Insect sampling biases of automated and manual methods informed by circadian flight activity

By

KIMBERLY GIBSON
THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Horticulture and Agronomy

in the

OFFICE OF GRADUATE STUDEIS

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

____
Dr. Paul Gepts, Chair

____
Dr. Li Tian

____
Dr. Jay Rosenheim

Committee in Charge

2023

Abstract

Monitoring and sampling insects provide critical data for decision making in integrated pest management strategies and plant breeding. Manual sampling and monitoring methods may be biased by interactions between the method, the cropping system, and the size and behavior of the target insects, as well as circadian and seasonal dynamics. These methods may also be limited by the quantity of time and labor required for collection and processing. Novel autonomous insect sensors resolve some of these limitations but may present new challenges. Understanding the limitations and biases of established and novel methods will improve the quality of available data. This study examines the biases of circadian flight activity and seasonal dynamics on water trap, vacuum, and autonomous sensor data with a focus on the crop pest, the Western Tarnished Plant Bug, *Lygus hesperus* (Hemiptera: Miridae Knight) in fields of Lima beans, *Phaseolus lunatus*. The study also includes sensor measurements of *L. hesperus* in a controlled laboratory environment. A strong multimodal circadian effect was observed in sensed and vacuumed *L. hesperus.* The laboratory results also revealed a similar pattern of daytime and crepuscular activity but lacked a midday activity peak observed in the field. Correlations were strong between total sensed insects and water trap data from over the course of the entire season, indicating that the novel autonomous insect sensor could be a good substitute for conventional monitoring methods. More research will be needed to determine the thresholds of sensed insect data that correspond with established measures for integrated pest management and breeding. Consideration of these biases will result in higher quality data that will be useful for advancing the development of insect resistant crop varieties and improving pest management.

Co-authors: K. Gibson[1], L. Sigsgaard[2], L. Still[3*], G. Frascarelli[1,4¥], F. Thoresen[3ŧ], A. Palkovic[1,4], P. Gepts[1], E. Bick[2,3,5]

[1]Department of Plant Sciences, University of California Davis, 1 Shields Avenue, Davis, CA 95616, USA

[2]Section for Organismal Biology, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, DK, Denmark

[3]FaunaPhotonics A/S, Støberigade 14, 2450 Copenhagen SV, Denmark

[4]Department of Agricultural, Food and Environmental Sciences, Marche Polytechnic University, 60131 Ancona, Italy

[5]Department of Entomology and Nematology, University of California Davis, 1 Shields Avenue, 367 Briggs Hall, Davis, CA 95616, USA

Acknowledgements

Thank you to the Godfrey Lab and Grettenberger Lab of UC Davis for loaning the portable aspirator for this work. Thank you to undergraduate interns Marissa Guzman, Treavor H., Christina Saldana, and Edward Sher for helping maintain the lab colony of *Lygus hesperus*. Thanks to Rami El Rashid and Nikolaj Sheller for technical support. Thanks for the Biology team at FaunaPhotonics who labeled much of the species. Thanks to Jesper Lemmich for supervision and FaunaPhotonics A/S for support of the project.

This work was supported by the California Dry Bean Advisory Board, NIFA National Needs Fellowship, CDFA Specialty Crops Block Grant, University of California, Davis Department of Plant Sciences, Università Politecnica Delle Marche in Ancona, Italy, and the Danish Innovation Fund [9066-00051A, 2020].

Table of Contents

1. Introduction

Insect data drives decision making for a wide range of agricultural tasks including integrated pest management (Binns and Nyrop 1992), pesticide development, plant disease vector monitoring, and phenotyping for plant breeding. Primary sources of insect data are generated from sampling (Kuno 1991) and monitoring (Montgomery et al. 2021) with manual and automated methods (Rydhmer et al. 2021).

Sampling for insects results in discrete temporal counts of insects, while monitoring collects accumulated temporal insect data. Common manual sampling methods include sweep netting, plant taping, and vacuuming (McCravy 2018; Hillhouse and Pitre 1974). Manual monitoring methods include water traps, funnel traps, and sticky cards. Automated monitoring is an emerging field that encompasses a variety of sensors including audio recordings, camera traps, and reflected light-based sensors. Each of these methods presents limitations and biases.

Interactions between sampling or monitoring methods, the size and behavior of target insects, and the cropping system can result in biased samples and other adverse effects (Binns and Nyrop 1992). As a consequence, pests may be over-represented while natural enemies are under-represented or vice versa by a given collection method in a certain crop (Bannerman et al. 2015). Additionally, sweep netting can damage crop canopies and over-represent large insects (McEwen and Hervey 1960; Osborne and Allen 1999). Conversely, vacuuming and flower cluster taping may result in samples with an over-representation of small insects and nymphal stages (Rancourt, Vincent, and de Oliveira 2000; Doxon, Davis, and Fuhlendorf 2011). Traps, used for monitoring, have wide-ranging efficacy. The placement, color, shape, and bait may over- or under-represent the prevalence of pests and beneficial insects in the field (Ikemoto et al. 2021; Lewis 1959). In all cases, manual sampling and monitoring take significant time and expertise to count and identify the collected insects.

Novel unsupervised near-infrared sensor technology may save significant time but is not immune to sampling interactions and biases (Rydhmer et al. 2021). These sensors measure insects flying above the crop canopy and therefore fail to capture economically important immature stages. Such sensors may also over-represent some species and will be biased toward measuring larger insects. While these issues may be outweighed by the accuracy and time-saving features of autonomous sensors, it is important that they be understood and accounted for.

Legumes are a globally important family of crops that have many pests (Sharma and et al 2005) but there are few established pest management protocols in grain legumes and successful insect-resistance breeding programs (Clement et al. 2000). Given their adaptation to a wide range of environments, ability to fix nitrogen, and the richness of their grains in proteins and micronutrients, legumes will likely play an important role in the effort to feed a growing human population and adapt to climate change (Mabhaudhi et al. 2019; Cheng et al. 2019; Semba et al. 2021). The historic and ongoing deficiencies in research on plant-insect interactions in legumes, especially in underutilized legume crops like Lima bean, preclude researchers from adequately accounting for sampling biases and sample method interaction with environments (Edwards and Singh 2006; Rubiales et al. 2015). Ensuring that producers and researchers of legumes have the best data with which to manage, study, and develop new varieties of legumes will help ensure the sustainable future of global food security.

