Evaluating the accuracy of hake abundance index predictions using different smoothing techniques

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

Summary

Alternative projection approaches based on linear and quadratic smoothing are applied retrospectively to abundance indices for hake to ascertain whether they can provide more accurate predictions of resource status one and two years ahead, in line with ideal needs for the empirical OMP used to set hake TACs. The results suggest that there is probably little if anything to be gained from attempting more complex formulations than three-year averages of abundance indices for input to a target-based empirical OMP for SA hake.

Introduction

The current OMP for South African hake is empirical, with estimates of trends of abundance indices over the past five-to-six years, and average values over the past three years, being used as input for computing TAC recommendations. For the future, it seems most likely that a "target-based" empirical OMP will be used, which adjusts the TAC based on the difference between the "current" and some target abundance index value, as this approach tends to provide more stable TACs than do calculations based on abundance index trends. Nevertheless, to reduce the impact of noise in any "current" measure of abundance, it is necessary to average over a number of years' data (typically at least three) to reduce variability.

A concern has been raised that using such an average, which necessarily has to be developed from data increasingly further back in time, results in a delay in identifying current stock status, and in particular fails to identify turning points in trends in abundance indices. This then results in TACs which do not correspond as closely as would be desirable to the current status of the resource. The aim of the retrospective analyses performed here is to explore the performance of alternative smoothing techniques to the averaging approach used at present. What is of particular interest is identifying which smoothing functions are best able to predict future abundance.

Methods

Five abundance index series are available for each of the two hake species, *Merluccius paradoxus* and *M. capensis*: a commercial CPUE series for the West and the South Coast, a survey abundance series for the West and the South Coast, and an overall index of abundance which is a weighted composite of the four CPUE and survey series (Table 1).

Three smoothing methods have been applied to each of these series:

1. **Average**: the average of the CPUE values from the last three years is the predicted value for the following year.
2. **Linear**: a linear local regression (LOESS[1]) applied to the previous five years' data to obtain a predicted value for the following year.
3. **Quadratic:** a quadratic local regression applied to the previous five years' data to obtain the predicted value for the following year.

In order to assess how well each of these three smoothing methods are able to predict, segments of historic CPUE data were used to make predictions for subsequent years for which CPUE data are available. In other words, data from 1988 to 1992 were used to predict a CPUE value for 1993, which was then compared to the 1993 observed value; 1989–1993 data points were used to predict the CPUE for 1994, and so on.

Results

Quantitative comparison of the three smoothing methods is based on the root mean square errors (RSME) between the predicted and observed CPUE values (Table 2). A graphic presentation of these RMSE values is given in Figure 1. Figure 2 shows plots of the predicted and observed values for the overall composite abundance index for each species, while Figure 3 shows similar plots for the four separate indices contributing to that composite index.

Discussion

TAC recommendations for the hake resource have to be developed in October of the calendar year before that for which they are implemented. Survey abundance indices are available for the year in which computations are made, and so involve only a one-year time lag, but CPUE data are available only for the preceding year, so that two years' delay is involved. For the composite index involving a weighting across coasts of both survey and CPUE indices, there is thus a two year delay. Retrospective estimates of accuracy (RMSE values) have thus been shown for both one and two year projections.

It is probably easiest to assimilate the results of these calculations which are listed in Table 2 through inspection of the plot in Figure 1. For single year projections, there is little to choose between the three methods for CPUE and for the composite index (in which CPUE is more heavily weighted than surveys) for *M. paradoxus,* though for *M. capensis* the average method performs worse. For surveys however, the quadratic method is notably worse, particularly for the south coast survey.

For two year projections, however, on which OMP input relies, the performance of quadratic smoothing is consistently the worst, with the average better than the linear approach in nine out of ten cases. The reasons for quadratic smoothing poorer performance is readily evident from comparisons for individual years in Figures 2 and 3, which show that the quadratic projections can sometimes show high variability.

These results suggest that there is probably little if anything to be gained from attempting more complex formulations than three-year averages of abundance indices for input to a target-based empirical OMP for SA hake.

---

[1] Given *n* data points, a locally weighted regression, or LOESS, (Cleveland and Devlin 1988), fits a polynomial of specified degree to each of the *n* points using the surrounding data points. The proportion of the *n* data points used in the fit is given by the smoothing parameter α; the nearest αn points to any point x_i are used to fit the polynomial. Points closest to x_i are given the highest weight.

References

Cleveland WS and Devlin SJ. 1988. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83:596–610.

