2023

A statistical analysis of the population of Barbel in the River Teme

Keeling, A.

Keeling, A. and Oben, P. (2023) 'A statistical analysis of the population of Barbel in the River Teme', The Plymouth Student Scientist, 16(2), pp. 157-174.
https://pearl.plymouth.ac.uk/handle/10026.1/21842

The Plymouth Student Scientist
University of Plymouth

A statistical analysis of the population of Barbel in the River Teme

Ava Keeling and Paula Oben

Project Advisor: Michele Lundy, School of Computing, Engineering and Mathematics, University of Plymouth, Drake Circus, Plymouth, PL4 8AA

Abstract
The focus of this project is on the population of barbel (*Barbus barbus*) living in the River Teme, over a number of years with the objective of examining the fish population size specifically investigating any temporal trend in fish population numbers and, if found, with the further aim of examining any covariates that may be causing the trend. In order to answer these questions, two datasets were investigated and rigorously analysed in order to ascertain whether stocks of barbel are in decline and, if so, to further examine which covariates could be contributing to their depletion. A number of statistical methods were used to examine the datasets including descriptive statistics, correlation analyses and statistical modelling techniques. The statistical models included Poisson and Negative Binomial regression which, once fitted to the data were used to predict recorded Barbel numbers. The results suggest that the fish population is in decline.

Keywords: Barbel fish, River Teme, Ecology, Statistical Analysis, Data Modeling, Generalised Linear Model, Environmental

Introduction

Within the discipline of ecology, the topic of freshwater fish populations has attracted considerable interest for a number of years [1]. The main concern being the plummeting numbers of fish found in rivers and what may be the cause of this decline [2]. This report looks into a specific freshwater fish to discover if there is any statistically significant evidence of a temporal trend in the population, whether that be stable, declining or increasing, as well as to study the relationship between this result and a number of potentially influential covariates.

Barbel fish are native to freshwater rivers and streams in Europe and their population numbers have experienced instability [3]. Whilst there have been several articles and evidence in recent and past times that point towards a decline in population size, in this project the given datasets were used to decipher whether that is true [3, 4, 5]

This project is based on the research carried out by Dr. Catherine Roberts, an ecologist at the University of Plymouth, who has been studying the population of barbel, a common freshwater fish, in the river Teme for a number of years.

The Collection of Data

The data used in this article was provided by Dr Catherine Roberts who obtained this data from 2 sources. One was from angler fishing (subsequently referred to as The Angler Dataset), which consisted of anglers (fishermen) catching barbel and recording information about the catch - including numbers of fish caught but also other useful information seen in Table 1 The second dataset was produced by the Environmental Agency (subsequently referred to as The Environmental Dataset), which used electrofishing to carry out surveys and record information. Both sets of data were formatted into multiple Excel spreadsheets, meaning that they had to be organized and compiled to ensure that they were clear, concise and suitable for subsequent analysis.

Angler Dataset

The finalised Angler Dataset consisted of the variables shown in Table 1. It was apparent that there were a group of anglers that fished together (referred to here as 6 Group Anglers) all of which had recorded one additional variable - as can be seen in Table 1.

Throughout the report, the 6 anglers that fished together will be referred to as Group Anglers 1-6, and the others as AC2, AC3, and AC4.

Table 1: Variables in Angler Dataset

Variables	Description	9 anglers	6 of 9 Anglers
Year	The year the data was recorded	✓	✓
Temperature	Temperature of the day measured in °C	✓	✓
Effort	The length of time fished in hours	✓	✓
Number of Barbel	Total number of barbel caught	✓	✓
Flooded	The area fished was flooded The area fished was not flooded	✓	✓
Time of Day	What time of day fishing took place (AM/PM/Evening)		✓

Environmental Dataset

The second set of data was produced by the Environmental Agency using electrofishing along the River Teme. This river begins in mid-Wales, making its way southeast towards the district of Shropshire and is home to many species including barbel fish. 14 sites along this river were taken over by the Environmental Agency in order to collect relevant information regarding the catchment of barbel. A total of 61 entries were recorded between 1975 and 2022. This dataset contained different variables compared to the previous Angler Dataset, as shown in Table 2.

Table 2: Environmental Dataset Variables

Variables	Description
Date	The date data was recorded
Site	The location along the Teme at which the catchings proceeded
Area Fished	The area in m^2 in which a survey was taken
Survey Strategy	How the fish were counted (Catch Depletion/Single Catch/Catch PUE/T) See below for fuller description
Survey Method	The way fish were surveyed (PDC/DC) See below for fuller description
Barbel Total	The total number of Barbel caught

According to the environmental agency it is best to survey using direct current (DC) whenever possible and, in conditions where DC cannot be, pulsed direct current (PDC) fields should be used [6]. These 2 types of surveying were used when collecting information for this dataset.

There were three different survey strategies used when collecting this data:

- Catch Depletion - Where there were initially two catchings and in the second catch if there was more than 50% of the 1st catch then a 3rd catch will be conducted to get the final total.
- Single Catch Sample - This is just simply fishing in the area fished required and using however much fish caught as the total number.
- Catch Per Unit Effort Sample (Catch PUE/T) - Catch PUE/T is calculated by dividing the catch of each fishing trip by the total number of hours (Effort) fished

during that period. This gives Catch PUE/T in units of the number of barbel per hour.

The survey strategy will play an important part in later results.

