APPLICATION

`TropFishR`: an `R` package for fisheries analysis with length-frequency data

Tobias Karl Mildenberger*,[1], Marc Hollis Taylor[2] and Matthias Wolff[3]

[1]*National Institute of Aquatic Resources, Technical University of Denmark, Kemitorvet, Building 202, 2800 Kgs. Lyngby, Denmark;* [2]*Johann Heinrich von Thünen Institute, Palmaille 9, 22767 Hamburg-Altona, Germany;* and [3]*Leibniz Centre for Tropical Marine Research (ZMT), Fahrenheitstr. 6, 28359 Bremen, Germany*

Summary

1. The `R` package `TropFishR` is a new analysis toolbox compiling single-species stock assessment methods specifically designed for data-limited fisheries analysis using length-frequency data.

2. It includes methods for (i) estimating biological stock characteristics such as growth and mortality parameters, (ii) exploring technical aspects of the fisheries (e.g. exploitation rate and selectivity characteristics), (iii) assessing size and composition of a fish stock by means of virtual population analysis (VPA), and (iv) assessing stock status with yield prediction and production models.

3. This paper introduces the package and demonstrates the functionality of a selection of its core methods.

4. `TropFishR` modernises traditional stock assessment methods by easing application and development and by combining it with advanced statistical approaches.

Key-words: data-poor, ELEFAN, FiSAT II, fish stock assessment, small-scale fisheries, tropics

Introduction

The scientific analysis and evaluation of fisheries provides the basis for sustainable exploitation of marine resources in ecological, economic and social terms. The knowledge of current fishing pressure and stock biomass in respect to biological reference levels supports the development of management strategies for sustainable harvest. For management purposes, single-species models are mainly used to derive these reference levels (Hilborn & Walters 1992; Skern-Mauritzen *et al.* 2015). The variety of potential models is large, but their application is strongly limited by the type of available data. Fish stocks for which only catch or length-frequency (LFQ) data is available, are categorised as data-poor and require a particular set of models (ICES 2012). Fish stocks under data-poor conditions are prominent in all regions and can even be influential in the management of data-rich fish stocks (e.g. choke species in multi-species fisheries). In particular, small-scale fisheries are prone to be data-poor (e.g. based on catch only) due to a difficulty in monitoring; these are typically fleets consisting of a large number of small boats with many landing sites and targeting a variety of species. These aspects make data-collection and monitoring for adequate stock management extremely difficult. Although small-scale fisheries comprise more than half of the global landings (Kolding, Bene & Bavinck 2014), most are neither assessed quantitatively nor qualitatively (Costello *et al.* 2012). For these fisheries, bulk catch data (no species identification) or LFQ data, from a limited time period, may be the only data available. In this work, we introduce the `R` package `TropFishR`, which compiles a wide range of stock assessment methods specifically designed for data-limited fisheries assessment using LFQ data and demonstrate the application of a selection of core methods.

The `TropFishR` package

Different `R` packages are currently available for fisheries analysis, such as `fishmethods` (Nelson 2016), `FSA` (Ogle 2016), or `FLR` (Kell, Mosqueira & Grosjean 2007). These packages offer a broad catalogue of general methods of fisheries science (e.g. natural mortality estimators) and age-based stock assessment tools. Furthermore, other `R` packages with length-structured methods and suitable for the assessment of data-poor fisheries are available, such as `DLMtool` (Carruthers & Hordyk 2016), offering a management strategy evaluation framework, `s6model` (Kokkalis *et al.* 2015), offering a length-based assessment model, or `LBSPR` (Hordyk *et al.* 2016), offering a length-based version of the spawning potential ratio method. The `TropFishR` package uniquely adds further data-limited method capacity (Table 1) by including traditional and updated versions of the Electronical LEngth Frequency ANalysis (ELEFAN) method, used in growth parameter estimation, with new optimisation techniques (Taylor & Mildenberger 2017), Millar's nonlinear selectivity models (Millar & Holst 1997), and a complete set of methods for fisheries analysis with LFQ data. This compilation allows a stock assessment routine to derive reference levels (e.g. F_{MSY}, $F_{0.1}$) by means of yield per recruit modelling, which may be

*Correspondence author. E-mail: t.k.mildenberger@gmail.com

Table 1. Overview of main functions of the `TropFishR` package and required data types