Rademeyer RA, Fairweather T, Glazer JP, Leslie RL, and Butterworth DS. 2010. The 2010 operational management procedure for the South African *Merluccius paradoxus* and *M. capensis* resources. Unpublished report, Department of Agriculture, Forestry and Fisheries, South Africa. Report No. FISHERIES/2010/OCTOBER/SWG-DEM/59.

Table 1: Hake abundance indices: CPUE and survey by coast, and the weighted composite of those indices for each species which is used in the hake OMP control rule (R. Rademeyer, pers. comm.)

	M. paradoxus					*M. capensis*				
Year	WC-CPUE	SC-CPUE	WC-survey	SC-survey	Overall	WC-CPUE	SC-CPUE	WC-survey	SC-survey	Overall
1979	0.885	0.423				0.731	0.942			
1980	0.881	0.419				0.961	0.986			
1981	0.858	0.484				0.996	1.045			
1982	0.833	0.476				1.115	1.106			
1983	0.807	0.506				1.031	1.135			
1984	0.822	0.487				1.096	1.207			
1985	0.845	0.554	0.298			1.174	1.362	0.647		
1986	0.922	0.673	0.321			1.349	1.635	0.629		
1987	0.934	0.769	0.380			1.329	1.712	0.551		
1988	0.887	0.820	0.374	0.190	0.694	1.240	1.673	0.450	0.876	1.106
1989	0.787	0.737	0.389	0.190	0.633	1.066	1.550	0.370	0.876	1.012
1990	0.751	0.659	0.386	0.190	0.594	1.003	1.574	1.356	0.876	1.150
1991	0.788	0.710	0.534	0.167	0.652	0.938	1.753	1.384	1.448	1.351
1992	0.852	0.803	0.488	0.160	0.695	0.977	1.837	1.088	1.088	1.227
1993	0.857	0.951	0.519	0.522	0.784	1.074	1.830	0.459	1.004	1.132
1994	0.847	0.938	0.521	0.700	0.794	1.194	1.558	0.535	0.804	1.057
1995	0.833	0.873	0.578	0.764	0.787	1.358	1.489	0.715	0.976	1.172
1996	0.820	0.703	0.634	0.527	0.718	1.460	1.442	0.699	1.128	1.237
1997	0.824	0.670	0.774	0.577	0.743	1.672	1.512	0.934	1.167	1.366
1998	0.779	0.730	0.877	0.693	0.775	1.691	1.314	0.884	1.128	1.307
1999	0.800	0.799	0.993	1.433	0.901	1.769	1.197	1.184	0.988	1.307
2000	0.747	0.826	0.986	2.017	0.946	1.715	1.104	1.032	1.009	1.252
2001	0.700	0.748	0.986	2.017	0.903	1.706	1.179	1.032	1.009	1.266
2002	0.585	0.712	0.469		0.602	1.453	1.095	0.552		1.134
2003	0.518	0.611	0.595	0.683	0.578	1.237	1.065	0.473	0.678	0.908
2004	0.504	0.644	0.556	0.503	0.556	1.004	1.037	0.761	0.669	0.871
2005	0.497	0.626	0.576	0.394	0.541	0.908	1.040	0.730	0.615	0.821
2006	0.494	0.617	0.520	0.239	0.510	0.749	0.972	0.752	0.619	0.761
2007	0.475	0.557	0.604	0.417	0.520	0.658	0.777	0.484	0.556	0.627
2008	0.526	0.559	0.580	0.445	0.539	0.591	0.652	0.441	0.624	0.592
2009	0.599	0.599	0.599	0.599	0.599	0.667	0.667	0.667	0.667	0.667
2010	0.659	0.674	0.699	0.668	0.672	0.930	1.014	0.772	0.836	0.896
2011	0.702	0.767	0.761	0.634	0.726	1.130	1.245	0.856	0.856	1.030
2012	0.707	0.892	0.790	0.612	0.770	1.349	1.475	0.677	0.875	1.129
2013	0.685	0.908	0.668	0.159	0.696	1.288	1.238	0.590	0.773	1.011

Table 2: Root mean square prediction error values are shown as a measure of accuracy of the retrospective predictions made using three smoothing techniques, applied to the five hake abundance indices. For each year for which sufficient past data are available, predictions for the next year (as well as for two years ahead) are made based on previous years' data. "Average" takes the average of the last three years' data; the other two methods apply linear and quadratic LOESS smoothing functions to the previous five years' data to obtain the predicted value.

