The Presence of Hellbender Salamander Populations on the Environmental Composition of Streams Over Time

Christian Conroy

Introduction

A study was conducted in six streams in central Pennsylvania to assess stream factors that contribute to analyze whether environmental factors including mean current, dissolved oxygen, and conductivity levels change with time as well as with the status of the Hellbender population. The primary question of interest is concerned with whether changes in these longitudinally sampled factors are the same for streams with known Hellbender populations and those without. This will provide us insight into whether the presence of salamanders in the stream changes the composition of the stream in a way that may or may not merit intervention. At the same time, understanding how the three factors change over time in streams with and without known hellbender salamander populations will also give us clues into the environmental factors that underlay hellbender salamander site occupancy.

The remainder of this analysis report is organized as follows: In section two we discuss the exploratory analysis used in building separate independence models for the three longitudinally sampled stream factors. Section three describes the models used in the analysis and the results of those models. Lastly, section four compares the results between the two groups and draws conclusions related to our primary question of interest.

Methods

The analysis is based on balanced data comprised of measurements of longitudinally sampled variables including current, dissolved oxygen, and conductivity taken twice a month on six streams – four with known Hellbender salamander populations and two without – from June through October. The exploratory analyses for the independent models built for the mean response profiles on the three respective stream factors are included below.

With a balanced design on bimonthly measurements and only two groups, we should choose a saturated model for the mean with a total of 10 parameters for the response profiles for the two treatment groups that includes time and all other main effects in addition to two and higher way interactions. To model the outcome for each respective longitudinally sampled mean, we therefore use the independence model below:

$$\begin{aligned} Y_{ij} = {} & \beta_0 + \beta_1 G_{ij} + \beta_2 1(t_{ij} = 1) + \beta_3 1(t_{ij} = 2) + \beta_4 1(t_{ij} = 3) + \beta_5 1(t_{ij} = 4) + \beta_6 1(t_{ij} = 5) + \beta_7 1(t_{ij} = 6) \\ & + \beta_8 1(t_{ij} = 7) + \beta_9 1(t_{ij} = 8) + \beta_{10} 1(t_{ij} = 9) + \beta_{11} 1(t_{ij} = 10) + \beta_{12} G_{ij} * 1(t_{ij} = 1) \\ & + \beta_{13} G_{ij} * 1(t_{ij} = 2) + \beta_{14} G_{ij} * 1(t_{ij} = 3) + \beta_{15} G_{ij} * 1(t_{ij} = 4) + \beta_{16} G_{ij} * 1(t_{ij} = 5) + \beta_{17} G_{ij} \\ & * 1(t_{ij} = 6) + \beta_{18} G_{ij} * 1(t_{ij} = 7) + \beta_{19} G_{ij} * 1(t_{ij} = 8) + \beta_{20} G_{ij} * 1(t_{ij} = 9) + \beta_{21} G_{ij} \\ & * 1(t_{ij} = 10) + e \end{aligned}$$

where β_0 is our intercept, β_1 through β_{11} are the coefficients on each time interval with the associated X indicating known or unknown populations of Hellbender salamanders, and β_{12} through β_{21} represent the interaction between each time period and having known or unknown populations of Hellbender salamanders.

Current

As Figure 1 in the appendix shows, there is no clear pattern in how current changes over the measurement intervals. However, as shown by the response profiles of the two groups, it does look as though there may be similar patterns between the two stream groups such as relative flatness between baseline and July and consistent decline in August. That being said, given how small our sample size is, it cannot be ruled out that upward or downward spikes are the result of extreme values or outliers in a stream caused by high flooding, for example.

Given that there are far too many parameters to fit the model using an unstructured covariance structure – there are 55 – we can use the raw estimates of covariance and correlation to give us a sense of where to start in making assumptions about the covariance structure. While there is no consistent change in correlation as the distance between time measurement increases, the fact that there is such variability will likely lead us to reject a structure like compound symmetry that assumes correlations between time intervals will be constant regardless of how far apart they are.

Likelihood ratio tests (LRTs) to compare the homogeneous and heterogeneous versions of each respective covariance structure show that the homogeneous variance-covariance models performed the best regarding the mean response profile for current. While a strict interpretation of the LRT and a comparison of the AIC criterion would cause us to conclude that the homogeneous compound symmetric variance structure is preferable to both the homogeneous Toeplitz and AR(1) structures, given what we know about our longitudinally sampled data from the raw covariance and correlation matrices and how the compound symmetric structure assumes the same amount of correlation between measurement intervals regardless of the distance between those measurements, we will effectively ignore these results and instead focus on patterns like Toeplitz and AR(1) that better reflect the nature of the data. While the LRT between the homogeneous Toeplitz and AR(1) covariance patterns suggests that the AR1 structure is most appropriate for fitting the data, the homogeneous Toeplitz structure remains preferable given that AR(1) is much more restrictive of an assumption. Given that our simple size is only six streams and that the unstructured model could not even converge given the high amount of parameters in the model, the Toeplitz model serves as the most conservative option and provides smaller standard errors and p-values that allow us to make more valid inferences from our data. We therefore will use the homogeneous Toeplitz covariance structure to model the data.

Dissolved Oxygen

As evidenced by the plot showing the observed trajectories for each stream grouped by those with known and unknown populations of hellbender salamanders (Figure 2), there is no clear pattern in how dissolved oxygen levels change over the measurement intervals, though it does appear as if there is greater volatility between bimonthly periods as compared to the mean response profile for current. The response profile comparing the two groups leads us to believe it is likely that analysis will reveal the two groups to demonstrate both parallel and coincident changes in dissolved oxygen levels over time. Given the small sample size, we must exercise caution in interpreting the high volatility as outliers are likely to have a large impact on the plot trajectories.

Unlike the raw covariance provided for the current variable, our raw covariance matrix for dissolved oxygen contains several pairs with negative covariance, meaning that some pairs of underlying variables may be negatively correlated with each other. There also seems to be significant variability in the variance as indicated by the matrix diagonal, which may affect whether we accept models with constant variance.

LRTs to compare the homogeneous and heterogeneous versions of each respective covariance structure again showed that the homogeneous variance-covariance models performed the best in regarding the mean response profile for dissolved oxygen. While a strict interpretation of the LRT and a comparison of the AIC criterion would cause us to conclude that the homogeneous compound symmetric variance structure is preferable to both the homogeneous Toeplitz and AR(1) structures again here, given what we know about our longitudinally sampled data from the raw covariance and correlation matrices and how the compound symmetric structure assumes the same amount of correlation between measurement intervals regardless of the distance between those measurements, we will effectively ignore these results and instead focus on patterns like Toeplitz and AR(1) that better reflect the nature of the data. Given that the LRT concludes that the AR(1) covariance pattern is most appropriate and the fact that the standard errors do not differ to a large degree between the homogeneous Toeplitz and AR(1) models (in contrast to the mean response profile for current), we apply the AR(1) covariance pattern to the response profile for dissolved oxygen.

