Bulletin of Mathematical Biology Vol. 47, No. 2, pp. 305-316, 1985.
Printed in Great Britain

0092-8240/85$3.00 + 0.00
Pergamon Press Ltd.

TOPOLOGICAL ANALYSIS OF BINARY TREE STRUCTURES WHEN OCCASIONAL MULTIFURCATIONS OCCUR

■ RONALD W. H. VERWER and JAAP VAN PELT
Netherlands Institute for Brain Research,
Meibergdreef 33,
1105 AZ Amsterdam, The Netherlands

The occurrence of multifurcations in essentially binary trees is investigated with respect to two methods for testing growth models, *viz.* subtree partition analysis and vertex analysis. It is shown that under certain conditions multifurcations may be incorporated in the analysis. Although the conditions are more restrictive for subtree partition analysis only minor loss of information occurs if forbidden multifurcations are simply ignored.

1. Introduction. Recently we proposed a new method for the analysis of binary tree structures, which enabled the application of statistical tests to the topological properties of the structure (Verwer and Van Pelt, 1983). Essential to the method is a classification which is performed according to the partition of terminal segments over the pairs of subtrees emanating from the bifurcation points (Van Pelt and Verwer, 1984a). Accordingly we call the method subtree-partition analysis (SPA). The method is very useful in testing growth models since the frequencies of occurrence of the partitions are highly dependent on the mode of growth. The classification is only applicable to binary branching structures. In neurobiology dendritic arborizations are predominantly binary, however, sometimes trifurcations occur (Fig. 1). For instance, Bradley and Berry (1976) reported that in normal Purkinje cell dendrites about 7% of the branching points consists of trifurcations. When the length of the segment connecting two bifurcations is smaller than its diameter the actual sequence of two bifurcations cannot be discerned and a trifurcation is observed. Accordingly, it has been suggested that trifurcations occurring occasionally in essentially binary trees may be considered as an aggregation of two consecutive bifurcations (Shreve, 1966; Horsfield and Cummings, 1968; Percheron, 1979). Also another method to analyse topological tree structures, called vertex analysis (VA), which would be applicable in the presence of trifurcations was proposed (Berry and Flinn, 1984; Sadler and Berry, 1983). In the present paper we investigate how to interpret an occasional trifurcation in terms of bifurcations and the implications for both SPA and VA.

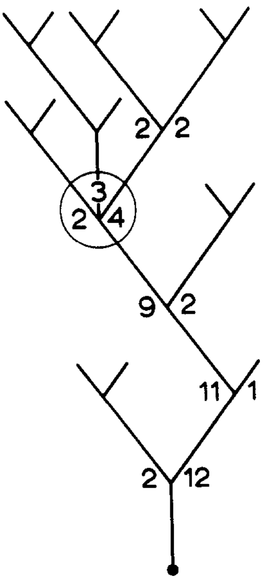


Figure 1. Illustration of a trifurcation occurring in a binary tree.

