A comparison of macroinvertebrate rockbag samples from Terra Nova National Park in summer of 2003.

Submitted
to Dr. Schneider
in partial fulfillment of the requirements for
BIOL 7720

Robyn Auld
200398444

Memorial University of Newfoundland
Fall 2004

Introduction:

Benthic macroinvertebrate stream communities have been the subject of extensive study, and have recently received particularly intensive interest with respect to the way in which they respond to environmental conditions (Wetzel 2001). Research proceeds to uncover the way in which community dynamics and production are affected by both biotic and abiotic parameters (Lomond and Colbo 2000). This information may lead to a better understanding of how natural and anthropogenic influences may affect aquatic ecosystems.

Within Newfoundland and Labrador, and northeastern North America in general, there is still a need for community and population data with respect to aquatic insect communities. While Larson and Colbo (1983) have made great strides in providing a general overview of the macroinvertebrate taxa of Newfoundland, knowledge of the effects that various environmental parameters have on these populations is still required. The aquatic diversity of Newfoundland is particularly low as compared to most mainland regions, and it is of interest whether differences in regional patterns and richness will be detectable.

The analysis data in this project serves two purposes. The first is a contribution towards a larger collection of available aquatic insect data for Newfoundland and Labrador. One goal in this sampling data is to discover what insects are present in the two watersheds and to characterize the population structures that exist within those streams. This data will be used to compare insect population structure under the influence of several variables by looking at the composition of communities in two different watersheds, at both upstream and downstream locations. In addition, the data will be used to compare community structure at two different times during the summer: July and August.

Methods:

The data used in this analysis was taken from a larger set of data collected by Colbo, Kendall and Cote in Terra Nova National park during the summer of 2003. Terra Nova park is located on the eastern coast of Newfoundland, and borders on the Atlantic Ocean. The scale of the study was limited to the interior of the park and so the watershed used in this analysis is considered a random variable. Sampling locations included Big Pond and Minchin Pond (Watershed 1), and Blue Pond and Wing Pond (Watershed 2). Upstream and downstream locations were sampled within each of the watersheds. This was repeated in both June and July.

Sampling consisted of using rock bag samples, similarly to previous work by Lomond and Colbo (2000) and Muzaffar and Colbo (2002). Rockbags were constructed from expandable plastic bags with 1 cm openings. The bags measured 10-15 cm across and 30 cm long. They were filled with 2 L of course gravel and placed in the streams for 3 weeks. The original placement was in fast riffle habitats only. However, sometimes after three weeks the stream water level dropped and they were in much slower water. Sampling dates were June 18-20 and July 15-16. On sampling day, each bag was removed from the sampling location using a sweep net. The rocks were brushed cleaned and returned to the site, while the debris and insects were sieved and placed in 95% ethanol. Specimens were sorted and counted in the lab.

The response variable in this analysis is count (C), as measured by sorting or rock bag samples. This ratio scale variable has no units. Analysis of deviance was preformed using the Generalized Linear Model to test whether the number of M, S, and C differed significantly between watersheds, upstream and downstream locations, as well as between sampling dates and order. Statistical analysis was preformed using S-Plus 6.1 Academic Site Edition.

Results:

The data that I used was taken from a larger data set that included insect count data for 4 watersheds, 6 habitats and two date codes. However, as I later discovered, not every habitat was present in each of the watersheds, causing problems within my modeling. I reduced my working data to include 2 watersheds, 2 habitats and 2 date codes. The original data also contained several insect orders, each of which was divided into numerous species. In order to make the data manageable, I focused on the total count in three orders: Ephemeroptera, Plecoptera and Trichoptera, which will here forth be referred to by their common names: Mayflies, Stoneflies and Caddisflies.

This data is count data, meaning that it has non-negative integer values. Because of this, its error structure is non-normal. As a result, I used a generalized linear model to evaluate this data. I began by testing my model with a poisson log distribution but found some minimal cone structure. I experimented with the gamma identity function but the model would not compute. A gamma inverse function produced a worse cone structure and so I chose to return to the poisson log model, which appeared to produce the most homogeneous deviance residuals. This infers that the residuals arise from a Poisson distribution.

Originally, I conducted three analyses; one for each of the three orders. Using the number of individuals in each group as a response variable, I entered watershed, habitat and date as explanatory variables. However, based on the knowledge that the ratio of three orders is often compared in diversity studies (REF), I realized that it would be very difficult to compare these groups through three separate analyses. I therefore adapted my model so that order became additional explanatory variable for count. The following is a full account of the analysis of the data from that point forward.

Through this analysis I am interested in determining the following: Does macroinvertebrate count depend on watershed, habitat, collection date and order? Of particular interest is whether or not watershed, habitat or collection date interacts with order.

The variables are as follows:

Variable	Variable Type	Random/ Fixed	Scale	Units/Categories	Symbol
Count	Response	N/A	Ratio	None	C
Watershed	Explanatory	Random	Nominal	Watershed 1 Watershed 2	W
Habitat	Explanatory	Fixed	Nominal	Habitat 1 (Upstream) Habitat 2 (Downstream)	H
Date	Explanatory	Fixed	Nominal	Date 1 (June) Date 2 (July)	D
Order	Explanatory	Fixed	Nominal	Order 1 (Mayfly) Order 2 (Stonefly) Order 3 (Caddisfly)	O

Figures 1, 2, 3 and 4 present count data as related to each individual explanatory variable.

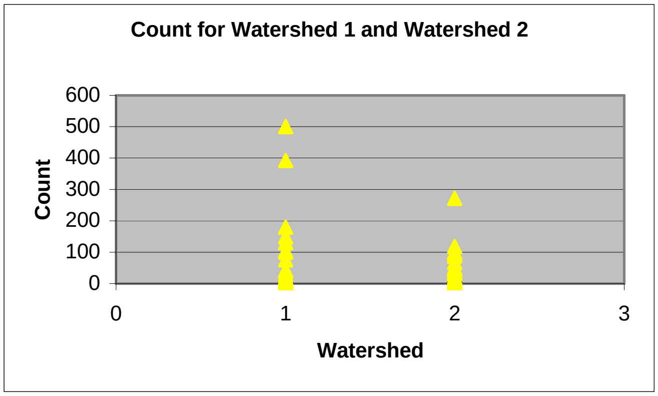


Figure 1. Counts for the two watersheds chosen as a random variable.

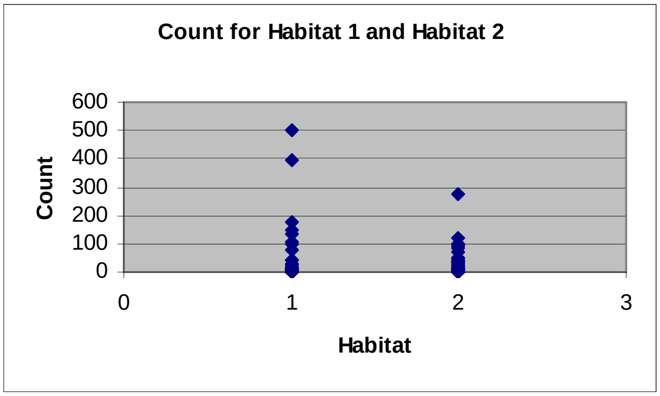


Figure 2. Counts for two habitat types used in the analysis. Habitat 1 is an upstream location. Habitat 2 is a downstream location.

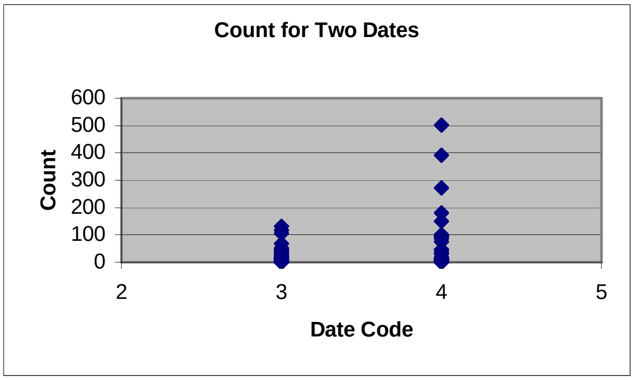


Figure 3. Counts for two date codes. Date code 3 represents samples taken in June. Date code 4 represents samples taken in July.

