Smart crop row detection in a soybean crop – results experiments in 2024

Janne Kool[1], Eva de Jonge[1], Ard Nieuwenhuizen[1]

1 Wageningen University & Research

This study was carried out by the Wageningen Research Foundation (WR) business unit Agrosystems Research. and was commissioned and financed by Topsector TKI under grant agreement number LWV20.242 "Smart technology for soybean production" Projectcode: BO-69-001-005

WR is part of Wageningen University & Research, the collaboration of Wageningen University and Wageningen Research Foundation.

Wageningen, March 2025

Janne Kool, Ard Nieuwenhuizen, Eva de Jonge, 2025. Smart crop row detection in a soybean crop – results 2024. Wageningen Research

Photo cover: Vinicius Marini

Contents

Summary

The objective of WP1 in the final year 2024 of the project is to investigate if it is possible to detect soybean crop rows, for different angles between the driving direction and the crop rows using the WEED-IT sensor from partner Rometron. In the previous report published for this project (Janne Kool, 2023) the distinctiveness of the measurements with this sensor is introduced. The distinctiveness measures how well the sensor can distinguish between soil and crop row when driving skew over crop rows in the field. This distinctiveness is dependent on the width of the field of view of the sensors. To compare the distinctiveness and to assess the possibilities to determine the driving angle from the data, a soybean trial field was planted in Brazil. A quad vehicle equipped with the WEED-IT sensor placed half the height as in earlier seasons was used for the measurements, and these took place at 0- , 15- , 30- and 45- degree angles towards the crop rows. In the data analysis a Fourier approach was used, analogue to the approach in 2023. There is no clear peak in the Fourier transform around the expected frequencies, so it is not possible to find the unique crop row positions. The percentage of soil measured, as approximation of the distinctiveness, is compared with the measurements in 2023 around the same growing stage. This percentage is indeed higher for the measurements done with a mounted lower sensor.
On the dataset collected in the 2022-2023 season in Brazil an algorithm has been developed to predict the locations of the crop rows two meters ahead using a shorter distance. This is based on Fourier analysis and the concatenating of channels behind each other. It is evaluated how well the predictions are based on the driving distance and several down sampling rates.

1 Introduction

The goal of green-on-green measurements is to detect weeds in a crop field, in this case a soybean field using the WEED-IT sensor. Since the WEED-IT sensor is based on a technology that can only distinguish living plant material from no living plant material, it does not distinguish weeds from crop plants. Therefore, distinction has to be based on plant location. That is, plants that are located outside the plant rows are considered weed, and vice versa: plants in the crop rows are considered crop.

It is common practice in Brazil to drive with an angle towards the rows when broadcast spraying is applied. In the report published earlier in this project (Janne Kool, 2023) a method using the Fourier transform is explored to detect weeds between the rows. In that research, a measure called distinctiveness is developed with which it can be assessed how well the sensor can distinguish between soil and crop row when driving skew over the field. For convenience of the reader the explanation about the distinctiveness has been included again in this report.

In the growing season of 2023-2024, an experiment has been conducted to evaluate whether it is possible to detect these weeds between the crop rows using the WEED-IT sensor, while driving in different angles towards the crop rows. Also the height of the sensor has been halved to increase the distinctiveness.

In the data collected in the 2023-2024 season there is no clear peak in the region of interest in the frequency domain, that is the frequency in which the crop rows are expected to appear, for most of the measurements. Hence it is not possible to the deduce the location of the rows.

Using the data of the 2022-2023 season an algorithm to predict the crop rows two meters ahead is developed. For this algorithm it is assessed how many meters of data is needed to make good predictions, and, moreover, how much down sampling is allowed to do so.

The work in this report is made possible with the help of the following colleagues and partners in the project. Mainly to help with the field trials in Brazil, as well as with providing feedback on the methods and results generated. Vinicius Marini, Milena Martinez Albeche, Fernando Furian Cossentin, Bruno Garlet Alberti, Jeferson Picetti Fontoura dos Santos, Angelo Roberto Mohr, Telmo Amado, Lucio Amarol, all from Universidade Federal de Santa Maria, Hendrik Braam, Alwin Meenderink, both from Rometron B.V., Thomas Liska, Leonardo Seibel Sander, Tiago Rosado from STARA, Mateus Tonini Eitelwein from Smart Agri Sensing and Willem de Visser, Annelies Berniers and Jean Marie Michielsen from Wageningen University and Research.

Chapter 2 explains the distinctiveness results on 2024 experiment in Brazil, with a lowered sensor system, half the height of 2023. Chapter 3 explains the algorithm developed for predicting the crop rows two meters ahead with a short distance driven, and the evaluation of these prediction with several distances and down sampling rates.

2 Soybean field experiment Brazil 2024

In the Brazilian soybean season of '23 –'24, experiments were conducted in Brazil. In the soybean field where the experiments are done, weeds had to be controlled using herbicides. Therefore, the measurements were conducted the day before the herbicide application. This leads to a very realistic measurement, as we are in a 'normal' field, just before a 'standard' herbicide application event. The weather conditions at the start of the season were challenging for soy production and seedlings did not emerge sufficiently. A few weeks after planting, all seedlings are removed and the sowing is done again.

Measurements with the WEED-IT took place on February 1st and 2nd, 2024. The weed locations were annotated a couple of days later. A second measurement round is performed on February 20th 2024.

2.1 Background: distinctiveness explained

In Brazil it is common practice to drive in an angle of approximately 30 degrees towards the crop rows instead of parallel. This generates a possibility to see a pattern in the crop rows with the WEED-IT sensor.

The WEED-IT sensor, and the datalogger connected to the sensor, both measure with a frequency of 14 kHz in the driving direction. The sensor also has a width: it has four channels, which measure beside each other. With the sensor hanging at 110cm above the soil, the measuring width of one channel is 25 cm.

Because of this width, the boundaries of a channel can be above a crop row while the center of the channel isn't. This is illustrated in Figure 1. The diagonal arrow pointing to the upper right is the driving direction and the two lines perpendicular to the arrow in the bottom illustrate that the measurement channels have a width (SW: sensor width) that can partly cover a crop row.

When a measurement channel *partly* covers a crop row, the sensor reacts the same as when the *entire* channel covers a crop row. The channel either detects plant material or it doesn't, as the signal is the average of all occurrences of crop material in the width of the channel.

We therefore define:

- *D*: The *distinctiveness* of the setup is the *fraction of soil* measured in a channel in which no part of the channel is above the crop. Distinctiveness is expressed as a number between 0 and 1, and calculated from the parameters below.
- *d*: The *distance*, in the driving direction, between the foliage of the two crop rows
- *sd*: Is the distance that the centre of the sensor has driven, while *a part of the sensor is above the crop row*. The higher *sd*, the lower the distinctiveness *D*.
- *SW*: the *sensor width*, the width of the sensor beam (per channel). SW varies with the height at which the sensor is mounted (the width increases with the distance between sensor and the soil).
- *a*: the *distance* between the canopies of two crop rows. *a* is by definition smaller than the row distance, and further decreases as the plants grow and the canopy closes the area between two rows.
- α: the angle between the driving direction and the crop rows.

Note that *D, d* and *sd* depend on the driving angle α of the setup, distance *a* between the crop foliage, and sensor width *SW*.

And note that the closer the distinctiveness D of the measurement setup reaches 1, the higher the fraction of 'bare soil' that is recorded (and the lower the fraction of crop canopy). A distinctiveness of 0 means that the sensor is always above a crop row (so driving parallel, or above a closed canopy), while 1 means that the setup would be driving perpendicular to the rows, so the rows are recorded by all channels at exactly the same time.

A higher distinctiveness equals a more precise detection of weeds, because it makes distinguishing weeds closer to the crop rows possible. Note that this means that the bigger the angle α, the higher the distinctiveness and the easier it is to detect weeds between the crop rows.

All parameters are illustrated in Figure 1.

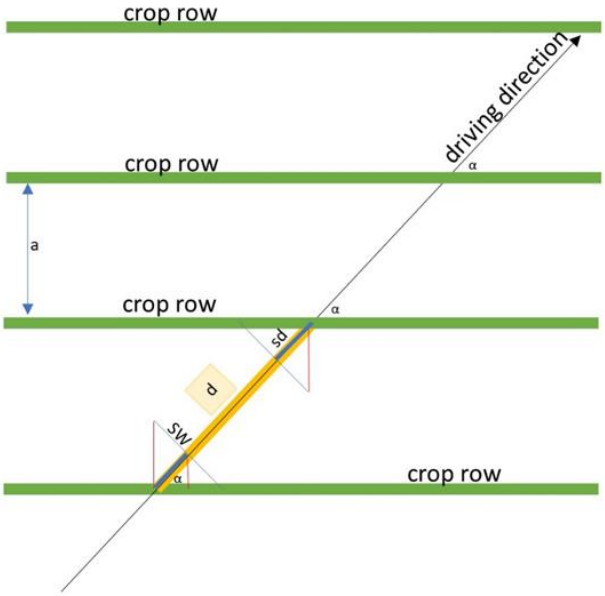


Figure 1: A schematic sketch of the crop rows and the sensor moving, illustrating the main parameters for the calculation of the distinctiveness D.

The distance in driving direction *d* can be calculated: $d = a/\sin(\alpha)$.
The distance where a part of the sensor is above the crop row *sd* = $0.5*SW*\cos(\alpha)/\sin(\alpha)$.

The distinctiveness can then be calculated as follows:

$$D = (d-2*sd)/d$$
$$= 1-SW\cos(\alpha)/a$$

Note that when the driving direction is parallel to the crop rows, α=0. Distances *d* and *sd* cannot be defined in this case and D cannot be defined either. Moreover, if $2*sd>d$, there is always some canopy under the sensor. *D* attains negative values as a result.

