# The Relationship between the Southern Oscillation and Tropical Cyclone Frequency in the Australian Region

ANDREW SOLOW

*Woods Hole Oceanographic Institution, Woods Hole, Massachusetts*

NEVILLE NICHOLLS

*Bureau of Meteorology Research Centre, Melbourne, Victoria*

(Manuscript received 7 September 1989, in final form 28 February 1990)

ABSTRACT

A statistical model of the relationship between tropical cyclone frequency in the Australian region and an index of the strength and phase of the southern oscillation is developed for the period 1910–88. The modeling is nonstandard because the cyclone record is incomplete early in this period. The fitted model indicates that the mean annual number of cyclones during a major cold event is twice that during a major warm event.

## 1. Introduction

In Fig. 1, a plot is shown of an annual record of the number of tropical cyclones observed in the Australian region (5°–32°S, 105°–165°E) for the period 1910–88. The cyclone season in this region runs from October to May, so the observation for 1910 (for example) actually covers the period July 1909 to June 1910. These data have been analyzed by Nicholls (1979, 1984, 1985). Nicholls (1985) attributed the apparent trend in these counts to improvements in observing capability. Solow (1989) considered the problem of reconstructing the true record of cyclone counts using a simple model of the true record and of the way in which the true record has been observed through time.

Nicholls (1979, 1984, 1985) demonstrated that the annual cyclone count in the Australian region is related to El Niño–Southern Oscillation (ENSO) activity. Revell and Goulter (1986a, 1986b) and Dong (1988) have confirmed this relationship. In particular, Revell and Goulter (1986a) demonstrated that tropical cyclone activity around the northeast Australian coast weakens, while activity east of 107°E strengthens, during El Niño events—i.e., when the east equatorial Pacific is anomalously warm. The opposite pattern occurs during La Niña events—i.e., when the east equatorial Pacific is anomalously cool.

The purpose of this study is to construct a statistical model of the relationship between Australian tropical cyclone frequency and ENSO activity using the method of Solow (1989) to account for incomplete observations of cyclone numbers early in the record.

## 2. Approach

Let $Y_t$ be the true number of tropical cyclones in year $t$ ($t = 0, \cdots, n$). A natural statistical model is that ($Y_t$, $t = 0, \cdots, n$) is a sequence of independent Poisson random variables with

$$E(Y_t) = \mathrm{Var}(Y_t) = m_t \quad t = 0, \cdots, n \qquad (1)$$

where $E$ denotes expected value and Var denotes variance. Let $X_t$ be the value of a regressor variable related to ENSO activity in year $t$ ($t = 0, \cdots, n$). The precise definition of $X_t$ is given in the next section. The dependence of $Y_t$ on $X_t$ can be expressed in the following form:

$$m_t = m(X_t; \alpha) \quad t = 0, \cdots, n$$

where $\alpha$ is a vector of parameters. A simple choice of the form of $m(x; \alpha)$ is also given in the next section.

In year $t$, each cyclone is observed with probability $p_t$. Let $Z_t$ be the recorded number of cyclones in year $t$. It is straightforward to show that ($Z_t$, $t = 0, \cdots, n$) is also a sequence of independent Poisson random variables with:

$$E(Z_t) = \mathrm{Var}(Z_t) = p_t m_t \quad t = 0, \cdots, n. \qquad (2)$$

Assume that the sequence ($p_t$, $t = 0, \cdots, n$) is nondecreasing and that it reaches 1 at $t = n_0 (0 < n_0 < n)$. Assume that $n_0$ is known. A flexible, parsimonious family of such probability sequence is:

$$p_t = \begin{cases} p_0 + (1 - p_0)F(t/n_0; a, b) & t = 0, \cdots, n_0 - 1 \\ 1 & t = n_0, \cdots, n. \end{cases}$$

*Corresponding author address:* Dr. Andrew R. Solow, Woods Hole Oceanographic Institution, Woods Hole, MA 02543.

