Progress working report Badminton spatial modeling workshop: 19-21 May 2010

Workshop objective

The objective of the data workshop (WP1) is to collate discard data and store in a similar format. The modelling workshop attempted to combine the data for one species: Cod (*Gadus morhua*) in the North Sea and set up a statistical framework which can model the spatial distribution in space and time.

Data input

Currently the discard data from Denmark, England and the Netherlands have been included in a single framework (Fig. 1). One of the challenges in using the data into a single framework, is that the different discard data not only show little spatial overlap, also the vessel and gear characteristics differ. Consequently, a multivariate analysis cannot differentiate between spatial and gear effects. In statistical terms, this problem is known as multi-collinearity. To deal with this issue, IBTS survey has been included as well. This will allow for modeling the species distribution and country specific deviations much be due to differences in catchability (Fig 1).

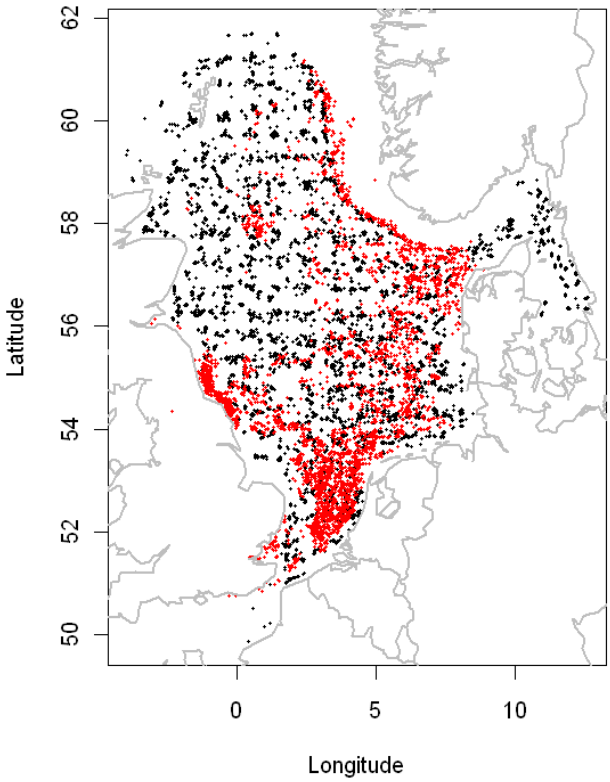


Figure 1. Spatial distribution of discard data (red) and IBTS (black).

Statistical modeling framework

The statistical model framework will be based on a Generalized Additive Mixed Model (GAMM). GAMMs allow for non-normal distributed data, non-linear effects of covariates (such as space and time) and covariance structures that capture the non-independence in the data, such as within-vessel correlation in the data and spatial or temporal correlation. For more details see Zuur et al. () and Wood (2006).

Defining the response and error structure

The raw data consists of numbers by length measured for each sample of the haul. Because only a fraction of the total discards have been sampled, these numbers are multiplied by the sub sampling factor resulting in total numbers by length in the haul. The count data will have a Poisson or quasi-Poisson distribution. The latter accounts for possible over- or under dispersion. Finally, the log of haul duration is included in the model as an offset. So the model will quantify the influence of covariates on the discard numbers per minute.

Factors influencing discard numbers (per unit effort)

Several environmental, temporal and gear specific covariates may influence the observed discard numbers. For more details see Rochet & Trenkel (). Below we describe several of these factors and illustrate how they may be included as smooth terms in the GAMM (see Figure 2).

- Spatial heterogeneity in discards which may be length specific. This may differ between years and seasons: s(lon, lat, length, year, season)
- Quota restrictions. This probably only applies to the larger individuals. If quota restrictions apply, individuals (just) above MLS may be discarded more. Quota restrictions may differ between countries and vessels, but probably no data on vessel specific quota restrictions. This does not apply to survey. To capture this in the model: s(percent quota used, length of fish or above/below MLS, day of year, market price, by country)
- Vessel capacity. Discarding may increase if the vessel reaches its capacity. Perhaps little data: s(vessel capacity)
- Small individuals are not retained by the gear, large individuals may end up in the landings. This effect will differ between mesh sizes and whether a sorting panel/selectivity device is used: s(mesh, length, sorting panel). If good information is available on the effect of those panels (or mesh size selectivity), this could be incorporated into the model offset
- Environmental variables: bottom type, depth. Also, time to low water may influence where they distribute in the water column: s(depth) + s(bottom) + s(time to low/high tide) + s(local current velocity)

