Data-driven modeling of zebrafish behavioral response to acute caffeine administration

Daniel A. Burbano-L.[a], Maurizio Porfiri[a,b,*]

[a] *Department of Mechanical and Aerospace Engineering*
New York University, Tandon School of Engineering, New York, USA
[b] *Department of Biomedical Engineering and Department of Civil and Urban Engineering*
New York University, Tandon School of Engineering, New York, USA

Abstract

Over the last thirty years, we have witnessed a dramatic rise in the use of zebrafish in preclinical research. Every year, more than 5,000 technical papers are published about zebrafish, many of them seeking to explain the underpinnings of anxiety through animal testing. In-silico experiments could significantly contribute to zebrafish research and welfare, by offering new means to support the 3Rs principles of replacement, reduction, and refinement. Here, we propose a data-driven modeling framework to predict the anxiety-related behavioral response of zebrafish to acute caffeine administration. The modeling framework unfolds along a two-time-scale dichotomy to capture freezing behavior along a slow temporal scale and burst-and-coast locomotion at a fast time-scale. Anchored in the theory of Markov chains and stochastic differential equations, we demonstrate a parsimonious, yet robust, modeling framework to accurately simulate experimental observations of zebrafish treated at different caffeine concentrations. Our results complement recent modeling efforts, laying the foundations for conducting in-silico experiments in zebrafish behavioral pharmacology.

Keywords: anxiety, *Danio rerio*, in-silico, pharmacology, stochastic differential equation

1. Introduction

Zebrafish (*Danio rerio*) is a freshwater species that is employed as a model organism in a wide array of scientific disciplines, ranging from toxicology (Hill et al., 2005) and behavioural genetics (Norton and Bally-Cuif, 2010), and on to translational neuroscience (Stewart et al., 2014) and evolutionary ecology (Spence et al., 2008). Every year, more than 5,000 technical papers are published about zebrafish (Meyers, 2018). Easy maintenance, knowledge of their genome, and psychological and neurological homologies with humans are some of the key advantages that have promoted zebrafish use in preclinical research (Nusslein-Volhard and Dahm, 2002).

Experiments with zebrafish could help clarify the underpinnings of anxiety in humans, an emotional state that is associated with alarming and troubling feelings about potential threats, as defined by the Psychiatric Association (2013). Zebrafish exhibits a complex behavioral repertoire, which is modulated by acute and chronic administration of psychoactive compounds that can be simply released in water, such as ethanol, caffeine, citalopram, and cocaine, to name a few; in this vein, several pharmacological studies have utilized the zebrafish as an animal model to investigate the effect of psychoactive compounds on anxiety-related behavioral response (López-Patiño et al., 2008; Egan et al., 2009; Maximino et al., 2011).

Zebrafish anxiety-related behavioral phenotype (Maximino et al., 2010a,b; Kalueff et al., 2013) includes erratic activity (repeated darting accompanied by sudden changes in direction or velocity), geotaxis (preference to stay at the bottom of the tank), thigmotaxis (preference to stay closer to the walls), and freezing (complete cessation of movement except for eyes and gills). High-throughput pharmacological assays measure the behavior of fo-

*Corresponding author.
Email address: email address:mporfiri@nyu.edu (Maurizio Porfiri)

cal subjects along these metrics, or a subset of them, in response to acute or chronic administration at chosen concentrations. These studies typically require a large number of individuals (Khan et al., 2017) and many substances that could be harmful to the animals (Jayne and See, 2019; Sloman et al., 2019).

Although animal experiments are the final cornerstone against which hypotheses shall be tested, computational modeling represents a viable means to accelerate the process of scientific discovery and contribute to animal well-being. In particular, computational, in-silico, models of zebrafish behavioral response to psychoactive compounds could benefit all of the three pillars of 3Rs endeavors, namely, to reduce the number of subjects, refine experimental designs, and replace the use of live animals (Russell et al., 1959). Empowered with accurate and reliable computational tools, researchers could: (i) conduct in-silico pilot tests to isolate critical experimental conditions that should be tested with live animals; (ii) improve on the analysis of their in-vivo experimental data to garner further insight on zebrafish response; and (iii) inform initial power analysis to optimize the design of their in-vivo experiment.

Caffeine is the most consumed psychoactive compound in the world, with an estimate of 80% of the population consuming a caffeinated product every day (de Paula Lima and Farah, 2019). One of the side-effects of caffeine in humans relates to anxiety-related response (Fredholm et al., 1999). Experimental evidence on zebrafish suggests that acute caffeine treatment can stimulate fish activity at low doses, while evoking robust anxiety-related response in the form of increased freezing response and reduced locomotion at higher concentrations (Santos et al., 2017; Rosa et al., 2018; Neri et al., 2019).

Data-driven computational modeling has been recently shown to be an effective approach to describe zebrafish swimming in isolation (Gautrais et al., 2009; Mwaffo et al., 2015a; Mwaffo and Porfiri, 2015; Zienkiewicz et al., 2015b; Mwaffo et al., 2017b) and in groups (Gautrais et al., 2012; Calovi et al., 2015; Zienkiewicz et al., 2015a; Butail et al., 2016; Collignon et al., 2016; Mwaffo et al., 2017a,c; Calovi et al., 2018; Zienkiewicz et al., 2018). In particular, the jump persistent turning walker (JPTW) model has been proposed as a versatile paradigm to capture the burst-and-coast swimming style of zebrafish, consisting of isolated tail bursts that are followed by coasting phases without tail beating (Mwaffo et al., 2015a). In its original formulation, this model describes the turn rate dynamics of an individual swimming at a constant speed in shallow water via a stochastic mean reverting jump diffusion process (Klebaner, 2012). Building upon the JPTW, we have recently incorporated diving along the water column and speed modulation through a Cox-Ingersoll-Ross process (Mwaffo et al., 2017b).

While this modeling framework constitutes a promising approach toward studying zebrafish locomotion, its systematic use in the pharmacological study of anxiety-related response has yet to be demonstrated. Thus far, to the best of our knolwedge, the only study about modeling zebrafish response to psychoactive compounds is the paper by Mwaffo and Porfiri (2015), where the original formulation of the JPTW for the turn rate dynamics was calibrated on experimental data of zebrafish exposed to acute ethanol administration.

Several technical steps should be undertaken toward a parsimonious, accurate, and robust modeling framework for zebrafish behavioral pharmacology. First, existing modeling schemes should be extended to describe freezing behavior, which might dominate the response of treated individuals in many assays (Maximino et al., 2010a,b; Kalueff et al., 2013). Second, the description of the speed modulation should be refined to encapsulate transitions between freezing and swimming episodes. Third, the effect of pharmacological manipulations on all the model parameters should be robustly identified from experimental data of animal trajectories to allow for mapping treatment into phenotype. In this work, we seek to address these technical steps in the context of zebrafish response to acute caffeine treatment in a shallow water tank.

Our modeling framework predicts zebrafish behavior along two time-scales: a slow time-scale, where zebrafish alternates between freezing and swimming, and a fast time-scale, associated with the evolution of speed and turn rate during locomotion. The dynamics unfolding on the slow time-scale is described by a discrete-time Markov chain (Brémaud, 2013), with two states that are associated with swimming and freezing. The dynamics on the fast time-scale follows an improved version of the JPTW model, in which we employ a stochastic logistic equation (Gard, 1988), rather than a Cox-Ingersoll-Ross process to study speed modulation. The model has two constant parameters associated with the speed variability and its rate of growth. When compared to the Cox-Ingersoll-Ross process,

such a logistic equation offers improved robustness in handling the transitions between swimming and freezing, along with a reduced number of model parameters to be experimentally calibrated.

Model parameters are estimated using data obtained from over-head recordings of zebrafish behavior in a shallow-water tank. The parameters of the Markov chain are estimated from the frequency of the transitions between swimming and freezing (Bowman et al., 2013), while maximum likelihood (Lo, 1988) is utilized to calibrate the stochastic differential equations of the locomotion model. We use our previous dataset (Neri et al., 2019) consisting of subjects acutely treated at four different caffeine concentrations; that is, 0 mg/l, 25 mg/l, 50 mg/l, and 70 mg/l. We carry out a detailed statistical analysis to explore the dependence of the model parameters on caffeine treatment, thereby shedding light on the potential value of data-driven computational modeling in zebrafish pharmacology.

The rest of the paper is organized as follows. The description of the experiment along with the data collection and analysis are presented in Section 2. Our novel approach to modeling zebrafish behavior is presented in Section 3. Technical details on model calibration are summarized in Section 4. A parametric analysis of the model for different caffeine concentrations is presented in Section 5, along with in-silico experiments that illustrate the potential value of the approach. The main results along with concluding remarks and avenues for future research are discussed in Section 6.

2. Experiments: description, data collection, and analysis

2.1. Experiment description

We use the dataset from our previous experiment in Neri et al. (2019), approved by the Animal Welfare Committee of the New York University Tandon School of Engineering (protocol number 13-1424). The experiment consisted of recording the trajectories of individual fish acutely-exposed to different caffeine concentrations in a circular arena of 90 cm in diameter and 10 cm in depth. The arena was covered with white contact paper allowing high-contrast background that facilitated visual tracking, see Neri et al. (2019) for further details.

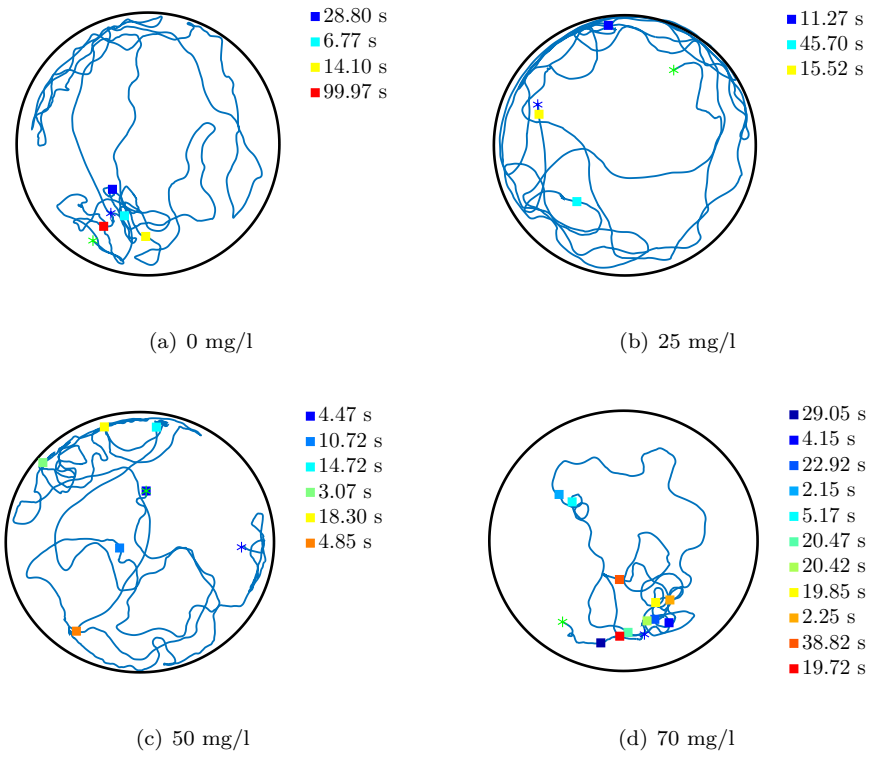
Four different caffeine concentrations C were considered: 0 mg/l (control condition), 25 mg/l, 50 mg/l, and 70 mg/l. Ten naïve fish were tested for each caffeine concentration (forty fish in total). The sex was randomized in the experiments and the average body length (BL) was approximately 3 cm. The pharmacological manipulation consisted of: (i) randomly picking one fish from the holding tank using a hand net and transferring it to a 500 ml glass beaker with a solution of water (from the experimental tank) and caffeine with concentration C; (ii) placing the beaker with the fish in the tank for one hour; and, finally, (iii) transferring the fish from the beaker to the arena using again a hand net.

