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

Progress Report to SCRS

Doug Beare

14 September 2015

Executive Summary

ICCAT contracted Doug Beare (Globefish Consultancy Services, GCS) to develop a modeling approach to estimate overall Atlantic fishing effort exploiting the spatio-temporal information available in the data. The project began at the end of May 2015. In this document we describe the development of software for analysing the EFFDIS data and how it can be used to 'raise' fishing effort using the Japanese longline fleet as an example. The methods have been presented so far at two International fora: (i) The Sub-Committee on Ecosystems by Dr L. Kell in June 2015; (ii) and by GCS (Doug Beare) at the Blue Shark Stock Assessment Meeting in Lisbon in July 2015. Both groups approved the methodology overall, and the feedback is reproduced verbatim in Appendices I and II below.

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'. 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 in two main groups termed, 'Task 1' and 'Task 2'. Task 1 data are annual totals for catch (eg. tons bluefin tuna caught in 1999 by Japan) by gear in the various relevant 'regions' (Atlantic & Mediterranean) and are believed to be totally comprehensive. Task 2 data, on the other hand, are much more detailed, available at greater spatial (e.g. 5^ox5^o 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, potentilly be made by 'raising' the Task 2 estimates by those from Task 1. The EFFDIS database thus represents a rich and valuable source of information on fishing activity in the Atlantic and Mediterranean since 1950. It has the potential to reveal both seasonal and long-term changes in the distributions of the fisheries, and their target species in addition to exposing the vulnerability of various by-catch taxa such as turtles and seabirds.

Here we describe the development of a statistical modeling approach to estimating overall Atlantic long-line (LL), and purse seine (PS) effort by time-area strata for the EFFDIS database which is critical, especially with regard to by-catch evaluations. The software developed is written in R linked by SQL to a PostGreSQL database. Its use 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.

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 has also been installed on this machine 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 during the project are being backed up on GitHub (https://github.com/bearedo/effdis). Furthermore all reports and presentations, including this one, are being done with Rmarkdown, linked to GitHub which will facilitate straightforward future modification and updating.

The *effdis-tuna-cc1* server also hosts an online geographic information system being trialled 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 below in Figure 1.

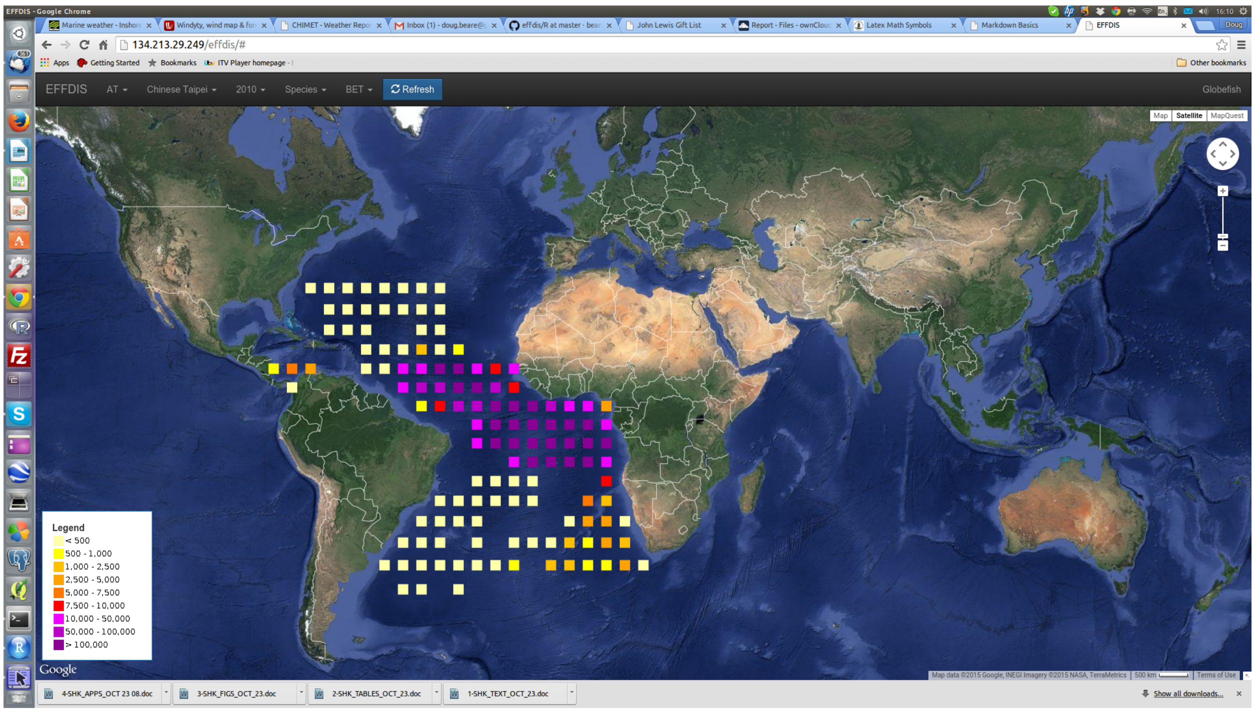

Figure 1. *Screenshot of EFFDIS geo-database - Bigeye catches by Chinese Taipei in 2010 reported to ICCAT as Task II*

EFFDIS effort estimation

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)
```

```r
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-scripts which is in the process of being loaded into an R-package (*effdisR*) which will soon render this step redundant.

Exploratory data analysis

For any data analysis some initial exploration is essential. When data are distributed non-randomly in space and time, for example, spurious results can easily be obtained. Our software for analysing EFFDIS includes a suite of tools for plotting and examining the EFFDIS data which are described below.

Data screening

Once the RODBC library is installed and the /etc/odbc.ini file modified to provide the necessary parameters for connection to the effdis database (see above) the data can be accessed via R according to the following code from the RODBC R package:

```r
chan <- odbcConnect("effdis-local", case="postgresql",
                    believeNRows=FALSE)
```

In Table 1 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. The table illustrates the sort of problems that exist with these data. Chinese Taipei, for example, has supplied 'NO. HOOKS' (37795) only while Belize has supplied both 'NO. HOOKS' (629) and 'D.FISH' (485). Similarly in Table 2 we summarise the different types of fishing effort that have been recorded for purse-seiners. Spain, for example, which has an important Atlantic purse-seining fleet has sent data as, 'D.AT SEA', 'D.FISH', and 'FISH.HOUR'. It is also worth noting that, in many cases, no effort data have been submitted with catches for Task II at all, see '-none-' (Tables 1 & 2).

Table 1. Effort type sampling by flag for longline in EFFDIS Task II database

```r
effort_type_by_flag_ll <- sqlQuery(chan, "SELECT flagname AS Flag,
eff1type AS Effort_type, count(eff1type) as No_records
                 FROM t2ce
                 WHERE region ='AT' AND geargrpcode = 'LL'
                 GROUP BY flagname, eff1type
                 ORDER BY flagname, eff1type, No_records;")
pander(effort_type_by_flag_ll)
```

