



Research

Cite this article: Mwaffo V, Anderson RP, Butail S, Porfiri M. 2015 A jump persistent turning walker to model zebrafish locomotion. *J. R. Soc. Interface* **12**: 20140884. <http://dx.doi.org/10.1098/rsif.2014.0884>

Received: 8 August 2014

Accepted: 17 October 2014

Subject Areas:

computational biology

Keywords:

zebrafish, behaviour, stochastic process, locomotion, jump process, modelling

Author for correspondence:

Maurizio Porfiri

e-mail: mporfiri@nyu.edu

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2014.0884> or via <http://rsif.royalsocietypublishing.org>.

A jump persistent turning walker to model zebrafish locomotion

Violet Mwaffo¹, Ross P. Anderson¹, Sachit Butail² and Maurizio Porfiri¹

¹Department of Mechanical and Aerospace Engineering, New York University Polytechnic School of Engineering, Brooklyn, NY, USA

²Indraprastha Institute of Information Technology Delhi (IIITD), New Delhi, India

Zebrafish are gaining momentum as a laboratory animal species for the investigation of several functional and dysfunctional biological processes. Mathematical models of zebrafish behaviour are expected to considerably aid in the design of hypothesis-driven studies by enabling preliminary *in silico* tests that can be used to infer possible experimental outcomes without the use of zebrafish. This study is motivated by observations of sudden, drastic changes in zebrafish locomotion in the form of large deviations in turn rate. We demonstrate that such deviations can be captured through a stochastic mean reverting jump diffusion model, a process that is commonly used in financial engineering to describe large changes in the price of an asset. The jump process-based model is validated on trajectory data of adult subjects swimming in a shallow circular tank obtained from an overhead camera. Through statistical comparison of the empirical distribution of the turn rate against theoretical predictions, we demonstrate the feasibility of describing zebrafish as a jump persistent turning walker. The critical role of the jump term is assessed through comparison with a simplified mean reversion diffusion model, which does not allow for describing the heavy-tailed distributions observed in the fish turn rate.

1. Introduction

Zebrafish (*Danio rerio*) are rapidly emerging as an experimental species for the investigation of functional and dysfunctional biological processes, owing to their sequenced genome, high reproduction rate, short intergeneration time, prominent shoaling tendency and elevated stocking density compared with laboratory mammals [1–4]. Furthermore, with the aim of generating high-throughput behavioural data [5–7], considerable research is being performed to integrate computer-animated images and robotic replicas of conspecifics, heterospecifics and predators [8–12]. Zebrafish have been used to study the fundamental mechanisms governing the exhibition of emotional patterns [13,14], individual response to alcohol and drugs of abuse [15–17], and higher order brain functions, such as memory and learning [18]. Zebrafish behaviour is typically measured from locomotory patterns [19,20], which are detected in the form of changes in speed and turn rate [5,6]; these patterns are annotated via human observation [21] or classified on the basis of sequences of trajectory data obtained using vision-based tracking [22,23].

While building of a comprehensive ethogram of zebrafish behaviour from video tracking data is well on its way [5], progress in the opposite direction, that is, generating realistic trajectory data through mathematical modelling of a given behaviour, is far from being realized. The ability to simulate such complex behaviours may inform experimental protocols and hypotheses by enabling *in silico* experiments that can be used to predict possible experimental outcomes without animal use. Within the three R's principle (replacement, reduction and refinement) [24], such virtual experiments can be used to perform *a priori* power analysis on the number of subjects, maximize the biological output of experimental observations and reduce the effect of biological confounds. All of these factors can contribute to a reduction in animal use and suffering.

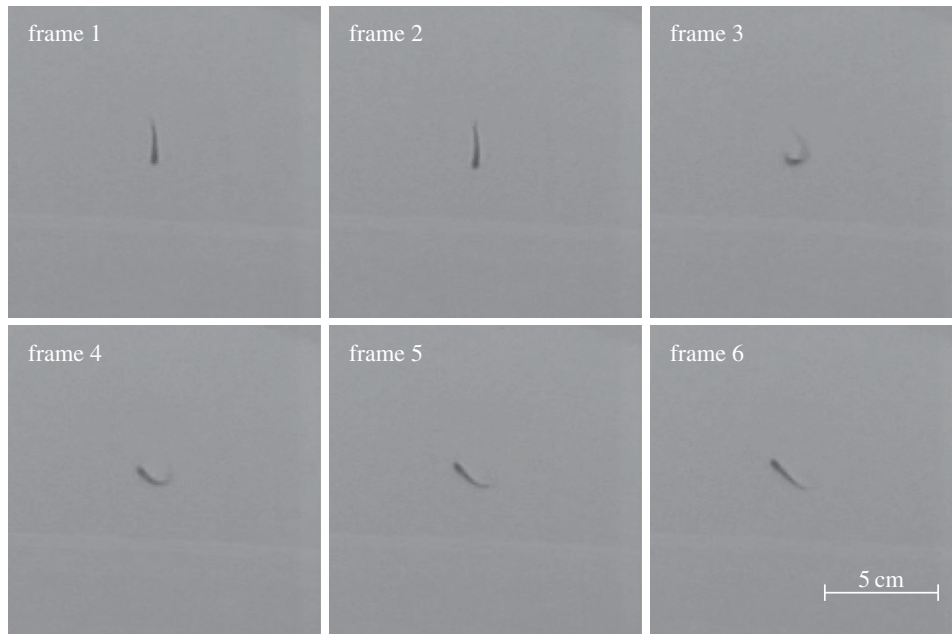


Figure 1. Fast zebrafish manoeuvre. Consecutive images of a zebrafish swimming over a 250 ms video sequence showing a sudden, high turn rate. Images are from a video in [34].

Moreover, models of individual fish motion form an important building block in modelling and analysis of fish collective behaviour [25–27].

Several models for the locomotion of an individual animal have incorporated uncorrelated and unbiased random walks, where increments to an animal's position are assumed to be independent [28,29]. Recent works have also addressed animal movement at the individual level using correlated random walks (CRWs) [26,30–33]. In these studies, successive steps taken by the animal are assumed to be correlated, as in the persistent random walker (PRW) model proposed in [30] to study ant displacement and the CRW model proposed in [31] to study caribou movement. These types of processes have recently been adapted to describe fish turn rate in the form of a persistent turning walker (PTW) [33]. In particular, the tendency of a barred flagtail fish (*Kuhlia mugil*) to turn both ways—left and right—was captured by modelling the turn rate as a stochastic mean reverting process with a drift toward zero [33].

