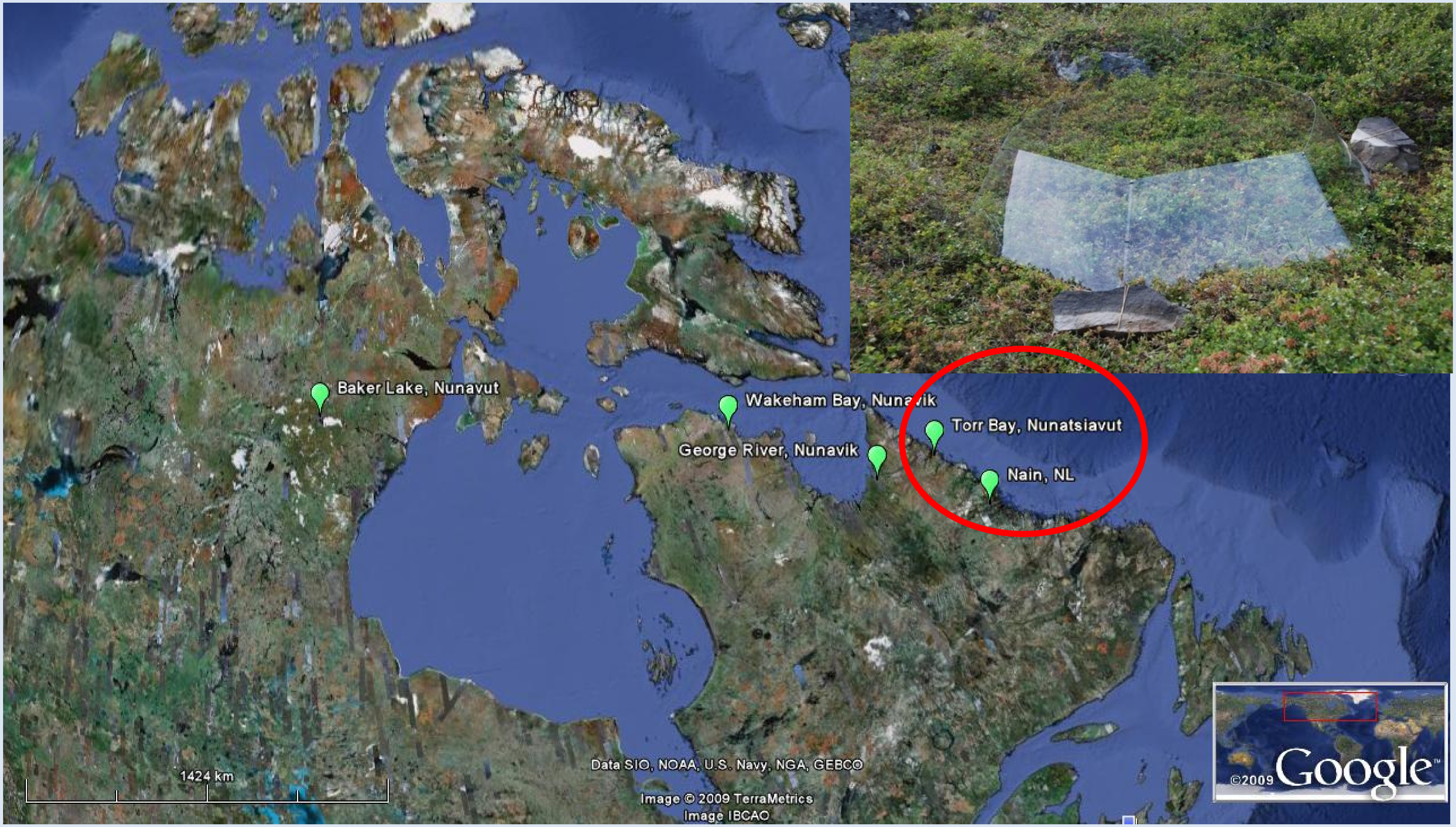
Investigating shrub associations with continuous vs. categorical cover data

Laura Siegwart-Collier
Department of Biology

Climate warming in the Arctic Tundra

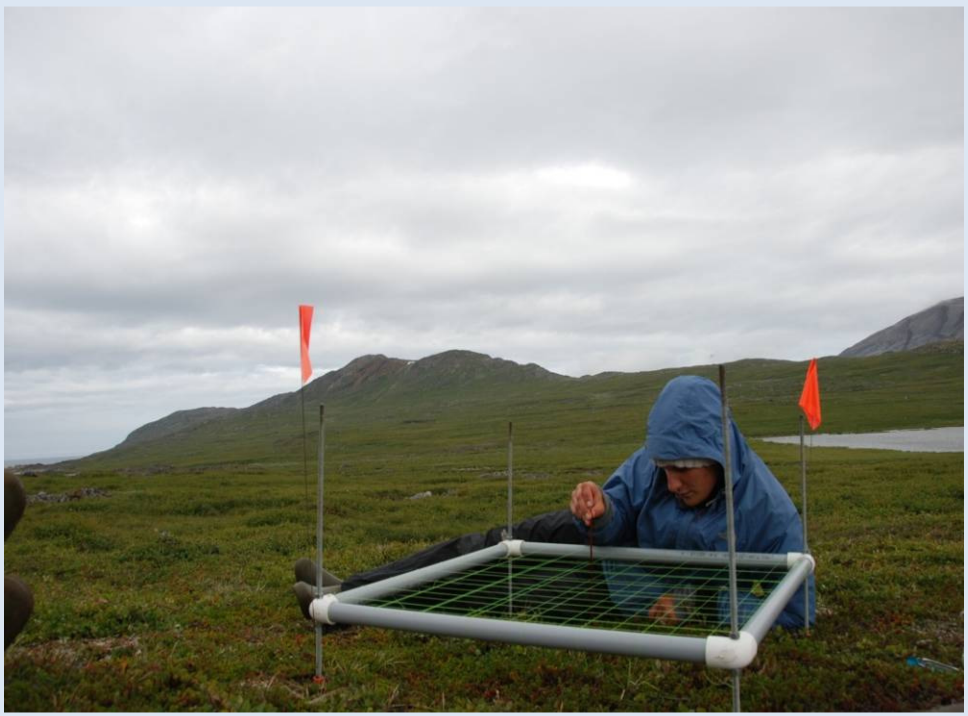
- **Importance of tundra ecosystems**
  - **Warming predictions**
  - **Shrub expansion and impacts**
- **Are berries a changing resource?**

