Homework 6 - Answer Key

Data Set Up

```r
require(MARSS)
```

Loading required package: MARSS

```r
require(forecast)
```

Loading required package: forecast

```r
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
```

```
  Cryptomonas        Diatoms         Greens       Unicells    Other.algae
 2.329499e-17   1.463504e-17  -2.761472e-19   3.546358e-17   1.507041e-18
```

```r
apply(dat, 1, var, na.rm = TRUE)  #this should be 1
```

```
Cryptomonas     Diatoms      Greens    Unicells Other.algae
          1           1           1           1           1
```

Set up constants.

```r
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.

```r
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)
```

```
Success! abstol and log-log tests passed at 486 iterations.
Alert: conv.test.slope.tol is 0.5.
Test with smaller values (<0.1) to ensure convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
Estimation converged in 486 iterations.
Log-likelihood: -651.766
AIC: 1455.532    AICc: 1478.04
```

	Estimate
R.diag	0.0958
B.(X.Cryptomonas,X.Cryptomonas)	0.3121
B.(X.Diatoms,X.Diatoms)	0.4657
B.(X.Greens,X.Greens)	0.3204
B.(X.Unicells,X.Unicells)	0.6495
B.(X.Other.algae,X.Other.algae)	0.2015
Q.(X.Cryptomonas,X.Cryptomonas)	0.5537
Q.(X.Diatoms,X.Diatoms)	0.2421
Q.(X.Greens,X.Greens)	0.4664
Q.(X.Unicells,X.Unicells)	0.4141
Q.(X.Other.algae,X.Other.algae)	0.4312
x0.X.Cryptomonas	-0.0766

```
x0.X.Diatoms                 -1.3587
x0.X.Greens                   2.7993
x0.X.Unicells                 1.2676
x0.X.Other.algae             -1.5274
C.(X.Cryptomonas,Jan)        -0.7928
C.(X.Diatoms,Jan)            -0.0783
C.(X.Greens,Jan)             -0.7265
C.(X.Unicells,Jan)           -0.2588
C.(X.Other.algae,Jan)        -1.0087
C.(X.Cryptomonas,Feb)         0.2178
C.(X.Diatoms,Feb)             0.3275
C.(X.Greens,Feb)             -0.5705
C.(X.Unicells,Feb)            0.2111
C.(X.Other.algae,Feb)        -0.7783
C.(X.Cryptomonas,Mar)         0.2678
C.(X.Diatoms,Mar)             0.7709
C.(X.Greens,Mar)             -0.3081
C.(X.Unicells,Mar)            0.2141
C.(X.Other.algae,Mar)        -0.3459
C.(X.Cryptomonas,Apr)         0.7667
C.(X.Diatoms,Apr)             0.7427
C.(X.Greens,Apr)             -0.0183
C.(X.Unicells,Apr)            0.7463
C.(X.Other.algae,Apr)        -0.0156
C.(X.Cryptomonas,May)         0.5519
C.(X.Diatoms,May)             0.6458
C.(X.Greens,May)              0.3305
C.(X.Unicells,May)           -0.1908
C.(X.Other.algae,May)         0.1830
C.(X.Cryptomonas,Jun)         0.2017
C.(X.Diatoms,Jun)            -0.0433
C.(X.Greens,Jun)              1.0091
C.(X.Unicells,Jun)           -0.2287
C.(X.Other.algae,Jun)        -0.1981
C.(X.Cryptomonas,Jul)        -0.7215
C.(X.Diatoms,Jul)            -0.2118
C.(X.Greens,Jul)              0.6025
C.(X.Unicells,Jul)            0.2395
C.(X.Other.algae,Jul)         0.5061
C.(X.Cryptomonas,Aug)        -0.2006
C.(X.Diatoms,Aug)            -0.7375
C.(X.Greens,Aug)              0.1143
C.(X.Unicells,Aug)            0.1380
C.(X.Other.algae,Aug)         0.6597
C.(X.Cryptomonas,Sep)        -0.1057
C.(X.Diatoms,Sep)            -0.9919
C.(X.Greens,Sep)             -0.1624
C.(X.Unicells,Sep)           -0.0550
C.(X.Other.algae,Sep)         0.5215
C.(X.Cryptomonas,Oct)         0.2331
C.(X.Diatoms,Oct)            -0.1915
C.(X.Greens,Oct)              0.0885
C.(X.Unicells,Oct)           -0.2098
C.(X.Other.algae,Oct)         0.7259
```

```
C.(X.Cryptomonas,Nov)              0.1383
C.(X.Diatoms,Nov)                  0.0206
C.(X.Greens,Nov)                   0.0318
C.(X.Unicells,Nov)                -0.2794
C.(X.Other.algae,Nov)              0.4489
C.(X.Cryptomonas,Dec)             -0.5750
C.(X.Diatoms,Dec)                 -0.2049
C.(X.Greens,Dec)                  -0.7270
C.(X.Unicells,Dec)                -0.4392
C.(X.Other.algae,Dec)             -0.6936
Initial states (x0) defined at t=0

Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.
```

Get the appropriate seasonal effects:

```
seas.a = coef(q1a, type = "matrix")$C
rownames(seas.a) = phytos
colnames(seas.a) = the.months
```
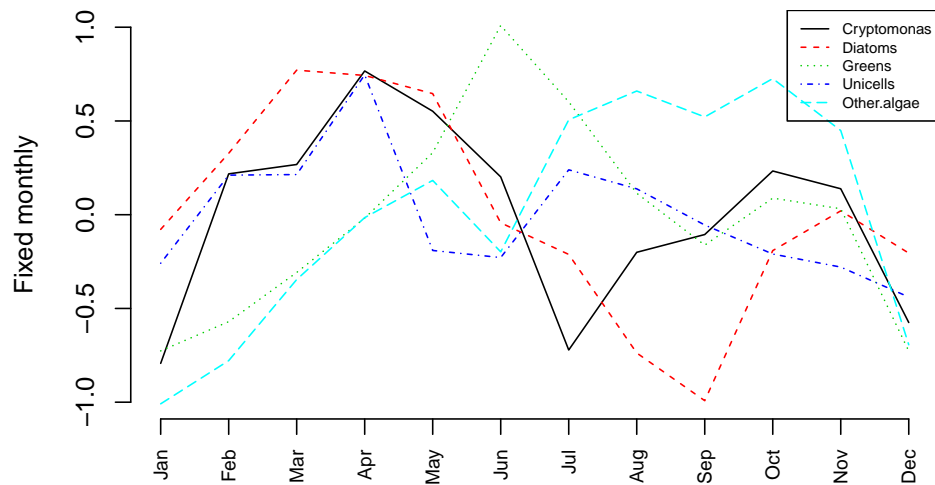
We could plot the seasonal effect by species with this code:

```
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 $\mathbf{C}$ 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)
```

Warning! Abstol convergence only. Maxit (=500) reached before log-log convergence.

