**Supplementary material for the paper:**

**Seed tuber microbiome is a predictor of next-season potato vigor**

---

# Quantification of Potato Plant Vitality in Field Trials

---

**Yang Song, Roeland Berendsen**

Plant-Microbe Interactions
Department of Biology
Faculty of Science
Utrecht University
P.O. Box 800.56
3508 TB Utrecht
Netherlands

**Elisa Atza, Neil Budko**

Numerical Analysis, DIAM, EEMCS
Delft University of Technology
Mekelweg 4
2628 CD Delft
Netherlands

April 10, 2024

**Abstract**

This document describes the field trials for measuring the vitality of potato plants conducted in 2019 and 2020, as well as the methods and techniques pertaining to the extraction and spatial correction of the vitality data.

# 1 Field trials and available data

The field trials for measuring the vitality of potato plants have been conducted during two consecutive years, 2019 and 2020, in the same three geographical locations[1]:

- Montfrin (M), in the South of France (54.4980 N, 5.1090 E)
- Kollumerwaard-SPNA (S), in the North of the Netherlands (ca. 70.4325 N, 6.9825 E)
- Veenklooster (V), in the North of the Netherlands (70.3935 N, 6.7080 E)

Each year the experiments involved 180 batches belonging to 6 varieties (30 batches per variety), labeled as listed in Table 1. Although the same 6 genotypes were studied in both years, the seed tuber batches that represented these genotypes had a different production origin among the years. Observing and explaining the differences between the batches were the main goals of this project.

From the batches that were tested in 2019, only a restricted subset was made available for the microbiome analysis, with the labels of the selected batches given in the second column of Table 1. The selection was made *a posteriori*, after a preliminary analysis of the vitality measurements, to ensure sufficient contrast in vitality among the restricted sets of batches for each genotype.

| genotype | batch labels, 2019 | batch labels, 2020 |
|---|---|---|
| A | 2, 6, 8, 10, 17, 19, 23, 26, 28, 30 | 1-30 |
| B | 31, 34, 38, 47, 50, 51, 52, 56, 58, 60 | 31-60 |
| C | 214, 216, 220, 223, 247, 249, 250, 258, 259, 260 | 61-90 |
| D | 63, 65, 75, 76, 77, 79, 80, 86, 88, 89 | 91-120 |
| E | 91, 96, 98, 100, 103, 104, 107, 115, 117 | 121-150 |
| F | 202, 205, 217, 219, 224, 229, 231, 232, 237, 238 | 151-180 |

Table 1: Labelling of the genotypes, the ranges of the corresponding batch labels, and the batches that were available for the microbiome analysis in 2019 (in 2020 all batches were available).

From each batch of potato seed tubers, 96 tubers were randomly selected for a field trial and planted in four different plots of 24 plants. Thus, with 180 batches, this resulted in 720 plots distributed over the trial field according to a complete randomized block design. For example, the experimental design realized in the Veenklooster test field is shown in Figure 1.

The development of potato plants was documented with images of the complete field taken by a drone mounted camera at certain moments after the planting of the seed tubers up until (and sometimes also after) the canopy closure, i.e., the moment when the leaf canopies of the neighboring plants begin to overlap. The exact dates of the drone images per year and field can be found in Table 2.

The next sections of this document detail our methodology of measuring the plant vitality, with particular emphasis on the following stages:

1. RGB image post-processing and plot localization
2. image segmentation and raw canopy area estimation
3. spatial effect removal
4. choice of the vitality measure
5. correlations in vitality across test fields

[1] All raw, pre- and post-processed vitality data described in this document can be obtained from N.B. (n.v.budko@tudelft.nl)

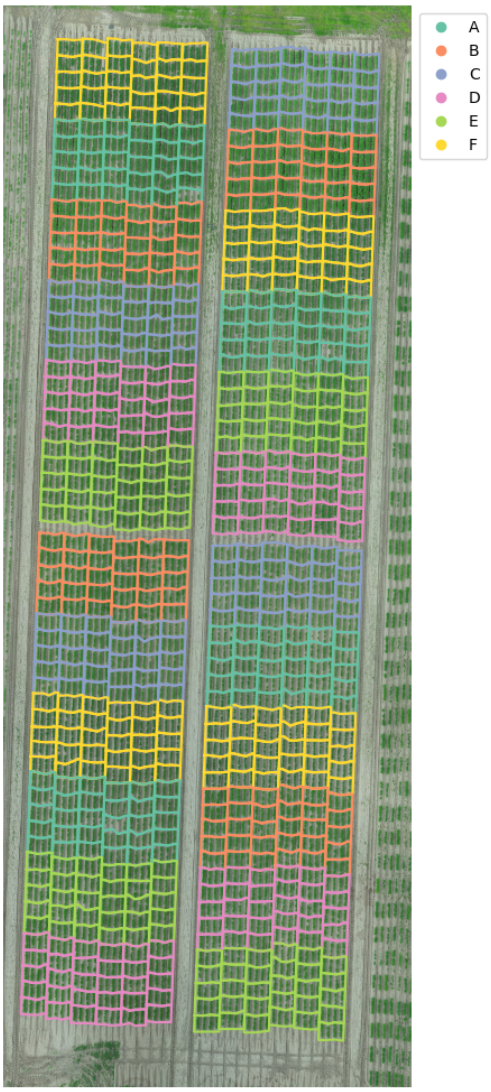


Figure 1: Complete randomized block design in the Veenklooster trial field in 2020. Each genotype (A–F, colors) is repeated four times as four randomly located compact blocks. Each batch has one plot inside the corresponding genotype block (small polygon), i.e., there are four plots of each batch in total.

