Review Comments

This manuscript by Yang et al. explores the association between land use, hydrological, and climate properties on watershed nitrate export. They mainly explore the spatial and temporal patterns of synchrony between annual peak nitrate concentrations and peak flow, with the aim of understanding the typologies of watersheds where peak N aligns with maximum discharge (QMax-Synced), minimum discharge (QMin-Synced), or neither (Asynced). I believe that this work contributes to a better understanding of nitrate export across land-use gradients, and the exploratory nature of the paper provides a foundation for future investigation of the mechanistic drivers of watershed seasonality patterns. However, I believe revisions are needed to make the manuscript publication ready.

We thank the reviewer for the constructive and encouraging comments on our manuscript. We are pleased that you recognise the contribution to improving understanding of nitrate export across land-use and hydro-climatic gradients, and that the synchrony framework provides a useful basis for mechanistic investigation.

General comments

- The current analyses using descriptive statistics and RF model are finding correlations and associations between catchment feature and synchrony class. Throughout the manuscript and in the title, "drivers" is used which implies a mechanistic association. I think it would be more accurate to use "associations" or "correlations" in place of "drivers".

  We agree and thank the reviewer for this helpful point. We will revise the title as: "When does nitrate peak in rivers and why? Catchment traits and climate relate to synchrony with discharge".

- Please clarify there was any pre-processing and outlier detection of the water quality data. For example, NW-88004024 with $R^2 = 0.11$ appears to be influenced by a single extreme outlier (~16 mg/L when all other observations are <4 mg/L). Is this observation is reliable? Additionally, was there any gap filling of flow in the cases where flow data was missing? Was there a screening for catchments based on proportion of high flow events that had sampled N data?

  We thank the reviewer for this insightful comment. As correctly pointed out, R^2 is highly influenced by outliers. Following suggestions from reviewer 1, we will present Spearman's Rho instead of R2 in the revised manuscript. The results show that the temporal structure is well preserved in the reconstructed series (median Spearman's ρ = 0.72, IQR = 0.134).

  We applied the threshold of 90% available over the 20-year period on screening the flow data mentioned in the Method 2.1. Short missing segments were filled using simple linear interpolation to produce the continuous daily series.

  We'll add a new sentence in the beginning of Result: "To assess the temporal consistency, spearman correlations between observed and reconstructed concentrations were calculated showing a median ρ of 0.72 (IQR = 0.134). No sites were excluded based on model performance." And we will revise the sentences in the methods: "For the synchrony analysis, both reconstructed concentrations and observed discharge were subsequently aggregated to monthly mean values, and all synchrony classifications were determined at the monthly scale based on the timing of monthly maxima and minima. To assess the adequacy of WRTDS-estimated concentrations for identifying monthly peaks, we computed the temporal Spearman correlation between observed and modelled monthly concentrations to assess agreement in temporal patterns relevant to the synchrony analysis."

We will add the gap filling method about the discharges on L96: "Short missing segments of daily discharges were filled using simple linear interpolation to produce the continuous daily series."

Station NW-88003442 exhibits markedly higher concentration variability before ~2004 compared to after, suggesting a potential shift in catchment or monitoring conditions. WRTDS assumes gradual evolution of concentration-discharge relationships and may not perform well when abrupt changes occur (Hirsch et al., 2010). The authors should investigate whether a known change (e.g., dam construction, wastewater treatment upgrades, monitoring protocol change) occurred around this time, and discuss whether modeling this site as a single continuous record is appropriate or whether the pre- and post-2004 periods should be treated separately. While model performance statistics suggest WRTDS handled this transition adequately, if the catchment underwent a structural change, the extracted β_2 coefficients and synchrony classification for this site may reflect a blend of two distinct periods rather than a coherent long-term signal?

We agree and many thanks for this thoughtful comment. The site NW-88003442 shows a structural change in nitrate behaviour around 2003–2004. However, based on available public records, we were unable to identify a clearly attributable intervention.

