UNIVERSITY OF WISCONSIN-LA CROSSE

Graduate Studies

A COMPARISON OF STATISTICAL MODELING

APPROACHES FOR HIERARCHICAL DATA: A SIMULATION

STUDY AND APPLICATION TO ECOLOGY DATA

A Chapter Style Project Submitted in Partial Fulfillment of requirements for the
Degree of Master of Science

Jillian L. Hebert

College of Science and Health
Applied Statistics

May 2021

A COMPARISON OF STATISTICAL MODELING

APPROACHES FOR HIERARCHICAL DATA: A SIMULATION

STUDY AND APPLICATION TO ECOLOGY DATA

By Jillian L. Hebert

We recommend acceptance of this project in partial fulfillment of the candidate's requirements for the degree of Master of Science in Applied Statistics.

The candidate has completed the oral defense of the project.

_____   _____

Barbara Bennie, Ph.D.   Date
Project Committee Chairperson

On behalf of the committee members named below:

Richard Erickson, Ph.D.

Aaron Cupp, M.S.

Project accepted

_____   _____

Meredith Thomsen, Ph.D.   Date
Director of Graduate Studies

ABSTRACT

Hebert, Jillian. <u>A comparison of statistical modeling approaches for hierarchical data: A simulation study and application to ecology data</u>. MS in Applied Statistics, May 2021, 47pp. (B. Bennie)

Variable specification, declaring variables as fixed or random, and model selection are intricate parts of many statistical analyses. This project aims to investigate the differences in four model's results when variable specification and model selection differ for hierarchical data. The first model pooled over the random effects and fit the fixed. The second model specified all variables as fixed. The third model specified variables appropriately as fixed or random and used a maximum likelihood algorithm. The fourth model specified variables appropriately and used a restricted maximum likelihood algorithm. Two simulated datasets, one with no treatment effect and one with a treatment effect, were used to make comparisons about model estimate accuracy and power. Two empirical datasets were used to compare models, with special attention to the significance values and degrees of freedom reported. Results exemplified that the fixed model inappropriately handled variables and produced results that inaccurately reflected the data. The pooled and restricted maximum likelihood models had paralleling results due to the numerical methods but differed when repeated measures were included in the data structure. Finally, the maximum likelihood and restricted maximum likelihood models had negligible differences, but conclusions shifted when near the border of the statistical significance.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Section 1. Introduction

Variable specification, declaring variables as fixed or random, and model selection are intricate parts of many statistical analyses. These choices can have drastic effects on model performance and subsequent inferences. This is a vital but difficult part of the modeling process and uninformed choices can lead to results that do not accurately reflect the context of the data. To prevent conclusions being drawn based off erroneous statistical models, companies and individuals with less statistical knowledge must ensure that their model selection and variable specifications are correct. Many individuals applying statistical methods to their field of study have received heavy critiques or rejected papers because of their variable specification or model choice, but how significant are those choices? Should journal reviewers reject submitted papers if the only major disagreements are over the variable specification or model selection – even if the full consequences of these choices are unknown to both the reviewer and the submitting author?

1.1. Background

Mixed effects models are a common model type that include fixed and random effect terms; mixed effects models allow for modeling and estimation of random effects. Fixed effects are comparing means of levels while random effects are variations across levels. Mixed effects models can be exemplified through a hierarchical design. Hierarchical designs are seen in many fields of study, especially the sciences, but a specific case of a hierarchical design is the nested design. A nested design refers to a hierarchical model with an experiment structure where the treatment and experimental units are sub-sampled. An example of a nested design would be a researcher having six tanks with fifty fish in each tank; each tank receives a treatment level, and

a measurement is recorded from five fish in every tank. In this design, the tank would be the random effect with unique fish nested in tank. These types of nested designs can be tricky when it comes to variable specification because the models must account for the random effects that occur naturally within the design structure.

When variables are appropriately specified, how much of an impact do random effects have on the results? This project aims to investigate the differences in results when variable specification and model selection differ. Using simulated and empirical datasets, four different modeling methods will be compared to evaluate the major differences in results.

1.2. Pooled Modeling

Pooled modeling is an extreme application of complete data pooling. When pooling data, all subjects within a study will be averaged – or pooled – together, creating one identical measurement for every experimental unit, ignoring the random effects. Referencing the nested example previously mentioned, instead of having five unique measurements from every tank, those five measurements would be averaged together so each tank only had one pooled measurement. The major drawback to this modeling approach is the loss of information, or variation, within the random effect.

1.3. Fixed Effect Modeling

Fixed effect modeling is a type of modeling that purposefully excludes all randomization. Instead of specifying variables as fixed or random, all variables are specified as fixed in the model. Referencing the nested example, instead of specifying that the tank variable is random, tank will be included as a fixed effect variable. A major drawback to this modeling approach is

the artificial inflation of sample size, which will be reflected in the degrees of freedom used for significance estimations. Declaring all variables as fixed effects will lead to unimportant results, using the degrees of freedom to fit differences in variables that do not address a study's objectives.

1.4.   Maximum Likelihood Modeling

Maximum likelihood (ML) modeling is a modeling method that estimates parameters by maximizing a likelihood function. This results in an estimation that has the highest probability based off the data. For fixed effects designs, the most common maximum likelihood estimates are least squares estimates – or estimates that minimize the sum of the squares of the residuals. For mixed effect designs, designs that include fixed and random effects, the likelihood function is commonly based off the numerical log likelihood. The drawback to maximum likelihood estimates is the biased variance estimator (Oehlert, 2012).

1.5.   Restricted Maximum Likelihood Modeling

Restricted maximum likelihood (REML) modeling is a modeling method that estimates parameters by maximizing a likelihood function. This results in an estimation that has the highest probability based off the data. For fixed effects designs, the most common maximum likelihood estimates are least squares estimates – or estimates that minimize the sum of the squares of the residuals. For mixed effect designs, designs that include fixed and random effects, the likelihood function is commonly based on the best linear unbiased prediction (BLUP), allowing for estimations of random effects (Cesarani et. al, 2018). Restricted maximum likelihood modeling

is a way to produce unbiased estimates of variance components which is why it is most often

associated with linear mixed models (Oehlert, 2012).

Section 2. Methods

Two simulated and two empirical datasets were used to assess the impacts of variable specification and model selection.

2.1. Comparison Procedure

Two simulated datasets that emulate nested designs were used to make comparisons between each model. Comparisons were also made between the two simulated datasets; one dataset had no shift between treatment means, or no treatment effect, while the other simulated dataset had a shift of half the standard deviation between treatment means, creating a treatment effect. Investigations expanded to the impacts of four conditions – between subject standard deviation, within subject standard deviation, number of subjects, and number of repetitions – on the relative power of each model. Model's effect size, or confidence intervals, were also investigated to identify if certain models had tighter intervals and under what conditions.

Two empirical datasets were used to make comparisons on computed estimates, significance values (p-values), and confidence intervals. Special attention was paid to the degrees of freedom reported by each model because the values represent whether the correct level of replication was used to estimate the error variance. If the denominator degrees of freedom were unexpectedly large, that could indicate inappropriate variable specification or model selection, which would artificially amplify statistical results that were not truly supported by the data.

A linear mixed effects model was used as the base model for all analysis. Different variations of the linear mixed effects model was used to accommodate for the different variable specifications. Using the same base model allowed for easier comparisons between model

results. All analysis was done in R and RStudio using a significance level of $\alpha = 0.05$ for all tests (R Core, 2020).

2.2. Simulated Data

The simulated datasets being used were identical, except for the treatment shift. One simulation had no treatment effect; the other simulation had a treatment effect of half the standard deviation. The data's structure was generated from a gamma distribution that emulated a nested design, allowing for variations in the number of subjects, the number of repetitions, the between subject standard deviation, and the within subject standard deviation. The gamma distribution allowed for the addition of skewness which matches the structure of the empirical datasets. By changing one condition at a time, comparisons were made between size structure and variance structure. Size structure could vary as balanced, having a larger number of subjects, or having a larger number of repetitions. Variance structure could vary as balanced, having a larger standard deviation between subjects, or having a larger standard deviation within subjects. Nine unique combinations of size structure and variance structure were used to make comparisons for each model method.

Table 1: Nine Cases of Unique Condition Combinations

Case	Alpha	Between Subject SD	Within Subject SD	Number of Subjects	Number of Repetitions	Variance Structure	Sample Size Structure
1	0.05	2	5	10	10	Large Within	Balanced
2	0.05	5	2	10	10	Large Between	Balanced
3	0.05	3	3	10	10	Balanced	Balanced
4	0.05	2	5	5	20	Large Within	Large Repetition
5	0.05	5	2	5	20	Large Between	Large Repetition
6	0.05	3	3	5	20	Balanced	Large Repetition
7	0.05	2	5	20	5	Large Within	Large Subject
8	0.05	5	2	20	5	Large Between	Large Subject
9	0.05	3	3	20	5	Balanced	Large Subject

Once the size and variance structure were specified, each model was fit to the nested

structure created, and the significance value for the treatment effect was pulled. The lower and

upper 95% confidence interval values for estimating treatment effect were also isolated for each

model. This modeling was repeated 1,000 times, creating a table with the significance value,

lower confidence interval value, and upper confidence interval value for each model method for

every repetition. To calculate the power of each model method, the significance value was

identified as significant or not (using a significance level of $\alpha = 0.05$), and the number of

significant treatment values were divided by the total number of repetitions ($Power =$

$\frac{sum(significant\ treatment)}{total\ repetitions}$). The 1,000 recorded lower and upper 95% confidence intervals were

averaged to identify the overall lower and upper value for each model and the center of the

interval was used for the estimate. The widths of the confidence intervals were compared by

model type, and special attention was paid to the appropriateness of the models estimate.

2.3.   Exposure Related Effects of Zequanox

This nested empirical dataset was obtained from U.S. Geological Survey and analyzed

the impacts of different levels of Zequanox (0 mg/L, 50 mg/L, and 100 mg/L) on juvenile Lake

Sturgeon and Lake Trout. Zequanox is a chemical applied to water to control invasive muscles,

but the impacts to fish were evaluated to determine potential non-target effects on native species.

