**The School of Mathematics**


THE UNIVERSITY
*of* EDINBURGH

# Freshwater Macroinvertebrate Time Trends in England's Rivers: A Seasonal Censored-Normal GAMM Approach

**by**

**Tilak Pundalik Heble**
**s2563426**

Dissertation Presented for the Degree of
**MSc in Statistics with Data Science**

July 15th – August 15th 2025

Supervised by
**Simon Wood, Mike Dunbar and Nicole Augustin**

# Executive Summary

The Environment Agency has collected river invertebrate records across England for more than 30 years. Our goal was to turn these records into clear, national-level lines showing how the abundance of key invertebrate groups has changed over time, in a way that matches how the data were actually recorded.We focus on Aphelocheiridae, Brachycentridae, Odontoceridae, and Cordulegastridae. In practice each of these is represented by a single species in the dataset, which keeps identification consistent through time. They also have enough records since 1990 to support reliable national summaries. (Potamanthidae was left out because there were too few records.)

**What's special about the data ?** Records come mostly from spring and autumn, when surveys are routinely done. Abundance was recorded in two ways: either as broad categories (0, 1–9, 10–99, 100–999, ...) or as numbers that are often rounded and sometimes based on sub-sampling. Because of this, many values are better treated as ranges rather than exact counts.

**How we prepared the data!!** We built a clean list of "site × date" visits from EA open data, kept the standard field method, focused on years 1990 onward (when procedures were aligned nationally), and kept just one sample per site and day. Using the detailed taxon table, we labelled each sample as categorical or numeric and handled them accordingly. For each family, we then joined the abundances and added back zeros where the family wasn't found at sites where it appears at least once.

**How we analysed it ?** We built up the story in four easy steps. First, we checked whether each family was simply found or not found over time. Then we looked at whether samples were shifting into higher or lower abundance bands. Next, where the data had numbers, we treated those numbers as sensible ranges (because many are rounded) and looked for patterns. Finally, we combined these ideas in our main approach, which (1) uses ranges, (2) lets spring and autumn have their own patterns, and (3) prevents a few busy sites from dominating the national picture. We compared results from all four steps to make sure our conclusions didn't rely on just one way of looking at the data.

**What we found ?** Across the four families we see a common picture: a decline into the early 2010s, followed by a recovery through the 2020s. Levels are generally higher in autumn than in spring. Where data are sparse (early years and outside spring/autumn), the shaded uncertainty bands widen, as expected. The three cross-checks tell a consistent story in both direction and overall shape.

**How to use these results ?** For routine reporting we recommend showing the national season-average line with its uncertainty band, and adding separate spring and autumn lines where seasonal differences matter. These are national indicators, meant for tracking broad change and comparing families—not for site-by-site decisions.

**Points to keep in mind!!** Recording practices varied over time and region; some counts are rounded or sub-sampled; surveys outside spring/autumn are limited; and the monitoring network has changed. We account for these factors as far as possible, but they do increase uncertainty in places.

**Where to go next ?** Future improvements include adding environmental information (flow, temperature, nutrients) to explain differences, bringing in location-aware summaries to separate regional from national change, calibrating the categories using more recent numeric data, and running sensitivity checks to show the results are stable under reasonable alternatives.

Despite quirks in how the data were recorded, the national picture is clear—many river invertebrates dipped around the early 2010s and are now recovering—giving the EA a practical, season-aware indicator to track progress and target action.

# Acknowledgments

I would like to thank **Mike Dunbar** (Environment Agency scientist) for his domain expertise and guidance on interpreting the Environment Agency macroinvertebrate monitoring protocols and data. I am also grateful to my instructors, **Simon N. Wood** and **Nicole Augustin**, for their methodological direction, clear teaching, and constructive feedback throughout this project. Special thanks to **Antoni Sieminski** (PhD student) for practical assistance and timely advice on data processing and model diagnostics.

I gratefully acknowledge the **Environment Agency (England)** for maintaining and providing access to the national macroinvertebrate datasets used in this study, and the **School of Mathematics, University of Edinburgh**, for supporting this work. Finally, I thank my family and friends for their encouragement and support.

*AI Acknowledgment:* I occasionally used OpenAI's ChatGPT to interpret R error messages, clarify modelling concepts, suggest code refactoring, draft/review inline comments, and help with LaTeX formatting (section structure, tables, and figure placement). All analyses, modelling decisions, and final code are my own; no data were shared with the tool.

Word Count of Report is 4991 words.

# University of Edinburgh – Own Work Declaration

Name: **Tilak Pundalik Heble**

Matriculation Number: **s2563426**

Title of work: **Freshwater Macroinvertebrate Time Trends in England's Rivers: A Seasonal Censored-Normal GAMM Approach**

I confirm that all this work is my own except where indicated, and that I have:

- Clearly referenced/listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Not sought or used the help of any external professional academic agencies for the work
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Complied with any other plagiarism criteria specified in the Course handbook

I understand that any false claim for this work will be penalised in accordance with the University regulations (`https://teaching.maths.ed.ac.uk/main/msc-students/msc-programmes/statistics/data-science/assessment/academic-misconduct`).

Sincerely,

Signature: Tilak Pundalik Heble

Date: 15 August 2025

# Contents

# List of Tables



# List of Figures

# 1 Introduction

## 1.1 Background

Macroinvertebrates are widely used as sentinels of river health because community composition and abundance integrate the effects of nutrients, organic enrichment, habitat alteration, and toxicants over ecologically meaningful timescales. The Environment Agency (EA) operates a comprehensive, long-running biomonitoring programme for rivers in England, yielding multi-decadal observations suitable for trend assessment and operational decision-making. This study estimates national time trends in macroinvertebrate abundance for four **focal families**—Aphelocheiridae, Brachycentridae, Odontoceridae, and Cordulegastridae—selected to minimise taxonomic aggregation artefacts and to provide interpretable indicators for management. We analyse a multi-decadal record spanning a nationwide network of river sites with repeated spring and autumn campaigns.

(a) Kick-sampling during an EA field survey.

(b) Bottle sampling during routine monitoring.

Figure 1: Field procedures for EA river biomonitoring.

Two features of the monitoring data shape the statistical problem. First, **measurement heterogeneity**: records combine ordinal AB classes (e.g., AB1–AB4+) with numeric counts commonly recorded to **one significant figure**—often via subsampling—so many observations are naturally **interval-valued** (e.g., a recorded 17 corresponds to a band in the mid-teens to mid-twenties). Second, **seasonality**: most sampling occurs in **spring (Mar–May)** and **autumn (Sep–Nov)**, and life-cycle dynamics (emergence, growth, cohort turnover) mean trajectories can differ by season. Ignoring either feature risks biased inference or overconfident conclusions. In addition, site-level heterogeneity is substantial—catchments differ in climate, chemistry, size, and habitat—and effort is uneven across space and time. Any national analysis must therefore control for site effects while producing **population-level** summaries that are actionable for surveillance.

## 1.2 Scientific and Practical Significance

Our aim is to estimate national macroinvertebrate abundance trends with uncertainty that respects mixed data types and temporal structure, and to report population-level curves suitable for surveillance. To address the challenges above, the primary analysis uses a **censored-normal generalized additive model (GAM)** with **seasonal smooths**—the **seasonal cnorm GAM**. Each observation is treated as an **interval** on a variance-stabilising $p$**-power scale** (a transformation $x \mapsto x^p$ chosen to regularise mean–variance behaviour) so that interval censoring naturally reflects one-significant-figure rounding. On this transformed scale we model the response with (i) a **season effect** and a

**season-varying smooth** of time to capture distinct spring and autumn dynamics, and (ii) a **random intercept for site** to absorb persistent spatial heterogeneity. For national trend estimation we **exclude random effects at prediction time**, yielding **site-marginal (population-level)** curves with **95% confidence bands** that are comparable across families and seasons. The value of $p$ is selected by a small grid search to stabilise residual spread while maintaining interpretability. This framework accommodates the mixture of categorical and rounded numeric data, respects interval uncertainty, and represents temporal structure flexibly without imposing parametric trend shapes (8; 9). The pipeline uses EA open data and modern R tooling for modelling and diagnostics (2; 6) and is fully reproducible.

(a) Crustacean macroinvertebrate (shrimp).

(b) Freshwater hydra (Cnidaria).

Figure 2: Example macroinvertebrate taxa encountered in routine samples.

Because phenology and sampling windows differ, a single smooth risks averaging away genuine seasonal signals. We therefore compare a **single-trend** specification with the **season-varying** specification using information criteria and explained deviance, retaining the seasonal model where fit and diagnostics improve materially. Goodness-of-fit is evaluated using **randomised quantile residuals** (RQRs), QQ plots, residual autocorrelation, and effective-degrees-of-freedom checks; detailed panels are provided in the appendix (1; 9). To increase confidence that conclusions are not artefacts of distributional assumptions, we also fit complementary models for **triangulation**: a **presence/absence (binomial) GAMM**, an **ordered-categorical (OCAT) GAMM**, and a **censored Poisson** GAMM. Concordance in trend direction and shape across these models strengthens inference.

### 1.3 Report Structure

**Report Section Overview**

**The remainder of this report is structured as follows:**

- **Section 2: Data—Description, Cleaning, and Preprocessing**
  Datasets and scope; visit roster construction; family abundances and site metadata; BIN/-COUNT classification via `TAXA` with overrides; season coding; per–family modelling tables; and quality checks.
- **Section 3: Exploratory Data Analysis (EDA)**
  Sampling effort, family frequencies, recording–mode proportions, seasonal coverage, rounding fingerprints, and presence/abundance summaries that motivate the modelling choices.
- **Section 4: Model Methodology**
  Common GAM structure; *PA* (binomial), *OCAT* (ordered categorical), *cPois* (interval–censored Poisson), and *cNorm* (variance–stabilised censored normal); single vs. season–varying smooths; estimation, prediction, diagnostics, and "what each model answers".
- **Section 5: Results**
  Primary seasonal trends from the cNorm model (worked example and family summary), diagnostics for the exemplar family, fit metrics across families, triangulation with PA/OCAT/cPois, and a brief note on cNorm variants and model selection.
- **Section 6: Discussion**
  Interpretation of national trends; strengths of the design; limitations and cautions; implications and recommended use; and priorities for future work.
- **Section 7: Conclusion**
  Concise take-aways for surveillance and next steps.
- **References and Appendices**
  Bibliography followed by appendices with extended cNorm outputs for remaining families and per-family results for PA, OCAT, and cPois, plus supplementary figures and tables.

# 2 Data: Description, Cleaning, and Preprocessing

## 2.1 Datasets and Scope

We use four Environment Agency (EA) open datasets plus a harmonised family-level table. Each is included for a specific reason, and all are reduced to the minimal fields needed so joins are stable and the analysis remains auditable.

- **METRICS (INV_OPEN_DATA_METRICS)** — used to define the *visit roster* (one potential row per site–date). Establishing this roster first ensures that every downstream join and inclusion rule is anchored to a reproducible set of sampling occasions.
- **WHPT family abundances (R_INV_WHPT_METRICS_B)** — used as the source of family-level responses for modelling. Using the harmonised WHPT table avoids inconsistencies from taxonomic practice changes and ensures that each focal family has a single, consistent abundance field.
- **SITES (INV_OPEN_DATA_SITE)** — included to provide reporting area, catchment, water body, and BNG coordinates for summaries and maps. Retaining only these fields keeps interpretation clear while enabling optional spatial context.
- **TAXA (INV_OPEN_DATA_TAXA)** — consulted solely to determine whether a sample was recorded categorically (AB classes) or numerically (counts). This step is essential because modelling must respect interval uncertainty when counts are rounded or categorical.

Table 1 summarises the inputs used in this section.

Table 1: EA datasets used and their roles in this analysis.

| Dataset | Role in pipeline |
|---|---|
| INV_OPEN_DATA_METRICS | Visit roster (site–date, methods); S3PO-only; ≥1990; de-dup same-day. |
| R_INV_WHPT_METRICS_B | Family-level abundances for the four focal families. |
| INV_OPEN_DATA_SITE | Site metadata for summaries/maps (area, catchment, BNG coords). |
| INV_OPEN_DATA_TAXA | Determines recording mode per sample (BIN vs COUNT; non-zero, non-NA). |

**Focal families :** We restrict attention to four one-species families (**Aphelocheiridae, Brachycentridae, Odontoceridae, Cordulegastridae**) to minimise aggregation bias and keep ecological interpretation straightforward. Potamanthidae is excluded because it is too rare for stable trend estimation.

**Units of analysis :** The core unit is a **site–date** visit; family-level responses are then joined to this unit to maintain a consistent analysis frame across data types.

## 2.2 Data Cleaning

Data cleaning aligns heterogeneous records into a single, reproducible analysis frame by harmonising methods, dates, and identifiers, so downstream models reflect biological signal rather than artefacts of sampling and recording. Accordingly, we standardise typing in Arrow, restrict to EA's core analysis methods and S3PO, apply a ≥1990 date floor, and de-duplicate same-day replicates to obtain one unique, auditable visit per site–date.

### 2.2.1 Visit roster (`metrics_clean`)

We retain only `ANAA`, `ANLA`, and `ANLE` because other analysis codes are not comparable for this task and would introduce uncontrolled heterogeneity. Arrow typing is enforced before `collect()` so that integer keys and dates are consistent across joins and so that subsequent processing is deterministic. The S3PO filter is applied by default to standardise field gear and reduce confounding from sampling

method imbalance. A date floor at 1990 is used because the national network and procedures were standardised around then; earlier data are sparse and unevenly distributed and can bias trends. Same-day replicates are de-duplicated to avoid pseudo-replication; keeping the earliest `SAMPLE_ID` provides a deterministic rule that preserves one observation per site–date.

#### 2.2.2 Family abundances (`whpt_clean`)

Keys and `TOTAL_NUMBER` are cast to integers to align types with the roster. Restricting to `SITE_ID`, `SAMPLE_ID`, `EQ_TAXON_UNIT`, and `TOTAL_NUMBER` keeps the pipeline explicit and prevents unintended use of fields not required for modelling.

#### 2.2.3 Site metadata (`sites_clean`)

Only identification and mapping fields are retained so that summaries or maps can be created without adding extraneous covariates. This keeps the analysis focused and the joins lightweight.

### 2.3 Data Preprocessing

#### 2.3.1 Classifying samples as BIN (categorical) or COUNT (numeric)

We classify recording mode at the *sample* level using TAXA, inspecting all taxa within each `SAMPLE_ID` because this reveals how the laboratory actually recorded abundances. Non-zero, non-NA values are examined: if all positive entries are AB placeholders ($\{1, 3, 33, 333, 3333, 33333\}$), the sample is treated as categorical; if both placeholders and other numbers occur, we conservatively treat the sample as categorical; otherwise it is treated as numeric. Zeros and NAs are ignored for classification because they provide no positive evidence of recording mode. We then apply analysis-method overrides: `ANLE` is treated as categorical, `ANAA` as numeric, and `ANLA` follows the TAXA-based decision, defaulting to categorical when ambiguous or when TAXA is missing for that sample. Finally, any residual unknowns after joining back to the roster default to categorical. This conservative policy avoids overstating precision, accommodates regional and temporal method variation, and encodes known EA practices about categorical versus numeric recording.

#### 2.3.2 Season coding and off-season policy

We focus on spring (Mar–May) and autumn (Sep–Nov) because most sampling occurs in these windows and because phenology differs by season. Off-season data are dropped by default to reduce imbalance and avoid extrapolating from sparse winter/summer coverage. Encoding season as a factor at preprocessing ensures that modelling can estimate separate seasonal trajectories without ad hoc later manipulation.

#### 2.3.3 Constructing per-family modelling tables

Four aligned tables are built for each focal family to triangulate trends across complementary likelihoods while preserving the same roster and covariates.

**Presence/absence (`pa_tbl`).** We limit the roster to sites where the family has ever been observed so that occurrence models are not diluted by sites with no plausible presence. Zeros are then re-injected within those sites to obtain a realistic absence–presence structure. Adding `decimal_date` facilitates smooths, and a site factor enables a random-effect structure at fitting time.

**Ordered categorical (`ord_tbl`).** Categorical samples are mapped to ordinal classes (`AB0`, `AB1`, `AB2`, `AB3`, `AB4`). Rows that do not map cleanly to these classes are dropped, and the highest class is collapsed into `AB3p` to mitigate sparsity at the top end, stabilising ordinal thresholds while preserving ecological signal in the common classes. The ordered factor levels are set to `c("AB0","AB1","AB2","AB3p")`.

**Censored Poisson (`cpois_tbl`).** Numeric samples are converted to intervals that reflect one significant–figure practice: exact counts 0–9 are represented as tight point–intervals ($\pm 0.5$), while values $\geq 10$ are represented as decade blocks with non–integer bounds (e.g., $17 \rightarrow [9.5, 19.5]$, $20 \rightarrow [19.5, 29.5]$). Using non–integer limits avoids ambiguity in the censored–count likelihood and encodes measurement uncertainty consistently with EA rounding.

**Censored normal (`cnorm_tbl`).** The same intervals are reused with a variance–stabilising transform (e.g., square–root/$p$–power) so that a censored–normal likelihood can handle over–dispersion induced by subsampling and ecological variability. This provides a defensible alternative to Poisson–based models while retaining the same information about interval bounds.

### 2.4 Quality checks and design trade-offs

- **Sanity checks.** Cross-tabulating `ANALYSIS_METHOD` against data type confirms that overrides behave as intended; inspecting the spring/autumn balance verifies that the season filter has not introduced artefacts; and verifying de-duplication ensures that each site–date is unique.
- **Conservative classification.** Treating ambiguous `ANLA` samples and rare mixed cases as categorical avoids injecting spurious precision and aligns with guidance on method heterogeneity.
- **Season focus.** Concentrating on spring and autumn improves comparability of seasonal trends and stabilises uncertainty; richer month-wise structures are possible but are not required for surveillance aims.
- **Sparse top categories.** Collapsing `AB4` into `AB3p` addresses sparsity without discarding signal, improving ordinal fit stability.
- **Population-level interpretation.** When reporting national trends, we marginalise over site random effects at prediction to provide curves that reflect the population-level signal rather than site-specific deviations.

The pipeline yields (i) a clean, deduplicated visit roster (`metrics_clean`); (ii) typed family abundances (`whpt_clean`); (iii) site metadata (`sites_clean`); (iv) an annotated roster with data-type and season flags (`roster_flagged`); and (v) four aligned modelling tables per family for presence/absence, ordered categories, censored Poisson, and censored normal analyses. This foundation is reproducible, scientifically defensible, and tailored to the EA data's mixed recording modes—so the models can focus on *signal* rather than artefacts of measurement.

