Bulletin of Mathematical Biology (2000) **62**, 657-674

Statistical analysis and parsimonious modelling of dendrograms of in vitro neurones

J.M. Devaud [1]*, B. Quenet [2]*, J. Gascuel [3] and C. Masson [3]

[1] Instituto Cajal (CSIC)
Avenida Doctor Arce, 37 28002 Madrid – Spain

[2] Laboratoire d'Electronique (ESPCI)
10 Rue Vauquelin, 75005 Paris – France

[3] Centre d'Etude des Sciences du Goût (CNRS)
15 Rue Hugues Picardet, 21000 Dijon - France

* These authors contributed equally to this work.

ABSTRACT

The processes whereby developing neurones acquire morphological features that are common to entire populations (thereby allowing the definition of neuronal types), are still poorly understood. A mathematical model of neuronal arborisations may be useful to extract basic parameters or organisation rules, hence helping to achieve a better understanding of the underlying growth processes.

We present a parsimonious statistical model, intended to describe the topological organisation of neuritic arborisations with a minimal number of parameters. It is based on a probability of splitting which depends *only* on the centrifugal order of segments. We compare the predictions made by the model of several topological properties of neurones with the corresponding actual values measured on a sample of honeybee (olfactory) antennal lobe neurones grown in primary culture, described in a previous study. The comparison is performed for three populations of segments corresponding to three neuronal morphological types previously identified and described in this sample. We show that simple assumptions together with the knowledge of a very small number of parameters, allow the topological reconstruction of representative (bi-dimensional) biological neurones. We discuss the biological significance (in terms of possible factors involved in the determinism of neuronal types) of both common properties and cell-type specific features, observed on the neurones and predicted by the model.

1 Introduction

The dendritic trees of neurones exhibit a large range of sizes, shapes and ramification patterns. Nevertheless, common morphological characteristics shared by groups of neural cells can be identified, and may serve as the basis for the identification of neuronal types. However, the developmental mechanisms underlying the acquisition of cell-type specific features are still poorly understood. The development of mathematical descriptions of neuronal arborisations and their application to build functional models and growth models, have contributed to our knowledge of the physiological and developmental significance of morphological regularities (Burke *et al*., 1992; Nowakowski *et al*., 1992; Dityatev *et al*., 1995). The recent development of elaborate models of dendritic arborisations, make it possible to infer, from the modelled properties of mature neurones, some aspects of their growth (Dityatev *et al*., 1995; Van Pelt *et al*., 1997; Van Pelt, 1997). In the present study, we wish to shed light on some of the possible mechanisms whereby neurones produce neuritic arborisations in culture. We develop a simple, parsimonious, statistical model, based on a splitting probability distribution only, intended to give a quantitative description of neuronal trees. The model is parsimonious in that it requires *only two parameters* in order to account for the splitting probability distribution. These two parameters of the model are then fitted to data obtained from a population of cultured neurones taken from a previous study (Devaud *et al.,* 1994) and the predictions of the resulting model are tested by comparing the topological properties of the segments that make up the experimental arborisations, i.e. distributions of some relevant morphometrical parameters or relationships between some of them, with the computed equivalents. In this approach, we analyse separately the different morphological types of the neuronal sample, and check that the model can be easily adapted to each of them, despite its small number of parameters. The detailed mathematical relations about the different probability laws and the different means used in this study are presented in Appendix.

2 Morphometric characterisation of the neuritic trees

The study was done on a large data set obtained from 361 neurones of the honeybee antennal lobe, cultured for one week at a low density in a conditioned medium (at that time most growth processes had stopped). A detailed description of the culture preparation and of the procedure for acquiring images of the neurones has been given elsewhere (Devaud *et al.*, 1994).The neuritic arborisation of the cultured bi-dimensional neurones is described as a set of hierarchically-arranged segments. A *segment* is defined as a portion of neurite comprised between two ramification nodes (intermediate segments), or between a node and a tip (terminal segments) (Van Pelt & Verwer, 1983). On the whole, the sample contains more than 7,000 segments.

In contrast to the first approach used on this sample (Devaud *et al.*, 1994), this work focuses on the population of *segments* which compose the neurones, instead of the neurones themselves. Hence new morphometric parameters have been used specifically for the characterisation of the segments:

- ***Order (q).*** It represents the topological distance from the soma. Its value is an integer incremented at every bifurcation ('centrifugal order', for review see Uylings *et al.*, 1975). An order value of 1 is assigned to the primary segments, i.e. those emerging directly from the soma (Capowski, 1989).

- ***Degree (D).*** It represents the number of tips of a subtree (or partition) stemming from a segment (Verwer & Van Pelt, 1987). In a binary tree, it is related to N, the number of segments of the partition, by $N = 2D - 1$ (Verwer *et al.*, 1992). For a terminal segment, $D = 1$, and for an intermediate segment, $D \geq 2$.

- ***Asymmetry index (A)*** (Van Pelt *et al.*, 1992). It can be associated to *intermediate segments only* ($D \geq 2$): if D_1 and D_2 are the degrees of two "sister" segments beginning at the bifurcation of the considered segment, the asymmetry index of this bifurcation is given by the following formula:

$$A = \left| \frac{D_1 - D_2}{D_1 + D_2 - 2} \right| \qquad \text{if } D_1 + D_2 = D > 2,$$

$$A = 0 \qquad \text{if } D_1 = D_2 = 1.$$

The larger the value of A, the larger the asymmetry of the partition. A belongs to [0,1], and $A = 0$ when the arborisation is symmetrical ($D_1 = D_2$), whereas A=1 when the arborisation is completely asymmetrical ($D_1 = 1$ and $D_2 = D - 1$).

We had previously been able to identify, using a detailed morphometric description, three neuronal types (referred as types 1, 2 and 3) (Devaud *et al*., 1994). Both metrical and topological differences were observed between these types on the basis of measurements of, respectively, the maximal lengths, total lengths and areas of the dendritic trees, and their neuronal degrees and maximal segment orders. Thus, type 1 neurones are 'elongated', 'large' and 'highly ramified'; type 2 neurones are 'poorly elongated', 'small' and 'poorly ramified'; type 3 neurones are 'elongated', 'large' and 'very poorly ramified'. In this study, we will only focus on the topological properties of these types (the mean values of the correspondent parameters are given in Table I). In order to take into account these specificities in the present study, the segments were grouped according to the types of the neurones they belong to: the numbers of segments of type 1, 2 and 3 are respectively 3911, 1485 and 1732. The parameters of the model were fitted to the measured values *separately* for the segments belonging to neurones of each type.

3 Assumptions of the model

The basic assumptions of the model are:

1. All neuritic arborisations are binary (Verwer *et al*., 1992), i.e. the rare multiple ramifications (only 0.7% of the ramifications in our experimental sample) are considered as successive bifurcations separated by very short segments whose length is smaller than the microscope resolution.

2. The splitting probability P_q, i.e. the probability for a segment to bifurcate into two new segments, depends on its order q only. In particular, the splitting probability of a given segment is supposed to be independent of that of its "sister" segment, which means that no explicit competition between the segments growing from the same bifurcation is introduced in the model ("independence hypothesis"). Under this assumption, it is legitimate to work on sets of neuritic segments, not on individual neurones.

3. A segment can be either an intermediate segment or a terminal one: as a consequence, the probability for a segment being terminal is simply $(1-P_q)$.

These first three assumptions mean that we consider the neuronal trees under study as random binary trees grown according to a Bienaymé-Galton-Watson process (Athreya & Ney, 1972, Kliemann, 1987).

4. To adapt the expression of the splitting probability as a function of the segment order to the experimental measures, we looked for the function that would give the best fit to the dependence of the ratio of the number of intermediate segments (N_q^I) to the total number of segments (N_q), on their order. Since a segment is defined as intermediate if there is a bifurcation at its extremity, the number of intermediate segments at a given order q equals the number of ramifications (Verwer *et al*., 1992), each of them being the point of departure of two segments of order q+1:

$$N_q^I = \frac{N_{q+1}}{2}.$$

It can be proved simply (see Appendix 1, and Burke *et al*., 1992) that this ratio is an unbiased estimator of the splitting probability :

$$P_q = \lim_{N_q, N_{q+1} \to \infty} \frac{N_{q+1}}{2 \cdot N_q} = \lim_{N_q, N_q^I \to \infty} \frac{N_q^I}{N_q}.$$

We assume that our data set of segments is large enough to get a good estimate P_q^* of P_q, which is denoted as:

$$P_q^* = \frac{N_q^I}{N_q}.$$

The observed evolution of this ratio as a function of the order is shown in Fig. 1 for each neuronal type. As the order increases, the relative number of intermediate segments decreases regularly, indicating that more and more segments stop their growth. A non-linear regression shows *that the functions that best fit these three functions are exponential*, of the form:

$$\boxed{P_q^* = \frac{1}{2} \cdot e^{(\alpha - \beta \cdot q)}} \qquad [1]$$

where the *two* parameters α and β (at the origin of the parsimony of the model) take on different positive values for each of the three neuronal types (see Table II) .

