Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2025 (43)** : 1–11.
ISSN-0037-8712
doi:10.5269/bspm.76009

Optimization of Maggot Mass Rearing via Topological Statistics Approach *

Thiago de Melo†, Rodrigo Rosa, Jamil Viana Pereira, Washington Mio, Claudio José von Zuben and Alice Kimie Miwa Libardi

ABSTRACT: In this work, we present an investigation of mass rearing via topological data analysis and statistical methods. We study the optimization of mass production and develop methods for interpretation and visualization of contrasts in mass rearing results under different combinations of larval densities and availability of food resources. The approach considers receiver operating characteristic (ROC) curves to define distances between histograms. It is used to compare results under different conditions based on survival rates and properties of pupal weight and size distributions. The method is effective and practical for optimizing mass rearing and identification of cost-effective approaches.

Key Words: Applied Entomology, *Chrysomya megacephala*, Insect Mass Rearing, ROC Analysis, Topological Data Analysis.

Contents

1. Introduction

Chrysomya megacephala [10] is a mechanical vector of several pathogens [21] that can be used as a biological indicator in some medical legal investigations involving forensic entomology [13], and has already been used in biotherapy. In the laboratory, this species is usually reared with the use of artificial diets [20]. Taking into account other calliphorids, Firoozfar et al. [12] have investigated the best way of mass producing *Lucilia sericata*, which can also be used in maggot therapy. This blowfly has been eradicated from certain areas using the sterile insect technique. According to Chen et al. [7], eradication success has been possible due to the mass production of this blowfly using artificial diets.

For some flies, a major challenge in laboratory rearing is the need for an alternative substrate to the host. Mass rearing has been applied to various insect groups, usually with the goal of mass release of sterile insects or natural enemies for biological pest control. This allows insects to be reared under conditions that will make them equally, if not more, vigorous and adaptable to environmental conditions than wild populations [8,16]. According to Lapointe et al. [19], mass rearing of insects on artificial diets is often an essential component of research and development of pest control strategies. Furthermore, Lapointe et al. [17] proposed that many insect-rearing programs would benefit from the application of n dimensional mixture design methods to situations where diet optimization is desired according to the criteria selected by the researcher.

* This work was supported in part by FAPESP grants 2016/24707-4 and 2022/16455-6, Projeto Temático: Topologia Algébrica, Geométrica e Diferencial, and NSF grant DMS-1722995. The second author was supported by CAPES.

† Corresponding author

Submitted March 06, 2025. Published August 12, 2025
2020 *Mathematics Subject Classification:* 62P10, 62R40.

Several authors investigated insect mass rearing utilizing different methods. Bautista et al. [2] developed a mass-rearing method that facilitated large scale production of the parasitoid *Fopius arisanus* (Hymenoptera: Braconidae). Lapointe et al. [18] used a geometric design combined with response surface models to evaluate the effects of the amount and proportion of diet ingredients to optimize the production of normative adults of *Diaprepes abbreviates* (Coleoptera: Curculionidae). Assemi et al. [1] applied a design of experimental methodology using a Taguchi orthogonal matrix with the aim of improving the mass rearing feasibility of tobacco budworm, *Helicoverpa armigera* Hübner (Lepidoptera: Noctuidae).

Von Zuben et al. [24] investigated the effects of spatial distribution of immatures in patchy resources on larval competition for food, in experimental populations of *C. megacephala*, proposing a theoretical model of intraspecific competition considering a hypothetical previous random adult oviposition in a system of homogeneous patches. Subsequently, Bianconi et al. [3] proposed three artificial neural networks to predict the number of resultant adults in experimental populations of *C. megacephala*, based on initial larval density, amount of food available, and duration of immature stages. However, a more detailed study of optimal mass rearing of this species was still lacking. Hence, the importance of studying the effect on the mass rearing of different levels of larval aggregation in competition for food is demonstrated in the present work. Using a topological statistics method, the present work maps the complex combined effects of aggregation and availability of food resources in development of *C. megacephala*, providing a panoramic view of the variation in mass rearing results under multiple different conditions.

2. Objectives

This study investigates how variation in larval density (D) and amount of food (F) affects mass rearing using data collected by Von Zuben et al. [3] for experimental populations of *C. megacephala*. The primary goals are the following.

(i) to identify optimal combinations of D and F for mass rearing, including ways of accounting for cost effectiveness, as mass rearing may target different objectives such as optimizing the cost/benefit ratio or the sheer amount of mass produced;

(ii) to develop techniques, via topological data analyses and statistic methods, for interpreting and visualizing how variation in D and F affect developmental outcomes by constructing a cohesive map of information residing in the entire dataset.

Thus, a key objective is to develop a representation and visualization of the data as a network in which the nodes correspond to different (D, F)-combinations and the edges connect combinations that produce similar results. Such networks are very informative and lead to a more thorough understanding and visualization of the entire data landscape. We compare the outcomes of larval rearing under different conditions based on survival rates and properties of the distributions of pupal weight and size, beyond their means. The results can be easily interpreted and visualized and the method is effective and practical for optimizing mass rearing and identifying cost-effective approaches.