```
MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
WARNING: Abstol convergence only no log-log convergence.
 maxit (=500) reached before log-log convergence.
 The likelihood and params might not be at the ML values.
 Try setting control$maxit higher.
Log-likelihood: -691.2546
AIC: 1454.509   AICc: 1459.266
```

	Estimate
R.diag	0.061143
B.(X.Cryptomonas,X.Cryptomonas)	0.244085
B.(X.Diatoms,X.Diatoms)	0.365194
B.(X.Greens,X.Greens)	0.318704
B.(X.Unicells,X.Unicells)	0.589248
B.(X.Other.algae,X.Other.algae)	0.155030
Q.(X.Cryptomonas,X.Cryptomonas)	0.736705
Q.(X.Diatoms,X.Diatoms)	0.346130
Q.(X.Greens,X.Greens)	0.573557
Q.(X.Unicells,X.Unicells)	0.518047
Q.(X.Other.algae,X.Other.algae)	0.519240
x0.X.Cryptomonas	-0.933557
x0.X.Diatoms	-1.577440
x0.X.Greens	3.352402
x0.X.Unicells	1.088766
x0.X.Other.algae	-2.387367
C.(X.Cryptomonas,1)	-1.398504
C.(X.Diatoms,1)	-1.219105
C.(X.Greens,1)	-1.465674
C.(X.Unicells,1)	-0.422665

```
C.(X.Other.algae,1)            -0.961017
C.(X.Cryptomonas,2)             0.970363
C.(X.Diatoms,2)                 1.296856
C.(X.Greens,2)                  0.587269
C.(X.Unicells,2)                0.338610
C.(X.Other.algae,2)            -0.044937
C.(X.Cryptomonas,3)            -0.157394
C.(X.Diatoms,3)                -0.254391
C.(X.Greens,3)                 -0.047262
C.(X.Unicells,3)               -0.050380
C.(X.Other.algae,3)             0.078025
C.(X.Cryptomonas,4)             0.007135
C.(X.Diatoms,4)                 0.012965
C.(X.Greens,4)                  0.000351
C.(X.Unicells,4)                0.001860
C.(X.Other.algae,4)            -0.005862
Initial states (x0) defined at t=0

Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.


Convergence warnings
 Warning: the  R.diag  parameter value has not converged.
 Type MARSSinfo("convergence") for more info on this warning.
```

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):

```r
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

```r
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.

```r
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)
```

```
Warning! Abstol convergence only. Maxit (=500) reached before log-log convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
WARNING: Abstol convergence only no log-log convergence.
```

```
 maxit (=500) reached before log-log convergence.
 The likelihood and params might not be at the ML values.
 Try setting control$maxit higher.
Log-likelihood: -698.8021
AIC: 1449.604    AICc: 1452.067


                                         Estimate
R.diag                                     0.0502
B.(X.Cryptomonas,X.Cryptomonas)            0.2940
B.(X.Diatoms,X.Diatoms)                    0.3974
B.(X.Greens,X.Greens)                      0.2777
B.(X.Unicells,X.Unicells)                  0.5876
B.(X.Other.algae,X.Other.algae)            0.1833
Q.(X.Cryptomonas,X.Cryptomonas)            0.7785
Q.(X.Diatoms,X.Diatoms)                    0.3569
Q.(X.Greens,X.Greens)                      0.5815
Q.(X.Unicells,X.Unicells)                  0.5444
Q.(X.Other.algae,X.Other.algae)            0.5759
x0.X.Cryptomonas                          -2.5409
x0.X.Diatoms                              -2.4183
x0.X.Greens                                2.7543
x0.X.Unicells                              1.0305
x0.X.Other.algae                          -3.9474
C.(X.Cryptomonas,cos.t)                   -0.2059
C.(X.Diatoms,cos.t)                       -0.1414
C.(X.Greens,cos.t)                        -0.5800
C.(X.Unicells,cos.t)                      -0.2153
C.(X.Other.algae,cos.t)                   -0.3124
C.(X.Cryptomonas,sin.t)                    0.2351
C.(X.Diatoms,sin.t)                        0.6836
C.(X.Greens,sin.t)                        -0.2524
C.(X.Unicells,sin.t)                       0.1505
C.(X.Other.algae,sin.t)                   -0.6188
Initial states (x0) defined at t=0


Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.


Convergence warnings
 Warning: the  R.diag  parameter value has not converged.
 Type MARSSinfo("convergence") for more info on this warning.
```

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
```

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
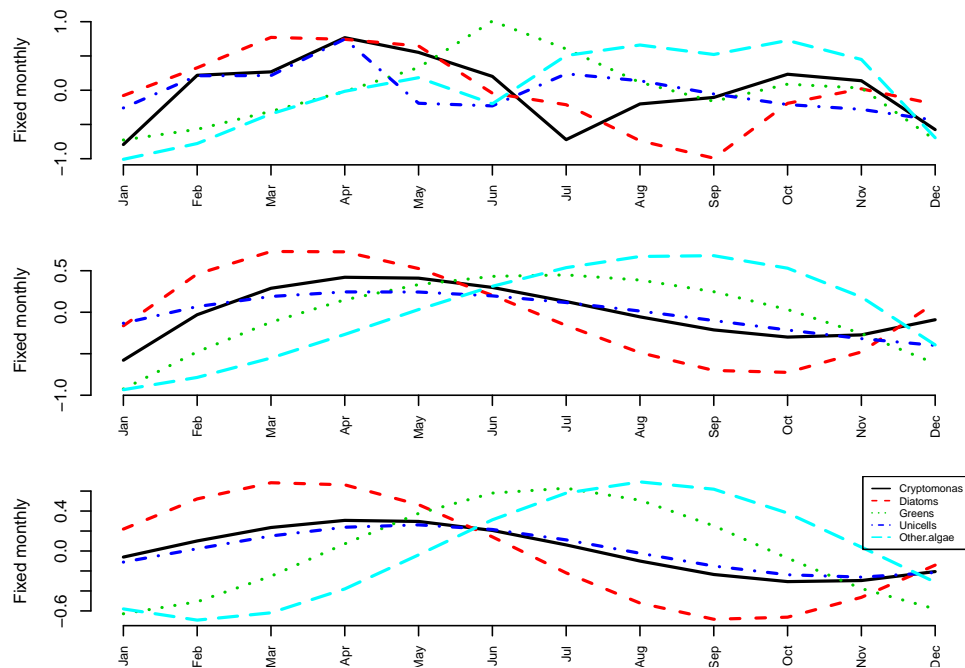
Figure 1: Our model had the season effect in the process part of the model. That means this graph is showing when growth rates, not abundance, is highest.