flag	effort_type	no_records
Belize	D.FISH	485
Belize	NO.HOOKS	629

flag	effort_type	no_records
Brasil	NO.HOOKS	9390
Brasil	-none-	27
China P.R.	NO.HOOKS	1249
China P.R.	-none-	43
Chinese Taipei	NO.HOOKS	37795
Cuba	NO.HOOKS	2298
Cuba	-none-	6
EU.España	NO.HOOKS	19721
EU.España	-none-	3152
EU.Malta	D.FISH	3657
EU.Malta	NO.HOOKS	203
EU.Malta	-none-	432
EU.Portugal	NO.HOOKS	1693
EU.Portugal	-none-	2758
EU.Portugal	NO.TRIPS	13
Japan	NO.HOOKS	34770
Korea Rep.	NO.HOOKS	9390
Maroc	NO.HOOKS	70
Maroc	-none-	12
Mexico	NO.HOOKS	601
Mexico	NO.SETS	12
Mexico	SUC.SETS	61
Namibia	NO.HOOKS	861
Namibia	-none-	6
Other	D.AT SEA	51
Other	D.FISH	295
Other	NO.HOOKS	8684
Other	-none-	1419
Other	NO.SETS	286
Other	NO.TRIPS	65
Other	SUC.D.FI	4
Panama	NO.HOOKS	547
Philippines	NO.HOOKS	497
Philippines	-none-	804
South Africa	D.AT SEA	10
South Africa	D.FISH	13
South Africa	NO.BOATS	3
South Africa	NO.HOOKS	1510
South Africa	-none-	4
St. Vincent and Grenadines	NO.HOOKS	1112
St. Vincent and Grenadines	-none-	1
Trinidad and Tobago	NO.BOATS	1
Trinidad and Tobago	NO.HOOKS	84
Trinidad and Tobago	-none-	1
Trinidad and Tobago	NO.TRIPS	12
Uruguay	NO.HOOKS	1919
Uruguay	-none-	17
U.S.A.	NO.HOOKS	78698
U.S.A.	-none-	2
U.S.S.R.	D.FISH	61
U.S.S.R.	NO.HOOKS	166
U.S.S.R.	-none-	26

flag	effort_type	no_records
Vanuatu	NO.HOOKS	19083
Venezuela	NO.HOOKS	18753

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

```
effort_type_by_flag_ps <- sqlQuery(chan, "SELECT flagname AS Flag,
eff1type AS Effort_type, count(eff1type) as No_records
                 FROM t2ce
                 WHERE region ='AT' AND geargrpcode = 'PS'
                 GROUP BY flagname, eff1type
                 ORDER BY flagname, eff1type, No_records;")
pander(effort_type_by_flag_ps)
```

flag	effort_type	no_records
Belize	D.FISH	57
Brasil	D.FISH	99
Brasil	FISH.HOUR	17
Brasil	NO.SETS	8
Brasil	NO.TRIPS	58
EU.España	D.AT SEA	54
EU.España	D.FISH	22662
EU.España	FISH.HOUR	33878
EU.España	-none-	61
EU.Malta	D.FISH	1
EU.Portugal	D.FISH	25
EU.Portugal	-none-	318
EU.Portugal	NO.TRIPS	30
EU.Portugal	SUC.SETS	8
Japan	D.FISH	1498
Japan	NO.SETS	1443
Maroc	NO.BOATS	12
Maroc	-none-	6
Other	D.AT SEA	63
Other	D.FISH	35229
Other	FISH.HOUR	65448
Other	HOURS.SEA	12
Other	NO.BOATS	90
Other	-none-	213
Other	NO.SETS	1651
Other	NO.TRIPS	7
Other	SUC.D.FI	92
Panama	D.AT SEA	117
Panama	FISH.HOUR	3556
South Africa	D.AT SEA	28
South Africa	D.FISH	11
South Africa	NO.BOATS	5
South Africa	NO.SETS	10
U.S.A.	D.AT SEA	49
U.S.A.	D.FISH	7057
U.S.A.	-none-	7

flag	effort_type	no_records
U.S.A.	SUC.D.FI	3
U.S.S.R.	D.FISH	165
U.S.S.R.	-none-	8
Venezuela	D.FISH	6800
Venezuela	NO.BOATS	4

After examining Tables 1 and we made the decision to examine records with 'NO. HOOKS' only for long-liners 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 re-inforces the point that data need to be submitted using the same variables.

Catchunit is another important variable in the Task II data, denoting whether the catch was recorded in terms of total numbers (nr) or total weight (kg). Table 3 illustrates that there are many records (121225) in the Task II data with catchunit = '–' and these rows were also, perforce, removed from subsequent analyses.

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

```
catchunit <- sqlQuery(chan, "SELECT catchunit AS catchunit, count(catchunit) as No_records
                 FROM t2ce
                 WHERE region ='AT'
                 GROUP BY catchunit
                 ORDER BY catchunit, No_records;")
pander(catchunit)
```

catchunit	no_records
–	121225
kg	258997
nr	171613

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 *yr.month.coverage.task2.r* are plotted in Figure 2 for longline between 1960 and 2010 for Japan, Chinese Taipei, Brasil, 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 (Fig. 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
WHERE region ='AT' AND timeperiodid < 13 AND eff1type='NO.HOOKS' AND geargrpcode = 'LL'
AND flagname IN ('Japan', 'Chinese Taipei','Brasil', 'U.S.A.','China P.R.')
AND catchunit != '--' ;")
```

```r
par(mfrow=c(2,2),mar=c(2,2,2,2),oma=c(1,1,1,1))
yr.month.coverage.task2.r(tdata=ll,which.gear='LL',
              start.year=1960,end.year=2010,which.flag='Japan')
yr.month.coverage.task2.r(tdata=ll,which.gear='LL',
              start.year=1960,end.year=2010,which.flag='Chinese Taipei')
yr.month.coverage.task2.r(tdata=ll,which.gear='LL',
              start.year=1960,end.year=2010,which.flag='Brasil')
yr.month.coverage.task2.r(tdata=ll,which.gear='LL',
              start.year=1960,end.year=2010,which.flag='U.S.A.')
```

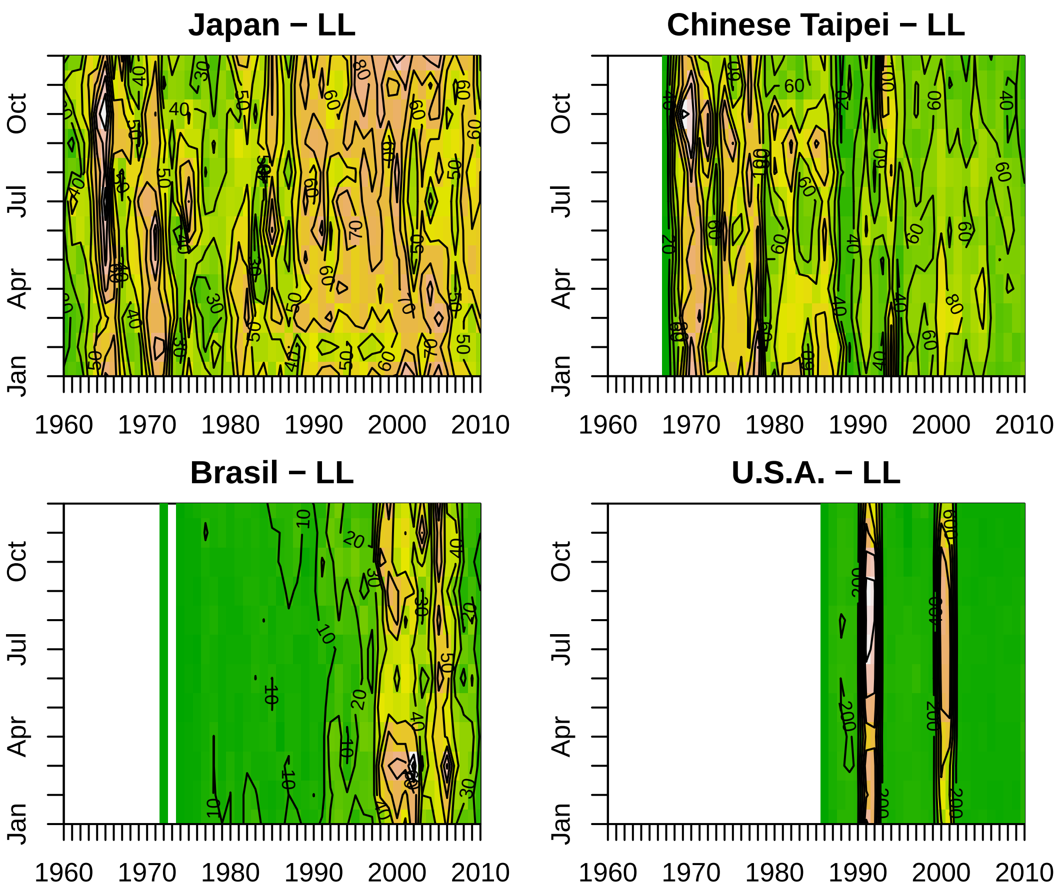

Figure 2. *Temporal (by year and month) sampling distribution of long-liners in Task II database by Japan, Chinese Taipei, Brazil, and U.S.A.*

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 for longliners for two arbitarily selected flags and time-periods in Figures 3 and 4.

