Supplementary Material for ”Space-time modelling of blue ling for fisheries stock management”

Nicole H. Augustin[1],Verena M. Trenkel[2], Simon N. Wood[1] and Pascal Lorance[2]
[1]Department of Mathematical Sciences, University of Bath, Bath, UK.
[2]Ifremer, rue de l’Ile d’Yeu, BP 21105, 44311 Nantes Cedex 03, France.
nha20@bath.ac.uk.

November 20, 2012

1 Model checking results

The residuals plots were qualitatively similar for all considered models. We present the results for the best model (Model 2a). The histogram of residuals as well as the qq-plot show no anomalies (Figure 1). This indicates that the selected value of the index for the Tweedie distribution (p=1.5) was suitable, otherwise the histogram of residuals would not be symmetrical and could even be bimodal (V. Trenkel, unpublished results). As the model does not include a vessel effect but takes account of inter-vessel differences via vessel engine power, we considered the distribution of deviance residuals by vessel (Figure 2 left). The median and spread of residuals was similar across vessels and generally negative. Deviance residuals had similar spread and medians for all years (Figure 2 right). The temporal variograms per vessel indicated that there was no within month temporal correlation pattern in residuals (flat variograms), though the temporal correlation varied substantially between vessels (Figure 3). Spatial variograms for residuals outside and during the spawning season within each year showed no spatial autocorrelation of residuals, except for slight autocorrelation for the residuals in year 2003 and 2006 during the spawning season (results not shown). There was no systematic difference in spatial autocorrelation of residuals during and outside the spawning season.

Cross-validation provided prediction errors by fishing area and by year. Boxplots showed that areas new5 and ref5 had larger predictions errors compared to the other three areas (Figure 4 left). The prediction errors were similar for all years (Figure 4 right).

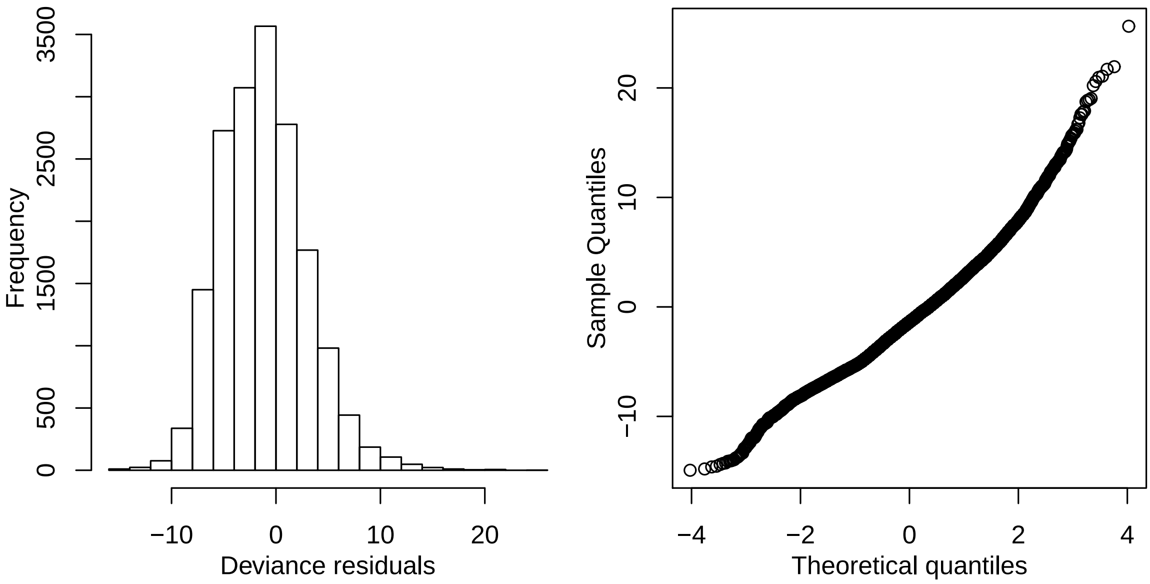

Figure 1: Histogram of deviance residuals (left) and QQ-plot for full space-time model with soap film spatial smooth (right) (model 2a).

