Understanding New Zealand River Water Quality during 1980-2017

Group 6: Georgia Shank, Ryan Feng, Angelina Chen

2022-10-16

ABSTRACT

To model turbidity as a measure of river water quality using nitrogen content, phosphorus content, E.coli content, and types of surrounding land covers, we retrieve data from the Ministry of Environment which covers over 1000 rivers throughout New Zealand across 3 time periods and was collected in 2017, base on this set of data we conduct different statistical tests such as principal component analysis, linear discriminant analysis, and factor analysis to find out the relationships between variables. In conclusion, our study shows that there is a correlation between nitrogen and turbidity, and that phosphorus has changed linearly over the past years which is then proved to be valid for prediction.

INTRODUCTION

New Zealand as the number 14th country in the world with the cleanest water, (Tiseo, 2022) has always had a focus on maintaining water quality at a high standard. Having high-quality water is not only crucial to humans, as waterways also support a diverse range of plants and animals. (Why is water quality important?, 2022) There are large volume of observational studies and their corresponding data available around this topic, and we are interested in constructing models for water quality prediction based on different variables such as the content of waterborne bacteria, the concentration of trace elements within, time, and the surrounding environment, using data that were not collected for this purpose. By having a model as such, we will be able to produce reliable predictions on water quality using relevant elements and therefore deploy water treating schemes before the quality degrades, hence maintaining it at a high standard.

In this study, we purpose that water quality can be quantified by variable Turbidity and predicted by the levels of 3 numerical variables: phosphorus, nitrogen, and E. coli, and different types of dominant land cover and different periods the measurements belong to. Turbidity refers to light scattering by suspended particles, E.coli plays as an indicator of human or animal faecal contamination, the percentage of change comes from the percent change in a water quality variable per year, (Scott Larned, 2018), and the dominant landcover describes the main landcover near the river.

In the following part of the report, we will test our hypothesis by exploring the correlation between variables, and conducting multiple statistical tests.

METHODOLOGY

Data Cleansing

The raw data retrieved from the Ministry of Environment contains over 10,000 entry and 40 columns where the 1,000 rivers are recorded repeatedly for different variables and time. We first extracted the columns which contain information we need, and called this dataset "cleaned". Following the first extraction, we would likt to expand the "np_id_name" column which contained most of the useful variables we need. for example, to make turbidity of a river an individual column from the extracted dataset, we repeat this until all the

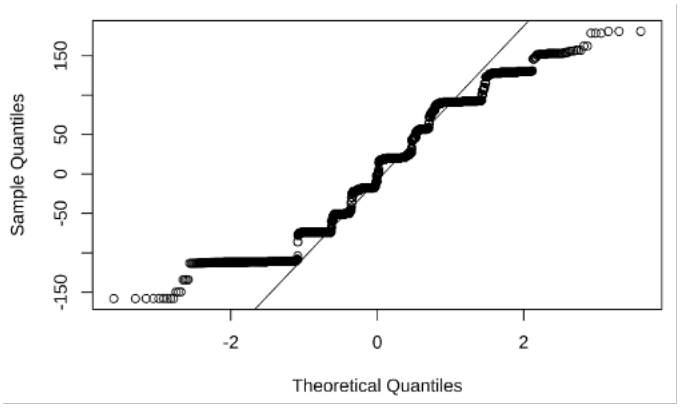
useful variables are expanded and categorised to their own columns, by the same time. After obtaining a wide format of data, we rename the columns and merge and stack them into one ready-to-use dataset.

Exploratory Data Analysis

The Exploratory Data Analysis is used for summarizing the variables, we also visualize variety plots to check if the dataset contains normal distributions and constant variance.

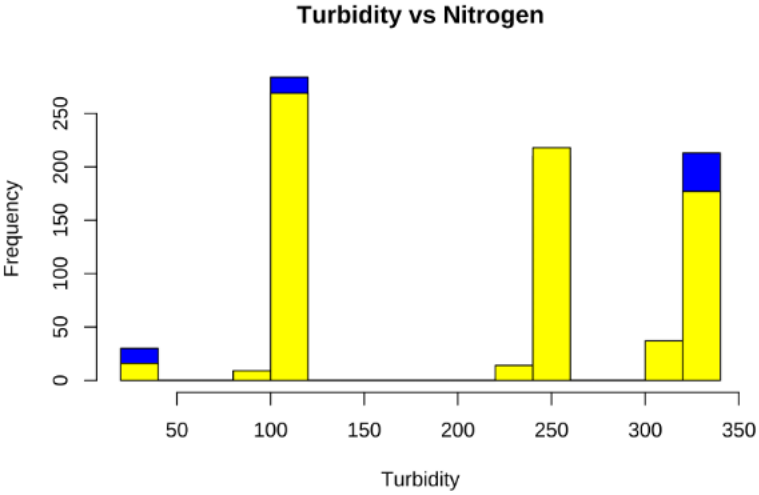
The qqplot below shows a slight departure from the assumption of normality and constant variance.

```
knitr::include_graphics("../Figures/qqplotEDA.png")
```



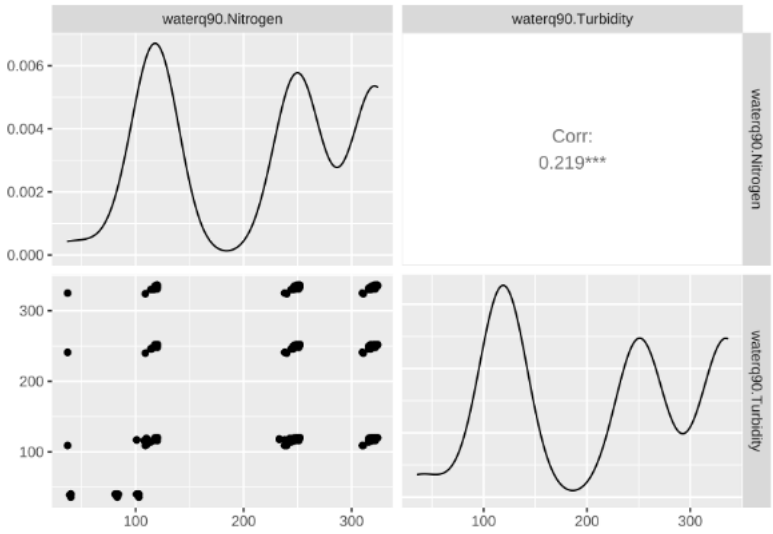
We found an interesting relationship from Coloured Histogram, both Nitrogen and Turbidity has a concided shape, there are strong evidence of a relevant relationship between Nitrogen and Turbidity.

```
knitr::include_graphics("../Figures/ColouredHisto.png")
```



We also found some differences in 3 periods of time across the waterquality.

```
knitr::include_graphics("../Figures/PairsplotEDA.png")
```



Principal Component Analysis

EDA visualized a relevant connection between Turbidity and Nitrogen, which reminds us of the hypothesis regarding as a potential factor, if Nitrogen can affect the water quality in New Zealand? Therefore, we ran a Principal analysis among Turbidity and 3 potential factors to find more evidence.

