Master's Thesis, 30 ECTS

Evaluation of Statistical Moments for Wingbeat Detection in Entomological LiDAR Signals

Division of Combustion Physics
Department of Physics
Faculty of Engineering LTH
Lund University

LUND UNIVERSITY

Author:
Mira Håkansson

Supervisor:
Meng Li
Co-supervisors:
David Dreyer
Mikkel Brydegaard

VT 2026

Abstract

This thesis evaluates whether statistical moments can be used to identify oscillatory wingbeat signals in entomological Light Detection and Ranging (LiDAR) data. The LiDAR data analyzed in this thesis was acquired using a Scheimpflug LiDAR configuration in Yamoussoukro, Ivory Coast during four days at a rice/tomato field, yielding more than 300 000 insect observations in total. The statistical moments kurtosis and skewness were applied to the intensity-distribution histograms of three representations of the backscattered signal; the integrated intensity, envelope-normalized intensity and the envelope. The statistical moments were applied on a manually labeled subset of 591 signals. To extend the test to the entire set of 300 000 signals, an automatic filtering method based on the Autocorrelation Function (ACF) was developed. The method was validated against the manually filtered subset at 96% agreement, and then used to label the rest of the dataset. In both the manually labeled dataset and the full dataset, the two classes overlap in the skewness-kurtosis values, thus the results show that these statistical moments could not differentiate between the classes. The likely reason is that the intensity-distributions histograms discard the temporal periodic structure that distinguishes oscillatory signals. Oscillations were also detected in the time-resolved center-of-mass position of the insect through the beam, which could provide an independent check on wingbeat frequency estimates. The daily activity of the oscillatory signals displayed hotspots near sunrise and sunset, consistent with swarming at specific positions along the LiDAR beam.

List of Abbreviations

ACF	Autocorrelation Function
CMOS	Complementary Metal-Oxide Semiconductor
CoM	Center of Mass
HCA	Hierarchical Cluster Analysis
LiDAR	Light Detection and Ranging
WBF	Wingbeat Frequency

Contents

Chapter 1

Introduction

1.1 Background and Motivation

Recent long-term studies across protected areas report that insect biomass has decreased by about 75% over the last 25 years[1]. The driving factors that contribute to decline in both insect biomass and diversity can be summed up by the concept in Figure 1.1 [2]. Factors such as these contributed to the encroachment of 30-50% of all natural ecosystems by the end of the 20th century [3], and overall, over 40 percent of insect species are threatened with extinction[4]. One such contributing factor to this decline is climate change. It causes changes within plant compositions in ecosystems, which in turn causes species compositions to change[5]. Global warming has also enabled the range expansion of high disease vector species such as the Asian tiger mosquito[6]. By tracking the distribution and migration patterns of disease vectors like mosquitoes, these monitoring systems can inform public health strategies and help predict high-risk exposure windows[7]. On top of this, as many of insects have seasonally varying life-stages, fast-reproducing species such as the diamondback moth are favored by the rise in temperature[8]. Due to warmer temperatures, insect hatching events can also occur earlier in the season, which is not always matched by the timing of food sources, causing a mismatch that could lead to a reduction in survival rates[9].

Effects such as these lead to a redistribution of species populations in ecosystems across the world[10]. Understanding these effects is crucial for us, since humans rely on insects as the primary pollinators of crops[11], a service worth approximately $235 to $577 billion every year[12]. Monitoring insect migration can aid in understanding the connection between ecosystems, through for example nutrient transfer and pollination across long ranges[13]. This type of monitoring over long ranges and time periods is made possible through advances in for example entomological Light Detection and Ranging (LiDAR)[14]. Changes in insect migration, species abundance, declines in biomass, as well as tracking disease vectors all call for effective methods within both instrumentation as well as computation in order to detect and distinguish high numbers of insects simultaneously. In Ivory Coast, at

which this study's data was collected, mosquitoes are a main disease vector as some carry malaria. A quarter of a million people are infected yearly by the mosquito-borne disease[7]. By utilizing entomological LiDAR in settings such as these, the activity of mosquitoes and other flying insects can be monitored in these areas and aid in developing and advancing knowledge in vector ecology[7].

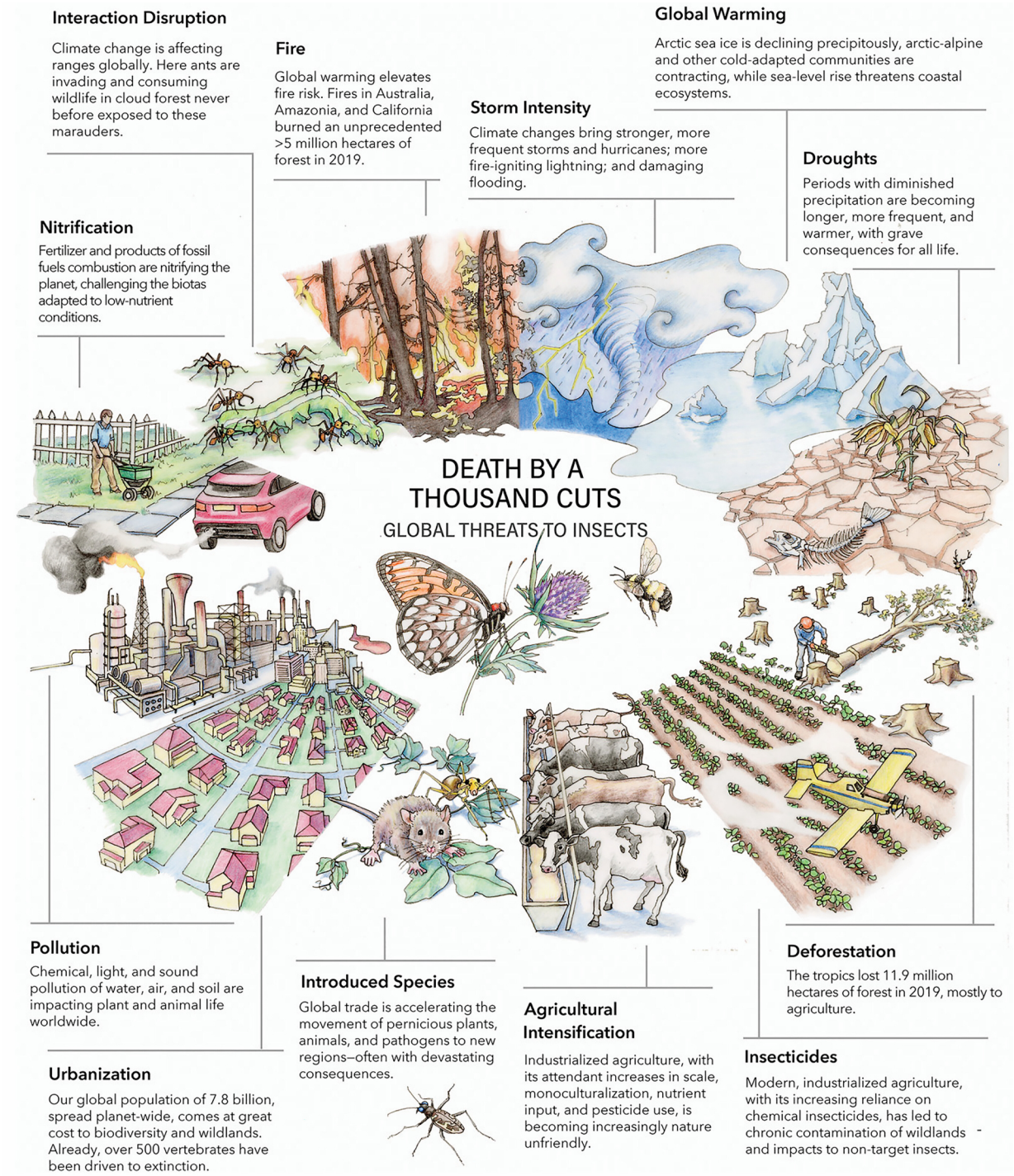

Figure 1.1: *The "Death by a Thousand Cuts" concept, illustrating the several stressors contributing to the global decline of insect biodiversity and biomass. Figure originated from [2].*

1.2 Why LiDAR?

To study insects at this large scale and over long periods, manual labeling and tracking are highly impractical. Therefore, using techniques for long-term, automated monitoring, such as insect radar[15], automated camera traps[16], and entomological LiDARs[17], [7] is essential. Using these techniques, we can also study changes in insect migration as a result of climate change, along with finding biomarkers indicating where decline could be mitigated before extinction occurs[17]. This localized information is also crucial for precision agriculture and conservation, potentially enabling targeted pest management and pollinator conservation, protecting fields more effectively, reducing chemical use, and improving the management of habitats facing urbanization and deforestation. Another important application of entomological LiDAR is the ability to track and monitor disease vectors such as mosquitoes[7]. This information of when and where increased disease vector activity occurs could be used to prevent risks of high exposure and increase public safety.

A central parameter in automated insect monitoring is determining the wingbeat frequency (WBF), as this varies for different species, and can therefore be used for classification. The success of these automated systems relies heavily on accurately determining this frequency. Acoustic methods measure the WBF by recording and studying the sound of the insect during flight, with early methods dating back to the 1950s[18]. More recently, this approach has expanded into highly accessible smartphone applications, such as the recent HumBug project[19], where standard mobile phone microphones are utilized to collect acoustic recordings of free-flying mosquitoes and classify them by their WBF.

The WBF can also be determined using optical and visual methods. Early studies used stroboscopic techniques, with such measurements dating back to the 1940s[20]. Later, optical photosensors have also been utilized to record wingbeat-induced variations in light intensity, producing digitized waveforms of the wingbeats[21] that can be analyzed for WBF determination. Although these methods have been important for determining the WBF, they are not suitable for monitoring insect species populations and migration patterns, as many require trapping the insects, performing close-range measurements as well as removing them from their natural habitats. Additionally, extracting and deducing a generalized mathematical formula applicable to acquire the WBF has been extremely challenging[22], with early work dating back to the 1950s[23]. There is a need for more scalable methods. This is where the long range of LiDAR is useful, and entomological LiDAR also enables small scale feature extraction from the insect signals.

A recently developed non-invasive technology that uses atmospheric electromagnetic fields to detect insect activity and extract the WBF without physical trapping are the autonomous sensors developed by Evolito[24]. Additionally, photonic approaches such as e-traps[25] and entomological LiDAR[26][14], can estimate the fundamental flight frequency by analyzing the transient backscattered light produced as an insect flies through an emitted light beam [27]. The entomological LiDAR enables the insects to be recorded in their

natural habitat, without the need for any capturing or collecting. A disadvantage of the autonomous sensors developed by Evolito is that their recordings exhibit a size-dependent sensitivity pattern, creating a bias for recording larger insects[24]. With entomological LiDAR, one can obtain high-resolution insect recordings for the entire observation range and probing volume[14], with nearly 350 000 observations during a single day[28]. Entomological LiDAR also retrieves information on the physical features of each insect, such as body size, wing size, WBF and flight trajectory[29]. With more spectral bands and polarization included in hardware upgrades, wing thickness and melanization can also be measured [29].

Entomological LiDAR provides a non-invasive approach to insect observations compared to the other aforementioned methods, and enable a large number of insect observations in their natural habitat, which can be utilized to study insect biodiversity [30] [31]. Using non-invasive methods and systems such as this one, we can investigate key biological events such as insect migration and hatching, and gain insight into the role that insects play in the ecosystems they inhabit. The core principle behind the analysis is that a diverse insect assemblage should naturally exhibit a diverse range of LiDAR signals.

1.3 The Wingbeat Frequency Problem in LiDAR Signals

To determine the WBF in LiDAR signals, the standard approach is to identify the highest peak in the power spectrum among the fundamental frequency and its harmonic overtones. However, the back-scattered insect signals depend on the orientation of the body relative to the laser beam. Because of this, the second harmonic overtone can often exhibit a higher amplitude than the fundamental tone [32], [33], leading to the misidentification of the true fundamental frequency. Rather than attempting to extract a discrete WBF value in the face of this orientation issue, previous studies within the research group employed unsupervised hierarchical cluster analysis (HCA) on the power spectra[30][31]. By grouping distinct spectral signatures, they derived a proxy for insect diversity. This method was validated by comparing the signal clusters to co-located Malaise trap identifications, establishing a $\sim$70% correspondence with insect families[34].

However, HCA still operates in the frequency domain, and one issue with this is that the power spectrum discards relative phase between harmonics, and distinct time-domain waveforms such as

$$x_1(t) = \sin(2\pi f_1 t) + \sin(2\pi f_2 t) \tag{1.1}$$

$$x_2(t) = \sin(2\pi f_1 t) + \cos(2\pi f_2 t) \tag{1.2}$$

yield identical power spectra despite representing temporally different signals (see Figure 1.2). Using only the power spectrum could mean grouping two distinctly different signals in the time-domain since identical in frequency domain, hiding important species or ecological

information. Therefore it is in our interest to expand how we determine signal features

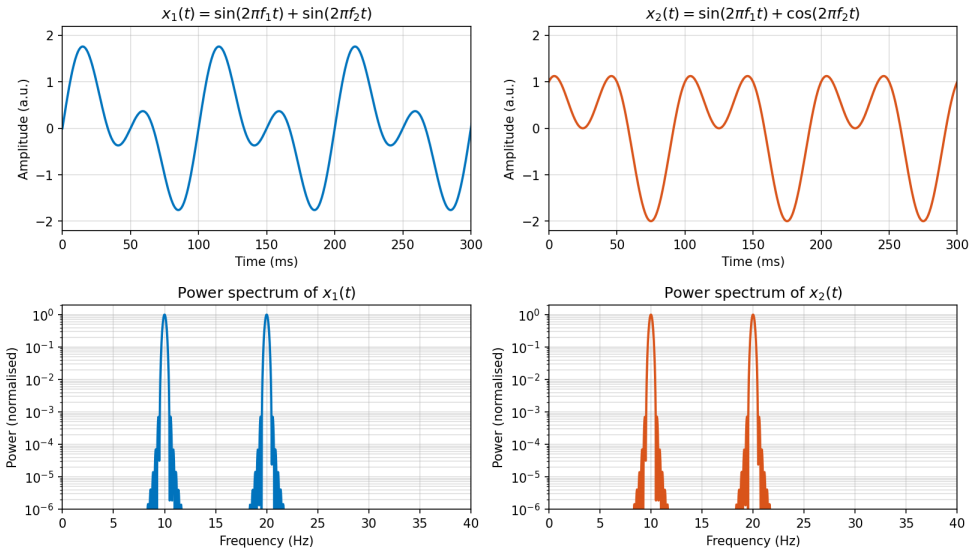

Figure 1.2: *Demonstration of how distinct time domain waveforms can have identical power spectra.*

by looking beyond the power spectrum and into the time domain. The goal is to extract additional features from the data without the need to change the instrument's hardware. In this thesis, I will explore whether statistical moments[29] can serve as a robust method for identifying oscillatory signals, on which WBF and other wing-related features can then be analyzed.

1.4 Aim of Thesis

This project analyzes a dataset collected from the Yamoussoukro region in the Ivory Coast, originally recorded during the experiment detailed in [28]. The complete dataset contains approximately 314 470 insect observations, collected with a single-band entomological Scheimpflug LiDAR system at a rice/tomato field site during four consecutive days. A key objective is to maintain low instrumental cost by utilizing existing data without the need for additional adjustments or improvements to the current LiDAR setup.

The aim of this thesis is to evaluate whether statistical moments, inspired by the multi-dimensional framework of [29], can be used to identify oscillatory signals within the LiDAR dataset. Once the oscillatory signals are detected, further analysis on wing-related features such as determining wing-to-body ratio, WBF and wingflash pattern can be conducted. These wing-features provide clues that can aid in grouping insect signals into functional or taxonomic categories. This can in turn aid in the research mentioned previously, such as monitoring disease vectors and species redistribution. As higher-order moments may

exhibit waveforms in their signals, caused by the physical oscillations of the insect's body during flight, this effect can provide an additional qualitative control of a WBF. On top of this, another goal is to study daily activity patterns of insects within the microhabitats in question.

Chapter 2

Entomological LiDAR

2.1 Entomological LiDAR Setup

The LiDAR system discussed within this thesis is in fact not the traditional Time-of-Flight (ToF) LiDAR, but a Scheimpflug LiDAR, also referred to as entomological LiDAR[14][26]. Traditional ToF LiDAR determines distance by measuring the travel time of light pulses. In contrast, a Scheimpflug LiDAR, relies on optical triangulation, where range is deduced from the observation angle. By aligning the object plane (the laser beam), the lens plane, and the image plane (the sensor) so that they intersect at a single line, the system achieves sharp focus across the entire monitoring range simultaneously despite large telescope apertures. Because of this unique geometry, it can capture continuous, high-resolution signals across both range and time, and improve light collection efficiency. With a probing range typically between less than 100 meters and 1 kilometer, the LiDAR can continuously monitor insect activity across multiple distinct microhabitats. The combination of high temporal and spatial resolution data, such as milliseconds and centimeters in range, from entomological LiDAR can provide ecological perspectives and quantify populations and distributions more than what manual sampling can achieve. Furthermore, it is significantly cheaper and easier to deploy than a ToF LiDAR of equivalent resolution.

The Scheimpflug LiDAR setup used to collect the data follows the presentation in [35], and a schematic setup is presented in Figure 2.1. It consists of a 3.2W, near infrared 808nm laser diode which generates the laser beam defining the probing volume. The light is backscattered into the receiver, a Newtonian telescope(f=800mm, D = 200mm) equipped with a $\theta = 45$ tilted linear CMOS line camera with number of pixels

$$N_p = 2048.$$

The pixel size is 14x 200μm, with each pixel corresponding to a range bin relating to the distance from LiDAR according to

$$r_J = \frac{L\left[P_J(\sin\theta - \cos\theta\tan\phi) + f\right]}{P_J(\cos\theta + \sin\theta\tan\phi)}, \tag{2.1}$$

where L is the separation between the transmitter and receiver, f is the focal length of the telescope, θ is the sensor tilt, ϕ is the slant angle[35] and P_J is the position of pixel J. As an insect crosses through the laser beam an observation is recorded. The Scheimpflug configuration ensures intersection between laser beam, lens plane, and detector plane enabling sharp focus over the entire range, which for the rice/tomato setup was up to 414 m [28]. The camera runs at a line rate of

$$f_{line} = 3.5kHz. \tag{2.2}$$

Due to laser on/off modulation that enables real time background subtraction, the effective sampling frequency achieved is

$$f_s = \frac{f_{line}}{2} = \frac{3.5\text{kHz}}{2} = 1.75 \text{ kHz}. \tag{2.3}$$

This in turn results in a time resolution of

$$\Delta t = \frac{1}{f_s} = 0.57ms, \tag{2.4}$$

and a Nyquist frequency of

$$f_{Nyq} = \frac{f_s}{2} = 875Hz, \tag{2.5}$$

which sets an upper limit on the measurable wingbeat frequencies [35]. The ability to successfully resolve different insect species relies heavily on the LiDAR configuration, since the backscattered light varies depending on both the chosen wavelength and the insect's physical characteristics. Furthermore, to successfully resolve all insect species present, the sampling frequency (Eq.2.3) must exceed twice the maximum wingbeat frequency of the insects being observed. This is known as the Nyquist criterion. If the sampling frequency is too low, the wingbeat frequencies will appear aliased to lower apparent frequencies, and the true WBF can not be determined. For example, accurately resolving a mosquito with a WBF of up to 800 Hz requires a sampling frequency of 1600 Hz. With the 1750 Hz sampling frequency used to collect the data analyzed in this thesis, the male mosquito WBF(600-800 Hz) is at the upper limit of what can be resolved. Other factors that could influence the insect behavior and detection within LiDAR signals include weather, such as rain[36], temperature and humidity[37] as well as time of day and type of habitat [35].

