Hazard function models
a new tool to estimate fish mortality rates from age data
with application to Qld mullet (*Mugil cephalus*)

Marco Kienzle

Qld Department of Agriculture, Fisheries - Dutton Park
UQ's school of Agriculture and Food Sciences - St Lucia

March 2, 2016

Contributions acknowledgements

The development of this method, presentation and paper is the result of many discussions and inputs from a range of people (in no particular order)

C. Millar (FRS-Scotland), W.N. Venables (CSIRO), N. White (QUT), P. Baker and R. Ware (UQ-School of Public Health), D. Mayer and J. McGilvray (DAF), 4 anonymous reviewers

Some publications about survival analysis:

Cox, D.R. and Oakes, D. (1984) **Analysis of survival data**, *Chapman and Hall Ltd, London*
Kleinbaum, D.G. and Klein, M. (2005) **Survival Analysis: A Self-Learning Text** *Springer, New York*
Ferrandis, E. and Hernández, P. (2007) **Direct Survival Analysis: a new stock assessment method** *Scientia Marina*, 71:1

Motivations for this collaboration between biometry and LTMP

- convert LTMP data into useful scientific information for fisheries managers
- apply statistical methods to estimate mortality rates from age data
- choose methods capable of updating mortality estimates yearly as new data become available

Methods applied to perform this work

- critical evaluation of available methods to estimate mortality rates (a4a, CASAL, SS3, etc...)
- determine the accuracy and precision of each method using Monte Carlo simulations (quality control, testing)
- comparison of statistical methods using the likelihood approach (R.A. Fisher) and selection of the maximum likelihood mortality rates estimates

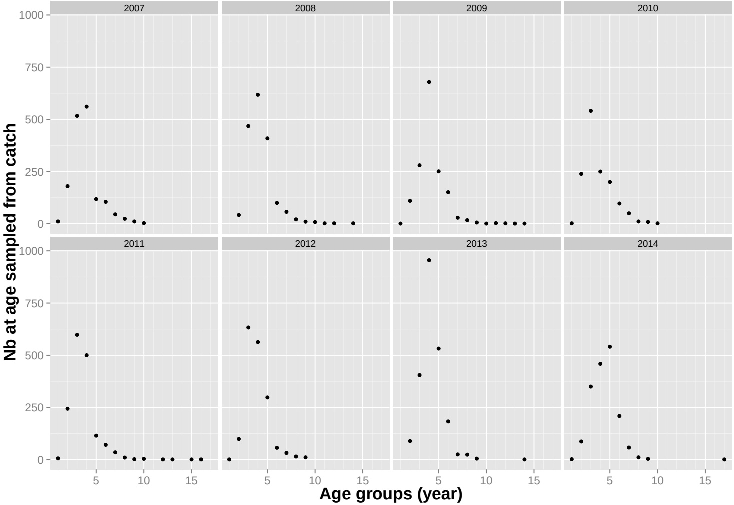
"... it will be found that the method here outlined is illuminating in all similar cases when the same quantity may be ascertained by more than one statistical formula" (R.A. Fisher introducing the likelihood method in 1920)

The problem of estimating mortality rates in the age-structure context

Estimate natural and fishing mortality rates for all ages and years from age data
For example,

		age − groups						
		1	2	3	$\dots$	j	$\dots$	n
years	1	0.3	0.35	0.4	$\dots$	$\dots$		$\dots$
	$\vdots$	$\vdots$	$\vdots$	$\dots$		$\dots$		$\dots$
	i	$\vdots$	$\vdots$	0.65		$M_{i,j} + F_{i,j}$		$\dots$
	$\vdots$			0.7		$\dots$		0.8
	p			$\dots$		0.8		0.9

Mullet age data from the long term monitoring program (LTMP)



3-5 years old mullet were most frequent in the sample

Deterministic theory for fishing a single cohort

Exponential decline a cohort through time from Quinn and Deriso [1999]

