Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2018 September ; 10779: . doi:10.1117/12.2323432.

Analysis of predictor variables for mosquito species identification from dual-wavelength polarization-sensitive lidar measurements

Adrien P. Genoud^a, Roman Basistyy^a, Gregory M. Williams^b, Benjamin P. Thomas^a

^aNew Jersey Institute of Technology, 323 Dr Martin Luther King Jr Blvd, Newark, NJ, USA

^bCenter for Vector Biology, Rutgers, The State University of New Jersey, 57 US Highway New Brunswick, NJ, USA.

Abstract

Mosquito-borne diseases are a major challenge for Human health as they affect nearly 700 million people every year. Monitoring insects is generally done through trapping methods that are tedious to set up, costly and present scientific biases. Entomological lidars are a potential solution to remotely count and identify mosquito species and gender in real-time. In this contribution, a dual-wavelength polarization sensitive lidar is used in laboratory conditions to retrieve the wingbeat frequency as well as optical properties of flying mosquitoes transiting through the laser beam. From the lidar signals, predictive variables are retrieved and used in a Bayesian classification. This paper focuses on determining the relative importance of the predictive variables used in the classification. Results show a strong dominance of the wingbeat frequency, the impact of predictive variables based on depolarization and backscattering ratios are discussed, showing a significant increase in classification accuracy.

Keywords

Remote sensing; Laser; Classification; Mosquito; *Aedes Albopictus*; *Aedes Vexans*; Entomology; Lidar

1. INTRODUCTION

Mosquitoes are considered one of the most dangerous animal and poses major health concern in many countries [1]. They affect around 700 million people annually resulting in over 1 million premature deaths [2]. Therefore, mitigation methods are of the utmost importance to reduce mosquitoes' impact on human population. To maximize the efficiency of vector control strategies, a fine understanding of mosquitoes dynamic and population is paramount [3].

Presently, the majority of mosquito population monitoring campaigns are done through trapping methods using bait such as pheromone, light, food or CO₂ [4,5]. The widely use of traps is not without reason as they allow for a thorough study of the captured specimens, with nearly 100% identification accuracy, but their use is limited by the cost required to set and collect the traps as well as the laboratory analysis by qualified personnel. Moreover,

traps can be biased regarding species and gender [6,7] and they provide little information that are temporarily resolved or spatially resolved above a few meters around the trap. Because of the aforementioned reasons, wide scale study of mosquito population can be challenging and costly despite being incredibly important for the development of effective vector control strategies and action threshold that protects human populations.

An alternative solution to study mosquitoes or other flying insects resides in entomological lidars [8,9,10,11] or optical sensors [12,13,14]. For optical instruments to provide data for eco-epidemiologic model and risk assessment, they must be able to differentiate mosquito species and, for an even better efficiency, gender. The wingbeat frequency is known to be a valuable predictive variable to differentiate between gender and to a lesser extent species of mosquitoes [10,12,13,15]. Additional predictors such as wing and body polarimetric and spectral ratios have also been studied to enhance the identification with promising results [11].

In this paper the experimental setup allows for the recording of the backscattered light by mosquitoes at two wavelengths, 924 and 1320 nm as well as polarimetric study of the later. Mosquito interaction with light is dependent on both wavelength and polarization [11,16] therefore the effectiveness of the polarimetric and spectral features of this optical instrument to improve species and gender differentiation of mosquitoes can be evaluated. This paper will investigate what predictor variables, obtained from the raw data recorded by the three detection channels, are the most appropriate to separate three different species of mosquitoes with both genders. For the classification tasks, a Naïve Bayes classifier was chosen for its implementation simplicity while maintaining good performances [13,17,18].

The paper is organized as follow. First the methodology to retrieve the predictors and their associated probability functions is explained together with the classification. Secondly the experimental results are presented with an example of raw data, probability density function and importance of predictors. Finally, we discuss the results and the more important predictors for species and gender identification are highlighted.

2. METHODOLOGY

2.1 Experimental methodology

In this section, the experimental setup is presented along with the method to measure the optical signals and to retrieve the optical information on the studied mosquito specimen. Figure 1 presents the laboratory layout used to record the optical characteristics of the mosquito with three different optical channels. In this setup, two continuous wave (CW) infrared laser diode source emitting at 1320 and 924 nm wavelength are employed. The continuity of those lasers allows for a constant monitoring of the specimen thus avoiding the dead time that a pulsed laser would introduce. Both lasers are linearly polarized, yet only the depolarization of the 1320 nm wavelength laser is measured. A half-wave plate is used to rotate the laser beam polarization to coincide with the axis of the polarization beam splitter (PBS) of the detection system and is filtered with a polarizer. For both lasers, their spatially Gaussian beams are collimated and expanded to reach 2 cm full width half max to increase the likeliness of interaction with the free flying mosquitoes while ensuring that transit times

