ORIGINAL ARTICLE

Estimating 3D Movements from 2D Observations Using a Continuous Model of Helical Swimming

Eliezer Gurarie · Daniel Grünbaum · Michael T. Nishizaki

Received: 9 April 2010 / Accepted: 6 July 2010 © Society for Mathematical Biology 2010

Abstract Helical swimming is among the most common movement behaviors in a wide range of microorganisms, and these movements have direct impacts on distributions, aggregations, encounter rates with prey, and many other fundamental ecological processes. Microscopy and video technology enable the automated acquisition of large amounts of tracking data; however, these data are typically two-dimensional. The difficulty of quantifying the third movement component complicates understanding of the biomechanical causes and ecological consequences of helical swimming. We present a versatile continuous stochastic model-the correlated velocity helical movement (CVHM) model-that characterizes helical swimming with intrinsic randomness and autocorrelation. The model separates an organism's instantaneous velocity into a slowly varying advective component and a perpendicularly oriented rotation, with velocities, magnitude of stochasticity, and autocorrelation scales defined for both components. All but one of the parameters of the 3D model can be estimated directly from a two-dimensional projection of helical movement with no numerical fitting, making it computationally very efficient. As a case study, we estimate swimming parameters from videotaped trajectories of a toxic unicellular alga, Heterosigma akashiwo (Raphidophyceae). The algae were reared from five strains originally collected from locations in the Atlantic and Pacific Oceans, where they have caused Harmful Algal Blooms (HABs). We use the CVHM model to quantify cell-level and strain-level differences in all movement parameters, demonstrating the utility of the model for identifying strains that are difficult to distinguish by other means.

E. Gurarie (🖂)

Department of Biosciences, University of Helsinki, Box 64, Viikinkaari 1, 00014 Helsinki, Finland e-mail: eliezer.gurarie@helsinki.fi

D. Grünbaum School of Oceanography, University of Washington, Seattle, USA

Keywords Helical motion · Continuous movement models · Heterosigma akashiwo · Microvideography

1 Introduction

Many microorganisms, including bacteria, most freely swimming protists, spores of many species of fungi and plants, spermatozoa of many marine invertebrates, and some larval invertebrates swim in helical trajectories (Crenshaw 1996). The ubiquity of helical movements has been recognized for a century (Jennings 1901), and considerable work has been conducted developing a formalism for quantitatively describing helical movements (Crenshaw 1993; Crenshaw et al. 2000). Mathematical models of helical individual movement tend to focus on characterization of tangential and rotational velocities of individuals. Less emphasis has been placed on intrinsic stochasticity or population level variability (Crenshaw 1996; Crenshaw et al. 2000), both of which can have important impacts on the population scale (Gurarie et al. 2009a). On the other hand, empirical descriptions of microorganism movement tend to deemphasize the helical component, focusing on displacement of the central axis to extrapolate diffusion parameters (Bearon and Grünbaum 2008).

In fact, helical movement might have important and unexpected consequences for dispersal and concentration of organisms. For example, it has been shown that vertical movement of micro-organisms in shears and turbulence can lead to concentrations and downward fluxes (Kessler 1985a, 1985b; Mitchell et al. 1990; Visser and Jonsson 2000). However, the fact that much of a helically swimming organism's movement may occur on a plane perpendicular to that of the axis might considerably alter expected responses to shears and stresses. Furthermore, variation in the oscillatory component of movement likely reflects physiological or genetic differences between populations, information that has significant ecological value.

An example of an ecological phenomenon influenced by individual-level microorganism movements is Harmful Algal Blooms (HABs) (Kessler 1985a; Franks 1997; Bearon and Grünbaum 2008). HABs are often characterized by a rapid and highly localized accumulation of toxic cells at high population densities and quantitative swimming characteristics likely influence their occurrence. Several studies have suggested that motility contributes to HABs by concentrating algal cells with ambient flows (Kessler 1985a; Franks 1997; Bearon et al. 2004) or enabling cells to locate favorable environments that enhance growth rates (Watanabe et al. 1988; Liu et al. 2001; Bearon et al. 2004). However, the contribution of swimming behavior is perhaps among the least understood mechanism that influences the formation of HABs (Bearon et al. 2006). For example, the HAB causing algae Heterosigma akashiwo, is a biflaggelate radiophyte (Smayda et al. 1998) that typically swims with a vertically oriented helical trajectory. While variations in mean vertical velocities and swimming behaviors from geographically distinct strains have been identified from two-dimensional videography data and the potential impacts on larger temporal, spatial and population-level scales have been considered (Bearon et al. 2004), the consequences of helicity of the movement and its ecological consequences have yet to be examined closely.

The overarching goal in this paper is to characterize helical microorganism movement from two-dimensional projections of swimming data. Though threedimensional swimming data of microorganisms exist (e.g., Sheng et al. 2007; Polin et al. 2009; Menden-Deuer and Grünbaum 2006), the methods used to collect them are substantially more difficult and costly than conventional two-dimensional micro-videography. Tractable and statistically robust methods of translating two-dimensional observations into estimates of three-dimensional helical swimming data available for ecological analysis.

To characterize helical trajectories, we developed a three-dimensional model that captures the main features of helical microorganism movement: oscillation around a net advective axis, with considerable stochasticity in both movement components. We apply the estimation methods to data obtained from *Heterosigma* strains sampled from a wide geographical range and consider potential interpretations of the results and applications of the model.

2 Methods and Materials

2.1 Three-Dimensional Helical Movement Model

We developed a stochastic, autocorrelated, continuous-time model of three-dimensional helical movement consisting of a slowly varying advective component \mathbf{X}^a and an oscillating component \mathbf{X}^o occurring on a two-dimensional plane (x', y') perpendicular to \mathbf{X}^a .