In this paper, we develop a stochastic mean reverting jump diffusion model to study zebrafish locomotion. To better represent the swimming movements of zebrafish, we extend the PTW model [33] to allow for the possibility of sudden jumps in the fish turn rate (figure 1). We refer to this dynamic model as the jump persistent turning walker (JPTW). The jumps are modelled through a stochastic mean reverting jump diffusion process, which captures large deviations in turn rate in an otherwise well-behaved random process. Jump diffusion processes are commonly used to model sudden changes in stock prices and interest rates in finance [35,36], considerable deviations in particle kinematics in statistical physics [37–39], and the dispersal of microorganisms and foraging of animals in biology [40–42]. We evaluate the JPTW model on an experimental dataset [34] consisting of trajectories of single zebrafish in a shallow water tank. We identify the parameters of our model and of the PTW model from the observed trajectories, and statistically compare the goodness of fit of the models, revealing the importance of incorporating large-deviation jumps in zebrafish locomotion.

2. Results

2.1. Zebrafish exhibits a complex behavioural response during experimental observations

Visual inspection of 5-min video recordings revealed typical zebrafish behavioural responses. Specifically, in figure 2, traces of these fish trajectories show that zebrafish direction of motion is highly variable, while the speed displays more modest variations (see also the electronic supplementary material). We identify segments of a fish trajectory in which the direction of motion changes abruptly with black circles. These circles are placed where the absolute value of the instantaneous turn rate was found to be at least three times larger than the turn rate standard deviation; the prevalence of such circles emphasizes the rate at which large deviations in the turn rate can occur.

The zebrafish turn rate has an average value of $-0.062 \pm 0.098 \text{ rad s}^{-1}$ for the entire group, suggesting that, on average, fish have no preference to turn in either direction. In figure 3, we present a representative zebrafish turn rate time series with a corresponding histogram of the increments to this process. Consistent with the circles in figure 2, the instantaneous turn rate can exceed three times the turn rate standard deviation (figure 3*a*). Meanwhile, the distribution of turn rate increments exhibits heavy tails and a peakedness that cannot be described by a Gaussian distribution (figure 3*b*). In particular, the heavy tails observed in the histogram are due to the presence of numerous spikes in the turn rate time series.

Figure 4 shows that the increments to the turn rate are correlated with their history over the duration of a few seconds. While indicative of an inertia in the fish turn rate, the exponential decay seen in the autocorrelations also justifies the representation of the turn rate as an autoregressive (AR) process [43]. Assuming a first-order AR(1) model, for a given time step Δt corresponding to the video acquisition sampling time, the one-step correlation coefficient [44] of turn rate is approximately $e^{-\theta \Delta t}$, where the parameter θ is defined such that $1/\theta$ is the autocorrelation time. This parameter is included

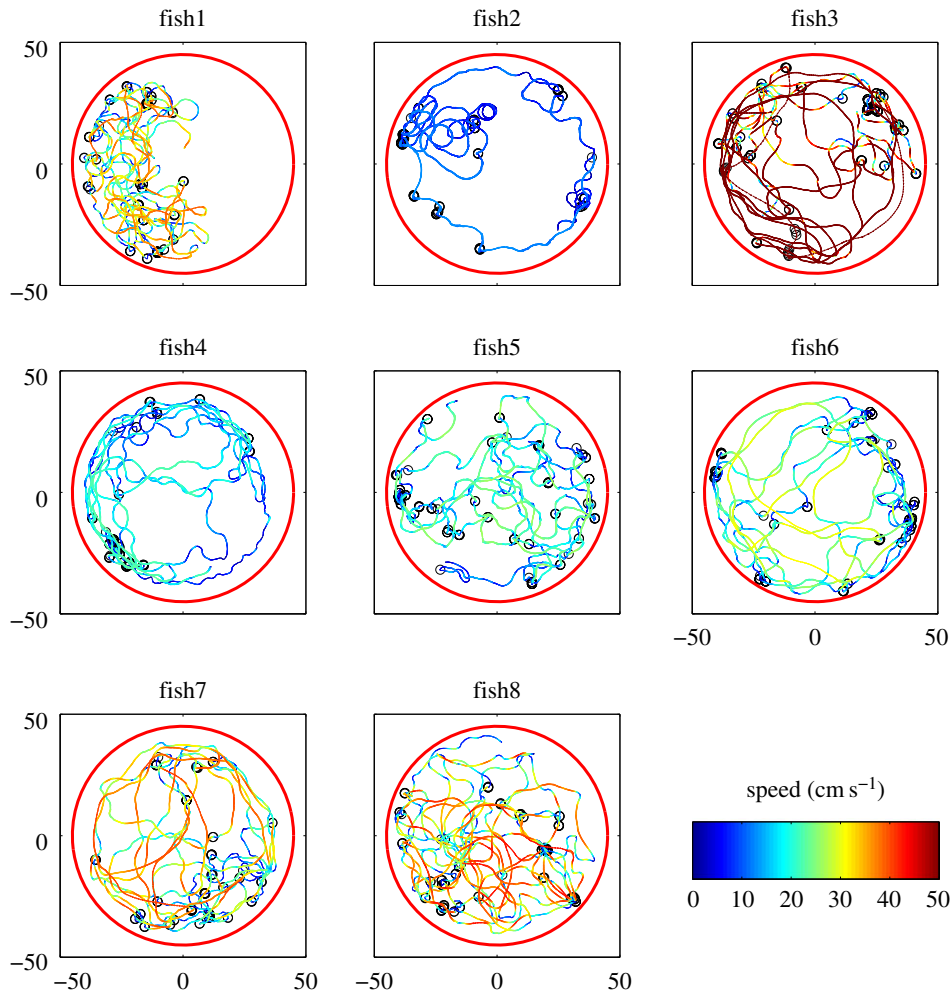


Figure 2. Zebrafish trajectories over 5 min for all the fish considered in the analysis. The circular contour of the shallow tank is plotted in red and the colour of the trajectories indicates instantaneous speed. Zebrafish are seen to not engage in wall-following behaviour and are instead observed to cover a large distance, while often changing directions with sharp turns, which are indicated with dark circles.

