

Analysis of NSLMP dataset

Bela Klimesova

2025-04-10

Preparation of data for analysis

```
library(ggplot2)
library(gridExtra)
library(grid)
```

Data loading (NSLMP)

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

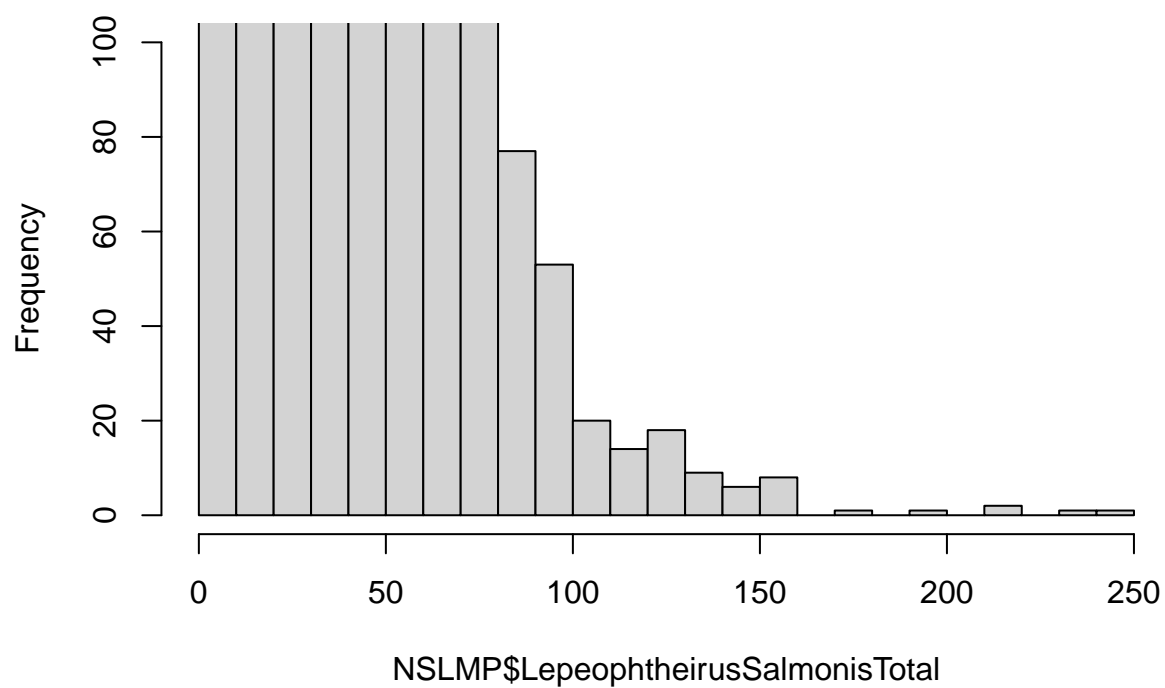
Exploration of the raw dataset and excluding outliers from analysis (3 datapoints)

```
nrow(NSLMP)
```

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

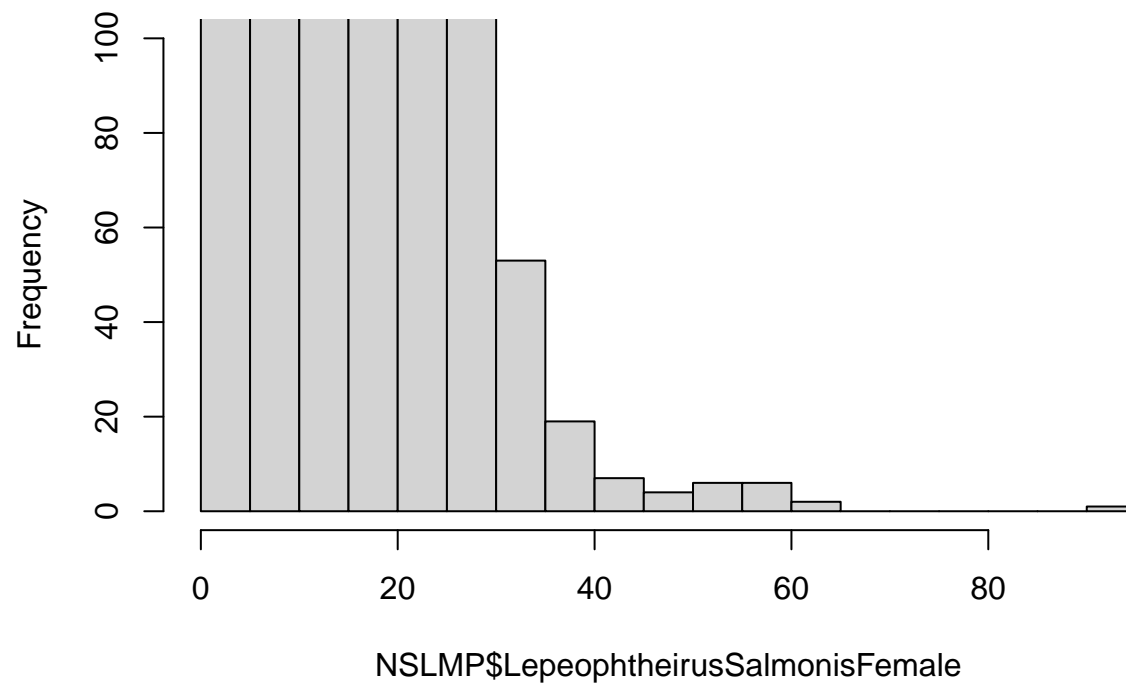
```
hist(NSLMP$LepeophtheirusSalmonisTotal,ylim=c(0,100))
```