M. paradoxus		Overall	WC CPUE	SC CPUE	WC Survey	SC Survey
One year	**Average**	0.1203	0.0759	0.1177	0.1556	0.4344
	Linear	0.1036	0.0545	0.0953	0.1482	0.4167
	Quadratic	0.1234	0.0547	0.0962	0.2317	1.0034
Two year	**Average**	0.1545	0.1039	0.1525	0.2001	0.6131
	Linear	0.1899	0.1073	0.1872	0.2319	0.8452
	Quadratic	0.3137	0.1497	0.2616	0.5059	2.3571

M. capensis		Overall	WC CPUE	SC CPUE	WC Survey	SC Survey
One year	**Average**	0.1706	0.2426	0.2207	0.3587	0.1428
	Linear	0.1325	0.1299	0.1837	0.405	0.2716
	Quadratic	0.1401	0.1258	0.1920	0.5469	0.5848
Two year	**Average**	0.2305	0.3454	0.2888	0.3775	0.1761
	Linear	0.2360	0.2526	0.3237	0.6840	0.4489
	Quadratic	0.3780	0.2992	0.4963	1.3368	1.3742

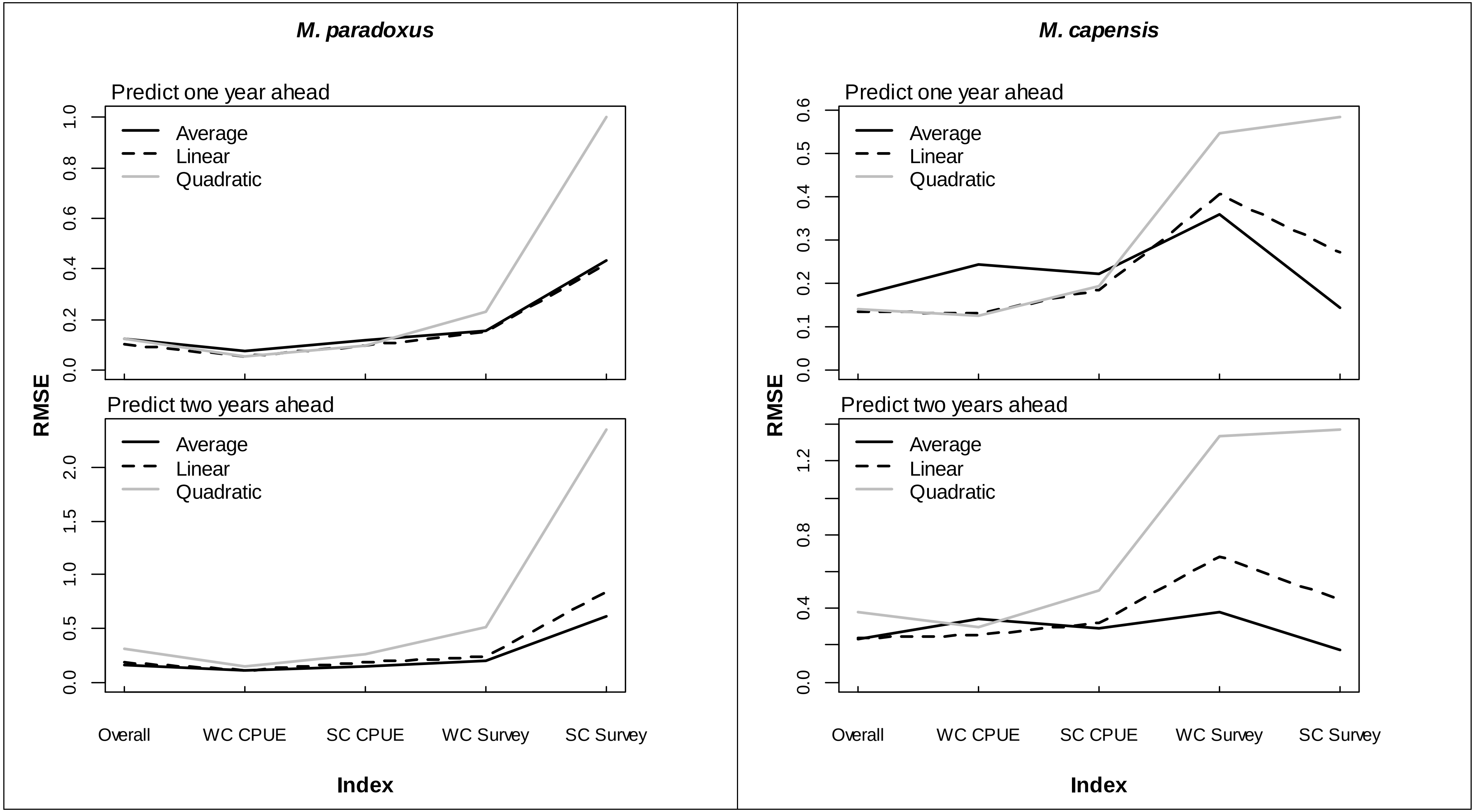

Figure 1: Root mean square error values are shown as a measure of accuracy of the retrospective predictions made using three smoothing techniques, applied to the five hake abundance indices. For each year for which sufficient past data are available, predictions for the next year (as well as for two years ahead) are made based on previous years. "Average" takes the average of the last three years' data; the other two methods apply linear and quadratic LOESS smoothing functions to the previous five years' data to obtain the predicted value.