As is to be expected with information collected from real data sources; some preparatory data cleaning had to be carried out to deal with issues before the analysis stage could begin in earnest.

Methods

Preliminary Methods

Angler Data

In order to investigate patterns in the number of barbel it was important to examine the relationship between the collected covariates and the number of barbel (both at a univariate but also multivariate level - initially as plots and ultimately through modelling). Since the covariate 'Year' is key to understanding the presence or otherwise of a temporal trend in the barbel population, initially a basic plot of the number of barbel caught vs. Year was produced for all anglers together (see Figure 4). From this plot, it became apparent that the vast majority of data points fell between 2005 and 2015. This could be skewing the data and causing any trends seen in this first plot and so this led to the decision of further segmentation of the data into individual anglers. This plot was then reproduced for each individual angler.

The number of hours spent fishing, as well as an individual angler's ability, could well affect the numbers of barbel caught and so the Number of Barbel Trend graphs were segmented first by angler (see Figure 1). Then separately by Effort levels (low: session length $<$ 2 hours, med: 2 hours $<$ session length $<$ 4 hours, high: session length $>$ 4 hours) (see Figure 5 with Effort measured in terms of hours fished) before effectively combining these separate segmentations together by producing Barbel Rate Trends (segmented by angler) (see Figure 6, with Barbel Rate measured by Number of Barbel/Effort and therefore taking Effort into account in addition to the angler segmentation). It was clear from these graphs that Effort and angler are important covariates to consider when moving into the modelling stage. The former will be accounted for within the dependent variable and the latter as one of the many independent variables.

Environmental Data

The Environmental Dataset was initially treated very similarly to the 9 Angler Dataset, starting with a plot showing the number of barbel caught vs. Year. After this, the data was then divided into different sites. First, a plot was produced showing the number

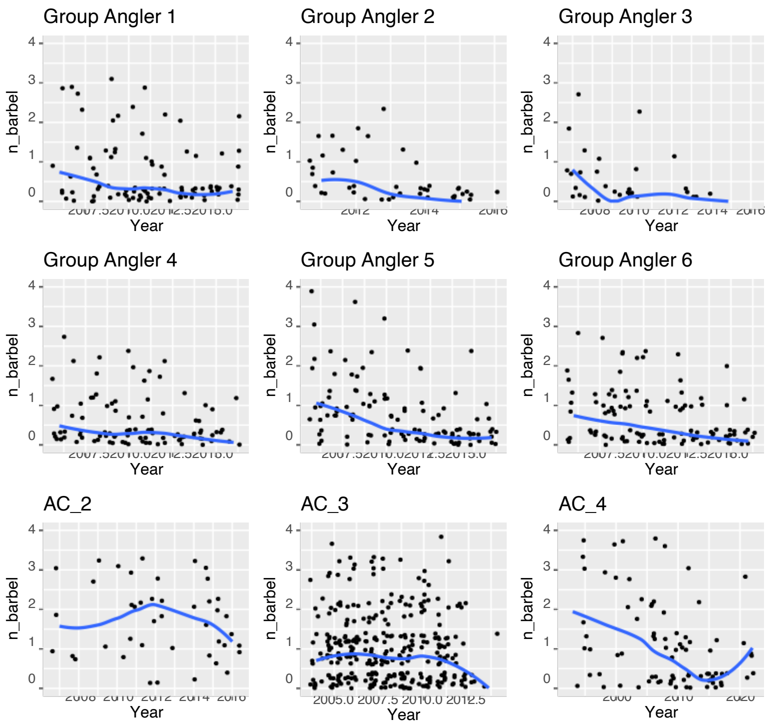


Figure 1: Number of Barbel vs Year for each individual angler

of barbel against site number, in order to easily spot any obvious trends. Then, the mean 'efficiency' of each site was plotted to determine the river sites with the most successful catchment rates (where efficiency = $\frac{NumberofBarbel}{areafished}$). It was found that sites 5 and 9 were substantially more efficient than the others.

Model Selection

Poisson

Due to a difference in the types of variables contained within the different datasets, 3 separate models were developed; one for the Environmental Data ('Environmental Model'), one for the 6 anglers that fished together ('6 Angler Model'), and one for all 9 of the anglers ('9 Angler Model').

Since the data was 'count data', a generalised linear model [7] (`glm` in `R` [8]) with Poisson family was deemed appropriate initially (though for reasons explained in 3.3.2, this choice was later amended). First, univariate models were developed to examine the relationship of each covariate independently of the others on the rate of catch. Then, a model that included all possible covariates was developed, different terms were removed one by one and the change in the AIC was noted to find the best model possible. Also, one-way ANOVA tests were performed on each model to further confirm the best choice

As discussed in 3.1, Effort needs to be included in the dependant variable, in order to account for the effect that it is having on the data. This is done using an 'offset' term within both the angler models (`offset = log(Effort)`) [8]. In effect, it adds $\log(\text{Effort})$ as a term to the model but ensures that the corresponding β value of the $\log(\text{Effort})$ term is set $= 1$. Originally, an 'offset' term was also included in the Environmental data model (`offset = log(AreaFished)`). But, after considering AIC values, it was determined that the model without this term provided a better fit.

Using `glm` with `family = Poisson` assumes that a Poisson distribution is appropriate for the data. But, as the AIC values for these models were still relatively high, other distributions were considered. Using a Poisson distribution assumes that the conditional distribution of the barbel counts has a mean which is equal to it's variance [9]. This is unlikely to be true for the data.

