1122

NIRS Model Trained on Lab-reared Mosquitoes Estimates Age of Wild Mosquitoes with 67% Similarity to Detinova

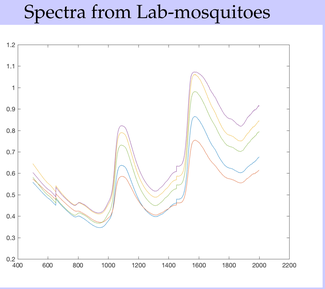
Masabho P. Milali[1,2], Maggy Sikulu-Lord[2,3], Samson S. Kiware[1,2], Richard J. Povinelli[1], and George F. Corliss[1]

[1]Marquette University USA, [2]Ifakara Health Institute Tanzania, [3]QIMR Berghofer Medical Research Institute Australia

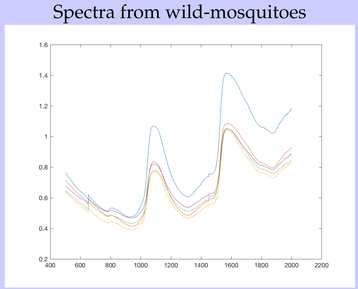
> Introduction

No clear difference between spectra collected from lab-reared mosquitoes and those collected from wild mosquitoes (Milali.2016). We applied NIRS model trained on lab-reared mosquitoes to estimate age of wild mosquitoes. We compare model estimates with Detinova estimates. The model estimated age with 67% similarity to Detinova.The on going practice of applying model trained on the lab-reared mosquitoes might be appropriate [3].

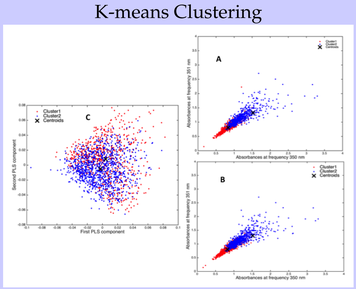
Spectra from Lab-mosquitoes



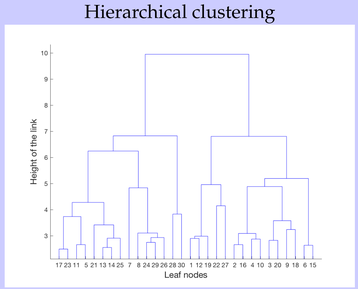
Spectra from wild-mosquitoes



K-means Clustering



Hierarchical clustering



K-means clustering:

Cluster	Lab-reared	Wild	Total	Av.SC	P-value
1	337	495	832	0.75	
2	261	432	693	0.62	0.258
Total	598	927	1525		

Hierarchical clustering:

Cluster	Leaf node	Lab-reared	Wild	Total	P-value
1	17,23,11,5,21,13,14, 25,7,8,24,29,26,28,30	132	175	307	
2	1,12,19,22,27,2,16,4, 10,3,20,9,18,6,15	466	752	1218	0.129
Total		598	927	1525	

Financial support: We thank Grand Challenge Canada for funding the study produced the data, The Marquette University GasDay and Graduate school, for sponsoring studies of the first author, which resulted into the knowledge to perform the analysis of this study.

Project supported by: Grand Challenges Canada® Grands Défis Canada

GASDAY MARQUETTE UNIVERSITY

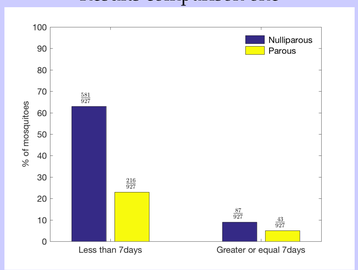
MARQUETTE UNIVERSITY BE THE DIFFERENCE.

> Method and Results

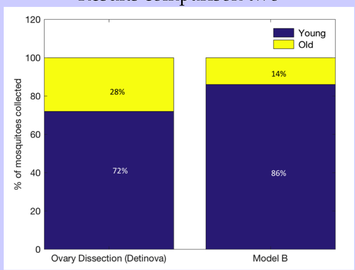
We applied a classification model trained on lab-reared An. arabiensis to classify the ages of wild *An. arabiensis*. Because we lack age labels of wild *An.arabiensis*, we indirectly validated our model by:

1. Analyzing the distribution of nulliparous and parous mosquitoes in each age-class from the model;
2. Comparing the number of mosquitoes in each age-class obtained when classification is done using a model and when done using Detinova ovary dissection; and
3. Relating with the historical studies conducted to determine the age structure of wild mosquito populations.