To get a feeling on how the distinctiveness is dependent on the parameters, see Figure 2 and Figure 3. In Figure 2 the distance between the canopy of two rows *a*=30 cm. In Figure 3, *a*=20 cm, and *D* can become negative. So while the crop grows, distance *a* shrinks, reaching 0 when the canopy closes.

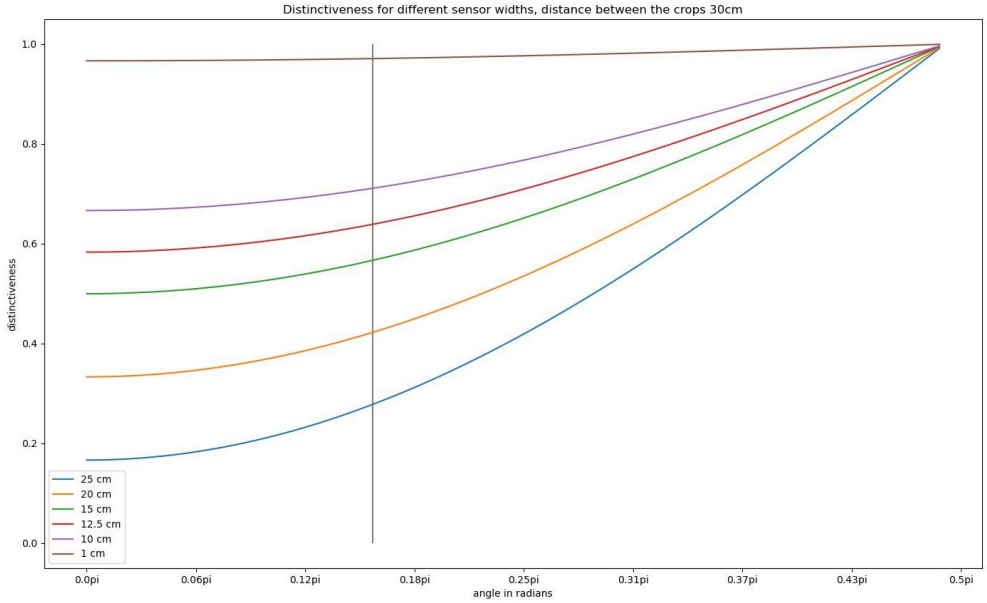


Figure 2: The distinctiveness plotted against the angle between the driving direction and the crop row when the distance between the foliage of the crop rows is 30cm. The vertical line is the angle which was used during the experiments. A distinctiveness of 0.5 means that half of the distance that the center of the sensor is above the soil, part of the sensor is above the crop. The sensor is entirely above the soil for the remaining distance (time in case of equal speed).

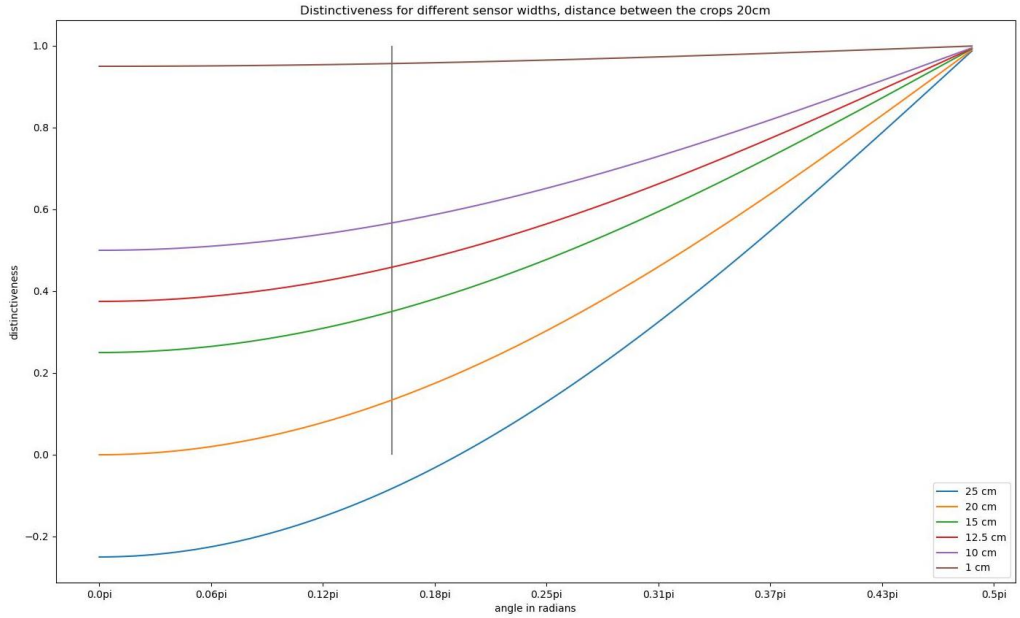


Figure 3: The distinctiveness plotted against the angle between the driving direction and the crop row when the distance between the foliage of the crop rows is 20cm. The vertical line is the angle which was used during the experiments. A negative distinctiveness means that the sensor is always measuring crop and will not measure any bare soil.

2.2 Experimental setup

The distinctiveness (D) of the measurements in 2022-2023 was rather low. The sensor was mounted 110 cm above the soil that season, which corresponds to a measurement width of 25 cm per channel. Since the row distance is 45 cm, crop canopy was detected in most of the measurements. Therefore, in this experiment the height of the sensor is set to 55 cm, and consequently the measurement width of each channel to 12.5 cm.

A smaller measurement width (or sensor width SW) increases the distinctiveness, and we assume that it increases the possibility of detecting weeds between the crop rows.

The WEED-IT sensor, together with a GoPro camera, was mounted on a quad. The GoPro recordings were used as reference data.

As said in paragraph 2.1, spraying is often done in an angle towards the crop rows. In practical conditions, angle can differ between- and within fields however. Therefore measurements were done in three different angles toward the crop rows, and parallel to the crop rows. This was done to see if the angle can be detected from the signal. To separate the plots, plants were removed between the plots. Figure 4 shows a schematic sketch of the experimental set-up.

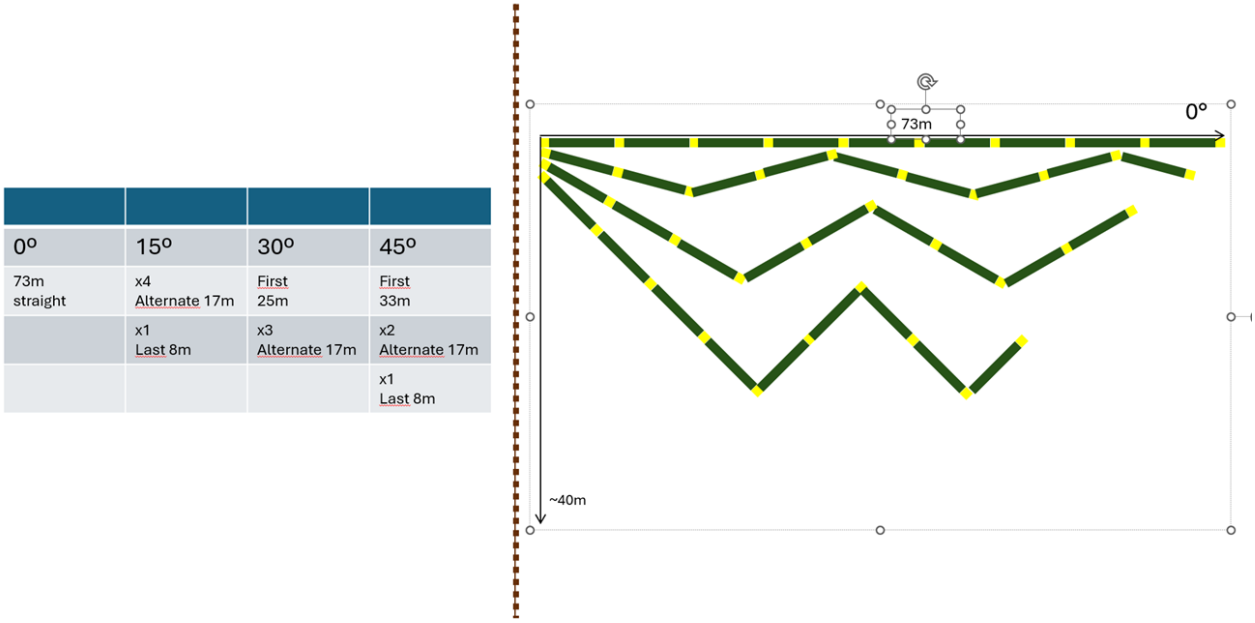


Figure 4: A schematic sketch of the experimental setup. By Vinicius Marini.

Table 1 below shows the effect of each driving angle on the distinctiveness (expressed in percent). The distinctiveness is computed with the assumption that the width of the crop row is 0. In reality crop rows have a positive width (increasing over the season) and the distinctiveness is smaller.

Table 1: The distance d *between the rows in the driving direction and the theoretical distinctiveness for each driving angle α.*

Angle between crop rows and driving direction *α*	**Distance in m between the rows in the driving direction *d***	**Distinctiveness (%)**
0	n.a.	n.a.
15	1.73	73
30	0.9	76
45	0.64	80

The plots were about 17 meters long, with a 45 cm row distance. The size and growing stage of the soybean plants varied a lot at the beginning of the season, and emergence was not good. Some pictures of the experimental fields are in Figure 5, showing plot numbers, a measurement tape. The figure also shows the irregular growth stage of the crop.