2. Multifurcations in Topological Analysis

2.1. Subtree-partition analysis. When the topological properties of observed branching patterns are investigated there are several ways in which the observations can be classified (Van Pelt and Verwer, 1984a). The very problem that arises when a multifurcation occurs is due to its interference with the classification of the observations as will be shown below. In Fig. 1 a tree with a trifurcation is shown. For classification according to the degrees of the subtree pairs of bifurcation points we can distinguish the following partitions: 14(2, 12), 12(1, 11), 11(2, 9) and 4(2, 2). But we would be uncertain about how to classify the trifurcation 9(2, 3, 4). This trifurcation might have arisen from the aggregation of the pairs of bifurcations [(2, 7)(3, 4)], [(3, 6)(2, 4)] or [(4, 5)(2, 3)] and there is no way to determine which one applies to this particular observation. Similarly, classification according to ambilateral type (i.e. the three-dimensional characterization of a tree which is invariant to internal rotations) would invoke the dilemma to which of the three ambilateral type numbers (1115, 1120 or 1127) the tree belongs (cf. Van Pelt and Verwer, 1983, 1984a). It may be noted that there are 2179 different ambilateral types with 14 terminal segments. The purpose of classification is to enable either comparison between different groups of binary branching structures or testing against growth models. We have recently worked out a range of growth models, which includes terminal and segmental growth as particular cases (Van Pelt and Verwer, 1985). Characteristic of the growth models is that two kinds of segments are distinguished (i.e. terminal and intermediate segments) and that all segments of a kind have the same probability to form a new segment. For instance, in the case of terminal growth all intermediate segments have probability zero and all terminal segments have probability $1/n$, if the tree has degree n, to branch. For this range of growth models the partition of subtrees at each bifurcation is completely independent of all the other partitions in the same tree. Thus, each partition can be considered as an independent event in the studied growth processes of binary trees; and the probability of a tree is given by the product of all partition probabilities corrected for the occurrence of bifurcations whose subtrees are of 'equal degree but of unequal type' (EDUT) (Van Pelt and Verwer, 1984a). Consequently, the frequencies of the partitions in a sample of observed binary trees form the basis for testing the growth hypotheses. From this point of view we could simply ignore the multifurcations and use only proper bifurcations. On the other hand, this implies that to study multifurcations we only need to consider the product of partition probabilities that can be formed from the subtrees involved in the multifurcations corrected for the occurrence of EDUT-subtrees. Therefore, investigation of multifurcations proceeds in the same way for both ambilateral types and subtree partitions.

In Fig. 2 a tree with a trifurcation at the first branching point is shown, together with the schematic representation of the three possible ambilateral types ($\alpha^n_{k_1}$, $\alpha^n_{k_2}$, $\alpha^n_{k_3}$) that cannot be distinguished as a result of the trifurcation. Here, n denotes the degree and k_m (m = 1, 2, 3) is the particular number in the set of nth-degree trees (Van Pelt and Verwer, 1983). The subtrees originating from the trifurcation have ambilateral type $\alpha^{r_1}_{i_1}$, $\alpha^{r_2}_{i_2}$ and $\alpha^{r_3}_{i_3}$ with the putative corresponding complementary subtrees $\alpha^{r_2+r_3}_{j_1}$, $\alpha^{r_1+r_3}_{j_2}$ and $\alpha^{r_1+r_2}_{j_3}$.

In its general form the partition probability (cf. Van Pelt and Verwer, 1985) of an nth-degree tree, with first-order subtrees of degree r and $n-r$, can be reformulated as:

$$p(r, n-r; Q) = \frac{\prod_{m=1}^{r-1}\left(\frac{m-Q}{m}\right)\cdot\prod_{m=1}^{n-r-1}\left(\frac{m-Q}{m}\right)}{\prod_{m=1}^{n-1}\left(\frac{m-Q}{m}\right)}\cdot\left[1+Q\left\{\frac{n(n-1)}{2r(n-r)}-2\right\}\right] \times \frac{2^{1-\delta_{r,n-r}}}{(n-1)}. \tag{1}$$

Here $\delta_{x,y} = 0$ if $x \neq y$ and $\delta_{x,y} = 1$ if $x = y$.

