Maladaptive plastic responses of flowering time to geothermal heating (Cerastium 2)

Repeat and extend analyses done by Johan

Alicia Valdés

Contents

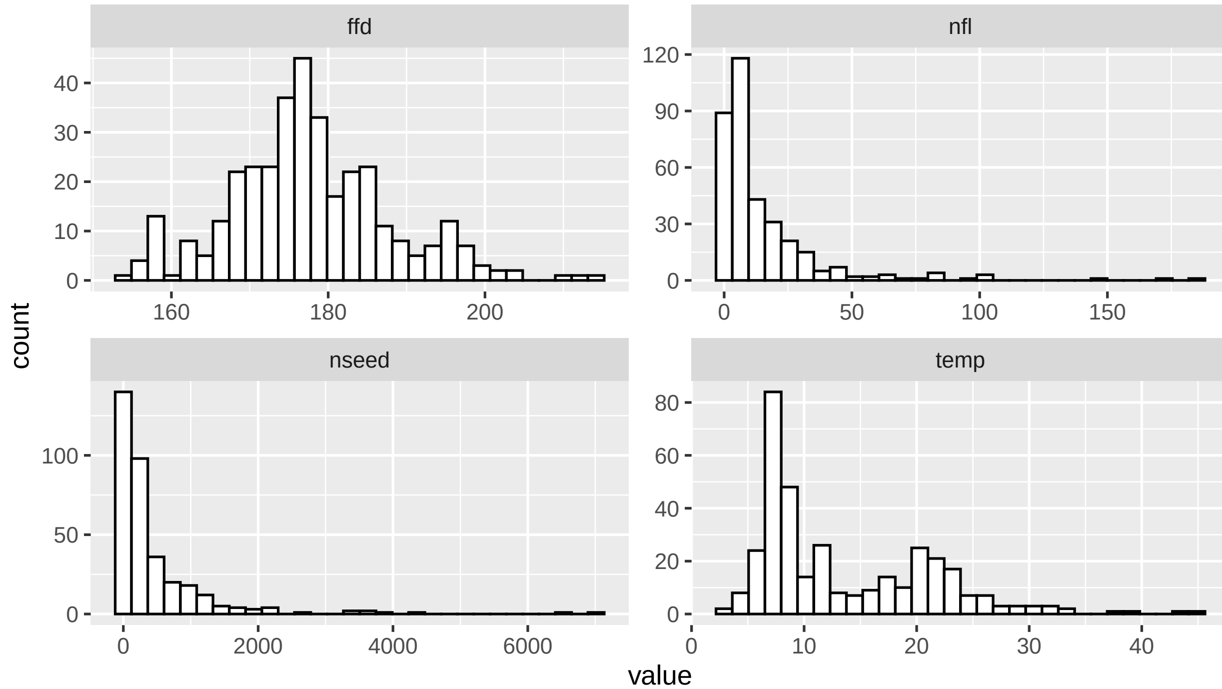
Data preparation

Load data, keep variables needed and merge

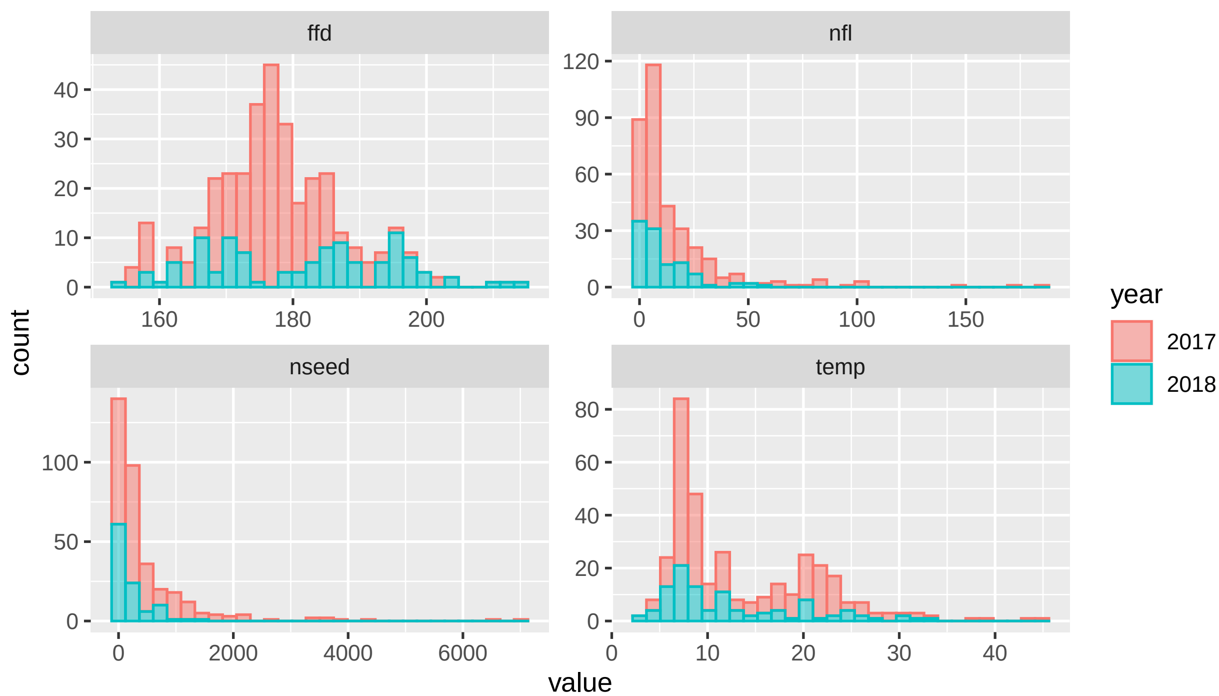
```
##        id id_original temp   ffd nfl    nseed year    ffd_std     nfl_std
## 1 2017_01          H1 19.4 172.0   6 269.3333 2017 -0.4912772 -0.34533654
## 4 2017_04          H5 31.0 163.5   4 106.0000 2017 -1.5514686 -0.71175481
## 6 2017_06          H7 29.4 170.0   4 105.0000 2017 -0.7407340 -0.71175481
## 7 2017_07          H8 29.0 170.0   9 150.0000 2017 -0.7407340  0.02108173
## 8 2017_08         H12 21.7 170.0   4 141.0000 2017 -0.7407340 -0.71175481
## 9 2017_09         H13 22.6 170.0   4 126.0000 2017 -0.7407340 -0.71175481
##   nseed_rel
## 1 0.4769565
## 4 0.1877131
## 6 0.1859422
## 7 0.2656317
## 8 0.2496938
## 9 0.2231307
```

Distributions

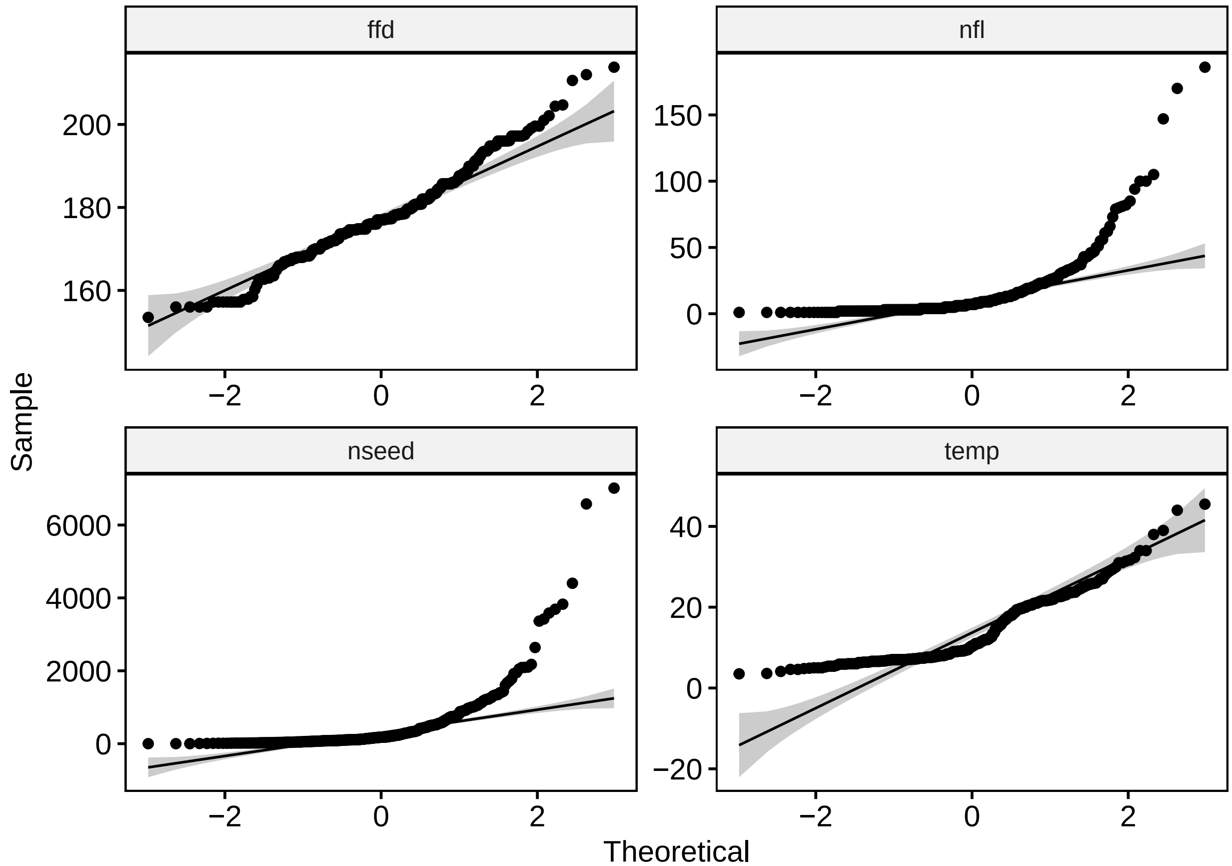
Histograms



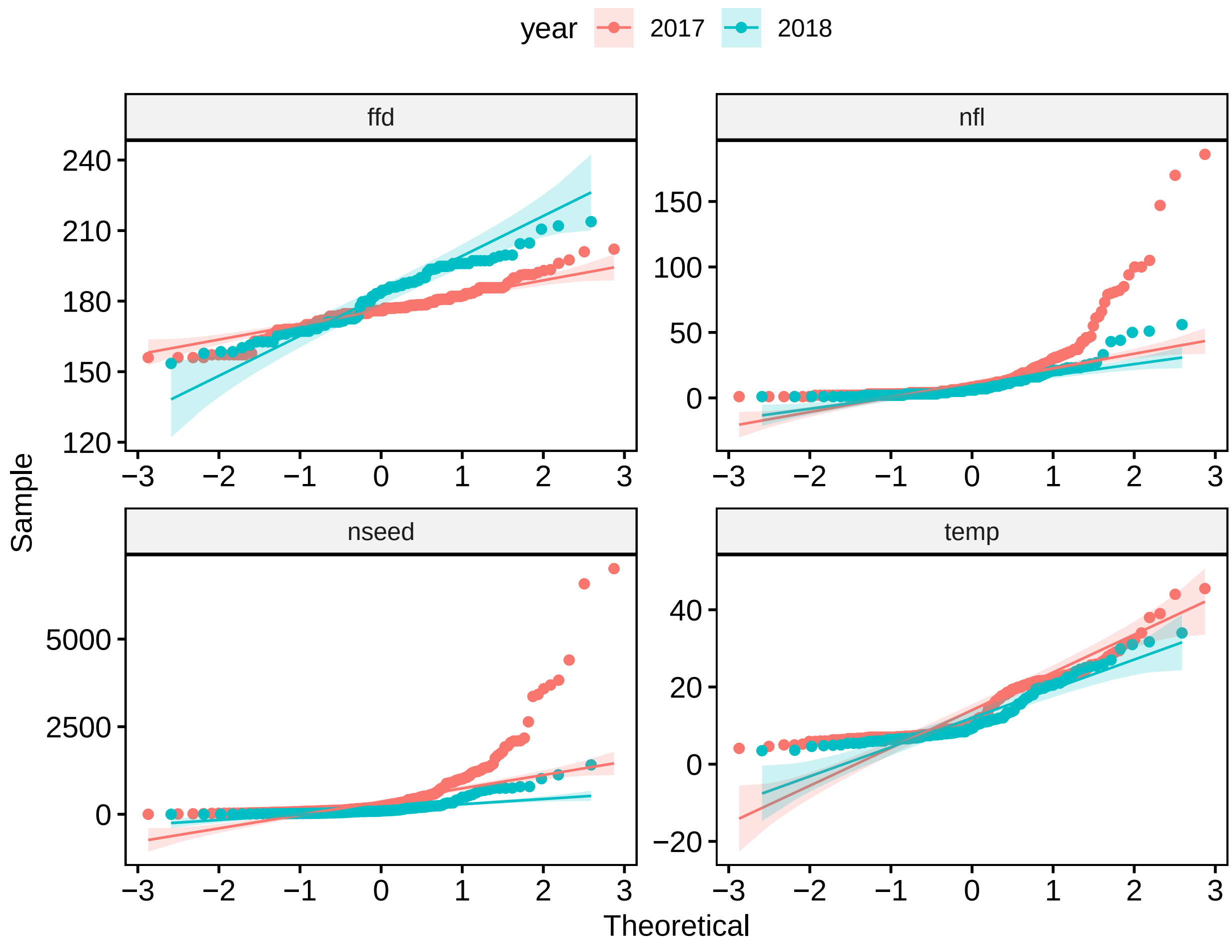
Histograms by year



QQplots



QQplots by year



1. Effect of temperature on FFD

Models as fit by Johan

Only linear

```
FFD_2017_1<-lm(ffd~temp+log(nfl),subset(mydata,year==2017))
summ(FFD_2017_1,vif=T,scale=T) # scale=T reports standardized coefs
```