There were twelve tanks throughout the experiment with 100 fish in each tank. Due to mortality,

certain tanks had less than 100 fish by the conclusion of the study. All remaining fish had their

length (mm) and weight (g) recorded. A condition factor was created using the subject's length

and weight and was multiplied by a factor of 100,000 to allow for proper model convergence

($\frac{weight}{length^3} * 100{,}000$). This condition factor was used as the response variable to assess the impacts

of Zequanox (Luoma et. al, 2018). There were 2,203 observations with seven variables: tank number, unique fish number, species, treatment level, weight, length, and condition factor.

For the pooled model, a pooled condition factor was created by averaging the condition factor of all the fish in each tank by species. This created one response for every tank and species and was used as the response variable in the linear model (lm). Treatment level and species were used as predictor variables and were specified as fixed effects in the linear model.

For the fixed effects model, all predictor variables were specified as fixed effects in the model. A linear model (lm) was fit with the original condition factor as the response variable with treatment level, species, and tank as predictor variables.

For the maximum likelihood (ML) model, a linear mixed effects model (lme) was fit with the original condition factor as the response variable with treatment level and species as fixed effects. The model also included tank number and unique fish number as random variables with specifications that unique fish number was nested in tank number. The maximum likelihood (ML) method was also specified within the model.

For the restricted maximum likelihood (REML) model, a linear mixed effects model (lme) was fit with the original condition factor as the response variable with treatment level and species as fixed effects. The model also included tank number and unique fish number as random variables with specifications that unique fish number was nested in tank number. The restricted maximum likelihood (REML) method was also specified within the model.

2.4. Acoustic Telemetry Evaluation of Carbon Dioxide

This nested experimental design was also obtained from the U.G. Geological Survey and focused on the impacts of different levels of CO_2 (low, medium, or high) from different pond

locations (east or west side) on fish movements at two different periods (acclimation and treatment). Nine 48-hour trials (24-hours acclimation and 24-hours treatment) were conducted, monitoring the reactions of ten fish (five Bighead Carp and five Grass Carp) in a large U-shaped outdoor tank. The location of each unique fish was monitored throughout the 48-hour period in relation to the reference point at the center of the pond (Cupp et. al, 2020). The original dataset contained the movements of each fish every three seconds which resulted in an extremely large dataset of 2,567,428 observations.

To make movement comparisons easier, an absolute distance from the reference point was created as opposed to the directed distance originally available in the dataset. To retain the movement within the tank, a crossing variable was used to monitor when the fish crossed the reference point from one side of the tank to the other. Using the R package *adehabitatLT* (Calenge, 2006), the absolute distance was calculated for each fish in the tank every three seconds. To condense the dataset, the data was grouped by date, trial, period (acclimation or treatment), and unique fish, and the absolute distance was averaged over the hour, so each fish had one absolute distance measurement per hour, resulting in 3,697 observations. The total crossings were maintained by taking the summation of crosses over the hour for each fish.

To parallel the nested structure of the simulated datasets and the Exposure Related Effects of Zequanox dataset, the two species were separated and handled separately to maintain a nested design. If the two species were handled simultaneously, the dataset would be a cross-nested design, adding an extra layer that did not exist in the other datasets. It is important to note that this empirical dataset also included a repeated measure, where each fish was measured twice, which was not in the design of the other datasets. Once the species were separated, the absolute distance was averaged over period to also maintain the same structure as the other three

9

datasets, resulting in the final two datasets of 86 observations with six variables: trial, unique fish number, period, treatment level, treatment side, and absolute distance.

For the pooled model, a pooled absolute distance was created by averaging the absolute distance of all the fish in each tank. This created one response for every tank and was used as the response variable in the linear model (lm). Period, level, and treatment side were used as predictor variables and were specified as fixed effects in the linear model.

For the fixed effects model, all predictor variables were specified as fixed effects in the model. A linear model (lm) was fit with the original absolute distance as the response variable with period, level, and their interaction as predictor variables. Trial nested in level was also used as a fixed predictor to include trial as a fixed effect. Since there was only one treatment side and one level per trial, level was nested in trial to properly specify the data structure.

For the maximum likelihood (ML) model, a linear mixed effects model (lme) was fit with the original absolute distance as the response variable and period, level, and treatment side as fixed effects. The model also included the interaction between period and level as a fixed effect, as well as unique fish code nested in trial as random effects. The maximum likelihood (ML) method was also specified within the model.

For the restricted maximum likelihood (REML) model, a linear mixed effects model (lme) was fit with the original condition factor as the response variable and period, level, and treatment side as fixed effects. The model also included the interaction between period and level as a fixed effect, as well as unique fish code nested in trial as random effects. The restricted maximum likelihood (REML) method was also specified within the model.

Section 3. Results

The results from the four datasets gave insight to how the model's results differed when variables were specified differently. Differences in models with varying data structures were also highlighted to give more clarity to which conditions models perform best under.

3.1. Simulated Data

The simulated datasets were used to highlight differences between models based on certain dataset conditions. Before reviewing the results, it is important to note the drawbacks of the simulated datasets. Both simulated datasets had balanced sample sizes across treatments and experimental units. Furthermore, the simulations were based off a fixed skewness of 0.5 from the gamma distribution and does not account for changes in skewness.

3.1.1. No Treatment Effect

The first simulated dataset had no treatment effect. Each model had 1,000 significance values that were used to calculate the power for each model. Table 2 shows the power by model, which should be close to 0.05 since there was no treatment effect in the simulated dataset. Each row represents one of the nine cases with different specified conditions. These conditions are the between subject standard deviation, the within subject standard deviation, number of subjects, and number of repetitions.

Table 2: Model Powers from No Treatment Effect Simulate Dataset

Case	Alpha	Between Subject SD	Within Subject SD	Number of Subjects	Number of Repetitions	Pooled Power	Fixed Power	ML Power	REML Power
1	0.05	2	5	10	10	0.056	0.778	0.065	0.056
2	0.05	5	2	10	10	0.060	0.907	0.076	0.060
3	0.05	3	3	10	10	0.053	0.864	0.066	0.053
4	0.05	2	5	5	20	0.050	0.831	0.076	0.050
5	0.05	5	2	5	20	0.050	0.929	0.072	0.050
6	0.05	3	3	5	20	0.053	0.900	0.077	0.053
7	0.05	2	5	20	5	0.050	0.698	0.054	0.050
8	0.05	5	2	20	5	0.048	0.842	0.051	0.048
9	0.05	3	3	20	5	0.041	0.757	0.043	0.041

The differences in model powers by condition can be seen in the table, but overall, the models continue to keep a power between 0.043 and 0.82 apart from the fixed model power. Due to the artificial inflation of sample size in the fixed model, the number of significant values out of the 1,000 simulations is much larger than the other three models, leading to a higher power for each case. Since there is no treatment effect in this dataset, the high powers for the fixed effect model highlights the fact that the fixed model is determining the treatment effect as significant at a higher rate than what is expected as well as what the dataset truly supports.

Confidence intervals for each model for each case were also calculated. There is no treatment effect in this dataset, therefore, the estimates for each interval should be centered at zero. Figure 1 shows the intervals for each model, allowing for visual comparisons of the interval widths. From the figure, it can be seen that the interval for the fixed model always had the shortest width. All intervals for every model and case were centered close to zero, demonstrating that all models properly identified the correct treatment estimate; the fixed model tended to be

the most inaccurate, with an interval centered farther away from zero than the other three
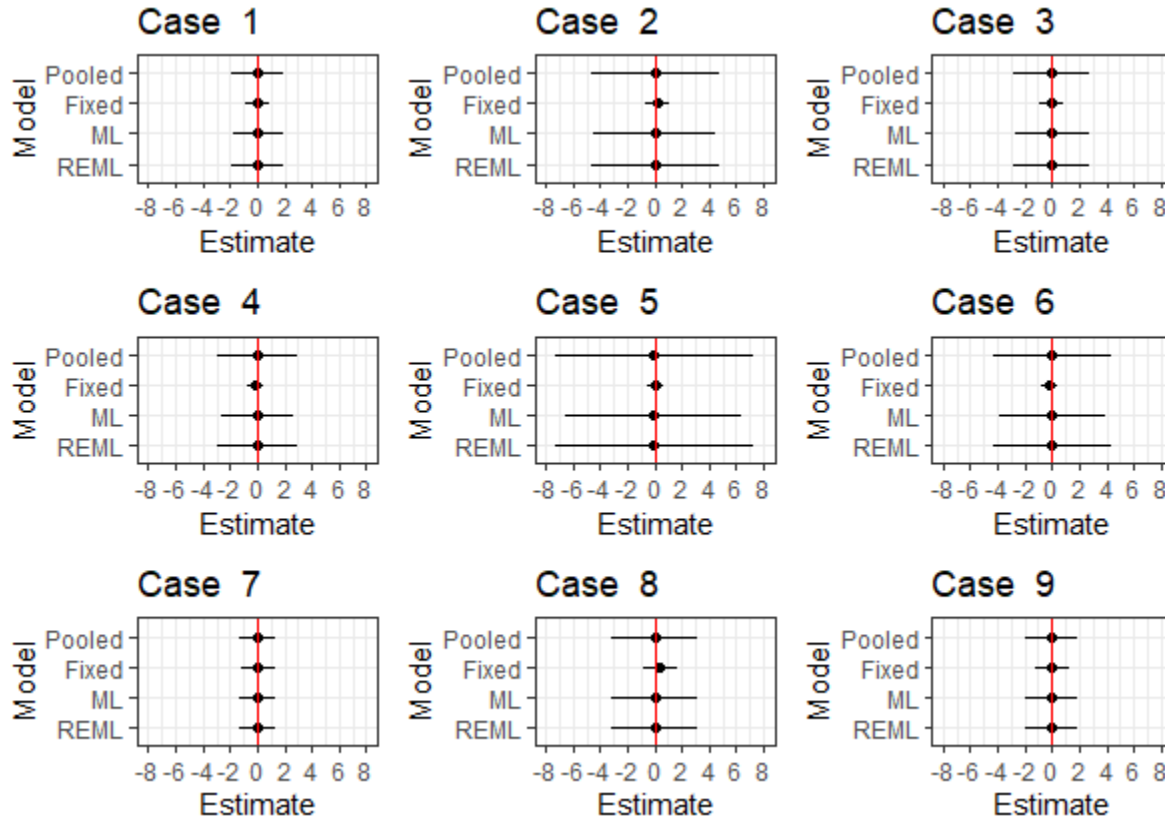
models.