are long enough to retrieve the wingbeat frequency. Using a dichroic mirror, both laser beams are superimposed and follow the same optical path. The two wavelengths were chosen to fit specific requirement. First, they ensure a non-negligible backscatter power [16]. Secondly, they are outside of the visual perception of the mosquitoes [19,20] which will eliminate a potential bias as mosquitoes can be influenced by visual stimuli. The beams are transmitted through a Plexiglas tube (1.25 m long, 12.5 cm diameter) located 3 m away from the output mirror. The beams enter the tube through a borosilicate glass tilted 5° downward at one end of the tube and are stopped by an Infrared (IR) beam-stop at the other end. Mosquitoes are introduced in the tube through small apertures. A 7 cm diameter off-axis parabolic gold mirror collects the backscattered light from the mosquito and reflects it into the detection system. There, a dichroic mirror separates both wavelengths. The 924 nm wavelength laser beam is transmitted by the dichroic mirror and is focused onto the 3.14 mm^2 active area of an InGaAs amplified near-infrared (NIR) photodetector with a 67 kHz DC bandwidth. The 1320 nm wavelength laser beam is reflected by the dichroic mirror onto a PBS that separates the collected light between parallel and perpendicular polarization. Both polarizations of the laser beam are then focused on two 3.14 mm^2 active area InGaAs amplified NIR photodetectors with a 67 kHz DC bandwidth. All three signals are recorded on three channels using a 16 bit 250 MS/s 125 MHz bandwidth digitizer.

This experimental setup allows for the recording of three different optical signals, the 924 nm wavelength noted I_{924} , the 1320 nm wavelength with parallel polarization and the 1320 nm wavelength with perpendicular polarization, respectively noted $I_{1320, //}$ and $I_{1320, \perp}$. As described in length in Genoud et al. [10], the contribution of the wings and the body of the mosquito to the signal can be differentiated on all of those three channels. This gives, overall, 6 different components that can yield information about the mosquito optical characteristics. In addition, the wingbeat frequency of the mosquito can be retrieved by analysis of the fast Fourier transform of each signal [10]. From the aforementioned information, differences between species and/or gender of mosquitoes are expected and used as predictors in a Naïve Bayes Classifier [17,18]. In this paper, the goal is to evaluate the contribution and dominance of these predictors when used to differentiate species and gender of mosquitoes.

In this study, 3 mosquito species of both genders were studied:

- *Aedes albopictus*, also known as the Asian tiger mosquito, this species originates from tropical and subtropical areas in Southeast Asia and has recently spread into Western countries such as the United States. This species is of particular importance in respect to health prevention as they can be an epidemiological vector for diseases such as dengue fever, chikungunya fever, yellow fever or Zika virus.
- *Aedes vexans*, this is the most common mosquito in Europe and has a worldwide distribution. This species is known for aggressive biting behavior towards human and is suspected to be a contributor to the spread of Zika virus. Furthermore, it can transmit the West Nile virus, Rift valley fever virus or St. Louis encephalitis virus.

- *Culex Genus (unknown species)*, from which several species are vectors of disease such as West Nile virus and multiple forms of encephalitis. *Culex* are widely geographically spread and one of the most encountered mosquito genera in North America

2.2 Calibrations and predictive variables

Equation 1 presents the general expression of the measured backscattered intensity by a mosquito flying through a laser beam at a specific distance d from the detection system, for any of the three channels.

$$I = \frac{K}{d^2} \cdot I_0 \cdot \beta$$

(1)

Where I is the recorded intensity by the detector of this channel, I_0 is the initial intensity of the laser beam, d is the distance between the detector and the mosquito at the time of the measurement and K a constant taking into account the size of the off-axis parabolic mirror, quantum efficiency of the detector, and optical transmission or reflection coefficients of the transmitting and collecting optics. This equation holds true for all three channels and both the wings and body components. The optical transmission of the air being considered as negligible for the range of distance of 3-4.25 m. Considering that the duration of a mosquito transit through the laser beam is in the order of 100-200 ms, the contribution of backscattering from gas molecules or particles in the probed volume of air and potential other light sources to the recorded signal I can be treated as constant and easily subtracted from the signal. Since this system uses CW diode lasers, the evaluation of the distance through time of flight methods, as in most lidar studies, is inapplicable. However, by considering the ratio between signals, the influence of the distance can be removed. Using a near Lambertian gold target, each detection channel was aligned so that the ratio between signals is constant with range, for instance, $I_{924}(d) / I_{1320, //}(d)$ is constant with the distance d varying from 3 to 4.25 m using the gold target. The consistency of these ratios with distance was determined and the less constant ratio had less than 1.6% variation over the entire range. This ensure that, within a reasonable error margin, signal ratios are independent from the actual distance at which the mosquito transit through the beam. In essence, signal ratios are dependent of the specificity of this optical setup and therefore, in an effort to broaden the scope of this publication, a calibration has been performed to retrieve backscattering coefficient ratios (e. g. $\beta_{924} / \beta_{1320, //}$) from the signal ratios (e.g. $I_{924} / I_{1320, //}$). The following paragraphs describes the calibration for depolarization ratios and the calibration between the two considered spectral ranges.

