Australian Government
Bureau of Rural Sciences

Work requested 23 June 2010

Veronica Boero Rodriguez

June 2010

ISSN obtain from Bureaux Communications Team

and/or

ISBN obtain from Bureaux Communications Team

Boero Rodriguez, V 2010, *Southern Bluefin Tuna. Exploration of catch per unit effort standardisation models*, Bureau of Rural Sciences, Canberra, March.

Postal address:
Bureau of Rural Sciences
GPO Box 858
Canberra, ACT 2601

Ph: 1800 020 157
Fax: 02 6272 2330
Email: info@brs.gov.au
Internet: http://www.brs.gov.au

Table of contents

Figures

.

Models

Six models were fitted and their residuals diagnostics, AIC and BIC compared; as well as the effect they produced in the standardised series compared to the raw series.

- Model 1: year + month + area + Lat5
- Model 2: year + month + area + Lat5 + BET_CPUE + YFT_CPUE
- Model 3: year + month + area + Lat5 + BET_CPUE + YFT_CPUE + month*area
- Model 4: year + month + area + Lat5 + BET_CPUE + YFT_CPUE + year*Lat5
- Model 5: year + month + area + Lat5 + BET_CPUE + YFT_CPUE + year*area
- Model 6: year+month+area+Lat5+BET_CPUE+YFT_CPUE+month*area+year*Lat5+year*area

All models have similar fit in terms of residuals diagnostics since the residual versus fitted values and residuals qq-plots are fairly similar for all models. However, model 1 residual plot shows two clear groups and models 4 and 6 residuals qq-plots present slightly more curvature than the other models.

Model 6 has the lowest AIC while model 3 has the lowest BIC. This is because BIC penalise more heavily the number of parameter in the model than AIC.