Experimental warming sites



Baseline data

Continuous cover (%) OR ordinal cover class ?



Percent Cover	Cover Class
0 %	0
> 0 ≤ 5 %	1
6-10 %	2
11-25 %	3
26-50 %	4
51-75 %	5
76-100 %	6
>100 %	7

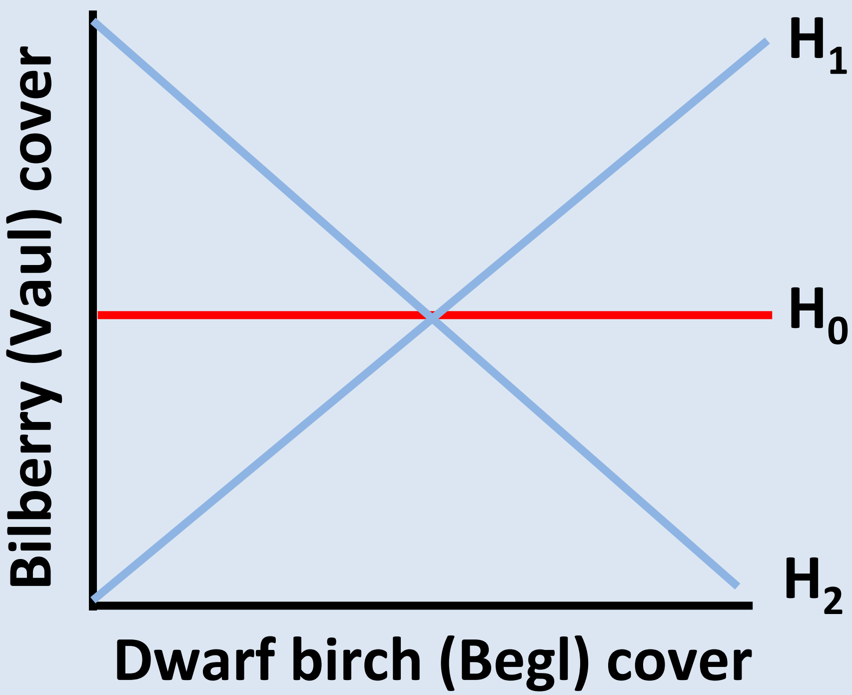
Research context

<u>Context</u> Long-term warming favours shrubs over ground cover, including lowbush *Vaccinium* (berries). Shrubs are thought to reduce greenhouse effects, relative to ground cover.

<u>Question</u> What is the effect of dwarf birch cover (Begl) and community (Nain vs Torr Bay) on bilberry cover before experimental warming?

About the data.....

<u>Question</u> What is the effect of dwarf birch cover (Begl) and community (Nain vs Torr Bay) on bilberry cover before experimental warming?



$H_0 : C_{Nain} = C_{TorrBay}$
$H_A : C_{Nain} \neq C_{TorrBay}$

Explanatory variables:

- Begl cover (% and CC; N = 78)
- Community; N=2

Response variable:

- Vaul cover (% and CC; N=78)

Model 1: Continuous response

- Quantitative (% cover)
- Gaussian (normal) error
- Identity link

Model:

$H = \mu_i + \varepsilon_i$

$\mu_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots \beta_p X_{ip}$

(same as two-way ANOVA, under general linear model)

Model 2: Multicategorical ordinal response

- Natural order (Cover class)
- Binomial family/extension of logistic regression
- Logit link; cumulative probabilities
- Assumes equal effect
- Fit using maximum likelihood

Proportional odds model expression:

$L_j(x_i) = \log (F_j(x_i)/1\text{-}F_j(x_i)) = \beta_{0j} + \beta_{1j}X_{1i} + \beta_{2j}X_{2i} + \ldots\beta_{pj}X_{pi} + \varepsilon_i$

For j = 1,2 ....., (k-1)

Descriptive statistics

Correlation matrix between Begl (shrub birch cover) and Vaul (Bilberry growing on ground....