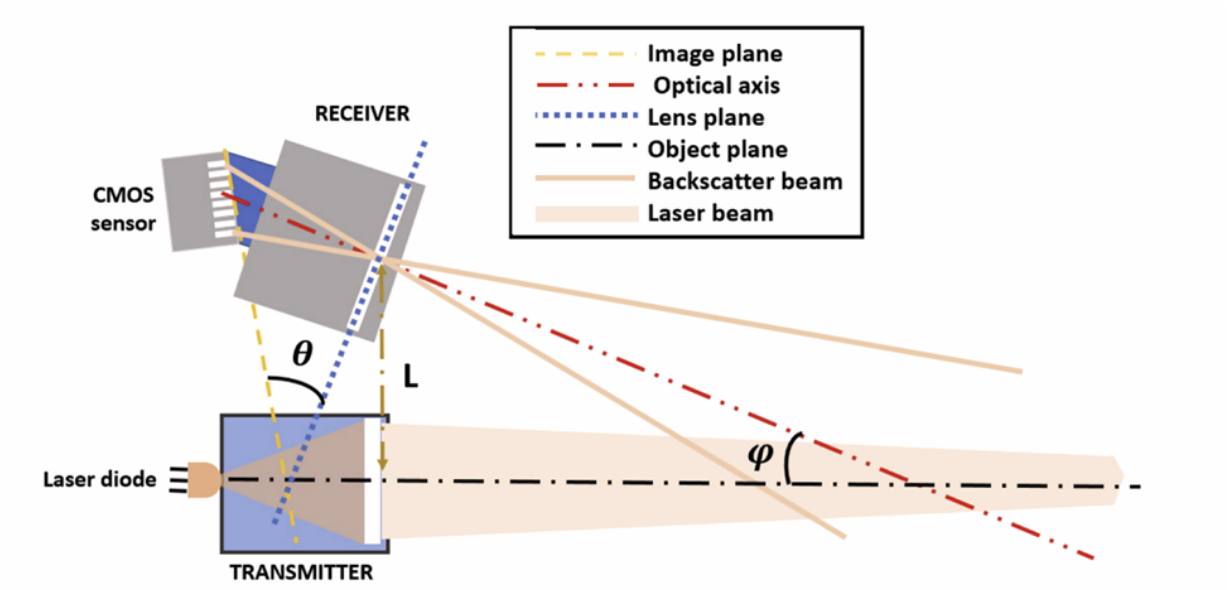

Figure 2.1: *A schematic setup of the Scheimpflug principle. The Scheimpflug principle is utilized in order to obtain sharp focus along the entire laser beam length. The principle is explained in detail in the text. This specific figure is from the paper first presenting the data from the Ivory Coast[35].*

2.2 The Yamoussoukro LiDAR Dataset

The dataset from Yamoussoukro, Ivory Coast, was first presented in [28]. The data was collected from four distinct habitat: A bush habitat, a pond site, a rice/tomato field, and a lake site. The data was collected at each site during four consecutive days during the dry season from November 2021 to March 2022. A total number of $N_{Obs} = 1.7 \cdot 10^6$ individual observations were recorded across all sites. A motivation for this project is the unique nature of the dataset as it has the highest total number of recorded signals among all entomological LiDAR datasets from the research group, which can aid in performing substantial statistical analysis and efficiently evaluating the effectiveness of the statistical methods for identifying oscillatory signals. In this thesis we have focused on the rice/tomato field dataset and developing a methodology to be applicable on all datasets. A map of Yamoussoukro and the setup for the focal site of the rice/tomato field is seen in Figure 2.2. The LiDAR system is pointed in the arrow's direction, and the sites have varying ranges with the tomato/rice field laser terminated at $r_{term} = 414$ m, positioned at $h_0 = 3$ m and $h_{term} = 7$ m.. The rice/tomato site is a small-scale agricultural habitat. The habitat is mainly characterized by cultivated vegetation and open fields. The LiDAR beam is passed close to cultivated patches of rice and tomato plants. In the paper[28] it is stated that it may exhibit humidity gradients, and fine-scale microhabitats associated with the plant and vegetation structure. They also state that due to the elevation of the LiDAR beam,

$r = 0$ and $r_{term} = 3$ m in the image shown in Figure 2.2.b), it may be harder to resolve small-scale variations within the topography or microhabitat[28].

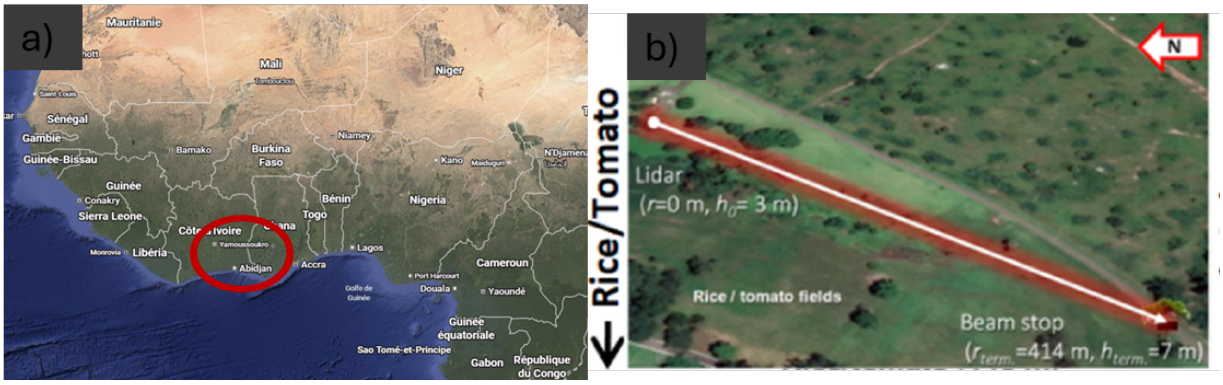

Figure 2.2: *a) A map of the west coast of Africa showing the Yamoussoukro region in Ivory Coast from Google Earth, imagery from 1/1/2021. b) LiDAR setup at the rice/tomato agricultural field in Yamoussoukro, Ivory Coast [28].*

The rice/tomato field dataset was chosen mainly due to its somewhat smaller size of around 300 000 observations, enabling faster data processing with less computational power required. The rice/tomato datasets are here referred to using the same date convention as in the paper [28], i.e Nov07, Nov09, Nov10, Nov12, although the file timestamps differ slightly from these labels. The actual timestamps for each dataset is presented in Appendix A.1.

2.3 The LiDAR Data Structure

The LiDAR data analyzed in this thesis was collected, pre-processed and cropped for insect observations by the research group[28]. This cropping is based on the groups cropping pipeline in p.29-33 in [38]. This cropping resulted in a reduction in the number of total observations by 1:1000[28]. For each insect observation, a matching noise segment was extracted from the same 10 s raw time-range backscatter intensity file, using the same range and mask but taken at a different time when no insect signal was present. This noise was then removed from the insect observation in question. The resulting cropped observations were provided as MATLAB `.mat` files with the insect observations stored in MATLAB structures, referred to as *obs*. Each field in the *obs* structure describes the insect observations, such as range, timestamp, backscatter intensity, and corresponding noise signal. The fields of the *obs* structure data can be seen in Figures A.13 and A.14 in Appendix A.6.

Each detection is stored as an observation in the *obs* data structure. A full detailed summary of each parameter and description is presented in Table 2.1. The fields relating to

the time and geometry of the cropped observation are `ft`, `rInd` and `tInd`. `ft` is the timestamp of the observation, while `tInd` and `rInd` are the time and range indices, respectively. Together, they provide the spatial and temporal coordinates of the observation. The signal data is stored in the fields `rt` and `bs`. The field `rt` is the 2D time-range backscatter intensity matrix. This matrix is referred to as $I(p, t)$ in the analysis(p : range, or pixel, and t : time). `bs` is the corresponding 1D backscattered time series. The backscattered intensity values are stored as 16-bit values. `dt` is the transit time of the insect across the beam, and `dsamp` is the transit time in exposures. The samples are related to the time by

$$dt = \frac{dsamp}{f_s}, \tag{2.6}$$

where f_s is the sampling frequency of the LiDAR. Lastly, the noise-separated signals `rtnoise` and `bsnoise` represent the background noise subtracted from the signal. *ext* is the extinction, or the light which has not reached termination and is subsequently lost along the path instead of being backscattered. Lastly, `satPix` is the number of saturated pixels.

Table 2.1: *Description of* ***obs*** *data fields.*

Field	**Description**
fno	File number of the observation.
filename	File name including an uncalibrated timestamp. The actual timestamp is given by `ft`.
ft	Absolute timestamp of the observation.
rInd	Range indices of the observation measured in pixels.
tInd	Time indices of the observation measured in exposures.
dt	Total transit time (seconds): duration the insect stays in the laser.
dsamp	Transit time measured in number of samples/exposures.
rt	Time–range intensity data: a 2D matrix of 16-bit backscattered light intensity values with dimensions `rInd` $\times$ `tInd`, where each element represents intensity at a given range bin (pixel) and time sample (exposure).
bs	Backscattered time series.
ext	Extinction: portion of light not reaching termination.
satPix	Number of saturated pixels.
rtnoise	Isolated time–range noise, subtracted from the backscattered intensity data.
bsnoise	Noise corresponding to the backscattered time series.

Chapter 3

Methodology and Data Processing

3.1 Signal Envelope Fitting

From the 2D LiDAR signal $I(p,t)$, where p is the range pixel, and t is the time, the sum over the range is calculated.

$$\Sigma p(t) = \sum_{p=p_{\text{near}}}^{p_{\text{far}}} I(p,t). \tag{3.1}$$

This results in the total backscattered intensity from the insect at each time step. An example of this signal $\Sigma p(t)$ displaying wingflashes can be found in Figure 3.1, which likely shows a Diptera signal, due to the WBF of around 280 Hz. The rapid peaks indicate a clear wing. The nature of the current method for analysis on the insect signals builds on separating the insect's body (also referred to as envelope) from the oscillatory flashes in backscatter signal intensity. These are attributed to the insect's wingflashes. This is done using a method presented in [39] and [40]. The intensity-sum, $\Sigma p(t)$ is first normalized by its maximum value. Then, the fitting is performed by estimating the signal's center time and time-width. The center is estimated using the intensity-weighted mean and the width is estimated using the spread around this center. These values are then used as a starting point for the fit. The envelope fit $E(t)$ is described by the equation:

$$E(t) = A \exp\left[s \ \text{sign}(t-t_0) \left| \frac{t-t_0}{\Delta t} \right|^{2/g-1} - \left| \frac{t-t_0}{\Delta t} \right|^{2/g} \right]. \tag{3.2}$$

A sets the amplitude of the fit, t_0 gives the position in time, Δt is the temporal width, g controls the shape of the fit, and s controls the asymmetry.

Despite the envelope being attributed to the insect's body signal, some factors could cause a loss of intensity in this signal. Due to the angle of the wings with respect to the laser, light could be redirected before reaching the insect's body when in a position such as shown in Figures 3.2.a) and 3.2.c). No redirection occurs when the wings are parallel to the LiDAR as in Figure 3.2.b)[41].

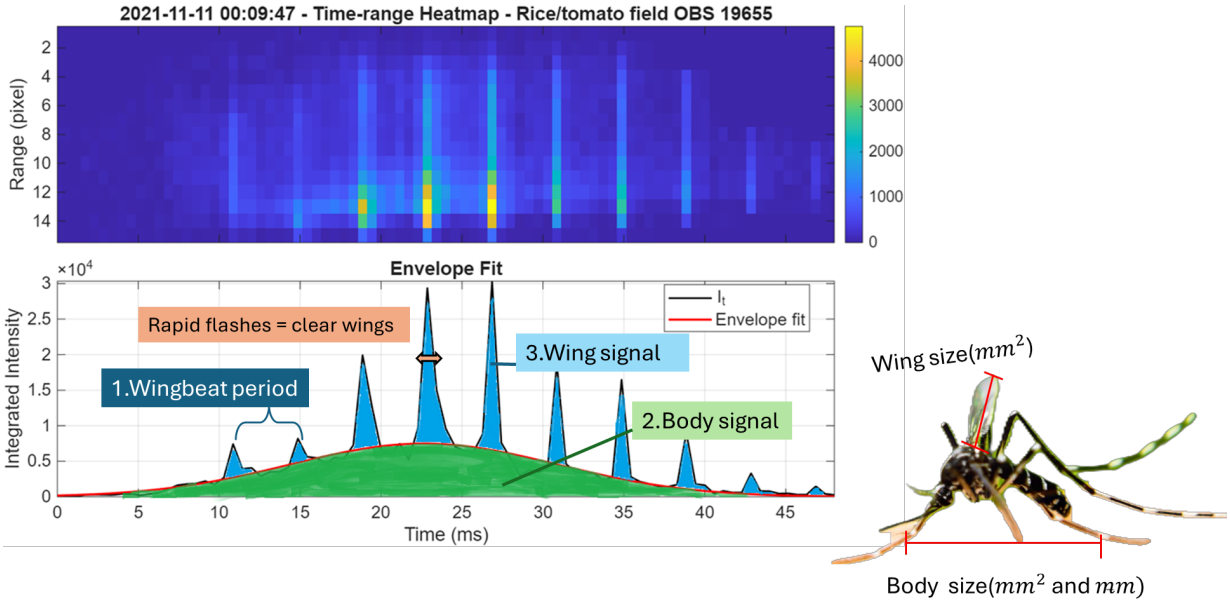

Figure 3.1: *An observation from November 11th at the Rice/tomato field around midnight. Top panel:Time-range heatmap with absolute range values on the y-axis, and time on the x-axis in milliseconds (ms). The duration of the observation is around 50 ms. The wingbeats are seen as the flashes in intensity. Bottom panel: The range-sum of the intensity; integrated intensity as a function of time. The signal is decomposed into the body (or envelope) in the image colored green, and the wings colored blue, using the envelope fit.*

Figure 3.2: *Simple visual guide to the wings angle compared to the laser light(red arrows), and when it is blocked as in a) and c), and when no blocking occurs, such as in b). Image from [41].*

3.2 Multidimensional Statistical Framework

In the data analyzed in this these, not all 314 470 cropped observations display wingbeat flashes such as in the observation in Figure 3.1. Without these oscillations, wing-related features such as the WBF, wing-to-body ration, and wingflash patterns can not be extracted. Recent work within the group [29] has used multidimensional statistical moments as a robust approach for determining the trajectory and heading of each LiDAR insect

observation. Therefore, this framework, presented in Table 3.1, is considered in order to distinguish signals displaying oscillatory behavior from those without. The flight trajectory related moments are highlighted in Table 3.1.

Table 3.1: *Matrix of primary and secondary statistical moments extracted from the time-dependent backscattered LiDAR signal [29]. The primary moments describe the spatial structure of the signal, in the pixel or range domain. The secondary moments describe the time variation of these quantities. These multidimensional moments map directly to specific physical and biological insect traits, such as body cross-section, apparent size, and wing amplitude.*

Primary Statistical Moments	Secondary Statistical Moments						
		Pixel to vector	0^{th}	1^{st}	2^{nd}	3^{rd}	4^{th}
	0^{th}	$\Sigma p_{(t)} \rightarrow \sigma_{(t)}$ ***Sum***	σ_0^0	σ_1^0 **Body cross section**	σ_2^0 **Wing cross section**	σ_3^0 **Glossiness**	σ_4^0
			t_0^0	t_1^0 **Absolute time**	t_2^0 **Transit time**	t_3^0 **Signal decrease**	t_4^0 **Beam shape**
	1^{st}	$p_{CoM(t)} \rightarrow r_{(t)}$ ***Center of Mass***	p_0^1, r_0^1	p_1^1, r_1^1 **Mean range**	p_2^1, r_2^1 **Displacement**	p_3^1, r_3^1 **Curvilinearity**	p_4^1, r_4^1 **Erraticness**
			t_0^1	t_1^1	t_2^1	t_3^1 **Distance/altitude decrease**	t_4^1 **Tailedness**
	2^{nd}	$p_{spread(t)} \rightarrow \delta_{(t)}$ ***Spread***	p_0^2, δ_0^2	p_1^2, δ_1^2 **Apparent size**	p_2^2, δ_2^2 **Wing size**	p_3^2, δ_3^2	p_4^2, δ_4^2
			t_0^2	t_1^2	t_2^2	t_3^2 **Size decrease**	t_4^2
	3^{rd}	$p_{skew(t)} \rightarrow \delta_{(t)}$ ***Skewness***	p_0^3	p_1^3	p_2^3 Wing amplitude	p_3^3	p_4^3
			t_0^3	t_1^3	t_2^3	t_3^3	t_4^3
	4^{th}	$p_{kurt(t)} \rightarrow \delta_{(t)}$ ***Kurtosis***	p_0^4	p_1^4	p_2^4	p_3^4	p_4^4
			t_0^4	t_1^4	t_2^4	t_3^4	t_4^4

Table 3.1 shows the matrix of the primary and secondary statistical moments[29]. The primary moments, denoted by $p^i(t)$ are computed for each timestep of the backscattered intensity signal. These are the rows in the table, and the $i = 0, 1, 2, 3, 4$ correspond to the sum, mean(center of mass), spread, skewness and kurtosis, respectively. The secondary moments are the columns, and are computed on the primary moments $p^i(t)$. They are denoted by t_j^i and p_j^i, where the superscript i indicates the order of the primary moment, and j the order of the secondary. The secondary moments essentially summarize the behavior of $p^i(t)$ over the entire observation interval. Since they describe the overall behavior of the signal, they are also referred to in this thesis as global moments. The primary moments reduce the 2D signal into a 1D signal, and the secondary moments reduce the 1D signal to a scalar value.

The primary moments are calculated as follows:

$p^0(t) = \Sigma p(t)$ is the sum over the range indices, or pixels p, of the backscattered signal, and represents the total backscattered intensity at time t. described by:

$$\Sigma p(t) = \sum_{p=p_{\text{near}}}^{p_{\text{far}}} I(p,t). \tag{3.3}$$

$p^1(t) = p_{CoM}(t)$ is the intensity-weighted mean pixel position(center of mass) over time:

$$p_{\text{CoM}}(t) = \frac{\sum_{p=p_{\text{near}}}^{p_{\text{far}}} p\, I(p,t)}{\Sigma p(t)}. \tag{3.4}$$

$p^2(t) = p_{spread}(t)$ is the pixel spread around the center of mass over time:

$$p_{\text{spread}}(t) = \sqrt{\frac{\sum_{p=p_{\text{near}}}^{p_{\text{far}}} \left(p - p_{\text{CoM}}(t)\right)^2 I(p,t)}{\Sigma p(t)}}. \tag{3.5}$$

$p^3(t) = p_{skew}$ is the skewness, which describes the asymmetry of the signal profile around the center of mass:

$$p_{\text{skew}}(t) = \frac{\sum_{p=p_{\text{near}}}^{p_{\text{far}}} \left(\frac{p - p_{\text{CoM}}(t)}{p_{\text{spread}}(t)}\right)^3 I(p,t)}{\Sigma p(t)}. \tag{3.6}$$

$p^4(t) = p_{kurt}$ is the kurtosis. Like the skewness, it also describes the shape of the signal profile around the center of mass. It is an indication of how peaked, or concentrated the signal is compared with its tails.

$$p_{\text{kurt}}(t) = \frac{\sum_{p=p_{\text{near}}}^{p_{\text{far}}} \left(\frac{p - p_{\text{CoM}}(t)}{p_{\text{spread}}(t)}\right)^4 I(p,t)}{\Sigma p(t)}. \tag{3.7}$$

A secondary moment can be applied to each of these time-dependent vectors in two ways. It can either describe the distribution of the vector values, denoted by p^i_j, or describe the temporal shape of the vector, t^i_j, which is reflected in Table 3.1. For example: $p_{CoM}(t)$ is the primary moment. Then: p^1_1 is the mean value of $p_{CoM}(t)$ and t^1_1 is the mean time. It is to be noted that since $p_{CoM}(t)$ is intensity-weighted, this will not always be the midpoint of the selected pixel interval. In Table 3.1, σ and δ are the optical cross section in mm^2 and apparent size in mm, respectively.

In this thesis, the statistical moments of main interest are the skewness and kurtosis. These will be applied to the intensity-distribution of the backscattered signal $\Sigma p(t)$, the envelope-normalized $\Sigma p(t)$, and the envelope. In an oscillatory signal, the shape of the intensity-distribution will be given by the wingflashes of high intensity, with low intensity between flashes. For a non-oscillatory signal, the distribution will be different since the high-intensity flashes will not be present. Since kurtosis and skewness describe the shape of the distribution, these quantities could be able to differentiate between the two classes.

