

Solent Seals

Sarah Marley

21 July 2020

This code is supplementary to the manuscript “*Increasing numbers of harbour seals and grey seals in the Solent*”, which has been published in *Ecology & Evolution* (see: <https://doi.org/10.1002/ece3.8167>).

Summary: This study utilised a 20-year dataset of seal counts from two neighbouring harbours in the Solent region of southeast England. Generalised additive models (GAMs) showed a significant increase in the numbers of harbour and grey seals utilising Chichester Harbour. Conversely, in Langstone Harbour there has been a significant decrease in the number of harbour seals.

Packages

This code requires four packages:

```
# ggplot2 for data visualisation
library("ggplot2")

# gridExtra for arranging figures
library("gridExtra")

# Use mgcv for GAMs
library("mgcv")

# Use MuMIn for AICc
library("MuMIn")
```

Data

Visual surveys were undertaken in Chichester Harbour from 1999 to 2012 and 2015 to 2019, with surveys also undertaken in Langstone Harbour from 2009 to 2017 and 2019. Surveys were timed to approximately overlap with low tide, when the tidal sand bars and mud flats were at their maximum availability for seal haul-outs. However, in some cases surveys were limited by poor weather.

Counts of the total number of harbour and grey seals were collected. From 2015, the number of adults / juveniles / pups were also indicated. Note that there was no data collection in Langstone Harbour in 2018.

The supporting dataset is “*AllSealCounts.csv*”

```
# Read CSV from saved location on local drive
Seal <- read.csv("AllSealCounts.csv")

# Check first few rows
head(Seal)
```

```
##           Site Month Year Harbour Grey
## 1 Chichester      1 1999         1    0
## 2 Chichester      2 1999         2    0
## 3 Chichester      3 1999         4    0
## 4 Chichester      4 1999         6    0
## 5 Chichester      5 1999         5    0
## 6 Chichester      6 1999         5    0
```

The “*AllSealCounts.csv*” contains all counts for both species from both sites. But we want to do species-specific models from each site separately. So let’s subset the data to create a dataset for each site.

```
# Make data site-specific to Chichester Harbour
Ch.S <- Seal[ which(Seal$Site=='Chichester'),]

# Check first few rows
head(Ch.S)
```

```
##           Site Month Year Harbour Grey
## 1 Chichester      1 1999         1    0
## 2 Chichester      2 1999         2    0
## 3 Chichester      3 1999         4    0
## 4 Chichester      4 1999         6    0
## 5 Chichester      5 1999         5    0
## 6 Chichester      6 1999         5    0
```

```
# Make data site-specific to Langstone Harbour
La.S <- Seal[ which(Seal$Site=='Langstone'),]

# Check first few rows
head(La.S)
```

```
##           Site Month Year Harbour Grey
## 159 Langstone     2 2009         5    0
## 160 Langstone     3 2009         7    0
## 161 Langstone     4 2009         5    0
## 162 Langstone     5 2009         6    0
## 163 Langstone     6 2009         3    0
## 164 Langstone     7 2009         5    0
```

Data Exploration

In this project, we are interested in investigating temporal trends in seal counts. The first step is to visualise the data.

First, we’ll look at annual patterns:

```
# Harbour Seals by Year
H1 <- ggplot(data = Seal) +
  geom_point(mapping = aes(x = Year, y = Harbour, colour = Site)) +
  labs(x = "Year", y = "Numbers of Harbour Seals") +
  labs(title = "Harbour Seals by Year") +
  theme(legend.position="bottom") +
```