To apply the principal components, we used the function "prcomp" to load the PCA. Since too many observations could differ the useful visualizations, we reduced the sample of variables to 300.

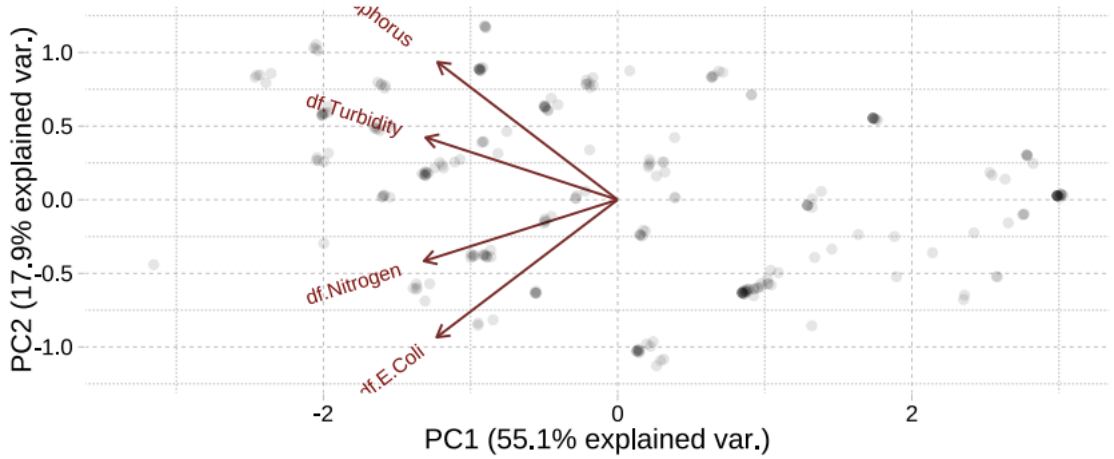
```
knitr::include_graphics("../Figures/Reduceddata.png")
```

```
set.seed(300399182)
sample.ind<-sample(length(wq$df.Nitrogen),size=300,replace = FALSE)
Reducedwq<-wq[sample.ind,]
PCA.Reducedwq<-prcomp(Reducedwq, center=TRUE, scale=TRUE)
summary(PCA.Reducedwq)
```

```
## Importance of components:
##                           PC1   PC2   PC3   PC4
## Standard deviation     1.484 0.846 0.743 0.728
## Proportion of Variance 0.551 0.179 0.138 0.132
## Cumulative Proportion  0.551 0.730 0.868 1.000
```

Thereafter, we made a few biplots to see the relationship among principal components. Retrieving the ggbiplot for reduced data (300 samples), we can find the biplot interpretation: 1. Smallest cosine of angle of contributions between corresponding variables (Nitrogen and Turbidity) in a 2-dimension graphic. 2. Uncorrelated variables are closed to each other 3. Similar direction of correlated variables

```
knitr::include_graphics("../Figures/BiplotsAllPeriods.png")
```

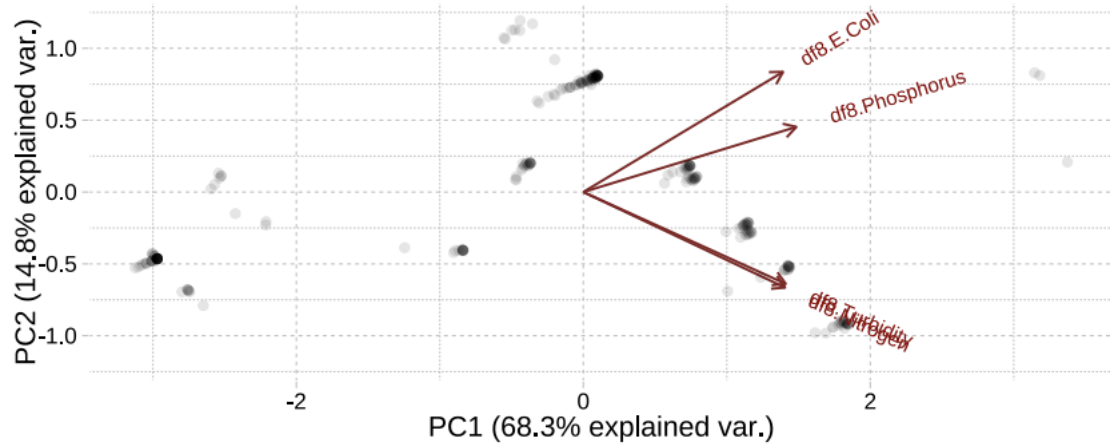


We also plot corresponding variables in the year 2008, beside the interpretation we found from all periods, interpretation in the year 2008 is:

1. The cosines of angle of contributions between corresponding variables (Nitrogen and Turbidity) are literally coincided in a 2-dimension graphic (smaller cosines than all periods)
2. Uncorrelated variables are closed to each other
3. Similar direction of correlated variables

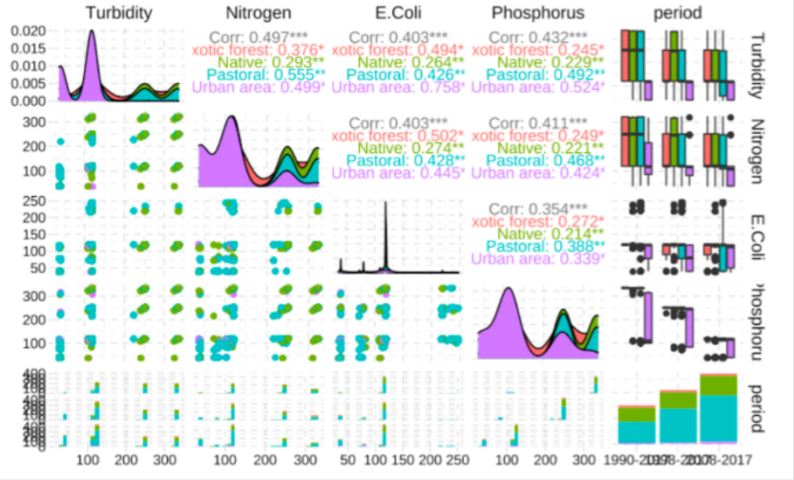
According to the biplots above ,we have corresponding variables Nitrogen and Turbidity has highly relevant and positive relationship.

```
knitr::include_graphics("../Figures/Biplots08.png")
```