To assess the impact on our synchrony results, we re-computed peak concentration months and synchrony classification using only the post-2004 data. The result of synchrony categories is also consistent with the previous result, which does not alter any of the aggregated patterns or conclusions reported in the manuscript. We therefore retain the full record in the main analysis for consistency across sites and add the sensitivity analysis for this site.

We will include the sentence at the end of Method 2.3: "One catchment (NW-88003442) exhibited a clear structural change in the early 2000s. As a robustness check, we recalculated its synchrony using only post-2004 data; the dominant synchrony category remained unchanged, so the full record was retained for consistency across sites."

Lines 190-194, 225-226: The finding that concentration magnitudes across Qmax and Qmin classes do not differ is surprising given that arable land, and thus I assume nitrogen inputs, is significantly higher in QMax-Synced catchments. This is not a typical pattern I've seen across other catchments. Some discussion of why in these catchments higher agricultural inputs don't translate to higher concentrations would strengthen the interpretation.

We agree and thank the reviewer for this comment. The observation that peak concentrations do not differ significantly across synchrony classes, despite higher arable cover in QMax-Synced catchments, is indeed important and requires further clarification. Across many UK studies, river nitrate concentrations have been shown to decouple from agricultural nitrogen inputs because winter high-flow conditions, groundwater mixing, and subsurface denitrification collectively suppress concentration peaks even in highly arable catchments (Bowes et al., 2014; Hiscock et al., 2023). By contrast, although their overall nitrogen inputs are lower, QMin-Synced catchments may reach nitrate peaks comparable to, or even exceeding, those in agricultural QMax catchments because summer low-flow conditions sharply reduce dilution and amplify the influence of continuous urban and wastewater inputs (Cooper et al., 2022).

We will add following sentence to the Discussion 4.1 : "Although agricultural catchments receive larger nitrogen inputs, winter high-flow conditions, groundwater mixing and subsurface denitrification can suppress concentration peaks (Bowes et al., 2014; Hiscock et al., 2023), whereas in urban catchments summer low-flow periods amplify continuous wastewater and urban drainage inputs (Cooper et al., 2022); as a result, QMin-Synced catchments may exhibit peak nitrate concentrations comparable to those in QMax-Synced systems"

The interpretation of QMin-Synced catchments as uniformly "urban-dominated" may oversimplify what appears to be a mechanistically heterogeneous group. I understand that you are looking at dominant behavior and broad stroke patterns, however, I think attributing QMin-Synced catchment behavior exclusively to urban dynamics leads to missing important nuance. For instance, while arable land is significantly higher in QMax-Synced catchments on average, Figure 5 shows that approximately 25% of QMin-Synced catchments have >20% arable land cover. Depending on agricultural intensity, proximity to the watershed outlet, etc., this fraction could meaningfully contribute to nitrate dynamics. The possibility that a subset of QMin catchments reflects agricultural legacy contributions, particularly in catchments with lower drainage density, deserves consideration before concluding that QMin synchrony is exclusively urban-driven. Consider stratifying QMin-Synced catchments by WWTP density to distinguish (1) High-WWTP catchments, high urban land use, and high population density where urban point sources likely dominate, and low-WWTP catchments where other mechanisms (legacy groundwater, forested catchment dynamics, or non-WWTP urban sources) may drive QMin behavior.

We thank the reviewer for this thoughtful insight and suggestion to consider the heterogeneity within QMin-Synced catchments. Additional Spearman analyses were conducted focusing specifically on QMin catchments. These analyses show that QMin catchments form a continuous land-use gradient but no discrete subgroups. Median nitrate concentrations may increase with arable cover ($\rho = 0.41$, $p=0.11$) and urban land ($\rho = 0.42$, $p=0.10$) and decrease with grassland cover (–0.50, $p \leq 0.05$), indicating that background concentration levels are affected primarily by land use composition. The reason why we have not used the WWTPs density is because it does not directly correlate with urban land cover or population density. High densities of treatment work often occur in rural areas, whereas large cities are often served by a small number of large capacity treatment works.

