Optimal Estimation of Cell Movement Indices from the Statistical Analysis of Cell Tracking Data

Richard B. Dickinson and Robert T. Tranquillo
Dept. of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455

Active cell migration is essential in many physiological processes and in the function of some bioartificial tissues. Therefore, several investigators have recently attempted to quantitatively characterize random cell movement on isotropic substrata in vitro. A popular approach is to fit a theoretical expression for mean-squared cell displacement deriving from correlated random walk models to cell tracking data, yielding three objective cell movement indices: root-mean-squared speed, directional persistence time, and random motility coefficient (analogous to a molecular diffusion coefficient). The data are obtained typically by averaging cell displacements over a cell track composed of cell positions measured at equal time increments and frequently by further pooling such displacement data from tracks of different cells from the same population. We identify pitfalls introduced if an ordinary nonlinear least-squares regression analysis is used to fit the theoretical expression to the data as is commonly done and propose a generalized least-squares regression analysis as a remedy. This method estimates the cell movement indices and associated uncertainties much more accurately. It also predicts the precision of the indices based on their assumed true values and provides a means to address such issues as optimal sampling methods for data acquisition from cell tracks and handling errors associated with measuring cell position.

Introduction

Cell migration is essential to many important physiological processes such as inflammation, wound healing, embryogenesis, and tumor cell metastasis (Trinkaus, 1984). It also plays a major role in the successful function of many bioartificial tissues and organs, such as skin-equivalents (Doillon et al., 1988) and vascular grafts (Greisler, 1990). Consequently, numerous investigators have attempted to characterize cell movement *in vitro*. The central goals of these studies are variously to examine the differences in cell movement between cell types, to elucidate the mechanisms of cell locomotion, and to determine how cell migration is modulated by the cell's environment (such as medium and substratum composition). Although qualitative studies contribute progress toward these goals, quantitative studies based on the measurement of objective cell movement *indices* offer truly valid characterization and comparison of cell movement behavior.

Recently, several investigations have involved statistical characterization of cell tracks obtained from time-lapse observation of cell movement in an isotropic environment (Wilkinson et al., 1984; Takle and Lackie, 1986; Dunn and Brown, 1987; Burton et al., 1987; de Boisfleury-Chevance et al., 1989; Glasgow et al., 1989; Stokes and Lauffenburger, 1991; Saltzman et al., 1991; Buettner and Pittman, 1991; DiMilla et al., 1992). Common features revealed by cell tracks are that cell movement persists in the same direction over short times, but that significant random directional changes occur over longer times. This observation has led several investigators to mathematically model random cell movement as a correlated random walk (Gail and Boone, 1970; Nossal and Weiss, 1974; Hall, 1977; Dunn, 1983; Othmer et al., 1988). Although some of the underlying assumptions of these models vary slightly, a common feature of each is that the cell velocity, $\mathbf{v}$, at time t, has a correlation which decays exponentially with time, that is, the velocity autocorrelation function, $G_v(\tau) \equiv \langle \mathbf{v}(t+\tau) \cdot \mathbf{v}(t) \rangle$, has the form (Alt, 1990):

$$G_v(\tau) = S^2 e^{-\tau/P} \tag{1}$$

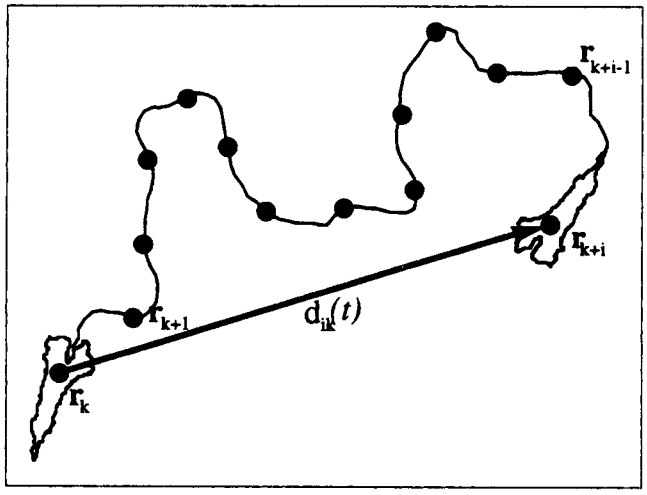


Figure 1. Discretization of a hypothetical cell track (cell outline is from Dunn and Brown, 1987).

The true cell track is measured at various time points, t_i, spaced by time increment Δt. The displacement vector $\mathbf{d}_{ik}$ is defined by $\mathbf{r}_{k+i} - \mathbf{r}_k$. Note the path exhibits directional persistence over shorter times and random walk properties over longer times.

where S and P are cell movement indices. In the terminology of Dunn (1983), S is the root-mean-squared speed, and P is the directional persistence time, the characteristic time in which cell movement persists in the same direction. S and P provide objective measures of the rate of translational movement and the characteristic time of directional persistence, respectively.

Generally, of primary interest is knowing how the displacement of the cell during a known time interval, $\mathbf{d}(t)$, is related to S and P. $\mathbf{d}(t)$ is measured from cell tracks like that depicted in Figure 1. The mean-squared displacement, $\langle d^2(t)\rangle$ can be obtained from (Alt, 1990):

$$\langle d^2(t)\rangle = \langle \mathbf{d}(t)\cdot\mathbf{d}(t)\rangle = 2\int_0^t dt' \int_0^{t'} d\tau G_v(\tau) \qquad (2)$$

which, upon substitution of Eq. 1 and integration, becomes:

$$\langle d^2(t)\rangle = 2S^2P[t - P(1-e^{-t/P})] \qquad (3)$$

This expression provides the essential relationship between the statistics of the cell track and the cell movement indices S and P. Equation 3 has two interesting limits: for $t \ll P$, the displacement is determined by purely unidirectional deterministic motion (that is, no random turning) and $\langle d^2(t)\rangle \cong S^2t^2$. For $t \gg P$, $\langle d^2(t)\rangle$ becomes linearly dependent on t with slope $2S^2P$. In an analogy to a molecular diffusion coefficient, the random motility coefficient, μ, can be defined from this limiting slope (Gail and Boone, 1970; Alt, 1990):

$$\mu = \lim_{t\to\infty} \frac{\langle d^2(t)\rangle}{2n_d t} = \frac{1}{n_d} S^2P \qquad (4)$$

where n_d is the number of dimensions that define $\mathbf{d}(t)$ (for example, $n_d = 2$ for cell movement on a planar surface, $n_d = 3$ for that in a three-dimensional matrix). Equation 4 implies

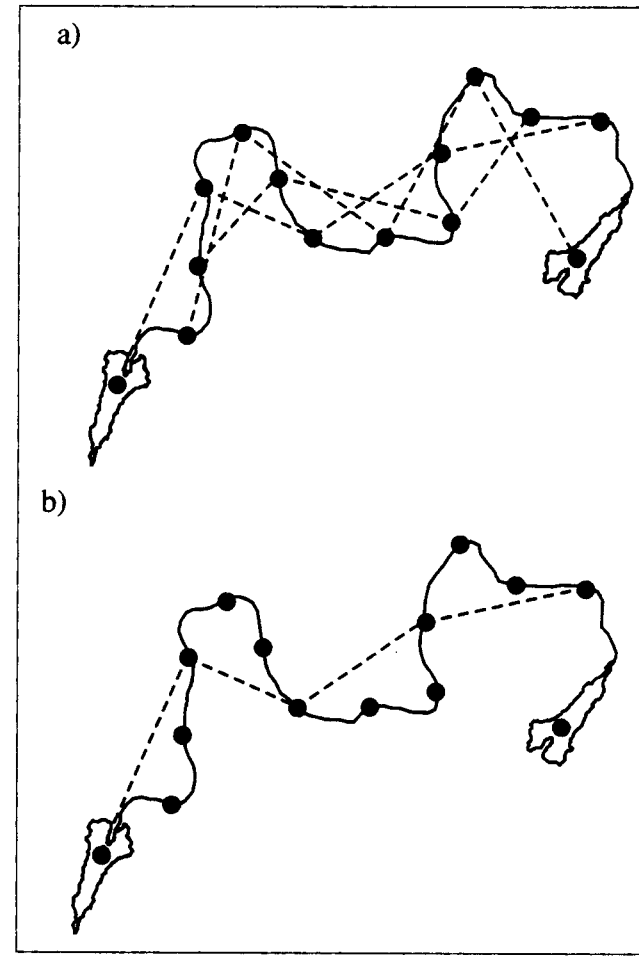


Figure 2. Methods of obtaining displacement data ($i = 3$) averaging (a) overlapping and (b) nonoverlapping intervals.

The dashed lines represent samples of displacements used in the averaging process. Note that in (a), more intervals are included in the averaging process, although the displacements overlap and, therefore, are not independent.

that Eq. 3 can also be written as:

$$\langle d^2(t)\rangle = 2n_d\mu[t - P(1-e^{-t/P})] \qquad (5)$$

To obtain estimates for S, P, and/or μ, Dunn (1983) proposed fitting Eq. 3 (or an approximation thereof) to displacement data from cell tracks with a least-squares regression analysis. As he discussed (and we have experienced), obtaining valid estimates for the indices from experimental cell tracks is not straightforward. In fact, although several investigators have employed this method, there exist several unresolved issues and potential pitfalls in its application. As we demonstrate in this work, unless these pitfalls are circumvented, such an analysis can lead to spurious results and invalid conclusions about the movement of the cells under observation.

In this article, we identify the issues and pitfalls concerning the optimal estimation of cell movement indices using ordinary nonlinear least-squares regression analysis. Specifically, we demonstrate with simulated cell tracks how standard averaging methods for obtaining mean-squared displacement data can

result in large and systematic deviation from fitted Eq. 3, and how they result in correlated and nonidentically distributed residuals that violate the assumptions behind an ordinary regression analysis. A generalized regression analysis is proposed and shown to remedy these pitfalls. We also apply this analysis to resolve such issues as optimal methods for acquiring data from cell tracks and handling measurement error in locating the cell position. Further, we apply this analysis to a case study set of tracking data for melanoma cells migrating on a type I collagen gel.

Generalized Nonlinear Least-Squares Regression Analysis

In this section, we identify pitfalls in applying an ordinary nonlinear least-squares regression (OLSR) analysis (Seber and Wild, 1989) of Eq. 3 or 5 to estimate P, S, and μ from mean-squared displacement data. We then show how these pitfalls can be remedied using a generalized nonlinear least-squares regression (GLSR) analysis which accounts for correlated and nonidentically distributed residuals. First, however, we introduce the random variables and indices of interest in the statistical analysis of cell tracks.

The cell track is assumed to consist of a sequence of n cell positions associated with a series of increasing time points differing by a constant time increment, Δt. As depicted in Figure 1, if $\mathbf{r}_k$ represents the position vector at the kth time point, then the cell displacement over time interval $t_i \equiv i\Delta t$, from $\mathbf{r}_k$ to $\mathbf{r}_{k+i}$, is:

$$\mathbf{d}_{ik} \equiv \mathbf{r}_{k+i} - \mathbf{r}_k \tag{6}$$

Let x_{ik} be the squared displacement from $\mathbf{r}_k$ to $\mathbf{r}_{k+i}$:

$$x_{ik} \equiv \mathbf{d}_{ik} \cdot \mathbf{d}_{ik} \tag{7}$$

Then, x_{ik} is considered a random variable with expected value $\eta_i \equiv \langle x_{ik} \rangle = \langle d^2(t_i) \rangle$, where η_i is the *theoretical* mean-squared displacement over t_i given by Eq. 3 or 5.

To obtain the *measured* mean-squared displacement, $\bar{x}_i$, corresponding to time interval t_i, we average several squared displacements over the cell track. The two obvious and commonly used sampling methods, overlapping and nonoverlapping time intervals, are illustrated in Figure 2. The total number of samples available from a single track, n_i, is maximized by averaging squared displacements from overlapping time intervals ($x_{i,k+1}$, $x_{i,k+2}$, $x_{i,k+3}$, and so on), such that:

$$\bar{x}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} x_{ik} \tag{8}$$

where $n_i = (n - i)$ in this case.