Figure 5: Pictures of the experimental fields on February 1st, 2024. Irregularity in the crop is visible.

2.3 Description of the data

Due to the weather conditions, the soybean plants did not emerge evenly across the fields (see also Figure 5). The field design, which involved removing green vegetation between plots to accurately detect plot locations in the data, did not work as intended during the first measurement on February 1st. Moreover, some plot areas remained empty, further complicating the process. Additionally, there was a gap of several days between the WEED-IT sensor measurements and the weed annotation. As a result, some weeds that were found during the annotation, were not found in the GoPro footage.

For each measurement, graphs are made of (see for example Figure 6):

1. The original signal (left, blue line);
2. The absolute value of the Fourier transform (right, blue);
3. The filtering of the Fourier transform (right, orange);
4. The Fourier inverse of the Fourier transform (left, orange).

For an explanation of how these graphs are made, see (Janne Kool, 2023), section 4.2.

In the frequency domain for the Fourier filtering, a 'region of interest' was selected. This was done using the expected frequencies, based on the angle α and crop row distance. The distances between the crop rows in the driving direction determine an expected frequency: when the distance is d, the expected frequency is 1/d. The region of interest was selected to be between 1/(d+0.2) and 1/(d-0.2). The distances in the driving direction can be found in Table 1.

For the measurements of February 2nd, a piece of the 15, 30 and 45 degree angle measurements is shown as an example in Figure 6, Figure 7 and Figure 8 respectively.

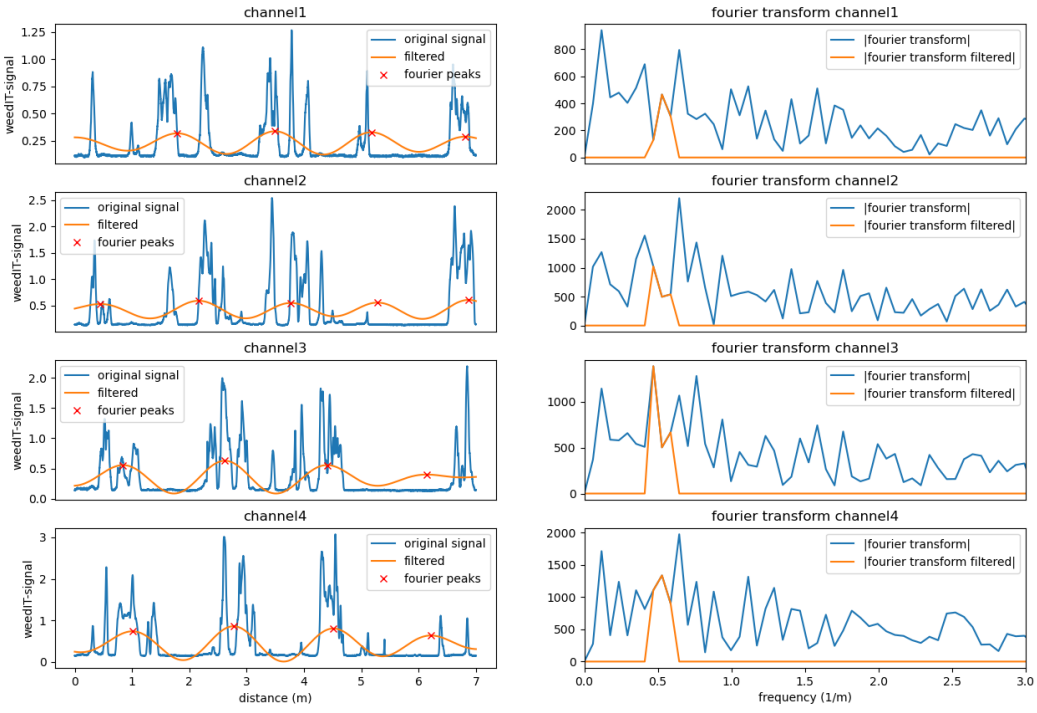


Figure 6: (Left) the signal, the Fourier filtered signal and the peaks of a plot measured on the 2nd February 2024. (Right) the absolute value of the Fourier transform and the filtered domain. The angle between the crop and the driving direction was 15 degrees.

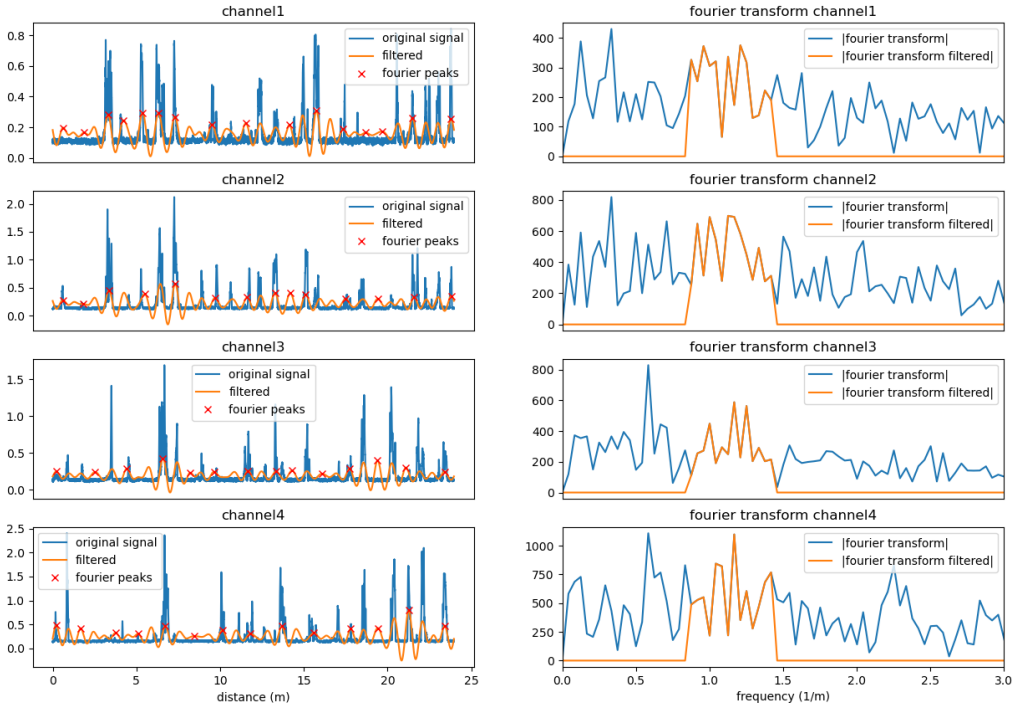


Figure 7:(Left) the signal, the Fourier filtered signal and the peaks of a plot measured on the 2nd February 2024. (Right) the absolute value of the Fourier transform and the filtered domain. The angle between the crop and the driving direction was 30 degrees.

Figure 8:(Left) the signal, the Fourier filtered signal and the peaks of a plot measured on the 2nd February 2024. (Right) the absolute value of the Fourier transform and the filtered domain. The angle between the crop and the driving direction was 45 degrees.

The soybean crop growth stage increased a lot between February 2nd and 20th. As the sensor was placed relatively low, the sensor touched the crop during measurement on February 20th. This led to complete saturation of the signal (see Figure 4, Figure 9 and Figure 10).

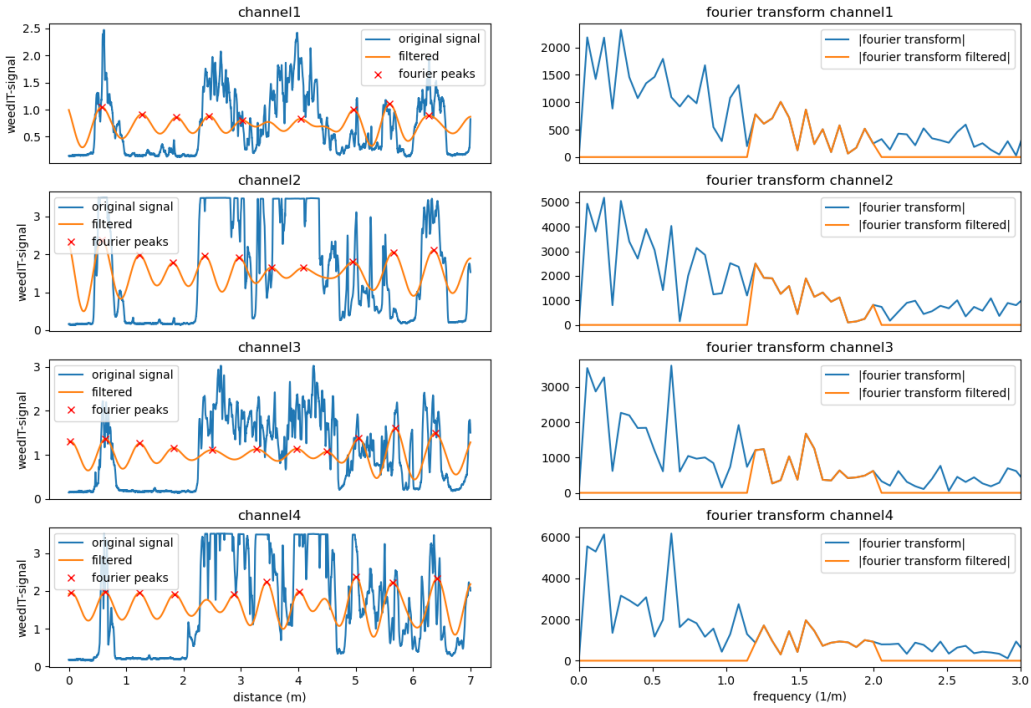


Figure 9: (Left) the signal, the Fourier filtered signal and the peaks of a plot measured on the 20th February 2024. (Right) the absolute value of the Fourier transform and the filtered domain. The angle between the crop and the driving direction was 15 degrees.

