

R Notebook

Contents

Overall approaches to data analysis	1
Not accounting for error	1
Multiple linear regression with parameters	1
Figures showing main effects of scaling, ex, plateau, and termination	2
Multiple linear regression with PCs	2
Accounting for error	3
Estimates of the parameters are correlated with the error in those estimates	4
Multiple linear regression with parameters	4
Main effect of host and baseline	5
host by termination interaction	6
Multiple linear regression with PCs	6
Conclusions	7
Table summary of stats	7

Overall approaches to data analysis

We're framing the study as diapause consists of modules that comprise of modular components, so how do Rhagoletis hosts assemble their developmental modules to exit dormancy/diapause?

We argue that the parameters of the bi/tri-phasic metabolic rate trajectory models are modular components, which can assemble into modules. They can act independently or in concert.

To capture the interdependence of metabolic rate trajectory parameters, we should do a PCA to estimate their correlational structure and evaluate their(PCs) importance in a multiple linear regression. On the other hand, these parameters could be acting independently, and we can evaluate this with a multiple linear regression using the parameters. Eclosion timing in days is the response variable in all of these models.

One problem that we encounter is that parameter estimates have error and it was suggested that we should account for this. So here is a report laying out the analysis correcting for error and without.

Here are the parameter names:

- b= baseline metabolic rate
- term = termination timing
- plateau = metabolic rate at which it plateaus after termination
- Ex = post diapause development in days or distance between termination and when MR exponentially increases
- c1 = scaling parameter

Not accounting for error

Multiple linear regression with parameters

Model construction: testing host x parameters interaction on eclosion timing

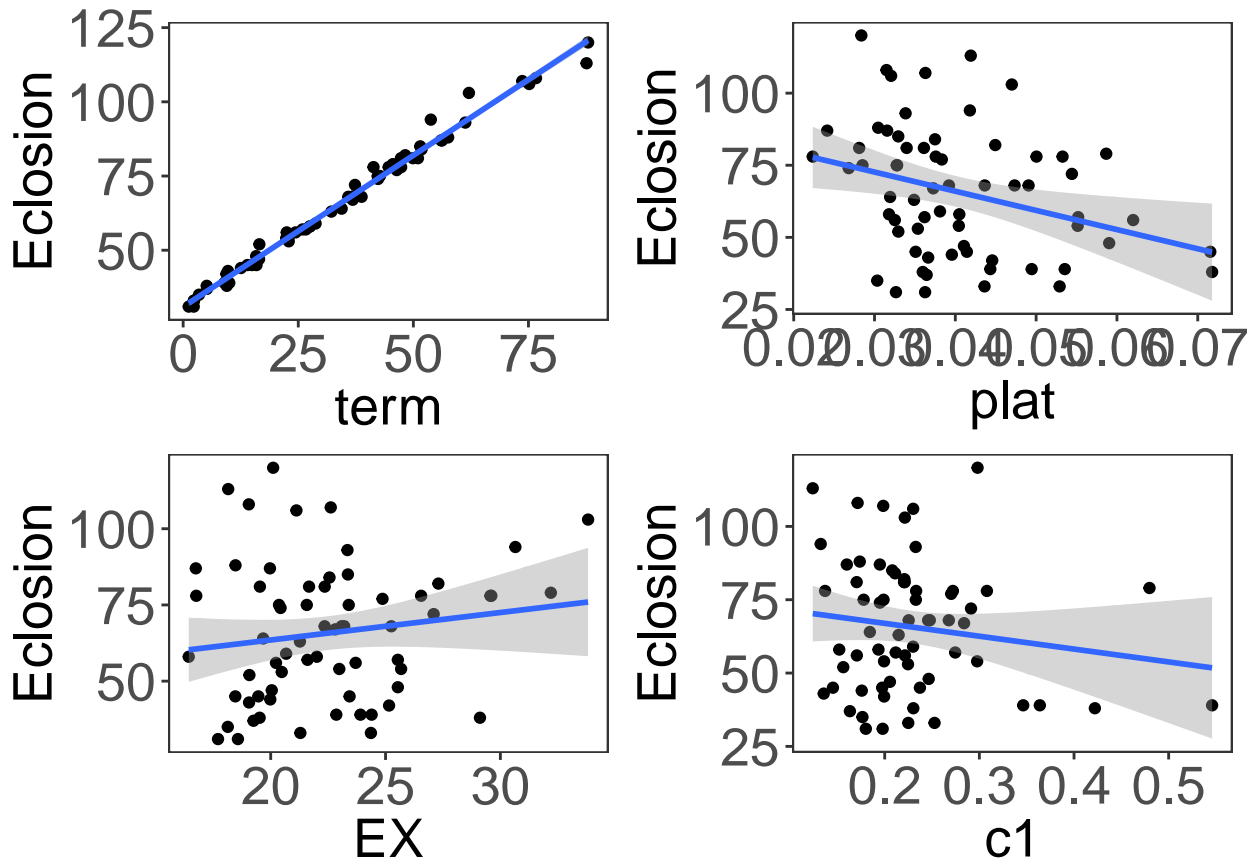
- parameters are scaled so that slope coefficients are comparable.
- Model selection based on AIC.

```
full.mod3<-lm(Eclosion~host*b+host*c1+host*EX+host*plat+host*term,data=uncor3)
#summary(full.mod3)
x<-stepAIC(full.mod3,direction="both")
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	65.603	0.387	169.522	0.000
hostH	-0.369	0.581	-0.635	0.528
c1	-1.067	0.228	-4.673	0.000
EX	2.058	0.540	3.814	0.000
plat	-0.507	0.428	-1.183	0.242
term	21.718	0.273	79.430	0.000
hostH:EX	1.067	0.632	1.689	0.097
hostH:plat	-1.275	0.672	-1.897	0.063

There are main effects of scaling, Ex, plateau, and termination

Figures showing main effects of scaling, ex, plateau, and termination



Multiple linear regression with PCs

Grab PC's

