Analysis of Microtubules using Growth Curve Modeling

Md. Aleemuddin Siddiqi* S. Rao Jammalamadaka*

Statistics and Applied Probability, University of California, Santa Barbara

March 1, 2006

1 Introduction

The activity of microtubules in cell structure has been of great importance in studying certain diseases and their treatments. For instance, promising anti-cancer drugs block cell division by stabilizing microtubule activity. This work analyzes microtubule activity using statistical techniques that have been developed for "Growth Curve" modeling.

2 Microtubules

Microtubules are sub-cellular structures in most plant and animal cells and play a major part in cell locomotion, cell transport and cell division. Microtubule images in living cells are acquired by fluorescence microscopy. The images are in the form of video which is essentially a stack of images taken over successive time intervals.

In each stack a certain number of microtubules are selected and tracked, i.e. their individual co-ordinate locations are noted for each image in the stack by the experimenter. It is of interest to study the dynamicity of these

*This work is supported by the National Science Foundation Grant #0331697

microtubules under different treatments and conditions. Dynamicity is a biological term to describe changes in the lengths of microtubules. So a microtubule which grows and/or shortens very rapidly would be said to be highly "dynamic". We develop rigorous statistical analyses and tests to confirm or refute various biological hypotheses about the microtubules under differing treatment conditions. For our analysis, we use growth curve modeling (see eg. [4]), which is discussed briefly in the following section.

Figure 1: *Microtubule tips are tracked manually or automatically.*

3 The Growth Variable

By observing a stack/video of the microtubules, the experimenter chooses some microtubules whose tips are easy to track. These microtubule tips are then tracked in each frame. For each microtubule in each frame the tip's pixel location $\mathbf{p}_i = (a_i, b_i)$ (where i is the frame number) is noted. The experimenter also adds the co-ordinates of the other end of the microtubule, $\mathbf{p}_0 = (a_0, b_0)$ (i.e. where the microtubule is attached and seems to grow out of) as the initial observation or the origin. Thus the Euclidean distance between the points $\mathbf{p}_0$ and $\mathbf{p}_i$ viz, $\| \mathbf{p}_i - \mathbf{p}_0 \|$ is approximately the length of the microtubule in frame number i.

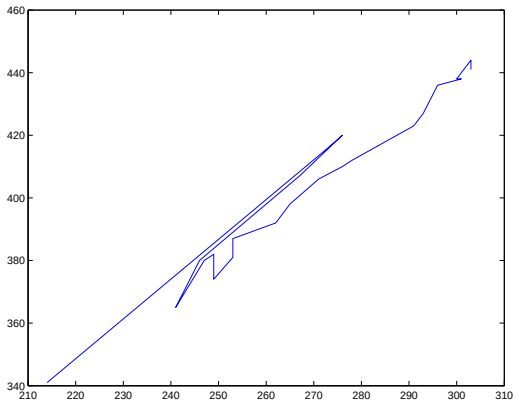

Figure 2: *The path traversed by a typical microtubule tip. The first point(origin) is the location of one end of the microtubule.*

Figure 2 is the plot of the path given by the sequence $\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, ...$ for a typical microtubule. It can be seen that the tip locations, $\mathbf{p}_1, \mathbf{p}_2, ...$ are reasonably along a straight line. Not all microtubules can be tracked for the same number of frames. Some microtubules are lost or become denatured before the others. Hence for each tip, the tracking sequence is of different lengths. As a preliminary analysis, we chose only those microtubules which could be tracked for the full 25 frames. For each microtubule we compute the length sequence $\{l_i\}$ given by,

$$l_i = \{(a_i - a_0)^2 + (b_i - b_0)^2\}^{\frac{1}{2}}; \quad i = 1, 2, ..., 25. \tag{1}$$

Now we define our "growth variable" x_i, $i = 1, 2, ..., 25$ as the cumulative change in length up till frame i i.e.,

$$x_i = \sum_{j=1}^{i} \mid l_j - l_{j-1} \mid, \tag{2}$$

where we take $l_0 = l_1$, making $x_1 = 0$. We use this variable for our growth curve analyses. Our growth curve is defined by x_i for the 25 time points, $i = 1, 2, 3, ..., 25$.