Histogram of NSLMP\$LepeophtheirusSalmonisTotal



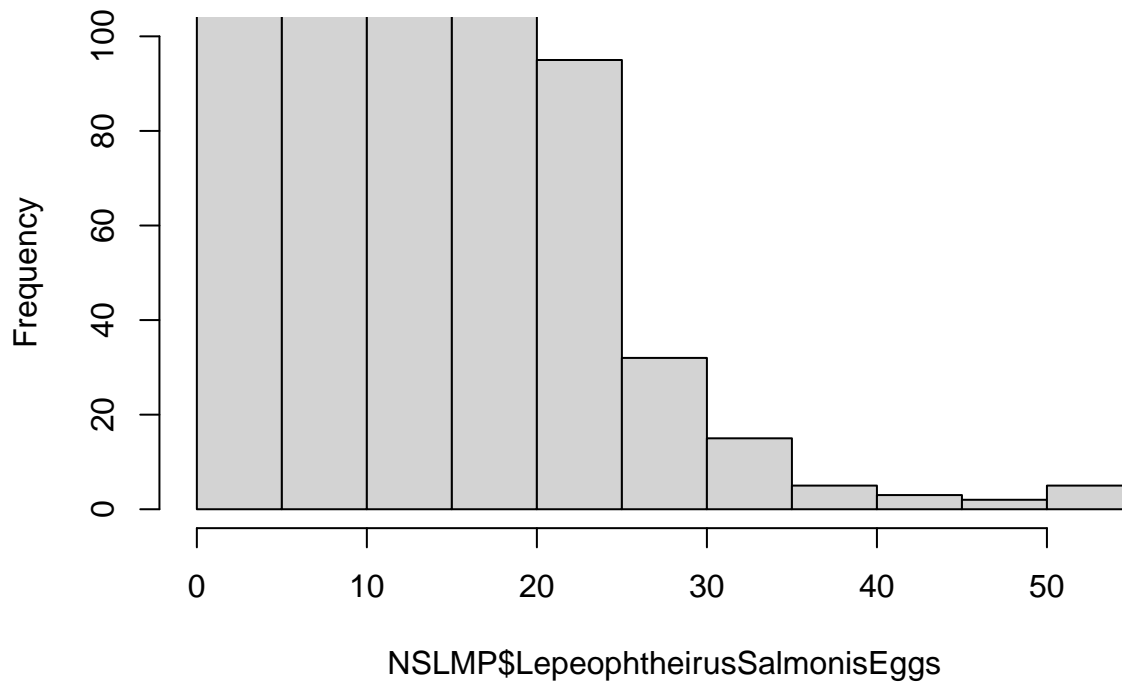
```
NSLMP<-NSLMP[NSLMP$LepeophtheirusSalmonisTotal<160,]  
hist(NSLMP$LepeophtheirusSalmonisFemale,ylim=c(0,100))
```

Histogram of NSLMP\$LepeophtheirusSalmonisFemale



```
NSLMP<-NSLMP[NSLMP$LepeophtheirusSalmonisFemale<80,]  
hist(NSLMP$LepeophtheirusSalmonisEggs,ylim=c(0,100))
```