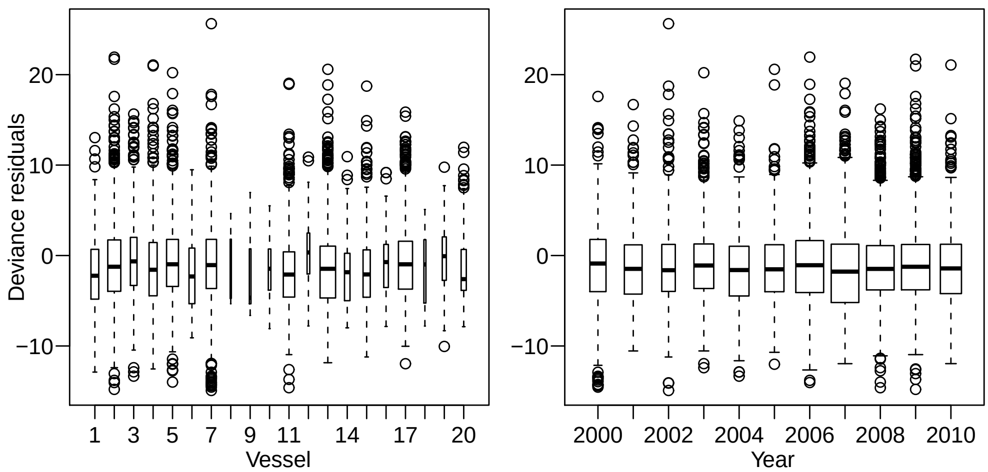

Figure 2: Boxplots of deviance residuals per vessel (left) and year (right) (model 2a). Box width is proportional to square-root of the number of observations.

