Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 233 (2005) 573–588
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Sampling rate effects on measurements of correlated and biased random walks

E.A. Codling[a,*,1], N.A. Hill[b]

[a]Department of Applied Mathematics, University of Leeds, Leeds LS2 9JT, UK
[b]Department of Mathematics, University of Glasgow, Glasgow G12 8QQ, UK

Received 27 July 2004; received in revised form 4 November 2004
Available online 15 December 2004

Abstract

When observing the two-dimensional movement of animals or microorganisms, it is usually necessary to impose a fixed sampling rate, so that observations are made at certain fixed intervals of time and the trajectory is split into a set of discrete steps. A sampling rate that is too small will result in information about the original path and correlation being lost. If random walk models are to be used to predict movement patterns or to estimate parameters to be used in continuum models, then it is essential to be able to quantify and understand the effect of the sampling rate imposed by the observer on real trajectories. We use a velocity jump process with a realistic reorientation model to simulate correlated and biased random walks and investigate the effect of sampling rate on the observed angular deviation, apparent speed and mean turning angle. We discuss a method of estimating the values of the reorientation parameters used in the original random walk from the rediscretized data that assumes a linear relation between sampling time step and the parameter values.

Keywords: Biased random walk; Velocity jump process; Sampling rate; Animal dispersal

1. Introduction

The simple random walk model has been frequently used to describe the movement and dispersal of groups of animals or microorganisms (Skellam, 1951, 1973; Okubo, 1980). Possibly the simplest form of the two-dimensional random walk is when a walker is restricted to changing positions on a square lattice where there is equal probability of moving up, down, left and right at each step. Such a walk is uncorrelated as the direction of movement is completely independent of the previous direction moved and results in *Brownian* motion (Brown, 1828; Einstein, 1906). Bias can be introduced into the random walk by making the probability of moving in one particular direction more likely and a drift in this direction will be observed (Berg, 1983). A problem with these simple uncorrelated models is that they allow for effectively infinite propagation speeds (Okubo, 1980; Othmer et al., 1988), and the resulting diffusion equations are only valid as long time approximations to the true underlying behaviour.

A more realistic random walk model is one that includes correlation between successive steps, so that the random walk is in the *velocity* rather than the position. In one dimension it is possible to set up a simple correlated random walk on a line (*velocity jump process*) and derive the *telegraph equation* to describe the population density (Goldstein, 1951; Kac, 1974; Okubo, 1980). However, a similar method does not work in higher dimensions and it is not possible to derive an equation for the population density directly (Othmer et al., 1988; Hillen and Othmer, 2000; Codling and Hill, 2004).

*Corresponding author. Tel.: +353 91 730400; fax: +353 91 730470.
E-mail addresses: edd.codling@marine.ie (E.A. Codling), n.a.hill@maths.gla.ac.uk (N.A. Hill).
[1]Current address: Department of Zoology, University College Cork, Cork, Ireland

0022-5193/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2004.11.008

It is not necessary to restrict the two-dimensional correlated random walk to a lattice and it is more realistic to use a continuous circular probability distribution for the choice of direction at each step. Given such a circular probability distribution for the reorientation at each step, it is possible to calculate the properties of a two-dimensional correlated and *unbiased* random walk at a given time such as the mean squared displacement (Tchen, 1952; Nossal and Weiss, 1974; Kareiva and Shigesada, 1983), the sinuosity or rate of turning (Dunn, 1983; Bovet and Benhamou, 1988), and the mean dispersal distance (Bovet and Benhamou, 1988; McCulloch and Cain, 1989; Byers, 2001; Codling, 2003).

In general however, similar methods do not work with *biased* and correlated two-dimensional random walks and it is much harder to calculate their statistics. Using a linear transport equation to describe the velocity jump process, Othmer et al. (1988) show how it is possible to calculate equations for the spatial moments of a random walk that has separate probability distributions for bias and correlation effects in the reorientation. This method has been extended by Codling and Hill (2004) to include a more realistic reorientation model that implicitly includes bias and correlation effects in the same probability distribution, although this method requires making several moment closure assumptions. A general theory for using velocity jump processes and deriving transport equations is discussed in Hillen and Othmer (2000); Othmer and Hillen (2002) and Hillen (2002).

If such random walk models are to be used to predict or analyse the spatial properties of real animal populations, then it is essential that we use accurate and realistic models for the movement and reorientation of each individual. Both the position jump and velocity jump processes result in discrete step-wise movement paths that are not necessarily realistic. Certain animals have been observed to move in a step-wise fashion, such as ovipositing butterflies moving from site to site (Kareiva and Shigesada, 1983), but most movements are observed to have a more continuous path (e.g. Hill and Häder, 1997). However, when observing and recording the trajectory of an animal it is usually necessary to impose a fixed sampling rate (the number of observations in a given time) or sampling length (the time or distance between successive observations) to discretize the continuous path. The best results will obviously come from observations using the smallest possible sampling lengths but it may not be possible to use such small sampling lengths due to experimental constraints or practical considerations. It is therefore of intrinsic interest to understand how the sampling rate imposed by an observer will affect the properties of the random walk and the subsequent conclusions that may be drawn. In general, by increasing the sampling length, the trajectory will initially appear more random as correlation effects are lost, while smoothing of the path will mean that the apparent speed will also decrease as the total length of the walk decreases. The fact that the apparent randomness in turning increases with sampling length led to the definition of sinuosity (Bovet and Benhamou, 1988) (see Section 4) and has also been exploited by Hill and Häder (1997) in their random walk on a circle (see Section 6). Hill and Häder (1997) used a range of sampling time steps to estimate the reorientation parameters of a continuously turning walk, such as the angular deviation per unit time (or sinuosity) and the amplitude of the mean turning angle (which is used to calculate the mean reorientation time). These parameters can then be used in continuum models for the behaviour of populations of swimming microorganisms where bioconvection patterns occur (Kessler, 1986; Hill et al., 1989; Vincent and Hill, 1996; Hill and Pedley, 2004). However, the method of estimating the parameters used by Hill and Häder (1997) relies on an ad hoc assumption of linear relations between the sampling time step and both the angular deviation and the amplitude of the mean turning angle. At small sampling time steps this is likely to be true but at larger sampling time steps this assumption will not hold, so it is essential to be able to quantify at what point the sampling time step becomes too large.

In this paper we investigate the effect of changing the sampling rate on the observed properties of both biased and unbiased simulated velocity jump processes and in particular how the sampling rate affects the angular deviation (sinuosity) and the apparent speed. We demonstrate that the method of Hill and Häder (1997) to estimate the reorientation parameters of a biased and continuously turning random walk is valid for small sampling time steps, and quantify when the method is likely to fail.

2. Spatial and temporal sampling

We are concerned with discrete *temporal* sampling of the trajectory, i.e. an observation is made every τ_s time units (where τ_s is the *sampling time step* and the *sampling rate* is given by $1/\tau_s$). This is in contrast to discrete *spatial* sampling where a section of the trajectory that is already known is split into steps of a fixed length L, as used by Bovet and Benhamou (1988). In situ observations of the movement of animals and microorganisms are likely to rely on temporal sampling. Spatial sampling can only be used when a sufficient length of the complete trajectory has already been observed and, as the data is purely spatial, it is not possible to observe waiting times or variable speeds of movement that may have occurred in the original movement. In both temporal and spatial sampling, as the sampling time step or length (τ_s or L) increases, an observer will lose more information about

the original trajectory. In a correlated random walk, the initial correlation between successive steps is lost as the sampling length increases so that the observed walk appears less correlated (or completely random), while the apparent speed will be less as the observed trajectory will cover less total distance in the same time. Bovet and Benhamou (1988) note that many examples of biological movement has been observed to approximately fit the simple diffusion model (which is based on a purely random uncorrelated walk, see Okubo, 1980), and the reason for this may be that the sampling lengths are too large to distinguish the correlations.

In a correlated and biased random walk (where there is a preferred direction of movement), correlation between successive steps will be lost as the sampling length increases. The apparent speed will also decrease, but the motion will appear more like a straight line in the preferred direction. Bovet and Benhamou (1988) argue that a purely spatial measure of the sinuosity is useful because temporal sampling will result in an amalgamation of time and space so that it is not possible to distinguish between the components representing the animal's path and those representing its velocity. This blending of spatial and temporal components is most significant when there is likely to be a correlation between the animal's velocity and the structure of its path, such as with foraging animals that show area-restricted searching behaviour. We do not try and address this problem here, so care should be taken in generalizing any results to such cases.

3. The velocity jump process and reorientation model

3.1. The velocity jump process

In their 'random walk in external field', Othmer et al. (1988) describe a simple two-dimensional velocity jump process where a population of individuals all move with a fixed speed s and the probability of turning is governed by a Poisson process with turning frequency λ. As there is a fixed speed in this process, the only change in the velocity is in the direction of movement. The probability of turning from the previous direction of movement θ' to a new direction of movement θ (so that $\delta = \theta - \theta'$ is the turning angle, where θ and θ' are absolute angles measured in relation to some fixed origin) is governed by a reorientation kernel $T(\theta, \theta')$. Under these assumptions, the velocity jump process can be described by a linear transport equation (see Eq. (71) in Othmer et al., 1988), which can be used to derive a system of equations for the spatial moments. In this particular example of a velocity jump process, Othmer et al. (1988) used a superposition of two separate probability distributions for $T(\theta, \theta')$ to represent bias and correlation effects (see Fig. 4 in Othmer et al., 1988). Using this choice of $T(\theta, \theta')$, they then derived a simple closed system of differential equations for the spatial statistics of the population of random walkers. More recently, Hillen and Othmer (2000), Othmer and Hillen (2002) and Hillen (2002) discuss the general properties of velocity jump processes and describe how, for most choices of $T(\theta, \theta')$, it is necessary to use an appropriate moment closure or approximation method to derive a closed system of equations for the spatial moments.

Note that the spatial moments such as the average position and spread will be unaffected by changing the temporal sampling rate as each individual will still be observed in the same position at the same time, although as discussed earlier, the apparent speed will decrease.

3.2. Circular distributions

The simplest random walk models are fixed on a square lattice, and in two dimensions there is only a choice of four possible directions of movement. A more realistic model is not restricted to a lattice and allows for movement in any direction at each step. This requires a probability distribution function (p.d.f.) for the turning angle, δ.

The simplest unimodal circular distributions to use for the p.d.f. of the turning angle are the wrapped normal and von Mises distributions (Batschelet, 1981; Mardia and Jupp, 1999). Both distributions have a similar shape and, as they differ by only a few percent for appropriate choices of parameter values, they are often assumed to be equivalent. The mean of both distributions is given by μ_δ, while the spread is measured in slightly different ways. The wrapped normal has an angular variance, σ_δ^2, that is analogous to the variance of the linear normal distribution; while the von Mises distribution has a *concentration parameter*, κ. A relation can be found between these two parameters by equating the first moments (mean resultant lengths) of the distributions; see Section 3.4.