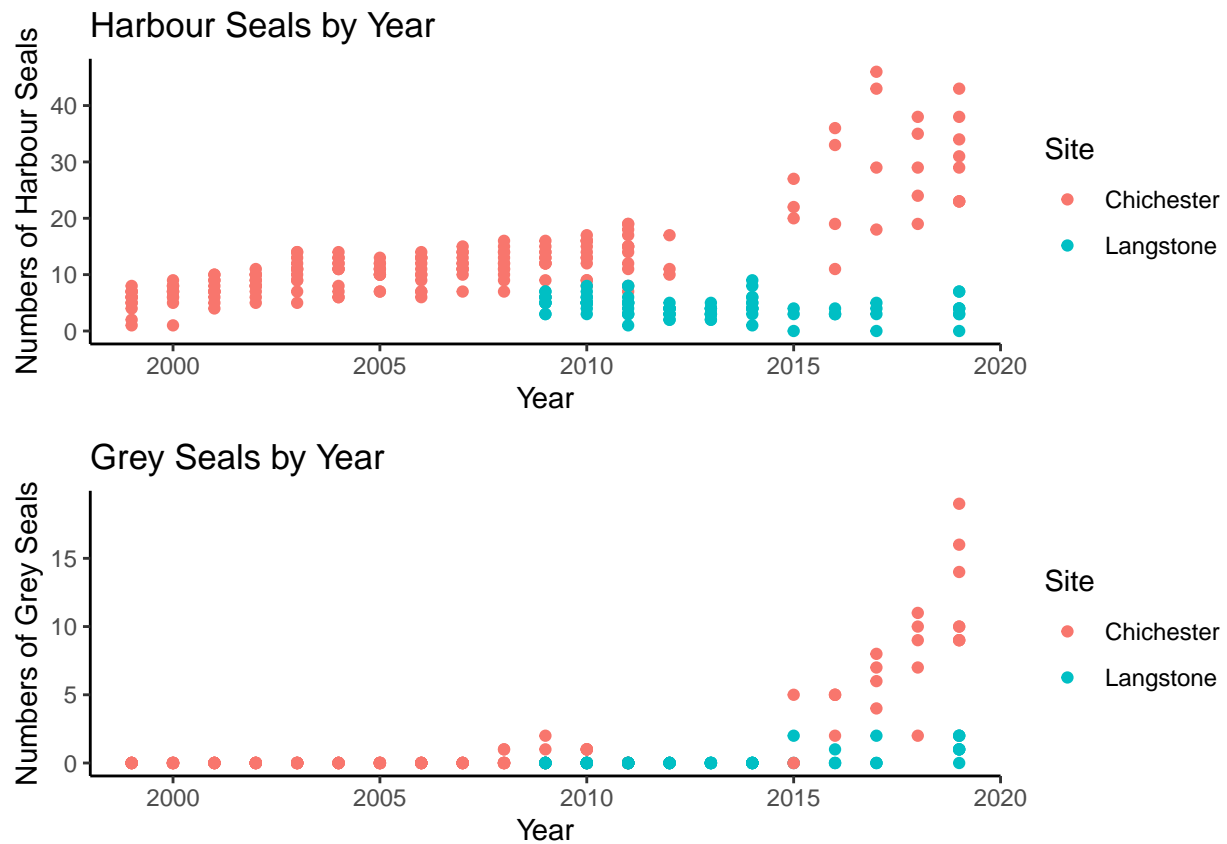
```

theme_classic()

# Grey Seals by Year
G1 <- ggplot(data = Seal) +
  geom_point(mapping = aes(x = Year, y = Grey, colour = Site)) +
  labs(x = "Year", y = "Numbers of Grey Seals") +
  labs(title = "Grey Seals by Year") +
  theme(legend.position="bottom") +
  theme_classic()

grid.arrange(H1, G1, nrow=2)

