Solutions Homework 4

Data Set Up

```
require(MARSS)
require(forecast)
phytos = c("Cryptomonas", "Diatoms", "Greens",
           "Unicells", "Other.algae")
yrs = lakeWAplanktonTrans[,"Year"]%in%1985:1994
dat = t(lakeWAplanktonTrans[yrs,phytos])
#z-score the data
avg = apply(dat, 1, mean, na.rm=TRUE)
sd = sqrt(apply(dat, 1, var, na.rm=TRUE))
dat = (dat-avg)/sd
rownames(dat)=phytos
#z-score the covariates
covars = rbind(Temp=lakeWAplanktonTrans[yrs,"Temp"],
               TP=lakeWAplanktonTrans[yrs,"TP"])
avg = apply(covars, 1, mean)
sd = sqrt(apply(covars, 1, var, na.rm=TRUE))
covars =  (covars-avg)/sd
rownames(covars) = c("Temp","TP")
#
#always check that the mean and variance are 1 after z-scoring
apply(dat,1,mean,na.rm=TRUE) #this should be 0
apply(dat,1,var,na.rm=TRUE) #this should be 1
```

Set up constants.

```
TT = dim(dat)[2] # length of time series
m = n = dim(dat)[1] # number of states & obs
period = 12 # data were collected monthly
```

Set up the model structures that will be used across all the questions.

```
common=list(
  B = "diagonal and unequal",
  U = "zero",
  Q = "diagonal and unequal",
  Z = "identity",
  A = "zero",
  R = "diagonal and equal",
  tinitx = 0
)
ctl = list(maxit=500) #in case we want to compare
```

Note the models do not converge in the default of `maxit=500`. You could either set `maxit` higher by passing in `control=list(maxit=2000)` or try `method="BFGS"`. However, here results are shown with the defaults and in some cases, the models have not converged.

Problem 1

How does month affect the mean phytoplankton population growth rates? Show a plot of mean growth rate versus month. Estimate seasonal effects without any covariate (Temp, TP) effects.

Since the covariate affects growth rate, it appears in the **x** equation not the **y** equation. Thus we want to set up **C** and **c**.

Option 1: Treat month as a factor and you will estimate a month effect for each month. You need a row for each month (Jan, Feb, ...) and a column for each time point. You put a 1 in the Jan row for each Jan time point, repeat for the other months. The following code will create such a **c** matrix.

```
c.fac = diag(period)
for(i in 2:(ceiling(TT / period))) { c.fac = cbind(c.fac, diag(period)) }
the.months = month.abb
rownames(c.fac) = the.months
```

We want each month to have separate effect on each taxon (60 effects), so our **C** matrix can just be specified as `"unconstrained"`. Our covariates are only in **x** so **D** and **d** are zero.

We set up the model list and fit:

```
C = "unconstrained"; c=c.fac
D="zero"; d="zero"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q1a = MARSS(dat, model=model.list, control=ctl)
```

Get the appropriate seasonal effects:

```
seas.a = coef(q1a,type="matrix")$C
rownames(seas.a) = phytos
colnames(seas.a) = the.months
```

We could plot the seasonal effect by species with this code:

```
matplot(t(seas.a),type="l",bty="n",xaxt="n", ylab="Fixed monthly", col=1:5)
axis(1,labels=the.months, at=1:12,las=2,cex.axis=0.75)
legend("topright", lty=1:5, legend=phytos, cex=0.6, col=1:5)
```

Note, if we had set `U="unequal"`, we would need to set one of the columns of **C** to zero because the we would have created an under-determined problem (infinite solutions). Basically, we have a problem like $10 = y + z$. You can't solve for both y and z in that case.

Option 2: The factor approach required estimating 60 effects. Another approach is to model the temperature effect as a 3rd order (or higher) polynomial $bm + cm^2 + dm^3$. This approach has less flexibility but requires only 20 **C** parameters.