| field | date | DAP | zero ridges (%) |
|---|---|---|---|
| M | 2019-04-10 | 36 | 630 (21.1 %) |
| M | 2019-04-19 | 45 | 23 (0.8 %) |
| M | 2019-04-26 | 52 | 0 (0.0 %) |
| V | 2019-05-24 | 36 | 134 (4.7 %) |
| V | 2019-05-29 | 41 | 0 (0.0 %) |
| V | 2019-06-07 | 50 | 0 (0.0 %) |
| S | 2019-06-07 | 36 | 12 (0.4 %) |
| S | 2019-06-19 | 48 | 0 (0.0 %) |

| field | date | DAP | zero ridges (%) |
|---|---|---|---|
| M | 2020-04-10 | 35 | 634 (22.0 %) |
| M | 2020-04-13 | 38 | 326 (11.3 %) |
| M | 2020-04-16 | 41 | 48 (1.7 %) |
| M | 2020-04-18 | 43 | 20 (0.7 %) |
| M | 2020-04-22 | 47 | 1 (0.0 %) |
| M | 2020-04-25 | 50 | 0 (0.0 %) |
| V | 2020-05-27 | 35 | 29 (1.0 %) |
| V | 2020-05-30 | 38 | 6 (0.2 %) |
| V | 2020-06-07 | 46 | 2 (0.1 %) |
| V | 2020-06-10 | 49 | 0 (0.0 %) |
| V | 2020-06-12 | 51 | 0 (0.0 %) |
| S | 2020-06-03 | 35 | 41 (1.4 %) |
| S | 2020-06-10 | 42 | 11 (0.4 %) |
| S | 2020-06-12 | 44 | 4 (0.1 %) |
| S | 2020-06-15 | 47 | 4 (0.1 %) |
| S | 2020-06-19 | 51 | 0 (0.0 %) |

Table 2: The dates (year-month-day) of the drone images of the three test fields (M, V, and S) in 2019 (left table) and in 2020 (right table). The 'DAP' column shows the time of the drone image in Days After Planting (DAP). The column 'zero ridges (%)' gives the number of ridges (each plot has four ridges) with no measurable canopy and their fraction among all ridges in the field.

# 2 RGB image post-processing and canopy measurement

## 2.1 Plot localization

Due to the relatively large scale of the trials and the chosen planting technique, the trial fields did not exhibit the usual regular structure with easily identifiable rows and columns of plots. Therefore, we have developed an in-house standardized procedure with minimal manual interaction to detect and identify plots in the trial fields' images. The main steps of the plot-detection algorithm are:

1. From the provided row-column plot labeling and schematics, the expected number $N$ of plots along the ridges of the trial fields is identified
2. The field image where the canopy size allows to detect the gaps between the plots along the ridges is selected from the available images of the trial field. Such an image is usually found towards the end of the canopy growth season, where canopies inside a plot are touching, but have not yet grown as to bridge the gaps to neighbouring plots
3. The beginning and the end of the trial field along each ridge is interactively determined in the selected image
4. The expected number of $N-1$ inter-plot gaps is automatically detected in the images along each ridge. To this end:
   (a) The field image is binarized using a strict green filter such that decidedly green regions will be white and soil will be black in the resulting image. Since the purpose is to find gaps between lots the usage of a strict filter is not detrimental in this case
   (b) The morphological operation of closure is applied to the binarized image to discard possible noise left over from the binarization procedure
   (c) Each ridge is uniformly divided into $N$ sub-intervals
   (d) Image intensity is extracted along three lines parallel and in the neighborhood of the mid-ridge line
   (e) The regions for which the intensity along all three lines is zero (black, i.e. soil in our binarization) are considered the inter-plot gaps and the image coordinates of the midpoint of these regions is saved for further processing
5. The detected plot polygons are displayed and inspected for eventual remaining inaccuracies and distortions and the wrongly identified plot boundary points are corrected interactively

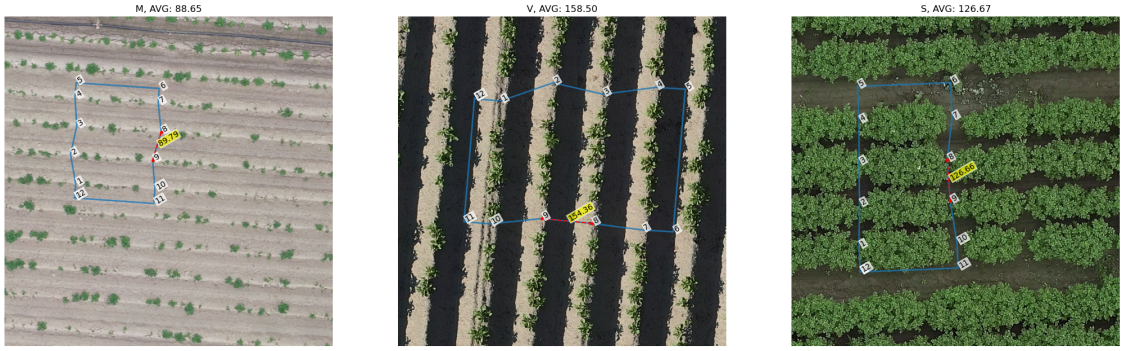


Figure 2: Examples of polygon boundary vertices found in different fields at different times throughout the season with vertex labels. Numbers inside yellow rectangles are the distances between the two vertices in pixels. Image titles show the field average of the inter-ridge distance in pixels, which is subsequently used for the conversion of the canopy area measurements to $cm^2$.

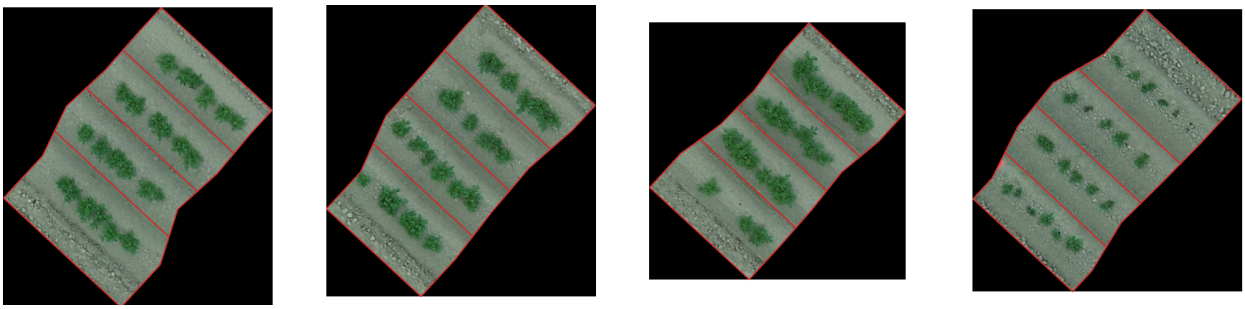

Figure 3: The four plots of the same batch detected in the field. Also shown are the three inter-ridge boundaries within the plots that allow estimating the canopy area on the subplot level thereby contributing to the estimate of the canopy variance.