```



Now let's look at seasonal patterns:

```

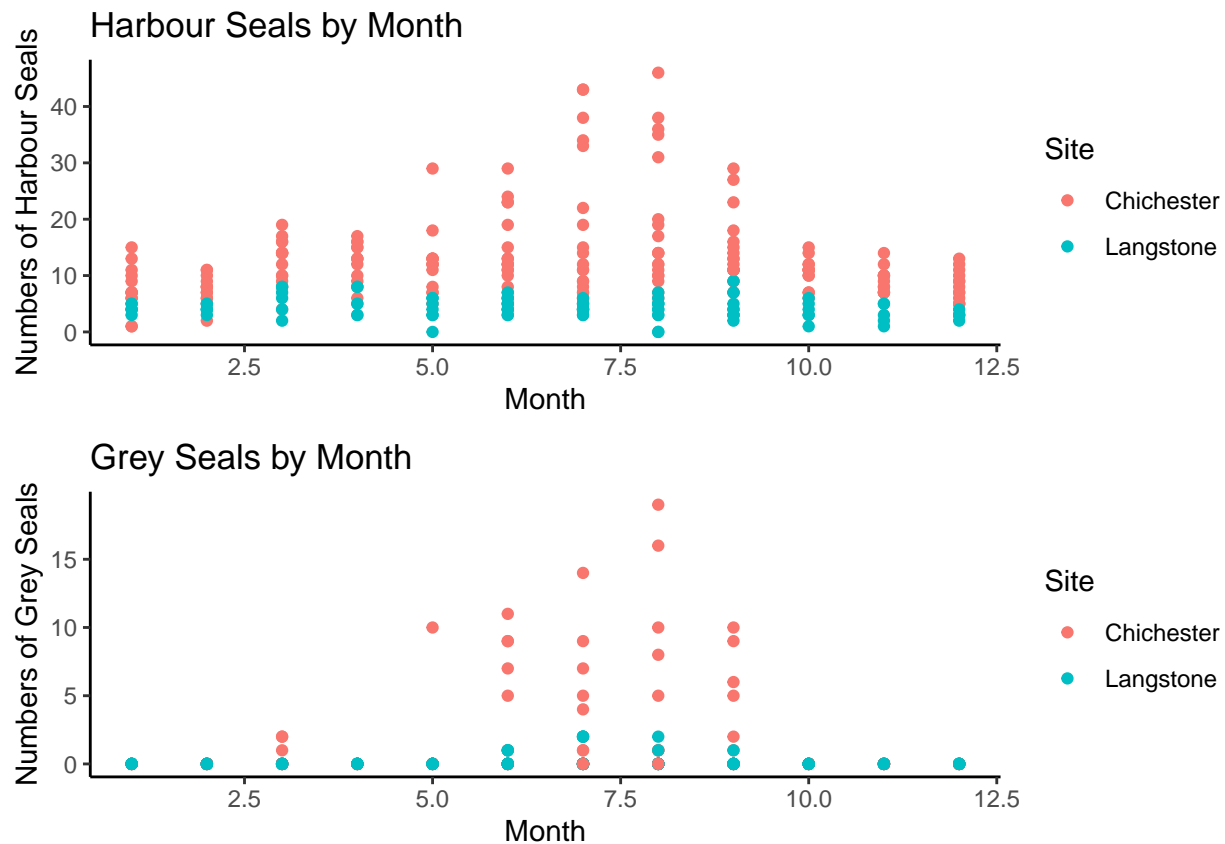
# Harbour Seals by Month
H2 <- ggplot(data = Seal) +
  geom_point(mapping = aes(x = Month, y = Harbour, colour = Site)) +
  labs(x = "Month", y = "Numbers of Harbour Seals") +
  labs(title = "Harbour Seals by Month") +
  theme_classic()

# Grey Seals by Month
G2 <- ggplot(data = Seal) +
  geom_point(mapping = aes(x = Month, y = Grey, colour = Site)) +
  labs(x = "Month", y = "Numbers of Grey Seals") +
  labs(title = "Grey Seals by Month") +

```

```
theme_classic()

grid.arrange(H2, G2, nrow=2)
```



It looks like there are some patterns worth investigating - but they don't look linear!

Build GAMs

Seal counts relative to temporal variables were assessed using separate Generalised Additive Models (GAMs) for each species and site. Exploratory analyses indicated that the temporal variables under consideration required fitting as non-linear terms; GAMs thus allowed smoothing functions to be fitted to these explanatory variables.

Four models were developed:

1. Harbour seals (Chichester Harbour)
2. Grey seals (Chichester Harbour)
3. Harbour seals (Langstone Harbour)
4. Grey seals (Langstone Harbour)

To illustrate the process of model fitting, a walkthrough of the Harbour Seal (Chichester Harbour) data is provided below. The other three models followed the same process process.

Example: Harbour Seals in Chichester Harbour

Temporal variation in the number of seals present was examined for *Month* and *Year*. Cyclic cubic splines were used as the basis type for *Month* to ensure there was no discontinuity between January and December,

and the basis size of this smooth was restricted to $k = 12$ to account for the number of unique values. Restricted maximum likelihood (REML) was used to minimise over-fitting