The Western Tarnished Plant Bug, *Lygus hesperus* Knight (Hemiptera: Miridae), is a major agricultural pest of multiple crops in the Western United States (Metcalf and Flint 1951). Lygus can cause up to 82% yield loss, as measured between sprayed and unsprayed plots of sensitive legumes like Lima beans, *Phaseolus lunatus* (Bushing, Burton, and Tucker 1974). Sweep netting has been the standard sampling method for economically important IPM and plant breeding decisions for the system of *Lygus hesperus* in Lima beans despite its limitations, including the limited time and special representation (McEwen and Hervey 1960; Bushing, Burton, and Tucker 1974; Gavloski 2018; Hagel 1978). Understanding the circadian and seasonal flight activity of *L. hesperus* in Lima beans will inform research on this system with all sampling methods and improve future research.

To elucidate some of the biases of insect sampling in legumes, we studied circadian flight activity and seasonal dynamics using sensors, water traps, and vacuuming in Lima beans. Specifically, we tested circadian flight activity of *Lygus hesperus* in a laboratory with a near-infrared sensor and in lima bean fields with near-infrared sensors, water taps, and a vacuum. We also tested the seasonal dynamics of *L. hesperus* in a Lima bean field with near-infrared sensors and water traps.

2. Materials and methods

2.1 Vacuum time of day sampling

Lima beans (commercial variety: UC Haskell) were planted at the Plant Sciences Field Facility (PSFF) of the University of California Davis on June 7, 2017, and June 9, 2019 at 38°32'16.9"N 121°47'19.2"W and 38°32'03.3"N 121°46'44.9"W, respectively. The planting consisted of a 70 m long and 4 row wide strip planted with 5-cm between plants. The strip was on the edge of a larger planting of Lima beans, which were drip irrigated and conventionally managed. No insecticides were applied to the strip or the adjacent field.

On July 23, 2017, July 24, 2017, September 9, 2017, and July 30, 2019, eight to 12 samples from randomly selected 6.1-meter-long plots in the strip were obtained using a portable aspirator built from a Stihl SH 86 C-E leaf Shredder vacuum (Stihl Incorporated, Virginia Beach, VA.), operated for one minute per sample. Samples were between four time-windows (6 am – 8 am, 10 am to 12 pm, 2 pm – 4 pm, and 6 pm – 8 pm) for each sampling date. The collected adult *L. hesperus* insects were bagged, frozen, and then counted. Plots were not resampled.

2.2 Circadian flight activity

2.2.1 Laboratory evaluation

Collection of data occurred at the University of California, Davis in the spring of 2021. To evaluate circadian flight activity in a controlled environment, a *L. hesperus* colony founded by individuals collected from Lima bean and alfalfa fields in the fall of 2019 and 2020 was used. The colony insects were maintained at 20°C and a 12 h photoperiod. Adults were held in 30.5cm cube collapsible cage (BioQuip Products, Rancho Dominguez, CA) with a water-soaked organic cotton round (Swisspers®, Parkdale Inc., Gastonia, NC), hulled sunflower seeds, and fresh organic green beans supplied three times a week. Beans with *L. hesperus* eggs were moved to rearing tubs where *L. hesperus* nymphs emerged in approximately 7-10 days. Nymphs were supplied with green beans three times per week and moved to adult cages when they reached the adult stage.

An autonomous near-infrared insect sensor (described in Rydhmer et al. 2022, FunaPhotonics ApS., Copenhagen SV, Denmark) with a 17.5 L measurement volume monitored one week old male and female *L. hesperus* from the colony and field-collected from alfalfa (38°31'28.2"N 121°46'18.4"W). *Lygus* flights were recorded in a label cage (a black neoprene walled cage with floor strip LED lighting on a timer and an attached sensor as described in Bick, Edwards and Licht, 2021) with the top of the cage replaced by mesh. Flight data consisted of a one-dimensional matrix of the intensity of backscattered light resulting from an insect flight. A nearby window provided additional light with a photoperiod ranging from 14h : 24m to 14h : 52m, and LED light strips placed along the cage floor were illuminated 24 hours a day. Measurements for a total of 1415 flights were collected and analyzed. During the measurement period, fresh organic green beans, hulled sunflower seeds, and a water-soaked cotton round were available to the insects in the cage.

2.2.2 Machine learning species classification

Machine learning algorithms can classify insects to species (Kirkeby et al. 2021). Light backscattered by an insect flight is recorded as a 1-dimensional time series of light intensity (Rydhmer et al. 2021; Bick, Edwards, and de Fine Licht n.d.), termed 'events.' A machine learning algorithm was trained, validated, and tested on sensor-recorded flights from 14 insect groups. Specifically, we trained the model to compare *L. hesperus* Knight (Hemiptera: Miridae) to the insects listed in Table 1, and non-insect noise signals.

Ten thousand random events per group were fed into an 8-layer Convolutional Neural Network (CNN) (80% training, 10% validation, and 10% test sets) implemented in Python 3.7 using Keras, set up as specified in (Bick, Edwards, and Licht 2021). A t-SNE (t-distributed stochastic neighbor embedding), implemented in Python 3.7 using the sklearn package, was used on the trained features of the test set to visualize the separation between the statistical representations of insects within each group (Fig. 1). The algorithm was then applied to field data to extract events classified as *L. hesperus.*

Figure 1. t-SNE plot showing the separation of insect flight events in the test set in the learned feature space. Different insect groups are represented by different colors to aid interpretation.

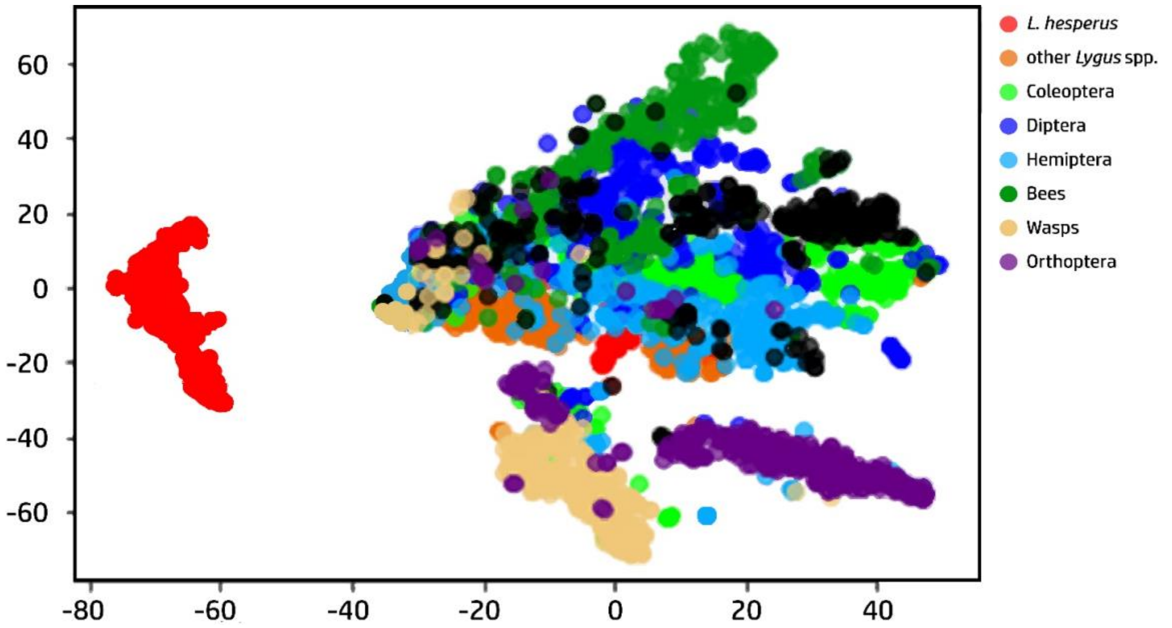


Table 1: List of insect species whose sensor recorded flights were used for model training