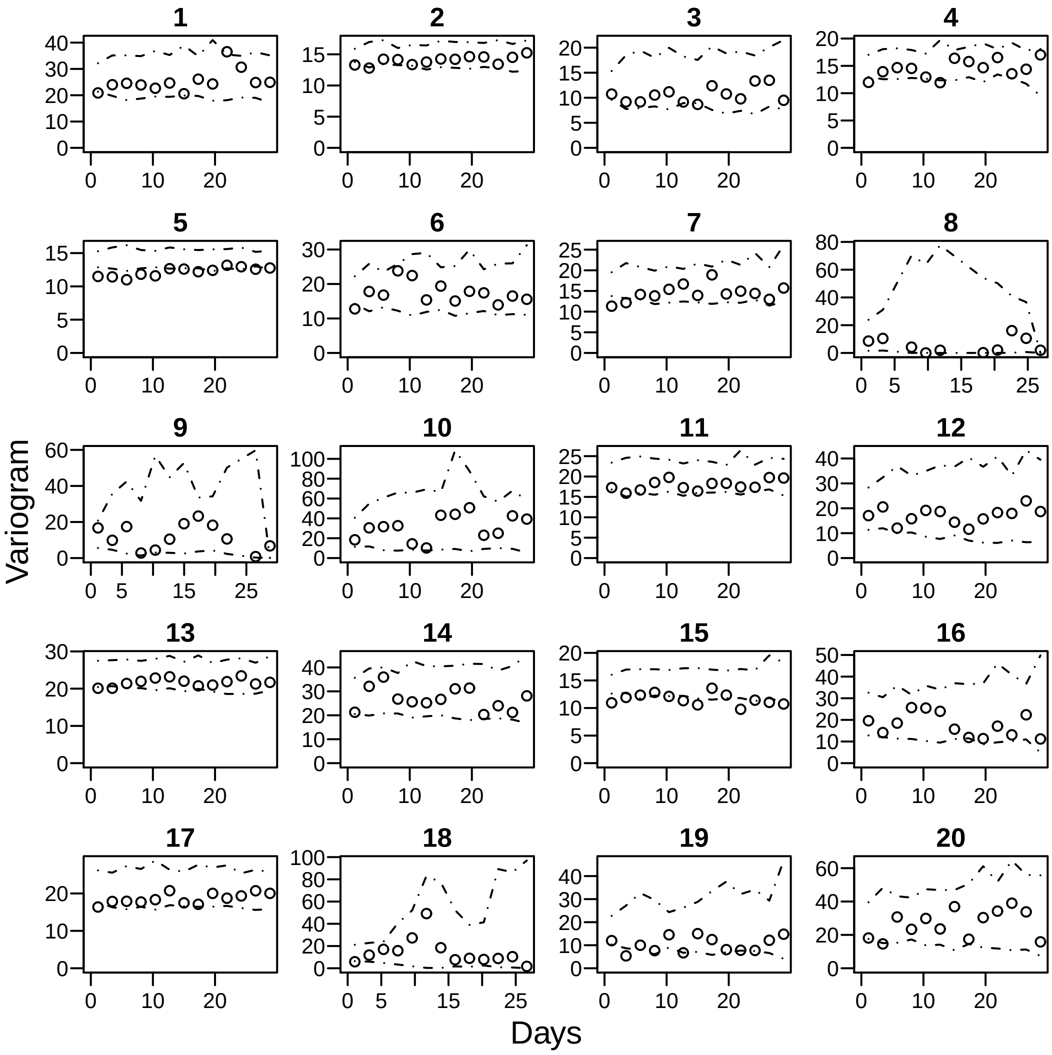

Figure 3: Empirical variograms of Deviance residuals (model 2a) per vessel along the time-axis (points). Numbers are vessel identifiers. The envelops represent minimum and maximum values derived by permuting data points between times for each vessel. The variograms indicate model mis-specification if the empirical points lie outside the envelops.

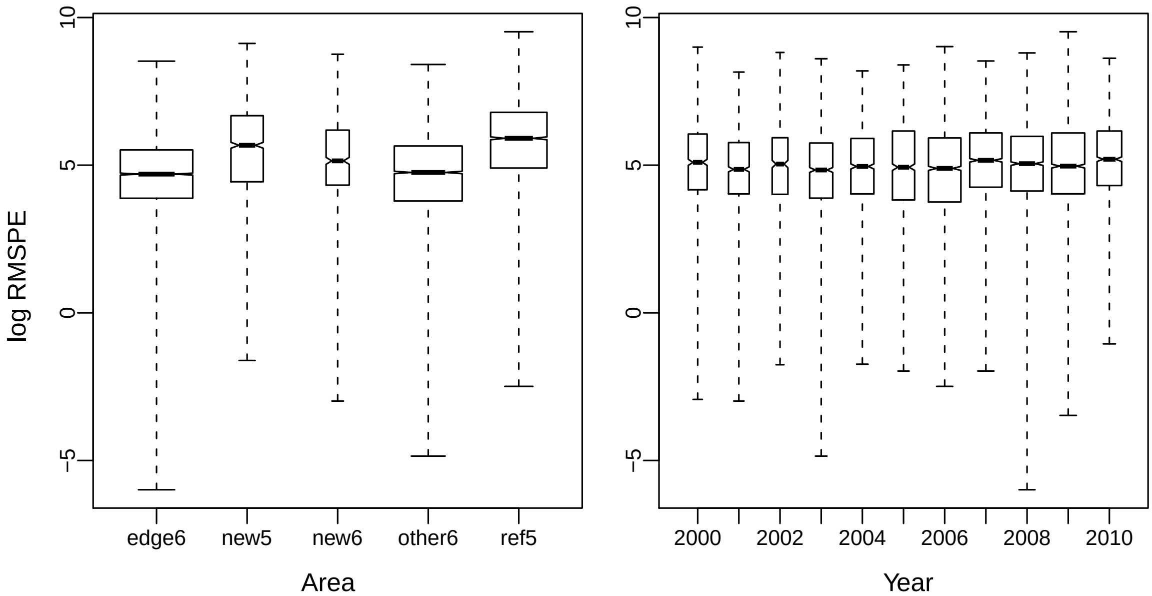

Figure 4: Boxplots of log root mean square prediction error (RMSPE) for 11-fold crossvalidation (random vessel-year selection) by area (left) and by year (right) (model 2a). The width of the boxes is proportion to the square-root of the number of observations in the respective group.

