

PENNSTATE



Center for **S**tatistical **E**cology and **E**nvironmental **S**tatistics

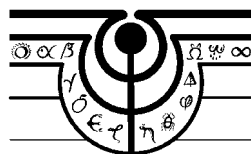
DIVERSITY PROFILES

By G. P. Patil

Center for Statistical Ecology and Environmental Statistics
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

[In Encyclopedia of Environmetrics, A. El-Shaarawi and W. W. Piegorsch, eds. John Wiley & Sons, UK. pp. 555-561, 2002]

Technical Report Number 2001-0206
TECHNICAL REPORTS AND REPRINTS SERIES
February 2001



Department of Statistics
The Pennsylvania State University
University Park, PA 16802

G. P. Patil
Distinguished Professor and Director
Tel: (814)865-9442 Fax: (814)865-1278
Email: gpp@stat.psu.edu
<http://www.stat.psu.edu/~gpp>

Diversity Profiles

Introduction

Community diversity, diversity measurement, and diversity comparisons have been important in community ecology studies for more than four decades.

Diversity is a generic term referring to the condition of being different; a useful synonym for diversity is variety. It summarizes for qualitative characteristics (e.g., tree species) what the variance summarizes for quantitative measurements (e.g., tree diameter) (Pielou 1975, p.5). Diversity is looked upon as a community property and, accordingly, the trophic, taxonomic or structural levels that comprise the community must be specifically delimited.

In making diversity comparisons, it is not uncommon to find that diversity has increased according to one index, but decreased according to a second index. This merely reflects the fact that diversity is a complex multidimensional property of a community. To view diversity through the lens of a single index is to project that multidimensional complexity onto a one-dimensional ordinal scale with distorted perception and possibly misleading conclusions. It is important to confront the multidimensional character of diversity by establishing an intrinsic, and index-free, diversity ordering. An implication of the preceding discussion is that a given pair of communities may not be comparable according to this intrinsic diversity ordering. In effect, diversity may appear to have decreased when viewed from one vantage point (i.e., index), and increased when viewed from a different perspective.

In view of the inadequacy of a single index, it is desirable to quantify diversity by

means of *diversity profiles*. A diversity profile is a curve depicting the simultaneous values of a large collection of diversity indices. Thus, the profile portrays the views of diversity from many different vantage points simultaneously and in a single picture.

Changes, or differences, in community diversity are studied by comparing profiles. If the two communities are intrinsically comparable, then one profile will lie uniformly above the other. Conversely, when the communities are not intrinsically comparable, their profiles may intersect. But even here, the profiles can reveal which portions of the community have undergone opposing diversity changes.

Measuring Diversity—An Overview

The literature on diversity measurement is immense with numerous indices and other methods available for use in measuring diversity. In addition, problems with misapplication of diversity measures evidently are commonplace in the literature. Other sources such as Dennis and Patil (1986), Gove et al. (1992, 1994, 1995, 1996), Grassle et al. (1979), Hurlbert (1971), Magurran (1988), Myers et al. (1995), Patil and Taillie (1979ab, 1982) and Pielou (1974, 1975, 1977) discuss many of these other measures and their correct use.

Semanticswise, the term *species* as used when referring to diversity is simply a convenient label for a set of distinct and countable categories comprising the community; therefore, “species” could mean a set of vegetative height strata, forest stands, or true biological species. The *community*, then, is used in an equally general sense to mean a recognizable collection of species in a space-time continuum.

Average Species Rarity

The diversity of a community consists of two components: the number of species, or species richness, and the evenness, or equitability of species as given by their relative abundance. Let s be the species richness for a community, and define the relative abundance vector for this community as $\pi = (\pi_1, \dots, \pi_s)$, such that $\sum_{i=1}^s \pi_i = 1$. The relative abundance vector may be computed from any one of a number of different measures such as number of individuals, percent cover, or biomass per unit area. The quantity used will depend on the definition of species (i.e., the composition of the s categories of interest) and there is, as of yet, no consensus on which measure to use in a given situation (Pielou 1974, Magurran 1988). A conceptual community may be written as $C(\mu)$, where μ may be some function of s and $\underline{\pi}$, or μ may have some space-time connotation. The *completely even community* is a special case where $\pi_i = \frac{1}{s} = \pi_E$ for all i , so that s alone determines the relative abundance vector. The completely even community is denoted as $C_E(s)$.

