Rachel Prokopius

Regression Analysis

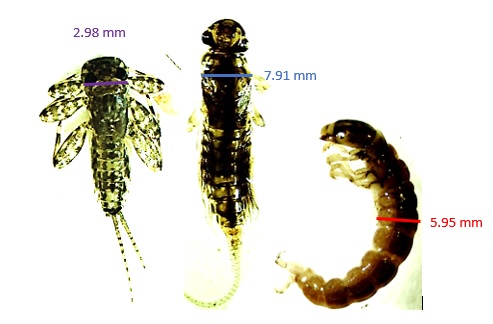
Final Project

17 April 2020

***Introduction***

Macroinvertebrates, the larval stage of an insect's life cycle, contribute immensely to the ecosystem energetics of marine and aquatic systems. Because of this, measuring and interpreting the energetic content of these water-based insect larvae can greatly assist scientists in studying energy-associated ecosystem processes such as predator-prey selectivity and trophic-level energetic transfer (Gilinsky 1984, Oertli 1993, Turesson et al. 2002, Zhao et al. 2005, Helmus et al. 2013). The standard length-to-mass relationship has been used since the 1980's to relate macroinvertebrate size to dry mass, which is then converted to caloric content via relationships measured and recorded since the 1960's (Cummins and Wuycheck 1971, Smock 1980, Brodmann and Reyer 1999, Wilt et al. 2014).

What is not standard, however, is the protocol for measuring the length of the macroinvertebrate specimen; measurements may include, but are not limited to, the length from the anterior labium to posterior of the last abdominal segment of the insect larvae, the width of the mesothorax, labium and head capsule, and any other measurements the researcher deems it fit to add (Sample et. al 1993, Johnston and Culjak 1999, Eklöf et al. 2017).Lengths can be determined by hand-measuring, using simple digital imaging and analysis software to compute the measurements or by taking measurements from previous literature (Sample et. al 1993, Paavo et al. 2008, Velghe and Gregory-Eaves 2013, Pond et al. 2016).Because of the wide variety of unique body parts and length-mass relationships between macroinvertebrate taxa, it is generally accepted that length-dry mass relationships must be taxon specific at the very least to the family level (Smock 1980, Mährlein et al. 2016, Eklöf et al. 2017). Because of the wide variety of measurement techniques for macroinvertebrate length-dry mass relationships that are oftentimes researcher-specific, getting accurate, reusable relationships has proven incredibly difficult. Additionally, the wide variety of morphological characteristics between macroinvertebrate families makes standardizing length measurement protocols nearly impossible (Image 1).

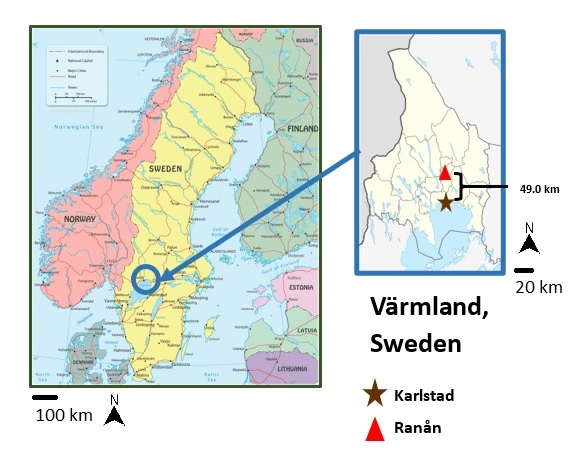


*Image 1. Image illustrating the unique morphologies of different macroinvertebrate families. Images taken at 1x power with a Motic 3.0 Camera. Size bars for each macroinvertebrate are included. Macroinvertebrate familes from left to right: Ephemeroptera, Plecoptera, Trichoptera. Images taken by Rachel Prokopius using a Motic 3.0 camera.*

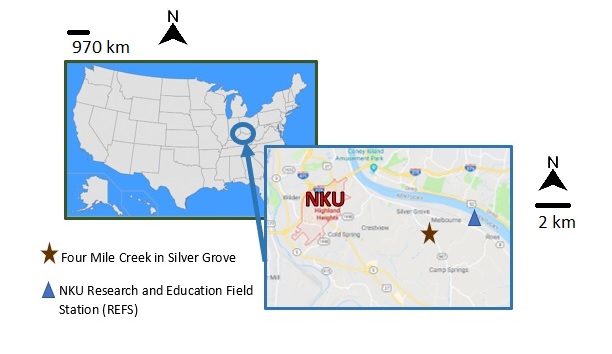
With advancements in technology and increased access to inexpensive computer programs for analyzing images (for example, ImageJ), it is possible to obtain potentially more-accurate measurements for aquatic organisms, namely the surface-area of the organism. Because of the variety of morphological characteristics unique to certain macroinvertebrate families and the relative subjectivity of length measurements, measurement of surface area from macroinvertebrate images may be a better measurement when compared to the standard length measurements. Previous analysis of the dataset obtained for this study have suggested this hypothesis to be true. However, visual analysis of the relationship between both surface area and length to macroinvertebrate dry mass appear to be more curvilinear than linear. Therefore, this project aims to determine whether a transformed model is statistically a better predictor of dry mass from surface area and length of macroinvertebrates. It is hypothesized that the comparison of measurement methods for caddisfly larvae (Trichoptera) will show that surface area is a stronger predictor (i.e. has a higher R2 value) of macroinvertebrate dry mass than length when the surface area and dry mass measurements are transformed using a natural log function.