Continuous cover (%)

Species	Begl	Vaul
Begl	1.00	-0.36
Vaul	-0.36	1.00

Cover Class

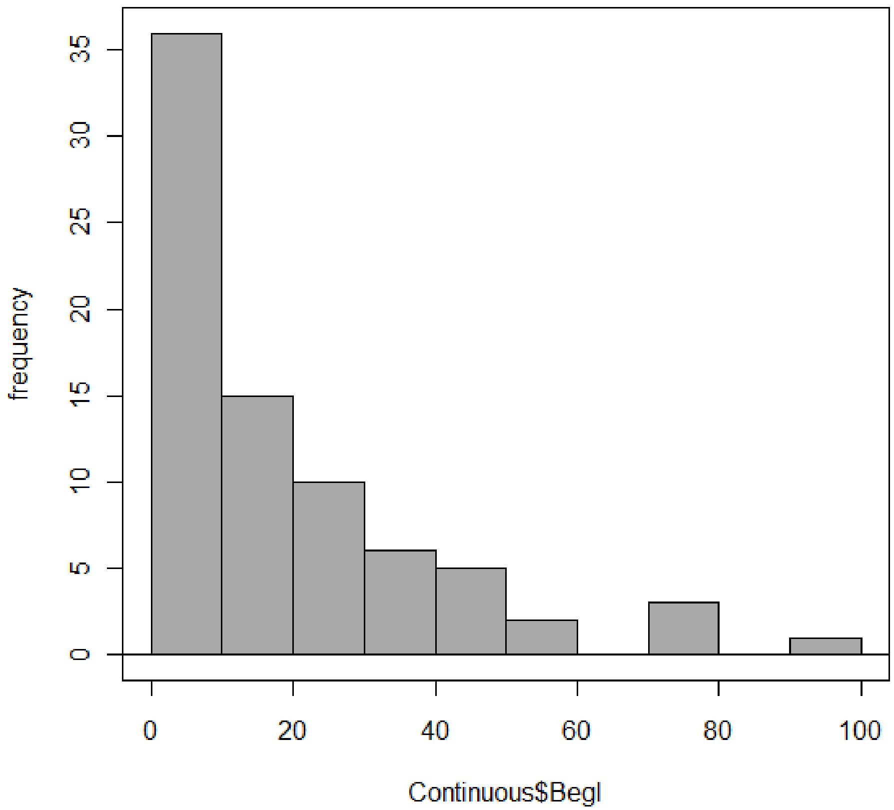
	BeglCC	VaulC
BeglCC	1.00	-0.34
VaulCC	-0.34	1.00

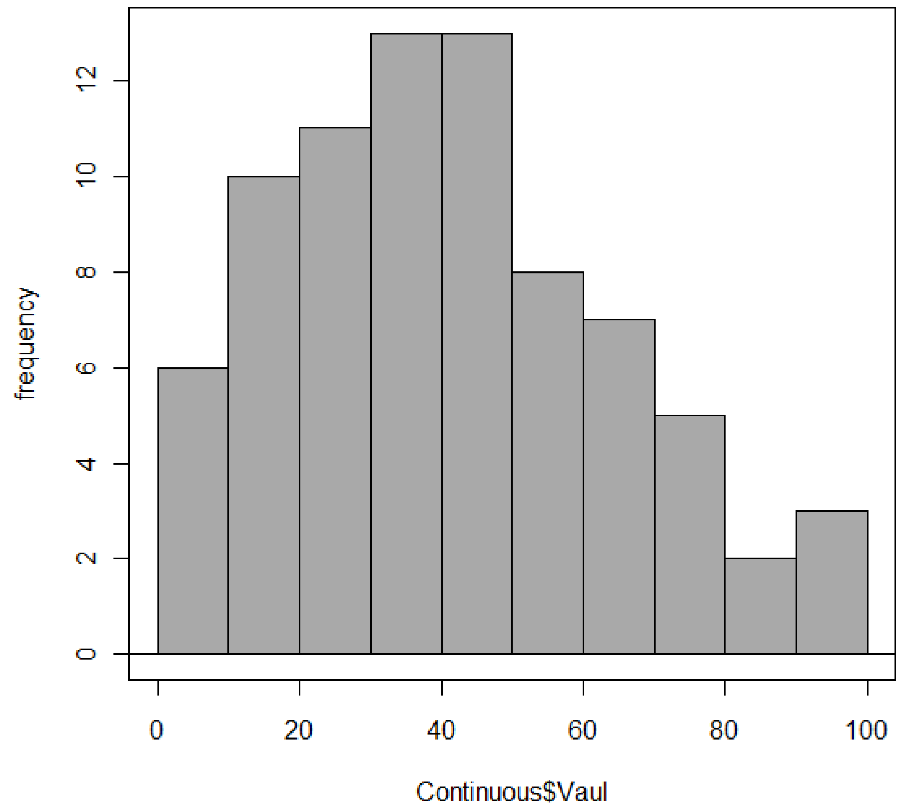
R-Code: > rcorr.adjust(Cover [,c("BeglCC","VaulCC")], type="pearson")