The manuscript states that higher CV_C/CV_Q in QMin catchments signals "stronger and more dynamic anthropogenic pressures." and "increased urbanisation and population density are likely the main drivers of QMin-Synchrony, reflecting the dominance of continuous anthropogenic nitrate inputs." The first sentence is vague and overall I'm not sure if this interpretation is well-supported:

WWTP effluent loads are typically constant, which would produce dilution-driven concentration variability (high CVc) but wouldn't necessarily indicate "dynamic" pressures because it would indicate consistent point-source loading being diluted by variable flow. Are the authors referring to other urban pressures? If so, the specific pressures are nebulous, which weakens the argument.

The wide distribution of CV_C/CV_Q values within QMin-Synced (Figure S9) that overlaps with QMax-Synced suggests this class may contain distinct subgroups operating under different mechanisms.

High CVc/CVq can also occur in natural forested catchments (Ehrhardt et al. 2019). Given that woodland percentage appears higher in QMin than QMax catchments (Figure 5), could some of this variability reflect natural catchment dynamics rather than urban watershed dynamics?

Second, we tested whether the high CVc/CVq values in QMin-Synced catchments might be explained by agricultural legacy or forested dynamics, as suggested. Legacy-dominated agricultural systems generally exhibit weak or chemostatic C–Q behaviour and low temporal variability, because groundwater or soil-nitrogen stores release nitrate gradually over time (Winter et al., 2021). Johnson & Stets (2020) similarly show that legacy nitrate elevates winter low-flow concentrations but does not produce sharp annual peaks or strong flow sensitivity. These signatures differ from our QMin-Synced catchments, which in general exhibit steep negative C–Q slopes, high CVc/CVq, and a single annual concentration peak aligned with the minimum-flow month, patterns consistent with hydrological dilution of continuous urban inputs rather than gradual legacy drainage. Our correlation analyses further support this distinction: CVc/CVq increases strongly with urban land and population density ($\rho = 0.55$ and 0.52, $p < 0.05$) and decreases with arable land ($\rho = -0.59$, $p < 0.05$) and woodland cover ($\rho = -0.61$, $p < 0.01$). Woodland cover is low (median = 8%, IQR=4%) across all QMin sites. Thus, all these results could suggest legacy nitrogen likely contributes to concentration levels in some QMin catchments but does not dominate the temporal variability or the synchrony mechanism. These additional analyses support that QMin-Synced catchments are heterogeneous, but this heterogeneity is best described as a gradient in urbanisation and land-use composition, and the urbanisation may remain the primary influencing impacts on the QMin-Synchrony behaviour.

We will add the following sentence in the Method 2.4: "To test the possible internal heterogeneity of QMin-Synced catchments, we also calculated Spearman correlations among land-use variables and two nitrate metrics (median concentrations and CVc/CVq)."

We will add more details in the Result 3.3: "For a further spearman correlation within QMin-synced catchments, median nitrate concentrations showed positive tendencies with arable land ($\rho = 0.41$, $p = 0.11$) and urban land ($\rho = 0.42$, $p = 0.10$), and a significant negative association with grassland cover ($\rho = -0.50$, $p = 0.04$). In contrast, CVc/CVq was most strongly correlated with urban land ($\rho = 0.55$, $p < 0.05$), population density ($\rho = 0.52$, $p < 0.05$), arable land ($\rho = -0.59$, $p < 0.05$) and woodland cover ($\rho = -0.61$, $p < 0.01$). All the other correlation results are shown in Fig. S6."

