Chapter 7: Interactions

Make a note of the followings:

| Topic | Description | Comment |
| --- | --- | --- |
| Centering predictor variables | Centering predictors have two advantages:  1. much easier to compare models with and without interaction, as predictor means are zero.  2. A model is less prone to convergence error, as MAP value can be easily searched |  |
| Interaction between multiple continuous predictors | Use **tripletych** for graphing.  Make one predict constant and graph the relationship between the other predictor and the outcome variable. | Make sure to center predictors for ease of model’s interpretation and model comparisons |
| Interaction between categorical predictor and continuous predictor | Graph by holding one categorical variable constant | note that interaction can be interpreted in both directions for x1 and x2 predictors. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Practice

**#7H1**

m7.9=map(

alist(

blooms~dnorm(mu,sigma),

mu<-a+bW\*water.c+bS\*shade.c+bWS\*water.c\*shade.c,

a~dnorm(130,100),

c(bW,bS,bWS)~dnorm(0,100),

sigma~dunif(0,100)

),data=d,method="Nelder-Mead",start=list(a=mean(d$blooms),bW=0,bS=0,bWS=0,sigma=sd(d$blooms))

)

**#7H2**

m7h1=map(

alist(

blooms~dnorm(mu,sigma),

mu<-a[bed.i]+bW\*water.c+bS\*shade.c+bWS\*water.c\*shade.c,

a~dnorm(0,100),

c(bW,bS,bWS)~dnorm(0,100),

a[bed.i]~dnorm(0,100),

sigma~dunif(0,100)

),data=d,method="Nelder-Mead",start=list(a=mean(d$blooms),bW=0,bS=0,bWS=0,sigma=sd(d$blooms)),control=list(maxit=1e4)

)

precis(m7h1,depth=2)

precis\_plot(precis(m7h1,depth=2))

compare(m7.9,m7h1)

plot(compare(m7.9,m7h1))

post=extract.samples(m7h1)

str(post)

dens(post$a[,1],xlim=c(40,200))

dens(post$a[,2],col=rangi2,add=TRUE)

dens(post$a[,3],col=col.alpha("green",0.9),add=TRUE)

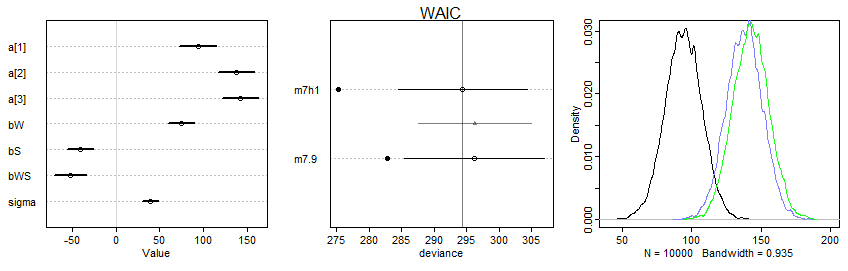


Figure. Density function of bed coefficients (a[1], a[2], and a[3]) (right figure) and m7h2 model coefficients (left)

Answer: WAIC results (left figure) matches with probability density functions (right) for each bed coefficient.

**#7H3**

(a)

data(rugged)

d <- rugged

d <- d[complete.cases(d$rgdppc\_2000),]

d$log\_gdp <- log(d$rgdppc\_2000)

d$rugged.c <- d$rugged - mean(d$rugged)

# (a)fit the model with and without Seychelles country

# with Seychelles

m7h3.s=map(

alist(

log\_gdp~dnorm(mu,sigma),

mu<-a+bRR\*rugged.c+bA\*cont\_africa+bAR\*cont\_africa\*rugged.c,

a~dnorm(0,10),

c(bRR,bA,bAR)~dnorm(0,1),

sigma~dunif(0,10)

),data=d

)