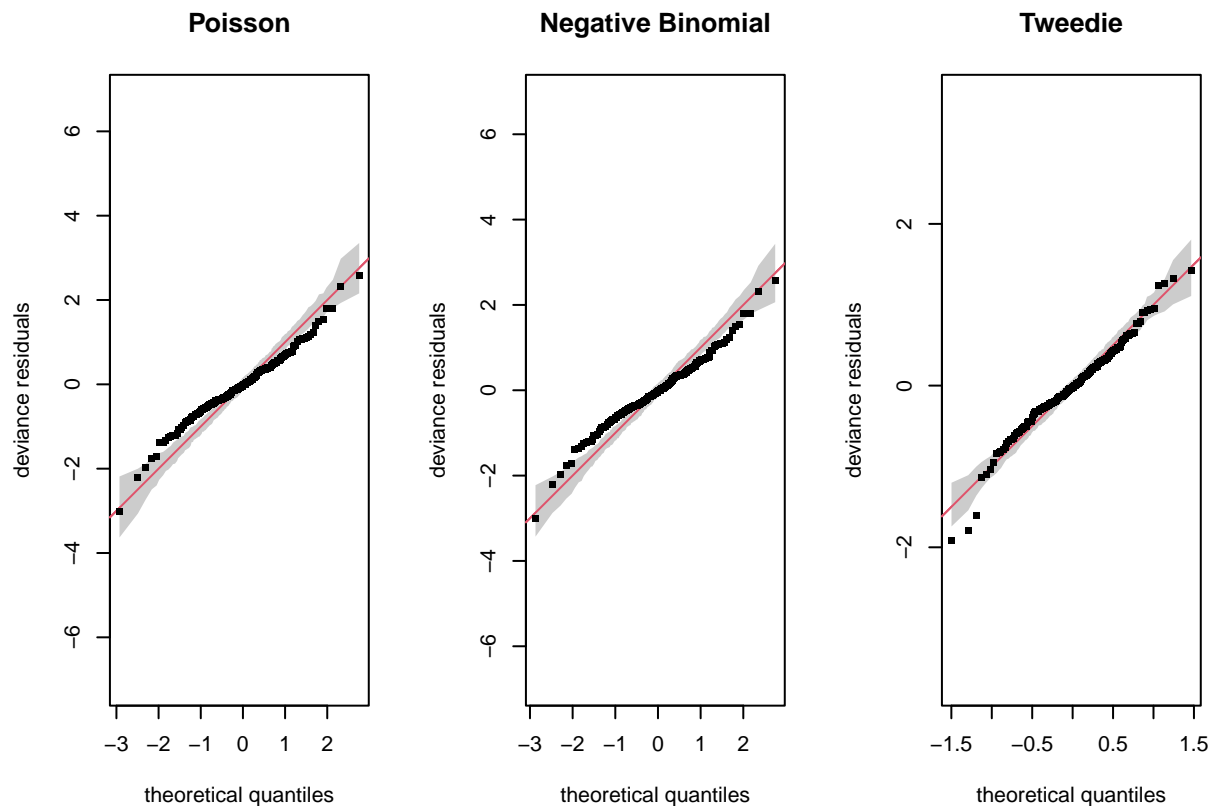
Dealing with count data, so start off trying a *Poisson* distribution. But also suspect overdispersion, so try models with *Negative Binomial* and *Tweedie* distributions too.

```
# Poisson distribution
H.po <- mgcv::gam(Harbour ~ s(Year) + s(Month, bs = "cc", k = 12),
  data = Ch.S, family = poisson(), method = "REML")

# Negative Binomial distribution
H.nb <- mgcv::gam(Harbour ~ s(Year) + s(Month, bs = "cc", k = 12),
  data = Ch.S, family = mgcv::nb(), method = "REML")

# Tweedie distribution
H.tw <- mgcv::gam(Harbour ~ s(Year) + s(Month, bs = "cc", k = 12),
  data = Ch.S, family = mgcv::tw(), method = "REML")

# Check QQ Plots
par(mfrow = c(1, 3))
mgcv::qq.gam(H.po, asp = 1, main = "Poisson", cex = 5, rep = 100)
mgcv::qq.gam(H.nb, asp = 1, main = "Negative Binomial", cex = 5, rep = 100)
mgcv::qq.gam(H.tw, asp = 1, main = "Tweedie", cex = 5, rep = 100)
```



Looking at all three QQ Plots, the *Poisson* and *Negative Binomial* models look over-dispersed. But the *Tweedie* looks a lot better!

