

Assessing Dynamics of Lake Huron Fish Communities using Dynamic Factor Analysis

QFC Technical Report T2014-01

Prepared for:
Ontario Ministry of Natural Resources

Prepared by:
Brian A. Maurer^{1,2}, James R. Bence^{2,3}, and Travis O. Brenden^{2,3}
¹Center for Statistical Training and Consulting, Michigan State University, East Lansing, MI 48824
²Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824
³Quantitative Fisheries Center, Michigan State University, East Lansing, MI 48824

Introduction

An important component of understanding how human activities impact the biological diversity of Lake Huron is assessing how fish communities change over decade-long periods of time at different locations in the Lake Huron basin. In this report we analyze five time series of data obtained by the Ontario Ministry of Natural Resources (OMNR) over the past 30 years. We focus on describing common trends in relative abundance of the most abundant species found these five locations where OMNR has sampled as part of a larger series of surveys and stock assessments done by OMNR's Upper Great Lakes Management Unit.

Our goals for this report are twofold. First, we examine the utility of using dynamic factor analysis (DFA) to represent changes in fish communities over time in Lake Huron. Second, we present the results of DFA for five time series to describe patterns variation in changes that have occurred in fish communities at different sample locations in Lake Huron. We present these results with the intent of providing information that OMNR personnel can use to understand past changes in fish biodiversity and to formulate strategies and recommendations on how to encourage the preservation of fish diversity in the future.

Methods

Data on fish communities at five different locations in Lake Huron (Fig. 1) were provided by OMNR. Each location was sampled a different number of years between 1979 and 2012. Each year for which sampling was done, nets were set multiple days. The number of net sets (sample) per year varied within and among the five locations. For this reason, we took the average number of fish caught for each species across all samples in a given year as the basic response. For each sample, eight nets with different mesh sizes were used. For this analysis we

summed the catch for all species across all mesh sizes to calculate averages per sample. For all the sampling locations except for the eastern North Channel site, the type of nets used changed during the middle of the time series. In several years two different types of nets were used to facilitate comparison of catches between the two different types of nets. In our analyses, we did not account for possible differences in catch between the two net types. Although there are some differences between the two net types, those differences do not appear to be sufficiently large to change the overall patterns we found.

Dynamic factor analysis (DFA) – DFA is a technique that models multivariate time series as linear combinations of underlying latent trends common to all time series. In the present context, this means that DFA searches for underlying common trends of covariation among fish species across years. The assumption is that there may be environmental factors operating in a part of the lake that affect at least some species in similar ways. DFA can identify more than one latent trend, so it is possible that for some time series, different sets of species may react to different environmental factors in the same way, so that the overall trends in abundance in the fish community can be decomposed into a small number of “guilds” of species each reacting to different aspects of environmental change.

The model can be written formally as follows (Holmes et al 2012; Zuur et al 2003a,b; Zuur and Pierce 2004). If we let the abundances of S species of fish in year t be represented by an $S \times 1$ column vector, \mathbf{y}_t , then

$$\mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \boldsymbol{\mu} + \mathbf{v}_t \quad (1)$$

is the DFA model. Here, the values of the m latent trends at time t are in the $m \times 1$ column vector \mathbf{x}_t . The $S \times m$ matrix \mathbf{Z} contains the *loadings* for each species on each of the m trends. A loading is a measure of the relationship between a species and a particular underlying trend. If a loading

is relatively large and positive, it indicates that the species in question is closely associated with a particular trend. Loadings are generally between -1 (indicating the abundance of a species changes in the opposite direction of a trend) and +1. A loading near zero indicates that fluctuations in abundance for a species are not related to the trend in question. The $S \times 1$ column vector $\boldsymbol{\mu}$ contains the average abundances of the species across the time series. Thus $\mathbf{Z}\mathbf{x}_t + \boldsymbol{\mu}$ gives the overall trends for each species. Finally, \mathbf{v}_t models random deviations around the means and trends. This last term is assumed to come from a multivariate normal distribution with a mean vector containing zeroes and a covariance matrix \mathbf{R} . To complete the description of the DFA model, the latent trends in \mathbf{x}_t are assumed to follow a simple random walk so that the value of the trend in a given year depends only on the value of that trend in the previous year plus some random variation. This gives

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{w}_t \quad (2)$$

Here \mathbf{w}_t represents random variation and is also assumed to be from a multivariate normal distribution with zero mean vector and covariance matrix \mathbf{Q} . Finally, since the initial value of the trend, \mathbf{x}_0 , is unknown, it is modeled as coming from a multivariate normal distribution with a mean vector $\boldsymbol{\pi}$ and covariance matrix \mathbf{P} . The unknown parameters that need to be estimated from the data, then, are \mathbf{Z} , $\boldsymbol{\mu}$, \mathbf{R} , $\boldsymbol{\pi}$, and \mathbf{P} . For these analyses, we assumed no structure to the covariance matrices \mathbf{R} and \mathbf{P} . The latent trends are then modeled using equation (2), starting at the estimate of the mean vector $\boldsymbol{\pi}$. The estimates of the loadings describe which species are associated with which trends. The model given in equation (1) can be expanded to include environmental measurements; however, since no such measurements were available for these time series, that term was not included.

We used the expectation-maximization (EM) algorithm as implemented in the MARRS package (Holmes et al. 2012) to obtain maximum likelihood estimates for the DFA model. Obtaining estimates for all of the parameters in the DFA model cannot be achieved by conventional statistical techniques, such as least squares. The problem is that the best fit depends on both how observed data differ from observed values, and how unobserved (latent) variables differ from their expected values. It is known that least squares and similar methods cannot simultaneously find the best fitting parameters that influence both the distribution of the observed data and the distribution of the latent variables. Many models that include latent variables, like DFA can be fit using the expectation-maximization (EM) algorithm (Dempster et al. 1977, Millar 2011). Basically, the EM algorithm is an iterative procedure with two alternating steps. One of the steps is to fix the latent variables (or parameters and disturbances determining them and their distribution) and estimate the other parameters. The other step is to then fix these other parameters and obtain updated estimates of latent variables and their associated parameters. The process is iterated until all parameter values converge. The EM algorithm has not been widely available until recently with the inclusion of the MARRS package in R (Holmes et al 2012). This package calculates parameter estimates and provides a method for testing how many latent trends are necessary to give the best description of the data. An additional advantage of the EM algorithm is that it implicitly imputes missing values. This was important because the time series for three of the locations had years in which no surveys were obtained. The eastern North Channel site had no data for 2010, the southern main basin site had no surveys during 1996, and the Owen Sound site was not sampled in 1995.

For each of the five time series we fit DFA models assuming $m = 1$, $m = 2$, and $m = 3$ latent trends. Models were compared using Akaike Information Criterion (AIC) values based on

the logarithm of the likelihood for each model, adjusted by the number of parameters (Burnham and Anderson 2002, Millar 2011). Models with the smallest AICs are defined as having the best support given the data. Based on AICs, it is possible to weight models, with the best model having the greatest relative support by the data. Due to the time series being relatively short, we used a correction to the AIC for small sample sizes (Burnham and Anderson 2002). Since we are more concerned with estimation of trends rather than providing population estimates for each species, we transformed the time series for each species in each location to have a mean of zero and standard deviation of one. Thus, the latent trends are represented as “standardized” trends, and the population trends for each species are represented as standardized deviations away from the mean. In some sense this means each species is on an equal footing with respect to its influence on the analysis regardless of how abundant it is. The loadings for each trend can be rotated using a standard factor rotation algorithm (Holmes et al 2013). This has the effect of maximizing the loadings for species that weight heavily on a given trend while minimizing the loadings for species with small weights. We used a varimax rotation for the loadings estimated from the EM algorithm in R.

Overall measures of goodness of fit for these models are not directly available from the MARRS algorithm. However, in order to examine how closely the models represented the data for each species, we used a relative mean squared deviation of the predicted abundance for each species as a measure of goodness of fit. If z_{it} is the standardized abundance for species i at time t , then the relative mean squared error for species i (RelMSE $_i$) is

$$\text{RelMSE}_i = 1 - \frac{\sum_{t=1}^T \frac{(z_{it} - z_{it}^*)^2}{T}}{s_i^2} \quad (3)$$

where z_{it}^* is the standardized abundance predicted from the best DFA model for species i , T is the length of the time series, and s_i^2 is the variance in abundance for species i . Since the

abundances for species were standardized, $s_i^2 \equiv 1$ for all species. The RelMSE is similar to R^2 as a measure of goodness of fit. It is close to 1 when the model fits the data well and close to 0 when the model fits the data poorly.

Results

Estimates of latent trends for the fish community time series in the two main basin sampling locations were similar. In each case, the best DFA model had two separate latent trends (Table 1). The first trend identified by DFA in the central basin was qualitatively similar to the second southern basin trend, while the second central basin trend was qualitatively similar to the first trend in the southern basin (Fig. 2). The loadings of different species on qualitatively similar trends at each sampling location were also similar (Table 2). We interpret this to mean that both sites show the same two trends in fish community structure. Henceforth we will treat both sampling sites as indicative of two general eastern main basin trends for Lake Huron.

The first main basin trend evident in the data for the two sampling sites was a decline that began around 1985 and bottomed out around 1995. Species that were most closely associated with this trend were burbot, lake whitefish, round whitefish, and yellow perch (Table 2). Chinook salmon were also associated with this trend at the central sampling site, but not the southern site. At the southern site, rainbow smelt were also associated with this trend. The second main basin trend was an initial increase that peaked around 1990, then declined to reach approximately the same level as the first trend by 2000. Species associated with this trend included bloater, lake chub (only at the central sampling location), long-nosed sucker, rainbow smelt, and white sucker (also only at the central sampling location).

The three sampling locations that were not in the main basin of Lake Huron were best modeled by a single trend (Table 1). The two Georgian Bay time series both showed trends that were qualitatively similar to the first main basin trend (Fig. 3). This trend showed a peak during the early 1980's followed by a decline that ended around 1995. The Owen Sound site wasn't sampled across the entire length of the trend. Species associated with this trend at the two Georgian Bay stations were generally similar to those that were associated with the similar trend in the main basin (Table 3). In addition, bloater was associated with this trend in Owen Sound, and both splake and yellow perch were associated with this trend in both Georgian Bay sites, but not in the main basin. Lake chub were also associated with this trend at the South Georgian Bay station. The trend identified for the eastern North Channel sampling location was distinctive (Fig. 3). When sampling began around 1995, populations were highest, but then began to decline and then increase again by 2002. After that, there was a substantial decline that leveled off around 2007. Alewife and white sucker were associated positively with this trend (Table 3), and round whitefish was negatively associated with it (that is, the whitefish increased during the time that the other two species were decreasing).

Discussion

The results of the DFA show that there were substantial declines in relative abundances of many species of fish in Lake Huron in the past several decades. The first decline began around 1980 and ended around 1995. This decline occurred in both the main basin and Georgian Bay, and was associated with bloater, the two suckers, and rainbow smelt. In Georgian Bay, in addition to these species, the decline involved splake and yellow perch. There was a second decline of different species in the main basin that began in 1990, after a period of increase, and

ended around 2000. The species associated with this main basin decline included burbot, the two species of whitefish, and yellow perch. Note that in some locations, some species either did not decline, or in some cases, increased. What this implies is that there have been substantial shifts in the structure of these fish communities over time so that current fish communities in Lake Huron are substantially different from those that existed at the time sampling began. For the most part this reorganization of community structure has involved substantial declines of species previously common, rather than increases of some species coupled with decreases of others. That is, species that were rare earlier in the sampling period did not increase in absolute abundance, rather, the increases in their relative abundance occurred because common species had declined.

Without additional information on changes in environmental conditions in Lake Huron over this time period, it is difficult to pinpoint why exactly these changes in fish communities occurred. There have been substantial biological changes in the lakes due to influx of invasive species, such as quagga and zebra mussels or sea lampreys, that may have significantly altered food webs that these species rely on. Changes in water chemistry due to pollution or climate change may have also had some impact, but without solid data on such factors, it is difficult to speculate on what specific biological or environmental factors were involved. A next step in attempting to understand these patterns might be to identify additional data on changes in the Lake Huron biological and/or physical environment that could be included in DFA models to help identify what changes in the lake were most likely to have caused the observed patterns of community reorganization documented here.

References

- Burnham, K.P. and D.R. Anderson. 2002. Model selection and multimodel inference. Springer, New York.
- Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B* 39: 1–38.
- Holmes, E. E., E. J. Ward and M. D. Scheuerell. 2012. Analysis of multivariate time-series using the MARSS package. NOAA Fisheries, Northwest Fisheries Science Center, 2725 Montlake Blvd E., Seattle, WA.
- Millar, R.B. 2011. Maximum likelihood estimation and inference. Wiley, Chichester, United Kingdom.
- Zuur, A.F., R.J. Fryer, I.T. Jolliffe, R. Dekker, and J.J. Beukema. 2003a. Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics* 14: 665-685.
- Zuur, A.F., and G.J. Pierce. 2004. Common trends in northeast Atlantic squid time series. *Jour of Sea Research* 52: 57-72.
- Zuur, A.F., I.D. Tuck, and N. Bailey. 2003b. Dynamic factor analysis to estimate common trends in fisheries time series. *Canadian Journal of Fisheries and Aquatic Sciences* 60: 542-552.

Table 1. Corrected AIC values for DFA models assuming different numbers of latent trends for each of the five time series analyzed here. T represents the number of years each time series was sampled.

Location	T	One trend		Two trends		Three trends	
		AICc	weight	AICc	weight	AICc	weight
Central basin	32	896	0.00	871	1.00	882	0.00
South basin	28	744	0.02	740	0.98	765	0.00
Southern Georgian Bay	34	988	1.00	1002	0.00	1020	0.00
Owen Sound	13	370	1.00	448	0.00	*	
Eastern north channel	19	533	1.00	561	0.00	598	0.00

* A three trend model was not fit to the Owen Sound data because the time series was so short

Table 2. Rotated trend loadings and goodness of fit measures (RelMSE) for species in fish communities in the central and southern sampling stations in the Lake Huron main basin. Loadings in bold indicate that a species weighted strongly on that particular trend.

Species	Central basin			Southern basin		
	Trend 1	Trend 2	RelMSE	Trend 1	Trend 2	RelMSE
Alewife	0.20	0.04	0.27	-0.06	-0.02	0.05
Bloater	0.06	0.35	0.84	0.37	0.07	0.74
Burbot	0.29	0.14	0.63	0.18	0.21	0.37
Chinook salmon	0.20	-0.16	0.40	0.14	0.18	0.25
Lake chub	0.04	0.20	0.30			
Lake whitefish	0.25	0.14	0.54	0.02	0.45	0.78
Long-nosed sucker	0.08	0.27	0.52	0.24	-0.04	0.28
Rainbow smelt	0.06	0.27	0.53	0.27	0.26	0.66
Round whitefish	0.28	0.01	0.46	0.02	0.32	0.39
White sucker	0.05	0.25	0.44	-0.02	-0.12	0.09
Yellow perch	0.30	0.15	0.66	0.08	0.36	0.54

Table 3. Trend loadings and goodness of fit measures (RelMSE) for species in fish communities in the southern Georgian Bay, Owen Sound and eastern north channel sampling stations in the Lake Huron upper basin. Loadings in bold indicate that a species weighted strongly on that particular trend.

