

TREE ASYMMETRY—A SENSITIVE AND PRACTICAL MEASURE FOR BINARY TOPOLOGICAL TREES

- JAAP VAN PELT, HARRY B. M. UYLINGS and
RONALD W. H. VERWER
Netherlands Institute for Brain Research,
Meibergdreef 33, 1105 AZ Amsterdam,
The Netherlands
- ROBERTA J. PENTNEY
Department of Anatomical Sciences,
Medical School, Sherman Hall,
State University of New York,
Buffalo, NY 14214, U.S.A.
- MICHAEL J. WOLDENBERG
Department of Geography,
Wilkeson Quad.,
State University of New York,
Buffalo, NY 14261, U.S.A.

The topological structure of a binary tree is characterized by a measure called tree asymmetry, defined as the mean value of the asymmetry of its partitions. The statistical properties of this tree-asymmetry measure have been studied using a growth model for binary trees. The tree-asymmetry measure appears to be sensitive for topological differences and the tree-asymmetry expectation for the growth model that we used appears to be almost independent of the size of the trees. These properties and the simple definition make the measure suitable for practical use, for instance for characterizing, comparing and interpreting sets of branching patterns. Examples are given of the analysis of three sets of neuronal branching patterns. It is shown that the variance in tree-asymmetry values for these observed branching patterns corresponds perfectly with the variance predicted by the used growth model.

1. Introduction. Natural branching patterns generally have common features in their shapes, notwithstanding the great variability in these shapes. In studying groups of branching patterns one is interested in the quantitative characterization of such common features, with the aim of searching for significant differences between groups, to detect effects of experimental conditions or to establish modes of growth. The question of which features most prominently define such differences is not trivial if the number of morphological features required to completely reconstruct the tree structure is large. Facing such questions with respect to the analysis of neuronal branching patterns (dendritic and axonal trees) we have split up the morphological features into metrical (spatial) and topological ones.

In this paper we will focus on the tree's topology (tree type), defined by the connectivity pattern of the segments in the tree (Van Pelt and Verwer, 1983). Topological variability in sets of natural trees is completely defined by the frequency of occurrence of the tree types. However, the number of different tree types increases rapidly with the size of the trees (number of segments). For instance, there exist about 10^{192} tree types with 500 terminal segments. Evidently, only a fraction of this number will occur in observed tree sets making tree-type frequency distributions not appropriate for statistical analysis. Alternatively, one may use a measure of tree topology, to describe the topological variability. Measures of tree topology can be defined in many different ways, dependent on the particular aspects of interest. Especially in the geomorphological and neuromorphological literature such measures have critically been evaluated (e.g. Werner and Smart, 1973; Jarvis and Werrity, 1975; Uylings *et al.*, 1989a; Van Pelt *et al.*, 1989a; and Verwer *et al.*, 1992). The last two studies have indicated that each of the evaluated measures is based on one of only three different ordering schemes of elements in a tree (bifurcation points and segments) i.e.: (1) the Strahler-ordering scheme with the branching ratio as main measure (e.g. Jarvis and Woldenberg, 1984); (2) the centrifugal ordering scheme with measures as mean order or maximal order, which is equivalent to the topological diameter; and (3) the ordering based on the partitions in a tree with the partition asymmetry and tree asymmetry as main measures (Verwer and Van Pelt, 1986). A partition is an integer number pair indicating the division of terminal segments over the two subtrees of a bifurcation point. Each of the three ordering schemes stresses different aspects of the tree's topology although they are not independent of each other. In the above mentioned studies the measures are compared on the basis of their ability to distinguish different tree types (i.e. having unique values for different trees).

Using the third ordering scheme a tree can be represented by its tree partition set, being the set of partitions from all bifurcation points. Apart from the tree type itself, the tree partition set appears to be the most unique representation of the tree's topology with the tree-asymmetry measure as the most discriminative one (Van Pelt *et al.*, 1989a). Important for the ability to characterize common topological features in the tree shapes are the statistical properties of the measures, which must be studied on the basis of tree-type frequency distributions. For the analysis of geological (two-dimensional) channel networks (river patterns), Werner and Smart (1973) have evaluated the branching ratio and measures based on the centrifugal ordering scheme using a uniform distribution of (two-dimensional) tree types ("random topology model"). Jarvis and Werrity (1975) have used this uniform distribution to compare several measures on the basis of the Shannon information content. Horsfield *et al.* (1987) studied the branching ratio using their synchronous

growth model for binary trees. They showed how the asymptotic value of the expected branching ratio depends on the model parameters. Additionally, they relate the outcomes to those from the sequential growth model of Van Pelt and Verwer (1986). This model, indicated by its two parameters as the *QS* growth model and summarized in Section 3, describes growth as a sequence of single branching events. Van Pelt *et al.* (1989b) used the *QS* growth model in a detailed study of the mean-order measure (from the second ordering scheme). The expected values of both the mean and the variance of this measure appeared to correspond perfectly to those from observed neuronal branching patterns.

Although measures of tree topology should primarily be sensitive for the connection pattern of the segments, their outcomes also depend on the size of the tree (the number of segments) or the degree (the number of terminal segments). The immediate comparison of trees of different sizes is then not at all trivial. Using growth models, the relation between trees of different size is implicitly defined by the growth rules and one can determine how the outcomes of a measure, calculated for a group of trees of equal size, depend on the size. Shreve (1966) showed for the “random topology model” that the branching ratio becomes asymptotically stable for large trees. Horsfield and Woldenberg (1986) demonstrated for both the random terminal and the random segmental growth model an oscillatory behavior of the branching ratio with decreasing amplitude for increasing size of the trees. Horsfield *et al.* (1987) showed how the asymptotic value of the branching ratio relates to the parameters in both the sequential and the synchronous growth model. Van Pelt *et al.* (1989b) fitted the parameters from the *QS* growth model on the observed data points in a plot of mean-order *vs* degree. This (regression) method can be used for any size distribution of the trees and has the advantage that the observed data set is finally described by only two (best-fitting) model parameters. However, the required use of a growth model including inherent assumptions may be considered as a disadvantage if only description is aimed at. A general rescaling principle to relate the expected centrifugal-order distributions or the expected partition distributions of trees of different sizes has unfortunately not been found. We found, however, in preliminary studies that the tree-asymmetry measure was surprisingly stable for a large range of degrees and under several model conditions.

Because of the attractive properties of the partition ordering scheme and the tree-asymmetry measure we considered it worthwhile to study in detail the statistical behavior of the tree-asymmetry measure for the *QS* growth model. It is shown that the expectation of the tree asymmetry strongly depends on the mode of growth. In contrast, the tree-asymmetry expectation and variance appear to be almost independent of the size of the tree. Only small trees show larger variance which rapidly becomes smaller with increasing tree size. An

analytical proof of the asymptotic stability could only be given for the random terminal growth model. The results for the other growth modes are based on sets of randomly generated trees.

To exemplify the method a set of rat Purkinje cell dendritic trees has been analysed. The asymmetry values of these large dendritic trees indeed show a high stability of the mean asymmetry per size class. The observed variance per size class also appears to correspond perfectly with the model predictions. Because of the minimal variance between the mean values per size class, the mean tree asymmetry for the whole set is an efficient descriptor for the topological structure of the whole tree set, without reference to any growth model and its implicit assumptions. This is of practical importance for the comparison of groups of trees. In a second example two sets of small dendritic trees from pyramidal and multipolar non-pyramidal cells are analysed. This example also shows stable mean values and a perfect agreement between the predicted and observed variances.

The partition-asymmetry and the tree-asymmetry measures are defined in Section 2. Section 3 is dedicated to the statistical properties of these measures for the *QS* growth model. Section 4 describes the tests of the observed data sets. A discussion of the obtained results is given in Section 5.

2. Topological Trees, Partition Asymmetry and Tree Asymmetry. The topology of a tree is determined by the connectivity pattern of its segments. Trees can then be identified by formal representations of such patterns, called tree types (Van Pelt and Verwer, 1983). The degree of a tree denotes its number of terminal segments. A tree with n terminal segments (degree n) has $n-1$ intermediate segments (and thus $n-1$ bifurcation points) and in total $2n-1$ segments. We shall restrict ourselves to binary three-dimensional (3D) trees and will indicate the corresponding set of "3D" tree types of degree n by $\{\alpha_i^n | i=1, \dots, N_\alpha^n\}$. The number of different "3D" tree types N_α^n is given by the recurrent expression:

$$N_\alpha^n = \frac{1}{2} \left(\sum_{r=1}^{n-1} [N_\alpha^r N_\alpha^{n-r} + (1 - \varepsilon(n)) N_\alpha^{n/2}] \right) \quad \text{with} \quad N_\alpha^1 = 1 \quad (1)$$

(Harding, 1971). The quantity $\varepsilon(n)$ equals 0 for even n and equals 1 for odd n . A reasonable approximation for the order of magnitude is given by $N_\alpha^n \approx 2.4^n$. For binary 2D trees, like rivers, a distinction can be made between "left" and "right" at bifurcation points, resulting in a larger set of "2D" tree types $\{\tau_i^n | i=1, \dots, N_\tau^n\}$ (Shreve, 1966; Smart, 1969). The number of "2D" tree types is given by:

$$N_\tau^n = \frac{1}{2n-1} \binom{2n-1}{n} \quad (2)$$

(Caley, 1859; cf. Shreve, 1966).

