Green on Green weed detection

Finding weeds in a soybean crop in Brazilian fields with the Rometron WEED-IT sensor. Intermediary report.

Janne Kool, Eva de Jonge, Ard Nieuwenhuizen, Hendrik Braam

1 Wageningen University & Research

2 Rometron

This study was carried out by the Wageningen Research Foundation (WR) business unit Agrosystems Research.

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

Wageningen, December 2023

Report WPR-10.18174/649472

WAGENINGEN
UNIVERSITY & RESEARCH

Janne Kool, Ard Nieuwenhuizen, Eva de Jonge, 2023. *Green on Green weed detection; Finding weeds in a soybean crop in Brazilian fields.* Wageningen Research, Report WPR-10.18174/649472.

This report can be downloaded for free at https://doi.org/10.18174/649472,

Summary Nem omnis enis gerces et quos caessed cressul hiliusat pos, poritanduc tus camdientriam lostes vendac omnihilia nos tescii trum hostorio vere fica ingulicur, et adducon Ita num no. Serehendam iu consum nonsus cotam identem hos condacto crehebe factebatiam medeliquit oca nos hae atraedeffrem es facte ad si tem publis occhuit. Locturnit intim isse tractuis Catquon sulvitat. Perunce ssentif erfecul tertuius, mo videffrentem huius curi peret ad C. Gitam rem, compotatus am se nox nonem supio unum pora addum, consus consuli ia talernimo comnes omner inte pra? qua imum imilicut atusquastra desilin guleribut Catuis ces con.

Keywords: goose, deterrence, quantity regulation

Chamber of Commerce no. 09098104 at Arnhem
VAT NL no. 8065.11.618.B01

Report WPR-10.18174/649472

Contents

Contents 3

Summary 6

1 Introduction 8

2 Experimental setup 9

3 Expected resolution of the WEED-IT sensor 11

4 Weed detection algorithms 14

4.1 The general idea behind the algorithms 14
4.2 Algorithm 2: Finding the locations of the crop rows 14
4.3 Algorithm 1: Distinguish the difference between plant and mulch 19

5 Finding the weeds 22

6 Discussion 24

Summary

The objective of WP1 this season was to further investigate whether weeds can be identified in a soybean field using the Rometron WEED-IT sensor and an algorithm. The WEED-IT sensor can distinguish living from non-living plant material but cannot differentiate between weed- and crop plants. Therefore the study relies on plant location: plants outside the rows are deemed weeds, and those within are considered crops. The experiment, conducted in the 2022-2023 growing season, aims to assess the WEED-IT sensor's effectiveness in detecting weeds between crop rows.

In Brazil, a soybean trial was planted and analysed for the study. Measurements took place at a 30-degree angle to the crop rows, as this is common practice in Brazil. A quad with a metal frame carried the WEED-IT sensor and a GoPro camera alongside recording the weeds and crop rows for reference. Weed locations were noted in the field using a measurement tape as reference.

Two algorithms were developed: one, using a Gaussian Mixture Model, detects plant material (as the sensor does), and the second (Fourier analysis) identifies the location of the crop rows. When combined, these algorithms create a decision system capable of determining whether to initiate spraying based on the presence of plant material and the location of crop rows.

For the Fourier approach to be effective, timing is crucial—it requires measuring when the crop is sufficiently large to be detected while maintaining positive distinctiveness. If the crop is too small, the Fourier approach is not effective.
Also, the Fourier analysis relies on a key assumption: the rows are regularly planted and emerge uniformly. Significant irregularities in row spacing can result in errors in the detection process.

The Fourier model enables the identification of row locations. This offers the possibility to spray only when the spray nozzle is between the rows, reducing chemical use even if weed detection is imprecise.

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 will have 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. This means that weeds in the row are not recognized and therefore this study focuses on weeds between the rows.

In the growing season of 2022-2023, an experiment has been conducted to evaluate whether it is possible to detect these weeds between the crop rows using the WEED-IT sensor.

2 Experimental setup

In Brazil, 36 plots with a length of 14 meter were planted on a soybean field. The corn was planted on the 6th of December 2022, later than usual due to weather conditions. The soybean was planted with a 50 centimeter row distance.

It is common practice in Brazil that during spraying the driving angle with the crop rows is approximately 30 degrees. So the measurements were done in a 30-degree angle from the plot rows as well, like a conventional spraying operation in the field. The metal frame on the quad on the picture is 390 cm wide, and the WEED-IT sensor, colored green, is hanging at 120 cm from the center. The sensor is placed 110 cm above the ground (see Figure 1). The sensor data of the WEED-IT signal comprises 4 channels, next to each other. The width of each sensor channel when mounted on this height corresponds to 25cm at ground level. Next to the WEED-IT sensor a GoPro camera was mounted, such that the measured ground locations were also visibly recorded in a digital video. A measurement tape was laid on the plots, and the location of the weeds was recorded in the field using the measurement tape as a reference.

Often, only one location was recorded for the weeds, but sometimes an interval was indicated. A second check on the weeds locations was done using the GoPro videos. Based on the videos, the channel of the WEED-IT sensor in which the weed was expected was recorded. On these videos more weeds were found and indicated, also some weeds recorded in the field were not found on the videos.

Figure 1: The quad, with a metal frame and the WEED-IT sensor (green) hanging on the right.

Apart from the WEED-IT signal also a wheel speed sensor has constantly monitored the speed of the setup. Using this speed, the raw WEED-IT data was transformed to data on millimeter intervals, by taking the average of the raw data obtained per mm.

After germination, data was recorded on the fields 5 times. The dates and the growing stages are presented in Table 1. In Figure 2 to 6, the growing stages are illustrated, using the GoPro footage.

Table 1 *The dates of the measurements and the growing stages of the soybean.*

Date	**Growing stage of the soybean**
21-12-2023	Just germinated
27-12-2022	Just germinated
3-1-2023	5 leaves, usual spraying
10-1-2023	Late spraying
14-1-2023	Full canopy

Figure 2: growth stage 21-12-2022 field 105

Figure 3: Growth stage at 27-12-2022 field 101

Figure 4: Growth stage at 3-1-2023 field 201

Figure 5: Growth stage at 10-1-2023 filed 102

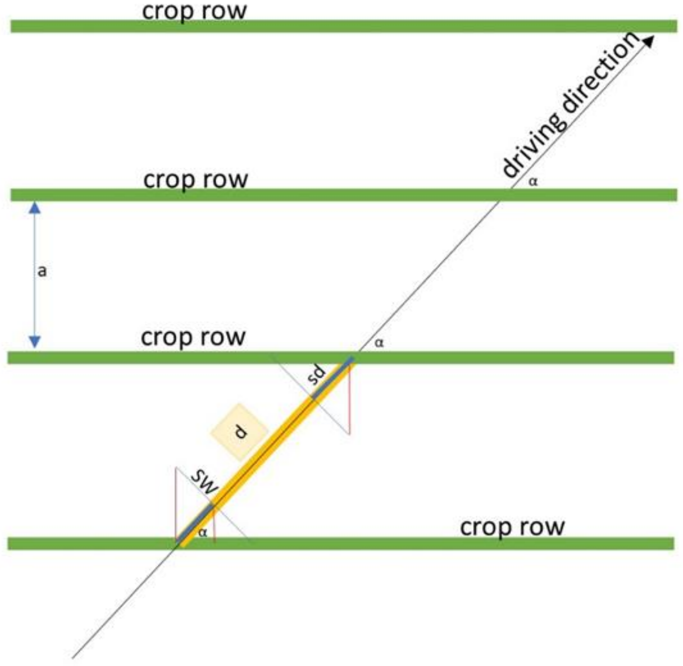
Figure 6: Growth stage at 14-1-2023 field 203