Conductivity

Based on the mean response trajectory for both groups show in Figure 3, we suspect that it is likely that the analysis will show parallelism, though the apparent outlier at the interval at month 10 may affect results. Given that the mean trajectory between the two demonstrates far less volatility (again with the exception of the interval at month 10 for the Known group), we suspect that a parametric linear model may need to be considered in this case.

There also appears to be a clearer pattern existing within the variance diagonal of the raw covariance structure. Namely, apart from a dip in variance between baseline and the first measurement period, the variance continuously rises from measurement period 6.5 through 9.5 before the outlier evident in the mean response profile trajectory decreases the variance drastically at measurement period 10. While this change in variance may lead us to believe a constant variance structure is not the most appropriate, there appears to be consistency in correlations regardless of the distance between measurements. Not only are the correlations consistent, but they appear to be high in most cases.

LRTs to compare the homogeneous and heterogeneous versions of each respective covariance structure again showed with the exception of the homogeneous compound symmetry pattern that the heterogeneous variance-covariance models performed the best in modeling the variance-covariance matrix regarding the mean response profile for dissolved oxygen. Interpreting the LRT p-values and comparing AIC criterion causes us to conclude that the heterogeneous Toeplitz variance structure is preferable to both the homogeneous Toeplitz and AR(1) structures here. Given that our simple size is only six streams and that the unstructured model could not even converge given the high amount of parameters in the model, the Toeplitz model serves as the most conservative option and provides smaller standard errors and p-values that allow us to make more valid inferences from our data, as made evident in the exploratory analysis for the conductivity mean response profile.

Results

The results of all three regressions are shown in Figure 4 in the appendix with empirically estimated standard errors included for each model in Figures 5-7. To ensure robustness with the models, we defend against potential model mis-specification by correcting the standard errors using the empirical or so-called robust variances. The empirical variance of $\hat{\beta}$ is estimated by:

$$\left[\sum_{i=1}^{n} X_i' \hat{\Sigma}^{-1} X_i\right]^{-1} \sum_{i=1}^{n} (X_i' \hat{\Sigma}^{-1} \hat{V} \hat{\Sigma}^{-1} X_i) \left[\sum_{i=1}^{n} (X_i' \hat{\Sigma}^{-1} X_i\right]^{-1}$$

Given that the empirical estimator of Cov($\hat{\beta}$) provides a consistent estimator of the variance, the standard errors in our model will be inflated compared to if we had simply used the results from the model specified with the respective covariance patterns selected to model the three response profiles.

Given the sequence of repeated measures on our two groups, there are three main hypotheses we pose regarding the mean response profiles. The test of the group * time interaction assesses whether changes in the mean for current are parallel between the two groups. A second question concerns the time effect, where the focus is on comparing the mean response at each occasion averaged over the groups to assess whether the respective trajectories for change in mean response are flat. Lastly, assuming that the population mean response profiles are parallel, we can look at the group effect to assess whether the profiles are at the same level in the sense that the profiles for the groups coincide. We can use the results of such analysis to inform our interpretation of the coefficients produced by our independence model.

<u>Current</u>

We can conclude from the p-values on the coefficients of the group, time, and group * time effect that the two groups do not demonstrate flat trajectories, but that they are parallel and coincide. Because the stream groups do not demonstrate a statistically significant difference in change over time, we can conclude that the presence of hellbender salamanders does not appear to be having any differential effect on the environmental composition of the streams.

Having established that the two mean response profiles demonstrate no statistically significant differences, we can use our independence model specified with time as a factor variable and using a homogeneous Toeplitz covariance structure to estimate the regression coefficients demonstrating the relationship between the time measurement intervals and the presence of Hellbender salamanders in our streams.

As the results in Figure 4 show, there are no statistically significant coefficients for either the main effects or the interaction. While our plot of the mean response profile trajectory appeared to show that specific time periods may have resulted in significant declines in current for our stream groups, the fact that there are no statistically significant coefficients for the time variables indicates that those spikes are not statistically significant and may have occurred largely due to extreme outlier values as opposed to any systemic reason.

<u>Dissolved Oxygen</u>

In conducting LRTs, we find statistically significant p-values demonstrating statistical difference between those models that include group, time, and group * time effects, and those that do not. The conclusion from the statistical difference and the associated AIC criterion show that the mean response profile for the two groups are not coincident, parallel, or flat, a slightly surprising result given the plot above.

Having established that the two mean response profiles demonstrate statistically significant differences, we can use our independence model specified with time as a factor variable and using a homogeneous Toeplitz covariance structure to estimate the regression coefficients demonstrating the relationship between the time measurement intervals and the presence of Hellbender salamanders in our streams.

As the results in Figure 4 show, there is only one statistically significant coefficient, which is the coefficient that demonstrates the relationship between time at measurement interval 9 and the presence of a known population of salamanders. This statistical significance is unsurprising as the plot showed a divergent spike between periods 8.5 and 9, with the trajectory increasing sharply for streams without known populations and decreasing slightly for those with known populations of Hellbender salamander. The fact that this is the only coefficient that is statistically significant based on our model may lead us to believe that the divergent spike was the result of an external shock such as intensified chemical or agricultural runoff and not evidence of any specific trend connected to the presence of the salamander population.

<u>Conductivity</u>

In conducting LRTs, we find statistically significant p-values between only those models that include time and those that do not. However, though without statistical significance demonstrating a difference in the models, we can conclude from the AIC criterion that the mean response profiles are roughly parallel and coincident.

We can use our independence model specified with time as a factor variable and using a heterogeneous Toeplitz covariance structure to estimate the regression coefficients demonstrating the relationship between the time measurement intervals and the presence of Hellbender salamanders in our streams.

As the results in Figure 4 show, there is only one statistically significant coefficient, which is the coefficient that demonstrates the correlation between the interval leading up to the last measurement period and conductivity This statistical significance is unsurprising as the plot showed a wide divergent spike between periods 10 and 10.5 for the group without a known population of hellbender salamanders The fact that this is the only coefficient that is statistically significant based on our model may again lead us to believe that the divergent spike was the result of an external shock and not evidence of any specific trend connected to the presence of the salamander population.

