# Direct Survival Analysis: a new stock assessment method

EDUARDO FERRANDIS and PILAR HERNÁNDEZ

Departamento de Ciencias del Mar y Biología Aplicada, Universidad de Alicante, España. Carretera S. Vicente s/n.
Alicante 03080. E-mail: Eduardo.Ferrandis@ua.es

SUMMARY: In this work, a new stock assessment method, Direct Survival Analysis, is proposed and described. The parameter estimation of the Weibull survival model proposed by Ferrandis (2007) is obtained using trawl survey data. This estimation is used to establish a baseline survival function, which is in turn used to estimate the specific survival functions in the different cohorts considered through an adaptation of the separable model of the fishing mortality rates introduced by Pope and Shepherd (1982). It is thus possible to test hypotheses on the evolution of survival during the period studied and to identify trends in recruitment. A link is established between the preceding analysis of trawl survey data and the commercial catch-at-age data that are generally obtained to evaluate the population using analytical models. The estimated baseline survival, with the proposed versions of the stock and catch equations and the adaptation of the Separable Model, may be applied to commercial catch-at-age data. This makes it possible to estimate the survival corresponding to the landing data, the initial size of the cohort and finally, an effective age of first capture, in order to complete the parameter model estimation and consequently the estimation of the whole survival and mortality, along with the reference parameters that are useful for management purposes. Alternatively, this estimation of an effective age of first capture may be obtained by adapting the demographic structure of trawl survey data to that of the commercial fleet through suitable selectivity models of the commercial gears. The complete model provides the evaluation of the stock at any age. The coherence (and hence the mutual "calibration") between the two kinds of information may be analysed and compared with results obtained by other methods, such as virtual population analysis (VPA), in order to improve the diagnosis of the state of exploitation of the population. The model may be drawn up in a deterministic format, but the main concepts may be interpreted as expectations if stock and catch are considered as stochastic processes.

Keywords: marine population dynamics, stock assessment, survival analysis, mortality models.

RESUMEN: ANÁLISIS DIRECTO DE SUPERVIVENCIA: UN NUEVO MÉTODO DE EVALUACIÓN DE STOCKS. – En el presente trabajo se propone y elabora un nuevo método para la evaluación de poblaciones marinas: DSA, o análisis directo de supervivencia. Utilizando datos proporcionados por las campañas de prospección de arrastre, se obtiene la estimación de los parámetros del modelo de supervivencia de Weibull propuesto por Ferrandis (2007) en un trabajo precedente. Dicha estimación permite establecer una función global de supervivencia y, a partir de ella, la supervivencia relativa a las distintas cohortes consideradas en el período de estudio a través de una adaptación del modelo separable de la mortalidad por pesca introducido por Pope y Shepherd. Y ello permite contrastar hipótesis acerca de la evolución de la supervivencia durante el período así como identificar posibles tendencias en el reclutamiento. Se establece un nexo entre este análisis de los datos procedentes de campañas de prospección directa y los datos procedentes de las capturas comerciales estructurados por edades, generalmente obtenidos para la evaluación de la población por métodos analíticos. Ello permite la estimación de la supervivencia correspondiente a las capturas comerciales, el tamaño inicial de la cohorte y la edad efectiva de primera captura. Lo que completa la estimación de parámetros relativos a la mortalidad y supervivencia y los parámetros de referencia utilizados en la gestión de recursos marinos explotados. Esta estimación de una edad efectiva de primera captura puede obtenerse, alternativamente, relacionando la estructura demográfica proporcionada por las campañas de prospección de arrastre con la relativa a la flota comercial a través de modelos adecuados de la  selectividad de las artes utilizadas por la flota comercial. El modelo proporciona la estimación del stock a cualquier edad de la cohorte. Y la coherencia (por tanto, la mutua calibración) entre las dos fuentes de información puede ser descrita y analizada, así como la comparación de los resultados con los obtenidos por otros métodos (VPA), para profundizar en el diagnóstico acerca del estado de explotación de la población. La elaboración del modelo puede establecerse en un enfoque determinista. Pero todos sus conceptos pueden ser interpretados como valores esperados si el stock y la captura se consideran como procesos estocásticos en la línea desarrollada por Ferrandis (2007).

Palabras clave: dinámica de poblaciones marinas, evaluación de stocks, análisis de supervivencia, modelos de mortalidad.

INTRODUCTION

Results produced by trawl surveys have often been criticised on the grounds of their lack of robustness based on their high variability. However, the experimental data relative to size and ages of some important target species in the Spanish Mediterranean have shown an excellent statistical behaviour in comparison with the proposed probabilistic models. This reinforces the capabilities of trawl surveys as fundamental sources of information for the management of marine resources.

The present work describes new approaches and models for interpreting the global evolution of the mortality and survival of an exploited marine resource. The modelling of survival allows a bridge to be built between the two classical and often conflicting methods: direct and indirect evaluation.

In this first extension of the classical models, the scenario is restricted to the assumption of a constant exploitation rate from the age of first capture that is coherent with mixed Weibull models for lifetime distribution.

In the Weibull model proposed in this work, four parameters are used to establish the mortality throughout the entire life of the exploited resource: the mortality coefficients, M and Z, of the natural and total mortality rates respectively, the exponent $\alpha$ (the shape of the lifetime distribution) and the effective age of first capture, $t_c$. In this work, the authors do not impose artificial conditions on either the age (whose range is $[0, \infty)$) or the stock (which is finite and bounded, even for age 0), nor do they use an assumed terminal fishing mortality rate or the mortality model to produce a mortality vector. On the contrary, we demonstrate that one can work directly with the Continuous Survival Model. The name Direct Survival Analysis is therefore not artificial at all.