```r
par(mfrow=c(4,4),mar=c(0,0,1,0))
spatial.coverage.by.year.task2.r(tdata=ll,
        start.year=1975,end.year=1990,which.gear='LL',which.flag='Brasil')
```

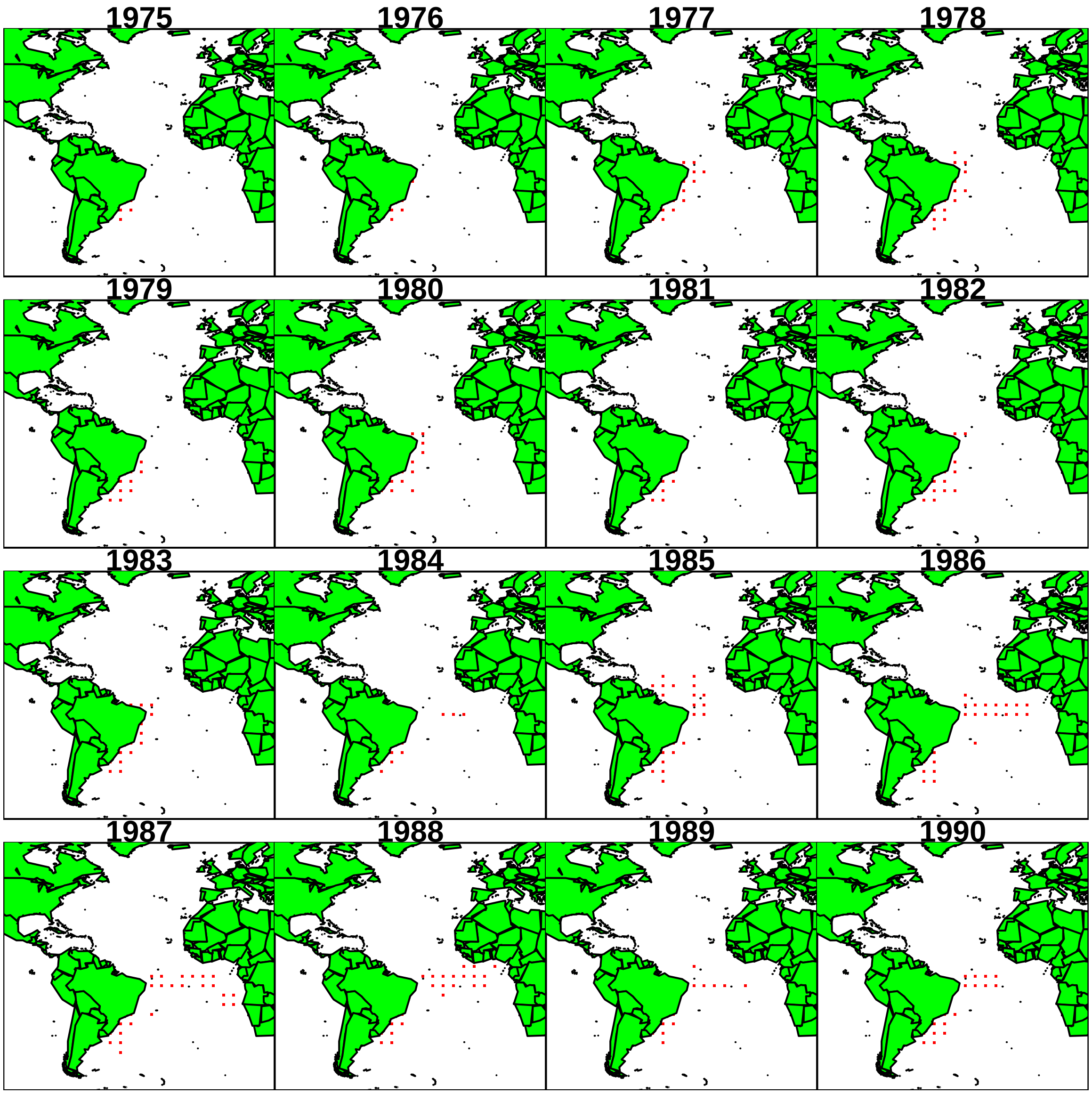

Figure 3. *Spatial sampling by Brazilian long-liners between 1975 and 1990*

The extent of Brazilian longlining activity has, for example, spread out from the coast of South America between 1975 and 1990 (Fig. 4). In contrast fleets from both Japan and Chinese Taipei cover substantial areas of the Atlantic Ocean (Figs. 5 and 6).

```
par(mfrow=c(4,4),mar=c(0,0,1,0))
spatial.coverage.by.year.task2.r(tdata=ll,
        start.year=1985,end.year=2000,which.gear='LL',which.flag='Japan')
```

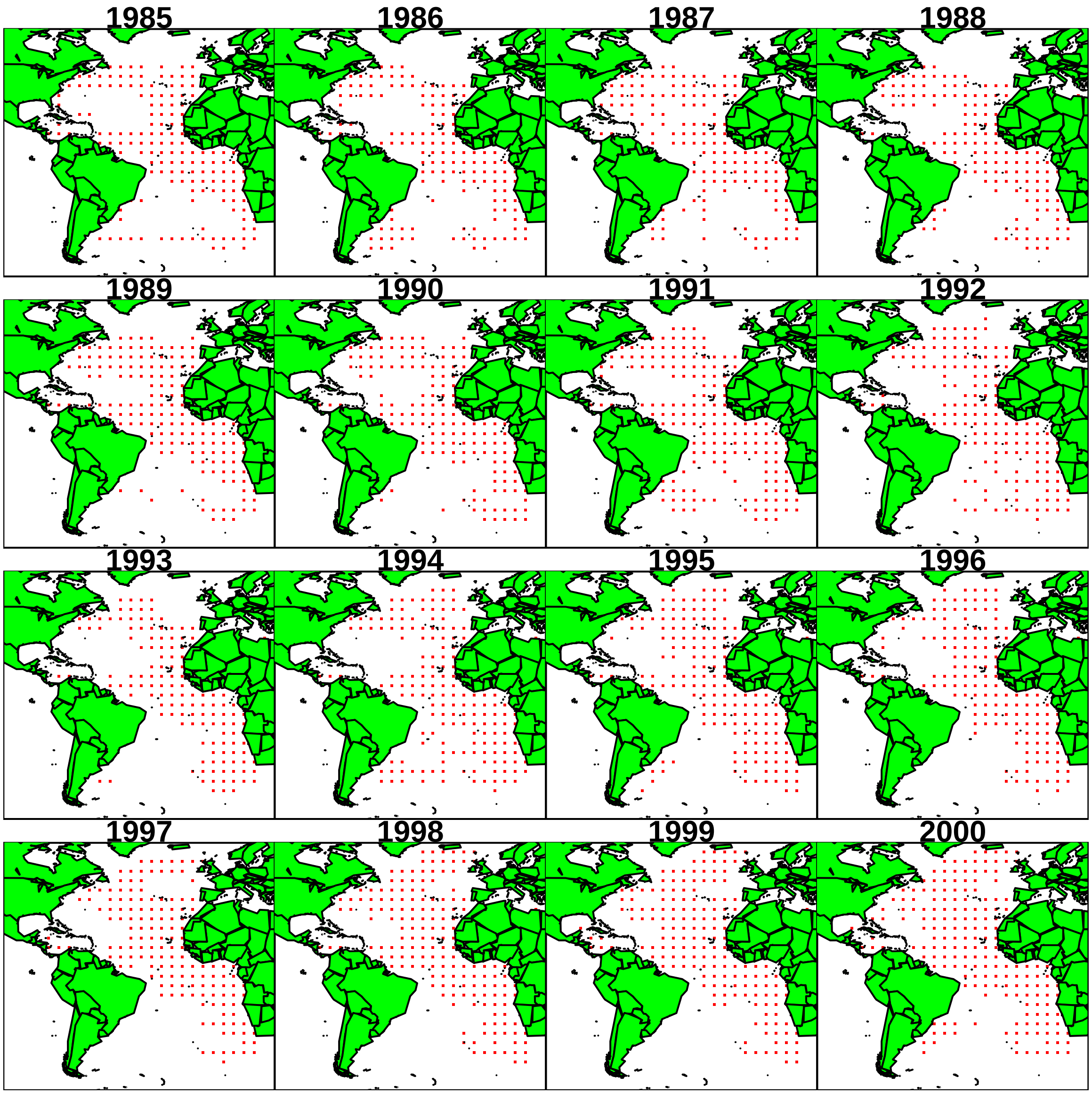

Figure 4. *Spatial sampling by Japanese long-liners between 1985 and 2000*