4 Analytical properties

4.1 Order distribution

The comparison between the computed order distributions and the experimental ones is a simple way of checking that the splitting probabilities P_q^* (one per

neuronal type) have been correctly estimated. The computation of the order distributions is based on equation [3], a consequence of the recurrent relation [2], where N_1 is the number of primary neurites:

$$N_q^* = 2 \cdot N_{q-1}^* \cdot P_{q-1}^* \qquad \text{for } q > 1 \qquad [2]$$

Substituting P_q^* by its expression (equation (1)), N_q^* becomes:

$$\Rightarrow \qquad \boxed{N_q^* = N_1 \cdot \mathrm{e}^{(q-1)\left(\alpha - \frac{\beta}{2} q\right)}} \qquad [3]$$

Function N_x^* is maximum when $x = \frac{1}{b}\left(\alpha + \frac{\beta}{2}\right)$, *independently of* N_1, and $N_{\max}^* = N_1 \cdot e^{\frac{1}{2b}\left(\alpha - \frac{\beta}{2}\right)^2}$. As x and $N_{\max}^*$ are not necessarily integers, the maximal generation size (or 'width' of the tree) can be simply defined by taking the integer part of $N_{\max}^*$.

If the probability P_q is strictly lower than 0.5, the extinction probability of a Galton-Watson tree is equal to 1(Athreya & Ney, 1972); therefore, a splitting probability that decreases exponentially leads necessarily to such a situation. We define an integer limit value for q, $q_{\max}$, such as $N_{q_{max}}^* \geq 1$ and $N_{q_{max}+1}^* < 1$. This limit can be estimated as follows: if $x_{\max}$ is the solution of $N_{x_{max}}^* = 1$, then

$$x_{\max} = \frac{1}{b}\left(a + \frac{\beta}{2} + \sqrt{2 \cdot \beta \cdot \ln(N_1) + \left(\alpha - \frac{\beta}{2}\right)^2}\right) \qquad [4]$$

and $q_{\max}$ can be defined as the integer part of $x_{\max}$: it represents the 'height' of the tree. In the following sections, all equations should be considered as valid for $q \leq q_{\max}$.

4.2 Distribution of the degree and evolution of the mean degree with order

In Appendix 2, we show that the probability for a segment of order q to have a degree $D = d$ can be expressed by the following recurrent relation:

$$\boxed{\begin{aligned} &p_q(D=1) = (1 - P_q) \\ &p_q(D = d > 1) = P_q \cdot \left(\sum_{k=1}^{d-1} p_{q+1}(D=k) \cdot p_{q+1}(D=d-k)\right) \qquad \text{if } d \leq 2^{q_{\max}-q} \\ &p_q(D = d > 1) = 0 \qquad \text{if } d > 2^{q_{\max}-q} \end{aligned}} \qquad [5]$$

In order to perform the computation of the degree distribution, we simply apply the recurrent rule to the estimated probabilities and we choose to initialise this rule with : $p_{q_{\max}}^*(D=1) = (1 - P_{q_{\max}}^*)$ and $p_{q_{\max}}^*(D = d > 1) = 0$. The estimation of

the number of segments with degree d at order q is $n_q^*(d) = N_q^* \cdot p_q^*(D = d)$; hence the number $n^*(d)$ of segments with degree d, irrespective of the order is:

$$n^*(d) = \sum_{q=1}^{q_{\max}} n_q^*(d) = \sum_{q=1}^{q_{\max}} N_q^* \cdot p_q^*(D = d) \qquad [6]$$

We also show (see Appendix 2) the following recurrent relation for the mean degree:

$$\boxed{E_q[D] = (1 - P_q) + 2 \cdot P_q \cdot E_{q+1}[D]} \qquad [7]$$

valid for $q \in [1,.., q_{\max} - 1]$. If we denote by $\overline{D}_q^*$ the estimation of the mean degree at order q, then [7] becomes: $\overline{D}_q^* = (1 - P_q^*) + 2 \cdot P_q^* \cdot \overline{D}_{q+1}^*$. Let us now suppose that: $\overline{D}_{q_{\max}}^* = 1$; then, using equation [6], we can write:

$$\overline{D}_q^* = 1 + \frac{1}{2} \cdot \sum_{k=q}^{q_{\max}-1} \prod_{n=q}^{k} 2 \cdot P_n^* \qquad [8]$$

Replacing P_q^* by its expression in [1], we finally obtain:

$$\overline{D}_q^* = 1 + \frac{1}{2} \cdot \sum_{k=q}^{q_{\max}-1} \prod_{n=q}^{k} e^{(\alpha - \beta \cdot n)} = 1 + \frac{1}{2} \cdot \sum_{k=q}^{q_{\max}-1} e^{\left(\sum_{n=q}^{k} (\alpha - \beta \cdot n) \right)}$$

The mean value of the degree as a function of the order q is finally:

$$\boxed{\overline{D}_q^* = 1 + \frac{1}{2} \cdot \sum_{k=q}^{q_{\max}-1} e^{\left(\alpha (k - q + 1) - \frac{\beta}{2} (k(k+1) - q(q-1)) \right)}} \qquad [9]$$

which is not only dependent on α and β, the two parameters of the splitting probability function, but also depends on N_1, the number of primary neurites, through $q_{\max}$, computed with equation [4].

4.3 Distribution of the asymmetry index and evolution of the mean asymmetry index with order.

We show in Appendix 3 that the conditional probability for an *intermediate* segment at order q to have both a degree d and an asymmetry index a can be expressed as a function of the probabilities of its "daughter" segments to have some particular degrees (δ is the Kronecker symbol):

$$\boxed{\begin{aligned} & p_q(A = a, D = d \,|\, d > 1) \\ & = 2^{1 - \delta(a,0)} \cdot p_{q+1}\left(D = \frac{d \cdot (1 - a)}{2} + a \right) \cdot p_{q+1}\left(D = \frac{d \cdot (1 + a)}{2} - a \right) \end{aligned}} \qquad [10]$$

As previously, the computation is made by applying this rule to the estimated probabilities (labelled with *). The estimation of the number of segments that,

at order q, have both degree d and asymmetry index a is $n_q^*(d,a) = N_q^{I*} \cdot p_q^*(A=a, D=d|d>1)$, where $N_q^{I*} = N_q^* \cdot P_q^*$ is the estimation of the number of intermediate segments at order q. In other terms, $n_q^*(d,a) = N_q^* \cdot p_q^*(A=a, D=d|d>1) \cdot P_q^*$. Thus, the number of segments of asymmetry index a, irrespective of the order and the degree, is:

$$n^*(a) = \sum_{q=1}^{q_{\max}} \sum_{d=2}^{d_{\max_q}} n_q^*(d,a) = \sum_{q=1}^{q_{\max}} \sum_{d=2}^{d_{\max_q}} N_q^* \cdot p_q^*(A=a, D=d|d>1) \cdot P_q^* \qquad [11]$$

where $d_{\max_q} = 2^{q_{\max}-q}$.

The mean value of the asymmetry index A for a segment of order q can be estimated using the expression derived in Appendix 3:

$$E_q[A] = \frac{\sum_a \sum_d a \cdot p_q(A=a, D=d|d>1)}{\sum_a \sum_d p_q(A=a, D=d|d>1)}$$

The mean value $\overline{A}_q^*$ of A, is given by:

$$\boxed{\overline{A}_q^* = \frac{\sum_a \sum_d a \cdot p_q^*(A=a, D=d|d>1)}{\sum_a \sum_d p_q^*(A=a, D=d|d>1)}} \qquad [12]$$

5 Comparison with the measurements performed on actual neurones

In order to test the validity of the model, we compared its predictions with the experimental measures made on the sample of cultured honeybee neurones. The distribution of the proportion of intermediate segments were fitted by the exponential function for the three populations of segments corresponding to the three neuronal types present in the sample (Fig. 1). The values of the parameters α and β of the exponential function and N_1 the number of primary neurites were then used to reconstruct the order distribution, in order to asses the quality of the exponential fit. The analytical properties presented in the previous section were then used to compute data that *had not been exploited* to define the parameters of the model: the degree and the asymmetry index distributions, as well as the evolution of the mean degree and the mean asymmetry index with the order. These data predicted by the model were compared with the experimental data considering the neurone types separately in each case.

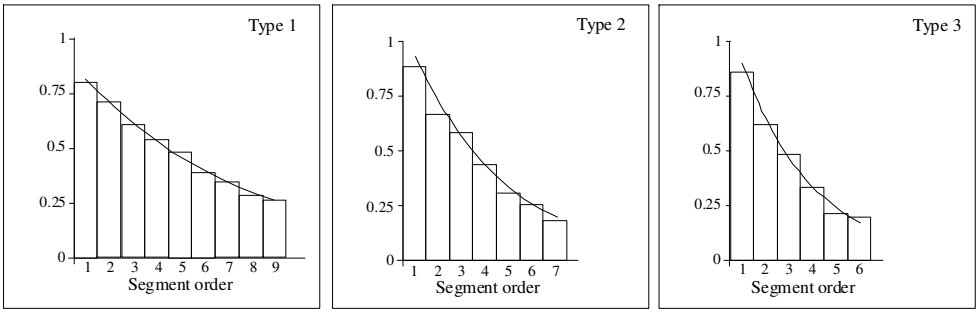

Figure 1. Evolution of the splitting probability with centrifugal order. Each panel corresponds to one of the three neuronal types; it displays the observed distribution of the proportion of *intermediate* neuritic segments at different orders (N_q^I / N_q), plotted as a histogram, and the decreasing exponential function, P_q that best fits the distribution. The parameters α and β, and R-squared are given in Table II.