Community diversity is defined here as the average species rarity within the community. If a numerical measure of rarity can be associated with each species in the community, then the community diversity can be determined as its expected value. Thus, if the rarity of the i th species in community $C(\underline{\pi})$ is given as $R(i; \underline{\pi})$, then

$$\Delta = \Delta(\underline{\pi}) = \sum_{i=1}^s R(i; \underline{\pi})\pi_i \quad (1)$$

is the community diversity. The choice of how to assign a rarity value to each individual species will determine the diversity measure.

Patil and Taillie (1979a, 1982) discuss two types of rarity measures: *dichotomous-type* and *rank-type*. With the dichotomous-type rarity measures, the rarity of the i th species depends only on its own relative abundance π_i . In this case, rarity is denoted $R(\pi_i)$, and the measure of community diversity is defined as $\Delta = \sum_{i=1}^s R(\pi_i)\pi_i$. For rank-type measures, however, species rarity depends only on its rank. Consider the *ranked* relative abundance vector, $\underline{\pi}^\#$, which is determined by ranking the components of $\underline{\pi}$ from greatest to least so that $\underline{\pi}^\# = (\pi_1^\#, \dots, \pi_s^\#)$, where $\pi_1^\# \geq \dots \geq \pi_s^\#$. Then if the rarity of the i th-ranked species is denoted as $R(i)$, community diversity is given as $\Delta = \sum_{i=1}^s R(i)\pi_i^\#$. Examples of both types of rarity measures and their associated diversity functions are given in the following sections.

Diversity Indices

Several of the most frequently used diversity indices may be conveniently expressed under the umbrella of average species rarity through judicious choice of rarity functions. Species richness, species count, Shannon's, and Simpson's indices all may be derived from this theory as follows

$$\Delta_{SR} = \sum_{i=1}^s \left(\frac{1}{\pi_i} \right) \pi_i = s \quad \text{species richness,} \quad (2)$$

$$\Delta_{SC} = \sum_{i=1}^s \left(\frac{1}{\pi_i} - 1 \right) \pi_i = s - 1 \quad \text{species count,} \quad (3)$$

$$\Delta_{Sh} = \sum_{i=1}^s (-\log \pi_i) \pi_i = \sum_{i=1}^s \pi_i \log \pi_i \quad \text{Shannon,} \quad (4)$$

$$\Delta_{Si} = \sum_{i=1}^s (1 - \pi_i) \pi_i = 1 - \sum_{i=1}^s \pi_i^2 \quad \text{Simpson} \quad (5)$$

where the term in parentheses denotes the species rarity function used in each case.

Table 1 presents a hypothetical example of three forest stands composed of just five or fewer species of trees. The relative abundances of these tree species based on some quantitative measure of abundance are given, and the diversity indices (2) through (5) calculated from these relative abundances also are shown for each community. The example clearly shows the inconsistency of the different indices in their ranking of these three communities. For example, $\Delta_{SC}(\text{Stand 1}) > \Delta_{SC}(\text{Stand 2})$, but $\Delta_{Sh}(\text{Stand 1}) < \Delta_{Sh}(\text{Stand 2})$ and $\Delta_{Si}(\text{Stand 1}) < \Delta_{Si}(\text{Stand 2})$. This is an interesting comparison because it illustrates how one may be lead to the conclusion that a community with fewer species (Stand 2) can be more diverse than one with more species (Stand 1) using either Shannon's or Simpson's index. Similar inconsistencies among the indices may be found by comparing Stands 1 and 3. The only comparison that is consistently ordered with all indices is $\Delta(\text{Stand 2}) > \Delta(\text{Stand 3})$. This inconsistency of different diversity indices evidently is quite common when making comparisons between communities (Hurlbert 1971) and arises from a lack of intrinsic diversity ordering between the communities being compared (See the following section).