Function name	Description	Data category*
Bhattacharya()	Resolves observed LFQ distribution into Gaussian components	LFQ
powell_wetherall()	Estimates total mortality (Z) and asymptotic length (L_{inf})	LFQ
ELEFAN()	Estimates growth parameters	LFQ
ELEFAN_SA()	Estimates growth parameters with simulated annealing	LFQ
ELEFAN_GA()	Estimates growth parameters with genetic algorithm	LFQ
VBGF()	Translates length to age or vice-versa	–
growth_length_age()	Estimates growth parameters from length-at-age data	–
growth_tagging()	Estimates growth parameters from tagging data ('Munro plot')	–
M_empirical()	Applies 12 different empirical formulas to estimate natural mortality (M)	–
Z_BevertonHolt()	Applies Beverton & Holt method to estimate total mortality (Z)	CAA, LFQ
Z_CPUE()	Estimates total mortality (Z) from catch-per-unit-effort (CPUE) data	CPUE
catchCurve()	Applies the (length-converted) linearised catch curve	CAL, CAA, LFQ
recruitment()	Estimates recruitment patterns	LFQ
select_ogive()	Estimates the fraction per length group retained in the net	SEL
select()	Estimates the selectivity of gillnets and trawl nets	SEL
select_Millar()	Estimates the selectivity of different gears	SEL
VPA()	Applies virtual population analysis (VPA) or cohort analysis (CA)	CAL, CAA
predict_mod()	Applies yield-per-recruit (YPR) models	CAL, CAA
prod_mod()	Applies the equilibrium approach of Schaefer and Fox production models	CPUE
prod_mod_ts()	Applies Schaefer and Fox production models with time series fitting	CPUE

*Data types and available example datasets in parentheses: LFQ = Length frequency data (alba, synLFQ1 – syLFQ7 datasets), CAL = catch at length data (goatfish, hake), CAA = catch at age data (shrimps, synCAA1, synCAA2, whiting), SEL = selectivity data (bream, gillnet, haddock, tilapia, trammelnet), CPUE = catch and effort data (emperor, synCPUE, trawl_fishery_Java).

based on a single year of LFQ data. Until now the preferred software for single species stock assessment with length-frequency data has been the windows-based programme FiSAT II (Gayanilo, Sparre & Pauly 1996) due to its user-friendly, click-based interface. The software is, however, limited in its ability to import data and perform automated analyses. The `TropFishR` package aims to remedy these shortcomings by allowing further expansion and flexibility. Although wider in scope, the main methods follow those outlined in the FAO manual 'Introduction to tropical fish stock assessment' (Sparre & Venema 1998). Many of the same examples and datasets featured therein are included in the package (Table 1) and documented in accompanying help files, which facilitates use in training and teaching. Finally, output from various functions can be passed to plotting functions, allowing for export as publication-quality figures.

Data-poor stock assessment using `TropFishR`

For historical reasons, and the link to the above-mentioned book by Sparre & Venema (1998), the package's name reflects the fact that the methods have often been applied to tropical fisheries, although they are equally applicable to other regions with data-poor stocks for which LFQ data is available. Typically, the workflow of a data-poor stock assessment with LFQ data would include: (i) estimation of biological stock characteristics (growth and natural mortality), (ii) fisheries performance aspects (exploitation rate and selectivity), and (iii) stock size and status. The order of the methods is important as they build upon each other in a sequential way. If some or all of the vital parameters for stock assessment are already known, the user may skip the data-poor approaches for their assessment and can directly

proceed with yield modelling applications. For a full tutorial of fisheries analysis with LFQ data, see the tutorial vignette: `vignette("tutorial", package = "TropFishR")`. Here, we demonstrate the application of a selection of the functions of Table 1 with the use of a generated LFQ dataset with known parameters. The data simulation procedure is based on an individual-based model that creates length-at-age variability through differences in individual growth parameters (for further information see supplementary online material and Taylor 2015). The generated LFQ data (Fig. 1a)

Table 2. Parameters used for data generation of length-frequency data

Abbreviation	Parameter	Value
L_{inf}	Infinite length	33·7 cm
K	Growth coefficient	0·47 year^{-1}
repro_wt	Indicating time of year of reproduction	0·25
C	Intensity of seasonality	0
ts	Summer point	0
a, b	Coefficients of allometric length-weight	0·017 [cm, g]
	Relationship	3·042
L_m	Length at maturity	25·3 cm
L_{50}	Length at first capture	10 cm
wqs	Width of selectivity ogive	3 cm
M	Natural mortality	0·8 year^{-1}
F	Fishing mortality	1·3 year^{-1}
rmax	Maximum recruitment level of Beverton and Holt's recruitment model	10 000
beta	Steepness of Beverton and Holt's recruitment model	1

