# LAKES WITHOUT WELL-DEFINED OUTLETS – TOPOGRAPHIC CATCHMENTS AND EFFORTS

Scientific Report from DCE – Danish Centre for Environment and Energy No. 698 2026

AARHUS UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

# LAKES WITHOUT WELL-DEFINED OUTLETS – TOPOGRAPHIC CATCHMENTS AND EFFORTS

Scientific Report from DCE – Danish Centre for Environment and Energy No. 698 2026

Tuba Bucak
Robert Ladwig
Eti E. Levi
Liselotte Sander Johansson
Ane Kjeldgaard
Martin Søndergaard

Aarhus University, Department of Ecoscience

AARHUS UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

# Data sheet


| | |
|---|---|
| Series title and number: | Scientific Report from DCE – Danish Centre for Environment and Energy no. 698 |
| Category: | Scientific advisory report |
| Title: | Lakes without well-defined outlets – topographic catchments and efforts |
| Authors: | Tuba Bucak, Robert Ladwig, Eti E. Levi, Liselotte Sander Johansson, Ane Kjeldgaard, Martin Søndergaard |
| Institution: | Aarhus University, Department of Ecoscience |
| Publisher: | Aarhus University, Danish Centre for Environment and Energy © |
| URL: | https://dce.au.dk/en |
| Year of publication: | February 2026 |
| Editing completed: | 9 February 2026 |
| Referee: | Mikkel Boel |
| Quality assurance, DCE: | Signe Jung-Madsen |
| Linguistic QA: | Anne Mette Poulsen |
| External comments: | The comments can be found here: |
| Financial support: | Styrelsen for Grøn Arealomlægning og Vandmiljø (SGAV) |
| Please cite as: | Bucak, T., Ladwig, R., Levi, E.E., Johansson, L.S., Kjelgaard, A., Søndergaard, M. 2026. Lakes without well-defined outlets – topograhic catchments and efforts. Aarhus University, DCE – Danish Centre for Environment and Energy, 47 pp. – Scientific report no. 698 |





| | |
|---|---|
| Abstract: | Many Danish lakes included in the 3rd River Basin Management Plan (VP3, 2021–2027) lack well-defined catchments, preventing load assessments and targeted restoration despite failing to achieve good ecological status. Using multivariate statistics (RDA) and machine-learning approaches (Random Forest and Conditional Inference Trees), this study evaluated how landscape characteristics influence lake water quality, ecological status, and nutrient limitation. All methods showed low explanatory power, indicating that lake conditions are only partly explained by surface catchment characteristics and are likely influenced by internal loading, historical pressures, and hydrological processes. Nevertheless, poorer ecological conditions were consistently linked to organic soils, intensive drainage, and wetter catchments, while non-cultivated land and deeper lakes were associated with improved ecological status. These findings suggest that reducing artificial drainage and protecting natural areas can support improvements in water quality and ecological status. |



| | |
|---|---|
| Keywords: | Outlet, ecological status; nutrient limitation; landscape characteristics; drainage; Random Forest; Conditional Inference Trees; RDA |
| Front page photo: | Lake Sem Sø near Skrødstrup, Northern Jutland (Source: GeoDanmark Orthophoto (spring 2025)) |
| ISBN: | 978-87-7648-030-1 |
| ISSN (elektronic): | 2244-9981 |
| Number of pages: | 47 |
| Supplementary information: | The report is also published in Danish – see it here: |

# Content

# Preface

This report examines how landscape characteristics might affect Danish lakes without well-defined topographic catchment boundaries. The analyses are conducted for both estimated catchments and for different buffer regions around the lakes to elucidate the relationship between landscape characteristics, lake water quality, and ecological status. The report was prepared by Aarhus University at the request of The Danish Agency for Green Transition and Aquatic Environment (Styrelsen for Grøn Arealomlægning og Vandmiljø, SGAV), which had the opportunity to comment on a draft report.

As part of the project deliverables, SGAV has received the background data and calculations on which the report is based for each individual lake.

This report is partly an update of an earlier publication by Søndergaard et al. (2023).

# Sammenfatning

Mange søer i Danmark, som er omfattet af den tredje vandområdeplan (VP3) 2021–2027, i alt 418 ud af 985 søer, har ikke et klart defineret opland på grund af manglende eller dårligt definerede tilløb og afløb. Som følge heraf har det ikke været muligt at gennemføre belastningsopgørelser og dermed heller ikke at estimere den nødvendige indsats for at forbedre disse søers økologiske tilstand, selv om størstedelen af dem ikke opfylder målet om god økologisk tilstand. Et tidligere studie udført af Søndergaard et al. (2023) viste, at der kun var en svag sammenhæng mellem de nuværende oplandskarakteristika og den økologiske tilstand i disse søer.

På baggrund af ovenstående er analyserne i dette studie blevet opdateret i forhold til Søndergaard et al. (2023) ved anvendelse af en kombination af multivariate statistiske metoder (redundansanalyse, RDA) og maskinlæringsmetoder (Random Forest, RF og Conditional Inference Trees, CIT) med henblik på at undersøge vigtigheden af landskabskarakteristika for vandkvalitet, økologisk tilstand og klassificering af næringsstofbegrænsning. Tilsammen giver disse metoder komplementære indsigter i de potentielle effekter af landskabskarakteristika på søernes næringsstofniveauer og økologiske tilstand. Derudover inkluderede analyserne flere bufferstørrelser (10, 50, 100 og 250 m) for at kunne sammenligne med – eller understøtte – resultaterne baseret på de estimerede oplande.

Alle tre analytiske tilgange havde en forholdsvis lav forklaringsevne, hvilket stemmer overens med det tidligere studie baseret på lineær regression (Søndergaard et al., 2023). Den lave forklaringsevne tyder på, at andre faktorer også spiller ind, såsom intern næringsstofbelastning, historiske næringstilførsler, vandsøjlens opblandingsdynamik og grundvandsbidrag, som ikke opfanges ved overfladebaserede oplandsafgrænsninger og -karakteristik.

Alle tilgange identificerede de samme landskabskarakteristika som centrale determinanter for søernes økologiske tilstand (baseret på næringsstofkoncentrationer eller økologisk status). Dårligere økologisk tilstand (højere næringsstofkoncentrationer og ringe/dårlig økologisk status) var forbundet med højere andele af organiske jordtyper, større drænede arealer og vådere oplande. Omvendt var større roddybde og højere andele af ikke-dyrkede arealer forbundet med bedre økologisk tilstand.

RF indikerede endvidere, at dybere og større søer med lav sammenhæng mellem vandløbsafstrømning og nedbør er karakteriseret ved en forholdsvis god økologisk tilstand. CIT understøttede disse resultater og identificerede klare tærskler, idet dybere søer (især >5,7 m) eller lavvandede søer med >60 % ikke-dyrkede arealer havde lavere næringsstofkoncentrationer og en relativt god økologisk tilstand. CIT viste desuden, at fosforbegrænsede søer hyppigst forekom i sandede, svagt drænede oplande, mens kvælstofbegrænsede søer forekom i stærkt drænede systemer.

Samlet set viser resultaterne, at selv om landskabskarakteristika kun forklarer en del af variationen i søernes tilstand, spiller processer på oplandsniveau alligevel en væsentlig rolle for næringsstofniveauer og økologisk tilstand i tilsyneladende hydrologisk isolerede søer. På baggrund af vores resultater kan søer uden veldefinerede afløb have gavn af at kunstig dræning bliver reduceret samt af at øge eller beskytte naturlige (ikke-dyrkede) arealer. Da højere næringsstofniveauer var forbundet med intensiv dræning og organiske jordtyper, kan søer beliggende i sådanne oplande forventes at reagere bedre på målrettede indsatser. Andre faktorer som fx intern belastning, samspil med grundvand og søspecifikke processer vil dog sandsynligvis også være af betydning. Palæolimnologiske analyser kan desuden bidrage til at identificere historiske drivkræfter bag økologiske ændringer (og dermed forklare noget af den resterende uforklarede variation), hvilket kan understøtte beslutningsprocesser og mere målrettede indsatser.

# Summary

Many lakes in Denmark included in the 3rd River Basin Management Plan (VP3) 2021-2027, 418 out of 985 lakes, lack a clearly defined catchment due to missing or poorly defined inflows and outflows. As a result, load assessments could not be carried out, and thus it has not been possible to estimate the effort needed to improve their ecological conditions, even though the majority of these lakes fail to meet the environmental target of good ecological status. A previous study by Søndergaard et al. (2023) found that the current catchment characteristics and ecological status of these lakes only showed weak correlations.

Considering the above information, the analysis in the current study was updated compared to Søndergaard et al. (2023) using a combination of multivariate statistical (redundancy analysis, RDA) and machine learning approaches (Random Forest, RF, and Conditional Inference Trees, CIT) to investigate the influence of landscape characteristics on water quality, ecological status, and nutrient limitation classification. Together, these methods provide complementary insights into the potential effects of landscape features on lake nutrients and ecological status. Additionally, the analyses were performed using several buffer zones (10, 50, 100, and 250 m) to compare with, or support, the results obtained from the estimated catchment area.

All three analytical approaches had relatively low explanatory power, which was consistent with the previous linear regression-based study (Søndergaard et al., 2023). This low explanatory power likely points to the influence of additional factors, such as internal nutrient loading, historical nutrient inputs, mixing dynamics, and groundwater inputs, which were not captured by surface-based catchment delineations.

All approaches identified the same set of landscape characteristics as key determinants of lake ecological conditions (based on nutrient concentrations or ecological status). Poorer ecological conditions (higher nutrient concentrations and Poor/Bad ecological status) were associated with higher proportions of organic soils, larger drained areas, and wetter catchments. In contrast, greater root depth and larger proportions of non-cultivated land were linked with improved ecological conditions.

RF further indicated that deeper and larger lakes with low streamflow precipitation elasticity are characterized by relatively good ecological conditions. CIT supported these findings and suggested clear thresholds, showing that deeper lakes (especially >5.7 m), or shallower lakes with >60 % non-cultivated land, had lower nutrient concentrations and relatively good ecological status. CIT also showed that P-limited lakes were most common in sandy, weakly drained catchments, whereas N-limited lakes occurred in heavily drained systems.

In summary, although landscape characteristics explain only part of the variation in lake conditions, catchment-scale processes still play a major role in shaping nutrient levels and ecological status in hydrologically isolated lakes. Based on our results, lakes without well-defined outlets could benefit from reducing artificial drainage and increasing or protecting natural (non-cultivated) areas. Because higher nutrient levels were associated with intensive drainage and organic soils, lakes situated in such catchments may respond better to targeted interventions. However, other factors, such as internal

loading, groundwater interactions, and lake-specific processes are also likely to be important. Palaeolimnological analyses may further help identify historical drivers of ecological change (thus explaining some of the remaining unexplained variation), thereby contributing to decision-making and targeted interventions.

# 1 Background and objectives

## 1.1 Background

Approximately half of all lakes included in the review of the 3rd River Basin Management Plan 2021-2027 (VP3) are lakes without well-defined catchment areas within the size range 1-629 ha, and most of these (approximately 80%) do not meet the environmental objective of good ecological status. In VP3, it has not been possible to calculate the required effort to achieve improved ecological conditions in these lakes, because a prerequisite for preparing a load assessment is that a topographic catchment area can be determined.

In connection with a former report from DCE on lakes without well-defined catchment areas (Søndergaard et al., 2023), the topographic catchment area was estimated for 418 lakes. However, the report emphasized that the lake catchment estimates were uncertain, and analyses of correlations between current catchment characteristics and the lake ecological status only showed weak correlations. The recommendations in the report included, among others, assessment of the importance of phosphorus and nitrogen in the lakes and, in connection with preliminary studies of lakes designated for restoration, undertaking of palaeolimnological analyses to describe the lakes' development and previous conditions.

## 1.2 Objectives

The purpose of this project is to apply new methods to re-examine the empirical correlations identified by Søndergaard et al. (2023) between lake ecological status and land use in both estimated catchments and buffer zones of varying width around the individual lakes. These analyses, together with results from an ongoing project on the importance of nitrogen in lakes, will form the basis for assessing which efforts may be relevant to implement to improve the ecological quality of these lakes.

The project's results and conclusions can contribute to SGAV's efforts to identify the need for action for lakes without well-defined catchments and outlets, as well as to determine measures required to achieve the environmental goal of good ecological status in the River Basin Management Plans 2027-2033.

# 2 Data

## 2.1 In-lake data

A list of the lakes in question was provided by SGAV. Water chemistry data were obtained from the national lake monitoring program, NOVANA, stored in the VanDa database (https://vanda.miljoeportal.dk/) and accessed through Overfladevanddatabasen (ODA – ODA.dk). For lakes without well-defined catchment areas, total phosphorus (TP), total nitrogen (TN), and chlorophyll *a* (Chla) concentrations were calculated as yearly summer means (May–September) for use in subsequent analyses. The lake data used in the analysis covered the period 2010-2023. For further analysis, the average summer value of each in-lake parameter over the years was used for each lake. The ecological status classes of the lakes were based on Chla values provided by SGAV. This classification does not necessarily reflect the final ecological state as it does not consider the other biological quality elements required under the Water Framework Directive.

## 2.2 Landscape characteristics

### Land-use/field data

In this report we used the *Fields and Field Blocks* dataset (IMK-markkort) from the Danish Agricultural Agency, which provides official delineations of agricultural field blocks in Denmark. Data is available as from 2010 and has been updated every year since 2018. The data covers > 350 field block classes. However, in this project, we have aggregated these into four, following the approach outlined in Søndergaard et al. (2023). Short names used in the analysis are given in parentheses:

**Non-cultivated land:** These areas are not cultivated, neither in rotation nor with permanent grass. Forest and natural areas are also included in this class (*non-cultivated*).

**Cultivated land in rotation:** These areas are regularly cultivated by ploughing or harrowing (1–2-year crop rotations), including areas used for vegetable production and horticulture (*agriculture*).

**Permanent grassland:** These areas are never turned over, i.e. not ploughed nor harrowed. The grass must be grazed or mowed annually so that it does not turn into bushes (*grassland*).

**Agri-environment management areas:** These areas cover a variety of conditions, such as the maintenance of grassland and natural areas, typically involving grazing and no fertilization; environmentally friendly agricultural practices, usually reduced fertilization, maintenance of wetlands, and upkeep of modified drainage systems, which may also include wet cultivation (*agri-environment*).

### Soil data

The Danish 10x10m soil type maps used in the current study (version 1.1; Møller et al., 2024) were produced by DCA, Aarhus University. We used

‘JB2024_opdelt’ data, which includes a revised map of topsoil (0–30 cm) soil types. In this version, the JB4 class was further subdivided based on subsoil texture, following the classification described in Greve (2024). The original soil classes in the database were aggregated into the main classes:

**Clay:** Clayey soil and heavy clayey soil or silty soil

**Sand:** Coarse sand, fine sand, clay mixed sandy soil

**Organic:** Humic, organic-rich soil

### Potential drained areas

For calculation of the proportion of the catchment that is likely to be drained (%), a dataset produced by Møller et al. (2018a) was used. The dataset has a resolution of 30 m, and drainage percentages are calculated as an area-weighted average of the catchment land area, excluding lake area.

### Slope

Slope of the catchment and buffer regions was calculated based on DHM (Danmarks HøjdeModel) 10 m resolution data. Average catchment and buffer slopes were assigned to each lake.

### CAMELS-DK dataset

To integrate catchment characteristics, we used the CAMELS-DK dataset (Liu, et al., 2025). The CAMELS (Catchment Attributes and Meteorological Time Series for Large Samples) dataset is a framework to advance hydrological and catchment analysis (Addor et al., 2017; Coxon et al., 2017; Höge et al., 2023) and has been developed for various regions world-wide. CAMELS-DK was developed by researchers from GEUS and Aarhus University using data from DMI, GEUS, VanDa, and BASEMAP, as well as open-source datasets such as CORINE Land Cover. The dataset includes 304 gauged basins and 3026 un-gauged catchments, covering a total of 3300 Danish ID-15 catchments.

Dynamic catchment attributes include climate variables derived from gridded daily datasets provided by the Danish Meteorological Institute (DMI), stream-flow observations from the Danish surface water database (VanDa), and simulated data from the DK-model, such as streamflow and groundwater heads. The dataset also includes soil, geology, land-use and topography data for each ID-15 catchment. In this project, we included only climate, hydrology, and soil attributes in our analysis.

For CAMELS-DK variables, we did not compute separate values for each buffer distance (10–250 m) because of the resolution of CAMELS-DK dataset which is ~15 km$^2$. We therefore extracted CAMELS-DK values once using a 1000 m buffer around each lake, which was sufficient to intersect the relevant CAMELS catchment unit(s). The resulting area-weighted CAMELS-DK values were then used uniformly for all buffer distances (10, 50, 100, 250), as well as for the catchment-scale variable set, because they do not vary with buffer distance. A list of variables from the CAMELS-DK dataset and their descriptions are given in Table 2.1.

