CUT TREES IN THE TOPOLOGICAL ANALYSIS OF BRANCHING PATTERNS

J. Van Pelt and R. W. H. Verwer
Netherlands Institute for Brain Research,
P.O. Box 41850,
1009 DB Amsterdam,
The Netherlands

In this paper the effects of the occurrence of cut trees in the topological analysis of branching patterns have been studied. It is assumed that branches are removed at random from the trees. We prove that, for both the segmental and terminal growth models, the probability distributions of the cut trees are identical to those of complete trees.

1. Introduction. Recent studies of the growth of branching patterns in terms of their topological properties have resulted in new methods for the analysis of observed tree structures (Van Pelt and Verwer, 1983; Verwer and Van Pelt, 1983) and the application of hypothesis tests to the experimental data (Uylings et al., 1983). One of the implied assumptions in these studies was that the observed trees involved in the analysis were complete, i.e. that the structure was not affected by factors not related to the growth model. In practice, however, many trees are demonstrably incomplete and as such do not meet this criterion. In neurological studies, for instance, sectioning or incomplete staining may cause trees to be incomplete. Reconstruction of a complete neuron from its sectioned parts is possible in principle but is very time consuming. For instance, computer reconstruction of a neuron defined by 7900 coordinates from eight sections takes about 40 hr (Capowski and Sedivec, 1981). Deleting incomplete trees from the analysis forms an alternative, although undesirable in view of the amount of effort to be spent in revealing the structure of neuronal branching patterns. To recover properties of lost parts one might try to estimate these properties by extrapolation from the observed data. For instance, in a study to quantify the number of segments of a given order, Samuels et al. (1977) have compensated for cut branches in a statistical way. The correction factors could not be inferred from the actual structures, however, and were highly dependent on assumptions concerning the likelihood of cutting intermediate and terminal segments. Attempts to recover lost information from cut trees will generally be based upon assumptions concerning the lost parts and thus can strongly influence the result. Finally, one can try to calculate or estimate the amount of lost information and
determine if changes in properties of branching patterns due to cutting are detectable at all in the particular experiments. This approach seems to be most appropriate as a first step and will be followed in this study. Hollingworth and Berry (1975) have studied the effect of random loss of portions of large dendritic trees on the frequency distribution of Strahler orders and found that, for both the terminal and segmental growth models, random loss did not lead to significant differences in the slopes of the curves (i.e. bifurcation ratios). These results have been obtained from simulating random loss in large trees (i.e. ≈ 500 terminal segments). It is not clear, however, whether the results also apply to the distribution of ambilateral types or subtree pairs, or if the number of terminal segments is of importance to their conclusion. The aim of this study was to quantify mathematically the distortion of the ambilateral-type probability distributions (cf. Van Pelt and Verwer, 1983) by the incorporation of cut trees in the analysis. Since the terminal and segmental growth models are most frequently used in other studies, they are used to define the probability distribution of the trees that are cut. To describe the cutting process we have assumed that each tree is cut at random, i.e. one of its subtrees is selected at random and removed from the tree. Under these conditions it appeared for both models that distributions of cut trees of a particular degree were identical to the complete-tree distributions for the same degree. The practical consequence of these results is that cut trees can be incorporated in the analysis if the observed distributions are actually generated by the terminal or segmental growth model and the trees are assumed to be cut randomly. 

Exact probabilities of cut trees within the context of terminal or segmental growth have been derived mathematically. Expressions are derived for the probability of occurrence of subtrees of a particular degree in a set of trees. With these formulae we are able to prove the equality of cut-tree and complete-tree probability distributions.