Histogram of NSLMP\$LepeophtheirusSalmonisEggs



```
nrow(NSLMP)
```

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

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

```
## [1] 39
```

Calculation of dispersion index

Ovigerous females

```
# Overall dispersion index
```

```
var(NSLMP$LepeophtheirusSalmonisEggs)/mean(NSLMP$LepeophtheirusSalmonisEggs)
```

```
## [1] 6.381062
```

```
# Graph illustrating the relationship between abundance and dispersion
```

```
A<-tapply(NSLMP$LepeophtheirusSalmonisEggs,paste(NSLMP$Year,NSLMP$SiteName),mean)
```

```
B<-tapply(NSLMP$LepeophtheirusSalmonisEggs,paste(NSLMP$Year,NSLMP$SiteName),  
          function(x){var(x)/mean(x)})
```

```
C<-data.frame(ID=names(A),Mean=as.vector(A),Disp=as.vector(B))
```

```
C[C$Disp=='NaN',]$Disp<-0
```

```
Eggs<-ggplot(C,aes(Mean,Disp))+
  geom_point()+
  labs(title='Ovigerous females',x='Mean abundance',
        y='Index of overdispersion')+
  theme_minimal()+
  geom_smooth(method='lm')
```

Total adult females

```
# Overall dispersion index
var(NSLMP$LepeophtheirusSalmonisFemale)/mean(NSLMP$LepeophtheirusSalmonisFemale)
```

```
## [1] 7.887911
```

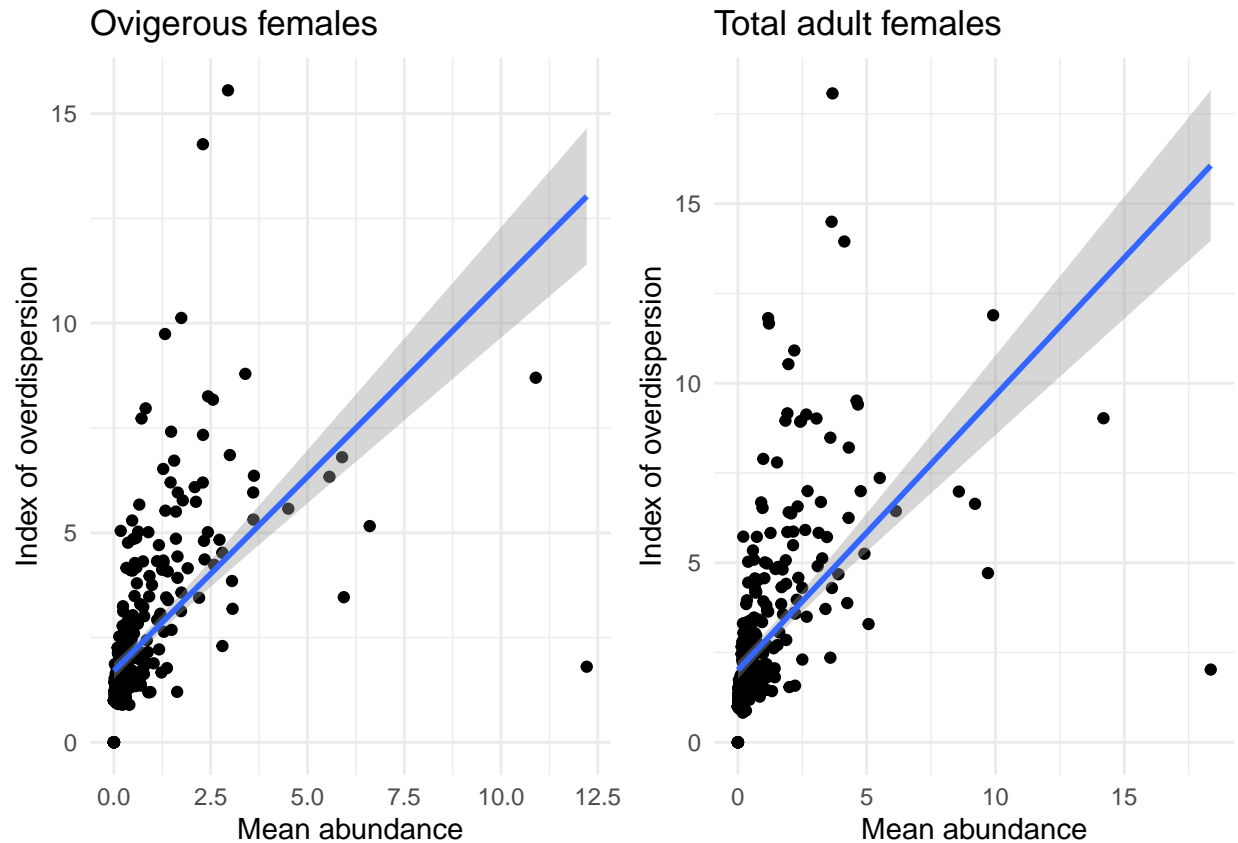
```
A<-tapply(NSLMP$LepeophtheirusSalmonisFemale,paste(NSLMP$Year,NSLMP$SiteName),mean)
B<-tapply(NSLMP$LepeophtheirusSalmonisFemale,paste(NSLMP$Year,NSLMP$SiteName),
  function(x){var(x)/mean(x)})
C<-data.frame(ID=names(A),Mean=as.vector(A),Disp=as.vector(B))
C[C$Disp=='NaN',]$Disp<-0

# Graph illustrating the relationship between abundance and dispersion
Females<-ggplot(C,aes(Mean,Disp))+
  geom_point()+
  labs(title='Total adult females',x='Mean abundance',
        y='Index of overdispersion')+
  theme_minimal()+
  geom_smooth(method='lm')
```