Scientific Name	Common Name	Order
Acleris comariana	Strawberry Tortrix moth	Lepidoptera
Adalia bipunctata	two spotted ladybird	Coleoptera
Adoxophyes orana	summer fruit tortrix	Lepidoptera
Aedes aegypti	yellow fever mosquito	Diptera
Agelastica alni	alder leaf beetle	Coleoptera
Aleochara bilineata		Coleoptera
Aleyrodes proletella	cabbage whitefly	Hemiptera
Andrena vaga	grey-backed mining bee	Hymenoptera
Aphelinus abdominalis		Hymenoptera
Aphidius colemani		Hymenoptera
Aphidius matricariae		Hymenoptera
Aphidius ervi		Hymenoptera
Aphidoletes aphidimyza	gal midge	Diptera
Aphis fabae	black bean aphid	Hemiptera
Aphis gossypii	cotton aphid	Hemiptera
Apis mellifera	European honey bee	Hymenoptera
Aulacorthum solani	foxglove aphid	Hemiptera
Bombus impatiens	common eastern bumble bee	Hymenoptera
Bombus pascuorum	common carder bee	Hymenoptera
Bombus terrestris	buff-tailed bumblebee	Hymenoptera
Brassicogethes aenus	pollen beetles	Coleoptera
Brevicoryne brassicae	cabbage aphid	Hemiptera
Ceutorhynchus obstrictus		Coleoptera
Ceutorhynchus pallidactylus		Coleoptera
Cladius pectinicornis		Hymenoptera
Cryptolaemus montrouzieri		Diptera
Cydia pomonella	codling moth	Lepidoptera
Dacnusa sibirica		Hymenoptera
Dalotia coriaria	greenhouse rove beetles	Coleoptera
Dasineura brassica	brassica pod midge	Diptera
Delia antiqua	onion fly	Diptera
Diplolepis rosae		Hymenoptera
Drosophila melanogaster	fruit fly	Diptera
Drosophila suzukii	spotted wing drosophila	Diptera
Encarsia formosa		Hymenoptera
Episyrphus balteatus	marmalade hoverfly	Diptera
Eretmocerus eremicus		Hymenoptera
Eupedes corrolae		Diptera
Feltiella acarisuga		Diptera
Graphosoma italicum	European striped shield bug	Hemiptera
Helicoverapa armigera	cotton bollworm	Lepidoptera
Hydrotaea aenescens	black dump fly	Diptera
Hypera meles	clover head weevil	Coleoptera
Ischnura elegans	blue-tailed damselfly	Odonata
Leptomastix dactylopii		Hymenoptera
Lucilia sericata	common green bottle fly	Diptera
Lygorcoris pabulinus	common green capsid	Hemiptera
Musca domestica	house fly	Diptera
Myzus persicae	Peach potato aphid	Hemiptera
Orius laevigatus		Hemiptera
Orius majusculus		Hemiptera
Osmia rufa		Hymenoptera
Plodia interpunctella	Indianmeal moth	Lepidoptera
Plutella xylostella	diamondback moth	Lepidoptera
Popillia japonica	Japanese beetle	Coleoptera
Rhagonycha fulva	common red soldier beetle	Coleoptera
Scaphoideus titanus	American grapevine leafhopper	Hemiptera
Scaptomyza flava	turnip leafminer	Diptera
Spalangia cameroni		Hymenoptera
Sphaerophoria rueppellii		Diptera

2.2.3 Sensor field evaluation

On May 15, 2021, the commercial variety of Lima bean, UC Haskell, was planted at the PSFF (38°32'18.8"N 121°47'19.9"W). The planting consisted of a 42 meter long and 4 row wide strip planted with 5 cm spacing. The strip was on the edge of a larger planting of lima beans that were drip irrigated and conventionally managed. No insecticides were applied to the strip or the adjacent field.

The sensor was placed on the edge of the planting, 28 meters in from the end of the strip. The sensor's field of view was trained just above canopy height over the four rows of the planting (Fig. 2). The height of the sensor was adjusted as needed on a weekly basis to maintain consistent height above the canopy. The sensors measured insect flights in this location continuously from the 17th of June 2021 to the 1st of September 2021. Flowering started on the 7th of July 2021.

Figure 2: Autonomous near-infrared insect sensors (FaunaPhotonics ApS., Copenhagen SV, Denmark) installed over Lima Beans at the PSFF in June 2021. Data from the one sensor observing insects above the variety UC Haskell were used for this study.

2.3 Seasonal dynamics

Seasonal dynamics of *L. hesperus* were evaluated with both water traps and the sensor. From July 2, 2021, to August 28, 2021, two green water traps made from 37oz green plastic cereal bowls (Room Essentials ™, Target Corporation, Minneapolis, MN.) mounted with black zip ties on bamboo stakes were positioned in the strip planting of UC Haskell lima beans described above. The water traps containing water and a drop of dish soap to reduce surface tension were positioned at the level of the crop canopy and repositioned as needed on a weekly basis to retain this relative height. Every three days, the water traps were emptied and strained through a mesh sieve. The collected insects were then transferred to vials of ethanol, identified, and counted.

2.4 Analysis

Data visualizations were conducted in Python 3.7 using the matplotlib package. To quantify correlations between laboratory and field circadian flight activity, and between predicted *L. hesperus* and general insect activity in the sensor, the Kolmogorov-Smirnov test was run using the SciPy stats package (Virtanen et al. 2020).

3. Results

3.1 Vacuum time of day sampling

The *L. hesperus* densities approximated a normal distribution with the greatest proportion occurring between 12:00 to 14:00 in Pacific Daylight Time. There was a 2.8-fold difference seen between the highest and lowest temporal blocks (12:00 to 14:00 and 6:00 to 8:00, respectively) (Fig. 3).

