



**SCIENTIFIC COMMITTEE  
TENTH REGULAR SESSION**

Majuro, Republic of the Marshall Islands  
6-14 August 2014

---

**Guidelines for presenting CPUE indices of abundance for WCPFC stock assessments**

---

**WCPFC-SC10-2014/ SA-IP-11**

**Simon D Hoyle<sup>1</sup>, Adam D Langley<sup>2</sup>, and Robert A Campbell<sup>3</sup>**

---

<sup>1</sup> Consultant

<sup>2</sup> Consultant

<sup>3</sup> The Commonwealth Scientific and Industrial Research Organisation (CSIRO)

# GUIDELINES FOR PRESENTING CPUE INDICES OF ABUNDANCE FOR WCPFC STOCK ASSESSMENTS

Simon D Hoyle, Adam D Langley, Robert A Campbell

This paper presents guidelines for information that should be included in papers that describe abundance indices developed for WCPFC stock assessments from catch and effort data. The paper includes four sections:

- A) A prioritized list of items that cover the recommended information, including a section on model diagnostics.
- B) Examples of each of the suggested diagnostic figures, based on modelling of simulated data. In each case we fitted the appropriate model given the distribution of the data, so the diagnostics show examples of good fit.
- C) The R code that generated the figures above, and simulated the data used in the figures.
- D) A table reviewing papers providing CPUE indices for a number of recent WCPFC stock assessments, referencing their contents against the items on the list.

## A. LIST OF RECOMMENDED INFORMATION TO SUPPORT CPUE ANALYSES

This list identifies items that would assist review of CPUE abundance indices submitted for inclusion in WCPFC stock assessments. The items are grouped into categories according to subject area. We have allocated priorities (Priority 1: essential; Priority 2: preferable; Priority 3: additional) to identify the most important items. These should be taken as a guide rather than an authority, since the relative importance of the items will vary case-by-case, depending on the issues being considered. This list was informed by a list developed for the South East Data, Assessment, and Review Program (SEDAR 2009).

### 1. BACKGROUND

This section describes the fishery and the associated data to let the reader judge the utility of the index. It is often preferable to include relevant information in the paper being assessed. However in some cases it will be necessary to refer the reader to other papers, which can be made available online if not otherwise accessible.

#### 1.1. Fishery and management

- A. Describe the fishery and the nature of the vessels involved in the fishery, and any changes through time (e.g. catch and effort, number of vessels, and changes in spatial size of fishery). Priority 1.
- B. Identify the species targeted in the fishery, any seasonal or spatial changes in targeting, and any long term changes in targeting, and long term changes in the species composition of the catch. Priority 1.
- C. Provide (or cite) history of management regulations (e.g. spatial limits, licensing changes, gear restrictions, vessel size limits, effort or catch limits, etc.). Priority 2.
- D. Describe how management regulations may have changed fishing practises and/or fleet behaviour and how these changes may have impacted on CPUE. Priority 2.
- E. Discuss methods used (if any) to account for the effects of management measures on the CPUE series. Priority 2.

#### 1.2. Data

- A. Identify the area covered by the dataset, and how it relates to the known distribution of the stock, and/or the fishery that the index applies to. Priority 1.
- B. Describe the data source (e.g. aggregated data (and resolution), vessel logbooks or observer data), and significant stages of data management between initial recording and data analysis. Priority 1.
- C. Describe any changes through time to reporting, coverage, variables reported, etc. Priority 1.
- D. Describe the variables reported in the data set (e.g. location, time, temperature, catch, effort, fishing gear etc.). Priority 1.
- E. Describe the size/age range that the index applies to. Include supporting figures (e.g. size comp) if available. Report any significant size/age variation in time (seasonal / long term). Priority 1.
- F. Report any significant size/age variation in space. Priority 1.
- G. Report the proportion of the annual total catch (by the specific method and overall) included within the specific fishery of interest. Priority 2.
- H. Review reliability of reported catch (in weight) data, where possible. For example, the approach used to determine the catch weight/number used in the dependent variable of the CPUE analysis (e.g. conversion factors from processed to whole fish weight). Priority 2.
- I. Review the availability of additional data to verify (or corroborate) the catch and effort data. Catch and effort data could be verified from observer data and effort data (days of fishing and location) could be verified from VMS data. Priority 3.

## 2. METHODS

### 2.1. Data Reduction and Exclusions

- A. Describe any procedures used to 'clean' the data to identify and remove possible data errors and

- outliers. How many were identified? Were they excluded or substituted? Priority 1.
- B. Describe the subset (if any) of the total data selected for standardisation. Describe any data exclusions (e.g. gears, fishing modes, vessels, sampling areas etc.). Report the number of records removed and justify removal. Priority 1.
  - C. Describe data reduction techniques (if any) used to address targeting (e.g. selection based on gear configuration, species assemblage, cluster analysis etc.). Priority 1.
  - D. Describe any other processing of the data prior to standardization, such as data aggregation or imputation. Priority 1.
  - E. Present the proportion of the annual catch and/or effort records excluded/retained at each step of the data exclusion process. Priority 2.

## **2.2. Describe Analysis Dataset (after exclusions and other treatments)**

- A. Provide tables and/or figures of the total number of observations by year. Priority 1.
- B. Provide tables and/or figures of the total number of observations by model factors (including area, season and important interactions). Priority 2.
- C. Describe and justify the formulation of any species target variables, if appropriate (e.g. cluster variable, species composition). Priority 1.
- D. Describe and justify the formulation of any spatial stratification in relation to the distribution of catch, effort and CPUE from the fishery (and size structure of the catch). Priority 1.
- E. Include tables and/or figures of the number (or proportion) of positive CPUE observations by model factors and interaction terms. Give specific attention to key model variables that may have had a strong influence on catchability, such as substantial changes in fishing gear and searching technology. Priority 2.
- F. Identify levels in main model categorical variables (e.g. spatial-temporal effects) and associated interactions that have low or missing data coverage. For levels with low data coverage assess whether data may be uncharacteristic (e.g. for application of spatial imputation methods). Priority 1.
- G. Include time series plots of the main model factors. Give specific attention to key model variables that may have had a strong influence on catchability, such as substantial changes in fishing gear and searching technology. Priority 1.
- H. Include tables and/or figures of nominal and/or average (unstandardized) CPUE by main categorical variable and interaction terms. Summarize where necessary to save space. Priority 2.
- I. Include a time series of maps of locations of fishing effort and associated catch rates. Priority 1.
- J. Define the effort variable and the units. If more than one effort variable is present in the dataset, justify selection. Priority 1.
- K. Define the units of catch (e.g. numbers or biomass, whole weight, gutted weight, kilograms, metric tonnes). Where appropriate, note any changes in processing techniques over time and the application of processed-whole weight conversion factors. Priority 1.
- L. Specify all other factors included in the model (e.g. vessel-id, area effects, gear categories, etc) and any auxiliary data sets used to source values for any of the model factors (e.g. environmental data sets). Priority 1.
- M. Present simple spatial metrics to monitor concentration of fishing effort over time. Priority 2.

## **2.3. Model Standardization**

- A. Describe the class of model (e.g. GLM, GAMM, GEE). Priority 1.
- B. Specifically define the dependent variable in the model; i.e. catch (weight, number) or CPUE ratio. Priority 1.
- C. Describe the treatment of zero catch records. Priority 1.
- D. Describe and, where appropriate, justify the model structure and distributional assumptions (e.g. delta-lognormal, zero-inflated negative binomial). Priority 1.

- E. Describe the method used to adjust for changes in effort concentration through time (e.g. data weighting). Priority 1.
- F. Describe the assumptions about abundance changes through time in areas not fished by the effort included in the index (e.g. spatial imputation). Priority 1.
- G. Describe the parameterisation of variables in the model, especially the continuous variables; e.g. x-order polynomial, transformation of the variable. Priority 1.
- H. Identify the base levels used for the categorical variables in the model. Priority 2.
  - I. Describe the range of potential explanatory variables, including the resolution of the spatial variables, and interaction terms considered and offered to the model. Priority 1.
- J. Explore potential interactions with the main time effect. If these interactions are included in the final model, describe how they are included (e.g. fixed effect, random effect). Report how random effects were tested for significance. Priority 2.
- K. Describe the method for selecting model components (e.g. AIC, forward selection from null etc.), and inclusion criteria for factors and interaction terms. Priority 1.
- L. Estimate variance inflation factors (vif) to identify confounded variables. Priority 2.
- M. Provide a table summarizing the components of the final model, including the improvement in the explanatory power (AIC, r-squared, etc) with the inclusion of each variable. Priority 2.
- N. Describe how the abundance index is constructed from the model components, especially if any interactions with the main time effect are included and how missing levels, if any, in such interactions were modelled. Priority 1.
- O. For two part models (e.g. delta-lognormal), describe how the combined abundance index is constructed from the separate model components, especially how the base proportion of positive records is determined (specific to binomial model). Priority 1.
- P. Plot the standardized CPUE stepwise, showing changes as each model factor is added. Describe the influence of each model factor (e.g. using influence plots *influ* package). Priority 1.
- Q. Report convergence statistics, where appropriate (e.g. non linear zero inflated model). Priority 1.
- R. Describe the method used to estimate confidence intervals. Priority 1.
- S. Describe any sensitivity analyses conducted; e.g. to determine the influence of different vessel subsets, assumptions regarding the parameterisation of key variables in the model. Priority 2.

### 3. MODEL DIAGNOSTICS

*Comment: Other model structures are possible and acceptable. Please provide appropriate diagnostics.*

#### 3.1. General

- A. Include a design plot to give an overview of the distributions of the unstandardized variable levels. Priority 2. (Figure 1).
- B. Include interaction plots. Priority 2. (Figure 2).

#### 3.2. Binomial Component

- A. If appropriate, include plots of the chi-square residuals by factor, aggregated as necessary. Priority 1. (Figure 3 and Figure 4).
- B. Include plots of predicted and observed proportions of positive observations by year and factor (e.g. year\*area). Priority 2. (Figure 5).
- C. Report over-dispersion parameter and other fit statistics (e.g. chi-square / degrees of freedom). Priority 2.
- D. Include influence plots. Priority 1. (Figure 6).

#### 3.3. Lognormal/Gamma Component

- A. Include histogram of log(CPUE) or a histogram of the residuals of the model on CPUE. Overlay

- the expected distribution. Priority 1. (Figure 7).
- B. Include plots describing error distribution (e.g. Studentized residuals vs. linear predictor. Priority 2. (Figure 7).
- C. Include plots of the parameterisation of the key model parameters and the associated confidence intervals. Priority 1. (Figure 8).
- D. Covariance matrix. Look for strong correlations between main variables, especially highlight strong correlations with year effects. Priority 2.
- E. Include QQ-plot – (e.g. Student deviance residuals vs. theoretical quantiles), Overlay expected distribution. Priority 1. (Figure 8).
- F. Include diagnostic plot for variance function (e.g. square root of std residuals vs. fitted values). Overlay expected distribution. Priority 2. (Figure 9).
- G. Include diagnostic plot for link function (e.g. linear response variable vs. linear predictor). Overlay expected distribution. Priority 3. (Figure 9).
- H. Include plots of the residuals by factor and time. Priority 1. (Figure 10 and Figure 11).
- I. Include influence plots. Priority 1. (Figure 12).

### **3.4. Poisson / Negative Binomial Component**

- A. Report over-dispersion parameter and other fit statistics (e.g. chi-square / degrees of freedom). Priority 2.
- B. Include plots describing error distribution (e.g. Studentized residuals vs. linear predictor. Priority 1. (Figure 13).
- C. Include QQ-plot – (e.g. Student deviance residuals vs. theoretical quantiles), Overlay expected distribution. Priority 1. (Figure 13).
- D. Include diagnostic plot for variance function (e.g. square root of std residuals vs. fitted values). Overlay expected distribution. Priority 2. (Figure 14).
- E. Include diagnostic plot for link function (e.g. linear response variable vs. linear predictor). Overlay expected distribution. Priority 3. (Figure 14).
- F. Include plots of the residuals by factor and time. Priority 1. (Figure 15 and Figure 16).
- G. Include Influence plots. Priority 1. (Figure 17).

### **3.5. Zero-inflated model**

- A. Include plots of the distribution of zeroes. Priority 1. (Figure 18).
- B. Include an ROC curve to quantify goodness of fit. Priority 1. (Figure 19).
- C. Plot residuals to check for outliers. Priority 2. (Figure 19).
- D. Include plots describing error distribution (e.g. Residuals vs. linear predictor). Priority 1. (Figure 20).
- E. Include plots of the residuals by factor and time. Priority 1. (Figure 21 and Figure 22).
- F. Include a plot showing the influence of each variable in the model, and variables of interest that were not included. Variables may be either added progressively (Figure 23) or dropped from the full model. Priority 1.

## **4. MODEL RESULTS**

- A. Tables of Nominal CPUE, Standardized CPUE, Observations, Positive Observations, Proportion Positive Observations and Coefficients of Variation (CVs). Other statistics may also be appropriate to report. Priority 1.
- B. Figure of Nominal and Standardized Indices with measure of variance (i.e. CVs). Priority 1.

## **5. IF MULTIPLE MODEL STRUCTURES WERE CONSIDERED:**

*(Note: this is recommended, particularly when model diagnostics are poor.)*

- A. Plot of resulting indices and estimates of variance. Priority 1.
- B. Table of model statistics (e.g. AIC criteria). Priority 1.
- C. Identify and apply criteria that can be used to assess the utility (or merit) of each of the resulting indices. Priority 1.

## B. EXAMPLES OF DIAGNOSTIC PLOTS

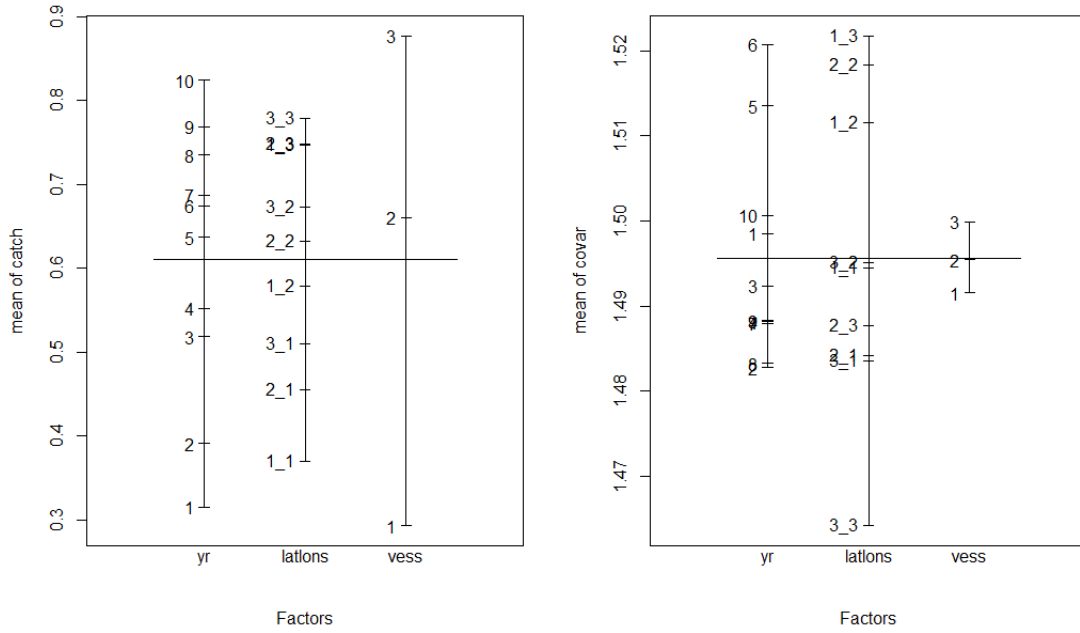


Figure 1: Design plot, showing the mean values of the response variable catch (left) and the covariate (right) for each factor submitted to the function.

