Supplementary material: Tables and figures

Duncan Golicher

1/8/2019

Contents

Peak assemblage

Methodology

The peak assemblage table was derived from applying a consistent algorithm to the pooled data. The abundances were summed over all the sites for each month for all of the 19 species. The maximum count, rather then the low water count was used in the case of the observations of dunlin in 2017/18. This was considered to most likely match the method used for the data that was obtained from previous years surveys. The species abundances were first summed over all sites to produce an assemblage count for all winter months. For each winter season the month with the maximum assemblage value was found. The abundances of each of the species recorded in this month was taken as a measure of their individual contributions to this peak assemblage. This is a consistent methodology. However it differs from the data tables produced in the report in 2017. In the previous analysis the peak species abundances referred to the maximum observed count across all the months of each season. A proportional contribution could not be calculated using this approach, as the sum of the species counts exceeded the total for the peak assemblage.

Data

Counts data

Copy CSV Show 10 entries Search:

	Species	1999	2000	2001	2002	2007	2008	2010	2011	2012	2013	2014	2015	2016	2017
	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All
1	Avocet	31	0	0	161	1945	990	260	8	1900	1190	65	1074	599	1315
2	Bar-tailed godwit	0	0	0	0	0	240	10	0	1	0	0	0	5	2
3	Black-tailed godwit	53	0	0	0	1460	1800	600	632	2760	3602	361	961	419	1277
4	Cormorant	0	17	5	11	145	182	1	15	0	15	0	1	13	0
5	Curlew	0	9	1	3	56	56	40	114	154	57	27	77	77	170
6	Dark-bellied Brent goose	0	0	0	0	0	0	0	0	0	0	0	0	0	14
7	Dunlin	7750	7650	3677	9432	3433	1700	750	2722	1545	2690	6425	679	5256	7140
8	Gadwall	0	0	0	0	0	0	0	0	0	0	0	0	0	7
9	Golden plover	0	0	0	0	0	0	0	0	0	0	2	0	300	355
10	Greater white-fronted goose	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Showing 1 to 10 of 28 entries Previous 1 2 3 Next

Table of percent contribution to assemblage

Copy CSV Show 10 entries Search:

	Species	1999	2000	2001	2002	2007	2008	2010	2011	2012	2013	2014	2015	2016	2017
	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All
1	Avocet	0.4	0	0	1.6	22.5	14.4	10.2	0.2	21.9	14.5	0.9	32.2	5.9	9.1
2	Bar-tailed godwit	0	0	0	0	0	3.5	0.4	0	0	0	0	0	0	0
3	Black-tailed godwit	0.6	0	0	0	16.9	26.2	23.5	15.4	31.8	43.9	4.9	28.8	4.1	8.9
4	Cormorant	0	0.2	0.1	0.1	1.7	2.6	0	0.4	0	0.2	0	0	0.1	0
5	Curlew	0	0.1	0	0	0.6	0.8	1.6	2.8	1.8	0.7	0.4	2.3	0.8	1.2
6	Dark-bellied Brent goose	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1
7	Dunlin	90.1	91.1	87.8	91.9	39.8	24.7	29.4	66.1	17.8	32.8	86.7	20.3	52	49.7
8	Gadwall	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	Golden plover	0	0	0	0	0	0	0	0	0	0	0	0	3	2.5
10	Greater white-fronted goose	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Showing 1 to 10 of 28 entries Previous 1 2 3 Next

Peak assemblage values

Copy CSV Show 10 entries Search:

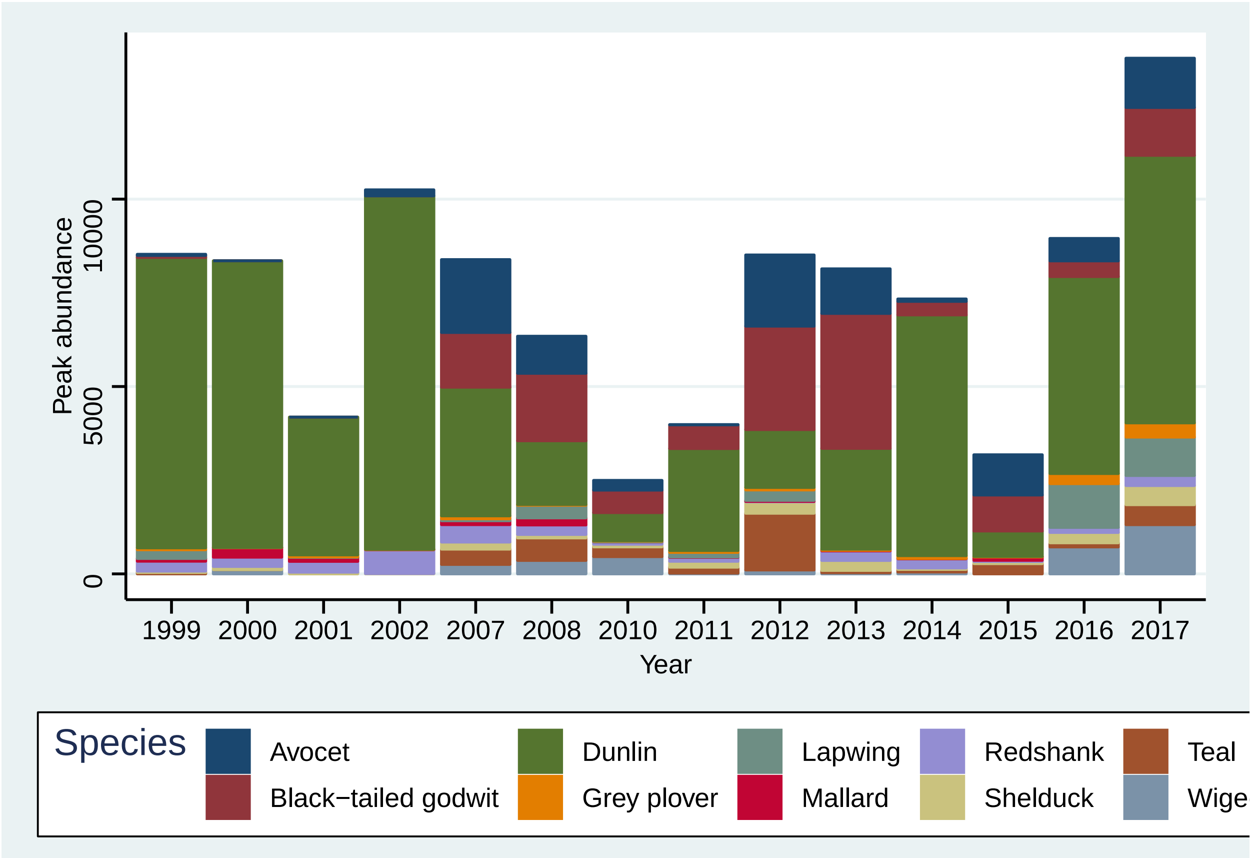
	Year	Sum
	All	All
1	1999	8606
2	2000	8395
3	2001	4190
4	2002	10263
5	2007	8635
6	2008	6869
7	2010	2554
8	2011	4115
9	2012	8666
10	2013	8213

Showing 1 to 10 of 14 entries Previous 1 2 Next

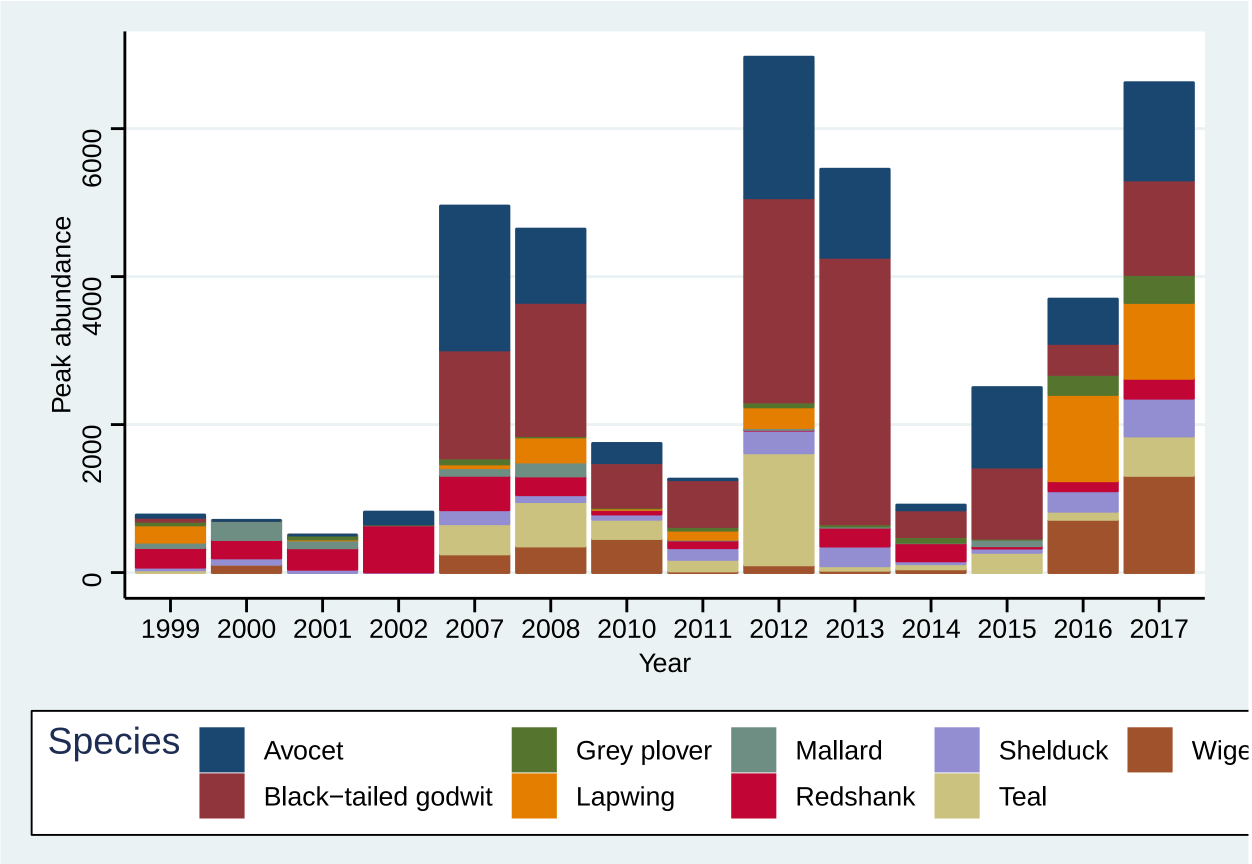
Barcharts

London Gateway

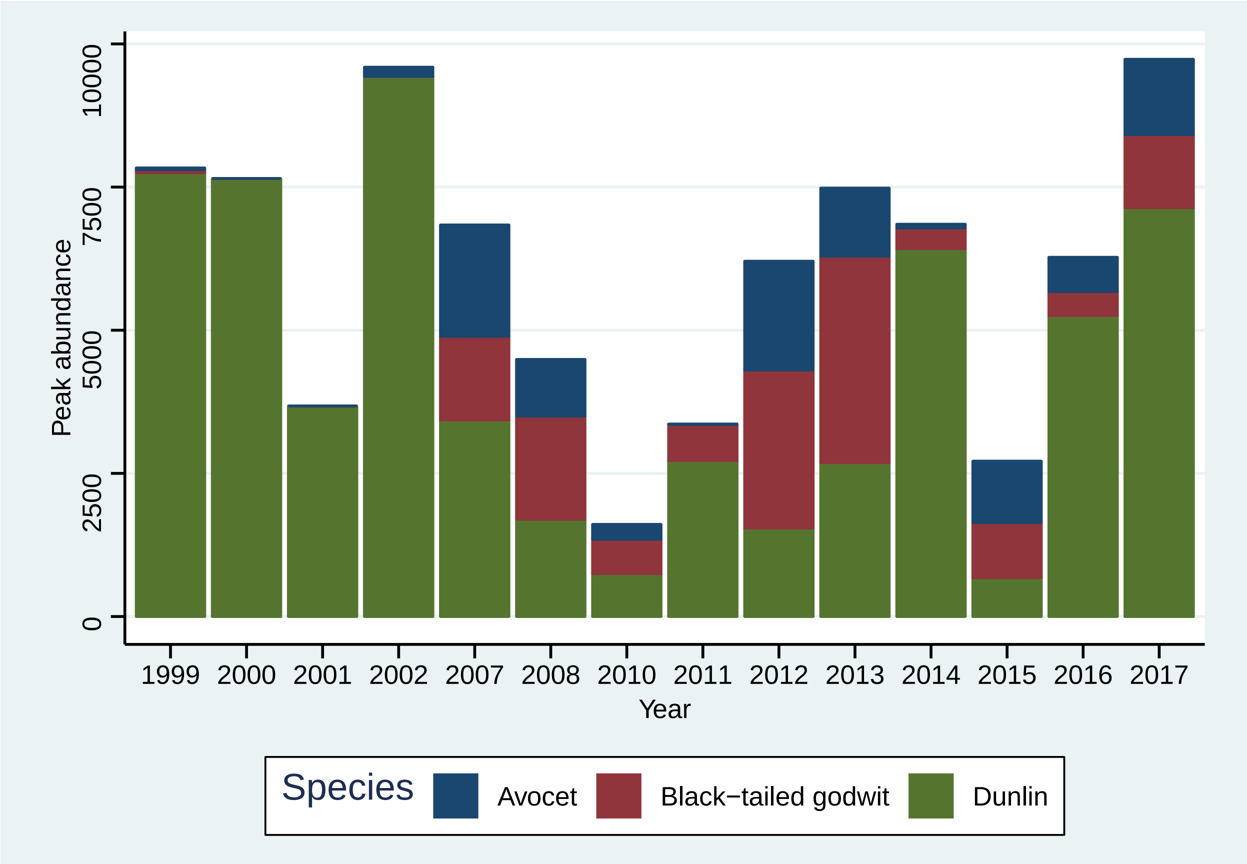
Top ten species



Excluding Dunlin



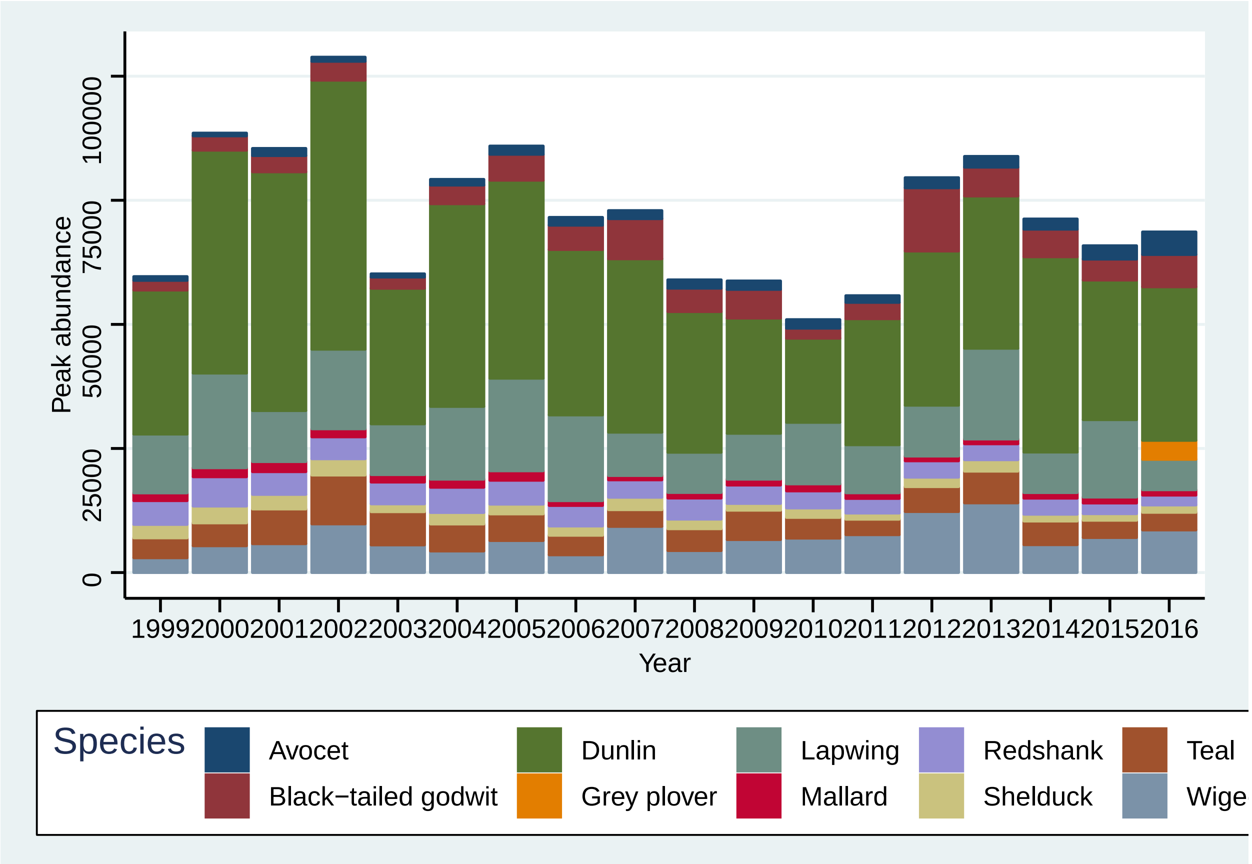
Top three species (key species)



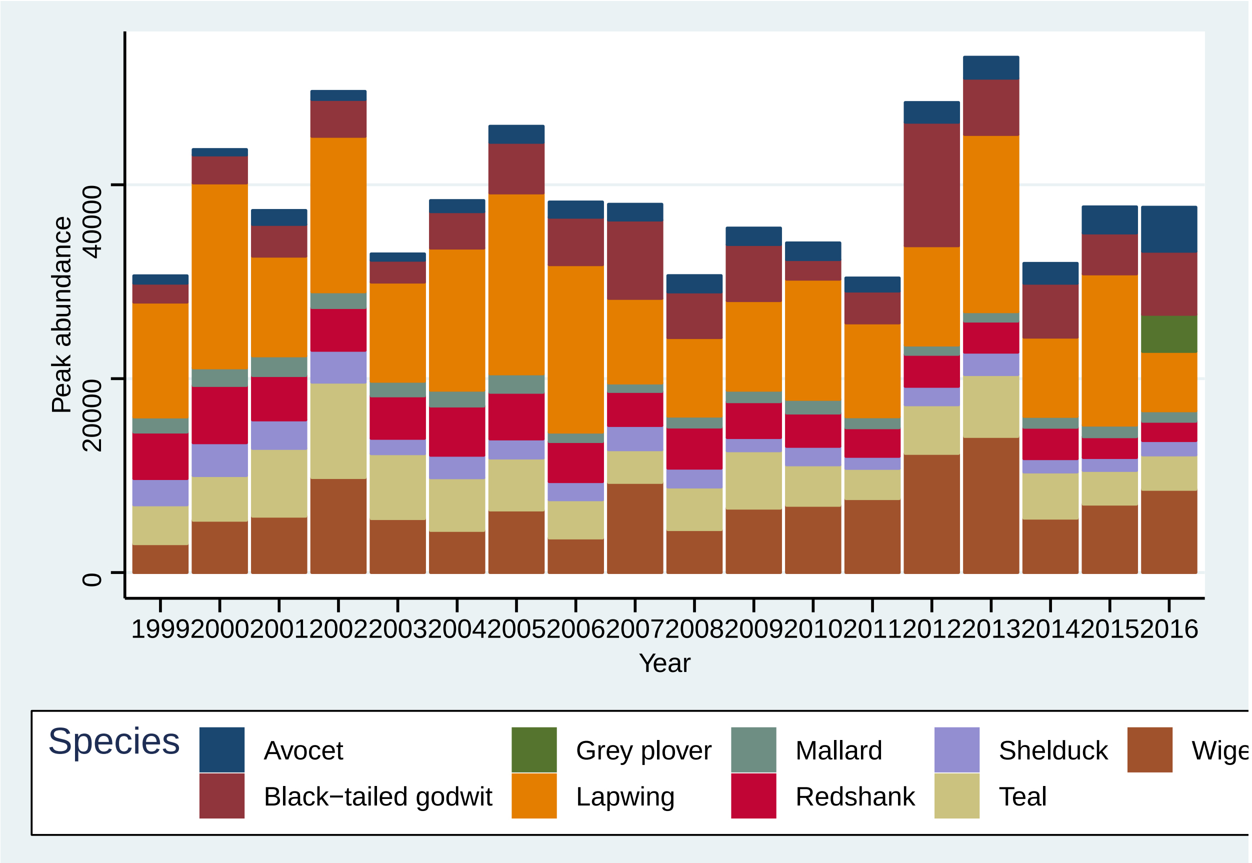
The top three most abundant species are Dunlin, Black-tailed godwit and Avocet.

Thames estuary

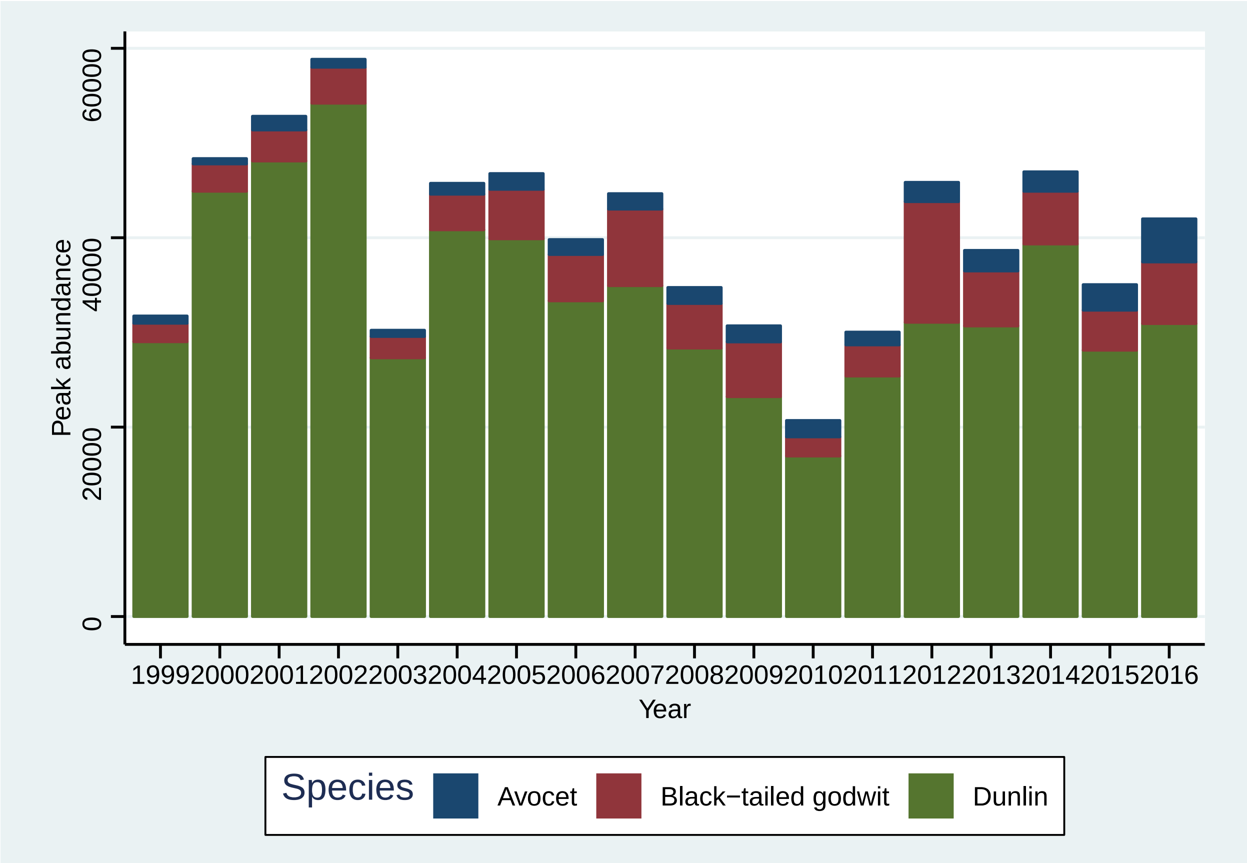
Top 10 species



Excluding Dunlin



Top three species (key species)



The top three most abundant species are Dunlin, Black-tailed godwit and Avocet.