3.2 Circadian flight activity

3.2.1 Laboratory evaluation

In the laboratory evaluation, the sensor recorded a total of 404 *L. hesperus* flights over 17 days of monitoring. Flights were distributed in two large and two small peaks. The large peaks occurred first between 04:00 and 05:00 with 44 flights and the second between 21:00 and 22:00 with 51 flights. The small peaks occurred between 11:00 and 12:00 with 19 flights and between 16:00 and 17:00 with 15 flights (Fig. 4).

Figure 3. Overlapping histograms of Lygus hesperus (1) sensed and (2) vacuumed. Insect data is averaged per hour and presented graphically in two-hour blocks

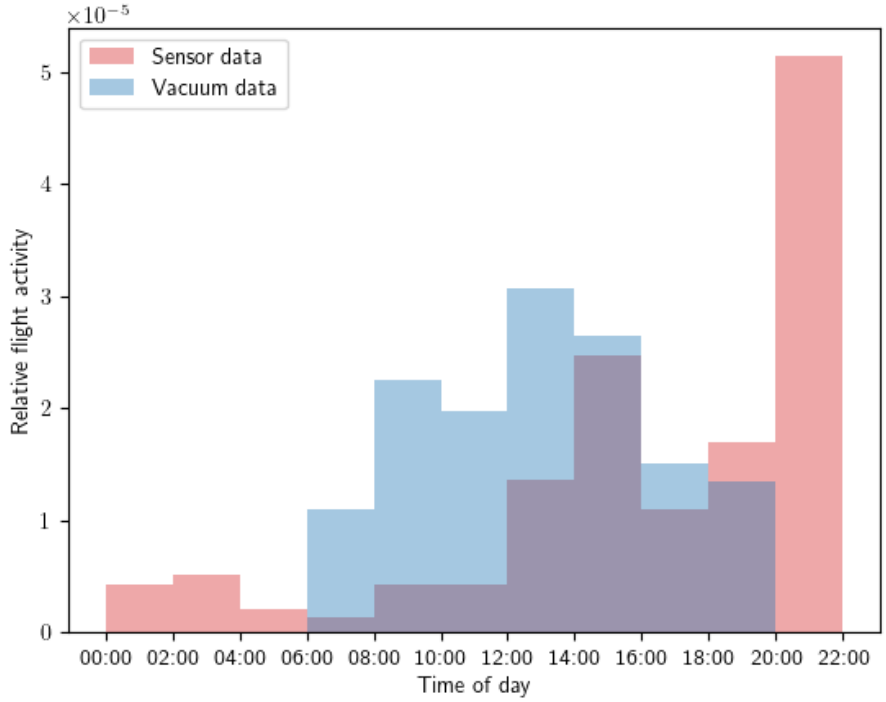
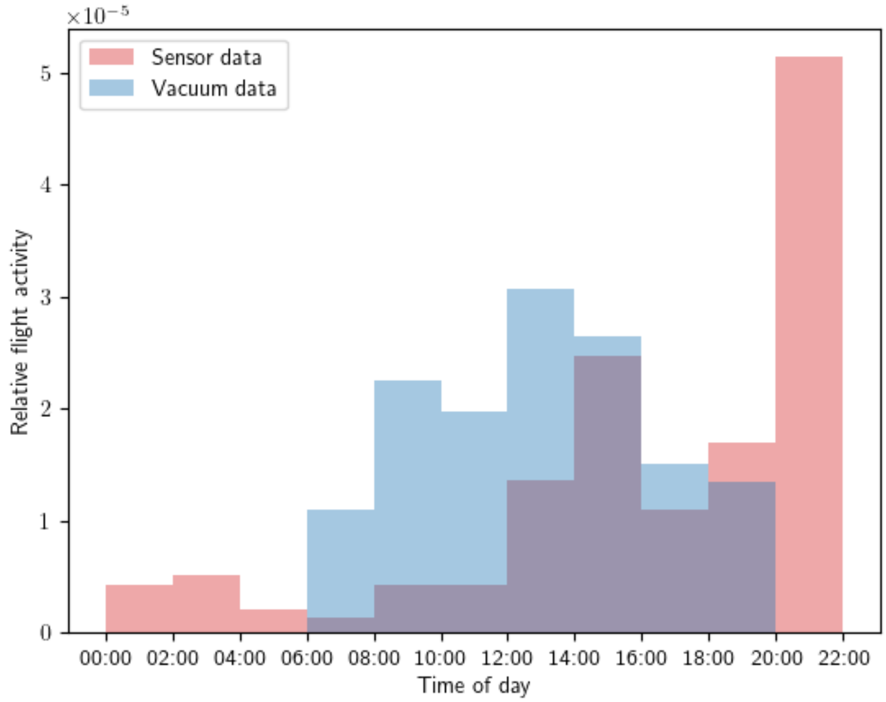


Figure 4. Overlapping histograms of sensed Lygus hesperus from the field (1) and sensed Lygus hesperus from the lab (2). Insect data is averaged per hour and presented graphically in one-hour blocks.



3.2.2 Machine learning species classification

The machine learning algorithm was evaluated using an unseen test set consisting of 229 *L. hesperus* events and 3535 events from other insects, including 194 European *Lygus* spp. events (meaning *Lygus* that were not *L. hesperus*). When applied to the test set, the algorithm correctly identified 90.0% (206) of *L. hesperus* events (Fig. 10). The test observed one Type 1 error, also known as false negative, in which 0.03% *L. hesperus* events classified as another species. The test observed 23 Type II errors, also known as false negative, in which 10.0% of *L. hesperus* events identified as other species, of which 13 were classified as European *Lygus* spp. The t-SNE analysis of the feature space (Fig. 10) shows that *L. hesperus* events are fully separated from other insects, further indicating that training was successful in regard to the target group. However, more work is needed to ensure that insect events classified as '*L. hesperus*' are indeed those insects.

