Hydroecology and Engineering

SCIEPublish

Article

Ecological Modelling of Physico-Chemical Parameters and EPT Macroinvertebrate Abundance in River Kiminini, Trans-Nzoia County, Kenya

Charity Nafuna Barasa [1,*], Gelas Simiyu Muse [1], Judith Ashindu Khazenzi [1] and Luckson Solicitor Maurice Gondwe [2]

[1] Department of Environmental Sciences, University of Eldoret, Eldoret P.O. Box 1125-30100, Kenya; gmuse@uoeld.ac.ke (G.S.M.); judith.khazenzi@uoeld.ac.ke (J.A.K.)

[2] Department of Aquaculture and Fisheries Science, Lilongwe University of Agriculture and Natural Resources, Lilongwe P.O. Box 219, Malawi; 200101527@luanar.ac.mw (L.S.M.G.)

* Corresponding author. E-mail: senvebhp00125@uoeld.ac.ke (C.N.B.)

Received: 24 June 2026; Revised: 22 July 2026; Accepted: 21 October 2026; Available online: 11 November 2026

ABSTRACT: Freshwater macroinvertebrates are widely used in biomonitoring due to their tendency to change in abundance and distribution with changing environmental conditions. Different taxonomic groups vary in response to individual physicochemical variables. This study assessed the association of physicochemical variables and the abundance of Ephemeroptera, Plecoptera, and Trichoptera (EPT) in River Kiminini, Trans-Nzoia County, Kenya. The sampling was done during the wet season from July to September 2023. The field data was collected twice a month for three months across 10 sampling stations, resulting in a total of 60 observations. Ephemeroptera, Plecoptera, and Trichoptera counts were used as response variables, while temperature, total dissolved solids (TDS), electrical conductivity (EC), pH (potential of hydrogen), nitrate, and phosphate were used as predictors. There was a significant overdispersion among the three taxa from a Poisson distribution, with a variance to mean ratio of 16.49, 5.73 and 5.48 for Ephemeroptera, Plecoptera and Trichoptera, respectively. The negative binomial generalized linear mixed models (NB2 GLMMs) with station as a random intercept were fitted, and predictors were standardized to 1 standard deviation. The NB2 (Negative binomial type 2 parameterization) specification was supported by model comparison over NB1 (Negative binomial type 1 parameterization) for all three taxa, especially for the Ephemeroptera. Ephemeroptera abundance was negatively associated with TDS (incidence-rate ratio [IRR] = 0.564, 95% confidence interval (CI): 0.350–0.907) and pH (IRR = 0.679, 95% CI: 0.511–0.902), but positively associated with nitrate (IRR = 1.372, 95% CI: 1.052–1.788). Trichoptera abundance was negatively associated with TDS (IRR = 0.322, 95% CI: 0.119–0.867) and pH (IRR = 0.387, 95% CI: 0.204–0.733). No statistically significant relationships were found for Plecoptera. Simulation-based residual diagnostics indicated no evidence of excess zero inflation for any taxon. Ephemeroptera demonstrated a modest residual-dispersion signal that warrants caution. The findings illustrate taxon-specific associations between physicochemical gradients and EPT abundance and provide a regional application of count-data modelling for freshwater biomonitoring in Kenya.

© 2026 The authors. This is an open access article under the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

Keywords: Physico-chemical variables; Freshwater macroinvertebrates; Negative-binomial GLMM; Overdispersion; Biomonitoring; EPT

1. Introduction

Freshwater ecosystems support a variety of macroinvertebrate species which play a key role in aquatic food webs, nutrient cycling and ecosystem function [1]. Benthic macroinvertebrate larvaes have been used extensively in biomonitoring. Their abundance, distribution, and community composition in response to changes in water chemistry and physical habitat indicate the ecological status of a water system [2]. Ephemeroptera, Plecoptera, and Trichoptera (EPT) larvae are vital in freshwater biomonitoring because of their sensitivity to environmental change [3].

Different EPT taxa vary in their response to environmental gradients [4]. Taxon-specific responses may be as a result of physiological tolerance, respiration, feeding strategy, and habitat requirements [5]. For instance, total dissolved solids (TDS) affect aquatic invertebrates through alterations on ionic and osmoregulatory conditions [6]. An experimental study by [7] reported that increased concentrations of total dissolved solids affect growth and may lead to mortality in aquatic [8]. Equally, studies modelling EPT taxa have showed that explanatory variables and environmental preferences can vary among taxa, supporting the use of taxon-specific analyses rather than assuming a common response across EPT groups [9].

Nutrients may correlate with macroinvertebrate abundance, but these relationships can be hard to interpret because nutrient levels may be associated with land use, primary production, habitat condition, and other environmental factors [10]. Thus, a statistical relationship between nutrients and macroinvertebrate abundance does not necessarily imply a cause/effect relationship. The same consideration applies to pH, electrical conductivity (EC), and TDS, which can act as correlated indicators of water chemistry rather than independent ecological processes.

