Analysis of NSLMP and Industry dataset

Bela Klimesova

2025-04-11

Preparation of data for analysis

```r
library(ggplot2)
library(gridExtra)
library(grid)
library(rnaturalearth)
```

Data loading (NSLMP)

```r
NSLMP<-read.table("NSLMP.csv",
    header = TRUE,sep = ',',fill = TRUE,dec = ".",na.strings = "NA")

Industry<-read.table("Industry.csv",
    header = TRUE,sep = ',',fill = TRUE,dec = ".",na.strings = "NA")
```
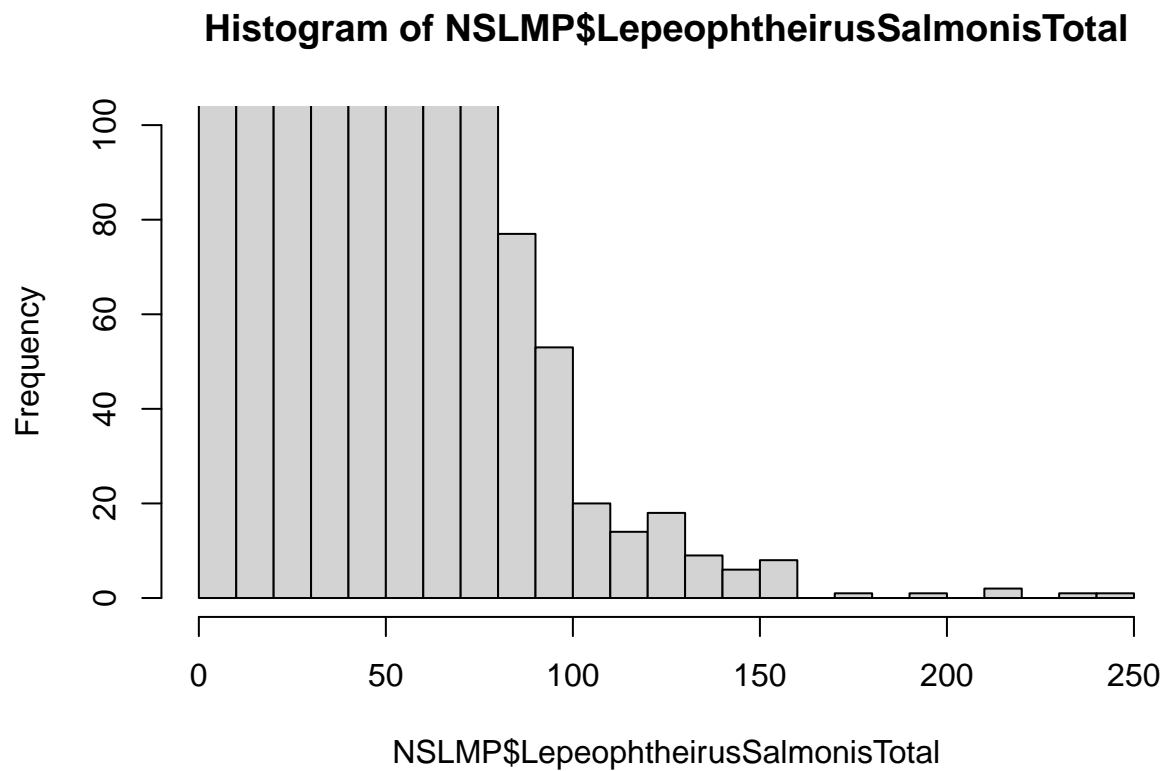
Exploration of the raw dataset and excluding outliers from analysis (3 datapoints)
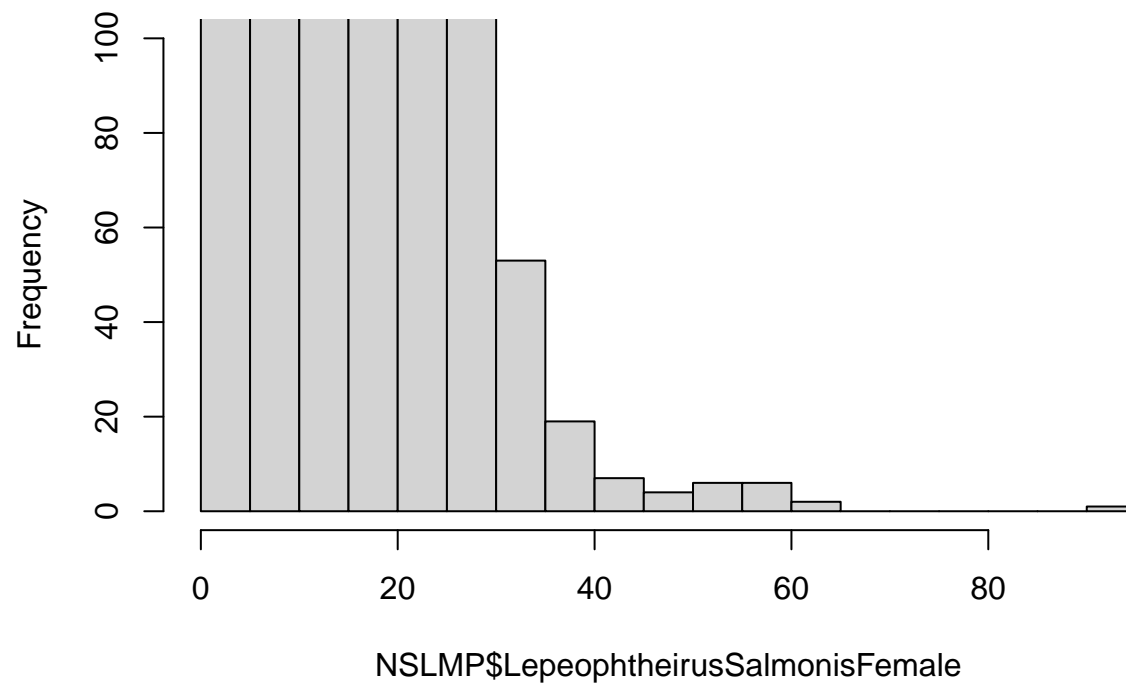
NSLMP

```r
nrow(NSLMP)
```

```
## [1] 178981
```

```
hist(NSLMP$LepeophtheirusSalmonisTotal,ylim=c(0,100))
```

**Histogram of NSLMP$LepeophtheirusSalmonisTotal**


```
NSLMP<-NSLMP[NSLMP$LepeophtheirusSalmonisTotal<160,]

hist(NSLMP$LepeophtheirusSalmonisFemale,ylim=c(0,100))
```