Discussion

While we may be tempted to use the analysis above to conclude that the presence of a hellbender salamander population has no apparent influence on environmental streams over time or that current conditions have little influence on hellbender salamander site occupancy, it is important to note that we have a very small sample size, meaning that any inferences here must be interpreted carefully. In all three analyses, it appeared as though the few statistically significant coefficients reflected extemporaneous spikes in one of the longitudinally sampled factors at one time point and not trends showing the differential impact of hellbender salamanders occupying a stream or not. It appears from the limited data that the presence of hellbender salamander may have no significant impact on the environmental composition of the stream. Similarly, the parallelism and coincidence of the mean response profiles for current and conductivity do little to shed light on to the stream occupancy habits of the hellbender salamander. While the analysis showed differing mean response profile trajectories in regards to dissolved oxygen for the two groups, a larger sample would be needed before making any environmental policy based on the observed relationship between the presence of a hellbender salamander population and the change in dissolved oxygen levels in the streams over time.

Appendix

Figure 1: Mean response profile trajectories for two groups – current

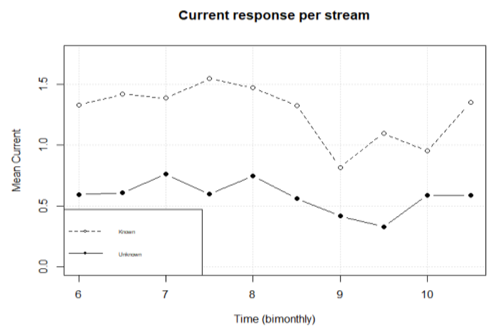

Figure 2: Mean response profile trajectories for two groups – dissolved oxygen

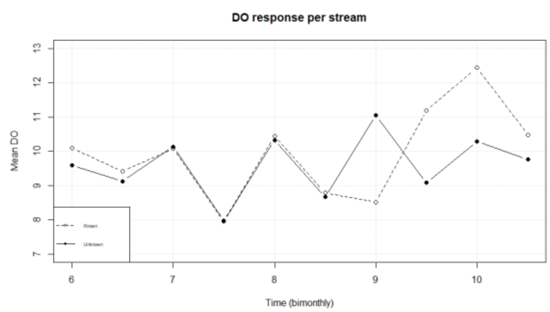

Figure 3: Mean response profile trajectories for two groups – dissolved oxygen

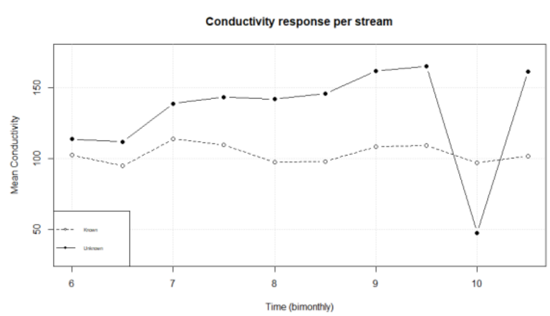

Figure 4: Mean response profile trajectories for two groups - conductivity

```
Figure 1: Analysis of Mean Response Profile Results
=====================================================================
Dependent variable:
------------------------------------
Current       DO      Conductivity
(1)          (2)          (3)
---------------------------------------------------------------------
factor(Known)1                     0.7345      0.4967      -11.2500
(0.7321)     (0.9417)     (35.0395)
t = 1.0032   t = 0.5274   t = -0.3211
p = 0.3158   p = 0.5979    p = 0.7482
factor(Month)6.5                   0.0115      -0.4660     -2.0000
(0.3722)     (1.0369)     (46.2630)
t = 0.0309   t = -0.4494 t = -0.0432
```

```
p = 0.9754  p = 0.6532   p = 0.9656
factor(Month)7                      0.1655      0.5419      25.0000
(0.2988)    (1.0829)    (18.1628)
t = 0.5540  t = 0.5005   t = 1.3764
p = 0.5797  p = 0.6168   p = 0.1687
factor(Month)7.5                    0.0020      -1.6242     29.5000
(0.3450)    (1.0869)    (49.1086)
t = 0.0058  t = -1.4943  t = 0.6007
p = 0.9954  p = 0.1351   p = 0.5481
factor(Month)8                      0.1515      0.7278      28.0000
(0.3702)    (1.0873)    (46.1023)
t = 0.4093  t = 0.6693   t = 0.6073
p = 0.6824  p = 0.5033   p = 0.5437
factor(Month)8.5                    -0.0335     -0.9315     32.0000
(0.3741)    (1.0873)    (41.4752)
t = -0.0895 t = -0.8567  t = 0.7715
p = 0.9287  p = 0.3917   p = 0.4404
factor(Month)9                      -0.1775     1.4591      48.0000
(0.3992)    (1.0873)    (125.3192)
t = -0.4446 t = 1.3419   t = 0.3830
p = 0.6566  p = 0.1797   p = 0.7018
factor(Month)9.5                    -0.2650     -0.5171     51.5000
(0.4102)    (1.0873)    (32.7743)
t = -0.6460 t = -0.4756  t = 1.5714
p = 0.5183  p = 0.6344   p = 0.1162
factor(Month)10                     -0.0070     0.6976      -66.5000
(0.4252)    (1.0873)    (58.4364)
t = -0.0165 t = 0.6415  t = -1.1380
p = 0.9869  p = 0.5212   p = 0.2552
factor(Month)10.5                   -0.0060     0.1696      47.5000*
(0.4391)    (1.0873)    (25.2222)
t = -0.0137 t = 0.1560   t = 1.8833
p = 0.9891  p = 0.8761   p = 0.0597
factor(Known)1:factor(Month)6.5     0.0752      -0.2174     -5.8000
(0.4558)    (1.2699)    (56.6604)
t = 0.1651  t = -0.1712 t = -0.1024
p = 0.8689  p = 0.8641   p = 0.9185
factor(Known)1:factor(Month)7       -0.1106     -0.5506     -13.7500
(0.3659)    (1.3262)    (22.2448)
t = -0.3021 t = -0.4151 t = -0.6181
p = 0.7626  p = 0.6781   p = 0.5365
factor(Known)1:factor(Month)7.5     0.2130      -0.5205     -22.2500
(0.4225)    (1.3312)    (60.1455)
t = 0.5041  t = -0.3910 t = -0.3699
p = 0.6142  p = 0.6958   p = 0.7115
factor(Known)1:factor(Month)8       -0.0093     -0.3692     -33.2500
(0.4533)    (1.3317)    (56.4636)
t = -0.0204 t = -0.2772 t = -0.5889
p = 0.9838  p = 0.7816   p = 0.5560
factor(Known)1:factor(Month)8.5     0.0285      -0.3777     -36.5000
(0.4582)    (1.3317)    (50.7965)
t = 0.0622  t = -0.2836 t = -0.7186
p = 0.9505  p = 0.7768   p = 0.4725
factor(Known)1:factor(Month)9       -0.3345    -3.0313**    -42.2500
(0.4889)    (1.3317)    (153.4840)
t = -0.6841 t = -2.2762 t = -0.2753
p = 0.4939  p = 0.0229   p = 0.7832
factor(Known)1:factor(Month)9.5     0.0315      1.6211      -44.7500
(0.5024)    (1.3317)    (40.1402)
t = 0.0627  t = 1.2173  t = -1.1148
p = 0.9501  p = 0.2236   p = 0.2650
factor(Known)1:factor(Month)10      -0.3708     1.6542      60.7500
(0.5207)    (1.3317)    (71.5697)
t = -0.7120 t = 1.2422   t = 0.8488
p = 0.4765  p = 0.2142   p = 0.3960
factor(Known)1:factor(Month)10.5    0.0252      0.2216      -48.2500
(0.5378)    (1.3317)    (30.8908)
t = 0.0470  t = 0.1664  t = -1.5620
p = 0.9626  p = 0.8679   p = 0.1183
Constant                            0.5965     9.5941***  114.0000***
(0.5978)    (0.7689)    (28.6096)
t = 0.9979  t = 12.4781  t = 3.9847
p = 0.3184  p = 0.0000   p = 0.0001
--------------------------------------------------------------------
Observations                          60          60          60
Log Likelihood                     -31.6030    -70.3559    -197.0264
```

```
Akaike Inf. Crit.                 111.2060    184.7117     460.0528
Bayesian Inf. Crit.               151.7391    221.8671     515.7858
=====================================================================
Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

Figure 2: Empirical Estimator – Standard Errors (Current)