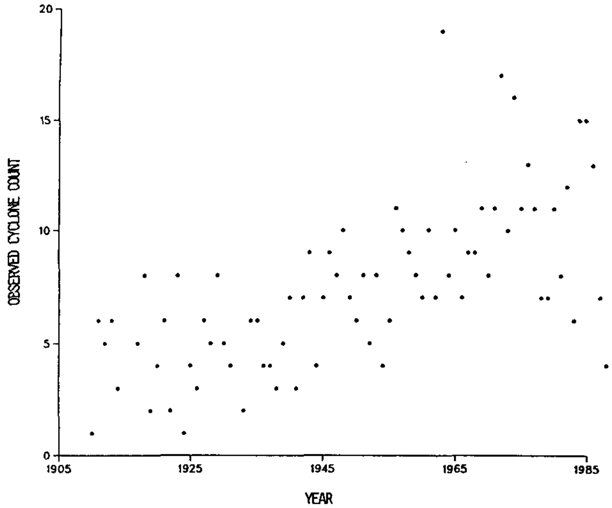


FIG. 1. Annual record of observed tropical cyclone counts ($Z_t$) for the Australian region, 1910–88.

Here $p_0$ is the observation probability in year 0 and $F(t; a, b)$ is the cumulative distribution function for a random variable with a beta distribution with parameters $a$ and $b$:

$$F(t; a, b) = (\Gamma(a+b)/\Gamma(a)\Gamma(b)) \int_0^t u^{a-1}(1-u)^{b-1}du$$

where $\Gamma$ is the gamma function (Epstein 1985). This integral can be evaluated numerically using the subroutine BETDF contained in the IMSL (1987) library of FORTRAN subroutines. The choice of this form for the probability sequence is based entirely on its flexibility and parsimony. This is illustrated in Fig. 2, where $F(t; a, b)$ is plotted against $t$ for selected values of $a$ and $b$.

Let $z_t$ be the observed value of $Z_t$. Under this model, the probability that $Z_t = z_t$ is given by the Poisson probability:

$$pr(Z_t = z_t) = \exp(-p_t m_t)(p_t m_t) z_t / z_t! \quad (3)$$

This probability depends on the unknown vector of parameters $\beta = (p_0\ a\ b\ \alpha)$. Because the observed counts are assumed to be independent, the joint probability that $Z_0 = z_0, \cdots, Z_n = z_n$ is given by the product over $t$ of these probabilities. Regarded as a function of $\beta$, this product is called the likelihood function. The maximum likelihood estimates of $\beta$—which is denoted by $\beta^*$—can be found by maximizing the likelihood (or its logarithm) over $\beta$ (Silvey 1970). This maximization must be done numerically.

## 3. Application

The first step in applying the approach outlined in the previous section to the cyclone counts in Fig. 1 is the specification of $n_0$, which is chosen to be 1965. This corresponds more or less to the advent of comprehensive meteorological satellite coverage (Nicholls 1985). It is possible to extend the maximum likelihood approach to the case where $n_0$ is treated as an unknown parameter, although this increases the amount of computation.

The second step in this application is to define a suitable measure of ENSO activity. A common index of ENSO activity is the Southern Oscillation Index (SOI), which is defined as the atmospheric pressure differential at sea level between Darwin and Tahiti. The version of this index used in this paper has been standardized to have a mean of 0 and a standard deviation of 10. Large positive values of this index correspond to the cold phase of the Southern Oscillation and large negative values correspond to the warm phase. For this paper, $X_t$ was taken to be the average value of the index for the month of September preceding the onset of the Australian cyclone season in October. For the period 1965–88, this month has the maximal correlation with cyclone count (0.68) of any month. The values of $X_t$ are shown in Fig. 3. Values were missing for the years 1915–16 and 1932; therefore these years were omitted from the analysis.

The third step in this application is the specification of the form of the dependence of $Y_t$ on $X_t$. In Fig. 4, $Y_t$ is plotted against $X_t$ for the period 1965–88. From this figure, it seems reasonable to assume that $Y_t$ and $X_t$ are approximately linearly related. In fact, from a statistical point of view, it is more appropriate to assume that:

$$m_t = \exp(\alpha_0 + \alpha_1 X_t) \quad t = 0, \cdots, n \quad (4)$$

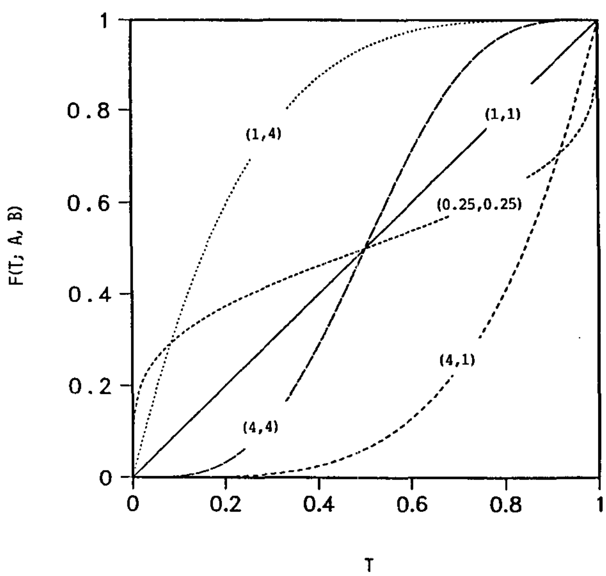


FIG. 2. The incomplete beta ratio $F(t; a, b)$ v $t$ for selected values of $a$ and $b$.