# 3 Exploratory Data Analysis (EDA)

EDA had two aims: (i) characterise how the EA monitoring programme has changed through time and space, and (ii) check that the recording rules used in preprocessing match the empirical fingerprints in the data.

Sampling effort has declined from an early peak and the mix of analysis methods has shifted markedly (Figure 3). ANLE dominates in the 1990s, ANLA increases around 2000, and ANAA appears intermittently thereafter. This validates treating method codes as informative but imperfect labels and motivates our rule-based BIN/COUNT classification using the TAXA table rather than method code alone.

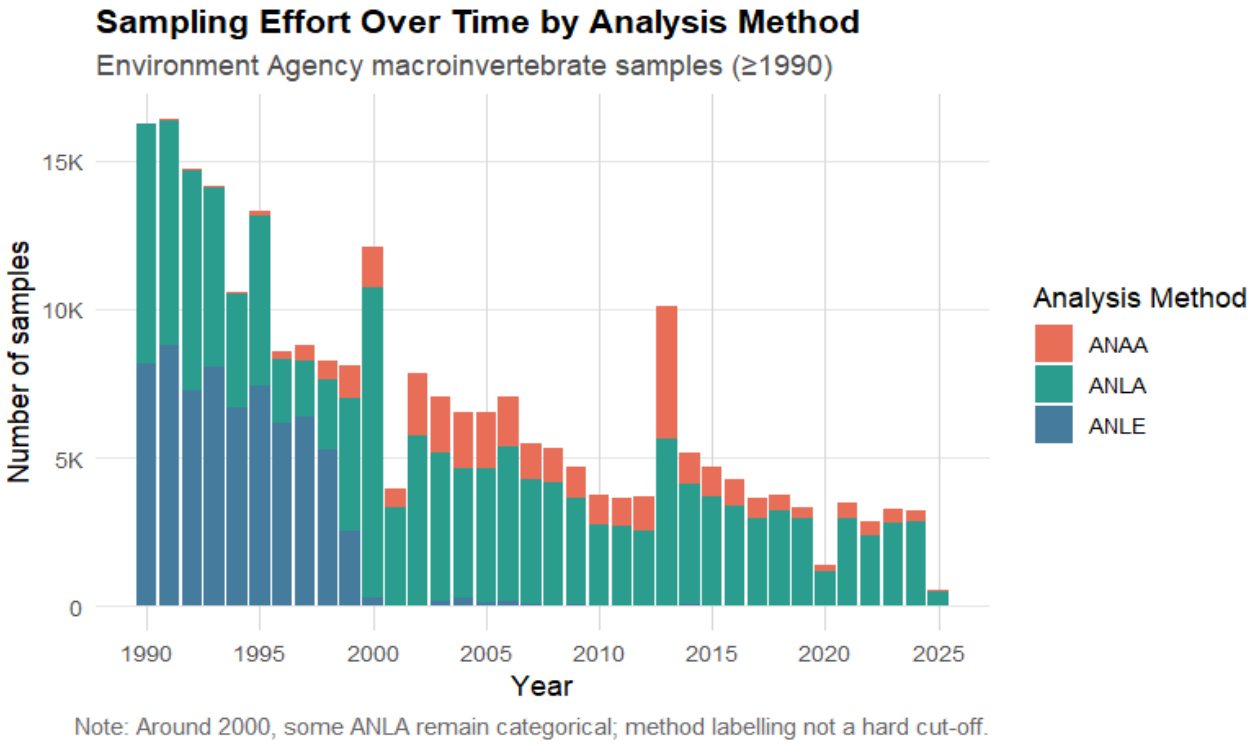



Figure 3: Sampling effort over time by analysis method (samples ≥1990).

**Recording mode and rounding :** The proportion of categorical ("bin") versus numeric ("count") samples varies by family and year (Figure 4). Mixed cases occur—especially around the 2000 transition—and are conservatively treated as categorical. For samples tagged numeric, last-digit distributions for values ≥10 show strong spikes at 0 (and often 5), while single-digit counts are spread across 1–9 (Figure 5). This is the classic signature of one-significant-figure recording and confirms the need to represent numeric observations as intervals on the count scale (e.g., 20 → 19.5–29.5) rather than as exact values. These fingerprints are stable across families, strengthening the rationale for an interval-censored likelihood.

**Categorical structure :** For categorical records, AB0 (absence) dominates, with AB1 and AB2 contributing most of the non-zero mass and very few observations in the highest class (Figure 6). Because AB4 is extremely sparse, we merge AB4 into AB3 (reported as AB3+) to stabilise threshold estimation without losing ecological signal in the common classes. This decision is borne out across families: the zoomed view shows AB3+ typically at ∼1% or less of all samples, whereas AB1–AB2 carry the informative variation.

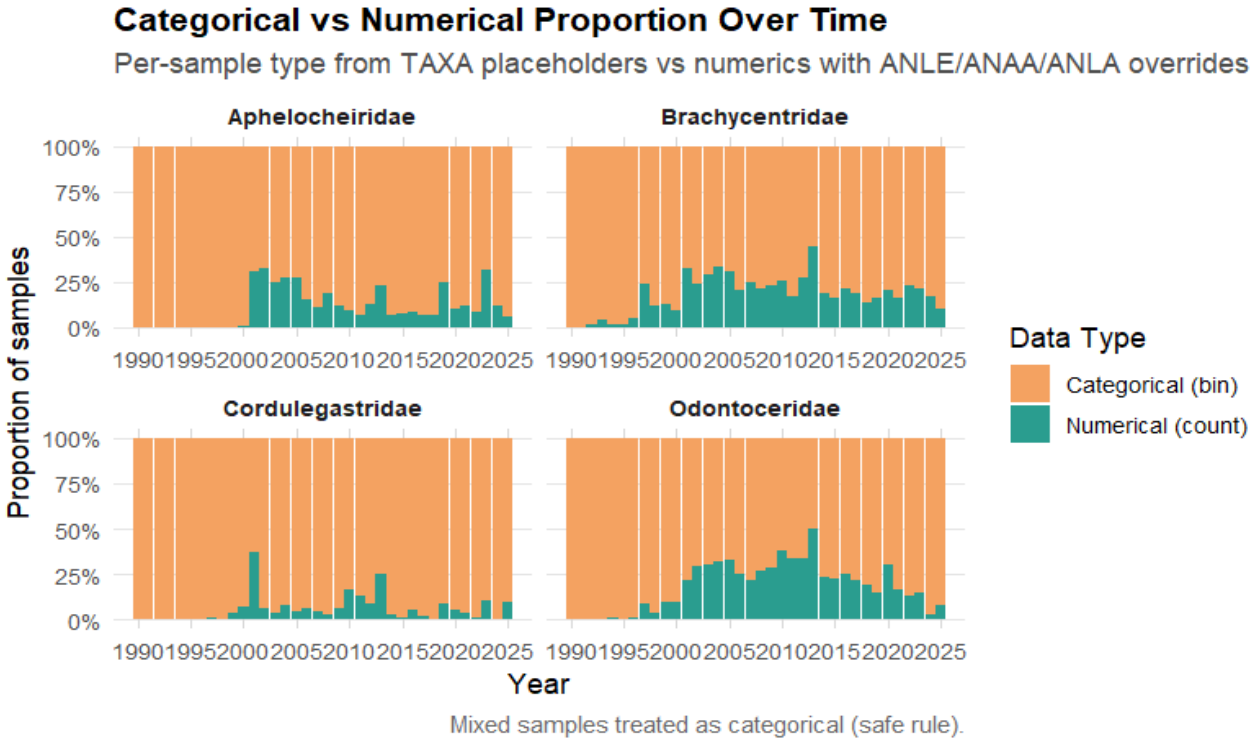



Figure 4: Categorical vs. numeric proportion over time by family (BIN/COUNT classification).

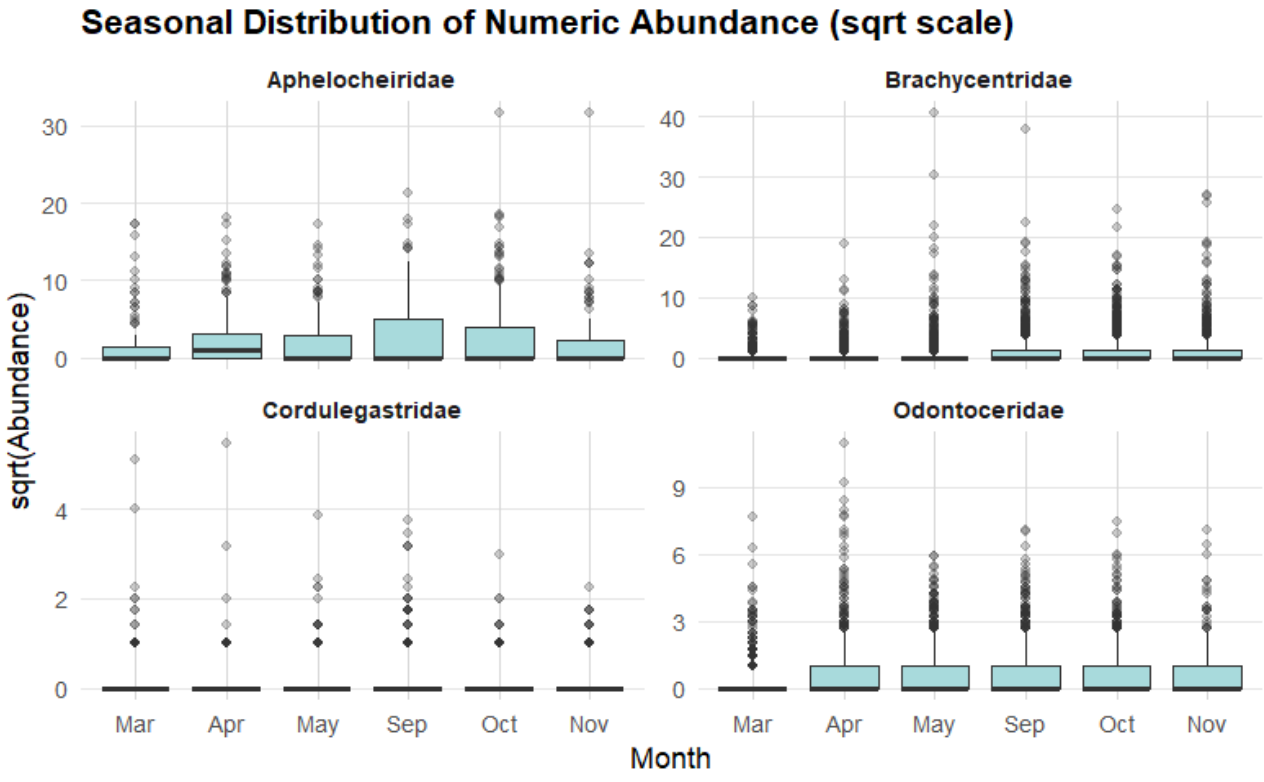
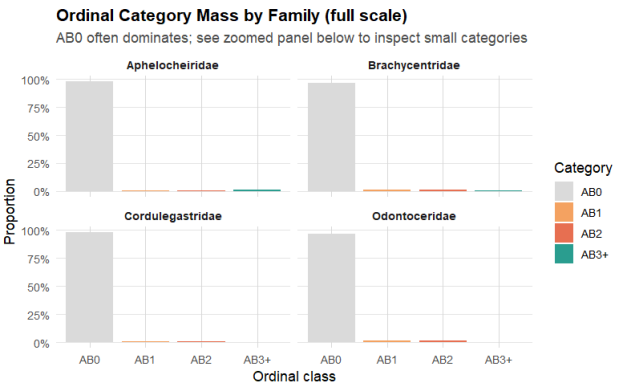
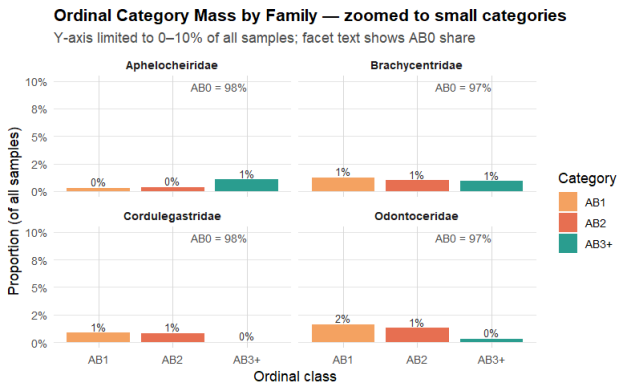



Figure 5: Rounding fingerprint for numeric counts (distribution of last digit for totals $\geq$10).




(a) Full scale.




(b) Zoomed to small categories (AB1–AB3+).

Figure 6: Ordinal category mass by family: AB0 dominates; AB4 is extremely sparse (merged into AB3+).

**Seasonality :** Sampling is concentrated in spring (Mar–May) and autumn (Sep–Nov) across all methods (Figure 7). This justifies focusing the analysis on those seasons and modelling season-varying smooths. Presence rates plotted separately by season reveal systematic differences: for several families autumn presence is higher and less volatile than spring, but the gap is not constant through time (Figure 8). These features argue against a single, season-agnostic trend and in favour of allowing the temporal smooth to vary by season.

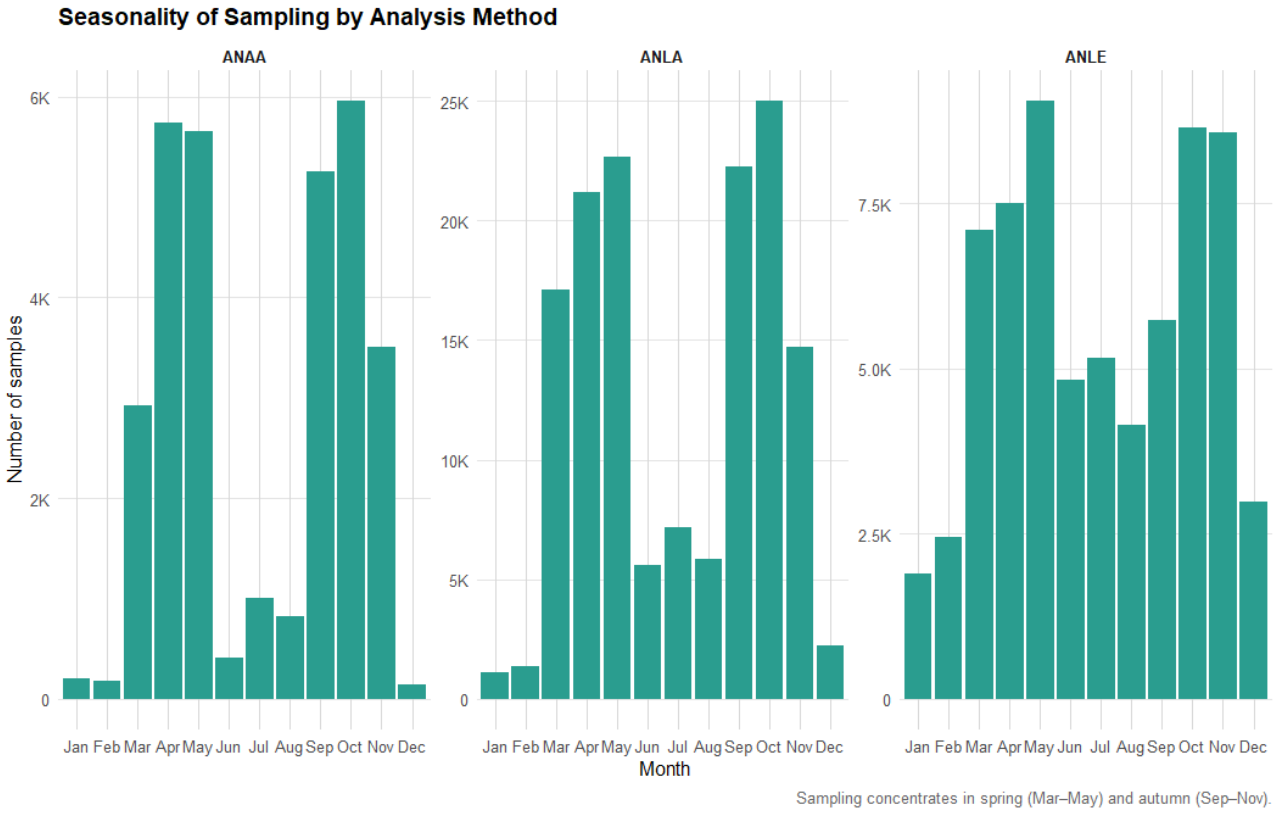



Figure 7: Seasonality of sampling by analysis method (counts by month).

**Spatial coverage and site effort :** A map of one focal family illustrates heterogeneous spatial effort and detection (Figure 9). Some catchments have many revisits and high presence rates, while others are sparsely sampled or dominated by non-detections. This heterogeneity supports including a site random effect to capture persistent local differences and reporting national, population-level trends by marginalising over sites at prediction time.

