

Monthly Nitrate Concentration and Load Estimation in the Choptank River (Chesapeake Bay Watershed) Using Regression and Machine Learning Models

Banafsheh Faramarzpour Darzini¹, Fatemeh Khamespanah²

¹ M.Sc. in Industrial Engineering, Head of Applied Research Group, Sistan and Baluchestan Regional Water Company, Zahedan, Iran.

² M.Sc. in Surveying Engineering (GIS), GIS Expert, Sistan and Baluchestan Regional Water Company, Zahedan, Iran.
Corresponding author email: (Faramarzpour.banafsheh@gmail.com)

Received: 13 November 2025

Accepted: 06 Desember 2025

Available online: 05 January 2026

Abstract: Riverine nitrate loads are usually estimated from sparse concentration samples and continuous discharge records with regression models, and machine learning (ML) is increasingly proposed as an alternative. This study compared the seven-parameter LOADEST regression, an extended multiple linear regression (MLR), random forest (RF), extreme gradient boosting (XGBoost), support vector regression (SVR) and a multilayer perceptron (MLP) for estimating monthly nitrate-plus-nitrite concentrations and loads in the Choptank River near Greensboro, Maryland (USGS 01491000; 292.7 km²). The data comprised 606 samples and daily discharge for water years 1980–2011, aggregated to 345 months with paired observations. Predictors were monthly discharge statistics, a digital-filter baseflow index (BFI), time and seasonal terms. Concentrations averaged 1.20 mg L⁻¹ as N, showed a weak negative relation with discharge and a winter maximum, and increased significantly over time (seasonal Kendall $Z = 6.56$, $p < 0.001$; about 1% per year, also after flow adjustment). In eight-fold blocked cross-validation, all models explained only a modest part of the concentration variance (NSE = 0.21–0.38; RMSE = 0.25–0.29 mg L⁻¹). SVR performed best, followed by MLP and XGBoost, and all ML models outperformed LOADEST. In a forward test (2005–2011), the tree ensembles underestimated the rising concentrations by 11–13% because they cannot extrapolate the time trend, whereas MLP (NSE = 0.30) and MLR (0.28) were the most robust. Permutation importance ranked seasonality and time above discharge, and partial dependence showed higher concentrations under baseflow-dominated conditions. Annual loads (mean 134 t N yr⁻¹, 4.6 kg ha⁻¹ yr⁻¹) from all models agreed closely with the WRTDS reference (NSE = 0.81–0.92) but were 6–8% higher because monthly aggregation ignores the within-month covariance of concentration and discharge. Monthly load estimates are thus controlled mainly by discharge, whereas concentration trends require models that can represent time explicitly.

Keywords: Nitrate, Load estimation, LOADEST, WRTDS, Machine learning, Choptank River

1. Introduction

Excess nitrogen delivered by rivers is a leading cause of coastal eutrophication and hypoxia. In the Chesapeake Bay, nutrient loading from the watershed is regulated through a total maximum daily load programme, and progress is judged largely from riverine concentration and load records (USEPA, 2010). Loads cannot be measured continuously in most rivers. They are estimated by combining discrete concentration samples with continuous discharge records through a statistical model, so the choice of model directly affects the reported loads and trends.

Rating-curve regressions relating the logarithm of concentration to discharge, season and time have been the standard approach for decades (Cohn et al., 1989). The USGS LOADEST program implements a family of such models (Runkel et al., 2004), and the Weighted Regressions on Time, Discharge, and Season (WRTDS) method estimates these relationships locally to capture changing concentration–discharge behaviour (Hirsch et al., 2010). WRTDS, implemented in the EGRET package (Hirsch and De Cicco, 2015), is now the reference method for trend reporting in the Chesapeake Bay watershed. Because these models are fitted in log

space, back-transformed predictions must be bias-corrected, for example with the smearing estimator (Duan, 1983).

Machine learning (ML) methods such as random forests and neural networks can represent non-linear interactions between discharge, season and antecedent conditions without a prescribed functional form (Breiman, 2001; Tyralis et al., 2019). They have been used to predict nitrogen and phosphorus concentrations from high-frequency sensor data (Harrison et al., 2021) and water quality at continental scales (Zhi et al., 2021). Few studies, however, have compared ML with established load regression models on long, sparsely sampled records, and even fewer have asked whether ML models can reproduce the long-term trends that load programmes are designed to detect.

The Choptank River on the Delmarva Peninsula is a well-studied agricultural watershed in which a large fraction of streamflow is groundwater discharge. Nitrate in this water can be years to decades old (Lindsey et al., 2003; Sanford and Pope, 2013), so legacy nitrogen may delay the response of stream concentrations to management (Chang et al., 2021). Concentrations in such catchments often vary little with discharge, a

pattern described as chemostatic behaviour (Godsey et al., 2009; Basu et al., 2010).

The objectives of this study were to (i) characterise the seasonal, discharge-related and long-term behaviour of monthly nitrate concentrations in the Choptank River from 1980 to 2011; (ii) compare LOADEST, MLR and four ML models (RF, XGBoost, SVR and MLP) for monthly concentration estimation under interpolation (blocked cross-validation) and extrapolation (forward test); (iii) interpret the ML models; and (iv) compare the resulting annual loads with the published WRTDS estimates.