3 Expected resolution of the WEED-IT sensor

The WEED-IT sensor, and the datalogger that was connected to the sensor, both measure with a frequency of 14 kHz; a very high resolution in the driving direction. However, the sensor also has a width: there are four channels. In these channels the measurements are taken next to each other. With the sensor hanging at 110cm, the width of one channel is 25 cm. Because of this width, the boundaries of the channel can be above the crop row when the center of the channel is not. In this case, plant material is detected anyway, as the signal is the average of all occurrences of crop material in the width of the field of view of the channel.

We, therefore, define the *distinctiveness, D,* of the setup to be the part of soil measured in a channel in which nothing of the channel is still measuring crop. The distance between the point that the centre of the sensor is not above the crop and the tip of the sensor is not above the crop anymore, is called the separation distance, and is denoted with *sd*. The distance that the centre of the sensor is above the soil is denoted with *d.*

Note that *D* and *sd* depend on the angle α between the crop row, the direction of driving, the distance *a* between the foliage of the crop rows (which is by definition smaller than the row distance) and the width of the sensor, *SW*. All parameters needed for the expression are illustrated in



. The distinctiveness is given by:

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

Note that when the driving direction is parallel to the crop rows and hence α =0, both *d* and *sd* are not defined. Moreover, if 2*sd>d*, *D* attains negative values, meaning that there is always the canopy under the sensor. To get a feeling on how the distinctiveness is dependent on the parameters, see Figure 8 and Figure 9. In Figure 9 the *a*=20 cm and *D* is negative. With the growth of the crop *a* will shrink, being 0 at the closure of the canopy.

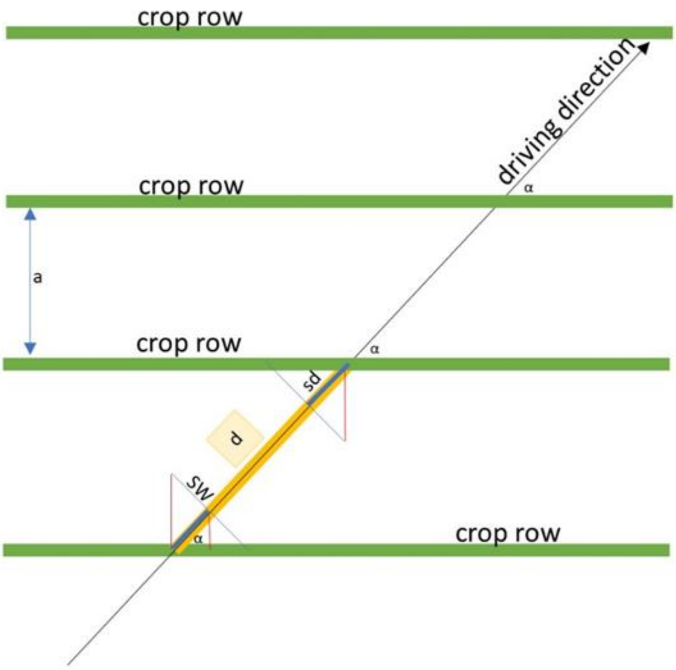


*Figure 7: A schematic sketch of the crop rows and the sensor movement, in which the main parameters for the calculation of the distinctiveness D are illustrated. a) the distance between the foliage of the crop rows, α is the angle between the crop rows and the driving direction, d) is the distance in the driving direction between the foliage of the crop rows, so d = a/sin(α), SW) is the sensor width, in this experiment 25cm, sd) is the separation distance, that is the distance the center of the sensor had driven until the whole sensor is not above the foliage of the crop anymore. sd = 0.5*SW* cos(α)/sin(α).*

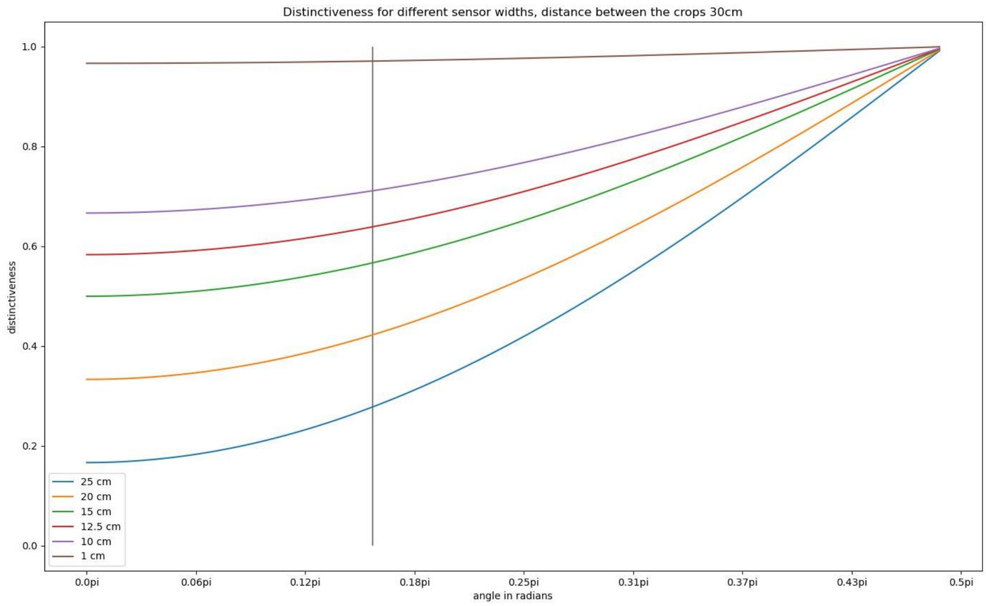


Figure 8: The distinctiveness plotted against the angle between the driving direction and the crop row when the distance between the foliage of the crop rows is 30cm. The vertical line is the angle which was used during the experiments. A distinctiveness of 0.5 means that half of the distance that the 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 rest of this distance (time in case of equal speed).