$$\frac{dN}{dt} = -FN - MN \tag{1}$$

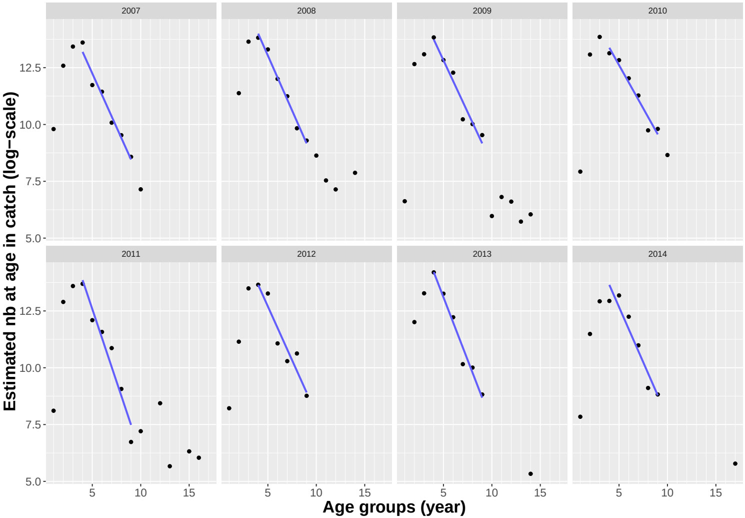
$$N(t) = N_0 e^{-(F+M)t} \tag{2}$$

Baranov catch equation (1910)

$$\frac{dC}{dt} = FN \tag{3}$$

$$C_{0\to\tau} = \frac{F}{M+F} N_0 (1 - e^{-(M+F)\tau}) \tag{4}$$

Cross sectional approach to estimate total mortality (M+F) for mullet

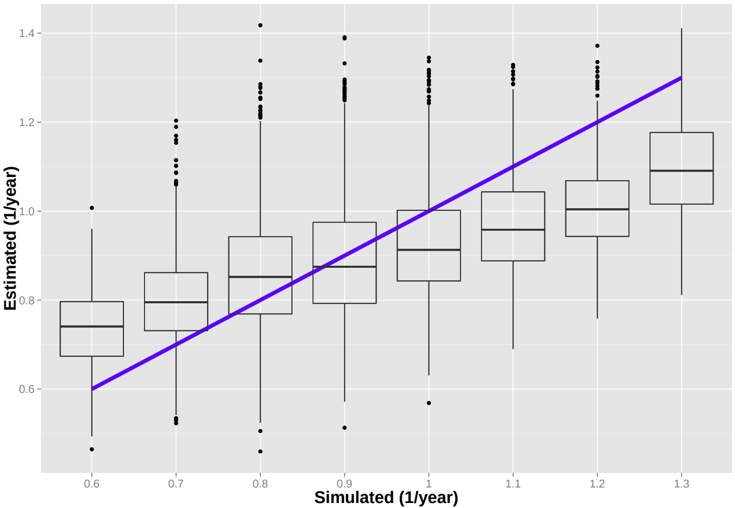


Monte Carlo simulations to test statistical methods

- Validate methods and their computer implementations (regression tests, software eng.)
- Quantify statistical properties of methods (bias, variance, sentitivity, ...)
- Compare alternative methods (in stock assessment: National Research Council (1999) Improving Fish Stock Assessments, National Academic Press.)
- Simulated 8000 age sample with parameters consistent with sea mullet fishery ($0 \leq$ age ≤ 16 years)

Variable type	Distribution	Parameters
recruitment	uniform	min=4e6, max=8e6
natural mortality	uniform	min=0.30, max=0.36
catchability	uniform	min=1.5e-4, max=2.5e-4
fishing effort	uniform	min=2e3, max=4e3
gear selectivity α	uniform	min=7.5, max=8.5
gear selectivity β	uniform	min=2, max=3

Monte Carlo simulations:
cross sectional provides inaccurate estimates of mortality



Shortcomings of cross sectional analysis

- provides an inaccurate estimate of total mortality
- ignores the fact that catches are made of several cohorts
- it mixes recruitment and mortality signals into a noisy estimator
- uses only a subset of the data (48/88=54% in the case of mullet)
- it doesn't offer a framework where we can compare different ideas (take it or leave it)

Survival analysis: a probabilistic theory for fishing a single cohort

Survival analysis [Cox and Oakes, 1984, Ferrandis and Hernández, 2007] using a constant hazard function

$$h(t;\theta) = M + F \tag{5}$$

The probability density function (pdf)

