CC/NUMBER 16
APRIL 19, 1982

# This Week's Citation Classic

**Williams W T & Lambert J M.** Multivariate methods in plant ecology. I. Association-analysis in plant communities. *J. Ecology* **47**:83-101, 1959. [Botany Dept., University of Southampton, England]

**Given a set of presence-or-absence records of species in quadrats, $X^2$ is calculated between all species-pairs. The species with max. $\Sigma x^2$ carries most information, and $k \neq j$ subdivision on this species provides valuable ecological information. [The *SCI*® indicates that this paper has been cited over 145 times since 1961.]**

W.T. Williams Australian
Institute of Marine Science
Cape Ferguson
Townsville M.S.O., Queensland 4810
Australia

December 21, 1981

"Ecologists go into the field, lay down sample areas called quadrats, and list the species in them; they finish with a table of the presence or absence of many species in even more quadrats. They then want to sort the quadrats into groups representing different types of vegetation. Existing European hand-sorting methods were uncomfortably subjective, and I felt they were no concern of mine. However, in 1953 I appointed a new member to the staff of the Southampton botany department —a short, square, vigorous ecologist named Joyce Lambert. She was convinced that there could be a mathematical means of doing the job, and insisted that I find it. I remembered that Goodall had tackled the problem, and reread his paper.[1] It was brilliant; it suggested a means by which the problem might be solved, without quite solving it.

"Goodall calculated a measure of association, $x^2$, between all pairs of species. If there were enough high values, the set was heterogeneous, and should be divided; he aimed to find the most informative species, and to divide the set into those quadrats which had, and those which had not, this species —what we now call a 'divisive monothetic' classification. He used the most abundant species, which is usually the wrong one. I suspected that the right species was the one with the highest $X^2$ column total, though I could give no mathematical reason. I wanted a simple test problem; Lambert found me an area in the New Forest, burnt in patches by passing trains. Could the method find the burnt patches? It could, branding them as of minor importance, making its main division along a line which proved to be a major soil difference whose existence we had not suspected. We wrote a computer program and tried it on other areas, and it always worked admirably. But why did it work? I approached several statisticians who told me that the method was invalid and the results would be meaningless. But they weren't, and I eventually thought up an implausible explanation which an editor was willing to publish. Other people tried it, wrote programs for their own computers, and the method became well known. In 1963 we even discovered *why* it worked, and we became respectable.

"I wouldn't advise reading the paper now. The algebra is not quite accurate, the explanation given is unsound, and $x^2$ is not the best model —it was later superseded by an information statistic. So why is the paper still cited? I think it is simply a respect for history. It was the first method of its kind, and has had many successors.[2] Its use is now worldwide, but this paper is where it began, and citation represents a kindly acknowledgement of the fact. It brought me no immediate honours, but it launched me on a career of numerical classification which took me to Australia and later brought me an honorary doctorate and an O.B.E. But the credit really belongs to Lambert, without whose infuriating persistence I would never have started the work."

1. **Goodall D W.** Objective methods for the classification of vegetation. I. The use of positive interspecific correlation. ***Aust.. J. Bot.*** **1**:39-63, 1953.
2. **Williams W T & Lance G N.** Hierarchical classificatory methods. (Enslein K, Ralston A & Wilf H S, eds.) *Statistical methods for digital computers.* New York: Wiley, 1977. p. 269-95.