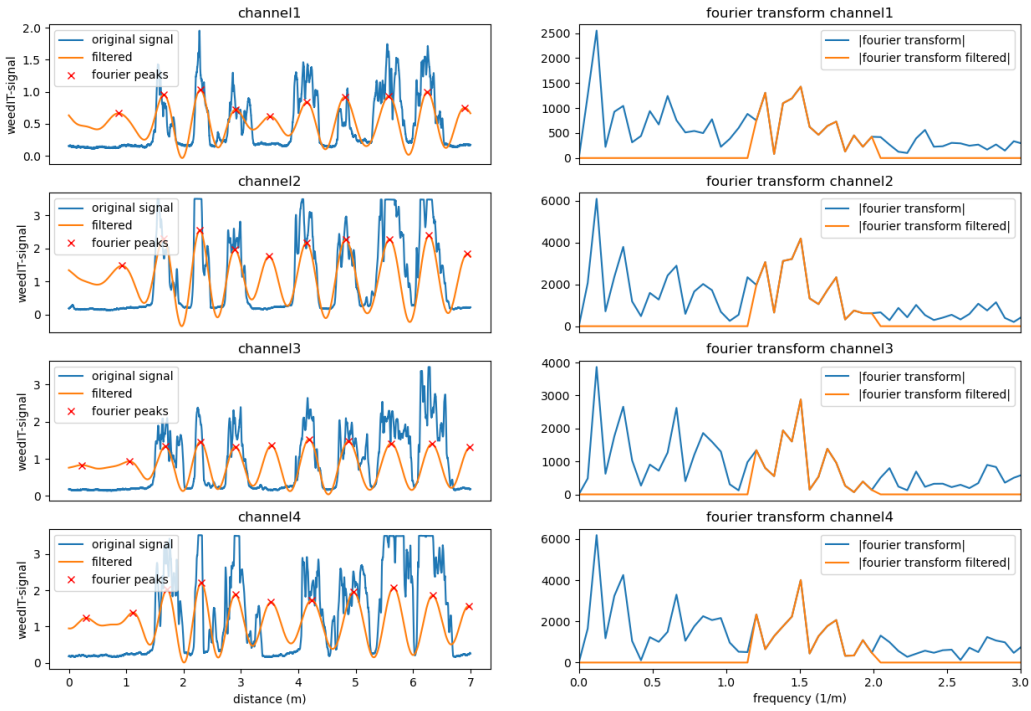


Figure 10:(Left) the signal, the Fourier filtered signal and the peaks of a plot measured on the 20th February 2024. (Right) the absolute value of the Fourier transform and the filtered domain. The angle between the crop and the driving direction was 45 degrees.

In contrast to the measurements in the season '22-'23 there is no clear peak in the region of interest in the frequency domain for most of the measurements, hence it is not possible to deduce the frequencies an the locations of the crop rows.

2.4 Quantifying distinctiveness

To see the difference in distinctiveness with the different heights of the sensor, and hence the widths of the channels, the percentage of the plant material is quantified during the measurement on January 3th 2023 and February 1st 2024. As approximation for the distinctiveness, 100 minus the percentage is used, which is the part of the measurements where no plant material is detected. The days are chosen such that the growth stage is comparable between the two years, but in 2024 some parts of the plots did not emerge well, or the plants between the plots were not removed sufficiently to be able to distinguish between plots.

To determine the amount of plant material, the Gaussian Mixture model approach was used as described in section 4.3 of (Janne Kool, 2023), with a look back distance of 0.1 meter. The percentages of plant material and the approximations of the distinctiveness are summarized in Table 2.

Table 2: The measured distances and the percentage of plant material on January 3th 2023 and February 1st 2024. As the sensor has 4 channels, the distance measured is four times the accumulated plot length, distance plant is the distance on which the GMM model detects plant material, the percentage plant is distance plant/distance measured, and approximation distinctiveness is 100 minus percentage plant.

Degrees	Accumulated plot length (m)	Distance measured (m)	Distance plant (m)	Percentage plant (%)	Approximation distinctiveness (%)
15 (2024)	76	304	90.55	29	71
30 (2024)	72.1	291.4	44.57	15	85
45 (2024)	73.7	294.8	53.23	18	82
30 (2023)	110	440	210	47	53

2.5 Conclusions on distinctiveness based on field experiment Brazil 2024

The crop did not emerge well in 2024, making the rows difficult to distinguish. Also, the annotation of the weeds was not precise enough to be useful for weed detection. However, the differences in the angles between crop rows and driving direction are clearly visible in the plots.

The distinctiveness was higher in 2024 than in 2023, as was expected due to the lower height of the sensor, creating a narrower measuring width, resulting in more measurements where no part of the sensor was above a crop row. The expectation was that the distinctiveness of the measurements taken under a 45 degree angle would be higher than under 30 degrees, but that was not the case. Both 45 and 30 degrees had a higher distinctiveness then the 15 degrees.

3 Detecting crop rows in 2022 and 2023 field experiment data

3.1 Introduction

To detect where there are supposed to be crop plants, and where there are not supposed to be plants, it is important to locate the crop rows. Doing so based on the frequency, it is desirable to predict the locations of the rows about two meters ahead. That is expected to be a workable distance for operating the spraying devices on broadcast field sprayers.

In reality, the angle between the crop and the driving direction will vary a bit during the operation. Therefore, it is desirable to collect the important parameters for row detection in a short period of driving, preferably within a few meters. To do so, a measure is developed, the so called *signal-to-noise-ratio* (SNR). With this SNR the dominancy of the dominant frequency can be assessed within a relatively short amount of data gathered.

When driving, the length of the measured signal can be made four times longer by combining the signals of the four channels in one WEED-IT sensor. However, the *phase* of each of the channels differs a bit, depending on the angle α between the crop row and the driving direction; when channel1 is on top of a certain crop row, channel4 is not, except if the angle is exactly 90 degrees and we drive perpendicular to the crop rows. The SNR is theoretically a tool to determine the *real* driving angle α; when the SNR is computed for several angles, it will peak at the real angle. However, the assessment of this method led to the conclusion that this method is not accurate enough yet in practice conditions.

Given the inverse Fourier transform and the dominant frequency, a prediction is made two meters ahead of the sensor and spray boom, by simply using a cosine function as prediction, with the dominant frequency two meters ahead, starting from the last peak found in the inverse-Fourier transform. The score evaluating this prediction evaluated how much plant material is around the peaks of the cosine. A summary of the algorithm is explained in pseudo-code in Figure 11. The details of the algorithm follow in the paragraphs below. Using the algorithm, an assessment is carried out to evaluate the driving distance needed and the down sampling rate allowed, to come to good predictions. For this analysis, data collected in field experiments on 27-12-2022, 03-01-2023 and 10-01-2023 is used.

Algorithm summary

Notation The signal $X = [c_1, c_2, c_3, c_4]$ comprises four channels, with for each mm a signal value. For a channel c_i, the last m meters are given by $c_i[-m:]$. Two channels c_i, c_j put behind each other are denoted by $c_i + c_j$. The Fourier transform of a channel is denoted by $\tilde{c}_i = \text{FFT}(c_i)$, likewise the inverse Fourier transform is given by $\text{IFFT}(f_i)$. Values in the distance domain are denoted by x, and values in the frequency domain are denoted by f.

Input:	$X = [c_1, c_2, c_3, c_4]$, meter (m)
For $\alpha \in (15, 45)$ **compute:**	φ_α =phase shift. $X_\alpha = c_1[-m:] + c_2[-m + \varphi_\alpha :] +$ $c_3[-m + 2\varphi_\alpha :] + c_4[-m + 3\varphi_\alpha :]$ (s_α, e_α) region of interest in frequency domain $\tilde{X}_\alpha = \text{FFT}(X_\alpha)$ filtered to (s_α, e_α)
Determine optimal α:	
$\bar{\alpha} = \text{argmax}_\alpha\{\text{SNR}(\tilde{X}_\alpha)\}$	the angle with the largest SNR
$f_{\text{dom}} = \text{argmax}_f\{\tilde{X}_{\bar{\alpha}}\}$	the dominant frequency for $\bar{\alpha}$
$lp_i = \max\{x \mid \text{IF}\tilde{\text{F}}\text{T}(X_{\bar{\alpha}})(x) \text{ is a peak}\}$	distance of the last peak in filtered inverse Fourier transform in channel c_i
Predict and score:	
$p_i = \cos(2\pi f_{\text{dom}}(x - \text{lp}_\text{i})),$ $lp_i < x < lp_i + 2$	2 meter prediction in channel c_i is a cosine with the dominant frequency starting at last peak.
$q_i = \{x \in c_i \mid\mid argmax(p_i) - x\| < 0.1,$ $\|lp_i - x\| > 0.1\}$	the locations x for which the distance between the non-starting peaks of p_i and x is smaller than 10 cm
$\text{score}_i = \frac{\#mm(x \in q_i \text{and } c_i(x) > 0.18)}{\#mm(x \in q_i)}$	the percentage of mm in q_i for which $c_i > 0.18$

Figure 11: A summary of the algorithm in pseudo-code. All details are explained in de subsequent paragraphs.

3.2 Thresholding plant material and background

To evaluate whether or not there is plant material under the sensor a threshold value of 0.18 is used. To determine the threshold, histograms of the measurements are made. The threshold chosen is a bit larger than the inflection point after the big peak comprising the datapoints corresponding to soil, see Figure 12. The inflection point is determined for five plots per measurement day, and is very stable.

