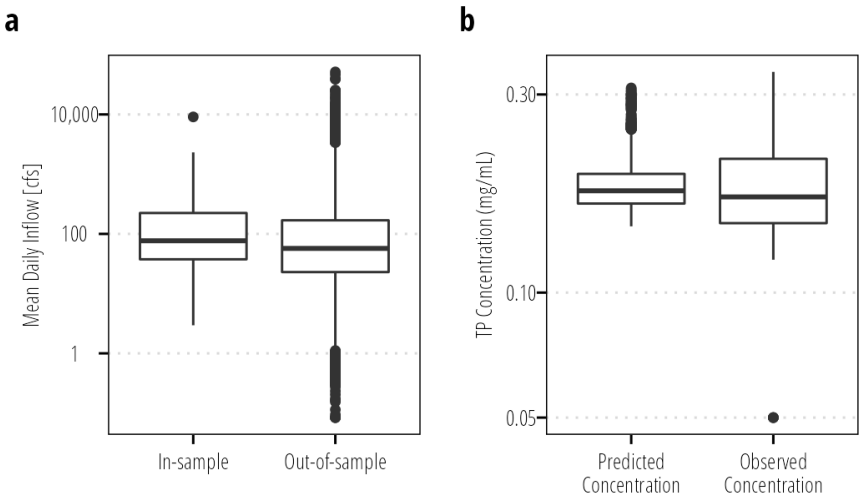

Figure 47: Comparisons of (a) in-sample and out-of-sample mean daily inflows and (b) predicted daily concentration (for both sampled and non-sampled days) and measured daily concentration.

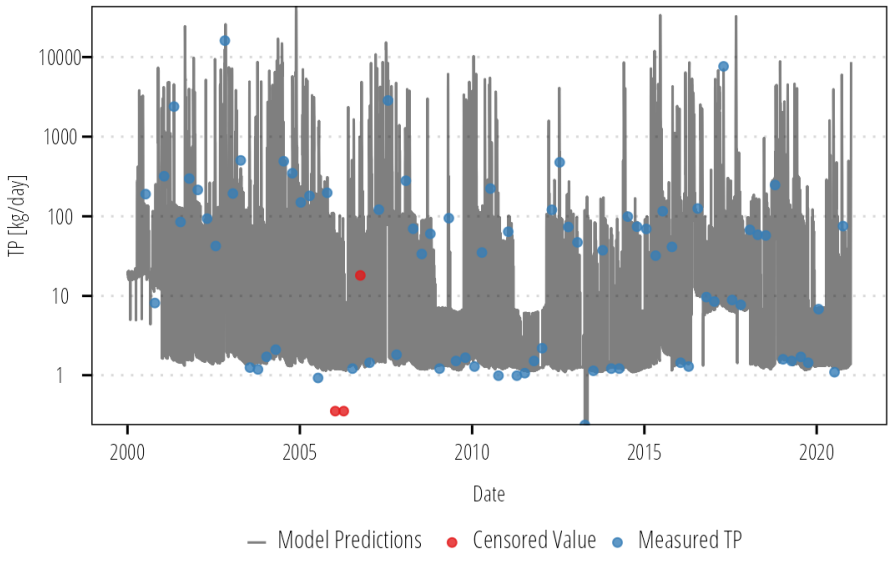

Figure 48: Time series plot of TP model predictions and observed values below Palmetto Bend Dam.

References

Beck, M.W., and Murphy, R.R. 2017. Numerical and qualitative contrasts of two statistical models for water quality change in tidal waters. JAWRA Journal of the American Water Resources Association 53 (1): 197–219. https://doi.org/10.1111/1752-1688.12489.

Bergbusch, N.T., Hayes, N.M., Simpson, G.L., and Leavitt, P.R. 2021. Unexpected shift from phytoplankton to periphyton in eutrophic streams due to wastewater influx. Limnology and Oceanography 66 (7): 2745–61. https://doi.org/10.1002/lno.11786.

Kuhnert, P.M., Henderson, B.L., Lewis, S.E., Bainbridge, Z.T., Wilkinson, S.N., and Brodie, J.E. 2012. Quantifying total suspended sediment export from the Burdekin River catchment using the loads regression estimator tool: REGRESSION ESTIMATOR TOOL FOR POLLUTANT LOADS. Water Resources Research 48 (4). https://doi.org/10.1029/2011WR011080.

McDowell, R.W., Simpson, Z.P., Ausseil, A.G., Etheridge, Z., and Law, R. 2021. The implications of lag times between nitrate leaching losses and riverine loads for water quality policy. Scientific Reports 11 (1): 16450. https://doi.org/10.1038/s41598-021-95302-1.

Robson, B.J., and Dourdet, V. 2015. Prediction of sediment, particulate nutrient and dissolved nutrient concentrations in a dry tropical river to provide input to a mechanistic coastal water quality model. Environmental Modelling & Software 63 (January): 97–108. https://doi.org/10.1016/j.envsoft.2014.08.009.

Zhang, Q., and Ball, W.P. 2017. Improving riverine constituent concentration and flux estimation by accounting for antecedent discharge conditions. Journal of Hydrology 547 (April): 387–402. https://doi.org/10.1016/j.jhydrol.2016.12.052.