The authors emphasise the role of the Weibull distribution in fishery science without expecting it to be generally applicable. The only condition for its use should be the goodness of fit with respect to the empirical results. The Weibull distribution here applied, is a natural extension of the generally accepted exponential model underlying the classical Virtual Population Analysis (Gulland, 1965) and its further modifications (Separable VPA by Pope and Shepherd, 1982, and XSA by Shepherd, 1991, 1999). It produces a natural adaptation of the stock and catch equations that makes it possible to "rewrite" the theory and computational tools of pop-

ulation dynamics in a relatively easy way, therefore, understandable to all the fisheries scientist who are familiar with widespread software such as FISAT (Gayanilo and Pauly, 1997), VIT (Lleonart and Salat 1997), VPA Suite (Darby and Flatman, 1994). It also provides an admissible model of mortality rate that makes it possible to consider the stochastic character of the survival, and hence, to incorporate the likelihood methodology (which is strong when one is dealing with large samples as is generally the case in fishery applications) and the very powerful tools of survival analysis developed in many biological and epidemiological studies.

Thus, the present work not only provides a new tool for evaluating exploited marine populations, but further provides new approaches and perspectives to marine population dynamics.

Some initial steps of this promising method will be presented in future applications recommended by the recently created Permanent Working Group on Stock Assessment Methodology (PWGAM) of the General Fisheries Commission of the Mediterranean (GFCM).

## 1. ESTIMATING THE MODEL WITH TRAWL SURVEY DATA

### 1.1. Trawl survey data and the proportional hazard survival model

The so-called "direct methods", and in particular trawl surveys, are evaluation procedures independent of the fishery activity and hence not affected by the questionable trustworthiness and lack of coherence of catch data provided by fleets and/or fishery administrations belonging to different countries (or even regions in the same country). On the contrary, they allow a temporal coordination and standardization of gears and sampling effort and methods between different research vessels. They can thus be used to establish historical series of coherent data that are suitable for identifying spatio-temporal trends and are very useful for both theoretical research and management purposes.

The assumed survival model has been proposed by Ferrandis (2006) as a battery of Weibull distributions corresponding to the unexploited phase subjected only to natural mortality (with a coefficient of mortality $M$) and the exploited phase subjected to total mortality (with a higher mortality coefficient $Z=M+F$) from an age of first capture, $t_c$:

$$\left. \begin{aligned} & if\ t \leq t_c \Rightarrow S(t) = \Pr\left[T \geq t\right] = e^{-Mt^\alpha} \\ & if\ t \geq t_c \Rightarrow S(t) = \Pr\left[T \geq t\right] = e^{-Mt_c^\alpha} e^{-Z(t^\alpha - t_c^\alpha)} \Leftrightarrow \\ & \qquad \Leftrightarrow S(t)\,/\,S(t_c) = e^{-Z(t^\alpha - t_c^\alpha)} \end{aligned} \right\}$$

$$(1.1.1)$$

The exponent $\alpha$ is the main parameter related to the shape of the mortality, so it is natural to assume that this exponent is quite stable for a given species and a given region. Therefore, it could be considered as a population parameter.

## 1.2. The growth equation

In general, surveys of marine resources provide size-frequency distributions that may be converted into age distributions through the inverse growth equation if adequate estimations of its parameters are available.

$$t = t_0 - \frac{1}{k}\ln(1 - \frac{l}{L_\infty}) \qquad (1.2.1)$$

where $l$ is the size at age $t$ and $t_0$, $k$, $L_\infty$ are the Von Bertalanfy growth equation parameters. Otherwise, the conversion from size to age may be performed if a table of transition probabilities between size and age classes is available.

## 1.3. The identification of a representative age interval (baseline interval)

Most of the demersal resources investigated in trawl surveys are not fully recruited to the gear used and/or to the trawled area. Because of this, for most of the resources the youngest individuals are not present in the capture, and older individuals are often underrepresented in the sample. Therefore, it is necessary to identify an age interval $[t_1, t_2]$ for which it can be assumed that the obtained sample is representative of the population and which should be contained in the exploited phase, that is to say, in the second interval of ages considered by the Weibull model (1.1.1). In the identification of this baseline interval, the analysis of the age and size distribution with the estimation of their first mode and the knowledge of the size range caught by the fishery fleet may be helpful.

Assuming that the equivalence of the demographic structure in the sampled baseline interval is representative of the population is equivalent to assuming that the mortality rate caused by the survey is proportional to the general mortality in that interval (Cox, 1959) and also that, in this interval, the catchability of the sample gear used in the survey should be independent of age.

## 1.4. Estimation of the conditional survival in the selected interval: the baseline survival

In the probabilistic framework presented by Ferrandis (2006), the estimation of the survival restricted to the baseline established interval may be considered as a conditional probability (denoted by "|")

$$for\ t \in [t_1, t_2] \Rightarrow S(t\,|\,t_1 \leq T \leq t_2) =$$
$$= \Pr\left[T \geq t\,|\,t_1 \leq T \leq t_2\right] \qquad (1.4.1)$$

This conditional survival, $S(t|t_1 \leq T \leq t_2)$, is estimated by the empirical survival function, that is, the quotient between the number of observed ages greater than $t$ and less than $t_2$ and the total number of observed ages in the baseline interval. This can be expressed formally by:

$$\hat{S}(t\,|\,t_1 \leq T \leq t_2) = \frac{1}{n}\sum_i I_{[t,t_2]}(t_i) \qquad (1.4.2)$$

where n is the number of observed ages in the interval $[t_1, t_2]$ and $I_{[t,t_2]}$ is the indicator function of the interval $[t,t_2]$, defined, for any set A, by

$$I_A(x) = \left\{ \begin{aligned} & 1\ if\ x \in A \\ & 0\ otherwise \end{aligned} \right\}$$

In general, this estimation will be obtained from the historical series of trawl surveys data restricted to the baseline interval and on a cohort basis: i.e. for the estimation of this survival function a cohort, or a set of cohorts, will be considered. Also, these estimations will be obtained on a catch per unit effort basis because the survey effort is not constant during the study period. This is easily done by weighting the frequency of the ages by the particular and measured trawled area (multiplying the frequency by the mean swept area and dividing by the area swept on each trawl).