Check the Akaike's Information Criterion for small sample sizes (AICc) to compare models:

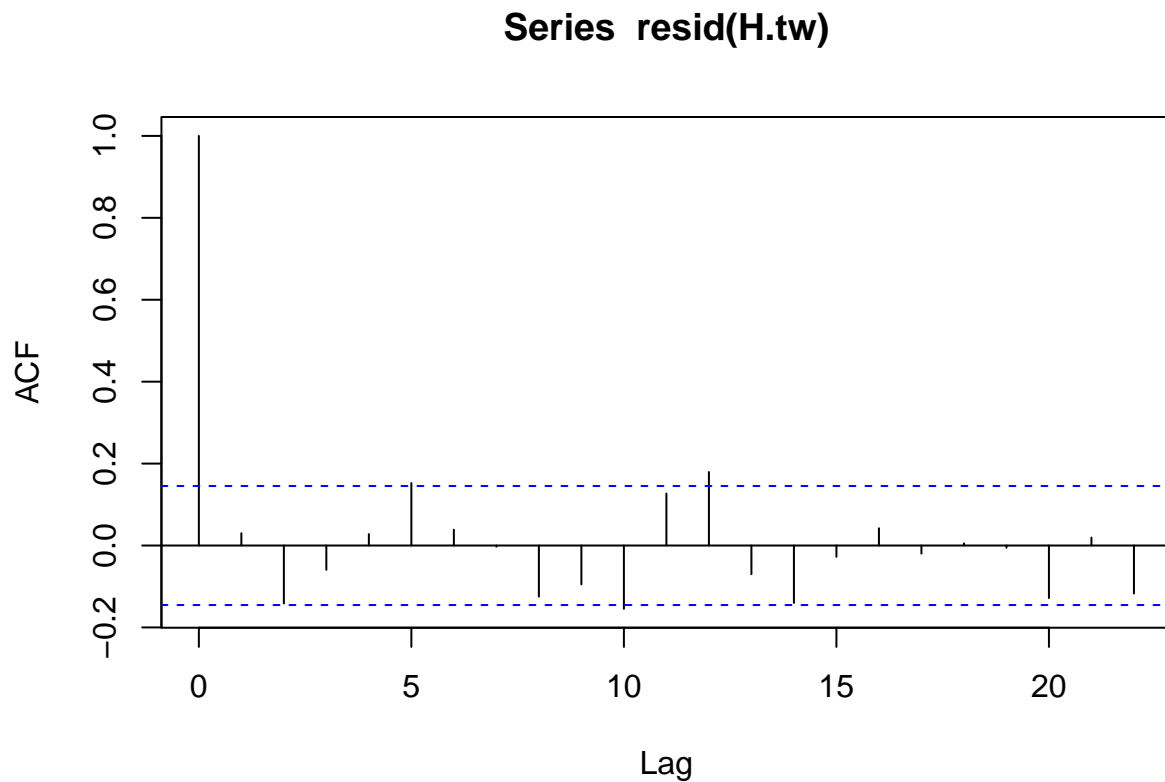
```
MuMIn::AICc(H.po, H.nb, H.tw)
```

```
##           df      AICc
## H.po 13.04790 910.1049
## H.nb 14.04794 912.4562
## H.tw 16.11859 887.0814
```

Again, the *Tweedie* model comes out the best with the lowest AICc score. We will use this model going forward.

Also want to check for the presence of temporal autocorrelation. The `acf()` function plots autocorrelation versus timelag.

```
acf(resid(H.tw))
```



The black bars are mostly under the blue line, so we can be reasonably confident that autocorrelation is not an issue.

