Australian Government
Department of Agriculture, Fisheries and Forestry
ABARES

A generalised additive model for southern bluefin tuna catch per unit effort (CPUE)

Mark Chambers

Research by the Australian Bureau of Agricultural and Resource Economics and Sciences

August 2013

Working Paper CCSBT-ESC/1309/13 prepared for the CCSBT Extended Scientific Committee for the 18th Meeting to the Scientific Committee

2–7 September 2013, Canberra, Australia

Department of Agriculture, Fisheries and Forestry
Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES)
Postal address GPO Box 1563 Canberra ACT 2601
Switchboard +61 2 6272 2010|
Facsimile +61 2 6272 2001
Email info.abares@daff.gov.au
Web daff.gov.au/abares

Acknowledgements
This work was supported by ABARES and the Fisheries Resources Research Fund.

Contents

Summary ... 1
1 Method ... 2
Calculation of the Index ... 2
Maps of Modelled CPUE ... 2
2 Results ... 4
3 Discussion ... 6
The GAM Model and Index ... 6
Maps of modelled CPUE ... 6
Appendix A - CPUE Prediction Grid ... 8
Appendix B – Model Diagnostics ... 10
Appendix C - CPUE Maps ... 12
R-Code used to calculate GAM Index ... 17
References ... 19

Summary

In response to requests for alternative longline catch per unit effort (CPUE) monitoring series made at meetings of the CPUE Working Group during the CCSBT ESC17 an index based on a generalised additive model (GAM) of catch and effort data is described.

Consistent with other CPUE indices for southern bluefin tuna (SBT), CPUE is defined as the number of SBT aged four years and older captured per thousand longline hooks set. The fitted GAM allows CPUE to vary smoothly over space at each point in time and the spatial distribution of CPUE is allowed to change smoothly over time. An index is calculated by using the fitted model to estimate CPUE over a constant spatial grid each year.

Reassuringly the resulting index is not very different to those used in the operating model to assess the global population of SBT and the management procedure that sets the global TAC.

Generalised additive models of CPUE with spatial covariates, such as the one described, can be used to produce CPUE maps which might help better understand how the spatial distribution of longline CPUE has changed over the history of the fishery. The ability to model spatial CPUE by age class is likely to be still more informative. However, the scope for modelling the spatial distribution of CPUE is currently limited by the spatial resolution of the CPUE data currently provided to the CCSBT.

1 Method

Calculation of the Index

Definition of CPUE

For the purpose of this analysis CPUE is defined as number of southern bluefin tuna (SBT) reported per thousand hooks as recorded in the database SEC_CPUEInputs_65_12.mdb. Accordingly the fitted data comprise CPUE aggregated monthly to five degree square of longitude and latitude. To accommodate the GAM structure it is just necessary to transform the longitude values as recorded in the database so that longitude is continuous across the international dateline. Observations from CCSBT Statistical Areas 1, 10, 11, 12 and 13 are excluded from analysis.

The fitted GAM

At any point in time it is assumed that CPUE varies smoothly over space and the spatial distribution of CPUE is allowed to change smoothly over time. Smoothing is achieved via two multivariate tensor-product splines. A YEAR factor allows mean CPUE to vary each year with reasonable freedom despite the spatial smoothing. A quasi-Poisson error structure is assumed allowing cells with zero catch and non-zero effort to be handled naturally. The aggregated observations are weighted by the number of hooks set with weights capped at 500 thousand hooks. The model is fitted in R using the gam() function within the R package mgcv{} (Wood, 2006) and can be specified with the command:

```
CPUE.GAM <- gam(CPUE ~ te(LONG, LAT, MONTH) + te(LONG, LAT, YEAR) + as.factor(YEAR),
family = quasipoisson(link = log),weights = NHOOKS.CAPPED)
```

Tensor product splines are used in this case because of the differences in scales of the spatial and temporal variables.

Observation weights are capped corresponding to 500 thousand hooks which is around the 90th percentile. The decision to cap weights was made because it was observed that a few observations with exceptionally high weights in the early 1970s led to distorted maps of CPUE. Laslett (2001) specified a minimum variance for the spline model used in the Laslett Core Area index which might be considered a similar strategy.

Calculation of the Index

The fitted GAM model is used to predict CPUE on the 'Laslett Core Area' (Laslett, 2001) grid cells. Annual index values are calculated as the mean of predicted CPUE in the 277 core area grid cells for the corresponding year between 1969 and 2012. The grid cells that comprise the Laslett Core Area are shown in the Appendix. Full R code used to calculate the index is also included in the appendix.

Maps of Modelled CPUE

The fitted GAM can be used to produce maps of modelled longline CPUE for SBT. These can be easier to interpret than maps showing raw CPUE in each grid cell, although both are useful. Since the tensor product splines used in the GAM for CPUE incorporate temporal as well as spatial variables, the vis.gam() function can be used to produce contours of modelled CPUE at different points in time. The maps also serve as an informal model diagnostic allowing model fitted CPUE to be compared with what is currently understood about SBT CPUE.

In practice it is difficult to fit GAM models to the SBT CPUE data due to the coarseness of the aggregated data. In particular there are too few unique latitude values to permit reasonable interpolation between the monthly five degree grid points.

2 Results

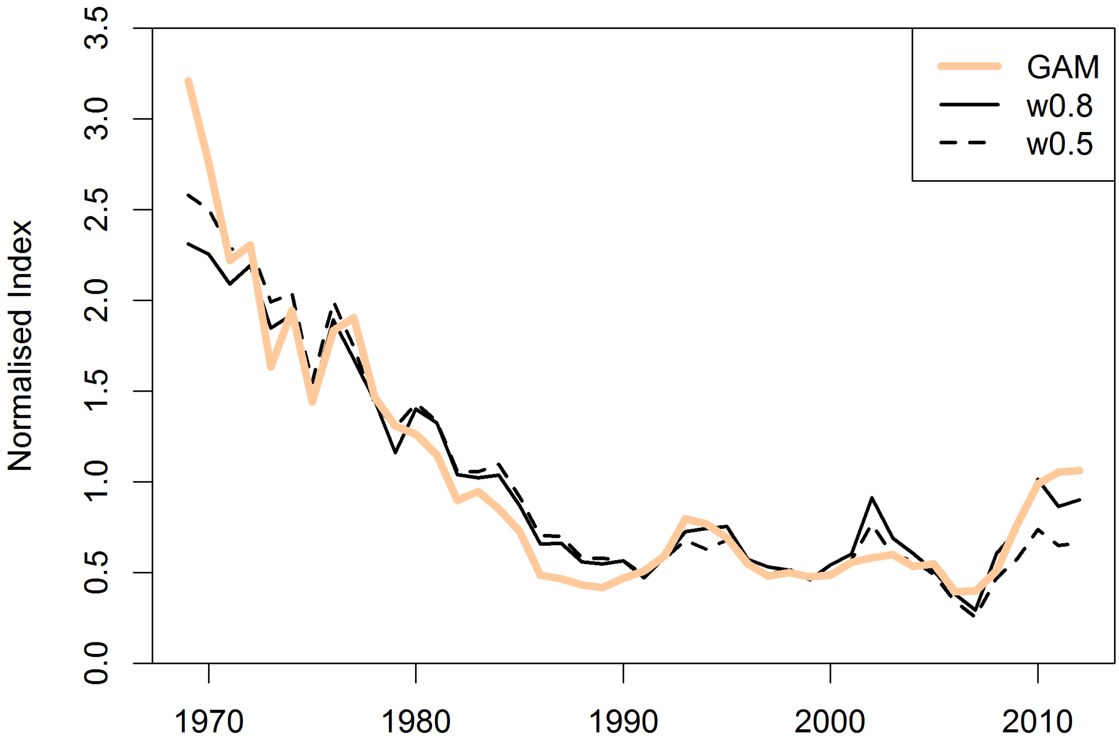

Figure 1: GAM based CPUE index plotted with base w0.8 and base w0.5 indices for comparison. All indices have been normalised to have an average of one between 1969 and 2012.