2.2. Data collection and processing

Fish activity was recorded for 5 min ($T_{\text{exp}} = 300$ s) at 40 frame/s. Videos were processed in MATLAB, employing a multi-target tracking system (Ladu et al., 2014) to obtain individual $(x(k\Delta), y(k\Delta))$ centroid positions in the tank, with respect to a global Cartesian reference frame, where $\Delta = 1/40$ s is the sampling time and $k \in \{1, \dots, N = 12,000\}$ the time-stamp of the recording. The centroid positions were processed using a Daubechies wavelet filter (Mwaffo et al., 2015b). The filtered outputs were used for the identification of freezing episodes and for the computation of speed and turn rate, as illustrated in what follows.

Following Kopman et al. (2013), we identified a freezing episode when the fish centroid trajectories were enclosed in a circle of 2 cm in radius for 2 s. In particular, freezing can be described through a binary variable, denoted by $G(i\tilde{\Delta})$ taking values every 2 s, so that $\tilde{\Delta} = 80\Delta$ and $i \in \{1, \dots, \tilde{N} = 150\}$. The binary values that the variable can take are denoted with F and S, identifying freezing or swimming episodes. Specifically, we set $G(i\tilde{\Delta}) = \text{F}$ if the centroid positions $(x(i\tilde{\Delta}), y(i\tilde{\Delta}))$ along with all the subsequent centroid positions within a 2 s time window $(x(i\tilde{\Delta} + \Delta), y(i\tilde{\Delta} + \Delta)), \dots, (x(i\tilde{\Delta} + 79\Delta), y(i\tilde{\Delta} + 79\Delta))$, were contained in a circle of 2 cm in radius, and $G(i\tilde{\Delta}) = \text{S}$ otherwise. Not only does this definition of freezing include complete cessation of movement, but also it comprises akinesia episodes, where the fish exhibits an abnormal reduction of its swimming speed (Kalueff et al., 2013).

Exemplary fish trajectories for different caffeine concentrations are illustrated in Figure 1. Freezing instances with their duration are indicated in the figure to help appreciate typical patterns in zebrafish locomotion, alternating smooth swimming



(a) 0 mg/l (b) 25 mg/l

(c) 50 mg/l (d) 70 mg/l

Figure 1: Exemplary experimental trajectories at different caffeine concentrations. Each square represents a freezing episode, whose duration is indicated on the side legend. Green and blue asterisks mark the start and end of each trajectory, respectively.

segments with freezing episodes. As one might expect, the frequency of freezing bouts seems to be controlled by caffeine concentration, such that fish spend more time freezing at higher concentrations (Neri et al., 2019).

The fish speed $v(k\Delta)$ was obtained according to $v(k\Delta) := \sqrt{v_x^2(k\Delta) + v_y^2(k\Delta)}$, where $v_x(k\Delta)$ and $v_y(k\Delta)$ are two components of the velocity vector along the x- and y-axes, computed through numerical differentiation (using the first-order forward difference method).

In addition, we calculated the turn rate $\omega(k\Delta)$ by mapping the filtered centroid positions to intrinsic coordinates, as described by Gautrais et al. (2009). The method estimates the curvilinear abscissa $\mathcal{S}((k+1)\Delta)$ and the heading angle $\phi((k+1)\Delta)$ using the following three steps: (i) fitting three subsequent centroid positions $(x(k\Delta), y(k\Delta))$, $(x((k+1)\Delta), y((k+1)\Delta))$, and $(x((k+2)\Delta), y((k+2)\Delta))$ on a circle; (ii) calculating the increment of the heading angle $\delta\phi((k+1)\Delta)$, which is the angle between the two lines from the center of the circle to the first and third centroid positions, at the kth and $(k+2)$th steps, respectively; and (iii) approximating the turn rate as $\omega((k+1)\Delta) = \delta\phi((k+1)\Delta)/2\Delta$. No post-processing was conducted on the turn rate, except for removing a few false peaks whose magnitude exceeded 20 rad/s – supplementary analysis not included in this work suggests that local smoothing of the data would neither change the features of the time-series nor confound the effect of caffeine, but it may lower the values of the calibrated model parameters.

An example of the time-evolution of the speed and turn rate is shown in Figures 2 and 3, along with their corresponding histograms. Subscript S indicates that the computations have been executed only when the fish was swimming, that is, discarding freezing from the original measurements of $\omega(k\Delta)$ and $v(k\Delta)$ – the same notation will be used in the development of the data-driven mathematical model to distinguish swimming from freezing. The speed for the four different caffeine concentrations fluctuates around a mean value of 6.68 cm/s, 6.27 cm/s, 4.30 cm/s, and 3.41 cm/s, respectively.

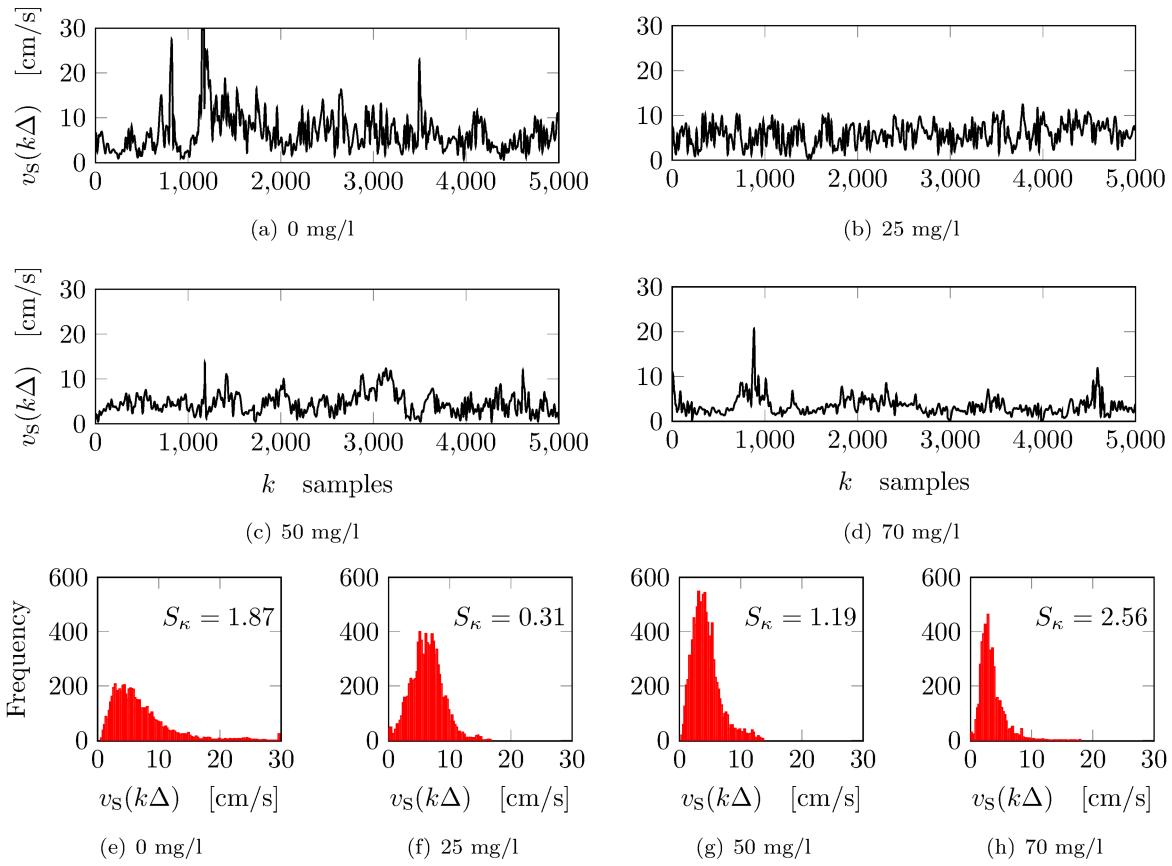


Figure 2: Time-trace and histograms of the speed for each fish trajectory from Figure 1 neglecting freezing episodes. The top four panels are the time- evolution of the speed (for visualization purposes, only the first 125 s are displayed) and the bottom row panels are their corresponding histograms. The bin size for all histograms is 0.25 cm/s, and the skewness is indicated by S_κ.

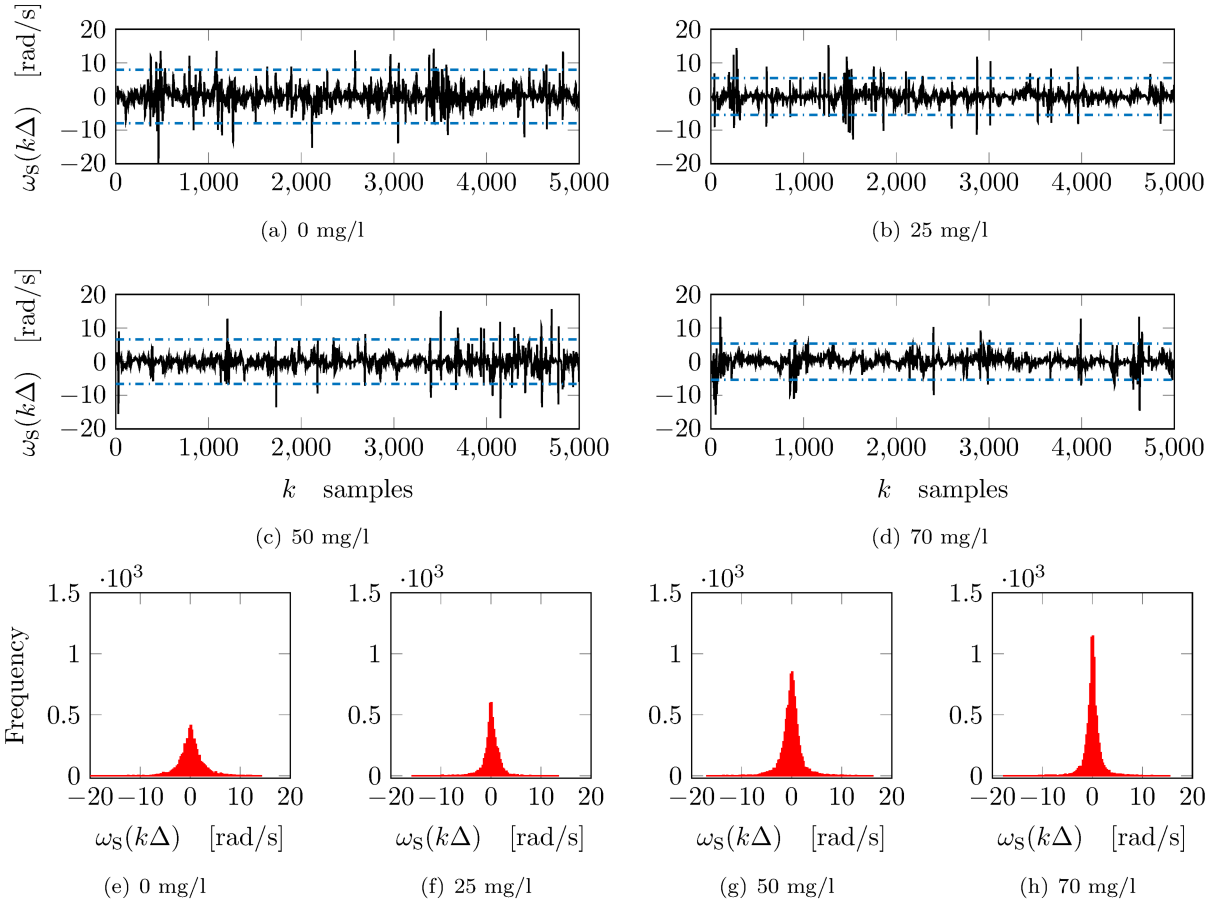
The distribution of the speed is positively skewed, indicating that there are more deviations from the mean towards higher speeds. Moreover, the turn rate fluctuates around zero, suggesting that fish do not have a preferential swimming direction in the tank. As expected from our previous work (Mwaffo et al., 2015a), the turn rate displays a sequence of spikes (or jumps), corresponding to values exceeding three times the standard deviation of the turning rate trajectory. These spikes correspond to C- and U-turns, typical bouts of burst-and-coast locomotion (Danos and Lauder, 2007), and they manifest into heavy tails in the distributions shown in the panels on the right-hand-side of Figure 3. We comment that the turn rate for the control subject shows more variability than the caffeine-treated subjects, whereby increasing C from 0 mg/l to 70 mg/l, the standard deviation of the turn rate attains the following values: 2.65 rad/s, 1.81 rad/s, 2.20 rad/s, and 1.78 rad/s, respectively.