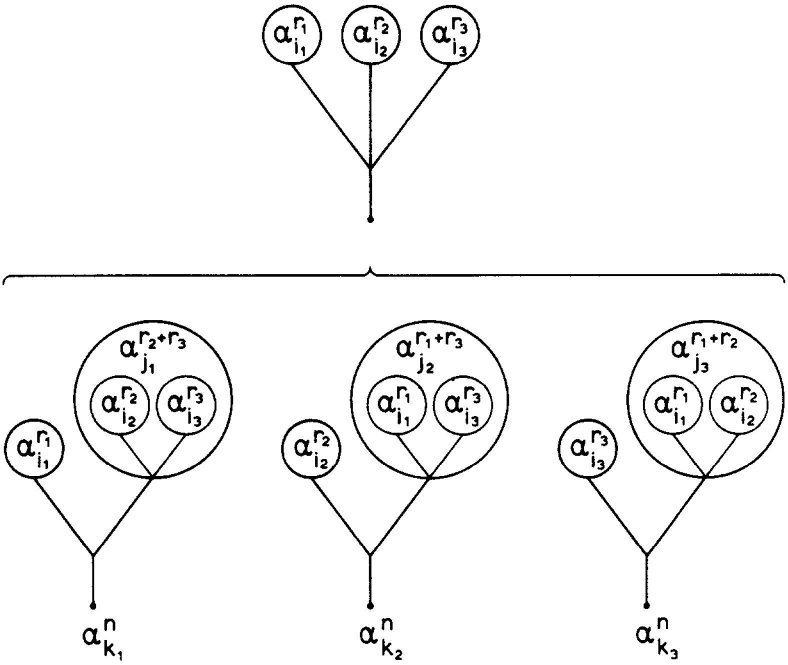


Figure 2. Three different arrangements of subtrees that could produce the presented trifurcation when the two subsequent bifurcation points are very close. Explanation is given in the text.

When Q is in the half-open interval $[0, 1)$, then the partition probabilities are interpretable in terms of growth modes (Van Pelt and Verwer, 1985) such that different values of Q correspond to different growth modes. For example, for terminal growth $Q = 0$ while for segmental growth $Q = 1/2$.

The correction factor for EDUT-subtrees is $2^{\delta_{r,n-r}(1-\delta_{i,j})}$ (Van Pelt and Verwer, 1983), where i and j are the ambilateral type numbers of the subtrees of degree r and $n - r$ respectively. The probability of the particular combination $(r_1, r_2 + r_3)(r_2, r_3)$ in Fig. 2. is

$$\mathrm{Prod}(1; Q) = p(r_1, n - r_1; Q)p(r_2, r_3; Q) \cdot 2^{\delta_{r_1,n-r_1}(1-\delta_{i_1,j_1})} \cdot 2^{\delta_{r_2,r_3}(1-\delta_{i_2,i_3})}.$$

Let

$$G(r, n - r) = \frac{n(n-1)}{2r(n-r)} - 2,$$

and

$$F(r_1, r_2, r_3) = \frac{\prod_{m=1}^{r_1-1}\left(\frac{m-Q}{m}\right) \cdot \prod_{m=1}^{r_2-1}\left(\frac{m-Q}{m}\right) \cdot \prod_{m=1}^{r_3-1}\left(\frac{m-Q}{m}\right)}{\prod_{m=1}^{n-1}\left(\frac{m-Q}{m}\right)} \cdot \frac{1}{n-1},$$

then the products of the partition probabilities for the three situations can be written as:

$$\mathrm{Prod}(1;Q) = F(r_1, r_2, r_3)[1 + QG(r_1, r_2 + r_3)]\,[1 + QG(r_2, r_3)] \times \frac{2^{2-\delta_{r_1,r_2+r_3}\delta_{i_1,j_1}-\delta_{r_2,r_3}\delta_{i_2,i_3}}}{(r_2 + r_3 - 1)} \tag{2a}$$

$$\mathrm{Prod}(2;Q) = F(r_1, r_2, r_3)[1 + QG(r_2, r_1 + r_3)]\,[1 + QG(r_1, r_3)] \times \frac{2^{2-\delta_{r_2,r_1+r_3}\delta_{i_2,j_2}-\delta_{r_1,r_3}\delta_{i_1,i_3}}}{(r_1 + r_3 - 1)} \tag{2b}$$

$$\mathrm{Prod}(3;Q) = F(r_1, r_2, r_3)[1 + QG(r_3, r_1 + r_2)]\,[1 + QG(r_1, r_2)] \times \frac{2^{2-\delta_{r_3,r_1+r_2}\delta_{i_3,j_3}-\delta_{r_1,r_2}\delta_{i_1,i_2}}}{(r_1 + r_2 - 1)}. \tag{2c}$$