Each of these processes was modeled separately using an autocorrelated stochastic model of velocity adapted from the continuous stochastic movement models of Dunn and Brown (1987) and Alt (1988), which were themselves specific applications of an Ornstein–Uhlenbeck process. The advective velocity is described by the stochastic differential equation

$$d\mathbf{V}^{a} = A_{a}\left(\boldsymbol{\mu} - \mathbf{V}^{a}\right)dt + \sigma_{a}\,d\mathbf{W}_{t}^{3} \tag{1}$$

where μ is a mean drift (here, a vector with only a *z*-component of magnitude μ_a), A_a is a matrix related to the time scale of the autocorrelation in the velocities process, σ_a is a scalar parameter representing the strength of the stochasticity, and $d\mathbf{W}_t^3$ is the derivative of a three-dimensional Wiener process, i.e., a vector whose elements are drawn from an multivariate normal distribution with correlation zero. We assumed that the three velocities can be modeled independently, and that A_a and σ_a are isotropic in all three dimensions. Under these assumptions, the autocorrelation matrix can be expressed $A_o = \frac{1}{\tau_a}I$, where *I* is the 3 × 3 identity matrix and τ_a is a *characteristic time scale* (or *relaxation time*), which corresponds to the time interval at which the autocorrelation function of the velocity falls by a factor of e^{-1} . The advective velocity can therefore be separated into three independent orthogonal components

$$dV_x^a = -\frac{1}{\tau_a} V_x^a dt + \sigma_a dW_t$$

$$dV_y^a = -\frac{1}{\tau_a} V_y^a dt + \sigma_a dW_t$$

$$dV_z^a = \frac{1}{\tau_a} (\mu_a - V_z^a) dt + \sigma_a dW_t$$
(2)

This model generates a relatively versatile family of smooth three-dimensional trajectories with potential for a net bias in one direction (Figs. 1a, 1b and 1c). The equivalence of this model to a discrete first-order auto-regressive (AR(1)) time-series process is illustrated by discretizing the differential equations of the form in (2) by setting dt = 1 and rearranging terms, yielding

$$V_t = \mu + \left(1 - \frac{1}{\tau}\right)(V_{t-1} - \mu) + \varepsilon_i \tag{3}$$

where ε_i is an independent random Gaussian variable with variance σ . This is the standard discrete AR(1) (e.g. Cryer and Chan 2008) with mean μ and first order autocorrelation parameter $\gamma = 1 - \frac{1}{\tau}$. It should be noted that this discretization is only valid for values of $\tau \gg 1$.

The oscillatory component $\mathbf{V}^{o}(t)$ takes place in plane $\{X', Y'\}$ orthogonal to the orientation of the advective component at time *t*. In that plane, it is described by a two-dimensional autocorrelated velocity process with rotation but no advection:

$$d\mathbf{V}^{o} = A(\mathbf{V}^{o}) dt + \sigma_{o} d\mathbf{W}_{t}^{2}$$

$$\tag{4}$$

Here, $A = \begin{vmatrix} 1/\tau_o & \omega_o \\ \omega_o & 1/\tau_o \end{vmatrix}$ is a matrix in which $\tau_o > 1$ is the characteristic time scale of autocorrelation as in the advective case, the off-diagonal coefficient ω_o corresponds to the average angular velocity of rotation, σ_o is the strength of stochasticity and $d\mathbf{W}_t^2$ is a two-dimensional, uncorrelated Wiener process. A decomposition into x' and y' components yields:

$$dV_{x'}^{o} = -\left(\frac{1}{\tau_{o}}\right)V_{x'}dt + \omega_{o}V_{y'}dt + \sigma_{o}dW_{t}$$

$$dV_{y'}^{o} = -\left(\frac{1}{\tau_{o}}\right)V_{y'}dt + \omega_{o}V_{x'}dt + \sigma_{o}dW_{t}$$
(5)

Again, three parameters describe the oscillatory component of movement: τ_o , ω_o , and σ_o (Figs. 1d, 1e and 1f). Greater values of τ_o lead to more autocorrelated movement and higher values of ω_o yield more rotations per unit time (compare Figs. 1f to 1e).

The two velocity components are integrated to obtain their respective position vectors

$$\mathbf{X}^{a}(t) = \int_{t'=0}^{t} \mathbf{V}^{a}(t') dt'$$
(6)

$$\mathbf{X}^{o}(t) = \int_{t'=0}^{t} \mathbf{V}^{o}(t') \, dt'$$
(7)



Fig. 1 Realizations of trajectories for (**A**–**C**) advective $\mathbf{X}^{a}(t)$, (**D**–**F**) oscillatory $\mathbf{X}^{o}(t)$, and (**G**–**I**) combined helical movements. For all three-dimensional $\mathbf{X}^{a}(t)$ simulations $\mu_{a} = 5$. Other parameters are (**A**) $\tau_{a} = 10$ and $\sigma_{a} = 2$; (**B**) $\tau_{a} = 1.4$ and $\sigma_{a} = 2$; (**C**) $\tau_{a} = 1.4$ and $\sigma_{a} = 5$. Two-dimensional trajectories for $\mathbf{X}^{o}(t)$ have parameter values; (**D**) $\omega_{o} = 2$, $\tau_{o} = 100$, $\sigma_{o} = 0.05$; (**E**) $\omega_{o} = 2$, $\tau_{o} = 2$, $\sigma_{o} = 0.1$; (**F**) $\omega_{o} = 4$, $\tau_{o} = 100$, $\sigma_{o} = 0.1$. The three oscillatory components (**D**), (**E**), and (**F**) are combined with the smoothest advective axis (**A**) to create the helical movements **X** in (**G**), (**H**), and (**I**), respectively

To combine the two trajectories into a helix, the oscillatory component is rotated at each moment *t* along a plane perpendicular to $\mathbf{X}^{a}(t)$. This transformation consists of a rotation around the *x*-axis ($R_{x}(t)$) and one around the *y*-axis ($R_{y}(t)$) determined by the orientation of $\mathbf{V}^{a}(t)$. The rotation matrices are given by



$$R_{x}(t) = \begin{cases} 1 & 0 & 0\\ 0 & \cos(\theta_{1}(t)) & \sin(\theta_{1}(t))\\ 0 & -\sin(\theta_{1}(t)) & \cos(\theta_{1}(t)) \end{cases}$$
(8)

and

$$R_{y}(t) = \begin{cases} 1 & 0 & 0\\ 0 & \cos(\theta_{2}(t)) & \sin(\theta_{2}(t))\\ 0 & -\sin(\theta_{2}(t)) & \cos(\theta_{2}(t)) \end{cases}$$
(9)

where $\theta_1(t) = \cos^{-1}(V_z^a(t)/\sqrt{V_z^a(t)^2 + V_y^a(t)^2})$ and $\theta_2(t) = \cos^{-1}(V_z^a(t)/\sqrt{V_z^a(t)^2 + V_x^a(t)^2})$ (Fig. 2). The final expression for the helical model is

$$\mathbf{X}(t) = \mathbf{X}^{a}(t) + R(t) \cdot \mathbf{X}^{o}(t)$$
(10)

where $R(t) = R_x(t) \cdot R_y(t)$. We refer to the final trajectory as a *correlated velocity helical movement*, or CVHM (Figs. 1g, 1h and 1i).

