Comment on "Recent changes in climate extremes in the Caribbean region by Peterson et al. (2002). Journal of Geophysical Research, 107(D21), 4601, doi:10.1029/ 2002JD002251, 2002" by R.J. Stone.

Peterson et al. (2002) conducted a study using Caribbean meteorological data by performing ordinary least squares regression analyses on several indices of extremes derived from daily temperature and rainfall observations for the period 1958 to 1999. The authors reported several specific findings with respect to observed trends in these indices and concluded that "the climate of the Caribbean is changing" and that "these changes generally agreed with what is observed in many other parts of the world". We re-visited the study due to apparent methodological flaws with respect to sample selection and data analysis which cast suspicion on their findings.

To investigate the apparent flaws, we used, as an illustrative example, the finding of an increasing linear trend in the number or percent of very warm days during the study period. This particular finding was chosen because it is now widely used to support the popular belief that Caribbean climate is changing and has changed dramatically during the past few decades due to human-induced greenhouse warming. For example, based on this study, the Joint Nature Conservation Committee, the statutory adviser to the UK Government on national and international nature conservation, states on its official website (www.jncc.gov.uk/page-4381), referring to the Caribbean area: "The number of very warm days in the region is increasing, but the number of very cold nights is decreasing (1950s to present)".

The authors did not define the area referred to as the Caribbean region but, based on the location of the stations used to constitute the sample, we surmised that the area was bounded on the North by the 26.55°N latitude line (Freeport, Bahamas), on the South by the 10.37°N latitude line (Piarco, Trinidad), on the East by the 59.59°W longitude line (Husbands, Barbados) and on the West by the 88.30°W longitude line (PSWGIA, Belize), comprising an area approximately 4.5 million square kilometers which therefore represented the statistical population. A warm day was defined as a day in which the daily maximum temperature was greater that the 90th percentile computed using data from the period 1977 to 1997.

Sample selection was based on the following procedure. Invitations were sent out to the personnel at the various meteorological stations in the region requesting their attendance with their data at a Regional Climate Change Workshop. Volunteers from 30 stations, representing 14 countries, responded. The data were subjected to quality tests, and in the case of percent warms days, data from only 13 stations, representing 9 countries, were deemed to be acceptable for calculating the percent warm days. Nine of these stations were located at airports in urbanized areas. Not all stations, however, had available data for every year during the entire period; the number varied from 5 to 13, with a median of 10. The medians for the first half and the second half of the study period were 7 and 12 respectively. No definitive statement was made on whether complete metadata were available for any of these stations.

It is thus evident that convenience sampling, also referred to as accidental or haphazard sampling, was used in the study. Convenience samples, since they are based on non-probability sampling, can only be considered representative if the population is homogeneous. This was clearly not the case, for in addition to measurements being taken mainly at airports, the authors stated that: "…indices of extremes show considerable variability from station to station". This considerable variability is reflected in the following distribution of standard errors of the sample means for the 42 years: 9 values were between 0 and 1, 22 values were between 1.01 to 2, 6 values were between 2.01 to 3, and 5 values were greater than 3 with the maximum value = 4.00. A similar pattern of considerable variability and imprecision is also indicated in the distribution of the coefficients of variation: 6 values were between 0 to 50%, 28 values were between 50.1 to 100% and 8 values were greater than 100% with the maximum value = 174.3%.

The importance of selecting representative samples of adequate size in good statistical practice to avoid incorrect conclusions was emphasized by Chatfield (1991). Haining (2003) discussed the various spatial sampling plans available and necessary for obtaining adequately-sized representative samples to compute an unbiased estimate of the population mean with the desired precision.

We illustrate the lack of sample representativeness more specifically by looking at 1958 and 1959, the first two years of the time series. The sample means were computed using values from two stations in Florida (Key West international airport and Everglades city), values from two stations in Belize (PSWG international airport and Central Farm district) and one value from a station in the Dominican Republic (Las Américas international airport). It was unreasonable to assume that just these 5 stations constituted a representative sample of the Caribbean region while measurements over the Caribbean Sea (approximately 2.75 million square kilometers), and the numerous other islands including Cuba, the largest Caribbean island, were excluded.

Added to this is uncertainty about the reliability of inferring the direction and magnitude of regional climate trends from single-site records even over relatively homogeneous terrain (Pielke Sr. et al., 2000). Also, the spatial representativeness of a particular station is not only a function of regional climatic variability but also observational practices, and the changing microclimate within its immediate surroundings which limits its utility in representing climate over larger spatial scales (Janis and Robeson, 2004).