However, since the samples from overlapping time intervals are not statistically independent, an alternative method is to average only nonoverlapping intervals. In this case, the estimate of η_i is:

$$\bar{x}_i = \frac{1}{n_i} \sum_{k=0}^{n_i - 1} x_{i,1+ki} \tag{9}$$

where now

$$n_i = \left[\frac{n-i}{i} \right] \tag{10}$$

(the square brackets indicate the largest integer less than or equal to the quotient therein). In Figure 3, plots of $\bar{x}_i$ vs. t_i using "data" obtained from a simulated cell track by averaging displacements using nonoverlapping and overlapping intervals are compared to the theoretical curve of $\langle d^2(t_i) \rangle$ for the simulated track. (See Appendix I for a discussion of the simulation of cell tracks. Hereafter, the axes of all such plots are made dimensionless by scaling t and t_i to P, $\bar{x}_i$ to the "mean persistence length," $L \equiv SP$, and $\langle d^2 \rangle$ to L^2, where in accord with convention, the time interval t_i is implied when writing $\langle d^2(t) \rangle$.) This comparison demonstrates that the correspondence of the data to $\langle d^2(t) \rangle$ decreases with increasing t. Since these simulated data are generated from the equations of a correlated random walk model, this decrease is entirely due to less data being averaged for increasing t (n_i decreases as i increases) as well as a larger sample variance associated with the larger displacements for increasing t. This decreasing correspondence of data to $\langle d^2(t) \rangle$ is also observed for data from real cell tracks and is due *at least* to these two factors, which are inherent to cell track displacement data. Figure 3 also suggests that averaging overlapping intervals provides, in general, a better fit of $\langle d^2(t) \rangle$ to the data, but a systematic deviation for larger t. The trade-off between using overlapping vs. nonoverlapping intervals is discussed in more detail below.

To estimate S, P and μ, either Eq. 3 or 5 can be fitted to

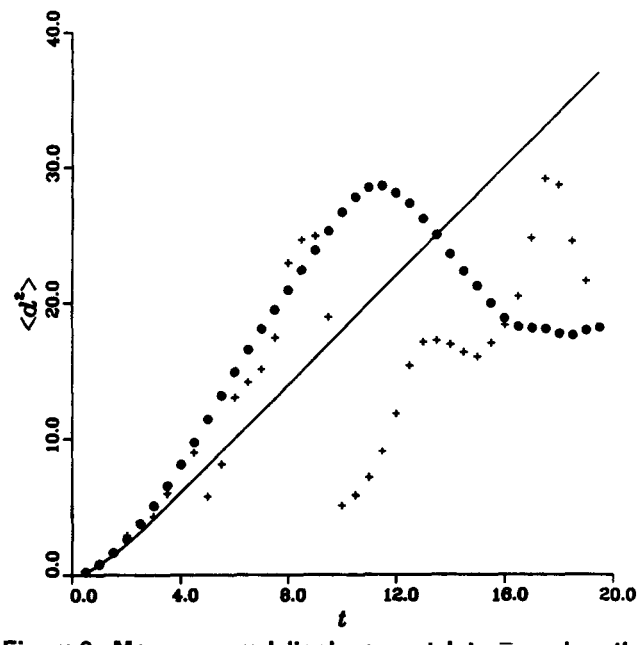


Figure 3. Mean-squared displacement data, $\bar{x}_i$, vs. length of time interval, t_i, obtained from averaging-squared displacements using overlapping (•) and nonoverlapping (+) time intervals.

Data were obtained from simulated cell tracks and are compared to the theoretical $\langle d^2(t) \rangle$ curve ($n = 40$). Note the asymptotic limits in the theoretical curve: for $t/P \gg 1$, $\langle d^2(t) \rangle$ becomes proportional to μt, and for $t \ll P$, $\langle d^2(t) \rangle$ becomes proportional to $S^2 t^2$.

the data $[\bar{x}_i, t_i]$. Defining $\boldsymbol{\Theta}$ as the vector of parameters, $\boldsymbol{\Theta} = [S\ P]$ if Eq. 3 is used in the regression analysis and $\boldsymbol{\Theta} = [\mu\ P]$ if Eq. 5 is used. OLSR yields $\boldsymbol{\Theta}$ that minimizes the sum of the square of the residuals, that is, minimizes the function:

$$F(\boldsymbol{\Theta}) = \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} \tag{11}$$

where $\boldsymbol{\epsilon}(\mathbf{t}, \boldsymbol{\Theta})$ is the vector of residuals defined by:

$$\epsilon_i \equiv \bar{x}_i - \eta_i(t_i, \boldsymbol{\Theta}) \tag{12}$$

The $\boldsymbol{\Theta}$ that minimizes Eq. 11 is the optimal estimate of the parameters assuming that the residuals are independent and identically distributed (i.i.d.) random variables (Seber and Wild, 1989). As seen in Figure 3, this assumption must be clearly violated. A striking feature of this plot is that $\bar{x}_i$ corresponding to small t_i agrees well with the theoretical $\langle d^2(t) \rangle$ curve; however, $\bar{x}_i$ corresponding to larger t_i appears to deviate significantly from $\langle d^2(t) \rangle$. This implies that including the large time data in OLSR may actually *reduce* the accuracy of the estimate of $\boldsymbol{\Theta}$. In fact, as is apparent from this example, the systematic deviation of large time data can be so severe that one might conclude that the cells do not exhibit correlated random walk behavior. The increasing deviation implies that the variance of ϵ_i is a strong, increasing function of t_i; therefore, the assumption that $\langle \epsilon_i^2 \rangle$ is equal for all i is clearly violated.

A second concern of Figure 3 is that the residuals appear to be correlated, that is, if ϵ_i is positive, then $\epsilon_{i\pm1}$ also tends to be positive. This effect is more prominent in the data from overlapping intervals; however, it is still apparent in the data from nonoverlapping intervals. The correlation between ϵ_i and ϵ_j $(i \neq j)$ results from averaging over the same portions of the cell track to obtain $\bar{x}_i$ and $\bar{x}_j$. This is evident if the displacements are constructed for $i = 4$ in Figure 2.

For these reasons, the assumption of i.i.d. residuals is strongly violated, and applying OLSR to cell displacement data can therefore result in highly inaccurate estimates of cell movement indices. Furthermore, as shown below, correlation between residuals decreases the variance of the data about the fitted curve and can result in the underestimation of the associated uncertainty in these indices.

These problems can be remedied by applying GLSR, where the possibility of correlated and nonidentically distributed residuals is rigorously handled by deriving the correlation matrix of the residuals, $\mathbf{V}$, where $v_{ij} \equiv \langle \epsilon_i \epsilon_j \rangle$. GLSR yields $\boldsymbol{\Theta}$ which minimizes the function:

$$F(\boldsymbol{\Theta}) = \boldsymbol{\epsilon}^T \mathbf{V}^{-1} \boldsymbol{\epsilon} \tag{13}$$

and does not require the assumption that the residuals are i.i.d.

The matrix $\mathbf{V}$ can be derived for overlapping intervals by noting from Eqs. 8 and 12 that:

$$v_{ij} = \frac{1}{n_i}\frac{1}{n_j}\sum_{k=1}^{n_i}\sum_{l=1}^{n_j}\omega_{l-k}^{(i,j)} = \frac{1}{n_i}\frac{1}{n_j}\sum_{k=1}^{n_i}\sum_{p=1-k}^{n_j-k}\omega_p^{(i,j)} \tag{14}$$

where $\omega_p^{(i,j)}$ is the covariance of x_{ik} and $x_{j,k+p}$ defined by:

$$\omega_p^{(i,j)} \equiv \langle x_{ik} x_{j,k+p} \rangle - \eta_i \eta_j \tag{15}$$

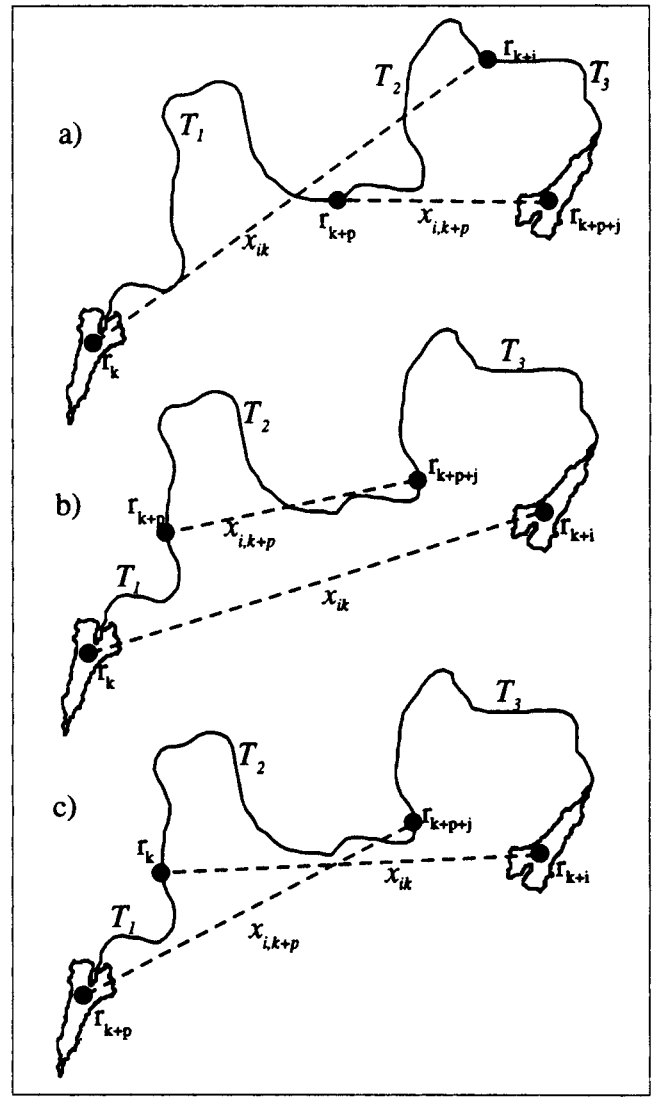


Figure 4. Definition of T_1, T_2, T_3 in evaluation of $\omega_p^{(i,j)}$ in Eqs. 17-24.

Dashed lines represent intervals with associated squared displacements x_{ik} and $x_{j,k+p}$. For $i \geq j$, three combinations of indices can lead to overlap: a) $p > 0$ and $p + j > i$, therefore time segment of overlap, T_2, equals $(i-p)\Delta t$, and those of nonoverlap, T_1 and T_3, equal $p\Delta t$ and $(p+j-i)\Delta t$, respectively (c.f. Eq. 17); b) $p > 0$ and $p + j < i$, therefore $T_2 = j\Delta t$, $T_1 = p\Delta t$, and $T_3 = (i-j-p)\Delta t$ (c.f. Eq. 18); and c) $p < 0$ therefore $T_2 = (j+p)\Delta t$, $T_1 = -p\Delta t$, and $T_3 = (i-j-p)\Delta t$ (c.f. Eq. 19).

Similarly, for nonoverlapping intervals, we have:

$$v_{ij} = \frac{1}{n_i n_j}\sum_{k=0}^{n_i-1}\sum_{l=0}^{n_j-1}\omega_{ki+1,lj+1}^{(i,j)} \tag{16}$$

To evaluate, $\omega_p^{(i,j)}$, we note that from the assumptions of the correlated random walk, x_{ik} and $x_{j,k+p}$ are correlated if, and only if, they correspond to overlapping time intervals. As shown in Figure 4, defining T_1 and T_3 as the times corresponding to nonoverlapping regions of the cell track included in samples x_{ik} and $x_{j,k+p}$, and T_2 as the time period of overlap, then we have the following possibilities (for $i \geq j$):

$$T_1 = p\Delta t \qquad T_2 = (i-p)\Delta t \qquad T_3 = (p-j+i)\Delta t$$

$$\text{if } i > p > i-j \quad (17)$$

$$T_1 = p\Delta t \qquad T_2 = j\Delta t \qquad T_3 = (i-j-p)\Delta t$$

$$\text{if } 0 \le p \le i-j \quad (18)$$

$$T_1 = -p\Delta t \qquad T_2 = (j+p)\Delta t \qquad T_3 = (i-j-p)\Delta t$$

$$\text{if } -j > p > 0 \quad (19)$$

$$T_1 = 0 \qquad T_2 = 0 \qquad T_3 = 0$$

$$\text{if } p \le -j \text{ or } p \ge i \quad (20)$$

Since $\omega_p^{(i,j)}$ is the covariance of the squared displacements with an overlapping interval of time length T_2, and nonoverlapping intervals of time lengths T_1 and T_3, then

$$\omega_p^{(i,j)} = \langle d^2(T_1+T_2)d^2(T_2+T_3)\rangle - \langle d^2(T_1+T_2)\rangle\langle d^2(T_2+T_3)\rangle \quad (21)$$

which can be evaluated by integrating the cell velocity over the cell track:

$$\omega_p^{(i,j)} = 4\int_0^{T_1+T_2} dt_1 \int_{t_1}^{T_1+T_2} dt_2 \int_{T_1}^{T_1+T_2+T_3} dt_3 \int_{t_3}^{T_1+T_2+T_3} dt_4 \, \Phi(t_1, t_2, t_3, t_4; P, S) \quad (22)$$

where

$$\Phi(t_1, t_2, t_3, t_4; P, S) = \langle [\mathbf{v}(t_1)\cdot\mathbf{v}(t_2)][\mathbf{v}(t_3)\cdot\mathbf{v}(t_4)]\rangle - \langle \mathbf{v}(t_1)\cdot\mathbf{v}(t_2)\rangle\langle \mathbf{v}(t_3)\cdot\mathbf{v}(t_4)\rangle \quad (23)$$