Graph of abundance and dispersion for both ovigerous and total adult females

```
grid.arrange(Eggs,Females,ncol = 2)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
## 'geom_smooth()' using formula = 'y ~ x'
```



All analyses focused on the average sea lice count per pen, either for ovigerous females or total adult females. Therefore, the datasets were adjusted to the pen level by calculating the mean of both abundance and prevalence per Sample ID and retaining only the necessary variables.

Ovigerous females

```
OF1<-aggregate(NSLMP$LepeophtheirusSalmonisEggs,list(NSLMP$SampleID,NSLMP$Month),mean)
colnames(OF1)[3]<-'AbunOF'
OF2<-aggregate(NSLMP$Prev_Eggs,list(NSLMP$SampleID,NSLMP$Month),mean)
colnames(OF2)[3]<-'PrevOF'
OF<-merge(OF1,OF2,by=c('Group.1','Group.2'),all=TRUE)
colnames(OF)[c(1:2)]<-c('ID','Month')
```

Total adult females

```
TF1<-aggregate(NSLMP$LepeophtheirusSalmonisFemale,list(NSLMP$SampleID,NSLMP$Month),mean)
colnames(TF1)[3]<-'AbunTF'
TF2<-aggregate(NSLMP$Prev_TF,list(NSLMP$SampleID,NSLMP$Month),mean)
colnames(TF2)[3]<-'PrevTF'
TF<-merge(TF1,TF2,by=c('Group.1','Group.2'),all=TRUE)
colnames(TF)[c(1:2)]<-c('ID','Month')
```

The ovigerous female and total female sea lice datasets were merged into a new combined dataset.

```
OF_TF<-merge(OF,TF,by=c('ID','Month'),all=TRUE)
```

Prevalence as a metric for sea lice infestation

To assess the suitability of prevalence as a metric for sea lice infestation levels, we analyzed the relationship between prevalence and abundance. The mean abundance was plotted against the mean prevalence separately for ovigerous and total adult female sea lice. Abundance values higher than 4 were excluded from the final graphs to improve clarity.

Ovigerous females

```
Graph_OF<-ggplot(OF_TF, aes(x = AbunOF, y =PrevOF)) +  
  geom_point(size=0.3,alpha=0.4)+  
  coord_cartesian(xlim =c(0,4))+  
  xlab('Abundance')+  
  ylab('Prevalence')+  
  theme_minimal()+  
  ggtitle('Ovigerous females (OF)')+  
  theme(text = element_text(size =12))+  
  stat_smooth(method = "gam",  
              formula = y ~ s(x, bs = "cs"),  
              linewidth=1.2) # +  
  # geom_vline(xintercept =0.5)+  
  # geom_vline(xintercept =2)+  
  # geom_hline(yintercept =0.37)
```

Total adult females

```
Graph_TF<-ggplot(OF_TF, aes(x=AbunTF,y=PrevTF)) +  
  geom_point(size=0.3,alpha=0.4)+  
  coord_cartesian(xlim =c(0,4))+  
  xlab('Abundance')+  
  theme_minimal()+  
  ggtitle('Total adult females (TF)')+  
  theme(text = element_text(size =12),  
        axis.title.y=element_blank())+  
  stat_smooth(method = "gam",  
              formula = y ~ s(x, bs = "cs"),  
              linewidth=1.2) # +  
  # geom_vline(xintercept =0.6)+  
  # geom_vline(xintercept =1.75)+  
  # geom_hline(yintercept =0.42)
```

The total number of observations in the dataset and proportion of discarded values (abundance higher than 4) was determined.

```
nrow(OF_TF)
```

```
## [1] 6098
```

```
# Ovigerous females
round(nrow(OF_TF[OF_TF$AbunOF>4,])/nrow(OF_TF),3)
```

