DOES PARASITE INFECTION DEFINITELY INCREASE FOR SARDINE AGED 2 AND ABOVE ON THE SOUTH COAST?

Andrea Ross-Gillespie and Doug S. Butterworth
Marine Resource Assessment and Management Group (MARAM)
Department of Mathematics and Applied Mathematics
University of Cape Town, Rondebosch 7701, South Africa

Summary

The trend with length in parasite prevalence in sardine on the South Coast is modelled in a manner that isolates behaviour at larger lengths to enable a determination of whether the trend continues to increase from age one to ages of two and above. The results indicate that a continued increase is robustly confirmed at the 5% level of significance, and hence that there must be some movement of sardine of ages greater than one in at least one direction between the West and the South Coasts under the hypothesis that infection by the parasite can occur only on the West Coast.

Introduction

Data measuring prevalence of a parasite infection of sardine, when fit to a logistic model, indicate a continued upward trend in prevalence as length increases (van der Lingen and Hendricks, 2014; van der Lingen *et al.*, 2014; van der Lingen and Winker, 2014). This has important implications for two-stock models of the sardine population, the current version of which assume no movement of fish after their first birthday in either direction between the South and West Coasts. This assumption would be invalidated if prevalence on the South Coast continues to increase above this age, given the hypothesis that the parasite infection can be contracted only on the West Coast (van der Lingen and Hendriks, 2014).

The aim of this analysis is to investigate whether this upward trend is significantly different from zero for the South Coast over the length range where ages of 2 and older sardine predominate. It is possible that earlier results reported by van der Lingen and colleagues reflect the extrapolation of trends through data at lower lengths dominated by fish of age 0 and 1, which at larger lengths swamp any signal from the smaller samples of older fish. The method utilised for the South Coast therefore models the parasite prevalence dependence on length as a logistic up to length 18cm, and thereafter as a straight line (in contrast, for the West Coast, the dependence on length was modelled with a logistic for all lengths). The choice of 18 cm was based on inspection of length-at-age distributions for sardine, which suggested that the contributions of 1-year-old fish above this length would not be large (in November at least) (de Moor and Butterworth 2012, 2013). This approach was chosen because of its simplicity, compared to taking the mixture of 1-year-old and older fish at different lengths into account at a more detailed level, which would require incorporating this analysis within the assessment model itself to be able to take account of changing cohort strength and fishery impacts on age and hence length structure from year to year. The output statistic of interest is the slope parameter of the straight line for the South Coast data, and whether this value is significantly different from zero, indicating an increase in prevalence from age 1 to age 2+ year old sardine on the South Coast.

Data and Methods

The available data on parasite prevalence from 2011 to 2013were kindly provided by Carl van der Lingen.

The logistic function utilised is of the form $P_l = a/\left[1+e^{-b*l+c}\right]$, where P_l denotes prevalence of the parasite infection in fish of length class l, and *a*, *b* and *c* are estimable parameters. A binomial error distribution is assumed for the likelihood maximised in the fitting process, and as such the negative log likelihood is given by:

$$-\ln L = -\sum_l \left\{k_l \ln P_l + (n_l - k_l)\ln(1-P_l)\right\} \tag{1}$$

where n_l is the total number of observations for length class l, k_l is the total number of infected fish observed for length class l, and P_l is the model-estimated prevalence.

Sensitivity tests repeat the analysis for the assumed 1-2 year transition length approximation of 18cm by alternative choices of 17 and 19cm. This is to check the robustness of the result to the choice of 18 cm, given uncertainty about the most appropriate length to choose.

Results

The estimates for the parameters of the logistic function for the West Coast, along with associated standard errors, are given in Table 1. The estimates for the logistic function for the South Coast, as well as the estimates of the slope parameter for the straight line relation at larger lengths, are given with their standard errors in Table 2.

Figure 1 shows plots of the South Coast data with the fits for both the case where a logistic was fit for all lengths, and where a straight line was fit instead for lengths greater than 18cm, with continuity but not derivative continuity at the 18 cm transition point.

The over-dispersion parameter *D* for the data used in these South Coast fits was estimated as follows:

$$D = \sum_l \left(k_l - n_l P_l\right)^2 / \sum_l n_l P_l (1 - P_l) \qquad (2)$$

Estimates of *D* do not differ significantly between coasts for a specific year, but do differ from year to year. The values for the South Coast, with standard errors in parenthesis, are: 0.50 (0.09); 1.52 (0.20) and 0.84 (0.17) for 2011, 2012 and 2013 respectively.

Discussion

The key concern of this paper is whether or not the data for the South Coast are sufficient to confirm a continuing increase in prevalence with length for lengths corresponding to age 2+ sardine, which translates into the question of whether or not the slope of the straight line relationship above a certain length is significantly different from zero. For the baseline choice of 18 cm to separate sardine of age 2 and above from those of age 1, and if adding the assumption of distribution normality, the slope estimate is clearly significant at the 5% level for all three years considered (see Table 2). This remains so even when the alternative separation lengths of 17 and 19 cm are used, with one exception. That is for the case of 19 cm in 2012, for which the slope estimate, though still positive, is not significantly different from zero at the 5% level. However reference to the plot of fits for the South Coast in 2012 in Fig. 1 shows that this is merely the inevitable result of few data being available beyond 19 cm, with these being unable to provide the necessary information content required to lead to sufficiently precise estimates.