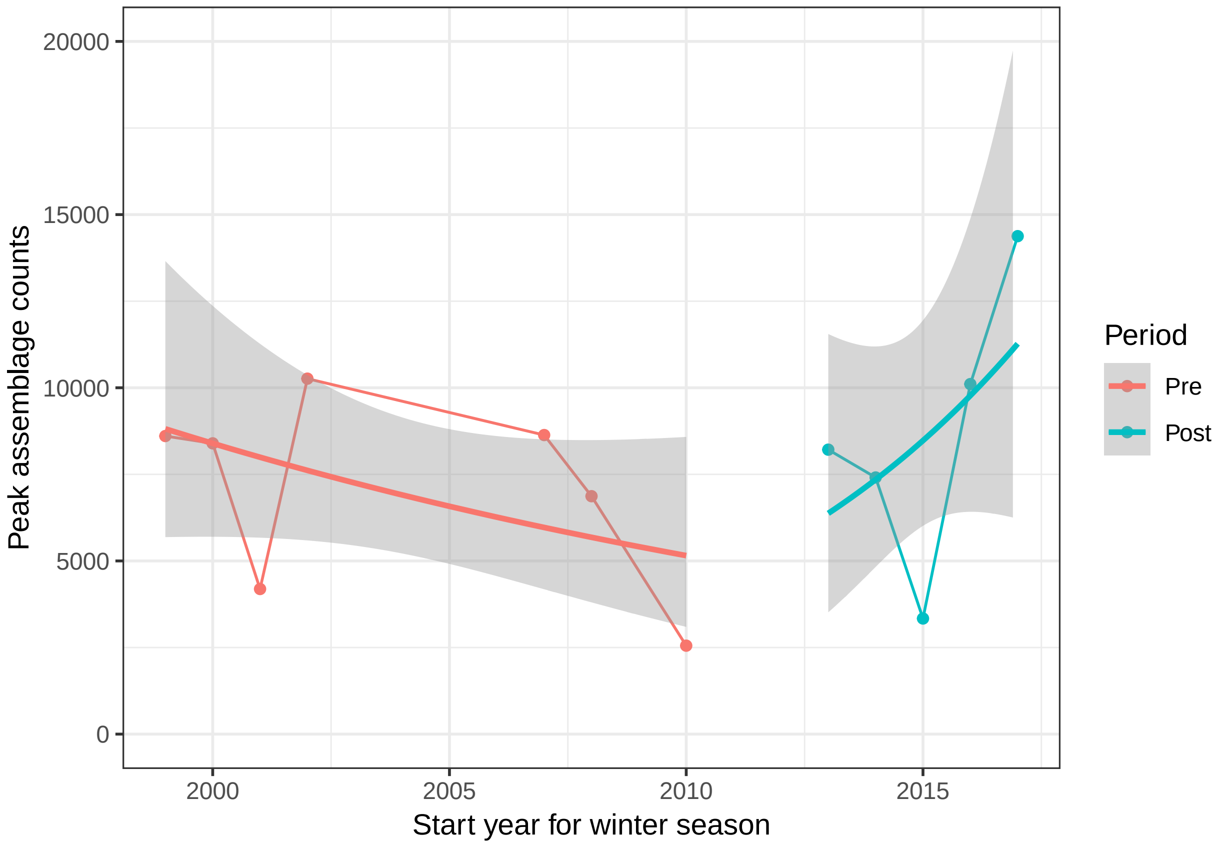
Generalised linear models for trends

Methodology

The use of regression as a method to formally analyse trends is greatly limited by the low sample size and by the high variability between count values. A longer time series of observations could show serial autocorrelation between the values forming a time series. Additional co-variates such as climatic effects might also be taken into account. As there are only a small number of data points available, a regression analysis could only provide evidence of a significant trend in the unlikely case of a monotonic year on year increase or decrease. In the case of counts that may take low numbers regression analysis using normally distributed errors may produce confidence intervals that fall below zero, which is impossible. Generalised linear models of the negative binomial family which account for overdispersion are prefereable in this case. Plotting lines derived from a negative binomial GLM with confidence intervals for the pre and post development periods shows no indication of such a consistent trend.

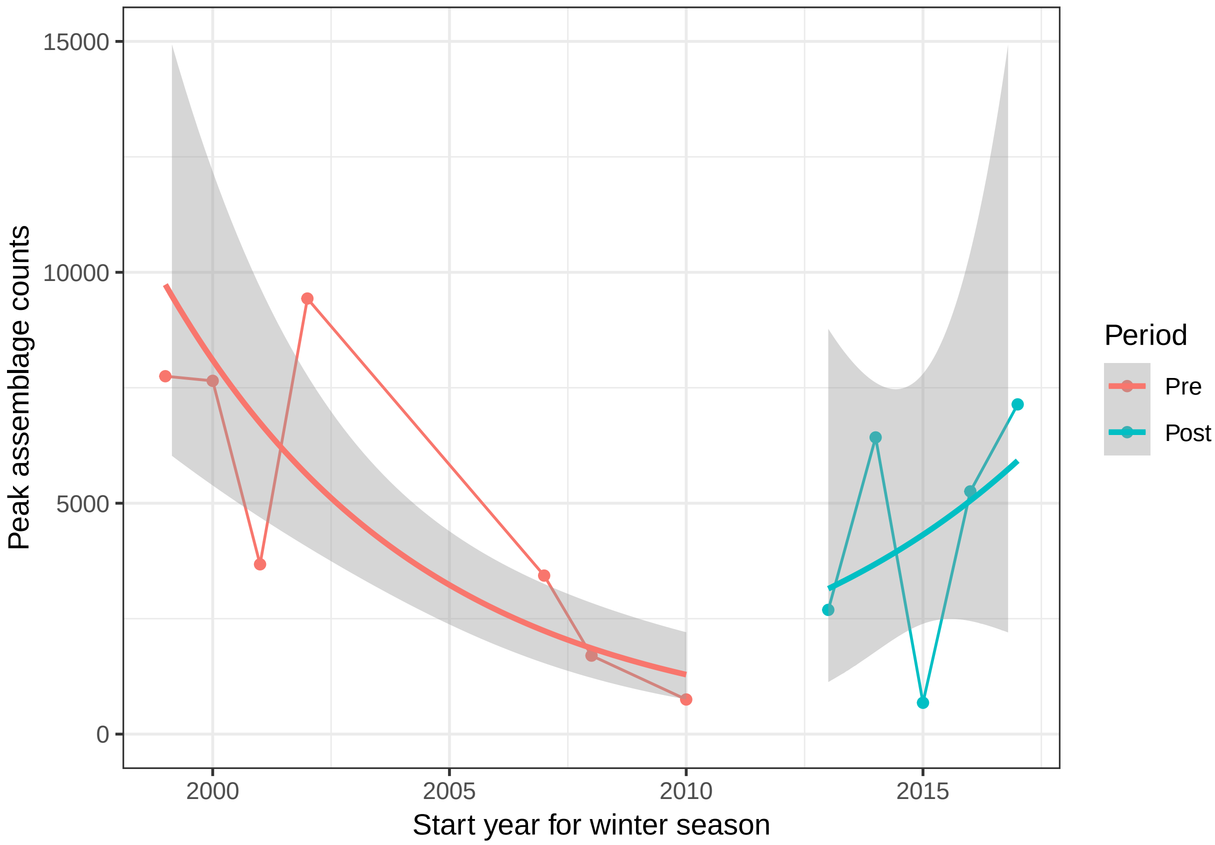
London gateway

All species

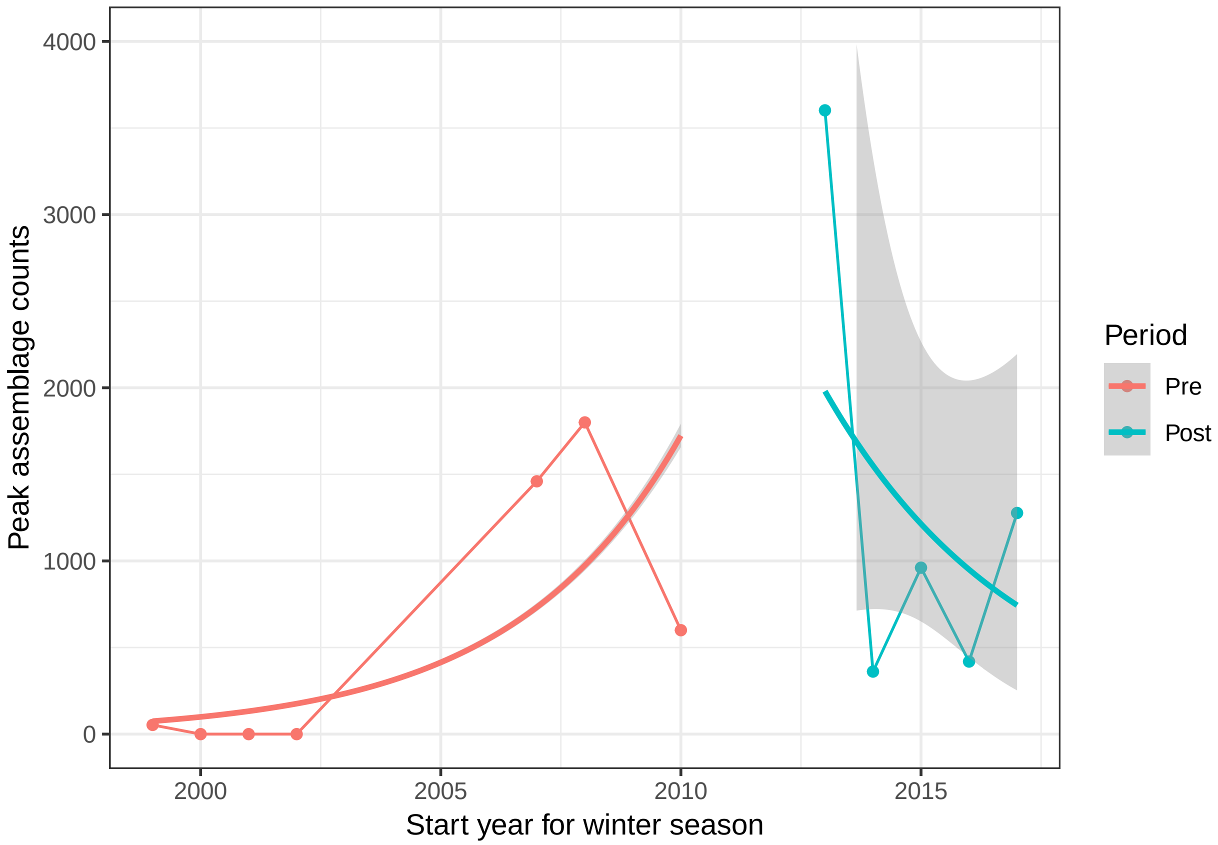


Peak assemblage counts
20000
15000
10000
5000
0
2000
2005
2010
2015
Start year for winter season
Period
Pre
Post

Dunlin

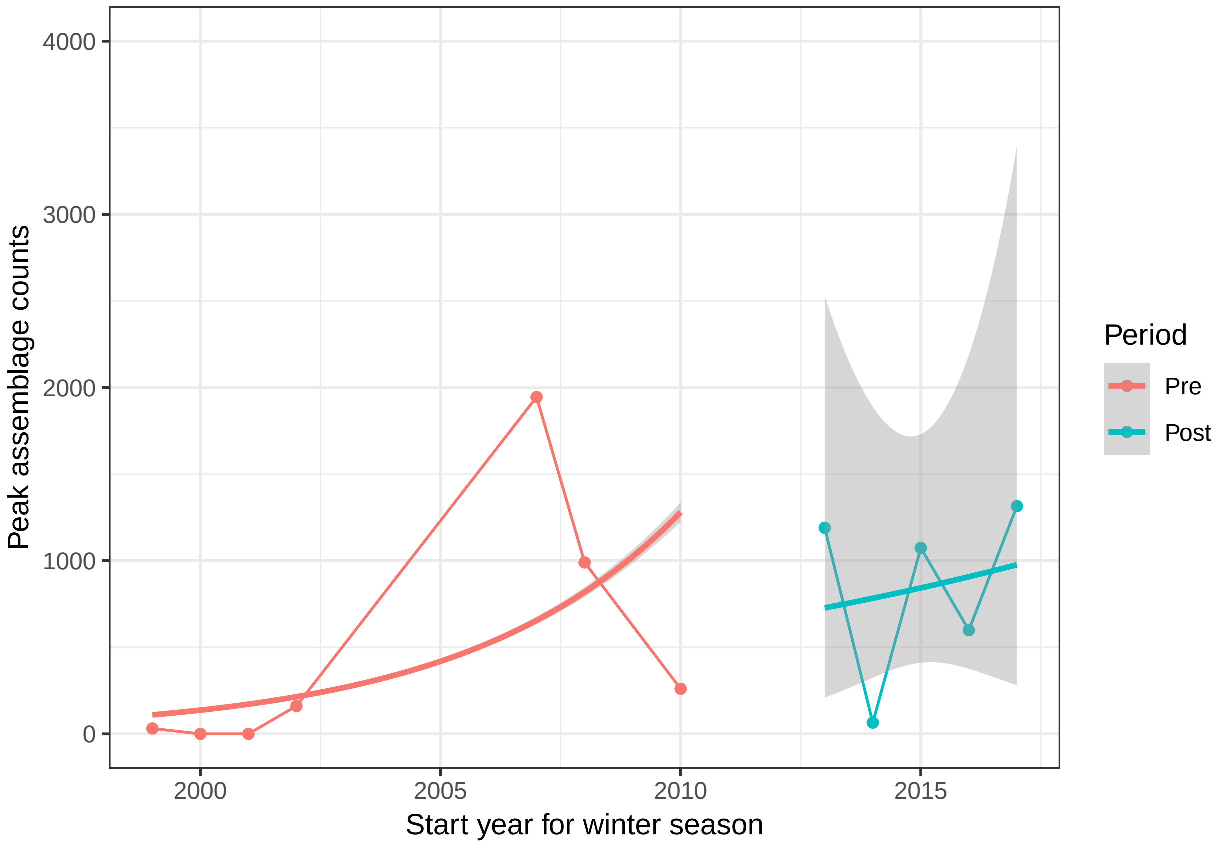


Black tailed godwit



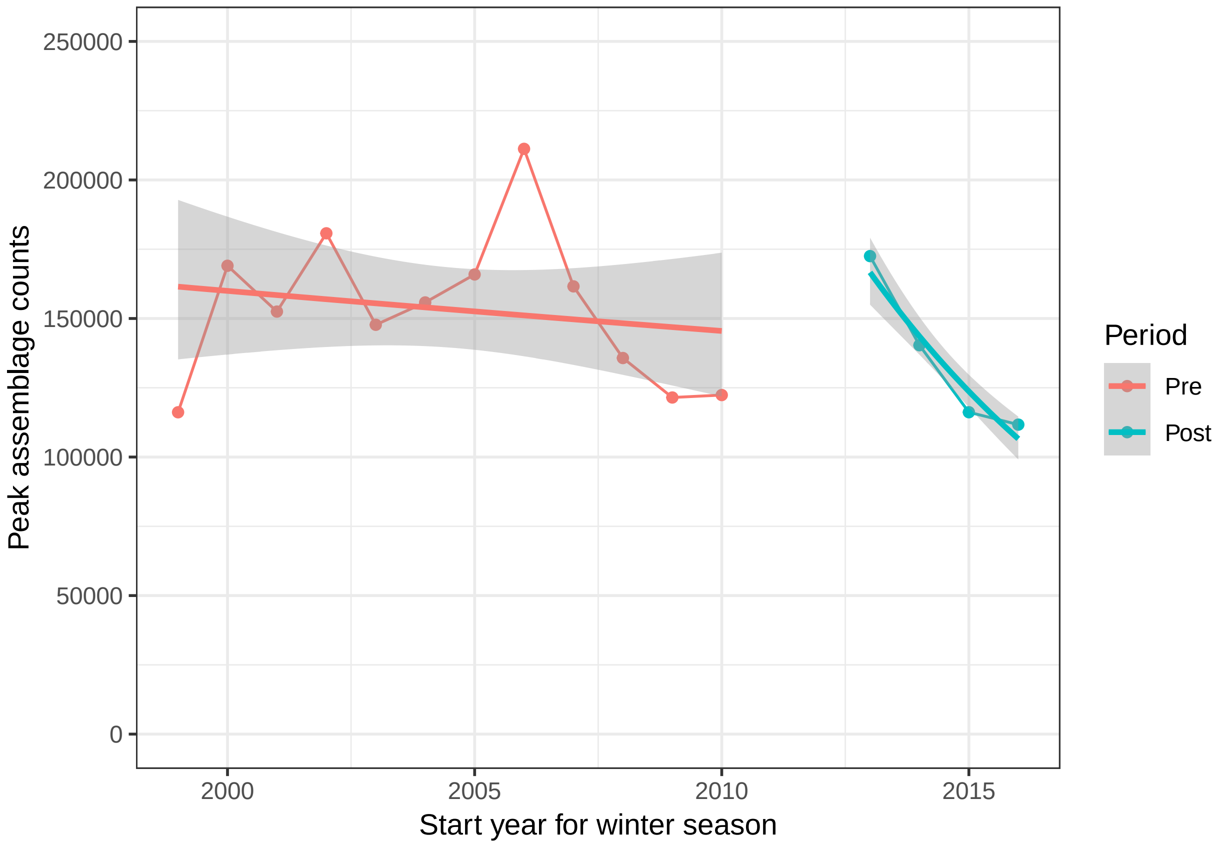
Peak assemblage counts
4000
3000
2000
1000
0
2000
2005
2010
2015
Start year for winter season
Period
Pre
Post

Avocet

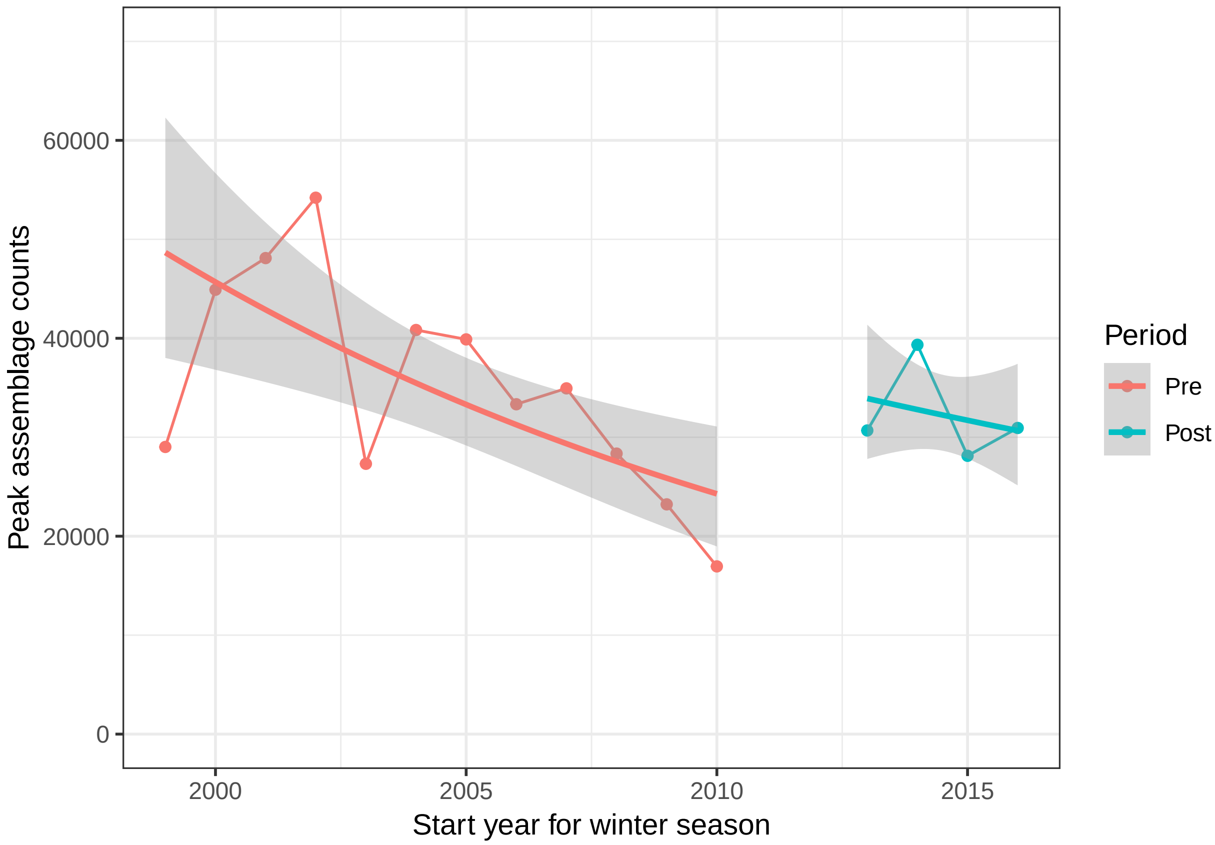


Thames estuary

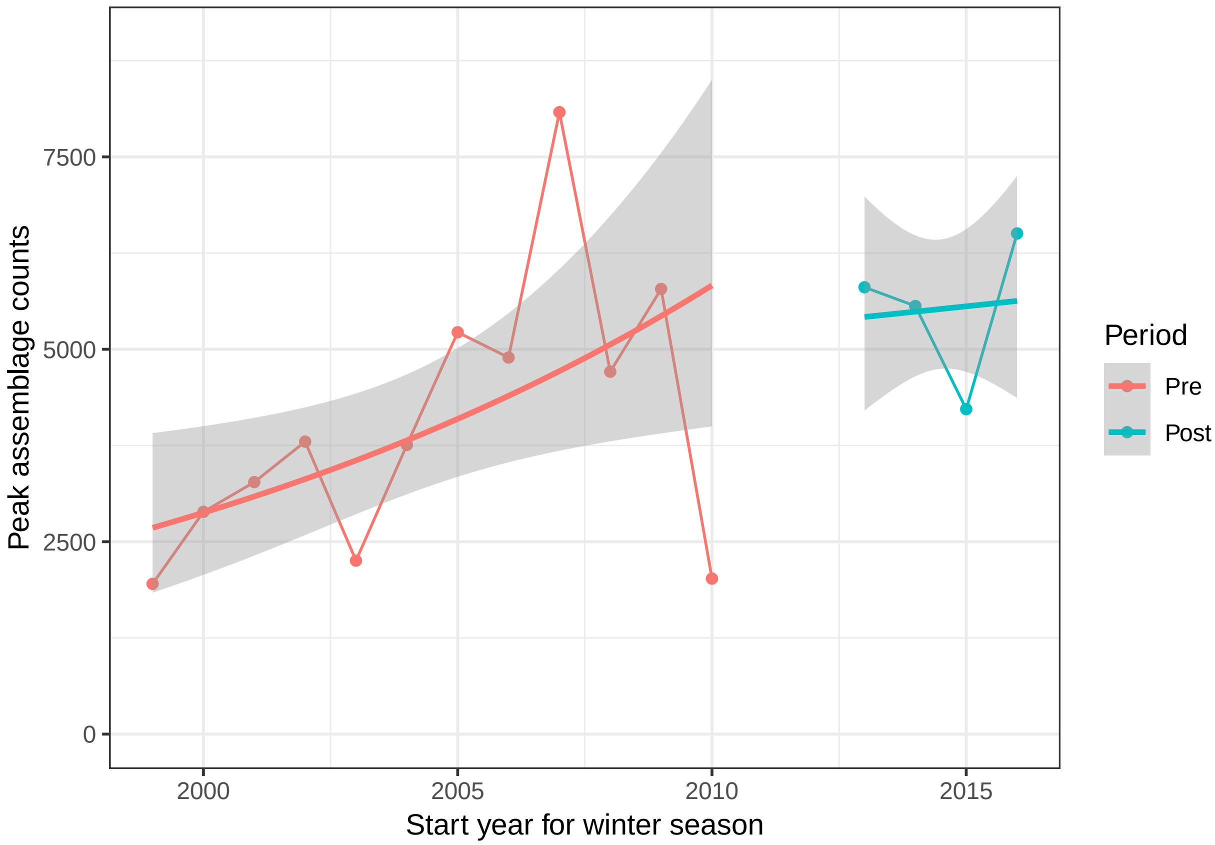
All species


Peak assemblage counts
250000
200000
150000
100000
50000
0
2000
2005
2010
2015
Start year for winter season
Period
Pre
Post