In this report we analyze the microtubule dynamics for 4 different classes -the control or uninjected, and 3 treatments.

1. **control**: These are for the normal cells with no treatments. There are 27 observations in this class.
2. **3R**: These are for cells injected with 3R tau. There are 22 observations in this class.
3. **4R**: These are for cells injected with 4R tau. There are 20 observations in this class.
4. **4RGV**: These are for cells with mutation. There are 16 observations in this class.

Figure 3 shows the plots for the growth variables for these different classes.

4 Growth Curve Modeling

Suppose that there are r different treatments or groups and x is the real valued growth variable measured at p different time points: $t_1, t_2, ..., t_p$ for n_j individuals chosen at random from the jth treatment. Potthoff and Roy [4] first introduced a nice way to treat such correlated data and labeled

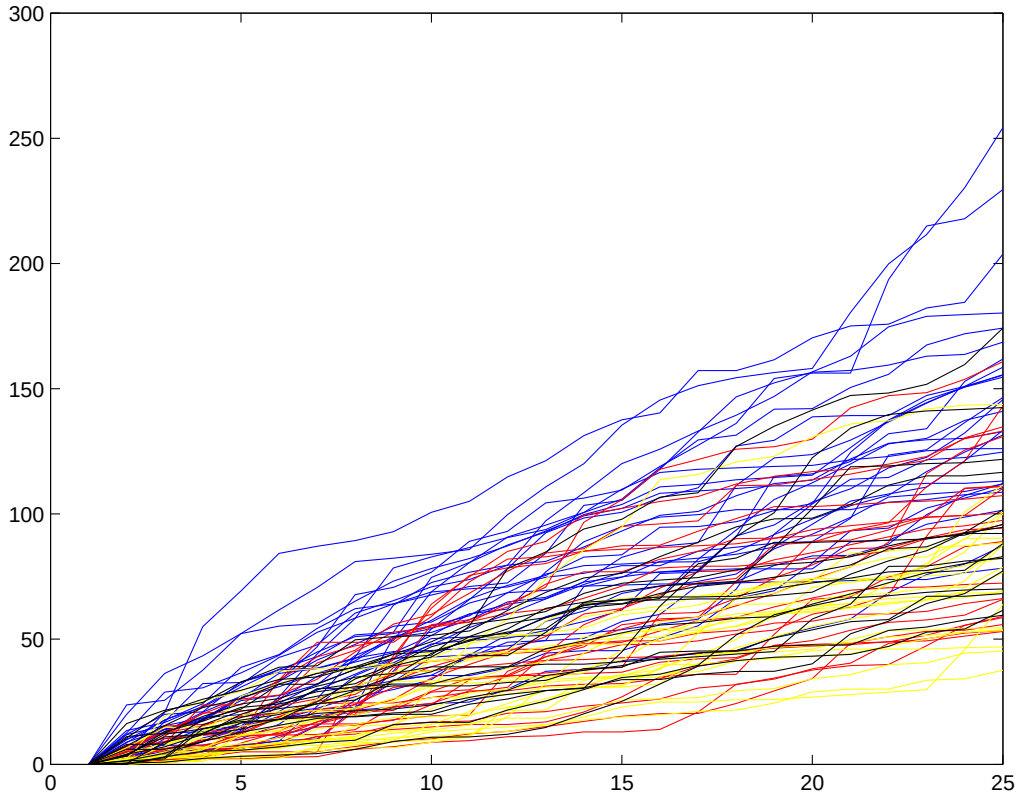

Figure 3: *blue: control, red: 3R, yellow: 4R, black: 4RGV.*

it Growth Curve Analysis. It was discussed and analyzed later by several authors significant among them, Khatri [3] and Grizzle and Allen [1]. See also Rao [5]. We specify the following polynomial regression model of degree $(q-1)$ for the growth x on the time variable t,

$$\begin{aligned} E(x_t) \quad &= \psi_{j0}t^0 + \psi_{j1}t^1 + ... + \psi_{jq-1}t^{q-1}; \\ &(t = t_1, ..., t_p; p > q-1; j = 1, 2, ..., r). \end{aligned} \tag{3}$$