We will revise the sentences in the Discussion: "A legacy-nitrogen explanation for QMin-synchrony, whereby slower drainage of stored soil or groundwater nitrate could elevate concentrations during low-flow periods, would be consistent with the large literature on this topic (Johnson and Stets, 2020). Our Spearman correlations suggest that agricultural legacy may indeed contribute to background nitrate levels in some QMin-Synced catchments. Legacy-dominated agricultural systems typically show weak or chemostatic C–Q behaviour and comparatively low temporal variability because nitrate is released gradually from subsurface stores (Winter et al., 2021).

At the same time, several characteristics of our QMin-Synced catchments point to a stronger influence of continuous urban inputs on the synchrony pattern. These catchments feature steeply negative C–Q slopes, consistent with hydrological dilution of relatively stable urban point sources. Moreover, if both diffuse and point sources were active, we would likely expect dual peaks, one during winter flushing (as with QMax-Synched catchments) and another at low flow, yet only a single low-flow maximum is observed. This pattern further implies that urban land and population density associated inputs have largely displaced the diffuse, winter-mobilisation behaviour typical of QMax-Synched catchments, creating an engineered inversion where nitrate concentrations peak only under low-flow conditions and are otherwise easily diluted (Kaushal and Belt, 2012; Kaushal et al., 2011). Our correlation analyses further support the interpretation that variability (CVc/CVq) increases with urbanisation, whereas arable land is associated with reduced variability, suggesting a more stable, weakly flow-responsive behaviour typical of legacy-influenced systems. Together, while legacy nitrate contributes to background concentration levels in some QMin-Synced catchments, the observed QMin-synchrony may be shaped primarily by flow-dependent dilution of continuous urban inputs."

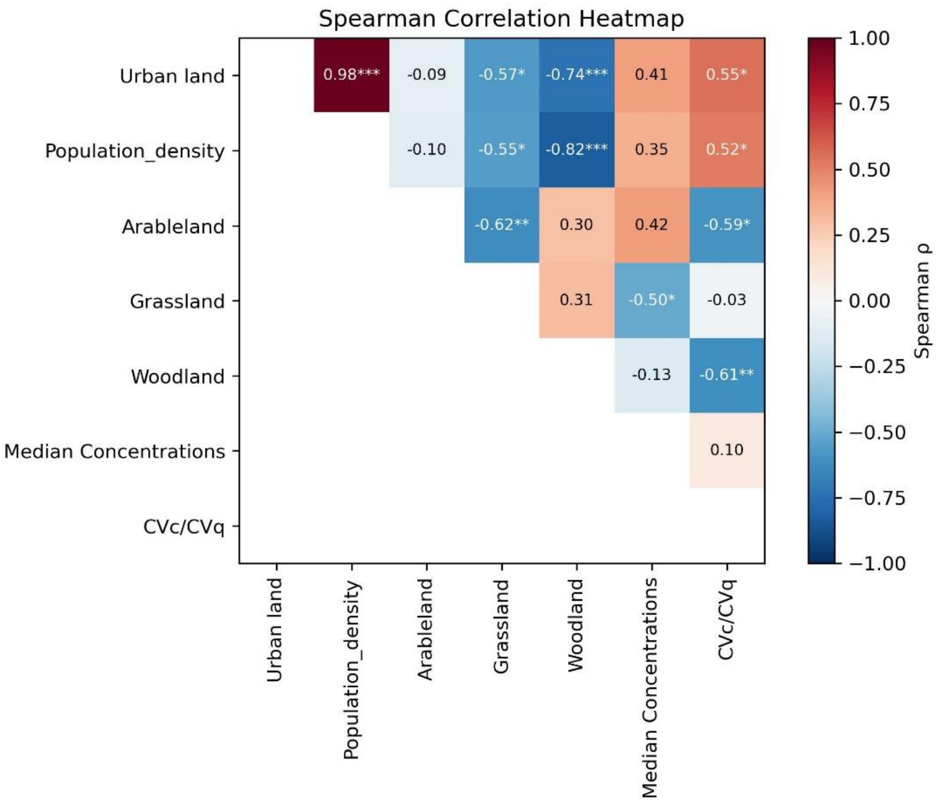

Figure S6: Spearman Correlation matrix for land-use variables and nitrate metrics in QMin-Synced catchments; Colours indicate correlation strength and direction. Asterisks denote significance: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