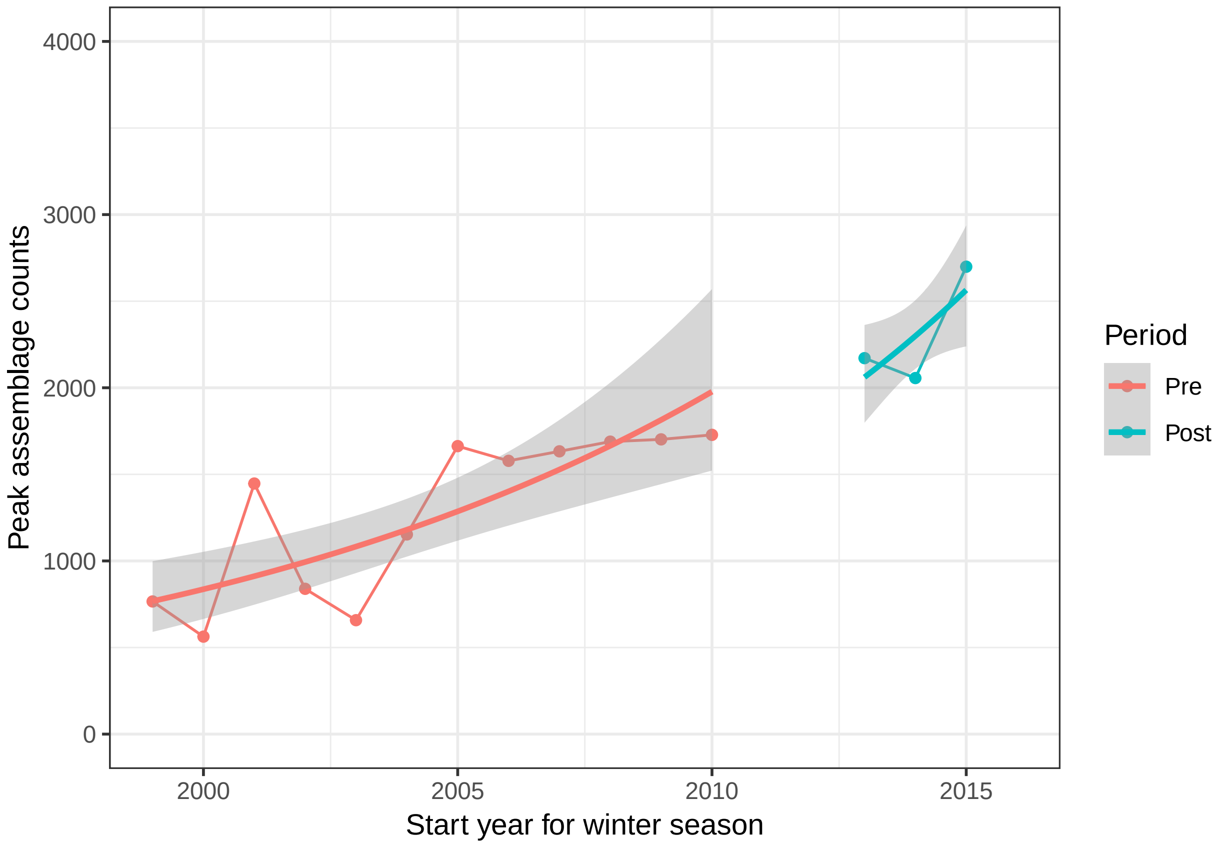
Dunlin



Black tailed godwit


Peak assemblage counts
0
2500
5000
7500
2000
2005
2010
2015
Start year for winter season
Period
Pre
Post

Avocet



Mean differences

Methodology

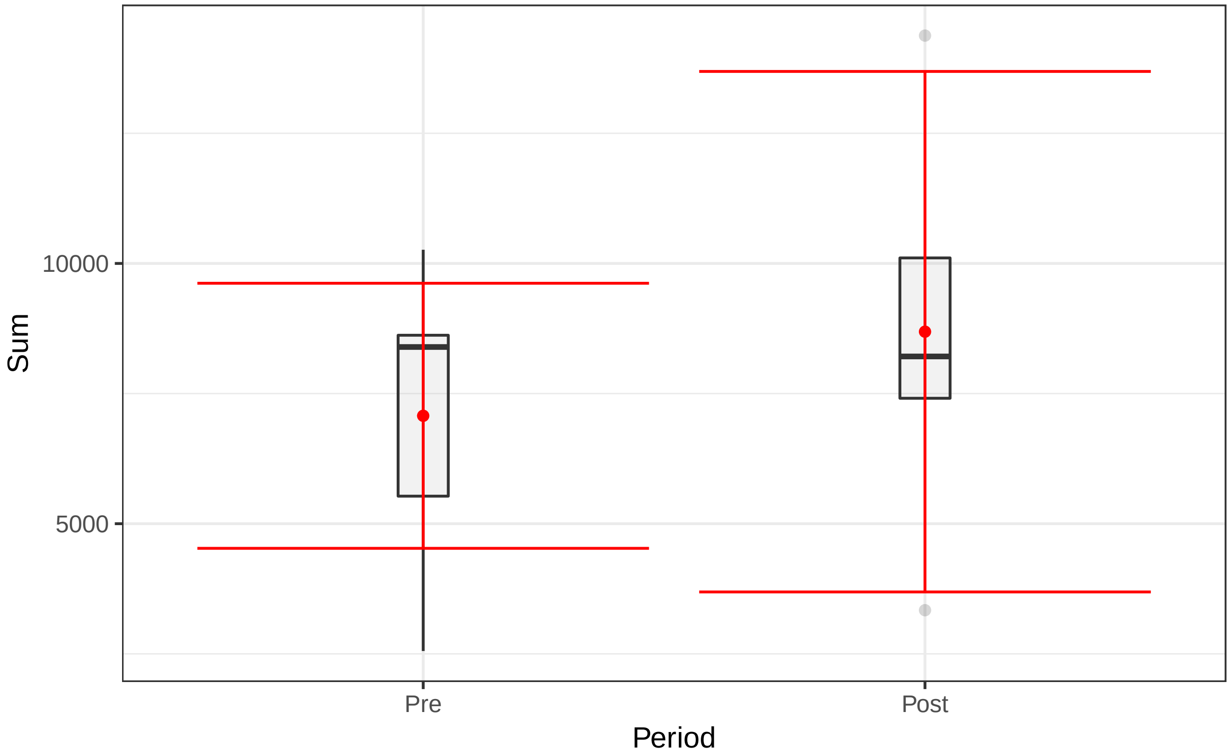
As there is no evidence of consistent trends, the variability between years can be treated as if the were a set of independent observations. This does not imply that there is no underlying relationship between the total population of birds in consecutive, simply that random variability due to movements of flocks and changes in the observability of the birds in combination with stochastic population fluctuations are adequate explanations for the observed variability.

In this case the period becomes a factor with two levels. Data can be visualised as boxplots and means with confidence intervals calculated from the variability around the mean values.

Aggregation

Boxplots and confidence intervals

Boxplot stats for peak aggregations in the two contrasted periods, showing superimposed means and 95% confidence intervals



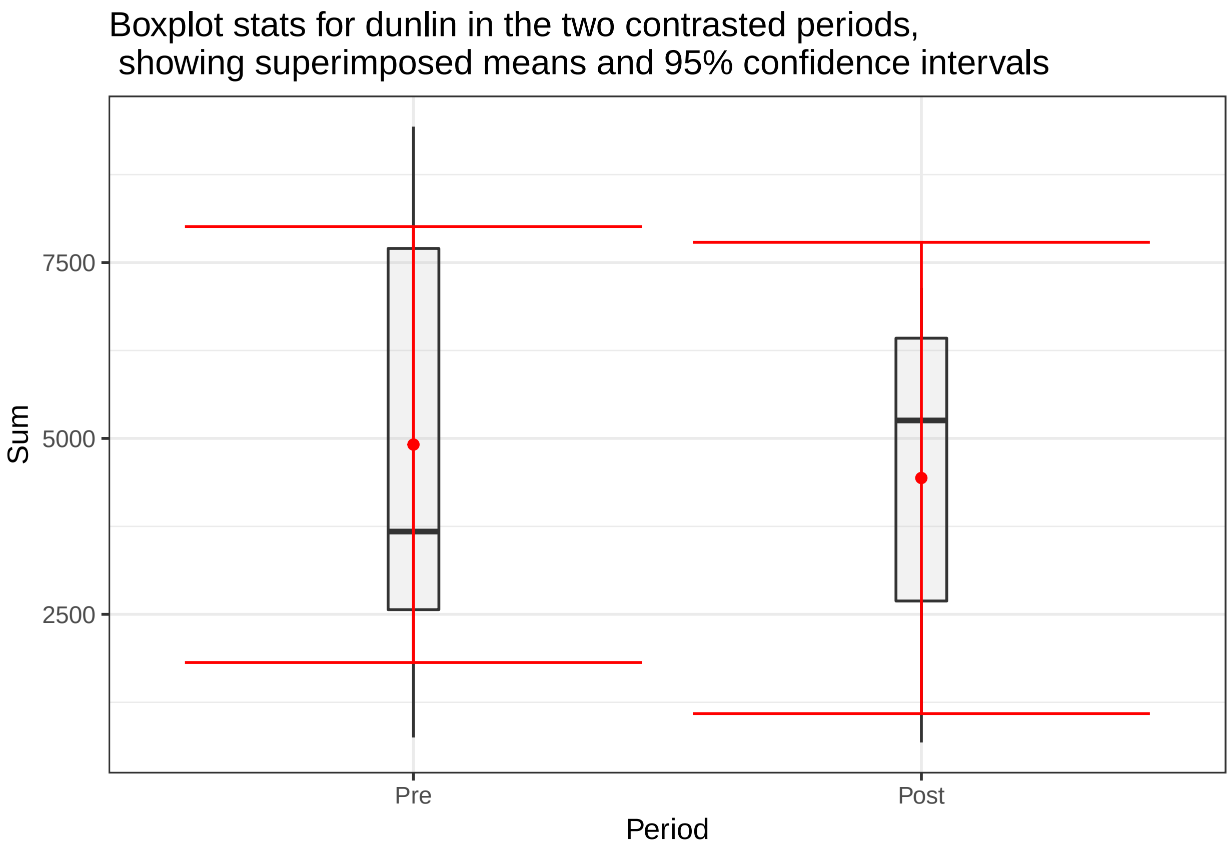
T-test

```
##
##  Welch Two Sample t-test
##
## data:  Sum by Period
## t = -0.7768, df = 6.6254, p-value = 0.4641
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6590.732  3359.418
## sample estimates:
##  mean in group Pre mean in group Post
##           7073.143           8688.800
```

The t-test provides no evidence of a significant difference between the mean peak assemblages pre and post works (p = 0.46).

Dunlin

Boxplots and confidence intervals



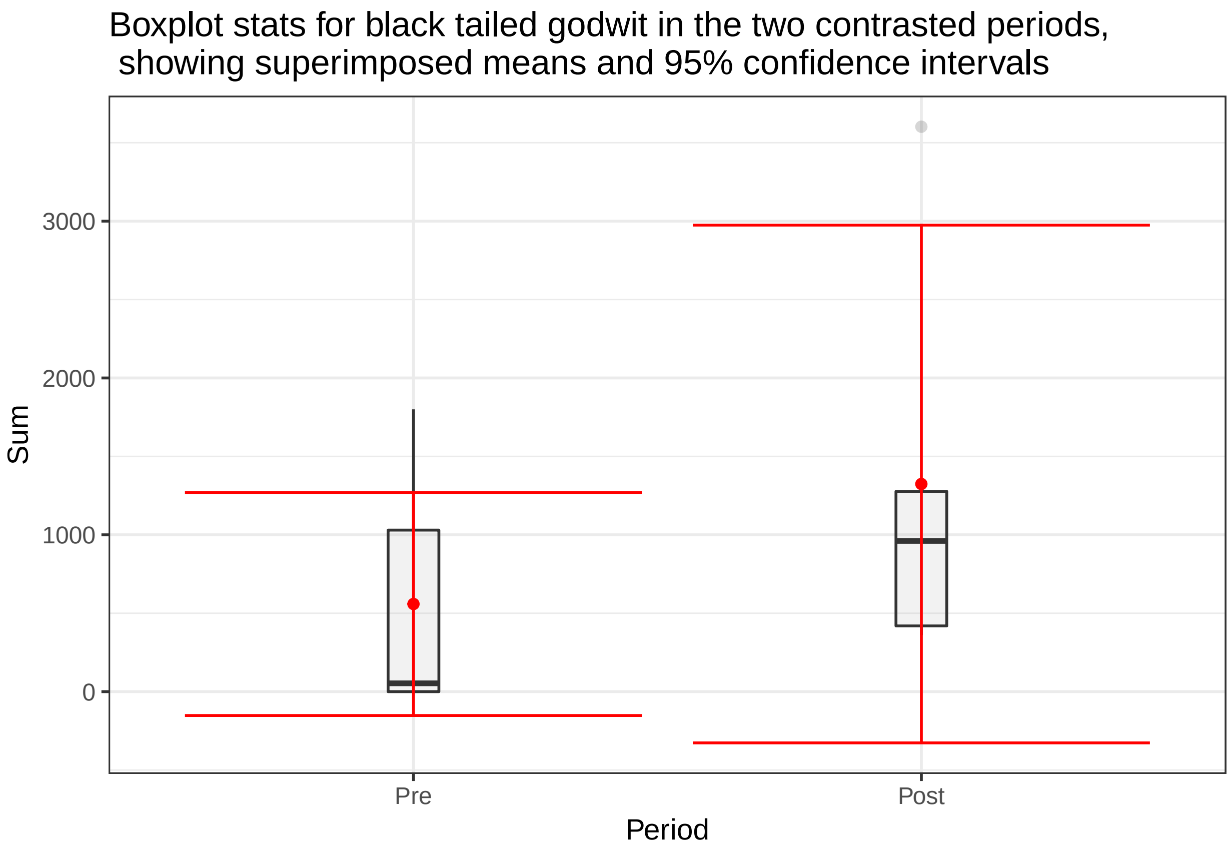
T-test

```
##
##  Welch Two Sample t-test
##
## data:  Sum by Period
## t = 0.27169, df = 9.7662, p-value = 0.7915
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3434.144  4384.429
## sample estimates:
##  mean in group Pre mean in group Post
##           4913.143           4438.000
```

The t-test provides no evidence of a significant difference between the mean peak assemblages pre and post works (p = 0.79).

Black tailed godwit

Boxplots and confidence intervals



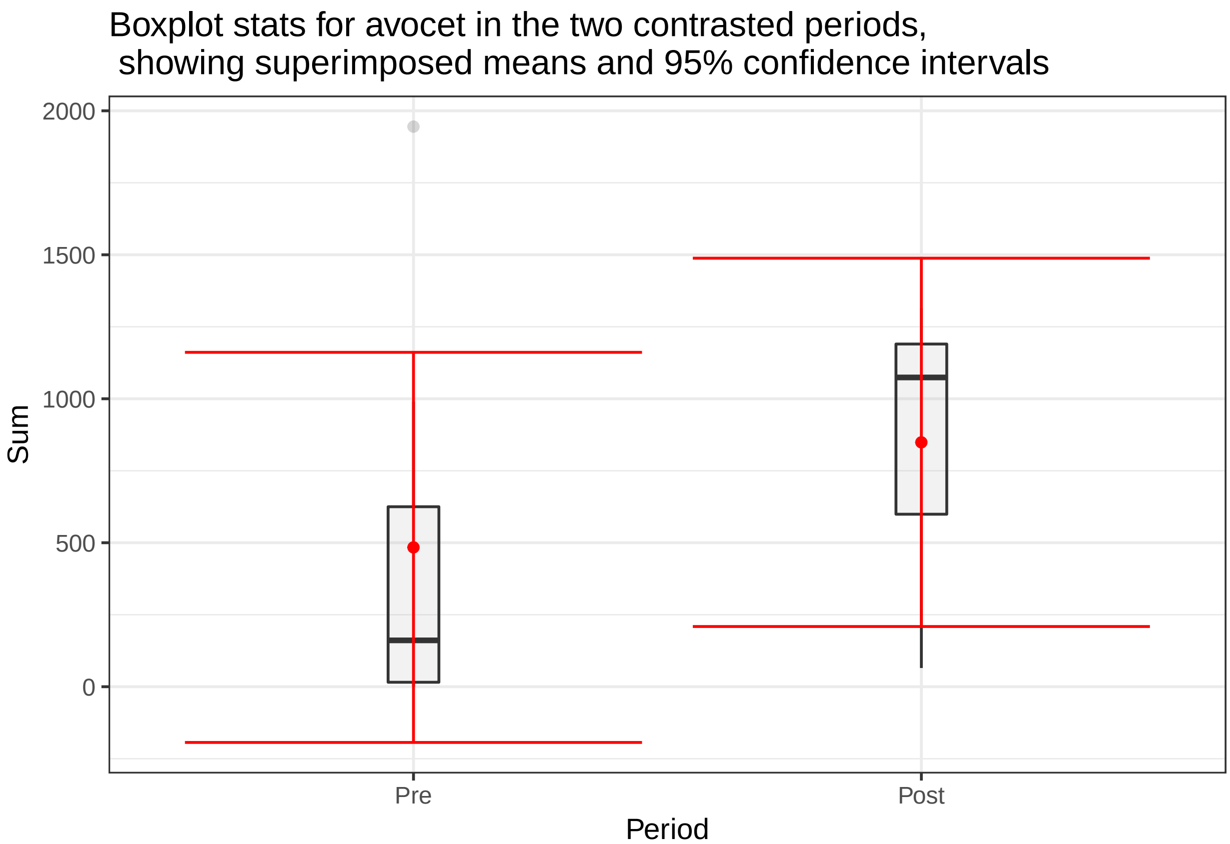
T-test

```
##
##  Welch Two Sample t-test
##
## data:  Sum by Period
## t = -1.156, df = 5.9146, p-value = 0.2922
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2389.9715   859.9715
## sample estimates:
##  mean in group Pre mean in group Post
##                559               1324
```

The t-test provides no evidence of a significant difference between the mean peak assemblages pre and post works (p = 0.29).

Avocet

Boxplots and confidence intervals



T-test

```
##
##  Welch Two Sample t-test
##
## data:  Sum by Period
## t = -1.0128, df = 9.9965, p-value = 0.335
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1167.2154   437.7296
## sample estimates:
##  mean in group Pre mean in group Post
##           483.8571           848.6000
```

The t-test provides no evidence of a significant difference between the mean peak assemblages pre and post works (p = 0.34).

Bayesian t-test

Methodology

There is an issue with regard to the interpretation of the result of such a test. Under null hypothesis significance testing (NHST) it is not possible to accept the null hypothesis of no difference (Anderson et al. 2000). NHST simply fails to reject the null. The p value represents the probability of obtaining the data, or data more extreme, given that the null hypothesis is in fact true . However the precise point null hypothesis of exactly no difference between assemblage counts before and after the works is not a credible one. There must be some differences. The issue is whether any differences fall within acceptable bounds. Thus NHST is problematic when the decision rule in question involves looking at the evidence **in favour** of the null hypothesis. This is the case here as shown in the description of the decision rule provided in the report.

*The initial target against which the success of the mitigation and compensation will be assessed shall be that the sites in combination support an assemblage of wintering waterfowl at low tide comprising, on a 5-year mean peak basis at least **7900 birds** made up of, in particular, avocet, dunlin and black-tailed godwit in similar proportions to those supported by North Mucking during the winters of 1999/2000 to 2002/2003 (considered in the context of the wider population trends)*

The alternative to NHST is to adopt a Bayesian approach to inference. Under this approach he 5 year means for peak abundances are not considered to be fixed quantities, but are themselves treated as random variables with distributions. Bayes' formula provides a formal mechanism of providing probabilities for unknown quantities of interest. In this case the difference between μ_1 and μ_2 (the mean peak assemblages before and after the works) is an unknown quantity.

Bayes theorem states.

$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$ Where $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, v)$

So, the posterior credibility of the combination of values for $(\mu_1, \mu_2, \sigma_1, \sigma_2, v)$ is the likelihood of that combination times the prior credibility of the combination, divided by the constant p(D). When it is assumed that the data are independently sampled, the likelihood is the multiplicative product across all the data values of the probability density of a t distribution. The prior is the product of the five independent parameter distributions. The constant p(D) is the marginal likelihood, which may be obtained by integrating the product of the likelihood and prior over the entire parameter space. This integral is difficult to compute analytically. This difficulty limited the application of Bayesian methods before computational solutions using simulation became available. However this limitation no longer exists. It is now computationally simple to fit the true Bayesian model using tools supplied through R (Plummer 2018). This allows the full posterior distributions of the parameter values to be obtained, leading to a richer and more informative analysis (Kruschke 2013).

Providing that uninformative prior probabilities for the parameters are used, applying Bayes theorem in the context of a t-test will then provide credible intervals for the differences between means (Ellison 2004) . Although the estimates may be numerically very similar to those derived from the confidence intervals of a traditional t-test (Colegrave and Ruxton 2003 , Edwards (1996)), the interpretation of the result now directly maps onto the required decision rule (RUBIN 1984).

The traditional t-test found that the best estimate for the pre-works mean as 484 and the post works mean as 849 giving a point estimate difference between the means of 365.

The original target value of 7900 for the overall assemblage were derived from low water count data for the four winter periods 1999/2000 to 2002/2003. The pre works data used in the t-test included some additional observations as, These observations are helpful in establishing the range of variability for inference so have been included. If it were desirable these values could be excluded and the analysis re-run without them.

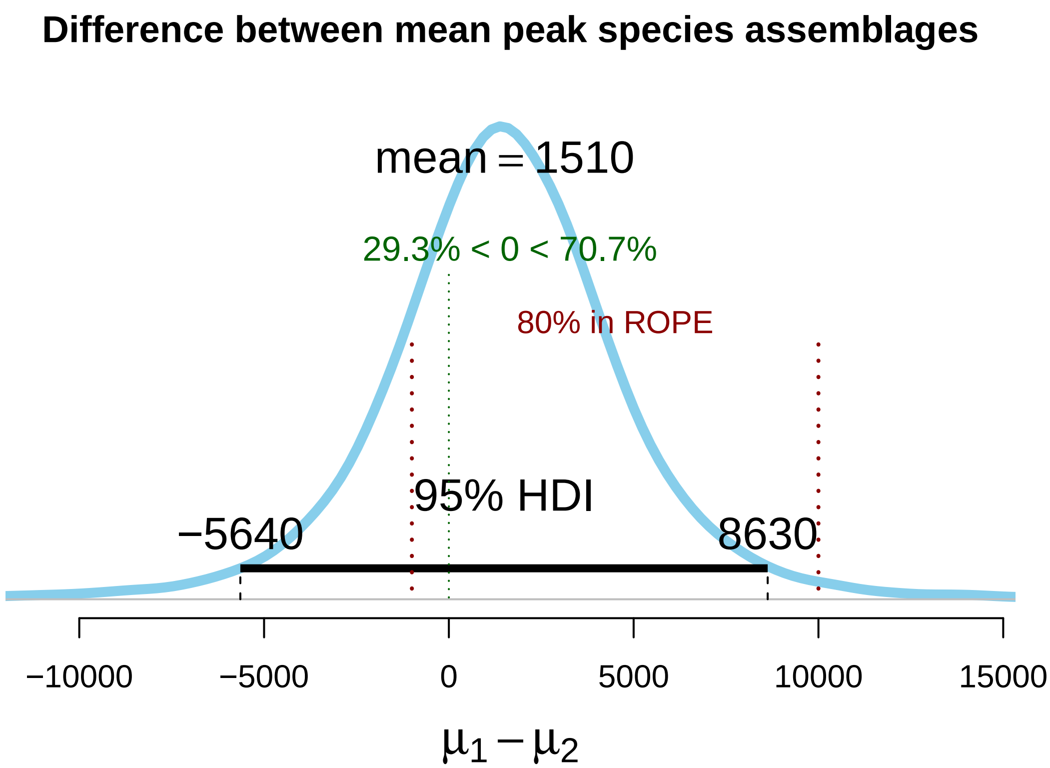
Bayesian model fitting allows a formal evaluation of a decision rule based on the concept of the **region of practical equivalence** (ROPE). This is an area around the null value of no difference which encloses those values of the parameter that are deemed to be not importantly different from the null value for the practical purposes of the study.

In this case any increase in assemblage numbers, even if not statistically significant, are of no practical importance in evaluating whether Clause 10.5.4 has been met. The ROPE can therfore extend to the right almost indefinitely. The choice of a left hand boundary for the ROPE has to be considered through a careful evaluation of the available data. The target value was originally set at around 8000 birds. This is around 1000 higher than the first estimate of the pre-works mean. It would thus seem reasonable to set a ROPE lying between -1000 and 1000.

The bayesian t-test is then run using the package BEST {Kruschke and Meredith (2018)} in R. The model used completely non-informative vague priors for the parameters of interest in order to avoid subjectivity.