3. Materials and Methods

3.1. The Dataset

Our analysis of the insect *C. megacephala* was based on data collected in a study by Von Zuben et al. [3] in which larvae were reared to pupation in combinations $n = 20$ of initial larval density (D = number of specimens) and amount of food (F), labeled 1–20 in Table 1. Several combinations had the same ratio D/F to allow comparison of results at different levels of larval aggregation for the same initial amount of food per sample. The data set also contained the pupal weight (W) and the pupal size (S) of a sample of 60 survivors from each group, except for four groups that had measurements for all survivors (< 60). For each group, the table also shows the mean pupal weight ($\overline{W}$) and the mean pupal size ($\overline{S}$) estimated from the sample, and the survival rate (SR).

Table 1: Data for larval *Chrysomya megacephala*: larval density (D), amount of food (F), D/F ratio, mean pupal weight ($\overline{W}$), mean pupal size ($\overline{S}$), and survival rate (SR).

Group	D	F (g)	D/F	$\overline{W}$ (mg)	$\overline{S}$ (mm)	SR
1	75	15	5	44.42	8.89	15%
2	150	30	5	36.92	8.29	15%
3	300	60	5	49.73	9.01	41%
4	450	90	5	51.50	9.59	25%
5	150	15	10	28.02	7.57	10%
6	300	30	10	42.67	8.76	30%
7	600	60	10	28.62	7.45	58%
8	900	90	10	31.52	7.94	55%
9	300	15	20	19.55	7.19	9%
10	600	30	20	28.21	7.87	26%
11	1200	60	20	28.66	7.99	25%
12	1800	90	20	31.61	7.68	32%
13	450	15	30	23.53	7.17	15%
14	900	30	30	28.58	7.81	14%
15	1800	60	30	25.62	7.79	20%
16	2700	90	30	31.31	7.96	15%
17	600	15	40	23.28	7.24	12%
18	1200	30	40	27.18	7.27	7%
19	2400	60	40	17.82	6.62	4%
20	3600	90	40	21.06	7.16	12%

3.2. ROC Curves and AUC

The receiver operating characteristic (ROC) methodology originates from signal detection theory. It was used to quantify the ability of a radar operator to discern between information-bearing patterns (signal) and random patterns that distract from the information (noise). Its application and relevance have been notably expanded to fields such as machine learning and topological data analysis.

ROC analysis is also useful for evaluating the precision of statistical models that classify subjects into two distinct categories.

Let X and Y be two random variables with distribution functions F and G and survival functions $\tilde{F} = 1 - F$ and $\tilde{G} = 1 - G$. We say that X is smaller than Y in the stochastic order, and we write $X \leq_{st} Y$ if $\tilde{F}(x) \leq \tilde{G}(x)$, for all $x \in (-\infty, +\infty)$.

To compare classifiers, we may want to reduce the ROC performance to a single scalar value representing the expected performance. A common method is to calculate the area under the ROC curve, abbreviated AUC (Bradley, 1997; Hanley and McNeil, 1982). It is the most important numerical index used to describe the behavior of the ROC curve and is defined by

$$\text{AUC} := \int_0^1 \text{ROC}(t)dt.$$

Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1. Since random guessing produces a diagonal line with an AUC of 0.5, a practical classifier should exhibit an AUC that exceeds this value. The AUC has an important statistical property: the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [4].

For the scope of this analysis, we restrict our attention to random variables with support on a compact subset of $\mathbb{R}$. For an ordered pair (x_1, x_2) of compactly supported random variables with distributions P_1

and P_2, let $F_1(t) = P_1[x_1 \leq t]$ and $F_2(t) = P_2[x_2 \leq t]$ be their cumulative distribution functions. The ROC curve for the pair (x_1, x_2) is the curve in $[0,1] \times [0,1]$ given by $\{(F_2(t), F_1(t)); t \in \mathbb{R}\}$. Clearly, the curve depends only on the underlying probability distributions, and the points $(0,0)$ and $(1,1)$ always lie on the ROC curve. It follows directly from the definition that the ROC curve for (x_2, x_1) is the mirror image about the diagonal line of the ROC curve for (x_1, x_2). Histogram discretizations of the distributions induce a discretization of the ROC curve as a polygonal curve. The AUC quantifies the degree to which the distribution of x_2 is shifted towards higher values relative to x_1.