6. For each plot a set of image coordinates of the plot polygonal boundary is saved

Examples of detected inter-plot gaps along the ridges can be found in Figure 2. The four plots of the same batch detected in a trial field are shown in Figure 3.

## 2.2 Time alignment and distortion correction

The raw RGB data consisting of stitched orthophotographic images of the trial fields obtained at several dates during the growth season are not spatially aligned between the dates and sometimes contain significant spatial distortions.

In the first post-processing step, we searched for the transformations between the reference frames of all images of the same trial field collected during the growth season. More specifically, we were interested in the transformation between the reference frame of one particular orthophoto and all other orthophoto's that preceeded and followed in time. The need for such transformations stems from the fact that the polygonal plot boundaries described in the previous section can be successfully detected only at certain dates during the growth season, when the plants have just the right size. Suppose, the polygonal boundaries of the plots have been identified on some reference date. To avoid having to detect the plots on another date, which is also often impossible, we would like to simply superimpose the previously detected boundary polygons on the second orthophoto. To do that we need to find a transformation of the reference orthophoto to the pixel coordinate system of the second orthophoto. We call this procedure the *time alignment* of the two orthophoto's. While in general this transformation can be rather complicated, in the present case, an affine transformation turns out to be completely sufficient, as was confirmed by the visual inspection of the transformed superimposed polygonal plot boundaries.

Time alignment relies on the presence of time-invariant features with the same physical position in the images. In the 2019 trials we had to rely on natural time-invariant features, such as the irrigation pipes. In the 2020 trials we had artificial, square, 10 cm $\times$ 10 cm, red-colored markers installed in the fields that we could

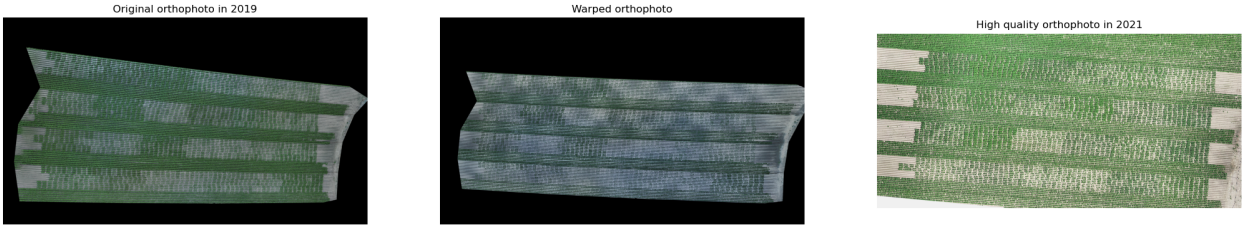


Figure 4: Original distorted 'orthophoto' of the Montfrin field taken in 2019 (left). The same image with distortion corrected by warping to achieve approximately the same inter-ridge distance across the field (middle). High-quality orthophoto of the Montfrin field taken in 2021 (right) confirms that the distortion in 2019 was and artifact of image stitching.

subsequently find in the images and use as reference points.

The detection of these red markers in the 2020 data was aided by the pre-processing of the images, in which bright red pixels are filtered and grouped so that their position can be highlighted on an interactive image plot. The user is shown the image in question with dots in contrasting color plotted at the positions of the algorithmically detected markers. The user has to zoom in on the relevant portions of the field and manually select the middle of the markers.

To understand the simple mathematics behind the recovery of the affine transformation, let there be $N$ markers with their (pixel) coordinates $(x_i, y_i)$, $i = 1, \ldots, N$, on the reference date stored in the vector $\mathbf{p} \in \mathbb{R}^{2N}$ as $\mathbf{p}^T = [x_1, y_1, \ldots, x_N, y_N]$. Suppose that the same $N$ markers have the coordinates $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, \ldots, N$ in another orthophoto taken on another date.

Since a two-dimensional affine transformation is defined by just four numbers, two or more markers with known (same) physical positions in both orthophoto's are sufficient. Although the more markers one has, the more robust is the recovery of the transformation against the noise. With $N$ markers the four transformation elements stored in the vector $\mathbf{t} \in \mathbb{R}^4$ can be recovered by solving the following linear algebraic problem:

$$\begin{bmatrix} 1 & 0 & x_1 & -y_1 \\ 0 & 1 & y_1 & x_1 \\ \ldots & & & \\ 1 & 0 & x_N & -y_N \\ 0 & 1 & y_N & x_N \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{y}_1 \\ \vdots \\ \tilde{x}_N \\ \tilde{y}_N \end{bmatrix}. \tag{1}$$

Given a sufficient number of markers, there exists the unique least-squares solution $\hat{\mathbf{t}}^T = [\hat{t}_1, \hat{t}_2, \hat{t}_3, \hat{t}_4]$ to this problem, which can be used to determine the location of each vertex of the polygonal plot boundary in the second orthophoto. Let the (pixel) vertex coordinates on the reference date be $(x, y)$, then the pixel coordinates $(\tilde{x}, \tilde{y})$ of this vertex in the orthophoto from another date can be obtained as:

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} \hat{t}_1 \\ \hat{t}_2 \end{bmatrix} + \begin{bmatrix} \hat{t}_2 & -\hat{t}_3 \\ \hat{t}_3 & \hat{t}_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{2}$$

The stitched orthophoto images of the Montfrin test field in 2019 contained an additional strong distortion, see Figure 4, making the inter-ridge distance measured on one side of the field strongly differ from the inter-ridge distance measured on the other side. The fact that it was a distortion of the image rather than the natural shape of the field is obvious from the high-quality orthophoto of the same field taken in 2021 (Figure 4, right). Since this distortion has consequences for both the estimation of the green area and the spatial correction, the original orthophoto was re-processed. The field image was warped so that the inter-ridge distance in the "left" part of the image, as measured with the help of selected polygonal vertices, became closer to the inter-ridge distance measured on the "right". The same transformation was used to warp the polygonal plot boundaries. No such distortions were observed or had to be corrected in other fields and years.