Additionally, it is important to point out that the sample mean is a design-based unbiased estimator of the population mean under random sampling assuming all constituents of the population have equal probability of inclusion (Brus and de Gruijter, 1997). Convenience sampling does not satisfy these assumptions and introduces selection bias in the sample mean. Consequently, we cannot know whether the fluctuations and variations observed were due to variations in weather and climate, at a micro- or macro-scale, or to variations in selection bias due to changes in sample size and sample constituents thereby resulting in a non-homogeneous time series. Inferences about climate change based on non-homogeneous time series are, of course, invalid.

The authors also performed an ordinary least squares (OLS) regression analysis and reported that the increasing trend detected was significant at the 1% level but, unfortunately, did not report standard summary statistics such as the slope, intercept, confidence interval or coefficient of determination, R^2, or more importantly, indicate whether any diagnostic checks were done to verify that the underlying assumptions of OLS regression analysis were satisfied. It must be noted that a significant P-value does not insure that the data have been fitted well (Chatterjee and Price, 1977; Montgomery and Peck, 1982)

Wilks (1995) cautioned and advised: "It is not sufficient to blindly feed data to a computer regression package and uncritically accept the results. Some of the results can be misleading if the assumptions underlying the computations are not satisfied. Since the assumptions pertain to residuals, it is important to examine the residuals for consistency with the assumptions made about their behaviour".

Following the advice above, the OLS regression was redone and the results presented in Table 1 which, apparently, confirmed the significant trend reported. However, if the assumptions were satisfied, the residual plot shown in Figure 1 should reveal a random scatter about the zero residual horizontal line. Clearly, the scatter is not random but exhibits some degree of curvature thereby demonstrating violation of the linearity assumption. In OLS regression, linearity means that there is a linear relationship between the dependent variable (Y) and the independent variable (X) which implies a constant rate of change of Y with respect to X over all values of X. Therefore, the observed violation of linearity suggests that a linear relationship between Y (percent warm days) versus X (year) is inappropriate.

To confirm the violation of linearity, the half-slope ratio method (Wilcox, 2005) was employed. The time series was divided into two half-periods, 1958-1978 and 1979-1999, and the least squares summary statistics were computed – see Table 1. A linear regression line is only considered appropriate if the half-slope ratio, the ratio of the larger slope to the smaller slope, is close to 1. The computed half-slope ratio was found to be -5.0, far from 1, thereby confirming that the OLS regression line for the period 1958 to 1999 is inappropriate and thus invalidates the finding of an increasing linear trend. Chatfield (1991) described the inappropriate P-value and the fitting of a linear regression line to nonlinear data as common statistical pitfalls – "examples of horror stories which are part of statistical folklore".

It is interesting to note that the slopes of the two half-periods were both not significantly different from zero at the 5% level. This demonstrates that, individually, the two half-periods do not exhibit departure from randomness thereby implying that the entire time series may be essentially random. To explore this phenomenon further, OLS regression analyses were performed for varying time periods starting from 1958 to 1979, then from 1958 to 1980 and so on until the end of the series. The results are shown in Table 1.

It is evident that the slopes were not significantly different from zero up to and including 1996 implying that the percent warm days showed no trend but varied randomly during the period 1958-1996. It is also evident that the last three years were responsible for the apparent increasing long-term trend reported with 1998 being a possible outlier (indicated by its large residual on the residual plot) due to the extremely strong 1997/1998 El Niño event. The apparent significant trend, as was shown earlier, was thus not real but simply an artefact of nonlinearity.

Moreover, this situation, where a relatively small percent of the data has such a significant impact on the regression line is undesirable since the regression relationship is expected to be embedded in all the observations and not merely be an artifice of a few points (Montgomery and Peck, 1982).

Draper and Smith (1998) concurred and advised: "In any data set where the estimation of one or more parameters depends heavily on a small number of observations, problems of interpretation arise. One way to anticipate such problems is to check whether the deletion of observations greatly affects the fit of the model and the subsequent conclusions. If it does, the conclusions are shaky and more data are probably needed".

We conclude that both the sample selection methodology and the subsequent statistical analysis performed by the authors were flawed and, consequently, invalidate the finding of an increasing linear trend in the number or percent warm days in the Caribbean region during the period 1958-1999.

References

Brus, D.J. and J.J. deGruijter. 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma 80: 1-44.

Chatfield, C. 1991. Avoiding statistical pitfalls (with discussion). Statistical Science 6: 240-268.

Chatterjee, S. and B. Price. 1977. Regression analysis by example. John Wiley and Sons, New York.

Draper, N.R. and H. Smith. 1998. Applied regression analysis. John Wiley and Sons, Inc., New York.

Haining, R. 2003. Spatial data analysis: Theory and practice. Cambridge University Press, Cambridge.

Janis, M.J. and S.M. Robeson. 2004. Determining the spatial representativeness of air-temperature records using variogram-nugget time series. Physical Geography 25: 513-530.

Montgomery, D.C. and E.A. Peck. 1982. Introduction to linear regression analysis. John Wiley and Sons, Inc., New York.