The set $\{\alpha\}$ is the subset of tree types in $\{\tau\}$ which cannot be transformed into each other by rotation of the subtrees at a bifurcation point. Henceforth we will refer to a “3D” tree type when using the term tree type. Each bifurcation in a binary tree corresponds to a pair of subtrees. The degrees of both subtrees, say, r and s , constitute a pair of numbers (r, s) , called a partition. The sum of both numbers is called the degree of the partition. A partition will be indicated as (r, s) with $r \leq s$, or as $(r, m - r)$ with m indicating its degree. Evidently, there are $[m/2]$ possible partitions of degree m , i.e. $(1, m - 1), \dots, ([m/2], m - [m/2])$. The symbol $[x]$ denotes the greatest integer, smaller than or equal to the number x . We will also use the notation m_e for the greatest even integer, smaller than or equal to m . Thus, for the integer m we have $m_e = m - \varepsilon(m)$ and $[m/2] = m_e/2$. In a tree of degree n , the root segment will end in a partition of degree n , all other partitions will have smaller degrees. The number of possible partitions in the range of degrees $(2, 3, \dots, n)$ is equal to $\sum_{i=2}^n [i/2] = (n^2 - \varepsilon(n))/4$. This number is smaller than the number of tree types of degree n .

2.1. Partition asymmetry. A binary partition of degree N can be considered as a division of N elements into two groups with N_1 and N_2 elements ($N = N_1 + N_2$). To indicate the deviation from an equal division we can use a normalized dispersion measure, say A_p , defined as $A_p = D/D_{\max}$, with value domain $[0,1]$. Here, D indicates a dispersion measure and D_{\max} its maximum value. For the dispersion measure D we have chosen the absolute difference between both values, $D = |N_2 - N_1|$ (Van Pelt and Verwer, 1986). The size of a binary tree can be expressed by the total number of its segments $(2n - 1)$, by the number of its terminal segments (n) or by the number of its bifurcation points $(n - 1)$. Any of these tree elements can be used for the calculation of the dispersion measure. For the normalized dispersion measure, however, all three options give the same value. Expressing the partition as a division of terminal segments we obtain generally for the partition $(r, m - r)$:

$$A_p(r, m - r) = \frac{m - 2r}{m - 2}, \quad \text{if } m > 2, \text{ with } r \leq m - r, \tag{3}$$

with the maximum dispersion occurring for the partition $(1, m - 1)$. The normalized dispersion measure is not defined for $m = 2$ because $D_{\max} = 0$ for the partition $(1,1)$. By definition, however, we take $A_p(1,1) = 0$. The chosen normalized dispersion measure can be generalized to multifurcations, occurring in trees with branching points from which more than two subtrees emerge, and is called proportional sum of absolute deviations (PSAD) (Verwer and Van Pelt, 1986). For bifurcations we shall denote the normalized dispersion measure as an asymmetry measure, having a value of 1 for the most asymmetric partition $(1, m - 1)$ and a value of 0 for the symmetric one $(m/2, m/2)$.

2.2. *Tree asymmetry.* Let α_i^n denote a tree with number i from the set of trees of degree n . A measure for the asymmetry of the tree can be defined as the weighted mean value of all the $n-1$ partition asymmetries in the tree as:

$$A_t(\alpha_i^n) = \frac{1}{w(\alpha_i^n)} \sum_{j=1}^{n-1} w_{ij} A_p(r_{ij}, s_{ij}), \quad \text{with } w(\alpha_i^n) = \sum_{j=1}^{n-1} w_{ij} \quad (4)$$

indicating the sum of the weight factors for this tree. Here, the index j runs over all $n-1$ partitions (r_{ij}, s_{ij}) in tree α_i^n with n, r_{ij} and s_{ij} indicating the degree of the tree and the degrees of both subtrees at partition j , respectively. The degree m_{ij} of partition j is equal to $m_{ij} = r_{ij} + s_{ij}$. Each partition has a weight w_{ij} that still has to be defined. Several options for the weight factors will be investigated with the aim of obtaining a measure that is optimally sensitive for topological differences between trees.

Option 1. All $n-1$ partitions have equal weight factor, for example one. Then, the sum of weight factors is equal to $n-1$ for any tree of degree n , and equation (4) reduces to:

$$A_t(\alpha_i^n) = \frac{1}{n-1} \sum_{j=1}^{n-1} A_p(r_{ij}, s_{ij}). \quad (5)$$

Option 2. One may argue that it is sufficient to sum only over the partitions of degree $m_{ij} > 3$ because the partitions of degree 2 and 3 are trivial ones, i.e. (1,1) and (1,2), and do not further discriminate between the tree types. Then, the weight factors will be defined as $w_{ij} = 1$ if $m_{ij} > 3$ and $w_{ij} = 0$ if $m_{ij} \leq 3$. The number of (1,1) and (1,2) partitions in a tree of degree n depend on the tree type and the tree asymmetry will be defined on a varying number of relevant partitions.

Option 3. Additionally, one may argue that relevant partitions ($m > 3$) of different degree should have different weights because they are defined on a different number of tree elements that has to be distributed over the two subtrees. For example, the two subtrees in a partition of degree m contain $m-2$ bifurcation points and we may therefore assign a weight of $m-2$ to a partition of degree m .

Option 4. Finally, one may argue that in choosing how to distribute $m-2$ bifurcation points over two subtrees the first choice is trivial and only $m-3$ choices are meaningful. So as to say, there are $m-3$ degrees of freedom and it makes sense to take $m-3$ as a weight factor.

In Options 2, 3 and 4, the number of contributing partitions and the sum of their weight factors per tree depend on the tree type. An example of the asymmetry values for trees of degree 7, calculated according to the four options, is given in Table 1. Option 1, including the partitions (1,1) and (1,2), lets the tree asymmetry values range from a minimum of 0.2 up to a maximum of 0.833. The most asymmetric tree of degree 7 does not get an asymmetry value of 1 because the (1,1) partition has arbitrarily been defined as a symmetric one (i.e. has an asymmetry value of 0). For the most asymmetric trees of larger degree the effect of the single (1,1) partition will become negligible with respect to the other (1, $m - 1$) partitions and their asymmetry values will approach the value 1. Options 2, 3 and 4 exclude the (1,1) and (1,2) partitions and the most asymmetric tree indeed has an asymmetry value of 1. The most symmetric tree does not have an asymmetry of zero in all four options. Actually, a tree only has an asymmetry value of zero if all partitions in the tree are symmetric, which can only occur if the degree of the tree is a power of 2. The coefficients of variation (cv) in asymmetry values appear to differ only slightly between the four options, indicating an equal sensitivity of the measures for topological differences. This finding also holds for other degrees. The four different measures for tree asymmetry result in different rankings of trees, indicating that asymmetry is a relative concept specified only by the choice of the measure. The four options also result in measures with different ability to distinguish tree types. The number of unique values per degree is given in Table 2 and appears to be low only for Option 3. None of the options result in a measure that distinguishes all the tree types, as already has been shown for Option 2 by Uylings *et al.* (1989a) and Van Pelt *et al.* (1989a). In conclusion, Options 1, 2 and 4 result in a measure that is sensitive of topological differences and may therefore be used. A final choice on practical grounds will be made for Option 1, as is discussed in the next section.