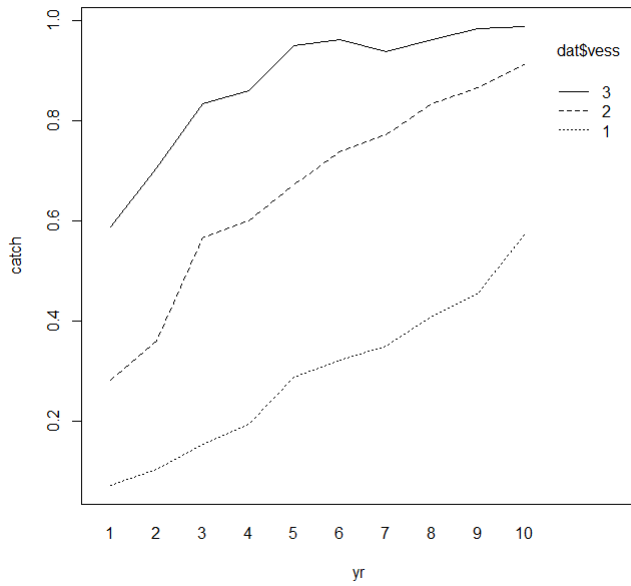


Figure 2: Interaction plot, showing the unstandardized distribution of catch by vessel and year.



# Binomial

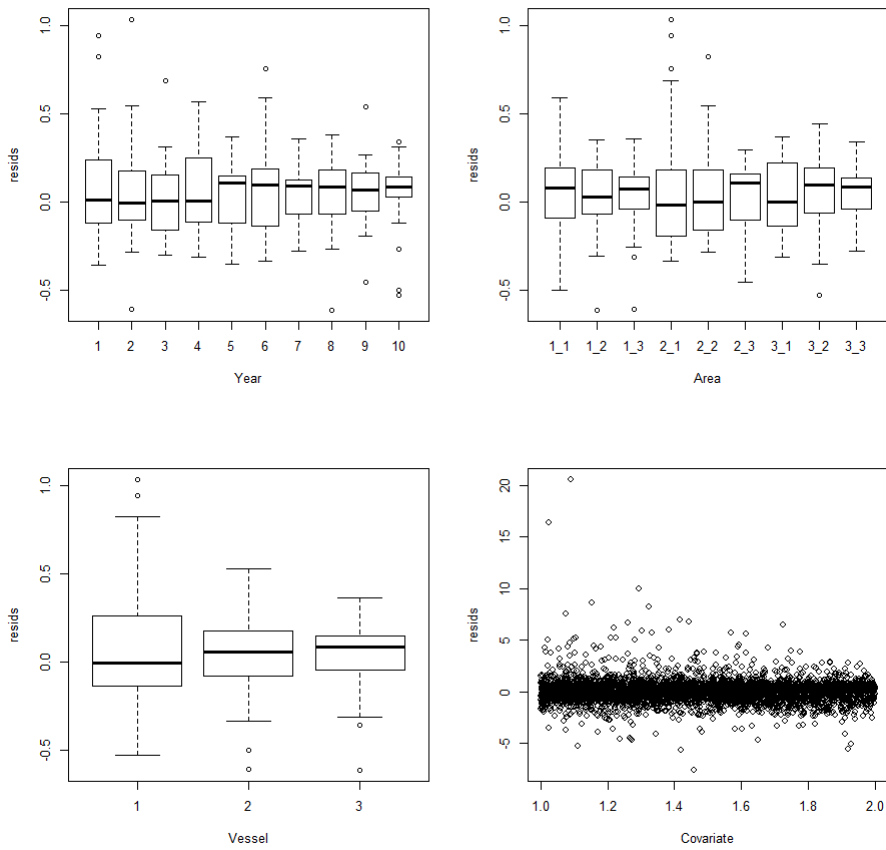


Figure 3: BIN\_A Boxplots of residuals by factor value, and raw values versus the covariate.

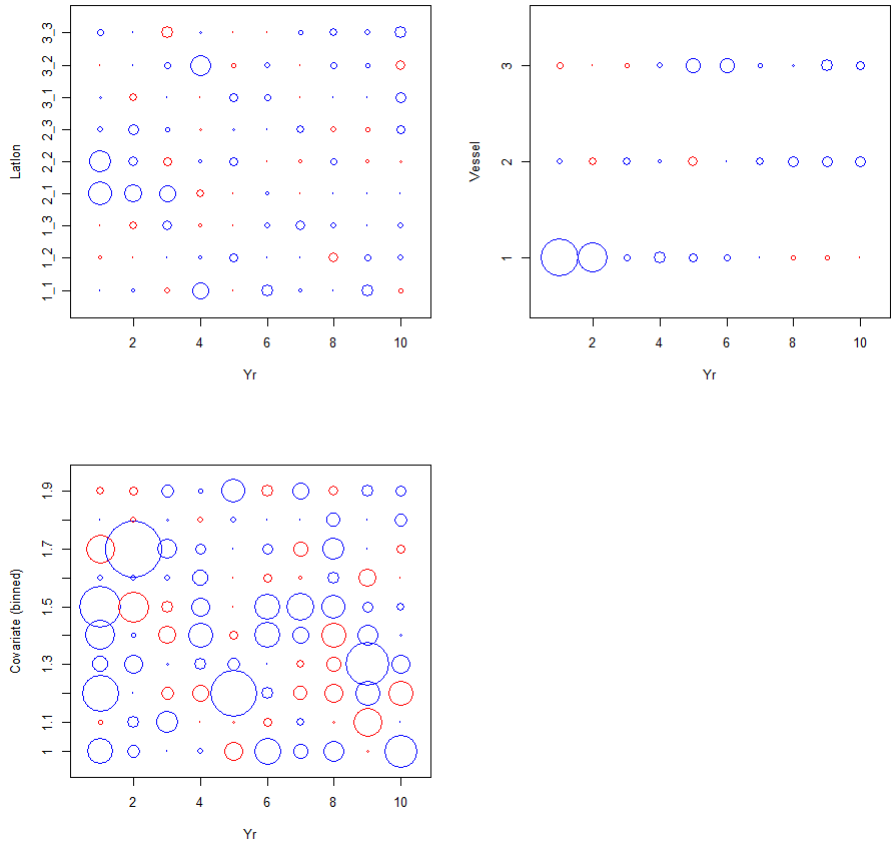


Figure 4: BIN\_A Bubble plots of residuals by effect and year, with the covariate binned to facilitate plotting.

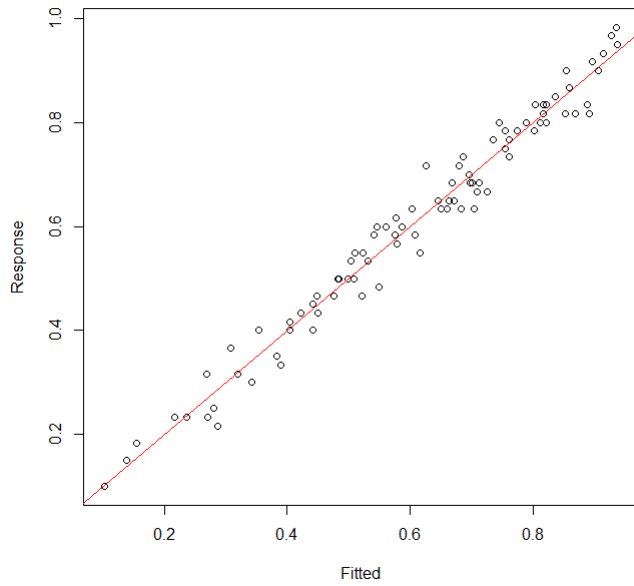


Figure 5: BIN\_B\_Predicted proportion nonzero vs response.

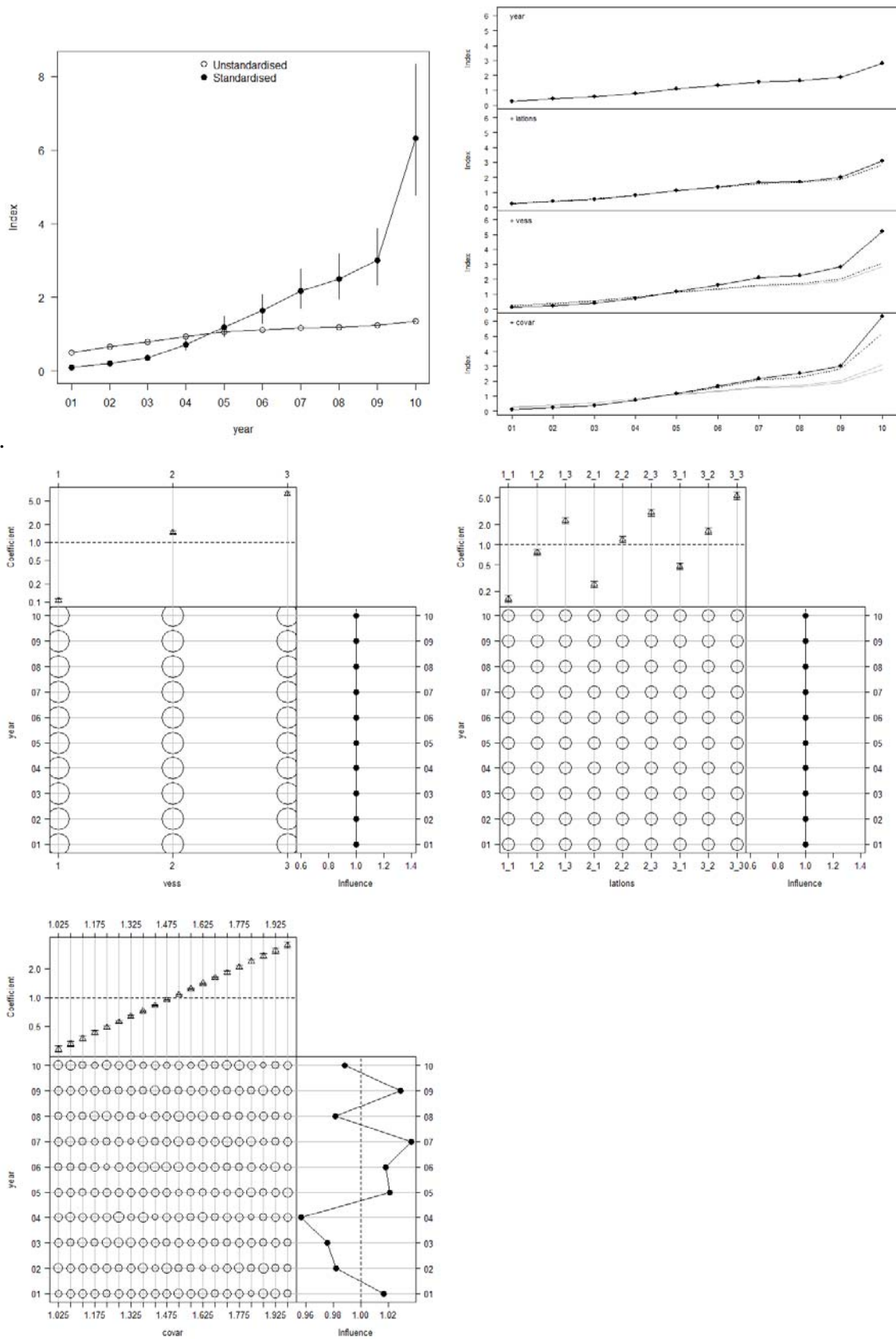


Figure 6: BIN\_D Influence plots showing the unstandardized and standardized index, progressive changes as each covariate is added, and a plot demonstrating the influence of each covariate.

# Lognormal

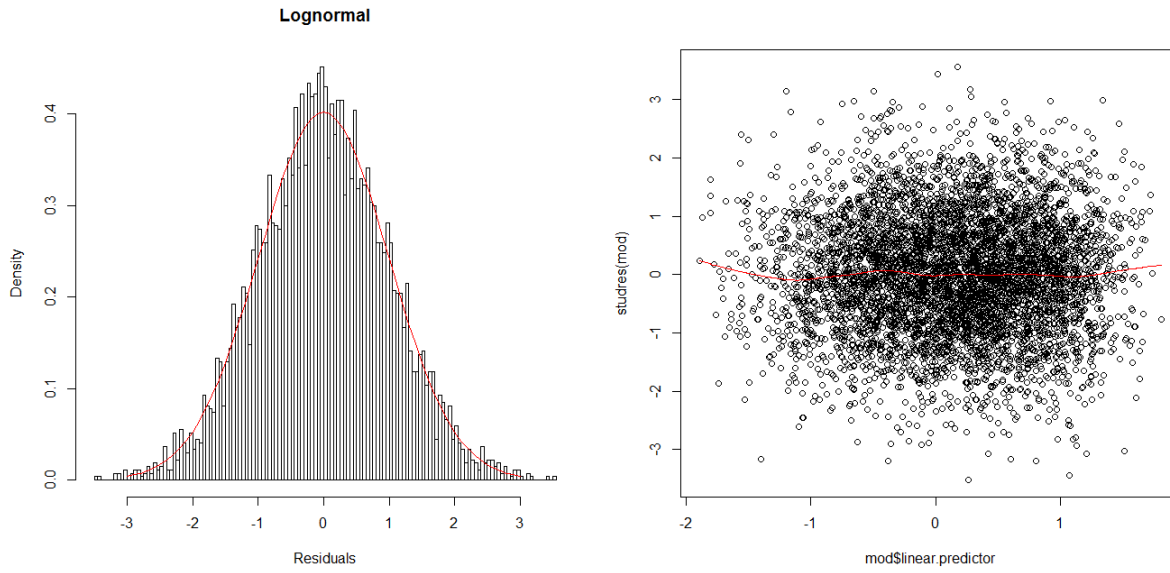


Figure 7: LN\_A Observed and expected frequency distributions (left), and LN\_B Residuals versus predictor (right).