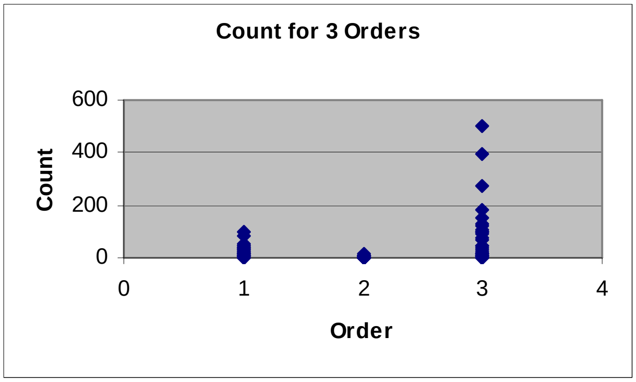


Figure 4. Counts for 3 insect orders. Order 1 is mayflies, Order 2 is stoneflies and Order 3 is caddisflies.

The formal model for the evaluation of this data is:

$f = e^\wedge(ß_{ref}) + e^\wedge(ß_W * W) + e^\wedge(ß_H * H) + e^\wedge(ß_D * D) + e^\wedge(ß_O * O) + e^\wedge(ß_{O*H} * O * H)$

$+ e^\wedge(ß_{O*D} * O * D) + e^\wedge(ß_{D*H} * D * H) + e^\wedge(ß_{O*D*H} * O * D * H) + \text{error}.$

There is a three-way interaction term in this model between habitat, date and order. Watershed is a random variable and is not considered in any interaction terms. Figure 5 shows the residual deviations of this model.

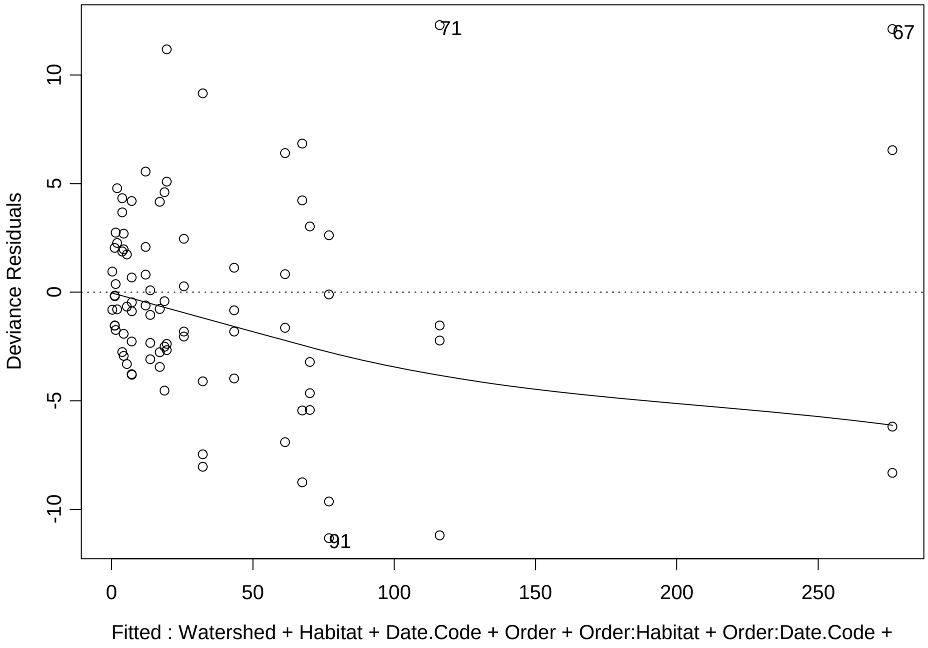


Figure 5. The deviance residuals for the formal model.

There is a slight conical shape to the residuals. However, compared to other fits of this data, the cone shaped is minimal. The model, which is a poisson distribution with poisson error is deemed appropriate.

The samples were taken at random points within each habitat, and are considered representative of the population in the area.

Since I am interested in knowing if count is affected by the above-mentioned variables, and if there is an interaction among the variables, hypothesis testing is appropriate. It should be mentioned that I expect count to vary with order and am therefore not interested in determining if this is true through hypothesis testing.

Although the model will be run in it's full form, the first relationship to be examined will be the 3-way interaction between habitat, date and order. Other hypotheses will be written as appropriate though the analyses.

HA: $ß_{O*D*H}$ = 0 (counts depends on order, date and habitat)

HO: $ß_{O*D*H}$ = 0 (the cross product ratio is equal to unity)

The statistic is G. The probability distribution is chi-square, and $\alpha = 0.05$.

The following is the model output and analysis of deviance table:

```
Coefficients:
                             Value Std. Error    t value
           (Intercept)  2.51780788 0.04770564  52.777992
             Watershed -0.63943574 0.02028443 -31.523470
               Habitat  0.23162980 0.04632573   5.000025
             Date.Code  0.33013160 0.04632573   7.126312
                Order1 -1.07293367 0.06853172 -15.656015
                Order2  0.75961193 0.02409441  31.526474
         HabitatOrder1 -0.07304022 0.06853172  -1.065787
         HabitatOrder2 -0.23590642 0.02409441  -9.790917
       Date.CodeOrder1  0.27578313 0.06853172   4.024168
       Date.CodeOrder2  0.09086204 0.02409441   3.771083
     Date.Code:Habitat -0.05714604 0.04632573  -1.233570
HabitatOrder1Date.Code -0.26600253 0.06853172  -3.881451
HabitatOrder2Date.Code -0.06798006 0.02409441  -2.821403

Analysis of Deviance Table

Poisson model

Response: Count

Terms added sequentially (first to last)
                        Df Deviance Resid. Df Resid. Dev
                   NULL                    95   7667.952
              Watershed  1 1197.714        94   6470.239
                Habitat  1   56.072        93   6414.167
              Date.Code  1  621.188        92   5792.979
                  Order  2 3227.907        90   2565.071
          Order:Habitat  2  419.237        88   2145.835
        Order:Date.Code  2  115.427        86   2030.408
      Date.Code:Habitat  1   18.551        85   2011.857
Order:Date.Code:Habitat  2   99.466        83   1912.390
```

ΔG (improvement of fit) for the three-way interaction is 99.466. I compute the p value for this statistic by using MiniTab.

```
Chi-Square with 2 DF

     x  P( X <= x )
99.466            1
```

Using this value, I calculate the other end of the tail, but subtracting from 1. The result is that the p value < 0.0001. There is no need to recompute the p value. There is a high n, and the number is very significant (randomization is not likely to change the result).

$p < 0.0001$ and therefore $p < 0.05$. Hence, reject HO and accept HA. There is an interaction between habitat, date and order. $\Delta G = 99.466$, df =2, $p < 0.0001$.