Diversity Profiles

Diversity profiles allow the graphical comparison of diversity between communities. One set of profiles that incorporates indices (3) through (5) as point estimates along the curve are the so-called Δ_β profiles of Patil and Taillie (1979a, 1982). Since the Δ_β profile incorporates indices developed from dichotomous-type rarity measures, it too may be developed

Table 1: Three hypothetical forest stand communities composed of five or fewer species of trees.

Species	Stand		
	1	2	3
	<i>Relative abundance</i>		
<i>Pinus strobus</i>	0.50	0.25	0.35
<i>Quercus rubra</i>	0.30	0.25	0.35
<i>Tsuga canadensis</i>	0.10	0.25	0.30
<i>Acer rubrum</i>	0.05	0.25	0.00
<i>Betula papyrifera</i>	0.05	0.00	0.00
Total:	1.00	1.00	1.00
	<i>Diversity index</i>		
Δ_{SR}	5	4	3
Δ_{SC}	4	3	2
Δ_{Sh}	1.24	1.39	1.10
Δ_{Si}	0.65	0.75	0.67

in the same manner:

$$\Delta_{\beta} = \sum_{i=1}^s \frac{(1 - \pi_i^{\beta})}{\beta} \pi_i = \frac{1 - \sum_{i=1}^s \pi_i^{\beta+1}}{\beta}, \beta \geq -1. \quad (6)$$

The restriction that $\beta \geq -1$ assures that Δ_{β} has certain desirable properties (Patil and Taillie 1979a, 1982). The species count, Shannon and Simpson indices are related to Δ_{β} by $\Delta_{SC} = \Delta_{-1}$, $\Delta_{Sh} = \Delta_0$, and $\Delta_{Si} = \Delta_1$.

The Δ_{β} diversity profiles for the three stands in Table 1 are presented in Figure 1. Note that the profile for Stand 1 crosses both profiles for Stands 2 and 3. The profile for Stand 1 crosses that of Stand 2 at $\beta = -0.45$, which explains why both Δ_{Sh} and Δ_{Si} rank diversity of these two communities differently from Δ_{SC} . On the other hand, the

profiles for Stands 1 and 3 cross at $\beta = 0.62$ showing how Δ_{SC} and Δ_{Sh} rank these two communities differently from Δ_{Si} . In general, it also is possible for two Δ_β profiles to cross at $\beta > 1$ or for them to cross more than once; in either case, even calculating all three indices (Δ_{SC} , Δ_{Sh} , and Δ_{Si}) alone may not be enough to show the inconsistent ranking of communities at larger β . Calculating and plotting Δ_β profiles for $\beta > 1$ may not be helpful either because the profiles tend to converge quickly beyond this point and intersections do not resolve—an algorithm for numerically finding the intersections of any two Δ_β profiles is required in this case.

Figure 1: Δ_β profiles for the three hypothetical forest stand communities in Table 1

Perhaps the most useful way to compare diversity between communities is by the concept of *intrinsic diversity ordering*. This concept may be defined as follows (Patil and

Taillie 1982)

Community C' is *intrinsically more diverse* than community C (written $C' \overset{imd}{>} C$) provided C leads to C' by a finite sequence of 1. introducing a species, 2. transferring abundance from more to less abundant species without reversing the rank-order of the species, and 3. relabeling species (i.e., permuting the components of the abundance vector).

Patil and Taillie (1982) note that this ordering is only partial and two given communities need not be intrinsically comparable.

A diversity profile approach has been developed by Patil and Taillie (1979a,b, 1982) using a rank-type rarity measure on $\underline{\pi}^\#$ that incorporates the concepts of intrinsic diversity ordering defined above. Let

$$R(i) = \begin{cases} 1 & \text{if } i > j; \\ 0 & \text{if } i \leq j, \end{cases}$$

for $1 \leq j \leq s$. Then average species rarity is given as

$$T_j = \sum_{i=j+1}^s \pi_i^\#, \quad j = 1, \dots, s-1, l; \quad (7)$$

where $T_s = 0$ and $T_0 = 1$. The quantity in (7) is termed the right tail-sum of the ranked relative abundance vector $\underline{\pi}^\#$, and when a plot of the (j, T_j) pairs is constructed for each community, the resulting profiles are termed *intrinsic diversity profiles* (Patil and Taillie 1979b). Any intrinsic orderings of the communities, if they exist, can be determined with the intrinsic diversity (T_j) profiles.