Specific comments

- Section 2.2: WRTDS is a complex model, and given that the authors use its model parameters directly, the manuscript would benefit from subtle but important clarification on the model methods. For instance, "fitted through regression at each time point" is ambiguous. The readers might not understand that coefficients are estimated for each modeled day, meaning each station has thousands of β_2 values. Additionally, there is no mention of the time, season, and discharge window widths used, which affect how rapidly the estimated C-Q relationship can change over time. Did you choose default window sizes or select each catchment window them based on model fit?

We agree and thank the reviewer for highlighting this important point. We will clarify the WRTDS fitting procedure in Sect. 2.2. Specifically, we now explain that WRTDS performs a locally weighted regression for each modelled day, such that every catchment has thousands of sets of coefficients that vary smoothly over time. We will explicitly report the window widths governing the local weighting. In this study, we used the default EGRET settings, namely half-window widths of 7 years in time, 0.5 years in season, and 2 natural-log units in discharge. These values are the recommended defaults in EGRET and are widely applied in WRTDS studies, ensuring methodological consistency across all catchments.

The revised text will clarify and resolve the ambiguity noted by the reviewer: "The WRTDS method estimates a smooth concentration surface in time–discharge–season space by performing a locally weighted regression for every modelled day, producing thousands of smoothly varying coefficient vectors for each catchment rather than a single global fit. The local weighting is controlled by three half-window widths: 7 years in time, 0.5 years in season, and 2 natural-log units in discharge."

- Line 169-172: Confidence in the models could be strengthened through more detailed model assessment. While observed vs. modeled concentration plots are provided in supplementary material, the main text should clarify whether any sites were excluded based on fit criteria and whether poor model performance at certain sites affects interpretation of results. Additionally, RMSE is difficult to evaluate without knowing mean concentrations at each site. Using a normalized metric such as KGE, NSE, or PBIAS would aid interpretation in the manuscript.

We agree and thank the reviewer for this helpful suggestion. Following the other reviewer's recommendation, we will use the Spearman temporal correlation as a metric to assess whether WRTDS captures the temporal pattern relevant for synchrony detection. Because our analysis focuses on the timing of monthly concentration peaks, temporal correlations could provide a more appropriate assessment.

We will revise the results section 3.1 as follows: "To assess the temporal consistency, spearman correlations between observed and reconstructed concentrations were calculated showing a median ρ of 0.72 (IQR = 0.134). No sites were excluded based on model performance."

- Line 117: Was distinct unimodal pattern determined visually? Did any sites not have unimodal pattern at all sites and were therefore removed from the dataset? Or were these part of the Asynced class?

We agree and thank the reviewer for raising this question. The statement regarding a "distinct unimodal pattern" refers to the dominant seasonal cycle that characterises temperate UK catchments. Prior to the synchrony analysis, we verified that all sites exhibited a single, well-defined seasonal peak–trough structure based on their long-term monthly means and that annual maxima and minima in discharge and nitrate concentration could be consistently identified.

We will add the following clarification: "We confirmed that both discharge and nitrate concentration exhibit a clear seasonal peak–trough structure at all sites, based on their long-term monthly means."

- Section 2.4: Clarification on the RF model would strengthen the analysis. First, was there any hyperparameter tuning done or were default parameters used? Second, it is unclear whether the permutation importance rankings were averaged across the k-fold cross validation, or whether k-fold cross validation was used only to assess model performance and a single final model fit to all data used to generate permutation importance rankings.

