Science of the Total Environment 636 (2018) 767–774

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Escherichia coli's water load affects zebrafish (*Danio rerio*) behavior

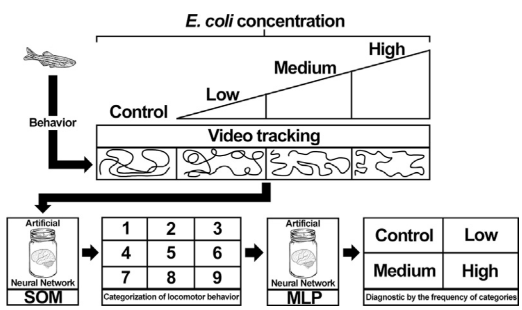
João Amorim [a,b,*], Miguel Fernandes [a], Isabel Abreu [b], Fernando Tavares [a,c], Luis Oliva-Teles [a,b]

[a] Faculdade de Ciências da Universidade do Porto, Departamento de Biologia, Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal
[b] Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR) da Universidade do Porto, Terminal de Cruzeiros do Porto de Leixões, Avenida General Norton de Matos, s/n, 4450-208 Matosinhos, Portugal
[c] Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), InBIO Laboratório Associado, Universidade do Porto, Rua Padre Armando Quintas, no 7, 4485-661 Vairão, Portugal

HIGHLIGHTS

- Analysis of zebrafish's locomotor behavior was used for the detection of *E. coli*.
- Video tracking and two artificial neural networks were used for the diagnostic.
- *E. coli* water loads of 600, 1800 and 5000 CFU/100 mL altered zebrafish's behavior.
- The system's accuracy, sensitivity and specificity values ranged from 89 to 100%.
- The system can thus be potentially applied to infer microbiological contamination.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 23 February 2018
Received in revised form 11 April 2018
Accepted 24 April 2018
Available online xxxx

Editor: Daniel Wunderlin

Keywords:
Biological early warning system
Microbiological contamination
Water quality
Artificial neural networks
Biomonitoring
Biosensor

ABSTRACT

Traditional physico-chemical sensors are becoming an obsolete tool for environmental quality assessment. Biomonitoring techniques, such as biological early warning systems present the advantage of being sensitivity, fast, non-invasive and ecologically relevant. In this work, we applied a video tracking system, developed with zebrafish (*Danio rerio*), to detect microbiological contamination in water. Using the fishs' behavior response, the system was able to detect the presence of a non-pathogenic environmental strain of *Escherichia coli*, at three different levels of contamination: 600, 1800 and 5000 CFU/100 mL (colony forming units/100 mL). Data was collected during 50 min of exposure and analyzed with the artificial neural networks Self-organizing Map and Multi-layer Perceptron. The behavior of exposed fish was more erratic, with pronounced and rapid changes on movement direction and with significant less exploratory activity. The accuracy, sensitivity and specificity values regarding the detection capability (distinction between presence or absence of contamination) ranged from 89 to 100%. Regarding the classification capability (distinction between experimental conditions), the values ranged from 67 to 89%. This research may be a valuable contribution to improve water monitoring and management strategies, by taking as reference the effects on biosensors, without a biased anthropocentric perspective.

1. Introduction

The continuous increase of pollution sources, resulting from the uncontrolled development of multiple human activities (Gavrilescu et al., 2015), makes water quality management fundamental in today's society (Bae et al., 2012). Traditionally, this task is based on physico-

* Corresponding author at: Faculdade de Ciências da Universidade do Porto, Departamento de Biologia, Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal.
E-mail addresses: jamorim@ciimar.up.pt, (J. Amorim), ftavares@fc.up.pt, (F. Tavares), loteles@fc.up.pt (L. Oliva-Teles).

https://doi.org/10.1016/j.scitotenv.2018.04.316

chemical sensors that dismiss the microbiotic component, require expensive maintenance costs and only reflect specific and momentary variations within aquatic ecosystems. Furthermore, these sensors are unable to provide information regarding synergistic and antagonistic effects of chemical mixtures, as well as its ecological impacts in the ecosystems (Bae and Park, 2014). Unlike traditional methods, biomonitoring has been appointed as a reliable, cost-effective and faster tool to detect early environmental dysfunctions, which might be ecologically relevant (Liu et al., 2011).

Recently, many biological early warning systems (BEWS) have been developed to evaluate the organisms' physiological and behavioral responses to water quality (Wu and Yang, 2011). These systems have a great potential for detecting a wide range of contaminants, in a short period of time (Bae and Park, 2014). BEWS' applications are vast and include monitoring of drinking water intakes, water distribution systems, wastewater effluents, effluents from contamination remediation sites, and river basin monitoring programs (Allan et al., 2006). Furthermore, numerous organisms, such as bacteria, algae, invertebrate and fish, can be applied in this type of methodology (Bae and Park, 2014). Indeed, fish were the first organisms used in BEWSs, in the early 1970s (Døving et al., 1977), and they continue to be an excellent choice as water quality indicators, mainly due to their remarkable chemoreception, able to sense the proprieties of their surrounding environment (Døving et al., 1977). Video tracking systems are a type of BEWS, using video cameras and digital image processing to track and record the organisms' movement. This technique is particularly useful to determine spatial variables such as distance travelled, average speed of movement, among others, that cannot be consistently evaluated using other approaches (Spink et al., 2001). Commonly used as an ecotoxicological model, zebrafish (*Danio rerio*) is a very active species, moving almost continuously during light period, being therefore effectively used for video tracking (Grillitsch et al., 1999).

Behavioral responses to contamination can be assessed by BEWS, since they result from the integration of biochemical and physiological processes (Hellou, 2011), and can be related to the conditions to which the organisms are exposed. In fact, they have been used for decades in environmental monitoring (Melvin and Wilson, 2013). It has been acknowledging that behavior can be more sensitive than survival as a toxicological endpoint (Gerhardt, 2007), with a sensitivity 10 to 1000 times higher than the conventional LC_{50} (Hellou et al., 2008). Moreover, lethality represents the ultimate stress test and may not be protective (Beitinger, 1990), nor the best predictor of survival or damage in organisms (Little and Brewer, 2001).

