

A MODELING APPROACH TO ESTIMATE OVERALL ATLANTIC FISHING EFFORT BY TIME-AREA STRATA (EFFDIS)

D.J. Beare¹, C. Palma¹, P. de Bruyn¹ and L. Kell¹

SUMMARY

In this document we describe the development of spatio-temporal statistical models for estimating fishing effort from ICCAT databases. The methods described have been packaged into a fully documented R-library (effdisR) for ease of use. Further information is available from the author (doug.beare@iccat.int). The Japanese longline fleet is used here as an illustration of the approach but the models and software developed work equally well for other fleets and métiers (e.g. purse seine and baitboat). The methods have been presented so far at two international fora: (i) The Sub-Committee on Ecosystems in June 2015 and; (ii) at the Blue Shark Stock Assessment Meeting in Lisbon in July 2015. Both Groups approved the methodology overall, and the feedback is described in the discussion section.

RÉSUMÉ

Dans le présent document, nous décrivons le développement de modèles statistiques spatio-temporels pour estimer l'effort de pêche à partir des bases de données de l'ICCAT. Les méthodes décrites ont été rassemblées dans une R-bibliothèque entièrement documentée (effdisR) afin d'en faciliter l'utilisation. Des informations complémentaires sont disponibles auprès de l'auteur (doug.beare@iccat.int). La flottille palangrière japonaise est utilisée ici à titre d'illustration de l'approche, mais les modèles et le logiciel développés fonctionnent tout aussi bien pour d'autres flottilles et métiers (p. ex. senneurs et canneurs). Les méthodes ont été présentées jusqu'à présent dans deux enceintes internationales : (i) au sous-comité des écosystèmes en juin 2015 et (ii) à la réunion d'évaluation du stock de requin peau bleue à Lisbonne en juillet 2015. Les deux groupes ont approuvé la méthodologie globale et les commentaires fournis sont décrits dans la section consacrée aux discussions.

RESUMEN

En este documento describimos el desarrollo de modelos estadísticos espacio-temporales para estimar el esfuerzo pesquero a partir de las bases de datos de ICCAT. Los métodos descritos se han incluido en una biblioteca de R completamente documentada (effdisR) para mayor comodidad. El autor dispone de más información (doug.beare@iccat.int). La flota de palangre japonés se usa aquí como ilustración del enfoque pero los modelos y el software desarrollados trabajan igualmente bien para otras flotas y métiers (por ejemplo, cerco y cebo vivo). Hasta ahora, los métodos se han presentado en dos foros internacionales: (i) el Subcomité de ecosistemas en junio de 2015 y (ii) la Reunión de evaluación del stock de tinterera de Lisboa en julio de 2015. Ambos grupos aprobaron la metodología general y los comentarios aportados se describen en la sección de discusión.

KEYWORDS

Tuna fisheries, Catch/effort, Computer programs, Longlining, Purse seining

¹ ICCAT Secretariat, C/Corazón de María, 8. 28002 Madrid, Spain; Doug.Beare@iccat.int; Phone: +34 914 165 600 Fax: +34 914 152 612.

1. Introduction

The International Commission for the Conservation of Atlantic Tunas (ICCAT) (www.iccat.int) maintains a database of fishing effort and catches distributed by time-area strata which is known as 'EFFDIS' (de Bruyn *et al.* 2014). A total of 27 different fishing nations submit catch and effort data to ICCAT for the main gears they use for targeting tuna and tuna-like species within the ICCAT convention area. EFFDIS data are available as 'Task I' and 'Task II' data. Task I data are annual totals for catch (e.g. tons bluefin tuna caught in 1999 by Japan) by gear in the various relevant 'regions' (e.g. Atlantic and Mediterranean) and are believed to be totally comprehensive. Task II data, on the other hand, are much more detailed, available at greater spatial (e.g. 5°x5° degree square grid) and temporal (e.g. month and year) resolution. The negative side is that they tend to be only partially complete.

Comprehensive estimates of fishing effort can, therefore, be made by 'raising' the Task II estimates by those from Task I. The Task I and II databases thus represent rich and valuable sources of information on fishing activity in the Atlantic and Mediterranean since 1950. These data can reveal both seasonal and long-term changes in the distributions of the fisheries, and their target species, in addition to exposing the vulnerability to fishing mortality of various by-catch taxa such as turtles and seabirds.

Here we describe the development of a statistical modeling approach to estimating overall Atlantic effort by time-area strata for the Task I and II databases which is critically important, particularly with regard to by-catch evaluations. As mentioned the software developed (*effdisR*) was written in R (Team and others 2010) linked by SQL to a PostgreSQL database "in the cloud". The use of the software is described below using examples from a range of countries with particular focus on EU Spain for the purse-seine gear and Japan for the long-line.

1.1 Data, servers, and version control

Note that an Ubuntu cloud server with a static IP address (134.213.29.249, *effdis-tuna-cc1*) was set up by the ICCAT Secretariat specifically for the current work. A PostGIS-enabled PostgreSQL server was also installed where all data related to the project are being stored and retrieved. The database can be accessed directly from the command line of any computer with the PostgreSQL client installed (`psql -h 134.213.29.249 -d effdis -U postgres`) or using the ODBC (Open Database Connectivity) protocol via the R-library, RODBC. All scripts (R, Rmarkdown, Shell, PHP) developed for this work are under sub-version control (see <https://github.com/bearedo/effdis>) facilitating straightforward and transparent modification and updating.

The *effdis-tuna-cc1* server also hosts an online geographic information system (still very much a beta version) being trialed for EFFDIS (<http://134.213.29.249/effdis/#>). Although the plotting and modeling work is being done in R there are features of bona fide databases like PostgreSQL that are particularly useful, e.g. the SQL language, very rapid searches, and functionality for linking directly with GIS software such as QGIS. An example screenshot of bigeye tuna catch distribution by Chinese Taipei from the beta version of this database is shown in **Figure 1**.

2. Methods: EFFDIS effort estimation with ICCAT Task I and II using *effdisR*

The first step is to install (not shown) and then attach the relevant R-libraries. *rgdal*, for example, is used for converting between standard spatial formats while *RODBC* is necessary for connecting to the PostgreSQL database.

```
library(rio)
library(spatial)
library(sp)
library(doBy)
library(rgdal)
library(RODBC)
library(RColorBrewer)
library(ggplot2)
library(vmstools)
library(mgcv)
library(maps)
library(mapdata)
```

```
library(reshape2)
library(rgeos)
library(lattice)
library(pander)
library(kfigr)
```

The next step is to load the R-package (*effdisR*) which contains all the scripts needed for the analyses.

```
library(effdisR)
```

2.1 Exploratory data analysis

For any data analysis some initial exploration is essential. Where data are distributed non-randomly in space and time, typical in databases such as these, spurious results can easily be obtained. Our software for analysing Task I and II data includes a suite of tools for plotting and examining data and their use is described below.