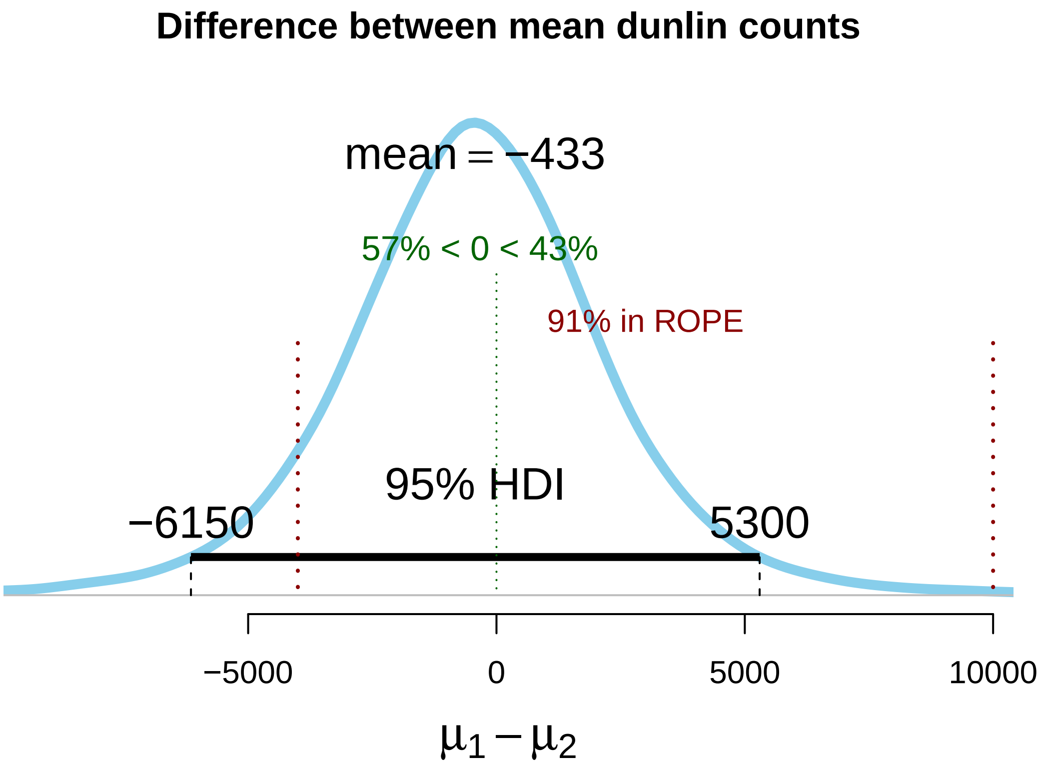
The resulting simulation provided the full posterior distribution for the differences between the two means.

Aggregation



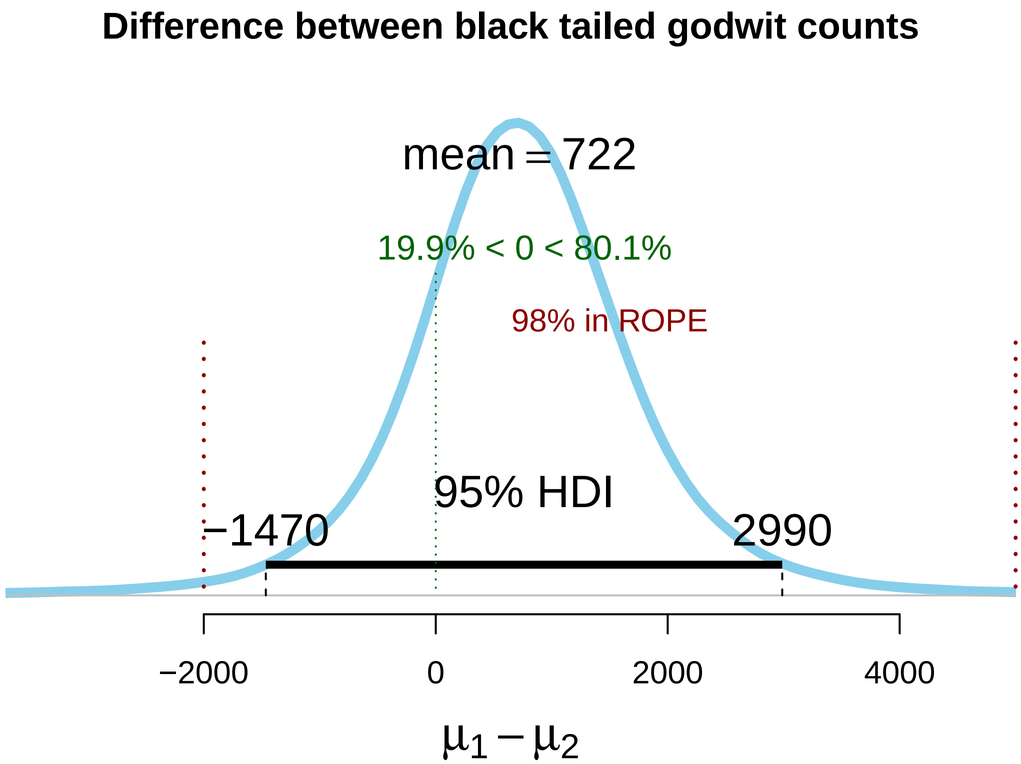
The figure shows the full posterior distribution for the difference between the two means, which is treated as a random variable and takes a t-distribution. The interpretation of the figure in terms of a decision rule of the Bayesian t-test analysis is clear when the ROPE is superimposed on the distribution. Around 80% of the posterior distribution for the difference between the two means lies within the ROPE. Although this does imply that there is a 20% chance that the value could stil lie outside the ROPE, the analysis is still being based on limited data. Intuitively it would be impossible to decide with certainty that the criteria had been met based on many fewer data points. The analysis formalises the strength of the currently avaialable evidence. As more data becomes available the probability that the criteria would be met becomes higher. Bayesian analysis allows for updating posterior distributions through additional data.

Dunlin



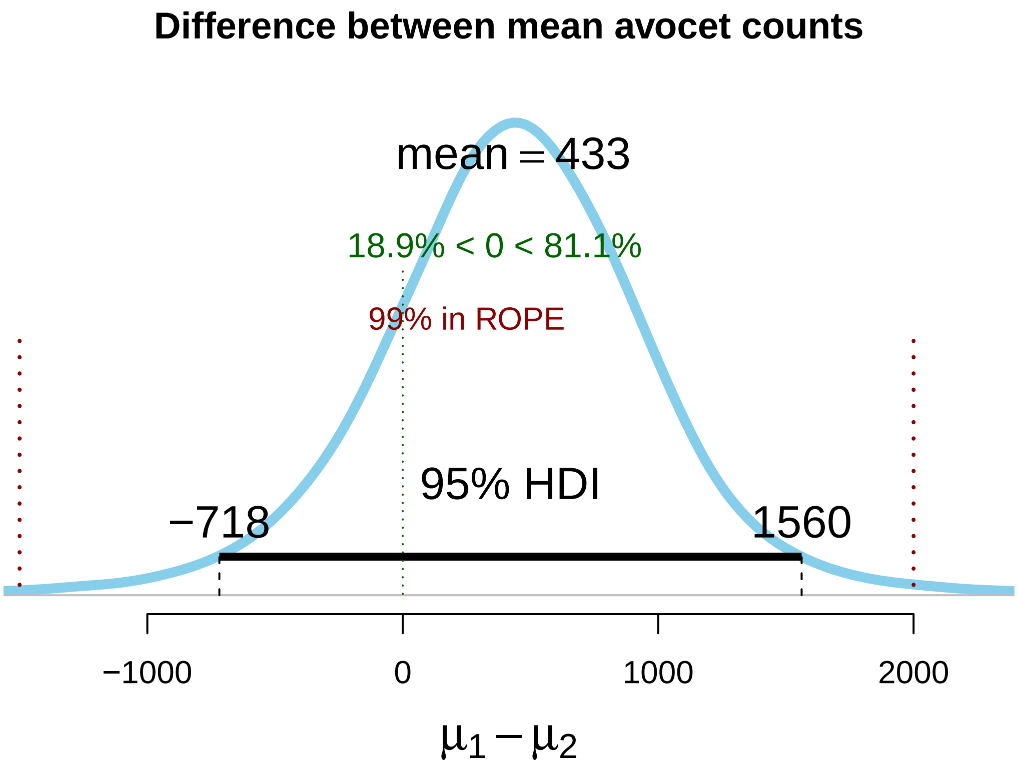
The analysis shows that around 90% of the posterior distribution falls within the ROPE. Thus there is very strong evidence that the works have had no practical impact on overall dunlin numbers.

Black tailed godwit



In this case a small amount (around 2%) of the ROPE actually falls below the posterior 95% highest density interval for the differences between the two means. The practical equivalence criteria is met to a very high degree of certainty, given the addditional evidence that black tailed godwit numbers have significantly increased over the period from 1998.

Avocet



The practical equivalence criteria is again met to a very high degree of certainty, given the addditional evidence that avocet numbers have significantly increased over the period from 1998.

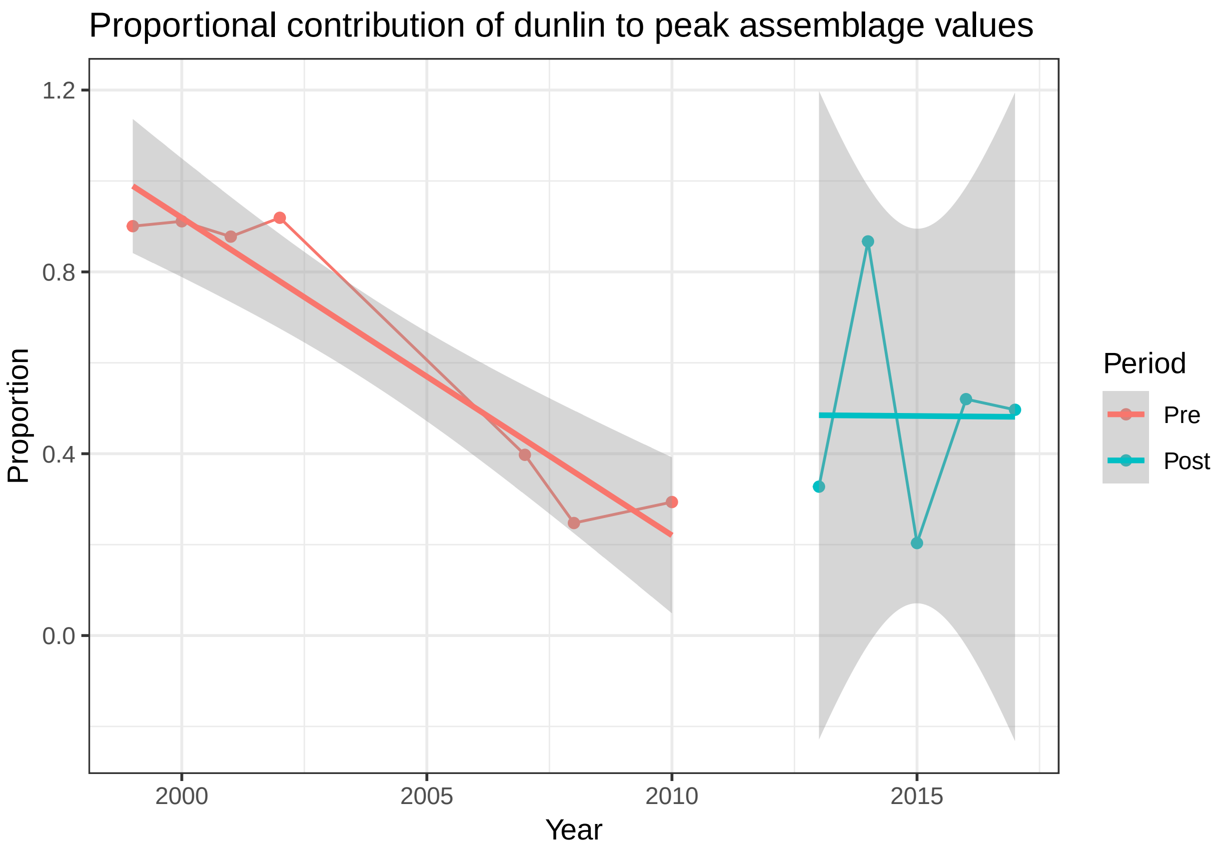
Differences in proportional abundance

Methodology

The target goal was also stated in terms of proportional abundance. *At least 7900 birds made up of, in particular, avocet, dunlin and black-tailed godwit in similar proportions to those supported by North Mucking during the winters of 1999/2000 to 2002/2003*

Inspection of the stacked bar charts and the raw data shows that the proportion of dunlin in the assemblage was higher between 1999 and 2002 than at present. As dunlin are small common waders a decrease in their proportional contribution would be interpreted as a positive effect, rather than a negative one.

Changes in proportional abundance of dunlin.



Trend analysis for proportional abundance of dunlin using beta regression

In order to establish the signficance of the change generalised linear modelling based on the beta distribution would provide the most robust approach. Proportions cannot be modelled with normally distributed errors. The betareg package in R allows this {Grün et al. (2012)}

Yearly trend

```
##
## Call:
## betareg(formula = Proportion ~ Year, data = dunlin)
##
## Standardized weighted residuals 2:
##     Min      1Q  Median      3Q     Max
## -1.5243 -0.9514  0.3101  0.6394  2.1324
##
## Coefficients (mean model with logit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 210.8212    84.1799   2.504   0.0123 *
## Year         -0.1048     0.0419  -2.501   0.0124 *
##
## Phi coefficients (precision model with identity link):
##       Estimate Std. Error z value Pr(>|z|)
## (phi)    4.299      1.625   2.646  0.00815 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 3.661 on 3 Df
## Pseudo R-squared: 0.439
## Number of iterations: 141 (BFGS) + 4 (Fisher scoring)
```

Beta regression produces evidence of a statistically signficant (p=0.012) reduction in the proportional contribution of dunlin to the species assemblage between 1999 and present. However the trend occurred prior to the works cmencing.

Changes in species diversity

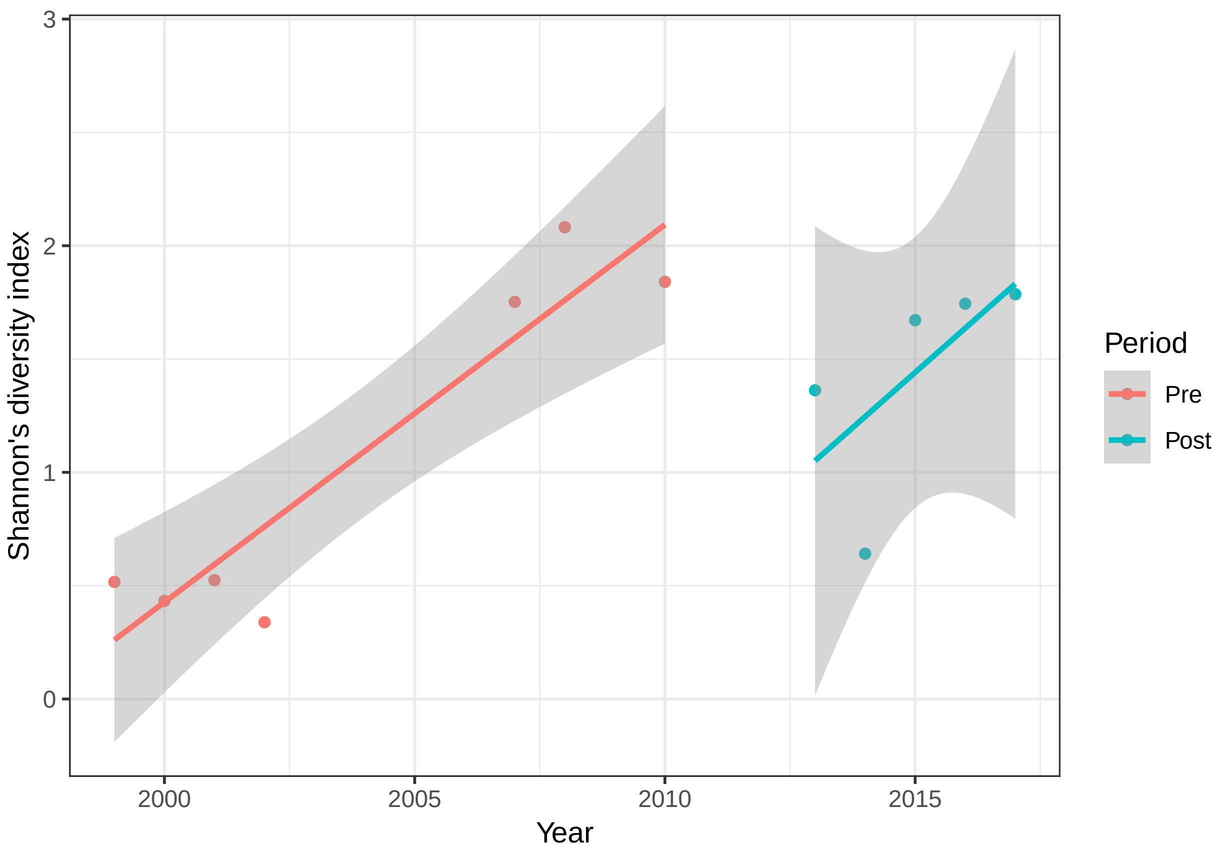
Although the criteria used to evaluate the impact of the works aimed to ensure a comparable mix of species abundances, the decline in relative abundance of dunlin and the increase in the relative contribution of other species may have increased species diversity. This is generally considered to be a positive outcome for conservation.

In order to evaluate changes in species diversity a commonly used diversity index was calculated for the assemblage. Shannnon's index is based on proportional contributions of each species to the assemblage.

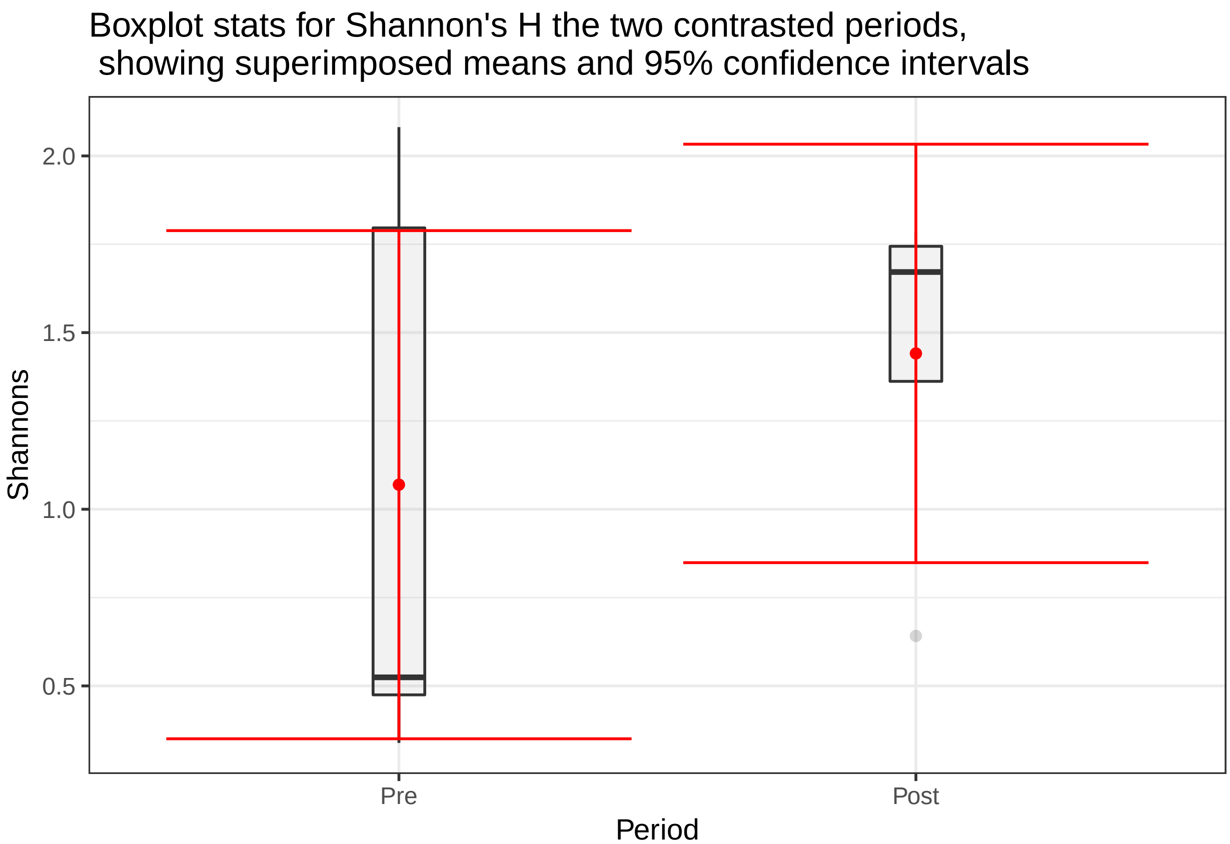
$H = -\sum_{i=1}^{N} p_i \ln(p_i)$

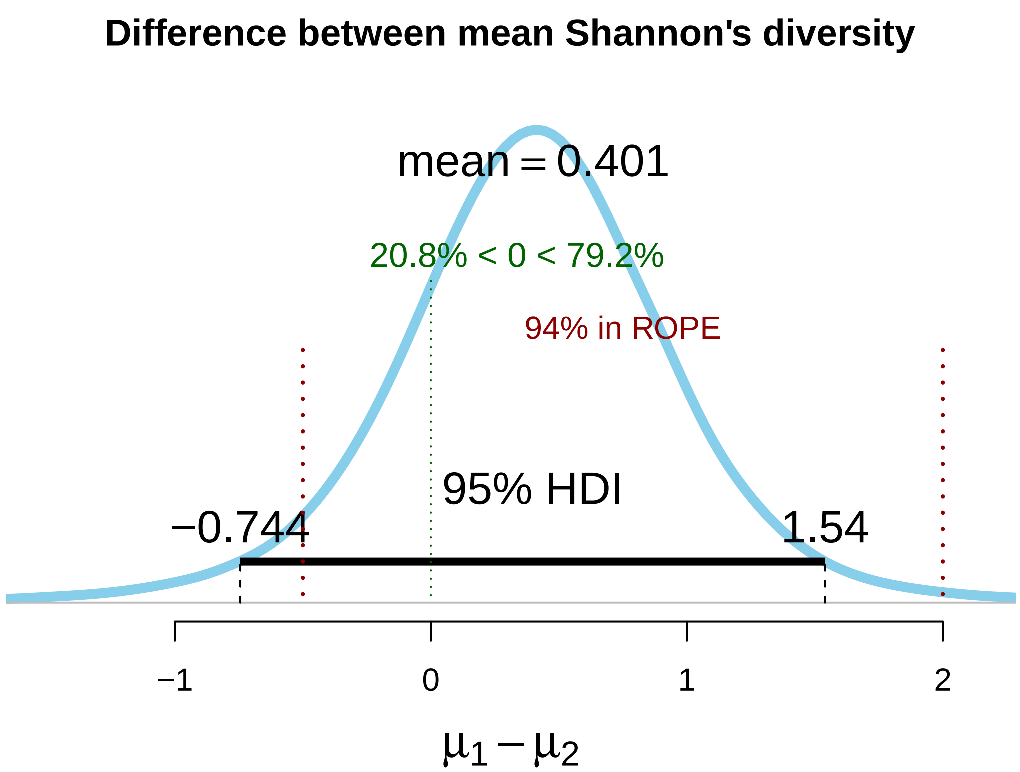
Where p_i is the proportional abundance of each species in an assemblage consisting of N species.

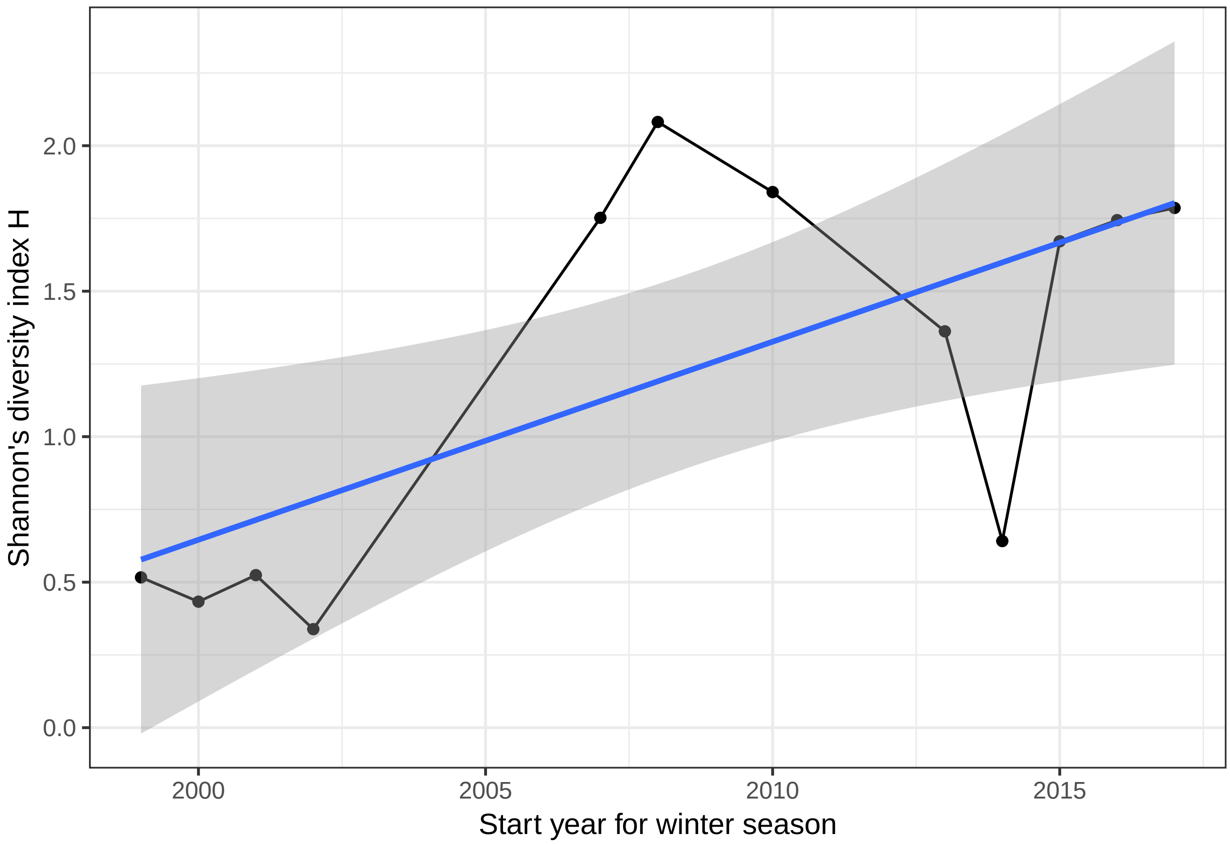
Shannon's index was calculated by transforming the table of counts into a matrix and applying the diversity function in the R package vegan {Oksanen et al. (2018)}



Changes in mean Shannon's index







Peak abundances

Methodology

The peak abundances table was derived from applying a consistent algorithm to the pooled data. The abundances were summed over all the sites for each month for all of the 19 species. The maximum count, rather then the low water count was used in the case of the observations of dunlin in 2017/18. This was considered to most likely match the method used for the data that was obtained from previous years surveys. The maximum species abundances in any month within the winter season was taken as the peak abundance. The peak assemblage in this case was calculated as the sum of the peak abundances over all the months. This Produces a rather higher estimate of the peak assemblage than the previous methodology. In both cases the proportional abundances of the species are calculated in realation to the sum. The two methodologies have been included in order to assess the sensitivity of the conclusions to the method used to calculate the peak assemblage. The substantive conclusions are unaltered and are not sensitive to the choice of methododlogy, although some of the quantitative results are slightly different.