With the development of new methods that allow the study of animal behavior (i.e. video tracking), it has become essential to develop tools to effectively interpret the tremendous amount of obtained data. Artificial neural networks (ANNs) are one of the most suitable methods for this task. Within ANNs, the Self-organizing Map (SOM) has been used in several ecotoxicological studies with aquatic organisms, including *Daphnia magna*, *Chironomus samoensis*, *Oryzias latipes* and *Danio rerio*, in order to classify behavior patterns and organize them into different groups (Kim et al., 2006; Liu et al., 2011; Liu et al., 2012; Park et al., 2005). SOM works with unsupervised learning processes and can interpret, organize and classify the input data without the need of an external agent. Also, SOM is particularly useful in behavioral studies, because it allows for a non-linear projection of the dense and complex data, into a two-dimensional space. Within this topological organization, the network detects similarities and tries to establish increasingly complex relationships, being able to detect patterns in the inputs (Qiao and Han, 2010). Another ANN, used in several studies to detect behavioral changes caused by toxicants in organisms, including non-biting midges, medaka and zebrafish (Ji et al., 2006; Kim and Cha, 2002; Kwak et al., 2002; Lee et al., 2015) is the Multi-layer Perceptron (MLP). This ANN contains three or more layers: an input, an output and one or more hidden layers (Poulton, 2001). The signals are propagated from the input layer, across the hidden ones, to the output layer via the network connections. Through backpropagation (supervised learning process), the resulting error, produced by comparing the obtained network output with the desired one, is propagated through the network backwards, during which successive adjustments are made to the synaptic weights (Haykin, 2009; Yu and Tan, 2014).

Thermotolerant coliform bacteria in general, and *Escherichia coli* in particular, are frequently used as indicators for fecal contamination of recreational water and assessment of its sanitary condition, according to national regulations aligned with international directives, namely the EU Water Framework Directive (2000/60/EC of 23rd October) and the WHO Guidelines for Safe Recreational Water Environments (2009). The quantitative microbiological criteria proposed by these regulations address essentially human health risks, dismissing its putative biohazards effects on water ecosystems. The development and optimization of a monitoring system, capable of evaluating microbiological disruptions in water ecosystems, beyond the detection threshold of the standard methods used to determine human health biohazards, is therefore urgently needed.

The main goal of the present work was to use a video tracking-based BEWS, to detect the effects of a non-pathogenic *E. coli* river isolate on zebrafish, below the currently used concentrations, to evaluate the water sanitary risk for human health. This microbiological monitoring system would allow a meaningful biohazard assessment of water ecosystems in real time, and contribute to review the currently used microbiological criteria, biased for human health risk assessment.

2. Material and methods

2.1. Test organism

Healthy wild-type adult zebrafish (*Danio rerio*) were purchased from ORNI-EX, Lda (Arcozelo, Vila Nova de Gaia, Portugal). The fish were acclimated for one month under laboratory conditions, prior to any testing. They were fed once a day with Tetra Min Bio Active Formula and kept in a 12/12 h (light/dark) photoperiod. Stock fish were maintained in aerated tanks, supplied with recirculated water (28 ± 0.5 °C). Healthy adult zebrafish (2.5 ± 0.2 cm body length, mean ± SD; 0.9 ± 0.1 g body weight, mean ± SD) were randomly selected from a stock population, to be used in the trials. This study complied with the European Union Council (2010/63/EU, of 22nd September) and the National Guidelines (Decreto-Lei 113/2013 of 7th August) on animal experiments, as well as with the 3R principles' guidelines (refinement, re-placement, and reduction of animal use).

2.2. Escherichia coli concentrations

A previously characterized *E. coli* isolate, from Rio Febros (Norwest of Portugal) (Cabral and Marques, 2006), was cultured in Nutrient Broth (Difco) and incubated overnight at 28 °C. *E. coli* density was determined by spectrophotometry (OD 600 nm) and added to 1.5 L of recirculating tap water at 28 °C, at three density levels (low, medium and high) corresponding to 600, 1800 and 5000 CFU/100 mL (colony forming units/100 mL), respectively. These values represented an *E. coli* cell load below the range recorded for environmental fresh water (Ratajczak et al., 2010) and, both below and above the recommended maximum level of 2000 CFU/100 mL for thermotolerant coliforms, according to the Portuguese legislation for recreational waters (Decreto-Lei 236/1998 of 1st August).

2.3. Recording areas

The behavior observation system (recording areas) used in this study was similar to the one reported in Amorim et al. (2017) and consisted of: three main transparent aquaria 35 × 20 × 15 cm (length, width and height), each one internally divided into four equal sections (arenas); three video cameras (Flow Electronics 540L IR cameras (model CACO0008)) with super high color resolution, placed above each one of the main aquaria; a computer (Intel® Pentium® Dual CPU (2:00 E2180@2.00 GHz, 1.87 GB RAM)); and white LED lights to improve the quality of the footage taken. The recording areas were surrounded by plates of expanded cork (2 cm thick), to eliminate or minimize external noise and vibrations. Each aquarium was in a water bath, equipped with an electrical resistance to maintain water temperature at 28 °C and a semi-transparent acryl-plate base (2 mm thick) to provide an even distribution of LED light intensity to the arenas. The aquaria outer glass was frosted, to maximize the footage's contrast, while white plastic boards were glued to the inner glass to eliminate visual contact and completely isolate the fish from each other. The depth of water in the arenas was reduced to a minimum value (10 cm–1.5 L) to further reduce the vertical movement without affecting the overall level of fish activity. The culture medium consisted of tap water, aged for at least 24 h, in a 50 L aquarium with recirculating water at 28 °C. Given that zebrafish's activity and movement is higher in the horizontal than in the vertical plane (Grillitsch et al., 1999), an horizontal view was chosen for the recordings.