2.2 Data screening

Once the RODBC library is installed and the `/etc/odbc.ini` file modified to provide the necessary parameters for connection to the Task I and II databases the data can be imported directly and conveniently into R according to the following code from the RODBC R package:

```
chan <- odbcConnect("effdis-tuna-cc1", case="postgresql", believeNRows=FALSE)
```

In **Table I** we use this link to the database to count the frequencies with which each effort-type has been recorded in the Task II data for longliners in the Atlantic Ocean in the year 2010. **Table I** illustrates the sort of problems that exist with these data. The United Kingdom, for example, has supplied 'D.FISH' (4) only, while France has supplied both 'D.AT SEA' (4) and 'D.FISH' (20).

```
effort_type_by_flag_II <- sqlQuery(chan, "SELECT flagname AS Flag,
eff1type AS Effort_type, count(eff1type) as No_records
FROM t2ce_new
WHERE yearc = 2010 AND region ='AT' AND geargrpcode = 'LL'
GROUP BY flagname, eff1type
ORDER BY flagname, eff1type, No_records;")
```

Similarly in **Table 2** we summarize the different types of fishing effort that have been recorded for purse-seiners. Ghana, for example, has sent data as, 'FISH.HOUR' and 'SUC.D.FISH'. It is also worth noting that, in many cases, no effort data are submitted for Task II at all, see '-none-' (**Tables 1 and 2**).

```
effort_type_by_flag_ps <- sqlQuery(chan, "SELECT flagname AS Flag,
eff1type AS Effort_type, count(eff1type) as No_records
FROM t2ce_new
WHERE yearc = 2010 AND region ='AT' AND geargrpcode = 'PS'
GROUP BY flagname, eff1type
ORDER BY flagname, eff1type, No_records;")
```

After examining **Tables 1 and 2** we made the decision to examine records with 'NO. HOOKS' only for longliners and 'FISH HOUR' and 'D.FISH' for the purse-seiners, removing all other rows. Obviously this represents a potentially important loss of information but it simply reinforces the point that data need to be submitted using the same measurement variables.

Catchunit is another important variable in the Task II data, denoting whether the catch was recorded in terms of total numbers (nr) of fish or their total weight (kg). **Table 3** illustrates that in 2010, for example, there were many records (2779) in the Task II data with `catchunit = '--'` and these rows were, perforce, removed from subsequent analyses.

```
catchunit <- sqlQuery(chan, "SELECT catchunit AS catchunit, count(catchunit) as No_records
FROM t2ce_new
WHERE yearc = 2010 AND region ='AT'
GROUP BY catchunit
ORDER BY catchunit, No_records;")
```

2.3 Data coverage

It is important to understand how the distribution of samples in the Task II database varies with respect to location and time (long-term and seasonal). The function `yr.month.coverage.task2.r` available in *effdisR* counts the number of samples by year and month for any "strata" (gear, flag etc) and displays the results as a "3D plot". This type of plot reveals non-random sampling in time. It is possible, for example, that sampling might have concentrated on the first part of the year for a decade, and then switched to the latter part of the year. 'Trends' estimated from data collected in such a manner will clearly be spurious. Examples of the output of the function `yr.month.coverage.task2.r` are plotted in **Figure 2** for longliners between 1980 and 2015 for Japan, Chinese Taipei, Brazil, and U.S.A. Clearly the extent of the data available varies substantially between flags. There are no obvious seasonal biases in the data, but the amount of reporting has changed with long-term time. Japan has been particularly consistent (**Figure 2**, top right), while the U.S.A. has been inconsistent.

```
ll <- sqlQuery(chan,"SELECT yearc AS year, trend, timeperiodid AS month,
flagname, region, geargrpcode,longitude,latitude, catchunit, dsettype, eff1, eff1type
FROM t2ce_new
WHERE yearc > 1979 AND yearc < 2016 AND region ='AT' AND timeperiodid < 13 AND eff1type='NO.HOOKS' AND geargrpcode = 'LL'
AND flagname IN ('Japan', 'Chinese Taipei','Brazil', 'U.S.A.','China P.R.')
AND catchunit != '--' ;")
```

```
par(mfrow=c(2,2),mar=c(2,2,2,2),oma=c(1,1,1,1))
yr.month.coverage.task2(tdata=ll,which.gear='LL',start.year=1980,end.year=2015,which.flag='Japan')
yr.month.coverage.task2(tdata=ll,which.gear='LL',start.year=1980,end.year=2015,which.flag='Chinese Taipei')
yr.month.coverage.task2(tdata=ll,which.gear='LL',start.year=1980,end.year=2015,which.flag='Brazil')
yr.month.coverage.task2(tdata=ll,which.gear='LL',start.year=1980,end.year=2015,which.flag='U.S.A.')
```

2.4 Sampling in space by year - Brazil, and Japan

The distribution of data/samples in space is similarly important. The function `spatial.coverage.by.year.task2.r` plots the distribution of Task II data by location for any combination of flag, gear etc. Output is illustrated below for longliners for two arbitrarily selected flags and time-periods (see Figs 3 and 4).

```
par(mfrow=c(3,3),mar=c(0,0,2,0))
spatial.coverage.by.year.task2(tdata=ll,start.year=1982,end.year=1990,which.gear='LL',which.flag='Brazil')
```

The extent of Brazilian longlining activity has, for example, spread out from the coast of South America between 1981 and 1990 (Fig. 3). In contrast the Japanese longline fleet covers substantial areas of the Atlantic Ocean (Fig. 4).

```
par(mfrow=c(3,3),mar=c(0,0,1,0))
spatial.coverage.by.year.task2(tdata=ll,start.year=1992,end.year=2000,which.gear='LL',which.flag='Japan')
```

```
ps <- sqlQuery(chan,"SELECT yearc AS year, trend, timeperiodid AS month,
flagname, region, geargrpcode,lon,lat,longitude,latitude,quadid AS quad, squaretypecode AS square, catchunit, d
settype, eff1, eff1type
FROM t2ce_new
WHERE yearc > 1999 AND region ='AT' AND timeperiodid < 13 AND eff1type IN ('D.FISH','FISH.HOUR')
AND geargrpcode = 'PS' AND flagname IN ('EU.España','Japan','Other','Panama')
AND catchunit != '--' ;")
```

```
par(mfrow=c(2,3),mar=c(1,1,2,1),oma=c(3,3,3,3))
spatial.coverage.by.year.task2(tdata=ps,start.year=2009,end.year=2014,which.gear='PS',which.flag='EU.España')
```

Most of the purse seine data have been supplied at a higher spatial resolution (1x1) than is available for the longliners. This is illustrated in Figure 5 which summarizes the extent of Spanish purse-seining activity off the West Coast of Africa between 2009 and 2014.

2.5 Effort by year and location (raw data, no modeling)

Given that we can now determine the timing and location of fishing activities (**Figures 2-5**) we now need to know its intensity, i.e. how many hooks were set or days fished at a particular location? Task II effort data of any type can be plotted spatially using the R function *three.d.effort.by.year.r* and example output is shown in **Figures 6 to 9**.