Observations	245
Dependent variable	ffd
Type	OLS linear regression

F(2,242)	44.424
R^2	0.269
Adj. R^2	0.263

	Est.	S.E.	t val.	p	VIF
(Intercept)	175.939	0.440	399.972	0.000	NA
temp	-1.636	0.461	-3.549	0.000	1.093
‘log(nfl)‘	-3.371	0.461	-7.316	0.000	1.093

Standard errors: OLS; Continuous predictors are mean-centered and scaled by 1 s.d.

```
FFD_2018_1<-lm(ffd~temp+log(nfl),subset(mydata,year==2018))
summ(FFD_2018_1,vif=T,scale=T) # scale=T reports standardized coefs
```

Observations	104
Dependent variable	ffd
Type	OLS linear regression

F(2,101)	57.208
R^2	0.531
Adj. R^2	0.522

	Est.	S.E.	t val.	p	VIF
(Intercept)	181.876	0.926	196.402	0.000	NA
temp	-7.436	0.932	-7.983	0.000	1.002
‘log(nfl)‘	-6.276	0.932	-6.737	0.000	1.002

Standard errors: OLS; Continuous predictors are mean-centered and scaled by 1 s.d.

Similar results. I guess it is good to include number of flowers to evaluate the effect of temperature on phenology independent of number of flowers. But we didn’t do that in the GCB paper.

Questions: Try models without number of flowers? (also significant effects of temperature) Do we need to take log of number of flowers? Should we report standardized estimates (i.e. with scaled predictors)?

Models without number of flowers

```
FFD_2017_2<-lm(ffd~temp,subset(mydata,year==2017))
summ(FFD_2017_2,scale=T)
```

Observations	245
Dependent variable	ffd
Type	OLS linear regression

F(1,243)	29.049
R^2	0.107
Adj. R^2	0.103

	Est.	S.E.	t val.	p
(Intercept)	175.939	0.485	362.693	0.000
temp	-2.620	0.486	-5.390	0.000

Standard errors: OLS; Continuous predictors are mean-centered and scaled by 1 s.d.

```
FFD_2018_2<-lm(ffd~temp,subset(mydata,year==2018))
summ(FFD_2018_2,scale=T)
```

Observations	104
Dependent variable	ffd
Type	OLS linear regression

F(1,102)	48.102
R^2	0.320
Adj. R^2	0.314

	Est.	S.E.	t val.	p
(Intercept)	181.876	1.109	163.946	0.000
temp	-7.731	1.115	-6.936	0.000

Standard errors: OLS; Continuous predictors are mean-centered and scaled by 1 s.d.

Models with number of flowers, untransformed

```
FFD_2017_3<-lm(ffd~temp+nfl,subset(mydata,year==2017))
summ(FFD_2017_3,vif=T,scale=T)
```

Observations	245
Dependent variable	ffd
Type	OLS linear regression

F(2,242)	25.107
R^2	0.172
Adj. R^2	0.165

	Est.	S.E.	t val.	p	VIF
(Intercept)	175.939	0.468	375.895	0.000	NA
temp	-2.144	0.482	-4.454	0.000	1.054
nfl	-2.100	0.482	-4.360	0.000	1.054

Standard errors: OLS; Continuous predictors are mean-centered and scaled by 1 s.d.

```
FFD_2018_3<-lm(ffd~temp+nfl,subset(mydata,year==2018))
summ(FFD_2018_3,vif=T,scale=T)
```

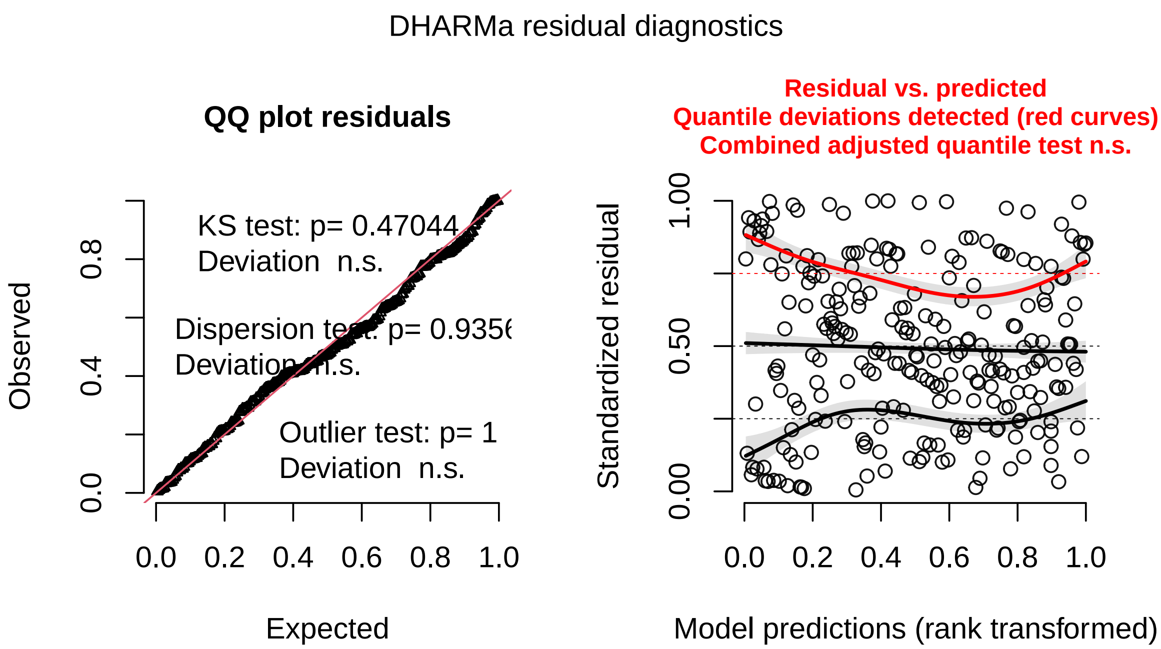
Observations	104
Dependent variable	ffd
Type	OLS linear regression

F(2,101)	39.717
R^2	0.440
Adj. R^2	0.429

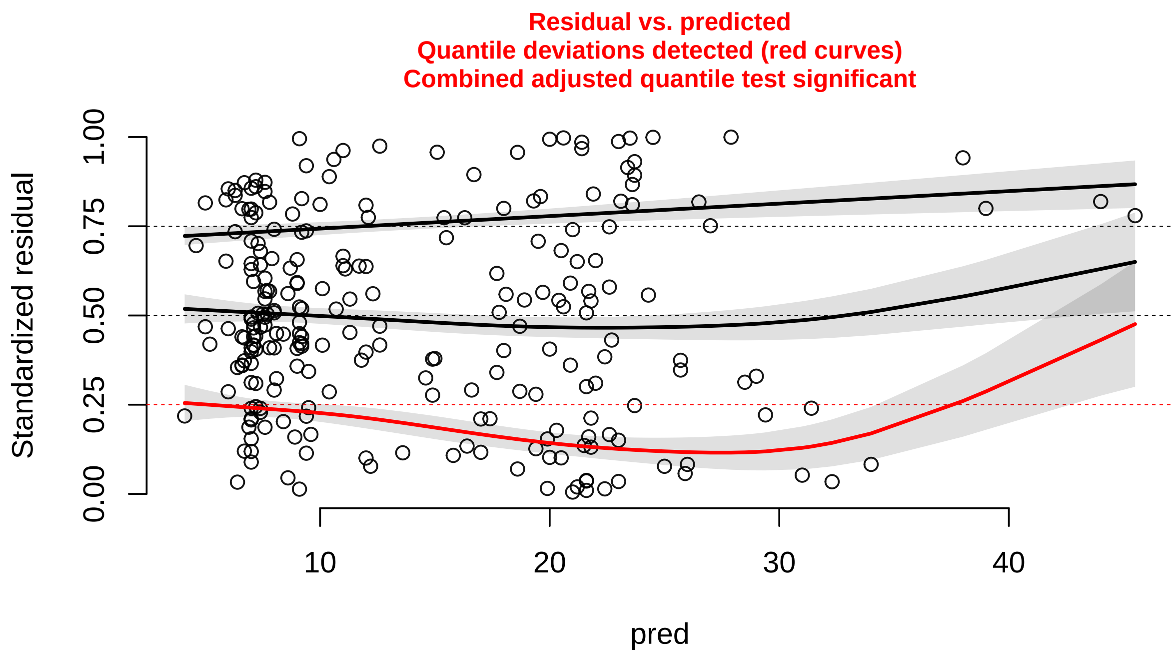
	Est.	S.E.	t val.	p	VIF
(Intercept)	181.876	1.012	179.749	0.000	NA
temp	-7.744	1.017	-7.616	0.000	1.000
nfl	-4.727	1.017	-4.649	0.000	1.000

Standard errors: OLS; Continuous predictors are mean-centered and scaled by 1 s.d.

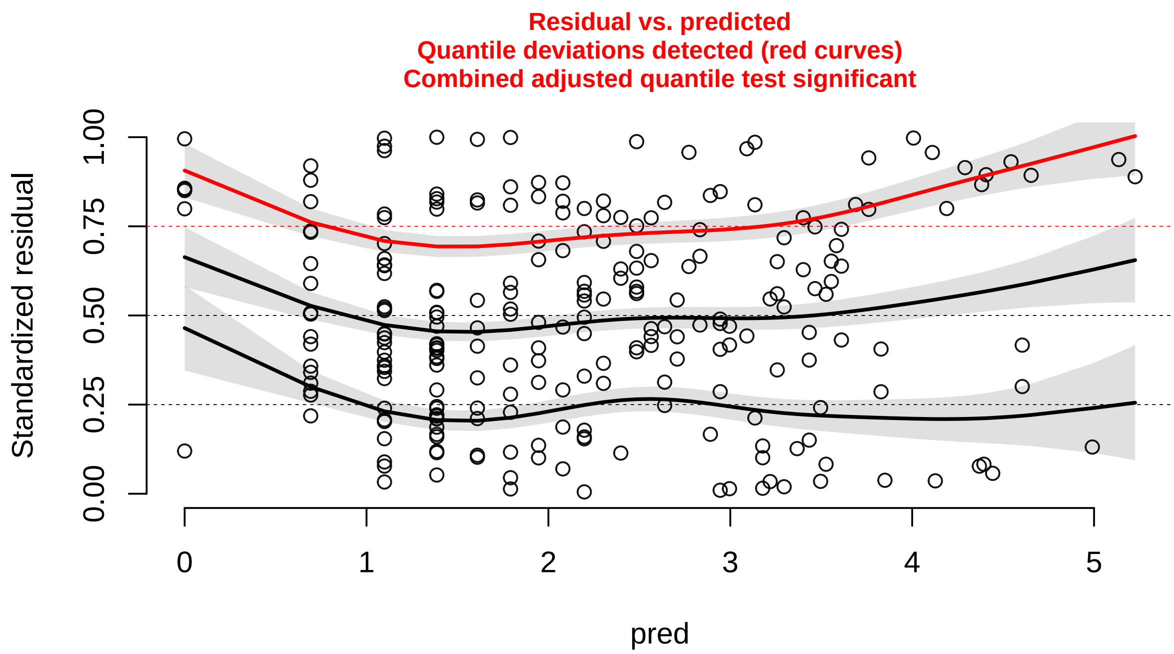
Model diagnostics FFD_2017_1



Residuals against temp



Residuals against log(nfl)