3.3. Weight and Size Histograms

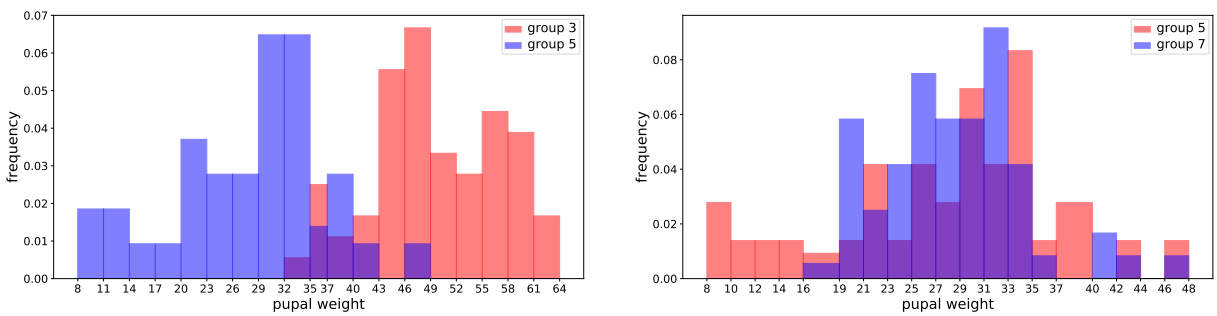
A histogram representation of a sample set with respect to measurement represents the frequency values of that measurement among the samples [6].

To compare developmental outcomes, we constructed weight and size histograms for each group using Python 2.7 libraries Matplotlib, NumPy, Scikit-learn and SciPy [14,15,22,23]. Fig. 1 illustrates contrasts in weight distributions for the pair of groups $(3,5)$ and $(5,7)$, respectively. As explained in the following, the histograms were used for:

(i) quantifying differences for any pair (i,j), $1 \leq i,j \leq 20$, via distances (measures of similarity/dissimilarity) between their histograms;

(ii) identifying clusters formed under these distance data combined with information about survival rates, as they also reflect success in rearing;

(iii) examining how clusters formed under these distance data relate to those obtained from the simpler, baseline statistics $\overline{W}$ and $\overline{S}$.



(a) Groups $(3,5)$ with $(D/F, D) = (5,300)$ and $(10,150)$

(b) Groups $(5,7)$ with $(D/F, D) = (10,150)$ and $(10,600)$

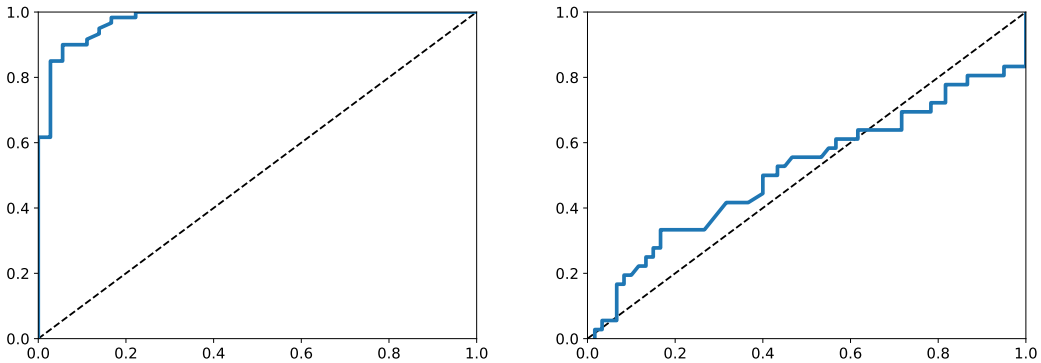
Figure 1: Histograms for weight distributions.

Finding the distance, or similarity/dissimilarity, between two histograms is relevant to classification of the patterns and clustering. Several measures for computing distance have been proposed and used [6].

We used receiver operating characteristic (ROC) curves (cf. [11]) to define the distances between the histograms. For each pair ordered (i,j), $1 \leq i,j \leq 20$, we computed the ROC curve for their weight histograms and the area under the curve, denoted $\mathrm{AUC}\,(i,j)$. Identical distributions have $\mathrm{AUC} = 1/2$, while $\mathrm{AUC}\,(i,j) = 0$ (or 1) indicates a sharp separation with the histogram of group j lying completely to the left (or right) of the histogram for group i, with no overlaps. The values $1/2 < \mathrm{AUC}(i,j) \leq 1$ indicate that, without taking survival rates into account, the group j was more successful than the group i, more so for larger AUC values. Fig. 2 shows the ROC curves for the pairs $(5,3)$ and $(5,7)$ with $\mathrm{AUC}\,(5,3) = 0.98$ and $\mathrm{AUC}\,(5,7) = 0.51$.

The ROC curve for (j,i) is the mirror image of the ROC curve for (i,j) about the diagonal line, so that $\mathrm{AUC}\,(j,i) = 1 - \mathrm{AUC}\,(i,j)$. Thus, we define the weight-based distance between groups i and j to be

$$d_w(i,j) = \max\{\mathrm{AUC}\,(i,j), \mathrm{AUC}\,(j,i)\} - 1/2\,.$$



(a) AUC $(5,3) = 0.9755$ (b) AUC $(5,7) = 0.5148$

Figure 2: ROC curves for weight distributions.

Note that $d_w(i,j) = d_w(j,i)$ and $0 \leq d_w(i,j) \leq 1/2$. A similar distance $d_s(i,j)$ was defined using size distributions. We integrate d_w and d_s into a single ws-distance by taking the average

$$d_{ws}(i,j) = (d_w(i,j) + d_s(i,j))/2\,.$$

To compare survival rates, we employed the distance $d_{sr}(i,j) = |SR(i) - SR(j)|$. Our analysis and visualization of the variation in results were performed using d_{ws} and d_{sr}. Table 2 shows $d_{ws}(i,j)$ for all pairs of groups.