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 2). 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)
```

```
Success! abstol and log-log tests passed at 198 iterations.
Alert: conv.test.slope.tol is 0.5.
Test with smaller values (<0.1) to ensure convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
Estimation converged in 198 iterations.
Log-likelihood: -712.7631
AIC: 1477.526    AICc: 1479.989


                                 Estimate
R.diag                             0.1074
B.(X.Cryptomonas,X.Cryptomonas)    0.3367
```

8

```
B.(X.Diatoms,X.Diatoms)            0.6430
B.(X.Greens,X.Greens)              0.2871
B.(X.Unicells,X.Unicells)          0.6487
B.(X.Other.algae,X.Other.algae)    0.2224
Q.(X.Cryptomonas,X.Cryptomonas)    0.7409
Q.(X.Diatoms,X.Diatoms)            0.2964
Q.(X.Greens,X.Greens)              0.5782
Q.(X.Unicells,X.Unicells)          0.5031
Q.(X.Other.algae,X.Other.algae)    0.5187
x0.X.Cryptomonas                  -2.3631
x0.X.Diatoms                      -1.6727
x0.X.Greens                        2.5002
x0.X.Unicells                      0.8301
x0.X.Other.algae                  -3.2930
C.(X.Cryptomonas,Temp)            -0.2489
C.(X.Diatoms,Temp)                -0.5425
C.(X.Greens,Temp)                  0.1097
C.(X.Unicells,Temp)               -0.1152
C.(X.Other.algae,Temp)             0.4555
C.(X.Cryptomonas,TP)              -0.2215
C.(X.Diatoms,TP)                  -0.2057
C.(X.Greens,TP)                   -0.2926
C.(X.Unicells,TP)                 -0.0738
C.(X.Other.algae,TP)              -0.0174
Initial states (x0) defined at t=0

Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.
```

```r
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)")
```

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.

```r
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:

```r
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
```
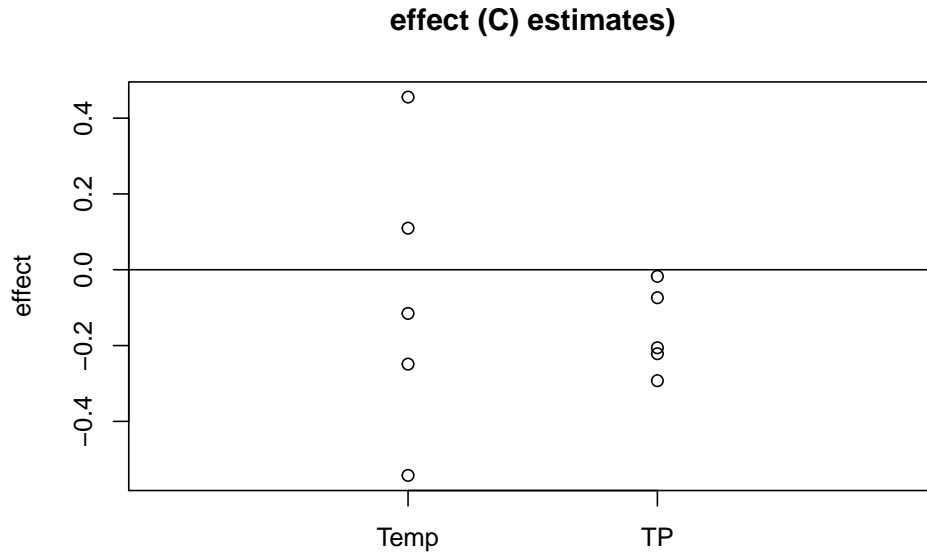
effect (C) estimates)

Figure 2: Estimated effect sizes for all taxon.

```
[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)
```

```
Success! abstol and log-log tests passed at 176 iterations.
Alert: conv.test.slope.tol is 0.5.
Test with smaller values (<0.1) to ensure convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
Estimation converged in 176 iterations.
Log-likelihood: -718.6454
AIC: 1479.291    AICc: 1480.898


                                     Estimate
R.diag                                 0.1386
B.(X.Cryptomonas,X.Cryptomonas)        0.3908
B.(X.Diatoms,X.Diatoms)                0.7173
B.(X.Greens,X.Greens)                  0.2706
B.(X.Unicells,X.Unicells)              0.6825
B.(X.Other.algae,X.Other.algae)        0.2323
Q.(X.Cryptomonas,X.Cryptomonas)        0.7180
Q.(X.Diatoms,X.Diatoms)                0.2577
Q.(X.Greens,X.Greens)                  0.5835
Q.(X.Unicells,X.Unicells)              0.4660
Q.(X.Other.algae,X.Other.algae)        0.4855
x0.X.Cryptomonas                      -2.3169
```

```
x0.X.Diatoms                    -1.5995
x0.X.Greens                      2.2047
x0.X.Unicells                    0.7925
x0.X.Other.algae                -3.2470
C.Temp1                         -0.0778
C.Temp2                         -0.3772
C.Temp3                          0.3493
C.Temp4                         -0.0631
C.Temp5                          0.4649
Initial states (x0) defined at t=0
```

```
Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.
```

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)
```

```
Success! abstol and log-log tests passed at 249 iterations.
Alert: conv.test.slope.tol is 0.5.
Test with smaller values (<0.1) to ensure convergence.
```

```
MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
Estimation converged in 249 iterations.
Log-likelihood: -734.5589
AIC: 1511.118    AICc: 1512.725
```