Analyzing the abundance of macroinvertebrates is also statistically challenging because the counts are positive and typically have higher variance than the mean [11]. Poisson regression assumes the equality of the conditional mean and variance and may thus fail to yield adequate inference for overdispersed count data [12]. Alternatively, negative-binomial models are used for count data with additional variation beyond the Poisson variance structure, and are widely adopted for ecological count data [13]. If observations are recorded multiple times at the same sampling station, a mixed-effects approach can also explicitly capture the station-level variation and prevent multiple observations within a station from being treated as completely independent [14].

EPT taxa have been widely applied to freshwater biomonitoring [15], but quantitative modelling of taxon-specific responses to physicochemical gradients is poorly understood for Kenyan and other Afrotropical river systems. In Kenya, recent studies have shown associations between EPT distribution and human activities [3], further analyses are required to determine the association of each physicochemical gradient with EPT abundance while considering the overdispersion and repeated sampling. This study therefore represents a regional application of established ecological count-data statistical modelling approach.

This study aimed at quantification of the relationships between the abundance of Ephemeroptera, Plecoptera and Trichoptera (EPT) in addition to temperature, total dissolved solids (TDS), electrical conductivity (EC), pH, nitrate and phosphate concentrations in River Kiminini. We hypothesized that abundance of Ephemeroptera and Trichoptera would decline along increasing TDS, EC, and pH gradients, but that nutrient effects would be weaker or taxon-specific relative to ionic variables. We also anticipated weaker relationships for Plecoptera because we did not measure key habitat attributes, including substrate, flow, and habitat complexity. Collectively, these analyzes provide an understanding of how physicochemical gradients may structure EPT communities in a tropical river system.

2. Materials and Methods

2.1. Study Area

The study was conducted along River Kiminini in Trans-Nzoia County, Kenya, at an elevation of approximately 1400 m above sea level. Ten sampling stations were established along the river to represent variation in degradation, prevailing anthropogenic activities, and station accessibility. Sampling was conducted during the wet season from July to September 2023.

2.2. Study Design and Sampling

Sampling was conducted twice a month for three months, resulting in six sampling occasions at each of the 10 stations. There were 60 observations in total across all sampled stations. Physicochemical parameters and macroinvertebrate measurements were collected simultaneously during each sampling occasion.

At each station, water samples were collected in duplicate 500 mL polyethylene containers, labelled, and transported in an ice-filled cooler box to the University of Eldoret laboratory within 48 h for nutrient analysis. Macroinvertebrates were sampled from the riffles, runs, and pools, using a 500-µm mesh kick net. Samples were sorted in white trays, transferred to labelled vials, and preserved in 70% ethanol until laboratory processing.

2.3. Physicochemical Measurements

Temperature, pH, electrical conductivity (EC), and total dissolved solids (TDS) were measured *in situ* using a YSI Hydrolab Model 650 MDS multiparameter display system (YSI Incorporated, Yellow Springs, OH, USA). Nitrate and phosphate concentrations were determined from the duplicate water samples using standard colorimetric analytical procedures. Analyses were conducted at the University of Eldoret biotechnology laboratory using a DR 3000 spectrophotometer (Hach Company, Loveland, CO, USA) of approximately 340–1000 nm wavelength range and analytical-grade reagents. Nitrate analysis was determined using the Brucine method, while Phosphate analysis used the Ascorbic acid method.

2.4. Macroinvertebrate Sampling

Macroinvertebrates collected from riffle, run, and pool habitats were sorted and identified using updated taxonomic keys to obtain abundance counts for Ephemeroptera, Plecoptera, and Trichoptera (EPT). The three orders were analyzed separately because the study hypotheses predicted potentially different responses to physicochemical gradients.

The response variables were therefore Ephemeroptera, Plecoptera, and Trichoptera abundance, while temperature, TDS, EC, pH, nitrate, and phosphate were treated as explanatory variables. Since measurements were repeated at the same stations across sampling occasions, station identity was retained in the statistical analysis as a random intercept to account for non-independence among repeated observations.

2.5. Exploratory Analysis

The data set consisted of 60 observations and 10 variables. There were no missing values, and each of the stations contributed six observations. We examined the mean-variance relationship by evaluating the mean and variance of each EPT response. The ratios of variances to means were: Ephemeroptera (16.49), Plecoptera (5.73), and Trichoptera 5.48. The proportion of zero observations was 6.7% for Ephemeroptera, 21.7% for Plecoptera, and 56.7% for Trichoptera. These results indicated significant overdispersion compared to a Poisson distribution, especially for Ephemeroptera. Full Poisson model estimates of Pearson dispersion were 8.05, 5.26 and 3.16 for Ephemeroptera, Plecoptera and Trichoptera, respectively.

2.6. Assessment of Multicollinearity

The associations among the physicochemical predictors were assessed using Pearson correlations and variance inflation factors (VIFs). The strongest correlation was between TDS and EC (r = −0.828). TDS was also correlated with pH (r = −0.502) and nitrate (r = −0.538), while EC was positively correlated with nitrate (r = 0.627).

