4.4 CPUE and LPUE as relative abundance indices

Section 4.2.4 described sources of information about tuna fisheries and pointed out that these are mostly linked with commercial fisheries. This section comments on the use of fishery dependent data for landings or catch per unit effort (LPUE or CPUE respectively) as indices of relative abundance of fish. A concluding paragraph comments very briefly on fishery independent CPUE data.

CPUE is usually taken to be proportional to numbers of fish, N, in the stock present in an area:

$$CPUE = q.N$$

The constant of proportionality, q, is called the 'catchability'. The equation could be re-written with subscripts, l, to refer to specific length classes if required. Strong assumptions are inherent in the general relationship (Paloheimo and Dickie, 1964; Maunder and Punt, 2004), e.g.

- Mean CPUE is estimated for the same time period, depths, and geographic region as those supporting the N fish of the stock.
- q is constant under all fishing conditions.
- q does not vary with N.

The difference between LPUE and CPUE unfortunately creates further uncertainty if no information about discarding or other losses of fish at sea is available. When estimating abundance as an index based on mean LPUE or CPUE by time-area strata, it is necessary to consider many factors, e.g.

- whether fishing covered the same area as the stock;
- whether fishing covered the same depths as the stock;
- what the effects of migrations, both horizontally and vertically, would be on local abundance (or q);
- whether fish aggregate and become less catchable at low stock numbers; and
- whether the technologies and strategies being used by the fishing fleet are sufficiently stable to assume that q is constant. A gradual improvement in the fishing power of a vessel is often observed as the captain develops fishing skills, and as the vessel is fitted with better fish finding equipment and possibly more engine power etc. This is referred to as 'technical creep'.

Other potential specific issues for tuna include:

- seasonal migration effects on the CPUE data from a single nation;
- the effect of Fish Aggregation Devices (FADs) on CPUE;
- co-operation between different gear types when fishing on FADs[4];
- calculating CPUE as an index of population abundance for a schooling species;

[4] Purse seine vessels may hold a large school of tuna in place while baitboats take part of the fish. In this case, when the baitboats are sampled in port, the composition of their catch will be different from that achieved by baitboats under normal circumstances. The effort to catch these fish will also be different to that under normal circumstances. This might require the addition of a gear category for purse-seine co-operating baitboats, and the issue returns to the clear definition of the fishery to be sampled. Another example is the co-operation of purse seine vessels to search for and catch bluefin in the Mediterranean. CPUE of individual vessels are then inconsistent.

- calculating CPUE where tuna are caught for farming. Length and weight measured at market (e.g. Japan) will not be comparable to that of 'wild fish'.

Clear answers to these questions are seldom available so it will be necessary either to accept the proportionality assumption with great caution, or to undertake modelling to try to improve LPUE or CPUE as an index of abundance (Xiao *et al.*, 2004, part I). Regression trees offer another, less prescriptive, model based approach (Watters and Deriso, 2000).

Modelling of CPUEs is a research exercise. The predictor variables usually have to be selected from a long list of possibilities that should include the interactions among the variables (e.g. Rodríguez-Marín *et al.*, 2003). Omission of one important variable could cause the model to perform erratically when used to predict outside the time or space frame of the observations used to fit the model. Prior biological knowledge is the best guide for an initial selection of predictor variables which can subsequently be refined by statistical methods (Burnham and Anderson, 2002). An approach to avoid is that of stepwise selection through all available variables. This is because the statistical significance of a predictor with one set of data and one set of additional predictors will often change substantially when slightly different conditions prevail. The distribution of 'error' (=observed – fitted) values around the model has to be chosen from several statistical possibilities which include allowance for zero CPUE values (Ortiz and Arocha, 2004). The modelling method has to be chosen to suit the error distribution. The simplest situation is when log(CPUE) can be treated as approximately normally distributed around a linear model with zeros ignored; least squares linear regression methods, described in many textbooks, are then suitable. Other distributions, e.g. Poisson, would require a Generalised Linear model (McCullagh and Nelder, 1989). Non-linear relationships can be estimated with Generalised Additive models (Hastie and Tibshirani, 1990). They require a decision on the degree of flexibility to be allowed in the fitted curves, in addition to specification of the model function. Differential weighting of observations having different degrees of reliability is another consideration for modelling (Cotter and Buckland, 2004). A useful general summary of modelling theory in a fisheries context is by Venables and Dichmont (2004).