```
                                   Estimate
R.diag                               0.0917
B.(X.Cryptomonas,X.Cryptomonas)      0.3545
B.(X.Diatoms,X.Diatoms)              0.7334
B.(X.Greens,X.Greens)                0.3193
B.(X.Unicells,X.Unicells)            0.6505
B.(X.Other.algae,X.Other.algae)      0.3356
Q.(X.Cryptomonas,X.Cryptomonas)      0.7793
Q.(X.Diatoms,X.Diatoms)              0.4221
Q.(X.Greens,X.Greens)                0.5942
Q.(X.Unicells,X.Unicells)            0.5267
Q.(X.Other.algae,X.Other.algae)      0.5910
x0.X.Cryptomonas                    -2.2095
x0.X.Diatoms                        -1.4527
x0.X.Greens                          2.1540
```

```
x0.X.Unicells                              0.8032
x0.X.Other.algae                          -2.4403
C.TP1                                     -0.0144
C.TP2                                      0.2629
C.TP3                                     -0.3691
C.TP4                                      0.0227
C.TP5                                     -0.3520
Initial states (x0) defined at t=0


Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.
```

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)
```

```
Success! abstol and log-log tests passed at 198 iterations.
Alert: conv.test.slope.tol is 0.5.
Test with smaller values (<0.1) to ensure convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
Estimation converged in 198 iterations.
Log-likelihood: -712.7631
AIC: 1477.526    AICc: 1479.989


                                         Estimate
R.diag                                     0.1074
B.(X.Cryptomonas,X.Cryptomonas)            0.3367
B.(X.Diatoms,X.Diatoms)                    0.6430
B.(X.Greens,X.Greens)                      0.2871
B.(X.Unicells,X.Unicells)                  0.6487
B.(X.Other.algae,X.Other.algae)            0.2224
Q.(X.Cryptomonas,X.Cryptomonas)            0.7409
Q.(X.Diatoms,X.Diatoms)                    0.2964
Q.(X.Greens,X.Greens)                      0.5782
Q.(X.Unicells,X.Unicells)                  0.5031
Q.(X.Other.algae,X.Other.algae)            0.5187
x0.X.Cryptomonas                          -2.3631
x0.X.Diatoms                              -1.6727
x0.X.Greens                                2.5002
x0.X.Unicells                              0.8301
x0.X.Other.algae                          -3.2930
C.(X.Cryptomonas,Temp)                    -0.2489
C.(X.Diatoms,Temp)                        -0.5425
C.(X.Greens,Temp)                          0.1097
C.(X.Unicells,Temp)                       -0.1152
C.(X.Other.algae,Temp)                     0.4555
C.(X.Cryptomonas,TP)                      -0.2215
C.(X.Diatoms,TP)                          -0.2057
C.(X.Greens,TP)                           -0.2926
C.(X.Unicells,TP)                         -0.0738
```

```
C.(X.Other.algae,TP)              -0.0174
Initial states (x0) defined at t=0
```

```
Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.
```

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))
```

```
  Model   AICc
1  Temp 1480.9
2    TP 1512.7
3  Both 1480.0
```

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)
df <- data.frame(Model = mod.names, delta.AICc = delta.AICc)[order(delta.AICc),
    ]
kable(df, row.names = FALSE)
```

Model	delta.AICc
Fourier	0.0
Cubic	9.7
Fixed	28.4
Both	30.4
Temp	31.3
TP	63.1

Diagnostics Temperature and TP model residuals for all taxa except Greens appear to show significant negative autocorrelation at lag=1 (Figure 3 and Figure 4), suggesting that both models are inadequate to capture all of the systematic variation in phytoplankton abundance.

```
par(mfrow = c(5, 2), mai = c(0.25, 0.6, 0.2, 0.2), omi = c(0,
    0, 0, 0))
for (i in 1:5) {
    plot.ts(residuals(q2.Temp)$model.residuals[i, ], ylab = "Residual",
        main = phytos[i])
    abline(h = 0, lty = "dashed", xlab = "")
    acf(residuals(q2.Temp)$model.residuals[i, ], xlab = "")
}
```
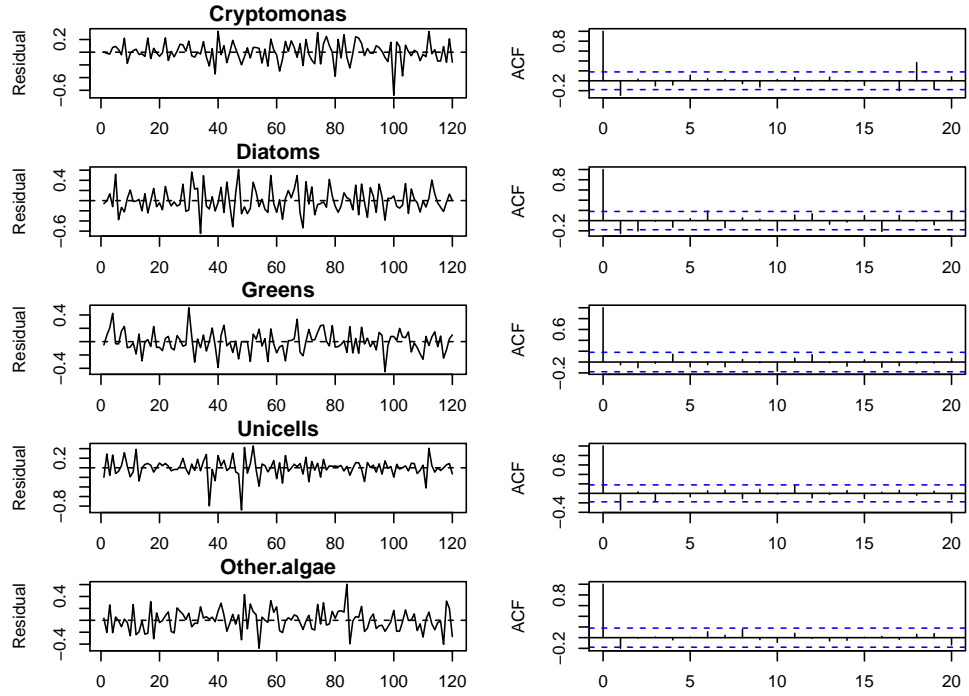
Figure 3: Plot the temperature (q2.Temp) model residuals and ACF's for all 5 taxa.

```r
par(mfrow = c(5, 2), mai = c(0.25, 0.6, 0.2, 0.2), omi = c(0,
    0, 0, 0))
for (i in 1:5) {
    plot.ts(residuals(q2.TP)$model.residuals[i, ], ylab = "Residual",
        main = phytos[i], xlab = "")
    abline(h = 0, lty = "dashed")
    acf(residuals(q2.TP)$model.residuals[i, ], xlab = "")
}
```
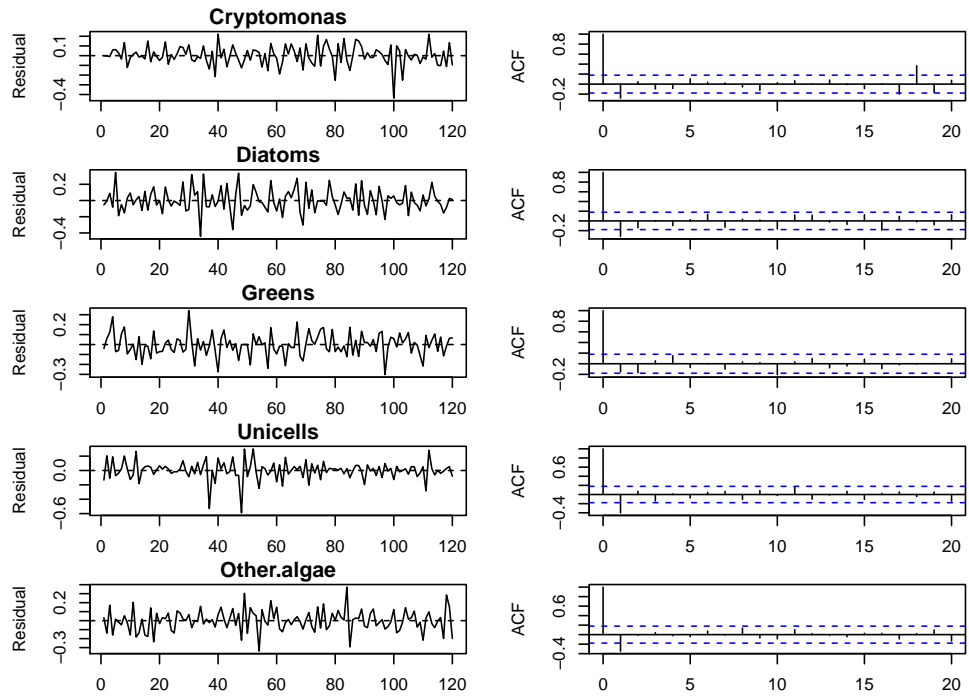
14