```
ps <- sqlQuery(chan,"SELECT yearc AS year, trend, timeperiodid AS month,
flagname, region, geargrpcode,longitude,latitude, catchunit, dsettype, eff1, eff1type
FROM t2ce
WHERE 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.r(tdata=ps,start.year=2005,end.year=2010,
                                 which.gear='PS',which.flag='EU.España')
```

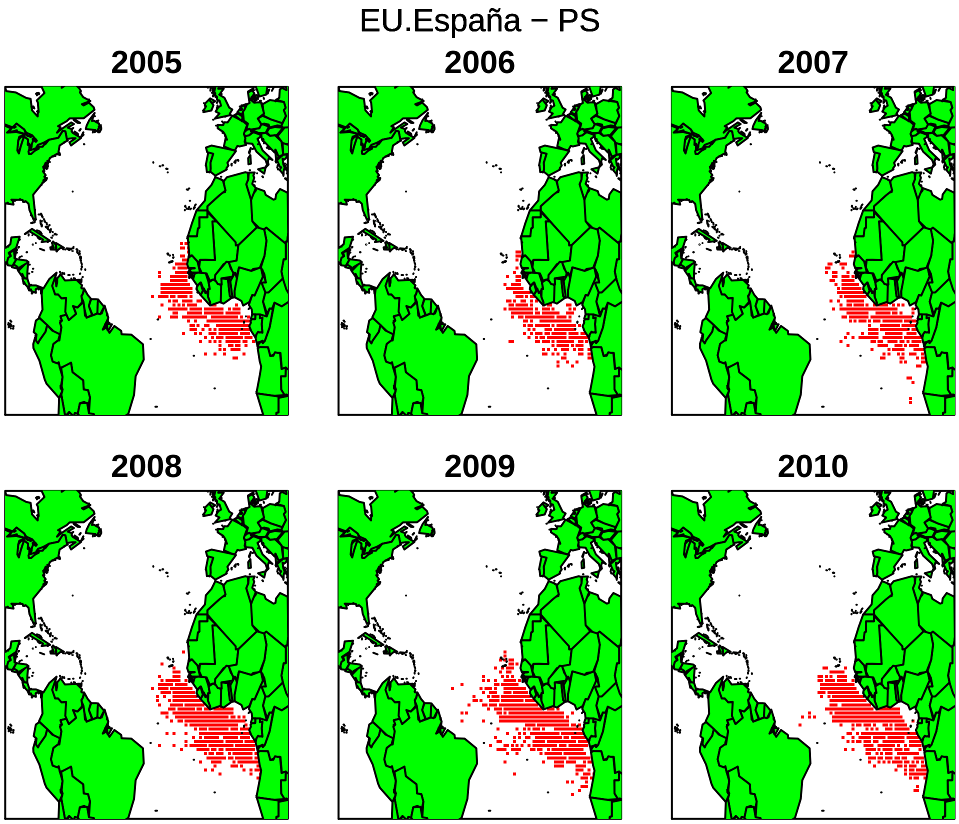

Figure 5. *Spatial sampling by Spanish purse-seiners between 1995 and 2010*

Most of the purse-seining data have been supplied at a higher spatial resolution (1x1) than is available for the long-liners. This is illustrated in Figure 5 which shows the distribution of Spanish purse-seining activity off the West Coast of Africa between 2005 and 2010.

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

Given that we can now determine the timing and location of fishing activities (Figs. 2-6) we now need to know its intensity, ie. 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 2006, for example, longlining effort by Japan focused on the North and Eastern Atlantic (Fig. 6). Chinese Taipei flagged long-liners, on the other hand also worked in the North Atlantic but their effort appears greater in the South Atlantic (Fig. 7). Purse-seining data are submitted to ICCAT at a higher spatial resolution (1x1 grid) and examples for Spain and Panama in 2009 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 (Figs 8 & 9).

```
three.d.effort.by.year.r(tdata=ll,what.year=2006,
                         what.flag='Japan',scaling.f=10000,effort.type='NO.HOOKS')
```

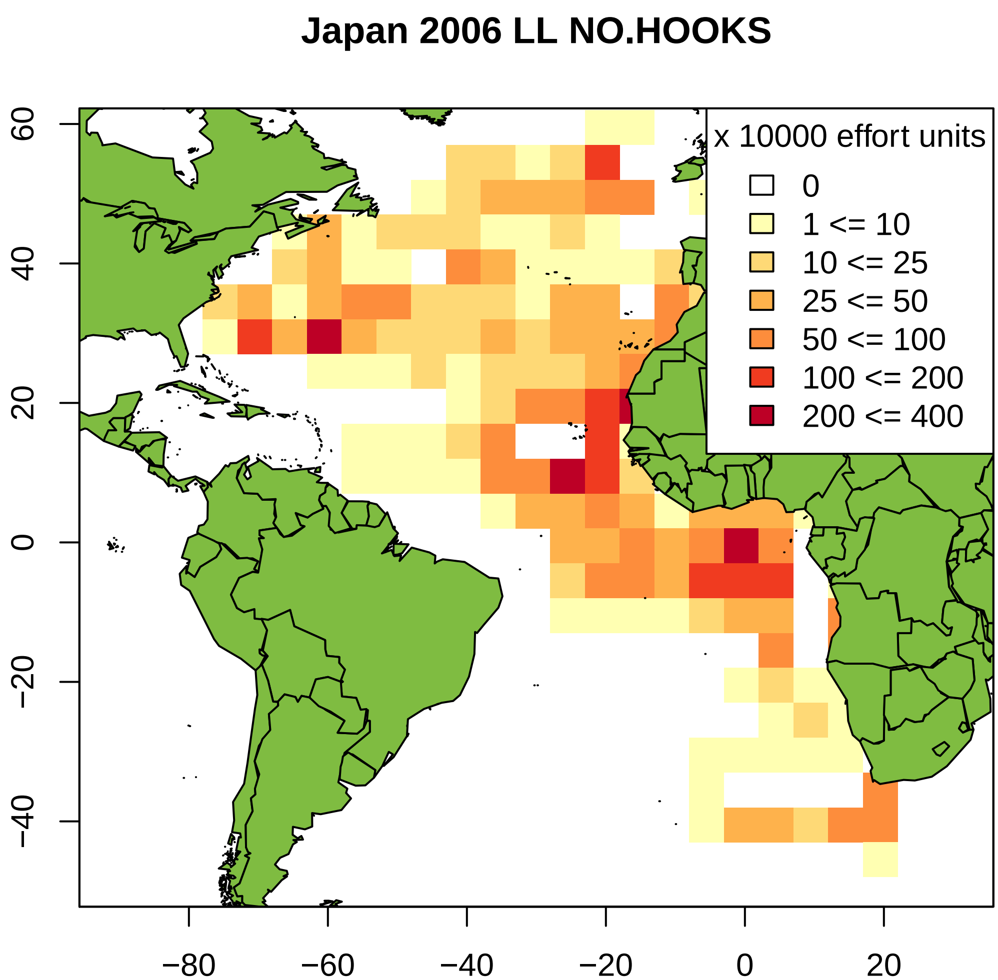

Figure 6. *Total number of hooks set by Japanese fleet (Task II only) in 2006*