2. Materials and Methods

2.1. Study Area and Data

The Choptank River near Greensboro, Maryland (USGS station 01491000; 38.997°N, 75.786°W) drains 292.7 km² of the Atlantic Coastal Plain upstream of the Choptank estuary, a tributary of the Chesapeake Bay (Figure 1). The watershed is flat, is dominated by agriculture, and is underlain by permeable sediments that support a large baseflow component (Lindsey et al., 2003). Daily mean discharge (Q , m³ s⁻¹) for 1 October 1979 – 30 September 2011 (water years 1980–2011) and 606 filtered nitrate-plus-nitrite samples (USGS parameter 00631, mg L⁻¹ as N; hereafter nitrate) were taken from the Choptank example dataset of the EGRET package (Hirsch and De Cicco, 2015). This dataset also contains WRTDS daily concentration and flux estimates, which were used as the load reference. One sample was censored and was represented by its interval midpoint, following EGRET. Discharge was complete, and 213 of the 11,688 daily values were flagged as estimated.

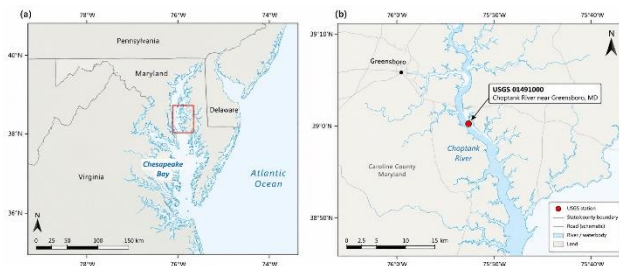


Figure 1. Location of the Choptank River monitoring station (USGS 01491000) in the Chesapeake Bay region (schematic coastline)

Monthly means of daily discharge and of all samples collected in a month were computed. This gave 384 monthly discharge values, of which 347 had at least one concentration sample (202 months with one sample, 93 with two and 52 with three or more). After removal of the first months, for which antecedent discharge was undefined, 345 paired months were used for modelling (Table 1).

Table 1. Characteristics of the monitoring data (water years 1980–2011)

Item	Value
Station	Choptank River near Greensboro, MD (USGS 01491000)
Drainage area	292.7 km ²
Daily discharge records	11,688 (213 estimated)
Nitrate samples	606 (1 censored)
Months with samples / used	347 / 345
Mean (min–max) monthly Q	4.10 (0.16–23.40) m ³ s ⁻¹
Mean annual runoff	442 mm
Long-term baseflow index	0.48 (daily filter)
Mean (min–max) monthly C	1.20 (0.38–2.36) mg L ⁻¹ as N

2.2. Baseflow Separation and Predictors

Baseflow b was separated from daily discharge with the recursive digital filter of Lyne and Hollick (1979), applied in three forward–backward–forward passes with $\alpha = 0.925$ (Nathan and McMahon, 1990):

$$q_{f,i} = \alpha q_{f,i-1} + \frac{1+\alpha}{2}(Q_i - Q_{i-1}), \quad b_i = Q_i - \max(q_{f,i}, 0) \quad (1)$$

where $q_{f,i}$ is the quickflow. The monthly baseflow index (BFI) is the ratio of mean baseflow to mean discharge in the month. Nine predictors were derived for each month (Table 2): the logarithm of monthly mean discharge ($\ln Q$), of the previous month's mean discharge and of the antecedent three-month mean discharge; the monthly change in $\ln Q$; the logarithm of the maximum daily discharge; BFI; decimal time t ; and $\sin(2\pi t)$ and $\cos(2\pi t)$.

Table 2. Predictors of the monthly nitrate models

Predictor	Description
$\ln Q$	Log of monthly mean discharge
$\ln Q(t-1)$	Log of previous-month mean discharge
$\ln \bar{Q}(3 \text{ mo})$	Log of mean discharge over the three preceding months
$\Delta \ln Q$	$\ln Q - \ln Q(t-1)$
$\ln Q_{\max}$	Log of maximum daily discharge in the month
BFI	Monthly baseflow index (Lyne–Hollick filter)
t	Decimal time (years)
$\sin, \cos$	$\sin(2\pi t), \cos(2\pi t)$

2.3. Regression and Machine Learning Models

The LOADEST seven-parameter model (Cohn et al., 1989; Runkel et al., 2004) was fitted by ordinary least squares:

$$\ln C = a_0 + a_1 q + a_2 q^2 + a_3 s + a_4 c + a_5 \tau + a_6 \tau^2 + \varepsilon \quad (2)$$

where $q = \ln Q - \text{mean}(\ln Q)$ is centred log discharge, τ is centred decimal time, and $s = \sin(2\pi t)$ and $c = \cos(2\pi t)$. An extended MLR model used all nine predictors linearly. The ML models were RF (Breiman, 2001), XGBoost (Chen and Guestrin, 2016), SVR with a radial basis function kernel (Vapnik, 1995) and an MLP trained with the L-BFGS algorithm (Rumelhart et al., 1986). All models were fitted to $\ln C$, and their predictions were back-transformed with Duan's (1983)

smearing factor estimated from the training residuals $e_{\{i\}}$:

$$\hat{C} = \exp(\ln \hat{C}) \cdot \frac{1}{n} \sum_{i=1}^n \exp(e_i) \quad (3)$$

The hyperparameters of the ML models were selected by grid search (Table 3) with five-fold cross-validation grouped by water year. Inputs were standardised, and scikit-learn (Pedregosa et al., 2011) and xgboost were used.