The ggpairs plot reveals correlations for 4 majority landcovers in New Zealand, but we do not find strong difference of water quality among 4 landcovers. Furthermore, comparing of other 2 variables, we find both Turbidity and Nitrogen has highly related curve.

```
knitr::include_graphics("../Figures/ggpairsPCA.png")
```



Factor Analysis

We conduct factor analysis using different numbers of factos ranging from 0-3. Although p-values show an increase trend from 0 to 3.44e-05, it is still too small to not reject the null hypothesis to conclude a valid factor analysis.

Linear Discriminant Analysis

```
summary(data)
```

```
##    Turbidity        Nitrogen         E.Coli        Phosphorus        period
##  Min.   : 36.0   Min.   : 36.0   Min.   : 36.0   Min.   : 36   1990-2017: 740
##  1st Qu.:118.0   1st Qu.:119.0   1st Qu.:109.0   1st Qu.:119   1998-2017:1049
##  Median :120.0   Median :120.0   Median :119.0   Median :216   2008-2017:1365
##  Mean   :178.6   Mean   :181.3   Mean   :109.2   Mean   :193
##  3rd Qu.:252.0   3rd Qu.:251.0   3rd Qu.:120.0   3rd Qu.:252
##  Max.   :336.0   Max.   :324.0   Max.   :247.0   Max.   :336
```

Pairs plot

```
pairs.panels(data[, -5],
gap=0,
bg=c("red", "green", "blue")[data$period],
pch=21)
```

We can see scatterplots of each combination of variables and their correlation coefficients. Observations seem to be grouped into periods, however, some pairs of variables exhibit overlap between periods. Turbidity and nitrogen appear to have the highest correlation of 0.50 (Fig. . .).

The data was split into subsets of 60% and 40% of the data, to be used as training and testing data, respectively.

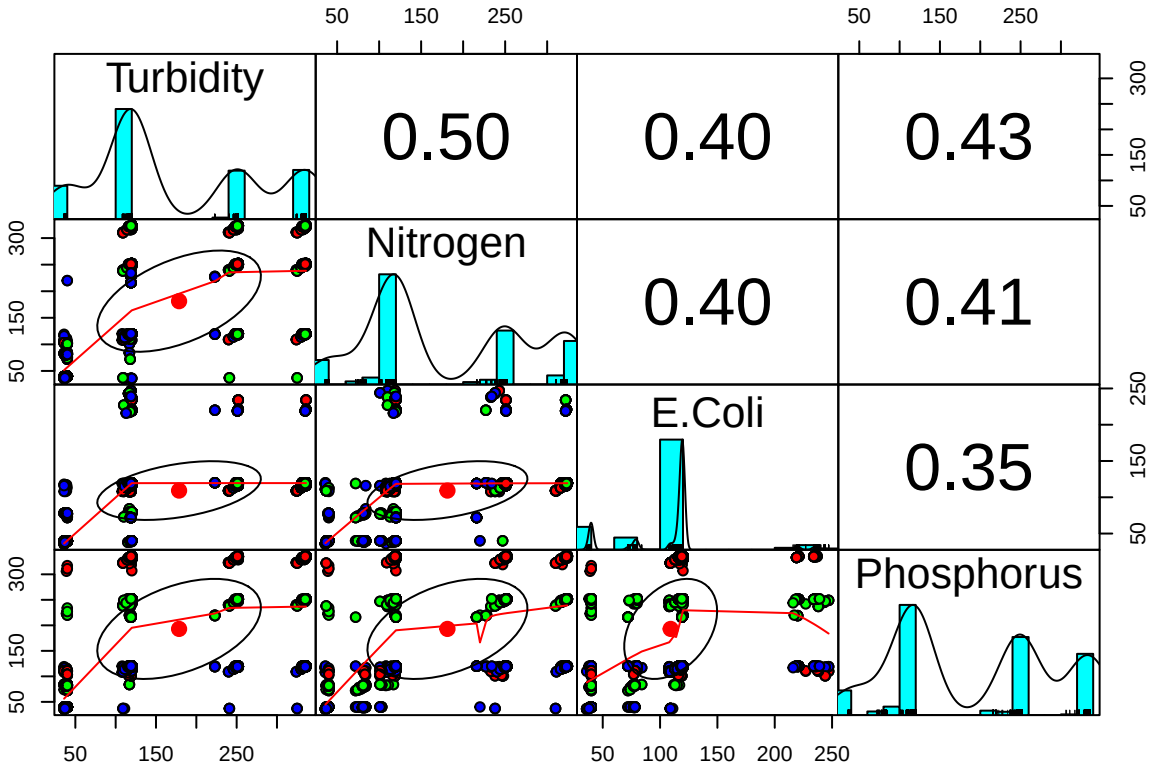


Figure 1: Pairs plot of river health indicators throughout the three periods; 1990-2017 (red), 1998-2017 (green) and 2008-2017 (blue)

```
set.seed(1234567890, kind="Mersenne-Twister")
ind <- sample(c("Train", "Test"),
              nrow(data),
              replace=TRUE,
              prob=c(0.6, 0.4))
Train <- data[ind=="Train",]
Test <-  data[ind=="Test",]
```

Linear Discriminant Analysis

```
(LDA <- lda(period ~ Turbidity+Nitrogen+E.Coli+Phosphorus, data=Train))
```

```
## Call:
## lda(period ~ Turbidity + Nitrogen + E.Coli + Phosphorus, data = Train)
##
## Prior probabilities of groups:
## 1990-2017 1998-2017 2008-2017
## 0.2288714 0.3328084 0.4383202
##
## Group means:
##           Turbidity Nitrogen   E.Coli Phosphorus
## 1990-2017  218.0344 212.4266 123.0826   312.6055
## 1998-2017  177.6041 185.3880 106.8344   228.6167
## 2008-2017  159.5114 160.3317 104.0491   103.8287
##
## Coefficients of linear discriminants:
##                      LD1          LD2
## Turbidity   0.002984324 -0.005721030
## Nitrogen    0.002018386  0.001963659
## E.Coli      0.006720757 -0.020306350
## Phosphorus -0.025056680  0.002501191
##
## Proportion of trace:
##    LD1    LD2
## 0.9965 0.0035
```

The estimates of prior probabilities show that 22.89% of the training data corresponds to the period 1990-2017, 33.28% to the period 1998-2017, and 43.83% to the period 2008-2017.