Table 2: d_{ws}-distance for all pairs of groups.

0.00	0.13	0.04	0.12	0.31	0.04	0.33	0.25	0.38	0.29	0.25	0.27	0.40	0.28	0.34	0.24	0.40	0.32	0.46	0.42
	0.00	0.31	0.42	0.29	0.18	0.32	0.22	0.37	0.25	0.14	0.21	0.40	0.23	0.31	0.18	0.39	0.27	0.46	0.41
		0.00	0.15	0.46	0.18	0.48	0.43	0.49	0.44	0.35	0.42	0.50	0.43	0.46	0.41	0.50	0.43	0.50	0.49
			0.00	0.48	0.31	0.49	0.48	0.49	0.48	0.43	0.46	0.50	0.47	0.49	0.47	0.50	0.47	0.50	0.50
				0.00	0.41	0.06	0.14	0.15	0.08	0.06	0.04	0.20	0.04	0.10	0.12	0.20	0.05	0.34	0.24
					0.00	0.44	0.38	0.46	0.39	0.28	0.35	0.48	0.37	0.43	0.35	0.48	0.36	0.49	0.47
						0.00	0.22	0.14	0.12	0.08	0.08	0.20	0.10	0.18	0.18	0.19	0.04	0.41	0.28
							0.00	0.30	0.09	0.05	0.09	0.37	0.11	0.20	0.01	0.37	0.15	0.47	0.40
								0.00	0.22	0.15	0.21	0.02	0.19	0.17	0.27	0.04	0.02	0.26	0.09
									0.00	0.03	0.10	0.28	0.04	0.12	0.07	0.28	0.12	0.42	0.33
										0.00	0.09	0.16	0.04	0.04	0.05	0.15	0.14	0.34	0.21
											0.00	0.25	0.09	0.14	0.07	0.23	0.12	0.42	0.31
												0.00	0.23	0.24	0.33	0.03	0.02	0.28	0.08
													0.00	0.04	0.10	0.22	0.12	0.40	0.28
														0.00	0.16	0.23	0.09	0.42	0.31
															0.00	0.32	0.15	0.44	0.37
																0.00	0.01	0.30	0.10
																	0.00	0.17	0.02
																		0.00	0.23
																			0.00

The weight and size histograms contain finer information than the mean values $\overline{W}$ and $\overline{S}$, which are coarser summaries of the distributions. We showed that the information carried by histograms and mean values is largely in agreement, other than some fine details, as follows. First, we assign each of the twenty groups a score of W_1 ranging from 0 to 19 via AUC. The W_1 score of a group j is the number of groups for which AUC $(i,j) > 1/2$. As explained above, this means that group j was more successful than group i as measured by AUC. Second, we also assigned each group an S_1 score based on size adopting a similar scoring system. Third, we assign each group a total score $\lambda_1 = W_1 + S_1$. We also linearly ranked the twenty groups by mean weight and mean size assigning W_2 and S_2 scores ranging from 0 to 19, and a total score $\lambda_2 = W_2 + S_2$. Fig. 3 shows rank orderings based on λ_1 and λ_2, confirming that they only differ in finer details. The colors in the figures represent the intervals in which the λ scores fall.

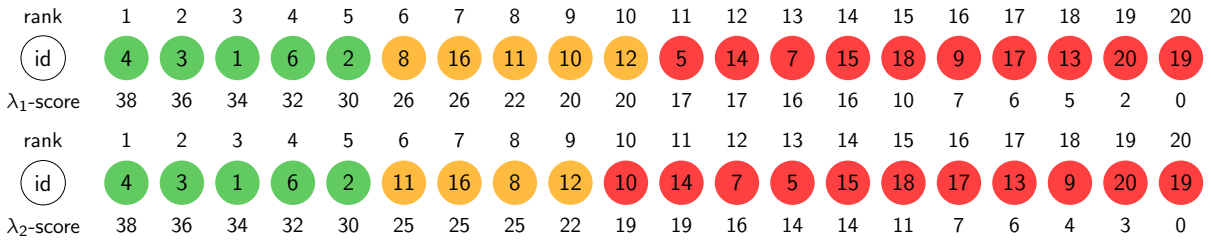


Figure 3: Rank ordering by λ_1 and λ_2 scores.

3.4. Statistical Significance

Before using ws-distances for data analysis, we investigated their statistical significance to establish a threshold value above which, at a given significance level, $d_{ws}(i,j)$ quantifies a meaningful difference in the outcomes for groups i and j. For each (i,j), $1 \leq i < j \leq 20$, we used 5000 permutations of the samples for groups i and j, recalculating $d_{ws}(i,j)$ using the permuted data. Fig. 4 shows scatter plots and histograms of ws-distances obtained for all permutations for the pairs $(3,5)$ and $(5,7)$, respectively, with red lines highlighting the true values of the ws-distance. The resulting p-values were $p = 0$ and $p = 0.28$, respectively.