Histogram of NSLMP$LepeophtheirusSalmonisFemale


```
NSLMP<-NSLMP[NSLMP$LepeophtheirusSalmonisFemale<80,]

hist(NSLMP$LepeophtheirusSalmonisEggs,ylim=c(0,100))
```

Histogram of NSLMP$LepeophtheirusSalmonisEggs
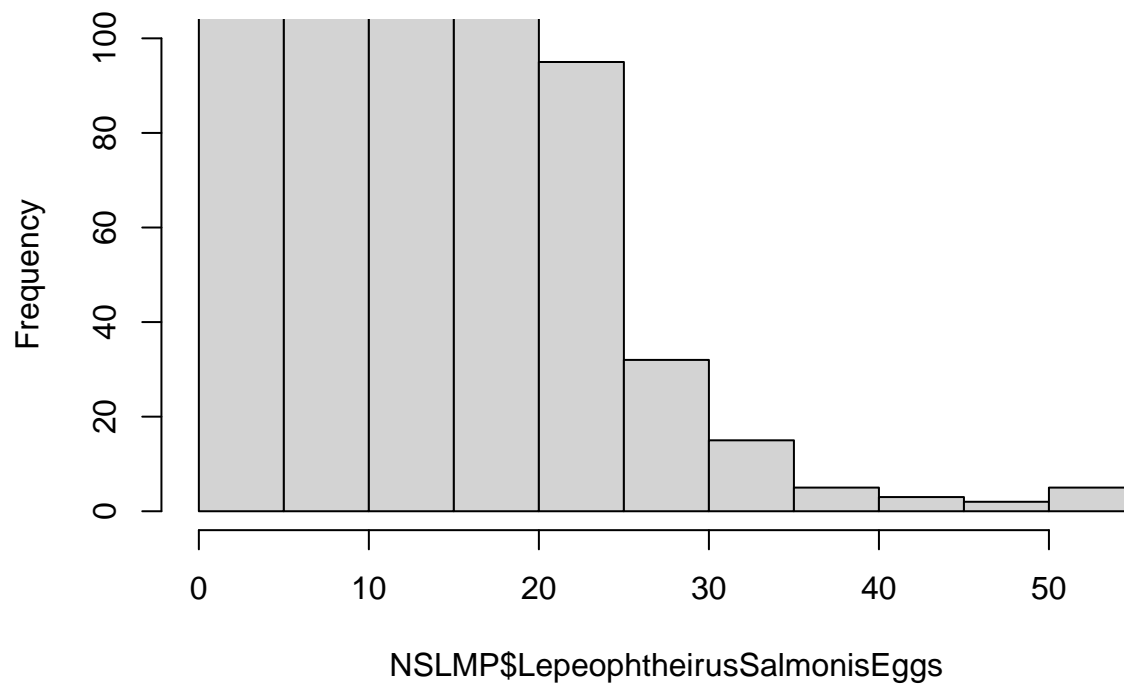

```r
nrow(NSLMP)
```

```
## [1] 178974
```

```r
length(unique(NSLMP$SiteName))
```

```
## [1] 39
```
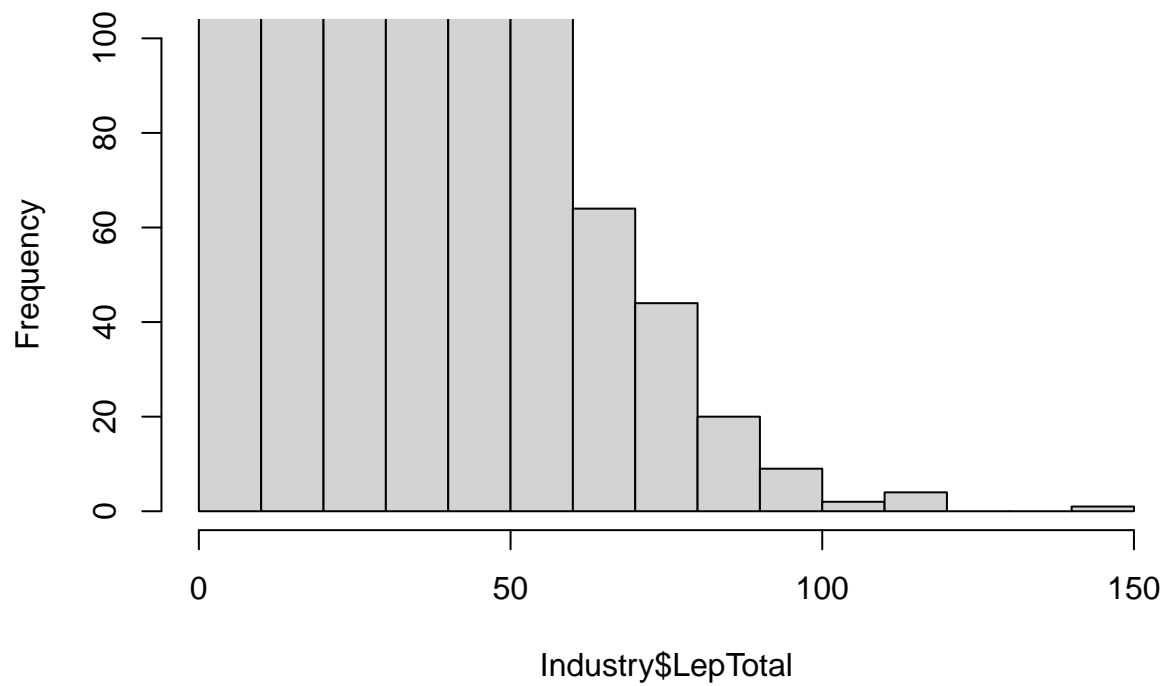
Industry
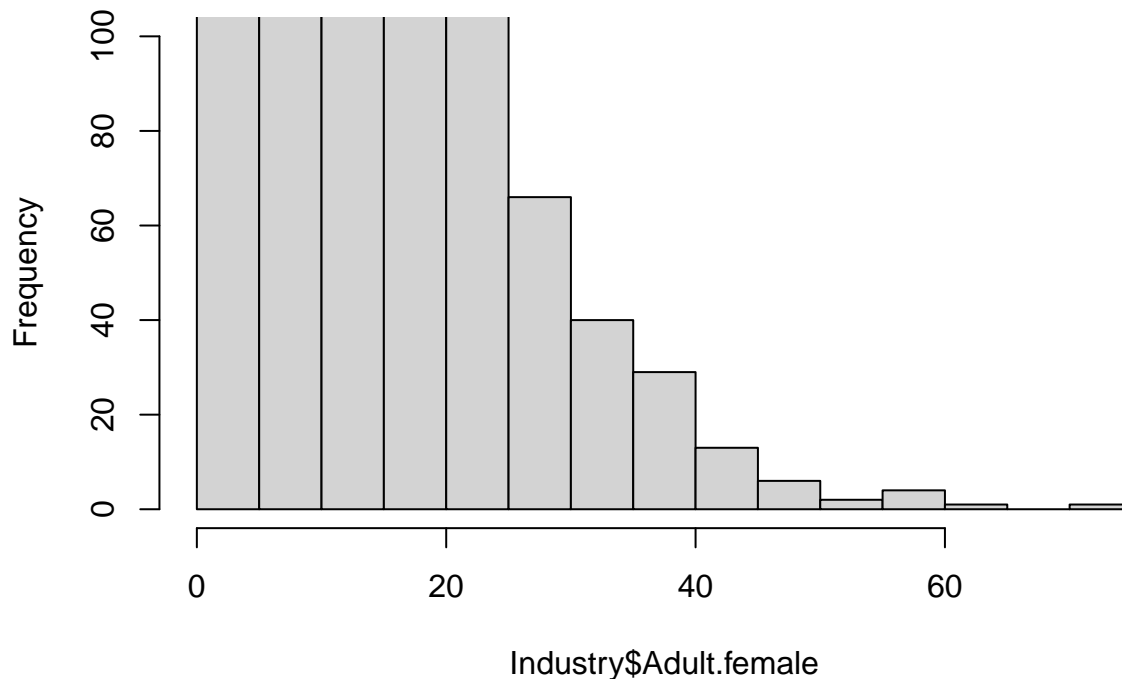
```r
nrow(Industry)
```

```
## [1] 119722
```

```r
hist(Industry$LepTotal,ylim=c(0,100))
```

