COMMENT ON "COASTAL CAVES IN BAHAMIAN EOLIAN CALCARENITES: DIFFERENTIATING BETWEEN SEA CAVES AND FLANK MARGIN CAVES USING QUANTITATIVE MORPHOLOGY"

RANE CURL

2805 Gladstone Ave., Ann Arbor, MI 48104, ranecurl@umich.edu

Waterstrat et al. (JCKS (2010), **72** (2) 61–74) compared means of geometric measurements of sea and flank margin caves with the Student t-test and concluded that these different types of caves could be "statistically differentiated." However, their use of the t-test is dependent on the data being normally distributed. The data that were analyzed, cave area to perimeter ratio, entrance-width to maximum-width ratio and short-axis to long-axis ratio, are all skewed to the right, generally have large standard deviations relative to their means, and are all non-negative. The data for at least some of the cases they analyzed may be normalized by a logarithmic transformation, yielding a close fit to log-normal distributions. This is shown in Figure 1 as log-normal probability plots of data for Flank Margin caves (FM) and San Salvador sea caves (SC) from Waterstrat (2007 – cited by the authors), where the natural logarithm of the area to perimeter ratio, $log_e(A/P)$, is plotted versus the standard unit-normal variable z corresponding to the unbiased estimator for the cumulative of the data, $F(k) = k/(n+1)$, where k is an order statistic of the data and n is the number of caves in a sample. Log-normal distributions of the data would yield straight lines. The linearity of the data is quite good, as shown by the correlation coefficients.

For these data, the respective means for $log_e(A/P)$ for FM and SC caves are 0.581 and 0.392, and their corresponding standard deviations are 0.682 and 0.411. A two-sided t-test of the difference of the these means is not significant at the 5% level of significance (los), contrary to the authors' conclusion based on applying the t-test to the unnormalized data. An additional interesting comparison can be made for the standard deviations of $log_e(A/P)$ for FM and SC caves. The slopes of the regression lines in Figure 1, which approximate the respective standard deviations, are clearly different. The variance ratio of the two data sets is 2.75, which is significant at a 1% los in a Snedecor F-test.

Waterstrat (2007) sorted data for flank margin caves and sea caves based on the general presence of speleothems in the former and their lack in the later, along with various dissolutional features, although the sorting of caves between the two types was also, in part, subjective. However the authors' statements that these statistical comparisons differentiate between sea caves and flank margin caves requires examination, if by differentiate they mean to use the comparisons for sorting the two types of caves into sea or flank margin caves. If a cave has, say, a value of $A/P = 1$ (m)

($log_e(A/P) = 0$), it is not possible to decided whether it is a sea cave or flank margin cave from the distribution data in Figure 1. This will be true in all cases where the statistical distributions of data sets overlap. For the data in Figure 1, a differentiation of type may be possible for large $log_e(A/P)$, where only flank margin caves have values over about 1.5, but even then, one needs to be concerned about the reasons for sea caves not being observed above that value, which may be because of inadequate discovery. This problem is even more obvious in the comparison of the variances of $log_e(A/P)$ of the two groups of caves. Even though the variances of the data for the two cave types are significantly different, this is no help in deciding whether a particular cave is one or the other type: both the mean and variance of the data for a cave type are not properties of individual caves, but rather of cave populations.

Still, both the mean and variance of data for a particular cave type, differentiated by other observations, are characteristics of the each cave type, and would be bases for testing geomorphic process models for their development. For example, how do the dissolutional and erosional processes make the variance of the $log_e(A/P)$ data for flank margin caves about 2.75 times greater than that for the San Salvador sea caves, or, fundamentally, why are these data apparently log-normally distributed?
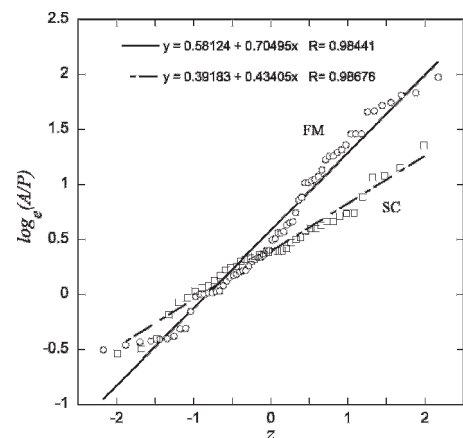

Figure 1. Log-normal probability plots of area/perimeter (*A/P*) data for flank margin caves (FM) and San Salvador sea caves (SC), from Waterstrat (2007). *Z* is the standard unit-normal variable corresponding to the estimated cumulative probability for each data point.