Species	S. Georgian Bay		Owen Sound		Eastern N. channel	
	Trend	RelMSE	Trend	RelMSE	Trend	RelMSE
Alewife	-0.33	0.04	0.11	0.18	0.29	0.49
Bloater	0.26	0.52	0.20	0.40		
Burbot	0.07	0.06	0.28	0.76	-0.12	0.13
Chinook salmon	-0.08	0.07				
Lake chub	0.31	0.72			0.12	0.12
Lake whitefish	-0.16	0.20	-0.15	0.26	0.02	0.06
Long-nosed sucker	0.15	0.20	0.27	0.68	0.20	0.25
Rainbow smelt	0.35	0.86	0.25	0.59	0.06	0.07
Round whitefish	-0.05	0.05	-0.21	0.43	-0.29	0.48
Splake	0.32	0.73	0.24	0.58		
White sucker	0.32	0.74	0.25	0.61	0.31	0.56
Yellow perch	0.24	0.43	0.23	0.50	0.19	0.23

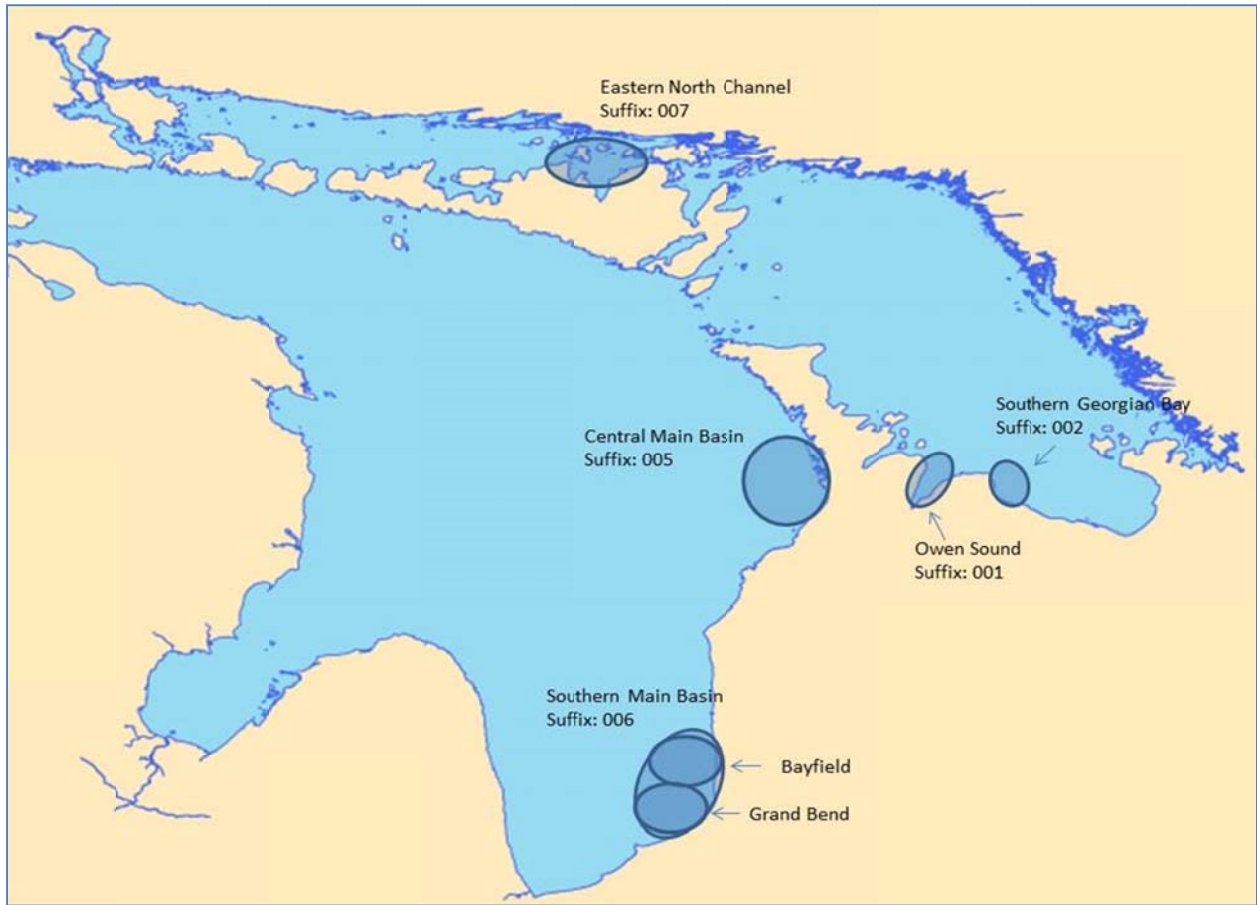


Figure 1. Five sampling locations on Lake Huron where data were collected on fish communities by Ontario Ministry of Natural Resources between 1979 and 2012. Each location was sampled a different number of times, and over a different span of years.

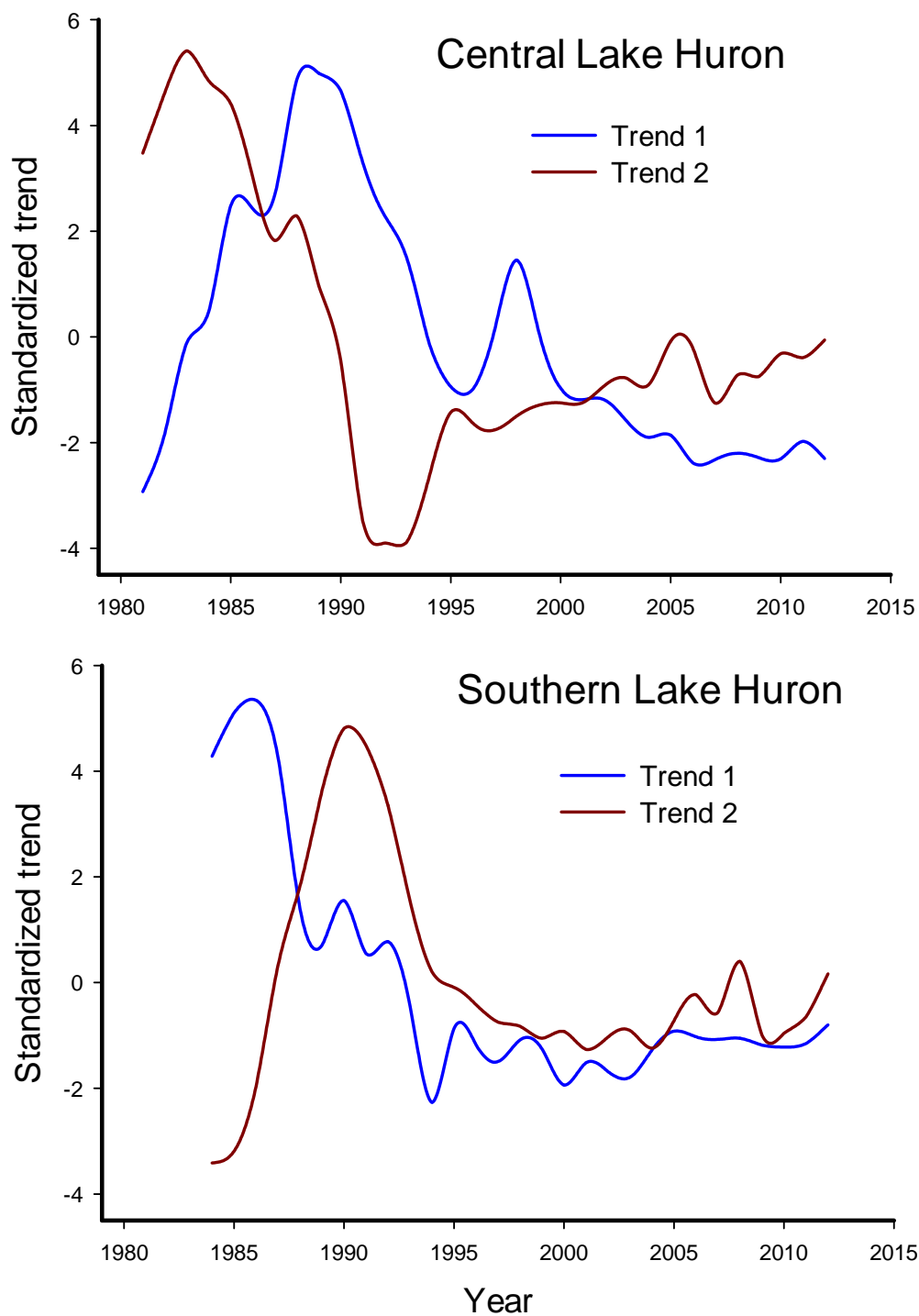


Figure 2. Estimated latent trends for the two Lake Huron main basin sampling stations. Note that the first trend in the central site is qualitatively similar to the second trend in the southern site, and the second trend in the central site is qualitatively similar to the first trend in the southern site.

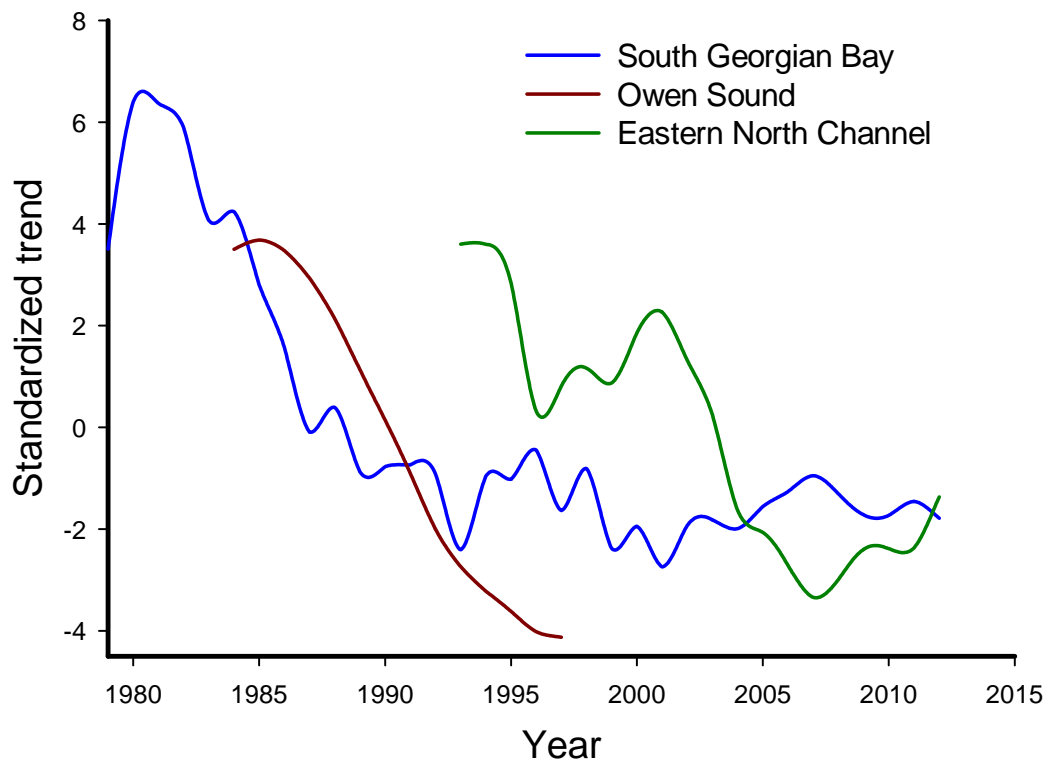
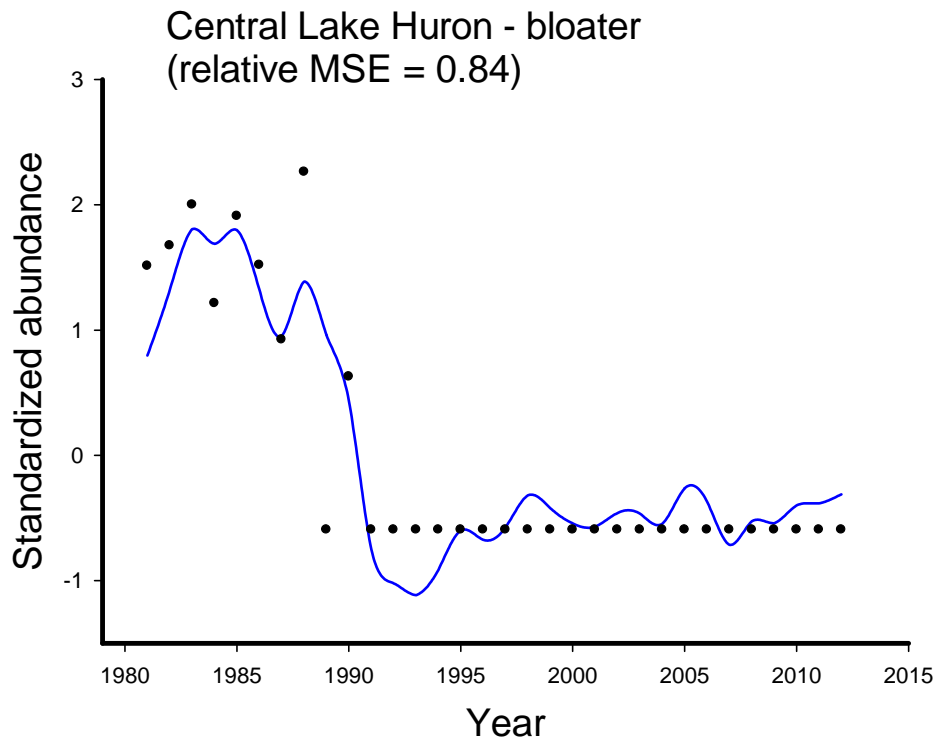
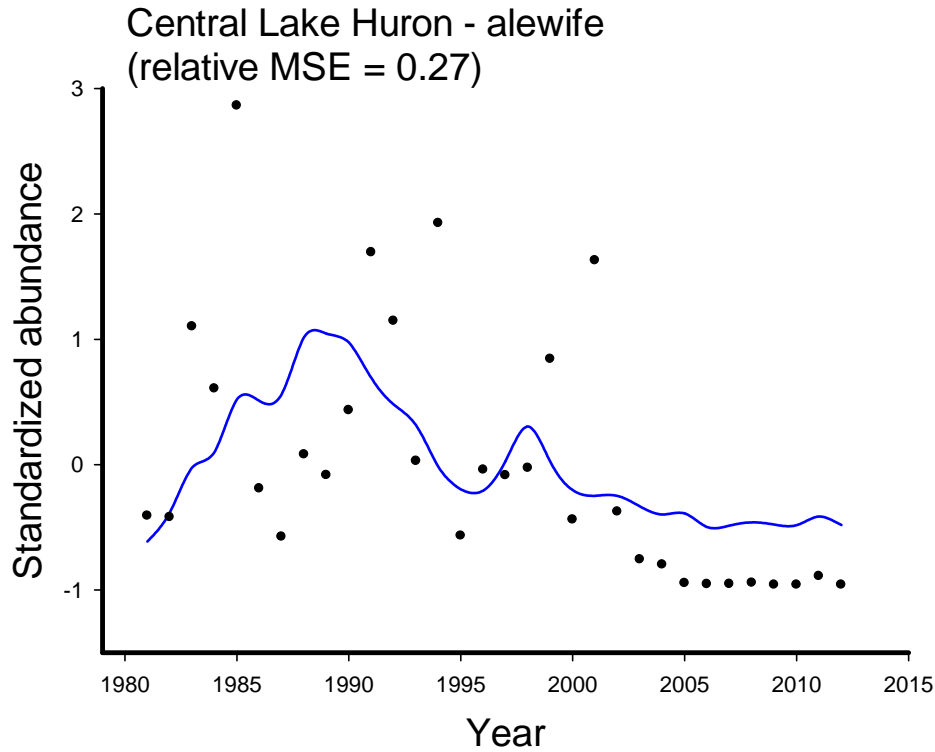
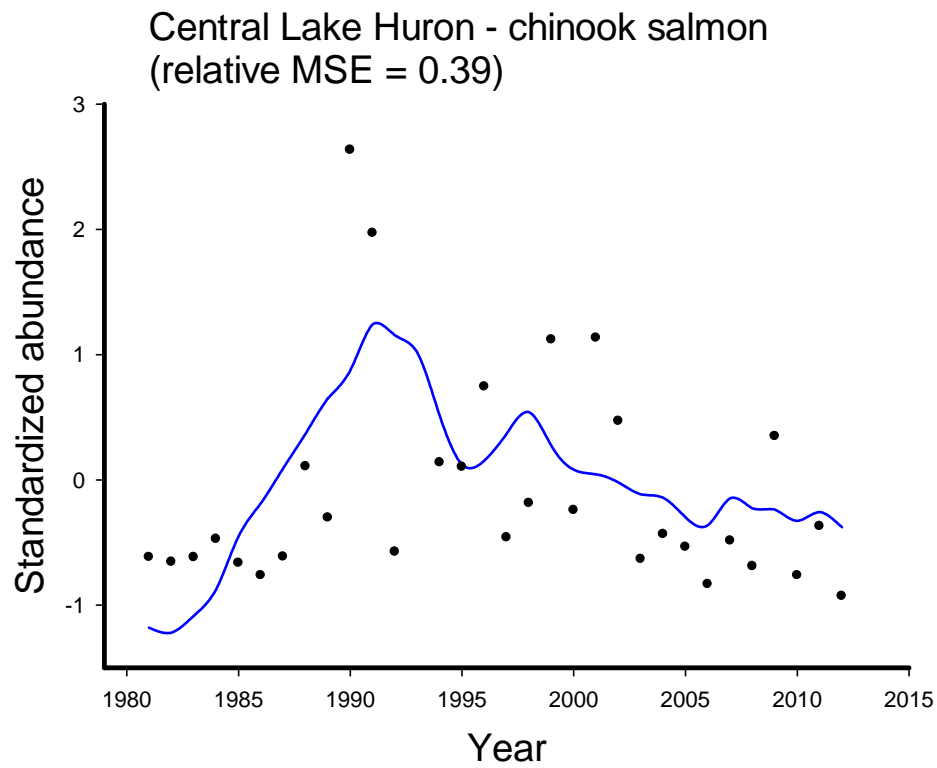
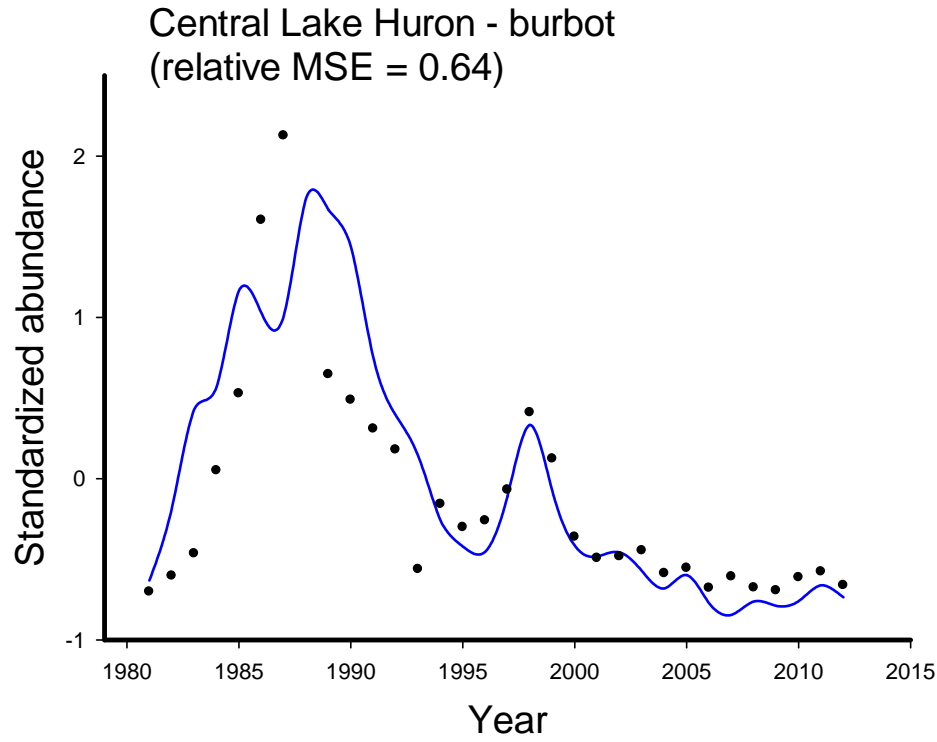
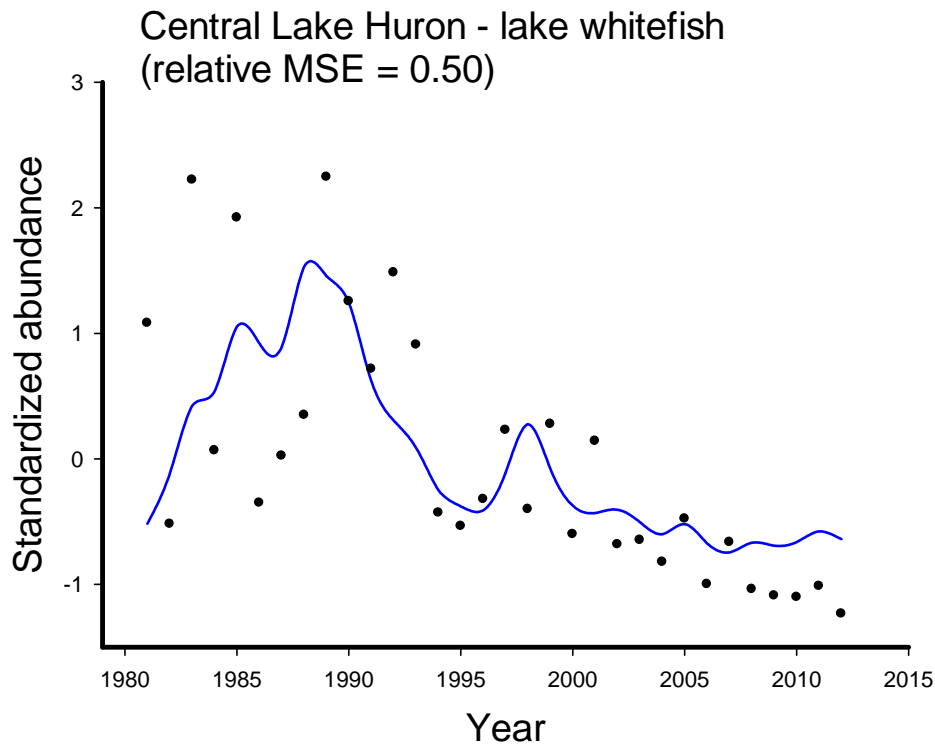
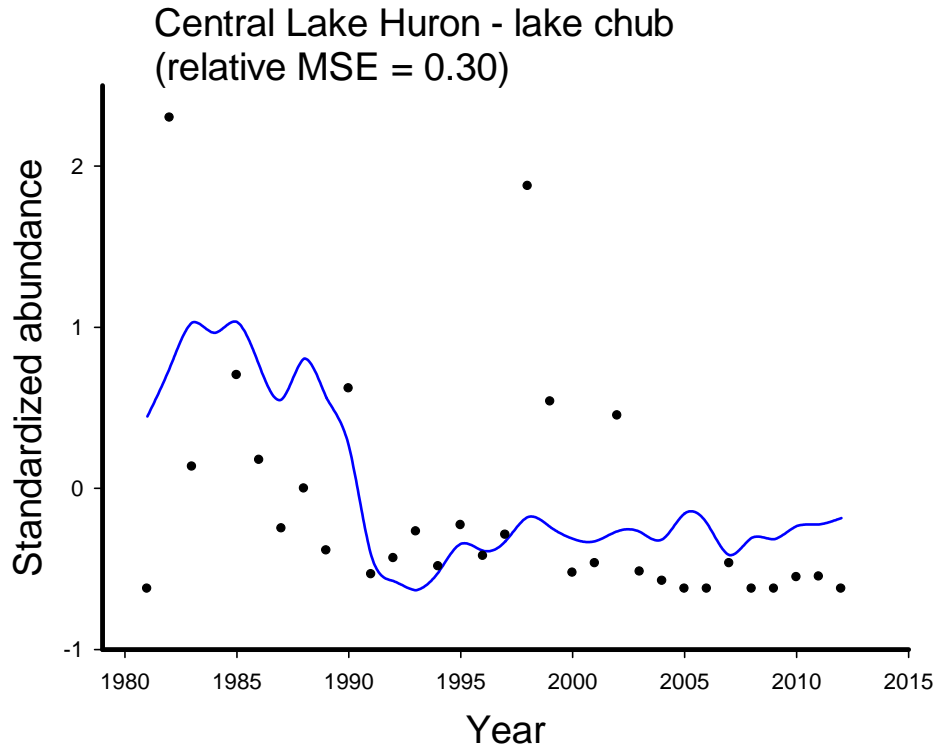


