APPENDIX A

Statistical Trend Analysis

Coordinating Lead Author: Richard L. Smith, Univ. N.C., Chapel Hill

In many places in this report, but especially in Chapter 2, trends have been calculated, either based directly on some climatic variable of interest (*e.g.*, hurricane or cyclone counts) or from some index of extreme climate events. Statistical methods are used in determining the form of a trend, estimating the trend itself along with some measure of uncertainty (*e.g.*, a standard error), and in determining the statistical significance of a trend. A broad-based introduction to these concepts has been given by Wigley (2006). The present review extends Wigley's by introducing some of the more advanced statistical methods that involve time series analysis.

Some initial comments are appropriate about the purpose, and also the limitations, of statistical trend estimation. Real data rarely conform exactly to any statistical model, such as a normal distribution. Where there are trends, they may take many forms. For example, a trend may appear to follow a quadratic or exponential curve rather than a straight line, or it may appear to be superimposed on some cyclic behavior, or there may be sudden jumps (also called changepoints) as well or instead of a steadily increasing or decreasing trend. In these cases, assuming a simple linear trend (equation (1) below) may be misleading. However, the slope of a linear trend can still represent the most compact and convenient method of describing the overall change in some data over a given period of time.

In this appendix, we first outline some of the modern methods of trend estimation that involve estimating a linear or nonlinear trend in a correlated time series. Then, the methods are illustrated on a number of examples related to climate and weather extremes.

The basic statistical model for a linear trend can be represented by the equation

(1) $y_t = b_0 + b_1 t + u_t$

where t represents the year, y_t is the data value of interest (*e.g.*, temperature or some climate index in year t), b_0 and b_1 are the intercept and slope of the linear regression, and u_t represents a random error component. The simplest case is when u_t are uncorrelated error terms with mean 0 and a common variance, in which case we typically apply the standard ordinary least squares (OLS) formulas to estimate the intercept and slope, together with their standard errors. Usually the slope (b_1) is interpreted as a trend, so this is the primary quantity of interest.

The principal complication with this analysis in the case of climate data is usually that the data are autocorrelated; in other words, the terms cannot be taken as independent. This brings us within the field of statistics known as time series analysis, see *e.g.*, the book by Brockwell and Davis (2002). One common way to deal with this is to assume the values form an autoregressive, moving average process (ARMA for short). The standard ARMA(p,q) process is of the form

(2) $u_t - \varphi_1 u_{t-1} - \ldots - \varphi_p u_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$

where $\varphi_1 \ldots \varphi_p$ are the autoregressive coefficients, $\theta_1 \ldots \theta_q$ are the moving average coefficients, and the ε_t terms are independent with mean 0 and common variance. The orders p and q are sometimes determined empirically, or sometimes through more formal model-determination techniques such as the Akaike Information Criterion (AIC) or the Bias-Corrected Akaike Information Criterion (AICC). The autoregressive and moving average coefficients may be determined by one of several estimation algorithms (including maximum likelihood), and the regression coefficients b_0 and b_1 by the algorithm of generalized least squares (GLS). Typically, the GLS estimates are not very different from the OLS estimates that arise when autocorrelation is ignored, but the standard errors can be very different. It is quite common that a trend that appears to be statistically significant when estimated under OLS regression is not statistically significant under GLS regression, because of the larger standard error that is usually though not invariably associated with

GLS. This is the main reason why it is important to take autocorrelation into account.

An alternative model which is an extension of (1) is

(3) $y_t = b_0 + b_1 x_{t1} + \ldots + b_k x_{tk} + u_t$

where $x_{t1} \ldots x_{tk}$ are k regression variables (covariates) and $b_1 \ldots b_k$ are the associated coefficients. A simple example is polynomial regression, where $x_{tj}=t^j$ for j=1,…,k. However, a polynomial trend, when used to represent a nonlinear trend in a climatic dataset, often has the disadvantage that it behaves unstably at the endpoints, so alternative representations such as cubic splines are usually preferred. These can also be represented in the form of (3) with suitable $x_{t1} \ldots x_{tk}$. As with (1), the u_t terms can be taken as uncorrelated with mean 0 and common variance, in which case OLS regression is again appropriate, but it is also common to consider the u_t as autocorrelated.