The von Mises and wrapped normal distributions have been used as the p.d.f. for the turning angle, δ, to model *correlated* and *unbiased* random walks (e.g. Siniff and Jessen, 1969; Kareiva and Shigesada, 1983; Bovet and Benhamou, 1988). In these models the mean turning angle, μ_δ, is assumed to be zero so that there is a tendency to continue moving in the same direction—a realistic model for animal motion where there is no overall preferred direction of movement. If the mean turning angle, μ_δ, is made dependent on the direction of movement, θ', then *bias* can be introduced into the turning angle distribution as we show in Section 6. If the bias introduced into the system is fixed and only dependent on the direction of movement (spatially independent) such as in phototaxis (Hill and Vincent, 1993) or gyrotaxis (Kessler, 1986), then our reorientation kernel is arguably more realistic than Othmer et al.

(1988), who used a superposition of two separate probability distributions for correlation and bias in their random walk in an external field.

3.3. The von Mises distribution

For our reorientation kernel, $T(\theta,\theta')$, we use the von Mises distribution as it easier to simulate and to work with analytically than the wrapped normal distribution. This is defined as

$$T(\theta,\theta') = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\delta - \mu_\delta)), \tag{1}$$

where

$$\int_{-\pi}^{\pi} T(\theta,\theta')\,\mathrm{d}\theta = 1 \tag{2}$$

and I_n denotes the modified Bessel function of the first kind and order n. $\delta = \theta - \theta'$ is the turning angle and the mean turning angle, $\mu_\delta(\theta')$, is either zero (in the case of an unbiased walk) or takes a sinusoidal or linear form to introduce bias (Hill and Häder, 1997) see Section 6. When $\kappa = 0$ the von Mises distribution equals the uniform distribution, and as $\kappa \to \infty$ the distribution becomes sharply peaked about the mean turning angle μ_δ (Batschelet, 1981; Mardia and Jupp, 1999).

Note that, because of the dependence of $\mu_\delta(\theta')$ on θ' that we introduce in Section 6, (1) violates assumption T4 of Hillen and Othmer (2000) and Hillen (2002) in their general theory of velocity jump processes. That is, a simple calculation shows that when $\mu_\delta(\theta') \neq 0$,

$$\int_{-\pi}^{\pi} T(\theta,\theta')\,\mathrm{d}\theta' \neq 1, \tag{3}$$

for our velocity jump process and the general results of Hillen and Othmer (2000) and Hillen (2002) may not hold.

3.4. Relation between the wrapped normal and von Mises distributions

If the first moment (mean resultant length), R, of the wrapped normal and von Mises distributions are equal then the relation between σ_δ and κ is given by

$$R = A_1(\kappa) = I_n(\kappa)/I_0(\kappa) = \exp(-\sigma_\delta^2/2) \tag{4}$$

and the two distributions only differ by a few percent, so that in applications it is convenient to treat their properties as being the same (Mardia and Jupp, 1999). The function $A_1(\kappa)$ and its inverse $A^{-1}(\kappa)$ are readily computed and can be found from tables of the inverse Bessel functions (Batschelet, 1981; Mardia and Jupp, 1999). Assuming (4) holds then, as $\kappa \to 0$, $\sigma_\delta^2 \to \infty$, and $\sigma_\delta^2 \to 0$ as $\kappa \to \infty$.

In the subsequent simulations, we set up a random walk that uses the von Mises distribution for the p.d.f. of the turning angle δ, and exploit the fact that this distribution is approximately equal to the wrapped normal in order to study the angular deviation σ_δ.

3.5. Simulating the velocity jump process

To simulate the random walk of an individual organism as a velocity jump process we use the following algorithm (a full description of the algorithm is given in Codling, 2003):

1. The random walker starts at the origin $(x,y) = (0,0)$ facing a direction, θ', chosen at random from the uniform distribution.
2. The walker then moves with a fixed speed, s, in the current direction, θ', for a random time period, τ, before turning. The turning frequency is drawn from a Poisson process with intensity λ. In such a Poisson process the times between events are exponentially distributed with mean $1/\lambda$ (e.g. Grimmett and Stirzaker, 2001), so the *average* time between turning events is given by $\bar{\tau} = 1/\lambda$.
3. The walker then turns to a new direction, θ, where θ is randomly drawn from a von Mises distribution with parameters κ and μ_δ. The parameters κ and d_τ (the amplitude of the mean turning angle, μ_δ) are fixed at the start of the simulation. If $d_\tau = \mu_\delta = 0$ then the walk is unbiased; realistic models for $\mu_\delta(\theta')$ that introduce bias are discussed in Section 6. The von Mises distribution is much simpler to simulate than the wrapped normal, and we use an algorithm by Fisher and Best (1979). The value of κ used is related to the angular deviation σ_δ by (4).
4. Steps 2–3 are then repeated for the required time.

In the algorithm there are no other external effects on the movement of the walker, e.g. environmental factors, interactions or flow. In the form described above, the algorithm is suitable for simulating the movement of walkers where the preferred direction of movement and the reorientation parameters are always fixed. The algorithm is easily adapted to account for a changing preferred direction or spatially dependent parameters (Codling et al., 2004), but the results presented here assume that these are fixed for all time for all walkers.

4. The sinuosity of an unbiased velocity jump process

For a correlated random walk Bovet and Benhamou (1988) defined sinuosity as

$$S = \sigma_\delta/\sqrt{P}, \tag{5}$$

where σ_δ and P are the angular deviation and step length used in the original random walk. From simulation results they found that the observed sinuosity took

the form

$$S^* = z\ \sigma_L^*/\sqrt{L}, \tag{6}$$

where z is a constant to be found that takes into account smoothing of the trajectories and σ_L^* is the angular deviation observed when a fixed spatial step length of L is imposed on the random walk. Thus, there is a linear relationship between the observed angular variance $(\sigma_L^*)^2$ and the step length used to observe the walk. Dunn (1983) also looked at the mean square turning angle per unit time which is a measure of the temporal sinuosity, while Hill and Häder (1997) used the assumption that the relationship between the observed angular variance and the sampling time step is linear to estimate the reorientation parameters of a continuously turning random walk from experimental data. In a similar manner to Bovet and Benhamou (1988), we define the *temporal sinuosity* to be

$$S_\tau = \sigma_\delta/\sqrt{\bar{\tau}} \tag{7}$$

and the *observed* temporal sinuosity to be

$$S_\tau^* = \xi\ \sigma_{\tau_s}^*/\sqrt{\tau_s}, \tag{8}$$

where $\bar{\tau}$ is the average time step between turns in the original random walk or velocity jump process, ξ is a constant that accounts for smoothing of the trajectories, and τ_s is the fixed sampling time step imposed by the observer. If we use a fixed speed s in the original random walk then, in a fixed time step τ_s, a walker will move a maximum distance of $D = s\tau_s$, so that it may appear that (5) and (7) are the same. This is not true however, as the underlying sampling processes are different. Spatial sampling imposes a fixed distance between separate points of the random walk, while with temporal sampling there is only a *maximum* distance between points, and the observed distance between points can be anything less than this maximum depending on the rate of turning.

Bovet and Benhamou (1988) completed simulations of unbiased and correlated random walks with a *fixed* step length between turns to find the value of z in their model. For $\sigma_L^* < 1.2$ rad they found that

$$\sigma_L^* = 0.85\ \sigma_\delta\sqrt{L/P} \tag{9}$$

and thus $z = 1.18$ in (6). For values of $\sigma_L^* > 1.2$ rad this relationship breaks down as, when the angular variance is large, the distribution of turning angles appears almost completely random and all original correlation effects are lost.

4.1. The observed sinuosity

In order to study the relationship between the observed angular deviation, $\sigma_{\tau_s}^*$, and the sampling time step, τ_s, we have simulated unbiased velocity jump processes using the algorithm given in Section 3.5 (where the mean turning angle $\mu_\delta = 0$). In each simulation we run 1000 walkers for 200 time units, all with a fixed angular deviation σ_δ (from which we use (4) to calculate a value for κ in the simulation), a fixed speed $s = 1$ and turning rate $\lambda = 1$. The time between turning events varies but the *average* time between turns is given by $\bar{\tau} = 1$. This is equivalent to a non-dimensionalized system. Each original trajectory is then rediscretized using linear interpolation with a given sampling time step τ_s and the angular deviation is calculated for each trajectory and averaged over the population.

Fig. 1 shows plots of the observed angular deviation, $\sigma_{\tau_s}^*$, against $\sqrt{\tau_s/\bar{\tau}}$ for various values of σ_δ. It is clear from Fig. 1 that for $\sigma_{\tau_s}^* < 1.2$ rad there is a linear relation between $\sigma_{\tau_s}^*$ and $\sqrt{\tau_s}$, but for larger values of $\sigma_{\tau_s}^*$ this linear relation starts to break down and the sinuosity cannot be measured. This is the same behaviour as Bovet and Benhamou (1988) observed. Using the method of least squares, we have fitted linear functions to the observed data for which $\sigma_{\tau_s}^* < 1.2$ rad, and for each value of σ_δ we find

$$\sigma_{\tau_s}^* = 0.79\ \sigma\sqrt{\tau_s/\bar{\tau}}. \tag{10}$$

This is similar to (9) except for the smaller value for $1/\xi$, the constant that gives a measure of the smoothing effect of the sampling time step on the original random walk.

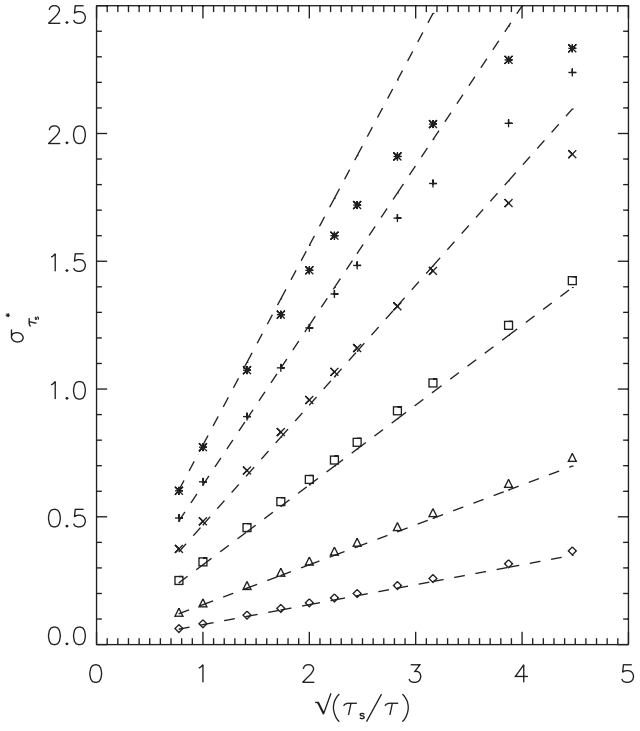

Fig. 1. Plots of observed $\sigma_{\tau_s}^*$ against $\sqrt{\tau_s/\bar{\tau}}$ for an unbiased velocity jump process with $\sigma_\delta = 0.1$ ($\diamond$), $\sigma_\delta = 0.2$ ($\triangle$), $\sigma_\delta = 0.4$ ($\square$), $\sigma_\delta = 0.6$ ($\times$), $\sigma_\delta = 0.8$ ($+$), and $\sigma_\delta = 1.0$ ($*$). The dashed lines are plots of $0.79\ \sigma_\delta\sqrt{\tau_s/\bar{\tau}}$. It is clear that when $\sigma_{\tau_s}^* < 1.2$ rad the linear relation between $\sigma_{\tau_s}^*$ and $\sqrt{\tau_s}$ holds but, when $\sigma_{\tau_s}^* > 1.2$ rad, the linear relation starts to break down and the sinuosity cannot be measured.