The results for all pairs of groups are summarized in Fig. 5. Each point in the plot represents a pair (i,j), $1 \leq i < j \leq 20$. The first coordinate gives the distance $d_{ws}(i,j)$ and the second coordinate gives the p-value for the pair. The red horizontal line highlights the level $p = 0.05$, showing that $p \leq 0.05$ for $d_{ws}(i,j) > 0.09$. Moreover, p quickly decays toward zero as d_{ws} grows. This means that if $d_{ws}(i,j) \leq 0.09$, we do not have enough evidence to reject the hypothesis that the results for the pair (i,j) are the same in significance 5%, whereas d_{ws} quantifies a significant difference in results for higher values of d_{ws}. In particular, in cluster analysis, a pair for which $d_{ws}(i,j) \leq 0.09$ should not be placed in different clusters, at 5% significance, based on this metric.

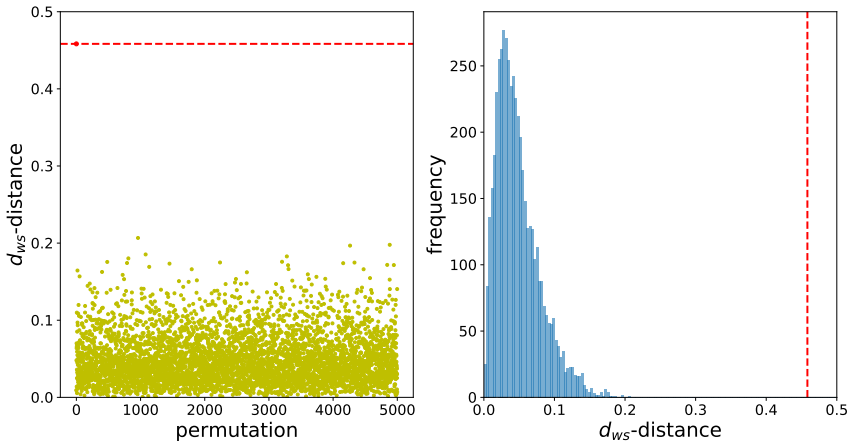
3.5. Cluster Analysis

Cluster analyses are machine-learning algorithms which has the property to delineate subgroups in datasets, characterized by discrete differences [9]. To interpret and visualize the information embedded in the distance data $d_{ws}(i,j)$ and $d_{sr}(i,j)$, we constructed Vietoris–Rips (VR) networks in which the nodes represent the twenty groups and a pair of nodes is connected by an edge according to a neighboring criterion based on the two distances.

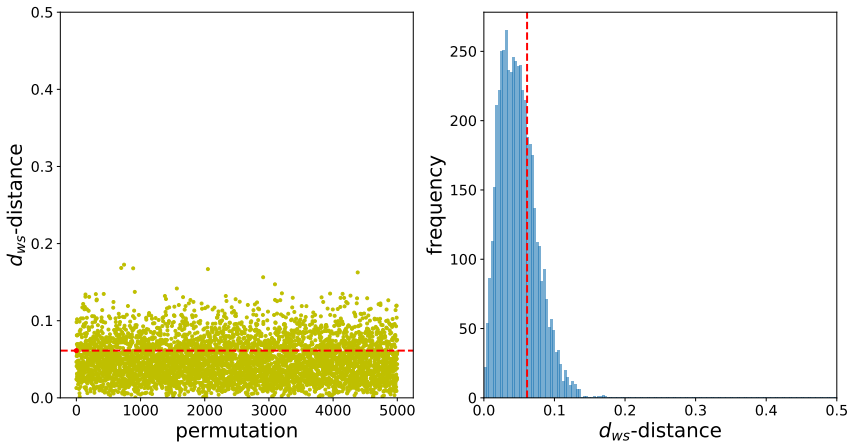
For a pair of positive numbers (a,b), the nodes representing the groups i and j are connected if $d_{ws}(i,j) \leq a$ and $d_{sr}(i,j) \leq b$. This hierarchy of networks, illustrated in Fig. 6, gives a multi-scale view of the evolution of the clusters as we relax the thresholds a and b for similarity of phenotypic outcomes and survival rates. As our analysis of d_{ws} indicated that at 5% significance we should have $a > 0.09$, we set the initial value $a = 0.10$ (a value just above the significance threshold). The parameter a increases along columns and b along rows. As we traverse a row from left to right, the maximum allowed difference in phenotypic outcomes (measured by d_{ws}) remains fixed, but larger differences in survival rates are allowed between neighboring nodes. A dual interpretation applies to columns of the grid. A diagonal NW–SE path in the grid corresponds to gradually allowing larger differences in both measures. Thus, in the northwest quadrant of the grid, groups connected by an edge exhibit high similarity in weight and size distributions, as well as survival rates.

4. Results and Discussion

The proposed multi-scale representation of results in *C. megacephala* mass rearing, under different combinations of larval aggregation and food resources, makes the information contained in the experimental data more interpretable and easier to visualize. The method produces a full map of mass rearing outcomes and let us identify the most successful groups in a quantitative, objective manner.