2. Nomenclature.
Ambilateral type
2D-topological type
$A^n$
Root segment
Intermediate segment
Terminal segment

representation of the three-dimensional (3D) topological structure of a tree,
representation of the two-dimensional (2D) topological structure of a tree,
set of all ambilateral types with $n$ terminal segments,
ambilateral type with number $i$ from set $A^n$,
first segment of a tree,
segment between two consecutive bifurcation points,
segment between a bifurcation point and a terminal point,
3. Subtree Probabilities. The quantity \( p(m|n) \) denotes the probability that a subtree within a tree of degree \( n \), randomly selected from a set of \( n \)th-degree trees, is of degree \( m \) (\( m \leq n \)). If \( p(m|\alpha_0^n) \) equals the probability of finding an \( m \)th-degree subtree in ambilateral type \( \alpha_0^n \), then we have

\[
p(m|n) = \frac{\sum_{i=1}^{N^n_\alpha} p(m|\alpha_0^i) \cdot p(\alpha_0^i)}{N^n_\alpha}.
\] (1)

Here \( N^n_\alpha \) is equal to the total number of ambilateral types with \( n \) terminal segments and \( p(\alpha_0^i) \) is the probability of occurrence of ambilateral type \( \alpha_0^i \) in the set \( A^n \). By this definition the probability \( p(m|n) \) is determined by the probabilities of occurrence of the \( n \)th-degree trees in the set \( A^n \). The random selection of a subtree in a tree corresponds to the random selection of a segment that acts as the root segment of the subtree. The probability \( p(m|n) \) has some basic properties: an \( n \)th-degree tree has \((2n - 1)\) segments and \( n \) terminal segments. This means, for instance, that if a subtree is selected at random, the probability to select a subtree of degree one (terminal segment) is equal to \( n/(2n - 1) \),

\[
p(1|n) = n/(2n - 1).
\] (2)

Moreover, an \( n \)th-degree tree has one subtree of degree \( n \), i.e. the tree itself. The probability of selecting the root segment and thus selecting the tree itself is equal to \( 1/(2n - 1) \),

\[
p(n|n) = 1/(2n - 1).
\] (3)

The probability \( p(m|n) \) can be expressed in terms of the corresponding probabilities for the first-order subtrees of the \( n \)th degree tree by the equation
\[ p(m|n) = \sum_{r=m}^{n-1} p(r, n-r) \cdot \frac{2r-1}{2n-1} \cdot (1 + \delta_{r,n-r}) \cdot p(m|r) \]  

(4)

for \( m < n \).

The summation runs over all degrees of first-order subtrees which, in their turn, could contain an \( m \)th-degree subtree. Each term in the summation describes the probability of finding an \( m \)th-degree subtree in a first-order subtree. Here, \( p(r, n-r) \) denotes the probability of a tree having first-order subtrees of degrees \( r \) and \( n-r \), and the term \( (2r-1)/(2n-1) \) describes the probability of selecting one of the segments in the \( r \)th-degree subtree. The quantity \( p(m|r) \) denotes the probability of finding the \( m \)th-degree subtree in the \( m \)th-degree first-order subtree. The Kronecker delta \( \delta_{r,n-r} \) equals zero if \( r \neq n-r \) and equals one if \( r = n-r \). The term \( (1 + \delta_{r,n-r}) \) accounts for the situation where both first-order subtrees are of equal degree, which doubles the probability. In equation (4) the term \( p(r, n-r) \) is determined by the probabilities of occurrence of all trees in set \( A^n \) and as such is dependent on the way the trees have been grown. We shall derive expressions for the probability \( p(m|n) \) for trees grown via terminal or via segmental growth.

Subtree probabilities in terminally grown trees. For trees grown via terminal growth (random branching of only terminal segments) we have (Van Pelt and Verwer, 1983)

\[ p_t(r, n-r) = \frac{1}{n-1} \cdot 2^{1-\delta_{r,n-r}}. \]  

(5)

Substitution of equation (5) in equation (4) gives

\[ p_t(m|n) = \frac{1}{(n-1) \cdot (2n-1)} \cdot \sum_{r=m}^{n-1} 2^{1-\delta_{r,n-r}} \cdot (1 + \delta_{r,n-r}) \cdot (2r-1) \cdot p_t(m|r) \]  

(6)

for \( m < n \).