Figure 1: 95% Confidence Intervals for No Treatment Effect Simulated Dataset

3.1.2. Treatment Effect

The second simulated dataset had a treatment effect of half the standard deviation which

equated to a difference of three. Each model had 1,000 simulated significance values that were

used to estimate the power for each model. Table 3 shows the power by model. Each row

represents one of the nine cases with different specified conditions. These conditions are the

between subject standard deviation, the within subject standard deviation, number of subjects,

and number of repetitions.

Table 3: Model Powers from Treatment Effect Simulate Dataset

Case	Alpha	Between Subject SD	Within Subject SD	Number of Subjects	Number of Repetitions	Pooled Power	Fixed Power	ML Power	REML Power
1	0.05	2	5	10	10	0.887	0.858	0.902	0.887
2	0.05	5	2	10	10	0.221	0.894	0.254	0.221
3	0.05	3	3	10	10	0.513	0.871	0.566	0.513
4	0.05	2	5	5	20	0.451	0.907	0.541	0.451
5	0.05	5	2	5	20	0.112	0.927	0.144	0.112
6	0.05	3	3	5	20	0.241	0.905	0.299	0.241
7	0.05	2	5	20	5	0.997	0.806	0.997	0.997
8	0.05	5	2	20	5	0.456	0.874	0.473	0.456
9	0.05	3	3	20	5	0.855	0.831	0.862	0.855

The differences in model powers by criteria can be seen in the table, with higher powers representing more significant values out of the 1,000 simulations. Models with higher powers indicate better performance. Table 3 shows that a model's power will deviate depending on the conditions specified. Apart from the fixed model, the powers for each model are relatively similar, with the maximum likelihood model having slightly higher powers than the pooled and restricted maximum likelihood models.

Confidence intervals for each model for each case were also calculated. Since there is a treatment effect, the estimates for each interval should now be centered at three. Figure 2 shows the intervals for each model, allowing for visual comparisons of the interval widths. From the figure, it can be seen that the interval for the fixed model still had the shortest width. All intervals for every model and case were centered near three, showing that all four models did well in identifying the correct treatment estimate; the fixed model tended to be the most inaccurate, with an interval centered farther away from three than the other models for specific cases.

14
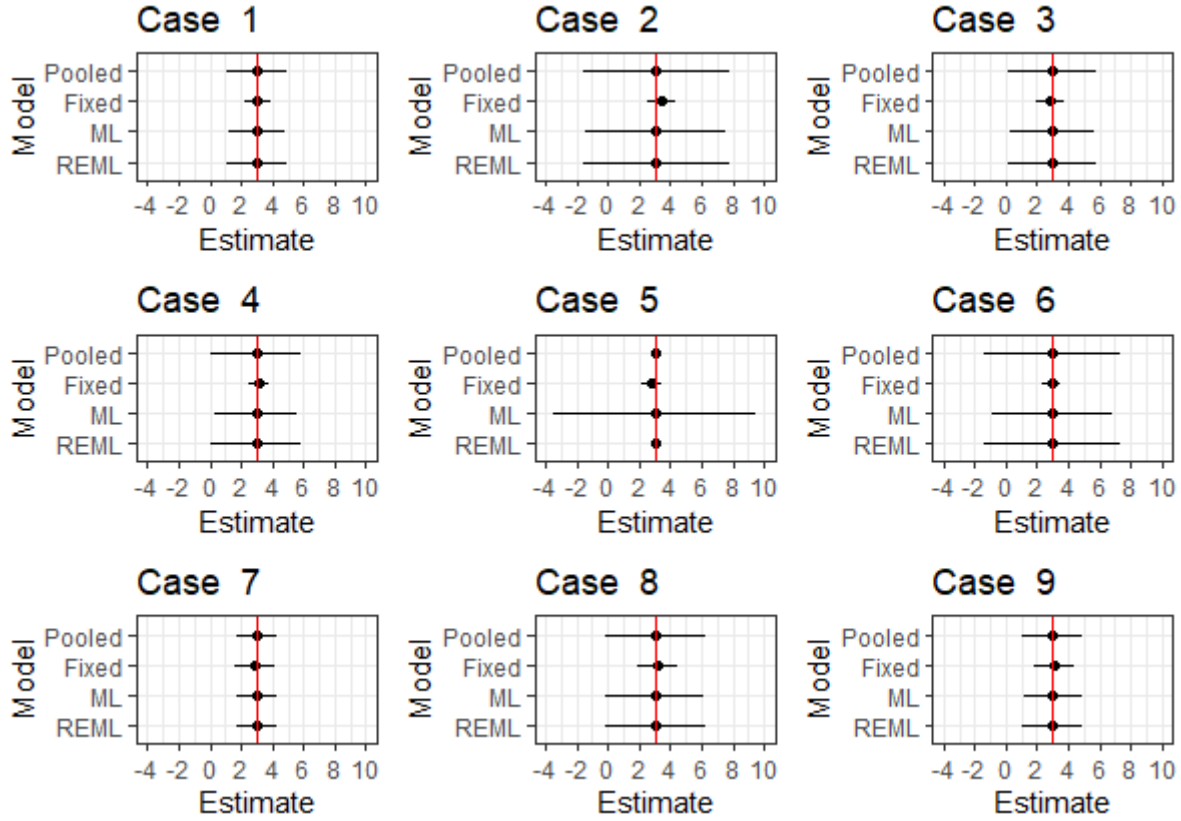
Figure 2: 95% Confidence Intervals for Treatment Effect

3.2. Exposure Related Effects of Zequanox

This empirical dataset was used to highlight differences in model results based on variable specification. Before reviewing the results, it is important to note that this empirical dataset had unbalanced sample sizes by tank and by species.

3.2.1. Descriptive Statistics

The Zequanox dataset had 2,203 observations with seven variables: tank number, unique fish number, species, treatment level, weight, length, and condition factor. The minimum, maximum, and average values by tank and species are shown in table 4. The average length and

weight of lake sturgeon is greater than lake trout, but the condition factor for lake trout is greater than the condition factor for lake sturgeon. This is due to difference in the relationship between length and weight for each species.

When comparing the condition factors by treatment and species, it can be seen from figure 3 that lake trout has more variation in condition factor than lake sturgeon as well as having higher condition factor values. The treatment differences also had more variance in lake trout than the differences in treatment for lake sturgeon.

Figure 3: Condition Factory by Treatment and Species

Note that the condition factor axis is different for each species.

Table 4: Summary Statistics

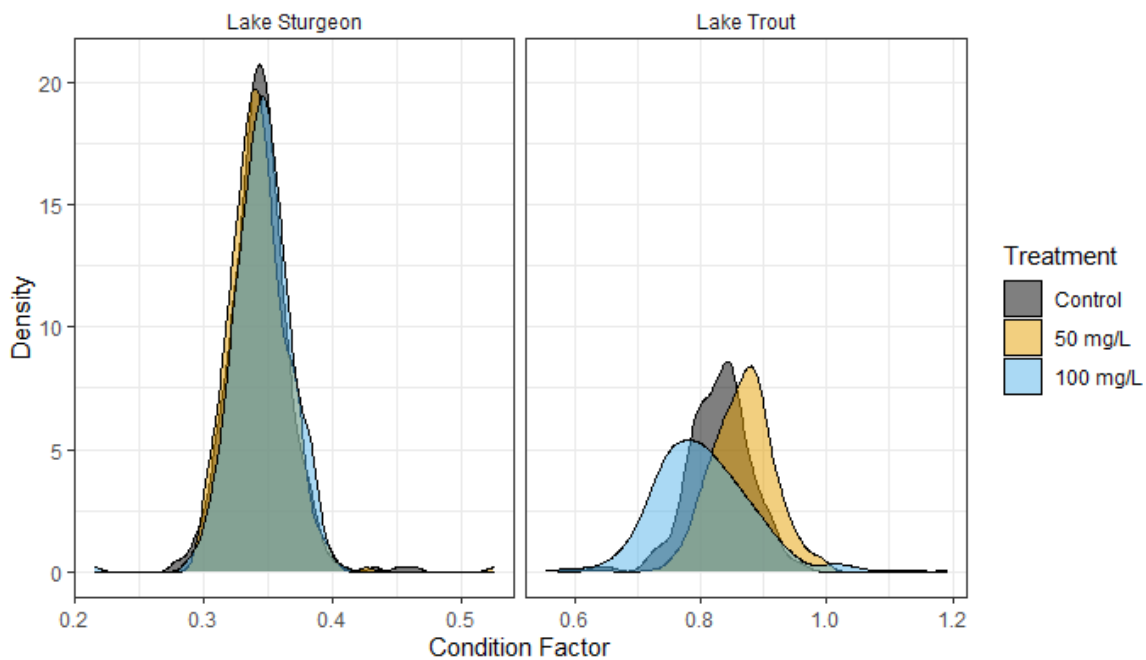| Tank	Species	Minimum Weight	Average Weight	Maximum Weight	Minimum Length	Average Lenght	Maximum Length	Minimum Condition Factor	Average Conition Factor	Maximum Condition Factor	Count
1	Lake Trout	3.20	6.2872	10.16	75	91.5000	104	0.670	0.8022	0.944	50
1	Lake Sturgeon	14.07	21.7716	32.00	145	184.0800	209	0.278	0.3462	0.463	100
2	Lake Trout	5.94	9.8469	19.05	94	105.0400	128	0.715	0.8366	0.969	100
2	Lake Sturgeon	11.02	20.3645	32.20	151	180.8300	201	0.293	0.3411	0.399	100
3	Lake Trout	3.93	6.0560	9.80	82	90.2400	104	0.687	0.8142	1.130	50
3	Lake Sturgeon	12.74	20.8258	29.55	152	183.0800	206	0.303	0.3372	0.388	100
4	Lake Trout	5.49	8.8747	14.09	87	100.5000	113	0.764	0.8629	1.000	100
4	Lake Sturgeon	5.49	19.5513	30.32	117	176.9000	211	0.301	0.3468	0.392	100
5	Lake Trout	5.26	9.0260	13.47	87	103.0600	117	0.631	0.8141	0.954	100
5	Lake Sturgeon	12.06	20.7352	34.04	152	181.4900	209	0.301	0.3435	0.399	100
6	Lake Trout	3.27	6.0415	9.93	75	90.5185	104	0.583	0.7983	1.030	54
6	Lake Sturgeon	13.05	21.7574	36.93	158	184.4100	223	0.292	0.3436	0.432	100
7	Lake Trout	3.97	9.2675	15.13	82	102.4200	119	0.569	0.8499	1.190	100
7	Lake Sturgeon	9.97	20.7803	30.07	140	180.6211	205	0.215	0.3484	0.408	95
8	Lake Trout	5.81	8.8383	14.15	88	100.1212	116	0.766	0.8694	0.990	99
8	Lake Sturgeon	7.49	20.5356	30.60	129	180.3061	202	0.299	0.3465	0.388	98
9	Lake Trout	4.48	9.1132	16.18	85	103.4343	123	0.555	0.8098	0.923	99
9	Lake Sturgeon	11.02	20.3974	31.28	128	180.1500	208	0.301	0.3463	0.525	100
10	Lake Trout	6.08	10.2105	18.96	90	105.2600	129	0.782	0.8589	0.957	100
10	Lake Sturgeon	13.37	21.6723	32.56	151	184.3900	212	0.279	0.3427	0.391	100
11	Lake Trout	4.88	8.5393	13.35	84	98.3838	113	0.756	0.8821	0.996	99
11	Lake Sturgeon	8.19	19.8497	30.66	135	179.3500	205	0.292	0.3394	0.384	100
12	Lake Trout	3.30	6.3403	10.32	75	91.8644	107	0.610	0.8036	1.050	59
12	Lake Sturgeon	11.39	20.8653	30.60	150	182.8500	207	0.290	0.3389	0.451	100