For more information about specific parameters refer to the help file of the `lfqGen` function of the fishdynr package (`help(lfqGen)`, Taylor 2015).
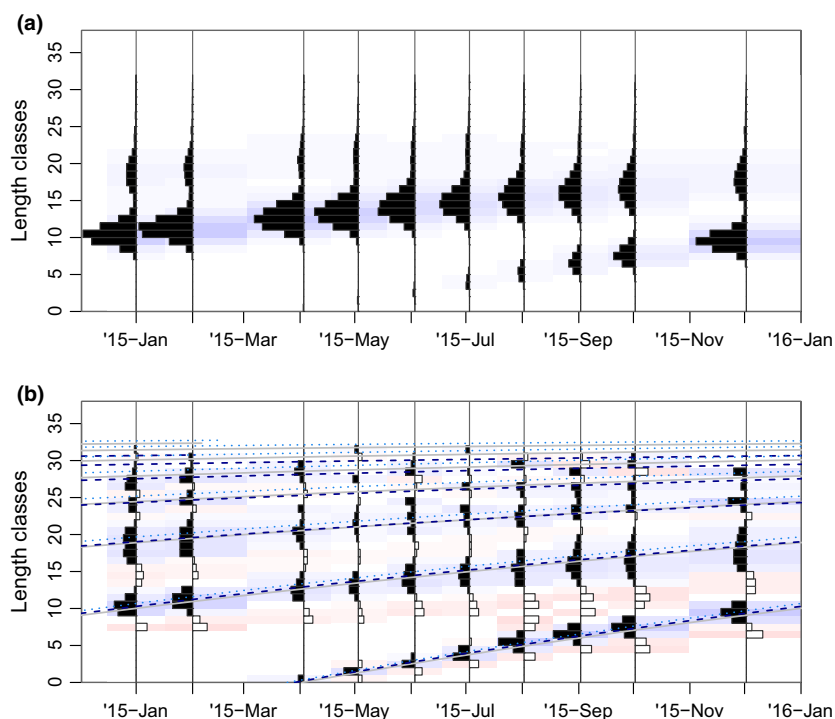
Fig. 1. Length frequency data and growth curves. LFQ data visualised in terms of (a) catches and (b) restructured data with a moving average setting of MA = 5. Lines show estimated (blue) and true (grey) growth curves plotted through the length frequency data. The dashed dark blue and dotted light blue curves represent the curves of the traditional ELEFAN method and ELEFAN_SA, respectively.

is based on life history traits of the Thumbprint Emperor *Lethrinus harak* (Forsskål, 1775); parameters represent the average of estimates provided in FishBase (Froese & Pauly 2016) (Table 2), while the length-at-first-maturity is the average of results estimated by Kulmiye, Ntiba & Kisia (2002). The resulting LFQ dataset is provided in **TropFishR** under the name 'synLF6'. It is structured into 32 length classes with a bin size of 1 cm and 10 sampling dates within 1 year.

AVAILABLE METHODS

ELEFAN

ELEFAN is a method to derive growth parameters of the von Bertalanffy growth function (VBGF) from LFQ data (Pauly 1980). It requires a vector with the mid-lengths of defined length classes, a matrix with catches in numbers per length class (rows) and per sampling time (columns), and a vector with the dates of the sampling times (see synLFQ6).

ELEFAN can be separated into three steps: (i) 'restructuring' of LFQ data according to a procedure that scores length bins based on deviations from a moving average across neighbouring bins, (ii) calculation of the cumulative score for a given set of VBGF parameters based on the bin scores that are intersected by resulting growth curves, and (iii) search for VBGF parameters that result in the maximum score value. The following code shows how to restructure the data and visualise the positive and negative bins (compare Fig. 1b). The arguments 'MA' (moving average) and 'addl.sqrt' (additional square root transformation) allow for further adjustments to the default restructuring process.

```
lfqbin <- lfqRestructure(param = synLFQ6, MA = 5, addl.sqrt = TRUE)
plot(lfqbin)
```

For the third step, traditional search procedures include: (i) response surface analysis (RSA) varies both the growth coefficient (K) and asymptotic length (L_{inf}) of the VBGF, and (ii) K-Scan, which holds L_{inf} constant while varying K (Pauly 1986). In addition to these traditional methods, two new optimisation procedures are offered that search over all parameters simultaneously: ELEFAN_SA and ELEFAN_GA, which are based on simulated annealing and genetic algorithms, respectively (Taylor & Mildenberger 2017). Here, we compare the performance of the RSA and the simulated annealing approach by the use of the generated dataset. Both methods require the same LFQ data list and some confinement in the search space for the growth parameters. There are many recommendations on how to define initial estimates of L_{inf}, such as using the maximum length class (or the average of the few largest classes) as a reference (Taylor 1958; Beverton 1963), but it is notable that large length classes can be underrepresented due to high exploitation, high natural mortality (i.e. high M/K ratio), or selectivity patterns of the gear, e.g. dome-shaped selection ogive of gillnets. Another approach is to use the Powell–Wetherall method (Wetherall, Polovina & Ralston 1987), which fits a linear regression to $\bar{L} - L'$ against L', where $\bar{L}$ is the mean length of all individuals larger than L' and L' is any length equal and above the length under full exploitation. Here, we used this approach to estimate L_{inf} (ca. 36 cm, see supplementary online material) and, subsequently, restricted the search space range to ±5 cm. Without prior information, the search for K may be done on a relatively wider range of values (e.g. 0·1–1·0).

```
ELEFAN(x = synLFQ6, MA = 5, addl.sqrt = TRUE,
       Linf_range = seq(30,40,0.5), K_range = seq(0.1,1,0.1))
```

When using RSA, we recommend analysing several local maxima of the score function with a finer resolution for both parameters and also comparing the calculated score values and fit graphically (comparable to Fig. 1b, see tutorial vignette for automated algorithm).