2.3. Experiments and analysis

Four trials were discarded due to a recording issue that was detected in the post-processing phase, one for 25 mg/l, one for 50 mg/l, and two for 70 mg/l. Three additional trials, two for $C = 50$ mg/l and one for $C = 70$ mg/l, were dismissed, since the fish froze more than 95% of the time, thus failing to provide sufficient data for swimming trajectory analysis and parameter calibration.

The effect of caffeine concentration on the fish swimming behavior was quantified through four salient metrics, namely, (i) time spent freezing t_F, (ii) distance traveled D, (iii) mean speed V_S, and (iv) mean absolute turn rate Ω_S.

The time spent freezing was computed by summing all the occurrences of freezing, that is, aggregating all the numbers reported in the legends



(a) 0 mg/l (b) 25 mg/l (c) 50 mg/l (d) 70 mg/l

(e) 0 mg/l (f) 25 mg/l (g) 50 mg/l (h) 70 mg/l

Figure 3: Time-evolution and histograms of turn rate for each fish trajectory from Figure 1 neglecting freezing episodes. The top four panels are the time- evolution of the speed (for visualization purposes, only the first 125 s are displayed) and the bottom row panels are their corresponding histograms, plotted with a bin size of 0.25 rad/s. The dashed-dotted blue lines are three times the standard deviation of the turn rate ($\pm 3\text{std}[\omega_S(k\Delta)]$, where std$(\cdot)$ is the standard deviation).

of Figure 1. The distance traveled was obtained by summing the Euclidean distances between two consecutive centroid positions, $(x(k\Delta), y(k\Delta))$ and $(x((k+1)\Delta), y((k+1)\Delta))$, respectively. The mean speed was calculated by averaging the speed during swimming, that is, discarding freezing episodes. The mean absolute turn rate was equivalently computed by excluding freezing episodes when averaging the absolute value of the turn rate. We note that, for each trial, the number of samples used to compute V_S and Ω_S could be different, as fish spend different times swimming or freezing across trials. Further, we comment that dividing D by the time of the experiment (300 s) does not exactly yield V_S since in the latter quantity we exclude freezing episodes, where the speed should be zero in principle.

We compared the four metrics t_F, D, V_S, and Ω_S across different caffeine concentrations using a one-way analysis of variance (ANOVA) with C as the independent variable (Navidi, 2008). One outlier was identified and discarded from the analysis for $C = 70$ mg/l using the interquartile range rule (Navidi, 2008) on the average distance traveled. Once a significant difference was found with ANOVA, we applied the Fisher's least significant difference (LSD) method (Navidi, 2008) for conducting post-hoc analysis that could reveal differences between caffeine concentrations.

The overall results of the ANOVA on the considered four metrics are shown in Figure 4. As reported by Neri et al. (2019), caffeine concentration influences the time spent freezing ($F(3, 28) = 3.27$, $P = 0.035$). Specifically, post-hoc comparisons re-

vealed that subjects treated at $C = 70\,\text{mg/l}$ spent more time freezing than subjects in any other condition: control condition ($P = 0.014$), $25\,\text{mg/l}$ ($P = 0.002$), and $50\,\text{mg/l}$ ($P = 0.009$). Similarly, we found that the distance traveled D was affected by the caffeine concentration ($F(3, 30) = 3.27$, $P = 0.031$). Post-hoc analysis indicated that the distance traveled by individuals treated at the highest caffeine concentration of $70\,\text{mg/l}$ was lower than that of control subjects ($P = 0.012$) and fish treated at $50\,\text{mg/l}$ ($P = 0.008$).

In agreement with previous pharmacological studies (Santos et al., 2017; Rosa et al., 2018; Neri et al., 2019), we noted that the mean speed decreased as the the caffeine concentration increases. In addition, we found that the mean absolute turn rate had an inverted U-shape with the highest activity attained at $C = 50\,\text{mg/l}$. However, no significant differences were registered through ANOVA on either metric.

3. Data-driven modeling framework

3.1. Zebrafish kinematics

Zebrafish can be modeled as a rigid body moving on a horizontal plane. Fish motion is measured in the global reference frame $\{\mathcal{X}_I, \mathcal{Y}_I\}$ with origin $\mathcal{O}$, as shown in Figure 5. We also define a local reference frame $\{\mathcal{X}_R, \mathcal{Y}_R\}$ relative to the fish centroid position in the global frame as $P_C(t) = [x(t), y(t)]^\top$ at time t, where the superscript $\top$ is matrix transposition. Then, the pose of the fish with respect to the global reference can described by the three-dimensional vector $[x(t), y(t), \theta(t)]^\top$ where $\theta(t) \in [-\pi, \pi)$ is the angular rotation from the global to local reference frame. The pose in either the local or global reference can be mapped onto the other through a linear transformation (de Wit et al., 2012).

The fish pose in the global frame can be computed by integrating the following system of ordinary differential equations:

$$\begin{bmatrix} \frac{\mathrm{d}x(t)}{\mathrm{d}t} \\ \frac{\mathrm{d}y(t)}{\mathrm{d}t} \\ \frac{\mathrm{d}\theta(t)}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} v(t)\cos(\theta(t)) \\ v(t)\sin(\theta(t)) \\ \omega(t) \end{bmatrix}, \tag{1a}$$

with initial conditions

$$x(0) = x_0, \quad y(0) = y_0, \quad \theta(0) = \theta_0, \tag{1b}$$

where $v(t)$ and $\omega(t)$ are the instantaneous speed and turn rate of the fish, and $x_0, y_0, \theta_0 \in \mathbb{R}$ are the initial conditions.

Variables $v(t)$ and $\omega(t)$ constitute the key components of the kinematic model in Eq. (1); modeling their time-evolution with respect to caffeine administration is the chief objective of this study.

3.2. Zebrafish dynamics

To establish a valid modeling framework to study zebrafish response to caffeine exposure, it is important to account for the possibility of freezing. Toward this aim, we split the modeling problem into two different sub-problems. First, we examine the transitions between freezing and swimming, and then, during the swimming segments of the trajectory, we characterize the subject locomotion in terms of speed and turn rate. The resulting modeling architecture unfolds along a two time-scales dichotomy.

3.2.1. Transitioning between freezing and swimming

Toward modeling the transitions between freezing and swimming, we use a discrete-time Markov chain, as sketched in Figure 6. The red circle represents a freezing state, and the blue circle is associated with swimming. Each state has a probability of persistence, given by p_F and p_S, and a probability of transitioning to the other state, given by $p_\text{FS} = 1 - p_\text{F}$, and $p_\text{SF} = 1 - p_\text{S}$, respectively.

Letting $\Gamma(iT)$ be the binary random variable taking the values F (freezing) or S (swimming), the speed $v(t)$ and turn rate $\omega(t)$ can be written as

$$v(t) = \begin{cases} 0, & \text{if } \Gamma(iT) = \text{F} \\ v_\text{S}(t), & \text{if } \Gamma(iT) = \text{S} \end{cases}, \tag{2a}$$

$$\omega(t) = \begin{cases} 0, & \text{if } \Gamma(iT) = \text{F} \\ \omega_\text{S}(t), & \text{if } \Gamma(iT) = \text{S} \end{cases}, \tag{2b}$$

$$v(0) = v_0, \quad \omega(0) = \omega_0, \tag{2c}$$

where $t \in [0, T_\text{exp}]$, $i = \{1, \cdots, \widetilde{N}\}$ and $v_0, \omega_0 \in \mathbb{R}$ are the initial conditions. The realizations of $\Gamma(iT)$ are drawn from the Markov model shown in Figure 6 with $T > 0$ being a constant time period and initial condition given by

$$\Gamma(0) = \Gamma_0 \in \{\text{S}, \text{F}\}. \tag{3}$$

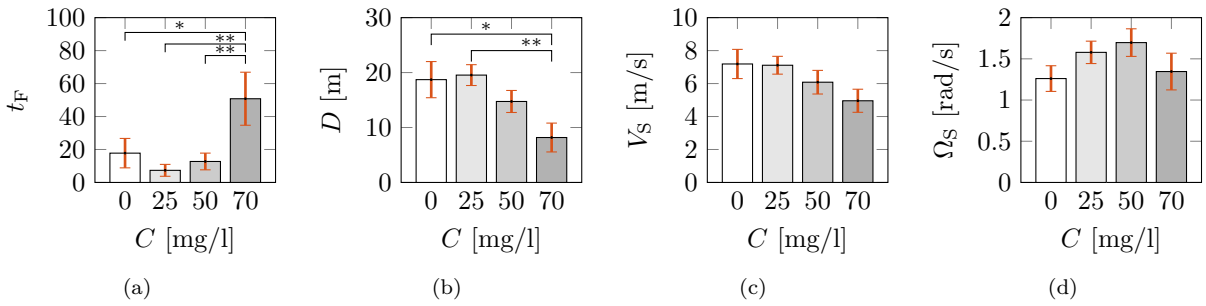


Figure 4: Effect of caffeine treatment on zebrafish behavior, scored through four metrics: (a) time spent freezing, as a fraction of the total experimental time of 300 s, (b) distance traveled, (c) mean speed, and (d) mean absolute turn rate. Vertical red lines represent standard error of the means (SEM), and symbols $**$ and $*$ indicate significant differences in post-hoc analysis, with P-values satisfying $P < 0.01$ and $P < 0.05$, respectively.