In 2014, for example, longlining effort by Japan focused in the Eastern Atlantic (**Figure 6**). Chinese Taipei flagged longliners, on the other hand also worked in the East Atlantic but also more in the South Atlantic (**Figure 7**). Purse seine data tend to be submitted to ICCAT at a higher spatial resolution (1x1 grid) than is the case for longliners, and examples for Spain and Panama in 2014 are plotted in **Figures 8 and 9**. Both fleets focus their activity along the West Coast of Africa but the Panamanian fleet tends not to venture as far inshore as the Spanish fleet, except in the case of the Mauritanian EEZ (**Figures 8 and 9**).

```
par(mfrow=c(1,1))
three.d.effort.by.year(tdata=ll,which.year=2014,which.flag='Japan',scaling.f=10000,effort.type='NO.HOOKS')

three.d.effort.by.year(tdata=ll,which.year=2014,which.flag='Chinese Taipei',scaling.f=10000,gridx=5,gridy=5,
,effort.type='NO.HOOKS',which.gear='LL')

three.d.effort.by.year(tdata=ps,which.year=2014,which.flag='EU.España',gridx=1,gridy=1,scaling.f=2,effort.type='FISH.HOUR',which.gear='PS')

three.d.effort.by.year(tdata=ps,which.year=2014,which.flag='Panama',gridx=1,gridy=1,scaling.f=1,effort.type='FISH.HOUR',which.gear='PS')
```

2.6 Catch weights by year and location - Japanese longliners, and Spanish Purse-seiners

With *circa* 27 flags, multiple species, 12 months, more that 60 years of data, and different effort submissions (e.g. days at sea, number of hooks set) there is a large number of possible combinations for examining the Task I and II data. Some examples are shown in Figure 10 using the *effdisR* function, *three.d.catch.by.year.r* linked to the PostgreSQL database via an SQL script (see below for *alb*). In this example we extracted data for albacore tuna caught by longline for Japan, and by purse seine for Spain. [Note that observations were ignored where the catchunit is unknown ('-'). Numbers or kilograms caught can also be selected depending on availability].

```
alb <- sqlQuery(chan,"SELECT yearc AS year, trend,
timeperiodid AS month, flagname, region, geargrcode,longitude,latitude,lon,lat,
quadid AS quad, squaretypecode AS square,
catchunit, dsetype, eff1, eff1type, alb as measured_catch
FROM t2ce
WHERE yearc > 1991 AND region ='AT' AND timeperiodid < 13 AND eff1type IN ('NO.HOOKS','FISH.HOUR') AND
geargrcode IN ('LL','PS') AND flagname IN ('Japan','EU.España') AND catchunit != '-' ;")
alb$species <- 'alb'
```

```
par(mfrow=c(2,2),mar=c(2,1,2,1))
three.d.catch.by.year(tdata=alb,which.year=2006,which.gear='LL',which.species='alb',which.flag='Japan',scaling.f=10,catchunit='nr')
three.d.catch.by.year(tdata=alb,which.year=2009,which.gear='LL',which.species='alb',which.flag='Japan',scaling.f=10,catchunit='nr')
three.d.catch.by.year(tdata=alb,which.year=1992,which.gear='PS',which.species='alb',gridx=1,gridy=1,which.flag='EU.España',scaling.f=25,catchunit='kg')
three.d.catch.by.year(tdata=alb,which.year=1995,which.gear='PS',which.species='alb',gridx=1,gridy=1,which.flag='EU.España',scaling.f=25,catchunit='kg')
```

3. Worked example using Japanese longline gear: calculating fishing effort from Task I and II data

3.1 Step 1

Get Task II data from the database for each dsetype using *get.effdis.t2.data.r* and then combine them using *rbind*.

```
ll_n <- get.effdis.t2.data(which.dsn='effdis-tuna-cc1',which.gear='LL',which.flag='All',which.dsettype = 'n-')
ll_nw <- get.effdis.t2.data(which.dsn='effdis-tuna-cc1',which.gear='LL',which.flag='All',which.dsettype = 'nw')
ll_w <- get.effdis.t2.data(which.dsn='effdis-tuna-cc1',which.gear='LL',which.flag='All',which.dsettype = '-w')
long_line <- rbind(ll_n,ll_nw,ll_w)
```

3.2 Step 2

Find and extract those EFFDIS data that are in the Atlantic using *find.ocean.r*. This function imports a shapefile and, in addition to the polygons for the Atlantic Ocean proper, also includes the Caribbean Sea, The Gulf of St. Lawrence, the Gulf of Guinea, and the Gulf of Mexico. This aspect, however, is entirely flexible and can easily be changed. Data on land (the Secretariat is working to remove these), the Mediterranean and Pacific can be extracted using the strings, 'land', 'med', and 'pac' respectively in the place of 'atl' below.

```
long_line<-find.ocean(long_line)
long_line <- long_line[long_line$which.ocean == 'atl',]
```

3.3 Step 3

Make sure the data are 'clean' using *prepare.effdis.data.r* as follows. Sometimes downloading from a database creates 'factor' objects in R which can cause problems. This function converts them to character strings.

```
long_line<-prepare.effdis.data(input=long_line)
```

3.4 Step 4

Convert data from 'short format' to 'long format' using *convert2long.format.t2.r*. We do this because the long format is required for the subsequent regression modeling.

```
long_line_lf <- convert2long.format.t2(input =long_line)
```

3.5 Step 5

Estimate catch weights from numbers where none are supplied using the functions *model.nos.kgs.r* (fits the model) and *kgs.from.nos.r* which imputes weights for countries that supply only numbers caught. Note that some countries report catches by total weight, some by total numbers and some by both. For the purposes of the effort estimations we are attempting to make here, it is essential that the catch data are available in the same unit of measurement. The data for countries that have reported both weights and numbers were, therefore, extracted and examined together. The relationships between them are highly linear and we decided to model the weight caught as a (linear) function of number caught plus other useful, predictive covariates (e.g. flag, species, and trend). A stepwise model selection procedure chooses the 'best' model (*bm*) which is then used to impute catch weights in kgs for Task II in cases where total numbers only were supplied, e.g. U.S.A. The model below (*bm*) fits the data well and explains most of the variance ($R^2 = 83\%$; **Table 4**). This part of the procedure/model is included in the error variance estimation.

```
bm <- model.nos.kgs(input=long_line_lf,which.gear='LL')
long_line_lf <- kgs.from.nos(long_line_lf)
```

3.6 Step 6

Fit regression models to the Task II catch and effort data using *fit2stageGAMtoCatch.r* and *fitGAMtoEffort.r*. This step is illustrated here for speed and convenience with bluefin tuna only ('bft') and only between 1990 and 2000. To get a global estimate of effort, however, you would need to do the same for all nine species although it would be straightforward to add any others. The functions fit and test a suite of generalized additive models (GAMs) fitted to the Task II catch and effort data. GAMs were selected because they are highly flexible, impose no particular functional form on the data, and they can deal with skew distributions and high prevalence of zeros (Wood 2006). The models take the relevant variables (eg. number of hooks set, weight of fish caught) and model them as smooth functions of various combinations of covariates of location (latitude, longitude, bottom depth) and time (year and month).

The first stage models the probability of recording a catch using a GAM from the Quasibinomial family (Model 1), where P is the probability of catching a fish.