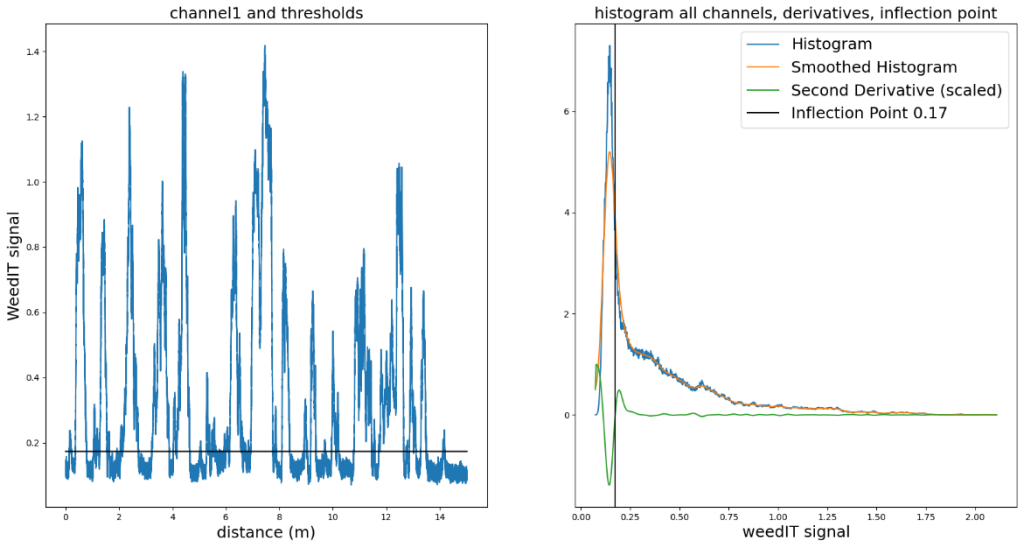


Figure 12: On the left the plot of a channel in blue and the threshold value in black. Signals higher than 0.18 are considered plant material, and the lower signals are considered background. On the right the histogram of the signal of all four channels, its smoothed version, the second derivative and the location of the inflection point of the smoothed histogram. This location was used to determine the threshold value.

3.3 Short distance frequency determination

It is desirable to get the dominant frequency in which the crop rows appear under the sensor in a short distance. Therefore, an exploration was carried out to concatenate the four channel signals after each other, taking the phase shift into account.

The four channels of one measurement are plotted below each other in Figure 13. The red lines illustrate the same crop rows. The lines are not vertical, the slope of the line illustrates the difference in phase between the channels. The distances between the peaks are plotted on the right, where the distance is quite stable, with an exception of row 11, which might be due to a connection row; a row with a different distance because seeding happened with another pass of the seeding machine.

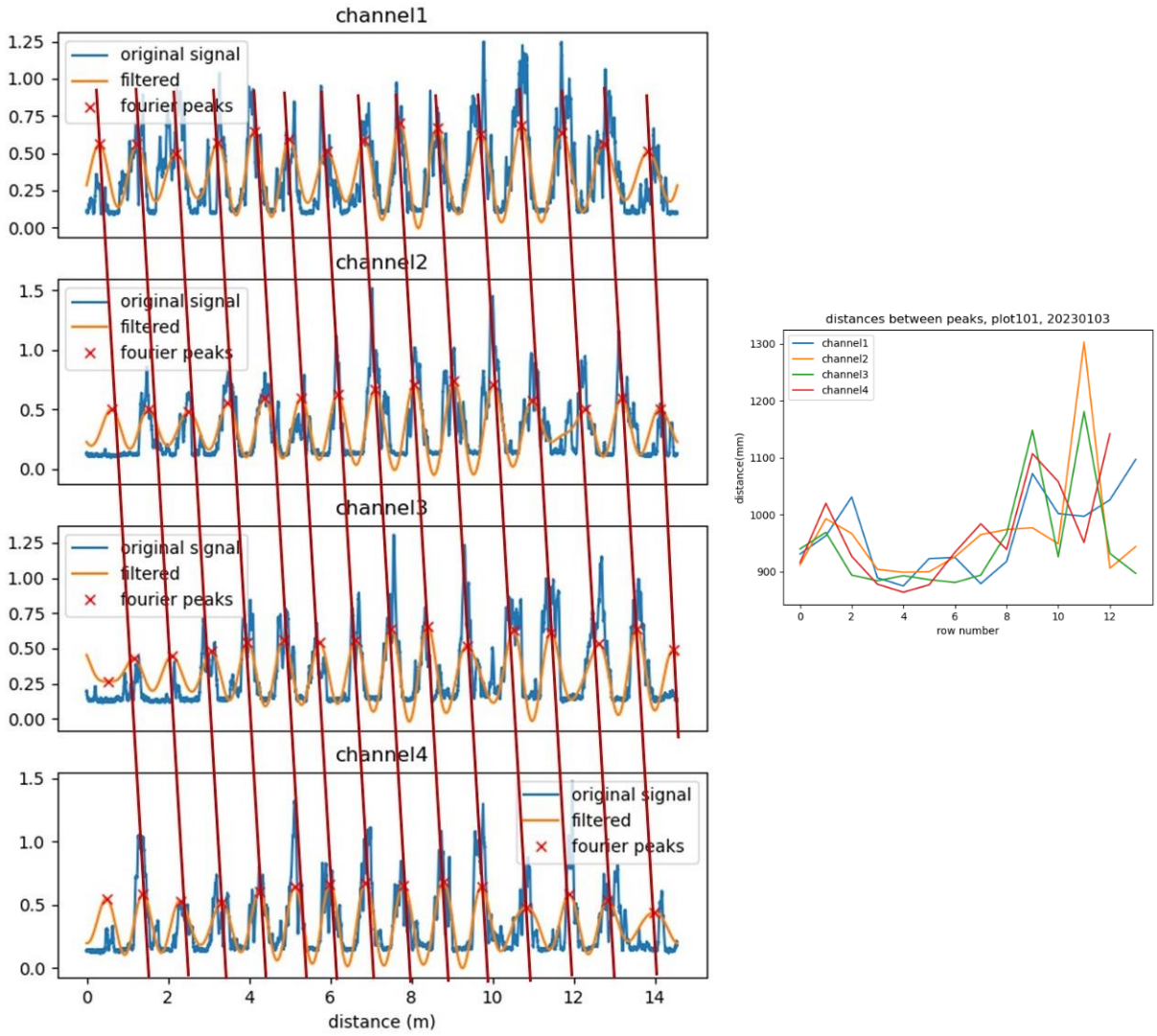


Figure 13: The four channels of one plot number of the 2023 dataset below each other, including the inverse Fourier transform of the filtered signal, and its peaks. With the dark red lines the signals coming from the same crop row are connected. Right) The distances between the peaks of the filtered signals for each of the channels.

3.3.1 Dependence on the phase shift

The different channels have a width of 25cm. Therefore, when driving non-parallel to the rows, there is a phase shift when the different crop rows appear under the sensor. This phase shift is dependent on the angle between the row and the driving direction and is given by:

$$\varphi_\alpha = \frac{0.25}{\tan(\alpha)}$$

3.3.2 Region of interest in the frequency domain

The region of interest in the frequency domain is the *interval in which the peak is expected*. Given the angle and the row distance of 0.45 meter the peak is expected to be around:

$$\text{distance in driving direction}(\alpha) = \frac{0.45}{\sin(\alpha)}$$

As region of interest the frequencies corresponding to 20 cm shorter to 20 cm longer than the distance in driving direction was used. That is:

$$(s_\alpha, e_\alpha) = \left(\frac{\sin(\alpha)}{0.45 + 0.2\sin(\alpha)}, \frac{\sin(\alpha)}{0.45 - 0.2\sin(\alpha)}\right)$$

This region is used for filtering the Fourier transform and determination of the SNR.

3.3.3 Signal to noise ratio to quantify peaks

In general there is an obvious peak in the region of interest in the frequency domain, which is related to the distance between the crop rows. The peak can be more or less pronounced. For instance in Figure 10, the peak in channel 1 is not that clear, while the peak in channel 3 is more clear. To quantify the importance of the peak, the signal-to-noise-ratio (SNR) is computed. Concretely, in our setting, this ratio is computed as follows:

$$\text{SNR}(\tilde{X}_\alpha) = \frac{\max \tilde{X}_\alpha}{\text{mean}(\tilde{X}_\alpha \setminus \max \tilde{X}_\alpha)}$$

where a backslash denotes exclusion, that is, in the denominator the maximum value is excluded when computing the mean. For evaluation of the suitability of the algorithm for determination of the driving angle 0.55 to 2.5 m^{-1} was used, as region of interest in the frequency domain. This region corresponds to the expected distances between the crop rows of 0.4 to 1.88m. When the driving direction is 90 degrees, the expected driving distance is +/- 0.45 cm. With 15 degrees it is +/- 1.9 meter. For evaluation of the predictions the (s_α, e_α) for α=30 degrees as described in the previous paragraph is used.

In Figure 14, the effect of the distance on the signal to noise ratio is illustrated; pieces of 5, 3 and 2.5 meter from the same measurement are concatenated and the height of the peak in the frequency domain is plotted. The longer the analyzed distance, the larger the height of the peaks, and hence the SNR.