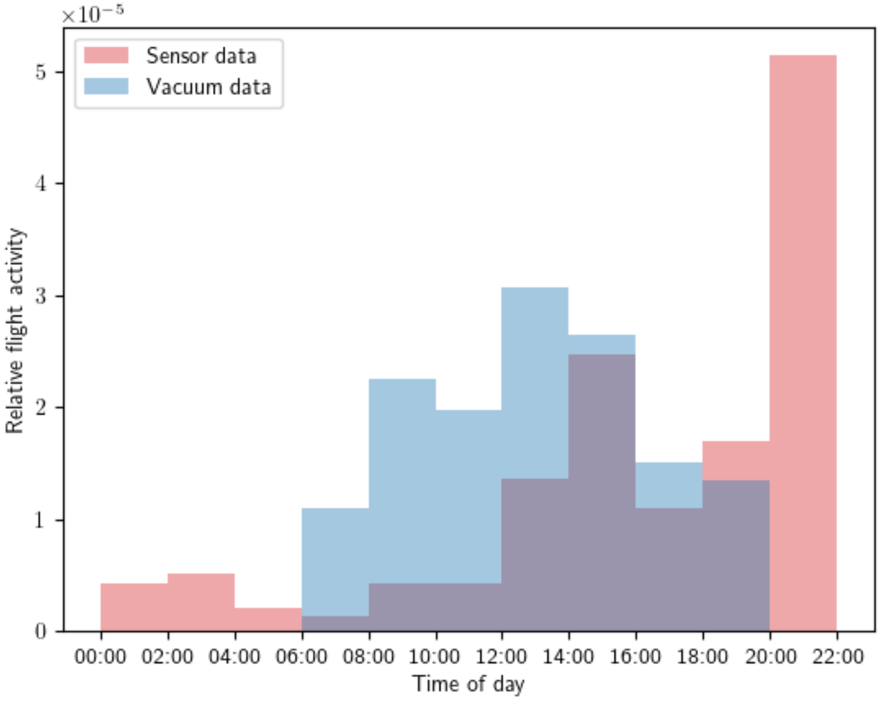
3.2.3 Sensor field evaluation

In the field, the sensor recorded 119,664 insect flights of which 87 were classified at a greater than 0.75 likelihood as *L. hesperus*. All insect flights were distributed between one large, one broad, and one small peak (Fig. 14). The large peak was narrow-shaped and occurred between 20:00 and 21:00 with 18,673 insect flights. The broad peak occurred between 14:00 and 15:00 with 11,302 flights. The small peak occurred between 02:00 and 03:00 with 3,728 flights.

L. hesperus-classified flights were distributed in three large and one small peaks. The large peaks occurred between 09:00 and 10:00 with eight flights, 13:00 and 14:00 with nine flights, and between 20:00 and 21:00 with 11 flights. The small peak was uniformly distributed between 01:00 and 04:00 am with four flights per hour.

A distribution analysis using a Kolmogorov-Smirnov test indicated that insects captured using the vacuum method did not exhibit a significantly different distribution to flying *L. hesperus* identified by the sensor, D(7) = 0.57, p=0.21), indicating that field results were consistent across years and methods.

Figure 5. Overlapping histograms of sensed Lygus hesperus (1) and all sensed insects (2). Insect data is averaged per day and presented graphically in one-day blocks.



3.3 Seasonal dynamics

The total number of recorded insect flights increased consistently, albeit with fluctuations, over the season, with a low of 479 insect events recorded over the 21st/22nd of June 2021, and a high of 6,240 insect events recorded over the 26th/27th of August 2021 (Fig. 5).

Insect events classified as *L. hesperus* also showed an increase in activity over the season, generally exhibiting very low (<5 per two-day period) activity, with a peak of 4 *L. hesperus* flights on the 15th/16th July 2021, until the end of August 2021, where an exponential-like increase was observed concluding with the highest peak of 28 flights on the 28th/29th August.

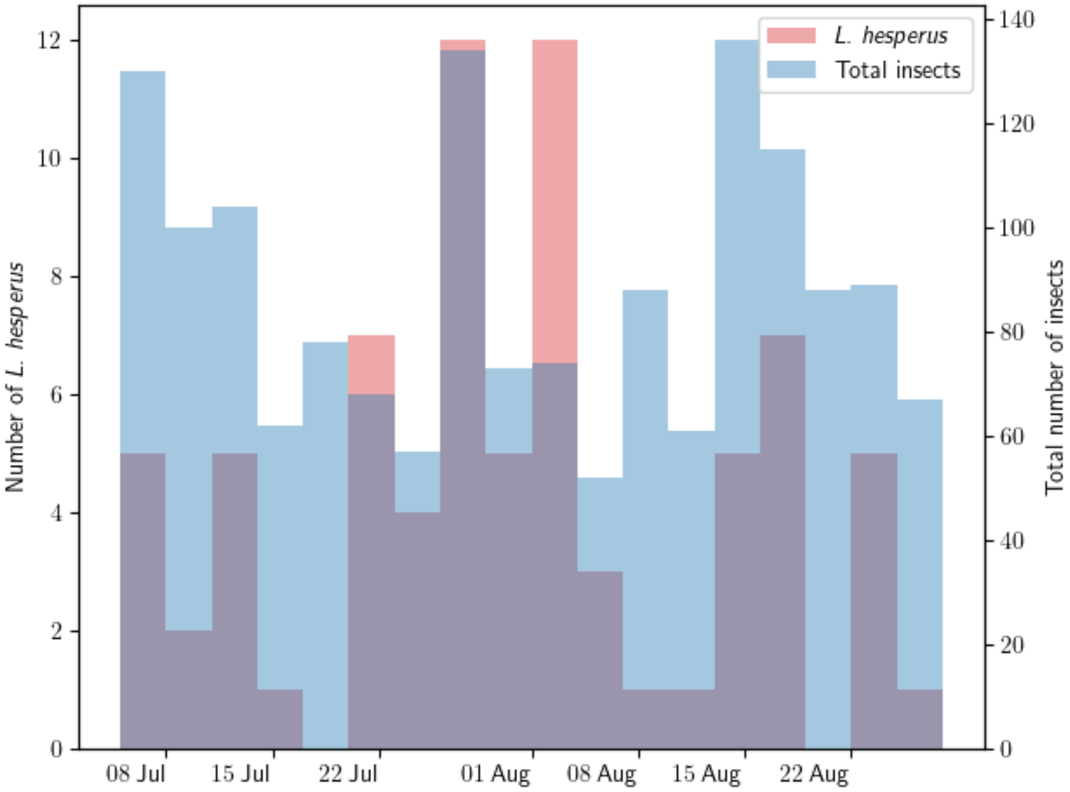
A comparison of the flight activity distribution of insects detected by the sensor between 25th June 2021, and the 27th August, 2021, determined using the Kolmogorov-Smirnov test that the distribution of *L. hesperus* activity and the general insect population over the season differ significantly, $D(38) = 0.63$, $p = 1.99e-07$.

Water traps identified a total of 1610 insects, of which 77 were *L. hesperus* (Fig. 6.). The general insect population showed three peaks within this time, one early in the season with 130 insects collected between the 3rd to the 5th of July 2021, a narrow second peak of 134 insects between the 26th and the 28th July, 2021, and a final peak of 136 insects between the 13th and the 15th of August, 2021.

A Kolmogorov-Smirnov test comparing the distribution of *L. hesperus* and the general insect population in water traps did not find a significant difference, $D(18) = 0.39$, $p = 0.13$.