5.1 Exponential fit and order distribution

For each neuronal type, the values of the parameters α and β, for which the highest goodness-of-fit was given by an exponential regression, are presented in Table II, with the value of the R-squared (higher to 95% in the three cases). The order distributions computed according to equation [3], with the estimated values of the two parameters α and β, have been tested as theoretical distributions for the experimental ones with a χ^2 test, whose values and probabilities are also given in table II. The measured values of N_1 (the number of primary segments) for the three types being 211, 107 and 186, the predicted values of $q_{\max}$, according to equation [4], are 14, 10 and 9, respectively. Experimental and theoretical order distributions are plotted in Fig. 2, which shows a good correlation for the three types.

Table II: Estimated coefficients α and β (mean ± standard error) of the fitted exponential function modelling the probability of splitting, P_q.

Neuronal type	α (±std error)	β (±std error)	R-squared	χ^2 test (d.f., p)
Type 1	**0.609** ± 0.001	**0.137** ± 0.0002	98.9 %	5.28 (14, p = 0.98)
Type 2	**0.848** ± 0.002	**0.247** ± 0.0008	98.6 %	3.15 (10, p = 0.97)
Type 3	**0.862** ± 0.001	**0.313** ± 0.0007	99.2 %	5.46 (9, p = 0.79)

The regression has been performed with Advanced Regression of Statgraphics Plus, on all the values of P_q corresponding to the orders where intermediate segments were experimentally observed ($N^I_q > 0$), weighted by the number of segments at each order (N_q). The χ^2 test results from the comparison of the computed order distributions (equation [3]) with the experimental ones: the degree of freedom corresponds to the computed value of q_{max} , and p is the probability that the experimental distribution may derive from the computed one.

5.2 Degree distribution and evolution of the mean degree with order

The computation of the degree distributions of the three segment types was performed from equation [6], and the simulated data were compared to the experimental ones with a χ^2 test. Each test, for each neuronal type, was applied to a number of bins corresponding to the maximal degree observed experimentally, this number also defined the degree of freedom for the test (see Table III).

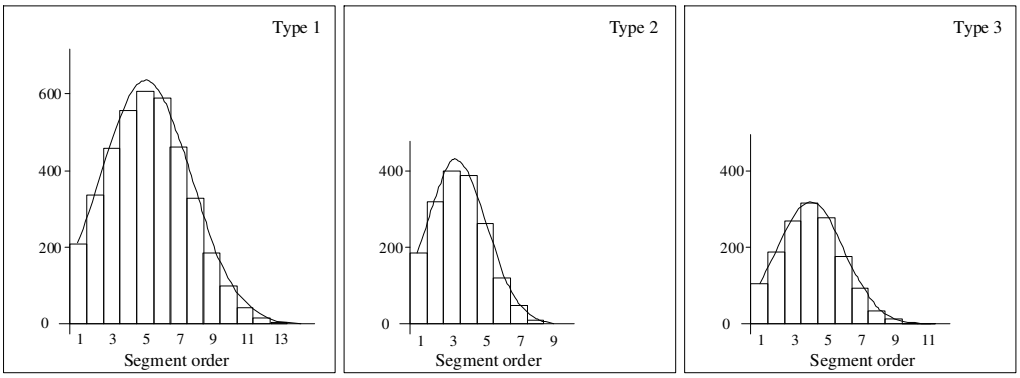

Figure 2. Frequency distributions of the order. Each graph corresponds to one of three neuronal types; it displays the observed frequency distribution of the centrifugal order of the neuritic segments, plotted as a histogram, and the distribution predicted with equation [3], as a continuous line.

The evolution of the mean degree with order q, $\overline{D}^*_q$, computed from equation [9], is shown in Fig. 3. For the three neuronal types, the magnitude of the difference between the computed mean degree and the experimental one, averaged on different q, was computed in order to compare it with the error around the experimental estimation of the value of the mean degree per order. These values are also given in Table III. As expected from the different

intensities of ramification previously measured for the three types (see the values of the mean degree per neurone in Table I), the highest degree values are found in segments from type 1 neurones while those from type 3 neurones exhibit the lowest values.

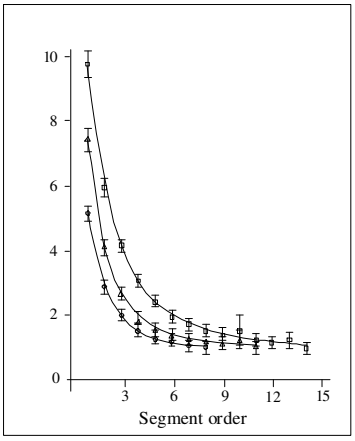

Figure 3. Evolution of the mean degree with order . The mean degree (± standard error) of the neuritic segments is plotted versus their centrifugal order for the three types of neurones, labelled as squares (type 1), triangles (type 2) and circles (type 3). The values expected from the model and computed with equation [9] are plotted as continuous lines.

Table III: Comparison of the computed and experimental degree distributions by a χ^2 test.

Neuronal type	χ^2 (d.f., p)	Mean ($\left\|\overline{D}_q^* - \overline{d}_q\right\|$)	Mean (std error ($\overline{d}_q$))
Type 1	29 (26, p = 0.3)	0.046	0.14
Type 2	4.2 (15, p = 0.99)	0.088	0.18
Type 3	7.14 (14, p = 0.91)	0.038	0.07

The probability p of the χ^2 test corresponds to the probability that the experimental degree distribution and the computed one (equation [9]) derive from the same distribution. The degree of freedom corresponds to the maximal experimental degree observed in each neuronal type. The mean values derived from experimental data ($\overline{d}_q$ is the mean experimental degree of the segments of order q) and from the comparison of these data with the computed ones ($\overline{D}_q^*$) provide a quantitative estimate of the quality of the model. Here the mean standard error around the mean experimental degree is much larger than the mean absolute difference between the experimental mean and the computed one.

5.3 Asymmetry index distribution and evolution of the mean asymmetry index with order

Equation [11] allows the computation of the asymmetry index distributions. As previously, the only parameters that are necessary to compute these

distributions are α, β and N_1 through the value of q_{max}. However, the number of the possible terms in the sum, where d goes from 2 to $2^{q_{max}-q}$, may be very large (up to 2^{13}). This is much larger than the experimental maximal degree observed (26, for a segment of a type 1 neurone). In order to avoid the computation of unnecessary terms, a limit of the degree was chosen arbitrarily equal to 40 (higher than 26 but not too high). In other terms, in the sum of equation [11], d goes from 2 to min(40, $2^{q_{max}-q}$). This limited degree leads to a limitation of the possible values for the asymmetry index, since it is a fraction with (d - 2) as denominator (see Appendix 3).

A statistical comparison between the computed asymmetry index distributions and the experimental ones is more complex than for the degree, since the asymmetry index is neither discrete nor really continuous. As an example, we consider the intermediate segments of type 1 neurones: there are 1850 of them, split up into 44 different values of $A = a$, while the corresponding computed distributions are spread over more than 200 possible values of a. If we consider only the asymmetry indices for which $n^*(a)$ is equal to or larger than 1 (according to equation [11], $n^*(a)$ is not necessarily an integer), their number is reduced to 40 values. Some of them are equal to the experimental values (typically: 0, 0.3333 and 1), but the majority of them are not. We can now consider three ways of processing the computed and experimental data in order to compare them. The first one consists in trying a comparison using a test usually applied to continuous distributions : the Kolmogorov-Smirnov test, knowing that it may be particularly severe in the case of semi-continuous data. The second one consists in building a histogram with $n^*(a)$ limited to the 44 values of the experimental asymmetry index and comparing it with the experimental histogram with a χ^2 test. The third one consists in the reconstruction of experimental and computed histograms with an arbitrary number of bins, for instance 100, distributed regularly over the interval [0,1]; a χ^2 test could also be applied to these histograms. These two χ^2 tests did not reveal any significant difference between the computed and the experimental histogram, nor did the Kolmogorov-Smirnov test. The results of the latter are presented for the three neuronal types in Table IV.

Table IV: Comparison of the computed and experimental asymmetry index distributions by a Kolmogorov-Smirnov test.

Neuronal type	D_{KS} (var, prob.)	Mean($\left\|\overline{A}_q^* - \overline{a}_q\right\|$)	Mean(std error($\overline{a}_q$))
Type I	0.038 (1.15, p = 0.14)	0.060	0.052
Type II	0.067 (1.22, p = 0.10)	0.038	0.047
Type III	0.054 (0.99, p = 0.27)	0.051	0.028

D_{KS} is the maximum value of the absolute difference between the respective cumulative functions of the computed distribution and of the experimental one. The Kolmogorov-Smirnov test is then applied by computing a convenient function of 'var', itself a function of D_{KS} (Press *et al*. 1992). Values of p lower than 0.05 would indicate a significant difference between the experimental and computed (equation [11]) distributions; as it is not the case here we conclude that the experimental distributions are correctly represented by the computed ones. As previously for the degree, the mean values derived from experimental data ($\overline{a}_q$ is the mean experimental asymmetry index of the segments of order q) and from the comparison of these data with the computed ones ($\overline{A}_q^*$) provide a quantitative estimate of the quality of the model. Here the mean standard error around the mean experimental asymmetry index is close to the mean absolute difference between the experimental mean and the computed one.