In general Prod(1; Q), Prod(2; Q) and Prod(3; Q) will have different values for a given Q and moreover these values will change if Q is changed. Now suppose that for each Q $(0 \leqslant Q < 1)$

$$\mathrm{Prod}(1;Q) = \mathrm{Prod}(2;Q) = \mathrm{Prod}(3;Q) \tag{3}$$

and

$$\mathrm{Prod}(1;Q_1) \neq \mathrm{Prod}(1;Q_2) \quad \text{if} \quad Q_1 \neq Q_2.$$

If we evaluate equations (2a–c) it appears that only two situations correspond to relation (3). One situation is that all three subtrees are of the same ambilateral type and the other is that all three subtrees are of the same degree, but each having a different ambilateral configuration (see also Table I; situations 2 and 4). In both situations the classification is the same for all three combinations, which enables further analysis. In principle it would also be possible to use any trifurcation for which the relative proportions of Prod(1; Q), Prod(2; Q) and Prod(3; Q) would be constant for each Q $(0 \leqslant Q < 1)$, i.e.

$$\mathrm{Prod}(1;Q) : \mathrm{Prod}(2;Q) : \mathrm{Prod}(3;Q) = \text{constant} \tag{4}$$

and

$$\mathrm{Prod}(1;Q_1) \neq \mathrm{Prod}(1;Q_2) \quad \text{if} \quad Q_1 \neq Q_2.$$

However, this has no practical significance since, apart from the special relation (3), the trifurcation must be divided over the three combinations in accordance with the relative proportions of the products of the partition probabilities in relation (4). For instance, suppose that we observe a trifurcation of which two subtrees have the same ambilateral type and the third only has the same degree as the others (cf. situation 3 in Table I). Then, unless we have observed a multiple of five of this trifurcation we end up with classes containing a fractional number of observations, which is very uncommon in statistical practice. The alternative of lumping these classes not very practical either, since when a number of different trifurcations is observed the number of classes may be strongly reduced. If the relative proportions of Prod(1; Q), Prod(2; Q) and Prod(3; Q) vary with different growth hypotheses it is completely impossible to incorporate the trifurcation in further analysis. Prior knowledge of Q would be required to classify such a trifurcation, which prohibits subsequent testing of the hypotheses. In Table I the relative proportions for the products of the partition probabilities under terminal and segmental growth have been elaborated for a number of distinct situations to illustrate the above arguments. Thus, in practice only situations 2 and 4 in Table I are useful. For multifurcations involving more than three subtrees the situation becomes more complicated, even if all subtrees are identical. Imagine a quadrifurcation with all subtrees having ambilateral type (α_1^4), then one group of 12 trees consists of bifurcation combination (4, 12) (4, 8) (4, 4) and another group of three trees consists of bifurcation combination (8, 8) (4, 4) (4, 4). Proceeding in the same fashion as with equations (2a–c) it turns out that the relative proportions of the partition probability products still contain terms with Q. Therefore they vary with changing growth hypotheses and are useless in further analysis. However, it is clear that a multifurcation in a higher-order node will not affect the analysis of subtree pairs of first-order bifurcation points. If the data conform to one of the hypothetical growth modes defined by equation (1) this also holds for higher-order branching points. Note that the bifurcation points before and after the multifurcation are not affected. Thus, all branching points not included in a multifurcation can be used for SPA to enlarge the number of observations.

2.2. Vertex analysis. In vertex analysis the number of closed vertices (Va) (i.e. bifurcations into two terminal segments) is divided by the number of half-open vertices (Vb) (i.e. bifurcations into an intermediate segment and a terminal segment) to provide a test value (cf. Berry and Flinn, 1984). In order to determine the effect of multifurcations on vertex analysis we must first establish some relations. Let $f(m \mid \alpha_i^n)$ be the number of mth-degree subtrees in an ambilateral type with n terminal segments whose rank is i. Of course, $f(m = 1 \mid \alpha_i^n)$ equals n, since each terminal segment is a subtree of degree one. It is easily verified that the number of closed vertices is identical to the number of second-degree subtrees and therefore,

$$Va = f(2 \mid \alpha_i^n).$$

TABLE I

Illustration of the Relative Proportions of the Products of the Partition Probabilities of Subtrees involved in a Trifurcation under Terminal and Segmental Growth Hypothesis for a Number of Distinct Situations