Table 3. Hyperparameter search spaces and values selected for the forward-test models

Model	Search space	Selected (forward test)
RF	trees = 500; max depth {4, 8, none}; min leaf {2, 5, 10}; max features {0.33, 0.66, 1}	8 / 5 / 0.66
XGBoost	trees {200, 500}; learning rate {0.02, 0.05}; depth {2, 3, 5}; min child weight {1, 5}	200 / 0.02 / 3 / 5
SVR	C {0.1, 1, 10}; γ {0.01, 0.05, 0.1, 0.5}; ϵ {0.01, 0.05, 0.1}	1 / 0.05 / 0.01
MLP	hidden {(4), (8), (8, 4)}; α {0.01, 0.1, 1, 10}; activation {tanh, ReLU}	(8, 4) / 1.0 / relu

2.4. Validation, Trend Analysis and Load Estimation

Two validation schemes were used. In *blocked cross-validation*, the record was divided into eight blocks of four water years, and each block was predicted by models trained on the other seven. This tests interpolation within the observed period. In the *forward test*, models were trained on water years 1980–2004 and applied to 2005–2011 (83 months). This tests extrapolation into a later period. Performance was measured by R^2 , RMSE, MAE, NSE and PBIAS (Nash and Sutcliffe, 1970; Moriasi et al., 2007).

Monotonic trends in monthly concentration were tested with the seasonal Kendall test (Hirsch et al., 1982):

$$S_k = \sum_{g=1}^{12} \sum_{i < j} \text{sgn}(x_{g,j} - x_{g,i}) \quad (4)$$

where $x_{\{g,i\}}$ is the value for month g in year i . The test was applied to raw concentrations and to flow-adjusted residuals obtained from a LOWESS fit of $\ln C$ on $\ln Q$ (span 0.5). The same test was applied to $\ln Q$. Monthly loads L (kg month^{-1}) were computed for all 381 months with complete predictors from each model fitted to the full record:

$$L_m = 86.4 \hat{C}_m \bar{Q}_m N_m \quad (5)$$

where $N_{\{m\}}$ is the number of days in the month and 86.4 converts $\text{mg L}^{-1} \times \text{m}^3 \text{s}^{-1}$ to kg d^{-1} . Annual loads for water years 1981–2011 were compared with the sums of the WRTDS daily fluxes. Model interpretation used permutation importance (50 repetitions) and partial dependence plots.

3. Results

3.1. Concentration Patterns and Trends

Monthly nitrate concentrations ranged from 0.38 to 2.36 mg L^{-1} as N (mean 1.20, CV 0.27). They were much less variable than discharge (CV 0.95) and nearly symmetric (skewness 0.19) (Table 4; Figure 2). Concentrations were highest in winter (January 1.43, February 1.40 and December 1.38 mg L^{-1}) and lowest in April (0.97 mg L^{-1}), whereas discharge peaked in March (Figure 3). The concentration–discharge relation was weak and negative (log–log slope -0.053, $r = -0.19$, $p < 0.001$). Concentrations were positively correlated with the monthly BFI ($r = 0.35$), which points to a baseflow-borne nitrate source.

Table 4. Descriptive statistics of monthly nitrate concentration and discharge

Variable	N	Min	Max	Mean	Median	SD	CV	Skew
C (mg L^{-1})	34 5	0.3 8	2.36	1.20	1.20	0.3 2	0.2 7	0.19
Q ($\text{m}^3 \text{s}^{-1}$)	38 4	0.1 6	23.4 0	4.10	2.92	3.9 0	0.9 5	1.72

