Length-weight analysis for southern hake stock

Cousido-Rocha, M., Izquierdo, F., Mendes, H., Silva, C., Silva, A.V., Cerviño, S.

2022-03-09

Contents

Nowaday, the length-weight relationship carried out in 1999 is used. This relationship provides global (not year specific) estimates of the a and b parameters ($W_i = aL_i^b$, $i = 1, \ldots, N$, being N the total sample size). More precisely, the actual values are $a = 0.00659$ and $b = 3.01721$. Hence, in the current study a review of length-weight relationship is done addressing the following tasks:

- Estimation of global (not year specific) a and b using the updated length-weight data base.
- Estimation of year specific b (and common a).
- Estimation of global a and b by sex.
- Analysis of previous results and derived proposal.

Models: The typical length-weight model is $W_i = aL_i^b$, where a and b are parameters to be estimated. If we take logarithms (with base 10) in both sides, we obtain $\log_{10}(W_i) = \log_{10}(a) + b\log_{10}(L_i)$. This model can be fitted in R using the common functions for linear models:

$\log_{10}(W_i) = \log_{10}(a) + b\log_{10}(L_i) + \epsilon_i$,

where ϵ_i normally distributed with mean 0 and variance σ^2.

Prepare data

We read the data files: Portugal and Spain data. We have then three data sets, **dataS** corresponds to the Spain data, **dataP** corresponds to Portugal data, and finally **data** contains both data sets.

```
##   year month   tl  tw gw sex source       area prof M    Y     logL     logW
## 1 2009     7 36.7 385 NA   F market 9a-S_cadiz   NA 7 2009 1.564666 2.585461
## 2 2009     7 29.0 210 NA   F market 9a-S_cadiz   NA 7 2009 1.462398 2.322219
## 3 2009     7 36.0 330 NA   F market 9a-S_cadiz   NA 7 2009 1.556303 2.518514
## 4 2009     7 35.2 414 NA   F market 9a-S_cadiz   NA 7 2009 1.546543 2.617000
## 5 2009     7 30.9 245 NA   F market 9a-S_cadiz   NA 7 2009 1.489958 2.389166
## 6 2009     7 35.6 345 NA   F market 9a-S_cadiz   NA 7 2009 1.551450 2.537819
```

We unify the *area* variable defining the categories northwestern Cantabrian fishing grounds (Spain data, termed "CNO") and Portugal.

```
##
##      CNO Portugal
##    41276    48331
```

The time series is reduced starting in 1982. Note that data for 2020 is not considered since not enough sampling has been carried out in such year.