However, if in this interval the sample is representative of the population, the conditional survival is related to the total survival in the following way:

$$S(t\,|\,t_1 \leq T \leq t_2) =$$
$$= \Pr\left[T \geq t\,|\,t_1 \leq T \leq t_2\right] = \frac{S(t) - S(t_2)}{S(t_1) - S(t_2)} \qquad (1.4.3)$$

Dividing by the survival at the initial age of the interval, $S(t_1)$

$$S(t \mid t_1 \le T \le t_2) = \frac{S(t)/S(t_1) - S(t_2)/S(t_1)}{S(t_1)/S(t_1) - S(t_2)/S(t_1)} =$$

$$= \frac{S(t)/S(t_1) - S(t_2)/S(t_1)}{1 - S(t_2)/S(t_1)} \qquad (1.4.4)$$

and, according to the assumed model, the quotients only depend on the two parameters $Z$ and $\alpha$.

$$S(t \mid t_1 \le T \le t_2) =$$

$$= \frac{e^{-Z(t^{\alpha} - t_1^{\alpha})} - e^{-Z(t_2^{\alpha} - t_1^{\alpha})}}{1 - e^{-Z(t_2^{\alpha} - t_1^{\alpha})}} \qquad (1.4.5)$$

Then, the estimation of the parameters of the model, $Z$ and $\alpha$, may be obtained from the estimation of the conditional survival. This can be achieved in two ways: by non-linear regression, taking the unit values as the seed, and alternatively and complementarily by maximum likelihood.

In fact, the density function corresponding to the conditional survival is:

$$f(t \mid t_1 \le T \le t_2) = \frac{Z\alpha t^{(\alpha-1)} e^{-Z(t^{\alpha} - t_1^{\alpha})}}{1 - e^{-Z(t_2^{\alpha} - t_1^{\alpha})}} \qquad (1.4.6)$$

The log-likelihood is thus:

$$l = n\left[\ln(Z) + \ln(\alpha)\right] + (\alpha - 1)\sum_i \ln(t_i) - Z\sum_i \left(t_i^{\alpha} - t_1^{\alpha}\right) -$$

$$- n\ln\left[1 - e^{-Z\left(t_2^{\alpha} - t_1^{\alpha}\right)}\right]$$

$$(1.4.7)$$

whose derivatives

$$\begin{cases} \dfrac{\partial l}{\partial Z} = 0 \\[2mm] \dfrac{\partial l}{\partial \alpha} = 0 \end{cases} \qquad (1.4.8)$$

provide a system of two non-linear equations with the two unknowns $Z$ and $\alpha$.

However, as is shown in the annex, an explicit equation relates both parameters:

$$Z = \frac{A + B(\alpha) - C(\alpha)}{D(\alpha) - E(\alpha)C(\alpha)} \qquad (1.4.9)$$

where

$$\begin{cases} A = \dfrac{\sum_i \ln(t_i)}{n}; \quad B(\alpha) = \dfrac{1}{\alpha}; \quad C(\alpha) = \dfrac{t_2^{\alpha}\ln(t_2) - t_1^{\alpha}\ln(t_1)}{t_2^{\alpha} - t_1^{\alpha}} \\[4mm] D(\alpha) = \dfrac{\sum_i \left(t_i^{\alpha}\ln(t_i) - t_1^{\alpha}\ln(t_1)\right)}{n}; \quad E(\alpha) = \dfrac{\sum_i \left(t_i^{\alpha} - t_1^{\alpha}\right)}{n} \end{cases}$$

$$(1.4.10)$$

the sum being extended on all the observed ages in the baseline interval.

Hence, the system of equations (1.4.8) is reduced to one non-linear equation with the single unknown $\alpha$, which can be solved computationally by the "regula falsi" method. (Hildebrand, 1974; Curtis, 1994). The authors have developed suitable software to obtain both the non-linear regression (least squares estimation) and the maximum likelihood estimation with a prefixed precision.

These parameter estimation methods provide an estimate of the survival in the exploited phase, which, in accordance with usual concepts in survival analysis, will be called the estimated "baseline survival" (Smith, 2002).

Likewise, the estimation of the baseline survival makes it possible to specify the natural mortality, as described in Ferrandis (2006). For an assumed constant natural mortality rate $M_0$ and an estimated exponent $\widehat{\alpha}$, the corresponding coefficient of natural mortality is

$$M = \left(\frac{M_0 \Gamma_{(1/\alpha)}}{\alpha}\right)^{\alpha} \qquad (1.4.11)$$

## 1.5. Estimation of the survival in the different cohorts

In order to specify the survival for each of the cohorts considered from the general baseline survival, an adaptation of the Separable Model (Pope and Shepherd, 1982) is proposed. The separability hypothesis expresses the fishing mortality rates as a product of age selection effects correlated with the exploitation pattern and temporal effects correlated with the intensity of exploitation. In this adaptation, the temporal effects will be set as a cohort effect which will become a year effect, when a pseudocohort is considered.

Let

$$Z_0 = M + F \qquad (1.5.1)$$

be the mortality coefficient corresponding to the baseline survival, which provides the mortality rate

$$Z_0(t) = (M + F)\alpha t^{\alpha-1} =$$

$$= M\alpha t^{\alpha-1} + F\alpha t^{\alpha-1} = M(t) + F(t) \qquad (1.5.2)$$

and that corresponding to the i-th cohort

$$Z_i = M + k_i F \qquad (1.5.3)$$

$$Z_i(t) = (M + k_iF)\alpha t^{\alpha-1} =$$
$$= M\alpha t^{\alpha-1} + k_iF\alpha t^{\alpha-1} = M(t) + k_iF(t) \quad (1.5.4)$$

In this context, $M(t) = M\alpha t^{\alpha-1}$ is the natural mortality rate, $F(t) = F\alpha t^{\alpha-1}$ the age selection effect of the fishing mortality rate and $k_i$ the cohort effect.