(a) Permutations for pair $(3, 5)$ with $d_{ws}(3, 5) = 0.46$ and $p = 0$



(b) Permutations for pair $(5, 7)$ with $d_{ws}(5, 7) = 0.06$ and $p = 0.28$

Figure 4: Permutations and p-values for d_{ws}.

The colors of the nodes in Fig. 6 show the level of success from the point of view of weight and size among survivors, but the grid and the networks summarize important additional information about similarity in survival rates and the development of phenotypic traits. The fact that there are no edges between the green nodes in the graph at position $(1, 1)$ (first row, first column) indicates that all the groups colored green exhibit differences in the measured traits or survival rates, possibly both. As we move along the first row, the groups 1 and 6 connect, indicating that the most pronounced difference between them is in survival rates, 15% and 30%, respectively. Similarly, going down the first column, we see that the dominant differences between Groups 1 and 4 are in measured phenotypes. The northwest quadrant of the grid is particularly informative as it reveals the structure among the green groups, sharply separating them from the less successful ones, except for the group 2, which received the lowest λ_1 score in the green cluster.

Note that all green groups had 5 or 10 larvae/gram of food, indicating that this is the optimal range for the D/F ratio. In addition, the two green groups with the highest survival rates were 3 and 6, both with larval density equal to 300 with different amounts of food (60 and 30 grams, respectively). Thus, considering mass rearing in practice, it would be more interesting to choose the first (60 g) for a not so expensive food and the second for the case of a higher cost food. As noted in the Introduction, a minimum level of aggregation is needed to allow larvae to increase feeding efficiency by producing a healthy supply

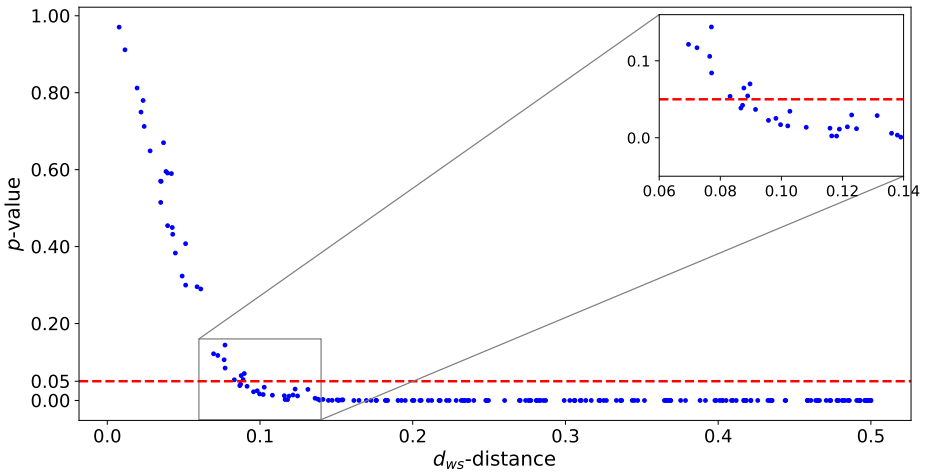


Figure 5: Statistical significance of d_{ws}-distance for all pairs of groups.

of digestive fluids. An aggregation level of $D = 300$ emerged as ideal for mass rearing as higher levels of competition for food proved detrimental to larval development. Although not high performers from the standpoint of weight and size, the grid shows that the groups 7 and 8 also distinguish themselves by not connecting to other groups in the VR networks until the survival rate threshold is significantly relaxed. This is caused by their much higher survival rates of 58% and 55%, respectively. Thus, cost effectiveness in mass rearing, which depends on both the distributions of mass produced and survival rates, requires a more detailed analysis. For this reason, we considered two basic measures of cost efficiency.

(i) a primary measure $R_1 = \overline{W} \times SR$, the average amount of mass produced per specimen in the original population;

(ii) a secondary measure $R_2 = \overline{W} \times SR \times D/F = R_1 \times D/F$, the average amount of mass produced per gram of food available.

Table 3 shows the values of R_1 and R_2 for all groups. As the values indicate, the average mass return per larva as measured by R_1 is highest for the group 3 with a density of 300 larvae and 60 grams of food. The average mass yield per gram of food, measured by R_2, is higher for some other groups that had higher survival rates, but at the expense of lower R_1 values.

Table 3: R_1: mass produced per larva in the original population; R_2: mass produced per gram of food available.