***Methods***

Samples were taken once a month from November 2017 through May 2018 from Ranån, a stream about forty minutes north of Karlstad, Sweden that connects to Klarälven, the large river present in the area (Figure 1). Similar data was collected and analyzed in the Northern Kentucky Area, in a tributary of the Ohio River known as Four Mile Creek. Data was collected in October of 2018 (Figure 2).



*Figure 1. Maps of study area in Värmland, Sweden. Scale bars and north arrows are supplied for each map portion in the figure. Maps supplied by http://www.geographicguide.com/europe-maps/sweden.htm and https://en.wikipedia.org/wiki/V%C3%A4rmland\_County.*



*Figure 2. Maps of study area in Northern Kentucky. Red "NKU" symbolizes the university. Scale bars and north arrows are supplied for each map portion in the figure. Maps supplied by https://www.sporcle.com/games/Matt/find\_the\_states and Google Maps*

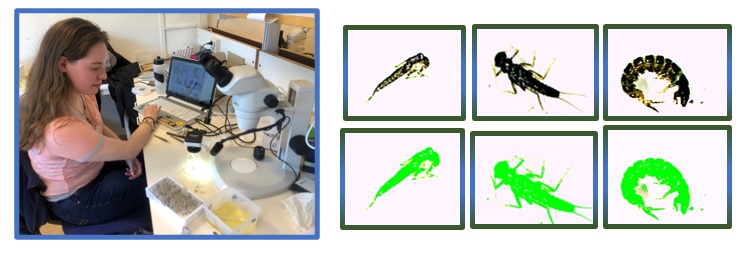
Each sampling excursion included analysis of stream conditions such as water height, temperature, dissolved oxygen, etc. Macroinvertebrate sampling comprised of one individual positioning a fine-mesh seine towards the flow of water and another individual using his/her feet and macroinvertebrate extraction device to dislodge debris from up to one meter in front of the seine. The individual would work his/her way towards the seine, continuing the kicking and dislodging of rocks and debris. Once at the seine, both individuals would lift the seine so it stayed taut and parallel to the surface of the water. The seine would then be carefully exposed to the water to wash the collected specimens towards the center of the seine, which made collection of specimens in a bucket much easier and quicker (Image 2). This process was repeated four more times for a total of five samples from the sample area. Additional collection methods included using D-nets to dislodge macroinvertebrates from the rocks and debris at the edges of the stream, and placing drift-foraging collection apparatuses in the water column for a couple of hours to catch macroinvertebrates traveling freely through the water.



*Image 2. Research at the Ranån study site in Sweden. Upper left: holding kicknet for catching macroinvertebrates. Upper right: research students collecting macroinvertebrates. Lower left: setting up the drift net system. Lower right: learning to use field equipment. Photos taken by Dr. Richard Durtsche and Rachel Prokopius.*

The bucket containing the specimens and water from the sample area were then taken to Karlstad Universitet and Northern Kentucky respectively, placed in a refrigerator in the lab analysis area with a bubbler running constantly to provide aeration. This method kept specimens alive for days or even weeks, depending on the time available to analyze the sample each day. During sample analysis, the monthly sample was removed from the bucket and picked through with little water in order to better see specimen movement. Specimens were removed from the debris and detritus in the bucket and sorted by Order, the majority of which were further identified to Family and Genus, when possible.

Imaging of specimens was performed using a Motic Images Plus 3.0 multi output digital microscope camera and accompanying software. The camera attached to an imaging cone with 1x power, which combined with the 1x power of the camera itself to have an overall power of 1x. On rare occasions an individual specimen would be analyzed under the dissecting microscope; the camera would attach to a 10x lens that fit into one of the eyepiece holes of the microscope, and the microscope would be set at 0.63x, giving an overall power of 6.3x. The cone and microscope were each calibrated in the imaging software, and each calibration was used accordingly. When using the cone, it would be placed with lights over a petri dish of specimens. Anywhere from ten to twenty specimens could be analyzed in a single image, depending on the size of the specimens. Lighting, color, contrast, exposure, etc. were manipulated on the program in order to isolate each specimen from the dish and from one another (Image 1). Two images were taken of each dish, one for measuring and one for reference. The specimens in the measuring image were labeled with numbers and surface area, perimeter and length were either recorded on the spot or done at a later time with the imaging software. The measurements could take place at a later time because the program could upload the saved image and still recognize the individual specimens, which allowed for many specimens to be imaged in one day, preserved and measured later. The specimens would then be placed into individual capped vials and suspended either in water and frozen or suspended in 70% ethanol for at least 24 hours.