# without Seychelles

m7h3=map(

alist(

log\_gdp~dnorm(mu,sigma),

mu<-a+bRR\*rugged.c+bA\*cont\_africa+bAR\*cont\_africa\*rugged.c,

a~dnorm(0,10),

c(bRR,bA,bAR)~dnorm(0,1),

sigma~dunif(0,10)

),data=d[(d$country != "Seychelles"),]

)

# Tripletych plot

precis\_plot(precis(m7h3.s))

precis\_plot(precis(m7h3))

plot(log\_gdp~rugged.c,dd)

# with seychelles

par(mfrow=c(2,2))

tripletych.plot=function(model,text){

cont.fix=0:1

for(i in cont.fix){

dt=d[d$cont\_africa==i,]

rugged.seq=seq(from=-1.5,to=5,length.out = 50)

pred.data=data.frame(rugged.c=rugged.seq,cont\_africa=i)

mu=link(model,data=pred.data)

mu.mean=apply(mu,2,mean)

mu.PI=apply(mu,2,PI,prob=0.89)

plot(log\_gdp~rugged.c,dt,col=ifelse(i==0,col.alpha("black",0.9),rangi2),

main=ifelse(i==0,"Non-Africa","Africa"),

sub=ifelse(text=="Seychelles","Seychelles",""))

lines(rugged.seq,mu.mean,col=ifelse(i==0,col.alpha("black",0.9),rangi2))

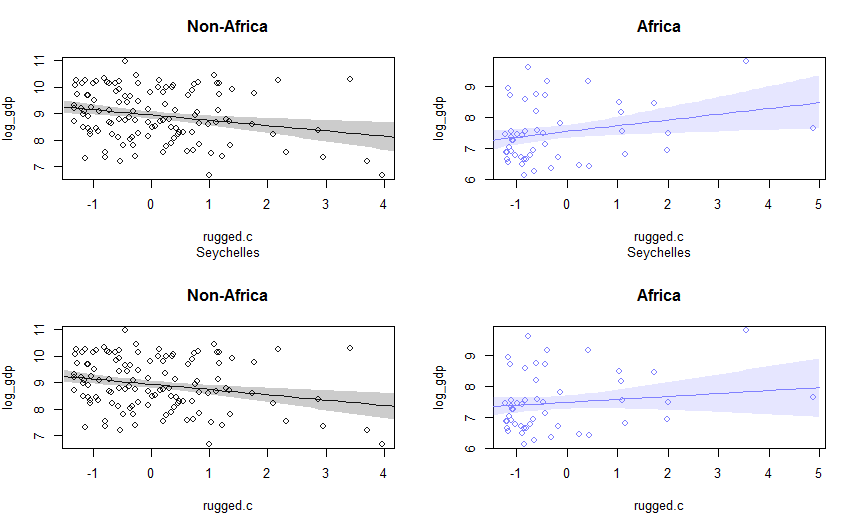
shade(mu.PI,rugged.seq,col=ifelse(i==0,col.alpha("black",0.2),col.alpha(rangi2,0.2)))

}

}

tripletych.plot(m7h3.s,"Seychelles")

tripletych.plot(m7h3,"")

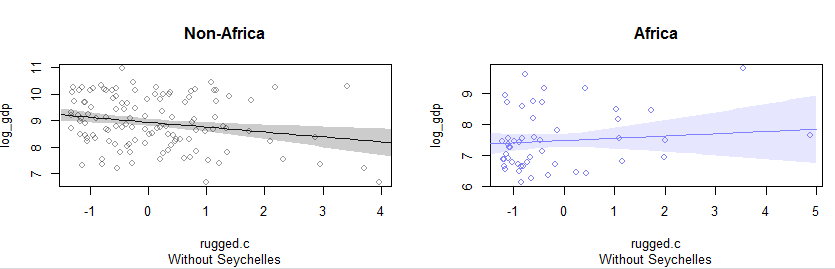


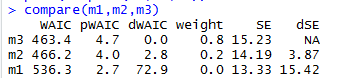
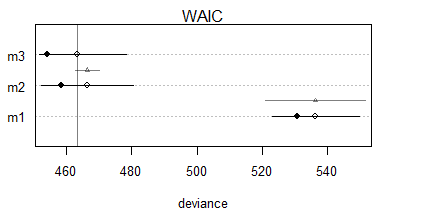
(b)

Answer:

Yes, the effect of ruggedness depend on continent, regardless of the presence of Seychelles. The expected relationship between ruggedness and log GDP changed little between models with and without Seychelles.