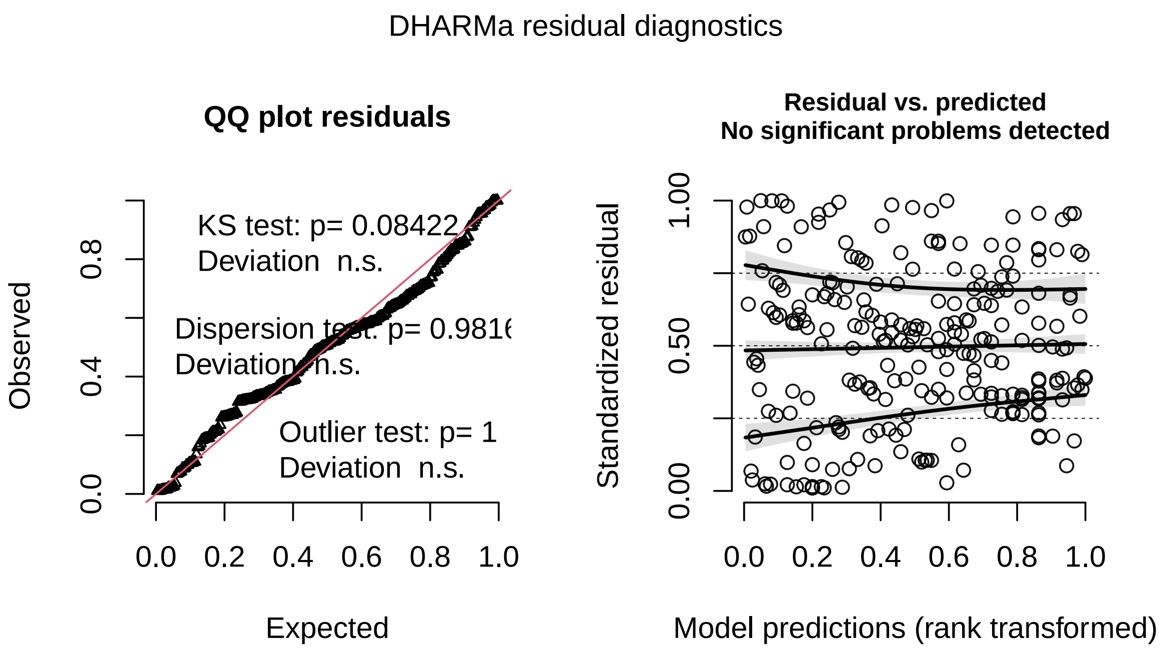


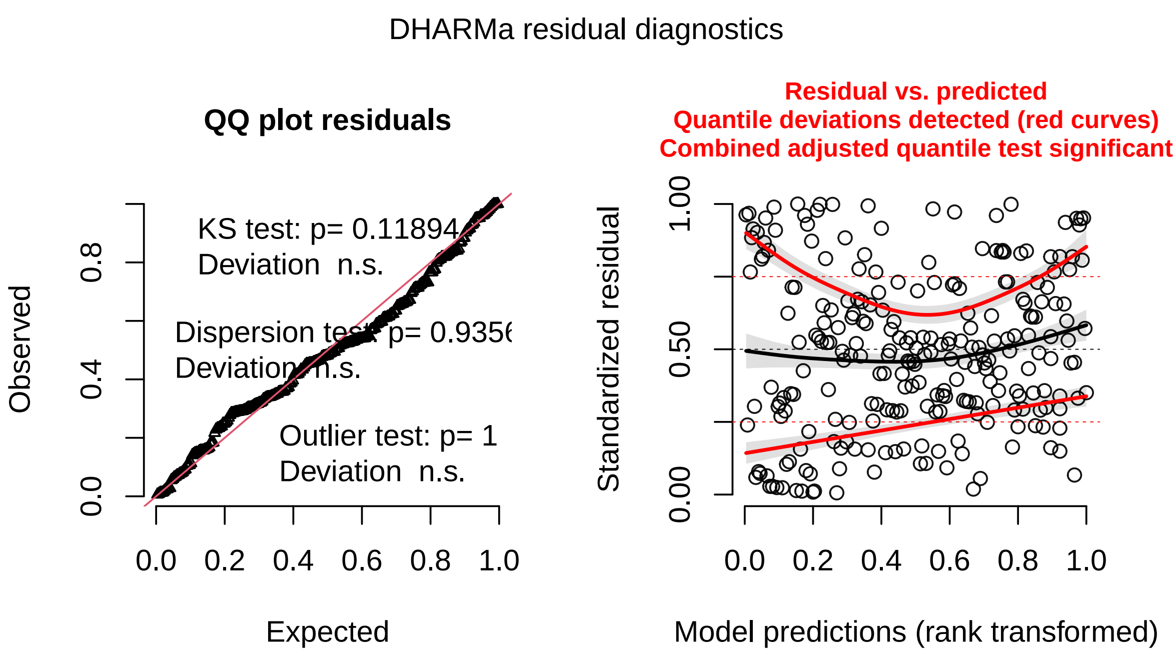
Heteroskedasticity problem: Construct heteroskedasticity consistent standard errors, or robust standard errors (White's standard errors)

```
##
## t test of coefficients:
##
```