*Image 3. Analysis of macroinvertebrates in the lab. Left: setup of imaging and data collection from macroinvertebrates. Photo taken by Dr. Richard Durtsche. Right: Surface area of macroinvertebrates highlighted in green and calculated in square mm by the Motic 3.0 Imaging software. Photos taken by Rachel Prokopius.*

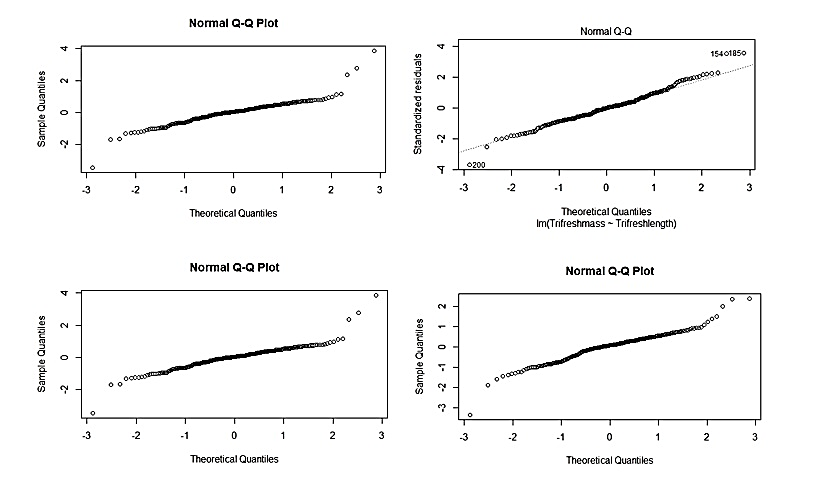
After specimens were preserved for at least 24 hours in their respective mediums, they were reimaged using the process illustrated above and their new measurements recorded for comparison to fresh specimen measurements. Each specimen was returned to its respective vial without a preservation medium and dried uncapped in a drying oven at 60°C for at least 24 hours. Once dried, the specimens were removed from the drying oven and quickly capped in order to minimize condensation on the specimens as they were moved from a warmer to a cooler environment. Each specimen was massed in milligrams on a microgram scale that measured to two places passed the decimal.

Statistical analyses were performed with the data collected from the macroinvertebrate specimens. Normality plots and residual plots were created for the Order Trichoptera to identify any patterns in the data pre and post-transformation. Regressions were created for surface area-mass relationships and length-mass relationships for specimens in different preservative mediums (fresh, frozen or alcohol), and an ANOVA run on each regression to test for significance.. All statistical analysis was performed with R 3.6.2.

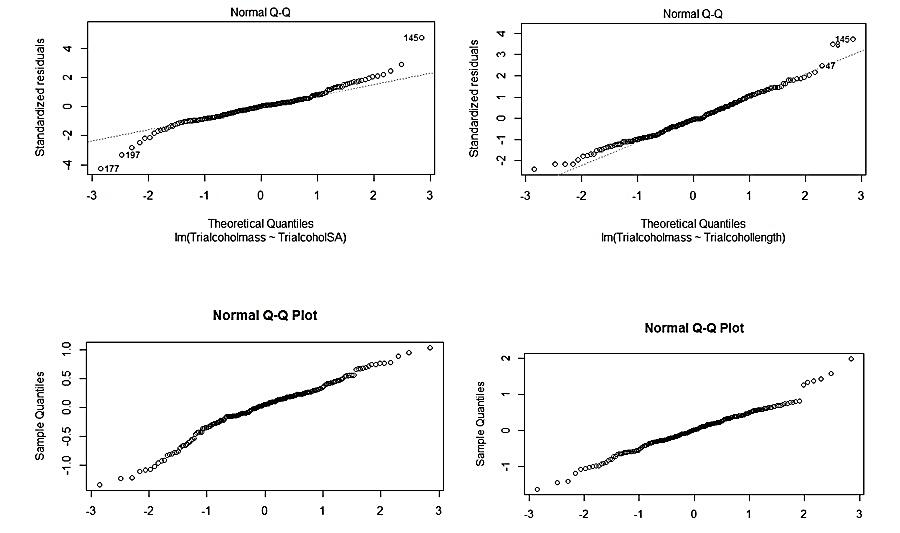
***Results***

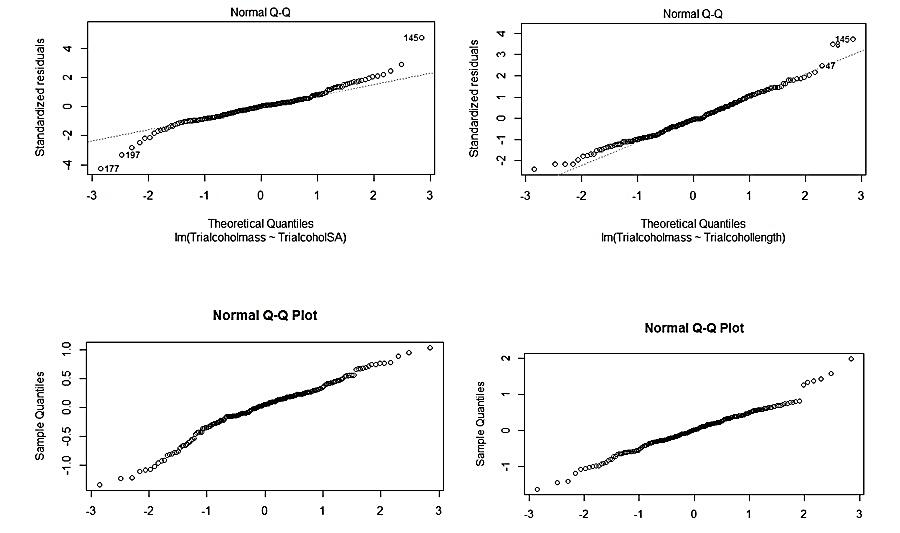
Out of all of the data collected on the Order Trichoptera in Sweden and Northern Kentucky, the Family Hydropsychidae was most-often found. Therefore, analysis of results will include the entire Order Trichoptera and the individual Family Hydropsychidae in order to apply the technique to different taxonomic levels.