Description of the situation	Under terminal growth hypothesis	Under segmental growth hypothesis
1 All three subtrees have a different degree ($r_1 \neq r_2 \neq r_3$)	$\frac{1}{r_2+r_3-1} : \frac{1}{r_1+r_3-1} : \frac{1}{r_1+r_2-1}$	1 : 1 : 1
2 All three subtrees have the same ambilateral type ($r_1 = r_2 = r_3; i_1 = i_2 = i_3$)	1 : 1 : 1	1 : 1 : 1
3 Only two subtrees have same ambilateral type; third has same degree ($r_1 = r_2 = r_3; i_1 = i_2 \neq i_3$)	2 : 2 : 1	2 : 2 : 1
4. All subtrees have same degree, but different ambilateral type ($r_1 = r_2 = r_3; i_1 \neq i_2 \neq i_3$)	1 : 1 : 1	1 : 1 : 1
5 Two subtrees have same ambilateral type ($r_1 = r_2 \neq r_3; i_1 = i_2$)	$\frac{2}{r_2+r_3-1} : \frac{2}{r_1+r_3-1} : \frac{1}{r_1+r_2-1}$	2 : 2 : 1
6 Two subtrees have same degree ($r_1 = r_2 \neq r_3; i_1 \neq i_2$)	$\frac{1}{r_2+r_3-1} : \frac{1}{r_1+r_3-1} : \frac{1}{r_1+r_2-1}$	1 : 1 : 1
7 One subtree has same ambilateral type as the combination of the two others ($r_1 + r_2 = r_3; i_3 = j_3$)	$\frac{2}{r_2+r_3-1} : \frac{2}{r_1+r_3-1} : \frac{1}{r_1+r_2-1}$	2 : 2 : 1
8 Two subtrees have same ambilateral type; the third subtree has same ambilateral type as the combination of the former subtrees ($r_1 = r_2; r_1 + r_2 = r_3; i_1 = i_2; i_3 = j_3$)	$\frac{4}{r_2+r_3-1} : \frac{4}{r_1+r_3-1} : \frac{1}{r_1+r_2-1}$	4 : 4 : 1

Since each terminal segment that is not part of a pair belongs to a half-open vertex, the number of half-open vertices is equal to the total number of terminal segments minus twice the number of second-degree subtrees, which can be written as

$$Vb = n - 2 \cdot f(2 \mid \alpha_i^n).$$

If $E\{f(2 \mid n)\}$ denotes the mean number of second-degree subtrees in an nth-degree tree

$$E\{f(2 \mid n); Q\} = \sum_{i=1}^{N_\alpha^n} f(2 \mid \alpha_i^n) \cdot p(\alpha_i^n; Q) \tag{5}$$

where N_α^n is the total number of possible ambilateral types with n terminal segments and $p(\alpha_i^n; Q)$ is the probability of occurrence of ambilateral type α_i^n within the set of nth-degree trees, then the vertex ratio $R = Va/Vb$ can be estimated by

$$\hat{R}(Q) = \frac{E(Va; Q)}{E(Vb; Q)} = \frac{E\{f(2 \mid n); Q\}}{n - 2 \cdot E\{f(2 \mid n); Q\}} \tag{6}$$