Histogram of Industry$LepTotal


```
Industry<-Industry[Industry$LepTotal<130,]

hist(Industry$Adult.female,ylim=c(0,100))
```

Histogram of Industry$Adult.female


```r
nrow(Industry)
```

```
## [1] 119721
```

```r
length(unique(Industry$SiteName))
```

```
## [1] 14
```

Accuracy of sea lice counts: NSLMP vs. industry producer's data

Sea lice data from the NSLMP and industry producer monitoring were analyzed to assess differences in reported sea lice counts. All analyses focused on the average sea lice count per pen, either for total mobile sea lice or total adult females. Therefore, the datasets were adjusted to the pen level by calculating the mean of abundance per inspection (pen on salmon farm and date) and retaining only the necessary variables.

Industry

```r
Ind1<-aggregate(Industry$Total,list(Industry$SiteName,Industry$Date,Industry$Pen.ID,
                          Industry$Month,Industry$Year,Industry$Pen.no),mean)
colnames(Ind1)[7]<-'LepTotal'
Ind2<-aggregate(Industry$Adult.female,list(Industry$SiteName,Industry$Date,Industry$Pen.ID,
                          Industry$Month,Industry$Year,Industry$Pen.no),mean)
colnames(Ind2)[7]<-'Females'
```

```
Industry1<-merge(Ind1,Ind2,by=c('Group.1','Group.2','Group.3','Group.4',
                                          'Group.5','Group.6'),all=TRUE)
colnames(Industry1)[c(1:6)]<-c('Site','Date','Pen','Month','Year','PenID')
Industry1$Origin<-'Industry'
```

NSLMP

```
NSL1<-aggregate(NSLMP$LepeophtheirusSalmonisTotal,list(NSLMP$SiteName,NSLMP$Date,
             NSLMP$Month,NSLMP$Year,NSLMP$S.R,NSLMP$SampleID,NSLMP$Cage.No),mean)
colnames(NSL1)[8]<-'LepTotal'
NSL2<-aggregate(NSLMP$LepeophtheirusSalmonisFemale,list(NSLMP$SiteName,NSLMP$Date,
             NSLMP$Month,NSLMP$Year,NSLMP$S.R,NSLMP$SampleID,NSLMP$Cage.No),mean)
colnames(NSL2)[8]<-'Females'

NSLMP1<-merge(NSL1,NSL2,by=c('Group.1','Group.2','Group.3','Group.4',
                                     'Group.5','Group.6','Group.7'),all=TRUE)
colnames(NSLMP1)[c(1:7)]<-c('Site','Date','Month','Year','Pen','SampleID','PenID')
NSLMP1$Origin<-'NSLMP'
```

The NSLMP dataset was filtered to include only data from the same salmon farms and years as the industry producer's dataset (Oct 2015 – Dec 2019).

```
# Keep only data from the same sites, year, and month
NSLMP1$Ident<-paste(NSLMP1$Site,NSLMP1$Year,NSLMP1$Month)
Industry1$Ident<-paste(Industry1$Site,Industry1$Year,Industry1$Month)

NSLMP2<-NSLMP1[NSLMP1$Ident %in% Industry1$Ident,]
Industry2<-Industry1[Industry1$Ident %in% NSLMP2$Ident,]
```

A variable of day for date of inspection and ID for every individual mean sea lice count per pen was added to the dataset. A copy of the dataset was created.

```
# ID
Industry2$ID<-c(1:nrow(Industry2))
NSLMP2$ID<-c(1:nrow(NSLMP2))

# Day
Industry2$Day<-as.numeric(substr(as.character(Industry2$Date),9,10))
NSLMP2$Day<-as.numeric(substr(as.character(NSLMP2$Date),9,10))

# Copy of the datasets
Industry3<-Industry2
NSLMP3<-NSLMP2
```

To align the datasets, sea lice counts from individual pens in the NSLMP dataset were matched to the closest available inspection date in the industry producer's dataset. If multiple pens were sampled on the same farm and date, a pen different from the one inspected by the NSLMP was randomly selected. Sample points with inspection dates more than one month apart were excluded from the analysis.