Plots of normality were constructed for the linear and natural log-transformed models for each of the preservation techniques (fresh, alcohol and frozen) for both the Order Trichoptera and the Family Hydropsychidae (Figures 3-8). The original normality plots for the fresh and alcohol-preserved specimens at both the Order and Family levels do not appear to change drastically, with a few outliers on the edges of the normality plots but overall, the residuals falling on a straight line. For the frozen-preserved specimens, the data points with higher standardized residuals do not fall on a straight line for the linear model. However, once the data is log-transformed these residuals fall on the straight line of the normality plot.

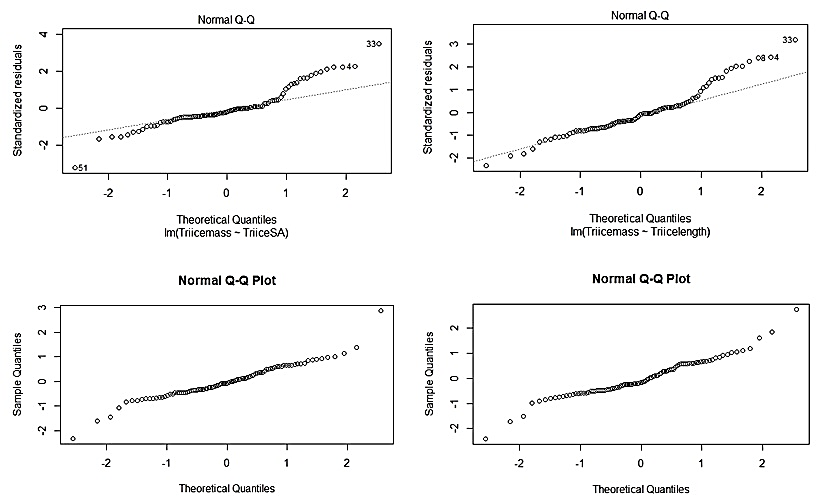


*Figure 3: Plots of normality for the residuals of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of fresh specimens in the Order Trichoptera. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*

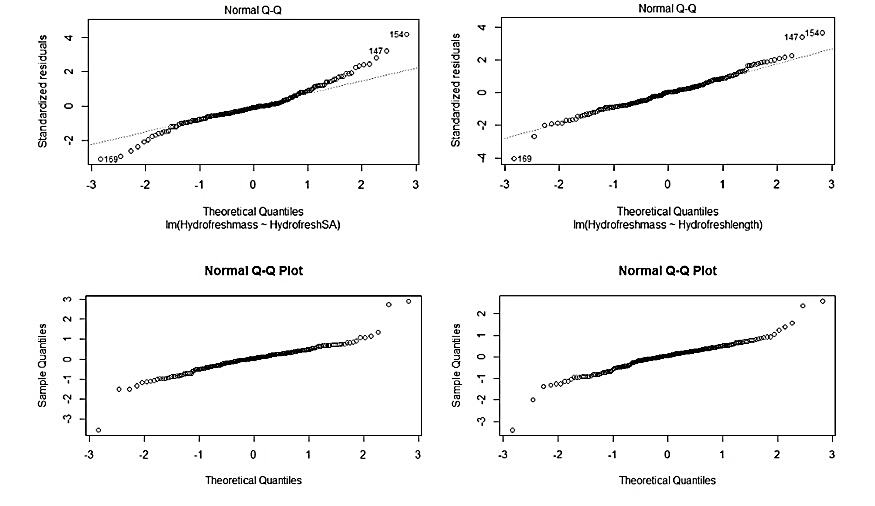




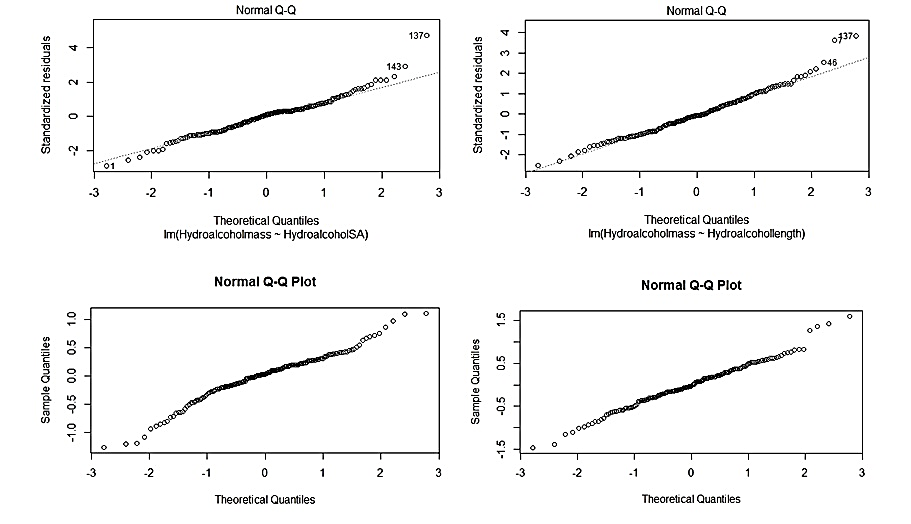
*Figure 4: Plots of normality for the residuals of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of alcohol-preserved specimens in the Order Trichoptera. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*