```

uncor.pca.param<-princomp(scale(uncor[,3:7]))
#summary(uncor.pca.param)
knitr::kable(round(uncor.pca.param$loadings[,1:3],3))

```

	Comp.1	Comp.2	Comp.3
b	0.498	0.361	0.026
plat	0.543	-0.095	0.372
term	-0.298	-0.703	-0.029
EX	0.404	-0.547	0.389
c1	0.452	-0.258	-0.842

```

uncor4<-cbind(uncor,uncor.pca.param$loadings[,1:3]) #just links up data

```

the model: Testing effect of PC1 and PC2, and their interaction with host on eclosion timing

```

pc.mod1<-lm(Eclosion~host*Comp.1+host*Comp.2,data=uncor4)
knitr::kable(round(summary(pc.mod1)$coefficients,3))

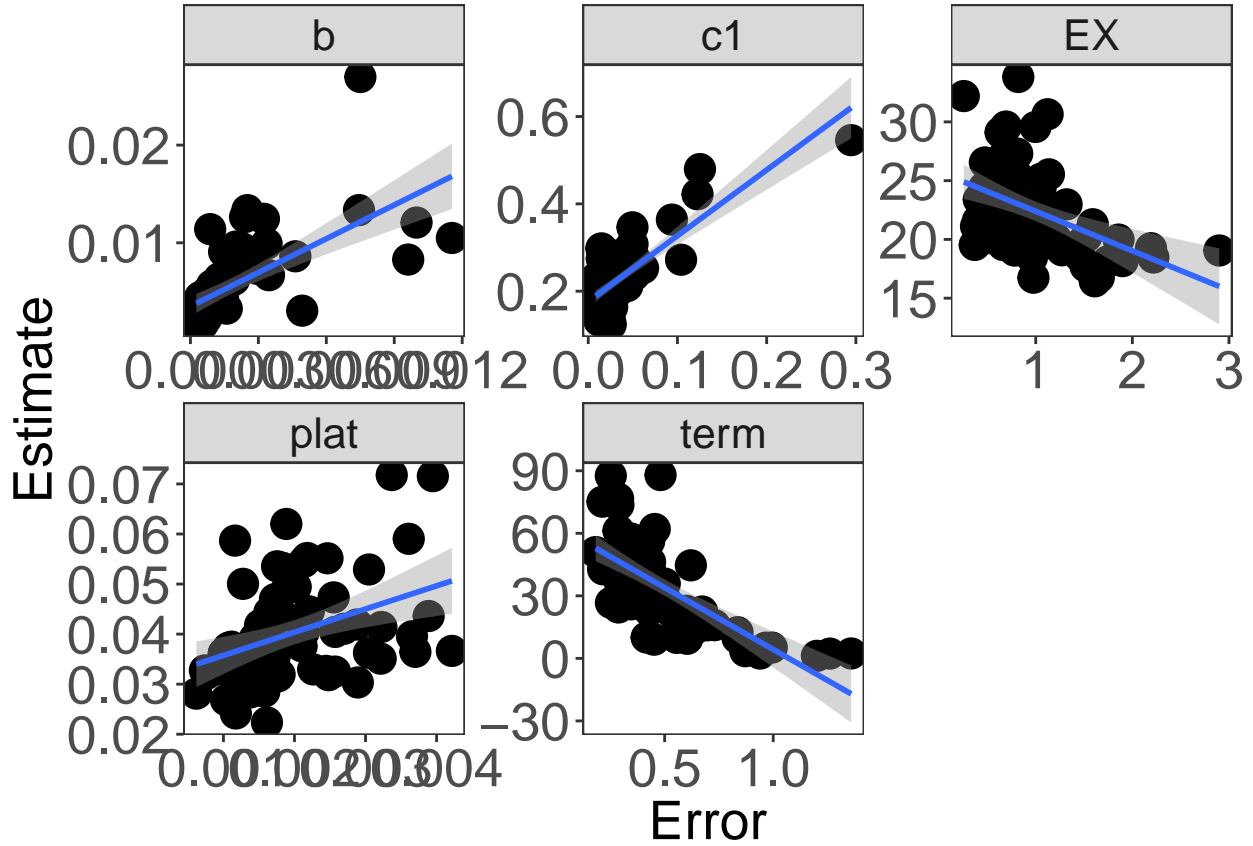
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	46.453	6.745	6.887	0.000
hostH	37.808	9.192	4.113	0.000
Comp.1	8.166	12.540	0.651	0.517
Comp.2	3.754	10.538	0.356	0.723
hostH:Comp.1	-12.728	17.089	-0.745	0.459
hostH:Comp.2	4.416	14.360	0.308	0.760

Accounting for error

The general approach to account for error was to regress parameters against the error of the parameters and take those residuals. These residuals should represent error free estimates of the parameters.

Estimates of the parameters are correlated with the error in those estimates



Multiple linear regression with parameters

Model construction: testing host x parameters interaction on eclosion timing

* Model selection based on AIC.