Shapiro-Wilk test of normality……

Explanatory	Df	W	P-Value
Begl %	1/76	0.8174	2.04e-08

Response	Df	W	P-Value
Vaul %	1/76	0.9716	0.07882

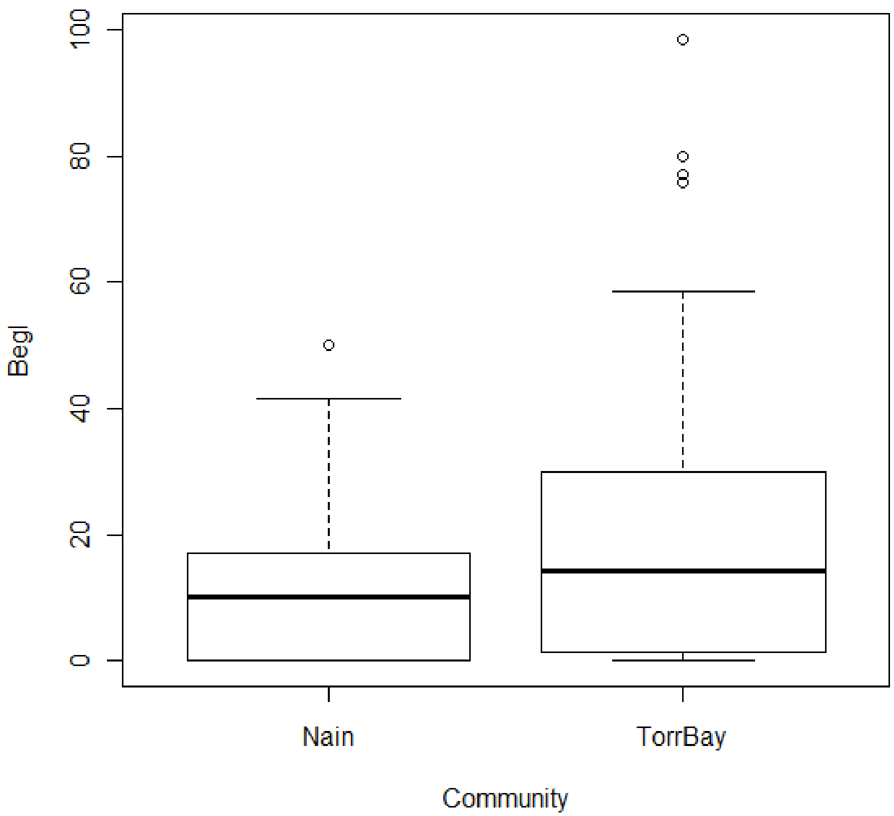


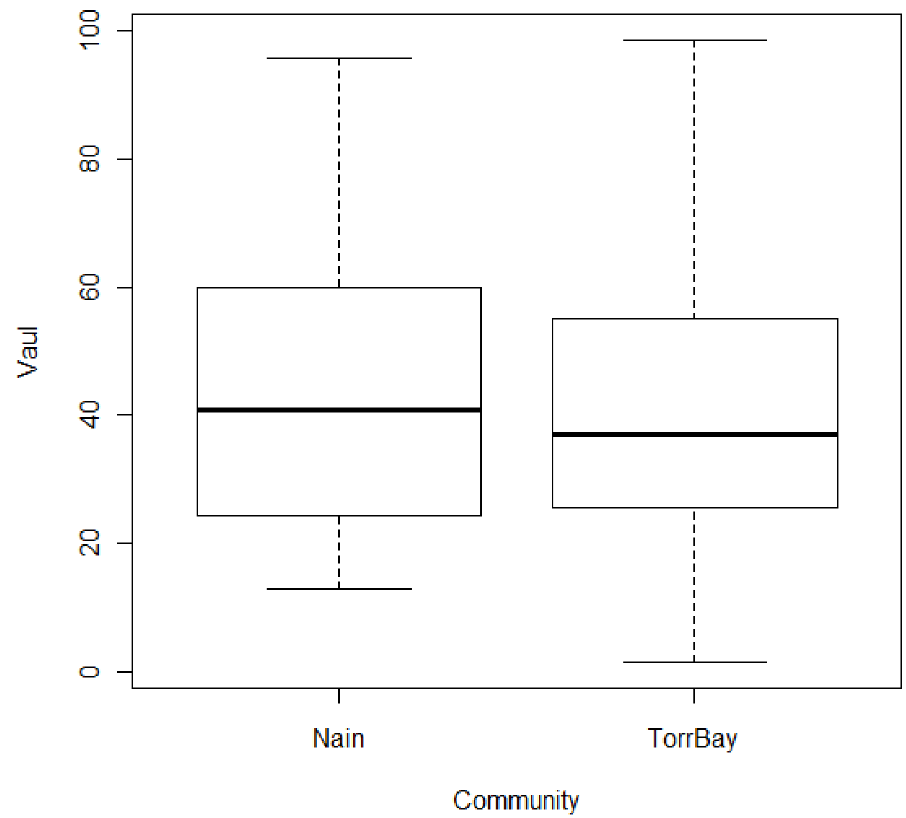


R-Code: > shapiro.test(Seminar1$x)

Levene's variance test..........

Explanatory	Df	F-Value	Pr(>F)	Response	Df	F-Value	Pr(>F)
Begl %	1/76	1.1823	0.2803	Vaul %	1/76	0.0208	0.8857





R Code: > levene.test(Seminar1$x, Seminar1$factor)

Descriptive statistics

Conclusions

- Overall, there is a negative relation between bilberry and birch (shrub)
- Response variable (Vaul = bilberry) shows exponential distribution
- Explanatory variable (Begl = birch) shows slightly right skewed distribution
- Because response variable is exponentially distributed count, we might expect heterogeneous errors.

GLM 1: Continuous

R-Code: > GLM.1 <- glm(Vaul ~ Begl*Community, gaussian, data=Seminar1Raw)

Formal Model: $Vaul\% = \beta_0 + \beta_{Begl}Begl + \beta_{Community}Community + \beta_{Begl*Community}Begl*Community$