The resulting index exhibits the same major trends as the two CPUE indices (base w0.8 and base w0.5, defined in CCSBT/ESC/1208/35) used in the operating model (Figure 1). The most notable difference is that the GAM model index is relatively higher at the beginning of the series in 1969.

Differences between the series in recent years are more apparent if the series is plotted between 1980 and 2012 as shown in Figure 2.

The new GAM index is compared with the Laslett Core Area index in Figure 3.

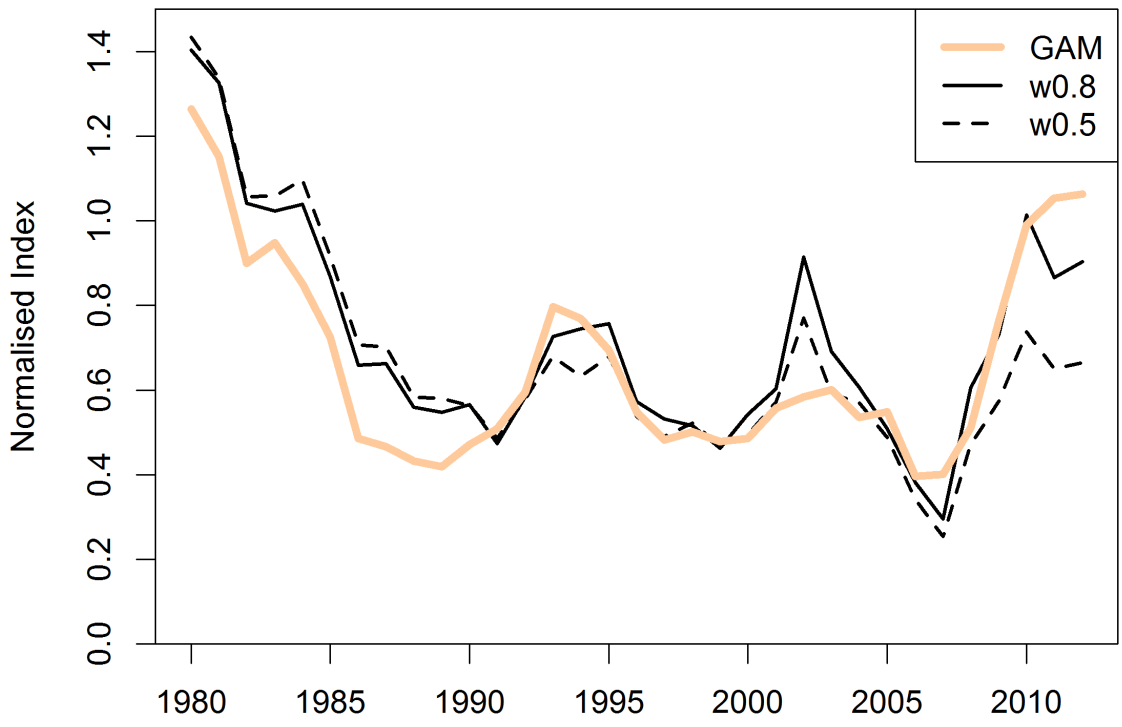

Figure 2: GAM based CPUE index plotted between 1980 and 2012 with base w0.8 and base w0.5 indices for comparison. All indices have been normalised to have an average of one between 1969 and 2012.

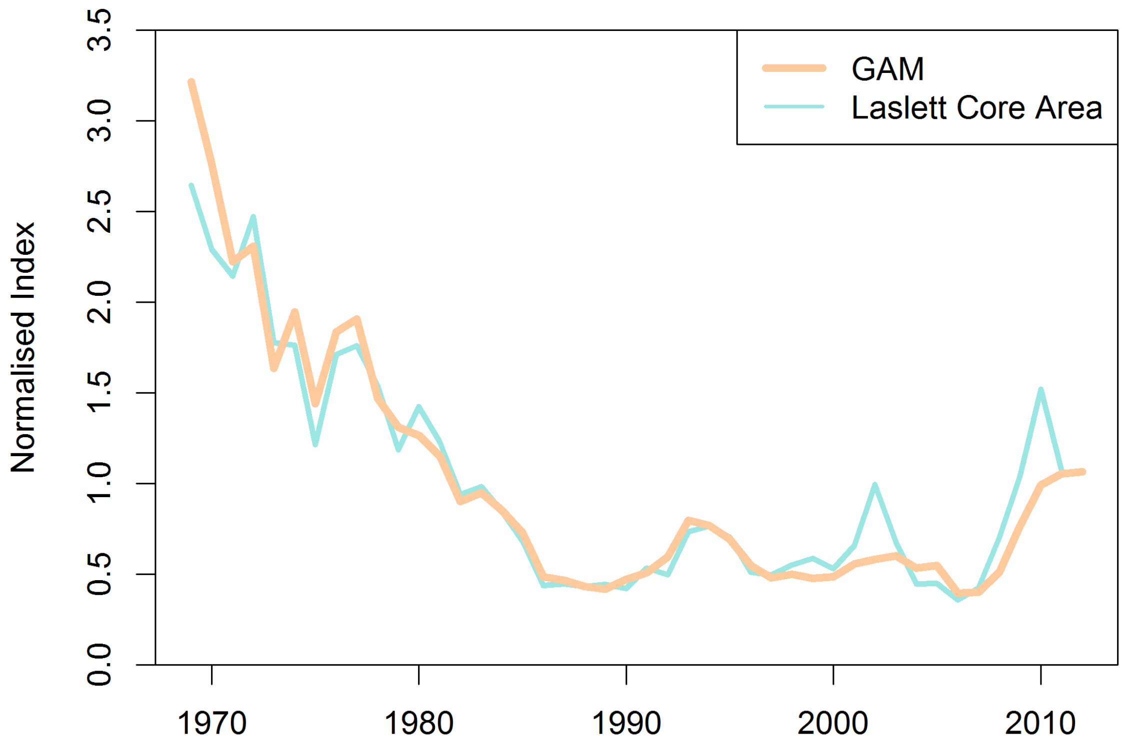

Figure 3: GAM based CPUE index plotted with the Laslett Core Area Index (1969-2011). Both indices have been normalised to have an average of one between 1969 and 2011.

3 Discussion

The GAM Model and Index

The series derived from the GAM model shares the basic features of the w0.5 and w0.8 indices used in the SBT operating model.

Conceptually the model described here is similar to the Laslett Core Area index previously proposed for SBT (Laslett, 2001). In particular the incorporation of a 3-variable spline in latitude, longitude and month is borrowed from Laslett's model. The grid of prediction points over which annual index values are calculated is also the same as the grid used to calculate the Laslett Core Area index.

The most significant difference between the new GAM index and the Laslett Core Area index is the Laslett model was fitted to each year's data separately whereas the GAM model is fitted to all years' data simultaneously and includes predictors involving YEAR. Another difference is the assumed error structure. The new GAM model assumes a quasi-Poisson error structure whereas the Laslett model is log-normal with variance modelled on effort and expected value and adjustments for cells where a CPUE of zero is observed. The ability to fit the new GAM model in R is an additional benefit.

Fitting the model separately each year allows the Laslett CPUE model to be very flexible to changes in the distribution of CPUE between years. On the other hand it might lack robustness in years when effort is low and the spatial distribution of fishing is limited. This might explain the very high level of the Laslett Core Area index in 2002 and 2010 for instance (Figure 3). The incorporation of the second spline involving YEAR in the GAM model, which constrains year-to-year changes in the distribution of CPUE should increase the robustness of the new model relative to the Laslett Core Area index in years when effort is low.

The effect of weighting the aggregated CPUE observations in the GAM model by the number of hooks is noticeable. A quasi-Poisson model weighted by the number of hooks is quite consistent with the assumptions about variance of the lognormal Laslett CPUE model. The most recent (2012) index point is particularly sensitive to data weighting.