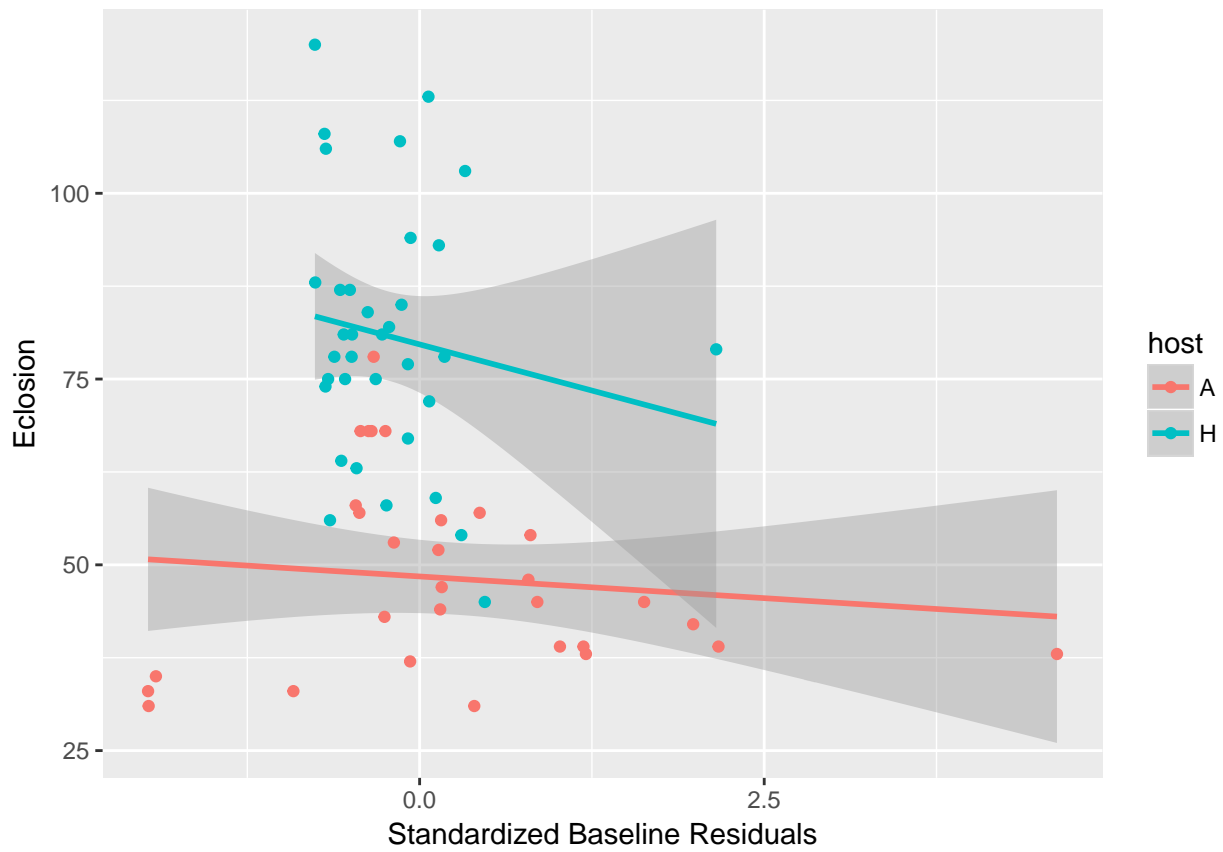
```
full.mod<-lm(Eclosion~host*b+host*c1+host*EX+host*plat+host*term,data=comb.dat.wide)
mod.sel1<-stepAIC(full.mod,direction="both")
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	47.544	2.181	21.799	0.000
hostH	26.359	2.882	9.146	0.000
b	-4.641	1.583	-2.932	0.005
c1	3.857	2.143	1.800	0.077
EX	3.339	3.218	1.037	0.304