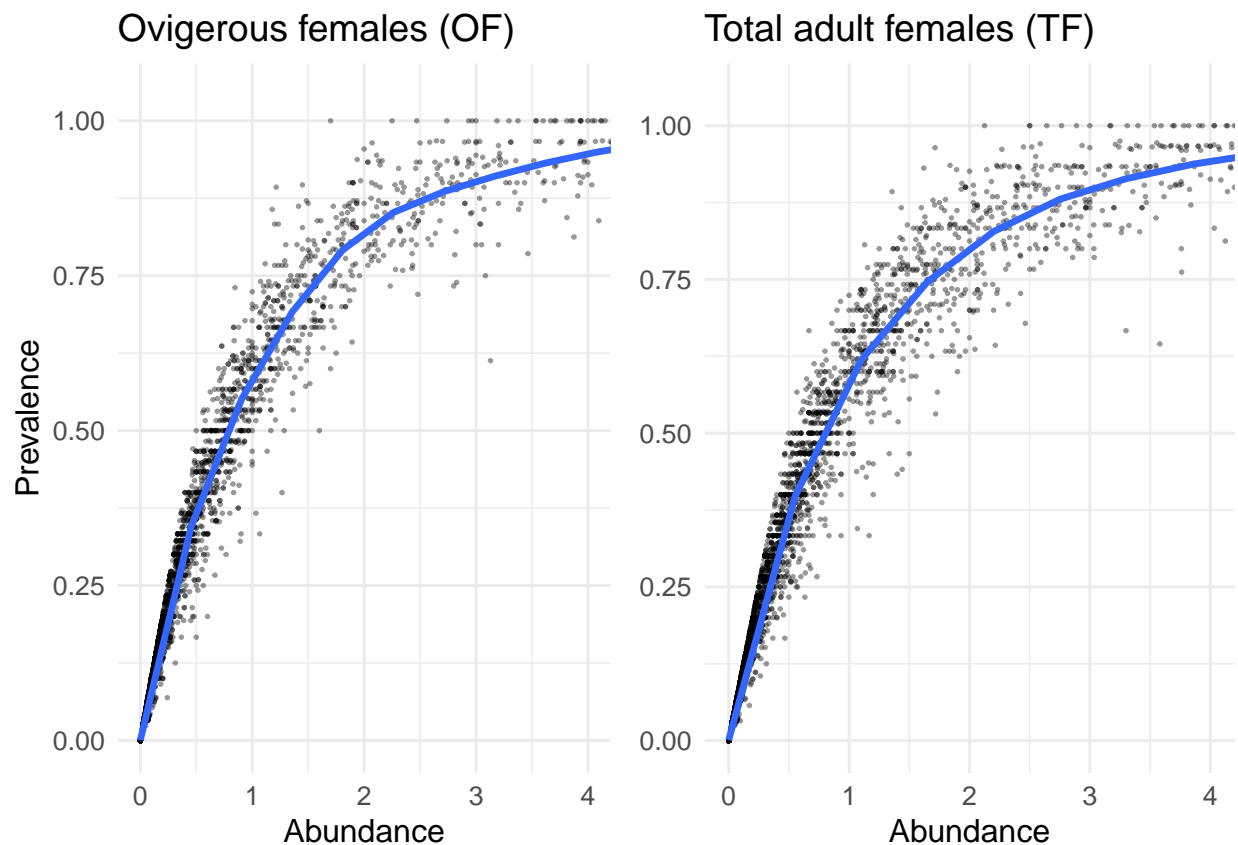
```
## [1] 0.036
```

```
# Total adult females
round(nrow(OF_TF[OF_TF$AbunTF>4,])/nrow(OF_TF),3)
```

```
## [1] 0.054
```

The graphs illustrating the relationship between prevalence and abundance for both ovigerous and total adult female sea lice were combined into a single plot.

```
grid.arrange(Graph_OF,Graph_TF,ncol = 2)
```



Relationship between ovigerous females and total adult females

For the exploration of the relationship between ovigerous females and total adult females, the dataset of mean abundance at pen level (OF_TF) was used. Two separate linear regression models were used to examine the relationship between ovigerous females and total adult females. The fitted regression equations were determined and used to calculate the corresponding treatment trigger levels between life stages. The treatment trigger levels were calculated separately for the spring period (0.5) and summer to winter (2).