Data

Counts data

Copy CSV Show 10 entries Search:

	Species	1999	2000	2001	2002	2007	2008	2010	2011	2012	2013	2014	2015	2016	2017
	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All
1	Avocet	399	7	379	323	1945	2300	260	361	2780	1378	913	1175	1507	1905
2	Bar-tailed godwit	2	7	0	0	1	240	10	1	1	1	70	3	39	24
3	Black-tailed godwit	53	0	470	57	1460	1800	600	720	2760	3602	2430	961	2919	5372
4	Cormorant	0	17	17	22	233	182	1	17	32	18	27	18	30	0
5	Curlew	8	29	63	17	75	74	40	119	173	163	123	77	183	275
6	Dark-bellied Brent goose	0	0	0	0	0	0	0	0	0	0	0	0	0	68
7	Dunlin	7750	7650	3677	9432	4105	1700	750	2722	5077	3580	6425	1024	5256	7140
8	Gadwall	0	0	0	0	0	0	0	0	0	0	0	0	0	29
9	Golden plover	0	0	0	0	0	0	0	0	0	0	2	0	360	470
10	Greater white-fronted goose	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Showing 1 to 10 of 28 entries Previous 1 2 3 Next

Table of percent contribution to assemblage

Copy CSV Show 10 entries Search:

	Species	1999	2000	2001	2002	2007	2008	2010	2011	2012	2013	2014	2015	2016	2017
	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All
1	Avocet	4.3	0.1	6.5	3	18.8	25.9	10.2	7.2	20.3	12	8.3	27.5	9.9	9
2	Bar-tailed godwit	0	0.1	0	0	0	2.7	0.4	0	0	0	0.6	0.1	0.3	0.1
3	Black-tailed godwit	0.6	0	8.1	0.5	14.1	20.3	23.5	14.4	20.1	31.4	22.1	22.5	19.2	25.4
4	Cormorant	0	0.2	0.3	0.2	2.3	2	0	0.3	0.2	0.2	0.2	0.4	0.2	0
5	Curlew	0.1	0.3	1.1	0.2	0.7	0.8	1.6	2.4	1.3	1.4	1.1	1.8	1.2	1.3
6	Dark-bellied Brent goose	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3
7	Dunlin	83.5	88.7	63.3	87.6	39.6	19.1	29.4	54.5	37	31.3	58.5	24	34.5	33.8
8	Gadwall	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1
9	Golden plover	0	0	0	0	0	0	0	0	0	0	0	0	2.4	2.2
10	Greater white-fronted goose	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Showing 1 to 10 of 28 entries Previous 1 2 3 Next

Peak assemblage values

Copy CSV Show 10 entries Search:

	Year	Sum
	All	All
1	1999	9284
2	2000	8628
3	2001	5812
4	2002	10766
5	2007	10355
6	2008	8887
7	2010	2554
8	2011	4994
9	2012	13718
10	2013	11455

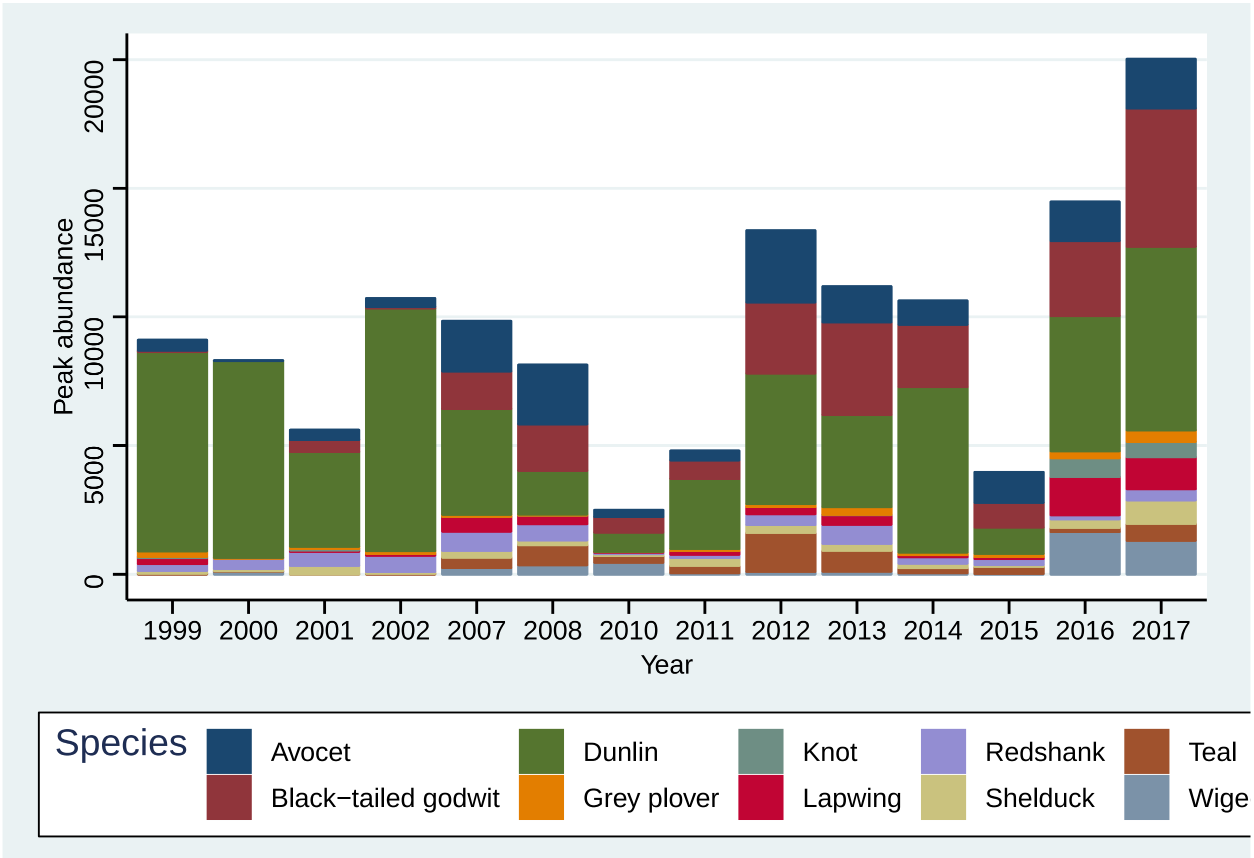
Showing 1 to 10 of 14 entries

Previous 1 2 Next

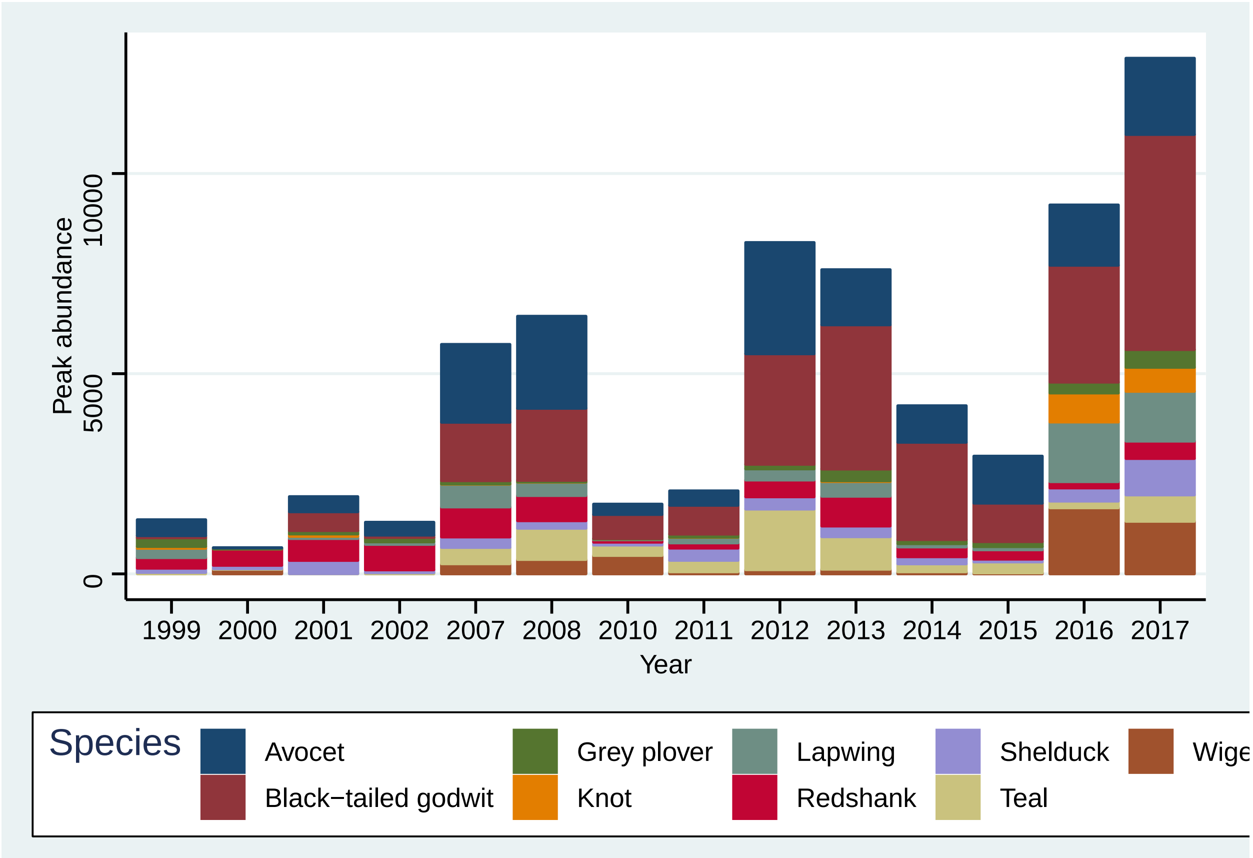
Barcharts

London Gateway

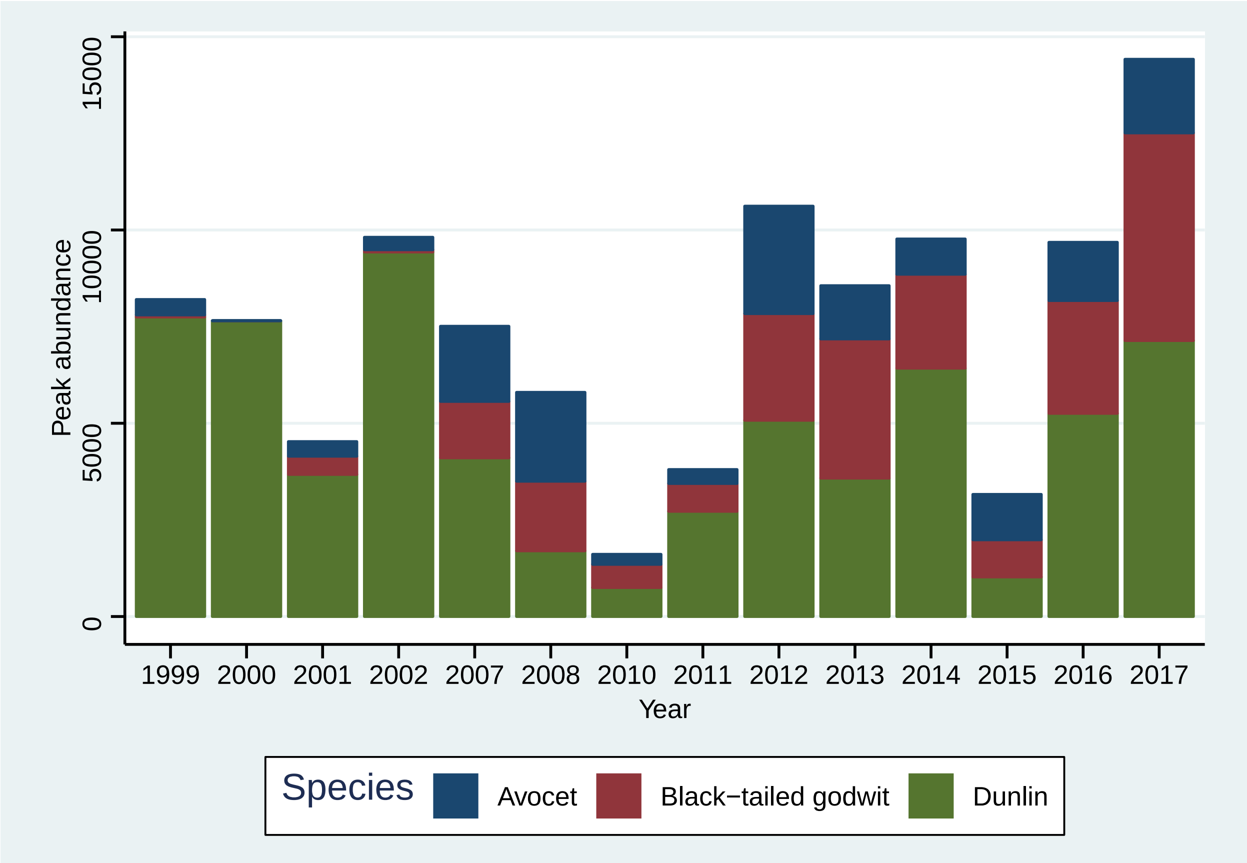
Top ten species



Excluding Dunlin



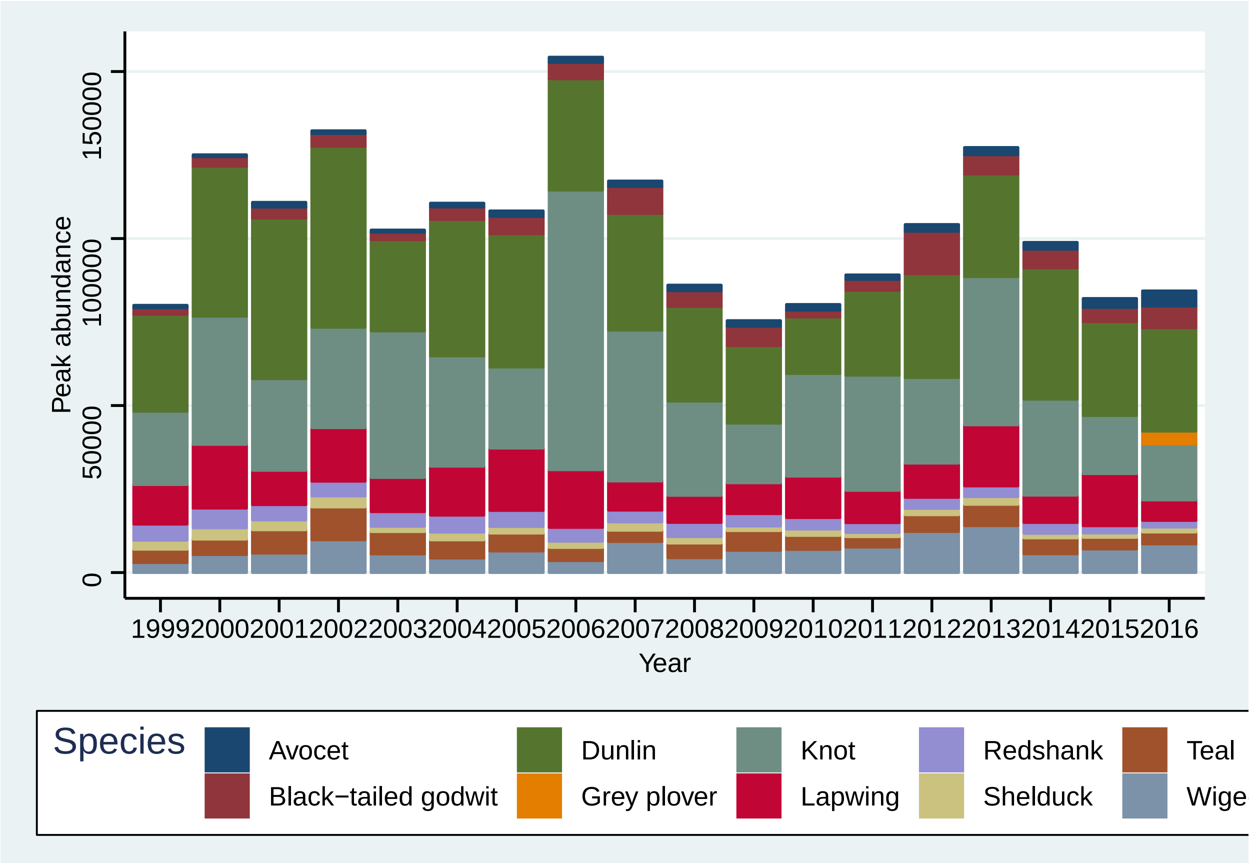
Top three species (key species)



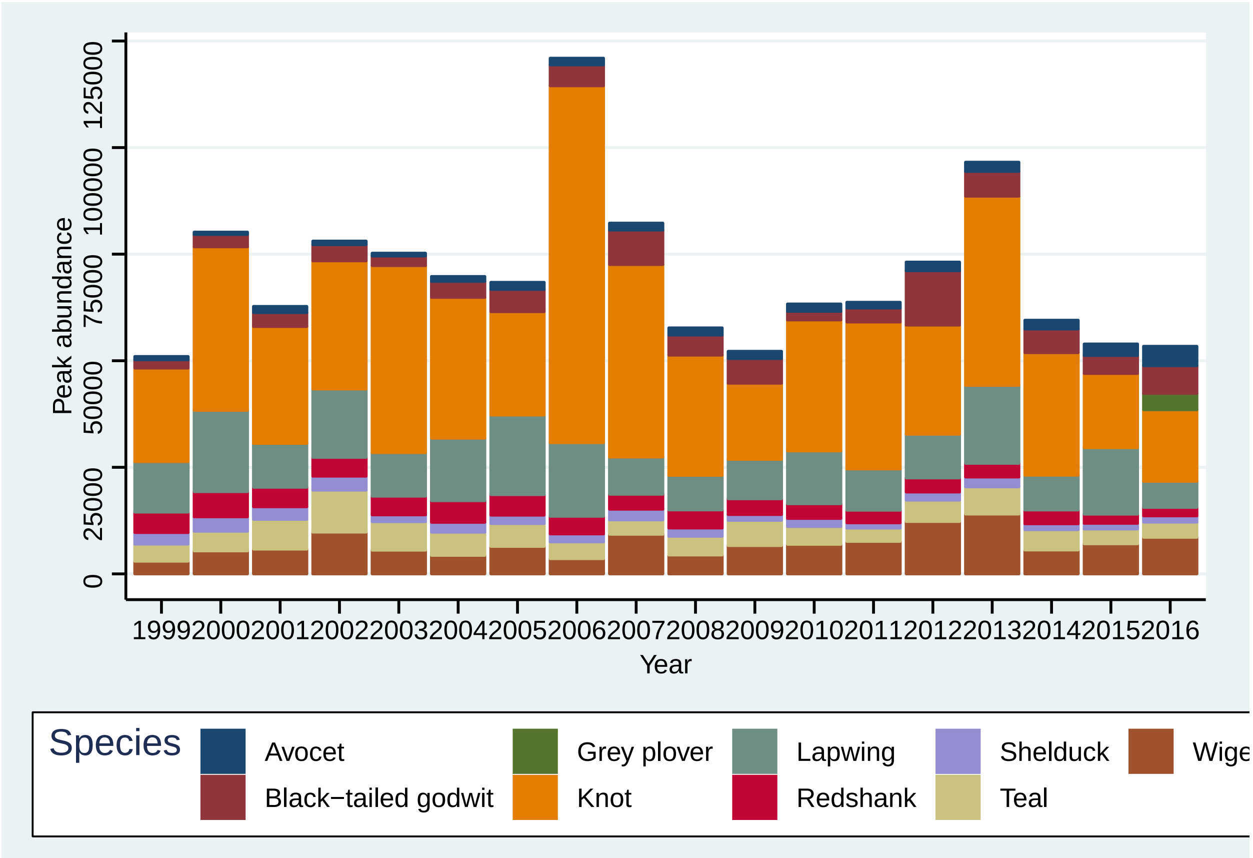
The top three most abundant species are Dunlin, Black-tailed godwit and Avocet.