2.3. *Mean tree asymmetry in a set of trees.* A set of trees can now be described by the distribution of the tree-asymmetry values, for instance, by the mean and variance. In averaging the tree-asymmetry values, however, one should also consider the use of weight factors. For instance, we may use the tree weight factor from the calculation of the tree asymmetry (i.e. the sum of the partition weights). In doing so, one obtains effectively a mean asymmetry value of all the partitions in all the trees:

$$\bar{A}_t(\text{set}) = \frac{\sum_{i=1}^N w(\alpha_i^{n_i}) A_t(\alpha_i^{n_i})}{\sum_{i=1}^N w(\alpha_i^{n_i})} = \frac{\sum_{i=1}^N \sum_{j=1}^{n_i-1} w_{ij} A_p(r_{ij}, s_{ij})}{\sum_{i=1}^N w(\alpha_i^{n_i})} \tag{6}$$

Table 1. Decomposition of trees into their partitions

Branching codes	Partition frequencies										Tree asymmetry (four options)			
	Degree 2 Partition 1,1 Partition asymmetry 0	3 1,2 1	4 1,3 1	4 2,2 0	5 1,4 1	5 2,3 0,33	6 2,4 0,5	6 3,3 0	7 2,5 0,6	7 3,4 0,2	1 $m > 1$ $w = 1$	2 $m > 3$ $w = 1$	3 $m > 3$ $w = m - 2$	4 $m > 3$ $w = m - 3$
7(1 6(1 5(1 4(1 3(1 2(1 1))))))	1	1	1	1	1	1	1	1	1	0.833	1.000	1.000	1.000	
7(1 6(1 5(1 4(2(1 1) 2(1 1))))	2	1	2	1	1	1	1	1	1	0.500	0.750	0.857	0.900	
7(1 6(1 5(2(1 1) 3(1 2(1 1))))	2	1	1	1	1	1	1	1	1	0.556	0.778	0.833	0.852	
7(1 6(2(1 1) 4(1 3(1 2(1 1))))	2	1	1	1	1	1	1	1	1	0.583	0.833	0.818	0.813	
7(1 6(2(1 1) 4(2(1 1) 2(1 1))))	3	1	1	1	1	1	1	1	1	0.250	0.500	0.636	0.688	
7(1 6(3(1 2(1 1) 3(1 2(1 1))))	2	2	1	1	1	1	1	1	1	0.500	0.500	0.556	0.571	
7(2(1 1) 5(1 4(1 3(1 2(1 1))))	2	1	1	1	1	1	1	1	1	0.600	0.867	0.800	0.771	
7(2(1 1) 5(1 4(2(1 1) 2(1 1))))	3	1	1	1	1	1	1	1	1	0.267	0.533	0.600	0.629	
7(2(1 1) 5(2(1 1) 3(1 2(1 1))))	3	1	1	1	1	1	1	1	1	0.322	0.467	0.500	0.511	
7(3(1 2(1 1) 4(1 3(1 2(1 1))))	2	2	1	1	1	1	1	1	1	0.533	0.600	0.429	0.360	
7(3(1 2(1 1) 4(2(1 1)2(1 1)))	3	1	1	1	1	1	1	1	1	0.200	0.100	0.143	0.160	
										Mean	0.468	0.630	0.659	
										SD	0.181	0.238	0.237	
										cv	0.387	0.378	0.360	

m = degree of included partitions; w = weight per partition.

Table 2. Number of unique tree-asymmetry values per degree

Degree	Option Number of tree types	1	2	3	4
		$m > 1$ $w = 1$	$m > 3$ $w = 1$	$m > 3$ $w = m - 2$	$m > 3$ $w = m - 3$
4	2	2	2	2	2
5	3	3	3	3	3
6	6	5	5	6	6
7	11	10	10	11	11
8	23	18	18	20	23
9	46	38	40	32	44
10	98	70	78	48	93
11	207	145	154	75	192
12	451	266	298	108	409
13	983	608	690	150	858
14	2 179	950	1 105	207	1 782
15	4 850	2 376	2 820	275	3 882
16	10 905	4 339	5 300	369	8 243
17	24 631	7 569	9 185	470	17 202
18	56 011	15 660	19 666	597	37 031
19	127 912	36 904	47 692	744	81 760

m = degree of included partitions; w = weight per partition.

for a set of trees $\{\alpha_i^{n_i}, i = 1, \dots, N\}$ with n_i the degree of tree α_i . Then, it makes no difference if the mean asymmetry in a set of trees is obtained by immediately averaging all the partition asymmetries in the set or doing it in two steps via the individual trees. For practical purposes, however, it is not convenient to work with a complicated scheme of weight factors. Weight factors can be eliminated if all trees have equal weights. This is assured if: (1) the partitions in the tree have equal weight; (2) an equal number of partitions per tree contribute in the calculation of the tree asymmetry; and (3) the mean tree asymmetry is calculated for trees of the same degree. The first two requirements are only realized in Option 1 for the definition of tree asymmetry and equation (6) reduces to:

$$\bar{A}_t(\text{set} | n) = \frac{1}{M} \sum_{i=1}^M A_t\{\alpha_i^n\} = \frac{1}{M(n-1)} \sum_{i=1}^M \sum_{j=1}^{n-1} A_p(r_{ij}, s_{ij}) \tag{7}$$

for a set of M trees of degree n . Note, that this choice differs from the one made in the studies by Verwer and Van Pelt (1986), Uylings *et al.* (1989a) and Van Pelt *et al.* (1989a) in which Option 2 was used. The elimination of weight factors for trees of equal size also permits the formulation of a simple relation between the expectation of tree asymmetry and partition asymmetry as is shown in the next sections. No choice has yet been made for weighing trees of

different size. To avoid any use of weight factors in practice we use equal weights and obtain finally:

$$\bar{A}_t(\text{set}) = \frac{1}{N} \sum_{i=1}^N A_t\{\alpha_i^{n_i}\}, \quad \text{with } A_t\{\alpha_i^{n_i}\} \text{ given by equation (5).} \quad (8)$$

3. Growth Models, Expectations and Variances. The tree asymmetry expectation and variance is determined by the probabilities of the tree types. These probabilities have been obtained for the so called *QS* growth model (Van Pelt and Verwer, 1986). This model assumes that a tree grows via a sequence of branching events, one at a time. During a branching event one segment is divided by a branching point from which a new terminal segment protrudes. A branching event may occur at any segment in a tree with a certain probability. The branching probability of a terminal segment has been defined as $p_t = C2^{-S\gamma}$ and of an intermediate segment as $p_i = Rp_t$. The centrifugal order of the segment is denoted by γ , the parameter S defines the strength of the order dependency and the parameter R defines the ratio of branching probabilities of an intermediate and a terminal segment of equal order. The parameter $Q = R/(R+1) = p_i/(p_i + p_t)$ with value domain $[0,1]$ is preferably used rather than R which may attain values in $[0,\infty]$. The normalization constant C must be chosen such that the branching probabilities of all segments sum to one. The model describes random terminal growth if only terminal segments are allowed to branch ($Q=0$) without order-dependency ($S=0$). The model describes random segmental growth if both terminal and intermediate segments are allowed to branch with equal probability without order-dependency ($S=0$ and $Q=0.5$). For a given model, each tree type will have a particular probability of occurrence and for each size of the trees (i.e. after a given number of branching events) a probability distribution of tree types can be obtained. Essential in the definition of the growth rules is that the development of a subtree is independent of its position in its parent tree. Any subtree pair will also develop independently of other subtree pairs. We may therefore consider a partition as a random variable and derive for any degree a partition probability distribution which describe the statistics of the partitions anywhere in the tree. For instance, the partition probability $p(r, m-r)$ denotes the probability that a partition of degree m is of the type $(r, m-r)$. These partition probability distributions are functions of the two free parameters Q and S . Additionally, the probability of occurrence of a tree type is simply the product of the probabilities of occurrence of the partitions in the tree and a factor of two for each occurrence of an equal-degree unequal-type subtree pair (Van Pelt and Verwer, 1985). Consequently, it is sufficient to know the partition probabilities for the calculation of tree-type probabilities which is advantageous because the

number of possible partitions is much smaller than the number of possible tree types (Section 2). Therefore, the partition probability distributions, calculated on the basis of the QS growth model, form the basis for the calculation of the tree-type probability distributions. In the following we implicitly refer to the QS model in the calculation of the probabilities and expectations.

3.1. *Partition-asymmetry expectations.* The expectation of the partition asymmetry $E\{A_p^m | Q, S\}$, or in short $E\{A_p^m\}$, of a partition of degree m can now be defined as:

$$E\{A_p^m\} = \sum_{r=1}^{[m/2]} p(r, m-r) A_p(r, m-r) = \sum_{r=1}^{[m/2]} p(r, m-r) \frac{m-2r}{m-2}$$

if $m > 2$. (9)

For the trivial partitions (1,1) and (1,2) we have $p(1,1)=1$, $A_p(1,1)=0$ and $p(1,2)=1$ such that:

$$E\{A_p^2\}=0 \quad \text{and} \quad E\{A_p^3\}=1.$$

3.1.1. *QS growth model.* In general, the partition probabilities are not available in closed form but have to be obtained by means of two coupled recursive equations (Van Pelt and Verwer, 1986). For several values of the model parameters Q and S , the expectations of the partition asymmetry have been calculated and are plotted vs degree in Fig. 1. The curves in Fig. 1 show