VIFs remained below 5 for all predictors in the three taxon-specific models. The highest VIFs were recorded for TDS and EC, with an approximate range of 3.2 to 4.2. These values indicate moderate or elevated predictor overlap but do not provide evidence of severe multicollinearity according to the commonly used VIF = 5 threshold. The TDS–EC relationship was therefore considered when interpreting individual coefficients.

2.7. Statistical Modelling

The response variables were counts and exhibited substantial overdispersion; therefore, Poisson models were not considered adequate as the final models. Poisson and negative-binomial models were compared using model fit statistics. The Poisson models had significantly larger Akaike Information Criterion (AIC) values than the negative-binomial models for all three taxa.

For repeated observations within stations, negative-binomial generalized linear mixed models (NB-GLMMs) were fitted using the glmmTMB package in R [16]. Station was included as a random intercept.

The general model was:

$$Y_i \sim NB(\mu_i, \phi)$$

with

$$\log(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_6 X_{6i} + b_{station(i)}$$

where Y_i is the observed abundance, μ_i is the expected abundance, $X_1, \ldots, X_6$ are the six physicochemical predictors, β_0 is the fixed-effect intercept, $\beta_1, \ldots, \beta_6$ are fixed-effect coefficients, and $b_{station(i)}$ is the station-specific random intercept.

The final models used the NB2 parameterization, in which the variance increases quadratically with the mean:

$$Var(Y_i) = \mu_i + \frac{\mu_i^2}{\phi}$$

Predictors were standardized to mean zero and standard deviation one before modelling. Consequently, each coefficient represents the change in log expected abundance associated with a one-standard deviation increase in the corresponding predictor while the other predictors and station-level variation are held constant.

NB1 and NB2 models were compared using AIC. NB2 produced lower AIC values for all three taxa. The improvement was significant for Ephemeroptera (ΔAIC ≈ 13.75) and smaller for Plecoptera (ΔAIC ≈ 1.36) and Trichoptera (ΔAIC ≈ 1.43). NB2 was therefore retained consistently across taxa.

Quasi-Poisson models were not used for likelihood-based model selection because their quasi-likelihood framework does not provide an AIC comparable with the likelihood-based Poisson and negative-binomial models. The choice of the final model was therefore based on the observed overdispersion, likelihood-based model comparison, and the repeated station sampling structure.

2.8. Model Interpretation

Regression coefficients were converted to incidence-rate ratios (IRRs) using:

$$IRR = e^{\beta}$$

An IRR below 1 indicates a lower expected abundance per one-standard deviation increase in the predictor, whereas an IRR above 1 indicates a higher expected abundance. These estimates demonstrate conditional statistical relationships and are not interpreted as causal effects.

Wald 95% confidence intervals were calculated for the coefficients and transformed to the IRR scale. Because the predictors were standardized, the intercept was not interpreted ecologically.

2.9. Model Diagnostics

Model adequacy was assessed using simulation-based residual diagnostics implemented in DHARMa. The diagnostics examined:

- residual uniformity;
- residual dispersion;
- zero inflation;
- outliers.

The use of simulation-based diagnostics was preferred to normal Q-Q plots because normality of residuals is not a standard distributional assumption of negative-binomial generalized linear mixed models.

For Ephemeroptera, residual uniformity was not significantly different from the expected distribution (p = 0.205), and there was no evidence of excess zero inflation (p = 1.000) or outliers (p = 1.000). The dispersion test demonstrated a modest residual-dispersion signal (dispersion = 2.19, p = 0.044), which was treated as a diagnostic caveat rather than evidence for automatically changing the model.

For Plecoptera, residual uniformity (p = 0.312), dispersion (p = 0.918) and zero-inflation tests (p = 0.222) provided no evidence of major model inadequacy.

For Trichoptera, residual uniformity (p = 0.750), dispersion (p = 0.454) and zero-inflation tests (p = 1.000) likewise provided no evidence of major model inadequacy.

Because sampling dates were not retained in the analytical dataset, explicit temporal autocorrelation could not be evaluated. Station-level repeated sampling was instead accounted for using the station random intercept.

2.10. Statistical Software

All analyses were conducted in R version 4.5.0 using relevant packages including MASS, car, glmmTMB, DHARMa and dplyr.

3. Results

3.1. Characteristics of the Response Variables

The dataset included 60 observations from 10 stations, with six observations per station. No missing values were present. Ephemeroptera abundance ranged from 0 to 79 individuals, with a mean of 7.83 and variance of 129.19. Plecoptera abundance ranged from 0 to 23, with a mean of 4.57 and variance of 26.18. Trichoptera abundance ranged from 0 to 18, with a mean of 1.23 and variance of 6.76, as shown in Table 1.