2.4. Experimental plan

Each of the three recording areas was comprised of four individual arenas, with one fish in each one, representing the four experimental conditions, namely the three *E. coli* densities (low – C1, medium – C2 and high – C3) and a control (C0) to which no bacteria was added. Three independent trials were carried out to allow for a significant number of replicates and statistical validation. A total of thirty-six organisms were used, nine per experimental condition. The number of replicates was adjusted to support the capabilities of the system, presenting a compromise between statistical significance and ideal methodological approach. Thus, this number was selected based on the recording cameras' maximum resolution (number of organisms it can reliably detect and track simultaneously), the amount of data collected during the recordings (six frames per second (FPS), over a fifty-minute period) and the amount of analyzed variables (nine). The main objective of the experimental plan was to determine the effect of different *E. coli* water densities, on zebrafish's swimming behavior. The trials started when the fish were transferred to the individual arenas, and each of the three different cell densities of *E. coli*, was added to its respective arena. Declorinazed aged tap-water was added to the controls. The fish were given 10 min to acclimatize and then their behavior was recorded for 40 min, always from 2 PM to 2:40 PM, within the zebrafish's period of increased activity (Pather and Gerlai, 2009). At the end of the assays, fish were isolated in a new aquarium (both the exposed and the control) and maintained under the same conditions as described above. No harmful effects were caused by the exposure, and the organisms were fully recovered.

2.5. Video processing

The recorded videos were processed using multiple plugins of ImageJ, a public domain, Java-based image processing program (version 1.40g, Rasband, W.S., U. S. National Institutes of Health, Maryland, USA). The main tool used was the MultiTracker plugin (version 2001.7.17, Jeffrey Kuhn, University of Texas, Austin, USA). In general, the purpose of the video processing was to individualize each fish and transform its location in the arena in coordinates that could then be analyzed. The visual reference of the fish was considered a “dot” and its coordinates were calculated for each frame of the footage. These coordinates were then exported into an X and Y spreadsheet, to proceed with the statistical analysis. The cameras' resolution allowed the system to use increments of 1 mm for the movement analysis and the calculation of the movement descriptor parameters. The refresh rate was set at six FPS, meaning a sample rate of the fish's position at every 1/6 of a second.

2.6. Statistical analysis

The zebrafish's locomotor behavior characterization was made by SOM. The objective was to define different classes of behavior (based on organisms' locomotion characteristics), by grouping data similarities. The analysis then assigned the behavior to one of nine classes, whose number was established in preliminary assays. Nine movement descriptors, most of them already characterized (Chen et al., 2011; Chon et al., 2005; Williams et al., 2012), were used as input variables: mean value of the X coordinate (mm); mean value of the Y coordinate (mm); mean linear velocity (mm/s); mean linear acceleration (mm/s^2); mean angular velocity (degrees/s); mean angular acceleration (degrees/s^2); absolute meander (degrees/mm), which is a curvature measure in the trajectory; dispersion (mm), measured as the square root of the product between the standard deviation of the X and Y coordinates; and the angle (degrees) between the segments, formed by the fish's coordinates in three successive frames. The setting parameters of the neural network, available in the software herein used (Statistica version 10 (StatSoft, 2011)) were adjusted through error analysis of the training and test subsets, being defined as follows: topological dimensions – 3 × 3 (nine behavior classes); start learning rate – 0.01; end learning rate – 0.0001; start neighborhoods – 2; end neighborhoods – 0; random sampling method of data subsets – 34% “training”, 33% “test”, 33% “validation”; training cycle – 1000; network randomization – normal. The remaining parameters were set as they were, by default. The use of independent (not overlapping) data subsets, although rarely applied to ecological studies, was fundamental for the correct evaluation of the developed model's accuracy (diagnostic performance), particularly when considering the overfitting problem, typically observed in the application of ANNs (Özesmi et al., 2006). The training and test subsets were used by the algorithm to learn, adjust and calibrate the model (unsupervised learning). The validation subset was used to validate the results and achieve the diagnostic in conjunction with MLP (see below). The values of the different variables herein studied were not instantaneous, but rather the average of 64 consecutive frames (approximately 10 s, at six FPS), repeated in regular intervals in order to avoid overlapping values, thereby decreasing the results' variance, making the analysis more sensitive. Since the fish's behavior was not constant during trials, occupying one of the nine behavior categories at different times (10 s intervals), each experimental condition was not represented by the categories themselves, but rather by their frequency.

In order to verify if the exposure to *E. coli*, at different cell densities, affected the frequency of the behavior classes, the final diagnosis was made using MLP. In this analysis, each object (fish) was classified into one of four exposure conditions (C0, C1, C2 or C3). This type of ANN uses a supervised learning process called backpropagation. This process works by changing connection weights between the multiple layers, after each single piece of data in the “Training” subset is processed, based on the amount of error found in the output layer compared to the expected result. The subsets used were the same defined previously by SOM, once again avoiding overlapping. The MLP used as input values the frequencies of the nine behavior classes, defined by SOM, and as output targets the four exposure conditions. This analysis used 2/3 of the data for training (1/3) and required supervision (1/3). Nonetheless, the final diagnosis/classification (validation) was completely independent. This method was unbiased, since the remaining 1/3 of the used data were not subjective to manipulation by any external agent and were used exclusively by the MPL to achieve the validation diagnosis.

Finally, a descriptive analysis of the nine behavioral variables was performed, according to the different experimental conditions, by Analysis of Variance (ANOVA) and Tukey Honest Significance Test, to allow a greater understanding of the behavior, under different conditions.