2.2 Estimating \mathbf{V}^a and \mathbf{V}^o from data

Consider a regular subsampling \mathbf{X}_i of the continuous three-dimensional CVHM $\mathbf{X}(t)$, where $i = \{1, 2, ..., n\}$ and the sampling occurs at a fixed interval Δt . Because the model is defined in terms of the velocity dynamics, we perform the estimations on the estimated measured velocity vector $\mathbf{V}_i = (\mathbf{X}_{i+1} - \mathbf{X}_i)/\Delta t$.

To estimate the underlying parameters of the process, V must first be decomposed into estimates of the advective random axis \widehat{V}^a and the oscillatory component \widehat{V}^o .

The advective axis is estimated by averaging V over a period of oscillation:

$$\widehat{\mathbf{V}}_{i}^{a} = \frac{1}{\widehat{P}} \sum_{j=i-\widehat{P}/2}^{i+\widehat{P}/2} \mathbf{V}_{j}$$
(11)

The period \widehat{P} can be estimated in several ways. A simple estimate is given by the first maximum of the auto-correlation function (acf) of any of the components of the velocity data (V_x, V_y, V_z) . A more robust estimate uses the *complex velocity auto-correlation function* (CVAF) sensu Alt (1988). For a complex vector $\widetilde{\mathbf{V}}$ with real and imaginary elements V_x and V_z respectively, the CVAF is defined as

$$\mathcal{G}_V(t) = \left\langle \mathbf{\tilde{V}}(t'+t) \cdot \mathbf{\tilde{V}}^*(t') \right\rangle$$
(12)



Fig. 3 Example of estimating parameters using the complex velocity auto-correlation velocity (CVAF). In this example, an oscillating continuous random walk **X** was generated with parameters $\omega_o = 2$, $\tau_o = 5$, $\sigma_o = 1$, $T_{\text{max}} = 300$, dt = 0.1. Panel (**A**) shows the first 30 time units of the walk. The empirical CVAF of the velocity **V** is illustrated in (**B**) in the complex plane. Panel (**C**) illustrates the real component of the CVAF (*black line*), the *grey line* is the fitted curve, and *vertical lines* represent the periodicity estimate. In this example, the estimated parameters were $\omega_o = 2.05$, $\tau_o = 7.87$, $\sigma_o = 1.08$

where $\widetilde{\mathbf{V}}^*(t')$ is the complex conjugate of $\widetilde{\mathbf{V}}$, the product is a complex product, and the angle bracket notation refers to the expectation over all t'. In the case of periodic or oscillatory movement, the real and imaginary components of the CVAF have a well defined and easily identified periodicity (Fig. 3), and is therefore the method we used in this analysis.

The oscillatory component was estimated in two steps. The "raw" oscillatory component $\widetilde{\mathbf{X}}^o$ was obtained by subtracting the estimated advective component from the data (Figs. 4b and 4c)

$$\widetilde{\mathbf{X}}^o = \mathbf{X} - \widehat{\mathbf{X}}^a \tag{13}$$

The final oscillatory component was estimated by rotating $\widetilde{\mathbf{X}}^o$ back by the spherical angle determined by the angle between $\widehat{\mathbf{V}}^a$ and the *z*-axis (Figs. 4d and 4e):

$$\widetilde{\mathbf{X}}^o = \widehat{R}^{-1} \widetilde{\mathbf{X}}^o \tag{14}$$

where \widehat{R} is the spherical angle defined as in (8) and (9), with \mathbf{V}^a replaced by $\widehat{\mathbf{V}}^a$ (Fig. 4).



Fig. 4 Estimation of \mathbf{X}^a and \mathbf{X}^o : (A) Simulated CVHM X (from Fig. 1G); the *dark grey line* represents the estimate $\mathbf{\hat{X}}^a$ obtained by averaging X over the estimated period of oscillations; the *light grey line in the center* of the helix is the true \mathbf{X}^a (from Fig. 1A). (B) and (C) are the raw oscillatory component, obtained by $\mathbf{\tilde{X}}^o = \mathbf{X} - \mathbf{X}^a$. (D) and (E) are the final estimate $\mathbf{\hat{X}}^a$, obtained by rotating $\mathbf{\tilde{X}}^o$ by the angular deviation of \mathbf{V}^a from the *z*-axis. In (C) and (E), the oscillatory components are presented on the *x*-*z* and *x'*-*z'* plane, respectively, to illustrate the flattening of the estimate after rotation

2.3 Estimating parameters for V^a and V^o

Estimates for most movement parameters can be obtained directly, with no fitting or numerical estimation, from $\widehat{\mathbf{V}}^o$ and $\widehat{\mathbf{V}}^a$. For the advective axis:

$$\hat{\mu}_a = \overline{V}_z^a \tag{15}$$

$$\hat{\tau}_a = -\Delta t / \log(\hat{\gamma}) \tag{16}$$

where $\hat{\gamma}$ is the mean of the estimated first-order autocorrelation coefficients of the velocities for all three velocity components, given by $\hat{\gamma}_k = \sum_{i=1}^{n-1} (V_{k,i+1} - \overline{V}_k) \times (V_{k,i} - \overline{V}_k)/[(n-1)s_x^2]$ where *k* represents the dimension *x*, *y*, or *z*. The relationship in (16) is obtained by noting that the autocorrelation of the velocities decays exponentially as $\exp(-t/\tau)$, while the discrete AR(1) autocorrelation decays as γ^t . The final advective parameter, σ_a , can be obtained by estimating the variance of the discrete velocity differences in the data. Thus, if we define the velocity difference vector in dimension *k* such that the *i*th term is given by $\Delta \mathbf{V}_{k,i} = (\mathbf{V}_{k,i+1}^a - \mathbf{V}_{k,i}^a)$, then:

$$\hat{\sigma}_{a}^{2} = \frac{1}{3} \sum_{k=x,y,z} \operatorname{Var}\left[\frac{(\Delta \mathbf{V}_{k} - (\frac{1}{\hat{\tau}})(\hat{\mu}_{k} - \mathbf{V}_{k}^{a}))}{\Delta t^{1/2}}\right]$$
(17)

The estimate is the mean of the estimates for each of the three axes, each of which is obtained directly by discretizing and rearranging the equations in (2). The 1/2 exponent on the Δt reflects the appropriate scaling of the white noise process, in which the expected squared displacement increases linearly with time.