Matching a single pen from NSLMP dataset to a pen in the Industry dataset with the closest inspection date

```r
set.seed(1991)

NSL_Sel<-NSLMP3[NSLMP3$ID %in% unique(NSLMP3$ID)[1],]
                # taking data about the first pen from NSLMP dataset
Ind_Sel<-Industry3[Industry3$Site==unique(NSL_Sel$Site),]
                # taking all data from Industry from the same farm as the NSLMP pen
Dates<-which(abs(difftime(Ind_Sel$Date,unique(NSL_Sel$Date),units="days"))== min(
      abs(difftime(Ind_Sel$Date,unique(NSL_Sel$Date),units="days"))))
                # Identifying which industry inspection date is the closest
                # to the NSLMP pen inspection date
Ind_Sel1<-Ind_Sel[Dates,]
                # Taking all data with the closest inspection dates
Ind_Sel2<-Ind_Sel1[!(Ind_Sel1$PenID %in% c(NSL_Sel$PenID)),]
                # Discarding observation from the same pen
Selection<-Ind_Sel2[sample(nrow(Ind_Sel2),1), ]
                # Randomly sampling one pen with the closest inspection date
Selection$MiID<-NSL_Sel$ID
Dat_Fin<-Selection
                # Creating dataset with selected Industry data
Industry3<-Industry3[!(Industry3$ID %in% Selection$ID),]
                # Discarding the selected pen from the original Industry data
```

The same process was then repeated for all pens in the NSLMP dataset with a for loop

```r
for (i in 2:length(unique(NSLMP3$ID))){
    NSL_Sel<-NSLMP3[NSLMP3$ID %in% unique(NSLMP3$ID)[i],]
    Ind_Sel<-Industry3[Industry3$Site==unique(NSL_Sel$Site),]
    Dates<-which(abs(difftime(Ind_Sel$Date,unique(NSL_Sel$Date),units="days"))==
                min(abs(difftime(Ind_Sel$Date,unique(NSL_Sel$Date),units="days"))))
    Ind_Sel1<-Ind_Sel[Dates,]
    Ind_Sel2<-Ind_Sel1[!(Ind_Sel1$PenID %in% c(NSL_Sel$PenID)),]
    if(nrow(Ind_Sel2)==0){
        Industry3<-Industry3[!(Industry3$ID %in% Ind_Sel1$ID),]
        Ind_Sel<-Industry3[Industry3$Site==unique(NSL_Sel$Site),]
        Dates<-which(abs(difftime(Ind_Sel$Date,unique(NSL_Sel$Date),units="days"))==
                min(abs(difftime(Ind_Sel$Date,unique(NSL_Sel$Date),units="days"))))
        Ind_Sel1<-Ind_Sel[Dates,]
    Ind_Sel2<-Ind_Sel1[!(Ind_Sel1$Pen %in% c(NSL_Sel$PenID)),]
        Selection<-Ind_Sel2[sample(nrow(Ind_Sel2),1), ]
        Selection$MiID<-NSL_Sel$ID
        Dat_Fin<-rbind(Dat_Fin,Selection)
        Industry3<-Industry3[!(Industry3$ID %in% Selection$ID),]
    }else{
        Selection<-Ind_Sel2[sample(nrow(Ind_Sel2),1), ]
        Selection$MiID<-NSL_Sel$ID
        Dat_Fin<-rbind(Dat_Fin,Selection)
        Industry3<-Industry3[!(Industry3$ID %in% Selection$ID),]
    }}
```

The new Industry dataset has much lower number of means pen pen compared to the original Industry dataset, because every mean per pen in the Industry dataset is matched to one NSLMP observation of mean per pen.

```r
# New Industry dataset
nrow(Dat_Fin)
```

```
## [1] 662
```

```r
# Original Industry dataset
nrow(Industry3)
```

```
## [1] 5001
```

```r
# NSLMP dataset
nrow(NSLMP3)
```

```
## [1] 662
```

The new Industry dataset was merged with the NSLMP dataset.

```r
# Unite datasets
Ind_select<-Dat_Fin[c(2,7,8,13)]
colnames(Ind_select)<-c('Mowi_Date','Mowi_Total','Mowi_Females','MiID')
NLS_select<-NSLMP3[c(2,8,9,12)]
colnames(NLS_select)<-c('MI_Date','MI_Total','MI_Females','MiID')
IND_NSL<-merge(Ind_select,NLS_select,by='MiID')
```

The final dataset was then checked for the differences in IDs between the Industry and NSLMP data and number of observations in the final dataset.

```r
# Check new dataset
setdiff(Ind_select$MiID,NLS_select$MiID)
```

```
## integer(0)
```

```r
setdiff(NLS_select$MiID,Ind_select$MiID)
```

```
## integer(0)
```

```r
nrow(IND_NSL)
```

```
## [1] 662
```

The number of days between matched pairs of pens between Industry and NSLMP inspection was determined and the differences explored. Dates of inspection more than 30 days apart were identified as outliers and excluded from the analysis.

```r
IND_NSL$Day_diff<-as.Date(IND_NSL$MI_Date)-as.Date(IND_NSL$Mowi_Date)
```

```
IND_NSL1<-IND_NSL[!(abs(IND_NSL$Day_diff)>7),]
# Number of discarded rows
nrow(IND_NSL[(abs(IND_NSL$Day_diff)>7),])
```

```
## [1] 53
```

```
# Histogram of the differences in days between samplings
hist(as.numeric(IND_NSL1$Day_diff))
```

Histogram of as.numeric(IND_NSL1$Day_diff)
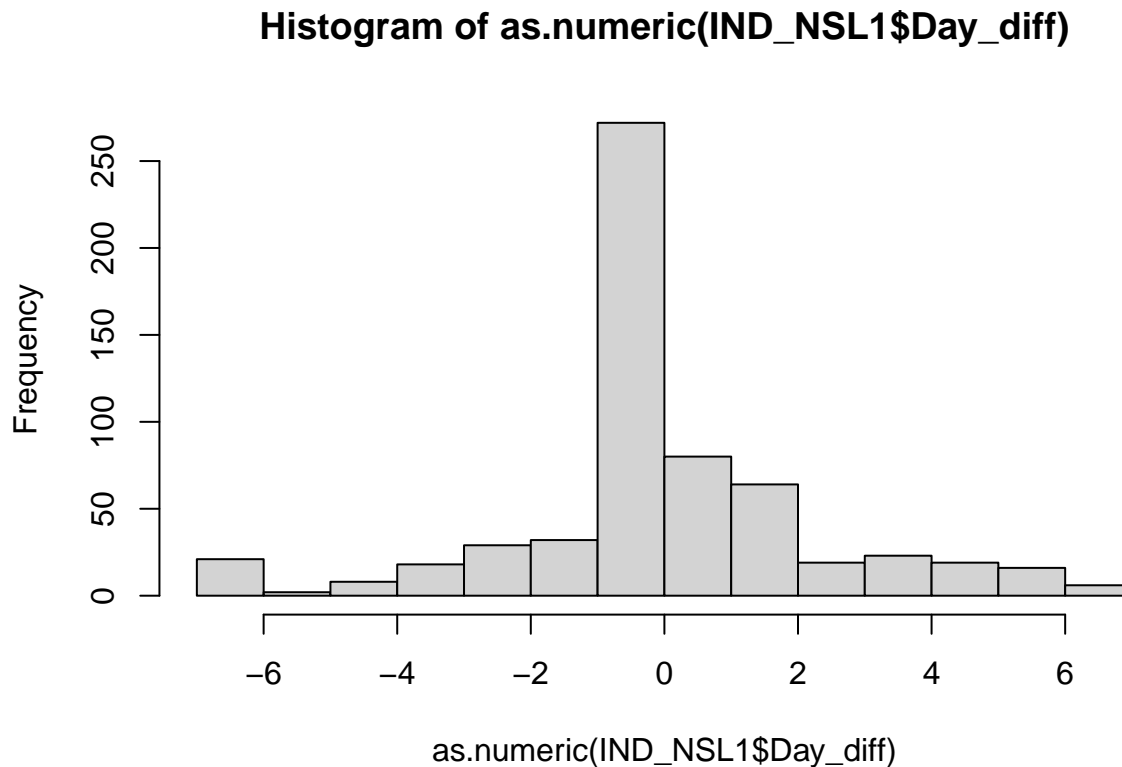

```
# Mean of differences
round(mean(abs(IND_NSL1$Day_diff)),2)
```

```
## Time difference of 1.49 days
```