The coefficients of the linear discriminant functions show the linear combinations of predictor variables comprising the LDA. LD1 = 0.003xTurbidity + 0.002xNitrogen + 0.0067xE.Coli + -0.025xPhosphorus. Phosphorus appears to be the greatest contributor to LD1. LD2 = -0.006xTurbidity + 0.002xNitrogen + -0.02xE.Coli + 0.0025xPhosphorus.

The proportion of the trace shows that 99.65% of variance between periods can be explained by the first Linear Discriminant (LD1), and 0.35% by the second Linear Discriminant (LD2).

```
Pred <- predict(LDA)
par(mar=c(1,1,1,1))
ldahist(data=Pred$x[,1], g=Train$period)
```

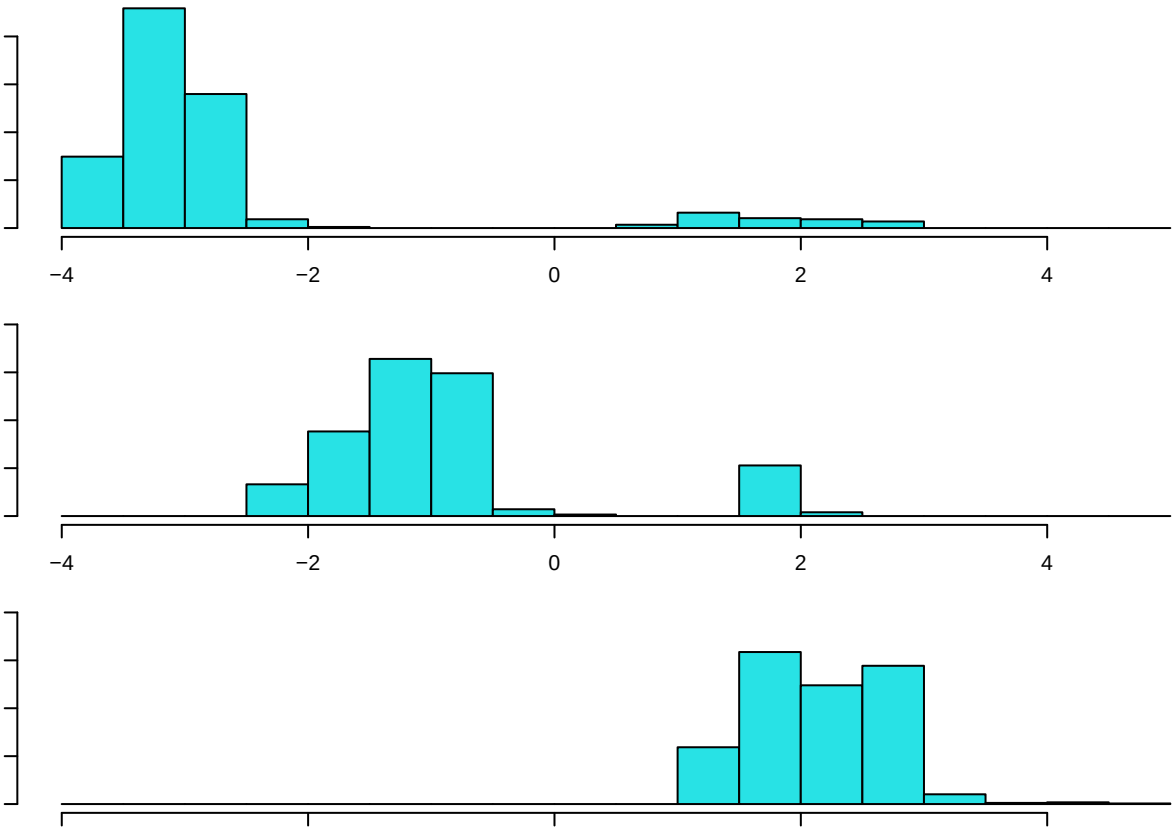


Figure 2: Separation between periods 1990-2017, 1998-2017 and 2008-2017 for each Linear Discriminant

```
ldahist(data=Pred$x[,2], g=Train$period)
```