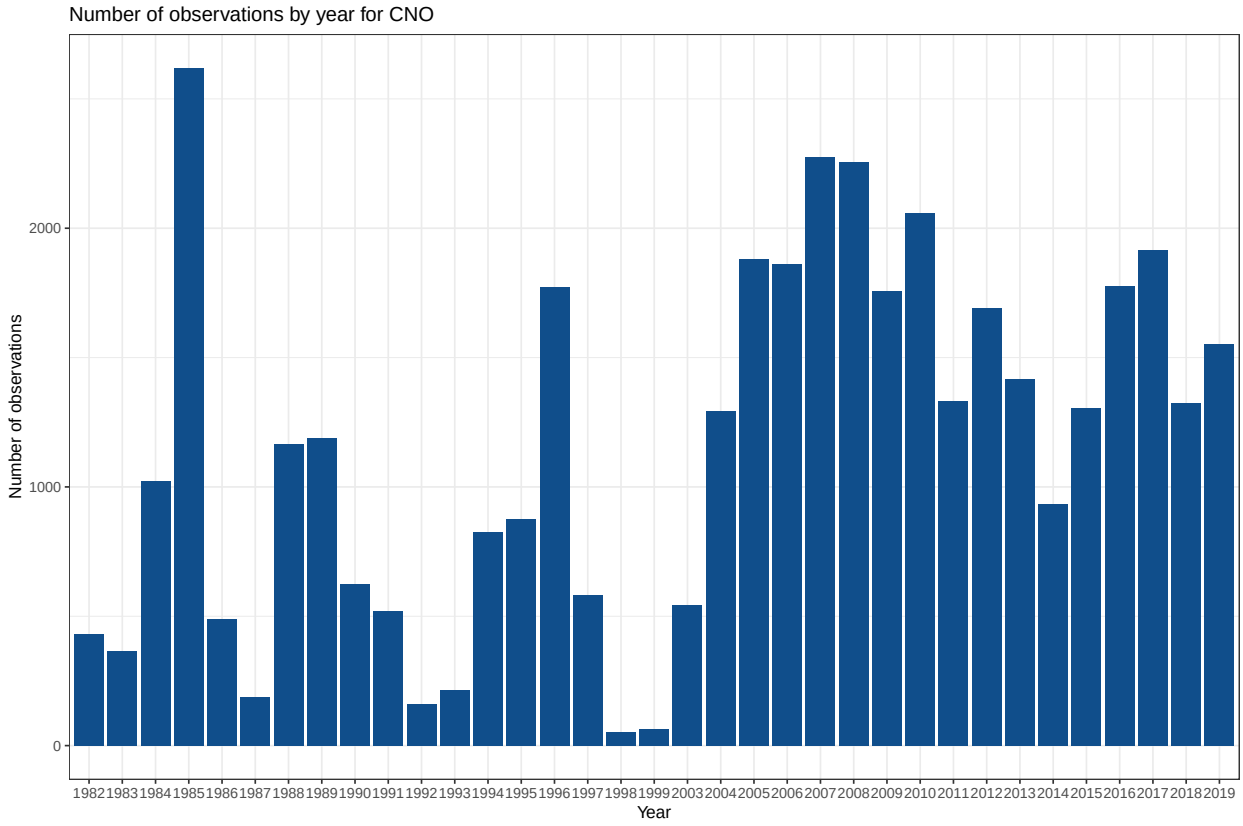
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1982    2004    2009    2007    2013    2019
```

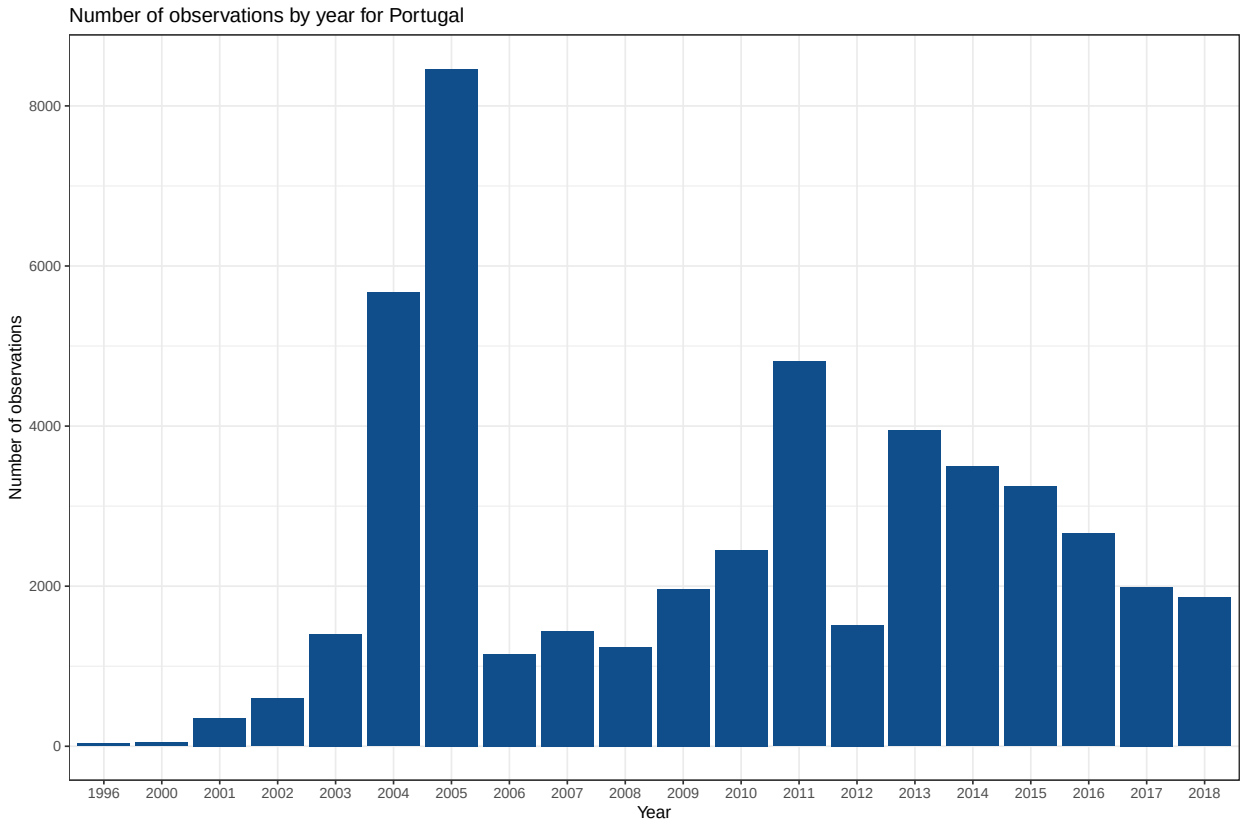
Exploratory

General

We check the number of individuals by area. Portugal data cover from 1996 to 2018, whereas Spain data cover from 1982 to 2019. Note that in 2000, 2001, 2002 there are not individuals in CNO (Cantabrian fishing grounds), also no data is available for Portugal for 1997-1999 and 2019.

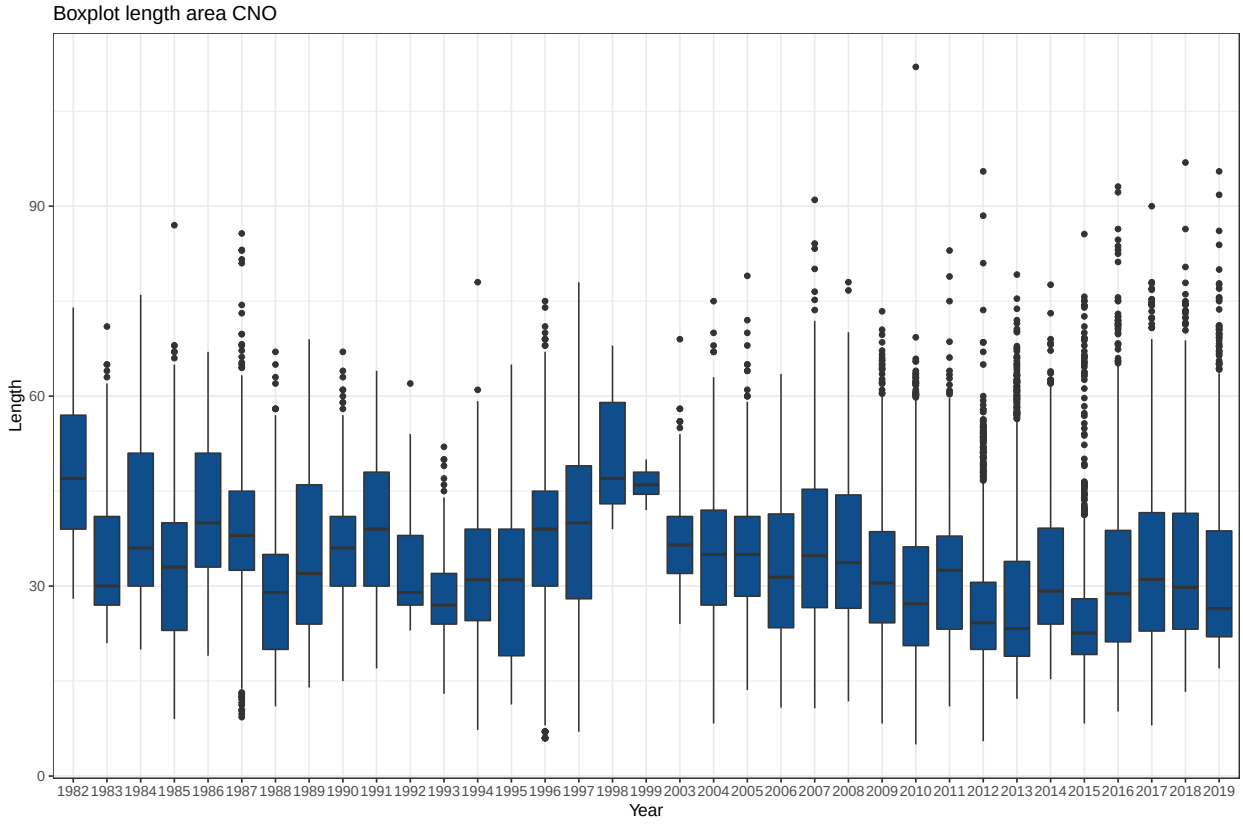
```
##
##         CNO Portugal
##   1982  431        0
##   1983  364        0
##   1984 1023        0
##   1985 2619        0
##   1986  490        0
##   1987  187        0
##   1988 1165        0
##   1989 1187        0
##   1990  623        0
##   1991  519        0
##   1992  160        0
##   1993  214        0
##   1994  827        0
##   1995  874        0
##   1996 1773       35
##   1997  581        0
##   1998   53        0
##   1999   63        0
##   2000    0       47
##   2001    0      350
##   2002    0      596
##   2003  542     1401
##   2004 1295     5670
##   2005 1880     8464
##   2006 1862     1150
##   2007 2275     1443
##   2008 2254     1243
##   2009 1757     1963
##   2010 2058     2455
##   2011 1333     4807
##   2012 1693     1516
##   2013 1418     3947
##   2014  934     3495
##   2015 1306     3243
##   2016 1777     2658
##   2017 1914     1991
##   2018 1323     1857
##   2019 1552        0
```

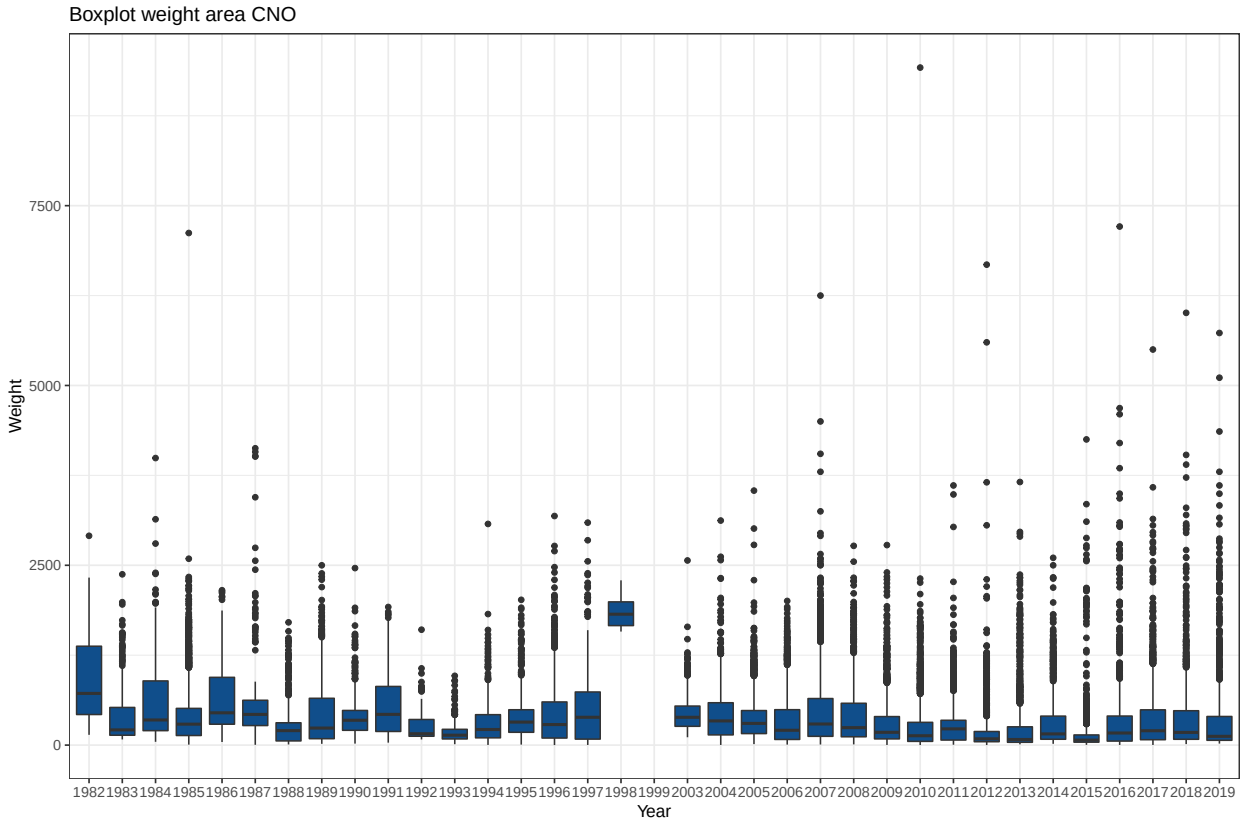




We look at length and weight along time (by grouped area). Two outliers in weight have been detected in

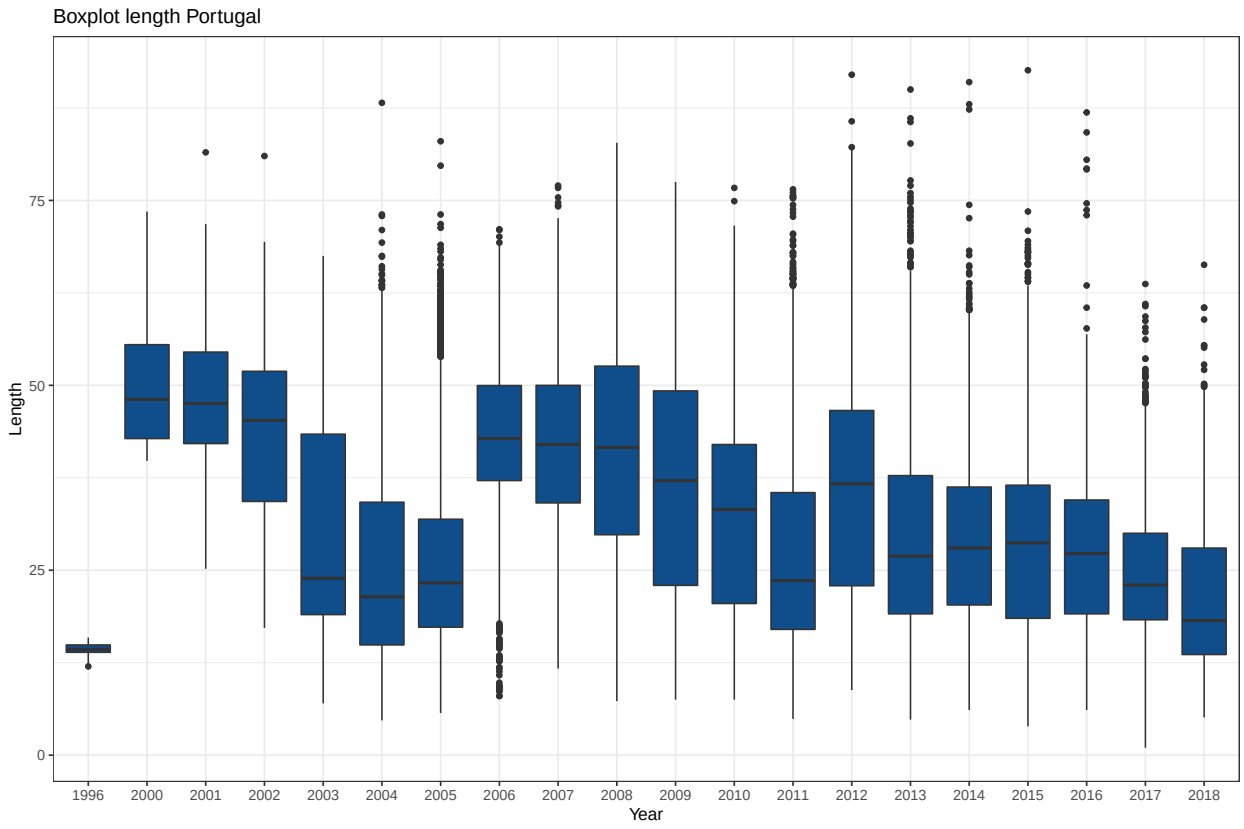
2010 (CNO area), and hence they have been removed (their values were 110211, 116611).

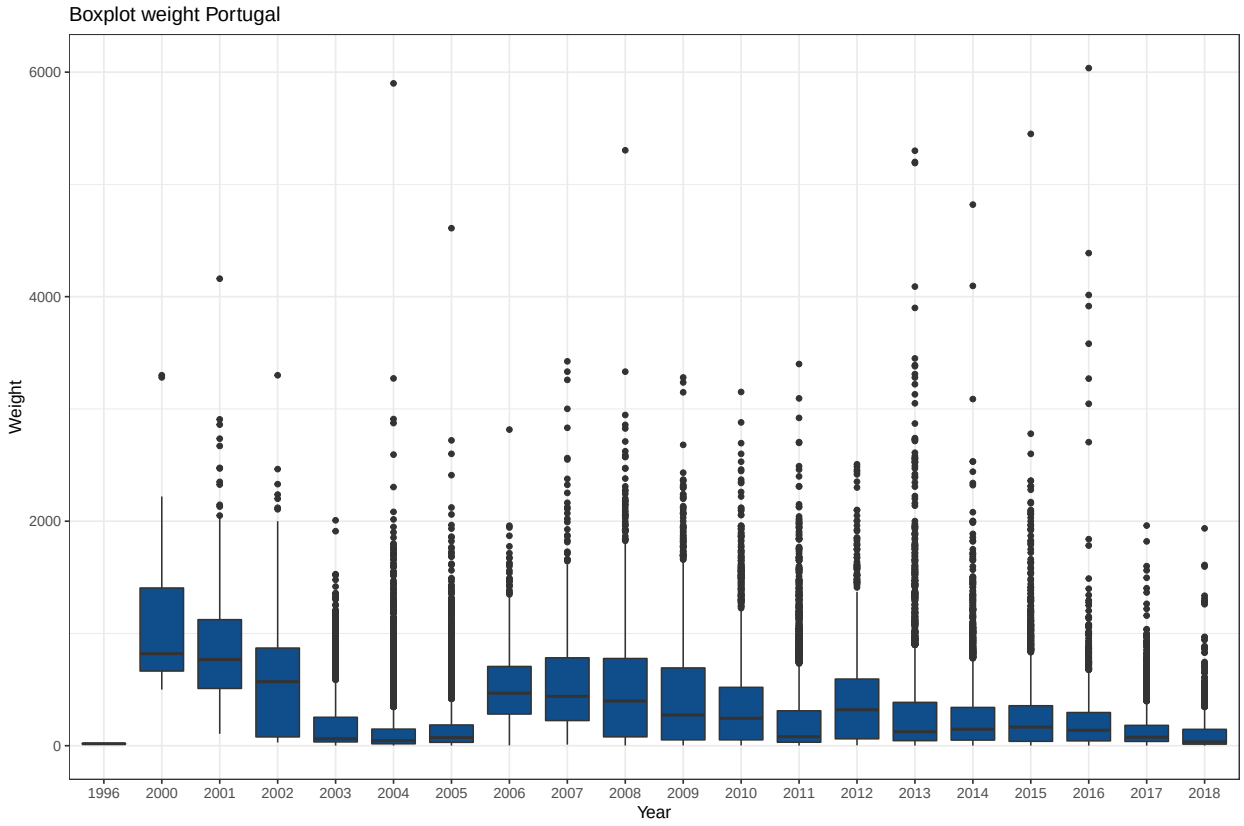




Year 1999 does not contains the weights for the corresponding lengths. Hence, 1999 for CNO is removed from the data base. Year 1998 has a particular behaviour and a small sample size (53 individuals).