## 2.3 Leaf canopy segmentation and estimation

To measure the canopy area within the polygonal plot boundary, the image pixels are segmented into two disjoint sets: pixels of the canopy and pixels of the surrounding soil. Then, the canopy pixels are counted and the result converted to the $\text{cm}^2$ units.

While a human operator is usually very successful in segmenting an image, a fully automated segmentation procedure that would work in all circumstances is not available. The present dataset featured a variety of illumination and moisture conditions, both of which affect the color of pixels. Also, leaf canopy colors have systematic differences between genotypes, ranging from light green to almost purple. Therefore, every orthophoto had to be processed individually, resulting in different segmentation filters with date- and field-specific parameters. In all cases, the quality of segmentation has been confirmed by visual inspection of randomly selected plots of each genotype.

For all fields and years, except Montfrin 2019, the following simple segmentation procedure gave satisfactory segmentation results:

1. Convert the image to HSV format
2. Find date-specific range of the Hue and Saturation channels for the canopy (by visual inspection) and set all the other pixels to 'black'
3. Filter the Value channel by allowing only the pixels with the value below a threshold and setting all other pixels to 'black'
4. Binarize the image (white pixels – canopy, black pixels – soil)

Orthophoto's of the Montfrin field in 2019 have a lower spatial resolution and overall quality than all other orthophoto's (later images have been acquired with a higher resolution camera and by a more experienced drone operator). Therefore, the segmentation of the Montfrin 2019 orthophoto's required additional effort and contained some additional steps:

1. Convert the image to HSV format
2. Find date-specific range of the Hue channel for the canopy. This was done by analyzing the Hue channel histogram, which showed two peaks, canopy and soil, with the canopy peak growing in time. Set all the other pixels to 'black'
3. Equalize Saturation and Value channels
4. Filter the Value channel by allowing only the pixels with the value below a threshold and setting all other pixels to 'black'
5. Obtain a grayscale image by applying a weighted combination (weights were manually adjusted to achieve the best segmentation) of some standard agricultural 'green' filters. The Normalized Difference Red/Green Redness Index (RI) [5], [2], Excess Green (EXG) [3], and Brightness Index (BI) [4]:

$$\mathrm{BI} = \sqrt{\frac{r^2 + g^2}{2}} \tag{3}$$

$$\mathrm{EXG} = 2g - (r + b) \tag{4}$$

$$\mathrm{RI} = \frac{r - g}{r + g} \tag{5}$$

6. Normalize the image, set all pixels below a threshold to 'black', and binarize the image
7. Remove 'salt and pepper' noise by a morphological transformation
8. White pixels – canopy, black pixels – soil

The parameters of the above segmentation procedures are provided in Table 3 for each field and date.

## 2.4 Conversion from pixels to cm$^2$

After segmentation, the mean canopy area $S_{\mathrm{px}}$ (in pixels) over each ridge of each plot was determined by summing all white pixels within the geometrical boundaries of the ridge and dividing by 6 – the number of plants in each ridge. To convert a canopy area in pixels to its area in cm$^2$, we use the fact that the distance $d_{\mathrm{cm}}$ between the ridges in the field is determined by the planting device and is $d_{\mathrm{cm}} = 75$ cm in the Veenklooster (V) and Kollumerwaard-SPNA (S) fields, and $d_{\mathrm{cm}} = 74$ cm in the Montfrin (M) field.

| field/date | H range | S range | S,V equalize | V filter | weights (BI,RI,EXG) | thresh. | morph. |
|---|---|---|---|---|---|---|---|
| M-2019-04-10 | [40, 80] | [0, 255] | yes | [0, 100] | 0.8, 0.8, 1.0 | 0.3 | yes |
| M-2019-04-19 | [25, 100] | [0, 255] | no | [0, 110] | 0.1, 0.3, 0.6 | 0.25 | yes |
| M-2019-04-26 | [25, 80] | [0, 255] | no | no | 0.0, 0.0, 1.0 | 0.35 | no |
| V-2019-05-24 | [30, 125] | [63, 255] | – | – | – | – | – |
| V-2019-05-29 | [30, 100] | [63, 255] | – | – | – | – | – |
| V-2019-06-07 | [50, 125] | [0, 255] | – | – | – | – | – |
| S-2019-06-07 | [30, 110] | [30, 150] | – | – | – | – | – |
| S-2019-06-19 | [30, 100] | [55, 255] | – | – | – | – | – |
| M-2020-04-10 | [40, 68] | [86, 162] | – | – | – | – | – |
| M-2020-04-13 | [40, 68] | [86, 162] | – | – | – | – | – |
| M-2020-04-18 | [40, 68] | [86, 162] | – | – | – | – | – |
| M-2020-04-22 | [40, 68] | [86, 162] | – | – | – | – | – |
| M-2020-04-25 | [50, 80] | [60, 255] | – | – | – | – | – |
| V-2020-05-27 | [35, 120] | [59, 255] | – | – | – | – | – |
| V-2020-05-30 | [35, 80] | [60, 255] | – | – | – | – | – |
| V-2020-06-07 | [50, 120] | [100, 140] | – | – | – | – | – |
| V-2020-06-10 | [50, 120] | [100, 140] | – | – | – | – | – |
| V-2020-06-12 | [50, 120] | [100, 140] | – | – | – | – | – |
| S-2020-06-03 | [40, 100] | [80, 255] | – | – | – | – | – |
| S-2020-06-10 | [35, 120] | [80, 160] | – | – | – | – | – |
| S-2020-06-12 | [35, 120] | [80, 160] | – | – | – | – | – |
| S-2020-06-15 | [35, 120] | [100, 200] | – | – | – | – | – |
| S-2020-06-19 | [40, 100] | [80, 255] | – | – | – | – | – |

Table 3: Field- and date-specific parameters for canopy segmentation in orthophoto's.