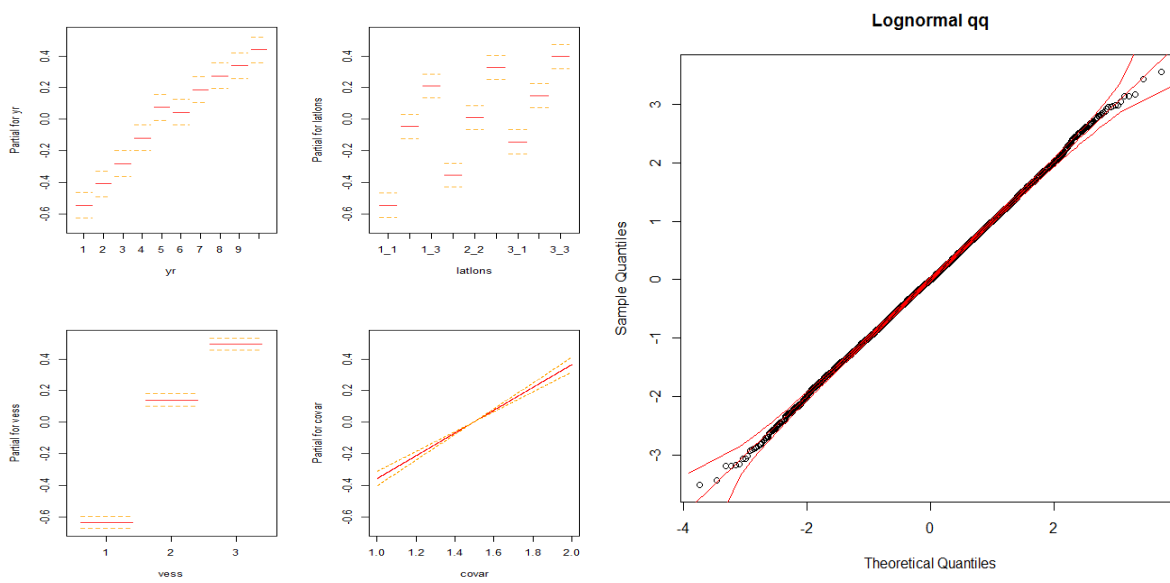


Figure 8: LN\_C Plots of estimated terms and confidence intervals (left) and LN\_E, QQ plot (right).

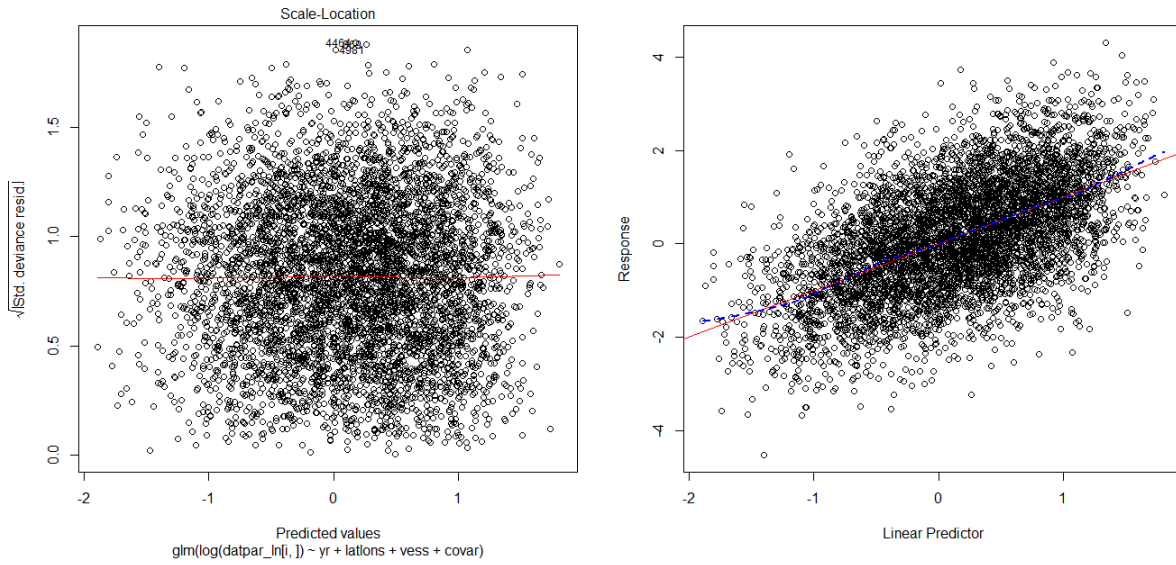


Figure 9: LN\_F Residuals versus linear predictor with a smoother in red (left) and LN\_G Response versus linear predictor with smoothed mean marked in blue and a line of equality in red (right).

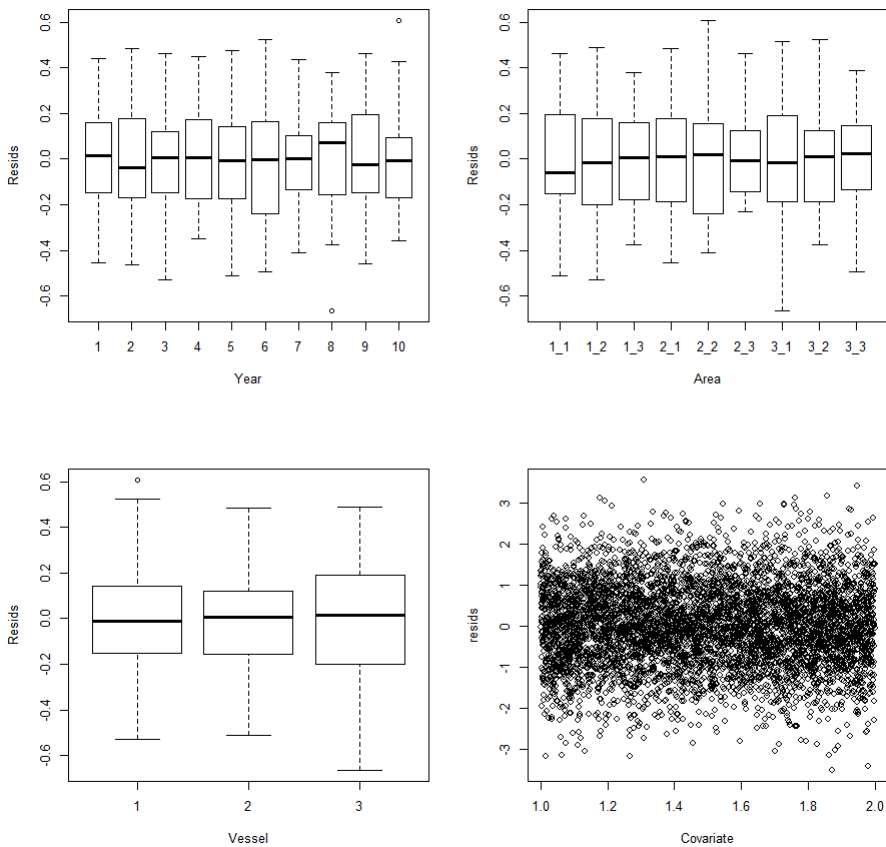


Figure 10: LN\_H Boxplots of residuals by factor value, and raw values versus the covariate.

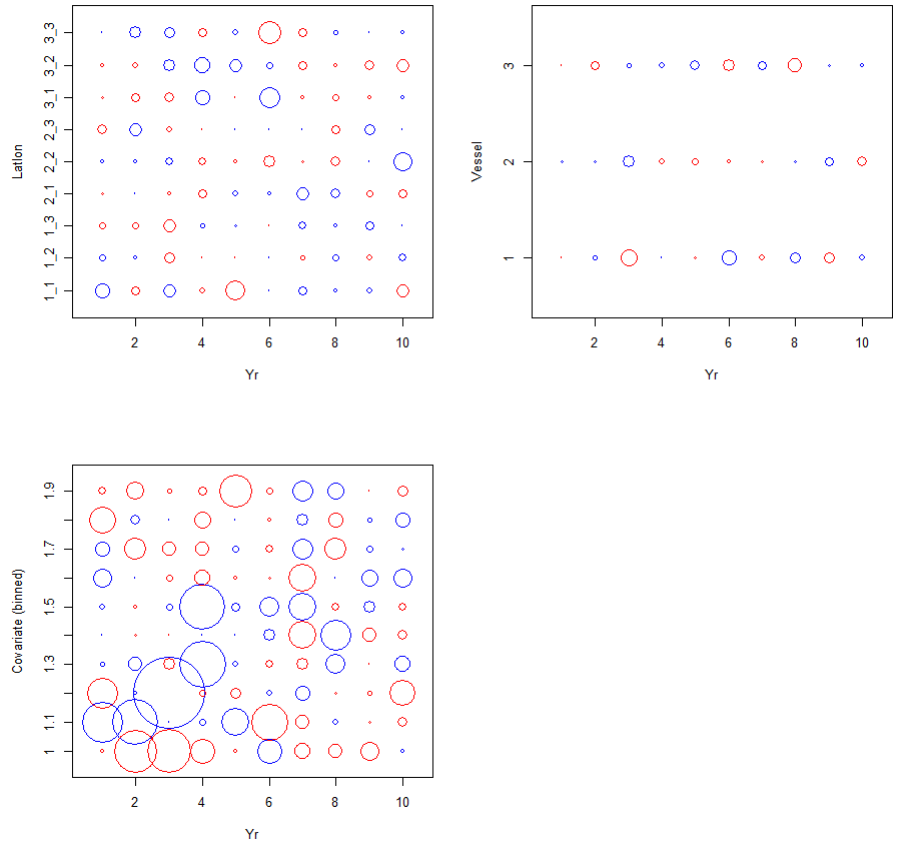


Figure 11: LN\_H Bubble plots of residuals by effect and year, with the covariate binned to facilitate plotting.

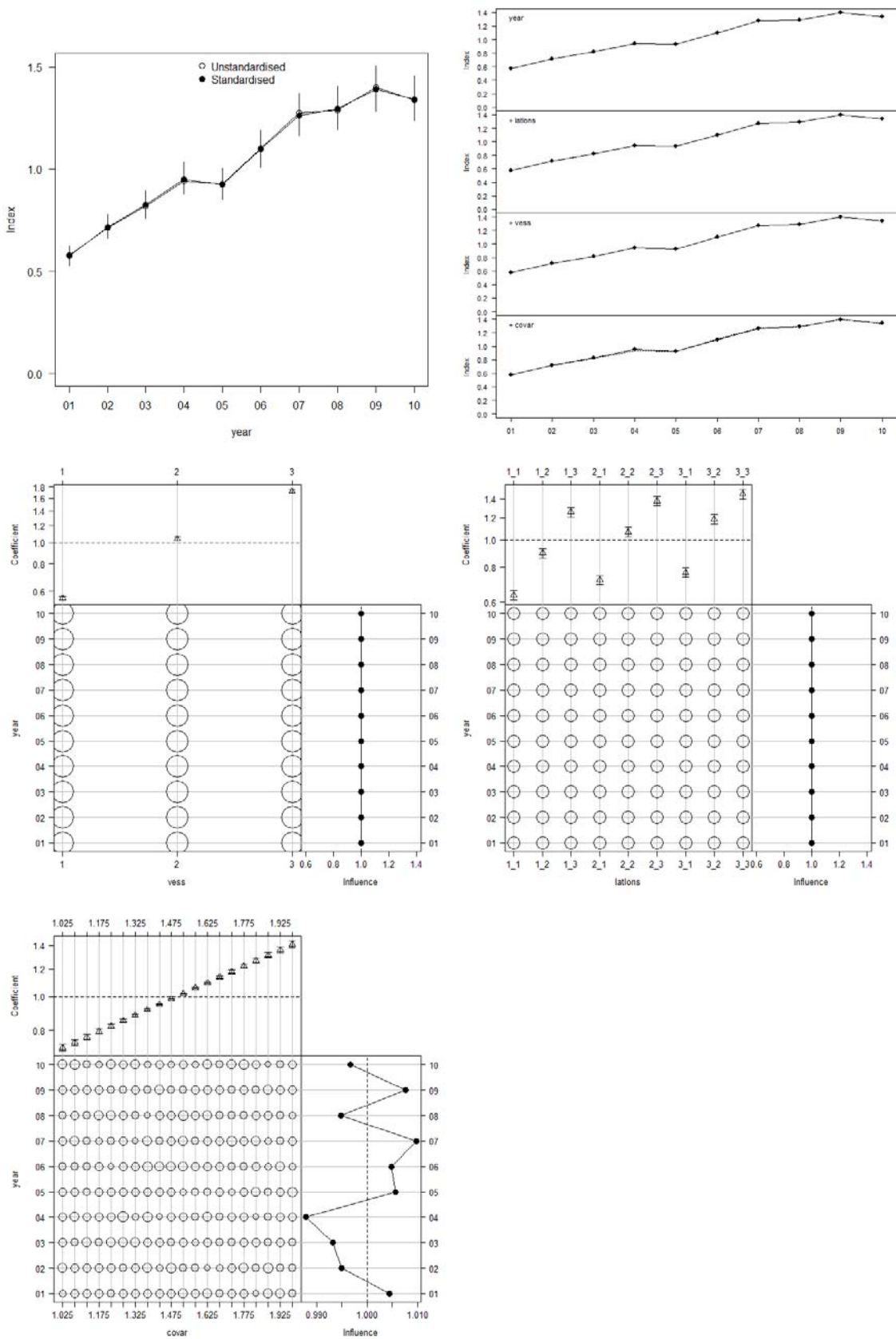


Figure 12: LN\_I Influence plots showing the unstandardized and standardized index, progressive changes as each covariate is added, and a plot demonstrating the influence of each covariate.