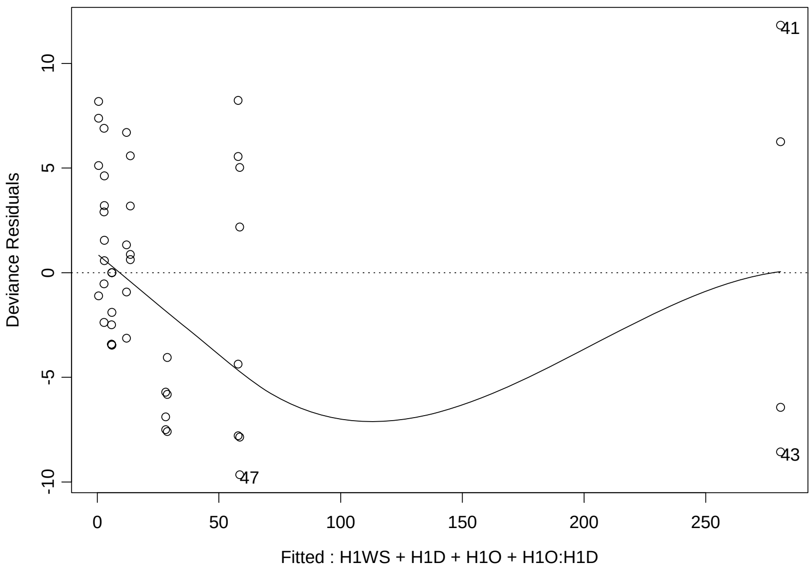
I cannot proceed to examine the other interactions because I cannot interpret one main effect independent of another. I must start breaking up the model to evaluate interactions and single variables separately. The first step in to examine the 2-way interactions for each of the other variables. The rest of this write up shows the analyses through which I proceeded along this course. Figure 6 shows a summary of these interactions. Below are the models, hypotheses, plots of deviance residuals, analysis of deviance tables, and accompanying statistics for each test.

I first look at the interaction between date and order for each of the two habitats.

Habitat One:

Model: $f = e\text{^}(ß_{ref}) + e\text{^}(ß_W * W) + e\text{^}(ß_D * D) + e\text{^}(ß_O * O) + + e\text{^}(ß_{O*D} * O * D)$ + error.
Hypotheses: HA: $ß_{O*D} = 0$
HO: $ß_{O*D} = 0$
The statistic is G. The probability distribution is chi-square, and $\alpha = 0.05$.

Plot of deviance residuals:



The residuals are homogeneously distributed. The model is appropriate.

```
Analysis of Deviance Table

Poisson model

Response: H1C

Terms added sequentially (first to last)
        Df Deviance Resid. Df Resid. Dev
   NULL                    47   5284.165
   H1WS  1  929.359        46   4354.806
    H1D  1  557.834        45   3796.973
    H1O  1 2118.622        44   1678.351
H1O:H1D  1  260.747        43   1417.604


 Chi-Square with 1 DF

     x  P( X <= x )
639.03           1
```

The p value is therefore < 0.0001 and is significant. $\Delta G = 260.747$, df =1, $p < 0.0001$.

The next step is to examine whether count is affected by order, for each date. However, I am not interested in hypothesis testing for this purpose. I know that there will be different counts for various orders. I was only interested in whether date had an interaction with the count for each order, which it does. At this point I will not test whether dates differ for orders, because I will be testing that with more data further on in the analysis. The following are the summary statistics for counts grouped by order under habitat 1 (date 1), and under habitat 1 (date 2).

```
H1D1O:1
             H1D1C
     Min:  0.00000
    Mean: 16.37500
  Median: 16.50000
     Max: 39.00000
Std Dev.: 12.96079
-----------------------------------------------------------
H1D1O:2
            H1D1C
     Min: 0.00000
    Mean: 0.75000
  Median: 0.00000
     Max: 4.00000
Std Dev.: 1.38873
-----------------------------------------------------------
H1D1O:3
              H1D1C
     Min:   3.00000
    Mean:  43.12500
  Median:  22.50000
     Max: 131.00000
Std Dev.:  48.30687
```

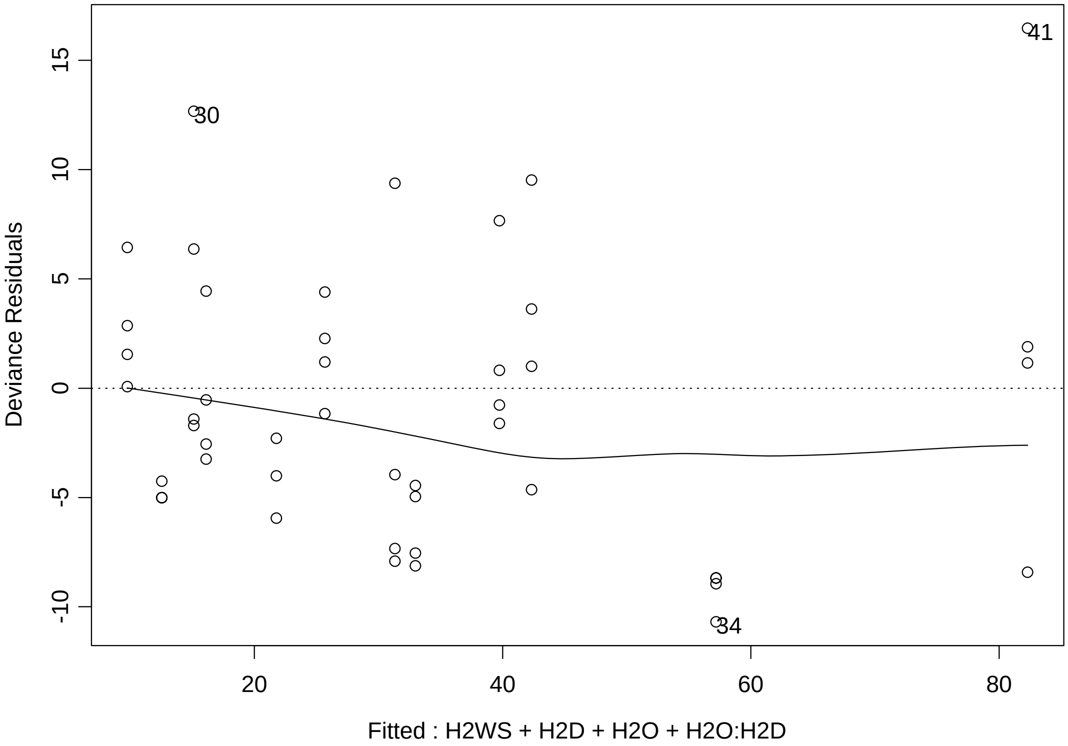
```
H1D2O:1
              H1D2C
     Min:  0.000000
    Mean:  8.750000
  Median:  9.000000
     Max: 15.000000
Std Dev.:  5.257647
-----------------------------------------------------------
H1D2O:2
              H1D2C
     Min:  0.000000
    Mean:  3.500000
  Median:  3.000000
     Max: 10.000000
Std Dev.:  3.664502
-----------------------------------------------------------
```

```
H1D2O:3
            H1D2C
     Min:   3.000
    Mean: 176.625
  Median: 125.500
     Max: 501.000
Std Dev.: 179.714
------------------------------------------------------------
```

Habitat 2:

Model: *f* = e^($ß_{ref}$) +e^($ß_W$* W) +e^($ß_D$*D) +e^($ß_O$*O) + +e^($ß_{O*D}$*O*D) + error.
Hypotheses: HA: $ß_{O*D} = 0$
HO: $ß_{O*D} = 0$
The statistic is G. The probability distribution is chi-square, and α = 0.05.

Plot of deviance residuals:



The deviance residuals are homogeneously distributed. No cone shape. Accept model.

```
Analysis of Deviance Table

Poisson model

Response: H2C

Terms added sequentially (first to last)
        Df Deviance Resid. Df Resid. Dev
   NULL                    47   2327.716
   H2WS  1 322.4982        46   2005.217
    H2D  1 122.1739        45   1883.044
    H2O  1 104.4971        44   1778.546
H2O:H2D  1   2.9530        43   1775.593

Chi-Square with 1 DF

    x  P( X <= x )
2.953     0.914282
```

The p value is 0.085718 for $\Delta G = 2.9530$, df =1; $p > 0.05$. The interaction is not significant. I must examine the other variables now. The hypotheses are as follows:

HA: $ß_O = 0$
HO: $ß_O = 0$

HA: $ß_D = 0$
HO: $ß_D = 0$

HA: $ß_W = 0$
HO: $ß_W = 0$

I look at other p values.