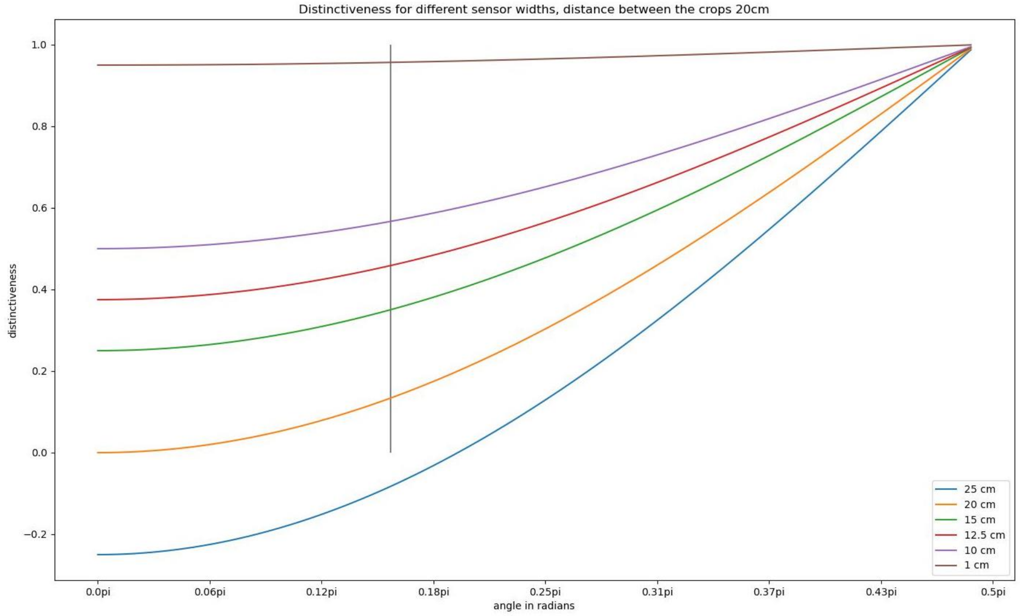


Figure 9: The distinctiveness plotted against the angle between the driving direction and the crop row when the distance between the foliage of the crop rows is 20cm. The vertical line is the angle which was used during the experiments. A negative distinctiveness means that the sensor is always measuring crop and will not measure any bare soil.

4 Weed detection algorithms

4.1 The general idea behind the algorithms

To find weeds it is necessary to find plant material on locations were you do not expect plant material. The first step is to detect living plant material and the second step is to know the locations where plant material is expected. So two algorithms are developed, the first one to detect plant material, something the WEED-IT sensor is already specialized in. The second algorithm detects the location of the crop rows. Together they could function as a decision system to decide whether to spray or not (Figure 10).

In the next paragraphs we explain both algorithms, starting with the second algorithm as knowing the location of the crop rows is the first step, before identifying plants as weeds.

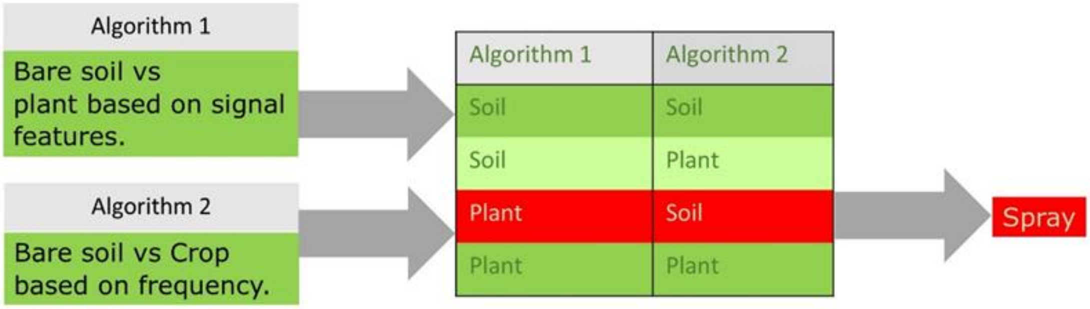


Figure 10: An illustration of the algorithm rationale. In algorithm one the distinction between plant and non-plant material is made and in algorithm two the location of the crop rows is determined.

4.2 Algorithm 2: Finding the locations of the crop rows

When driving in an angle of 30 degrees with respect to the crop rows, the crop rows will pass by on a regular frequency. To find the exact locations of the crop rows in the signal, a Fourier transform analysis is used. Fourier analysis is a well developed and established method to analyze data in which frequencies play a major role.

Let us briefly explain how Fourier analysis works. In general, the following steps are performed in the analysis: the signal is multiplied with $\sin(2\pi\omega x + p)$, the basic periodic function, with frequency ω and phase p, and then the area under the curve, or integral, is computed. An illustration of these steps can be found in Figure 11. The greater the integral the more prominent that frequency and phase combination is part of the signal.

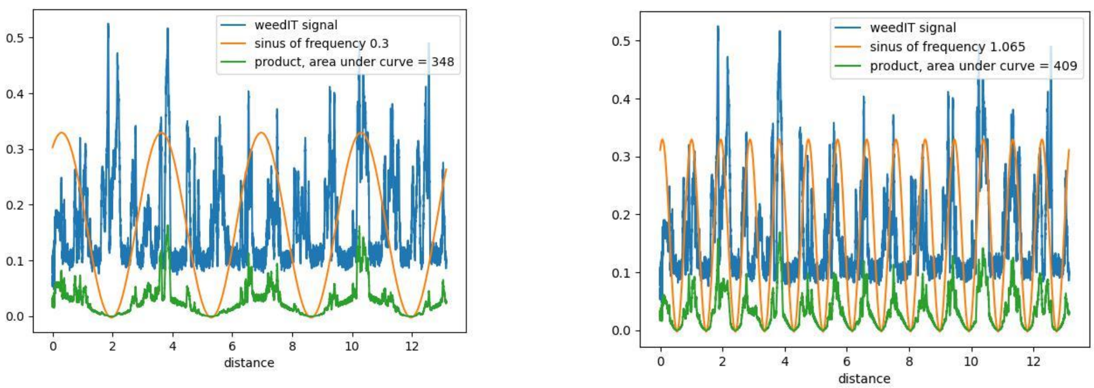


Figure 11: The WEED-IT signal multiplied with two different sine functions. On the left, the signal does not have a frequency that is very prevalent in the signal, while on the right the frequency is very prevalent in the signal. In green the product of the orange and the blue lines. The y-axes here represent the WEED-IT signal. The area under the green curve on the right is larger than on the left.

If you take the normalized values of these areas for each ω an p, you get what is called the Fourier transform of the signal. The efficient way to compute the product of $sin(2\pi\omega x + p)$ with the signal for each frequency ω and phase p, is to multiply with $exp(i\ \omega\ x) = cos(\omega) + i\ sin(\omega)$. So the Fourier transform of the WEED-IT signal *s,* is a complex-valued series $\hat{s}$, meaning it attains values in the complex numbers.

We work with the discrete version of the Fourier transform and then it can be written as a series expansion. A nice property of the Fourier transform is that if you take the Fourier transform of the Fourier transform, the original signal is returned, that means that the Fourier transform is the inverse of itself. This allows for filtering out the most interesting part of the signal, by considering only those frequencies, that is that part of the series expansion, in which the crop rows are likely to pass by, and setting the coefficients of all the other frequencies to zero. The high frequencies correspond to noise, while the low frequencies indicate, slowly changing effects in the signal due to changes in environment and growth conditions of the crop.