```
##       year month tl tw  gw sex source area prof M    Y     logL logW areaU
## 14363 1999     2 48 NA 648   I market   8c   NA 2 1999 1.681241   NA   CNO
## 14364 1999     2 49 NA 714   I market   8c   NA 2 1999 1.690196   NA   CNO
## 14365 1999     2 48 NA 686   I market   8c   NA 2 1999 1.681241   NA   CNO
## 14366 1999     2 50 NA 750   I market   8c   NA 2 1999 1.698970   NA   CNO
## 14367 1999     2 47 NA 693   I market   8c   NA 2 1999 1.672098   NA   CNO
## 14368 1999     2 46 NA 605   F market   8c   NA 2 1999 1.662758   NA   CNO
```

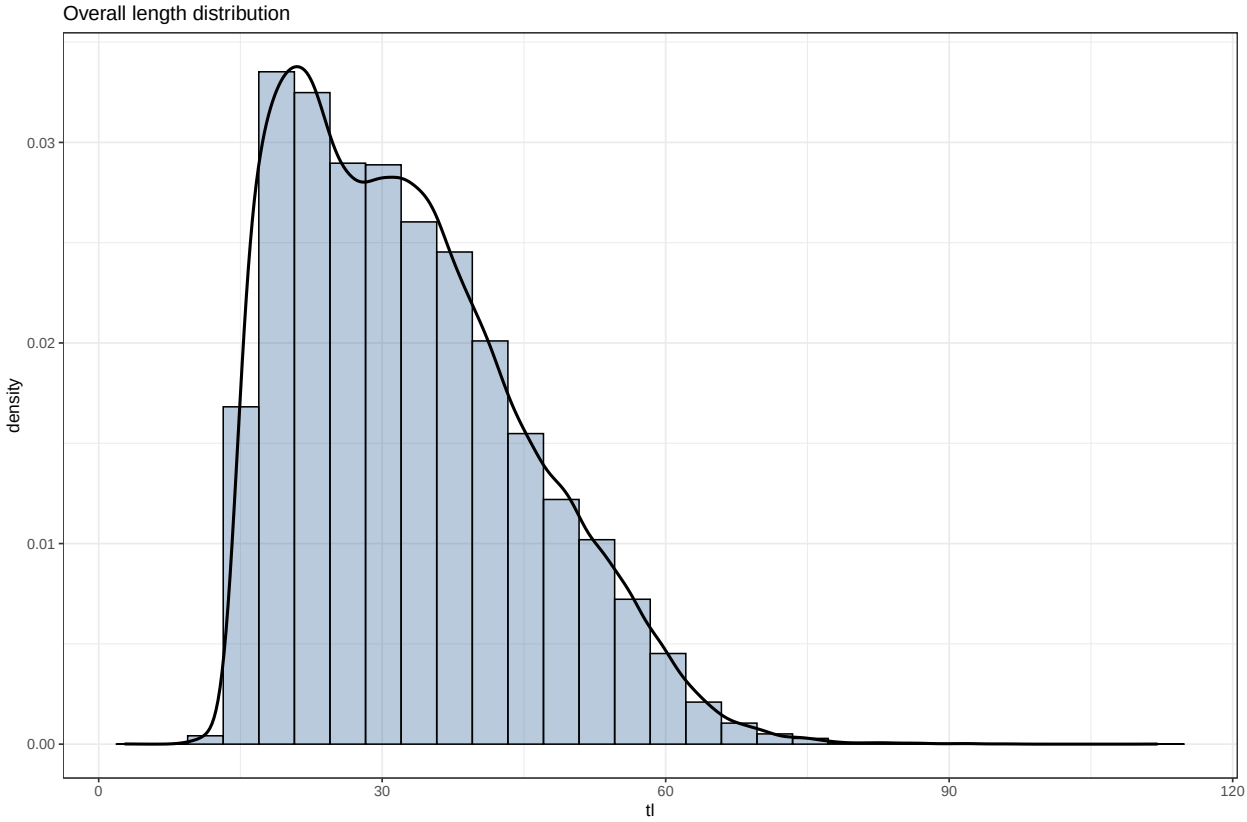


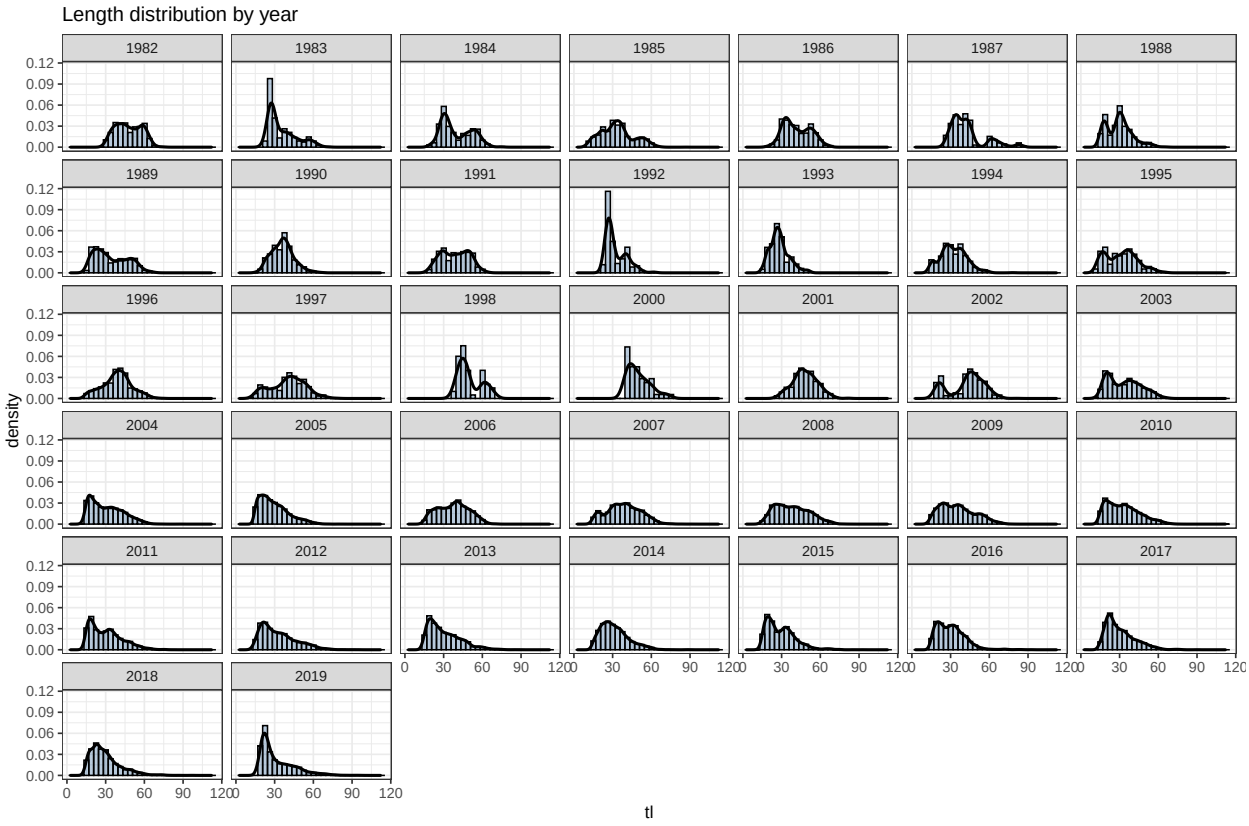


Year 1996 has a clear strange behaviour, then we eliminate the data of this year whose sample size is 35 (too small).

Finally, we decide to eliminate all individuals whose weight is less or equal than 18 gr since the measures for weights smaller than these ones are not trustworthy.

Length distributions





Length-weight model

Global model

Estimation of global (not year specific) a and b using the updated length-weight data base.