Given all this flexibility associated with modelling approaches to standardisation of LPUE and CPUE, it is **essential** that those reporting the results of modelling work to ICCAT should summarise all the choices and assumptions made and, so far as possible, explain the reasons for them. The resulting diagnostic plots (e.g. residual, QQ plots) should also be presented to demonstrate appropriate selection of model and error structure. General understanding of the foundations of a modelling study and of its strengths and weaknesses is of considerable assistance when weighing up the information it produces for the purposes of assessment and management of a stock.

LPUEs of fishing vessels can vary by orders of magnitude from set to set. As a result, it is important to use the right estimator for average LPUE in a time-area stratum. For simplicity, consider just two sets labelled $i = 1,2$ in which L_i fish were retained for landing following application of E_i units of fishing effort. Two different estimators for average LPUE are:

$$mean_1(LPUE) = \frac{(L_1/E_1) + (L_2/E_2)}{2} \tag{1}$$

and

$$mean_2(LPUE) = \frac{L_1 + L_2}{E_1 + E_2} \tag{2}$$

Suppose, for illustrative purposes, that contrasting catches of fish occurred such that $L_1 = 1$, $L_2 = 100$, $E_1 = 1$, and $E_2 = 2$. Then

$$mean_1(LPUE) = 25.5$$

and

$$mean_2(LPUE) = 33.66.$$

The first estimator is the unweighted average of the two point values of LPUE, one for each set. This estimator uses the information of which sets provided each pair of *L* and *E* values (cf. second bullet point under *Information*, above) and is the recommended estimator for average LPUE because each set, whatever the catch, is an equally valid observation of fishing success. In contrast, the second estimator gives more weight to the larger landing figure, L_2. This estimator has to be used when total landings and total effort for multiple sets are the only available data. The bias for the example data is +32%.

CPUE data from fishery independent sources such as surveys by research vessels or spotter planes may be available. The advantage of these is that they are not influenced by commercial decisions about fishing locations and times, or, if well standardised and documented in SOPs, by changes of fishing gear and technique over time. The disadvantages of such surveys are that they are unlikely to cover the whole area occupied by a stock, and that the degree of overlap may itself vary with season, migrations, and possibly from year to year. The design of the survey is also important. A systematic grid, for example, will be poor for finding fish when the stock is low and aggregated in small, localised concentrations that fall between the nodes of the grid. Generally, survey abundance indices are likely to have higher variance than mean LPUE values from a widespread commercial fishery. They are also likely to be biased due to the mismatch of locations of fish and survey observation points. Use of a time-series of survey results requires the strong assumption that the survey bias is constant over time.

4.4.1 Specific ICCAT issues

A growing issue concerns the overlap between the time-space ‘sampled’ by the gear and the time-space inhabited by the fish; does the degree of overlap change through time?

For bycatch species such as Atlantic white marlin, the only available time series of relative abundance are fishery-derived CPUE indices. Commercial indices come from wide-ranging fisheries, but these may have changed in spatial distribution, gear (moving from shallow to deeper longline sets) or target species over time (deeper set hooks indicate a change in targeting to bigeye tuna). Other CPUE data come from more localized sport fisheries that have always targeted marlins. Alternative GLM formulations have been put forward in an attempt to remove biases caused by changes in fishing depth over time in the fishery (Babcock and McAllister, 2004).