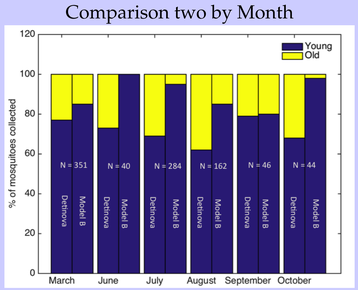
Results comparison one



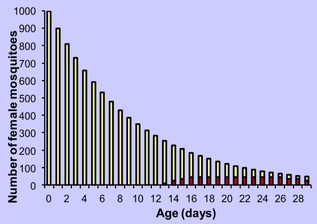
Results comparison two



Comparison two by Month



Comparison three (Brownstein et al).



NIRS model classified 86% of the total 927 wild mosquitoes as young (less than seven days old) and 14% as old (greater or equal to seven days old). Detinova ovary dissection classified 72% of the same number (927) of mosquitoes as young (not laid eggs) and 28% as old (laid eggs). A Jaccard similarity analysis comparing Detinova ovary dissection and our model trained on lab-reared mosquitoes shows there is a 67% chance that the two methods will classify a mosquito into the same age class and a 33% chance they will classify a mosquito into different age classes. Hence, a classification model trained on lab-reared mosquitoes and Detinova ovary dissection are more similar than they are different.

Acknowledgment:
We thank Andrew Kafwenji and Paulina Kasanga for help maintaining the mosquito colony, Floyd E. Dowell (Engineering and Wind Erosion Research Unit, Grain Marketing and Production Research Center, U.S. Department of Agriculture, Agricultural Research Service, Manhattan, KS, USA) for loaning us the Near Infrared Spectrometer used to scan the mosquitoes, Marta F. Maia, Fredros O. Okumu, and Sheila Ogoma for participating in grant writing and managing of the project produced the data used in this study.

> Maths

Equation to calculate silhouette coefficient:
Let

- s(o) = Silhouette coefficient of a single object 'o'
- s(o) = Average distance of object 'o' to the other objects in its cluster
- b(o) = Average distance of object 'o' to the other objects in the nearest cluster.

Then

$$s(o) = \frac{b(o) - a(o)}{max(a(o), b(o)).} \quad (1)$$

Jaccard Coefficient: Jaccard similarity coefficient is a static measure for comparing the similarity diversity of sample sets. Assume A and B are two different sets. Then

$$J(A,B) = \frac{A \cap B}{A \cup B} \quad (2)$$

Interpretation:

- Jaccard coefficient(JC) ranges from 0 to 1.
- High JC = The sets are similar (JC = 1, sets are same)
- Low JC = The sets are dissimilar (JC = 0, sets are different)

If we translate this to our problem, A can be Detinova and B can be NIRS. Therefore,

$$J(Detinova, NIRS) = \frac{Detinova \cap NIRS}{Detinova \cup NIRS} \quad (3)$$

> Conclusion

While further studies may be required to explore a more appropriate way to estimate age of wild mosquitoes, these results strengthen the ongoing practice of training models to estimate age of wild mosquitoes using spectra collected from lab-reared mosquitoes [3]. The reliability of the age estimates from the model might still be questioned, which is acceptable as model estimates are not always expected to be accurate [2, 4, 5]. Despite of this known caveat, the most important advantage of using models is to give insight to situations where it is difficult to get the truth [4, 1]. Getting actual age in days of wild mosquitoes is almost impractical as it is very difficult, tedious, time inefficient, and expensive. Therefore, the ongoing practice of applying a model trained on laboratory-reared mosquitoes to estimate wild mosquitoes might not be an ideal, but the results from this study show that it might be reliable enough to give an insight on age structure of wild mosquito population, especially when complemented with other existing knowledge on age structure of wild mosquitoes.

References:

1. Jong MCM, de,1995. Mathematical Modelling in Vertinary Epidemiology. Why Model Building is Important. Preventive Veterinary Medicine. 1995;25
2. Poeter E, 2007. All Models are Wrong, How do We Know Which are Useful? Ground Water. 45(4):390-1
3. Sikulu-Lord MT, Milali MP, Henry M, Wirtz RA, Hugo LE, Dowell FE, Devine GJ, 2016. Near-Infrared Spectroscopy, a Rapid Method for Predicting the Age of Male and Female Wild-Type and Wolbachia Infected *Aedes Aegypti*. PloS
4. Sterman JD, 2002. All Models are Wrong: Reflections on Becoming a Systems Scientist. System Dynamics Review.
5. Wit E, Van den Heuvel E, Romeijn J, 2012. All Models are Wrong: An Introduction to Model Uncertainty.Statistica