plat	2.917	2.528	1.154	0.254
term	-1.750	2.902	-0.603	0.549
hostH:c1	-4.139	3.108	-1.332	0.189
hostH:EX	-6.483	3.963	-1.636	0.108
hostH:plat	6.176	4.704	1.313	0.195
hostH:term	18.417	3.324	5.541	0.000

There is a main effect of baseline and host; host by termination interaction

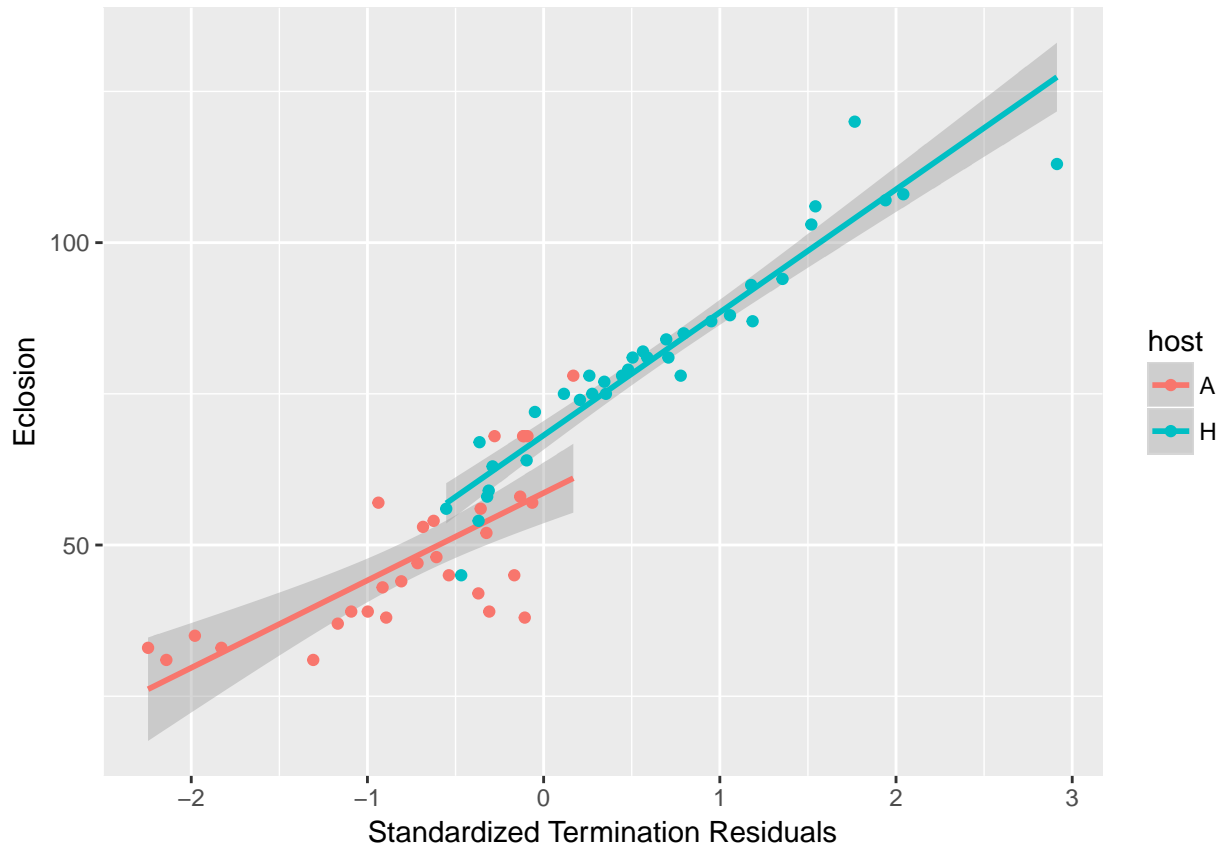
Main effect of host and baseline