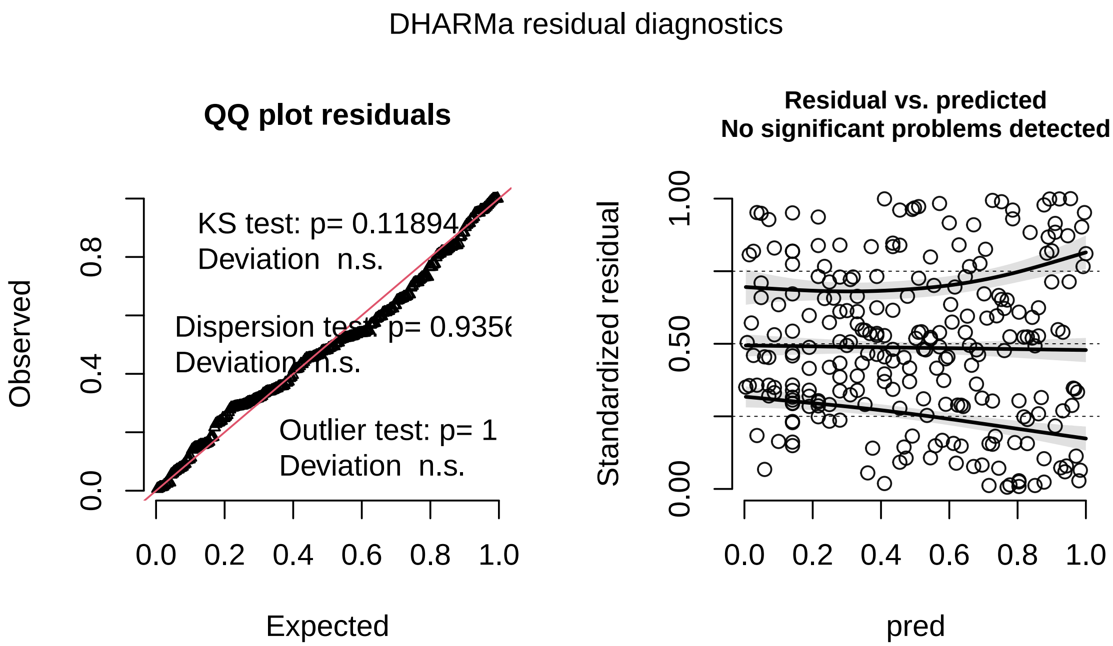
```
##                 Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 185.488273   1.111882 166.8237 < 2.2e-16 ***
## temp         -0.207221   0.065476  -3.1648  0.001751 **
## log(nfl)     -3.046813   0.459734  -6.6273 2.202e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

FFD_2017_2

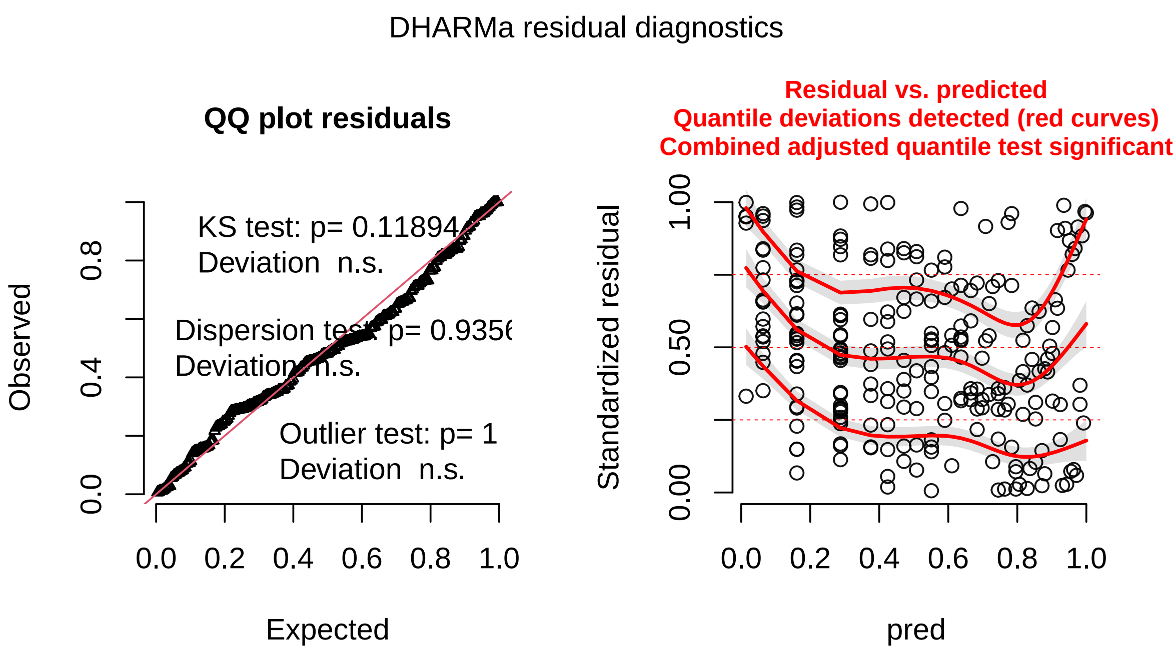




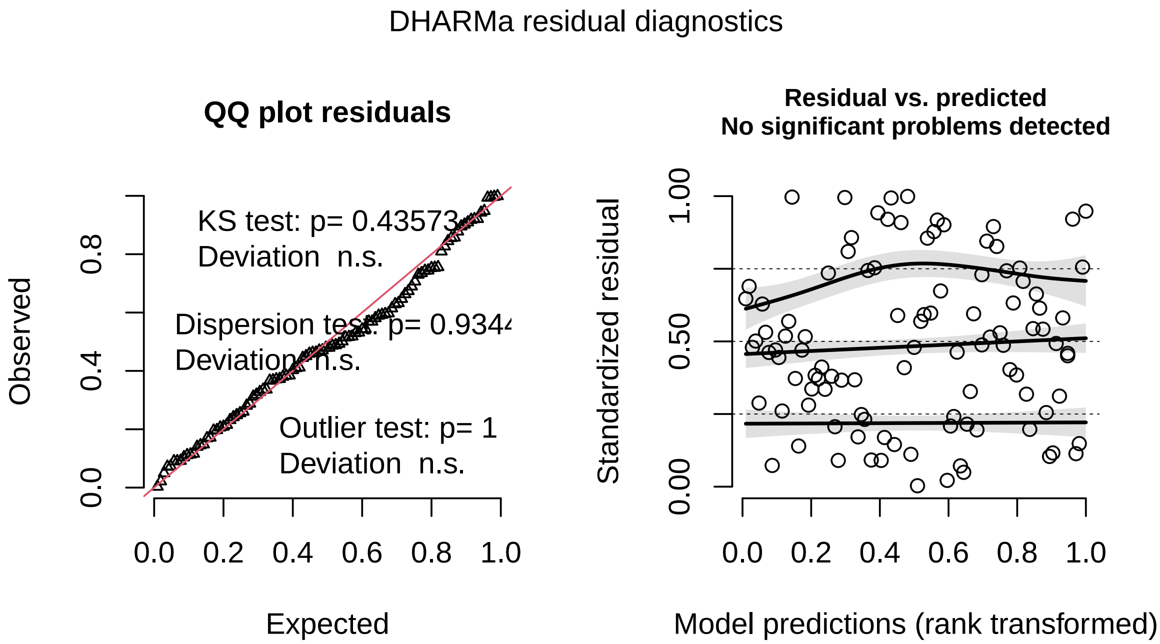
Residuals against temp

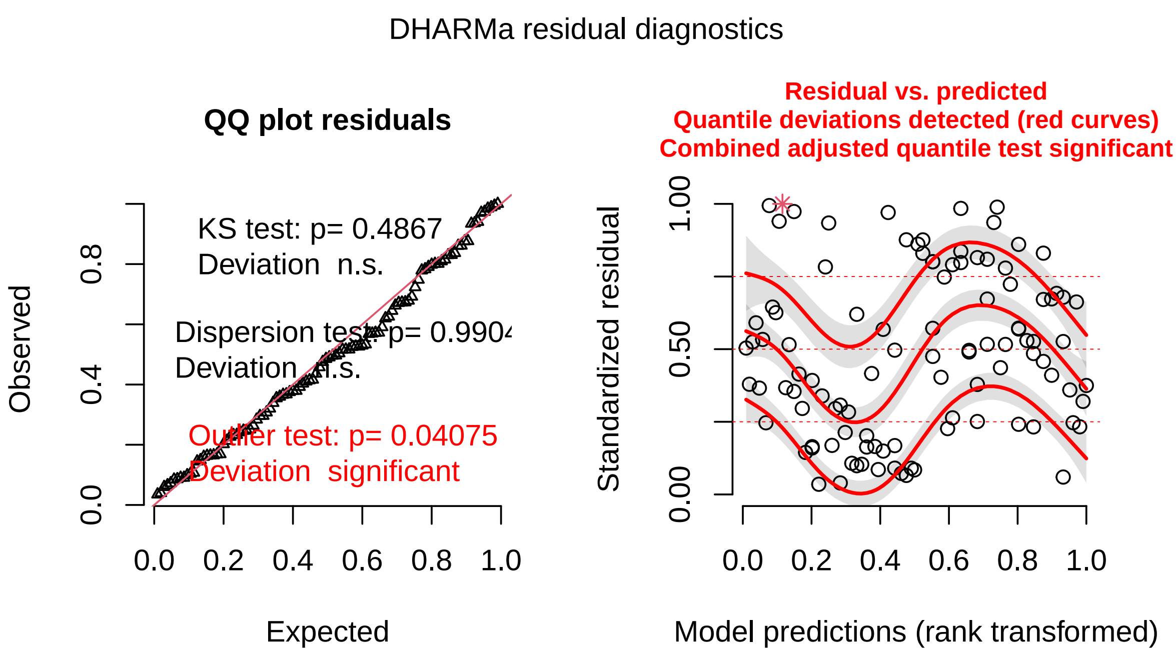


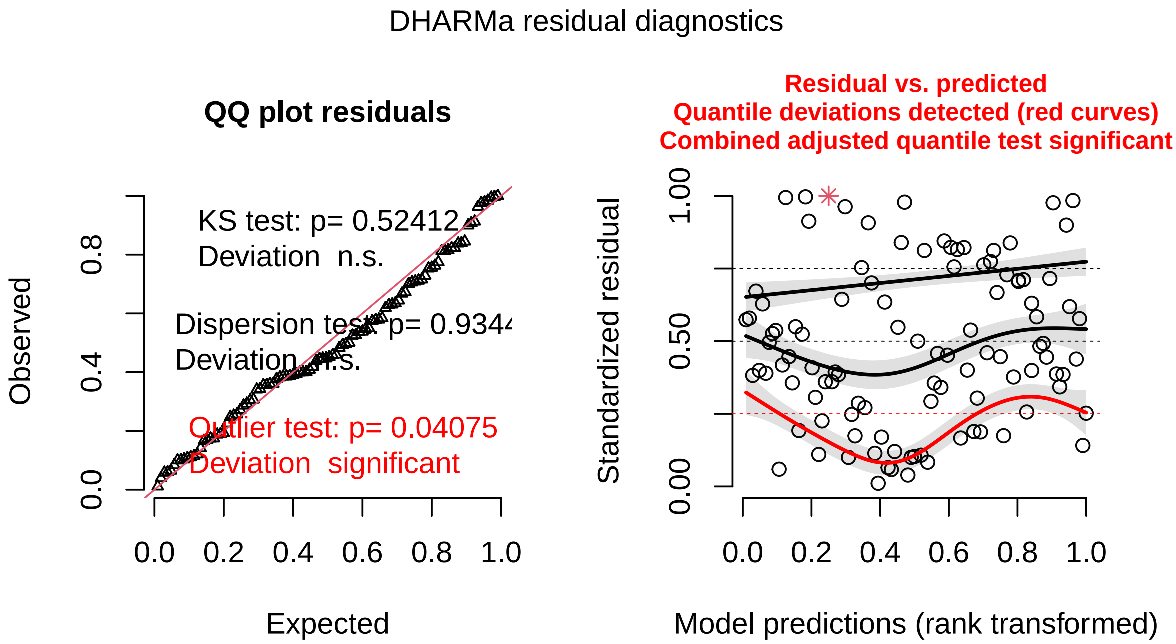
Residuals against nfl



2018:


DHARMa residual diagnostics
QQ plot residuals
KS test: p= 0.43573
Deviation n.s.
Dispersion test: p= 0.9344
Deviation n.s.
Outlier test: p= 1
Deviation n.s.
Observed
Expected
Residual vs. predicted
No significant problems detected
Standardized residual
Model predictions (rank transformed)


DHARMa residual diagnostics
QQ plot residuals
KS test: p= 0.4867
Deviation n.s.
Dispersion test: p= 0.9904
Deviation n.s.
Outlier test: p= 0.04075
Deviation significant
Observed
Expected
Residual vs. predicted
Quantile deviations detected (red curves)
Combined adjusted quantile test significant
Standardized residual
Model predictions (rank transformed)



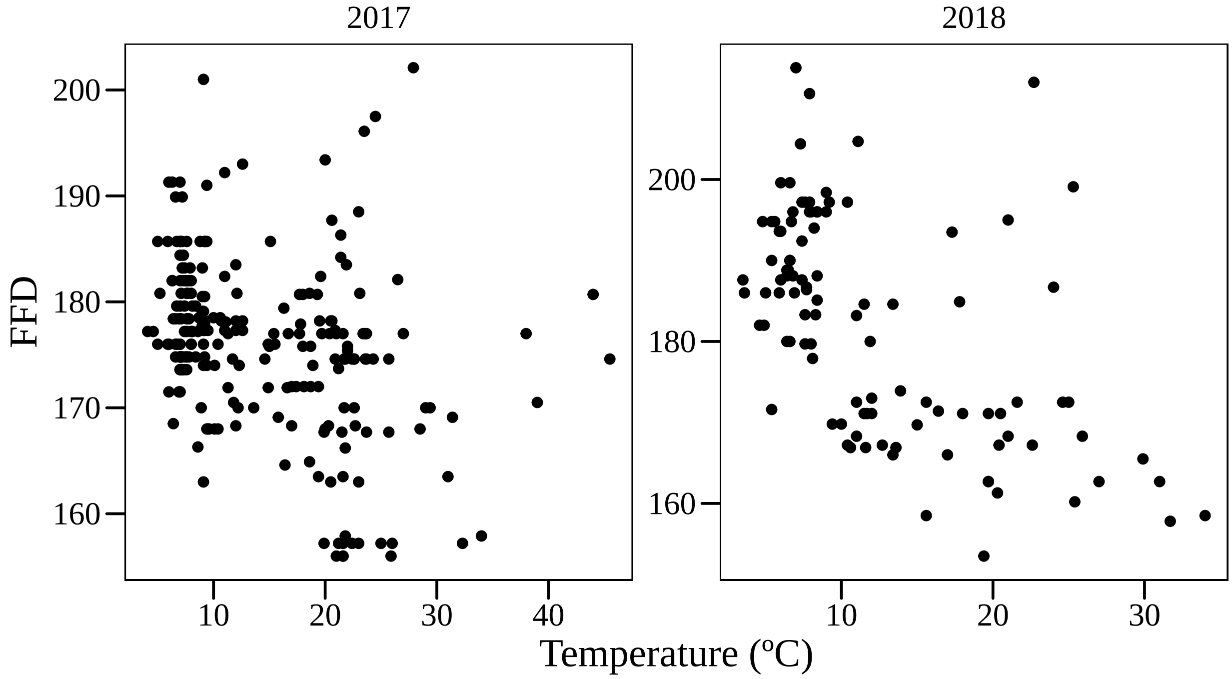
If using models nr 1 for both 2017 and 2018, calculate BCa intervals at least for 2017, where there is heterokedasticity.

BCa intervals for 2017

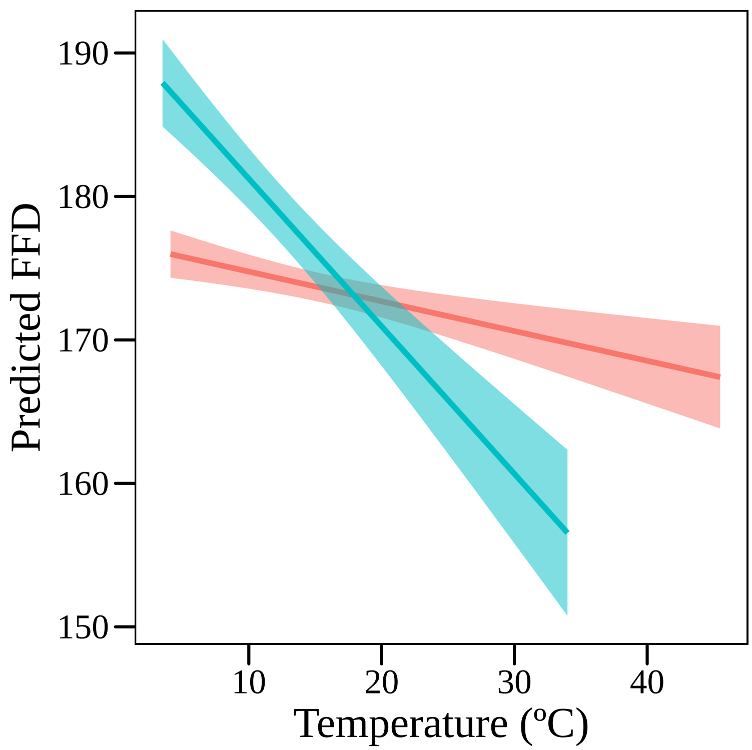
```
##              lower       upper
## temp   -0.3304317 -0.07174499
## lognfl -4.0279429 -2.18180262
```

Both are “significant” according to BCa intervals.

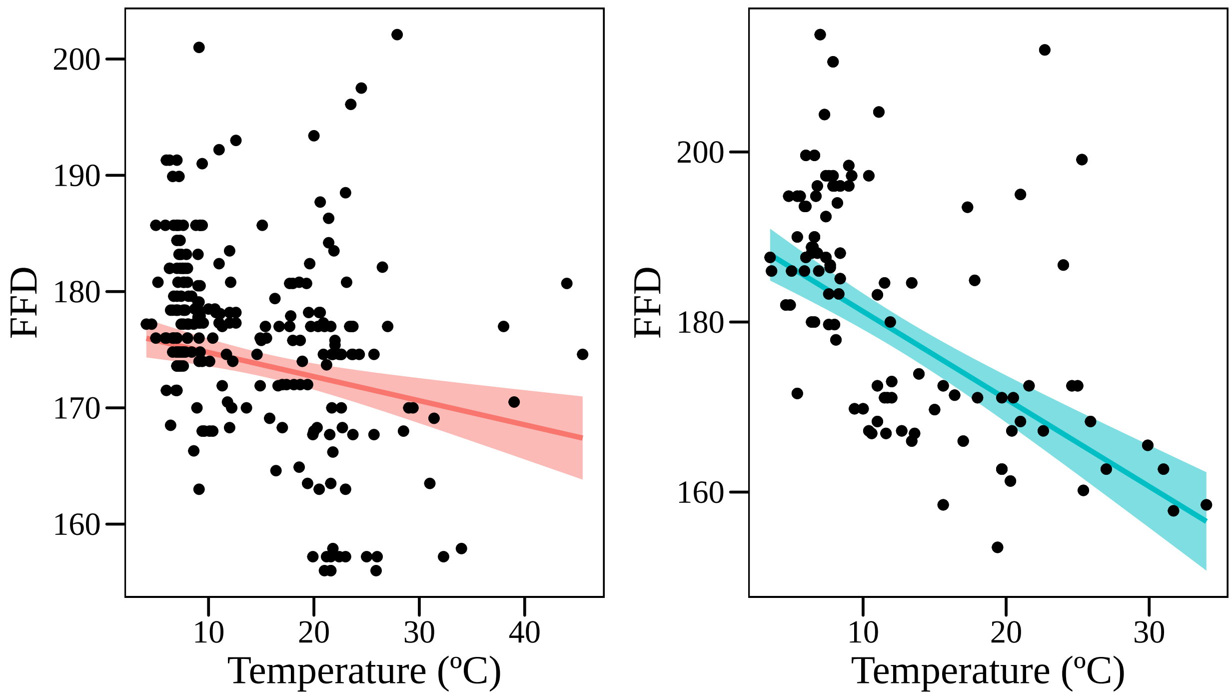
Plots Raw data



Model predictions



Raw data + model predictions



Quadratic