- Different discards per unit effort due to difference in year class strength: factor(year class) or factor(year) or s(length, year)
- Perhaps include the effect of catch composition in model (e.g. Nephrops). This is vessel specific, so perhaps no data to include in the model. However it may be reflected in the quota restrictions
- When the catch is large, this may lead to more discarding. Also catch composition may have an influence. E.g. dogfish may deteriorate the fish quality, potentially leading to more discards: s(catch per haul) + s(species specific catch (i.e. choose one or more species)).
- Day-night effect and day length: s(day or night), s(day length), s(day/night, day length)
- Weather conditions

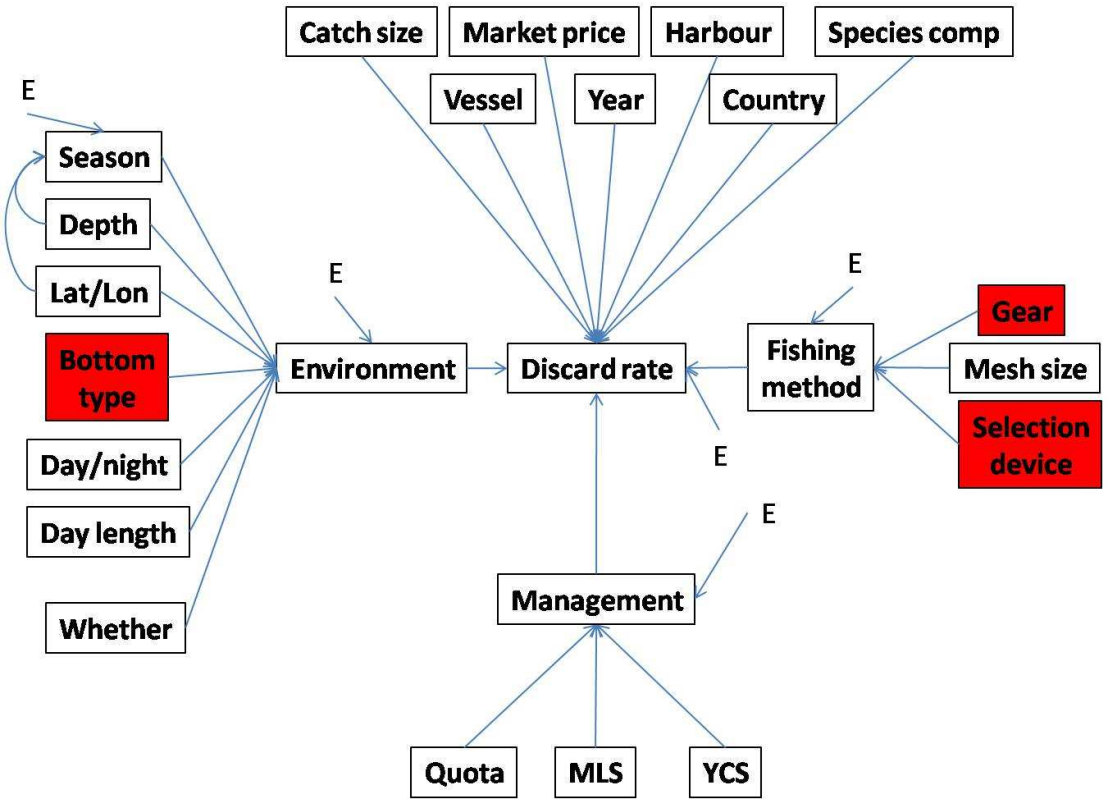


Figure 2. Diagram showing the different covariates which may influence the discard numbers. Data on red boxes may be difficult to obtain.

Accounting for unexplained variability

Some important covariates may not be included in the model. For example, if cod distribution heavily depends on bottom structure, small scale patchiness exist, but may not be capture by the model. This will lead to spatial correlation in the residuals. Not dealing with remaining non-independence in the data will often lead to a model which is too complex and underestimates of the parameter standard errors.