Table 1
Variation of the behavioral categories' frequency, measured by the difference between the expected and observed values, and its association with the four exposure conditions (C0, C1, C2 and C3). Dark shade values represent strong association with the condition (|x| ≥ 50%), light shade values represent moderate association with the condition (20% ≤ |x| < 50%), and no-shade values represent weak association with the condition (0 ≤ |x| < 20%).

Behavioral category	Observed minus expected frequencies (%)				Exposure conditions association					
	C0	C1	C2	C3	Positively related		Negatively related			Type
1	81	−21	−1	−62	C0	–	C3	C1	C2	A
2	67	3	−5	−67	C0	C1	C3	C2	–	A
3	63	−32	2	−37	C0	C2	C3	C1	–	A
4	−87	66	49	−26	C1	C2	C0	C3	–	B1
5	−87	−9	49	48	C2	C3	C0	C1	–	B2
6	−62	−12	−42	121	C3	–	C0	C2	C1	B3
7	−38	−36	−27	103	C3	–	C0	C1	C2	B3
8	−59	37	−49	75	C3	C1	C0	C2	–	B3
9	−5	−29	29	3	C2	C3	C1	C0	–	C

3. Results and discussion

3.1. Descriptive analysis

Correspondence analysis was performed in order to evaluate the association between the nine behavioral categories, automatically generated by SOM, and the different experimental conditions. This analysis was made without discriminating the three data subsets. The results were highly significant (total inertia = 0.27382; $chi^2 = 1224.0$; df = 24; p = 0.0000), being possible to evaluate the association between the behavior categories and the exposure to the bacteria water loads (C0, C1, C2, and C3). This scrutiny was made by determining the frequency variation of the behavioral categories, for each experimental condition (Table 1 and Fig. 1).

According to the intensity (strong, moderate or weak) and nature (positive or negative) of association, between the nine behavioral categories and the four experimental conditions (C0, C1, C2 and C3), it was possible to establish five association types, designated by: A, B1, B2, B3 and C (Table 1 and Fig. 1). The A and B association types represent an antagonistic relationship. Type A refers to a strong positive association with the control condition (C0) and a strong to moderate negative association with the highest exposure concentration (C3), while type B represents a strong to moderate positive association with at least one of the exposure conditions with bacteria (C1, C2 and C3) and a strong to moderate negative association with C0. On the contrary, type C is characterized by a null association with C0 and C3. Instead, this association type reveals a moderate positive association with the intermediate exposure condition (C2) and a moderate negative association with the lowest exposure condition (C1). The type B behavior categories were further divided into three subtypes (B1, B2 and B3), according to the strongest positive association with the corresponding exposure condition, i.e. B1 with C1, B2 with C2 and B3 with C3.

ANOVA and Post-Hoc were used to further characterize the behavioral classes, particularly each one of the nine movement descriptors

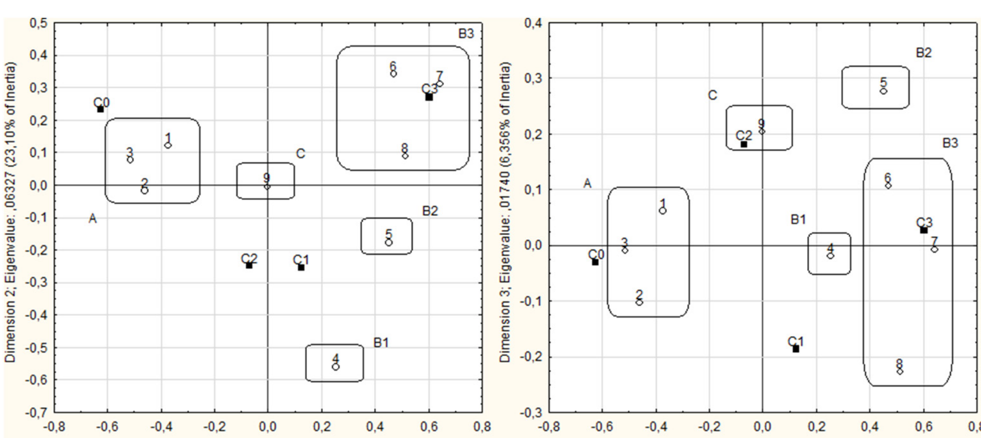


Fig. 1. Correspondence analysis plots of the 1–2, and 1–3 dimensions. The categories' (1 through 9) relative position and proximity to the exposure condition points (C0, C1, C2 and C3) indicates their level of association with said conditions. Association types A, B1, B2, B3 and C are highlighted by rectangles.

used. This was performed with the average values, obtained in each category for the different experimental conditions, replicas and repetitions, without discriminating the three data subsets. All movement descriptors herein presented had highly significant differences ($p \leq 0.000000$, except for the "Angle (degrees)": $p = 0.003897$).