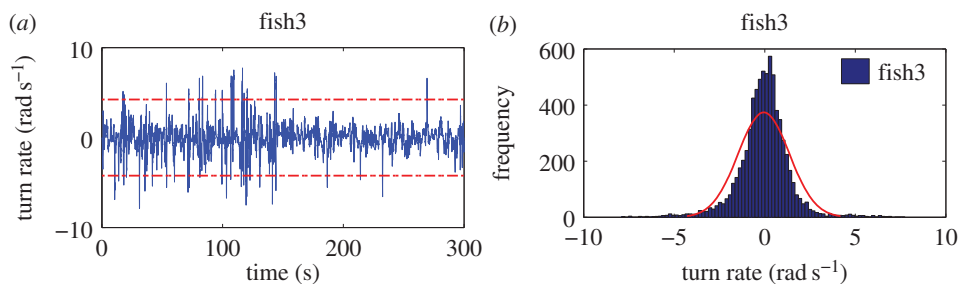


Figure 3. Representative fish turn rate time series (a) and best Gaussian fit constructed with the MATLAB (R2011a; MathWorks, Natick, MA, USA) routine *histfit* (b). Turn rate can spike over three times the standard deviation of the time series ($3 \times 1.4 \text{ rad s}^{-1} = 4.2 \text{ rad s}^{-1}$), which is represented with the dashed red lines in (a). In (b), these large deviations are responsible for the heavy tails observed in the histogram, which are not well described by a Gaussian distribution in red.

in the proposed turn rate model and has an average value $1.451 \pm 0.245 \text{ s}^{-1}$ across the whole group.

2.2. The JPTW turn rate distribution fits the empirical zebrafish turn rate

A comparison on the basis of a quantile–quantile (Q–Q) plot [45] of the empirical data against the predictions by the JPTW model shows a linear agreement (figure 5). In particular, the JPTW model presents a better agreement for all experimental datasets (figure 5a) compared with the PTW model (figure 5b). Closer analysis of the upper and lower quantiles

shows that the tails of the probability distribution remain aligned with the theoretical tails in the JPTW model, but do not match for the PTW model. Specifically, the PTW model assumes a turn rate where random increments are Gaussian, and, as a consequence, the upper and lower tails of the experimental data, corresponding to large changes in direction, are not matched by the theoretical quantiles of the Gaussian distribution. The PTW model, shown in green in figure 5c, does not match the observed distribution of turn rate, both in the tails of the distribution and for small turn rates, owing to the nature of the observed data, which show a distinct peakedness and heavy tails.

Table 1. Estimated parameters for the PTW (second and third columns) and JPTW (fourth, fifth, sixth, and seventh columns) models. The mean and standard deviation of the model parameters are computed at the bottom of the table. The value of the log-likelihood ratio used in the LRT test is presented in the last column of the table. Note that λ denotes the average number of jumps between two successive steps.

ID	θ (s^{-1})	σ ($rad\ s^{-1}$)	θ (s^{-1})	σ ($rad\ s^{-1}$)	γ ($rad\ s^{-1}$)	λ (s^{-1})	LRT
1	1.582	3.260	1.470	3.001	3.064	0.011	1071
2	1.249	1.435	1.013	1.145	1.565	0.015	3111
3	1.642	2.559	1.307	1.976	2.323	0.023	3454
4	1.603	2.673	1.460	2.491	2.725	0.012	1260
5	2.101	3.751	1.779	3.390	3.451	0.012	2353
6	1.708	2.617	1.299	2.197	2.422	0.016	3528
7	1.726	2.973	1.594	2.665	2.777	0.012	1210
8	1.870	3.611	1.688	3.415	3.206	0.012	1367
mean \pm s.d.	1.685 ± 0.245	2.860 ± 0.732	1.451 ± 0.245	2.535 ± 0.765	2.692 ± 0.594	0.014 ± 0.004	—

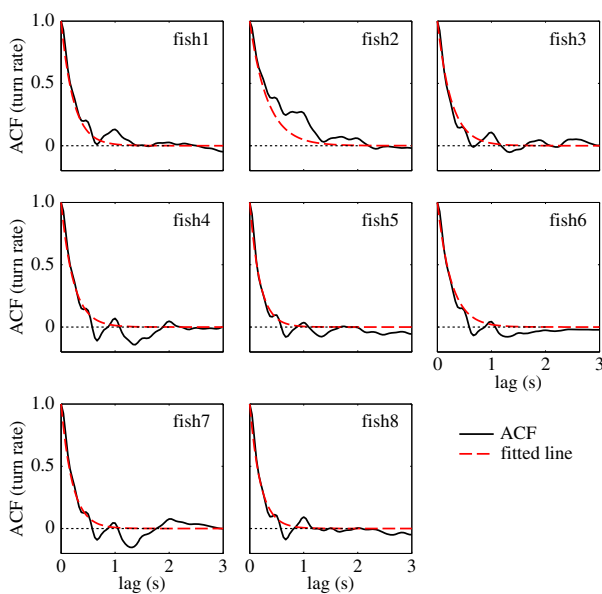


Figure 4. Zebrafish turn rate autocorrelation. Computed autocorrelations of the turn rates of eight fish (in solid black) show an exponential decay toward zero. Red dashed lines indicate the best exponential fit, computed using a linear regression on the logarithm of the ACF in MATLAB. (Online version in colour.)

2.3. Including jumps in the PTW model improves the statistical fit on empirical data

The proposed JPTW model introduces two parameters, in addition to the relaxation rate θ and turn rate variability σ defined in the PTW model, for describing the large, occasional changes in direction. A likelihood ratio test (LRT) comparing the two models (table 1), PTW and JPTW, shows that the incorporation of jumps in the former improves the statistical fit ($p < 0.001$, $LRT > \chi^2(2\text{ d.f.})$). One-way ANOVA comparisons on model parameters common to the PTW and JPTW models indicate that neither the relaxation rate nor the turn rate variability is differently estimated by the two models ($p = 0.077$, $F_{1,12} = 3.65$ relaxation rate; $p = 0.400$, $F_{1,12} = 0.75$, turn rate variability).

3. Discussion

In behavioural studies of zebrafish, methods that rely on human observation comprise an important tool-set used in

analysis [5,6]. Coupled with recently introduced tracking-based tools [23,46], these methods have led to considerable advancements in our understanding of zebrafish behaviour. Besides serving to facilitate the behavioural annotation process, the trajectory data made available through these tools can be further exploited to formulate and calibrate mathematical models of zebrafish behaviour. Such models can, in turn, be used to perform highly tailorable exploratory *in silico* experiments before animal testing. Computational trials can benefit both the biological output of animal tests, by maximizing the information content of trials, and the welfare of subjects, by improving the reliability of power analysis and reducing the role of biological confounds.