First we make month cov matrix with our m, m^2, and m^3 in different rows:

```
poly.order = 3
month.cov = matrix(1,1,period)
for(i in 1:poly.order) { month.cov = rbind(month.cov,(1:12)^i)  }
# for c, month.cov is replicated 10 times (once for each year)
c.m.poly = matrix(month.cov, poly.order+1, TT, byrow=FALSE)
```

Everything except **c** remains the same.

```
C = "unconstrained"; c=c.m.poly
D="zero"; d="zero"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q1b = MARSS(dat, model=model.list, control=ctl)
```

The effect of month m $C_1 m + C_2 m^2 + C_3 m^3$. We can make our seasonal effect matrix (like seas above):

```
C.b = coef(q1b,type="matrix")$C
seas.b = C.b %*% month.cov
rownames(seas.b) = phytos
colnames(seas.b) = the.months
```

Option 3: The factor approach required estimating 60 effects, and the 3rd order polynomial model was an improvement at only 20 parameters. Another option is using a combination of sine and cosine waves, which would require only 10 parameters.

Begin by defining the seasonal covariates as

```
cos.t = cos(2 * pi * seq(TT) / period)
sin.t = sin(2 * pi * seq(TT) / period)
c.Four = rbind(cos.t,sin.t)
```

Everything except **c** remains the same.

```
C = "unconstrained"; c=c.Four
D="zero"; d="zero"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q1c = MARSS(dat, model=model.list, control=ctl)
```

Now make our seasonal effect matrix (like seas above). The net seasonal effect is then $C_a \cos(2\pi t/12) + C_b \sin(2\pi t/12)$.

```
C.c = coef(q1c,type="matrix")$C
seas.c = C.c %*% c.Four[,1:period]
rownames(seas.c) = phytos
colnames(seas.c) = the.months
```
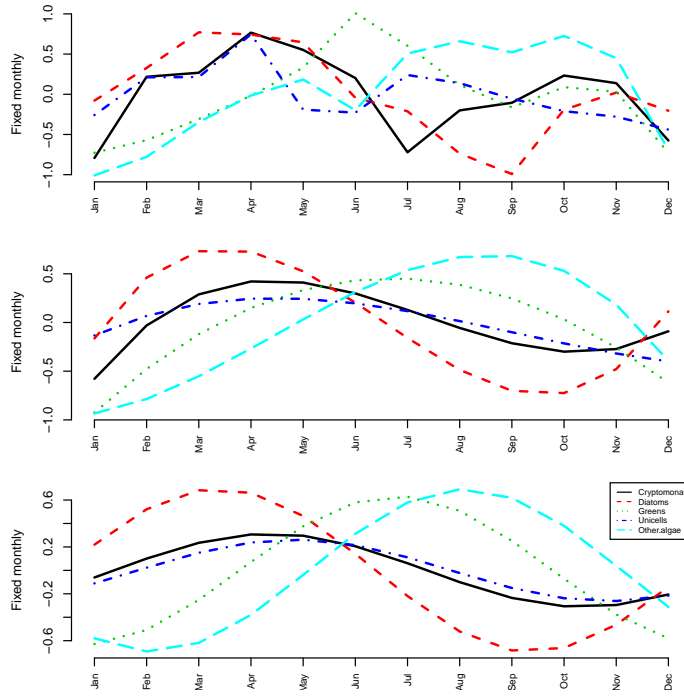

Fig. 0.1. Our model had the season effect in the process part of the model. That means this graph is showing when growth rates, not abundance, is highest.

Problem 2

It is likely that both temperature and total phosphorus (TP) affect phytoplankton population growth rates. Evaluate which is the more important driver or if both are important. Leave out the seasonal covariates from question 1, i.e. only use Temp and TP as covariates.

There are two ways that you might approach this. One is to fit a model with both covariates and look at the **C** matrix. The other is to fit different models and compare model support.

Option 1: Fit a model with both covariates and look at the estimated effects (Figure). We get the estimates but it is not obvious how to compare which is more important. That takes us to option 2.

```
C = "unconstrained"
c = covars
D = d = "zero"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q2.both = MARSS(dat, model=model.list, control=ctl)
```

Option 2: Use model selection to compare support. The idea here is to fit different MARSS models: one where temperature affects growth rates, one where TP affects growth rates, and one where both affect growth rate. The question did not specify what (if any) lags to use for your covariates. We are going to use lag-0 for both temperature and TP. However we will see at the end this is not a a very good model.