To find the pixel-to-cm conversion factor, we compute the average pixel distance $d_{\text{px}}$ between the adjacent ridges in the field for a specific date (see Figure 2). Then, the area $S_1$ of a single pixel in $\text{cm}^2$ is given by:

$$S_1 = \left(\frac{d_{\text{cm}}}{d_{\text{px}}}\right)^2. \tag{6}$$

Thus, the canopy area $S$ in $\text{cm}^2$ is obtained from the canopy area $S_{\text{px}}$ in pixels as:

$$S = S_1 S_{\text{px}}. \tag{7}$$

# 3 Spatial effect removal

The ridge-mean plot canopies obtained after the transformation, plot localization, and segmentation procedures described above constitute the raw data and cannot be used to estimate the mean batch canopy, since the test fields are usually spatially nonuniform, which may systematically increase or decrease the canopy size in certain areas of the field. The spatial effect is well-illustrated in Figure 3, showing the four plots of the same batch, with the plot depicted in the right-most image having significantly smaller canopies than the other three plots. Obviously this plot has been affected by some unfavourable growth conditions and would pull down the estimate of the mean canopy if included in the calculations as it is.

```
library(statgenSTA)
library(SpATS)

CanopyMeasurement <- read.csv("/Path_to_Canopy_measurement.csv", header = TRUE, sep = ",")

TableToFit <- createTD(data = CanopyMeasurement, genotype = "Batch",
                       repId = "Block", subBlock = "Block",
                       rowCoord = "Row", colCoord = "Col")

FittedModel <- fitTD(TD = TableToFit, trials='trialname',traits = "Canopy",
                     design = "rcbd")

BLUEsTable <- extractSTA(STA = FittedModel, what = "BLUEs", keep = "Variety")
```

Listing 1: Code to produce the array of spatially corrected canopies from the raw measured array in R

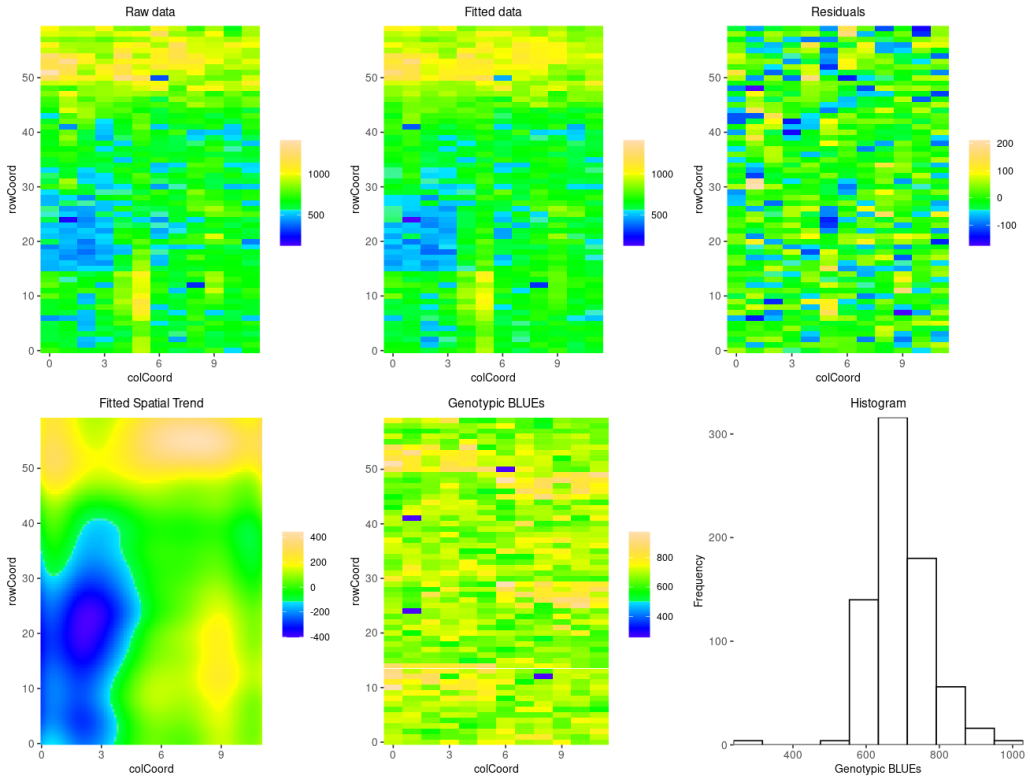


Figure 5: Spatial plot produced by the package SpaTS as an illustration of spatial effect removal.

We use the state of the art spatial effect removal method [6], implemented in the R-package SpaTS [1], that removes both random and fixed spatial effects and provides the Best Linear Unbiased Estimate (BLUE) of the mean batch canopy size. The typical output of the SpaTS package is illustrated in Figure 5. We apply the spatial effect removal to the raw canopy data obtained from all available orthophoto's.

# 4 Example of the canopy area estimation procedure

The following Python code will produce the Figure...

# 5 Choice of the vitality measure

There is no universally accepted measure of the plant vitality. In fact, although it may be a matter of semantics, one could argue that the term 'vigor' is more applicable in the present situation. In any case, the size of the leaf canopy to a large extent determines the yield of potato plants, which is the ultimate goal of potato farmers. Therefore, the sooner a plant can develop a large canopy the more vigorous or vital is this plant. From this point of view, it appears that considering the time evolution of the canopy could eventually lead to a measure of vitality. Unfortunately, systematic analysis of the canopy time evolution could not be achieved in this project due to sparsity and inconsistency of the imaging dates. Nevertheless, a well-chosen single-time measurement of the canopy size may be indicative of the plant vitality as well. One should be aware though that a large canopy at a certain point in time could be either due to an early-emergent but relatively slow-growing plant or due to a late-emergent and fast-growing plant. Nevertheless, a relatively large canopy somewhere in the middle of the growth season, after the early 'transients' related to the emergence time and the initial rate of growth have passed, is surely indicative of the plant vitality. Indeed, as long as a plant acquires a large canopy at a useful point in time, i.e., not too late in the season, it can be considered a vital/vigorous plant, and it does not really matter whether it is due to an early emergence time or a fast initial growth rate.