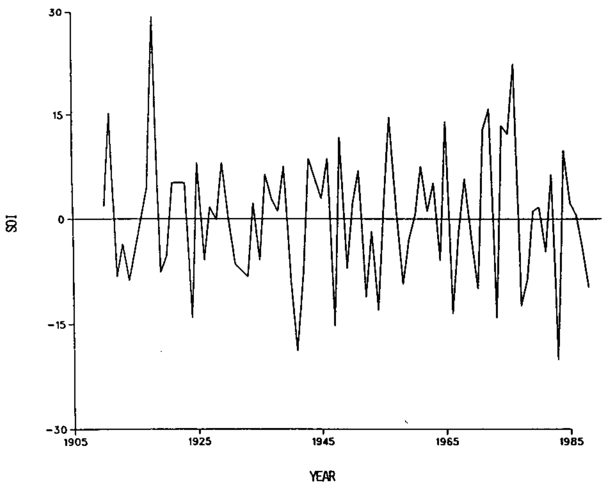


FIG. 3. September southern oscillation index ($X_t$), 1910–88.

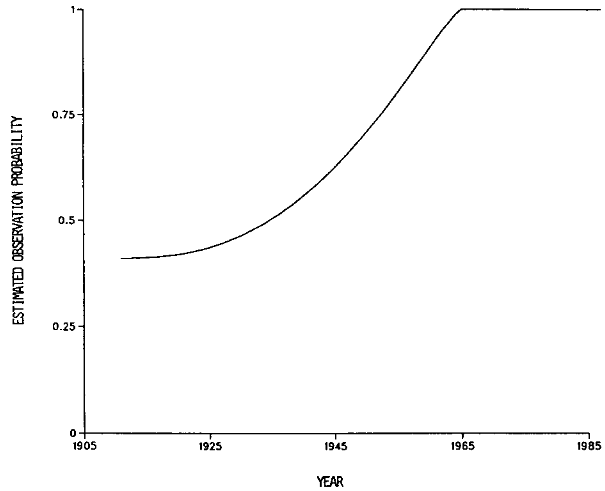


FIG. 5. Estimated observation probabilities ($p_t^*$).

as this ensures that $m_t$ is positive. In any case, the exponential linear model is close to the linear model for small values of $\alpha_1$.

The maximum likelihood estimates of the model parameters were found to be $p_0^* = 0.41$, $a^* = 2.7$, $b^* = 1.2$, $\alpha_0^* = 2.30$ and $\alpha_1^* = 0.021$. The estimated observation probability, which is shown in Fig. 5, increases slowly from 0.41 in 1910 to 0.50 in 1935, and then more rapidly, reaching 0.80 in 1956.

Approximate confidence intervals for the elements of $\beta$ can be found by exploiting the connection between confidence intervals and hypothesis tests. For example, a 0.95 confidence interval for $\alpha_1$ contains the value $A_1$ provided that the null hypothesis $H_0$: $\alpha_1 = A_1$ cannot be rejected at the 0.05 significance level. The set of all such $A_1$'s constitutes a 0.95 confidence interval for $\alpha_1$. This null hypothesis can be tested using the generalized likelihood ratio test (Silvey 1970). The test statistic is twice the difference of the maximized log-likelihood and the maximized log-likelihood with $\alpha_1$ fixed at $A_1$. Under $H_0$, this statistic has an approximate chi-squared distribution with one degree of freedom. Approximate 0.95 confidence intervals for $\alpha_0$ and $\alpha_1$ found in this way are (2.23, 2.39) and (0.013, 0.029), respectively.

In Fig. 6 the estimated values of $m_t$ are plotted against $t$. This figure indicates that years of relatively high cyclone activity occurred in 1911, 1918, 1948 and 1956, and that years of relatively low activity occurred in 1924, 1941, 1947, 1952 and 1954. It is interesting to note that, although the estimated mean function exhibits no secular trend, there appears to be a slight increase in variability with time. This is due to an apparent increase in the variability of $X_t$ with time that is discernible in Fig. 3.