## Poisson and Negative Binomial

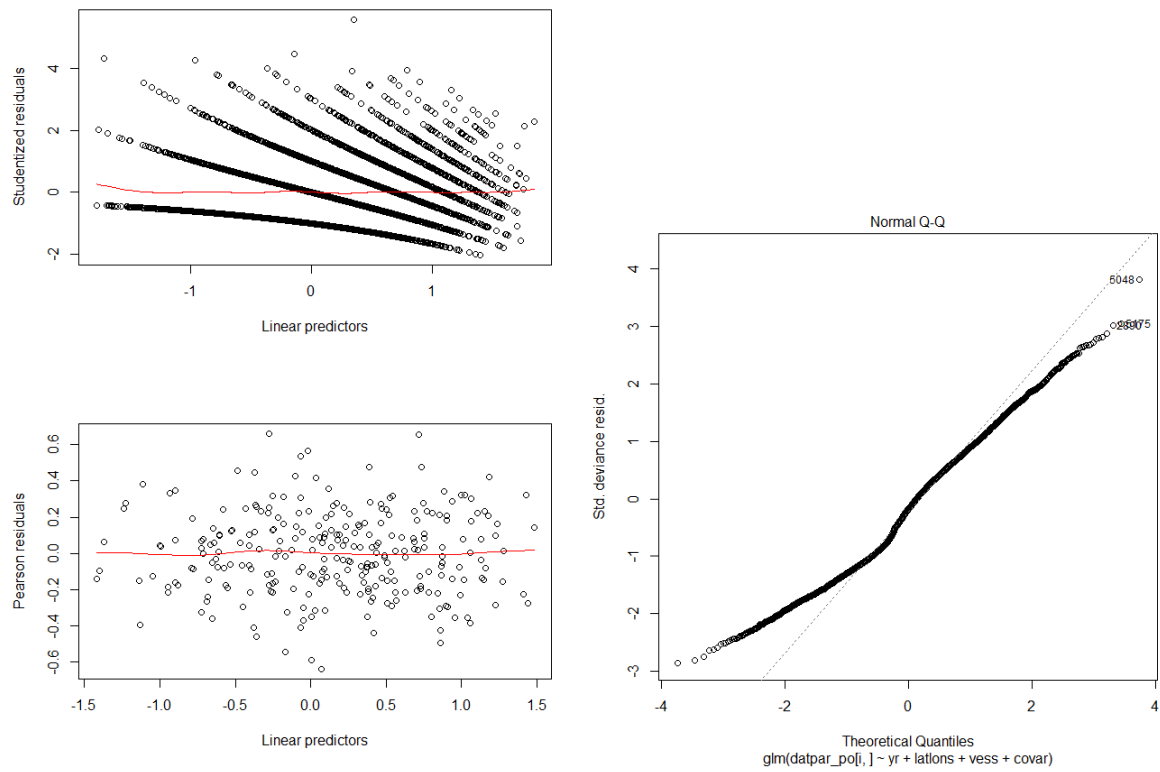


Figure 13: POI\_B Residuals versus linear predictor, by set and aggregated, with smoothed mean in red (left), and POI\_C QQ plot (right).

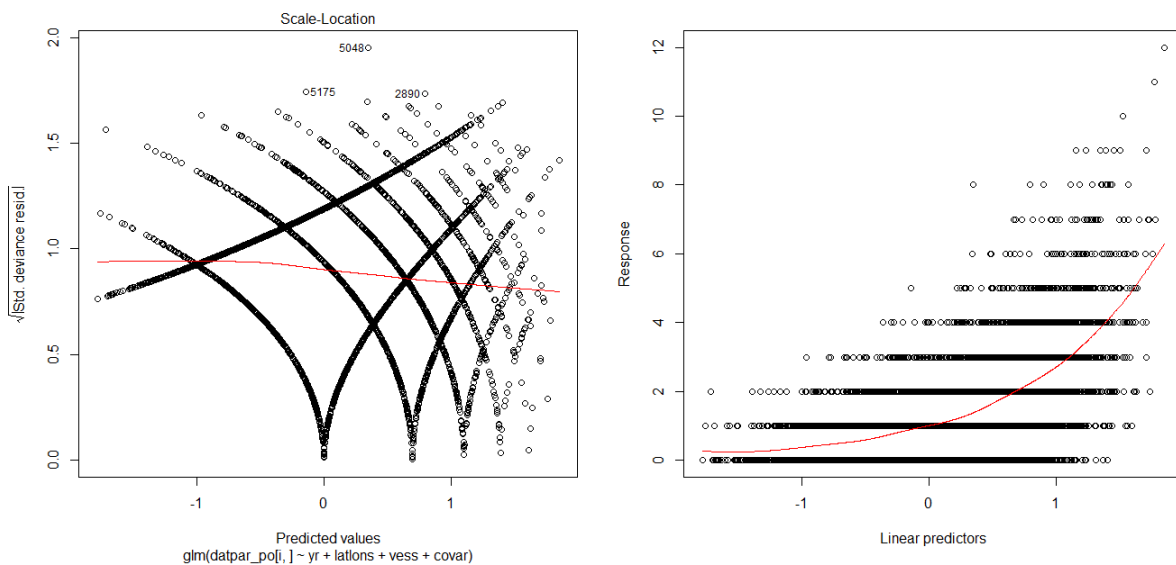


Figure 14: POI\_D Square root of standardized residuals versus fitted values with smoothed mean in red (left) and POI\_E Response variable versus linear predictor, with smoothed mean in red (right).



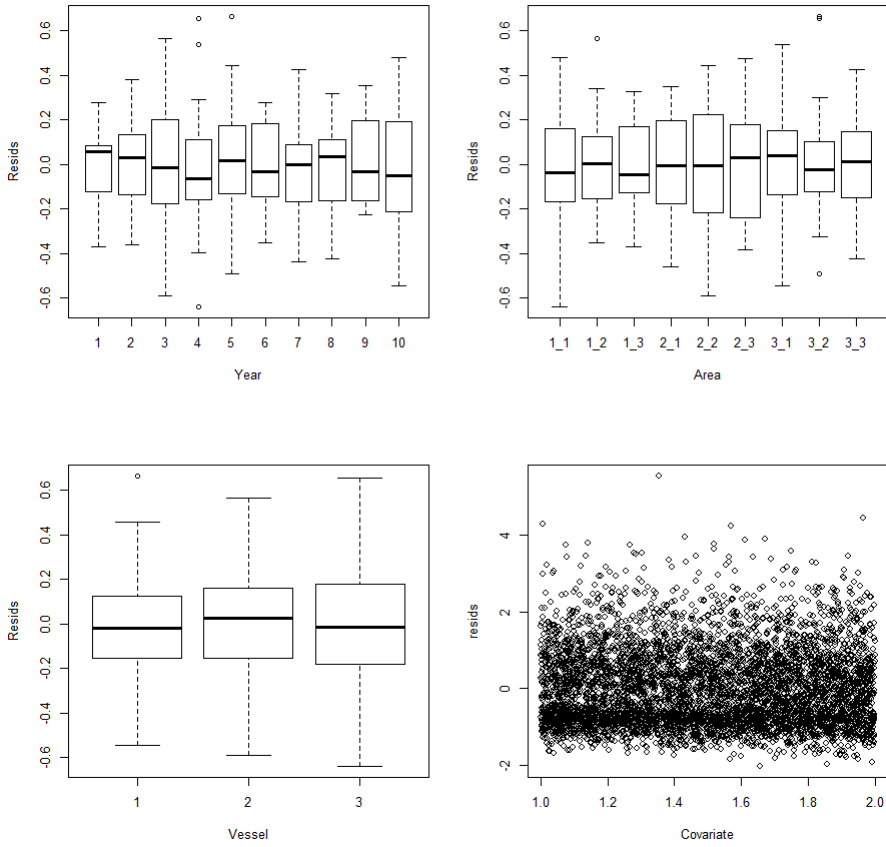


Figure 15: POI\_F Boxplots of residuals by factor value, and raw values versus the covariate.

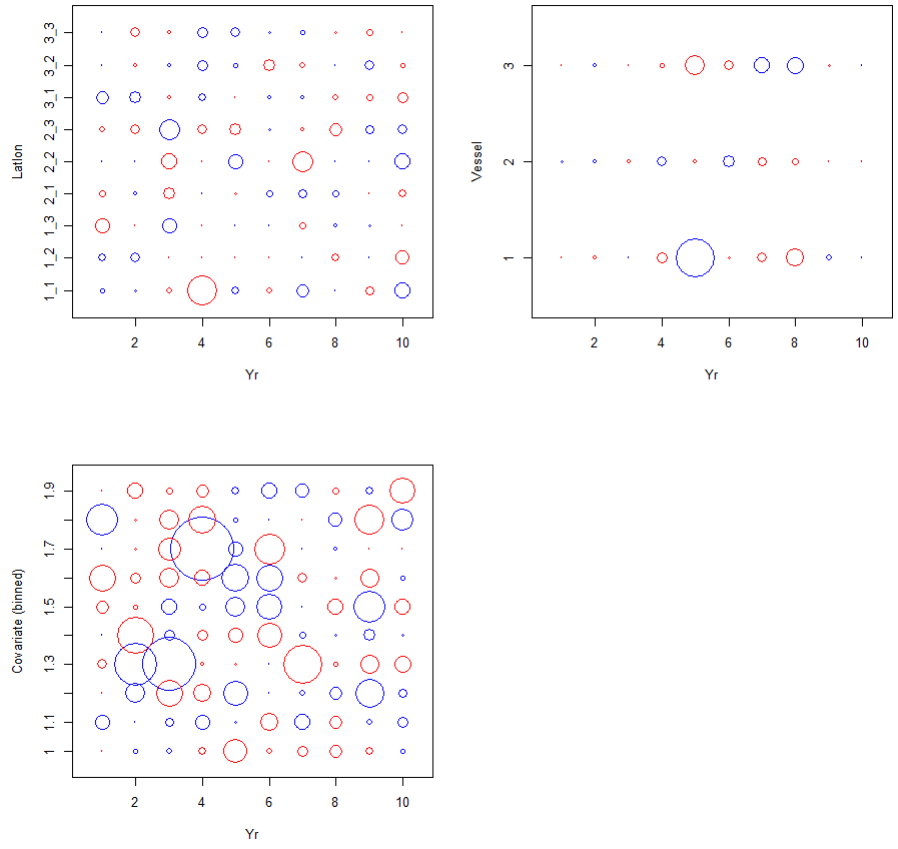


Figure 16: POI\_F Bubble plots of residuals by effect and year, with the covariate binned to facilitate plotting.

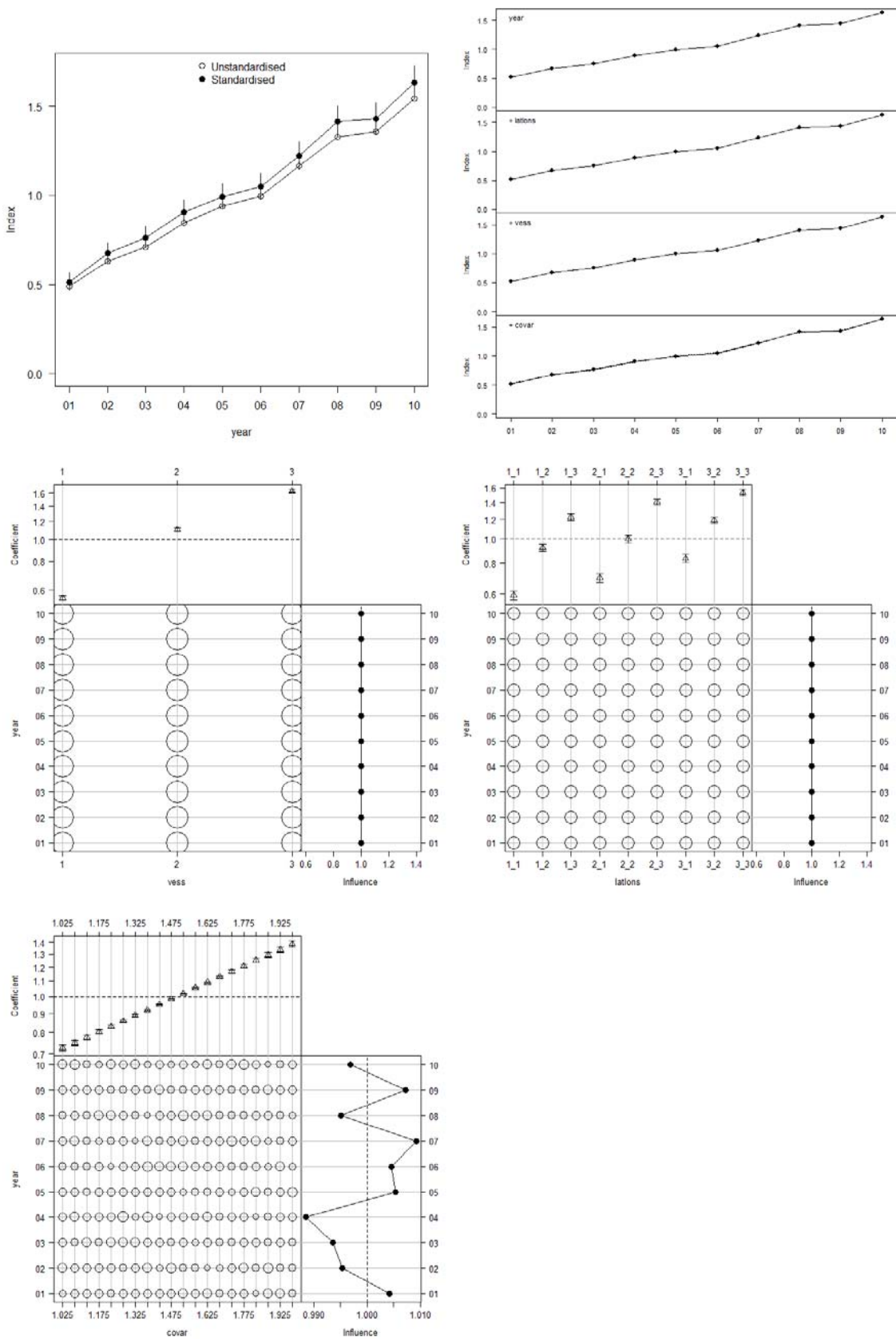


Figure 17: POI\_G Influence plots showing the unstandardized and standardized index, progressive changes as each covariate is added, and a plot demonstrating the influence of each covariate.

## Zero Inflated Poisson and Zero Inflated Negative Binomial

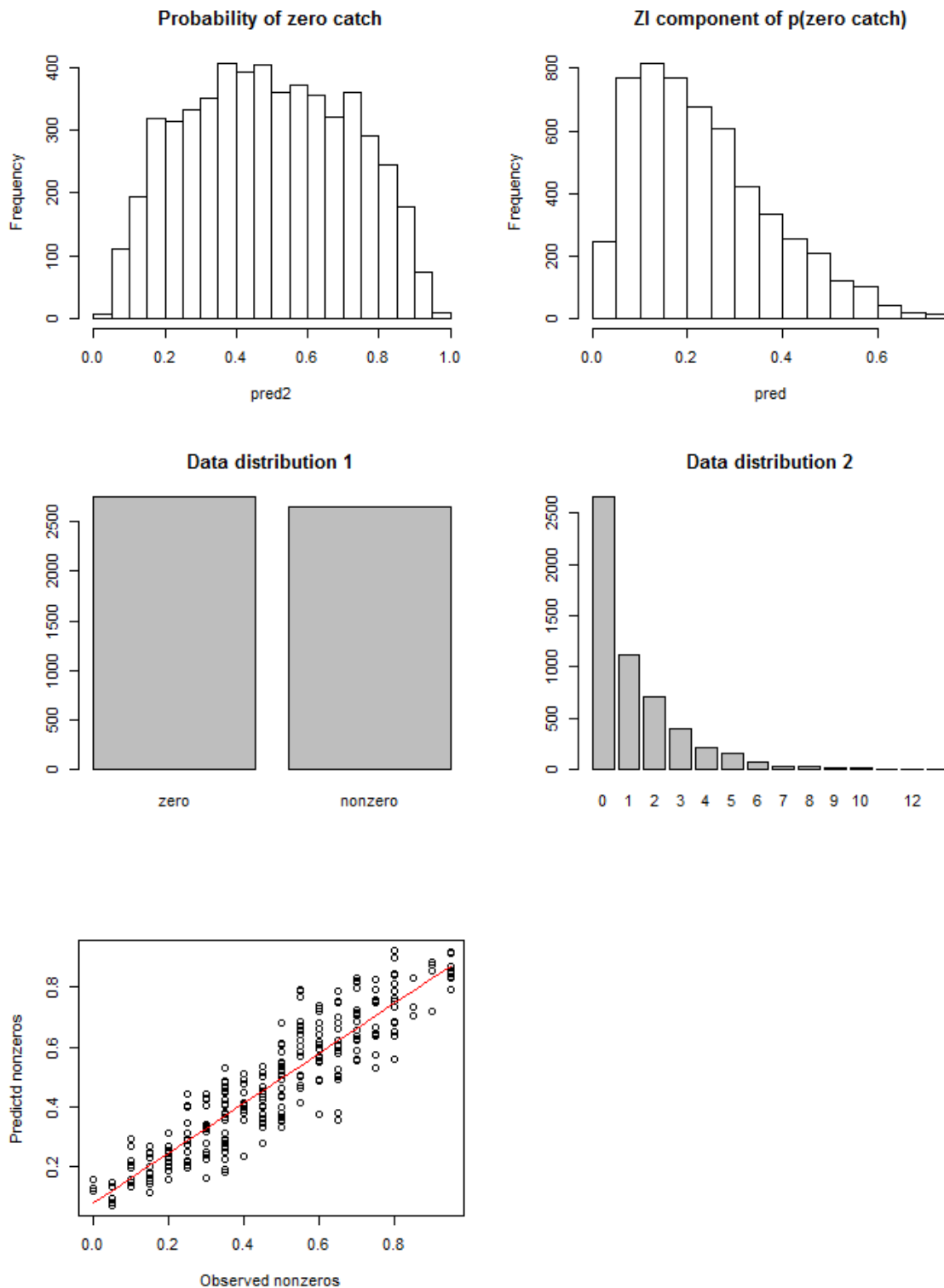


Figure 18: ZINB\_A Plots showing the distributions of observed and expected zeroes.