Figure 6 and Figure 7 show the summary of the BLUE canopy data for each of the six genotypes at the available time points for each year and field. As one can see, the growth pattern and the attained canopy size are different between the fields and years. This can be attributed to the varying weather conditions and soil types, caused mainly by the difference in geographic location and the different times of the year the seed tubers

were planted in each field (see Table 2). Hence, a single *a priori* chosen day after planting cannot be used as a time point to take the vitality measurement. Therefore, our choice of the canopy measurement date is year- and field-specific and is guided by the data and the following basic principles:

- It should be possible to compare different batches within each genotype
- The plant-plant and batch-batch interactions should be avoided

From this point of view, there is just one or two dates per year and field among the available data that can be used as the vitality measure. Indeed, in the early measurements not all plants have emerged yet, which does not allow for a fair comparison of the batches. In the late measurements, the canopies of rapidly growing genotypes start "merging" not only inside the plots, but also with those of the neighboring plots, thus introducing plant-plant and batch-batch interactions.

We have noticed that from 47 to 50 DAP's all plants have emerged and the canopies are not yet overlapping, making this a good time range to consider and reducing the choice to one or two dates per field in each year. From Figure 6 and Figure 7 it is clear that this corresponds to the end of the exponential growth period and precedes the period of 'saturation' in the expected sigmoid growth curves. The remaining choice between the few acceptable dates was made by taking into account the quality of the orthophoto's, i.e., the precision of the canopy area measurements. It should be mentioned that these final choices do not alter the conclusions of the subsequent association/regression studies in any significant way, since the data at these time points are very highly correlated, Table 4.

| | M 52 | S 48 | V 50 |
|---|---|---|---|
| M 45 | 90% | - | - |
| S 36 | - | 98 % | - |
| V 41 | - | - | 90% |

| | M 47 | M 50 | S 47 | S 51 | V 49 | V 51 |
|---|---|---|---|---|---|---|
| M 43 | 99% | 98 % | - | - | - | - |
| M 47 | - | 100% | - | - | - | - |
| S 44 | - | - | 95% | 91 % | - | - |
| S 47 | - | - | - | 97 % | - | - |
| V 46 | - | - | - | - | 90% | 87% |
| V 49 | - | - | - | - | - | 94% |

Table 4: Correlations of canopy size at two points in time in 2019 (left) and in 2020 (right).

Our final choices of the dates for the vitality measurement are highlighted with green boxes along the $x$-axis of In Figure 6 and Figure 7.

# 6 Correlations in vitality across fields

After all the necessary steps for the vitality data extraction and removal of known effects, an exploratory analysis of the vitality data has been performed to ascertain the plausibility of the main project hypothesis that the vitality of a plant is determined, at least to a certain extent, by the seed tuber from which the plant has grown. The presence of such dependence could be inferred if the vitality of the plants produced by the seed tubers of a given batch, relative to other batches, was consistent for repetitions inside the field and, especially, across different fields.

Pearson correlation of the vitality data provides an adequate measure of consistency. Correlation analysis can only be applied within a given year, where tubers of the same batches were planted in three different fields. As was mentioned above, while the same genotypes were tested in both years, the seed-tuber batches had a different production origin in each year and cannot be directly compared for consistency.

Correlations between the raw and spatially corrected (BLUE) vitality data across fields in each year are shown in Figure 8 and Figure 9. The symbols along the horizontal and vertical axes denote the test fields. The range of statistical significance ($p$-value) of these correlation results is indicated with the star symbols next to the value: three stars, $p < 0.001$, i.e., very significant, one or no stars – not significant. One can see that both the raw and the BLUE data significantly correlate across the fields in both years, with the spatial effect removal leading to a slight increase in the correlation. The fact that the correlations are not perfect can be explained by the influence of the weather and soil conditions, inconsistent application of the herbicide, and the aging of the seed tubers.

Figure 10 presents the correlations individually for each genotype and shows a much more varying picture with the lack of correlation for some genotypes in some years. For instance, the seed batches of the genotype D performed consistently in 2019 across all three fields, whereas, in 2020, the only strong correlation was between the Montfrin and the Kollumerwaard-SPNA fields, and the other correlations were weak and insignificant. At the same time, the seed batches of the genotype C show persistent correlations across all fields in both years.