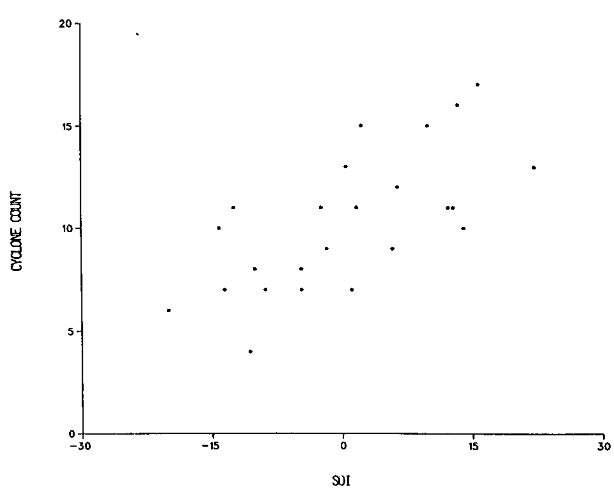


FIG. 4. Annual cyclone count ($Y_t$) v southern oscillation index ($X_t$) for the period 1965–88 during which the observed counts are assumed to be complete.

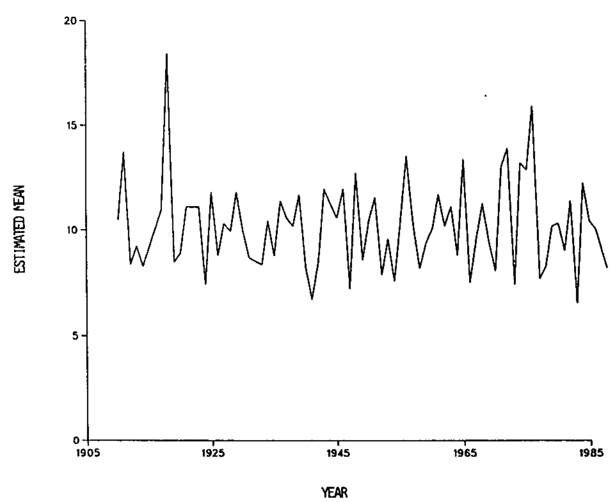


FIG. 6. Estimated mean function of true cyclone counts ($m_t^*$).

## 4. Diagnostics

In this section some diagnostics based on residuals are presented. The aim of these diagnostics is to validate the model under which the analysis of the previous section was carried out. The model for the observed counts ($Z_t, t = 0, \cdots, n$) consists of a systematic component and a random component. The systematic component refers to the estimated mean function ($p_t^* m_t^*, t = 0, \cdots, n$). A plot of the estimated mean function is given in Fig. 7 along with the actual observed counts. The random component refers to the assumption that the observed counts are independent Poisson variates.

Goodness-of-fit measures for models of counts are discussed by Pregibon (1979) and McCullagh and Nelder (1983). The basis of these measures is the deviance residual, which is a generalization of the ordinary residual to situations involving nonnormal data. For a Poisson observation $z$ and its estimate $z^*$, the deviance residual is given by:

$$d = \text{sgn}(z - z^*)(2(z \log(z/z^*) - (z - z^*)))^{1/2}. \tag{6}$$

In Fig. 8, the deviance residuals ($d_t, t = 0, \cdots, n$) are plotted against the fitted values ($p_t^* m_t^*, t = 0, \cdots, n$). This plot is useful for detecting inadequacies in the systematic component of the model, which appear as trend or curvature in this plot. No such structure appears in Fig. 8. This plot is also useful for detecting departures from the assumed mean-variance relationship, which appear as a trend in the dispersion of the points. There appears to be a slight tendency for dispersion to be lower for higher fitted values. It is important to note however, that even if the assumed mean-variance relationship is correct, intervals on the abscissa of this plot with high density of points will tend to produce a wider range of residuals than intervals with low density. In Fig. 8 the density of points decreases as the fitted value increases, and this may explain any apparent decrease in dispersion.

