Documention of what is implemented so far

Olav Nikolai Breivik

January 13, 2018

1 Sampling procedures

This section describes what is implemented in our procedure for estimating CPUE's with uncertainty. All the code is given in the R-folder in the package.

1.1 Sampling ALK

First of all let just state that the ALK is a deterministic function of the CA-data in an RFA. We have implemented two procedures for sampling from the CA-data for each RFA, and thereby sampling ALK. We call the two method the *simple ALK simulating procedure* and the *stratified ALK simulating procedure*.

1.1.1 Simple ALK simulating procedure

Assume there are N samples of otholits in an RFA. The simple procedure sample N rows in the CA-data randomly with replacement.

1.1.2 Stratified ALK simulating procedure

The IBTS tries to sample otholits from a certain number of fish of each length class. Further more, only one fish per length class is sampled from each trawl haul (this is information Olav heard at the yearly meeting with IMR and OVT were a presentation of the IBTS-data was held, but he has no documentation of this yet). One could the argue that the simple ALK simulating procedure given in **??** overestimate the variance since the the CA-data is intended to be distributed evenly, with natural limitation, over the length classes. To compensate for that the CA-data should consist of observations in a wide specter of length class, we can simulate the CA-data in a stratified way. Currently we have implemented (January 2017) a suggestion for a stratified procedure. Let $\mathbf{L}$ be all the length classes for a given species, and let $l_1, l_2 \in L$ be two length classes. Define $d(l_1,l_2)$ to be the number of length classes in L between l_1 and l_2. The procedure is like this: Let L_{obs} be all the length classes we have observations of otholits.

1. Let $i = \min(\mathbf{L}_{obs})$

2. Assume length class l_i has n_i observations of otholits. If $n_i > 1$ we sample with replacement n_i of them. If $n_i = 1$ we sample with replacement either that observation or an observatin from a length class l_j, where l_j is such that $d(l_i,l_j)$ is minimised given that there is an observations in length class l_j.

3. If $i < max(L_{obs})$ set i equal to $i + 1$ and go to step 2.

4. Repeat step 1-3 B times.

1.2 Sampling CPUE

Sampling the CPUE per length class can be divided into three steps. First the ALK is sampled, then the CPUE per length class is sampled, and lastly those two samples are combined. In section **??** we documented the first part, that is simulating the ALK. We are now going to document the sumulation procedure for CPUE per length class.

Note first that the CPUE per length class i deterministic function of the HL-data. As for the ALK, we have implemented two procedures for simulating the HL-data, and thereby the CPUE. We call these two method the *simple procedure for simulating CPUE* and *a stratified procedure for simulating CPUE.*

1.2.1 Simple CPUE simulating procedure

Assume there are N_h trawl hauls in an RFA. The simple procedure samples with replacement N_h trawl hauls in the RFA.

1.2.2 Stratified CPUE simulating procedure

Let $\mathbf{N}_{rec} = \{1,2...,n_{rec}\}$ be the statistical rectangles in the RFA.

- Let $i = \min(\mathbf{N}_{rec})$.

- Assume there is n trawl hauls in ith statistical rectangle. If $n > 1$ we sample with replacement n trawl huals in that statistical rectangle. Assume $n = 1$, we then sample with replacent either that trawl haul or the trawl haul closest in air distance.

- If $i < \max(\mathbf{N}_{rec})$, set i equal to $i+1$ and go to step 2.
- Repeat step 1-3 B times.

2 Suggestion for a spatio-temporal-length model

Let us just focus on a model for calculation of the CPUE per length class in the beginning.

Let $\mathbf{Y}(s,t) = (Y_{l_1}(s,t),...,Y_L)$ be a vector with the number of fish for each length class of a given species is caught at location $\mathbf{s}$ at time t. Assume a that each element of $\mathbf{Y}(s,t)$ is zero-inflated negative binomial distributed. We assume that the additional zero-proability can be described by a logitlink. In the beginning we can assume that there are no link between the linear predictor for the negative binomial distribution and the zero-probability. Assume that the log-linear predictor of the negative binomial part, $\boldsymbol{\mu}(\mathbf{s},t)$ is given by

$$\boldsymbol{mu}(\mathbf{s},t) = \boldsymbol{\beta}_0 + \boldsymbol{\gamma} + \boldsymbol{\delta}, \tag{1}$$

were $\boldsymbol{\gamma}$ is an spatio-temporal Gaussian field, and $\boldsymbol{\delta}$ is a "spatio-temporal-length" Gaussian field. The term "spatio-temporal-length" is never used before as I know, but i believe this idea can be good. Assume that the covariance of $\boldsymbol{\delta}$ is separable and can be divided into a spatio-temporal part, represented by covariance matrix $\boldsymbol{\Sigma}_{ST_l}$, and a part describing the correlation between the fish lengths, represented by covariance matrix $\boldsymbol{\Sigma}_l$. We have then that the covariance matrix representing $\boldsymbol{\delta}$ is given by $\boldsymbol{\Sigma}_\delta = \boldsymbol{\Sigma}_{ST_l} \otimes \boldsymbol{\Sigma}_l$.

We can the try out for example an exponential correlation structure in both space, time and length class and see what happens. This is quite straight forward to achieve in TMB i believe. We may want to model both small and large scale spatio-temporal structures.

We can futher define something similar for the zero-probability, but I guess that it is not so important to include complex structures for the zero-proabability, but it may be important to include them so that they do not make much noize for the negative binomial part.