Negative Binomial

Firstly, to justify the use of the Negative Binomial distribution, it was necessary to check if the data was over-dispersed ($Var(Y_i|X_i) > E(Y_i|X_i)$) [9] [1]. This particular distribution was considered appropriate as it is another counting model but includes an extra term to account for the excess variance [10]. The variance of this model is a quadratic function of the mean [11]. By using the `dispersiontest` function [12] in `R`, the following `R` output for the 9 angler model was obtained:

```
        Dispersion test

data:  nineanglers
z = 3.4899, p-value = 0.0004833
alternative hypothesis: true dispersion is not equal to 1
sample estimates:
dispersion
  1.691628
```

Figure 2: `R` output for all 9 Angler Dispersion Test

This output shows that the dispersion value (1.691628) is greater than 1, therefore indicating over-dispersion is present. However, it is necessary to test the hypothesis of $H_0 : \theta = 1$ vs. $H_1 : \theta \neq 1$, where $\hat{\theta} = 1.691628$ to confirm whether this assumption is statistically significant.

As the p-value = 0.0004833 (as seen in Figure 2), there is overwhelming evidence to reject H_0 and accept H_1 [2]. Therefore, using the Negative Binomial distribution will provide a better model for this dataset. This dispersion test was then repeated for each of the Poisson models, yielding similar results (see Appendix).

After verifying that the data was over-dispersed for all 3 models, using a Generalized Linear Model [7], Negative Binomial regression was performed for all 3 data-sets in order to find a model with a stronger fit to the data. In a similar way to when using the Poisson Distribution, a model containing all possible covariates was created and then used to investigate which terms were statistically significant, by performing hypothesis tests on each of the coefficients. Insignificant terms were removed from the model and

[1] Where *E* means the Expected Value.

[2] Throughout this report a significance level of 5% is used

the AIC was used to determine which model provided the best fit. Again, 'Effort' was offset in these models for the same reasons previously stated.

To further confirm the final models selected, they were used to predict barbel numbers for the same covariate values as the original datasets. These predictions were then compared to the real values of number of barbels caught in order to see if the model provided a good fit to the data. The predicted values were found using the `predict` function in `R` [8]. For the 6 Angler model, a plot was produced of the Number of Barbel vs. Year for all cases in the dataset. For simplicity, only Effort $= 2, 4, 5$ were included, as this was where most of the data points were (Figure 3).

The blue line shows the predicted Number of Barbel and the red lines shows the 95% confidence interval for these values. It can be seen that the predicted number of barbel closely follows the trend in the real data, and therefore it is apparent that the model appears to offer a good fit to the data.

Similar plots were also produced for the Environmental Data model (see Appendix).

Unfortunately, due to the number of different continuous variables within the 9-angler model, it was not possible to produce similar plots for this model.

Quasi-Poisson

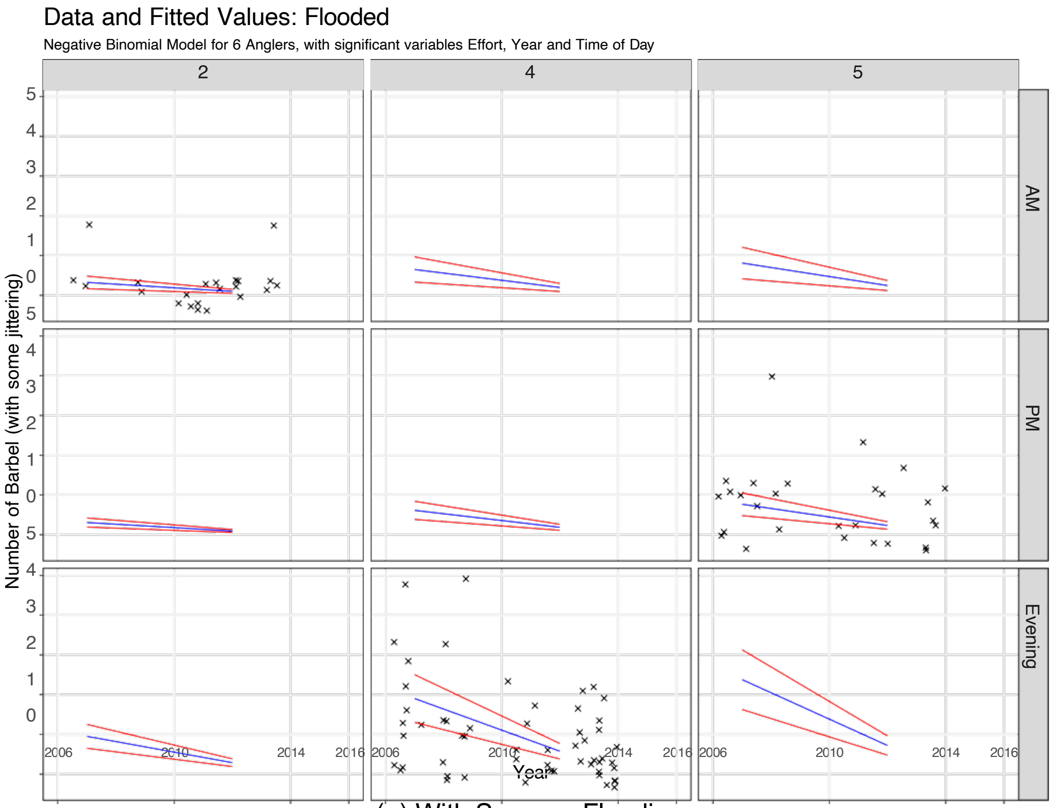
Negative Binomial regression is not the only way of dealing with over-dispersion and an alternative approach is to use Quasi-Poisson regression. The Quasi-Poisson distribution is another statistical distribution that is often used in modelling count data. It is similar to the Negative Binomial distribution in that it has the same number of terms (one more than Poisson to account for the extra variance). However, the variance of this model is a linear function of the mean [11]. This model is an extension of the Poisson Distribution in that it assumes proportionality rather than equality between the conditional mean and variance (Var$(Y_i|X_i) \propto$ E$(Y_i|X_i)$) [9]. This was applied to the 6 angler model using the `glm` [8] function in `R`, setting `family = quasipoission`.