1. $P_{xytm} = s(x,y) + s(t) + s(m) + \epsilon$

We then model the positive component of the catch, C , with a GAM from the Gamma family (Model 2).

$$2. \quad C_{xytm} = s(x, y) + s(t) + s(m) + \epsilon$$

And finally the fishing effort (number of hooks) is modeled using a GAM from the QuasiPoisson family (Model 3).

$$3. \quad E_{xytm} = s(x, y) + s(t) + s(m) + \epsilon$$

In all 3 models, x, y, t , and m are longitude, latitude, trend, and month respectively and s are spline smooth functions fitted by generalised-cross-validation using the MGCV R-library, see <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>. ϵ is different, as appropriate, for each model.

```
j_bft_ll <- fit2stageGAMtoCatch(input=long_line_ll, which.flag='Japan',which.species='bft',start.year=1990,
end.year=2010)
j_emod_ll <- fitGAMtoEffort(input=long_line_ll,which.flag='Japan',which.effort='NO.HOOKS',start.year=1990,
end.year=2010)
```

3.7 Step 7

Use the models to predict values over a grid of 'new data'. Currently this takes the range of locations *ever* recorded in the data, constructs a grid for each time-step (1950 to present by month) and makes the predictions with the model. The function *predict.effdis.t2.data.r* also identifies and flags up those points in space and time where data were actually collected.

```
j_bft.ll.pred <- predict.effdis.t2.data(cmod=j_bft_ll,effmod=j_emod_ll,grid.res=5,start.year=1990,end.year=2010,
which.flag='Japan')
```

Once assessed for adequacy of fit the model parameters are used to 'predict' values of catch, effort and catch-per-unit-effort as functions on a grid of all combinations of the selected covariates, together with error or variance if required using the function *predict.effdis.t2.data.r*. Note that total Task II catch is calculated by multiplying the fits from Models 1 and 2, ie. 'given that fish were caught, how many/much', in a 'two-stage process'. The function *plot.mods.r* can be used to plot excerpts from the model output. The 'probability of catching a blue fin tuna in January 1995 is, for example, created using the following code:

```
par(mfrow=c(1,2))
plot.mods(input=j_bft.ll.pred,cmod=j_bft_ll,which.year = 1995,which.month=1,which.value = 'prob',grid.res=5,
which.gear='LL')
plot.mods(input=j_bft.ll.pred,cmod=j_bft_ll,which.year = 1995,which.month=1,which.value = 'prob',grid.res=5,
which.gear='LL',plot.samples.only = FALSE)
```

Figure 11 shows predictions for bluefin tuna from the binary (Bernoulli) model for January 1995. The probability of catching one is highest in the Caribbean, Gulf of Mexico and North Atlantic. The left-hand plot shows the model output for grid cells where a real observation exists, while the right-hand plot is the interpolation based on the area of the entire dataset.

3.8 Step 8

Obtain (and aggregate for later use) Task I long-line data from the database using *get.effdis.t1.data.r*. Note that Task I are annual catch totals which are thought to be comprehensive.

```
long_line.t1 <- get.effdis.t1.data(which.dsn='effdis-tuna-cc1',which.gear = 'LL',which.region='AT',which.flag='
Japan')
long_line.sum.t1 <- aggregate(list(qty_t=long_line.t1$qty_t),list(year=long_line.t1$yearc),sum,na.rm=T)
```

3.9 Step 9

Collect model data together into a single database. In our method you bind up the model estimates for all 9 species as follows. Remember that you need to run Steps 6 and 7 for each species (code not shown here).

```
model.data <- rbind(j_alb.ll.pred,j_bft.ll.pred,j_bet.ll.pred,j_bum.ll.pred,j_skj.ll.pred,j_yft.ll.pred,j_swo.ll.pred,j
_sai.ll.pred,j_whm.ll)
```

3.10 Step 10

Convert Task II catches from kgs to tonnes to match with the Task I data.

```
model.data$catch <- model.data$catch/1000
```

3.11 Step 11

Sum Task II data (catch and effort) for all 9 species.

```
model.data.totals <- aggregate(list(catch=model.data$catch,catch=model.data$catch,eff=model.data$eff),by=list(year=model.data$year),sum,na.rm=T)
```

3.12 Step 12

Merge the Task I and modeled Task II totals, calculate a global, modeled Task II CPUE and raise by Task I CPUE to give raised effort.

```
t1.t2.merged <- merge(model.data.totals,long_line.sum.t1)
t1.t2.merged$cpue <- t1.t2.merged$catch/t1.t2.merged$eff
t1.t2.merged$raised.effort <- t1.t2.merged$qty_t/t1.t2.merged$cpue
```

3.13 Step 13

Plot the raised effort as a function of year.

```
effdis <- read.table('C:\\Users\\DBeare\\effdis\\effdis\\data\\japan-effdis-estimate.csv',sep=',',header=T)
par(mfrow=c(1,1),mar=c(5,5,4,4))
plot(effdis$year,effdis$raised.effort/1000000,xlab='',ylab='hooks',type='l',lwd=3,xlim=c(1970,2010))
abline(v=seq(1960,2010,by=5),lty=2,col='blue')
title('Task I Catch / Task II2 CPUE for Japanese long-liners')
```

4. Discussion

The use of the *effdisR* code for raising Task II EFFDIS effort estimates by Task I totals has been demonstrated here for Japanese longliners. The code above also works for the purse-seiners and baitboats but obviously different input parameters need to be inserted into the functions. The estimate shown here for the Japanese long-line fleet is very similar to that calculated by de Bruyn *et al* (2014). To get global estimates the code above should be re-run for each fleet or flag and the estimates summed. All the analyses are available on a github repository <http://iccat-stats.github.io/>. The method described here is also "modular" in nature so it can be altered to use information from any aggregation, e.g. either fleet or flag.

In the past the ICCAT Sub-Committee on Ecosystems and the Working Group on Stock Assessment Methods have both made a number of recommendations for updating and improving the estimation of fishing effort from Task I and II (EFFDIS), most of which have been incorporated in the new estimates. EFFDIS estimates are reliant on Task II catch and effort information, and it is known that there are errors in these data. Data submission and screening should be improved to eliminate problems such as effort duplication. Such improvements will reduce the amount of problematic data which effect accurate effort estimation.

The original purpose of the work was to generate spatio-temporally resolved (month, year and fine-scale location) effort estimates needed to estimate fishing impact on target and by-catch species, and the new estimates are already being distributed to interested parties/groups.

References

- de Bruyn, Paul, Carlos Palma, Laurie Kell, and Juan Gallego. 2014. "Proposals for the Improvement of the Estimation of the Overall Longline Effort Distribution (EffDIS) in the ICCAT Area." *SCRS*, sec. 026.
- Team, R., and others. 2010. "R: A Language and Environment for Statistical Computing." *R Foundation for Statistical Computing Vienna Austria*, no. 01/19.
- Wood, Simon. 2006. *Generalized Additive Models: An Introduction with R*. CRC Press.

Table 1. Effort type sampling by flag for Atlantic longline during 2010 in ICCAT Task II database

<i>Flag</i>	<i>Effort_type</i>	<i>No_records</i>
Barbados	NO.HOOKS	12
Belize	NO.HOOKS	151
Brazil	NO.HOOKS	172
Canada	NO.HOOKS	42
Canada	-none-	9
China PR	NO.HOOKS	73
Chinese Taipei	NO.HOOKS	632
Côte D'Ivoire	NO.HOOKS	36
EU.Denmark	D.AT SEA	1
EU.España	NO.HOOKS	822
EU.France	D.AT SEA	4
EU.France	D.FISH	20
EU.Portugal	NO.HOOKS	309
EU.Portugal	-none-	1
EU.United Kingdom	D.FISH	11
FR.St Pierre et Miquelon	D.FISH	5
Japan	NO.HOOKS	656
Korea Rep.	NO.HOOKS	142
Maroc	-none-	12
Maroc	SUC.D.FI	12
Mexico	NO.HOOKS	56
Namibia	NO.HOOKS	111
Philippines	NO.HOOKS	54
Senegal	D.FISH	22
South Africa	NO.HOOKS	256
St. Vincent and Grenadines	NO.HOOKS	186
Trinidad and Tobago	NO.HOOKS	12
UK.Bermuda	NO.HOOKS	10
Uruguay	-none-	17
U.S.A.	NO.HOOKS	95
Vanuatu	NO.HOOKS	162
Venezuela	NO.HOOKS	80

Table 2. Effort type sampling by flag for Atlantic purse-seiners in ICCAT Task II database in 2010.