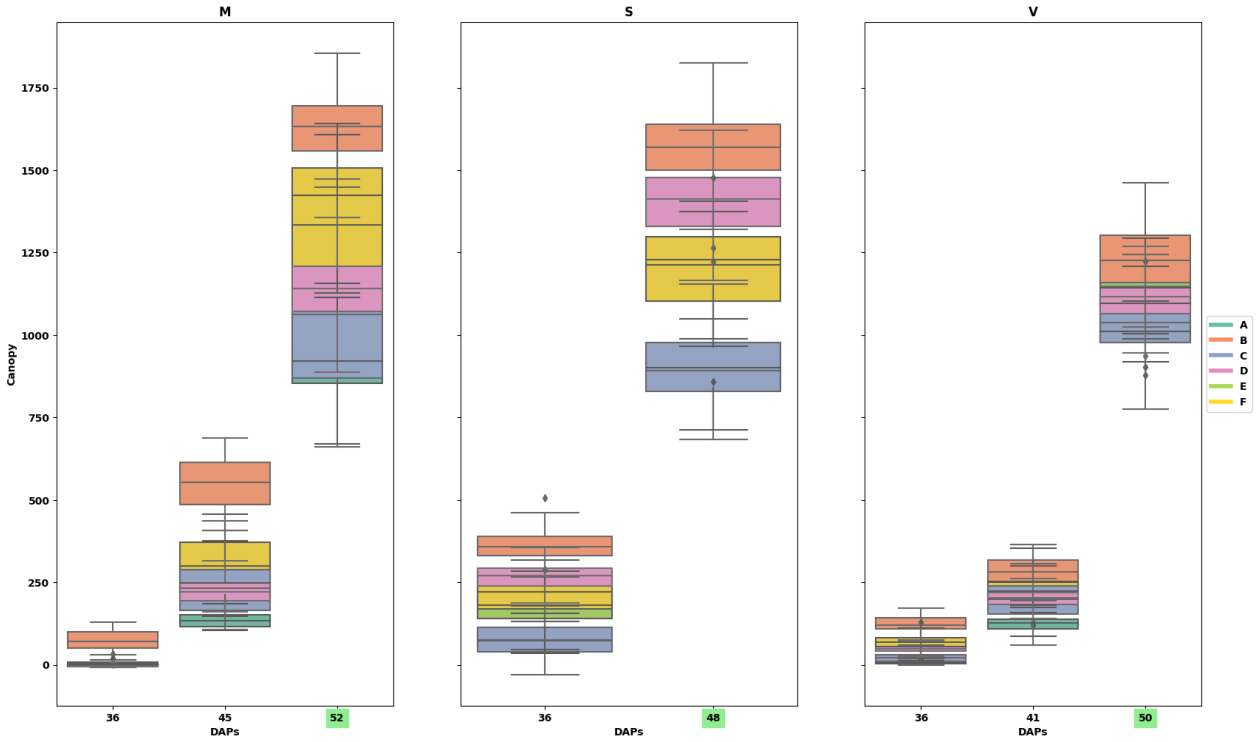


Figure 6: Summary of all available measurements in 2019 per genotype as boxplots. The choice of the date for the vitality measurement is highlighted in green color on the $x$-axis.

# 7 Precision measure for microbiome predictions and confusion matrix

The random forest (RF) algorithm allows to make a prediction of the continuous vitality parameter based on the microbial composition of different seedlots. While the accuracy of this prediction can sometimes be low, it may still be sufficient to categorize the vitality of a seedlot as belonging to one of the three practically important classes or quantiles: low (L), middle (M), or high (H) vitality. We introduce the three quantiles of interest as the values that divide the vatality range into three portions of equal probability. In the absence of a well-defined *a-priori* probability distribution, one can estimate these quantile boundaries by putting them between the three equal-size portions of the ordered data vector.

After assigning a quantile-based class to each predicted vitality value, i.e., quantization of the results, the regression problem turns into a classification problem, where other measures of accuracy should be applied. To measure the performance of the classification model, we compute the class-specific precision measure for a class $C_i$ as:

$$\text{Precision}(C_i) = \frac{\text{True positives}(C_i)}{\text{True positives}(C_i) + \text{False positives}(C_i)}, \tag{8}$$

which is the ratio of correctly classified samples in $C_i$ over the total number of samples classified as $C_i$. Note, that this is an estimate of the conditional probability:

$$\text{Precision}(C_i) \approx P(C_i|C_i) = \mathbb{P}(x \text{ measured } C_i | x \text{ predicted } C_i) = \frac{\mathbb{P}(x \text{ predicted } C_i \cap x \text{ measured } C_i)}{\mathbb{P}(x \text{ predicted } C_i)}$$

A comprehensive measure of the model performance is the confusion matrix which contains all possible conditional probabilities:

$$\begin{bmatrix} P(L|H) & P(M|H) & P(H|H) \\ P(L|M) & P(M|M) & P(H|M) \\ P(L|L) & P(M|L) & P(H|L) \end{bmatrix},$$

where the precision measures are on the diagonal and the cross-class conditional probabilities are off diagonal and are estimated as:

$$P(C_i|C_j) \approx \frac{\text{Number of measured } C_i \text{ predicted as } C_j}{\text{True positives}(C_j) + \text{False positives}(C_j)}. \tag{9}$$

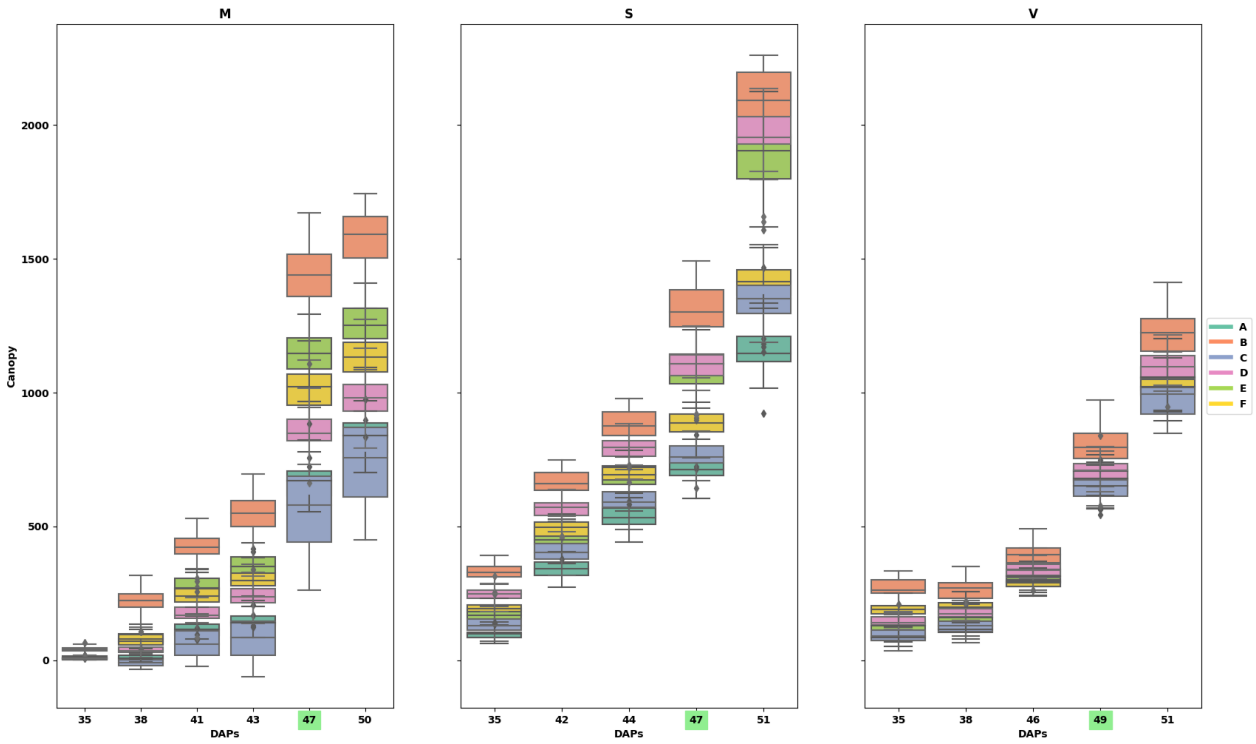


Figure 7: Summary of all available measurements in 2020 per genotype as boxplots. The choice of the date for the vitality measurement is highlighted in green color on the $x$-axis.