As a Quasi-Poisson model doesn't have a likelihood function (due to not 'necessarily having a distributional form'), there is a distinct lack of information critereon available to directly compare this to the previous models used [11]. However, the standard error and t-value columns can be viewed to assess if it is a better distribution for the data.

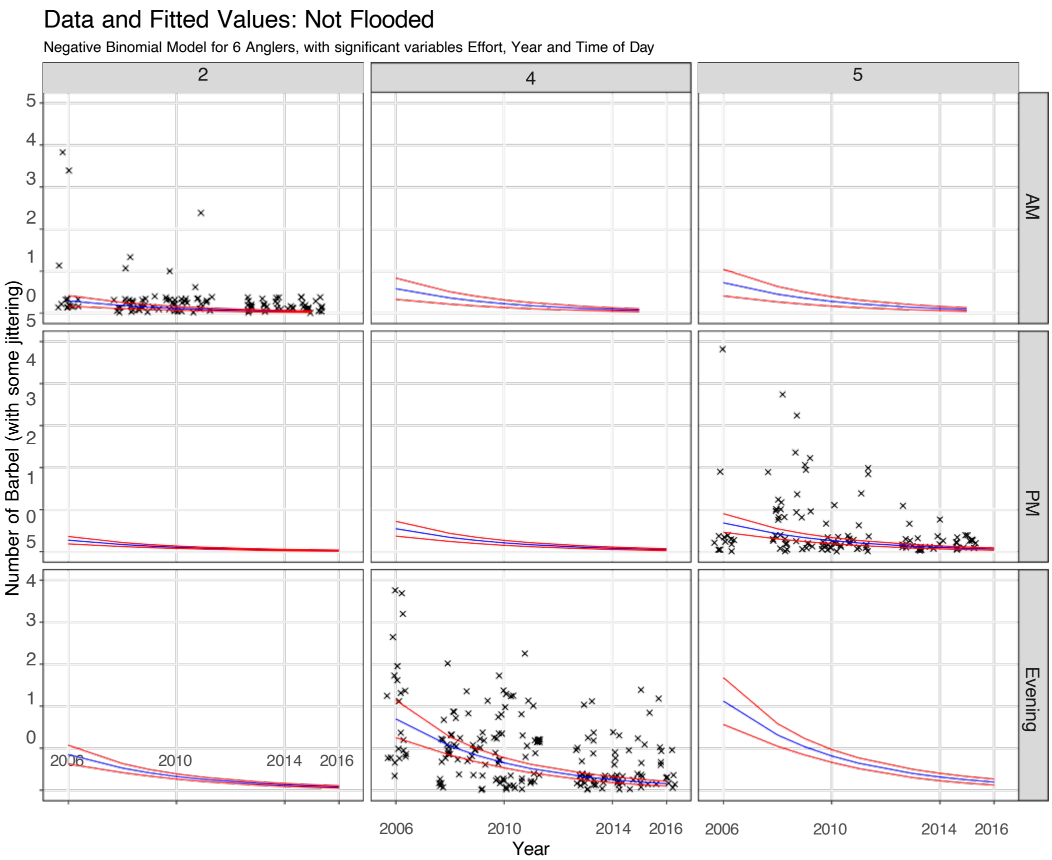
Results and Discussion

Angler Data

It can be seen from Figure 4 that there appears to be a negative correlation between the number of Barbel caught and the Year. It is also apparent that most of the data fell between 2005 and 2015, and could be interpreted that this is the cause of the apparent trend. As mentioned in 3.1, this led to the decision to further segmentation of the data into individual anglers and Effort levels.



(a) With Summer Flooding



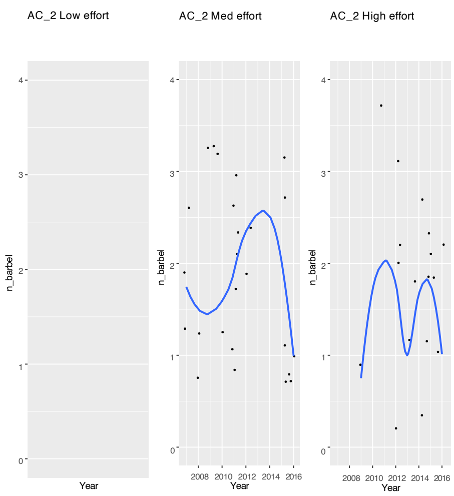
(b) Without Summer Flooding

Figure 3: Fitted values of Number of Barbel from the Negative Binomial model and real data points, plotted against Year.