There are, by now, several algorithms available that fit these models in a semiautomatic fashion. The book by Davis and Brockwell (2002) includes a CD containing a time series program, ITSM, that among many other features, will fit a model of the form (1) or (3) in which the u_t terms follow an ARMA model as in (2). The orders p and q may be specified by the user or selected automatically via AICC. Alternatively, the statistical language R (R Development Core Team, 2007) contains a function "arima" which allows for fitting these models by exact maximum likelihood. The inputs to the arima function include the time series, the covariates, and the orders p and q. The program calculates maximum likelihood/GLS estimates of the ARMA and regression parameters, together with their standard errors, and various other statistics including AIC. Although R does not contain an automated model selection procedure, it is straightforward to write a short subroutine that fits the time series model for various values of p and q (for example, all values of p and q between 0 and 10), and then identifies the model with minimum AIC. This method has been routinely used for several of the following analyses.

However, it is not always necessary to search through a large set of ARMA models. In very many cases, the AR(1) model in which p=1, q=0, captures almost all of the autocorrelation, in which case this would be the preferred approach.

In other cases, it may be found that there is cyclic behavior in the data corresponding to large-scale circulation indexes such as the Southern Oscillation Index (SOI – often taken as an indicator of El Niño) or the Atlantic Multidecadal Oscillation (AMO) or the Pacific Decadal Oscillation (PDO). In such cases, an alternative to searching for a high-order ARMA model may be to include SOI, AMO or PDO directly as one of the covariates in (2).

Two other practical features should be noted before we discuss specific examples. First, the methodology we have discussed assumes the observations are normally distributed with constant variances (homoscedastic). Sometimes it is necessary to make some transformation to improve the fit of these assumptions. Common transformations include taking logarithms or square roots. With data in the form of counts (such as hurricanes), a square root transformation is often made because count data are frequently represented by a Poisson distribution, and for that distribution, a square root transformation is a so-called variance-stabilizing transformation, making the data approximately homoscedastic.

The other practical feature that occurs quite frequently is that the same linear trend may not be apparent through all parts of the data. In that case, it is tempting to select the start and finish points of the time series and recalculate the trend just for that portion of the series. There is a danger in doing this, because in formally testing for the presence of a trend, the calculation of significance levels typically does not allow for the selection of a start and finish point. Thus, the procedure may end up selecting a spurious trend. On the other hand, it is sometimes possible to correct for this effect, for example, by using a Bonferroni correction procedure. An example of this is given in our analysis of the heatwave index dataset below.

EXAMPLE 1: COLD INDEX DATA (SECTION 2.2.1)

The data consist of the "cold index," 1895-2005. A density plot of the data shows that the original data are highly right-skewed, but a cube-root transformation leads to a much more symmetric distribution (Figure A.1).

We therefore proceed to look for trends in the cube root data.

A simple OLS linear regression yields a trend of -.00125 per year, standard error .00068, for which the 2-sided p-value is .067. Recomputing using the minimum-AIC ARMA model yields the optimal values p=q=3, trend -.00118, standard error .00064, p-value .066. In this case, fitting an ARMA model makes very little difference to the result compared with OLS. By the usual criterion of a .05 significance level, this is not a statistically significant result, but it is close enough that we are justified in concluding there is still some evidence of a downward linear trend. Figure A.2 illustrates the fitted linear trend on the cube root data.