(c)

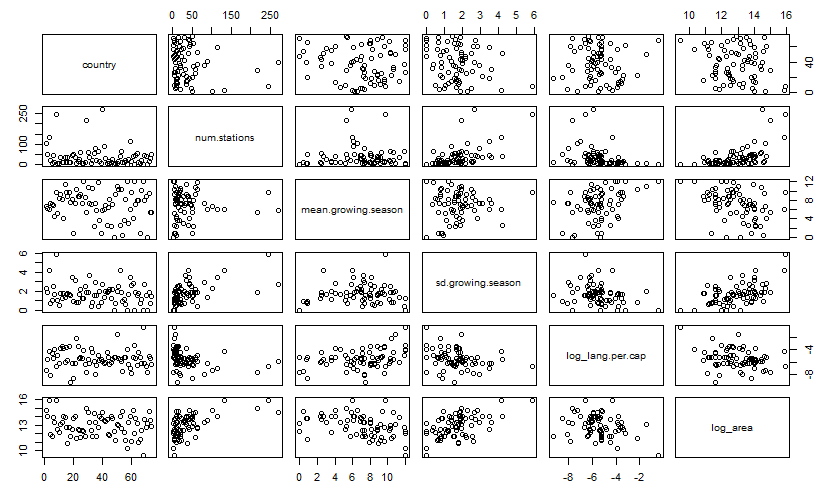




Answer:

In comparison to an interaction model without Seychelles (b) above, an ensemble model has little effect on the relationship between ruggedness and GDP for non-African nation. In contrast, uncertainty in the posterior mean has slightly increased for African nations under the ensemble. This is because a model m1 that has very high dSE with m3, the best model has led to widen the uncertainty when combined.

**#7H4**



(a)

Code:

data("nettle")

d=nettle

d$lang.per.cap=d$num.lang/d$k.pop

d$log\_lang.per.cap=log(d$lang.per.cap)

d$log\_area=log(d$area)

d$m.grow.season.c=d$mean.growing.season-mean(d$mean.growing.season)

pairs(d[,c(1,5,7,9,10,11)])

m\_a=map(

alist(

log\_lang.per.cap~dnorm(mu,sigma),

mu<-a+bG\*m.grow.season.c+bA\*log\_area,

a~dnorm(0,10),

c(bG,bA)~dnorm(0,1),

sigma~dunif(0,10)

),data=d)

precis(m\_a)

precis\_plot(precis(m\_a))

# Counterfactual plots

# holding log\_area constant at its log mean

log\_area.ave=mean(d$log\_area)

grow.s.seq=seq(from=-8,to=6,length.out = 50)

pred.data=data.frame(m.grow.season.c=grow.s.seq,log\_area=log\_area.ave)

mu=link(m\_a,data=pred.data)

mu.mean=apply(mu,2,mean)

mu.PI=apply(mu,2,PI,prob=0.89)

plot(log\_lang.per.cap~m.grow.season.c,d,col=rangi2)

mtext("Holding log\_area constant")

lines(grow.s.seq,mu.mean)

shade(mu.PI,grow.s.seq)