The consistent variance-to-mean ratios were 16.49 for Ephemeroptera, 5.73 for Plecoptera, and 5.48 for Trichoptera, demonstrating clear overdispersion.

Trichoptera had the highest proportion of zero observations (56.7%), followed by Plecoptera (21.7%) and Ephemeroptera (6.7%).

Table 1. Distribution of EPT abundance and evidence of overdispersion.

Taxon	Mean	Variance	Variance/Mean	Zero Counts (%)	Poisson Dispersion
Ephemeroptera	7.833	129.192	16.493	6.7	8.05
Plecoptera	4.567	26.182	5.733	21.7	5.263
Trichoptera	1.233	6.758	5.48	56.7	3.161

3.2. Physicochemical Predictors

Temperature ranged from 7.8 to 24.7 °C, TDS from 70 to 140, EC from 50 to 200, pH from 7.0 to 8.7, nitrate from 1.75 to 19.25, and phosphate from 0.25 to 3.15. TDS and EC showed the strongest pairwise correlation (r = −0.828). Other notable correlations included EC and nitrate (r = 0.627), TDS and nitrate (r = −0.538), and TDS and pH (r = −0.502).

All VIFs were below 5, although TDS and EC consistently had the largest VIFs. These results indicated that individual TDS and EC coefficients should be interpreted with some caution because the variables contain overlapping information.

3.3. Comparison of Count-Data Models

Poisson models showed substantial overdispersion for all three taxa. The Poisson AIC values were 587.76 for Ephemeroptera, 430.01 for Plecoptera and 216.05 for Trichoptera. The corresponding negative-binomial GLM AIC values were considerably lower: 365.91, 324.42 and 179.01, respectively (Table 2). When the repeated station structure was included, the NB2 GLMMs had AIC values of 358.31, 316.28 and 180.99 for Ephemeroptera, Plecoptera and Trichoptera, respectively. NB2 also had a lower AIC than NB1 for all taxa.

Table 2. Model comparison.

Taxon	Poisson AIC	NB-GLM AIC	NB1 GLMM AIC	NB2 GLMM AIC	Final Model
Ephemeroptera	587.76	365.91	372.06	358.31	NB2 GLMM
Plecoptera	430.01	324.42	317.64	316.28	NB2 GLMM
Trichoptera	216.05	179.01	182.41	180.99	NB2 GLMM

The station random-intercept variance was 0.306 for Ephemeroptera, 0.388 for Plecoptera, and 0.029 for Trichoptera, indicating residual station-level variation in the first two taxa and comparatively little station-level variation for Trichoptera.

3.4. Ephemeroptera

The final NB2 GLMM showed statistically supported negative associations between Ephemeroptera abundance and TDS and pH, and a positive association with nitrate. A one-SD increase in TDS was associated with an estimated 43.6% lower expected Ephemeroptera abundance (IRR = 0.564, 95% CI: 0.350–0.907). A one-SD increase in pH was associated with an estimated 32.1% lower expected abundance (IRR = 0.679, 95% CI: 0.511–0.902). A one-SD increase in nitrate was associated with an estimated 37.2% higher expected abundance (IRR = 1.372, 95% CI: 1.052–1.788) (Table 3). Temperature, EC, and phosphate were not statistically supported as predictors.

Table 3. Final NB2 GLMM results for EPT abundance.

Taxon	Predictor	Estimate	SE	Z	*P*	IRR	95% CI
Ephemeroptera	Temperature	−0.089	0.12	−0.741	0.459	0.915	0.724–1.157
	TDS	−0.573	0.243	−2.362	0.018	0.564	0.350–0.907
	EC	−0.258	0.225	−1.148	0.251	0.772	0.497–1.201
	pH	−0.388	0.145	−2.672	0.008	0.679	0.511–0.902
	Nitrate	0.316	0.135	2.335	0.02	1.372	1.052–1.788
	Phosphate	0.167	0.104	1.6	0.11	1.182	0.963–1.450
Plecoptera	Temperature	−0.183	0.126	−1.454	0.146	0.833	0.650–1.066
	TDS	−0.257	0.293	−0.880	0.379	0.773	0.436–1.372
	EC	0.015	0.267	0.054	0.957	1.015	0.601–1.713

	pH	0.052	0.186	0.278	0.781	1.053	0.731–1.517
	Nitrate	0.02	0.159	0.123	0.902	1.02	0.747–1.393
	Phosphate	0.152	0.129	1.178	0.239	1.164	0.904–1.499
Trichoptera	Temperature	−0.130	0.2	−0.652	0.514	0.878	0.594–1.298
	TDS	−1.134	0.506	−2.241	0.025	0.322	0.119–0.867
	EC	−0.593	0.459	−1.293	0.196	0.552	0.225–1.358
	pH	−0.950	0.326	−2.911	0.004	0.387	0.204–0.733
	Nitrate	0.084	0.28	0.299	0.765	1.087	0.628–1.882
	Phosphate	−0.050	0.273	−0.182	0.856	0.952	0.557–1.625

Note: Predictors were standardized before modelling. Estimate values are on the log-abundance scale. NB2 GLMM = negative binomial type 2 generalized linear mixed model; SE = standard error; Z = Wald z-statistic; *P* = p-value; TDS = total dissolved solids; EC = electrical conductivity; pH=Potential of hydrogen; IRR = incidence rate ratio, calculated as exp (Estimate); CI = confidence interval. IRR values represent the multiplicative change in expected abundance associated with a one-standard-deviation increase in the predictor, conditional on the other predictors and station random effect. The 95% CI represents the Wald 95% confidence interval for the IRR.