The estimates for the angular velocity of the oscillatory component $\widehat{\mathbf{V}}^o$ are

$$\hat{\omega}_o = 2\pi/\widehat{P} \tag{18}$$

In order to estimate τ_o , we take advantage of a useful result by Alt (1988): A twodimensional movement process (4), has an expected CVAF given exactly by

$$\mathcal{G}_{V}(t) = \frac{1}{2}\tau\sigma^{2}\exp\left[\left(\frac{1}{\tau} + i\omega\right)t\right]$$
(19)

which is an exponentially decaying, oscillating function with frequency ω and decay rate $1/\tau$. We calculated the empirical CVAF: $\widehat{\mathcal{G}}_V(dt) = E[\widehat{\mathbf{V}}_t^o \cdot \widehat{\mathbf{V}}_{t+dt}^{o*}]$ and fitted the empirical curve to (19) using least squared fitting with an exponential weighting to obtain $\widehat{\tau}^o$. Finally, we estimated σ_o using expression (17)

$$\hat{\sigma}_o = \sqrt{\frac{|\mathbf{V}^o|^2}{(2\hat{\tau}_o)}} \tag{20}$$

The use of the CVAF to estimate the parameters of this process is illustrated in Fig. 3.

2.4 Converting Two-Dimensional Data to Three-Dimensions

We consider now the problem of estimating parameters for the three-dimensional CVHM model from two-dimensional data. This three-dimensionalization of the two-dimensional track is performed, as before, by analyzing the velocities V rather than the positions X. Given V_x and V_z , our specific goal is to derive a reasonable estimate for the unobserved velocity component V_y .

We note first that the (unknown) magnitude of the total velocity is given by

$$|\mathbf{V}| = \sqrt{|V_x|^2 + |V_y|^2 + |V_z|^2}$$
(21)

At any point where the organism is moving perpendicularly to the camera, V_y is equal to zero and the total velocity is given by $\sqrt{V_x^2 + V_z^2}$. If we assume that the movement of the microorganism has zero net drift in the y-direction, this leads to movement that is perpendicular to the camera exactly twice at each rotation of the helix. Thus, the total magnitude of velocity can be estimated as a spline of the peaks of the $\sqrt{V_x^2 + V_z^2}$ curve (Fig. 5b). At each peak, V_y is exactly zero, but its sign must change for the rotation around the axis to be completed (Fig. 5c). The final inferred y component of velocity can be integrated to obtain the final three-dimensionalized trajectory (Fig. 5d).

Several important movement features cannot be captured with this method. Notably, if the organism spends any significant time moving away from or toward the



Z

Deringer

Fig. 5 Illustration of the three-dimensionalization procedure of estimating a third velocity component from two-dimensional data. (A) Plot of raw movement data in the x-z plane for a single Heterosigna cell. (B) Squared sum velocities of the x and z components (dotted and grev lines). The points indicate the maxima at each rotation, and the dark line indicates the estimated magnitude of total velocity [V]. (C) Final estimates for V_v (*thick line*), with the *points* indicating locations where the sign is flipped. (D) Final three-dimensionalized trajectory

camera with a net positive or negative y component of movement, this behavior cannot be accounted for or estimated from x-z data without specific and usually unjustified constraints on swimming characteristics. Furthermore, the parameters cannot be meaningfully estimated if the trajectory is not recognizably helical and does not contain a sufficient number of data points to estimate and subtract a periodic component. Despite these drawbacks, the procedure produces three-dimensional helical trajectories that look qualitatively realistic.

It should be further noted that it is impossible to infer the handedness of a helix from two-dimensional data alone. The handedness obtained in our three-dimensional visualizations is arbitrarily chosen based on the sign of the first "flip" of the estimated *y* component. However, for many helically swimming organisms the helicity is genetically predetermined and experimentally known (Pedley and Kessler 1992), including for *Heterosigma*, which swims in a counterclockwise direction relative to its long axis (Han et al. 2002).

2.5 Simulation Study

We explored the precision and accuracy of the estimation procedure outlined in Sect. 2.3 with a simulation experiment. We simulated 200 CVHM, each of 30 s duration with a sampling rate of 1/30 s. The values chosen for the six parameters reflected the distributions of those obtained for the actual *Heterosigma* data (Sect. 3.2, Table 3). Values for μ_a , ω_o , σ_a , σ_o , $\tau_o^{1/2}$ and $\tau_a^{1/2}$ were randomly drawn from normal distributions with means and standard errors summarized in Table 2, with negative values truncated away. The square of the normal distribution was chosen for τ_o and τ_a to reflect the skew of the data estimated parameters. Similarity between simulation parameters and their estimates were assessed using least squared fits (Fig. 6).

2.6 Data

We estimated the CVHM parameters for individual tracks of *Heterosigma akashiwo* from five strains originating in diverse geographic locations and different years where *Heterosigma* HABs have been reported (Table 3). Two strains were collected in the Atlantic Ocean (in Long Island Sound in 1952 and in Narragansett Bay, Rhode Island in 1991), two in Puget Sound in Washington State (in 2002 and 2007), and one in the Sea of Japan (1984). For each strain, triplicate cultures were grown in an artificial seawater medium (McIntosh and Cattolico 1978).