Factor	ΔG	df	p
Order	104.4971	1	<0.0001
Date	122.1739	1	<0.0001
Watershed	322.4982	1	<0.0001

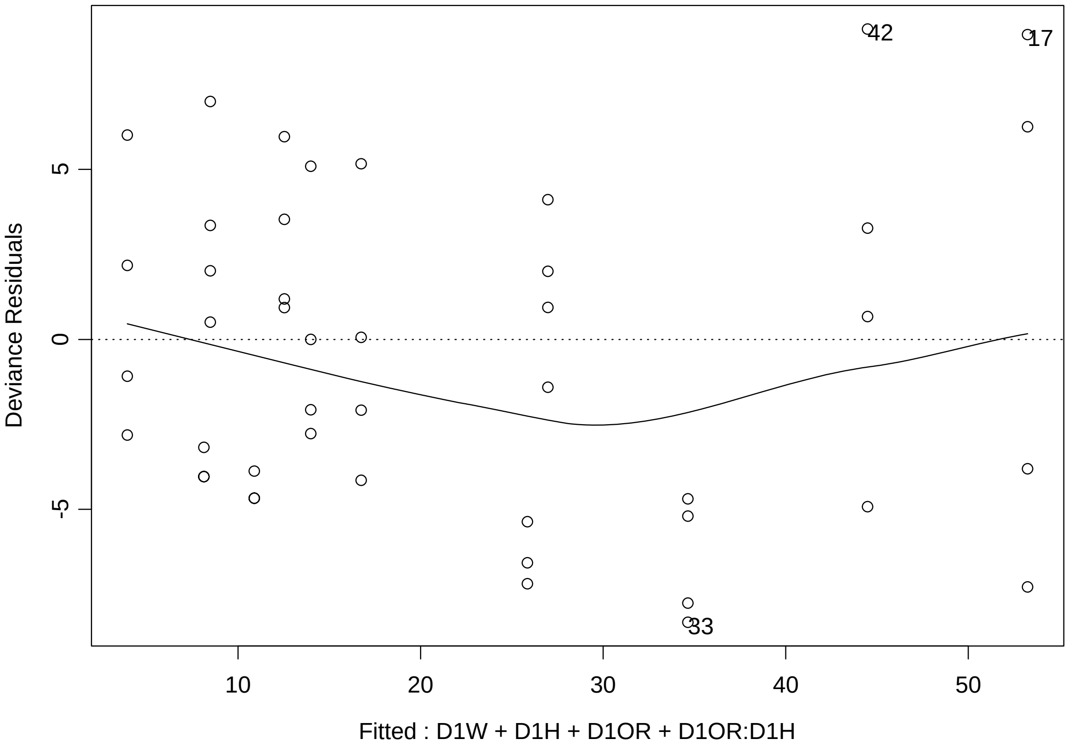
The other factors are significant for Habitat 2. In all cases reject HO and accept HA for Habitat 2.

Now I will look at the second interaction term in the 3-way interaction. I will look at the interaction between habitat and order for each of the two dates.

Date 1 (June):

Model: $f = e^{\wedge}(ß_{ref}) + e^{\wedge}(ß_W * W) + e^{\wedge}(ß_H * H) + e^{\wedge}(ß_O * O) + e^{\wedge}(ß_{O*H} * O * H)$ + error.
Hypotheses: HA: $ß_{O*H} = 0$
HO: $ß_{O*H} = 0$
The statistic is G. The probability distribution is chi-square, and $\alpha = 0.05$.

Plot of deviance residuals:



There is a slight bowl, but no cone shape, and the deviance residuals are fairly homogeneous.

```
Analysis of Deviance Table

Poisson model

Response: D1CT

Terms added sequentially (first to last)
         Df Deviance Resid. Df Resid. Dev
    NULL                    47   1547.053
     D1W  1 296.8850        46   1250.168
     D1H  1   5.5588        45   1244.609
    D1OR  1 137.3289        44   1107.280
D1OR:D1H  1  33.8240        43   1073.456

Chi-Square with 1 DF

     x  P( X <= x )
33.824      1.00000
```

The p value is <0.0001 for $\Delta G = 33.824$, df =1; $p < 0.05$. The p value, as calculated from this statistic is significant. Therefore reject HO and accept HA. Habitat and order affect counts for Date 1. Since the interaction term is significant we cannot proceed to evaluate the individual effects from this model. We will not proceed to break this down further, for I already know that count is affected by order and am not interested is proving this relationship. The summary statistics for Order under Date 1 (Habitat 1) and Date 2 (Habitat 2) are as follows:

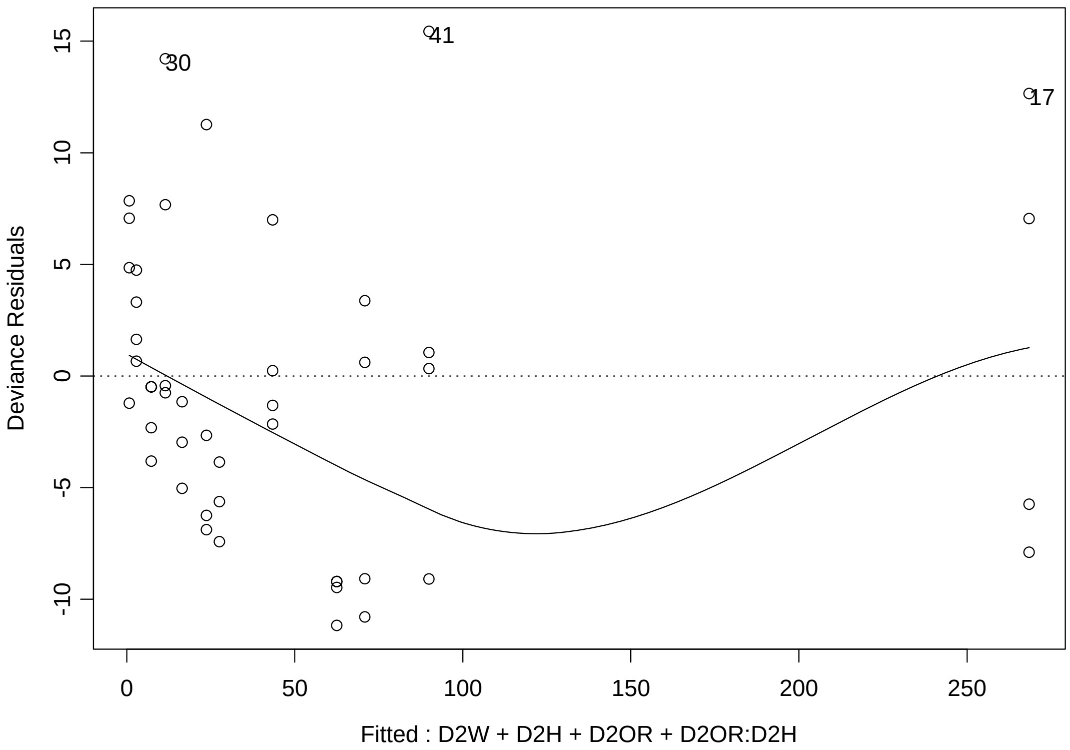
```
D1H1O:1
             D1H1C
     Min:  0.00000
    Mean: 16.37500
  Median: 16.50000
     Max: 39.00000
Std Dev.: 12.96079
-----------------------------------------------------------
D1H1O:2
            D1H1C
     Min: 0.00000
    Mean: 0.75000
  Median: 0.00000
     Max: 4.00000
Std Dev.: 1.38873
-----------------------------------------------------------
```

```
D1H1O:3
               D1H1C
     Min:   3.00000
    Mean:  43.12500
  Median:  22.50000
     Max: 131.00000
Std Dev.:  48.30687
-----------------------------------------------------------
D1H2O:1
              D1H2C
     Min: 10.00000
    Mean: 27.75000
  Median: 26.00000
     Max: 51.00000
Std Dev.: 13.78146
-----------------------------------------------------------
D1H2O:2
               D1H2C
     Min:  0.000000
    Mean:  2.750000
  Median:  0.500000
     Max: 11.000000
Std Dev.:  4.527693
-----------------------------------------------------------
D1H2O:3
               D1H2C
     Min:   5.00000
    Mean:  39.25000
  Median:  26.50000
     Max: 118.00000
Std Dev.:  38.76578
-----------------------------------------------------------
```

Date 2 (July):

Model: $f = e^{\wedge}(ß_{ref}) + e^{\wedge}(ß_W * W) + e^{\wedge}(ß_H * H) + e^{\wedge}(ß_O * O) + e^{\wedge}(ß_{O*H} * O * H)$ + error.
Hypotheses: HA: $ß_{O*H} = 0$
HO: $ß_{O*H} = 0$
The statistic is G. The probability distribution is chi-square, and $\alpha = 0.05$.