3.2.2. Pooled Modeling

The pooled model used a pooled condition factor as the response variable. The pooled condition was created by averaging the condition factor of all the fish in each tank by species. The linear model (lm) output is shown in table 5 when using treatment level and species as fixed predictor variables.

Table 5: Pooled Model Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.8314	0.0074	0.8160	0.8468	112.4413	<0.0001
Treatment50 mg/L	0.0177	0.0091	-0.0012	0.0366	1.9552	0.0647
Treatment100 mg/L	-0.0114	0.0091	-0.0303	0.0075	-1.2596	0.2223
SpeciesLST	-0.4901	0.0074	-0.5055	-0.4747	-66.2862	<0.0001

The intercept is the control treatment and lake trout species. LST stands for lake sturgeon.

Using a significance value of $\alpha = 0.05$, there was a significant difference in the pooled condition factor estimate when moving from one species to another, with a lower pooled condition factor estimate for lake sturgeon than for lake trout. The pooled model determined that were no significant differences when moving from one treatment level to the next.

3.2.3. Fixed Effect Modeling

The fixed effects model specified all predictor variables as fixed effects. Treatment level, species, and tank were used as predictor variables with the original condition factor as the response variable in the linear model (lm). The output from the model is shown in table 6.

Table 6: Fixed Model Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.8098	0.0043	0.8014	0.8182	189.5437	<0.0001
Treatment50 mg/L	-0.0078	0.0095	-0.0265	0.0109	-0.8190	0.4129
Treatment100 mg/L	-0.0076	0.0074	-0.0221	0.0069	-1.0288	0.3037
SpeciesLST	-0.4557	0.0074	-0.4702	-0.4412	-61.6828	<0.0001

The intercept is the control treatment and lake trout species. LST stands for lake surgeon.
Not all output is shown in the table.

Using a significance value of $\alpha = 0.05$, there was a significant difference in the estimated condition factor when moving from one species to another, with a lower condition factor estimate for lake sturgeon than for lake trout. The complete model output was not included, but multiple rows continued to make comparisons between tanks. These outputs were excluded from the table since the main interest is in the significance by treatment level and species.

3.2.4. Maximum Likelihood Modeling

The maximum likelihood (ML) model uses the original condition factor as the response variable in a linear mixed effects model (lme) with method specified as ML. Treatment level and species were used as fixed predictor variables; tank number and unique fish number were random variables with specifications that unique fish number was nested in tank number. The output from the maximum likelihood (ML) model is shown in table 7.

19

Table 7: ML Model Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.8315	0.0067	0.8184	0.8447	123.5276	<0.0001
Treatment50 mg/L	0.0177	0.0082	0.0005	0.0349	2.1495	0.0440
Treatment100 mg/L	-0.0110	0.0083	-0.0283	0.0063	-1.3240	0.2004
SpeciesLST	-0.4904	0.0068	-0.5045	-0.4763	-72.5871	<0.0001

The intercept is the control treatment and lake trout species. LST stands for lake sturgeon.

Using a significance value of $\alpha = 0.05$, there was a significant difference in the condition factor estimate when moving from one species to another. There was also a significant difference in condition factor when the treatment moved from control to 50 mg/L, but there were no significant differences in the condition factor estimate when moving from the control treatment to the 100 mg/L treatment. The condition factor estimate was lower for lake sturgeon than for lake trout, and the condition factor estimate increased when moving from the control treatment to the 50 mg/L treatment.

3.2.5. Restricted Maximum Likelihood Modeling

The restricted maximum likelihood (REML) model uses the original condition factor as the response variable in a linear mixed effects model (lme) with method specified as REML. Treatment level and species were used as fixed predictor variables; tank number and unique fish number were random variables with specifications that unique fish number was nested in tank number. The output from the restricted maximum likelihood (REML) model is shown in table 8.

Table 8: REML Model Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.8315	0.0074	0.8171	0.8460	112.7938	<0.0001
Treatment50 mg/L	0.0177	0.0090	-0.0011	0.0365	1.9625	0.0638
Treatment100 mg/L	-0.0111	0.0091	-0.0300	0.0079	-1.2179	0.2374
SpeciesLST	-0.4904	0.0074	-0.5058	-0.4749	-66.3148	<0.0001

The intercept is the control treatment and lake trout species. LST stands for lake sturgeon.

Using a significance value of α = 0.05, there was a significant difference in the condition factor estimate when moving from one species to another, with a lower condition factor estimate for lake sturgeon than for lake trout. The restricted maximum likelihood model determined that were no significant differences when moving from one treatment level to the next.

3.3. Acoustic Telemetry Evaluation of Carbon Dioxide

This empirical dataset was used to highlight differences in model results based on variable specification. Before reviewing the results, it is important to note that this empirical dataset had repeated measures since the same fish were measured once during the acclimation period and measured again during the treatment period.

3.3.1. Descriptive Statistics

The distribution of the original distance metric from the dataset was graphically explored to evaluate the differences by period and species. Negative distances represent movement to the west of the reference point in the center of the pond, while positive distances represent movement to the east. Figure 4 shows the distributions by trial, and figure 5 shows the densities

by level. Figure 4 shows that during the acclimation period there was a larger spread over the

distance metric, meaning fish were moving more freely throughout the pond.
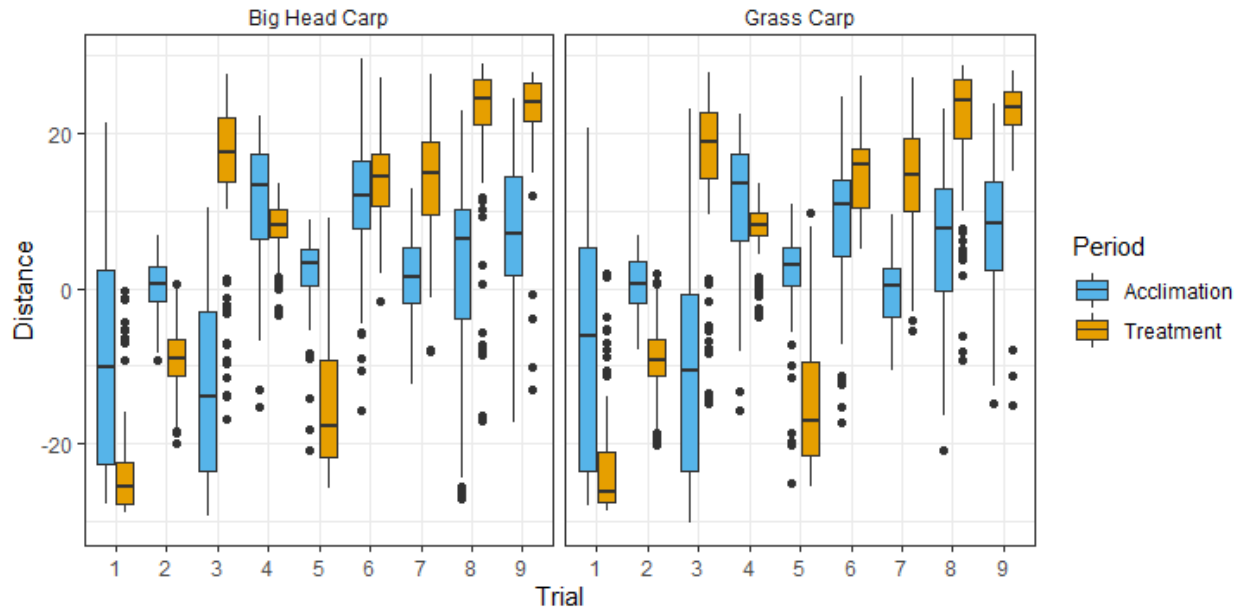
Figure 4: Spread of Distance by Period and Species for Each Trial