**Table 2.1.** Overview of catchment and streamflow variables derived from the CAMELS-DK dataset, including precipitation, soil, and hydrological indices used in the study.

| Variable | Description |
|---|---|
| high_prec_freq | Frequency of high precipitation days (≥5 times mean daily precipitation) (day* $yr^{-1}$). |
| high_prec_dur | Average duration of high precipitation events (number of consecutive days ≥5 times mean daily precipitation). |
| p_seasonality | Seasonality and timing of precipitation, where positive values indicate summer peaks, negative values indicate winter peaks, and values near zero reflect uniform precipitation throughout the year. |
| root_depth | Depth available within the soil profile from which roots can extract water (m). Deep rainfall infiltration and well-drained conditions favor deeper root systems, enhanced percolation, and reduced surface runoff, whereas poorly drained soils constrain roots to shallow depths, increasing soil saturation and runoff generation (Fan, et al., 2017). |
| tawc | Total available water content (mm). |
| high_Q_duration | Average duration of high-flow events (number of consecutive days > 9 times the median daily flow (d). |
| low_Q_duration | Average duration of low-flow events (number of consecutive days < 0.2 times the mean daily flow (d). |
| Q_var | Variance of streamflow ($mm^2 \cdot d^{-2}$). |
| BFI | Baseflow index (ratio of mean daily baseflow to mean daily streamflow). |
| QP_elasticity | Annual streamflow precipitation elasticity (sensitivity of streamflow to changes in precipitation using mean daily streamflow as reference). |
| StorageFraction | Ratio between active and total storage |

# 3 Methods

## 3.1 Nutrient limitation classification

For classification of nutrient limitation, phytoplankton-nutrient stoichiometric imbalance methodology was used, following Moon, Scott, & Johnson (2021) and a recent implementation for continental US (Rock & Collins, 2024). Details on this methodology are summarized in Ladwig et al. (2026, in press.). The method relies on quantile regressions and is built on the assumption that phosphorus-limited samples exhibit a higher Chla yield per unit P, while nitrogen-limited samples yield more Chla per unit N.

## 3.2 Extraction of landscape attributes

The aim of this report is to explore the relation of landscape characteristics to lake ecological status, and for this purpose we extracted landscape attributes for four buffer distances 10 m, 50 m, 100 m, and 250 m – around the lakes in addition to the previously calculated catchment boundary. As part of the project assignment, DCE was tasked with evaluating whether the catchment definitions used in Søndergaard et al. (2023) could be improved. Because no new digital elevation model data were available, the same catchment delineation was applied in this report, while additional landscape information was obtained using buffer zones.

For land-use and soil maps, class percentages were calculated for the total catchment or buffer area after removing the lake area. For potential drained areas, the percentage of the total potential drainage area was calculated, again excluding the lake area inside the catchment. For slope, average slope values were calculated for each buffer distance and the catchment.

## 3.3 Redundancy analysis (RDA)

Redundancy analysis (RDA) was conducted to assess general in-lake response patterns by investigating the influence and relative importance of four predictor types: land-use, soil, drainage, catchment characteristics (CAMELS-DK), and lake area on in-lake variables (Table 3.1). Separate analyses were carried out for each buffer zone and for the estimated catchment.

**Table 3.1.** List of predictor and response variables used in the RDA. Secchi/max_depth: Ratio of Secchi depth to lake maximum depth. See explanation for other abbreviations in Section 2.

| Predictors | | | | | | | Response |
|---|---|---|---|---|---|---|---|
| **Land-use** | **Soil** | **Drainage** | **Camels_DK** | **Area slope** | **Storage Fraction** | **Lake related** | **In-lake parameters** |
| non_cultivated | organic | drained | t_mean | slope | storage_fraction | lake_area | summer_TN |
| grassland | clay | | p_seasonality | | | | summer_TP |
| agriculture | sand | | high_prec_freq | | | | Chla |
| agri-environment | | | high_prec_dur | | | | Secchi/max_depth |
| | | | high_Q_dur | | | | |
| | | | low_Q_dur | | | | |
| | | | Q_var | | | | |
| | | | root_depth | | | | |
| | | | tawc | | | | |
| | | | BFI | | | | |
| | | | QP_elasticity | | | | |

RDA was carried out using the *vegan* package (v. 2.7-2; Oksanen et al., 2025) in R Statistical Software (v. 4.5.1; R Core Team, 2025). To determine statistically significant predictors and to simplify the models, backward selection was applied using the *ordiR2step* function. Adjusted $R^2$ values for each RDA are also reported in the results. To account for multicollinearity among predictors, variables with variance inflation factors (VIF) greater than 10 were removed prior to analysis. Following backward selection, statistical significance of the predictors was determined using permutation-based ANOVA (anova.cca, by = "term").

Response variables were $\log_{10}(x + 1)$ transformed after removing TP>3 and TN>5 mg $L^{-1}$ observations from the data. These thresholds correspond to approximately the 95th–99th percentiles of our dataset (TN > 5 mg L-1 ≈ 95th percentile; TP > 3 mg $L^{-1}$ ≈ 99th percentile), and thus they represent extreme outliers. Catchment characteristics (CAMELS-DK) and drainage variables were centered and scaled. Compositional predictors (land-use and soil variables) were transformed using the centered log-ratio (CLR) method with the *compositions* R-package (van den Boogaart et al., 2025). CLR transformation involves dividing each component by the geometric mean of all components in a sample and then taking the natural logarithm. This reduces false or misleading correlations that often arise in compositional data when Euclidean distance-based analysis is conducted (e.g. redundancy analysis).

Before CLR transformation, zeros in the compositional datasets were handled by removing variables containing many zeros and replacing the remaining zeros using the *cmultRepl* function from the *zCompositions* package, which applies a Bayesian-multiplicative replacement method (Palarea-Albaladejo & Martín-Fernández, 2025). Following the CLR transformation, compositional predictors were centered and scaled.

## 3.4 Random forest analyses

Random forests are a robust choice for analyzing large, multivariate datasets in classification problems as they construct multiple independent decision tree models (Breiman, 2001). For random forest analysis, catchment characteristics and general lake information data were used (see Section 2 for predictors).

Data were split into training-testing sets with a ratio of 4:1 (80-20), stratified by ecological status. Before training, data covariates with zero variance were removed, and data were normalized. The model targets were the categorical ecological status values. To ensure optimal setup, hyperparameters were fitted using k-fold cross-validation, also stratified by ecological status. The optimized hyperparameters were then used to run the final classification model using the *parsnip* and *ranger* R-packages (Kuhn, 2025; Wright and Ziegler, 2017). We evaluated model performance on the test dataset by calculating the ROC (receiver operating characteristic) and AUC (area under ROC curve). ROC highlights the true positive against the false positive rates, indicating how well the model learned to classify the categorical targets. AUC is a key performance metric used in classification problems that measures a model's ability to distinguish between classes, where a value of 0.5 indicates no better discrimination than random expectation, and values approaching 1 indicate increasing discriminatory ability.

To assess the final model's sensitivity to the input data and thereby identify which covariates are important for classifying a lake's ecological status, we ran a feature importance analysis using the *iml* R-package (Molnar, Bischl and Casalicchio, 2018). We undertook a global permutation importance analysis to quantify how sensitive the baseline loss quantified by cross-entropy is to changes in model input features. For local importance, we calculated the accumulated local effects (ALE) of each variable exceeding the 90%-quantile in the global permutation importance analysis. ALE illustrates the average effect of each feature on the model's predicted probabilities. The random forest analysis was performed separately for each buffer zone category.

## 3.5 Conditional inference trees

We constructed conditional inference trees (CIT, or decision trees) using the *partykit* R-package (Hothorn, Seibold, & Zeileis, 2015) to identify important landscape predictors and their potential thresholds. Decision trees were applied to the full multivariate dataset with TN, TP, ecological status, and limitation type as target variables for each buffered region and catchment border. This approach identifies the most influential catchment and water-quality thresholds that best separate the response variables.

CITs are particularly effective for capturing hierarchical structures and nonlinear relationships among predictors, providing interpretable decision rules that show how combinations of nutrient concentrations and catchment characteristics influence lake state. We conducted CIT for explanatory purposes, although predictive performance metrics were also calculated to evaluate the model performance. Accuracy and Kappa value (measuring agreement between predictions and observations beyond what would be expected by random guessing; <0 as no agreement; 0.01-0.20 as none to slight; 0.21-0.40 as fair, 0.41-0.60 as moderate; 0.61-0.80 as substantial; 0.81-1.00 as perfect agreement (McHugh (2012)) were calculated for categorical responses (ecological status, nutrient limitation type), and $R^2$ was calculated for continuous response variables (TN, TP) to assess the performance of decision trees. For the CIT analysis, outlier values were filtered: TP>3 and TN>5 mg $L^{-1}$ were removed from the analysis to focus on the most relevant nutrient levels.

The analyses were repeated for four different buffer zones as well as for the full catchment area. However, only the results based on the catchment area are presented in the main text, as the explained variance was generally higher for landscape predictors derived at the catchment scale. Results for the buffer-zone analysis are provided in the appendix.

# 4 Results

## 4.1 Overview of the lakes

The main characteristics of the lakes included in the analyses in this report are presented in Table 4.1. Total phosphorus (TP) ranged from 0.003 to 16.7 mg $L^{-1}$, with a mean of 0.3 mg $L^{-1}$, while total nitrogen (TN) ranged from 0.3 to 11.4 mg $L^{-1}$, with a mean of 1.8 mg $L^{-1}$.

Based on ecological status classes, of the 418 lakes, 74 were classified as having high status, 50 as good status, 53 as moderate status, 43 as poor status, and 65 as bad status. A total of 60 lakes did not have a defined ecological status, hence they were not included in the analysis. After removing the data prior to 2010 and excluding lakes with unusually high TP and TN values (TN > 5 mg $L^{-1}$ and TP > 3 mg $L^{-1}$), 331 lakes remained in the dataset.

**Table 4.1.** The full range of in-lake characteristics in the unfiltered dataset. Outliers were excluded from the analyses but are shown here to reflect the original data variability.

| Variables | Mean | Median | Min. | Q25 | Q75 | Max. |
|---|---|---|---|---|---|---|
| Summer TN (mg $L^{-1}$) | 1.8 | 1.3 | 0.3 | 0.8 | 2.1 | 11.4 |
| Summer TP (mg $L^{-1}$) | 0.3 | 0.1 | 0.003 | 0.03 | 0.2 | 16.7 |
| Maximum depth (m) | 3.9 | 1.9 | 0.1 | 0.9 | 5.7 | 45.0 |
| Secchi depth (m) | 1.5 | 0.8 | 0.1 | 0.4 | 2.0 | 7.3 |
| Lake area (ha) | 11.8 | 5.9 | 1.0 | 2.5 | 9.4 | 576.7 |

Both TN and TP concentrations increased consistently from high-status lakes toward degraded lakes (bad) (Figure 4.1). Lakes in the bad class showed clearly elevated nutrient levels, while high and good status lakes generally exhibited low TN and TP concentrations with narrow ranges.

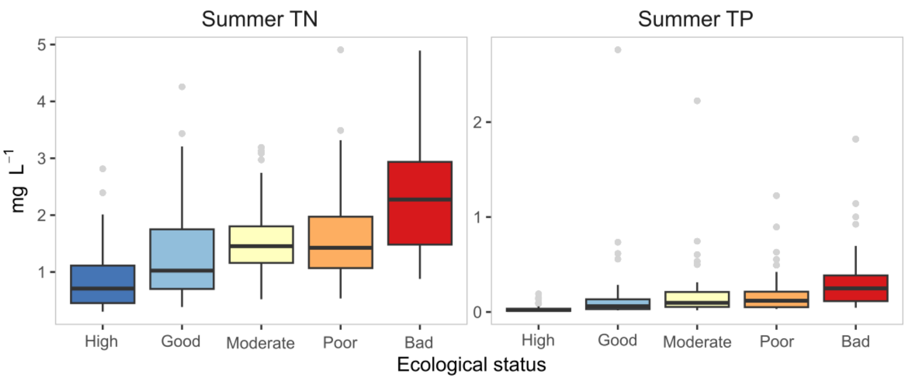



**Figure 4.1.** Boxplots of summer total nitrogen and summer total phosphorus concentrations across ecological status classes. In boxplot, horizontal solid line indicates median of the distribution, the box represents the lower to upper quartile values of the data, the whiskers extend to the last data point beyond 1.5 * the interquartile range, circles represent outliers beyond this range.

## 4.2 Overview of landscape attributes

### Land Use/ Fields

According to the 'Land Use/Fields' data, non-cultivated areas have remained relatively stable over the past 15 years (Figure 4.2). In contrast, marked changes occurred in the *agricultural*, agri-environment management areas (*agri-environment*), and permanent grassland (*grassland*) categories. Agricultural and grassland areas showed a decreasing trend over time, whereas the proportion of areas under the *agri-environment* increased steadily. The figure also shows the average proportion of major land-use categories within four buffer distances (10 m, 50 m, 100 m, 250 m) and in the full catchment for the years 2010 and 2018–2024. On average, the *non-cultivated* class – which represents natural areas – had the highest percentage among the categories, although its proportion decreased with increasing buffer size. *Agriculture* displayed the opposite pattern, with its percentage increasing as buffer size increased.

Figure 4.3 shows the land-use percentages of different classes across the ecological status classes for lakes at various buffer zones. In general, across all buffer distances, high-quality lakes had a higher percentage of *non-cultivated* areas, whereas lakes with bad status had lower percentages of *non-cultivated* areas. For *agricultural* areas and permanent *grassland* areas, the pattern was reversed at the 250-m buffer and catchment scale, and lakes with poor/bad ecological status generally had higher percentages of *agriculture* and permanent *grassland* than high-quality lakes. The pattern for areas under the environmental scheme was less clear, although lakes with good status had slightly higher percentages of *agri-environment* areas.

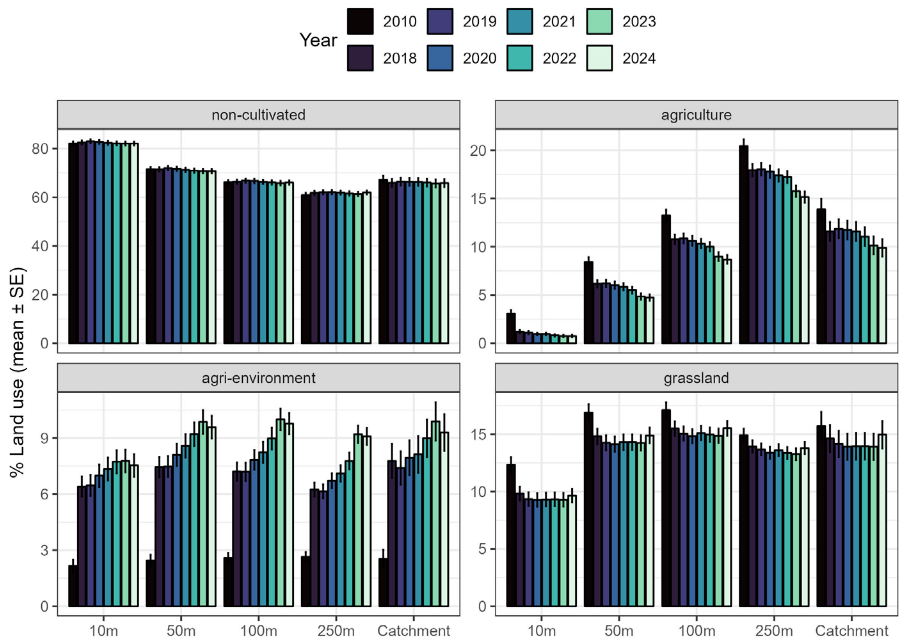



**Figure 4.2.** Mean percentages of land-use types within different buffer distances (10 m, 50 m, 100 m, 250 m) and estimated catchments across eight years (2010–2024). Bars show mean ± SE across lakes for each land-use class. See Section 2 for variable explanations.

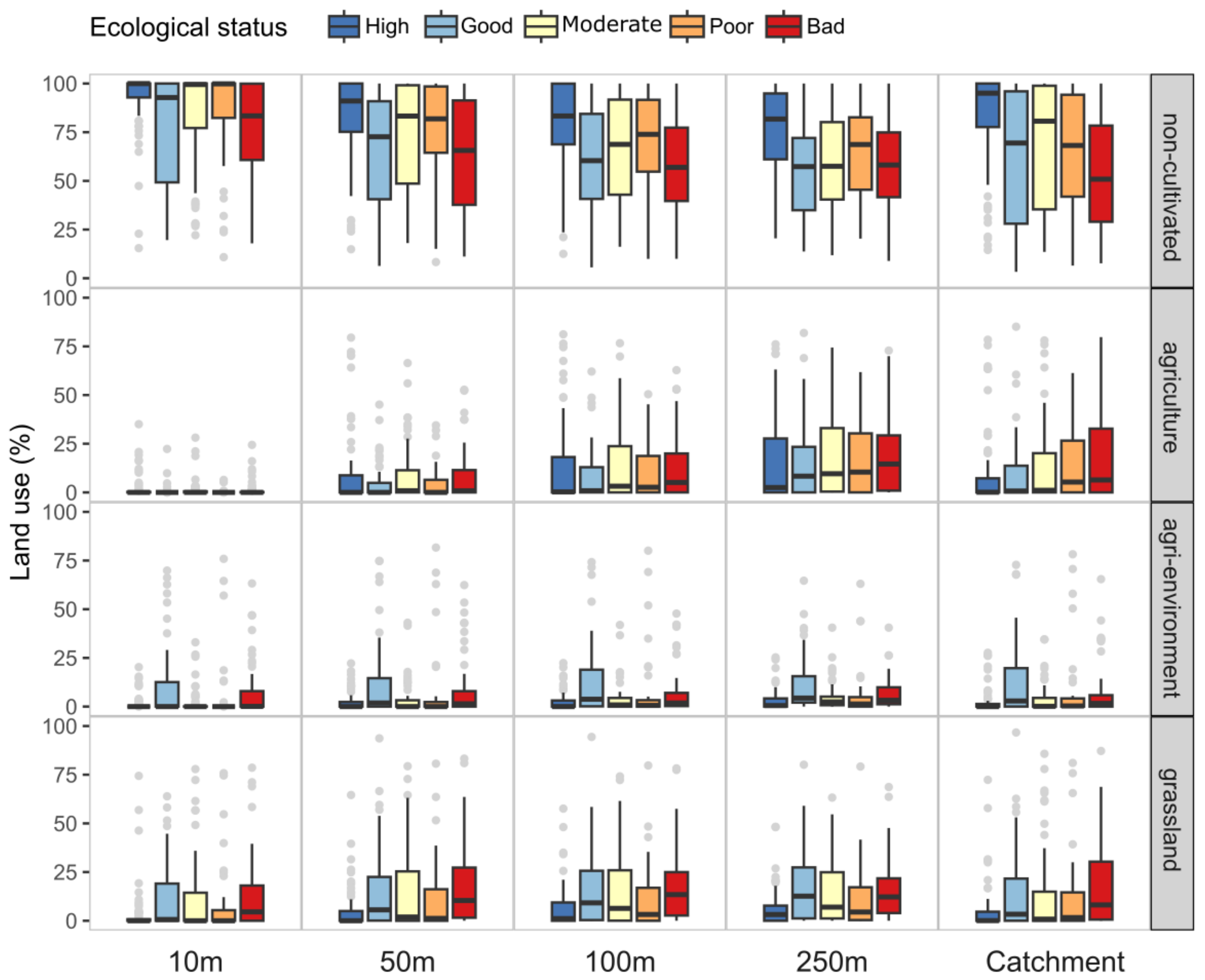



**Figure 4.3.** Land-use percentages (non-cultivated, agriculture, permanent grassland (grassland), agri-environmental management scheme (agri-environment)) within lake buffer zones and catchments across ecological status classes (High, Good, Medium, Poor, bad). In boxplot, horizontal solid line indicates median of the distribution, the box represents the lower to upper quartile values of the data, the whiskers extend to the last data point beyond 1.5 * the interquartile range, circles represent outliers beyond this range. See Section 2 for variable explanations.