```
fit.0 <- lm(logW ~ logL, data = data)
summary(fit.0)
```

```
##
## Call:
## lm(formula = logW ~ logL, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.90013 -0.03208 -0.00432  0.02778  2.99730
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.398095   0.001881   -1275   <2e-16 ***
## logL         3.153611   0.001268    2488   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05646 on 76590 degrees of freedom
##   (4983 observations deleted due to missingness)
## Multiple R-squared:  0.9878, Adjusted R-squared:  0.9878
## F-statistic: 6.188e+06 on 1 and 76590 DF,  p-value: < 2.2e-16
```

The diagnosis plots can be improved eliminating the values whose cook's distance is above its 0.99775 percentile.

The model is adjusted again.

```
fit.0 <- lm(logW ~ logL, data = data)
summary(fit.0)
```

```
##
## Call:
## lm(formula = logW ~ logL, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34822 -0.03128 -0.00371  0.02813  0.35800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.414365   0.001580   -1528   <2e-16 ***
## logL         3.164206   0.001065    2972   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04723 on 76417 degrees of freedom
##   (4983 observations deleted due to missingness)
## Multiple R-squared:  0.9914, Adjusted R-squared:  0.9914
## F-statistic: 8.834e+06 on 1 and 76417 DF,  p-value: < 2.2e-16
```

Model by year (common intercept, year specific slope)

Estimation of year specific b (and common a). Below the summaries report the model results using two different parametrizations of the same model.

```
fit.YY <- lm(logW ~ logL:Y +logL, data = data)
summary(fit.YY)
```

```
##
## Call:
## lm(formula = logW ~ logL:Y + logL, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34050 -0.03008 -0.00318  0.02734  0.35558
##
## Coefficients:
##               Estimate Std. Error   t value Pr(>|t|)
## (Intercept) -2.4146726  0.0016067 -1502.863  < 2e-16 ***
## logL         3.1672825  0.0010830  2924.651  < 2e-16 ***
## logL:Y.L    -0.0323537  0.0010835   -29.859  < 2e-16 ***
## logL:Y.Q    -0.0051941  0.0017287    -3.005 0.002660 **
## logL:Y.C     0.0103104  0.0012312     8.374  < 2e-16 ***
## logL:Y^4    -0.0075119  0.0016830    -4.463 8.08e-06 ***
## logL:Y^5     0.0029580  0.0012980     2.279 0.022679 *
## logL:Y^6    -0.0152811  0.0015889    -9.618  < 2e-16 ***
## logL:Y^7    -0.0037772  0.0013198    -2.862 0.004212 **
## logL:Y^8     0.0111245  0.0015192     7.323 2.46e-13 ***
```