The evolution of the mean asymmetry index, $\overline{A}_q^*$, with order q, computed from equation [12], is represented in Fig. 4 together with the experimental values. As in the case of the degree, the magnitude of the difference between the computed mean asymmetry index and the experimental one, averaged over different q, has been computed in order to compare it with the error around the experimental estimation of the mean value of the asymmetry per order. These values are also displayed in Table IV. In Fig. 4, it can be observed that, whatever the neuronal type, the common tendency is a decrease of $\overline{A}_q^*$ with the order, which means that on the average the segments of higher order are more symmetric. This feature is more pronounced in segments taken from type 3 neurones, which is in keeping with their general morphology consisting of a main neurite with essentially single neurites emerging from it. If, on the whole, the computed function appears as a good fit of the experimental values, it is noticeable that,

for the three types, the value of $\overline{A}_1^*$ underestimates the experimental value. This is especially the case for type 1 neurones: this can be understood as the consequence of the dominance of a main primary neurite, with a few single emerging neurites at low order, and many ramified segments at higher order.

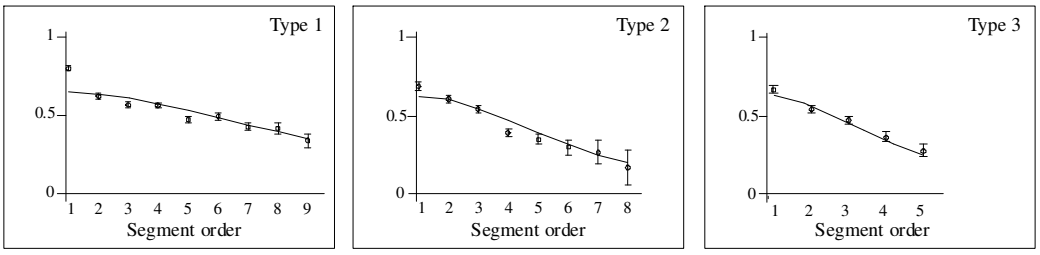

Figure 4. Evolution of the mean asymmetry index with order. Each graph corresponds to one of the three neuronal types; it displays the values of the asymmetry index (mean ± std error) of the neuritic segments versus their order. The values expected from the model and computed with equation [12] are plotted as continuous lines.

6 Discussion

We have presented a quantitative study of some basic parameters of neuronal trees over a large population of cultured neurones, and developed a simple model based on an order-dependent splitting probability. This model is parsimonious since we are able to characterise topological regularities in the organisation of the neuritic trees, and to predict them, with functions derived *only* from the exponential function that relates the splitting probability to the segment order. We did not address the question of individual morphologies; instead we intended to extract relevant characteristics of a large population from the strong "noise" produced by individual morphological variations. We are able to reconstruct several morphometric characteristics of the whole population without considering the characteristics of each cell, but taking into account, when necessary, the particularities of each neuronal type. It should be noticed that, using the mathematical rules presented before and the parameters that characterise the exponential splitting probability as a function of the segment order, we are not only able to derive the main topological properties of a large set of *segments*, provided the number N_1 of primary segments of such a set is given, but also a good representation of some topological features of a

typical neurone that would be representative of the type of neurones the set of segment is extracted from, provided the mean number of primary neurites per neurone is given. Such features are given in Table V, and can be compared with the values in Table I. Previous studies provided a demonstration that a branching probability (applied to branching events occurring during neuritic growth) is an important parameter for a parsimonious description and modelling of growing neuronal trees (Van Pelt *et al*., 1986; Kliemann, 1987; Verwer & Van Pelt, 1992; Carriquiry *et al*., 1992). Although we use a *splitting* probability (deduced from fully grown neurones), we also find that introducing an order-dependent relationship broadens the range of predictable morphologies (Van Pelt *et al*., 1986; Kliemann, 1987; Verwer & Van Pelt, 1990; Dityatev *et al*., 1995). It should be noted that this dependency upon the order is not an absolute requisite, as some neuronal morphologies can be predicted - at least as far as their topological organization is concerned - with a good degree of confidence with a model based on an order-independent probability (Verver & Van Pelt, 1990). The most realistic neuronal shapes were obtained with decreasing branching probabilities with increasing order (Van Pelt *et al*., 1986; Van Veen & Van Pelt, 1994; Dityatev *et al*., 1995; Uemura *et al*., 1995). Here, the excellent correlation between the distributions predicted from the probability law and the data obtained on actual segments strongly support the assumption of an exponential dependence.

Table V: Topological characteristics of a typical neurone for each neuronal type, according to the exponential splitting probability law.

Topological parameters	Type		
	1	**2**	**3**
Neuronal degree	**14**	**9**	**7**
Maximal segment order	**9**	**7**	**6**

Taking into account the mean number of primary neurites per neurone, as given in Table I, introduced as N_1 (which is not an integer value in that case) in equation [4], with α and β, the

parameters of the exponential splitting probability law, we define q_{max}, the maximal segment order for a typical neurone. It is then easy to compute with equation [9] the value of the mean degree of a primary segment: $\overline{D}_1^*$. Finally, as the degree of a neurone is the sum of the degrees of all its primary neurites, the typical neuronal degree can simply be obtained by rounding off the multiplication of $\overline{D}_1^*$ by N_1.

The very low proportion of multifurcations in our sample of neurones, as in many neuronal types (Percheron, 1979; Verwer & Van Pelt, 1985; Dityatev *et al*., 1995), led us to consider the neuronal trees under study as binary trees (Burke *et al*., 1992). In this respect, we followed most previous studies in which they were considered as the result of a limited spatial resolution rather than an inherent growth process (e.g. Stuermer, 1984; Hollingworth & Berry, 1975; Verwer & Van Pelt, 1990). Verwer & Van Pelt (1985) showed that the calculation of a branching probability without taking into account the multifurcations provided very poorly biased predictions, provided the frequency of multifurcations is low. These conclusions, along with our own results showing a high goodness-of-fit between our predicted data and the measured values, lead us to believe that our approximation results in a minor loss of information, and thus does not yield a strong bias.

Our results also suggest that, at least in our sample, the topology of neurones can be modelled without having to consider competitive processes other than the ones which may have been involved to give the observed distribution of terminal and intermediate segments, which is used for the adjustment of the splitting probability. Since all the theoretical distributions were obtained under the assumption that the growth of a segment does not depend on that of its "sister" segment (independence hypothesis), the high level of accuracy with which the experimental distributions were estimated strongly supports its validity. In particular, this holds true for the asymmetry index, a parameter sensitive to the developmental history of the tree (Van Pelt *et al*., 1992). This is in accordance with the results obtained in various studies from growth models, based, either explicitly (Kliemann, 1987; Nowakowski *et al*., 1992) or implicitly (Ireland *et al*., 1985; Li & Qin, 1996), on assumptions similar to our independence hypothesis. In contrast, a lateral inhibition has been considered by some authors (Li *et al*., 1992; 1995) to account for the internal limitation of material necessary to the extension of neurites at many growth cones. However,

such a limitation can be introduced without considering a lateral inhibition, by taking into account the finite length of the neuronal trees (Van Veen & Van Pelt, 1993), which has been done here by introducing a maximal value for the order (q_{max}).

Previous studies have shown that the topological variations observed in actual dendritic arborisations are highly dependent on the growth rules followed during development (Van Pelt *et al.*, 1986; 1992; 1997). Hence modelling "established" morphologies of neurones which have terminated their growth may allow one to infer some of the processes involved in the acquisition of basic properties of the arborisations. In particular, our approach based on the comparison of neurones which have been shown previously to belong to separate types, on the basis of morphological characteristics including metrical ones, may provide insights into the determinism of common and different aspects of type-specific neuronal shapes. First, we show that the three neuronal types under investigation can be reliably described in a unique way by the model presented here: the exponential probability law depending on the order gives an excellent prediction of their topological properties, independently of their strong differences in shape (see Devaud *et al.*, 1994). This common feature may be related to the similar conditions of growth in culture, since, *in vivo,* these neurones frequently present morphologies which do not correspond to a continuous decrease in the splitting probability. Most of them develop very dense distal ramifications (Fonta *et al.*, 1993; Devaud & Masson, 1999), which suggests that their splitting probability does not decrease as quickly as *in vitro*. This is consistent with the demonstration that non-stationarity in the dependence on order caused by environmental factors can result in major changes in the topology of arborisations (Van Pelt, 1997). Such factors capable of altering the law of branching probability include contacts made with afferents (Sadler & Berry, 1989; Carriquiry *et al.*, 1992) and growth factors (Dityatev *et al.*, 1995).

Still, our finding that the constraints defined by a common topological organisation allow an important variability between neuronal types supports the observations of others (Nowakowski *et al.*, 1992; Van Pelt *et al.*, 1997). At least in the present case, this variability is more likely to be related to intrinsic differences between neurones, since their development occurred in a homogeneous cell-culture medium and without cell-to-cell contacts (Devaud *et*

al., 1994). In this respect, the values of the parameters α and β of the exponential function describing the splitting probability, more than the type of function itself, seem to be relevant for the study of such differences. Using this criterion, neurones of types 2 and 3 have quite similar characteristics, including in terms of asymmetry, while type 1 neurones appear to be different. Hence the α and β parameters may be related to neurone-type specific processes of growth. They appear to be poorly correlated to the metrical features of neurones, since, for instance, type 1 and 3 neurones were shown to have similar sizes (as measured by the total area occupied by the arborisation and the maximal neuritic length; see Devaud et al, 1994).