### Soil data

Across all buffer distances and catchment areas, clay soils were the dominant soil type, typically comprising more than half of the surrounding area (Figure 4.4). Organic soils accounted for a substantial but smaller portion, while sandy soils constituted the lowest share across all ecological status classes. The proportion of organic soil increased as ecological status declined, with the highest values generally occurring in lakes with poor or bad ecological status (poor and bad). Conversely, lakes with high or good ecological status (high and good) tended to have lower proportions of organic soil in their surroundings. The percentage of soil classes did not vary across different buffer zones. Lakes in better ecological states had higher percentages of sandy soil than lakes in poorer ecological conditions.

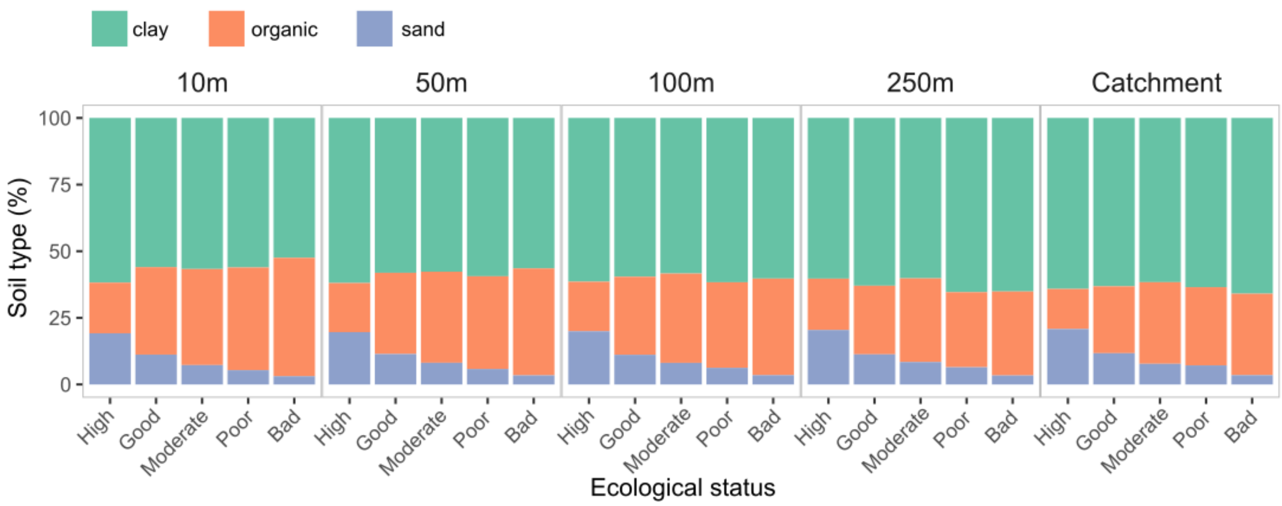



**Figure 4.4.** Soil percentages within lake buffer zones and catchments across ecological status classes.

### Potential drainage area

Drainage intensity (*drained*%) across various buffer regions for ecological status are given in Figure 4.5. Lakes with high and good ecological status generally had a lower proportion of drained areas in their surrounding catchments. In contrast, lakes with poorer ecological status tended to be in catchments with a higher percentage of drained land.

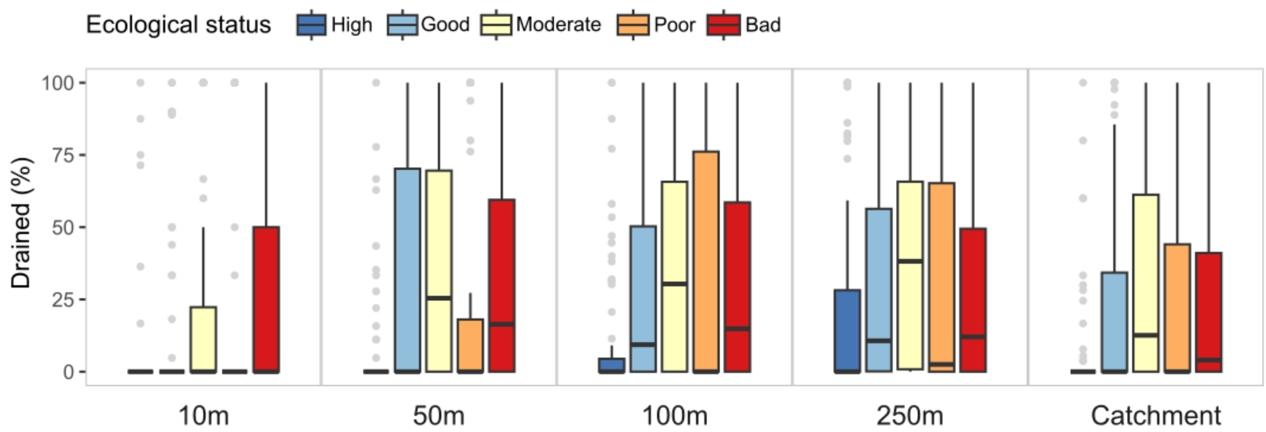



**Figure 4.5.** Potential drainage area percentages (Drained %) within lake buffer zones and catchments across ecological status classes.

The relationship between drainage intensity (*drained*% ) and lake nutrient concentrations is shown in Figure 4.6. With increasing drainage intensity in the catchment, the peak of the TN and TP density distributions shifts to the right in density distribution plots, indicating higher nutrient concentrations in lakes with more extensive drainage from the buffer area/catchment. This shift is accompanied by a widening of the right tail, showing that lakes in highly drained catchments are more likely to reach elevated TP and TN levels. For TN, lakes with 0–25% drainage showed a unimodal distribution with a peak around ~1 mg $L^{-1}$, whereas higher drainage classes exhibited right-shifted peaks and in some cases a bimodal distribution. Together, these patterns suggest that lakes located in catchments with extensive artificial drainage (75–100%) exhibited higher nutrient concentrations and greater variability compared to lakes in less-drained catchments. This pattern was especially pronounced for TN, where highly drained catchments displayed long right-hand tails, indicating higher TN values.

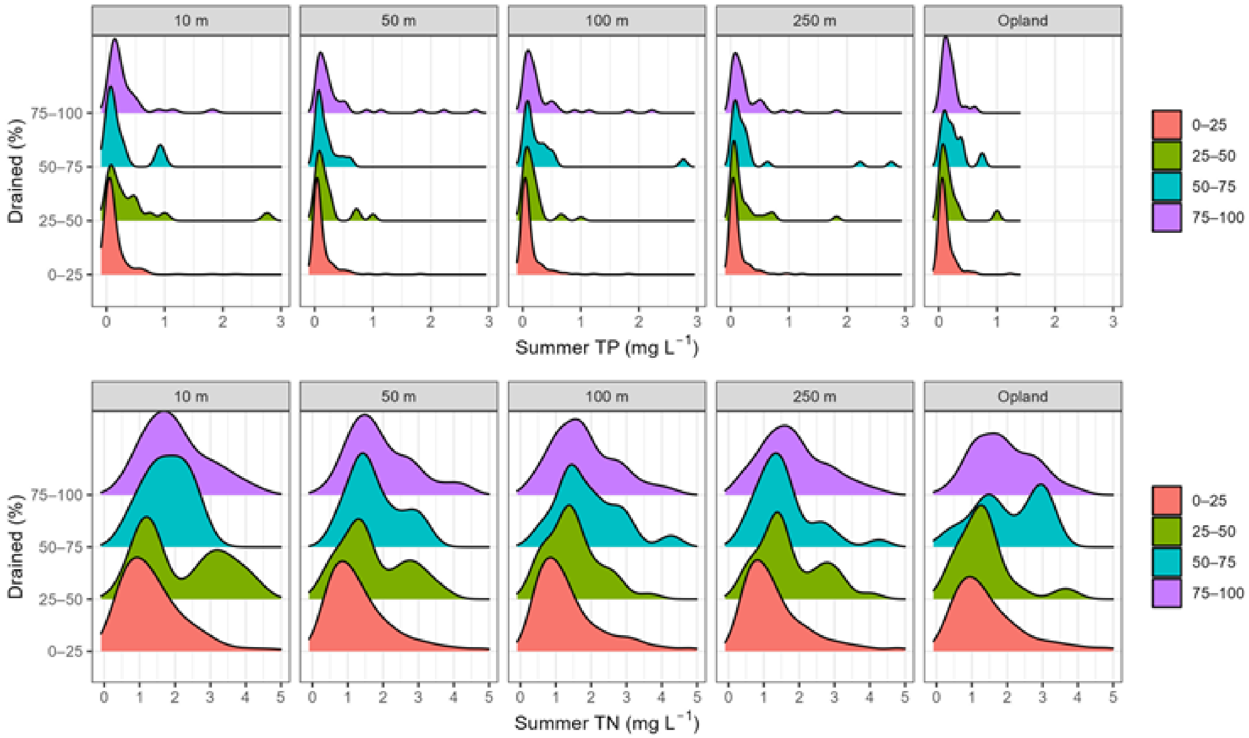



**Figure 4.6.** Density distributions of total phosphorus (TP) and total nitrogen (TN) concentrations across drainage classes surrounding each lake. Drainage classes represent the percentage of catchment or buffer area covered by artificial drainage (0–25%, 25–50%, 50–75%, and 75–100%). In the plots TP > 3 mg $L^{-1}$ and TN > 5 mg $L^{-1}$ are not shown.

## 4.3 Effects of landscape attributes on lake state

### RDA results

RDA was used to assess the general relationship of land-use, soil, drainage, catchment characteristics, and lake area variables with in-lake responses. namely TN, TP, Chla concentrations, and Secchi/Maximum Depth ratio, representing lake turbidity. Observations with missing values were removed from the analysis, because RDA does not accept missing data. Consequently, sample sizes (numbers of lakes) for the RDAs ranged from 222 to 231 depending on buffer distance or catchment.

All global RDA models and their first two axes were significant, the adjusted $R^2$ values showing that the selected predictors explained 25–32% of the constrained variation in in-lake water quality (Figure 4.7). RDA1 was the dominant axis and explained 22–30% of the variation, whereas RDA2 contributed only 3.0–3.4%. The RDA models were refined using backward selection, and all predictors not listed in the permutation test table (Table 4.2) were removed as either not significant or due to multicollinearity.

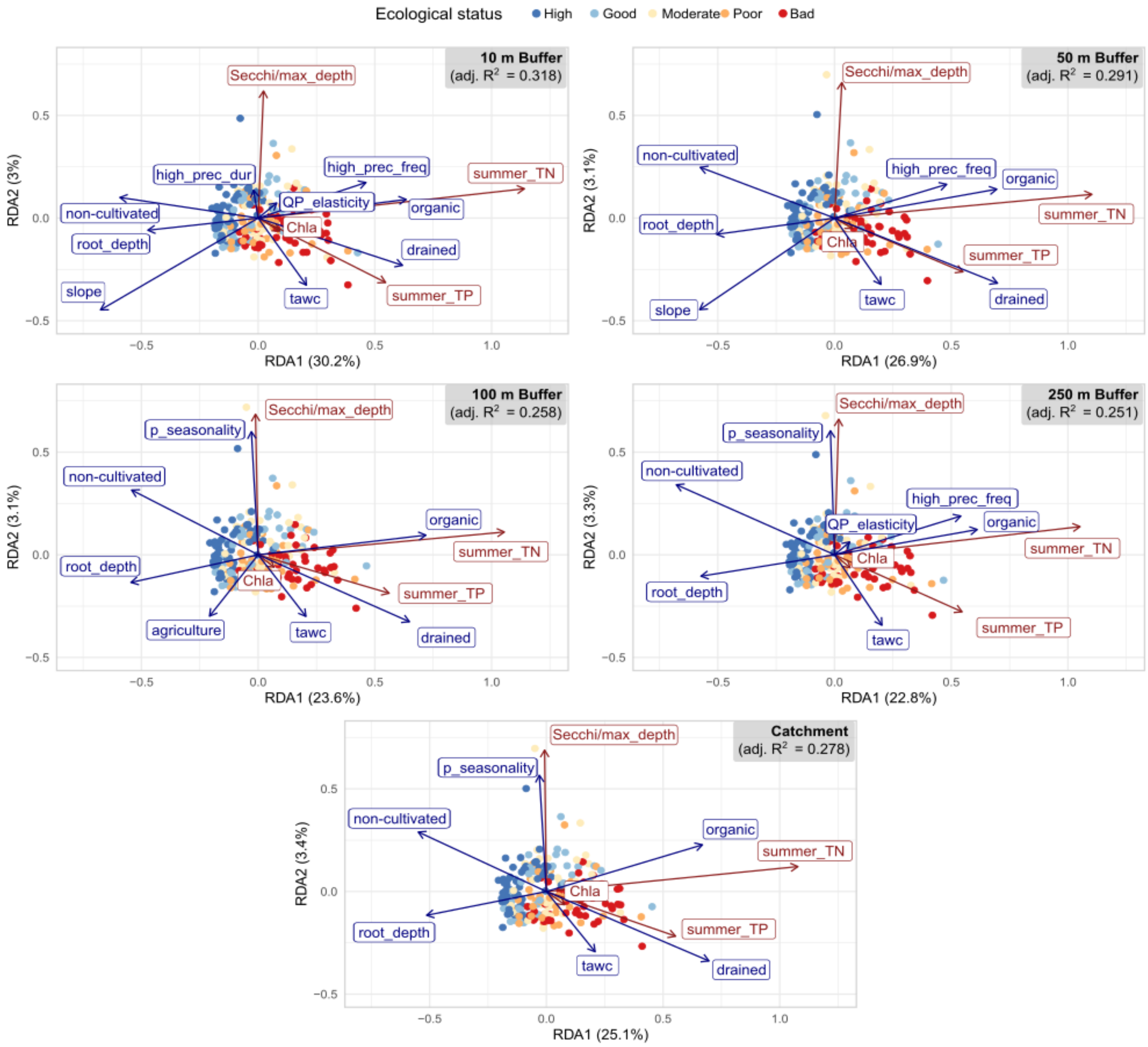



**Figure 4.7.** RDA analysis results for different buffer distances (10 m, 50 m, 100 m, 250 m) and catchments. Blue arrows are statistically significant predictors automatically selected with backward selection, and red arrows are in-lake response variables. Points represent lakes, while colours represent ecological status from bad (red) to high (dark blue). See Section 2 for variable explanations.

Because land-use and soil data were CLR-transformed, their effects should be interpreted relative to alternative classes within each compositional group rather than individual effects, even if some classes were not individually significant in the RDAs. For example, higher proportions of *non-cultivated* to other land-use categories (e.g., *agriculture, grassland, agri-environmental*) were associated with better lake quality (ecological status class: high, good). On the other hand, higher organic soil relative to sand and clay indicated a greater risk of eutrophication (ecological status class: poor, bad).

At the 10 m and 50 m buffer scales, as well as for the full catchment, RDA1 represented a eutrophication gradient. The in-lake response variables summer TN, summer TP and summer Chla were strongly and positively associated with this axis, indicating that RDA1 reflects increasing nutrient concentrations. These responses were positively related to explanatory variables, including the percentage of potential drained area (*% drained*), proportion of organic soils (*organic*), and higher total available water capacity (*tawc*), and, to a lesser extent, to the frequency of intense precipitation periods (*high_prec_freq*).

In contrast, *root_depth* showed a strong negative association with this gradient, and *non-cultivated* areas displayed a similar but weaker negative relationship. Average *slope* in these buffers also showed a strong opposite relationship with RDA1, especially with *high_prec_freq*. Lakes, colored by ecological status, showed that predictors negatively associated with RDA1 were located towards lakes with better ecological status.

Collectively, the results in Figure 4.7 suggest that wetter catchments with more *organic* soils, less frequent precipitation (*high_prec_freq*) and larger *drained* areas are associated with poorer lake ecological quality.