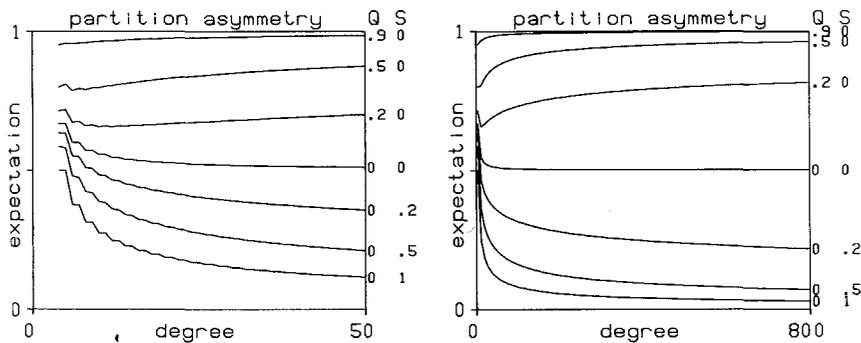


Figure 1. Expected values of the partition asymmetry, calculated for several growth models and plotted vs the degree in the range (4, 50) in the left figure and in the range (4, 800) in the right figure. Each curve is labelled by the values of the model parameters Q and S .

that the expected partition asymmetry strongly depends on the mode of growth. Asymmetric partitions are highly probable for Q -values near one, i.e. if branching occurs predominantly at intermediate segments. Symmetric partitions are highly probable for positive S -values and $Q=0$, i.e. if branching occurs predominantly at low-order terminal segments. The expected values change with the degree in a stepwise pattern, which is most prominent at small degrees. In the Section 3.1.2 it is shown that this pattern is caused by the number of possible partitions per degree $[m/2]$, that is equal for a successive even-odd pair of degrees. In the right figure it is shown that the expected values for the random terminal growth model $(Q, S) = (0, 0)$ stabilize rapidly at a value of 0.5. For all other growth modes the expected values appear to increase or decrease monotonously over the studied range of degrees.

3.1.2. Random terminal growth model. An analytical result for the partition asymmetry expectation can only be obtained for the random terminal growth model $(Q, S) = (0, 0)$ for which we have:

$$p(r, m-r) = (2 - \delta(r, m-r))/(m-1), \quad \text{with } r \leq m \quad (10)$$

(Van Pelt and Verwer, 1983). The Kronecker delta $\delta(r, m-r)$ equals zero if $r \neq m-r$ and equals one if $r = m-r$. Inserting this equation into (9) we obtain:

$$\begin{aligned} E\{A_p^m\} &= \frac{1}{(m-1)(m-2)} \sum_{r=1}^{[m/2]} (m-2r)(2 - \delta(r, m-r)) \\ &= \frac{2}{(m-1)(m-2)} \sum_{r=1}^{[m/2]} (m-2r) = \frac{1}{2 - 1/[m/2]} = \frac{m_e}{2(m_e - 1)} \quad (11) \end{aligned}$$

for $m > 2$ and $E\{A_p^2\} = 0$. The value m_e is equal to m for the even-odd number pair m and $m+1$, explaining the stepwise pattern in $E\{A_p^m\}$ in Fig. 1. Equation (11) has as limit:

$$\lim_{m \rightarrow \infty} E\{A_p^m\} = 1/2, \quad (12)$$

proving that for random terminal growth the expected partition asymmetry goes to 0.5 for large trees.

3.1.3. Random segmental growth model. The partition probabilities for the random segmental growth model $(Q, S) = (0.5, 0)$ are described by:

$$p(r, m-r) = (2 - \delta(r, m-r)) \frac{N_\tau^r N_\tau^{m-r}}{N_\tau^m}. \quad (13)$$

(Van Pelt and Verwer, 1983). For the expectation $E\{A_p^m\}$ we obtain:

$$E\{A_p^m\} = \sum_{r=1}^{\lfloor m/2 \rfloor} (2 - \delta(r, m-r)) \frac{N_\tau^r N_\tau^{m-r}}{N_\tau^m} \frac{m-2r}{m-2}. \tag{14}$$

Neither an analytical expression nor a limit expression for $m \rightarrow \infty$ could be obtained for this sum series.

3.2. *Tree-asymmetry expectations.* The expectation $E\{A_t^n | Q, S\}$ of the asymmetry of a tree of degree n for a particular growth mode (Q, S) , or in short $E\{A_t^n\}$, is defined as:

$$E\{A_t^n\} = \sum_{i=1}^{N_\alpha^n} p(\alpha_i^n) A_t(\alpha_i^n) \tag{15}$$

if N_α^n denotes the number of topological tree types of degree n in the set $\{\alpha_i^n | i = 1, \dots, N_\alpha^n\}$ and $p(\alpha_i^n)$ is the probability of occurrence of tree type α_i^n in this set. Inserting equation (5) into (15) gives:

$$E\{A_t^n\} = \sum_{i=1}^{N_\alpha^n} p(\alpha_i^n) \frac{1}{n-1} \sum_{j=1}^{n-1} A_p(r_{ij}, s_{ij}). \tag{16}$$

In this formula we have to sum over all partitions in all trees of the set of degree n . All partitions in a tree will have asymmetry expectations as given by equation (9) and shown in Fig. 1.

As mentioned in Section 3, the probability of a tree for a particular growth mode is equal to the product of the probabilities of all its partitions, including correction factors for equal-degree unequal-type subtree pairs (Van Pelt and Verwer, 1985). The probability of the partitions in a tree, however, highly depends on the mode of growth. For instance, the model with $(Q, S) = (0.8, 0)$ results in a high probability of asymmetric partitions whereas the model with $(Q, S) = (0, 1)$ results in a high probability of symmetric partitions. Consequently, an asymmetric tree will occur with a high probability under the $(0.8, 0)$ growth model but with a low probability under the $(0, 1)$ model. Therefore, the asymmetry of a tree is a property that is indicative for the probability of occurrence of this tree for a particular QS growth mode.

For the calculation of $E\{A_t^n\}$ via equation (16) one has to sum over all the N_α^n tree types of degree n , being an unfeasibly large number even for medium-sized trees. The tree-asymmetry expectation $E\{A_t^n\}$, however, can also be expressed in terms of the partition-asymmetry expectations $E\{A_p^m\}$ via:

$$E\{A_t^n\} = \frac{2n-1}{n-1} \sum_{m=2}^n p(m|n) E\{A_p^m\}, \tag{17}$$

reducing the summation to only $n-1$ terms. The function $p(m|n)$ denotes the probability that a uniformly random chosen segment in a tree of degree n , at random chosen according to the tree-type probability distribution, is of degree m . The derivation of this equation is given in Appendix A.

3.2.1. *QS* growth model. The probability function $p(m|n)$ is generally expressed by the recursive equation:

$$p(m|n) = \frac{1}{2n-1} \sum_{r=m}^{n-1} p(r, n-r) (2r-1) (1 + \delta(r, n-r)) p(m|r) \quad \text{for } m < n, \quad (18)$$

with $p(1|n) = n/(2n-1)$ and $p(n|n) = 1/(2n-1)$ (Van Pelt and Verwer, 1984). The partition probabilities, required to solve equation (18) and to calculate the partition-asymmetry expectations are also only expressed in recursive equations, as mentioned in Section 3.1.1. Therefore, the tree-asymmetry expectations cannot generally be obtained in closed form. Only for the particular growth modes of random terminal growth $(Q, S) = (0, 0)$ and random segmental growth $(Q, S) = (0.5, 0)$ is the function $p(m|n)$ explicitly known.

3.2.2. Random terminal growth model. The probability function $p(m|n)$ for this growth model is equal to:

$$p(m|n) = \frac{2n}{m(m+1)(2n-1)} \quad \text{for } m < n \quad \text{and} \quad p(n|n) = 1/(2n-1), \quad (19)$$

(Van Pelt and Verwer, 1984). Inserting equation (19) and the expected partition asymmetry as calculated in equation (11) into equation (17) results in:

$$E\{A_t^n\} = \frac{n}{n-1} \left\{ \frac{2-n_e/n}{2(n_e-1)} - \frac{1}{3} + \sum_{k=1}^{n_e/2-1} \frac{1}{(k+1)(2k-1)} \right\}, \quad (20)$$

as is shown in Appendix B. The general outcome for a finite series of the type in this equation can be expressed in terms of the Psi or Digamma function Ψ by:

$$\sum_{k=1}^m \frac{1}{k(kx+y)} = \frac{1}{y} \{ \Psi(m+1) - \Psi(1) + \Psi(y/x+1) - \Psi(y/x+m+1) \}, \quad (21)$$

(Hansen, 1975; Section 6.1.7). This equation can be reformulated by taking $x=2$, $y=-3$, $\Psi(-0.5) - \Psi(1) = 2(1 - \ln 2)$ and using the relations $\Psi(m+0.5) = 2\Psi(2m) - \Psi(m) - 2 \ln 2$ and $\Psi(m+2) = \Psi(m) + 1/m + 1/(m+1)$ (Abramowitz and Stegun, 1965) into:

$$\sum_{k=1}^m \frac{1}{(k+1)(2k-1)} = \frac{1}{3} \left\{ 1 - \frac{1}{m} - \frac{1}{m+1} + 2(\Psi(2m) + \Psi(m)) \right\}, \tag{22}$$

or, using the definition for the Digamma or Psi function Ψ into:

$$\sum_{k=1}^m \frac{1}{(k+1)(2k-1)} = \frac{1}{3} \left\{ 1 - \frac{1}{m} - \frac{1}{m+1} + 2 \sum_{k=m}^{2m-1} (1/k) \right\}. \tag{23}$$

Replacing m by $n_e/2 - 1$ and inserting equation (23) into equation (20) we obtain

$$E\{A_t^n\} = \frac{2n}{3(n-1)} \left\{ \frac{2-3n_e/n}{4(n_e-1)} - \frac{2}{n_e} + \sum_{k=n_e/2}^{n_e} (1/k) \right\}. \tag{24}$$

No further reduction of this equation could be obtained. For large n , however, equation (20) reduces to:

$$\lim_{n \rightarrow \infty} E\{A_t^n\} = -\frac{1}{3} + \sum_{k=1}^{\infty} \frac{1}{(k+1)(2k-1)} = \frac{2}{3} \ln 2 = 0.4621, \tag{25}$$

(Hansen, 1975; Section 5.5.24), proving that the tree-asymmetry expectation for the random terminal growth model goes to a stable value for large trees.