Now, the filtering is done by taking the Fourier transform of this altered series $\hat{s}$. When the Fourier transform is filtered, the inverse Fourier transform will again be complex-valued. Because we are interested in filtering the original real-valued function, the filtered signal is the real part of that function. The filtering procedure is illustrated in Figure 12.

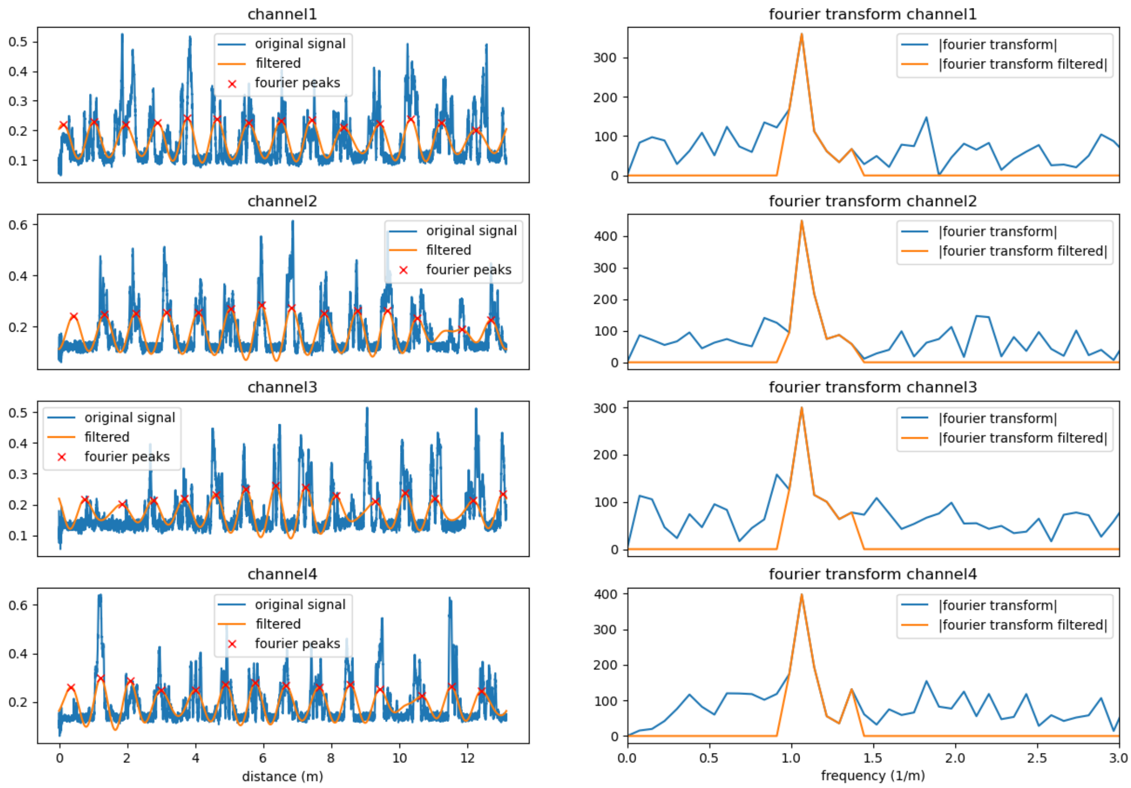


Figure 12: Plot 101 measured on 27 December 2022. On the left for each of the four channels the WEED-IT signal in blue, and on the right in blue the absolute value of the Fourier transform. In orange on the right the filtered absolute value of the Fourier transform, where the frequencies that smaller than 0.9 and larger, then 1.4 are set to 0. On the left in orange the Real part of the inverse Fourier transform of the filtered Fourier transform, with the peaks marked red. The rows in Brazil were planted 45 cm apart from each other, the driving angle was 30 degree, hence the distance a was 90 cm and the peak in the frequency is expected to be around 1.1.

Using this Fourier analysis, the locations of the crop rows in the signal can be found. They correspond to the peaks of the filtered functions, where the found peaks were forced to be more than 60 cm apart.
The planting distance was 45 cm, and hence most crop rows are between 41 and 49 cm apart, corresponding. The distance between the rows in the driving direction hence was between 82 and 98 cm. The expected peak in the frequency was around 1/0.9 = 1.1. We filtered the signal to observe only the frequency between 0.9 and 1.4, corresponding to a row distance in the driving direction 0.7 and 1.1 meter.

To validate this measurement, we counted the number of peaks passing during the measurements using the GoPro videos, in which we could see the red light produced by the WEED-IT sensor. This is not a very precise method, as it is sometimes unclear if the first and last row on the video is indeed in the measurement. An illustration of this can be found in Figure 13. Therefore, a difference of one row is considered allowable, this is used to compute the accuracies.

Figure 13: At the end of a measurement it is not always clear if a row is measured or not, especially not if a row is failing at the end or beginning of the measurement. This led to an uncertainty in the counting of the peaks.

Figure 14: The rows are not always regularly apart from each other. Sometimes they approach each other very closely. This is plot 203.

The accuracy of the Fourier method increases over time. On December 27th, for 62% of the measurements the counts of the Fourier peaks and the counts of the go-pro peaks did not differ more than 1. On January 1st, 90% of the measurements had an allowable difference, and on January 10th, 100% had an allowable difference.

This is due to the fact that during the first measurement not all soybean had germinated, and sometimes rows were partly missing. The germination increased over time and was full on 10-01-2023. The exact measurements are in **Error! Reference source not found.**, 3 and 4. The number of rows per plot on the v ideo's varied between 14 and 19.

Table 2: The measurements of the difference in number of peaks found between the Fourier method and hand counting's using the GoPro videos, on 27-12-2022. 62% of the counts are correct.

Plot number	Channel	Fourier	GoPro	Difference
102	channel1	7	8	-1
102	channel2	7	9	-2
102	channel3	7	8	-1
102	channel4	7	8	-1
103	channel1	15	18	-3
103	channel2	15	18	-3
103	channel3	15	17	-2
103	channel4	15	18	-3
105	channel1	17	15	2
105	channel2	17	14	3
105	channel3	18	14	4
105	channel4	18	15	3
106	channel1	15	16	-1
106	channel2	15	15	0
106	channel3	15	15	0
106	channel4	15	15	0
201	channel1	15	16	-1
201	channel2	15	17	-2
201	channel3	15	16	-1
201	channel4	15	16	-1
202	channel1	16	15	1
202	channel2	16	15	1

202	channel3	16	15	1
202	channel4	16	15	1
203	channel1	14	17	-3
203	channel2	14	18	-4
203	channel3	15	17	-2
203	channel4	14	17	-3
204	channel1	16	17	-1
204	channel2	16	18	-2
204	channel3	16	17	-1
204	channel4	16	17	-1
205	channel1	15	15	0
205	channel2	16	15	1
205	channel3	15	15	0
205	channel4	16	16	0
806	channel1	16	16	0
806	channel2	16	16	0
806	channel3	17	16	1
806	channel4	15	16	-1

Table 3: The measurements of the difference in number of peaks found between the Fourier method and hand counting's using the GoPro videos, on 03-01-2023, 90% of the counts are correct.