The simulated annealing approach requires the user to define the maximum time of the optimisation procedure (argument 'SA_time'; here 2 min). The graph produced by this method can help to find an adequate computing time (graph not shown here). The extent of stochasticity in the search process can be varied with the argument 'SA_temp', representing the initial 'temperature' of the annealing process.

```
ELEFAN_SA(x = synLFQ6, SA_time = 60*2, SA_temp = 6e5,
          MA = 5, addl.sqrt = TRUE,
          low_par  = list(Linf = 30, K = 0.1, t_anchor = 0),
          init_par = list(Linf = 35, K = 0.5, t_anchor = 0.5),
          up_par   = list(Linf = 40, K = 1,   t_anchor = 1))
```

If prior information about the maximum age is available the argument 'agemax' can be used to fix the number of fitted cohorts, otherwise the maximum age is set to the age when expected length is $0.95 L_{inf}$. The ELEFAN methods cannot estimate t_0 (theoretical age at length zero) from LFQ data alone (Pauly 1980). However, this parameter does not influence results of the length-based assessment methods included in **TropFishR** (catch curve, virtual population analysis [VPA], and yield per recruit model) and can thus be set to zero (Mildenberger, unpublished). Furthermore, the ELEFAN methods in this package do not return starting points as FiSAT II users might be used to; instead, they return the parameter called 't_anchor', which describes the fraction of the year where yearly repeating growth curves cross length equal to zero; for example a value of 0·25 refers to April 1st of any year (Taylor & Mildenberger 2017).

Both ELEFAN methods were able to reproduce true parameters used for data generation quite accurately (Table 3). Both scores are higher than the score of the original parameters, which can be attributed to the noise introduced in the data simulation process. The fit of estimated growth parameters can also be explored visually (Fig. 1b) and shows high similarity to the true growth curves (grey) and a good fit through the positive scored peaks of the restructured LFQ data.

```
plot(lfqbin, Fname = "catch", date.axis = "modern")
plot(lfqbin, Fname = "rcounts", date.axis = "modern")
lt1 <- lfqFitCurves(lfq = lfqbin, par = list(Linf=33.7, K=0.47,
    t_anchor=0.25), draw = TRUE, col = "grey", lty = 1, lwd=1.5)
lt2 <- lfqFitCurves(lfq = lfqbin, par = list(Linf=32.5, K=0.5,
    t_anchor=0.24), draw = TRUE, col = "darkblue", lty = 2, lwd=1.5)
lt3 <- lfqFitCurves(lfq = lfqbin, par = list(Linf=33.9, K=0.45,
    t_anchor=0.23), draw = TRUE, col = "dodgerblue2", lty = 3, lwd=1.5)
```

An advantage of the ELEFAN_SA and ELEFAN_GA approaches is the possibility to estimate all parameters of the seasonalised VBGF (soVBGF) simultaneously (with the additional argument: `seasonalised = TRUE`). Until now, this was only possible to a limited extent within the 'Automatic Search' approach of FiSAT II, and required 'understanding of how to conduct searches in multidimensional space, in the presence of multiple local optima' (Gayanilo & Pauly 1997). With appropriate settings, the alternate optimisation approaches show promise in estimating the parameters of the soVBGF (Taylor & Mildenberger 2017). As an illustration of the expandability that an R-based implementation affords, the confidence intervals for the growth parameters can be easily calculated with the jack knife technique – a resampling method with replacement (Quenouille 1956; Tukey , 1958, 1986), – by incorporating ELEFAN_SA in a for-loop (supplementary online material).

Cohort analysis

Jones' length-converted cohort analysis (CA; Jones 1984) – a modification of Pope's VPA for LFQ data – estimates the trend of the fishing mortality across length classes of the stock as well as a reconstruction of overall stock size. Beside the growth and mortality parameters, CA also requires the parameters '*a*' and '*b*' of the allometric length–weight relationship, which can be calculated from weight-at-length data. The function 'lfqModify' allows for the addition of a 'plus group' to the data (i.e. the largest group includes all sizes above a given value), changing the bin size, and/or transforming the catch matrix to a single catch vector (per year) as required by CA (with the argument 'vectorise_catch').

```
synLFQ6 <- lfqModify(lfq = synLFQ6, vectorise_catch = TRUE)

# growth parameters estimated by 'ELEFAN_SA' (see above)
synLFQ6$Linf <- 33.9
synLFQ6$K <- 0.45
synLFQ6$t_anchor <- 0.23
```