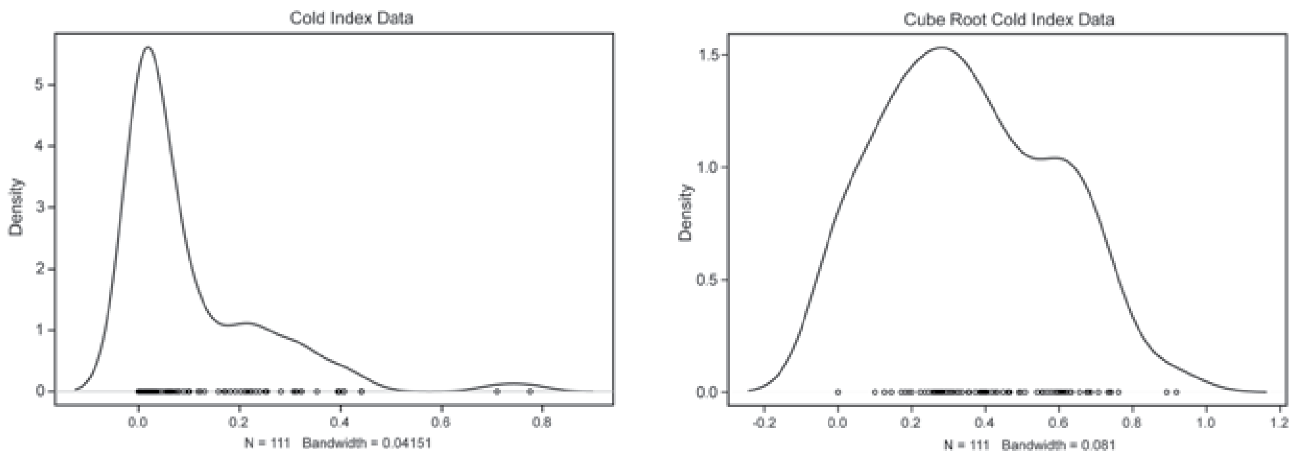

Figure A.1 Density plot for the cold index data (left), and for the cube roots of the same data (right).

EXAMPLE 2: HEAT WAVE INDEX DATA (SECTION 2.2.1 AND FIG. 2.3(A))

This example is more complicated to analyze because of the presence of several outlying values in the 1930s which frustrate any attempt to fit a linear trend to the whole series. However, a density plot of the raw data show that they are very right-skewed, whereas taking natural logarithms makes the data look much more normal (Figure A.3). Therefore, for the rest of this analysis we work with the natural logarithms of the heat wave index.

In this case, there is no obvious evidence of a linear trend either upwards or downwards. However, nonlinear trend fits suggest an oscillating pattern up to about 1960, followed by a steadier upward drift in the last four decades. For example, the solid curve in Figure A.4, which is based on a cubic spline fit with 8 degrees of freedom, fitted by ordinary linear regression, is of this form.

Motivated by this, a linear trend has been fitted by time series regression to the data from 1960-2005 (dashed straight line, Figure A.4). In this case, searching for the best ARMA model by the AIC criterion led to the ARMA(1,1) model being selected. Under this model, the fitted linear trend has a slope of 0.031 per year and a standard error of .0035. This is very highly statistically significant. Assuming normally distributed errors, the probability that such a result could have been reached by chance, if there were no trend, is of the order 10^{-18}.

We should comment a little about the justification for choosing the endpoints of the linear trend (in this case, 1960 and 2005) in order to give the best fit to a straight line. The potential objection to this is that it creates a bias associated with multiple testing. Suppose, as an artificial example, we were to conduct 100 hypothesis tests based on some sample of data, with significance level .05. This means that if there were in fact no trend present at all, each of the tests would have a .05 probability of incorrectly concluding that there was a trend. In 100 such tests, we would typically expect about 5 of the tests to lead to the conclusion that there was a trend.

A standard way to deal with this issue is the Bonferroni correction. Suppose we still conducted 100 tests, but adjusted the significance level of each test to .05/100=.0005.
Then even if no trend were present, the probability that at least one of the tests led to rejecting the null hypothesis would be no more than 100 times .0005, or .05. In other words, with the Bonferroni correction, .05 is still an upper bound on the overall probability that one of the tests falsely rejects the null hypothesis.