In this work, we present a model of fish turn rate temporal evolution in the form of a stochastic mean reverting jump diffusion process. In addition to offering a more robust model of zebrafish locomotion, the inclusion of jumps is intended to capture sudden and abrupt changes in direction [5,23]. Similar to previous studies on individual animal motion [32,33], zebrafish turn rate is found to have correlated time increments. Unlike the smooth trajectories seen in representative traces of larger fish, such as the barred flagtail fish in [33], the direction of motion of zebrafish is found to change abruptly. These occasional large deviations are responsible for the heavy tails observed in the probability distribution of the turn rate, and they suggest the use of a turn rate model that allows for such extreme events.

The fact that zebrafish locomotion is better described through the inclusion of jumps highlights well-known effects of morphology on swimming [47,48]. In particular, fish must accommodate the constraints imposed by viscous and inertial forces of the water. Zebrafish are thought to transition from a viscous to an inertial flow regime as they develop from larvae to adults, and its burst-and-coast swimming style may be a consequence of such a transition [49]. In adults, the burst phase of motion is often powered by a single tail flick, which changes the direction of the fish, and subsequent bursts can then occur before the resulting coasting phase is complete [49]. It is therefore tenable that zebrafish turn rate is better described through jumps. Larger fish, on the other hand, which spend more time in the inertial flow regime, tend to employ a more continuous swimming style [50]. In [26,33], for instance, a model of the turn rate of *K. mugil* without jumps, yields good predictions of its locomotion.

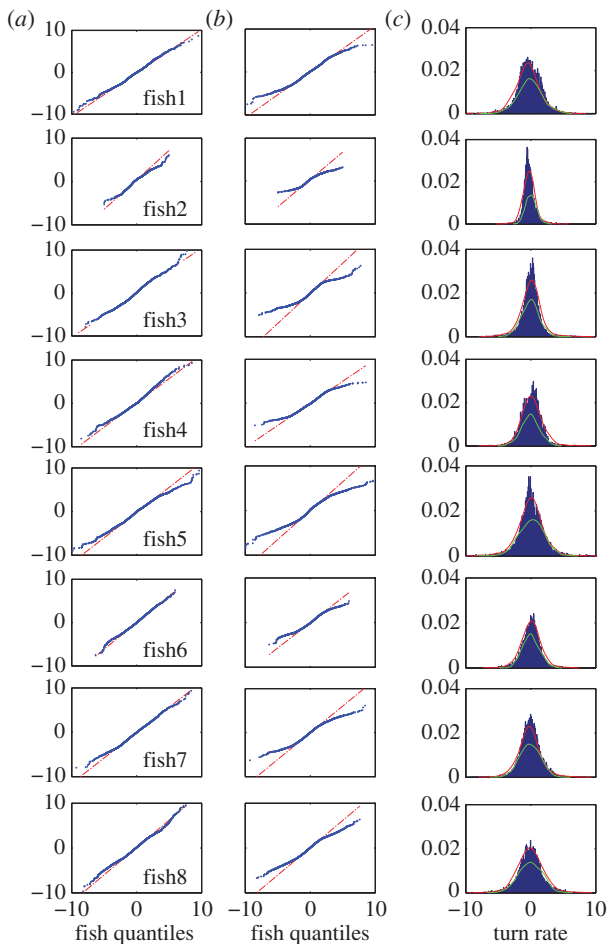


Figure 5. Comparison of the Q–Q plot of the seven fish turn rates based on both the JPTW (a) and PTW models (b) and normalized histograms for each (c). The quantiles for JPTW model are similar or linearly related to the real fish data, while the quantiles of the PTW model increments are not aligned with the tails of the empirical turn rate distribution. Histograms of each simulated model (red for JPTW and green for PTW) also show a better fit for the JPTW model.

However, *K. mugil* is at least six times larger than zebrafish and has a larger mass, which affects the interplay of viscous and inertial forces on the locomotion of the fish.

The parameters common to both the PTW and JPTW model include the relaxation rate, which measures the rate at which a turning fish ceases to turn, and the variability of the turn rate. Both parameters are not significantly different between models, suggesting that the inclusion of jumps in the PTW may not affect its baseline parameters, especially the turn rate variability. The parameters unique to the JPTW model, namely, the frequency of large turns and their magnitude, show that while jumps occur infrequently (approx. one large turn every 3 s), such turns are quick ($2.69 \pm 0.59 \text{ rad s}^{-1}$) compared with the average turn rate ($-0.062 \pm 0.098 \text{ rad s}^{-1}$). These large infrequent jumps are probably indicative of the high degree of manoeuvring control, coupled with an increased rate of axial undulations by the tail and fins by zebrafish adults [51]. Such manoeuvres, which are well below the magnitude of a typical escape response [5], are commonly exhibited by zebrafish in their natural habitat as part of an ontogenic shift in their foraging behaviour to avoid inter- and intra-specific competition [51].

Overall, these findings suggest that the inclusion of sudden jumps in a stochastic mean reverting model can aid in the quantification of zebrafish locomotion. The resulting JPTW

model enables a robust interpretation of the components of fish turn rate, while improving the fit of the experimental data to the model with respect to existing models of fish locomotion, where jumps are discarded. In addition to informing the design of zebrafish behavioural studies, the new model is fully customizable as a baseline for more complex studies. In particular, similar to [26,33], the effects of stimuli (e.g. a predator or conspecifics), environmental factors (e.g. wall or obstacle, light/dark) and possible variations of speed [52] can be incorporated in the model as control parameters. With respect to the latter aspect, we have recently been collaborating to assess the possibility of extending the PTW model to account for speed variations [53]. Finally, under the framework proposed here, collective behaviour can be studied using the individual kinematics of the JPTW model alongside defined information sharing rules [26].

4. Material and methods

4.1. Experimental apparatus and procedure and data collection

The data used in this paper are derived from the trajectory dataset published in [34]. The experimental apparatus consisted of a large $120 \times 120 \text{ cm}$ square tank with a 90 cm diameter ring to partition off the experimental region. The water was maintained at a constant depth of 10 cm. A high-resolution wide-angle camera (Hero 3, GoPro, San Mateo, CA, USA) was mounted 150 cm above the water surface to observe the fish. The tank was lit from above with four 25 W fluorescent lamps mounted approximately 100 cm above the water surface. The set-up was isolated with dark curtains on each side of the tank.