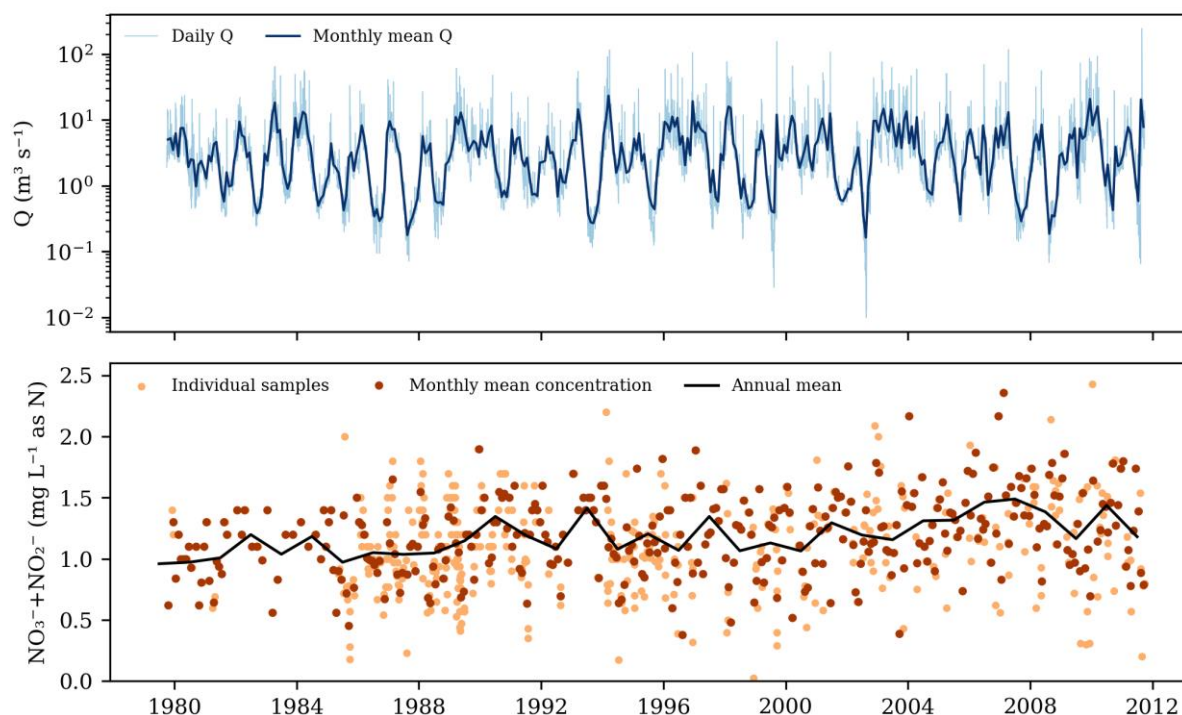


Figure 2. Daily and monthly discharge (top) and nitrate concentrations of individual samples, monthly means and annual means (bottom), water years 1980–2011

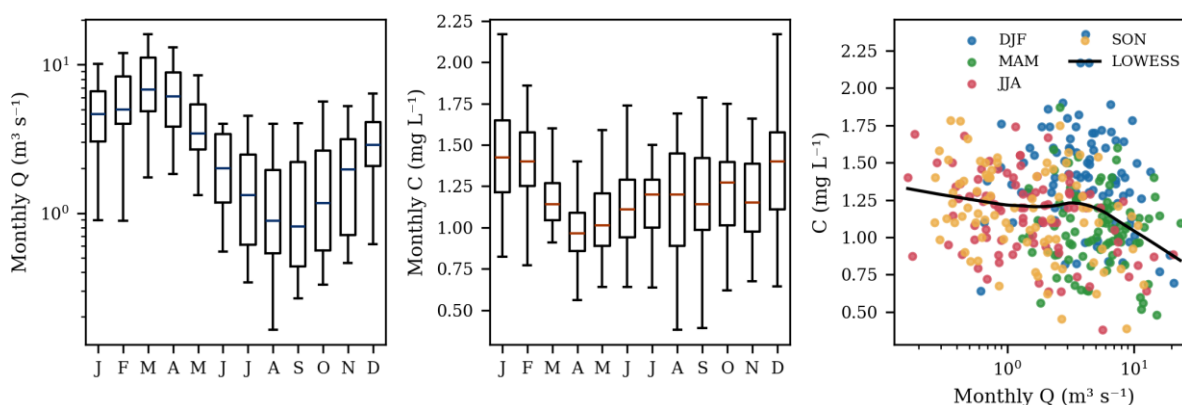


Figure 3. Monthly distributions of discharge and nitrate concentration, and the concentration–discharge relation by season with a LOWESS fit

Nitrate increased steadily over the record (Figure 2; Table 5). Five-year mean concentrations rose from 1.06 mg L^{-1} (1980–1984) to 1.36 mg L^{-1} (2005–2009). The seasonal Kendall test found a highly significant upward trend in raw concentrations ($Z = 6.56$, $p < 0.001$; Sen slope $0.0117 \text{ mg L}^{-1} \text{ yr}^{-1}$, or 0.97% of the median per year). The trend remained after flow adjustment ($Z = 7.33$, $1.03\% \text{ yr}^{-1}$; Theil–Sen slope of all residuals $0.95\% \text{ yr}^{-1}$; Figure 4). The increase is therefore not an artefact of changing discharge, which had only a weak, non-significant upward tendency ($Z = 1.81$, $p = 0.07$).

Table 5. Seasonal Kendall trend tests (water years 1980–2011)

Series	Z	p	Sen slope	Change (% yr ⁻¹)
C (mg L ⁻¹)	6.56	<0.001	$0.0117 \text{ mg L}^{-1} \text{ yr}^{-1}$	0.97
Flow-adjusted ln C	7.33	<0.001	0.0102 yr^{-1}	1.03
ln Q	1.81	0.070	0.0082 yr^{-1}	–

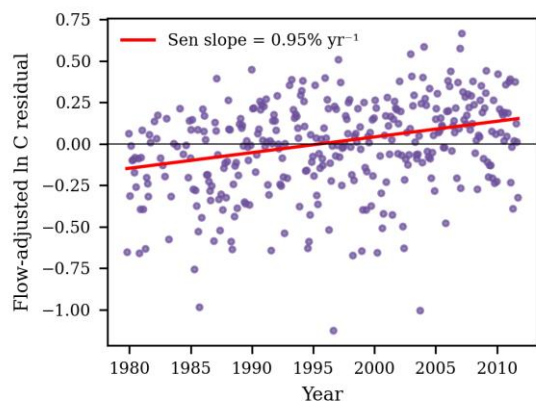


Figure 4. Flow-adjusted nitrate residuals and their Theil–Sen trend

3.2. Model Performance

Table 6. Performance of the concentration models (C in mg L^{-1} as N)

Scheme	Model	Train NSE	R^2	RMS E	MA E	NS E	PBIAS (%)
Blocked CV	LOADEST	0.29	0.23	0.285	0.228	0.21	+0.7
	MLR	0.39	0.33	0.262	0.211	0.33	+0.1
	RF	0.79	0.34	0.262	0.210	0.33	-1.5
	XGBoost	0.82	0.36	0.256	0.206	0.36	-0.1
	SVR	0.58	0.38	0.253	0.199	0.37	+0.2
	MLP	0.56	0.38	0.256	0.202	0.36	+0.8
Forward test	LOADEST	0.23	0.27	0.292	0.229	0.20	-6.1
	MLR	0.35	0.34	0.277	0.224	0.28	-5.6
	RF	0.70	0.41	0.307	0.248	0.12	-13.1
	XGBoost	0.72	0.37	0.296	0.241	0.18	-10.7
	SVR	0.56	0.39	0.291	0.240	0.21	-10.0
	MLP	0.57	0.37	0.273	0.214	0.30	-1.8

Blocked CV: 8 folds of 4 water years (345 months). Forward test: training WY1980–2004, test WY2005–2011 (83 months). Bold: best NSE per scheme.