```
## logL:Y^9     0.0009670  0.0016171     0.598 0.549845
## logL:Y^10   -0.0012902  0.0013215    -0.976 0.328906
## logL:Y^11    0.0026899  0.0016729     1.608 0.107865
## logL:Y^12   -0.0044380  0.0012643    -3.510 0.000448 ***
## logL:Y^13    0.0050320  0.0017317     2.906 0.003663 **
## logL:Y^14    0.0062401  0.0012337     5.058 4.24e-07 ***
## logL:Y^15   -0.0106238  0.0016891    -6.290 3.20e-10 ***
## logL:Y^16    0.0004268  0.0011648     0.366 0.714058
## logL:Y^17    0.0023624  0.0017300     1.366 0.172081
## logL:Y^18   -0.0119752  0.0012886    -9.293  < 2e-16 ***
## logL:Y^19    0.0103504  0.0015851     6.530 6.62e-11 ***
## logL:Y^20    0.0035704  0.0014325     2.493 0.012686 *
## logL:Y^21   -0.0037021  0.0015072    -2.456 0.014039 *
## logL:Y^22   -0.0048811  0.0016558    -2.948 0.003201 **
## logL:Y^23    0.0096612  0.0015244     6.338 2.34e-10 ***
## logL:Y^24    0.0035673  0.0017844     1.999 0.045598 *
## logL:Y^25   -0.0150287  0.0014054   -10.694  < 2e-16 ***
## logL:Y^26    0.0088529  0.0017529     5.051 4.42e-07 ***
## logL:Y^27   -0.0002690  0.0013994    -0.192 0.847542
## logL:Y^28    0.0090714  0.0020866     4.348 1.38e-05 ***
## logL:Y^29    0.0098937  0.0018067     5.476 4.36e-08 ***
## logL:Y^30    0.0017274  0.0020300     0.851 0.394808
## logL:Y^31   -0.0083472  0.0014994    -5.567 2.60e-08 ***
## logL:Y^32    0.0064917  0.0017542     3.701 0.000215 ***
## logL:Y^33   -0.0053534  0.0020271    -2.641 0.008271 **
## logL:Y^34    0.0066544  0.0030359     2.192 0.028390 *
## logL:Y^35   -0.0033692  0.0027129    -1.242 0.214275
## logL:Y^36   -0.0022713  0.0006514    -3.487 0.000489 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04599 on 76381 degrees of freedom
##   (4983 observations deleted due to missingness)
## Multiple R-squared:  0.9919, Adjusted R-squared:  0.9919
## F-statistic: 2.519e+05 on 37 and 76381 DF,  p-value: < 2.2e-16
```

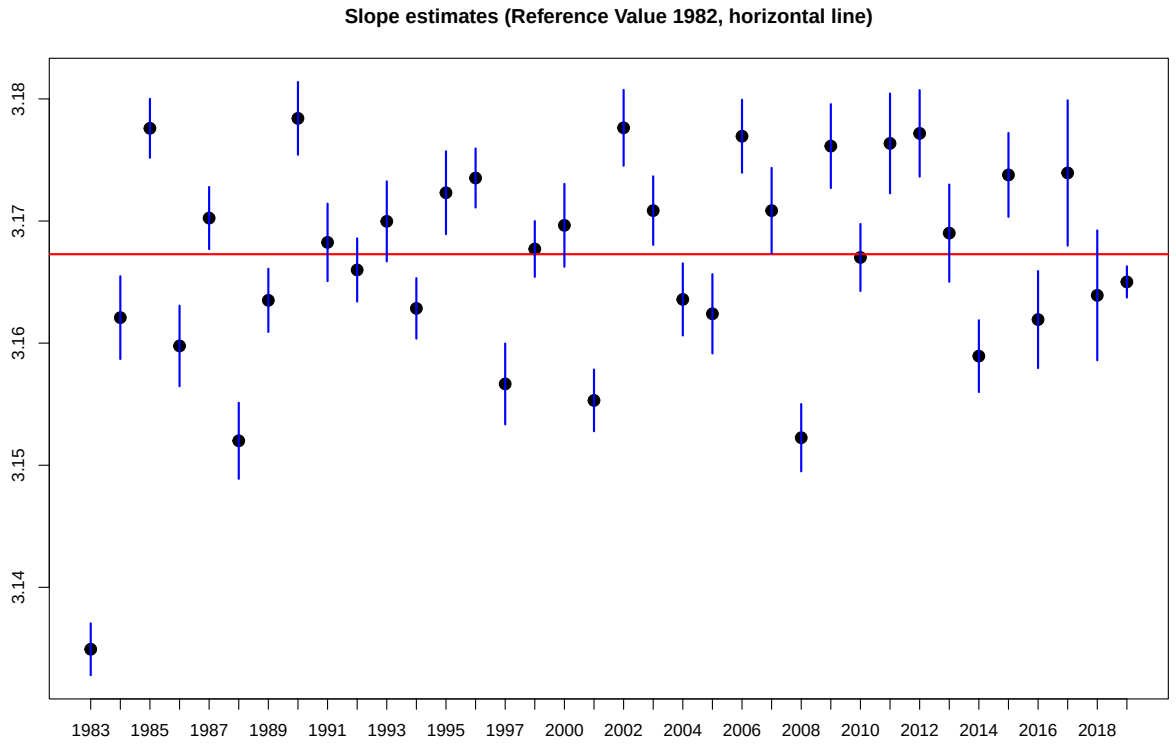
```
# alternative parameterization
fit.YY.bis <- lm(logW ~ logL:Y, data = data)
summary(fit.YY.bis)
```