The calibration for the ratio $I_{1320, \perp} / I_{1320, //}$ is a relatively usual situation in polarization sensitive lidar studies. A thorough and effective calibration method, described in Alvarez et al. [21], was used in this work. The advantages of this calibration method are that in addition

to the retrieval of the calibration constant between the parallel channel and the perpendicular channel, it also gives the depolarization ratio of the target as well as an information on the possible misalignment between the PBS and the emitted parallel polarization. To briefly resume the methodology described fully in Alvarez et al. [21], the orientation of the emitted polarization is manually rotated away from its original alignment by the use of a half-wave plate. For each rotated state, characterized by the angle (Υ), the ratio of the intensity of the channel $I_{1320,\perp}$ over channel $I_{1320,\parallel}$, noted δ^* , is measured and fitted using equation 2 from Alvarez et al [21]. Figure 2 presents the measurements as well as the fit obtained, which provides the best suited values of depolarization ratios (δ), calibration constant (G) and reference angle (Υ_0).

$$\delta^* = G \frac{\delta + \tan(2(\Upsilon - \Upsilon_0))^2}{1 + \delta(\tan(2(\Upsilon - \Upsilon_0))^2)}$$

(2)

The calibration constant between the 924 nm channel and both polarizations of the 1320 nm channel can be found using the known reflectance of the gold target. In addition, the total backscattering coefficient at 1320 nm is in turn equal to the sum of the backscattering coefficient of the perpendicular and parallel polarization ($\beta_{1320,\parallel}$ and $\beta_{1320,\perp}$). As presented in equations 3 and 4, the calibration constants ($G_{\parallel/\parallel}$ and $G_{\perp/\parallel}$) can be found by measuring signal ratios between channels while knowing the depolarization ratio of the gold target ($\delta = \beta_{1320,\perp} / \beta_{1320,\parallel}$).

$$G_{\parallel/\parallel} = \frac{I_{1320,\parallel/\parallel}}{I_{924}} * \frac{\beta_{924}}{\beta_{1320,\parallel/\parallel}} = \frac{I_{1320,\parallel/\parallel}}{I_{924}} * \frac{\beta_{1320,\parallel/\parallel}}{I_{1320,\parallel/\parallel}} = \frac{I_{1320,\parallel/\parallel}}{I_{924}} * \frac{\beta_{1320,\parallel/\parallel} + \beta_{1320,\perp/\parallel}}{\beta_{1320,\parallel/\parallel}} = \frac{I_{1320,\parallel/\parallel}}{I_{924}} (1 + \delta)$$

(3)

$$\begin{aligned}
 G_{\frac{\lambda}{924}} &= \frac{I_{1320, \lambda}}{I_{924}} * \frac{\beta_{924}}{\beta_{1320, \lambda}} = \frac{I_{1320, \lambda}}{I_{924}} * \frac{\beta_{1320}}{I_{1320, \lambda}} = \frac{I_{1320, \lambda}}{I_{924}} * \frac{\beta_{1320, \lambda} + \beta_{1320, \lambda}}{\beta_{1320, \lambda}} \\
 &= \frac{I_{1320, \lambda}}{I_{924}} \left(1 + \frac{1}{\delta}\right)
 \end{aligned}$$

(4)

The aim of this paper is to study the relevance of the predictive variables to the accuracy of differentiation between species and gender of mosquitoes. With the experimental setup previously detailed, there are multiple predictors available. First and foremost, there is the wingbeat frequency which is known to be an effective predictor of gender and species of flying insects including mosquitoes [10,12,13,15]. In addition, measured signals from the 3 channels for both wings and body yield a total of 15 possible backscattering ratios. Furthermore, the ratio between the total backscattering coefficient (body + wings) at each wavelength was added, along with the ratio of the contribution of both wavelengths to the wings component of the signal by the contribution of both wavelengths to the body component of the signal. Mosquito absorption is known to be wavelength dependent [16] and the wing surface and body surface may be different from one species/gender to the other, therefore it is reasonable to assume that either could be an effective predictor, bringing the total number of studied predictors to a total of 18.

2.3 Event classification and predictor importance

For every mosquito event, understood as a recorded mosquito transit through the laser beams, the 18 predictors are determined and used in a Naïve Bayesian classification to predict from which of the 6 classes (both genders of the three species) the event is most likely to have originated from. This determination of the predictors allows for the construction of histograms describing the value distribution of the predictors. Those histograms will in turn allow for the evaluation of a distribution function describing the likeliness that a given predictor will take a specific value.