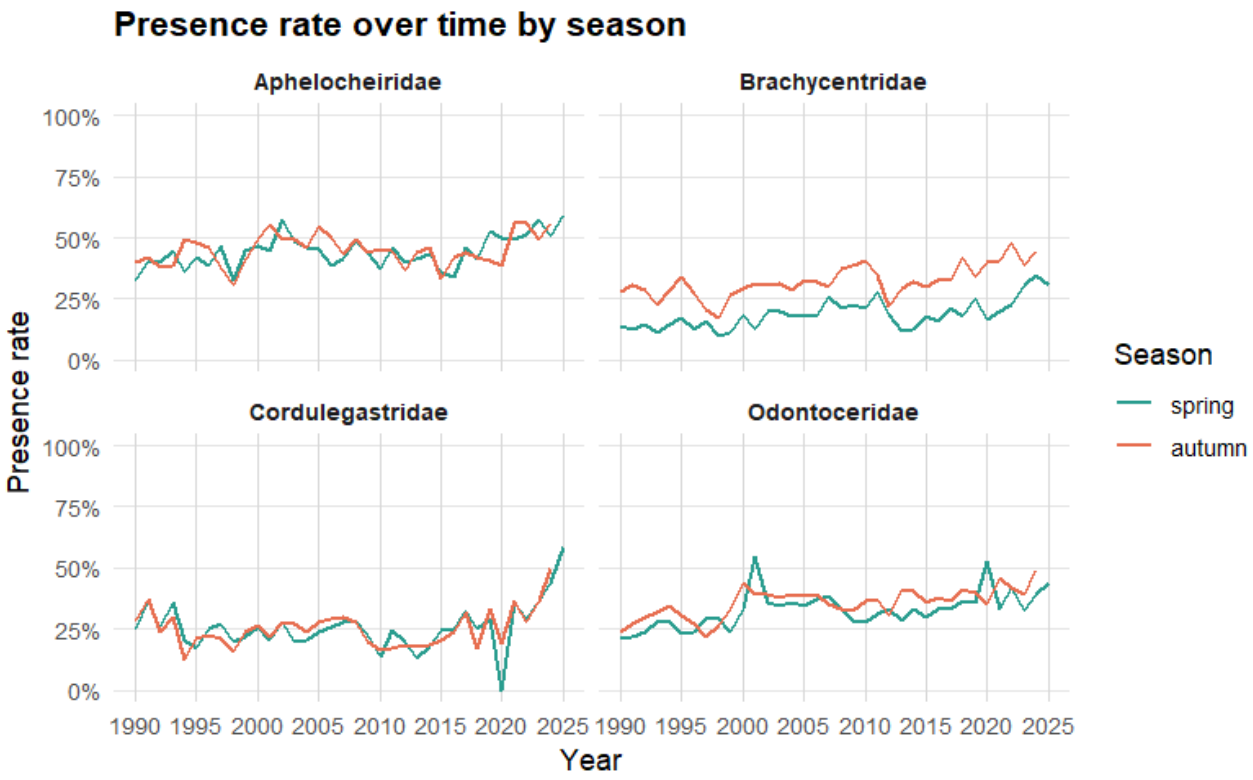



Figure 8: Presence rate over time by season for each family (zeros re-injected within the family's site roster).

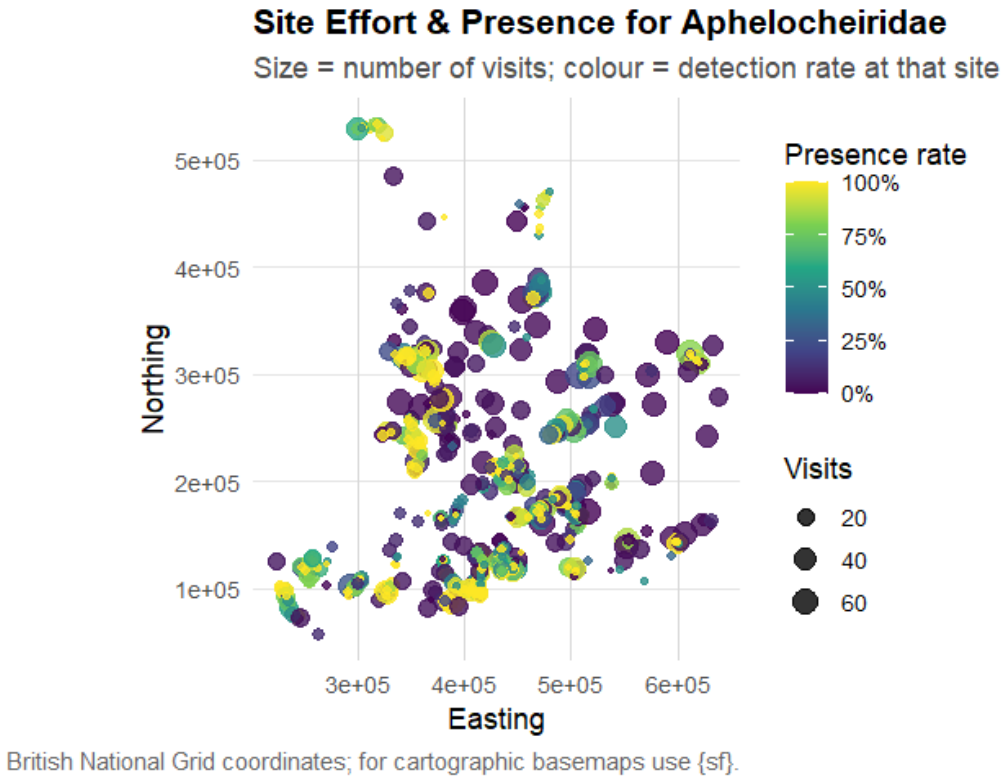



Figure 9: Site effort and presence for Aphelocheiridae (size = visits; colour = detection rate).

**Preliminary trend signals :** Simple presence-rate trajectories (zeros re-injected within each family's site roster) show broadly increasing occurrence over the long run for all families, with family-specific plateaus or dips (Figure 8). The pattern is consistent with known improvements in water quality in the 1990s, but also reflects changes in recording practice; hence our formal models rely on triangulation (presence/absence, ordered categorical, and interval-censored abundance) rather than on any single exploratory statistic.

Further exploratory figures—including presence-rate trajectories, numeric-only medians, rounding fingerprints, seasonal zero proportions, regional data-type mixes, and replicate-day counts—are provided in Appendix C.

### Key Takeaways from EDA

EDA establishes that (i) recording mode is mixed and time-varying, (ii) numeric counts carry clear 1-s.f. rounding, (iii) categorical classes are highly imbalanced with sparse upper bins, (iv) sampling and biology are seasonal, and (v) effort and detection vary strongly by site. These findings directly motivate our preprocessing (BIN/COUNT flags, interval bounds, AB3+ merge, spring/autumn focus) and our modelling choices (season-varying smooths, site random effects, and triangulation across likelihoods).

Further exploratory figures—including presence-rate trajectories, numeric-only medians, rounding fingerprints, seasonal zero proportions, regional data-type mixes, and replicate-day counts—are provided in Appendix C.

# 4 Model methodology

Our modelling pipeline is designed to (i) respect the EA recording system (categorical AB classes vs one-significant-figure counts), (ii) capture seasonal dynamics, and (iii) report **population-level** national trends by family. We therefore build four aligned models per focal family—using the *same* roster, covariates, seasons, and site structure—and compare them for triangulation.

### Common structure and notation

Let $i = 1, \ldots, n$ index site–date samples after preprocessing, with:

- $t_i$: decimal date,
- $s_i \in \{\text{spring}, \text{autumn}\}$: season factor,
- $j(i)$: site id for sample $i$,
- $b_j$: site random intercept,
- $f_{\text{sp}}(t), f_{\text{au}}(t)$: season-specific smooths of time.

All models share the same additive predictor

$$\mu_i \;=\; \beta_0 \;+\; \beta_{\text{season}}(s_i) \;+\; \underbrace{f_{\text{sp}}(t_i)\mathbb{1}[s_i = \text{spring}] + f_{\text{au}}(t_i)\mathbb{1}[s_i = \text{autumn}]}_{\text{season-varying smooth of time}} \;+\; b_{j(i)},$$

with $b_j \sim \mathcal{N}(0, \sigma^2_{\text{site}})$. The smooths are thin-plate regression splines with basis size $k$ chosen large enough to avoid over-smoothing (checked via the $k$-index), and smoothing penalties selected by **REML**. We fit with **mgcv::bam** (large-data optimisations enabled). For reporting *national* trends, we set $b_j = 0$ at prediction, i.e., we marginalise over sites to obtain population-level curves (8; 9).

Below, $Y_i$ denotes the family-level response after aligning to the visit roster, with zeros re-injected where the family was not recorded at sites where it has ever been observed.

## 4.1 Presence/absence (binomial GAMM)

We first ask whether occurrence is changing through time. Define $Y_i \in \{0, 1\}$ as presence/absence.

$$Y_i \sim \text{Bernoulli}(p_i), \qquad \text{logit}(p_i) = \mu_i.$$

This model uses the full roster restricted to sites where the family has been observed at least once (thereby avoiding dilution by implausible sites) and includes the season-varying smooth plus a site random effect. The PA fit provides an interpretable baseline: trends in occupancy independent of abundance conditional on presence.

**Diagnostics and use :** We inspect deviance residual QQ/ACF, the $k$-index for smooth adequacy, and concurvity. The PA trend gives a first indication of recovery/decline and is later compared with abundance-based fits (8; 9).

## 4.2 Ordered categorical abundance (OCAT GAMM)

For samples classified as **categorical** (data_type = "bin"), we map WHPT placeholders to classes AB0, AB1, AB2, AB3p (with **AB4 merged into AB3** to mitigate sparsity). Let $Y_i \in \{0, 1, 2, 3\}$ denote these ordered levels.

We fit an ordered-categorical model (mgcv's **ocat** family) in threshold form:

$$\Pr(Y_i \le k) \;=\; g^{-1}\big(\alpha_k - \mu_i\big), \qquad k = 0, 1, 2, 3,$$

where $g$ is the cumulative link (default logistic), $\alpha_k$ are ordered cut-points ($\alpha_0 < \alpha_1 < \alpha_2 < \alpha_3$), and $\mu_i$ is the same season-varying additive predictor as above. Category probabilities follow by differencing consecutive cumulative probabilities. The linear predictor is on the latent *abundance propensity* scale (8; 9).

**Why this model :** It uses all categorical information without inventing pseudo-numeric scores, and by merging the rarest top class we stabilise threshold estimation while retaining signal in AB1–AB2.

### 4.3 Interval-censored Poisson abundance (cPois GAMM)

For **numeric** samples (data_type = "count"), EA practice implies one-significant-figure uncertainty. We therefore convert each observed count to a **real-valued interval** on the **count** scale:

$$y \in \{0, \ldots, 9\} \;\mapsto\; (y - 0.5,\, y + 0.5], \qquad y \geq 10 \;\mapsto\; (10m - 0.5,\, 10m + 9.5] \;\text{ for }\; m = \lfloor y/10 \rfloor.$$

Let $(L_i, U_i]$ denote these bounds. With $Y_i \sim \text{Poisson}(\mu_i^{(\text{pois})})$ and $\log \mu_i^{(\text{pois})} = \mu_i$, the censored likelihood contribution is

$$\Pr\left(L_i < Y_i \leq U_i \,\middle|\, \mu_i^{(\text{pois})}\right) \;=\; F_{\text{Pois}}\big(\lfloor U_i \rfloor;\, \mu_i^{(\text{pois})}\big) \;-\; F_{\text{Pois}}\big(\lfloor L_i \rfloor;\, \mu_i^{(\text{pois})}\big).$$

We fit this with **mgcv::bam(..., family = cpois())**, using the same season-varying smooth and site random effect (9).

**Role in triangulation :** The count-scale censoring respects measurement rules, but pure Poisson can show over-dispersion (from sub-sampling and ecological heterogeneity). We therefore use cPois for comparison and then move to a variance-stabilised censored normal.

### 4.4 Variance-stabilised censored normal abundance (cNorm GAM)

To address over-dispersion while keeping interval information, we transform counts using a ***p*-power variance-stabilising transform**, $g(x) = x^p$ (square-root is the special case $p = 1/2$). We apply $g$ to the **interval bounds**:

$$Z_i \in [g(L_i),\, g(U_i)],$$

and assume

$$Z_i \sim \mathcal{N}(\mu_i^{(\text{norm})},\, \sigma^2), \qquad \mu_i^{(\text{norm})} = \mu_i.$$

The likelihood contribution is

$$\Pr\big(g(L_i) \leq Z_i \leq g(U_i)\big) = \Phi\Big(\tfrac{g(U_i) - \mu_i}{\sigma}\Big) - \Phi\Big(\tfrac{g(L_i) - \mu_i}{\sigma}\Big).$$

We fit with **mgcv::bam(..., family = cnorm())**, estimating $\sigma$ and the smoothing penalties by REML. A short grid search over $p$ (e.g., $p \in \{0.30, 0.35, 0.40\}$) selects the value that yields the flattest residual spread versus fitted values/time; $p \approx 0.35$ typically worked well. This **seasonal cNorm GAM** is our **primary** model, because it (i) encodes interval uncertainty, (ii) reduces over-dispersion relative to cPois, and (iii) provides smooth, interpretable population-level trends (8; 9).

Since $g(x) = x^p$ is monotone increasing on $x \geq 0$, applying $g$ to $(L_i, U_i]$ preserves the censoring order (including cases with $L_i = 0$).

### 4.5 Single-trend vs season-varying smooth

To verify that separate spring/autumn trajectories are warranted, we compare:

$$\textit{Single-trend:} \quad \mu_i = \beta_0 + \beta_{\text{season}}(s_i) + f(t_i) + b_{j(i)}.$$

$$\textit{Season-varying:} \quad \mu_i = \beta_0 + \beta_{\text{season}}(s_i) + f_{\text{sp}}(t_i)\mathbb{1}[s_i = \text{sp}] + f_{\text{au}}(t_i)\mathbb{1}[s_i = \text{au}] + b_{j(i)}.$$

We retain the season-varying model where it offers materially better fit (lower REML/AIC, higher explained deviance) and cleaner residual diagnostics. Given phenology and EDA evidence, season-varying smooths are usually preferred (8). Both seasonal smooths use the same basis size $k$ for comparability.

### 4.6 Diagnostics and robustness

Across models we check:

1. **Residuals.** QQ plots for deviance or working residuals; for censored models we compute **randomised quantile residuals (RQRs)** to assess distributional adequacy under censoring. Residuals are also stratified by data type (exact 0–9, 1-s.f. $\geq 10$, categorical) to avoid confounding changes in residual patterns with changes in recording mode over time (1; 6).

2. **Smoothing adequacy.** $k$-index (from `gam.check`) and effective degrees of freedom vs upper bound; increase $k$ only if there is evidence of over-smoothing (8).

3. **Temporal dependence.** Residual ACF; if material autocorrelation were present we would consider adding a correlation structure, but fits were acceptable after season-varying smooths.

4. **Model comparison.** Explained deviance and REML/AIC across PA, OCAT, cPois and cNorm. Agreement in trend direction and shape increases confidence that conclusions are not artefacts of any single likelihood (9).

### 4.7 What each model answers

| Model | Question answered |
|---|---|
| **PA (binomial)** | Is the *probability of occurrence* changing, by season? |
| **OCAT** | Within categorical samples, are *abundance classes* shifting over time? |
| **cPois** | Given count-scale intervals, what is the *mean count* trend assuming Poisson errors? |
| **cNorm** | Given interval uncertainty and variance-stabilisation, what is the *population-level abundance* trend with stable dispersion and seasonal dynamics? |

Table 2: What each model answers.

By fitting the four models on the same roster with the same seasonal/site structure, we obtain a coherent triangulation. The **seasonal cNorm GAM** is used for the headline national trends, while PA and OCAT provide complementary evidence that is robust to how the data were recorded, and cPois documents any sensitivity to the Poisson assumption.

**Formula summary (Seasonal Models).**

$$\begin{aligned}
\text{PA:} \quad & Y_i \sim \text{Bern}(p_i), && \text{logit}(p_i) = \mu_i. \\
\text{OCAT:} \quad & \Pr(Y_i \leq k) && = g^{-1}(\alpha_k - \mu_i), \;\; k = 0, 1, 2, 3. \\
\text{cPois:} \quad & \Pr(L_i \leq Y_i \leq U_i) && = F_{\text{Pois}}(U_i;\, e^{\mu_i}) - F_{\text{Pois}}(L_i - 1;\, e^{\mu_i}). \\
\text{cNorm:} \quad & \Pr(g(L_i) \leq Z_i \leq g(U_i)) && = \Phi\Big(\tfrac{g(U_i)-\mu_i}{\sigma}\Big) - \Phi\Big(\tfrac{g(L_i)-\mu_i}{\sigma}\Big),
\end{aligned}$$

with $\mu_i = \beta_0 + \beta_{\text{season}}(s_i) + f_{\text{sp}}(t_i)\mathbb{1}[s_i = \text{sp}] + f_{\text{au}}(t_i)\mathbb{1}[s_i = \text{au}] + b_{j(i)}$, and $b_j \sim \mathcal{N}(0, \sigma^2_{\text{site}})$. This framework directly operationalises our preprocessing choices (BIN/COUNT flags, interval bounds, AB3+ merge, spring/autumn focus) and yields seasonal, site-marginal national trends that are scientifically defensible and reproducible (8; 9).

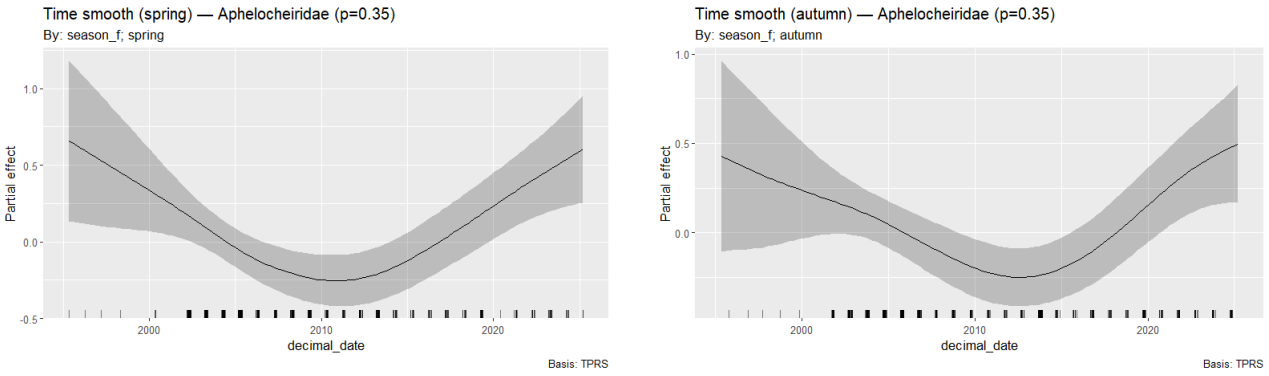
# 5 Results

We report national, population-level trends using the seasonal censored-normal (cNorm) GAM as the primary analysis, and summarise diagnostics and fit quality. Trends are evaluated with site random effects set to zero to provide site-marginal curves that are comparable across families and seasons. Presence/absence, ordered-categorical and censored-Poisson fits were used for triangulation; their qualitative agreement with cNorm increases confidence that conclusions are not artefacts of any single likelihood.

## 5.1 Primary seasonal trends (cNorm)

**Aphelocheiridae, cNorm** $p = 0.35$ **:** The seasonal censored-normal GAM explains **72.7%** of deviance (**AIC** $= 5164.8$; $n = 1089$), with near-ideal dispersion (SD(RQR) $\approx 1.00$). Autumn shows a higher mean level than spring ($\hat{\beta}_{\text{autumn}} > 0$). The spring smooth is relatively simple whereas autumn exhibits richer temporal structure; the site effect is substantial, indicating marked spatial heterogeneity. The figures below present the season-specific smooths and site-marginal national trends with 95% intervals, followed by residual diagnostics for model adequacy.