$$f(t;\theta) = (M+F)\ e^{-(M+F)t} = M\ e^{-(M+F)t} + F\ e^{-(M+F)t} \tag{6}$$

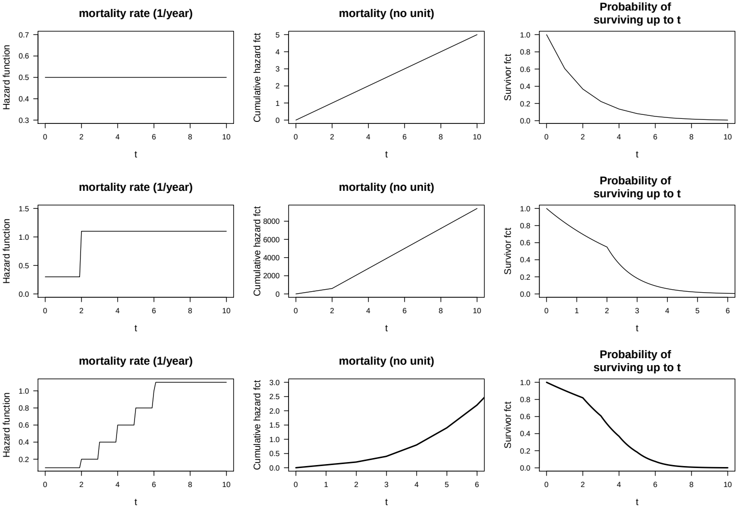
The likelihood [Edwards, 1992] of a sample of fish caught in the fishery (S_i) using the probability of dying in the interval $[a_i; a_{i+1}]$.

$$\mathcal{L} = \prod_{i=1}^{n} \Big(\int_{t=a_i}^{t=a_{i+1}} f(t;\theta)\ dt \Big)^{S_i} \tag{7}$$

The logarithm of the likelihood

$$\log(\mathcal{L}) = \sum_{i=1}^{n} S_i \log\big(e^{-(M+F)\times a_i} - e^{-(M+F)\times a_{i+1}} \big) \tag{8}$$

Illustration of survival analysis concepts



Hazard models for more complicate situations

Fishing mortality is a function of effort (E)
$F(t) = q\ E(t) \rightarrow h(t;\theta) = M + q\ E(t)$

Gear selectivity (s) varies with age
$F(t) = q\ s(t)\ E(t) \rightarrow h(t;\theta) = M + q\ s(t)\ E(t)$

Survival analysis is a flexible method to analyze age data

- restricted range of ages $\rightarrow$ truncated pdf
- +group $\rightarrow$ integrate over a larger interval $[a_+; \infty[$

Applying hazard model to data from multiple cohort

		age − groups				
		1	$\dots$	j	$\dots$	n
	1	$\dots$		$\dots$		$\dots$
	$\vdots$	$\dots$		$\dots$		$\dots$
years	i	$\dots$		$S_{i,j}$		$\dots$
	$\vdots$	$\dots$		$\dots$		$\dots$
	p	$\dots$		$\dots$		$\dots$

- $n + p - 1$ cohorts, separability $F_{i,j} = q\ E_i \otimes s_j$

The likelihood

$$\mathcal{L} = \prod_{k=1}^{n+p-1} \prod_{l=1}^{r_k} \left(\int_{t=a_{k,l}}^{t=a_{k,l+1}} g_k(t;\theta)\ dt \right)^{S_{k,l}} \tag{9}$$

Computer implementation of these methods

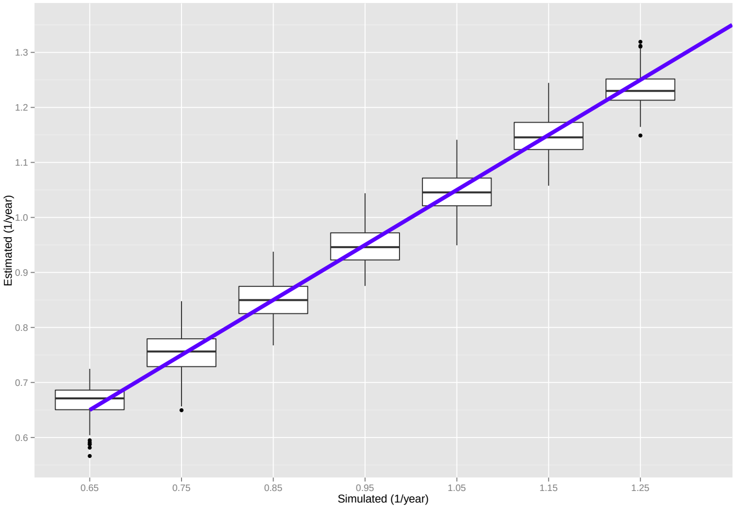
Methods implemented in R [R Core Team, 2015] in a package called Survival Analysis for Fisheries Research (SAFR)