Figure 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 **D** and **d** instead of **C** and **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)
```

Warning! Abstol convergence only. Maxit (=500) reached before log-log convergence.

```
MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
WARNING: Abstol convergence only no log-log convergence.
 maxit (=500) reached before log-log convergence.
 The likelihood and params might not be at the ML values.
 Try setting control$maxit higher.
Log-likelihood: -728.5586
AIC: 1499.117    AICc: 1500.724
```

	Estimate
R.diag	0.0276
B.(X.Cryptomonas,X.Cryptomonas)	0.3378
B.(X.Diatoms,X.Diatoms)	0.5965
B.(X.Greens,X.Greens)	0.3415
B.(X.Unicells,X.Unicells)	0.6043
B.(X.Other.algae,X.Other.algae)	0.1790
Q.(X.Cryptomonas,X.Cryptomonas)	0.8485
Q.(X.Diatoms,X.Diatoms)	0.5240
Q.(X.Greens,X.Greens)	0.6475
Q.(X.Unicells,X.Unicells)	0.6067
Q.(X.Other.algae,X.Other.algae)	0.6128
x0.X.Cryptomonas	-2.5703
x0.X.Diatoms	-2.0283
x0.X.Greens	2.2103
x0.X.Unicells	0.9552
x0.X.Other.algae	-3.3616
D.Cryptomonas	-0.0505
D.Diatoms	-0.3893
D.Greens	0.5072
D.Unicells	0.0923
D.Other.algae	0.5646

```
Initial states (x0) defined at t=0

Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.

Convergence warnings
 Warning: the  R.diag  parameter value has not converged.
 Type MARSSinfo("convergence") for more info on this warning.
```

Now use AIC to compare the 2 models.

```r
df <- data.frame(Model = c("proc", "obs"), AICc = round(c(q2.Temp$AICc,
    q3.Tobs$AICc), 1))
kable(df, row.names = FALSE)
```

Model	AICc
proc	1480.9
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.

```r
# 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)
```

```
Warning! Abstol convergence only. Maxit (=500) reached before log-log convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
WARNING: Abstol convergence only no log-log convergence.
 maxit (=500) reached before log-log convergence.
 The likelihood and params might not be at the ML values.
```

```
 Try setting control$maxit higher.
Log-likelihood: -758.4122
AIC: 1550.824    AICc: 1551.881


                                      Estimate
R.diag                                 0.02395
B.(X.Cryptomonas,X.Cryptomonas)        0.33549
B.(X.Diatoms,X.Diatoms)                0.63113
B.(X.Greens,X.Greens)                  0.46323
B.(X.Unicells,X.Unicells)              0.60355
B.(X.Other.algae,X.Other.algae)        0.41653
Q.(X.Cryptomonas,X.Cryptomonas)        0.85369
Q.(X.Diatoms,X.Diatoms)                0.57679
Q.(X.Greens,X.Greens)                  0.77365
Q.(X.Unicells,X.Unicells)              0.61444
Q.(X.Other.algae,X.Other.algae)        0.79059
x0.X.Cryptomonas                      -2.40741
x0.X.Diatoms                          -1.15107
x0.X.Greens                            0.25745
x0.X.Unicells                          0.75183
x0.X.Other.algae                      -3.14216
C.1                                   -0.00193
Initial states (x0) defined at t=0


Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.


Convergence warnings
 Warning: the  R.diag  parameter value has not converged.
 Type MARSSinfo("convergence") for more info on this warning.
```

Use AIC to compare the 2 Temp models

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

```
           Effect    AICc
1 taxon-specific 1480.9
2          equal 1551.9
```

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.