For illustration we show the full set of cNorm outputs for **Aphelocheiridae**. The model estimates separate smooths for spring and autumn, allowing distinct seasonal trajectories while sharing the same site structure and roster.




(a) Spring time smooth (partial effect).

(b) Autumn time smooth (partial effect).

Figure 10: Season-varying smooths for Aphelocheiridae (cNorm, $p = 0.35$). Rugs show sample-time density; ribbons are 95% intervals.

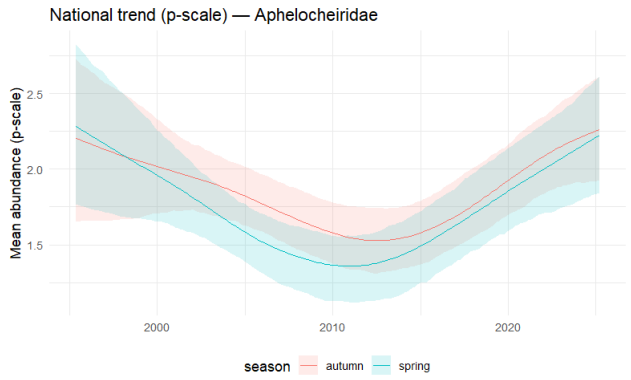
The seasonal smooths (Fig. 10) show a broadly U-shaped pattern from the 1990s to early 2010s followed by recovery, with slightly stronger autumn amplitude. Translating partial effects into site-marginal trends yields the national trajectories below.

The season-specific curves (Fig. 11a) and their season-average (Fig. 11b) indicate declining abundance through the early 2010s and subsequent increase into the 2020s, consistent with EDA patterns and with the notion that water quality improvements are reflected in macroinvertebrate abundance signals. Confidence bands widen where sampling density was lower (see rugs in Fig. 10).

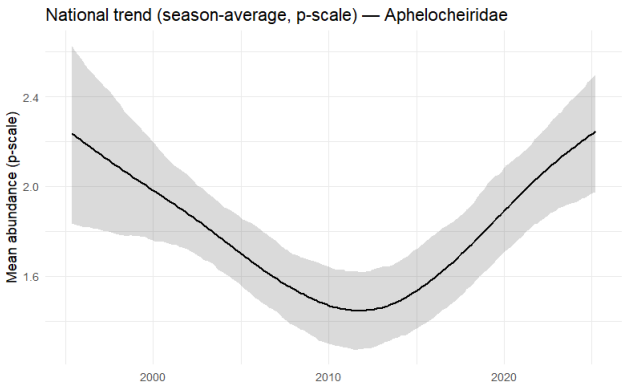
## 5.2 Model diagnostics (Aphelocheiridae)

We assess distributional adequacy using randomised quantile residuals (RQRs), along with QQ, residual-vs-fitted, histogram, and ACF checks. The panels below show broadly acceptable behaviour, with mild tail deviations but no concerning autocorrelation after accounting for season-varying smooths.

Overall, the residual spread is approximately constant on the transform scale; slight upper-tail inflation is expected given occasional large counts and heterogeneous effort, and motivates our use of

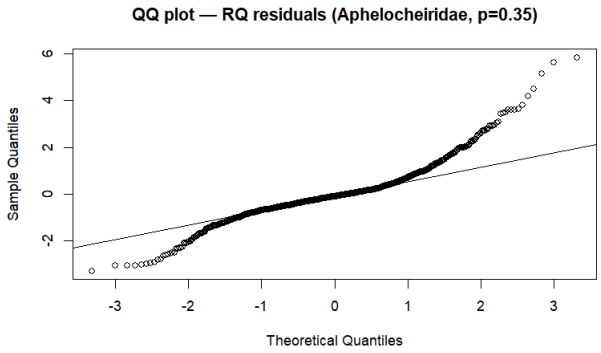



(a) Seasonal trends (site-marginal).

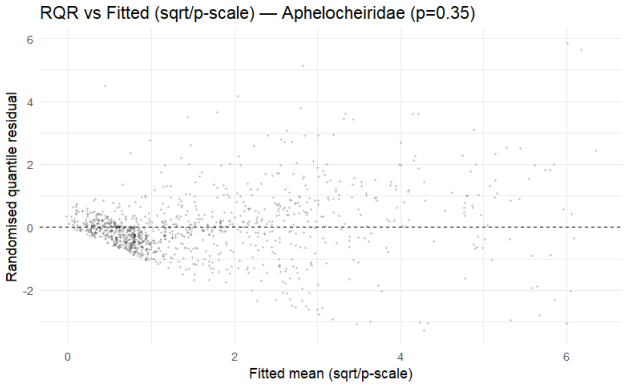



(b) Season-average trend (site-marginal).

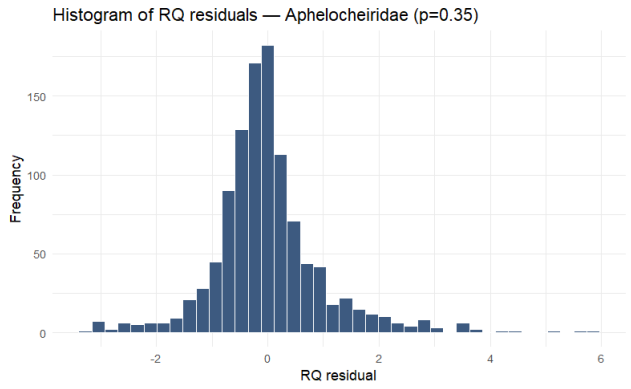
Figure 11: Aphelocheiridae national trends on the $p$-scale (cNorm, $p = 0.35$). Shaded ribbons are 95% pointwise intervals.



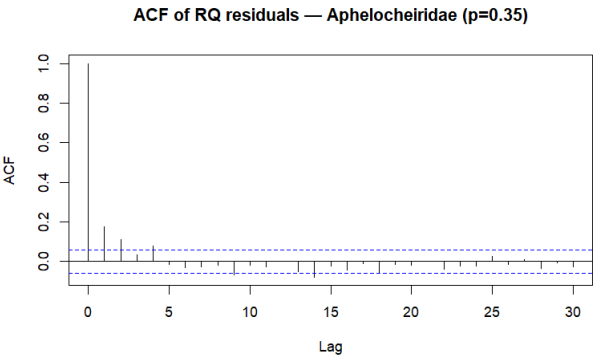

(a) QQ plot of RQRs.




(b) RQR vs fitted (on $p$-scale).




(c) Histogram of RQRs.




(d) ACF of RQRs.

Figure 12: Aphelocheiridae diagnostics for the seasonal cNorm GAM ($p = 0.35$).

the variance-stabilised censored-normal likelihood rather than a pure Poisson alternative. Diagnostics are acceptable overall (Fig. 12a–d): the QQ plot is close to the 1–1 line, residuals vs. fitted show no strong structure, the histogram is centred, and the ACF shows no material autocorrelation.

### 5.3 Fit summary across families

Table 3 summarises key fit statistics for the seasonal cNorm model for all four families. The $p$ values were chosen by a small grid search to stabilise residuals; effective degrees of freedom ($k_{\text{time}}$) were kept below their basis limits; the standard deviation of RQRs is close to 1, indicating reasonable dispersion on the transform scale; and explained deviance is high for Aphelocheiridae and Cordulegastridae and moderate for the other two families.

| Family | $p$ | $n$ | $k_{\text{time}}$ | SD(RQR) | Dev. Expl. (%) | AIC |
|---|---|---|---|---|---|---|
| Aphelocheiridae | 0.35 | 1089 | 25 | 1.00 | 72.7 | 5164.8 |
| Brachycentridae | 0.35 | 5917 | 29 | 0.97 | 49.8 | 21869.1 |
| Odontoceridae | 0.40 | 6674 | 28 | 1.00 | 56.2 | 17987.6 |
| Cordulegastridae | 0.40 | 1064 | 25 | 0.98 | 72.1 | 1552.7 |

Table 3: Seasonal cNorm GAM summary across focal families. SD(RQR) near 1 indicates adequate dispersion on the transform scale; deviance explained is computed on the working scale.

Figure 13 summarises the site–marginal seasonal trends from the cNorm model for all four families, displayed on the back–transformed count scale for readability. Shaded ribbons are 95% pointwise intervals and vertical dotted lines mark EA QC milestones (1990, 1995, ˜2000). The panel corroborates the family–specific results: a decline into the early 2010s followed by recovery, with autumn typically higher than spring (see attached image `Trend_Plot.png`).

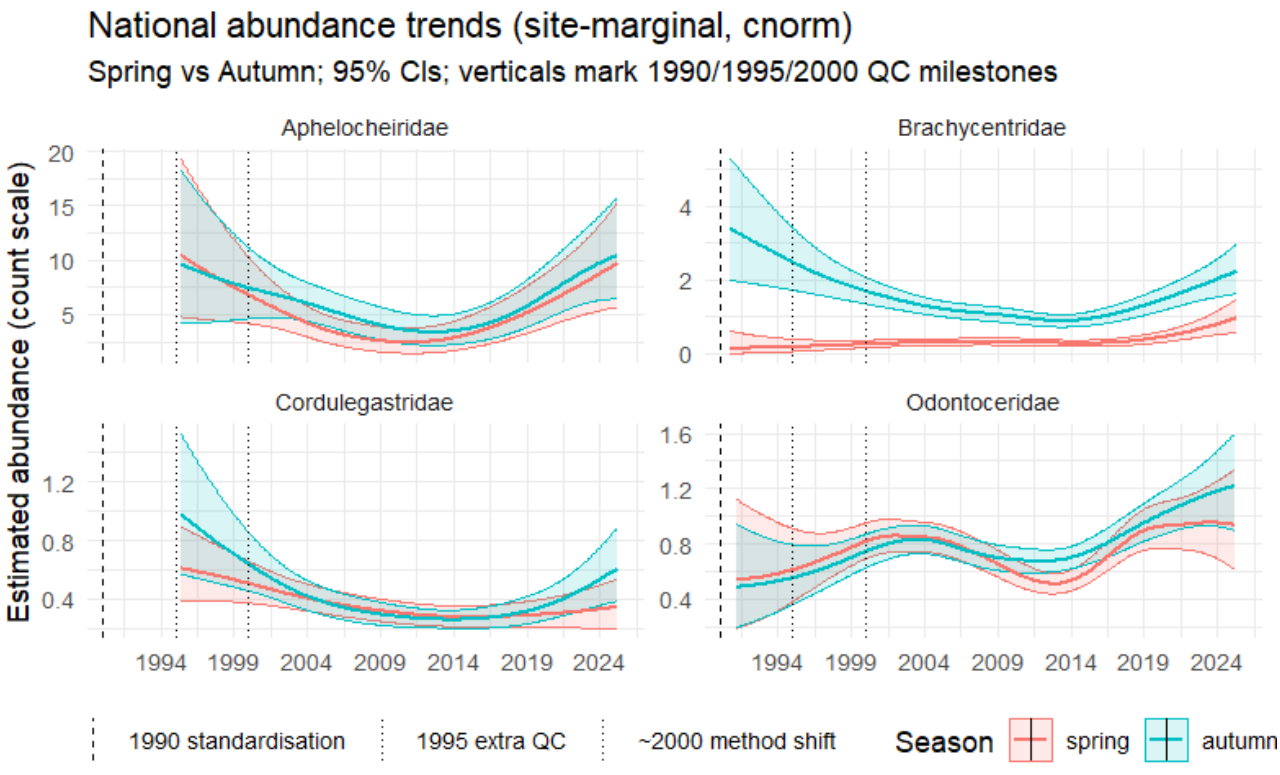



Figure 13: National, site–marginal seasonal trends from the cNorm model across the four focal families (back–transformed count scale). Shaded ribbons show 95% pointwise intervals; vertical dotted lines indicate EA QC/standardisation milestones (1990, 1995, ˜2000).

*Complete cNorm outputs (seasonal smooths, national curves, and RQR-based diagnostics) for the remaining families—Brachycentridae, Odontoceridae, and Cordulegastridae—are provided in Appendix D.*

### 5.4 Triangulation and interpretation

Checks against the binomial PA and OCAT models show consistent directional change by season: increasing occurrence and upward shifts in categorical classes align with the cNorm trend. The censored-Poisson fits capture similar multi-decadal shapes but tend to show heavier-tailed residuals, reinforcing the choice of a variance-stabilised censored-normal likelihood for headline inference. Taken together, the four models provide a coherent picture: a decline into the early 2010s followed by an increase through the 2020s, with modest seasonality in amplitude. These population-level curves are therefore suitable for surveillance and for comparing trajectories across families. **Per-family outputs for the PA (binomial), OCAT, and cPois models are provided in Appendix E, where the PA and OCAT summary tables appear immediately after each model's diagnostic figures.**

### 5.5 cNorm variants and Model Selection

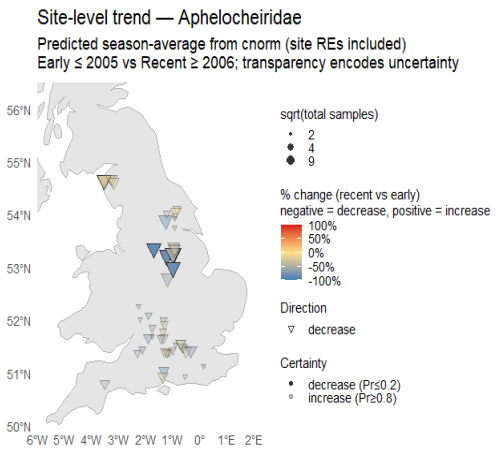
We compared censored-normal specifications that differ only in the variance–stabilising exponent $p \in \{0.30, 0.35, 0.40\}$ and whether the time smooth is *single* or *season-varying*, under the same roster, censoring rules, and site random effects. Selection prioritised well-calibrated dispersion (SD of randomised quantile residuals $\approx 1$), improved QQ/ACF behaviour, and competitive explained deviance/AIC. Consistent with phenology and EDA, season-varying smooths were retained where they materially improved fit and diagnostics. Family-specific $p$, diagnostics, and a worked comparison are provided in Appendix F. National curves are reported as site-marginal predictions ($b_j = 0$).

# 6 Discussion

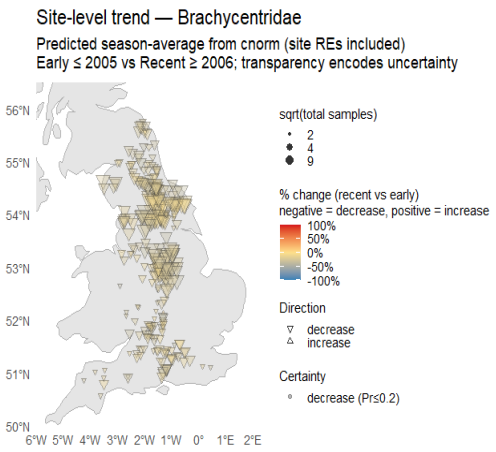
## 6.1 Interpreting the national trends

Across families, the seasonal censored–normal (cNorm) GAM yields coherent, site–marginal national trajectories (Section 5). For *Aphelocheiridae* (our worked example), the spring and autumn smooths show a decline into the early 2010s followed by recovery through the 2020s, with autumn levels consistently higher than spring. The shape is mirrored qualitatively in the PA and OCAT fits, and the cPois model recovers similar multi–decadal structure despite heavier–tailed residuals. Confidence bands widen where sampling density is lower (see rugs in Fig. 10), and the main features are robust to likelihood choice and to modest changes in the variance–stabilising transform. These results are *consistent* with the expectation that macroinvertebrate abundance responds positively to improvements in water quality, although they do not in themselves establish causation.

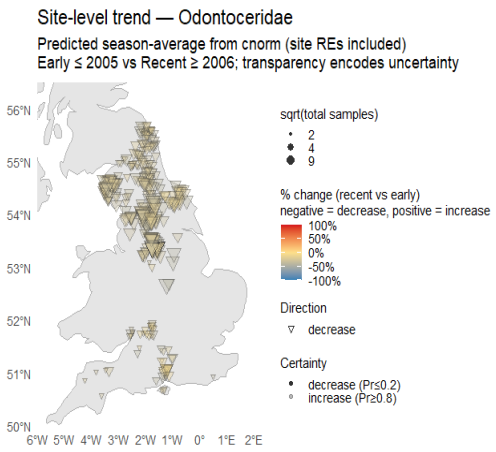
**Spatial Context (Site-Level Change) :** To complement the national trends, Fig. 14 maps the *site-level* percent change (recent $\geq$ 2006 vs. early $\leq$ 2005) in season-average cNorm abundance for each family. Symbols encode direction and magnitude; transparency reflects uncertainty. Broad spatial coherence is evident, with pockets of increase and decrease varying by family.



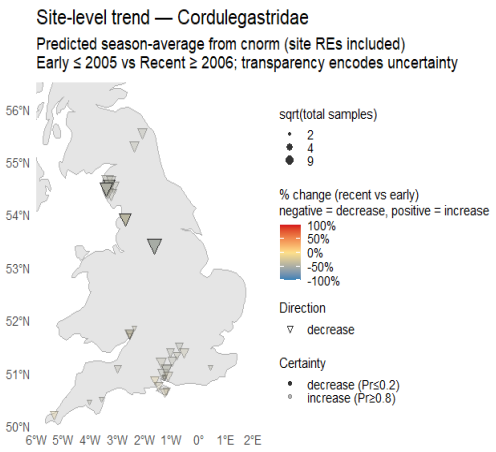

(a) Aphelocheiridae




(b) Brachycentridae




(c) Odontoceridae




(d) Cordulegastridae

Figure 14: Site-level trend maps: percent change in season-average cNorm abundance (recent vs. early). Direction (triangle up/down), magnitude (fill), uncertainty (alpha), and sampling intensity (point size).

### 6.2 Strengths of the approach

Three design choices underpin credibility. First, the data model respects *how* the EA data were recorded: categorical AB classes and one–significant–figure numeric counts are both represented as *intervals*, so the likelihood "knows" about measurement uncertainty. Second, the *season–varying* smooths allow spring and autumn to exhibit different temporal trajectories, avoiding the averaging–away of seasonal signals; diagnostics and AIC comparisons preferred this structure in our use–case. Third, site heterogeneity is handled via random intercepts, while national trends are reported as *population–level* (site–marginal) curves, which are the most relevant for surveillance and comparison across families. The pipeline is reproducible end–to–end: we fix types up front, build a deterministic visit roster, classify samples conservatively (ANLE $\rightarrow$ categorical; ANAA $\rightarrow$ numeric; ANLA resolved from TAXA, defaulting to categorical), and construct aligned modelling tables for triangulation. Together, these choices reduce the risk of spurious precision and ensure that inference targets the population signal rather than local idiosyncrasies.