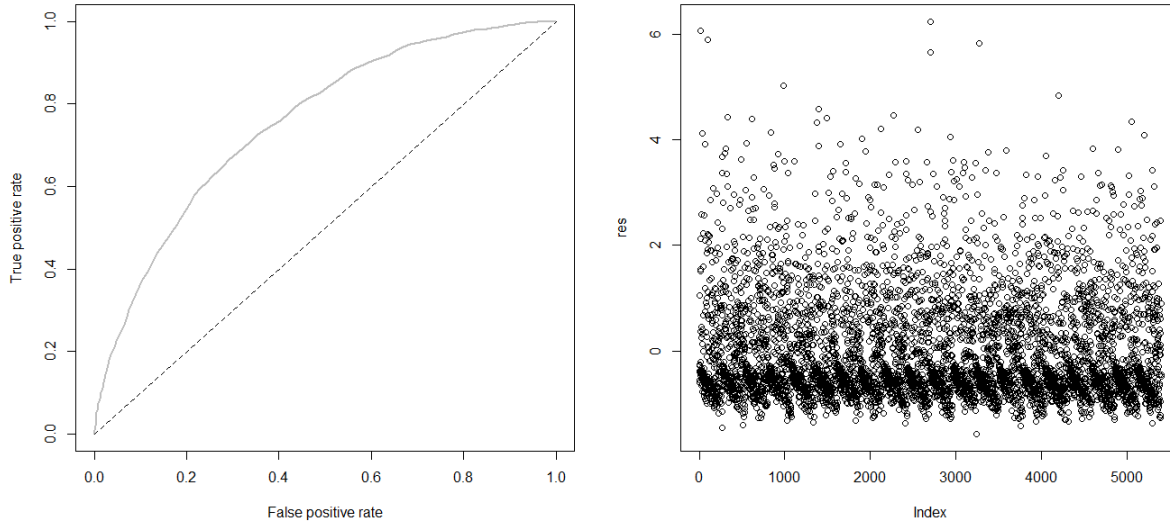


Figure 19: ZINB\_B Receiver operating characteristic (ROC) curve showing the proportions of true and false positives (left), and ZINB\_C Raw residual plot (right).

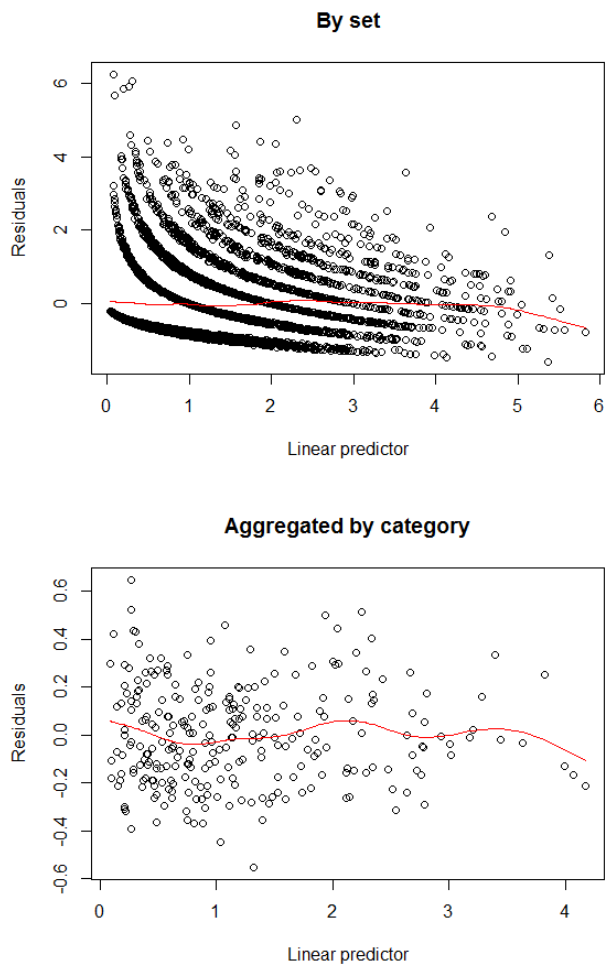


Figure 20: ZINB\_E Plots of residuals versus linear predictor, both by set and aggregated by category, with smoothed mean in red.

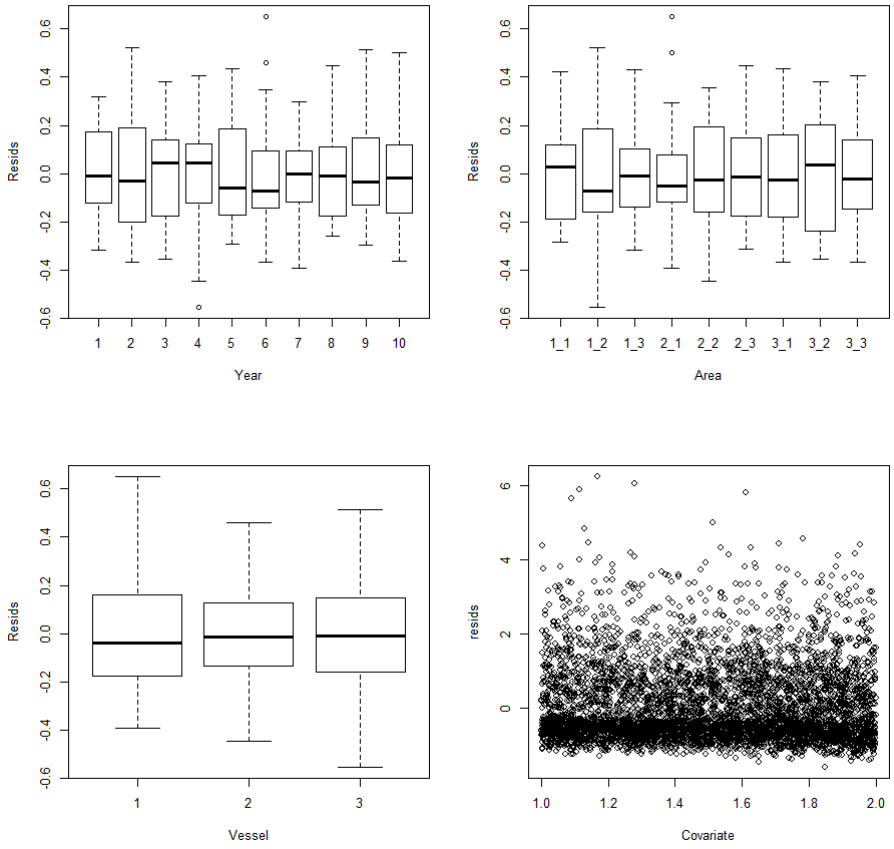


Figure 21: ZINB\_E Boxplots of residuals by factor value, and raw values versus the covariate.

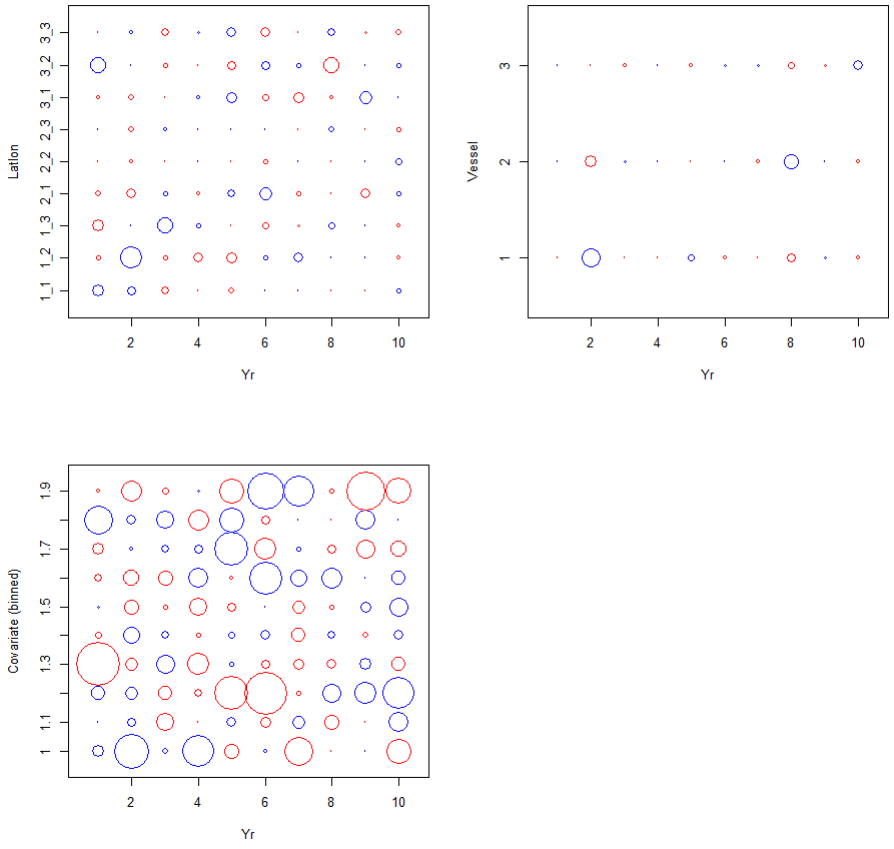


Figure 22: ZINB\_E Bubble plots of residuals by effect and year, with the covariate binned to facilitate plotting.

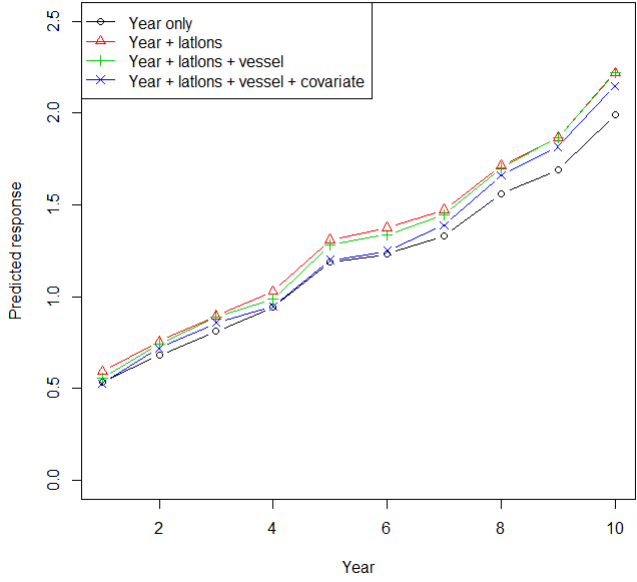


Figure 23: ZINB\_F Plot showing the influence of each variable as they are added progressively to the model.

## C. R CODE TO PRODUCE DIAGNOSTIC PLOTS

The following R code was used to generate the diagnostic plots in the previous section. The code contains three sections: first it simulates multiple realisations of simple catch and effort data from various distributions; secondly it validates the data by fitting the appropriate model to each dataset and plotting the indices; third it generates diagnostic plots for the appropriate model fitted to one dataset from each distribution.

```
### R script with diagnostic plot examples
setwd("C:/Users/Username/CPUE")
require(boot)
require(MASS)
require(pscl)
require(influ)

### Operating model
## Set up
# numbers and labels
nyr <- 10; nlat <- 3; nlon <- 3; n vess <- 3; nset=20; nsim=25
yr <- as.factor(2001:(2000+nyr))
yr <- 1:nyr
lats <- as.factor(1:nlat)
lons <- as.factor(1:nlon)
latlons <- expand.grid(lats,lons)
latlons <- as.factor(paste(latlons$Var1,latlons$Var2,sep="_"))
nlatlons <- length(latlons)
vess <- as.factor(1:n vess)

# parameter values
dens_ll <- 1:nlatlons + nlatlons/2
dens_ll <- dens_ll/mean(dens_ll)
abund_yr <- 0.4 + 1:nyr/10
q_vess <- 1:n vess / mean(1:n vess)
dat <- expand.grid(yr=yr,lat=lats,lon=lons,vess=vess,set=1:nset,hooks=1000)
dat$year <- as.factor(formatC(dat$yr,width=2,flag="0"))
dat$covar <- runif(n=length(dat$lat),min=1,max=2)
dat$latlons <- as.factor(paste(dat$lat,dat$lon,sep="_"))
datparmn <- matrix(nrow=nsim,ncol= dim(dat)[1])

## Simulate
# Expected catch per set for each vessel in each area each year
for(i in 1:nsim) datparmn[i,] <- abund_yr[match(dat$yr,yr)] * dens_ll[match(dat$latlons,latlons)] *
q_vess[match(dat$vess,vess)] * dat$covar

## Add distribution
# binomial
datpar_bin <- matrix(rbinom(datparmn,size=1,prob=inv.logit(4*datparmn-4)),nrow=nsim)
```



```

# lognormal
lnsd=1
datpar_ln <- matrix(datparmn * exp(rnorm(datparmn,mean=0,sd=lnsd)),nrow=nsim)

# poisson
datpar_po <- matrix(rpois(datparmn,lambda=datparmn),nrow=nsim)

# negbin
datpar_nb <- matrix(rnbinom(datparmn,size=5,mu=datparmn),nrow=nsim)

# ZINB
zi <- exp(datparmn)/(1+exp(datparmn))
#zi <- 0.8
datpar_zi <- matrix(rbinom(datparmn,size=1,prob=zi),nrow=nsim)
datpar_znb <- matrix(rnbinom(datparmn,size=5,mu=datparmn),nrow=nsim)
datpar_zinb <- datpar_zi * datpar_znb
#rbind(datpar_zi[1,],datpar_znb[1,],datpar_zinb[1,])

# ZANB
datpar_zanb <- datpar_zinb
graphics.off()

### Test Parameter Estimation
# Lognormal models
windows(); par(mfrow=c(3,2))
plot(yr,abund_yr * dens_ll[5] * q_vess[2] * 1.5,ylim=c(0,3),main="Lognormal")
mod <- list()
newdat <- expand.grid(yr=sort(unique(dat$yr)),latlons=latlons[5],vess=vess[2],covar=1.5)
for(i in 1:nsim) {
  mod[[i]] <- glm(log(datpar_ln[i,]) ~ yr + latlons + vess + covar,data=dat)
  newdat$par <- exp(predict(mod[[i]],newdata=newdat,type="response"))
  points(jitter(as.numeric(newdat$yr)),newdat$par,col=i,cex=1)
}
predict(mod[[i]],newdata=newdat,type="terms",se.fit=T)
predict(mod[[i]],newdata=newdat,type="response",se.fit=T)

# Poisson models
plot(yr,abund_yr * dens_ll[5] * q_vess[2] * 1.5,ylim=c(0,3),main="Poisson")
mod <- list()
for(i in 1:nsim) {
  mod[[i]] <- glm(datpar_po[i,] ~ yr + latlons + vess + covar,data=dat,family=poisson)
  newdat$par <- predict(mod[[i]],newdata=newdat,type="response")
  points(jitter(as.numeric(newdat$yr)),newdat$par,col=i,cex=1)
}

# Negative binomial models
plot(yr,abund_yr * dens_ll[5] * q_vess[2] * 1.5,ylim=c(0,3),main="Negative binomial")
mod <- list()
for(i in 1:nsim) {