Chapter 4

Results

4.1 Manually Filtered Oscillatory Signals

In order to apply the statistical moments, and compare these between oscillatory and non-oscillatory, it was recognized that the first step would be to manually filter a subset of observations. The smallest dataset, the rice/tomato dataset Nov10, was chosen. 591 signals were manually filtered into oscillatory and non-oscillatory, and resulted in 46 oscillatory and 545 non-oscillatory.

For each signal, intensity-distribution histograms of; $\Sigma p(t)$, the envelope-normalized $\Sigma p(t)$, and envelope was formed. The envelope-normalization of $\Sigma p(t)$ is done by division of $\Sigma p(t)$ by its envelope fit. This division must be performed pointwise in time, since the goal is to remove the time-dependent body signal from the time-dependent wingbeat signal. The separated wingbeat flashes can then be analyzed independently of the insect's body contribution. The time-dependent signals and their intensity-distributions are presented for an oscillatory signal in Figure 4.1. The intensity signals $\Sigma p(t)$, envelope-normalized $\Sigma p(t)$ and the envelope are shown in Figure 4.1.a), c), and e), respectively. The envelope-normalized $\Sigma p(t)$ is also referred to as the wingbeat signal. Their corresponding intensity-distribution histograms are shown in Figure 4.1.b), d), and f). The intensity-distribution histograms can be interpreted as how often different intensity values occur in the signal. An example of these histograms for a non-oscillating signal is shown in Figure 4.2, with the same layout in a-f as Figure 4.1.

One difference between the oscillating and non-oscillating signals is seen in the $\Sigma p(t)$ intensity-distribution histograms in Figures 4.1.b) and 4.2.b). The oscillating signal in Figure 4.1.b) displays a skewed intensity-distribution, increasing closer to 0, while the non-oscillating example in Figure 4.2b) displays a more symmetric distribution over the full intensity range. Another difference is seen in the intensity-distribution histogram of the envelope-normalized $\Sigma p(t)$ in Figures 4.1.d) and 4.2.d). The histogram from the non-oscillating signal in Figure 4.2.b) is centered around 0, while the histogram from the

oscillating signal in Figure 4.1.d) displays a somewhat smoothed distribution.

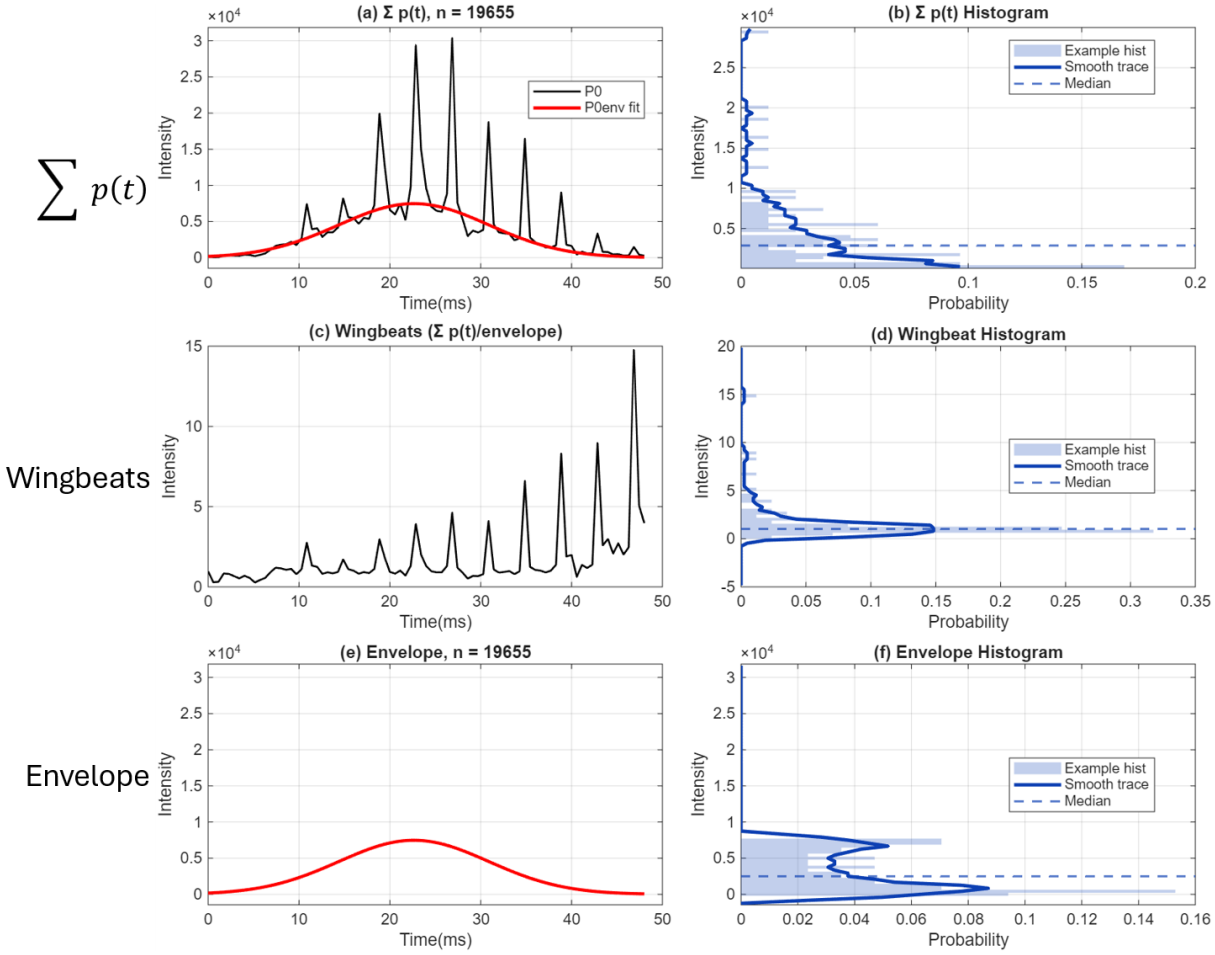

Figure 4.1: *Example oscillatory signal from Nov10. In panel a), c), and e) we have its* $\Sigma p(t)$*, the envelope-normalized* $\Sigma p(t)$*, and envelope, respectively. In panel b), d), and f) their corresponding intensity-distribution histograms are presented. The median values are represented by dashed lines in each of the histograms.*

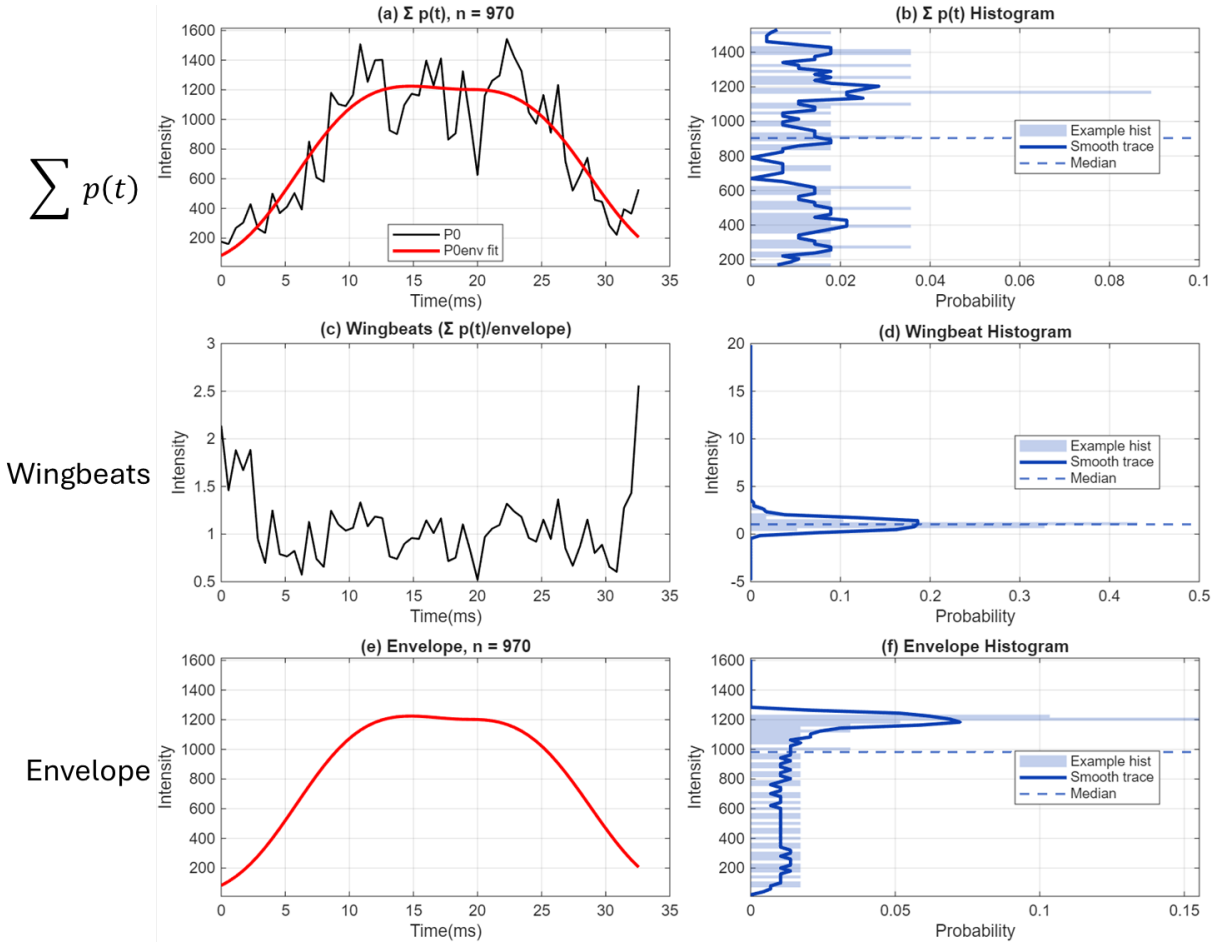

Figure 4.2: *Example non-oscillatory signal from Nov10. In panels a), c), e) the intensity sum $\Sigma p(t)$, the wingbeat signal, and the envelope of the observation $n = 970$ is plotted, respectively. Their corresponding intensity-distribution histograms are presented in panels b), d) and f).*

4.2 Evaluation of Statistical Moments for Wingbeat Detection

To quantify these differences, the skewness and kurtosis of the intensity-distribution histograms presented in Figures 4.1b), d) and f) and 4.2b), d) and f) are calculated and plotted for all 591 signals. Skewness measures whether the histogram is asymmetric. A positive value corresponds to a longer tail towards higher intensity values, whilst a negative value corresponds to a longer tail towards lower intensity values. The kurtosis measures how concentrated or tail-heavy the histogram is. A flat, uniform histogram will have a low kurtosis, while a more concentrated histogram, such as in Figure 4.2.d), will have a high kurtosis. The skewness-kurtosis plots can be seen in Figure 4.3 and summarize the shape

of the intensity-distribution histograms.

Skewness and kurtosis are not independent, the moment pairs are constrained by the Pearson inequality

$$K \geq S^2 + 1. \tag{4.1}$$

This implies that all distributions of these values must lie above a parabola. This also implies that for a large absolute skewness, the kurtosis is forced to increase. Therefore, the parabolic shape in the Figure 4.3 is partly caused by this lower bound. To distinguish real clustering from this geometric constraint, an adjusted quantity was also considered,

$$\Delta_P = K - S^2 - 1, \tag{4.2}$$

which measures the vertical distance above the Pearson lower bound. In this plot, the parabolic lower bound is mapped to the line $\Delta_P = 0$. This result is shown in Figure 4.4.

The important thing to note is that the distributions for oscillatory and non-oscillatory signals look very similar in both cases and show no clustering in the overlay comparison panel, thus it is not possible to tell from these values whether a signal is oscillatory or not.

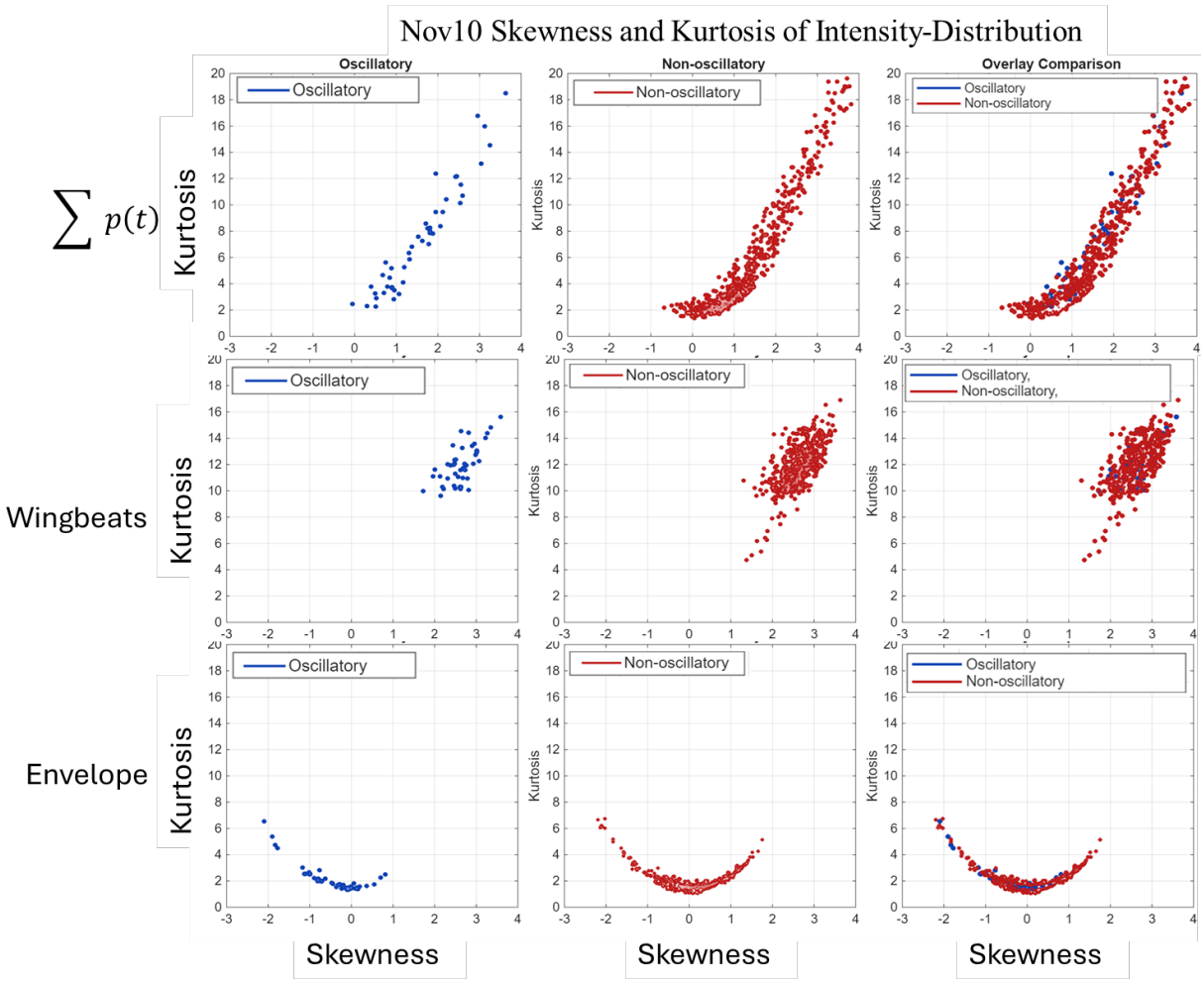

Figure 4.3: *The skewness(x-axis) and kurtosis(y-axis) plotted for the manually filtered 591 signals from Nov10. In the first row, the kurtosis and skewness is shown for the* $\Sigma p(t)$ *intensity-distribution(also shown in Figure 4.2.a), b) and 4.1a), b) for one example signal. The second row is the same plots for the wingbeat signal, and the third row is the same plots for the envelope.*

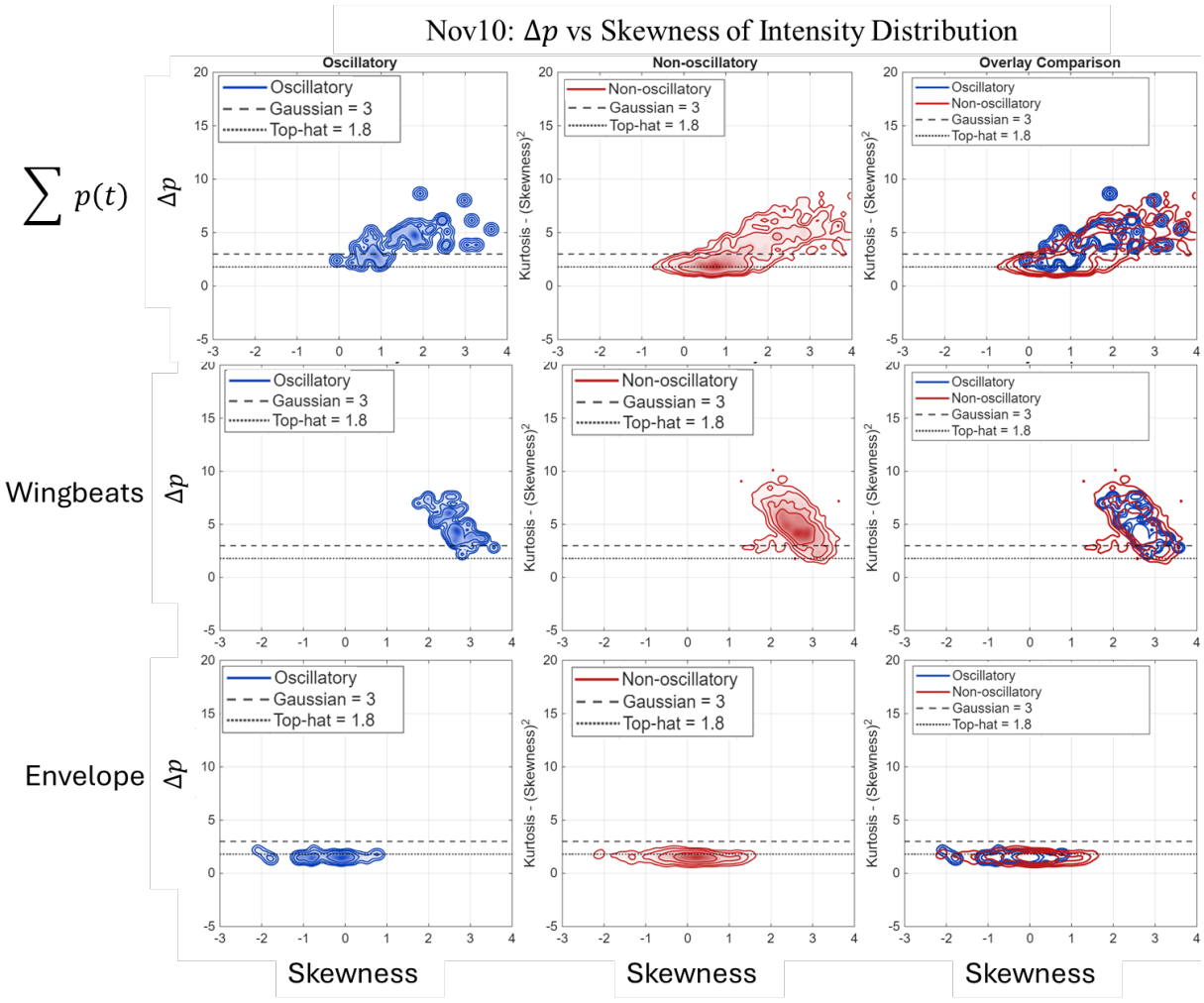

Figure 4.4: *The figure shows the deviation from theoretical minimum of* $\Delta p = 0$ *of the intensity-distributions. The typical values for a Gaussian and a Tophat distribution are also shown. The top row is the* $\Sigma p(t)$*, the middle row is the wingbeat signal, and the last row is the envelope.*

The overlay comparison in the last column of Figure 4.4 does not show any clear separation between oscillatory and non-oscillatory signals either. This indicates that despite the qualitative differences in example histograms in Figures 4.1 and 4.2 panels b), d), and f), these differences do not appear to show any strong clustering when all observations are considered. However, this dataset only contains 46 oscillatory signals compared to 545 non-oscillatory signals. This class imbalance could limit the reliability of a visual cluster comparison.

