Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Drivers of zooplankton species richness in Chilean lakes: integrating Poisson regression and explainable machine learning

Patricio R. De los Ríos-Escalante

prios@uct.cl

Universidad Católica de Temuco

Hamdi Raissi

Pontificia Universidad Católica de Valparaíso

Research Article

Keywords: Zooplankton, Species richness, Poisson regression, Bootstrap, Trophic interactions, Top-down control, Bottom-up control, Freshwater ecology

Posted Date: May 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9709478/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Zooplankton species richness in freshwater ecosystems is regulated by the combined action of biotic interactions and environmental filtering. We analysed crustacean zooplankton richness in 56 Chilean lakes using a Poisson generalized linear model and a complementary Poisson-XGBoost model. The Poisson GLM was validated using residual simulation, dispersion tests, zero-inflation assessment, residual uniformity tests, and comparison with a Negative Binomial alternative. The model showed no evidence of overdispersion, zero inflation, or residual lack of fit, supporting the use of the Poisson distribution. Fish presence was the only statistically significant predictor of richness, increasing expected richness by approximately 67% relative to fishless lakes. Chlorophyll, depth, conductivity, and surface area did not show significant linear effects. In contrast, the Poisson-XGBoost model indicated that fish presence and conductivity were the most relevant predictors, suggesting that conductivity may operate through non-linear environmental filtering not fully captured by the GLM. Together, these results indicate that zooplankton richness in Chilean lakes is primarily associated with fish presence, while conductivity may play a secondary context-dependent role. This combined inferential and machine-learning framework provides a robust and interpretable approach for evaluating top-down and environmental drivers of lake zooplankton richness.

Introduction

Zooplankton communities in freshwater ecosystems are structured by the interplay between bottom-up (resource availability) and top-down (predation) processes (Carpenter et al. 1985; Dodson et al. 2000). Primary productivity, commonly approximated by chlorophyll concentration, is a key driver of species richness through its effects on energy availability and competitive interactions. Previous studies have shown that water quality variables can explain a large proportion of the variability in aquatic community composition, particularly in zooplankton, highlighting strong bottom-up control along trophic gradients (Dodson et al. 2000; Smith & Kirkwood, 2025).

In contrast, fish presence can exert strong top-down control on zooplankton communities through direct predation and indirect effects on species interactions and size structure (Brooks and Dodson, 1965; Jeppesen et al. 2010). Experimental evidence indicates that predator effects may vary depending on intraspecific traits, such as sexual dimorphism, leading to differential impacts on prey communities (Davenport et al. 2025). Moreover, predator effects are often context-dependent, potentially varying in both magnitude and direction under similar environmental conditions (Peacor et al. 2025).

Environmental drivers such as temperature, thermal variability, and extreme events further modulate zooplankton community dynamics by altering species composition, abundance, and dominance patterns (Devkota et al. 2025; Hubbard et al. 2025). In addition, anthropogenic disturbances, including nutrient enrichment and contamination, can reduce biodiversity and shift community structure (Larson et al. 2025).

The relative importance of top-down and bottom-up processes may vary across regions depending on environmental conditions and ecological context (Shurin et al. 2002; Chase, 2003). Therefore, comparative analyses across geographic regions provide an opportunity to assess the generality of these mechanisms.

In this study, we evaluated the relative effects of fish presence and chlorophyll concentration on zooplankton species richness in lakes from Chile using Poisson regression models and XGBoost procedures to assess the robustness of estimated effects.

Materials and Methods

Study data and preprocessing

The dataset included lakes located in Chile and contained information on zooplankton species richness (count data), fish presence/absence, and environmental variables including chlorophyll concentration, depth, conductivity, and surface area (Table 1; references are provided in the Supplementary Material).

Table 1
Data of Chilean lakes considered for present study (references see supplementary material).