Plot number	Channel	Fourier	GoPro	Difference
101	channel1	15	16	-1
101	channel2	15	15	0
101	channel3	15	16	-1
101	channel4	15	15	0
102	channel1	15	15	0
102	channel2	15	15	0
102	channel3	15	15	0
102	channel4	15	15	0
103	channel1	16	16	0
103	channel2	15	16	-1
103	channel3	16	17	-1
103	channel4	16	16	0
104	channel1	16	16	0
104	channel2	17	17	0
104	channel3	17	16	1
104	channel4	16	16	0
106	channel1	14	16	-2
106	channel2	14	15	-1
106	channel3	15	15	0
106	channel4	14	16	-2

Table 4: The measurements of the difference in number of peaks found between the Fourier method and hand counting's using the GoPro videos, on 10-01-2023 100% of the counts are correct.

Plot number	Channel	Fourier	GoPro	Difference
101	channel1	15	15	0

101	channel2	15	15	0
101	channel3	15	15	0
101	channel4	15	15	0
102	channel1	15	15	0
102	channel2	15	15	0
102	channel3	15	14	1
102	channel4	15	15	0
103	channel1	15	15	0
103	channel2	15	14	1
103	channel3	16	15	1
103	channel4	15	14	1
104	channel1	16	16	0
104	channel2	16	17	-1
104	channel3	17	16	1
104	channel4	16	16	0
106	channel1	15	16	-1
106	channel2	15	16	-1
106	channel3	15	15	0
106	channel4	15	16	-1

To find were the crop is to be expected, not only the location of the crop row has to be determined, but also the row width. In consult with Rometron it has been decided to park this issue for now, as the row width could theoretically be measured by hand.

The Nyquist-Shannon sampling theorem states that for the determination of the frequencies at least two times the main frequency should be measured, meaning that to get the Fourier frequencies right in this case at least 90cm should be measured. So the practical distance of about 14 meter used in this experiment is enough to get the Fourier model. Note that it is very useful to get the planting distance beforehand.

By not setting the Fourier filter too narrow, as is illustrated in Figure 12, small irregularities are captured in this procedure of finding the peaks.

4.3 Algorithm 1: Distinguish the difference between plant and mulch

To make the distinction between plant material and mulch, we made the following assumptions:

1. Both the signals for plant material and for mulch material are normally distributed.
2. When the signal is larger than the mean signal of plant material then the signal comes from plant material.

Given these assumptions, we developed an algorithm based on a Gaussian mixture model. In a Gaussian mixture model a number of normal distributions is fitted to the data, in this case two. From the data collected in Wageningen in 2022, a set of measurements with mostly soil and a dataset with mostly crop material was obtained. Figure 15 illustrates that the assumptions are reasonable. We took the histograms from a measurement day three as an example to explain the method. The other histograms look very similar.

In Figure 15B, the crop histogram has two peaks, one falling together with the peak of the soil data and one for what is actually plant material. This contamination of the data is probably due to the sampling method of the crop and plant material, as the crop was not completely covering the soil. Due to the high frequency of the measurements, we assume that some of the datapoints in the plant dataset did not cover plant material.

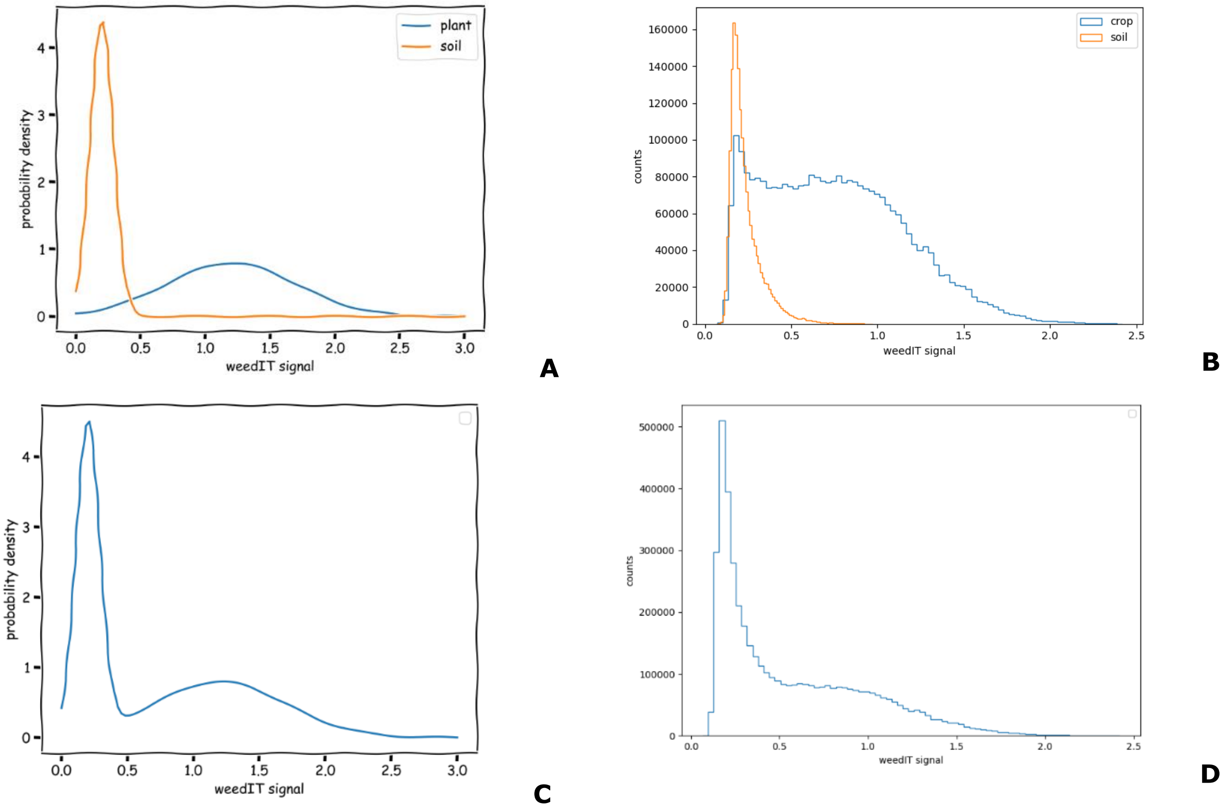


Figure 15: In A: an illustration of two normal distributions with a mean and variance as found in the data. In C: the addition of the two distributions as made in A. In B: the histograms of the crop and soil data as collected on the third measurement day of the measurements done in Wageningen in 2022. In D: the addition of the histograms found in B.