Zebrafish were procured from an online aquaria (Liveaquaria.com Rhinelander, WI, USA) and were kept in the housing facility for at least 12 days prior to the experiment. The water temperature and pH in the housing tanks was maintained at $25 \pm 1^\circ\text{C}$ and 7.2, respectively. The stocking density was maintained at no more than 1 fish per litre and the lighting was controlled according to a 12 L : 12 D circadian rhythm [54].

The experimental procedure consisted of introducing a single zebrafish in a large tank and recording videos of its swimming movements. Ten experimentally naive fish were observed for 5 min at 24 frames per second, so that $\Delta t = 1/24 \text{ s}$. Further details of the experimental procedure are provided in [34].

4.2. Data processing

Fish motion was tracked using an automated tracking algorithm [34] developed in MATLAB that used a Kalman filter to estimate the two-dimensional position. Briefly, the Kalman filter used the fish blob centroid, after background subtraction, as measurement and a constant velocity assumption between successive frames to optimize the position and velocity estimates in a minimum mean-square sense. Fish trajectories were further verified and repaired, if needed, on a graphical user interface (GUI) to ensure the availability of full-length tracks for each experiment. Similar to [33], oscillations due to the beating of the fish tail, which occur at approximately 2 Hz on average for zebrafish [55], were removed from the data using a Daubechies wavelet filter [56] implemented with the *wden* routine in MATLAB.

Trajectory data were analysed to exclude trials where the fish demonstrated abnormal freezing or wall-following. Freezing was quantified on the basis of [57], where the fish was considered freezing if it remained within a circle of 2 cm radius for more than 2 seconds. A threshold of 10% of time freezing was used to dismiss trials for abnormal stress levels [11,21]. To avoid

numerical confounds associated with the interaction with the wall, fish spending more than 35% of the experimental time within two body lengths of the circular partition were also dismissed. Future work will seek to enrich the proposed JPTW with a model of wall interactions, following a similar line of arguments as [53], to predict both orientation and position of zebrafish in tanks of varying size and geometry. We anticipate that such interactions will be more relevant as the tank size is decreased and corners are present.

Following the methodology of [33], trajectory data were mapped to the intrinsic coordinates $(S(k), \psi(k))$, where $S(k)$ is the path length at discrete time k , and $\psi(k)$ is the direction of motion. The latter is defined such that $\psi(k) - \psi(k-1)$ is the angle between the velocity vectors of the fish at the video frames k and $k-1$, for $k=2, \dots, T$, where T is the total number of frames. Fish instantaneous speed $v(k)$ (cm s^{-1}) and turn rate $\omega(k)$ (rad s^{-1}) were initially computed with central difference approximations; notably, no filtering was performed on the turn rate, except for treating a few false readings (with values greater in magnitude than 20 rad s^{-1}), for which the turn rate at the previous time step was then used. The intrinsic coordinates $(S(k), \psi(k))$ were then obtained using finite differences. We comment that the procedure to construct $\psi(k)$ does not wrap the fish direction of motion around $\pm \pi$, leaving the variable $\psi(k)$ free of artificially induced jumps.

4.3. Fish model

We extend the work in [33] to model zebrafish turn rate in the form of a mean reverting stochastic differential equation (SDE) with jumps, which are intended to capture sudden changes in the turn rate. Our model consists of three components in a continuous-time formulation. The first component drives the turn rate to its mean value, which is chosen to be zero. We refer to this process as mean reversion. The second component is standard Brownian motion, which represents continuous but well-behaved variation in the turn rate, and is modelled using a Wiener process [58]. The final component is a jump process whose mechanism is derived in what follows. In brief, the jump term will occasionally increment the turn rate by a potentially larger amount. More formally, the turn rate ω_t is described by the following SDE in time t

$$d\omega_t = -\theta\omega_t dt + \sigma dB_t + dJ_t, \quad (4.1)$$

where dt (s) is an infinitesimal time increment; θ (s^{-1}) denotes the mean reversion coefficient, also known as the relaxation rate [32] that characterizes the tendency of a turning fish to return to straight swimming (its inverse $\tau = 1/\theta$ (s) is the relaxation time [26]); σ (rad s^{-1}) captures the variability of the turn rate increments; B_t is a standard Wiener process [58] whose increments dB_t are mutually independent and sampled from a normal distribution $\mathcal{N}(0, dt)$; J_t is the jump process [59]; and dJ_t is an infinitesimal jump increment. Here and in what follows, we use a subscript t to identify continuous-time stochastic processes.

The jump diffusion term J_t in equation (4.1) is modelled as a compounded Poisson process [35], that is, $J_t = \sum_{j=1}^{v_t} Y_j$, where Y_j , for $j=1, 2, \dots$, are independent and identically distributed (i.i.d.) Gaussian random variables with variance γ^2 . The value of the jump process at time t depends on the number of terms Y_j that are included in the summation, which is determined by v_t . The number of jumps v_t is a counting process [60], defined so that for any r and t such that $r \leq t$, $v_t - v_r$ is a Poisson random variable with parameter $\lambda(t-r)$. Hence, v_t controls the number of discrete changes to the fish turning rate through J_t , and the rate of the occurrence of these jumps is λ . When a change in the fish turn rate due to a jump occurs, the value of the jump is normally distributed with variance γ^2 . Based on this approach, the model (4.1) is observed to capture two

distinct types of dynamics, one based on the Wiener process, and another that additionally considers larger changes in turn rate occurring at a rate λ .

Using Itô's integral formula [58,61], for any r and t such that $r \leq t$, a solution to equation (4.1) is

$$\begin{aligned} \omega_t = & m(\omega_r, t-r) + \sigma e^{-\theta(t-r)} \int_r^t e^{\theta(z-r)} dB_z \\ & + e^{-\theta(t-r)} \int_r^t e^{\theta(z-r)} dJ_z, \end{aligned} \quad (4.2)$$

where $m(\omega_r, t-r) \equiv \mathbf{E}[\omega_t | \omega_r, t-r]$ and $v(t-r) \equiv \mathbf{Var}[\omega_t | \omega_r, t-r]$ are the conditional mean and variance of the stochastic process ω_t without the jump diffusion term and given a previous turn rate ω_r . They are given by the following analytical expressions:

$$m(\omega_r, t-r) = \omega_r e^{-\theta(t-r)}, \quad (4.3a)$$

and

$$v(t-r) = \frac{\sigma^2}{2\theta} (1 - e^{-2\theta(t-r)}). \quad (4.3b)$$

For small time increments Δt , a discrete-time approximation [62,63] of equation (4.2) is

$$\begin{aligned} \omega(k) = & m(\omega(k-1), \Delta t) + \sqrt{v(\Delta t)\Delta t} \varepsilon(k) + \gamma(v(k\Delta t) \\ & - v((k-1)\Delta t))\zeta(k), \end{aligned} \quad (4.4)$$

where $\omega(k)$, $k=1, \dots, T$, is a discrete-time approximation that converges weakly to the continuous time process ω_t [64]; $\varepsilon(k)$ and $\zeta(k)$, for $k=1, \dots, T$, are normally distributed random variables with zero mean and unit variance; and $v(k\Delta t) - v((k-1)\Delta t)$ is a Poisson random variable with intensity $\lambda\Delta t$.