For every event, the most likely class C_j will be predicted by the classifier. The correct class of all the recorded events is known a priori since the experiment was conducted in a control environment and the species and gender of the mosquitoes determined before they are released into the measurement enclosure. This allows for the verification of the prediction made by the classifier. In Bayesian statistics, the probability that an event E_i belong to the class C_j is described by equation 5 [18].

$$P(C_j | E_i) = \frac{P(C_j)P(E_i | C_j)}{P(E_i)}$$

(5)

Where $P(C_j)$ is the prior probability of the class C_j , $P(E_i | C_j)$ is the probability of obtaining the value E_i in the class C_j and $P(E_i)$ is the prior probability of the observed event E_i . For each event E_i , 18 predictive variables can be retrieved. The prior probabilities of every class $P(C_j)$ will be chosen as equal since prior knowledge about the relative likeliness of a class is usually unknown. Equation 5 can be applied to every predictor individually. The next step is to obtain, for each predictor, a probability density function (PDF) that can be used to express $P(E_i | C_j)$. Those probability density functions were estimated by using a Kernel density estimation (KDE). For a given event E_i and a given predictor a score $S_{C_j}(E_i)$ will be attributed to every class C_j as described by equation 6.

$$S_{C_j}(E_i) = \frac{P(C_j | E_i)}{\sum_{C_k \in C} P(C_k | E_i)} = \frac{P(E_i | C_j)}{\sum_{C_k \in C} P(E_i | C_k)}$$

(6)

All prior probabilities $P(C_j)$ and $P(E_i)$ are equal for all class, either because they were chosen equal for all classes: $P(C_j)$ or because they are independent from the class: $P(E_i)$. Thus, they cancel out and the only remaining terms are the one that can be estimated by the Kernel PDFs. For a given predictor the sum of all the score of an event is equal to 1. So, the score of the class C_j can be regarded as the probability of this class in respect to all the other possible classes for the evaluated predictor. This methodology is then applied to every predictor individually and the total score $TS_{C_j}(E_i)$ of a class for the event E_i is defined by equation 7.

$$TS_{C_j}(E_i) = \sum_{p=1}^{18} S_{C_j}(E_i)_p$$

(7)

Where $S_{C_j}(E_i)_p$ is the score of the class C_j for the predictor p of the considered event E_i . Then the predicted class P_{C_j} is the class C_j for which the total score is the greatest. This class is, for the Bayesian classifier, the most likely class but it is not necessarily the correct one since there is some overlap between the different Kernel PDFs.

Finally, the overall accuracy of the classification is equal to the percentage of correctly predicted events. This statistic will allow to determine the importance of each predictor to the classification. To evaluate the importance of a predictor, the difference in the overall accuracy between the classification that includes all the predictors $a(\sum_j p_j)$ minus the one without the studied predictor p_k $a(\sum_{j \neq k} p_j)$ is calculated [22] as described in equation 8. Hence, if the predictor variable enhances the overall accuracy, the difference $\Delta_a(p_k)$, hereafter referred to as predictor contribution, will be positive, on the contrary if the predictor variable decreases the overall accuracy the difference will be negative, and the predictor variable should not be used.

$$\Delta_a(p_k) = a(\sum_j p_j) - a(\sum_{j \neq k} p_j)$$

(8)

3. RESULTS

3.1 Data analysis

This section presents an example of the optical signal recorded by the three channels for a mosquito transiting through the laser beams out of the total 672 events recorded. Then, the KDE of the predictor $\beta_{1320,\perp,b} / \beta_{1320,\parallel,b}$ (see table 1) for all 6 classes of mosquitoes is displayed, Figure 4, as an illustration of the methodology detailed in section 2.2. This methodology is then applied to all mosquito transits for every predictor. Results that are used later as PDFs in the Bayesian classification.

As displayed in Figure 3, when a mosquito transit through the laser beams, the raw signals display a Gaussian shape due to the spatial profile of the Gaussian laser beams and sharp intensity peaks due to the wings' orientation rapidly changing. Those two different contributions are then separated by using an interpolation of local minimums of the raw data as described in Genoud et al. [10].

The histograms presented in Figure 4 are used to apply a KDE and the fitted results are displayed in the same figure. The area under the curve is then evaluated by trapezoidal numerical integration and used to normalize the area of the KDE fit to an area under the curve of one for all predictors and all classes. For any events, the value of those normalized KDE fit is evaluated in order to provide the probability that an event belong to any of the class used in the classification. The value distribution of these predictors has non-negligible fluctuation that probably originates from the intrinsic variability between different specimens, the uncertainty of measurement but also from the fact that these ratios, like most of the predictors in this paper, may be dependent on the orientation of the mosquito during a transit.

Table 1 presents the list of all predictor variables that have been studied in this work. For the sake of clarity those predictor variables are expressed by a simplified expression.

3.2 Classification and predictors importance results

In this section, the importance of every predictor variable to the classification is evaluated by following the methodology described in section 2.3. The overall accuracy is evaluated for multiple combinations of predictor variables. In some cases, the probability density function of predictor variables of different classes may overlap. For instance, the wings to body ratio for the 924 nm wavelength is very similar between species and gender, making this predictor variable ineffective as it does not increase the overall accuracy of the classification. In addition, some predictor variables have a negative contribution to the overall accuracy, as they may increase the score of the wrong class for events that are difficult to classify and lead to a decrease of the overall accuracy.

Table 2 presents the confusion matrix when all the predictor variables are considered, even the ones with a null or negative contribution. Results are obtained from the average of 10,000 iterations where 50 events of each class (300 total events) are randomly chosen and classified.