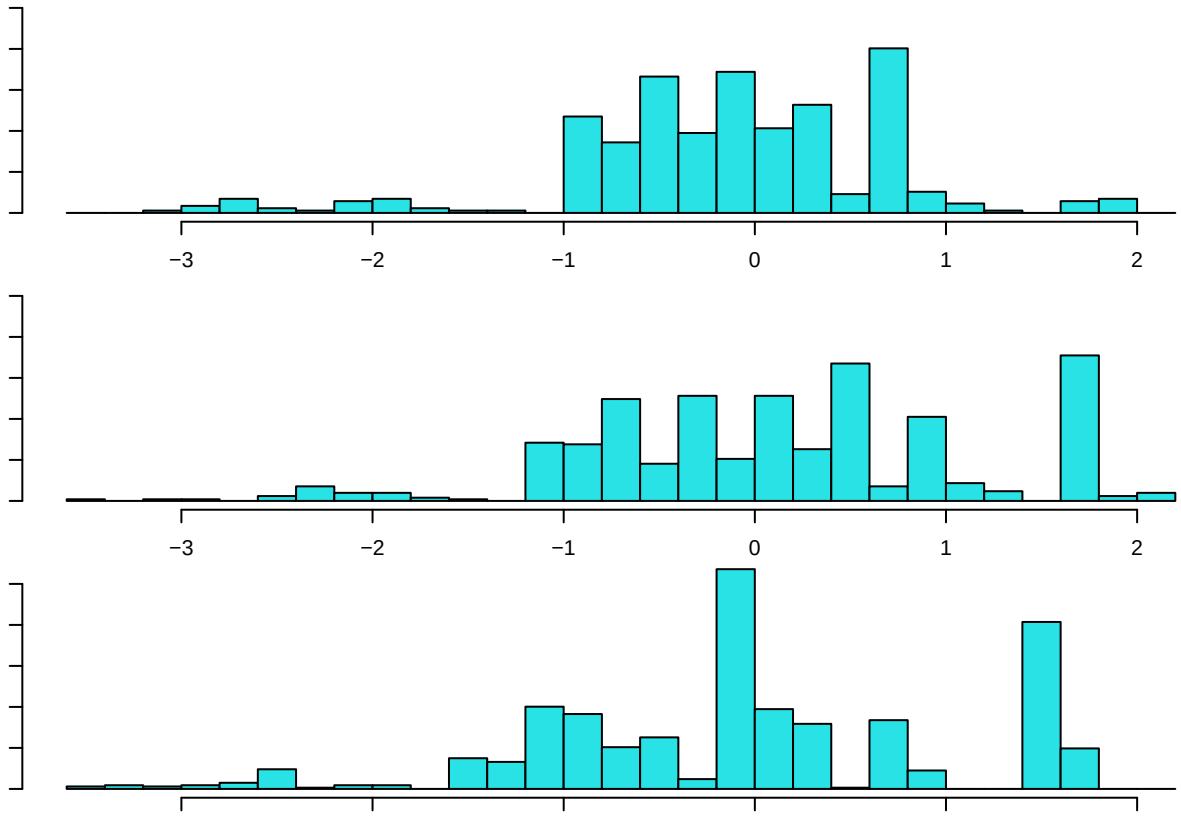
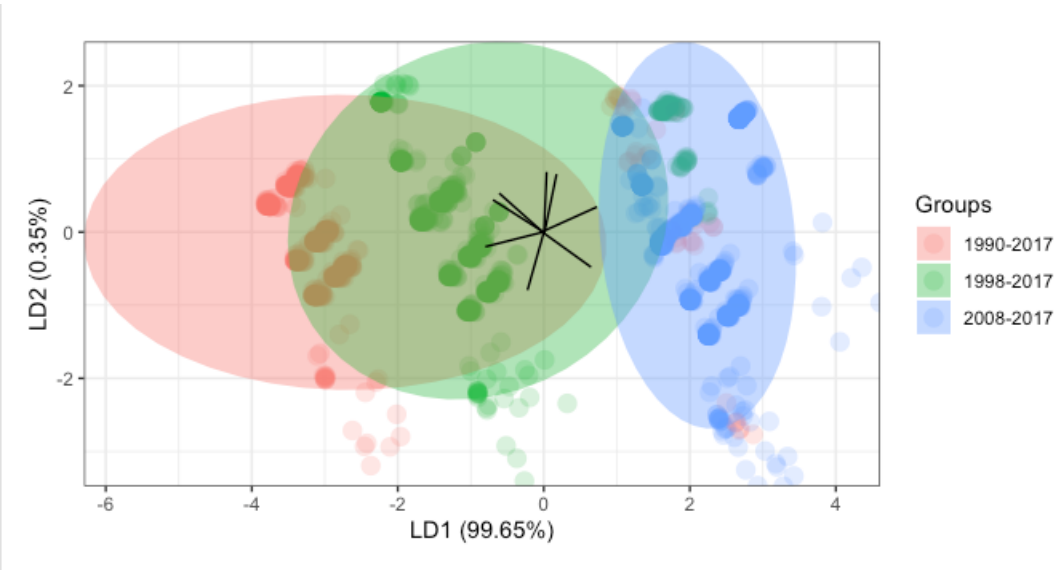


Figure 3: Separation between periods 1990-2017, 1998-2017 and 2008-2017 for each Linear Discriminant

For LD1, there is large separation and little overlap between periods, reflecting 99.65% separation (Fig . . .). Conversely, we can see that there is little separation (0.35%) and substantial overlap between periods in LD2.

```
knitr::include_graphics("../Figures/ggord.png")
```



We can see that three clusters of observations exist, indicating the three time periods. However, there is some overlap on the right side, in which observations in periods 1990-2017 and 1998-2017 overlap with the cluster for 2008-2017. (Fig. . .)

```
partimat(period~., data=Train, method="lda")
```

The partition plot displays the classification of each observation in the training dataset, based on the LDA model, for each pair of variables. The top row exhibits much greater error rates (>0.5), indicating the greater occurrence of observations predicted to be in the incorrect periods. The bottom row, interestingly all pairs including phosphorus, shows much smaller error rates, indicating a lower amount of observations falling outside their predicted period (Fig . . .).

```
#Optimistic
Optimistic <- predict(LDA, Train)$class
(OCM <- table(Optimistic, Actual=Train$period))
```

```
##             Actual
## Optimistic  1990-2017 1998-2017 2008-2017
##   1990-2017       396         0         0
##   1998-2017         0       562         0
##   2008-2017        40        72       835
```

```
sum(diag(OCM))/sum(OCM)
```

```
## [1] 0.9412073
```

```
#Confusion matrix using test data
Realistic <- predict(LDA, Test)$class
(RCM <- table(Realistic, Actual=Test$period))
```

```
##             Actual
## Realistic   1990-2017 1998-2017 2008-2017
##   1990-2017       267         0         0
##   1998-2017         0       357         0
##   2008-2017        37        58       530
```