Ovigerous females


```
lm_TF_OF<-lm(AbunTF~AbunOF,data=OF_TF)
lm_TF_OF
```

```
##
## Call:
## lm(formula = AbunTF ~ AbunOF, data = OF_TF)
##
## Coefficients:
## (Intercept)      AbunOF
##      0.07369      1.33843
```

```
# Spring
round(( 1.33954*0.5) + 0.07362,2)
```

```
## [1] 0.74
```

```
# Summer to winter
round(( 1.33954*2) + 0.07362,2)
```

```
## [1] 2.75
```

Total adult females

```
lm_OF_TF<-lm(AbunOF~AbunTF,data=OF_TF)
lm_OF_TF
```

```
##
## Call:
## lm(formula = AbunOF ~ AbunTF, data = OF_TF)
##
## Coefficients:
## (Intercept)      AbunTF
##      -0.02481      0.71146
```

```
# Spring
round((0.71039*0.5) -0.02424,2)
```

```
## [1] 0.33
```

```
# Summer to winter
round((0.71039*2) -0.02424,2)
```

```
## [1] 1.4
```

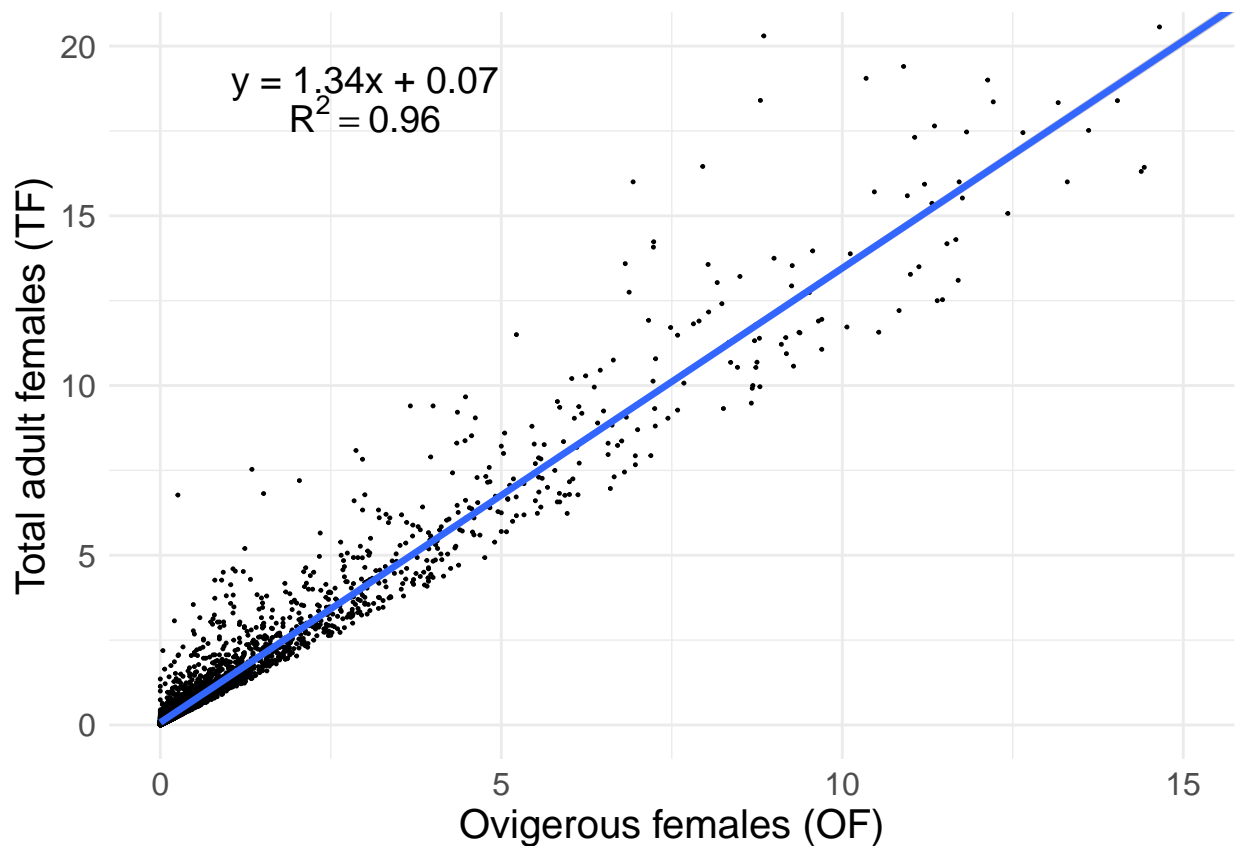
The relationship between the abundance of ovigerous females and total adult females was visualised in a graph. Abundance values higher than 15 for ovigerous females and 20 for total adult females were excluded from the final graphs to improve clarity.

```

newdata1<-predict(lm_TF_OF,interval = "confidence")
mydata <- cbind(OF_TF,newdata1)

ggplot(mydata,aes(AbunOF,AbunTF)) +
  geom_ribbon(aes(ymin=lwr,ymax=upr),colour='grey80',fill= "grey80", alpha = 0.5)+
  geom_point(size=0.2) +
  coord_cartesian(xlim =c(0,15),ylim=c(0,20))+
  ylab('Total adult females (TF)')+
  xlab('Ovigerous females (OF)')+
  annotate('text',x=3,y=19,size=5,label='y = 1.34x + 0.07')+
  annotate('text',x=3,y=18,size=5,label="R^2==0.96",parse=TRUE)+
  stat_smooth(method = lm,formula = 'y ~ x', se = TRUE,linewidth=1.2)+
  theme_minimal()+
  theme(text = element_text(size = 15))