With the following change of parameter

$$k_i^* = \frac{M + k_iF}{M + F} \Leftrightarrow k_i = \frac{k_i^*(M+F) - M}{F} \quad (1.5.5)$$

the conditional survival for a specific cohort and the baseline survival are related as:

$$\frac{S_i(t)}{S_i(t_1)} = e^{-Z_i(t^\alpha - t_1^\alpha)} = e^{-(M+k_iF)(t^\alpha - t_1^\alpha)} =$$

$$= e^{-k_i^*(M+F)(t^\alpha - t_1^\alpha)} =$$

$$= \left(e^{-(M+F)(t^\alpha - t_1^\alpha)}\right)^{k_i^*} = \left(\frac{S_0(t)}{S_0(t_1)}\right)^{k_i^*} \quad (1.5.6)$$

and the mortality rates as

$$Z_i(t) = M(t) + k_iF(t) = (M + k_iF)\alpha t^{\alpha-1} =$$
$$= k_i^*(M+F)\alpha t^{\alpha-1} = k_i^* Z_0(t) \quad (1.5.7)$$

Hence, the separable model may be characterised by the relation between the specific mortality rates and the baseline mortality that corresponds to the proportional hazard model (Smith, 2002) considered in the Survival Analysis framework. The specific survival of a particular cohort is a power of the baseline survival.

As in the case of the baseline survival, the estimation of the proportional parameter $k_i^*$ may be obtained from the estimated conditional baseline survivals in two ways: by non-linear regression, taking the unit value as the seed, and alternatively and complementarily by maximum likelihood. The likelihood solution is

$$\frac{\partial l}{\partial k_i^*} = 0 = \frac{1}{k_i^*} - Z_0 \frac{\sum_i (t_i^\alpha - t_1^\alpha)}{n} - \frac{e^{-k_i^* Z_0\left(t_2^\alpha - t_1^\alpha\right)} Z_0\left(t_2^\alpha - t_1^\alpha\right)}{1 - e^{-k_i^* Z_0\left(t_2^\alpha - t_1^\alpha\right)}}$$
$$(1.5.8)$$

where the sum is extended to the n ages of the cohort considered in the selected baseline interval.

Again, this non linear equation with the single unknown, $k_i^*$, can be solved computationally by the "regula falsi" method. (Hildebrand, 1974, Curtis, 1994). As with other required computing facilities, the authors have developed adequate and specific software to obtain the solution of (1.5.8).

## 1.6. Hypotheses on the evolution of survival

An increasing/decreasing trend on $\{k_i^*\}$ corresponds to a decreasing/increasing trend on survival $\{S_i\}$. The hypothesis of a stable survival during the period may be tested by likelihood ratio test:

$$\lambda = \frac{L(S)}{\prod_i L(S_i)} \quad (1.6.1)$$

$L(S)$ being the likelihood corresponding to the baseline survival and the whole data set of ages in the baseline interval, and $L(S_i)$ the likelihood corresponding to the $i$-th cohort.

The statistic $-2\ln(\lambda)$ follows asymptotically a chi-squared distribution with $(c-1)$ degrees of freedom, $\chi^2_{(c-1)}$, c being the number of cohorts considered.

## 1.7. Identification of recruitment trends

Splitting the baseline interval into a partition of subintervals $\{[t_i, t_{i+1}), i = 1...k\}$ the catch equation will provide the expected number of individuals obtained in each subinterval of the cohort:

$$E\left[C(t_i, t_{i+1})\right] = E\left[N(t_i)\right](F/Z)(1 - S(t_{i+1})/S(t_i)) =$$
$$= E\left[N(t_i)\right](F/Z)(1 - e^{-Z(t_{i+1}^\alpha - t_i^\alpha)})$$

where the expected number of individuals at the initial age of the interval, $E[N(t_i)]$, is related to the recruitment as:

$$E\left[N(t_i)\right] = R(e^{-M(t_c^\alpha - t_r^\alpha)})e^{-Z(t_i^\alpha - t_c^\alpha)} =$$
$$= R(e^{-M(t_c^\alpha - t_r^\alpha)})e^{Zt_c^\alpha}e^{-Zt_i^\alpha}$$

where $t_r$ is the age at recruitment.

The particular coefficient of mortality of the cohort and the survival of the cohort, by the proportional hazard model, are

$$Z = k_i^* Z_0 \text{ and } S = S_0^{k_j^*}$$

and the fishing mortality rate produced by the survey is

$$F = \sum_h q_h A_h,$$

where $q_h$ is the catchability coefficient and $A_h$ is the area swept on trawl $h$. The catchability coefficient will be proportional to the corresponding accessibility.

If it can be assumed that the accessibility is stable in a geographical or bathymetric range, there

will be a constant catchability, $q_h = q$, in this range and in the considered baseline age interval.

Then, in general, the catch equation becomes

$$E\big[C(t_i, t_{i+1})\big]=$$

$$= qR(e^{-M(t_c{}^\alpha - t_r{}^\alpha)})e^{Zt_c{}^\alpha}e^{-Zt_i{}^\alpha}\frac{\sum_h A_h}{k_j{}^* Z_0}(1 - e^{-k_j{}^* Z_0(t_{i+1}{}^\alpha - t_i{}^\alpha)})$$

$$(1.7.1)$$

Now, taking the set of observed catches

$$Y = \{y_i = C(t_i, t_{i+1})\}$$

as the dependent variable and

$$X = \{x_i\} =$$

$$= \left\{ e^{-k_j{}^* Z_0(t_i{}^\alpha - t_c{}^\alpha)}\frac{\sum_h A_h}{k_j{}^* Z_0}(1 - e^{-k_j{}^* Z_0(t_{i+1}{}^\alpha - t_i{}^\alpha)}) \right\}$$

$$(1.7.2)$$

as the independent variable, the catch equation provides a linear regression $Y = (qe^{-M(t_c{}^\alpha - t_r{}^\alpha)}R)X$.

This is a regression without a constant term whose regression coefficient, $qe^{-M(t_c{}^\alpha - t_r{}^\alpha)}R$, will be proportional to the recruitment of the cohort, that is to say, it will constitute a recruitment index. Therefore, the trends in these regression coefficients reflect trends in the recruitment.