```
=====================================================================
              (Intercept)               factor(Known)1
                0.3128948                    0.5275432
         factor(Month)6.5               factor(Month)7
                0.5052362                    0.5536796
         factor(Month)7.5               factor(Month)8
                0.4814626                    0.5855793
         factor(Month)8.5               factor(Month)9
                0.4969518                    0.4110567
         factor(Month)9.5              factor(Month)10
                0.3766421                    0.4782460
        factor(Month)10.5  factor(Known)1:factor(Month)6.5
                0.4771774                    0.7835907
   factor(Known)1:factor(Month)7  factor(Known)1:factor(Month)7.5
                0.7836101                    0.8176100
   factor(Known)1:factor(Month)8  factor(Known)1:factor(Month)8.5
                0.8080577                    0.7798258
   factor(Known)1:factor(Month)9  factor(Known)1:factor(Month)9.5
                0.6528185                    0.6518196
  factor(Known)1:factor(Month)10 factor(Known)1:factor(Month)10.5
                0.6913483                    0.7836577
```

Figure 3: Empirical Estimator – Standard Errors (Dissolved Oxygen)

```
=====================================================================
              (Intercept)               factor(Known)1
              0.009350639                  0.439577186
         factor(Month)6.5               factor(Month)7
              0.466265229                  0.672500932
         factor(Month)7.5               factor(Month)8
              0.830646420                  0.078515189
         factor(Month)8.5               factor(Month)9
              0.853509555                  0.536817917
         factor(Month)9.5              factor(Month)10
              0.698496919                  0.481849127
        factor(Month)10.5  factor(Known)1:factor(Month)6.5
              0.549557527                  1.044889779
   factor(Known)1:factor(Month)7  factor(Known)1:factor(Month)7.5
              0.959115129                  0.980341204
   factor(Known)1:factor(Month)8  factor(Known)1:factor(Month)8.5
              0.481565067                  0.968867013
   factor(Known)1:factor(Month)9  factor(Known)1:factor(Month)9.5
              0.848378781                  1.027578151
  factor(Known)1:factor(Month)10 factor(Known)1:factor(Month)10.5
              0.740659508                  0.767333615
```

Figure 4: Empirical Estimator – Standard Errors (Conductivity)

```
=====================================================================
              (Intercept)               factor(Known)1
                 43.84062                     44.49631
         factor(Month)6.5               factor(Month)7
                 61.50203                     72.12836
         factor(Month)7.5               factor(Month)8
                 74.10887                     82.59843
         factor(Month)8.5               factor(Month)9
                 83.80036                     87.44427
         factor(Month)9.5              factor(Month)10
                 88.97823                     45.75068
        factor(Month)10.5  factor(Known)1:factor(Month)6.5
                 79.32922                     62.22860
   factor(Known)1:factor(Month)7  factor(Known)1:factor(Month)7.5
                 72.97035                     74.98864
   factor(Known)1:factor(Month)8  factor(Known)1:factor(Month)8.5
                 83.27730                     84.45320
   factor(Known)1:factor(Month)9  factor(Known)1:factor(Month)9.5
                 88.76787                     90.09916
  factor(Known)1:factor(Month)10 factor(Known)1:factor(Month)10.5
                 47.06030                     80.19147
```

```
hellb <- read.table("/Users/chris/Documents/GeorgetownMPPMSFS/McCourtMPP/Semester3Fall2017MPP/Math426ALA/hellbender.txt", header
= T)
attach(hellb)
library(nlme)
################################################################## #### Longitudinally Sampled Variable 1 - Current #####

# First restrict the data to the first longitudinally sampled variable we are interested in and then convert to wide format.
install.packages('tidyverse', repos = "http://cran.us.r-project.org")
library(tidyverse)
current <- hellb[c(1:4)]
attach(current)
current_wide <- spread(current, key=Month, value=Current)

# Descriptive statistics for the two groups
library(psych)
describeBy(current_wide, current_wide$Known)

# Plot the observed trajectories for the individual streams in each group
current_wide_k <- current_wide[which(current_wide$Known == 1),3:12]
current_wide_u <- current_wide[which(current_wide$Known == 0),3:12]

matplot(c(6,6.5,7,7.6,8,8.5,9,9.5,10,10.5), t(current_wide_k), type='b', pch=20, lty=1, main= "Current by Bimonth (Known)",
col='gray', xlab =
          "Bimonth", ylab = "Current")
matplot(c(6,6.5,7,7.6,8,8.5,9,9.5,10,10.5), t(current_wide_u), type='b', pch=20, lty=1, main= "Current by Bimonth (Unknown)",
col='gray', xlab =
          "Bimonth", ylab = "Current")