As shown in Appendix II, Eq. 22 can be integrated to obtain:

$$\begin{aligned}\omega_p^{(i,j)} = {} & 16S^4P^4(2-e^{-T_1/P}-e^{-T_3/P})\left[T_2/P(1+2e^{-T_2/P}) - 2(1-e^{-T_2/P}) - \frac{1}{2}(1-e^{-2T_2/P})\right] \\ & + 8S^4P^4(1-e^{-T_1/P})(1-e^{-T_2/P})[1-e^{-2T_2/P}-2T_2e^{-T_2/P}] \\ & + 32S^4P^4\left[\frac{1}{2}T_2^2/P^2 - T_2\left(\frac{5}{2}+2e^{-T_2/P}\right) + \frac{1}{4}(1-e^{-2T_2/P}) + 4(1-e^{-T_2/P})\right] \quad (24)\end{aligned}$$

Evaluating the double summation in Eq. 16 for larger i and j can be computationally expensive and is unnecessary since $\mathbf{V}$ is symmetric. Therefore, the following alternative form of Eq. 14 which contains only single summations is used in practice (from evaluating $\mathbf{V}$ only for $i \ge j$ and then setting $v_{ji} = v_{ij}$):

$$v_{ij} = \frac{1}{n_j}\omega_0^{(i,j)} + \frac{1}{n_i}\frac{1}{n_j} \times \left(\sum_{p=1}^{n_i-1}(n_i-p)(\omega_p^{(i,j)} + \omega_{-p}^{(i,j)}) + \sum_{p=n_i}^{n_j-1}(n_j-p)\omega_p^{(i,j)}\right) + \frac{1}{n_i}\frac{1}{n_j}\left(\sum_{p=1}^{i-j} p\omega_p^{(i,j)} + (i-j)\sum_{p=i-j+1}^{n_i}\omega_p^{(i,j)}\right) \quad \text{for } n_i \ge i-j \quad (25)$$

and

$$v_{ij} = \frac{1}{n_j}\omega_0^{(i,j)} + \frac{1}{n_i}\frac{1}{n_j} \times \left(\sum_{p=1}^{n_i-1}(n_i-p)(\omega_p^{(i,j)} + \omega_{-p}^{(i,j)}) + \sum_{p=i-j}^{n_j-1}(n_j-p)\omega_p^{(i,j)}\right) + \frac{1}{n_i}\frac{1}{n_j}\left(\sum_{p=1}^{n_i} p\omega_p^{(i,j)} + n_i\sum_{p=n_i+1}^{i-j-1}\omega_p^{(i,j)}\right) \quad \text{for } n_i < i-j, \quad (26)$$

In Figure 5, we compare the plots of $\bar{x}_i$ for 30 simulated cell tracks to $\langle d^2(t_i)\rangle \pm \sqrt{v_{ii}}$ for both overlapping (Figure 5a) and nonoverlapping (Figure 5b) intervals to show that Eqs. 14–16 (for $i=j$) provides an accurate estimate of the variance of the data about the theoretical $\langle d^2(t)\rangle$ curve. Our algorithm for GLSR is presented in Appendix III.

The superiority of GLSR is shown in Figure 6, where Eq. 3 is fitted to a set of squared displacement data (overlapping intervals) using OLSR, using GLSR with weighted (but not correlated) residuals (only diagonal elements of $\mathbf{V}$), and GLSR with weighted and correlated residuals (full form of $\mathbf{V}$). The solid curve is computed from Eq. 3 with the true values of the indices. The dashed curves are computed from Eq. 3 with the estimated values obtained from each method. As expected, these curves show an increasing superiority of the estimates from OLSR to GLSR (weighted residuals) to GLSR (weighted and correlated residuals). In fact, we have found that GLSR using Eq. 14 with Eqs. 25–26 will handle data for which OLSR fails due to a large systematic deviation from the theoretical $\langle d^2(t)\rangle$ curve such as that shown in Figure 3.

Before extending this analysis for multiple cell tracks, we note that, in general, a *consistent* and *unbiased estimator* (Steen, 1982) for the mean-squared displacement can be written in the form:

$$\bar{x}_i = \sum_{k=1}^{n_i} w_k^{(i)} x_{ik} \quad (27)$$

with the constraint

$$\sum_{k=1}^{n_i} w_k^{(i)} = 1 \quad (28)$$

In writing Eq. 8, we conform to the convention typically used when averaging over overlapping intervals, setting all the weighting factors, $w_k^{(i)}$, equal to $1/n_i$. However, since intervals including the two end segments of a cell track overlap with fewer adjacent intervals than those contained within the interior that do not include an end segment, their corresponding squared displacements, x_{ik}, are less correlated with other squared displacements obtained from the cell track. This implies that Eq. 8 is not necessarily the *best unbiased estimator* for the mean-squared displacement, that is, the chosen vector of weighting factors, $\mathbf{w}^{(i)} \equiv \{1/n\}$, may not minimize the *mean-squared error* of the estimator, given by the function $v_{ii}(\mathbf{w}^{(i)}) \equiv \langle(\bar{x}_i - \eta_i)^2\rangle$. The weighting factors for the *best unbiased estimator* can be determined by solving the following system of algebraic equations along with the condition of Eq. 28:

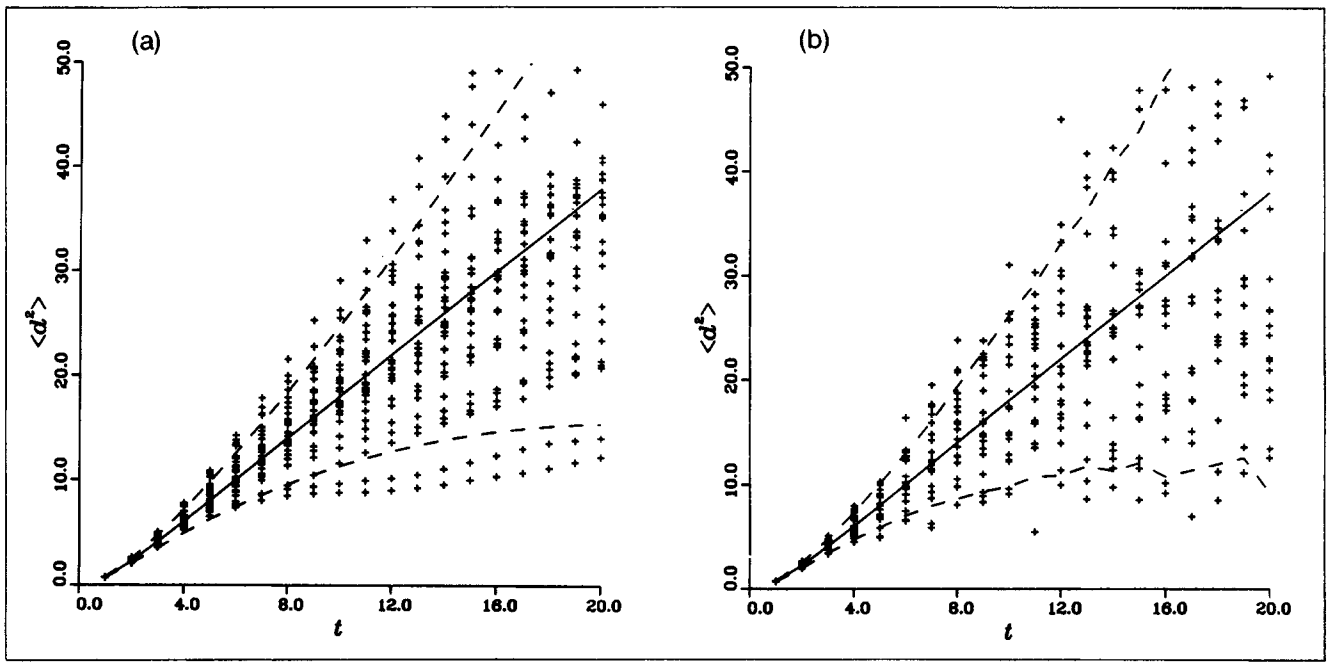


Figure 5. Mean-squared displacement for 30 simulated cell tracks where $\bar{x}_i$ data (+) is obtained by averaging displacements from (a) overlapping and (b) nonoverlapping intervals.

The solid line represents the theoretical $\langle d^2(t)\rangle$ curve, and the dashed lines represent $\langle d^2(t_i)\rangle \pm \sqrt{v_{ii}}$, showing that the scatter in the data agrees with the expected deviation of the data from the theoretical curve ($n=80$, $\Delta t=1.0$).

$$\frac{\partial v_{ii}}{\partial \mathbf{w}^{(i)}} = \mathbf{0} \tag{29}$$

Unfortunately, v_{ii} is also a function of the unknown parameters, $\boldsymbol{\theta}$; therefore, using the *best unbiased estimator* in GLSR requires an update of $\mathbf{w}^{(i)}$ and a recomputation of the data, $\bar{x}_i$, via Eq. 27 for each new estimate of $\boldsymbol{\theta}$ in the iterative process required to obtain the optimal estimate of $\boldsymbol{\theta}$ (see Appendix III). The additional computational expense thus required is substantial. However, in the case where the time interval of interest is small relative to the entire tracking time ($i \ll n$), then most intervals are overlapped to the same degree, and $w_k^{(i)}$ approaches $1/n_i$ for all k. Since the shorter intervals (large n_i) are weighted more than longer intervals (small n_i) in estimation of the cell movement indices, we do not expect the improvement in the precision of the estimates from finding the *best unbiased estimator* to be significant enough to merit the computational increase required in evaluating $\mathbf{w}^{(i)}$.

We also note that an alternative to using averaged data of the form Eq. 27 which circumvents this question is to consider *all* squared displacements, x_{ik}, as data in the regression analysis. In this case, the residuals are defined as $\epsilon_{ik} \equiv (x_{ik} - \eta_i)$, and the covariance is given by:

$$v_{i,k,j,k+p} \equiv \langle \epsilon_{ik}\epsilon_{j,k+p}\rangle = \omega_p^{(i,j)} \tag{30}$$

where $\omega_p^{(i,j)}$ is defined by Eq. 21. To incorporate this approach into a GLSR algorithm, we define $\epsilon \equiv [\epsilon_1 \ldots \epsilon_q \ldots \epsilon_{(n-2)(n+1)/2+1}]$ and $\mathbf{V} \equiv \{\epsilon_q\epsilon_r\}$ where we introduce a new numbering scheme for the residuals, $q = (i-1)(2n-i)/2 + j$. Since i ranges from 1 to $n-1$ and j from 1 to $n-i$, then the indices q and r range from 1 to $(n-2)(n+1)/2+1$.

To facilitate comparison to previous studies, we limit our analysis hereafter to the estimator defined by Eq. 8, assuming that it is sufficiently appropriate for typical applications and noting that the analysis to follow can be easily extended to the more general methods defined by Eqs. 27 and 30.

Pooled Data from Multiple Cell Tracks

To maximize the available data for a given tracking time or to characterize the average migration of a cell population, it is often necessary to obtain estimates of cell movement indices from a pooled average of displacements from multiple cell tracks (this assumes that the cell population is sufficiently homogeneous for pooling to be valid). In this case, the estimated *pooled mean-squared displacement*, $\hat{x}_i$, can be taken from averaging displacements over all cell tracks. For M cell tracks, $\hat{x}_i$ is given by:

$$\hat{x}_i = \sum_{m=1}^{M} w_{m,i}\,\bar{x}_i^{(m)} \tag{31}$$

where $\bar{x}_i^{(m)}$ is the average squared displacement from cell track m, M is the total number of tracks, and w_m is a weighting factor to account for tracks of unequal time length. This weighting factor is simply the fraction of total averaged intervals contributed from track m:

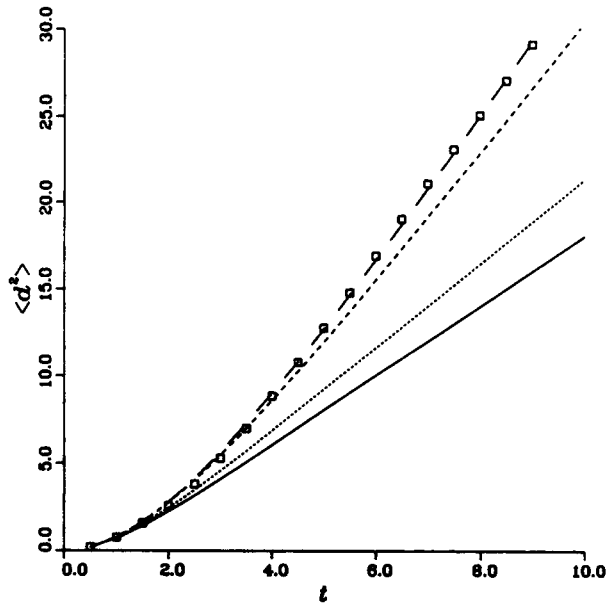


Figure 6. Mean-squared displacement computed from Eq. 3 using cell movement indices obtained from: OLSR (long dashes, $S=0.95\pm0.01$, $\mu=1.10\pm0.01$, and $P=2.43\pm0.08$); GLSR accounting for weighted but not correlated residuals (medium dashes, $S=0.97\pm0.02$, $\mu=0.95\pm0.18$, and $P=2.01\pm0.44$), and GLSR accounting for weighted and correlated residuals (short dashes, $S=1.00\pm0.02$, $\mu=0.61\pm0.20$, and $P=1.22\pm0.42$) of simulated data (unfilled squares).