Figure 5 also supports the conclusion that fish are moving more freely throughout the

pond during the acclimation period, but when the CO_2 was coming from the east side, there was

less movement from fish in the low and medium levels. When looking at the treatment period,

the movement of the fish paralleled what was expected, with them moving towards the opposite

side of the tank from where the CO_2 was being released. What was unexpected was the spread of

the fish during the medium level. When the level was high and low, the fish had less spread, but

the medium level had a significantly wider spread across the tank during the treatment period.
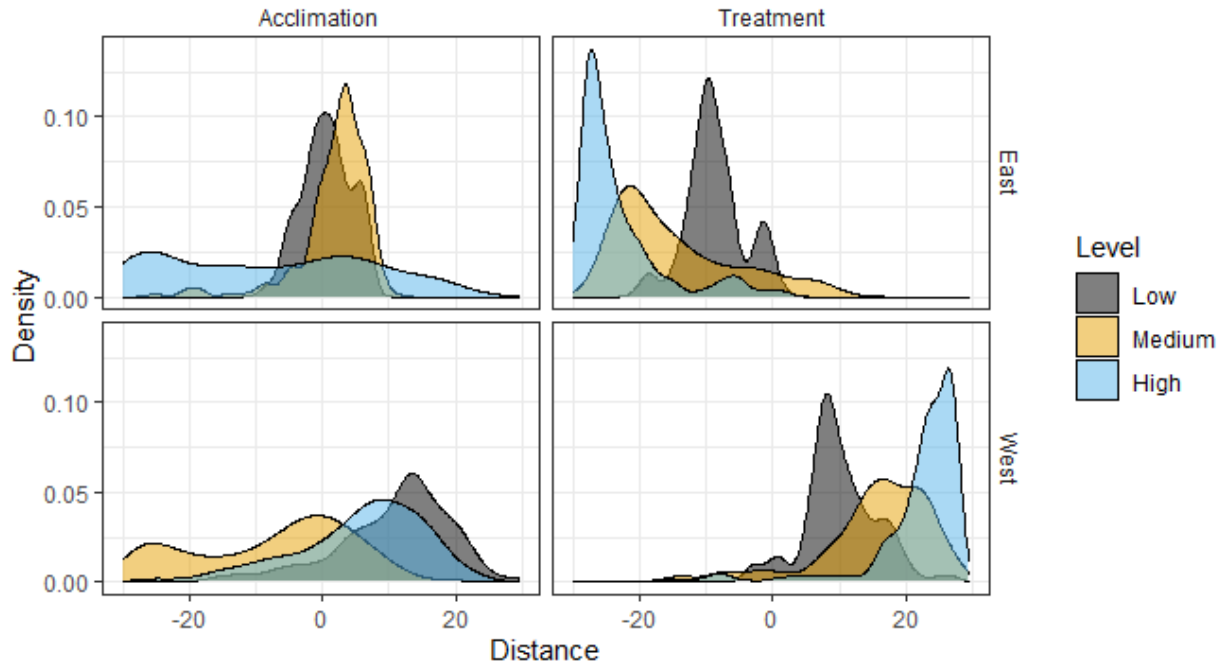
Figure 5: Distribution of Original Distance Metric

After applying the *adehabitatLT* package to the dataset to create an absolute distance, figure 5 was recreated. Figure 6 shows the absolute distance making visual comparisons by treatment side easier. With the absolute distance metric, the variation in distance during the medium level is still the greatest, but when the CO_2 treatment is coming from the east side, the high level has less variation in distance than when the CO_2 treatment is coming from the west side. Figure 6 also highlights the large differences in distribution by level during the acclimation treatment. There was very little movement during the medium level with a peak much taller than the low and high levels for the east side.
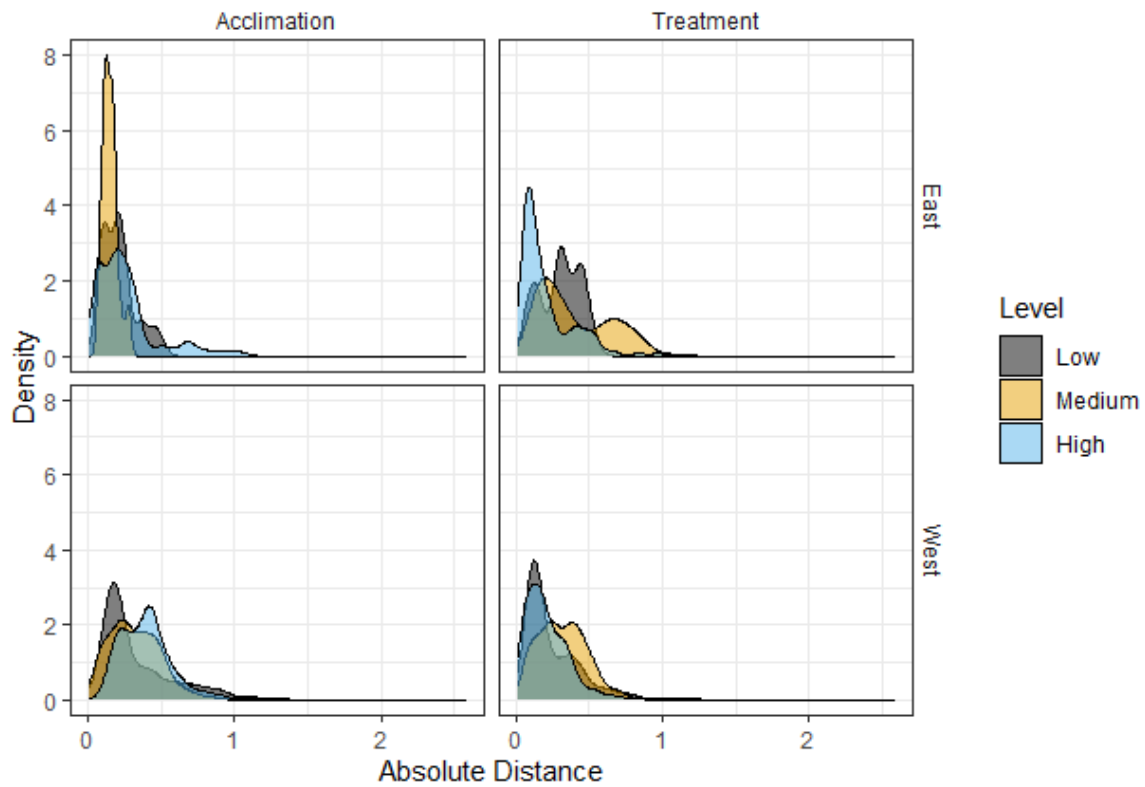
Figure 6: Distribution of Absolute Distance from *adehabitatLT* Package

A visualization of the total number of crossings in the pond was also created. The crossings were determined as every time the fish crossed the reference point at the center of the pond. Figure 7 shows the total number of crossings by period and species. It can be seen that the two species do not seem to be significantly different, but there are fewer crossings during the treatment period in comparison to the acclimation period.

Figure 7: Number of Total Crossings by Period and Species
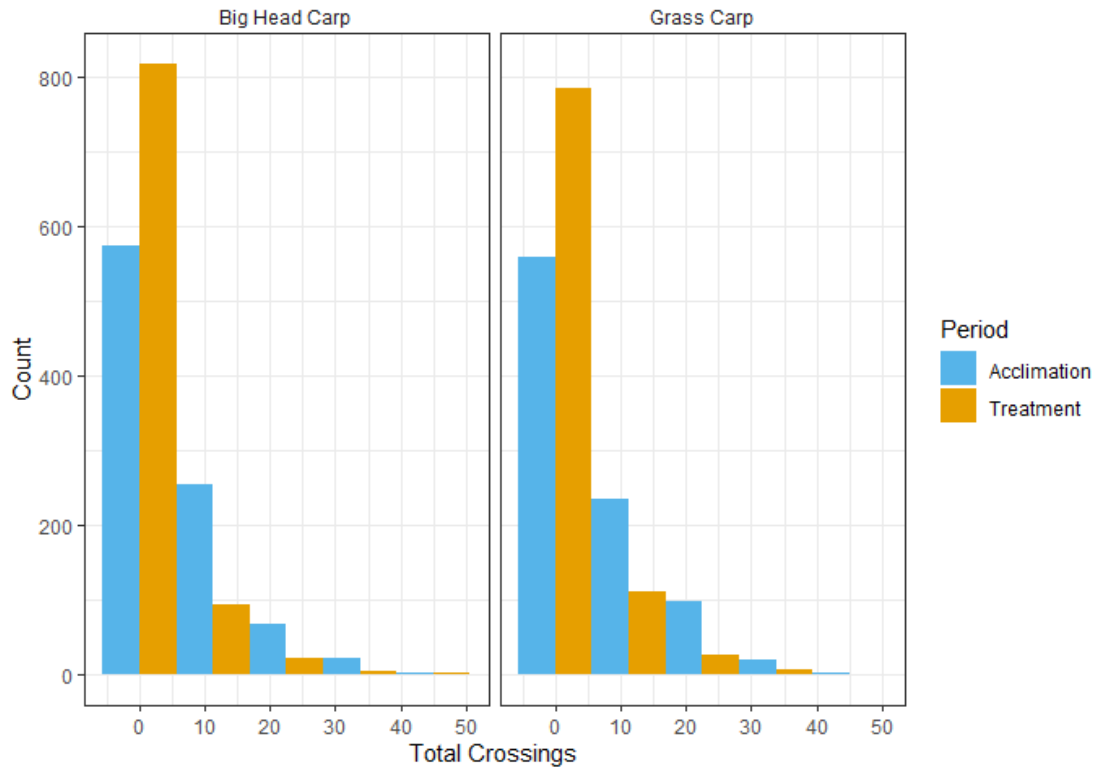
3.3.2. Bighead Carp

The species were handled separately to maintain a nested structure. The following results only include bighead carp.

3.3.2.1.   Pooled Modeling

The pooled model used the pooled absolute distance, which was created by averaging the absolute distance of all the fish in each tank, as the response variable. The linear model (lm) output is shown in table 9 when using period, level, treatment side, and the interaction between level and treatment as fixed predictor variables.

Table 9: Bighead Carp Pooled Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.2619	0.0545	0.1217	0.4021	4.8027	0.0049
trtsideWest	0.0757	0.0412	-0.0303	0.1817	1.8356	0.1259
levelMedium	-0.0559	0.0714	-0.2395	0.1276	-0.7832	0.4690
levelHigh	0.0303	0.0714	-0.1533	0.2138	0.4237	0.6894
periodTreatment	0.0198	0.0714	-0.1638	0.2033	0.2771	0.7928
levelMedium:periodTreatment	0.1005	0.1010	-0.1590	0.3601	0.9956	0.3651
levelHigh:periodTreatment	-0.1352	0.1010	-0.3948	0.1244	-1.3389	0.2383

The intercept is the acclimation period, east treatment side, and low level.