# PLot the mean response profiles for Current for each respective group (With Hellbender population and without)
plot(c(6,6.5,7,7.5, 8, 8.5, 9, 9.5, 10, 10.5), apply(current_wide_k, 2, mean),  ylim = c(0,1.75), xlab = 'Time (bimonthly)',
ylab = 'Mean Current', main = "Current response per stream", type = 'n')
abline(v = axTicks(1), h = axTicks(2), col = rgb(0.75, 0.75, 0.75, alpha = 0.5), lty = 3)
lines(c(6,6.5,7,7.5, 8, 8.5, 9, 9.5, 10, 10.5), apply(current_wide_k, 2, mean), type = 'b', pch = 1, lty = 2)
lines(c(6,6.5,7,7.5, 8, 8.5, 9, 9.5, 10, 10.5), apply(current_wide_u, 2, mean), type = 'b', pch = 16)
legend('bottomleft', legend = c('Known', 'Unknown'), lty =
         c(2,1), pch = c(1,16), cex = 0.5)

# Under certain conditions, we can look at the raw matrix to give me a sense of where to start with making assumptions about the
covariance.
cov(current_wide[,c(3:12)])
cor(current_wide[,c(3:12)])

#### Assessing Covariance Structures ####
# Setting up the data
attach(current)

current$occur <- rep(1:10, length(unique(current$ID)))

head(current)

current <- current[order(current$ID),]

# Unstructured Covariance - Does not work as there are too many parameters (55).

# Assessing the compound symmetry structure
#Homogeneous
model2 <- gls(Current ~ factor(Known) * factor(Month), data = current, correlation = corCompSymm(form = ~ occur | ID))
summary(model2)

getVarCov(model2)
cov2cor(getVarCov(model2))

# Heterogeneous
model3 <- gls(Current ~ factor(Known) * factor(Month), data = current, correlation = corCompSymm(form = ~ occur | ID), weights =
varIdent(form = ~ 1 | occur))
summary(model3)

getVarCov(model3)
cov2cor(getVarCov(model3))

anova(model2, model3)
# In this case it looks like, based on the p-value of 0.5736, we fail to reject the null and would conclude that homogeneous is
a better fit here (Model2).

# Toeplitz

# Homogeneous
model4 <- gls(Current ~ factor(Known)*factor(Month), data = current, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0))
summary(model4)

getVarCov(model4)
cov2cor(getVarCov(model4))

# Heterogeneous
model5 <- gls(Current ~ factor(Known)*factor(Month), data = current, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0),
              weights = varIdent( form = ~ 1 | occur))
summary(model5)
```

```
getVarCov(model5)
cov2cor(getVarCov(model5))

anova(model4, model5)
# Based on the p-value of 0.3719, we fail to reject the null and conclude that homogeneous Toeplitz is more appropriate for
fitting the model (Model4).

# Checking out AR1 Structure
# Homogeneous
model6 <- gls(Current ~ factor(Known) * factor(Month), data = current, correlation = corAR1(form = ~ occur | ID))
summary(model6)

getVarCov(model6)
cov2cor(getVarCov(model6))

# Heterogeneous
model7 <- gls(Current ~ factor(Known) * factor(Month), data = current, correlation = corAR1(form = ~ occur | ID), weights =
varIdent(form = ~ 1 | occur))
summary(model7)

getVarCov(model7)
cov2cor(getVarCov(model7))

anova(model6, model7)
# In this case,based on the p-value of 0.2979, it looks like we fail to reject the null and would conclude that homogeneous is a
better fit here (Model6).

# Homogeneous Toeplitz and Homogeneous AR1
anova(model4,model6)
# These are nested. Based on the p-value of 2e-04, it would appear that Homogeneous AR1 is the most appropriate model. However,
Ar1 is more restrictive of an assumption. Given that the unstructured model doesn't converge, the Toeplitz model is the most
conservative. You might be making too strong of an assumption with AR1, so we wouldn't be able to trust the inferences. In fact,
we can confirm this in comparing the standard errors on the coefficients between the heterogeneous Toeplitz model and the
homogeneous AR1 model. Overall, the standard errors are smaller on the majority of the coefficients in the Toeplitz model.
Similarily, the p-values are also generally lower on coefficients in the model with the Toeplitz covariance structure.
Therefore, we will choose to use the Toeplitz covariance structure to fit the model.If we had sampled more streams, potentially
we would have a clearer picture.

# Homogeneous Compound and Homogeneous AR1 - AIC/BIC
anova(model2, model6)
# These are not nested. But in looking at the AIC and BIC, it looks like compound symmetry may be better because we are being
constrained to just two parameters. However, this is not a test with null hypotheses and we can therefore not make strong
conclusions from just looking at the AIC and BIC. Compound symmetric suggests that successive measurements have the same amount
of correlation between them.Is there any affect of time on the correlation as the measurements get farther apart? Yes, and
therefore compound symmetry is likely not the most appropriate.

# Homogeneous Compound and Homogeneous Toeplitz
anova(model2, model4)
# These are nested. As made evident by the p-value of 0.3077, the compound symmetric model appears more appropriate.

# Homogeneous Toeplitz and Homogeneous AR1
anova(model4,model6)
# These are nested. Based on the p-value of 2e-04, it would appear that Homogeneous AR1 is the most appropriate model. However,
Ar1 is more restrictive of an assumption. Given that the unstructured model doesn't converge, the Toeplitz model is the most
conservative. You might be making too strong of an assumption with AR1, so we wouldn't be able to trust the inferences. In fact,
we can confirm this in comparing the standard errors on the coefficients between the heterogeneous Toeplitz model and the
homogeneous AR1 model. Overall, the standard errors are smaller  on the majority of the coefficients in the Toeplitz model.
Similarily, the p-values are also generally lower on coefficients in the model with the Toeplitz covariance structure.
Therefore, we will choose to use the Toeplitz covariance structure to fit the model.If we had sampled more streams, potentially
we would have a clearer picture.


###### Assessing whether the parametric model with time as linear might be appropriate ########
model7 <- gls(Current ~ factor(Known) * factor(Month), data = current, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0),
method = 'ML')

model8 <- gls(Current ~ factor(Known) + Month, data = current, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0), method =
'ML')

anova(model7, model8)

# As evident by the p-value of 0.1403 and lower AIC for Model8, a non-parametric linear model may have been appropriate here.

###### Assessing quadratics #######
current$Month2 <- current$Month^2

model9 <- gls(Current ~ Known*Month + Known*Month2 - Known,
              data = current, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0), method = 'ML')

anova(model7, model9)
# the p-value of 0.0798 indicates that we can reject the null, and therefore may need to consider the non-parametric model with
quadratics.

# It is very clear from the plot that a linear splice model would not be appropriate here given how many potential  "knots"
there would be. We therefore do not attempt to fit a linear splice model here. We're modeling the means, so we'd want to think
about if we were to fit a linear model to this and treat it linearly, maybe it seems like it's flat and then a slight increase
```

```
after a time-point, then that is the way to think about linear spllicing. When we pick a linear model also, we're trying to
smooth out the variability in the means.