Our **c**, **D** and **d** matrices will stay the same, so we set-up those up once.

```
c = covars
D = d = "zero"
```

Create a **C** matrix where temperature (only) has a linear effect on population growth. That means the effect of TP is 0. Then fit:

```
C = matrix(list(0),5,2); C[,1]=paste("Temp",1:5,sep="")
C

     [,1]     [,2]
[1,] "Temp1" 0
[2,] "Temp2" 0
[3,] "Temp3" 0
[4,] "Temp4" 0
[5,] "Temp5" 0

model.list = c(common, list(C=C,c=c,D=D,d=d))
q2.Temp = MARSS(dat, model=model.list, control=ctl)
```

```
plot(c(rep(1,5),rep(2,5)),coef(q2.both)$C,xlab="",ylab="effect",xaxt="n",xlim=c(0,3))
axis(1, at=c(1,2),labels=c("Temp","TP"))
abline(h=0)
title("effect (C) estimates)")
```
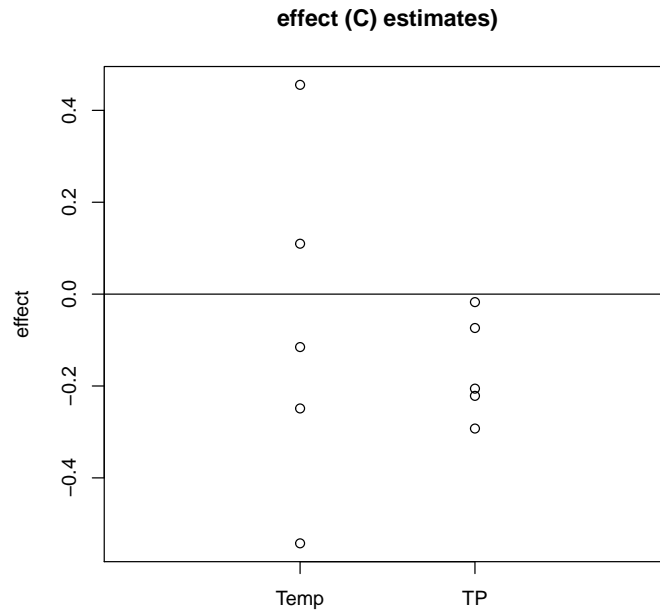
effect (C) estimates)

Fig. 0.2. Estimated effect sizes for all taxon.

Create a model where TP (only) has a linear effect on population growth and fit:

```
C = matrix(list(0),5,2); C[,2]=paste("TP",1:5,sep="")
C

      [,1] [,2]
[1,] 0    "TP1"
[2,] 0    "TP2"
[3,] 0    "TP3"
[4,] 0    "TP4"
[5,] 0    "TP5"
```

```
model.list = c(common, list(C=C,c=c,D=D,d=d))
q2.TP = MARSS(dat, model=model.list, control=ctl)
```

Create a model where both temperature and TP have linear effects:

```
C = "unconstrained"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q2.both = MARSS(dat, model=model.list, control=ctl)
```

Use AIC to compare the 3 models:

```
data.frame(Model=c("Temp", "TP", "Both"),
           AICc=round(c(q2.Temp$AICc, q2.TP$AICc, q2.both$AICc), 1))
```

The AICc for the Temp model is 31.83 units lower than that for TP, offering support for lag-0 temperature being more important than lag-0 TP in explaining phytoplankton community abundance. The model including both Temp and TP is <2 different in AICc units than the Temp-only model, which suggests that adding an extra covariate (and parameter) to the model does not really improve model fit.

However, none of the Temp or TP models are particularly good compared to any of the purely seasonal (month) effect models. It's easy to compare all of them using AICc.

```
mod.names = c("Temp", "TP", "Both", "Fixed", "Cubic", "Fourier")
AICc.all = c(q2.Temp$AICc,q2.TP$AICc,q2.both$AICc,q1a$AICc,q1b$AICc,q1c$AIC)
delta.AICc = round(AICc.all - min(AICc.all), 1)
# model selection results sorted from best (top) to worst (bottom)
data.frame(Model=mod.names, delta.AICc=delta.AICc)[order(delta.AICc),]
```

```
    Model delta.AICc
6 Fourier        0.0
5   Cubic        9.7
4   Fixed       28.4
3    Both       30.4
1    Temp       31.3
2      TP       63.1
```