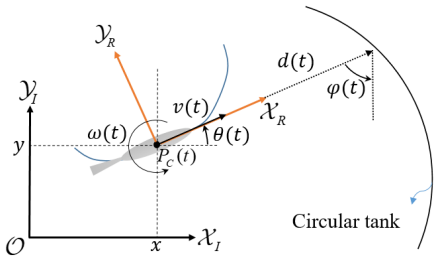


Figure 5: Zebrafish kinematics, depicting the global and local reference frames, along with the fish trajectory (blue line) and the interaction with the tank wall.

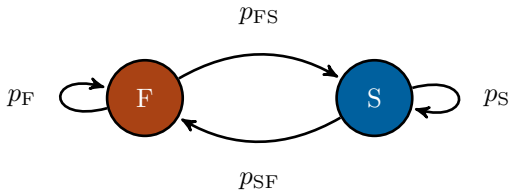


Figure 6: Markov model of swimming and freezing behavior of zebrafish.

We note that $\Gamma(iT)$ should predict the evolution of $G(i\tilde{\Delta})$ obtained from the experimental data. Equation (2) has two different time-scales associated with the transitions process $\Gamma(iT)$ that is updated with T, and the speed and turn rate that evolve continuously in time. Following the definition of freezing by Kopman et al. (2013), we set $T = 2\,\mathrm{s}$, so that $\widetilde{N} = 150$.

We acknowledge that in its present formulation, the model defines freezing as a complete cessation of movement. As a result, the theoretical definition of freezing excludes akinesia episodes that are instead retained in the analysis of experimental data as part of freezing, due to inherent uncertainty in the tracking software. A potential way to reconcile experimental constraints with mathematical assumptions might be to add a nonzero threshold in both the speed and the turn rate in Eqs. (2a) and (2b).

3.2.2. Locomotion

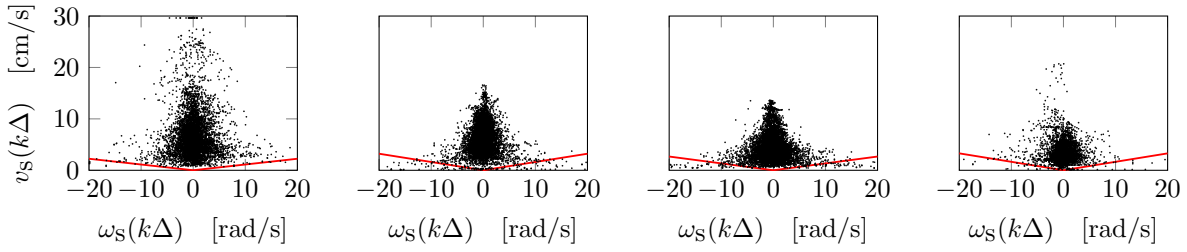
To describe the evolution of the speed and turn rate once the fish enters the swimming state ($\Gamma(iT) = \mathrm{S}$), we use a set of continuous-time stochastic differential equations that extend previous work on zebrafish swimming in shallow water (Mwaffo et al., 2015a; Mwaffo and Porfiri, 2015; Zienkiewicz et al., 2015b).

Experimental observations of fish speed (see Figure 2) for different caffeine concentrations C indicate that the probability distribution of the speed is positively skewed and has a shape similar to a gamma distribution. To encapsulate these features, we propose the use of a stochastic logistic equation (Pasquali, 2001), as follows:

$$\begin{aligned} dv_S(t) &= (\eta v_S(t) - g(\omega_S(t))v_S^2(t))dt \\ &\quad + \sigma_v v_S(t) dW_v(t), \end{aligned} \tag{4}$$

where η [s^{-1}] and σ_v [$\mathrm{s}^{-1/2}$] are positive parameters of the logistic model, measuring the linear rate of growth of the speed and the strength of the added noise, respectively; $W_v(t)$ is a standard Wiener process; and the nonlinear function $g(\omega_S(t))$ [cm^{-1}] regulates the increase or decrease of the speed according to the turn rate $\omega_S(t)$.

To adequately select the function $g(\omega_S(t))$, we first recall that the expected value of Eq. (4) for $\sigma_v^2 < 2\eta$, and a given value of the turn rate, say



(a) 0 mg/l (b) 25 mg/l (c) 50 mg/l (d) 70 mg/l

Figure 7: Phase plots of turn rate versus speed for each fish trajectory from Figure 1 neglecting freezing episodes. The red line represents the function $g(\omega_S)$ that couples the turn rate with the speed modulation process.

$\omega_S(t) = \omega^*$, is given by (Pasquali, 2001)

$$\mathbb{E}[v_S(t)] = \frac{1}{g^*}\left(\eta - \frac{\sigma_v^2}{2}\right), \tag{5}$$

where $g^* = g(\omega^*)$ quantifies the influence of the turn rate on the expected value of the speed, such that fish should decrease their speed when experiencing large turn rates (for example, during C- and U-turns), and, vice versa, they will be able to attain larger speeds during straight swimming.

Next, we plot phase portraits of turn rate versus speed for the fish trajectories of Figure 1, corresponding to different caffeine concentrations, as shown in Figure 7. Therein, we identify a clear triangular-like region with higher turn rates corresponding to lower speeds, and vice versa. This relation can be captured by considering the following equation:

$$g(\omega_S(t)) = \frac{1}{\mathrm{std}_\omega \mathrm{BL}}|\omega_S(t)|. \tag{6}$$

Here, $\mathrm{std}_\omega = \mathrm{std}[\omega_S(t)]$ is the experimental value of the standard deviation in the turn rate, which is treated as a fish-specific parameter (as indicated for example in Figure 1), and $\mathrm{BL} = 3\,\mathrm{cm}$ is the average body length of the animals. Equation (6) does not require any parameter for calibration, whereby it assumes a simple linear relationship on the absolute turn rate with a coefficient of proportionality that is given by experimental data. By scaling the turn rate by its standard deviation and the body length, we recognize that C- and U-turns are associated with extreme values of the turn rate and are defined with respect to the characteristic size of the animal.

According to the JPTW model, the turn rate evolution is governed by the following mean reversion jump diffusion process (Mwaffo et al., 2015a):

$$\begin{aligned} d\omega_S(t) = (-\alpha\omega_S(t) + f(\varphi(t), d(t)))dt \\ + \sigma_\omega dW_\omega(t) + dJ(t), \end{aligned} \tag{7}$$

where α [s^{-1}] and σ_w [$\mathrm{rad\,s}^{-3/2}$] are positive parameters representing the relaxation rate associated with the ability of fish to resume straight swimming and the strength of the added noise, respectively; $W_\omega(t)$ is a standard Wiener process; and $J(t)$ is a jump process capturing sudden turn rates that are typical of zebrafish swimming style (Mwaffo et al., 2015a). In particular, the jump diffusion term is a compound Poisson process $J(t) = \sum_{j=1}^{m(t)} Y_j$, where Y_j are independent and identically distributed Gaussian random variables with zero mean and variance γ^2. The intensity of the jump term at time t is determined by the stochastic process $m(t)$, whose increments $m(t'') - m(t')$ are Poisson random variables with parameter $\lambda(t'' - t')$ for any $t', t'' \in [t_1, t_2]$ with $t'' > t'$.

The nonlinear function $f(\varphi(t), d(t))$ [rads^{-2}] represents the interaction with the wall, as illustrated in Figure 5, and can be approximated following Gautrais et al. (2009) and Zienkiewicz et al. (2015b) through

$$f(\varphi(t), d(t)) = a\,\mathrm{sgn}(\varphi(t))e^{-bd(t)} \tag{8}$$

where $\mathrm{sgn}(\cdot)$ is the sign function, a [rads^{-2}] and b [cm^{-1}] are positive parameters, $d(t)$ is the distance from the wall, and $\varphi(t)$ is the projected angle to collision, as illustrated in Figure 5.

4. Computational implementation of the modeling framework

4.1. Calibration of the transition probabilities of the Markov model

From the real time-series of the binary variable $G(i\tilde{\Delta})$ for each fish, we estimate the transition probabilities through

$$p_{\mathrm{SF}} = \frac{N_{\mathrm{SF}}}{N_{\mathrm{SS}} + N_{\mathrm{SF}}}, \quad p_{\mathrm{FS}} = \frac{N_{\mathrm{FS}}}{N_{\mathrm{FF}} + N_{\mathrm{FS}}}, \tag{9}$$

where N_{SF} and N_{FS} are the number of transitions from swimming to freezing and freezing to swimming, respectively. Similarly, N_{SS} and N_{FF} are the number of times the fish remains in the swimming or freezing state, respectively. The estimated probabilities for all of the tested concentrations are reported in Table A.1.

4.2. Calibration of the parameters of the locomotion model

The calibration of the model is organized in two steps. First, we examine the segments of the trajectories where fish swim away from the wall to estimate the parameters η, σ_v, α, σ_w, γ, and λ using the maximum likelihood estimation method. Then, we calibrate the wall avoidance function in Eq. (8), using the data corresponding to the fish swimming near the wall via a weighted least squares method.

4.2.1. Discrete approximation of the locomotion equations

Following Zienkiewicz et al. (2015b), we consider that the fish is away from the wall if the distance from the circular arena is greater than 6 cm (or 2 BL). Next, we obtain discrete approximations of Eq. (4) and Eq. (7). In particular, we use the Euler-Maruyama method (Higham, 2001) to discretize Eq. (4), which upon replacing Eq. (6), yields

$$\begin{aligned} v_{\mathrm{S}}((k+1)\Delta) = (1+\eta\Delta)v_{\mathrm{S}}(k\Delta) \\ - \frac{\Delta}{\mathrm{BL}\,\mathrm{std}_\omega}|\omega_{\mathrm{S}}(k\Delta)|v_{\mathrm{S}}^2(k\Delta) \\ + \sigma_v\sqrt{\Delta}v_{\mathrm{S}}(k\Delta)\varepsilon_v(k), \end{aligned} \tag{10}$$

where $\varepsilon_v(k)$ is a standard Gaussian random variable, with zero mean and unit variance. Similarly, the discrete approximation of the turn rate without wall avoidance function is given by (Mwaffo et al., 2015a)

$$\begin{aligned} \omega_{\mathrm{S}}((k+1)\Delta) = e^{-\alpha\Delta}\omega_{\mathrm{S}}(k\Delta) + \sqrt{s(k\Delta)}\varepsilon_{\omega_1}(k) \\ + \gamma\zeta(k)\varepsilon_{\omega_2}(k), \end{aligned} \tag{11a}$$

$$s(k\Delta) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha\Delta}), \tag{11b}$$

where $\varepsilon_{\omega_1}, \varepsilon_{\omega_2}$ are independent standard Gaussian random variables and $\zeta(k)$ is a Bernoulli random variable with probability $\lambda\Delta$. In Eqs. (10) and (11), we have used the same time-step as the experimental recordings Δ; however, one might also consider alternative discretizations of the system of continuous-time stochastic differential equations for fish swimming.

4.2.2. Maximum likelihood estimation

We introduce the two vectors of unknown parameters $\Theta_1 = [\eta, \widetilde{\sigma}_v]^\top$ and $\Theta_2 = [\alpha, \sigma, \gamma, \lambda]^\top$. We define $\widetilde{\sigma}_v = \kappa\sigma_v$, with κ being a positive known constant, to avoid singularities on the objective function of the optimization problem. Estimates of Θ_1 and Θ_2 can be obtained by solving two independent optimization problems formulated within two separate admissible sets of parameter values χ_1 and χ_2, respectively, which were selected from previous work (Mwaffo et al., 2015a, 2017b). These optimization problems take as input real time-series of the speed and turn rate, $v_{\mathrm{S}}(k\Delta)$ and $\omega_{\mathrm{S}}(k\Delta)$, to return the calibrated model parameters.