3.5. Plecoptera

Not any of the six physicochemical predictors showed a statistically reinforced association with Plecoptera abundance. The projected relations were negative for temperature and TDS and positive for EC, pH, nitrate and phosphate, but all confidence intervals included an IRR of 1. The station random intercept variance was 0.388, indicating residual station-level variation after accounting for the measured physicochemical predictors.

3.6. Trichoptera

Trichoptera abundance indicated statistically supported negative associations with TDS and pH. A one-SD increase in TDS was associated with an estimated 67.8% lower expected Trichoptera abundance (IRR = 0.322, 95% CI: 0.119–0.867). A one-SD increase in pH was associated with an estimated 61.3% lower expected Trichoptera abundance (IRR = 0.387, 95% CI: 0.204–0.733). Temperature, EC, nitrate, and phosphate were not statistically supported as predictors.

3.7. Model Diagnostics

The DHARMa simulation-based diagnostics showed tolerable residual homogeneity and no evidence of excess zero inflation for all three taxa. For Ephemeroptera, residual uniformity was supported ($p = 0.205$), and no excess zero increase ($p = 1.000$) or outliers ($p = 1.000$) were noticed. Nevertheless, the distribution test showed a modest residual-dispersion signal ($p = 0.044$). For Plecoptera, there was no evidence of non-uniform residuals ($p = 0.312$), residual dispersion ($p = 0.918$), or excess zero inflation ($p = 0.222$). For Trichoptera, there was no evidence of non-uniform residuals ($p = 0.750$), residual dispersion ($p = 0.454$) or excess zero inflation ($p = 1.000$). The high proportion of zero Trichoptera counts therefore did not, by itself, provide evidence that a zero-inflated model was required.

4. Discussion

4.1. Taxon-Specific Responses to Physicochemical Gradients

The outcomes show that the three EPT orders did not display a common response to the dignified physicochemical variables. Ephemeroptera abundance was related to TDS, pH, and nitrate; Trichoptera abundance was associated with TDS and pH; whereas Plecoptera showed no statistically supported association with any measured predictor.

The taxon-specific pattern is dependable with prior modelling studies showing that different EPT taxa can have diverse environmental relationships and preferences [17]. It also facilitates the value of analyzing individual taxa rather than treating EPT as a single homogeneous biological response.

The absence of a statistically supported association should not be deduced as evidence that a variable has no ecological importance. Somewhat, in the range of conditions sampled in River Kiminini, the existing data did not offer sufficient evidence of an association after concurrently accounting for the other measured predictors and station-level variation.

4.2. TDS and Ephemeroptera

TDS was adversely associated with Ephemeroptera abundance. A one-SD increase in TDS corresponded to an estimated 43.6% reduction in anticipated abundance.

The outcome is consistent with experimental and field evidence demonstrating that dissolved-solids concentration can impact freshwater macroinvertebrate performance and distributions [7]. TDS embodies the concentration of dissolved substances in water and can therefore echo changes in ionic conditions that might be physiologically pertinent to aquatic insects [18].

Though, TDS should not habitually be deduced as a straight measure of sediment disturbance. The current study did not measure suspended sediment or turbidity. The detected association consequently shows an association with dissolved-solids conditions rather than indicating that sediment disturbance caused the observed reduction in Ephemeroptera abundance.

The negative association should also be inferred in aggregation with the correlation between TDS and EC. Since TDS and EC were strongly correlated, their individual coefficients might partly echo overlapping information about the chemical environment.

4.3. pH and Ephemeroptera

Ephemeroptera abundance was also negatively associated with pH. The IRR of 0.679 describes that a one-SD upsurge in pH was associated with about 32% lower probable abundance, conditional on the other predictors. The outcome is supported by the broader literature demonstrating that aquatic insect propagations can deviate along pH inclines [7,19,20]. However, this analysis does not inaugurate a biological mechanism for the association. In specific, pH should not be inferred as a direct measurement of alkalinity since alkalinity was not measured in this study.

The association can also echo associations between pH and other environmental physiognomies not involved in the model. Consequently, the outcome is best inferred as a statistical association between pH and Ephemeroptera abundance in the sampled environmental range.