The product \( 2^{1-\delta_{r,n-r}} \cdot (1 + \delta_{r,n-r}) \) equals 2 irrespective of whether \( r = n-r \) or \( r \neq n-r \), so we get for equation (6)

\[ p_t(m|n) = \frac{2}{(n-1) \cdot (2n-1)} \cdot \sum_{r=m}^{n-1} (2r-1) \cdot p_t(m|r) \]  

(7)

for \( m < n \).

Straightforward evaluation of this expression for \( p_t(m|n+1) \) gives
for \( m < n \) or, written differently,

\[
\frac{pt(m|n + 1)}{n + 1} = \frac{2(n + 1) - 1}{n + 1} = \frac{2n - 1}{n} 
\]

for \( m < n \).

Equation (9) leads to the conclusion that the expression \( pt(m|n) \cdot \frac{(2n - 1)}{n} \) is independent of the value of \( n \) \((m < n)\), and therefore we can also write

\[
pt(m|n) \cdot \frac{(2n - 1)}{n} = pt(m|m + 1) \cdot \frac{(2(m + 1) - 1)}{m + 1}. \tag{10}
\]

According to (7) we have (if \( n = m + 1 \))

\[
pt(m|m + 1) = \frac{2}{m} \cdot \frac{2m - 1}{2(m + 1)} \cdot pt(m|m),
\]

and further reduction using (3) gives

\[
p(m|m + 1) = 2/(m \cdot (2m + 1)). \tag{11}
\]

Inserting (11) into (10) results in

\[
pt(m|n) = 2n/((2n - 1) \cdot (m + 1) \cdot m) \tag{12}
\]

for \( m < n \); for \( m = n \), (3).

**Subtree probabilities in segmentally grown trees.** Shreve (1967) has derived that the probability of selecting a subtree of degree \( m \) in a set of 2D-topological types of degree \( n \), grown via segmental growth, is described by

\[
pt(m|n) = (n - m + 1) \cdot N_{T}^{n-m+1} \cdot N_{T}^{m}/((2n - 1) \cdot N_{T}^{n}). \tag{13}
\]

The quantity \( N_{T}^{n} \) denotes the number of 2D-topological types of degree \( n \). This formula can be explained as follows. The denominator describes the total number of segments in all 2D-topological types of degree \( n \). The numerator describes the number of ways in which an \( m \)-th-degree 2D-topological type (\( N_{T}^{m} \) different types) can be appended to one of the terminal segments of an \((n - m + 1)\)-th-degree 2D-topological type [the number of terminal segments of all \((n - m + 1)\)-th-degree trees is equal to \((n - m + 1) \cdot N_{T}^{n-m+1}\)]. All these combinations result in \( n \)-th-degree 2D-topological types that do contain a subtree of degree \( m \). The probability of a subtree of a particular degree in a set of trees must, of course, be independent of the choice to describe the trees in a 2D-projection plane (2D-topological types) or in the 3D-space (ambilateral types). This property can be verified as follows.
In analogy to equation (1), the quantity \( p(m|n) \) derived by Shreve can be written as

\[
p(m|n) = \sum_{j=1}^{N^n_T} p(m|\tau^n_j) \cdot p(\tau^n_j).
\] (14)

The expression can be rewritten as the sum of all 2D-topological types that are of the same ambilateral type and a summation over these groups (in fact, over all ambilateral types):

\[
p(m|n) = \sum_{i=1}^{N^n_m} \sum_{j=1}^{N_{eq}(\alpha^n_i)} p(m|\alpha^n_i) \cdot p(\tau^n_j).
\] (15)

Here, \( \tau^n_j \) denotes a 2D-topological type \( \tau^n_j \) that corresponds to the ambilateral type \( \alpha^n_i \) and \( N_{eq}(\alpha^n_i) \) denotes the number of 2D-topological types corresponding to \( \alpha^n_i \) (Van Pelt and Verwer, 1983).