Figure 1: Residuals diagnostics - model 1

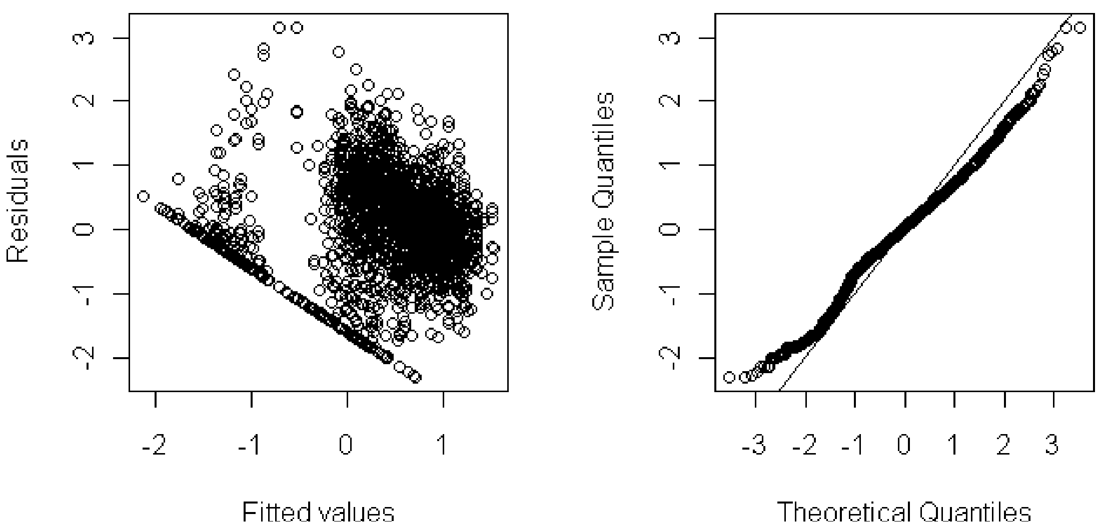


Figure 2: Residuals diagnostics – model 2