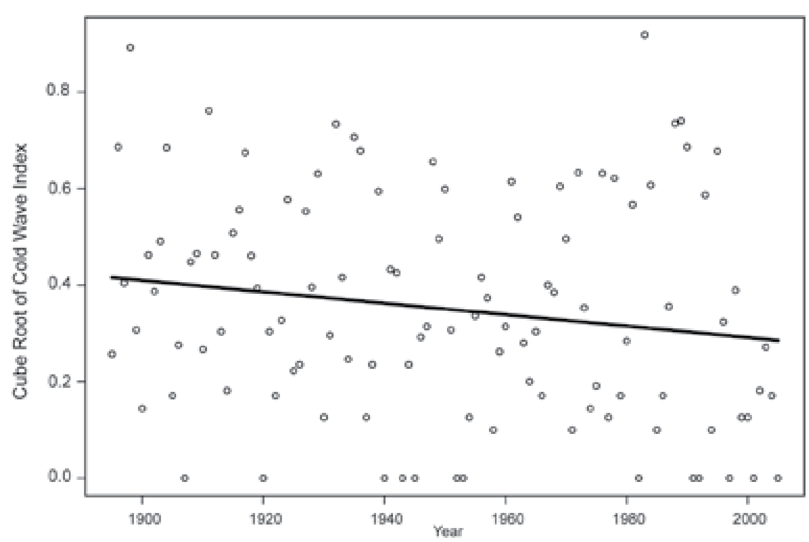

Figure A.2 Cube root of cold wave index with fitted linear trend.

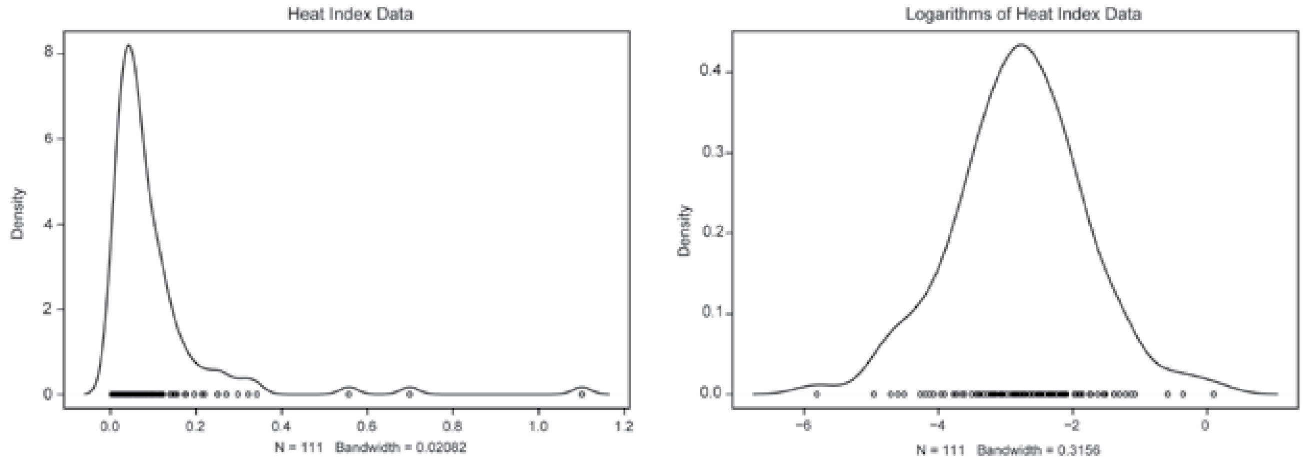

Figure A.3 Density plot for the heat index data (left), and for the natural logarithms of the same data (right).

In the case under discussion, if we allow for all possible combinations of start and finish dates, given a 111-year series, that makes for 111x110/2=6105 tests. To apply the Bonferroni correction in this case, we should therefore adjust the significance level of the individual tests to .05/6105=.0000082. However, this is still very much larger than 10^{-18}. The conclusion is that the statistically significant result cannot be explained away as merely the result of selecting the endpoints of the trend.

This application of the Bonferroni correction is somewhat unusual. It is rare for a trend to be so highly significant that selection effects can be explained away completely, as has been shown here. Usually, we have to make a somewhat more subjective judgment about what are suitable starting and finishing points of the analysis.