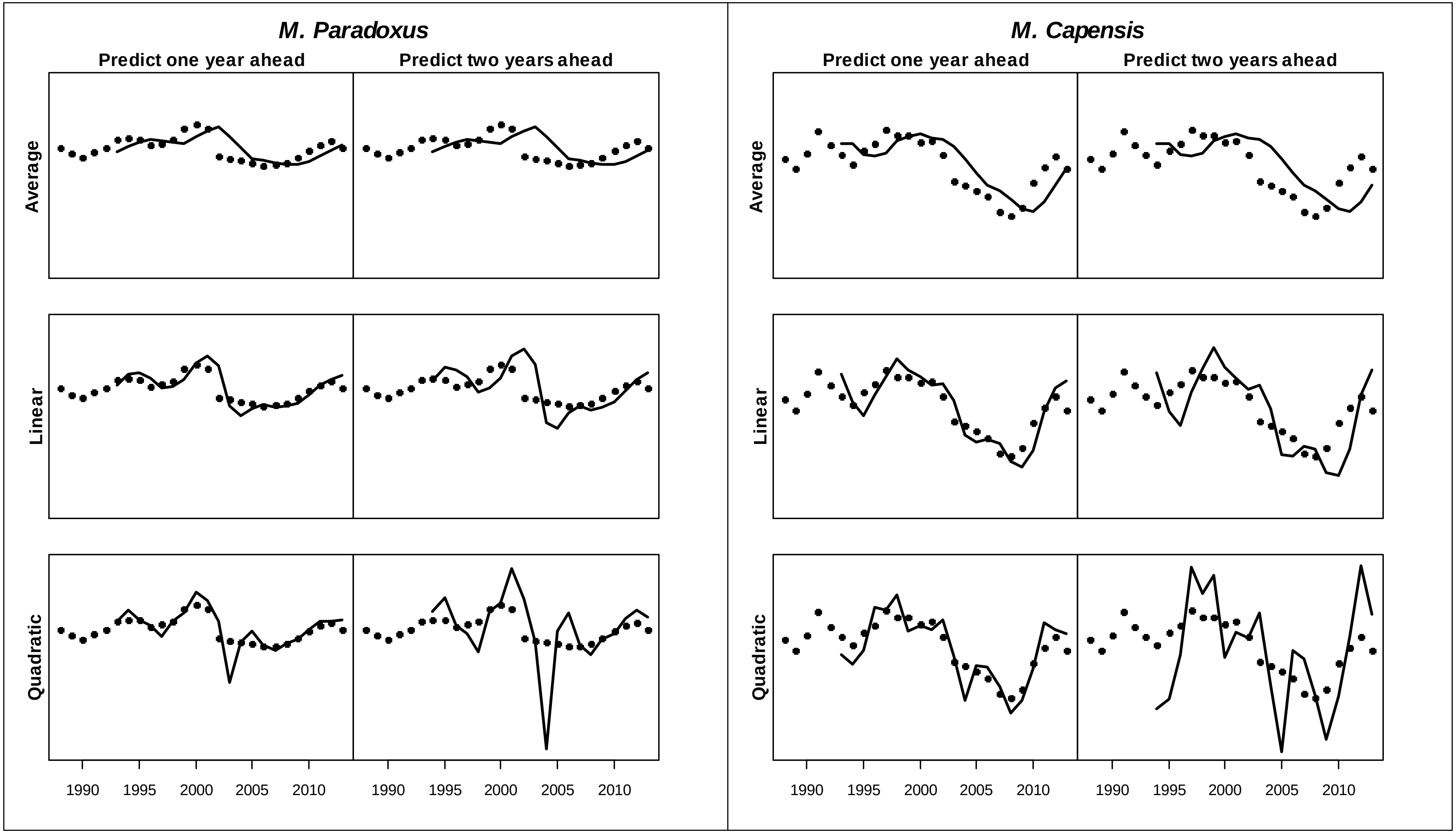

Figure 2: Three different smoothing techniques applied to the overall hake abundance index to test the accuracy of predictions. The smooth lines show the predicted values: each year has a corresponding value predicted from previous years. “Average” takes the average of the last three years’ data; the other two methods apply linear and quadratic LOESS smoothing functions to the previous five years’ data to obtain the predicted value. Predictions are made one and two years into the “future”.