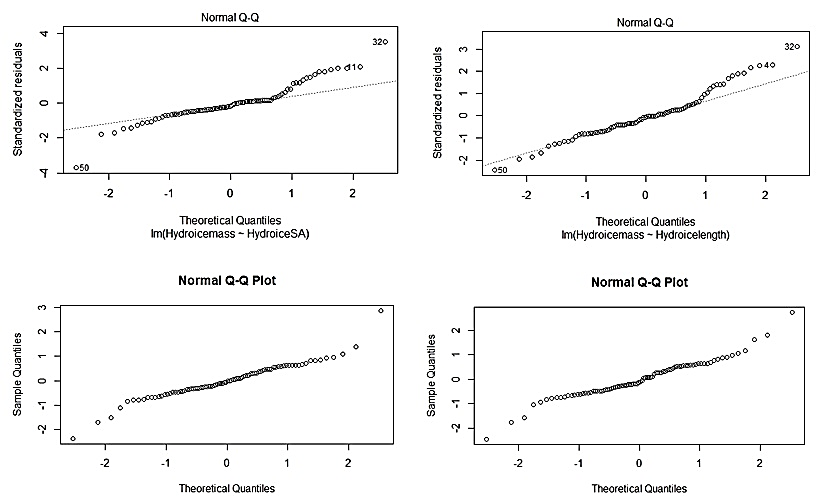
*Figure 5: Plots of normality for the residuals of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of frozen-preserved specimens in the Order Trichoptera. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*



*Figure 6: Plots of normality for the residuals of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of fresh specimens in the Family Hydropsychidae. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*

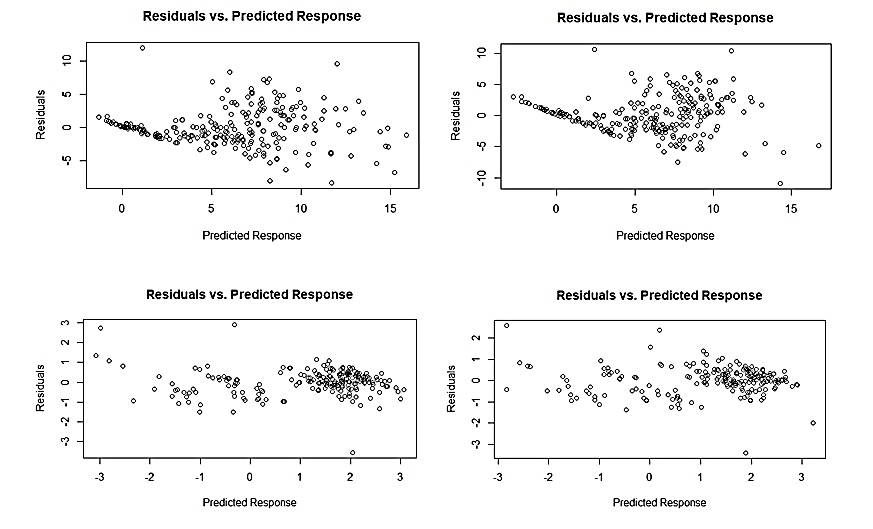


*Figure 7: Plots of normality for the residuals of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of alcohol-preserved specimens in the Family Hydropsychidae. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*

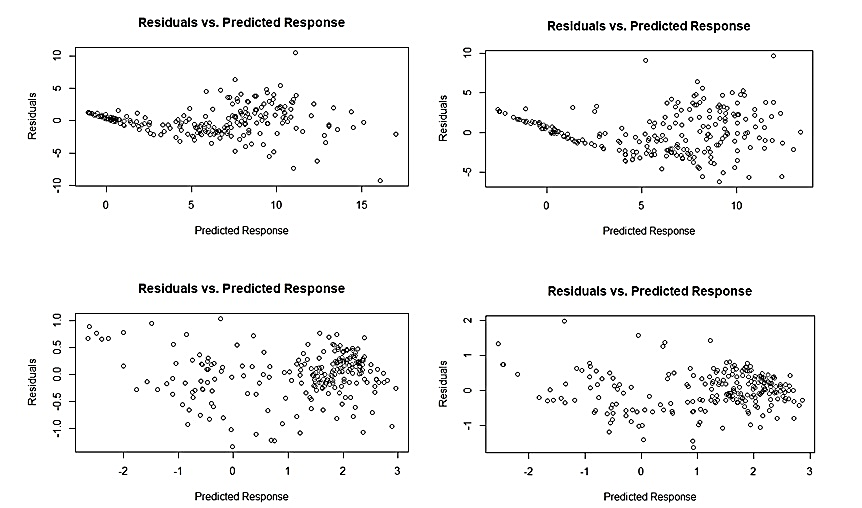


*Figure 8: Plots of normality for the residuals of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of frozen-preserved specimens in the Family Hydropsychidae. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*