A normal distribution is determined by a mean value and the variance. When fitting the model, mean values and the variances are found, moreover the percentage of samples coming from each of the distributions is found. Let π be the percentage of points coming from the soil and (*1-π*) the percentage of points coming from plant material. Let *s* be the probability distribution function for the distribution of the soil data and let *q* be the probability distribution function for the plant signal. Then, given n data points,

$$X = x_1, \ldots, x_n$$

it is possible to compute that these points come from a plant or from the soil, using the following formulas:

$$p(X = soil) = \frac{\pi \prod_{i=1}^{n} s(x_i)}{\pi \prod_{i=1}^{n} s(x_i) + (1-\pi) \prod_{i=1}^{n} q(x_i)}$$

$$p(X = plant) = \frac{(1-\pi) \prod_{i=1}^{n} q(x_i)}{\pi \prod_{i=1}^{n} s(x_i) + (1-\pi) \prod_{i=1}^{n} q(x_i)}$$

To really compute these probabilities one quickly runs into problems when the number of sample points, *n*, increases, because the products of probabilities quickly run to zero. However, it is not necessary to compute the probabilities, as we only need to determine which of the two is larger. The denominators are the same, hence it is only necessary to determine which of the two numerators is larger.

To do so, the logarithm of numerators is computed. The logarithm is a strictly increasing function, so the inequality we are after stays unaltered. The logarithm transforms the multiplications to additions keeping all computations doable for a machine.

The algorithm works as follows:

1. A Gaussian mixture model is fitted to the data of one plot, using Python *Scikit-learn* (Sklearn) version 1.0.2. The means are initiated as 0.1 for the soil and the mean value of channel1 for the plant material. This can be done in the software.
    a. There is some randomness in the procedure so the answers are not always the same, and there is a risk that one percentage becomes zero. So iterate until a solution is found in which the order of the means is not changed compared to the initialization, that is smallest mean is the mean corresponding to the soil, and both means are positive. Sometimes Sklearn finds solutions in which the means with the lowest index corresponds to the plant material, while the initiation is the other order. This gives nonsense outcomes.
    b. The mean values of channel1 increased a lot over the days, and therefore it is important that initialization happens using the data, initialization with a too large number might lead to non-valid solution, like the one discussed above.
2. Set the number of mm needed to determine if there is plant or soil data, this sets the number of sample points. This means that you pick a number of points on which bases the algorithm makes the decisions.
3. Compute the logarithms of the soil numerators for the previous number of point, altering the distribution functions such that assumption 2 is satisfied, that is, large signals come with probability 1 from plant material. If this step is left out than very high signals will be considered as coming from the soil. In Figure 15A this is illustrated.
4. Determine which of the probabilities is largest.

The model depends on the distance that is used to determine if there is plant material or not. Figure 16 shows predictions for one plot and several channels. One can observe that if there are less millimeters to look back on, the model operates more precisely, but there is also more noise. For example, on the top plot in figure 16 the model shifts more frequently between plant material and soil, than in the lower figures. This means smaller plants and soil areas can be observed (higher precision) but high frequency in changes can be considered noise.

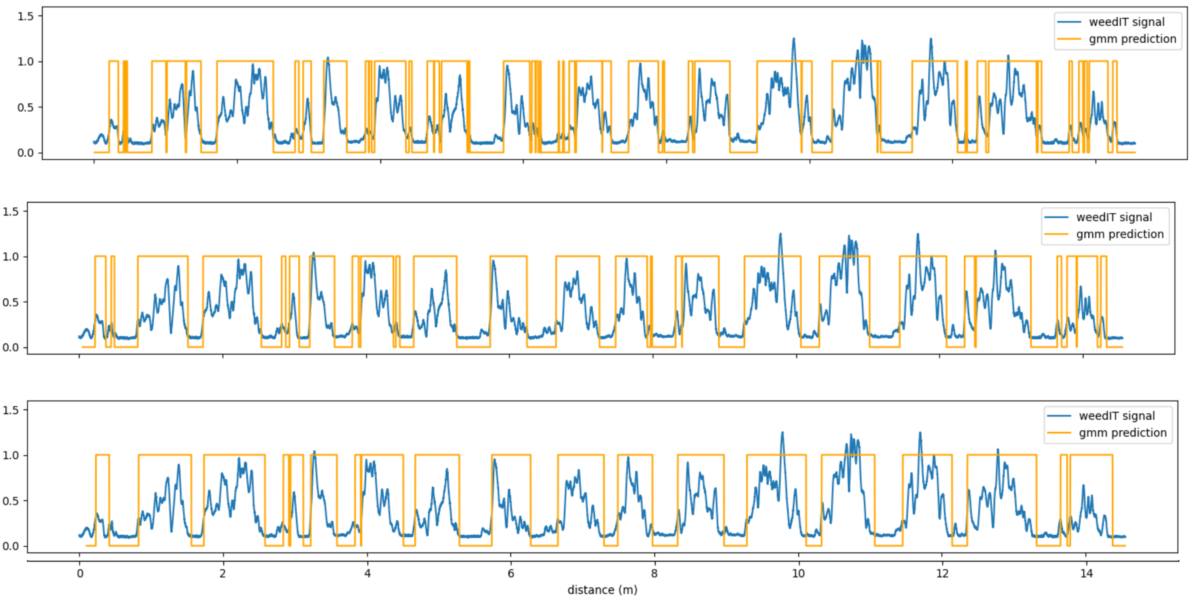


Figure 16: The Gaussian Mixture Model (GMM) in action on the first channel of plot 101 on 1-3-2023. The Y-axis is the voltage of the WEED-IT sensor. The model is based on if the probability is larger to be mulch or plant based on the signal received on the last chosen distance signalled. The chosen distance from top to bottom were 0.01, 0.05, 0.1 meter respectively. The model returns 1 if the signal is more likely to come from plants and 0 if it is more likely to come from mulch.

It has been validated if the parameters (mean and variance) varied a lot between the different measurement. To do so a Gaussian mixture model was trained and the variance and the means were collected. Of those variances and means, again the variances and means have been computed. They are presented in Table 5. All of those variances are smaller than 0.001, except for the measurements on 21-12-2022 and 27-12-2022 when the crop was still very low, and on the last measurement on 14 January when the canopy was closing.

The model itself could not and has not been validated in the sense that we know exactly where there is plant material and where there isn't.

Table 5: In this table the means and the variances over each data collection day of the means and the variances found for the GMM fitted on the plots. The folder corresponds to the name of the data folder where the data has been collected during the measurements. Conclusions about this values are in the text above.

folder	mu_cov_soil	mu_cov_plant	cov_cov_soil	cov_cov_plant	mu_mu_soil	mu_mu_plant	cov_mu_soil	cov_mu_plant
20221221 Data collection	0.00048	0.00601	0.00031	0.00126	0.09869	0.22183	0.06464	0.03744
20221227 Data collection	0.00038	0.00987	0.00017	0.00165	0.10896	0.21523	0.04970	0.02278
20230103 Data collection	0.00246	0.05966	3.84E-06	1.36E-05	0.16227	0.52091	3.84E-05	0.00019
20230110 Data collection	0.00061	0.14426	6.43E-06	8.44E-05	0.14034	0.67751	0.00010	0.00046
20230114 Data collection	0.10599	0.68942	0.00050	0.0015	0.49864	2.25957	0.00106	0.00165

5 Finding the weeds

The two models and the annotations of the weeds have been combined in one plot, to observe the performance. The performance is also measured by counting weeds sprayed that are not in the row. The results are in Table 6. The weeds were counted by hand from the go-pro footage. Two different people have been doing the countings leading to some variety in the counts, therefore it is difficult to draw conclusions from this table.