Available at: `https://github.com/mkienzle/SurvivalAnalysisForFisheries`

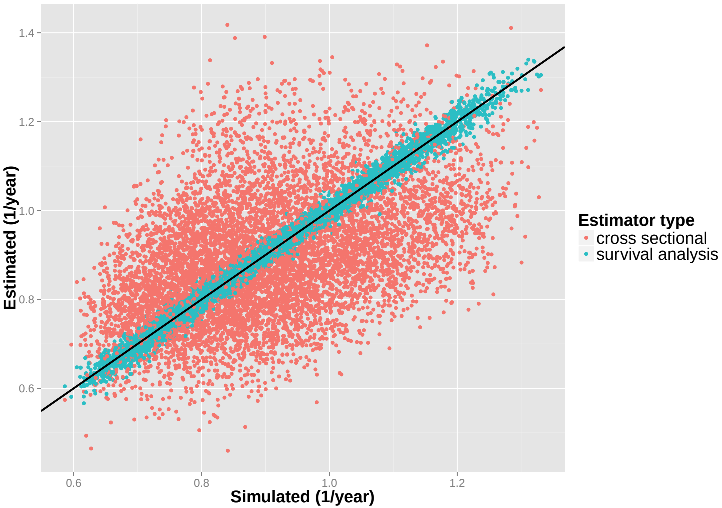
Manuscript was published after a double blind reviewing process:
Kienzle, M. (2015) **Hazard Function Models to Estimate Mortality Rates Affecting Fish Populations with Application to the Sea Mullet (Mugil cephalus) Fishery on the Queensland Coast (Australia)** *Journal of Agricultural, Biological, and Environmental Statistics*, 21(1), p. 76–91, doi:10.1007/s13253-015-0237-y

One of the reviewer's comment was : "The manuscript is well written and statistically sound."

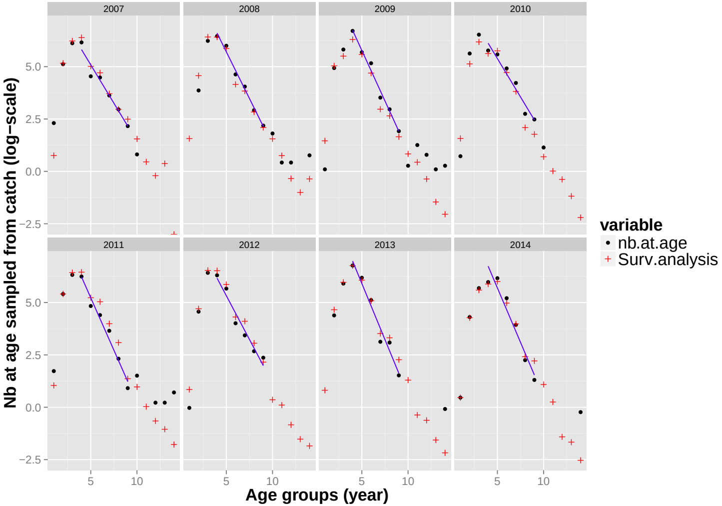
Survival analysis provides accurate estimates of mortality



Survival analysis is more precise than cross sectional analysis



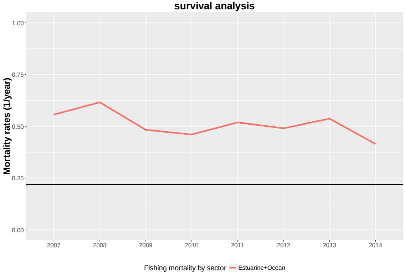
Survival analysis provides a model of all data

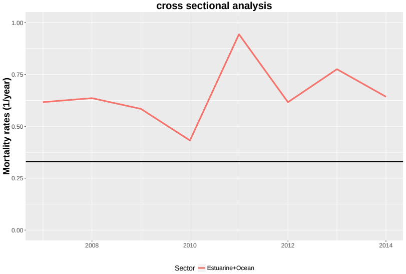


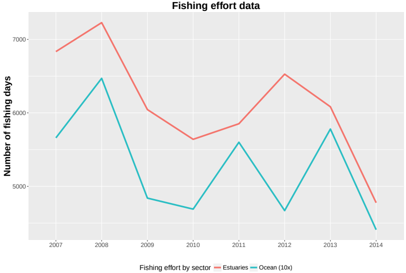
Summary of methods comparison:
cross sectional is less accurate at estimating mortality rates

Characteristics	Cross sectional	Survival analysis
accuracy	less	more
precision	less	more
deals with cohorts	no	yes
use all data	no	yes
likelihood based	no	yes
tested with MC	yes	yes
estimate nat. mort.	no	yes
-log($\mathcal{L}$)	22148.44	18795.28

Qld mullet fishery mortality rates estimates







Conclusions

- There exists an accurate and more precise method to estimate mortality rates from LTMP age data than currently used in DAF
- Survival analysis
  - is fast, robust, provides accurate estimates of mortality on a yearly basis
  - is published [Kienzle, 2015, Ferrandis and Hernández, 2007, Cox and Oakes, 1984]
  - is operational to analyze DAF's mullet data
  - is more precise than cross sectional analysis
  - is a general age-structured method for stock assessment applicable to other stocks

Thank you for your attention

- We value your feedback