University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Papers in Natural Resources | Natural Resources, School of

2022

Rapid online plant leaf area change detection with high-throughput plant image data

Y. Zhan

R. Zhang

Y. Zhou

V. Stoerger

J. Hiller
University of Nebraska - Lincoln

See next page for additional authors

Follow this and additional works at: https://digitalcommons.unl.edu/natrespapers

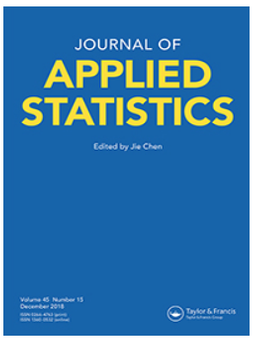
 Part of the Natural Resources and Conservation Commons, Natural Resources Management and Policy Commons, and the Other Environmental Sciences Commons

Zhan, Y.; Zhang, R.; Zhou, Y.; Stoerger, V.; Hiller, J.; Awada, T. N.; and Ge, Y., "Rapid online plant leaf area change detection with high-throughput plant image data" (2022). *Papers in Natural Resources*. 1561.
https://digitalcommons.unl.edu/natrespapers/1561

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors
Y. Zhan, R. Zhang, Y. Zhou, V. Stoerger, J. Hiller, T. N. Awada, and Y. Ge

This article is available at DigitalCommons@University of Nebraska - Lincoln: https://digitalcommons.unl.edu/natrespapers/1561



Taylor & Francis
Taylor & Francis Group

Journal of Applied Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Rapid online plant leaf area change detection with high-throughput plant image data

Yinglun Zhan, Ruizhi Zhang, Yuzhen Zhou, Vincent Stoerger, Jeremy Hiller, Tala Awada & Yufeng Ge

To cite this article: Yinglun Zhan, Ruizhi Zhang, Yuzhen Zhou, Vincent Stoerger, Jeremy Hiller, Tala Awada & Yufeng Ge (2022): Rapid online plant leaf area change detection with high-throughput plant image data, Journal of Applied Statistics, DOI: 10.1080/02664763.2022.2150753

To link to this article: https://doi.org/10.1080/02664763.2022.2150753

Published online: 05 Dec 2022.

Submit your article to this journal

Article views: 60

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=cjas20

JOURNAL OF APPLIED STATISTICS
https://doi.org/10.1080/02664763.2022.2150753

Taylor & Francis
Taylor & Francis Group

Check for updates

Rapid online plant leaf area change detection with high-throughput plant image data

Yinglun Zhan[a], Ruizhi Zhang[a*], Yuzhen Zhou[a**], Vincent Stoerger[b], Jeremy Hiller[c], Tala Awada[c] and Yufeng Ge[d]

[a]Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE, USA; [b]Agricultural Research Division, University of Nebraska-Lincoln, Lincoln, NE, USA; [c]School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA; [d]Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE, USA

ABSTRACT

High-throughput plant phenotyping (HTPP) has become an emerging technique to study plant traits due to its fast, labor-saving, accurate and non-destructive nature. It has wide applications in plant breeding and crop management. However, the resulting massive image data has raised a challenge associated with efficient plant traits prediction and anomaly detection. In this paper, we propose a two-step image-based online detection framework for monitoring and quick change detection of the individual plant leaf area via real-time imaging data. Our proposed method is able to achieve a smaller detection delay compared with some baseline methods under some predefined false alarm rate constraint. Moreover, it does not need to store all past image information and can be implemented in real time. The efficiency of the proposed framework is validated by a real data analysis.