Cell swimming was observed in a cylindrical acrylic tank (1.5L 31.8 cm height and 8 cm diameter) filled with autoclaved seawater. A weak linear salinity gradient (28 to 30 ppt) was created to suppress ambient fluid motion. *Heterosigma* cells were visualized under dark-field illumination from an infra-red light source in order to minimize phototactic behavioral responses and temperature was maintained at 20°C. Algal cells from exponentially growing cultures were slowly introduced into the bottom of the tank and allowed to swim freely to the surface. Cell motion was recorded with a CCD video camera (COHU 4815-3000, Nikon Nikkor 60 mm f/2.8D lens) at 30 Hz. 180 seconds of video was captured for each trial and three trials were conducted for each of the five strains. The pixel locations of free-swimming cells

Symbol	Definition	Meaning		
Continuous mode	el			
$\mathbf{X}(t)$		three-dimensional continuous helical movement model		
$\mathbf{V}(t)$	$d\mathbf{X}/dt$	three-dimensional continuous velocity vector		
$\mathbf{V}^{a}(t)$	see (1)	velocity vector of advective axis		
$\mathbf{V}^{o}(t)$	see (4)	velocity vector of oscillation perpendicular to \mathbf{V}^a		
V_x, V_y, V_z		x, y and z components of $\mathbf{V}(t)$		
V_x^a, V_y^a, V_z^a		x, y and z components of $\mathbf{V}^{a}(t)$		
V_x^o, V_y^o, V_z^o		x, y and z components of $\mathbf{V}^{o}(t)$		
Discrete data				
Δt		time interval between steps		
\mathbf{X}_i		discrete movement data, $t = (1, 2,, n)$		
\mathbf{V}_i	$(\mathbf{X}_{t+1} - \mathbf{X}_t) / \Delta t$	velocity data vector		
$V_{t,x}, V_{t,y}, V_{t,z}$		x, y and z components of data V		
Parameters				
μ_a		mean vertical velocity of \mathbf{V}^a		
$ au_a$		characteristic time scale of \mathbf{V}^a		
σ_a		magnitude of stochasticity for \mathbf{V}^a		
ω_o		characteristic angular velocity of \mathbf{V}^o		
P_o	$1/\omega_o$	characteristic period of rotation of \mathbf{V}^{o}		
$ au_O$		characteristic time scale of \mathbf{V}^o		
σ_o		magnitude of stochasticity for \mathbf{V}^{o}		
\overline{V}_t		mean tangential velocity		
Θ	$\sin^{-1}(\mu_a/\overline{V}_t)$	mean tangential angle		

Table 2 Parameters used togenerate 200 realizations of the	Parameter	Mean	s.d.
CVHM for the simulation experiment. Random values	μ_a	0.12	0.03
were drawn from normal	σ_a	0.015	0.004
standard deviations tabulated	$\tau_a^{1/2}$	3	1
here. The source distributions	ω_a	3.8	0.5
for $\mu_a, \omega_o, \sigma_a$, and σ_o were	σ_o	0.04	0.02
truncated to exclude negative values	$\tau_{o}^{1/2}$	3	1

were determined using an open-source processing software package (Avidemux2) that enabled the removal of background noise and stationary objects. From these cell locations, 2D swimming paths were generated using an in-house MATLAB-based motion-analysis program (Tracker3D).

A total of 749 trajectories containing at least 510 frames (17 s) were recorded. Of these, 445 were excluded from the analysis because the parameter estimation failed,



(A) Strains					
Location	Narraganset	Long Island	Puget Sound,	Puget Sound,	Onagawa Bay,
	Bay, RI	Sound, NY	WA	WA	Sea of Japan
Year	1991	1952	2002	2007	1984
Ν	55	131	31	54	33
(B) Parameters					
$\mu_a (\mathrm{mms^{-1}})$	0.069	0.036	0.047	0.049	0.043
	(0.022–0.175)	(0.005-0.129)	(0.007 - 0.088)	(0.021-0.118)	(0.013-0.094)
$ au_a$	4.46	3.64	3.68	6.45	6.30
	(1.54–13.84)	(1.21–14.75)	(1.58–13.48)	(2.22–21.86)	(1.98–18.73)
$\sigma_a \; (\mathrm{mm}\mathrm{s}^{-1})$	0.008	0.006	0.005	0.004	0.002
	(0.003-0.034)	(0.002-0.011)	(0.002-0.009)	(0.002-0.009)	(0-0.007)
$\omega_o (s^{-1})$	2.227	1.878	1.923	1.558	1.904
	(1.344–3.624)	(1.041-3.351)	(1.095–2.543)	(0.97–3.08)	(1.06–3.36)
$ au_o$	4.66	3.68	2.66	5.81	3.51
	(0.35–21.38)	(0.15–18.39)	(0.07–19.08)	(1.47–47.81)	(1.01–37.57)
$\sigma_o (s^{-1})$	0.037	0.043	0.024	0.021	0.009
	(0.007 - 0.95)	(0.005 - 0.48)	(0.006-0.618)	(0.004–0.312)	(0.003-0.029)
$\overline{V}_t \; (\text{mm s}^{-1})$	0.141	0.114	0.055	0.097	0.101
	(0.085 - 0.383)	(0.05–0.155)	(0.022-0.103)	(0.053-0.12)	(0.061-0.153)
Θ_t (rad)	0.52	0.376	0.872	0.575	0.618
	(0.108–1.051)	(0.036–0.805)	(0.379–1.212)	(0.124–1.173)	(0.098–0.92)

Table 3 (A) Locations, collection years and sample sizes for five strains of videotracked *Heterosigma akashiwo* that were analyzed using the CVHM, and (B) median and 95% range (i.e., 2.5 to 97.5 quantiles) for all parameter estimates for each of the five strains

reflecting the fact that the tracks were not detectably helical according to our algorithm. Thus, our final sample size was 304 cell trajectories (Table 3).

In addition to the CVHM estimates, we report the mean tangential velocities (\overline{V}_t) , obtained directly from the measured displacements according to

$$\overline{V}_t = \frac{1}{n-1} \sum_{i=2}^n |\mathbf{X}_i - \mathbf{X}_{i-1}| dt$$
(22)

and the mean tangential angles estimated as $\theta_t = \sin^{-1}(\mu_a/\overline{V}_t)$. A tangential angle of 0 corresponds to circular movement in a plane, whereas an angle of $\pi/2$ represents vertical linear movement.

All parameter estimates were compared between strains using two-sided nonparametric multiple rank comparison tests, which are robust against non-normal distributions and unbalanced data (Munzel and Hothorn 2001). The tests were implemented using the "npmc" package in the R open source programming language (Helms and Munzel 2008; R Development Core Team 2009).

All of the symbols representing the CVHM movement process, the data components of measured trajectories, and the parameters estimated are summarized and defined in Table 1.