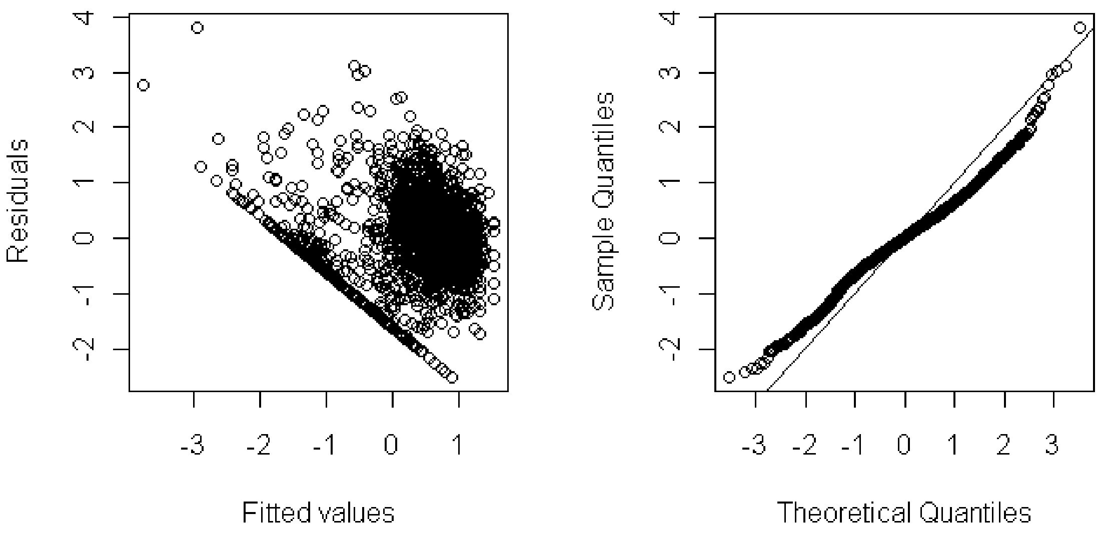


Figure 3: Residuals diagnostics – model 3

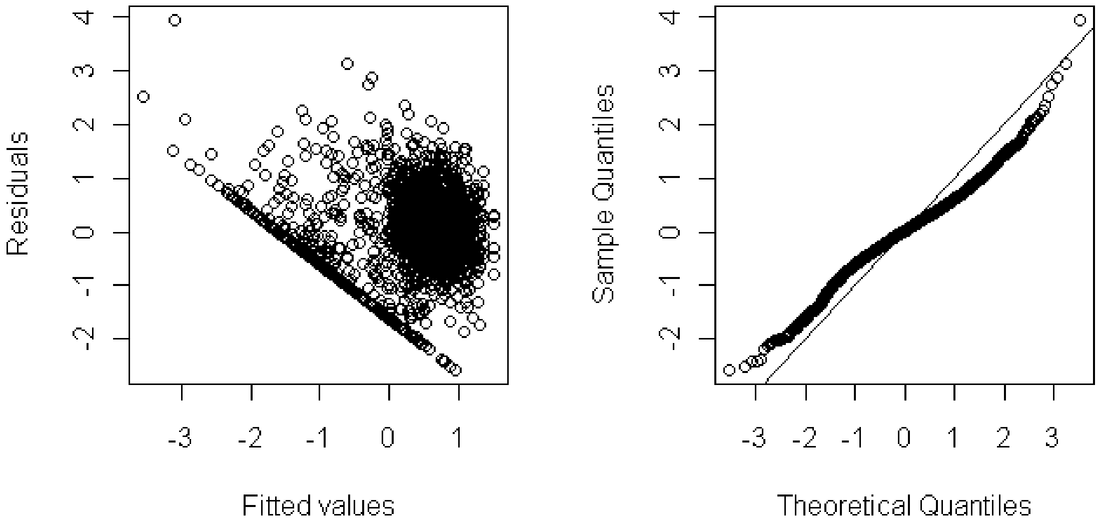