The right tail-sum profiles for the three stands in Table 1 are plotted in Figure 2. Notice that the profile for Stand 1 crosses both those for Stands 2 and 3, but that the

profile for Stand 2 is everywhere above that for Stand 3. It follows that the only intrinsic diversity ordering for these stands is $C(\text{Stand 2}) \stackrel{imd}{>} C(\text{Stand 3})$. This is consistent with the findings of the indices in the section on Average Species Rarity and the Δ_β profiles. The Δ_β profiles are isotonic to intrinsic diversity ordering in that, if an intrinsic diversity ordering exists, they will preserve it. However, the Δ_β profiles may not cross even if the T_j profiles do; therefore, the Δ_β profiles do not necessarily reflect intrinsic diversity ordering. Since the diversity indices discussed have the same properties as the Δ_β profiles, it should be emphasized that, of the methods presented thus far, the T_j profiles are the most reliable measure of intrinsic diversity ordering between communities.

Figure 2: Right tail-sum (T_j) profiles for the three hypothetical forest stand communities in Table 1.

Analysis Methods

Diversity indices such as Shannon's index and diversity profiles such as T_j and Δ_β were developed for use when the $\underline{\pi}$ have been calculated from complete enumeration of the population. Patil and Taillie (1979b) note that "most diversity indices, when evaluated on a sample, consistently underestimate community diversity by an amount varying with sample size." They presented a jackknifing method that may help to correct for this bias and, in addition, provide estimates of precision that can be used for hypothesis testing. In particular, the jackknifing method was used to test whether intersections that occur in the T_j profiles are real or simply a sampling effect. Patil and Taillie's (1979b) jackknife approach is used on all Δ_β profiles for bias correction and intersection testing. The procedure is summarized as follows:

- (1) Calculate and plot the T_j profiles with all plots included for the community of interest and identify any intersections.
- (2) Calculate and plot the Δ_β profiles with all plots included.
- (3) Jackknife the Δ_β profiles for each community as follows
 - (i) Calculate $\Delta_\beta^{(-i)}$, the Δ_β profile with the i^{th} plot data omitted.
 - (ii) Calculate a pseudo value as given in (3.1) at each β :

$$\psi_\beta^{(-i)} = n\Delta_\beta - (n-1)\Delta_\beta^{(-i)}. \quad (8)$$

- (iii) Repeat steps (i) and (ii) n times, where n is the number of plots.

(iv) Calculate the average of the pseudo values and the estimated variance at each β :

$$\hat{\Delta}_\beta = \frac{1}{n} \sum_{i=1}^n \psi_\beta^{(-i)}, \quad (9)$$

$$\widehat{Var}(\hat{\Delta}_\beta) = \frac{1}{n(n-1)} \sum_{i=1}^n (\psi_\beta^{(-i)} - \hat{\Delta}_\beta)^2. \quad (10)$$

(4) Plot the jackknifed $\hat{\Delta}_\beta$ profiles for each community.

The inference procedure involves forming the composite null hypothesis given by Patil and Taillie (1979b). To illustrate, Figure 3 shows $\hat{\Delta}_\beta$ profiles for two communities (C_1 and C_2) with an intersection found at the point β^* . In this case, the procedure would be to choose β' and β'' so that $\beta' < \beta^* < \beta''$, and form the composite null hypothesis:

Figure 3: Δ_β profiles illustrating an intersection and hypothetical points (B', β^*, β'') for the composite null hypothesis described in the text.

$$\mathcal{H}'_0 : \hat{\Delta}_{\beta'}(C_1) \geq \hat{\Delta}_{\beta'}(C_2) \quad (11)$$

$$\mathcal{H}''_0 : \hat{\Delta}_{\beta''}(C_1) \leq \hat{\Delta}_{\beta''}(C_2). \quad (12)$$

Simultaneous rejection of the composite null hypothesis at some specified level of significance would establish that $\Delta_{\beta'}(C_1) < \Delta_{\beta'}(C_2)$ and $\Delta_{\beta''}(C_1) > \Delta_{\beta''}(C_2)$; that is, the two community profiles $\hat{\Delta}_{\beta}(C_1)$ and $\hat{\Delta}_{\beta}(C_2)$ *do* intersect and, therefore, the two communities are *not* intrinsically comparable.