Through the Post-Hoc test, it is clear that categories 1, 2 and 3 have some common characteristics, since they share the majority of the positions in the homogenous groups, all of them positively associated with C0 (light shade values). These were the categories that presented a strong association to the control condition (type A in the correspondence analysis - Table 1). These categories presented low to average values of angular velocity (120–195 degrees/s), angular acceleration (659–1176 degrees/s^2), meander (2–4 °/mm), and linear velocity (50–65 mm/s). This behavior could be described as representing moderate swimming activity, with low intensity changes of direction. Also, high values of dispersion (21–25 mm) and average central values of X (71–73 mm) and Y (38–51 mm), indicated a wide and uniform circulation through space, which has been defined as normal in zebrafish (Collignon et al., 2016). In contrast, the values of at least one or more variables, of the remaining behavior categories (all types B and C association in the correspondence analysis - Table 1), were outside the range described above for type A association. Fish exposed to the bacteria showed a typical stress response, when compared to zebrafish's normal behavior. The X and Y coordinates had average values outside the ones observed in type A, which roughly defined the central zone of the arenas, in categories 4, 5 and 8, associated with types C1, C2 and C3, respectively. Particularly, X = 116 mm in category 4, Y = 28 mm in category 5, and Y = 70 mm in category 8. These results seem to indicate that the bacteria presence increased the fish escape response, as the animals were more likely to move away from the center of the arena and look for more remote areas. This behavior is considered an adaptive response and is used as a strategy to facilitate the search for shelter, escape routes and protection from predators (Ahmad and Richardson, 2013). Furthermore, this is often observed when animals are exposed to stress as, for instance, the adjustment to new environments (Sharma et al., 2009). Indeed, a significant increase of the distance to arenas' center has also been described for killifish (*Fundulus heteroclitus*) exposed to tricane methanesulfonae (Kane et al., 2004). The dispersion values herein recorded, allowed for a distinct separation of types A, B and C (Table 2). With the exception of category 5, all the other categories negatively associated with the control (association type B) and, association type C category 9 (null association with the control), presented mean values of dispersion significantly lower than the categories positively associated with the control (association type A). Category 6 (association type B3), which is positively associated with the highest contamination level (C3), presented the lowest average dispersion value. This category is also characterized by extremely low average linear velocity (4 mm/s) and extremely high meander values (59 degrees/mm). Concerning these variables, only two behavior categories presented values outside the range of those observed for association type A. In fact, category 4 (association type B1), positively associated with C1, has a high meander value (33 degrees/mm), and category 7 (association type B3), positively associated with C3, reveals the highest value for linear velocity (106 mm/s). It is interesting to note that association type B3 categories 6 and 7 display antagonistic behaviors regarding linear velocity. In fact, the fish are almost motionless (4 mm/s) in category 6, while they are extremely agitated in category 7 (106 mm/s). These contrasting behaviors (freeze and escape) were associated to the same (and highest) *E. coli* density (C3). These two distinct responses may be associated with the animals' alert mechanisms, present at early stages of exposure (freeze response) and the subsequent avoidance behavior (escape response). The angular velocity was shown to be the main behavior feature of category 4 (association type B1), with the highest average value recorded (328 degrees/s), followed by categories 7 and 8 (association type B3), with angular velocities of 280 and 272 degrees/s, respectively. These were significantly higher than the values recorded for the

Table 2
Average values of the nine movement descriptors (mean value of the X coordinate (mm); mean value of the Y coordinate (mm); mean linear velocity (mm/s); mean linear acceleration (mm/s^2); mean angular velocity (degrees/s); mean angular acceleration (degrees/s^2); absolute meander (degrees/mm); dispersion (mm); and the angle (degrees) between the segments formed by the fish's coordinates in three successive frames), in the nine behavior categories defined by SOM. The position in the different homogeneous groups was given by Tukey Unequal N HSD test. Dark shade values represent homogeneous groups negatively associated with C0, and light shade values represent homogeneous groups positively associated with C0.

Behavioral category	Type*	X (mm)		Y (mm)		Velocity (mm/s)		Acceleration (mm/s²)		Angular velocity (degrees/s)		Angular acceleration (degrees/s²)		Meander (degrees/mm)		Dispersion (mm)		Angle (degrees)	
		Average	Group	Average	Group	Average	Group	Average	Group	Average	Group	Average	Group	Average	Group	Average	Group	Average	Group
1	A	71	1	38	2	63	4 5	2.5	1 2 3	152	1 2	882	2	2	1	26	6	−1.2	1 2
2	A	73	1	44	2 3	50	2 3	2.3	1 2	120	1	659	1	3	1	21	5	4.6	2
3	A	73	1	51	4	65	4 5	4.1	3 4	195	3 4	1176	3	4	1	25	6	2.0	1 2
4	B1	116	2	46	3 4	38	2	2.9	1 2 3	328	7	1643	5	33	2	11	2	0.6	1 2
5	B2	83	1	28	1	70	5	3.3	2 3	220	4	1378	4	4	1	20	4 5	−5.0	1
6	B3	72	1	46	3 4	4	1	0.6	1	229	4 5	582	1	59	3	2	1	0.1	1 2
7	B3	81	1	45	2 3 4	106	6	5.7	4	280	6	1735	5	7	1	18	3 4	0.8	1 2
8	B3	70	1	70	5	60	3 4 5	3.6	2 3	272	5 6	1733	5	6	1	12	2	3.2	2
9	C	74	1	41	2 3	55	3 4	2.7	1 2 3	178	2 3	1025	2 3	9	1	17	3	−2.2	1 2

* Association type in Table 1.

behavior categories 1, 2 and 3 of association type A. Finally, categories 4, 5, 7 and 8 (association type B) presented values of angular acceleration well above the ones registered for the C0 condition, ranging from 1378 to 1735 degrees/s^2. Linear acceleration (mm/s^2) and angle (degrees) were not described, since they did not present discriminatory capacity. All homogeneous groups defined by these two parameters had behavior categories associated with the control and other two or three experimental conditions. In general, the behavior of fish exposed to *E. coli* could be described as more erratic, with pronounced and rapid changes on movement direction (higher values of angular velocity and angular acceleration), as well as with significant less exploratory activity (lower values of dispersion). These results are in agreement with previous studies with fish exposed to toxicants or in stressful situations (Egan et al., 2009; Amorim et al., 2017).

Two studies, both using bacterial fish pathogens, have addressed the effect of bacterial water contamination on fish's behavior. One of these studies also reported erratic swimming, as well as other unspecific clinical signs, such as anorexia, anemia, exophthalmia and high mortality on tilapia (*Oreochromis niloticus*) infected with *Francisella* sp. (Soto et al., 2009). The second study, which used MLP to assess fish behavior, reported stiff movement and frequent stops in zebrafish, at an initial stage of exposure to *Edwardsiella tarda* (Lee et al., 2015). Still, the exposed fish did not show erratic movement, nor pronounced and rapid changes on movement direction. Instead, they showed minimum turning rate and meander levels, even at early infection stages, which maybe explained by a severe infection with *E. tarda*, that lead to tail necrosis, abdominal distention, petechial hemorrhaging, and eventually the death of some animals during trials (Lee et al., 2015). Contrary to the studies mentioned above, which aimed to examine changes in fish's behavior during infection stages, our research addressed the effect of *E. coli* for the purpose of microbiological water load determination.