Plot of deviance residuals:



The deviance residuals appear homogeneously distributed. Accept the model.

```
Analysis of Deviance Table

Poisson model

Response: D2CT

Terms added sequentially (first to last)
         Df Deviance Resid. Df Resid. Dev
    NULL                    47   5499.711
     D2W  1  904.842        46   4594.869
     D2H  1  109.332        45   4485.537
    D2OR  1 1629.430        44   2856.107
D2OR:D2H  1  686.235        43   2169.871


Chi-Square with 1 DF

      x  P( X <= x )
686.235            1
```

The p value is <0.0001 for $\Delta G = 686.235$, df =1; $p < 0.05$. The p value, as calculated from this statistic is significant. Therefore reject HO and accept HA. Habitat and Order affect counts for Date 2. Since the interaction term is significant we cannot proceed to evaluate the individual effects from this model. We will not proceed to break this down further, for I already know that count is affected by order and am not interested is proving this relationship. The summary statistics for Order under Habitat 1 and Habitat 2 for Date 2 are as follows:

```
D2H1O:1
              D2H1C
     Min:  0.000000
    Mean:  8.750000
  Median:  9.000000
     Max: 15.000000
Std Dev.:  5.257647
-----------------------------------------------------------
D2H1O:2
              D2H1C
     Min:  0.000000
    Mean:  3.500000
  Median:  3.000000
     Max: 10.000000
Std Dev.:  3.664502
-----------------------------------------------------------
D2H1O:3
            D2H1C
     Min:   3.000
    Mean: 176.625
  Median: 125.500
     Max: 501.000
Std Dev.: 179.714
-----------------------------------------------------------
D2H2O:1
             D2H2C
     Min:  9.00000
    Mean: 44.87500
  Median: 40.00000
     Max: 97.00000
Std Dev.: 32.32618
-----------------------------------------------------------
```

```
D2H2O:2
              D2H2C
     Min:  0.000000
    Mean:  4.625000
  Median:  5.500000
     Max: 12.000000
Std Dev.:  3.925648
------------------------------------------------------------
D2H2O:3
              D2H2C
     Min:   0.00000
    Mean:  74.25000
  Median:  56.00000
     Max: 272.00000
Std Dev.:  91.37325
------------------------------------------------------------
D2H2O:NA
```

I now look at the third term in the 3-way interaction, which is order. This will allow me to see if habitat and date were interactive for each of the three orders.

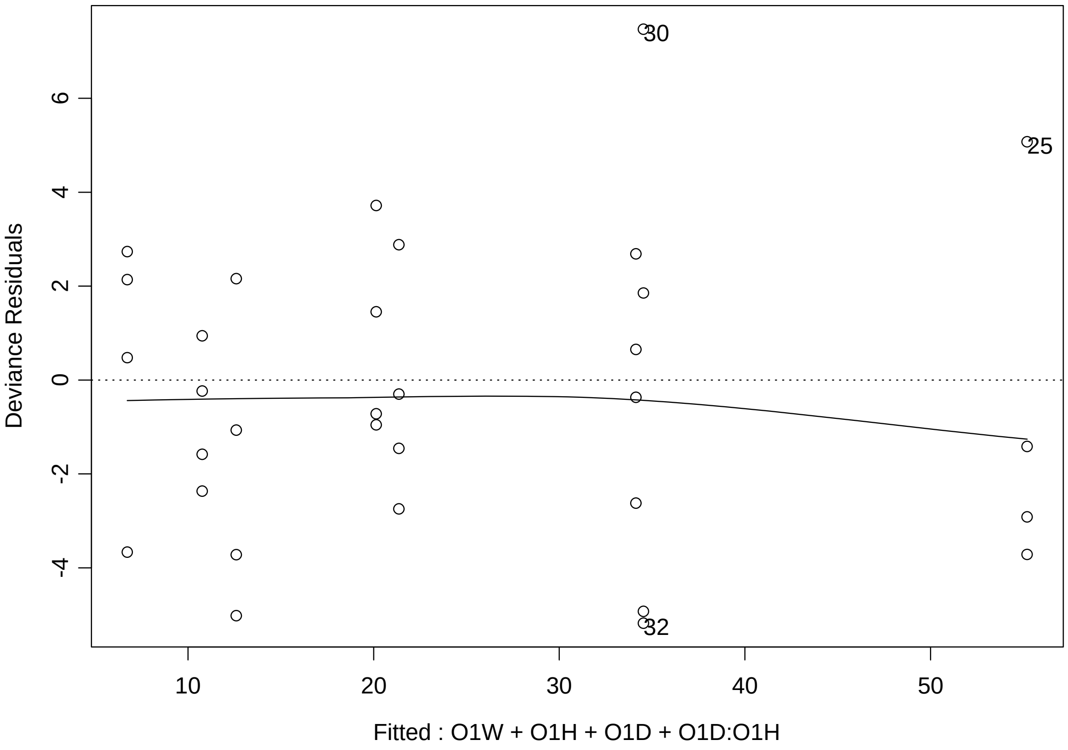
Order 1 (Mayflies):

Model: *f* = $e^\wedge(ß_{ref})$ +$e^\wedge(ß_W$* W) +$e^\wedge(ß_H$*H)+$e^\wedge(ß_D$*D) + $e^\wedge(ß_{D*H}$*D*H) + error.

Hypotheses: HA: $ß_{D*H} = 0$
HO: $ß_{D*H} = 0$

The statistic is G. The probability distribution is chi-square, and $\alpha = 0.05$.

Plot of deviance residuals:



There is no cone shape. The residuals are homogeneously distributed. The model is deemed appropriate.

```
Analysis of Deviance Table

Poisson model

Response: O1C

Terms added sequentially (first to last)
        Df Deviance Resid. Df Resid. Dev
   NULL                    31   576.2054
    O1W  1  41.8061        30   534.3993
    O1H  1 192.7103        29   341.6890
    O1D  1   7.3979        28   334.2911
O1D:O1H  1  44.0206        27   290.2705
```

The p value is <0.0001 for $\Delta G = 44.0206$, df =1; $p < 0.05$. The p value, as calculated from this statistic is significant. Therefore reject HO and accept HA. Habitat and date affect counts for Order 1 (Mayflies). Since the interaction term is significant we cannot proceed to evaluate the individual effects from this model. I will evaluate date under each of the two habitats and habitat under each of the two dates. However, these results will be presented all together, in the interest of avoiding excessive repetition (all results were significant). All plots were homogeneously distributed with no cone shape, rendering the model suitable.

For Dates 1 and 2:
Model: $f = e^\wedge(ß_{ref}) + e^\wedge(ß_W * W) + e^\wedge(ß_H * H)$ + error.

For Habitats 1 and 2:
Model: $f = e^\wedge(ß_{ref}) + e^\wedge(ß_W * W) + e^\wedge(ß_D * D)$ + error.

Hypotheses:

For Dates 1 and 2:

HA: $ß_H = 0$
HO: $ß_H = 0$

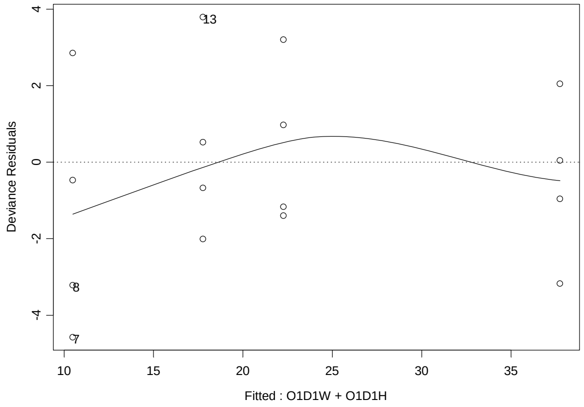
For Habitats 1 and 2:

HA: $ß_D = 0$
HO: $ß_D = 0$

The statistic is G. The probability distribution is chi-square, and $\alpha = 0.05$.