3.2.3. Random segmental growth model. The probability function $p(m|n)$ for this growth mode is equal to:

$$p(m|n) = 2 \frac{2(n-m)-1}{2n-1} \frac{N_\tau^m N_\tau^{n-m}}{N_\tau^n}, \tag{26}$$

(Van Pelt and Verwer, 1984). The expected tree asymmetry can be formulated using equations (14), (17) and (26) resulting in:

$$E\{A_t^n\} = \frac{2}{(n-1)N_\tau^n} \sum_{m=2}^n \frac{2(n-m)-1}{m-2} N_\tau^{n-m} \sum_{r=1}^{[m/2]} \{2 - \delta(r, m-r)\} N_\tau^r N_\tau^{m-r} (m-2r). \tag{27}$$

It was, unfortunately, not possible to study from this equation, analytically, the limit behavior of the expectation.

3.2.4. Tree-asymmetry means and variances in sets of randomly sampled trees. Approximate values for tree-asymmetry expectations can be obtained from random samples of trees. A random tree of degree n on the basis of the QS

growth model can be obtained using the partition probability distributions in the following way. Take a random partition of degree n according to the partition probability distribution of degree n , resulting in, say, $n(r, s)$. Subsequently, take random partitions of degree r and degree s according to their partition probability distributions, resulting in the partitions, say, $r(t, u)$ and $s(v, w)$. Repeating this sequence generates finally a tree $n(r(t(\dots)u(\dots))s(v(\dots)w(\dots)))$, that has been sampled with the correct probability for this growth model. This probability in its turn can be calculated as described in Section 3. Such a sampling procedure also gives an estimate of the variances in the tree-asymmetry values. For several growth modes we have created groups of 100 trees per degree and plotted the mean tree asymmetry and the 1σ standard deviation interval per group for degree 4 up to 800 (Fig. 2).

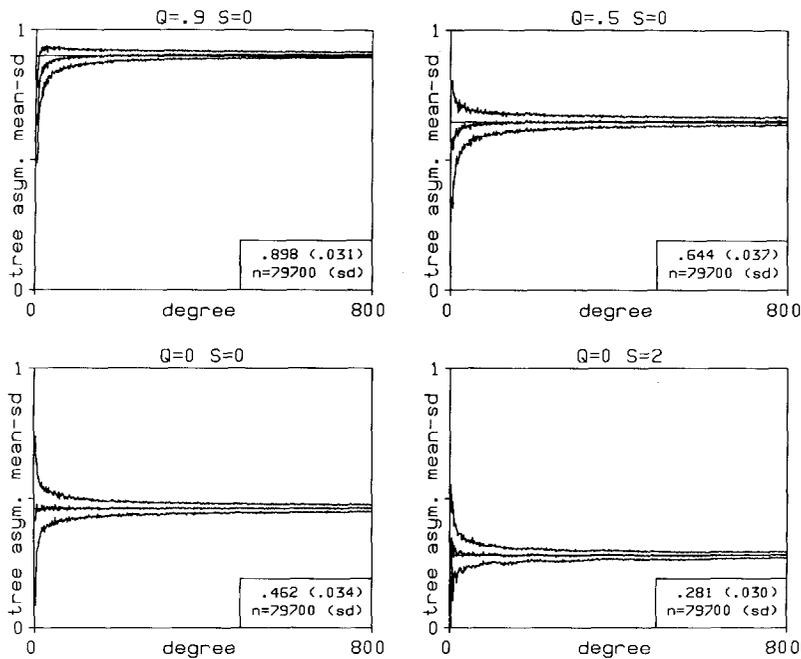


Figure 2. Mean and standard deviation of tree-asymmetry values, plotted vs degree in the range (4, 800). The data are obtained by generating per degree a group of 100 trees for four different growth models (see figures). The grand mean over the 79 700 tree-asymmetry values per figure is plotted as a horizontal line and printed in the small box together with the grand SD value.

As for the expected partition asymmetries, also the expected tree asymmetries appear to depend strongly on the mode of growth. It appears, however, that the expected tree asymmetries are very stable for the whole range of degrees for any growth mode. This is in contrast with the partition asymmetries whose

expectations were monotonously increasing or decreasing over the studied range of degrees, except for the random terminal growth mode. The variance in tree asymmetries is greatest for small trees but becomes readily smaller for increasing tree size and quickly tends to a rather stable value. In Table 3 are listed the grand mean and grand standard deviation (SD) of the tree-asymmetry values for four sets of trees of different ranges of degrees, with a group of 100 trees per degree. The first set consists of 79 700 trees in the range of degrees 4–800, the second set has 70 100 trees in the range 100–800, the third set has 9 700 trees in the range 4–100 and the fourth has 1 200 trees in the range 4–15. Both the SD per group (i.e. per degree) and the SD in the group means contribute to the grand SD. The within-group SD is equal to the square root of the mean of all the group variances and represents the intrinsic variance caused by the random choices during growth of the trees. The between-group standard deviation indicates the spread of the group means around the grand mean and is a combination of the standard deviation in the mean per group and the degree dependency of the expected asymmetry value. For all four degree ranges it can be concluded that the within-group SD plays a dominant role in the grand SD. In fact, the between-group SD appears to be negligibly small in most cases, even if the restriction is made for small trees as in the 4–100 and in the 4–15 degree set. The final conclusion is that, given the intrinsic SD, the tree-asymmetry expectations can be considered as independent of the degree. The standard deviation in the mean per group can be estimated for the set of trees in the degree range 100–800 to be 0.002 [= standard deviation (0.02) divided by $\sqrt{100}$] and appears to be equal to the between-group SD. That means that for such large trees the degree-dependency in the asymmetry expectation is approximately zero. An analytical proof for the asymptotic stability of the tree-asymmetry expectation has only been given explicitly for the random terminal growth model (see Section 3.2.2). The random tree sampling procedure allows us to calculate for each point in the Q - S parameter space the expected tree asymmetry. Connecting points of equal expectation results in a family of isoclines. The size-insensitivity of the tree-asymmetry expectations implies that these isoclines are also highly insensitive for the degree. This is shown in Fig. 3 in which the tree-asymmetry expectations are calculated for degree 50, 200 and 800.

4. Test of Observed Data Sets. In a first example to illustrate the usefulness of the proposed measure, a sample of rat Purkinje cell dendritic trees (from R. Pentney) is used. The data set consists in total of 82 trees taken from four age classes (i.e. 1, 10, 18 and 28 months). A plot of the asymmetry of every tree vs its degree is given in Fig. 4a. The mean asymmetry of all the trees, calculated according to equation (8), is given by \bar{A} (SD) = 0.494 (0.020). The best-fitting Q - S model, predicting tree sets with the same mean asymmetry, was found for