3.2. Diagnosis

Multi-layer Perceptron was able to classify each of the nine individual fish utilized in each experimental conditions (C0, C1, C2, and C3), a total of thirty-six animals (Table 3).

This data classification was instrumental to determine the system's capability to correctly detect behavioral changes, after exposure to different bacteria water loads (C0, C1, C2, and C3). Concerning the nine fish that were not exposed to bacteria (C0), MLP was able to correctly classify eight of those cases as not exposed to *E. coli*. For the nine fish exposed to a low bacterial density (C1), corresponding to 600 CFU/100 mL, MLP was able to detect a behavior pattern, consistent with the bacteria presence, in eight cases, seven of those correctly diagnosed as low bacterial load. Within the experimental condition C2 (medium level of exposure to *E. coli*, i.e. 1800 CFU/100 mL), the system was capable of correctly identify a behavior pattern for eight out of nine tested fish. C2 was, however, the experimental condition for which the MLP revealed higher classification difficulty, with only five fish being correctly differentiated from the other experimental conditions (C0, C1 and C3). Finally, for the nine fish of experimental condition C3, corresponding to a high bacterial density of 5000 CFU/100 mL, MLP was able to detect the presence of *E. coli* in all of them, and to correctly classify the bacterial load in six cases.

3.3. Diagnostic performance

The accuracy, sensitivity, specificity, false positives and negatives, and positive and negative predictive values (PPV or NPV, respectively) (Table 4), of BEWS method, were calculated based on the validation subset data. These parameters are required for the validation of any diagnostic tool and important to demonstrate the model's success or failure.

Regarding performance (Table 4), the system showed a very good diagnostics capability, and achieved robust results regarding the distinction between different experimental conditions (C0, C1, C2 and C3). Concerning the ability to detect bacterial water contamination, an accuracy of 89% was achieved for the low and medium *E. coli* densities (C1 and C2), while the highest *E. coli* density (C3) displayed an accuracy of 94%. Accuracy is an indicator of the overall quality of the method, because it corresponds to the proportion of correct diagnostics over the entire set of results. Sensitivity values were similar to those observed for accuracy, with a sensitivity of 89% for the two lowest *E. coli* densities (C1 and C2) and a sensitivity of 100% for the highest density (C3), since the MLP system was always able to sense the bacteria in the water. The sensitivity measures the method's ability to correctly identify the toxicant presence. The specificity was the same for all experimental conditions with a value of 89%. This parameter measures the ability to correctly identify the absence of a toxicant. Regarding MLP's ability to differentiate between different experimental conditions, the results showed to be promising, with an accuracy of 78%, a sensitivity of 67% and a specificity of 89%.

For the development, implementation and validation of new diagnostic tools, the two most important parameters that need to be considered are the PPV and NPV. These values are particularly informative to evaluate the overall performance of the system. The PPV refers to the probability that a positive screening (given by the method) is in fact a true positive event. On the contrary, the NPV refers to the probability that a negative screening is in fact a true negative case. According to recommendations issued by the European Centre for the Validation of Alternative Methods (Genschow et al., 2002), the predictive capacity of the proposed system, in regards to the detection results, was classified as "excellent" (≥85%), with values ranging from 89% to 90%, for PPV,

Table 3
Diagnostic results achieved by classification with MLP, using only the validation subset data. All thirty-six fish and four experimental conditions are represented simultaneously. The results are shown as absolute values. Correct diagnoses are represented in dark grey shading.

	Diagnostic			
		Positive		
Experimental condition	**Negative**	**Low concentration**	**Medium concentration**	**High concentration**
Control	8	1	0	0
Low concentration	1	7	1	0
Medium concentration	1	2	5	1
High concentration	0	1	2	6

Table 4
Quality of the diagnostic results, achieved by classification with MLP using only the validation subset: accuracy, sensitivity, specificity, false positives and negatives, and positive and negative predictive values. The first three results (for low, medium and high concentration) are related to the detection capacity of the system (distinction between presence or absence of contamination), while the last one is related to the model's classification capacity (distinction between the four experimental conditions).

	Diagnostic quality			
	Detection			Classification
	Low concentration	Medium concentration	High concentration	
Accuracy[a]	89	89	94	78
Sensitivity	89	89	100	67
Specificity	89	89	89	89
Positive predictive value	89	89	90	69
Negative predictive value	89	89	100	80

[a] The accuracy values are balanced, given the same weight to the events with and without toxicant.

and 89% to 100%, for NPV. The predictive capacity of the system, when concerning the classification results, was lower (69% for PPV and 80% for NPV), but still in the "acceptable" (≥65%>) and "good" (≥75%>) levels, respectively. From the obtained results, it is clear that the system presented very good results in regards to *E. coli*'s detection. The classification capability was understandingly lower, since the complexity of the diagnostic increased from two possible outcomes (presence or absence of bacteria), to four possible final results (C0, C1, C2 and C3).

4. Conclusion

In this work we propose a BEWS to assess water quality beyond the microbiological criteria that has been establish by international and national regulatory bodies. These criteria mainly take into account public human health concerns, therefore neglecting the noxious effect on other ecosystem organisms, some of which more sensitive than humans to biotic and abiotic disruptions. By recording fish's behavior disturbances as a response to *E. coli*, a bioindicator commonly used to determine fecal contamination in water, at three distinct cell densities (two considered acceptable: C1 and C2, and one above the maximum limit defined by Portuguese legislation for recreational water: C3), we were able to determine fish behavioral categories for each experimental condition. Although, several studies have been published using fish as a biosensor to detect the presence of numerous toxicants, to the best of our knowledge this is the first research addressing the effect of water bacterial contamination on fish's behavior. Contrary to the certified methods of bacterial detection to evaluate water quality, which are culture dependent and require at least 24 to 48 h, the BEWS approach described in this study may allow a real time water quality monitoring, unbiased from anthropocentric use, by taking as reference autochthonous or model organisms. In this context, and having in mind the increasing concern regarding water policies, this research may be a valuable contribution to improve water monitoring and management strategies. Efforts to implement the BEWS outside the described laboratory controlled conditions, are absolutely necessary to evaluate how the environmental factors affect BEWS performance.