(see Cochran, 1977). Computer simulations of large networks have shown that the vertex ratio equals 1 for terminally grown trees and 0.5 for segmentally grown trees (Berry and Flinn, 1984; Sadler and Berry, 1983). To show that equation (6) precisely corresponds to the estimator proposed by Berry and coworkers we must evaluate $\hat{R}(Q)$ for terminal ($Q = 0$) and segmental ($Q = 0.5$) growth respectively. Van Pelt and Verwer (1984b) have derived that the probability that a subtree of degree m occurs in a terminally grown tree of degree n is given by

$$p(m \mid n; Q = 0) = \frac{2n}{(2n - 1)(m + 1)m}.$$

Therefore, the mean number of second-degree subtrees in a tree of degree n equals

$$E\{f(2 \mid n); Q = 0)\} = \frac{n}{3}. \tag{7}$$

If this is substituted into equation (6) we arrive at

$$\hat{R}(Q = 0) = \frac{n/3}{n - 2n/3} = 1.$$

For segmentally grown two-dimensional topological trees Shreve (1967) deduced the probability of an mth-degree subtree occurring in an nth-degree tree. It was proved that his formula also holds for three-dimensional topological trees and reformulation (Van Pelt and Verwer, 1984b) results in

$$p(m \mid n; Q = 0.5) = \frac{(4n - 4m - 2)}{(2n - 1)} \cdot \frac{N_\tau^{n-m} \cdot N_\tau^m}{N_\tau^n}$$

where

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

(cf. Shreve, 1966). The mean number of second-degree subtrees in a segmentally grown nth-degree tree then is

$$E\{f(2 \mid n); Q = 0.5\} = \frac{(4n - 10)(2n - 1)}{3 \cdot (2n - 5)} \cdot \frac{\binom{2n-5}{n-2}\binom{3}{2}}{\binom{2n-1}{n}}$$

which can be rewritten as

$$E\{f(2 \mid n); Q = 0.5\} = \frac{n(n-1)}{2(2n-3)}. \tag{8}$$

Substitution into equation (6) and dividing both numerator and denominator by $E\{f(2 \mid n); Q = 0.5\}$ gives