Table 3. Mean asymmetries in simulated tree sets

Model <i>Q</i>	Degree range 4–800 (797 groups)						Degree range 100–800 (701 groups)						Degree range 4–15 (12 groups)						Degree range 4–100 (97 groups)							
	Within group			Between group			Within group			Between group			Within group			Between group			Within group			Between group				
	Grand mean	SD	SD	Grand mean	SD	SD	Grand mean	SD	SD	Grand mean	SD	SD	Grand mean	SD	SD	Grand mean	SD	SD	Grand mean	SD	SD	Grand mean	SD	SD		
0.9	0	0.898	0.031	0.025	0.018	0.003	0.902	0.016	0.016	0.003	0.785	0.128	0.112	0.062	0.869	0.071	0.059	0.869	0.071	0.059	0.869	0.071	0.059	0.869	0.071	0.059
0.8	0	0.819	0.035	0.031	0.016	0.004	0.823	0.020	0.020	0.004	0.722	0.146	0.140	0.041	0.791	0.079	0.073	0.791	0.079	0.073	0.791	0.079	0.073	0.791	0.079	0.073
0.7	0	0.753	0.036	0.033	0.013	0.004	0.756	0.022	0.021	0.004	0.670	0.151	0.149	0.028	0.727	0.081	0.077	0.727	0.081	0.077	0.727	0.081	0.077	0.727	0.081	0.077
0.6	0	0.695	0.036	0.034	0.012	0.004	0.698	0.023	0.022	0.004	0.630	0.156	0.154	0.021	0.672	0.081	0.079	0.672	0.081	0.079	0.672	0.081	0.079	0.672	0.081	0.079
0.5	0	0.644	0.037	0.035	0.010	0.003	0.647	0.023	0.022	0.003	0.586	0.164	0.163	0.023	0.625	0.084	0.082	0.625	0.084	0.082	0.625	0.084	0.082	0.625	0.084	0.082
0.4	0	0.600	0.036	0.035	0.008	0.003	0.602	0.023	0.022	0.003	0.553	0.164	0.163	0.015	0.585	0.083	0.081	0.585	0.083	0.081	0.585	0.083	0.081	0.585	0.083	0.081
0.3	0	0.560	0.036	0.035	0.007	0.003	0.561	0.022	0.022	0.003	0.527	0.167	0.166	0.017	0.548	0.082	0.081	0.548	0.082	0.081	0.548	0.082	0.081	0.548	0.082	0.081
0.2	0	0.524	0.035	0.035	0.005	0.002	0.525	0.022	0.022	0.002	0.501	0.169	0.168	0.016	0.517	0.082	0.081	0.517	0.082	0.081	0.517	0.082	0.081	0.517	0.082	0.081
0.1	0	0.492	0.034	0.034	0.004	0.002	0.492	0.021	0.021	0.002	0.476	0.164	0.163	0.014	0.488	0.080	0.080	0.488	0.080	0.080	0.488	0.080	0.080	0.488	0.080	0.080
0	0	0.462	0.034	0.034	0.003	0.002	0.462	0.021	0.021	0.002	0.453	0.168	0.168	0.016	0.460	0.080	0.080	0.460	0.080	0.080	0.460	0.080	0.080	0.460	0.080	0.080
0	0.2	0.426	0.033	0.033	0.005	0.002	0.426	0.021	0.020	0.002	0.433	0.159	0.157	0.027	0.429	0.077	0.076	0.429	0.077	0.076	0.429	0.077	0.076	0.429	0.077	0.076
0	0.4	0.397	0.033	0.033	0.004	0.002	0.397	0.020	0.020	0.002	0.415	0.159	0.159	0.013	0.403	0.077	0.077	0.403	0.077	0.077	0.403	0.077	0.077	0.403	0.077	0.077
0	0.6	0.374	0.032	0.032	0.004	0.002	0.373	0.020	0.020	0.002	0.385	0.159	0.158	0.015	0.379	0.076	0.076	0.379	0.076	0.076	0.379	0.076	0.076	0.379	0.076	0.076
0	0.8	0.354	0.032	0.032	0.004	0.002	0.353	0.020	0.019	0.002	0.372	0.159	0.158	0.018	0.359	0.076	0.075	0.359	0.076	0.075	0.359	0.076	0.075	0.359	0.076	0.075
0	1	0.337	0.032	0.031	0.004	0.002	0.337	0.019	0.019	0.002	0.350	0.155	0.154	0.021	0.343	0.074	0.074	0.343	0.074	0.074	0.343	0.074	0.074	0.343	0.074	0.074
0	2	0.281	0.030	0.030	0.005	0.002	0.280	0.018	0.018	0.002	0.301	0.149	0.147	0.029	0.287	0.072	0.071	0.287	0.072	0.071	0.287	0.072	0.071	0.287	0.072	0.071

100 simulated trees per group.

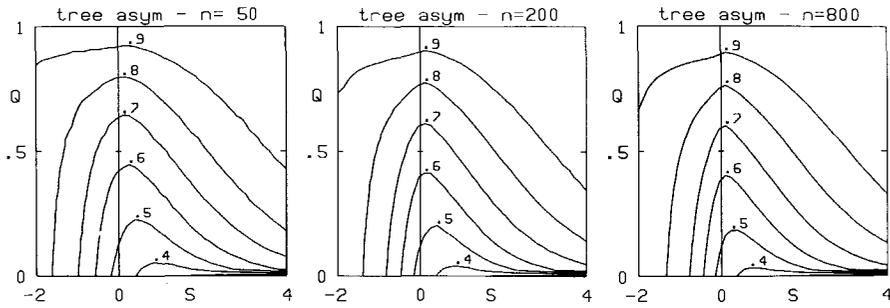


Figure 3. Families of isoclines, connecting points in the Q - S parameter space which give rise to equal tree-asymmetry expectations (see label at each isocline). Each figure is obtained by calculating for each point of a 41×41 grid on the parameter space the mean of the asymmetry values of 1000 (for degree $n = 50$) or 500 (for degree $n = 200$ and $n = 800$) randomly sampled trees of the indicated size. The three figures show that the isocline family is highly stable for the different degrees.

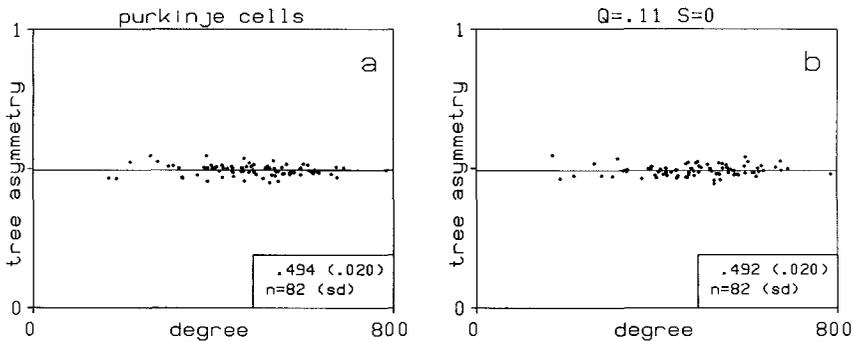


Figure 4. Plot of tree-asymmetry values vs degree for: (a) a set of 82 rat Purkinje cell dendritic trees provided by R. Pentney and M. J. Woldenberg; and (b) a set of simulated trees with the same degree distribution for the best-fitting growth model $Q = 0.11$ and $S = 0$. The observed and the simulated tree set appear to have equal standard deviations.

the model parameters $Q = 0.11$ and $S = 0$, resulting in a reduced chi-square value of 1.07. In searching the best-fitting model, we have restricted ourselves to a subspace in the Q - S parameter plane, i.e. the Q -axis ($S = 0$) and the positive S -axis ($Q = 0$). In previous studies it has been shown that this subspace gives rise to a full range of tree-type probability distributions, while only a minor improvement in the fit could be obtained outside this subspace (Van Pelt *et al.*, 1986;1989b). Also Fig. 3 shows that on this subspace the tree-asymmetry expectation can adopt all values in $(0,1)$. With the best-fitting model a

sample of random trees has subsequently been generated with the same degree distribution as observed. The asymmetry values of these random trees are displayed in Fig. 4b. The similarity between the asymmetry values of the observed and the randomly generated trees is striking. The correspondence between the mean values was expected because it was optimized for the best-fitting model. Especially the correspondence in the variance of the asymmetry values around the mean value and in the stability of the asymmetry values for different degrees is of major importance. Both the observed and the simulated data set have a standard deviation of 0.02. From these results it can be concluded that: (1) the variance in the data set is completely explained by the *QS* growth model; and (2) also in this set of natural trees the asymmetry measure does not show any degree dependency.

To test the consistency of this outcome we have also analysed this data set using the mean-order measure and have used the procedure described in Van Pelt *et al.* (1989b). The best-fitting *QS* model, predicting optimally the observed mean-order vs degree relationship, was found for the parameters $Q=0.11$ and $S=0$ and a reduced chi-square value of 0.91. Thus both procedures give consistent results.