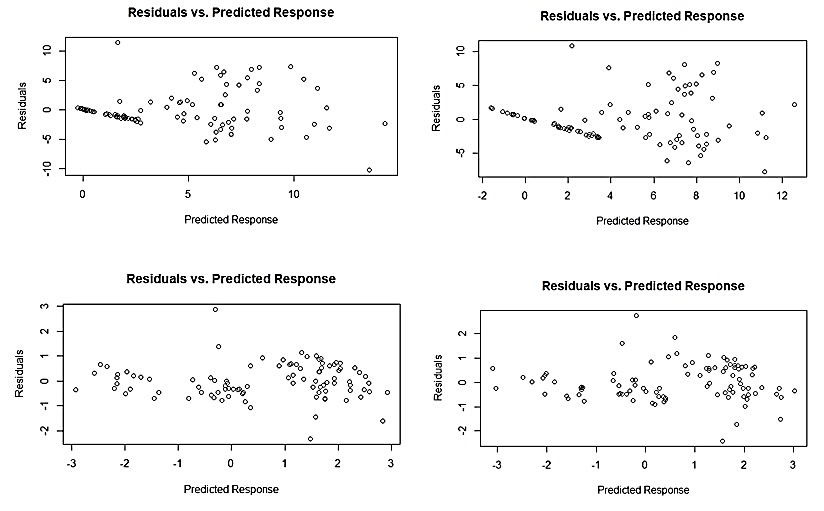
Plots of residuals versus predicted response values were constructed for the linear and natural log-transformed models for each of the preservation techniques (fresh, alcohol and frozen) for both the Order Trichoptera and the Family Hydropsychidae (Figures 9-14). A pattern is seen for smaller predicted responses in all of the linear models, with less of a pattern for higher predicted responses. For the natural log-transformed models, there is less of a pattern seen for the plots overall regardless of the size of the predicted response.



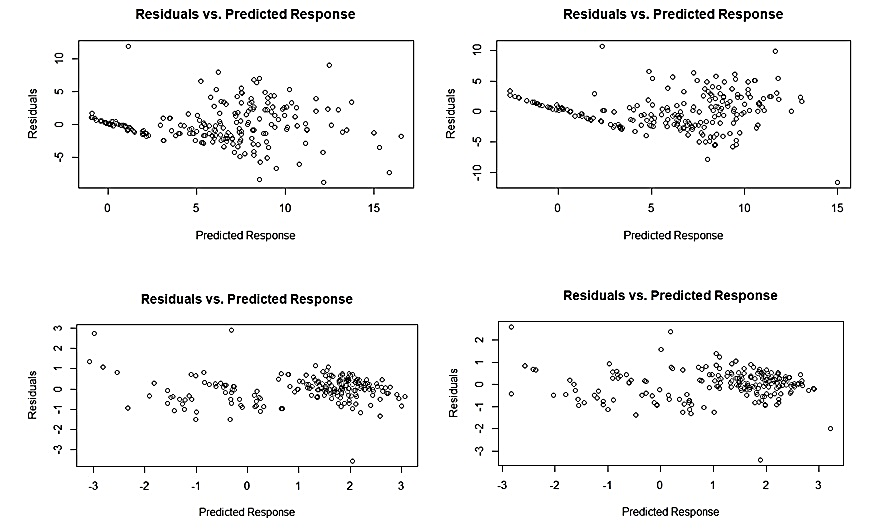
*Figure 9: Plots of residuals versus predicted response values of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of fresh specimens in the Order Trichoptera. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*



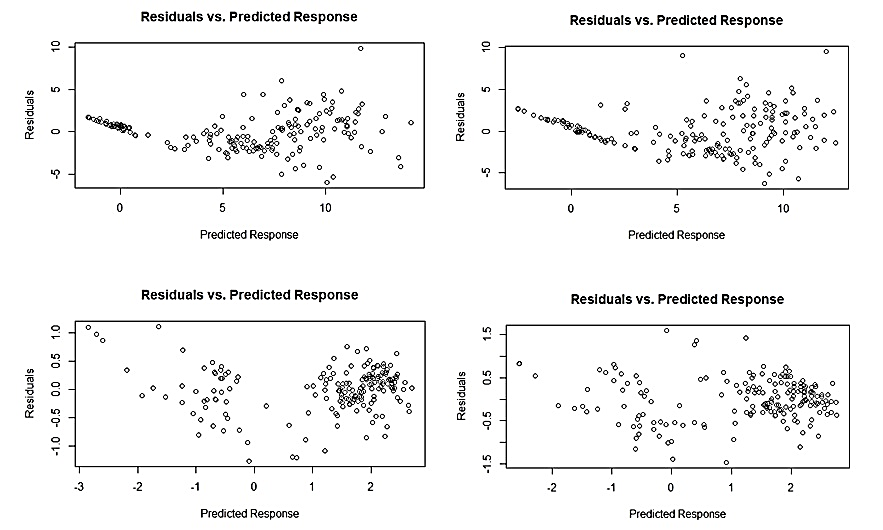
*Figure 10: Plots of residuals versus predicted response values of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of alcohol-preserved specimens in the Order Trichoptera. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*