Thames estuary

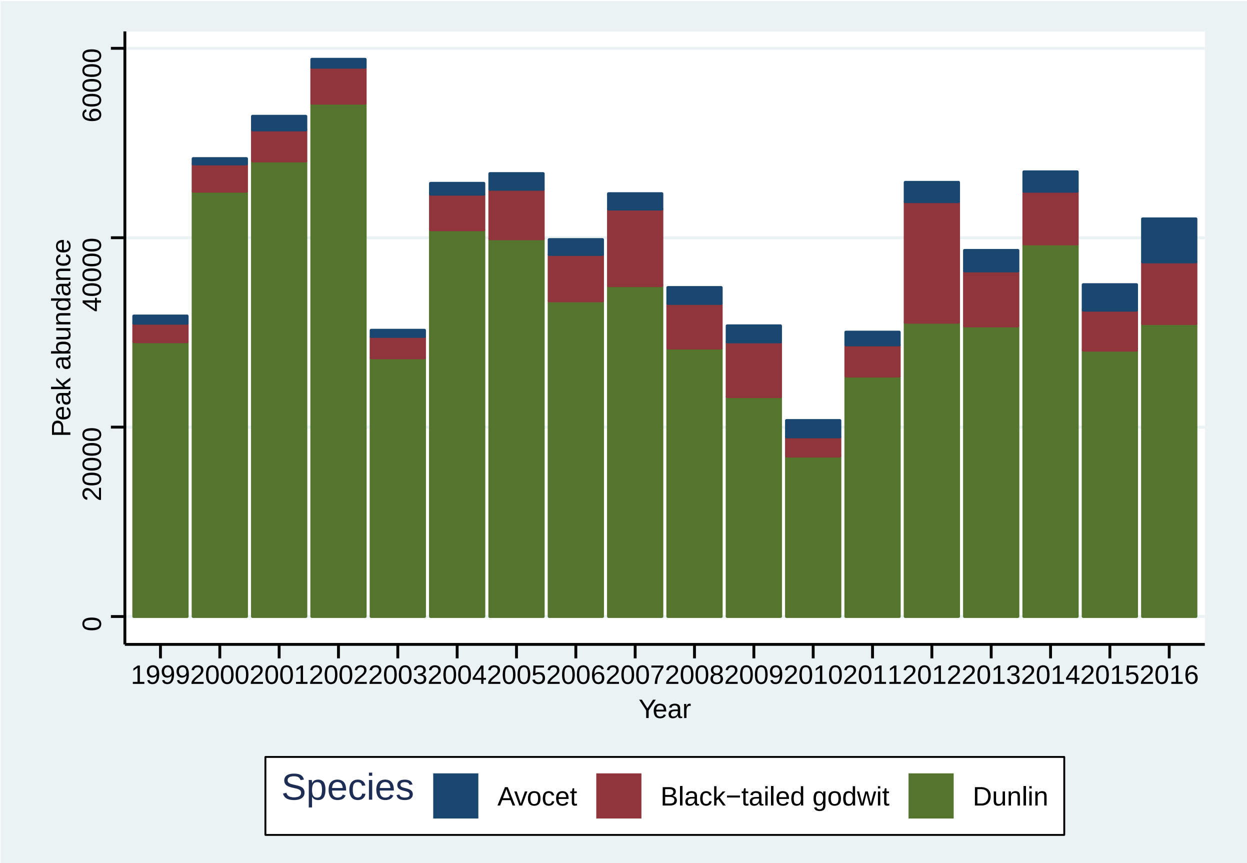
All species



Excluding Dunlin



Top three species (key species)



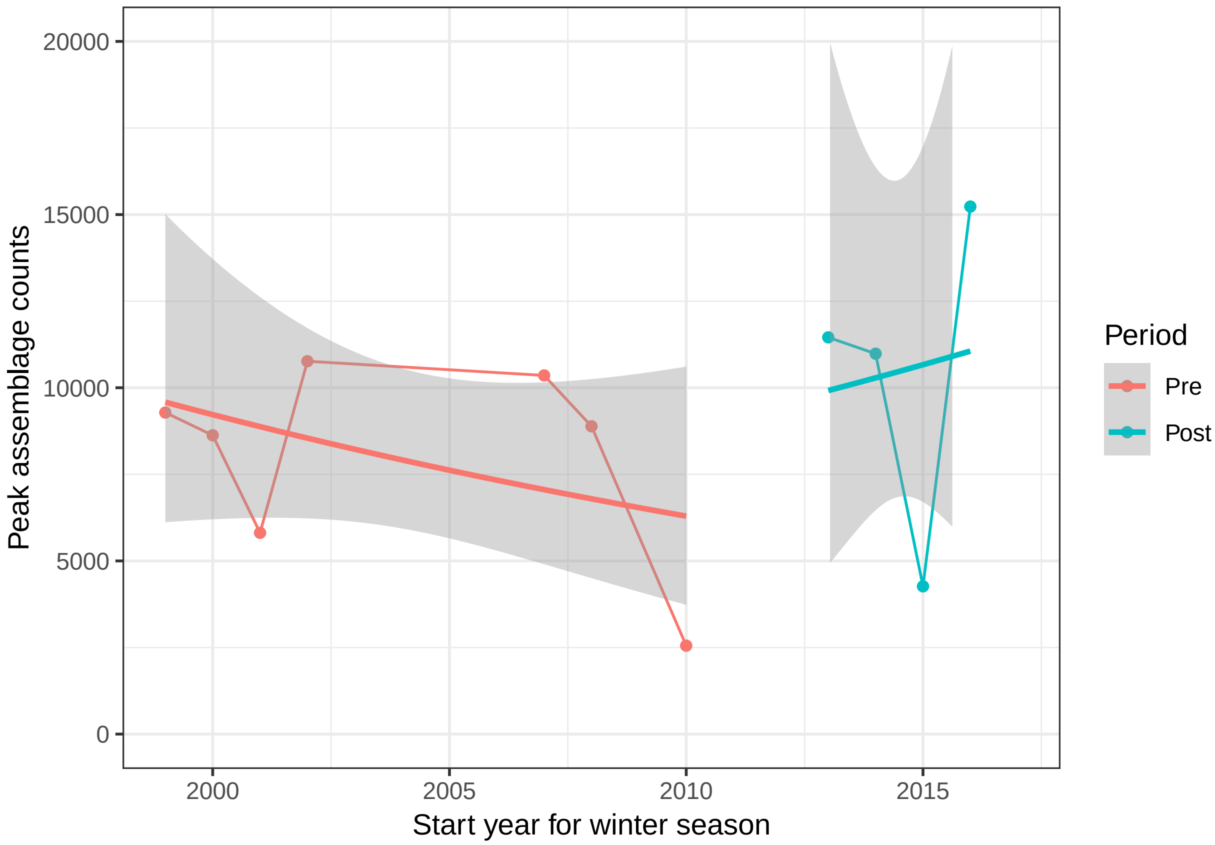
The top three most abundant species are Dunlin, Black-tailed godwit and Avocet.

Generalised linear models for trends

Methodology

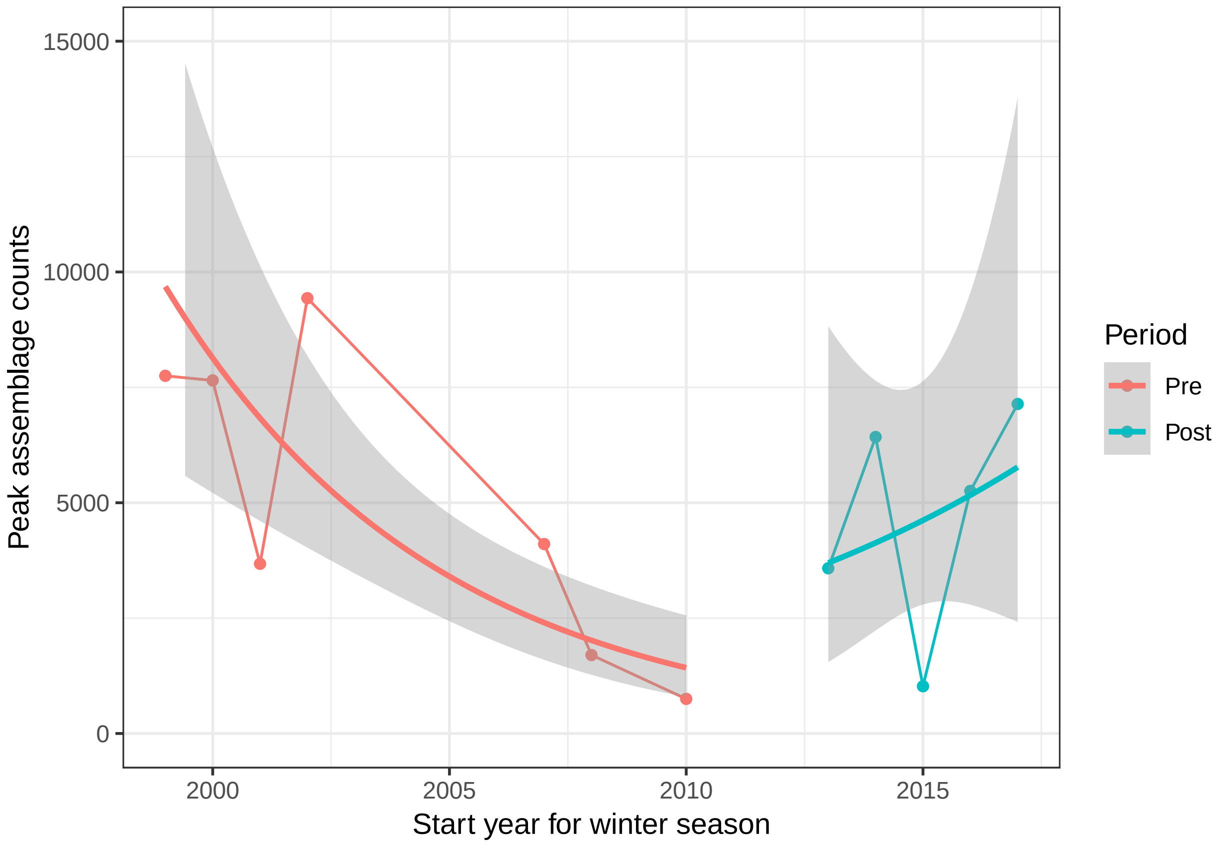
The use of regression as a method to formally analyse trends is greatly limited by the low sample size and by the high variability between count values. A longer time series of observations could show serial autocorrelation between the values forming a time series. Additional co-variates such as climatic effects might also be taken into account. As there are only a small number of data points available, a regression analysis could only provide evidence of a significant trend in the unlikely case of a monotonic year on year increase or decrease. In the case of counts that may take low numbers regression analysis using normally distributed errors may produce confidence intervals that fall below zero, which is impossible. Generalised linear models of the negative binomial family which account for overdispersion are prefereable in this case. Plotting lines derived from a negative binomial GLM with confidence intervals for the pre and post development periods shows no indication of such a consistent trend.

All species



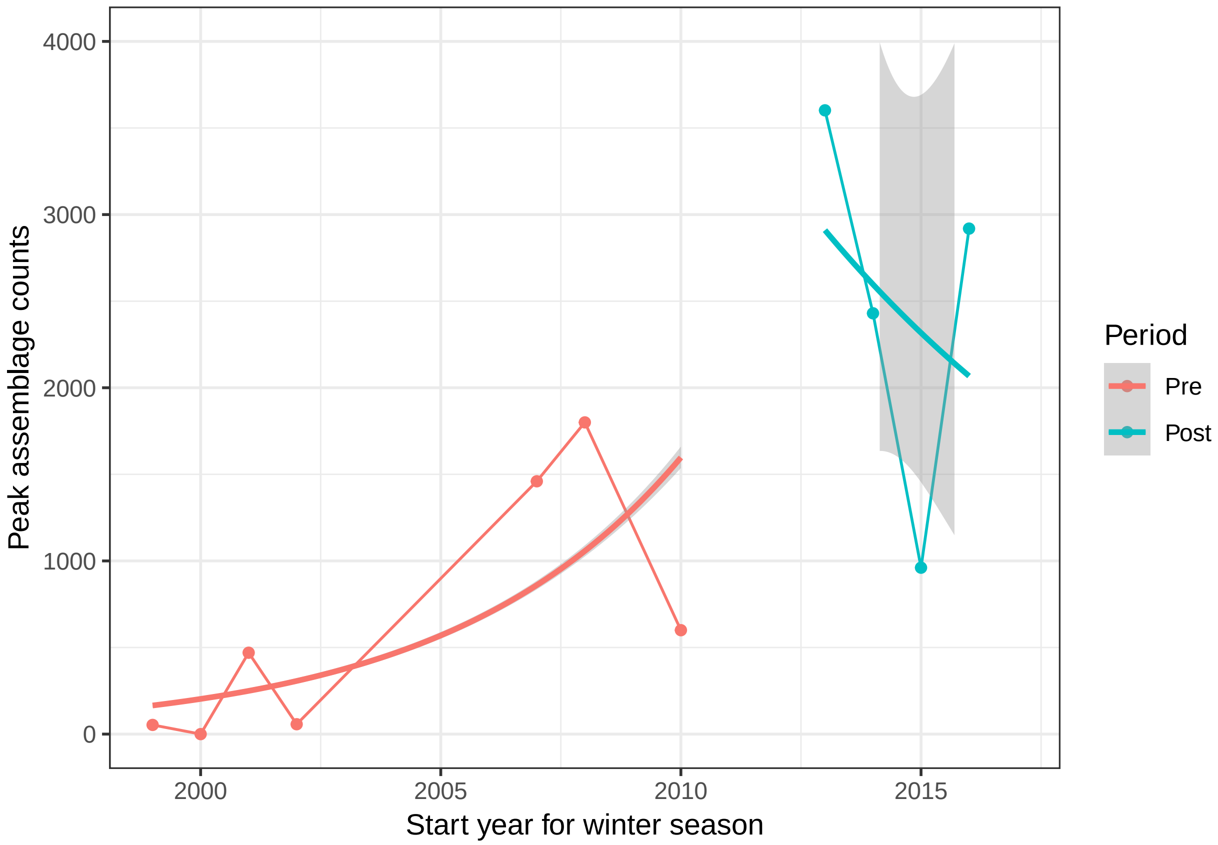
Peak assemblage counts
20000
15000
10000
5000
0
2000
2005
2010
2015
Start year for winter season
Period
Pre
Post

Dunlin

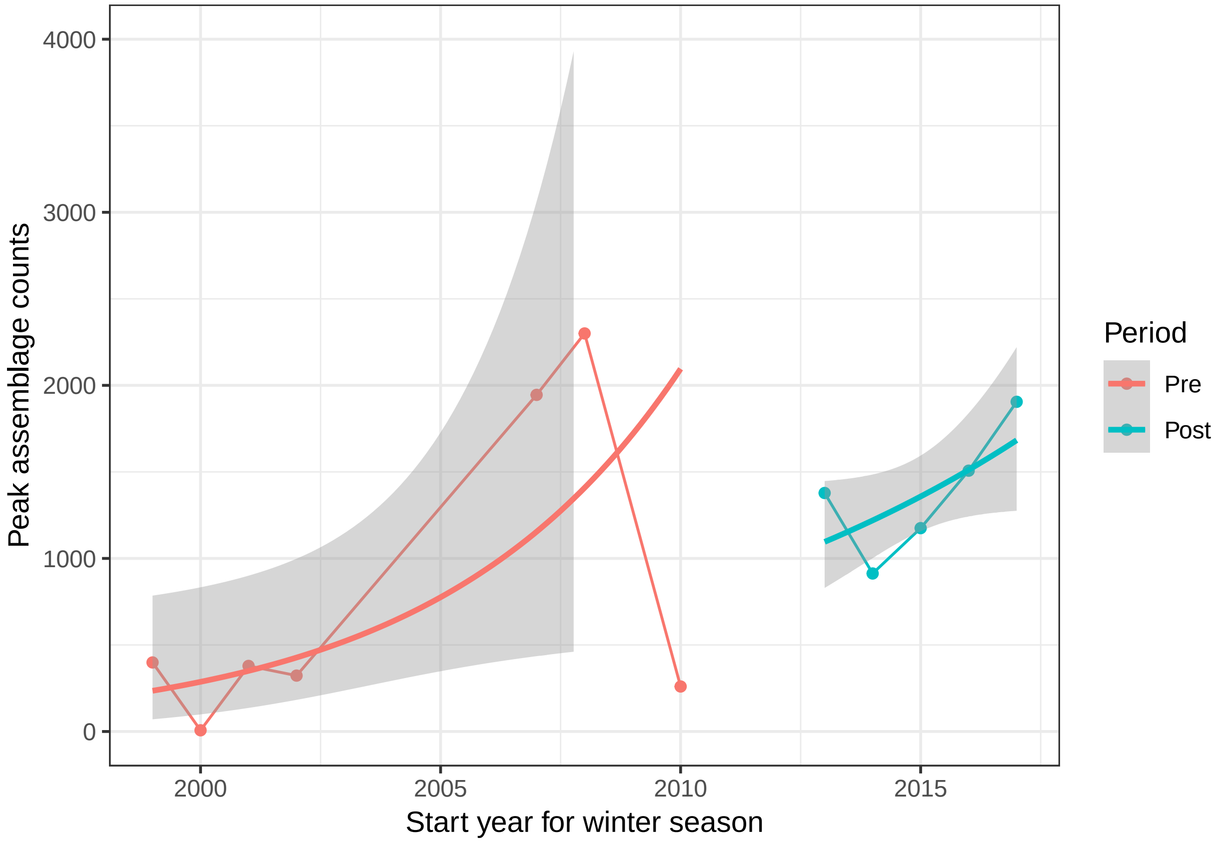


15000
10000
5000
0
Peak assemblage counts
2000
2005
2010
2015
Start year for winter season
Period
Pre
Post

Black tailed godwit



Avocet



Mean differences

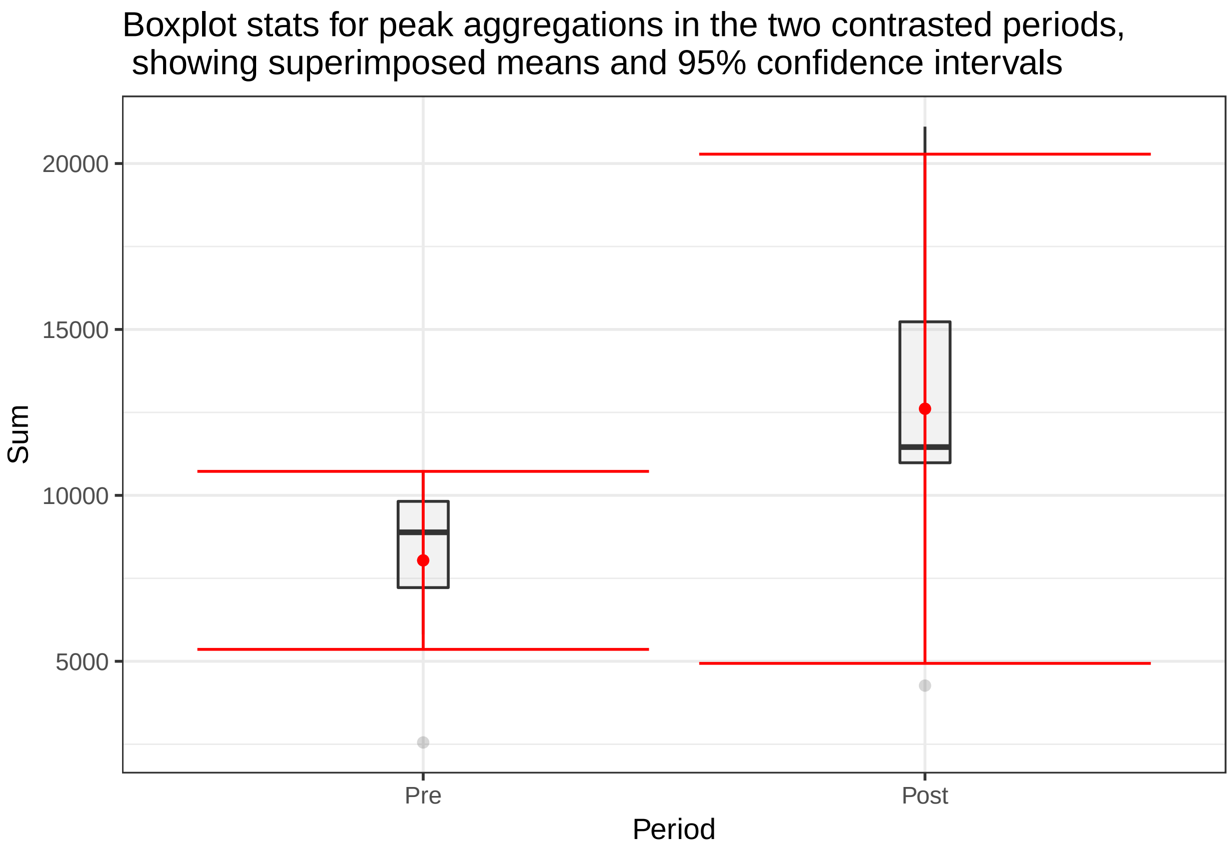
Methodology

As there is no evidence of consistent trends, the variability between years can be treated as if the were a set of independent observations. This does not imply that there is no underlying relationship between the total population of birds in consecutive, simply that random variability due to movements of flocks and changes in the observability of the birds in combination with stochastic population fluctuations are adequate explanations for the observed variability.

In this case the period becomes a factor with two levels. Data can be visualised as boxplots and means with confidence intervals calculated from the variability around the mean values.

Aggregation

Boxplots and confidence intervals



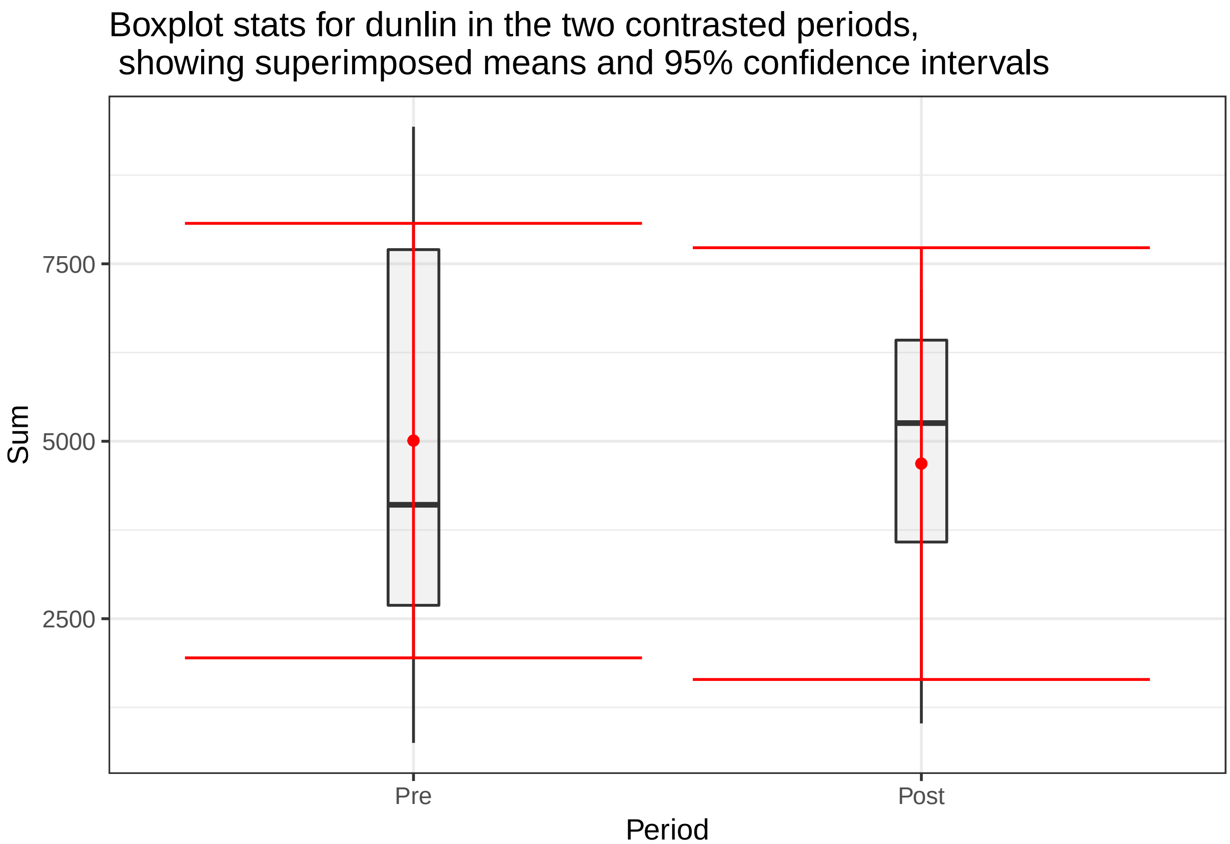
T-test

```
##
##  Welch Two Sample t-test
##
## data:  Sum by Period
## t = -1.5368, df = 5.2706, p-value = 0.182
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -12094.457   2956.971
## sample estimates:
##  mean in group Pre mean in group Post
##           8040.857          12609.600
```

The t-test provides no evidence of a significant difference between the mean peak assemblages pre and post works (p = 0.18).

Dunlin

Boxplots and confidence intervals



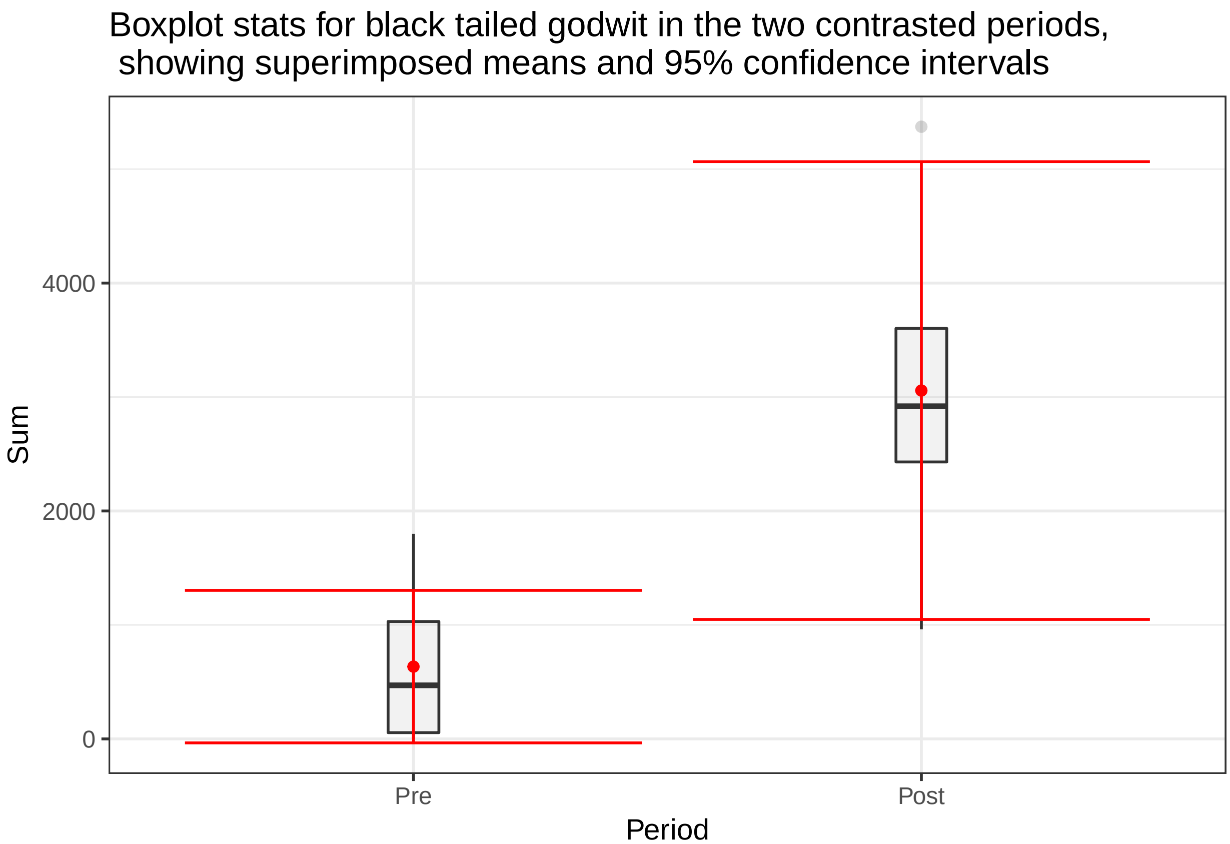
T-test

```
##
##  Welch Two Sample t-test
##
## data:  Sum by Period
## t = 0.19493, df = 9.952, p-value = 0.8494
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3383.466  4031.752
## sample estimates:
##  mean in group Pre mean in group Post
##           5009.143           4685.000
```

The t-test provides no evidence of a significant difference between the mean peak assemblages pre and post works ($p = 0.85$).