Table 6: The raw data and accuracies of the overall algorithm. Note there has not been optimized for the GMM distance and row width. Here to spray means that the algoritm asks to spray and acc is defined as to spray/(number of weeds – number of weeds in row).

date	number of weeds	weeds in row	to spray	missed weeds	acc	gmm dist (m)	row width (m)	number of plots
20221227	216	38	34	144	0.19	0.05	0.15	24
20230103	42	17	10	15	0.4	0.05	0.3	23
20230110	131	53	41	37	0.52	0.05	0.45	12

No optimization has been performed for the GMM distance, and the row widths have been guessed, optimization and especially more width between the rows would improve the performance. From Figure 17 and Figure 18 it can be seen that often we sprayed while it was not necessary. The amount of unnecessary spraying decisions has not been quantified in this dataset.

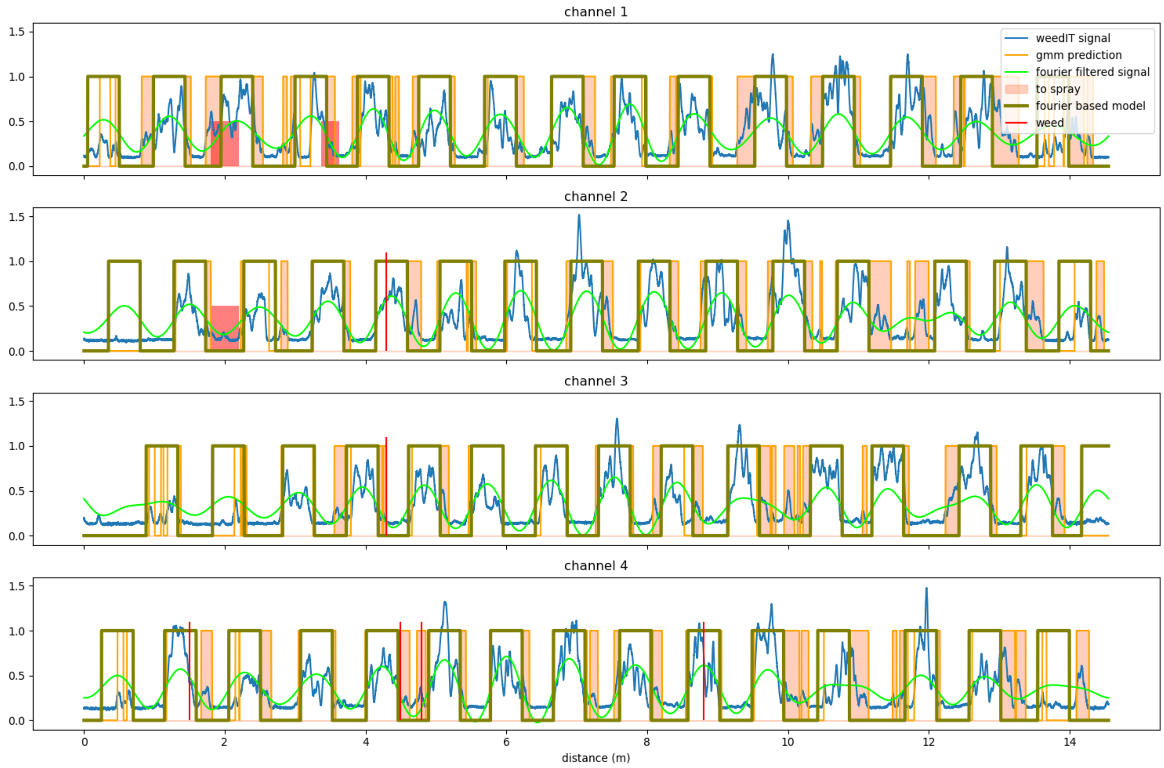


Figure 17: All elements of the algorithm in one figure. Plot 101 on 3-1-2023 row width was set to be 45 cm and the GMM was based on 5 m.

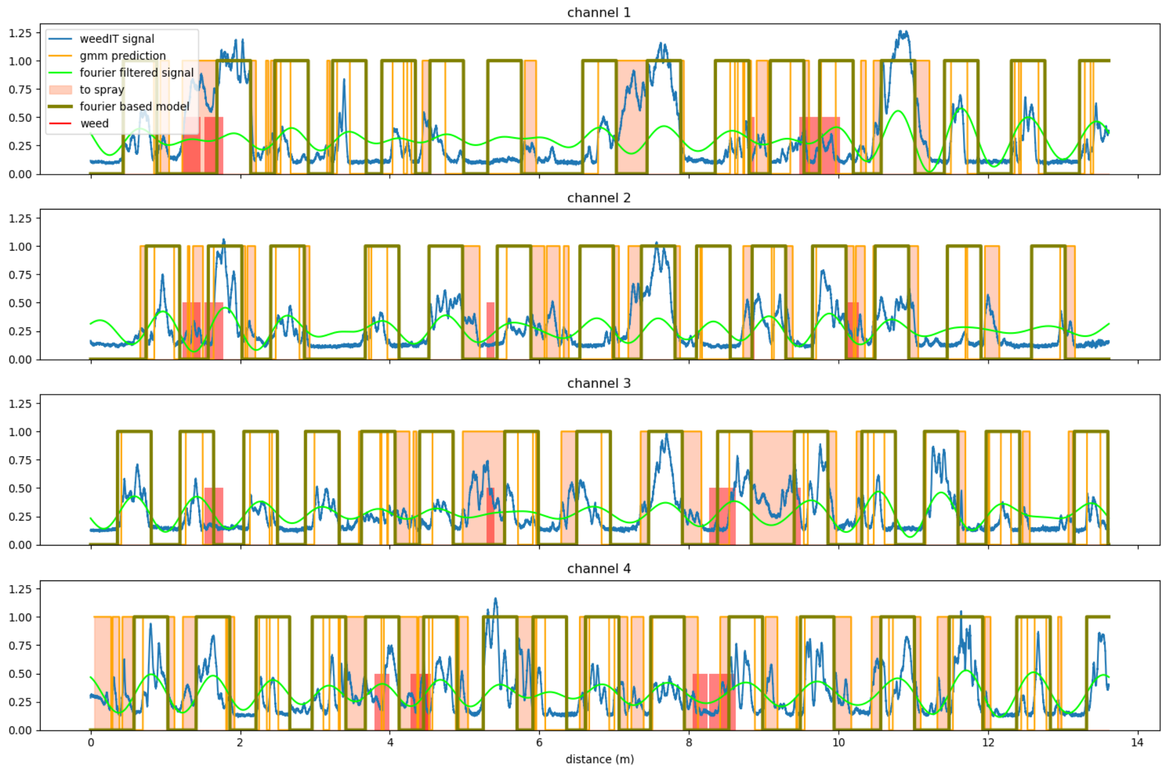


Figure 18: All elements of the algorithm in one figure. Plot 803 on 3-1-2023, the GMM is based on 5 cm, and the row width was set to be 45 cm.