Using a significance value of α = 0.05, there was no significant difference in the pooled absolute distance estimate when moving from one period to another, when moving from the east treatment side to the west, or when moving from one treatment level to the next regardless of period.

3.3.2.2. Fixed Effect Modeling

The fixed effects model specified all predictor variables as fixed effects. Level, trial nested in level, period, and the interaction between period and level were used as predictor variables with the original absolute distance as the response variable in the linear model (lm). The output from the model is shown in table 10.

Table 10: Bighead Carp Fixed Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.4846	0.0217	0.4413	0.5279	22.3080	<0.0001
levelMedium	-0.1892	0.0306	-0.2501	-0.1283	-6.1861	<0.0001
levelHigh	-0.1892	0.0307	-0.2505	-0.1280	-6.1603	<0.0001
periodTreatment	-0.0016	0.0223	-0.0460	0.0428	-0.0708	0.9438
levelMedium:periodTreatment	0.0803	0.0310	0.0185	0.1420	2.5904	0.0115
levelHigh:periodTreatment	-0.1308	0.0315	-0.1936	-0.0680	-4.1495	0.0001

The intercept is the acclimation period and low level.

Using a significance value of $\alpha = 0.05$, the fixed model found that there was a significant difference when moving levels in both the acclimation and treatment period. The differences in estimated absolute distance was greater when moving from one level to the next during the acclimation period than the treatment period. There was also a significant difference in estimated absolute distance when moving from the acclimation to treatment period.

3.3.2.3.   Maximum Likelihood Modeling

The maximum likelihood (ML) model used the original absolute distance as the response variable in a linear mixed effects model (lme) with method specified as ML. Period, level, treatment side, and the interaction between period and level were used as fixed predictor variables with unique fish number nested in trial as random variables. The output from the maximum likelihood (ML) model is shown in table 11.

Table 11: Bighead Carp ML Model Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.2680	0.0549	0.1618	0.3743	4.8862	<0.0001
trtsideWest	0.0706	0.0509	-0.0549	0.1960	1.3860	0.2244
levelMedium	-0.0416	0.0609	-0.1917	0.1085	-0.6831	0.5249
levelHigh	0.0376	0.0611	-0.1129	0.1880	0.6151	0.5654
periodTreatment	-0.0016	0.0228	-0.0457	0.0426	-0.0692	0.9452
levelMedium:periodTreatment	0.0803	0.0317	0.0189	0.1417	2.5319	0.0154
levelHigh:periodTreatment	-0.1308	0.0322	-0.1932	-0.0683	-4.0557	0.0002

The intercept is the acclimation period, east treatment side, and low level.

Using a significance value of $\alpha = 0.05$, there was no significant difference in the absolute distance estimate when moving from one period to another, when moving from the east treatment side to the west, or when moving from one treatment level to the next during the acclimation period. The absolute distance estimate was significantly different when moving from one treatment level to the next during the treatment period, however, the difference was greater when moving from level low to high during the treatment period than when moving from level low to medium.

3.3.2.4.   Restricted Maximum Likelihood Modeling

The restricted maximum likelihood (REML) model used the original absolute distance as the response variable in a linear mixed effects model (lme) with method specified as REML. Period, level, treatment side, and the interaction between period and level were used as fixed predictor variables with unique fish number nested in trial as random variables. The output from the restricted maximum likelihood (REML) model is shown in table 12.

Table 12: Bighead Carp REML Model Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.2678	0.0701	0.1262	0.4094	3.8221	0.0005
trtsideWest	0.0704	0.0656	-0.0983	0.2391	1.0724	0.3325
levelMedium	-0.0412	0.0774	-0.2401	0.1576	-0.5332	0.6168
levelHigh	0.0379	0.0775	-0.1613	0.2371	0.4894	0.6452
periodTreatment	-0.0016	0.0223	-0.0466	0.0435	-0.0708	0.9439
levelMedium:periodTreatment	0.0803	0.0310	0.0176	0.1429	2.5902	0.0133
levelHigh:periodTreatment	-0.1308	0.0315	-0.1945	-0.0671	-4.1491	0.0002

The intercept is the acclimation period, east treatment side, and low level.

Using a significance value of $\alpha = 0.05$, there was no significant difference in the absolute distance estimate when moving from one period to another, when moving from the east treatment side to the west, or when moving from one treatment level to the next during the acclimation period. The absolute distance estimate was significantly different when moving from one treatment level to the next during the treatment period, but the difference was greater when moving from level low to high during the treatment period than when moving from level low to medium.

3.3.3. Grass Carp

The species were handled separately to maintain a nested structure. The following results only include grass carp.

3.3.2.1. Pooled Modeling

The pooled model used the pooled absolute distance, that was created by averaging the absolute distance of all the fish in each tank, as the response variable. The linear model (lm)

output is shown in table 13 when using period, level, treatment side, and the interaction between level and treatment as fixed predictor variables.

Table 13: Grass Carp Pooled Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.2648	0.0621	0.1053	0.4244	4.2668	0.0080
trtsideWest	0.0656	0.0469	-0.0550	0.1862	1.3973	0.2212
levelMedium	-0.0666	0.0813	-0.2755	0.1423	-0.8199	0.4496
levelHigh	0.0388	0.0813	-0.1701	0.2477	0.4771	0.6534
periodTreatment	-0.0137	0.0813	-0.2226	0.1952	-0.1685	0.8728
levelMedium:periodTreatment	0.1464	0.1149	-0.1490	0.4418	1.2737	0.2588
levelHigh:periodTreatment	-0.0949	0.1149	-0.3903	0.2005	-0.8256	0.4466

The intercept is the acclimation period, east treatment side, and low level.

Using a significance value of $\alpha = 0.05$, there was no significant difference in the pooled absolute distance estimate when moving from one period to another, when moving from the east treatment side to the west, or when moving from one treatment level to the next regardless of period.

3.3.3.2.   Fixed Effect Modeling

The fixed effects model specified all predictor variables as fixed effects. Level, trial nested in level, period, and the interaction between period and level were used as predictor variables with the original absolute distance as the response variable in the linear model (lm). The output from the model is shown in table 14.

30

Table 14: Grass Carp Fixed Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.5078	0.0252	0.4575	0.5581	20.1291	<0.0001
levelMedium	-0.2309	0.0357	-0.3020	-0.1598	-6.4718	<0.0001
levelHigh	-0.2042	0.0360	-0.2760	-0.1324	-5.6688	<0.0001
periodTreatment	-0.0488	0.0252	-0.0991	0.0015	-1.9341	0.0569
levelMedium:periodTreatment	0.1389	0.0357	0.0678	0.2100	3.8927	0.0002
levelHigh:periodTreatment	-0.0750	0.0370	-0.1488	-0.0013	-2.0263	0.0463

The intercept is the acclimation period and low level.

Using a significance value of $\alpha = 0.05$, the fixed model found that there was a significant difference when moving levels in both the acclimation and treatment period. The differences in estimated absolute distance was greater when moving from one level to the next during the acclimation period than the treatment period. There was also a significant difference in estimated absolute distance when moving from the acclimation to treatment period.

3.3.3.3. Maximum Likelihood Modeling

The maximum likelihood (ML) model used the original absolute distance as the response variable in a linear mixed effects model (lme) with method specified as ML. Period, level, treatment side, and the interaction between period and level were used as fixed predictor variables with trial as a random variable. The output from the maximum likelihood (ML) model is shown in table 15.

Table 15: Grass Carp ML Model Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.2846	0.0583	0.1717	0.3975	4.8829	<0.0001
trtsideWest	0.0655	0.0540	-0.0674	0.1985	1.2141	0.2789
levelMedium	-0.0699	0.0649	-0.2297	0.0899	-1.0774	0.3305
levelHigh	0.0291	0.0653	-0.1317	0.1899	0.4453	0.6747
periodTreatment	-0.0488	0.0258	-0.0988	0.0012	-1.8915	0.0658
levelMedium:periodTreatment	0.1389	0.0365	0.0682	0.2095	3.8068	0.0005
levelHigh:periodTreatment	-0.0750	0.0379	-0.1484	-0.0017	-1.9816	0.0544

The intercept is the acclimation period, east treatment side, and low level.

Using a significance value of $\alpha = 0.05$, there was no significant difference in the absolute distance estimate when moving from one period to another, when moving from the east treatment side to the west, or when moving from one treatment level to the next during the acclimation period. The absolute distance estimate was significantly different when moving from level low to medium during the treatment period.

3.3.3.4. Restricted Maximum Likelihood Modeling

The restricted maximum likelihood (REML) model used the original absolute distance as the response variable in a linear mixed effects model (lme) with method specified as REML. Period, level, treatment side, and the interaction between period and level were used as fixed predictor variables with trial as a random variable. The output from the restricted maximum likelihood (REML) model is shown in table 16.

Table 16: Grass Carp REML Model Output

	Estimate	Std. Error	Lower 95% CI	Upper 95% CI	t-Value	p-Value
(Intercept)	0.2846	0.0741	0.1348	0.4344	3.8408	0.0004
trtsideWest	0.0655	0.0693	-0.1126	0.2437	0.9456	0.3878
levelMedium	-0.0699	0.0819	-0.2805	0.1407	-0.8529	0.4327
levelHigh	0.0291	0.0822	-0.1823	0.2405	0.3537	0.7380
periodTreatment	-0.0488	0.0252	-0.0998	0.0022	-1.9343	0.0602
levelMedium:periodTreatment	0.1389	0.0357	0.0668	0.2110	3.8931	0.0004
levelHigh:periodTreatment	-0.0750	0.0370	-0.1498	-0.0002	-2.0265	0.0494

The intercept is the acclimation period, east treatment side, and low level.

Using a significance value of $\alpha = 0.05$, there was no significant difference in the absolute distance estimate when moving from one period to another, when moving from the east treatment side to the west, or when moving from one treatment level to the next during the acclimation period. The absolute distance estimate was significantly different when moving from one treatment level to the next during the treatment period, but the difference was greater when moving from level low to medium during the treatment period, than when moving from level low to high during the treatment period.

Section 4. Discussion

As exemplified in the results section, variable specification and model selection will change the conclusions. Each dataset highlighted the different ways these choices impact results and how models compare to one another.