This difference arises because of the variable time between each turning event in the velocity jump process. When imposing a fixed sampling time step it is likely that, if the time between two particular turning events in the original walk is large, then the discretization process will introduce artificial 'zero turns' of $\delta = 0$ that were not in the original walk. This will lead to an underestimation of σ_δ, and it is the reason that our value of $1/\xi$ is found to be smaller than that found by Bovet and Benhamou (1988) in their spatial sampling of random walks that have a fixed step length between turns.

5. The effect of sampling rate on the apparent speed of an unbiased velocity jump process

The apparent speed of an unbiased and correlated random walk will decrease as the sampling time step increases since the apparent distance moved will get smaller as the walk is rediscretized. Averaging over the whole population, the average total displacement will always be zero in an unbiased random walk. This is in contrast to a biased random walk where there is an average displacement in the preferred direction that is dependent on the reorientation parameters κ (or σ_δ^2) and μ_δ and can be calculated (Codling and Hill, 2004). If we increase the sampling time step in an unbiased random walk, we expect the observed apparent speed to tend asymptotically to zero. It is also clear that the more sinuous the trajectory is, the more the apparent speed will decrease, and vice versa.

Simulations of 1000 walkers have been completed as described in the previous section. The trajectories are rediscretized with different sampling time steps, τ_s, and the observed apparent speed (averaged over the whole population), s^*, and the standard deviation of the apparent speed, s_σ^*, are calculated.

5.1. The observed speed

Fig. 2 shows plots of the negative natural logarithm of the observed speed, $-\log s^*$, against $\tau_s/\bar{\tau}$ for various values of the angular deviation, σ_δ, for simulations using a *fixed* speed. Using the method of least squares, we have fitted linear functions to the observed data for which $s^* > 0.75$. This value of s^* corresponds to the sampling time steps in the previous section where the observed angular deviation becomes too large, $\sigma^*_{\tau_s} < 1.2$ rad (compare Figs. 1 and 2). The resulting relation between s^* and $\tau_s/\bar{\tau}$ is given by

$$s^* = \exp(-0.074\ \sigma_\delta^2 \tau_s/\bar{\tau}). \tag{11}$$

Similar simulations were also completed with a *variable* speed across the population. In these simulations, each walker still has a fixed speed, s, but this is now randomly drawn for each walker from a normal distribution,

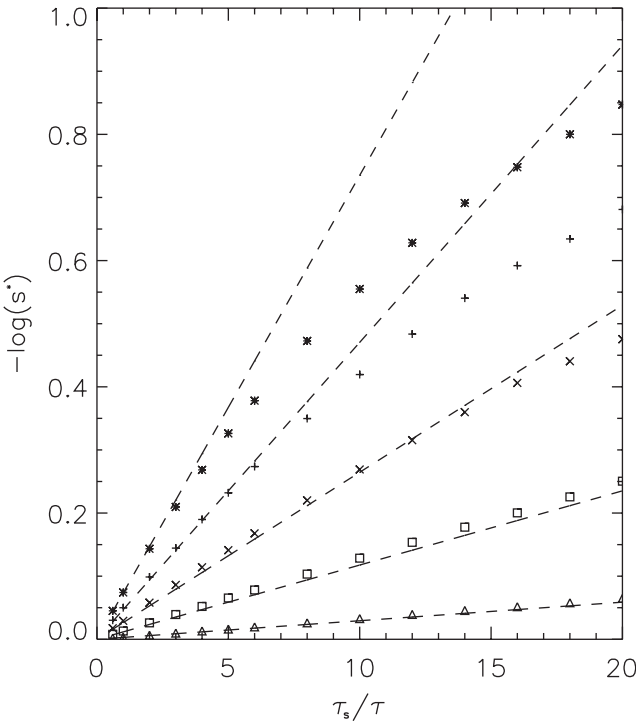

Fig. 2. Plots of the natural logarithm of the apparent speed $-\log(s^*)$ against $\tau_s/\bar{\tau}$ for an unbiased velocity jump process with a speed of $\hat{v} = 1$ fixed over the population. Simulations have been run with $\sigma_\delta = 0.2$ (△), $\sigma_\delta = 0.4$ (□), $\sigma_\delta = 0.6$ (×), $\sigma_\delta = 0.8$ (+), and $\sigma_\delta = 1.0$ (✻). The dashed lines are plots of $0.074\ \sigma^2\ \tau_s/\bar{\tau}$. Simulations using a variable speed across the population produce the same results.

$N(1, 0.5)$, centred on 1 with standard deviation 0.5. We further restrict these speeds by rejecting any draws where $s \leqslant 0$ or $s \geqslant 2$ to ensure there are no non-sensical negative speeds and a symmetric distribution. This results in a symmetric distribution of speeds, with an average of 1 but with a standard deviation ≈ 0.44. From simulations it is found that the observed speed, s^*, is unaffected by the introduction of a variable speed into the population of walkers and the results are thus omitted.

5.2. The observed standard deviation of the speed

The standard deviation of the apparent speed, s_σ^*, measures the variability of the observed speed. It is not immediately clear how the standard deviation of the observed speed would be expected to change as the sampling time step increases for a random walk with a *fixed* speed. However, even though the speed is constant in the original random walk, if $s_\sigma^* > 0$ then rediscretizing the trajectories produces variability in the observed speed. Fig. 3(a) shows plots of the observed standard deviation of the apparent speed, s_σ^*, against $\tau_s/\bar{\tau}$ for various values of the angular deviation, σ_δ, for simulations using a fixed speed. Using the method of least squares, we have fitted linear functions of the form

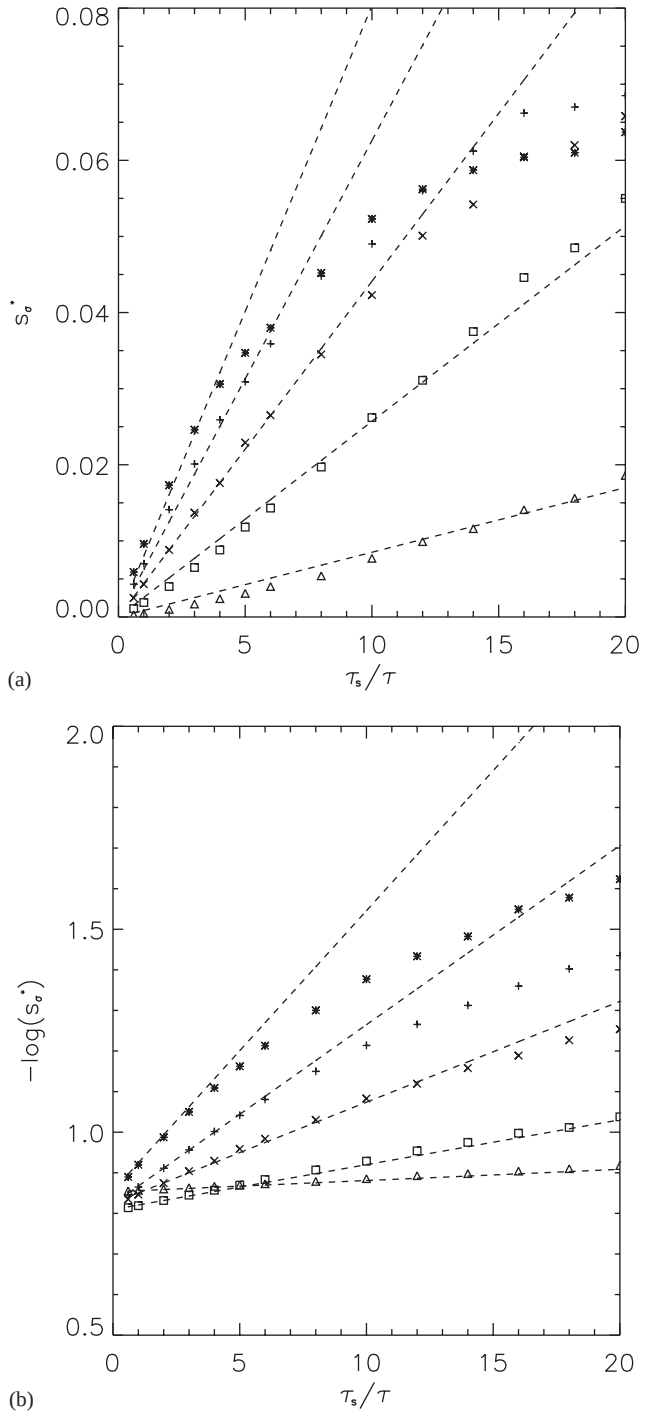

Fig. 3. Plots of (a) the standard deviation of the apparent speed s_σ^* and (b) the negative natural logarithm of the standard deviation of the apparent speed $-\log(s_\sigma^*)$, against $\tau_s/\bar{\tau}$ for an unbiased velocity jump process with (a) a fixed speed and (b) a variable speed over the population. Simulations have been run with $\sigma_\delta = 0.2$ (△), $\sigma_\delta = 0.4$ (□), $\sigma_\delta = 0.6$ (×), $\sigma_\delta = 0.8$ (+), and $\sigma_\delta = 1.0$ (✻). In (a) the dashed lines are plots of $f(x) = a\,\tau_s/\bar{\tau}$, where a has been calculated by the method of least squares for $s_\sigma^* < 0.04$ and $a = 0.00085$ for $\sigma_\delta = 0.2$, $a = 0.00257$ for $\sigma_\delta = 0.4$, $a = 0.00441$ for $\sigma_\delta = 0.6$, $a = 0.00626$ for $\sigma_\delta = 0.8$, and $a = 0.00803$ for $\sigma_\delta = 1.0$. In (b) the dashed lines are plots of $-\log(s_\sigma) + 0.069\,\sigma_\delta^2\,\tau_s/\bar{\tau}$, where s_σ is the standard deviation of the speed in the original velocity jump process.