<i>Flag</i>	<i>Effort_type</i>	<i>No_records</i>
Argentina	D.AT SEA	22
Belize	D.FISH	10
Brazil	FISH.HOUR	18
Cape Verde	-none-	12
Curaçao	FISH.HOUR	637
EU.Denmark	D.AT SEA	10
EU.España	FISH.HOUR	1984
EU.France	FISH.HOUR	1010
EU.Portugal	D.FISH	13
EU.United Kingdom	D.FISH	12
Ghana	FISH.HOUR	1352
Ghana	SUC.D.FI	12
Guatemala	FISH.HOUR	242
Maroc	SUC.D.FI	20
NEI (ETRO)	FISH.HOUR	957
Panama	FISH.HOUR	546
Venezuela	D.FISH	133

Table 3. Catch unit sampling by flag in EFFDIS Task II database in 2010.

Catchunit	No_records
--	2779
kg	8881
nr	1564

Table 4. Linear model summarizing the relationship between weights and numbers for countries that sent both to ICCAT for the Task II database.

	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr(> t)</i>
lnr	0.8	0.007	122	0
trend	-5e-04	1e-04	-3	0.001
speciesbft	1	0.3	3	0.001
speciesbet	0.4	0.06	7	4e-11
speciesskj	-0.9	0.1	-8	4e-15
speciesyft	0.3	0.06	5	7e-07
speciesswo	0.6	0.05	12	1e-33
speciesbum	0.5	0.07	7	1e-11
speciessai	-0.2	0.1	-2	0.1
specieswhm	-0.6	0.07	-9	2e-18
(Intercept)	5	0.1	47	0

Fitting linear model: lkg ~ lnr + trend + species

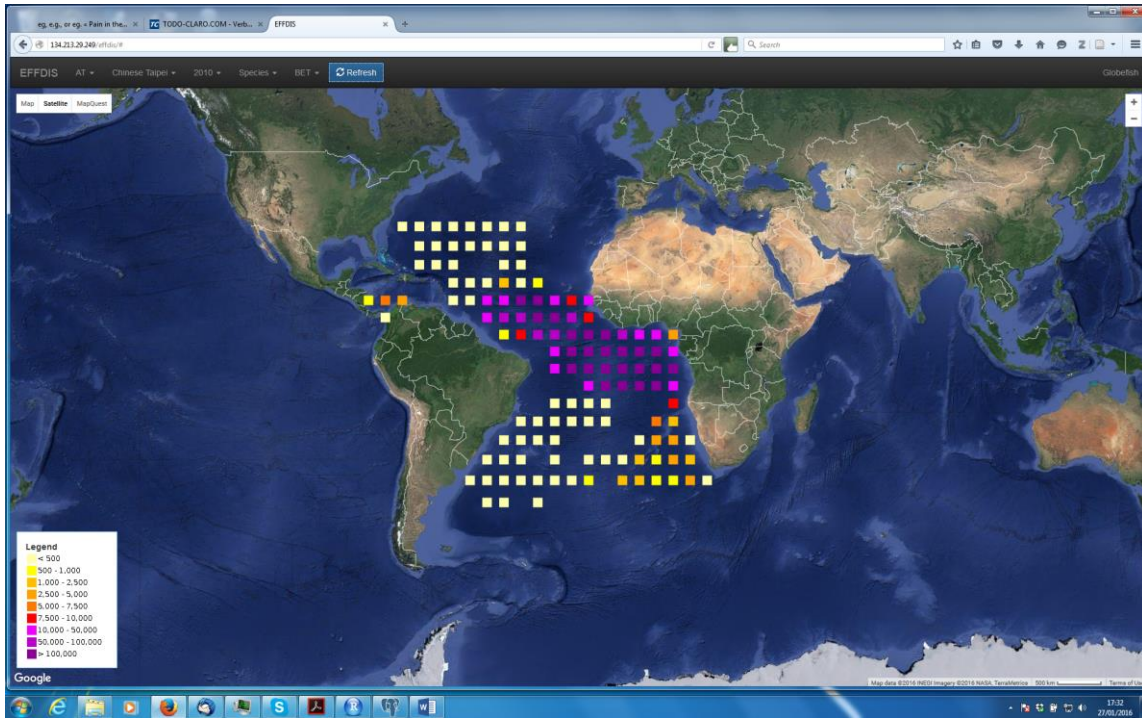


Figure 1. Screenshot of EFFDIS geo-database - Bigeye catches by Chinese Taipei in 2010 reported to ICCAT as Task II.

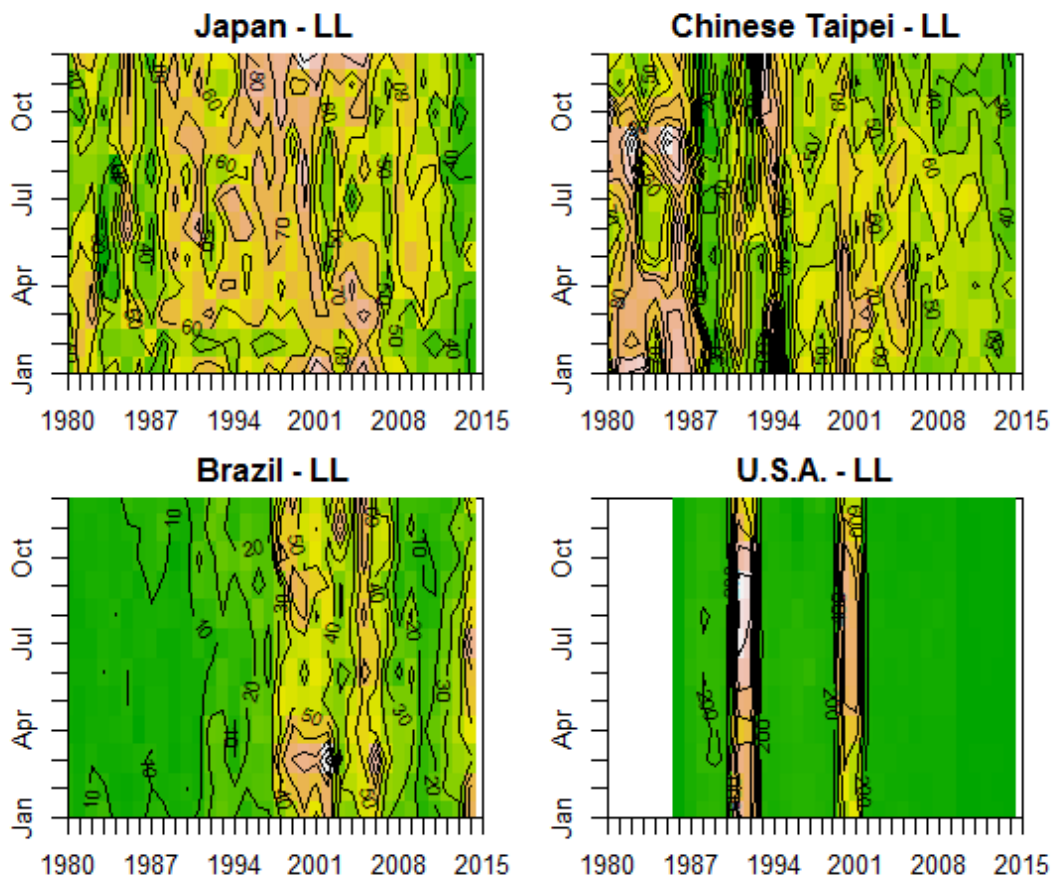


Figure 2. Temporal (by year and month) distribution of longliner data in Task II database: Japan, Chinese Taipei, Brazil, and U.S.A. between 1980 and 2015.