Furthermore, for all neurones, irrespective of their type, the model underestimates the asymmetry index at the first order, while the predictions fit very well the measurements made at higher orders. Although we cannot discard the possibility that it might be an effect of the regeneration of a neuritic stump already grown by some neurones at the time of plating, this is unlikely given the low occurrence of this process - less than 15% (Devaud, 1997). Alternatively, this result may suggest that the ramification of the primary neurite obeys different mechanisms than the other segments, at least in the sample under study.

In summary, we have presented a simple, parsimonious model which allows a good characterisation of the topology of neurones. To address the question of the determinism of neuronal types, we applied the model to a sample in which three different types of neurones had been previously characterised. We show that these types share in common a basic topological organisation (represented by a splitting probability decreasing exponentially with the segment order), but exhibit specific variations (in particular in the coefficients of the exponential function). Our hypothesis that the exponential dependence on the order may be determined by the conditions of growth provided by the *in vitro* environment, while the different values of the parameters may represent cell-type specific properties, should be tested in the near future by modifying the environment, and by modelling the topology of the same neurones *in vivo*.

Acknowledgements

We would like to thank Prof. Gérard Dreyfus and Dr. Irina Kopysova for their critical reading of the manuscript, and Dr. Pascal Monestiez for our fruitful discussions. The anonymous reviewers are also thanked for their comments. We are also indebted to Steven Finch, whose website on random binary trees proved to be very informative (http://www.mathsoft.com/asolve/trees/trees.html). This work was supported by grants from the Swiss National Found for Scientific Research (FNRS), the Institut Français de la Recherche Agronomique (INRA) and the French Ministry of Science and Education.

References

Athreya, K. B. & Ney, P. (1972) *Branching processes*. Springer-Verlag.

Burke, R. E., Marks, W. B. & Ulfhake, B. (1992) A parsimonious description of motoneuron dendritic morphology using computer simulation. *J. Neurosci.* **12 (6)**, 2403-2416.

Capowski, J. J. (1989) *Computer Techniques in Neuroanatomy*. New York, Plenum.

Carriquiry, A. L., Ireland, W. P., Kliemann, W. & Uemura, E. (1992) Statistical evaluation of dendritic patterns. *Bull. Math. Biol.* **53**, 579-589.

Devaud, J. M., Quenet, B., Gascuel, J. & Masson, C. (1994) A morphometric classification of pupal honeybee antennal lobe neurones in culture, *NeuroReport* **6**, 214-218.

Devaud, J. M. (1997) Morphogenèse de neurones impliqués dans la discrimination olfactive: étude *in vivo* et *in vitro*. PhD Thesis, Université Pierre-et-Marie-Curie, Paris.

Devaud, J.M. & Masson, C. (1999) Dendritic pattern development of the honeybee antennal lobe neurons: a laser scanning confocal microscopic study in the honeybee *J. Neurobiol.* **34 (9)**, 461-474.

Dityatev, A. E., Chmykova, N. M., Studer L., Karamian, O., Kozhanov, V. M. & Clamann, H. P. (1995) Comparison of the topology and growth rules of motoneuronal dendrites. *J. Comp. Neurol*. **363**, 505-516.

Fonta C., Sun, X. J. & Masson, C. (1993) Morphology and spatial distribution of bee antennal lobe interneurones responsive to odours. *Chem. Senses*, **18(2)**, 101-119.

Hollingworth, T. & Berry, M. (1975) Network analysis of dendritic fields of pyramidal cells in neocortex and Purkinje cells in the cerebellum of the rat, *Phil.Trans.R.Soc.(Lond) B* **270**, 227-264.

Ireland, W., Heidel, J. & Uemura, E. (1985) A mathematical model for the growth of dendritic trees. *Neurosci. Lett.* **54**, 243-249.

Johnson, R. A. & Bhattacharyya, G. K. (1996) Statistics, principles and methods. John Wiley & Sons.

Kliemann, W. (1987) A stochastic dynamical model for the characterization of the geometrical structure of dendritic processes. *Bull. Math. Biol.* **49**, 135-152.

Li, G. H., Qin, C., & Wang, L. W. (1992) Neurite branching pattern formation : modelling and computer simulation. *J. Theor. Biol*. **157**, 463-486.

Li, G. H., Qin, C. & Wang, L. W. (1995) Computer model of growth cone behavior and neuronal morphogenesis. *J. Theor. Biol.* **174**, 381-389.

Li, G. H. & Qin, C. D. (1996) A model for neurite growth and neuronal morphogenesis. *Math. Biosci*. **132**: 97-110.

Nowakowski, R. S., Haye, N. L. & Egger, M. D. (1992) Competitive interactions during dendritic growth: a simple stochastic growth algorithm. *Br. Res*. **576**, 152-156.

Percheron, G. (1979) Quantitative analysis of dendritic branching. A Simple formulae for the quantitative analysis of dendritic branching. *Neurosci*. *Lett.* **14**, 287-293.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992) *Numerical Recipes in C*. Cambridge University Press.

Sadler, M. & Berry, M. (1989) Topological link-vertex analysis of the growth of Purkinje cell dendritic trees in normal, Reeler and Weaver mice. *J*. *Comp. Neurol*. **289**, 260-283.

Stuermer, C. A. O. (1984) Rules for retinotectal terminal arborizations in the goldfish optic tectum. *J. Comp. Neurol*. **229**, 214-232.

Uemura, E., Carriquiry, A., Kliemann, W. & Goodwin, J. (1995) Mathematical modelling of dendritic growth in vitro. *Brain Res*. **671**, 187-194.

Uylings, H. B. M., Smit, G. J. & Weltman, W. A. M. (1975) Ordering methods in quantitative analysis of branching structures of dendritic trees. *Adv. Neurol*. **12**, 247-254.

Van Pelt, J. & Verwer, R.W.H. (1983) The exact probabilities of branching patterns under terminal and segmental growth hypotheses. *Bull. Math. Biol*. **48**, 197-211.

Van Pelt , J., Verwer, R. W. H., & Uylings, H. B. M. (1986) Application of growth models to the topology of neuronal branching patterns. *J. Neurosci. Meth.* **18**, 153-165.

Van Pelt, J., Uylings, H. B. M., Verwer, R. J., Pentney, R. J. & Woldenberg, M. J. (1992) Tree asymmetry : a sensitive and practical measure for binary topological trees. *Bull. Math. Biol*. **54**, 759-784.

Van Pelt, J., Dityatev, A. & Uylings, H. B. M. (1997) Natural variability in the number of dendritic segments : model-based inferences about branching during neurite outgrowth. *J. Comp. Neurol*. **387**, 325-340.

Van Pelt, J. (1997) Effect of pruning on dendritic topology. *J. Theor. Biol*. **186**, 17-32.

Van Veen, M. P. & Van Pelt, J. (1993) Terminal and intermediate segment lengths in neuronal trees with finite length. *Bull. Math. Biol*., **55**, 277-294.

Van Veen, M. P. & Van Pelt, J. (1994) Dynamic mechanisms of neuronal outgrowth. *Prog. Br. Res*. **102**, 95-108.

Verwer, R. W. H. & Van Pelt, J. (1985) Topological analysis of binary tree structures when occasional multifurcations occur. *Bull. Math. Biol*., **47(2)**, 305-316.

Verwer, R. W. H. & Van Pelt, J. (1987) Multifurcations in topological trees: growth models and comparative analysis. *Acta Stereol.* **6/III**, 399-404.

Verwer, R. W. H. & Van Pelt, J. (1990): Analysis of binary trees when occasional multifurcations can be considered as aggregates of bifurcation. *Bull. Math. Biol.* **52**, 629-641.

Verwer, R. W. H., Van Pelt, J. & Uylings, H. B. M. (1992): An introduction to topological analysis. In: *Quantitative method in neuroanatomy* (M. G. Stewart, ed) pp. 295-323. Chichester: John Wiley & Sons.

APPENDIX

1 Estimation of the splitting probability

1.1 Definitions

We suppose that the neuronal trees are random binary trees grown following a Bienaymé-Galton-Watson branching process. The splitting probability of a segment of order q is P_q, the probability of generation of two segments of order q+1, probability which depends on the order. The only alternative for such a segment is not to branch, with probability (1 - P_q). The number of segments at order q is labelled N_q.