### 6.3 Limitations and Cautions

(i) **Analysis–method reliability.** ANLA semantics vary historically and regionally; our TAXA–based classification plus conservative fallbacks mitigate, but cannot eliminate, misclassification.

(ii) **Rounding/sub–sampling.** We treat numeric data as accurate to $\sim$1 s.f., using decade–scale censoring for $\geq$10; this is principled, but small departures from the assumed rounding rule are inevitable.

(iii) **Seasonal and temporal imbalance.** Off–season samples are sparse and excluded; early–period coverage is thinner. Both issues widen uncertainty and can subtly affect smoothness.

(iv) **Spatial confounding.** Random intercepts capture persistent site–level shifts, but not spatially structured trends or region–specific temporal slopes; changes in the monitoring network over time can also influence national curves.

(v) **Model choices.** Collapsing AB4 into AB3p stabilises thresholds but coarsens the upper tail; the $p$–power transform aids dispersion for cNorm at the cost of a non–count scale for fitted values.

(vi) **Roster conditioning.** For PA we condition on sites where the family has ever been recorded; this avoids obvious dilution, but implies that absolute occupancy across *all* EA sites is not our target estimand.

### 6.4 Model Diagnostics

Panels in Fig. 12 (a–d) show QQ, residual–vs–fitted, histogram, and ACF for the cNorm fit. Behaviour is broadly acceptable, with mild tail deviations but no concerning autocorrelation after accounting for season–varying smooths.

### 6.5 Implications and Recommended use

Given these constraints, the cNorm seasonal curves provide defensible, population–level indicators of change, suitable for operational surveillance and communication. Agreement in direction and shape across PA, OCAT, and cPois strengthens confidence that the headline conclusions are not artefacts of any single likelihood or data type. For management use, we recommend reporting the *season–average* cNorm trend with 95% intervals, supplemented by spring/autumn curves where seasonal differences matter ecologically. Note that PA inference is *conditional* on the set of sites where each family has ever been recorded. Full per–family PA, OCAT, and cPois outputs (with summary tables immediately following each model's diagnostic figures) are provided in Appendix E.

### 6.6 Future Work

(a) **Covariates and seasonality.** Include month as a cyclic smooth or a season factor–by–smooth interaction with environmental drivers (flow, temperature, conductivity, nutrients) to explain residual variance and reduce over–dispersion.

(b) **Spatial structure.** Add spatial smooths (e.g., Gaussian process on BNG coordinates) and/or region–specific time smooths to separate broad–scale from local change.

(c) **Class calibration.** Following the instructors' suggestion, use post–2000 numeric data to estimate the empirical mean (and width) of each AB class and propagate those as informative intervals for pre–2000 categorical data.

(d) **Joint or hierarchical models.** Explore multi–response formulations to borrow strength across families, or two–part occupancy–abundance models linking PA and cNorm.

(e) **Sensitivity studies.** Systematically vary the ANLA classification rules, the AB3/AB4 merge, the $p$ grid, and the roster definition (e.g., all sites vs. ever–present sites) to quantify impact on trend estimates.

(f) **Alternative error structures.** Where cPois residuals are heavy–tailed, consider location–scale cNorm (time–varying $\sigma$), observation–level random effects, or Tweedie–type approximations for counts.

In summary, the modelling choices adopted here directly reflect EA recording practices, deliver seasonally resolved national trends with calibrated uncertainty, and provide a transparent basis for ongoing surveillance. The proposed extensions offer a clear path to finer ecological interpretation without sacrificing reproducibility.

# 7 Conclusion

This study set out to produce defensible national trends in river macroinvertebrate abundance from the Environment Agency's long-running monitoring programme while respecting how the data were actually recorded. We built a reproducible pipeline that (i) constructs a one-row-per-visit roster, (ii) classifies samples as categorical or numeric using TAXA with conservative overrides, (iii) represents both categorical AB classes and one–significant–figure counts as intervals, and (iv) models seasonally varying temporal patterns with site random effects. The primary analysis—seasonal censored–normal (cNorm) GAMs on a variance–stabilising $p$–scale—delivers population-level (site–marginal) curves with calibrated uncertainty.

Across the four focal families, the cNorm fits explain substantial deviance with randomised-quantile residuals close to unit dispersion. For the worked example (*Aphelocheiridae*) we find a decline into the early 2010s followed by recovery through the 2020s, with autumn levels typically higher than spring. Presence/absence and ordered–categorical models show the same directional changes, and interval–censored Poisson fits recover similar long-term structure albeit with heavier tails—triangulation that increases confidence that the conclusions are not an artefact of a single likelihood or data type.

The main scientific message is that national abundance signals can be extracted from heterogeneous records by (a) treating measurements as intervals reflecting their recording rules and (b) allowing seasonally distinct smooth trends while marginalising over site effects. The outputs are directly usable for surveillance: season-average national trends with 95% intervals, complemented by spring and autumn curves where phenology matters. Because our estimates are site-marginal, they are suited to comparing trajectories across families and to communicating change to non-specialists.

Two practical implications follow. First, continuing to standardise field gear (S3PO) and to record analysis practice will further stabilise inference; where rounding or sub-sampling occurs, explicit flags would improve future modelling. Second, adding routinely available covariates (flow, temperature, nutrients) and modest spatial structure would likely reduce residual variation and sharpen intervals without altering the core pipeline.

Limitations remain—most notably historical ambiguity in ANLA semantics, seasonal/temporal imbalance, and potential spatial confounding—but the modelling choices adopted here are transparent, reproducible, and aligned with EA practice. Overall, the seasonal cNorm framework provides a robust statistical foundation for ongoing national surveillance of macroinvertebrate abundance, with clear avenues for extension as metadata and covariates improve.

## References


[1] P. K. Dunn and G. K. Smyth. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244, 1996.

[2] Environment Agency. Environment agency open data: Ecology explorer. `https://environment.data.gov.uk/ecology/`, 2025. Accessed: 15 August 2025.

[3] A. C. Johnson, Y. Qu, O. Charlotte, N. Isaac, K. Powell, I. Bishop, D. Roy, J. I. Jones, J. Murphy, I. P. Vaughan, and S. J. Ormerod. What has happened to river macroinvertebrate biodiversity in england and wales over the past 30 years?, 2024. SSRN preprint.

[4] E. Pharaoh, M. Diamond, S. J. Ormerod, G. Rutt, and I. P. Vaughan. Evidence of biological recovery from gross pollution in english and welsh rivers over three decades. *Science of the Total Environment*, 878:163107, 2023.

[5] Y. Qu, V. Keller, N. Bachiller-Jareno, M. Eastman, F. Edwards, M. D. Jürgens, J. P. Sumpter, and A. C. Johnson. Significant improvement in freshwater invertebrate biodiversity in all types of english rivers over the past 30 years. *Science of the Total Environment*, 905:167144, 2023.

[6] G. L. Simpson. *gratia: Graceful 'ggplot'-based Graphics and Utilities for GAMs*, 2024. R package.

[7] I. P. Vaughan and S. J. Ormerod. Large-scale, long-term trends in british river macroinvertebrates. *Global Change Biology*, 18(7):2184–2194, 2012.

[8] S. N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, Boca Raton, 2 edition, 2017.

[9] S. N. Wood. *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*, 2023. R package, version 1.9-0 or later.

# Appendices

## A Rmd Code and Knitted Pdf Files

The **R Markdown source code** (`FreshWater_Macroinvetebrates_v9.Rmd`) and its **knitted pdf output** (`FreshWater_Macroinvetebrates_v9.pdf`) used to generate all figures, tables, and statistical models in this report are available at the following GitHub Repo:

`https://github.com/TilakHeble/FreshWater_MacroInvetebrates`

This decision was made to avoid cluttering the appendix with long code blocks. All scripts are reproducible and commented for transparency.

## B Literature Review

Macroinvertebrates are long-established indicators of river condition in the UK because their life cycles integrate water quality, habitat, and disturbance over months to years. National monitoring by the Environment Agency (EA) builds on indices such as BMWP/ASPT and, more recently, WHPT, providing a basis for trend detection across wide spatial and temporal scales. Several syntheses report broad recovery in English and Welsh rivers from the early 1990s onward, coincident with improvements in wastewater treatment and regulation, although trajectories vary among regions and taxa (7; 4; 5; 3).

A practical complication in these archives is measurement heterogeneity. Historically, laboratories recorded taxon abundance using categorical "AB" classes (e.g., 1, 3, 33, 333), while later workflows increasingly reported numerical counts, often to one significant figure and sometimes derived from sub-sampling. These practices imply interval uncertainty around many observations and motivate modelling strategies that respect mixed data types. Two strands of analysis are common: (i) presence/absence (occupancy) models that are robust to counting details, and (ii) abundance models that either treat categories as ordered outcomes or treat rounded counts as censored intervals on the count scale.

Generalized additive models (GAMs) are widely used for ecological time-series because they flexibly represent smooth, non-linear change while accommodating covariates and random effects (8; 9). Within this framework, ordered-categorical likelihoods capture AB-class shifts without inventing pseudo-scores, while censored-count formulations (e.g., Poisson or normal families with interval bounds) propagate rounding and sub-sampling uncertainty through to inference. Randomised quantile residuals provide a convenient, distribution-agnostic diagnostic for both ordered and censored responses (1). Seasonality and spatial heterogeneity are typically handled via factor-by-smooth interactions and site-level random effects, respectively, and model "triangulation" (fitting complementary likelihoods on the same design) is recommended to check that substantive conclusions do not hinge on one error assumption.

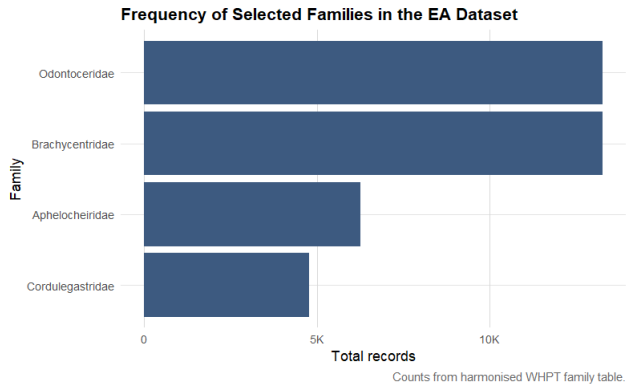
An additional emerging idea is "class calibration": using later, numeric data to estimate the empirical means (and widths) of historic AB classes, thereby refining interval definitions for retrospective analyses. Together, these developments support defensible, reproducible assessment of long-term macroinvertebrate trends from heterogeneous monitoring records.

## C Additional EDA Figures

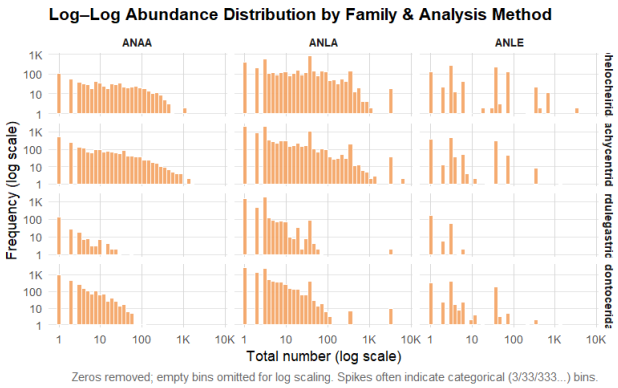
This appendix expands the exploratory analysis with figures that diagnose coverage, recording practice, and data quality. Captions explain how each figure should be read and why it matters for the modelling choices in Sections 4–5.

## C.1 Coverage and abundance distributions

Fig. 15a shows sample counts for the four focal families to give basic scale and balance. Fig. 15b presents log–log histograms of total abundance by analysis method (ANAA/ANLA/ANLE). Long right tails and spikes at powers of ten motivate the use of interval censoring and variance–stabilising transforms.




(a) Total records by family (dataset balance).

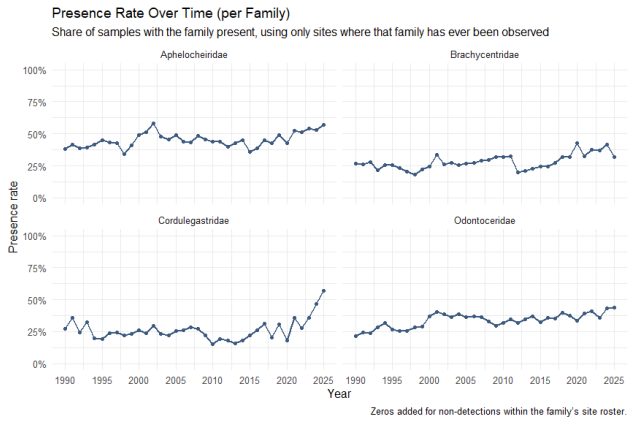



(b) Log–log abundance histograms by family and analysis method.

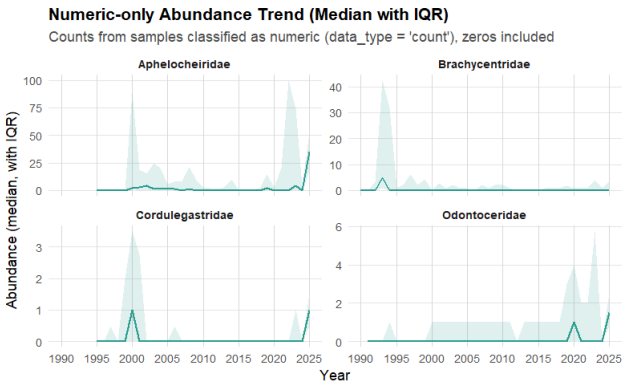
Figure 15: Dataset coverage and abundance distributions.

## C.2 Occurrence and Numeric-only summaries

Fig. 16a gives presence rates through time using only sites where a family has ever been observed (zeros injected for non-detections at those sites). Fig. 16b summarises numeric-only samples via the yearly median and IQR, showing that raw counts are highly skewed and sparse—another rationale for modelling on a transformed scale.




(a) Presence rate over time (site roster restricted to ever-present sites).

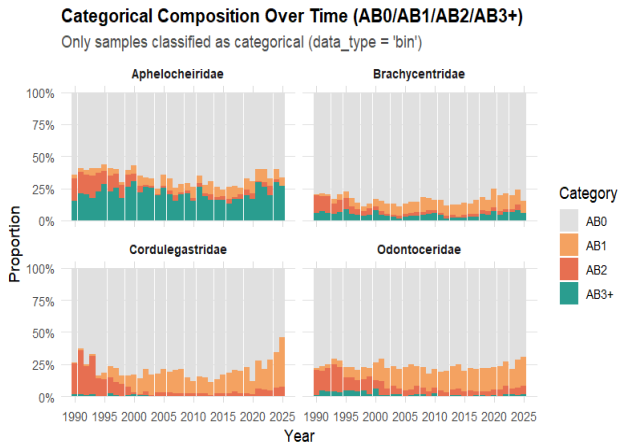



(b) Numeric-only median abundance with IQR (zeros included).

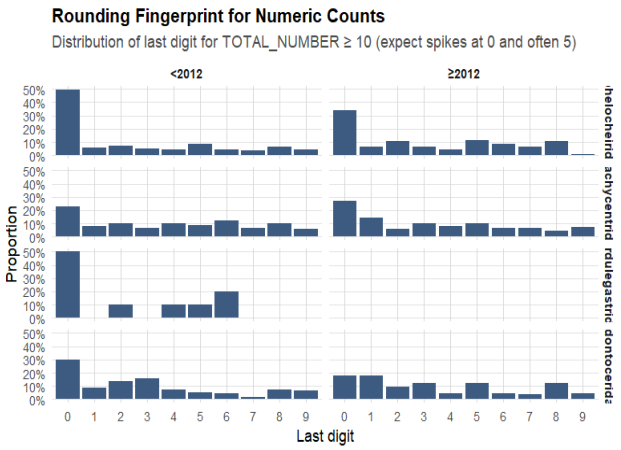
Figure 16: Occurrence and numeric-only summaries.

## C.3 Categorical composition and Rounding fingerprint

Fig. 17a shows the composition of AB classes through time for categorical (BIN) samples; stability of AB0 and rare upper classes motivated merging AB4 into AB3+. Fig. 17b shows last-digit distributions for numeric counts before/after 2012; spikes at $\{0, 5\}$ are consistent with one–significant–figure reporting.




(a) AB-class composition over time (categorical samples only).

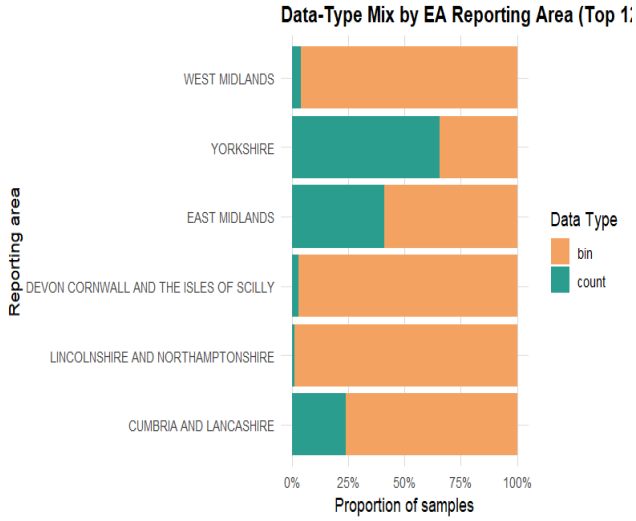



(b) Last-digit fingerprint for numeric counts (pre/post 2012).