```
three.d.effort.by.year.r(tdata=ll,what.year=2006,
                         what.flag="Chinese Taipei",scaling.f=10000,gridx=5,gridy=5,
                         effort.type='NO.HOOKS',what.gear='LL')
```

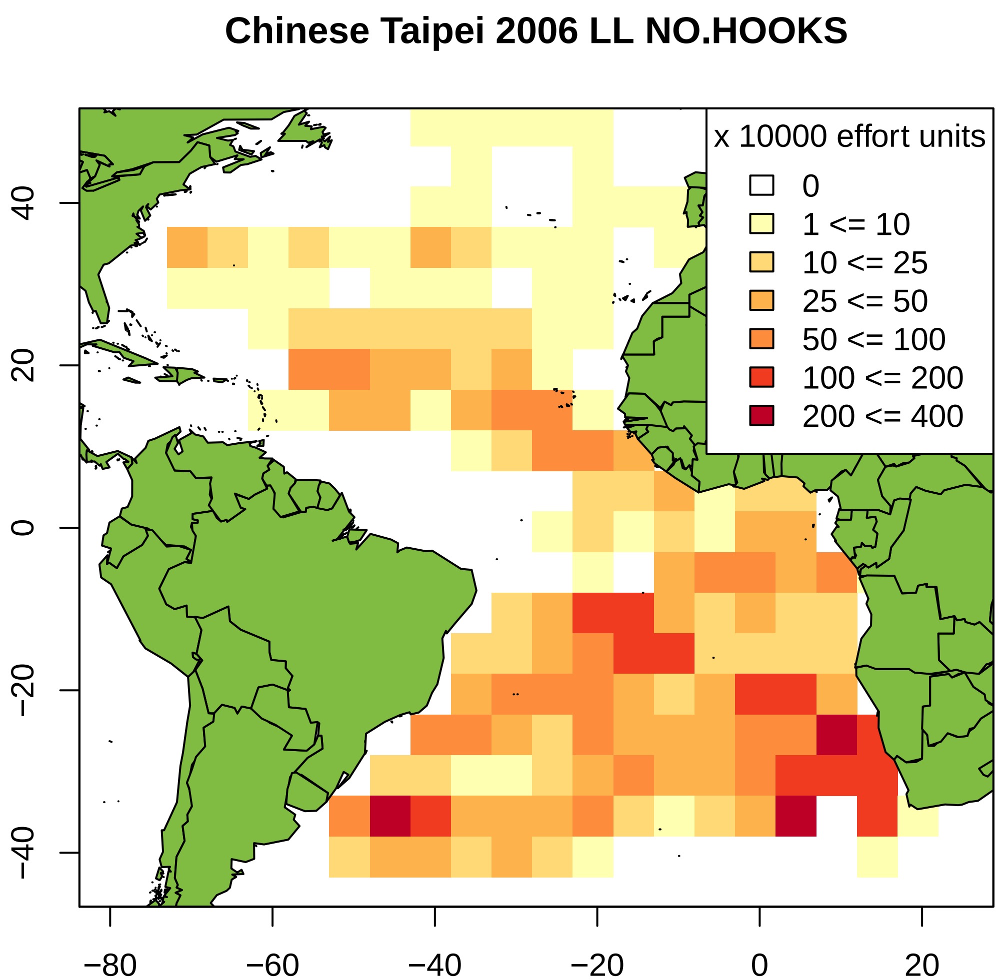

Figure 7. *Total number of hooks set by Chinese Taipei (Task II only) in 2006*

```
three.d.effort.by.year.r(tdata=ps,what.year=2009,
                         what.flag='EU.España',gridx=1,gridy=1,
                         scaling.f=2,effort.type='FISH.HOUR',what.gear='PS')
```

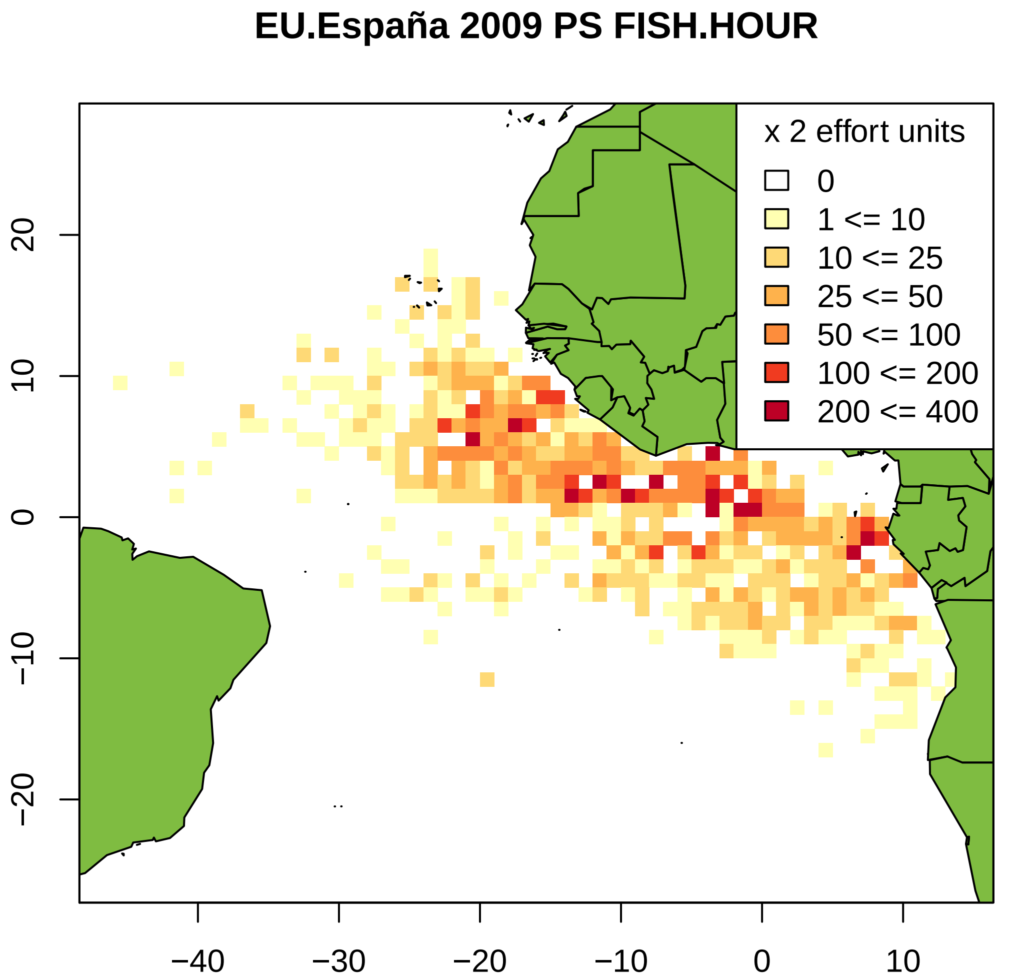

Figure 8. *Total fishing hours reported by Spanish purse-seiners (Task II only) in 2009*