```r
# 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)
```

```
Warning! Abstol convergence only. Maxit (=500) reached before log-log convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
```

```
WARNING: Abstol convergence only no log-log convergence.
 maxit (=500) reached before log-log convergence.
 The likelihood and params might not be at the ML values.
 Try setting control$maxit higher.
Log-likelihood: -756.5644
AIC: 1547.129   AICc: 1548.186


                                  Estimate
R.diag                             0.0200
B.(X.Cryptomonas,X.Cryptomonas)    0.3138
B.(X.Diatoms,X.Diatoms)            0.6121
B.(X.Greens,X.Greens)              0.4291
B.(X.Unicells,X.Unicells)          0.5728
B.(X.Other.algae,X.Other.algae)    0.3899
Q.(X.Cryptomonas,X.Cryptomonas)    0.8613
Q.(X.Diatoms,X.Diatoms)            0.6241
Q.(X.Greens,X.Greens)              0.7360
Q.(X.Unicells,X.Unicells)          0.6274
Q.(X.Other.algae,X.Other.algae)    0.7506
x0.X.Cryptomonas                  -2.2167
x0.X.Diatoms                      -1.0087
x0.X.Greens                        0.5386
x0.X.Unicells                      0.9780
x0.X.Other.algae                  -3.0629
C.1                               -0.0758
Initial states (x0) defined at t=0


Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.


Convergence warnings
 Warning: the  R.diag  parameter value has not converged.
 Type MARSSinfo("convergence") for more info on this warning.
```

Now use AIC to compare the 2 TP models

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

```
           Effect   AICc
1 taxon-specific 1512.7
2          equal 1548.2
```

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).

```r
# 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)
```

Success! abstol and log-log tests passed at 198 iterations.
Alert: conv.test.slope.tol is 0.5.
Test with smaller values (<0.1) to ensure convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
Estimation converged in 198 iterations.
Log-likelihood: -712.7631
AIC: 1477.526   AICc: 1479.989

	Estimate
R.diag	0.1074
B.(X.Cryptomonas,X.Cryptomonas)	0.3367
B.(X.Diatoms,X.Diatoms)	0.6430
B.(X.Greens,X.Greens)	0.2871
B.(X.Unicells,X.Unicells)	0.6487
B.(X.Other.algae,X.Other.algae)	0.2224
Q.(X.Cryptomonas,X.Cryptomonas)	0.7409
Q.(X.Diatoms,X.Diatoms)	0.2964
Q.(X.Greens,X.Greens)	0.5782
Q.(X.Unicells,X.Unicells)	0.5031
Q.(X.Other.algae,X.Other.algae)	0.5187
x0.X.Cryptomonas	-2.3631
x0.X.Diatoms	-1.6727
x0.X.Greens	2.5002
x0.X.Unicells	0.8301
x0.X.Other.algae	-3.2930
C.(X.Cryptomonas,Temp)	-0.2489
C.(X.Diatoms,Temp)	-0.5425
C.(X.Greens,Temp)	0.1097
C.(X.Unicells,Temp)	-0.1152
C.(X.Other.algae,Temp)	0.4555
C.(X.Cryptomonas,TP)	-0.2215
C.(X.Diatoms,TP)	-0.2057
C.(X.Greens,TP)	-0.2926
C.(X.Unicells,TP)	-0.0738
C.(X.Other.algae,TP)	-0.0174

Initial states (x0) defined at t=0

Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.

```
# 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)
```

```
Success! abstol and log-log tests passed at 219 iterations.
Alert: conv.test.slope.tol is 0.5.
Test with smaller values (<0.1) to ensure convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
Estimation converged in 219 iterations.
Log-likelihood: -732.5287
AIC: 1509.057   AICc: 1510.82


                                        Estimate
R.diag                                    0.1449
B.(X.Cryptomonas,X.Cryptomonas)           0.3634
B.(X.Diatoms,X.Diatoms)                   0.7502
B.(X.Greens,X.Greens)                     0.3762
B.(X.Unicells,X.Unicells)                 0.6672
B.(X.Other.algae,X.Other.algae)           0.3999
Q.(X.Cryptomonas,X.Cryptomonas)           0.7032
Q.(X.Diatoms,X.Diatoms)                   0.3011
Q.(X.Greens,X.Greens)                     0.5516
Q.(X.Unicells,X.Unicells)                 0.4603
Q.(X.Other.algae,X.Other.algae)           0.5629
x0.X.Cryptomonas                         -2.1869
x0.X.Diatoms                             -1.3909
x0.X.Greens                               1.8029
x0.X.Unicells                             0.8766
x0.X.Other.algae                         -2.1783
C.Temp                                   -0.1350
C.TP 1                                   -0.1229
C.TP 2                                    0.1613
C.TP 3                                   -0.4573
C.TP 4                                   -0.0862
C.TP 5                                   -0.4469
Initial states (x0) defined at t=0

Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.
```

```r
# 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)
```

```
Success! abstol and log-log tests passed at 182 iterations.
Alert: conv.test.slope.tol is 0.5.
Test with smaller values (<0.1) to ensure convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
Estimation converged in 182 iterations.
```

```
Log-likelihood: -714.4338
AIC: 1472.868    AICc: 1474.631


                                     Estimate
R.diag                                  0.116
B.(X.Cryptomonas,X.Cryptomonas)         0.352
B.(X.Diatoms,X.Diatoms)                 0.661
B.(X.Greens,X.Greens)                   0.281
B.(X.Unicells,X.Unicells)               0.633
B.(X.Other.algae,X.Other.algae)         0.238
Q.(X.Cryptomonas,X.Cryptomonas)         0.731
Q.(X.Diatoms,X.Diatoms)                 0.284
Q.(X.Greens,X.Greens)                   0.578
Q.(X.Unicells,X.Unicells)               0.499
Q.(X.Other.algae,X.Other.algae)         0.515
x0.X.Cryptomonas                       -2.348
x0.X.Diatoms                           -1.651
x0.X.Greens                             2.348
x0.X.Unicells                           0.936
x0.X.Other.algae                       -2.853
C.Temp 1                               -0.201
C.Temp 2                               -0.505
C.Temp 3                                0.218
C.Temp 4                               -0.180
C.Temp 5                                0.335
C.TP                                   -0.159
Initial states (x0) defined at t=0


Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.
```

```r
# 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)
```

```
Warning! Abstol convergence only. Maxit (=500) reached before log-log convergence.