The overall accuracy for species and gender classification is equal to $64.34 \pm 1.84 \%$. This result is better than the $55.58 \pm 1.65 \%$ overall accuracy obtained when only considering the wingbeat frequency. Yet, including all the predictors does not provide the highest overall accuracy. The overall accuracy difference Δ_a is calculated, as detailed by equation 8 in section 2.3 and the results are presented in Table 3. This approach uses the difference in overall accuracy with and without a predictor variable as a way to determine its effectiveness in the classification.

Of the 18 predictor variables, 8 have a negative contribution, meaning that without them the classification accuracy increases. The remaining 10 predictors have a contribution ranging from 15.4% to 0.06%. As expected, the wingbeat frequency is the one with the highest impact on classification and its importance far exceed the one from other positive predictors. The knowledge about the importance of each predictor allows for the removal of the negative predictors from the classification, improving the overall accuracy of the classification. After removing every predictor variable with negative contribution, the overall accuracy reaches $70.65 \pm 1.67 \%$ which is 6.31% more than with all the predictor variables and 15.07% more than with the wingbeat frequency alone.

Yet, one can note that the increase in overall accuracy is superior to the absolute sum of all the negative contribution. Similarly, it is observed that the predictor contribution varies if other predictor variables are discarded. This demonstrate the complex dynamics between predictor variables. While the predictor contribution is a good indication of the quality of the predictor variable, the highest overall accuracy is not obtained when each predictor variable with negative contribution is removed. The highest overall accuracy is found to be $74.95 \pm 1.62 \%$ with a total of 6 predictor variables:

- F_{wb}
- $\beta_{1320,\perp,b} / \beta_{1320,\parallel,b}$
- $\beta_{1320,\perp,w} / \beta_{1320,\parallel,w}$

- $\beta_{1320,\perp,w} / \beta_{1320,\perp,b}$
- $\beta_{1320,\parallel,w} / \beta_{924,b}$
- $\beta_{1320} / \beta_{924}$

Table 4 displays the confusion matrix obtained when using only these 6 predictor variables, yielding the highest overall accuracy.

4. DISCUSSION AND CONCLUSION

In this contribution, we present an infrared CW optical sensing system able to perform spectral and polarimetric study of free flying mosquitoes. These optical information are used to create predictor variables that can differentiate between species and gender of mosquitoes. The system is used in laboratory conditions, where male and female mosquitoes of three different species, *Ae. albopictus*, *Ae. vexans* and an unknown species of the *Culex* genus, are known a priori and studied from a distance between 3 and 4.25 m. A total of 18 predictors (Table 1) have been studied and their importance for the accuracy of the Bayesian classification evaluated. The wingbeat frequency along with spectral and polarimetric characteristics of mosquitoes contain information that are specific to species and gender of mosquitoes. This study evaluates which ones are the most relevant and what combinations of predictors yields the best accuracy for species and gender classification. In this context, the accuracy of the Naïve Bayes classification was evaluated for different combination of predictors. The average overall accuracy of the classification evolved from 64.34% when all predictors were considered to 74.95% when only the 6 best predictor variables are used, which is an increase of 19.37% from the average overall accuracy using the wingbeat frequency alone. This confirms the benefits of using dual-wavelengths and polarization sensitive lidars in the field of entomology, as stated in Gebru et al. [11].

These predictor variables have a well-defined physical meaning of the kind that could be expected to be different between species and/or gender. Those predictor variables include the already well-known wingbeat frequency, the depolarization ratios of the body and wings of the mosquitoes: $\beta_{1320,\perp,b} / \beta_{1320,\parallel,b}$; $\beta_{1320,\perp,w} / \beta_{1320,\parallel,w}$, the ratio between wings and body backscattering coefficient of the perpendicular polarization at 1320nm: $\beta_{1320,\perp,w} / \beta_{1320,\perp,b}$, and also the ratio between the total backscattering coefficient at both wavelengths: $\beta_{1320} / \beta_{924}$. Still, one unexpected predictor variable turned out to be relevant: $\beta_{1320,\parallel,w} / \beta_{924,b}$ which is the ratio between the wings backscattering coefficient of the parallel polarization for the 1320 nm wavelength and the body backscattering coefficient for the 924 nm wavelength. In this regard, this approach to select the best predictive variables is closer to the methodology used in machine learning where predictive variables are selected not for the physical meaning but solely on their efficiency.

Bayes classification assumes that all predictor variables are completely independent which is not true in this study. This interdependence is known to degrade Bayes performance [23]. In practice, Bayes classifier still performs relatively well [24] so other consideration must be envisioned. Two predictors together may positively contribute to the accuracy of the classification while being unproductive when considered alone. As an illustration, one predictive variable could be very effective at differencing 2 classes but ineffective for all

other 4 while a second predictive variable could behave similarly for 2 other classes. In this case, when both are considered simultaneously, they improve the accuracy by compensating the flaws of the other but when considered alone they appear as contributing negatively. The relation between predictors could be even more complex than this simple illustration. Hence, in the spectrum of all possible predictor variable combinations, finding the best one can be challenging.