The proportion of inspection in both datasets conducted on the same day was determined.

```
nrow(IND_NSL1[IND_NSL1$Day_diff==0,])/nrow(IND_NSL1)
```

```
## [1] 0.4466338
```

The mean and differences between the mean values per pen recorded by the Industry and by the NSLMP for both total mobiles and total adult females were calculated.

```
IND_NSL1$Total_diff<-IND_NSL1$MI_Total-IND_NSL1$Mowi_Total
IND_NSL1$Females_diff<-IND_NSL1$MI_Females-IND_NSL1$Mowi_Females
IND_NSL1$Total_mean<-apply(IND_NSL1[,c(3,6)],1,mean)
IND_NSL1$Females_mean<-apply(IND_NSL1[,c(4,7)],1,mean)
```

Test the normal distribution of differences between methods

```
shapiro.test(IND_NSL1$Total_diff)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  IND_NSL1$Total_diff
## W = 0.43106, p-value < 2.2e-16
```

```
shapiro.test(IND_NSL1$Females_diff)
```
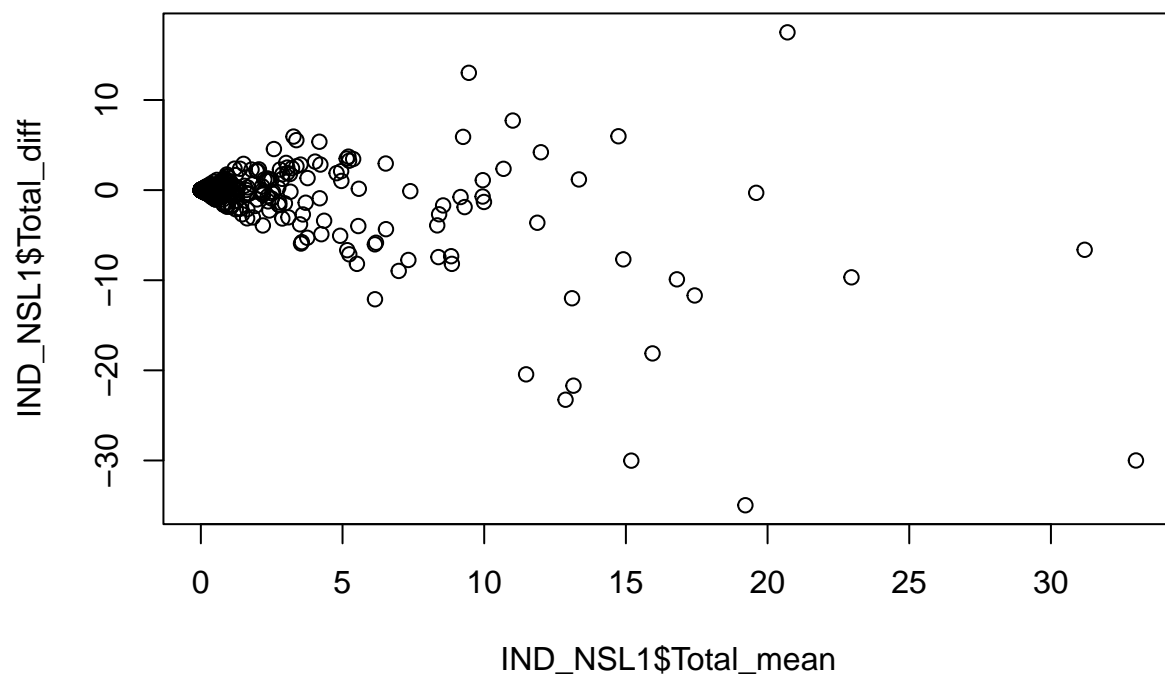
```
##
##  Shapiro-Wilk normality test
##
## data:  IND_NSL1$Females_diff
## W = 0.31269, p-value < 2.2e-16
```

Total mobiles

```
plot(IND_NSL1$Total_mean,IND_NSL1$Total_diff)
```

```r
# Model the mean difference as a function mean
fit1 <- lm(Total_diff ~ Total_mean, data = IND_NSL1)
b <- coef(fit1)
summary(fit1)
```

```
##
## Call:
## lm(formula = Total_diff ~ Total_mean, data = IND_NSL1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.7924  -0.3048  -0.2255   0.0682  27.4087
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.26704    0.13146   2.031   0.0427 *
## Total_mean  -0.49131    0.03412 -14.399   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.992 on 607 degrees of freedom
## Multiple R-squared:  0.2546, Adjusted R-squared:  0.2534
## F-statistic: 207.3 on 1 and 607 DF,  p-value: < 2.2e-16
```

```r
# Model heteroscedasticity
residuals1 <- abs(resid(fit1))
fit2 <- lm(residuals1 ~ Total_mean, data = IND_NSL1)
summary(fit2)
```

```
##
## Call:
## lm(formula = residuals1 ~ Total_mean, data = IND_NSL1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.4517  -0.2859  -0.0688   0.0073  14.9503
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.28272    0.07890   3.583 0.000366 ***
## Total_mean   0.58813    0.02048  28.720  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.796 on 607 degrees of freedom
## Multiple R-squared:  0.5761, Adjusted R-squared:  0.5754
## F-statistic: 824.9 on 1 and 607 DF,  p-value: < 2.2e-16
```

```r
c <- coef(fit2)

# Calculation of LoA: limits of agreement
A_grid <- seq(min(IND_NSL1$Total_mean), max(IND_NSL1$Total_mean), length.out = 200)
```

```
grid <- data.frame(Total_mean = A_grid)
grid$Dhat <- b[1] + b[2] * grid$Total_mean
grid$LoA_lower <- grid$Dhat - 2.46 * (c[1] + c[2] * grid$Total_mean)
grid$LoA_upper <- grid$Dhat + 2.46 * (c[1] + c[2] * grid$Total_mean)
```