The GAM model and the Laslett model would both be fitted more reliably and with greater flexibility if the CPUE data were available at a finer scale. With the current 5-degree square level of aggregation, only three or sometimes four unique latitude values are fished.

Maps of modelled CPUE

Caveats

Maps of the spatial distribution of CPUE are provided for April, June-July and September for the years 1972, 1982, 1992, 2002 and 2012. To allow for reasonable contrast between areas, CPUE is mapped on the log scale. A consequence of this is that small differences in modelled log CPUE reflect relatively large differences in modelled CPUE on the natural scale.

A degree of caution should be exercised when interpreting the maps. Firstly, remembering that the spatial resolution of the input data is aggregated to 5 degrees of latitude and longitude, interpolation between data points less reliable than would be the case if the model were fitted to CPUE at a finer spatial scale. Strictly speaking the location of observations should probably be translated from the north-west corner of each grid square to its centroid before fitting the GAM model. However, we expect the effect of this spatial bias to be fairly minor. It is likely also that the data are "over smoothed" temporally such that real differences in the year-to-year distribution of CPUE are sometimes filtered out.

Results

The maps suggest that the spatial distribution of CPUE varies between years. Perhaps the first point to notice is that the spatial extent of moderate-high CPUE (above 1.5 on the log scale or approximately 4.5 fish per thousand hooks on the natural scale) is predicted to have been far more widespread in 1972 than in later decades.

High CPUE around the south island of New Zealand in September seems to be a fairly constant feature. However, in 2012 high CPUE is predicted to extend west of Tasmania and begins earlier in the season than suggested for the other years considered. Conversely, unusually low CPUE is suggested for the southern Indian Ocean in June-July of 1982 and 1992. We reiterate that temporal smoothing is likely to result in misleading predictions for the spatial distribution of CPUE at times. If the data were available at a finer scale, a similar model could be fitted separately each year as is done for the Laslett Core Area Index to give a more reliable representation of the distribution of CPUE.

To meaningfully compare CPUE between decades, changes in the size classes of fish targeted should also be taken into account since changes in selectivity of longline fleets catch SBT are assumed in the stock assessment model for SBT.

Finally it is impossible to know what the effect of the historic overcatch might be on the CPUE indices because of uncertainty around both the catch and effort over the period affected. Unfortunately this problem affects all CPUE indices for SBT.

Appendix A - CPUE Prediction Grid

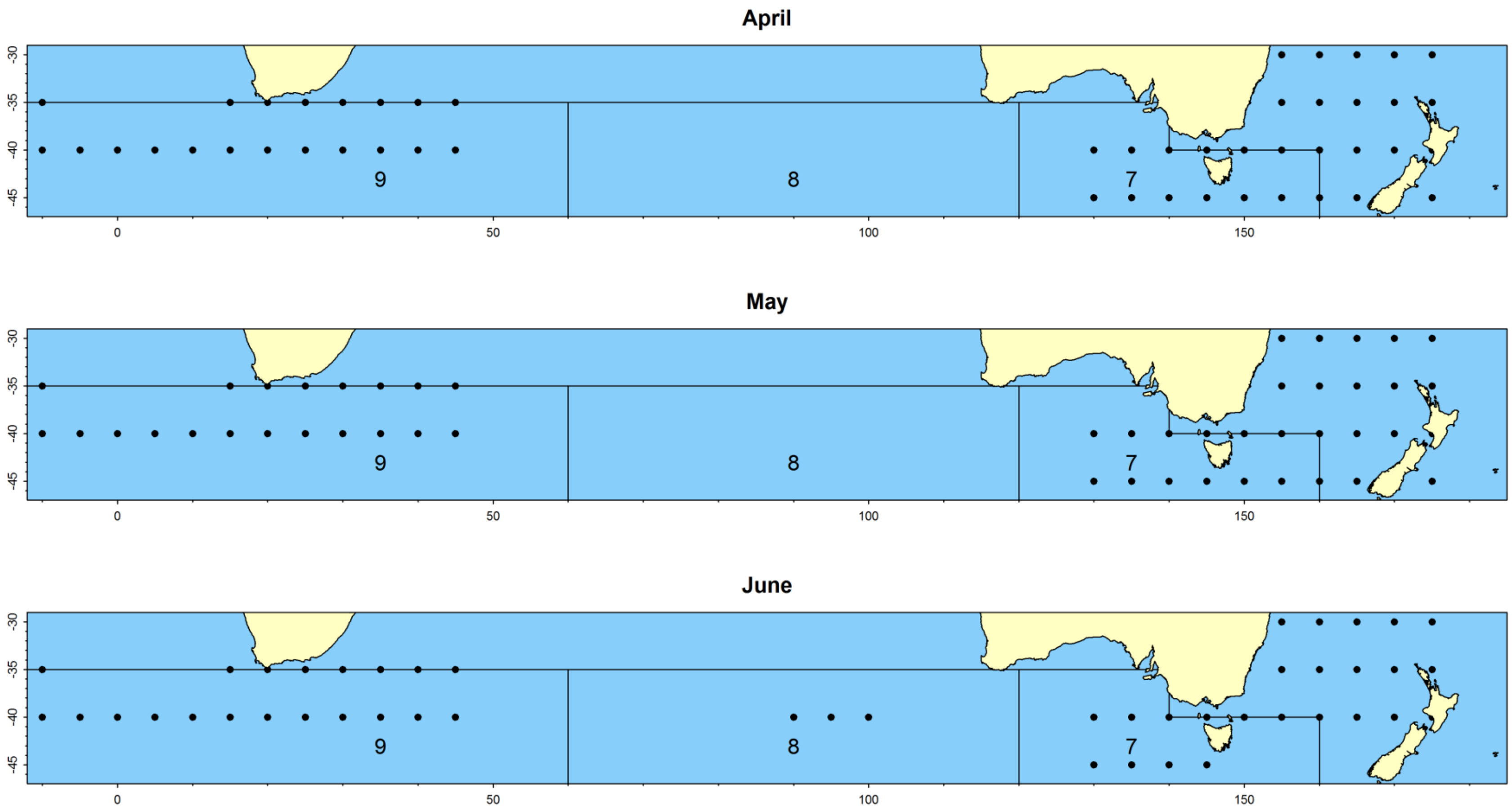

Figure 4: Laslett Core Area grid points (April-June) used to calculate GAM index. Numerals denote CCSBT statistical areas.