$$\hat{R}(Q = 0.5) = \left\{ \frac{2(2n-3)}{(n-1)} - 2 \right\}^{-1}. \tag{9}$$

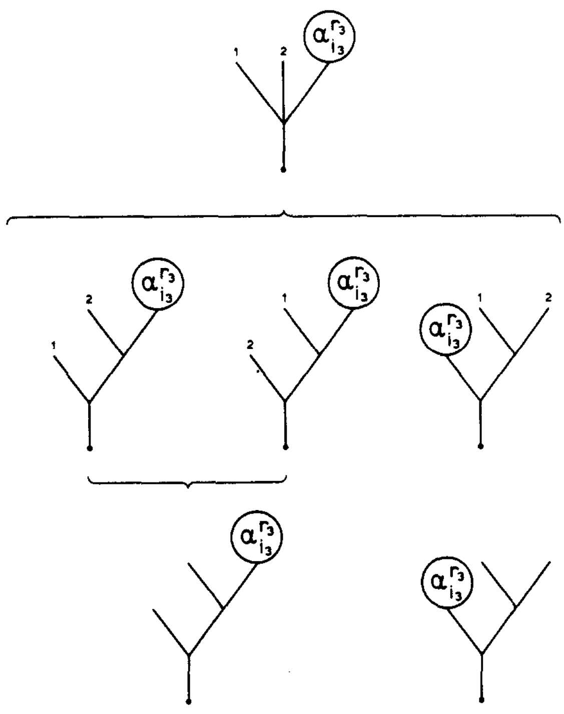
Since,

$$\lim_{n \to \infty} \frac{2(2n-3)}{(n-1)} = 4$$

it follows, that

$$\lim_{n \to \infty} \hat{R}(Q = 0.5) = 0.5.$$

Thus, our exact expression of the vertex ratio estimate is in accordance with the computer simulations of Berry and coworkers (cf. Sadler and Berry, 1983). In vertex analysis it is necessary and sufficient to determine the degree of observed trees, count the number of observed second-degree subtrees, subsequently calculate the vertex ratio estimate and compare this with the expected vertex ratio values for terminal or segmental growth respectively. Consequently, trifurcations that contain two first-degree subtrees impede the counting of second-degree subtrees and the tree in which they occur must be deleted from the analysis. In Fig. 3 a situation is presented in which the number of second-degree subtrees cannot be established unambiguously. This is a special case of situation 5 in Table I. It is clear that if one would try to complete the count of second-degree subtrees this results in favouring one growth model over others. Generally, each x-multifurcation ($x > 3$) that involves at least two subtrees of degree one will invalidate the tree that contains the multifurcation for analysis. As far as we know at present no growth models other than terminal and segmental growth have been investigated with vertex analysis. It may be noted that in vertex analysis trees of different degree can be analysed together. This means that, if the analysis is not restricted to trees of one specific degree n, subtrees of a tree containing a multifurcation which themselves do not contain the multifurcation can be treated as separate trees in the analysis.



Terminal growth	$\frac{4}{n-2}$	:	1
Segmental growth	4	:	1

Figure 3. A trifurcation that cannot be included in vertex analysis. See text for explanation.

3. Conclusions. We have investigated in which way the occurrence of multifurcations interfered with subtree partition analysis and vertex analysis. When essentially binary trees contain occasional multifurcations, such that the underlying growth mechanism [defined by equation (1)] is not affected, all bifurcation points not included in a multifurcation can be used in the subtree partition analysis. If the subtrees that are involved in any trifurcation conform to certain constraints (cf. situations 2 and 4 in Table I), that trifurcation may be included in the subtree partition analysis as well. This requires that all subtrees of the existing trifurcations are checked. However, if the number of normal bifurcations in the observed sample is sufficiently large it seems unnecessary to invest so much effort. For vertex analysis the situation is different, only one possible trifurcation (see Fig. 3) cannot be analysed and the extra check may be worthwhile. Multifurcations with more than three subtrees involved cause extra complications. From our results it appears that attempting to reconstruct the lost arrangement of bifurcations may result in manipulating the data in favour of some growth model. However, it should be realized that in neuronal trees multifurcations occur infrequently. Since SPA uses all available topological information, loss of data due to occasional multifurcations seems of minor importance. Although VA uses less topological information trifurcations with precisely two first-degree subtrees or multifurcations with two or more first-degree subtrees will form only a small fraction of all possible multifurcations. Probably this will not cause serious problems in practice.

We gratefully acknowledge H. M. B. Uylings for critical commenting on the manuscript and H. Stoffels for drawing the figures.

LITERATURE

Berry, M. and R. Flinn. 1984. "Vertex Analysis of Purkinje Cell Dendritic Trees in the Cerebellum of the Rat." *Proc. R. Soc.* **221B**, 321–348.

Bradley, P. M. and M. Berry. 1976. "The Effects of Reduced Climbing and Parallel Fibre Input on Purkinje Cell Dendritic Growth." *Brain Res.* **109**, 133–151.

Cochran, W. G. 1977. *Sampling Techniques*, 3rd edn. New York: J. Wiley.

Horsfield, K. and G. Cummings. 1968. "Morphology of the Bronchial Tree in Man." *J. appl. Physiol.* **24**, 373–383.

Percheron, G. 1979. "Quantitative Analysis of Dendritic Branching. I. Simple Formulae for the Quantitative Analysis of Dendritic Branching." *Neurosci. Lett.* **14**, 287–293.

Sadler, M. and M. Berry. 1983. "Morphometric Study of the Development of Purkinje Dendritic Trees in the Mouse using Vertex Analysis." *J. Microsc.* **131**, 341–354.

Shreve, R. L. 1966. "Statistical Law of Stream Numbers." *J. Geol.* **74**, 17–37.

——. 1967. "Infinite Topologically Random Channel Networks." *J. Geol.* **74**, 178–186.

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. 1984a. "New Classification Methods of Branching Patterns." *J. Microsc.* **136**. 23–34.

—— and ——. 1984b. "Cut Trees in the Topological Analysis of Branching Patterns." *Bull. math. Biol.* **46**, 283–294.

—— and ——. 1985. "Growth Models (including Terminal and Segmental Branching) for Topological Binary Trees." *Bull. math. Biol.* **47**.

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.

RECEIVED 7-30-84