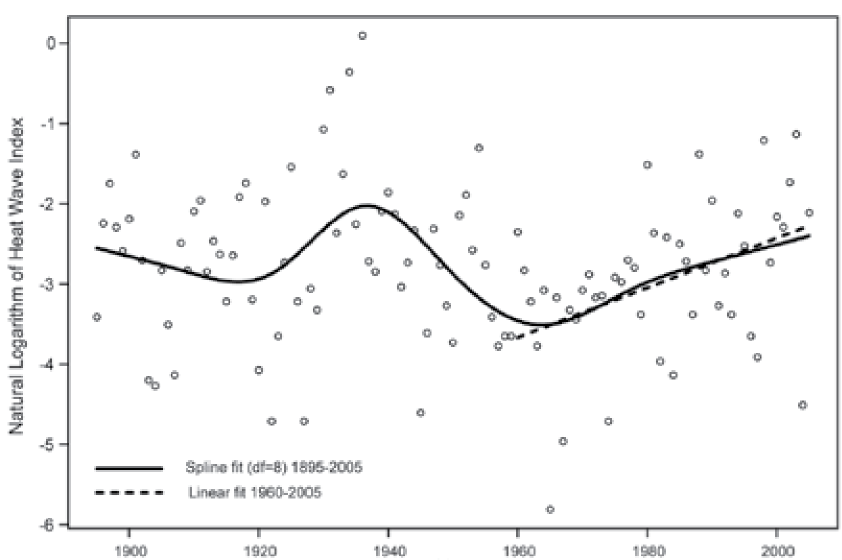

Figure A.4 Trends fitted to natural logarithms of heat index. Solid curve: nonlinear spline with 8 degrees of freedom fitted to the whole series. Dashed line: linear trend fitted to data from 1960-2005.

EXAMPLE 3: 1-DAY HEAVY PRECIPITATION FREQUENCIES (SECTION 2.1.2.2)

In this example, we considered the time series of 1-day heavy precipitation frequencies for a 20-year return value. In this case, the density plot for the raw data is not as badly skewed as in the earlier examples (Figure A.5, left plot), but is still improved by taking square roots (Figure A.5, right plot). Therefore, we take square roots in the subsequent analysis.

Looking for linear trends in the whole series from 1895-2005, the overall trend is positive but not statistically significant (Figure A.6). Based on simple linear regression, the estimated slope is .00023 with a standard error of .00012, which just fails to be significant at the 5% level. However, time series analysis identifies an ARMA (5, 3) model, when the estimated slope is still .00023, the standard error rises to .00014, which is again not statistically significant.

However, a similar exploratory analysis to that in Example 2 suggested that a better linear trend could be obtained starting around 1935. To be specific, we have considered the data from 1934-2005. Over this period, time series analysis identifies an ARMA(1,2) model, for which the estimated slope is .00067, standard error .00007, under which a formal test rejects the null hypothesis of no slope with a significance level of the order of 10^{-20} under normal theory assumptions. As with Example 2, an argument based on the Bonferroni correction shows that this is a clearly significant result

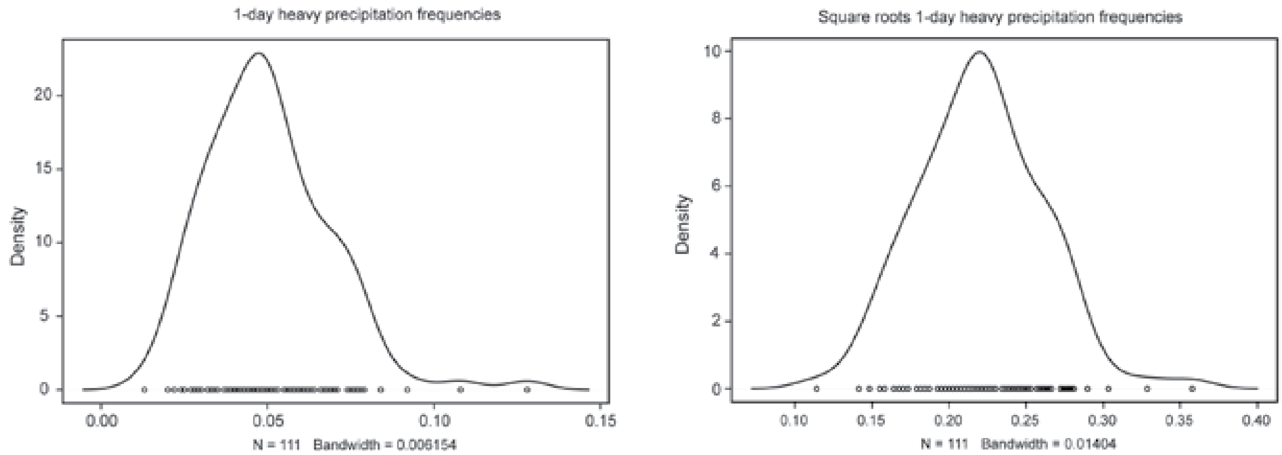

Figure A.5 Density plot for 1-day heavy precipitation frequencies for a 20-year return value (left), and for square roots of the same data (right).

even allowing for the subjective selection of start and finish points of the trend.