```

```

mod[[i]] <- glm.nb(datpar_nb[i,] ~ yr + latlons + vess + covar,data=dat)
newdat$par <- predict(mod[[i]],newdata=newdat,type="response")
points(jitter(as.numeric(newdat$yr)),newdat$par,col=i,cex=1)
}

```

# Zero adjusted negative binomial models

```

plot(yr,abund_yr * dens_ll[5] * q_vess[2] * 1.5 * inv.logit(abund_yr * dens_ll[5] * q_vess[2] *
1.5),ylim=c(0,2.5),main="ZANB")
mod <- list()
for(i in 1:nsim) {
  mod[[i]] <- hurdle(datpar_zanb[i,] ~ yr + latlons + vess + covar,data=dat,dist="negbin")
  newdat$par <- predict(mod[[i]],newdata=newdat,type="response")
  points(jitter(as.numeric(newdat$yr)),newdat$par,col=i,cex=1)
}

```

# Zero inflated negative binomial models

```

plot(yr,abund_yr * dens_ll[5] * q_vess[2] * 1.5 * inv.logit(abund_yr * dens_ll[5] * q_vess[2] *
1.5),ylim=c(0,2.5),main="ZINB")
mod <- list()
for(i in 1:nsim) {
  res <- try(zeroinfl(datpar_zinb[i,] ~ yr + latlons + vess + covar,data=dat,dist="negbin"))
  if(inherits(res, "try-error")) {
    print(paste("Not fitted",i))
    flush.console()
  }
  else {
    mod[[i]] <- res
    newdat$par <- predict(mod[[i]],newdata=newdat,type="response")
    points(jitter(as.numeric(newdat$yr)),newdat$par,col=i,cex=1)
  }
}

```

savePlot("Simulation\_plots",type="png")

#####

## **# Diagnostics**

### **# Some functions**

# histogram of log(CPUE) or a histogram of the residuals of the model on CPUE. Overlay the expected distribution

```

qqDist <- function (x, standardise = F, add.median = F, ti="", ...)

```

```

{
  n <- length(x)
  seq.length <- min(1000, n)
  if (standardise) {
    SEQ <- seq(1, 2 * n + 1, length = seq.length)/2
    U <- qnorm(qbeta(0.975, SEQ, rev(SEQ)))
    L <- qnorm(qbeta(0.025, SEQ, rev(SEQ)))
    if (add.median)
      M <- qnorm(qbeta(0.5, SEQ, rev(SEQ)))

```

```

}
else {
  SD <- sqrt(var(x) * (n + 1)/n)
  SEQ <- seq(1, 2 * n + 1, length = seq.length)/2
  U <- mean(x) + SD * qt(qbeta(0.975, SEQ, rev(SEQ)), n -
    1)
  L <- mean(x) + SD * qt(qbeta(0.025, SEQ, rev(SEQ)), n -
    1)
  if (add.median)
    M <- mean(x) + SD * qt(qbeta(0.5, SEQ, rev(SEQ)),
      n - 1)
}
X <- qnorm((SEQ - 0.25)/(n + 0.5))
qqnorm(x, main = ti, ...)
lines(X, U, type = "l",col=2)
lines(X, L, type = "l",col=2)
if (add.median)
  lines(X, M, type = "l",col=2)
invisible()
}

```

```

plotlognorm <- function(res,ti="") {
  hist(res,nclass=200,freq=F,xlab="Residuals",main=ti)
  lines((-30:30)/10,dnorm((-30:30)/10,sd=sd(res)),col=2)
}

```

```

plotqq <- function(res,ti="") {
  sdres <- res/sd(res)
  qqDist(sdres,add.median=T,ti=ti)
}

```

## Plots to be used for all cases

```

mod <- glm(datpar_bin[1,] ~ yr + latlons + vess + covar,data=dat,family="binomial")

```

```

windows(width=11,height=7); par(mfrow=c(1,2))
dat$catch <- datpar_bin[1,]
plot.design(catch ~ yr + latlons + vess + covar,data=dat)
savePlot("GEN_A_Design_plot",type="png")
windows()
interaction.plot(dat$yr, dat$vess,dat$catch, xlab="yr",ylab="catch")
savePlot("GEN_B_interaction_plot",type="png")

```

## Binomial model

```

mod <- glm(datpar_bin[1,] ~ yr + latlons + vess + covar,data=dat,family="binomial")

```

#A. If appropriate, include plots of the chi-square residuals, using box plots or bubble plots, by factor.  
Priority 1.

```

windows(height=10,width=10);par(mfrow=c(2,2))
b <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$latlons +
dat$vess,FUN=mean) ;names(b) <- c("yr","latlons","vess","y","res") # Pearson chi-square residuals
boxplot(b$res ~ b$yr,xlab="Year",ylab="resids")
boxplot(b$res ~ b$latlons,xlab="Area",ylab="resids")
boxplot(b$res ~ b$vess,xlab="Vessel",ylab="resids")
plot(residuals(mod,type='pearson') ~ dat$covar,xlab="Covariate",ylab="resids")
savePlot("BIN_A_resids_by_variable",type="png")

```

```

windows(height=10,width=10);par(mfrow=c(2,2))
bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$latlons,FUN=mean)
;names(bb) <- c("yr","latlons","y","res") # Pearson chi-square residuals
plot(as.numeric(as.character(bb$yr)),bb$latlons,xlim=c(0.5,10.5),ylim=c(0.5,length(unique(bb$latlons))+0.5),type="n",xlab="Yr",ylab="Latlon",yaxt="n")
axis(2,at=unique(bb$latlons),labels=unique(bb$latlons))
a <- bb[bb$res >= 0,]
symbols(as.numeric(as.character(a$yr)),a$latlons,circles=a$res,add=T,fg=4, inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$latlons,circles=-a$res,add=T,fg=2, inches=F)

```

```

bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$vess,FUN=mean)
;names(bb) <- c("yr","vess","y","res") # Pearson chi-square residuals
plot(as.numeric(as.character(bb$yr)),bb$vess,xlim=c(0.5,10.5),ylim=c(0.5,length(unique(bb$vess))+0.5),type="n",xlab="Yr",ylab="Vessel",yaxt="n")
axis(2,at=unique(bb$vess),labels=unique(bb$vess))
a <- bb[bb$res >= 0,]
symbols(as.numeric(as.character(a$yr)),a$vess,circles=2*a$res,add=T,fg=4, inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$vess,circles=-2*a$res,add=T,fg=2, inches=F)

```

```

dat$covar_b <- floor(dat$covar*10)/10
bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$covar_b,FUN=mean)
;names(bb) <- c("yr","covar_b","y","res") # Pearson chi-square residuals
edge <- min(bb$covar_b)*0.05
plot(as.numeric(as.character(bb$yr)),bb$covar_b,xlim=c(0.5,10.5),ylim=c(min(bb$covar_b)-edge,max(bb$covar_b) + edge),type="n",xlab="Yr",ylab="Covariate (binned)",yaxt="n")
axis(2,at=unique(bb$covar_b),labels=unique(bb$covar_b))
a <- bb[bb$res >= 0,]
symbols(as.numeric(as.character(a$yr)),a$covar_b,circles=2*a$res,add=T,fg=4, inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$covar_b,circles=-2*a$res,add=T,fg=2, inches=F)
savePlot("BIN_A_bubble_plots_by_variable",type="png")

```

#B. Include plots of predicted and observed proportions of positive observations by year and factor (e.g. year\*area). Priority 1.

```

windows(height=7,width=7)
a <- aggregate(mod$fitted.values ~ dat$yr + dat$latlons,FUN=mean);names(a) <-
c("yr","latlons","fitted")

```

```

b <- aggregate(mod$y ~ dat$yr + dat$latlons,FUN=mean) ;names(b) <- c("yr","latlons","y")
plot(a$fitted,b$y,xlab="Fitted",ylab="Response"); xx <- (min(a$fitted)-
2):(2+max(a$fitted));lines(xx,xx,col=2)
savePlot("BIN_B_Response_vs_fitted",type="png")

```

#C. Report overdispersion parameter and other fit statistics (e.g. chi-square / degrees of freedom).  
Priority 1.

```

pearson.deviance <- sum((mod$y-mod$fitted.values)^2/mod$fitted.values)
pearson.deviance / mod$df.residual

```

```

mod <- glm(datpar_bin[1,] ~ year + latlons + vess + covar,data=dat,family="binomial")

```

#D. influence plots

```

graphics.off()
infYear <- Influence$new(mod,data=dat)
infYear$calc()
infYear$summary
windows(height=7,width=7)
infYear$stanPlot()
savePlot("BIN_D1_Influence plots",type="png")
infYear$stepPlot()
savePlot("BIN_D2_Influence plots",type="png")
infYear$cdiPlot('vess')
savePlot("BIN_D3_Influence plots",type="png")
infYear$cdiPlot('latlons')
savePlot("BIN_D4_Influence plots",type="png")
infYear$cdiPlot('covar')
savePlot("BIN_D5_Influence plots",type="png")
graphics.off()

```

## Lognormal/gamma model

```

mod <- glm(log(datpar_ln[i,]) ~ yr + latlons + vess + covar,data=dat)

```

# A. Histogram of the residuals of the model on CPUE. Overlay expected distribution.

```

windows(height=7,width=7)
plotlognorm(res=mod$residuals,ti="Lognormal")
savePlot("LN_A_Frequency distribution",type="png")

```

#B. Include plots describing error distribution (e.g. Studentized residuals vs. linear predictor. Priority 1.

```

windows(height=7,width=7)
plot(mod$linear.predictor,studres(mod));
lines(smooth.spline(mod$linear.predictor,studres(mod),spar=1),col=2)
savePlot("LN_B_Residuals_vs_predictor",type="png")

```

#C. Include plots of the parameterisation of the key model parameters and the associated confidence intervals. Priority 1.

```

windows(height=10,width=7);par(mfrow=c(2,2))

```

```
termplot(mod,se=T)
savePlot("LN_C_Termplot",type="png")
```

```
# E. QQ-plot. Overlay expected distribution
windows(height=7,width=7)
plotqq(res=mod$residuals,ti="Lognormal qq")
savePlot("LN_E_Proportions_by_variable",type="png")
```

```
#F. Include diagnostic plot for variance function (e.g. square root of std residuals vs. fitted values).
Overlay expected distribution. Priority 1.
windows(height=7,width=7)
plot(mod,which=3)
savePlot("LN_F_Resids_vs_linear_predictor",type="png")
```

```
#G. Include diagnostic plot for link function (e.g. linear response variable vs. linear predictor).
Overlay expected distribution. Priority 1.
windows(height=7,width=7)
plot(mod$linear.predictors,mod$y,xlab="Linear Predictor",ylab="Response"); xx <-
(min(mod$linear.predictors)-2):(2+max(mod$linear.predictors));lines(xx,xx,col=2)
lines(smooth.spline(mod$linear.predictors,mod$y,spar=1),col=4,lwd=2,lty=2)
savePlot("LN_G_Response_vs_linear_predictor",type="png")
```

```
#H. Include plots of the residuals by factor and time. Priority 1.
windows(height=10,width=10);par(mfrow=c(2,2))
b <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$latlons +
dat$vess,FUN=mean) ;names(b) <- c("yr","latlons","vess","y","res") # Pearson chi-square residuals
boxplot(b$res ~ b$yr,xlab="Year",ylab="Resids")
boxplot(b$res ~ b$latlons,xlab="Area",ylab="Resids")
boxplot(b$res ~ b$vess,xlab="Vessel",ylab="Resids")
plot(residuals(mod,type='pearson') ~ dat$covar,xlab="Covariate",ylab="resids")
savePlot("LN_H_Resids_by_variable",type="png")
```

```
windows(height=10,width=10);par(mfrow=c(2,2))
bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$latlons,FUN=mean)
;names(bb) <- c("yr","latlons","y","res") # Pearson chi-square residuals
plot(as.numeric(as.character(bb$yr)),bb$latlons,xlim=c(0.5,10.5),ylim=c(0.5,length(unique(bb$latlon
s))+0.5),type="n",xlab="Yr",ylab="Latlon",yaxt="n")
axis(2,at=unique(bb$latlons),labels=unique(bb$latlons))
a <- bb[bb$res >= 0,]
symbols(as.numeric(as.character(a$yr)),a$latlons,circles=a$res,add=T,fg=4, inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$latlons,circles=-a$res,add=T,fg=2, inches=F)
```

```
bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$vess,FUN=mean)
;names(bb) <- c("yr","vess","y","res") # Pearson chi-square residuals
plot(as.numeric(as.character(bb$yr)),bb$vess,xlim=c(0.5,10.5),ylim=c(0.5,length(unique(bb$vess))+
0.5),type="n",xlab="Yr",ylab="Vessel",yaxt="n")
axis(2,at=unique(bb$vess),labels=unique(bb$vess))
a <- bb[bb$res >= 0,]
```

```

symbols(as.numeric(as.character(a$yr)),a$vess,circles=2*a$res,add=T,fg=4, inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$vess,circles=-2*a$res,add=T,fg=2, inches=F)

dat$covar_b <- floor(dat$covar*10)/10
bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson'))) ~ dat$yr + dat$covar_b,FUN=mean)
;names(bb) <- c("yr","covar_b","y","res") # Pearson chi-square residuals
edge <- min(bb$covar_b)*0.05
plot(as.numeric(as.character(bb$yr)),bb$covar_b,xlim=c(0.5,10.5),ylim=c(min(bb$covar_b)-
edge,max(bb$covar_b) + edge),type="n",xlab="Yr",ylab="Covariate (binned)",yaxt="n")
axis(2,at=unique(bb$covar_b),labels=unique(bb$covar_b))
a <- bb[bb$res >= 0,]
symbols(as.numeric(as.character(a$yr)),a$covar_b,circles=2*a$res,add=T,fg=4, inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$covar_b,circles=-2*a$res,add=T,fg=2, inches=F)
savePlot("LN_H_bubble_plots_by_variable",type="png")