Some examples:

- Gear/vessel specific variability not included in the model. Include in model random effect: random = list(code_tripnr=~1) or list(code_vessel=~1)
- Small scale local variation (e.g. due to local environmental conditions, e.g. bottom type), not included in the model. Include spatial correlation in the model (see table below)
- Small scale temporal correlation. This can be very similar to the spatial component. Include temporal correlations structure. E.g. corAR1 or even more complex: correlation=corCAR1(~d_date|trip_no)

Available correlations structures in R:

corAR1	autoregressive process of order 1.
corARMA	autoregressive moving average process, with arbitrary orders for the autoregressive and moving average components.
corCAR1	continuous autoregressive process (AR(1) process for a continuous time covariate).
corCompSymm	compound symmetry structure corresponding to a constant correlation.
corExp	exponential spatial correlation.
corGaus	Gaussian spatial correlation.
corLin	linear spatial correlation.
corRatio	Rational quadratics spatial correlation.
corSpher	spherical spatial correlation.
corSymm	general correlation matrix, with no additional structure.

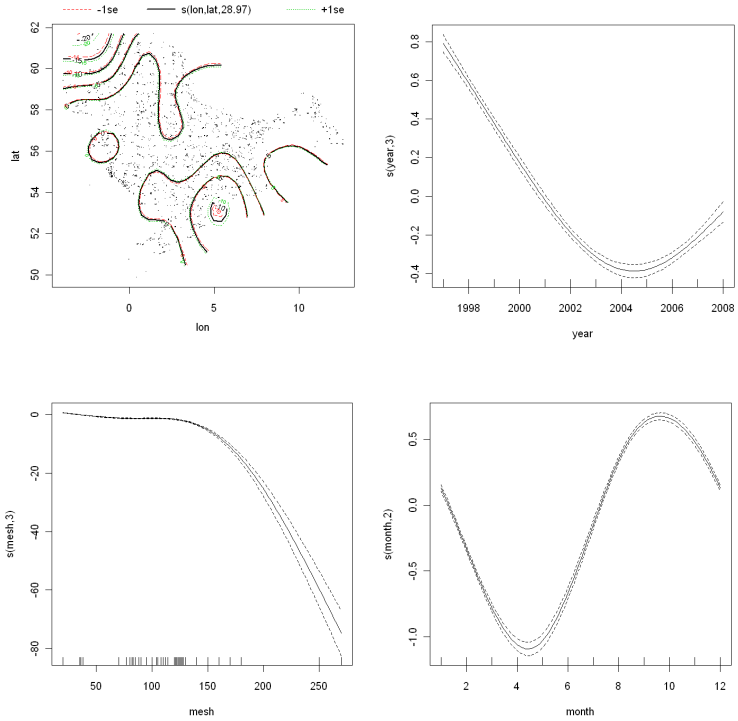
Model construction and variable selection

Model selection is based on the procedure described in Zuur et al. ()

1. **Fit a complex 'near-optimal' gam model:**

```
gam1<-gam(number~s(lon,lat)+s(year,k=4,fx=T)+s(mesh,k=4,fx=T)+
          factor(country)+s(month,bs="cc",k=4,fx=T)+
          offset(offsetter),data=h.sub2,family=poisson)
```

This model is most likely over-parameterized (i.e. too complex) and the standard errors are underestimated. This is reflected in the narrow confidence intervals around the response curves.



2. Choose correlation structure. For example:

Random effect for trip

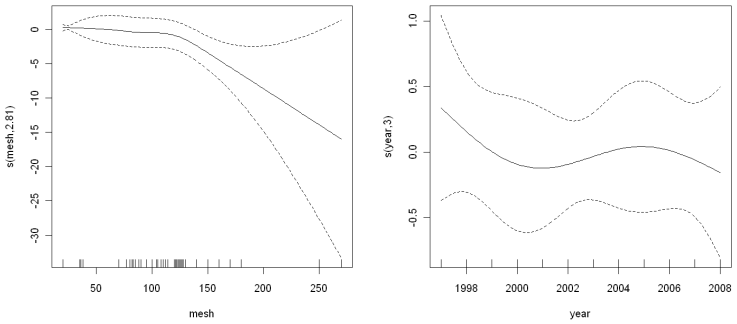
```
gamm1a<-gamm(number~s(lon,lat)+s(year,k=4,fx=T)+s(mesh)+
             factor(country)+s(month,bs="cc")+
             offset(offsetter),data=h.sub2,
             family=poisson,random=list(trip_no=~1),
             method="REML"))
```