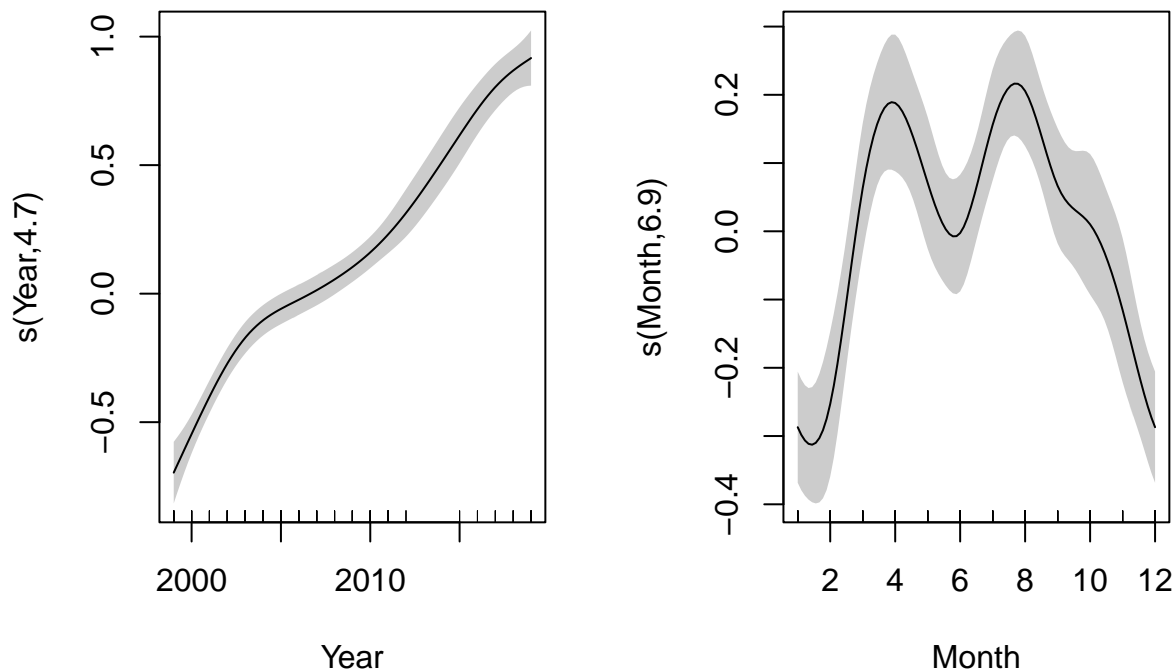
Now let's check the results:

```
# Model Output
summary(H.tw)
```

```
##
## Family: Tweedie(p=1.324)
```

```
## Link function: log
##
## Formula:
## Harbour ~ s(Year) + s(Month, bs = "cc", k = 12)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.42182    0.01767   137.1  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df      F p-value
## s(Year)      4.699  5.729 103.605 <2e-16 ***
## s(Month)     6.898 10.000   8.835 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.829   Deviance explained = 84.3%
## -REML = 452.48   Scale est. = 0.2792    n = 182

# Model Plots
par(mfrow = c(1, 2))
plot(H.tw, scale = 0, shade = TRUE)
```



Both *Year* and *Month* appear to be highly significant (both $p < 0.001$), and these terms explain a good

deal (84%) of the variation in this model. Overall, the number of harbour seals in Chichester Harbour is increasing over time. Within this site there are also seasonal patterns, with two distinct peaks around March and August.

If there had been any non-significant variables, we could have tried removing them to see if model fit improved. Here, this was not the case, so we are all good! However, for Langstone Harbour dataset, *Month* was not significant for either species. Removing this variable improved the harbour seal model, but not the grey seal model.

Final Models

The above process was repeated for each of the four models. The final models are shown below:

```
# Harbour seals (Chichester Harbour)
hs.ch <- gam(Harbour ~ s(Year) + s(Month, bs = "cc", k = 12),
            data = Ch.S, family = tw(), method = "REML")

# Grey seals (Chichester Harbour)
gr.ch <- gam(Grey ~ s(Year) + s(Month, bs = "cc", k = 12),
            data = Ch.S, family = tw(), method = "REML")

# Harbour seals (Langstone Harbour)
hs.lh <- gam(Harbour ~ s(Year),
            data = La.S, family = tw(), method = "REML")

# Grey seals (Langstone Harbour)
gr.lh <- gam(Grey ~ s(Year) + s(Month, bs = "cc", k = 12),
            data = La.S, family = tw(), method = "REML")
```

Model Outputs:

```
# Harbour seals (Chichester Harbour)
summary(hs.ch)

##
## Family: Tweedie(p=1.324)
## Link function: log
##
## Formula:
## Harbour ~ s(Year) + s(Month, bs = "cc", k = 12)
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.42182    0.01767   137.1   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df      F p-value
## s(Year)      4.699  5.729 103.605 <2e-16 ***
## s(Month)     6.898 10.000   8.835 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```
##
## R-sq.(adj) = 0.829   Deviance explained = 84.3%
## -REML = 452.48   Scale est. = 0.2792   n = 182
```

```
# Grey seals (Chichester Harbour)
summary(gr.ch)
```

```
##
## Family: Tweedie(p=1.01)
## Link function: log
##
## Formula:
## Grey ~ s(Year) + s(Month, bs = "cc", k = 12)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.1054      0.3427  -9.061 2.36e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df      F p-value
## s(Year)    1.000      1 171.632 < 2e-16 ***
## s(Month)   2.842     10   1.164 0.00404 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.929   Deviance explained = 92.1%
## -REML = 69.636   Scale est. = 0.99226   n = 182
```

```
# Harbour seals (Langstone Harbour)
summary(hs.lh)
```