Table 3. Growth parameters used for data generation (True), compared to values and scores of the response surface analysis (RSA) and ELEFAN with simulated annealing (ELEFAN_SA)

Parameter	True	RSA	ELEFAN_SA
L_{inf} in cm	33·7	32·5 (−3·7%)	33·9 (0·6%)
K in year^{-1}	0·47	0·5 (6%)	0·45 (−4·4%)
phi$'$	2·73	2·72 (−0·4%)	2·71 (−0·7%)
t_{anchor}	0·25	0·24 (−4·2%)	0·23 (−8·7%)
Score	0·703	0·781	0·773

Values in brackets show the deviation from true values in percentage terms. The score value of the true parameters can be calculated with lfqFitCurves (more information in help(lfqFitCurves)). Score value can only be compared relatively, changing MA or the bin size can influence the score values (for more information see Taylor & Mildenberger 2017).

```
# natural mortality estimated by 'M_empirical' (not shown here)
synLFQ6$M <- 0.72
# total and fishing mortality estimated by 'catchCurve'
(not shown here)
synLFQ6$Z <- 1.78
synLFQ6$FM <- 1.06
# length-weight parameters from FishBase
synLFQ6$a <- 0.000017
synLFQ6$b <- 3.042
```

Cohort analysis further requires an estimate of the terminal fishing mortality (argument 'terminal_F'), representing the fishing mortality for the largest length class (see Hilborn & Walters (1992) for a discussion of this parameter). Since the catches by length class should represent an entire year, the argument 'catch_corFac' can be used to add a correction factor in cases where under-sampling occurred. Assuming that the catch vector is representative for all sampled months and that catch levels were similar for the two missing months in the example data, we set 'catch_corFac' to $1 + 2/12$ in order to correct for the under-reporting.

```
CA <- VPA(param = synLFQ6, terminalF = 1.06,
          analysis_type = "CA",
          plot=TRUE, catch_corFac = (1+2/12))
```

The graphical results in Fig. 2 generally reflect the logistic-shaped fishing pattern used in the data generation (red line in Fig. 2a,b), although larger classes seem to be underrepresented in the catches, which results in lower fishing mortality values. The overall stock size in numbers and biomass can be calculated via the following:

```
sum(CA$annualMeanNr, na.rm = TRUE)
sum(CA$meanBiomassTon, na.rm = TRUE) * 1e3
```

According to this method, the size of the stock is around 78 435 individuals or 2217 kg. These values deviate from the true values of 78 863 individuals and 1858 kg by -0.5% and 16% respectively. The deviations might here be related to the value of the terminal fishing mortality, which besides natural mortality and the length-weight conversion, has been shown to be an influential input parameter for VPA (Pelletier 1990; Hilborn & Walters 1992).

Yield per recruit modelling

Yield per recruit models (YPR), also called prediction models (e.g. Thompson and Bell model), allow for the evaluation of the stock's status in relation to reference levels as well as the impact of a given management control measure, such as regulations to fishing effort or gear selectivity (e.g. net mesh size). These models require growth and mortality parameters (comparable to CA) as well as selectivity information about the gear. In the first application of the YPR model, selectivity is represented by the vector of fishing mortality by length class estimated by the previous CA:

```
synLFQ6$FM<-CA$FM_calc
```

Holding this selectivity pattern constant, a vector of fishing mortality ('FM_change') is used to simulate changes in fishing intensity scaled to the fishing mortality of the fully-exploited length class(es). In the second application of the YPR model, the selectivity characteristics are described by a more generalised trawl-like selection ogive[1] (argument 's_list'), which allows for change in selectivity characteristics ('Lc_change') simultaneously with fishing mortality ('FM_change') or exploitation rate ('E_change'), respectively. The resulting estimates are presented as an isopleth graph showing yield per recruit isoclines (Fig. 3b). By setting the argument 'stock_size_1' to 1, all results are per recruit. If the number of recruits (recruitment to the fishery) is known, the exact yield and biomass can be estimated. The arguments 'curr.E' (fishing mortality divided by total mortality) and 'curr.Lc' (estimated with the length-converted catch curve) allow for the estimation and visualisation of yield and biomass (per recruit) values for current fishing patterns (Fig. 3).

```
TB1 <- predict_mod(param = synLFQ6, type = "ThompBell",
             FM_change = seq(0,5,0.1),  stock_size_1 = 1,
             curr.E = 0.6)

TB2 <- predict_mod(param = synLFQ6, type = "ThompBell",
             s_list = list(selecType = "trawl_ogive",
                       L50 = 9.65, L75 = 10.83),
             FM_change = seq(0,5,0.1), Lc_change = seq(3,30,0.1),
             stock_size_1 = 1, curr.E = 0.6, curr.Lc = 9.65)
```