The deviance residuals are also useful for detecting departures from the random component of the model. Provided that the model assumptions are approximately correct and that the systematic component of the model is adequate, the deviance residuals are approximately independent unit normal. For the fitted model in Fig. 7, the deviance residuals exhibit no important serial correlation—e.g., the estimated correlation at a lag of one year is −0.04. To check for unit normality, the normal probability plot of the deviance residuals is shown in Fig. 9. This is a plot of the ordered

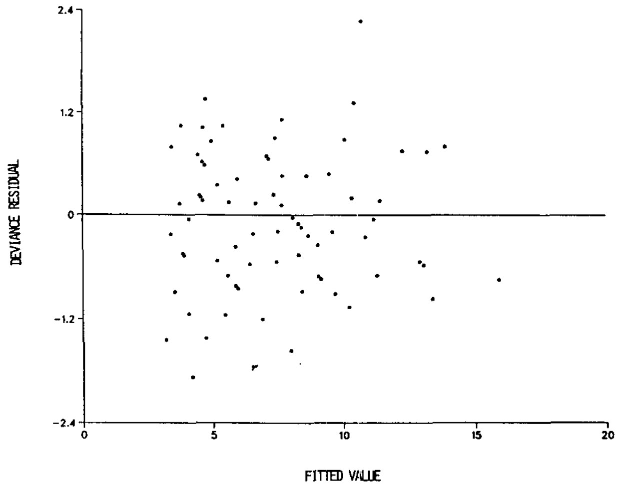


FIG. 8. Deviance residuals ($d_t$) v fitted values ($p_t^* m_t^*$).

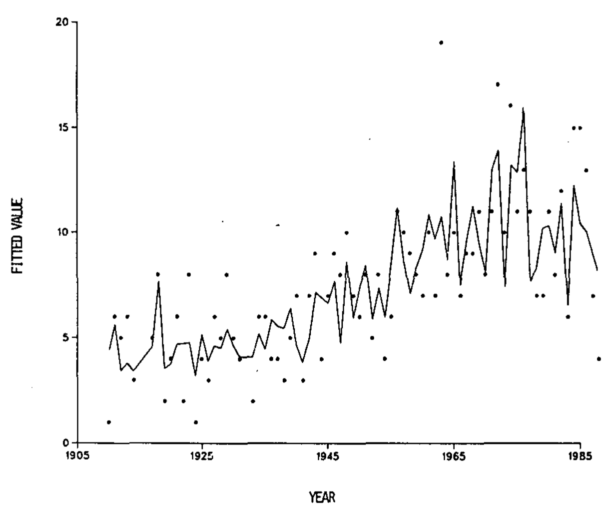


FIG. 7. Fitted values of observed cyclone counts ($p_t^* m_t^*$).

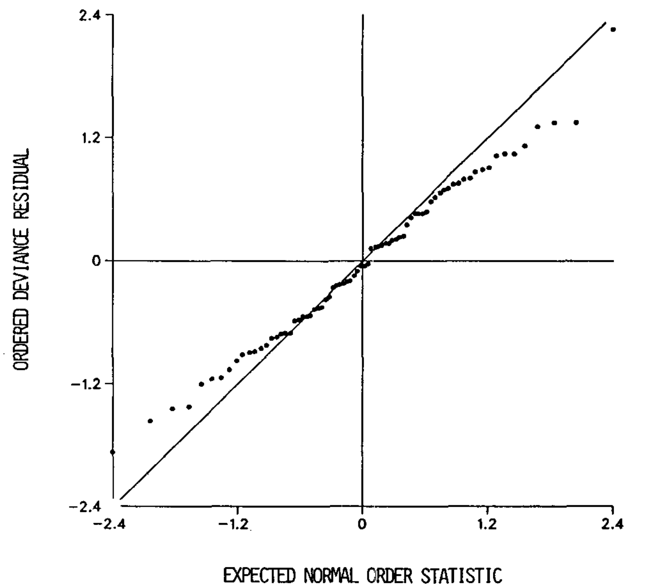


FIG. 9. Normal probability plot of deviance residuals ($d_t$).

deviance residuals against the corresponding expected order statistics for the unit normal distribution (David 1981). This plot shows no important departures from linearity. The sample standard deviation of the deviance residuals is 0.83. This underdispersion, which is not very large, indicates that the cyclone counts are slightly less variable than expected under the Poisson model.

## 5. Discussion

The purpose of this paper has been to construct a statistical model relating the annual number of tropical cyclones in the Australian region to an index of ENSO activity. The approach to fitting this model is nonstandard because the record of cyclone counts is incomplete, particularly early in the record. The fitted model appears to perform well at describing the observed record.