4.4. Nitrate and Ephemeroptera

Nitrate was positively associated with Ephemeroptera abundance. A one-SD upsurge in nitrate was associated with a 37.2% higher expected abundance. This association must be deduced carefully. It does not show that nitrate improvement directly benefits Ephemeroptera. Nitrate might covary with land use, primary production, organic matter availability, season, and other environmental conditions, as supported by [21]. The present dataset did not involve measurements of dissolved oxygen, primary productivity, habitat structure, flow, substrate, or land-use intensity that can be used to discrete these mechanisms. The positive link consequently classifies nitrate as a statistically informative environmental gradient in the sampled system rather than establishing a beneficial causal effect.

4.5. Trichoptera Responses

Trichoptera abundance demonstrated a negative relationship with both TDS and pH. The TDS link was moderately strong: a one-SD escalation was associated with an estimated 67.8% lower anticipated abundance. The consistent pH association showed roughly 61.3% lower estimated abundance per one-SD increase.

The TDS outcome is largely well-matched with evidence that dissolved-solids concentrations could impact aquatic insect growth, survival, and distributions [22]. Previous EPT modelling has likewise validated taxon-specific relationships with physicochemical conditions [17].

These study analysis does not, conversely, establish that TDS or pH caused decreases in Trichoptera abundance. The connotations might epitomize direct physiological reactions, indirect effects through habitat or food resources, or correlations with unmeasured environmental conditions.

4.6. Plecoptera and the Absence of Supported Associations

No physicochemical predictor was statistically supported for Plecoptera. The outcome varies from the clearer associations observed for Ephemeroptera and Trichoptera. Numerous enlightenments are possible. Plecoptera abundance was moderately low as compared with Ephemeroptera, and 21.7% of Plecoptera observations were zero. The research also involved only 60 observations and six physicochemical predictors. Subsequently, the analysis might have had limited power to detect relatively small associations.

An additional prospect is that variables not measured in this study, like substrate composition, flow velocity, channel structure, habitat heterogeneity, or riparian condition, might clarify some of the spatial variation in Plecoptera. This clarification remains a hypothesis because these variables were not involved in the present dataset. The station random-effect variance for Plecoptera was 0.388, showing residual station-level variant after accounting for the measured physicochemical variables. This offers supplementary rationalization for future analyses that integrate habitat and spatial characteristics.

4.7. Multicollinearity and Interpretation of Physicochemical Predictors

The solid negative correlation between TDS and EC ($r = -0.828$) is significant when inferring individual coefficients. Although all VIFs were below 5, TDS and EC demonstrated greater predictor overlay than the other variables.

This does not invalidate the model, but it means that the coefficient for one chemical variable should not be construed as demonstrating a wholly independent environmental process. In particular, the negative TDS association should not be interpreted as evidence that TDS alone is responsible for variation in EPT abundance.

The use of standardized predictors permits the magnitude of associations to be compared across variables dignified in different units. Conversely, standardization does not remove correlation among predictors.

4.8. Statistical Implications for Freshwater Biomonitoring

The analysis shows the significance of matching the statistical distribution to the structure of ecological count data. All three taxa exhibited variance significantly greater than their means, and Pearson dispersion statistics from Poisson models were greater than 3 for all taxa. The negative-binomial models significantly reduced AIC relative to the matching Poisson models. Integrating station as a random effect was also supported by lower AIC for Ephemeroptera and Plecoptera relative to the corresponding fixed-effects negative-binomial models, while Trichoptera showed slight station-level variation.

The final NB2 specification was selected constantly because it provided the lowest AIC among the evaluated NB1 and NB2 mixed models. The enhancement was strongest for Ephemeroptera and more modest for Plecoptera and Trichoptera. These results support the use of negative-binomial mixed models when freshwater abundance data are overdispersed, and sampling involves repeated observations from the same

stations. They do not support a blanket recommendation that negative-binomial models should always be used. Model choice should instead be based on the distribution and sampling structure of the particular dataset.

4.9. Zero Observations and Model Diagnostics

Trichoptera had 56.7% zero observations, which initially raises the possibility of zero inflation. However, the DHARMa zero-inflation test did not designate important excess zeros. This discrepancy is significant because a high ratio of zeros does not mechanically infer a zero-inflated data-generating process. In the current data, the observed number of zeros was well-matched with the fitted negative-binomial model.

The Ephemeroptera model demonstrated a modest residual-dispersion signal in spite of otherwise satisfactory simulation-based diagnostics. The outcome designates that the model should not be described as having perfect fit. In its place, the NB2 GLMM offered a valuable representation of the data while retaining a diagnostic caveat for Ephemeroptera.

4.10. Regional Contribution and Study Limitations

The primary influence of this study is regional and practical rather than methodological. Negative-binomial mixed modelling is an established statistical method. The contribution is applying it to repeated EPT abundance observations from River Kiminini and identifying taxon-specific associations within River Kiminini, Kenya.

Numerous limitations in the study were:

Primary, the analysis comprised physicochemical variables but did not involve thorough habitat characteristics like substrate composition, flow velocity, channel morphology, or riparian condition. These variables could help expound variation that was not captured by water chemistry, particularly for Plecoptera.