Figure 4: Residuals diagnostics – model 4

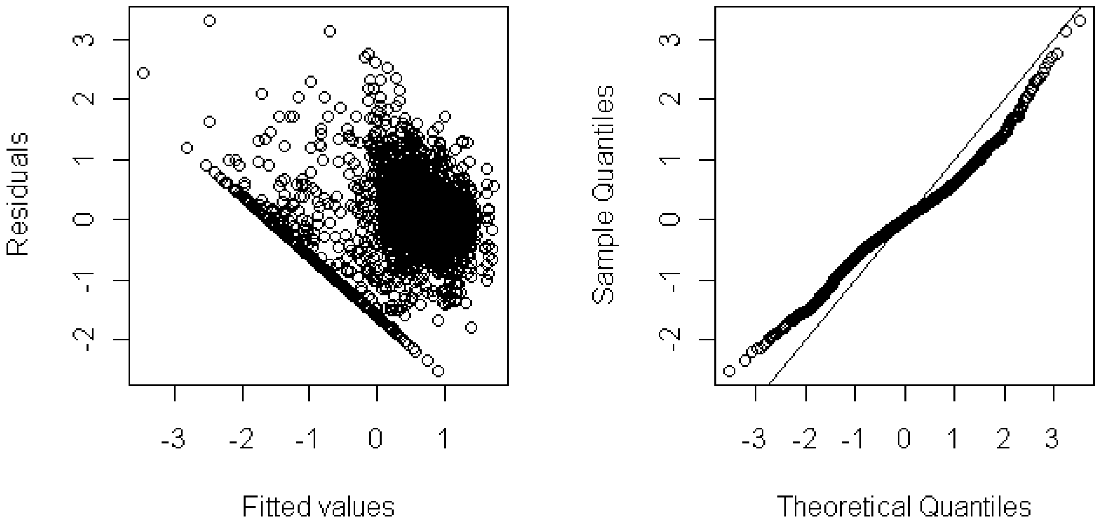


Figure 5: Residuals diagnostics – model 5

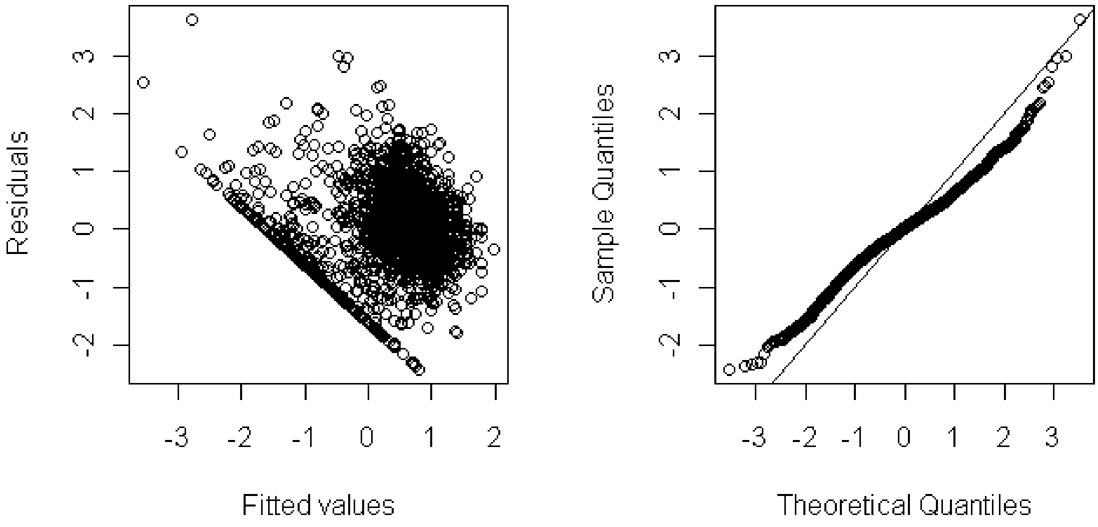