Site	Surface (km^2)	Maximum Depth (m)	Chlorophyll a (µg/l)	Conductivity (µS/cm)	Fish presence/absence	Species number	References
Cejas	1	10	1	205.53	Absence	1	De los Ríos-Escalante, 2011
Paposo	1	0.1	1	7.4	Absence	1	De los Ríos-Escalante, 2011
Miscanti	13.4	9	1	7.73	Absence	3	De los Ríos-Escalante, 2011
Miniques	1.6	5	1	15.42	Absence	1	De los Ríos-Escalante, 2011
Rungue	0.48	15	65.9	0.33	Presence	4	Schmid-Araya & Zúñiga, 1992
Peñuelas	19	15	57.4	0.12	Presence	11	Schmid-Araya & Zúñiga, 1992
Lanalhue	32	26	5.2	0.01	Presence	5	De los Ríos-Escalante et al., 2011
Lleulleu	40	46	1.2	0.01	Presence	4	De los Ríos-Escalante et al., 2011
Icalma	10	135	0.8	0.01	Presence	3	De los Ríos-Escalante et al., 2011
Galletué	13	43	2.8	0.01	Presence	4	De los Ríos-Escalante et al., 2011
Negrita	0.1	1	0.3	2.7	Absence	5	De los Ríos & Roa, 2010
De_los_Patos	0.1	0.1	0.2	10.6	Absence	4	De los Ríos & Roa, 2010
Escondida	0.1	0.1	0.7	12.4	Absence	5	De los Ríos & Roa, 2010
Seca	0.1	0.1	0.5	12.3	Absence	5	De los Ríos & Roa, 2010

Site	Surface (km^2)	Maximum Depth (m)	Chlorophyll a (µg/l)	Conductivity (µS/cm)	Fish presence/absence	Species number	References
Negra	0.1	0.1	0.5	0.7	Absence	2	De los Ríos & Roa, 2010
Bella	0.1	0.1	0.2	0.7	Absence	2	De los Ríos & Roa, 2010
Los Pastos	0.1	0.1	0.5	0.5	Absence	4	De los Ríos & Roa, 2010
Vaca_Hundida	0.1	0.1	1.2	1.1	Absence	4	De los Ríos & Roa, 2010
Tinquilco	10	40	2	0.01	Presence	5	De los Ríos-Escalante et al., 2011
Verde	1	3	0.1	0.01	Absence	2	De los Ríos-Escalante et al., 2012a
Captren	1	6	0.1	0.01	Absence	3	De los Ríos-Escalante et al., 2012a
Villarrica	175.9	185	0.4	0.06	Presence	7	Campos et al., 1983
Pirihueico	30.5	145	0.6	0.05	Presence	5	Wölfl, 1996
Riñihue	77.5	323	1.2	0.06	Presence	7	Wölfl, 1996
Ranco	442.6	199	0.8	0.07	Presence	9	Campos et al., 1992a
Puyehue	165.4	123	2.1	0.08	Presence	7	Campos et al., 1989
Rupanco	235	273	1.2	0.06	Presence	4	Campos et al., 1992b
Llanquihue	870.5	317	0.5	0.11	Presence	3	Campos et al., 1988
Todos_los_Santos	178.5	335	0.4	0.04	Presence	4	Campos et al., 1990
Sargazo	1.397	10	0.9	0.01	Absence	4	De los Ríos-Escalante et al., 2012b
Chaiquenes	0.582	10	0.6	0.01	Absence	2	De los Ríos-Escalante et al., 2012b