Secondly, the analytical dataset did not retain exact sampling dates as variables. While repetitive sampling was embodied through the station random effect, explicit temporal autocorrelation could not be evaluated.

Moreover, the observational design does not permit causal implication. Connotations between physicochemical variables and abundance might echo direct reactions, indirect ecological pathways or confounding by unmeasured environmental factors.

Future research should consequently assimilate physicochemical measurements with habitat, land-use, hydrological and temporal variables. Larger datasets could also support spatial and temporal validation of predictive models.

5. Conclusions

The study found that EPT macroinvertebrate orders indicated different statistical associations with physicochemical gradients in River Kiminini.

Ephemeroptera abundance was negatively associated with TDS and pH and positively associated with nitrate. Trichoptera abundance was negatively associated with TDS and pH. No statistically supported associations were detected between the measured physicochemical variables and Plecoptera abundance.

The outcomes indicate that TDS and pH were significant statistical gradients associated with variation in Ephemeroptera and Trichoptera abundance in the sampled river system. The positive nitrate association with Ephemeroptera should not be interpreted as evidence of a beneficial causal effect because other environmental variables that might covary with nitrate were not measured.

The lack of supported physicochemical associations for Plecoptera suggests that additional variables, including physical habitat characteristics, may be important for explaining its distribution, though this remains a hypothesis requiring direct testing.

From a statistical point of view, the significant overdispersion of all three response variables facilitated negative-binomial modelling over Poisson regression. The recurrent sampling of the same stations further

justified a mixed-effects framework. The final NB2 GLMMs offered a consistent framework for approximating taxon-specific associations while accounting for station-level variation.

The study offers regional evidence that EPT abundance in River Kiminini is connected with physicochemical gradients in a taxon-specific manner. The results support the continuous use of EPT organisms in freshwater biomonitoring while emphasizing the need to assimilate water chemistry with habitat and temporal information in future ecological assessments.

6. Recommendations

Negative-binomial mixed models should be taken aboard when freshwater macroinvertebrate reactions are overdispersed count data and sampling includes repeated observations from the same sites. Model selection should rely on the observed distribution, sampling design, and diagnostic evidence rather than applying a single model universally.

For River Kiminini, future monitoring should syndicate physicochemical variables with measurements of substrate composition, flow, channel characteristics, riparian condition, land use and other habitat variables. Such an expanded dataset would provide a stronger basis for determining whether the associations identified in the present analysis persist after accounting for physical habitat and temporal variation.

Future studies should also retain exact sampling dates and, where adequate repeated observations are available, explicitly evaluate temporal autocorrelation and seasonal effects.

Supplementary Materials

The following supporting information can be found at: https://www.sciepublish.com/article/pii/1309, Table S1: Raw data of the sampling stations, physical chemical parameters and EPT abundance.

Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript, the authors used Deekseek.com to rate the flow of work, improve coherence, check grammatical errors and to generate codes of data analysis. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Author Contributions

C.N.B: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Visualization, Writing original draft, Review & Editing. G.S.M: Conceptualization, Data Curation, Investigation, Methodology, Project Administration, Software, Supervision & Validation. J.A.K: Conceptualization, Data Curation, Investigation, Methodology, Project Administration, Software, Supervision & Validation. L.S.M.G: Methodology, Project Administration, Software, Validation, Review & Editing.

Ethics Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The dataset analyzed in this study is provided as supplementary material in Table S1.

Funding

This research received no external funding.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