Figure 6: Residuals diagnostics – model 6

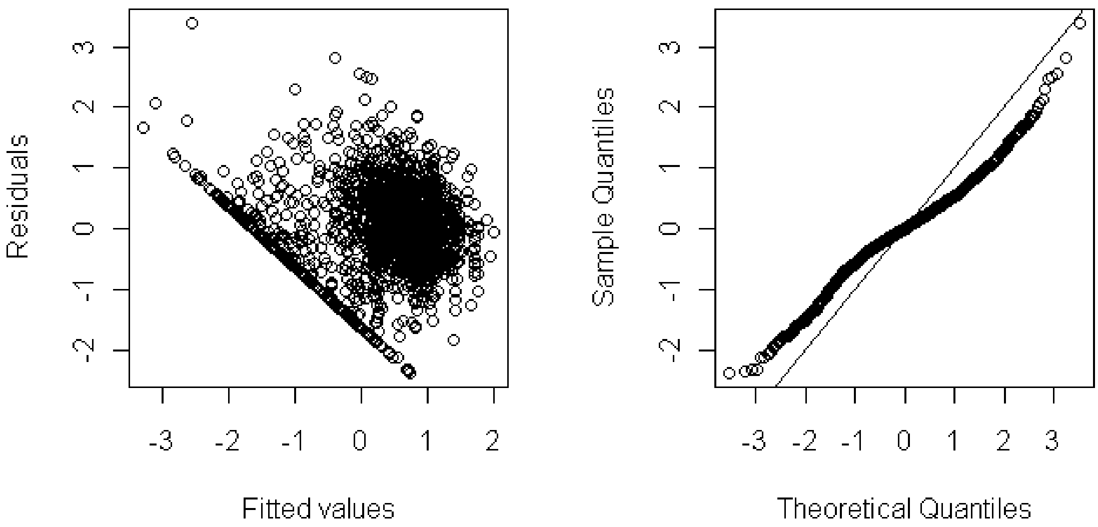
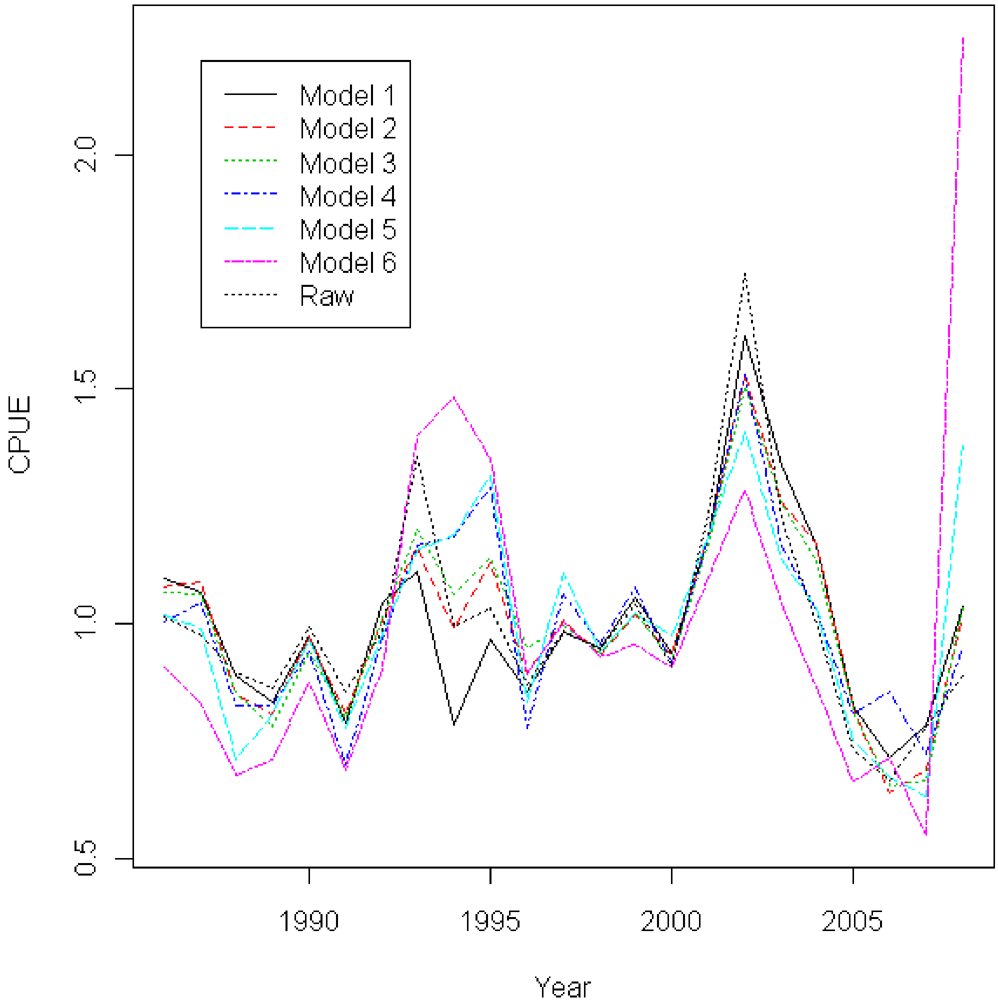