MARSS fit is
Estimation method: kem
Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
WARNING: Abstol convergence only no log-log convergence.
 maxit (=500) reached before log-log convergence.
 The likelihood and params might not be at the ML values.
 Try setting control$maxit higher.
Log-likelihood: -754.1381
AIC: 1544.276    AICc: 1545.46


                                     Estimate
R.diag                                 0.0247
B.(X.Cryptomonas,X.Cryptomonas)        0.3067
B.(X.Diatoms,X.Diatoms)                0.5926
```

```
B.(X.Greens,X.Greens)             0.4713
B.(X.Unicells,X.Unicells)         0.5693
B.(X.Other.algae,X.Other.algae)   0.4251
Q.(X.Cryptomonas,X.Cryptomonas)   0.8366
Q.(X.Diatoms,X.Diatoms)           0.5598
Q.(X.Greens,X.Greens)             0.7464
Q.(X.Unicells,X.Unicells)         0.6145
Q.(X.Other.algae,X.Other.algae)   0.7965
x0.X.Cryptomonas                 -2.3718
x0.X.Diatoms                     -1.0903
x0.X.Greens                       0.4251
x0.X.Unicells                     0.9384
x0.X.Other.algae                 -2.8902
C.Temp                           -0.1469
C.TP                             -0.1799
Initial states (x0) defined at t=0


Standard errors have not been calculated.
Use MARSSparamCIs to compute CIs and bias estimates.


Convergence warnings
 Warning: the  R.diag  parameter value has not converged.
 Type MARSSinfo("convergence") for more info on this warning.
```

```r
df <- data.frame(Effect = c("unconstrained", "TP constrained",
    "Temp constrained", "both constrained"), AICc = round(c(q4.unc$AICc,
    q4.Tempunc$AICc, q4.TPunc$AICc, q4.cons$AICc), 1))
kable(df, row.names = FALSE)
```

Effect	AICc
unconstrained	1480.0
TP constrained	1474.6
Temp constrained	1510.8
both constrained	1545.5

Problem 5

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

This question is open-ended. We were looking for you to try to answer questions 2-4 using `lm()`. There are many approaches you could take.

Observation model is vague in the context of this question. What we had in mind was using the following model:

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

if you are investigating a month effect or

$$y(t) = \beta_{temp}T + \beta_{tp}TP + w(t) \tag{2}$$

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

A *process model* would try to explain the error between y_t and βy_{t-1}.

$$y(t) = \beta y(t-1) + \alpha_{month} + w(t) \tag{3}$$

if you are investigating a month effect or

$$y(t) = \beta y(t-1) + \beta_{temp}T + \beta_{tp}TP + w(t) \tag{4}$$

Data set-up

`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.

```
kable(dat.frame[118:125, 1:5], row.names = FALSE)
```

abund	abund.lag1	taxon	month	temp
-0.8623450	-0.2102360	Cryptomonas	Oct	0.8466326
0.7092607	-0.8623450	Cryptomonas	Nov	0.0015140
-0.7070257	0.7092607	Cryptomonas	Dec	-0.9059344
-0.7413776	NA	Diatoms	Jan	-1.2330640
0.2085792	-0.7413776	Diatoms	Feb	-1.4642953
1.0206142	0.2085792	Diatoms	Mar	-1.3406134
0.8039275	1.0206142	Diatoms	Apr	-0.8464835
1.5061781	0.8039275	Diatoms	May	-0.1651877

Question 1

Question 1 was not part of Problem 5, but let's address it. Q1 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, 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)
```

```
Loading required package: 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 5 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)
```
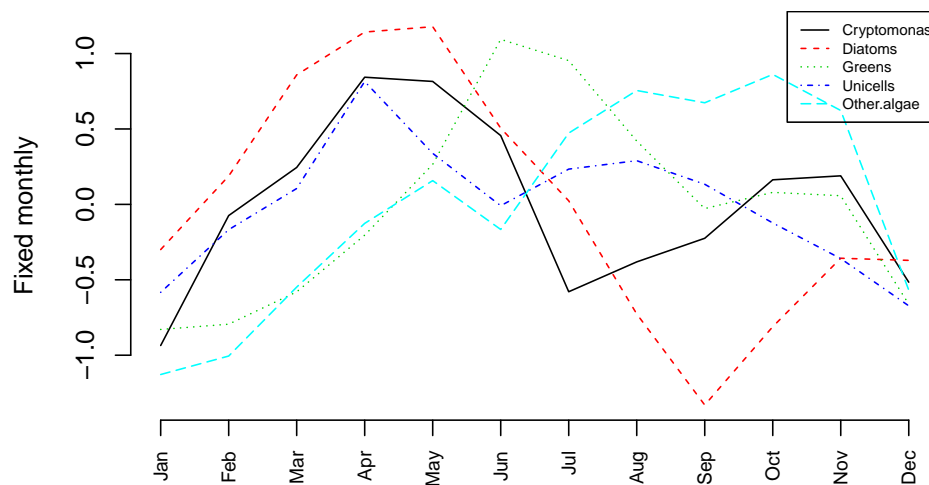

The results are quite similar to the state-space model.

```
par(mfrow = c(2, 1), mar = c(2, 2, 2, 2))
matplot(t(seas.a), type = "l", bty = "n", xaxt = "n", ylab = "Fixed monthly",
    col = 1:5, ylim = c(-1.1, 1.1))
abline(h = 0)
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)
title("MARSS estimates")

matplot(t(seaslm), type = "l", bty = "n", xaxt = "n", ylab = "Fixed monthly",
    col = 1:5, ylim = c(-1.1, 1.1))
```
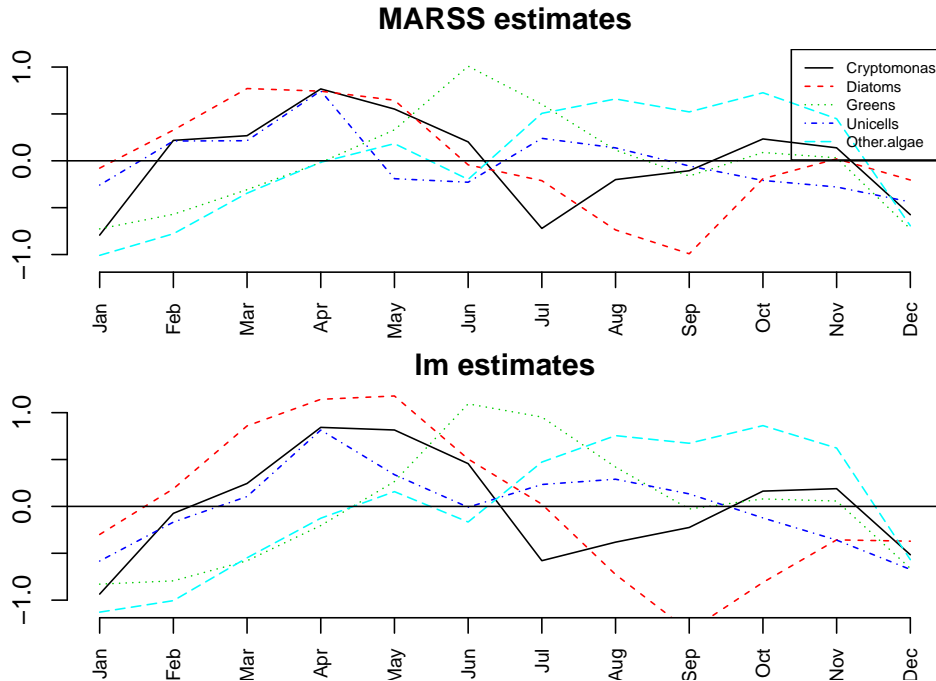
Figure 5: Comparison of the monthly population growth rate effect sizes.

```
abline(h = 0)
axis(1, labels = month.abb, at = 1:12, las = 2, cex.axis = 0.75)
title("lm estimates")
```