Site	Surface (km^2)	Maximum Depth (m)	Chlorophyll a (µg/l)	Conductivity (µS/cm)	Fish presence/absence	Species number	References
Triangulo	0.577	10	1.4	0.01	Absence	2	De los Ríos-Escalante et al., 2012b
Cucao	10.6	25	3.4	156.85	Presence	5	De los Ríos-Escalante & Kies, 2018
Huillinco	19.1	47	2.7	27.33	Presence	7	De los Ríos-Escalante & Kies, 2018
Tarahuin	7.7	33	8.4	50.9	Presence	10	De los Ríos-Escalante & Kies, 2018
Natri	33.2	58	16.4	41.4	Presence	7	De los Ríos-Escalante & Kies, 2018
Tepuhueico	14.3	25	3.6	30.2	Presence	7	De los Ríos-Escalante & Kies, 2018
Los_Palos	59	5	0.8	0.02	Presence	4	Villalobos, 1999
Riesco	14.7	130	0.9	0.02	Presence	4	Villalobos, 1999
Escondida	43	7	0.5	0.01	Presence	5	Villalobos, 1999
Larga	0.1	5	0.9	3.45	Absence	6	Soto & De los Ríos, 2006
De_los_Cisnes	0.1	1	93.7	16.51	Absence	6	Soto & De los Ríos, 2006
Redonda	0.1	3	2.3	1.49	Absence	8	Soto & De los Ríos, 2006
Juncos	0.1	3	1.8	2.29	Absence	6	Soto & De los Ríos, 2006
Nordensjold	25	200	0.3	0.13	Presence	2	Soto & De los Ríos, 2006

Site	Surface (km^2)	Maximum Depth (m)	Chlorophyll a (µg/l)	Conductivity (µS/cm)	Fish presence/absence	Species number	References
Paso_de_la_Muerte	0.1	3	10	0.8	Absence	7	Soto & De los Ríos, 2006
Jovito	0.1	3	4	1.38	Absence	5	Soto & De los Ríos, 2006
Melliza_Este	0.13	25	3.2	0.66	Presence	8	Soto & De los Ríos, 2006
Meliza_Oeste	0.12	16	5.3	0.8	Presence	8	Soto & De los Ríos, 2006
Sarmiento	86	312	0.3	0.85	Presence	5	Soto & De los Ríos, 2006
Pehoe	15	200	3.2	0.2	Presence	2	Soto & De los Ríos, 2006
Del_Toro	196	300	0.4	0.07	Presence	4	Soto & De los Ríos, 2006
Kon Aiken	0.1	1	2.7	0.02	Absence	4	De los Ríos et al., 2008
Laredo 1	0.1	1	4.4	6.14	Absence	3	De los Ríos et al., 2008
Laredo 2	0.1	1	5.4	2.29	Absence	3	De los Ríos et al., 2008
Rio Seco	0.1	1	28.6	2.3	Absence	3	De los Ríos et al., 2008

Statistical modelling

Species richness was modelled as count data using a generalized linear model with a Poisson error distribution and a log-link function (McCullagh and Nelder, 1989; Ver Hoef and Boveng, 2007). Fish presence/absence was included as a categorical predictor, whereas chlorophyll-a concentration, maximum depth, conductivity, and lake surface area were standardized prior to analysis to allow direct comparison among regression coefficients. Model adequacy was evaluated through Pearson dispersion, simulated residual diagnostics using DHARMa, tests for dispersion, zero inflation, residual uniformity, and outliers, and assessment of multicollinearity using variance inflation factors. A Negative Binomial GLM was fitted as an alternative model to evaluate whether additional dispersion was required. Model performance was compared using AIC, AICc, BIC, Nagelkerke pseudo-R2, and RMSE. Incidence rate ratios were obtained by exponentiating model coefficients and are reported with 95% confidence intervals. Model fitting and coefficient extraction were conducted using base R (R Core Team, 2025) and the broom package (Robinson et al. 2026). Graphical outputs were generated using ggplot2 (Wickham, 2016).

Machine-learning modelling: Zooplankton species richness was analysed using an extreme gradient boosting model adapted for count data. The response variable was species richness, expressed as the number of zooplankton species recorded in each lake. Predictor variables included fish presence/absence and environmental variables such as conductivity and other limnological descriptors available in the dataset. Prior to modelling, the first column containing lake names was removed, and the response variable was separated from the predictor matrix. The dataset was divided into training and testing subsets using an 80:20 split with a fixed random seed to ensure reproducibility.