We agree and thank the reviewer for this useful comment. We did not perform hyperparameter tuning. Apart from setting num.trees = 500, all parameters followed the ranger defaults. The repeated 10-fold cross-validation was used solely to evaluate model performance and stability. Permutation importance was not averaged across cross-validation folds. Consistent with the variable-importance framework of Altmann et al. (2010), CV was used solely to assess classification performance, after which a final RF model was trained on the full dataset, and computed permutation importance and permutation-based p-values from this final model, using 1000 label permutations to generate the null distribution.

We will clarify this as follows: "To understand the catchment controls on the two synchrony patterns, a series of catchment descriptors (Table 1) were selected, and a Random Forest (RF) model was applied to relate synchrony type (QMax-Synced vs QMin-Synced) to catchment characteristics. RF analysis was chosen for its robustness and ability to handle complex interactions within the data (Breiman, 2001). Asynced catchments were excluded as this group represents catchments with no dominant synchrony pattern and high variability, which would reduce the clarity and interpretability of the results. A Random Forest classifier was implemented using the ranger (Version 0.16.0) engine via mlr3 (Version 0.22.1) in R (Lang et al., 2019; Wright and Ziegler, 2017). No hyperparameter tuning was performed; ranger default settings were used except that the number of trees was set to 500. A three-repeated 10-fold cross-validation procedure was solely used to evaluate model performance and its generalizability. Following cross-validation, a final RF model was trained on the full dataset, and permutation-based variable importance values were extracted from the underlying ranger model, following Altmann et al. (2010). Empirical permutation p-values (<0.05) were obtained from 1000 random permutations of the response variable to identify descriptors whose importance exceeded the null distribution."

- Section 3.4: This section is challenging to understand. In general, ternary plots are challenging to parse, and the current presentation does not provide sufficient guidance for interpretation. I suggest either providing more detailed guidance in the text walking the reader more carefully through the interpretation, and/or supplementing the plots with simpler plots to isolate the key relationships. Additionally, Figure S11 caption is not clear and should be expanded upon to help readers understand.

We thank the reviewer for this helpful comment. We agree that ternary plots can be difficult to interpret without explicit guidance.

We refined the paragraph with the starting sentence: While Section 3.3 identified catchment characteristics that explain dominant synchrony types across space, here we examine (i) what controls interannual variability in synchrony within catchments, and (ii) how long-term variability in synchrony composition relate to catchment attributes.

And more explicit guidance related to interpretation of the ternary plots will be included: "The ternary plots (Fig. 6) summarise the long-term synchrony composition at each site by showing the relative proportion of QMax-, QMin- and Asynced years. In these diagrams, each vertex represents 100% dominance of one synchrony type; positions along the edges indicate mixtures of two types; and points near the centre represent a balanced mixture of all three. The colour scale applied to each point represents values for a selected catchment descriptor (e.g., arable land percentage, Drainage Path Slope), allowing the ternary diagram to visualise how synchrony composition varies along environmental gradients. These colour gradients are consistent with the spatial correlations shown in Fig. S7. No catchment sits at a single vertex; instead, most exhibited a mix of synchrony types, with Asynced catchments spanning a particularly wide range. No sites are located near both the QMax and QMin vertices simultaneously, indicating that switching occurs mainly between a dominant mode and Asynced behaviour rather than directly between the two synchronous modes."

The caption of Figure S7 will be “Figure S7: Heatmaps showing Spearman correlations between catchment descriptors and the proportion of years classified as QMax-Synced, QMin-Synced, or Asynced within three groups of catchments. Coloured cells indicate the direction and strength of correlation (ρ), with warm colours denoting positive and cool colours denoting negative correlations. Asterisks mark statistically significant correlations ($p < 0.05$)”.

Technical corrections

- *QMax-Synched* and *QMax-Synced* are both used throughout the manuscript.

We thank the reviewer for raising this point. We will standardise the terminology throughout the manuscript and revise all occurrences to “QMax-Synced” for consistency.

- Line 84: Fix table and citation parentheses.