These plots are compared to the theoretical $\langle d^2(t)\rangle$ curve (solid line, $S=1.000$, $\mu=0.5000$, and $P=1.000$) to show the increasing superiority of the parameter estimation ($n=40$ and $\Delta t=0.5$). Note that although OLSR-estimated indices generate a curve which closely fits the data in this case, GLSR-estimated indices generate a curve which is much closer to the theoretical curve. Also note from the above values that OLSR vastly underestimates the uncertainty in the estimates of the indices since the true values are far outside the uncertainty limits.

$$w_{m,i}=\frac{n_i^{(m)}}{\left(\sum_{k=1}^{M} n_i^{(k)}\right)} \tag{32}$$

where $n_i^{(m)}$ is the number of intervals from track m used in the averaging process. In the special case where all tracks are of equal time length, then

$$w_{m,i}=\frac{1}{M} \tag{33}$$

Since the tracks of different cells are independent, it is simple to show that the *pooled correlation matrix*, $\hat{\mathbf{V}}$, is given by:

$$\hat{v}_{ij}\equiv\langle(\hat{x}_i-\eta_i)(\hat{x}_j-\eta_j)\rangle=\sum_{m=1}^{M} w_{m,i}w_{m,j}v_{ij}^{(m)} \tag{34}$$

Again, for tracks of equal time length,

$$\hat{\mathbf{V}}=\frac{1}{M}\mathbf{V} \tag{35}$$

In the results given hereafter, $\mathbf{V}$ refers to the correlation of the residuals for $M\geq 1$.

Uncertainty in Parameter Estimates

The evaluation of $\mathbf{V}$ also provides a direct calculation of the variance-covariance matrix for the parameter estimates, $\mathbf{D}$,

$$\mathbf{D}(\mathbf{t},\boldsymbol{\Theta})\equiv\begin{bmatrix}\sigma_{\theta_1}^2 & \sigma_{\theta_1\theta_2}\\ \sigma_{\theta_2\theta_1} & \sigma_{\theta_2}^2\end{bmatrix}=(\mathbf{J}^T\mathbf{V}^{-1}\mathbf{J})^{-1} \tag{36}$$

where $\mathbf{J}(\mathbf{t};\boldsymbol{\Theta})$ is the Jacobian matrix defined by $J_{ij}=\partial\eta_i(t_i;\boldsymbol{\Theta})/\partial\theta_j$ for $i=1, n$, $j=1, 2$. The variance in the third, dependent parameter (S if $\boldsymbol{\Theta}=[\mu\ P]$, or μ if $\boldsymbol{\Theta}=[\mu\ P]$) can also be calculated from $\mathbf{D}$. For example, if S is the dependent parameter, then $\theta_1=\mu$, $\theta_2=P$, and σ_S^2 is given by:

$$\sigma_S^2=\left(\frac{\partial S}{\partial\boldsymbol{\Theta}}\right)^T\mathbf{D}\,\frac{\partial S}{\partial\boldsymbol{\Theta}} \tag{37}$$

where

$$\frac{\partial S}{\partial\boldsymbol{\Theta}}=\left[\frac{\partial S}{\partial\mu}\ \frac{\partial S}{\partial P}\right]^T.$$

Thus, the uncertainty in all three cell movement indices can be directly obtained from Eqs. 36-37. $\mathbf{D}(\mathbf{t},\boldsymbol{\Theta})$ is computed from $\boldsymbol{\Theta}$ following convergence, as indicated in Appendix III.

Since $\mathbf{D}$ can be evaluated as a function of $\boldsymbol{\Theta}$, Eqs. 36-37 allow comparison of the uncertainties in the cell movement indices using overlapping vs. nonoverlapping intervals (by choosing the appropriate form of $\mathbf{V}$, Eqs. 25-26 or Eq. 16, respectively), and provide predictions of how these uncertainties depend on M, n, and Δt. We immediately note from Eqs. 35 and 36 that for multiple cell tracks of equal time length, the elements of $\mathbf{D}$ are proportional to $1/M$; therefore, the uncertainties in the indices are proportional to $\sqrt{1/M}$. In the sections to follow, we examine the issue of optimal sampling methods with respect to the other averaging parameters (n and Δt) and then address the consequences of finite precision in measuring cell position.

Nonoverlapping vs. overlapping intervals

As stated above, the benefit of averaging overlapping intervals over nonoverlapping intervals (more samples being available for averaging) is questionable since samples obtained from overlapping intervals are clearly correlated. However, when averaging squared displacements from nonoverlapping intervals, which are not correlated, the number of samples available for averaging, n_i, decreases rapidly with increasing time interval size, i (Figure 7). Using GLSR, we plot the relative uncertainty in the estimates of μ, P, and S vs. n in Figure 8 for averaging overlapping and nonoverlapping intervals. These plots indicate that averaging overlapping intervals provides only a modest increase in precision over nonoverlapping intervals, at the expense, however, of significantly more com-

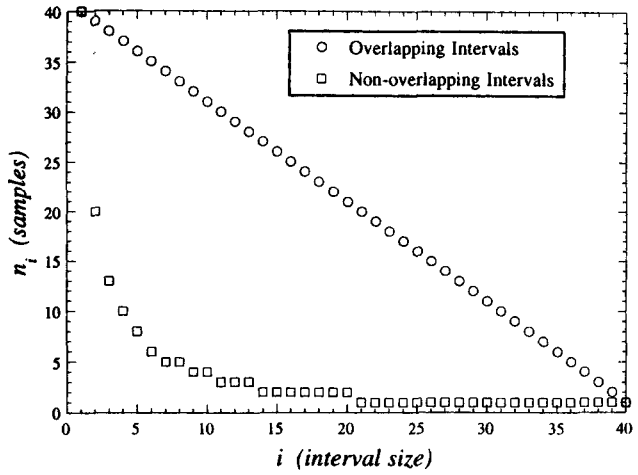


Figure 7. Number of sample intervals available for averaging, n_i vs. number of time increments in the time interval, i, for averaging overlapping and nonoverlapping intervals.

The plot shows that n_i decreases with i and is greater for overlapping intervals than nonoverlapping intervals ($n=40$).

putations in averaging the $\bar{x}_i$ and evaluating **V**. For example, the ratio of computation time required to calculate **V** for overlapping intervals using Eqs. 25–26 to nonoverlapping intervals using Eq. 14 is approximately five to one for $n=40$. For typical database sizes obtainable by current imaging technology, the required computational time for either case is significant, so the fivefold greater than required for overlapping intervals must be weighed against the added precision of the estimates.

Optimal time increment

We can also use GLSR to find the optimal time increment for the estimation of the cell movement indices of interest for the case of fixed n. In Figure 9, the relative uncertainties in the estimates of the indices are plotted vs. $\Delta t/P$ (fixing $n=40$). This plot suggests that for fixed n, the optimal time increment for estimating P and μ is $\Delta t \sim P$. However, if an optimal estimation of S is desired, then Δt should be made as small as possible. That σ_p exhibits a minimum is expected since in the limits as $t \to 0$ and $t \to \infty$, $\langle d^2(t) \rangle = S^2 t^2$ and $\langle d^2(t) \rangle = \mu t$, respectively, that is, $\langle d^2(t) \rangle$ is directly dependent on P at intermediate times only. (For $t \to \infty$, $\langle d^2(t) \rangle$ depends on P in the product $\mu = (1/n_d)S^2P$, but since only one parametric degree of freedom is allowed in this limit, S^2 and P cannot be evaluated independently.)

Positioning error

Another relevant issue is the effect of error in measuring the cell position or "positioning error," on the estimation of the cell movement indices. The derivation of Eq. 3 assumes the cell position is identifiable as a precisely measurable point which is constant within the reference frame of the moving cell. In reality, there is always error in measuring this point, and a cell in active locomotion displays morphological changes which may alter the reference frame. Therefore, positioning error may result from a limited precision in measuring the cell's reference point as well as from fluctuations in the position of this point within the reference frame of the cell. For example,

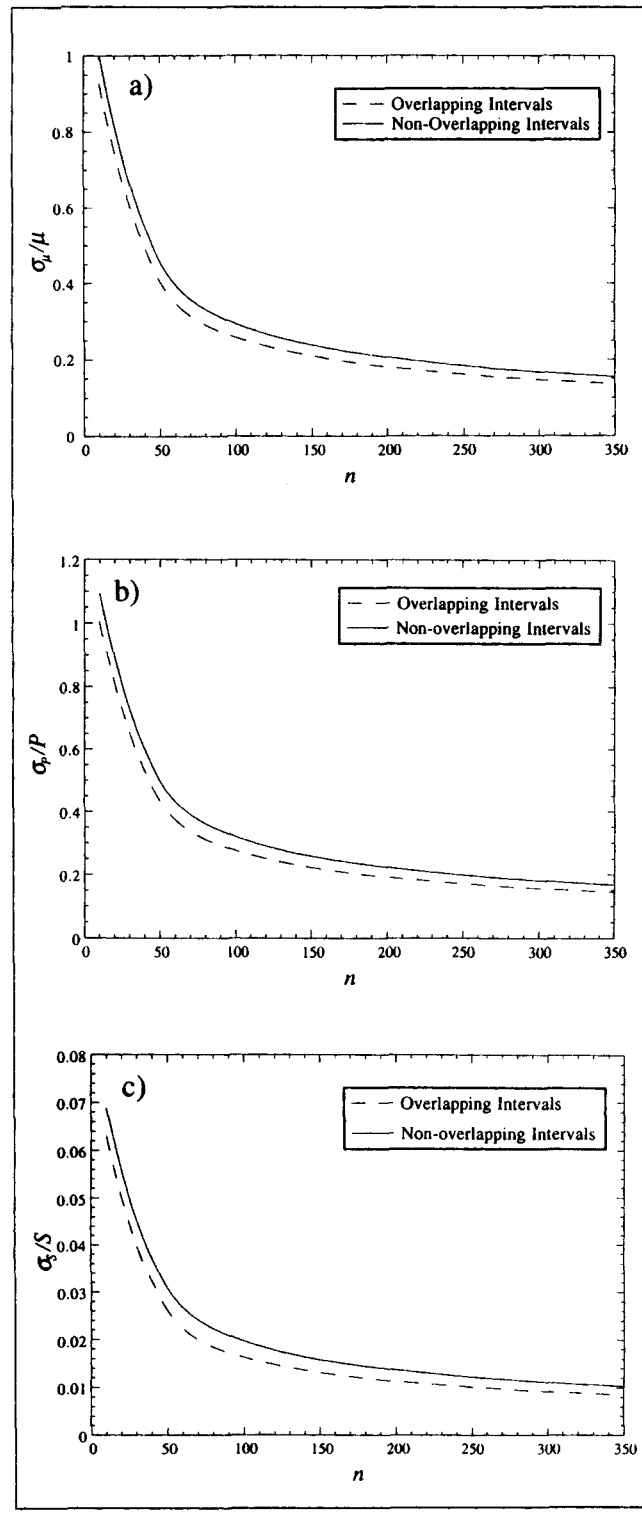


Figure 8. Relative uncertainties in GLSR estimates of (a) μ, (b) P, and (c) S vs. total number of intervals, n, for displacement data obtained from averaging nonoverlapping (solid lines) and overlapping intervals (dashed lines) ($\Delta t = P$).

Precision of the estimates increases with n, and averaging displacements over nonoverlapping intervals offer only a slight increase in the precision of the parameter estimates. $S=P=1$ are presumed parameter values to calculate **D** in Eq. 36.

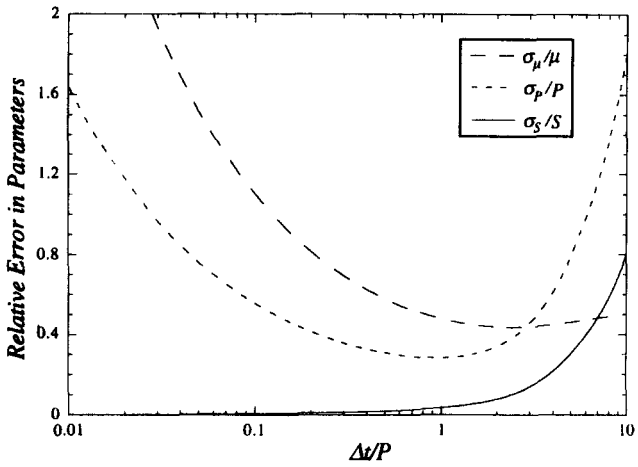


Figure 9. Relative uncertainties in GLSR estimates of μ, P and S vs. time increment, $\Delta t/P$.