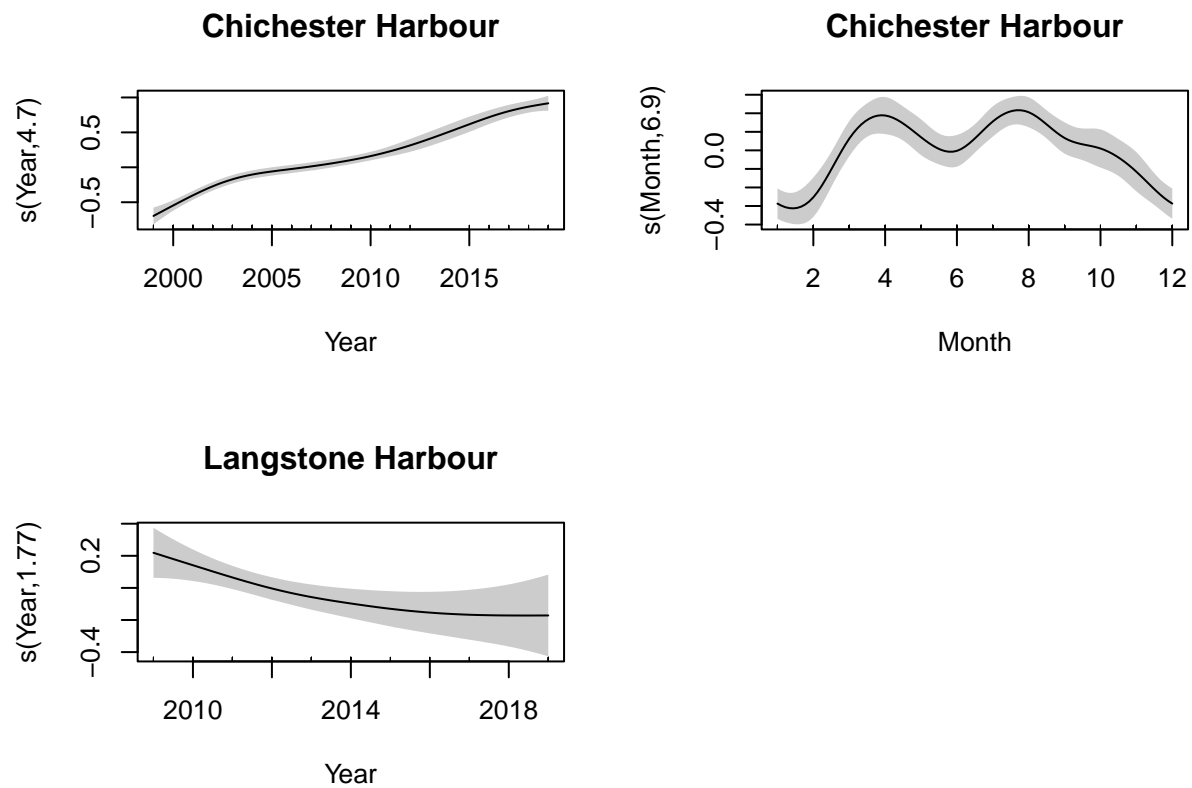
```
##
## Family: Tweedie(p=1.022)
## Link function: log
##
## Formula:
## Harbour ~ s(Year)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.44383    0.04655  31.02  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df      F p-value
## s(Year)    1.768    2.185  3.984 0.0198 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.101   Deviance explained = 9.86%
## -REML = 178.19   Scale est. = 0.77754   n = 88
```

```
# Grey seals (Langstone Harbour)
summary(gr.lh)
```

```
##
## Family: Tweedie(p=1.01)
## Link function: log
##
## Formula:
## Grey ~ s(Year) + s(Month, bs = "cc", k = 12)
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.701      1.242   -3.785 0.000289 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df      F p-value
## s(Year)      1.449  1.761  8.454 0.00681 **
## s(Month)     2.136 10.000  0.531 0.06078 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.684   Deviance explained = 78.4%
## -REML = 10.45   Scale est. = 1.0117      n = 88
```