In comparing water traps and sensor data, using a Kolmogorov-Smirnov test, no significant difference in distribution was found in the general insect population, D(18) =0.11, p = 1.0; however, a significant difference was observed in the distribution of *L. hesperus* caught in water traps compared to those observed in the sensor, D(18) = 0.5, p = 0.02.

Figure 6. Overlapping histograms of the total number of insects and the L. hesperus collected using water traps.



4. Discussion

A clear circadian rhythm is observed in both the vacuum samples and sensed flights classified as *L. hesperus*; hereafter termed 'sensed *Lygus*'. This indicates that circadian rhythm synchronizes both the activity on plants and in the air, resulting in a correlation between vacuum and sensor measurements over time (3.2.3). As the vacuum sampling has greater access to insects on the outer canopy and the sensor only measures insects in flight, this synchronization seems to be a result of the insect behavior that leads to both movement to the outer canopy and flight. This movement could be driven by access to a more optimal microclimate, need for food, or mate seeking. A previous study examining *L. hesperus* flight periodicity found the insects had primarily daytime and crepuscular flight activity (Blackmer, Naranjo, and Williams 2004). While this reported periodicity was observed for the sensed *Lygus* in the controlled environment, the field sensed *Lygus* also included increased peak flight activity from 12:00 to 13:00. As this period of increased circadian activity is observed in both sets of field data but obtained in different years (vacuum sampling and sensed *Lygus*), it is likely driven by field abiotic conditions. *Lygus* spp. are sensitive to both light and temperature (Blackmer, Naranjo, and Williams 2004; Spurgeon and Cooper 2012) with *L. hesperus* flights in response to light (Blackmer, Naranjo, and Williams 2004). The closely related species *Lygus lineolaris* (Hemiptera: Miridae) in strawberries also exhibited three periods of circadian flight activity in the field, matching the timeframes of field sensed *Lygus* (Bick 2019). This indicates findings are likely applicable to other agricultural ecosystems.

Insect population estimates based on sampling and monitoring drive decisions (Dornelas and Daskalova 2020; Didham et al. 2020; Greenwood et al. 2005; Dangles and

Casas 2019). However, sampling and monitoring methods are sensitive to circadian rhythm, increasing variation in insect counts. There was a 2.8-fold increase in vacuumed *L. hesperus* between the lowest and highest temporal periods (3.1). The sensor was even more sensitive to *Lygus* temporal dynamics, with some time periods detecting no *Lygus* flights. These findings extend Rancourt, Vincent and de Oliveira, 2000's recommendation that sampling of *L. lineolaris* should account for flight activity of other *Lygus* spp. Moreover, insect circadian flight patterns have been observed in many insect orders including Lepidoptera (Wikström, Milberg, and Bergman 2009), Diptera (Zahn and Gerry 2020), Coleoptera (Mauchline et al. 2017), Hemiptera (Rancourt, Vincent, and de Oliveira 2000; Sétamou et al. 2012), and many others. Therefore, generally sampling practices of flying insects should account for the target's circadian flight activity.

No significant difference was seen between the seasonal distribution of sensed insects and total insects collected in the water traps. This supports the use of sensors as a suitable equivalent method in measuring general insect population (Rydhmer et al. 2021). A significant difference in seasonal dynamics was identified between *L. hesperus* collected in water traps and sensed *Lygus*, which could indicate that these two monitoring methods are observing distinct behaviors, and thus may be more suitable as complementary methods. However, it is relevant to note that previous work using these sensors (Rydhmer et al. 2021) has indicated that there is an expected delay on the order of days between flights observed by the sensor and insects collected by the water traps. As the water traps were concluded on the same day the sensor recorded the highest peak, it is possible that had the water traps been continued, a later peak could have been identified. Further research is needed to more explicitly determine the relationship between sensed flights and manual methods.

In both the water trap and the sensor, insect observations form multi-modal peaks with high activity periods followed by periods with very low or no activity. The presence of multi-modal peaks in both water traps and the sensor could be reflective of the known cycles of *L. hesperus* population dynamics due to having multiple in-field generations, whereby breeding cycles result in cyclical periods of high adult populations followed by periods dominated by nymphal stages (Leigh 1963). This breeding cycle is known to conclude with an exponential rise in population (Bick 2019), which can be observed in the sensor but not in the water traps.

A major limitation of this study is the uncertainty in accuracy of classifying flights detected by the sensor as *L. hesperus.* Specifically, light detection and ranging (LiDAR) sensors have limited capacity to validate the species classifications of machine learning algorithms in the field, due to their lack of trapping individuals in sync with signal events (with a camera or otherwise). However, the high accuracy of the algorithm to separate out *L. hesperus* from other *Lygus* spp. as well as the low occurrence of Type I and Type II errors (3.2.2) indicates the classifying algorithm has a high likelihood of success. Moreover, the alignment of the circadian rhythm in the vacuumed and sensed *Lygus* (Fig. 3, 3.2.2) increases the confidence in the likelihood that sensed flights are accurately classified as *L. hesperus*. Finally, the proportion of sensed insects classified as *L. hesperus* is low, as would be expected in an environment where *L. hesperus* do not comprise the majority of the insect population and varies independently from the seasonal dynamics of all sensed insect flights

(Fig. 5), indicating that the algorithm is not indiscriminately classifying an arbitrary subset of insects as *L. hesperus*.

Another study limitation is the low number of *L. hesperus* flights. This phenomenon occurred in the both the controlled environment and the field (3.2.1, 3.2.3). This data aligns with a flight mill study where 60-86% of *L. hesperus* were classified as nonflying and 60 - 97% of the flying individuals made only 'trivial flights'(Blackmer, Naranjo, and Williams 2004). As the sensor is only capable of detecting insects in flight, the low numbers of flight events highlight the difficulty of an optical sensor's application to reluctantly flying insects. The results obtained thus help to explore the limits of the sensor when applied to diverse species and behavioral patterns.

There are many potential applications of optical sensors to entomological problems. As circadian rhythm can be used to differentiate insects within a species by sex (houseflies and *L. lineolaris*), sensor data may be able to do so as well. Mosquitoes' sex can be identified by their wing beat frequency. A machine learning algorithm might be able to identify insects within species by sex with less obvious differentiation.

This study focused primarily on a single species, *L. hesperus*, contrasted with the general undifferentiated insect population. However, as indicated by the t-SNE plot (Fig. 1), using sensor data it is possible to distinguish multiple groups, and potentially to quantify group separation. This raises opportunities for using the sensor to measure biodiversity or to identify functional groups.

While this study only utilized a single sensor in the field, a network of sensors placed across a field, or a landscape, may be able to counter some of the limitations presented when monitoring low-flight insects. A network of sensors may also be able to provide more detailed information about the spatial and temporal dynamics of insects.