Black tailed godwit

Boxplots and confidence intervals



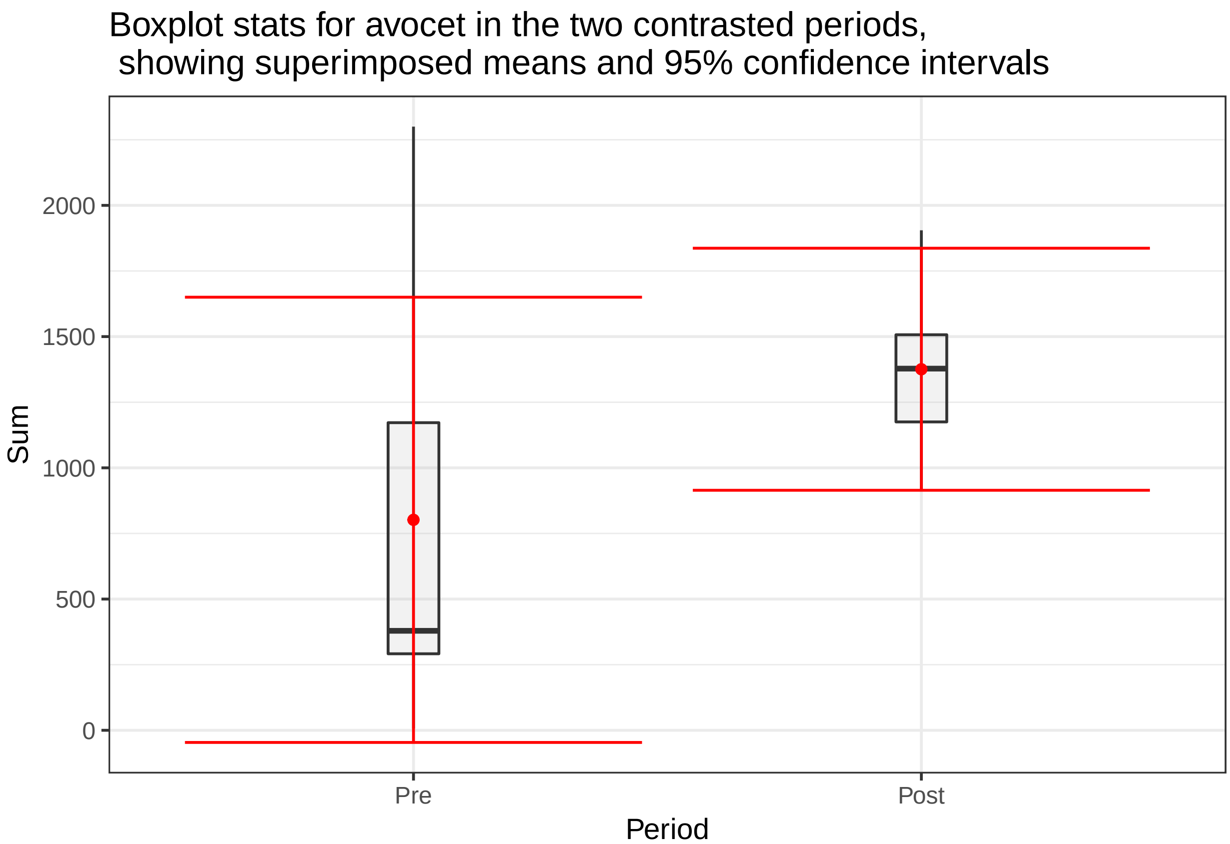
T-test

```
##
##  Welch Two Sample t-test
##
## data:  Sum by Period
## t = -3.1329, df = 5.1558, p-value = 0.02482
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4392.2623  -452.7663
## sample estimates:
##  mean in group Pre mean in group Post
##           634.2857          3056.8000
```

The t-test provides no evidence of a significant difference between the mean peak assemblages pre and post works (p = 0.02).

Avocet

Boxplots and confidence intervals



T-test

```
##
##  Welch Two Sample t-test
##
## data:  Sum by Period
## t = -1.4926, df = 8.4066, p-value = 0.1721
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1452.7407   305.2549
## sample estimates:
##  mean in group Pre mean in group Post
##           801.8571          1375.6000
```

The t-test provides no evidence of a significant difference between the mean peak assemblages pre and post works (p = 0.17).

Bayesian t.test

Methodology

There is an issue with regard to the interpretation of the result of such a test. Under null hypothesis significance testing (NHST) it is not possible to accept the null hypothesis of no difference. NHST simply fails to reject the null. The p value represents the probability of obtaining the data, or data more extreme, given that the null hypothesis is in fact true. However the precise point null hypothesis of exactly no difference between assemblage counts before and after the works is not a credible one. There must be some differences.

The issue is whether any differences fall within acceptable bounds. Thus NHST is problematic when the decision rule in question involves looking at the evidence **in favour** of the null hypothesis. This is the case here as shown in the description of the decision rule provided in the report.

*The initial target against which the success of the mitigation and compensation will be assessed shall be that the sites in combination support an assemblage of wintering waterfowl at low tide comprising, on a 5-year mean peak basis at least **7900 birds** made up of, in particular, avocet, dunlin and black-tailed godwit in similar proportions to those supported by North Mucking during the winters of 1999/2000 to 2002/2003 (considered in the context of the wider population trends)*

The alternative to NHST is to adopt a Bayesian approach to inference. Under this approach he 5 year means for peak abundances are not considered to be fixed quantities, but are themselves treated as random variables with distributions. Bayes' formula provides a formal mechanism of providing probabilities for unknown quantities of interest. In this case the difference between μ_1 and μ_2 (the mean peak assemblages before and after the works) is an unknown quantity.

Bayes theorem states.

$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$ Where $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, v)$

So, the posterior credibility of the combination of values for $(\mu_1, \mu_2, \sigma_1, \sigma_2, v)$ is the likelihood of that combination times the prior credibility of the combination, divided by the constant p(D). When it is assumed that the data are independently sampled, the likelihood is the multiplicative product across all the data values of the probability density of a t distribution. The prior is the product of the five independent parameter distributions. The constant p(D) is the marginal likelihood, which may be obtained by integrating the product of the likelihood and prior over the entire parameter space. This integral is difficult to compute analytically. This difficulty limited the application of Bayesian methods before computational solutions using simulation became available. However this limitation no longer exists. It is now computationally simple to fit the true Bayesian model using tools supplied through R (Plummer 2018). This allows the full posterior distributions of the parameter values to be obtained, leading to a richer and more informative analysis (Kruschke 2013).

Providing that uninformative prior probabilities for the parameters are used, applying Bayes theorem in the context of a t-test will then provide credible intervals for the differences between means (Edwards 1996). Although the estimates may be numerically very similar to those derived from the confidence intervals of a traditional t-test, the interpretation of the result now directly maps onto the required decision rule.

The traditional t-test found that the best estimate for the pre-works mean as 802 and the post works mean as 1376 giving a point estimate difference between the means of 574.

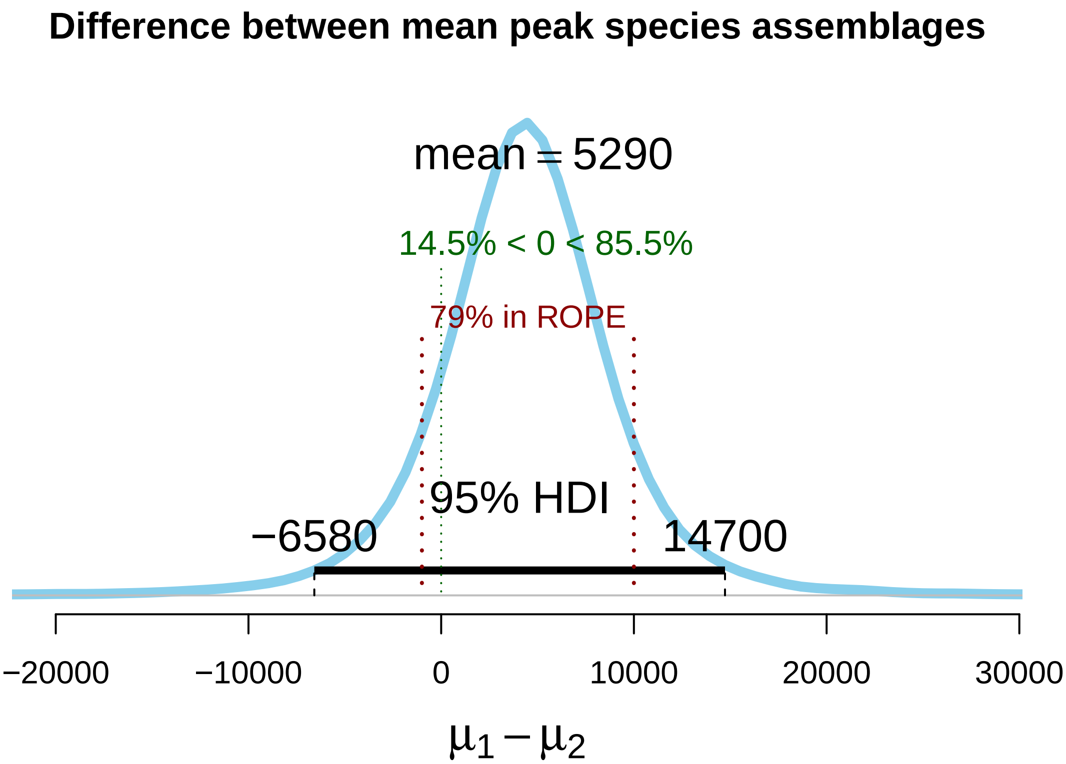
The original target value of 7900 for the overall assemblage were derived from low water count data for the four winter periods 1999/2000 to 2002/2003. The pre works data used in the t-test included some additional observations as, These observations are helpful in establishing the range of variability for inference so have been included. If it were desirable these values could be excluded and the analysis re-run without them.

Bayesian model fitting allows a formal evaluation of a decision rule based on the concept of the **region of practical equivalence** (ROPE). This is an area around the null value of no difference which encloses those values of the parameter that are deemed to be not importantly different from the null value for the practical purposes of the study.

In this case any increase in assemblage numbers, even if not statistically significant, are of no practical importance in evaluating whether Clause 10.5.4 has been met. The ROPE can therfore extend to the right almost indefinitely. The choice of a left hand boundary for the ROPE has to be considered through a careful evaluation of the available data. The target value was originally set at around 8000 birds. This is around 1000 higher than the first estimate of the pre-works mean. It would thus seem reasonable to set a ROPE lying between -1000 and 1000.

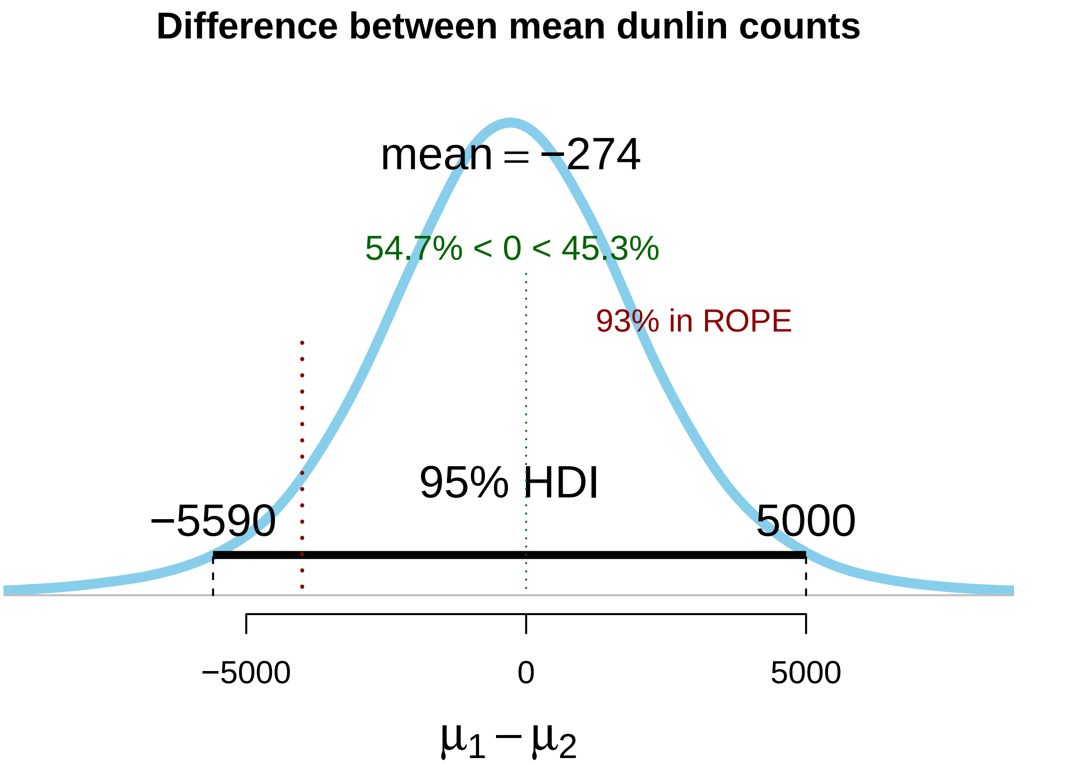
The bayesian t-test is then run using the package BEST {Kruschke and Meredith (2018)} in R. The model used completely non-informative vague priors for the parameters of interest in order to avoid subjectivity. The resulting simulation provided the full posterior distribution for the differences between the two means.

Aggregation



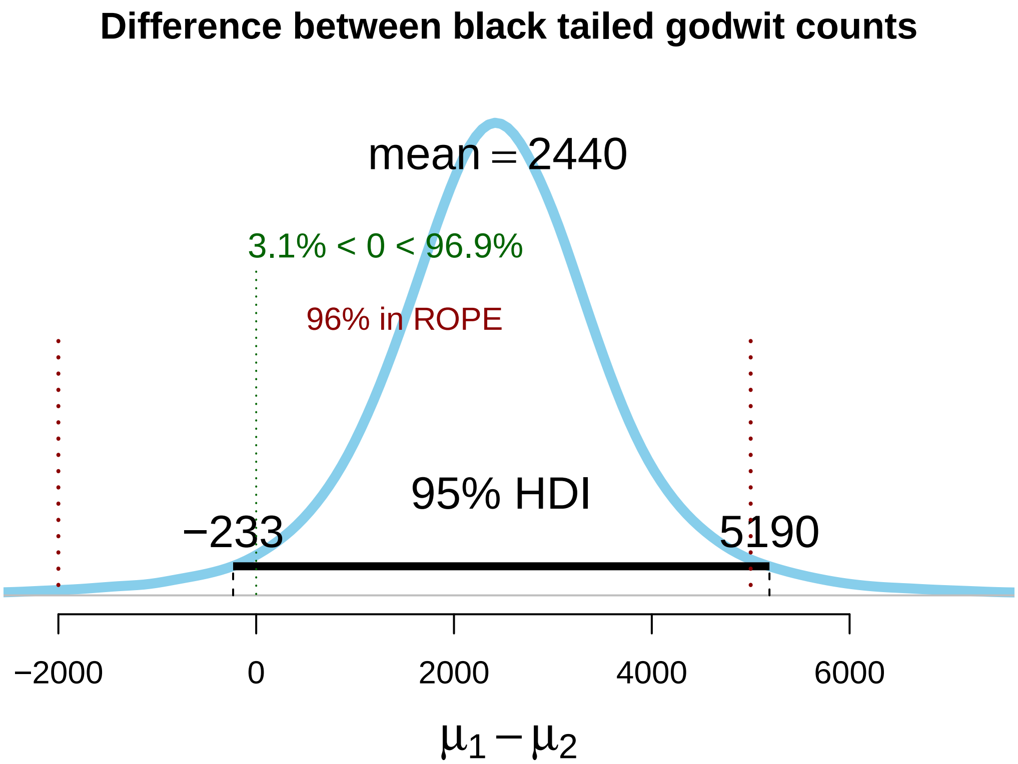
The figure shows the full posterior distribution for the difference between the two means, which is treated as a random variable and takes a t-distribution. The interpretation of the figure in terms of a decision rule of the Bayesian t-test analysis is clear when the ROPE is superimposed on the distribution. Around 80% of the posterior distribution for the difference between the two means lies within the ROPE. Although this does imply that there is a 20% chance that the value could stil lie outside the ROPE, the analysis is still being based on limited data. Intuitively it would be impossible to decide with certainty that the criteria had been met based on many fewer data points. The analysis formalises the strength of the currently avaialable evidence. As more data becomes available the probability that the criteria would be met becomes higher. Bayesian analysis allows for updating posterior distributions through additional data.

Dunlin



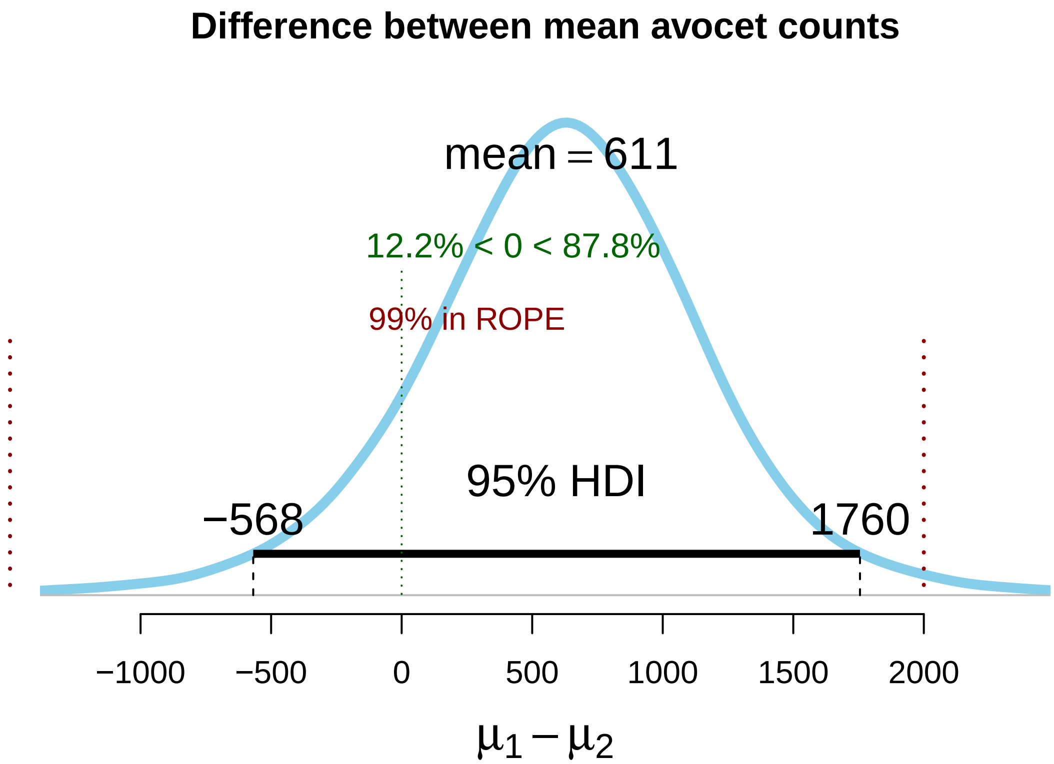
The analysis shows that around 90% of the posterior distribution falls within the ROPE. Thus there is very strong evidence that the works have had no practical impact on overall dunlin numbers.

Black tailed godwit



In this case a small amount (around 2%) of the ROPE actually falls below the posterior 95% highest density interval for the differences between the two means. The practical equivalence criteria is met to a very high degree of certainty, given the addditional evidence that black tailed godwit numbers have significantly increased over the period from 1998.

Avocet



The practical equivalence criteria is again met to a very high degree of certainty, given the addditional evidence that avocet numbers have significantly increased over the period from 1998.

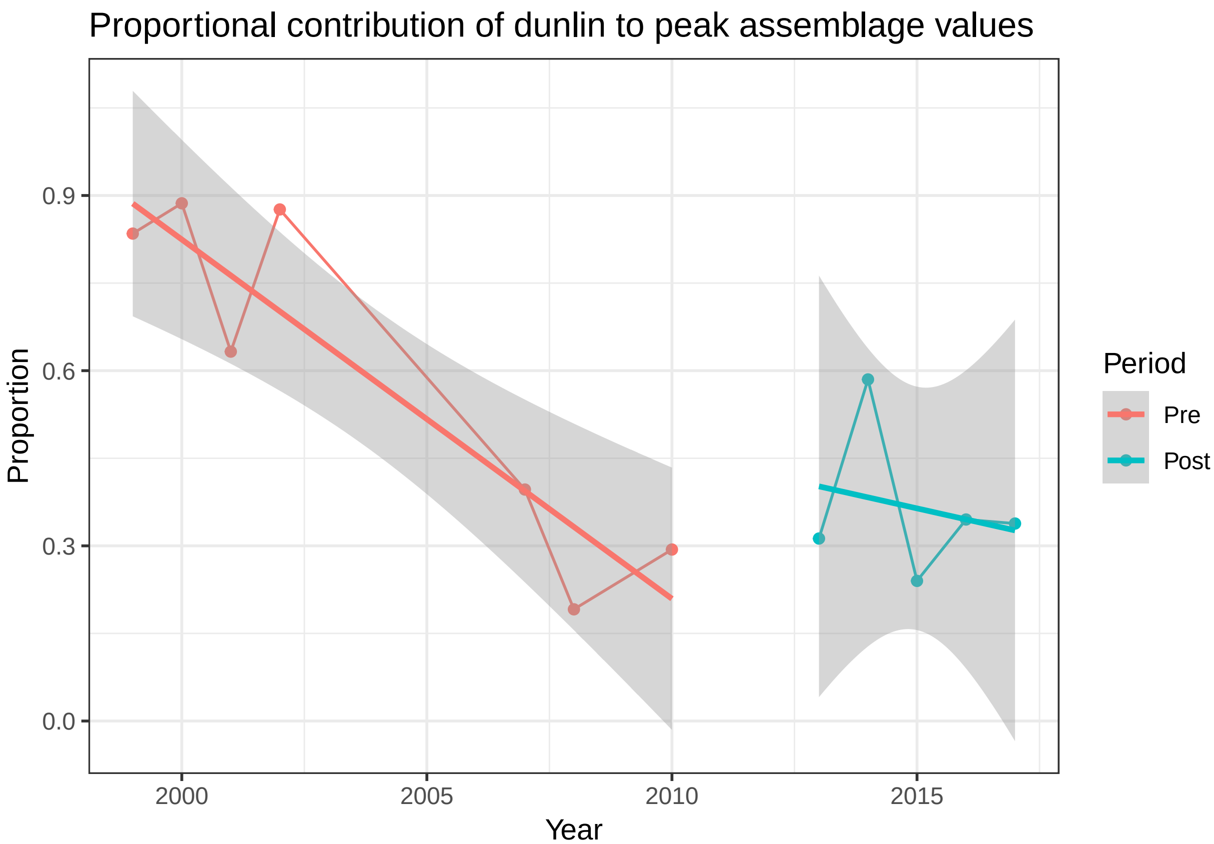
Differences in proportional abundance

Explanation

The target goal was also stated in terms of proportional abundance. *At least 7900 birds made up of, in particular, avocet, dunlin and black-tailed godwit in similar proportions to those supported by North Mucking during the winters of 1999/2000 to 2002/2003*

Inspection of the stacked bar charts and the raw data shows that the proportion of dunlin in the assemblage was higher between 1999 and 2002 than at present. As dunlin are small common waders a decrease in their proportional contribution would be interpreted as a positive effect, rather than a negative one.

Changes in proportional abundance of dunlin.