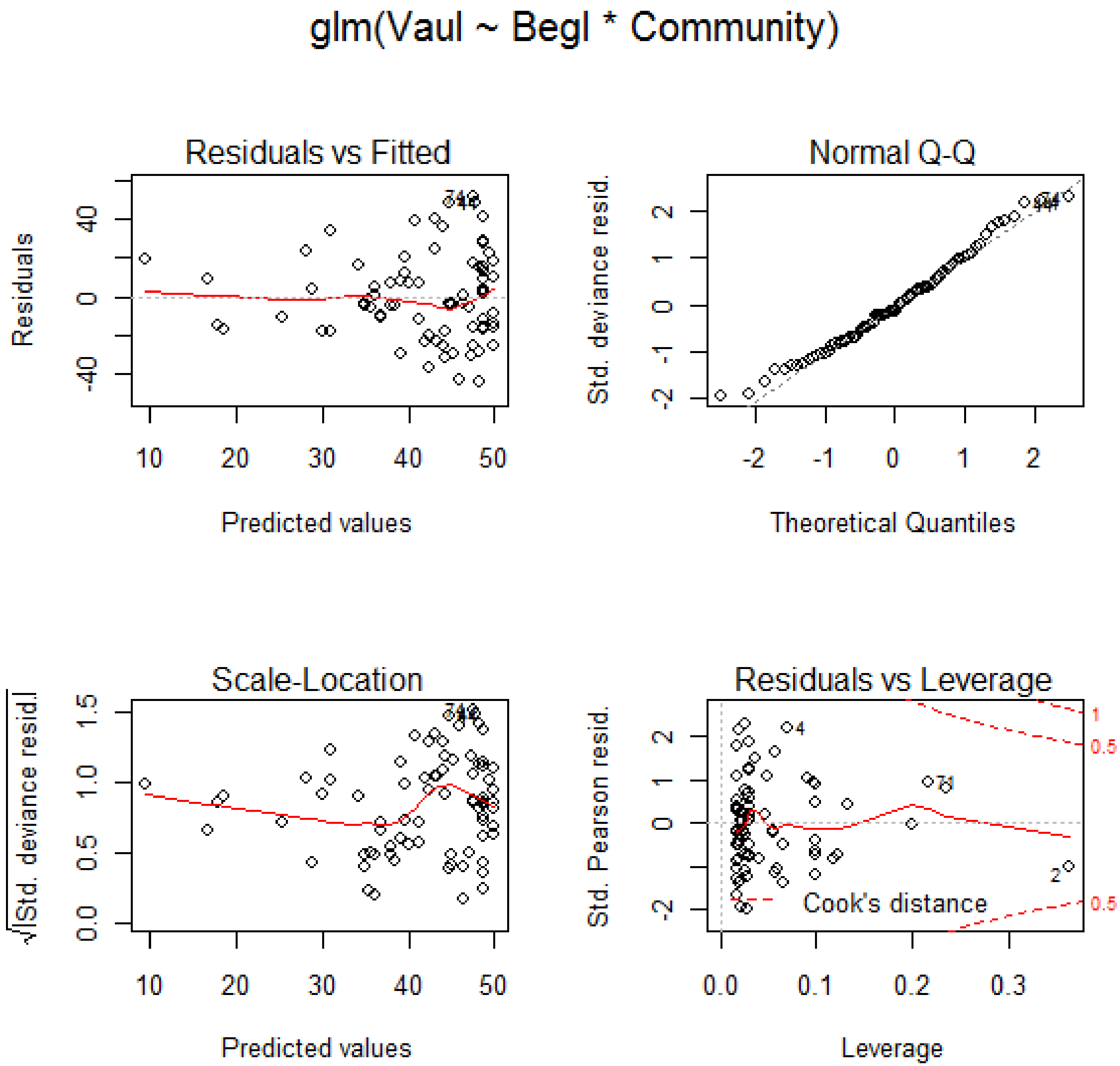
Deviance Residuals:

Min	1Q	Median	3Q	Max
-44.252	-16.307	-3.469	14.148	51.543

Coefficients	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	49.85491	7.05887	7.063	7.51e-10 ***
Begl	-0.37556	0.34227	-1.097	0.276
Community	-1.20486	8.04584	-0.150	0.881
Begl*Community	-0.02211	0.36525	-0.061	0.952

Dispersion parameter: 503.0409
Null deviance: 42950 on 77 degrees of freedom
Residual deviance: 37225 on 74 degrees of freedom
AIC: 712.46
Number of Fisher Scoring iterations: 2

Overdispersion!!



Plots indicate:

- Heterogeneous error
- At least 2-3 high leverage values

What happens if……

- Change the link?

– no change

- Remove leverage values and interaction term?

– weak

GLM 2: Continuous (leverages removed)

R-Code: > GLM.1 <- glm(Vaul ~ Begl+Community, gaussian, data=Seminar1Mod)

Formal Model: Vaul% = $\beta_0 + \beta_{Begl}\text{Begl} + \beta_{Community}\text{Community}$