In a second example a sample of 443 pyramidal (pyr) and a sample of 490 multipolar non-pyramidal (mpnp) cell dendritic trees from the rat is used, from P. McConnell and H. B. M. Uylings (Uylings *et al.*, 1989b). For both sets the averaged asymmetry values and standard deviations (SD's) per degree are plotted vs degree in Fig. 5a and b. The models that fit the asymmetry data best

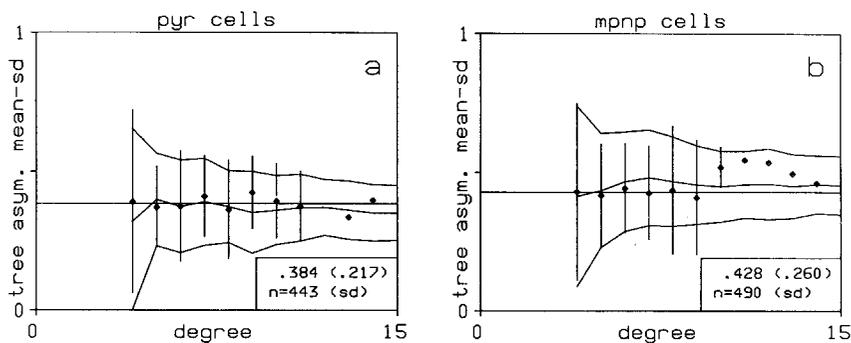


Figure 5. Plot of averaged tree-asymmetry values and SD's vs degree for: (a) a set of 443 pyramidal (pyr) cells; and (b) a set of 490 multipolar non-pyramidal (mpnp) cells provided by P. McConnell and H. B. M. Uylings. In the same figures are plotted the expected tree asymmetries and SD's per degree for the best-fitting growth models with parameter values $Q=0$ and $S=0.87$ for the pyr set and $Q=0$ and $S=0$ for the mpnp set (continuous curves). The correspondence between observed and expected SD values is very good.

were found for the parameters $Q=0, S=0.87$ for the pyr cells and $Q=0, S=0$ for the mpnp cells. In the same figures we have also plotted the expected means and SD intervals for these models. The figures show that the expected means and SD's correspond nicely to the observed ones. Additionally, we have simulated tree sets with the same degree distribution as observed and calculated the means and SD's which are given in Table 4. The grand SD's of the simulated sets appear to be in perfect agreement with those of the observed ones.

Table 4. Mean asymmetries (SD) in observed and simulated tree sets

pyr cells ($n=443$)		mpnp cells ($n=490$)	
Observed	Simulated	Observed	Simulated
	$Q=0, S=0.87$ (10 sets)		$Q=0, S=0$ (10 sets)
0.38 (0.22)	0.37 (0.21)	0.43 (0.26)	0.45 (0.26)

5. Discussion. A tree-asymmetry measure for the topological structure of a binary tree is proposed as the mean of its partition asymmetries. This measure uses much of the structural information in a topological tree because it is able to distinguish, (i.e. attains unique values for) most of the different tree types in a set. In comparative studies of measures of tree topology (Uylings *et al.*, 1989a; Van Pelt *et al.*, 1989a) this aspect has already been quantified. The conclusion was that none of the evaluated measures was able to distinguish all the tree types, but that the tree-asymmetry measure has highest discriminative power. The present study has demonstrated that the outcomes of this measure have also very attractive statistical properties for the QS growth model that we used. Firstly, the expected value depends strongly on the mode of growth. Secondly, the expected value is almost independent of the size of the tree. Thirdly, the intrinsic variance, caused by the random choices of the branching segments during growth, decreases rapidly with increasing size of the trees and already becomes stable and small for medium-sized trees. These properties are important for practical purposes because: (1) a set of trees can be characterized by one single number (i.e. the mean tree asymmetry), irrespective of the sizes of the trees; (2) this single number is probably highly sensitive for the developmental history of the tree set; (3) different sets of trees can easily be compared statistically without reference to models with their implicit assumptions; (4) all the calculations are simple and the analysis does not require heavy computer support. When a data set consists of a group of very small and a group of large trees it is expected that the variance in the mean tree asymmetry will be dominated by those in the small trees. In comparing two of

such sets it may be worthwhile to split up the data sets in two size classes and do the comparison per size class.

Model studies remain indispensable to interpret the mean and variance in observed tree asymmetries. Especially, if the observed variance deviates from the model predictions it may suggest that additional constraints or sources of variance were operational during development.

Some critical remarks can be made. One may argue that the statistical properties arise under the used *QS* growth model while it is uncertain if they also arise under natural conditions. This remark is certainly true but the outcomes of the analysis of observed neuronal dendritic trees have been shown to agree with the *QS* growth model findings, especially with respect to the stability of the expected tree asymmetries and the predicted variances. Apparently, the assumptions made in the model are sufficient to explain all the variance in the data set. Nevertheless, it is meaningful to also study the tree-asymmetry measure under alternative growth rules. An alternative would be the random synchronous growth model, proposed by Horsfield *et al.* (1987), in which none, one, or more than one branching events at a time step are allowed. This is in contrast with the *QS* model which describes growth essentially as a sequence of single-branching events (sequential growth model).

The analysis on the basis of tree asymmetries has been shown to give consistent results with previous studies using partition distributions and the mean-order measure. The use of the mean-order measure requires a growth model to interpolate the observed mean order vs degree relationship. The partition analysis requires a growth model to relate the partition frequency distributions of different degree. Both approaches therefore require considerable computer calculations, while the group of observed trees is finally described by the best-fitting model parameters. The mean-order and the tree-asymmetry measure also differ in their variance properties. The SD in asymmetry values does not depend much on the mode of growth and is monotonously decreasing for increasing size of the tree, as is shown in Table 3. The SD in mean-order values, however, depends strongly on the mode of growth, as do the mean-order values themselves, and also increase with increasing tree size (Van Pelt *et al.*, 1989b). In conclusion, the tree-asymmetry measure has attractive properties in comparison with other measures of tree topology. Therefore, it may play a valuable and practical role in descriptive studies to characterize a set of trees or to quantify structural changes which may occur during development or in experimental conditions. It may also be used in the evaluation of observed variances in terms of hypothesized growth rules (like the *QS* model, used in this paper).

The partition asymmetry has been defined as a normalized dispersion measure for binary partitions. Natural trees, however, may contain "complex" branching areas where a parent segment seemingly gives rise to more than two

daughter segments. Observers may then find it difficult to interpret such an area as a sequence of bifurcations and conclude to have observed a multifurcation. For the data sets used in this paper, however, the number of “complex” branching areas that could not be resolved in successive bifurcations was negligibly small. Nevertheless, subjective criteria may play a role in these reconstructions and it is an interesting question whether the tree asymmetry measure is sensitive for possible “arbitrary” choices in resolving “complex” branching areas.

Alternatively, a multifurcation may be treated as an observed entity and a generalized dispersion measure for multifurcations, called PSAD (Verwer and Van Pelt, 1986) may be used. The PSAD measure is defined as the proportional sum of absolute deviations of the number of tree elements in the subtrees of the multifurcation. The statistical properties of the PSAD measure, however, can only be studied if explicit assumptions are made about the developmental process leading to such multifurcations (Verwer and Van Pelt, 1990). If it can be assumed that the multifurcations arise because of merging of consecutive bifurcation points in a purely binary tree (e.g. by shrinkage of segments), it is possible to analyse such trees in the context of the (binary tree) *QS* growth model (Verwer and Van Pelt, 1985; 1990). If it is assumed that the multifurcations are generated inherent in the growth process itself, which is different from the *QS* model, the statistical behaviour of the PSAD must be studied under the proper model conditions. All these questions are part of current research.

The authors thank M. van Veen for critical reading of the manuscript and G. van der Meulen for photography. Especially, we are grateful to one of the referees for the detailed and valuable comments. One of the authors (R.J.P.) gratefully acknowledges the support of the National Institute for Alcoholism and Alcohol Abuse (NIAAA), Grant AA05592.

APPENDIX A

Proof that equation (17) is equal to equation (15). According to equations (15) and (5), the expected tree asymmetry is defined as:

$$E\{A_1^n\} = \sum_{i=1}^{N_2^n} p(\alpha_i^n) \frac{1}{n-1} \sum_{j=1}^{n-1} A_p(r_{ij}, s_{ij}). \tag{A1}$$

The $n-1$ partitions (r_{ij}, s_{ij}) in tree α_i^n are counted in an arbitrary order $j=1, n-1$, but may very well be counted according to their degree $(m_{ij}=r_{ij}+s_{ij})$ as:

$$E\{A_1^n\} = \sum_{i=1}^{N_2^n} p(\alpha_i^n) \frac{1}{n-1} \sum_{j=1}^{n-1} \sum_{m=2}^n \sum_{r=1}^{[m/2]} A_p(r, m-r) \delta(m_{ij}, m) \delta(r_{ij}, r). \tag{A2}$$