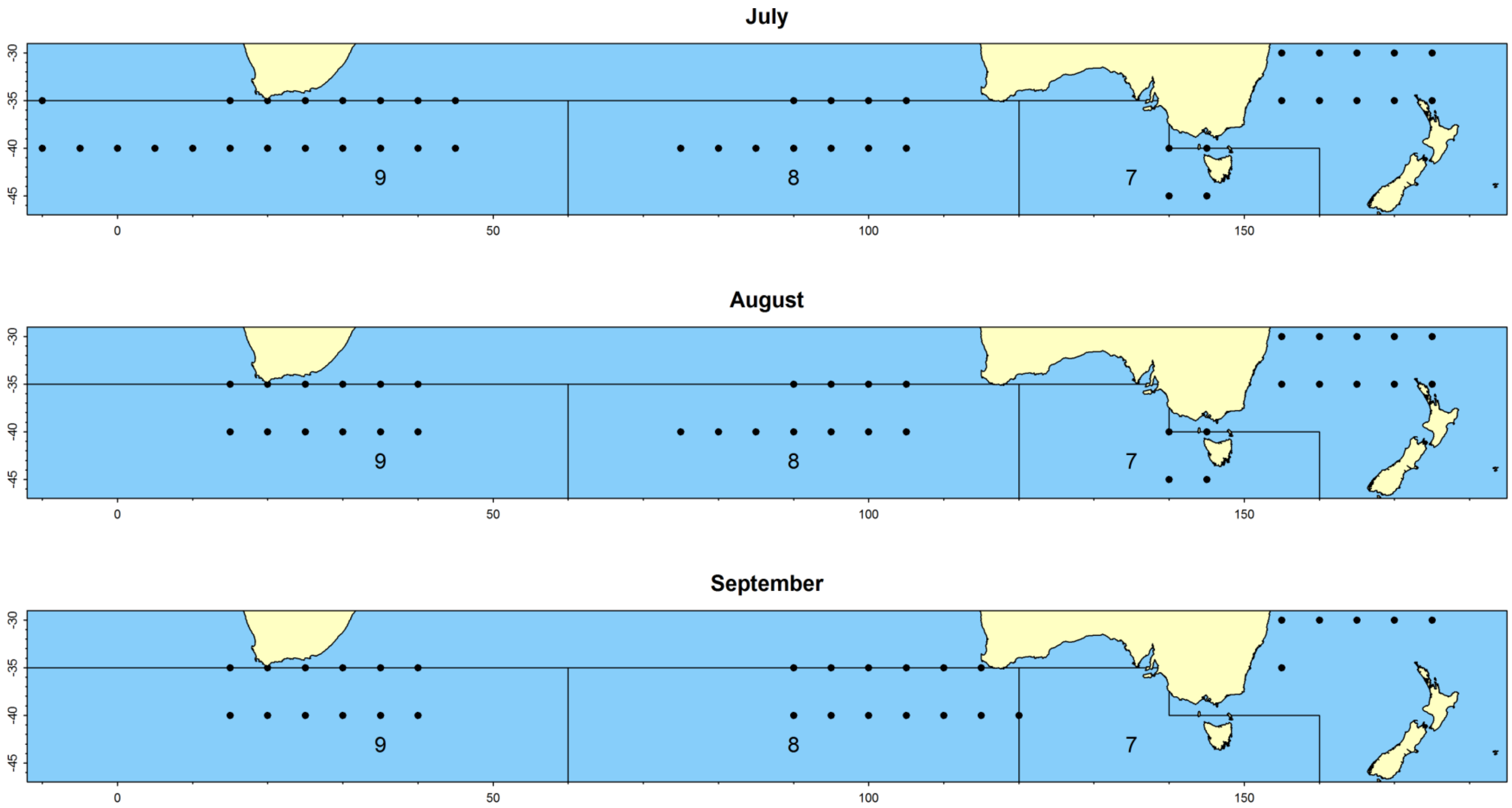

Figure 5: Laslett Core Area grid points (July - September) used to calculate the GAM index. Numerals denote CCSBT statistical areas.

Appendix B – Model Diagnostics

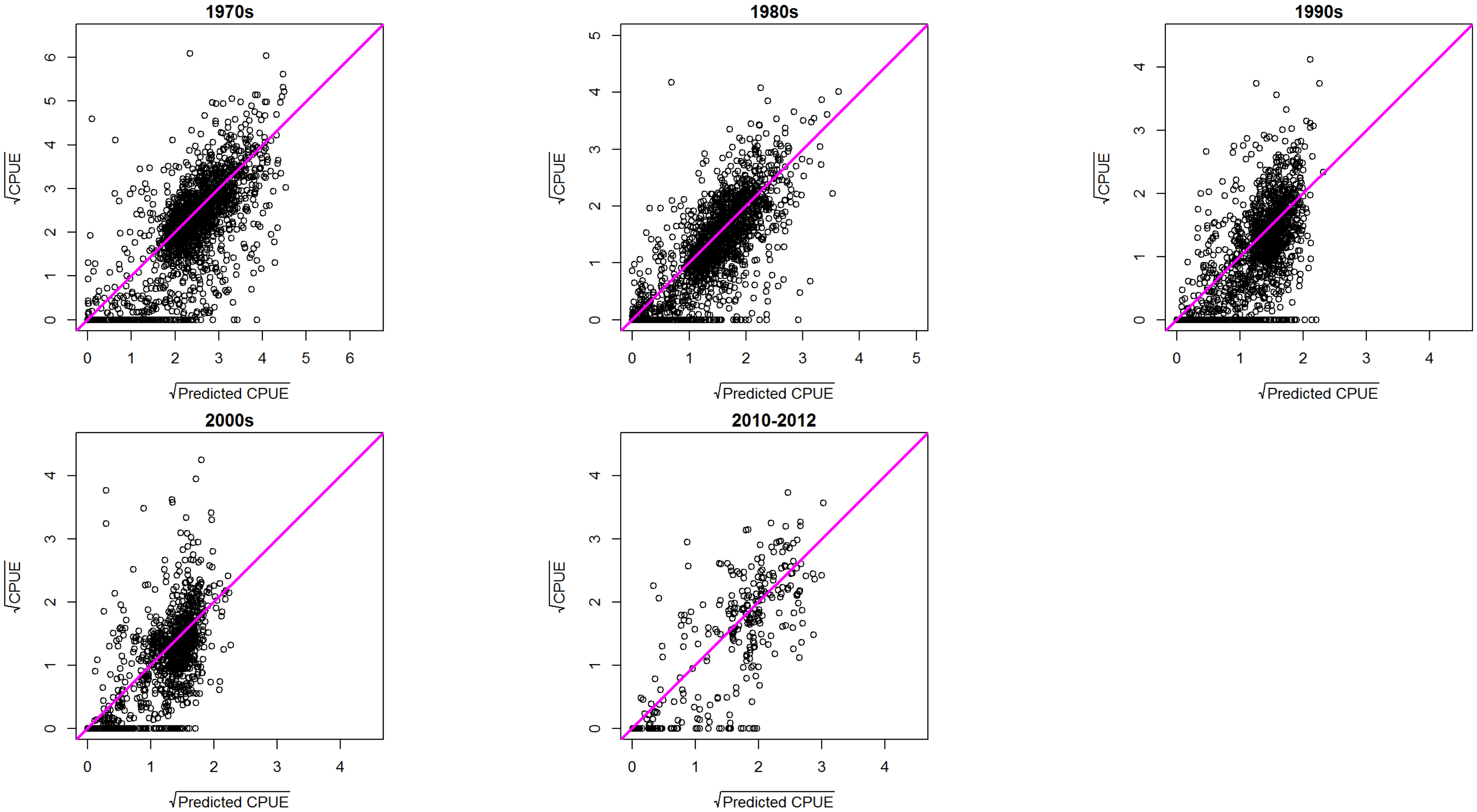

Figure 6: Plots of square root transformed CPUE versus square root transformed predicted CPUE for fished cells over the CPUE index prediction grid by decade. Note the changes in scale.