Peterson, T.C., M.A. Taylor, R. Demeritte, D.L. Duncombe, S. Burton, F. Thompson, A. Porter, M. Mercedes, E. Villegas, R.S. Fils, A.K. Tank, A. Martis, R. Warner, A. Joyette, W. Mills, L. Alexander, and B. Gleason. 2002. Recent changes in climate extremes in the Caribbean region. J.Geophys. Res., 107(D21), 4601, doi:10.1029/2002JD002251, 2002.

Pielke, Sr., R.A., T. Stohlgren, W. Parton, N. Doesken, J. Moeny, L. Schell and K. Redmond. 2000. Spatial representativeness of temperature measurements from a single site. Bull. Amer. Met. Soc. 81: 826-830.

Wilcox, R.R. 2005. Introduction to robust estimation and hypothesis testing. Academic Press, Boston.

Wilks, D.S. 1995. Statistical methods in the atmospheric sciences. Academic Press, San Diego.

Table1. Summary Statistics of OLS Regression Analyses for Different Periods

Period	Slope	Intercept	R^2	95% Confidence Interval for Slope	P-value
1958 -1999	0.134	-259	0.195	{0.047, 0.221}	0.003
1958 -1978	-0.063	129	0.062	{-0.180, 0.055}	0.277
1979 -1999	0.316	-620	0.177	{-0.011, 0.643}	0.057
1958 -1979	-0.054	111	0.052	{-0.161, 0.053}	0.307
1958 -1980	0.010	-14	0.001	{-0.114, 0.134}	0.868
1958-1981	0.051	-94	0.032	{-0.072,0.174}	0.402
1958-1982	0.051	-94	0.036	{-0.062,0.164}	0.363
1958-1983	0.085	-161	0.093	{-0.027,0.197}	0.130
1958-1984	0.063	-118	0.056	{-0.043,0.169}	0.234
1958 -1985	0.032	-58	0.015	{-0.073, 0.137}	0.535
1958-1986	0.005	-5	0.000	{-0.097,0.108}	0.914
1958-1987	0.025	-44	0.010	{-0.073,0.124}	0.604
1958-1988	0.026	-45	0.011	{-0.066,0.118}	0.571
1958-1989	0.024	-41	0.010	{-0.062,0.110}	0.578
1958-1990	0.037	-67	0.026	{-0.045,0.119}	0.366
1958-1991	0.038	-69	0.030	{-0.039,0.115}	0.326
1958-1992	0.026	-45	0.015	{-0.048,0.100}	0.484
1958-1993	0.023	-40	0.013	{-0.047,0.093}	0.505
1958-1994	0.025	-42	0.016	{-0.042,0.091}	0.457
1958-1995	0.062	-116	0.069	{-0.015,0.138}	0.110
1958-1996	0.060	-112	0.070	{-0.013,0.132}	0.103
1958-1997	0.077	-146	0.111	{0.005,0.148}	0.036
1958-1998	0.129	-249	0.173	{0.038,0.220}	0.007

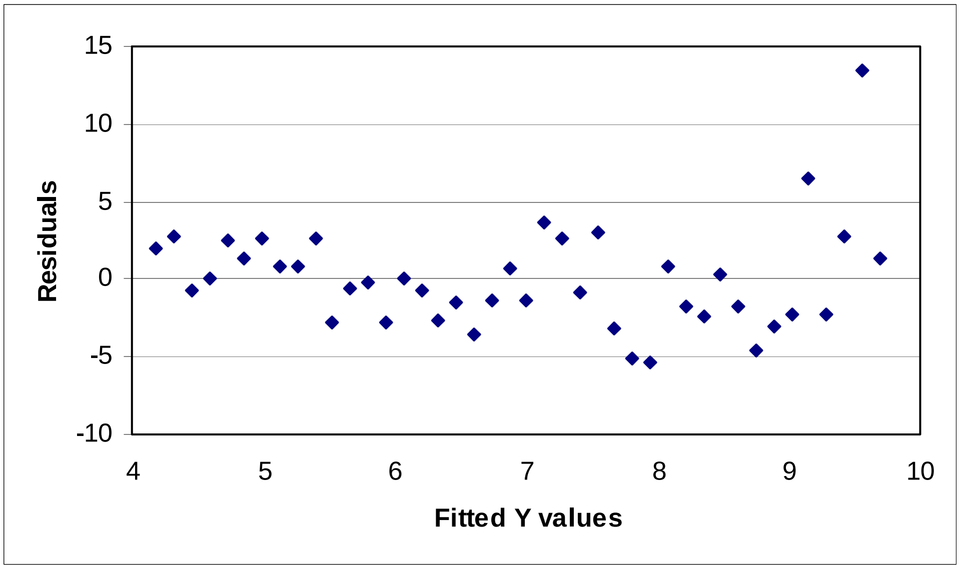

Figure1. Residual plot.