The probability of selecting an \( m \)th-degree subtree in a 2D-topological type \( \tau^n_j \) is equal to the probability that this subtree occurs in the ambilateral type \( \alpha^n_i \). Thus we get

\[
p(m|n) = \sum_{i=1}^{N^n_m} \sum_{j=1}^{N_{eq}(\alpha^n_i)} p(m|\alpha^n_i) \cdot p(\tau^n_j).
\] (16)

The second summation describes precisely the probability of occurrence of \( \alpha^n_i \), viz. \( p(\alpha^n_i) \), and we get

\[
p(m|n) = \sum_{i=1}^{N^n_m} p(m|\alpha^n_i) \cdot p(\alpha^n_i).
\] (17)

Comparison of this equation with equation (14) leads to the conclusion that the quantity \( p(m|n) \) can be defined on a set of 2D-topological types as well as on the corresponding set of ambilateral types. Further, it must be noted that this result is independent of the probability distributions of the types within their sets. According to this rule we can apply equation (13) also to the set of ambilateral types. Reformulation of equation (13), and making use of the expression for \( N^n_{eq} \):

\[
N^n_{eq} = (2n - 1)^{-1} \cdot \binom{2n - 1}{n}
\] (Shreve, 1966), results in
According to Van Pelt and Verwer (1983), the probability of a tree having first-order subtrees of degree $m$ and $n - m$ under segmental growth is equal to

$$p_s(m, n - n) = 2^{n-m} \cdot N_r^{n-m} \cdot N_r^m / N_r^n,$$

so we can also express the probability $p_s(m | n)$ in terms of $p_s(m, n - m)$:

$$p_s(m | n) = \frac{2(n - m) - 1}{2n - 1} \cdot 2^{n-m} \cdot N_r^m \cdot p_s(m, n - m).$$

4. Cutting in Trees. The expressions for the subtree probabilities now enable us to quantify the consequences of random cutting and in particular we shall calculate the influence of cutting on the distribution of subtree pairs. We shall define a cut tree as a tree from which an entire subtree, including its root segment, has been removed. If in practice after cutting a part of a segment is still present then this part is completely ignored in the analysis. According to this definition, the branching point acting as the root of the removed subtree will also disappear.

Random cutting. We shall assume that the trees are cut at random only once, i.e. from each tree one subtree is removed at random. Random means that each of the $2n - 1$ subtrees has equal probability to be removed. If we start with a set of trees of equal degree then the cutting process results in trees of any degree lower than the original one. Without loss of generality we can study the distribution in one of the lower degree sets and compare it with the distribution of complete trees in that set predicted by the growth model. Therefore we will only consider those cutting processes that involve the survival of an incomplete tree of a definite degree. Suppose all trees in the original set are of degree $k + m + q$ and from each tree one subtree is removed at random. All cases where a subtree of degree $q$ is removed result in an incomplete tree of degree $k + m$. For these incomplete trees we want to calculate the probability $p(k, m)$ of having first-order subtrees of degree $k$ and $m$. The probability $P_c(k, m)$ that a $(k, m)$ subtree pair is retained after cutting is equal to the probability $p(q | k + m + q)$ of selecting a $q$th-degree subtree from one of the original trees, multiplied by the probability $p(k, m)$, for the surviving tree to have first-order subtrees of degree $k$ and $m$:

$$P_c(k, m) = p(q | k + m + q) \cdot p(k, m).$$

Three situations in which the removal of a $q$th-degree subtree from an $(k + m + q)$-degree tree results in a cut tree with first-order subtrees of degree $k$ and $m$ can be distinguished (cf. Figure 1). The probability $P_c(k, m)$
Figure 1. Three different arrangements of subtrees with the property that the removal of a subtree of degree $q$ results in a tree with first-order subtrees of degrees $k$ and $m$. Each subtree is represented by a circle and the enclosed number indicates its degree (number of terminal segments). In configuration (b) the $q$th-degree subtree is located somewhere in the left first-order subtree of degree $k + q$ ($k = k_1 + k_2$). In configuration (c) the $q$th-degree subtree is located somewhere in the right first-order subtree of degree $m + q$ ($m = m_1 + m_2$). It is assumed that $k \leq m$, i.e. configuration (b) and (c) are indistinguishable if $k = m$.