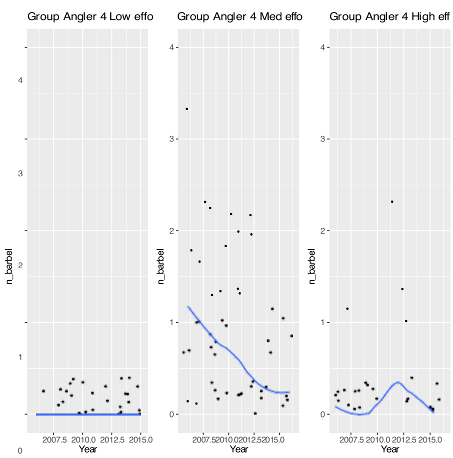
Figure 4: Number of Barbel vs Year for all anglers

.

Although all of the group angler's plots are remarkably similar to each other, these are considerably different from those of the other anglers (AC_2 and Group Angler 4's plots shown in Figure 5 for reference. All other angler's plots can be seen in Appendix).



(a) Plots for AC_2



(b) Plots for Group Angler 4

Figure 5: No. of Barbel vs. Year plotted for one of the group anglers and one of the separate anglers, with Effort levels segmented. All other angler's plots available in Appendix, these are just shown here for reference.

In the other angler's plots (Figure 5a), generally, there are no real trends and the data is very erratic. This could be because of the large spread of data for these anglers. For the group Anglers (Figure 5b), generally, at a low effort, the number of barbel being caught remains the same. At a medium effort, there is a clear decline. However, at a high effort, the results vary greatly from angler to angler. Some anglers are more inclined to put a high level of 'Effort' into their sessions than others, therefore leading to this variety.

These results show that 'Effort' was having a real effect on the number of barbel being caught and therefore needed to be considered when moving to the modelling stage. This led to the decision that the rate of barbel caught would be a more interesting statistic to look at. It was determined for each angler by dividing the number of barbel being caught by the Effort for each entry.

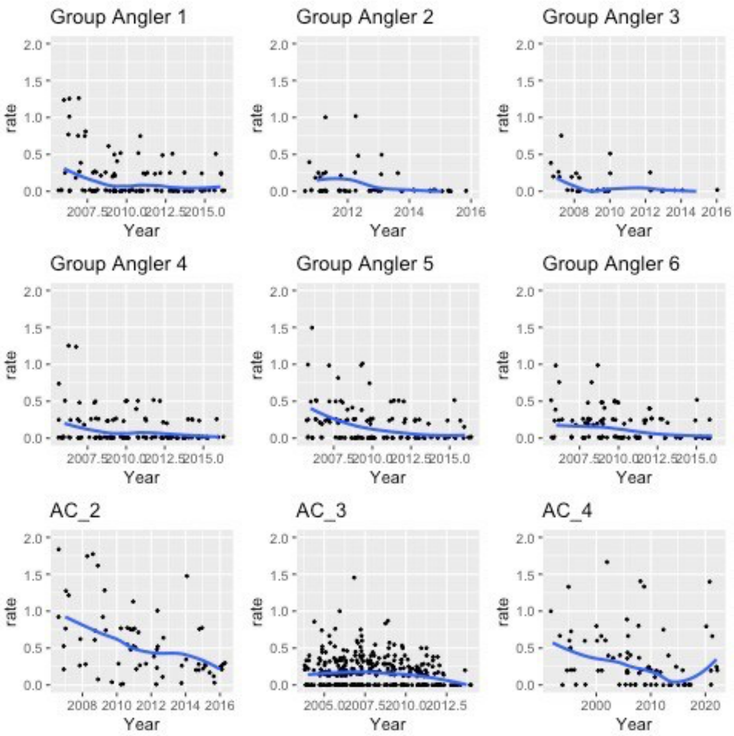


Figure 6: Rate vs Year for each individual angler

Looking at the Figure 6 indicates a decreasing trend in all of these graphs, therefore further confirming the merit of using the rate as the dependent variable.

Environmental Data

The first graph produced for the Environmental data shows that there appears to be a downward trend over time in the number of barbel caught. It was theorized that different sites could be influencing the number of barbel being caught, and so this led to plotting the number of barbel caught against the site number (see Figure 7), to see if there was any significant trend. There is no real correlation shown between these two and therefore it can be said that the different sites are not having a statistically significant effect on the number of barbel being caught.