$f(x) = ax$ (where $x = \tau_s/\bar{\tau}$) to the observed data for which $s_\sigma^* < 0.04$ (except for $\sigma_\delta = 1.0$ where we have fitted a linear function for $s_\sigma^* < 0.03$). The gradients of the linear functions are given by $a = 0.00085$ for $\sigma_\delta = 0.2$, $a = 0.00257$ for $\sigma_\delta = 0.4$, $a = 0.00441$ for $\sigma_\delta = 0.6$, $a = 0.00626$ for $\sigma_\delta = 0.8$, and $a = 0.00803$ for $\sigma_\delta = 1.0$. The sampling time steps where the linear relations seem to break down correspond to the same sampling time steps in the earlier results where the linear or exponential relations are no longer valid. There is no obvious relation between the angular deviation σ_δ and the gradient of the linear function fitted. For $s_\sigma^* > 0.04$ and the larger values of $\sigma_\delta \geqslant 0.6$, there is no obvious relationship and the data are noisy.

The increase in the standard deviation of the observed speed with the increase in sampling time step can be explained by the nature of the velocity jump process and the fact that the speed is constant. As the sampling time step increases and the turning events are further apart in time, there is greater variability in the observed spatial distance between turns and hence the standard deviation of the observed speed increases. If the walk is highly correlated and like a straight line (small σ_δ values), there will be less variability in the speed when the sampling time step is increased and hence the standard deviation will be small when compared to a more sinuous walk.

Fig. 3(b) shows plots of $-\log s_\sigma^*$ against $\tau_s/\bar{\tau}$ for various values of the angular deviation, σ_δ, for simulations using a variable speed. Using the method of least squares, we have fitted linear functions to the observed data for which $s_\sigma^* > 0.33$. As before, the sampling time steps where the linear relations seem to break down correspond to the same sampling time steps in the earlier results where the linear or exponential relations are no longer valid. The general form of the relation between s_σ^* and $\tau_s/\bar{\tau}$ is given by

$$s_\sigma^* = s_\sigma \exp(-0.069\,\sigma_\delta^2 \tau_s/\tau), \tag{12}$$

where s_σ is the standard deviation of the speed in the original velocity jump process. Note that due to the random nature of the simulations, s_σ is different for every value of σ_δ, which is why the plots in Fig. 3(b) have different initial values.

In contrast to the behaviour of s_σ^* for a random walk with a fixed speed, when there is a variable speed s_σ^* decreases exponentially with sampling time step for $s_\sigma^* > 0.33$. The fact that the observed variability in the speed *decreases* with sampling time step is almost certainly due to smoothing of the trajectories. This is consistent with the fact that one would expect less smoothing in the random walks that are highly correlated and more like straight lines (small σ_δ), as when these trajectories are rediscretized less information is lost about the original walk when compared to a highly sinuous walk.

6. The random walk on a circle

Hill and Häder (1997) carried out experiments to observe and analyse the motion of swimming micro-organisms such as the algae *Chlamydomonas nivalis* and compared results for swimming direction to a theoretical random walk on a unit circle. *C. nivalis* is known to be influenced by both *gyrotaxis* due to the balance between gravitational and viscous torques as the algae are bottom heavy (Kessler, 1986), and *phototaxis* as the algae are sensitive to light (Hill and Vincent, 1993; Vincent and Hill, 1996).

6.1. Reorientation models

Consistent with random walk theory, Hill and Häder (1997) assumed that the mean turning angle, μ_δ, and the variance of the turning angle, σ_δ^2, are proportional to the time step between turns, τ, such that as $\tau \to 0$

$$\mu_\delta(\theta,\tau) = \mu_0(\theta)\tau, \tag{13}$$

$$\sigma_\delta^2(\theta,\tau) = \sigma_0^2(\theta)\tau. \tag{14}$$

Based on their experimental data, they suggested the following forms for the mean turning angle, μ_δ:

$$\mu_\delta(\theta,\tau) = -d_\tau \sin\theta, \quad -\pi \leqslant \theta < \pi \tag{15}$$

for *sinusoidal* reorientation corresponding to gyrotaxis, and

$$\mu_\delta(\theta,\tau) = \begin{cases} -d_\tau\theta, & -\pi < \theta < \pi, \\ 0, & \theta = \pm\pi \end{cases} \tag{16}$$

for *linear* reorientation corresponding to phototaxis, where $d_\tau = \tau/B$ is a dimensionless parameter, τ is the time step between turning events, B is the typical reorientation time, and without loss of generality, the preferred direction is assumed to be $\theta = 0$. The fact that the mean turning angle, μ_δ, is dependent on the previous direction of movement, θ, introduces *bias* into the motion. Hill and Häder (1997) assumed that σ_0^2 is a constant, and independent of the absolute angle, θ.

6.2. Expected long time absolute angular distributions

Hill and Häder (1997) showed that for sinusoidal reorientation, where μ_δ is as defined in (15), the normalized *steady-state* solution for the probability density of the absolute angle, $f(\theta)$, is the von Mises distribution

$$f(\theta) = M(\theta;\theta_0,2z) = \frac{1}{2\pi I_0(2z)}\exp(2z\cos(\theta)), \tag{17}$$

where $z = (B\sigma_0^2)^{-1}$.

For linear reorientation, where μ_δ is as defined in (16), the normalized steady-state solution for the probability density of the absolute angle is

$$f(\theta) = N(z)\exp(-z\theta^2), \tag{18}$$

where $N(z)$ is the normalization function defined by

$$N(z) = \left(\int_{-\pi}^{\pi}\exp(-z\theta^2)\,d\theta\right)^{-1} = \sqrt{z}\,(\sqrt{\pi}\,\mathrm{erf}(\pi\sqrt{z}))^{-1}. \tag{19}$$

Table 1
Parameter values observed by Hill and Häder (1997) in experiments on the swimming algae *C. nivalis*

Data set	$\bar{s}$ (μms^{-1})	s_σ (μms^{-1})	$d_\tau = B^{-1}\tau$	$\sigma_\delta(\tau)$ (rad)
C1	55	31	0.37 (0.80) τ	1.3 (2.0)$\sqrt{\tau}$
C3	60	41	0.44 (0.62) τ	1.8 (2.1)$\sqrt{\tau}$
C4	59	47	0.19 (0.61) τ	0.9 (1.7)$\sqrt{\tau}$

Values in brackets are estimates using only the smaller sampling time steps, $\tau < 0.4$ s.

6.3. Experimental observations of the reorientation parameters

The values of the reorientation parameters observed in experiments on swimming algae by Hill and Häder (1997) are given in Table 1, together with the observed mean speed, $\bar{s}$, and the standard deviation of the speed, s_σ. Data set C1 corresponds to *C. nivalis* moving in a vertical plane and exhibiting gyrotaxis, while C3 and C4 correspond to *C. nivalis* moving in a horizontal plane and exhibiting phototaxis due to light sources of 80 and 200 klux, respectively.

More recently, Vladimirov et al. (2000) used a laser tracking algorithm to follow swimming algae. They impose a sampling time step of $\tau_s = 1.1$ s and calculate the swimming velocity over a period of 20 s. This is likely to give a very large underestimate of the instantaneous swimming speed of the algae, and is thus only useful as an estimate of the long time average velocity.

7. Parameter estimation for a biased velocity jump process

7.1. The biased and smooth random walk as a velocity jump process

The method used by Hill and Häder (1997) to estimate the reorientation parameters assumes that the underlying random walk is smooth so that the walker is continuously turning. The velocity jump process model described in Section 3.1 uses discrete steps, but if the mean turning angle at each step becomes small this

process becomes approximately smooth. Thus we can approximate a continuously turning random walk by having a velocity jump process that has a reorientation kernel, (1), with small values for the angular variance σ_δ^2 and the mean turning angle μ_δ, see (13) and (14). Hence, from (15) and (16), d_τ is also small. This results in a reorientation kernel that is very sharply peaked about the mean turning angle, which itself is very close to the previous direction of movement, producing a '*pseudo-smooth*' velocity jump process.

7.2. Sinusoidal reorientation

Simulations of a velocity jump process have been completed using the algorithm described in Section 3.5, together with the sinusoidal reorientation model for the mean turning angle (15). All simulations were completed with 1000 walkers for 200 time units, a fixed speed $s = 1$, a fixed average time between turns $\bar{\tau} = 1$ (so that the system is non-dimensionalized), and a fixed mean turning angle amplitude $d_\tau = 0.005$ (so that the average reorientation time is 200 time units). A fixed value of the angular deviation σ_δ was also used for each simulation, but it should be noted that for the larger values of σ_δ the approximation to a smooth random walk becomes less valid.

Each original trajectory was rediscretized using linear interpolation with a given sampling time step τ_s. For all the rediscretized trajectories, the angular data was split into 18 bins of size $\pi/9$ corresponding to the value of the absolute angle θ. The average angular variance and mean turning angle were calculated for each bin. This is exactly the same method as used by Hill and Häder (1997) for their experimental data, except that they used 12 bins since they had fewer data points. In fact, for larger sampling time steps, Hill and Häder (1997) had as few as 1 or 2 data points in each bin and their results were liable to error. In our simulations there are as many as 350,000 data points when $\tau_s = 0.6$ time units, and even at the largest sampling time steps, there are a large number of data points (an average of 2000 data points at $\tau_s = 100$). Thus some of the problems that Hill and Häder (1997) encountered can be avoided.

The plots in Fig. 4 show how the observed mean turning angle, $\mu_\delta(\theta)$, changes with θ for various sampling time steps applied to a random walk with $\sigma_\delta = 0.1$. It is immediately clear that there is a strong sinusoidal relation between $\mu_\delta(\theta)$ and θ and, using the method of least squares, we have fitted functions of the form $f(\theta) = -a \sin\theta$ for each sampling time step τ_s. Plotting these values of a against τ_s suggests a linear relationship that we can use to estimate the original value of d_τ used in the random walk, see Fig. 6(a). Using the method of least squares gives $f(x) = 0.0039x$ where $x = \sqrt{\tau_s/\tau}$, which underestimates the true value of $d_\tau = 0.005$ by 20%. This underestimate is likely to be due to smoothing effects as we have observed earlier.