ARTICLE HISTORY
Received 26 January 2022
Accepted 12 November 2022

KEYWORDS
Supervised learning; ADMM algorithm; online detection; adaptive cusum; plant leaf area; high-throughput plant phenotyping (HTPP)

1. Introduction

Agriculture is facing global challenges associated with meeting the demands of a growing population under dwindling natural resources, climate change, and increased frequency and intensity of extreme weather events. To meet the growing demands for food, fiber, biofuel, and feed, global crop production needs to double by 2050 [33]. One promising solution is to accelerate the development of crops with desirable traits (e.g. improved yield, and abiotic and biotic stress resistance) via traditional and emerging breeding and molecular technologies. Technological innovations in the past decade resulted in two major advances in plant sciences, the development of next-generation genetic sequencing and high-throughput phenotyping [10]. Although significant advancements have been made in sequencing the genetic material of economically important crops, the high-throughput assessment of phenomics traits associated with the expression of these genes has lagged

CONTACT Yuzhen Zhou zhouyuzh@msu.edu Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

*Present address: Department of Statistics, University of Georgia, Athens, GA, USA

**The work was done at the University of Nebraska-Lincoln prior joining to Amazon.

© 2022 Informa UK Limited, trading as Taylor & Francis Group

behind. In fact, plant phenotyping has become the bottleneck in the process, due to its costly and labor-intensive nature [11]. Also, efficient plant phenotyping is needed in crop management, such as irrigation and disease control [7]. Traditional low-throughput methods of measuring plant traits often rely on plant scientists and student workers manually measuring destructively or non-destructively crops. With the rapid advancements in imaging techniques, including visible (red, green, blue; RGB), multispectral, thermal, hyperspectral, and fluorescence imaging, high-throughput plant phenotyping (HTPP) has enabled new prospects for non-destructive plant traits measurement [40]. Without the need of destructive sampling, HTPP is able to acquire plant images regularly during the whole life cycle of the plant. The sampling frequency is usually much higher than the traditional methods. It makes it possible to monitor plant traits in real time and detect any significant change [9]. Yet, the new technology raises a challenge in the analysis, that is developing efficient statistical models to predict plant traits and monitor plant growth dynamics or stress based on a huge volume of plant imaging data [8].

To address this challenge, several efforts have emerged in recent years to address HTPP image analysis, modeling, and detection. Adams *et al.* [1] proposed a novel approach on image labeling to segment plants from the background with supervised learning methods using RGB images of maize plants. Xu *et al.* [42] used functional data analysis to study the maize plant growth dynamics with the extracted plant sizes from RGB images. Wang *et al.* [41] developed an R package called 'implant' for RGB plant image segmentation and functional data analysis of the extracted plant traits. Bashyam *et al.* [3] developed an automated leaf tracking method to determine corn plant growth based on image segmentation analysis. To evaluate whether the traits of plants under some treatment are significantly different from those measured in healthy plants at a given time point, say t_0, researchers usually use statistical hypothesis testings, such as two sample t test and ANOVA F test, or classification algorithms, such as support vector machines, *only* with image data collected at t_0. Behmann *et al.* [4] applied classification algorithms such as support vector machines to separate the drought-stressed plants from the well-watered plants with hyperspectral images. Asaari *et al.* [2] used a spectral similarity measure and F test to monitor water stress and recovery with hyperspectral images in real time. Romer *et al.* [36] used simplex volume maximization for feature extraction from hyperspectral images and applied MANOVA test for the real-time drought detection. However, without utilizing the historical plant traits information prior to the given time t_0, those methods are not optimal in terms of making the quickest detection of plant traits change under the false alarm rate constraint. In this paper, we focus on developing a real-time image-based plant traits monitoring and detection model, which is efficient in computing and can achieve a smaller detection delay by integrating the plant traits information up to the time t_0.