These plots reveal that for fixed n, the precision in estimating μ and P is maximized for $\Delta t \sim P$, but the precision in estimating S is minimized by minimizing Δt. $S = P = 1$ are presumed parameter values.

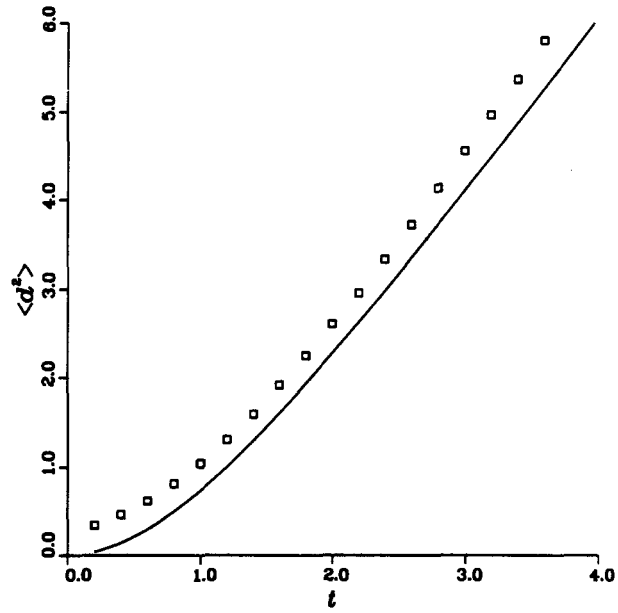


Figure 10. Effect of positioning error on the mean-squared displacement, $\bar{x}_i'$ (which are close to η_i' since n is large), vs. length of time interval, $t_i (n = 500, \Delta t = 0.2)$.

There is a systematic shift of displacement data above the theoretical $\langle d^2(t) \rangle$ curve computed from Eq. 3 which does not account for positioning error. $S = P = 1$ and $\gamma = 0.1$ were parameter values for simulating cell tracks.

if the centroid of the projected cell area is used to measure $\mathbf{r}_k$, morphological changes of the cell cause fluctuations in perfectly measured $\mathbf{r}_k$, even if the cell is not translocating. Therefore, positioning error is a function of the precision of measurement and of the dynamic morphology of the cell under observation.

We define the positioning error, $\mathbf{e}_k$, to be the difference between the measured position, $\mathbf{r}_k'$, and the position predicted by the model, $\mathbf{r}_k$:

$$\mathbf{e}_k = \mathbf{r}_k' - \mathbf{r}_k \tag{38}$$

Assuming that $\mathbf{e}_k$ are i.i.d. random variables with variance $\gamma = \langle \mathbf{e}_k \cdot \mathbf{e}_k \rangle$, the mean-squared *measured* displacement, η_i', can then be easily derived to be:

$$\eta_i' \equiv \langle x_{ik}' \rangle = \eta_i + 2\gamma \tag{39}$$

where $x_{ik}' \equiv (\mathbf{r}_{k+i}' - \mathbf{r}_k') \cdot (\mathbf{r}_{k+i}' - \mathbf{r}_k')$. Therefore, we expect the measured mean-squared displacement curve to be uniformly shifted upward by the amount 2γ, as shown in Figure 10. For data having significant positioning error, OLSR generally underestimates the directional persistence time, P, and overestimates the root-mean-squared cell speed, S.

The correlation matrix, $\mathbf{V}$, must be reformulated, noting that:

$$\langle x_{ik}' x_{jl}' \rangle - \langle x_{ik}' \rangle \langle x_{jl}' \rangle = \omega_{l-k}^{(i,j)} + (\delta_{k+i,j+l} + \delta_{k,l} + \delta_{k,l+j} + \delta_{k+i,l}) \sigma_\gamma^2 \tag{40}$$

where σ_γ^2 is defined by:

$$\sigma_\gamma^2 \equiv \langle ((\mathbf{e}_k \cdot \mathbf{e}_k)^2 - \gamma)^2 \rangle \tag{41}$$

We assume that $\mathbf{e}_k$ is normally distributed so that:

$$\sigma_\gamma^2 = 2\gamma^2 \tag{42}$$

From Eq. 40, the modified correlation matrix of the residuals, $\mathbf{V}'$, is then

$$v_{ij}' = \frac{1}{n_i} \frac{1}{n_j} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} \omega_{l-k}^{(i,j)} + (\delta_{k+i,j+l} + \delta_{k,l} + \delta_{k,l+j} + \delta_{k+i,l}) \sigma_\gamma^2 \tag{43}$$

which we find to be:

$$v_{ij}' = v_{ij} + \frac{8\gamma^2}{n-j} \tag{44}$$

Intuitively, large γ of order η_i will have a detrimental effect on the estimation of the cell movement indices, especially when using GLSR which preferentially weights smaller time data. One possible approach is to estimate 2γ and to subtract it from $\bar{x}_i'$, yielding an estimate of the "true" displacement, $\bar{x}_i''$:

$$\bar{x}_i'' = \bar{x}_i' - 2\hat{\gamma} \tag{45}$$

where $\hat{\gamma}$ is an *a priori* estimate of γ.

There is necessarily an associated uncertainty in $\hat{\gamma}$, $\sigma_{\hat{\gamma}} = \sqrt{\langle (\hat{\gamma} - \gamma)^2 \rangle}$, which must be accounted for by again modifying the correlation matrix of the residuals. Any element of the remodified correlation matrix, $\mathbf{V}''$, can be calculated to be:

$$v_{ij}'' \equiv \langle (\bar{x}_i'' - \eta_i)(\bar{x}_j'' - \eta_j) \rangle = \langle (\bar{x}_i' - 2\hat{\gamma} - \eta_i)(\bar{x}_j' - 2\hat{\gamma} - \eta_j) \rangle \tag{46}$$

Since

$$\langle \bar{x}_i' \rangle = \eta_i + 2\gamma \tag{47}$$

Eq. 46 becomes:

$$v_{ij}'' = v_{ij}' + 4\sigma_\gamma^2 \tag{48}$$

Although this method is a straightforward modification of the previous method, two parameter estimates are required *a priori*, $\hat{\gamma}$ and $\sigma_{\hat{\gamma}}^2$, and the accuracy of the regression estimates of the cell movement indices reflect the accuracy of these estimates.

A method we have found to be superior is to treat γ as a third parameter also to be estimated with GLSR. Therefore, $\Theta = [\mu, P, \gamma]$ or $[S, P, \gamma]$, and the model to be fitted becomes:

$$\eta_i'(t_i; \Theta) = 2n_d\mu[t_i - P(1 - e^{-t_i/P})] + 2\gamma \tag{49}$$

or equivalently,

$$\eta_i'(t_i; \Theta) = 2S^2P[t_i - P(1 - e^{-t_i/P})] + 2\gamma \tag{50}$$

The advantage of this method is that *a priori* estimates of γ and σ_γ^2 (that is, $\hat{\gamma}$ and $\sigma_{\hat{\gamma}}^2$) are not required (except for γ as an initial guess in the solution algorithm). Applying this method to data from simulated cell tracks shows that it provides accurate estimates for γ. In Figure 11, we plot the relative uncertainties in the indices vs. $\Delta t/P$ for both three parameter (S, P, and γ) and two parameter (S and P) fits. This plot shows that including γ as an additional parameter only slightly diminishes the precision of the estimates for $\Delta t < P$ when $\gamma^{1/2} < SP$. However, in Figure 12 we see that as $\gamma^{1/2}$ increases relative to SP, the uncertainties increase dramatically. We observe the same qualitative features of these results regardless of the value of $\Delta t/P$, reflecting the fact that for root-mean-squared positioning error, $\gamma^{1/2}$, larger than the "mean persistence length," SP, the random turning in the cell track is indistinguishable from the fluctuations in the measured cell position. In particular, the previous conclusion that the optimal Δt for estimation of S is Δt as small as possible remains valid.

Application to Melanoma Cell Tracks

We applied OLSR and GLSR to a set of 14 cell tracks obtained for melanoma cells migrating on a type I collagen gel using a cell tracking method previously described (Dickinson et al., 1993). Each track consists of 15 displacement intervals defined by a 10-min time-lapse increment ($n = 15$, $\Delta t = 10$ min). In Figure 13, fits of the theoretical $\langle d^2(t) \rangle$ curve to the $\bar{x}_i$ vs. t_i data using estimates of the cell movement indices obtained by OLSR and GLSR are presented for comparison. Displacements are averaged using overlapping intervals and all 14 tracks are pooled ($M = 14$). The nature of the fits is analogous to that seen in Figure 3 (where the "data" in that case was based on simulating a correlated random walk): OLSR yields an apparent better fit of the data than GLSR. Assuming, however, that the melanoma cell migration conforms to a correlated random walk, we again see that the estimates from OLSR are inaccurate and deceptively precise.

Discussion

Quantitative analysis of cell tracks provides objective com-

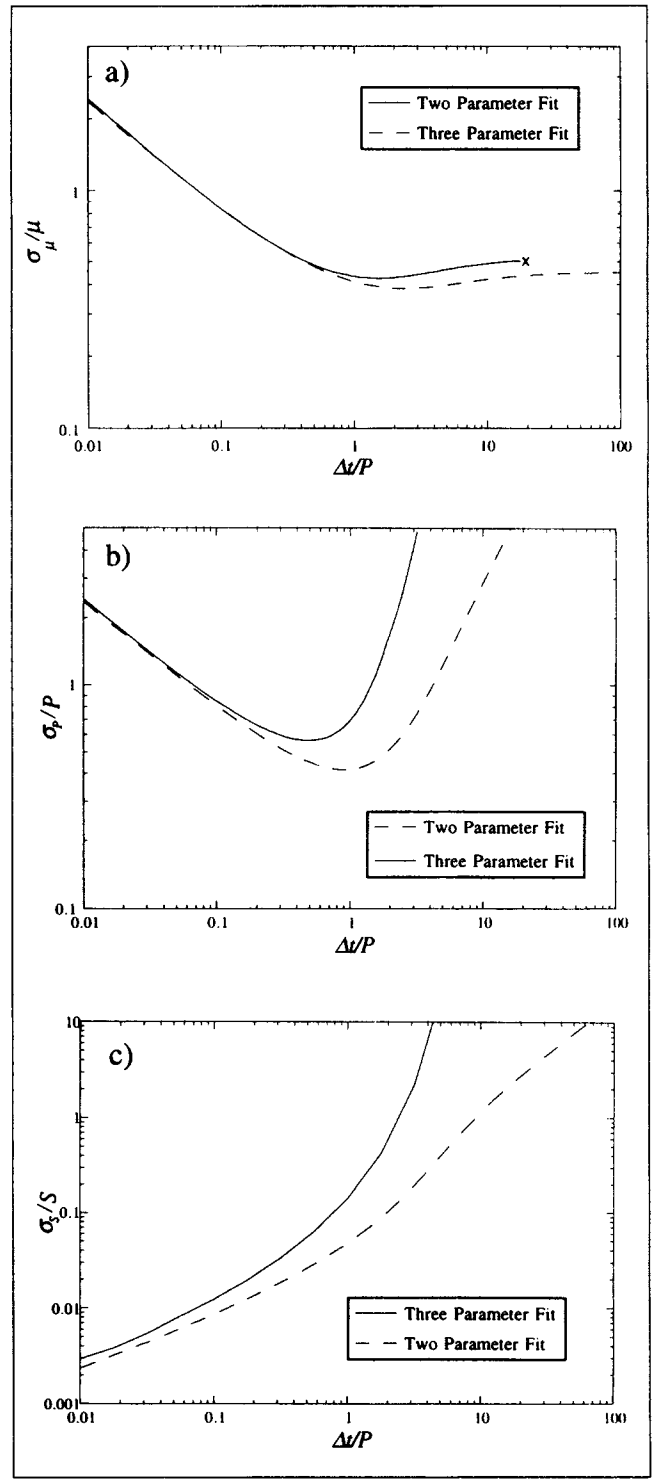


Figure 11. Relative uncertainties in GLSR estimates of (a) μ, (b) P, (c) S vs. $\Delta t/P$ for the three (Eq. 49) and two-parameter model (Eq. 3) ($n = 40$, overlapping intervals).

The additional uncertainty introduced when γ is included as a parameter in regression analysis is evident: For $\Delta t < P$, the additional uncertainty is small; however, for $\Delta t > P$, the additional uncertainty in S and P increases dramatically. The "X" in (a) indicates the point at which the uncertainties in S and P are too large for convergence of our GLSR algorithm. $S = P = 1$ and $\gamma = 0$ are presumed parameter values. The uncertainties are nearly identical to these values for nonzero γ satisfying $\gamma^{1/2} < SP$ (see Figure 12).