Let

$$\boldsymbol{\psi}'_j = [\psi_{j0}\psi_{j1}...\psi_{jq-1}] \tag{4}$$

denote the vector of the growth curve coefficients for the j^{th} treatment. The

observations $x_{t_1}, ..., x_{t_p}$ being on the same specimen are correlated, and we shall denote their variance-covariance matrix by $\mathbf{\Sigma}$. We assume $\mathbf{\Sigma}$ to be the same for all the r groups. Let $\mathbf{X}_j$ denote the $p \times n_j$ matrix of the observations for the j^{th} group. Since our "growth variable" sequence obtained for each of the microtubule starts with $x_1 = 0$, this first value is redundant. It will make the design matrix $\mathbf{X}$ (equation 5) singular and is omitted. Let

$$\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 ... \mathbf{X}_r]. \tag{5}$$

$\mathbf{X}$ is a $p \times N$ matrix of all the observations where,

$$N = n_1 + n_2 + ... + n_r. \tag{6}$$

Therefore from Equation (3) we get,

$$\begin{aligned} E(\mathbf{X}_j) &= [\mathbf{B}\boldsymbol{\psi}_j \mathbf{B}\boldsymbol{\psi}_j ... \mathbf{B}\boldsymbol{\psi}_j] \\ &= \mathbf{B}\boldsymbol{\psi}_j \mathbf{E}_{1n_j} (j = 1, 2, ..., r), \end{aligned} \tag{7}$$

where

$$\mathbf{B} = \begin{bmatrix} t_1^0 & t_1^1 & ... & t_1^{q-1} \\ t_2^0 & t_2^1 & ... & t_2^{q-1} \\ ... & ... & ... & ... \\ t_p^0 & t_p^1 & ... & t_p^{q-1} \end{bmatrix} \tag{8}$$

and $\mathbf{E}_{ab}$ denotes, a matrix of order $a \times b$ with all elements equal to 1. $\mathbf{B}_{p \times q}$ is called the "Design Matrix". From equation 7, we get

$$\begin{aligned} E(\mathbf{X}) &= [\mathbf{B}\boldsymbol{\psi}_1 \mathbf{E}_{1n_1} | \mathbf{B}\psi_2 \mathbf{E}_{1n_2} | ... | \mathbf{B}\boldsymbol{\psi}_r \mathbf{E}_{1n_r}] \\ &= \mathbf{B}\boldsymbol{\psi}\mathbf{A}, \end{aligned} \tag{9}$$

where

$$\boldsymbol{\psi} = [\boldsymbol{\psi}_1 ... \boldsymbol{\psi}_r] \tag{10}$$

is the $q \times r$ matrix of the growth curve coefficients and

$$\mathbf{A} = \text{diag}[\mathbf{E}_{1n_1}, \mathbf{E}_{1n_2}, ..., \mathbf{E}_{1n_r}], \tag{11}$$

a block diagonal matrix with $\mathbf{E}_{1n_j}(j = 1, 2, ..., r)$ along the diagonal blocks and zeros elsewhere. $\mathbf{A}$ is of order $r \times N$.

Let Vec$\mathbf{X}$, be defined as the column vector obtained by stacking the columns of $\mathbf{X}$ one below the other. Denoting Var(Vec$\mathbf{X}$) by Var($\mathbf{X}$) we find that,

$$\text{Var}(\mathbf{X}) = \mathbf{I}_N \otimes \mathbf{\Sigma}, \tag{12}$$

where $\otimes$ denotes the Kronecker product of two matrices. Equation (9) together with Equation (12) is what is called the Growth Curve Model.