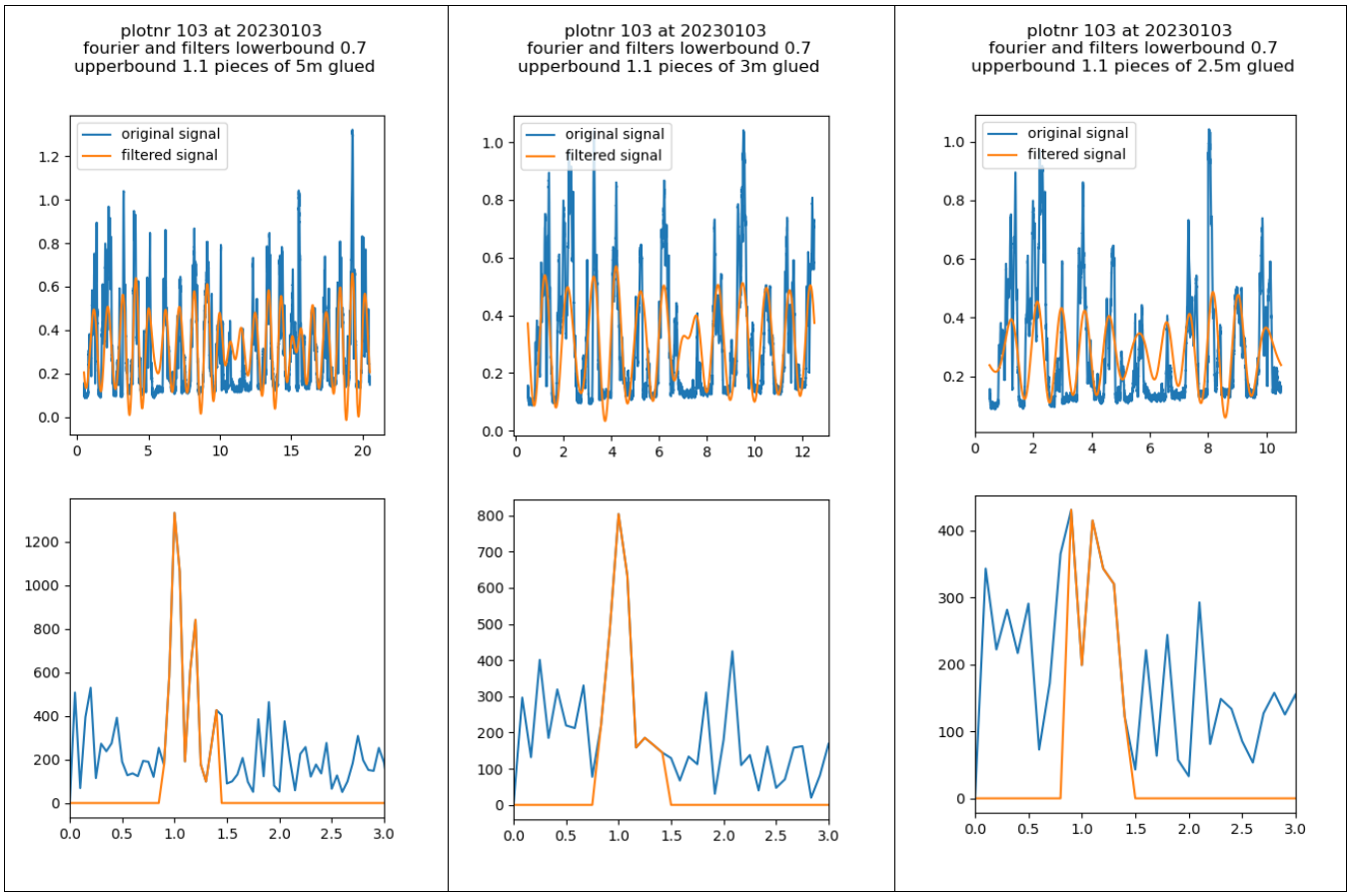


Figure 14: The four channels concatenated using 5 meter (left), 3 meter (middle) and 2.5 meter (right) of the measurements, on the x-axis the distance is measured in meter. On the bottom the respective Fourier transformed signals in blue. In orange the filtered signal, that is, the part that is outside the region of interest is set to zero. The height of the peak increases with the number of meters used, that is expressed in the SNR. The Fourier is complex-valued, hence the absolute values are mapped. The unit on the x-axis is 1/m.

3.3.4 The prediction

The locations of the crop rows are predicted two meters ahead from the last peak in the filtered inverse Fourier transform using a cosine starting at this last peak having as frequency the dominant frequency. That is the frequency corresponding to the peak in the Fourier transform within the region of interest. The locations of the peaks of this cosine in the following two meters are the predicted locations of the crop row. In Figure 15 an example of the predicting cosine on a 1cm down sampled signal.

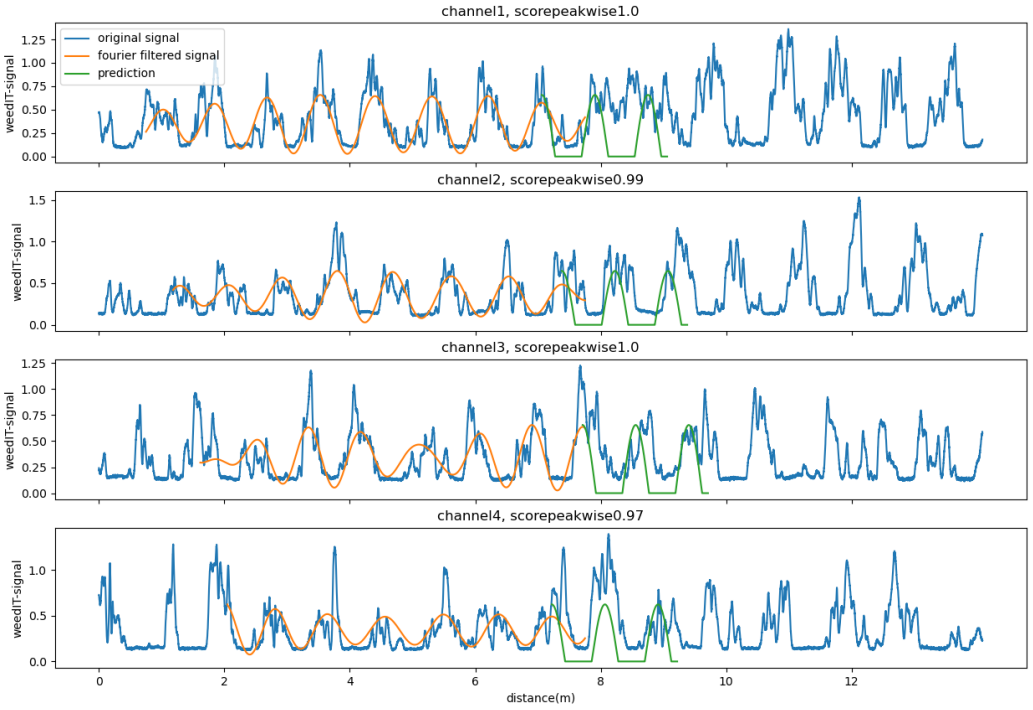
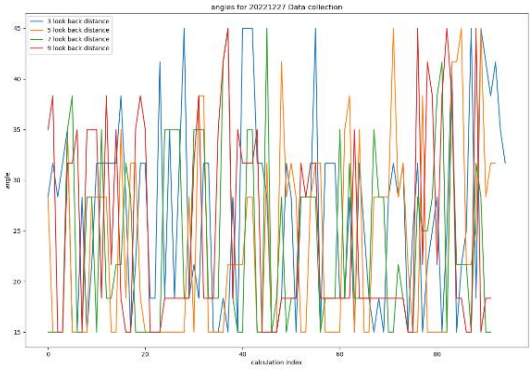


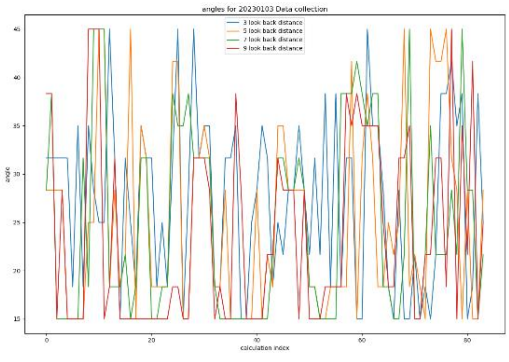
Figure 15: An example of a prediction made based on 7 meter driving on the 1cm down sampled signal. In blue the original signals, in orange the Fourier filtered signal, and in green the positive part of the predicting cosine 2 meters ahead. The scorepeakwise are the scores as discussed in the text.

3.3.5 Stability of the most dominant frequency

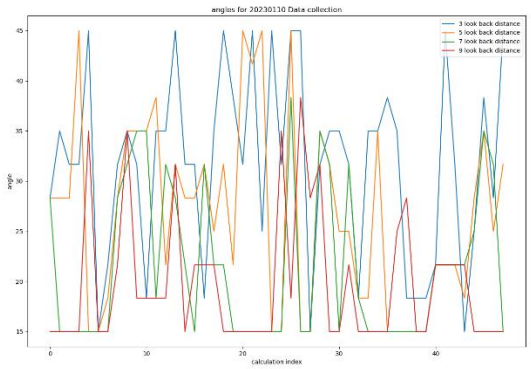
As said in the introduction of this chapter, the algorithm as presented in Figure 11 has the potential to determine the angle driven. The angle has been estimated to have a value between 15 and 45 degrees. However, the angle and dominant frequency using the algorithm is not stable. Starting the calculation from a later starting point in the same measurement the dominant frequency can change abruptly. In Figure 16 the angles found are plotted. On the x-axis, the calculation indices are shown, meaning that the angles are presented in the order that they are calculated. Calculations were done per measurement and then per starting point, so measurements done on the same plot are consecutive in the plots. Most angles found differ from the angle found in the previous calculation.