```
ggplot(comb.dat.wide,aes(x=b,y=Eclosion,colour=host))+stat_smooth(method="lm")+geom_point()+xlab("Standardized Baseline Residuals")
```



Slope should be the same even though it doesn't really look like it.

host by termination interaction



Multiple linear regression with PCs

Grab PC's

	Comp.1	Comp.2	Comp.3
b	0.535	0.052	0.137
c1	0.333	0.478	-0.698
EX	0.393	-0.713	0.043
plat	0.596	-0.100	-0.115
term	-0.305	-0.500	-0.692

```
## Warning in data.frame(..., check.names = FALSE): row names were found from
## a short variable and have been discarded
```

the model: Testing effect of PC1 and PC2, and their interaction with host on eclosion timing

```
full.mod3<-lm(Eclosion~host*Comp.1+host*Comp.2,data=cdw.pc)
knitr::kable(round(summary(full.mod3)$coefficients,3))
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	45.565	4.520	10.080	0.000
hostH	40.266	6.160	6.536	0.000
Comp.1	6.974	9.467	0.737	0.464

	Estimate	Std. Error	t value	Pr(> t)
Comp.2	-2.577	7.278	-0.354	0.725
hostH:Comp.1	-19.754	12.902	-1.531	0.131
hostH:Comp.2	9.558	9.918	0.964	0.339

Conclusions

PCs dont explain any variation but the individual parameters do, so the assembly of modular components may not be important and the parameters themselves are acting independently. The ones that are important depends on whether or not we correct for error in the estimates of the parameters.

Table summary of stats

Parameter	Uncorrected	Corrected
Baseline (b)	no	p <0.05
Termination (term)	p <0.05	p <0.05
Plateau (plat)	no	no
Post diapause development (EX)	p <0.05	no
Scaling (c1)	p <0.05	no
Host x Termination	no	p <0.05