The plots in Fig. 5 show how the observed angular variance, $\sigma_\delta^2(\theta)$, changes with θ for various sampling time steps applied to a random walk with $\sigma_\delta = 0.1$. In the original random walk, σ_δ^2 is independent of θ and, for smaller sampling time steps in Fig. 5, this is observed. However, at large sampling time steps ($\tau_s > 20$), there is a small observed variance for the angular bins where $|\theta| \approx 0$ and a much larger variance for the bins where $|\theta|$ is large. This effect was also seen by Hill and Häder (1997). It is likely that this effect occurs because at large sampling time steps the most likely direction to be moving in is the preferred direction $\theta = 0$. Simulation results show that for large sampling time steps there are significantly more data points in the corresponding angular bins where $|\theta| \approx 0$, resulting in a lower observed variance. This effect is not observed in unbiased random walk simulations where there is no preferred direction. Because of the non-uniform distribution of data points across the angular bins, the

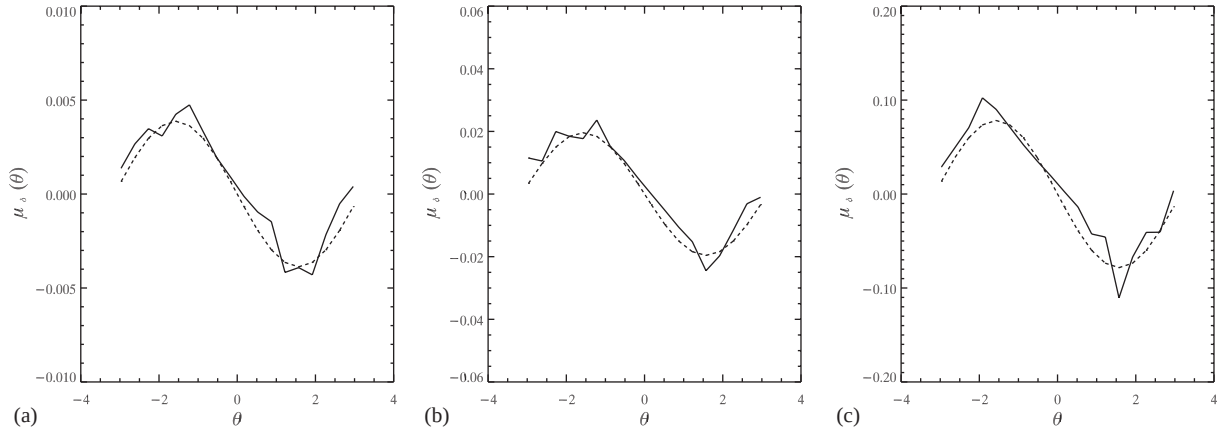

Fig. 4. Plots of the observed mean turning angle $\mu_\delta(\theta)$ vs. θ for simulations using the sinusoidal reorientation model with $d_\tau = 0.005$, $\sigma_\delta = 0.1$, $\bar{\tau} = 1$ and various sampling time steps τ_s. The dashed lines are sinusoidal functions of the form $f(\theta) = -a \sin(\theta)$ (where $a = d^*_{\tau_s}$) fitted using the method of least squares to find a. (The scale of each plot is different.) (a) $\tau_s = 1, \gamma = 0.0039$, (b) $\tau_s = 5, \gamma = 0.0196$ and (c) $\tau_s = 20, \gamma = 0.0783$.

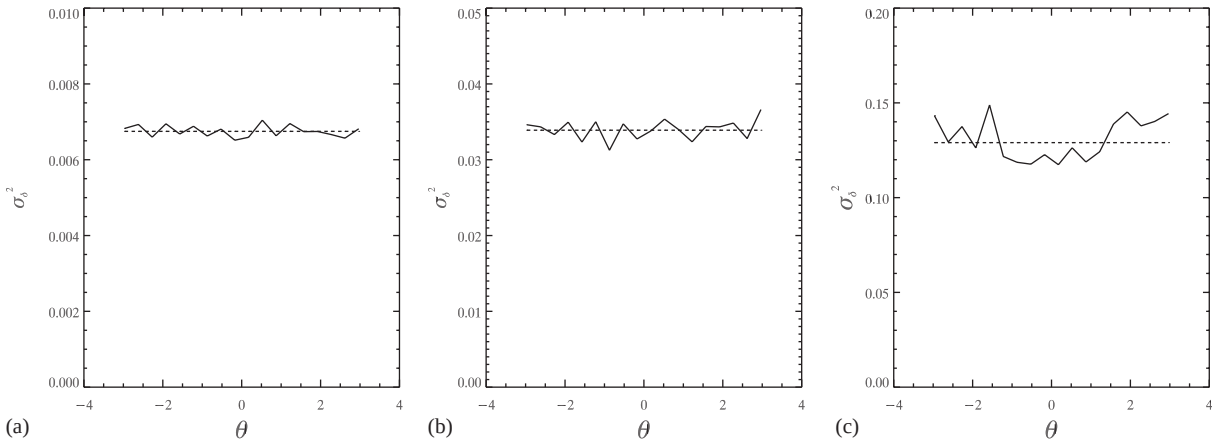

Fig. 5. Plots of the observed standard deviation of the turning angle $\sigma_\delta(\theta)$ vs. θ for simulations using the sinusoidal reorientation model with $d_\tau = 0.005$ and $\sigma_\delta = 0.1$ and various sampling time steps τ_s. The dashed lines are the value of $b = \sigma^*_{\tau_s}$. Calculated by averaging $\sigma_\delta(\theta)$ over all θ. (The scale of each plot is different.) (a) $\tau_s = 1, \sigma^*_{\tau_s} = 0.082$, (b) $\tau_s = 5, \sigma^*_{\tau_s} = 0.0184$ and (c) $\tau_s = 20, \sigma^*_{\tau_s} = 0.0359$.

average angular variance is close to the values in the bins where $|\theta| \approx 0$, and to find the average angular deviation, $\sigma^*_{\tau_s}$, we take the mean of the angular variance values over the whole population and all θ values. As in Section 4, plots of the observed values of $\sigma^*_{\tau_s}$ against τ_s suggest a linear relationship, see Fig. 6(b). Applying the method of least squares gives $f(x) = 0.79\sigma_\delta x$ where $x = \sqrt{\tau_s/\tau}$, which is identical to (10) for unbiased random walks.

A similar analysis of simulated data was completed for random walks with $d_\tau = 0.005$ and $\sigma_\delta = 0.2$, $\sigma_\delta = 0.4$ and $\sigma_\delta = 0.6$. The plots in Fig. 6 show the relations between (a) the observed value of the amplitude of the mean turning angle $d^*_{\tau_s}$ and τ_s, and (b) the observed angular deviation $\sigma^*_{\tau_s}$ and $\sqrt{\tau_s}$, for each of the original random walks.

From Fig. 6(a) it is clear that for $\sigma_\delta = 0.1$ there is a strong linear relation between $d^*_{\tau_s}$ and τ_s, but for larger values of σ_δ the linear relation does not hold and the data are very noisy, even with several thousand data points. Linear functions were fitted for the smaller values of τ_s where the data are less noisy ($\tau_s \leqslant 6$ for $\sigma_\delta = 0.2$ and $\tau_s \leqslant 3$ for $\sigma_\delta = 0.4$ and $\sigma_\delta = 0.6$), and the estimates for d_τ are given in Table 2. It appears that the underestimates of d_τ become worse as σ_δ increases, perhaps because the more sinuous walks (with large values of σ_δ) are affected by smoothing to a greater extent than walks that are straighter.

After least-squares analysis, all the plots in Fig. 6(b) are found to closely fit the relation given in (10) that was observed in unbiased random walks. There may be slightly less smoothing of the data in the biased walk, as the values in Table 2 are slightly larger than 0.79 (found from simulations with unbiased walks) but the difference does not appear to be significant. This suggests that, if there is only a small amount of bias present (as with the pseudo-smooth velocity jump process), then the observed sinuosity is unaffected and a linear relation between the observed angular variance, $\sigma^2_{\tau_s}$, and the sampling time step, τ_s, holds as long as the angular deviation is not too large ($\sigma^*_{\tau_s} > 1.2$).

The parameter estimates found from the data in Fig. 6 correspond to the variance per unit time, σ_0^2, and the inverse of the average reorientation time, $1/B$. However, the simulations are non-dimensionalized with respect to time as $s = 1$ and $\bar{\tau} = 1$, and from Section 6, $d_\tau = \bar{\tau}/B = 1/B$ and $\sigma^2_\delta = \bar{\tau}\sigma_0^2 = \sigma_0^2$ in this case. If the data are not scaled with respect to time then it is only possible to estimate d_τ and σ^2_δ for a discrete random walk if the average time between turns in the original trajectory, $\bar{\tau}$, is also known or estimated (see Section 6), which is a non-trivial problem. For continuum models for bioconvection (Hill and Pedley, 2004) or other examples that assume a smooth random walk, it is only necessary to know the unit parameters σ_0^2 and B, as $\bar{\tau}$ is assumed to tend to 0 in these models.

When analysing experimental data, Hill and Häder (1997) observed an approximately linear relationship between $d^*_{\tau_s}$ and τ_s and gave two possible estimates for B (recall that $d^*_{\tau_s} = \tau_s/B$) depending on whether they used only small sampling time steps or all sampling time steps. From our results, it seems their estimate using only the smaller sampling time steps is likely to be more accurate. However, Hill and Häder (1997) found that their data on the observed angular deviation $\sigma^*_{\tau_s}$ was too noisy to estimate σ_0 (the sinuosity or angular deviation per unit time). Using the observed long time angular distribution $f(\theta)$ together with the expected distribution given in (17) and their two estimates for B, they calculated two possible estimates for σ_0. We have completed a similar analysis using the simulated long time distribution for the absolute angle $f(\theta)$. The simulations were run for 1000 time units so the system is in an approximately steady state. The mean resultant length, R_f was calculated for the observed distribution