References

1. Lim S, Do Y. Macroinvertebrate conservation in river ecosystems: Challenges, restoration strategies, and integrated management approaches. *Entomol. Res.* **2023**, *53*, 271–290. DOI:10.1111/1748-5967.12665
2. Sripanya J, Vongsombath C, Vannachak V, Rattanachan K, Hanjavanit C, Mahakham W, et al. Benthic Macroinvertebrate Communities in Wadeable Rivers and Streams of Lao PDR as a Useful Tool for Biomonitoring Water Quality: A Multimetric Index Approach. *Water* **2023**, *15*, 625. DOI:10.3390/w15040625
3. Barasa CN, Simiyu Muse G, Khazenzi JA. Human activities influence on distribution of Ephemeroptera, Plecoptera, and Trichoptera (EPT) in a tropical riverine ecosystem in Kenya. *J. Freshw. Ecol.* **2025**, *40*, 2539125. DOI:10.1080/02705060.2025.2539125
4. McKenzie M, Brooks A, Callisto M, Collins AL, Durkota JM, Death RG, et al. Freshwater invertebrate responses to fine sediment stress: A multi-continent perspective. *Glob. Change Biol.* **2024**, *30*, e17084. DOI:10.1111/gcb.17084
5. Epele LB, Williams-Subiza EA, Bird MS, Boissezon A, Boix D, Demierre E, et al. A global assessment of environmental and climate influences on wetland macroinvertebrate community structure and function. *Glob. Change Biol.* **2024**, *30*, e17173. DOI:10.1111/gcb.17173
6. Nietch CT, Smucker NJ, Gains-Germain L, Peck CP, Guglielmi S, DeCelles S, et al. Using Single-Species and Whole Community Stream Mesocosm Exposures for Identifying Major Ion Effects in Doses Mimicking Resource Extraction Wastewaters. *Water* **2023**, *15*, 249. DOI:10.3390/w15020249
7. Olson JR, Hawkins CP. Effects of total dissolved solids on growth and mortality predict distributions of stream macroinvertebrates. *Freshw. Biol.* **2017**, *62*, 779–791. DOI:10.1111/fwb.12901
8. Lemessa F, Simane B, Seyoum A, Gebresenbet G. Assessment of the Impact of Industrial Wastewater on the Water Quality of Rivers around the Bole Lemi Industrial Park (BLIP), Ethiopia. *Sustainability* **2023**, *15*, 4290. DOI:10.3390/su15054290
9. Kopp DA, Stoddard JL, Hill RA, Doyle JM, Kaufmann PR, Herlihy AT, et al. Joint species distribution models reveal taxon-specific sensitivities to potential anthropogenic alteration. *Freshw. Sci.* **2023**, *42*, 268–284. DOI:10.1086/726283
10. Lee JW, Lee SW, Lee H, Park SR. Impact of Environmental Factors of Stream Ecosystems on Aquatic Invertebrate Communities. *Sustainability* **2025**, *17*, 1252. DOI:10.3390/su17031252
11. Bush A, Compson Z, Rideout NK, Levenstein B, Kattilakoski M, Hajibabaei M, et al. Replicate DNA metabarcoding can discriminate seasonal and spatial abundance shifts in river macroinvertebrate assemblages. *Mol. Ecol. Resour.* **2023**, *23*, 1275–1287. DOI:10.1111/1755-0998.13794
12. Khan MK, Hasan MT. A Poisson Regression Approach to Modeling Traffic Accident Frequency in Urban Areas. *Am. J. Interdiscip. Stud.* **2022**, *3*, 117–156. DOI:10.63125/wqh7pd07
13. Farida U, Fitriana L. Negative Binomial Regression Modeling of Stunting Cases in Central Java: Handling Overdispersion in Count Data. *ZERO J. Sains Mat. Dan Terap.* **2026**, *10*, 768–779. DOI:10.30829/zero.v10i2.29471
14. Wu R, Song Y, Head JR, Katz DSW, Peay KG, Shedden K, et al. Fungal Spore Seasons Advanced Across the US Over Two Decades of Climate Change. *GeoHealth* **2025**, *9*, e2024GH001323. DOI:10.1029/2024GH001323
15. Williams-Subiza E, Epele LB. Biodiversity surrogates in freshwater environments: What determines their effectiveness? *Sci. Total Environ.* **2026**, *1051*, 182289. DOI:10.1016/j.scitotenv.2026.182289
16. Brooks M, Kristensen K, Benthem K, Magnusson A, Berg C, Nielsen A, et al. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *R J.* **2017**, *9*, 378. DOI:10.32614/RJ-2017-066
17. Jerves-Cobo R, Everaert G, Iñiguez-Vela X, Córdova-Vela G, Díaz-Granda C, Cisneros F, et al. A Methodology to Model Environmental Preferences of EPT Taxa in the Machangara River Basin (Ecuador). *Water* **2017**, *9*, 195. DOI:10.3390/w9030195
18. Perera H, Jayawardana C, Chandrajith R. Freshwater salinisation: Unravelling causes, adaptive mechanisms, ecological impacts, and management strategies. *Environ. Monit. Assess.* **2024**, *196*, 1195. DOI:10.1007/s10661-024-13388-2
19. Arimoro FO, Keke UN. The intensity of human-induced impacts on the distribution and diversity of macroinvertebrates and water quality of Gbako River, North Central, Nigeria. *Energ. Ecol. Environ.* **2017**, *2*, 143–154. DOI:10.1007/s40974-016-0025-8

20. Pound KL, Larson CA, Passy SI. Current distributions and future climate-driven changes in diatoms, insects and fish in U.S. streams. *Global Ecol. Biogeogr.* **2021**, *30*, 63–78. DOI:10.1111/geb.13193
21. Larson JH, Bailey SW, Kreiling RM, Bartsch LA, Frost PC, Xenopoulos MA, et al. Comparison of sediment and water column nutrient processing rates in agricultural streams of contrasting buffer land use. *Ecosphere* **2024**, *15*, e4808. DOI:10.1002/ecs2.4808
22. Brent RN, Kunkel J, Tomek Z, Buchardt D, DeLisle PF, Sivers S. A Novel Approach to Developing Thresholds for Total Dissolved Solids Using Standardized and Experimental Toxicity Test Methods. *Environ. Toxicol. Chem.* **2022**, *41*, 2782–2796. DOI:10.1002/etc.5466