Therefore, our conclusion in this case is that there is an overall positive but not statistically significant trend over the whole series, but the trend post-1934 is much steeper and clearly significant.

EXAMPLE 4: 90-DAY HEAVY PRECIPITATION FREQUENCIES (SECTION 2.1.2.3 AND FIG. 2.9)

This is a similar example based on the time series of 90-day heavy precipitation frequencies for a 20-year return value. Once again, density plots suggest a square root transformation (the plots look rather similar to Figure A.5 and are not shown here).

After taking square roots, simple linear regression leads to an estimated slope of .00044, standard error .00019, based on the whole data set. Fitting ARMA models with linear trend leads us to identify the ARMA(3,1) as the best model under AIC: in that case, the estimated slope becomes .00046 and the standard error actually goes down, to .00009. Therefore, we conclude that the linear trend is highly significant in this case (Figure A.7).

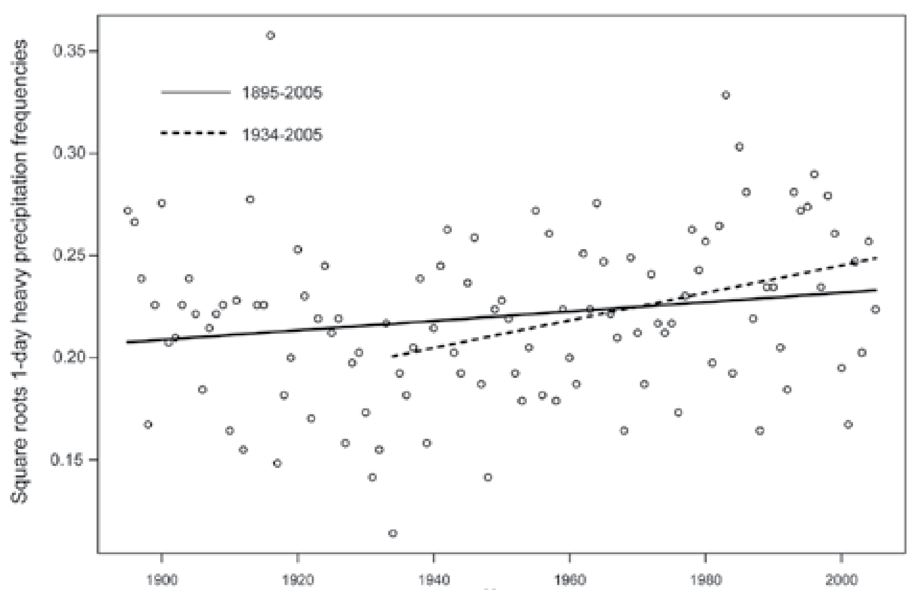

Figure A.6 Trend analysis for the square roots of 1-day heavy precipitation frequencies for a 20-year return value, showing estimated linear trends over 1895-2005 and 1934-2005.

EXAMPLE 5: TROPICAL CYCLONES IN THE NORTH ATLANTIC (SECTION 2.1.3.1)

This analysis is based on historical reconstructions of tropical cyclone counts described in the recent paper of Vecchi and Knutson (2008). We consider two slightly different reconstructions of the data: the "one-encounter" reconstruction in which only one intersection of a ship and storm is required for a storm to be counted as seen, and the "two-encounter" reconstruction that requires two intersections before a storm is counted. We focus particularly on the contrast between trends over the 1878-2005 and 1900-2005 time periods, since before the start of the present analysis, Vecchi and Knutson had identified these two periods as of particular interest.