The sensor is automated and standardized, which could make it a good tool for long-term commercial agricultural monitoring. Although the sensor's insect counts appear to correlate well with manual methods, it uses a unique metric (number of observed flights), which is not associated with an economic threshold. Further work is therefore required to develop a useful threshold from observed flights. Regardless of method, researchers should be consistent in the timing of sample collection or should use time of day to normalize counts since insect counts are often highly variable at different times of the day and season. For manual sampling methods, in which personnel are required to be present in the field, sampling during the midday peak may result in a higher insect count and thereby more easily meet the requirement for insecticide treatment or scoring as insect-infested in a breeding program. The threshold for insecticide application in Lima beans is quite low at 1.0-2.0 *L. hesperus* per sweep (Long et al. 2020). For automated methods, understanding the circadian flight activity could assist with the timing of non-continuous monitoring schemes for power saving. For example, if power supply limited sensor operation to a few hours a day, the optimal times to operate the sensor would be during the early morning, late evening, and midday peaks.

5. Conclusion

The understanding that circadian flight activity results in biased manual and automated sampling and monitoring will inform the use and collection of insect data. Additionally, confirming the utility of automated monitoring and beginning to explore its use in a variety of fields will greatly increase the effectiveness and efficiency of data collection. Here are many applications for this technology including as an early warning system for invasive species, a source of disease detection, and in IPM.

One of many potential uses for this technology best explored in this study is its use for research and specifically the development of insect tolerant crop varieties of legumes. To date, few breeding programs have successfully implemented breeding for arthropod field pest resistance in legumes (Edwards and Singh 2006). Sensors could be applied to improving phenotyping of arthropod field pest resistance traits. Specifically, continuous monitoring allows for analysis of insect behavior that will enable breeders to understand the mechanisms of resistance while also increasing the throughput. Current breeding programs predominantly identify field insect resistance through yield trials, manual sampling, and visual scoring. These methods may succeed when there is a clear qualitative, monogenic resistance but do not provide detailed enough phenotypes to differentiate genotypes by their quantitative resistance and, hence, capitalize on modern breeding techniques for complex quantitative traits. With more precise phenotype data, breeders could apply advanced techniques like marker -assisted selection, genomic selection, and interspecific hybridization (Rubiales et al. 2015). With better insect data for phenotyping, more rapid advances could be made in breeding legumes that are not reliant on insecticides, thereby meeting growing consumer demand for organic and less environmentally destructive agricultural systems.

References

Bannerman, J. A., Costamagna, A. C., McCornack, B. P., & Ragsdale, D. W. (2015). Comparison of relative bias, precision, and efficiency of sampling methods for natural enemies of soybean aphid (Hemiptera: Aphididae). *Journal of Economic Entomology* 108.3, 1381–97. doi:10.1093/JEE/TOV009.

Bick, E. (2019). Plant-insect interactions to promote pest management in strawberry and water hyacinth ecosystems. PhD Dissertation, University of California, Davis.

Bick, E., Edwards, S., & de Fine Licht, H. H. (2021). Detection of insect health with deep learning on near-infrared sensor data. BioRxiv. doi:10.1101/2021.11.15.468635.

Binns, M. R., & Nyrop, P. (1992). Sampling insect populations for the purpose of IPM decision making. *Annual Review of Entomology* 37.1, 427-453. doi: 10.1146/annurev.en.37.010192.002235

Blackmer, J. L., Naranjo, S. E., & Williams, L. H. (2004). Tethered and untethered flight by *Lygus hesperus* and *Lygus lineolaris* (Heteroptera: Miridae). *Environmental Entomology* 33.5, 1389–1400. doi: 10.1603/0046-225X-33.5.1389.

Bushing, R., Burton, V., & Tucker, C. (1974). Dry large Lima beans benefit from Lygus bug control. *California Agriculture* 28.5, 14–15.

Cheng, A., Raai, M. N., Zain, N. A. M., Massawe, F., Singh, A., & Wan-Mohtar, W. A. A. Q. I. (2019). In search of alternative proteins: unlocking the potential of underutilized tropical legumes. *Food Security* 11.6, 1205–15. doi: 10.1007/s12571-019-00977-0.

Clement, S. L., Wightman, J. A., Hardie, D. C., Bailey, P., Baker, G., & McDonald, G. (2000). Opportunities for integrated management of insect pests of grain legumes. In: Knight, K. (Ed.), Linking Research and Marketing Opportunities for Pulses in the 21st Century, Kluwer Academic Publishers, Adelaide, Australia, pp. 467–4480. doi:10.1007/978-94-011-4385-1_43.

Dangles, O., & Casas, J. (2019). Ecosystem services provided by insects for Achieving sustainable development goals. *Ecosystem Services* 35, 109–15. doi:10.1016/j.ecoser.2018.12.002

Didham, R. K., Basset, Y., Matilda Collins, C., Leather, S. R., Littlewood, N. A., Menz, M. H. M., Müller, J., Packer, L., Saunders, M. E., Schönrogge, K., Stewart, A. J. A., Yanoviak, S. P., Hassall, C. (2020). Interpreting insect declines: seven challenges and a way forward. *Insect Conservation and Diversity.* 13.2, 103–14. doi:10.1111/ICAD.12408.

Dornelas, M., & Daskalova, G.N. (2020). Nuanced Changes in Insect Abundance. *Science* 368.6489, 368–69. doi:10.1126/science.abb6861.

Doxon, E. D., Davis, C. A., & Fuhlendorf, S. D. (2011). Comparison of two methods for sampling invertebrates: vacuum and sweep-net sampling. *Journal of Field Ornithology* 82.1, 60–67. doi: 10.1111/j.1557-9263.2010.00308.x.

Edwards, O., & Singh, K. B. (2006). Resistance to insect pests: what do legumes have to offer? *Euphytica*. 147.1, 273–85. doi: 10.1007/s10681-006-3608-1

Gavloski, J. (2018). Integrated management of sap feeding insects of pulse crops. *Annals of the Entomological Society of America* 111.4, 184–94. doi: 10.1093/aesa/say008

Greenwood, B. M., Bojang, K., Whitty, C. J. M., Targett, G. A. T. (2005). Malaria. *Lancet*. 365.9469, 1487–98. doi: 10.1016/S0140-6736(05)66420-3.

Hagel, G. T. (1978). *Lygus spp.*: damage to beans by reducing yields, seed pitting, and control by varietal resistance and chemical sprays. *Journal of Economic Entomology* 71.4, 613–15. doi:10.1093/JEE/71.4.613.