The differences between sea lice counts recorded by the Industry and by the NSLMP were visualised using the Bland-Altmand plots, separately for total mobiles and total adult females.

```
TM<-ggplot(IND_NSL1, aes(x =Total_mean, y = Total_diff)) +
  geom_point(size = 1) +
  geom_line(data = grid, aes(x = Total_mean, y = Dhat), color = "blue", size = 1.2) +
  geom_line(data = grid, aes(x = Total_mean, y = LoA_lower), color = "red",
            linetype = "dashed", size = 1) +
  geom_line(data = grid, aes(x = Total_mean, y = LoA_upper), color = "red",
            linetype = "dashed", size = 1) +
  labs(title='Total mobiles (TM)',x = "Mean (NSLMP+industry)",
       y = "Difference (NSLMP-industry)") +
  theme_minimal(base_size = 14)
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

Total adult females

```
plot(IND_NSL1$Females_mean,IND_NSL1$Females_diff)
```
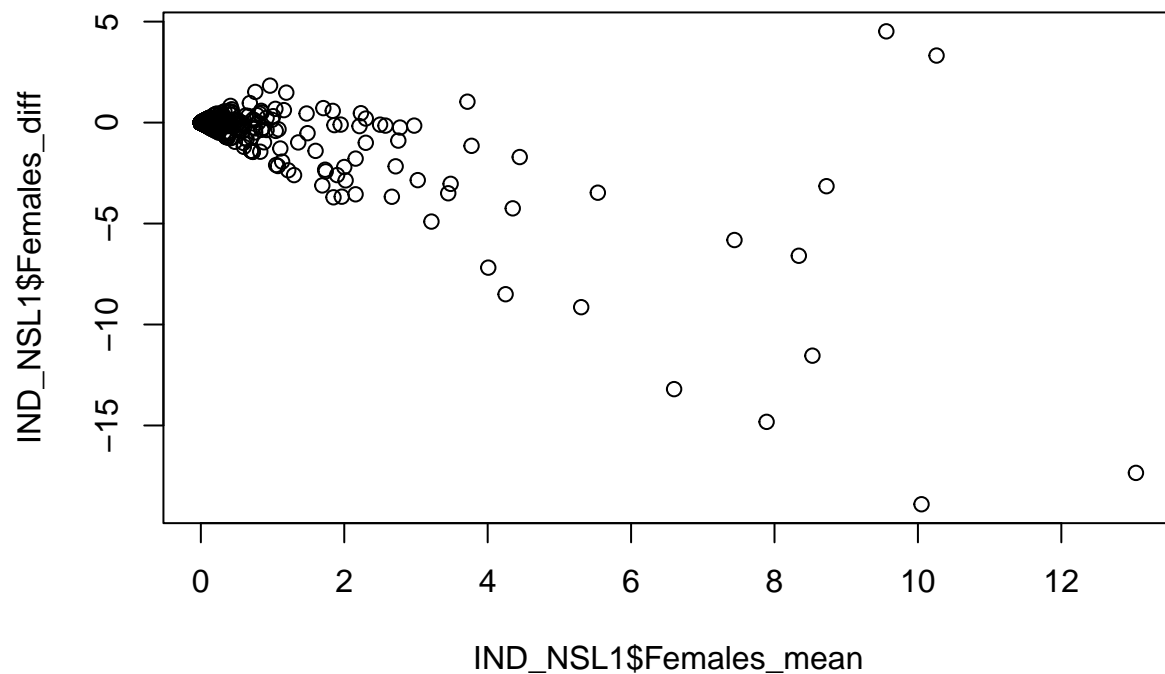
```r
# Model the mean difference as a function mean
fit1 <- lm(Females_diff ~ Females_mean, data = IND_NSL1)
b <- coef(fit1)
summary(fit1)
```

```
##
## Call:
## lm(formula = Females_diff ~ Females_mean, data = IND_NSL1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.0238  -0.0903  -0.0903   0.0425  12.9564
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.09035    0.05044   1.791   0.0738 .
## Females_mean -0.89219    0.03553 -25.114   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.177 on 607 degrees of freedom
## Multiple R-squared:  0.5096, Adjusted R-squared:  0.5088
## F-statistic: 630.7 on 1 and 607 DF,  p-value: < 2.2e-16
```

```r
# Model heteroscedasticity
residuals1 <- abs(resid(fit1))
fit2 <- lm(residuals1 ~ Females_mean, data = IND_NSL1)
summary(fit2)
```

```
##
## Call:
## lm(formula = residuals1 ~ Females_mean, data = IND_NSL1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0923 -0.0585  0.0195  0.0299  6.2640
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.07087    0.02587   2.739  0.00634 **
## Females_mean  0.69270    0.01822  38.021  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6037 on 607 degrees of freedom
## Multiple R-squared:  0.7043, Adjusted R-squared:  0.7038
## F-statistic:  1446 on 1 and 607 DF,  p-value: < 2.2e-16
```

```r
c <- coef(fit2)

# Calculation of LoA: limits of agreement
A_grid <- seq(min(IND_NSL1$Females_mean), max(IND_NSL1$Females_mean), length.out = 200)
grid <- data.frame(Females_mean = A_grid)
grid$Dhat <- b[1] + b[2] * grid$Females_mean
grid$LoA_lower <- grid$Dhat - 2.46 * (c[1] + c[2] * grid$Females_mean)
grid$LoA_upper <- grid$Dhat + 2.46 * (c[1] + c[2] * grid$Females_mean)
```

The differences between sea lice counts recorded by the Industry and by the NSLMP were visualised using the Bland-Altmand plots, separately for total mobiles and total adult females.

```r
TF<-ggplot(IND_NSL1, aes(x =Females_mean, y = Females_diff)) +
  geom_point(size = 1) +
  geom_line(data = grid, aes(x = Females_mean, y = Dhat), color = "blue", size = 1.2) +
  geom_line(data = grid, aes(x = Females_mean, y = LoA_lower), color = "red",
            linetype = "dashed", size = 1) +
  geom_line(data = grid, aes(x = Females_mean, y = LoA_upper), color = "red",
            linetype = "dashed", size = 1) +
  labs(title='Total adult females (TF)',x = "Mean (NSLMP+industry)",
       y = "Difference (NSLMP-industry)") +
  theme_minimal(base_size = 14)
```

The graphs illustrating the differences in reported sea lice counts by the Industry and NSLMP were combined into a single plot.