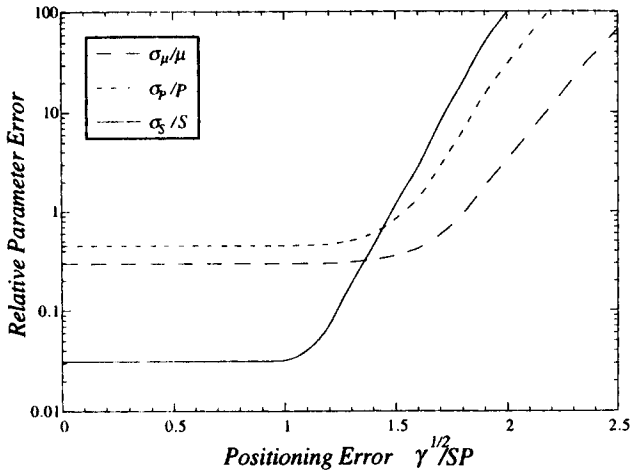


Figure 12. Relative uncertainties in GLSR estimates of μ, P and S vs. $\gamma^{1/2}/SP$.

This shows how large positioning error, $\gamma^{1/2} > SP$, induces large uncertainty in the estimates ($n = 40$, $\Delta t = 1.0$). For $\gamma^{1/2} < SP$, however, positioning error does not significantly affect when a three-parameter model (including γ) is used to fit the mean-squared displacement data.

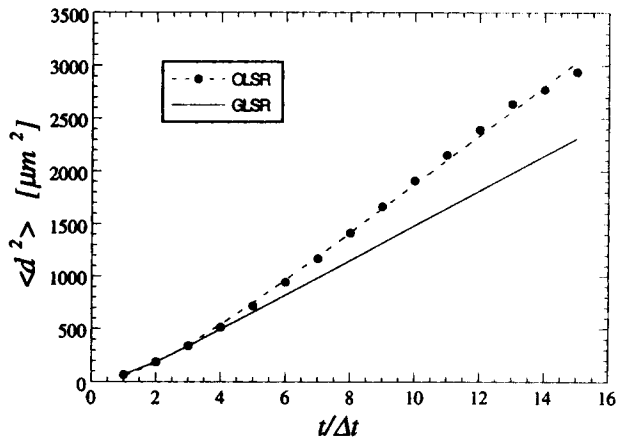


Figure 13. Mean-squared displacement (overlapping intervals) vs. time for a population of 14 murine melanoma cells moving on the surface of a reconstituted type-I collagen gel (circles) and tracked over 15 time increments ($\Delta t = 10$ min).

The solid line is the curve obtained using GLSR to fit the data, and the dashed line is that obtained using OLSR. Positioning error is taken into account via Eq. 49 in both cases. Even when pooling data from multiple cell tracks, OLSR leads to erroneous estimation of the cell movement indices. In this case, the OLSR analysis predicts $\mu = 5.80 \pm 0.14$ μm^2/min, $P = 19.8 \pm 3.5$ min, $S = 0.765 \pm 0.036$ μm/min, and $\gamma = 5.34 \pm 19.2$ μm^2; however, GLSR predicts $\mu = 4.14 \pm 0.76$ μm^2/min, $P = 11.2 \pm 3.1$ min, $S = 0.869 \pm 0.049$ μm/min, and $\gamma = 5.2 \pm 1.4$ μm^2. Reported uncertainties are the standard deviations of parameter estimates.

parisons of the migration properties of cells under different conditions and yields insight to the underlying mechanisms of locomotion. Dunn (1983) introduced a procedure to characterize tracks for cells exhibiting random migration in an isotropic environment in terms of three statistical quantities: the root-mean-squared cell speed, S, the directional persistence time, P, and the random motility coefficient, μ. These cell movement indices are defined in a theoretical expression (Eq. 3) for the mean-squared cell displacement, $\langle d^2(t) \rangle$, vs. time, t, that is derived from correlated random walk models. They can be estimated by fitting Eq. 3 to displacement data from cell tracks. Dunn (1983) described the fitting of approximations to Eq. 3 valid for certain ranges of t with linear regression. Most recent investigations are based on fitting Eq. 3 over the full range of t with an ordinary nonlinear least-squares regression analysis. However, several pitfalls exist because OLSR assumes independent and identically distributed residuals. We demonstrate herein how OLSR can result in spurious and inaccurate results because these assumptions are grossly violated. Also, issues concerning optimal acquisition of displacement data remain unresolved.

To remedy this situation, we propose here a generalized nonlinear least-squares regression analysis to account for unequally distributed and correlated residuals. This more accurate analysis requires the correlation matrix of the residuals, $\mathbf{V}$, which we derive from the same underlying stochastic model from which Eq. 3 is derived. The deceptively good fit that can be obtained from OLSR due to correlation of the residuals is striking, as revealed by GLSR performed on the same simulation data (Figure 6) and experimental data (Figure 12). That outcome, leading to estimates of cell movement indices of poor accuracy and deceptively high precision, is conclusive based on analysis of the correlated random walk simulation data. Analysis of the experimental data yields the same outcome, with OLSR yielding significantly different estimates of μ(+40%), P (+77%) and S (−12%) (derived from μ and P) and wrongly indicating much greater precision for μ (5 times smaller standard deviation) as compared to GLSR (in this case the analysis requiring the assumptions that the melanoma cell migration conforms to a correlated random walk and the cell population is homogeneous to validate pooling the tracks).

In addition to GLSR providing a superior accuracy in estimation of the cell movement indices and their associated uncertainties, this analysis also can predict these uncertainties based on presumed values of the indices, independent of data. This allows us to address unresolved data acquisition issues such as the optimal time increment and the relative advantage of averaging overlapping intervals vs. nonoverlapping intervals in obtaining displacement data. We find that for fixed n, the optimal time increment, Δt^*, for estimating P and μ is approximately $\Delta t^* \sim P$; however, the uncertainty in μ becomes asymptotically independent of Δt for $\Delta t \gg P$, but not significantly greater than that for $\Delta t \sim P$. In contrast, Δt^* for estimating S is Δt as small as possible. In addition, we find that averaging overlapping intervals provides only a modest improvement in the precision of the estimates, despite a much greater number of samples available in the averaging process. The trade-off in this improvement is an increased computation time in performing the regression analysis. Averaging overlapping intervals also introduces greater correlation in the residuals which we expect would decrease the accuracy of estimated cell movement indices if OLSR is used.

We also exploit GLSR to address how imprecise measurement of cell position or fluctuations of the reference point on the cell relative to its "true position" can impair estimation of cell movement indices. Two methods are proposed to address this problem. The first involves estimation of the positioning error *a priori*. However, this method has drawbacks

that the associated uncertainty in this estimate must be included in the evaluation of $\mathbf{V}$ and that the accuracy of the estimates depend strongly on the accuracy of the estimated mean-squared positioning error and its associated estimation uncertainty. The second proposed method is to treat the mean-squared positioning error, γ, as an additional parameter to be estimated in the regression analysis. We find that for $\Delta t < P$, this method introduces only slight additional uncertainty in the estimated cell movement indices, as long as $\gamma^{1/2} < SP$. For $\gamma^{1/2} > SP$, the meandering characteristics of the cell track becomes indistinguishable from mere fluctuations in the measured cell position, thus inducing large uncertainty in the estimates of the cell movement indices.

In addition to addressing the optimal time increment and preferred interval averaging method, we examine how precision of the estimates also depends on the total number of positions, n, defining the cell track and the number of cells tracked, M, if tracks are pooled. These results collectively provide a valuable tool in designing a cell tracking experiment to acquire the minimal optimal data for a desired precision of the estimates. When tracking multiple cells simultaneously using time-lapse video microscopy with on-line image analysis, the rate of measurement of cell positions is limiting. Therefore the investigator must weigh a desire for maximum precision of estimates of the indices against the tracking time and the number of cells tracked. Given a known rate of cell position measurement, one could use these results to design an optimal experiment by setting Δt, n, and M based on an *a priori* estimation of the cell movement indices.

Our GLSR formulae are not restricted to analysis of tracks of randomly migrating cells, but apply to tracks obtained for any species exhibiting a correlated random walk, that is, any substance exhibiting isotropic diffusion or any organism exhibiting random migration. An example of the former is diffusion of lipids and proteins in cell plasma membranes. An analysis analogous to ours has been presented motivated by membrane protein tracking data (Qian et al., 1991), including the possibility of simultaneous species convection, but restricted to the diffusion limit (to a pure random walk). Our analysis is valid for all observation times (for a correlated random walk), allowing estimation of the directional persistence time and root-mean-squared speed (in addition to the diffusion coefficient) when the interval between observation times is sufficiently short to make correlation of direction resolvable.

Acknowledgment

This work has been supported by a National Science Foundation PYI Award (BCS-8957736) to RTT. Many useful suggestions by the reviewers are gratefully acknowledged. The authors thank Elliot Elson for bringing the paper by Qian et al. to our attention during the revision of this manuscript.

Notation

$\mathbf{d}_{ik}$ = displacement vector from $\mathbf{r}_k$ to $\mathbf{r}_{k+i}$
$\langle d^2(t) \rangle$ = theoretical mean-squared displacement
$\mathbf{D}$ = variance-covariance matrix of parameter estimates
$\mathbf{e}_k$ = positioning error at $\mathbf{r}_k$
F = sum of the square of the residuals (weighted or nonweighted)
G_v = autocorrelation function of cell velocity
n = number of cell positions in a cell track
n_d = number of dimensions in which cell displacements are measured
n_i = number of sample intervals of time length $i\Delta t$ averaged from a cell track
P = directional persistence time
$\mathbf{r}_i$ = cell position vector
$\mathbf{r}_i'$ = cell position vector including positioning error
S = root-mean-squared cell speed
t = time
t_i = time interval of length $i\Delta t$
T_2 = overlapping time period of two overlapping displacements
T_1 or T_3 = nonoverlapping time periods of two overlapping displacements
$\mathbf{V}$ = matrix of the correlations of the residuals, ϵ
v_{ij} = element of $\mathbf{V}$
$\hat{\mathbf{V}}$ = $\mathbf{V}$ for multiple cell tracks
$\mathbf{V}''$ = $\mathbf{V}$ when estimating cell positioning error
$\mathbf{w}_m$ = vector of weighting factors for pooling data from multiple cell tracks
$w_{m,i}$ = element of $\mathbf{w}_m$
x_{ik} = square of displacement from $\mathbf{r}_k$ to $\mathbf{r}_{k+i}$
x_{ik}' = square of displacement from $\mathbf{r}_k'$ to $\mathbf{r}_{k+i}'$
x_{ik}'' = square of displacement from $\mathbf{r}_k'$ to $\mathbf{r}_{k+i}'$, subtracting $2\hat{\gamma}$
$\bar{x}_i$ = average of x_{ik} over a single cell track
$\hat{x}_i$ = average of $\bar{x}_i$ over multiple cell tracks

Greek letters

ϵ = vector of residuals in squared displacement data
ϵ_2 = element of ϵ
γ = mean-squared positioning error
$\hat{\gamma}$ = estimate of γ
Δt = time increment between measured cell positions
η_i = expected mean-squared displacement over time interval t_i
η_i' = expected mean-squared displacement over time intervals t_i including positioning error
μ = random motility coefficient
$\sigma_{\hat{\gamma}}$ = uncertainty in $\hat{\gamma}$
$\boldsymbol{\theta}$ = vector of parameter estimates
θ_i = element of $\boldsymbol{\theta}$

Literature Cited

Alt, W., "Correlation Analysis of Two-Dimensional Locomotion Paths," *Biological Motion*, W. Alt and G. Hoffmann, eds., p. 254, Springer-Verlag, Berlin (1990).

Buettner, H. M., and R. N. Pittman, "Quantitative Effects of Laminin Concentration on Neurite Outgrowth *in vitro*," *Dev. Biol.*, **145**, 266 (1991).

Burton, J. L., H. L. Bank, and P. Law, "Videoanalysis of Chemokinesis: Characterization of Speed, Persistence, and Orientation in an Agarose Assay," *Ann. Clin. Lab. Sci.*, **17**(6), 389 (1987).

de Boisfleury-Chevance, A., B. Rapp, and H. Gruler, "Locomotion of White Blood Cells: a Biophysical Analysis," *Blood Cells*, **15**(2), 315 (1989).

Dickinson, R. B., J. B. McCarthy, and R. T. Tranquillo, "Quantitative Characterization of Cell Invasion *in vitro*: Formulation and Validation of a Mathematical Model of the Collagen Gel Invasion Assay," *Ann. Biomed. Eng.*, **21**(6) (1993).