For field experiments, more mosquito species may have to be considered, which would result in a larger number of classes, and therefore, in a decrease of the overall accuracy of the classification process. Based on this work, we conclude that the efficiency of multi-wavelengths polarimetric entomological lidar to differentiate mosquito species can significantly be improved by further work on the analysis of transit signals and the study of additional relevant predictive variables.

REFERENCES

- [1]. World Health Organization, [World Malaria Report 2014] (2015).
- [2]. Caraballo H and King K, "Emergency department management of mosquito-borne illness: malaria, dengue, and West Nile virus." *Emerg Med Pract.* 16(5), 1–23 (2014).
- [3]. Eisen L and Eisen RJ, "Need for improved methods to collect and present spatial epidemiologic data for vectorborne diseases", *Emerg. Infect. Dis.* 13(12), 1816–1820 (2007). [PubMed: 18258029]
- [4]. McMeniman CJ, Corfas RA, Matthews BJ, Ritchie SA and Vosshall LB, "Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans", *Howard Hughes Med Inst.* 156, 1060–1071 (2014).
- [5]. Kawada H, Honda S and Masahiro T, "Comparative laboratory study on the reaction of aedes aegypti and aedes albopictus to different attractive cues in mosquito trap", *J. Med. Entomol.* 44(3), 427–432 (2007). [PubMed: 17547227]
- [6]. Crepeau TN, Unlu I, Healy SP, Farajollahi A and Fonseca DM, "Experiences with the large-scale operation of the Biogents Sentinel trap", *J. Am. Mosq. Control Assoc.* 29(2), 177–180 (2013). [PubMed: 23923335]
- [7]. Meeraus WH, Armistead JS and Arias JR, "Field comparison of novel and gold standard traps for collecting Aedes albopictus in Northern Virginia", *J. Am. Mosq. Control Assoc.* 24(2), 244–248 (2008). [PubMed: 18666532]
- [8]. Repasky KS, Shaw JA, Scheppele R, Melton C, Carsten JL and Spangler LH, "Optical detection of honeybees by use of wing-beat modulation of scattered laser light for locating explosives and land mines", *Applied Optics* 45(8):1839–43 (2006). [PubMed: 16572702]
- [9]. Brydegaard M, Gebru A and Svanberg S, "Super resolution laser radar with blinking atmospheric particles – application to interacting flying insects", *Prog. Electromagn. Res.* 147, 141–151 (2014).
- [10]. Genoud AP, Basistyy R, Williams GM and Thomas BP, "Optical remote sensing for monitoring flying mosquitoes, gender identification and discussion on species identification", *Appl. Phys. B*, 124(3), 46 (2018). [PubMed: 30505073]
- [11]. Gebru A, Jansson S, Ignell R, Kirkeby C, Prangma J and Brydegaard M, "Multiband modulation spectroscopy for determination of sex and species of mosquitoes in flight", *Journal of Biophotonics*. 10.1002/jbio.201800014 (2018).
- [12]. Batista GEAPA, Yuan H, Keogh E and Mafra-Neto A, "Towards automatic classification on flying insects using inexpensive sensors", *Proc. 10th International Conference on Machine Learning and Applications and Workshops*, vol. 1, pp. 364–369 (2011).
- [13]. Chen Y, Why A, Batista GEAPA, Mafra-Neto A and Keogh E, "Flying insect classification with inexpensive sensors", *J. Vis. Exp.* 92, 52111 (2014).

- [14]. Potamitis I, Rigakis I, Vidakis N, Petousis M and Weber M, “Affordable bimodal optical sensors to spread the use of automated insect monitoring”, *Journal of Sensors*, vol. 2018, Article ID 3949415 (2018).
- [15]. Potamitis I and Rigakis I, “Measuring the fundamental frequency and the harmonic properties of the wingbeat of a large number of mosquitoes in flight using 2D optoacoustic sensors”, *Appl. Acoust.* 109, 54–60 (2016).
- [16]. Mayagaya VS, Michel K, Benedict MQ, Killeen GF, Wirtz RA, Ferguson HM and Dowell FE, “Non-destructive determination of age and species of *Anopheles gambiae* s.l. using near-infrared spectroscopy”, *Am. J. Trop. Med. Hyg.* 81(4), 622–630 (2009). [PubMed: 19815877]
- [17]. Ratanamahatana CA and Gunopulos D, “Scaling up the naive Bayesian classifier: using decision trees for feature selection”, *Proc. Workshop Data Cleaning and Preprocessing at IEEE Intl Conf. Data Mining* (2002).
- [18]. Russell SJ and Norvig P [Artificial Intelligence A Modern Approach], Third Edition, Pearson Education, Upper Saddle River, chapters 13-14 (2010).
- [19]. Muir LE, Thorne MJ and Kay BH, “*Aedes aegypti* (Diptera: Culicidae) vision: spectral sensitivity and other perceptual parameters of the female eye”, *J. Med. Entomol.* 29(2), 278–281 (1992). [PubMed: 1495042]
- [20]. Gibson G, “A behavioural test of the sensitivity of a nocturnal mosquito, *Anopheles gambiae*, to dim white, red and infra-red light” *Physiol. Entomol.* 20(3), 224–228 (1995).
- [21]. Alvarez JM, Vaughan MA, Hostetler CA, Hunt WH and Winker DM “Calibration technique for polarization-sensitive lidars”, *J. Atmos. Ocean. Technol.* 23, 683–699 (2006).
- [22]. Retzer J, Soofi ES and Soyer R, “Information importance of predictors: Concept, measures, Bayesian inference, and applications”, *Computational Statistics & Data Analysis* 53(6):2363–2377 (2009).
- [23]. John GH, Kohavi R and Pfleger K, “Irrelevant features and the subset selection problem”, *Proc. Eleventh International Conference on Machine Learning*, pp. 121–129 (1994).
- [24]. Hand DJ and Yu K, “Idiot’s Bayes: Not so stupid after all?”, *International Statistical Review*, 69(3), 385–398 (2001).