To test the composite null hypothesis calculate the two-sample t statistics:

$$t_{\beta} = \frac{\hat{\Delta}_{\beta}(C_1) - \hat{\Delta}_{\beta}(C_2)}{\left[\widehat{Var}(\hat{\Delta}_{\beta}(C_1)) + \widehat{Var}(\hat{\Delta}_{\beta}(C_2))\right]^{\frac{1}{2}}}, \quad \beta = -1, -.97, -.94, \dots, 2. \quad (13)$$

If at least one t_{β} (*viz.*, the maximum t_{β}) on each side of the intersection is significant when compared to the one-tailed Bonferonni-adjusted $t(1 - \alpha/g; n - 1)$ value (where $g = 2$ and $\alpha = 0.05$ here), then the implication is that the profiles intersect, and the communities C_1 and C_2 are *not* intrinsically comparable. If none of the t_{β} values calculated are significant, then it may be inferred that there is no difference in diversity between the two communities being compared.

This procedure must be viewed as only an approximate test because it involves “difficult and unresolved questions of simultaneous inference” (Patil and Taillie 1979b). However, even with this qualification, it remains a valuable procedure over simply treating the $\underline{\pi}$ as population values. In addition, it by no means diminishes the usefulness of the bias correction aspect of the jackknife.

Diversity in Management Decisions

Structural diversity is a very important component of the community. The vertical diversity of foliage in spruce-fir forests, for example, was found by MacArthur and MacArthur (1961) to be a good predictor of bird species diversity. In uneven-aged forest management the shape of the diameter distribution, maximum tree diameter, and stocking level are of paramount importance to the forest manager.

Maximizing Diversity: General

For any given vector of relative abundances $\underline{\pi}$ a diversity profile can be generated by allowing β to vary in (6). In this section, $\Delta_\beta(\underline{\pi})$ is viewed in the opposite sense: β is held fixed and $\underline{\pi}$ is allowed to vary subject to the constraints that $\beta \geq -1$ and $\sum_{i=1}^s \pi_i = 1$. When this is done a *diversity surface* is generated at β .

That the completely even community maximizes diversity for a given number of species s is well known (Patil and Taillie 1979a, 1982, Pielou 1974, Solomon 1979). In general, the problem may be formulated for $\Delta_\beta(\underline{\pi})$ as

$$\begin{aligned} \text{Max}_{\{\underline{\pi}\}} \Delta_\beta(\underline{\pi}) \quad \text{St} : \sum_{i=1}^s \pi_i = \underline{\pi}'\underline{1} = 1. \end{aligned} \tag{14}$$

It is straightforward to show that the solution to (14) is $C_E(s)$. This is an interesting finding because it allows the introduction of an alternative objective function into model (14); maximizing (14) should, therefore, be equivalent to the following problem

$$\begin{aligned}
& \text{Min} && \sum_{i=1}^s |\pi_E - \pi_i| \\
& \{\underline{\pi}\} && \\
& \text{St :} && \sum_{i=1}^s \pi_i = \underline{\pi}' \underline{1} = 1.
\end{aligned} \tag{15}$$

The *unevenness* measure $\sum_{i=1}^s |\pi_E - \pi_i|$ in (15) is known as the Pietra index of income inequality in economics (Arnold 1987). The diversity surface for this formulation is presented in Figure 4. The level curves for (15) are hexagonal. This surface also is minimized at the completely even community, implying that diversity is at its maximum.

Figure 4: Model (15) diversity surface for all e-species communities. (Note that the surface is actually piecewise linear.)