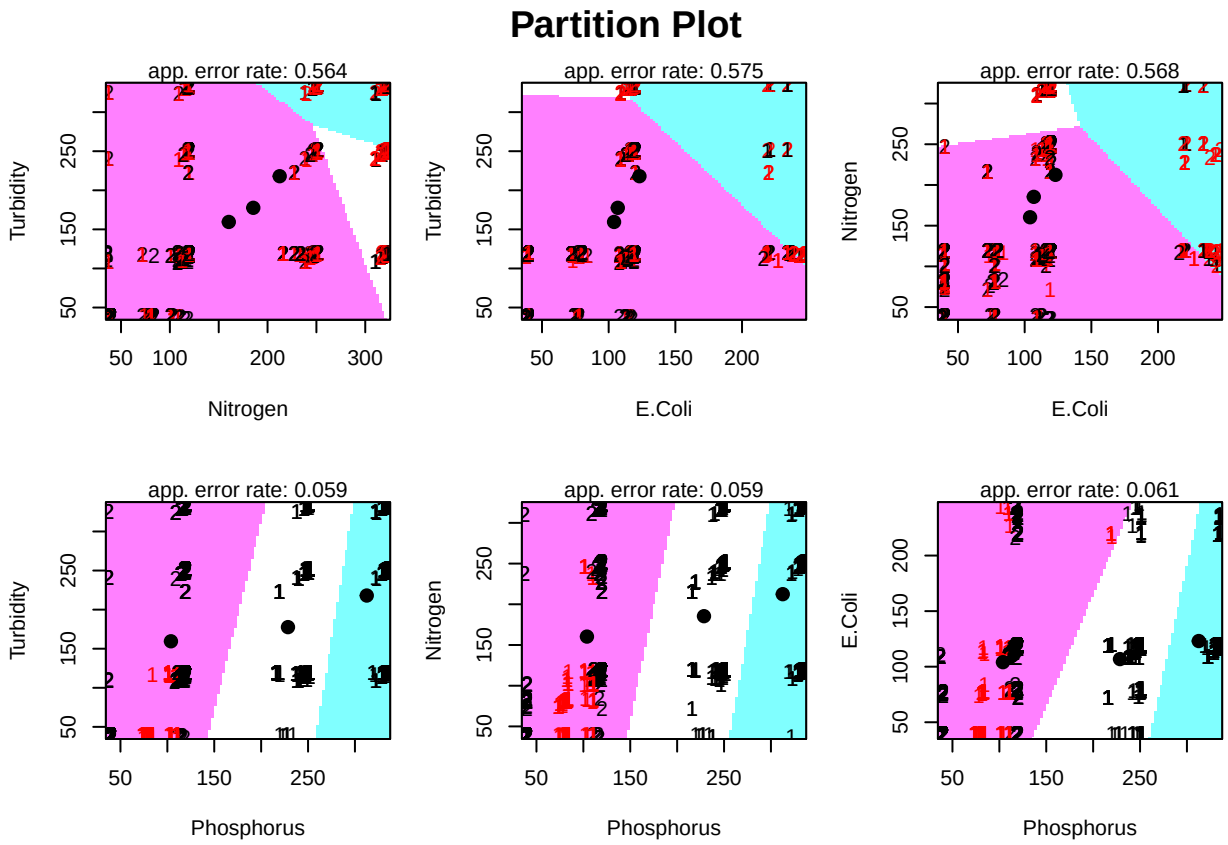


Figure 4: Partition Plot using the training data for river health observations over the three periods; 1990-2017, 298-2017 and 2008-2017

```
sum(diag(RCM))/sum(RCM)
```

```
## [1] 0.9239392
```

```
RCM <- data.frame(RCM)

#Tibble for test data
tab_fin_test=as_tibble(RCM)
colnames(tab_fin_test)=c("Target", "Prediction", "N")
plot_confusion_matrix(tab_fin_test, target_col = "Target",
  prediction_col = "Prediction",
  counts_col = "N")
```

```
## Warning in plot_confusion_matrix(tab_fin_test, target_col = "Target",
## prediction_col = "Prediction", : 'ggimage' is missing. Will not plot arrows and
## zero-shading.
```

```
## Warning in plot_confusion_matrix(tab_fin_test, target_col = "Target",
## prediction_col = "Prediction", : 'rsvg' is missing. Will not plot arrows and
## zero-shading.
```

Predictions using the LDA model proved to be 92.4% accurate.

The realistic confusion matrix using the test data shows that of all the observations where the target is 2008-2017, 84.8% were predicted correctly, 9.3% were predicted to belong to the period 1998-2017 and 5.9% to 1990-2017. However, 100% of the observations were correctly predicted for the targets 1998-2017 and 1990-2017.

Of all observations where the prediction is 2008-2017, 100% of them were actually 2008-2017. Where the prediction is 1998-2017, 14% of observations were from 2008-2017 and 86% were actually from 1998-2017. Finally, where the prediction was 1990-2017, 12.2% of observations were 2008-2017 and 87.8% were actually from 1990-2017.

Maximum Likelihood Classification Model (Naïve Bayes)

```
set.seed(1234567890, kind="Mersenne-Twister")
(model <- naive_bayes(period ~ Turbidity+Nitrogen+E.Coli+Phosphorus, data = Train, usekernel = T))
```

```
##
## ================================== Naive Bayes ==================================
##
##  Call:
## naive_bayes.formula(formula = period ~ Turbidity + Nitrogen +
##     E.Coli + Phosphorus, data = Train, usekernel = T)
##
## ---------------------------------------------------------------------------------
##
## Laplace smoothing: 0
##
## ---------------------------------------------------------------------------------
```