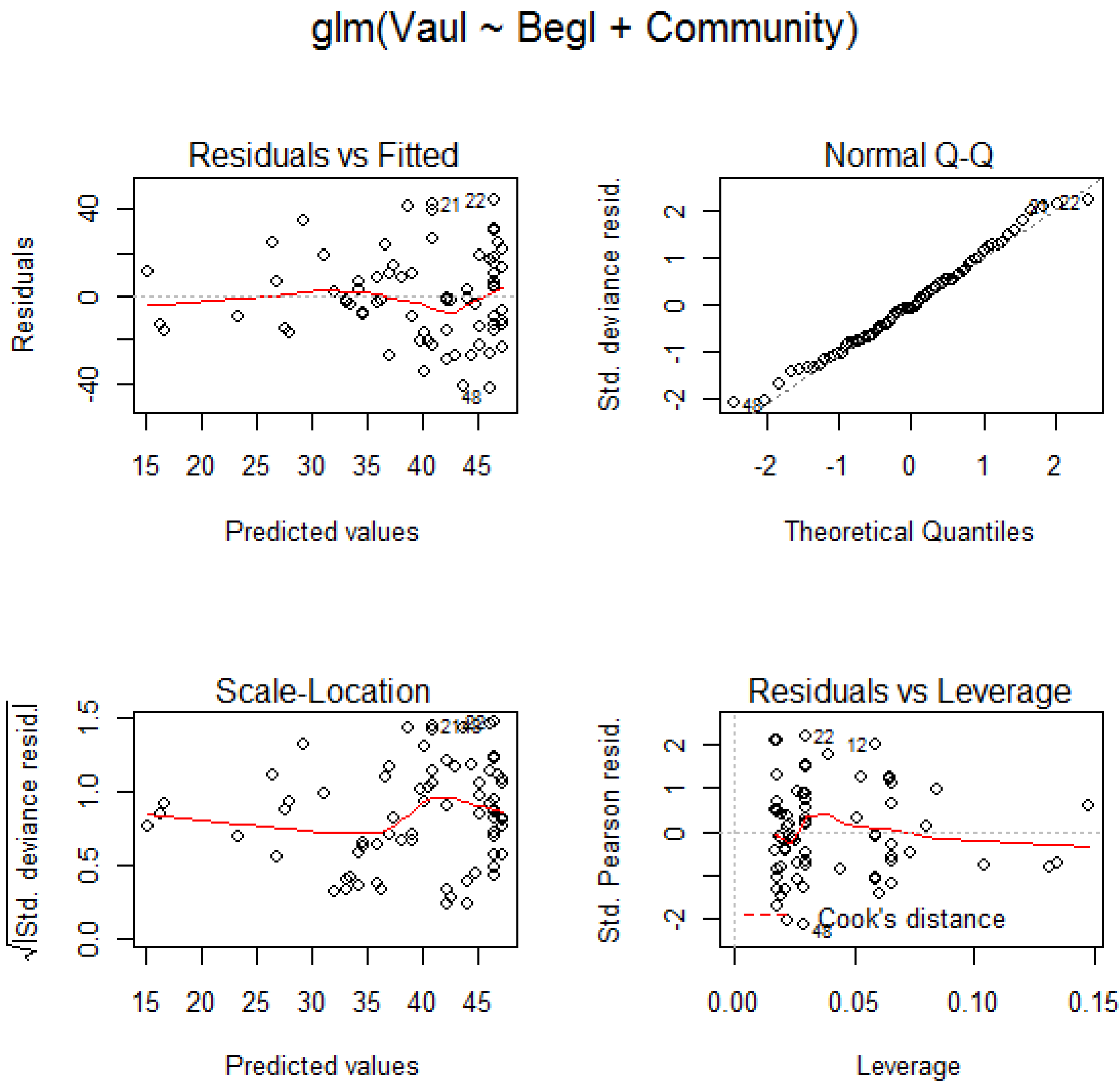
Deviance Residuals:

Min	1Q	Median	3Q	Max
-42.038	-14.564	-1.822	12.296	43.571

Coefficients	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	47.2294	5.1965	9.089	1.68e-13 ***
Begl	-0.3912	0.1200	-3.261	0.00171 **
Community	-0.8003	5.6297	-0.142	0.88736

Dispersion parameter: 410.2983
Null deviance: 33594 on 73 degrees of freedom
Residual deviance: 29131 on 71 degrees of freedom
AIC: 660.19
Number of Fisher Scoring iterations: 2

Poor model OR unmeasured variability?



Plots show:

- Heterogeneous error
- Influential points

What happens if……

- Change the link? – no change
- Change error structure– improved deviance

PropOdds 1: Proportional odds

Formal Model: $\text{logit}[P(Y \le j)] = \alpha_{j(7)} + \beta_{\text{Begl}}\text{Begl} + \beta_{\text{Community}}\text{Community}$

Residual Deviance: 217.8829

$R^2 = 0.178$

AIC: 241.8829

Coefficients	Value	Std. Error	t value	Wald Z	P-value
Begl Class 1	-0.67024170	0.6934556	-0.9665244	-0.97	0.3338
Begl Class 2	-2.30580738	1.0080601	-2.2873708	-2.29	0.0222
Begl Class 3	-1.47649932	0.6124566	-2.4107818	-2.41	0.0159
Begl Class 4	-1.27020795	0.5955910	-2.1326848	-2.13	0.0329
Begl Class 5	-2.74149918	1.3423741	-2.0422765	-2.04	0.0411
Begl Class 6	-3.13309510	1.2036705	-2.6029509	-2.60	0.0092
Community	0.03380395	0.4946482	0.1698115	0.17	0.8652

```
Library(MASS)
> ORM1 <- polr(VaulCC ~ BeglCC +  Community,
method="logistic", data=Seminar1, Hess=TRUE)
```

```
>Library(Design)
(OLR2<lrm(VaulCC~BeglCC+
Community,data=Laura))
```