```
##
## Call:
## lm(formula = logW ~ logL:Y, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34050 -0.03008 -0.00318  0.02734  0.35558
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.414673   0.001607 -1502.9   <2e-16 ***
## logL:Y1982   3.164556   0.001637  1932.9   <2e-16 ***
## logL:Y1983   3.170155   0.001926  1645.6   <2e-16 ***
## logL:Y1984   3.171420   0.001356  2339.2   <2e-16 ***
## logL:Y1985   3.179782   0.001216  2615.3   <2e-16 ***
```

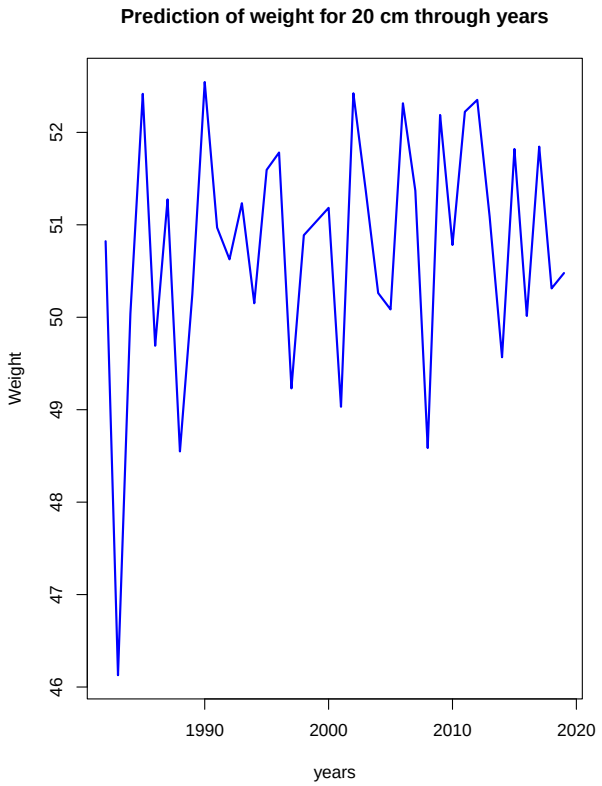
```
## logL:Y1986   3.173834   0.001631  1945.4   <2e-16 ***
## logL:Y1987   3.172121   0.002379  1333.1   <2e-16 ***
## logL:Y1988   3.195030   0.001454  2196.7   <2e-16 ***
## logL:Y1989   3.174024   0.001374  2309.6   <2e-16 ***
## logL:Y1990   3.178302   0.001579  2013.4   <2e-16 ***
## logL:Y1991   3.168156   0.001634  1939.2   <2e-16 ***
## logL:Y1992   3.157475   0.002647  1192.6   <2e-16 ***
## logL:Y1993   3.168621   0.002456  1290.1   <2e-16 ***
## logL:Y1994   3.175160   0.001532  2072.9   <2e-16 ***
## logL:Y1995   3.176700   0.001633  1945.2   <2e-16 ***
## logL:Y1996   3.171229   0.001387  2286.6   <2e-16 ***
## logL:Y1997   3.160179   0.001728  1828.8   <2e-16 ***
## logL:Y1998   3.161178   0.006667   474.2   <2e-16 ***
## logL:Y2000   3.187652   0.004069   783.3   <2e-16 ***
## logL:Y2001   3.170237   0.001847  1716.6   <2e-16 ***
## logL:Y2002   3.163135   0.001715  1843.9   <2e-16 ***
## logL:Y2003   3.169709   0.001350  2347.2   <2e-16 ***
## logL:Y2004   3.174690   0.001206  2632.7   <2e-16 ***
## logL:Y2005   3.165045   0.001168  2709.8   <2e-16 ***
## logL:Y2006   3.161249   0.001201  2632.0   <2e-16 ***
## logL:Y2007   3.161313   0.001151  2747.8   <2e-16 ***
## logL:Y2008   3.153412   0.001166  2704.9   <2e-16 ***
## logL:Y2009   3.158032   0.001185  2664.4   <2e-16 ***
## logL:Y2010   3.159046   0.001174  2691.2   <2e-16 ***
## logL:Y2011   3.161997   0.001180  2679.9   <2e-16 ***
## logL:Y2012   3.161874   0.001235  2560.2   <2e-16 ***
## logL:Y2013   3.159695   0.001182  2672.9   <2e-16 ***
## logL:Y2014   3.161606   0.001186  2665.1   <2e-16 ***
## logL:Y2015   3.166102   0.001211  2614.3   <2e-16 ***
## logL:Y2016   3.164287   0.001199  2638.4   <2e-16 ***
## logL:Y2017   3.159834   0.001214  2603.3   <2e-16 ***
## logL:Y2018   3.157208   0.001267  2490.9   <2e-16 ***
## logL:Y2019   3.155406   0.001340  2354.3   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04599 on 76381 degrees of freedom
##   (4983 observations deleted due to missingness)
## Multiple R-squared:  0.9919, Adjusted R-squared:  0.9919
## F-statistic: 2.519e+05 on 37 and 76381 DF,  p-value: < 2.2e-16
```

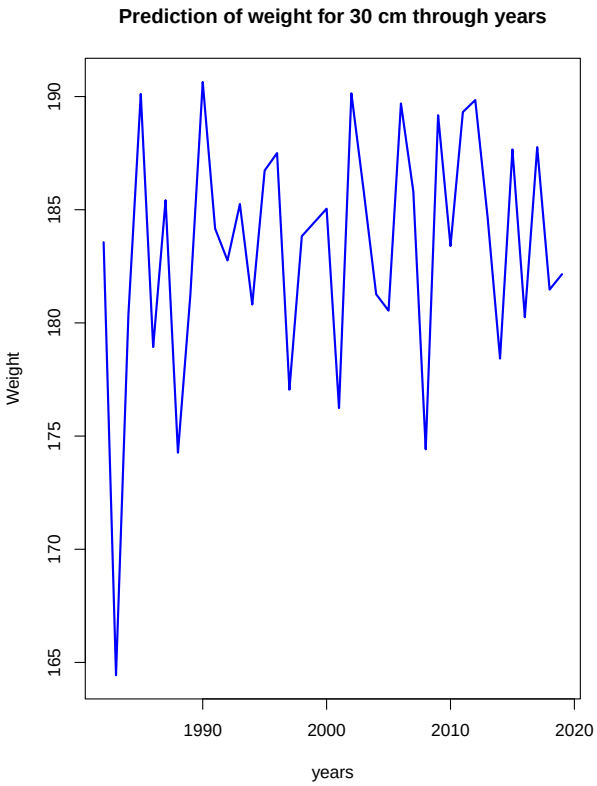
Plots (model by year)

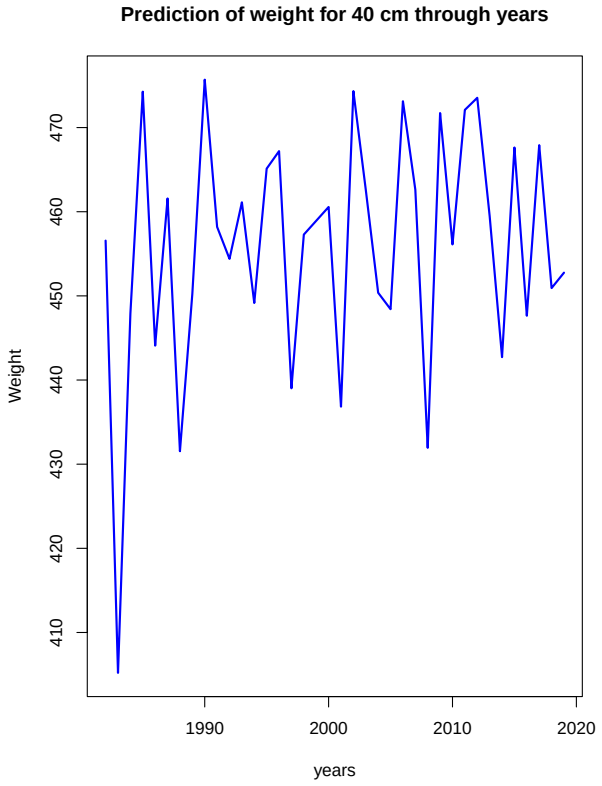
Comparison of the b year specific estimates (black point) respect to the b estimate at the first year of our time series (1982, red horizontal line). The 95% confidence intervals are plotted using a vertical line.

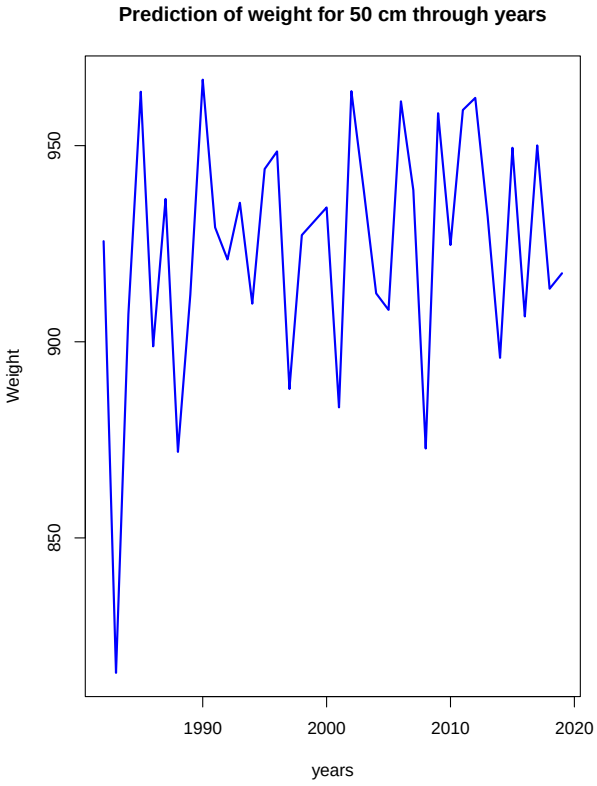


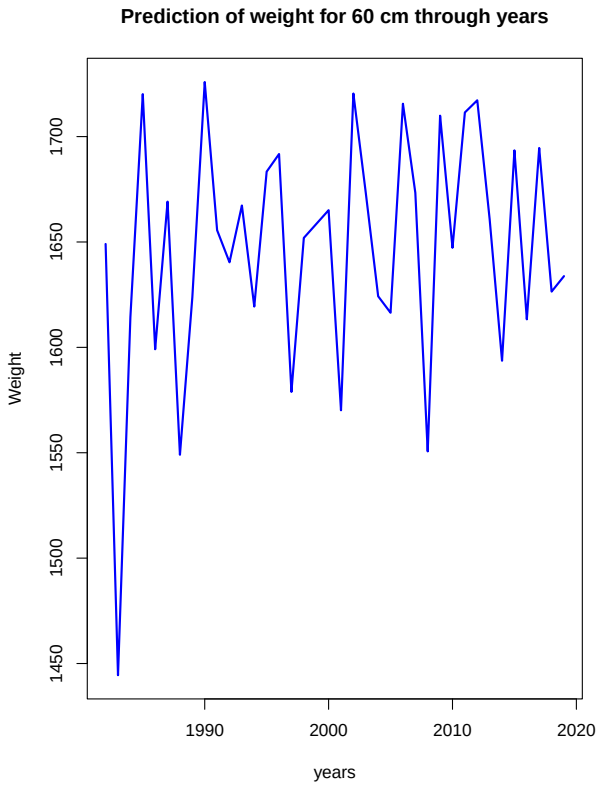
Predicted weights along the years for the length values of the sequence 20, 30, ..., 100 cm. The plots show that the predicted weight series are more or less stable for the last years (and for all the lengths). Then, it seems that year specific b estimates are not required.

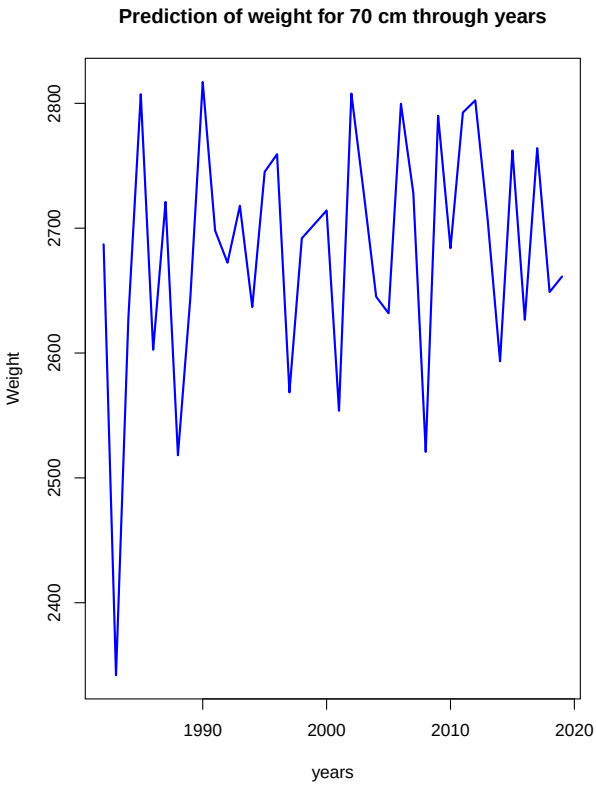

Prediction of weight for 20 cm through years
Weight
46
47
48
49
50
51
52
1990
2000
2010
2020
years

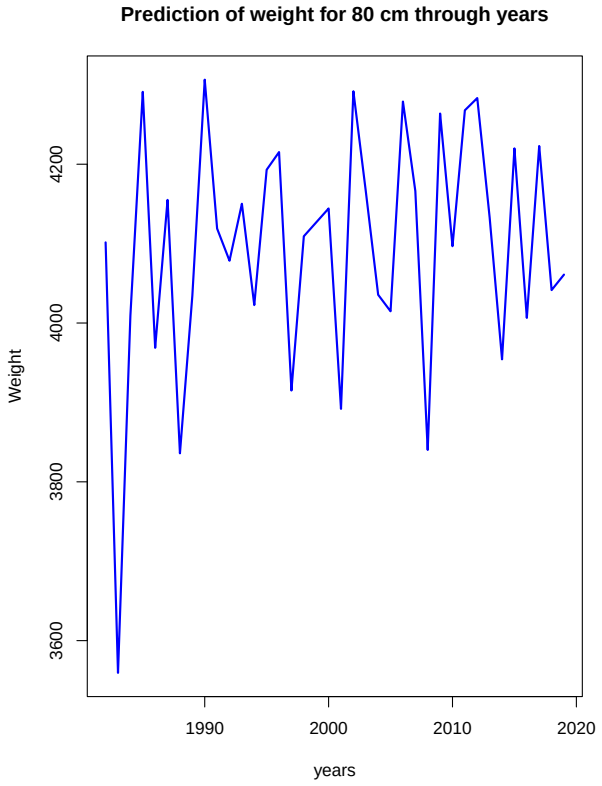

Prediction of weight for 30 cm through years
Weight
165
170
175
180
185
190
1990
2000
2010
2020
years

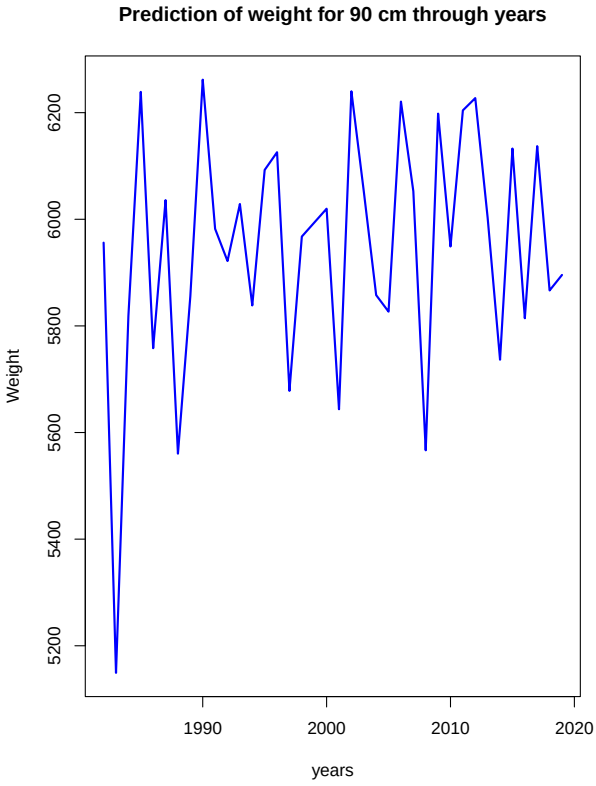

Prediction of weight for 40 cm through years
Weight
410
420
430
440
450
460
470
1990
2000
2010
2020
years

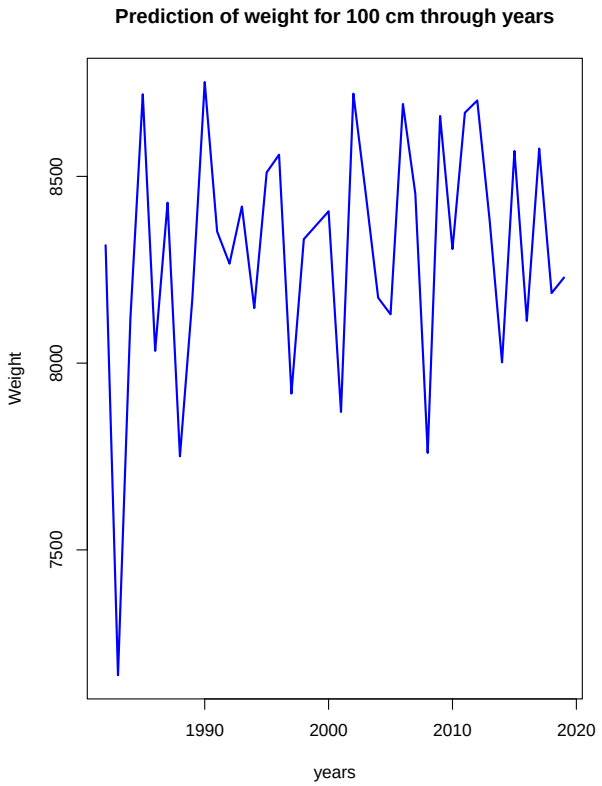

Prediction of weight for 50 cm through years
Weight
850
900
950
1990
2000
2010
2020
years


Prediction of weight for 60 cm through years
Weight
1450
1500
1550
1600
1650
1700
1990
2000
2010
2020
years

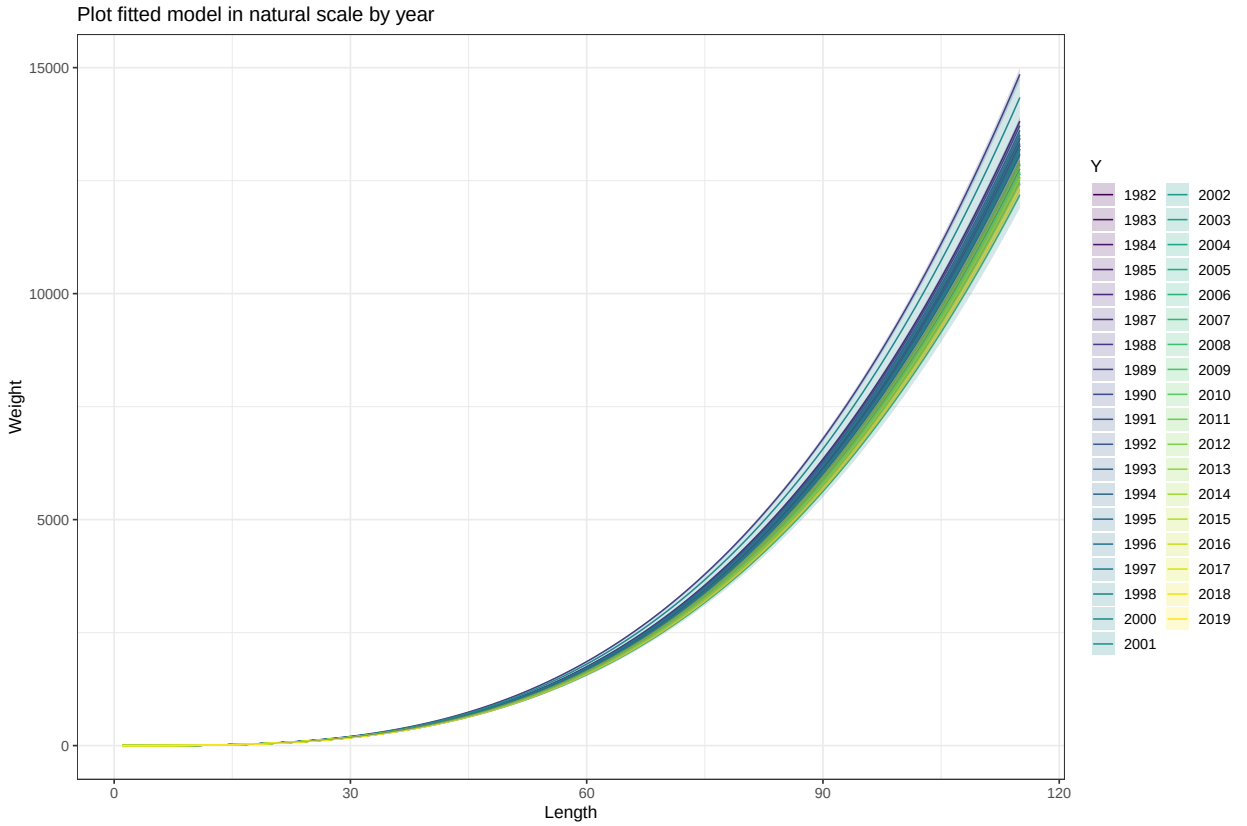

Prediction of weight for 70 cm through years
Weight
2400
2500
2600
2700
2800
1990
2000
2010
2020
years


Prediction of weight for 80 cm through years
Weight
3600
3800
4000
4200
1990
2000
2010
2020
years


Prediction of weight for 90 cm through years
Weight
5200
5400
5600
5800
6000
6200
1990
2000
2010
2020
years



Additionally to previous plots, the next plot show the weight-length relation curves for each year.



Model by sex

Estimation of global a and b by sex. Below the summaries report the model results using two different parametrizations of the same model.