5 Growth Curve Analysis

To fit the Growth Curve model and to test various hypotheses, we perform the following computations. Obtain matrix $\mathbf{B}_2$ of order $p \times (p-q)$ such that

$$\mathbf{B}_2'\mathbf{B} = 0, \tag{13}$$

where $\mathbf{B}$ is as in Equation 8. This can be done by choosing $(p-q)$ linearly independent columns of $[\mathbf{I}_p - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}']$.

Next we compute,

$$\mathbf{S} = \mathbf{X}(\mathbf{I} - \mathbf{A}'(\mathbf{A}\mathbf{A})^{-1}\mathbf{A})\mathbf{X}', \tag{14}$$

where $\mathbf{A}$ is defined in (11). We fit the mean growth curves for these classes using an algorithm described in [3]. We would like to test the hypothesis,

$$H_0^* : \text{the degree } q-1 \text{ of the growth curves is adequate.} \tag{15}$$

To test this hypothesis we construct the MANOVA table(Table 1).

To perform the test we construct the Wilks' Λ statistic defined by,

Source	d.f.	Dispersion, order $(p-q=22)$
H_0^*	$r=4$	$\mathbf{H}_0=\mathbf{B}_2'\mathbf{X}[\mathbf{A}'(\mathbf{A}(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}]\mathbf{X}'\mathbf{B}_2$
Error	$N-r=81$	$\mathbf{E}_0=\mathbf{B}_2'\mathbf{S}\mathbf{B}_2$
Total	$N=85$	$\mathbf{H}_0+\mathbf{E}_0=\mathbf{B}_2'\mathbf{X}\mathbf{X}'\mathbf{B}_2$

Table 1: *MANOVA for Test of Specification*

$$\Lambda_0=\frac{\mid \mathbf{E}_0 \mid}{\mid \mathbf{E}_0+\mathbf{H}_0 \mid} \text{ (see table 1).} \tag{16}$$

We need the following degrees of freedom,

$$\begin{aligned} d_m &= \text{order of the error or hypothesis matrix,} \\ d_E &= \text{d.f. associated with the error matrix,} \\ \text{and } d_H &= \text{d.f. associated with the hypothesis matrix.} \end{aligned}$$

For $q=1$ we get $\Lambda_0=0.0314$. To test the null hypothesis we use the test given by [5]; namely,

$$F=\frac{1-\Lambda_0^{\frac{1}{2}}}{\Lambda_0^{\frac{1}{2}}}.\frac{ms-2\lambda}{d_H.d_m} \tag{17}$$

is approximately an F_{df_1,df_2} where

$$\begin{aligned} df_1 &= d_m.d_H \\ df_2 &= ms-2\lambda \end{aligned}$$

and

$$\begin{aligned} m &= N-\tfrac{d_m+d_H+1}{2} \\ s &= (\tfrac{(d_m.d_H)^2-4}{d_m^2+d_H^2-5})^{\frac{1}{2}} \\ \lambda &= (d_m d_H-2)/4. \end{aligned}$$

Here $p = 24$ and $N = n_1 + n_2 + n_3 + n_4 = 27 + 22 + 20 + 16 = 85$. We construct our matrix of observations $\mathbf{X}_{24\times 85}$ and our design matrix $\mathbf{B}_{24\times q}$ is obtained from Equation (8).