```
FFD_2017_quad<-lm(ffd~temp+I(temp^2)+log(nfl),subset(mydata,year==2017))
summ(FFD_2017_quad)
```

Observations	245
Dependent variable	ffd
Type	OLS linear regression

F(3,241)	30.281
R^2	0.274
Adj. R^2	0.265

	Est.	S.E.	t val.	p
(Intercept)	187.128	1.676	111.646	0.000
temp	-0.464	0.204	-2.276	0.024
I(temp^2)	0.007	0.005	1.314	0.190
log(nfl)	-2.968	0.420	-7.066	0.000

Standard errors: OLS

```
FFD_2018_quad<-lm(ffd~temp+I(temp^2)+log(nfl),subset(mydata,year==2018))
summ(FFD_2018_quad)
```

Quadratic terms for temperature are not significant.

Observations	104
Dependent variable	ffd
Type	OLS linear regression

F(3,100)	39.956
R^2	0.545
Adj. R^2	0.532

	Est.	S.E.	t val.	p
(Intercept)	211.562	4.123	51.308	0.000
temp	-2.011	0.573	-3.508	0.001
I(temp^2)	0.030	0.017	1.757	0.082
log(nfl)	-5.957	0.882	-6.752	0.000

Standard errors: OLS

Piecewise regresssion

```
FFD_2017_segm<-segmented(FFD_2017_1,seg.Z=~temp,psi=20)
slope(FFD_2017_segm)
```

```
## $temp
##            Est.  St.Err.  t value CI(95%).l CI(95%).u
## slope1 -0.26504 0.065508 -4.04590  -0.39408   -0.1360
## slope2  0.69607 1.085000  0.64154  -1.44120    2.8334
```

```
AIC(FFD_2017_1,FFD_2017_segm)
```

```
##               df      AIC
## FFD_2017_1     4 1645.654
## FFD_2017_segm  6 1645.701
```

CIs for the slope of the second segment include zero (means it is not significantly different from zero?). Very little difference in AIC (<2), so the piecewise regression is not better.

```
FFD_2018_segm<-segmented(FFD_2018_1,seg.Z=~temp,psi=15)
slope(FFD_2018_segm)
```

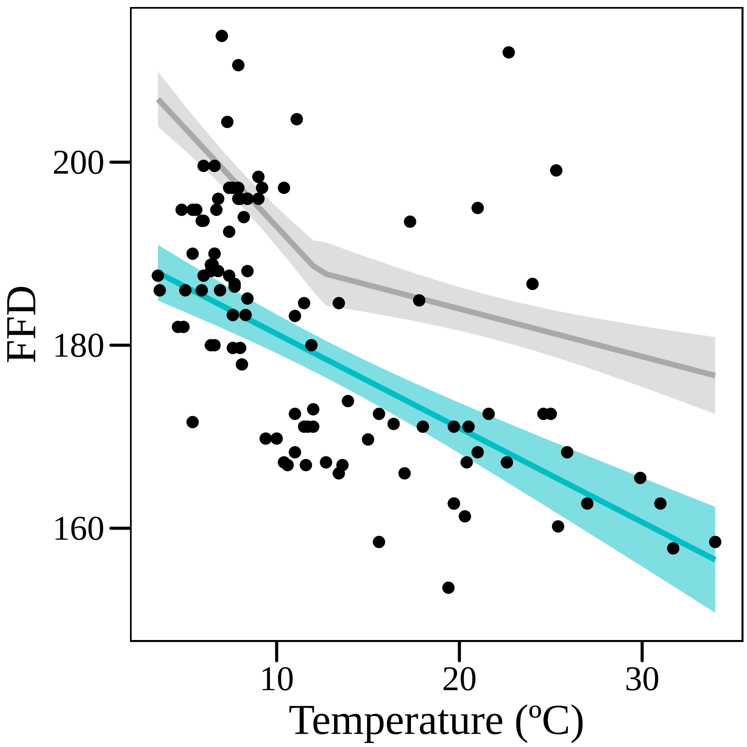
```
## $temp
##            Est. St.Err. t value CI(95%).l CI(95%).u
## slope1 -2.14560 0.51698 -4.1503   -3.1715 -1.119800
## slope2 -0.52127 0.28305 -1.8416   -1.0829  0.040354
```

```
AIC(FFD_2018_1,FFD_2018_segm)
```

```
##               df      AIC
## FFD_2018_1     4 767.1293
## FFD_2018_segm  6 763.0180
```

CIs for the slope of the second segment include zero (means it is not significantly different from zero?). Difference in AIC >2 so the piecewise regression seems to be better.

Plot 2018



I would probably keep the linear model, as the piecewise one does not seem to fit much better (the reduction in AIC was also small).

2. Effect of temperature on fitness

Models as fit by Johan

Only linear

```
fitness_2017<-lm(nseed~temp+log(nfl),subset(mydata,year==2017))
summ(fitness_2017,vif=T)
```

Observations	245
Dependent variable	nseed
Type	OLS linear regression

F(2,242)	99.840
R^2	0.452
Adj. R^2	0.448

	Est.	S.E.	t val.	p	VIF
(Intercept)	-413.970	110.157	-3.758	0.000	NA
temp	-19.588	5.737	-3.415	0.001	1.093
log(nfl)	577.419	40.918	14.112	0.000	1.093

Standard errors: OLS

```
fitness_2018<-lm(nseed~temp+log(nfl),subset(mydata,year==2018))
summ(fitness_2018,vif=T)
```

Observations	104
Dependent variable	nseed
Type	OLS linear regression

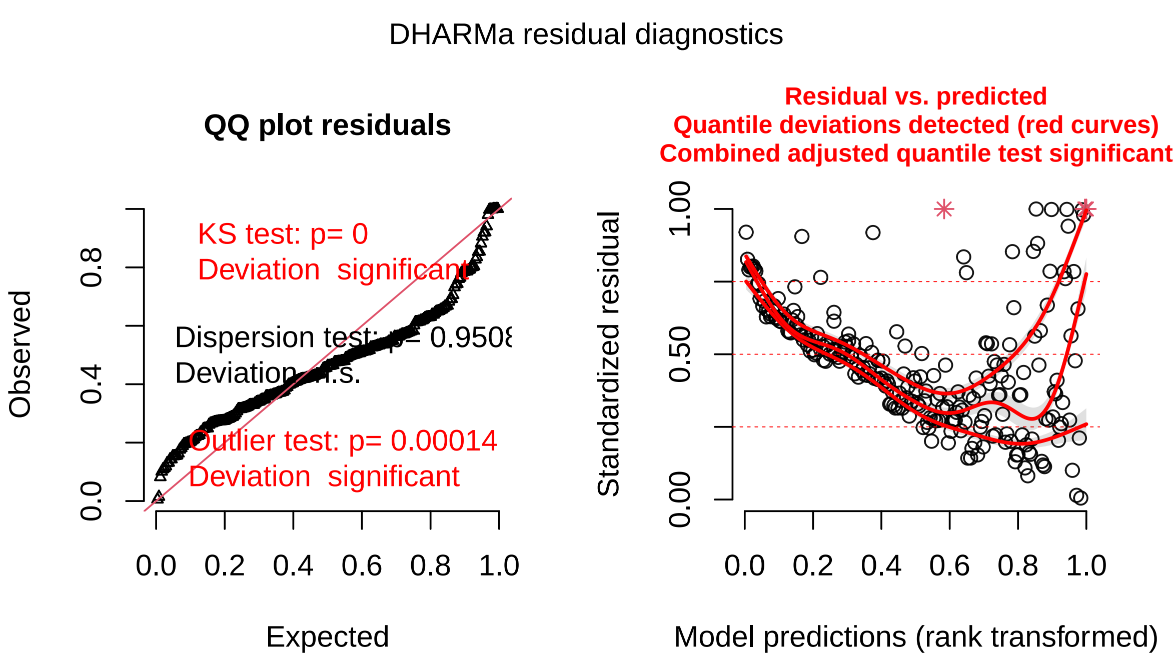
F(2,101)	64.443
R^2	0.561
Adj. R^2	0.552

	Est.	S.E.	t val.	p	VIF
(Intercept)	-40.374	46.970	-0.860	0.392	NA
temp	-8.006	2.503	-3.199	0.002	1.002
log(nfl)	190.899	17.305	11.031	0.000	1.002

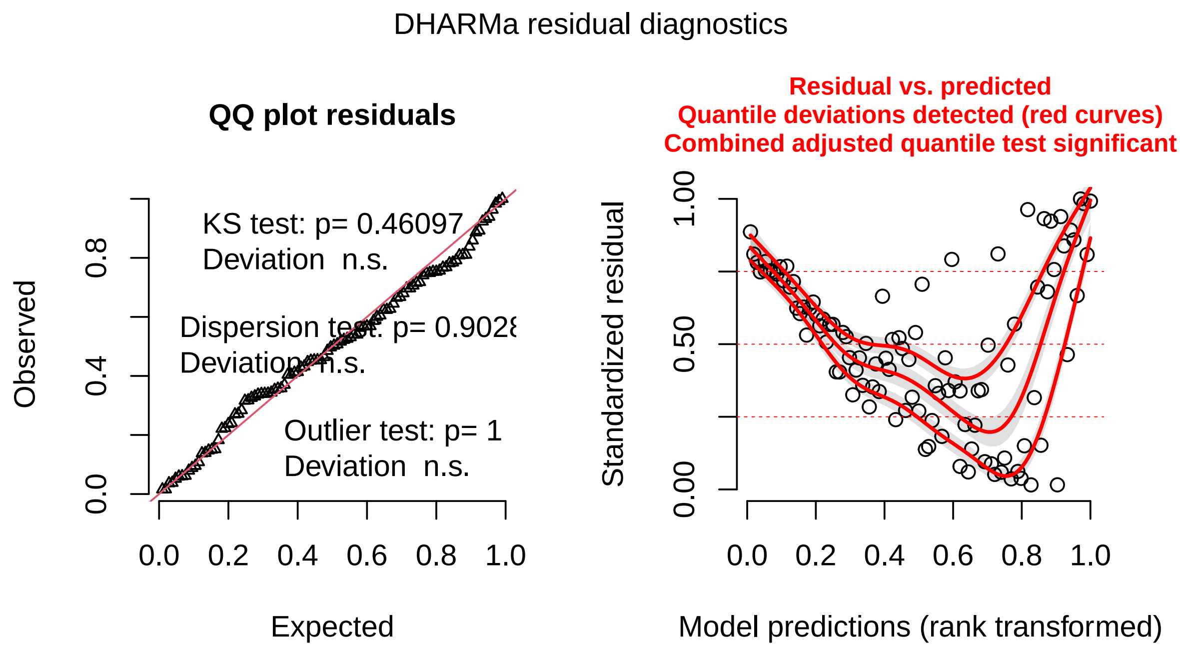
Standard errors: OLS

Model diagnostics qq-plot and plot of residuals vs. predicted:

2017:



2018:



Quite bad looking! We should try another distribution.

Quadratic