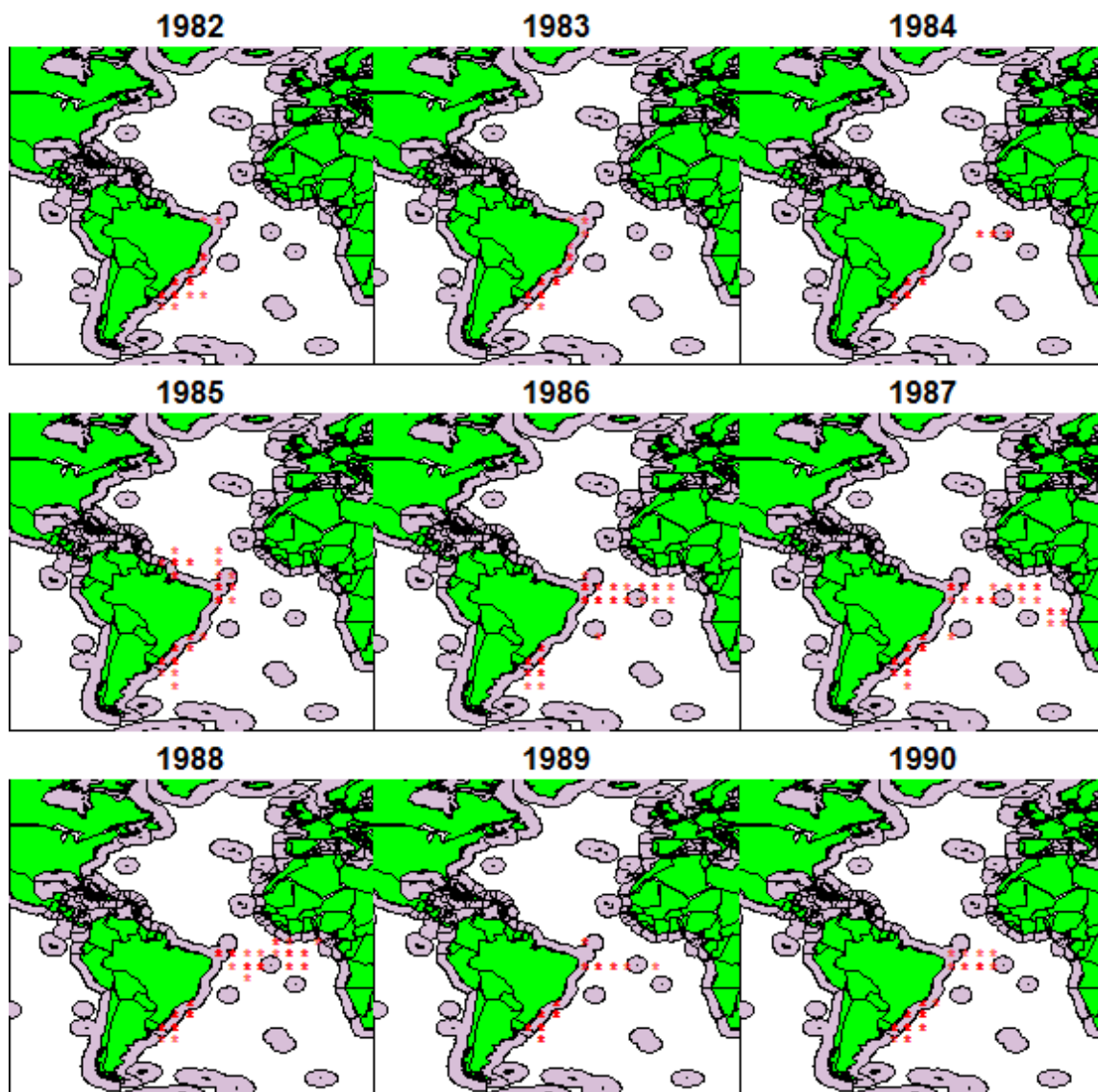


Figure 3. Distribution of Task II data: Brazilian long-liners between 1981 and 1990.

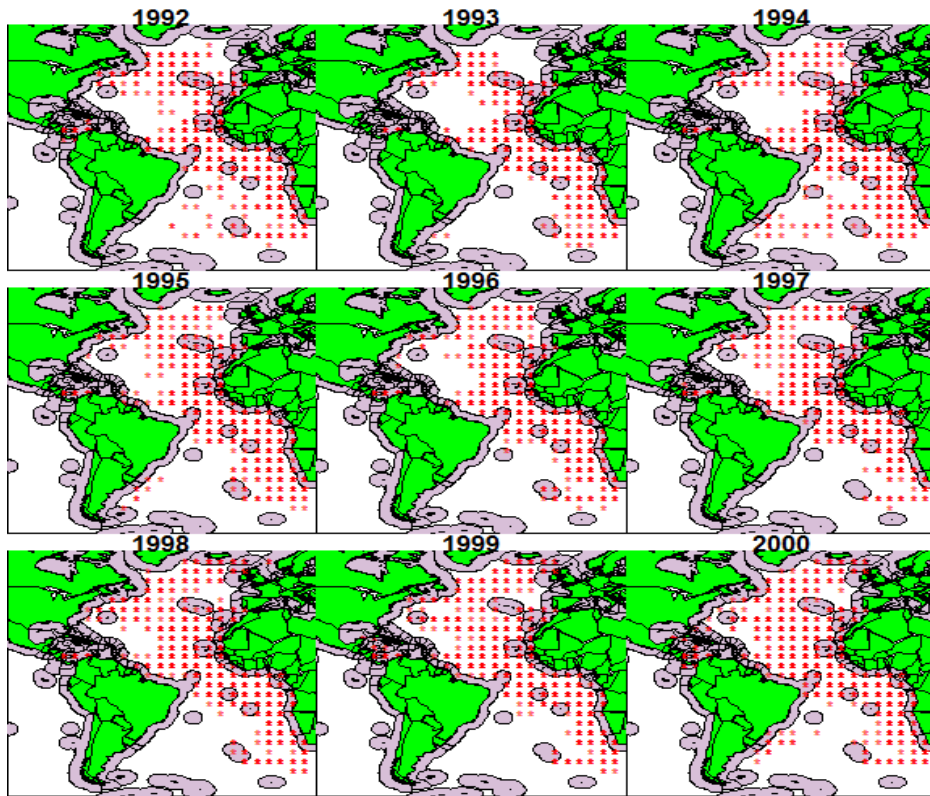


Figure 4. Distribution of Task II data: Japanese long-liners between 1992 and 2000.