Under blocked cross-validation, all models explained only a moderate part of the monthly concentration variance (Table 6; Figure 5). SVR performed best ($\text{NSE} = 0.37$, $\text{RMSE} = 0.253 \text{ mg L}^{-1}$), followed by MLP (0.36) and XGBoost (0.36). RF (0.33) and the extended MLR (0.33) were nearly identical, and LOADEST was the weakest (0.21). All models were nearly unbiased ($|\text{PBIAS}| < 2\%$). RF and XGBoost fitted the training folds much more closely ($\text{NSE} 0.79\text{--}0.82$) than they predicted the held-out blocks, which indicates overfitting of the sample-level noise.

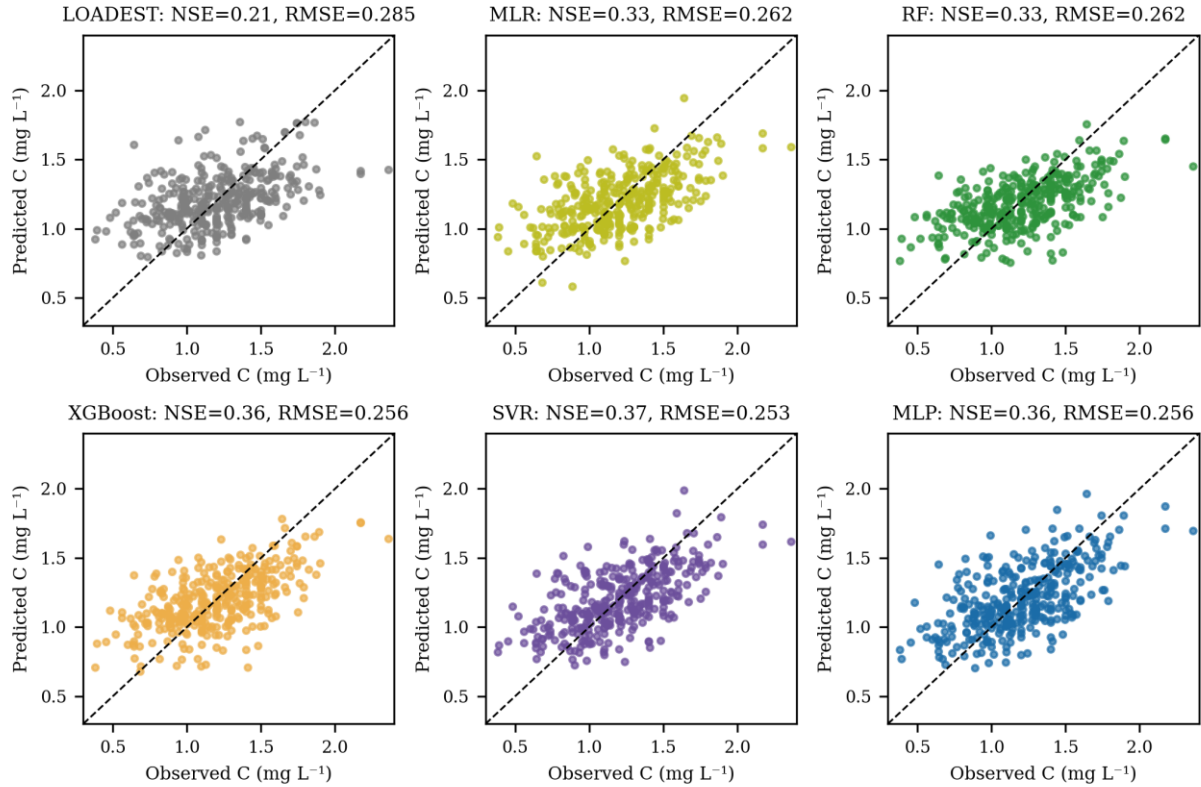


Figure 5. Observed and predicted monthly nitrate concentrations under blocked cross-validation

The forward test reversed part of this ranking (Figure 6). The MLP was the best model ($NSE = 0.30$, $PBIAS = -1.8\%$), followed by MLR (0.28). The tree ensembles had the lowest skill (RF 0.12, XGBoost 0.18) and the largest negative biases (-13.1% and -10.7%). The reason is that trees assign to all later dates the concentration

level of the last training years, so they cannot follow the continued increase in 2005–2011. SVR and LOADEST were also biased low (-10% and -6%). RF had the highest R^2 (0.41) despite its low NSE, which confirms that its errors were mainly systematic.