is also equal to the probability of one of the complete-tree configurations in the original set, displayed in Figure 1, times the probability of selecting precisely the $q$th-degree subtree. Without loss of generality we assume that $k \leq m$.

$$P^c(k, m) = p(q, k + m) \cdot p(k, m) \cdot 2^{q, k+m} \cdot \frac{1}{2(k + m + q) - 1}$$

$$+ \left[ p(k + q, m) \cdot \frac{2(k + q) - 1}{2(k + m + q) - 1} \cdot p(q|k + q) \cdot 2^{q, k+m} \right]$$

$$+ p(k, m + q) \cdot \frac{2(m + q) - 1}{2(k + m + q) - 1} \cdot p(q|m + q) \cdot 2^{k, m+q} \cdot 2^{-k,m}.$$  (23)

The first term in (23) describes configuration (a) in Figure 1 as the probability of selecting a $(k + m, q)$ tree in the original set times the probability that the $(k + m)$ subtree is of the type $(k, m)$ times the probability that just the $q$th-degree subtree is removed. The second term in (23) corresponds to configuration (b) in Figure 1 and describes the probability of selecting a $(k + q, m)$ tree in the original set multiplied by the probability that a segment in the $(k + q)$-degree first-order subtree is selected times the probability that this $(k + q)$-degree subtree has on its turn a subtree of degree $q$. The third term accounts for configuration (c) in Figure 1 in a corresponding way. All powers of two with Kronecker deltas account for configurations of equal-degree subtree pairs. The last one corrects for the situation when $k$ equals $m$, in which case configurations (b) and (c) in Figure 1 are indistinguishable. Rearranging the terms in (23) results in
For further evaluation of this equation we have to define the probability distribution of the trees in the original set. We shall assume in the next sections the segmental and terminal growth models for the original set of trees respectively.

**Random cutting of segments in segmentally grown trees.** The probability $p_s(m|n)$ has already been derived in equation (21) and can be inserted in equation (24) immediately:

\[
P^c_s(k, m) = \frac{1}{2(k + m + q) - 1} \times \left[ p_s(q, k + m) \cdot p_s(k, m) \cdot 2^{q, k + m} \right.
\]

\[
+ p_s(k + q, m) \cdot p_s(q|k + q) \cdot (2(k + q) - 1) \cdot 2^{q, k + q, m} \cdot 2^{-q, k, m}
\]

\[
+ p_s(m + q, k) \cdot p_s(q|m + q) \cdot (2(m + q) - 1) \cdot 2^{q, m + q} \cdot 2^{-q, k, m} \right].
\]  

The product $p_s(k + q, m) \cdot p_s(k, q)$ can be rewritten using the expression for these probabilities in equation (20) as

\[
p_s(k + q, m) \cdot p_s(k, q) = p_s(k, m) \cdot p_s(k + m, q) \cdot 2^{q, m + q, k} \cdot 2^{-q, k, q} \cdot 2^{-q, k + m, q} \cdot 2^{-q, k + m, q}.
\]  

A corresponding expression can be derived for $p_s(m + q, k) \cdot p_s(m, q)$ by interchanging $k$ and $m$. Substitution of these expressions into equation (25) results in

\[
P^c_s(k, m) = \frac{1}{2(k + m + q)} \cdot p_s(k, m) \cdot p_s(k + m, q) \cdot 2^{q, k + m} \cdot (2(k + m) - 1).
\]

According to equation (21) we have

\[
p_s(q|k + m + q) = \frac{2(k + m) - 1}{2(k + m + q) - 1} \cdot 2^{q, k + m} \cdot p_s(q, k + m),
\]

thus finally we have

\[
P^c_s(k, m) = p_s(k, m) \cdot p_s(q|k + m + q).
\]  


Comparing this equation with (22) leads to the result

\[ p_s(k, m)^c = p_s(k, m), \tag{29} \]

which shows that if in a set of segmentally grown trees from each particular tree one subtree is removed at random, the set of cut trees is also described by the segmental growth probability distribution.