In order to evaluate expression (1.7.2), the independent variable of the regression, an estimation of the age of first capture, $t_c$, is needed. This may be obtained from the established management routines of the fishery (such as the age corresponding to the minimum permitted catch size) or from the estimation of an effective age of first capture by adapting the approach introduced by J. Shepherd (1983). This estimation will be developed in detail in the next paragraphs, which links the model with commercial landing data.

The significance of possible recruitment trends can be tested by modelling the recruitment trends. For instance, a linear trend during the period may be formalised as $R = a + bY_c$, $R$ being the recruitment and $c$ the year of the cohort. Substituting this trend in the expected catches (1.7.1), we have a multiple linear regression of the catches as the dependent variable over the two independent variables, $X$, and the product, $XY_c$ as the second predictor. The expect-

ed catches (1.7.1) are then rewritten as

$$Y = (qe^{-M(t_c{}^\alpha - t_r{}^\alpha)}R)X =$$

$$= (aqe^{-M(t_c{}^\alpha - t_r{}^\alpha)})X + (bqe^{-M(t_c{}^\alpha - t_r{}^\alpha)})XY_c$$

Therefore, the significance of the coefficient $bqe^{-M(t_c{}^\alpha - t_r{}^\alpha)}$ of the predictor $XY_c$ implies the significance of the parameter $b$ of the considered linear trend in recruitment.

## 2. LINKING THE MODEL WITH COMMERCIAL LANDING DATA

### 2.1. Age structured commercial catch data

Commercial fleet catches provide biomass and density data for different species. When this information is complemented with biological samplings, size classes can be transformed into age classes which form the basis for the application of age-structured stock assessment models. A data series of successive years can be the source for a cohort analysis. If the number of years is low, an alternative is to analyse pseudo-cohorts (age structure of one-year catches) as if they represented the evolution of a cohort assuming equilibrium conditions.

In general, the catch-at-age data will be considered as a set of age intervals $\{[t_{i-1}, t_i); i = 1,...,k\}$ and a corresponding set of catch in number $\{C_i; i = 1,...,k\}$ for each age interval. In the most complete case, the initial and final ages will be: $t_0 = 0$ and $t_k = \infty$.

### 2.2. Parameter estimation

We use the baseline survival already estimated from trawl surveys, with parameters $Z_0$ and $\alpha$, in order to estimate the survival functions that correspond to the cohorts (or pseudo-cohorts) given by commercial catch-at-age data, considering the same species in the same or neighbouring regions in overlapping periods. Their parameters in the exploited phase should be $k^* Z_0$ and $\alpha$, where $k^*$ is a proportionality constant.

In order to estimate the proportionality constant, $k^*$, we identify the first age, $t_p$, in the set $\{t_0, t_1,...,t_k\}$ contained in the baseline interval in which the estimation of the survival and its parameters has been made. Let $N_p$ be the population size at this age cor-

responding to the cohort (or pseudo-cohort) that provides the catch-at-age data. That is: $N(t_p) = N_p$.

## 2.3. Estimation of $N_p$ and $k^*$

By applying the stock and catch equations proposed by Ferrandis (2006) to the successive age, $t_i$, such that $t_p \leq t_i \leq t_{k-1}$, we have the following conditional expectations:

$$E\left[N(t_i) \mid N(t_p)\right] =$$
$$= N(t_p)\frac{S(t_i)}{S(t_p)} = N(t_p)e^{-Z(t_i^\alpha - t_p^\alpha)} \qquad (2.3.1)$$

$$E\left[C(t_i, t_{i+1}) \mid N(t_p)\right] =$$
$$= N(t_p)\frac{(Z-M)}{Z}\frac{S(t_i)}{S(t_p)}\left(1 - \frac{S(t_{i+1})}{S(t_i)}\right) \qquad (2.3.2)$$

with

$$Z = k^* Z_0 \qquad (2.3.3)$$

Combining (2.3.1), (2.3.2) and (2.3.3) we obtain

$$E\left[C(t_i, t_{i+1}) \mid N(t_p)\right] =$$
$$= N(t_p)e^{-k^* Z_0(t_i^\alpha - t_p^\alpha)}\left(\frac{k^* Z_0 - M}{k^* Z_0}\right)\left(1 - e^{-Z(t_{i+1}^\alpha - t_i^\alpha)}\right)$$

$$(2.3.4)$$

where the unknowns are $N_p$ and $k^*$.

In general, catch-at-age data, $\{C(t_i, t_{i+1}); t_p \leq t_i \leq t_{k-1}\}$, allow the estimation of the parameters $N_p$ and $k^*$ from this set of catches and the expressions (2.3.4) by non linear regression, i.e. by least squares criteria. The estimates may be the initial values or complementary for the application of maximum likelihood as a method of estimation using the available stochastic models for the stock and the catch processes (Ferrandis, 2006). And the estimation of $k^*$ yields the estimation of the total mortality coefficient $Z = k^* Z_0$.

## 2.4. Estimation of an effective age of first capture

One way to estimate the effective age of first capture may be, following the work of Shepherd (1983), to consider the total catch of the cohort (pseudo-cohort), $C(0,\infty)$, from an effective age of first capture, $t_c$. The catch equation (2.3.2) becomes

$$C(0,\infty) = C(t_c,\infty) \approx E\left[C(t_c,\infty)\right] =$$
$$= (Z-M)E\left[N(t_c)\right]/Z \qquad (2.4.1)$$

and the stock equation (2.3.1) applied to the interval $[t_c, t_p]$ yields

$$E\left[N(t_p)\right] = E\left[N(t_c)\right]e^{-Z(t_p^\alpha - t_c^\alpha)} \Leftrightarrow$$
$$\Leftrightarrow E\left[N(t_c)\right] = E\left[N(t_p)\right]e^{Z(t_p^\alpha - t_c^\alpha)} \quad (2.4.2)$$