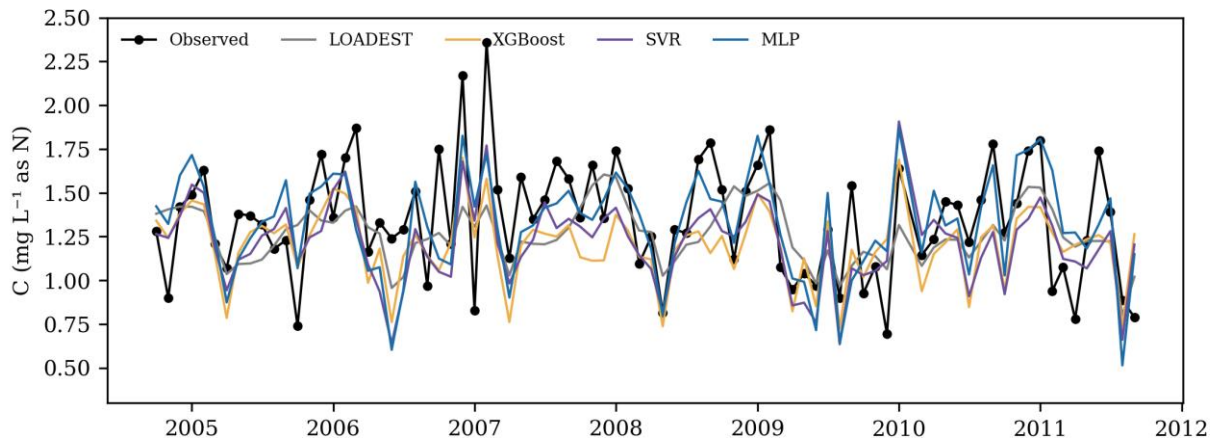


Figure 6. Observed and predicted monthly nitrate concentrations in the forward-test period (water years 2005–2011)

3.3. Model Interpretation

Permutation importance (Figure 7) ranked the cosine seasonal term and time as the most influential predictors in both RF and XGBoost, followed by BFI, the monthly change in discharge and $\ln Q$. The partial dependence of concentration on time rose from about 1.02 mg L^{-1} in the early 1980s to about 1.31 mg L^{-1} after 2006, with steps

around 1989–1990 and 2004–2006. Concentration decreased with increasing monthly discharge, from about 1.22 to 1.04 mg L^{-1} over the observed range. It increased with BFI, from about 1.05 mg L^{-1} at $BFI \approx 0.3$ to 1.19 mg L^{-1} at $BFI > 0.7$. These patterns show dilution of groundwater-derived nitrate by quickflow.

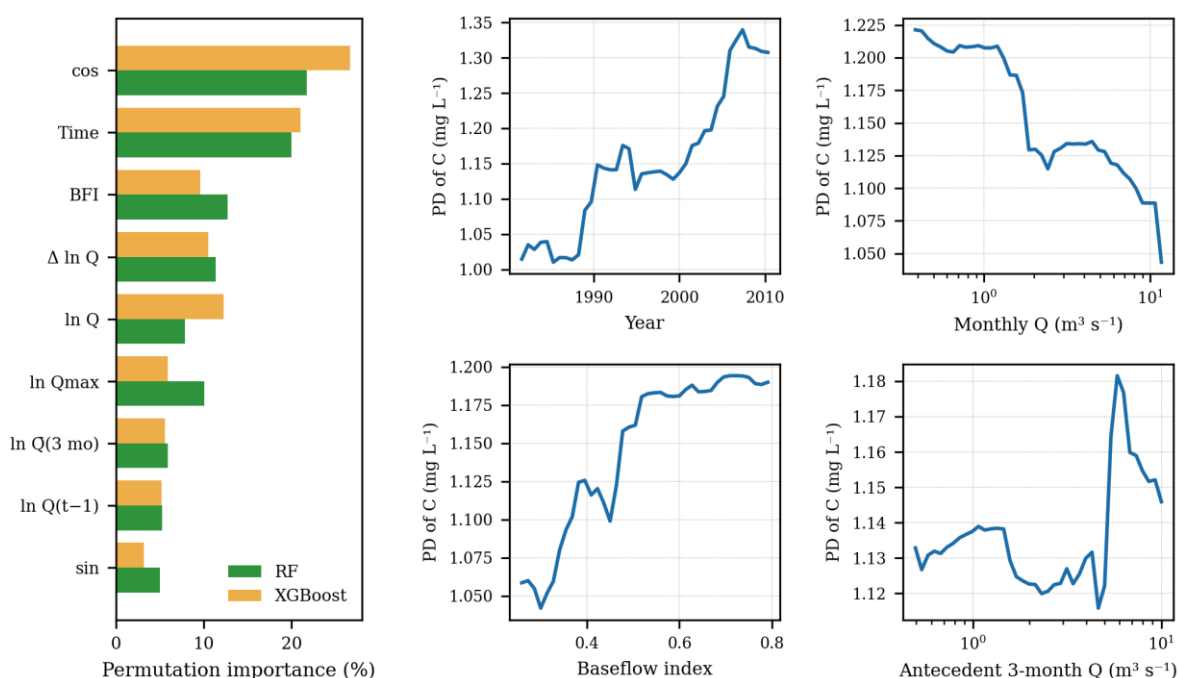


Figure 7. Permutation importance (RF and XGBoost) and XGBoost partial dependence of nitrate concentration on time, discharge, baseflow index and antecedent discharge

3.4. Annual Loads

Annual nitrate loads estimated by WRTDS averaged 134 t N yr^{-1} (range $48\text{--}266 \text{ t N yr}^{-1}$, lowest in water year 2002 and highest in 2003), which corresponds to a yield of $4.6 \text{ kg N ha}^{-1} \text{ yr}^{-1}$. The flow-normalized WRTDS load increased from 108 to 148 t N yr^{-1} over 1981–2011 (Figure 8). The annual loads from all six monthly models tracked the WRTDS values closely ($r \geq 0.98$; $\text{NSE} = 0.81\text{--}0.92$; Table 7). RF ($\text{NSE} = 0.92$), MLP (0.92) and LOADEST (0.92) agreed best. All models overestimated the WRTDS loads by 6.1–8.4%, with the largest differences in wet years such as 2003 and 2010.