Table 1: Models AIC and BIC

	AIC	BIC
Model 1	5558	5765
Model 2	5191	5409
Model 3	5064	5397
Model 4	5158	5756
Model 5	5182	5907
Model 6	**4939**	**6158**

Figure 7: CPUE series 1986-2008 – models 1 to 6 and raw series.



The year by area interaction seems to be causing the exacerbation of the trend change as that change increases the most when that interaction is added (model 5) and increases even further when all interaction are added (model 6).

Conclusions

When taking into account model fit, AIC, BIC and standardised series smoothness together, model 3 seems to be the most appropriate model because:

- All models have similar fit in terms of residuals diagnostics although model 1, 4 and 6 present slightly worse diagnostics.
- Model 3 has the lowest BIC and the second lowest AIC.
- Model 3 series is among the smoothest.

Appendix A

R code

```
#set the appropriate directory
setwd("P:\\Cross_Programme\\FisheryStats\\Southern Bluefin Tuna\\SBT CPUE Standardisation")
#libraries needed
library(nlme)
library(MASS)
#read the data
#RTMP data
RTMPdata<-read.csv("5x5core02112009RTMP.csv")
#delta constant
log.add <- 0.2
#factor variables
RTMPdata$year<-factor(RTMPdata$Year)
RTMPdata$month<-factor(RTMPdata$Month)
RTMPdata$area<-factor(RTMPdata$Area)
RTMPdata$lat5<-factor(RTMPdata$Lat5)
#create 5x5 square variable
RTMPdata$sq<-paste(RTMPdata$Lat5,RTMPdata$Lon5,sep="")
RTMPdata$sq<-factor(RTMPdata$sq)
#raw cpue RTMP
cpue.year<-aggregate(RTMPdata$CPUE,by=list(RTMPdata$year),mean)
cpue.r<-cpue.year$x/mean(cpue.year$x)
yearv<-c(seq(from=1986,to=2008,by=1))
#model 1
RTMPm1<-lm(logCPUE~year+month+area+lat5, data=RTMPdata)
#check residuals
par(mfrow=c(1,2), pty="s")
plot(RTMPm1$fitted,RTMPm1$resid, xlab="Fitted values", ylab="Residuals")
qqnorm(RTMPm1$resid, main=" ")
abline(a=0,b=1)
#constructing the standardised series
#create the matrix for prediction and convert NAs to zeros
pred1<-
expand.grid(year=levels(RTMPdata$year),month=levels(RTMPdata$month),area=levels(RTMPdat
a$area),lat5=levels(RTMPdata$lat5))
#calculate predictions
pred1$cpue<-exp(predict(RTMPm1,newdata=pred1)+((0.7953)^2)/2)-log.add
#calculate unweighted series
cpue<-aggregate(pred1$cpue,by=list(pred1$year),mean)
cpueUWm1<-cpue$x/mean(cpue$x)
```

```
#model 2
RTMPm2<-lm(logCPUE~year+month+area+lat5+BETcpue5+YFTcpue5, data=RTMPdata)
par(mfrow=c(1,2), pty="s")
plot(RTMPm2$fitted,RTMPm2$resid, xlab="Fitted values", ylab="Residuals")
qqnorm(RTMPm2$resid, main=" ")
abline(a=0,b=1)
BETav<-aggregate(RTMPdata$BETcpue5, by=list(RTMPdata$year,RTMPdata$month,
RTMPdata$area, RTMPdata$lat5), mean)
BET <- data.frame(
   year = BETav$Group.1,
   month = BETav$Group.2,
   area = BETav$Group.3,
   lat5 = BETav$Group.4,
   BETcpue5 = BETav$x)
YFTav<-aggregate(RTMPdata$YFTcpue5, by=list(RTMPdata$year,RTMPdata$month,
RTMPdata$area, RTMPdata$lat5), mean)
YFT <- data.frame(
   year = YFTav$Group.1,
   month = YFTav$Group.2,
   area = YFTav$Group.3,
   lat5 = YFTav$Group.4,
   YFTcpue5 = YFTav$x)
pred_data<-
expand.grid(year=levels(RTMPdata$year),month=levels(RTMPdata$month),area=levels(RTMPdat
a$area),lat5=levels(RTMPdata$lat5))
pred<-merge(pred_data,BET, all=T)
pred<-merge(pred,YFT, all=T)
for(i in 1:length(pred$BETcpue5)){
if(is.na(pred$BETcpue5[i])) pred$BETcpue5[i]<-0
if(is.na(pred$YFTcpue5[i])) pred$YFTcpue5[i]<-0}