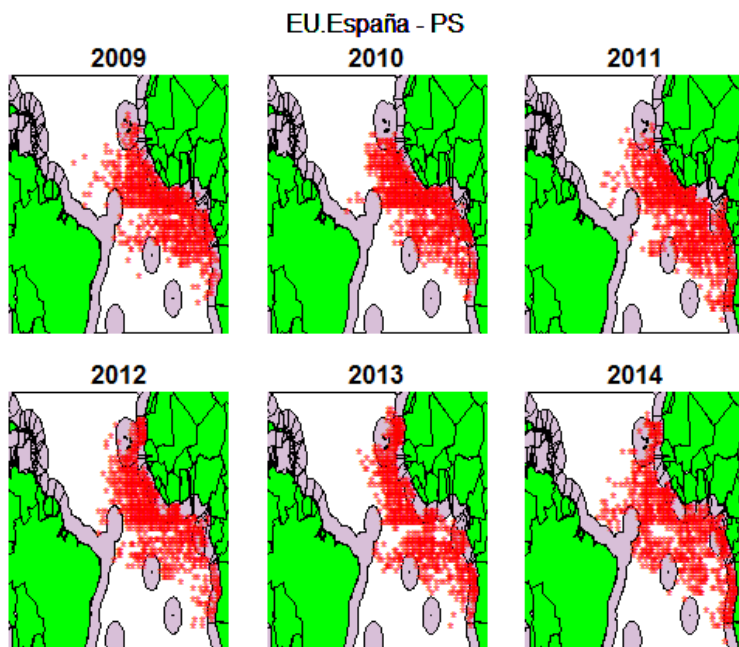


Figure 5. Distribution of Task II data: Spanish purse seiners between 2009 and 2010.

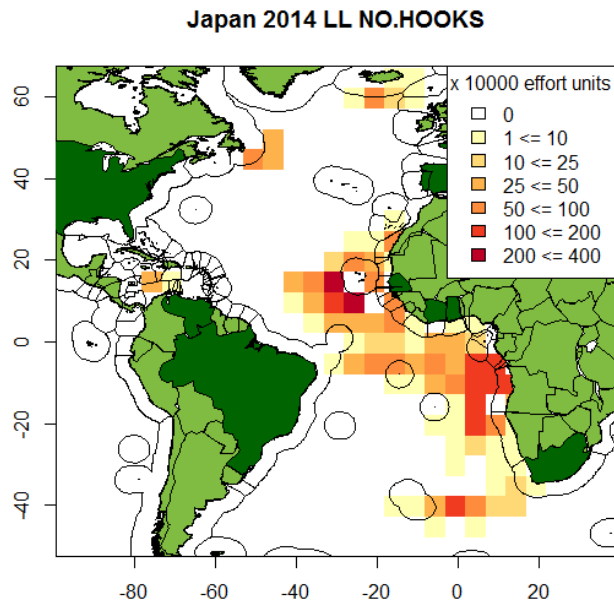


Figure 6. Total number of hooks set by Japanese fleet (Task II only) in 2014.

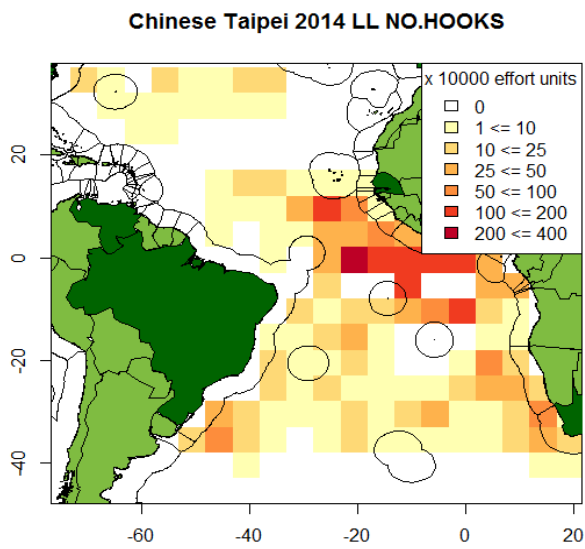


Figure 7. Total number of hooks set by Chinese Taipei (Task II only) in 2014.

EU.España 2014 PS FISH.HOUR

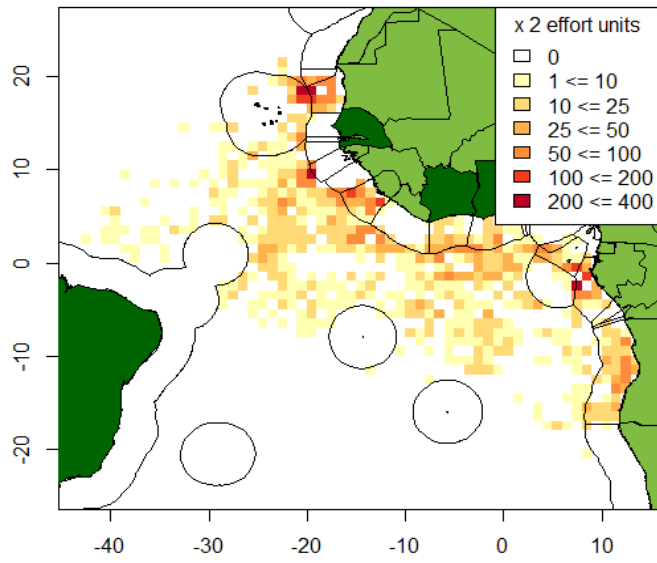


Figure 8. Total fishing hours reported by Spanish purse-seiners (Task II only) in 2014.

Panama 2014 PS FISH.HOUR

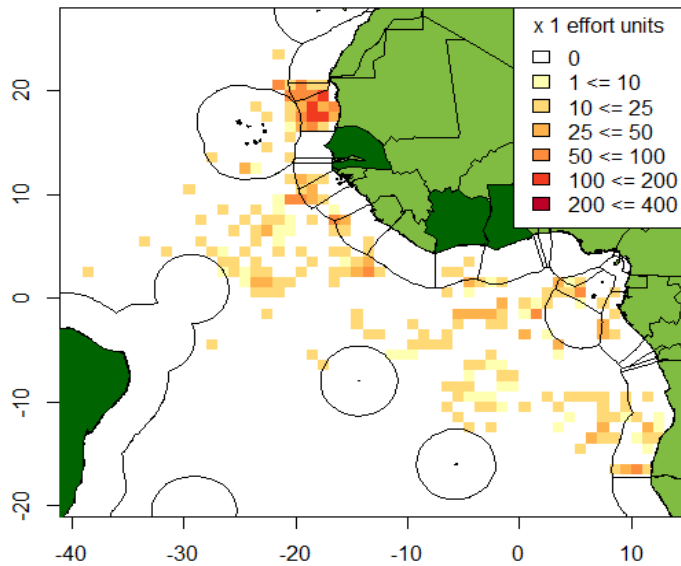


Figure 9. Total fishing hours reported by Panamanian purse-seiners (Task II only) 2014.