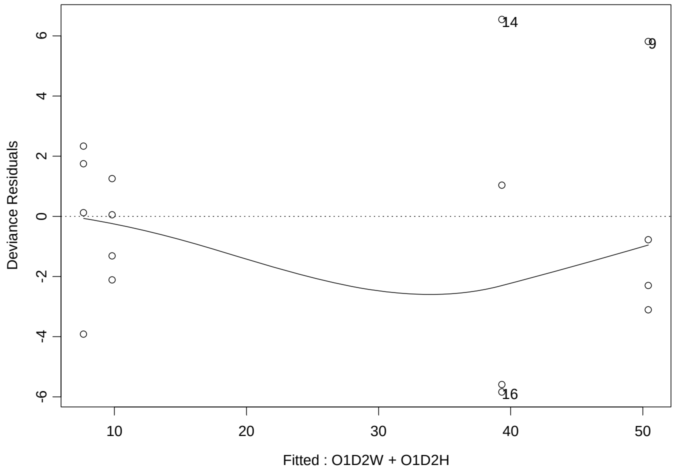
Plots of deviance residuals:

For Order 1, Date 1(testing for the effect of habitat and watershed):



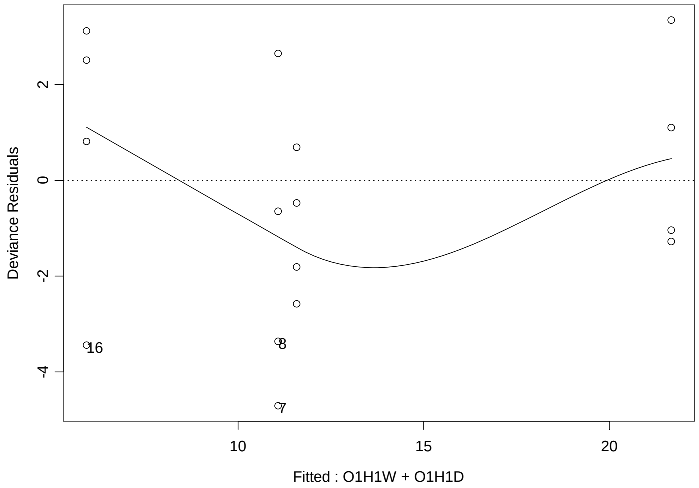
```
        *** Generalized Linear Model ***


Analysis of Deviance Table

Poisson model

Response: O1D1C

Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev
 NULL                     15   158.9739
O1D1W  1 46.73179         14   112.2421
O1D1H  1 23.72592         13    88.5162


Chi-Square with 1 DF

    x  P( X <= x )
23.72      1.00000

Chi-Square with 1 DF

    x  P( X <= x )
46.73      1.00000
```

For Order 1, Date 2 (testing for the effect of habitat and watershed):



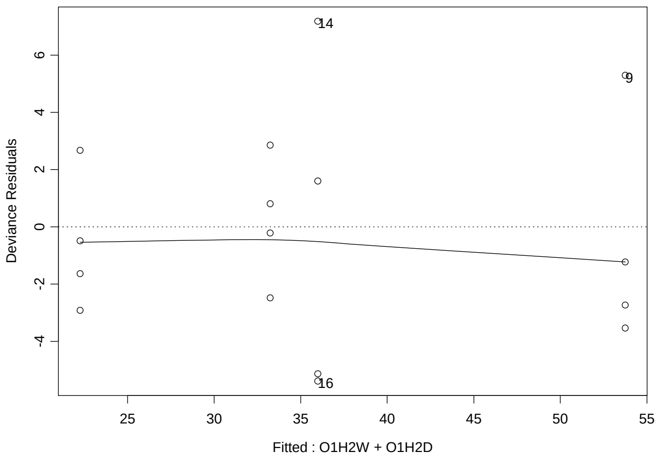
```
Analysis of Deviance Table

Poisson model

Response: O1D2C

Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev
 NULL                     15   409.8336
O1D2W  1   6.5645        14   403.2691
O1D2H  1 213.0051        13   190.2640

Chi-Square with 1 DF

     x  P( X <= x )
6.5645     0.989597
```

For Order 1, Habitat 1 (testing for the effect of date and watershed)



```
Analysis of Deviance Table

Poisson model

Response: O1H1C

Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev
 NULL                    15   135.3678
O1H1W  1 21.40251        14   113.9653
O1H1D  1 18.80763        13    95.1577
```

For Order 1, Habitat 2 (Testing for the effects of date)



```
Analysis of Deviance Table

Poisson model

Response: O1H2C

Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev
 NULL                   15   248.1273
O1H2W  1 22.91349       14   225.2138
O1H2D  1 32.61088       13   192.6029
```

The following is a summary for all results from these last analyses:

Test	ΔG	df	p	Decision
O1D1	H = 23.72	1	<0.0001	Significant. Reject HO, Accept HA
O1D1	W = 46.73	1	<0.0001	Significant. Reject HO, Accept HA
O1D2	H = 213.01	1	<0.0001	Significant. Reject HO, Accept HA
O1D2	W = 6.5645	1	0.0105	Significant. Reject HO, Accept HA
O1H1	D = 18.81	1	<0.0001	Significant. Reject HO, Accept HA
O1H1	W = 21.40	1	<0.0001	Significant. Reject HO, Accept HA
O1H2	D = 32.61	1	<0.0001	Significant. Reject HO, Accept HA
O1H2	W = 22.91	1	<0.0001	Significant. Reject HO, Accept HA

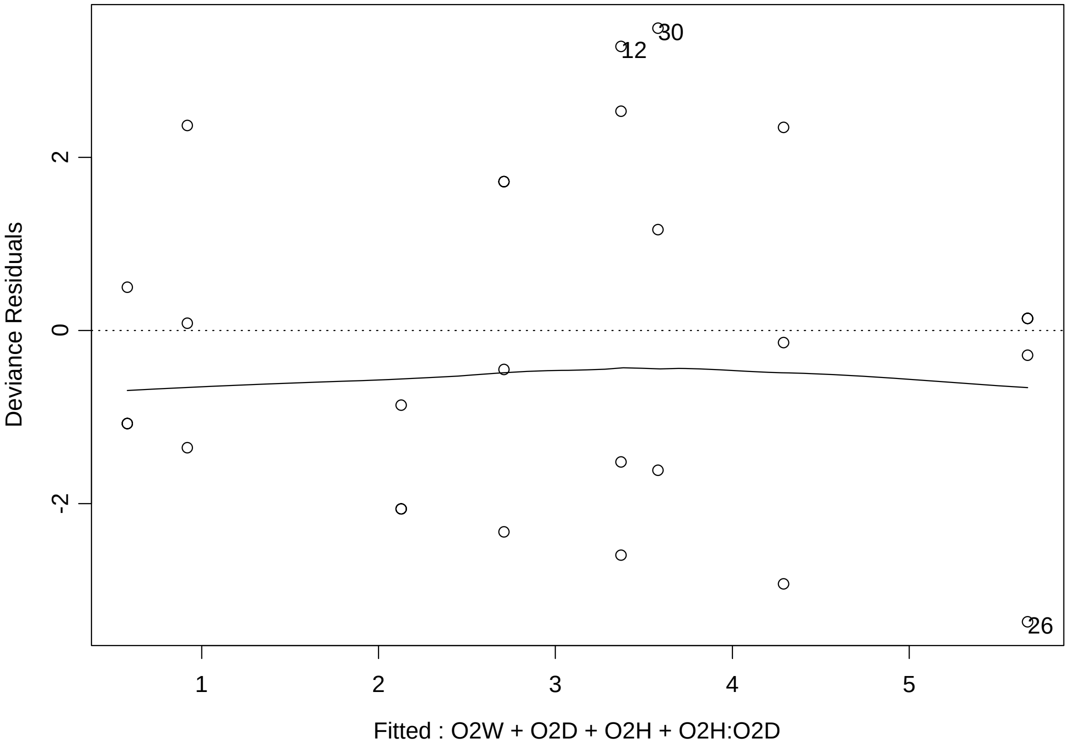
Now I will look at Order 2 (Stoneflies):

Model: $f = e^\wedge(\beta_{ref}) + e^\wedge(\beta_W * W) + e^\wedge(\beta_H * H) + e^\wedge(\beta_D * D) + e^\wedge(\beta_{D*H} * D * H) + \text{error}$.

