Trial harvest dynamics classification

C. Minto: discussion text only

Introduction

In the simulation-testing part of Group 1, we generated decision trees (what method performs best under what conditions?). The harvest dynamics were included as known variables (coupled: HD0.6; flat: HD0; roller coaster: RC; and one-way trip: OW). For the real data, we won't know this a-priori. We therefore need to discuss potential methods for classifying harvest dynamics based on catch dynamics. Given the simplified discrete catch equation

$$C = HB \tag{1}$$

where C is catch, H is the proportion of the population biomass B harvested; it is clear that the catch-dynamics are affected by both the harvest and biomass dynamics. In fact, it is the decoupling of these processes which lies at the heart of data-poor assessment methods. So from the beginning, classifying harvest dynamics based on catch dynamics is likely confounded by what's going on with biomass - without invoking embedding theorems. Nowithstanding, it is still interesting and potentially useful to attempt to classify catch to harvest dynamics broadly. That's the focus of this brief document, which is meant only as a discussion piece for the group, not as a definitive approach we'll take.

Methods

The simulated data were split into half for training (Figure 1) and half for testing (Figure 2). It can be seen that many of the HD0.6 series are in phase. Many time series clustering techniques focus on common factors or trends among time series referenced by specific years (e.g., dynamic factor analysis; time series factor analysis). We are not interested in the common factors but the dynamic features of the time series. For example, two out-of-phase cyclical time series might not share common factors but might have very similar underlying dynamics. One approach is to convert the time series from the time domain to the frequency domain, where the time series is represented by a spectrogram (Priestley, 1982). The spectrograms of multiple time series may then be clustered to common groupings of types of time series, irrespective of whether they are in phase or not.

There are multiple ways of deriving spectrograms, the method used here is a model-based AR(p) (using the `spec.ar` function in R). Given the time series length differs in the real and simulated time series, the default frequencies evaluated using `spec.ar` will differ, they were therefore linearly interpolated to a common set of frequencies (Figure 3).

The HD0.6 spectrograms are characterized by a peak at a frequency around 0.07 corresponding to a period of 1/0.07$\approx$14.3 years, though there is also higher density at lower frequencies capturing low frequency variation (Figure 1). Many of the individual ED0 spectrograms are flat as might be expected from non-directional white noise (Figure 3). The one-way trip and

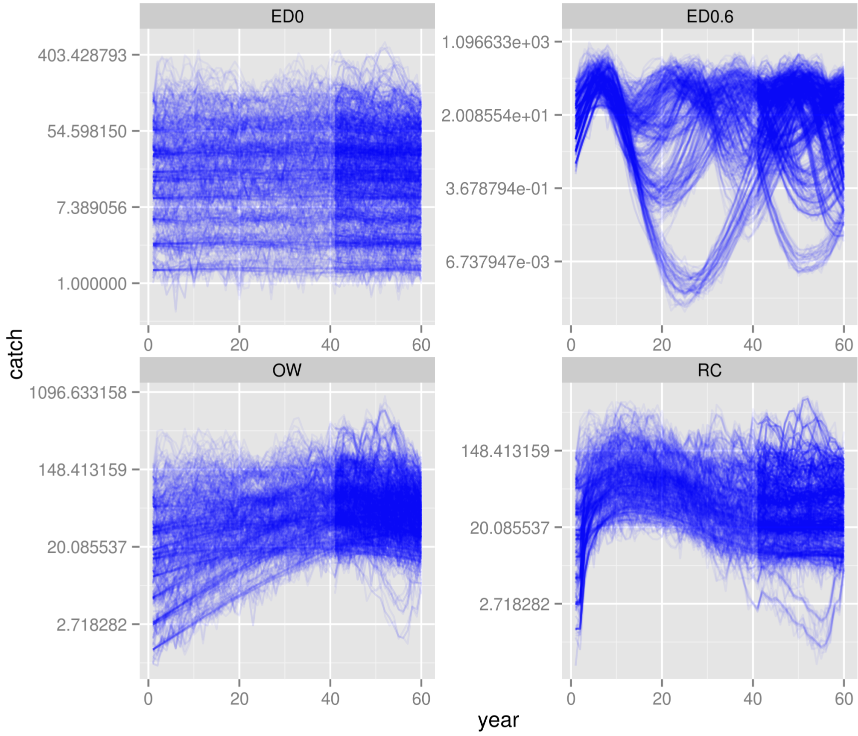

Figure 1: Training data used for generating a classification tree for harvest dynamics. Y-axis on natural logarithmic scale.

roller coaster spectrograms are similar with peaks at low frequencies representing longer term dynamics (Figure 1). Across harvest dynamics the spectrograms are similar in higher frequencies, likely representing the common recruitment and catch measurement errors added to the simulations.

A classification tree was built on the training dataset using known labels as the response and the estimated spectral densities at given frequencies as explanatory variables. The in-bag (same data train, same data fit) classifications accuracy was 53.7% (Table 1). One-way trip and roller coaster dynamics appear particularly difficult to separate in this manner. Out-of-bag classification is similar at 53.4% (Table 2). Under random assignment we would expect about 25% accuracy.

Code and data to conduct the analysis here is located in the dropbox folder:

```
FisheriesWorkingGroupPhaseII/Decision_trees/harvest_dynamics_classification
```

Table 1: In-bag classification of the training data harvest dynamics based on a classification tree of the catch time series spectrograms.

		Classified			
		ED0	ED0.6	OW	RC
True	ED0	1101	43	213	83
	ED0.6	38	1039	338	25
	OW	463	87	586	304
	RC	479	87	508	366

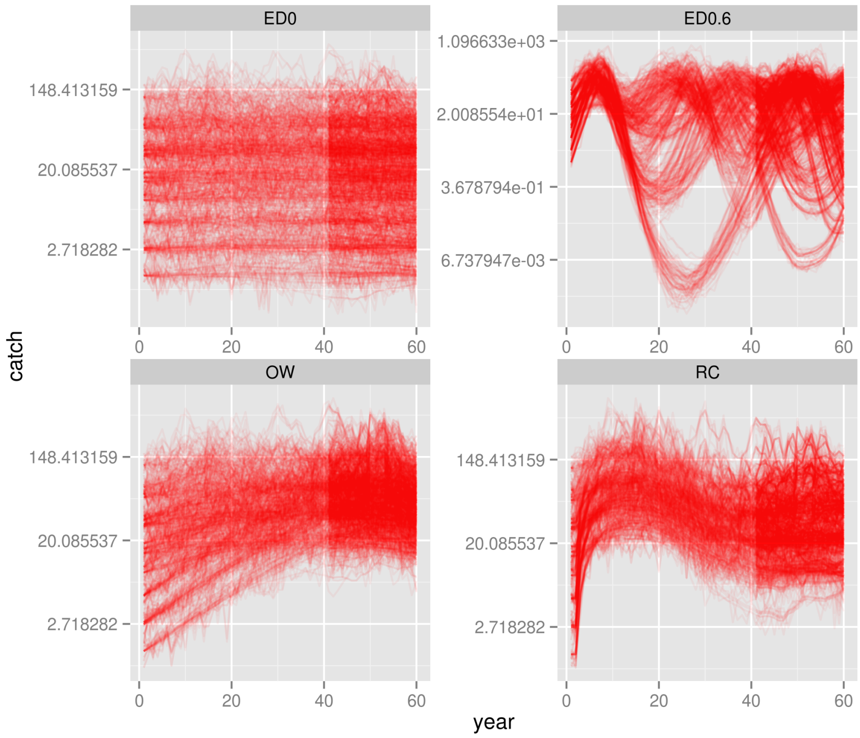

Figure 2: Testing data used for generating a classification tree for harvest dynamics. Y-axis on natural logarithmic scale.

Table 2: Out-of-bag classification of the testing data harvest dynamics based on a classification tree of the training catch time series spectrograms.

		Classified			
		ED0	ED0.6	OW	RC
True	ED0	1093	37	231	79
	ED0.6	43	1017	347	33
	OW	468	93	566	313
	RC	468	78	493	401

Applying the method to the FAO catch series resulted in classifications: HD0 (11%); HD0.6 (75%); OW (14%); RC (<1%) (Figure 4).

Discussion

Somehwat expectedly given the absence of biomass dynamics, the classification performance is not great; it is considerably better than chance but still miss-classifies many. Also, (not presented), the method is sensitive to decisions on the form of the spectrogram (e.g., using smoothed periodograms results in many more OW classifications). Other manifest variables could be included in the classification tree. Could perform similar analysis on lagged components of the catch series to see if the harvest component can be further isolated? Other approaches?

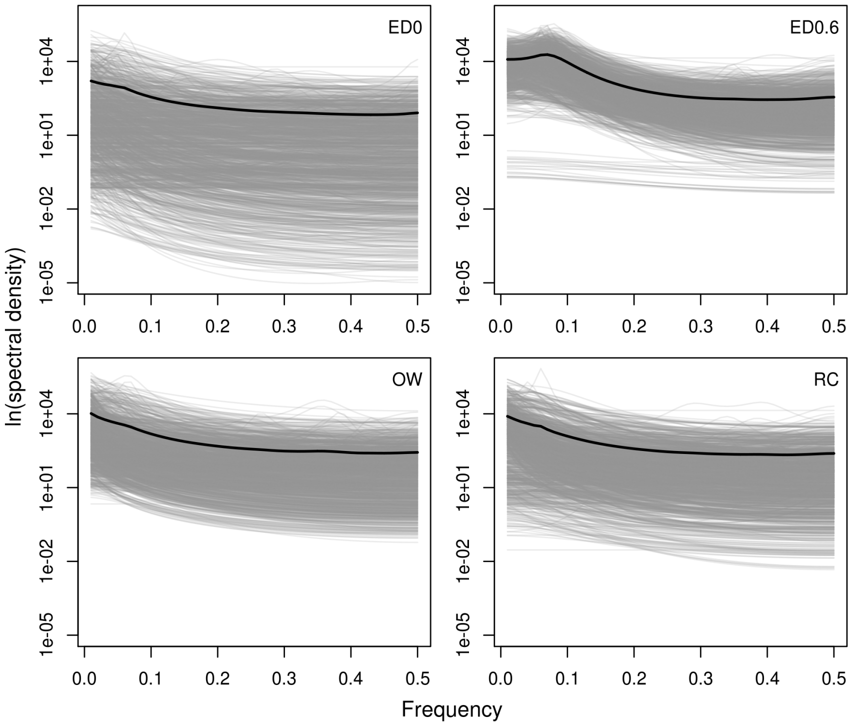

Figure 3: Estimated training data spectrograms of catch time series grouped by effort dynamics. Each grey line represents a catch time series and the black line represents the overall mean. Y-axis on natural logarithmic scale.

References

Priestley, B. (1982). Spectral analysis and time series. Probability and mathematical statistics. Academic Press.

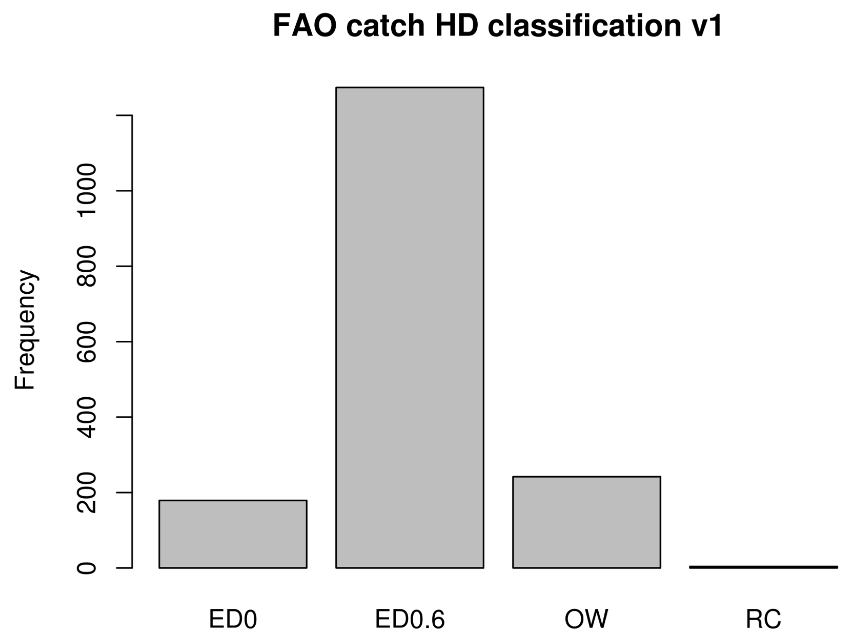

Figure 4: FAO classifications based on the spectrogram clustering method.