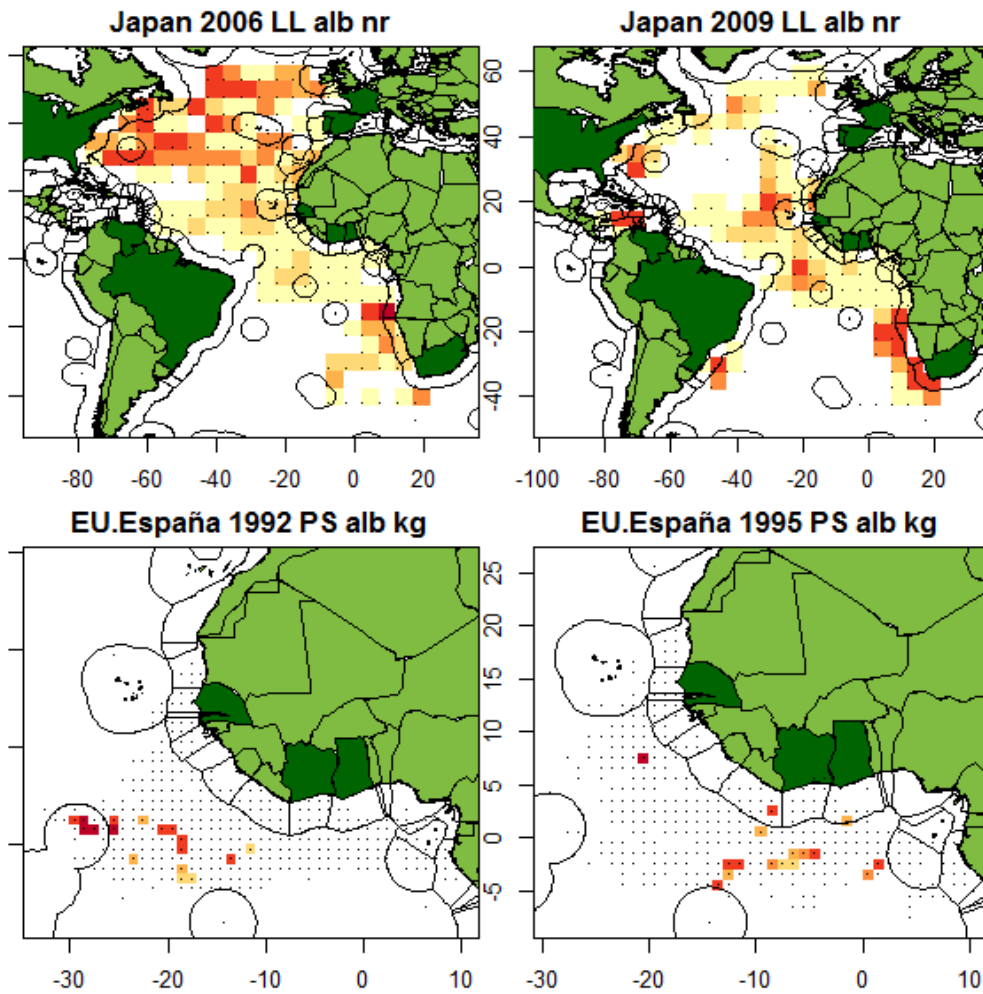


Figure 10. Weight of albacore tuna caught by longline by Japan (2006 and 2009) and purse-seine by Spain (1992 and 1995).

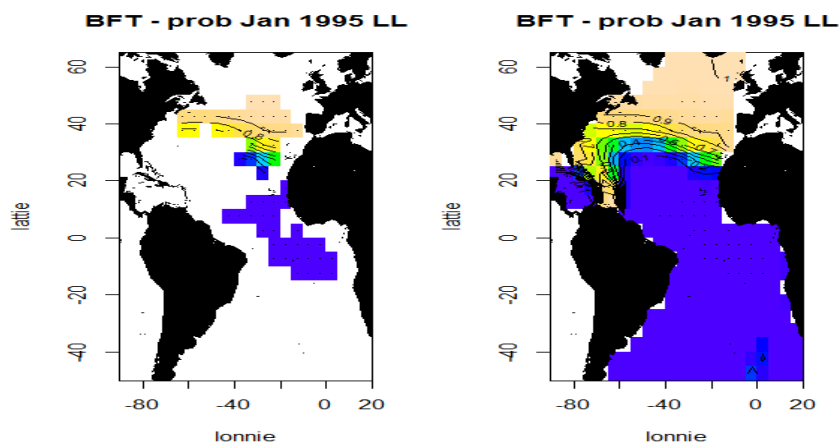


Figure 11. Probability of catching a bluefin tuna in January 1995 estimated using a Generalised Additive Model (Japanese longline fleet).

Task 1 Catch / Task 2 CPUE for Japanese long-liners

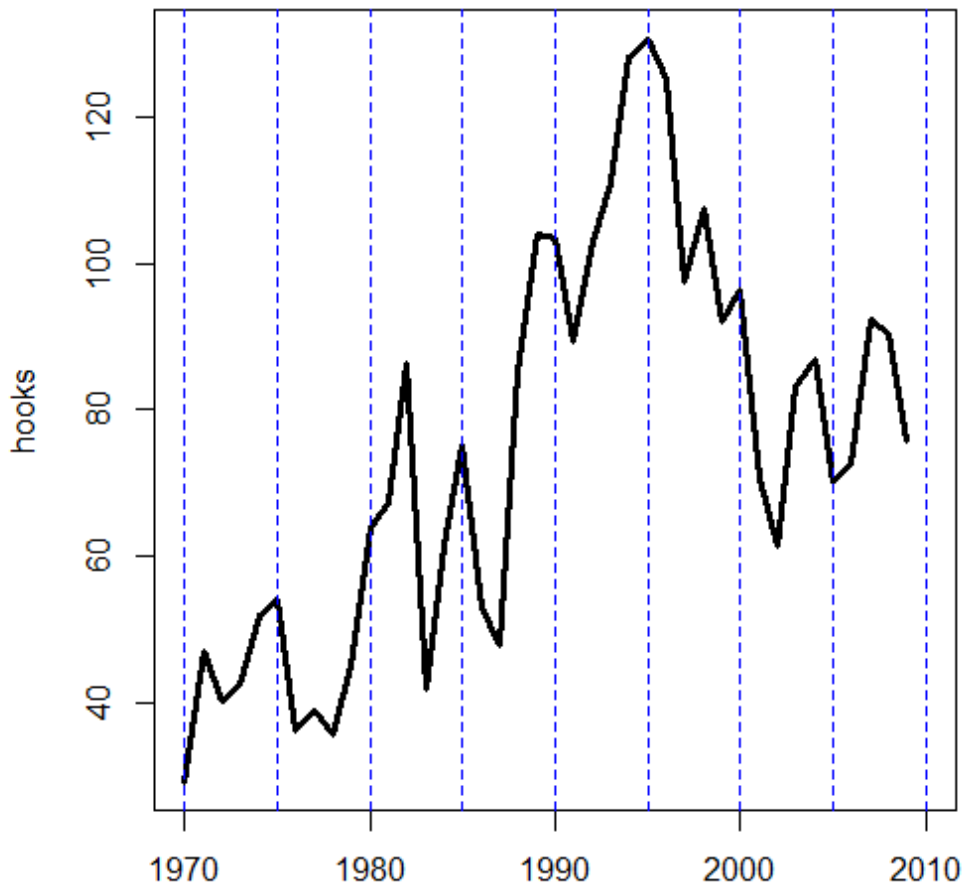


Figure 12. Estimate of total effort (no of hooks) calculated for Japanese long-liners according to $C_{\text{task1}}/U_{\text{task2}}$ where C=catch and U = catch-per-unit effort.