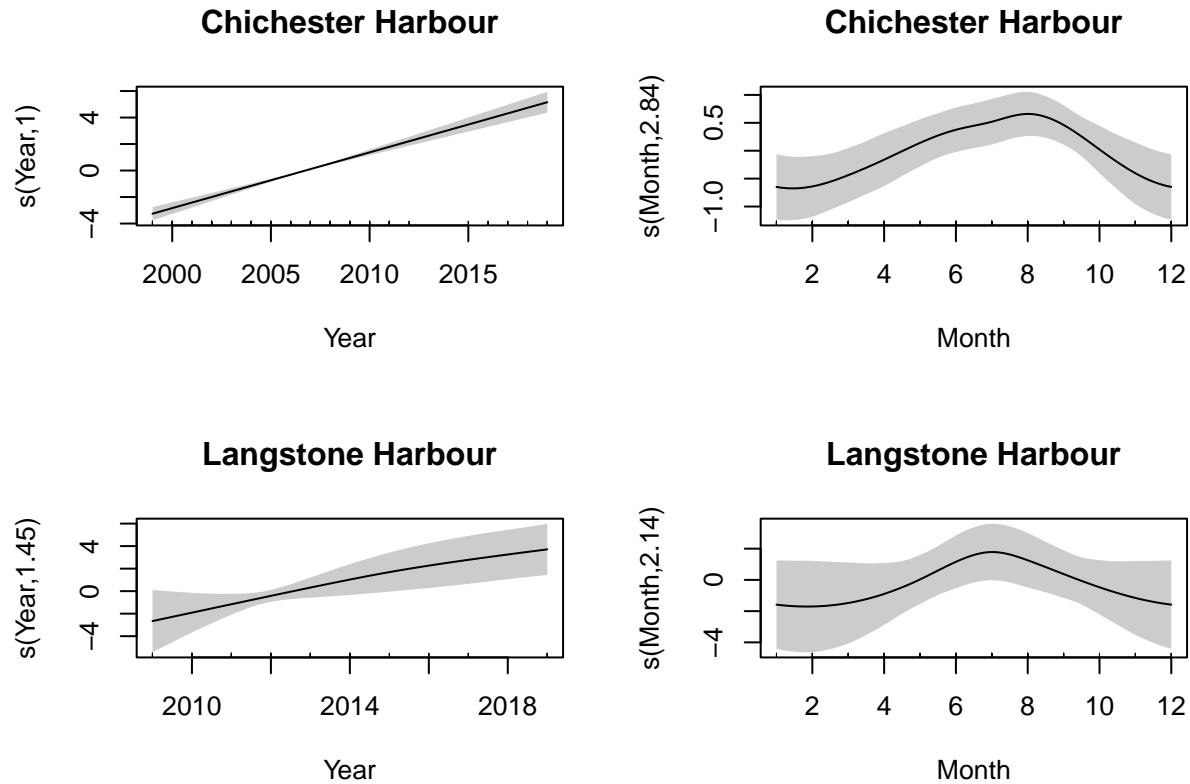
Final plots for harbour seals:

```
# Harbour seals from both harbours
par(mfrow = c(2, 2))
plot(hs.ch, scale = 0, shade = TRUE, main = "Chichester Harbour")
plot(hs.lh, scale = 0, shade = TRUE, main = "Langstone Harbour")
```



Final plots for grey seals

```
# Grey seals from both harbours
par(mfrow = c(2, 2))
plot(gr.ch, scale = 0, shade = TRUE, main = "Chichester Harbour")
plot(gr.lh, scale = 0, shade = TRUE, main = "Langstone Harbour")
```



Overall: in Chichester Harbour numbers of both seal species have increased. Between 1999 and 2019, mean number of harbour seals increased from 5.3 to 30.5, whilst grey seals increased from 0 to 12.0. There were also significant monthly trends in seal counts at this site. Peak numbers occurred in August for both harbour and grey seals, with harbour seals also showing a smaller peak in March.

In Langstone Harbour, there has been a significant decrease in the number of harbour seals (mean 5.3 to 4.0) between 2009 and 2019. However, the mean number grey seals has increased from 0 to 2.

Overall, this long-term study indicates an increasing number of both harbour and grey seals within the Solent. However, more research is required to identify the drivers of this trend.