```r
grid.arrange(TM,TF,ncol = 2)
```
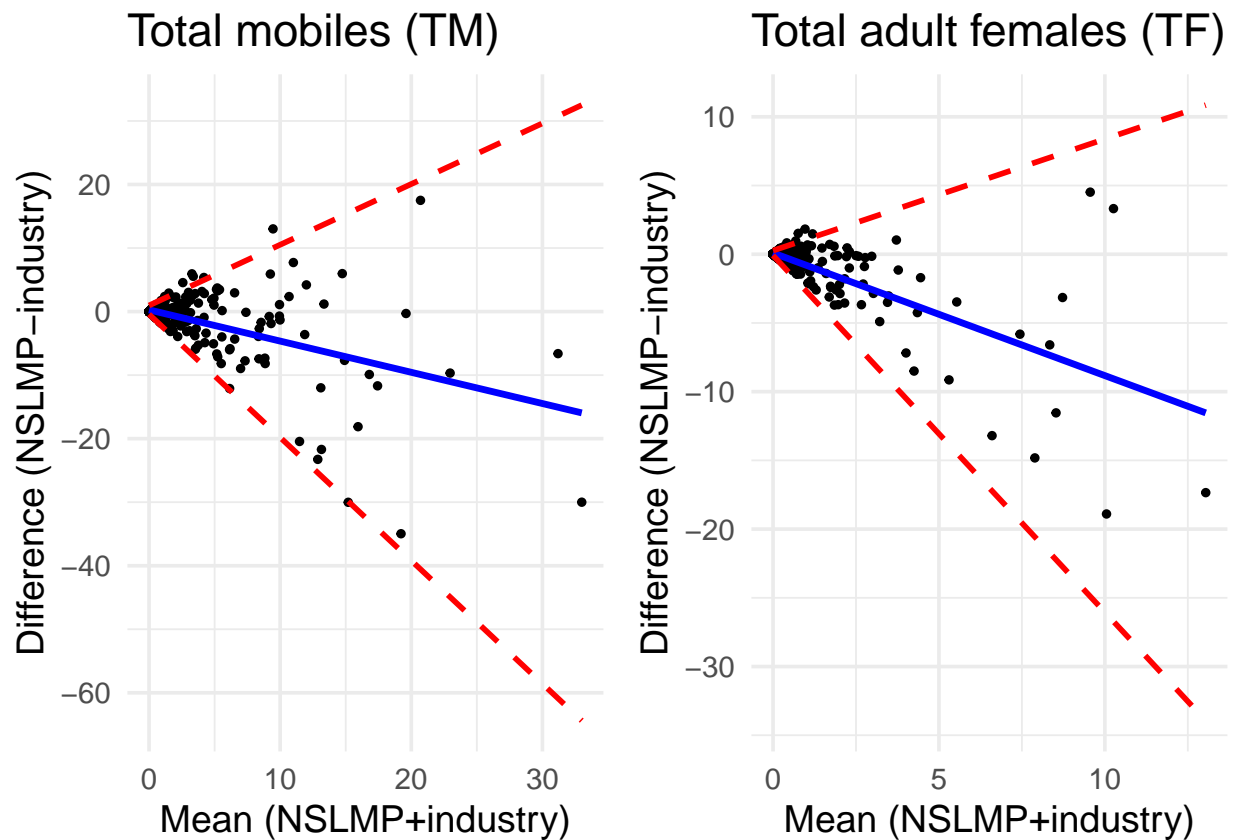

Testing of differences between sampling

The Shapiro-Wilk test was used to test the normality of the distribution between the differences of Industry and NSLMP for total mobiles and total adult females, separately.

Total mobiles

```r
shapiro.test(IND_NSL1$Total_diff)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  IND_NSL1$Total_diff
## W = 0.43106, p-value < 2.2e-16
```

The results of the Shapiro-Wilk test showed the differences between the two groups of sea lice sampling significantly deviated from a normal distribution. The distribution was visualised to determine the symmetricity of the data and identify possible outliers.

The non-parametric Wilcoxon matched pairs test was used to compare the differences in medians between the two groups of reported sea lice counts.

```
wilcox.test(IND_NSL1$Mowi_Total,IND_NSL1$MI_Total, paired = TRUE,
            alternative = "two.sided",conf.int = TRUE,detailed=TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  IND_NSL1$Mowi_Total and IND_NSL1$MI_Total
## V = 76772, p-value = 0.069
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -0.001886159  0.067820210
## sample estimates:
## (pseudo)median
##      0.02993277
```

Total adult females

```
shapiro.test(IND_NSL1$Females_diff)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  IND_NSL1$Females_diff
## W = 0.31269, p-value < 2.2e-16
```

The results of the Shapiro-Wilk test showed the differences between the two groups of sea lice sampling significantly deviated from a normal distribution. The distribution was visualised to determine the symmetricity of the data and identify possible outliers.

```
wilcox.test(IND_NSL1$Mowi_Females,IND_NSL1$MI_Females, paired = TRUE,
            alternative = "two.sided",conf.int = TRUE,detailed=TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  IND_NSL1$Mowi_Females and IND_NSL1$MI_Females
## V = 51616, p-value = 1.86e-06
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  0.03335807 0.09193496
## sample estimates:
## (pseudo)median
##      0.06003316
```

wilcox.test(IND_NSL1$Mowi_Females, IND_NSL1$MI_Females$, paired = TRUE, alternative = "greater", conf.int = TRUE,detailed=TRUE)

The correlation between the differences in dates of sampling and the differences in sea lice counts between the industry and NSLMP was determined separately for total mobiles and total adult females.

```
# Total mobiles
cor.test(as.numeric(abs(IND_NSL1$Day_diff)),abs(IND_NSL1$Total_diff),method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  as.numeric(abs(IND_NSL1$Day_diff)) and abs(IND_NSL1$Total_diff)
## t = 3.7629, df = 607, p-value = 0.0001842
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.0723989 0.2276996
## sample estimates:
##       cor
## 0.1509807
```

```
# Total adult females
cor.test(as.numeric(abs(IND_NSL1$Day_diff)),abs(IND_NSL1$Females_diff),method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  as.numeric(abs(IND_NSL1$Day_diff)) and abs(IND_NSL1$Females_diff)
## t = 3.6353, df = 607, p-value = 0.0003012
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.06730395 0.22283954
## sample estimates:
##       cor
## 0.1459737
```

Influence of treatment trigger levels on the accuracy of the reported sea lice counts

Industry data

To explore the influence of treatment trigger levels on the accuracy of the reported sea lice counts reported by the Industry, we used the pen-level mean abundance datasets (NSLMP2, Industry2). The treatment trigger levels were visualised by a red line for the Industry data.

To improve clarity, abundance values greater than 2 for both NSLMP and Industry data were excluded from the final graphs.