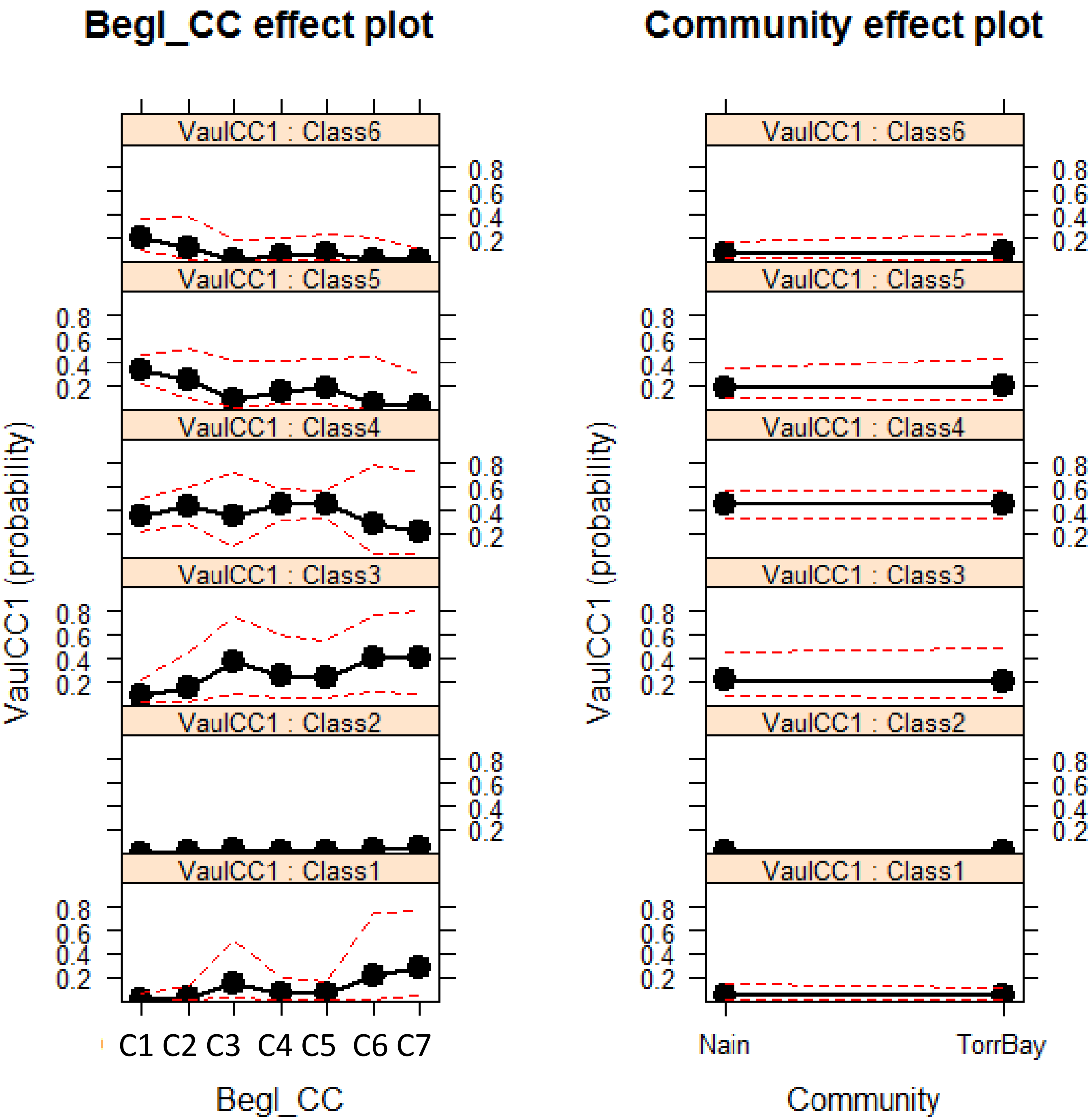
PropOdds 1: Proportional odds intercepts....

Cumulative logit	Value	Std. Error	t-value
Class1|Class2	-4.0190	0.7391	-5.4374
Class2|Class3	-3.8060	0.7123	-5.3429
Class3|Class4	-2.0655	0.5903	-3.4991
Class4|Class5	-0.1014	0.5316	-0.1908
Class5|Class6	1.4248	0.5832	2.4429

library(MASS)

Cumulative logit	Coefficient	Std. Error	Wald Z	P
y>=Class2	4.0190	0.7391	5.44	0.0000
y>=Class3	3.8060	0.7124	5.34	0.0000
y>=Class4	2.0655	0.5903	3.50	0.0005
y>=Class5	0.1014	0.5316	0.19	0.8487
y>=Class6	-1.4248	0.5832	-2.44	0.0146

library(Design)



Plots indicates:

- At low Birch cover (bottom panel) Increase in % cover by Bilberry with increase in Birch cover
- At high birch cover (top panels) decrease in bilberry with increase in birch.
- Community has no effect

What if we remove highly influential points?

PropOdds2: Proportional odds (highly influential points removed)

Formal Model: $\text{logit}[P(Y \leq j)] = \alpha_{j(7)} + \beta_{Begl}\text{Begl} + \beta_{Community}\text{Community}$

Residual Deviance: 194.8930
R^2 = 0.259
AIC: 218.8930

Coefficients	Value	Std. Error	t value	Wald Z	P-value
Begl Class 1	-1.4438179	0.7557413	-1.9104657	-1.91	0.0561
Begl Class 2	-3.3725031	1.0591110	-3.1842773	-3.18	0.0015
Begl Class 3	-1.6638144	0.6399771	-2.5998029	-2.60	0.0093
Begl Class 4	-1.4410459	0.6187642	-2.3289163	-2.33	0.0199
Begl Class 5	-3.0790585	1.4006119	-2.1983667	-2.20	0.0279
Begl Class 6	-4.6418434	1.4720168	-3.1533903	-3.15	0.0016
Community	0.2182535	0.5201189	0.4196224	0.42	0.6748