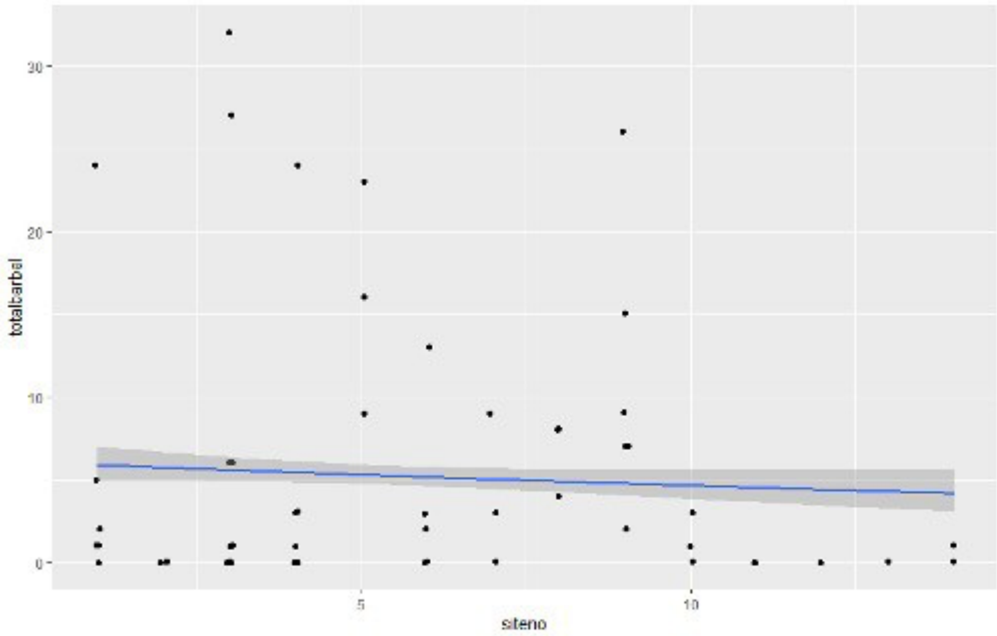


Figure 7: Number of Barbel caught vs. Site number

Testing Distributions

Poisson

The equation below shows the form of the final Poisson model for the 9 Angler Dataset where the β values and their standard errors are given in Table 3.

$$\log(E(\frac{\text{Number of Barbel}}{\text{Effort}})) = \beta_0 + \beta_1(\text{Year}) + \beta_2(\text{Temperature}) + \beta_3(\text{Flooded}) + \beta_{4i}(\text{Angler}_i) \quad (1)$$

Table 3: All 9 Angler model

Variable	β Value	SE(β)
Year	-0.09	0.00864
Temperature	0.06	0.0108
is Flooded	-0.33	0.0792
Angler (AC_3)	-1.52	0.0928
Angler (AC_4)	-1.28	0.132
Angler (Group Angler 1)	-1.67	0.136
Angler (Group Angler 2)	-1.95	0.240
Angler (Group Angler 3)	-2.31	0.268
Angler (Group Angler 4)	-1.99	0.153
Angler (Group Angler 5)	-1.57	0.129
Angler (Group Angler 6)	-1.65	0.135

The equation below shows the form of the final Poisson model for the 6 Angler Dataset where the β values and their standard errors are given in Table 4.

$$\log(E(\frac{\text{Number of Barbel}}{\text{Effort}})) = \beta_0 + \beta_1(\text{Year}) + \beta_2(\text{Temperature}) + \beta_3(\text{Flooded}) + \beta_{4i}(\text{Angler}_i) + \beta_{5i}(\text{Time of day}_i) \quad (2)$$

Table 4: 6 Angler model

Variable	β Value	SE(β)
Year	-0.24	0.0222
Temperature	1.00	0.198
is Flooded	-0.33	0.130
Time of day (Evening)	1.00	0.198
Time of day (PM)	-0.09	0.220
Angler (Group Angler 2)	-0.09	0.269
Angler (Group Angler 3)	-0.14	0.270
Angler (Group Angler 4)	-0.44	0.279
Angler (Group Angler 5)	-0.80	0.355
Angler (Group Angler 6)	-0.02	0.266

The equation below shows the form of the final Poisson model for the Environmental Dataset where the β values and their standard errors are given in Table 5.

$$\log(E(\text{Number of Barbel})) = \beta_0 + \beta_1(\text{Year}) + \beta_{2i}(\text{Survey Strategy}_i) + \beta_3(\text{Area Fished}) \quad (3)$$

Table 5: Environmental data model

Variable	β Value	SE(β)
Year	-0.04	0.0100
Survey Strategy Catch PUE/T Sample	-3.77	0.723
Survey Strategy Single Catch Sample	-0.61	0.125
Area Fished	-0.0001	0.0000306

The final models (for all 3 datasets) using the Poisson distribution (Equation 1) suggest that **all** covariates are having a statistically significant effect on the expected catch rate of barbel. This may be due to the Poisson distribution providing a poor fit to the data.

In all of these models, the β values that correspond to Year are negative, meaning that $E(\frac{\text{Number of Barbel}}{\text{Effort}})$ is decreasing by a factor of e^{β} for the angler models. This means that for every year increase, the 9 Angler Model decreases by a factor of 0.914 and the 6 Angler Model by 0.787. While 'Year' is also having a diminishing effect on the Environmental Model, here there is a different dependent variable in use and so $E(\text{Number of Barbel})$ is decreasing by $e^{\beta} = 0.961$ for each year. As all of the models show that Year has a negative coefficient, this further strengthens the argument that barbel numbers are decreasing over time, as seen in earlier plots.

Quasi-Poisson

When producing the 9 Angler Model using the Quasi-Poisson distribution it can be seen that, again all variables in the model are significant. From what we know about the Quasi-Poisson this is no surprise.

The 6 Angler Model, using the Quasi-Poisson distribution, shows that only 2 variables are significant enough to be used compared to the 5 variables in the Poisson distribution model. This is a much smaller model.

For the Environmental Data when applying the Quasi-Poisson to an all-variable model, it was apparent that no variable was statically significant at a 5% nor 10% significance level. In theory, this is saying that no covariate being recorded is having any impact on the number of barbel caught. Whilst this is a plausible model to have, it is necessary to assess and compare with the Negative Binomial Model for the same data, so the best fitting model can be chosen.

Since Quasi-Poisson models do not have a distributional form, it is not possible to use ANOVA or chi-squared tests to compare the two models. However, since Quasi-Poisson is just an extension of the Poisson distribution, this model has the same coefficients as the Poisson model. The only difference is in the standard errors, which are considerably larger [11]. As it has already been seen that the Poisson distribution provides a poor fit to the data, it is safe to assume that the Quasi-Poisson will not be superior. Therefore, the Negative Binomial models are preferred.