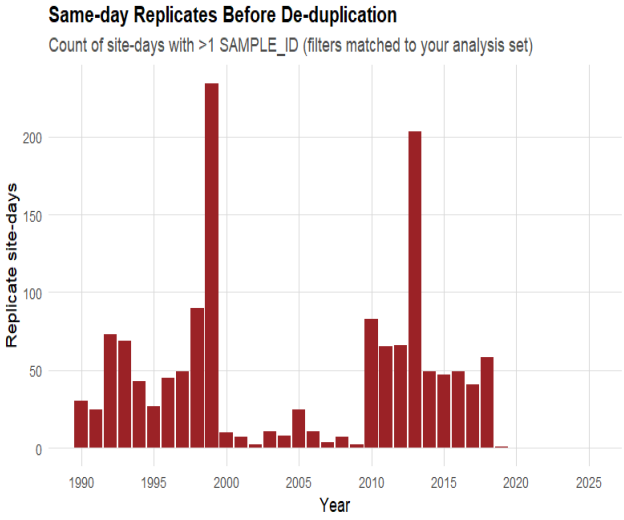
Figure 17: Categorical composition and numeric rounding evidence.

## C.4 Regional mix and Replicate sampling

Fig. 18a compares BIN vs. COUNT usage across the largest EA reporting areas, highlighting regional practice differences that the models must accommodate. Fig. 18b counts site-days with same-day replicates prior to de-duplication, indicating years where sampling effort or QA protocols changed.




(a) Data-type mix (BIN vs. COUNT) by reporting area (top areas).




(b) Same-day replicates before de-duplication (by year).

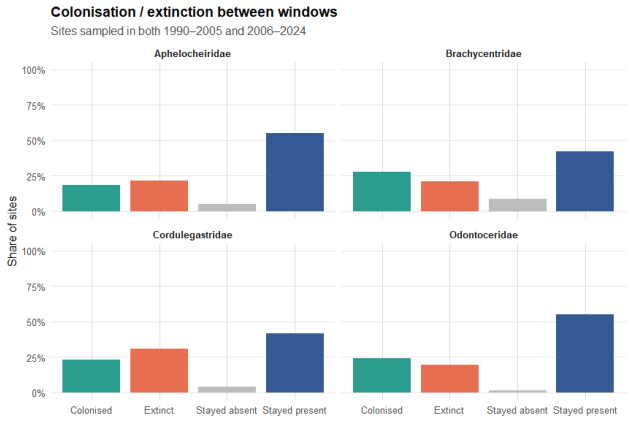
Figure 18: Regional data-type usage and replicate sampling.

## C.5 Site turnover and Zero prevalence

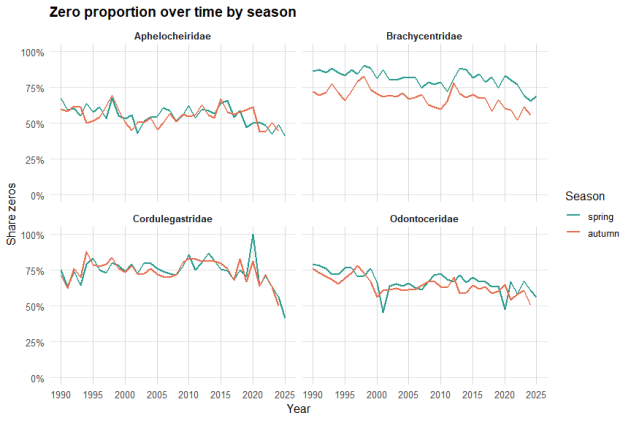
Fig. 19a compares presence between two multi-year windows (1990–2005 vs. 2006–2024) to summarise colonisation/extinction at sites re-visited in both periods. Fig. 19b shows the share of zeros by season, which informs detectability and motivates the PA model in our triangulation.

## C.6 Rounding Patterns over time

Fig. 20 categorises each numeric observation by its rounding signature (1–9 exact, ends with 5, other $\geq 10$, rounded to tens) and tracks proportions over time. The dominance of one–s.f. behaviour supports our decision to treat counts $\geq 10$ as decade intervals in the likelihood.




(a) Colonisation/extinction shares between early and recent windows.




(b) Zero proportion through time by season.

Figure 19: Site turnover and seasonal zero prevalence.

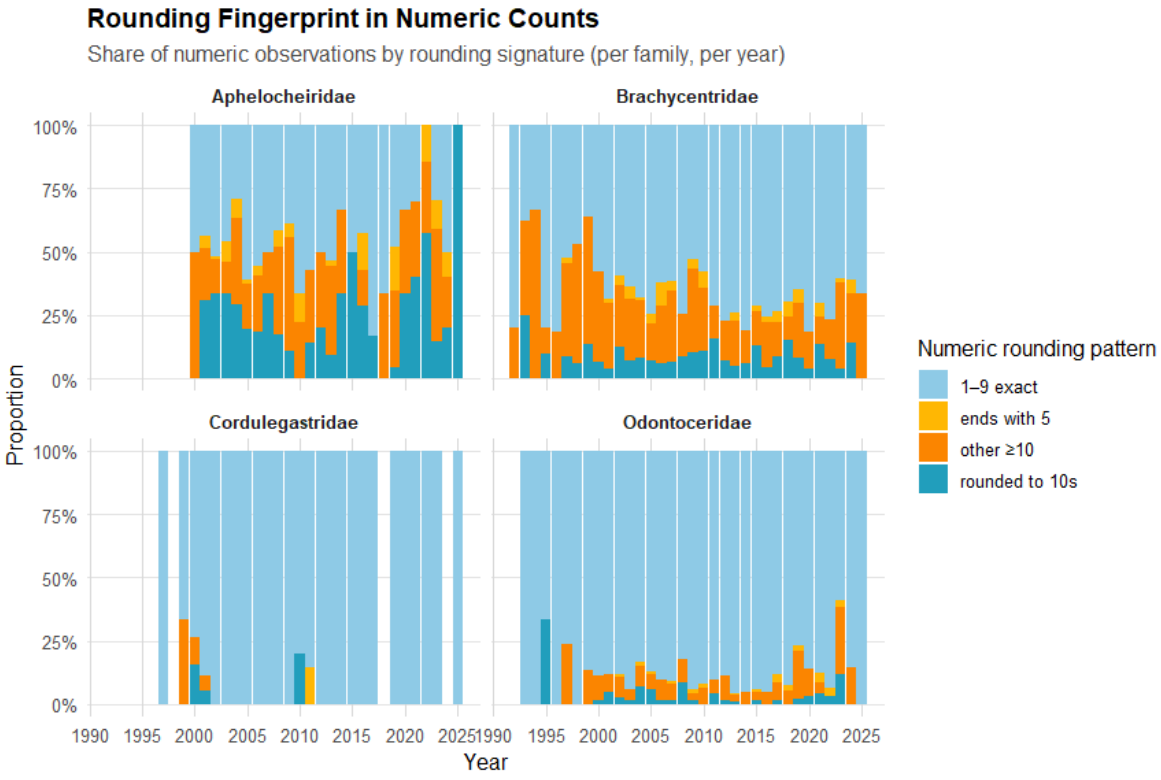



Figure 20: Share of numeric observations by rounding signature (per family, per year).

# D Additional cNorm results of other Families

**Overview.** The following figures replicate the seasonal cNorm analysis for the remaining families using the same roster, censoring rules, variance–stabilising transform, and site random effects described in Section 4. For each taxon we show: (i) season-specific time smooths; (ii) site-marginal national trends on the $p$-scale with 95% pointwise intervals (both by season and season-average); and (iii) randomised-quantile-residual diagnostics (QQ, residual–fitted, histogram, ACF). Transform exponents were $p = 0.35$ for *Brachycentridae* and $p = 0.40$ for *Odontoceridae* and *Cordulegastridae*.

## D.1 Brachycentridae (cNorm, $p = 0.35$)

**Trend :** The season-varying smooths (top row of Fig. 21) show a gradual multi-decadal fluctuation with a trough in the late 2000s followed by a sustained increase post-2010. Spring and autumn curves in the national trend panels (second row) are closely aligned with modest seasonal offsets; uncertainty bands widen in the early 1990s where sampling is sparse (tick marks on the x-axis indicate coverage).

**Diagnostics :** The QQ plot and histogram of randomised quantile residuals (bottom rows of Fig. 21) are close to normal with mild upper-tail departure at the largest fitted means. The residual–fitted panel shows roughly stable spread on the $p$-scale with a slight fan at high fitted values, and the ACF indicates negligible autocorrelation beyond the first few lags, supporting the adequacy of the seasonal cNorm specification.

## D.2 Odontoceridae (cNorm, $p = 0.40$)

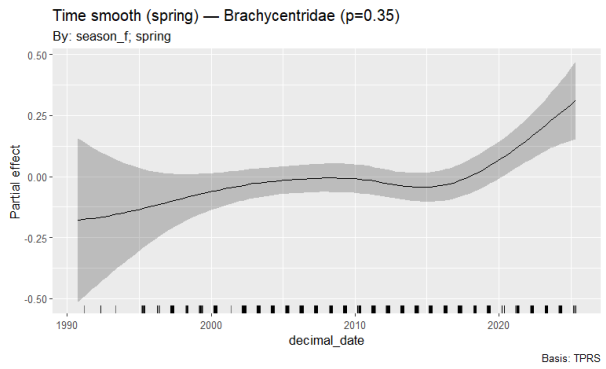
**Trend :** The time smooths (top row of Fig. 22) indicate low-frequency variation with broadly declining abundance into the late 2000s and a clear recovery thereafter. Seasonal national curves (second row) track each other closely; uncertainty narrows after 2000 as numeric data become more prevalent.

**Diagnostics :** Residual diagnostics (bottom rows of Fig. 22) are satisfactory: QQ and histogram are approximately normal with minor upper-tail deviation; residuals vs fitted show no strong structure on the $p$-scale; and the ACF is near zero beyond very short lags. These checks are consistent with an acceptable fit and stable dispersion under the variance-stabilising transform.

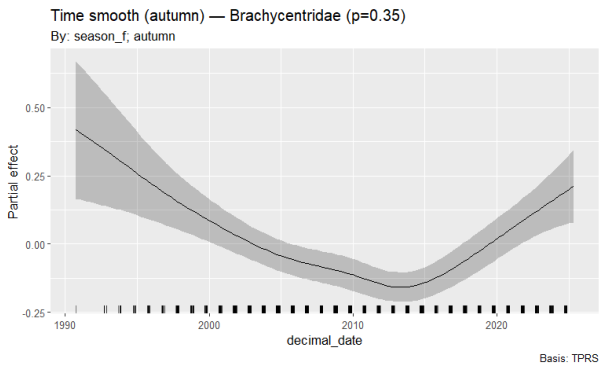
## D.3 Cordulegastridae (cNorm, $p = 0.40$)

**Trend :** With fewer observations for this family, credible bands are wider (Fig. 23). The seasonal smooths suggest a dip around the late 2000s and an increase from the mid-2010s, with spring and autumn exhibiting similar shapes. Early-period uncertainty reflects limited coverage in the 1990s.

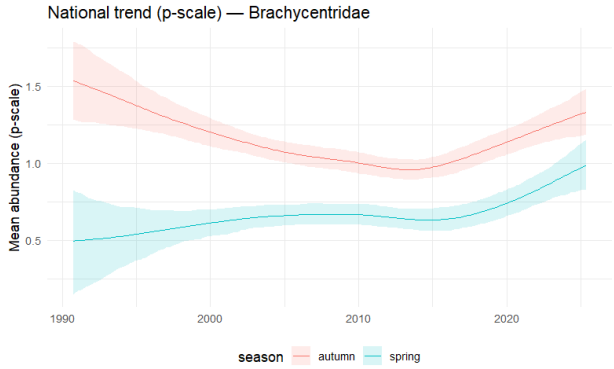
**Diagnostics :** Diagnostics (bottom rows of Fig. 23) remain acceptable given sample size: QQ and histogram show approximate normality with slightly heavier tails; residuals vs fitted reveal no systematic pattern; and ACF values are small beyond initial lags. Overall, model assumptions appear adequate for trend inference on the $p$-scale.



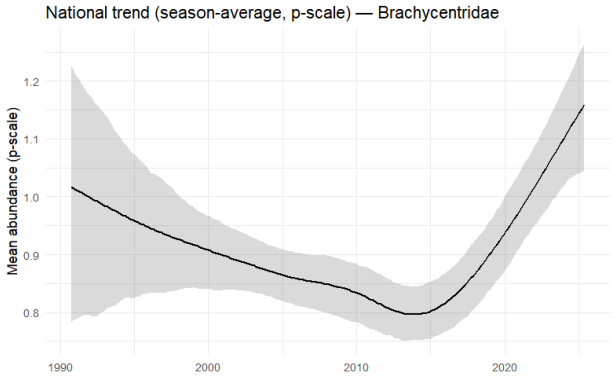

(a) Spring time smooth.



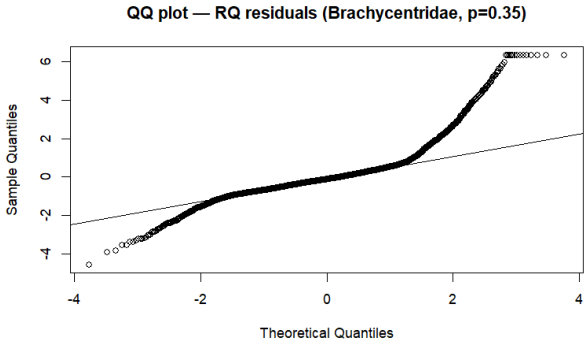

(b) Autumn time smooth.



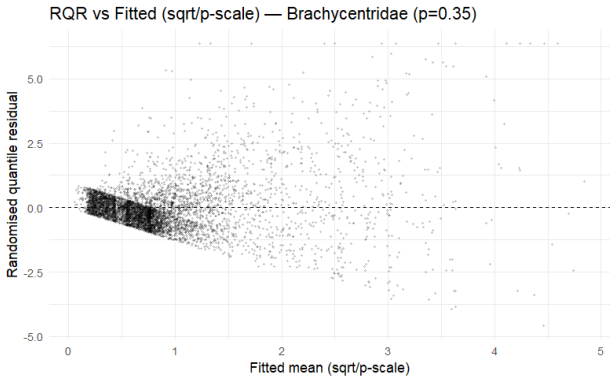

(c) Seasonal national trends (p-scale).



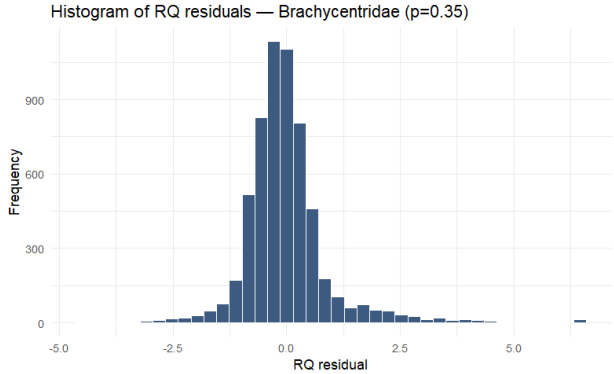

(d) Season-average national trend (p-scale).



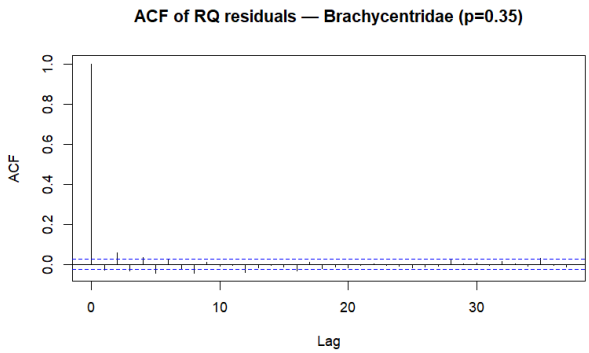

(e) QQ plot of randomised quantile residuals.




(f) RQR vs fitted (p-scale).

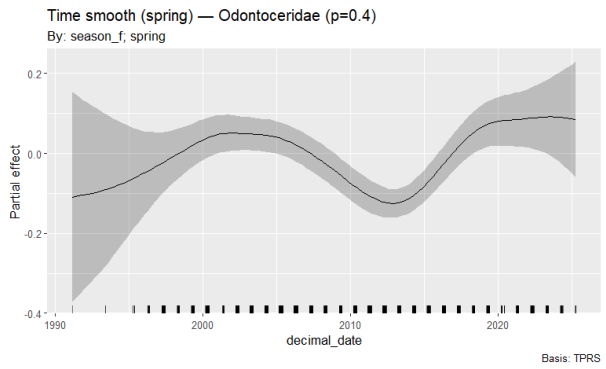



(g) Histogram of RQRs.

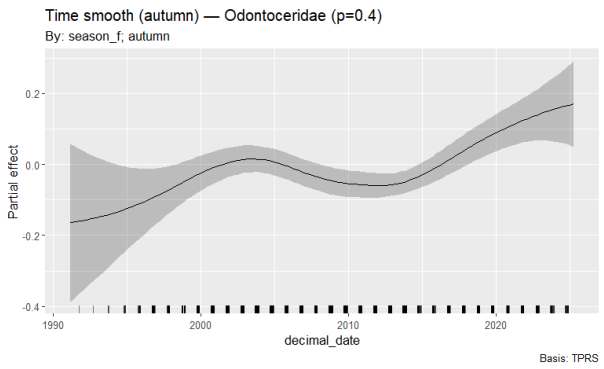



(h) ACF of RQRs.

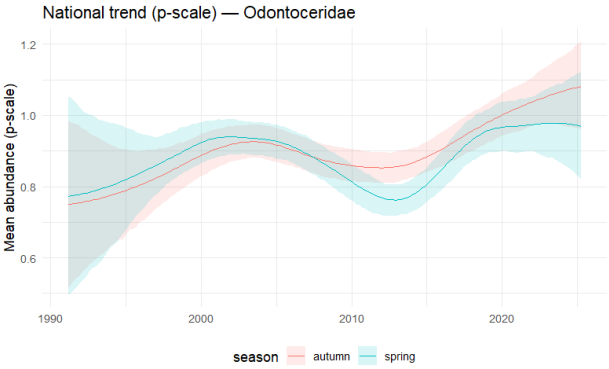
Figure 21: Brachycentridae: seasonal cNorm model outputs and diagnostics (one page).



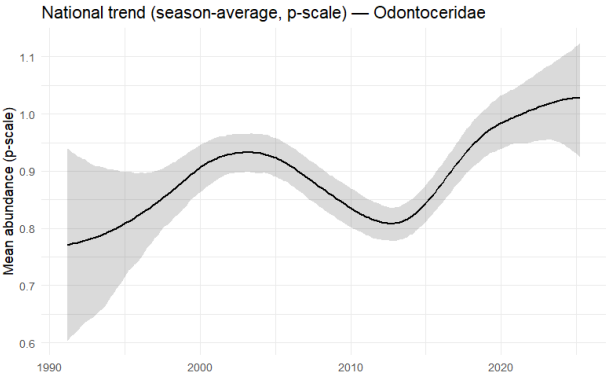

(a) Spring time smooth.