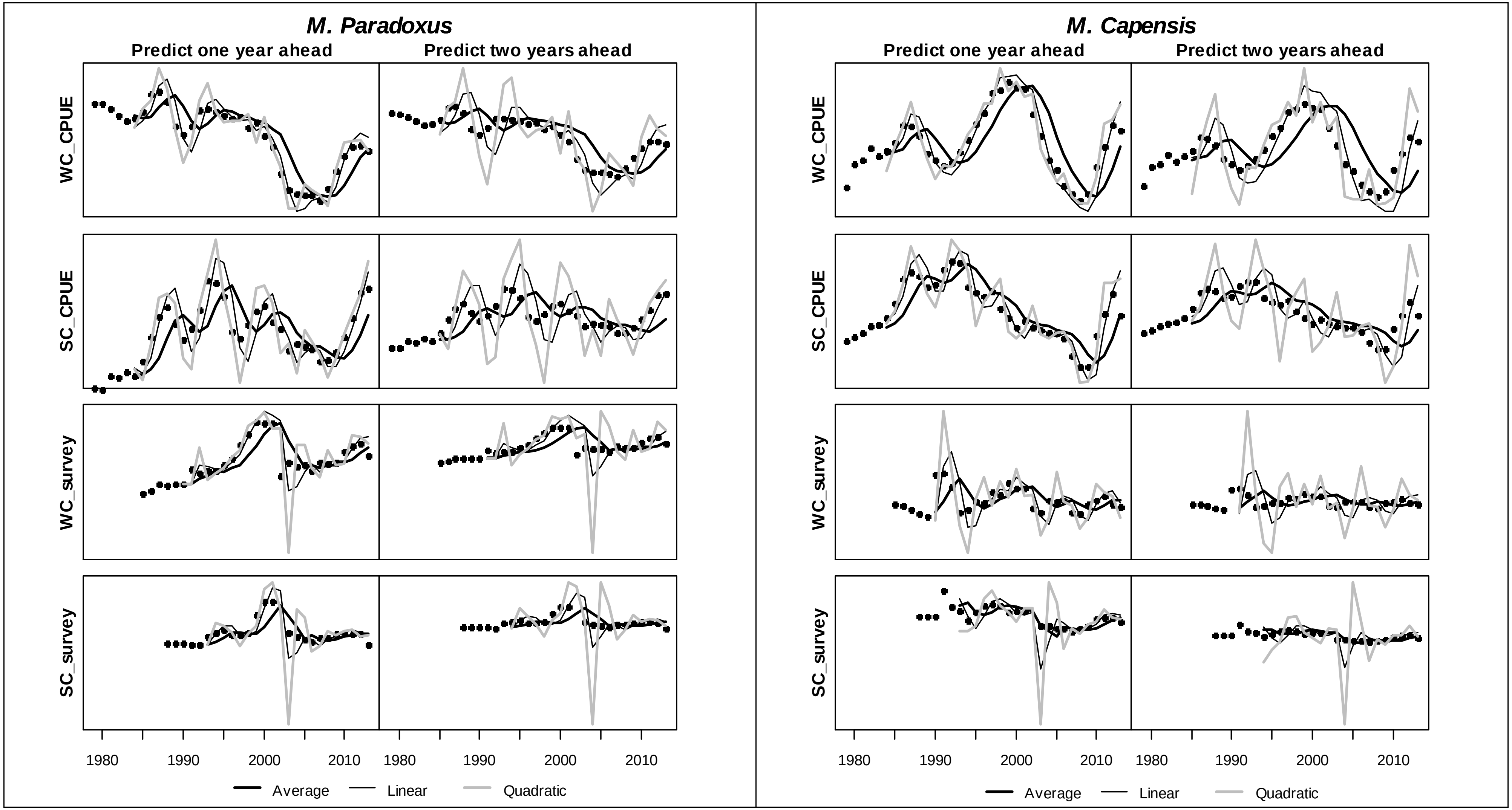

Figure 3: Three different smoothing techniques applied to the four basic hake abundance indices to test accuracy of predictions. The smooth lines show the predicted values: each year has a corresponding value predicted from previous years. "Average" takes the average of the last three years' data; the other two methods apply linear and quadratic LOESS smoothing functions to the previous five years' data to obtain the predicted value. Predictions are made one and two years into the "future".