**Random cutting of segments in terminally grown trees.** The subtree probability \( p_t(m|n) \) and the probability of first-order subtrees \( p_t(m, n - m) \) [equations (12) and (5) respectively] can be substituted immediately into equation (24), resulting in

\[ P_t(k, m) = \frac{2^{1-k,m}}{k + m - 1} \frac{2(k + m + q)}{(2k + m + q - 1) \cdot q \cdot (q + 1)}. \tag{30} \]

Realizing that the first and second terms describe exactly the probability of \( p_t(k, m) \) and \( p_t(q|k + m + q) \) according to equations (5) and (12), we have

\[ P_t^c(k, m) = p_t(k, m) \cdot p_t(q|k + m + q), \tag{31} \]

resulting finally in the equality

\[ p_t(k, m)^c = p_t(k, m). \tag{32} \]

Thus also for terminally grown trees we have the property that random cutting of segments results in sets of cut trees in which each tree has a probability as is predicted by the terminal growth model.

The results expressed by equations (29) and (32) concern the probabilities of subtree pairs, but can be shown to hold for the probabilities of ambilateral types as well. For both growth models the branching probability of a segment is independent of the order of the segment. Therefore the probability of an ambilateral type can be calculated by multiplying the probabilities of all its subtree pairs, including incidental corrections for symmetrical subtree pairs. If the subtree-pair probabilities remain consistent with the model after random cutting, then the same holds for the probabilities of ambilateral types. In the derivation of probabilities of cut trees we have assumed that the trees are cut only once. Evidently, however, repeated cutting will also result in consistent probability distributions for the terminal or segmental growth model. In conclusion, the property of having distributions of cut trees consistent with the segmental or terminal growth model does not depend on the number of times the trees are cut. The direct practical consequence of these findings is that incomplete (cut) trees can be involved in testing sets of observed trees against the terminal or segmental growth models, provided that the trees have been cut randomly.
5. Discussion. The results obtained by calculating the effect of incomplete trees on the statistical properties of its topological structures can be summarized as follows. Random removal of subtrees does not lead to model inconsistent probability distributions of ambilateral types or subtree pairs if the trees have actually grown via terminal or segmental growth.

However, one assumption initially stated may not be met in practice. The way the trees are cut may not be at random. This may seem to occur, for instance, if a substantial part of a tree is removed by sectioning. One has to note, however, that this type of cutting is certainly not random with respect to the geometrical properties, but the non-randomness in a topological sense may be much less severe. If non-randomness is expected, then the analysis must be restricted to only the complete subtrees of the observed trees. One can also artificially remove the smallest subtree containing all cut terminal points. Then, effectively, only one subtree is removed more or less at random in its original position. In general, however, it is not possible at present to give general rules, valid under all experimental (not-random) cutting conditions, when the distortion of the distribution by cut trees will become significant.

So far we have focused our attention on the probability distributions of cut trees. In this context it is also important to study the topological properties of the sets of the removed subtrees as trees with cut root points. For instance, we might pose the question: If, in a set of nth-degree trees, subtrees are removed at random from randomly chosen trees and again probability distributions are formed from all equal-degree subtrees, what is the relationship between these distributions and the one for the original trees?. For the terminal and segmental growth models this question can be answered immediately: the removed subtrees will again form distributions corresponding to the growth model. During the growth of a tree each terminal segment has the potential to become the root segment of a new subtree. Neither growth model distinguishes segments with respect to their order, and so the statistical properties of higher-order subtrees are identical to those of the complete trees. In fact, this statement is valid for any growth model defined by branching probabilities for segments that are independent of the order of the segments.

We thank Drs H. B. M. Uylings and A. J. Noest for critical reading of the manuscript and for their valuable suggestions. Furthermore, we are grateful to Mr. H. Stoffels and Mrs C. Sypkens for artwork and photography respectively.
LITERATURE


RECEIVED 6-17-83