Group	R_1	R_2	Group	R_1	R_2	Group	R_1	R_2	Group	R_1	R_2
1	6.5	33	6	12.6	126	11	7.1	143	16	4.6	138
2	5.4	27	7	16.6	166	12	10.1	202	17	2.7	107
3	20.2	101	8	17.4	174	13	3.4	103	18	1.9	76
4	13.1	66	9	1.7	33	14	3.9	116	19	0.7	27
5	2.7	27	10	7.4	148	15	5.0	150	20	2.5	102

In conclusion, the experimental results point to a combination of larval density $D = 300$ and ratio $D/F = 5$ as nearly optimal for development. However, cost-effectiveness considerations might influence these values to enhance survival rates. Furthermore, the method used in the analysis not only identifies

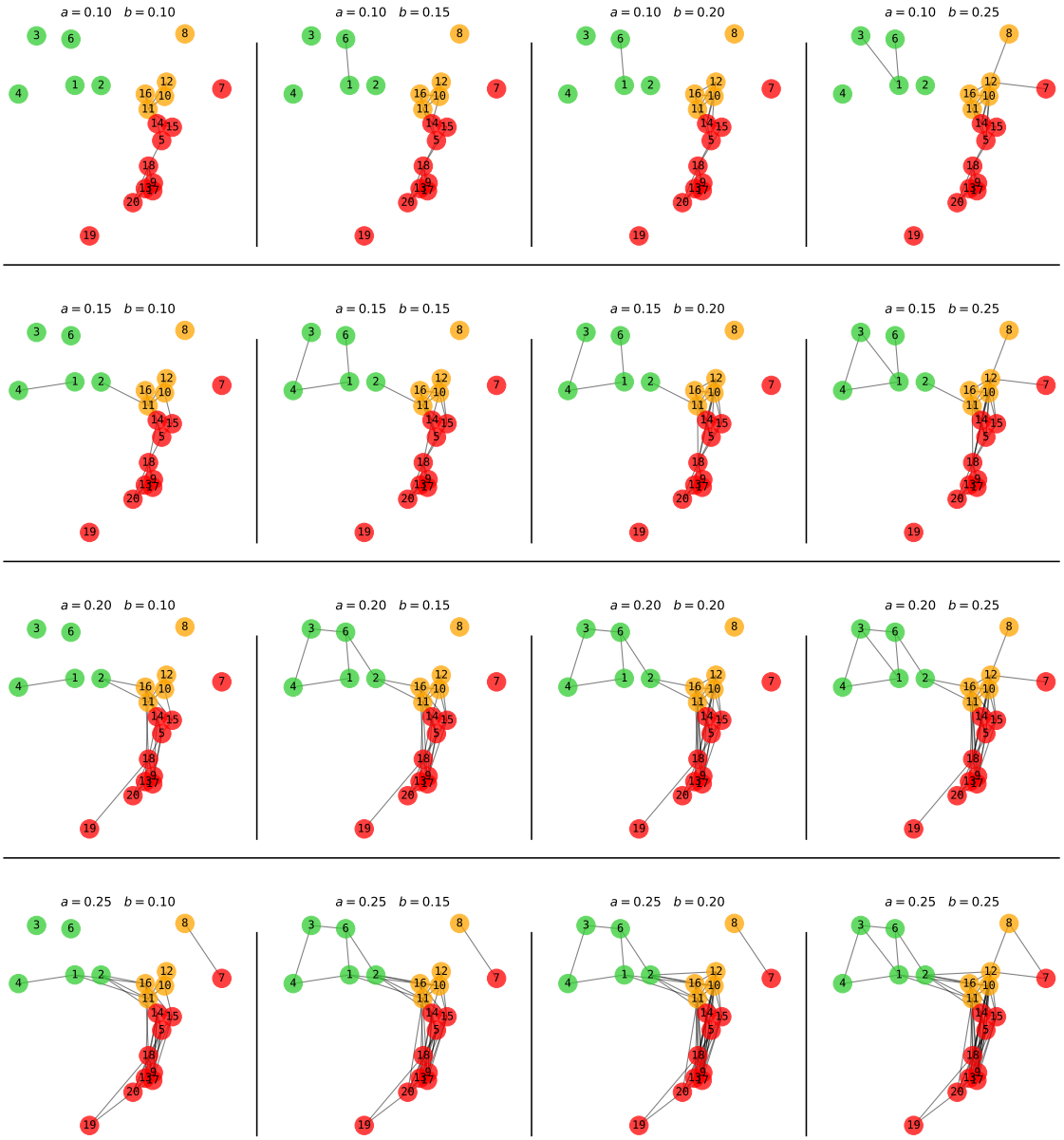


Figure 6: VR-networks with nodes numbered by group id (see Table 1) and colored by the λ_1 score.

optimal combinations of aggregation and food availability for mass rearing but also maps the relationships in developmental outcomes under different conditions.

Although this work focused on *C. megacephala*, the same topological and statistical methods apply to the analysis of the mass rearing of other species. The proposed summary and visualization of the experimental data used Vietoris–Rips graphs that capture the most basic relationships among the various groups as measured by similarity in phenotypic traits and survival rates. However, graph construction may be extended to higher-dimensional Vietoris–Rips simplicial complexes (cf. [5]), which capture higher-order relationships and can be useful as the data become more complex. In these simplicial complexes, three nodes that are pairwise connected are filled by a triangle, and, more generally, a k-simplex is assigned to any set of $k+1$ nodes that are pairwise connected by an edge.