Figure1.
Optical layout of the infrared system.

Figure 2.

Least square fit based on equation 2 presented in Alvarez et al. [21]. This fit allows for the determination of the calibration constant G between channel $I_{1320,\perp}$ and $I_{1320,\parallel}$ and the depolarization ratio of the gold target used during calibration, δ .

Figure 3.

Raw signals of one mosquito transiting through the laser beams as recorded by the three different channels. The channel for the laser at 1320 nm wavelength and parallel polarization (top left), the channel for the laser at 1320 nm wavelength and perpendicular polarization (top right), the channel for the laser at 924 nm wavelength (bottom left) and all three simultaneously displayed (bottom right).

Figure 4.

Histograms of the predictor $\beta_{1320,\perp,b} / \beta_{1320,\parallel,b}$ (see table 1) for the 6 classes of mosquitoes. The solid black lines are the Fits associated with the Kernel density estimation that are used as the descriptive function of probability for this predictor. In order to make the figure easier to read the normalization was not applied to the Fits displayed in this figure, unlike for the one used in the actual classification.

Table 1.

Predictor notation and explicit meaning.

Simplified expression	Explicit meaning
F_{wb}	Wingbeat frequency
$\beta_{1320, //, b} / \beta_{924, b}$	Ratio between the backscattering coefficient of the mosquito body for the parallel polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito body for the 924 nm wavelength laser
$\beta_{1320, \perp, b} / \beta_{924, b}$	Ratio between the backscattering coefficient of the mosquito body for the perpendicular polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito body for the 924 nm wavelength laser
$\beta_{1320, \perp, b} / \beta_{1320, //, b}$	Ratio between the backscattering coefficient of the mosquito body for the perpendicular polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito body for the parallel polarization of the 1320 nm wavelength laser
$\beta_{1320, //, w} / \beta_{924, w}$	Ratio between the backscattering coefficient of the mosquito wings for the parallel polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito wings for the 924 nm wavelength laser
$\beta_{1320, \perp, w} / \beta_{924, w}$	Ratio between the backscattering coefficient of the mosquito wings for the perpendicular polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito wings for the 924 nm wavelength laser
$\beta_{1320, \perp, w} / \beta_{1320, //, w}$	Ratio between the backscattering coefficient of the mosquito wings for the perpendicular polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito wings for the parallel polarization of the 1320 nm wavelength laser
$\beta_{1320, //, w} / \beta_{1320, //, b}$	Ratio between the backscattering coefficient of the mosquito wings for the parallel polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito body for the parallel polarization of the 1320 nm wavelength laser
$\beta_{1320, \perp, w} / \beta_{1320, \perp, b}$	Ratio between the backscattering coefficient of the mosquito wings for the perpendicular polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito body for the perpendicular polarization of the 1320 nm wavelength laser
$\beta_{924, w} / \beta_{924, b}$	Ratio between the backscattering coefficient of the mosquito wings for the 924 nm wavelength laser and the backscattering coefficient of the mosquito body for the 924 nm wavelength laser
$\beta_{1320, //, w} / \beta_{924, b}$	Ratio between the backscattering coefficient of the mosquito wings for the parallel polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito body for the 924 nm wavelength laser
$\beta_{1320, \perp, w} / \beta_{924, b}$	Ratio between the backscattering coefficient of the mosquito wings for the perpendicular polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito body for the 924 nm wavelength laser
$\beta_{1320, \perp, w} / \beta_{1320, //, b}$	Ratio between the backscattering coefficient of the mosquito wings for the perpendicular polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito body for the parallel polarization of the 1320 nm wavelength laser
$\beta_{924, w} / \beta_{1320, //, b}$	Ratio between the backscattering coefficient of the mosquito wings for the 924 nm wavelength laser and the backscattering coefficient of the mosquito body for the parallel polarization of the 1320 nm wavelength laser
$\beta_{924, w} / \beta_{1320, \perp, b}$	Ratio between the backscattering coefficient of the mosquito wings for the 924 nm wavelength laser and the backscattering coefficient of the mosquito body for the perpendicular polarization of the 1320 nm wavelength laser
$\beta_{1320, //, w} / \beta_{1320, \perp, b}$	Ratio between the backscattering coefficient of the mosquito wings for the parallel polarization of the 1320 nm wavelength laser and the backscattering coefficient of the mosquito body for the perpendicular polarization of the 1320 nm wavelength laser
I_w / I_b	Ratio between the sum of intensities of the wings contribution at both wavelengths by the sum of intensities of the body contribution at both wavelengths
$\beta_{1320} / \beta_{924}$	Ratio between the total backscattering coefficient for the 1320 nm wavelength laser by the total backscattering coefficient for the 924 nm wavelength laser