```
fitness_2017_quad<-lm(nseed~temp+I(temp^2)+log(nfl),subset(mydata,year==2017))
summ(fitness_2017_quad)
```

Observations	245
Dependent variable	nseed
Type	OLS linear regression

F(3,241)	67.240
R^2	0.456
Adj. R^2	0.449

	Est.	S.E.	t val.	p
(Intercept)	-260.385	164.726	-1.581	0.115
temp	-43.628	20.026	-2.179	0.030
I(temp^2)	0.650	0.519	1.253	0.211
log(nfl)	584.760	41.288	14.163	0.000

Standard errors: OLS

```r
fitness_2018_quad<-lm(nseed~temp+I(temp^2)+log(nfl),subset(mydata,year==2018))
summ(fitness_2018_quad)
```

Observations	104
Dependent variable	nseed
Type	OLS linear regression

F(3,100)	42.951
R^2	0.563
Adj. R^2	0.550

	Est.	S.E.	t val.	p
(Intercept)	-89.182	81.090	-1.100	0.274
temp	0.119	11.274	0.011	0.992
I(temp^2)	-0.249	0.337	-0.739	0.461
log(nfl)	190.526	17.352	10.980	0.000

Standard errors: OLS

Quadratic terms for temperature are not significant.

GLMs with poisson distribution

```r
fitness_2017_pois<-glm(round(nseed)~temp+log(nfl),subset(mydata,year==2017),family="poisson")
summ(fitness_2017_pois,vif=T)
```

Observations	245
Dependent variable	round(nseed)
Type	Generalized linear model
Family	poisson
Link	log

$\chi^2(2)$	148392.783
Pseudo-R^2 (Cragg-Uhler)	1.000
Pseudo-R^2 (McFadden)	0.710
AIC	60720.575
BIC	60731.078

	Est.	S.E.	z val.	p	VIF
(Intercept)	4.299	0.009	488.236	0.000	NA
temp	-0.034	0.000	-90.878	0.000	1.080
log(nfl)	0.922	0.002	383.357	0.000	1.080

Standard errors: MLE

```r
fitness_2018_pois<-glm(round(nseed)~temp+log(nfl),subset(mydata,year==2018),family="poisson")
summ(fitness_2018_pois,vif=T)
```

Observations	104
Dependent variable	round(nseed)
Type	Generalized linear model
Family	poisson
Link	log

$\chi^2(2)$	21060.607
Pseudo-R^2 (Cragg-Uhler)	1.000
Pseudo-R^2 (McFadden)	0.735
AIC	7612.990
BIC	7620.923

	Est.	S.E.	z val.	p	VIF
(Intercept)	3.511	0.025	137.814	0.000	NA
temp	-0.038	0.001	-34.179	0.000	1.001
log(nfl)	0.976	0.008	129.523	0.000	1.001

Standard errors: MLE

```r
overdisp_fun(fitness_2017_pois)
```

```
##      chisq      ratio        rdf          p
## 93055.8367   384.5283   242.0000     0.0000
```

```r
overdisp_fun(fitness_2018_pois)
```

```
##     chisq     ratio       rdf         p
## 6128.0539   60.6738  101.0000    0.0000
```

There is significant overdispersion.

GLMs with negative binomial distribution –> Keep these?

```r
fitness_2017_nb<-glm.nb(round(nseed)~temp+log(nfl),subset(mydata,year==2017))
summ(fitness_2017_nb,vif=T)
```

Observations	245
Dependent variable	round(nseed)
Type	Generalized linear model
Family	Negative Binomial(2.0993)
Link	log

$\chi^2()$	0.722	0.088	3273.106	3287.112
Pseudo-R^2 (Cragg-Uhler)	0.722	0.088	3273.106	3287.112
Pseudo-R^2 (McFadden)	0.722	0.088	3273.106	3287.112
AIC	0.722	0.088	3273.106	3287.112
BIC	0.722	0.088	3273.106	3287.112

	Est.	S.E.	z val.	p	VIF
(Intercept)	4.079	0.113	36.029	0.000	NA
temp	-0.030	0.006	-5.117	0.000	1.094
log(nfl)	0.982	0.042	23.377	0.000	1.094

Standard errors: MLE

```
fitness_2018_nb<-glm.nb(round(nseed)~temp+log(nfl),subset(mydata,year==2018))
summ(fitness_2018_nb,vif=T)
```

Observations	104
Dependent variable	round(nseed)
Type	Generalized linear model
Family	Negative Binomial(1.9313)
Link	log

$\chi^2()$	0.684	0.091	1203.369	1213.947
Pseudo-R^2 (Cragg-Uhler)	0.684	0.091	1203.369	1213.947
Pseudo-R^2 (McFadden)	0.684	0.091	1203.369	1213.947
AIC	0.684	0.091	1203.369	1213.947
BIC	0.684	0.091	1203.369	1213.947

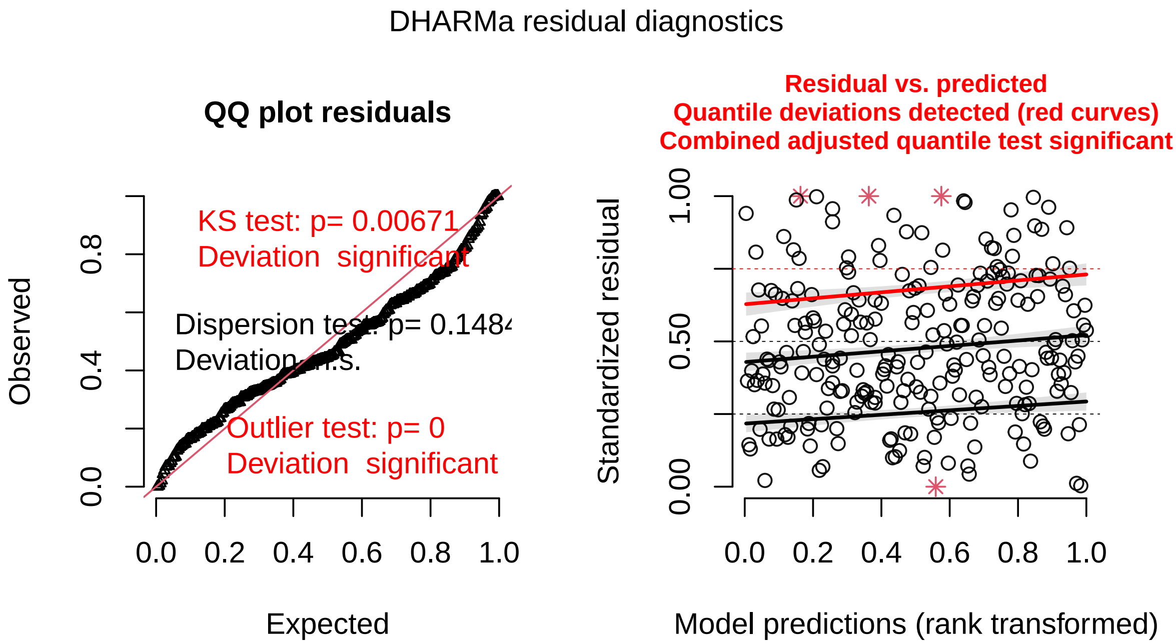
	Est.	S.E.	z val.	p	VIF
(Intercept)	3.543	0.188	18.878	0.000	NA
temp	-0.042	0.010	-4.166	0.000	1.003
log(nfl)	0.981	0.069	14.202	0.000	1.003

Standard errors: MLE

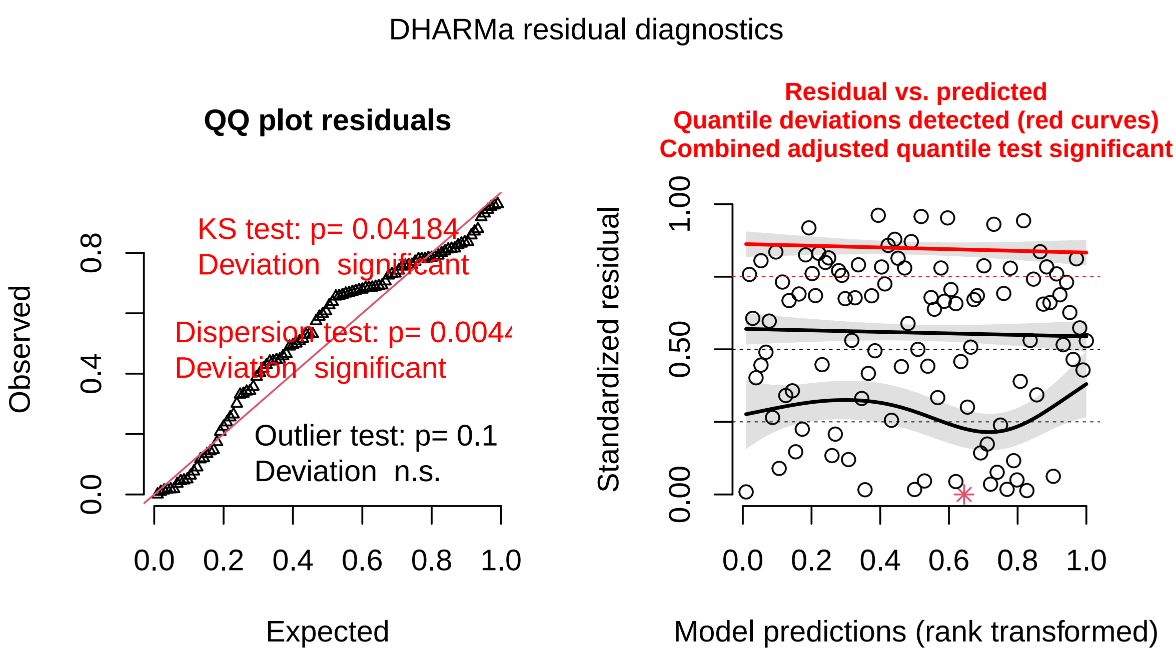
Model diagnostics

qq-plot and plot of residuals vs. predicted:

2017:



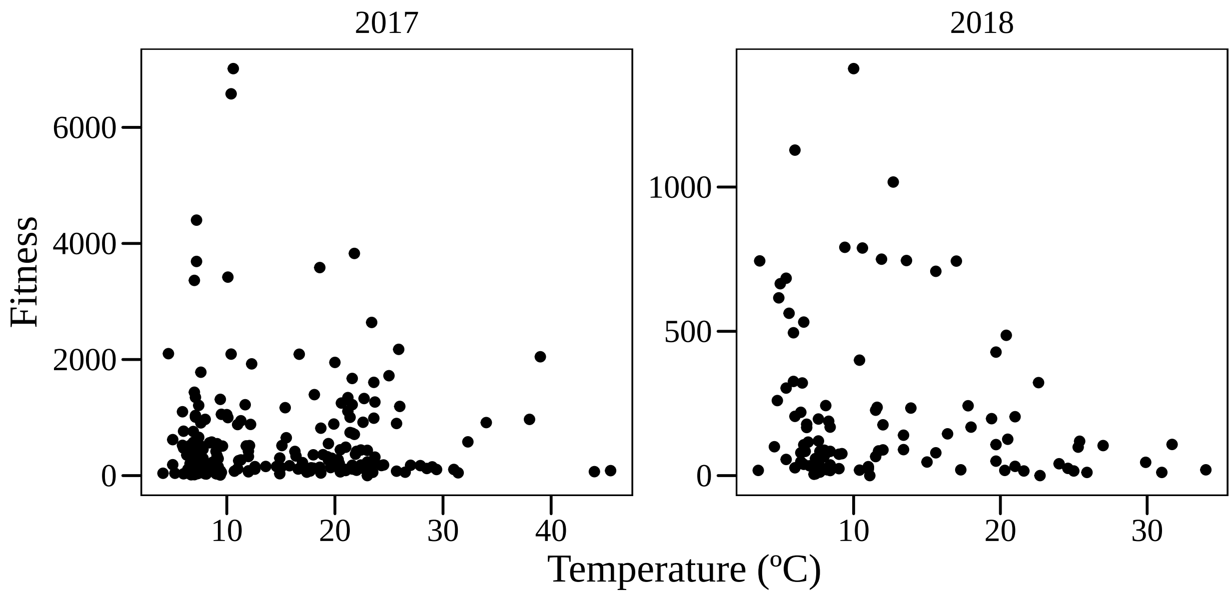
2018:



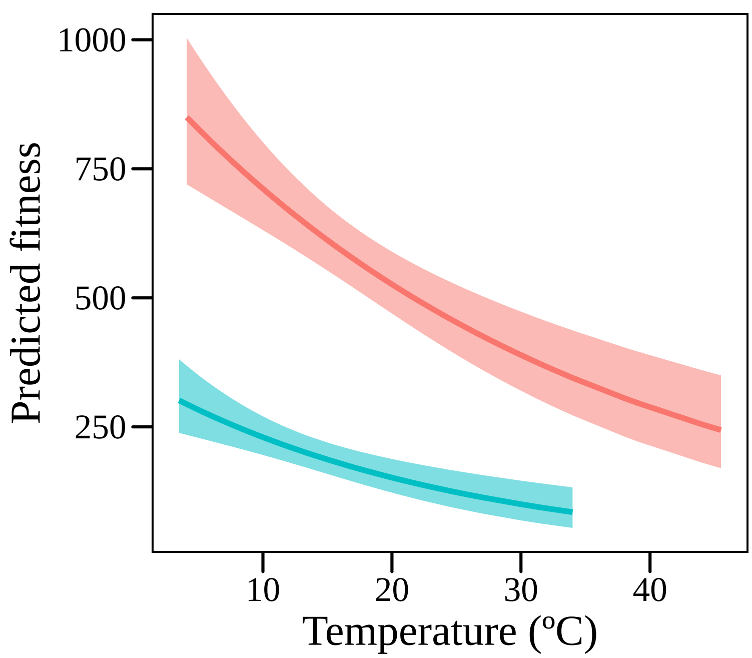
Some problems but maybe not so bad. Need to look a bit more into this later.

Plots

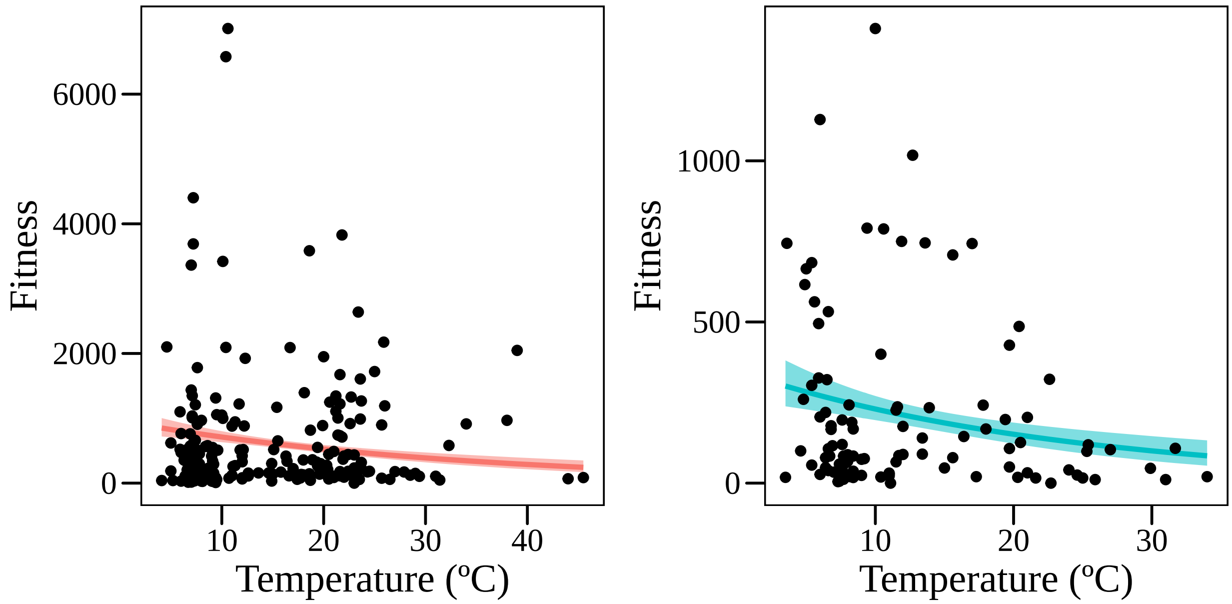
Graphs of raw data



Graphs of model predictions (negative binomial GLMs)



Raw data + model predictions



3. Effect of temperature on selection on FFD

Models as fit by Johan

Only linear