Substituting (2.4.2) in (2.4.1) we have

$$C(0,\infty) = C(t_c,\infty) \approx \frac{Z-M}{Z}N(t_p)e^{Z(t_p^\alpha - t_c^\alpha)} \quad (2.4.3)$$

which provides the estimation of an effective age of first capture

$$t_c \approx \left[t_p^\alpha - \frac{1}{Z}\ln\left[\frac{ZC(0,\infty)}{(Z-M)N(t_p)}\right]\right]^{1/\alpha} . \quad (2.4.5)$$

Alternatively, we can focus our attention on the early age catches, $C(t_c, t_p)$, obtained in the interval $[t_c, t_p]$. Again, applying the stock equations (2.3.1) and (2.4.2) and the catch equation (2.3.2), we obtain

$$C(0,t_p) = C(t_c,t_p) \approx E\left[C(t_c,t_p)\right] =$$
$$= \frac{Z-M}{Z}E\left[N(t_p)\right]\left(e^{Z(t_p^\alpha - t_c^\alpha)} - 1\right) \quad (2.4.6)$$

which yields

$$t_c \approx \left[t_p^\alpha - \frac{1}{Z}\ln\left[1 + \frac{ZC(0,t_p)}{(Z-M)N(t_p)}\right]\right]^{1/\alpha} . (2.4.7)$$

Finally, considering the two intervals $[0,t_p)$ and $[t_p,\infty)$, by (5.3.2)

$$C(t_p,\infty) \approx (Z-M)N(t_p)/Z \qquad (2.4.8)$$

and combining (2.4.6) and (2.4.8), we have the expression

$$\frac{C(0,t_p)}{C(t_p,\infty)} \approx e^{Z(t_p^\alpha - t_c^\alpha)} - 1 . \qquad (2.4.9)$$

as a ratio between the catches in the two intervals. This yields a new estimation of the age of first capture:

$$t_c \approx \left[t_p^\alpha - \frac{1}{Z}\ln\left[1 + \frac{C(0,t_p)}{C(t_p,\infty)}\right]\right]^{1/\alpha} . (2.4.10)$$

The above expressions may be computed from the information of the total or representative catches of

the fishery fleet or, alternatively, using selectivity studies of the commercial gears. The selectivity curve assigns a probability of catch depending on the size or the age of the individuals. Thus, weighting the individuals caught in the trawl survey by this probability we "translate" the demographic structure obtained in the survey to the demographic structure of the commercial landings. We use these weighted frequencies to apply expressions (2.4.5), (2.4.7) and (2.4.10).

## 2.5. Estimation of the initial population size

Once the age of first capture has been estimated, the expected catch until the first age considered in the analysis, $t_p$, is given by (2.4.6). This will fit the observed catch exactly if equation (2.4.7) is used to estimate the effective age of first capture.

By the stock equation (2.3.1), the initial size of the considered cohort (pseudo-cohort) will be obtained from the previous estimates:

$$E\left[N(t_p)\right] = E\left[N(t_c)\right]e^{-Z(t_p^\alpha - t_c^\alpha)} =$$
$$= N_0 e^{-M(t_c^\alpha)}e^{-Z(t_p^\alpha - t_c^\alpha)} \quad (2.5.1)$$

and therefore

$$N_0 = N(0) = E\left[N(t_p)\right]e^{M(t_c^\alpha) + Z(t_p^\alpha - t_c^\alpha)} \quad (2.5.2)$$

## 2.6. Goodness of fit

The estimation of the parameters from a non-linear regression on the model (2.3.2) makes it possible to calculate confidence intervals for these estimates and the correlation between them, as well as the coefficient of determination of the model, which will give a measure of its goodness of fit. The graphs of the observed versus the expected catches (given by (2.3.2)) will provide a visual description of the fit.

## 2.7. Stock evaluation

The survival model established (1.1.1) and its relation to the size of the population

$$E[N(t)] = N(0)S(t)$$

provide the estimation of the expected population in number of the considered cohort at any age. This evaluation may be compared with those obtained by other methods, such as virtual population analysis (VPA).

## 3. CONCLUSIONS

The identification of the baseline age interval (an interval in which the catches obtained in the trawl survey may be assumed to be representative of the studied cohorts of the population) makes it possible to estimate the parameters of the proposed Weibull model by maximum likelihood, and therefore to estimate the baseline survival during the study period.

The separable model of fishing mortality rates (Pope and Shepherd, 1984) has been adapted. Thus, the survival/mortality of each specific cohort may be expressed as a power/multiple of the baseline survival/mortality. Using this approach, the evolution of the survival and mortality of the successive cohorts during the study period becomes a particular case of the proportional hazard models considered in the general survival analysis framework. The exponent/factor of the power/multiple relationship may be estimated by non-linear regression or maximum likelihood.

The model makes it possible to test hypotheses about survival and mortality, and in particular the stability of survival, during the study period by the likelihood ratio method.

The extension of the catch equation established provides the expression of the catches in successive age subintervals as a linear regression on a covariate which depends on the survival. The regression coefficients are proportional to the recruitment of the cohort considered. The possible trends in these regression coefficients therefore reflect trends in the recruitment, whose statistical significance may be quantified.

The basic estimation of the baseline survival obtained from trawl survey data may be completed by analysing commercial catch-at-age data. This can help to check the coherence of the two kinds of information as well as the relationship between their respective survival and reference parameters.

The actual and potential uses of the proposed methodology include the following:

- Estimation of a general survival ("baseline survival") from a set of trawl surveys.

- Use of proportional hazard models related to the separable model of fishing mortality rate to estimate the specific survival for the various cohorts in the study period.

- Identification of trends in the age structure through the evolution of life expectancy.