Efficient online change detection methods with time-series data are well established and have been widely applied in many areas such as quality control [16], financial analysis [18], medical condition monitoring [5], and climate change detection [15]. The classical version of this problem, where one monitors independent and identically distributed (iid) univariate or low-dimensional multivariate observations, is a well-developed area, and many classical procedures have been developed such as the Shewhart chart [38], Page's CUSUM procedure [27], EWMA chart [34], and Shiryaev–Roberts procedure [35,39]. These procedures enjoy nice optimal properties when the pre-change and post-change distributions are

fully specified [22,29]. When the post-change parameters are unknown, generalized likelihood ratio based procedures and the adaptive CUSUM based procedures are often used to detect the possible change [17,21,23]. Yet, to the best of our knowledge, efficient CUSUM-based detection methods have not been established in the area of high-throughput plant phenotyping.

We proposed a two-step image-based online detection method to detect when an individual plant leaf area growth changes its dynamics, for instance, would slow down under a given stress, e.g. drought. First, we developed an efficient supervised learning algorithm to extract plant leaf area from the multiview RGB images. Then, we proposed to use the adaptive CUSUM procedure [23] to monitor the standardized relative change of the predicted plant leaf area. There are three advantages of our proposed method. First, our proposed framework of monitoring and quick detection of the individual plant leaf area via the real-time imaging data is general and can be used to monitor other features of the plant. Second, due to the recursive format of the adaptive CUSUM statistics, our proposed method does not need to store all past image information and can be implemented in real time. Therefore, our proposed method can be used to monitor a large number of plants simultaneously. Third, our proposed method does not rely on the true post-change information. Instead, given the minimal magnitude of the meaningful change based on some domain knowledge, our method can estimate the true post-change information adaptively and detect the change fast. To validate the efficiency of our proposed framework and method, we conducted a case study on the data sequence of multiview RGB images of 100 soybean plants, which were collected almost every 2 days over a 2-month period at the Nebraska Innovation Campus Greenhouse, High-Throughput Plant Phenotyping Core Facilities (Scanalyzer 3D, LemnaTec Gmbh, Aachen, Germany), University of Nebraska-Lincoln. The soybean plants were randomly split into a control group (watered, 80% field capacity) and a treatment group (water-stressed or drought, 40% field capacity). Our method allowed the detection of small and significant changes in plant leaf area growth over time under water-stress or drought treatment.

This paper consists of four sections. Section 1 is the introduction. Our proposed image-based plant monitoring and detection method is presented in Section 2. To validate the efficiency of the method, a case study is given in Section 3. We conclude the paper in Section 4.

2. Methods

Our proposed image-based plant monitoring and detection method include two steps, which will be presented in Section 2.1 and Section 2.2 respectively. First, an efficient supervised learning algorithm is developed to predict the plant leaf area with multiview RGB images. Second, we introduce the adaptive CUSUM procedure to detect the time when the individual plant leaf area growth has been affected by drought.

2.1. Model for plant leaf area

Suppose the study was conducted during the time period $[0, T]$ with n plants. Let $y_i(t)$ be the ith plant leaf area at time t, where $t \in [0, T]$, $i = 1, 2, \ldots, n$. To obtain plant leaf area measurements, a subset of plants was destructively sampled. Specifically, for the destructed

plant, all plant material above soil is harvested and fractionated into leaves, stems, and pods, and leaf area is then measured by a leaf area meter. Consequently, we can only have one leaf area measurement per destructed plant. For notation simplicity, we assume the first n_1 plants were harvested and their leaf areas were measured. For $1 \le i \le n_1$, denote by $y_i(t_i)$ the measured leaf area of the ith plant collected at time t_i. The destructive sampled time points t_is range from the early plant growth stage to the late stage.