##### Our tests for coicidence and parallel are below. #####
model4 <- gls(Current ~ factor(Known)*factor(Month), data = current, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0))
# Test the null that the mean response profiles of the two groups are idential. (Slide 12: Coincidence deals with group)
# Test the null that the mean response profiles of the two groups are flat (Slide 12: Coincidence deals with time)
# Test the null hypothesis that the pattern of means over bimonthly measurement intervals are parallel for the two groups.
(Slide 12: Parallel deals with group * time)

capture.output(anova(model4), file = "New.txt")

# As evidenced by the p-value of .2187 on factor(Known), we cannot reject the null and conclude that the two groups are
identical (coincicent).
# As evidenced by the p-value of .7115 on factor(Month), we can reject the null and conclude that the mean response profiles are
not flat.
# As evidenced by the p-value of .3964 on factor(Known):factor(Month), we cannot reject the null and conclude that the mean
response profiles are parallel
# Based on these results, we would conclude that the two groups are not flat, but they are parallel and coincident.

# The potential lack of significant result here could be due to some limitations of the data.

## Conclusion of the output for our selected model ##
install.packages('sandwich', repos = "http://cran.us.r-project.org")
library(sandwich)
indModel <- lm(Current ~ factor(Known) * factor(Month), data = current)
diag(sandwich(indModel))
sqrt(diag(sandwich(indModel)))

# Look at empirical estimators -> The most delicious of all statistical estimators. If we are overly concerend about fit. How do
the inferences change? We fit an independence model (Multiple linear regression). Do it without weighting, covariance structure.
This gives us the variances of the coefficients. So we'll look at the variances. Take the sqrt of the variances diagonal. The
sandwich estimator is robust to misspecification. If we're concerned that we've placed too much structure. It's not necessarily
that one will be stronger than the other, but putting too strict a structure on your model would lead to incorrect standard
errors which could lead to false inferences. If SE from compound are small, you might reject coefficients, but robust estimators
might be larger leading you to fail to reject. If you correctly specificy the model, you'll gain power. Sandwich is robust but
not powerful.
summary(model4)


# We do not have any statistically significant coefficients.The presence of the salamnnders in the stream does not appear to
have any influence on the environmental factor of the stream.

####################################################################
##Longitudinally Sampled Variable 2 - Dissolved Oxygen Level ##

disox <- hellb[c(1:3,5)]
attach(disox)
disox_wide <- spread(disox, key=Month, value=DO)

# Descriptive statistics for the two groups
library(psych)
describeBy(disox_wide, disox_wide$Known)

# Plot the observed trajectories for the individual streams in each group
disox_wide_k <- disox_wide[which(disox_wide$Known == 1),3:12]
disox_wide_u <- disox_wide[which(disox_wide$Known == 0),3:12]

matplot(c(6,6.5,7,7.6,8,8.5,9,9.5,10,10.5), t(disox_wide_k), type='b', pch=20, lty=1, main= "DO by Bimonth (Known)", col='gray',
xlab =
          "Bimonth", ylab = "DO")
matplot(c(6,6.5,7,7.6,8,8.5,9,9.5,10,10.5), t(disox_wide_u), type='b', pch=20, lty=1, main= "DO by Bimonth (Unknown)",
col='gray', xlab =
          "Bimonth", ylab = "DO")

# PLot the mean response profiles for Current for each respective group (With Hellbender population and without)
plot(c(6,6.5,7,7.5, 8, 8.5, 9, 9.5, 10, 10.5), apply(disox_wide_k, 2, mean),  ylim = c(7,13), xlab = 'Time (bimonthly)', ylab =
'Mean DO', main = "DO response per stream", type = 'n')
abline(v = axTicks(1), h = axTicks(2), col = rgb(0.75, 0.75, 0.75, alpha = 0.5), lty = 3)
lines(c(6,6.5,7,7.5, 8, 8.5, 9, 9.5, 10, 10.5), apply(disox_wide_k, 2, mean), type = 'b', pch = 1, lty = 2)
lines(c(6,6.5,7,7.5, 8, 8.5, 9, 9.5, 10, 10.5), apply(disox_wide_u, 2, mean), type = 'b', pch = 16)
legend('bottomleft', legend = c('Known', 'Unknown'), lty =
         c(2,1), pch = c(1,16), cex = 0.5)

# Under certain conditions, we can look at the raw matrix to give me a sense of where to start with making assumptions about the
covariance.
cov(disox_wide[,c(3:12)])
cor(disox_wide[,c(3:12)])

# No evident pattern from the covariance structure. However, similar to with Current, we suspect that we'll want a covariance
structure that does not have constant variance and that allows correlations to change as intervals between time measurements
being correlated increase. Interesting that there are negative values though.

#### Assessing Covariance Structures ####
# Setting up the data
attach(disox)
```

```
disox$occur <- rep(1:10, length(unique(disox$ID)))

head(disox)

disox <- disox[order(disox$ID),]

# Assessing the compound symmetry structure
#Homogeneous
model10 <- gls(DO ~ factor(Known) * factor(Month), data = disox, correlation = corCompSymm(form = ~ occur | ID))
summary(model10)

getVarCov(model10)
cov2cor(getVarCov(model10))

# Heterogeneous
model11 <- gls(DO ~ factor(Known) * factor(Month), data = disox, correlation = corCompSymm(form = ~ occur | ID), weights =
varIdent(form = ~ 1 | occur))
summary(model11)

getVarCov(model11)
cov2cor(getVarCov(model11))

anova(model10, model11)
# In this case, as made evident by the p-value of 0.3139, it looks like we fail to reject the null and would conclude that
homogeneous is a better fit here (Model10).

# Toeplitz

# Homogeneous
model12 <- gls(DO ~ factor(Known) * factor(Month), data = disox, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0))
summary(model12)

getVarCov(model12)
cov2cor(getVarCov(model12))

# Heterogeneous
model13 <- gls(DO ~ factor(Known) * factor(Month), data = disox, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0),
              weights = varIdent( form = ~ 1 | occur))
summary(model13)

getVarCov(model13)
cov2cor(getVarCov(model13))

anova(model12, model13)
# Based on the p-value of 0.3237, it looks again like the homogeneous Toeplitz structure is best here.

# Checking out AR1 Structure
# Homogeneous
model14 <- gls(DO ~ factor(Known) * factor(Month), data = disox, correlation = corAR1(form = ~ occur | ID))
summary(model14)

getVarCov(model14)
cov2cor(getVarCov(model14))

# Heterogeneous
model15 <- gls(DO ~ factor(Known) * factor(Month), data = disox, correlation = corAR1(form = ~ occur | ID), weights =
varIdent(form = ~ 1 | occur))
summary(model15)

getVarCov(model15)
cov2cor(getVarCov(model15))

anova(model14, model15)
# In this case, as made evident by the p-value of 0.3899, it looks like we fail to reject the null and would conclude that
homogeneous is a better fit here (Model14).