```



The total number of observations in the dataset and proportion of discarded values (abundance higher than 15 OF, 20 TF) was determined.

```

nrow(OF_TF)

## [1] 6098

# Ovigerous females
round(nrow(OF_TF[OF_TF$AbunOF>15,])/nrow(OF_TF),3)

```

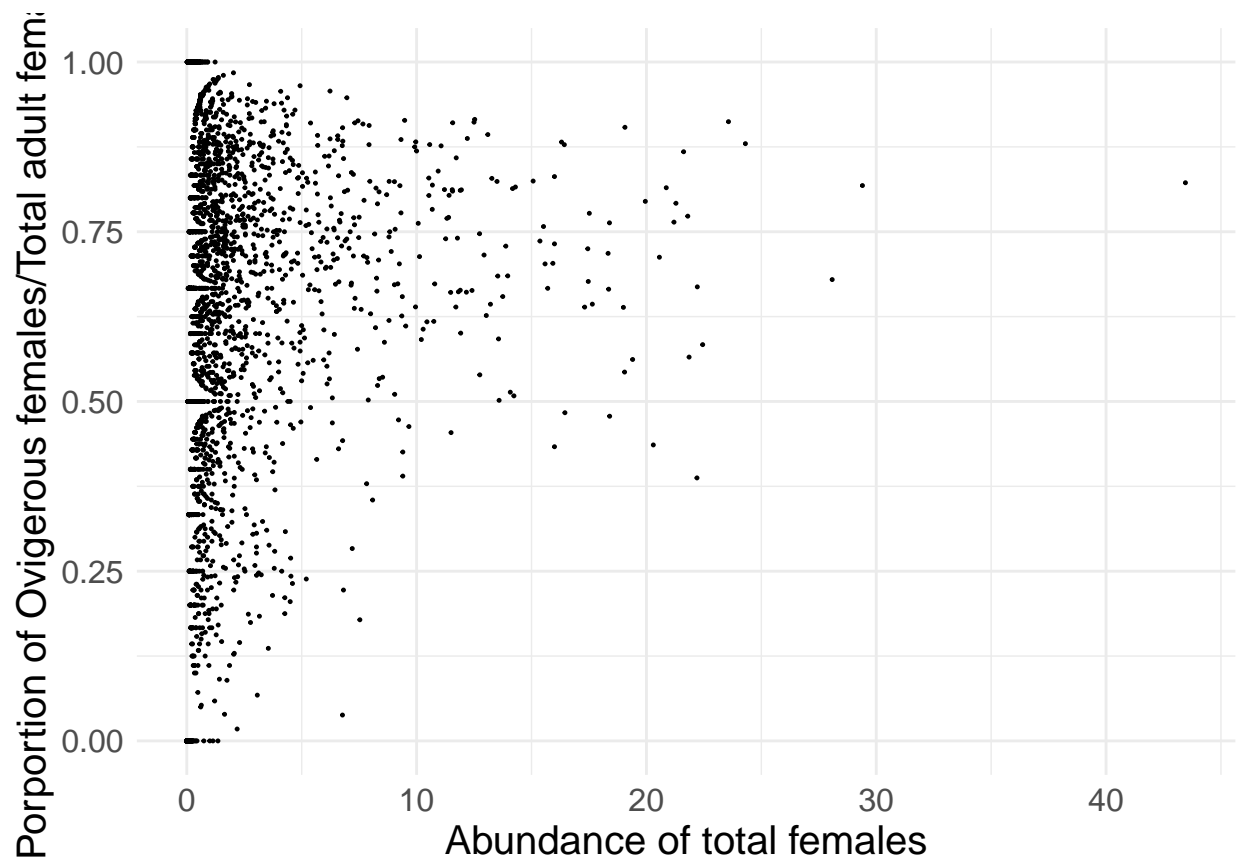
```
## [1] 0.002
```

```
# Total adult females  
round(nrow(OF_TF[OF_TF$AbunTF>20,])/nrow(OF_TF),3)
```

```
## [1] 0.003
```