3 Results

3.1 Simulation Study

Equations (15)–(20) provided reasonable estimates for most of the six parameters in the simulation study (Fig. 6). When regressed against the seeded values for the parameters, estimated slopes for $\hat{\mu}_a$, $\hat{\omega}_o$, and $\hat{\sigma}_o$ had slopes statistically indistinguishable from 1 and high r^2 values. The fit for $\hat{\tau}_o$ had a significant positive slope, although some of the actual estimated values tended to be extremely high (on the order of 10³). The estimates for $\hat{\sigma}_a$ were less accurate, with a low r^2 of 0.13, while the estimate for τ_a was completely uninformative ($r^2 = 0$, regression slope = 0).

3.2 Data Analysis

We estimated all six CVHM parameters and the tangential velocities and angles for each of 304 individual *Heterosigma* cells from the five strains and report median and 95% ranges (i.e. 0.025 and 0.975 quantiles of all estimates) in Table 3. The vertical speed μ_a ranged from 0.003 to 0.189 mm/s, with an overall median value of 0.044 mm/s. The highest values were recorded for the Rhode Island strain (0.069 mm/s median), well above the medians of the other four strains (p = 0.001, Table 3). Multiple rank comparisons indicate that the differences between the three Pacific Ocean strains were not significant (p > 0.1), while the Long Island strain (median 0.039 mm/s) was slightly slower than the slowest Pacific strain (Sea of Japan— 0.041, p = 0.01) and substantially slower than the remaining strains (p < 0.001).

The angular velocity ω_o ranged from 0 to 5.52 rotations s⁻¹, with a median value for all data of 1.92 and 95% of the values ranging between 1.01 and 3.23 s⁻¹. The angular velocities of the Rhode Island strain were significantly higher than all the other strains, with a median angular velocity of 2.23 s⁻¹, while the Puget Sound 1 strain was significantly lower at 1.558 s⁻¹. Other differences were not significant. There was a significant positive relationship between the estimates for ω_o and the estimates for μ_a (ANOVA p < 0.001).

The characteristic time scales for advection (τ_a) and oscillation (τ_o) had similar median values (4.33 and 3.77 s, respectively) and 95% ranges (1.28–16.7 and 0.14–22.1 s) across strains. There were no significant differences between strains in the oscillatory characteristic time scale τ_o (all p > 0.1). In the advective component, the time scales were longest in the Puget Sound 2 and Sea of Japan strain (medians around 6.4 s) and shortest for the Long Island and Puget Sound 1 strain (medians near 3.6 s), but we found no significant differences between the intermediate Rhode Island strain and any of the other strains. A regression of τ_a against τ_o yielded a significant positive relationship (doubly log transformed ANOVA p < 0.008, $r^2 = 0.19$).

The stochastic strength parameter for advection (σ_a) showed the greatest separation between strains of all parameters, with the highest values displayed by the strains from Rhode Island (median 0.008 mm s⁻¹) and Long Island (0.006 mm s⁻¹). The lowest was the Sea of Japan (0.002 mm s⁻¹), and the only two strains that were not significantly different were the two virtually identical Puget Sound strains (p = 0.90). The value of the oscillatory stochasticity (σ_o) was four to six times higher than the advective stochasticity, with a very similar ordering. Again, Long Island and Rhode Island strains displayed the highest values (median 0.037 and 0.043 mm s⁻¹), the Sea of Japan strain displayed the lowest (0.009 mm s⁻¹), while the strains from Puget Sound (around 0.021 mm s⁻¹) had intermediate values that were not statistically different than any of the other strains. A linear regression of σ_a against σ_o yielded a highly significant regression (log–log ANOVA p < 0.001, slope = 0.26 $r^2 = 0.56$).

The tangential velocities (\overline{V}_t) were highest for the Rhode Island strain, significantly lower for the Long Island strain (p < 0.001), and even lower for the Sea of Japan strain (p < 0.001). The two Puget Sound strains were not significantly different from each other and were slightly lower than the Long Island strain. The mean tangential angles θ_t varied considerably between 0.376 and 0.876 radians (21.5 and 50.0°). The angles were lowest in Long Island Sound and highest in the Sea of Japan (multiple paired comparison p < 0.001), and intermediate and indistinguishable among the Puget Sound and Rhode Island strains (p > 0.3).

4 Discussion

4.1 The CVHM Model

The CVHM is a versatile model for parameterizing and simulating a wide range of helical movements. An important feature of the CVHM is its definition in continuous space and time. Continuous models have several important advantages over discrete models of movement, as they accurately capture the continuous nature of real biological movements and allow for parameterizations of movement that are independent of sampling intervals (Johnson et al. 2008; Gurarie et al. 2009b). Another important feature is the explicit description of the magnitude and autocorrelation of the stochastic components of movement. These parameters are typically ignored or averaged away when characterizing helical movements, but may have intrinsic biological importance or diagnostic value. Stochasticity may reflect random fluctuations in the liquid medium or variations in the propulsive and orientation forces of the organisms. The strong relationship between the advective and oscillatory stochasticity may have a phenotypic source, related to size, configuration or propulsion of the individuals.

The structure of the model lends itself to a straightforward and efficient estimation of all parameters, all but one of which are obtained with direct computation from the data without relying on numerical fitting algorithms. The straightforwardness of the estimation despite the relative complexity of the movement model is due to the underlying Gaussian assumption of infinitesimal random fluctuations in velocity which (as the defining assumption of the Ornstein–Uhlenbeck process) is at the core of the CVHM. An additional vital tool is the analysis of autocorrelation structure, whether in the simple exponential model of the advective component or the periodic complex velocity autocorrelation of the oscillating component. In general, the basic elements of the model—a characteristic velocity, a scale of randomness, and a temporal scale of autocorrelation are very general features of continuous movement processes. We believe that the basic continuous velocity processes that we adapted to helical movement can be a useful and highly general paradigm for analyzing a wide range of movement data.

Our simulation study provided satisfactory support for the accuracy of the parameter estimates. However, the process of estimating and separating the oscillatory and advective components from a three-dimentional data had the effect of diluting the underlying stochasticity (σ_a) and autocorrelation time scale (τ_a) of the advective component, making those the most difficult to estimate accurately. Nonetheless, the final estimates are immediately practicable for the rapid and realistic generation of simulated helical movement tracks at arbitrary scales. For the purposes of differentiating strains, even the more poorly estimated parameters showed significant differences between the strains, suggesting that the method is robust.