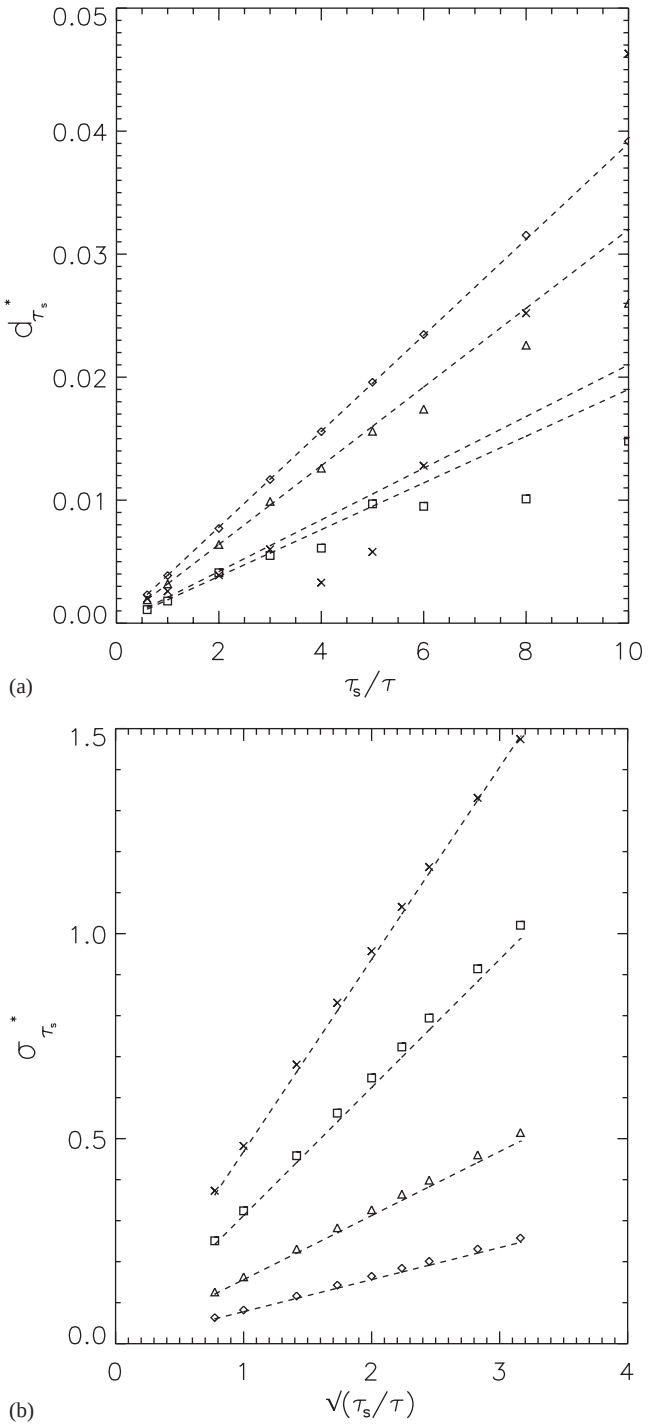

Fig. 6. Plots of (a) the observed amplitude of the mean turning angle, $d^*_{\tau_s}$ vs. $\tau_s/\bar{\tau}$ and (b) the observed average angular deviation $\sigma^*_{\tau_s}$ v $\sqrt{\tau_s/\bar{\tau}}$, for simulations using the sinusoidal reorientation model with $d_\tau = 0.005$ and $\sigma_\delta = 0.1$ ($\diamond$), $\sigma_\delta = 0.2$ ($\triangle$), $\sigma_\delta = 0.4$ ($\square$), $\sigma_\delta = 0.6$ ($\times$). The dashed lines have been fitted to the data using the method of least squares and take the form (a) $f(x) = a\,x$ and (b) $f(x) = b\,\sqrt{x}$, where $x = \tau_s/\bar{\tau}$. The values of a and b are given in Table 2.

$f(\theta)$, and a von Mises distribution with the corresponding value of R_f was fitted to the data, see Fig. 7. From (17), the concentration parameter of the fitted von Mises distribution is given by $2B\sigma_0^2 = 2d_\tau/\sigma_\delta^2$, so if the concentration parameter is calculated from R_f and either of the parameters B or σ_0 are known then the other parameter can be estimated. This may prove to be a useful technique in the situation described previously where the original random walk has a large value of σ_δ, and it is hard to estimate B but there is a good estimate of σ_0. Table 2 gives the estimated values of d_τ ($= 1/B$ in our non-dimensionalized simulations) calculated using a simple numerical solver routine from the mean resultant length of $f(\theta)$ and (17), assuming a perfect estimate of σ_0 (using (10) with our observed values of $\sigma^*_{\tau_s}$ does give a very good estimate of σ_0). Using this method with the random walks with small σ_δ, the estimates of d_τ are actually better than the estimates found using the least-squares analysis of the observed mean turning angle. However, for the larger values of σ_δ the estimates are less good because the data in the final angular distribution $f(\theta)$ are noisy, see Fig. 7. Similarly, we have calculated estimates of σ_0, assuming that we have a perfect estimate of d_τ, and the results are given in Table 2.

Table 2
Estimates of the parameters $d^*_{\tau_s}$ (the inverse of the observed average reorientation time) and $\sigma^*_{\tau_s}$ (the observed angular deviation) calculated using data from simulations of biased random walks using sinusoidal reorientation

Sim. d_τ	Sim. σ_δ	(a) Est. $d^*_{\tau_s}$	(a) Est. $\sigma^*_{\tau_s}$	R_f	(b) Est. $d^*_{\tau_s}$	(b) Est. $\sigma^*_{\tau_s}$
0.005	0.1	0.0039	$0.79\sigma_\delta$	0.4407	0.0049	0.1008
0.005	0.2	0.0032	$0.81\sigma_\delta$	0.1219	0.0049	0.2017
0.005	0.4	0.0019†	$0.81\sigma_\delta$	0.0342	0.0054	0.3822
0.005	0.6	0.0021†	$0.79\sigma_\delta$	0.0189	0.0068†	0.5143†

Estimates are calculated using (a) a least-squares fit to the linear relations between the observed parameters and sampling time step, τ_s; (b) a simple numerical solver routine where the parameters are found using the concentration parameter of $f(\theta)$, the steady-state density function for the absolute angle, and the mean resultant length, R_f. Estimates noted † are likely to be unreliable as the data is noisy due to the large value of σ_δ in the original walk.

7.3. Linear reorientation

Simulations have also been completed in exactly the same manner as described in the previous section with 1000 walkers for 200 time units, but using the linear reorientation model given in (16) in the pseudo-smooth velocity jump process.

The plots in Fig. 8 show how the observed mean turning angle, $\mu_\delta(\theta)$, changes with θ for various sampling time steps applied to a random walk with $\sigma_\delta = 0.1$. Allowing for noisy data at $|\theta| \approx \pi$, there is seen to be a strong linear relation between $\mu_\delta(\theta)$ and θ. Using the method of least squares, we have fitted functions of the

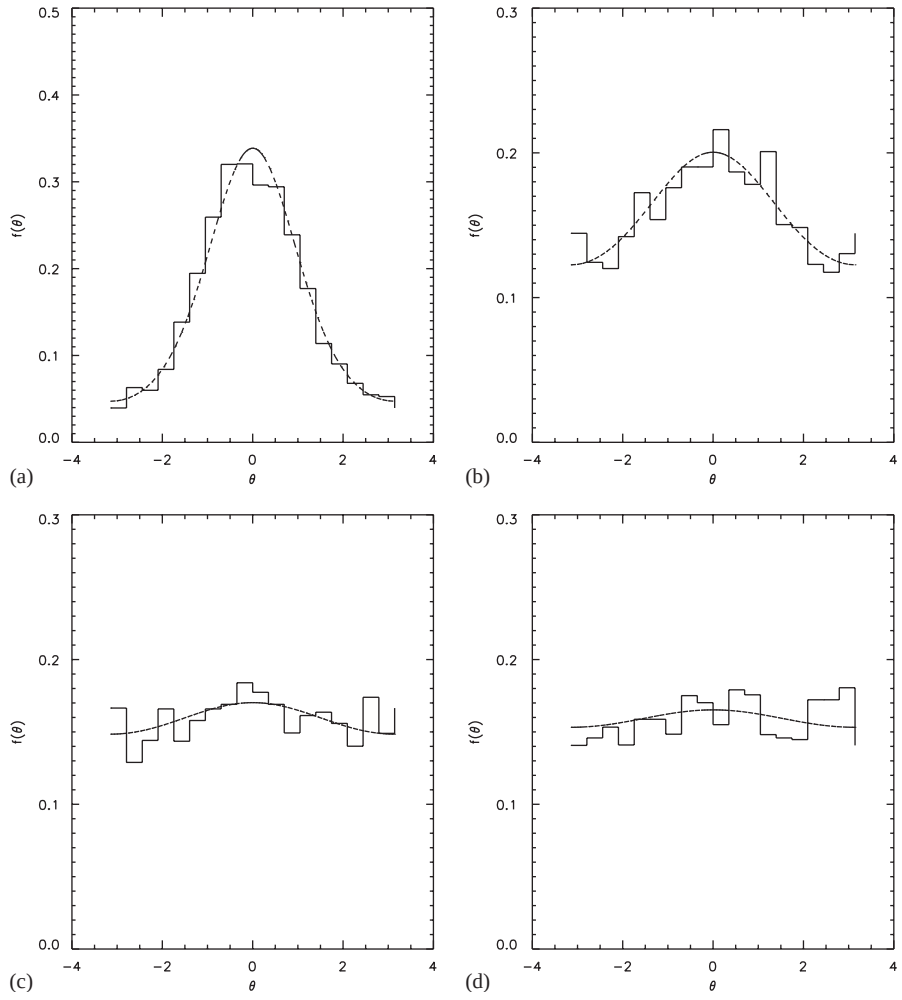

Fig. 7. Plots of the long time absolute angular distribution $f(\theta)$ after 1000 time units for the sinusoidal reorientation model with $d_\tau = 0.005$ and (a) $\sigma_\delta = 0.1$ and $R_f = 0.4407$, (b) $\sigma_\delta = 0.2$ and $R_f = 0.1219$, (c) $\sigma_\delta = 0.4$ and $R_f = 0.0342$, and (d) $\sigma_\delta = 0.6$ and $R_f = 0.0189$, where R_f is the mean resultant length of $f(\theta)$ calculated from the data. It is possible to fit a von Mises distribution with the value of R_f to the data (dashed lines in plots), and assuming σ_δ is known or has been estimated from the data one can use the corresponding concentration parameter to estimate d_τ using the theoretical equation for $f(\theta)$ (or vice versa).