6 Discussion

The distinctiveness calculations show that distinctiveness for this setup, at the growth stage of the crop that spraying is usually done, is actually too low since the distinctiveness becomes negative. Timing (measuring when the crop is big enough to be detected, but the distinctiveness is still positive) is key for the Fourier approach to work well. In our case the crop was still too small when the distinctiveness was not yet negative. In Figure 18 a measurement on 10-1-2023 is shown, where the distinctiveness is negative.

Still, there is some regularity visible in plots. Also a weed is marked (the red line that the arrow points at), that is only half on channel1 and out of sight for the other half. This leads to a decrease in the signal after all.

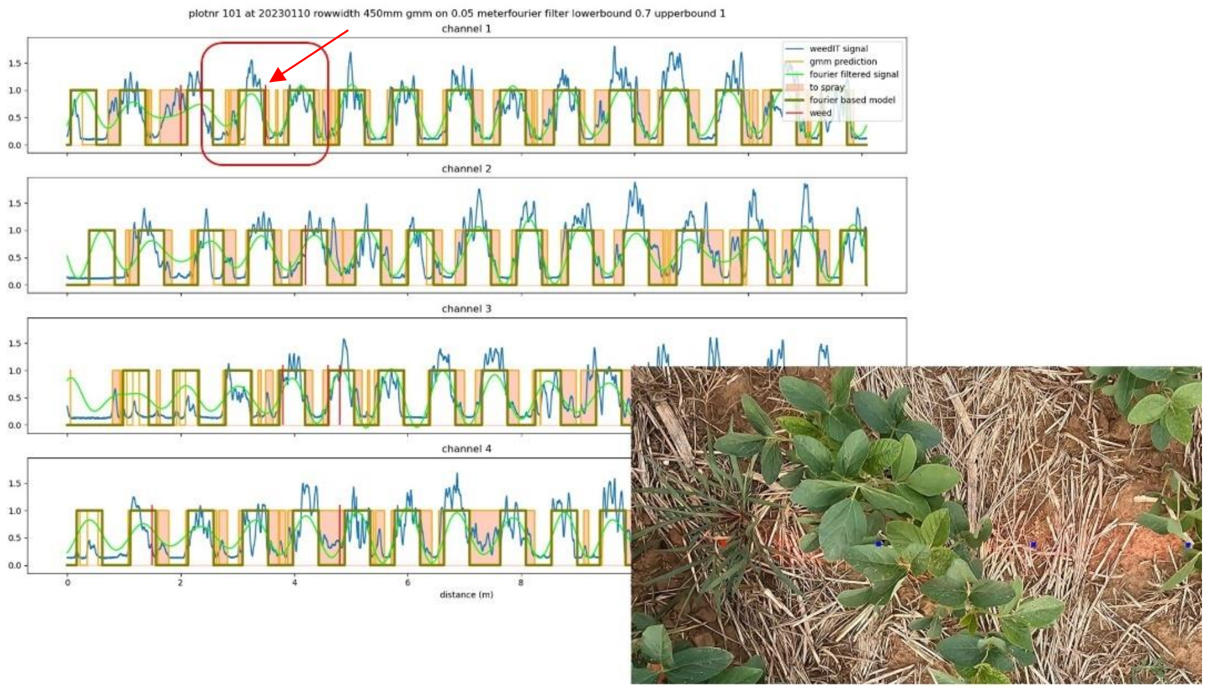


Figure 19: plot 101 on 10-1-2023, a marked weed is detected but leads to a decrease in the signal anyway. The weed on the picture is the weed marked in red on the graphs.

A key assumption in the method to detect the locations of the crop rows using Fourier analysis, is that the rows are regularly planted and emerge regularly. Too large irregularities inevitably lead to mistakes. In Figure 21 it is seen that two crop rows approximated each other very closely. In the red square in Figure 20, the Fourier filtered signal is indeed less regular for that location.

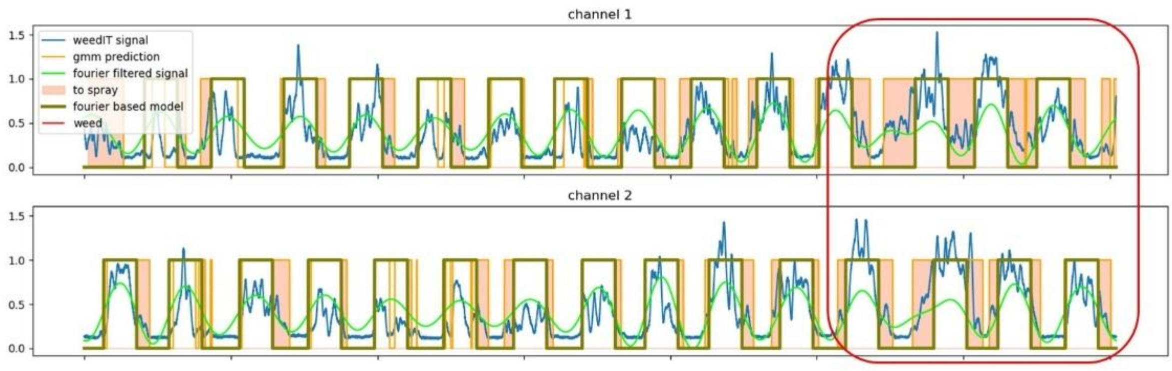


Figure 20: The models on plot 203, where to rows were approaching each other very closely. In the red line were this happened on the videos. This is from 3-1-2023.

Figure 21: A screenshot of the GoPro footage from the same moment as in figure 20.

It is Rometron’s core business to distinguish plant material from soil material, therefore the development of algorithm 1 will likely not be continued in 2024. A good follow-up would be to combine Rometrons algorithm with the Fourier algorithm.

The Fourier model allows for determination of the row locations. It might therefore also be possible to spray only when the center of the spray nozzle is not above the crop. Even if the weeds are not found exactly, this could lead to a reduction in chemical use.

An attempt has been made to find the row widths based on the GMM model. The idea was to find the widths of the crop rows using the GMM model. This turned out to be very subjective for parameter settings and in consult with Rometron this research attempt has been stopped.


Figure 22: Examples of weeds during the measurements on 3 January 2023. The weeds have approximately the same size as the soybean at this moment.


Figure 23: The aftermath of the experiment. Weeds outcompeted the soybean.

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

Report WPR-10.18174/649472