```
three.d.effort.by.year.r(tdata=ps,what.year=2009,
                         what.flag='Panama',
                         gridx=1,gridy=1,scaling.f=1,effort.type='FISH.HOUR',what.gear='PS')
```

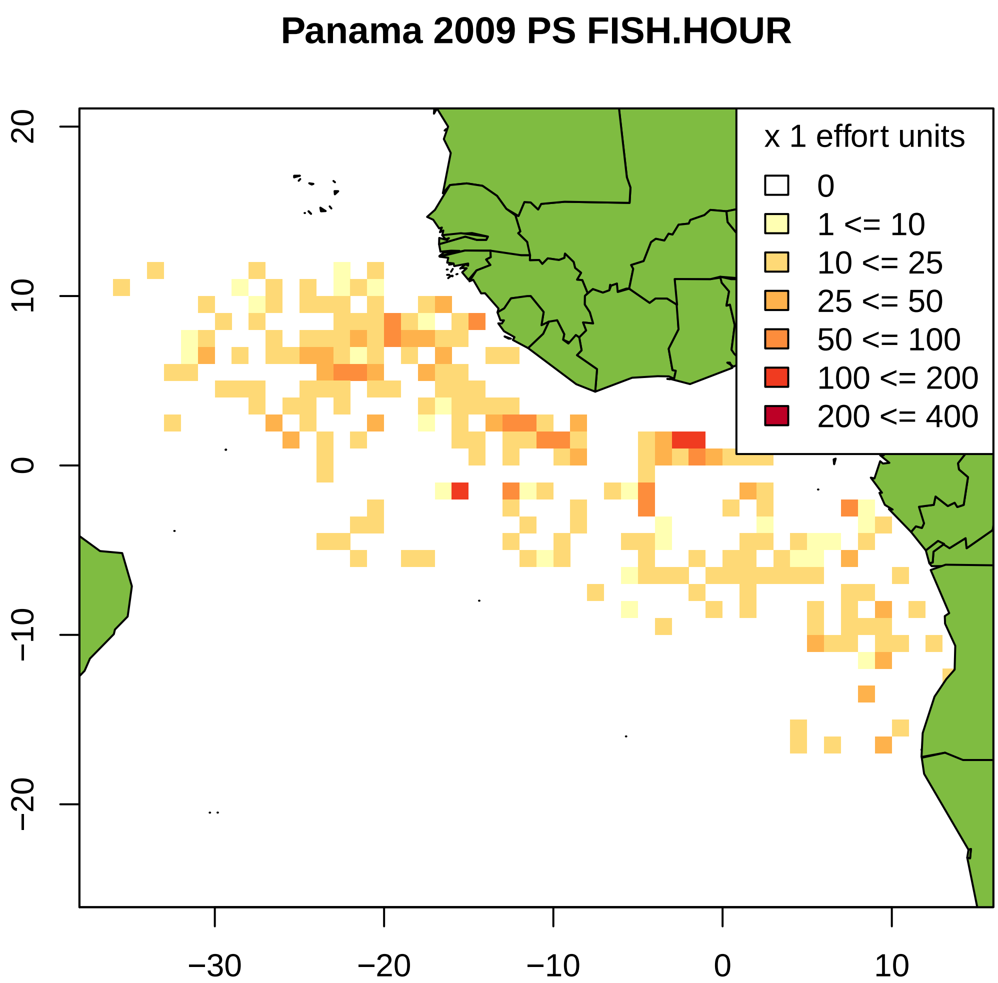

Figure 9. *Total fishing hours reported by Panamanian purse-seiners (Task II only) 2009*

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

With *circa* 27 flags, multiple species, 12 months, 61 years, and different effort submissions (e.g. days at sea, number of hooks set) there is a large number of possible combinations for examining the catches. Examples are illustrated in Figure 10 using the R function, *three.d.catch.by.year.r* linked to the PostgreSQL database via an SQL script (see below). In this example we extracted data for albacore tuna caught by long-line for Japan, and by purse-seine for Spain. [Note that observations were ignored where the catchunit is unknown ('–'). Numbers or kilograms caught can be selected depending on availability].

```
alb <- sqlQuery(chan,"SELECT yearc AS year, trend,
timeperiodid AS month, flagname, region, geargrpcode,longitude,latitude,
catchunit, dsettype, eff1, eff1type, alb as measured_catch
FROM t2ce
WHERE region ='AT' AND timeperiodid < 13 AND eff1type IN ('NO.HOOKS','FISH.HOUR') AND
geargrpcode 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.r(tdata=alb,what.year=2006,
                        what.gear='LL',what.species='alb',what.flag='Japan',scaling.f=10,catchunit='nr')
```

```
three.d.catch.by.year.r(tdata=alb,what.year=2009,
                        what.gear='LL',what.species='alb',what.flag='Japan',scaling.f=10,catchunit='nr')
three.d.catch.by.year.r(tdata=alb,what.year=1992,
                        what.gear='PS',what.species='alb',gridx=1,gridy=1,what.flag='EU.España',scaling
three.d.catch.by.year.r(tdata=alb,what.year=1995,
                        what.gear='PS',what.species='alb',gridx=1,gridy=1,what.flag='EU.España',scaling
```

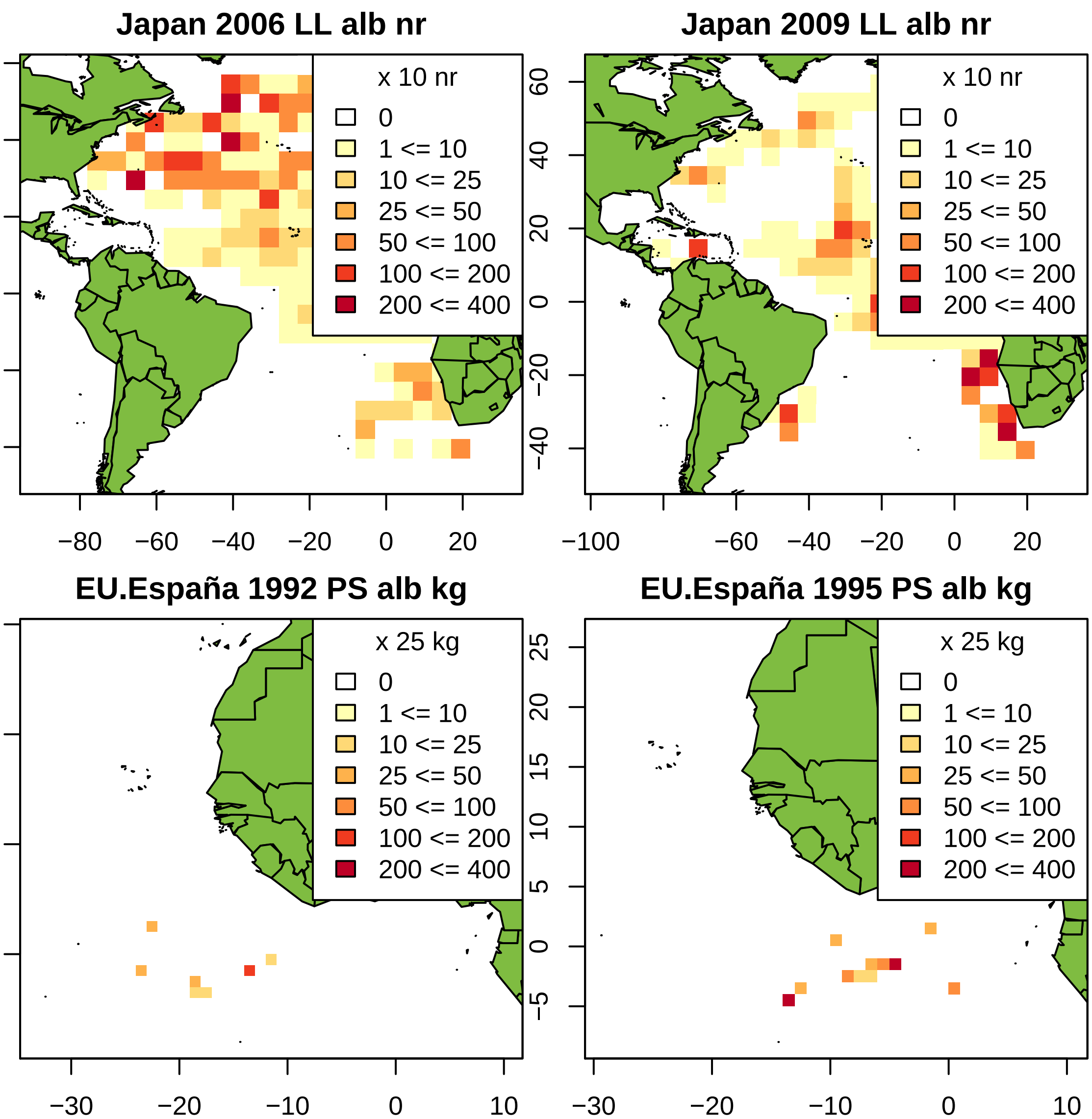

Figure 10. *Weight of albacore tuna caught by longline by Japan (2006 & 2009) and purse-seine by Spain (1992 & 1995)*

Worked example (Japanese longline)

Step 1

Get Task II data from the database for each dsettype using *get.effdis.t2.data.r* and then combine them using rbind. The which.dsn argument here is accessing a locally installed copy of the postgres database and this should be edited when using the ICCAT cloud server.

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

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

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.r(input=long_line)
```