Table 2.

This table display the actual class of every event versus their predicted class by the Bayesian classifier, also known as confusion matrix, data separated by gender and species. Each actual class have 50 events. All 18 predictors are used in the Bayesian classifier.

		Predicted						
		<i>Albopictus</i>		<i>Culex</i>		<i>Vexans</i>		
		Male	Female	Male	Female	Male	Female	
Actual	<i>Albopictus</i>	Male	35.2 ± 2.30	1.48 ± 0.86	5.43 ± 1.57	2.47 ± 1.09	4.96 ± 1.52	0.50 ± 0.50
		Female	1.74 ± 1.10	28.6 ± 2.96	6.09 ± 1.94	1.46 ± 0.10	0.88 ± 0.79	11.3 ± 2.49
	<i>Culex</i>	Male	1.58 ± 0.57	1.60 ± 0.56	30.9 ± 1.57	5.56 ± 1.02	0	10.3 ± 1.30
		Female	0.48 ± 0.35	1.43 ± 0.95	1.78 ± 1.05	38.6 ± 2.36	1.07 ± 0.82	6.79 ± 1.93
	<i>Vexans</i>	Male	4.71 ± 0.50	0.94 ± 0.23	0.94 ± 0.23	7.55 ± 0.61	30.2 ± 0.83	5.67 ± 0.53
		Female	1.04 ± 0.83	3.16 ± 1.38	3.52 ± 1.47	6.64 ± 1.96	4.21 ± 1.59	31.4 ± 2.77

Table 3.

Predictor contribution to the overall accuracy for all predictor variables.

Predictor variable										
	F_{wb}	$\beta_{1320, //b} / \beta_{924,b}$	$\beta_{1320, \perp b} / \beta_{924,b}$	$\beta_{1320, \perp b} / \beta_{1320, //b}$	$\beta_{1320, //w} / \beta_{924,w}$	$\beta_{1320, \perp w} / \beta_{924,w}$	$\beta_{1320, \perp w} / \beta_{1320, //w}$	$\beta_{1320, //w} / \beta_{1320, //b}$	$\beta_{1320, \perp w} / \beta_{1320, \perp b}$	$\beta_{1320, \perp w} / \beta_{1320, \perp b}$
a (%)	15.4 %	-1.42 %	0.27 %	1.48 %	1.28 %	1.81 %	0.61 %	-0.06 %	-0.32 %	
Predictor variable										
	$\beta_{924,w} / \beta_{924,b}$	$\beta_{1320, //w} / \beta_{924,b}$	$\beta_{1320, \perp w} / \beta_{924,b}$	$\beta_{1320, \perp w} / \beta_{1320, //b}$	$\beta_{924,w} / \beta_{1320, //b}$	$\beta_{924,w} / \beta_{1320, \perp b}$	$\beta_{1320, //w} / \beta_{1320, \perp b}$	I_w / I_b	$\beta_{1320} / \beta_{924}$	
a (%)	-0.23 %	0.93 %	1.09 %	-0.1 %	-0.55 %	-0.81 %	0.06 %	-0.6 %	0.83 %	

Table 4.

Confusion matrix, data separated by gender and species. Each actual class have 50 events. Only the 6 aforementioned predictors are used in the Bayesian classifier.

		Predicted					
		<i>Albopictus</i>		<i>Culex</i>		<i>Vexans</i>	
		Male	Female	Male	Female	Male	Female
Actual	<i>Albopictus</i>	Male	34.4 ± 2.12	0	0.99 ± 0.71	0	9.91 ± 2.01
		Female	0.64 ± 0.57	34.4 ± 2.76	6.04 ± 1.95	0.45 ± 0.29	0.85 ± 0.78
	<i>Culex</i>	Male	0.79 ± 0.41	2.38 ± 0.70	33.3 ± 1.54	0	2.38 ± 0.69
		Female	0.48 ± 0.36	0	0	46.0 ± 1.54	0.48 ± 0.36
	<i>Vexans</i>	Male	7.56 ± 0.36	0.95 ± 0.23	1.89 ± 0.32	0	36.8 ± 0.75
		Female	0	4.89 ± 1.70	2.44 ± 1.24	1.76 ± 1.04	5.27 ± 1.75