A Poisson objective function was used in the XGBoost regression model (count:poisson), which is appropriate for non-negative count data. Hyperparameters were optimized using Bayesian optimization implemented in Optuna. The optimized parameters included learning rate, maximum tree depth, L1 and L2 regularization terms, subsampling rate, column sampling rate, and number of estimators. Model performance during optimization was assessed using five-fold cross-validation and mean Poisson deviance as the loss function. The final model was fitted using the best hyperparameter combination identified by Optuna. Predictor importance was evaluated using XGBoost importance metrics, including weight, gain and cover. In addition, SHAP values were calculated to interpret the contribution of each predictor to model predictions. SHAP summary plots were used to visualize both global predictor importance and the direction of predictor effects.

All analyses were conducted in Python using pandas, numpy, xgboost, optuna, scikit-learn, and shap (Harris et al. 2020; Akiba et al. 2019; Lundberg and Lee, 2017; Chen and Guestrin 2016; Pedregosa et al. 2011; McKinney, 2010). The machine-learning approach was used as a complementary analysis to evaluate non-linear relationships and to assess the robustness of patterns identified by the generalized linear models.

Results

Model comparison supported the Poisson generalized linear model as the most parsimonious and statistically adequate approach for modelling zooplankton species richness (Table 2). The Poisson GLM explained a moderate proportion of richness variability (Nagelkerke pseudo-R2 = 0.363; Table 2). Residual diagnostics revealed no evidence of overdispersion, zero inflation, residual non-uniformity, outliers, or problematic multicollinearity (Table 3).

Table 2
Validation and comparison of count models for zooplankton species richness.

Model	AIC	AICc	BIC	Nagelkerke R2	RMSE	Dispersion	Conclusion
Poisson GLM	241.61	243.3	253.76	0.363	1.933	0.892	Supported
Negative Binomial GLM	243.61	245.9	257.79	0.363	1.933	NA	Not improved; theta ≈ 78,230
Quasi-Poisson	NA	NA	NA	NA	NA	0.892	Similar inference; no AIC available

Table 3
Simulated residual diagnostics for the Poisson GLM.

Diagnostic	Result	Interpretation
Pearson dispersion	0.892	No overdispersion
DHARMa dispersion test	p = 0.384	No significant dispersion problem
Zero-inflation test	p = 0.848	No evidence of excess zeros
Residual uniformity test	p = 0.912	Residuals consistent with expected distribution
Outlier test	p = 1.000	No significant outliers
Variance inflation factor	1.04–2.26	No problematic multicollinearity

Fish presence was the only statistically significant predictor in the Poisson GLM, increasing expected zooplankton richness by approximately 67% relative to fishless lakes (IRR = 1.67, p = 0.0007; Table 4). Chlorophyll concentration, conductivity, depth, and lake surface area did not show significant linear effects on richness.

Table 4
Poisson GLM coefficients expressed as incidence rate ratios.

Predictor	Beta	IRR	95% CI	p-value
Fish presence	0.514	1.67	1.24–2.25	0.0007
Chlorophyll (+ 1 SD)	0.078	1.08	0.96–1.20	0.160
Depth (+ 1 SD)	-0.142	0.87	0.72–1.05	0.140
Conductivity (+ 1 SD)	-0.049	0.95	0.82–1.08	0.488
Surface area (+ 1 SD)	0.042	1.04	0.89–1.20	0.589

The complementary Poisson-XGBoost analysis identified fish presence and conductivity as the most influential predictors of model performance (Table 5). SHAP analyses further indicated that conductivity contributed through heterogeneous non-linear responses among lakes, whereas fish presence exerted a more consistent directional effect across observations (Figs. 1–2). Differences between SHAP importance rankings and XGBoost gain metrics likely reflect distinct aspects of predictor contribution, where SHAP values capture overall prediction sensitivity whereas gain reflects tree-splitting optimization during model fitting. Observed and predicted richness values showed reasonable agreement, supporting the overall adequacy of the Poisson GLM for describing richness variability across lakes (Fig. 3). The discrepancy between GLM significance and SHAP importance for chlorophyll may indicate weak or non-linear productivity effects that are not adequately represented by linear parametric models.