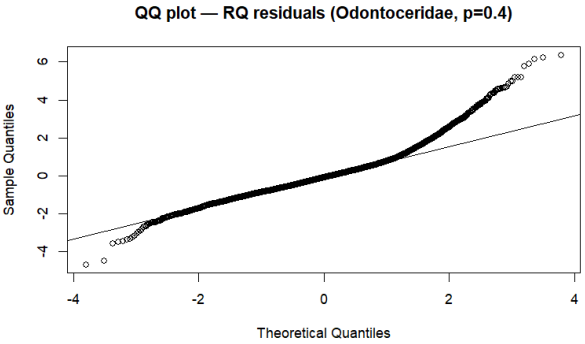

(b) Autumn time smooth.



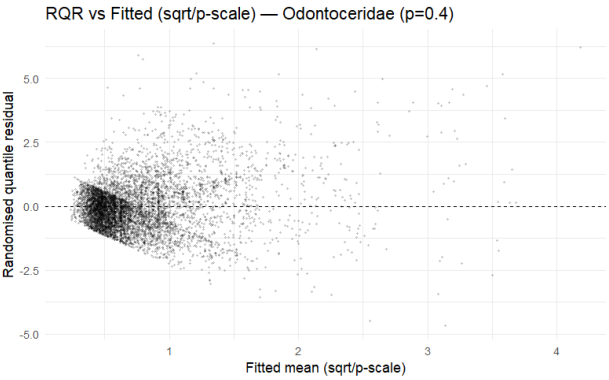

(c) Seasonal national trends (p-scale).



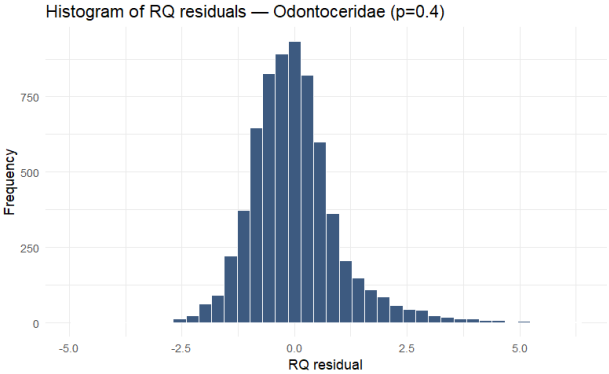

(d) Season-average national trend (p-scale).



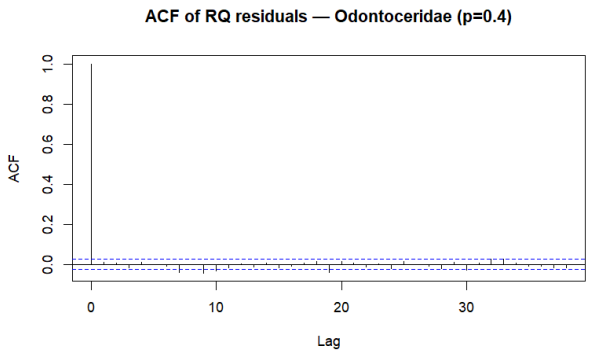

(e) QQ plot of RQRs.




(f) RQR vs fitted (p-scale).

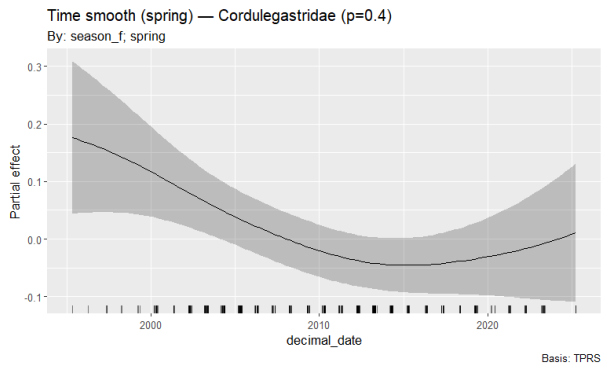



(g) Histogram of RQRs.

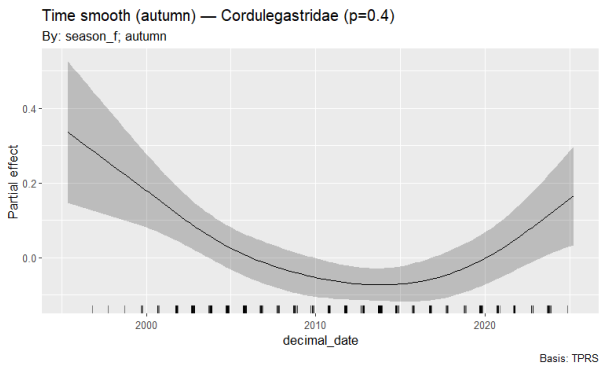



(h) ACF of RQRs.

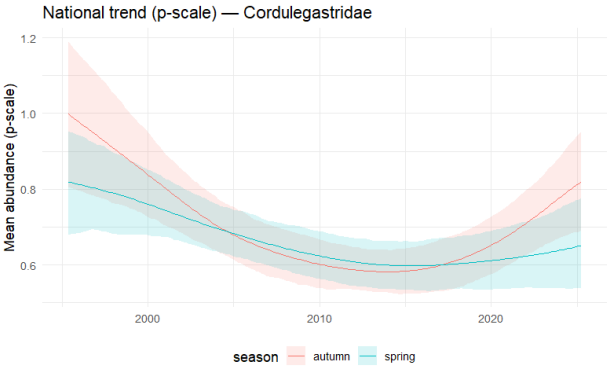
Figure 22: Odontoceridae: seasonal cNorm model outputs and diagnostics (one page).



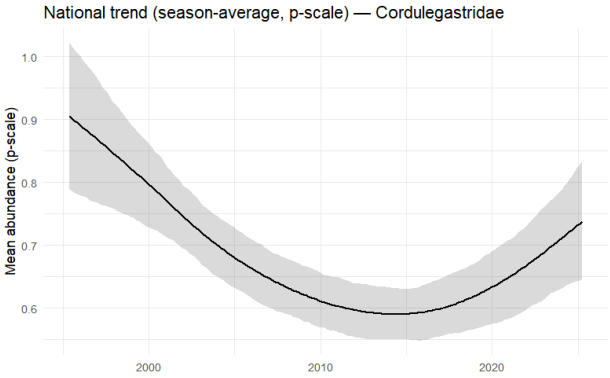

(a) Spring time smooth.



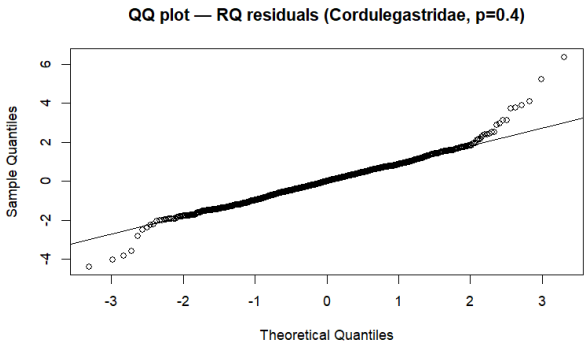

(b) Autumn time smooth.



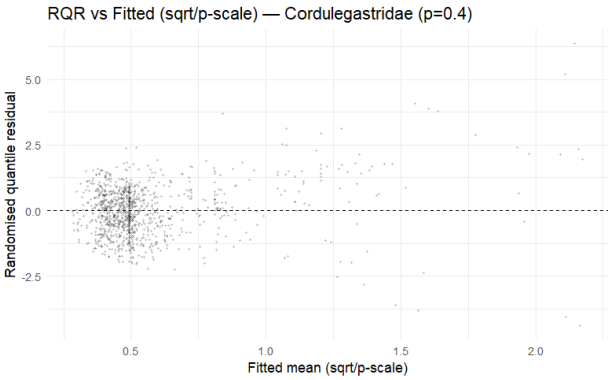

(c) Seasonal national trends (p-scale).



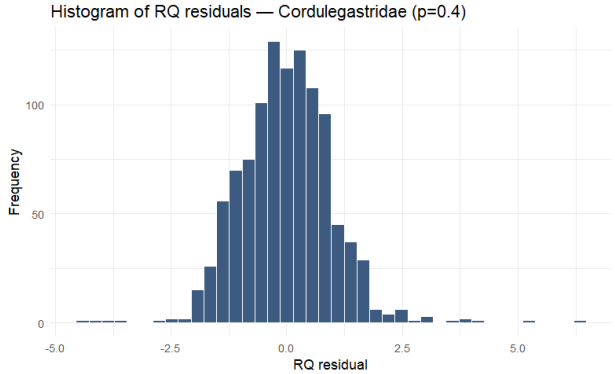

(d) Season-average national trend (p-scale).



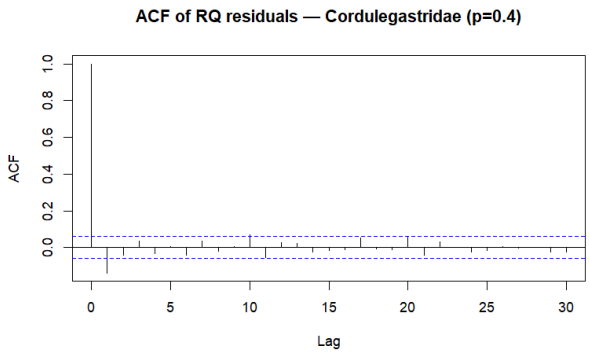

(e) QQ plot of RQRs.




(f) RQR vs fitted (p-scale).




(g) Histogram of RQRs.



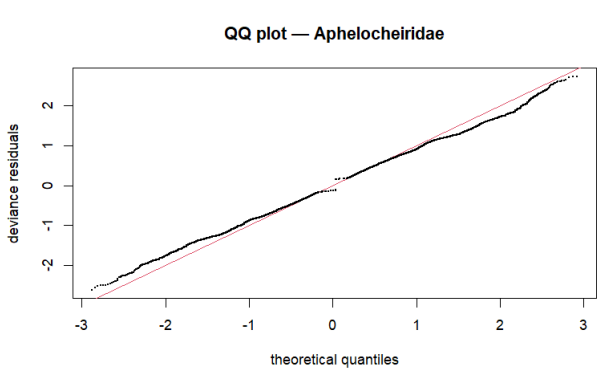

(h) ACF of RQRs.

Figure 23: Cordulegastridae: seasonal cNorm model outputs and diagnostics (one page).

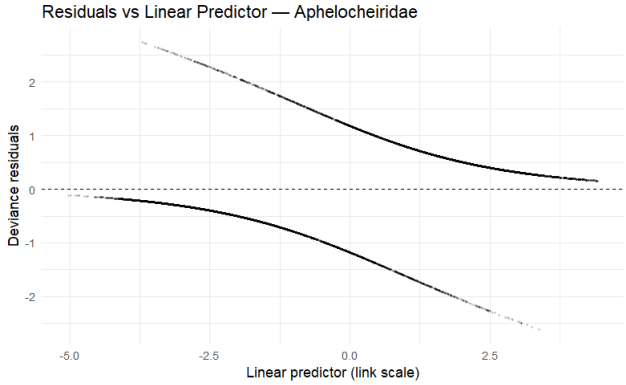
# E Alternative Models and Diagnostics fits (PA, OCAT, cPois)

## E.1 PA (Binomial) – Aphelocheiridae

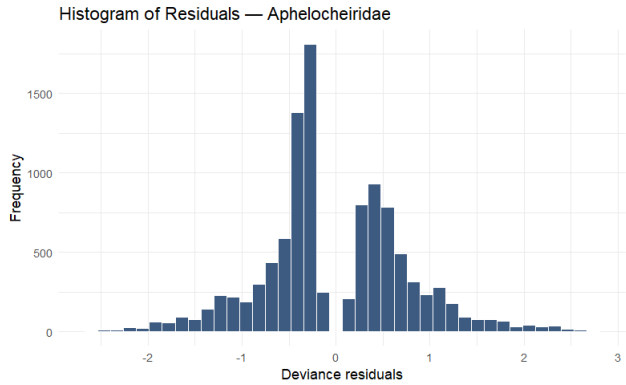
The PA model assesses seasonal change in *occurrence probability* using the same season-varying smooths and site random intercepts as the main analysis. Diagnostics are acceptable: the QQ plot (Fig. 24a) is close to the 1–1 line with mild tail deviation; residuals vs. linear predictor show no strong structure (Fig. 24b); the histogram is centred (Fig. 24c); and observed vs. fitted probabilities show good separation across the probability range (Fig. 24d). Seasonal smooths (Fig. 25a, b) indicate a rise through the late 1990s, a mid-2000s plateau/decline, and a modest recovery post-2015.



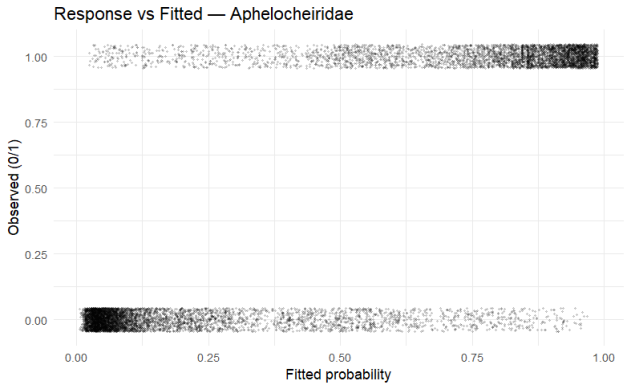

(a) QQ plot (deviance residuals).




(b) Residuals vs linear predictor.

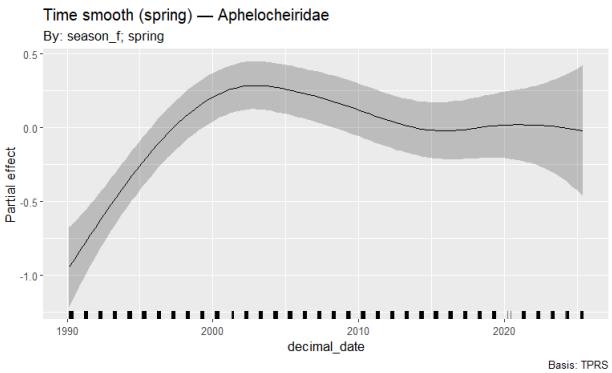



(c) Histogram of residuals.

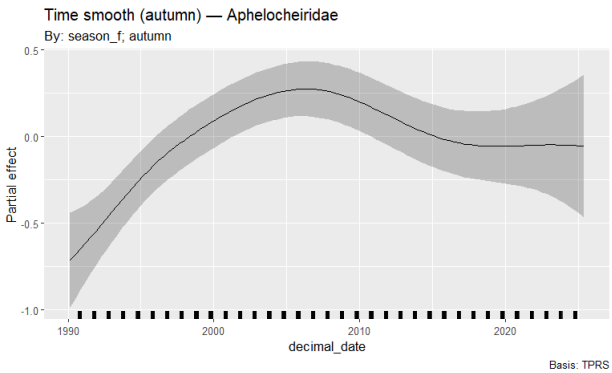



(d) Observed (0/1) vs fitted probability.

Figure 24: Aphelocheiridae PA (binomial) diagnostics.




(a) Spring time smooth.




(b) Autumn time smooth.

Figure 25: Aphelocheiridae PA (binomial): season-varying smooths.

*Cross-family summary.* A compact summary of the PA (binomial) fits—sample sizes, Brier scores, deviance explained, and AIC—is given in Table 4.

| **Family** | $n$ | **Brier** | **Deviance explained (%)** | **AIC** |
|---|---|---|---|---|
| Aphelocheiridae | 10,566 | 0.091 | 55.7 | 7,572.7 |
| Brachycentridae | 39,394 | 0.113 | 35.2 | 31,741.9 |
| Odontoceridae | 33,093 | 0.141 | 30.9 | 32,588.2 |
| Cordulegastridae | 15,451 | 0.122 | 30.6 | 13,953.2 |

Table 4: Presence/absence (PA, binomial) model summary by family, using the same roster, season-varying smooths, and site random intercepts.

## E.2 OCAT (Ordered Categorical) — Aphelocheiridae

The OCAT model uses categorical samples only (`AB0/AB1/AB2/AB3p`). Diagnostics in Fig. 26 are broadly satisfactory, with moderate tail deviations typical of ordered likelihoods. Season-specific smooths in Fig. 27 mirror the PA result: an early increase, a dip around 2010, then recovery. A cross-family summary of OCAT fits is provided in Table 5.

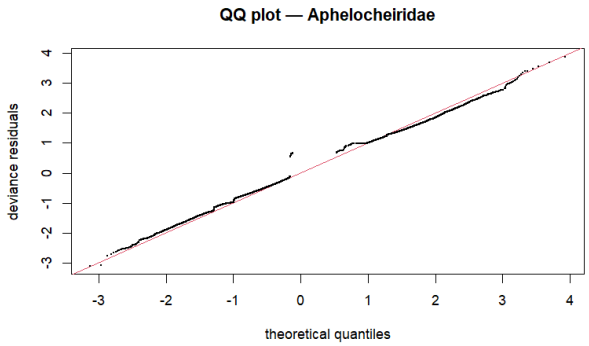
| **Family** | $n$ | $K$ | $k_{\mathbf{time}}$ | **Deviance explained (%)** | **AIC** |
|---|---|---|---|---|---|
| Aphelocheiridae | 9,477 | 4 | 20 | 48.96 | 13,197.25 |
| Brachycentridae | 33,477 | 4 | 20 | 30.87 | 37,341.18 |
| Odontoceridae | 26,419 | 4 | 20 | 30.83 | 32,038.23 |
| Cordulegastridae | 14,387 | 4 | 20 | 30.81 | 14,279.93 |

Table 5: Ordered categorical (OCAT) GAM summary by family (season-varying smooths; site random intercepts).