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 subsequent regression modeling.

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

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 betweem them are highly linear (see Interim Report) 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 selects 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%). This part of the procedure/model is included in the error variance estimation.

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

Table 3. Linear model summarising the relationship between weights and numbers for countries that sent both to ICCAT for the Task II database

```
panderOptions("digits",1)
pander(bm)
```

Table 4: Fitting linear model: lkg ~ lnr + trend + species

	Estimate	Std. Error	t value	Pr(>|t|)
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

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. The entire process takes too long to inclused as part of a markdown file. To get a global estimate of effort, however, you would, of course, 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 generalised 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 prevalences of zeros. 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. $C_{xytm} = s(x, y) + s(t) + s(m) + \epsilon$

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

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

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

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

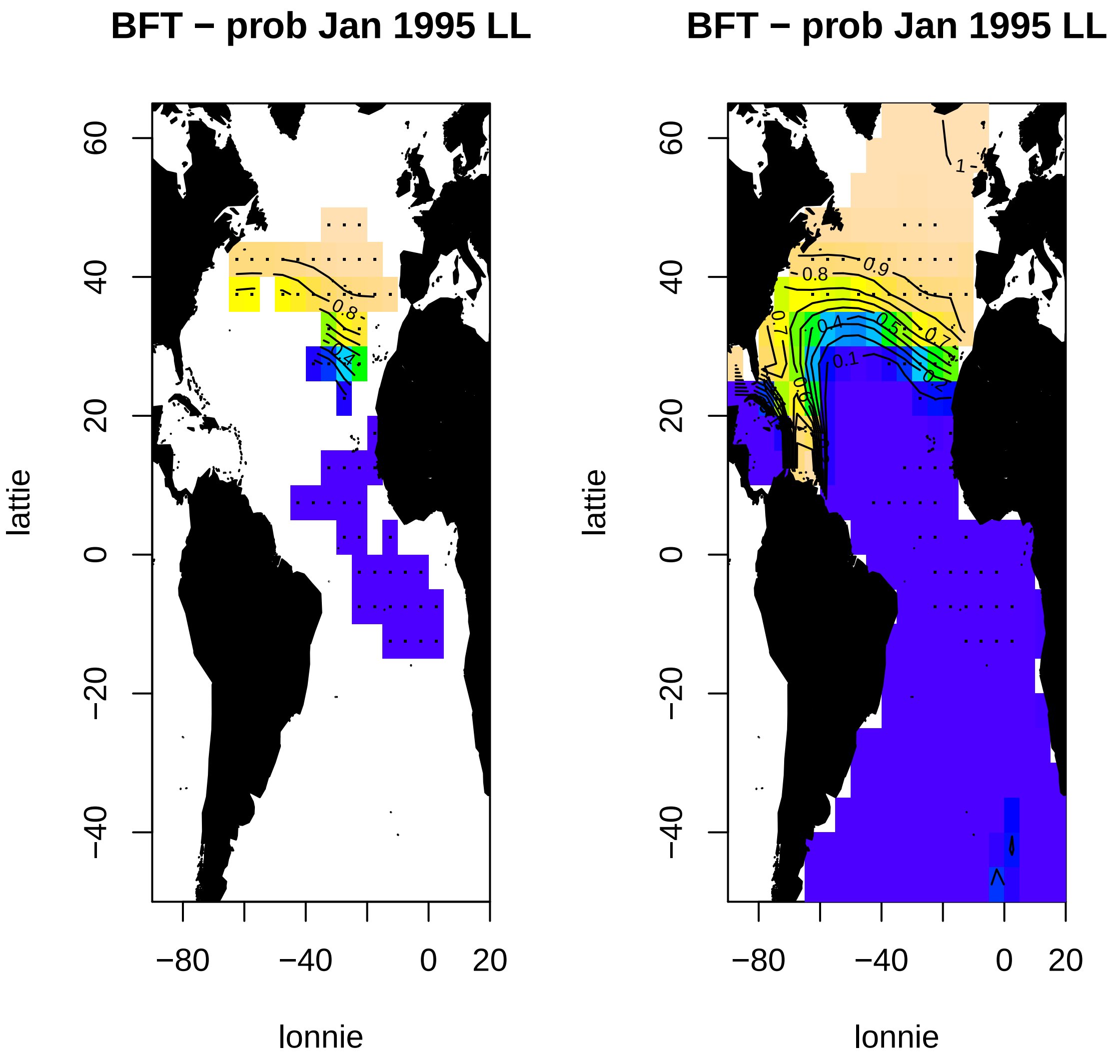

Figure 11. *Probability of catching a bluefin tuna (Japanese longline fleet)*

Figure 11 shows predictions for bluefin tuna from the binary (Bernouilli) 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.

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 1 are annual catch totals which are thought to be comprehensive.

```
long_line.t1 <- get.effdis.t1.data.r(which.dsn='effdis-local',
                                     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)
```

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)
```

Step 10

Block out places and times for which data were never collected. This is optional and if all the data are used (see Figure 11 above) the results are very similar.

```
model.data$catch[big$observation == FALSE] <- NA
model.data$prob[big$observation == FALSE] <- NA
model.data$measured_catch[big$observation == FALSE] <- NA
model.data$eff[big$observation == FALSE]  <- NA
```

Step 11

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

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

Step 12

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)
```

Step 13

Merge the Task 1 and modeled Task 2 totals, calculate a global, modeled Task II CPUE and raise by Task 1 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
```

Step 14

Plot the raised effort as a function of year.