We thank the reviewer for pointing this out. The parentheses will be corrected and the citation will be added in the table.

- Line 84: Please make sure the catchment features used are outlined more clearly. For example, looking at the NRFA citation, I find two BFI variables but the manuscript does not state which one is being used.

We thank the reviewer for highlighting this ambiguity. The catchment descriptor used in our analysis is BFIHOST, the FEH soil-based Baseflow Index derived from the Hydrology Of Soil Types (HOST) classification. It is the variable provided within the NRFA FEH catchment descriptor dataset. Table 1 will be updated to reflect this.

- Line 94: citation missing for WWTP data

We thank the reviewer for this kind reminding. The citation for WWTP data will be added as: “Environment Agency: Consented Discharges to Controlled Waters with Conditions. [dataset], available at: https://www.data.gov.uk/dataset/55b8eaa8-60df-48a8-929a-060891b7a109/consented-discharges-to-controlled-waters-with-conditions1, last access: [10 Oct 2024], 2024”

- Line 112: Subscript 2 in model coefficient

We thank the reviewer for pointing this out. We will correct the formatting of the coefficient and ensured that the subscript “2” appears properly in all instances of β_2 throughout the manuscript.

- Line 113: missing period

We thank the reviewer for noting this typographical issue. The missing period at Line 113 will be added.

- Line 305: spelling error in figure caption "(d) Population density Density of WWTPs" should be Density of WWTPs

We thank the reviewer for pointing this out. The error in the caption of Figure 6d will be corrected to “Density of WWTPs.”

- Line 354: "In our study, random forest analysis identified urban area as the strongest explanatory variable in these catchments." This sentence is misleading. The random forest ranked urban LU highest in variable importance for the classification. It the predictor helps most with prediction accuracy, but it does not provide explanatory inference. The significant differences between synchrony classes were shown in the Wilcoxon results, not by the RF. I'd recommend revising the phrasing to reflect this.

We agree thank the reviewer for this important clarification. The RF analysis provides a ranking of predictors based on their contribution to classification accuracy but does not itself imply explanatory causation. The statistical differences between synchrony classes are instead supported by the Wilcoxon rank-sum tests.

We will revise the sentence as follows: "In our study, Random Forest analysis identified urban land cover as the most influential predictor for distinguishing QMin- from QMax-synced catchments, with Wilcoxon tests confirming significant differences between the two groups."

References:

Bowes, M. J., Jarvie, H. P., Naden, P. S., Old, G. H., Scarlett, P. M., Roberts, C., Armstrong, L. K., Harman, S. A., Wickham, H. D., and Collins, A. L.: Identifying priorities for nutrient mitigation using river concentration–flow relationships: The Thames basin, UK, J. Hydrol., 517, 1-12, 10.1016/j.jhydrol.2014.03.063, 2014.

Cooper, R. J., Warren, R. J., Clarke, S. J., and Hiscock, K. M.: Evaluating the impacts of contrasting sewage treatment methods on nutrient dynamics across the River Wensum catchment, UK, Sci. Total Environ., 804, 150146, 10.1016/j.scitotenv.2021.150146, 2022.

Johnson, H. M. and Stets, E. G.: Nitrate in Streams During Winter Low-Flow Conditions as an Indicator of Legacy Nitrate, Water Resour. Res., 56, 10.1029/2019wr026996, 2020.

Hiscock, K. M., Cooper, R. J., Lovett, A. A., and Sünnenberg, G.: Export Coefficient Modelling of Nutrient Neutrality to Protect Aquatic Habitats in the River Wensum Catchment, UK, Environments, 10, 10.3390/environments10100168, 2023.

Winter, C., Tarasova, L., Lutz, S. R., Musolff, A., Kumar, R., and Fleckenstein, J. H.: Explaining the Variability in High-Frequency Nitrate Export Patterns Using Long-Term Hydrological Event Classification, Water Resour. Res., 58, 10.1029/2021wr030938, 2022.