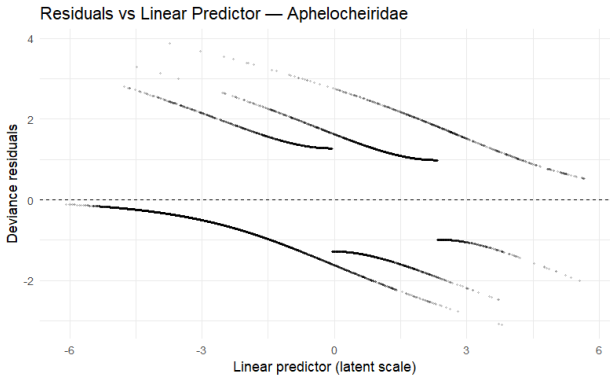
## E.3 cPois (Interval-censored Poisson) — Aphelocheiridae

The cPois model uses count-scale intervals reflecting one–significant-figure practice. For *Aphelocheiridae* ($n$ = 1,089), the fit explains **85.8%** deviance with **AIC = 9,738.0**; seasonal smooths pass the $k$-index check ($\approx 0.96$ for both spring and autumn), and the site random effect has large edf, indicating substantial spatial heterogeneity. However, residual dispersion is high: **SD(RQR) = 2.45** (ideal $\approx 1$) and a crude **Pearson** proxy $\chi^2/\mathrm{df} \approx$ **47.8**, evidencing overdispersion relative to Poisson. This supports using the variance-stabilised *cNorm* as the headline model. Despite heavier tails and some clustering at low fitted means (expected under discretisation), the temporal signal is consistent with PA/OCAT. Diagnostics are shown in Fig. 28.

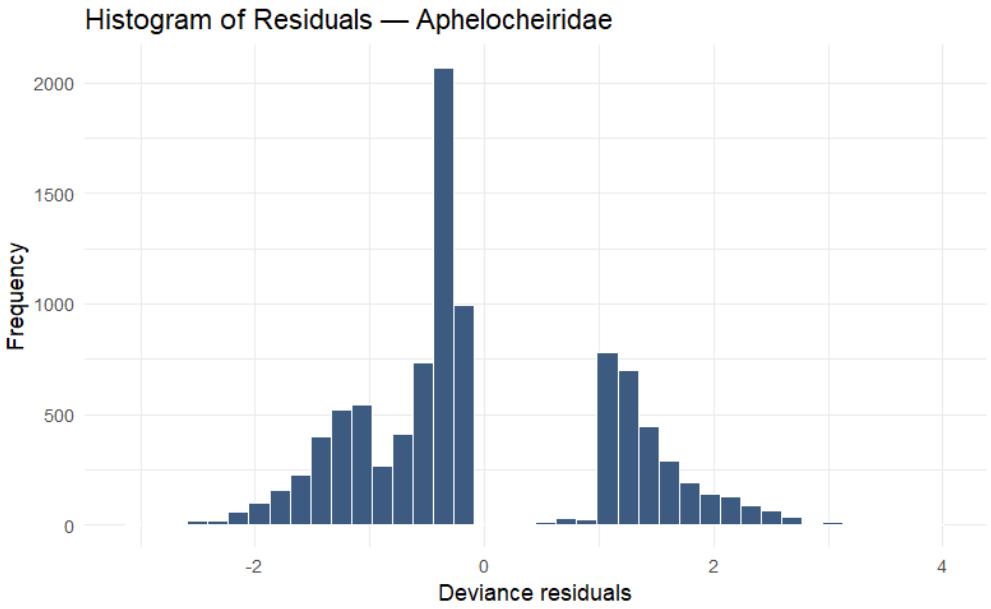
***Reproducibility note :*** To save space and clean report, detailed diagnostic and trend panels are shown in full only for *Aphelocheiridae*. For the other families (*Brachycentridae*, *Odontoceridae*, *Cordulegastridae*) we provide summary tables here; the complete figure sets (smooths, national trends, and diagnostics) can be regenerated from the project code and are also included in the knitted PDF artefacts. Please see Appendix A. for the GitHub repository link and pointers to the exact scripts used to produce all families' plots.




(a) QQ plot (deviance residuals).

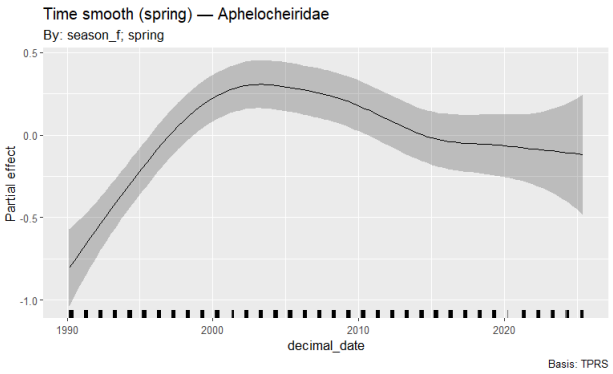



(b) Residuals vs linear predictor.

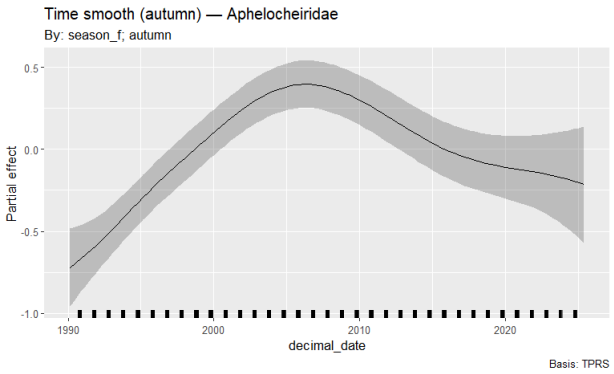



(c) Histogram of residuals.

Figure 26: Aphelocheiridae OCAT diagnostics (2–1 layout for readability).

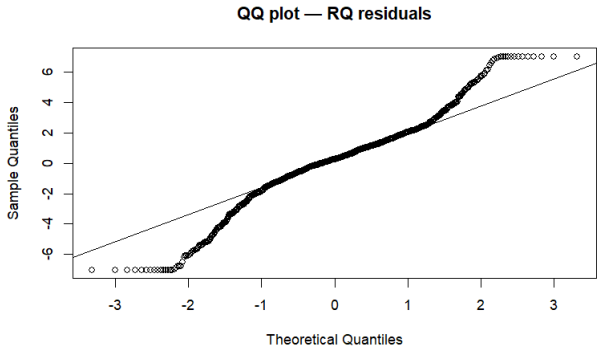



(a) Spring time smooth.

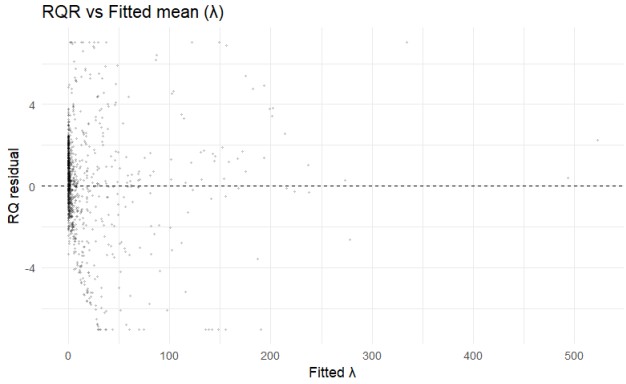



(b) Autumn time smooth.

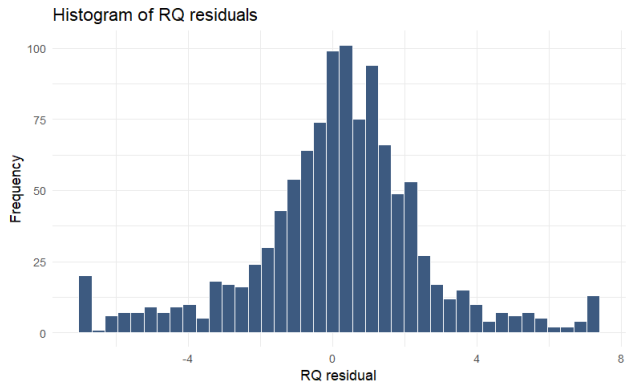
Figure 27: Aphelocheiridae OCAT: season-varying smooths.



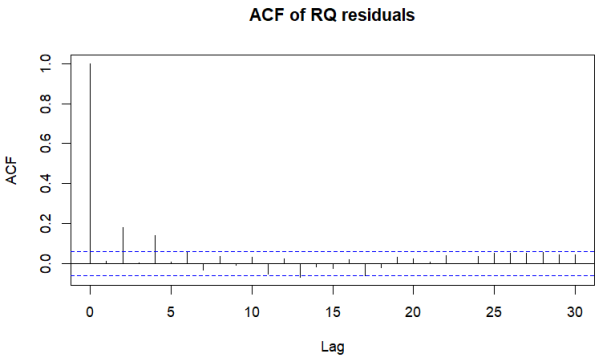

(a) QQ plot (randomised quantile residuals).




(b) RQR vs fitted mean ($\hat{\lambda}$).




(c) Histogram of RQRs.




(d) ACF of RQRs.

Figure 28: Aphelocheiridae cPois (interval-censored Poisson) diagnostics.

# F cNorm variants assessed and model selection (details)

We evaluated several censored-normal (*cNorm*) specifications that differ only in (i) the variance-stabilising transform exponent $p$ and (ii) whether the time smooth is shared or season-specific. A short grid over $p \in \{0.30, 0.35, 0.40\}$ was run for each family using the same roster, censoring rules, and site random effects. Selection favoured settings that (a) flattened residual spread (SD of randomised quantile residuals $\approx 1$), (b) improved QQ/ACF behaviour, and (c) offered competitive explained deviance and AIC. Basis sizes for the time smooths ($k_{\text{time}}$) were set large enough to avoid over-smoothing (checked via the $k$-index, typically 0.97–1.00), with smoothing chosen by REML. National curves are reported as *site-marginal* predictions ($b_j = 0$).

As a sanity test, we computed *Spearman(weight, year) = -0.120* ($|\rho|$ small is good), indicating negligible temporal association. For the cNorm $p = 0.35$ specification, the season-varying time smooth outperformed the single shared smooth: *Base (single time smooth): AIC = 5838.8, DevExpl = 78.3%; Alt (season-varying time): AIC = 5810.2, DevExpl = 79.2%; $\Delta$AIC (alt − base) = -28.6 → prefer season-varying.*

| **Family** | $p$ | $n$ | $k_{\text{time}}$ | **SD(RQR)** | **Dev. Expl. (%) / AIC** |
|---|---|---|---|---|---|
| Aphelocheiridae | 0.35 | 1,089 | 25 | 1.00 | 72.7 / 5,164.8 |
| Brachycentridae | 0.35 | 5,917 | 29 | 0.97 | 49.8 / 21,869.1 |
| Odontoceridae | 0.40 | 6,674 | 28 | 1.00 | 56.2 / 17,987.6 |
| Cordulegastridae | 0.40 | 1,064 | 25 | 0.98 | 72.1 / 1,552.7 |

Table 6: Selected cNorm settings and diagnostics by family (season-varying smooths with site random intercepts). SD(RQR) near 1 indicates well-calibrated dispersion under censoring; higher deviance explained and lower AIC are better.

# G Additional spatial context and time-slice maps

Fig. 29 shows presence-rate maps for *Aphelocheiridae* comparing early vs. recent windows; site size indicates sampling intensity. Fig. 30 contrasts mean cNorm abundance by era and their difference (recent−early). Full code links are provided in Appendix A.

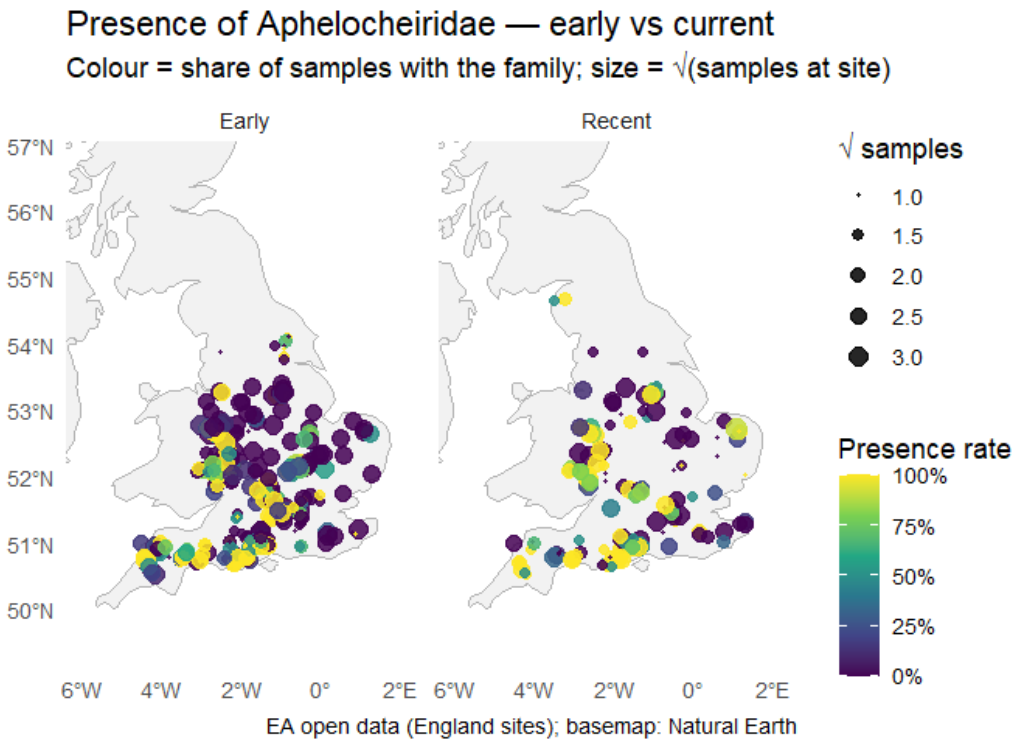
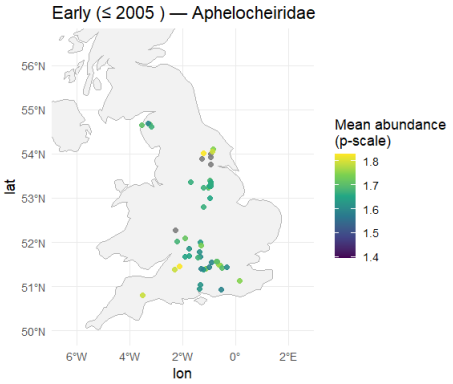
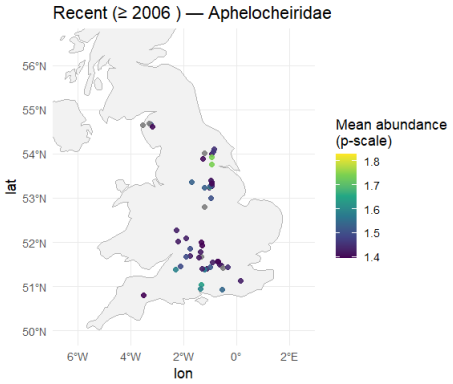


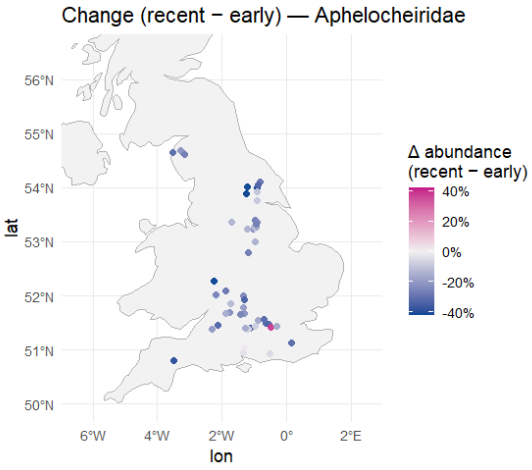

Figure 29: Presence of *Aphelocheiridae*, early vs. recent: colour = share of samples present, size = $\sqrt{\text{samples at site}}$.




(a) Mean abundance (early, cNorm $p$-scale).

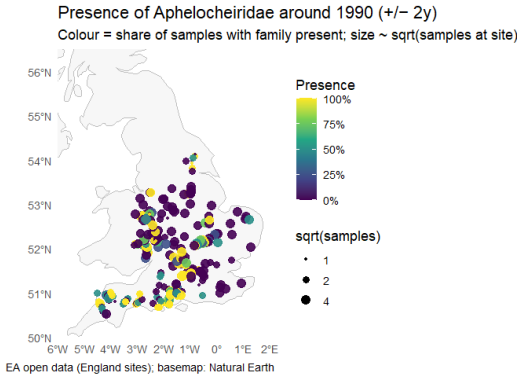



(b) Mean abundance (recent, cNorm $p$-scale).

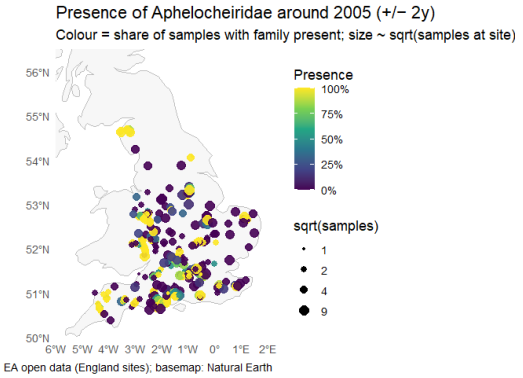



(c) Change (recent−early) in mean abundance.

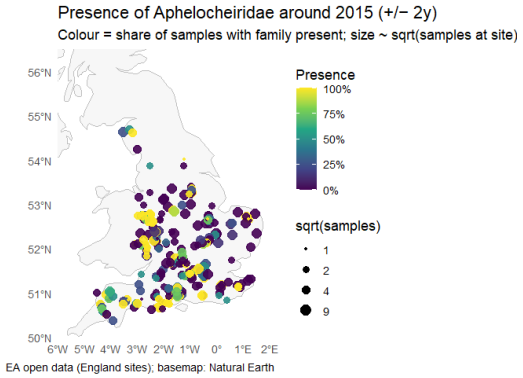
Figure 30: Aphelocheiridae: time-slice abundance maps and change surface.



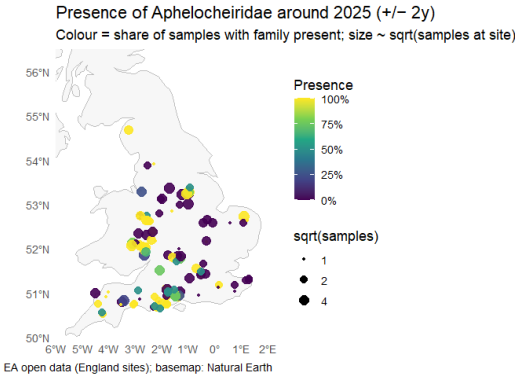

(a) 1990




(b) 2005




(c) 2015




(d) 2025

Figure 31: Presence-rate snapshots for *Aphelocheiridae* across eras.