1.2 Probability of observing $N_q = 2n$ segments at order q

The segments are supposed to grow *independently* of each other; this assumption allows us to express the conditional probability that the number of segments at order q+1 equals $2n$, knowing that at order q we have $2k$ segments, as:

$$\boxed{\begin{aligned} P\left(N_{q+1}=2n \middle| N_q=2k\right) &= \binom{2k}{n}\cdot P_q^n\left(1-P_q\right)^{2k-n} \quad \text{if } n \le 2k \\ P\left(N_{q+1}=2n \middle| N_q=2k\right) &= 0 \quad \text{if } n > 2k \end{aligned}} \qquad [1]$$

This probability corresponds exactly to the probability to observe n intermediate segments among the $2k$ segments at order q (binomial law). Relation [1] can be written as:

$$P\left(N_{q+1}=2n \middle| N_q=2k\right) = \mathbf{1}_{\{2k\ge n\}}\cdot\binom{2k}{n}\cdot P_q^n\left(1-P_q\right)^{2k-n} \qquad [2]$$

where $\mathbf{1}_{\{2k\ge n\}} = \begin{cases} 1 & \text{if } 2k \ge n \\ 0 & \text{elsewhere} \end{cases}$

This allows us to write the unconditional probability under the following recurrent form:

$$\boxed{P\left(N_{q+1}=2n\right) = \sum_{k=I\left(\frac{n+1}{2}\right)}^{N_1\cdot 2^{q-2}} P\left(N_q=2k\right)\cdot\binom{2k}{n}\cdot P_q^n\left(1-P_q\right)^{2k-n}} \qquad [3]$$

where N_1 is the number of primary neurites and ($N_1 \cdot 2^{q-1}$) is the maximal possible value for $N_q = 2k$, (k must be lower than ($N_1 \cdot 2^{q-2}$)); $I\left(\frac{n+1}{2}\right)$ is the integer part of $\frac{n+1}{2}$, the minimal possible value of k (which is $\frac{n}{2}$, if n is even

and $\frac{n+1}{2}$ when *n* is odd).

1.3 Recurrent expression of N_q and estimation of P_q

By definition, the first moment of N_{q+1} is:

$$E\left[N_{q+1}\right]=\sum_{n=0}^{\infty}2n\cdot P\left(N_{q+1}=2n\right) \quad [4]$$

remembering that: when $n > N_1 \cdot 2^{q-1}$, $P\left(N_{q+1}=2n\right)=0$.

Applying the recurrent relation [3], we have:

$$E\left[N_{q+1}\right]=\sum_{n=0}^{\infty}2n\cdot\sum_{k=0}^{\infty}1_{\{2k\geq n\}}P\left(N_q=2k\right)\cdot\binom{2k}{n}\cdot P_q^n\left(1-P_q\right)^{2k-n} \quad [5]$$

$$E\left[N_{q+1}\right]=\sum_{k=0}^{\infty}2P\left(N_q=2k\right)\sum_{n=0}^{\infty}1_{\{n\leq 2k\}}n\cdot\binom{2k}{n}\cdot P_q^n\left(1-P_q\right)^{2k-n} \quad [6]$$

$$E\left[N_{q+1}\right]=\sum_{k=0}^{\infty}2P\left(N_q=2k\right)\cdot\sum_{n=0}^{2k}n\cdot\binom{2k}{n}\cdot P_q^n\left(1-P_q\right)^{2k-n} \quad [7]$$

As: $$n\binom{2k}{n}=n\frac{2k!}{n!(2k-n)!}=2k\frac{(2k-1)!}{(n-1)!(2k-n)!}=2k\binom{2k-1}{n-1} \quad [8]$$

then: $$E\left[N_{q+1}\right]=\sum_{k=0}^{\infty}2P\left(N_q=2k\right)2k\cdot P_q\cdot\sum_{n=1}^{2k}\binom{2k-1}{n-1}\cdot P_q^{n-1}\left(1-P_q\right)^{2k-n} \quad [9]$$

and: $$\sum_{n-1=0}^{2k-1}\binom{2k-1}{n-1}\cdot P_q^{n-1}\left(1-P_q\right)^{2k-n}=1 \quad [10]$$

then: $$E\left[N_{q+1}\right]=2P_q\cdot\sum_{k=0}^{\infty}2k\cdot P\left(N_q=2k\right) \quad [11]$$

Finally:

$$\boxed{E\left[N_{q+1}\right]=2P_q\cdot E\left[N_q\right]} \quad [12]$$

Consequently, an *unbiased estimator* of P_q is $\frac{E\left[N_{q+1}\right]}{2E\left[N_q\right]}$

2 Estimation of the degree probability

2.1 Probability of having a given tree

Consider the case where P_q decreases with q, in such a way that: $\exists\,\tilde{q}\,/\,\forall\,q>\tilde{q}, P_q<\frac{1}{2}$. As a consequence, the tree is necessarily finite (Athreya & Ney, 1972) and we can define its maximal order $q_{\max}$. A binary tree can be described, for instance, by the number of its branches at each order: $\left\{N_1;N_2;N_3;.....;N_q;...N_{q_{max}}\right\}$. Note that *such a description is not univocal*:

different trees may have the same description. Now we want to express the conditional probability P of having a given set of tree, knowing $N_1 = n_1$:

$$P\left(N_2 = 2n_2; N_3 = 2n_3;...; N_q = 2n_q;...N_{q_{\max}} = 2n_{q_{\max}}; N_{q_{\max}+1} = 0 \middle| N_1 = n_1\right)$$

simplified as: $P\left(2n_2;2n_3;...;2n_q;...2n_{q_{\max}};0_{q_{\max}+1} \middle| n_1\right)$.

if $\exists i \in \{1,....,q\max\} \ / \ n_{i+1} > 2n_i$, $P\left(2n_2;2n_3;.....;2n_q;...2n_{q_{\max}};0_{q_{\max}+1} \middle| n_1\right) = 0$.

if $n_{i+1} \le 2n_i, \forall i \in \{1,....,q_{\max}\}$, then according to Bayes' theorem (see, e.g., Johnson & Bhattacharyya):

$$\begin{aligned} &P\left(2n_2;2n_3;..;2n_q;..2n_{q_{\max}};0_{q_{\max}+1} \middle| n_1\right) \\ &= P\left(0_{q_{\max}+1} \middle| 2n_{q_{\max}}\right) \cdot P\left(2n_{q_{\max}} \middle| 2n_{q_{\max}-1}\right) \cdot ..... \cdot P\left(2n_3 \middle| 2n_2\right) \cdot P\left(2n_2 \middle| n_1\right) \end{aligned} \quad [13]$$

As soon as $q_{\max} \ge 3$, this probability can be written as:

$$\begin{aligned} &P\left(2n_2;2n_3;..;2n_q;..2n_{q_{\max}};0_{q_{\max}+1} \middle| n_1\right) \\ &= \binom{n_1}{n_2} \cdot P_1^{n_2} \left(1-P_1\right)^{n_1-n_2} \cdot \binom{2n_2}{n_3} \cdot P_2^{n_3} \left(1-P_2\right)^{2n_2-n_3} ..... \\ &.. \cdot \binom{2n_{q_{\max}-1}}{n_{q_{\max}}} \cdot P_{q_{\max}}{}^{n_q} \left(1-P_{q_{\max}}\right)^{2n_{q_{\max}-1}-n_{q_{\max}}} \cdot \left(1-P_{q_{\max}}\right)^{2n_{q_{\max}}} \end{aligned} \quad [14]$$

or:

$$\boxed{\begin{aligned} &P\left(2n_2;2n_3;..;2n_q;..2n_{q_{\max}};0_{q_{\max}+1} \middle| n_1\right) \\ &= \binom{n_1}{n_2} \cdot P_1^{n_2} \left(1-P_1\right)^{n_1-n_2} \cdot \prod_{x=2}^{q_{\max}-1} \binom{2n_x}{n_{x+1}} \cdot P_x{}^{n_{x+1}} \left(1-P_x\right)^{2n_x-n_{x+1}} \cdot \left(1-P_{q_{\max}}\right)^{2n_{q_{\max}}} \end{aligned}} \quad [15]$$

2.2 Recurrent relation on the first moment of the degree

The degree D of a tree is the number of its extremities; in a binary tree which starts with a single segment, D is related to the total number N of segments by (Verwer *et al.*, 1992)

$$D = \frac{N+1}{2} \quad [16]$$

This rule can be easily generalised: let us take a binary tree characterised by the following sequence: $\{n_1;2n_2;2n_3;..;2n_{q_{\max}}\}$. The degree of such a tree is:

$$D = \left(n_1 - n_2\right) + \left(2n_2 - n_3\right) + \left(2n_3 - n_4\right) + .. + 2n_{q_{\max}} = n_1 + \sum_{x=2}^{q_{\max}} n_x \quad [17]$$

while the total number of segments is:

$$N = n_1 + 2\sum_{x=2}^{q_{\max}} n_x \quad [18]$$

thus :

$$D = \frac{N + n_1}{2} \qquad [19]$$

If $P\left\{2n_2;2n_3;..;2n_{q_{\max}} \middle| n_1\right\}$ is the probability of occurrence of a tree with the sequence $\left\{2n_2;2n_3;..;2n_{q\max} \middle| n_1\right\}$ knowing $N_1 = n_1$, the first moment of the degree of such a tree (with n_1 segments at order 1) is:

$$\boxed{E[D] = \sum_{q_{\max}=2}^{\infty} \sum_{n_2=1}^{n_1} \sum_{n_3=1}^{2n_2} .. \sum_{n_{q\max}}^{2n_{q\max-1}} P\left\{2n_2, 2n_3 .. 2n_{q_{\max}} \middle| n_1\right\} \cdot \left(n_1 + \sum_{x=2}^{q_{\max}} n_x \right) + P\left\{0_2 \middle| n_1\right\} \cdot n_1} \quad [20]$$

with, for $q_{\max} \geq 3$,

$$P\left\{2n_2, 2n_3 .. 2n_{q_{\max}} \middle| n_1\right\} = \binom{n_1}{n_2} P_1^{n_2} \left(1 - P_1\right)^{n_1 - n_2} \cdot \prod_{x=2}^{q_{\max}-1} \binom{2n_x}{n_{x+1}} P_x^{n_{x+1}} \left(1 - P_x\right)^{2n_x - n_{x+1}} \cdot \left(1 - P_{q_{\max}}\right)^{2n_{q_{\max}}} \qquad [21]$$

as previously seen (equation [15])

and , for $q_{\max} = 2$, $P\left\{2n_2 \middle| n_1\right\} = \binom{n_1}{n_2} P_1^{n_2} \left(1 - P_1\right)^{n_1 - n_2} \left(1 - P_{q_{\max}}\right)^{2n_{q_{\max}}}$.