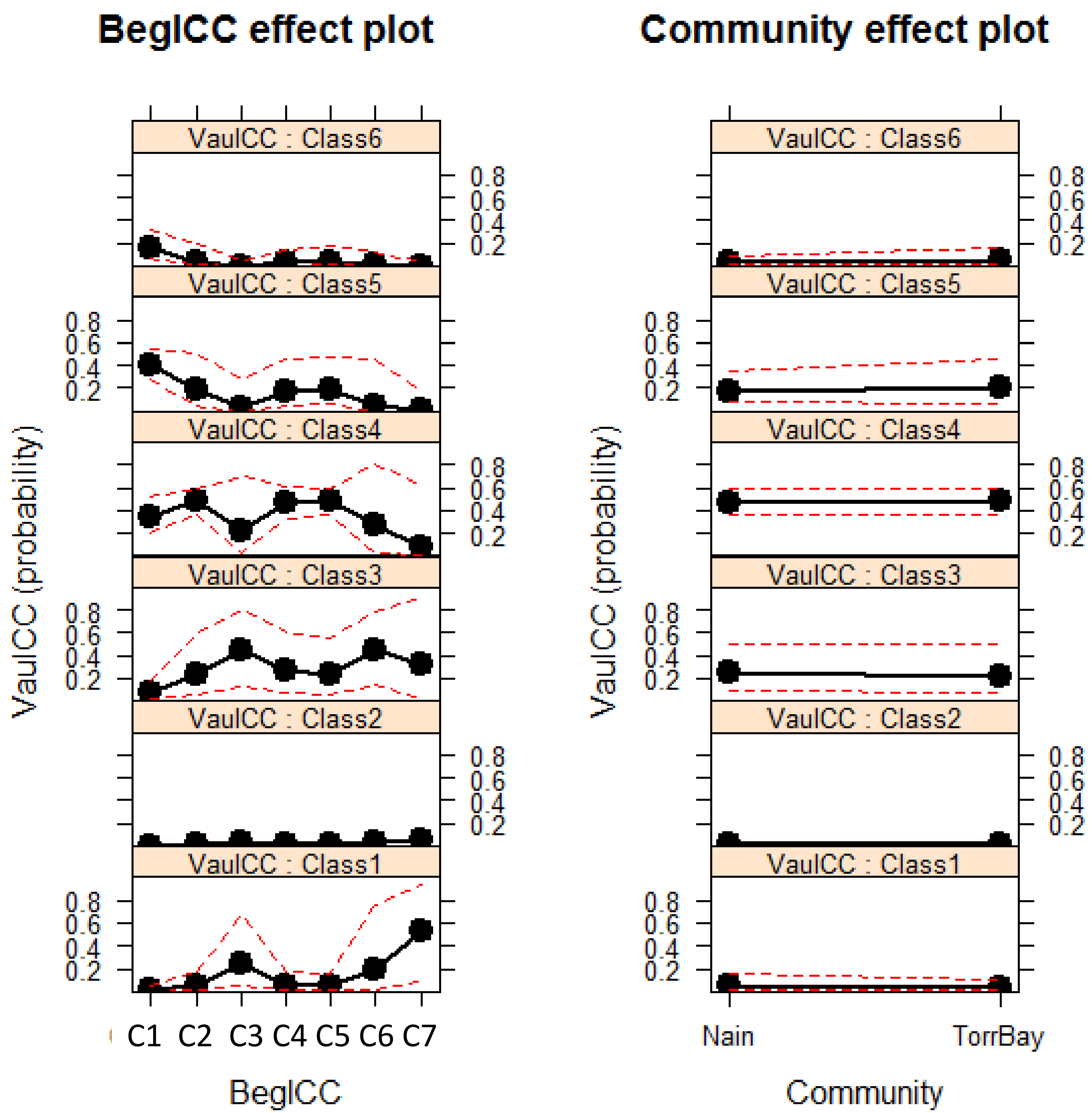
PropOdds2: Proportional odds intercepts (highly influential points removed)

Cumulative logit	Value	Std. Error	t-value
Class1|Class2	-4.3421	0.7924	-5.4799
Class2|Class3	-4.1146	0.7637	-5.3874
Class3|Class4	-2.1924	0.6192	-3.5408
Class4|Class5	-0.0991	0.5538	-0.1790
Class5|Class6	1.8406	0.6528	2.81194

Library(MASS)

Cumulative logit	Coefficient	Std. Error	Wald Z	P
y>=Class2	4.0190	0.7391	5.48	0.0000
y>=Class3	3.8060	0.7124	5.39	0.0000
y>=Class4	2.0655	0.5903	3.54	0.0004
y>=Class5	0.1014	0.5316	0.18	0.8579
y>=Class6	-1.4248	0.5832	-2.82	0.0048

Library(Design)


BeglCC effect plot
Community effect plot
VaulCC : Class6
VaulCC : Class5
VaulCC : Class4
VaulCC : Class3
VaulCC : Class2
VaulCC : Class1
0.8
0.6
0.4
0.2
VaulCC (probability)
C1 C2 C3 C4 C5 C6 C7
BeglCC
Nain
TorrBay
Community

Model comparison

Model Type	Residual Deviance	AIC/R^2
GLM 1 -normal error, identity link	37225 (df 74)	712.46
GLM 2 – as above influential points removed	29131 (df 71)	660.19
PropOdds 1 – proportional odds	217.88 (df 74?)	241.88/0.178
PropOdds 2 – proportional odds, influential points removed	198.89 (df = 71?)	218.89/0.259

- Begl: Reject H_0 and accept H_2 - Bilberry related to Birch cover
- Community: Accept H_0 - no difference in cover between the two locations.
- Influential (highly leveraged) points interpreted as strong microsite variability
- Ordinal model improves fit
- Unaccounted variability!! --additional biotic and abiotic variables ?

Important References

1. Agresti, A. 2002. Categorical data analysis, Second Edition, Wiley Series in Probability and Statistics.
2. Chatterjee, S. and A.S. Hadi. 2006. Regression Analysis by Example, Fourth Edition, Wiley Series in Probability and Statistics.
3. Crawley, M. J. 2007. The R Book, Wiley Publishing.
4. Thompson, L.A. 2007. R (and S-PLUS) Manual to Accompany Agresti's *Categorical Data Analysis (2002)* 2nd edition.