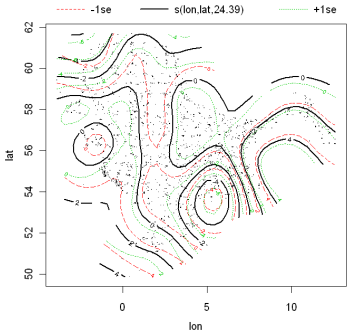
no random effect

```
gamm1d<-gamm(number~s(lon,lat)+s(year,k=4,fx=T)+
               factor(country)+s(mesh)+
               offset(offsetter),data=h.sub2,
               family=poisson, method="REML")
```

Check AIC; random effect for trip is better!

```
BIC(gamm1a$lme)
BIC(gamm1d$lme)
```





3. Forward model selection of covariates (see R-code)

	criterion	var_added	npar	L	AIC	BIC
1	BIC	Intercept	3	-11794.64	23595.28	23612.76
2	BIC	s(lon,lat)	6	-8337.37	16686.74	16721.69
3	BIC	s(mesh)	8	-8006.11	16028.22	16074.81
4	BIC	s(year,k=4,fx=T)	11	-8010.82	16043.64	16107.71
5	BIC	factor(country)	13	-8038.4	16102.8	16178.51
6	BIC	s(month,bs="cc")	14	-8094.31	16216.62	16298.16

4. Use final model (fitted using REML) to generate output

For example the model can be used to make spatial predictions. It should be noted that the model presented above is fitted to only a subset of the data, insufficient covariates and improper diagnostic have been generated. It is merely an illustration of the possibilities of the framework and a start-up for future, improved analysis.

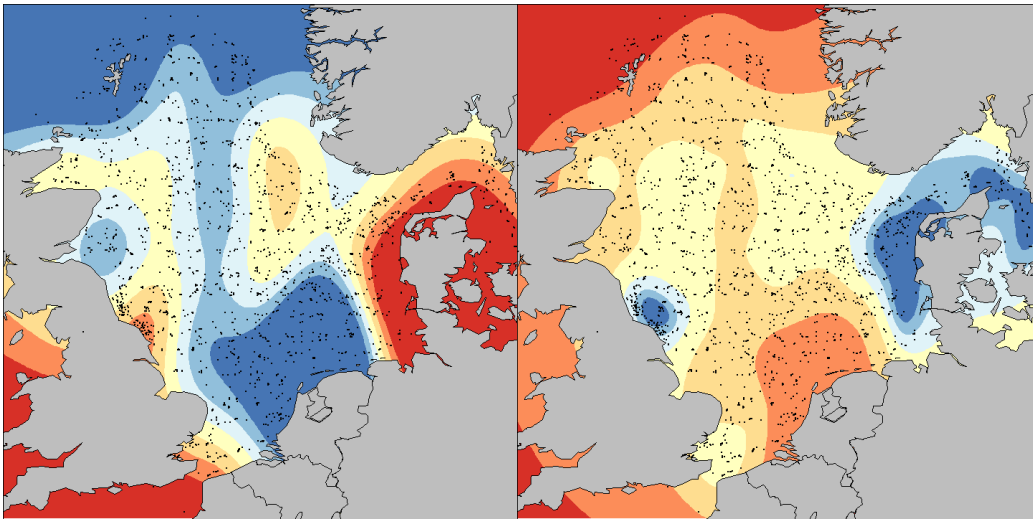


Figure 3. Spatial distribution of modelled cod discards (Danish, UK and IBTS data) and their corresponding standard errors (right). This figure cannot be trusted, but just illustrate the possibility (and limitations) of the framework.

References

Rochet, M-J & Trenkel, V.M. Factors for the variability of discards: assumptions and field evidence. 2005. Can. J. Fish. Aquat. Sci. **62**: 224–235.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M. 2009. Mixed effect models and extensions in Ecology. Springer.

Wood, S. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC Texts in Statistical Science