```

### #I Influence plots

```

graphics.off()
mod <- glm(log(datpar_ln[i,]) ~ year + latlons + vess + covar,data=dat)
infYear <- Influence$new(mod,data=dat)
infYear$calc()
infYear$summary
infYear$stanPlot()
savePlot("LN_I1_Influence_plots",type="png")
infYear$stepPlot()
savePlot("LN_I2_Influence_plots",type="png")
infYear$cdiPlot('vess')
savePlot("LN_I3_Influence_plots",type="png")
infYear$cdiPlot('latlons')
savePlot("LN_I4_Influence_plots",type="png")
infYear$cdiPlot('covar')
savePlot("LN_I5_Influence plots",type="png")
graphics.off()

```

### ##3.3. Poisson model

```

mod <- glm(datpar_po[i,] ~ yr + latlons + vess + covar,data=dat,family="poisson")

```

#A. Report overdispersion parameter and other fit statistics (e.g. chi-square / degrees of freedom).  
Priority 1.

```

pearson.deviance <- sum((mod$y-mod$fitted.values)^2/mod$fitted.values)
pearson.deviance / mod$df.residual

```

#B. Include plots describing error distribution (e.g. Studentized residuals vs. linear predictor. Priority 1.

```

windows(height=10,width=7);par(mfrow=c(2,1))
plot(mod$linear.predictors,studres(mod),xlab="Linear predictors",ylab="Studentized

```

```
residuals");lines(smooth.spline(mod$linear.predictors,studres(mod),spar=1),col=2)
```

```
b <- aggregate(cbind(mod$linear.predictors,mod$y,residuals(mod,type='pearson')) ~ dat$yr +  
dat$latlons + dat$vess,FUN=mean) ;names(b) <- c("yr","latlons","vess","linear.predictor","y","res")  
# Pearson chi-square residuals  
plot(b$linear.predictor,b$res,xlab="Linear predictors",ylab="Pearson  
residuals");lines(smooth.spline(b$linear.predictor,b$res,spar=1),col=2)  
savePlot("POI_B_Proportions_by_variable",type="png")
```

#C. Include QQ-plot – (e.g. Student deviance residuals vs. theoretical quantiles), Overlay expected distribution. Priority 1.

```
windows(height=7,width=7)  
plot(mod,which=2)  
savePlot("POI_C_Resids_vs_theor_quantiles",type="png")
```

#D. Include diagnostic plot for variance function (e.g. square root of std residuals vs. fitted values). Priority 1.Overlay expected distribution.

```
windows(height=7,width=7)  
plot(mod,which=3)  
savePlot("POI_D_sqrt_resids_vs_fitted",type="png")
```

#E. Include diagnostic plot for link function (e.g. linear response variable vs. linear predictor). Overlay expected distribution. The feasibility of this diagnostic is still under review. Priority 1.

```
windows(height=7,width=7)  
plot(mod$linear.predictors,mod$y,xlab="Linear predictors",ylab="Response");  
lines(smooth.spline(mod$linear.predictors,mod$y,spar=1),col=2)  
savePlot("POI_E_Response_vs_linear_predictor",type="png")
```

#F. Include plots of the residuals by factor and time. Priority 1.

```
windows(height=10,width=10);par(mfrow=c(2,2))  
b <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$latlons +  
dat$vess,FUN=mean) ;names(b) <- c("yr","latlons","vess","y","res") # Pearson chi-square residuals  
boxplot(b$res ~ b$yr,xlab="Year",ylab="Resids")  
boxplot(b$res ~ b$latlons,xlab="Area",ylab="Resids")  
boxplot(b$res ~ b$vess,xlab="Vessel",ylab="Resids")  
plot(residuals(mod,type='pearson') ~ dat$covar,xlab="Covariate",ylab="resids")  
savePlot("POI_F_Resids_by_variable",type="png")
```

```
windows(height=10,width=10);par(mfrow=c(2,2))  
bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$latlons,FUN=mean)  
;names(bb) <- c("yr","latlons","y","res") # Pearson chi-square residuals  
plot(as.numeric(as.character(bb$yr)),bb$latlons,xlim=c(0.5,10.5),ylim=c(0.5,length(unique(bb$latlon  
s))+0.5),type="n",xlab="Yr",ylab="Latlon",yaxt="n")  
axis(2,at=unique(bb$latlons),labels=unique(bb$latlons))  
a <- bb[bb$res >= 0,]  
symbols(as.numeric(as.character(a$yr)),a$latlons,circles=a$res,add=T,fg=4, inches=F)  
a <- bb[bb$res < 0,]  
symbols(as.numeric(as.character(a$yr)),a$latlons,circles=-a$res,add=T,fg=2, inches=F)
```



```

bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$vess,FUN=mean)
;names(bb) <- c("yr","vess","y","res") # Pearson chi-square residuals
plot(as.numeric(as.character(bb$yr)),bb$vess,xlim=c(0.5,10.5),ylim=c(0.5,length(unique(bb$vess))+
0.5),type="n",xlab="Yr",ylab="Vessel",yaxt="n")
axis(2,at=unique(bb$vess),labels=unique(bb$vess))
a <- bb[bb$res >= 0,]
symbols(as.numeric(as.character(a$yr)),a$vess,circles=2*a$res,add=T,fg=4, inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$vess,circles=-2*a$res,add=T,fg=2,inches=F)

```

```

dat$covar_b <- floor(dat$covar*10)/10
bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$covar_b,FUN=mean)
;names(bb) <- c("yr","covar_b","y","res") # Pearson chi-square residuals
edge <- min(bb$covar_b)*0.05
plot(as.numeric(as.character(bb$yr)),bb$covar_b,xlim=c(0.5,10.5),ylim=c(min(bb$covar_b)-
edge,max(bb$covar_b) + edge),type="n",xlab="Yr",ylab="Covariate (binned)",yaxt="n")
axis(2,at=unique(bb$covar_b),labels=unique(bb$covar_b))
a <- bb[bb$res >= 0,]
symbols(as.numeric(as.character(a$yr)),a$covar_b,circles=2*a$res,add=T,fg=4,inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$covar_b,circles=-2*a$res,add=T,fg=2,inches=F)
savePlot("POI_F_bubble_plots_by_variable",type="png")

```

```

#G Influence plots
graphics.off()
mod <- glm(datpar_po[i,] ~ year + latlons + vess + covar,data=dat,family="poisson")
infYear <- Influence$new(mod,data=dat)
infYear$calc()
infYear$summary
infYear$stanPlot() # Problem with year labels being resorted? Let Nokome know about bug.
savePlot("POI_G1_Influence_plots",type="png")
infYear$stepPlot()
savePlot("POI_G2_Influence_plots",type="png")
infYear$cdiPlot('vess')
savePlot("POI_G3_Influence_plots",type="png")
infYear$cdiPlot('latlons')
savePlot("POI_G4_Influence_plots",type="png")
infYear$cdiPlot('covar')
savePlot("POI_G5_Influence plots",type="png")
graphics.off()

```

### ##3.4. Zero-inflated model

```
dat$catch <- datpar_zinb[1,]
```

```

# A. Examine distribution of zeroes
cz <- as.numeric(dat$catch==0)
mod <- zeroinfl(catch ~ yr + latlons + vess + covar,data=dat,dist="negbin")
pred <- predict(mod,type="zero")
pred2 <- predict(mod,type="prob")[,1]

```

```

windows(9,12);par(mfrow=c(3,2))
hist(pred2,main="Probability of zero catch")
hist(pred,main="ZI component of p(zero catch)")
barplot(table(cz),names.arg=c("zero","nonzero"),main="Data distribution 1")
barplot(table(factor(dat$catch,levels=c(0:max(dat$catch))))),main="Data distribution 2")
a <- aggregate(cbind(cz,pred2) ~ yr + latlons + vess,data=dat,FUN=mean)
plot(a$cz,a$pred2,xlab="Observed nonzeros",ylab="Predictd
nonzeros");lines(smooth.spline(a$cz,a$pred2,spar=1),col=2)
savePlot("ZINB_A_zeroes",type="png")

```

#B. Include ROC curve to quantify goodness of fit. Priority 1.

```

roc.curve=function(s,print=FALSE){
  Ps=(pred2>s)*1
  FP=sum((Ps==1)*(cz==0))/sum(cz==0)
  TP=sum((Ps==1)*(cz==1))/sum(cz==1)
  if(print==TRUE){
    print(table(Observed=cz,Predicted=Ps))
  }
  vect=c(FP,TP)
  names(vect)=c("FPR","TPR")
  return(vect)
}
roc.curve(s=0.5,print=TRUE)
ROC.curve=Vectorize(roc.curve)
M.ROC=ROC.curve(seq(0,1,by=.01))
windows()
plot(M.ROC[1,],M.ROC[2,],col="grey",lwd=2,type="l",xlab="False positive rate",ylab="True
positive rate")
lines(c(0,1),c(0,1),lty=2,col=1)
savePlot("ZINB_B_ROC curve",type="png")

```

#C. Include plots describing error distribution (e.g. Residuals vs. linear predictor). Priority 1.

```

windows(height=11,width=7);par(mfrow=c(2,1))
linpred <- predict(mod,type="response")
plot(linpred,residuals(mod,type='pearson'),xlab="Linear predictor",ylab="Residuals",main="By
set");lines(smooth.spline(linpred,residuals(mod,type='pearson'),spar=1),col=2)
b <- aggregate(cbind(linpred,mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$latlons +
dat$vess,FUN=mean) ;names(b) <- c("yr","latlons","vess","linear.predictor","y","res") # Pearson chi-
square residuals
plot(b$linear.predictor,b$res,xlab="Linear predictor",ylab="Residuals",main="Aggregated by
category");lines(smooth.spline(b$linear.predictor,b$res,spar=1),col=2)
savePlot("ZINB_C_residuals_vs_linear_predictor",type="png")

```

# D Check for outliers

```

res <- resid(mod,type='pearson')
windows()
plot(res)
savePlot("ZINB_D_resid_outliers",type="png")

```

```

# E Plot residuals by variable
windows(height=10,width=10);par(mfrow=c(2,2))
b <- aggregate(cbind(mod$y,resid(mod,type='pearson')) ~ dat$yr + dat$latlons +
dat$vess,FUN=mean) ;names(b) <- c("yr","latlons","vess","y","res") # Pearson chi-square residuals
boxplot(b$res ~ b$yr,xlab="Year",ylab="Resids")
boxplot(b$res ~ b$latlons,xlab="Area",ylab="Resids")
boxplot(b$res ~ b$vess,xlab="Vessel",ylab="Resids")
plot(residuals(mod,type='pearson') ~ dat$covar,xlab="Covariate",ylab="resids")
savePlot("ZINB_E_resids_by_variable",type="png")
graphics.off()

windows(height=10,width=10);par(mfrow=c(2,2))
bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$latlons,FUN=mean)
;names(bb) <- c("yr","latlons","y","res") # Pearson chi-square residuals
plot(as.numeric(as.character(bb$yr)),bb$latlons,xlim=c(0.5,10.5),ylim=c(0.5,length(unique(bb$latlon
s))+0.5),type="n",xlab="Yr",ylab="Latlon",yaxt="n")
axis(2,at=unique(bb$latlons),labels=unique(bb$latlons))
a <- bb[bb$res >= 0,]
symbols(as.numeric(as.character(a$yr)),a$latlons,circles=a$res,add=T,fg=4, inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$latlons,circles=-a$res,add=T,fg=2,inches=F)

bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$vess,FUN=mean)
;names(bb) <- c("yr","vess","y","res") # Pearson chi-square residuals
plot(as.numeric(as.character(bb$yr)),bb$vess,xlim=c(0.5,10.5),ylim=c(0.5,length(unique(bb$vess))+
0.5),type="n",xlab="Yr",ylab="Vessel",yaxt="n")
axis(2,at=unique(bb$vess),labels=unique(bb$vess))
a <- bb[bb$res >= 0,]
symbols(as.numeric(as.character(a$yr)),a$vess,circles=2*a$res,add=T,fg=4,inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$vess,circles=-2*a$res,add=T,fg=2,inches=F)

dat$covar_b <- floor(dat$covar*10)/10
bb <- aggregate(cbind(mod$y,residuals(mod,type='pearson')) ~ dat$yr + dat$covar_b,FUN=mean)
;names(bb) <- c("yr","covar_b","y","res") # Pearson chi-square residuals
edge <- min(bb$covar_b)*0.05
plot(as.numeric(as.character(bb$yr)),bb$covar_b,xlim=c(0.5,10.5),ylim=c(min(bb$covar_b)-
edge,max(bb$covar_b) + edge),type="n",xlab="Yr",ylab="Covariate (binned)",yaxt="n")
axis(2,at=unique(bb$covar_b),labels=unique(bb$covar_b))
a <- bb[bb$res >= 0,]
symbols(as.numeric(as.character(a$yr)),a$covar_b,circles=2*a$res,add=T,fg=4,inches=F)
a <- bb[bb$res < 0,]
symbols(as.numeric(as.character(a$yr)),a$covar_b,circles=-2*a$res,add=T,fg=2,inches=F)
savePlot("ZINB_E_bubble_plots_by_variable",type="png")

#B. Plot progressive changes as more parameters are added. Priority 1.
newdat <- expand.grid(yr=sort(unique(dat$yr)),latlons=latlons[5],vess=vess[2],covar=1.5)
mod1 <- zeroinfl(catch ~ yr,data=newdat,dist="negbin")
mod2 <- zeroinfl(catch ~ yr + latlons,data=newdat,dist="negbin")

```

```
mod3 <- zeroinfl(catch ~ yr + latlons + vess,data=dat,dist="negbin")
mod4 <- zeroinfl(catch ~ yr + latlons + vess + covar,data=dat,dist="negbin")
newdat$par1 <- predict(mod1,newdata=newdat,type="response")
newdat$par2 <- predict(mod2,newdata=newdat,type="response")
newdat$par3 <- predict(mod3,newdata=newdat,type="response")
newdat$par4 <- predict(mod4,newdata=newdat,type="response")
plot(as.numeric(newdat$yr),newdat$par1,type="n",xlab="Year",ylab="Predicted
response",ylim=c(0,2.5))
lines(newdat$yr,newdat$par1,type="b",col=1,lty=1,pch=1)
lines(newdat$yr,newdat$par2,type="b",col=2,lty=1,pch=2)
lines(newdat$yr,newdat$par3,type="b",col=3,lty=1,pch=3)
lines(newdat$yr,newdat$par4,type="b",col=4,lty=1,pch=4)
legend("topleft",legend=c("Year only","Year + latlons","Year + latlons + vessel","Year + latlons +
vessel + covariate"),pch=1:4,lty=1,col=1:4)
savePlot("ZINB_F_Influence of each variable",type="png")
```



## D. REVIEW OF CPUE INDEX PAPERS FROM RECENT ASSESSMENTS

In the following table we examine a range of papers that describe CPUE indices used in recent stock assessments submitted to the WCOFC, and indicate whether we found the information suggested in the proposed list. We have chosen one or two papers for each species. For species where the indices in the assessment come from multiple papers, we have chosen the paper that provides the most influential indices used in the assessment.