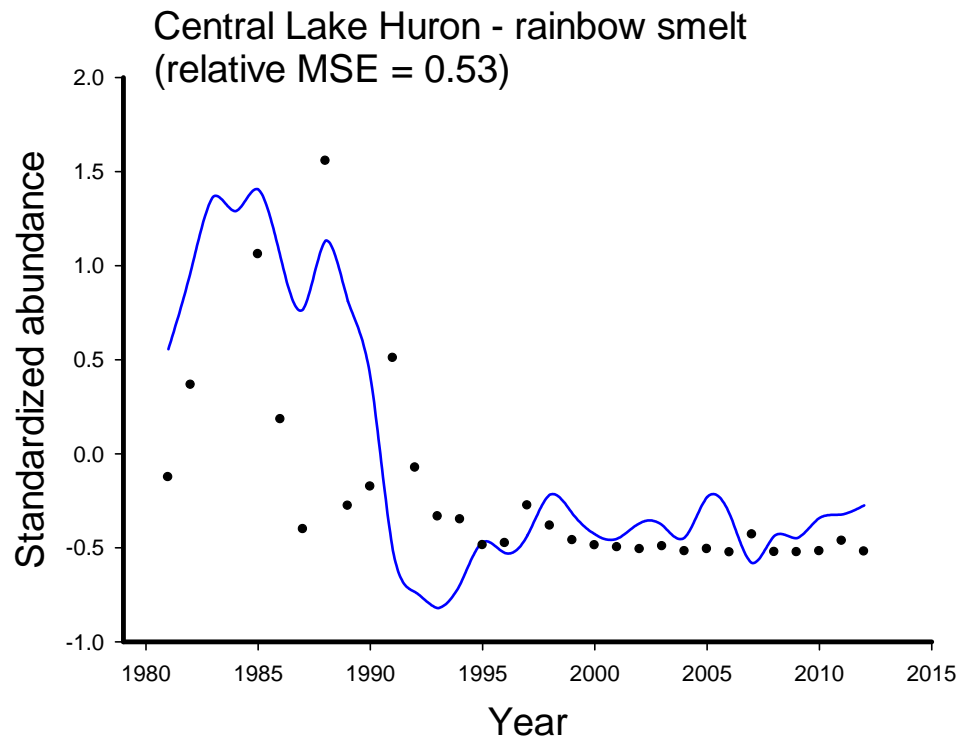
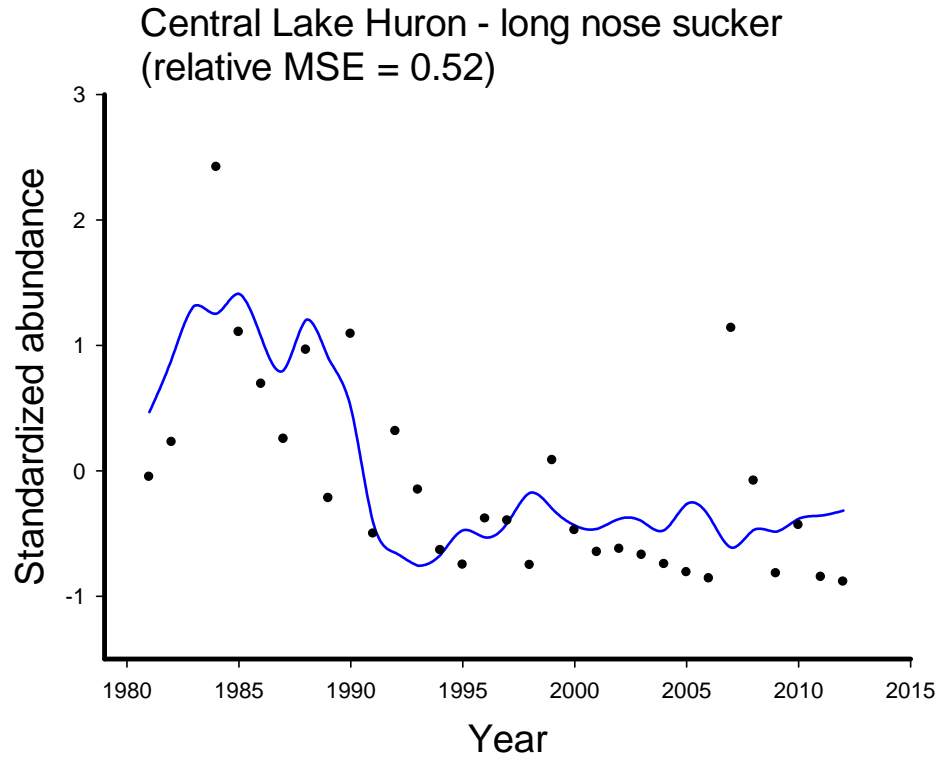
Figure 3. Estimated latent trends for the three sampling sites not located in the main basin of Lake Huron. Note the similarity of the two Georgian Bay trends.

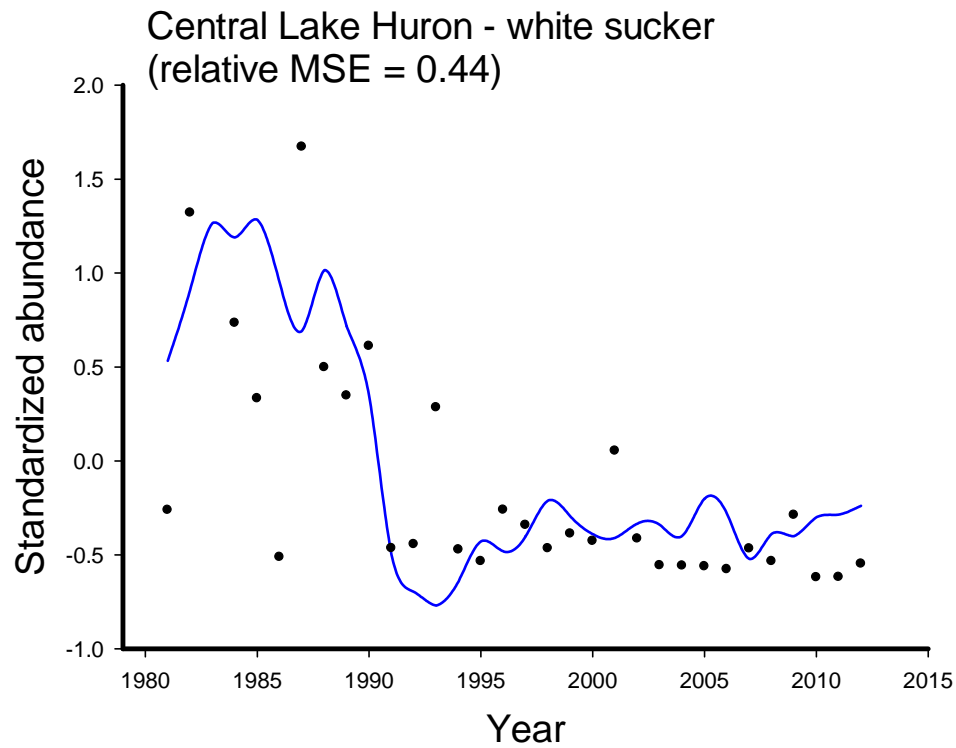
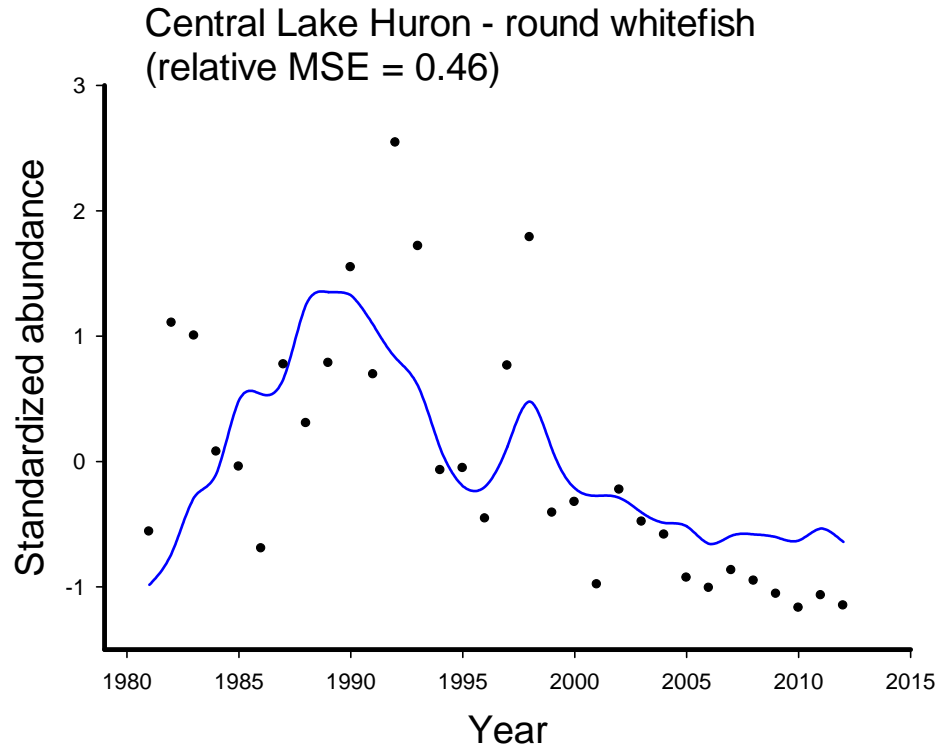
Appendix 1. Observed standardized abundance (dots) and estimated DFA trends (lines) for individual species from the central Lake Huron time series.