2 R code example

```
# Supporting material - Example R Code for models presented in:
#
# Nicole H. Augustin, Verena M. Trenkel,  Simon N. Wood and Pascal Lorance.
# Space-time modelling of blue ling for fisheries stock management.
# Environmetrics.
# The data is confidential and cannot be shared.

# preliminaries
# attach required libraries
.libPaths("/home/nha20/Rlocal")

load("Dat.rda") ### blue ling data
load("conBL.rda") ### data frame with polygons for boundary,
  ### variables longitude and latitude
knots<-read.table("Knots.csv",header=T,sep=";")
## file with variables longitude and latitude of 74 knot positions
knots[12,2] <- 6390 ## one knot is on boundary hence it is moved


library(mgcv)
library(soap)
library(MASS)

### function to print model summaries
printresults<-function(gobject){
edf=sum(gobject$edf)
nobs=length(gobject$y)
BIC<--2*logLik(gobject)+edf*log(nobs)
results<-c(sum(gobject$edf), gobject$deviance,summary(gobject)$r.sq,
gobject$aic,BIC)
results<-round(results,2)
names(results)<-c("edf","deviance","adj.R-sq","AIC","BIC")
results
}
dknots<- nrow(knots)


#Model 1a
#### full model with vessel as random effect
Dat$vessel<- factor(Dat$vessel)

mod1a <-gam(blue_ling~s(vessel,bs="re")+s(duration)+s(depth)+s(month,bs="cc")+
te(depth,month,bs=c("tp","cc"),k=c(10,10))
+te(depth,year,bs=c("tp","cr"),k=c(10,7))
+te(longitude,latitude,year,d=c(2,1),bs=c("sf","cr"),k=c(30,5),
xt=list(list(bnd=list(conBL)),NULL))+te(longitude,latitude,year,d=c(2,1),
bs=c("sw","cr"),k=c(dknots,5),xt=list(list(bnd=list(conBL)),NULL)),
knots=knots,data=Dat,family=Tweedie(p=1.5,link="log"),method="REML")
printresults(mod1a)

#Model 2a
```

```
#FULL model with depth (*year+month) + lat*long*year , fishing power
mod2a<-gam(blue_ling~s(duration)+s(depth)+s(month,bs="cc")+
te(depth,month,bs=c("tp","cc"),k=c(10,10))+te(depth,year,bs=c("tp","cr"),
k=c(10,7))
+te(longitude,latitude,year,d=c(2,1),bs=c("sf","cr"),k=c(30,5),
xt=list(list(bnd=list(conBL)),NULL))+
te(longitude,latitude,year,d=c(2,1),bs=c("sw","cr"),k=c(dknots,5),
xt=list(list(bnd=list(conBL)),NULL))+power,
knots=knots,data=Dat,family=Tweedie(p=1.5,link="log"),method="REML")

#Model 3a
#### additive model with vessel as random effect
mod3a <-gam(blue_ling~s(vessel,bs="re")+s(duration)+s(depth)+s(month,bs="cc")+
te(depth,month,bs=c("tp","cc"),k=c(10,10))+te(depth,year,bs=c("tp","cr"),
k=c(10,7))
+s(longitude,latitude,bs="so", k=30,xt=list(bnd=list(conBL)))+
s(year,bs="cr",k=5),knots=knots,data=Dat,family=Tweedie(p=1.5,link="log"),
method="REML")


#Model 4a
###soap additive with power
mod4a<-gam(blue_ling~s(duration)+s(depth)+s(month,bs="cc")+
te(depth,month,bs=c("tp","cc"),k=c(10,10))+te(depth,year,bs=c("tp","cr"),
k=c(10,7))+ s(year,bs="cr",k=5)+
s(longitude,latitude,bs="so",k=dknots,xt=list(bnd=list(conBL)))+power,
knots=knots,data=Dat,family=Tweedie(p=1.5,link="log"),method="REML")

#Model 1b