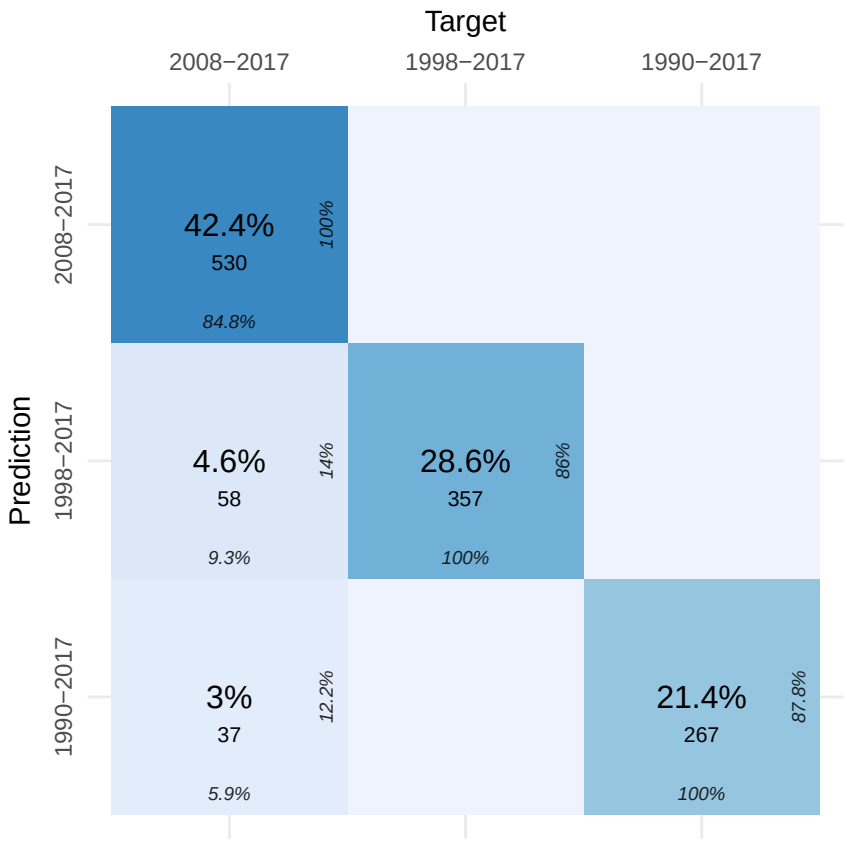


Figure 5: Realistic Confusion Matrix for the LDA using testing data for river health observations over three periods 1990-2017, 2998-2017 and 2008-2017

```
##
##  A priori probabilities:
##
## 1990-2017 1998-2017 2008-2017
## 0.2288714 0.3328084 0.4383202
##
## ------------------------------------------------------------------------------
##
##  Tables:
##
## ------------------------------------------------------------------------------
##  ::: Turbidity::1990-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (436 obs.);  Bandwidth 'bw' = 25.66
##
##        x                y
##  Min.   :-40.98   Min.   :4.688e-06
##  1st Qu.: 72.51   1st Qu.:5.456e-04
##  Median :186.00   Median :1.964e-03
##  Mean   :186.00   Mean   :2.200e-03
##  3rd Qu.:299.49   3rd Qu.:3.791e-03
##  Max.   :412.98   Max.   :5.798e-03
##
## ------------------------------------------------------------------------------
##  ::: Turbidity::1998-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (634 obs.);  Bandwidth 'bw' = 24.36
##
##        x                y
##  Min.   :-37.09   Min.   :1.478e-05
##  1st Qu.: 74.45   1st Qu.:7.725e-04
##  Median :186.00   Median :1.878e-03
##  Mean   :186.00   Mean   :2.238e-03
##  3rd Qu.:297.55   3rd Qu.:2.943e-03
##  Max.   :409.09   Max.   :7.831e-03
##
## ------------------------------------------------------------------------------
##  ::: Turbidity::2008-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (835 obs.);  Bandwidth 'bw' = 23.43
##
##        x                y
```

```
##  Min.   :-34.29   Min.   :0.0000244
##  1st Qu.: 75.86   1st Qu.:0.0007612
##  Median :186.00   Median :0.0019360
##  Mean   :186.00   Mean   :0.0022668
##  3rd Qu.:296.14   3rd Qu.:0.0028263
##  Max.   :406.29   Max.   :0.0083823
##
## ------------------------------------------------------------------------------
##  ::: Nitrogen::1990-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (436 obs.);  Bandwidth 'bw' = 23.68
##
##        x                y
##  Min.   :-34.05   Min.   :2.208e-06
##  1st Qu.: 73.22   1st Qu.:3.431e-04
##  Median :180.50   Median :1.926e-03
##  Mean   :180.50   Mean   :2.327e-03
##  3rd Qu.:287.78   3rd Qu.:4.104e-03
##  Max.   :395.05   Max.   :6.574e-03
##
## ------------------------------------------------------------------------------
##  ::: Nitrogen::1998-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (634 obs.);  Bandwidth 'bw' = 23.13
##
##        x                y
##  Min.   :-32.38   Min.   :9.210e-06
##  1st Qu.: 74.06   1st Qu.:6.863e-04
##  Median :180.50   Median :2.082e-03
##  Mean   :180.50   Mean   :2.346e-03
##  3rd Qu.:286.94   3rd Qu.:3.439e-03
##  Max.   :393.38   Max.   :7.523e-03
##
## ------------------------------------------------------------------------------
##  ::: Nitrogen::2008-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (835 obs.);  Bandwidth 'bw' = 22.12
##
##        x                y
##  Min.   :-30.35   Min.   :2.005e-05
##  1st Qu.: 74.82   1st Qu.:7.220e-04
##  Median :180.00   Median :2.108e-03
```

```
##  Mean   :180.00   Mean   :2.374e-03
##  3rd Qu.:285.18   3rd Qu.:2.847e-03
##  Max.   :390.35   Max.   :8.665e-03
##
## ------------------------------------------------------------------------------
##  ::: E.Coli::1990-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (436 obs.);  Bandwidth 'bw' = 0.3984
##
##        x               y
##  Min.   : 38.8   Min.   :0.00000
##  1st Qu.: 90.4   1st Qu.:0.00000
##  Median :142.0   Median :0.00000
##  Mean   :142.0   Mean   :0.00484
##  3rd Qu.:193.6   3rd Qu.:0.00000
##  Max.   :245.2   Max.   :0.44804
##
## ------------------------------------------------------------------------------
##  ::: E.Coli::1998-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (634 obs.);  Bandwidth 'bw' = 7.392
##
##        x                y
##  Min.   : 15.82   Min.   :0.000e+00
##  1st Qu.: 79.16   1st Qu.:4.222e-05
##  Median :142.50   Median :1.077e-03
##  Mean   :142.50   Mean   :3.943e-03
##  3rd Qu.:205.84   3rd Qu.:3.981e-03
##  Max.   :269.18   Max.   :3.557e-02
##
## ------------------------------------------------------------------------------
##  ::: E.Coli::2008-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (835 obs.);  Bandwidth 'bw' = 7.171
##
##        x                y
##  Min.   : 14.49   Min.   :0.000e+00
##  1st Qu.: 77.99   1st Qu.:3.687e-05
##  Median :141.50   Median :9.828e-04
##  Mean   :141.50   Mean   :3.933e-03
##  3rd Qu.:205.01   3rd Qu.:3.705e-03
##  Max.   :268.51   Max.   :3.490e-02
```

```
##
## ------------------------------------------------------------------------------
##  ::: Phosphorus::1990-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (436 obs.);  Bandwidth 'bw' = 0.5975
##
##        x                y
##  Min.   : 99.21   Min.   :0.000000
##  1st Qu.:158.85   1st Qu.:0.000000
##  Median :218.50   Median :0.000000
##  Mean   :218.50   Mean   :0.004186
##  3rd Qu.:278.15   3rd Qu.:0.000000
##  Max.   :337.79   Max.   :0.228799
##
## ------------------------------------------------------------------------------
##  ::: Phosphorus::1998-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (634 obs.);  Bandwidth 'bw' = 0.8778
##
##        x                y
##  Min.   : 70.37   Min.   :0.0000000
##  1st Qu.:116.43   1st Qu.:0.0000000
##  Median :162.50   Median :0.0000000
##  Mean   :162.50   Mean   :0.0054300
##  3rd Qu.:208.57   3rd Qu.:0.0006839
##  Max.   :254.63   Max.   :0.2141213
##
## ------------------------------------------------------------------------------
##  ::: Phosphorus::2008-2017 (KDE)
## ------------------------------------------------------------------------------
##
## Call:
##  density.default(x = x, na.rm = TRUE)
##
## Data: x (835 obs.);  Bandwidth 'bw' = 0.8745
##
##        x                y
##  Min.   : 33.38   Min.   :0.000000
##  1st Qu.: 55.69   1st Qu.:0.000000
##  Median : 78.00   Median :0.000000
##  Mean   : 78.00   Mean   :0.011187
##  3rd Qu.:100.31   3rd Qu.:0.002045
##  Max.   :122.62   Max.   :0.247198
##
## ------------------------------------------------------------------------------
```

```
#Optimistic
Optimistic <- predict(model, Train)
(OCM <- table(Optimistic, Actual=Train$period))
```

```
##            Actual
## Optimistic  1990-2017 1998-2017 2008-2017
##   1990-2017       416         0         5
##   1998-2017         1       625         0
##   2008-2017        19         9       830
```

```
sum(diag(OCM))/sum(OCM)
```

```
## [1] 0.9821522
```

```
#Confusion matrix using test data
Realistic <- predict(model, Test)
(RCM <- table(Realistic, Actual=Test$period))
```

```
##            Actual
## Realistic   1990-2017 1998-2017 2008-2017
##   1990-2017       275         0         7
##   1998-2017         0       407         0
##   2008-2017        29         8       523
```

```
sum(diag(RCM))/sum(RCM)
```

```
## [1] 0.9647718
```

```
RCM <- data.frame(RCM)

#Tibble for test data
tab_fin_test=as_tibble(RCM)
colnames(tab_fin_test)=c("Target", "Prediction", "N")
plot_confusion_matrix(tab_fin_test, target_col = "Target",
  prediction_col = "Prediction",
  counts_col = "N")
```