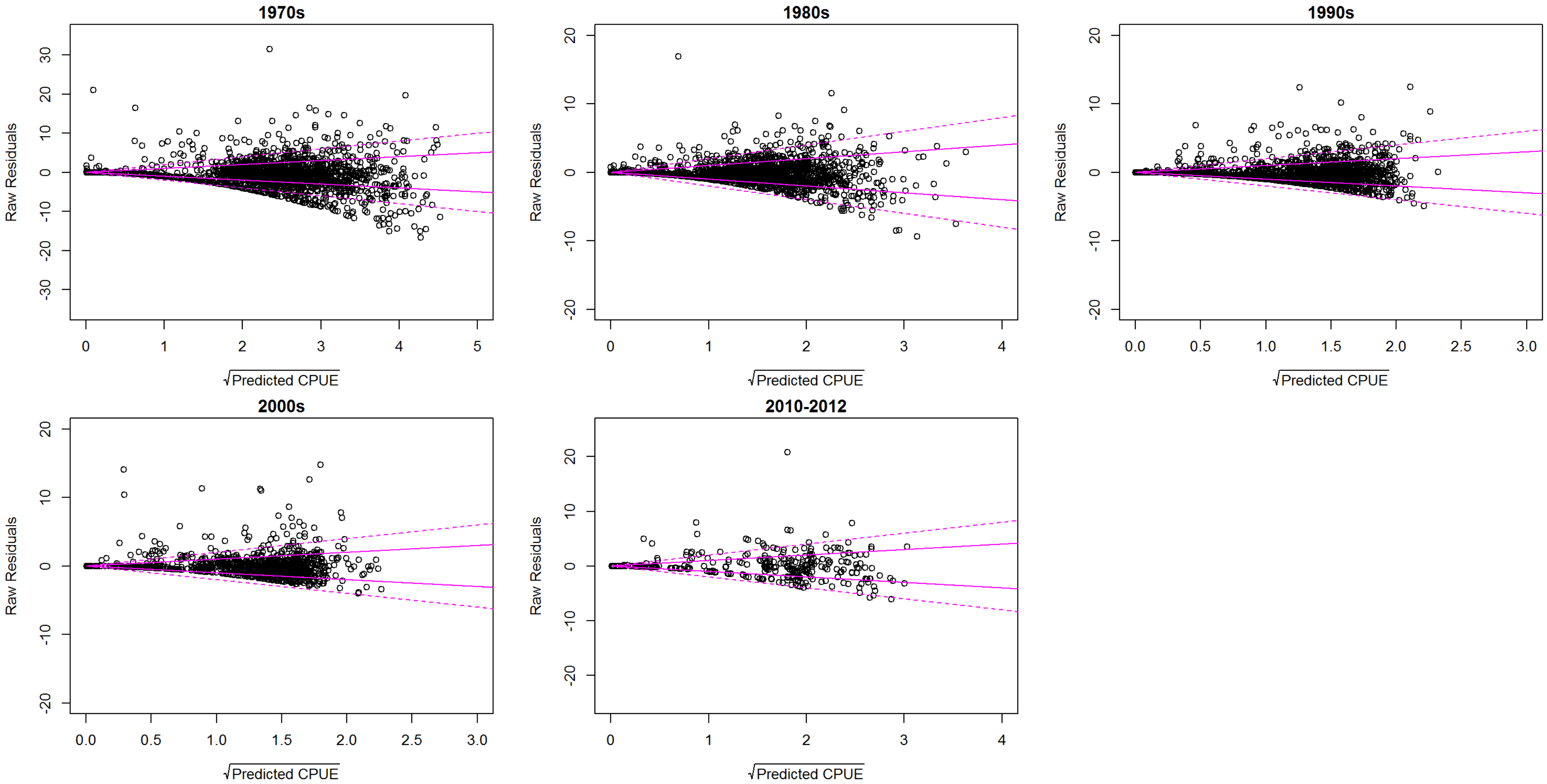

Figure 7: Raw residuals versus square root transformed predicted CPUE for fished cells over the CPUE index prediction grid by decade. Solid lines indicate expected ± 1 residual standard deviation, dashed lines indicate expected ± 2 residual standard deviations.

Appendix C - CPUE Maps

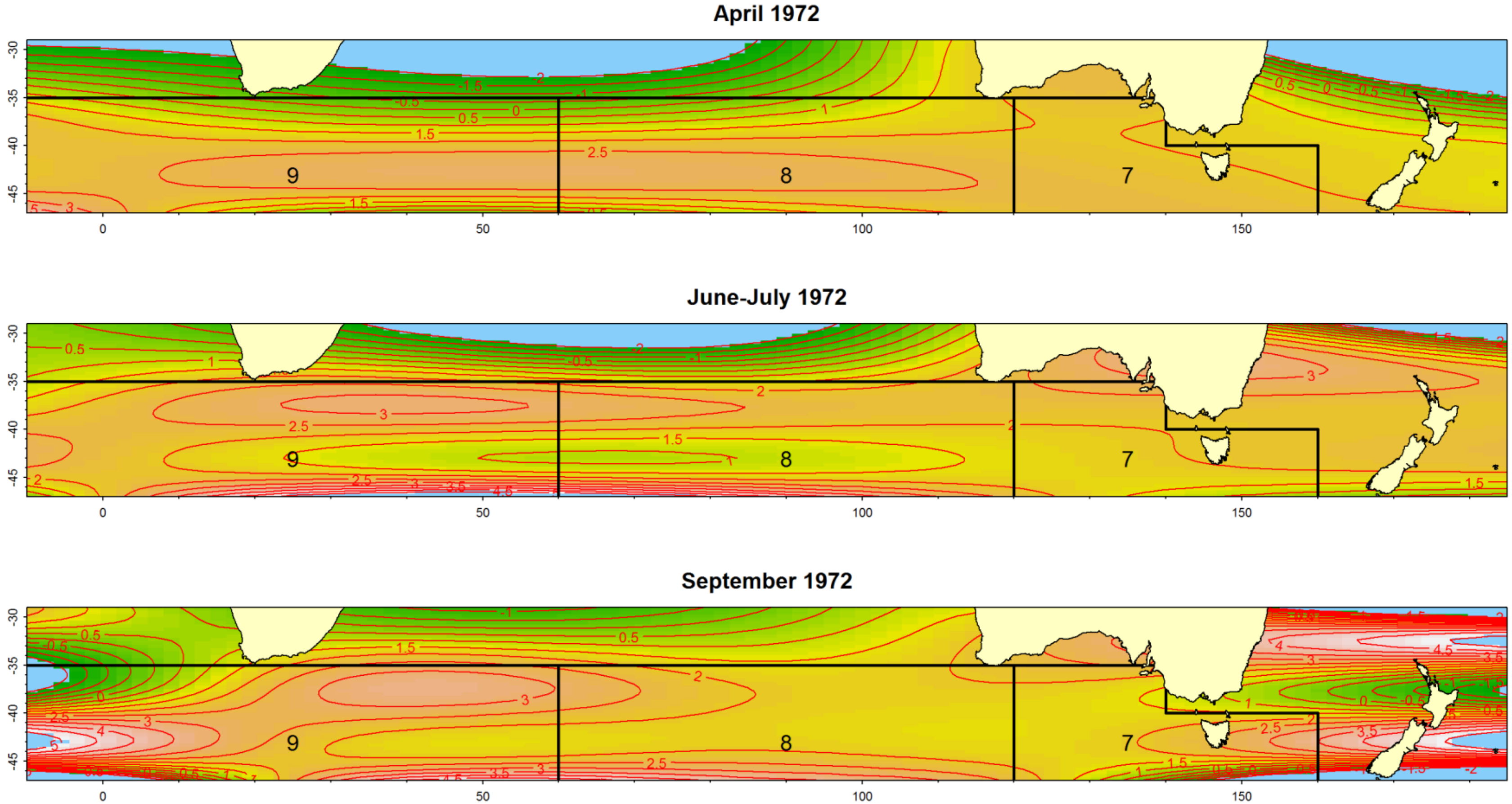

Figure 8: Maps of spatial distribution of predicted log CPUE in April, June-July and September 1972. The accuracy of interpolated predictions is limited by the spatial resolution of the CPUE data. Note: warmer colours reflect higher predicted CPUE. Blue reflects outside of area predicted.