Proportion between OF and TF

```
OF_TF$Abun_Prop<-OF_TF$AbunOF/OF_TF$AbunTF  
OF_TF[OF_TF$Abun_Prop=='NaN',]$Abun_Prop<-0  
  
ggplot(OF_TF,aes(AbunTF,Abun_Prop)) +  
  geom_point(size=0.2)+  
  xlab('Abundance of total females')+  
  ylab('Porportion of Ovigerous females/Total adult females')+  
  theme_minimal()+  
  theme(text = element_text(size = 15))
```



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

NSLMP

To explore the influence of treatment trigger levels on the accuracy of the reported sea lice counts, we used the pen-level mean abundance dataset (OF_TF). The dataset was split into two seasonal groups—spring, and summer to winter—reflecting the different treatment trigger levels applied in Ireland by the Marine Institute (MI).

```
# Spring
OF_TF_Spring<-OF_TF[OF_TF$Month %in% c(3,4,5),]
# Summer to winter
OF_TF_Winter<-OF_TF[!(OF_TF$Month %in% c(3,4,5)),]
```

To improve clarity, abundance values greater than 1.5 for spring and greater than 3 for summer to winter were excluded from the final graphs.

```
# Spring
OF_TF_Spring1<-OF_TF_Spring[OF_TF_Spring$AbunTF<1.5,]
OF_TF_Spring1$AbunOF<-round(OF_TF_Spring1$AbunOF,1)
# Summer to winter
OF_TF_Winter1<-OF_TF_Winter[OF_TF_Winter$AbunTF<2.5,]
OF_TF_Winter1$AbunTF<-round(OF_TF_Winter1$AbunTF,1)
```

Histograms of sea lice counts per pen for ovigerous and total adult females were plotted separately for spring and for summer to winter. The treatment trigger levels were visualised by a red line for ovigerous females.

```
# Ovigerous females, spring
OF_Spring<-ggplot(OF_TF_Spring1, aes(AbunOF)) +
  geom_histogram(binwidth = 0.1, color = "black", fill = "grey") +
  scale_y_sqrt(limits = c(0, 1750))+
  ggtitle('Ovigerous females (OF)')+
  ylab('Spring\nFrequency')+
  annotate('text',x=1.16,y=1600,label='Limit = 0.5',size=3)+
  geom_vline(xintercept = 0.55,color='red',linewidth=1.4)+ # intercept 0.55 to place the line on the
  theme_minimal()+
  theme(legend.position="none",
        text = element_text(size = 11),
        axis.title.x=element_blank())
```

```
# Total adult females, spring
TF_Spring<-ggplot(OF_TF_Spring1,aes(AbunTF)) +
  geom_histogram(color = "black",fill = "grey",binwidth = 0.1) +
  scale_y_sqrt(limits = c(0,1750))+
  ggtitle('Total adult females (TF)')+
  theme_minimal()+
  theme(legend.position="none",
        text = element_text(size = 11),
        axis.title.x=element_blank(),
        axis.title.y=element_blank())
```

```

# Ovigerous females, summer to winter
OF_Winter<-ggplot(OF_TF_Winter1,aes(AbunOF)) +
  geom_histogram(binwidth = 0.1, color="black",fill="grey") +
  scale_y_sqrt(limits = c(0,1750))+
  xlab('OF abundance')+
  ggtitle('')+
  ylab('Summer to Winter\nFrequency')+
  annotate('text',x=1.7,y=1600,label='Limit = 2',size=3)+
  geom_vline(xintercept =2.05,color='red',linewidth=1.4)+ # intercept 2.05 to place the line on the
  theme_minimal()+
  theme(legend.position="none",
        text = element_text(size = 11))

```

```

# Total adult females, summer to winter
TF_Winter<-ggplot(OF_TF_Winter1,aes(AbunTF)) +
  geom_histogram(binwidth = 0.1, color = "black",fill = "grey") +
  scale_y_sqrt(limits = c(0,1750))+
  xlab('TF abundance')+
  ggtitle('')+
  theme_minimal()+
  theme(legend.position="none",text = element_text(size =11),
        axis.title.y=element_blank())

```

The graphs illustrating the frequency of reported sea lice levels per pen for both ovigerous and total adult female sea lice were combined into a single plot.

```

grid.arrange(OF_Spring,TF_Spring,OF_Winter,TF_Winter,ncol = 2)

```