References

- Arnold, B. C. (1987). *Majorization and the Lorenz Order: A Brief Introduction*. Springer-Verlag, NY.
- Dennis, B., and Patil, G. P. (1986). Profiles of diversity. In *Encyclopedia of Statistical Sciences, Volume 7*, S. Kotz and N. L. Johnson, eds. John Wiley, NY. pp. 292–296.
- Gove, J. H., Martin, C. W., Solomon, D. S., and Hornbeck, J. W. (1992). Plant species diversity on even-aged harvests at the Hubbard Brook Experimental Forest: 10 year results. *Canadian Journal of Forest Research*, **22(11)**, 1800–1806.
- Gove, J. H., Patil, G. P., Swindel, B. F., and Taillie, C. (1994). Ecological diversity and forest management. In *Handbook of Statistics, Volume 12: Environmental Statistics*, G. P. Patil and C. R. Rao, eds. North Holland/Elsevier Science Publishers, New York and Amsterdam. pp. 409–462.
- Gove, J. H., Patil, G. P., and Taillie, C. (1995). A mathematical programming model for maintaining structural diversity in uneven-aged forest stands with implications to other formulations. *Ecological Modelling*, **79**, 11–19.
- Gove, J. H., Patil, G. P., and Taillie, C. (1996). Diversity measurement and comparison with examples. In *Biodiversity in Managed Landscapes: Theory and Practice*, R. C. Szaro, and D. W. Johnson, eds. Oxford University Press, NY. pp. 157–175.
- Grassle, J. F., Patil, G. P., Smith, W. K. and Taillie, C., eds., (1979). *Ecological diversity*

- in theory and practice*. International Co-operative Publishing House, Fairland, MD.
- Hurlbert, S. H. (1971). The nonconcept of species diversity: A critique and alternative parameters. *Ecology* **52**(4), 577–586.
- MacArthur, R. H. and MacArthur, J. W. (1961). On bird species diversity. *Ecology* **42**(3), 594–598.
- Magurran, A. E. (1988). *Ecological Diversity and its Measurement*. Princeton University Press, Princeton, NJ.
- Myers, W. L., Patil, G. P., and Taillie, C. (1995). Comparative paradigms for biodiversity assessment. In *Measuring and Monitoring Biodiversity in Tropical and Temperate Forests*, T. J. Boyle and B. Boontawee, eds. CIFOR, Bogor, Indonesia. pp. 67–85.
- Patil, G. P. and Taillie, C. (1979a). An overview of diversity. In: J. F. Grassle, G. P. Patil, W. K. Smith and C. Taillie, eds., *Ecological Diversity in Theory and Practice*. International Co-operative Publishing House, Fairland, Maryland, 3–27.
- Patil, G. P. and Taillie, C. (1979b). A study of diversity profiles and orderings for a bird community in the vicinity of Colstrip, Montana. In: G. P. Patil and M. L. Rosenzweig, eds., *Contemporary Quantitative Ecology and Related Ecometrics*. International Co-operative Publishing House, Fairland, Maryland, 23–48.
- Patil, G. P. and Taillie, C. (1982). Diversity as a concept and its measurement. *Journal of the American Statistical Association* **77**, 548–567.

Pielou, E. C. (1974). *Population and Community Ecology*. Gordon and Breach, New York.

Pielou, E. C. (1975). *Ecological Diversity*. John Wiley and Sons, NY.

Pielou, E. C. (1977). *Mathematical Ecology*. John Wiley and Sons, NY.

Solomon, D. L. (1979). A comparative approach to species diversity. In: J. F. Grassle, G. P. Patil, W. K. Smith, C. Taillie, eds., *Ecological diversity in theory and practice*. International Co-operative Publishing House, Fairland, Maryland, 29–35.