Hypotheses: HA: ß$_{D*H}$ = 0
HO: ß$_{D*H}$ = 0

The statistic is G. The probability distribution is chi-square, and α = 0.05.

Plot of deviance residuals:



There is some cone shape but it is not large enough to try a new model. The model is deemed acceptable.

```
Analysis of Deviance Table

Poisson model

Response: O2C

Terms added sequentially (first to last)
        Df Deviance Resid. Df Resid. Dev
   NULL                    31   422.7188
    O2W  1 13.78125        30   408.9375
    O2H  1 19.53125        29   389.4062
    O2D  1 42.78125        28   346.6250
O2D:O2H  1  1.53125        27   345.0938

Chi-Square with 1 DF

     x  P( X <= x )
1.5325     0.784262
```

The p value, as calculated from this statistic is not. The p value is 0.2157.for ΔG = 1.5325, df =1; $p > 0.05$. The interaction term is not significant (Thank goodness ☺). Therefore do not reject HO and do not accept HA. Habitat and date no not interact for Order 2 (Stoneflies). Since the interaction term is not significant we can proceed to evaluate the individual effects from this model. They are as follows

HA: $ß_W = 0$
HO: $ß_W = 0$

HA: $ß_H = 0$
HO: $ß_H = 0$

HA: $ß_D = 0$
HO: $ß_D = 0$

Term	ΔG	df	p
Watershed	13.78	1	<0.0001
Habitat	19.53	1	<0.0001
Date	47.78	1	<0.0001

Reject all HO. Accept all HA. All factors are individually significant in Order 2.

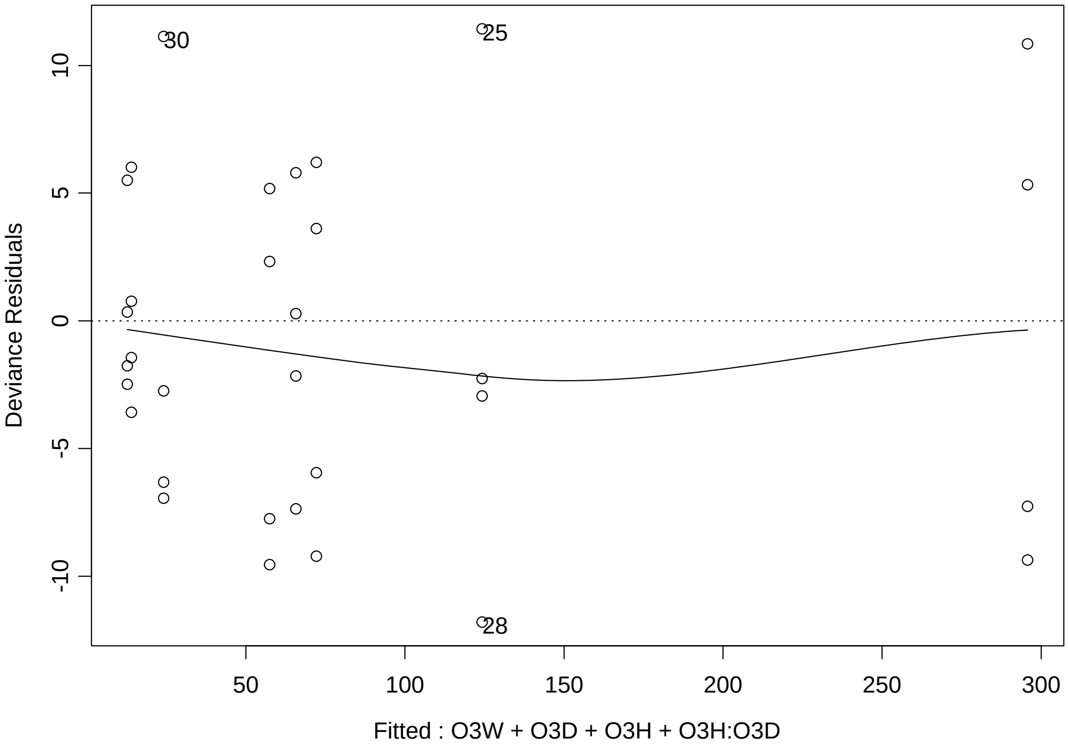
For Order 3 (Caddisfly):

Model: $f = e\hat{}(ß_{ref}) + e\hat{}(ß_W * W) + e\hat{}(ß_H * H) + e\hat{}(ß_D * D) + e\hat{}(ß_{D*H} * D * H) + error$.

Hypotheses: HA: $ß_{D*H} = 0$
HO: $ß_{D*H} = 0$

The statistic is G. The probability distribution is chi-square, and $\alpha = 0.05$.

Plot of deviance residuals:



The deviance residuals are fairly homogeneously distributed. The model is accepted

```
Analysis of Deviance Table

Poisson model

Response: O3C

Terms added sequentially (first to last)
        Df Deviance Resid. Df Resid. Dev
   NULL                    31   3715.730
    O3W  1 1327.020        30   2388.709
    O3D  1  714.082        29   1674.628
    O3H  1  275.794        28   1398.833
O3H:O3D  1   69.830        27   1329.003
```

The p value is <0.0001 for $\Delta G = 69.830$, df =1; $p < 0.05$. The p value, as calculated from this statistic is significant. Therefore reject HO and accept HA. Habitat and date affect counts for Order 3 (Caddisflies). Since the interaction term is significant we cannot proceed to evaluate the individual effects from this model. I will evaluate date under each of the two habitats and habitat under each of the two dates. Results are summarized in a table below the plots and deviance analyses.

For Dates 1 and 2:
Model: $f = e\hat{}(ß_{ref}) + e\hat{}(ß_W * W) + e\hat{}(ß_H * H)$ + error.

For Habitats 1 and 2:
Model: $f = e\hat{}(ß_{ref}) + e\hat{}(ß_W * W) + e\hat{}(ß_D * D)$ + error.

For Dates 1 and 2:

HA: $ß_H = 0$
HO: $ß_H = 0$

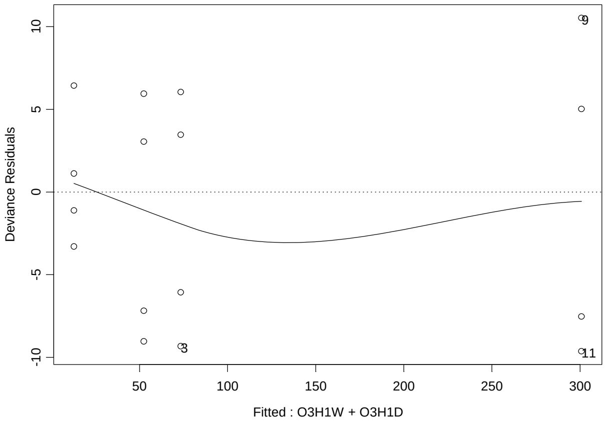
For Habitats 1 and 2:

HA: $ß_D = 0$
HO: $ß_D = 0$

The statistic is G. The probability distribution is chi-square, and $\alpha = 0.05$.