**Table 4.2.** Permutation test for RDA predictor variables; separate significance test for each predictor (p). Raw variance indicates the unadjusted variations in in-lake response data explained by each predictor. Residual variance is not shown. See Section 2 for variable explanations.

| Buffer /Catchment | Predictors | Raw variance | % Total variance | F | *p* |
|---|---|---|---|---|---|
| 10m | *slope* | 0.0043 | 10.9 | 34.8 | 0.001 |
| 10m | *drained* | 0.0031 | 7.9 | 25.2 | 0.001 |
| 10m | *tawc* | 0.0015 | 3.8 | 11.9 | 0.002 |
| 10m | *high_prec_freq* | 0.0012 | 3.0 | 9.5 | 0.001 |
| 10m | *organic* | 0.0011 | 2.9 | 9.2 | 0.001 |
| 10m | *root_depth* | 0.0008 | 2.0 | 6.5 | 0.002 |
| 10m | *non_cultivated* | 0.0008 | 1.9 | 6.2 | 0.004 |
| 10m | *high_prec_dur* | 0.0005 | 1.2 | 3.7 | 0.037 |
| 10m | *QP_elasticity* | 0.0004 | 1.0 | 3.2 | 0.049 |
| 50m | *drained* | 0.0038 | 9.8 | 30.4 | 0.001 |
| 50m | *slope* | 0.0029 | 7.4 | 23.0 | 0.001 |
| 50m | *root_depth* | 0.0016 | 4.2 | 13.1 | 0.001 |
| 50m | *organic* | 0.0015 | 3.7 | 11.6 | 0.001 |
| 50m | *tawc* | 0.0012 | 3.1 | 9.6 | 0.002 |
| 50m | *non_cultivated* | 0.0008 | 2.1 | 6.6 | 0.003 |
| 50m | *high_prec_freq* | 0.0004 | 1.0 | 3.3 | 0.036 |
| 100m | *organic* | 0.0033 | 8.4 | 25.3 | 0.001 |
| 100m | *root_depth* | 0.0032 | 8.1 | 24.4 | 0.001 |
| 100m | *tawc* | 0.0014 | 3.5 | 10.7 | 0.001 |
| 100m | *non_cultivated* | 0.0011 | 2.9 | 8.7 | 0.001 |
| 100m | *p_seasonality* | 0.0008 | 2.1 | 6.5 | 0.005 |
| 100m | *agricultural* | 0.0006 | 1.6 | 4.7 | 0.014 |
| 100m | *drained* | 0.0006 | 1.4 | 4.3 | 0.012 |
| 250m | *non_cultivated* | 0.0031 | 7.9 | 23.9 | 0.001 |
| 250m | *organic* | 0.0024 | 6.1 | 18.4 | 0.001 |
| 250m | *root_depth* | 0.0022 | 5.8 | 17.4 | 0.001 |
| 250m | *tawc* | 0.0012 | 3.1 | 9.5 | 0.001 |
| 250m | *p_seasonality* | 0.0008 | 2.0 | 6.1 | 0.006 |
| 250m | *high_prec_freq* | 0.0006 | 1.4 | 4.3 | 0.023 |
| 250m | *QP_elasticity* | 0.0004 | 1.0 | 3.0 | 0.065 |
| Catchment | *drained* | 0.0037 | 9.5 | 28.9 | 0.001 |
| Catchment | *organic* | 0.0031 | 8.0 | 24.3 | 0.001 |
| Catchment | *root_depth* | 0.0021 | 5.4 | 16.6 | 0.001 |
| Catchment | *tawc* | 0.0011 | 2.7 | 8.3 | 0.003 |
| Catchment | *non_cultivated* | 0.0010 | 2.5 | 7.5 | 0.002 |
| Catchment | *p_seasonality* | 0.0007 | 1.7 | 5.2 | 0.008 |

At the larger buffer scales (100 m and 250 m) (Appendix Figure 7.1), the eutrophication gradient remained, but the relative importance of the predictors shifted. At 100 m and 250 m buffers, the percentage of potential *drained* area contributed less than at smaller buffers, whereas *organic* soil and *root_depth* remained important.

Notably, *non-cultivated* land increased in explanatory power with buffer size, and its F-value and explained variance were highest at 250 m (Figure 4.8, Table 4.2). Despite this, the overall explanatory power of the RDA models decreased slightly with increasing buffer size. At 250 m scale, frequency of high-precipitation days was also one of the significant predictors associated with poorer ecological status and elevated nutrient concentrations, although its explanatory power was low.

The second RDA axis was dominated by the *Secchi/Max depth ratio* and was aligned with the duration of high frequency precipitation (*high_prec_dur*) and precipitation seasonality (*p_seasonality*; summer precipitation peak). This axis only explained ≤3.4% of the variation across all models, suggesting that this pattern may not reflect a strong ecological effect. Additionally, although agriculture areas were significant at 100 m (Table 4.2; $F = 4.7$, $p = 0.014$), its contribution was small compared with that of other predictors.

Overall, RDAs based on different buffer sizes and catchment areas indicate that poorer lake ecological quality, characterized by higher nutrient concentrations, is associated with higher total water availability, greater proportions of organic soils, and larger drained areas. In contrast, deeper root depth in the buffers or catchment points in the opposite direction, suggesting better ecological conditions. Furthermore, as the buffer area increases, the relative influence of *non-cultivated* area compared to other land-use parameters becomes stronger, with greater coverage of non-cultivated areas relative to other land-use classes being associated with improved ecological status.

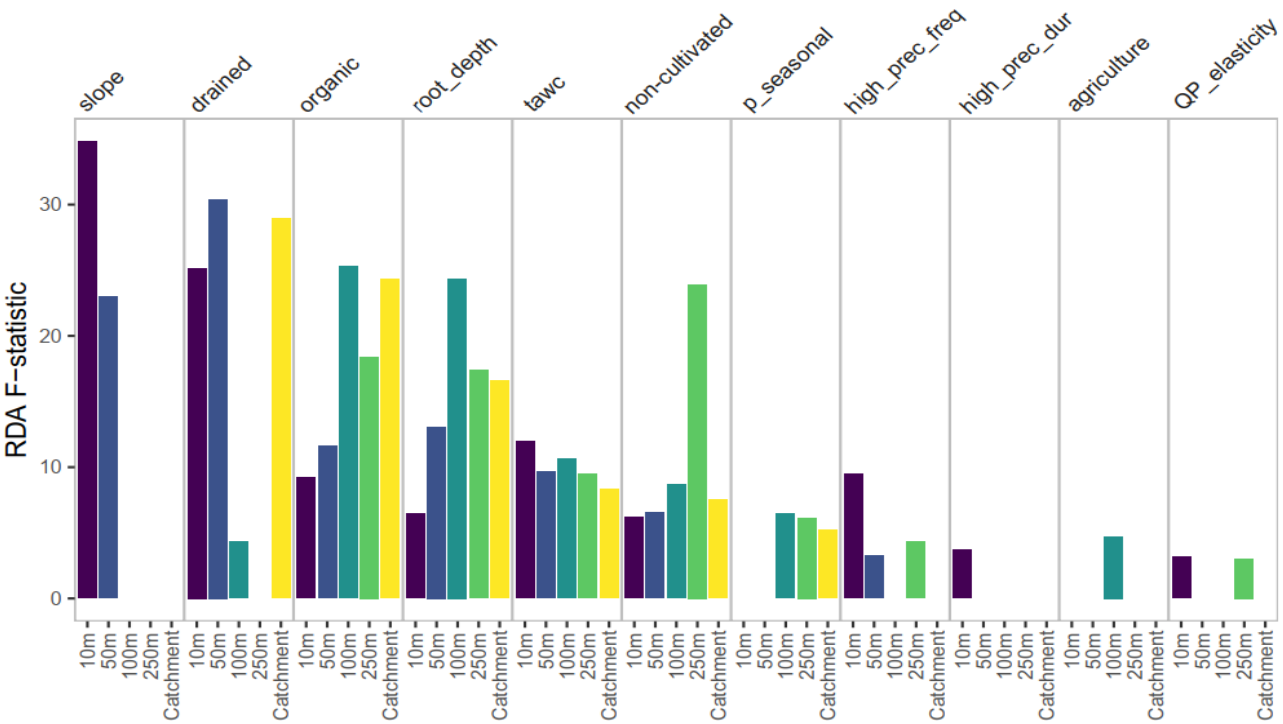



**Figure 4.8.** F-statistics from the permutation test for RDA predictor variables comparing the contribution of each predictor within each buff-er/catchment scale. See Section 2 for variable explanations.

### Random forest

For a buffer zone of 10 m (Figure 4.9), the random forest model obtained an AUC of 0.6 (individual AUCs: high = 0.88, good = 0.77, medium = 0.56, poor =0.47, bad = 0.77). The model showed poor prediction capacity for the moderate and poor ecological states, which is only slightly compensated for by a higher accuracy for lakes with good/high ecological status classes, resulting in the generally low AUC of 0.682.

Global feature importance (Figure 4.9A) ranked maximum lake depth, organic soil in the catchment, and lake area as important predictors of ecological status. Here, the probability of predicting better ecological states (high or good) increases with higher lake depth. Vice versa, higher lake depth decreases the chance of having worse ecological states which peak at low lake depth values. Highecological state probability peaks at very low soil organic values, whereas poor ecological status is associated with more organic matter. Similar to lake depth, increases in lake area are associated with high, good, and moderate ecological states (Figure 4.9B).

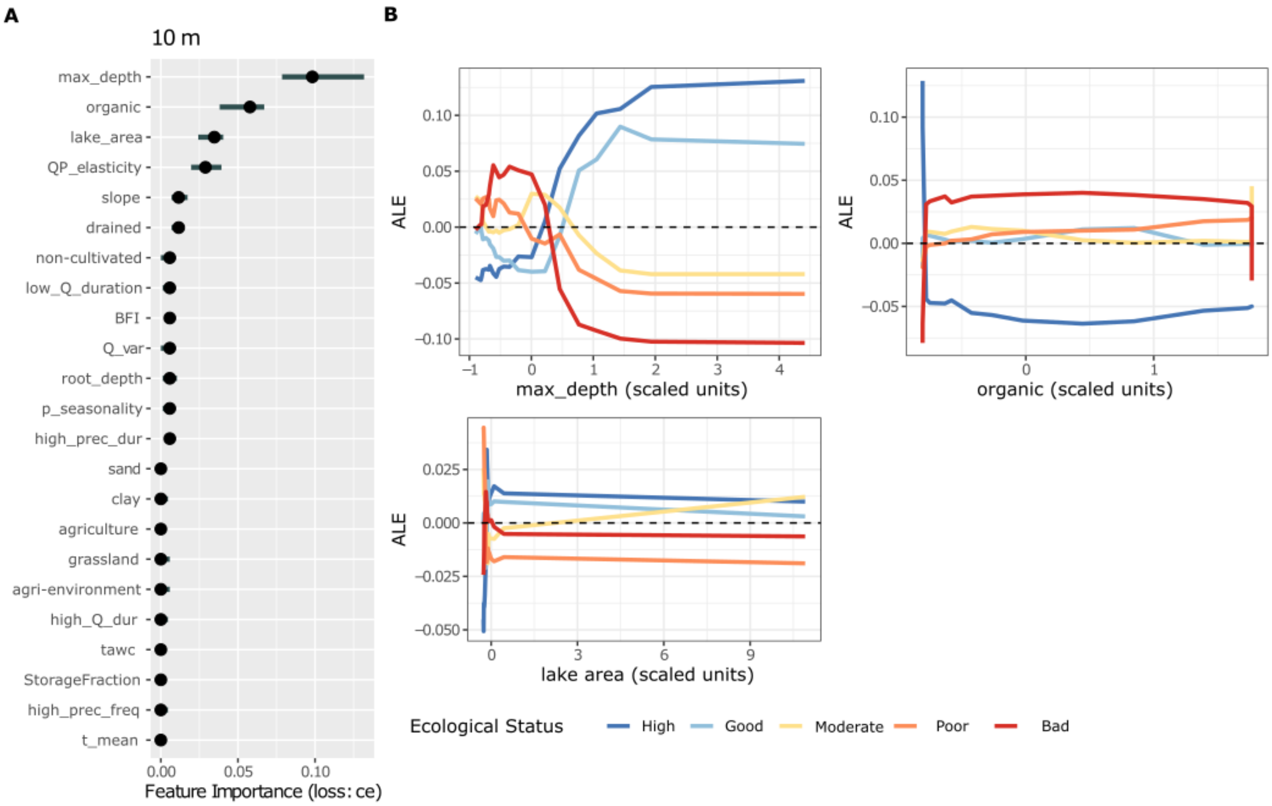



**Figure 4.9.** Feature importance for the 10 m buffer zone. A: Global feature importance (quantified using cross-entropy loss for classification). B: Accumulated local effects (ALE), which highlight the effects of specific covariates on the prediction of the random forest models. Positive ALE values (y-axis) highlight an above-average prediction of that state over the normalized range of the covariate (x-axis), vice versa, negative ALE values highlight a below-average contribution to that covariate. Here, the plot shows average model prediction over each feature stratified by ecological status. See Section 2 for variable explanations.

For a buffer zone of 50 m (Figure 4.10), the random forest model obtained an AUC of 0.6 (individual AUCs: high = 0.82, good = 0.68, medium = 0.49, poor =0.42, bad = 0.76). The model exhibited poor prediction capability for the moderate and poor ecological states, but higher accuracy for lakes with high/good and bad ecological status. Nonetheless, overall AUC is quite low, only slightly above a chance of a coin flip (0.5).

Global feature importance (Figure 4.10) ranked maximum lake depth, drainage percentage, and lake area as important covariates to classify ecological status. As above, improved ecological states are positively correlated with lake depth and area. The drainage percentage is positively related to bad ecological status and negatively to high ecological status. Nonetheless, the probability of a lake being classified as good/moderate increases with increasing drainage percentages, while the probability of high ecological state decreases. These results are contradictory and probably biased due to the model's low AUC for good, moderate, and poor ecological status.

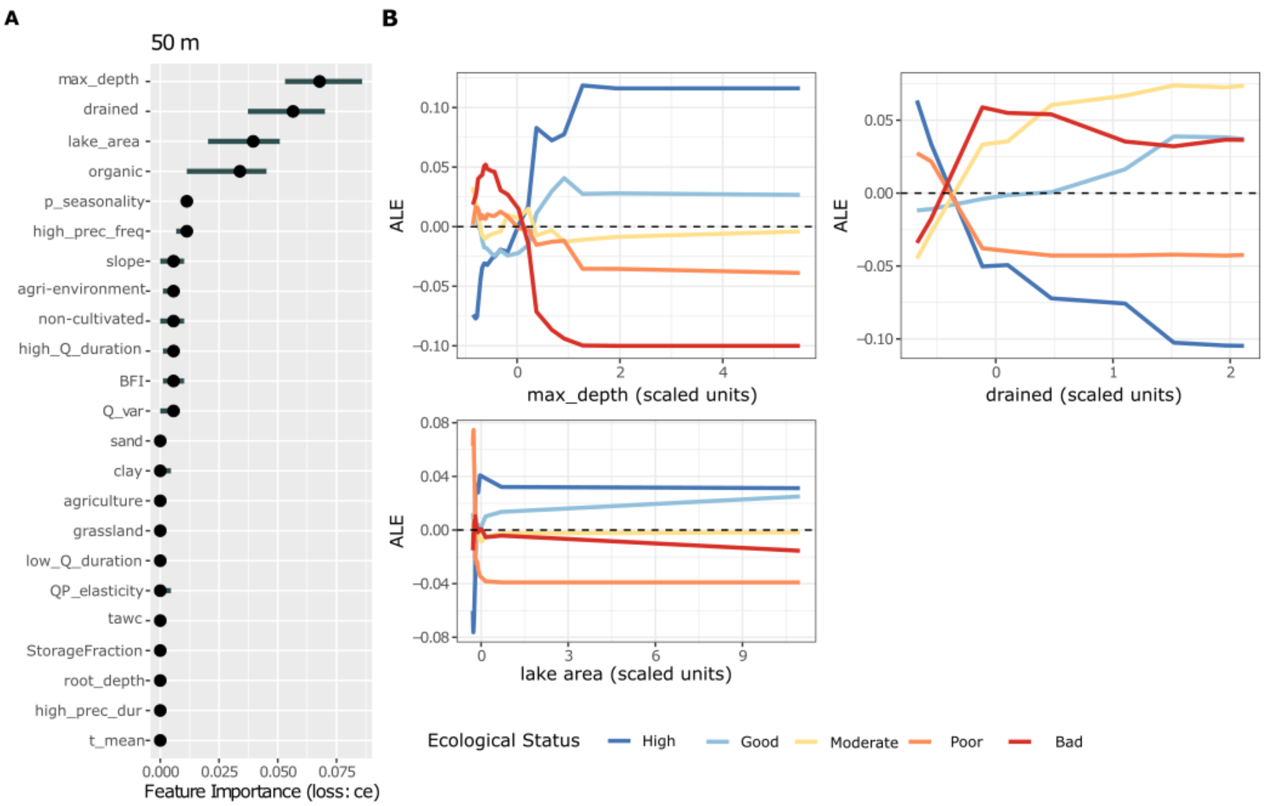



**Figure 4.10.** Feature importance for 50 m buffer zone. A: Global feature importance (quantified using cross-entropy loss for classification). B: Accumulated local effects (ALE), which highlight the effects of specific covariates on the prediction of the random forest models. See Section 2 for variable explanations and Figure 4.9 for figure explanations.

For a buffer zone of 100 m (Figure 4.11), the random forest model obtained an AUC of 0.6 (individual AUCs:high = 0.85, good = 0.64, medium = 0.68, poor =0.73, bad = 0.83). The model exhibited a good prediction capacity for all ecological states.

Global feature importance (Figure 4.11) identified maximum depth, grassland, streamflow precipitation elasticity (*QP_elasticity*), and non-cultivated land as important predictors of ecological status. Here, lakes with high ecological status were positively associated with greater maximum depths and higher proportions of non-cultivated land, whereas they were negatively associated with grassland and streamflow precipitation elasticity. The probability of a lake being classified as lake having bad ecological status increased with more grassland in the buffer zone but was low with higher streamflow precipitation indices, similar to high ecological status lakes. The likelihood of being in good, moderate, or poor condition increased with higher precipitation streamflow indices. Good, moderate, and poor ecological status lakes demonstrated contradictory dynamics of accumulated local effects, probably due to the lower AUC of the underlying model classification.