Diagnostics Temperature and TP model residuals for all taxa except Greens appear to show significant negative autocorrelation at lag=1 (Figure and Figure), suggesting that both models are inadequate to capture all of the systematic variation in phytoplankton abundance.

```
par(mfrow=c(5,2), mai=c(0.6,0.6,0.2,0.2), omi=c(0,0,0,0))
for(i in 1:5) {
  plot.ts(residuals(q2.Temp)$model.residuals[i,], ylab="Residual", main=phytos[i])
        abline(h=0, lty="dashed")
        acf(residuals(q2.Temp)$model.residuals[i,])
        }
```
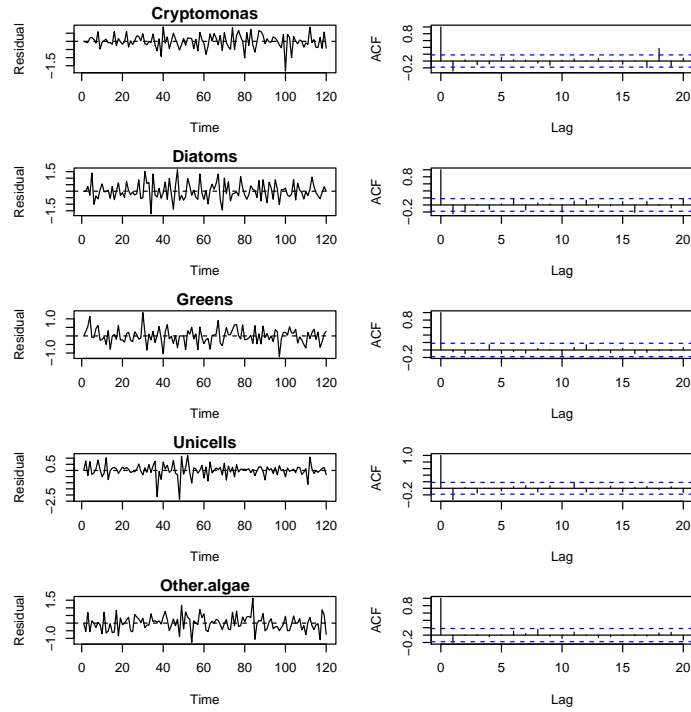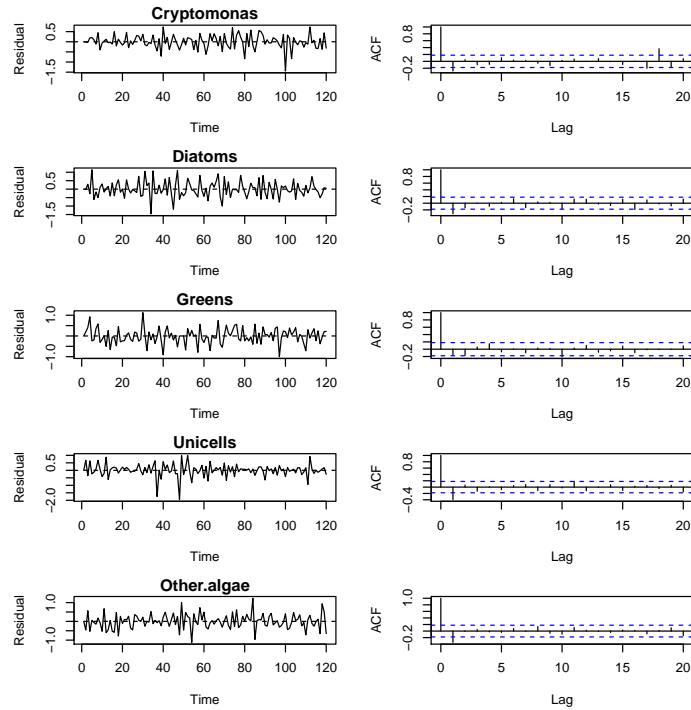

Fig. 0.3. Plot the temperature (q2.Temp) model residuals and ACF's for all 5 taxa.

```
par(mfrow=c(5,2), mai=c(0.6,0.6,0.2,0.2), omi=c(0,0,0,0))
for(i in 1:5) {
        plot.ts(residuals(q2.TP)$model.residuals[i,], ylab="Residual", main=phytos[i])
        abline(h=0, lty="dashed")
        acf(residuals(q2.TP)$model.residuals[i,])
        }
```


Fig. 0.4. Plot of TP model (q2.TP)residuals and ACF's for all 5 taxa.