The approach described in this paper has two advantages over the more traditional two-step approach under which a time trend is removed from the observed counts and the residuals are regressed against the SOI. The first advantage is formal. The approach described here is based on maximum likelihood estimation and shares the optimality properties and inferential machinery of that method. The two-step approach is clearly an approximation to the full maximum likelihood approach. The second advantage is practical. The effects of incomplete observation and ENSO activity operate jointly on the observed cyclone counts. The approach described here takes this into account, but the two-step approach does not. The first step in the two-step approach ignores the effect of ENSO activity in estimating the trend in observed counts. The trend that is estimated will almost certainly contain effects due to variations in ENSO activity and this will degrade the results of the second step in which the residuals from the estimated trend are regressed against the SOI.

The fitted model:

$$m_t^* = \exp(2.30 + 0.021X_t) \qquad (7)$$

can be used to predict $Y_t$ from $X_t$. For example, if $X_t = 20$ (which corresponds to a major cold event), the predicted value of $Y_t$ would be around 15 cyclones. If $X_t = -20$ (which corresponds to a major warm event), the predicted value of $Y_t$ would be around seven cyclones. To give an overall measure of the forecasting ability of the model, the following cross-validation exercise was performed. Each count for the period 1965–88 was omitted in turn, the model was re-fit using the remaining data, and the fitted value of the model for the omitted year was taken as the estimate of the omitted count. Cross-validation ensures that for each year the estimated count is independent of the true count. The mean squared error of predicting $Y_t$ in this way was 6.5. The corresponding figure using the overall cross-validated mean count was 11.7. This corresponds to a cross-validated $R^2$ of 0.44 although nonconstant variance of the true counts invalidates the usual interpretation in terms of explained variance.

The diagnostics presented in the previous section indicate a small number of anomalous years. For example, the observed count of one cyclone in 1910 appears to be rather low. Nicholls (1985) suggested that the low count in 1910 may be due to "start-up" problems in the recording system. Similarly, the observed count of 19 cyclones in 1963 appears to be rather high. This count has also been questioned on other grounds (Nicholls 1985). For example, the record shows that three cyclones were reported on New Year's Day in 1963, all within 150 km of each other. As the diameter of a tropical cyclone is typically around 200 km, this seems unlikely.

*Acknowledgments.* The authors thank the Editor and two anonymous reviewers for their helpful comments. This work was supported by NOAA, National Sea Grant College Program Office, Department of Commerce under Grant No. NA8-AA-D-SG090, WHOI Sea Grant Project No. R/0-4, The Pew Charitable Trusts, and the WHOI Marine Policy Center. This is Contribution 7291 of the Woods Hole Oceanographic Institution.

## REFERENCES

David, H. A., 1981: *Order Statistics.* Wiley, 360 pp.

Dong, K., 1988: El Niño and tropical cyclone frequency in the Australian region and the northwest Pacific. *Aust. Meteor. Mag.*, **36**, 219–225.

Epstein, E. S., 1985: *Statistical Inference and Prediction in Climatology: A Bayesian Approach.* Amer. Meteor. Soc., 199 pp.

IMSL, 1987: *Stat/Library: FORTRAN Subroutines for Statistical Analysis.* IMSL, Inc., 1300 pp.

McCullagh, P., and J. A. Nelder, 1983: *Generalized Linear Models.* Chapman and Hall, 280 pp.

Nicholls, N., 1979: A possible method for predicting seasonal tropical cyclone activity in the Australian region. *Mon. Wea. Rev.*, **107**, 1221–1224.

——, 1984: The southern oscillation, sea-surface temperature, and interannual fluctuations in Australian tropical cyclone activity. *J. Climatol.*, **4**, 661–670.

——, 1985: Predictability of interannual variations in Australian seasonal tropical cyclone activity. *Mon. Wea. Rev.*, **113**, 1144–1149.

Pregibon, D., 1979: Data analytic methods for generalized linear models. Ph.D. thesis, Department of Statistics, University of Toronto.

Revell, C. G., and S. W. Goulter, 1986a: South Pacific tropical cyclones and the southern oscillation. *Mon. Wea. Rev.*, **114**, 1138–1145.

——, and ——, 1986b: Lagged relations between the southern oscillation and numbers of tropical cyclones in the South Pacific region. *Mon. Wea. Rev.*, **114**, 2669–2670.

Silvey, S. D., 1970: *Statistical Inference.* Chapman and Hall, 191 pp.

Solow, A. R. 1989: Reconstructing a partially observed record of tropical cyclone counts. *J. Climate*, **2**, 1253–1257.