pred$cpue<-exp(predict(RTMPm2,newdata=pred)+((0.7345)^2)/2)-log.add
cpue<-aggregate(pred$cpue,by=list(pred$year),mean)
cpueUW2<-cpue$x/mean(cpue$x)

#model 3
RTMPm3<-lm(logCPUE~year+month+area+lat5+BETcpue5+YFTcpue5+month*area,
data=RTMPdata)
par(mfrow=c(1,2), pty="s")
plot(RTMPm3$fitted,RTMPm3$resid, xlab="Fitted values", ylab="Residuals")
qqnorm(RTMPm3$resid, main=" ")
abline(a=0,b=1)
BETav<-aggregate(RTMPdata$BETcpue5, by=list(RTMPdata$year,RTMPdata$month,
RTMPdata$area, RTMPdata$lat5), mean)
BET <- data.frame(
```

```
   year = BETav$Group.1,
   month = BETav$Group.2,
   area = BETav$Group.3,
   lat5 = BETav$Group.4,
   BETcpue5 = BETav$x)
YFTav<-aggregate(RTMPdata$YFTcpue5, by=list(RTMPdata$year,RTMPdata$month,
RTMPdata$area, RTMPdata$lat5), mean)
YFT <- data.frame(
   year = YFTav$Group.1,
   month = YFTav$Group.2,
   area = YFTav$Group.3,
   lat5 = YFTav$Group.4,
   YFTcpue5 = YFTav$x)
pred_data<-
expand.grid(year=levels(RTMPdata$year),month=levels(RTMPdata$month),area=levels(RTMPdat
a$area),lat5=levels(RTMPdata$lat5))
pred<-merge(pred_data,BET, all=T)
pred<-merge(pred,YFT, all=T)
for(i in 1:length(pred$BETcpue5)){
if(is.na(pred$BETcpue5[i])) pred$BETcpue5[i]<-0
if(is.na(pred$YFTcpue5[i])) pred$YFTcpue5[i]<-0}
pred$cpue<-exp(predict(RTMPm3,newdata=pred)+((0.7116)^2)/2)-log.add
cpue<-aggregate(pred$cpue,by=list(pred$year),mean)
cpueUW3<-cpue$x/mean(cpue$x)

#model 4
RTMPm4<-lm(logCPUE~year+month+area+lat5+BETcpue5+YFTcpue5+year*lat5,
data=RTMPdata)
par(mfrow=c(1,2), pty="s")
plot(RTMPm4$fitted,RTMPm4$resid, xlab="Fitted values", ylab="Residuals")
qqnorm(RTMPm4$resid, main=" ")
abline(a=0,b=1)
BETav<-aggregate(RTMPdata$BETcpue5, by=list(RTMPdata$year,RTMPdata$month,
RTMPdata$area, RTMPdata$lat5), mean)
BET <- data.frame(
   year = BETav$Group.1,
   month = BETav$Group.2,
   area = BETav$Group.3,
   lat5 = BETav$Group.4,
   BETcpue5 = BETav$x)
YFTav<-aggregate(RTMPdata$YFTcpue5, by=list(RTMPdata$year,RTMPdata$month,
RTMPdata$area, RTMPdata$lat5), mean)
YFT <- data.frame(
   year = YFTav$Group.1,
```

```
   month = YFTav$Group.2,
   area = YFTav$Group.3,
   lat5 = YFTav$Group.4,
   YFTcpue5 = YFTav$x)
pred_data<-
expand.grid(year=levels(RTMPdata$year),month=levels(RTMPdata$month),area=levels(RTMPdat
a$area),lat5=levels(RTMPdata$lat5))
pred<-merge(pred_data,BET, all=T)
pred<-merge(pred,YFT, all=T)
for(i in 1:length(pred$BETcpue5)){
if(is.na(pred$BETcpue5[i])) pred$BETcpue5[i]<-0
if(is.na(pred$YFTcpue5[i])) pred$YFTcpue5[i]<-0}
pred$cpue<-exp(predict(RTMPm4,newdata=pred)+((0.7194)^2)/2)-log.add
cpue<-aggregate(pred$cpue,by=list(pred$year),mean)
cpueUW4<-cpue$x/mean(cpue$x)

#model 5
RTMPm5<-lm(logCPUE~year+month+area+lat5+BETcpue5+YFTcpue5+year*area,
data=RTMPdata)
par(mfrow=c(1,2), pty="s")
plot(RTMPm5$fitted,RTMPm5$resid, xlab="Fitted values", ylab="Residuals")
qqnorm(RTMPm5$resid, main=" ")
abline(a=0,b=1)
BETav<-aggregate(RTMPdata$BETcpue5, by=list(RTMPdata$year,RTMPdata$month,
RTMPdata$area, RTMPdata$lat5), mean)
BET <- data.frame(
   year = BETav$Group.1,
   month = BETav$Group.2,
   area = BETav$Group.3,
   lat5 = BETav$Group.4,
   BETcpue5 = BETav$x)
YFTav<-aggregate(RTMPdata$YFTcpue5, by=list(RTMPdata$year,RTMPdata$month,
RTMPdata$area, RTMPdata$lat5), mean)
YFT <- data.frame(
   year = YFTav$Group.1,