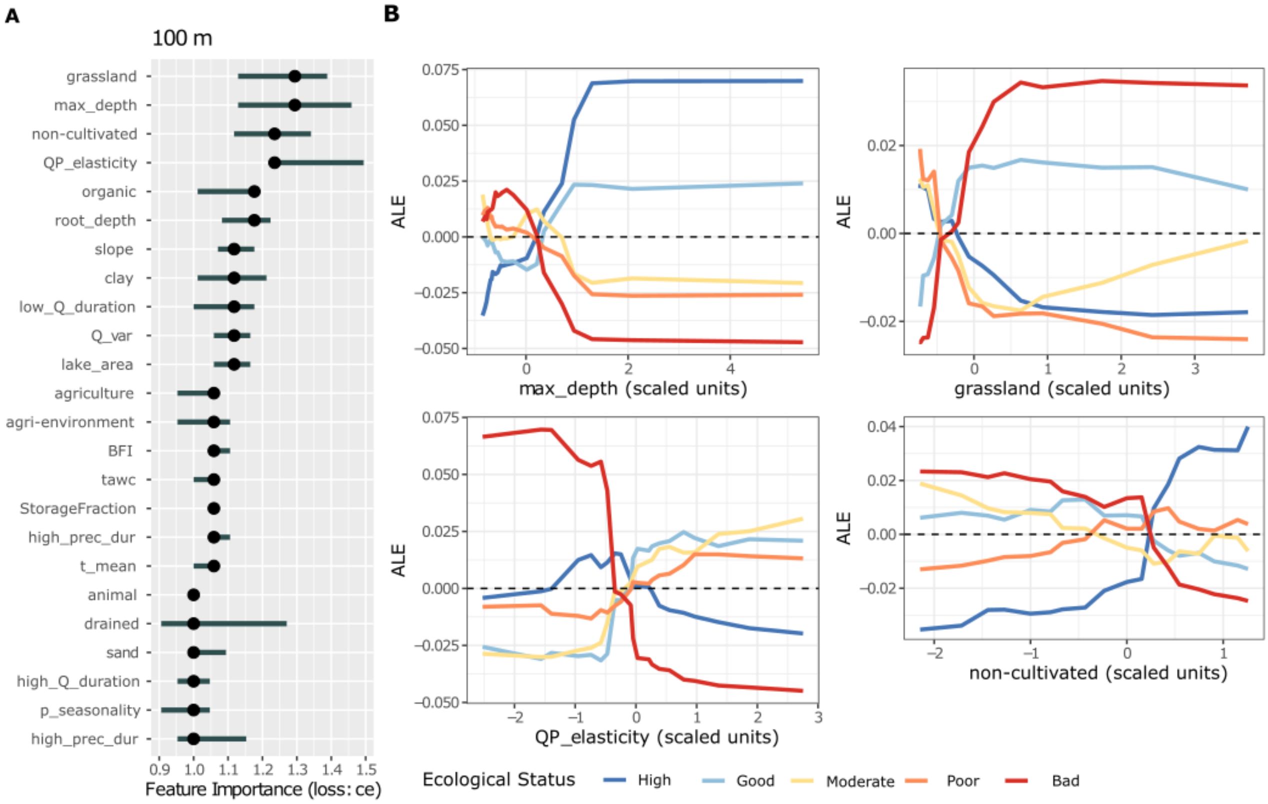



**Figure 4.11.** Feature importance for the 100 m buffer zone. A: Global feature importance (quantified using cross-entropy loss for classification). B: Accumulated local effects (ALE), which highlight the effects of specific covariates on the prediction of the random forest models. See Section 2 for variable explanations and Figure 4.9 for figure explanations.

For a buffer zone of 250 m (Figure 4.12), the random forest model achieved an AUC of 0.5 (individual AUCs: high = 0.69, good = 0.54, medium = 0.56, poor =0.51, bad = 0.86). The model had a poor to moderate classification for all states except bad ecological status. Here, overall AUC was low and near the chance of a coin flip (0.5).

Global feature importance (Figure 4.12) highlighted maximum depth, organic soil and streamflow precipitation elasticity as important, with the same relationships as explained above.

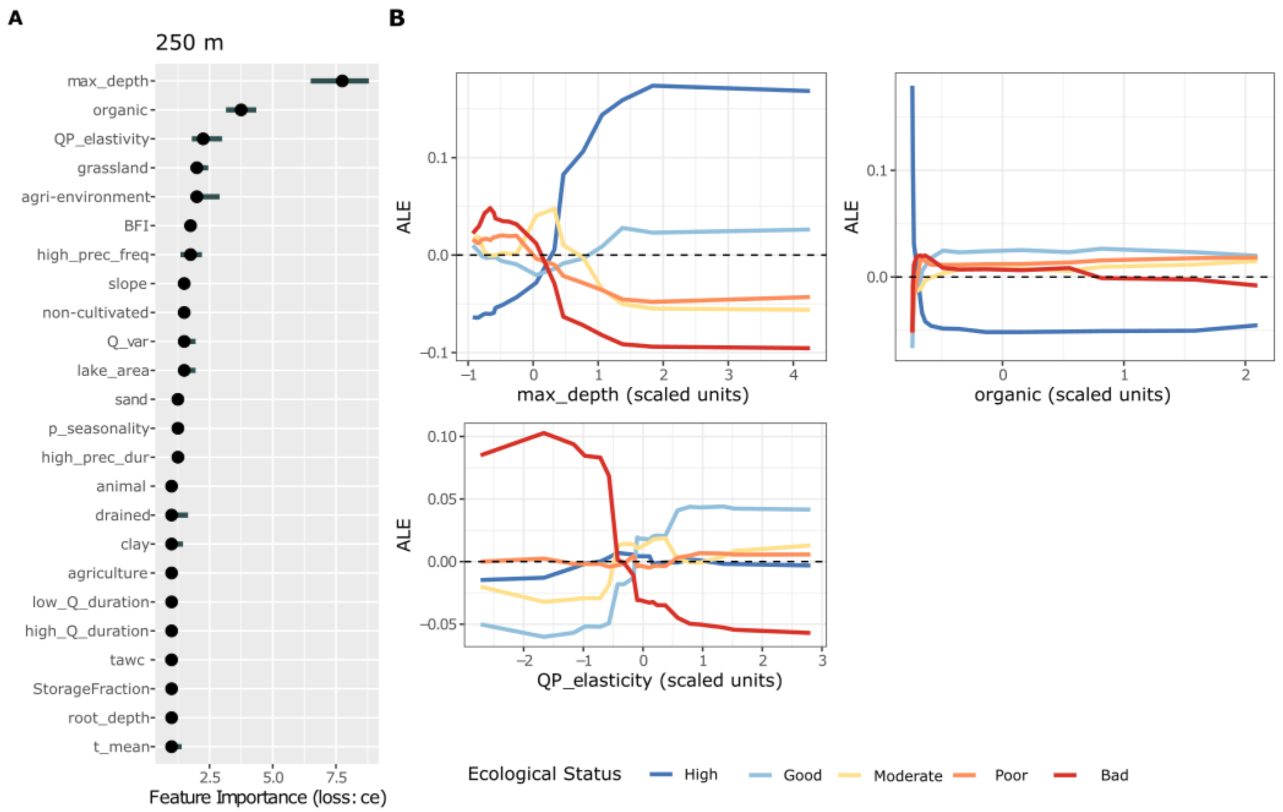



**Figure 4.12.** Feature importance for the 250 m buffer zone. A: Global feature importance (quantified using cross-entropy loss for classification). B: Accumulated local effects (ALE), which highlight the effects of specific covariates on the prediction of the random forest models. See Section 2 for variable explanations and Figure 4.9 for figure explanations.

For a buffer zone of the full catchment (Figure 4.13), the random forest model achieved an AUC of 0.6 (individual AUCs: high = 0.87, good = 0.67, medium = 0.53, poor =0.44, bad= 0.82). The model underperformed to classify the moderate and poor ecological states correctly. Overall, AUC was low and near the chance of a coin flip (0.5).

Global feature importance (Figure 4.13) ranked organic soil, maximum depth, streamflow precipitation elasticity, and non-cultivated land as important. The dynamics of the accumulated local effects were similar to those mentioned above.

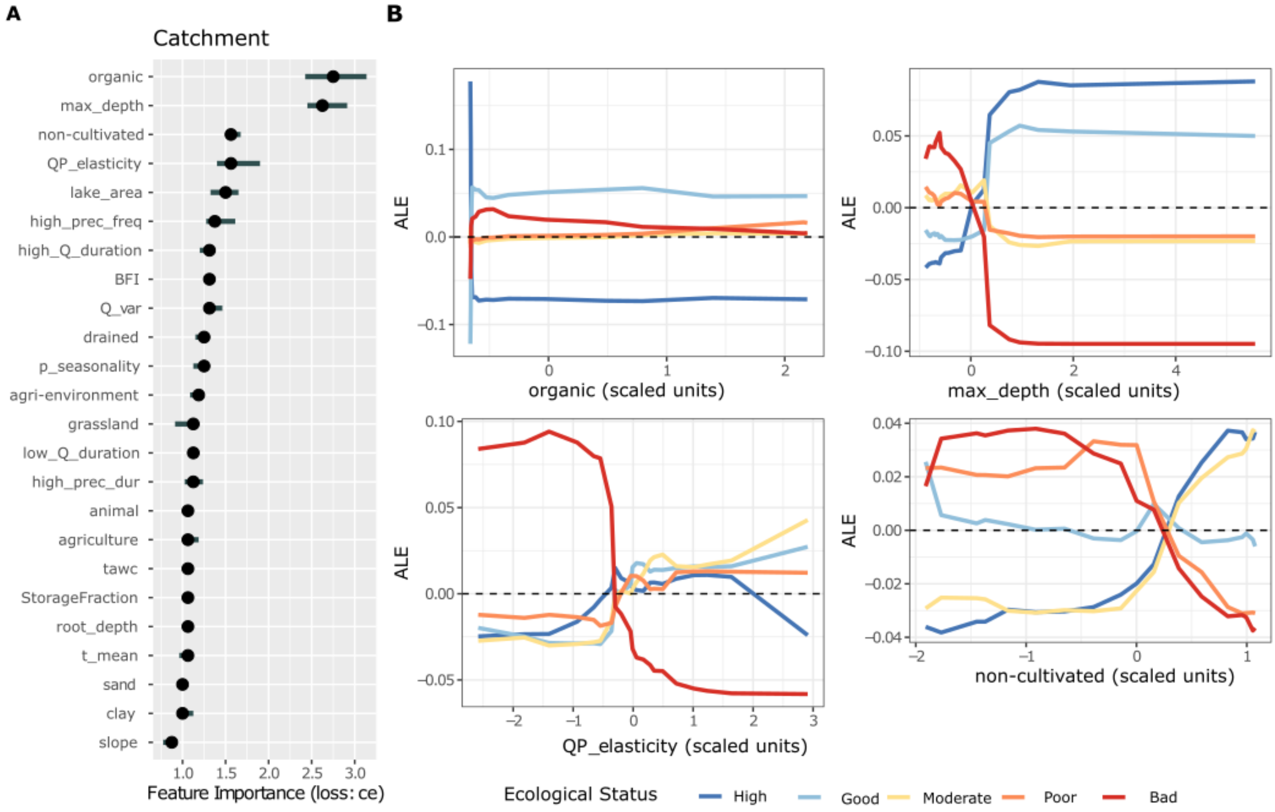



**Figure 4.13.** Feature importance for the catchment. A: Global feature importance (quantified using cross-entropy loss for classification). B: Accumulated local effects (ALE), which highlight the effects of specific covariates on the prediction of the random forest models. See Section 2 for variable explanations and Figure 4.9 for figure explanations.

### Conditional inference trees

To understand the factors and thresholds shaping summer TN, summer TP, ecological status, and nutrient-limitation in lakes without well-defined drainage areas, we applied conditional inference tree (CIT) analyses for each buffer and catchment area. The catchment area results are presented below, and buffer zone trees are shown in Appendix Figure 7.3-Figure 7.5.

#### Summer TN

Conditional inference trees were constructed to identify key parameters and threshold values controlling summer TN concentrations. Among all landscape and in-lake predictors, maximum depth, percentage of non-cultivated land, potential drained area percentage, and frequency of high precipitation events emerged as the most important variables (Figure 4.14).

Maximum depth was the strongest predictor of summer TN. In shallower lakes (< 5.7 m), additional splits occurred based on non-cultivated land area and high-precipitation frequency. Lakes experiencing fewer heavy precipitation events (< 17) had lower TN than those exposed to more frequent extreme rainfall events. Among shallow lakes with higher non-cultivated land cover (> 54%), TN levels were further influenced by catchment drainage characteristics. Specifically, lakes with drainage area percentages ≤ 3.8 showed relatively low TN values, while those above this threshold exhibited markedly higher TN concentrations.

In contrast, deeper lakes (>5.7 m) consistently showed low TN levels, and no additional predictors significantly improved the model's partitioning. This suggests that, for deeper systems, depth alone is sufficient to explain the observed variations in summer TN. The explained variance for this CIT was 0.31, reflecting a fair ability of the model to explain variability in TN concentrations.

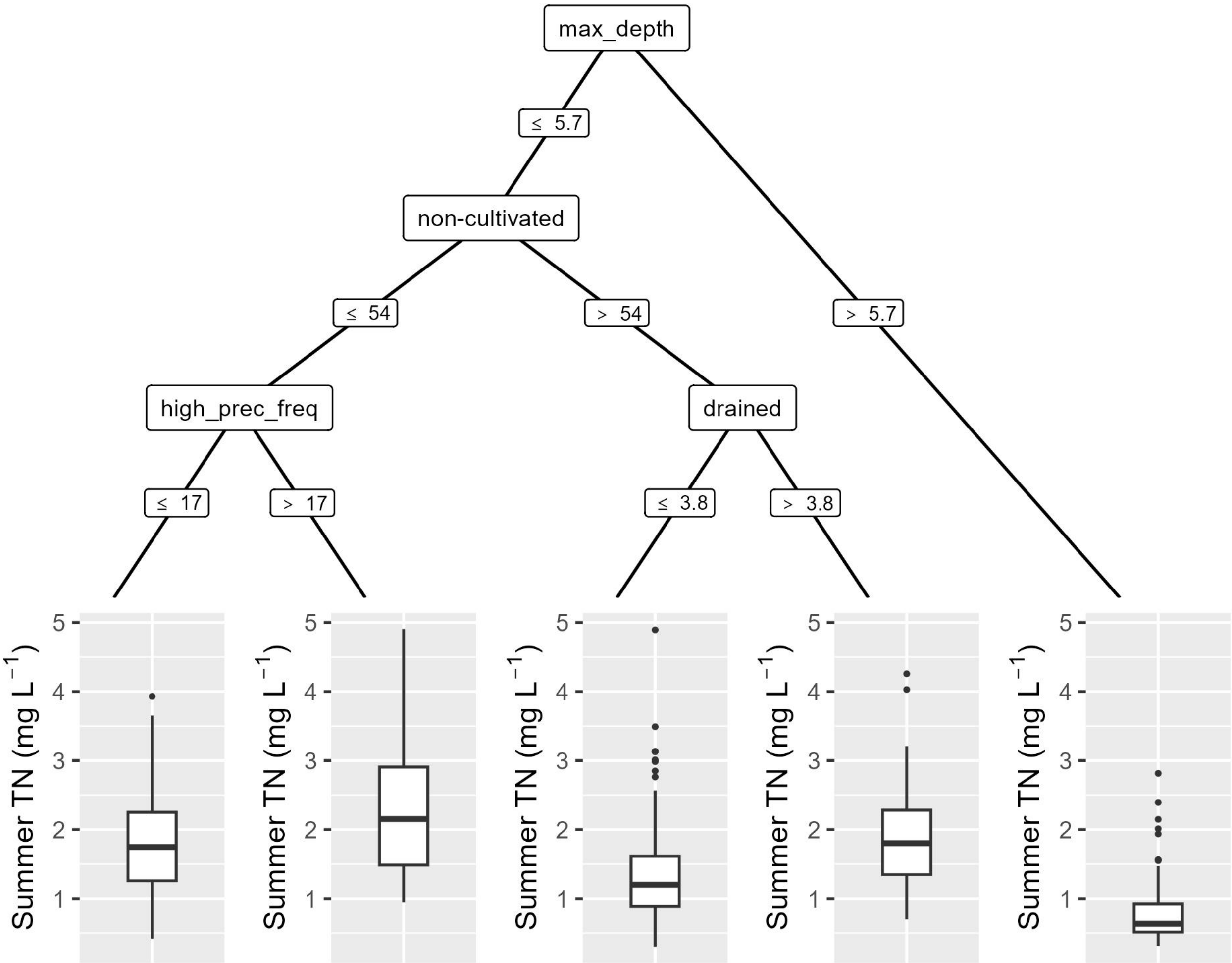



**Figure 4.14.** Conditional inference tree explaining summer Total Nitrogen concentrations in lakes. Landscape variables are extracted for catchment area. See Section 2 for variable explanations.

### Summer TP

Conditional inference trees for summer TP as response variable identified the percentage of potential drained area as the primary factor differentiating TP concentrations (Figure 4.15). Lakes with a higher drainage percentage (>71%) consistently exhibited higher TP concentrations. Among lakes with lower potential drained areas (71%), maximum depth further partitioned the lakes. Lakes deeper than 3.6 m had relatively lower TP concentrations, whereas shallower lakes showed higher summer TP concentrations. For shallower lakes, root depth (reflecting the catchment's soil infiltration capacity) was also significant. Lakes with catchments having shallower root depths (120 cm) tended to exhibit higher summer TP. The explained variance of the CIT was 0.16, indicating limited explanatory power of the model.