# Homogeneous Compound and Homogeneous AR1 - AIC/BIC
anova(model10, model14)
# These are not nested. But in looking at the AIC and BIC, it looks like compound symmetry may be marginally better because we
are being constrained to just two parameters. However, this is not a test with null hypotheses and we can therefore not make
strong conclusions from just looking at the AIC and BIC. Compound symmetric suggests that successive measurements have the same
amount of correlation between them. Espeically given that the AIC is only slightly better for homogeneous CS, we can reject the
conclusion that CS would be the best fit.

# Homogeneous Compound and Homogeneous Toeplitz
anova(model10, model12)
# The high p-value of 0.7245 would lead us to fail to reject the null and conclude that the homogeneous CS model would be the
best. However, again given that CS does not suit the longitudinal data, we will decide to fit with the Homogeneous Toeplitz.

# Homogeneous AR1 and Homogeneous Toeplitz
anova(model14, model12)
# The high p-value of 0.6639 would lead us to the conclusion that the AR1 model might be most appropriate here. However, Ar1 is
more restrictive of an assumption. Given that the unstructured model doesn't converge, the Toeplitz model is the most
conservative. You might be making too strong of an assumption with AR1, so we wouldn't be able to trust the inferences. In fact,
we can confirm this in comparing the standard errors on the coefficients between the heterogeneous Toeplitz model and the
homogeneous AR1 model. Overall, the standard errors are smaller  on the majority of the coefficients in the Toeplitz model.
Similarly, the p-values are also generally lower on coefficients in the model with the Toeplitz covariance structure.
```

Therefore, we will choose to use the Toeplitz covariance structure to fit the model.If we had sampled more streams, potentially we would have a clearer picture.

```
###### Assessing whether the parametric model with time as linear might be appropriate ########
model16 <- gls(DO ~ factor(Known) * factor(Month), data = disox, correlation = corAR1(form = ~ occur | ID), method = 'ML')

model17 <- gls(DO ~ factor(Known) + Month, data = disox, correlation = corAR1(form = ~ occur | ID), method = 'ML')

anova(model16, model17)

# As evident by the p-value of <.0001 and the lower AIC for model16, we can conclude that the non-parametric model is more appropriate for fitting the data.

# It is very clear from the plot that a linear splice model would not be appropriate here given how many potential  "knots" there would be. We therefore do not attempt to fit a linear splice model here. We're modeling the means, so we'd want to think about if we were to fit a linear model to this and treat it linearly, maybe it seems like it's flat and then a slight increase after a time-point, then that is the way to think about linear spllicing. When we pick a linear model also, we're trying to smooth out the variability in the means.

##### Our tests for coicidence and parallel are below. #####

# Test the null that the mean response profiles of the two groups are idential. (Slide 12: Coincidence deals with group)
# Test the null that the mean response profiles of the two groups are flat (Slide 12: Coincidence deals with time)
# Test the null hypothesis that the pattern of means over bimonthly measurement intervals are parallel for the two groups.
(Slide 12: Parallel deals with group * time)
# Coincidence
model19 <- gls(DO ~ factor(Month), data = disox, correlation = corAR1(form = ~ occur | ID), method = 'ML')

model20 <- gls(DO ~ factor(Known) * factor(Month), data = disox, correlation = corAR1(form = ~ occur | ID), method = 'ML')

anova(model19, model20)
# Based on the p-value of .0108 and the lower AIC for model20, we conclude that there is not coincidence.

# Flat
model21 <- gls(DO ~ factor(Known), data = disox, correlation = corAR1(form = ~ occur | ID), method = 'ML')

model22 <- gls(DO ~ factor(Known) * factor(Month), data = disox, correlation = corAR1(form = ~ occur | ID), method = 'ML')

anova(model21, model22)
# As evident by the p-value of < .00001 and the lower AIC for Model22, we can conclude that there is no flatness.

# Parallel
model23 <- gls(DO ~ factor(Known) + factor(Month), data = disox, correlation = corAR1(form = ~ occur | ID), method = 'ML')

model24 <- gls(DO ~ factor(Known) * factor(Month), data = disox, correlation = corAR1(form = ~ occur | ID), method = 'ML')

anova(model23, model24)
# As evident by the p-value of .0097 and the lower AIC for model24, we can conclude that they are not parallel.

# The bigger issue is just the small samples period period (One has 2 and the other only has 4)

## Conclusion of the output for our selected model ##
indModel2 <- lm(DO ~ factor(Known) * factor(Month), data = disox)
diag(sandwich(indModel2))
sqrt(diag(sandwich(indModel2)))


summary(model14)

###################################################################
# Longitudinally Sampled Variable 3 - Conductivity #

conductivity <- hellb[c(1:3, 6)]
attach(conductivity)
conductivity_wide <- spread(conductivity, key=Month, value=Conductivity)

# Descriptive statistics for the two groups
describeBy(conductivity_wide, conductivity_wide$Known)

# Plot the observed trajectories for the individual streams in each group
conductivity_wide_k <- conductivity_wide[which(conductivity_wide$Known == 1),3:12]
conductivity_wide_u <- conductivity_wide[which(conductivity_wide$Known == 0),3:12]

matplot(c(6,6.5,7,7.6,8,8.5,9,9.5,10,10.5), t(conductivity_wide_k), type='b', pch=20, lty=1, main= "Conductivity by Bimonth (Known)", col='gray', xlab =
          "Bimonth", ylab = "Conductivity")
matplot(c(6,6.5,7,7.6,8,8.5,9,9.5,10,10.5), t(conductivity_wide_u), type='b', pch=20, lty=1, main= "Conductivity by Bimonth (Unknown)", col='gray', xlab =
          "Bimonth", ylab = "Conductivity")

# PLot the mean response profiles for Current for each respective group (With Hellbender population and without)
plot(c(6,6.5,7,7.5, 8, 8.5, 9, 9.5, 10, 10.5), apply(conductivity_wide_k, 2, mean),  ylim = c(30,175), xlab = 'Time (bimonthly)', ylab = 'Mean Conductivity', main = "Conductivity response per stream", type = 'n')
abline(v = axTicks(1), h = axTicks(2), col = rgb(0.75, 0.75, 0.75, alpha = 0.5), lty = 3)
lines(c(6,6.5,7,7.5, 8, 8.5, 9, 9.5, 10, 10.5), apply(conductivity_wide_k, 2, mean), type = 'b', pch = 1, lty = 2)
lines(c(6,6.5,7,7.5, 8, 8.5, 9, 9.5, 10, 10.5), apply(conductivity_wide_u, 2, mean), type = 'b', pch = 16)
```