The inference of significance at the 5% level depends also on the assumption of data independence, and hence on the appropriateness of the binomial distribution assumption. The data proved too limited (lacking sufficient contrast) to estimate over-dispersion internally through the use of a beta-binomial distribution, for which convergence difficulties were encountered. However the estimates of *D* provide by equation (2) and reported above provide some insight. Certainly for 2011 and 2013, there is no indication that over-dispersion might invalidate the inferences drawn. However there is indication of over-dispersion for 2012, and adjustment for this may be made by dividing the negative log-likelihood of equation (1) by the point estimate of *D* of 1.52. If this is done, the standard error of the slope estimate for the straight line increases from 0.032 to 0.039, so that the estimate of 0.111 for the slope itself remains significant at the 5% level. Thus the inference above remains valid even when over-dispersion is taken into account.

Conclusion

A continued increase in parasite prevalence in sardine from age 1 to ages 2 and above in on the South Coast has been robustly demonstrated. It follows from this that, if parasite infection can occur only on the West Coast as hypothesised, there must be some movement (in at least one direction) of sardine of these ages between the West Coast and the South Coast.

Acknowledgements

Carl van der Lingen kindly provided the data upon which these analyses are based. We thank Andre Punt for discussions regarding the methodology to apply.

References

de Moor, C.L. and Butterworth, D.S. 2012. Assessment of the South African sardine resource using data from 1984-2011, with some results for a single stock hypothesis. FISHERIES/2012/SEP/SWG-PEL/48, 36pp.

de Moor, C.L. and Butterworth, D.S. 2013. Assessment of the South African sardine resource using data from 1984-2011: results for a two stock hypothesis at the posterior mode. FISHERIES/2013/AUG/SWG-PEL/20, 46pp.

van der Lingen, C.D. and Hendricks, M. 2014. Update on “tetracotyle” type metacercariae infection data and implications for sardine movement. FISHERIES/2014/MAR/SWG-PEL/08, 8pp.

van der Lingen, C.D., Weston, L.F., Reed, C.C., Hendricks, M. and H. Winker. 2014. Updated GLM analyses of parasite data corroborate the sardine multi-stock hypothesis. FISHERIES/2014/APR/SWG-PEL/27, 4pp.

van der Lingen, C.D. and Winker, H. 2014. Increasing prevalence of infection with increasing fish sizefor southern stock is not a model artefact. FISHERIES/2014/APR/SWG-PEL/30, 3pp.

Table 1: Estimates for the parameters of the logistic function for the West Coast data. Standard errors are given in parenthesis. Units of *b* are cm^{-1}.

West Coast 2011	a		b		c	
Binomial, logistic for all lengths	0.507	(0.045)	2.917	(1.037)	47.443	(16.742)
West Coast 2012	a		b		c	
Binomial, logistic for all lengths	1.000	(0.001)	0.120	(0.113)	0.663	(2.089)
West Coast 2013	a		b		c	
Binomial, logistic for all lengths	0.964	(0.144)	0.934	(0.231)	17.100	(3.949)

Table 2: Estimates for the parameters of the logistic function for the South Coast data, as well as the estimates of the slope *m* of the straight line into which the logistic function changes above the length indicated, which serves as a coarse separator of sardine above and below age 2. Results are given for the case where a logistic is fit for all lengths, as well as the cases where the logistic function is fit up to lengths 17cm, 18cm and 19cm respectively, and a straight line thereafter. In all cases a binomial error distribution has been assumed. Standard errors are given in parenthesis. Units of *b* are cm^{-1}.

South Coast 2011	a		b		c		**m**	
Binomial, logistic for all lengths	1.000	(0.001)	0.882	(0.175)	16.558	(3.202)	-	-
Binomial, straight line from 18cm	1.000	(0.008)	0.877	(0.420)	16.448	(7.373)	**0.198**	**(0.056)**
Binomial, straight line from 17cm	0.166	(0.047)	75.79	(43890)	1231	(713970)	**0.187**	**(0.030)**
Binomial, straight line from 19cm	1.000	(0.005)	0.914	(0.229)	17.121	(4.160)	**0.175**	**(0.029)**
South Coast 2012	a		b		c		**m**	
Binomial, logistic for all lengths	0.632	(0.142)	0.896	(0.387)	16.069	(6.540)	-	-
Binomial, straight line from 18cm	0.735	(1.987)	0.808	(1.057)	14.736	(14.452)	**0.111**	**(0.032)**
Binomial, straight line from 17cm	1.000	(0.055)	0.642	(0.678)	12.280	(11.411)	**0.120**	**(0.025)**
Binomial, straight line from 19cm	0.634	(0.324)	0.921	(0.552)	16.519	(8.908)	**0.081**	**(0.060)**
South Coast 2013	a		b		c		**m**	
Binomial, logistic for all lengths	1.000	(0.003)	0.601	(0.106)	12.993	(1.948)	-	-
Binomial, straight line from 18cm	0.051	(0.010)	52.374	(352760)	46.63	(1988200)	**0.139**	**(0.026)**
Binomial, straight line from 17cm	0.039	(0.010)	12.229	(151790)	1.553	(69360)	**0.075**	**(0.014)**
Binomial, straight line from 19cm	1.000	(0.003)	0.608	(0.164)	13.118	(2.918)	**0.105**	**(0.069)**