In particular, an estimate of Θ_1 can be obtained by solving

$$\widehat{\Theta}_1 = \underset{\Theta_1\in\chi_1\subset\mathbb{R}^2}{\arg\min} -\sum_{k=1}^{N^*}\log\ell_v(\Theta_1, v_{\mathrm{S}}(k\Delta), \omega_{\mathrm{S}}(k\Delta)), \tag{12a}$$

$$\text{subject to} \quad \sigma_v^2 < 2\eta. \tag{12b}$$

Here, $N^* < N$ is the number of time-steps where the fish is away from the wall of the arena in the experimental observations, and the function $\ell_v(\Theta_1, \omega_{\mathrm{S}}(k\Delta), v_{\mathrm{S}}(k\Delta))$ is the likelihood function obtained from the discrete model approximation in Eq. (10) as

$$\ell_v(\Theta, v_{\mathrm{S}}(k\Delta), \omega_{\mathrm{S}}(k\Delta)) = H(q(k\Delta), 0, \sqrt{\sigma_v^2\Delta}), \tag{13}$$

where $H(x, \mu, \sigma)$ is the Gaussian distribution at a value x with mean μ and variance σ^2, and we have introduced the aggregated quantity

$$q(k\Delta) = -\frac{\eta\Delta + 1}{\kappa} + \frac{v_{\mathrm{S}}(k\Delta)|\omega_{\mathrm{S}}(k\Delta)\Delta|}{\kappa \mathrm{BL}\,\mathrm{std}_{\omega}]} + \frac{v_{\mathrm{S}}((k+1)\Delta)}{\kappa v_{\mathrm{S}}(k\Delta)}. \quad (14)$$

The last term on the right-hand-side of Eq. (14) takes large values when the speed is near zero for $\kappa = 1$. Using values of κ larger than one allows for rescaling the denominator and avoiding numerical issues when solving the optimization problem. Heuristically, we found that setting $\kappa = 5$ is sufficient to avoid singularities in the whole dataset. The constraint $\sigma_v^2 < 2\eta$ in Eq. (12) guarantees the existence of a nontrivial solution of the stochastic logistic equation (Pasquali, 2001).

Similarly, the estimate of Θ_2 is given by

$$\widehat{\Theta}_2 = \underset{\Theta_2 \in \chi_2 \subset \mathbb{R}^4}{\arg\min} -\sum_{k=1}^{N^*} \log \ell_{\omega}(\Theta_2, \omega_{\mathrm{S}}(k\Delta)). \quad (15)$$

Here, the function $\ell_{\omega}(\Theta_2, \omega_{\mathrm{S}}(k\Delta))$ is obtained from Eq. (11) and is given by Mwaffo et al. (2015a)

$$\ell_{\omega}(\Theta, \omega_{\mathrm{S}}(k\Delta)) = \lambda\Delta H\left(z(k\Delta), 0, \sqrt{s(k\Delta) + \gamma^2}\right) + (1 - \lambda\Delta) H\left(z(k\Delta), 0, \sqrt{m(k\Delta)}\right), \quad (16)$$

where $z(k\Delta) = \omega_{\mathrm{S}}((k+1)\Delta) - s(k\Delta)$ and $s(k\Delta)$ is given in Eq. (11b). Working with real time-series of speed and turn rate, the optimization problems in Eqs. (12) and (15) are solved using the multistar solver of MATLAB for each fish; the results of the identification for each fish are presented in Table A.2.

4.2.3. Wall function calibration

Following Gautrais et al. (2009) and Zienkiewicz et al. (2015b), we first compute the wall-corrected turn rate from the real time-series of the turn rate

$$\omega_{\mathrm{c}}(k\Delta) = \begin{cases} |\omega_{\mathrm{S}}(k\Delta)|, & \text{if } \mathrm{sgn}(\omega_{\mathrm{S}}(k\Delta)) = \mathrm{sgn}(\varphi(k\Delta)) \\ -|\omega_{\mathrm{S}}(k\Delta)|, & \text{otherwise} \end{cases}, \quad (17)$$

where $\varphi(k\Delta)$ is the orientation of the fish with respect to the wall, as sketched in Figure 5. Next, we plot the projected distance $d(k\Delta)$ versus the corrected turn rate $\omega_{\mathrm{c}}(k\Delta)$ in Eq. (17). We only consider positive values of the corrected turn rate, as we are interested in turn rates associated with wall avoidance.

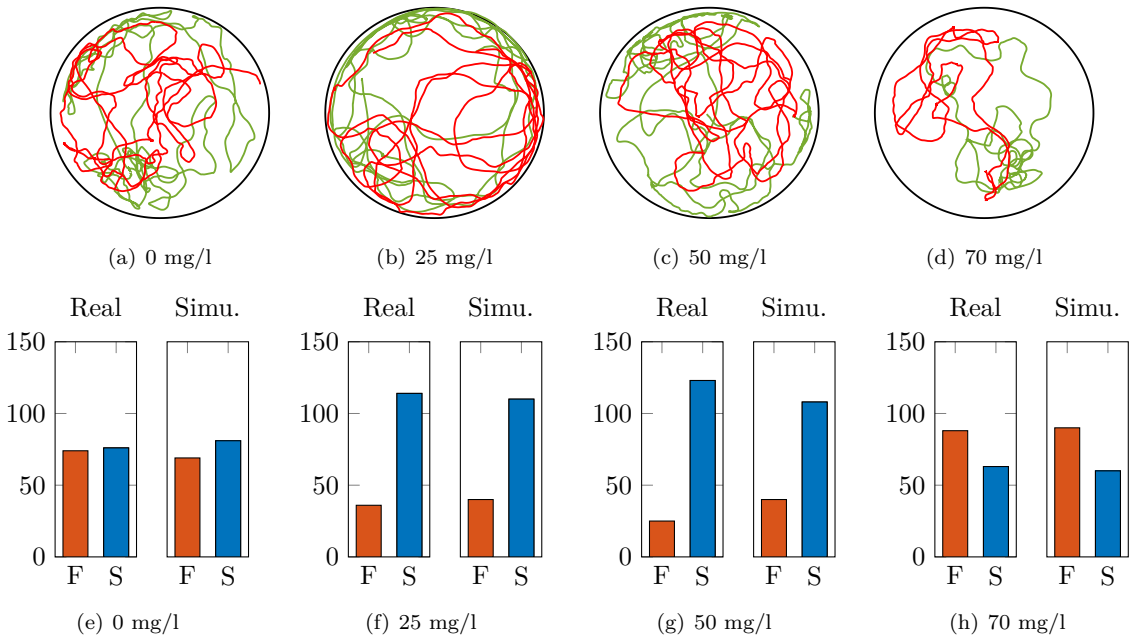
We implement a local regression smoothing to filter out noise. In particular, we use a robust non-parametric locally weighted least squares (RLOESS) function in MATLAB with a 10% span. The smoothed signal is utilized as an input to fit a parametric exponential function that provides the estimates of the parameters a and b in Eq. (8).

In the case when the number of available data points is limited or the noise is excessively high, the regression performance is low and the estimation of the wall parameters becomes inadequate, as shown by Zienkiewicz et al. (2015b). We excluded these instances from the analysis and we estimated common values of these parameters by averaging across all the identifications to obtain $a = 11.68\,\mathrm{rad/s^2}$ and $b = 0.19\,\mathrm{cm^{-1}}$.

4.3. Model simulation and validation

Once model parameters are calibrated from experimental data, one can simulate the behavioral response of zebrafish to caffeine exposure by integrating in time the governing equations of the modeling framework laid down in Section 3 for the desired experimental time ($T_{\mathrm{exp}} = 300\,\mathrm{s}$). From the Markov chain in Eq. (2), we simulate switching between freezing and swimming. In each swimming interval, at least $T = 2\,\mathrm{s}$-long, we integrate the locomotion model, composed of the logistic equation in Eq. (4) and the JPTW in Eq. (7). Integration of these continuous-time equations is performed using the Euler-Maruyama scheme, similar to the calibration process described in Section 4.2. However, to obtain an accurate solution of the stochastic differential equations, we discretize them with a finer time-step $\Delta_{\mathrm{E-M}} = 0.0063\,\mathrm{s}$, equal to one quarter of the experimental resolution. Simulation data are downsampled when needed for comparison with experimental observations.

To illustrate the accuracy of the model and the success of the identification procedure, we compare experimental and numerical results for the four animals presented in Figure 1. Simulations are initialized from experimental observations, such that the the initial condition of the Markov chain in Eq. (2) are chosen from the corresponding experiment, along with the initial position and heading in Eq. (1).



(a) 0 mg/l (b) 25 mg/l (c) 50 mg/l (d) 70 mg/l

(e) 0 mg/l (f) 25 mg/l (g) 50 mg/l (h) 70 mg/l

Figure 8: Comparison between empirical data and numerical simulations with respect to fish trajectories and time spent freezing or swimming for each fish trajectory from Figure 1. Top panels: overlay of real (green) and simulated (red) fish trajectories. Note that instances of live animals touching the tank wall are due to the fact that the experimental arena was not a perfect circle. Bottom panels: histograms of the number of occurrences of swimming (S) and freezing states (F) for experiments and simulations. Note that each pair of histograms sums to 150, whereby we examine 150 segments that are 2 s-long.

We start the comparison by examining the trajectories of the animals and their freezing response in Figure 8. For all of the four concentrations, we register very good agreement between model predictions and experimental observations, in terms of the similarity of the trajectories of the animals and their time budgeting freezing or swimming.

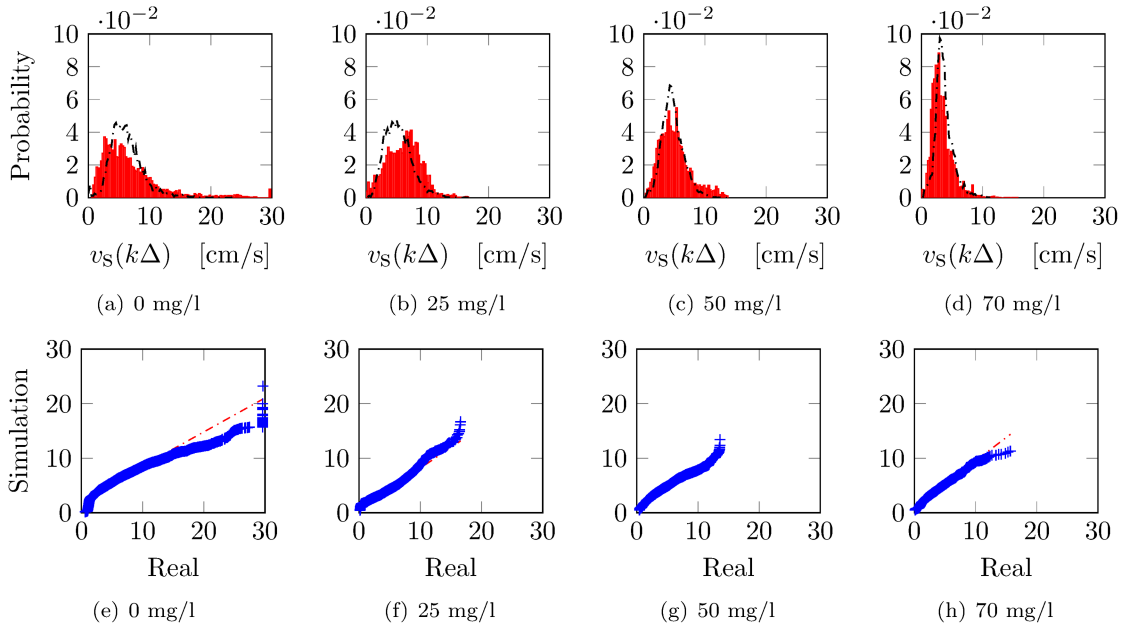
Next, we compare experimental results and computer simulations in terms of speed and turn rate, as shown in Figures 9 and 10. Therein, we report the distributions of the speed and turn rate from experiments and simulations, along with quantile-quantile (QQ) plots assessing the accuracy of the simulations against experimental observations. In agreement with our expectations, the logistic model in Eq. (4) is able to capture the wide speed variability of zebrafish. Similar to our previous work, we confirm that the JPTW model in Eq. (7) identifies the whole range of turn rates, including large values that are associated with C- and U-turns (Mwaffo et al., 2015a).