To further validate the results, ideally, the full rice/tomato dataset should be considered, and thus needs to be filtered for oscillatory signals. In the present analysis, filtering

591 signals took approximately 100 minutes. Scaling this to the entire rice/tomato dataset, consisting of around 300 000 observations across all four days, would require approximately 100 full-time working days, corresponding to half a year of manual work. This is therefore not feasible and motivates the need for an automatic filtering method.

4.3 The Autocorrelation-Based Alternative

Oscillatory signals are expected to display wingbeats with a resolvable wingbeat frequency. Current methods of WBF determination commonly rely on identifying the fundamental frequency and its harmonic overtones in the power spectrum of the time signal, where the fundamental frequency is attributed to the real WBF[27]. As stated in the introduction and in Eqs. 1.1 and 1.2, some information is discarded from the time-signal when applying this method. Thus, it is motivated to work with resolving the WBF within the time-signal.

A technique employed in previous work [42] utilized time-lag to compare the similarity between two different signals. However, the goal now is to compare the similarity of a signal and itself at a later time, in order to detect oscillations. The method will be described in detail with the corresponding mathematical formulas implemented in the MATLAB script in the Appendix A3-A.5 Instead, a shorter description that explains the intuition is presented here.

An autocorrelation function (ACF) is applied to the envelope-normalized $\Sigma p(t)$, the wingbeat signal. This measures the correlation between the signal and a time-shifted version of itself. When consecutive wingbeats align, the autocorrelation increases, producing peaks at time-lags corresponding to the wingbeat period. The corresponding physical time-lag τ_k is obtained from the lag index k by

$$\tau_k = \frac{k}{f_s}, \tag{4.3}$$

where f_s is the sampling frequency of the system. A more in depth explanation is found in Appendix A.3 to A.5.

Applying this method to the manually filtered dataset of 591 signals, the ACF-method was able to determine the oscillatory signals with 96% accuracy. The ACF-method was then applied to all signals in the rice/tomato dataset above 25 ms, as this is a lower bound used in previous works for insect transit times[7]. The number of oscillatory and non-oscillatory observations during the four days is shown in Table 4.1. The total number of observations, percentage of signals above 25 ms as well as oscillation percentage are presented for each day.

Table 4.1: *Results of oscillatory signals filtering using the ACF-method on all four days of observations at the rice/tomato site.*

Date	Observations	> 25 ms	> 25 ms (%)	> 25 ms + oscillating	> 25 + Oscillating(%)
Nov 7	121 919	100 469	82.40%	26 698	21.90%
Nov 9	75 203	64 315	85.52%	18 745	24.93%
Nov 10	22 238	17 751	79.82%	2 937	13.21%
Nov 12	95 110	73 926	77.72%	13 612	14.31%

The results show that the highest number of oscillatory observations was recorded on Nov 7, which also has the highest number of total observations. The oscillatory percentage was around 22%. On Nov 12, 95 110 observations were recorded, however only around 13% were oscillatory. It also had the lowest percentage of signals longer than 25 ms, which would affect the number of oscillatory signals. However, there is nearly double the fraction of oscillatory signals on Nov 7 as Nov 12, which is not explained by the difference in fraction of longer signals.

Nov 9 displayed the highest percentage oscillatory signals, and this day also had the highest number of signals longer than 25ms. When an insect stays in the beam longer, it has a higher chance of displaying oscillations simply due to the fact that more wingbeat periods can fit into the longer time window. November 10th had lowest percentage oscillatory signals, despite having a higher percentage longer signals. It is plausible that the number of observations could be affected by the fact that it was a rainy day[28]. Some research points in the opposite direction, that insect activity is actually increased during light rainfall[43]. However, the observations during this day was also interrupted due to the rain, leading to a lower number of observations. In total, oscillations were present in around 18.5%, out of the 314 470 observations.

Now that the entire rice/tomato dataset is filtered, the same skewness and kurtosis plots as in Figure 4.3 and 4.4 were produced for 1000 randomly picked oscillating and 1000 randomly picked non-oscillating signals for each of the four days in the dataset, a total of 4000 oscillating and 4000 non-oscillating signals. This plot can be found in Figure 4.5. Plots for each day are found in the Appendix A2. These displayed no clustering but was included to represent the entire dataset and for consistency. In Figure 4.6 the parabolic lower bound seen in Figure 4.5 is again mapped to the line $\Delta_P = 0$ to make visual comparison easier.

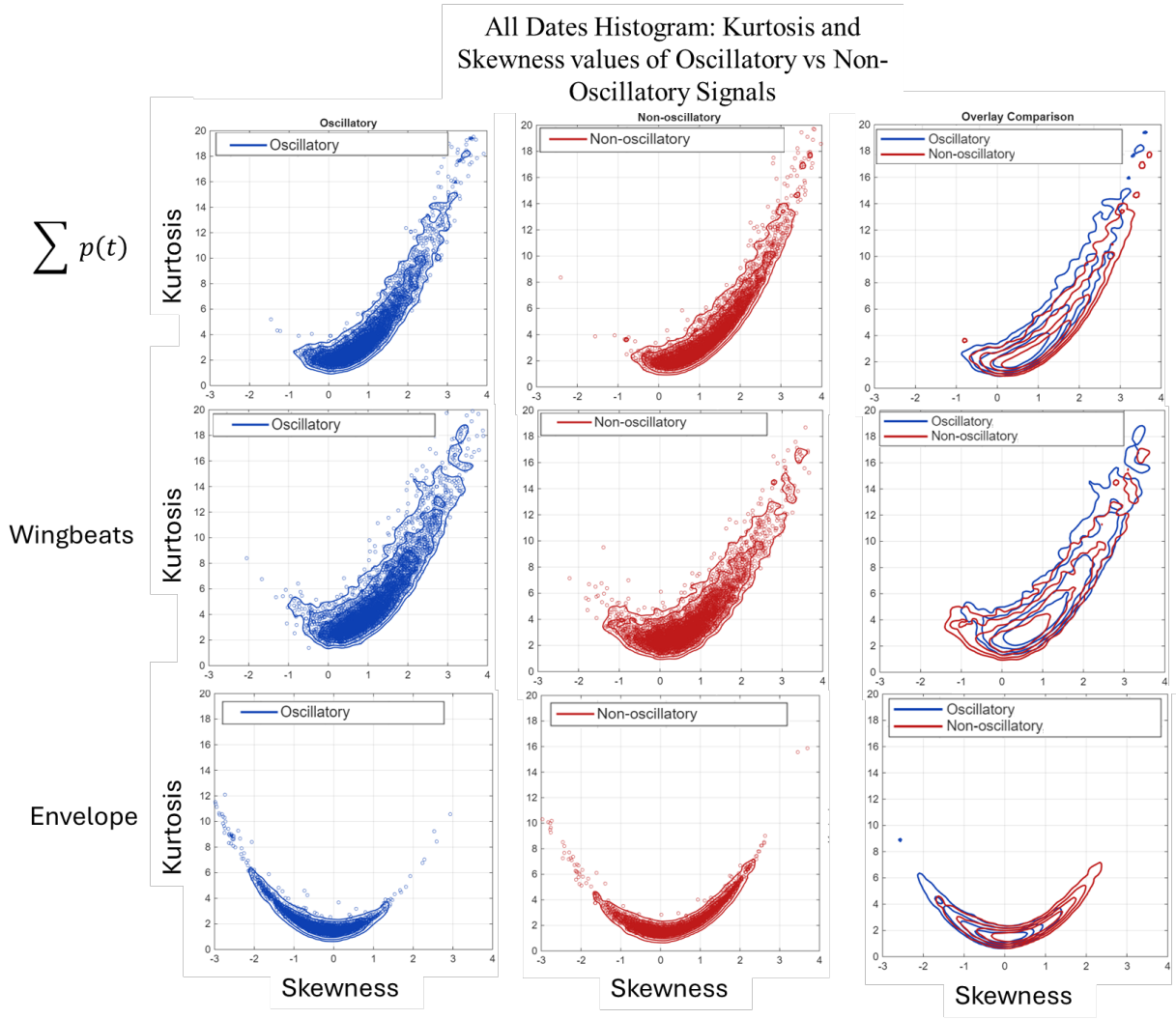

Figure 4.5: *The kurtosis vs skewness of the intensity-distribution histograms belonging to 1000 randomly chosen oscillating signals from each date and 1000 randomly chosen non-oscillating signals from each day. The first row displays $\Sigma p(t)$. Second row is the envelope-normalized $\Sigma p(t)$, and in the bottom row is the envelope. As the dataset is increased, no real clustering of the oscillatory and non-oscillatory signals is seen, and they display a large overlap.*

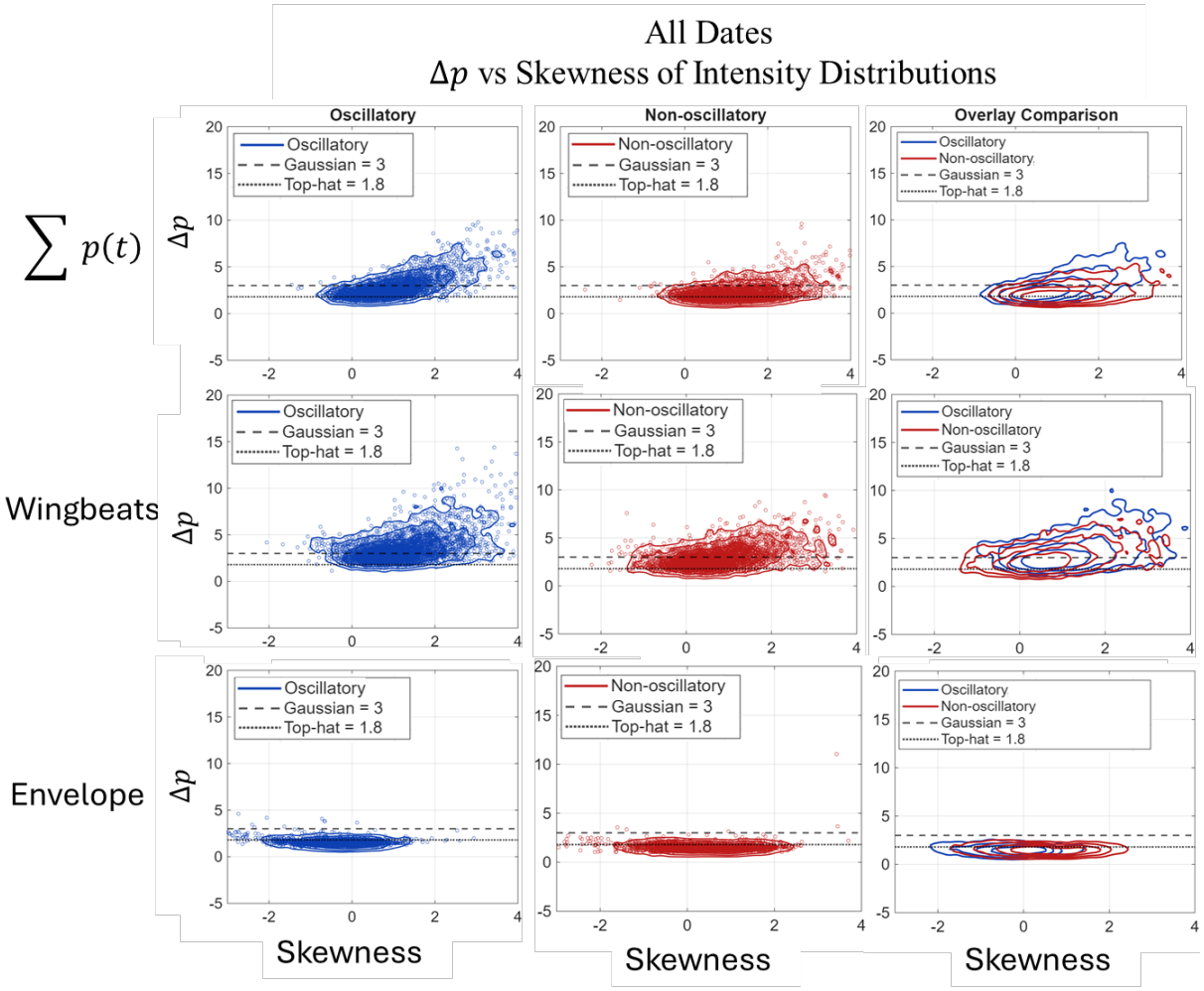

Figure 4.6: *The deviation from theoretical minimum of Δp of the intensity-distributions of $\Sigma p(t)$(top row), envelope-normalized $\Sigma p(t)$(middle row) and envelope(bottom row). 1000 randomly chosen oscillating signals from each date and 1000 randomly chosen non-oscillating signals from each day is plotted. The typical values for a Gaussian and a Tophat distribution are also shown.*

A pure envelope signal is expected to be close to a tophat, which is the case for the envelope plots seen in the bottom rows of Figures 4.4 and 4.6. In Figure 4.6 the $\Sigma p(t)$(top row) and envelope-normalized $\Sigma p(t)$ (middle row) show nearly the same distributions for oscillatory and non-oscillatory signals, with minor differences possibly in the right edge, where the oscillatory signals seem to lie higher in their Δp value. This is most noticeable in the overlay comparison in the last column. The overlap between the distributions is however still very large over the entire plane.

The results show that from these statistical moments on the intensity-distribution histograms, it is not possible to robustly differentiate and filter oscillatory signals from non-

oscillatory signals. Despite the histograms in the examples in Figures 4.1 and 4.2 showing differences in the intensity-distribution histograms, it is not possible to differentiate them using skewness and kurtosis in this way. This could be due to the oscillatory and non-oscillatory labeling done by the ACF-method, or that reducing the $\Sigma p(t)$, envelope-normalized $\Sigma p(t)$, and envelope signal into their intensity-distribution discards much of their periodic structure. The global moments are useful quantities as demonstrated in [29] for example extracting flight heading, but are unable to determine periodic structure within a signal in this case.

A reason could be that if the envelope fit of a signal is very low in intensity in the start and end of the signal, when the envelope-normalized $\Sigma p(t)$ is calculated, this produces high intensity peaks in the wingbeat signal at t_0 and t_{end} for non oscillatory signals. This would produce an intensity-distribution which is comparable to that of an oscillatory signal. Since the intensity-distribution is not time-resolved, the high intensity in the start and end of the signal would produce similar distributions as for a signal with periodic flashes. The temporal information that is lost is thus important for signals of this nature.

4.4 Case studies and Limitations of the Autocorrelation Method

From the results presented in the previous chapter in Figures 4.3 to 4.5, it is clear that the statistical moments regarded in this thesis are not enough to separate oscillatory signals from non-oscillatory signals. During the process of determining this, an automatic filtering method using ACF was developed. An additional feature of the ACF-method is the extraction of an approximate WBF for each oscillating signal. This is demonstrated in the chosen examples of oscillatory signals in Figures 4.7, 4.8, and 4.9. Despite detecting wingbeats with a high accuracy, the actual frequency value resolved is not always accurate. With the current implementation of the ACF-method, the search for WBF is carried out in integer lags $k = 1, 2, 3, ...$ It searches for peaks of high correlation at these k, which limits the number of different frequencies detectable to

$$\frac{f_s}{k} = \frac{1750}{k} \text{ Hz}, \tag{4.4}$$

where $f_s = 1750$ Hz is the sampling frequency of the system. This means that the ACF has set values of WBF that it evaluates, and the one with highest ACF-peak will be chosen as the WBF. However, there could exist a non-integer k that would resolve a WBF with a higher ACF-peak. This limitation is illustrated in the cases presented in Figures 4.7, 4.8 and 4.9, since manually measuring the WBF from the wingbeat flashes results in a different WBF than the one found automatically. The lag k search can be adjusted to non-integer steps in order to more accurately find the WBF. This will increase the computational time. Since the aim for using the ACF-method in this thesis was not to detect the WBF value, but to filter the oscillatory from non oscillatory, an integer lag value was chosen.

The ACF-method was also applied to detect oscillations within the insect's pixel center of mass, $p_{CoM}(t)$, through the LiDAR beam. Oscillations in the $p_{CoM}(t)$ were detected and are found in the Figures 4.7 and 4.9. In Figure 4.7, both the $\Sigma p(t)$ and p_{CoM} display the same oscillation frequency of 102.94Hz. This value corresponds to

$$\frac{f_s}{k} = \frac{1750}{17} = 102.94 \text{ Hz}. \tag{4.5}$$

The zoomed in panels of Figure 4.7.d) and 4.7.i) show that the real WBF is $\sim$ 205 Hz, double the detected value, highlighting the previously stated drawback of this thesis implementation of the ACF-method.

In Figure 4.7 an observation from Nov 10th is presented. This observation displays a WBF of 102.94Hz. In the zoomed panel next to panel d) we can see that this is incorrect. The true WBF can be calculated by manually measuring the distance between flashes. The real WBF is thus $\approx$ 205Hz. In Figure 4.7.i) it can be seen that the $p_{CoM}(t)$ also displays clear oscillatory behavior, but with the wrong WBF determined by the ACF. The wingbeat flashes in Figure 4.7.a) seem to be very long in range, indicating that the wingflash may take place somewhat further from the body backscatter signal than in the other observations in Figures 4.8 and 4.9. This could perhaps indicate an insect with longer wings.

In the heatmap of the Nov07 observation in Figure 4.8.a) The signal displays clearly resolved oscillations. However, the nature of the oscillations display an on/off periodicity instead of clear wingbeat flashes(compared to Figure 4.7), indicating that the signal is undersampled by the camera. Manually measuring the WBF results in a frequency of 870Hz, which is very close to the Nyquist frequency of 875Hz, calculated from the sampling frequency demonstrated in eq. 2.3. This observation demonstrates the instrumental limit of the system, and the types of WBF that cannot be resolved. With a higher sampling frequency for example 20 kHz or 40 kHz, as discussed in [44] the WBF could have been resolved.

In the observation from November 9th presented in Figure 4.9, we observe oscillations in both $\Sigma p(t)$ and $p_{CoM}(t)$, with different frequencies. The $\Sigma p(t)$ WBF is 92.11 Hz, while $p_{CoM}(t)$ oscillations display a frequency of 62.50 Hz. Similar to Figure 4.7, the ACF-determined WBF is incorrect, and measuring the flashes manually we find the correct frequency of 175 Hz. In this observation, every other wingbeat flash displays a higher intensity. Similar behavior of oscillatory insect signals have been mentioned before [33] and can be attributed to the direction which the insect is viewed from w.r.t the laser beam. When the insect is viewed from the side, the wingflashes are observed twice per wingbeat cycle, once for each direction of wing motion, but thus angled differently. This results in non-identical backscatter resonance with the laser light for the different directions of the wing motion, and thus results in half of the actual WBF[33]. In this case, the ACF-method would thus determine the correct WBF. The nature of the oscillations in the $p_{CoM}(t)$ shown in Figure 4.9.h) and j) are different from the oscillations in $\Sigma p(t)$ in Figure 4.9.b)-d). In

the $p_{CoM}(t)$ they display a descending behavior below the fit, followed by an overshoot, before slowly descending below the fit again. The body of the insect could be displaying a wingbeat at around $t \approx 22$ ms in Figure 4.9.h), followed by a decrease in altitude without a wingstroke, before a stronger wingstroke at $t \approx 28$ ms. The conclusion that it is stronger is motivated by the larger displacement in the p_{CoM} during this wingflash. At $t \approx 28$ ms, the highest intensity wingflash is displayed for $\Sigma p(t)$ as well, most clearly visible in panel c). This could further solidify that the smaller flash in intensity at $t \approx 22$ ms in both panel c) and h) is at half the real WBF[33]. In Figures 4.10 to 4.12 some more examples of signal types are presented, where the ACF-method has detected wingbeats with varying frequencies, as well as oscillations in the $p_{CoM}(t)$ with matching frequency to the WBF in Figure 4.10. The signal shape in the time-range map in panels 4.10.a), 4.11a), and 4.12a) are some examples of the many different shapes that the insect signal can have within the dataset.