Competing interests

The authors do not have any competing interests to declare.

Acknowledgements

The authors would like to thank the EU and FCT/UEFISCDI/FORMAS for funding, in the frame of the collaborative international consortium REWATER financed under the ERA-NET Cofund WaterWorks2015 Call. This ERA-NET is an integral part of the 2016 Joint Activities developed by the Water Challenges for a Changing World Joint Programme Initiative (Water JPI).

References

Ahmad, F., Richardson, M., 2013. Exploratory behaviour in the open field test adapted for larval zebrafish: impact of environmental complexity. Behav. Process. 92, 88–98.
Allan, I., Vrana, B., Greenwood, R., Mills, G., Roig, B., Gonzalez, C., 2006. A "toolbox" for biological and chemical monitoring requirements for the European Union's Water Framework Directive. Talanta 69, 302–322.
Amorim, J., Fernandes, M., Vasconcelos, V., Teles, L., 2017. Evaluation of the sensitivity spectrum of a video tracking system with zebrafish (*Danio rerio*) exposed to five different toxicants. Environ. Sci. Pollut. Res. 24, 16086–16096.
Bae, M., Park, Y., 2014. Biological early warning system based on the responses of aquatic organisms to disturbances: a review. Sci. Total Environ. 466–467, 635–649.
Bae, M., Kim, J., Park, Y., 2012. Evaluation of changes in effluent quality from industrial complexes on the Korean nationwide scale using a self-organizing map. Int. J. Environ. Res. Public Health 9, 1182–1200.
Beitinger, T., 1990. Behavioral reactions for the assessment of stress in fishes. J. Great Lakes Res. 16, 495–528.
Cabral, J., Marques, C., 2006. Faecal coliform bacteria in Febros river (northwest Portugal): temporal variation, correlation with water parameters, and species identification. Environ. Monit. Assess. 118, 21–36.
Chen, T., Wang, Y., Wu, Y., 2011. Developmental exposures to ethanol or dimethylsulfoxide at low concentrations alter locomotor activity in larval zebrafish: implications for behavioral toxicity bioassays. Aquat. Toxicol. 102, 162–166.
Chon, T., Chung, N., Kwak, I., Kim, J., Koh, S., Lee, S., Leem, J., Cha, E., 2005. Movement behaviour of Medaka (*Oryzias latipes*) in response to sublethal treatments of diazinon and cholinesterase activity in semi-natural conditions. Environ. Monit. Assess. 101, 1–21.
Collignon, B., Séguret, A., Halloy, J., 2016. A stochastic vision-based model inspired by zebrafish collective behaviour in heterogeneous environments. Royal Soc. Open Sci. 3.
Decreto-Lei 113/2013, 2013. do Ministério da Agricultura do Mar, do Ambiente e do Ordenamento do Território de 7 de Agosto de 2013 relativo à proteção dos animais utilizados para fins científicos.
Decreto-Lei 236/1998, 1998. do Ministério do Ambiente de 1 de Agosto de 1998 relativo ao estabelecimento de normas, critérios e objectivos de qualidade com a finalidade de proteger o meio aquático e melhorar a qualidade das águas em função dos seus principais usos.
Directive 2000/60/EC, 2000. of the European Parliament and of the Council of 23 October 2000 on establishing a framework for the Community action in the field of water policy.
Directive 2010/63/EU, 2010. of the European Parlament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes L 276/33.
Døving, K., Dubois-Dauphin, M., Holley, A., Jourdan, F., 1977. Functional anatomy of the olfactory organ of fish and the ciliary mechanism of water transport. Acta Zool. 58, 245–255.
Egan, R., Bergner, C., Hart, P., Cachat, J., Canavello, P., Elegante, M., Elkhayat, S., Bartels, B., Tien, A., Tien, D., Mohnot, S., Beeson, E., Glasgow, E., Amri, H., Zukowska, Z., Kalueff, A., 2009. Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish. Behav. Brain Res. 205, 38–44.
Gavrilescu, M., Demnerová, K., Aamand, J., Agathos, S., Fava, F., 2015. Emerging pollutants in the environment: present and future challenges in biomonitoring, ecological risks and bioremediation. New Biotechnol. 32, 147–156.
Genschow, E., Spielmann, H., Scholz, G., Seiler, A., Brown, N., Piersma, A., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., et al., 2002. The ECVAM international validation study on in vitro embryotoxicity tests: results of the definitive phase and evaluation of prediction models. European Centre for the Validation of Alternative Methods. Altern. Lab. Anim 30, 151–176.
Gerhardt, A., 2007. Aquatic behavioral ecotoxicology: prospects and limitations. Hum. Ecol. Risk Assess. 13, 481–491.
Grillitsch, B., Vogl, C., Wytek, R., 1999. Qualification of spontaneous undirected locomotor behavior of fish for sublethal toxicity testing. Part II. Variability of measurement parameters under toxicant-induced stress. Environ. Toxicol. Chem. 18, 2743–2750.
Haykin, S., 2009. Neural Networks and Learning Machines. vol. 3. Pearson Upper Saddle River, NJ, USA.
Hellou, J., 2011. Behavioural ecotoxicology, an "early warning" signal to assess environmental quality. Environ. Sci. Pollut. Res. Int. 18, 1–11.
Hellou, J., Cheeseman, K., Desnoyers, E., Johnston, D., Jouvenelle, M.L., Leonard, J., Robertson, S., Walker, P., 2008. A non-lethal chemically based approach to investigate the quality of harbour sediments. Sci. Total Environ. 389, 178–187.
Ji, C., Lee, S., Kwak, I., Cha, E., Lee, S., 2006. Computational analysis of movement behaviors of medaka (*Oryzias latipes*) after the treatments of copper by using fractal dimension and artificial neural networks. WIT Trans. Biomed. Health 10, 93–107.
Kane, A., Salierno, J., Gipson, G., Molteno, T., Hunter, C., 2004. A video-based movement analysis system to quantify behavioral stress responses of fish. Water Res. 38, 3993–4001.
Kim, C., Cha, E., 2002. Applications of wavelet transform and artificial neural networks to pattern recognition for environmental monitoring. In: McKay, B., Slaney, J. (Eds.), AI 2002: Advances in Artificial Intelligence: 15th Australian Joint Conference on Artificial Intelligence Canberra, Australia, December 2–6, 2002 Proceedings. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 385–394.
Kim, C., Kwak, I., Cha, E., Chon, T., 2006. Implementation of wavelets and artificial neural networks to detection of toxic response behavior of chironomids (Chironomidae: Diptera) for water quality monitoring. Ecol. Model. 195, 61–71.