Problem 3

Evaluate whether the effect of temperature on phytoplankton manifests itself via their underlying physiology (by affecting algal growth rates and thus abundance) or because physical changes in the water stratification makes them easier/harder to sample in some months. Leave out the seasonal covariates from question 1, i.e. only use Temp and TP as covariates.

The idea here is to fit two different MARSS models: one with the Temp effect in the process eqn and the other with it in the observation equation. We already fit the temperature-effect model for the process in the 1st part of question 2, so now fit one with temperature in the observation equation instead. Now we want $\mathbf{D}$ and $\mathbf{d}$ instead of $\mathbf{C}$ and $\mathbf{c}$ that we used for the process equation.

```
D = "unconstrained"
d = covars["Temp",,drop=FALSE]
C = c = "zero"
```

Create appropriate model list and fit the model.

```
model.list = c(common, list(C=C,c=c,D=D,d=d))
q3.Tobs = MARSS(dat, model=model.list, control=ctl)
```

Now use AIC to compare the 2 models.

```
data.frame(Model=c("proc", "obs"),
           AICc=round(c(q2.Temp$AICc, q3.Tobs$AICc), 1))

  Model   AICc
1  proc 1480.9
2   obs 1500.7
```

The AICc for the model with a temperature effect in the process equation is 19.83 units lower than the model with the effect in the observation equation. Thus, we conclude that the data support the hypothesis of a temperature effect on the physiology more than an observation/sampling phenomenon.

Diagnostics Similar to the temperature-effect in the process model, residuals for these models also show significant negative autocorrelation at lag=1 (Figure not shown but code is similar to that used in the previous question).

Problem 4

Is there support for temperature or TP affecting all functional groups' growth rates the same, or are the effects on one taxon different from another?

Option 1: Analyze temperature and TP effects one at a time. The problem with this is that perhaps including both in the model might change the conclusions. The idea here is to fit four different MARSS models (2 each for temperature and TP). Each one will assume covariates affect the process. Two models will have effects that vary by taxon as in Q2 (ie, C = "unconstrained") whereas the other 2 will assume the covariates affect all taxa equally.

```
# 1st: the Temp model
C = "equal"
c = covars["Temp",,drop=FALSE]
D = d = "zero"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q4.Temp = MARSS(dat, model=model.list, control=ctl)
```

Use AIC to compare the 2 Temp models

```
data.frame(Effect=c("taxon-specific", "equal"),
           AICc=round(c(q2.Temp$AICc, q4.Temp$AICc), 1))
```

The AICc for the model with taxon-specific temperature effects is greater than 2 lower than for the model that assumes a equal temperature effect, which supports the hypothesis of varying temperature effects by taxon.

```
# 2nd: the TP model
C = "equal"
c = covars["TP",,drop=FALSE]
D = d = "zero"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q4.TP = MARSS(dat, model=model.list, control=ctl)
```

Now use AIC to compare the 2 TP models

```
data.frame(Effect=c("taxon-specific", "equal"),
           AICc=round(c(q2.TP$AICc, q4.TP$AICc), 1))
```

The AICc for the TP model with taxon-specific effects is greater than 2 lower, which strongly supports the hypothesis of non-equal TP effects. Repeating the same diagnostic plot used in question 2 reveals that these models also have autocorrelation in the residuals suggesting that these models are also still inadequate.

Option 2: Analyze temperature and TP effects together in one model. The problem with this is that temperature and TP are correlated because they both have strong seasonal cycle. The idea here is to fit four different MARSS