We assume that at most one jump occurs in a small discrete time interval Δt . Therefore, the Poisson random variable $v(k\Delta t) - v((k-1)\Delta t)$ can be approximated by a Bernoulli-distributed random variable with parameter $\lambda\Delta t$ [65]. Thus, with probability $\lambda\Delta t$, an increment to the turn rate includes the jump term, the Brownian motion, and the mean reversion; with probability $1 - \lambda\Delta t$, the turn rate is affected by only the Brownian motion and mean reversion. Accordingly, we obtain the following conditional likelihood probability density function (pdf) for the turn rate in equation (4.4)

$$\begin{aligned} f_{\theta, \sigma, \lambda, \gamma}^{(k)}(\omega(k) | \omega(k-1)) = & (1 - \lambda\Delta t) \phi(\omega(k); m(\omega(k-1), \Delta t), v(\Delta t)) \\ & + \lambda\Delta t \phi(\omega(k); m(\omega(k-1), \Delta t), v(\Delta t) + \gamma^2), \end{aligned} \quad (4.5)$$

where $\phi(\omega; m, v)$ is the Gaussian pdf whose arguments are the mean m and variance v defined in equations (4.3a) and (4.3b), respectively, and $(\theta, \sigma, \gamma, \lambda)$ denotes the set of unknown parameters of the JPTW model.

4.4. Statistical analysis

The autocorrelation lag coefficients were computed in MATLAB according to the approach in [44]. Specifically, the autocorrelation for a corresponding time lag k is

$$\text{ACF}(k) = \frac{\sum_{s=1}^{T-k} (\omega(s) - \bar{\omega})(\omega(s+k) - \bar{\omega})}{\sum_{s=1}^T (\omega(s) - \bar{\omega})^2}, \quad (4.6)$$

where $\bar{\omega}$ is the average value of the turn rate time series $\omega(k)$, $k=1, \dots, T$. Using the expression of $\omega(k)$ in equation (4.4), the one-step correlation coefficient of the turn rate realizations is $\text{ACF}(1) \simeq e^{-\theta\Delta t}$ for a given time increment Δt .

Model parameters were estimated using a maximum-likelihood estimation (MLE) method [66]. Specifically, we numerically maximize the log-likelihood for the PTW model

$$\mathcal{L}_{\text{PTW}} = \sum_{k=1}^T \log f_{\theta, \sigma}^{(k)}(\omega(k) | \omega(k-1)),$$

with the respect to the parameters θ and σ , where T is the number of time steps. The initial value for the PTW model parameters are set to the values obtained using a simple linear regression, see for example [63]. We also numerically maximize the log-likelihood JPTW model

$$\mathcal{L}_{\text{JPTW}} = \sum_{k=1}^T \log f_{\theta, \sigma, \lambda, \gamma}^{(k)}(\omega(k) | \omega(k-1)),$$

with respect to the parameters θ , σ , λ and γ . To estimate the parameters of the JPTW model, the initial values of the parameters θ and σ are set to the MLE estimates [63] of the PTW model without jumps, the initial value for the jump deviation is $\gamma = 1$, and the initial value of the jump rate λ is set to the fraction of video frames in which the fish turn rate exceeds three times the standard deviation of the experimental turn rate. The optimization algorithm was implemented in MATLAB using the global search and multistart solvers available in the global optimization toolbox (details on the robustness of the identification can also be found in the electronic supplementary material).

To compare the goodness of fit of either the JPTW model or the PTW model to fish data, we used the LRT [67]. The null hypothesis is rejected if the value of $\text{LRT} = 2(\mathcal{L}_{\text{JPTW}} - \mathcal{L}_{\text{PTW}})$ was greater than the 5% quantile of a χ^2 distribution with

2 d.f. One-way ANOVA tests were used to analyse zebrafish turn rate and estimated model parameters for individual fish and their variations. The significance level was set to $p < 0.05$.

4.5. Model validation

Model validation was performed using a Q–Q plot [68]. When sample data arise from the modelled probability distribution, the Q–Q plot depicts a straight line [45]. The simulated turn rate time series with jumps was obtained using the discrete-time approximation in equation (4.4) with the MLE parameters estimated from experimental data, while simulated turn rates for the PTW model were obtained from equation (4.4) with the jump frequency λ set to 0.

Ethics statement. The experimental procedure in [34] is approved under the protocol number AWOC-2013–103 by the Animal Welfare Oversight Committee of the Polytechnic Institute of New York University (now New York University Polytechnic School of Engineering).

Acknowledgements. The authors would like to thank Dr Simone Macri for useful discussions.

Funding statement. This work was supported by the National Science Foundation under grant nos. CMMI-0745753, CMMI-1129820, and DGE-0741714.