```
legend('bottomleft', legend = c('Known', 'Unknown'), lty =
         c(2,1), pch = c(1,16), cex = 0.5)

# Under certain conditions, we can look at the raw matrix to give me a sense of where to start with making assumptions about the
covariance.
cov(conductivity_wide[,c(3:12)])
cor(conductivity_wide[,c(3:12)])

#### Assessing Covariance Structures ####
# Setting up the data
attach(conductivity)

conductivity$occur <- rep(1:10, length(unique(conductivity$ID)))

head(conductivity)

conductivity <- conductivity[order(conductivity$ID),]

# Assessing the compound symmetry structure
#Homogeneous
model25 <- gls(Conductivity ~ factor(Known) * factor(Month), data = conductivity, correlation = corCompSymm(form = ~ occur |
ID))
summary(model25)

getVarCov(model25)
cov2cor(getVarCov(model25))

# Heterogeneous
model26 <- gls(Conductivity ~ factor(Known) * factor(Month), data = conductivity, correlation = corCompSymm(form = ~ occur |
ID), weights = varIdent(form = ~ 1 | occur))
summary(model26)

getVarCov(model26)
cov2cor(getVarCov(model26))

anova(model25, model26)
# In this case it looks like, based on the p-value of 0.4967, we fail to reject the null and would conclude that homogeneous is
a better fit here (Smaller df) (Model9).

# Toeplitz

# Homogeneous
model27 <- gls(Conductivity ~ factor(Known)*factor(Month), data = conductivity, correlation = corARMA(form = ~ 1 | ID, p = 3, q
= 0))
summary(model27)

getVarCov(model27)
cov2cor(getVarCov(model27))

# Heterogeneous
model28 <- gls(Conductivity ~ factor(Known)*factor(Month), data = conductivity, correlation = corARMA(form = ~ 1 | ID, p = 3, q
= 0),
              weights = varIdent( form = ~ 1 | occur))
summary(model28)

getVarCov(model28)
cov2cor(getVarCov(model28))

anova(model27, model28)
# Given the p-value of 0.0028 and the lower AIC for model28, we conclude that the heterogeneous Toeplitz model is better here.

# Checking out AR1 Structure
# Homogeneous
model29 <- gls(Conductivity ~ factor(Known)*factor(Month), data = conductivity, correlation = corAR1(form = ~ occur | ID))
summary(model29)

getVarCov(model29)
cov2cor(getVarCov(model29))

# Heterogeneous
model30 <- gls(Conductivity ~ factor(Known)*factor(Month), data = conductivity, correlation = corAR1(form = ~ occur | ID),
weights = varIdent(form = ~ 1 | occur))
summary(model30)

getVarCov(model30)
cov2cor(getVarCov(model30))

anova(model29, model30)
# In this case, based on the p-value of 0.0026 and the lower AIC on model30, we can reject the null and would conclude that
heterogeneous is a better fit here (Model30).

# Homogeneous Compound and Heterogeneous Toeplitz
anova(model25, model28)
# Based on the p-value of .013 and the lower AIC on Model28, we'd conclude that the heterogeneous Toeplitz is a better fit.

# Homogeneous Compound and Heterogeneous AR1 - AIC/BIC
anova(model25, model30)
```

```
# In looking at the p-value of 0.3121 and the lower df on model25, we'd conclude that compound symmetric is best here.

# Heterogeneous Toeplitz and Heterogeneous AR1
anova(model28, model30)
# Based on the p-value of 0.0012 and the lower AIC on model28, we'd go with the heterogeneous toeplitz model here.


###### Assessing whether the parametric model with time as linear might be appropriate ########
model31 <- gls(Conductivity ~ factor(Known)*factor(Month), data = conductivity, correlation = corARMA(form = ~ 1 | ID, p = 3, q
= 0),
               weights = varIdent( form = ~ 1 | occur), method = 'ML')

model32 <- gls(Conductivity ~ factor(Known) + Month, data = conductivity, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0),
               weights = varIdent( form = ~ 1 | occur), method = 'ML')

anova(model31, model32)

# While the p-value of .0084 allows us to reject the null and conclude that the two models are statistically different, given
that the AICs are virtually similar, we are going to choode the the non-parametric model in order to better assess the potential
effect of time on how the treatment affects conductivity.

###### Assessing quadratics #######
conductivity$Month2 <- conductivity$Month^2

model33 <- gls(Conductivity ~ Known*Month + Known*Month2 - Known,
              data=conductivity, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0), weights = varIdent( form = ~ 1 | occur),
method = 'ML')


anova(model31, model33)
# The p-value demonstrates statistical significance and the AIC is slightly lower for the non-quadratic.

##### Our tests for coicidence and parallel are below. #####

# Test the null that the mean response profiles of the two groups are idential. (Slide 12: Coincidence deals with group)
# Test the null that the mean response profiles of the two groups are flat (Slide 12: Coincidence deals with time)
# Test the null hypothesis that the pattern of means over bimonthly measurement intervals are parallel for the two groups.
(Slide 12: Parallel deals with group * time)
# Coincidence
model34 <- gls(Conductivity ~ factor(Month), data = conductivity, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0),
               weights = varIdent( form = ~ 1 | occur), method = 'ML')

anova(model34, model31)
# Based on the p-value of .2036 and the lower df for model34, we conclude that there is coincidence.

# Flat
model35 <- gls(Conductivity ~ factor(Known), data = conductivity, correlation = corARMA(form = ~ 1 | ID, p = 3, q = 0),
               weights = varIdent( form = ~ 1 | occur), method = 'ML')

anova(model35, model31)
# As evident by the p-value of .0123 and the lower AIC for Model35, we can conclude that there is flatness.

# Parallel
model36 <- gls(Conductivity ~ factor(Known)+factor(Month), data = conductivity, correlation = corARMA(form = ~ 1 | ID, p = 3, q
= 0),
               weights = varIdent( form = ~ 1 | occur), method = 'ML')

anova(model36, model31)
# As evident by the p-value of 0.2619 and the lower AIC for model36, we can also conclude parallel.

## Conclusion of the output for our selected model ##
indModel3 <- lm(Conductivity ~ factor(Known) * factor(Month), data = conductivity)
diag(sandwich(indModel3))
sqrt(diag(sandwich(indModel3)))

summary(model28)

install.packages("stargazer", repos = "http://cran.us.r-project.org")
library(stargazer)
stargazer(model4, model14, model28, type="text",title="Analysis of Mean Response Profile Results", digits=4,style="all",
out="Model14dtdfg.txt")
```