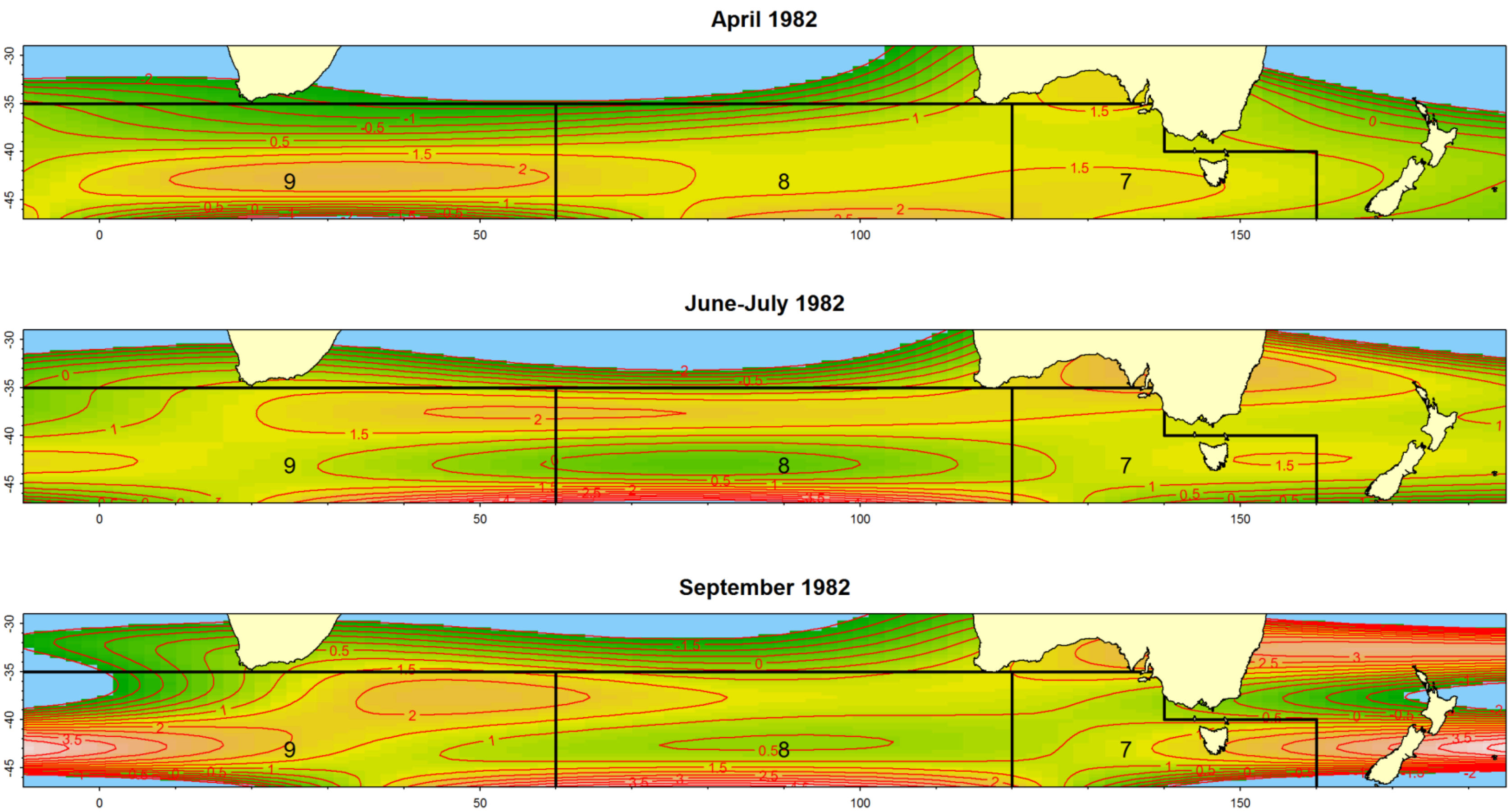

Figure 9: Spatial distribution of log CPUE in April, June-July and September 1982. The accuracy of interpolated predictions is limited by the spatial resolution of the CPUE data.

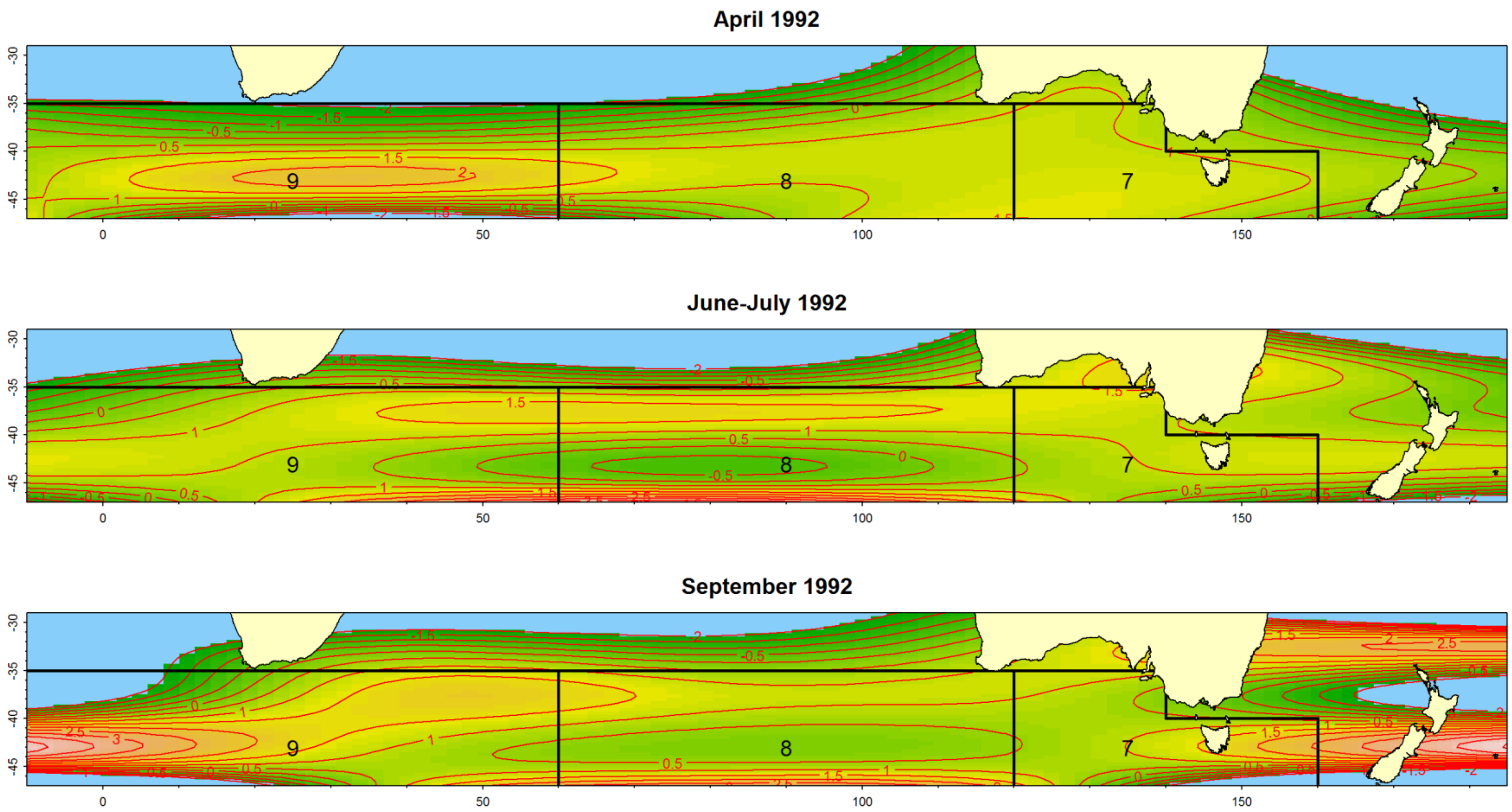

Figure 10: Spatial distribution of predicted log CPUE in April, June-July and September 1992. The accuracy of interpolated predictions is limited by the spatial resolution of the CPUE data.