References

- Lieschke GJ, Currie PD. 2007 Animal models of human disease: zebrafish swim into view. *Nat. Rev. Genet.* **8**, 353–367. (doi:10.1038/nrg2091)
- Spence R, Gerlach G, Lawrence C, Smith C. 2008 The behaviour and ecology of the zebrafish, *Danio rerio*. *Biol. Rev. Camb. Philos. Soc.* **83**, 13–34. (doi:10.1111/j.1469-185X.2007.00030.x)
- Barbazuk WB, Korf I, Kadavi C, Heyen J, Tate S, Wun E, Bedell JA, McPherson JD, Johnson SL. 2000 The syntenic relationship of the zebrafish and human genomes. *Genome Res.* **10**, 1351–1358. (doi:10.1101/gr.144700)
- Dooley K, Zon LI. 2000 Zebrafish: a model system for the study of human disease. *Curr. Opin. Genet. Dev.* **10**, 252–256. (doi:10.1016/S0959-437X(00)00074-5)
- Kalueff AV *et al.* 2013 Towards a comprehensive catalog of zebrafish behavior 1.0 and beyond. *Zebrafish* **10**, 70–86. (doi:10.1089/zeb.2012.0861)
- Maximino C, de Brito TM, da Silva Batista AV, Herculano AM, Morato S, Gouveia A. 2010 Measuring anxiety in zebrafish: a critical review. *Behav. Brain Res.* **214**, 157–171. (doi:10.1016/j.bbr.2010.05.031)
- Kalueff AV, Stewart AM, Gerlai R. 2014 Zebrafish as an emerging model for studying complex brain disorders. *Trends Pharmacol. Sci.* **35**, 63–75. (doi:10.1016/j.tips.2014.07.007)
- Saverino C, Gerlai R. 2008 The social zebrafish: behavioral responses to conspecific, heterospecific, and computer animated fish. *Behav. Brain Res.* **191**, 77–87. (doi:10.1016/j.bbr.2008.03.013)
- Polverino G, Abaid N, Kopman V, Macri S, Porfiri M. 2012 Zebrafish response to robotic fish: preference experiments on isolated individuals and small shoals. *Bioinspir. Biomim.* **7**, 036019. (doi:10.1088/1748-3182/7/3/036019)
- Luca R, Gerlai R. 2012 In search of optimal fear inducing stimuli: differential behavioral responses to computer animated images in zebrafish. *Behav. Brain Res.* **226**, 66–76. (doi:10.1016/j.bbr.2011.09.001)
- Cianca V, Bartolini T, Porfiri M, Macri S. 2013 A robotics-based behavioral paradigm to measure anxiety-related responses in zebrafish. *PLoS ONE* **8**, e69661. (doi:10.1371/journal.pone.0069661)
- Abaid N, Bartolini T, Macri S, Porfiri M. 2012 Zebrafish responds differentially to a robotic fish of varying aspect ratio, tail beat frequency, noise, and color. *Behav. Brain Res.* **233**, 545–553. (doi:10.1016/j.bbr.2012.05.047)
- Steenbergen PJ, Richardson MK, Champagne DL. 2011 The use of the zebrafish model in stress research. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **35**, 1432–1451. (doi:10.1016/j.pnpbp.2010.10.010)
- Stewart A, Gaikwad S, Kyzar E, Green J, Roth A, Kalueff AV. 2012 Modeling anxiety using adult zebrafish: a conceptual review. *Neuropharmacology* **62**, 135–143. (doi:10.1016/j.neuropharm.2011.07.037)
- Gerlai R, Lahav M, Guo S, Rosenthal A. 2000 Drinks like a fish: zebra fish (*Danio rerio*) as a behavior genetic model to study alcohol effects. *Pharmacol. Biochem. Behav.* **67**, 773–782. (doi:10.1016/S0091-3057(00)00422-6)
- Stewart A *et al.* 2011 Zebrafish models to study drug abuse-related phenotypes. *Rev. Neurosci.* **22**, 95–105. (doi:10.1515/rns.2011.011)
- Mathur P, Guo S. 2010 Use of zebrafish as a model to understand mechanisms of addiction and complex neurobehavioral phenotypes. *Neurobiol. Dis.* **40**, 66–72. (doi:10.1016/j.nbd.2010.05.016)
- Al-Imari L, Gerlai R. 2008 Sight of conspecifics as reward in associative learning in zebrafish (*Danio rerio*). *Behav. Brain Res.* **189**, 216–219. (doi:10.1016/j.bbr.2007.12.007)
- Stewart A *et al.* 2011 Neurophenotyping of adult zebrafish using the light/dark box paradigm. In *Zebrafish neurobehavioral protocols* (eds AV Kalueff, JM Cachat), pp. 157–167. Berlin, Germany: Springer.
- Egan RJ *et al.* 2009 Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish. *Behav. Brain Res.* **205**, 38–44. (doi:10.1016/j.bbr.2009.06.022)
- Speedie N, Gerlai R. 2008 Alarm substance induced behavioral responses in zebrafish (*Danio rerio*). *Behav. Brain Res.* **188**, 168–177. (doi:10.1016/j.bbr.2007.10.031)
- Blaser RE, Chadwick L, McGinnis GC. 2010 Behavioral measures of anxiety in zebrafish (*Danio rerio*). *Behav. Brain Res.* **208**, 56–62. (doi:10.1016/j.bbr.2009.11.009)
- Cachat J, Stewart A, Utterback E, Hart P, Gaikwad S, Wong K, Kyzar E, Wu N, Kalueff AV. 2011 Three-dimensional neurophenotyping of adult zebrafish behavior. *PLoS ONE* **6**, e17597. (doi:10.1371/journal.pone.0017597)
- Russell WMS, Burch KL. 1959 *The principles of humane experimental technique*. London, UK: Methuen & Co. Ltd.
- Sumpter D. 2010 *Collective animal behavior*. Princeton, NJ: Princeton University Press.
- Gautrais J, Ginelli F, Fournier R, Blanco S, Soria M, Chaté H, Theraulaz G. 2012 Deciphering interactions in moving animal groups. *PLoS Comput. Biol.* **8**, e1002678. (doi:10.1371/journal.pcbi.1002678)
- Herbert-Read JE, Perna A, Mann RP, Schaerf TM, Sumpter DJ, Ward AJ. 2011 Inferring the rules