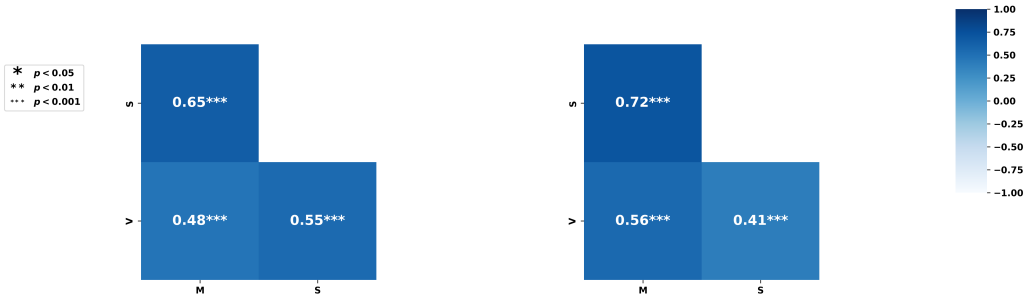


Figure 8: Correlations in raw vitality data between the fields in 2019 (left) and 2020 (right).

For example, in the first row of the matrix we see how likely it is for a sample, which we classify as highly vital, to have been measured as of low, medium, or high vitality.

Intuitively, a good classification model will have values close to one on the anti-diagonal of the confusion matrix and low values off the anti-diagonal, indicating that the probability to be misclassified is much lower than the probability to be classified correctly. In particular, the magnitude of the first element of the first row and the third element of the third row measure the probabilities of the most severe and practically dangerous misclassifications.

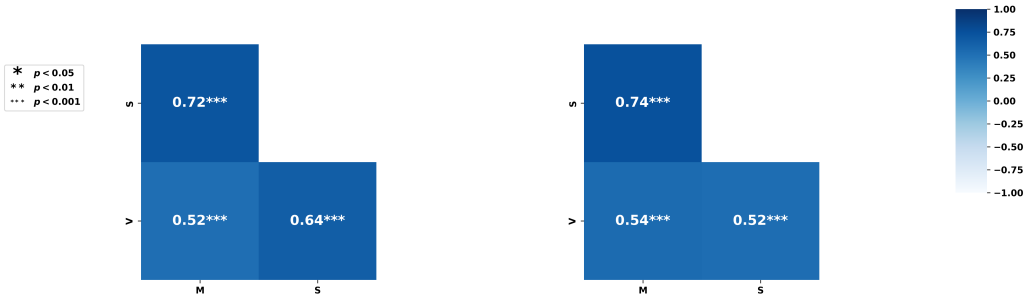


Figure 9: Correlations in spatially-corrected (BLUE) vitality data between the fields in 2019 (left) and 2020 (right).

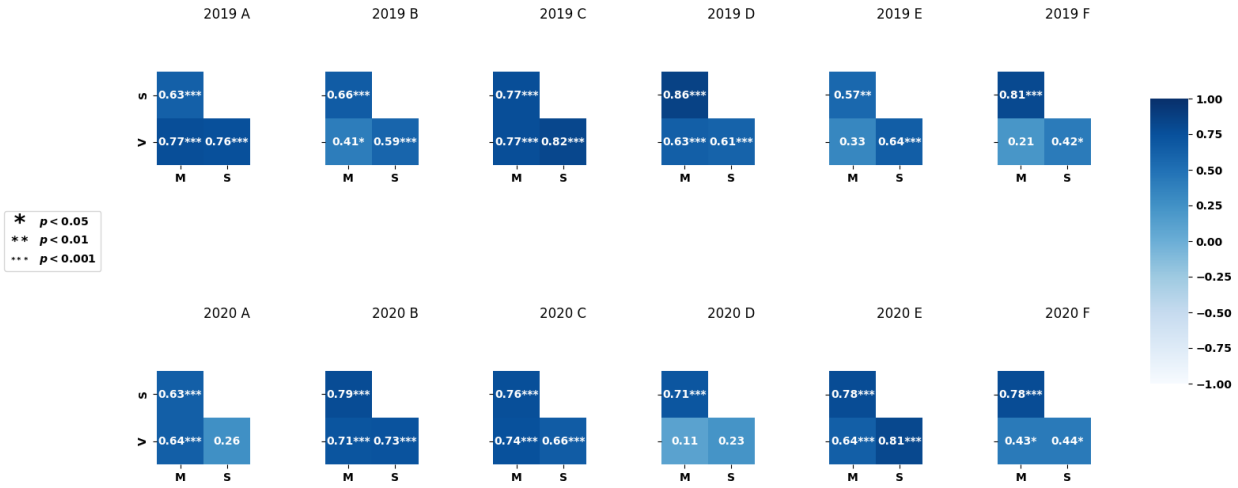


Figure 10: Correlations in spatially-corrected (BLUE) vitality data between the fields per genotype in 2019 (top row) and 2020 (bottom row).

# References


[1] https://cran.r-project.org/package=spats.

[2] https://www.indexdatabase.de/db/i-single.php?id=74.

[3] G. E. Meyer D. M. Woebbecke and K. Von Bargen et al. Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, 38(1):259–269, 1995.

[4] Richard Escadafal. Remote sensing of arid soil surface color with landsat thematic mapper. *Advances in space research*, 9(1):159–163, 1989.

[5] Escadafal R. and Huete A. Improvement in remote sensing of low vegetation cover in arid regions by correcting vegetation indices for soil noise. *Comptes Rendus de l'Academie des Sciences Serie 2*, 312(11):1385–1391, 1991.

[6] María Xosé Rodríguez-Álvarez, Martin P. Boer, Fred A. van Eeuwijk, and Paul H.C. Eilers. Correcting for spatial heterogeneity in plant breeding experiments with p-splines. *Spatial Statistics*, 23:52–71, 2018.