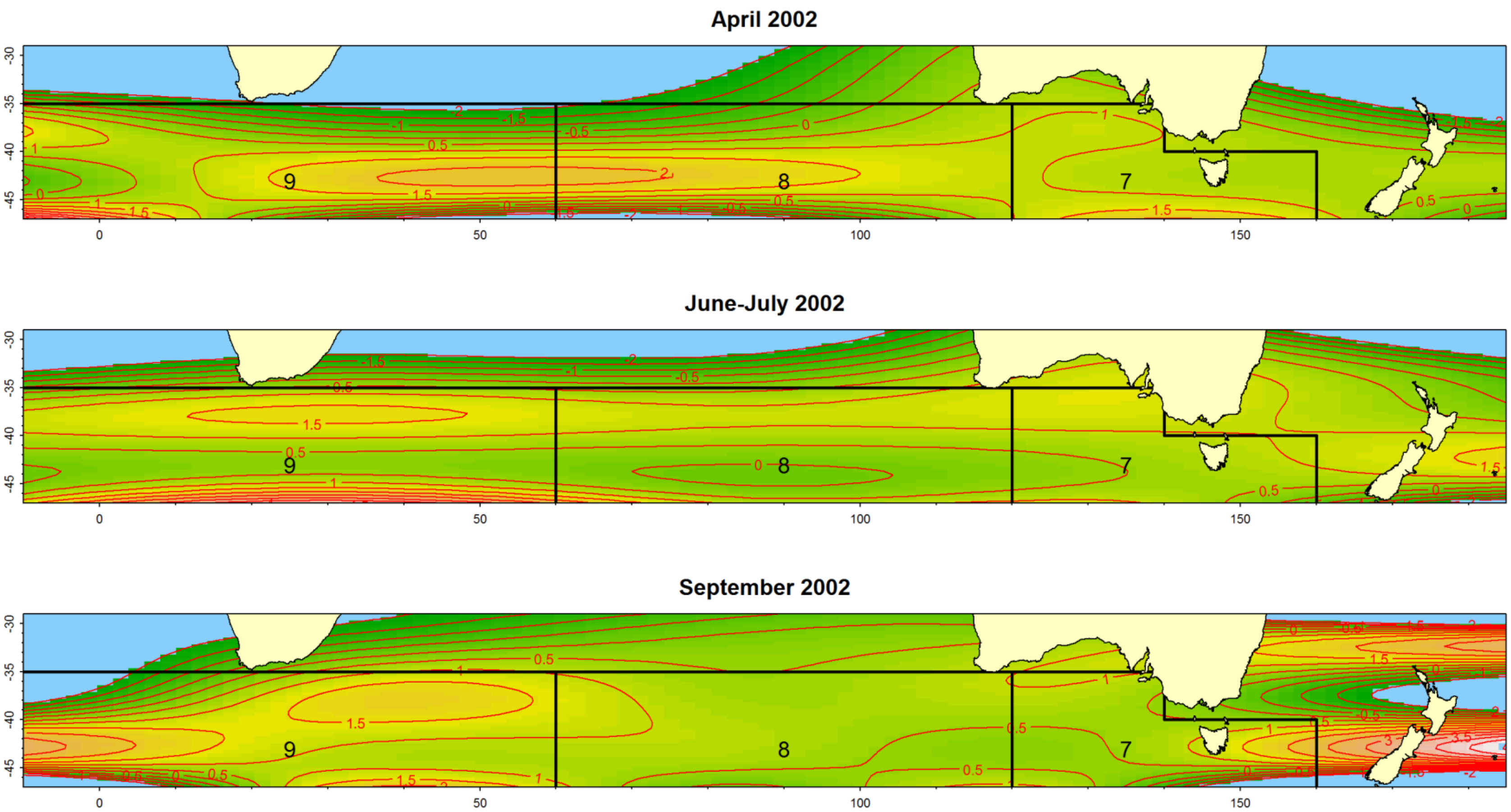

Figure 11: Spatial Distribution of predicted log CPUE in April, June-July and September 2002. The accuracy of interpolated predictions is limited by the spatial resolution of the CPUE data.

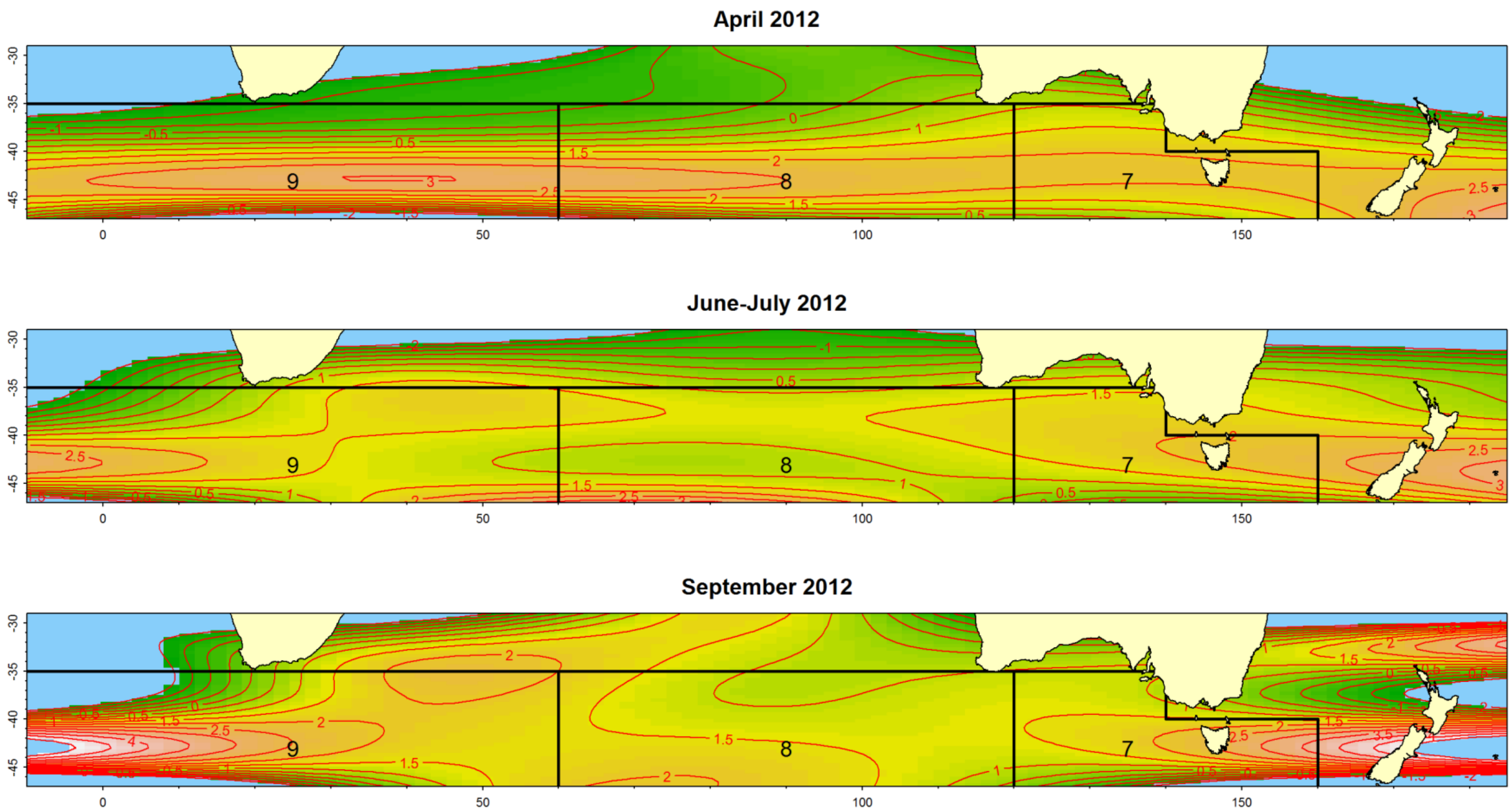

Figure 12: Spatial distribution of predicted log CPUE in April, June-July and September 2012. The accuracy of interpolated predictions is limited by the spatial resolution of the CPUE data.