5. In-silico analysis of the effect of caffeine administration

5.1. *Model parameters as functions of caffeine treatment*

Here, we investigate the effect of caffeine concentration on the values of all of the model parameters: p_{SF}, p_{FS}, η, σ_v, α, σ_w, γ, and λ. We carried out a one-way ANOVA choosing caffeine concentration as the independent variable. Results are synoptically presented in Figure 11.

We determined that the caffeine concentration has a strong effect on the transition probability from swimming to freezing p_{SF} ($F(3, 28) = 4.75$, $P = 0.008$). Post-hoc analysis revealed that fish exposed to 70 mg/l caffeine concentration have a higher chance of transitioning from swimming to freezing when compared to the control subjects ($P = 0.003$), as well individuals treated at $C = 25$ mg/l ($P = 0.001$) and $C = 50$ mg/l ($P = 0.011$). With respect to the probability of transitioning from freezing to swimming p_{FS}, we did not record a significant effect of the concentration ($F(3, 28) = 1.50$, $P = 0.212$), albeit larger values might be seen for intermediate caffeine concentrations.



(a) 0 mg/l (b) 25 mg/l (c) 50 mg/l (d) 70 mg/l

(e) 0 mg/l (f) 25 mg/l (g) 50 mg/l (h) 70 mg/l

Figure 9: Comparison between empirical data and numerical simulations with respect to the speed, captured through the stochastic logistic equation for each fish trajectory from Figure 1. Top panels are the histograms of the speed distributions, comparing experiments (red) with simulations (black dashed) and bottom panels are QQ plots. For all histograms the bin size was set to 0.25 cm/s. The dashed-dotted red line in each QQ plot joins the connects the first and third quartiles of the data.

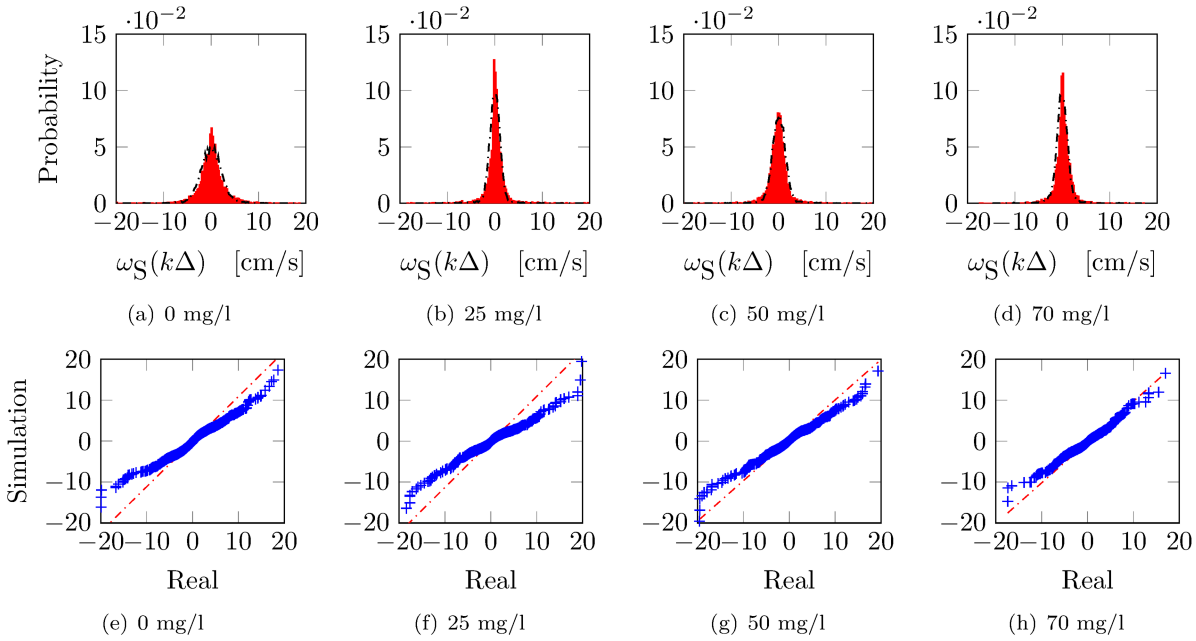
Examining the parameters of the locomotion model associated with the logistic equation for the speed evolution, both η ($F_{3,28} = 0.57$, $P = 0.637$) and σ_v ($F_{3,28} = 1.43$, $P = 0.253$) were indistinguishable across different caffeine treatments. These findings point at a secondary role of caffeine treatment on the logistic model for the speed evolution, such that during swimming, one could adopt the same stochastic differential equation to describe caffeine-treated subjects.

The analysis of the locomotion parameters associated with the JPTW for the evolution of the turn rate presents a richer picture on caffeine administration. Although all of the four parameters $(\alpha, \sigma_w, \gamma, \lambda)$ exhibited an inverted U-shape dependence on the caffeine concentration, a significant effect was registered only with respect to the parameters associated with the jump term in the JPTW. More specifically, we failed to identify a significant effect on either the relaxation rate α ($F_{3,28} = 2.06$, $P = 0.128$) or the turn rate variability σ_w ($F_{3,28} = 2.16$, $P = 0.114$), but we recorded a robust dependence of caffeine concentration on the intensity of jumps γ ($F_{3,28} = 4.56$, $P = 0.010$) and their frequency λ ($F_{3,28} = 3.80$, $P = 0.021$)

Post-hoc comparisons revealed that the jump intensity at the intermediate concentration $C = 50\,\text{mg/l}$ is higher than the control condition ($P = 0.005$) and $C = 70\,\text{mg/l}$ ($P = 0.002$). In addition, the jump frequency of individuals treated at the lowest caffeine concentration $C = 25\,\text{mg/l}$ was higher than those treated at the highest concentration $C = 70\,\text{mg/l}$ ($P = 0.002$).

Overall, our results indicate that caffeine treatment has a robust effect on the probability to initiate freezing, as well as on the frequency of sudden turns and their intensity. For all these three parameters, we registered a nonmonotonic dependence on caffeine concentration, such that: (i) the locomotion of individuals treated at low caffeine levels is characterized by more frequent turns of high intensity, and (ii) subjects treated at high caffeine levels are more prone to exhibit a freezing response and their swimming style features less sudden turns, which also have a smaller intensity.

The effect of caffeine administration on freezing is in line with the pharmacodynamics of this compound (Fredholm et al., 1999) and the literature on anxiety-related response of zebrafish (Stewart et al., 2010). In fact, high caffeine concentrations should elicit an anxiogenic response in the animals, due to antagonist compounds of adenosine A_1 re-



(a) 0 mg/l (b) 25 mg/l (c) 50 mg/l (d) 70 mg/l

(e) 0 mg/l (f) 25 mg/l (g) 50 mg/l (h) 70 mg/l

Figure 10: Comparison between empirical data and numerical simulations with respect to the turn rate, captured by the JPTW model for each fish trajectory from Figure 1. Top panels are the histograms of the turn rate distribution, comparing experiments (red) with simulations (black dashed) and bottom panels are QQ plots. For all histograms the bin size was set to 0.25 rad/s. The dashed-dotted red line in each QQ plot joins the connects the first and third quartiles of the data.

ceptors (Maximino et al., 2011). While heightened erratic behavior is often proposed as a measure of increased anxiety in zebrafish (Cachat et al., 2010), the concurrent pharmacological modulation of the freezing response favors the hypothesis of hyperlocomotion (Maximino et al., 2010a) in response to activity at adenosine A_{2A} receptors (Karcz-Kubicha et al., 2003).

5.2. Predicting zebrafish behavior in response to caffeine treatment

Finally, to demonstrate the predictive value of the proposed data-driven modeling framework, we conducted an in-silico experiment aimed at replicating the dependence of the behavioral metrics described in Section 2 on caffeine treatment. In particular, we performed ten simulations of our model for each caffeine concentration, using Euler-Mayurama discretization at the refined time-step $\Delta_{\mathrm{E-M}} = 0.0063\,\mathrm{s}$.

The simulation parameters were based on the calibration summarized in Figure 11, which indicated that caffeine concentration has a clear effect only on the probability of transitioning from swimming to freezing p_{SF}, the intensity of jumps γ, and the frequency of jumps λ. Hence, in our ten simulations, we drew the values of these three parameters from Gaussian distributions whose means and standard deviations corresponds to the data in Figure 11. Given that none of the other model parameters displayed a significant variation, we took them as the average across all the four caffeine concentrations. The initial conditions x_0, y_0, θ_0, Γ_0, v_0, and ω_0 were chosen randomly in their respective intervals. The body length of the fish was kept at $\mathrm{BL} = 3\,\mathrm{cm}$ and a common value for the standard deviation of the turn rate $\mathrm{std}_\omega = 2.487\,\mathrm{rad/s}$ was taken for all fish to describe the effect of the turn rate on the speed in Eq. (6).