In some cases where we did not find listed information in the paper, it is available in other papers. We have only indicated that the information has been provided if it is in the current paper, or if an appropriate source is directly referred to. Information can be provided at various levels of detail, comprehensiveness, and accuracy. The question whether the information provided is sufficient is often subjective. We have not judged the quality of the indices or the credibility of the information provided, noting only whether we found relevant information.

Yellowfin and bigeye tuna.

Assessments: YFT: WCPFC-SC7-2011/SA-WP-04 (Langley *et al.* 2011), BET: WCPFC-SC7-2011/SA-WP-02 (Davies *et al.* 2011)

CPUE papers: WCPFC -SC7-2011/SA-IP-01 (Hoyle and Okamoto 2011)

North Pacific blue shark:

Assessments: WCPFC-SC9-2013/SA-WP-11-ISC-Blue-Shark-NP-SA (ISC Shark Working Group 2013).

CPUE papers: ISC12-01-09, ISC13-02-02 (Hiraoka *et al.* 2012; Hiraoka *et al.* 2013).

Skipjack tuna:

Assessments: WCPFC-SC7-2011/SA-WP-04 (Hoyle *et al.* 2011)

CPUE papers: WCPFC-SC6-2010/SA-WP-08, WCPFC-SC7-2011/SA-IP-13 (Kiyofuji *et al.* 2011; Langley *et al.* 2010)

South Pacific albacore:

Assessments: WCPFC-SC8-2012/SA-WP-04 (Hoyle 2012)

CPUE papers: WCPFC-SC8-2012/SA-IP-14 (Bigelow and Hoyle 2012)

Southwest Pacific Swordfish:

Assessments: WCPFC-SC9-2013/SA-WP-05 (Davies *et al.* 2013)

CPUE papers: WCPFC-SC9-2013/SA-IP-03 (Hoyle *et al.* 2013)

Southwest Pacific striped marlin:

Assessments: WCPFC-SC8-2012/SA-WP-05 (Davies *et al.* 2012)

CPUE papers: WCPFC-SC8-2012/SA-IP-09 (Hoyle *et al.* 2012)

WCPO Silky shark:

Assessments: WCPFC-SC9-2013/ SA-WP-03 (Rice and Harley 2013)

CPUE papers: WCPFC-SC8-2012/ SA-IP-11, WCPFC-SC9-2013/SA-IP-02 (Rice 2012; Rice 2013)

Table 1: For selected stock assessments presented to the WCPFC, and the CPUE papers associated with those assessments that describe the most influential indices, the degree to which they supply the information listed above. Y= Yes, N=No, P=Partly, DE=Described Elsewhere, NA=Not Relevant.

Item	Priority	YFT & BET	NP BSH	SKJ	SP ALB	SWO	MLS	Silky
<b>1.1</b>	<b>A</b>	1 P	Y	P	P	P	P	Y
	<b>B</b>	1 Y	P	Y	Y	Y	Y	Y
	<b>C</b>	2 DE	DE	DE	N	N	N	N
	<b>D</b>	2 N	N	N	N	N	N	N
	<b>E</b>	2 N	N	N	N	N	N	N
<b>1.2</b>	<b>A</b>	1 Y	Y	Y	Y	Y	Y	Y
	<b>B</b>	1 Y	Y	Y	Y	Y	Y	Y
	<b>C</b>	1 Y	Y	Y	N	Y	Y	P
	<b>D</b>	1 Y	P	Y	Y	Y	Y	P
	<b>E</b>	1 DE	N	N	N	N	N	N
	<b>F</b>	1 DE	N	N	N	N	N	N
	<b>G</b>	2 P	N	N	N	N	N	N
	<b>H</b>	2 NA	P	NA	NA	NA	NA	NA
	<b>I</b>	3 N	Y	N	N	N	N	N
<b>2.1</b>	<b>A</b>	1 Y	Y	P	Y	P	P	Y
	<b>B</b>	1 Y	N	N	Y	P	P	Y
	<b>C</b>	1 Y	NA	Y	Y	Y	P	Y
	<b>D</b>	1 Y	Y	Y	Y	Y	Y	Y
	<b>E</b>	2 P	N	N	Y	N	N	Y
<b>2.2</b>	<b>A</b>	1 Y	N	Y	N	Y	Y	N
	<b>B</b>	2 N	N	N	P	N	N	P
	<b>C</b>	1 NA	Y	NA	Y	NA	NA	Y
	<b>D</b>	1 DE	DE?	DE	DE	DE	DE	DE
	<b>E</b>	2 Y	N	P	N	N	N	Y
	<b>F</b>	1 Y	N	N	N	P	P	P
	<b>G</b>	1 Y	N	N	N	N	N	Y
	<b>H</b>	2 N	N	N	N	Y	Y	Y
	<b>I</b>	1 Y	Y	Y	N	N	N	Y
	<b>J</b>	1 Y	Y	Y	Y	Y	Y	Y
	<b>K</b>	1 Y	Y	Y	Y	Y	Y	Y
	<b>L</b>	1 Y	Y	Y	Y	Y	Y	Y
	<b>M</b>	2 DE	N	N	N	N	N	N
<b>2.3</b>	<b>A</b>	1 Y	Y	Y	Y	Y	Y	Y
	<b>B</b>	1 Y	P	Y	Y	Y	Y	Y
	<b>C</b>	1 Y	Y	Y	Y	Y	Y	Y
	<b>D</b>	1 Y	Y	Y	Y	Y	Y	Y
	<b>E</b>	1 Y	N	N	N	N	N	N
	<b>F</b>	1 N	N	N	N	N	N	N

<b>G</b>	1	Y	Y	Y	Y	Y	Y	Y
<b>H</b>	2	N	N	Y	N	Y	Y	N
<b>I</b>	1	Y	P	Y	Y	Y	Y	Y
<b>J</b>	2	N	Y	N	N	N	N	N
<b>K</b>	1	Y	Y	N	Y	N	N	Y
<b>L</b>	2	N	N	N	N	N	N	N
<b>M</b>	2	N	P	N	N	N	N	Y
<b>N</b>	1	Y	Y	Y	Y	Y	Y	Y
<b>O</b>	1	N	NA	Y	NA	NA	NA	N
<b>P</b>	1	N	Y	P	N	N	N	N
<b>Q</b>	1	NA	Y	NA	N	NA	NA	Y
<b>R</b>	1	Y	Y	N	Y	Y	Y	P
<b>S</b>	2	NA	N	Y	N	N	N	N
<b>3.1 A</b>	2	N	N	N	N	N	N	N
<b>B</b>	2	N	N	N	N	N	N	N
<b>3.2 A</b>	2	N	NA	N	NA	NA	NA	NA
<b>B</b>	2	N	NA	N	NA	NA	NA	NA
<b>C</b>	2	N	NA	N	NA	NA	NA	NA
<b>D</b>	1	N	NA	P	NA	NA	NA	NA
<b>3.3 A</b>	1	Y	NA	N	NA	Y	Y	NA
<b>B</b>	2	N	NA	N	NA	N	N	NA
<b>C</b>	1	Y	NA	P	NA	P	P	NA
<b>D</b>	2	N	NA	N	NA	N	N	NA
<b>E</b>	1	Y	NA	N	NA	Y	Y	NA
<b>F</b>	2	N	NA	N	NA	N	N	NA
<b>G</b>	3	N	NA	N	NA	N	N	NA
<b>H</b>	2	N	NA	P	NA	N	N	NA
<b>I</b>	1	N	NA	N	NA	N	N	NA
<b>3.4 A</b>	2	NA	Y	NA	N	NA	NA	NA
<b>B</b>	1	NA	Y	NA	N	NA	NA	NA
<b>C</b>	1	NA	Y	NA	N	NA	NA	NA
<b>D</b>	2	NA	Y	NA	N	NA	NA	NA
<b>E</b>	3	NA	N	NA	N	NA	NA	NA
<b>F</b>	1	NA	N	NA	N	NA	NA	NA
<b>G</b>	1	NA	N	NA	N	NA	NA	NA
<b>3.5 A</b>	1	NA	NA	NA	NA	NA	NA	N
<b>B</b>	1	NA	NA	NA	NA	NA	NA	N
<b>C</b>	2	NA	NA	NA	NA	NA	NA	Y
<b>D</b>	1	NA	NA	NA	NA	NA	NA	Y
<b>E</b>	1	NA	NA	NA	NA	NA	NA	Y
<b>F</b>	1	NA	NA	NA	NA	NA	NA	P



<b>4</b>	<b>A</b>	1	Mostly	Mostly	Mostly	N	Y	Y	Y
	<b>B</b>	1	P	Y	Y	P	Y	Y	Y
<b>5</b>	<b>A</b>	1	NA	N	Y	NA	NA	NA	N
	<b>B</b>	1	NA	Y	N	NA	NA	NA	N
	<b>C</b>	1	NA	N	P	NA	NA	NA	Y

## References

- Bigelow, K.A.; Hoyle, S.D. Standardized CPUE for South Pacific albacore. 2012
- Davies, N.; Hoyle, S.D.; Hampton, W.J. Stock assessment of striped marlin (*Kajikia audax*) in the southwest Pacific ocean, WCPFC-SC8-2012/SA-WP-05. WCPFC Scientific Committee, Ninth Regular Session, 7-15 August 2012, Busan, Republic of Korea; 2012
- Davies, N.; Hoyle, S.D.; Harley, S.J.; Langley, A.D.; Kleiber, P.M.; Hampton, W.J. Stock Assessment Of Bigeye Tuna In The Western And Central Pacific Ocean. 2011
- Davies, N.; Pilling, G.M.; Harley, S.J.; Hampton, W.J. Stock assessment of swordfish (*Xiphias gladius*) in the southwest Pacific ocean, WCPFC-SC9-2013/SA-WP-05. WCPFC Scientific Committee, Ninth Regular Session, 6-14 August 2013, Pohnpei, Federated States of Micronesia; 2013
- Hiraoka, Y.; Kanaiwa, M.; Yokawa, K. Estimation of abundance indices for blue shark in the North Pacific. 2012
- Hiraoka, Y.; Kanaiwa, M.; Yokawa, K. Summary of estimation process of abundance indices for blue shark in the North Pacific. 2013
- Hoyle, S.; Davies, N.; Chang, S.-K. Analysis of swordfish catch per unit effort data for Japanese and Chinese Taipei longline fleets in the southwest Pacific Ocean, WCPFC-SC9-2013/SA-IP-03. WCPFC Scientific Committee, Ninth Regular Session, 7-15 August 2012, Busan, Republic of Korea; 2013
- Hoyle, S.; Kleiber, P.; Davies, N.; Langley, A.D.; Hampton, J. Stock assessment of skipjack tuna in the western and central Pacific Ocean, WCPFC-SC-7-SA/WP-04. WCPFC Scientific Committee Regular Session, Pohnpei, FSM, 9-17 August 2011, 7th; 2011
- Hoyle, S.D. Stock assessment of albacore tuna in the South Pacific ocean. Scientific Committee, Eighth Regular Session, 7-15 August 2012, Busan, Republic of Korea; 2012
- Hoyle, S.D.; Davies, N.; Pilling, G.M. CPUE standardisation for Striped Marlin in the Western and Central Pacific Ocean, WCPFC-SC8-2012/SA-IP-09. WCPFC Scientific Committee Sixth Regular Session, 7-15 August 2012, Busan, Republic of Korea; 2012
- Hoyle, S.D.; Okamoto, H. Analyses of Japanese longline operational catch and effort for bigeye and yellowfin tuna in the WCPO. Pohnpei, Federated States of Micronesia; 2011
- ISC Shark Working Group. Stock assessment and future projections of blue shark in the north Pacific ocean. WCPFC-SC9-2013/SA-WP-11. WCPFC-SC, Pohnpei, Federated States of Micronesia, 6-14 August; 2013
- Kiyofuji, H.; Uosaki, K.; Hoyle, S. Up-to-date CPUE for skipjack caught by Japanese distant and offshore pole and line in the western central Pacific Ocean. WCPFC-SC7-SA-IP-13; 2011
- Langley, A.; Uosaki, K.; Hoyle, S.; Shono, H.; Ogura, M. A standardized CPUE analysis of the Japanese distant-water skipjack pole-and-line fishery in the western and central Pacific Ocean (WCPO), 1972-2009. 2010
- Langley, A.D.; SD, H.; Hampton, W.J. Stock assessment for yellowfin tuna in the WCPO in 2011, WCPFC-SC-7-SA/WP-03. . Scientific Committee Regular Session, Pohnpei, FSM, 9-17 August 2011, 7th; 2011; 2011
- Rice, J. Catch per unit effort of silky sharks in the Western and Central Pacific Ocean. WCPFCSC8-SA-IP-11; 2012
- Rice, J. Catch and catch per unit effort of silky sharks in the Western and Central Pacific Ocean, WCPFC-SC9-2013/SA-IP-02. WCPFC Scientific Committee, Ninth Regular Session, 6-14 August 2013, Pohnpei, Federated States of Micronesia: WCPFCSC8-SA-IP-11; 2013
- Rice, J.; Harley, S. Updated Stock Assessment of Silky Sharks in the Western And Central Pacific Ocean, WCPFC-SC9-2013/SA-WP-03. . WCPFC Scientific Committee, Ninth Regular

Session, Pohnpei, Federated States of Micronesia, 6-14 August 2013. ; 2013  
SEDAR. Abundance Indices Workshop: Developing protocols for submission of abundance indices to the SEDAR process.  
[http://www.sefsc.noaa.gov/sedar/download/Abundance%20Indices%20Workshop%20Document\\_final.pdf?id=DOCUMENT](http://www.sefsc.noaa.gov/sedar/download/Abundance%20Indices%20Workshop%20Document_final.pdf?id=DOCUMENT): SEDAR Procedures Workshop 1, Oct 14-17, 2008, Miami, Florida, USA; 2009