```
fit.SSS <- lm(logW ~ logL*sex, data = data)
summary(fit.SSS)
```

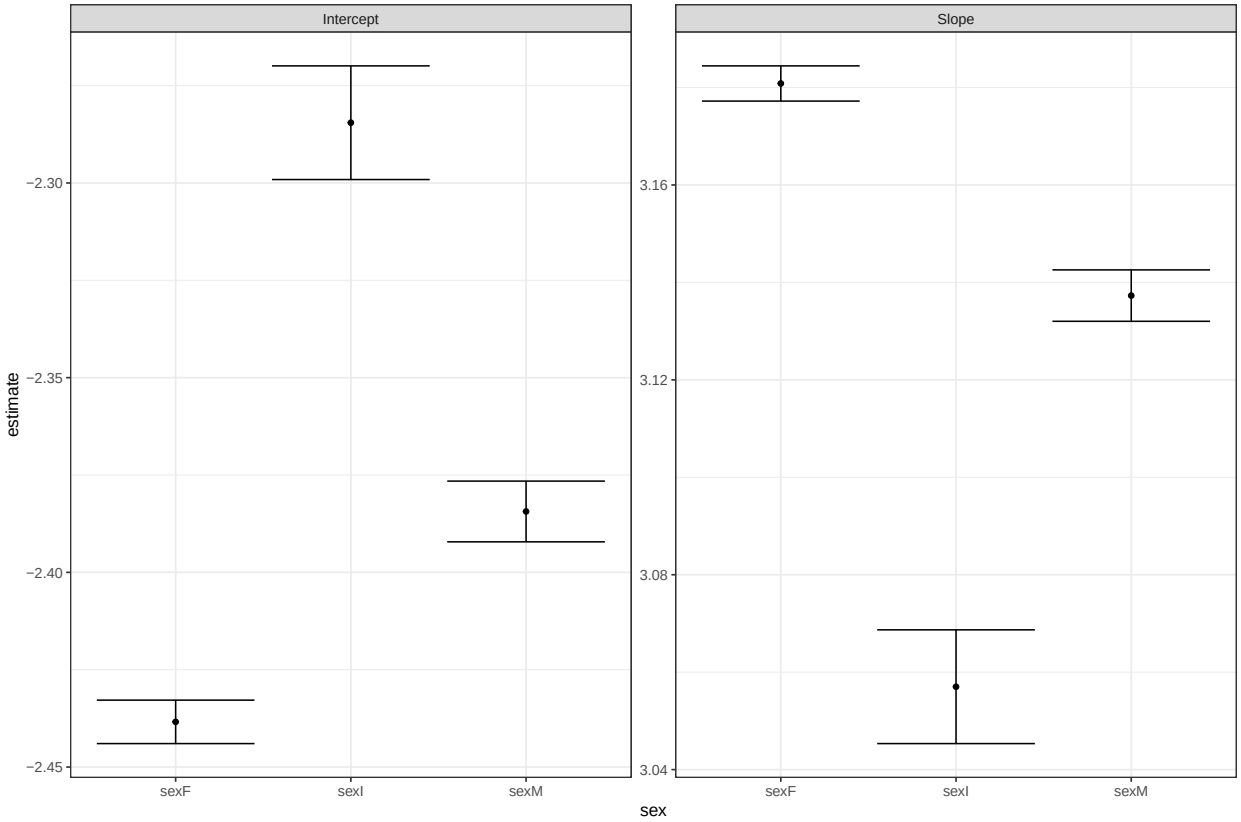
```
##
## Call:
## lm(formula = logW ~ logL * sex, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33954 -0.02925 -0.00326  0.02611  0.35808
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.438396   0.002847 -856.60   <2e-16 ***
## logL         3.180831   0.001844 1725.16   <2e-16 ***
## sexI         0.153870   0.007972   19.30   <2e-16 ***
## sexM         0.054033   0.004892   11.04   <2e-16 ***
## logL:sexI   -0.123814   0.006235  -19.86   <2e-16 ***
## logL:sexM   -0.043534   0.003263  -13.34   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0452 on 62656 degrees of freedom
```