Table 5
Predictor importance from the Poisson-XGBoost model.

Predictor	Weight	Gain	Cover	Relative gain (%)
Fish presence	363	0.531	329.986	60.15
Conductivity	3382	0.352	120.721	39.85
Chlorophyll, depth, surface area	Low/negligible	Low/negligible	Low/negligible	Low/negligible

Discussion

Our results revealed clear differences in the mechanisms structuring zooplankton richness across Chilean lakes. In Chilean lakes, fish presence had a strong and statistically significant positive effect on species richness, whereas chlorophyll showed a weaker and non-significant effect. This pattern suggests a predominance of top-down control, consistent with classical trophic cascade theory, where predation reduces competitive exclusion and promotes species coexistence (Brooks & Dodson, 1965; Carpenter et al. 1985).

This finding is consistent with previous experimental and observational studies showing that predators can indirectly influence biodiversity through trophic interactions (Davenport et al. 2025). Furthermore, the magnitude and direction of predator effects can depend strongly on environmental context, reinforcing the idea that these systems may indicate stronger top-down influence, although such effects are not universal and remain system-specific (Peacor et al. 2025).

The obtained results in the present study suggest a stronger role of top-down control associated with fish presence, whereas environmental variables such as conductivity and chlorophyll may exert secondary context-dependent influences on zooplankton richness patterns (Soto and Zuñiga, 1991; De los Ríos-Escalante, 2015). Similar patterns have been reported in studies where abiotic variables explain most of the variation in aquatic communities, particularly across trophic gradients (Smith and Kirkwood, 2025).

Although chlorophyll was not statistically significant in the GLM, its negative coefficient and elevated SHAP contribution may suggest weak or context-dependent productivity effects potentially associated with eutrophication processes, where increased

productivity leads to dominance by a few opportunistic species, reducing overall diversity (Dodson et al. 2000). This interpretation is consistent with studies showing declines in biodiversity under environmental stress and contamination (Larson et al. 2025).

Additionally, environmental factors such as temperature variability can influence zooplankton communities, potentially contributing to regional differences in community responses (Devkota et al. 2025; Hubbard et al. 2025). These findings highlight the importance of considering multiple interacting drivers when evaluating biodiversity patterns.

Overall, our results support the hypothesis that the relative importance of top-down and bottom-up processes varies among ecosystems, consistent with previous comparative studies (Shurin et al. 2002; Chase, 2003). Chilean lakes appear to be primarily structured by biotic interactions, whereas North American lakes are more strongly influenced by environmental conditions (Devkota et al. 2025; Hubbard et al. 2025).

These findings have important implications for ecosystem management. In systems where top-down processes predominate, fish community management may play a key role in maintaining biodiversity, although such effects are context-dependent and not universally observed. Conversely, in systems driven primarily by bottom-up processes, strategies focused on water quality and nutrient control are likely to be more effective.

Declarations

Funding

state: The present study was financed by CONICYT-Chile (Doctoral fellowship, Doctoral grant for support doctoral thesis, and grant finish doctoral thesis), also it was financed by projects DID-UACH d2001-11, IAI (Enhanced ultraviolet B radiation in natural ecosystems as an added perturbation due to ozone depletion), DGI-UCT CDA 2007-01, Tides Grant Foundation TRF13-03011, and MECESUP UCT 0804.

Author Contribution

PDE: wrote the main manuscript text and H.R., contributed to data analysis and interpretation, and revision of final version of the manuscript.

Acknowledgement

The present study was financed by CONICYT-Chile (Doctoral fellowship, Doctoral grant for support doctoral thesis, and grant finish doctoral thesis), also it was financed by projects DID-UACH d2001-11, IAI (Enhanced ultraviolet B radiation in natural ecosystems as an added perturbation due to ozone depletion), DGI-UCT CDA 2007-01, Tides Grant Foundation TRF13-03011, and MECESUP UCT 0804. The authors express their gratitude National Forestry Corporation (CONAF-Chile) for access authorizations for protected areas. Finally, the main author expresses their gratitude to M.I. and S.M.A for their valuable suggestions which improved the manuscript.