4.1. Simulated Data

When comparing the widths of the confidence intervals, table 17 shows that regardless of the treatment effect, the models will still produce intervals of the similar widths. Table 17 also shows that the intervals were the widest for cases with a large between subject standard deviation. The largest intervals occurred when there was a large between subject standard deviation, a high number of repetitions, and a low number of subjects (note: upper right cell of table 17).

Table 17: Confidence Interval Widths

	Large Number of Subjects	Balanced Size Structure	Large Number of Repetitions
Large Between Subject SD	Pooled: 6.4231/6.4278 Fixed: 2.5000/2.4989 ML: 6.2605/6.2650 REML: 6.4231/6.4278	Pooled: 9.4028/9.4180 Fixed: 1.7654/1.7613 ML: 8.9203/8.9347 REML: 9.4028/9.4180	Pooled: 14.5352/14.5935 Fixed: 1.2481/1.2476 ML: 13.0007/13.0528 REML: 14.5352/14.5935
Balanced Variance Structure	Pooled: 3.8835/3.8838 Fixed: 2.4930/2.4969 ML: 3.7852/3.7855 REML: 3.8835/3.8838	Pooled: 5.6696/5.6600 Fixed: 1.7645/1.7649 ML: 5.3786/5.3695 REML: 5.6696/5.6600	Pooled: 8.7645/8.7387 Fixed: 1.2476/1.2474 ML: 7.8392/7.8161 REML: 8.7645/8.7387
Large Within Subject SD	Pooled: 2.6251/2.6234 Fixed: 2.4964/2.4948 ML: 2.5586/2.5570 REML: 2.6251/2.6234	Pooled: 3.7883/3.7994 Fixed: 1.7647/1.7637 ML: 3.5939/3.6044 REML: 3.7883/3.7994	Pooled: 5.8485/5.8287 Fixed: 1.2464/1.2480 ML: 5.2310/5.2134 REML: 5.8485/5.8287

The first value is the width from the simulated dataset with no treatment effect. The second value is the width from the simulated dataset with a treatment effect.

Table 17 shows that the average confidence interval lengths are comparable except for the fixed model. Regardless of the treatment effect, the interval widths by model were very similar, but the fixed interval was so much smaller because of the incorrectly specified variables. The addition of random variables as fixed resulted in an artificial inflation of sample size, resulting in smaller confidence intervals. The comparison of intervals shows that having more subjects will tighten the confidence interval, while having more repetitions widens the interval.

Table 17 and 18 also highlight that the intervals for the pooled model and the restricted maximum likelihood model are the same. This is because the restricted maximum likelihood model maximizes a modified likelihood that is derived from the base regression model that the pooled model employs to remove the impacts of the mean components (Zhang, 2015).

Table 18: Average Confidence Interval Widths

Model	Average No Effect Width	Average Effect Width
Pooled	6.7715	6.7712
Fixed	1.8366	1.8362
ML	6.2856	6.2853
REML	6.7715	6.7712

Table 19 shows the percent difference between interval widths. The maximum likelihood and restricted maximum likelihood models were compared because the pooled and restricted maximum likelihood models have matching results. Table 19 highlights that the widths did not change depending on the treatment effect but did change depending on the data conditions. The widths were largest for case 5 when the data structure had a large between subject standard deviation, a high number of repetitions, and a lower number of subjects.

Table 19: Percent Difference for Interval Widths

Case	No Effect ML Width	No Effect REML Width	Effect ML Width	Effect REML Width	No Effect Percent	Effect Percent
1	3.6012	3.7960	3.6109	3.8062	2.9838	2.9962
2	8.9361	9.4194	8.9186	9.4010	7.4040	7.4003
3	5.3680	5.6584	5.3675	5.6579	4.4477	4.4538
4	5.2322	5.8498	5.2137	5.8291	9.4597	9.4400
5	13.0114	14.5472	12.9764	14.5080	23.5242	23.4952
6	7.8296	8.7538	7.7960	8.7162	14.1557	14.1156
7	2.5545	2.6208	2.5546	2.6209	1.0165	1.0180
8	6.2536	6.4161	6.2703	6.4332	2.4884	2.4987
9	3.7838	3.8821	3.7815	3.8798	1.5056	1.5069

When making comparisons between the power of the models, the maximum likelihood model tended to have the highest powers. The fixed model did report high powers, but as seen from the simulated dataset that had no treatment effect, the fixed model power was always greater than 0.7, even when the power should have been around 0.05, exemplifying the artificial inflation of significance. The other three models had lower powers when the dataset had a large between subject standard deviation or a large number of repetitions.

4.2. Exposure Related Effects of Zequanox

When comparing the variations in model results, different models came to conflicting conclusions when using a significance value of $\alpha = 0.05$. The pooled model found that there was a significant difference in estimated pooled condition factor when moving from one species to the other. The fixed model, maximum likelihood model, and restricted maximum likelihood

model echoed the conclusions of the pooled model and found that there was a significant difference in estimated condition factor when moving from one species to the other.

The pooled model did not find a significant difference in the pooled condition factor estimate when moving from one treatment level to another; the fixed and restricted maximum likelihood models also did not find a significant difference in the condition factor estimate when moving from one treatment level to another. However, the maximum likelihood model found that there was a significant difference in the estimated condition factor when moving from the control treatment to the 50 mg/L treatment.

While the pooled and restricted maximum likelihood models did not find the movement from the control treatment to the 50 mg/L treatment significant, their significant values were quite close to the significance level. The fixed model, however, did not have a low significance value when moving from the control treatment to the 50 mg/L treatment.

The Hasse Diagram in figure 8 shows the structure of the dataset. The values for each variable represent the proper degrees of freedom that the variable uses, with the superscript representing the number of unique combinations and the subscript representing the degrees of freedom needed.
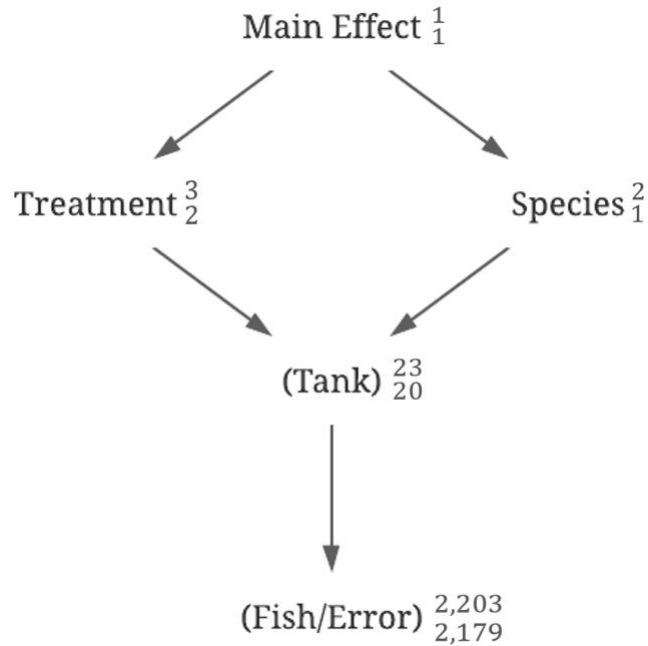
Figure 8: Hasse Diagram

Hasse Diagrams are used to check that the correct denominator degrees of freedom are being used in a model. Table 20 shows that the fixed model is using the error, or residual, degrees of freedom (2,179), when it should in fact be using the random variable tank as the denominator degrees of freedom (20). As a result, the fixed model has a much more significant effect, paralleling what was seen in the simulated datasets. Since the denominator degrees of freedom are artificially inflated, the treatment effect is determined to have a much more significant impact and at a much higher frequency, as shows by the simulated datasets.

Table 20: ANOVA Output

	Numerator DF	Denominator DF	F-Value	p-Value
Pooled Treatment	2	20	5.2481	0.0147
Fixed Treatment	2	2179	975.8404	<0.0001
ML Treatment	2	20	9.1762	0.0015
REML Treatment	2	20	7.2226	0.0044

4.3. Acoustic Telemetry Evaluation of Carbon Dioxide

4.3.1. Bighead Carp

When comparing the difference in model results, different models came to different conclusions when using a significance value of $\alpha = 0.05$. The pooled model found no significant differences between estimated pooled absolute distance when moving from one level to another during the treatment period. The other three models concluded that there was a significant difference in the absolute distance estimate when moving from one level to another during the treatment period.

While the conclusions of the maximum likelihood model and the restricted maximum likelihood model were the same, the fixed and restricted maximum likelihood models found that the difference in estimated absolute distance was greater when moving from the low level to the medium level during the treatment period than what the maximum likelihood model concluded.

The fixed model was the only model that found a significant difference in estimated absolute distance between treatment levels during the acclimation period. This highlights how the improper specifications of variables lead to more significant conclusions even when it is unsupported by the data.

The Hasse Diagram in figure 9 shows the structure of the dataset. The values for each variable represent the proper degrees of freedom that the variable uses, with the superscript representing the number of unique combinations and the subscript representing the degrees of freedom needed.
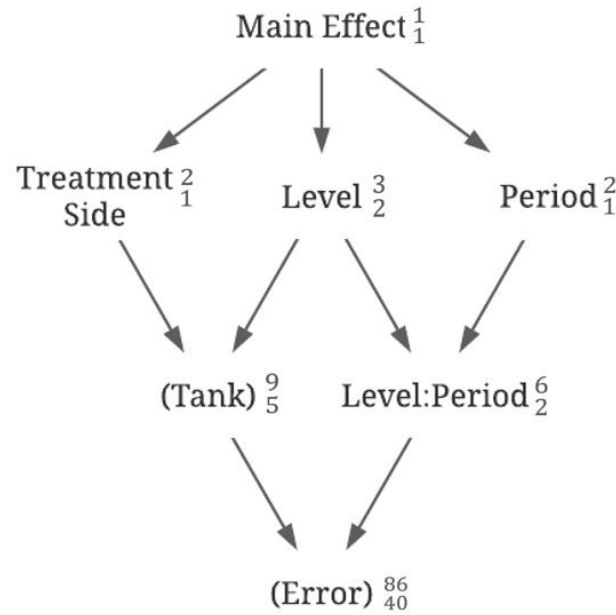

Figure 9: Hasse Diagram

Hasse Diagrams are used to check that the correct denominator degrees of freedom are being used in a model. The denominator degrees of freedom are different for the pooled model because the random term is excluded from the model. Table 21 shows that the fixed model is using the incorrect number of denominator degrees of freedom (74), which should be the error, or residual, degrees of freedom. As a result, the fixed model has an artificially inflated denominator degrees of freedom and is overly sensitive, but unlike the *Exposure Related Effects of Zequanox* dataset, the fixed model results in the same conclusions as the maximum likelihood model and the restricted maximum likelihood model. The pooled model was also using the

incorrect number of degrees of freedom (5), which lead to an underpowered model and subsequently a different result than the other models.