```
effdis <- read.table('/home/doug/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 1 Catch / Task 2 CPUE for Japanese long-liners')
```

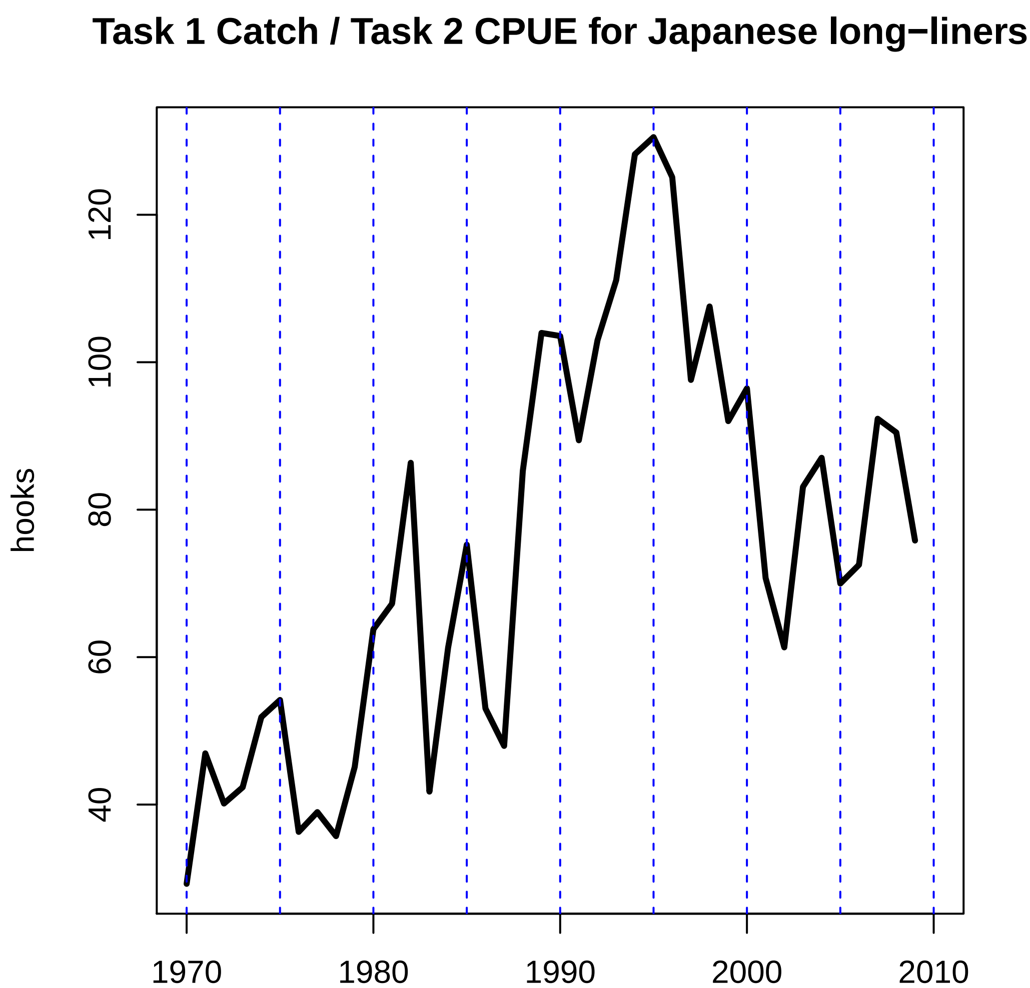

Figure 12. *Estimate of total effort (no of hooks) calculated for Japanese long-liners according to* C_{task1}/U_{task2} *where C=catch and U = catch-per-unit effort.*

Conclusions

The use of the code for raising Task 2 EFFDIS effort estimates by Task 1 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 actually 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.

Appendices

Appendix I. Recommendations of the Sub-Committee on Ecosystems (SCRS/P/2015/026)

In the past, the Sub-Committee on Ecosystems and the Working Group on Stock Assessment Methods have both made a number of recommendations for updating and improving EFFDIS, which will be incorporated in the new estimates. The Sub-Committee agreed that the EFFDIS data are complex, and difficult to analyse. GCS has been working to understand the data and identify issues related to non-random, non-representative sampling. All the analyses are available on a github repository http://iccat-stats.github.io/.

It was also clarified that the EFFDIS data are reliant on Task II catch and effort information, and it is known that there are errors in these data. The secretariat said that data screening would take place to eliminate problems such as effort duplication. This revision should reduce the amount of problematic data used for the EFFDIS estimation. The secretariat and GCS are also working to harmonise the very heterogeneous catch and effort data in order to make it comparable, and facilitate its use in the development of EFFDIS.

It was also discussed that the EFFDIS estimations rely on species composition information (for key target species). This is problematic when applying to by-catch species since the composition is biased towards target species and there are in-consistent historical trends in this bias. GCS is hoping to address this issue using cross-validation although non-random bias remains a complicated problem. The Sub-Committee also requested the addition of southern Bluefin tuna catch information into the estimation of EFFDIS.

Due to the fact that by-catch information is usually recorded on a set by set basis for purse seine, this unit of effort would be the most appropriate metric in the EFFDIS dataset for this gear. It is not, however, the most frequently reported unit of effort for purse seines, and thus GCS will have to evaluate the practicality of using this metric.

The Sub-committee also discussed the proposal by the 2013 Working group on stock assessment methods (WGSAM) regarding the additional gears that should be included in the EFFDIS estimation. Previously, it was requested that estimations be done for both purse seine and baitboat fleets. It was pointed out, however, that EFFDIS is only used to assess the fishing impacts of ICCAT fleets on by-catch species, and since by-catch in baitboat fisheries is minimal, there is no point in conducting this exercise for that gear. It was thus agreed that GCS should focus on the more important longline and purse seine estimations under the current contract.

The Sub-Committee also suggested that, when examining ‘fleet profiles’ for the purse seine fleet, instead of just separating the effort into ‘FAD’ or ‘Free school fishing’, an additional category, namely the Ghanaian purse seine/baitboat co-operative fishery should be considered. This is due to the different catchability apparent for this fleet due to the close co-operation in fishing operations between these two gear types and the sharing of catch, which could bias effort estimates. It was suggested that Ghanaian scientists be consulted to fully explore this unique sector.

Appendix II. Feedback and recommendations from The 2015 ICCAT Blue Shark Stock Assessment Session, Lisbon, 27-31 July 2015.

Presentation SCRS/P/2015/030 given at the Lisbon meeting detailed a statistical modeling framework approach, provided by an external contractor (GCS), to estimating overall Atlantic fishing effort on tuna and tuna-like species which is being developed using ‘Task I’ nominal catch and ’Task II’catch and effort data from the EFFDIS database. Initial findings are promising but problems of confounding (non-random sampling in both space and time) are substantial, and proving difficult to ignore. The purpose of the presentation was to describe the models, the outputs and the estimates of fishing effort made for the Atlantic thus far.

Feedback from the Group was positive and the overall modeling strategy/framework was approved. Some members of the group were, however, concerned about the treatment of the ‘fleet’ or ‘flag’. Aggregating the data by location and temporal variables could be too much of an oversimplification. Some fleets, for example, set surface longlines, others set them in mid or deepwater. Hook sizes, baits and targeting strategies all vary, and have varied substantially over time. Given that the data are particularly patchy prior to the 1960s it was suggested that the modeling framework could concentrate on more recent years only. This would substantially reduce the burden on computation. Also the contractor was asked to include data on artisanal fisheries and to consider ways to include information on fleet/flag combinations that report only Task 1 data. Data catalogues, prepared by the Secretariat are freely available for this.

The method being developed is modular in nature so it could easily be altered to include information from fleet or flag. Polygons could be set up around the data for each fleet and the same regression model (i.e. catch fitted to covariates of location and time) fitted to the data within each. ‘Surfaces’ estimated using the models could then be built up for each fleet, and effort estimated in the same manner as described above.

The contractor agreed that aggregation of data was probably only ‘hiding’ the underlying variability due to the fleet effect and agreed to experiment with this but noted that problems would arise because of: (i) non-random sampling in space and time; (ii) the fact that some fleets fail to report Task II data at all; and (iii) that the challenge of understanding the different fishing methods/activities is daunting.

The contractor was urged to remember the original purpose of the work. The main interest in the spatio-temporally resolved effort estimates is driven by the need to identify effort distribution by areas and time of year. This information is needed to estimate fishing impact on target and by-catch species. The Group discussed that, because fishing strategies are different among fleets, the estimation of EFFDIS by fleet is the preferable approach. It was also suggested that Task II data on their own would be enough for this and that the ‘raising’ to Task II might be unnecessary as an intermediate step. The contractor was also asked to consider the inclusion of artisanal fisheries which are important but it remains unclear where the data for this would come from and their likely quality.

In summary the contractor agreed to explore the effect of fleet/flag in more detail and make an effort to better understand the needs of the potential users for these data. The contractor is also extending the analysis too far south and the ICCAT secretariat agreed to provide more realistic boundaries within which interpolation would take place.