Question 2

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

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

Then fit to the differenced data and store the estimates. Figure 6 compares the estimates from the state-space model versus `lm()`. They are very similar.

```
fit = lm(abund ~ -1 + taxon:temp + taxon:tp, data = dat.frame)
require(stringr)
taxoncoef = str_detect(names(coef(fit)), ":temp")
Clm[, 1] = coef(fit)[taxoncoef]
taxoncoef = str_detect(names(coef(fit)), ":tp")
Clm[, 2] = coef(fit)[taxoncoef]
```

```
plot(c(rep(1, 5), rep(2, 5)), coef(q2.both)$C, xlab = "", ylab = "effect",
    xaxt = "n", xlim = c(0, 3), ylim = c(-1, 1))
points(c(rep(1, 5), rep(2, 5)) + 0.25, Clm, xlab = "", col = "red")
axis(1, at = c(1, 2), labels = c("Temp", "TP"))
abline(h = 0)
title("effect (C) estimates)")
```
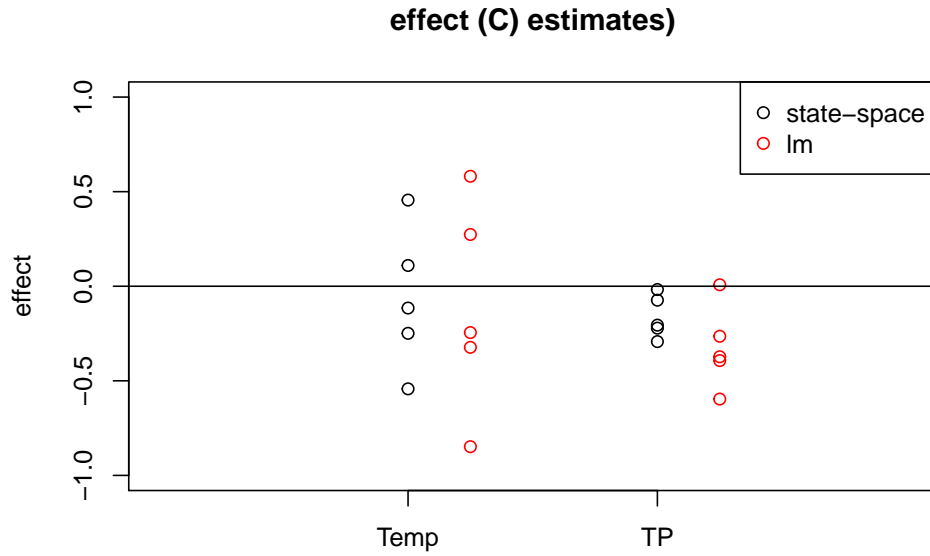
effect (C) estimates)

Figure 6: Estimated effect sizes for all taxon.

```
legend("topright", c("state-space", "lm"), pch = 1, col = c("black",
    "red"))
```

Another approach would be to compare AICs. We'll compare models with taxon specific effects of temperature and TP and non-specific effects.

```
fit = lm(abund ~ -1 + taxon:temp + taxon:tp, data = dat.frame)
fit.temp = lm(abund ~ -1 + taxon:temp, data = dat.frame)
fit.tp = lm(abund ~ -1 + taxon:tp, data = dat.frame)
fit.common = lm(abund ~ -1 + temp + tp, data = dat.frame)
fit.temp.common = lm(abund ~ -1 + temp, data = dat.frame)
fit.tp.common = lm(abund ~ -1 + tp, data = dat.frame)
df <- data.frame(model = c("taxon-specific both", "taxon-specific temp only",
    "taxon-specific TP only", "common both", "common temp only",
    "common TP only"), AIC = c(AIC(fit), AIC(fit.temp), AIC(fit.tp),
    AIC(fit.common), AIC(fit.temp.common), AIC(fit.tp.common)))
kable(df, row.names = FALSE)
```

model	AIC
taxon-specific both	1584.879
taxon-specific temp only	1610.999
taxon-specific TP only	1638.629
common both	1659.863
common temp only	1680.348
common TP only	1660.665

Question 3

We might address this by fitting a model with y_{t-1} as a covariate and comparing to the model without. We remove the NAs in the data frame so that we make sure to keep the response variable exactly the same between our fits. That allows us to compare AICs. The model with y_{t-1} as a covariate is much better.

```
dat.frame.no.NAs = na.omit(dat.frame)
fit.pros = lm(abund ~ -1 + taxon:abund.lag1 + taxon:temp + taxon:tp,
    data = dat.frame.no.NAs)
fit.obs = lm(abund ~ -1 + taxon:temp + taxon:tp, data = dat.frame.no.NAs)
c(process = AIC(fit.pros), obs = AIC(fit.obs))
```

```
 process      obs
1459.625 1569.131
```

Question 4

To answer this question, we can use AIC again. We'll use the model with y_{t-1} as a covariate since it is much better.

```
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)
df <- data.frame(Effect = c("unconstrained", "TP effect shared",
    "Temp effect shared", "both effects shared"), del.AIC.lm = round(AIC.lm -
    min(AIC.lm), 1), del.AICc.ss = round(AICc.ss - min(AICc.ss),
    1))
kable(df, row.names = FALSE)
```

Effect	del.AIC.lm	del.AICc.ss
unconstrained	4.7	5.4
TP effect shared	0.0	0.0
Temp effect shared	31.6	36.2
both effects shared	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.