Kwak, I., Chon, T., Kang, H., Chung, N., Kim, J., Koh, S., Lee, S., Kim, Y., 2002. Pattern recognition of the movement tracks of medaka (*Oryzias latipes*) in response to sub-lethal treatments of an insecticide by using artificial neural networks. Environ. Pollut. (Barking, Essex: 1987) 120, 671–681.

Lee, S., Choe, Y., Chon, T., Kang, H., 2015. Analysis of zebrafish (*Danio rerio*) behavior in response to bacterial infection using a self-organizing map. BMC Vet. Res. 11, 1–12.

Little, E., Brewer, S., 2001. Neurobehavioural toxicity in fish. In: Benson, D.S.a.W.H. (Ed.), Target Organ Toxicity in Marine and Freshwater Teleosts New Perspectives: Toxicology and the Environment. Taylor and Francis, London and New York, pp. 139–174.

Liu, Y., Lee, S., Chon, T., 2011. Analysis of behavioral changes of zebrafish (*Danio rerio*) in response to formaldehyde using Self-organizing map and a hidden Markov model. Ecol. Model. 222, 2191–2201.

Liu, Y., Wu, F., Ji, C., Chon, T., 2012. Movement patterning of *Daphnia* magna treated with copper based on self-organizing map. Procedia Environ Sci 13, 994–1002.

Melvin, S., Wilson, S., 2013. The utility of behavioral studies for aquatic toxicology testing: a meta-analysis. Chemosphere 93, 2217–2223.

Özesmi, S., Tan, C., Özesmi, U., 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol. Model. 195, 83–93.

Park, Y., Chung, N., Choi, K., Cha, E., Lee, S., Chon, T., 2005. Computational characterization of behavioral response of medaka (*Oryzias latipes*) treated with diazinon. Aquat. Toxicol. 71, 215–228.

Pather, S., Gerlai, R., 2009. Shuttle box learning in zebrafish (*Danio rerio*). Behav. Brain Res. 196, 323–327.

Poulton, M., 2001. Multi-layer perceptrons and back-propagation learning. In: Poulton, M. (Ed.), Handbook of Geophysical Exploration: Seismic Exploration. Pergamon, pp. 27–53.

Qiao, J., Han, H., 2010. An adaptive fuzzy neural network based on Self-Organizing Map (SOM). In: Matsopoulos, G.K. (Ed.), Self-Organizing Maps. InTech, Shanghai, pp. 1–14.

Ratajczak, M., Laroche, E., Berthe, T., Clermont, O., Pawlak, B., Denamur, E., Petit, F., 2010. Influence of hydrological conditions on the Escherichia coli population structure in the water of a creek on a rural watershed. BMC Microbiol. 10:222.

Sharma, S., Coombs, S., Patton, P., Perera, T., 2009. The function of wall-following behaviors in the Mexican blind cavefish and a sighted relative, the Mexican tetra (Astyanax). J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. 195, 225–240.

Soto, E., Hawke, J., Fernandez, D., Morales, J., 2009. *Francisella* sp., an emerging pathogen of tilapia, *Oreochromis niloticus* (L.), in Costa Rica. J. Fish Dis. 32, 713–722.

Spink, A., Tegelenbosch, R., Buma, M., Noldus, L., 2001. The EthoVision video tracking system: a tool for behavioral phenotyping of transgenic mice. Physiol. Behav. 73, 731–744.

WHO, 2009. Guidelines for Safe Recreational Water Environments. (Vol. 1 (2003), Vol. 2 (2006), and Addendum to Vol. 1).

Williams, L., Wong, K., Stewart, A., Suciu, C., Gaikwad, S., Wu, N., DiLeo, J., Grossman, L., Cachat, J., Hart, P., et al., 2012. Behavioral and physiological effects of RDX on adult zebrafish. Comp. Biochem. Physiol. C Toxicol. Pharmacol. 155, 33–38.

Wu, Y., Yang, H., 2011. Biological early warning system for prawn aquiculture. Procedia Environ Sci 10, 660–665.

Yu, S., Tan, K., 2014. Classification of lumbar ultrasound images with machine learning. In: Dick, G., Browne, W.N., Whigham, P., Zhang, M., Bui, L.T., Ishibuchi, H., Jin, Y., Li, X., Shi, Y., Singh, P., et al. (Eds.), Simulated Evolution and Learning: 10th International Conference, SEAL 2014, Dunedin, New Zealand, December 15–18, 2014 Proceedings. Springer International Publishing, Cham, pp. 287–298.