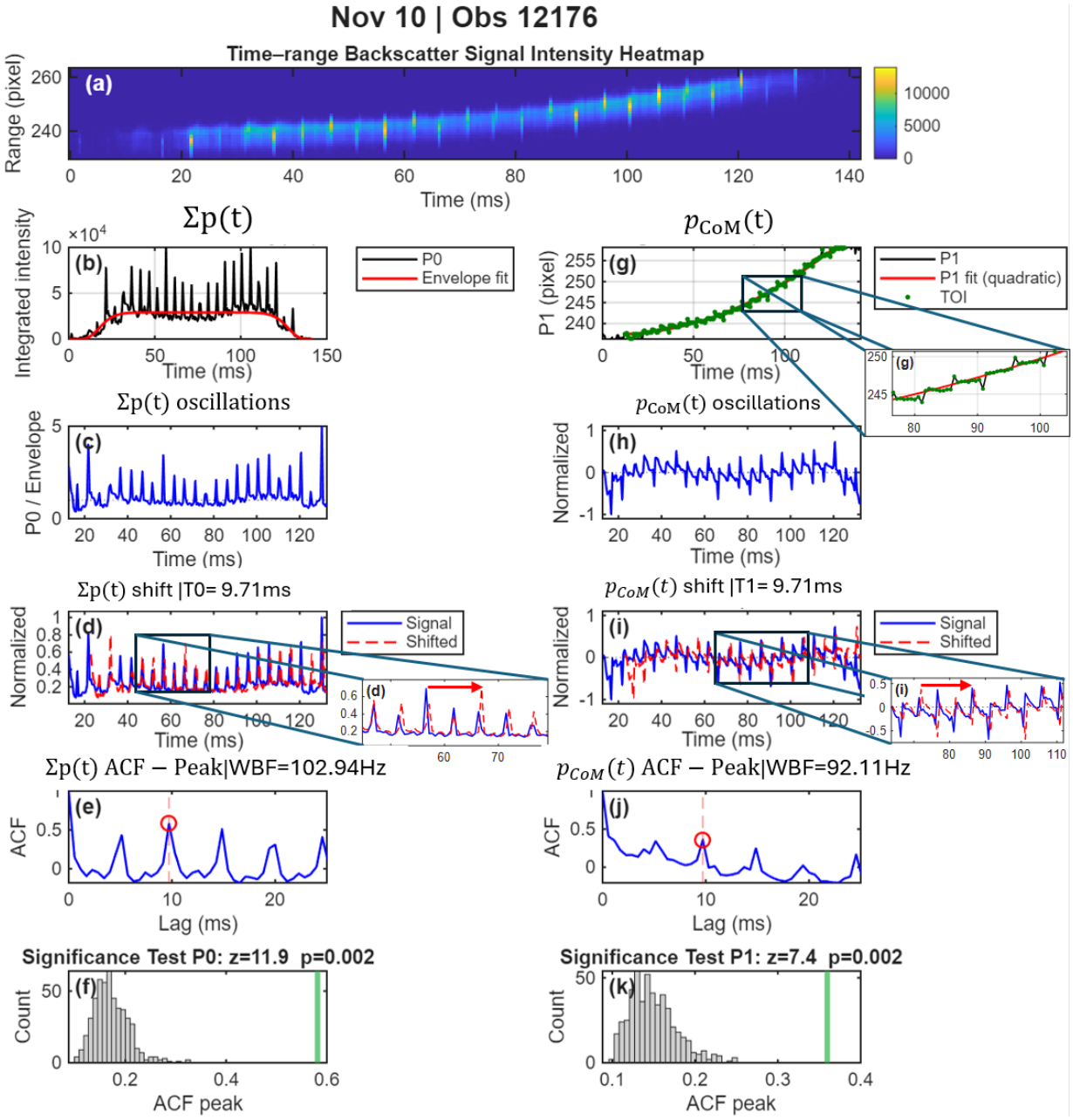

Figure 4.7: *An observation from November 10th with a WBF of 102.94Hz. a) The 2D time-range heatmap. b) $\Sigma p(t)$ with its envelope fit. c) Envelope-normalized $\Sigma p(t)$. d) ACF applied & the time-shift with highest correlation $T0 = 9.71ms$. The zoomed in panel illustrates the magnitude of the time-shift. e) Highest ACF-peak circled. This indicates which time-lag value(x-axis) corresponds to the highest correlation(y-axis). f) The significance test, with $Z - score$ and $P - value$. An explanation to these is found in Appendix A3 to A5. g) The $p_{CoM}(t)$ and a fit. The zoomed panel shows oscillations around the fit. The fit is subtracted from $p_{CoM}(t)$, the result is shown in h). In i) the time-shift of highest correlation is shown, $T1 = 9.71ms$. j) The highest ACF-peak is circled, as for e), indicating the time-lag value(x-axis) of highest correlation(y-axis). Panel k) is identical to panel e) but for $p_{CoM}(t)$.*

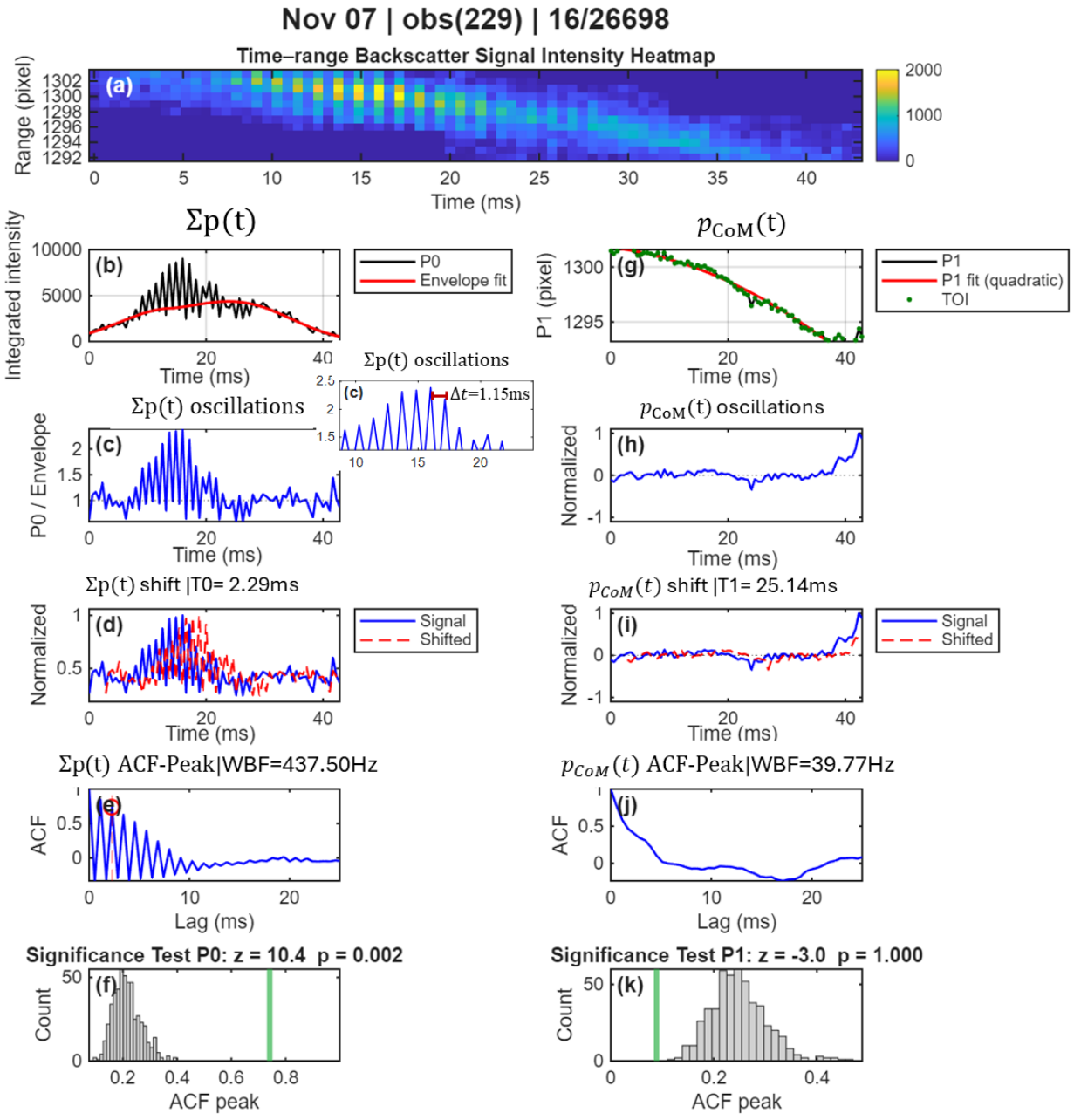

Figure 4.8: *An observation from Nov07 with WBF 437.5Hz. The panel structure is identical to that of Figure 4.7. Here, in panel c) the zoomed panel to the right shows the actual time-lag manually measured between two peaks as* $\Delta t = 1.15ms$ *which corresponds to a frequency of 870Hz. In panel e)the time-lag value T0 =2.29ms had highest correlation. In* $p_{CoM}(t)$ *panels g-k) there is no display of any clear oscillations.*

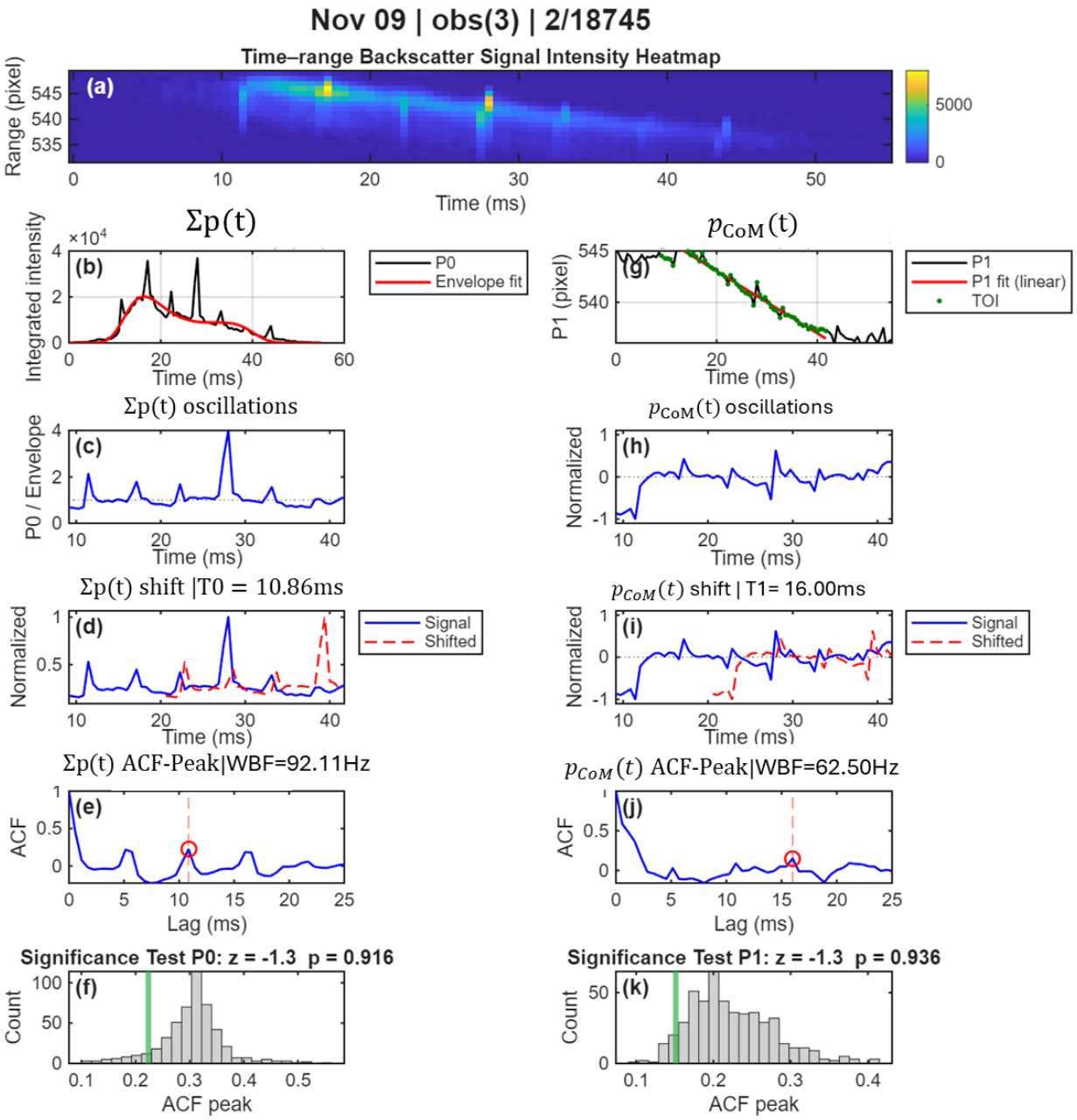

Figure 4.9: *An observation from Nov09. This figure also has identical panel structure as Figures 4.7 and 4.8. This observation displays clear wingbeat flashes, and as seen most clearly in panels a) and b), every other wingbeat flash in this observation has a higher intensity. In panels e) the WBF is determined to be 92.11Hz. In panel g) oscillations around the linear fit can be resolved. The $p_{CoM}(t)$ oscillations have a frequency of 62.50Hz, determined by the ACF in panel j).*

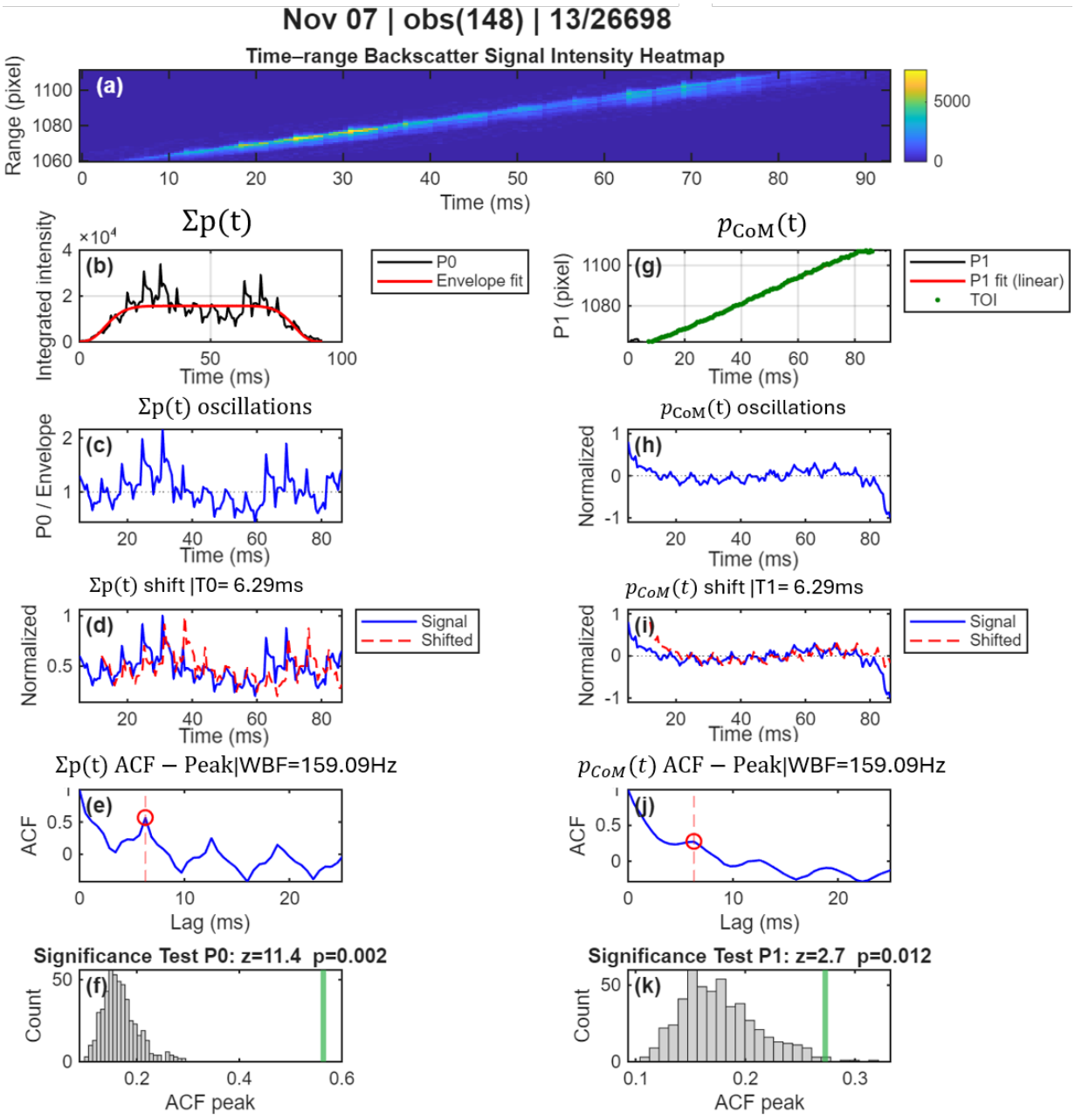

Figure 4.10: An observation from November 7th with WBF of around 160 Hz. The peak shapes seen in panels b)-d) are different from other presented examples, for example in Figure 4.12, where the peaks look more symmetric. In this figure the peaks seem to be cut off at thir tail, which may indicate some under-sampling.

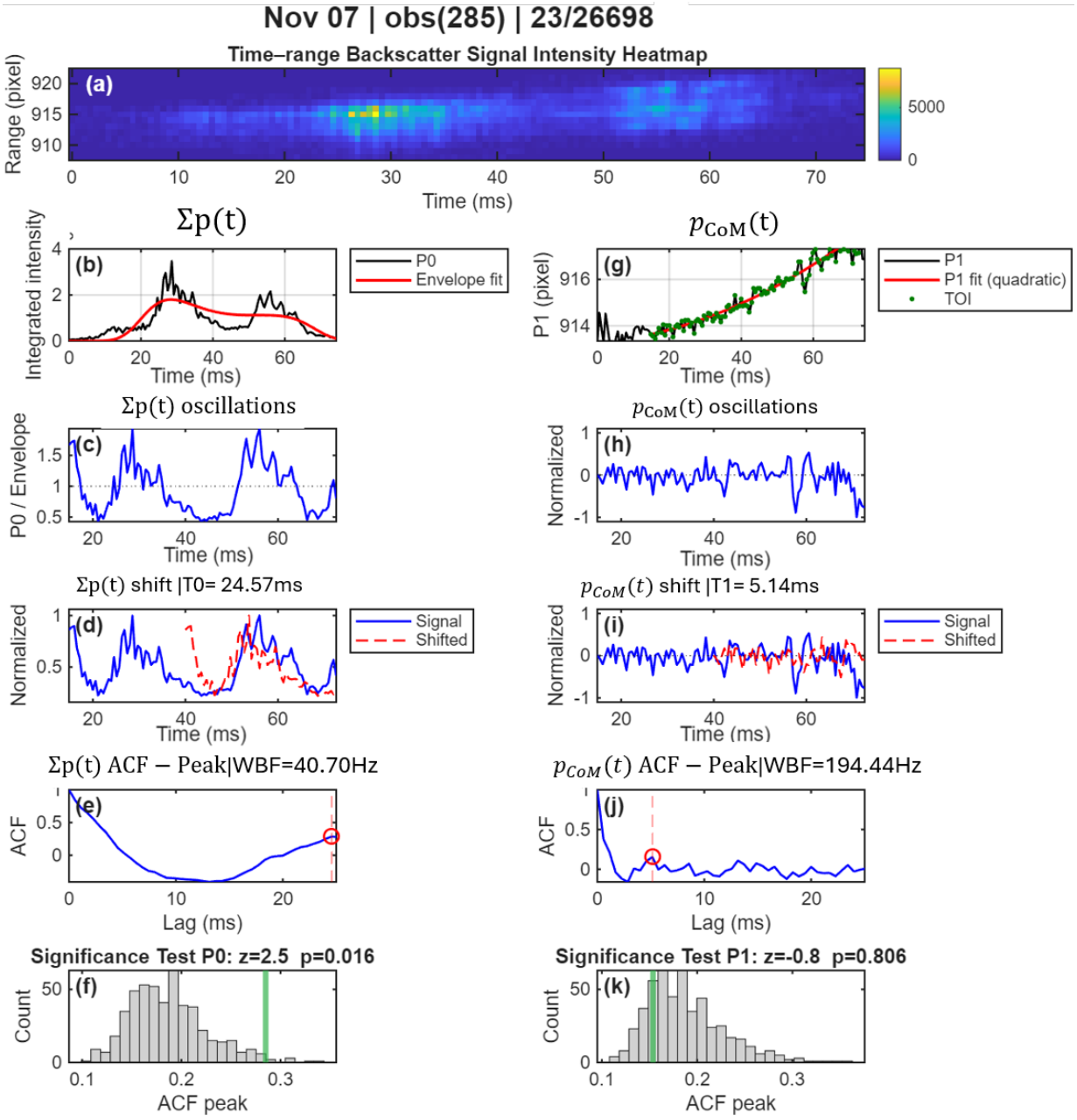

Figure 4.11: A signal from November 7th with a low wingbeat frequency of around 40 Hz. This signal shape is different than previously presented as it only has two high peaks of intensity, and the peaks are relatively broad, indicating an insect with more diffuse wings such as a moth. No oscillations in the $p_{CoM}(t)$.