The plotting functions allow control over details of the graphs:

```
plot(TB1, mark = TRUE)
plot(TB2, type = "Isopleth", xaxis1 = "E", mark = TRUE,
             contour = 6)
```

For this example, Fig. 3a shows that the yield could be increased when fishing mortality is reduced. The gear related analysis reveals that the current gear characteristics and exploitation rate produce a yield of above 25 g per recruit, which could be increased to above 30 g by increasing the mesh size (following vertical dashed line in Fig. 3b upwards).

```
# biological reference levels
TB1$df_Es
# current yield and biomass levels
TB1$currents
```

The results estimate that the current fishing mortality ($F = 1.08$) is higher than the fishing mortality for 'maximum sustainable yield' ($F_{MSY} = 0.9$), indicating overfishing.

---

[1]The selectivity parameters in this example were estimated by the length-converted catch curve analysis with simultaneous estimation of the selection ogive (`help(catchCurve)`).
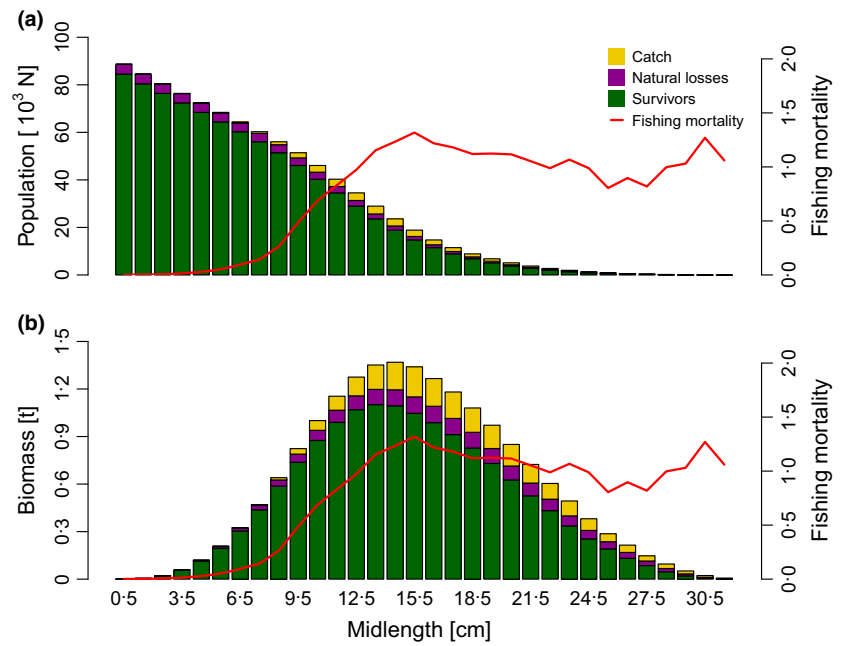
Fig. 2. Results of Jones' length-based cohort analysis (CA) with the reconstructed population structure (survivors, natural losses and catch) in numbers (a) and biomass (b) per length class and the fishing mortality rate by length class (red line in both plots).
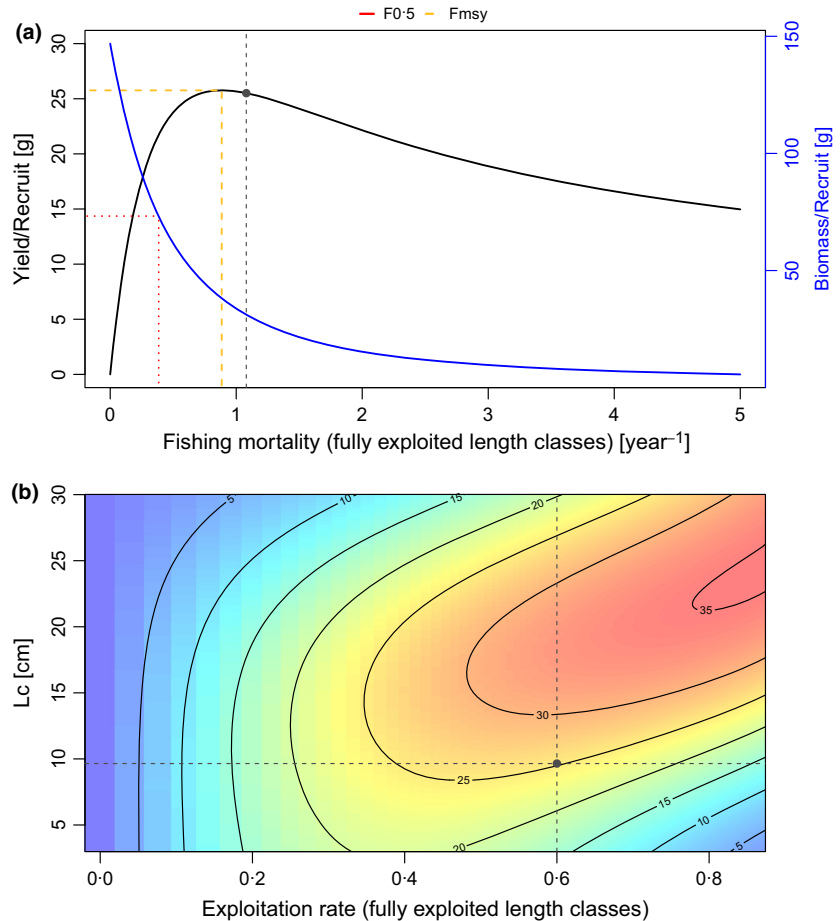
Fig. 3. Results of the Thompson and Bell model with the Thumbprint Emperor data: (a) Curves of yield and biomass per recruit with a L_c of 9·65 cm. The black dot represents yield and biomass per recruit under current fishing pressure. The yellow and red dashed lines represent fishing mortality for maximum sustainable yield (Fmsy) and fishing mortality associated with a 50% reduction relative to the virgin biomass ($F_{0·5}$), respectively. The x-axis corresponds to the fishing mortality of the fully exploited length class(es). (b) Exploration of impact of different exploitation rates and L_c values on the yield per recruit. The black dots represent the current (over-)fishing regime. The x-axis displays the exploitation rate (F/Z) of the fully-exploited length class(es).