Table 7. Agreement of annual loads (water years 1981–2011) with WRTDS estimates

Model	Mean load (t yr^{-1})	R^2	RMSE (t yr^{-1})	NSE	PBIAS (%)
LOADEST	143.7	0.979	14.3	0.917	+7.0
MLR	145.6	0.965	21.8	0.809	+8.4
RF	142.5	0.983	13.9	0.922	+6.1
XGBoost	143.2	0.979	15.3	0.905	+6.6
SVR	145.5	0.986	17.3	0.880	+8.4
MLP	143.7	0.989	14.3	0.918	+7.0
WRTDS	134.3	—	—	—	—

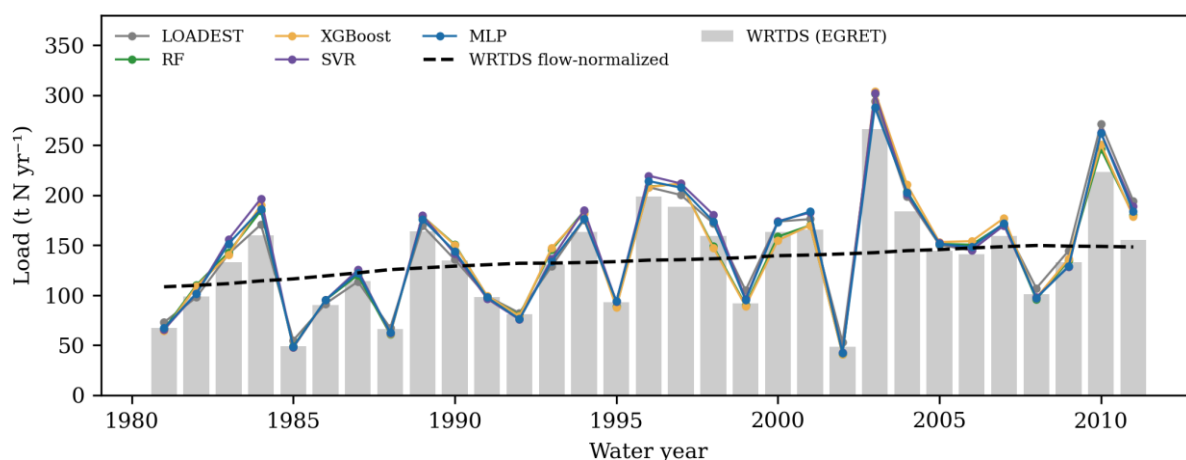


Figure 8. Annual nitrate loads from the monthly models compared with WRTDS and flow-normalized WRTDS loads

4. Discussion

Monthly nitrate concentrations in the Choptank River varied little and were only weakly related to

discharge. This near-chemostatic behaviour is typical of agricultural catchments with large legacy nitrogen stores (Basu et al., 2010) and is consistent with the long groundwater residence times reported for the Delmarva Peninsula (Lindsey et al., 2003; Sanford and Pope, 2013). The winter maximum, the positive relation with BFI and the dilution at high discharge all point to groundwater as the main nitrate pathway. The low winter uptake and denitrification also contribute to the seasonal pattern. The significant increase of about 1% per year after flow adjustment is consistent with the slow arrival of nitrogen applied in earlier decades and with the concerns about legacy effects delaying improvements in Chesapeake Bay water quality (Sanford and Pope, 2013; Chang et al., 2021).

Because of this chemostatic behaviour, the concentration models could explain only 21–38% of the monthly variance. Much of the remaining variance is probably sampling noise: most months had a single grab sample, which may not represent the monthly mean. The ML models improved on LOADEST mainly because they could use BFI and antecedent discharge, and SVR and MLP gained the most. The gain over the extended linear model was, however, small ($\Delta\text{NSE} \leq 0.05$), which suggests that the predictable part of the signal is largely captured by a few smooth relationships.

The forward test showed a practical weakness of tree-based models for water-quality trend applications. Trees cannot extrapolate a time trend beyond the training period, so RF and XGBoost underpredicted recent concentrations by 11–13%. Models with a smooth parametric or functional dependence on time (MLR, MLP) extrapolated better. This matters for load programmes, which often need to estimate the most recent years and in which a flattened trend could hide a continuing deterioration. WRTDS avoids this problem by re-estimating the model locally in time (Hirsch et al., 2010), and a similar local re-fitting or explicit trend terms would be needed if ML models were used operationally.

For annual loads, the model choice mattered much less than for concentrations, because discharge variability dominates the loads. All monthly models were 6–8% higher than WRTDS. Part of this difference is structural: multiplying monthly mean concentration by monthly mean discharge ignores the negative within-month covariance between concentration and discharge, which overweights the concentration during high flows. The smearing correction adds a small further positive bias. Daily estimation, as in WRTDS, is therefore preferable when accurate loads are required. The monthly models remain useful for gap filling and for exploring relationships. The study is limited to one watershed and one constituent, and the monthly aggregation of irregular samples introduces uncertainty that was not propagated to the load estimates.