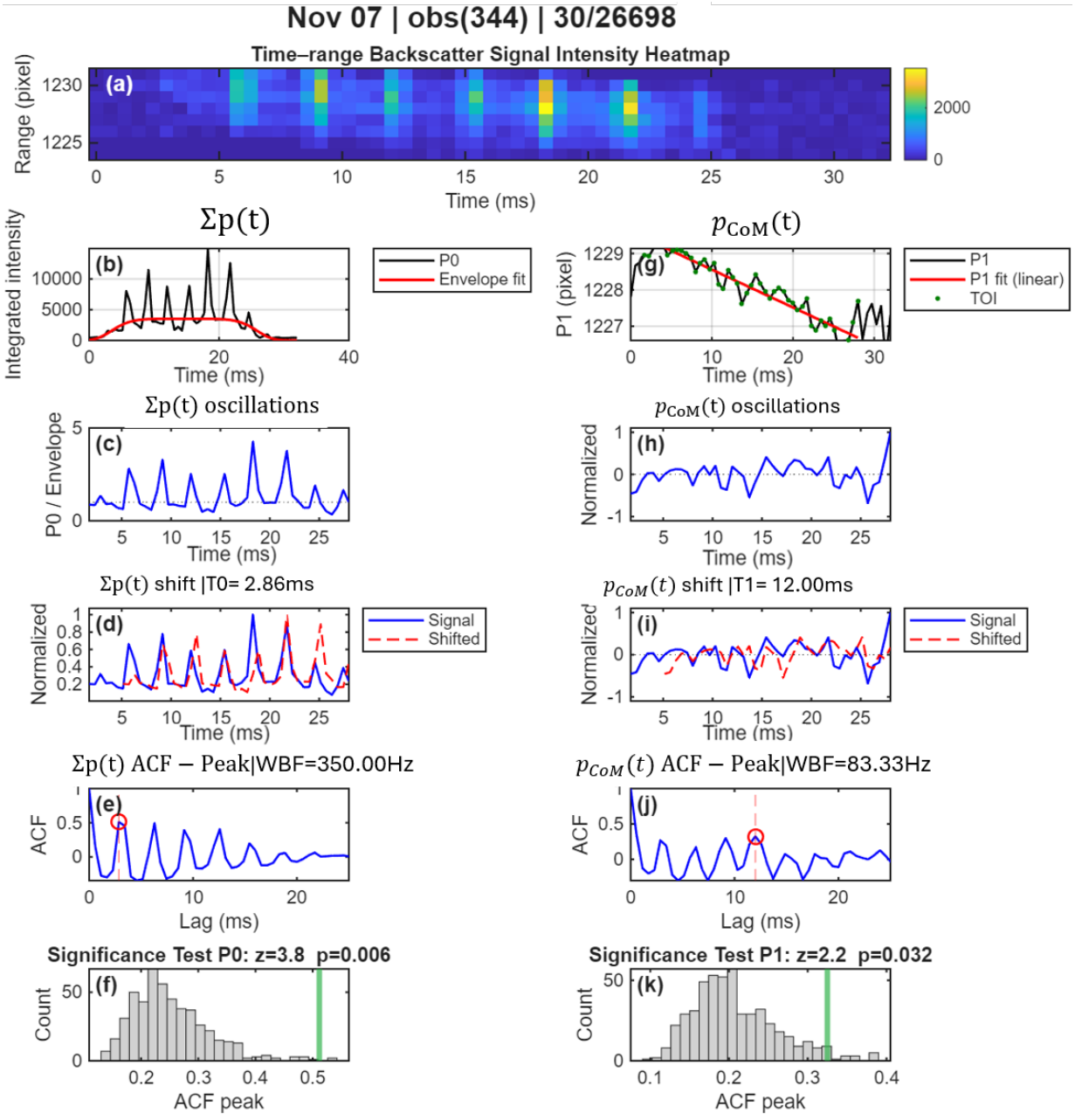

Figure 4.12: An observation from November 7th with a short transit time of around 30 ms. Despite this we see several clear wingflashes during the insect's transit through the beam. Oscillations in the $p_{CoM}(t)$ are also seen.

Chapter 5

Outlook

5.1 Daily Activity Patterns of Filtered Oscillatory Signals

By applying the ACF-method to automatically filter for large datasets, one can analyze the daily activity of known insect signals displaying wingbeats in order to extract ecological information. Such an ecological interpretation was not done in this thesis due to a limitation of time, but an initial example is presented here for the rice/tomato dataset for all four days. The x-axis shows the time of day, while the y-axis shows the range in pixels. This is translated (as introduced in the setup section) to a real physical position in the LiDAR beam. This dimensionality allows us to explore collective insect behavior, such as hotspots or swarming, where many insects gather at the same place at the same time. Such hotspots are seen in Figure 5.1 at 18:00 and 06:00, around sunset and sunrise. A similar behavior is observed for all signals, presented in Figure 5.2.

In Figure 5.1 we can clearly see a large increase in the number of observations as the sun rises at 06:00 on Nov09. During the day, the activity is constant until the afternoon. An increase in activity is observed for all days at around 18:00.

All signals also show increased activity during sunset and sunrise, seen in Figure 5.2 at 06:00 on Nov09 and around 18:00 on all days. A hotspot, is observed at Nov07 around 18:00 for both oscillating signals in Figure 5.1 and all signals in Figure 5.2. Another hotspot is visible during November 9th at 06:00, at a range of around 1500 pixels. A pattern that is noticeable for both Figure 5.1 and Figure 5.2 is that most hotspots occur at ranges $\sim$ 1500 pixels. This indicates that there are more swarming events or hotspots further away from the LiDAR setup, and thus further away from what appears to be a house nearby the LiDAR setup in Figure 2.2b) The vegetation around this area does not seem to be very different from the vegetation at closer ranges, other than perhaps some trees.

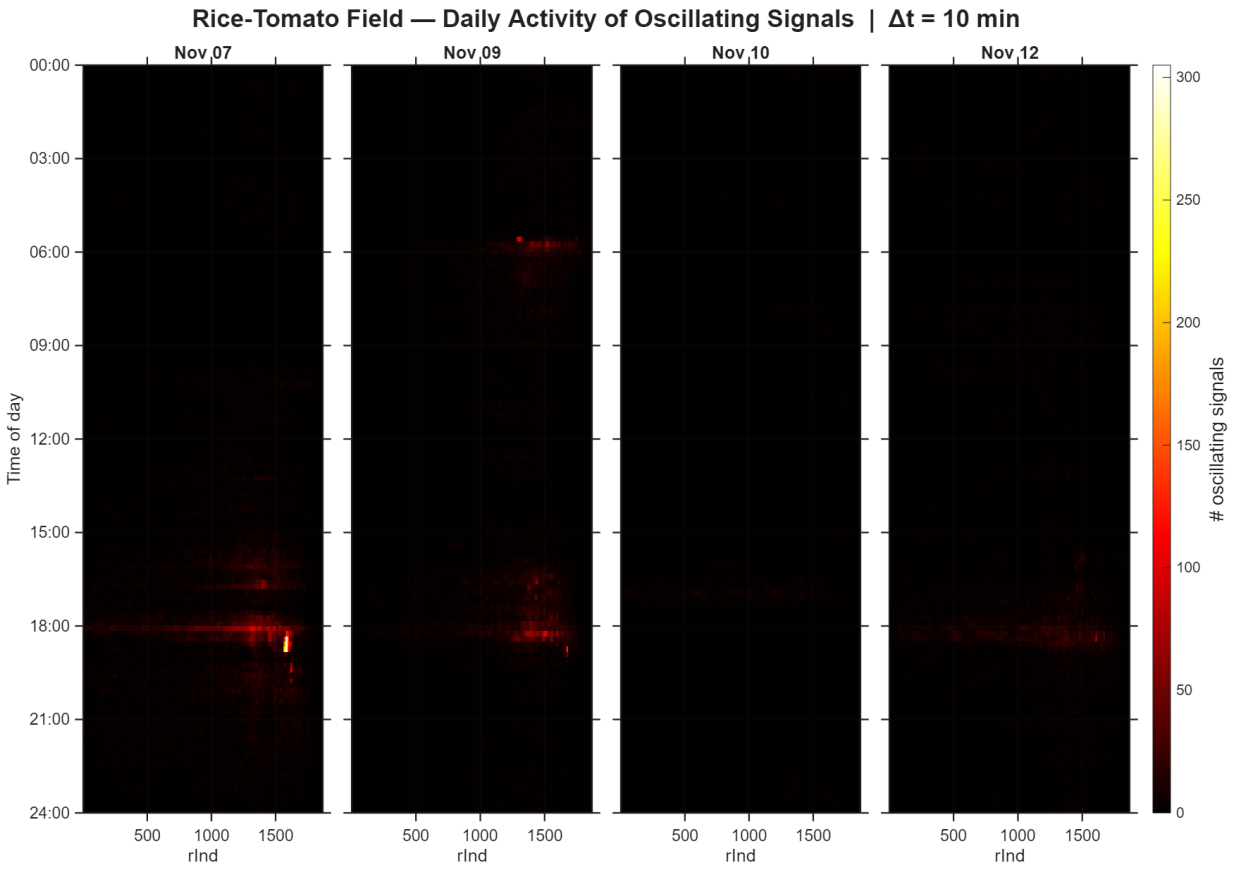

Figure 5.1: *Daily activity of oscillating signals filtered with ACF-method*

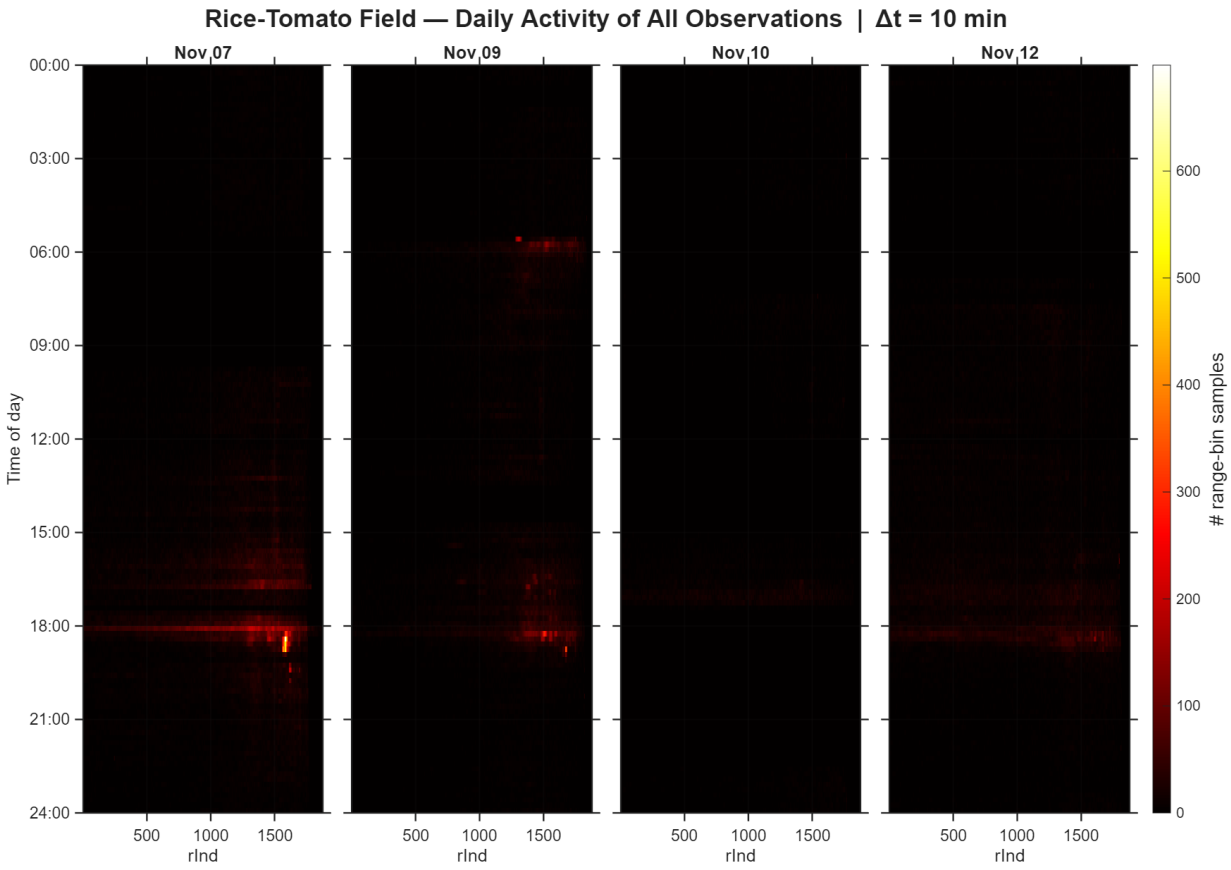

Figure 5.2: *Daily activity of all observations during the four days.*

Chapter 6

Conclusion

The aim of this thesis was to determine whether statistical moments could aid in determining the WBF of flying insects. In the results presented in Figures 4.3 and 4.4 the distribution between oscillatory(displaying wing-flashes) and non-oscillatory signals were very similar. Subsequent analysis on the entire dataset over the four days showed similar results. Thus it is concluded that the statistical moments applied in this thesis were not enough to be able to differentiate oscillatory signals from non-oscillatory signals in order to perform statistical analysis. A possible source of error could be high intensity flashes in the beginning and end of the signal envelope-normalized signal, generated by the envelope-normalization when the envelope is very low in intensity. A solution could be to limit the region of which the intensity-distribution is computed, to remove these edge effects. The dataset was filtered using the ACF-method, which could cause a bias in the filtered signals. One bias that is applied is the time limit of 25 ms, which can change the shape of the envelope. However, around 80% of the signals were above 25 ms, and only around 18% displayed oscillations, so a large number of non-oscillating signals were also above 25 ms. Another bias of the ACF-method could be the type of signal. For a high correlation value to be achieved, the signal must contain several clear wingbeat flashes within the transit time. This favors observations in which the insect remains long enough in the beam to record several wingbeat flashes. It also favors sharp wingbeat flashes, and well-defined insect bodies and wings, resulting in an under-representation of larger insects with diffuse wings and wingbeat modulations such as moths and butterflies, and favoring small insects with clear wings such as small flies.

Furthermore, another objective of this thesis was to explore hidden features within the data as an alternative to upgrading the hardware. The ACF-method has shown that the time domain contains previously hidden features capable of identifying wingbeats. This foundation is useful, despite the applied version of the ACF-method in this thesis being unable to determine the precise WBF of the flying insect signals. The unique approach presented in this thesis analyzes phase information of the signal within the time domain by removing the body contribution and calculating correlation within the wingflashes. The oscillations detected within the $p_{CoM}(t)$ can also be utilized in order to aid in verification

of a detected WBF.

The natural next step for this work is to improve the WBF determination within the ACF, and examining any bias the filtering method currently may have. Another outlook is to apply the ACF-method to all four sites, further analyze the daily activity patterns of these, and investigate the flight heading and trajectory information of the insect signals filtered with the ACF-method.

Chapter 7

Acknowledgements

A very big thank you to my wonderful supervisors Meng Li, Mikkel Brydegaard, and David Dreyer for the opportunity to work on this project, as well as always providing me with invaluable input and lending your knowledge and time to me, even on your busy schedules. I also want to give a big thank you to everyone in the Biophotonic Sensing Group. Everyone has always had helpful input and showed their support and interest during my work through our many interesting discussions together. You have always made me feel welcomed into the group and have made the work feel lighter and fun in so many ways. I also want to thank my wonderful colleagues outside the research group, with whom I have carried out my studies with these last six years, for your wonderful friendship and support.

Bibliography

[1] C. A. Hallmann, M. Sorg, E. Jongejans, H. Siepel, N. Hofland, H. Schwan, W. Stenmans, A. Müller, H. Sumser, T. Hörren, *et al.*, "More than 75 percent decline over 27 years in total flying insect biomass in protected areas," *PLOS ONE*, vol. 12, no. 10, p. e0185809, 2017.

[2] D. Wagner, E. Grames, M. Forister, M. Berenbaum, and D. Stopak, "Insect decline in the anthropocene: Death by a thousand cuts," *Proceedings of the National Academy of Sciences*, vol. 118, p. e2023989118, 01 2021.

[3] P. Vitousek, H. Mooney, J. Lubchenco, and J. Melillo, "Human domination of earth's ecosystems," *Science*, vol. 277, 07 1997.

[4] F. Sánchez-Bayo and K. A. Wyckhuys, "Worldwide decline of the entomofauna: A review of its drivers," *Biological Conservation*, vol. 232, pp. 8–27, 2019.

[5] B. R. Scheffers, L. D. Meester, T. C. L. Bridge, A. A. Hoffmann, J. M. Pandolfi, R. T. Corlett, S. H. M. Butchart, P. Pearce-Kelly, K. M. Kovacs, D. Dudgeon, M. Pacifici, C. Rondinini, W. B. Foden, T. G. Martin, C. Mora, D. Bickford, and J. E. M. Watson, "The broad footprint of climate change from genes to biomes to people," *Science*, vol. 354, no. 6313, p. aaf7671, 2016.

[6] I. Rochlin, D. V. Ninivaggi, M. L. Hutchinson, and A. Farajollahi, "Climate change and range expansion of the asian tiger mosquito (Aedes albopictus) in northeastern USA: Implications for public health practitioners," *PLOS ONE*, vol. 8, no. 4, p. e60874, 2013.

[7] M. B. et al., "Lidar reveals activity anomaly of malaria vectors during pan-african eclipse," *Science Advances*, vol. 6, no. 20, p. eaay5487, 2020.

[8] J. A. Harvey, R. Heinen, R. Gols, and M. P. Thakur, "Climate change-mediated temperature extremes and insects: From outbreaks to breakdowns," *Global Change Biology*, vol. 26, no. 12, pp. 6685–6701, 2020.

[9] M. E. Visser and L. J. M. Holleman, "Warmer springs disrupt the synchrony of oak and winter moth phenology," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, pp. 289–294, 2001.

[10] F. Ojija, M. Mng'ong'o, B. N. Aloo, G. Mayengo, and M. Helikumi, "Effect of global climate change on insect populations, distribution, and its dynamics," *Journal of Asia-Pacific Entomology*, vol. 28, no. 3, p. 102442, 2025.

[11] D. Goulson, "The insect apocalypse, and why it matters," *Current Biology*, vol. 29, no. 19, pp. R967–R971, 2019.