Negative Binomial

The equation below shows the form of the final Negative Binomial model for the 9 Angler Dataset where the β values and their standard errors are given in Table 6.

$$\log\left(\frac{\text{Number of Barbel}}{\text{Effort}}\right) = \beta_0 + \beta_1(\text{Year}) + \beta_2(\text{Temperature}) + \beta_{3i}(\text{Angler}_i) + \beta_{4i}(\text{Flooded}_i) \tag{4}$$

The equation below shows the form of the final Negative Binomial model for the 6 Angler Dataset where the β values and their standard errors are given in Table 7.

$$\log\left(\frac{\text{Number of Barbel}}{\text{Effort}}\right) = \beta_0 + \beta_1(\text{Year}) + \beta_{2i}(\text{Flooded}_i) + \beta_{3i}(\text{Time of day}_i) \tag{5}$$

The equation below shows the form of the final Negative Binomial model for the Environmental Dataset where the β values and their standard errors are given in Table 8.

$$\log\left(\frac{\text{Number of Barbel}}{\text{Effort}}\right) = \beta_0 + \beta_{1i}(\text{Survey Strategy}_i). \tag{6}$$

It is clear that for the 9 Angler Model, all p-values are < 0.001 (Table 6, and therefore there is overwhelming evidence that $\beta \neq 0$ for all covariates. This is similar to

Table 6: All 9 Angler Final Model

Variable	β Value	SE(β)	p-value
Year	-0.09596	0.01160	$< 2e-16$
Temperature	0.06581	0.01448	$5.52e-6$
is Flooded	-0.38819	0.10578	$2.43e-4$
Angler (AC_3)	-1.53190	0.14585	$< 2e-16$
Angler (AC_4)	-1.18729	0.18518	$1.44e-10$
Angler (Group Angler 1)	-1.73132	0.18517	$< 2e-16$
Angler (Group Angler 2)	-1.95960	0.27939	$2.32e-12$
Angler (Group Angler 3)	-2.38287	0.31376	$3.09e-14$
Angler (Group Angler 4)	-2.02657	0.19767	$< 2e-16$
Angler (Group Angler 5)	-1.59744	0.17863	$< 2e-16$
Angler (Group Angler 6)	-1.66536	0.18269	$< 2e-16$

Table 7: 6 Angler Final Model

Variable	β Value	SE(β)	p-value
Year	-0.23974	0.02505	$1.15e-13$
is Flooded	-0.35543	0.16273	0.0225
Time of day (Evening)	1.07468	0.21843	$4.70e-7$
Time of day (PM)	-0.06142	0.24083	0.8866

Table 8: Environmental Data Final Model

Variable	β Value	SE(β)	p-value
Survey Strategy Catch PUE/T Sample	-3.0812	0.9482	0.00116
Survey Strategy Single Catch Sample	-0.5225	0.4740	0.27029

when using Poisson distribution for the same dataset. The model produced for the 9 Angler Dataset using the Negative Binomial distribution has a lower AIC value (2889.8 compared to 3018.5 from the Poisson model) and therefore a higher associated log-likelihood, meaning that this model is preferred. Also, the β value corresponding to Year is, again, negative, further confirming the earlier point that Year is having a diminishing effect on the catch rate of barbel.

It can be seen in Table 6 that all estimated coefficients are negative, barring one. This coefficient relates to the temperature variable and equates to an increase of a factor of 1.07 in the number of barbel caught per hour for each Celsius increase in the temperature. When looking for a temporal trend, it is evident that the estimated coefficient corresponding to the Year variable is negative, further confirming the decreasing trend that has been found previously.

For the 6 Angler Model, in Table 7, it can be seen that the p-values are $>$ 0.05 for the corresponding coefficients for Temperature and all of the anglers, barring Group Angler 3. As only one out of 5 coefficients for the angler variables was significant, it was decided that this variable should be left out of the final model entirely. A likelihood ratio test was performed to further confirm the omission of the coefficient for the Temperature variable. The p-value obtained from this test was $0.6672 > 0.05$ indicating that the smaller model outperforms the larger model in terms of fit, and therefore this is the model that should be used. Setting the coefficients relating to temperature and angler = 0 gives the final model, 5.

Again, the AIC value for the 6 Angler Model was significantly smaller than that of the corresponding Poisson model, indicating that the Negative Binomial model offers a better fit to the data.

For the 6 Angler Model, it can be seen that the β corresponding to Year is negative which is important as this pushes towards an answer to the main objective of this report (Table 7). The negative value indicates that the $(\frac{\text{Number of Barbel}}{\text{Effort}})$ is decreasing by a factor of $e^{-0.23974}$. That is, for each additional year the number of barbel caught per hour fished is decreasing by a factor of 0.787. Another stand out β value is the coefficient corresponding to Time of day (Evening). This is the only positive β value and shows that fishing in the evening is having a very significant positive effect on $\log(\frac{\text{Number of Barbel}}{\text{Effort}})$. This result is further confirmed by Figure 3 where it can be seen that 'Evening' is having a large effect on the model. Therefore, fishing in the evening means that the number of barbel caught per hour increases by a factor of 2.93 in comparison to fishing in the AM.