The results from this dataset also highlight that the pooled model used a different denominator degree of freedom (5) than the maximum likelihood model and the restricted maximum likelihood model. This is most likely due to the difference in the repeated measure structure which occurred for this dataset but not in the *Exposure Related Effects of Zequanox* dataset.

Table 21: Bighead Carp ANOVA Output

	Numerator DF	Denominator DF	F-Value	p-Value
Pooled Level:Period	2	5	2.7445	0.1569
Fixed Level:Period	2	74	23.5108	<0.0001
ML Level:Period	2	40	22.4602	<0.0001
REML Level:Period	2	40	23.5068	<0.0001

4.3.2. Grass Carp

When comparing the difference in model results, different models came to different conclusions when using a significance value of $\alpha = 0.05$. The pooled model found no significant differences between estimated pooled absolute distance when moving from one level to another during the treatment period. The maximum likelihood model determined that there was a significant difference between the low treatment level and the medium treatment level during the treatment period, but there was not a significant difference between the low treatment level and the high treatment level. The fixed and restricted likelihood models concluded that there was a significant difference in the absolute distance estimate when moving from one level to another

41

during the treatment period. While the maximum likelihood model did not determine that there were no significant difference in estimated absolute distance between the low treatment level and the high treatment level during the treatment period, it is important to note that it was very close to being significant.

The fixed model was the only model that found a significant difference in estimated absolute distance between treatment levels during the acclimation period. This highlights how the improper specifications of variables lead to more significant conclusions even when it is unsupported by the data.

The Hasse Diagram in figure 10 shows the structure of the dataset. The values for each variable represent the proper degrees of freedom that the variable uses, with the superscript representing the number of unique combinations and the subscript representing the degrees of freedom needed.

Main Effect $_1^1$

Treatment Side $_1^2$ Level $_2^3$ Period $_1^2$

(Tank) $_5^9$ Level:Period $_2^6$
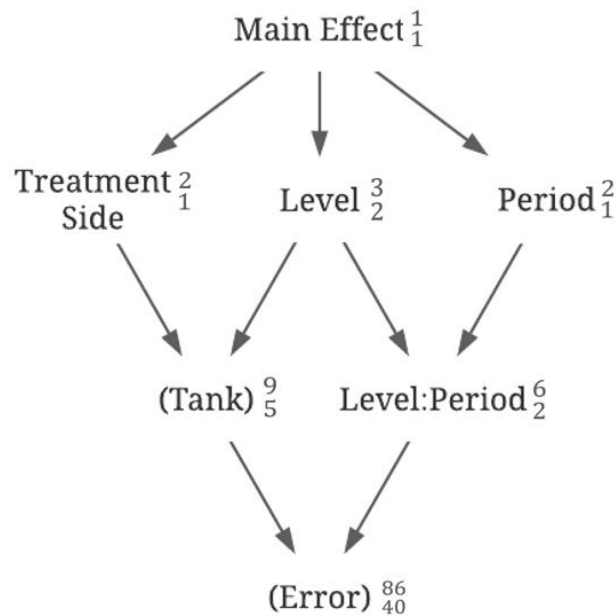
(Error) $_{40}^{86}$

Figure 10: Hasse Diagram

Hasse Diagrams are used to check that the correct denominator degrees of freedom are being used in a model. The denominator degrees of freedom are different for the pooled model because the random term is excluded from the model. Table 22 shows that the fixed model is using the incorrect number of denominator degrees of freedom (74), which should be the error, or residual, degrees of freedom. As a result, the fixed model has an artificially inflated denominator degrees of freedom and is overly sensitive, but unlike the *Exposure Related Effects of Zequanox* dataset, the fixed model results in the same conclusions as the maximum likelihood model and the restricted maximum likelihood model. The pooled model was also using the incorrect number of degrees of freedom (5), which lead to an underpowered model and subsequently a different result than the other models.

The results from this dataset also highlight that the pooled model used a different denominator degree of freedom (5) than the maximum likelihood model and the restricted maximum likelihood model. This is most likely due to the difference in the repeated measure structure which occurred for this dataset but not in the *Exposure Related Effects of Zequanox* dataset.

Table 22: Grass Carp ANOVA Output

	Numerator DF	Denominator DF	F-Value	p-Value
Pooled Level:Period	2	5	2.2371	0.2023
Fixed Level:Period	2	74	17.4903	<0.0001
ML Level:Period	2	40	16.7271	<0.0001
REML Level:Period	2	40	17.4938	<0.0001

When comparing the results between species, there were varying conclusions. The Bighead Carp and Grass Carp pooled models both found that there were no significant

differences when moving from the acclimation to treatment period, when moving from one treatment side to the next, or when moving from level to level regardless of period.

The Bighead Carp maximum likelihood model found that there were significant differences in estimated absolute distance when moving from one level to the next during the treatment period, but the difference was greater when moving from level low to level high. The Grass Carp maximum likelihood model only found the movement from the low treatment level to the medium treatment level during the treatment period significant, but the difference was greater when moving from level low to medium for Grass Carp than for Bighead Carp.

The Bighead Carp restricted maximum likelihood model found that there were significant differences in the absolute distance estimate when moving from one level to the next during the treatment period. The Grass Carp restricted maximum likelihood model also found that there were significant differences in the absolute distance estimate when moving from one level to another during the treatment period, but the Bighead Carp model found the difference between level low to high was larger while the Grass Carp model found the difference between level low to medium was larger.

While the absolute distance estimate was not significantly different when moving from the acclimation period to the treatment period for either species, the Grass Carp maximum likelihood and restricted maximum likelihood models found that there was a larger difference between periods than what was seen in the Bighead Carp maximum likelihood and restricted maximum likelihood models, leading to a lower significance value for the treatment period.

Section 5. Conclusion

As comparisons were made between the four models across four datasets, multiple significant differences came to light. The simulated datasets highlighted the artificial inflation of the degrees of freedom for the fixed model, leading to more determinations of a significant effect, even when there was no treatment effect. The simulated datasets also highlighted that the pooled model and restricted maximum likelihood model evaluated the data in the same ways. The pooled model pooled observations over the random effect, while the restricted maximum likelihood model used a method that separates by random effect and models over that specific section, resulting in the same conclusions.

By changing the conditions of the simulated dataset, comparisons could be made as the variance and size structure change. While models performed similarly, apart from the fixed model, the trends that emerged from the simulated datasets were that estimations were accurate, but the interval widths widened when the number of subjects were small and narrowed when the number of subjects was large. It was also found that having a larger within subject standard deviation narrowed intervals while having smaller within subject standard deviation widened intervals.

The empirical datasets highlighted the differences in results based on model. Changing the models changed the results and inferences, but the biggest differences were in the significance levels. Just switching the model method from the maximum likelihood method to the restricted maximum likelihood model changed whether an entire variable was significantly different. For the *Exposure Related Effects of Zequanox* dataset, the restricted maximum likelihood model concluded that there was a significant difference in estimated condition factor when moving from the control to the 50 mg/L treatment, but the maximum likelihood model

concluded that there was no significant difference between treatments. For the *Acoustic Telemetry Evaluation of Carbon Dioxide* dataset, the restricted maximum likelihood model concluded that every treatment level was significantly different for Grass Carp during the treatment period, but the maximum likelihood model concluded that there was only a significant difference when moving from the low level to the medium level.

The empirical datasets also highlighted the differences in the pooled and restricted maximum likelihood models. The simulated datasets framed how the two models were the same, but the empirical datasets highlighted how the two models result in different conclusions. The pooled models found little to no significant differences for both empirical datasets, but the restricted maximum likelihood model concluded that there were significant differences in estimates for certain variables. Finally, the empirical datasets highlighted how having repeated measures in a dataset changed the determined denominator degrees of freedom.

Variable specification and model selection can have big impacts on the conclusions drawn from results. When modeling data from a hierarchical design, it is important to properly specify variables to avoid unsupported results. The fixed model highlights how the degrees of freedom can be a great check to make sure variables are being properly modeled. Checking model structures and appropriately categorizing variables will prevent inaccurate results and incorrect inferences, which could make the biggest difference of all.

References

Calenge C (2006). "The package adehabitat for the R software: tool for the analysis of space and
habitat use by animals." *Ecological Modelling*, 197, 1035.

Cesarani, Alberto, Ivan Pocrnic, Nicolo Macciotta, Breno Fragomeni, Ignacy Misztal, and
Daniela Lourenco (2018). Bias in heritability estimates from genomic restricted
maximum likelihood methods under different genotyping strategies. Journal of Animal
Breeding and Genetics.136:40-50. DOI: 10.1111/jbg.12367.

Cupp, Aaron, Ashley Lopez, Justin Smerud, Jose Rivera, Nicholas Swyers, David Smith, and
Mark Gaikowski (2020). Acoustic telemetry evaluation of carbon dioxide as a behavioral
deterrent for invasive fishes: Data: U.S. Geological Survey data release,
https://doi.org/10.5066/P9QBSCIE.

Luoma, James, Todd Severson, Jeremy Wise, and Matthew Barbour (2018). Exposure-related
effects of Zequanox on juvenile lake sturgeon (Acipenser fulvescens) and lake trout
(Salvelinus namaycush). Management of Biological Invasions. 9(2):163-175. DOI:
10.3391/mbi.2018.9.2.09.

Oehlert, Gary (2012). A Few Words About REML. URL
http://users.stat.umn.edu/~gary/classes/5303/handouts/REML.pdf.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for
Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Zhang, Xiuming (2015). A Tutorial on Restricted Maximum Likelihood Estimation in Linear
Regression and Linear Mixed-Effects Model. A*Star-NUS Clinical Imaging Research
Center. 4-5.