References

1. Akiba T, Sano S, Yanase T, Ohta T, Koyama M (2019) Optuna: A Next-Generation Hyperparameter Optimization Framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, 4–8 August 2019, 2623–2631. https://doi.org/10.1145/3292500.3330701
2. Brooks JL, Dodson SI (1965) Predation, body size, and composition of plankton. Science 150:28–35. https://doi.org/10.1126/science.150.3692.28
3. Canty A, Ripley B (2024) boot: Bootstrap functions (original by Davison & Hinkley, 1997)

4. Carpenter SR, Kitchell JF, Hodgson JR (1985) Cascading trophic Interact BioScience 35:634–639. 10.2307/1309989
5. Chase JM (2003) Strong and weak trophic cascades. Oikos 101:187–195. https://doi.org/10.1034/j.1600-0706.2003.12062.x
6. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 785–794. https://doi.org/10.1145/2939672.2939785
7. Davenport JM, Babineau AM, Sloan RK, Groesbeck A, Montazeri AJ, Ramey M (2025) Sexual Dimorphic Effects of a Keystone Predator on Prey Communities. Ecology 106:e4530. https://doi.org/10.1002/ecy.4530
8. De los Rios-, Escalante P (2015) Fish predation effects on body length of planktonic cladocerans and copepods in Chilean lakes. Crustaceana 88:1193–1199. https://doi.org/10.1163/15685403-00003472
9. Devkota N, Salis RK, Hansson LA (2025) Warming of shallow temperate lakes: consequences for rotifer community composition and population dynamics. Hydrobiologia 852:971–985. https://doi.org/10.1007/s10750-024-05744-7
10. Dodson SI, Arnott SE, Cottingham KL (2000) The relationship in lake communities between primary productivity and species richness. Ecology 81:2662–2679. https://doi.org/10.1890/0012-9658(2000)081[2662:TRILCB]2.0.CO;2
11. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Fernandez J, Wiebe M, Peterson P, Gerard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE (2020) Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2
12. Hubbard JR, Albini D, Jackson MC (2025) Daily temperature variation explains zooplankton community differences in freshwater mesocosms recovering from a natural drought. Limnol Oceanogr 70:S300–S314. https://doi.org/10.1002/lno.70181
13. Larson JH, Lowe MR, Bailey SW, Bell AH, Cleveland DM (2025) Effect of copper mill waste material on benthic invertebrates and zooplankton diversity and abundance. PLoS ONE 20(3):e0318980. https://doi.org/10.1371/journal.pone.0318980
14. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Adv Neural Inf Process Syst 30
15. McCullagh P, Nelder JA (1989) Generalized Linear Models. Chapman & Hall
16. McKinney W (2010) Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, Austin, 28 June-3 July 2010, 56–61. https://doi.org/10.25080/Majora-92bf1922-00a
17. Peacor SD, Cressler CE, Pangle KL, Rafalski AV, Song C, Werner EE (2025) Similar conditions with opposite effects: predation-risk effects on prey abundance are highly contingent. Ecol Evol 15:e70861. 10.1002/ece3.70861
18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Perrot M, Duchesnay E (2011) Scikit-learn: Machine Learning in Python. J Machin Learn Res 12:2825–2830
19. R Core Team (2025) R: A language and environment for statistical computing. R Foundation for Statistical Computing
20. Robinson D, Hayes A, Couch S, Hvitfeldt E (2026) broom: Convert Statistical Objects into Tidy Tibbles (R package version 1.0.11). https://github.com/tidymodels/broom
21. Shurin JB, Borer ET, Seabloom EW, Anderson K, Blanchette CA, Broitman B, Cooper SD, Halpern BS (2002) A cross-ecosystem comparison of the strength of trophic cascades. Ecol Lett 5:785–791. https://doi.org/10.1046/j.1461-0248.2002.00381.x
22. Smith ED, Kirkwood AE (2025) Joint species distribution modelling of multiple taxonomic groups shows that water quality explains most of the variation in aquatic community composition along a lake trophic gradient. Aquat Ecol 59:823–835. https://doi.org/10.1007/s10452-025-10198-7
23. Soto D, Zúñiga L (1991) Zooplankton assemblages of Chilean temperate lakes: a comparison with North American counterparts. Rev Chil Hist Nat 64:569–581
24. Ver Hoef JM, Boveng PL (2007) Quasi-Poisson vs. Negative binomial regression: how should we model overdispersed count data? Ecology 88:2766–2772. https://doi.org/10.1890/07-0043.1

25. Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer

Figures

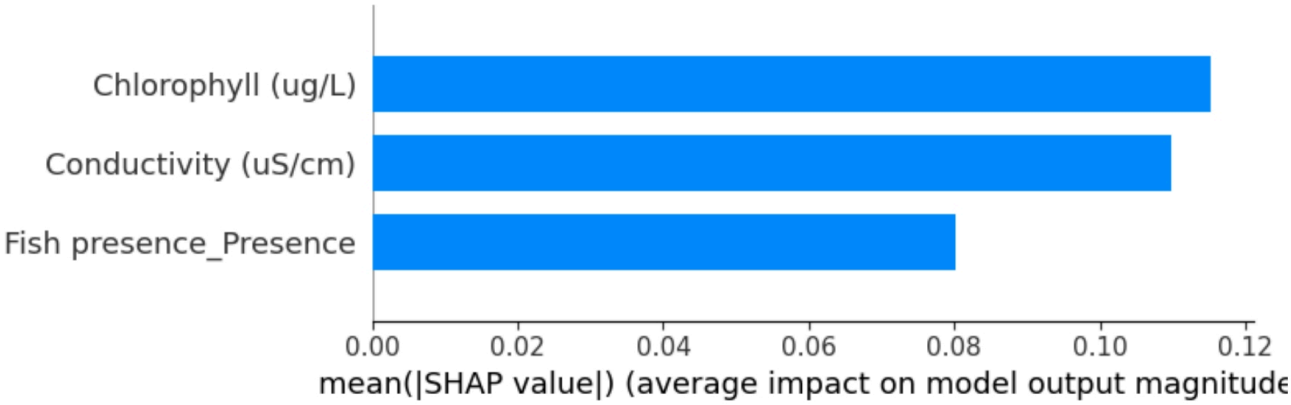


Figure 1

Global predictor importance based on SHAP values for the Poisson XGBoost model. Bars represent the mean absolute SHAP value for each predictor, indicating its overall contribution to model predictions of zooplankton species richness.