The Kronecker deltas $\delta(m_{ij}, m)$ and $\delta(r_{ij}, r)$ ensure that the partition $(r, m-r)$ is counted only if $r_{ij}=r$ and $m_{ij}=m$. We may interchange the summations and obtain:

$$E\{A_t^n\} = \sum_{m=2}^n \sum_{r=1}^{[m/2]} A_p(r, m-r) \sum_{i=1}^{N_p^n} \frac{p(\alpha_i^n)^{n-1}}{n-1} \sum_{j=1}^{n-1} \delta(m_{ij}, m) \delta(r_{ij}, r), \tag{A3}$$

which can be written as:

$$E\{A_t^n\} = \sum_{m=2}^n \sum_{r=1}^{[m/2]} A_p(r, m-r) p(r, m-r|n) \tag{A4}$$

if $p(r, m-r|n)$ denotes the probability that a uniformly random chosen bifurcation in a tree of degree n , randomly chosen according to the *QS* model tree-type probability distribution, is of degree m and has a partition of the type $(r, m-r)$. Because the partition probability does not depend on the position of the bifurcation in the tree the probability $p(r, m-r|n)$ can be written as:

$$p(r, m-r|n) = p^*(m|n) p(r, m-r|m, n) \tag{A5}$$

if $p^*(m|n)$ denotes the probability that a randomly chosen partition (or intermediate segment) in a randomly chosen tree of degree n is of degree m , defined by:

$$p^*(m|n) = \sum_{i=1}^{N_p^n} \frac{p(\alpha_i^n)^{n-1}}{n-1} \sum_{j=1}^{n-1} \delta(m_{ij}, m), \tag{A6}$$

and if $p(r, m-r|m, n)$ denotes the probability that a partition of degree m in a randomly chosen tree of degree n is of the type $(r, m-r)$. Insertion in equation (A4) gives:

$$E\{A_t^n\} = \sum_{m=2}^n \sum_{r=1}^{[m/2]} A_p(r, m-r) p^*(m|n) p(r, m-r|m, n). \tag{A7}$$

A basic property of the *QS* model is that the partition probability $p(r, m-r|m, n)$ does not depend on the degree of the tree (Van Pelt and Verwer, 1986). Showing the conditional degree m of the partition by the notation $(r, m-r)$, we have used the shorthand notation $p(r, m-r)$ for the partition probability $p(r, m-r|m, n)$ throughout this and previous papers. Equation (A7) can now be written as:

$$E\{A_t^n\} = \sum_{m=2}^n p^*(m|n) \sum_{r=1}^{[m/2]} A_p(r, m-r) p(r, m-r)$$

and using equation (9) as:

$$E\{A_t^n\} = \sum_{m=2}^n p^*(m|n) E\{A_p^m\}. \tag{A8}$$

The probability $p^*(m|n)$ is related to the function $p(m|n)$ denoting the probability that a randomly chosen segment in a tree of degree n is of degree m via:

$$p^*(m|n) = \frac{2n-1}{n-1} p(m|n), \quad \text{with } m > 1, \tag{A9}$$

because a tree of degree n has $2n-1$ segments and $n-1$ intermediate segments (or partitions, or bifurcation points). Finally, we get:

$$E\{A_1^n\} = \frac{2n-1}{n-1} \sum_{m=2}^n p(m|n)E\{A_p^m\}. \tag{17}$$

■

APPENDIX B

Proof of equation (20). For the expected tree-asymmetry value we have:

$$E\{A_1^n\} = \frac{2n-1}{n-1} \sum_{m=2}^n p(m|n)E\{A_p^m\}. \tag{17}$$

Insertion of equation (19) for the function $p(m|n)$ and equation (11) for the expected partition asymmetry for random terminal growth results in:

$$E\{A_1^n\} = \frac{n}{n-1} \left\{ \frac{n_e}{2n(n_e-1)} + \sum_{m=3}^{n-1} \frac{m_e}{m(m+1)(m_e-1)} \right\}. \tag{B1}$$

The summation can be split into terms with odd $m(m_e = m - 1)$ and terms with even $m(m_e = m)$ as:

$$\sum_{m=3}^{n-1} \frac{m_e}{m(m+1)(m_e-1)} = \sum_{m=3,5,\dots}^{n_e-1} \frac{m-1}{m(m+1)(m-2)} + \sum_{m=4,6,\dots}^{2n-2-n_e} \frac{1}{(m+1)(m-1)}$$

and by replacing m by $2k + 1$ for the odd- m terms and m by $2k$ for the even- m terms rewritten as:

$$\sum_{k=1,2,\dots}^{n_e/2-1} \frac{2k}{(2k-1)(2k+1)(2k+2)} + \sum_{k=2,3,\dots}^{n-1-n_e/2} \frac{1}{(2k-1)(2k+1)}$$

and by combining both summations formulated as:

$$\sum_{k=1}^{n_e/2-1} \frac{1}{(2k-1)(k+1)} - \frac{1}{3} + \frac{n-n_e}{n(n-2)}. \tag{B2}$$

Insertion of equation (B2) into equation (B1) and combining terms finally results in:

$$E\{A_1^n\} = \frac{n}{n-1} \left\{ \frac{2-n_e/n}{2(n_e-1)} - \frac{1}{3} + \sum_{k=1}^{n_e/2-1} \frac{1}{(2k-1)(k+1)} \right\}. \tag{20}$$

■

LITERATURE

Abramowitz, M. and I. A. Stegun (Eds). 1965. *Handbook of Mathematical Functions*. New York: Dover.
 Caley, A. 1859. On the analytical forms called trees. *Philos. Mag.* **18**, 374–378.
 Hansen, E. R. 1975. *A Table of Series and Products*. New York: Prentice Hall (series in automatic computation).

- Harding, E. F. 1971. The probabilities of rooted tree-shapes generated by random bifurcation. *J. appl. Prob.* **3**, 44–77.
- Horsfield, K. and M. J. Woldenberg. 1986. Branching ratio and growth of tree-like structures. *Resp. Physiol.* **63**, 97–107.
- Horsfield, K., M. J. Woldenberg and C. L. Bowes. 1987. Sequential and synchronous growth models related to vertex analysis and branching ratios. *Bull. math. Biol.* **49**, 413–429.
- Jarvis, R. S. and A. Werrity. 1975. Some comments on testing random topology stream network models. *Wat. Resour. Res.* **11**, 309–318.
- Jarvis, R. S. and M. J. Woldenberg (Eds). 1984. *River Networks. Benchmark Papers in Geology/80*. Stroudsburg, PA: Hutchinson Ross.
- Shreve, R. L. 1966. Statistical law of stream numbers. *J. Geol.* **74**, 17–37.
- Smart, J. S. 1969. Topological properties of channel networks. *Geol. Soc. Am. Bull.* **80**, 1757–1774.
- Uylings, H. B. M., J. van Pelt and R. W. H. Verwer. 1989a. Topological analysis of individual neurons. In *Computer Techniques in Neuroanatomy*, J. J. Capowski (Ed.), pp. 215–239. New York: Plenum.
- Uylings, H. B. M., J. van Pelt, R. W. H. Verwer and P. M. McConnell. 1989b. Statistical analysis of neuronal populations. In *Computer Techniques in Neuroanatomy*, J. J. Capowski (Ed.), pp. 241–263. New York: Plenum.
- Van Pelt, J. and R. W. H. Verwer. 1983. The exact probabilities of branching patterns under terminal and segmental growth hypotheses. *Bull. math. Biol.* **45**, 269–285.
- Van Pelt, J. and R. W. H. Verwer. 1984. Cut trees in the topological analysis of branching patterns. *Bull. math. Biol.* **46**, 283–294.
- Van Pelt, J. and R. W. H. Verwer. 1985. Growth models (including terminal and segmental branching) for topological binary trees. *Bull. math. Biol.* **47**, 323–336.
- Van Pelt, J. and R. W. H. Verwer. 1986. Topological properties of binary trees grown with order-dependent branching probabilities. *Bull. math. Biol.* **48**, 197–211.
- Van Pelt, J., R. W. H. Verwer and H. B. M. Uylings. 1986. Application of growth models to the topology of neuronal branching patterns. *J. Neurosci. Meth.* **18**, 153–165.
- Van Pelt, J., H. B. M. Uylings and R. W. H. Verwer. 1989a. Distributional properties of measures of tree topology. *Acta Stereol.* **8**, 465–470.
- Van Pelt, J., R. W. H. Verwer and H. B. M. Uylings. 1989b. Centrifugal-order distributions in binary topological trees. *Bull. math. Biol.* **51**, 511–536.
- Verwer, R. W. H. and J. van Pelt. 1983. A new method for the topological analysis of neuronal tree structures. *J. Neurosci. Meth.* **8**, 335–351.
- Verwer, R. W. H. and J. van Pelt. 1985. Topological analysis of binary tree structures when occasional multifurcations occur. *Bull. math. Biol.* **47**, 305–316.
- Verwer, R. W. H. and J. van Pelt. 1986. Descriptive and comparative analysis of geometrical properties of neuronal tree structures. *J. Neurosci. Meth.* **18**, 179–206.
- Verwer, R. W. H. and J. van Pelt. 1990. Analysis of binary trees when occasional multifurcations can be considered as aggregates of bifurcations. *Bull. math. Biol.* **52**, 629–641.
- Verwer, R. W. H., J. van Pelt and H. B. M. Uylings. 1992. An introduction to topological analysis of neurons. In *Quantitative Methods in Neuroanatomy*, M. G. Stewart (Ed.), in press. Milton Keynes: Open University Press.
- Werner, C. and J. S. Smart. 1973. Some new methods of topologic classification of channel networks. *Geog. Analysis* **5**, 271–295.

Received 11 December 1991

Revised 29 July 1991