Finally, for the Environmental Data, only one variable was statistically significant, leaving the model with 2 fewer terms than in the Poisson model. As can be seen in Table 8, Survey Strategy is the only variable left in the model, with p-values $>$ 0.05. Once more, a decrease in the AIC is evident, indicating that the Negative Binomial distribution produces a better model. Therefore, the Negative Binomial model was chosen as the final model.

When looking at the coefficients of the final environmental model, it can be seen that both survey strategy coefficients are negative (Table 8). The largest coefficient is the one associated with the PUE/T sample method, at $\hat{\beta} = -3.0812$. This suggests that using the PUE/T method decreases the number of barbel caught per area fished by a factor of 0.046, compared to when using the depletion method.

Conclusion

Working with 2 different datasets has allowed the exploration of different methods and ways of modelling data. Due to the dissimilarities in the data recorded, well-informed viewpoints can be formed on the disadvantages and advantages of each set of data. The angler Catch Data is a significantly larger dataset than the Environmental Data meaning that it was possible to gather more accurate results.

Altogether, the results found in each model suggest that the barbel fish population of the River Teme is in decline. While many covariates influence this, as can be seen by the many terms in the final models, it is hard to draw any meaningful conclusions about individual covariates from the previous analysis in this report. However, it can be said that the temperature covariate suggests conflicting conclusions for the 9 angler and the 6 angler models. In the final 9 angler model it is significant, causing an increase in number of barbel caught per hour, whereas in the 6 angler model the variable is not significant enough to be included in the model. However, the evidence provided in this paper further supports the growing body of evidence that fish populations nationwide for Barbel are still in decline.

Future Work

Given more time to work on this project, a Negative Binomial model with a spacial component could be implemented, similar to the one used by Alexander N. *et al* in 2000 [13] in order to look at how the use of different sites truly effects the number of barbel caught. Also, a Hurdle model could be implemented to possibly account for the excess of zeros in the data [14].

Acknowledgments

The authors would like to acknowledge Dr. Michele Lundy for her guidance on the project, Dr. Catherine Roberts for the research behind it, and Dr. Julian Stander for supplying all the material needed.

References

[1] Cooke, S.J., Fulton, E.A., Sauer, W.H.H. et al. (2023) 'Towards vibrant fish populations and sustainable fisheries that benefit all: learning from the last 30 years to inform the next 30 years.', *Rev Fish Biol Fisheries*, 33, Pp. 317–347. https://doi.org/10.1007/s11160-023-09765-8

[2] Environment Agency (2019), *Study to look at barbel and chub numbers.*. Available at: https://www.gov.uk/government/news/study-to-look-at-barbel-and-chub-numbers (Accessed: 2 May 2023)

[3] L. Vilizzi, G.H. Copp and J.R. Britton (2013). 'Age and growth of European barbel Barbus barbus (Cyprinidae) in the small, mesotrophic River Lee and relative to other populations in England' *Knowl. Managt. Aquatic Ecosyst.*, 409 (2013) 09 https://doi.org/10.1051/kmae/2013054

[4] Zieba, G., Stakėnas, S., Ives, M. J., Godard, M., Seymour, J., Carter, M. G. et al (2014). 'Long-term decline of barbel barbus barbus in the original course of the lower river lee (england), with particular reference to the survival of tagged fish during a water pollution incident.' *Fundamental and Applied Limnology*, 185(1), 43-53. https://doi.org/10.1127/fal/2014/0542

[5] Panchan, R., Pinter, K., Schmutz, S. et al. (2022 'Seasonal migration and habitat use of adult barbel (Barbus barbus) and nase (Chondrostoma nasus) along a river stretch of the Austrian Danube River.' *Environ Biol Fish*, 105, 1601–1616. https://doi.org/10.1007/s10641-022-01352-3

[6] Beaumont W. R. C., Taylor A. A. L., Lee M. J., Welton J. S. (2002) 'Guidelines for Electric Fishing Best Practice', *RD Technical Report*, W2-054/TR

[7] Nelder J. A., Wedderburn R. W. M. (1972) 'Generalized Linear Models' *Journal of the Royal Statistical Society, Series A (General)*, 135(3), Pp. 370-384

[8] R Core Team (2023) R: A Language and Environment for Statistical Computing, *R Foundation for Statistical Computing, Vienna, Austria*, https://www.R-project.org/

[9] Eales, J. (2023) 'Epidemiology', *MATH3614 Medical Statistics (University of Plymouth)*

[10] Malek-Ahmadi M. (2020) *'Regression Analysis with Right- Skewed Data: Applications for Pre-Clinical Alzheimer's Disease'* https://files.alz.washington.edu/presentations/2020/fall/Malek-Ahmadi.pdf (Accessed: 16 May 2023)

[11] Ver Hoef J. M., Boveng P. L. (2007) 'Quassi-Poisson vs Negative Binomial Regression: How Should We Model Overdispersed Count Data?', *Ecology*, 88(11), Pp. 2679-2950

[12] Kleiber C., Zeileis A. (2008) Applied Econometrics with R, *Springer-Verlag*, https://CRAN.R-project.org/package=AER

[13] Alexander N., Stander J., Moyeed R. (2000) 'Spatial modelling of individual-level parasite counts using the negative binomial distribution', *Biostatistics*, 1(4), Pp. 453-463

[14] Cragg, John (1971) 'Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods.', *Econometrica*, 39(5), Pp. 829-844

Appendices are provided separately as supplementary files (see additional downloads for this article).