Using the synthetic dataset, we calculated the four metrics introduced in Section 2.3, namely, the time spent freezing t_{F}, distance travelled D, mean speed V_{S}, and mean absolute turn rate Ω_{MtS} as reported in Figure 12. In-silico experiments predicted an effect of acute caffeine concentration that is consistent with the real experimental results for all the chosen metrics. In particular, ANOVA confirmed the differences found in the experiment for the time spent freezing ($F_{4,40} = 11.02$, $P < 0.001$) and distance travelled ($F_{4,40} = 7.47$, $P < 0.001$). Similar to real experiments, ANOVA on synthetic data failed to detect an effect of caffeine on the

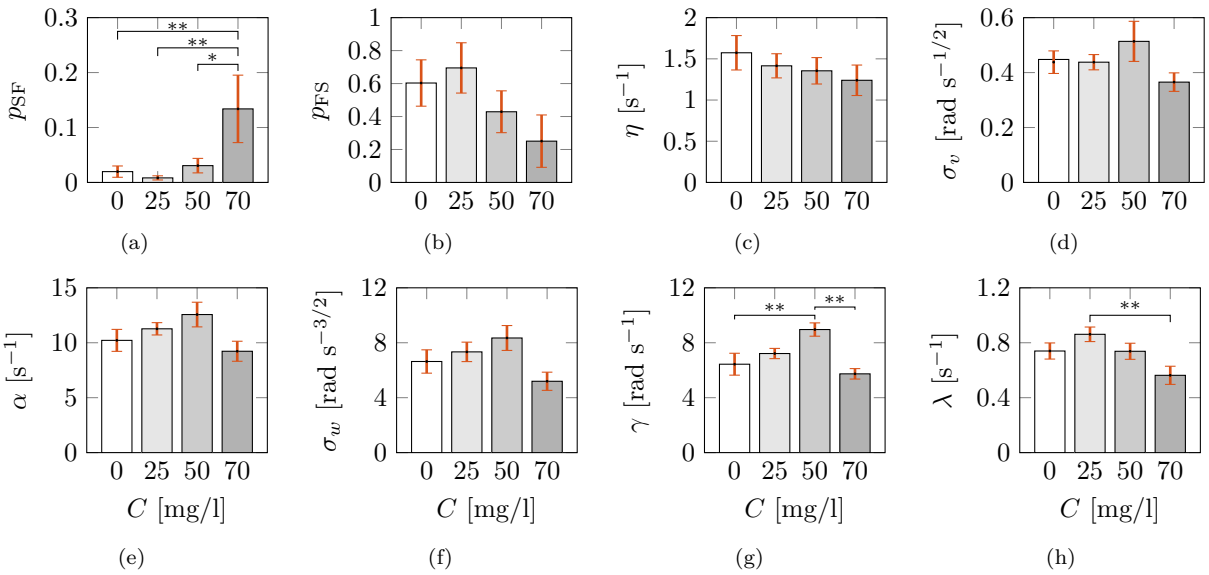


Figure 11: Calibrated parameters for different caffeine concentrations. Bars represent the average across all trials for each caffeine concentration for: (a) the probability of transitioning from swimming to freezing; (b) probability of transitioning from freezing to swimming; (c) logistic model parameter quantifying the linear rate of growth of the speed η; (d) logistic model parameter σ_v weighting the added noise on the speed evolution; (e) relaxation rate α of the turn rate measuring the rate at which the fish can resume straight swimming; (f) turn rate variability σ_w associated with the added noise in the turn rate evolution; (g) intensity of the jumps in the turn rate evolution γ, and (h) frequency of the jumps in the turn rate dynamics λ . Vertical red lines represent the SEM and symbols $**$ and $*$ indicate significant differences in post-hoc analysis, with P-values satisfying $P < 0.01$ and $P < 0.05$ respectively.

mean speed, but it revealed an effect on mean absolute turn rate ($F_{4,40} = 11.29$, $P < 0.001$) that was not found in real experiments. This was likely due to the limited inter-subject variability included in the in-silico experiment, where the entire logistic model and part of the JPTW were run with common parameters across the entire population.

6. Conclusions

Zebrafish exhibits a complex repertoire of locomotory patterns (Kalueff et al., 2013) and a comprehensive, all encompassing, data-driven model is far from being developed. Such a model could be central in supporting preclinical research with respect to three principles of replacement, reduction, and refinement.

In this work, we sought to contribute to this area of investigation by developing a data-driven modeling framework to examine anxiety-related behavioral response of zebrafish exposed to acute caffeine administration. Grounded in the theory of Markov chains and stochastic differential equations, our model predicts the transitions between freezing and swimming as well as locomotion in a shallow water tank.

The model is constructed upon two temporal scales to simultaneously examine freezing response and locomotion. Along the slower time-scale, we introduce a two-state Markov chain that encapsulates transitions between swimming and freezing. Along the fast time-scale, we complement the JPTW model introduced in our previous work (Mwaffo et al., 2015a) to study the dynamics of the turn rate with a novel logistic model to capture the time-evolution of the speed. Such a logistic model offers a parsimonious and robust representation of the speed dynamics, based on only two parameters that should be calibrated from the experiments.

Working with real trajectory data of zebrafish exposed to acute caffeine administration (Neri et al., 2019), we calibrate all the model parameters toward a first computational framework to examine pharmacological manipulations in-silico. Our results suggest that caffeine concentration has a strong effect on the parameters of the Markov model describing zebrafish freezing, where we registered a tenfold increase in the probability of transitioning from swimming to freezing between control subjects

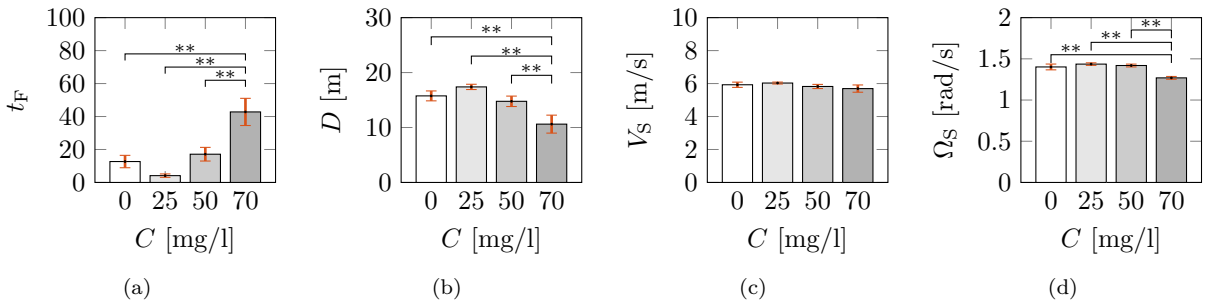


Figure 12: In-silico predictions of the effect of caffeine treatment on zebrafish behavior, scored through: (a) time spent freezing, (b) distance travelled, (c) mean speed, (d) mean absolute turn rate. Vertical red lines represent the SEM, and symbols $**$ and $*$ indicate significant differences in post-hoc analysis, with P-values satisfying $P < 0.01$ and $P < 0.05$, respectively.

and individuals treated at the highest caffeine concentration. The logistic model of the time-evolution of the speed is robust with respect to the level of caffeine treatment, supporting the use of the two-scale dichotomy on which the model is constructed. By evolving the speed only during the swimming episodes dictated by the underlying Markov chain, we identify model parameters that are common to all of the tested caffeine concentrations. The effect of caffeine treatment on zebrafish locomotion is encapsulated by the JPTW, whose jump intensity and frequency display the typical inverted U-shape with respect to caffeine treatment. The caffeine concentration seems to interfere with only three of the model parameters, which directly impinge on freezing and erratic activity – two of the typical behavioral measures of anxiety (Maximino et al., 2010b,a; Kalueff et al., 2013).

Our study is not free of limitations, which call for future work along a number of research directions. First, our model is limited to a shallow water tank that constraints animal locomotion in two dimensions, thereby preventing geotaxis (Maximino et al., 2010a,b; Kalueff et al., 2013). Extending our model to three dimensions is an ongoing research endeavor, which will result into a more complete prediction of anxiety-related phenotype. Second, we have focused on single psychoactive compound, as first step to demonstrate the possibility of modeling the effect of pharmacological manipulations. Future research should explore the effect of other substances, like citalopram, ethanol, and cocaine (López-Patiño et al., 2008; Egan et al., 2009; Maximino et al., 2011), toward a comprehensive toolbox to assist in-silico experimentation of anxiety-related response. Third, our model applies to animals swimming in isolation in an open-field environment, without external stimuli. Future work should seek to explore modeling of zebrafish groups, in which some of the individuals are pharmacologically-manipulated (Ladu et al., 2014; Neri et al., 2019) to determine how these manipulations modify social behavior, in terms of how untreated individuals appraise treated subjects and how treated subjects process information from untreated individuals. Along a similar line of inquiry, we also plan to explore zebrafish response to anxiety-evoking stimuli, such as the presence of predators (Maximino et al., 2010a), to widen the range of application of the proposed modeling framework.

Another avenue of future work should also seek to expand the number of states in the Markov chain to include other behavioral patterns, beyond swimming and freezing, such as thrashing (Kalueff et al., 2013) that could play a significant role on the quantification of the anxiety-related response of zebrafish. Accounting for thrashing would imply correcting the definition of freezing, such that we could isolate instances where the animal display locomotory patterns with large acceleration in the vicinity of a wall (Kopman et al., 2013).

Acknowledgements

This work was supported by the National Science Foundation under #Grant CMMI-1505832.

Appendix A. Calibrated parameters

Here, we detail the values of the parameters of the data-driven model calibrated from the experiments.

Table A.1: Transition probabilities of the Markov model for different caffeine concentrations. Note that the fish examined in Figures 1(a), 1(b), 1(c), and 1(d) correspond to trials 8 for $C = 0\,\mathrm{mg/l}$, 7 for $C = 25\,\mathrm{mg/l}$, 6 for $C = 50\,\mathrm{mg/l}$, and 4 for $C = 70\,\mathrm{mg/l}$, respectively.

	0 [mg/l]		25 [mg/l]		50 [mg/l]		70 [mg/l]	
Trial	p_{FS}	p_{SF}	p_{FS}	p_{SF}	p_{FS}	p_{SF}	p_{FS}	p_{SF}
1	0	1	0	1	0.019	0.208	0.007	0.006
2	0.052	0.047	0.008	0.048	0.003	0.333	0	1
3	0.012	0.004	0	1	0.008	0.4	0.082	0.054
4	0.006	0.20	0.003	0.333	0.004	0.013	0.071	0.011
5	0	1	0	1	0.051	0.102	0.022	0.166
6	0	1	0.012	0.050	0.020	0.107	0.121	0.014
7	0.020	0.013	0.013	0.027	0.007	0.250	-	-
8	0.003	0.500	0	1	-	-	-	-
9	0	1	0	1	-	-	-	-
10	0.003	0.200	-	-	-	-	-	-

References

Psychiatric Association, A., 2013. Diagnostic and statistical manual of mental disorders (DSM-5®). 5th ed., American Psychiatric Pub.

Bowman, G.R., Pande, V.S., Noé, F., 2013. An introduction to Markov state models and their application to long timescale molecular simulation. volume 797. Springer Science & Business Media.

Brémaud, P., 2013. Markov chains: Gibbs fields, Monte Carlo simulation, and queues. volume 31. Springer Science & Business Media.

Butail, S., Mwaffo, V., Porfiri, M., 2016. Model-free information-theoretic approach to infer leadership in pairs of zebrafish. Physical Review E 93, 042411.

Cachat, J., Canavello, P., Elegante, M., Bartels, B., Hart, P., Bergner, C., Egan, R., Duncan, A., Tien, D., Chung, A., et al., 2010. Modeling withdrawal syndrome in zebrafish. Behavioural Brain Research 208, 371–376.

Calovi, D.S., Litchinko, A., Lecheval, V., Lopez, U., Escudero, A.P., Chaté, H., Sire, C., Theraulaz, G., 2018. Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors. PLoS Computational Biology 14, e1005933.

Calovi, D.S., Lopez, U., Schuhmacher, P., Chaté, H., Sire, C., Theraulaz, G., 2015. Collective response to perturbations in a data-driven fish school model. Journal of The Royal Society Interface 12, 20141362.

Collignon, B., Séguret, A., Halloy, J., 2016. A stochastic vision-based model inspired by zebrafish collective behaviour in heterogeneous environments. Royal Society Open Science 3, 150473.

Danos, N., Lauder, G.V., 2007. The ontogeny of fin function during routine turns in zebrafish danio rerio. Journal of Experimental Biology 210, 3374–3386.