Final remarks

The R package **TropFishR** offers a new toolbox compiling fish stock assessment methods for data-poor fisheries, which are often, but not exclusively, encountered in the tropics. The package allows for the estimation of life history parameters and the development of management measures based on the analysis of catch-at-age data, bulk catch and effort data, or length-frequency data (Table 1). **TropFishR** and other length-based assessment methods (e.g. s6model, LBSPR) have the advantage of being applicable to data collected from relatively short time periods (e.g. 1 year), making them attractive for rapid assessments in data-poor situations. To the contrary, surplus production models typically require longer time series of 10 years or more. LFQ data may also originate from survey or catch data, making collection fairly easy and economical even for small-scale fisheries. When derived from catch data, only a representative subsample of the landings is required and data collection requires little more than a measuring tape. At present, the ubiquity of smart phones may both ease the collection of LFQ data and allow for supplementary information to be collected in parallel (i.e. 'metadata'); e.g. equipping the monitoring team with a smart phone application that automatically photo-measures fish, geo-references the sampling location and uploads the data directly into a shared database. The examples of this paper illustrate the accuracy of parameter estimation using a simulated dataset with known values. Nevertheless, further testing of the sensitivity of length-based methods is warranted, as has been done for related age-based approaches for VPA or YPR (Pelletier 1990). This further testing is needed to help identify influential input parameters and allow further quantification of uncertainty. The transparent and flexible toolbox offered by the **TropFishR** package should facilitate this work and allow for further development of LFQ methods.

Authors' contributions

All authors conceived the design of **TropFishR**; T.K.M. and M.H.T. implemented and programmed the R package; T.K.M. led the writing of the manuscript. All authors contributed critically to the drafts, tested and verified the software, and gave final approval for publication.

Acknowledgements

Availability

The examples presented in this paper are based on **TropFishR** (v1.1.3), which requires R >=3.0.0 and can be downloaded from GitHub as follows:

```
install.packages("devtools")
devtools::install_github("tokami/TropFishR", ref = "v1.1.3")
```

Data acessibility

The example data 'synLFQ6' is included in the package and can be loaded with `load(synLFQ6)`. The code for the generation of this dataset is given in the supplementary online material. The data can also be downloaded from GitHub at https://github.com/tokami/TropFishR/releases/tag/v1.1.3 (DOI: 10.5281/zenodo.495176).

References

Beverton, R.J. (1963) Maturation, growth and mortality of clupeid and engraulid stocks in relation to fishing. *Rapport Conseil International Exploration de la Mer*, **154**, 44–67.

Carruthers, T. & Hordyk, A. (2016) DLMtool: data-limited methods toolkit. R package version 3.2.2. Available at: https://cran.r-project.org/package=DLM-tool (accessed 16 February 2017).

Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O. & Lester, S.E. (2012) Status and solutions for the world's unassessed fisheries. *Science (New York, N.Y.)*, **338**, 517–20.

Froese, R. & Pauly, D. (2016) FishBase. World Wide Web electronic publication. Available at: http://www.fishbase.org (accessed 14 June 2016).

Gayanilo, F.C. & Pauly, D. (1997) FAO-ICLARM stock assessment tools Reference manual stock assessment tools Reference manual. 262.

Gayanilo, F.C.Jr, Sparre, P. & Pauly, P. (1996) FiSAT: FAO-ICLARM stock assessment tools. User's manual. FAO.