models. Each one will assume covariates affect the process. Two models will have effects that vary by taxon as in Q2 for one covariate. The other two models will have unconstrained **C** or **C** equal for both covariates (but not across covariates).

```
# 1st: unconstrained
C = "unconstrained"
c = covars
D = d = "zero"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q4.unc = MARSS(dat, model=model.list, control=ctl)

# 2nd: the TP unconstrained, Temp constrained
C = matrix("Temp",5,2)
C[,2] = paste("TP",1:5)
c = covars
D = d = "zero"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q4.TPunc = MARSS(dat, model=model.list, control=ctl)

# 3rd: the TP constrained, Temp unconstrained
C = matrix("TP",5,2)
C[,1] = paste("Temp",1:5)
c = covars
D = d = "zero"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q4.Tempunc = MARSS(dat, model=model.list, control=ctl)

# 4th: both constrained
C = matrix("TP",5,2)
C[,1] = "Temp"
c = covars
D = d = "zero"
model.list = c(common, list(C=C,c=c,D=D,d=d))
q4.cons = MARSS(dat, model=model.list, control=ctl)

data.frame(Effect=c("unconstrained", "TP cons", "Temp cons", "both cons"),
              AICc=round(c(q4.unc$AICc, q4.Tempunc$AICc,
                           q4.TPunc$AICc, q4.cons$AICc), 1))
```

```
        Effect   AICc
1 unconstrained 1480.0
2       TP cons 1474.6
3     Temp cons 1510.8
4     both cons 1545.5
```

Problem 5

Compare your results for questions 2-4 using a *process* error only model, by using the lm() function.

You were asked to use an observation error only model, but I don't think you can do that when you estimate b because it leads to an indeterminate problem (infinite solutions). If I'd had you fix $b = 1$ then you could have solved it by using $y(t) - y(t-1)$ as your data.

But if you assume your data have no observation error and that there is only process error, you can solve this. In this case, the model is

$$y(t) = y(t-1) + \alpha_{month} + w(t) \tag{0.1}$$

if you are investigating a month effect or

$$y(t) = y(t-1) + \beta_{temp}temperature + \beta_{tp}TP + w(t) \tag{0.2}$$

if you are investigating a temperature and tp effect. Notice I did not include a intercept since all the data are de-meaned.

lm() works best with a dataframe. We will have one column for the response variable (phytoplankton abundance), a factor column for taxon, a factor column for month, and two columns for temperature and TP (which are the same for each taxon). I need the abundance in the previous month as one of my predictors. Always be careful to match your covariate t to the right response t. How to match them up depends on your problem or question.

```
dat.frame = data.frame(
  abund = as.vector(t(dat)),
  abund.lag1 = as.vector(t(cbind(NA,dat[,1:119]))),
  taxon = rep(phytos,each=120),
  month = rep(month.abb, 5*10),
  temp = rep(covars[1,],5),
  tp = rep(covars[2,],5)
)
```

Take a look at the data frame as it changes to another taxon. Notice how the abundance in the previous month is set to NA as it should be in Jan at the start of the time series.

```
dat.frame[118:125,1:5]
```

```
        abund abund.lag1       taxon month        temp
118 -0.8623450 -0.2102360 Cryptomonas   Oct  0.846632613
119  0.7092607 -0.8623450 Cryptomonas   Nov  0.001513982
120 -0.7070257  0.7092607 Cryptomonas   Dec -0.905934438
121 -0.7413776         NA     Diatoms   Jan -1.233063992
122  0.2085792 -0.7413776     Diatoms   Feb -1.464295294
123  1.0206142  0.2085792     Diatoms   Mar -1.340613435
```

```
124  0.8039275  1.0206142    Diatoms    Apr -0.846483496
125  1.5061781  0.8039275    Diatoms    May -0.165187653
```

Question 1: Asks us to estimate the season effect on population growth rate. We will just do this for the case where we treat season as a factor. First let's set up a holder for the season effects.

```
seaslm = matrix(NA, 5,12, dimnames=list(phytos, month.abb))
```

Here is how we fit an lm() model with a month effect for each taxon. The `taxon:abund.lag1` part is the b estimates and the `taxon:month` part is the month effects by taxon.

```
fit = lm(abund ~ -1 + taxon:month + taxon:abund.lag1, data=dat.frame)
```

This gives us the month effects for each taxon but they are in alphabetical order. I'm going to use the stringr package to help get these coefficients out. Also I need to re-order the months from alphabetical to temporal.

```
require(stringr)
mon.order = c(5,4,8,1,9,7,6,2,12,11,10,3)
for(taxon in phytos){
  taxoncoef = str_detect(names(coef(fit)),paste(taxon,":month",sep=""))
  tmp = coef(fit)[taxoncoef]
  #order the months
  seaslm[taxon,] = tmp[mon.order]
}
```

We could plot the seasonal effect by species with this code. Figure compares the estimates with the state-space model.

```
matplot(t(seaslm),type="l",bty="n",xaxt="n", ylab="Fixed monthly", col=1:5)
axis(1,labels=month.abb, at=1:12,las=2,cex.axis=0.75)
legend("topright", lty=1:5, legend=phytos, cex=0.6, col=1:5)
```

The results are quite similar to the state-space model.