```
NSLMP4<-NSLMP2[NSLMP2$Females<2.05,]
nrow(NSLMP2[!(NSLMP2$Females<2.05),])/nrow(NSLMP2)
```

```
## [1] 0.03776435
```

```
Industry4<-Industry2[Industry2$Females<2.05,]
nrow(Industry2[!(Industry2$Females<2.05),])/nrow(Industry2)
```

```
## [1] 0.08980454
```

NSLMP

```
NSL<-ggplot(NSLMP4,aes(Females))  +
    geom_histogram(binwidth = 0.1, color="black",fill="grey") +
    scale_y_sqrt(limits = c(0,3500))+
    ylab('Frequency')+
    annotate('text',x=1.75,y=4000,label='Limit = 0.2',size=4)+
    geom_vline(xintercept = 0.26,color='red',linewidth=1.9)+
    ggtitle('NSLMP')+
    xlab('Total adult females (TF) abundance')+
    theme_minimal()+
    theme(legend.position="none",
        text = element_text(size =12))
```

Industry

```
Ind<-ggplot(Industry4,aes(Females))  +
    geom_histogram(binwidth = 0.1,color="black",fill="grey") +
    scale_y_sqrt(limits = c(0,3500))+
    xlab('Total adult females (TF) abundance')+
    ggtitle('Industry producer')+
    annotate('text',x=1.75,y=4000,size=4,label='Limit = 0.2')+
    geom_vline(xintercept = 0.25,color='red',linewidth=1.9)+
    theme_minimal()+
    theme(legend.position="none",
        text = element_text(size =12),
        axis.title.y=element_blank())
```

The graphs illustrating the frequency of reported sea lice levels per pen for both NSLMP and Industry were combined into a single plot.

```
grid.arrange(NSL,Ind,ncol = 2)
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_text()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_text()').
```
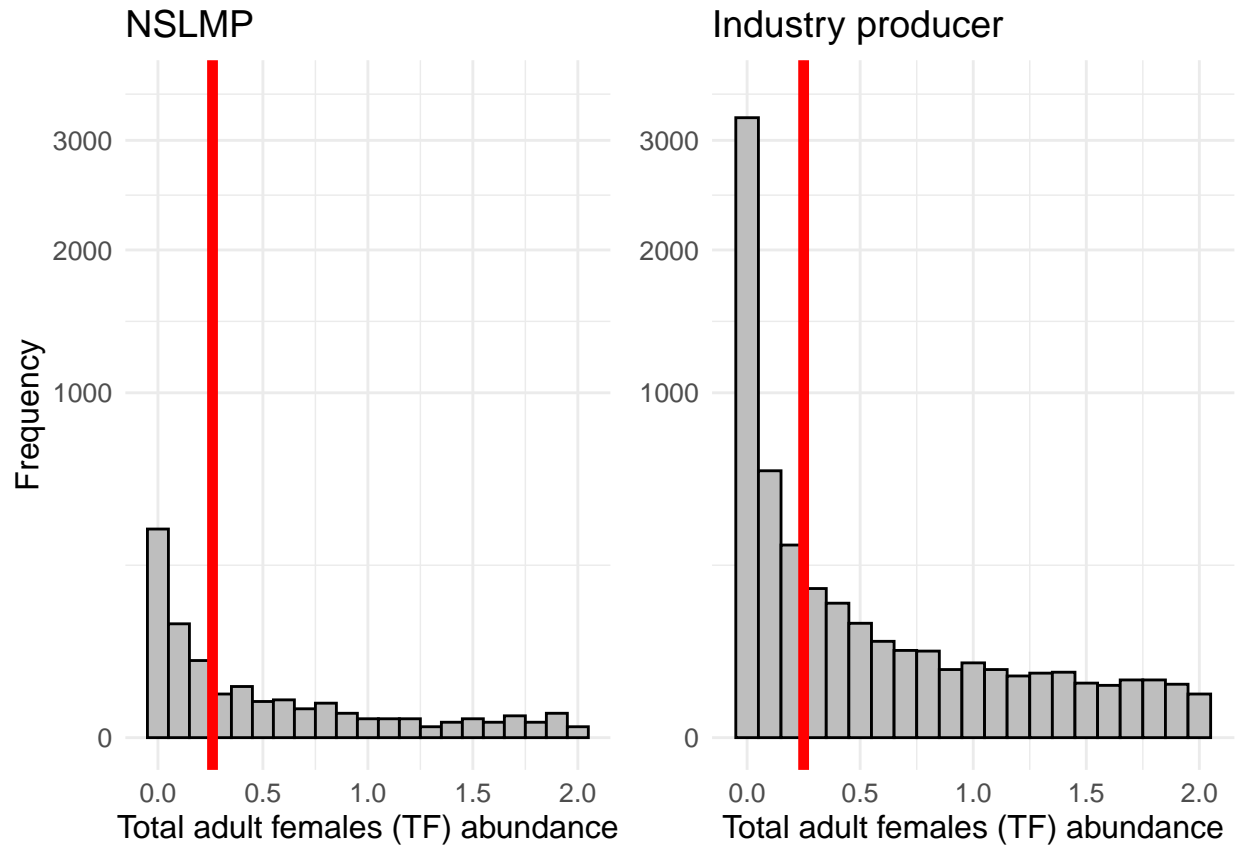
Location of salmon farms visualization

The coordinates of all salmon farms were extracted from the NSLMP dataset and farms that were monitored by both the Marine Insitute and by the Industry were identified.

```
Coordinates<- data.frame(Longitude = c(unique(NSLMP$Longitude)),
              Latitude = c(unique(NSLMP$Latitude)),
              Site=c(unique(NSLMP$SiteName)))
Coordinates$Monitoring<-'NSLMP'
Coordinates[Coordinates$Site %in% unique(Industry$SiteName),]$Monitoring<-'NSLMP and Industry'
```

The resulting dataset was then use to visualize the location of each salmon farm on the map of Ireland.

```
world <- ne_countries(scale = 10, returnclass = "sf")

ggplot(data = world) +
    geom_sf() +
    geom_point(data=Coordinates,aes(x=Longitude,y=Latitude,fill=as.factor(Monitoring)),
              size =4,shape = 23, alpha=0.8) +
    coord_sf(xlim = c(-10.5,-5.35), ylim = c(51.3,55.4), expand = FALSE) +
    theme_minimal()+
    scale_fill_manual(name='Monitoring',
              labels=c("NSLMP","NSLMP and
                    Industry producer"),
```

```
                   values=c('darkred','midnightblue'))+
    theme(text = element_text(size = 15))
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x9
```

```
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0x9
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0x9
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0x9
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0x9
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0x9
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0x9
```