Hilborn, R. & Walters, C.J. (1992) *Quantitative Fisheries Stock Assessment*. Chapman and Hall, New York, NY, USA.

Hordyk, A., Ono, K., Prince, J.D. & Walters, C.J. (2016) A simple length-structured model based on life history ratios and incorporating size-dependent selectivity: application to spawning potential ratios for data-poor stocks. *Canadian Journal of Fisheries and Aquatic Sciences*, **13**, 1–22.

ICES. (2012) ICES implementation of advice for data-limited stocks in 2012 in its 2012 advice. In: ICES CM 2012/ACOM 68. 42.

Jones, R. (1984) Assessing the effects of changes in exploitation pattern using length composition data (with notes on VPA and cohort analysis). FAO Fisheries Technical Paper, 118p.

Kell, L., Mosqueira, I., Grosjean, P.*et al.* (2007) FLR: an open-source framework for the evaluation and development of management strategies. *ICES Journal of Marine Science*, **64**, 640–646.

Kokkalis, A., Thygesen, U.H., Nielsen, A. & Andersen, K.H. (2015) Limits to the reliability of size-based fishing status estimation for data-poor stocks. *Fisheries Research*, **171**, 4–11.

Kolding, J., Bene, C. & Bavinck, M. (2014) Small-scale fisheries: Importance, vulnerability and deficient knowledge. *Governance of Marine Fisheries and Biodiversity Conservation: Interaction and Coevolution* (eds S.M. Garcia, J. Rice & A. Charles), John Wiley & Sons, Ltd., Chichester, UK.

Kulmiye, A. J., Ntiba, M.J. & Kisia, S.M. (2002) Some aspects of the reproductive biology of the thumbprint emperor, *Lethrinus harak* (Forsskål, 1775), in Kenyan coastal waters. *Western Indian Ocean Journal of Marine Science*, **1**, 135–144.

Millar, R.B. & Holst, R. (1997) Estimation of gillnet and hook selectivity using log-linear models. *ICES Journal of Marine Science*, **54**, 471–477.

Nelson, G.A.N. (2016) fishmethods: fishery science methods and models in R. R package version 1.10-0. Available at: https://CRAN.R-project.org/package=fishmethods (accessed 4 November 2016).

Ogle, D. (2016) FSA: fisheries stock analysis. R package version 0.8.9. Available at: https://cran.r-project.org/web/packages/FSA (accessed 16 February 2017).

Pauly, D. (1980) On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. *Journal du Conseil*, **39**, 175–192.

Pauly, D. (1986) On improving operation and use of the ELEFAN programs. Part II. Improving the estimation of L_{inf}. *Fishbyte*, **4**, 18–20.

Pelletier, D. (1990) Sensitivity and variance estimators for virtual population analysis and the equilibrium yield per recruit model. *Aquatic Living Resources*, **3**, 1–12.

Quenouille, M.H. (1956) Notes on bias in estimation. *Biometrika*, **43**, 353–360.

Skern-Mauritzen, M., Ottersen, G., Handegard, N.O., Huse, G., Dingsør, G.E., Stenseth, N.C. & Kjesbu, O.S. (2015) Ecosystem processes are rarely included in tactical fisheries management. *Fish and Fisheries*, **17**, 165–175.

Sparre, P. & Venema, S. (1998) *Introduction to Tropical Fish Stock Assessment. Part 1. Manual.* FAO fisheries technical paper. No. 306. 1, Rev. 2. FAO, Rome, Italy.

Taylor, C.C. (1958) Cod growth and temperature. *Journal du Conseil*, **23**, 366–370.

Taylor, M.H. (2015) fishdynr: an R package of fisheries science related population dynamics models. R package version 0.4. Available at: https://github.com/marchtaylor/fishdynr (accessed 21 October 2016).

Taylor, M.H. & Mildenberger, T.K. (2017) Extending electronic length frequency analysis in R. *Fisheries Management and Ecology*, 1–9. doi:10.1111/fme.12232.

Tukey, J. (1958) Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, **29**, 614.

Tukey, J. (1986) The future of processes of data analysis. *The Collected Works of John W. Tukey – Philosophy and Principles of Data Analysis: 1965–1986*, vol. 4 (ed. L.V. Jones), pp. 517–549. Wadsworth & Brooks/Cole, Monterey, CA, USA.

Wetherall, J.A., Polovina, J.J. & Ralston, S. (1987) Estimating growth and mortality in steady-state fish stocks from length-frequency data. *ICLARM Conference Proceedings*, **13**, 53–74.

Supporting Information

Details of electronic Supporting Information are provided below.

Appendix S1. Additional R-scripts that demonstrate aspects of the paper in further detail, including: generation of synthetic LFQ data, Powell-Wetherall plot for estimation of Linf, Jack knife technique for the calculation of confidence intervals for estimated growth parameters using ELEFAN.