Egan, R.J., Bergner, C.L., Hart, P.C., Cachat, J.M., Canavello, P.R., Elegante, M.F., Elkhayat, S.I., Bartels, B.K., Tien, A.K., Tien, D.H., et al., 2009. Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish. Behavioural Brain Research 205, 38–44.

Fredholm, B.B., Bättig, K., Holmén, J., Nehlig, A., Zvartau, E.E., 1999. Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. Pharmacological Reviews 51, 83–133.

Gard, T.C., 1988. Introduction to Stochastic Differential Equations. Pure and Applied Mathematics, Marcel Dekker Inc.

Gautrais, J., Ginelli, F., Fournier, R., Blanco, S., Soria, M., Chaté, H., Theraulaz, G., 2012. Deciphering interactions in moving animal groups. PloS Computational Biology 8, e1002678.

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. Journal of Mathematical Biology 58, 429–445.

Higham, D.J., 2001. An algorithmic introduction to numerical simulation of stochastic differential equations. SIAM review 43, 525–546.

Hill, A.J., Teraoka, H., Heideman, W., Peterson, R.E., 2005. Zebrafish as a model vertebrate for investigating chemical toxicity. Toxicological Sciences 86, 6–19.

Jayne, K., See, A., 2019. Behavioral Research on Captive Animals: Scientific and Ethical Concerns. Brill. chapter 21. pp. 517 – 547.

Kalueff, A.V., Gebhardt, M., Stewart, A.M., Cachat, J.M., Brimmer, M., Chawla, J.S., Craddock, C., Kyzar, E.J., Roth, A., Landsman, S., et al., 2013. Towards a comprehensive catalog of zebrafish behavior 1.0 and beyond. Zebrafish 10, 70–86.

Karcz-Kubicha, M., Antoniou, K., Terasmaa, A., Quarta, D., Solinas, M., Justinova, Z., Pezzola, A., Reggio, R., Müller, C.E., Fuxe, K., et al., 2003. Involvement of adenosine A1 and A2A receptors in the motor effects of caffeine after its acute and chronic administration. Neuropsychopharmacology 28, 1281.

Khan, K.M., Collier, A.D., Meshalkina, D.A., Kysil, E.V., Khatsko, S.L., Kolesnikova, T., Morzherin, Y.Y., Warnick, J.E., Kalueff, A.V., Echevarria, D.J., 2017. Zebrafish models in neuropsychopharmacology and cns drug discovery. British Journal of Pharmacology 174, 1925–1944.

Klebaner, F.C., 2012. Introduction to stochastic calculus with applications. World Scientific Publishing Company.

Kopman, V., Laut, J., Polverino, G., Porfiri, M., 2013.

Table A.2: Calibrated parameters of the locomotion model, composed of the logistic equation in Eq. (4) and the JPTW in Eq. (7), for different caffeine concentrations. Note that the fish examined in Figures 1(a), 1(b), 1(c), and 1(d) correspond to trials 8 for $C = 0\,\text{mg/l}$, 7 for $C = 25\,\text{mg/l}$, 6 for $C = 50\,\text{mg/l}$, and 4 for $C = 70\,\text{mg/l}$, respectively.

C	Trial	η [s^{-1}]	σ_v [$\text{s}^{-1/2}$]	α [s^{-1}]	σ_w [rad $\text{s}^{-3/2}$]	γ [rad s^{-1}]	λ [s^{-1}]
0 mg/l	1	1.141	0.326	11.375	5.856	6.873	0.656
	2	1.739	0.352	6.720	5.119	4.797	0.566
	3	0.992	0.383	7.681	3.648	4.251	0.630
	4	1.441	0.630	14.988	7.296	8.476	0.982
	5	0.906	0.205	5.426	1.403	1.307	0.669
	6	1.518	0.462	11.803	10.034	8.825	0.831
	7	3.314	0.537	7.509	7.364	5.248	0.411
	8	1.482	0.483	11.660	9.105	7.290	0.772
	9	1.583	0.519	11.550	7.301	7.794	0.955
	10	1.417	0.577	13.456	9.223	9.508	0.923
25 mg/l	1	1.4837	0.579	10.822	12.188	8.890	0.922
	2	1.265	0.501	12.838	6.581	7.632	1.105
	3	1.859	0.399	10.058	8.836	7.504	0.768
	4	1.040	0.318	9.377	7.398	5.898	0.780
	5	1.230	0.370	10.189	6.880	6.814	0.757
	6	0.941	0.451	11.326	7.215	8.945	1.036
	7	0.946	0.428	14.860	4.879	6.564	1.007
	8	1.699	0.373	10.402	5.848	6.081	0.703
	9	2.191	0.519	11.549	6.195	6.604	0.673
50 mg/l	1	0.998	0.361	10.797	11.899	9.815	0.680
	2	1.934	0.850	16.294	6.222	10.519	0.449
	3	1.196	0.437	9.630	8.399	7.508	0.784
	4	1.737	0.579	11.291	10.722	10.234	0.961
	5	0.921	0.331	10.602	9.172	8.659	0.746
	6	0.957	0.369	12.135	5.901	7.276	0.729
	7	1.394	0.666	17.216	6.113	8.721	0.814
70 mg/l	1	2.106	0.353	7.910	6.094	5.857	0.404
	2	1.427	0.415	10.285	5.658	5.856	0.694
	3	0.815	0.349	11.705	4.492	6.048	0.787
	4	0.987	0.490	5.775	7.710	6.288	0.512
	5	1.197	0.339	11.123	3.985	6.470	0.376
	6	0.834	0.243	8.558	3.233	3.899	0.602

Closed-loop control of zebrafish response using a bioinspired robotic-fish in a preference test. Journal of the Royal Society Interface 10, 20120540.

Ladu, F., Butail, S., Macrí, S., Porfiri, M., 2014. Sociality modulates the effects of ethanol in zebra fish. Alcoholism: Clinical and Experimental Research 38, 2096–2104.

Lo, A.W., 1988. Maximum likelihood estimation of generalized itô processes with discretely sampled data. Econometric Theory 4, 231–247.

López-Patiño, M.A., Yu, L., Cabral, H., Zhdanova, I.V., 2008. Anxiogenic effects of cocaine withdrawal in zebrafish. Physiology & Behavior 93, 160–171.

Maximino, C., de Brito, T.M., da Silva Batista, A.W., Herculano, A.M., Morato, S., Gouveia Jr, A., 2010a. Measuring anxiety in zebrafish: a critical review. Behavioural Brain Research 214, 157–171.

Maximino, C., De Brito, T.M., de Mattos Dias, C.A.G., Gouveia Jr, A., Morato, S., 2010b. Scototaxis as anxiety-like behavior in fish. Nature Protocols 5, 209–216.

Maximino, C., da Silva, A.W.B., Gouveia Jr, A., Herculano, A.M., 2011. Pharmacological analysis of zebrafish (*Danio rerio*) scototaxis. Progress in Neuro-Psychopharmacology and Biological Psychiatry 35, 624–631.

Meyers, J.R., 2018. Zebrafish: Development of a vertebrate model organism. Current Protocols Essential Laboratory Techniques 16, e19.

Mwaffo, V., Anderson, R.P., Butail, S., Porfiri, M., 2015a. A jump persistent turning walker to model zebrafish locomotion. Journal of The Royal Society Interface 12, 20140884.

Mwaffo, V., Butail, S., Di Bernardo, M., Porfiri, M., 2015b. Measuring zebrafish turning rate. Zebrafish 12, 250–254.

Mwaffo, V., Butail, S., Porfiri, M., 2017a. Analysis of pairwise interactions in a maximum likelihood sense to identify leaders in a group. Frontiers in Robotics and AI 4, 35.

Mwaffo, V., Butail, S., Porfiri, M., 2017b. In-silico experiments of zebrafish behaviour: modeling swimming in three dimensions. Scientific Reports 7, 39877.

Mwaffo, V., Korneyeva, V., Porfiri, M., 2017c. simUfish: An Interactive Application to Teach k-12 Students About Zebrafish Behavior. Zebrafish 14, 477–488.

Mwaffo, V., Porfiri, M., 2015. Turning rate dynamics of zebrafish exposed to ethanol. International Journal of Bifurcation and Chaos 25, 1540006.

Navidi, W.C., 2008. Statistics for engineers and scientists. McGraw-Hill Higher Education New York, NY, USA.

Neri, D., Ruberto, T., Mwaffo, V., Bartolini, T., Porfiri, M., 2019. Social environment modulates anxiogenic effects of caffeine in zebrafish. Behavioural Pharmacology 30, 45–58.

Norton, W., Bally-Cuif, L., 2010. Adult zebrafish as a model organism for behavioural genetics. BMC Neuroscience 11, 90.

Nusslein-Volhard, C., Dahm, R., 2002. Zebrafish. Oxford University Press.

Pasquali, S., 2001. The stochastic logistic equation: stationary solutions and their stability. Rendiconti del Seminario Matematico della Università di Padova 106, 165–183.

de Paula Lima, J., Farah, A., 2019. Chapter 14: Caffeine consumption, in: Coffee: Consumption and Health Implications. The Royal Society of Chemistry, pp. 313–339.

Rosa, L.V., Ardais, A.P., Costa, F.V., Fontana, B.D., Quadros, V.A., Porciúncula, L.O., Rosemberg, D.B., 2018. Different effects of caffeine on behavioral neurophenotypes of two zebrafish populations. Pharmacology Biochemistry and Behavior 165, 1–8.

Russell, W.M.S., Burch, R.L., Hume, C.W., 1959. The principles of humane experimental technique. volume 238. Methuen London.

Santos, L.C., Ruiz-Oliveira, J., Silva, P.F., Luchiari, A.C., 2017. Caffeine dose–response relationship and behavioral screening in zebrafish. The Question of Caffeine , 87–105.

Sloman, K.A., Bouyoucos, I.A., Brooks, E.J., Sneddon, L.U., 2019. Ethical Considerations in Fish Research. Journal of Fish Biology 94, 556–577.

Spence, R., Gerlach, G., Lawrence, C., Smith, C., 2008. The behaviour and ecology of the zebrafish, danio rerio. Biological Reviews 83, 13–34.

Stewart, A., Kadri, F., DiLeo, J., Min Chung, K., Cachat, J., Goodspeed, J., Suciu, C., Roy, S., Gaikwad, S., Wong, K., et al., 2010. The developing utility of zebrafish in modeling neurobehavioral disorders. International Journal of Comparative Psychology 23, 104–120.

Stewart, A.M., Braubach, O., Spitsbergen, J., Gerlai, R., Kalueff, A.V., 2014. Zebrafish models for translational neuroscience research: from tank to bedside. Trends in Neurosciences 37, 264–278.

de Wit, C.C., Siciliano, B., Bastin, G., 2012. Theory of robot control. Springer Science & Business Media.

Zienkiewicz, A., Barton, D., Porfiri, M., Di Bernardo, M., 2015a. Leadership emergence in a data-driven model of zebrafish shoals with speed modulation. The European Physical Journal Special Topics 224, 3343–3360.

Zienkiewicz, A., Barton, D.A., Porfiri, M., Di Bernardo, M., 2015b. Data-driven stochastic modelling of zebrafish locomotion. Journal of Mathematical Biology 71, 1081–1105.

Zienkiewicz, A.K., Ladu, F., Barton, D.A., Porfiri, M., Di Bernardo, M., 2018. Data-driven modelling of social forces and collective behaviour in zebrafish. Journal of Theoretical Biology 443, 39–51.