A



B



C

Figure 16: The angles found for the calculations done on the 1mm down sampled data on measurements done on A 27-12-2022, B 3-1-2023 and C 10-1-2023.

3.4 Quantitative evaluation of the predictions

3.4.1 Down sampling the data

During the measurements, the speed of the quad vehicle was monitored. Therefore, it was possible to assess the data in terms of distance. The measurements are down sampled to 1mm, 1cm, 5cm, 10cm. This was done by averaging. These down sampled signals have been used as input for the algorithm as outlined in Figure 11.

3.4.2 The score

The predictions are evaluated per channel as follows:

$$\text{score}_i = \frac{\#mm\{x \in c_i|\ |x - (\text{peak van cosine})| < 0.1 \text{ and signal}(x) > 0.18\}}{\#mm\{x \in c_i|\ |x - (\text{peak van cosine})| < 0.1\}}$$

In words, 10 cm around each peak of the predicting cosine, excluding the starting peak, the number of mm is determined of which the signal was larger than 0.18, and hence considered to be plant material. This is divided by the total number of mm that are closer than 10 cm to the peaks. All scores are calculated on the 1mm down sampled signals. In Figure 17 the calculation of the score is illustrated.

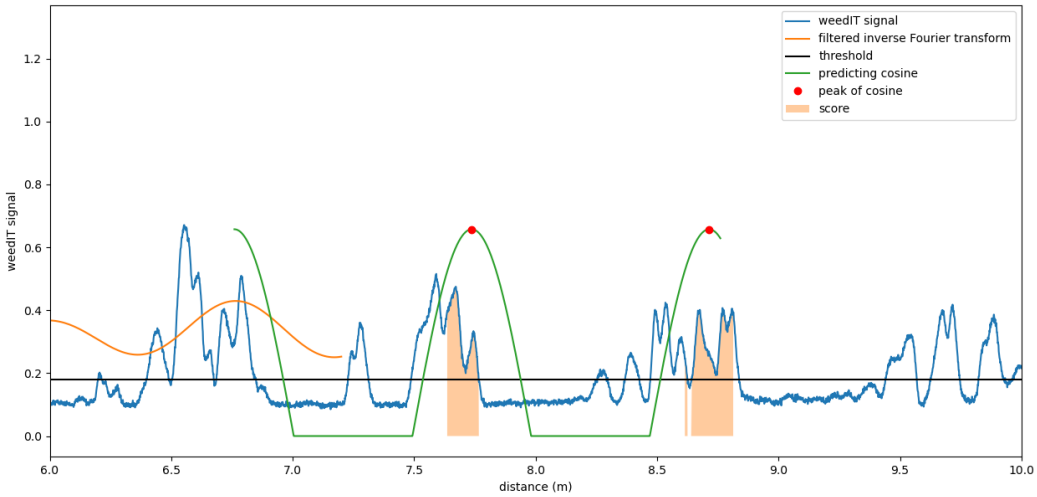


Figure 17: An illustration of how the score is defined. There are two peaks, the number of mm that the signal is larger than 0.18 around the first peak is 130, and around the second peak 183. The total number of peaks is two, so the score is (130+183)/400 = 0.78.

The algorithm described above is used to evaluate how large the distance in meters of the set-up has to measure, to get good predictions for the locations of the crop rows in the next two meters, and how far the signal can be down-sampled to still get good predictions.

3.4.3 Calculating the scores

To make a prediction, a certain distance has to be driven, such that the channels can be concatenated. Predictions and scores have been calculated based on 3, 5, 7 and 9 meter and on mm, 1cm, 5cm and 10cm down sampling rates. To generate these scores, all measurements have been used and for each measurement calculations have been carried out for starting points between 0.2 and 5.2 meters in steps of 0.55 meter. Some measurements were too short to do the predictions. Therefore, the number of calculated scores decreases with the number of meters signal used.

3.4.4 Evaluating the scores

To compare the scores over the different distances and different down sampling rates the Mann-Whitney-u test has been used. With this test it can be evaluated if the scores belong to the same distribution and if one has a larger median than the other. It is not assumed that the score is normally distributed. In Figure 18 it can be seen that the measurements are not having a Gaussian distribution.
Not all scores have been used in the evaluation. Scores from calculations for which the SNR<3 have been excluded. The threshold on SNR of 3 has been determined by visual assessment. In Figure 19 and Figure 20 two examples of plots and predictions and their SNR are shown. In the measurement plotted in Figure 19 the crop rows are not visible, so it is not expected that the algorithm can do proper predictions. This is expressed in a relatively low SNR. In the measurements plotted in Figure 20 the crop rows are visible and the SNR is 2.99.

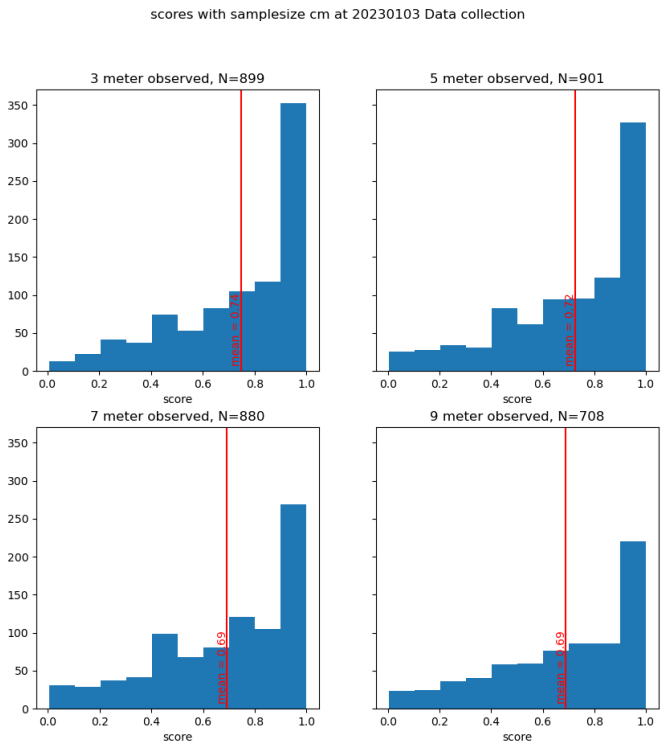


Figure 18: the scores, mean values and number of calculations done for the 3 January 2023 dataset, the four different distances for the 1 cm down sampled signals.

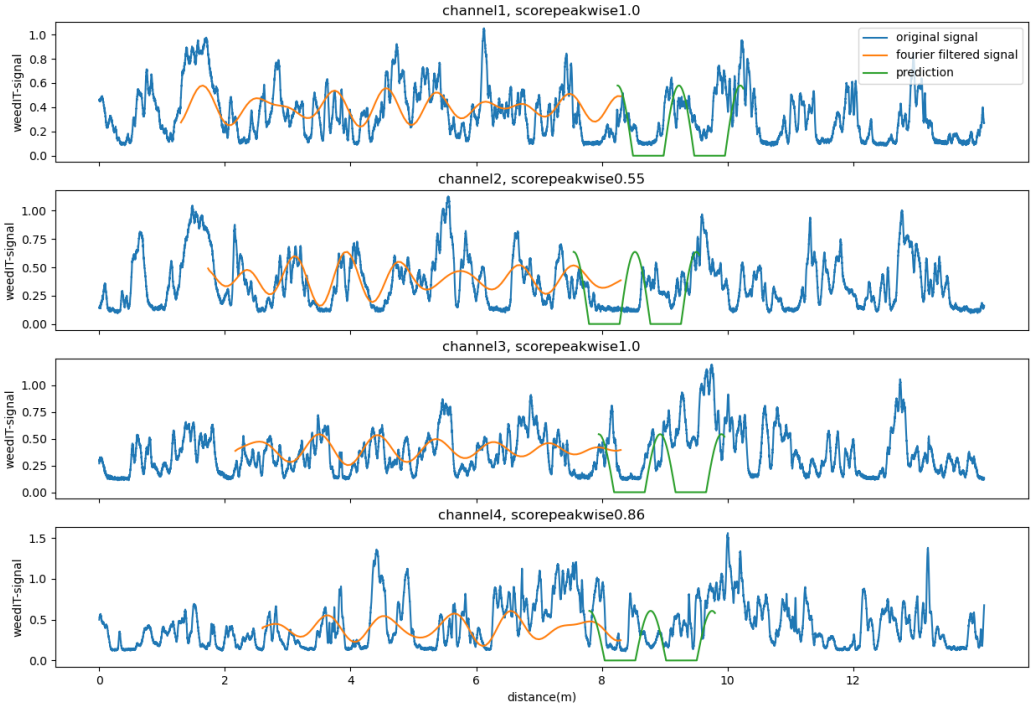


Figure 19: an example of a plot including the prediction which has been excluded from the evaluation. The SNR found in this calculation was 2.74. The scorepeakwise are the scores as discussed in the text.