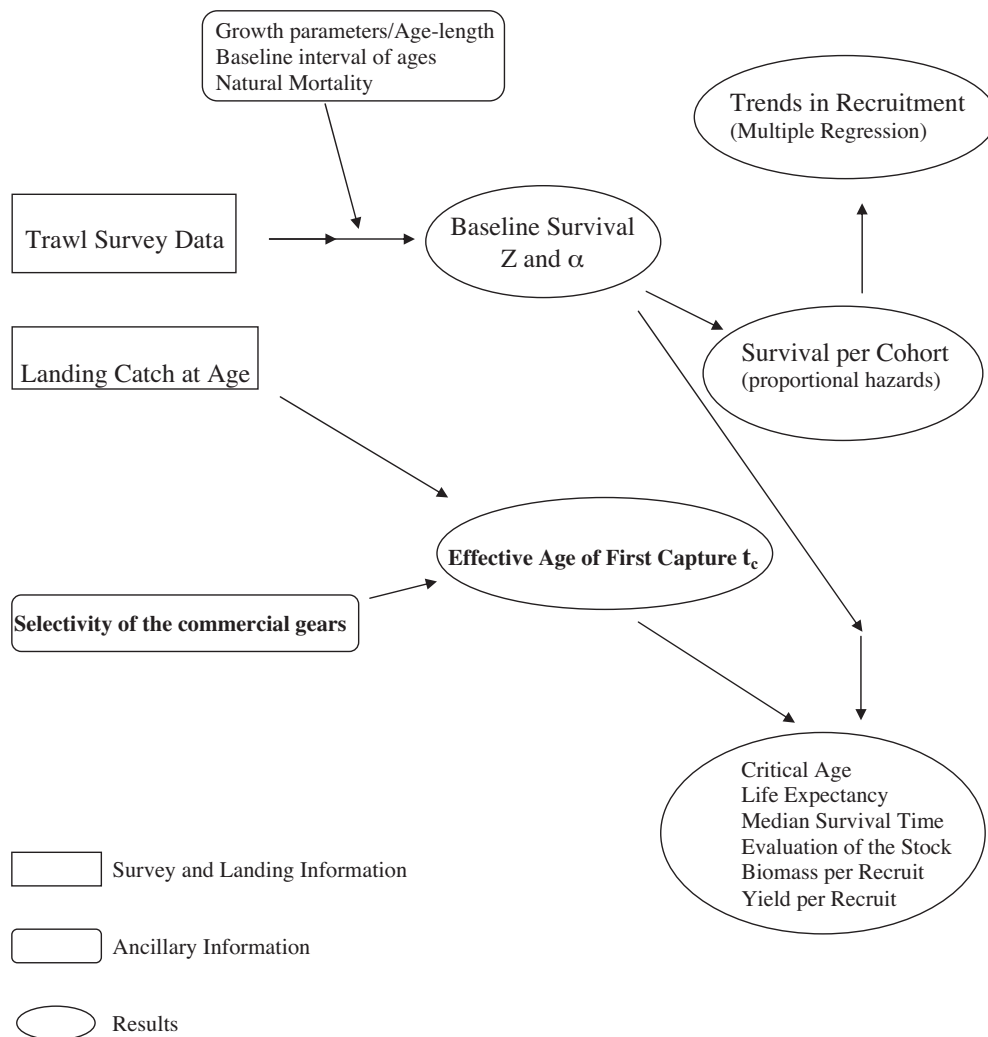- Use of the survival models to identify and test possible recruitment trends.

FIG. 1. – Flow chart of the application of the Direct Survival Analysis method in its current state of development.

- Verifying the coherence between the estimated baseline survival and other trawl surveys or age-structured data provided by indirect evaluations (VPA). This allows the comparison of the age structure corresponding to different regions and periods.

- Using the survival models to evaluate the stock without applying an arbitrary terminal fishing mortality rate.

- Establishing diagnoses relative to the management of the resource throughout the age-structure parameters (life expectancy) and the estimation of reference parameters derived from the mortality rates, the biomass per recruit and the yield per recruit.

The flow-chart of the operational tasks structuring the application of the proposed methodology is shown in Figure 1. These tasks are executed through a set of computer programs that constitute the implementation of the numerical methods described in the present work. The method is being applied to a set of important target demersal species, mainly in the Mediterranean Sea.

## REFERENCES

Cox, D.R. – 1959. The analysis of exponentially distributed life-times with two types of failure. *J. R. Stat. Soc. Ser. B (Methodological)*, 21(2): 411-421.

Curtis, F.G. and O.W. Patrick. – 1994. *Applied Numerical Analysis*. Addison –Wesley, New York.

Darby C.D. and S. Flatman. – 1994. Virtual Population Analysis: version 3.1 (Windows/DOS) user guide, Ministry of Agriculture, Fisheries and Food, Directorate of Fisheries Research Information Technology Series Number 1. *Info. Tech. Ser., MAFF Direct. Fish. Res.,* Lowestoft, (1): 85 pp.

Gulland, J.A.- 1965. Estimation of mortality rates. Annex to Arctic fisheries working group report ICES C.M./1965/D:3. (mimeo).

Reprinted as pp. 231-241. In: P.H. Cushing (ed), Key papers on fish populations. Oxford. IRL Press. 1983.

Ferrandis, E. – 2007. On the stochastic approach to marine population dynamics. *Sci. Mar.*, 71(1): 153-174.

Gayanilo, F.C. Jr. and D. Pauly (eds.). – 1997. The FAO-ICLARM Stock Assessment Tools (FiSAT) Reference Manual. FAO Computerized Information Series (Fisheries). No. 8. Rome, FAO, 262 p.

Hildebrandt, F.B. – 1974. *Introduction to Numerical Analysis.* McGraw-Hill, New York.

Lleonart, J. and J. Salat. – 1997. VIT: software for fishery analysis. User's manual. FAO Computerized Information Series (Fisheries). No. 11. Rome, FAO. 1997. 105 p.

Pope, J.G. and J.G. Shepherd. – 1982. A simple method for the consistent interpretation of catch-at-age data. *J. Cons. Expl. Mer.*, 40: 176-184.

Shepherd, J.G. – 1983. Two measures of overall fishing mortality. *J. Cons. Expl. Mer,* 41: 76-80.

Shepherd, J.G. – 1991. Simple methods for short term forecasting of catch and biomass. *ICES J. Mar. Sci.*, 48: 67-68.