In extracting three-dimensional movement statistics from two-dimensional data using the CVHM method, several conditions were necessary for obtaining robust estimates: sufficient length of the data (around 500 localizations in each individual trajectory), a minimum number of rotations in the helix (around 5), and an identifiably vertical long axis. Our three-dimensionalization technique was not able to capture any persistent movement towards or away from the video camera; this is not possible without obtaining data in the third dimension. Nonetheless, given the relative ease of collecting large samples of two-dimensional data from video analysis, and the largely vertical swimming movements of many microorganisms in still water, the approach may be an informative and efficient method for inferring the transient velocities in an unobserved third dimension. Our case study with *Heterosigma* illustrated the feasibility of such analysis, and the diagnostic insights that can result.

4.2 Differences Among Strains

The distinct types of helices observed in our *Heterosigma* strains likely correspond to physiological differences which are otherwise difficult to detect. Although phenotypic differences have been identified in *Heterosigma* strains from distinct geographic regions (Hosaka 1992; Smayda et al. 1998; Han et al. 2002; Bearon et al. 2004), it has to date not been possible to genetically identify strains (Connell 2000; Ki and Han 2007). Estimates of helical movement parameters derived from the CVHM provide another approach for identifying distinctive characteristics of populations. Of the known phenotypic differences among strains, movement characteristics (CVHM parameters and other velocity statistics) appear to have the most potential to be implemented in near real time and in field or remote sensing applications. These differences could provide a basis for a scheme for inferring strain identities from phenotypes.

The Rhode Island strain displayed the highest tangential and vertical velocities, the most rapid rotations and the greatest stochasticity of all the strains. It was notably distinct from the geographically proximate Long Island Sound strain, which displayed the slowest vertical velocities and intermediate angular velocities. The tangential velocity was, however, rather high for the Long Island strain, with the low vertical velocity explained by the shallowest helical angle. It is possible these differences reflect local adaptation to Narragansset Bay and Long Island Sound, or that they represent subsamples from a genetically diverse and spatially structured population. These two strains were collected 39 years apart (1952 and 1991, Table 3) and it is possible that a regional population has evolved or been displaced under local oceanographic or biochemical selection as conditions have changed in the Northwest Atlantic. It is also possible that swimming has changed in one or both strains under artificial selection in laboratory culture. To test these hypotheses and the sensitivity of the method more rigorously, the analysis should be applied to a more systematically designed sampling of strains.

The observed characteristic time scales indicated that both rotational and advective components of movement are autocorrelated over seconds to tens of seconds. From an ecological perspective, how these movement time scales compare to other time scales in the marine environment such as mean intervals between encounters with biotic or abiotic particles and Kolmogorov time scales of turbulence, is likely to have strong functional consequences.

The strong relationships between the estimates of advective and angular velocity (μ_a and ω_o), between the advective and oscillatory stochasticity (σ_a and σ_o), and between the characteristic time scales (τ_a and τ_o), despite the formal mathematical independence of the two kinds of movement, suggest that all of these parameters have a basis in the physiology of the organisms. That is, greater velocities, higher magnitudes of stochasticity and time scales of autocorrelation are linked to individual variation in the morphology or metabolism of the individual cells. Identifying the mechanistic sources of these variation, however, requires both closer analysis of the morphology of the cells and a model that is more directly parameterized in terms of the forces and torques that the cell experiences.

4.3 Potential Consequences of Helical Movement

Several explanations have been proposed for the ubiquity of helical movement among microorganisms. Over a century ago, Jennings (1901) postulated that the helical trajectory allows an otherwise asymmetric organism to move along a nearly straight trajectory. Directed movement is clearly beneficial for such helically moving organisms such as spermatozoa in pursuit of ova, or motile algae striving to be near the water surface to increase light uptake as turbulence drives them deeper into the water column (Kessler 1985b). Despite the fact that that loose helices (i.e., with a tangential velocity angle fairly close to $\pi/2$) have more rapid directed movement, very few of the individual *Heterosigma* we analyzed had tangential angles greater than $\pi/4$. This apparently excessive tightness in the helicity, as well as the considerable variability among strains, suggests additional evolutionary mechanisms encouraging helical movement.

It is important to note that the CVHM as it is presented here is a limited and idealized model in several key ways. It is limited to approximating variation around a known advective axis (here, a vertical axis, but easily modified by rotation for an arbitrary axis), making it difficult to apply directly to movements in more complex environments. The model can, however, be used as a basis for simulation-based experiments in which organisms are released in simulated complex shear flows to explore the consequences for aggregation and dispersal, as well as for exploring potential benefits of helical motion. A second limitation of the CVHM is that it describes behaviorally homogeneous movement processes. Many microorganisms, including *Heterosigma*, engage in multiple swimming modes with sudden transitions (Bearon et al. 2004; Polin et al. 2009). However, the CVHM can be considered a building block of helical movement, within which discrete behavioral changes can be identified as had been done in other continuous animal movement analyses (Gurarie et al. 2009b). Finally, the CVHM model describes trajectories without explicitly suggesting explanatory mechanisms. However, the values of the CVHM parameters are clearly rooted in the biomechanics and physical constraints of individual microscopic movement. Relating the CVHM movement parameters to physiological structure and physical principles would provide insight into the mechanics of individual movement and to the possible selective pressures that may explain variation between populations of helically swimming microorganisms.

Acknowledgements *Heterosigma* strains were provided by R. Cattolico. T. Clay, K. Chan, and E. Tobin contributed helpful discussions and expertise during the data collection. The development of the methods in this paper benefited greatly from discussions with O. Ovaskainen, J. Pennanen, and C. Zheng. We are in particular indebted to an acute observation by R. Bearon and a text review by K. Laidre. E.G. was supported in part by the European Research Council (ERC Starting Grant no. 205905). D.G. and M.T.N. gratefully acknowledge support from the NOAA Washington Sea Grant Program (no. NAO4DAR170032 to DG). Comments from three anonymous reviewers greatly improved the manuscript.