Question 2: For this one, we will fit models with temperature and TP as explanatory variables. First let's set up holders for the season effects.

```
Clm = matrix(NA, 5,2, dimnames=list(phytos, c("Temp","TP")))
```

Then fit to the differenced data and store the estimates. Figure compares the estimates from the state-space model versus lm(). They are similar but the non-state-space model has noisier estimates.

```
fit=lm(abund ~ -1 + taxon:abund.lag1 + taxon:temp + taxon:tp,
       data=dat.frame)
require(stringr)
```

MARSS estimates

lm estimates no observation error
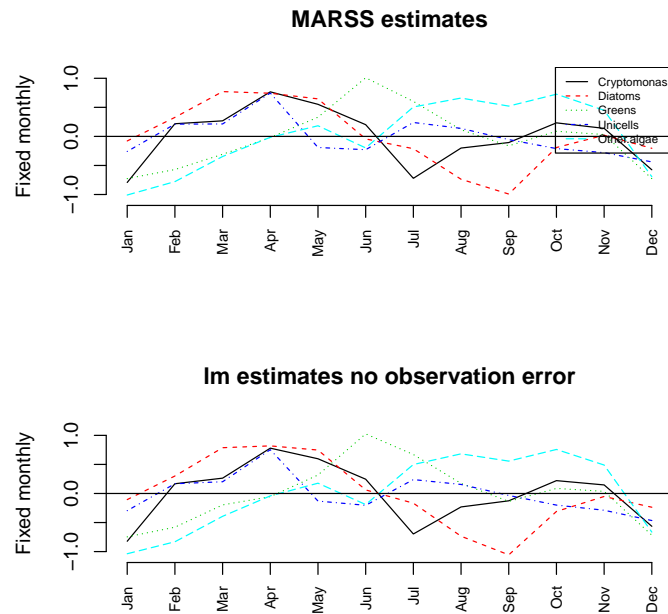
Fig. 0.5. Comparison of the monthly population growth rate effect sizes.

```
taxoncoef = str_detect(names(coef(fit)),":temp")
Clm[,1] = coef(fit)[taxoncoef]
taxoncoef = str_detect(names(coef(fit)),":tp")
Clm[,2] = coef(fit)[taxoncoef]
```

Another approach would be to compare AICs but we have to do this taxon by taxon.

```
fit=lm(abund ~ -1 + taxon:abund.lag1 + taxon:temp + taxon:tp,
        data=dat.frame)
fit.temp=lm(abund ~ -1 + taxon:abund.lag1 + taxon:temp,
        data=dat.frame)
fit.tp=lm(abund ~ -1 + taxon:abund.lag1 + taxon:tp,
        data=dat.frame)
c(both=AIC(fit), temp.only=AIC(fit.temp), tp.only=AIC(fit.tp))

    both temp.only   tp.only
1459.625   1461.234  1486.226
```
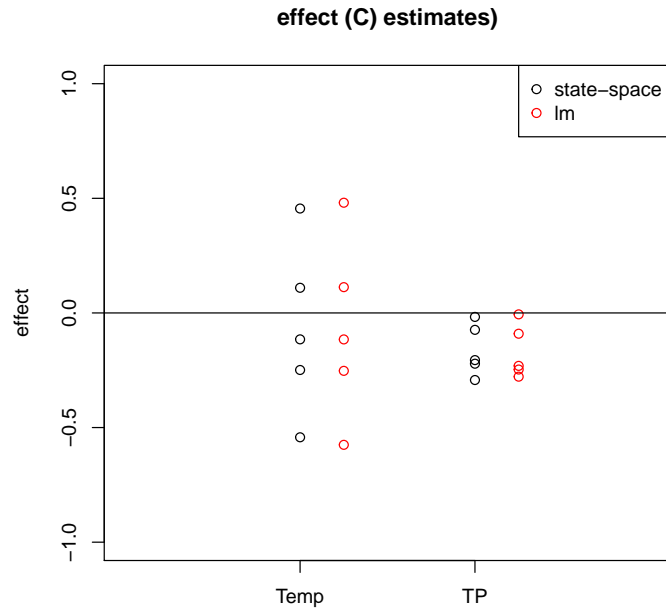
effect (C) estimates)

Fig. 0.6. Estimated effect sizes for all taxon.