# holding growing season constant at its mean (i.e., mean = 0)

log\_area.seq=seq(from=-12,to=17,length.out = 50)

grow.s.ave=mean(d$m.grow.season.c)

pred.data=data.frame(m.grow.season.c=grow.s.ave,log\_area=log\_area.seq)

mu=link(m\_a,data=pred.data)

mu.mean=apply(mu,2,mean)

mu.PI=apply(mu,2,PI,prob=0.89)

plot(log\_lang.per.cap~log\_area,d,col=rangi2)

mtext("holding m.growing.season.c constant")

lines(log\_area.seq,mu.mean)

shade(mu.PI,log\_area.seq)

# Prediction Residual plots

# observed vs predicted

mu=link(m\_a)

mu.mean=apply(mu,2,mean)

mu.PI=apply(mu,2,PI,prob=0.89)

plot(d$log\_lang.per.cap~mu.mean,col=rangi2,xlab="Observed",ylab="Predicted")

mtext("Log\_lang.per.cap")

abline(a=0,b=1)

# residuals vs other predictors (missing variable?)

resid=d$log\_lang.per.cap-mu.mean

names(d)

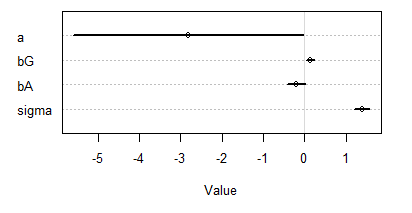
plot(resid~log(d$num.stations),col=rangi2);abline(h=0,lty=2)

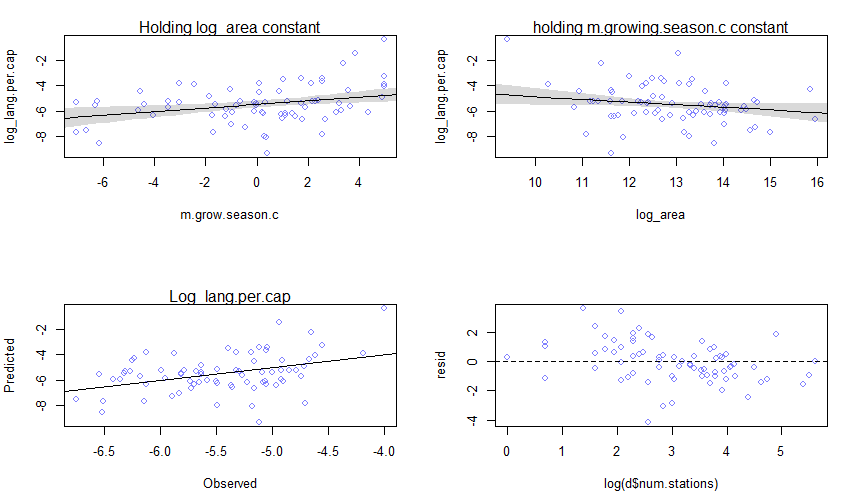
plot(resid~d$sd.growing.season,col=rangi2);abline(h=0,lty=2)

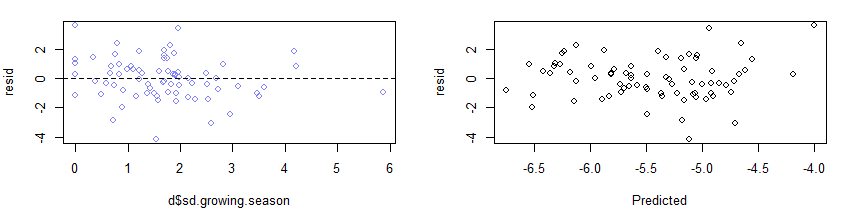
plot(resid~mu.mean,xlab="Predicted")

Answer:

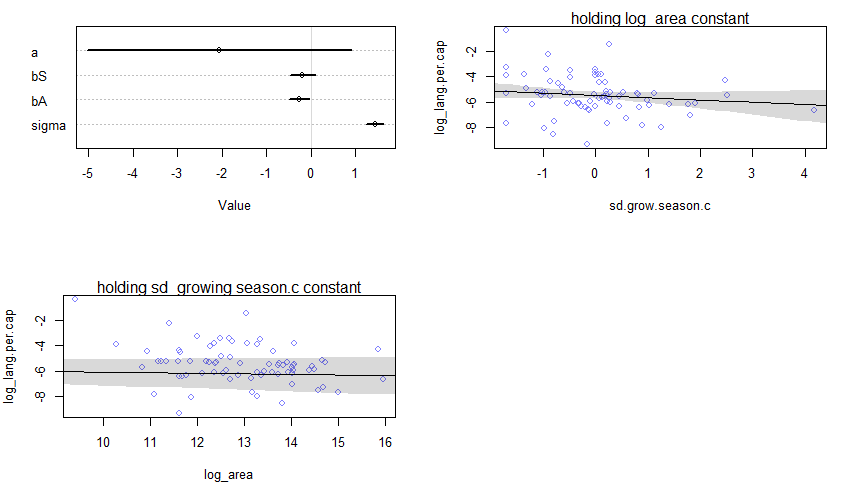
WAIC figure below displays a very wide WAIC for the intercept (a), while bG and bA (beta-coefficients) are close to zero, indicating both main terms have negligible effect on the language diversity. This is also supported by counterfacutal plots, with both log area and growing seasong constant their slopes are very gentle. In order to investigate any missing covaraites for this model, the other predictors such as the number of stations and sd growing season were graphed against the model. None of the predictors, however, appeared to have a relationship with the language diversity.







(b)



Answer:

WAIC indicates an extremely wide standard error for the intercept overlapping zero, while the model beta-coefficients (bS and bA) are very close to zero. But the SE of bS overlaps zero. Counterfactual plots show very weak relationship with the language diversity, as is supported by the WAIC.

(C)

Code:

m\_c=map(

alist(

log\_lang.per.cap~dnorm(mu,sigma),

mu<-a+bG\*m.grow.season.c+bS\*sd.grow.season.c+bGS\*m.grow.season.c\*sd.grow.season.c,

a~dnorm(0,5),

c(bG,bS,bGS)~dnorm(0,1),

sigma~dunif(0,10)

),data=d)

precis\_plot(precis(m\_c))

plot(d$sd.grow.season.c~d$m.grow.season.c,col=rangi2)

# holding m.grow.season constant

m.grow.s.ave=mean(d$m.grow.season.c)

sd.g.season.seq=seq(from=-2,to=5,length.out = 50)

pred.data=data.frame(m.grow.season.c=m.grow.s.ave,sd.grow.season.c=sd.g.season.seq)

mu=link(m\_c,data=pred.data)

mu.mean=apply(mu,2,mean)

mu.PI=apply(mu,2,PI,prob=0.89)

plot(log\_lang.per.cap~sd.grow.season.c,d,col=rangi2);mtext("holding mean growing season c constant")

lines(sd.g.season.seq,mu.mean)

shade(mu.PI,sd.g.season.seq)

# holding sd growins season constant

sd.g.season.ave=mean(d$sd.grow.season.c)

m.grow.s.seq=seq(from=-8,to=6,length.out = 50)

pred.data=data.frame(m.grow.season.c=m.grow.s.seq,sd.grow.season.c=sd.g.season.ave)

mu=link(m\_c,data=pred.data)

mu.mean=apply(mu,2,mean)

mu.PI=apply(mu,2,PI,prob=0.89)

plot(log\_lang.per.cap~m.grow.season.c,d,col=rangi2)

mtext("holding sd growing season c constant")

lines(m.grow.s.seq,mu.mean)

shade(mu.PI,m.grow.s.seq)

Answer:

Despite that beta-coefficients (bGS) is very close to zero, its standard error is nearly zero, indicating that the interaction between mean growing season and sd growing season appears to have a strong interaction relationship with the language diversity. Counterfactual plots displays that the language diversity declines with increasing variance of growing season length, while it linearly increases with mean growing season.