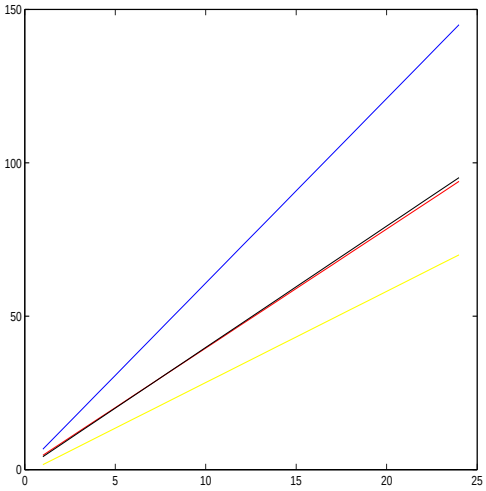

Figure 4: *blue: control, red: 3R, yellow: 4R, black: 4RGV.*

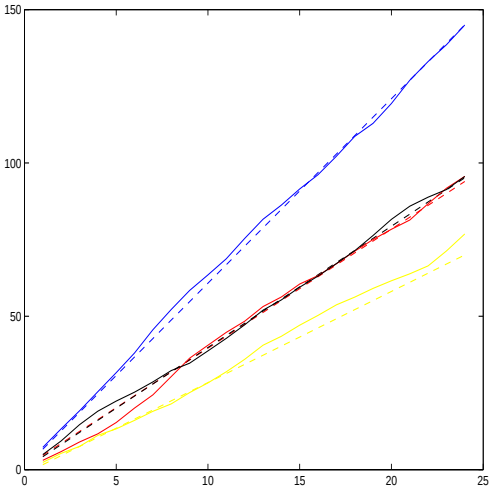

Figure 5: *The true means (solid lines) and the fitted means (dotted lines).*

For $q = 1$ we find $F = 3.5507$, with p-value$= 3.4417 \times 10^{-15}$. Hence we reject H_0^* which says that there is no growth or change over time. For $q = 2$ we find $\Lambda_0 = 0.3548$, $F = 0.8161$, with the p-value of 0.8654. This says that a linear fit for the growth-curve provides an adequate fit, and the model is well specified. Under this model that the mean growth function is linear, we show the plots for the mean curves in figure 4. Figure 5 shows how these linear fitted means for each class compare with the (nonparametric) observed means, computed as the mean at each time point for all the functions in that class. It turns out that higher order polynomial fits for the mean function do not provide any better fits and figure 6 shows the fitted curves for a 4^{th} degree polynomial. B-splines were also used to fit the data. The figure 7

shows the fits for B-splines of degree 2 and with 5 basis functions. Such spline bases are likely to be useful when dealing with non-steady-state data. For the steady state data that we have, linear fits seem to do fine as expected.

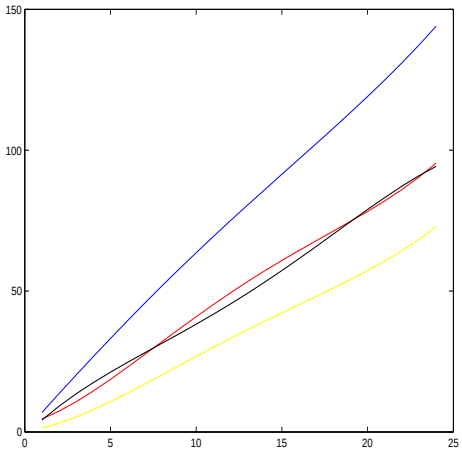

Figure 6: *The fits of order 4 ($q = 5$).*

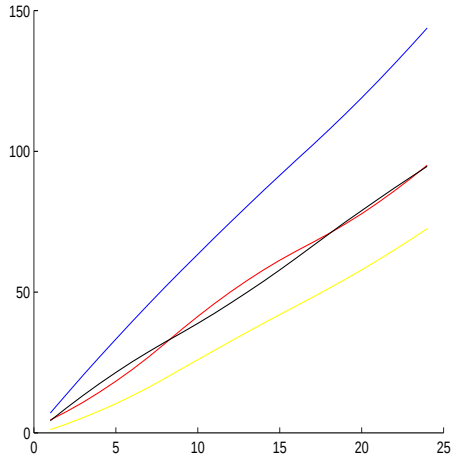

Figure 7: *The fitted spline curves.*

Next we test for differences between the classes, i.e. we test the hypothesis,

$$\begin{aligned} H_1^* \quad &: \quad \boldsymbol{\psi}_1 = \boldsymbol{\psi}_2 = ... = \boldsymbol{\psi}_r \text{ (here } r = 4) \\ &: \quad \mathbf{L}\boldsymbol{\psi}\mathbf{M} = 0; \end{aligned} \tag{18}$$