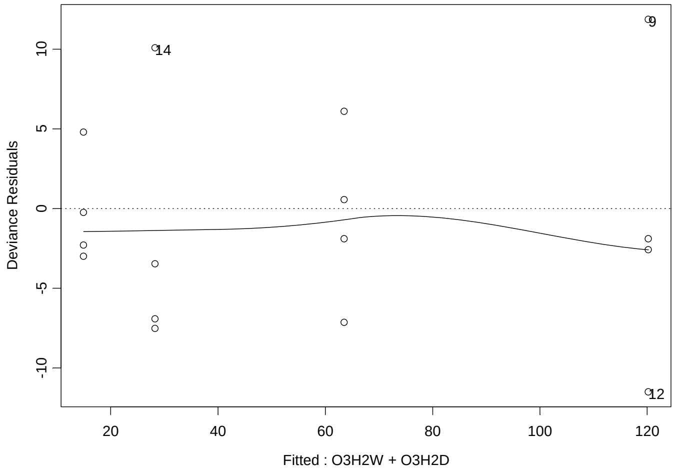
Plots of deviance residuals:

Habitat 1:



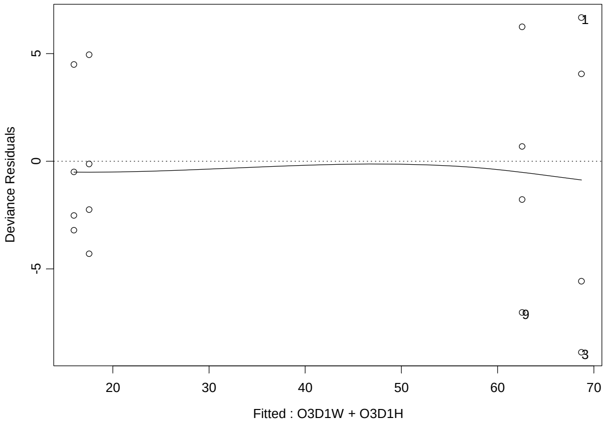
```
Analysis of Deviance Table

Poisson model

Response: O3H1C

Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev
 NULL                    15   2347.165
O3H1W  1 960.2598        14   1386.906
O3H1D  1 696.1452        13    690.760
```

Habitat 2:



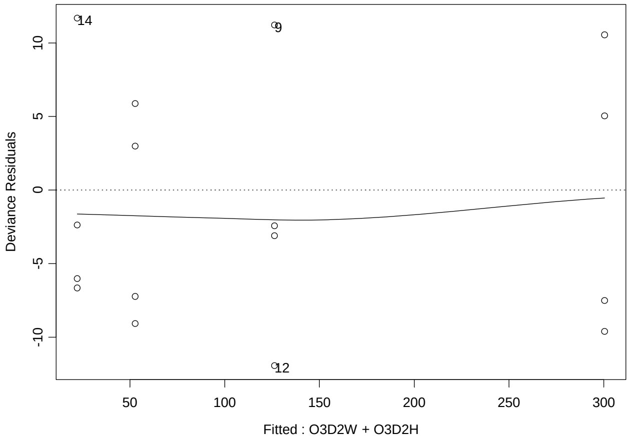
```
Analysis of Deviance Table

Poisson model

Response: O3H2C

Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev
 NULL                    15   1092.770
O3H2W  1 374.3840        14    718.386
O3H2D  1  87.7669        13    630.619
```

Date 1:



```
Analysis of Deviance Table

Poisson model

Response: O3D1C

Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev
 NULL                   15   597.5166
O3D1W  1 247.9836       14   349.5330
O3D1H  1   1.4588       13   348.0741
```

Date 2:



```
Analysis of Deviance Table

Poisson model

Response: O3D2C

Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev
 NULL                    15   2404.131
O3D2W  1 1089.185        14   1314.946
O3D2H  1  344.166        13    970.780
```

Results are as follows:

Test	ΔG	df	p	Decision
O3D1	H = 1.4588	1	0.2271	Not Significant. Do not reject HO
O3D1	W = 247.98	1	<0.0001	Significant. Reject HO, Accept HA
O3D2	H = 344.17	1	<0.0001	Significant. Reject HO, Accept HA
O3D2	W = 1089.19	1	0.0105	Significant. Reject HO, Accept HA
O3H1	D = 696.15	1	<0.0001	Significant. Reject HO, Accept HA
O3H1	W = 960.26	1	<0.0001	Significant. Reject HO, Accept HA
O3H2	D = 87.77	1	<0.0001	Significant. Reject HO, Accept HA
O3H2	W = 374.38	1	<0.0001	Significant. Reject HO, Accept HA

Discussion and Conclusions:

It is evident form the analysis that there was a lot of interaction within this model. When broken down into simpler models, almost all factors proved to be significant. Of particular importance is the watershed factor, which was assumed to be a random variable. However, in every case within this analysis, watershed proved to affect counts significantly. This shows that the two watershed sampled within Terra Nova National Park may are significantly different. As emphasized by Hauer and Lamberti (1996), one must be extremely careful when setting scale boundaries to include replicate watersheds.

In Newfoundland the macroinvertebrate diversity is low, which further exemplifies the importance of small differences across region (Lomond and Colbo 2000)

As expected, habitats 1 and 2 (upstream and downstream) were significantly different for all orders. This is not surprising. For example, generally, caddisflies prefer an upstream environment, while mayflies prefer a more lotic, downstream environment (Wetzel 2001). In the June sampling, the caddisflies did not differ significantly between habitat. This may be explained by Hynes (1970), who show that caddisfly number differ less between upstream and downstream location in the early part of the summer. The data here suggests that downstream communities are unique from upstream communities.

Significant differences between dates are logical. Not only to the number of insect in a habitat change over the summer, but so do the numbers within each order. According to Colbo (2004), June samples were intended to capture insects that had overwintered and emerged in the spring. The July sample was went to capture insect which emerge in July or which are present all year round. Therefore significant differences were expected. It is interesting to note the interaction between date and the other factors. These results show that differences in watershed counts, habitat counts and order counts were significant in many case. In habitat 2 however, date did not affect order, and when looking at stoneflies as an order, date did not affect habitat. This may support Dr. Colbo's comments that the June sample was later than he had hoped for, and that the two samples may not have been far enough apart to capture significant differences.

It was expected that order counts would vary. Mayfly, stonefly and caddisfly ratios are often compared to evaluate richness, composition and diversity of aquatic ecosystems (Lomond and Colbo 2000). This may have contributed to many significant interactions within my analyses. However, evaluating significance for habitat and date without the order interaction still yielded significant results, and in some cases the delta G value for order was less than that for habitat and date.

This is obviously a very complex data set. When broken down to one-way tests, there was often limited data to work with. A larger data set would be advantageous. However, based on these results, it appears that the two watershed differ significantly. Differences across region were detected despite typically low richness of aquatic insects in Newfoundland. The habitats show unique populations and there are some differences in community structure for sampling dates. The terms of this model were extremely interactive, suggesting a very complex ecological system.

References:

Colbo, M. 2004. Personal communication. Email on November 29, 2004 and December 2, 2004.

Hauer, F.R. and Lambertini, G.A. 1996. Methods in Stream Ecology. Academic Press. San Diego, California. pp 3-7, 339- 347.

Hynes, H.B.N. 1970. The Ecology of Running Water. Liverpool University Press. Bungay, Great Britain. pp 102.

Larson, D.J., and Colbo, M.H. 1983. The insects: biogeographics considerations. *In* Biogeography and ecology of the island of Newfoundland. *Edited by* R. South. Dr. W. Junk Publisher, The Hague. pp. 593-677.

Lomond, T.M., and Colbo, M.H. 2000. Variations in lake-outlet Ephemeroptera, Plecoptera, and Trichoptera communities amongst regions of eastern Newfoundland, Canada. Can. J. Zool. 78: 1536-1543.

Wetzel. 2001. Limnology: Lake and River Ecosystems. 3rd Edition. Academic Press. London. pp 696-697, 710.