[12] J. A. Noriega, J. Hortal, F. M. Azcárate, M. P. Berg, N. Bonada, M. J. Briones, I. Del Toro, D. Goulson, S. Ibanez, D. A. Landis, M. Moretti, S. G. Potts, E. M. Slade, J. C. Stout, M. D. Ulyshen, F. L. Wackers, B. A. Woodcock, and A. M. Santos, "Research trends in ecosystem services provided by insects," *Basic and Applied Ecology*, vol. 26, pp. 8–23, 2018. Insect Effects on Ecosystem services.

[13] K. R. Wotton, B. Gao, M. H. Menz, R. K. Morris, S. G. Ball, K. S. Lim, D. R. Reynolds, G. Hu, and J. W. Chapman, "Mass seasonal migrations of hoverflies provide extensive pollination and crop protection services," *Current Biology*, vol. 29, no. 13, pp. 2167–2173.e5, 2019.

[14] M. Brydegaard and S. Jansson, "Advances in entomological laser radar," *The Journal of Engineering*, vol. 2019, 11 2019.

[15] P. Aralimarad, A. M. Reynolds, K. S. Lim, D. R. Reynolds, and J. W. Chapman, "Flight altitude selection increases orientation performance in high-flying nocturnal insect migrants," *Animal Behaviour*, vol. 82, no. 6, pp. 1221–1225, 2011.

[16] J. C. Habel, M. J. Samways, and T. Schmitt, "Mitigating the precipitous decline of terrestrial European insects: Requirements for a new strategy," *Biodiversity and Conservation*, vol. 28, no. 6, pp. 1343–1360, 2019.

[17] H. Chen, M. Li, H. Månefjord, P. Travers, J. Salvador, L. Müller, D. Dreyer, J. Alison, T. Høye, G. Hu, E. Warrant, and M. Brydegaard, "Lidar as a potential tool for monitoring migratory insects: A field case study in sweden," *iScience*, vol. 27, p. 109588, 03 2024.

[18] C. M. Williams and R. Galambos, "Oscilloscopic and stroboscopic analysis of the flight sounds of drosophila," *The Biological Bulletin*, vol. 99, no. 2, pp. 300–307, 1950.

[19] I. Kiskin, M. Sinka, A. D. Cobb, W. Rafique, L. Wang, D. Zilli, B. Gutteridge, R. Dam, T. Marinos, Y. Li, G. Killeen, D. Msaky, E. Kaindoa, K. Willis, and S. J. Roberts, "HumbugDB: a large-scale acoustic mosquito dataset," 2021.

[20] S. C. Reed, C. M. Williams, and L. E. Chadwick, "Frequency of wing-beat as a character for separating species races and geographic varieties of drosophila," *Genetics*, vol. 27, pp. 349–361, 05 1942.

[21] A. Moore and R. H. Miller, "Automated identification of optically sensed aphid (homoptera: Aphidae) wingbeat waveforms," *Annals of the Entomological Society of America*, vol. 95, no. 1, pp. 1–8, 2002.

[22] M. Deakin, "Formulae for insect wingbeat frequency," *Journal of insect science (Online)*, vol. 10, p. 96, 07 2010.

[23] O. Sotavalta, "Flight-tone and wing-stroke frequency of insects and the dynamics of insect flight," *Nature*, vol. 170, pp. 1057–1058, 1952.

[24] F. Odgaard, P. V. Kjærbo, A. H. Poorjam, K. Hechmi, R. Luciano, and N. Krebs, "Automated insect detection and biomass monitoring via ai and electrical field sensor technology," *Scientific Reports*, vol. 15, 08 2025.

[25] I. Potamitis, I. Rigakis, and K. Fysarakis, "The electronic mcphail trap," *Sensors*, vol. 14, no. 12, pp. 22285–22299, 2014.

[26] M. Brydegaard and S. Svanberg, "Photonic monitoring of atmospheric and aquatic fauna," *Laser Photonics Reviews*, vol. 12, p. 1800135, 10 2018.

[27] A. K. Gebru, E. G. Rohwer, P. Neethling, and M. S. Brydegaard, "Investigation of atmospheric insect wing-beat frequencies and iridescence features using a multispectral kHz remote detection system," *Journal of Applied Remote Sensing*, vol. 8, no. 1, p. 083503, 2014.

[28] A. S. Doria Yamoa, B. K. Kouakou, A. Y. Gbogbo, A. Runemark, R. van Klink, J. T. Zoueu, and M. Brydegaard, "Lidar reveals distinct insect daily activity and diversity between habitats," *Scientific Reports*, vol. 15, p. 43544, Nov 2025.

[29] D. Dreyer, M. Li, H. Månefjord, A. S.-D. Yamoa, Y. A. Gbogbo, L. Müller, A. Runemark, B. K. Kouakou, R. Boateng, A. A. Huzortey, J. T. Zoueu, B. Anderson, and M. Brydegaard, "Robust and diverse multidimensional statistical moments in dual-band entomological lidar for improved real time insect monitoring," *Journal of Experimental Biology*, p. jeb.251761, 05 2026.

[30] D. Bernenko, M. Li, H. Månefjord, S. Jansson, A. Runemark, C. Kirkeby, and M. Brydegaard, "Insect diversity estimation in polarimetric lidar," *PLOS ONE*, vol. 19, 11 2024.

[31] H. Månefjord, A. Yamoa, A. Gbogbo, L. Müller, A. Runemark, B. Kouakou, R. Boateng, A. Huzortey, I. Kwame Badu, N. Wahlberg, M. Brydegaard, J. Zoueu, B. Anderson, and M. Li, "Stratification of insect diversity and daily activity patterns in the west african virgin forest taï assessed by entomological lidar," *Scientific Reports*, vol. 15, 07 2025.

[32] S. Jansson, A. Gebru, R. Ignell, J. Abbott, and M. Brydegaard, "Correlation of mosquito wing-beat harmonics to aid in species classification and flight heading assessment," p. 24, 07 2019.

[33] S. Jansson and M. Brydegaard, "Passive khz lidar for the quantification of insect activity and dispersal," *Animal Biotelemetry*, vol. 6, 05 2018.

[34] K. Rydhmer, S. Jansson, L. Still, B. D. Beck, V. Chatzaki, K. Olsen, B. Van Hoff, C. Grønne, J. K. Meier, M. Montoro, I. K. Schmidt, C. Kirkeby, H. G. Smith, and M. Brydegaard, "Photonic sensors reflect variation in insect abundance and diversity across habitats," *Ecological Indicators*, vol. 158, p. 111483, 2024.

[35] B. K. Kouakou, S. Jansson, M. Brydegaard, and J. T. Zoueu, "Entomological scheimpflug lidar for estimating unique insect classes in-situ field test from ivory coast," *OSA Continuum*, vol. 3, pp. 2362–2371, Sep 2020.

[36] S. Zhu, E. Malmqvist, W. Li, S. Jansson, Y. Li, Z. Duan, K. Svanberg, H. Feng, Z. Song, G. Zhao, M. Brydegaard, and S. Svanberg, "Insect abundance over chinese rice fields in relation to environmental parameters, studied with a polarization-sensitive cw near-ir lidar system," *Applied Physics B: Lasers and Optics*, vol. 123, no. 7, p. 211, 2017.

[37] A. S. D. Yamoa, B. K. Kouakou, A. Y. Gbogbo, and J. T. Zoueu, "Monitoring anopheles mosquitoes through lidar, weather, and vegetation index data analysis: A study in côte d'ivoire," *Journal of Materials and Environmental Science*, vol. 16, no. 9, pp. 1622–1640, 2025.

[38] S. Jansson, *Entomological Lidar: Target Characterization and Field Applications*. PhD thesis, Lund University, Lund, Sweden, 2020.

[39] M. Brydegaard, "Towards quantitative optical cross sections in entomological laser radar – potential of temporal and spherical parameterizations for identifying atmospheric fauna," *PLOS ONE*, vol. 10, pp. 1–15, 08 2015.

[40] E. Malmqvist, S. Jansson, S. Török, and M. Brydegaard, "Effective parameterization of laser radar observations of atmospheric fauna," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, pp. 1–1, 01 2015.

[41] Y. Li, Z. Han, R. Nessler, Z. Yi, P. Hemmer, R. Brick, A. V. Sokolov, and M. O. Scully, "Optical multiband polarimetric modulation sensing for gender and species identification of flying native solitary pollinators," *iScience*, vol. 26, no. 11, p. 108265, 2023.

[42] Z. Xu, "Insect diversity estimation in entomological lidar," master's thesis, Lund University, Department of Biology, 2021. Supervisors: Mikkel Brydegaard, Klas Rydhmer, Meng Li.

[43] Z. Song, B. Zhang, H. Feng, S. Zhu, L. Hu, M. Brydegaard, Y. Li, S. Jansson, E. Malmqvist, K. Svanberg, G. Zhao, J. Bood, S. Svanberg, and D. Li, "Application of lidar remote sensing of insects in agricultural entomology on the chinese scene," *Journal of Applied Entomology*, vol. 144, no. 3, pp. 161–169, 2020.

[44] M. Li, H. Månefjord, and M. Brydegaard, "Resolving fast wingbeat flashes in situ with entomological lidar," in *2024 IEEE Photonics Conference (IPC)*, pp. 1–2, 2024.

[45] L. J. Cator, B. J. Arthur, L. C. Harrington, and R. R. Hoy, "Harmonic convergence in the love songs of the dengue vector mosquito," *Science*, vol. 323, no. 5917, pp. 1077–1079, 2009.

[46] W. Yu, H. Zhang, R. Xu, Y. Sun, and K. Wu, "Characterization of wingbeat frequency of different taxa of migratory insects in northeast asia," *Insects*, vol. 13, no. 6, 2022.

Appendix A

Appendix

A.1 Date Inconsistency

The rice/tomato datasets are in the thesis referred to using the same date convention as in the paper [28], i.e Nov07, Nov09, Nov10, Nov12. However, direct conversion of the MATLAB serial date numbers using

$$t = \texttt{datetime}(\text{datenum}, \texttt{'ConvertFrom'}, \texttt{'datenum'})$$

shows that the timestamps in the files are not fully consistent with these labels: The file

Dataset label	Start time from file	End time from file
Nov07	2021-11-07 12:23:07	2021-11-08 11:08:48
Nov09	2021-11-09 07:32:33	2021-11-10 04:13:04
Nov10	2021-11-11 04:45:17	2021-11-11 15:06:34
Nov12	2021-11-12 07:36:45	2021-11-13 02:15:04

Table A.1: *Time intervals obtained from direct conversion of the MATLAB datenum values in the rice/tomato files. The dataset labels follow the convention used in the paper[28].*

timestamps indicate an inconsistency: the dataset labeled Nov10 converts to November 11th, and the labeled dates are not strictly four consecutive calendar dates. It is unclear whether the datasets are incorrectly timestamped on Nov10 or the label is incorrect. As stated, the labels used in the paper will also be used here in plots and in the discussion of the data in order to ensure consistency and comparability with the original paper[28].

A.2 Skewness Vs Kurtosis of Intensity-Distribution Results

In this section, skewness vs kurtosis of the intensity-distribution histograms for each date is plotted separately. In all figures (Figures A.1 to A.12, the signals display similar behavior

without any significant clustering seen in the overlay panels. The figures are included here for consistency in representing and analyzing all data.

A.2.1 Nov07

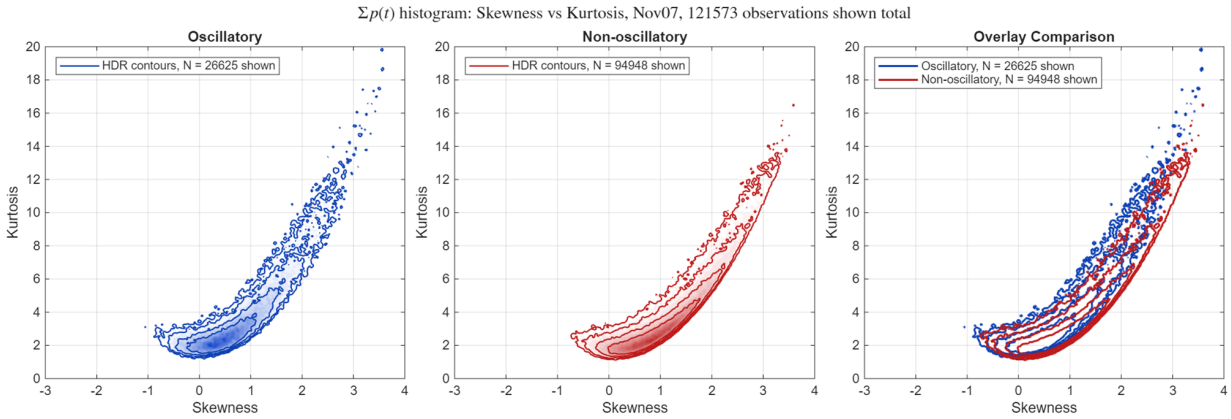

Figure A.1: *Kurtosis vs. skewness of the intensity-distribution histogram of* $\Sigma p(t)$ *on Nov. 7.*

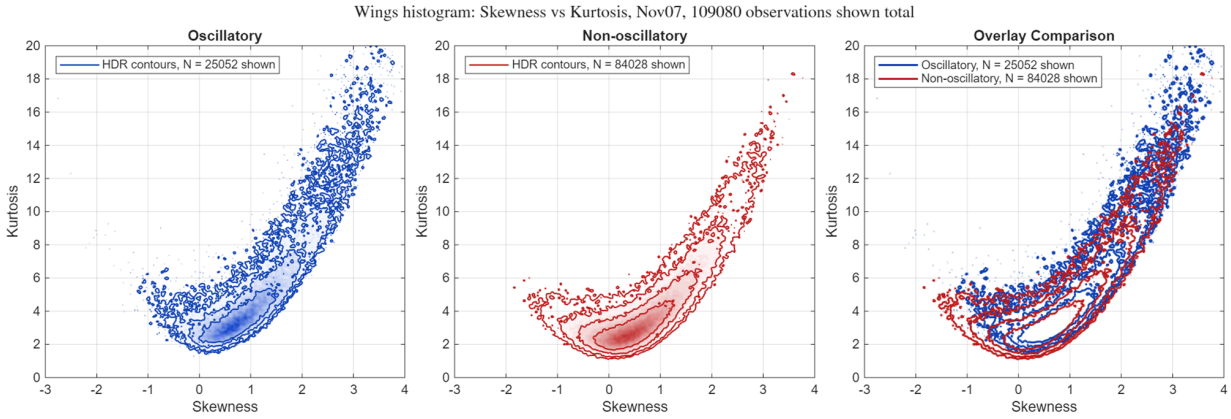

Figure A.2: *Kurtosis vs. skewness of the intensity-distribution histogram of the envelope-normalized* $\Sigma p(t)$ *on Nov. 7.*

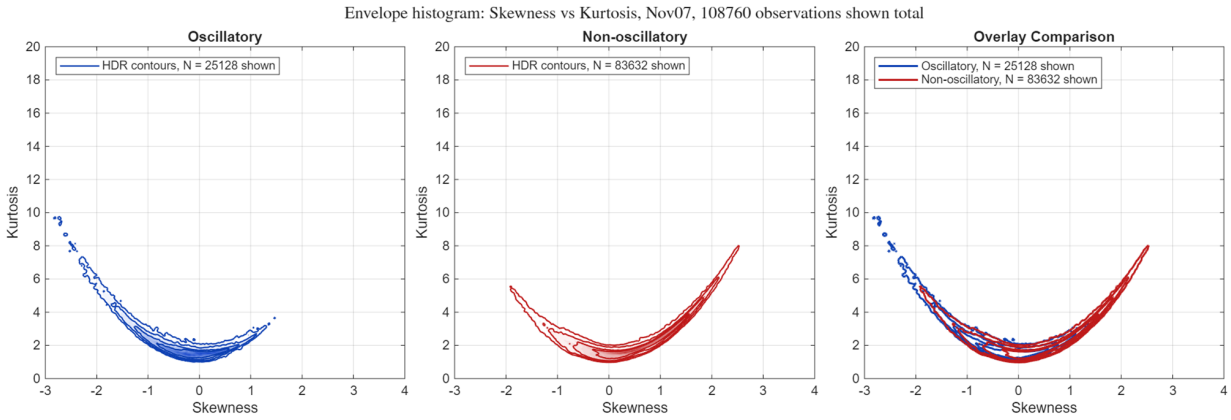

Figure A.3: *Kurtosis vs. skewness of the intensity-distribution histogram of the envelope on Nov. 7.*

A.2.2 Nov09

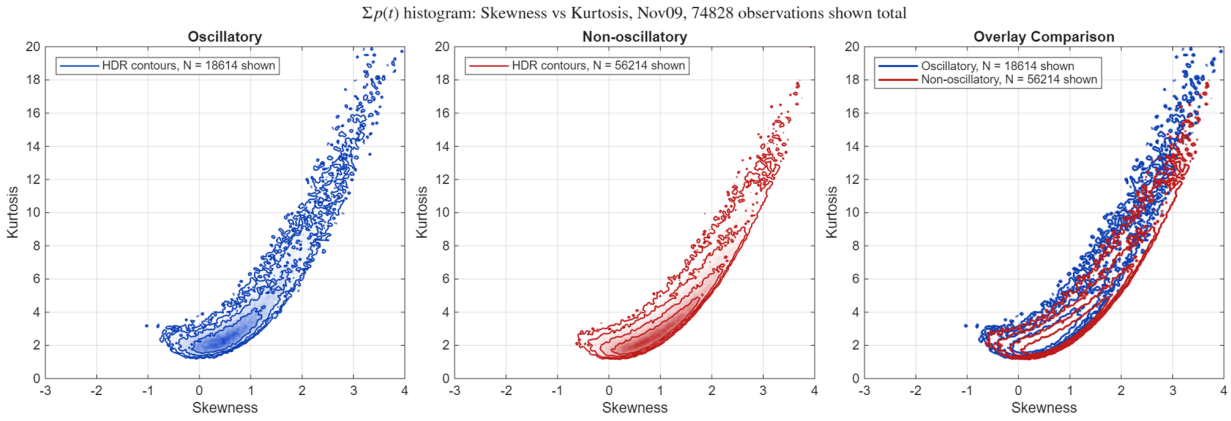

Figure A.4: *Kurtosis vs. skewness of the intensity-distribution histogram of* $\Sigma p(t)$ *on Nov. 9.*

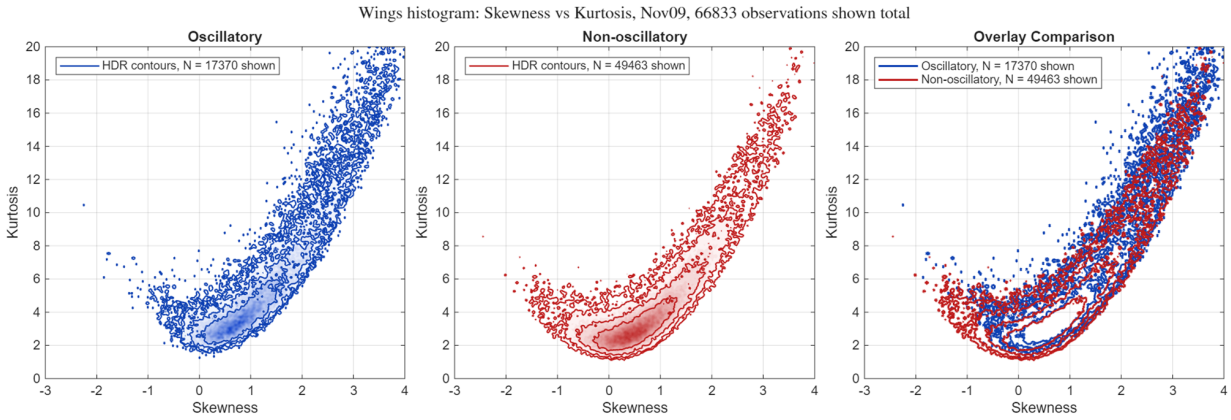

Figure A.5: *Kurtosis vs. skewness of the intensity-distribution histogram of the envelope-normalized $\Sigma p(t)$ on Nov. 9.*