Hillhouse, T. L., & Pitre, H. N. (1974). Comparison of Sampling Techniques to Obtain Measurements of Insect Populations on Soybeans. *Journal of Economic Entomology*. 67.3, 411–14. doi:10.1093/JEE/67.3.411.

Ikemoto, M., Kuramitsu, K., Sueyoshi, M., Seguchi, S., & Yokoi, T. (2021). Relative trapping efficiencies of different types of attraction traps for three insect orders in an agricultural field. *Applied Entomology and Zoology* 56.3, 393–405. doi: 10.1007/s13355-021-00748-4

Kirkeby, C., Rydhmer, K., Cook, S. M., Strand, A., Torrance, M. T., Swain, J. L., Prangsma, J., Johnen, A., Jensen, M., Brydegaard, & M., Græsbøll, K. (2021). Advances in automatic identification of flying insects using optical sensors and machine learning. *Scientific Reports*. 11.1, 1555. doi:10.1038/s41598-021-81005-0.

Kuno, E. (1991). Sampling and analysis of insect populations. *Annual Review of Entomology* 36, 285-304. doi: 10.1146/annurev.en.36.010191.001441

Leigh, T. F. (1963). Life history of *Lygus hesperus* (Hemiptera: Miridae) in the laboratory. *Annals of the Entomological Society of America* 56.6, 865–67. doi:10.1093/AESA/56.6.865.

Lewis, T. (1959). A comparison of water traps, cylindrical sticky traps and suction traps for sampling Thysanopteran populations at different levels*. Entomologia Experimentalis et Applicata* 2.3, 204–15. doi:10.1111/J.1570-7458.1959.TB00434.X.

Long, T., Hu, C., Wang, R., Zhang, T., Kong, S., Li, W., Cai, J., Tian, W., Zeng, T., 2020. Entomological Radar Overview: System and Signal Processing. IEEE Aerosp. Electron. Syst. Mag. 35.1, 20–32. doi:10.1109/MAES.2019.2955575.

Mabhaudhi, T., Chimonyo, V. G. P., Hlahla, S., Massawe, F., Mayes, S., Nhamo, L., & Modi, A. T. (2019). Prospects of orphan crops in climate change. *Planta* 250.3, 695–708. doi:10.1007/s00425-019-03129-y

Mauchline, A. L., Cook, S. M., Powell, W., Chapman, J. W., & Osborne, J. L. (2017). Migratory flight behaviour of the pollen beetle *Meligethes Aeneus*. Pest Management Science 73.6, 1076–82. doi:10.1002/PS.4550.

McCravy, K. W. (2018). A review of sampling and monitoring methods for beneficial arthropods in agroecosystems. *Insects* 9, 170. doi: 10.3390/insects9040170.

McEwen, F. L., & Hervey. G. E. R. (1960). The effect of Lygus bug control on the yield of Lima beans. *Journal of Economic Entomology* 53,513–16. doi:10.1093/JEE/53.4.513.

Metcalf, C. L. & Flint, W. P. (1951). Destructive and useful insects (3rd edition). New York, McGraw-Hill Book Co. Inc., 1071 pp. doi: 10.1086/334020

Montgomery, G. A., Belitz, M. W., Guralnick, R. P., & Tingley, M. W. (2021). Standards and best practices for monitoring and benchmarking insects. *Frontiers in Ecology and Evolution* 8, 513-513. doi:10.3389/FEVO.2020.579193/BIBTEX.

Osborne, K. H., & Allen, W. W. (1999). Allen-Vac: an internal collection bag retainer allows for snag-free arthropod sampling in woody scrub. *Environmental Entomology* 28.4, 594–96. doi:10.1093/EE/28.4.594.

Rancourt, B., Vincent, C., & de Oliveira, D. (2000). Circadian activity of *Lygus lineolaris* (Hemiptera: Miridae) and effectiveness of sampling techniques in strawberry fields. *Journal of Economic Entomology* 93.4, 1160–66. doi:10.1603/0022-0493-93.4.1160.

Rubiales, D., Fondevilla, S., Chen, W., Gentzbittel, L., Higgins, T. J. V., Castillejo, M. A., Singh, K.B., & Rispail, N. (2015). Achievements and challenges in legume breeding for pest and disease resistance. *Critical Reviews in Plant Sciences* 34, 195–236. doi:10.1080/07352689.2014.898445.

Rydhmer, K., Bick, E., Still, L., Strand, A., Luciano, R., Helmreich, S., Beck, B. D., Grønne, C., Malmros, L., Poulsen, K., Elbæk, F., Brydegaard, M., Lemmich, J., Nikolajsen, T. (2022). Automating insect monitoring using unsupervised near-infrared sensors. *Scientific Reports* 12.1, doi:10.1038/s41598-022-06439-6

Semba, R. D., Ramsing, R., Rahman, N., Kraemer, K., & Bloem, M. W. (2021). Legumes as a sustainable source of protein in human diets. *Global Food Security* 28, 100520. doi: 10.1016/J.GFS.2021.100520.

Sétamou, M., Sanchez, A., Patt, J. M., Nelson, S. D., Jifon, J., & Louzada, J. S. (2012). Diurnal patterns of flight activity and effects of light on host finding behavior of the Asian

Citrus Psyllid. *Journal of Insect Behavior* 25.3, 264–76. doi:10.1007/S10905-011-9295-3.

Sharma, H. C., Clement, S .L., Ridsdill-Smith, T. J., Ranga Rao, G. V., el Bouhssini, M., Ujagir, R., Spivastava, C. P., & Miles, M. (2005). Insect pest management in food legumes: the future strategies. Proceedings of the Fourth International Food Legumes Research Conference (IFLRC-IV), New Delhi, India. Accessed May 2, 2022. http://oar.icrisat.org/5589/1/ProceedingsFIFLRC_522-544_2005.pdf

Spurgeon, D. W., & Cooper, W.R. (2012). Temperature-dependent reproductive development of *Lygus hesperus* (Hemiptera: Miridae). *Environmental Entomology* 41.4, 941–49. doi:10.1603/EN11250

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van de Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Riberiro, A. H., Pedregosa, F., van Mulbregt, P., & SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods* 17.3, 261–72. doi:10.1038/s41592-019-0686-2.

Wikström, L., Milberg, P., & Bergman, K. O. (2009). Monitoring of butterflies in semi-natural grasslands: diurnal variation and weather effects. *Journal of Insect Conservation* 13.2, 203–11. doi:10.1007/S10841-008-9144-7.

Zahn, L. K., & Gerry, A. C. (2020). Diurnal flight activity of house flies (*Musca domestica*) is influenced by sex, time of day, and environmental conditions. *Insects* 11, 391-391. doi:10.3390/INSECTS11060391.