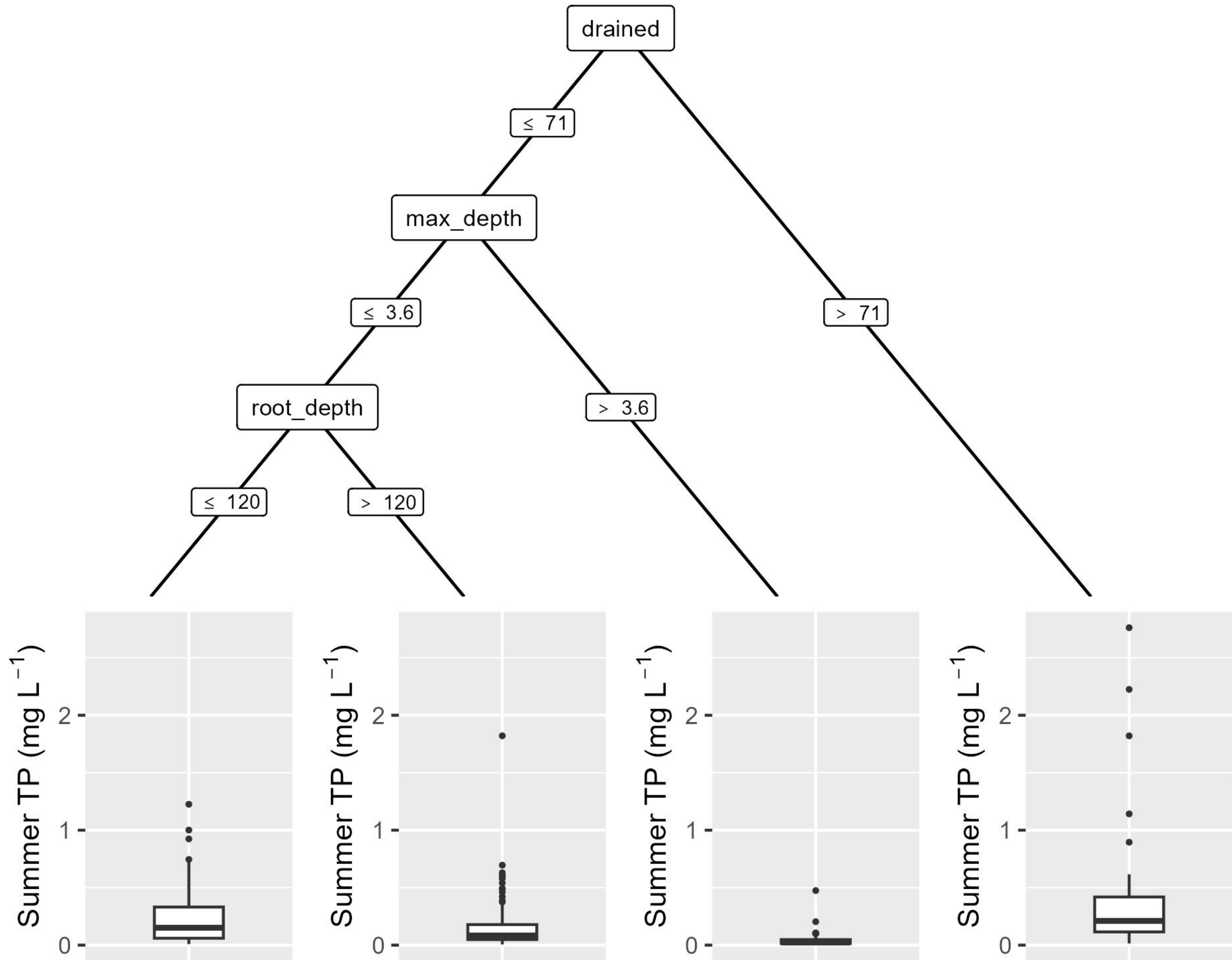



**Figure 4.15.** Conditional inference tree explaining summer Total Phosphorus concentrations in lakes. Landscape variables are extracted for catchment area. See Section 2 for variable explanations.

### Ecological status

Conditional inference trees using ecological status as response variable identified maximum depth as the primary factor influencing ecological status (Figure 4.16). Deeper lakes (>5.8 m) were generally characterized by better ecological status. Among shallower lakes, the percentage of non-cultivated areas further differentiated ecological status. Lakes with a higher proportion of non-cultivated areas (>83%) tended to have higher ecological status, whereas those with lower non-cultivated area percentages typically had lower ecological status. The overall accuracy of the tree was 0.35, with a kappa 0.16, indicating poor model performance.

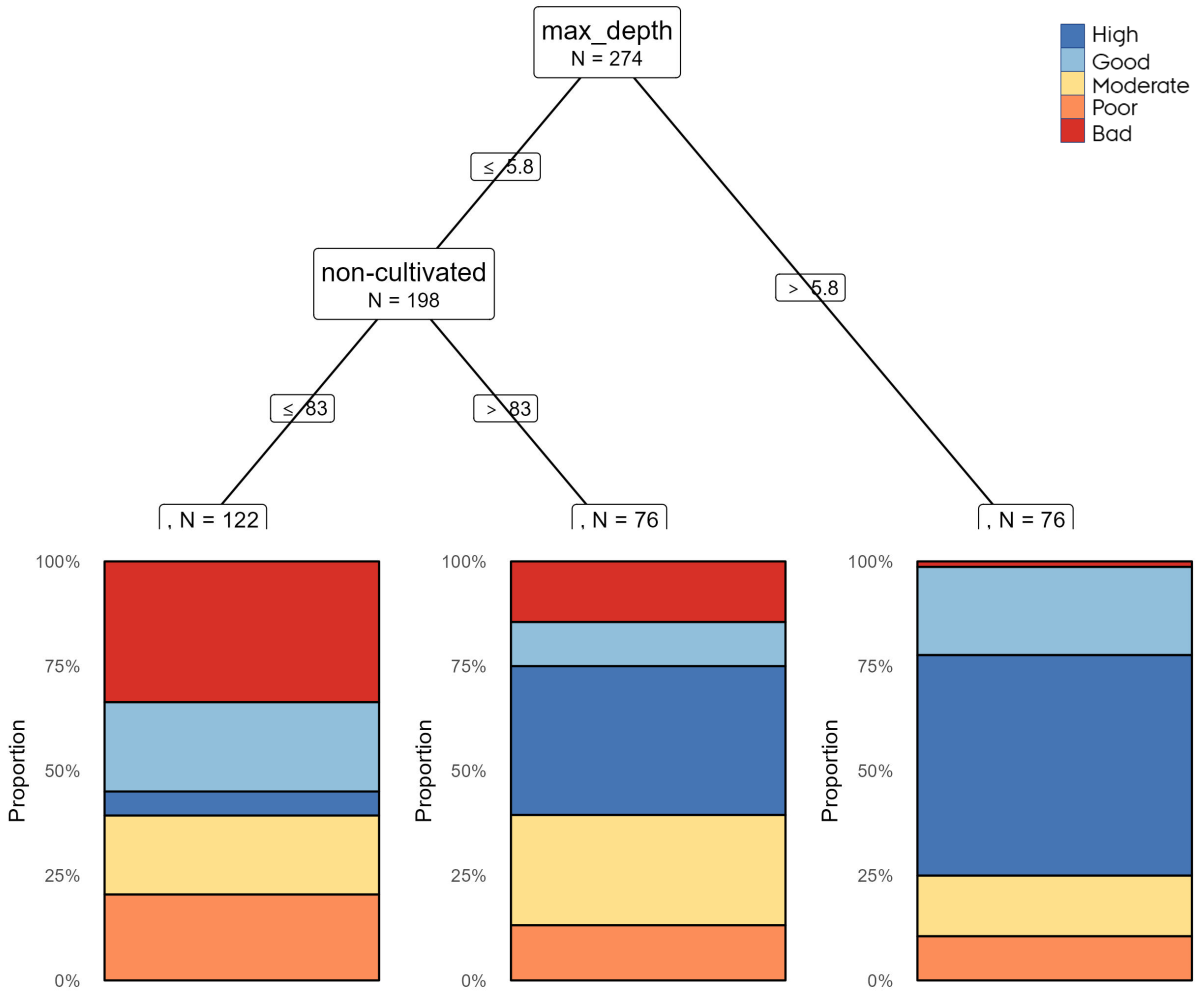



**Figure 4.16.** Conditional inference tree explaining ecological status in lakes. See Section 2 for variable explanations.

In summary, the conditional inference tree analyses revealed strong landscape effects despite low model accuracy. Maximum depth, percentage of drained area, and the proportion of non-cultivated land explained most of the variation in lake nutrient levels and ecological status. A higher proportion of non-cultivated land and lower drainage intensity were associated with reduced TN and TP concentrations. The relationship between landscape characteristics and ecological status was weaker, with only depth and the proportion of non-cultivated land distinguishing the lake groups. Overall, these findings highlight the role of lake depth and land use (% of non-cultivated land and drained area) in shaping nutrient dynamics.

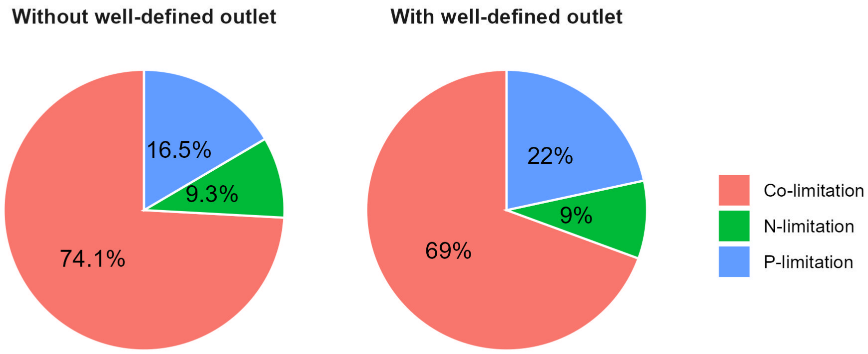
## 4.4 N or P limitation

### The stoichiometric approach

According to the nutrient limitation classification based on Ladwig et al. (in prep.), the majority of lakes without well-defined catchments (74%) were estimated to be co-limited, while 17% were P-limited and 9% were N-limited. In comparison, lakes with well-defined catchments had a higher proportion of P-limited lakes and a lower proportion of co-limited lakes than the lakes without well-defined catchments (Figure 4.17). In this group of lakes, 69% of the lakes were co-limited, 9% were N-limited lakes, and 22% were P-limited. According to Pearson's Chi-squared test, the difference between lakes with/without well-defined catchments was significant ($\chi^2$= 9.7632, $df$ = 2, $p$ =

0.0076). Although the difference between lakes with and without well-defined catchments is not large, these results confirm previous findings described in Søndergaard et al. (2023), suggesting that lakes without well-defined catchments are more likely to be N-limited.

**Figure 4.17.** Proportions of nutrient limitation classes for lakes without (left) and with well-defined drainage (right)




For lakes without well-defined catchments, N-limited lakes were predominantly associated with poor and bad ecological status (Figure 4.18). In contrast, P-limited lakes were more common among lakes with medium, good, and high ecological status. Co-limited lakes did not show a clear pattern across ecological status classes, occurring evenly across all categories. For lakes with well-defined catchments (Figure 4.18, right panel), the overall pattern was similar, with co-limitation strongly dominating all ecological states. However, differences between the two lake groups were noticeable. Lakes with well-defined catchments had fewer P-limited lakes in the high ecological status and more co-limited lakes across the moderate to bad states. This suggests that P-limitation is relatively more common in high-status lakes without clearly defined catchments, whereas in lakes with well-defined catchments, co-limitation dominated regardless of ecological status.

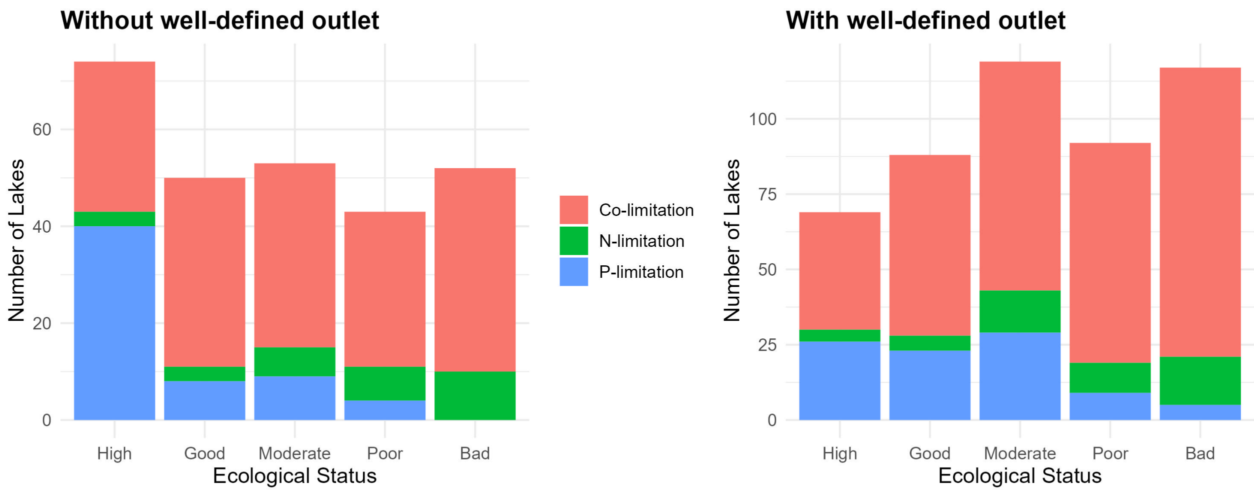



**Figure 4.18.** Nutrient limitation types across ecological status classes for lakes without (left) and with (right) well-defined catchments.

### The conditional inference tree approach

The conditional inference tree identified potential drainage percentage (*drained*) as the most important variable differentiating nutrient limitation classes (Figure 4.19). Lakes located in catchments with more than 71 % drained area exhibited a higher proportion of N-limited and co-limited lakes. For lakes with ≤71% drained area, the proportion of sandy soils in the catchment (*sand*) further structured the response. Among these lakes, those with ≤29% sandy soil were predominantly co-limited, whereas lakes with >29% sandy soil also included P-limited lakes along with co-limited lakes. For the lakes with <29% sandy soil, further split by the duration of high flow days (*high_Q_duration*) showed that the proportion of P-limited lakes increased with *high_Q_duration* >0.99. Overall, the model suggests that P-limited lakes are mostly found in catchments with higher percentages of sandy soils and lower drainage intensity, whereas N-limited lakes are associated with catchments characterized by higher drainage percentages. However, despite these patterns, the overall model performance was low, with an accuracy of 0.65 and a kappa value of -0.01, indicating little agreement beyond chance.

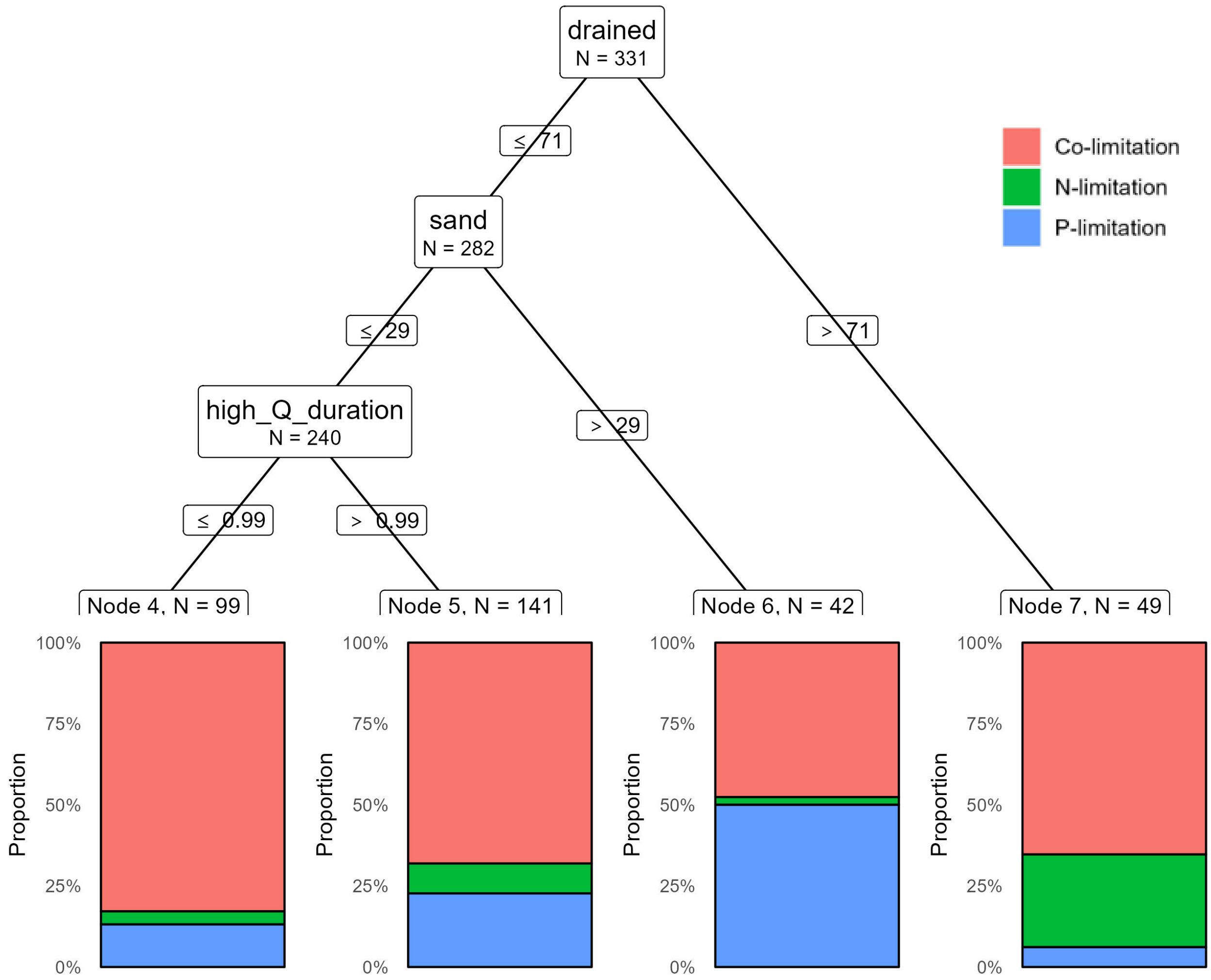



**Figure 4.19.** Conditional inference tree explaining limitation classes in lakes. See Section 2 for variable explanations.