A further extension for billfish species is the application of 'habitat-based' standardisation models (Hinton and Nakano, 1996). 'Habitat-based' models incorporate understanding of behavioural (depth and temperature preferences) and oceanographic parameters to standardise historical CPUE time series data, as well as accounting for significant gear changes over time. The basic idea is that if a hook is fished in an environment that is preferred by the species, it has a higher probability of capturing that species (Hinton and Maunder, 2004). Bigelow *et al*. (2002) have used the habitat based standardization method to create CPUE based indices of relative abundance for bigeye and yellowfin tuna in the Pacific Ocean. These indices have been used for assessments in both the western-central Pacific Ocean by SPC (Hampton, 2002) and the eastern Pacific Ocean by the IATTC (Maunder, 2002). The debate over the use of 'habitat-based' standardisation models is on-going.

4.4.2 Further reading

Babcock, E.A. and McAllister, M.K. (2004). Modelling biases and contradictions among catch rate indices of abundance for Atlantic white marlin (*Tetrapterus albidus*). Col. Vol. Sci. Pap. ICCAT 56, 110-119.

Bigelow, K.A., Hampton, J., and Miyabe, N. 2002. Application of a habitat-based model to estimate effective longline fishing effort and relative abundance of Pacific bigeye tuna (*Thunnus obesus*). Fish. Oceanogr. 11(3): 143-155.

Burnham, K. P. and D. R. Anderson (2002). Model Selection and Multimodel Inference. New York, Springer

Cotter, A. J. R. and S. T. Buckland (2004). Using the EM algorithm to weight data sets of unknown precision when modeling fish stocks. Mathematical Biosciences 190: 1-7.

Hampton, J. 2002. Stock assessment of bigeye tuna in the western and central Pacific. Secretariat of the Pacific Community, Noumea, New Caledonia, SCTB 15, BET– 1, 37 p.

Hastie, T. J. and R. J. Tibshirani (1990). Generalized Additive Models. Boca Raton, Chapman & Hall/CRC.335

Hinton, M.G. and Nakano, H. (1996). Standardizing catch and effort statistics using physiological, ecological or behavioural constraints and environmental data, with an application to blue marlin (*Makaira nigricans*) catch and effort data from Japanese longline fisheries in the Pacific. Bull. IATTC 21(4), 171-200.

Hinton, M.G. and Maunder, M.N. (2004). Methods for standardising CPUE and how to select among them. Col. Vol. Sci. Pap. ICCAT 56, 169-177.

Maunder, M.N. (2002). Status of yellowfin tuna in the eastern Pacific Ocean. Inter-Amer. Trop. Tuna Comm. Stock Assessment Report 3: 47-134.

Maunder, M. N. and A. E. Punt (2004). Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70(2-3): 141-159.

McCullagh, P. and J. A. Nelder (1989). Generalized Linear Models. Boca Raton, Chapman & Hall/CRC.511

Ortiz, M. and F. Arocha (2004). Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery. Fish. Res. 70: 275-297.

Paloheimo, J. E. and L. M. Dickie (1964). Abundance and fishing success. Rapports et Proces- verbaux des Reunions, Conseil Permanent International pour L'Exploration de la Mer 155(28): 152-163.

Rodríguez-Marín, E., Arrizabalaga, H., Ortiz, M., Rodríguez-Cabello, C., Moreno, G. and Kell, L.T. (2003). Standardisation of bluefin tuna, Thunnus thynnus, catch per unit effort in the baitboat fishery of the Bay of Biscay (Eastern Atlantic). ICES J. Mar. Sci. 60, 1216-1231.

Venables, W. N. and C. M. Dichmont (2004). GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fish. Res. 70(2-3): 319-337.

Watters, G. M. and R. B. Deriso (2000). Catch per unit of effort of bigeye tuna: a new analysis with regression trees and simulated annealing. Bulletin of the Inter-American tropical Tuna Commission 21(8): 531-571.

Xiao, Y. S., A. E. Punt, et al., Eds. (2004). Models in fisheries research: GLMs, GAMs and GLMMs., Fish. Res. 70:137-428