We can see that using Maximum Likelihood Classification (or Naïve Bayes), we achieve similar accuracy to the LDA model. This model achieved a 96.48% accuracy in predicting the time period of observations using the river health indicators, which is a higher accuracy rate than the LDA model (92.4%).

The realistic confusion matrix shows that 93.4% of observations with the target of 2008-2017 were predicted as such, 1.4% falsely predicted for 1998-2017 and 5.2% as 1990-2017. For the target of 1998-2017, 100% of observations were predicted correctly. With 1990-2017 as the target, 1990-2017, 2.5% of observations were falsely predicted as 2008-2017, and 97.5% were correctly predicted.

Observations predicted as 2008-2017, 98.7% of them were correct and 1.3% were actually belonging to 1990-2017. 1.9% of observations predicted as 1998-2017 were 2008-2017, and 98.1% were correctly predicted. Finally, 9.5% of observations predicted as 1990-2017 were instead 2008-2017 and 90.5% were correctly predicted.

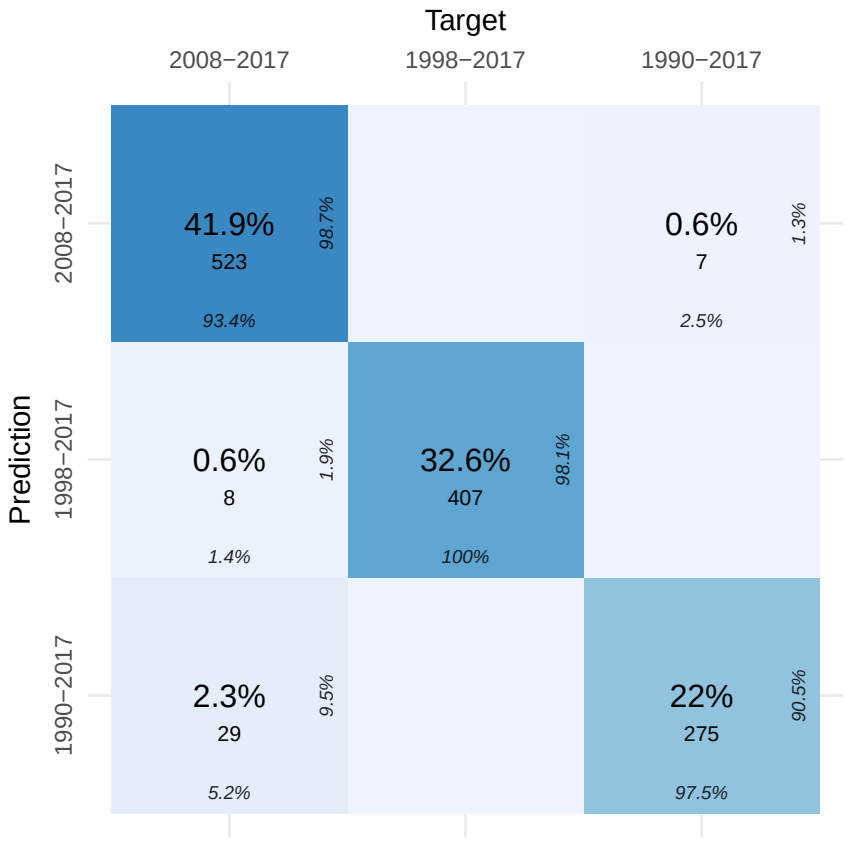


Figure 6: Realistic Confusion Matrix using Maximum Likelihood Classification for periods using river health indicators

CONCLUSION

Principal Component analysis reveals a highly relevant relationship between Nitrogen and Turbidity, since interpretation of biplots shows closesd angle between Nitrogen and Turbidity, in specifically period of 2008-2017, angle was nearly coincided, the correlation coefficient (cosine) was the largest out of 4 corresponding values, all angles among 4 variables are less than 90 degrees, therefore we can say nitrogen and turbidity are positive correlated, and Nitrogen can differ the water quality in New Zealand. The linear discriminant analysis showed that with 92.4% accuracy, we could classify observations from our data into time periods using river health indicators, with phosphorus being the greatest contributor. However, the Maximum Likelihood Classification model proved to show greater accuracy in predicting periods of observations, with an accuracy rate of 96.48%. This indicates that distinct patterns of river health indicators exist for each of the periods 1990-2017, 1998-2017 and 2008-2017. Thus, this shows that river health indicators could differ greatly over periods time.

BIBLIOGRAPHY

Scott Larned, A. W. (2018). Water Quality State and Trends in New Zealand Rivers. Christchunrch: Ministry for the Environment.

Tiseo, I. (2022, August 2). Leading clean waters scores worldwide as of 2021, by select country. Retrieved from Statista: https://www.statista.com/statistics/1143413/clean-water-index-region-globally/

Why is water quality important? (2022, August 10). Retrieved from ACT Government: https://www.environment.act.gov.au/water/act-healthy-waterways/water-quality/why-is-water-quality-important