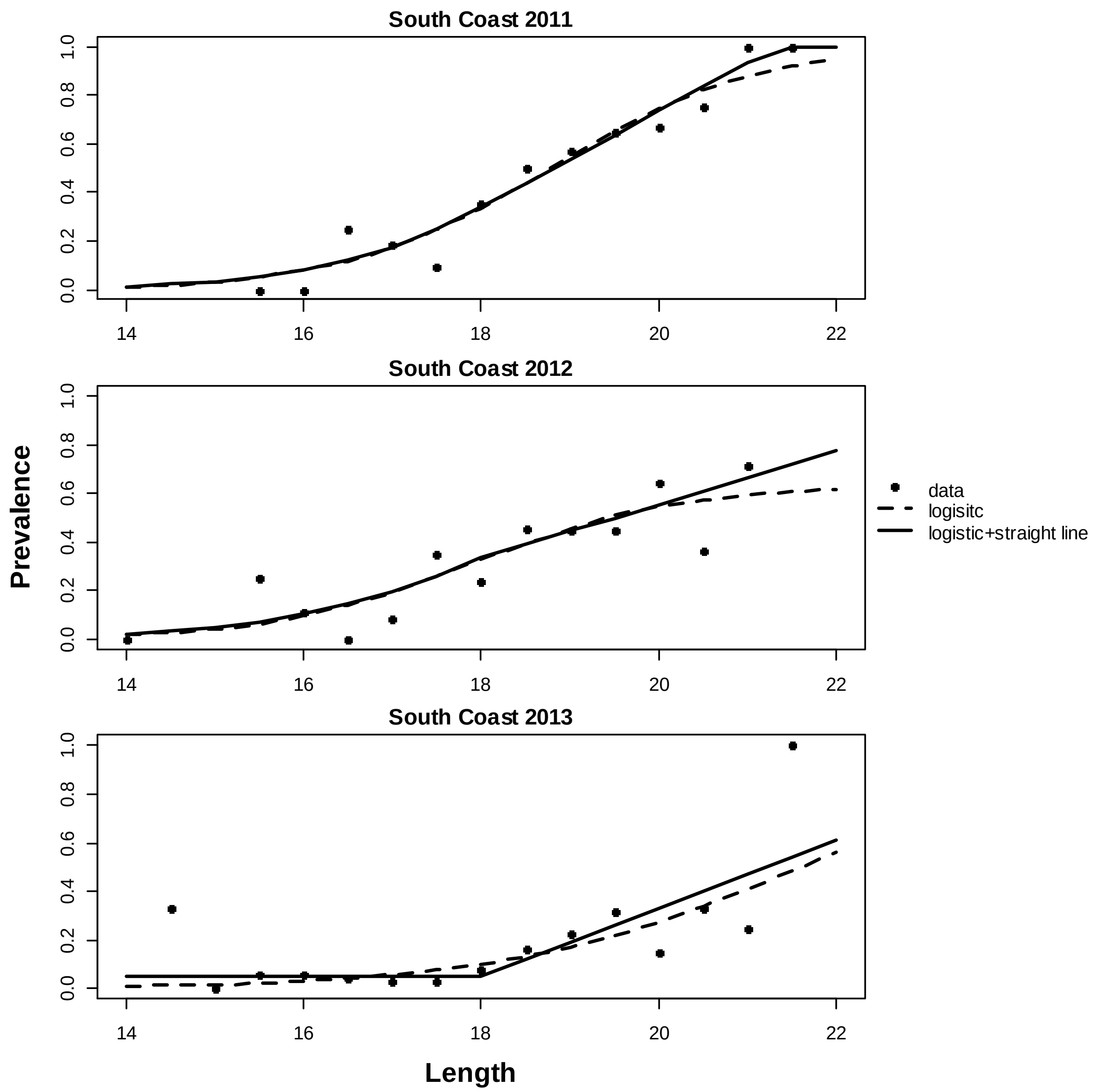

Figure 1: South Coast prevalence. Plots show the results for the case where the data are fit to a logistic function (dashed line), as well as for the case where they are fit to a logistic function up to length 18cm, with a straight line thereafter (solid line) (note that this line was replaced by 1 for lengths for which is would otherwise have exceeded 1). Note that a binomial error distribution was assumed for the fits shown here.