Question 3: You cannot address that question without a state-space model.

Question 4: To answer this question, we can use AIC again.

```
fit1=lm(abund ~ -1 + taxon:abund.lag1 + taxon:temp + taxon:tp,
        data=dat.frame)
fit2=lm(abund ~ -1 + taxon:abund.lag1 + taxon:temp + tp,
        data=dat.frame)
fit3=lm(abund ~ -1 + taxon:abund.lag1 + temp + taxon:tp,
        data=dat.frame)
fit4=lm(abund ~ -1 + taxon:abund.lag1 + temp + tp,
        data=dat.frame)
```

The best model is the same as what we got using the state-space model.

```
    AIC.lm=c(AIC(fit1), AIC(fit2),
                        AIC(fit3), AIC(fit4))
AICc.ss=c(q4.unc$AICc, q4.Tempunc$AICc,
                        q4.TPunc$AICc, q4.cons$AICc)
```

```
    data.frame(
    Effect=c("unconstrained", "TP cons", "Temp cons", "both cons"),
        del.AIC.lm=round(AIC.lm-min(AIC.lm), 1),
        del.AICc.ss=round(AICc.ss-min(AICc.ss), 1)
 )

        Effect del.AIC.lm del.AICc.ss
1 unconstrained        4.7         5.4
2       TP cons        0.0         0.0
3     Temp cons       31.6        36.2
4     both cons       60.4        70.8
```

Problems with the analysis and data

This homework is to get you digging into a data set and trying to tackle some semi-open-ended questions. There is no one way to answer the questions. The approaches taken in this answer key build on the material you have been learning so far in the class.

However, there are some problems with the approaches taken here. First, in all the analyses using temperature and TP, we did not account for seasonality in the data and the covariates. This seasonality causes high correlation between the two covariates which will problems in any analysis that includes both. Here are two common approaches to this problem:

- We could deseason both the response variables and covariates and proceed from there. This is the approach of using anomalies. We might do this if seasonality is not what we are trying to understand and we don't think the explanatory variable have different effect sizes in different seasons. We can remove seasonality by fitting a regression or ANOVA to the data and running the analysis on the residuals.
- Another approach would be to use month as an covariate and then use the residuals from say $Temp = a + \beta month$ as another covariate. The latter is then an indicator of months that are unusually hot or cold relative the month of the year.

If we use models with one covariate, say temperature only without TP, we don't have a problem of correlation between our covariates because there is only one of them. However even when we use models with one covariate, there are problems with leaving out the seasonality. You might think that by leaving out a season term in our model, we are exploring whether temperature or TP is more important in driving the seasonality in phytoplankton. But we saw in our seasonality analysis that growth rates are highest in spring and lowest in summer and winter. So growth rates are low when the temperature is both low and high. Clearly a model where temperature (and TP) has a linear effect doesn't make sense. We could try using a non-linear effect using polynomials of the covariates. What about using a temperature and month

effect? In lm notation this would be a `month:temperture` term where month is a factor. Now instead of estimating one temperature effect per taxon, we estimate 12 (one for each month). That is worth exploring. That would allow unconstrained non-linear temperature effects.

We only used lag-0 in our models. That is we used temperature at time t to drive growth rate from $t-1$ to t. Almost certainly some kind of cumulative temperature (degree days) is a better covariate. It may also that temperature and TP simply have a lagged response, so lag-2 or lag-3 would be better. You could use to explore this (`ccf()`), but you would have to deal with the seasonality first. The seasonality will mask other lagged correlations that you are looking for.

The similarity between the lm() and state-space model suggests some problems and mysteries:

1. Perhaps there is not much observation error in the data. That seems unlikely.
2. Perhaps the model is so flexible and the process variance estimates so large that it can fit the data with no observation error. This seems more likely.
3. The b terms need to be investigated. They are very close to zero indicating that $x(t)$ is almost independent of $x(t-1)$. Does that make sense or is there something wrong with the model? If b is close to zero, you have data that cannot be distinguished from white noise, no autoregression. This is odd for data that are counts of abundance.