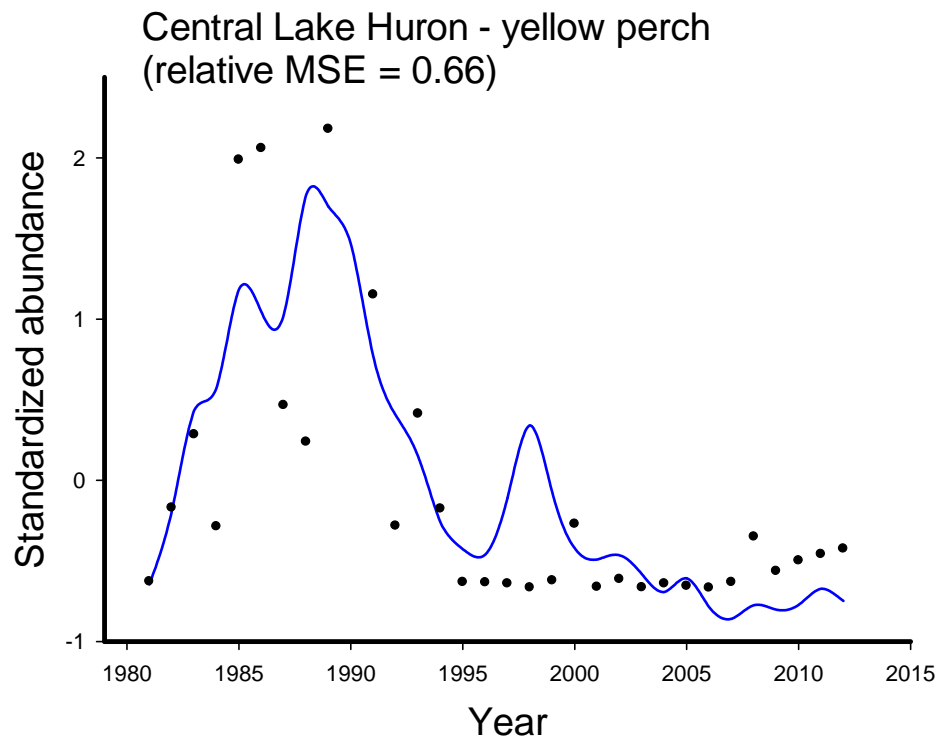






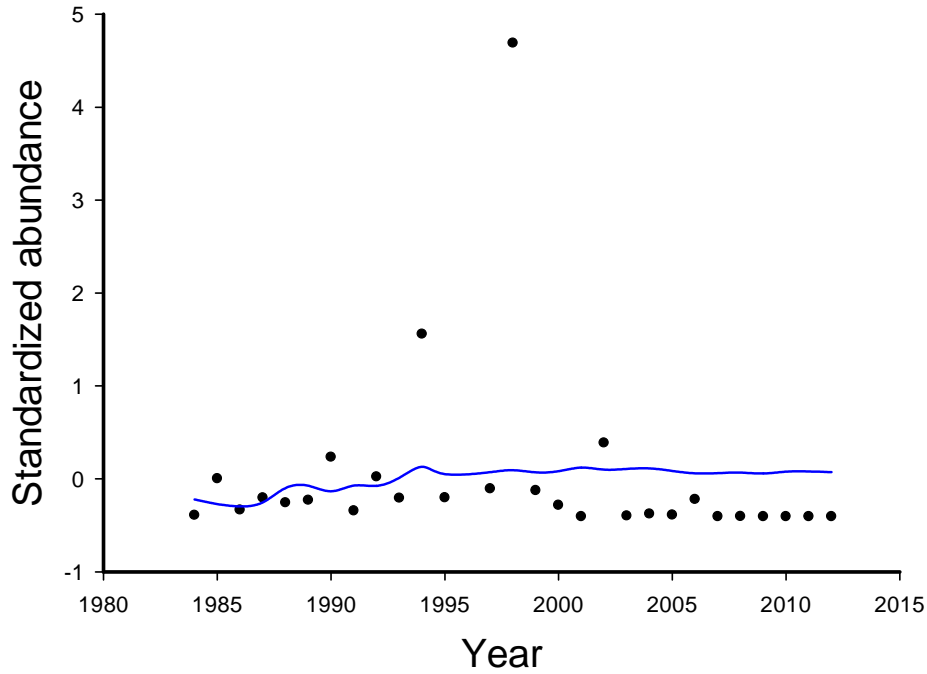




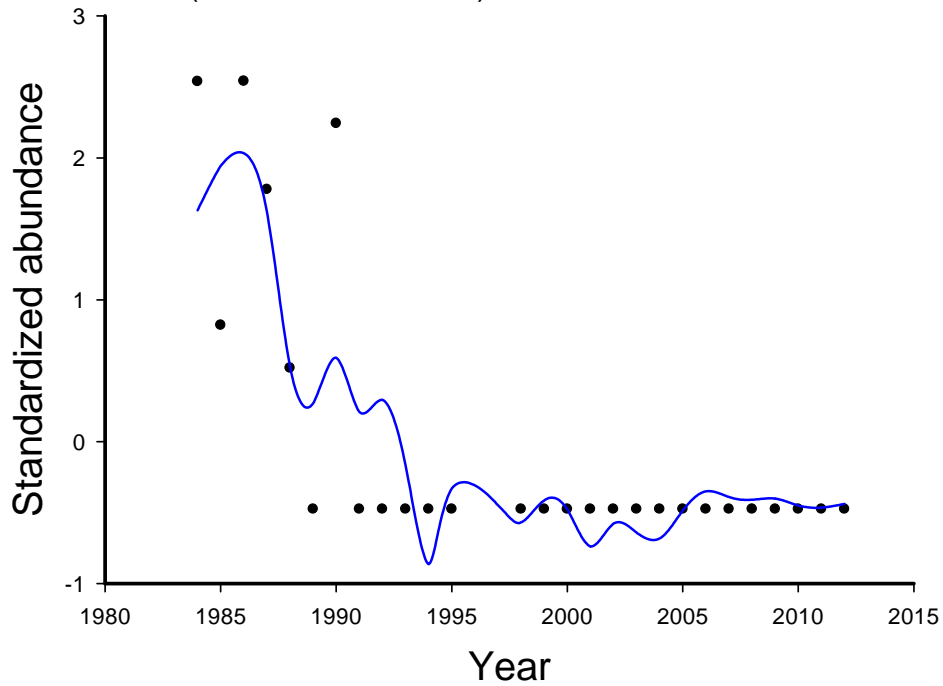


Appendix 2. Trends for individual species from the south Lake Huron time series

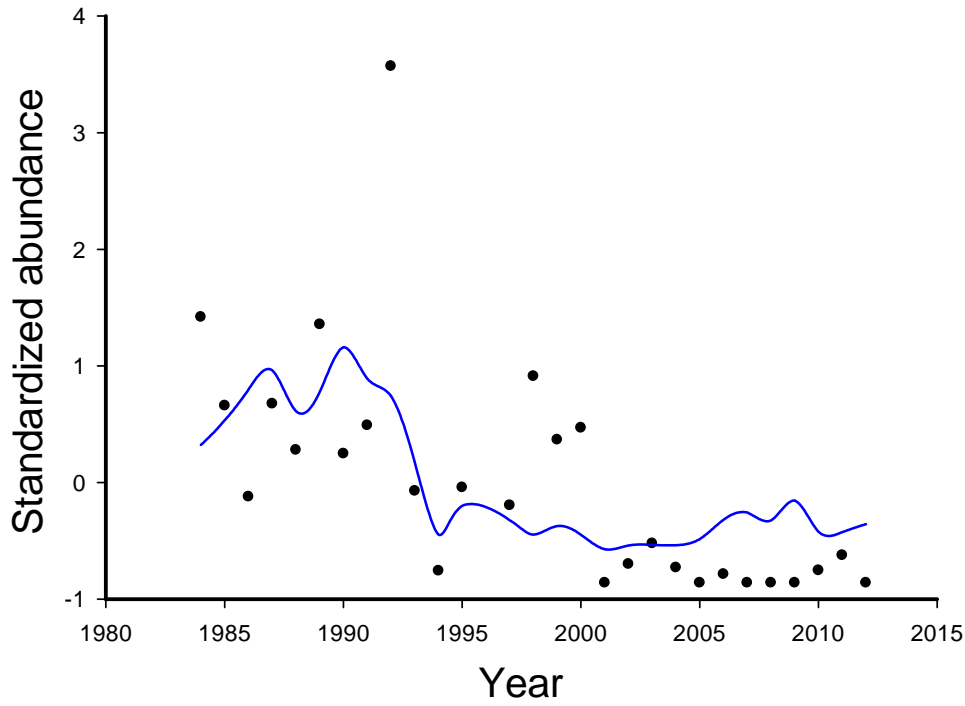
South Lake Huron - alewife
(relative MSE = 0.05)



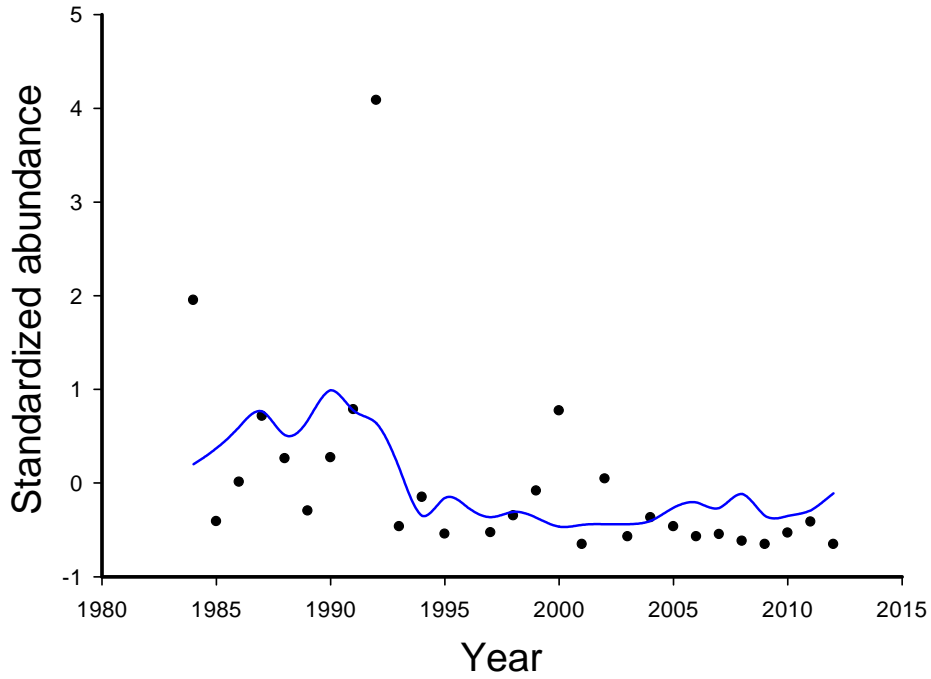
South Lake Huron - bloater
(relative MSE = 0.74)



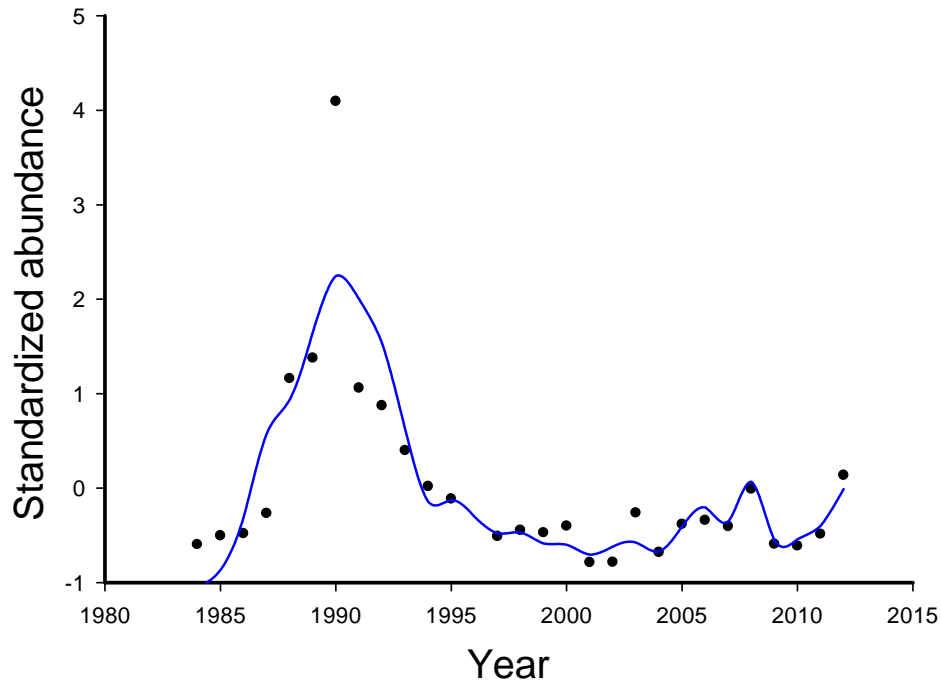
South Lake Huron - burbot
(relative MSE = 0.37)



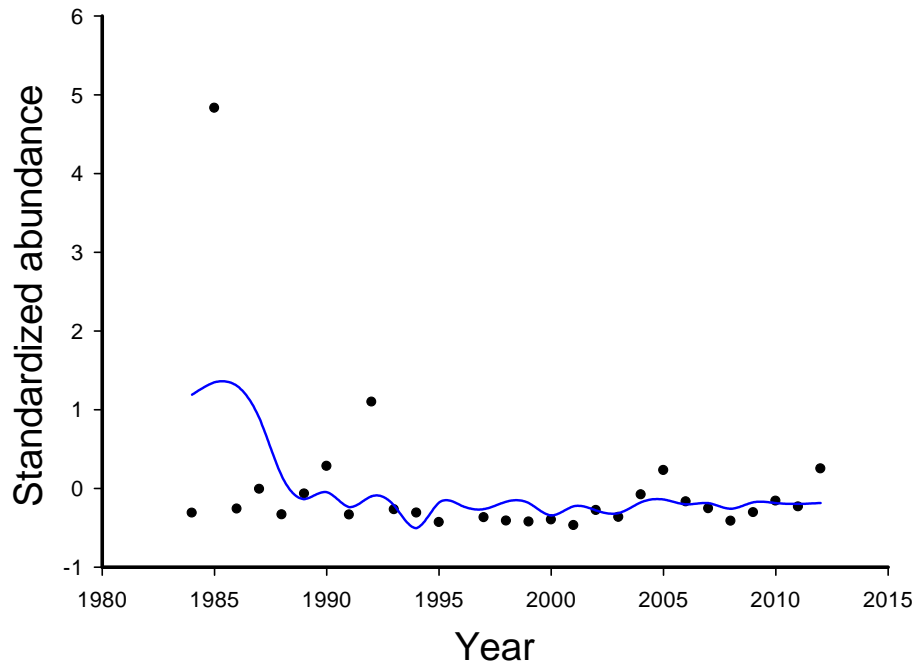
South Lake Huron - chinook salmon
(relative MSE = 0.25)



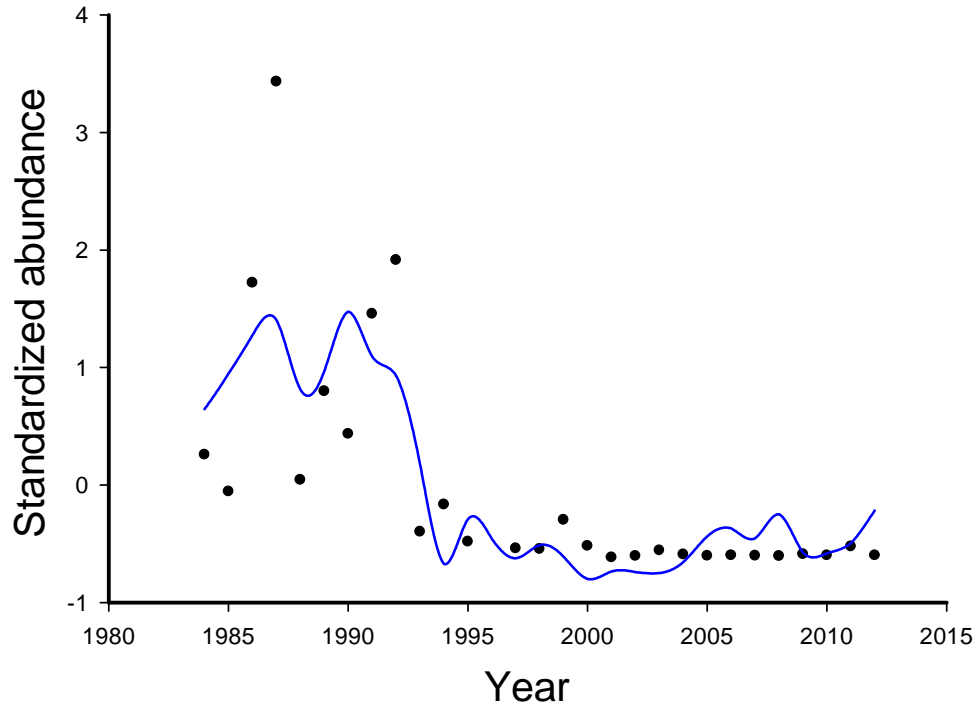
South Lake Huron - lake whitefish
(relative MSE = 0.78)



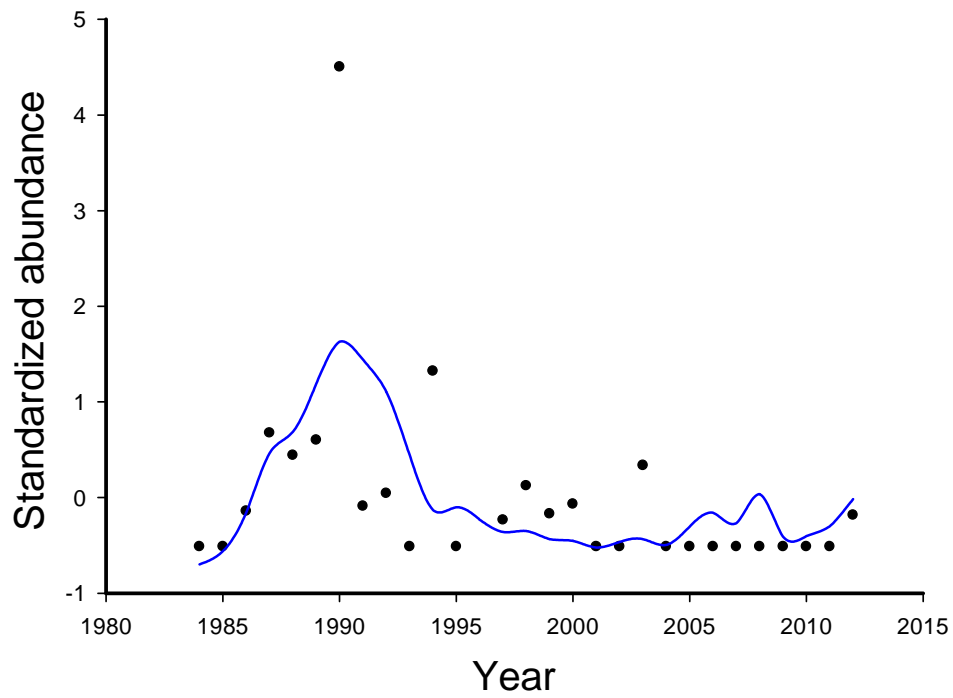
South Lake Huron - long-nosed sucker
(relative MSE = 0.28)



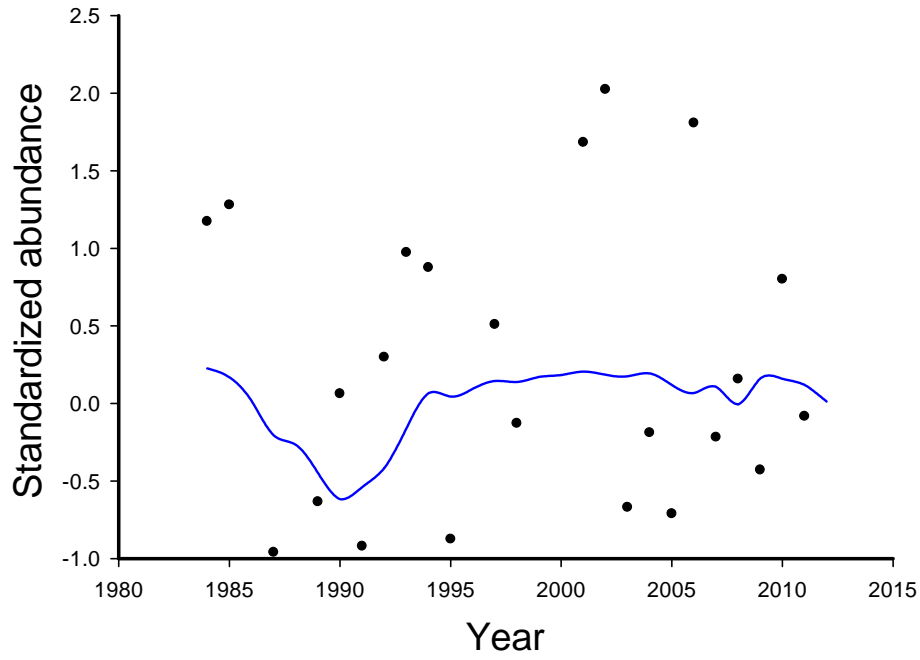
South Lake Huron - rainbow smelt
(relative MSE = 0.66)



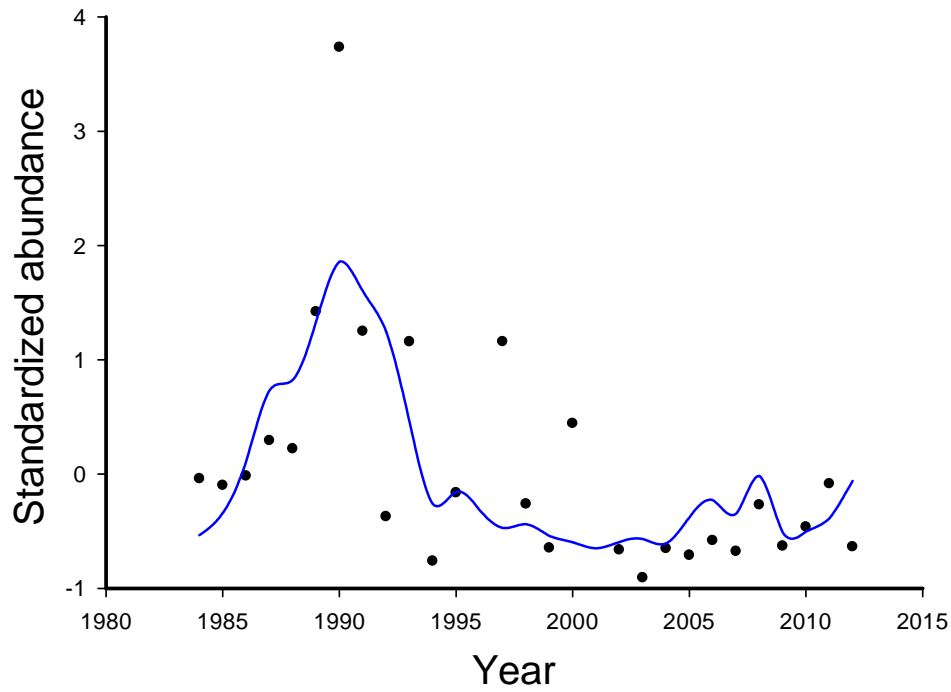
South Lake Huron - round whitefish
(relative MSE = 0.39)



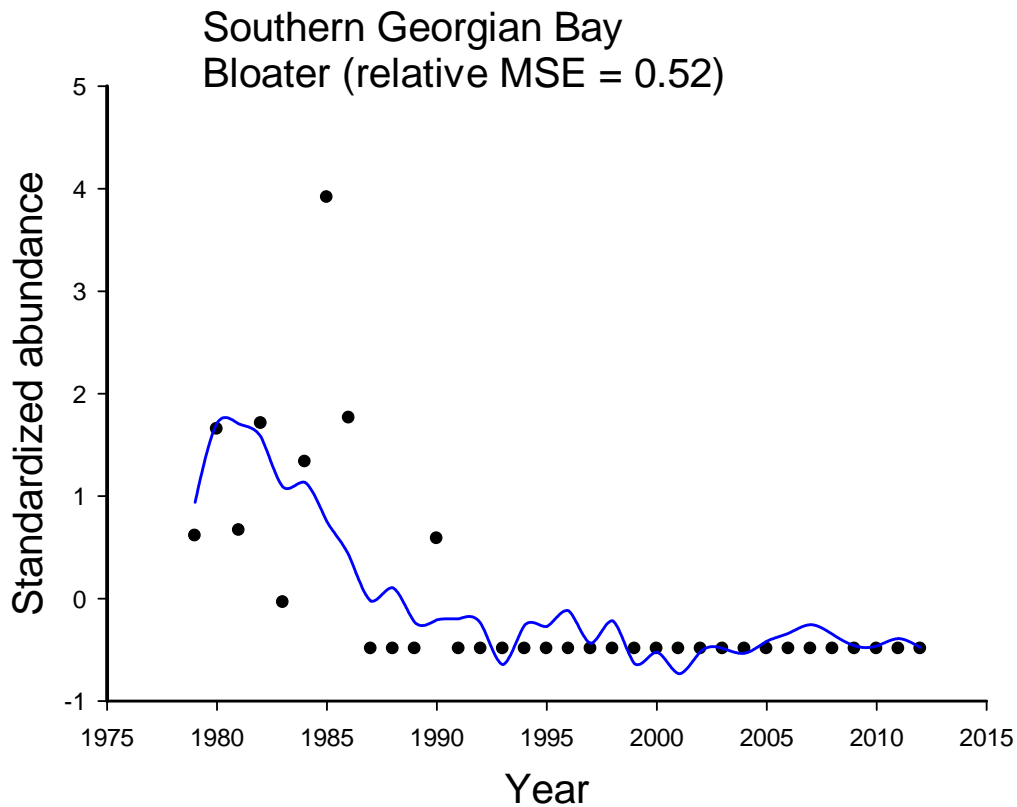
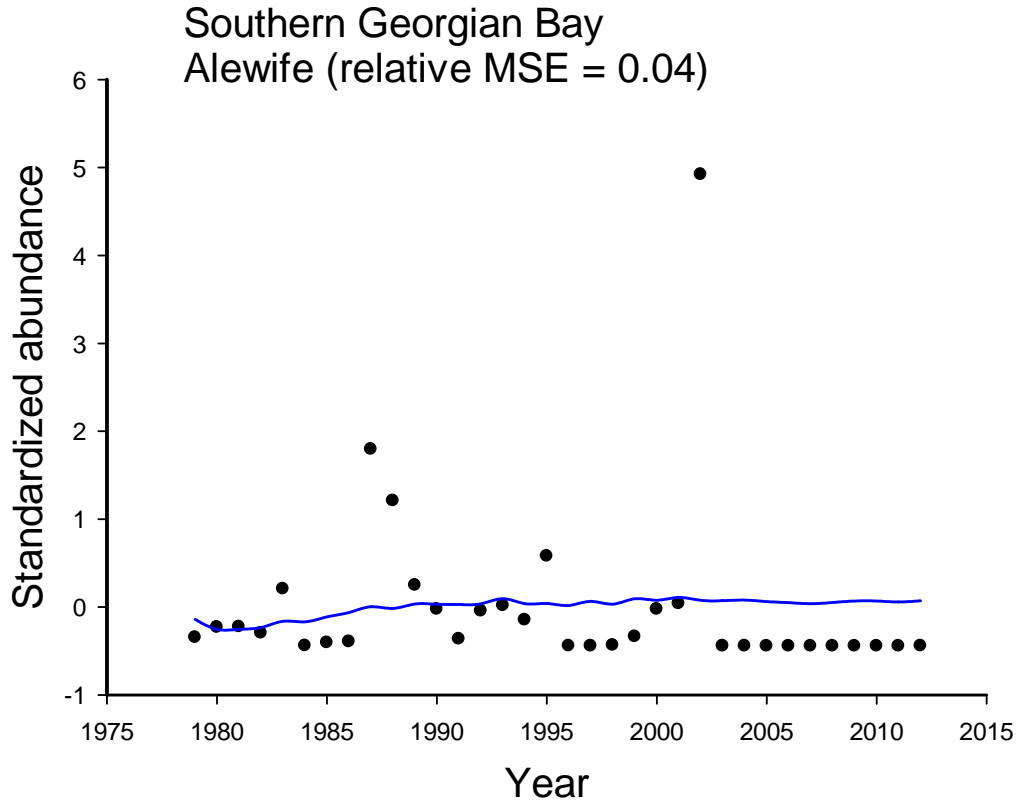
South Lake Huron - white sucker
(relative MSE = 0.09)



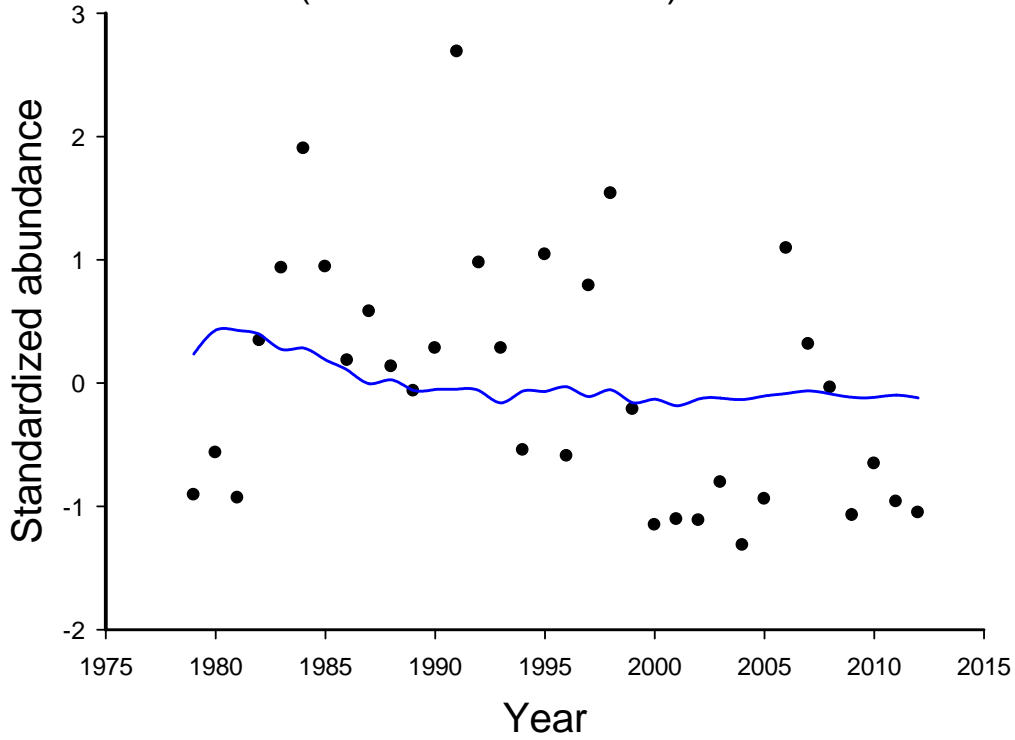
South Lake Huron - yellow perch
(relative MSE = 0.54)



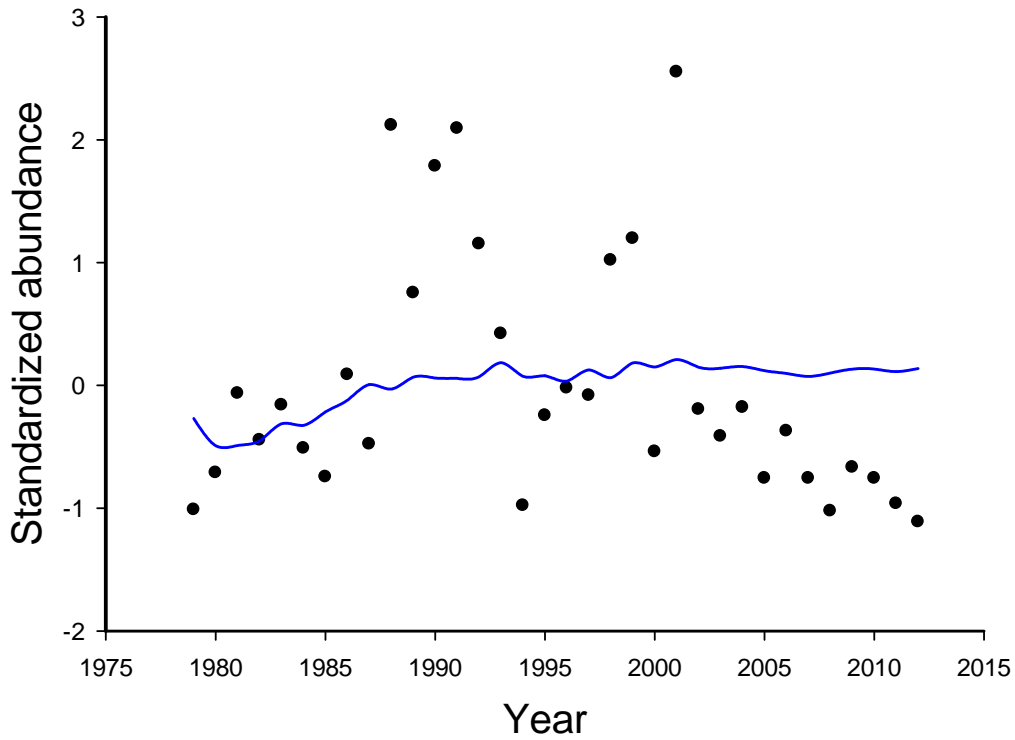
Appendix 3. Observed standardized abundance (dots) and estimated DFA trends (lines) for individual species from the south Georgian Bay time series

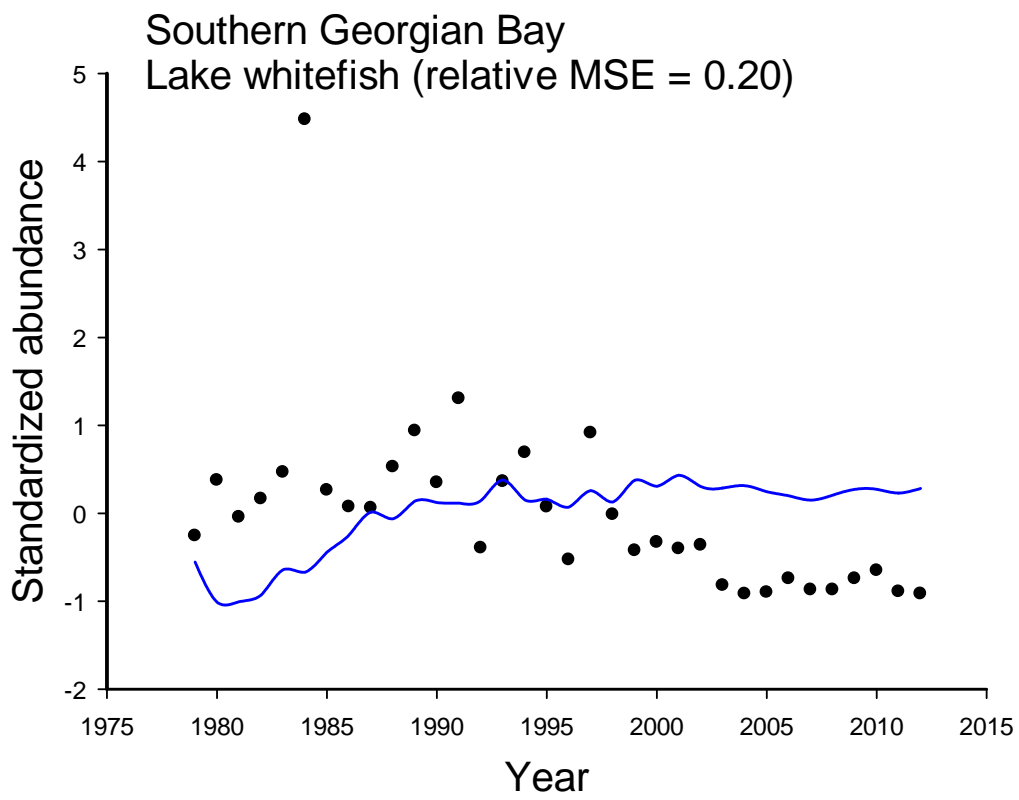
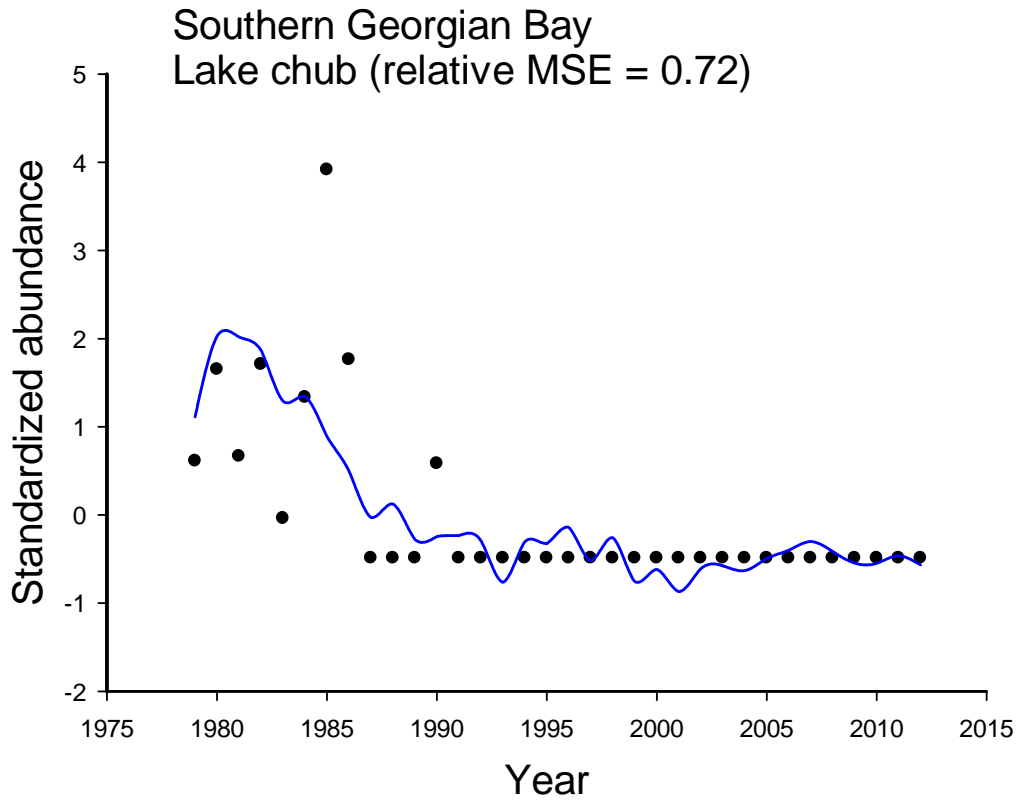


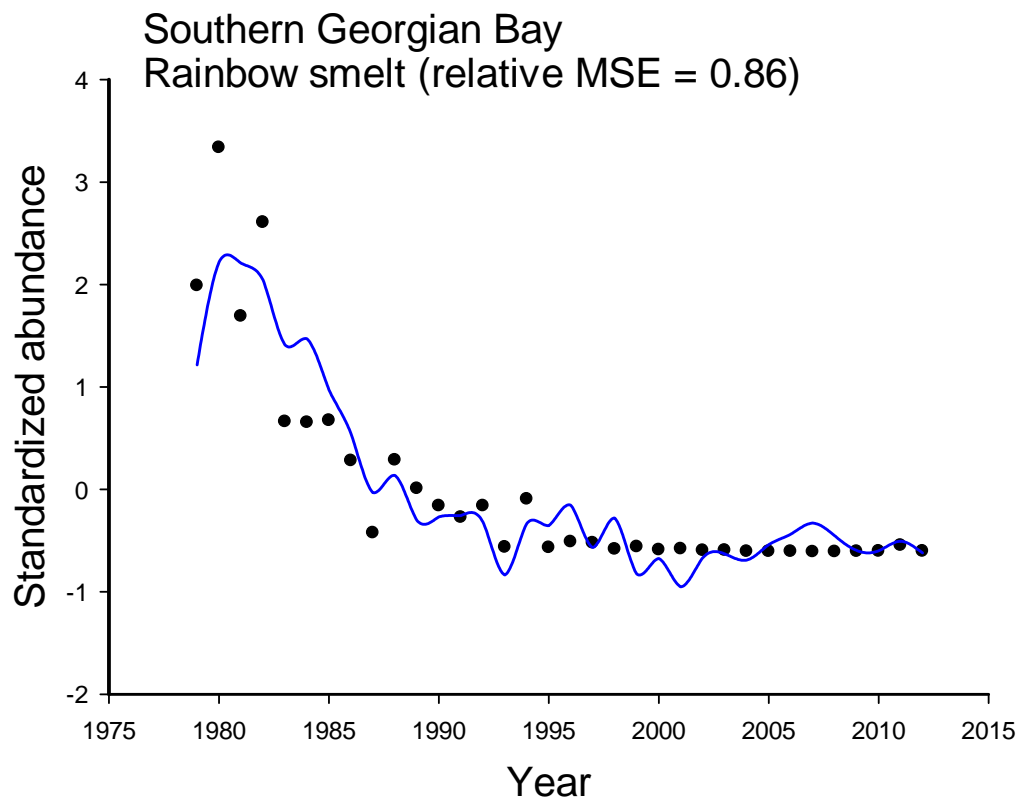
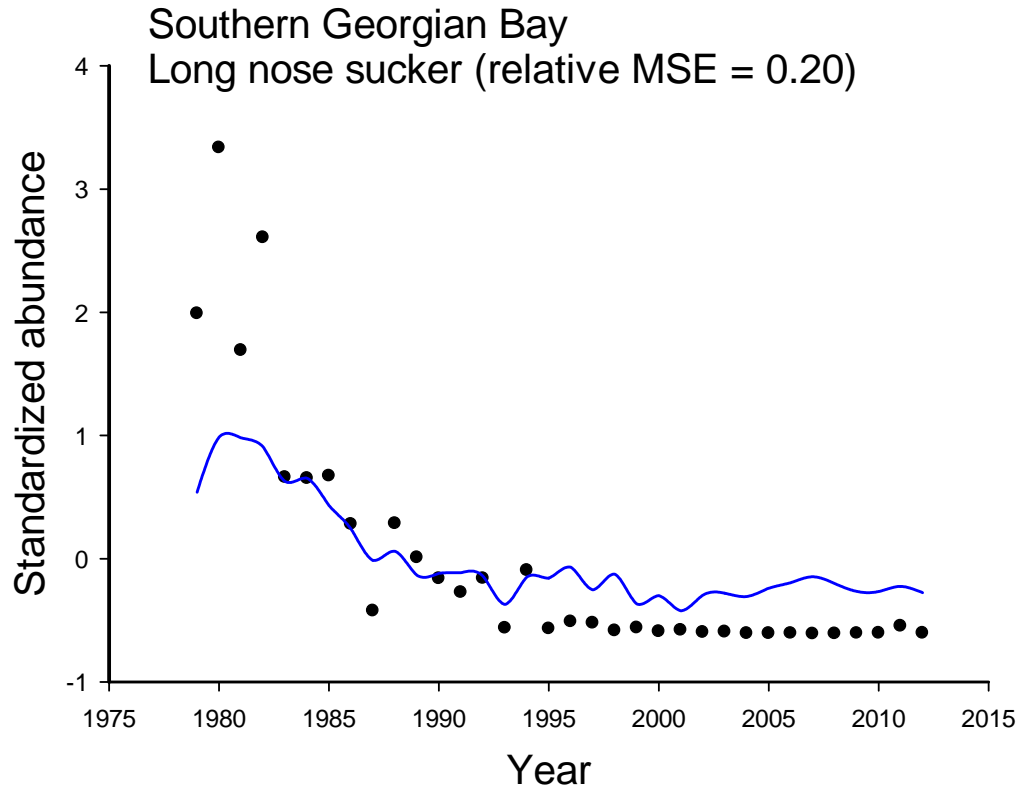
Southern Georgian Bay
Burbot (relative MSE = 0.06)

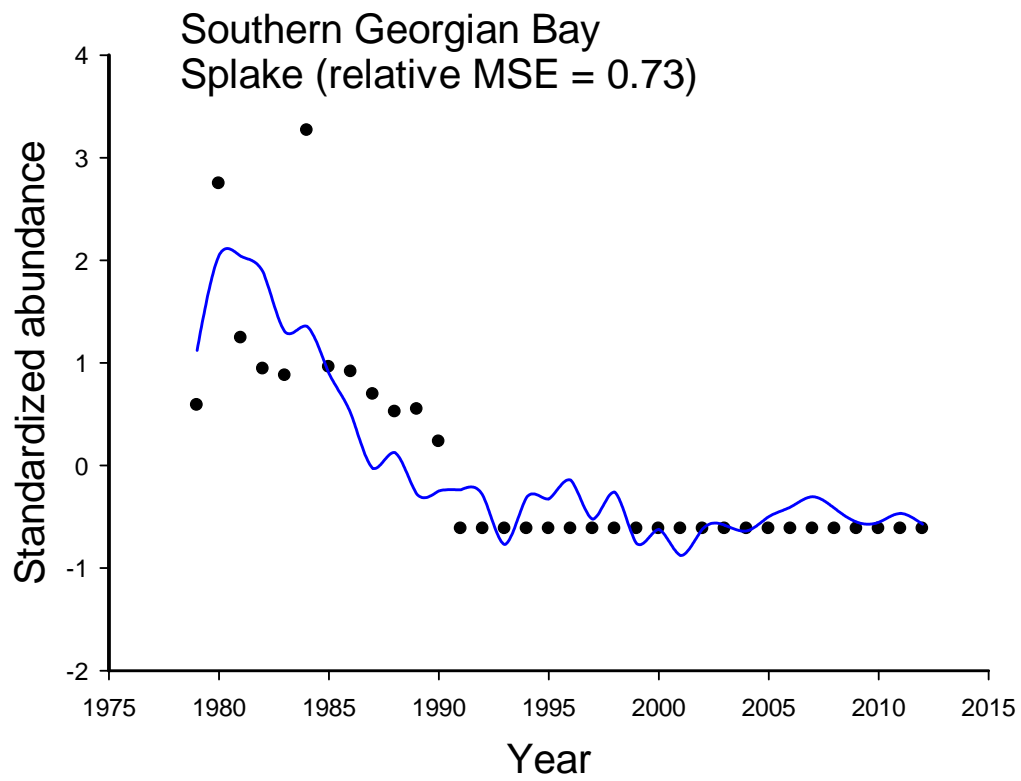
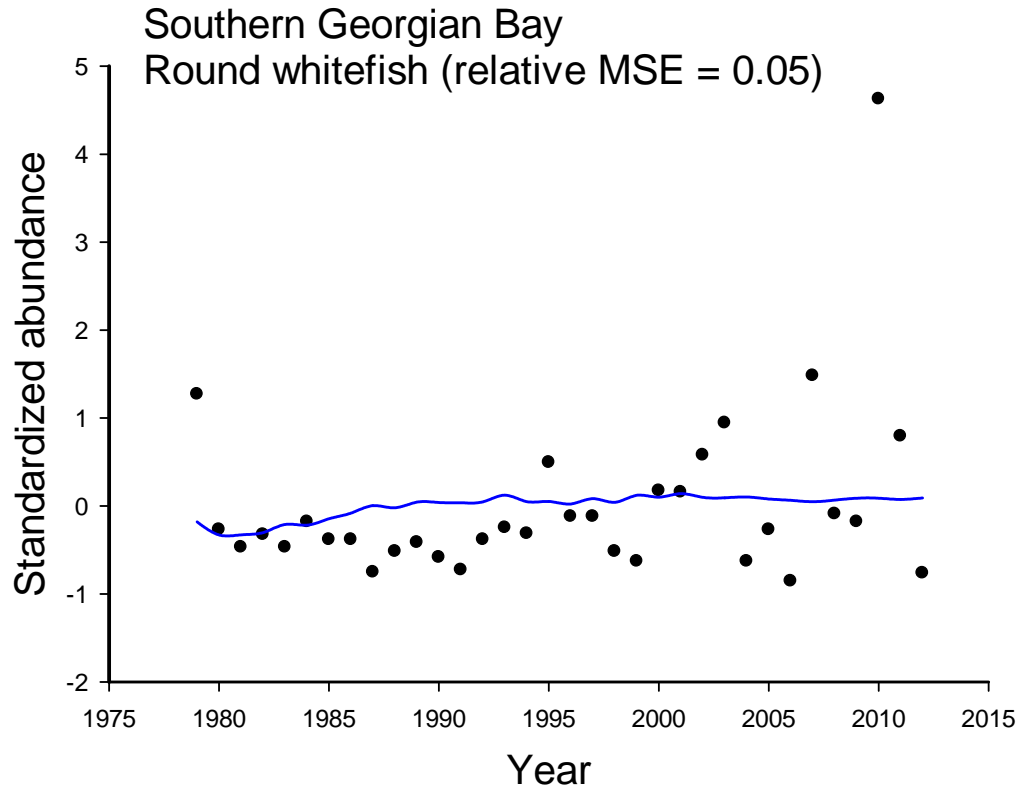


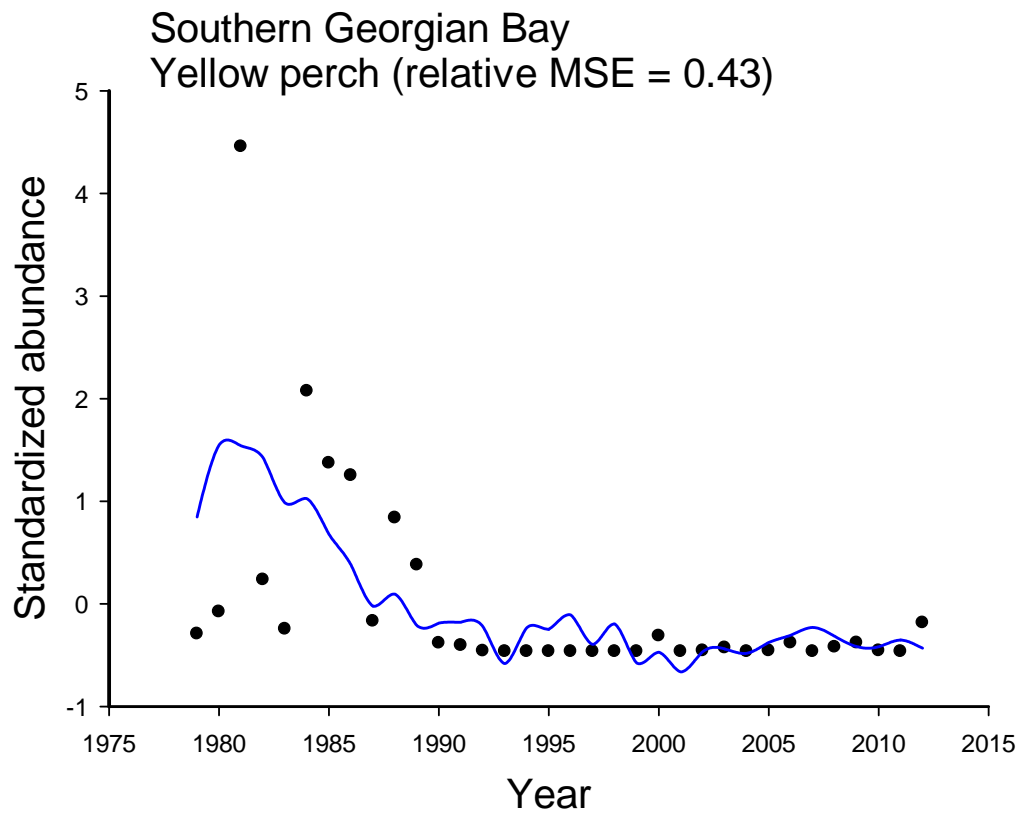
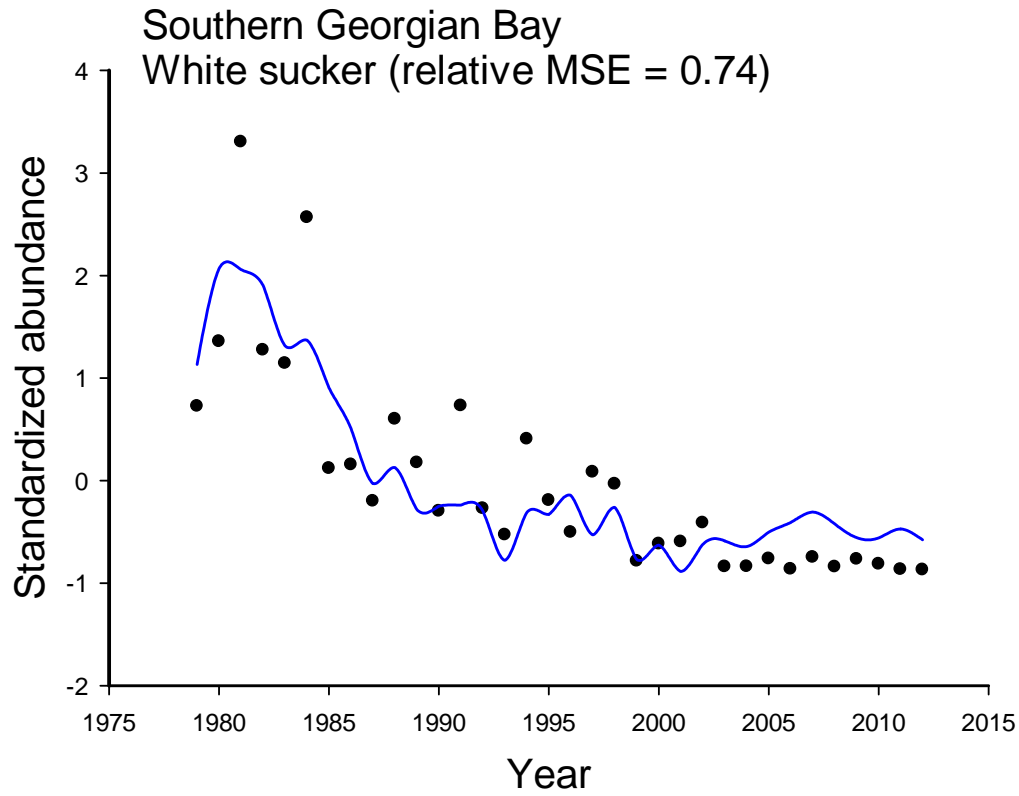
Southern Georgian Bay
Chinook salmon (relative MSE = 0.07)



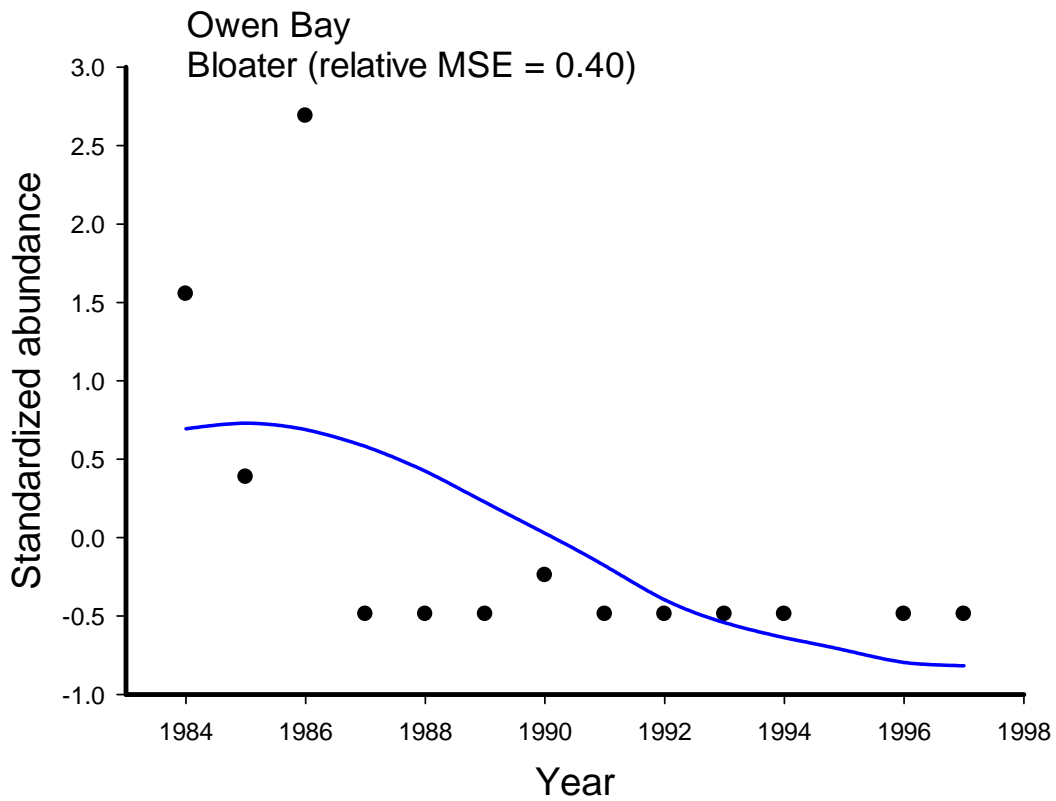
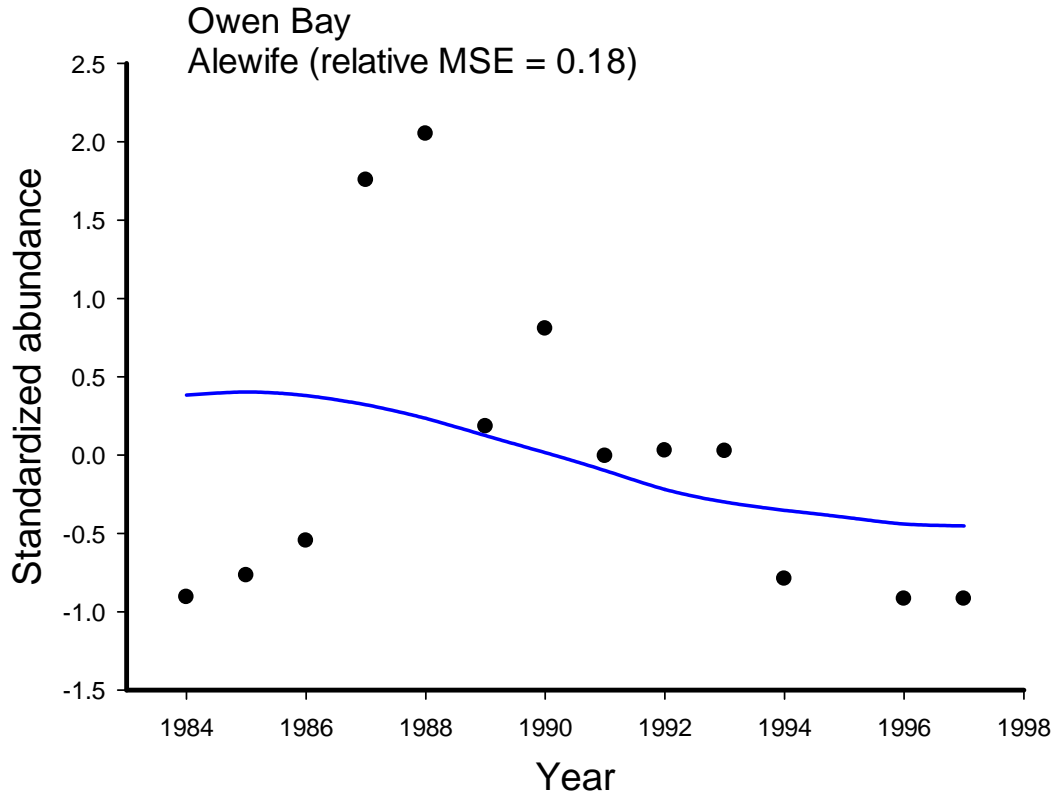


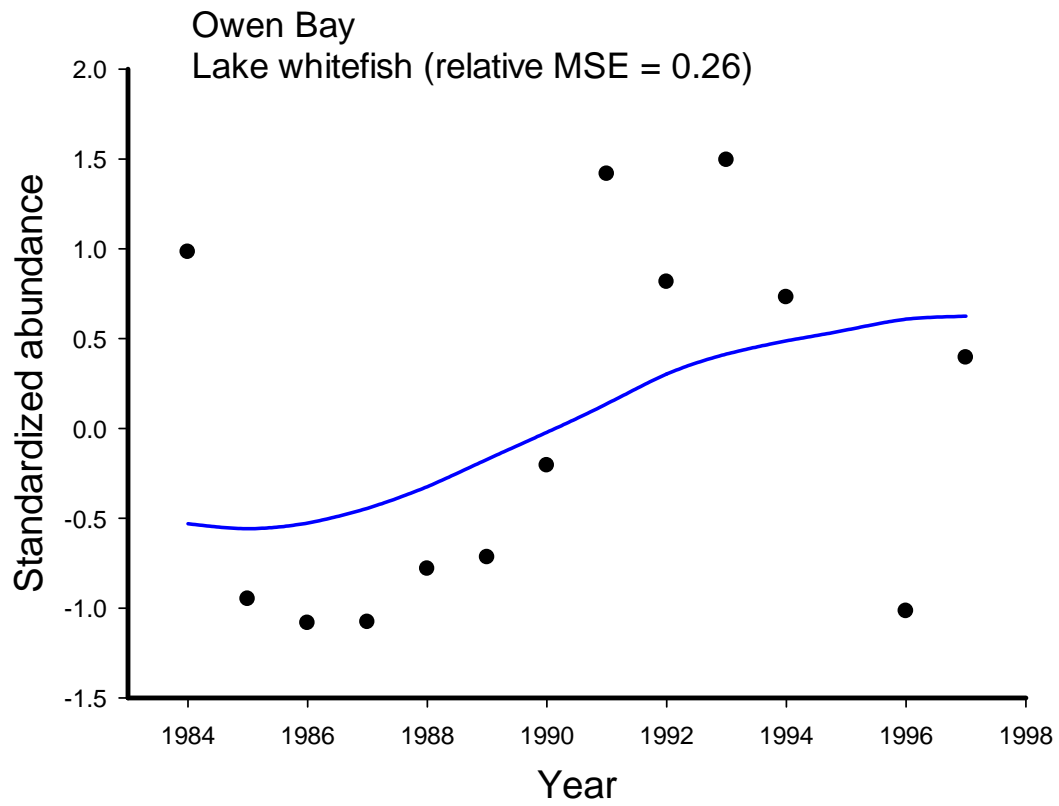
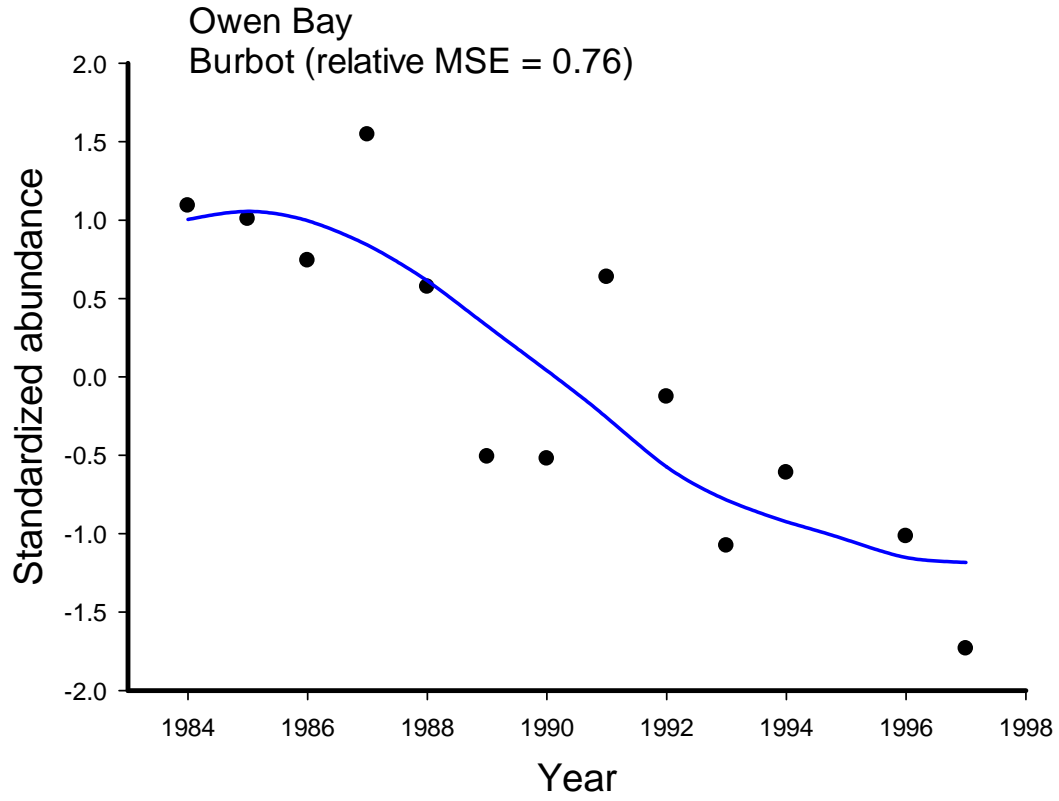


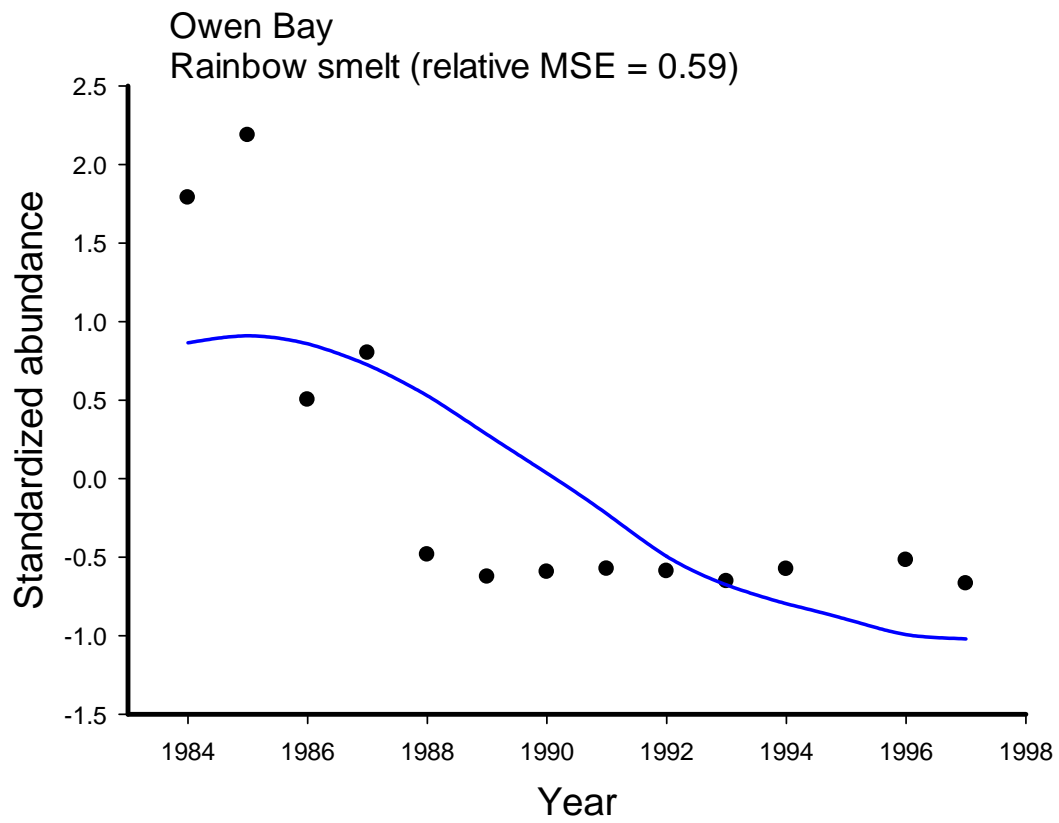
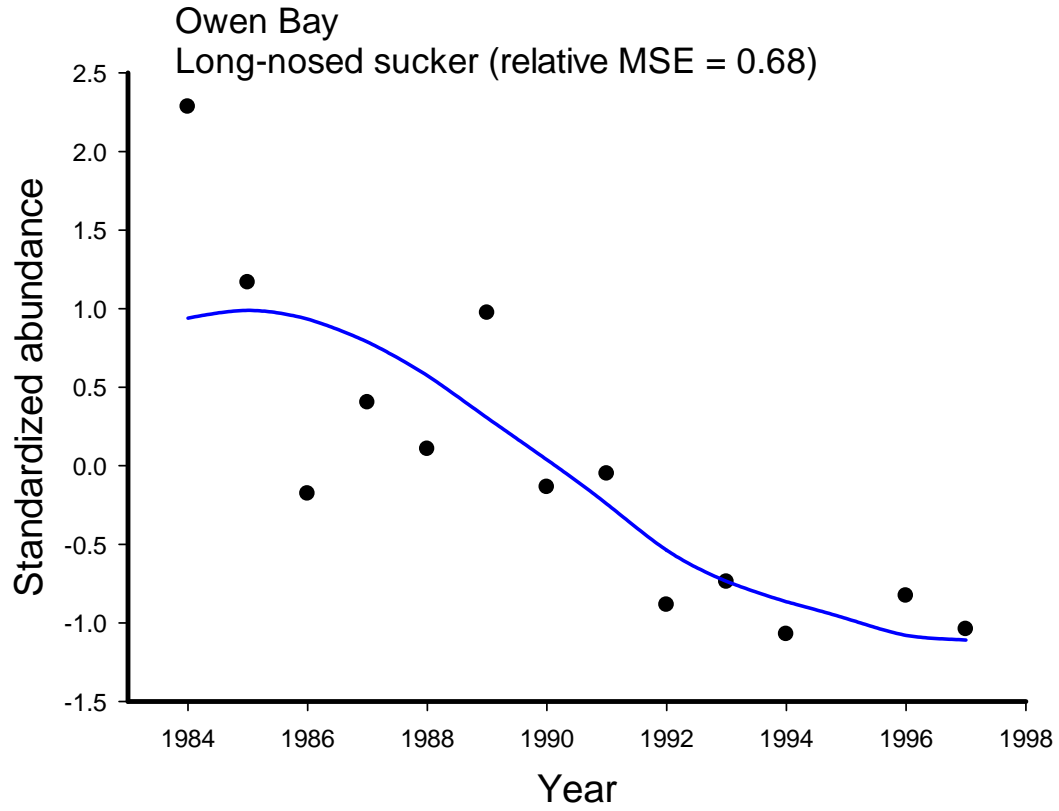


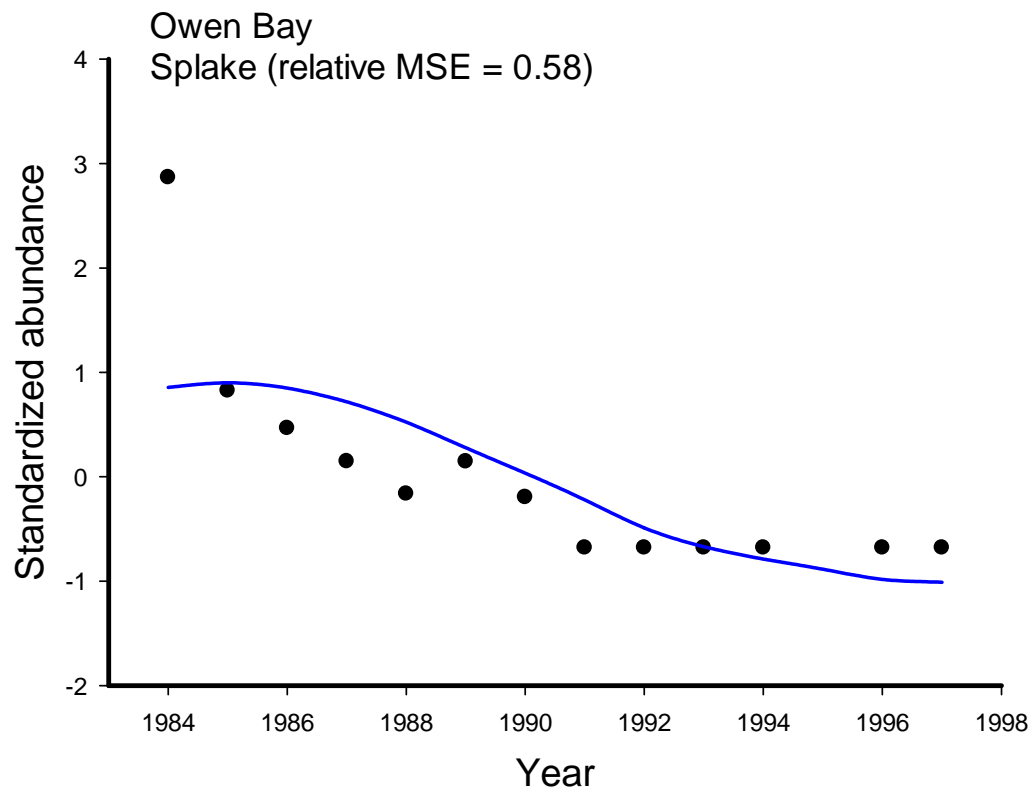
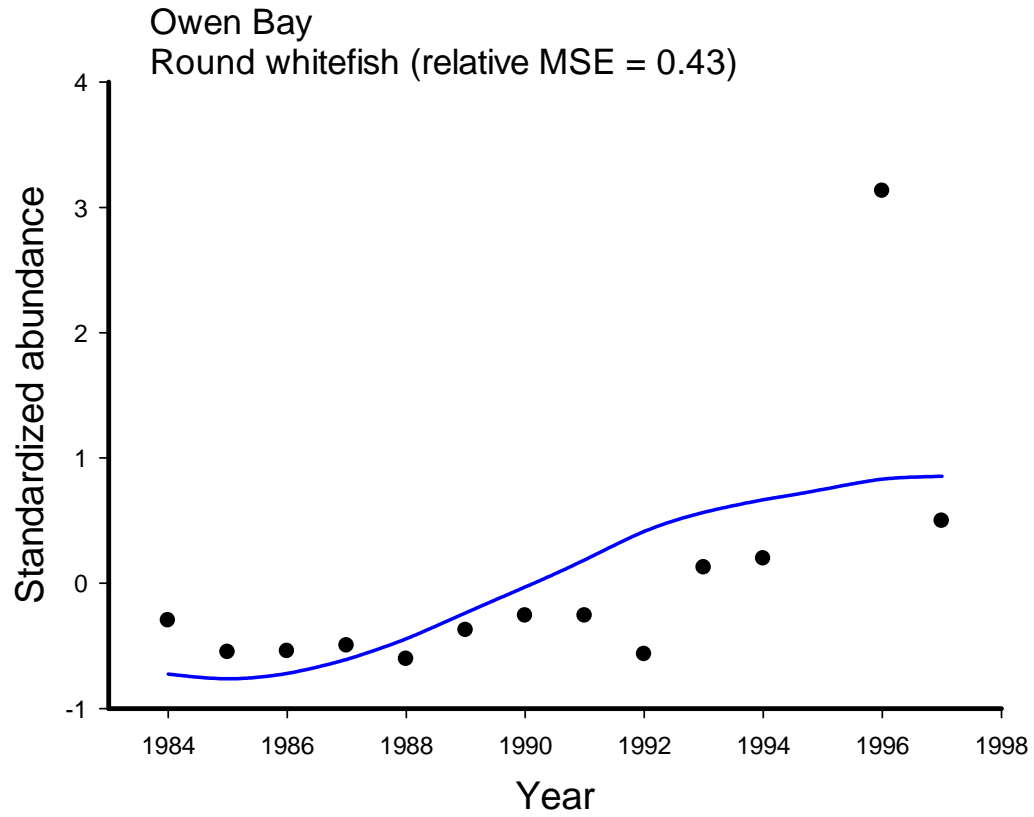


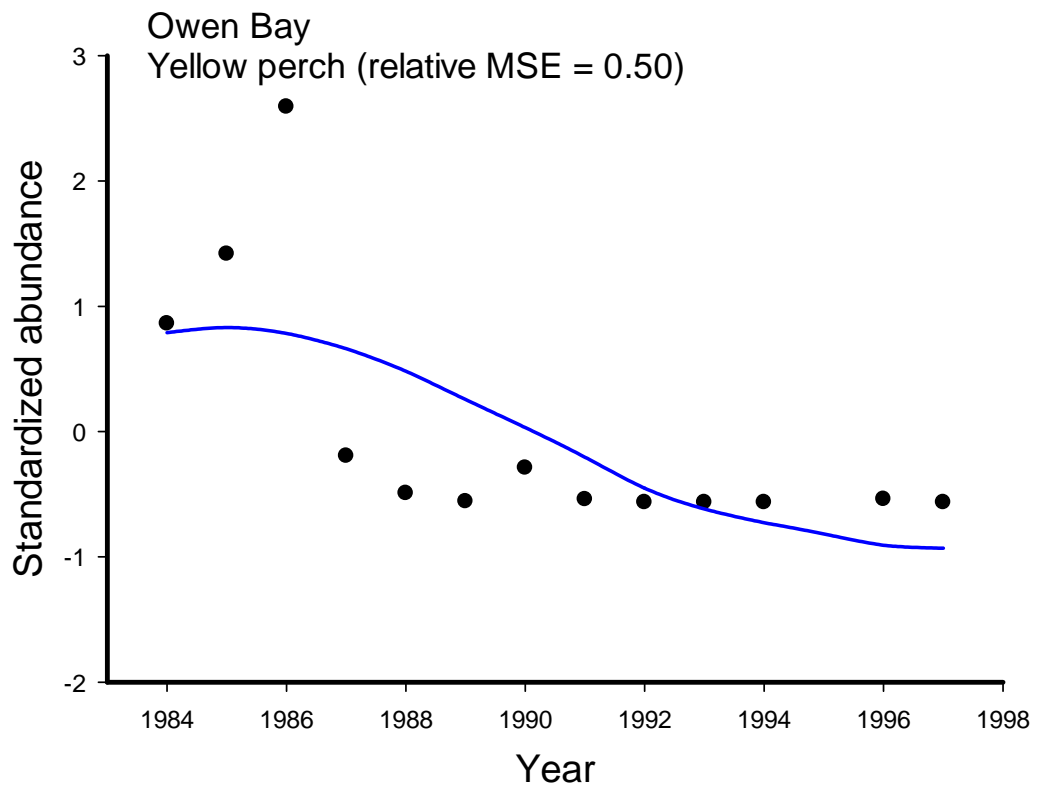
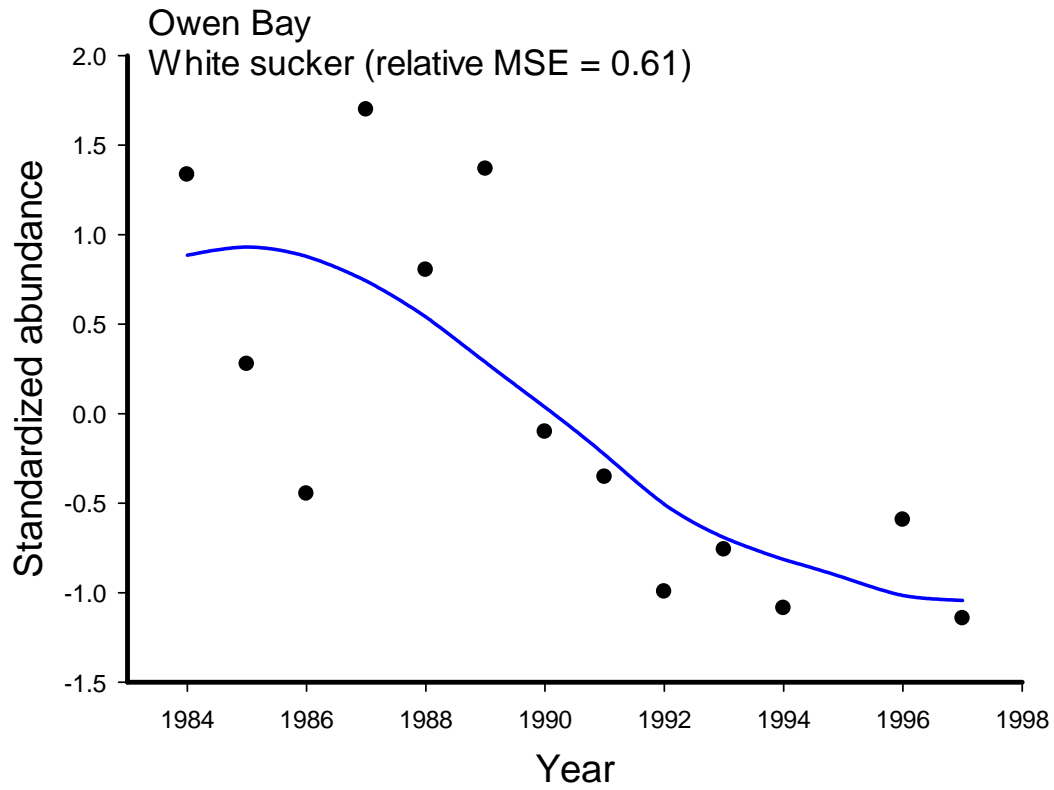
Appendix 4. Observed standardized abundance (dots) and estimated DFA trends (lines) for individual species from the Owen Sound time series











**Appendix 5. Trends for individual species from the eastern North Channel
time series**