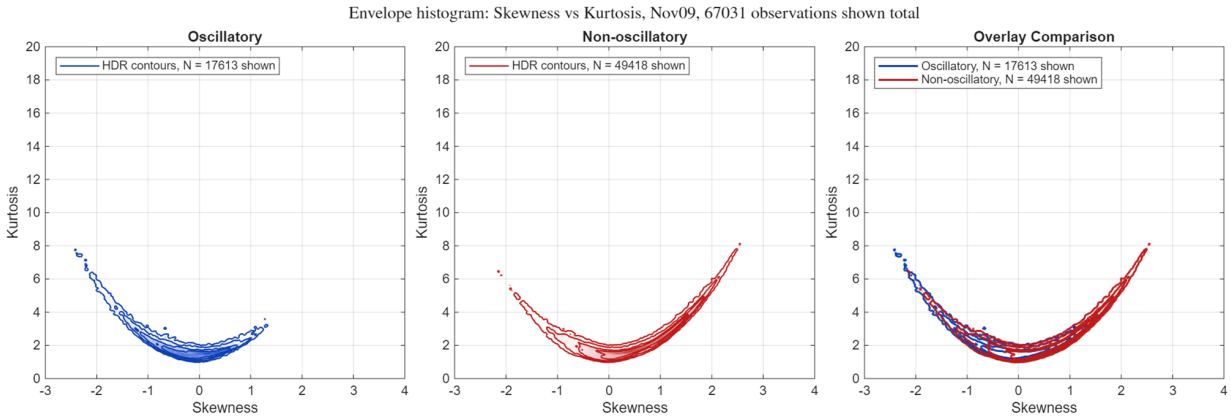

Figure A.6: *Kurtosis vs. skewness of the intensity-distribution histogram of the envelope on Nov. 9.*

A.2.3 Nov10

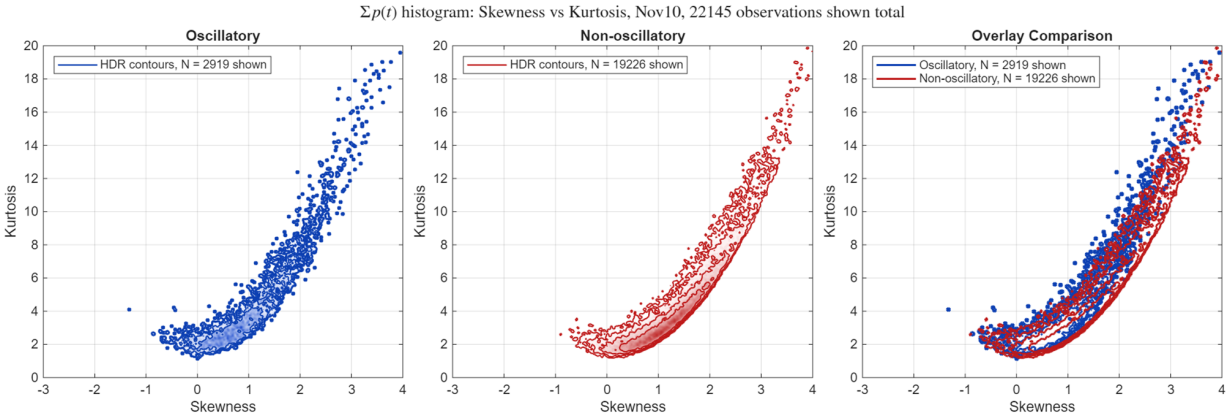

Figure A.7: *Kurtosis vs. skewness of the intensity-distribution histogram of* $\Sigma p(t)$ *on Nov. 10.*

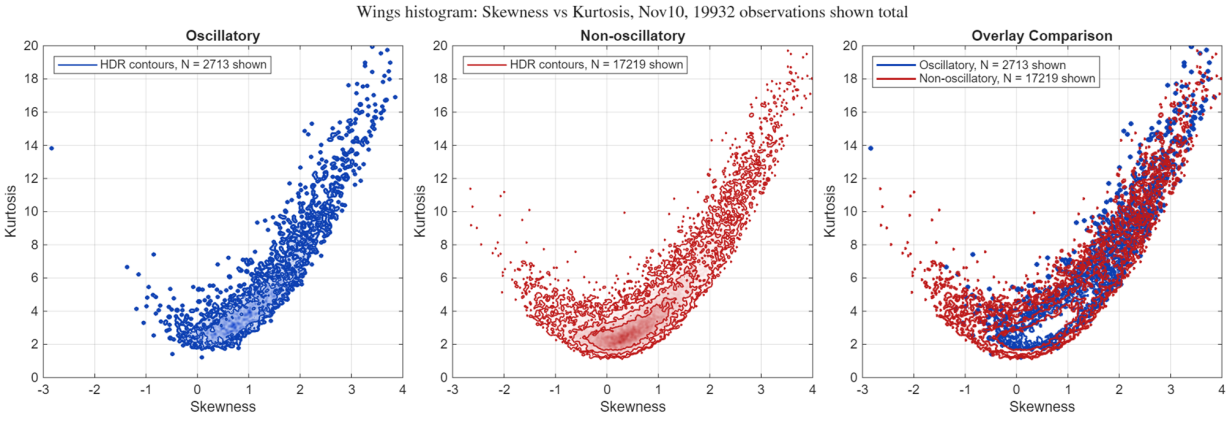

Figure A.8: *Kurtosis vs. skewness of the intensity-distribution histogram of the envelope-normalized* $\Sigma p(t)$ *on Nov. 10.*

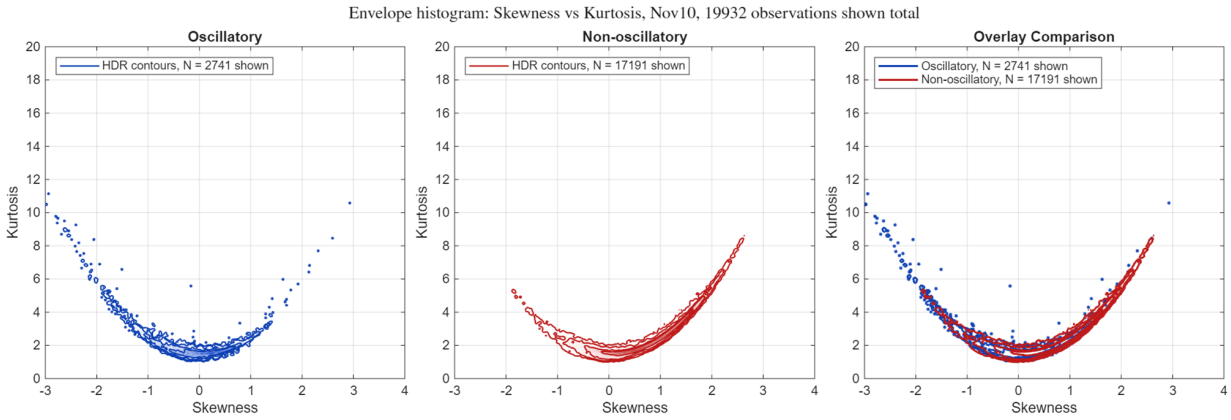

Figure A.9: *Kurtosis vs. skewness of the intensity-distribution histogram of the envelope on Nov. 10.*

A.2.4 Nov12

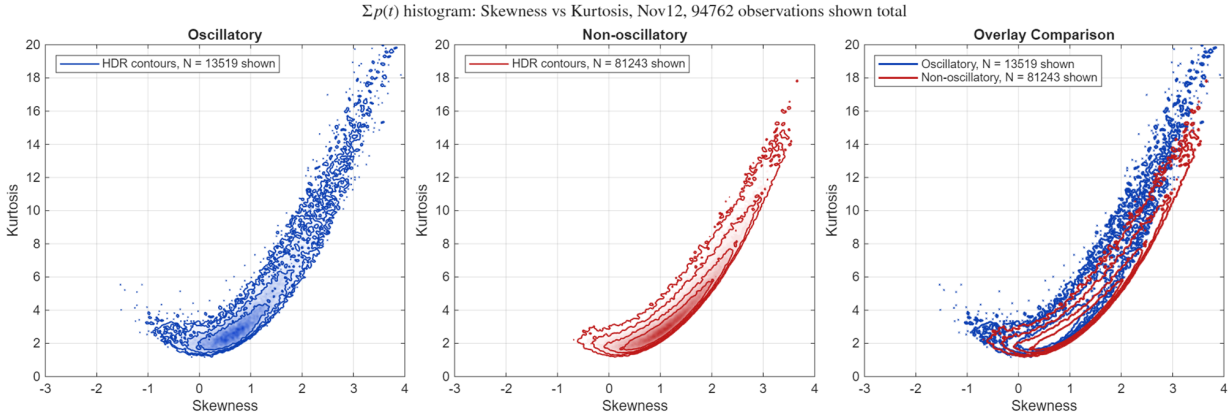

Figure A.10: *Kurtosis vs. skewness of the intensity-distribution histogram of* $\Sigma p(t)$ *on Nov. 12.*

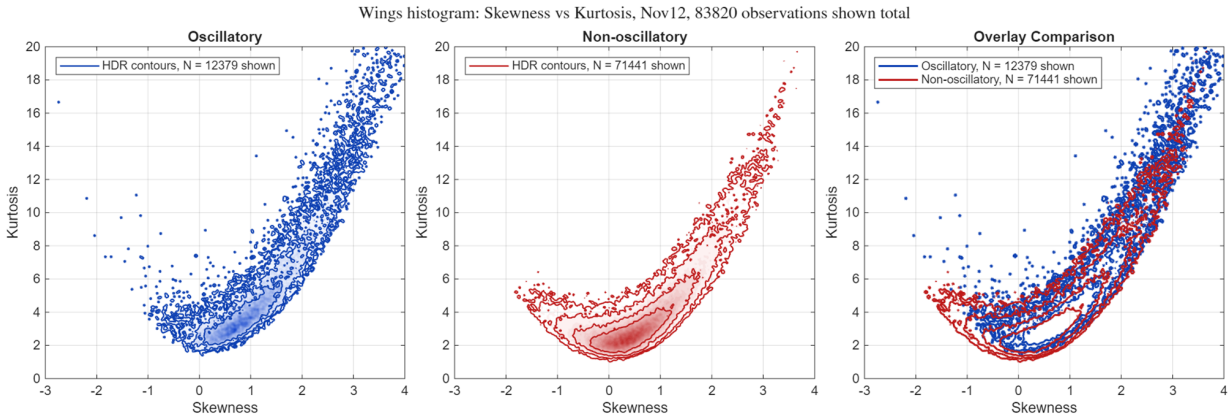

Figure A.11: *Kurtosis vs. skewness of the intensity-distribution histogram of the envelope-normalized $\Sigma p(t)$ on Nov. 12.*

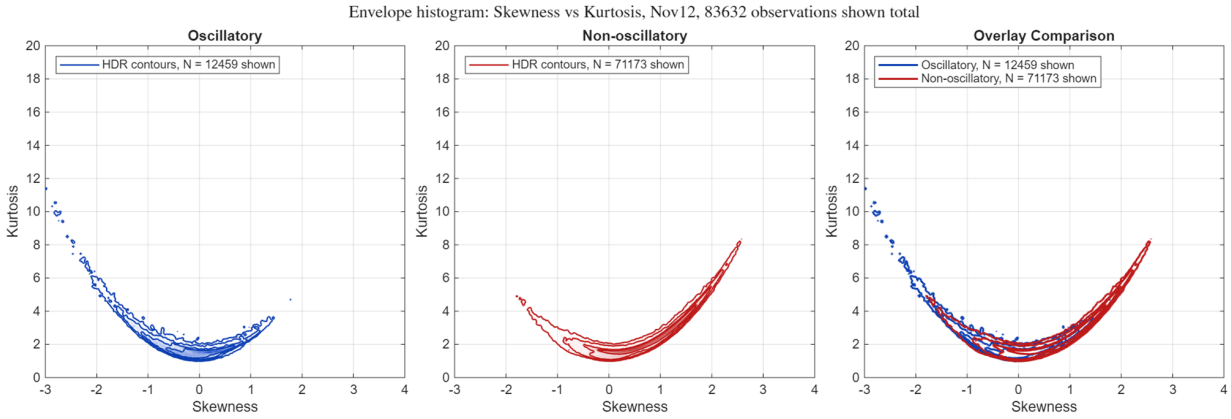

Figure A.12: *Kurtosis vs. skewness of the intensity-distribution histogram of the envelope on Nov. 12.*

A.3 Technical details of ACF method

A description of the ACF method in detail is presented in this section. The autocorrelation function (ACF) was utilized to determine periodic intensity modulations. In this thesis, and integer shift was chosen. The analyzed signal is shifted by an integer number of samples, denoted by lag index k. One sampling interval Δt is given by the inverse of the sampling frequency of the system:

$$\Delta t = \frac{1}{1750\text{Hz}} \approx 0.571 \text{ ms}. \tag{A.1}$$

A shift of k samples is thus a physical time-lag of

$$\tau_k = \Delta t \cdot k \approx 0.571 \cdot k \text{ ms}. \tag{A.2}$$

For an intensity signal,$I_{\text{wing}}(t)$, the normalized ACF is defined as

$$\text{ACF}(\tau) = \frac{\sum_t \left(I_{\text{wing}}(t) - \bar{I}_{\text{wing}}\right)\left(I_{\text{wing}}(t+\tau) - \bar{I}_{\text{wing}}\right)}{\sqrt{\sum_t \left(I_{\text{wing}}(t) - \bar{I}_{\text{wing}}\right)^2 \sum_t \left(I_{\text{wing}}(t+\tau) - \bar{I}_{\text{wing}}\right)^2}}, \tag{A.3}$$

where

$$\bar{I}_{\text{wing}} = \frac{1}{N}\sum_t I_{\text{wing}}(t), \tag{A.4}$$

is the sample mean of the signal. Subtracting the mean removes any constant offset of the signal which will ensure the correlation only responds to the wingbeat peaks.

Integer lag was implemented in this thesis, meaning the ACF searches through a discrete set of WBF, relating to the signal duration and sampling frequency. The first integers $k = 0, 1, 2$ were excluded as the criterion is that at least three wingflashes need to be present to resolve the WBF. The wingbeat period is found as:

$$T_{\text{WBF}} = \frac{T_{\text{samples}}}{f_s}. \tag{A.5}$$

For a signal detected with a period of T_{samples} samples, then

$$T_{\text{WBF}} = \frac{T_{\text{samples}}}{f_s}, \tag{A.6}$$

The corresponding WBF is then

$$f_{\text{WBF}} = \frac{1}{T_{\text{WBF}}} = \frac{f_s}{T_{\text{samples}}}. \tag{A.7}$$

A.4 Surrogate significance test

To determine whether the detected periodicity is significant, a surrogate test is performed. For each observation, the signal is randomly permuted N_{surr} times. The autocorrelation peak of the original signal is then compared to the distribution of autocorrelation peaks from the randomized signals.

The significance score is calculated as

$$Z = \frac{A_{\text{obs}} - \mu_{\text{surr}}}{\sigma_{\text{surr}} + \epsilon}, \tag{A.8}$$

where A_{obs} is the largest autocorrelation peak of the original signal, while μ_{surr} and σ_{surr} are the mean and standard deviation of the surrogate autocorrelation peaks. The P-score is calculated as:

$$p = \frac{\#\{A^*_m \geq A_{\text{obs}}\} + 1}{N_{\text{surr}} + 1}, \tag{A.9}$$

where A_m^* refers to the largest ACF peak of the mth surrogate signal.

$Z = n$ corresponds to non-permuted ACF-peak that is n standard deviations from the randomly permuted sample, providing strong evidence of periodic structure which cannot be attributed to noise. The $P - value$ measures the probability that a randomized version of the data also would produce a high-correlation overlap, such as the wingbeats lining up. A large z and small p indicate significant oscillatory structure.

The observation is then classified according to:

$$\begin{cases} Z > 3, & \text{YES}, \\ Z > 1.96, & \text{LIKELY}, \\ Z > 1, & \text{MAYBE}, \\ Z \leq 1, & \text{NO}. \end{cases} \tag{A.10}$$

The cases of YES and LIKELY were labeled as oscillatory.

A.5 WBF Limitations of ACF

Due to the nature of the ACF-method, it finds several peaks of high correlation. One of these is likely the WBF. The detection of oscillations is robust, while the exact WBF is not consistently picked. Due to the integer k lag time-shifts chosen, the detectable WBF are quantized according to

$$\text{WBF}_k = \frac{1750}{k}, \quad k = 2, 3, \ldots \tag{A.11}$$

For an observation of length T_{Obs}, the number of samples is

$$N_t = \lfloor T_{\text{obs}} \cdot f_s \rfloor + 1. \tag{A.12}$$

We evaluate the signal using time-shifts of lag k in range

$$k \in [3,\, N_t - 1]. \tag{A.13}$$

Thus, for increasing signal duration we search between more lags k, and are able to obtain Since the signals are chosen to be above 25ms, this corresponds to lower threshold of 44 samples. Thus we can obtain

$$\frac{1750}{3}, \frac{1750}{4}, \frac{1750}{5}, \ldots \frac{1750}{43} = 41 \text{ possible WBF values.} \tag{A.14}$$

If we suppose the insect signal is 100ms, we can obtain

$$N_t \approx 176 \implies k \in [3, 175] \implies 173 \text{ possible WBF values.} \tag{A.15}$$

The implementation utilized here searches for ACF-peaks within a *time of interest* (TOI), where the envelope signal(here referred to as *P0env*) is strong, and not too diffuse, according to

$$TOI = find((P0env > 0.1 * max(P0env)). * (P2 > 1)); \tag{A.16}$$

where the envelope is above 10% of its max value, and the spread(P2) is limited to decrease prevalence of noisy signals.
The consequence of these details are that longer signals have more possible WBF values since we search over more k. In the presented examples in Figure 4.7, the signal is $\approx$ 150 ms. This would mean WBF in the range

$$k = 3 : 538.3\text{Hz} \implies k = 262 : 6.7\text{Hz}. \tag{A.17}$$

For a signal such as in Figure 4.9, where the signal is $\approx$ 50ms, we can only search for 8 WBF values between 150 and 500;

$$\text{WBF}_k = \frac{1750}{k}, \quad k \in \{4, 5, 6, 7, 8, 9, 10, 11\} \implies \tag{A.18}$$

$$\text{WBF} \in \{437.5,\ 350,\ 291.7,\ 250,\ 218.75,\ 194.4,\ 175,\ 159.1\}\ \text{Hz} \tag{A.19}$$

As displayed above, we cannot detect any WBF between 350 Hz and 437.5Hz. However, mosquitoes exhibit WBF in ranges 400-500Hz in females, and 600-800Hz in males [45]. Thus it may still be feasible to attribute a WBF of 437.5Hz to be a probable female, and a 583.4 Hz to male. It is however still difficult to categorize insects based solely on the WBF, as the insects displaying the same WBF may differ in size [46]. Further analysis on the difference in body size, σ_1^0 in Table 3.1, between the signals with WBF 437.5Hz may be able to determine whether these differ significantly, which would indicate different species.

A.6 Example of data fields

The *obs* datafields are presented in two parts below.

1×22238 struct with 13 fields

Fields	fno	fname	ft	rInd	tInd
1	6000	'Frame Cam1 2020-11-11 15-06-21-248.dat'	7.3811e+05	[1457,1458,1459,1460,1461,1462,1463,1464]	*1×43 double*
2	6000	'Frame Cam1 2020-11-11 15-06-21-248.dat'	7.3811e+05	*1×18 double*	*1×40 double*
3	6000	'Frame Cam1 2020-11-11 15-06-21-248.dat'	7.3811e+05	*1×13 double*	*1×43 double*

Figure A.13: *First half-snapshot of the* ***obs*** *data, where three observations are seen.*

dt	dsamp	rt	bs	ext	satPix	rtNoise	bsNoise
0.0287	43	8×43 int16	1×43 single	1×43 single	0	8×43 single	1×43 single
0.0267	40	18×40 int16	1×40 single	1×40 single	0	18×40 single	1×40 single
0.0287	43	13×43 int16	1×43 single	1×43 single	0	13×43 single	1×43 single

Figure A.14: *Second half of the snapshot of the* ***obs*** *data of the same three observations as in Figure A.13.*