5. Conclusions

Monthly nitrate concentrations and loads in the Choptank River were modelled for water years 1980–2011 with LOADEST, an extended linear regression and four ML models. The main conclusions are:

(1) Nitrate concentrations were nearly chemostatic, highest in winter and under baseflow-dominated conditions, and increased significantly by about 1% per year, also after flow adjustment.

(2) In blocked cross-validation, ML models explained somewhat more of the concentration variance than LOADEST (NSE up to 0.38 for SVR vs. 0.21), but the gain over an extended linear model was small.

(3) In a forward test, tree ensembles underestimated recent concentrations by 11–13% because they cannot extrapolate trends, whereas the MLP and MLR models were the most robust.

(4) Annual loads from all models agreed well with WRTDS (NSE = 0.81–0.92) but were 6–8% higher because of monthly aggregation. Load estimates are therefore controlled mainly by discharge, whereas trend assessment requires models that represent time explicitly.

Data Availability Statement

The discharge, nitrate and WRTDS data are openly available in the EGRET R package (Choptank_eList; <https://CRAN.R-project.org/package=EGRET>), originally from the USGS National Water Information System. Scripts are available from the corresponding author on request.

Conflict of Interest

The authors declare no conflict of interest.

References

- Basu, N.B., Destouni, G., Jawitz, J.W., Thompson, S.E., Loukinova, N.V., Darracq, A., et al. (2010). Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity. *Geophysical Research Letters*, 37, L23404. <https://doi.org/10.1029/2010GL045168>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chang, S.Y., Zhang, Q., Byrnes, D.K., Basu, N.B., & Van Meter, K.J. (2021). Chesapeake legacies: the importance of legacy nitrogen to improving Chesapeake Bay water quality. *Environmental Research Letters*. <https://doi.org/10.1088/1748-9326/ac0d7b>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Cohn, T.A., DeLong, L.L., Gilroy, E.J., Hirsch, R.M., & Wells, D.K. (1989). Estimating constituent loads. *Water Resources Research*, 25(5), 937–942. <https://doi.org/10.1029/WR025i005p00937>
- Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, 78(383), 605–610.

- Godsey, S.E., Kirchner, J.W., & Clow, D.W. (2009). Concentration–discharge relationships reflect chemostatic characteristics of US catchments. *Hydrological Processes*, 23, 1844–1864. <https://doi.org/10.1002/hyp.7315>
- Harrison, J.W., Lucius, M.A., Farrell, J.L., Eichler, L.W., & Relyea, R.A. (2021). Prediction of stream nitrogen and phosphorus concentrations from high-frequency sensors using Random Forests Regression. *Science of the Total Environment*, 763, 143005. <https://doi.org/10.1016/j.scitotenv.2020.143005>
- Hirsch, R.M., & De Cicco, L.A. (2015). User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data. U.S. Geological Survey Techniques and Methods, book 4, chap. A10. <https://doi.org/10.3133/tm4A10>
- Hirsch, R.M., Moyer, D.L., & Archfield, S.A. (2010). Weighted Regressions on Time, Discharge, and Season (WRTDS), with an application to Chesapeake Bay river inputs. *Journal of the American Water Resources Association*, 46(5), 857–880. <https://doi.org/10.1111/j.1752-1688.2010.00482.x>
- Hirsch, R.M., Slack, J.R., & Smith, R.A. (1982). Techniques of trend analysis for monthly water quality data. *Water Resources Research*, 18(1), 107–121. <https://doi.org/10.1029/WR018i001p00107>
- Lindsey, B.D., Phillips, S.W., Donnelly, C.A., Speiran, G.K., Plummer, L.N., Böhlke, J.K., et al. (2003). Residence times and nitrate transport in ground water discharging to streams in the Chesapeake Bay Watershed. U.S. Geological Survey Water-Resources Investigations Report 03-4035. <https://doi.org/10.3133/wri034035>
- Lyne, V., & Hollick, M. (1979). Stochastic time-variable rainfall-runoff modelling. Institute of Engineers Australia National Conference, Perth, 89–93.
- Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., & Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900.
- Nash, J.E., & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290.
- Nathan, R.J., & McMahon, T.A. (1990). Evaluation of automated techniques for base flow and recession analyses. *Water Resources Research*, 26(7), 1465–1473. <https://doi.org/10.1029/WR026i007p01465>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- Runkel, R.L., Crawford, C.G., & Cohn, T.A. (2004). Load Estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers. U.S. Geological Survey Techniques and Methods, book 4, chap. A5. <https://doi.org/10.3133/tm4A5>
- Sanford, W.E., & Pope, J.P. (2013). Quantifying groundwater’s role in delaying improvements to Chesapeake Bay water quality. *Environmental Science & Technology*, 47(23), 13330–13338. <https://doi.org/10.1021/es401334k>
- Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, 11(5), 910. <https://doi.org/10.3390/w11050910>
- USEPA. (2010). Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus and Sediment. U.S. Environmental Protection Agency, Region 3, Philadelphia, PA.
- Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.
- Zhi, W., Feng, D., Tsai, W.-P., Sterle, G., Harpold, A., Shen, C., & Li, L. (2021). From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale? *Environmental Science & Technology*, 55(4), 2357–2368. <https://doi.org/10.1021/acs.est.0c06783>