Shepherd, J.G. – 1999. Extended survivors analysis: an improved method for the analysis of catch-at-age data and abundance indices. *ICES J. Mar. Sci.*, 56: 584-591.

Smith, P.J. – 2002. *Analysis of Failure and Survival Data.* Chapman & Hall / Crc., New York.

## APPENDIX

### Likelihood estimation of the Weibull survival model

The logarithmic transform ("ln") of the density function (4.4.6) of the observed ages in the selected interval is:

$$\ln\left(f\left(t \mid t_1 \leq T \leq t_2\right)\right) = \ln\left(\frac{Z}{1 - e^{-Z\left(t_2^{\alpha} - t_1^{\alpha}\right)}}\right) - Z\left(t^{\alpha} - t_1^{\alpha}\right) + \ln(\alpha) + (\alpha - 1)\ln(t) \qquad (A.1)$$

which provides the following log-likelihood function

$$l = \sum_i \ln\left(f\left(t_i \mid t_1 \leq T \leq t_2\right)\right)$$

$$l = n\ln\left(\frac{Z}{1 - e^{-Z\left(t_2^{\alpha} - t_1^{\alpha}\right)}}\right) - Z\sum_i \left(t_i^{\alpha} - t_1^{\alpha}\right) + n\ln(\alpha) + (\alpha - 1)\sum_i \ln(t_i) \,, \qquad (A.2)$$

whose derivative with respect to the mortality coefficient $Z$ is

$$\frac{\partial l}{\partial Z} = \frac{n}{Z} - \frac{n\left(t_2^{\alpha} - t_1^{\alpha}\right)}{e^{Z\left(t_2^{\alpha} - t_1^{\alpha}\right)} - 1} - \sum_i \left(t_i^{\alpha} - t_1^{\alpha}\right) \qquad (A.3)$$

and therefore

$$\frac{\partial l}{\partial Z} = 0 \iff \frac{1}{Z} - \frac{\left(t_2^{\alpha} - t_1^{\alpha}\right)}{e^{Z\left(t_2^{\alpha} - t_1^{\alpha}\right)} - 1} - \overline{\left(t^{\alpha} - t_1^{\alpha}\right)} = 0 \qquad (A.4)$$

which implies that

$$\frac{1}{e^{Z\left(t_2^{\alpha} - t_1^{\alpha}\right)} - 1} = \frac{1 - Z\overline{\left(t^{\alpha} - t_1^{\alpha}\right)}}{Z\left(t_2^{\alpha} - t_1^{\alpha}\right)} \qquad (A.5)$$

The derivative with respect to the exponent is

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} + \sum_i \ln(t_i) - Z\overline{\left(t^{\alpha}\ln(t) - t_1^{\alpha}\ln(t_1)\right)} - \frac{Z\left(t_2^{\alpha}\ln(t_2) - t_1^{\alpha}\ln(t_1)\right)}{e^{Z\left(t_2^{\alpha} - t_1^{\alpha}\right)} - 1} \,, \qquad (A.6)$$

which should be cancelled in the maximum likelihood

$$\begin{cases} \dfrac{\partial l}{\partial \alpha}=0 \Leftrightarrow \\[2mm] \dfrac{1}{\alpha}+\overline{\ln(t)}-Z\overline{\left(t^{\alpha}\ln(t)-t_1^{\alpha}\ln(t_1)\right)}-\dfrac{Z\left(t_2^{\alpha}\ln(t_2)-t_1^{\alpha}\ln(t_1)\right)}{e^{Z(t_2^{\alpha}-t_1^{\alpha})}-1}=0 \end{cases}. \tag{A.7}$$

Substituting (A.5) in (A.7) we have

$$\begin{cases} \dfrac{1}{\alpha}+\overline{\ln(t)}-\dfrac{\left(t_2^{\alpha}\ln(t_2)-t_1^{\alpha}\ln(t_1)\right)}{\left(t_2^{\alpha}-t_1^{\alpha}\right)}= \\[3mm] Z\left[\overline{\left(t^{\alpha}\ln(t)-t_1^{\alpha}\ln(t_1)\right)}-\overline{\left(t^{\alpha}-t_1^{\alpha}\right)}\dfrac{\left(t_2^{\alpha}\ln(t_2)-t_1^{\alpha}\ln(t_1)\right)}{\left(t_2^{\alpha}-t_1^{\alpha}\right)}\right] \end{cases}, \tag{A.8}$$

which is equivalent to

$$Z(\alpha)=\frac{A+B(\alpha)-C(\alpha)}{D(\alpha)-E(\alpha)C(\alpha)}, \tag{A.9}$$

where

$$\begin{cases} A=\dfrac{\sum_i \ln(t_i)}{n}; \quad B(\alpha)=\dfrac{1}{\alpha}; \quad C(\alpha)=\dfrac{t_2^{\alpha}\ln(t_2)-t_1^{\alpha}\ln(t_1)}{t_2^{\alpha}-t_1^{\alpha}} \\[4mm] D(\alpha)=\dfrac{\sum_i \left(t_i^{\alpha}\ln(t_i)-t_1^{\alpha}\ln(t_1)\right)}{n}; \quad E(\alpha)=\dfrac{\sum_i \left(t_i^{\alpha}-t_1^{\alpha}\right)}{n} \end{cases}, \tag{A.10}$$

And finally, replacing $Z$ in equation (A.4) by the expression (A.9) we obtain a single equation with a single unknown $\alpha$

$$\frac{1}{e^{Z(\alpha)(t_2^{\alpha}-t_1^{\alpha})}-1}=\frac{1-Z(\alpha)\overline{\left(t^{\alpha}-t_1^{\alpha}\right)}}{Z(\alpha)\left(t_2^{\alpha}-t_1^{\alpha}\right)}, \tag{A.11}$$

The solution of this equation and the corresponding value of $Z$ given by equation (A.9) provide the likelihood estimations, $\widehat{Z}$ and $\widehat{\alpha}$, of the parameters of the proposed Weibull survival model.