Trend analysis for proportional abundance of dunlin using beta regression

In order to establish the signficance of the change generalised linear modelling based on the beta distribution would provide the most robust approach. Proportions cannot be modelled with normally distributed errors. The betareg package in R allows this {Grün et al. (2012)}

Yearly trend

```
##
## Call:
## betareg(formula = Proportion ~ Year, data = dunlin)
##
## Standardized weighted residuals 2:
##     Min      1Q  Median      3Q     Max
## -2.1505 -0.7284  0.0682  0.8100  1.5386
##
## Coefficients (mean model with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 245.94963   65.31655   3.766 0.000166 ***
## Year         -0.12246    0.03252  -3.766 0.000166 ***
##
## Phi coefficients (precision model with identity link):
##       Estimate Std. Error z value Pr(>|z|)
## (phi)    8.387      3.272   2.564   0.0104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 6.138 on 3 Df
## Pseudo R-squared: 0.5831
## Number of iterations: 2563 (BFGS) + 8 (Fisher scoring)
```

Beta regression produces evidence of a statistically signficant (p=0.012) reduction in the proportional contribution of dunlin to the species assemblage between 1999 and present. However the trend occurred prior to the works cmencing.

Changes in species diversity

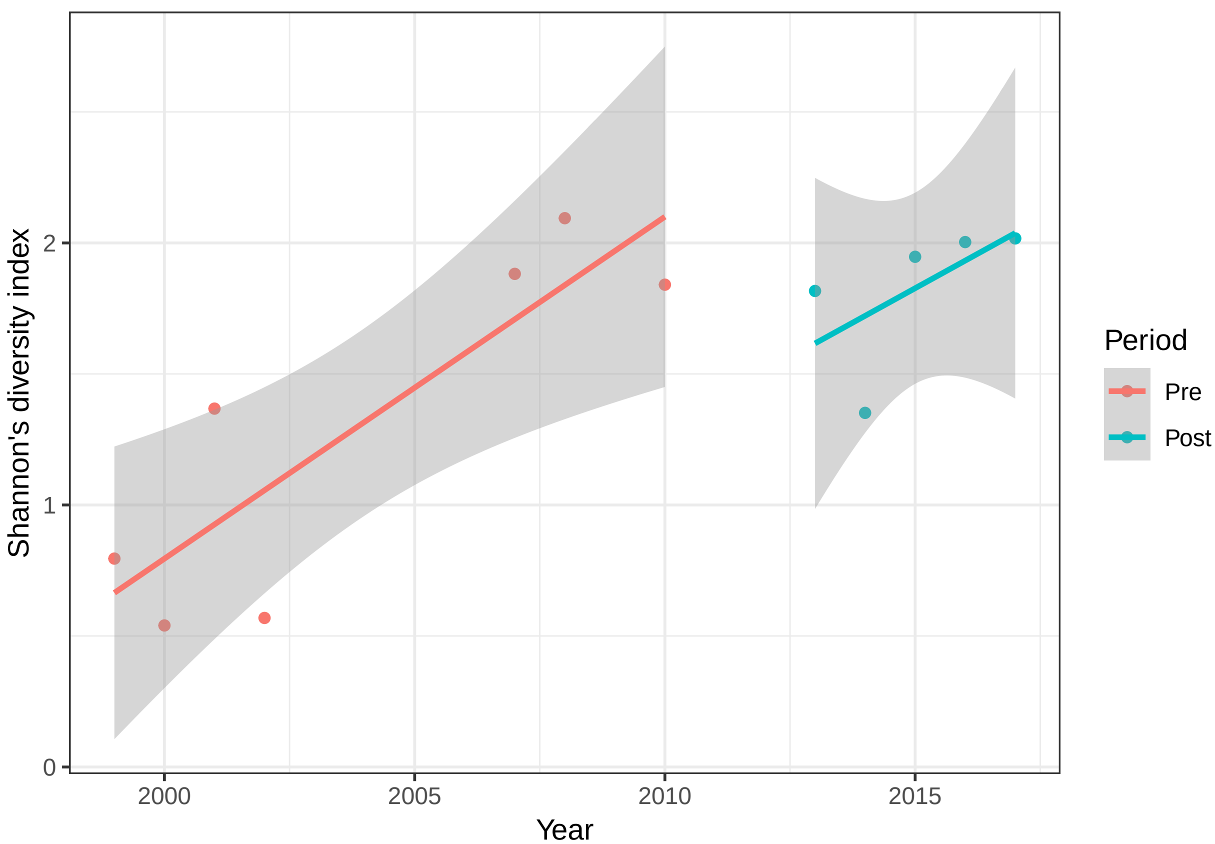
Although the criteria used to evaluate the impact of the works aimed to ensure a comparable mix of species abundances, the decline in relative abundance of dunlin and the increase in the relative contribution of other species may have increased species diversity. This is generally considered to be a positive outcome for conservation.

In order to evaluate changes in species diversity a commonly used diversity index was calculated for the assemblage. Shannnon's index is based on proportional contributions of each species to the assemblage.

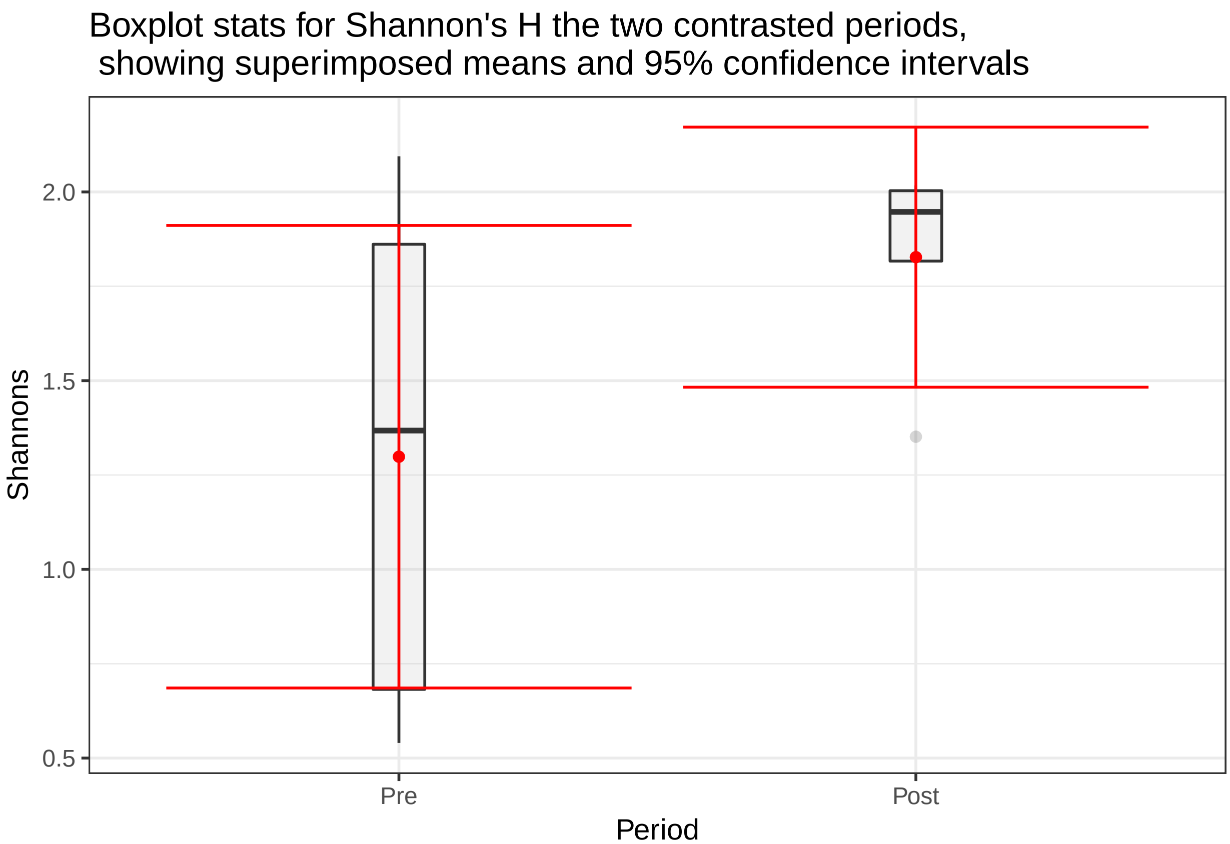
$H = -\sum_{i=1}^{N} p_i \ln(p_i)$

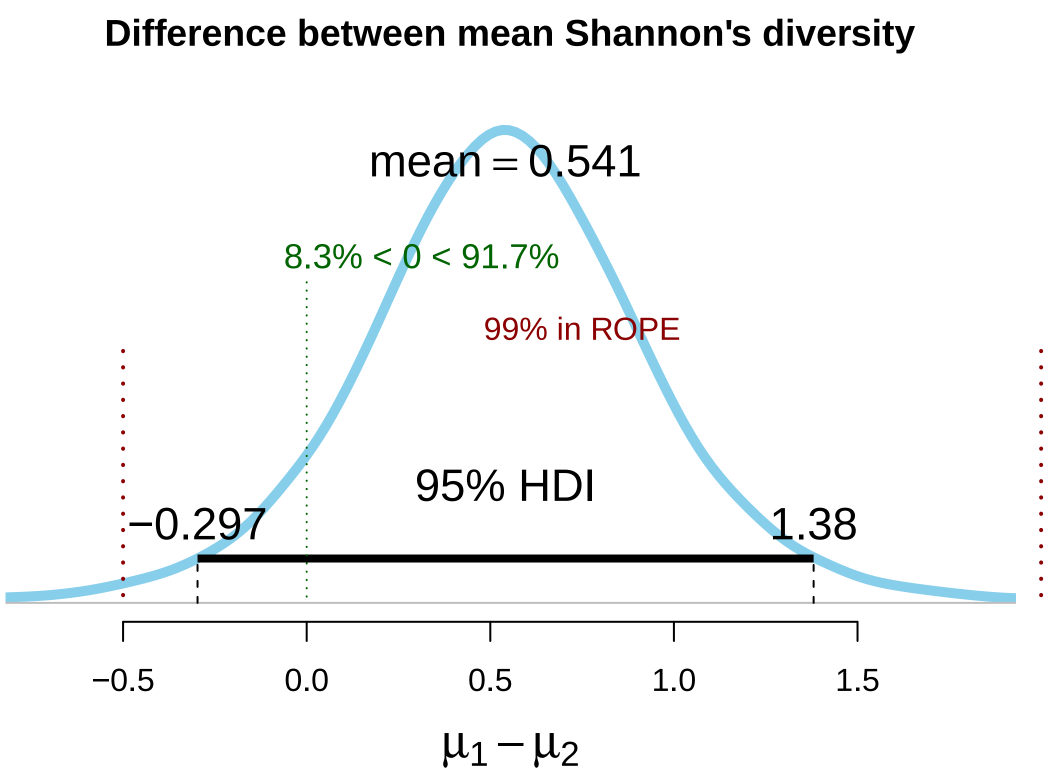
Where p_i is the proportional abundance of each species in an assemblage consisting of N species.

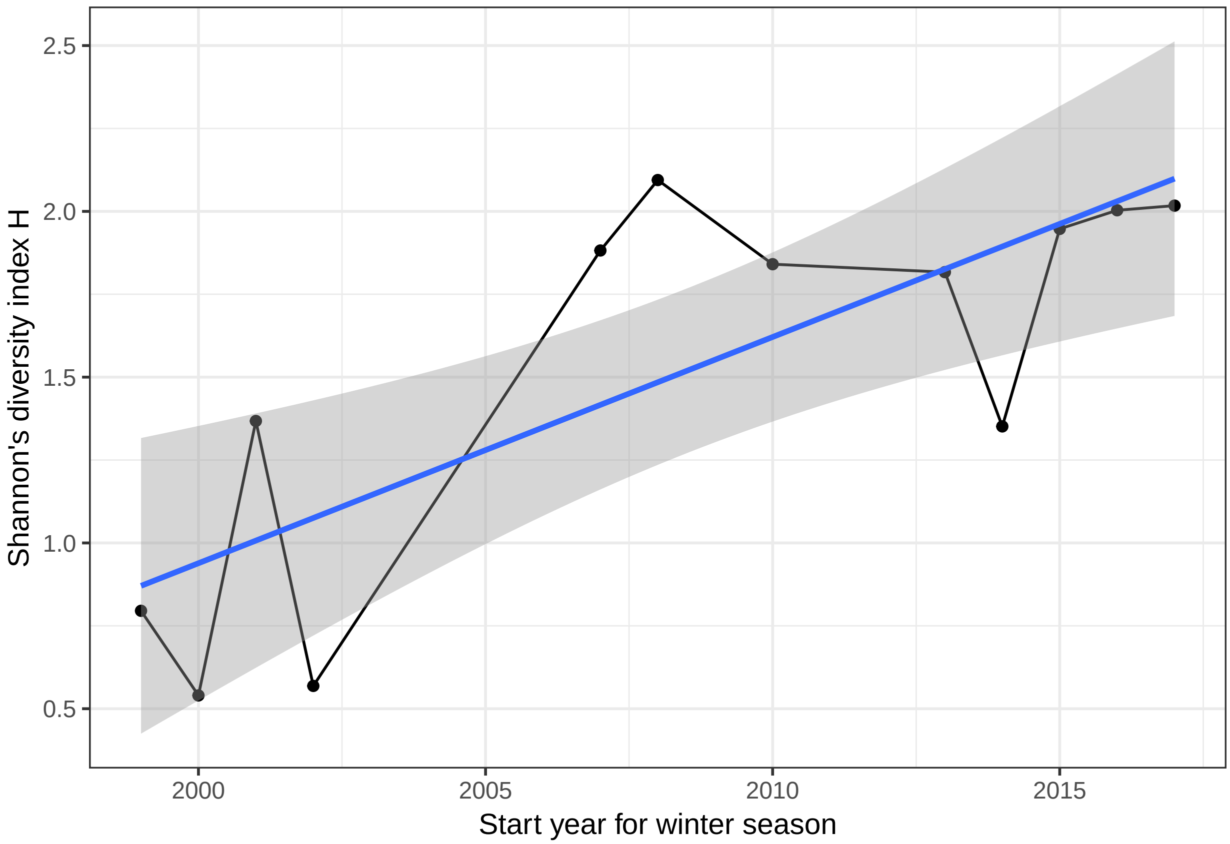
Shannon's index was calculated by transforming the table of counts into a matrix and applying the diversity function in the R package vegan {Oksanen et al. (2018)}



Changes in mean Shannon's index



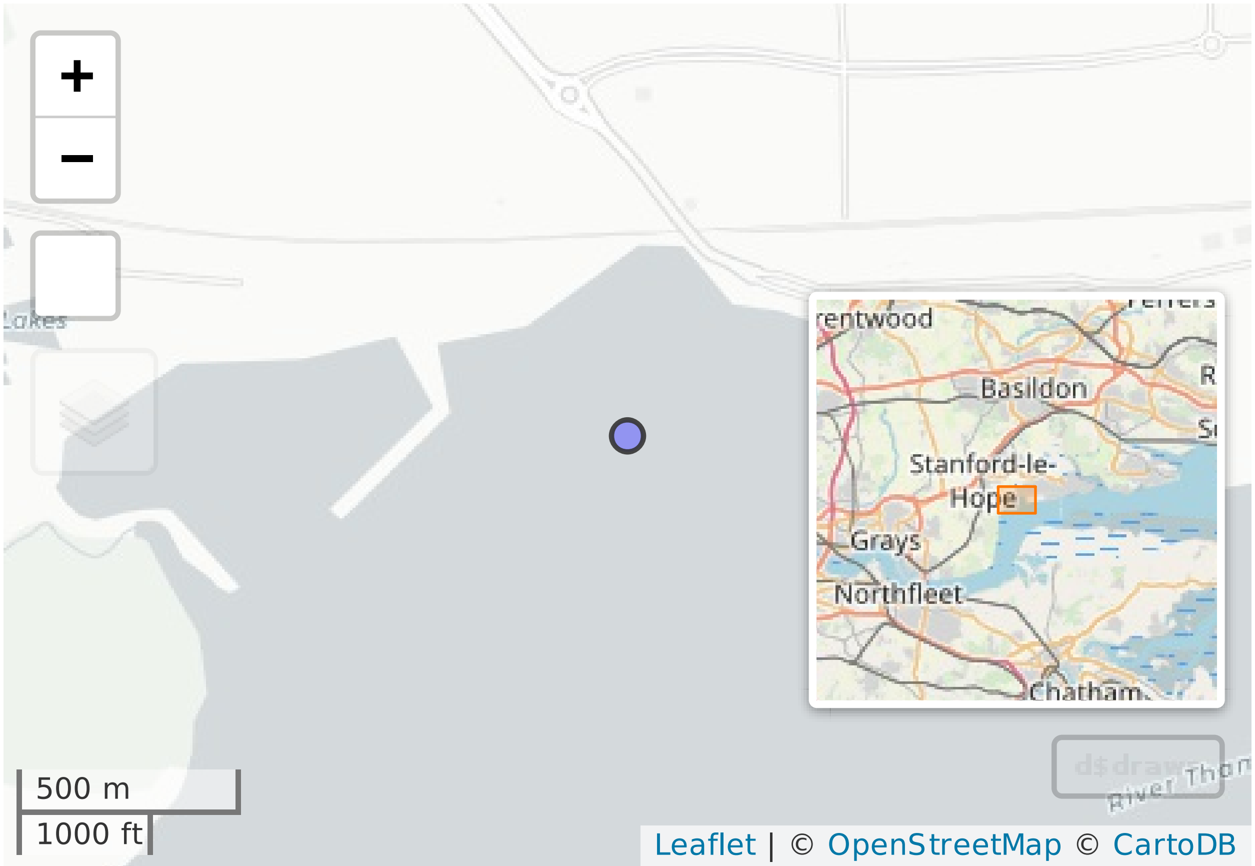




2.5
2.0
1.5
1.0
0.5
Shannon's diversity index H
2000
2005
2010
2015
Start year for winter season

Map of the area

Dynamic in HTML



Example screen shot for PDF

Example screen shot 2 for PDF

References

Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. Null hypothesis testing: Problems, prevalence, and an alternative. JOURNAL OF WILDLIFE MANAGEMENT 64:912–923.

Colegrave, N., and G. D. Ruxton. 2003. Confidence intervals are a more useful complement to nonsignificant tests than are power calculations. BEHAVIORAL ECOLOGY 14:446–4.

Edwards, D. 1996. Comment: The first data analysis should be journalistic. ECOLOGICAL APPLICATIONS 6:1090–1094.

Ellison, A. M. 2004. Bayesian inference in ecology. Ecology Letters 7:509–520.

Grün, B., I. Kosmidis, and A. Zeileis. 2012. Extended beta regression in r: Shaken, stirred, mixed, and partitioned. Journal of Statistical Software 48:1–25.

Kruschke, J. K. 2013. Bayesian estimation supersedes the t test. Journal of Experimental Psychology:

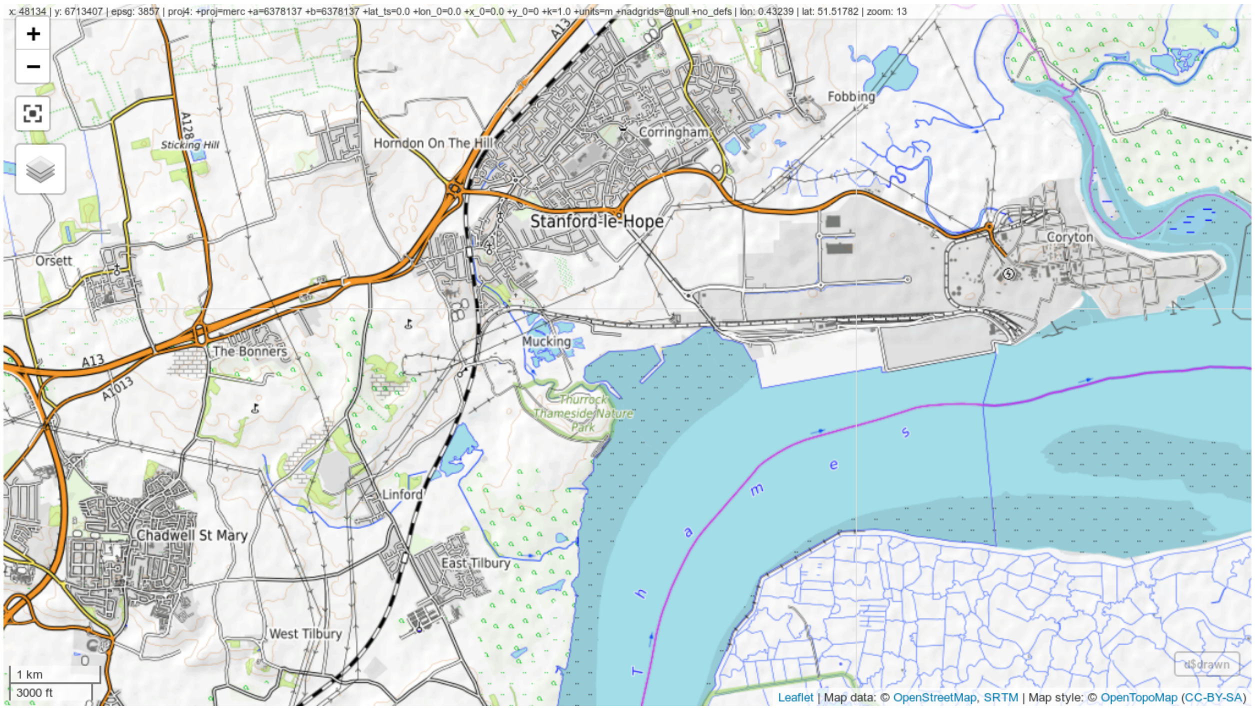


Figure 1:

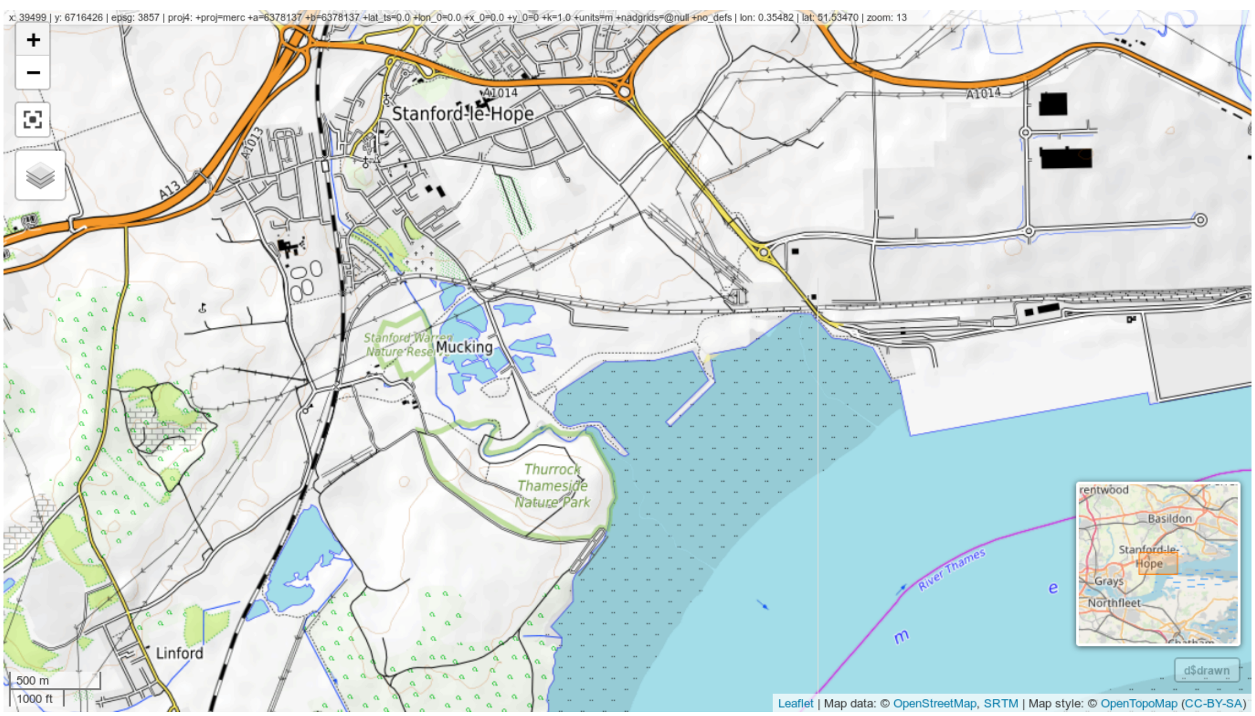


Figure 2:

General 142:573–588.

Kruschke, J. K., and M. Meredith. 2018. BEST: Bayesian estimation supersedes the t-test.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2018. Vegan: Community ecology package.

Plummer, M. 2018. Rjags: Bayesian graphical models using mcmc.

RUBIN, D. 1984. BAYESIANLY justifiable and relevant frequency calculations for the applied statistician. ANNALS OF STATISTICS 12:1151–1172.