To recover the full trajectory of the leaf area of any single plant in the study, we propose a statistical model to predict the plant leaf area with the non-destructive high-throughput multiview RGB images. The number of plant leaf region pixels from the RGB images is essential for the leaf area prediction. To extract this information, we use the enhanced green index $2G/(R+B)$, where R, G, and B represent red, green, and blue intensities of plant images at the pixel level, respectively [13]. Suppose that there are n plants in the study and plant images were taken in p different views with resolution $m_1 \times m_2$. Let $v_{i,j,k}(t)$ be the enhanced green index of the kth pixel of the ith plant image taken from view j at time t. The larger the enhanced green index of a pixel is, the more likely it is within the plant leaf region. The number of leaf region pixels can be appropriately estimated by the following threshold-based statistics:

$$x_{i,j}(t;c) = \sum_{k=1}^{m_1 m_2} \mathbf{1}_{\{v_{i,j,k}(t)>c\}}, \quad 1 \le j \le p, \tag{1}$$

where c is the threshold parameter. Indeed, those pixels with the enhanced green index above c will be counted as leaf region pixels.

By intuition, plant leaf area is related to the numbers of leaf region pixels from the multiview images. More specifically, it is positively correlated with the leaf pixel number from a single view, given those from all other views being the same. Hence, the expected plant leaf area can be modeled as a linear function of leaf pixel numbers from multiview images with non-negative coefficients. Specifically, we write the ith plant leaf area at time t as

$$y_i(t) = \sum_{j=1}^{p} \beta_j x_{i,j}(t;c) + \epsilon_i(t), \tag{2}$$

where $\beta_j \ge 0, 1 \le j \le p$ and $\epsilon_i(t)$s are independent stochastic process.

Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$. Since multiview images taken for a single plant are highly correlated, model parameters $\boldsymbol{\beta}$ would be poorly estimated with classical least-squares methods. We apply the penalized least squares method with the Lasso penalty for parameter estimation. Specifically, we will solve the following optimization problem:

$$(\hat{\boldsymbol{\beta}}, \hat{c}) = \arg\min_{\boldsymbol{\beta} \ge \mathbf{0}, c \in \mathbb{R}} \frac{1}{2} \|\mathbf{y} - \mathbf{X}(c)\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1, \tag{3}$$

where $\mathbf{y} = (y_1(t_1), \ldots, y_{n_1}(t_{n_1}))^\top$, $\mathbf{X}(c) = (x_{ij}(t_i;c))$ is the $n_1 \times p$ design matrix, λ is the complexity parameter, and $||\cdot||_2$ and $||\cdot||_1$ denote ℓ_2 and ℓ_1 norm, respectively.

Note that given c, the above optimization is a Lasso regression with the non-negative coefficient constraint, which can be solved by developing the alternating direction method of multipliers (ADMM) algorithm [6]. Specifically, given c, the $\boldsymbol{\beta}$ optimization in (3) can

Table 1. Algorithm: Parameter estimation based on ADMM algorithm.

- Set a dense discrete $c_1, \ldots, c_N$.
- For each given $c = c_i$, estimate $\boldsymbol{\beta}$ via the ADMM algorithm and denote by $\hat{\boldsymbol{\beta}}_{(i)}$. Let $\mathbf{I}$ be the identity matrix and $(x)_+ := x1_{\{x\geq 0\}}$, $\forall x \in \mathbb{R}$. For the kth iteration,
  - $\boldsymbol{\beta}^{(k+1)} = (\mathbf{X}(c)^\top \mathbf{X}(c) + \rho \mathbf{I})^{-1}(\mathbf{X}(c)^\top \mathbf{y} - \boldsymbol{\gamma}^{(k)} + \rho \boldsymbol{\kappa}^{(k)})$.
  - $\kappa_j^{(k+1)} = \left(\gamma_j^{(k)}/\rho + \beta_j^{(k+1)} - \lambda_i/\rho\right)_+$, $j = 1, ..., p$.
  - $\boldsymbol{\gamma}^{(k+1)} = \boldsymbol{\gamma}^{(k)} + \rho(\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\kappa}^{(k+1)})$.
- $\hat{c} = c_\ell$ and $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{(\ell)}$, where $\ell = \arg\min_{i=1,2,\ldots,N} \frac{1}{2}\|\mathbf{y} - \mathbf{X}(c_i)\hat{\boldsymbol{\beta}}_{(i)}\|_2^2 + \lambda_i \|\hat{\boldsymbol{\beta}}_{(i)}\|_1$, where λ_i is obtained by cross validation.