References

- Alt, W. (1988). Modelling of motility in biological systems. In J. McKenna & R. Temam (Eds.), *ICIAM* '87: Proceedings of the first international conference on industrial and applied mathematics (pp. 15– 30). Philadelphia: SIAM.
- Bearon, R. N., Grünbaum, D., & Cattolico, R. A. (2006). Effects of salinity structure on swimming behavior and harmful algal bloom formation in Heterosigma akashiwo, a toxic raphidophyte. *Mar. Ecol. Prog. Ser.*, 306, 153–163.
- Bearon, R. N., Grünbaum, D., & Cattolico, R. A. (2004). Relating cell-level swimming behaviors to vertical population distributions in Heterosigma akashiwo (Raphidophyceae), a harmful alga. *Limnol. Oceanogr.*, 49(2), 607–613.
- Bearon, R. N., & Grünbaum, D. (2008). From individual behaviour to population models: a case study using swimming algae. J. Theor. Biol., 251(4), 679–697.
- Connell, L. (2000). Nuclear ITS region of the alga Heterosigma akashiwo (Chromophyta: Raphidophyceae) is identical in isolates from Atlantic and Pacific basins. *Mar. Biol.*, 136(6), 953–960.
- Crenshaw, H. (1993). Orientation by helical motion—I. Kinematics of the helical motion of organisms with up to six degrees of freedom. *Bull. Math. Biol.*, *55*(1), 197–212.
- Crenshaw, H. (1996). A new look at locomotion in microorganisms: rotating and translating. *Integr. Comp. Biol.*, *36*(6), 608–618.
- Crenshaw, H., Ciampaglio, C., & McHenry, M. (2000). Analysis of the three-dimensional trajectories of organisms: estimates of velocity, curvature and torsion from positional information. J. Exp. Biol., 203(6), 961–982.
- Cryer, J., & Chan, K. (2008). Time series analysis: with applications in R. Berlin: Springer.
- Dunn, G. A., & Brown, A. F. (1987). A unified approach to analysing cell motility. J. Cell Sci. Suppl., 8, 81–102.
- Franks, P. J. S. (1997). Models of harmful algal blooms. Limnol. Oceanogr., 42, 1273-1282.
- Gurarie, E., Anderson, J. J., & Zabel, R. W. (2009a). Incorporating population heterogeneity into analysis of animal dispersal and movement. *Ecology*, *90*(8), 2233–2242.
- Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009b). A novel method for identifying behavioural changes in animal movement data. *Ecol. Lett.*, *12*(5), 395–408.

Han, M., Kim, Y., & Cattolico, R. (2002). Heterosigma akashiwo (Raphidophyceae) resting cell formation in batch culture: strain identity versus physiological response. J. Phychol., 38(2), 304–317.

Helms, J., & Munzel, U. (2008). npmc: Nonparametric multiple comparisons (R package version 1.0-7).

- Hosaka, M. (1992). Growth characteristics of a strain of Heterosigma akashiwo (Hada) isolated from Tokyo Bay, Japan. Bull. Plankton Soc. Jpn., 39, 49–58.
- Jennings, H. (1901). On the significance of the spiral swimming of organisms. Am. Nat., 35(413), 369-378.
- Johnson, D. S., London, J. M., Lea, M.-A., & Durban, J. W. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89(5), 1208–1215.
- Kessler, J. O. (1985a). Co-operative and concentrative phenomena of swimming micro-organisms. *Contemp. Phys.*, 26(2), 147–166.
- Kessler, J. O. (1985b). Hydrodynamic focusing of motile algal cells. Nature, 313(5999), 218-220.
- Ki, J., & Han, M. (2007). Nuclear rDNA and chloroplast rbcL, rbcS and IGS sequence data, and their implications from the Japanese, Korean, and North American harmful algae, Heterosigma akashiwo (Raphidophyceae). *Environ. Res.*, 103(3), 299–304.
- Liu, G., Janowitz, G., & Kamykowski, D. (2001). A biophysical model of population dynamics of the autotrophic dinoflagellate Gymnodinium breve. *Mar. Ecol. Prog. Ser.*, 210, 101–124.
- Menden-Deuer, S., & Grünbaum, D. (2006). Individual foraging behaviors and population distributions of a planktonic predator aggregating to phytoplankton thin layers. *Limnol. Oceanogr.*, 51(1), 109–116.
- McIntosh, L., & Cattolico, R. (1978). Preservation of algal and higher plant ribosomal RNA integrity during extraction and electrophoretic quantitation. *Anal. Biochem.*, 91(2), 600–612.
- Mitchell, J. G., Okubo, A., & Fuhrman, J. A. (1990). Gyrotaxis as a new mechanism for generating spatial heterogeneity and migration in microplankton. *Limnol. Oceanogr.*, 35(1), 123–130.
- Munzel, U., & Hothorn, L. A. (2001). A unified approach to simultaneous rank test procedures in the unbalanced one-way layout. *Biom. J.*, 43(5), 553–569.
- Pedley, T. J., & Kessler, J. O. (1992). Hydrodynamic phenomena in suspensions of swimming microorganisms. Ann. Rev. Fluid Mech., 24, 313–358.
- Polin, M., Tuval, I., Drescher, K., Gollub, J., & Goldstein, R. (2009). Chlamydomonas swims with two "gears" in a eukaryotic version of run-and-tumble locomotion. *Science*, *325*(5939), 487–490.
- R Development Core Team (2009). R: A language and environment for statistical computing R foundation for statistical computing (ISBN 3-900051-07-0). Retrieved from http://www.R-project.org.
- Sheng, J., Malkiel, E., Katz, J., Adolf, J., & Belas, R. (2007). Digital holographic microscopy reveals preyinduced changes in swimming behavior of predatory dinoflagellates. *Proc. Natl. Acad. Sci.*, 104(44), 17512–17517.
- Smayda, T. J., Anderson, D. M., Cembella, A. D., & Hallegraeff, G. M. (1998). Ecophysiology and bloom dynamics of Heterosigma akashiwo (Raphidophyceae). In D. M. Anderson, A. D. Cembellas & G. M. Hallegraeff (Eds.), *Physiological ecology of harmful algae blooms* (Vol. 41, pp. 115–131). Berlin: Springer.
- Visser, A., & Jonsson, P. (2000). On the reorientation of non-spherical prey particles in a feeding current. J. Plankton Res., 22(4), 761.
- Watanabe, M., Kohata, K., & Kunugi, M. (1988). Phosphate accumulation and metabolism by Heterosigma akashiwo (Raphidophyceae) during diel vertical migration in a stratified microcosm. J. Phychol., 24(1), 22–28.