Equation [20] defines the first moment of the degree for a tree starting at order 1 with n_1 segments. In the following, we consider the degree of a partition starting with *a single* segment at order *q*. The mean value of the degree at order *q* is labelled $E_q[D]$, which can be written as:

$$E_q[D] = \sum_{q_{\max}=q+1}^{\infty} \sum_{n_{q+1}=1}^{1} \cdots \sum_{n_{q_{\max}}=1}^{2n_{q_{\max}-1}} P\left\{2n_{q+1}, 2n_{q+2} .. 2n_{q_{\max}} \middle| 1_q\right\} \cdot \left(1 + \sum_{x=q+1}^{q_{\max}} n_x \right) + P\left\{0_{q+1} \middle| 1_q\right\} \cdot 1 \qquad [22]$$

In order to simplify this expression, let us write a sub-tree beginning at order *q* $\left\{2n_{q+1};2n_{q+2};..;2n_{q_{\max}} \middle| 1_q\right\}$ as $\left\{2n_x \middle| 1_q\right\}_{q_{\max}}$, and the set of all possible configurations of sub-trees starting at order *q* and *ending at order* $q_{\max} > q$, $\left\{n_x^q\right\}_{q_{\max}}$, then:

$$\sum_{q_{\max}=q+1}^{\infty} \left(\sum_{\left\{n_x^q\right\}_{q\max}} P\left\{2n_x \middle| 1_q\right\}_{q_{\max}} \right) + \left(1 - P_q\right) = \sum_{\left\{n_x^q\right\}} P\left\{2n_x \middle| 1_q\right\} = 1 \qquad [23]$$

where $\left\{2n_x \middle| 1_q\right\}$ represents one of *all* the possible sub-trees starting with one segment at order *q*, and $\left\{n_x^q\right\}$ represents the whole set of these sub-trees, with all the possible $q_{\max}$, from *q* to ∞.

We can rewrite the first moment of the degree as:

$$E_q[D] = \sum_{q_{\max}=q+1}^{\infty} \sum_{\{n_x^q\}_{q_{\max}}} P\{2n_x|1_q\}_{q_{\max}} \cdot \left(1 + \sum_{x=q+1}^{q_{\max}} n_x\right) + (1-P_q)\cdot 1$$

$$= \sum_{\{n_x^q\}} P\{2n_x|1_q\} \left(1 + \sum_{x=q+1}^{\infty} n_x\right)$$

From this expression, we can now derive the following recurrent relation:

$$\boxed{E_q[D] = (1-P_q) + 2P_q \cdot E_{q+1}[D]} \quad [24]$$

The expression for $E_{q+1}[D]$ is:

$$E_{q+1}[D] = \sum_{\{n_x^{q+1}\}} P\{2n_x|1_{q+1}\} \cdot \left(\sum_{x=q+2}^{\infty} n_x + 1\right) \quad [25]$$

which, taking into account equation [23], is equivalent to:

$$E_{q+1}[D] = \left(\sum_{\{n_x^{q+1}\}} P\{2n_x|1_{q+1}\} \cdot \left(\sum_{x=q+2}^{\infty} n_x + 1\right)\right) \cdot \sum_{\{m_x^{q+1}\}} P\{2m_x|1_{q+1}\} \quad [26]$$

then:

$$E_{q+1}[D] = \left(\sum_{\{n_x^{q+1}\}} \sum_{\{m_x^{q+1}\}} P\{2m_x|1_{q+1}\} \cdot P\{2n_x|1_{q+1}\} \cdot \left(\sum_{x=q+2}^{\infty} n_x + 1\right)\right) \quad [27]$$

which is exactly as:

$$E_{q+1}[D] = \left(\sum_{\{n_x^{q+1}\}} \sum_{\{m_x^{q+1}\}} P\{2n_x|1_{q+1}\} \cdot P\{2m_x|1_{q+1}\} \cdot \left(\sum_{x=q+2}^{\infty} m_x + 1\right)\right) \quad [28]$$

Consequently:

$$2E_{q+1}[D] = \sum_{\{n_x^{q+1}\}} \sum_{\{m_x^{q+1}\}} P\{2m_x|1_{q+1}\} \cdot P\{2n_x|1_{q+1}\} \cdot \left(\sum_{x=q+2}^{\infty} n_x + 1\right) + \sum_{\{n_x^{q+1}\}} \sum_{\{m_x^{q+1}\}} P\{2n_x|1_{q+1}\} \cdot P\{2m_x|1_{q+1}\} \cdot \left(\sum_{x=q+2}^{\infty} m_x + 1\right) \quad [29]$$

$$2E_{q+1}[D] = \sum_{\{n_x^{q+1}\}} \sum_{\{m_x^{q+1}\}} P\{2m_x|1_{q+1}\} \cdot P\{2n_x|1_{q+1}\} \cdot \left(\left(\sum_{x=q+2}^{\infty} n_x + 1\right) + \left(\sum_{x=q+2}^{\infty} m_x + 1\right)\right) \quad [30]$$

$$2E_{q+1}[D] = \sum_{\{n_x^{q+1}\}} \sum_{\{m_x^{q+1}\}} P\{2m_x|1_{q+1}\} \cdot P\{2n_x|1_{q+1}\} \cdot \left(\sum_{x=q+2}^{\infty} (n_x + m_x) + 2\right) \quad [31]$$

Then, since the numbers of segments *at each order* can be added and the splitting probabilities are *independent* for each segment:

$$\sum_{\{n_x^{q+1}\}}\sum_{\{m_x^{q+1}\}} P\{2m_x|1_{q+1}\}\cdot P\{2n_x|1_{q+1}\}\cdot\left(\sum_{x=q+2}^{\infty}(n_x+m_x)+2\right)$$
$$=\sum_{\{n_x^{q+1}+m_x^{q+1}\}} P\{2(m_x+n_x)2_{q+1}\}\cdot\left(\sum_{x=q+2}^{\infty}(n_x+m_x)+2\right) \quad [32]$$

Introducing the variable $l_x = m_x + n_x$:

$$2E_{q+1}[D]=\sum_{\{l_x^{q+1}\}} P\{2l_x|2_{q+1}\}\cdot\left(\sum_{x=q+2}^{\infty} l_x+2\right) \quad [33]$$

which is equivalent to:

$$2E_{q+1}[D]=\sum_{\{l_x^{q+1}\}} P\{2l_x|2_{q+1}\}\cdot\left(\sum_{x=q+2}^{\infty}(l_x)+1_{q+1}+1\right) \quad [34]$$

$P\{2l_x|2_{q+1}\}$, the probability of a given sub-tree starting with two segments at order q+1 is related to $P\{2_{q+1},2l_x|1_q\}$, the probability of the same sub-tree, but beginning with a single segment at order q *and* splitting into two segments of order q+1 by the following relation:

$$P\{2_{q+1},2l_x|1_q\}=P\{2l_x|2_{q+1}\}\cdot P\{2_{q+1}|1_q\}=P\{2l_x|2_{q+1}\}\cdot P_q \quad [35]$$

thus, as long as $P_q \neq 0$:

$$2E_{q+1}[D]=\frac{1}{P_q}\sum_{\{l_x^{q+1}\}} P\{2_{q+1},2l_x|1_q\}\cdot\left(\sum_{x=q+2}^{\infty} l_x+1_{q+1}+1\right)$$
$$=\frac{1}{P_q}\sum_{\{l_x^{q1},l_{q+1}\neq 0\}} P\{2l_x|1_q\}\cdot\left(\sum_{x=q+1}^{\infty} l_x+1\right) \quad [36]$$

as: $$\sum_{\{l_x^q\}} P\{2l_x|1_q\}=\sum_{\{l_x^q,l_{q+1}\neq 0\}} P\{2_{q+1},2l_x|1_q\}+P\{0_{q+1}|1_q\} \quad [37]$$

then:

$$P_q\cdot 2E_{q+1}[D]+(1-P_q)=\sum_{\{l_x^{q1},l_{q+1}\neq 0\}} P\{2l_x|1_q\}\cdot\left(\sum_{x=q+1}^{\infty} l_x+1\right)+P\{0_{q+1}|1_q\}$$
$$=\sum_{\{l_x^q,l_{q+1}\neq 0\}} P\{2l_x|1_q\}\cdot\left(\sum_{x=q+1}^{\infty} l_x+1\right)+P\{0_{q+1}|1_q\}\cdot(0_{x=q+1}+1) \quad [38]$$
$$=\sum_{\{l_x^q\}} P\{2l_x|1_q\}\cdot\left(\sum_{x=q+1}^{\infty} l_x+1\right)=E_q[D]$$

This is identical to equation [24].