DiMilla, P. A., J. A. Quinn, S. M. Albelda, and D. A. Lauffenburger, "Measurement of Individual Cell Migration Parameters for Human Tissue Cells," *AIChE J.*, **38**(7), 1092 (1992).

Doillon, C. J., A. J. Wasserman, R. A. Berg, and F. H. Silver, "Behaviour of Fibroblasts and Epidermal Cell Cultivated on Analogues of Extracellular Matrix," *Biomaterials*, **9**, 91 (1988).

Dunn, G. A., "Characterizing a Kinesis Response: Time Averaged Measures of Cell Speed and Directional Persistence," *Agents and Actions Suppl.*, **12**, 14 (1983).

Farrell, B. E., R. P. Daniele, and D. A. Lauffenburger, "Quantitative Relationships between Single-Cell and Cell-Population Model Parameters for Chemosensory Migration Responses of Alveolar Macrophages to C5a," *Cell Motil Cytoskeleton*, **16**(4), 279 (1990).

Gail, M. H., and C. W. Boone, "The Locomotion of Mouse Fibroblasts in Tissue Culture," *Biophys. J.*, **10**, 980 (1970).
Gard, T. C., "Introduction to Stochastic Differential Equations," *Monographs and Textbooks in Pure and Applied Mathematics*, Vol. 114, Marcel Dekker, New York (1988).
Gardiner, C. W., *Handbook of Stochastic Methods, Springer Series in Synergetics*, Vol. 13, Springer-Verlag, Berlin-Heidelberg (1983).
Glasgow, J. E., B. E. Farrell, E. S. Fisher, D. A. Lauffenburger, and R. P. Daniele, "The Motile Response of Alveolar Macrophages: an Experimental Study Using Single-Cell and Cell Population Approaches," *Amer. Rev. Respir. Dis.*, **139**(2), 320 (1989).
Greisler, H. P., "Interactions at the Blood/Material Interface," *Ann. Vasc. Surg.*, **4**(1), 98 (1990).
Hall, R. L., "Amoeboid Movement as a Correlated Walk," *J. Math. Biol.*, **4**, 327 (1977).
Nossal, R., and G. H. Weiss, "A Descriptive Theory of Cell Migration on Surfaces," *J. Theor. Biol.*, **47**, 103 (1974).
Othmer, H. G., S. R. Dunbar, and W. Alt, "Models of Dispersal in Biological Systems," *J. Math. Biol.*, **26**(3), 263 (1988).
Qian, H., M. P. Sheetz, and E. L. Elson, "Single Particle Tracking: Analysis of Diffusion and Flow in Two-Dimensional Systems," *Biophys. J.*, **60**, 910 (1991).
Saltzman, W. M., P. Parsons-Wingerter, K. W. Leong, and L. Shin, "Fibroblast and Hepatocyte Behavior on Synthetic Polymer Surfaces," *J. Biom. Mat. Res.*, **25**, 741 (1991).
Seber, G. A. F., and C. J. Wild, *Nonlinear Regression*, Wiley, New York (1989).
Steen, F. H., *Elements of Probability and Mathematical Statistics*, Duxbury Press, Boston (1982).
Stokes, C. L., S. K. Williams, and D. A. Lauffenburger, "Migration of Individual Microvessel Endothelial Cells: Stochastic Model and Parameter Measurement," *J. Cell Sci.*, **99**, 419 (1991).
Takle, G. B., and A. M. Lackie, "Chemokinetic Behavior of Insect Haemocytes *in vitro*," *J. Cell Sci.*, **85**, 85 (1986).
Trinkaus, J. P., *Cells Into Organs: the Forces That Shape the Embryo*, Prentice-Hall, Englewood Cliffs, NJ (1984).
Wilkinson, P. C., J. M. Lackie, J. V. Forrester, and G. A. Dunn, "Chemokinetic Accumulation of Human Neutrophils on Immune Complex-Coated Substrata: Analysis at a Boundary," *J. Cell Biol.*, **99**(5), 1761 (1984).

Appendix I: Simulation of Cell Tracks

A central focus of this article is the effects of data averaging methods and positioning error on mean-squared displacement data. These effects are explained by comparing sample data to the theoretical mean-squared displacement curve. This curve, however, is obtainable only if the values of the cell movement indices S (or μ) and P are known *a priori*. Unfortunately, these values are not exactly known nor easily manipulated for real cells. We can, however, numerically simulate cell movement with a model cell that obeys the assumptions of a correlated random walk exactly. Using simulated cell tracks has several advantages: we can track as many simulated cells as we want for as long as we want (within computational limits and tolerance for numerical error); the simulated cell movement is guaranteed to obey the assumptions of a correlated random walk, thus any deviations of squared displacement data from the theoretical mean-squared displacement curve are not due to violation of model assumptions; and we know precisely the values of S and P *a priori* (as shown below, these values are needed in the generation of the simulated tracks).

Since Eq. 3 can be derived from any stochastic process where the velocity autocorrelation function is of the form:

$$G_v \equiv \langle \mathbf{v}(t+\tau)\cdot\mathbf{v}(t)\rangle = S^2 e^{-\tau/P} \tag{AI1}$$

we can prescribe any stochastic process for the velocity of our simulated cell as long as it obeys Eq. AI1. We stipulate that $\mathbf{v} = S[\cos\theta, \sin\theta]$ and that θ obeys the simple stochastic differential equation (sense of Ito):

$$d\theta = \sqrt{2/P}\,dW \tag{AI2}$$

where dW is an increment in the Wiener process (Gardiner, 1983).

We show that this is sufficient for Eq. AI1 to hold by noting that the corresponding Fokker-Planck equation to Eq. AI2 for the conditional probability density, $p(\theta, \tau|\theta_0)$ is:

$$\partial_t p(\theta, \tau|\theta_0) = \frac{1}{P}\,\partial_\theta^2 p(\theta, \tau|\theta_0) \tag{AI3}$$

which, for an initial direction θ_0, has the solution:

$$p(\theta, \tau|\theta_0) = \frac{1}{2\pi}\left(1 + 2\sum_{n=1}^{\infty}\cos[n(\theta-\theta_0)]\exp(-n^2 t/P)\right) \tag{AI4}$$

As $t\to\infty$, $p(\theta, \tau|\theta_0)$ evolves to the stationary density:

$$p_s(\theta) = \frac{1}{2\pi} \tag{AI5}$$

therefore, the autocorrelation function of the velocity is:

$$G_v = \langle S^2\cos\theta\cos\theta_0 + S^2\sin\theta\sin\theta_0\rangle = S^2\langle\cos(\theta-\theta_0)\rangle$$

$$= S^2\int_0^{2\pi} d\theta \int_0^{2\pi} d\theta_0 \cos(\theta-\theta_0)p(\theta, \tau|\theta_0)p_s(\theta_0)$$

$$= S^2 e^{-\tau/P} \tag{AI6}$$

To obtain simulated cell tracks, Eq. AI2 along with the equations for cell position,

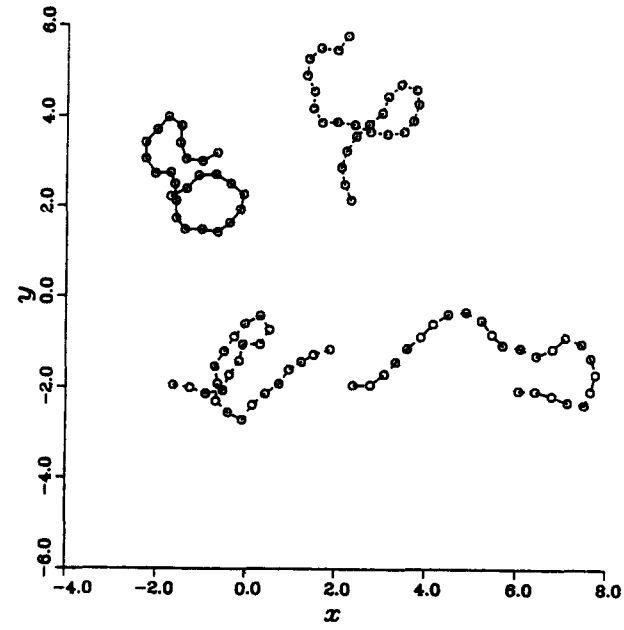
$$\frac{dx}{dt} = S\cos\theta \tag{AI8}$$

$$\frac{dy}{dt} = S\sin\theta \tag{AI9}$$

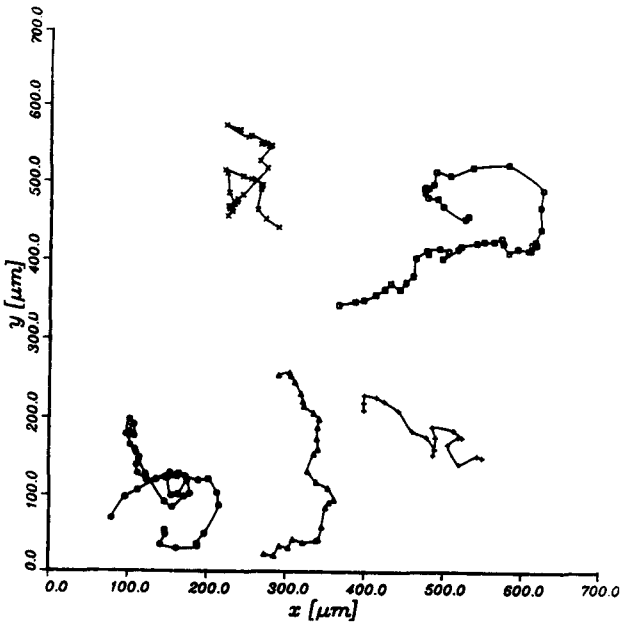
were integrated numerically with the Cauchy-Euler method (typical time step used was approximately $10^{-4}P$) (Gard, 1988).

Shown in Figure A1 are examples of simulated cell tracks with a comparison to actual cell tracks of murine melanoma cells on the surface of a collagen gel. This shows that the behavior of the simulated cells is similar to that of real cells. In Figure A2, mean-squared displacement data from a simulated cell track ($n = 1{,}000$, $\Delta t = 0.2$) obtained from the integration of Eqs. AI2, AI8 and AI9 is compared to the theoretical $\langle d^2(t)\rangle$ curve computed from Eq. 3. The agreement between the data and the curve implies that tracks obtained in this way simulate cell migration which obeys the correlated random walk assumption upon which Eq. 3 is based.

Positioning error is included by introducing additional position variables $[x', y']$ to represent the "measured" cell location. $[x', y']$ are generated by adding a bivariate random variable from the Gaussian distribution with variance γ^2 to $[x, y]$ at every time point.



(a)



(b)

Figure A1. (a) Simulated cell tracks ($\Delta t = 0.4P$) are compared to (b) real cell tracks ($\Delta t = 15$ min) of mouse melanoma cells migrating on the surface of a fibronectin-coated surface of a tissue culture chamber (unpublished data).

Note the similar track characteristics, that is, the directional persistence over short times, and the meandering path characteristic of a random walk over longer time.

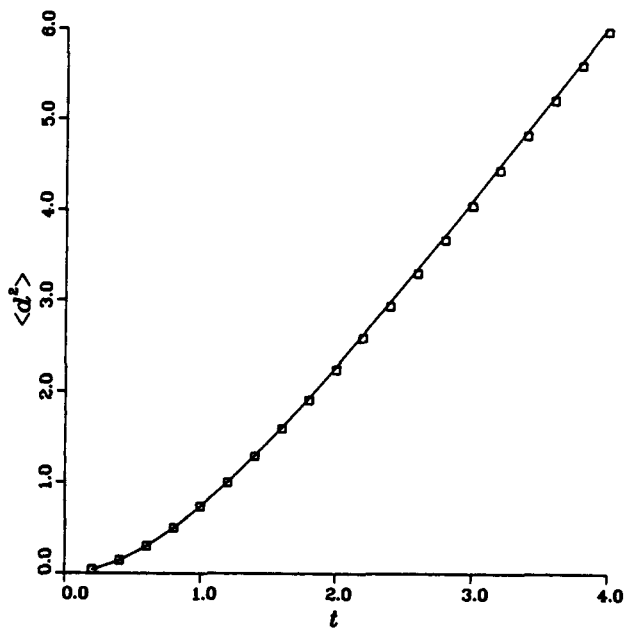


Figure A2. Mean-squared displacement data, $\bar{X}_t$, vs. time, t_t, for a simulated cell track obtained from the numerical integration of Eqs. AI2, AI8 and AI9 ($n = 500$, $\Delta t = 0.2$).

The correpondence of the data points with the theoretical $\langle d^2(t) \rangle$ curve implies that the simulated cell tracks obey the assumptions used in the derivation of Eq. 3.