- of interaction of shoaling fish. *Proc. Natl Acad. Sci. USA* **108**, 18 726–18 731. (doi:10.1073/pnas.1109355108)
28. Codling EA, Plank MJ, Benhamou S. 2008 Random walk models in biology. *J. R. Soc. Interface* **5**, 813–834. (doi:10.1098/rsif.2008.0014)
 29. Giuggioli L, Potts JR, Harris S. 2011 Brownian walkers within subdiffusing territorial boundaries. *Phys. Rev. E* **83**, 061138. (doi:10.1103/PhysRevE.83.061138)
 30. Kareiva PM, Shigesada N. 1983 Analyzing insect movement as a correlated random walk. *Oecologia* **56**, 234–238. (doi:10.1007/BF00379695)
 31. Bergman CM, Schaefer JA, Lutich S. 2000 Caribou movement as a correlated random walk. *Oecologia* **123**, 364–374. (doi:10.1007/s004420051023)
 32. Degond P, Motsch S. 2008 Large scale dynamics of the persistent turning walker model of fish behavior. *J. Stat. Phys.* **131**, 989–1021. (doi:10.1007/s10955-008-9529-8)
 33. Gautrais J, Jost C, Soria M, Campo A, Motsch S, Fournier R, Blanco S, Theraulaz G. 2009 Analyzing fish movement as a persistent turning walker. *J. Math. Biol.* **58**, 429–445. (doi:10.1007/s00285-008-0198-7)
 34. Ladu F, Butail S, Macrì S, Porfiri M. 2014 Sociality modulates the effects of ethanol in zebra fish. *Alcohol. Clin. Exp. Res.* **38**, 1–9. (doi:10.1111/acer.12432)
 35. Tankov P. 2003 *Financial modelling with jump processes*. Boca Raton, FL: CRC Press.
 36. Jorion P. 1988 On jump processes in the foreign exchange and stock markets. *Rev. Financial Stud.* **1**, 427–445. (doi:10.1093/rfs/1.4.427)
 37. Kampen V, Godfried N. 1992 *Stochastic processes in physics and chemistry*. Amsterdam, The Netherlands: Elsevier.
 38. Hall PL, Ross DK. 1981 Incoherent neutron scattering functions for random jump diffusion in bounded and infinite media. *Mol. Phys.* **42**, 673–682. (doi:10.1080/00268978100100521)
 39. Metzler R, Klafter J. 2000 The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **339**, 1–77. (doi:10.1016/S0370-1573(00)00070-3)
 40. Othmer HG, Dunbar SR, Alt W. 1988 Models of dispersal in biological systems. *J. Math. Biol.* **26**, 263–298. (doi:10.1007/BF00277392)
 41. Burrow JF, Baxter PD, Pitchford JW. 2008 Lévy processes, saltatory foraging, and superdiffusion. *Math. Model. Nat. Phenomena* **3**, 115–130. (doi:10.1051/mmnp:2008060)
 42. Ha S-Y, Levy D. 2009 Particle, kinetic and fluid models for phototaxis. *Discrete Continuous Dyn. Syst. B* **12**, 77–108. (doi:10.3934/dcdsb.2009.12.77)
 43. Dickey DA, Fuller WA. 1979 Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **74**, 427–431.
 44. Box GE, Jenkins GM, Reinsel GC. 2013 *Time series analysis: forecasting and control*. New York, NY: John Wiley & Sons.
 45. Wilk MB, Gnanadesikan R. 1968 Probability plotting methods for the analysis of data. *Biometrika* **55**, 1–17.
 46. Butail S, Bartolini T, Porfiri M. 2013 Collective response of zebrafish shoals to a free-swimming robotic fish. *PLoS ONE* **8**, e76123. (doi:10.1371/journal.pone.0076123)
 47. Hernández LP, Barresi MJF, Devoto SH. 2002 Functional morphology and developmental biology of zebrafish: reciprocal illumination from an unlikely couple. *Integr. Comp. Biol.* **42**, 222–231. (doi:10.1093/icb/42.2.222)
 48. Tytell ED, Borazjani I, Sotiropoulos F, Baker TV, Anderson EJ, Lauder GV. 2010 Disentangling the functional roles of morphology and motion in the swimming of fish. *Integr. Comp. Biol.* **50**, 1140–1154. (doi:10.1093/icb/icq057)
 49. Muller UK, Stamhuis EJ, Videler JJ. 2000 Hydrodynamics of unsteady fish swimming and the effects of body size: comparing the flow fields of fish larvae and adults. *J. Exp. Biol.* **203**, 193–206.
 50. Weihs D. 1980 Respiration and depth control as possible reasons for swimming of northern anchovy, *Engraulis mordax*, yolk sac larvae. *Fishery Bull.* **78**, 109–117.
 51. Danos N, Lauder GV. 2007 The ontogeny of fin function during routine turns in zebrafish *Danio rerio*. *J. Exp. Biol.* **210**, 3374–3386. (doi:10.1242/jeb.007484)
 52. Katz Y, Tunström K, Ioannou CC, Huepe C, Couzin ID. 2011 Inferring the structure and dynamics of interactions in schooling fish. *Proc. Natl Acad. Sci. USA* **108**, 18 720–18 725. (doi:10.1073/pnas.1107583108)
 53. Zienkiewicz A, Barton D, Porfiri M, di Bernardo M. In press. Data-driven stochastic modeling of zebrafish locomotion. *J. Math. Biol.*
 54. Cahill GM. 2002 Clock mechanisms in zebrafish. *Cell Tissue Res.* **309**, 27–34. (doi:10.1007/s00441-002-0570-7)
 55. Buss RR, Drapeau P. 2001 Synaptic drive to motoneurons during fictive swimming in the developing zebrafish. *J. Neurophysiol.* **86**, 197–210.
 56. Daubechies I, Sweldens W. 1998 Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.* **4**, 247–269. (doi:10.1007/BF02476026)
 57. Kopman V, Laut J, Polverino G, Porfiri M. 2013 Closed-loop control of zebrafish response using a bioinspired robotic-fish in a preference test. *J. R. Soc. Interface* **10**, 20120540. (doi:10.1098/rsif.2012.0540)
 58. Gardiner C. 1985 *Handbook of stochastic methods: for physics, chemistry, and the natural sciences*. Berlin, Germany: Springer.
 59. Klebaner FC et al. 2005 *Introduction to stochastic calculus with applications*, vol. 57. Singapore: World Scientific.
 60. Wolff RW. 1982 Poisson arrivals see time averages. *Oper. Res.* **30**, 223–231. (doi:10.1287/opre.30.2.223)
 61. Øksendal B. 2003 *Stochastic differential equations*. Berlin, Germany: Springer.
 62. Platen E, Bruti-Liberati N. 2010 *Numerical solution of stochastic differential equations with jumps in finance*, vol. 64. Berlin, Germany: Springer.
 63. Szimayer A, Maller R. 2004 Testing for mean reversion in processes of Ornstein–Uhlenbeck type. *Stat. Inference Stoch. Process.* **7**, 95–113. (doi:10.1023/B:SISP.0000026032.80363.59)
 64. Mikulevičius R, Platen E. 1988 Time discrete Taylor approximations for Itô processes with jump component. *Math. Nachr.* **138**, 93–104. (doi:10.1002/mana.19881380107)
 65. Ball CA, Torous WN. 1983 A simplified jump process for common stock returns. *J. Financial Quant. Anal.* **18**, 53–65. (doi:10.2307/2330804)
 66. Lo AW. 1988 Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econ. Theory* **4**, 231–247. (doi:10.1017/S0266466600012044)
 67. Vuong QH. 1989 Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica* **57**, 307–333. (doi:10.2307/1912557)
 68. Embrechts P, Klüppelberg C, Mikosch T. 1997 *Modelling extremal events: for insurance and finance*, vol. 33. Berlin, Germany: Springer.