2.3 Recurrent expression of the degree probability distribution per order

We can develop expression [24] in order to obtain a recurrent relation for the

probability of a segment of order q to have a degree $D = d$. The definition of the mean value of D for a segment of order q is:

$$E_q[D] = \sum_{d=1}^{\infty} d \cdot p_q(D = d) \qquad [39]$$

where $p_q(D = d)$ is the probability we are looking for: $\sum_{n=1}^{\infty} p_q(D = n) = 1$.

Applying equation [39] at the order q+1, and using the recurrent relation [24], we have:

$$E_q[D] = (1 - P_q) + P_q \cdot (E_{q+1}(D) + E_{q+1}(D)) \qquad [40]$$

$$E_q[D] = (1 - P_q) + P_q \left(\sum_{m=1}^{\infty} m \cdot p_{q+1}(D = m) + \sum_{n=1}^{\infty} n \cdot p_{q+1}(D = n) \right) \qquad [41]$$

$$E_q[D] = (1 - P_q) + P_q \left(\sum_{m=1}^{\infty} m \cdot p_{q+1}(D = m) \cdot \underline{\left(\sum_{n=1}^{\infty} p_{q+1}(D = n) \right)} + \sum_{n=1}^{\infty} n \cdot p_{q+1}(D = n) \cdot \underline{\left(\sum_{m=1}^{\infty} p_{q+1}(D = m) \right)} \right) \qquad [42]$$

then:

$$E_q[D] = (1 - P_q) + P_q \left(\begin{array}{l} \sum_{m=1}^{\infty} m \cdot \sum_{n=1}^{\infty} p_{q+1}(D = n) \cdot p_{q+1}(D = m) \\ + \sum_{n=1}^{\infty} n \cdot \sum_{m=1}^{\infty} p_{q+1}(D = m) \cdot p_{q+1}(D = n) \end{array} \right) \qquad [43]$$

and

$$E_q[D] = (1 - P_q) + P_q \left(\begin{array}{l} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} m \cdot p_{q+1}(D = n) \cdot p_{q+1}(D = m) \\ + \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} n \cdot p_{q+1}(D = m) \cdot p_{q+1}(D = n) \end{array} \right) \qquad [44]$$

If we change the variables: $d = n + m$ and $k = m$, then $d - k = n$ and we obtain:

$$E_q[D] = (1 - P_q) + P_q \left(\begin{array}{l} \sum_{k=1}^{\infty} \sum_{d=1+k}^{\infty} k \cdot p_{q+1}(D = k) \cdot p_{q+1}(D = (d - k)) \\ + \sum_{k=1}^{\infty} \sum_{d=1+k}^{\infty} (d - k) \cdot p_{q+1}(D = k) \cdot p_{q+1}(D = (d - k)) \end{array} \right) \qquad [42]$$

which can be rewritten as:

$$E_q[D] = (1 - P_q) + P_q \left(\begin{array}{l} \sum_{d=2}^{\infty} \sum_{k=1}^{d-1} k \cdot p_{q+1}(D = k) \cdot p_{q+1}(D = (d-k)) \\ + \sum_{d=2}^{\infty} \sum_{k=1}^{d-1} (d-k) \cdot p_{q+1}(D = (d-k)) \cdot p_{q+1}(D = k) \end{array} \right) \quad [43]$$

and

$$E_q[D] = (1 - P_q) + P_q \cdot \sum_{d=2}^{\infty} \left(\begin{array}{l} \sum_{k=1}^{d-1} k \cdot p_{q+1}(D = k) \cdot p_{q+1}(D = (d-k)) \\ + \sum_{k=1}^{d-1} (d-k) \cdot p_{q+1}(D = k) \cdot p_{q+1}(D = (d-k)) \end{array} \right) \quad [44]$$

$$= (1 - P_q) + P_q \cdot \sum_{d=2}^{\infty} \left(\sum_{k=1}^{d-1} p_{q+1}(D = k) \cdot p_{q+1}(D = (d-k)) \cdot (k + (d-k)) \right)$$

Thus:

$$E_q[D] = (1 - P_q) + P_q \cdot \sum_{d=2}^{\infty} d \cdot \left(\sum_{k=1}^{d-1} p_{q+1}(D = k) \cdot p_{q+1}(D = d-k) \right) \quad [45]$$

The only way to obtain equation [39] at order q is to identify the probability $p_q(D = d)$ to the following terms:

$$\boxed{\begin{array}{l} p_q(D = 1) = (1 - P_q) \\ p_q(D = d > 1) = P_q \cdot \left(\sum_{k=1}^{d-1} p_{q+1}(D = k) \cdot p_{q+1}(D = d-k) \right) \end{array}} \quad [46]$$

Note that for a finite binary tree, with the maximal order $q_{\max}$, this definition of $p_q(D = d)$ is valid only for $d \leq 2^{q_{\max} - q}$: for $d > 2^{q_{\max} - q}$, $p_q(D = d) = 0$.

3 Estimation of the asymmetry probability

3.1 Expression of the asymmetry probability per order

The probability for a segment at order q to have an asymmetry index a is: $p_q(A = a) = \sum_{d=2}^{d\max_q} p_q(A = a|d) \cdot p_q(D = d)$ [47]

where $d_{\max_q}$ is the maximal possible value of the degree of a segment of order q, i.e. $d_{\max_q} = 2^{q_{\max} - q}$, $p_q(D = d)$ is the probability for a segment of order q to have a degree d (see equation [46]), and $p(A = a|d)$ is the conditional probability for a segment of order q to have the asymmetry index a knowing that it has the degree d.

$$p_q(A=a)=\sum_{d=2}^{d\max_q} p_q(A=a|d)\cdot P_q\cdot\left(\sum_{k=1}^{d-1} p_{q+1}(D=k)\cdot p_{q+1}(D=d-k)\right) \quad [48]$$

$$p_q(A=a)=P_q\cdot\sum_{d=2}^{d\max_q} p_q(A=a|d)\cdot\sum_{k=1}^{d-1} p_{q+1}(D=k)\cdot p_{q+1}(D=d-k) \quad [49]$$

From relation [46], we write:

$$\sum_{k=1}^{d-1} p_{q+1}(D=k)\cdot p_{q+1}(D=d-k)=p_q(D=d|d>1) \quad [50]$$

Replacing $p_{q+1}(D=k)\cdot p_{q+1}(D=d-k)$ by $p_q((k,d-k)|d>1)$, the probability for a segment of order *q* which has two son-segments to have the first one of degree *k* and the second one of degree (*d* - *k*), we have:

$$p_q(A=a|d)\cdot p_q(D=d|d>1)=\sum_{k=1}^{d-1} p_q(A=a|(k,d-k)\cdot p_q((k,d-k)|d>1) \quad [51]$$

and:

$$p_q(A=a)=P_q\cdot\sum_{d=2}^{d_{\max_q}}\sum_{k=1}^{d-1} p_q(A=a|(k,d-k))\cdot p_{q+1}(D=k)\cdot p_{q+1}(D=d-k) \quad [52]$$

$p_q(A=a|(k,d-k))$ can be replaced by its value :

$$p_q(A=a|(k,d-k))=\delta\left(a,\frac{2k}{d-2}\right) \quad [53]$$

where δ is the Kronecker symbol : $\begin{cases}\delta(x,y)=1 & \text{if } x=y\\ \delta(x,y)=0 & \text{otherwise}\end{cases}$

Thus, there are *two* non-zero terms in the second sum if *a* > 0,

$$p_q(A=a>0)$$
$$=P_q\cdot\sum_{d=2}^{d\max_q} 2\cdot p_{q+1}\left(D=\frac{d\cdot(1-a)}{2}+a\right)\cdot p_{q+1}\left(D=\frac{d\cdot(1+a)}{2}-a\right) \quad [54]$$

and only *one* if *a* =0:

$$p_q(A=0)=P_q\cdot\sum_{d=2}^{d\max_q} p_{q+1}\left(D=\frac{d}{2}\right)\cdot p_{q+1}\left(D=\frac{d}{2}\right) \quad [55]$$

The conditional probability for a segment to be both of degree *d* and asymmetry index *a* at order *q*, knowing that the segment is an *intermediate* one (*d* > 1) can be written as:

$$\boxed{\begin{aligned}&p_q(A=a,D=d|d>1)\\&=2^{1-\delta(a,0)}\cdot p_{q+1}\left(D=\frac{d\cdot(1-a)}{2}+a\right)\cdot p_{q+1}\left(D=\frac{d\cdot(1+a)}{2}-a\right)\end{aligned}} \quad [56]$$

with the possible following values of a:

$$a = \frac{2 \cdot k}{(d-2)} \quad \text{with } k \in \left[0, \frac{d}{2} - 1\right], \qquad \text{if } d \text{ is even}$$

$$a = \frac{2 \cdot k + 1}{(d-2)} \quad \text{with } k \in \left[0, \frac{d-1}{2} - 1\right], \qquad \text{if } d \text{ is odd}$$