Appendix II: Derivation of the Correlation Matrix, V

To evaluate the correlation matrix of the residuals, **V**, we need to evaluate the covariance of squared displacements of overlapping intervals. Assume the period of overlap is T_2 and the two periods of nonoverlapping time are T_1 and T_3, as shown in Figure 4. The covariance of the squared displacements $d^2(T_1+T_2)$ and $d^2(T_2+T_3)$ is then defined by function ϕ as:

$$\phi(T_1, T_2, T_3; P, S) = \langle d^2(T_1+T_2) d^2(T_2+T_3) \rangle - \langle d^2(T_1+T_2) \rangle \langle d^2(T_2+T_3) \rangle \quad \text{(AII1)}$$

ϕ can be obtained from integrating the cell velocity over the corresponding intervals of the track:

$$\phi(T_1, T_2, T_3; P, S) = \int_0^{T_1+T_2} dt_1 \int_0^{T_1+T_2} dt_2 \int_{T_1}^{T_1+T_2+T_3} dt_3 \times \int_{T_1}^{T_1+T_2+T_3} dt_4 \, \Phi(t_1, t_2, t_3, t_4; P, S) \quad \text{(AII2)}$$

where the integrand is:

$$\Phi(t_1, t_2, t_3, t_4; P, S) = \langle [\mathbf{v}(t_1) \cdot \mathbf{v}(t_2)][\mathbf{v}(t_3) \cdot \mathbf{v}(t_4)] \rangle - \langle \mathbf{v}(t_1) \cdot \mathbf{v}(t_2) \rangle \langle \mathbf{v}(t_3) \cdot \mathbf{v}(t_4) \rangle \quad \text{(AII3)}$$

Noting in Eq. AII3 the symmetry of the dependence on t_3 and t_4, and on t_1 and t_2, the integral in Eq. AII2 can be written as:

$$\phi(T_1, T_2, T_3; P, S) = 4 \int_0^{T_1+T_2} dt_1 \int_{t_1}^{T_1+T_2} dt_2 \int_{T_1}^{T_1+T_2+T_3} dt_3 \times \int_{t_3}^{T_1+T_2+T_3} dt_4 \, \Phi(t_1, t_2, t_3, t_4; P, S) \quad \text{(AII4)}$$

where $t_4 > t_3$ and $t_2 > t_1$. We now need to derive the form of Φ for all possible permutations of t_1, t_3, t_2 and t_4. Immediately apparent is that if $t_3 > t_2$ or $t_1 > t_4$, then $\mathbf{v}(t_1)\cdot\mathbf{v}(t_2)$ and $\mathbf{v}(t_3)\cdot\mathbf{v}(t_4)$ are independent so that $\Phi = 0$. If, however, $t_1 < t_3 < t_2 < t_4$, then one can show that:

$$\Phi = \Phi_A(t_1, t_2, t_3, t_4; P, S) \equiv \langle \mathbf{v}(t_1)\cdot\mathbf{v}(t_3)\rangle\langle \mathbf{v}(t_2)\cdot\mathbf{v}(t_4)\rangle - \langle \mathbf{v}(t_1)\cdot\mathbf{v}(t_2)\rangle\langle \mathbf{v}(t_3)\cdot\mathbf{v}(t_4)\rangle \quad \text{(AII5)}$$

and from Eq. 3 is:

$$\Phi_A(t_1, t_2, t_3, t_4; P, S) = S^4 e^{-(t_4 - t_1)/P}[e^{(t_2 - t_3)/P} - e^{-(t_2 - t_3)/P}] \quad \text{(AII6)}$$

Likewise, if $t_3 < t_1 < t_4 < t_2$, then

$$\Phi_B(t_1, t_2, t_3, t_4; P, S) = S^4 e^{-(t_2 - t_1)/P}[e^{(t_4 - t_3)/P} - e^{-(t_4 - t_3)/P}] \quad \text{(AII7)}$$

If $t_3 < t_1 < t_2 < t_4$, then

$$\Phi_C(t_1, t_2, t_3, t_4; P, S) = S^4 e^{-(t_4 - t_3)/P}[e^{(t_2 - t_1)/P} - e^{-(t_2 - t_1)/P}] \quad \text{(AII8)}$$

Finally, if $t_1 < t_3 < t_4 < t_2$, then

$$\Phi_D(t_1, t_2, t_3, t_4; P, S) = S^4 e^{-(t_2 - t_3)/P}[e^{(t_4 - t_1)/P} - e^{-(t_4 - t_1)/P}] \quad \text{(AII9)}$$

Eq. AII4 can be written separating the integral into each possible permutation:

$$\begin{aligned}
\phi = {} & 4\int_0^{T_1} dt_1 \int_{T_1}^{T_1+T_2} dt_3 \int_{t_3}^{T_1+T_2} dt_3 \int_{t_2}^{T_1+T_2+T_3} dt_4\, \Phi_A \\
& + 4\int_{T_1}^{T_1+T_2} dt_1 \int_{t_1}^{T_1+T_2} dt_3 \int_{t_3}^{T_1+T_2} dt_3 \int_{t_2}^{T_1+T_2+T_3} dt_4\, \Phi_A \\
& + 4\int_0^{T_1} dt_1 \int_{T_1}^{T_1+T_2} dt_3 \int_{t_3}^{T_1+T_2} dt_4 \int_{t_4}^{T_1+T_2} dt_2\, \Phi_B \\
& + 4\int_{T_1}^{T_1+T_2} dt_1 \int_{t_1}^{T_1+T_2} dt_3 \int_{t_3}^{T_1+T_2} dt_4 \int_{t_4}^{T_1+T_2} dt_2\, \Phi_B \\
& + 4\int_{T_1}^{T_1+T_2} dt_3 \int_{t_3}^{T_1+T_2} dt_3 \int_{t_1}^{T_1+T_2} dt_2 \int_{t_2}^{T_1+T_2} dt_4\, \Phi_C \\
& + 4\int_{T_1}^{T_1+T_2} dt_3 \int_{t_3}^{T_1+T_2} dt_3 \int_{t_1}^{T_1+T_2} dt_4 \int_{t_4}^{T_1+T_2} dt_2\, \Phi_D \quad \text{(AII10)}
\end{aligned}$$

Note from Eqs. AII6-AII9 that, in general for any $\tau_1 < \tau_2 < \tau_3 < \tau_4$, then

$$\Phi(\tau_1, \tau_2, \tau_3, \tau_4; P, S) = S^4 e^{-(\tau_4 - \tau_1)/P}[e^{(\tau_3 - \tau_2)/P} - e^{-(\tau_3 - \tau_2)/P}] \quad \text{(AII11)}$$

Making this variable transformation into each of the integrals in Eq. AII10 and summing identical portions of the integrals yields:

$$\begin{aligned}
\phi(T_1, T_2, T_3; P, S) = {} & 8S^4 \int_0^{T_1} d\tau_1 \int_{\tau_2}^{T_1+T_2} d\tau_2 \int_{T_1}^{T_1+T_2} d\tau_3 \\
& \times \int_{\tau_3}^{T_1+T_2} d\tau_4\, \Phi(\tau_1, \tau_2, \tau_3, \tau_4; P, S) \\
& + 4S^4 \int_0^{T_1} d\tau_1 \int_{T_1}^{T_1+T_2} d\tau_2 \int_{\tau_2}^{T_1+T_2} d\tau_3 \\
& \times \int_{T_1+T_2}^{T_1+T_2+T_3} d\tau_4\, \Phi(\tau_1, \tau_2, \tau_3, \tau_4; P, S) \\
& + 16S^4 \int_{T_1}^{T_1+T_2} d\tau_1 \int_{\tau_1}^{T_1+T_2} d\tau_2 \int_{\tau_2}^{T_1+T_2} d\tau_3 \int_{\tau_3}^{T_1+T_2} d\tau_4 \\
& \times \Phi(\tau_1, \tau_2, \tau_3, \tau_4; P, S) \\
& + 8S^4 \int_{T_1}^{T_1+T_2} d\tau_1 \int_{\tau_1}^{T_1+T_2} d\tau_2 \int_{\tau_2}^{T_1+T_2} d\tau_3 \int_{T_1+T_2}^{T_1+T_2+T_3} d\tau_4 \\
& \times \Phi(\tau_1, \tau_2, \tau_3, \tau_4; P, S) \quad \text{(AII12)}
\end{aligned}$$

Upon integration, this becomes:

$$\begin{aligned}
\phi(T_1, T_2, T_3; P, S) = {} & 8S^4P^4(1 - \bar{e}^{T_1/P})\left[(T_2/P)(1 + 2e^{-T_2/P}) - 2(1 - e^{-T_2/P}) - \frac{1}{2}(1 - e^{-2T_2/P})\right] \\
& + 4S^4P^4(1 - e^{-T_1/P})(1 - e^{-T_2/P})[1 - e^{-2T_2/P} - 2(T_2/P)e^{-T_2/P}] \\
& + 16S^4P^4\left[\frac{1}{2}(T_2/P)^2 - (T_2/P)\left(\frac{5}{2} + 2e^{-T_2/P}\right) + \frac{1}{4}(1 - e^{-2T_2/P}) + 4(1 - e^{-T_2/P})\right] \\
& + 8S^4P^4(1 - e^{-T_3/P})\left[(T_2/P)(1 + 2e^{-T_2/P}) - 2(1 - e^{-T_2/P}) - \frac{1}{2}(1 - e^{-2T_2/P})\right] \quad \text{(AII13)}
\end{aligned}$$

in terms of S and P. Using the relation $\mu = 1/2\ S^2P$, this can also be written as:

$$\begin{aligned}
\phi(T_1, T_2, T_3; P, \mu) = {} & 32\mu^2P^2(2 - \bar{e}^{T_1/P} - e^{-T_3/P}) \\
& \times \left[T_2/P(1 + 2e^{-T_2/P}) - 2(1 - e^{-T_2/P}) - \frac{1}{2}(1 - e^{-2T_2/P})\right] \\
& + 16\mu^2P^2(1 - e^{-T_1/P})(1 - e^{-T_2/P})[1 - e^{-2T_2/P} - 2T_2e^{-T_2/P}] \\
& + 64\mu^2P^2\left[\frac{1}{2}T_2^2/P^2 - T_2\left(\frac{5}{2} + 2e^{-T_2/P}\right) + \frac{1}{4}(1 - e^{-2T_2/P}) + 4(1 - e^{-T_2/P})\right] \quad \text{(AII14)}
\end{aligned}$$

Appendix III: Generalized Nonlinear Regression Algorithm

We use the following algorithm based on the Gauss-Newton

method to perform the generalized nonlinear least-squares regression analysis. This algorithm is based on that of Seber and Wild (1989), with modification to reduce computational expense from evaluating $\mathbf{V}$ at each iteration in the Gauss-Newton method.

1. Provide an initial guess, $\boldsymbol{\theta}^{(0)}$, for the cell movement indices, where $\boldsymbol{\theta} = [\mu\ P]$ or $[S\ P]$.
2. Calculate $\mathbf{V}$ from $\boldsymbol{\theta}^{(0)}$.
3. Perform a Cholesky decomposition of $\mathbf{V}$ to obtain upper triangular matrix $\mathbf{U}$ where $\mathbf{V} = \mathbf{U}^T\mathbf{U}$.
4. Solve $\mathbf{U}^T\mathbf{z} = \bar{\mathbf{x}}$ and $\mathbf{U}^T\mathbf{k}(\boldsymbol{\theta}) = \boldsymbol{\eta}(\boldsymbol{\theta})$ for new vectors $\mathbf{z}$ and $\mathbf{k}(\boldsymbol{\theta})$.
5. Solve $\mathbf{U}^T\mathbf{K}(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta})$ for $\mathbf{K}(\boldsymbol{\theta})$.
6. Solve $\mathbf{K}^T\mathbf{K}\boldsymbol{\delta} = \mathbf{K}^T(\mathbf{z} - \mathbf{k})$ for $\boldsymbol{\delta} = (\boldsymbol{\theta}^{(i+1)} - \boldsymbol{\theta}^{(i)})$.
7. Update $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} + \boldsymbol{\delta}$.
8. Repeat steps 3-7 until $\|\boldsymbol{\delta}\| + \|\mathbf{K}^T(\mathbf{z} - \mathbf{k})\|$ is less than some small tolerance.
9. If $\|\boldsymbol{\theta}^{(i+1)} - \boldsymbol{\theta}^{(0)}\|$ is greater than some small tolerance, set $\boldsymbol{\theta}^{(0)} = \boldsymbol{\theta}^{(i+1)}$, and repeat steps 2-8. Otherwise, $\boldsymbol{\theta}^{(i+1)}$ is the optimal estimate of $\boldsymbol{\theta}$.
10. Compute $\mathbf{D}(\boldsymbol{\theta}) = (\mathbf{K}^T\mathbf{K})^{-1}$ from the optimal estimate of $\boldsymbol{\theta}$.

Note: Upon request, the authors will provide a FORTRAN program that implements GLSR, requiring the IMSL Math Library (Levenberg-Marquardt subroutine).

Manuscript received Sept. 1, 1992, and revision received Apr. 12, 1993.