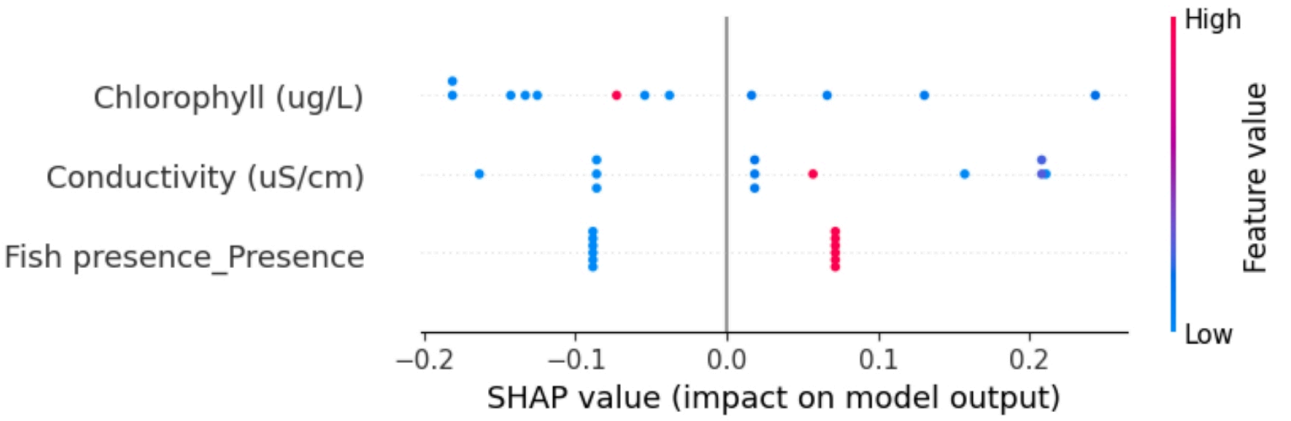


Figure 2

SHAP summary plot for the testing dataset. Each point represents an individual observation, showing the contribution of each predictor to predicted zooplankton richness. Positive SHAP values increase predicted richness, whereas negative values decrease it.

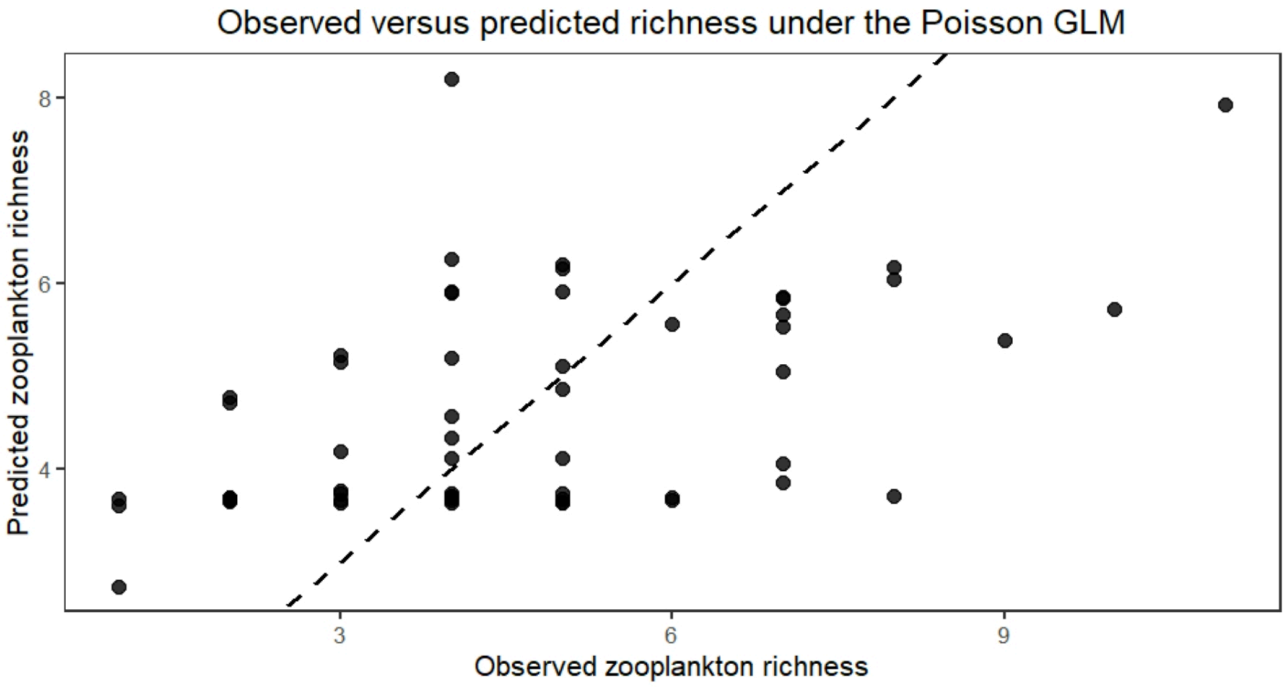


Figure 3

Observed versus predicted zooplankton species richness under the Poisson generalized linear model. The dashed line represents the 1:1 relationship between observed and predicted values. The overall distribution of points indicates a reasonable agreement between model predictions and observed richness values across Chilean lakes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx