# Investigation of the Association Between Hawaii Deep Slope Bottomfish CPUE and Environmental Variables[1]

**Hui-Hua Lee**
University of Hawaii
Joint Institute for Marine and Atmospheric Research
NOAA Pacific Islands Fisheries Science Center
250 Dole Street
Honolulu, HI 96822

**Jon Brodziak**
Pacific Islands Fisheries Science Center
NOAA Pacific Islands Fisheries Science Center
2570 Dole Street
Honolulu, HI 96822

[1] PIFSC Internal Report IR-11-019
Issued 28 October 2011

## Preface

In 2010 and early 2011, PIFSC researchers completed a new stock assessment of Hawaii bottomfish in the main Hawaiian Islands. The research was peer reviewed and, after revisions, released in October 2011 as a NOAA Technical Memorandum. In concert with the stock assessment, several supporting documents were drafted by PIFSC scientists to address ancillary information and technical issues. Because these informal documents were cited in the stock assessment report, they are being made available to the public. This is one of those documents. It is being released in its original form, with minimal editing.

Gerard DiNardo
Fisheries Research and Monitoring Division
Pacific Islands Fisheries Science Center
Gerard.DiNardo@noaa.gov

**Investigation of the association between Hawaii deep slope bottomfish CPUE and environmental variables**

**Hui-hua Lee and Jon Brodziak**

## Introduction

With recent increases in the exploitation of fish resources, understanding the influence of environmental on variable survival and abundance of Hawaii deep slope bottomfish community is necessary for management of both fisheries and waters in the Hawaiian Archipelago. In recent years, fisheries scientists and oceanographers have found evidence of slowly changing, long-term variability in these biological and physical variables such that values in a given year are closely related to values in previous years (i.e., the data contain positive autocorrelation). This variability has been exhibited in the Northeast Pacific Ocean, for example, by extended periods of unusually high sea-surface temperatures, El Nino/La Nina episode, unusually sea-surface height, and low biological productivity followed by periods of the opposite.

Correlation analysis has been a useful and widely applied tool for generating hypotheses about the effects of environmental or other variables on survival, abundance and recruitment at these various time scales (e.g., Myers *et al.* 1995). However, a major statistical challenge exists when time series of abundance and environmental data are strongly autocorrelated (e.g., Chelton 1984; Thompson and Page 1989). Such autocorrelation violates the assumption of serial independence required for most classical inference tests (Hurlbert 1984). This means that a sample correlation between two autocorrelated time series has fewer degrees of freedom (or a larger variance) than that assumed under the classical significance test. To address this problem, fisheries scientists and oceanographers have typically applied qualitatively approach. This approach is to modify the hypothesis testing procedure by computing either corrected degrees of freedom for the sample correlation (Garrett and Petrie 1981; Chelton 1984) or, equivalently, a corrected variance for the sample correlation (Kope and Botsford 1990). In addition, investigating the extent to which Hawaii deep slope bottomfish community dynamics covary with environmental conditions was one of the major recommendations by the Western Pacific Stock Assessment Review (WPSAR) panel.

The purpose of this study is to investigate the extent to which large scale oceanographic and meteorological conditions covary with catch-per-unit-effort (CPUE in lbs/trip) for Hawaii deep slope bottomfish community at main Hawaiian islands. The analysis will also incorporate autocorrelation and adjust the sample size.

## Methods

*Environmental variables*

Large scale oceanographic and meteorological conditions in the Pacific have been related to the southern oscillation index (SOI), the Pacific decadal oscillation (PDO), and sea surface height (SSH). First, the index of the SOI was derived from the Bureau of Meteorology, Australian (http://www.bom.gov.au/climate/current/soihtm1.shtml), which is the monthly fluctuations in the air pressure difference between Tahiti and Darwin. Second, the index of the PDO was obtained from the Joint Institute for the Study of the Atmosphere and Ocean (http://jisao.washington.edu/pdo/PDO.latest). Standardized values for the PDO index were derived as the leading principal component of monthly sea surface temperature (SST) anomalies in the North Pacific Ocean poleward of 20N (Mantua *et al.* 1997). The monthly mean global average SST anomalies are removed to separate this pattern of variability from any large-scale warming signal that may be present in the time-series data. In order to determine regional effects, average sea surface height anomalies (SSH) at Main Hawaii Islands were measured as the difference between the best estimate of the satellite-observed sea surface height and a mean sea surface. Due to the major fishing season for the bottom fish, monthly SOI, PDO and SSH data from January to March were averaged from 1949 to 2010 for SOI and PDO, from 1993 to 2010 for SSH (Table 1).

*Standard correlation analysis*

To examine possible relationships between CPUE and the environmental series, correlations were computed. Assume $N$ pairs of observations on two variables $x$ and $y$. The correlation coefficient between $x$ and $y$ is given by

$$r_{xy} = \frac{\sum_{t=1}^{N}(x_t - \bar{x})(y_t - \bar{y})}{\left[\sum_{t=1}^{N}(x_t - \bar{x})^2\right]^{\frac{1}{2}}\left[\sum_{t=1}^{N}(y_t - \bar{y})^2\right]^{\frac{1}{2}}} \tag{1}$$

*Autocorrelation analysis*

Autocorrelation refers to the correlation of a time series with its own past and future values. Autocorrelation is sometimes called "serial correlation", which refers to the correlation between members of a series of numbers arranged in time. Autocorrelation complicates the application of statistical tests by reducing the effective sample size. Autocorrelation can also complicate the identification of significant covariance or correlation between time series. The first-order autocorrelation coefficient is especially important because for physical and biological systems dependence on past values is likely to be strongest for the most recent past.

A similar idea to the correlation analysis can be applied to time series for which successive observations are correlated. Instead of two different time series, the correlation is computed between one time series and the same series lagged by one unit. The first-order autocorrelation coefficient is the correlation coefficient of the first *N-1* observations, $x_t$, $t$= 1,2,..., *N-1* and the next *N-1* observations, $x_t$ , $t$=2,3,...,$N$. The correlation between $x_t$ and $x_{t+1}$ is given by

$$r_{xx}(1) = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \tag{2}$$

Where $\bar{x} = \sum_{t=1}^{N} x_t$ is the overall mean.

*Cross-correlation analysis*

Cross-correlation incorporates autocorrelation in two different time series in the correlation analyses. Assume $N$ pairs of observations on two variables $x$ and $y$. the cross-correlation is computed between one time series and the other series lagged by one unit. The correlation between $x_{t+1}$ and $y_t$ is given by

$$r_{xy}(1) = \frac{\sum_{t=1}^{N-1}(x_t(1) - \bar{x})(y_t - \bar{y})}{\left[\sum_{t=1}^{N}(x_t(1) - \bar{x})^2\right]^{1/2}\left[\sum_{t=1}^{N}(y_t - \bar{y})^2\right]^{1/2}} \tag{3}$$

*Hypothesis test on* $r_{xy}$, $r_{xx}(1)$, $r_{xy}(1)$

The correlation coefficient, $r_{xy}$, the first-order autocorrelation coefficient, $r_{xx}(1)$, and the first-order cross-correlation coefficient, $r_{xy}(1)$, can be tested against the null hypothesis that the corresponding population value $\rho = 0$ . The critical value of $r_{xy}$, $r_{xx}(1)$ and $r_{xy}(1)$ for a given significance level (e.g., 95%) depends on whether the test is one-tailed or two-tailed. The choice of the alternative hypothesis depends on expected correlation. If there is reason to expect positive correlation, the one-sided test is selected. Otherwise, the two-sided test is selected. In this study, two-sided test was used. Then the alternative hypothesis is that the true $r_{xy}$, $r_{xx}(1)$ and $r_{xy}(1)$ is different from zero, with no specification of whether it is positive or negative:

$$H_1: \rho \neq 0 \tag{4}$$

If the true correlation between two variables $x$ and $y$ within the general population is $\rho$=0, and if the size of the sample, $N$, on which an observed value of $r_{xy}$ is based is equal to or greater than 6, then the student $t$ distribution can be obtained as follows.

$$t = \frac{r_{xy}}{\sqrt{\frac{1-r_{xy}^2}{df}}} \tag{5}$$

is distributed approximately as $t$ with $df$=$N$-2.

The Durbin-Watson ($d$) statistic is a test statistic used to detect the presence of autocorrelation (a relationship between values separated from each other by a given time lag) in the residuals (prediction errors) from a regression analysis. It tests the null hypothesis $H_0$ that the errors are uncorrelated against the alternative hypothesis $H_1$ that the errors are from the first-order autocorrelation model. Thus, if $\rho_e$ are the error autocorrelations, then we have $H_0: \rho_e = 0$ and $H_0: \rho_e \neq 0$. If $e_t$ is the residual associated with the observation at time $t$, then the test statistic is

$$d = \frac{\sum_{t=2}^{N} (e_t - e_{t-1})^2}{\sum_{t=1}^{N} e_t^2} \tag{6}$$

where $N$ is the number of observations. Since $d$ is approximately equal to 2(1-$r_{xy}(1)$), where $r_{xy}(1)$ is the sample autocorrelation of the residuals, $d$=2 indicates no autocorrelation. The value of $d$ always lies between 0 and 4. If the Durbin–Watson statistic is substantially less than 2, there is evidence of positive serial correlation.

*Adjusting the test procedure of a sample correlation*

Garrett and Petrie (1981), Chelton (1984), and Kope and Botsford (1990) provide similar methods for adjusting the null hypothesis test of a sample correlation between two autocorrelated time series, say $X$ and $Y$. These methods assume that the time series are stationary, such that their underlying means remain constant over time.

Furthermore, when testing a sample correlation at lag 1, denoted $r_{xy}(1)$, each of these methods assumes a null hypothesis of no correlation at all lags.

These methods can be summarized using the following theoretical approximation of the effective number of independent observations, $N^*$

$$\frac{1}{N^*} \approx \frac{1}{N} + \frac{2}{N}\frac{(N-1)}{N} r_{xx}(1) r_{yy}(1) \tag{7}$$

where $N$ is the sample size and $r_{xx}(1)$ and $r_{yy}(1)$ are the autocorrelations of $x$ and $y$ at lag 1. For example, Garrett and Petrie (1981) used a form of equation 7 where $r_{xx}(1)$ and $r_{yy}(1)$ was replaced by $r_{xy}$ (1). Given $N^*$, Garrett and Petrie (1981) used the standard critical value for $r_{xy}(1)$ at the α significance level that can be read from statistical tables or derived using the $t$ distribution for two-tailed tests (Zar 1984, p. 309).

**Results and discussion**

The association between CPUE for Hawaii deep slope bottomfish community at main Hawaiian islands and environmental variables were shown in Fig. 1. The hypothesis test indicated that CPUE were significantly negatively correlated with the Pacific decadal oscillation (PDO) ($r_{xy}$= -0.461, P<0.05; Table 2). Table 3 showed the statistical evidence of presence of autocorrelation for each variable. Two variables (CPUE and PDO) were significantly positively autocorrelated to the most recent past observation, implying that both CPUE and PDO were not independent series. A sample correlation between these two autocorrelated time series had fewer degrees of freedom than that assumed under the standard significance test (Table 4). The adjusting significance test showed the statistical significance of first-order cross-correlation between CPUE and PDO (($r_{xy}(1)$= -0.472, P<0.05). The cross-correlation between CPUE for Hawaii deep slope bottomfish community and the next year environmental variables were shown in Fig. 2. This also indicated the negative association between CPUE and 1-year lag PDO.

**References**

Chelton, D.B. 1984. Commentary: short-term climatic variability in the Northeast Pacific Ocean. In The influence of ocean conditions on the production of salmonids in the North Pacific. Edited by W. Pearcy. Oregon State University Press, Corvallis, Oreg. pp. 87–99.

Garrett, C. and Petrie, B. 1981. Dynamical aspects of the flow through the Strait of Belle Isle. J. Phys. Oceanogr. 11: 376–393.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 54: 187–211.

Kope, R.G. and Botsford, L.W. 1990. Determination of factors affecting recruitment of chinook salmon, *Oncorhynchus tshawytscha*, in central California. Fish. Bull. U.S. 88: 257–269.

Mantua, N.J. Hare, S.R., Zhang, Y., Wallace, J.M., and Francis, R.C. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society, 78, pp. 1069-1079.

Myers, R.A., Barrowman, N.J., and Thompson, K.R. 1995. Synchrony of recruitment across the North Atlantic: an update. ICES J. Mar. Sci. 52: 103–110.

Thompson, K.R. and Page, F.H. 1989. Detecting synchrony of recruitment using short, autocorrelated time series. Can. J. Fish. Aquat. Sci. 46: 1831–1838.

Zar , J.H.1984. Biostatistical analysis. Prentice-Hall, Englewood Cliffs, N.J.

Table 1. Annual time series of standardized catch-per-unit-effort (CPUE in lbs/trip) for Hawaii deep slope bottomfish community at main Hawaiian islands, the Southern Oscillation Index (SOI), the Pacific decadal oscillation (PDO), z-scores of sea surface height anomalies at main Hawaiian islands (SSH) for 1993-2010.

| Year | CPUE | SOI | PDO | SSH | Year | CPUE | SOI | PDO | SSH |
|---|---|---|---|---|---|---|---|---|---|
| 1949 | 192.803 | 0.100 | -2.203 | | 1980 | 180.152 | -1.400 | 0.767 | |
| 1950 | 196.854 | 13.433 | -2.057 | | 1981 | 185.155 | -5.700 | 1.013 | |
| 1951 | 207.238 | 8.233 | -1.500 | | 1982 | 167.580 | 4.133 | 0.243 | |
| 1952 | 241.552 | -5.633 | -1.033 | | 1983 | 170.844 | -30.633 | 1.270 | |
| 1953 | 233.366 | -3.200 | -0.587 | | 1984 | 137.469 | 0.433 | 1.493 | |
| 1954 | 284.531 | 0.500 | -1.150 | | 1985 | 151.253 | 0.400 | 0.927 | |
| 1955 | 376.935 | 4.233 | -0.860 | | 1986 | 151.744 | -0.633 | 1.637 | |
| 1956 | 293.728 | 11.033 | -2.593 | | 1987 | 173.126 | -11.833 | 1.910 | |
| 1957 | 333.878 | 0.833 | -0.823 | | 1988 | 199.679 | -1.233 | 1.197 | |
| 1958 | | -8.367 | 0.373 | | 1989 | 192.777 | 9.667 | -0.933 | |
| 1959 | | -4.767 | -0.230 | | 1990 | 172.048 | -8.967 | -0.523 | |
| 1960 | | 1.233 | 0.203 | | 1991 | 158.837 | -1.633 | -1.317 | |
| 1961 | 371.133 | -5.700 | 0.567 | | 1992 | 147.483 | -19.633 | 0.343 | |
| 1962 | 405.117 | 6.967 | -1.287 | | 1993 | 139.077 | -8.200 | 0.333 | 0.417 |
| 1963 | 255.235 | 6.567 | -0.343 | | 1994 | 153.245 | -3.867 | 0.867 | -0.534 |
| 1964 | 278.620 | 1.367 | -0.357 | | 1995 | 156.305 | -1.067 | 0.240 | -0.644 |
| 1965 | 349.054 | 0.167 | -0.787 | | 1996 | 141.086 | 5.233 | 0.783 | 0.846 |
| 1966 | 300.293 | -10.000 | -0.713 | | 1997 | 143.948 | 2.967 | 0.387 | 0.658 |
| 1967 | 285.515 | 11.767 | -0.527 | | 1998 | 135.568 | -23.733 | 1.467 | 0.961 |
| 1968 | 279.756 | 3.567 | -0.553 | | 1999 | 140.889 | 11.033 | -0.437 | -1.055 |
| 1969 | 263.112 | -6.200 | -0.903 | | 2000 | 162.380 | 9.133 | -0.847 | -2.734 |
| 1970 | 243.016 | -6.333 | 0.790 | | 2001 | 154.637 | 9.167 | 0.447 | 0.180 |
| 1971 | 219.975 | 12.533 | -1.773 | | 2002 | 144.013 | 1.733 | -0.267 | 0.403 |
| 1972 | 241.507 | 4.767 | -1.970 | | 2003 | 149.296 | -5.400 | 1.783 | 0.286 |
| 1973 | 226.161 | -5.233 | -0.523 | | 2004 | 139.561 | -0.933 | 0.507 | 1.153 |
| 1974 | 218.617 | 19.100 | -1.257 | | 2005 | 160.345 | -9.033 | 0.870 | 1.097 |
| 1975 | 214.761 | 4.000 | -0.687 | | 2006 | 151.894 | 8.867 | 0.580 | 0.869 |
| 1976 | 231.261 | 12.633 | -1.317 | | 2007 | 159.223 | -3.800 | -0.103 | 0.071 |
| 1977 | 183.192 | -1.933 | 1.160 | | 2008 | 192.167 | 15.867 | -0.827 | -0.033 |
| 1978 | 183.287 | -11.067 | 1.043 | | 2009 | 172.865 | 8.133 | -1.513 | -1.039 |
| 1979 | 209.865 | -0.100 | -0.537 | | 2010 | 146.035 | -11.733 | 0.697 | -0.902 |

Table 2. Correlations, $r_{xy}$, between Hawaii deep slope bottomfish CPUE and environmental variables (SOI = Southern Oscillation Index, PDO = Pacific decadal oscillation, SSH = z-scores of sea surface height anomalies at main Hawaiian islands). Asterisks refer to levels of significance using the student *t* test.

| Index | Environmental variables | $r_{xy}$ | *t* statistic | P-value |
|---|---|---|---|---|
| CPUE | SOI | 0.189 | 1.454 | 0.151 |
| | PDO | -0.461 | -3.925 | 0.000* |
| | SSH | -0.345 | -1.468 | 0.161 |

* p<0.05

Table 3. First-order autocorrelations, $r_{xx}(1)$, for Hawaii deep slope bottomfish CPUE, Southern Oscillation Index (SOI), Pacific decadal oscillation (PDO), and z-scores of sea surface height anomalies at main Hawaiian islands (SSH). Asterisks refer to levels of significance using the Durbin-Watson (*d*) test.

| Variables | $r_{xx}(1)$ | Modified $r_{xx}(1)$ | *d* statistic | P-value |
|---|---|---|---|---|
| CPUE | 0.863 | 0.878 | 0.258 | 0.000* |
| SOI | -0.029 | -0.030 | 2.032 | 0.899 |
| PDO | 0.473 | 0.481 | 0.983 | 0.000* |
| SSH | 0.348 | 0.368 | 1.246 | 0.091 |

* $p<0.05$

Table 4. Comparison of standard test procedure and adjusting test procedure for the first-order cross-correlations, $r_{xy}(1)$, between Hawaii deep slope bottomfish CPUE and environmental variables (SOI = Southern Oscillation Index, PDO = Pacific decadal oscillation, SSH = z-scores of sea surface height anomalies at main Hawaiian islands). Asterisks refer to levels of significance using the student $t$ test. The $N$ and $N^*$ are the number of observations and the effective number of independent observations from equation 7, respectively.

| | | | Standard test procedure | | | Adjusting test procedure | | |
|---|---|---|---|---|---|---|---|---|
| Index | Environmental variables | $r_{xy}(1)$ | $N$ | $t$ statistic | P-value | $N^*$ | $t$ statistic | P-value |
| CPUE | SOI | 0.186 | 59 | 1.429 | 0.159 | 43 | 1.212 | 0.233 |
| | PDO | -0.472 | 59 | 4.041 | 0.000* | 31 | 2.882 | 0.007* |
| | SSH | -0.002 | 18 | 0.007 | 0.994 | 18 | 0.007 | 0.994 |

* $p<0.05$

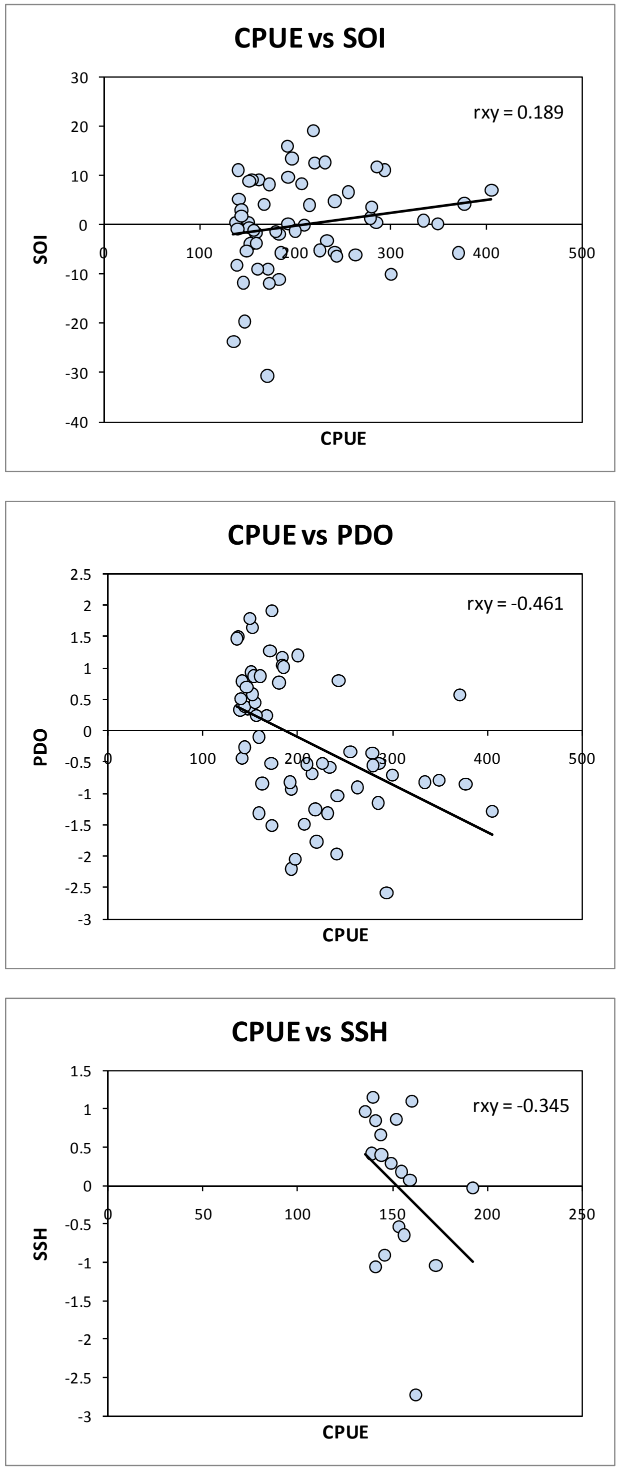


Figure 1. Scatter plots of between Hawaii deep slope bottomfish CPUE and environmental variables (SOI = Southern Oscillation Index, PDO = Pacific decadal oscillation, SSH = z-scores of sea surface height anomalies at main Hawaiian islands) where $r_{xy}$ represent correlation coefficient.

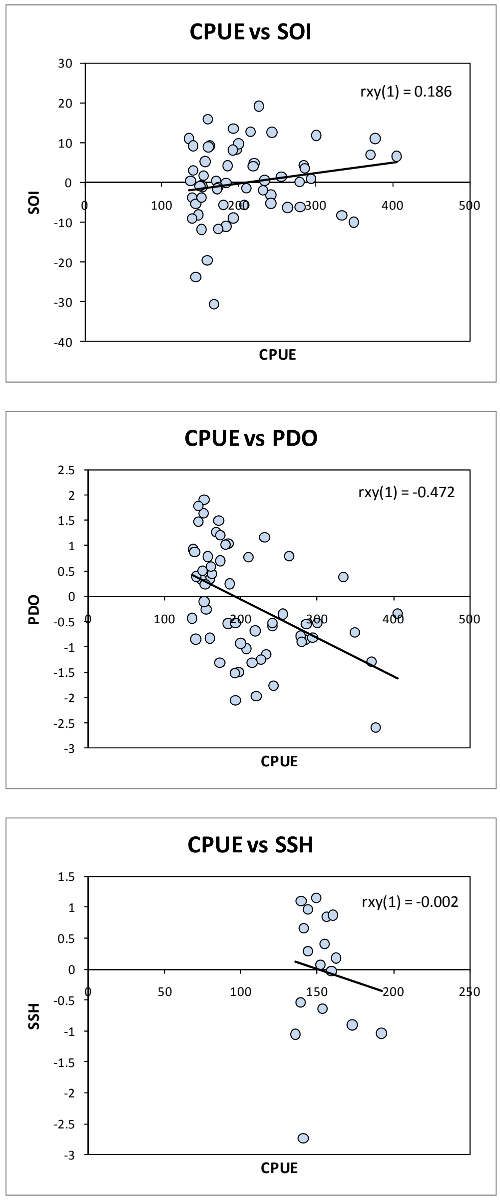


Figure 2. Scatter plots of between Hawaii deep slope bottomfish CPUE and 1-year lag environmental variables (SOI = Southern Oscillation Index, PDO = Pacific decadal oscillation, SSH = z-scores of sea surface height anomalies at main Hawaiian islands) where $r_{xy}(1)$ represent cross-correlation coefficient for the first-order autocorrelation model.