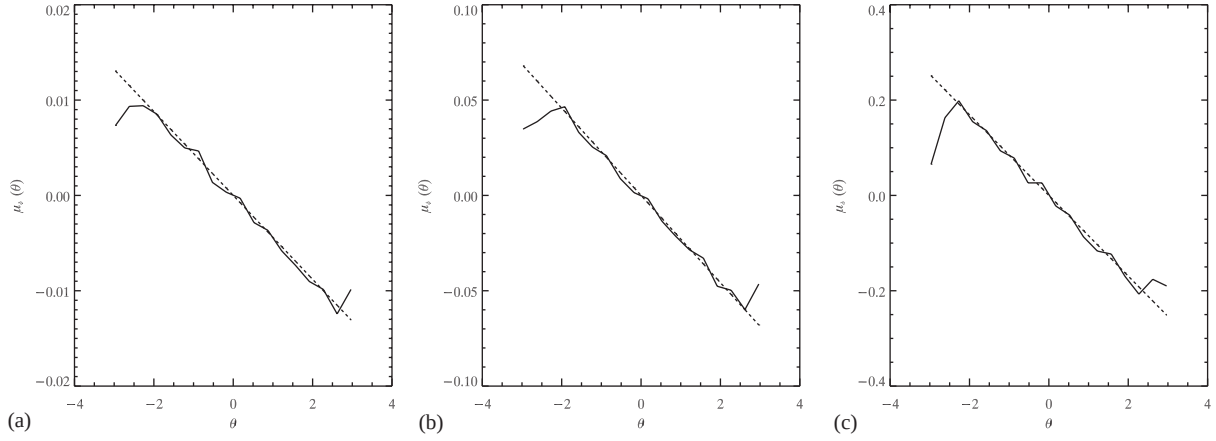

Fig. 8. Plots of the observed mean turning angle $\mu_\delta(\theta)$ vs. θ for simulations using the linear reorientation model with $d_\tau = 0.005$, $\sigma_\delta = 0.1$, $\bar{\tau} = 1$ and various sampling time steps τ_s. The dashed lines are linear functions of the form $f(\theta) = -a\,\theta$ fitted using the method of least squares to find a. (The scale of each plot is different.) (a) $\tau_s = 1, \gamma = 0.0044$, (b) $\tau_s = 5, \gamma = 0.0229$ and (c) $\tau_s = 20, \gamma = 0.0684$.

form $f(\theta) = -a\theta$ for each sampling time step τ_s. Plotting the values of a against the corresponding values of τ_s results in a linear relation that we can use to estimate the original value of d_τ used in the random walk (since we have a non-dimensionalized system), see Fig. 9(a). The linear relation was found to be given by $f(x) = 0.0043x$ where $x = \sqrt{\tau_s/\tau}$, which slightly underestimates the true value of $d_\tau = 0.005$ due to smoothing effects. Simulations were also completed for $d_\tau = 0.005$ and $\sigma_\delta = 0.2$, $\sigma_\delta = 0.4$ and 0.6, and the resulting data analysed in a similar way.

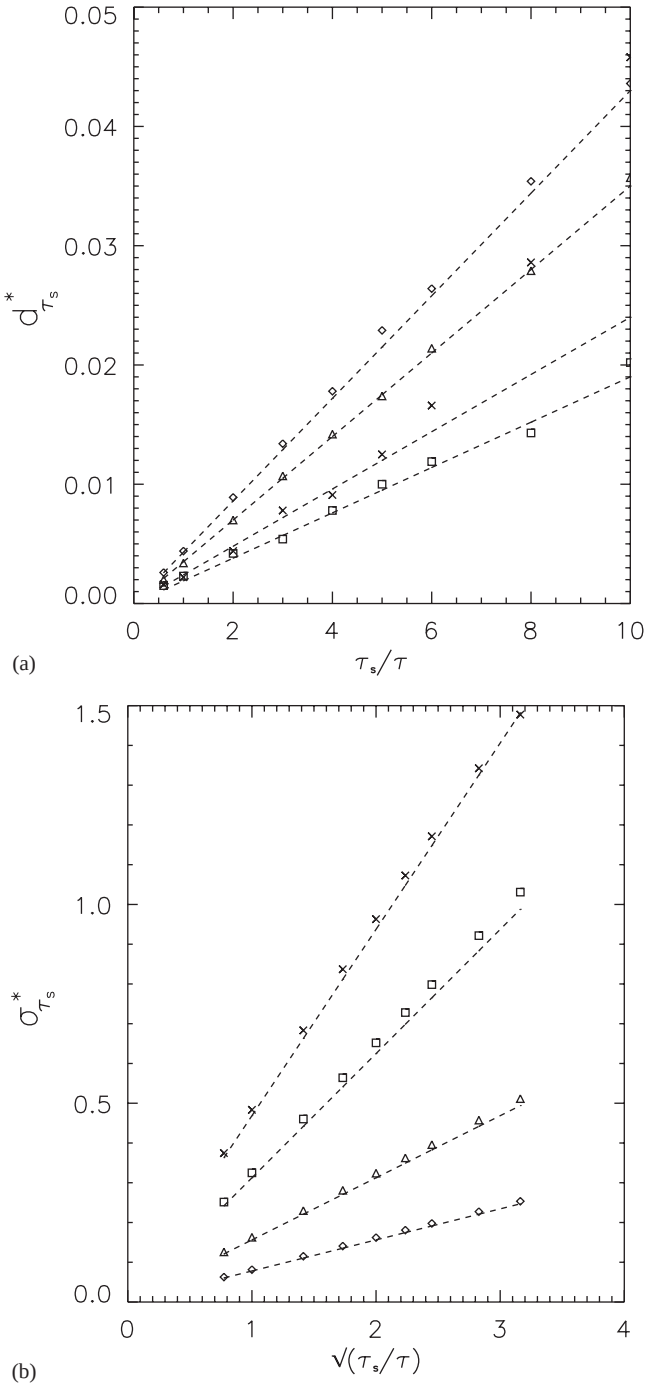

Fig. 9. Plots of (a) the observed amplitude of the mean turning angle, $d^*_{\tau_s}$ vs. $\tau_s/\bar{\tau}$ and (b) the observed average angular deviation $\sigma^*_{\tau_s}$ vs. $\sqrt{\tau_s/\bar{\tau}}$, for simulations using the linear reorientation model with $d_\tau = 0.005$ and $\sigma_\delta = 0.1$ ($\diamond$, $\sigma_\delta = 0.2$ ($\triangle$), $\sigma_\delta = 0.4$ ($\square$), $\sigma_\delta = 0.6$ ($\times$). The dashed lines have been fitted to the data using the method of least squares and take the form (a) $f(x) = a\,x$ and (b) $f(x) = b\,\sqrt{x}$, where $x = \tau_s/\bar{\tau}$. The values of a and b are given in Table 3.

As with the sinusoidal results, Fig. 9(a) shows that there is a strong linear relation between $d^*_{\tau_s}$ and τ_s, only for the smaller values of σ_δ and τ_s. At larger values there is no clear relation between the mean turning angle, $\mu_\delta(\theta)$, and the absolute angle, θ, and although we fit linear functions using the method of least squares this becomes almost arbitrary and produces large residuals. The estimated values of d_τ from the observed data are given in Table 3.

The observed angular variance, $\sigma^2_\delta(\theta)$, for the linear reorientation simulations is found to behave in a similar manner to the observed angular variance of the sinusoidal model as described previously and shown in Fig. 5. At larger sampling time steps, the angular variance is no longer independent of θ and is larger for the angular bins where $|\theta| \approx \pi$ because of the paucity of data points.

The plots in Fig. 9(b) clearly demonstrate the linear relation between the observed angular deviation, $\sigma^*_{\tau_s}$, and the square root of the sampling time step, τ_s. As with the sinusoidal model and the unbiased simulations, (10), the function $f(x) = 0.79\sigma_\delta x$ where $x = \sqrt{\tau_s/\tau}$ gives the best fit to the data using least-squares analysis (Table 3).

The expected long time distribution $f(\theta)$ for linear reorientation is given in (18). The mean resultant length, R_f, was calculated for the observed distribution, $f(\theta)$, after 1000 time units and a distribution with the form given in (18) was fitted to the data. From (18), the concentration parameter of the fitted distribution is given by d_τ/σ^2_δ (recall that we use non-dimensionalized simulations), so if the concentration parameter is calculated from R_f and either of the parameters d_τ or σ_δ are known then the other parameter can be estimated. Table 3 gives the estimated values of d_τ and σ_δ

Table 3
Estimates of the parameters $d^*_{\tau_s}$ and $\sigma^*_{\tau_s}$ calculated using data from simulations of biased random walks using linear reorientation

Sim. d_τ	Sim. σ_δ	(a) Est. $d^*_{\tau_s}$	(a) Est. $\sigma^*_{\tau_s}$	R_f	(b) Est. $d^*_{\tau_s}$	(b) Est. $\sigma^*_{\tau_s}$
0.005	0.1	0.0043	$0.79\sigma_\delta$	0.5999	0.0048	0.1016
0.005	0.2	0.0035	$0.81\sigma_\delta$	0.2108	0.0046	0.2085
0.005	0.4	$0.0019^\dagger$	$0.81\sigma_\delta$	0.0617	0.0050	0.3985
0.005	0.6	$0.0024^\dagger$	$0.79\sigma_\delta$	0.0331	$0.0060^\dagger$	$0.5468^\dagger$

Estimates are calculated using the methods as described for Table 2. Estimates noted † are likely to be unreliable as the data is noisy due to the large value of σ_δ in the original walk.

calculated using a simple numerical solver routine from the observed mean resultant length of $f(\theta)$ and (18), assuming that we have a perfect estimate of one of the parameters. As with the sinusoidal model, the estimates of d_τ using this method are better than the least-squares analysis except when the original random walks have a large angular deviation and the data then becomes noisy. For the same reorientation parameter values (d_τ and σ_δ), the mean resultant length is larger for the linear reorientation model. This can be explained by the fact that the linear model is a 'better' model for reorientation in the sense that in the linear model the maximum average reorientation back to the preferred direction ($\theta = 0$) is when facing directly away from the preferred direction ($\theta = \pi$), while the sinusoidal model will have a zero average reorientation when $\theta = \pi$.

7.4. Effects of sampling rate on apparent speed

Simulations have been completed in the same manner as described in Section 5 but using the sinusoidal and linear reorientation models. For the unbiased simulations, the apparent speed is related to the observed angular variance and sampling time step used, see (11). In the previous section, we demonstrated that when d_τ and σ_δ are both small, there is no difference between the observed angular deviation of biased and unbiased random walks, and thus we expect the apparent speed to behave in the same way as well. Simulations completed with $d_\tau = 0.005$ and various values of σ_δ have confirmed exactly the same behaviour as observed in Figs. 2 and 3 and the relations given in (11) and (12) are found to hold for $s^* > 0.75$.

When tracking the trajectories of swimming micro-organisms, Hill and Häder (1997) observed a steep decrease in the apparent speed of the organisms for small sampling time steps which they attributed to 'pixel noise' due to discretization of the video images at these small length scales. When looking at much larger sampling time steps, they then observed a linear decrease in the apparent speed which they used to extrapolate back and predict the original speed. While it is certainly true that pixel noise may have influenced the observations at small sampling time steps, our results suggest that perhaps Hill and Häder (1997) should have looked for a decaying exponential function similar to (11) rather than a linear function to extrapolate an estimate of the average speed (at very large sampling time steps an exponentially decaying function will appear approximately linear as it tends asymptotically to zero). This would result in a much larger estimate for the speed.

As with the speed, Hill and Häder (1997) observed a sharp decrease in the standard deviation of the speed at small sampling time steps, while the decrease appeared linear at larger sampling time steps. Again, this may be due to pixel noise, or it could be that they should have used a decaying exponential function similar to (12) rather than a linear function to extrapolate an estimate for the true standard deviation of the speed. The fact that Hill and Häder (1997) observed a decrease in the standard deviation with sampling time step implies that there was a smoothing of the apparent speeds and as we have shown previously, this suggests that there was a large variability in the true speed in the original trajectories (recall that when the speed is fixed over the population the standard deviation of the observed speed increased with sampling time step).