#### full model with vessel as random effect - NOSOAP

mod1b <-gam(blue_ling~s(vessel,bs="re")+s(duration)+s(depth)+s(month,bs="cc")+
te(depth,month,bs=c("tp","cc"),k=c(10,10))+te(depth,year,bs=c("tp","cr"),
k=c(10,7))
+te(longitude,latitude,year,d=c(2,1),bs=c("tp","cr"),k=c(dknots+30,5)),
data=Dat,family=Tweedie(p=1.5,link="log"),method="REML")

#Model 2b
#### FULL: in comparison without soap smooth
mod2b<-gam(blue_ling~s(duration)+s(depth)+s(month,bs="cc")+
te(depth,month,bs=c("tp","cc"),k=c(10,10))+te(depth,year,bs=c("tp","cr"),
k=c(10,7))
+te(longitude,latitude,year,d=c(2,1),bs=c("tp","cr"),k=c(dknots+30,5))
+power,
data=Dat,family=Tweedie(p=1.5,link="log"),method="REML")

#Model 3b
#### additive model with vessel as random effect - NOSOAP
mod3b<-gam(blue_ling~s(vessel,bs="re")+s(duration)+s(depth)+s(month,bs="cc")+
te(depth,month,bs=c("tp","cc"),k=c(10,10))+te(depth,year,bs=c("tp","cr"),
k=c(10,7))
+te(longitude,latitude,d=c(2),bs=c("tp"),k=c(dknots+30))
+s(year,bs="cr",k=5),
```

```
data=Dat,family=Tweedie(p=1.5,link="log"),method="REML")


#Model 4b
#### additive model with vessel power - NOSOAP
mod4b<-gam(blue_ling~power+s(duration)+s(depth)+s(month,bs="cc")+
te(depth,month,bs=c("tp","cc"),k=c(10,10))+te(depth,year,bs=c("tp","cr"),
k=c(10,7))
+te(longitude,latitude,d=c(2),bs=c("tp"),k=c(dknots+30))
+s(year,bs="cr",k=5),
data=Dat,family=Tweedie(p=1.5,link="log"),method="REML")



########### Computing time trends with Bayesian credible intervals

gobject<-mod1a

### obtain design/prediction matrix
ndat<-Dat

M<-predict(gobject,newdata=ndat,type="lpmatrix")
### ndat is the data with the spatial grid we want to predict for
### here we have set it to data as observed, for prediction need to
### need to set values in ndat as required

#### simulate 1000 fitted values from posterior distribution of parameters
simcoef <-mvrnorm(n=1000, coef(gobject), gobject$Vp)
simfit<-as.matrix(M)%*% t(simcoef)

simfit <- aggregate(simfit, by=list(ndat$year),mean,na.rm=TRUE)
years<-simfit[,1]
simfit<-simfit[,-1] ### exclude group index

## obtain quantiles for trendplot
simquant<-apply(simfit,1,quantile,p=c(0.025,0.5,0.975),na.rm=TRUE)

## basic time trend plot

plot(years,simquant[2,],type="o",pch=1,lty=1,col=1)
lines(years,simquant[1,],lty=2,col=1)
lines(years,simquant[3,],lty=2,col=1)

#### plot of spatial predictions
years<- sort(unique(Dat$year))
zlimmi<-range(gobject$linear.predictors)

par(mfrow=c(3,3))
par(pty="s")
oldpar<-par()
par(mar=oldpar$mar-c(4,3,3,2))

for (i in 1:length(years)) {
```

```
print(years[i])
vis.gam(gobject,view=c("longitude","latitude"),zlim=zlimmi,
cond=list(year=years[i]),plot.type="contour",type="link",
too.far=0.05, color="topo", main=" ", xlab="",ylab="")

text(x=600,y=6200,paste(years[i]))
points(Dat$longitude[Dat$year==years[i]],Dat$latitude[Dat$year==years[i]]
,pch=14,cex=0.3)
lines(conBL$longitude,conBL$latitude,type="l",xlim=range(Dat$longitude),
ylim=range(Dat$latitude)) ## add points where sample locations are
}
```