```
##   (3153 observations deleted due to missingness)
## Multiple R-squared:  0.9921, Adjusted R-squared:  0.9921
## F-statistic: 1.577e+06 on 5 and 62656 DF,  p-value: < 2.2e-16
```

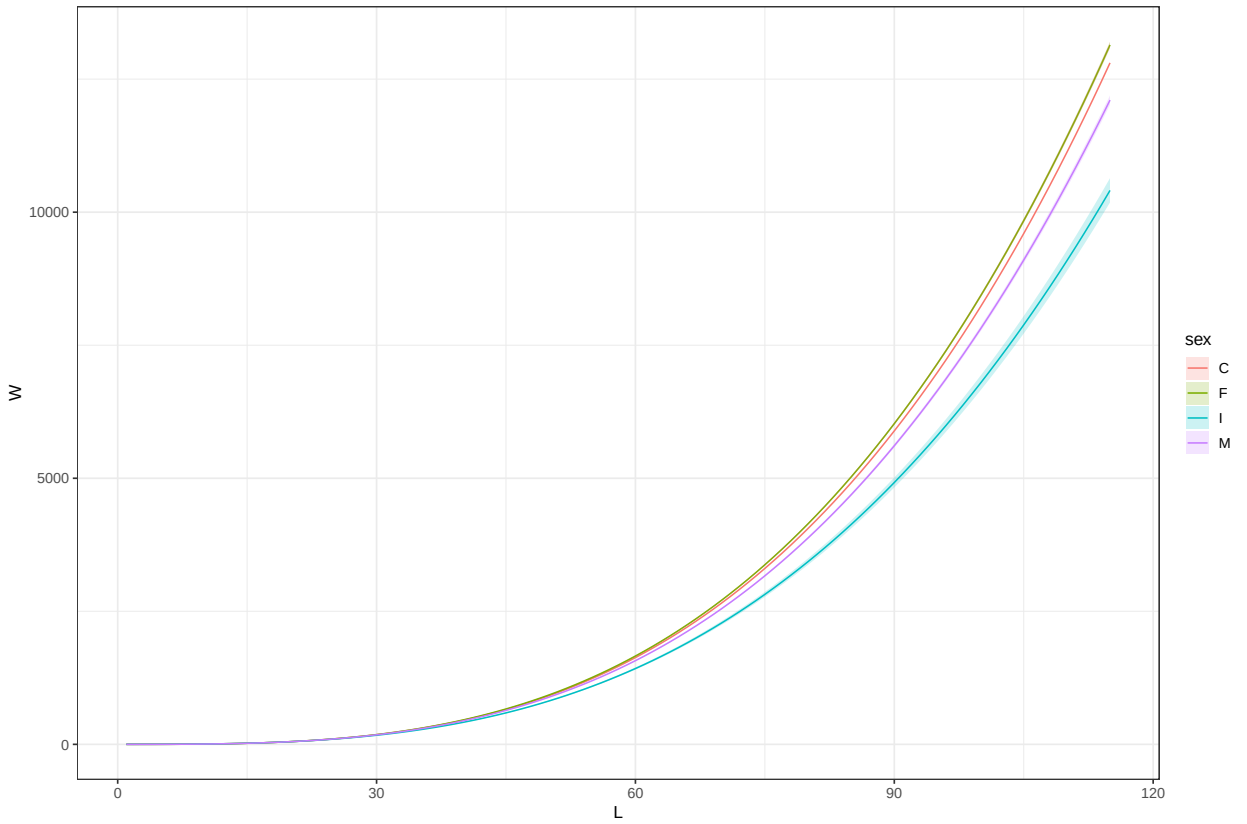
```
fit.SSS.bis <- lm(logW ~ logL*sex -1 -logL, data = data)
summary(fit.SSS.bis)
```

```
##
## Call:
## lm(formula = logW ~ logL * sex - 1 - logL, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33954 -0.02925 -0.00326  0.02611  0.35808
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## sexF      -2.438396   0.002847  -856.6   <2e-16 ***
## sexI      -2.284526   0.007446  -306.8   <2e-16 ***
## sexM      -2.384363   0.003979  -599.3   <2e-16 ***
## logL:sexF  3.180831   0.001844  1725.2   <2e-16 ***
## logL:sexI  3.057018   0.005956   513.3   <2e-16 ***
## logL:sexM  3.137297   0.002693  1165.1   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0452 on 62656 degrees of freedom
##   (3153 observations deleted due to missingness)
## Multiple R-squared:  0.9996, Adjusted R-squared:  0.9996
## F-statistic: 2.634e+07 on 6 and 62656 DF,  p-value: < 2.2e-16
```

Plot of the a and b estimates for each sex (F=females, M=males, I=inmatures) with the corresponding confidence interval.



Plot of the sex specific length-weight curves (C means global model, that is, no sex differences). The results show that the females and males curves are almost the same for the range of lengths for which data of both sexes is available, hence it lead to conclude that the sex differences are not relevant enough to propose sex-specific estimates.



Final proposal

We propose to update the a and b values using the estimates derived from the global model using data from 2003 to 2019. The selection of years has been carried out focusing on the sampling sizes of both countries.

1999 estimates:

a = 0.00659 b = 3.01721

Updated:

a=0.00377 b=3.16826

```
data=subset(data,data$year>2002)
fit.0 <- lm(logW ~ logL, data = data)
summary(fit.0)
```

```
##
## Call:
## lm(formula = logW ~ logL, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34459 -0.02976 -0.00350  0.02659  0.36163
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.423832   0.001678   -1445   <2e-16 ***
## logL         3.168263   0.001140    2780   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04565 on 62660 degrees of freedom
##   (3153 observations deleted due to missingness)
## Multiple R-squared:  0.992,  Adjusted R-squared:  0.992
## F-statistic: 7.727e+06 on 1 and 62660 DF,  p-value: < 2.2e-16
```

Comparison of the updated weight-length curve with the 1999 curve. The confidence interval is plotted in green. The grey points correspond to the observed data. The curves differs for large length values for which almost no data is available.