G. P. Patil

Center for Statistical Ecology and Environmental Statistics

The Pennsylvania State University

University Park, PA 16802 USA

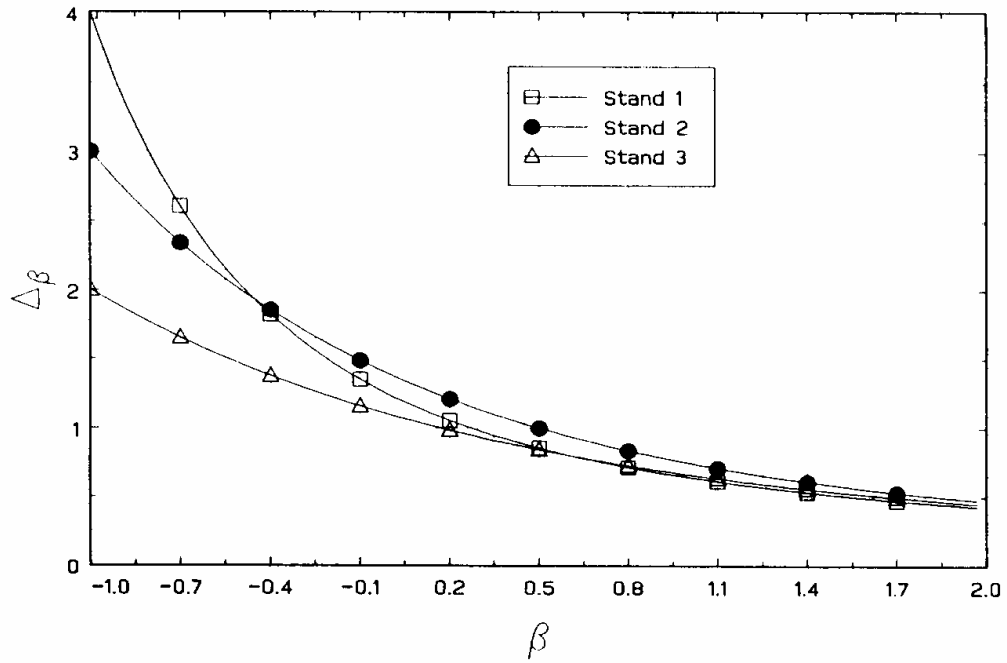