References

1. H. Assemi, M. Rezapanah, R. Vafaei-Shoushtari, and A. Mehrvar. Modified artificial diet for rearing of tobacco budworm, *Helicoverpa armigera*, using the Taguchi method and Derringer's desirability function. *J. Insect Sci.*, 12(100), 2012.
2. R. C. Bautista, N. Mochizuki, J. P. Spencer, E. J. Harris, and D. M. Ichimura. Mass-rearing of the tephritid fruit fly parasitoid *Fopius arisanus* (Hymenoptera: Braconidae). *Biol. Control*, 15:137–144, 1999.
3. A. Bianconi, C. J. Von Zuben, A. B. S. Serapião, and J. S. Govone. Artificial neural networks: A novel approach to analyzing the nutritional ecology of a blowfly species, *Chrysomya megacephala*. *Journal of Insect Science*, 10, 2010.
4. C. Calì and M. Longobardi. Some mathematical properties of the ROC curve and their applications. *Ricerche di Matematica*, 64(2):391–402, 2015.
5. G. Carlsson. Topology and data. *Bull. Amer. Math. Soc.*, 46:255–308, 2009.
6. S-H. Cha and S. N. Srihari. On measuring the distance between histograms. *Pattern Recognition*, (35):1355–1370, 2002.
7. H Chen. Artificial diets used in mass production of the New World screwworm, *Cochliomyia hominivorax*. *J. Appl. Entomol*, 138:708–714, 2014.
8. A. C. Cohen. Formalizing insect rearing and artificial diet technology. *Am. Entomol.*, 47(4):198–206, 2001.
9. E. S. Dalmaijer, C. L. Nord, and D. E. Astle. Statistical power for cluster analysis. *BMC Bioinformatics*, 23(1):205, 2022.
10. J. C. Fabricius. *Entomologia systematica emendata et aucta*, volume 4. Hafniae, 1794.
11. T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
12. F. Firoozfar, H. Moosa-Kazemi, M. Baniardalani, M. Abolhassani, M. Khoobdel, and J. Rafinejd. Mass rearing of *Lucilia sericata* Meigen (Diptera: Calliphoridae). *Asian Pac. J. Trop. Biomed.*, 1(1):54–56, 2011.
13. B. Greenberg. Flies as forensic indicators. *J. Med. Entomol.*, 28:565–577, 1991.
14. J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.
15. E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed February 13, 2018].
16. E. F. Knipling. *The basic principles of insect population suppression and management*. USDA Agric. Handb., 1979.
17. S. L. Lapointe, T. J. Evens, and R. P. Niedz. Insect diets as mixtures: Optimization for a polyphagous weevili. *J. of Insect Physiol.*, 54:1157–1167, 2008.
18. S. L. Lapointe, T. J. Evens, R. P. Niedz, and D. J. Hall. Artificial diet optimized to produce normative adults of *Diaprepes abbreviatus* (Coleoptera: Curculionidae). *Environ. Entomol.*, 39(2):670–677, 2010a.
19. S. L. Lapointe, R. P. Niedz, and T. J. Evens. An artificial diet for *Diaprepes abbreviatus* (Coleoptera: Curculionidae) optimized for larval survival. *Florida Entomol.*, 93(1):56–62, 2010b.
20. T. T. S. Leal, A. P. Prado, and A. J. Antunes. Rearing the larvae of the blowfly *Chrysomya chloropyga* (Wiedemann) (Diptera, Calliphoridae) on oligidic diets. *Rev. Bras. Zool.*, 1(1):41–44, 1982.
21. A. R. Olsen. Regulatory action criteria for filth and other extraneous materials. III. review of flies and foodborne enteric disease. *Regulatory Toxicology and Pharmacology*, 28(3):199–211, December 1998.
22. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
23. S. van der Walt, S. Chris Colbert, and G. Varoquaux. The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, 13:22–30, 2011.
24. C. J. Von Zuben, F. J. Von Zuben, and W. A. C. Godoy. Larval competition for patchy resources in *Chrysomya megacephala* (Dipt., Calliphoridae): implications of the spatial distribution of immatures. *J. Appl. Entomol.*, 125(1):537–541, 2001.

Thiago de Melo,
Department of Mathematics,
Institute of Geosciences and Exact Sciences,
São Paulo State University (Unesp),
Brazil.
E-mail address: thiago.melo@unesp.br

and

Rodrigo Rosa,
Department of Mathematics,
Institute of Biosciences, Letters and Exact Sciences,
São Paulo State University (Unesp),
Brazil.
E-mail address: rodrigo.rosa@unesp.br

and

Jamil Viana Pereira,
Department of Mathematics,
Institute of Geosciences and Exact Sciences,
São Paulo State University (Unesp),
Brazil.
E-mail address: jamil.v.pereira@unesp.br

and

Washington Mio,
Department of Mathematics,
Florida State University,
USA.
E-mail address: mio@math.fsu.edu

and

Claudio José von Zuben,
Department of Biodiversity, Biosciences Institute,
São Paulo State University (Unesp),
Brazil.
E-mail address: claudio.jv.zuben@unesp.br

and

Alice Kimie Miwa Libardi,
Department of Mathematics,
Institute of Geosciences and Exact Sciences,
São Paulo State University (Unesp),
Brazil.
E-mail address: alice.libardi@unesp.br