   month = YFTav$Group.2,
   area = YFTav$Group.3,
   lat5 = YFTav$Group.4,
   YFTcpue5 = YFTav$x)
pred_data<-
expand.grid(year=levels(RTMPdata$year),month=levels(RTMPdata$month),area=levels(RTMPdat
a$area),lat5=levels(RTMPdata$lat5))
pred<-merge(pred_data,BET, all=T)
pred<-merge(pred,YFT, all=T)
for(i in 1:length(pred$BETcpue5)){
```

```
if(is.na(pred$BETcpue5[i])) pred$BETcpue5[i]<-0
if(is.na(pred$YFTcpue5[i])) pred$YFTcpue5[i]<-0}
pred$cpue<-exp(predict(RTMPm5,newdata=pred)+((0.7199)^2)/2)-log.add
cpue<-aggregate(pred$cpue,by=list(pred$year),mean)
cpueUW5<-cpue$x/mean(cpue$x)

#model 6
RTMPm6<-
lm(logCPUE~year+month+area+lat5+BETcpue5+YFTcpue5+month*area+year*lat5+year*area,
data=RTMPdata)
#check residuals
par(mfrow=c(1,2), pty="s")
plot(RTMPm6$fitted,RTMPm6$resid, xlab="Fitted values", ylab="Residuals")
qqnorm(RTMPm6$resid, main=" ")
abline(a=0,b=1)
#constructing the standardised series
#calculate mean BET cpue for all combinations of year, month, area and Lat5 levels
BETav<-aggregate(RTMPdata$BETcpue5, by=list(RTMPdata$year,RTMPdata$month,
RTMPdata$area, RTMPdata$lat5), mean)
BET <- data.frame(
  year = BETav$Group.1,
  month = BETav$Group.2,
  area = BETav$Group.3,
  lat5 = BETav$Group.4,
  BETcpue5 = BETav$x)
#as per BET for YFT
YFTav<-aggregate(RTMPdata$YFTcpue5, by=list(RTMPdata$year,RTMPdata$month,
RTMPdata$area, RTMPdata$lat5), mean)
YFT <- data.frame(
  year = YFTav$Group.1,
  month = YFTav$Group.2,
  area = YFTav$Group.3,
  lat5 = YFTav$Group.4,
  YFTcpue5 = YFTav$x)
#create the matrix for prediction and convert NAs to zeros
pred_data<-
expand.grid(year=levels(RTMPdata$year),month=levels(RTMPdata$month),area=levels(RTMPdat
a$area),lat5=levels(RTMPdata$lat5))
pred<-merge(pred_data,BET, all=T)
pred<-merge(pred,YFT, all=T)
for(i in 1:length(pred$BETcpue5)){
if(is.na(pred$BETcpue5[i])) pred$BETcpue5[i]<-0
if(is.na(pred$YFTcpue5[i])) pred$YFTcpue5[i]<-0}
#calculate predictions
```

```
pred$cpue<-exp(predict(RTMPm6,newdata=pred)+((0.6715)^2)/2)-log.add
#calculate unweighted series
cpue<-aggregate(pred$cpue,by=list(pred$year),mean)
cpueUW6<-cpue$x/mean(cpue$x)

#plot all series
par(mfrow=c(1,1))
plot(yearv,cpueUWm1,ylim=c(min(cpueUWm1,cpueUW2,cpueUW3,cpueUW4,cpueUW5,
cpueUW6, cpue.r),max(cpueUWm1,cpueUW2,cpueUW3,cpueUW4,cpueUW5,
cpueUW6,cpue.r)),ylab="CPUE", xlab="Year",type="l",lty=1,col=1)
lines(yearv,cpueUW2,lty=2,col=2)
lines(yearv,cpueUW3,lty=3,col=3)
lines(yearv,cpueUW4,lty=4,col=4)
lines(yearv,cpueUW5,lty=5,col=5)
lines(yearv,cpueUW6,lty=6,col=6)
lines(yearv,cpue.r,lty=9,col=9)
legend(1987,2.2, c("Model 1","Model 2","Model 3","Model 4","Model 5", "Model 6","Raw"),
lty=c(1:6,9) ,col=c(1:6,9))

criterion_lm <- function(fit){
  # this function calculates the aic, aicc and bic for a linear model after
  # it has been fitted with lm()
  k <- length(fit$coefficients)+1; # no. of parameters, add one for the variance
  rss <- sum(fit$residuals^2); # residual sum of squares
  n <- length(fit$residuals); # no. of observations
  aic <- 2*k + n*(log(2*pi*rss/n)+1);
  aicc <- aic + 2*k*(k+1)/(n-k-1);
  bic <- aic - 2*k + k*log(n);
  return(list(aic=aic,aicc=aicc,bic=bic));
}

m1<-criterion_lm(RTMPm1)
m2<-criterion_lm(RTMPm2)
m3<-criterion_lm(RTMPm3)
m4<-criterion_lm(RTMPm4)
m5<-criterion_lm(RTMPm5)
m6<-criterion_lm(RTMPm6)
```