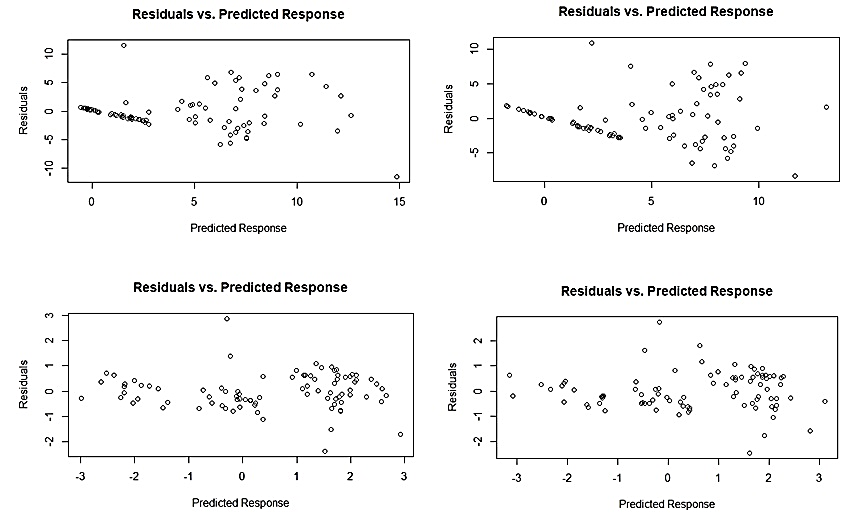
*Figure 11: Plots of residuals versus predicted response values of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of frozen-preserved specimens in the Order Trichoptera. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*



*Figure 12: Plots of residuals versus predicted response values of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of fresh specimens in the Family Hydropsychidae. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*



*Figure 13: Plots of residuals versus predicted response values of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of alcohol-preserved specimens in the Family Hydropsychidae. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*



*Figure 14: Plots of residuals versus predicted response values of the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of frozen-preserved specimens in the Family Hydropsychidae. Top left: linear surface area model. Top right: linear length model. Bottom left: natural log-transformed surface area model. Bottom right: natural log-transformed length model.*

All regression analyses for the Order Trichoptera and the Family Hydropsychidae were significant to an alpha level of 0.05, regardless of whether the model was transformed. Natural log transformation increased the R2 value for all of the regressions compared to their respective linear models. All surface area to dry mass models, except for the fresh surface area to dry mass linear model for the Order Trichoptera, had a larger R2 value than the corresponding length to dry mass relationships (Table 1, Figures 15-)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Taxonomic level** | **Model Type** | **Measurement** | **F-statistic** | **df** | **R2 value** |
| Order-Trichoptera | Linear | Surface Area | 478.8 | 1, 250 | 0.6569 |
| Order-Trichoptera | Natural log transformed | Surface Area | 878 | 1, 250 | **0.7784** |
| Order-Trichoptera | Linear | Length | 388 | 1, 250 | 0.6081 |
| Order-Trichoptera | Natural log transformed | Length | 888.8 | 1, 250 | **0.7805** |
| Family-Hydropsychidae | Linear | Surface Area | 392.1 | 1, 212 | 0.6491 |
| Family-Hydropsychidae | Natural log transformed | Surface Area | 904.9 | 1, 212 | **0.8102** |
| Family-Hydropsychidae | Linear | Length | 367.1 | 1, 212 | 0.6339 |
| Family-Hydropsychidae | Natural log transformed | Length | 870.7 | 1, 212 | **0.8042** |
| Order-Trichoptera | Linear | Surface Area | 808.7 | 1, 228 | 0.7801 |
| Order-Trichoptera | Natural log transformed | Surface Area | 2073 | 1, 228 | **0.9009** |
| Order-Trichoptera | Linear | Length | 520.7 | 1, 228 | 0.6955 |
| Order-Trichoptera | Natural log transformed | Length | 1236 | 1, 228 | **0.8443** |
| Family-Hydropsychidae | Linear | Surface Area | 713.4 | 1, 183 | 0.7958 |
| Family-Hydropsychidae | Natural log transformed | Surface Area | 1911 | 1, 183 | **0.9126** |
| Family-Hydropsychidae | Linear | Length | 450.8 | 1, 183 | 0.7113 |
| Family-Hydropsychidae | Natural log transformed | Length | 1180 | 1, 183 | **0.8657** |
| Order-Trichoptera | Linear | Surface Area | 113.7 | 1, 94 | 0.5474 |
| Order-Trichoptera | Natural log transformed | Surface Area | 457.1 | 1, 94 | **0.8294** |
| Order-Trichoptera | Linear | Length | 98.74 | 1, 94 | 0.5123 |
| Order-Trichoptera | Natural log transformed | Length | 379.9 | 1, 94 | **0.8016** |
| Family-Hydropsychidae | Linear | Surface Area | 112.6 | 1, 86 | 0.5671 |
| Family-Hydropsychidae | Natural log transformed | Surface Area | 421 | 1, 86 | **0.8304** |
| Family-Hydropsychidae | Linear | Length | 89.06 | 1, 86 | 0.5087 |
| Family-Hydropsychidae | Natural log transformed | Length | 347.1 | 1, 86 | **0.8014** |

*Table 1: Results of ANOVAs run on the linear and natural log-transformed models relating surface area (sq mm) and length (mm) to dry mass (mg) of specimens. Color corresponds to preservation technique (blue: fresh, red: alcohol-preserved, green: frozen-preserved). Models within preservation techniques are separated by taxonomic level (Order-Trichoptera and Family-Hydropsychidae) by darker lines. The higher R2 value for each model comparison is bolded, and the highest R2 value per taxonomic level per preservation technique is underlined. Every ANOVA result is significant to an alpha level of 0.05 and with a p-value less than 0.0001, and are not included with each ANOVA result for simplicity of table design.*