8. Conclusions and discussion

We have demonstrated that the method of calculating the sinuosity of a correlated and unbiased random walk given by Bovet and Benhamou (1988) can be extended to an *unbiased* velocity jump process for which temporal sampling is used. Due to the inherent variability of the time between turning events in the original random walk with a velocity jump process, we find a slightly smaller value for the constant that takes into account smoothing effects. The sinuosity is defined as the linear relation between the observed angular deviation $\sigma^*_{\tau_s}$ and the root of the sampling time step τ_s. If the linear relation exists then the sinuosity can be used to find the average angular deviation per unit time σ_0 which can be used in continuum models, or if the time between turns in the original random walk is known or can be estimated then the original angular deviation used in the walk can be calculated. As with Bovet and Benhamou (1988), we find that the linear relation between the angular deviation and the root of the sampling time step breaks down when the observed angular deviation becomes too large. Thus, if the original random walk is highly correlated then a larger sampling time step can be used before the linear relation breaks down, when compared to a more sinuous walk.

We have also demonstrated that the observed apparent speed fits a decreasing exponential function as the sampling time step increases, and this relation breaks down at the same sampling time steps where the linear relation between the angular deviation and the root of the sampling time step breaks down.

These results have been shown to also hold for a *biased* random walk that has been simulated using a pseudo-smooth velocity jump process. This model seems to be realistic approximation to the movement of continuously turning walkers such as swimming algae. We demonstrate that the method used by Hill and Häder (1997) is valid if the assumption that the original random walk is smooth holds (i.e. angular deviation and mean turning angle are small) and the sampling time step is not too large so that the linear relations between the observed angular variance, $(\sigma^*_{\tau_s})^2$, and the observed

amplitude of the mean turning angle, $d^*_{\tau_s}$ and the sampling time step, τ_s, do not break down. It is possible to use the observed long time distribution for the absolute angle, $f(\theta)$, together with the expected distribution to either check the estimates of the parameters found or to calculate an unknown parameter value if the other parameter value is the only estimate likely to be reliable.

Revisiting the results of Hill and Häder (1997), it seems likely that their sampling time step was too large because they observed linear relations only for the very smallest sampling time steps, and also their data was extremely noisy. However, their minimum sampling time step was determined by experimental constraints. Hill and Häder (1997) were unable to estimate smoothing effects, and we have demonstrated that even with a strong linear relation between the observed data and the sampling time step, the parameter estimate is likely to be approximately 80% of the actual value if the original random walk is a velocity jump process. Bovet and Benhamou (1988) showed that in a walk with a fixed step length the estimated angular deviation will be approximately 85% of the actual value. Thus the results of Hill and Häder (1997) are likely to underestimate the true reorientation parameters and also possibly the true speed and standard deviation of the speed.

The method of Hill and Häder (1997) is likely to break down if the reorientation parameters used in the original random walk are large, or if too great a sampling time step is used. We have demonstrated that, if the original random walk has a large angular deviation σ_δ, then the linear relations between the sampling time step τ_s and the observed values of $d^*_{\tau_s}$ and $(\sigma_\delta)^2$ break down. Such random walks can no longer be considered approximately smooth (continuously turning). Simulations have also been completed where the original random walk has a large value for the parameter d_τ, so that the average reorientation time is small (this also can no longer be considered an approximately smooth walk). We find that in such walks, when larger sampling time steps are applied and the trajectories are analysed the linear relation between $d^*_{\tau_s}$ and τ_s is either non-existent or only holds for a very small range of sampling time steps. In fact, in the case of linear reorientation, $d^*_{\tau_s}$ will never be observed to be greater than one, see Codling (2003). If in the original random walk $d_\tau \approx 1$, then $d^*_{\tau_s}$ is not proportional to τ_s for sampling times that are larger than the average time between turns in the original walk ($\tau_s > \bar{\tau}$). Similarly, simulations show that if the sampling time step applied to any biased random walk is large enough, then the observed angular deviation will decrease as τ_s increases. This is because at every sampling time step it becomes increasingly likely for a walker to be observed moving in the preferred direction, which reduces the angular deviation when averaging over each trajectory. Further discussion can be found in Codling (2003), but these are extreme cases.

The results we have demonstrated may not be relevant in all random walk models. We have assumed that all our walkers move with the same constant reorientation parameters and our results may not be valid if there is a large variation in turning behaviour between individuals. We have also only considered biased random walks where there is a fixed preferred direction and a fixed magnitude of bias. If there is a population of walkers moving to a point source (e.g. fish larvae recruiting to a reef in Codling et al., 2004), or moving up a gradient where the signal strength of the bias increases in magnitude (e.g. in chemotaxis models in Othmer and Hillen, 2002, etc), then the results given here may not generalize. Any similar analysis is likely to be complicated and would have to take into account the fact that the preferred direction and magnitude of bias will change with the position of each walker.

However, although any quantitative results are likely to be different in more complicated random walks, it is clear that care should be taken with any interpretation of data that has been collected by tracking the movements of animals and microorganisms. We have demonstrated how it is easy for the rediscretization of the movement path that is imposed by the observer to affect the measured properties of the trajectory and any possible estimates of parameters that are made. It is worth bearing this in mind, when considering the large number of experiments in the literature that have observed animal movement to fit the simple random walk and diffusion model, as these may have been carried out with too large a sampling time step (Bovet and Benhamou, 1988).

Acknowledgements

We would like to thank Dr. D.J. Read for helpful discussions and an anonymous referee for comments that helped improve the manuscript. E.A.C. has been supported in this research by a studentship from the E.P.S.R.C. in the UK.

References

Batschelet, E., 1981. Circular Statistics in Biology. Academic Press, London.

Berg, H.C., 1983. Random Walks in Biology. Princeton University Press, Princeton, NJ.

Bovet, P., Benhamou, S., 1988. Spatial analysis of animals' movements using a correlated random walk model. J. Theor. Biol. 131, 419–433.

Brown, R., 1828. A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and the general existence of active

molecules in organic and inorganic bodies. Philos. Mag. (new series) 4, 161–173.

Byers, J.A., 2001. Correlated random walk equations of animal dispersal resolved by simulation. Ecology 82, 1680–1690.

Codling, E.A., 2003. Biased random walks in biology. Ph.D. Thesis, University of Leeds. http://maths.leeds.ac.uk/Applied/phd/codling.html>.

Codling, E.A., Hill, N.A., 2004. Calculating spatial statistics for velocity jump processes with experimentally observed reorientation parameters. J. Math. Biol., in press.

Codling, E.A., Hill, N.A., Pitchford, J.W., Simpson, S.D., 2004. Random walk models for the movement and recruitment of reef fish larvae. Mar. Ecol. Prog. Ser., in press.

Dunn, G.A., 1983. Characterising a kinesis response: time averaged measures of cell speed and directional persistence. In: Keller, H.O., Till, G.O. (Eds.), Leukocyte Locomotion and Chemotaxis. Birkhauser, Basel.

Einstein, A., 1906. Zur Theorie der Brownschen Bewegung. Ann. Phys. 19, 371–381.

Fisher, N.I., Best, D.J., 1979. Efficient simulation of the von Mises distribution. Appl. Statist. 28, 152–157.

Goldstein, S., 1951. On diffusion by discontinuous movements, and on the telegraph equation. Quart. J. Mech. Appl. Math. 6, 129–156.

Grimmett, G., Stirzaker, D., 2001. Probability and Random Processes. Oxford University Press, Oxford.

Hill, N.A., Häder, D.P., 1997. A biased random walk model for the trajectories of swimming micro-organisms. J. Theor. Biol. 186, 503–526.

Hill, N.A., Pedley, T.J., 2004. Bioconvection. Fluid Dyn. Res., submitted.

Hill, N.A., Vincent, R.V.V., 1993. A simple model and strategies for orientation in phototactic micro-organisms. J. Theor. Biol. 163, 223–235.

Hill, N.A., Pedley, T.J., Kessler, J.O., 1989. Growth of bioconvection patterns in a suspension of gyrotactic micro-organisms in a layer of finite depth. J. Fluid Mech. 208, 509–543.

Hillen, T., 2002. Hyperbolic models for chemosensitive movement. Math. Model. Methods Appl. Sci. 12 (7), 1007–1034.

Hillen, T., Othmer, H.G., 2000. The diffusion limit of transport equations derived from velocity-jump processes. SIAM J. Appl. Math. 61(3), 751–775.

Kac, M., 1974. A stochastic model related to the telegraphers equation. Rocky Mountain J. Maths. 4, 497–509.

Kareiva, P.M., Shigesada, N., 1983. Analyzing insect movement as a correlated random walk. Oecologia 56, 234–238.

Kessler, J.O., 1986. Individual and collective fluid dynamics of swimming cells. J. Fluid Mech. 73, 191–205.

Mardia, K.V., Jupp, P.E., 1999. Directional Statistics. Wiley, Chichester.

McCulloch, C.E., Cain, M.L., 1989. Analyzing discrete movement data as a correlated random walk. Ecology 70, 383–388.

Nossal, R., Weiss, G.H., 1974. A descriptive theory of cell migration on surfaces. J. Theor. Biol. 47, 103–113.

Okubo, A., 1980. Diffusion and Ecological Problems: Mathematical Models. Springer, Berlin.

Othmer, H.G., Hillen, T., 2002. The diffusion limit of transport equations II: chemotaxis equations. SIAM J. Appl. Math. 62(4), 1222–1250.

Othmer, H.G., Dunbar, S.R., Alt, W., 1988. Models of dispersal in biological systems. J. Math. Biol. 26, 263–298.

Siniff, D.P., Jessen, C.R., 1969. A simulation model of animal movement patterns. Adv. Ecol. Res. 6, 185–219.

Skellam, J.G., 1951. Random dispersal in theoretical populations. Biometrika 38, 196–218.

Skellam, J.G., 1973. The formulation and interpretation of mathematical models of diffusionary processes in population biology. In: Bartlett, M.S., Hiorns, R.W. (Eds.), The Mathematical Theory of the Dynamics of Biological populations. Academic Press, New York.

Tchen, C.M., 1952. Random flight with multiple partial correlations. J. Chem. Phys. 20, 214–217.

Vincent, R.V.V., Hill, N.A., 1996. Bioconvection in a suspension of phototactic algae. J. Fluid Mech. 327, 343–371.

Vladimirov, V.A., Denissenko, P.V., Pedley, T.J., Wu, M., Moskalev, I.S., 2000. Algal motility measured by a laser-based tracking method. Mar. Freshwater Res. 51, 589–600.