```
selection_2017<-lm(nseed_rel~ffd_std*temp+nfl_std,subset(mydata,year==2017))
summ(selection_2017)
```

Observations	245
Dependent variable	nseed_rel
Type	OLS linear regression

F(4,240)	51.549
R^2	0.462
Adj. R^2	0.453

	Est.	S.E.	t val.	p
(Intercept)	1.408	0.166	8.460	0.000
ffd_std	0.331	0.206	1.602	0.110
temp	-0.030	0.010	-2.937	0.004
nfl_std	1.213	0.089	13.663	0.000
ffd_std:temp	-0.009	0.010	-0.848	0.397

Standard errors: OLS

```r
selection_2018<-lm(nseed_rel~ffd_std*temp+nfl_std,subset(mydata,year==2018))
summ(selection_2018)
```

Observations	104
Dependent variable	nseed_rel
Type	OLS linear regression

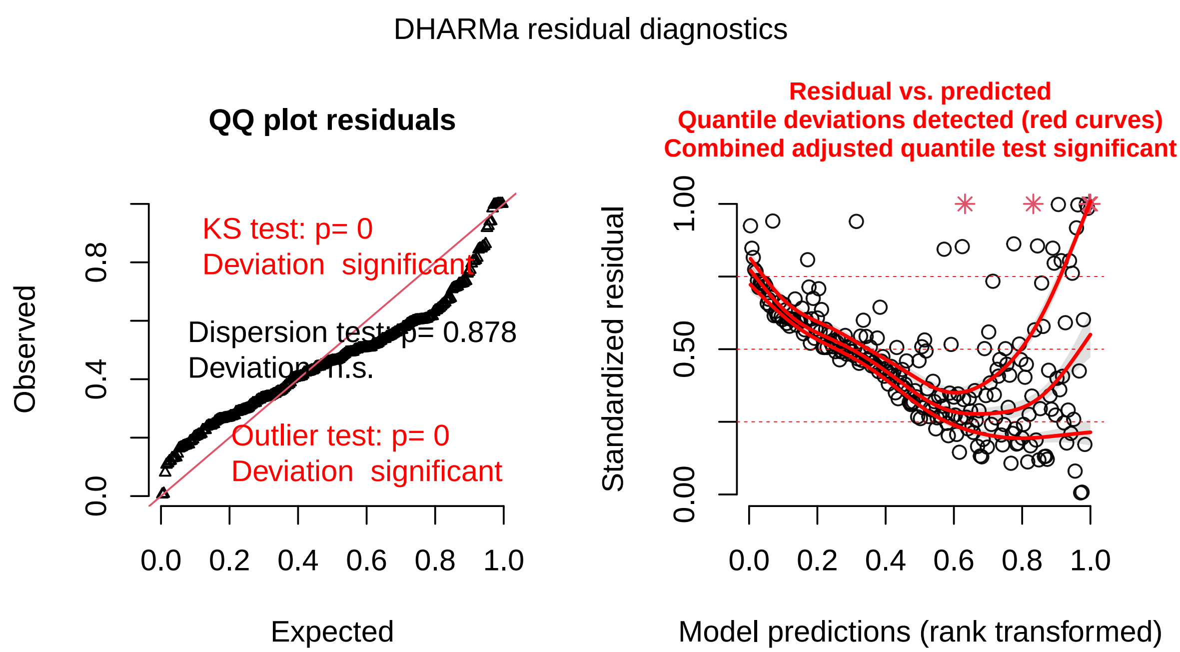
F(4,99)	34.660
R^2	0.583
Adj. R^2	0.567

	Est.	S.E.	t val.	p
(Intercept)	1.374	0.201	6.825	0.000
ffd_std	-0.293	0.205	-1.431	0.156
temp	-0.021	0.016	-1.368	0.174
nfl_std	0.976	0.101	9.679	0.000
ffd_std:temp	0.027	0.012	2.241	0.027

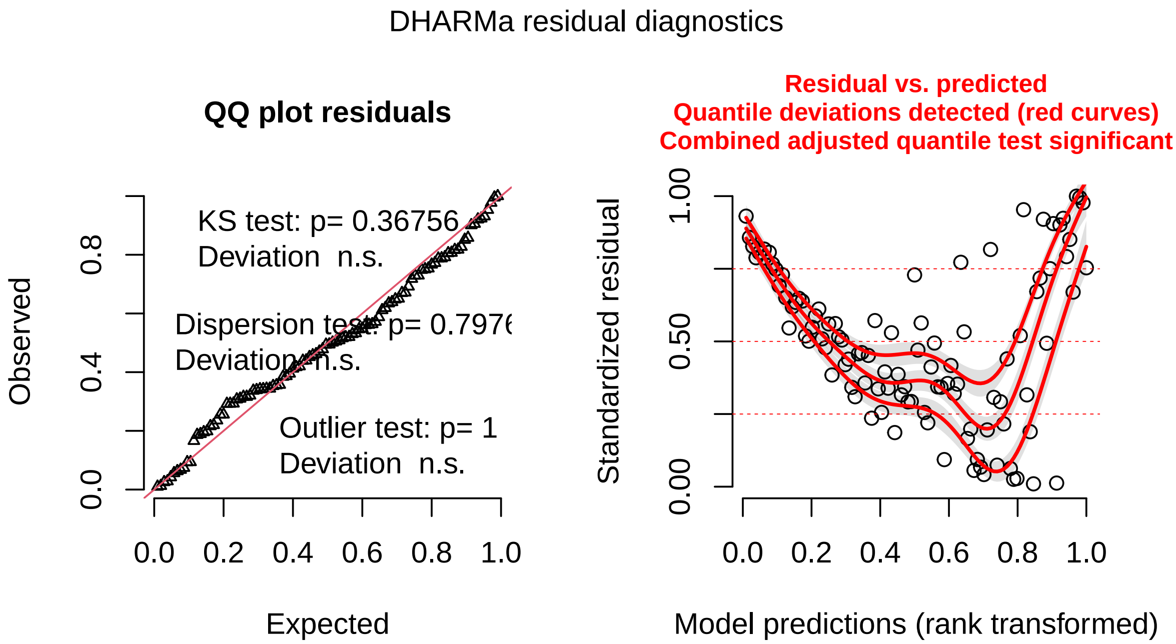
Standard errors: OLS

Model diagnostics qq-plot and plot of residuals vs. predicted:

2017:



2018:



Quite bad looking! If we want to keep the linear model with normal distribution (i.e. a "classic" selection model), we can assess significances using BCa intervals.

BCa intervals

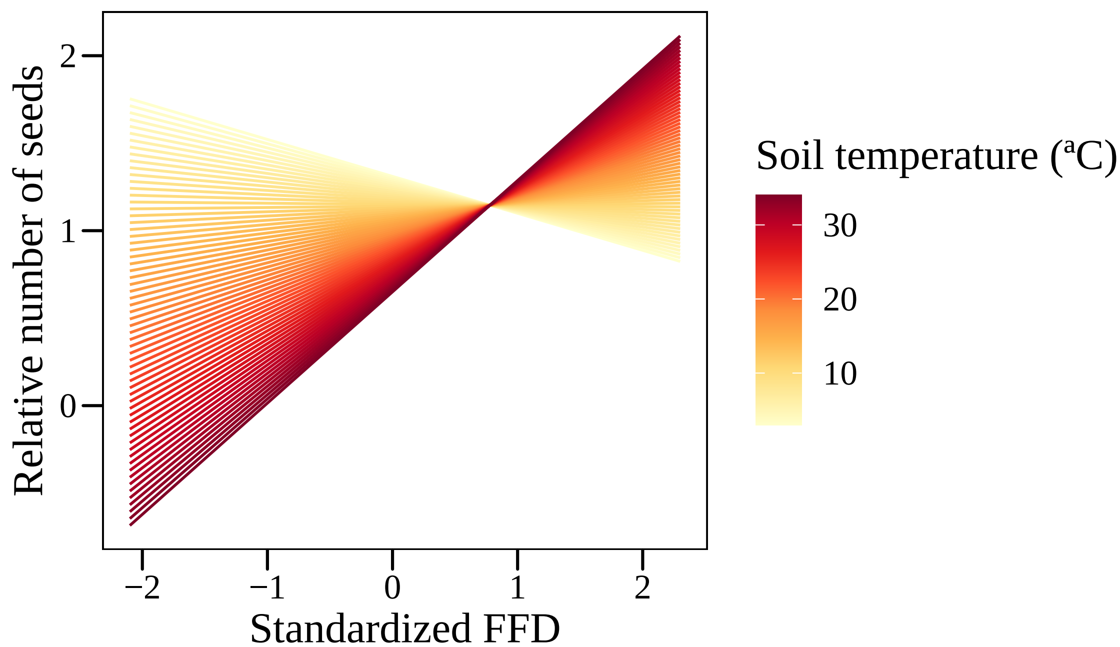
2017