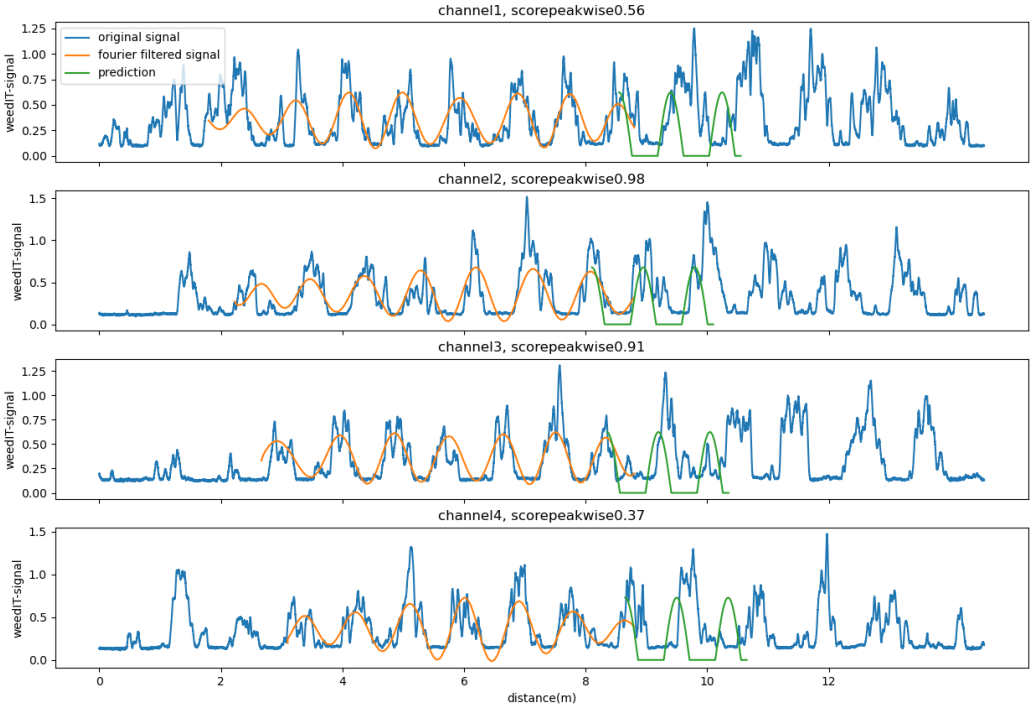


Figure 20: an example of a plot including the prediction which has just been excluded from the evaluation. The SNR found in this calculation was 2.99. The scorepeakwise are the scores as discussed in the text.

3.5 Results

3.5.1 Optimal look back distances

In Figure 21, the results of the statistical analysis of the calculations are shown. In the first column per down sample rate and day it is denoted whether a significant difference is found. That is, per group of four calculations there is no significant difference found in the distributions if the character in the first column is the same. For each down sample rate and day actually the shorter distance gives a bit higher score, however the number of calculations that pass the SNR>3 requirement is substantially smaller for this shorter distance.

significance	distance	mean	N	downsampling rate	date
a	3	0.5	262	mm	20221227
b	5	0.44	256	mm	20221227
b	7	0.45	356	mm	20221227
b	9	0.44	324	mm	20221227
a	3	0.8	218	mm	20230103
b	5	0.74	319	mm	20230103
b	7	0.71	382	mm	20230103
c	9	0.68	360	mm	20230103
a	3	0.9	96	mm	20230110
b	5	0.84	215	mm	20230110
b	7	0.83	284	mm	20230110
b	9	0.83	273	mm	20230110
a	3	0.5	263	cm	20221227
b	5	0.43	249	cm	20221227
b	7	0.45	349	cm	20221227
b	9	0.44	327	cm	20221227
a	3	0.8	220	cm	20230103
b	5	0.74	320	cm	20230103
b	7	0.71	369	cm	20230103
c	9	0.67	360	cm	20230103
a	3	0.91	84	cm	20230110
b	5	0.84	212	cm	20230110
b	7	0.83	288	cm	20230110
b	9	0.83	285	cm	20230110
a	3	0.5	254	5cm	20221227
b	5	0.43	241	5cm	20221227
b	7	0.45	357	5cm	20221227
b	9	0.44	341	5cm	20221227
a	3	0.8	209	5cm	20230103
b	5	0.74	319	5cm	20230103
b	7	0.71	382	5cm	20230103
c	9	0.67	363	5cm	20230103
a	3	0.9	88	5cm	20230110
a	5	0.85	215	5cm	20230110
a	7	0.84	273	5cm	20230110
a	9	0.83	276	5cm	20230110
a	3	0.52	242	10cm	20221227
b	5	0.45	235	10cm	20221227
b	7	0.44	346	10cm	20221227
b	9	0.44	357	10cm	20221227
a	3	0.81	191	10cm	20230103
b	5	0.75	323	10cm	20230103
c	7	0.71	375	10cm	20230103
c	9	0.67	378	10cm	20230103
a	3	0.89	88	10cm	20230110
a	5	0.86	222	10cm	20230110
a	7	0.83	273	10cm	20230110
a	9	0.82	282	10cm	20230110

Figure 21: A table with the results of the statistical analysis of the calculations. In the first column per down sample rate and day it is denoted wetter a significant difference is found. That is, per group of four calculations there is no significant difference found in the distributions if the character in the first column is the same.

3.5.2 Optimal down sampling rates

In Figure 22, the results of the statistical analysis of the calculations are shown. In the first column per distance and day it is denoted whether a significant difference is found. That is, per group of four calculations there is no significant difference found in the distributions if the letter in the first column is the same. Only the letter *a* appears in the first column, meaning that never any significant difference was found.

significance	distance	mean	N	downsampling rate	date
a	3	0.5	262	mm	20221227
a	3	0.5	263	cm	20221227
a	3	0.5	254	5cm	20221227
a	3	0.52	242	10cm	20221227
a	3	0.8	218	mm	20230103
a	3	0.8	220	cm	20230103
a	3	0.8	209	5cm	20230103
a	3	0.81	191	10cm	20230103
a	3	0.9	96	mm	20230110
a	3	0.91	84	cm	20230110
a	3	0.9	88	5cm	20230110
a	3	0.89	88	10cm	20230110
a	5	0.44	256	mm	20221227
a	5	0.43	249	cm	20221227
a	5	0.43	241	5cm	20221227
a	5	0.45	235	10cm	20221227
a	5	0.74	319	mm	20230103
a	5	0.74	320	cm	20230103
a	5	0.74	319	5cm	20230103
a	5	0.75	323	10cm	20230103
a	5	0.84	215	mm	20230110
a	5	0.84	212	cm	20230110
a	5	0.85	215	5cm	20230110
a	5	0.86	222	10cm	20230110
a	7	0.45	356	mm	20221227
a	7	0.45	349	cm	20221227
a	7	0.45	357	5cm	20221227
a	7	0.44	346	10cm	20221227
a	7	0.71	382	mm	20230103
a	7	0.71	369	cm	20230103
a	7	0.71	382	5cm	20230103
a	7	0.71	375	10cm	20230103
a	7	0.83	284	mm	20230110
a	7	0.83	288	cm	20230110
a	7	0.84	273	5cm	20230110
a	7	0.83	273	10cm	20230110
a	9	0.44	324	mm	20221227
a	9	0.44	327	cm	20221227
a	9	0.44	341	5cm	20221227
a	9	0.44	357	10cm	20221227
a	9	0.68	360	mm	20230103
a	9	0.67	360	cm	20230103
a	9	0.67	363	5cm	20230103
a	9	0.67	378	10cm	20230103
a	9	0.83	273	mm	20230110
a	9	0.83	285	cm	20230110
a	9	0.83	276	5cm	20230110
a	9	0.82	282	10cm	20230110

Figure 22: A table with the results of the statistical analysis of the calculations. In the first column per distance and day it is denoted wetter a significant difference is found. That is, per group of four calculations there is no significant difference found in the distributions if the letter in the first column is the same.

3.6 Discussion

The score is designed such that it is both independent of the strength of the signal and the dominant frequency found. The choice to look for plant material 10 cm around the found peak is somewhat arbitrary. In particular in an early growth stage the orientation of the seed planted will have a serious impact on where it appears, and therefore the plants are not always perfectly lined up.

It is possible to compare the scores calculated on the measurement data of one day. However, as the score will naturally increase with the growth of the crop, it is not useful to compare the scores between the different days.

Longer distances used as input for the algorithm do lead to higher values of the SNR, however, against the expectations, this does not lead to a higher scores. It is not clear what caused this.

The predictions made using the down sampled signals do not significantly differ per down sampling rate. This means that down sampling using averaging is a good way to down sample. It does not mean that lowering the sample rate will give enough information.

When the calculations are done in the setting where the angle could be between 15 and 45 degrees, the angles found are not stable, and this does not improve with the increasing distance. Once applied in the field, it might not be necessary to compute the SNR for angles between 15 and 45 degrees; it is not likely that the angle will change abruptly and an angle close to the previous one could be used. However, it cannot be excluded that abrupt changes in angle can appear.

3.7 Conclusion

A method has been explored to predict the location of the crop rows two meters ahead, and to quantify the predictions based on the measurements done in Brazil in December 2022 and January 2023. The method included combining the different sensor channels and a Fourier analysis approach. The algorithm as outlined in Figure 11 has been used with the starting assumption that the angles between the driving direction and the crop row to evaluate the ability to detect the angle. To evaluate the distance to be driven for doing predictions two meter ahead and different down sampling rates the algorithm has been used assuming that the angle was 30 degrees.

The performance of the different distances does not differ much, although 3 meter seems to perform slightly better. However, the number of calculations for which the SNR >3 is substantially higher for the longer distance. The optimal look back distance, therefore, is between 5 and 7 meter.

For the performance of the predictions based on the different down sampling rates, did not give any significant difference. That means that if down sampling happens using averaging the information on the peaks is still contained in the signal.

4 References

Janne Kool, E. d. (2023). *Green on Green weed detection: Finding weeds in a soybean crop in Brazilian fields with the Rometron WEED-IT sensor : intermediary report.* Wageningen: Wageningen Plant Research. doi:10.18174/649472

Corresponding address for this report:
P.O. Box 16
6700 AA Wageningen
The Netherlands
T +31 (0)317 48 07 00
wur.eu/plant-research