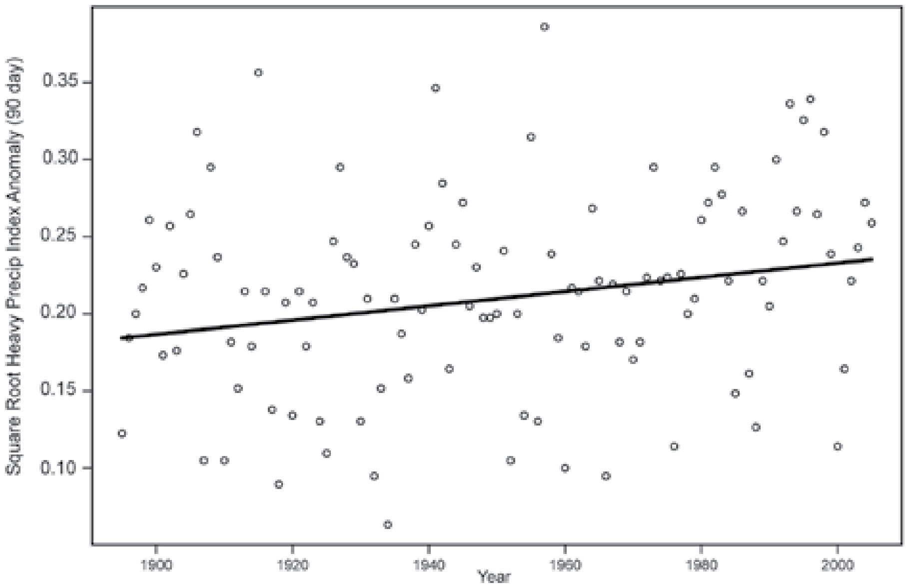

Figure A.7 Trend analysis for the square roots of 90-day heavy precipitation frequencies.

For 1878-2005, using the one-encounter dataset, we find by ordinary least squares a linear trend of .017 (storms per year), standard error .009, which is not statistically significant. Selecting a time series model by AIC, we identify an ARMA(9,2) model as best (an unusually large order of a time series model in this kind of analysis), which leads to a linear trend estimate of .022, standard error .022, which is clearly not significant.

When the same analysis is repeated from 1900-2005, we find by linear regression a slope of .047, standard error .012, which is significant. Time series analysis now identifies the ARMA(5,3) model as optimal, with a slope of .048, standard error .015–very clearly significant. Thus, the evidence is that there is a statistically significant trend over 1900-2005, though not over 1878-2005.

A comment here is that if the density of the data is plotted as in several earlier examples, this suggests a square root transformation to remove skewness. Of course the numerical values of the slopes are quite different if a linear regression is fitted to square root cyclones counts instead of the raw values, but qualitatively, the results are quite similar to those just cited–significant for 1900-2005, not significant for 1878-2005–after fitting a time series model. We omit the details of this.

The second part of the analysis uses the "two-encounter" data set. In this case, fitting an ordinary least-squares linear trend to the data 1878-2005 yields an estimated slope .014 storms per year, standard error .009, not significant. The time series model (again ARMA(9,2)) leads to estimated slope .018, standard error .021, not significant.

When repeated for 1900-2005, ordinary least-squares regression leads to a slope of .042, standard error .012. The same analysis based on a time series model (ARMA(9,2)) leads to a slope of .045 and a standard error of .021. Although the standard error is much bigger under the time series model, this is still significant with a p-value of about .03.

EXAMPLE 6: U.S. LANDFALLING HURRICANES (SECTION 2.1.3.1)

The final example is a time series of U.S. landfalling hurricanes for 1851-2006 taken from the website http://www.aoml.noaa.gov/hrd/hurdat/ushurrlist18512005-gt.txt. The data consist of annual counts and are all between 0 and 7. In such cases a square root transformation is often performed because this is a variance stabilizing transformation for the Poisson distribution. Therefore, square roots have been taken here.

A linear trend was fitted to the full series and also for the following subseries: 1861-2006, 1871-2006, and so on up to 1921-2006. As in preceding examples, the model fitted was ARMA (p,q) with linear trend, with p and q identified by AIC.

For 1871-2006, the optimal model was AR(4), for which the slope was -.00229, standard error .00089, significant at p=.01.

For 1881-2006, the optimal model was AR(4), for which the slope was -.00212, standard error .00100, significant at p=.03.

For all other cases, the estimated trend was negative, but not statistically significant.