***Discussion:***

Data collected from both Sweden and Northern Kentucky aquatic systems suggested that surface area measurements of macroinvertebrate taxa are more accurate than the traditional length-mass relationship previously established and used in literature (Smock 1980, Benke et al. 1999, Eklöf et al. 2017). Rsquared values in the 0.65-1.0 range are ideal for using regressions to make comparisons between variables, obviously with values closer to 1.0 being overall preferred. Rsquared values tended to be higher for Trichoptera Families that had larger sample sizes, which suggests that for certain Trichoptera Families (ex. Polycentropodidae, Rhyacophilidae) more data must be collected to determine accurate size to mass relationships for the particular families.

When considering nonlinear models, a lower standard error of regression is preferred because it suggests that the data points are closer to the model line. In every instance with the alcohol-preserved specimens,the standard error of regression is lower for the nonlinear model than the linear model (Table 1). However, certain nonlinear models do not represent significant relationships between the predictor and response variables. For example, the nonlinear model for the relationship between surface area and dry mass for Polycentropodidae Family is nonsignificant, while the linear model is (Table 1). This tended to occur when fewer data points were available for the macroinvertebrate family. For example, the Phryganeidae Family is not included in the table provided in this study because both the linear and nonlinear models relating size and mass were not significant. Only three members of this family were found when sampling, which shows that a certain number of specimens must be collected and analyzed in order to build an accurate model for relating size and mass.

It is also important to consider the efficiency and ease of linear versus nonlinear models. Though the standard error of regression for all of the nonlinear models for the Trichoptera Families listed in Table 1 was lower than the standard error of regression for the linear model, the errors of regression are not that different. In most cases, both models are significant and suggest there is a relationship between the size measurement and the dry mass of the macroinvertebrate. It must be considered whether a slight decrease in standard error of regression is worth the extra work of applying a nonlinear model and potentially transforming the data. Subsequent studies and additiions to the data set should include more specimens from the lower-represented Trichoptera Families, and other transformations such as log transformation could be applied to the data for potentially an even better fit to the data collected.

***Literature Cited***

Brodmann, P. A., and H. U. Reyer. 1999. Nestling provisioning in water pipits (*Anthus spinoletta*): do parents go for specific nutrients or profitable prey? Oecologia 120:506–514.

Cummins KW, Wuycheck JC. 1971. Caloric equivalents for investigation in ecological energetics. Mitt Int Vercin Limnol 18:1-158.

Eklöf J, Austin Å, Bergström, Donadi, Eriksson BDHK, Hansen J, Sunblad G. 2017. Size matters: relationships between body size and body mass of common coastal, aquatic invertebrates in the Baltic Sea. PeerJ 5, e2906.

Gilinsky E. 1984. The role of fish predation and spatial heterogeneity in determining benthic community structure. Ecology 65(2):455-68.

Helmus HR, Mercado-Silva N, Zanden MJV. 2013. Subsidies to predators, apparent competition and the phylogenetic structure of prey communities. Oecologia 173(3):997-1007.

Johnston TA, Cunjak RA. 1999. Dry mass-length relationships for benthic insects: a review with new data from Catamaran Brook, New Brunswick, Canada. Freshwater Biol 41:653-74.

Mährlein M, Pätzig M, Brauns M, Dolman AM. 2016. Length-mass relationships for lake macroinvertebrates corrected for back-transformation and preservation effects. Hydrobiologia 768:37-50.

Oertli B. 1993. Leaf litter processing and energy flow through macroinvertebrates in a woodland pond (Switzerland). Oecologia 96:466-77.

Paavo B, Ziegelmeyer A, Lavric E, Probert K. 2008. Morphometric correlations and body mass regressions for *Armandia maculate*, *Aglaophamus macrourna* (Polychaeta) and *Zethalia zelandica* (Gastropoda). New Zeal J Mar Fresh 42:85-91.

Pond GJ, Fritz KM, Johnson BR. 2016. Macroinvertebrate and organic matter export from headwater tributaries of a Central Appalachian stream. Hydrobiologia 779:75-91.

Sample BE, Cooper RJ, Greer RD, Whitmore RC. 1993. Estimation of insect biomass by length and width. The American Midland Naturalist 129(2):234-40.

Smock LA. 1980. Relationships between body size and biomass of aquatic insects. Freshwater Biol 10:375-83.

Turesson H, Persson A, Brönmark C. 2002. Prey selection of piscivorous pikeperch (*Stizostedion lucioperca*) includes active prey choice. Ecol Freshw Fish 11:223-33.

Velghe K, Gregory-Eaves I. 2013. Body size is a significant predictor of congruency in species richness patterns: a meta-analysis of aquatic studies. Plos One 8(2).

Wilt LM, Grebmeier JM, Miller TJ, Cooper LW. 2014. Caloric content of Chukchi Sea benthic invertebrates: modeling spatial and environmental variation. Deep-Sea Res Pt II 102:97-106.

Zhao X, Fox MG, Miller TJ, Lasenby DC. 2005. Effect of prey density, prey mobility and habitat structure on size selection and consumption of amphipods by a benthic feeding fish. Arch Hydrobiologia 165(2):269-88.