Source	d.f.	Dispersion, order $(p - q = 22)$
H_1^*	$m = 3$	$\mathbf{H}_1 = (\mathbf{L}\hat{\boldsymbol{\psi}}\mathbf{M})(\mathbf{M}'\mathbf{R}_{11}\mathbf{M})^{-1}(\mathbf{L}\hat{\boldsymbol{\psi}}\mathbf{M})'$
Error	$N - r - (p - q) = 59$	$\mathbf{E}_1 = \mathbf{L}(\mathbf{B}'\mathbf{S}^{-1}\mathbf{B})^{-1}\mathbf{L}'$
Total	$N - r - (p - q) + m = 62$	$\mathbf{H}_1 + \mathbf{E}_1$

Table 2: *MANOVA for Test of H_1^**

where,

$$\mathbf{L} = \mathbf{I}_2 \text{ and } \mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix}.$$

To test this hypothesis we construct the MANOVA table(Table 2). We calculate the following Wilks' Λ statistic,

$$\Lambda_1 = \frac{\mid \mathbf{E}_1 \mid}{\mid \mathbf{E}_1 + \mathbf{H}_1 \mid} \text{ (see Table 2)} = 0.5175. \tag{19}$$

$\mathbf{R}_{11}$ in Table 2 is defined by,

$$\mathbf{R}_{11} = (\mathbf{A}\mathbf{A}')^{-1}[\mathbf{I} + \mathbf{A}\mathbf{X}' \times \left\{\mathbf{S}^{-1}\mathbf{B}\left(\mathbf{B}\mathbf{S}^{-1}\mathbf{B}\right)^{-1}\right\}\mathbf{X}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}]. \tag{20}$$

Since $d_m = 2$, we compute the following F-statistic,

$$F_1 = \frac{1-\sqrt{\Lambda_1}}{\sqrt{\Lambda_1}}.\frac{(d_E-1)}{d_H} = 7.5428. \tag{21}$$

Under the H_1^* this has an F distribution with df=$2d_H, 2(d_E-1)$. The p-value= 7.6650×10^{-007}, which is very low. Hence H_1^* is rejected, i.e. the

Pair	Wilks' Λ	p-value
3R vs control	0.6839	0.0185
4R vs control	0.4112	1.3836×10^{-004}
4RGV vs control	0.4842	0.0043
3R vs 4R	0.8655	0.3150
3R vs 4RGV	0.9004	0.5616
4R vs 4RGV	0.8399	0.4179

Table 3: *Pair-wise comparisons*

growth curves are significantly different from each other across the various classes.

Next we perform pair-wise comparisons to see how the classes compare to one another. These were carried out by fitting the model to two classes at a time and testing the appropriate linear hypotheses for the equality of the polynomial coefficients (here linear). The table 3 summarizes the results.

The p-values show that there is significant difference between the following pairs: 3R vs control, 4R vs control and 4RGV vs control. More data might reveal differences among treatment groups like 3R vs 4R and 4R vs 4RGV.

References

[1] J. E. Grizzle and D. M. Allen (1969). *Analysis of growth and dose response curves*. Biometrics. 25, 357-381.

[2] D. G. Kabe (1974). *Generalized Sverdrup's lemma and the treatment of less than full rank regression model*. Canad. Math. Bull., 17L, 417-419.

[3] C. H. Khatri (1966). *A note on MANOVA model applied to problems in growth curves*. Ann. Inst. Stat. Math. 18, 75-86.

[4] R. Potthoff and S. N. Roy (1964). *A generalized multivariate analysis of variance models useful especially for growth curve problems*. Biometrika, 51, 313-326.

[5] C. R. Rao (1973). *Linear Statistical Inference, 2nd Edn.*. John Wiley & Sons, New York.