R-Code used to calculate GAM Index

```
########################################################
### R Code Used to Calculate GAM based CPUE index                    ###
### for Southern Bluefin Tuna                                        ###
########################################################
###---------------------- Load Required R Packages----------------- ###
library("RODBC")
library(mgcv)
###--------- Read in the CPUE database and prep the data -----------###
setwd("SEC_CPUEInputs_6512") # Change to the path with file CPUEInputs_6512_Revised.mdb
SBT.2013 <- odbcConnectAccess("CPUEInputs_6512_Revised.mdb")
CPUE.2013 <- sqlFetch(SBT.2013, "CPUE_INPUTS")
CPUE.2013 <- CPUE.2013[CPUE.2013$DATA_CODE == "COMBINED",] # Avoid double counting
CPUE.2013$LONG[CPUE.2013$LONG < -80] <- CPUE.2013$LONG[CPUE.2013$LONG < -80] +
360 # for continuity in longitude over dateline
CPUE.2013$YEAR.F <- as.factor(CPUE.2013$YEAR)
Sum.4plus <- function(X)
{
Four.Plus <- as.numeric(X[16:32])
SBT.4plus <- round(sum(Four.Plus),digits = 1)
return(SBT.4plus)
}
###---------- Define CPUE as number of 4+ SBT per thousand hooks set ----------###
CPUE.2013$SBT_4Plus <- apply(CPUE.2013,1,Sum.4plus)
CPUE.2013$CPUE <- 1000*CPUE.2013$SBT_4Plus/CPUE.2013$N_HOOKS
CPUE.2013$CPUE <- round(CPUE.2013$CPUE,3)
###----- Exclude records from Areas 1,10,11,12 and 13 ----###
CPUE.df <- CPUE.2013[,c(2,4,5,7,6,10,33,34,35)]
names(CPUE.df) <-
c("YEAR","MONTH","STAT_AREA_CODE","LONG","LAT","HOOKS","YEAR.F","SBT.4p","CPUE")
CPUE.df <- CPUE.df[CPUE.df$STAT_AREA_CODE %in% c(2,3,4,5,6,7,8,9,14,15),]
CPUE.df <- CPUE.df[CPUE.df$HOOKS > 0,]
CPUE.df$HOOKS.CAPPED <- ifelse(CPUE.df$HOOKS > 500000,500000,CPUE.df$HOOKS)
###----- Fit the model ----###
Monitor.GAM <- gam(CPUE ~ te(LONG,LAT,MONTH) + te(LONG,LAT,YEAR) + YEAR.F, data =
CPUE.df,
family=quasi(link = log,variance = "mu"),weights = HOOKS.CAPPED)
###---- Make the Laslett prediction grid ----###
Laslett.Months <- c(4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,5,
5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,6,6,6,6,6,6,6,
6,6,6,6,6,6,6,6,6,6,6,6,7,7,7,7,7,7,7,7,7,7,7,7,7,
7,7,7,7,7,7,7,8,8,8,8,8,8,8,8,8,8,8,8,9,9,9,9,9,9,9,
9,9,9,9,9,6,6,6,7,7,7,7,7,7,7,7,7,7,7,8,8,8,8,8,8,8,
8,8,8,8,9,9,9,9,9,9,9,9,9,9,9,9,9,4,4,4,4,4,4,4,4,4,
4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,5,5,5,
5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,
5,5,5,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,
6,6,6,7,7,7,7,7,7,7,7,7,7,7,7,7,7,7,7,7,8,8,8,8,8,8,8,
8,8,8,8,8,8,8,8,8,9,9,9,9,9,9,9,9)
Laslett.Longs <- c(-10,-5,0,5,10,15,20,25,30,35,40,45,
-10,15,20,25,30,35,40,45,-10,-5,0,5,10,15,20,25,30,35,
40,45,-10,15,20,25,30,35,40,45,-10,-5,0,5,10,15,20,25,
```

```
30,35,40,45,-10,15,20,25,30,35,40,45,-10,-5,0,5,10,15,
20,25,30,35,40,45,-10,15,20,25,30,35,40,45,15,20,25,30,
35,40,15,20,25,30,35,40,15,20,25,30,35,40,15,20,25,30,
35,40,90,95,100,75,80,85,90,95,100,105,90,95,100,105,
75,80,85,90,95,100,105,90,95,100,105,90,95,100,105,110,
115,120,90,95,100,105,110,115,150,155,160,165,170,175,
150,155,160,165,170,175,130,135,140,145,150,155,160,
165,170,175,130,135,140,145,150,155,160,165,170,175,
150,155,160,165,170,175,150,155,160,165,170,175,130,
135,140,145,150,155,160,165,170,175,130,135,140,145,
150,155,160,165,170,175,150,155,160,165,170,175,150,
155,160,165,170,175,130,135,140,145,150,155,160,165,
170,175,130,135,140,145,150,155,160,165,170,175,150,
155,160,165,170,175,140,145,140,145,150,155,160,165,
170,175,150,155,160,165,170,175,140,145,140,145,150,
155,160,165,170,175,150,155)
Laslett.Lats <- c(-40,-40,-40,-40,-40,-40,-40,-40,-40,-40,-40,
          -40,-35,-35,-35,-35,-35,-35,-35,-35,-40,-40,-40,-40,
          -40,-40,-40,-40,-40,-40,-40,-40,-35,-35,-35,-35,-35,
          -35,-35,-35,-40,-40,-40,-40,-40,-40,-40,-40,-40,-40,
          -40,-40,-35,-35,-35,-35,-35,-35,-35,-35,-40,-40,-40,
          -40,-40,-40,-40,-40,-40,-40,-40,-40,-35,-35,-35,-35,
          -35,-35,-35,-35,-40,-40,-40,-40,-40,-40,-35,-35,-35,
          -35,-35,-35,-40,-40,-40,-40,-40,-40,-35,-35,-35,-35,
          -35,-35,-40,-40,-40,-40,-40,-40,-40,-40,-40,-40,-35,
          -35,-35,-35,-40,-40,-40,-40,-40,-40,-35,-35,-35,
          -35,-40,-40,-40,-40,-40,-40,-40,-35,-35,-35,-35,-35,
          -35,-30,-30,-30,-30,-30,-30,-35,-35,-35,-35,-35,-35,
          -40,-40,-40,-40,-40,-40,-40,-40,-40,-40,-45,-45,-45,
          -45,-45,-45,-45,-45,-45,-45,-30,-30,-30,-30,-30,-30,
          -35,-35,-35,-35,-35,-35,-40,-40,-40,-40,-40,-40,-40,
          -40,-40,-40,-45,-45,-45,-45,-45,-45,-45,-45,-45,-45,
          -30,-30,-30,-30,-30,-30,-35,-35,-35,-35,-35,-35,-40,
          -40,-40,-40,-40,-40,-40,-40,-40,-40,-45,-45,-45,-45,
          -30,-30,-30,-30,-30,-30,-35,-35,-35,-35,-35,-35,-40,
          -40,-45,-45,-30,-30,-30,-30,-30,-30,-35,-35,-35,-35,
          -35,-35,-40,-40,-45,-45,-30,-30,-30,-30,-30,-30,-35,-35)
SBT.Years <- rep(1969:2012,length(Laslett.Lats))
SBT.Years <- sort(SBT.Years)
Laslett.grid <- data.frame(YEAR = SBT.Years,LONG = rep(Laslett.Longs,length(1969:2012)),
LAT = rep(Laslett.Lats,length(1969:2012)),MONTH = rep(Laslett.Months,length(1969:2012)))
Laslett.grid$YEAR.F <- as.factor(Laslett.grid$YEAR)
###---- Use the fitted model to predict CPUE over the Laslett grid ----###
Laslett.grid$pred <- predict(Monitor.GAM,Laslett.grid,type = "response")
###---- Compute the index with a mean of 1 ----###
GAM.Index <-
(tapply(Laslett.grid$pred,Laslett.grid$YEAR,mean)/mean(tapply(Laslett.grid$pred,Laslett.grid$
YEAR,mean)))
GAM.Index
```

References

Laslett, G. M. 2001, Exploratory analysis of the SBT CPUE data using smoothing splines, CCSBT-SC/0103/06.

Wood, S. N. 2006, Generalised Additive Models: An Introduction with R, Chapman & Hall/CRC, Boca Raton, Florida.