```
##                    lower        upper
## ffd_std       0.02310978  0.796295977
## temp         -0.05434295 -0.012941578
## nfl_std       0.90978664  1.697819198
## ffd_std:temp -0.02558533  0.005170813
```

2018

```
##                     lower       upper
## ffd_std      -0.734812519  0.21967818
## temp         -0.041653471 -0.00145628
## nfl_std       0.729117585  1.25825159
## ffd_std:temp  0.003800642  0.05221350
```

The significances according to the BCa intervals are similar to the ones given in the model summary.



Plot 2018

Quadratic

```
selection_2017_quad<-lm(nseed_rel~ffd_std*temp+ffd_std*I(temp^2)+nfl_std,subset(mydata,year==2017))
summ(selection_2017_quad)
```

Observations	245
Dependent variable	nseed_rel
Type	OLS linear regression

F(6,238)	34.508
R^2	0.465
Adj. R^2	0.452

	Est.	S.E.	t val.	p
(Intercept)	1.642	0.299	5.490	0.000
ffd_std	0.504	0.389	1.296	0.196
temp	-0.065	0.036	-1.808	0.072
I(temp^2)	0.001	0.001	1.033	0.303
nfl_std	1.226	0.090	13.667	0.000
ffd_std:temp	-0.035	0.045	-0.770	0.442
ffd_std:I(temp^2)	0.001	0.001	0.627	0.531

Standard errors: OLS

```
selection_2018_quad<-lm(nseed_rel~ffd_std*temp+ffd_std*I(temp^2)+nfl_std,subset(mydata,year==2018))
summ(selection_2018_quad)
```

Observations	104
Dependent variable	nseed_rel
Type	OLS linear regression

F(6,97)	22.809
R^2	0.585
Adj. R^2	0.560

	Est.	S.E.	t val.	p
(Intercept)	1.607	0.517	3.105	0.002
ffd_std	-0.455	0.499	-0.911	0.365
temp	-0.055	0.083	-0.662	0.510
I(temp^2)	0.001	0.003	0.369	0.713
nfl_std	0.973	0.103	9.437	0.000
ffd_std:temp	0.043	0.071	0.614	0.541
ffd_std:I(temp^2)	-0.000	0.002	-0.146	0.884

Standard errors: OLS

4. Effect of temperature on the relationship absolute fitness-FFD

Models as fit by Johan

Only linear

```
selection_2017_abs<-lm(nseed~ffd*temp+log(nfl),subset(mydata,year==2017))
summ(selection_2017_abs)
```

Observations	245
Dependent variable	nseed
Type	OLS linear regression

F(4,240)	51.549
R^2	0.462
Adj. R^2	0.453

	Est.	S.E.	t val.	p
(Intercept)	-4647.341	2606.895	-1.783	0.076
ffd	23.284	14.531	1.602	0.110
temp	88.601	124.771	0.710	0.478
log(nfl)	619.124	45.314	13.663	0.000
ffd:temp	-0.601	0.709	-0.848	0.397

Standard errors: OLS

```
selection_2018_abs<-lm(nseed~ffd*temp+log(nfl),subset(mydata,year==2018))
summ(selection_2018_abs)
```

Observations	104
Dependent variable	nseed
Type	OLS linear regression

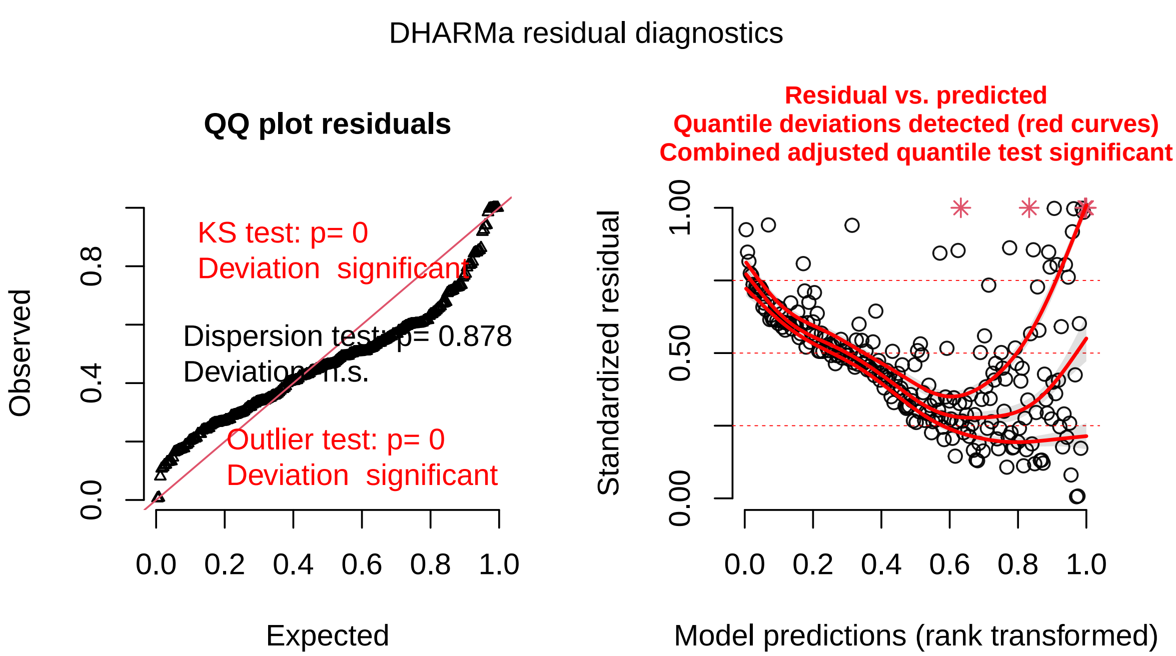
F(4,99)	34.660
R^2	0.583
Adj. R^2	0.567

	Est.	S.E.	t val.	p
(Intercept)	755.873	607.739	1.244	0.217
ffd	-4.561	3.188	-1.431	0.156
temp	-81.835	33.624	-2.434	0.017
log(nfl)	198.339	20.492	9.679	0.000
ffd:temp	0.425	0.190	2.241	0.027

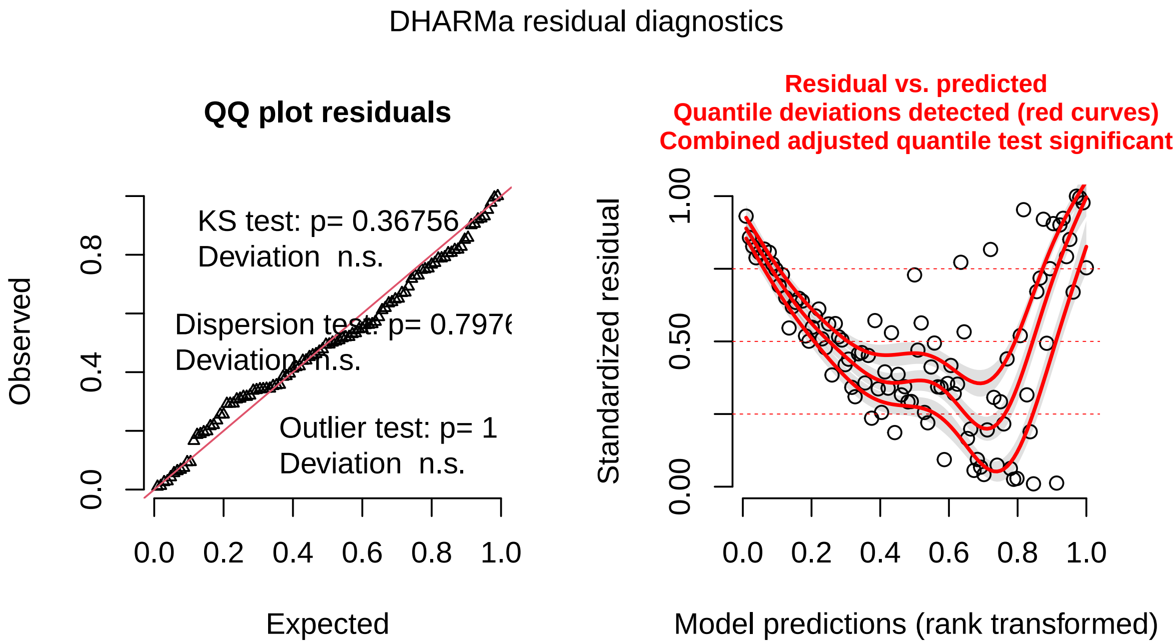
Standard errors: OLS

Model diagnostics qq-plot and plot of residuals vs. predicted:

2017:



2018:



Quite bad looking! If we want to keep the linear model with normal distribution (i.e. a "classic" selection model), we can assess significances using BCa intervals.

BCa intervals

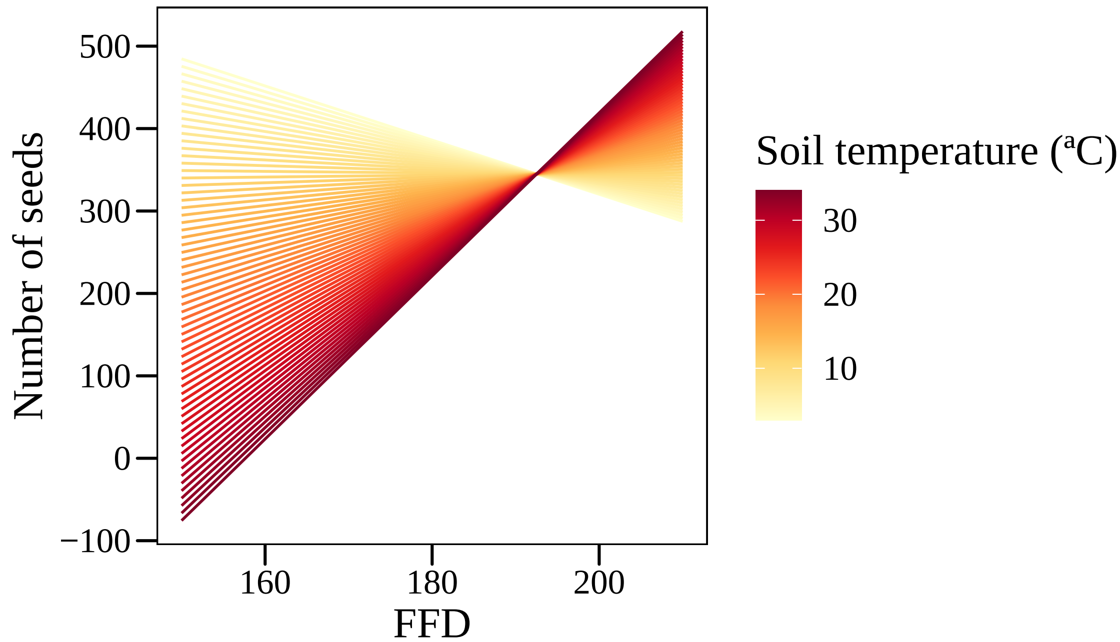
2017

```
##                  lower       upper
## ffd          0.5181807  55.7937405
## temp       -80.3479811 291.9942499
## nfl        463.0197317 868.0871957
## ffd:temp    -1.7758943   0.3872076
```

2018

```
##                    lower       upper
## ffd         -11.70286559   3.3960908
## temp       -151.86974365 -15.0743717
## nfl         147.68108072 254.1006289
## ffd:temp      0.06974127   0.8110243
```

The significances according to the BCa intervals are similar to the ones given in the model summary, with the exception that according to the BCa intervals ffd is significant in the model for 2017.



Plot 2018

We could also try this model with other distributions (Poisson, negative binomial), but I guess that keeping the "classic" approach with a normal distribution is OK if we show BCa intervals.