## 4.5 Palaeoecological investigations

In the previous report, Søndergaard et al. (2023) suggested that palaeoecological investigations could be a useful tool to identify reference or near-reference ecological conditions in lakes. This is important in the context of the Water Framework Directive, as it helps determine the ecological state to aim for in the management of lakes. By analyzing macrofossil remains of plants and animals in different sediment depths together with sediment dating, these can provide insights into past ecological conditions, for example 100 or 200 years ago.

The historical status and reference conditions of lakes without well-defined catchments (and without inflow and outflow) are particularly poorly known. Indeed, their current 'impacted' state could also be their natural state; however, there is limited evidence for this. Because monitoring records are often too short to capture long-term ecological variability, palaeolimnological approaches can help reconstruct past conditions and distinguish natural variability from human-induced changes. Such investigations would therefore provide insights not only into the historical state or reference condition of a lake or lake type, but also into the current level of impact.

A palaeoecological investigation could involve (see also comments below):

1. Collecting long, > 100 cm, sediment cores from lake(s) without well-defined catchment, ideally from the deepest area and the littoral zone.
2. Conducting X-ray fluorescence (XRF) on the core combined with dry weight, organic carbon (loss on ignition), and wet density analysis.
3. Dating cores using Pb210 and Cs137.
4. Analyzing the sedimentary pigment concentration of the core. This can be done as a comparison of bottom versus surface concentrations; however, regularly spaced samples is a better approach.
5. Analyzing plant macro remains in the sediments.
6. Analyzing cladoceran remains in the sediments.

Dry weight and organic matter analysis provides essential background data for sediment core analysis.

XRF provides high resolution elemental analysis of sediment cores, which can reveal historical pollution levels.

Dating the core allows age determination of the different sediment depths to establish the reference period and assess the rate and timing of key changes.

Sedimentary pigment analysis can indicate the historical levels of chlorophyll *a* and pigments of blue green algae, both of which are important for defining the past ecological conditions of the lakes.

Macrofossil remains of plants provide information on historical plant communities present in the lake and their changes over time, helping to interpret ecological shifts. This provides a unique historical perspective on the current status of the lake in the context of past changes.

Cladoceran communities are shaped both by bottom-up forces, such as nutrient loading, and top-down forces, such as fish predation. As a result, changes in cladoceran communities enable integrated assessment of lake ecological status, reflecting the balance of benthic versus pelagic production and the intensity of fish predation. The combination of plant macrofossils and cladoceran analysis is particularly powerful in tracking past ecological shifts.

A detailed palaeolimnological investigation would provide a better overview of historical changes. It is recommended to take at least two cores per lake – one from the littoral zone and one from the pelagic zone. Both cores should be analyzed for dry weight, loss on ignition, and elemental composition (XRF), and subsequently dated. The pelagic core should be analyzed for cladoceran remains and sedimentary pigments, while the littoral core should be analyzed for plant remains, all from approximately 10 depths in the core.

Considering cost and feasibility, a preliminary qualitative assessment could also be conducted using a subset of samples from each core (e.g. a few from the bottom core depths together with a couple from the surface). If potentially meaningful changes are identified, a more detailed analysis, as suggested above, can be performed. This preliminary approach can provide a rough indication of ecological changes over the last few hundred years in the lake and help determine whether a more detailed analysis is recommendable.

# 5 Conclusions and recommendations

This study provides an updated analysis of how landscape characteristics may influence water quality, ecological status, and nutrient limitation classification in Danish lakes without well-defined topographic catchment boundaries. A multivariate statistical approach (redundancy analysis) and two machine learning methods (Random Forest, Conditional Inference Trees) were applied to reveal the potential effects of landscape characteristics on lake nutrients as well as ecological status. Together, these methods provide complementary insights into the potential effects of landscape characteristics on lake nutrients and ecological status.

RDA showed that 25–32% of the variation in lake parameters was explained by landscape variables. The percentage of organic soil, root depth, percentage of non-cultivated land, and percentage of potentially drained areas emerged as important predictors of in-lake conditions. Organic soil and drained areas were consistently associated with higher nutrient concentrations and lower ecological status, whereas a higher proportion of non-cultivated areas and higher root depths were linked to lower nutrient levels and higher ecological status. In contrast, mainly agricultural areas and permanent grassland areas were associated with lakes with poor/bad ecological status. This is in line with the common findings from the literature, as agricultural and intensely managed grasslands increase nitrogen and phosphorus losses through fertilizer application, soil erosion, and tile drainage, subsequently elevating in-lake nutrient concentrations and impairing ecological status (Nielsen et al., 2012; Jeppesen et al., 2007). Studies from Danish and other European lakes also highlight that nutrient export from agricultural land is a major nutrient source to Danish surface waters, and that catchments dominated by intensive agriculture tend to support eutrophic conditions (Jeppesen et al., 2007).

Catchment characteristics, such as total water availability, were also significant at all buffer distances, and precipitation frequency was significant at 10 m, 50 m, and 250 m buffers. Overall, wetter catchments with a higher proportion of organic soils and greater drainage tended to be associated with poorer water quality. Similarly, a study of 52 temperate lakes, including Danish lakes, showed that catchment soil characteristics, particularly soil organic carbon, were strong predictors of lake dissolved organic carbon, total organic nitrogen, and TP, with higher proportions of organic soils associated with elevated nutrient levels, indicating that catchment soil type strongly shapes lake nutrient status (Sepp et al., 2022).

Buffer zone size influenced the significance levels of the predictors; however, the same predictors remained largely significant, and the overall patterns were consistent across scales. As buffer size increased, the effect of potential drainage area weakened, whereas the importance of non-cultivated areas became stronger. The effect of drainage has also been highlighted in previous publications on water quality (Søndergaard, 2023; Nielsen et al., 2012). Nielsen et al. (2012) showed that artificial tile drainage systems can account for a large share of nitrogen exported from agricultural fields, as a significant portion of Danish arable land is subsurface-drained, and such drains bypass natural soil and groundwater transport pathways (Olesen, 2009; Møller et al., 2018b).

The random forest model struggled to accurately classify lakes with good, moderate, and poor ecological status, but it generally succeeded in achieving a sufficient distinction between high and bad ecological status. This may reflect either lack of information on important covariates driving ecological conditions in these lakes or a high level of noise in the data, complicating accurate classification. Nonetheless, the models consistently identified important features for classification: maximum lake depth, organic soil, lake area, streamflow precipitation elasticity, drainage percentage, non-cultivated land, and grassland. Lakes in high ecological status were characterized by higher maximum lake depths, low organic soil (%), higher lake area, low streamflow elasticity, low amount of drained area, higher amount of non-cultivated land, and less grassland. Conversely, lakes in bad condition showed the opposite dynamics, except for streamflow elasticity.

Conditional inference tree analysis highlighted clear threshold relationships, although model accuracy and explained variance were low. Maximum depth was the strongest predictor of both TN and ecological status, with deeper lakes (>5.7–5.8 m) consistently exhibiting lower nutrient levels and better ecological conditions. Among shallower lakes, drainage intensity, non-cultivated land, and precipitation extremes further structured nutrient concentrations and ecological status. An increasing proportion of non-cultivated area, reduced drainage, and fewer heavy precipitation events were associated with lower TN concentrations. For TP, the percentages of drained area and depth were also important, as deeper lakes tended to have lower TP and a lower percentage of drained area. The relationship between landscape characteristics and ecological status was weaker, with only depth and percentage of non-cultivated land splitting lake groups. Lakes with a maximum depth greater than 5.8 m generally had better ecological status, while for shallower lakes, the percentage of non-cultivated land was important. Shallow lakes with less than 83% non-cultivated land had lower ecological status. Lake typology is important to understand the mechanisms since shallow lakes are typically polymictic, with frequent mixing and sediment–water interactions enhancing nutrient availability, whereas deeper lakes stratify for longer periods, altering internal nutrient dynamics. Because most lakes in our dataset were shallow, we examined whether depth was merely separating shallow and deep types, or if it was also informative within shallow systems. The conditional inference tree indicated that depth formed the first split, but that additional landscape and catchment variables structured variation among shallow lakes. This pattern suggests that depth organizes the primary response, while catchment controls modulate nutrient levels within shallow systems.

Nutrient limitation type (N- P-, or co-limited) was associated with drainage intensity, sandy soil percentage, and hydrological activity. P-limited lakes were most common in sandy, weakly drained catchments, while N-limited lakes primarily occurred in heavily drained systems, again in consistence with the expected hydrological controls on nutrient flows.

Martinsen & Sand-Jensen (2022) similarly identified clear landscape–lake relationships across Danish lakes in the national surface monitoring programme; however, the most important predictors in their analysis were primarily terrain-based variables such as curvature and terrain ruggedness, while the effects of land-use variables were relatively weak. This difference can be explained by the broader topographic gradients represented in their national dataset. In our study, we analyzed only a subset of Danish lakes located in low-relief landscapes, where variation in terrain variables is minimal.

In such settings, terrain-based predictors become less informative, and land-use and soil characteristics account for more of the observed variation. Consistent with the previous report using linear regression (Søndergaard et al., 2023), the use of non-linear machine learning also had low explanatory power. Overall, the results confirm that while landscape variables explain only part of the variation in lake conditions, several catchment-scale mechanisms strongly influence nutrient levels in lakes without well-defined catchment areas. These include the percentage of drained areas, percentage of non-cultivated areas, soil composition, precipitation extremes, and lake morphology. The lower explanatory power also underscores the continuing influence of other factors, such as internal loading, mixing dynamics, and groundwater contributions, which were not fully captured by surface-based catchment delineations. Furthermore, the present state of lakes could also be influenced by historical nutrient inputs, which are not reflected in the present catchment characteristics. However, it should also be emphasized that the ecological status used in this report was based solely on chlorophyll *a* value, whereas ecological status is formally assessed using all five biological quality elements. Due to the complex interactions within lake ecosystems, the relationships between ecological status and catchment characteristics can be obscured by internal processes and feedback, making statistical associations difficult to detect.

## Recommendations

Although the explanatory power of the models is low, the results consistently identify several landscape characteristics that are important for explaining water quality and ecological status in lakes without well-defined topographical catchments.

First, reducing drainage intensity – particularly in catchments with high proportions of organic soils – could help limit the nutrient transport to lakes, drained organic soils being repeatedly associated with elevated nutrient concentrations. A reduction of artificial drainage networks may therefore provide measurable benefits.

Second, increasing the extent of non-cultivated areas or protecting existing ones may also buffer nutrient inputs, given their association with lower nutrient levels and better ecological status.

Finally, management strategies should explicitly consider the composition of soils. Soils with a high content of organic substances combined with intensive drainage generated the highest nutrient concentrations, suggesting that lake restoration efforts in these areas may be especially effective in reducing nutrient loads and should be prioritized. Due to the relatively low proportion of explained variation, these results should be viewed as precautionary, complementary measures to be combined with assessments of internal loading, groundwater connectivity, lake-specific processes, and palaeolimnological evidence.

# 6 References


Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10), 5293–5313.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Coxon, G., et al. (2020). CAMELS-GB: Hydrometeorological time series and landscape attributes for 671 catchments in Great Britain. *Earth System Science Data*, 12(4), 2459–2483.

Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., & Otero-Casal, C. (2017). Hydrologic regulation of plant rooting depth. *Proceedings of the National Academy of Sciences*, *114*(40), 10572-10577.

Greve, M. B. (2024). *Revidering af jordbundstypekort 2024*. Rådgivningsnotat fra DCA – Nationalt Center for Fødevarer og Jordbrug, Aarhus Universitet, 4 sider. Leveret 17.05.2024.

Hothorn, T., Seibold, H., & Zeileis, A. (2015). *partykit: A toolkit for recursive partitioning*.

Höge, M., et al. (2023). CAMELS-CH: Hydrometeorological time series and landscape attributes for 331 catchments in Switzerland. *Earth System Science Data*, 15, 5755–5784.

Jeppesen, E., Søndergaard, M., Lauridsen, T. L., Kronvang, B., Beklioglu, M., Lammens, E., ... & Tátrai, I. (2007). Danish and other European experiences in managing shallow lakes. *Lake and Reservoir Management*, 23(4), 439-451.

Kuhn, M., & Vaughan, D. (2025). *parsnip: A common API to modeling and analysis functions* (R package version 1.4.0). https://github.com/tidymodels/parsnip

Ladwig, R., Bucak, T., Johansson, L.S., Berthelsen, A., Søndergaard, M. (in prep). Nitrogen as a limiting nutrient in Danish lakes. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi.

Liu, J., Koch, J., Stisen, S., Troldborg, L., Højberg, A. L., Thodsen, H., ... & Schneider, R. J. (2025). CAMELS-DK: hydrometeorological time series and landscape attributes for 3330 Danish catchments with streamflow observations from 304 gauged stations. *Earth System Science Data*, *17*(4), 1551-1572.

Martinsen, K. T., & Sand-Jensen, K. (2022). Predicting water quality from geospatial lake, catchment, and buffer zone characteristics in temperate lowland lakes. *Science of the Total Environment*, 851, 158090.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, *22*(3), 276-282.

Molnar C, Bischl B, Casalicchio G (2018). "iml: An R package for Interpretable Machine Learning." *JOSS*, **3**(26), 786. https://doi:10.21105/joss.00786, https://joss.theoj.org/papers/10.21105/joss.00786.

Moon, D. L., Scott, J. T., & Johnson, T. R. (2021). Stoichiometric imbalances complicate prediction of phytoplankton biomass in U.S. lakes: Implications for nutrient criteria. *Limnology and Oceanography*, 66(8), 2967–2978.

Møller, A. B., Greve, M. H., & Beucher, A. M. (2024). *Opdateret jordbundstype-kort*. Rådgivningsnotat fra DCA – Nationalt Center for Fødevarer og Jordbrug, Aarhus Universitet, 33 sider. Leveret 19.03.2024.

Møller, A. B., Beucher, A., Iversen, B. V., & Greve, M. H. (2018a). Predicting artificially drained areas by means of a selective model ensemble. *Geoderma*, 320, 30–42.

Møller, A. B., Børgesen, C. D., Bach, E. O., Iversen, B. V., & Moeslund, B. (2018b). *DCA-rapport nr. 135*. DCA – Nationalt Center for Fødevarer og Jordbrug.

Nielsen, A., Trolle, D., Søndergaard, M., Lauridsen, T. L., Bjerring, R., Olesen, J. E., & Jeppesen, E. (2012). Watershed land use effects on lake water quality in Denmark. *Ecological Applications*, 22(4), 1187-1200.

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., et al. (2025). *vegan: Community ecology package*. https://doi.org/10.32614/CRAN.package.vegan

Palarea-Albaladejo, J., & Martín-Fernández, J. A. (2015). zCompositions: R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems*, 143, 85–96. https://doi.org/10.1016/j.chemolab.2015.02.019

R Core Team (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rock, L. A., & Collins, S. M. (2024). A broad-scale look at nutrient limitation and a shift toward co-limitation in United States lakes. *Environmental Science & Technology*, 58(26), 11482–11491.

Sepp, M., Kõiv, T., Nõges, P., Nõges, T., Newell, S. E., & McCarthy, M. J. (2022). Catchment soil characteristics predict organic carbon, nitrogen, and phosphorus levels in temperate lakes. *Freshwater Science*, 41(1), 1–17.

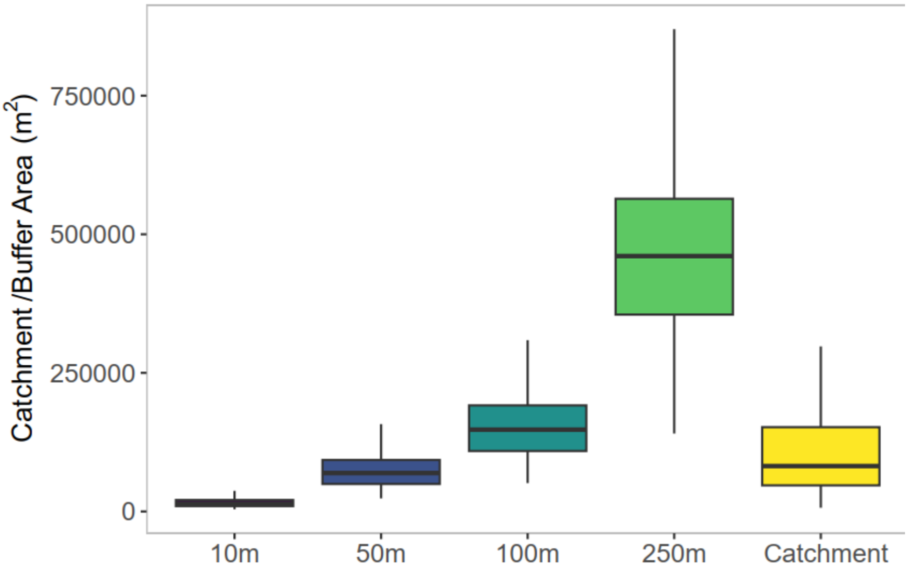
Søndergaard, M., Johansson, L. S., & Kjeldgaard, A. (2023). *Sammenhænge mellem oplandskarakteristika og tilstanden i søer uden veldefinerede afløb*. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 32 s. Videnskabelig rapport nr. 548.

van den Boogaart K.G., Tolosana-Delgado R. and Bren M. (2025). compositions: Compositional Data Analysis. doi:10.32614/CRAN.package.compositions. R package version 2.0-9.