Figure 1

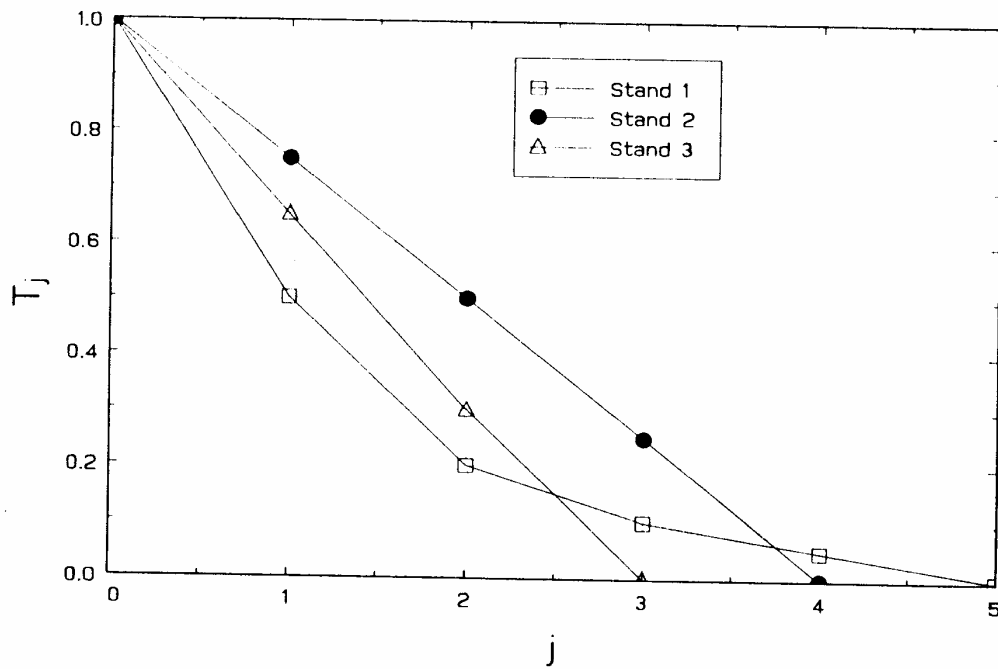


Figure 2

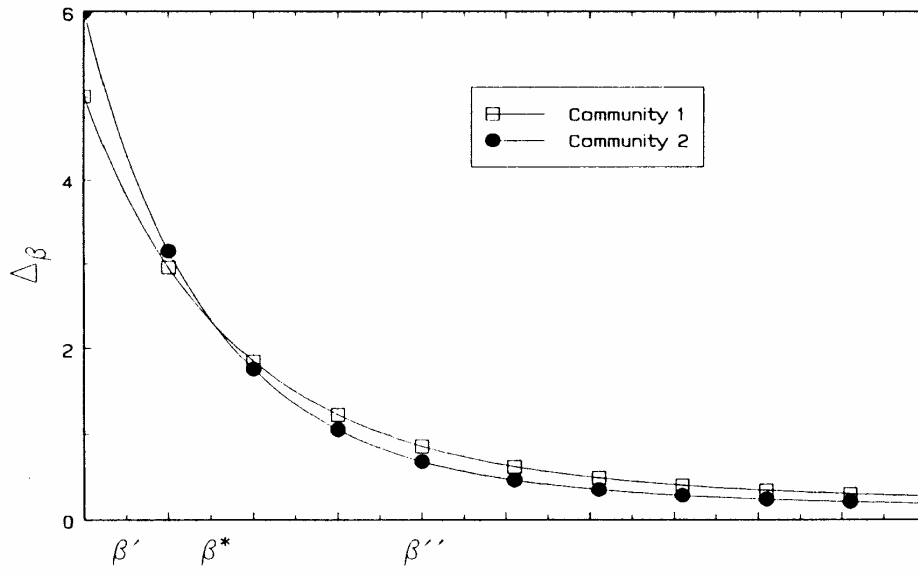


Figure 3

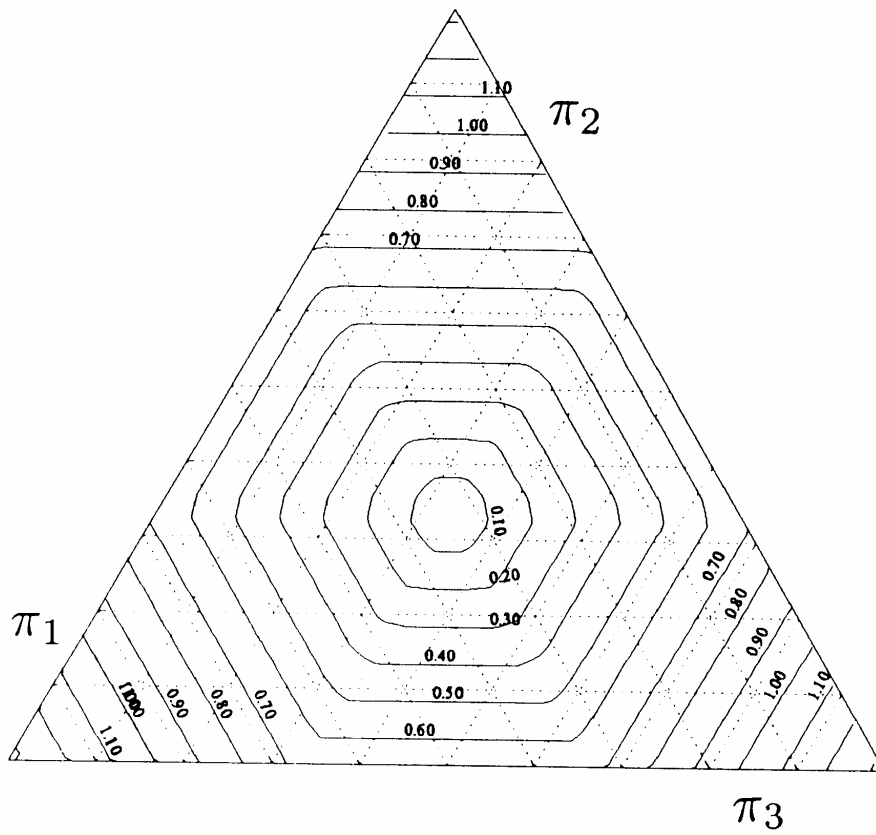


Figure 4