Wright MN, Ziegler A (2017). "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." *Journal of Statistical Software*, 77(1), 1–17. https://doi:10.18637/jss.v077.i01.

# 7 Appendix

**Figure 7.1.** Boxplots comparing the areas of buffer strips and catchments. Outliers are removed for visual purposes. In boxplot, horizontal solid line indicates median of the distribution, the box represents the lower to upper quartile values of the data, the whiskers extend to the last data point beyond 1.5 * the interquartile range, circles represent outliers beyond this range



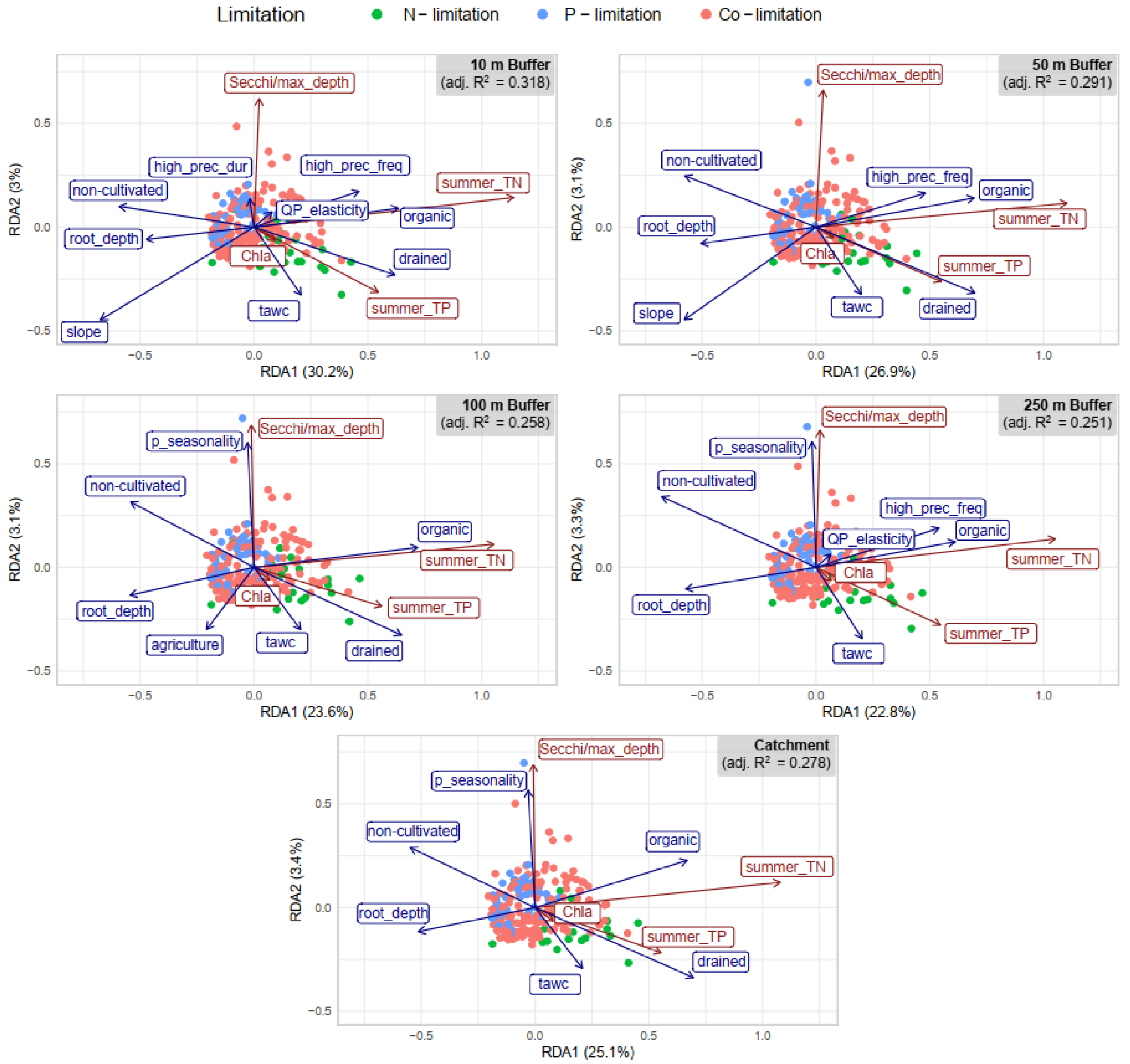



**Figure 7.2.** Results from RDA analysis for different buffer distances (10 m, 50 m, 100 m, 250 m) and catchments. Blue arrows indicate statistically significant predictors automatically selected using backward selection, while the red arrows represent in-lake response variables. Each point represents a lake, and the colors indicate nutrient limitation classes: N limitation (green), P limitation (blue), and co-limitation (red). See Section 2 for variable explanations.

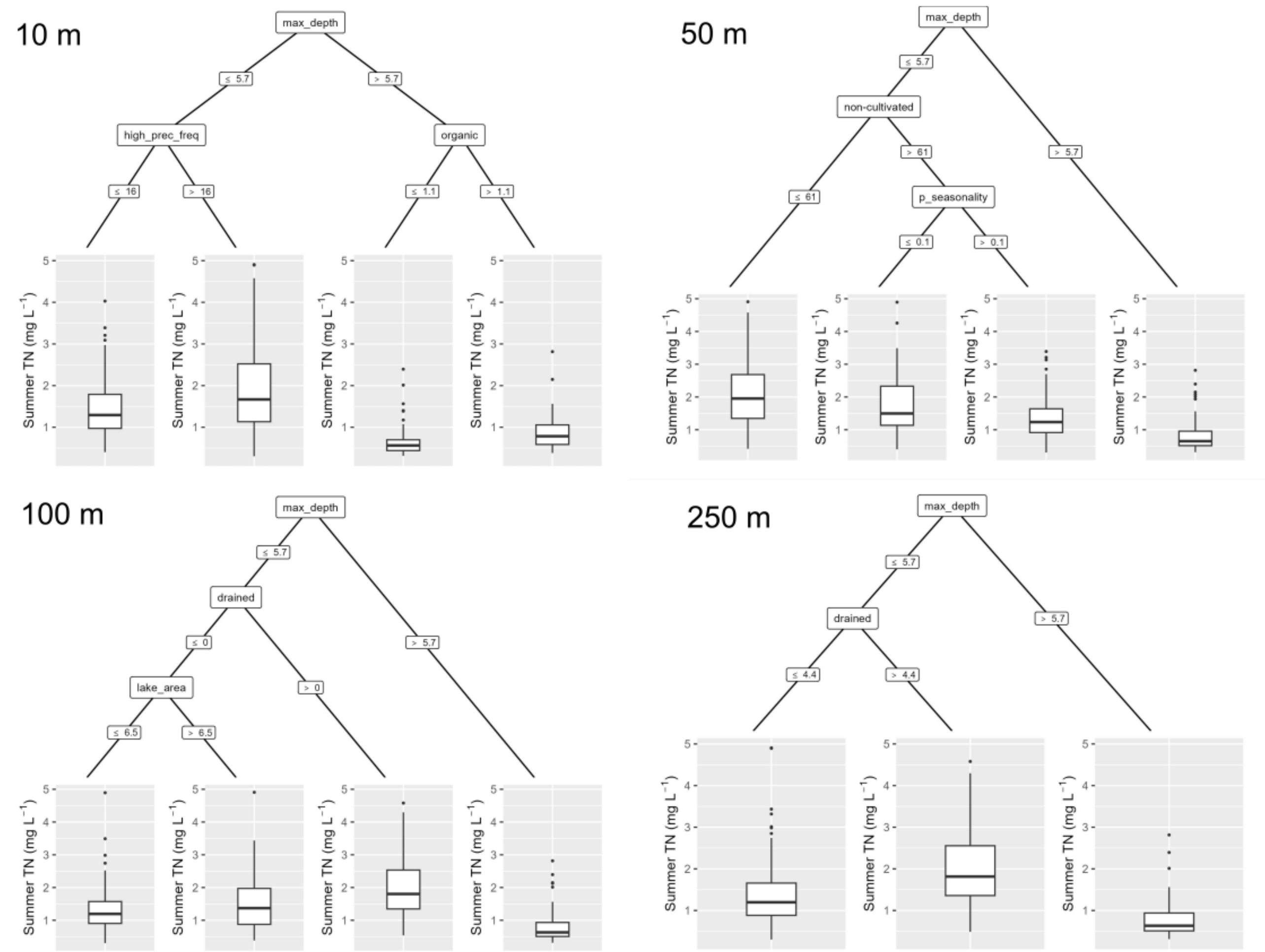



**Figure 7.3.** Conditional inference tree explaining summer total nitrogen (summer_TN) concentrations in lakes for different buffer distances (10m, 50m, 100m, 250m). See Section 2 variable explanations.

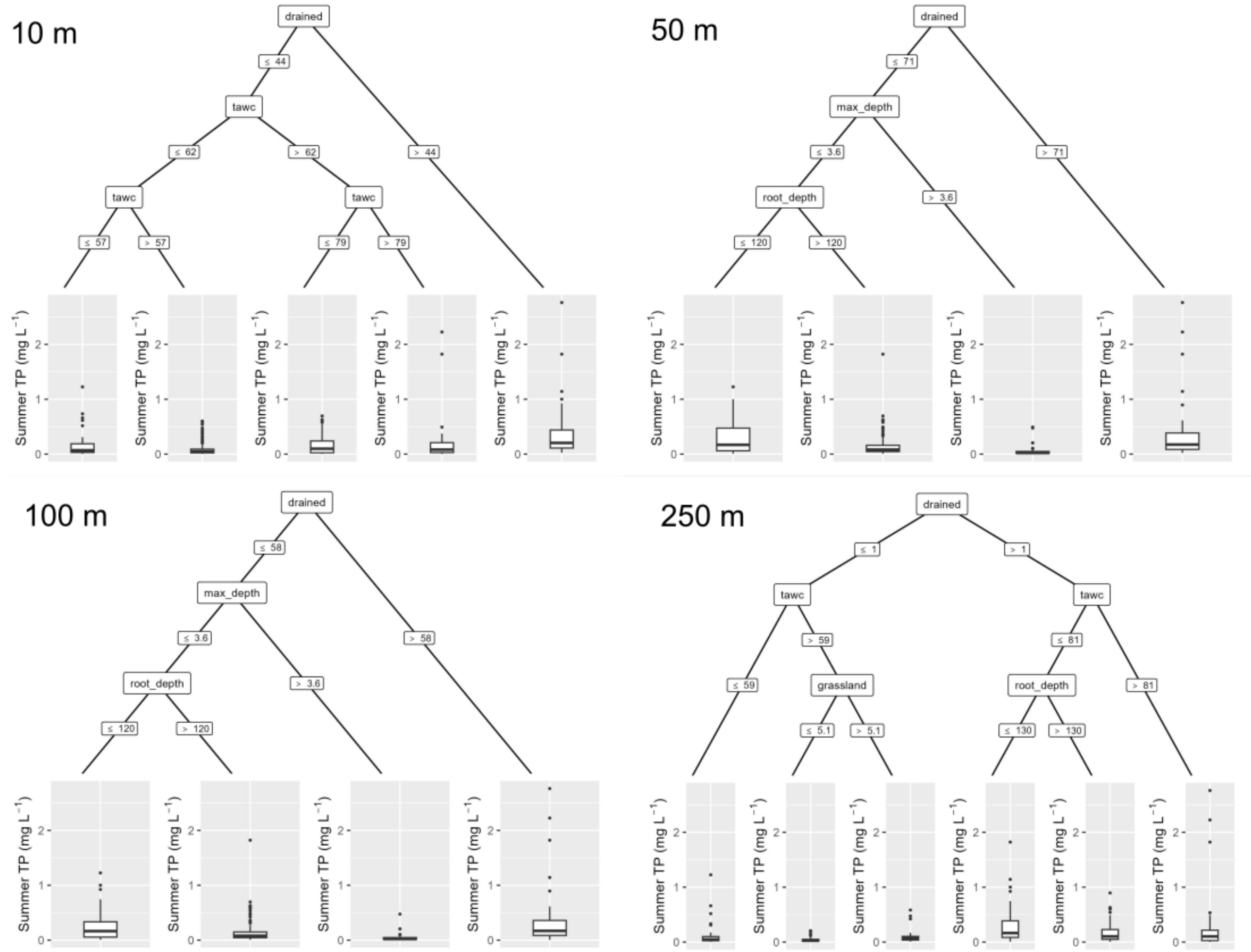



**Figure 7.4.** Conditional inference tree explaining summer total phosphorus concentrations in lakes. For different buffer distances (10m, 50m, 100m, 250m). See Section 2 variable explanations.

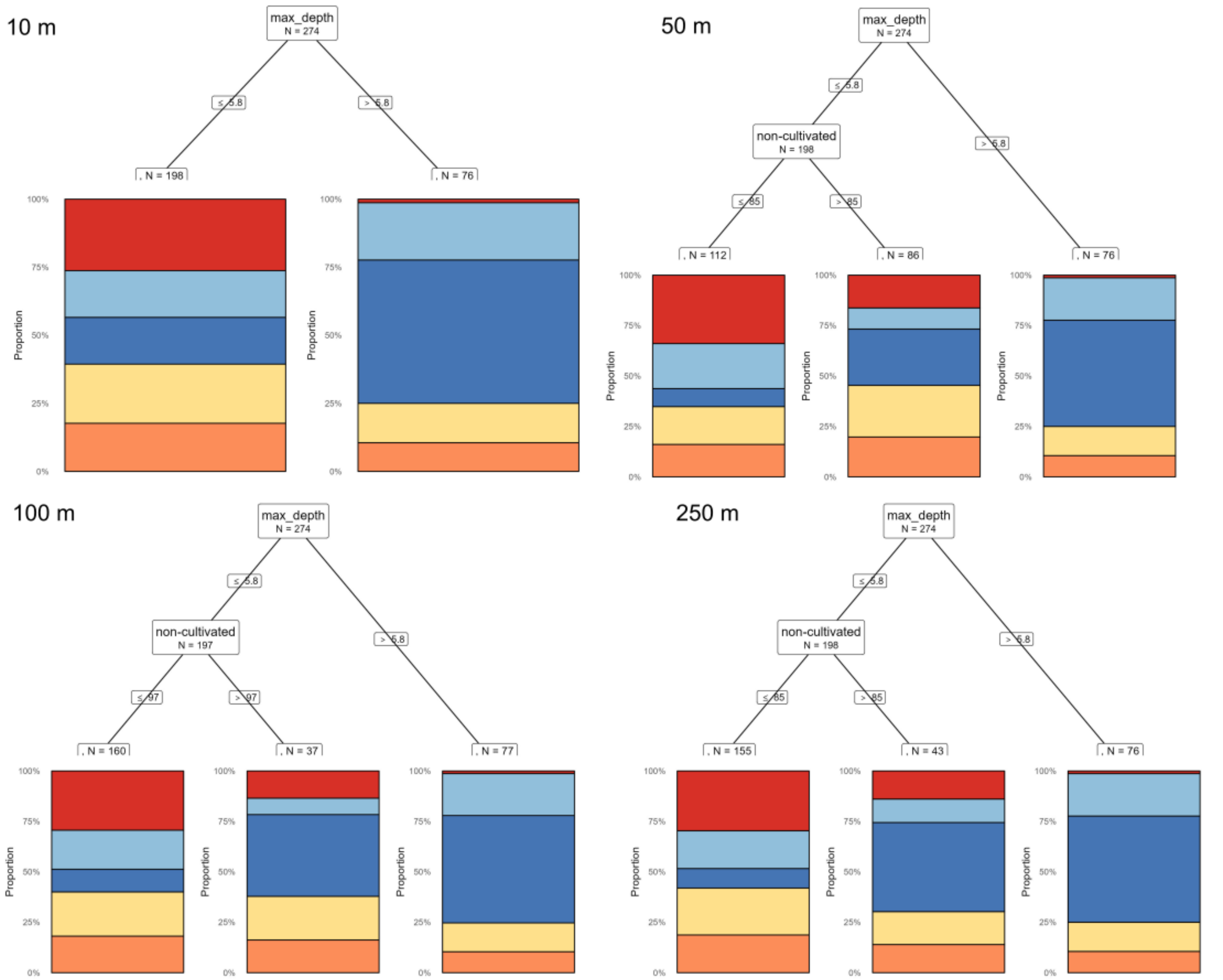



**Figure 7.5.** Conditional inference tree explaining ecological status in lakes in lakes for different buffer distances (10m, 50m, 100m, 250m). See Section 2 variable explanations.

## LAKES WITHOUT WELL-DEFINED OUTLETS – TOPOGRAPHIC CATCHMENTS AND EFFORTS

Many Danish lakes included in the 3rd River Basin Management Plan (VP3, 2021–2027) lack well-defined catchments, preventing load assessments and targeted restoration despite failing to achieve good ecological status. Using multivariate statistics (RDA) and machine-learning approaches (Random Forest and Conditional Inference Trees), this study evaluated how landscape characteristics influence lake water quality, ecological status, and nutrient limitation. All methods showed low explanatory power, indicating that lake conditions are only partly explained by surface catchment characteristics and are likely influenced by internal loading, historical pressures, and hydrological processes. Nevertheless, poorer ecological conditions were consistently linked to organic soils, intensive drainage, and wetter catchments, while non-cultivated land and deeper lakes were associated with improved ecological status. These findings suggest that reducing artificial drainage and protecting natural areas can support improvements in water quality and ecological status.

ISBN: 978-87-7648-030-1
ISSN: 2244-9981