be solved by a standard ADMM algorithm by introducing a new variable $\boldsymbol{\kappa} \in \mathbb{R}^p$, that is

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\kappa}}) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p,\, \boldsymbol{\kappa}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{X}(c)\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\kappa}\|_1 + I_+(\boldsymbol{\kappa}) :\ \boldsymbol{\beta} = \boldsymbol{\kappa} \right\}, \tag{4}$$

where $I_+(\cdot)$ is defined as, for any $\boldsymbol{\kappa} \in \mathbb{R}^p$,

$$I_+(\boldsymbol{\kappa}) = \begin{cases} 0, & \text{if } \boldsymbol{\kappa} \geq \mathbf{0}, \\ \infty, & \text{otherwise.} \end{cases}$$

To solve (4), we can minimize the augmented Lagrangian function given a penalty parameter $\rho \geq 0$,

$$\begin{aligned} L_\rho(\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\gamma}) =& \frac{1}{2}\|\mathbf{y} - \mathbf{X}(c)\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\kappa}\|_1 + I_+(\boldsymbol{\kappa}) \\ &+ \boldsymbol{\gamma}^\top(\boldsymbol{\beta} - \boldsymbol{\kappa}) + \frac{1}{2}\rho\|\boldsymbol{\beta} - \boldsymbol{\kappa}\|_2^2, \end{aligned} \tag{5}$$

where $\boldsymbol{\gamma}$ is the Lagrangian multiplier. The above optimization can be solved by running three-step iterations till convergence:

$$\boldsymbol{\beta} \text{ step}: \quad \boldsymbol{\beta}^{(k+1)} = \arg\min L_\rho\left(\boldsymbol{\beta}, \boldsymbol{\kappa}^{(k)}, \boldsymbol{\gamma}^{(k)}\right), \tag{6}$$

$$\boldsymbol{\kappa} \text{ step}: \quad \boldsymbol{\kappa}^{(k+1)} = \arg\min L_\rho\left(\boldsymbol{\beta}^{(k+1)}, \boldsymbol{\kappa}, \boldsymbol{\gamma}^{(k)}\right), \tag{7}$$

$$\boldsymbol{\gamma} \text{ step}: \quad \boldsymbol{\gamma}^{(k+1)} = \boldsymbol{\gamma}^{(k)} + \rho\left(\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\kappa}^{(k+1)}\right). \tag{8}$$

Moreover, both $\boldsymbol{\beta}$ step and $\boldsymbol{\kappa}$ step in (6) and (7) have the closed-form solutions, which will be given in Table 1. Hence, our proposed algorithm is efficient.

To sum up, we first pick a dense discrete set $c_1, \ldots, c_N$. Given each c_i, we estimate $\boldsymbol{\beta}$ with the above ADMM algorithm, where the tuning parameter λ is obtained by cross-validation. Then, $\hat{c}$, the estimate of c, will be the c_i that leads to the smallest value of the cost function, and $\hat{\boldsymbol{\beta}}$ will be the corresponding estimator obtained from the ADMM algorithm. The details of the algorithm are given in Table 1.

Suppose that the ith plant images were taken at ℓ_i time points, i.e. $t_{ik},\ k = 1, 2, .., \ell_i$. The ith plant leaf area at time t_{ik} can be predicted by

$$\hat{y}_i(t_{ik}) = \sum_{j=1}^{p} \hat{\beta}_j x_{i,j}(t_{ik}; \hat{c}). \tag{9}$$

Then, the full trajectory $y_i(t),\ t \in [0, T]$ of the ith plant leaf area is estimated using penalized smoothing spline [14,32], where the function $\hat{g}_i$ is a natural cubic spline with knots at

$t_{i1}, t_{i2}, \cdots, t_{i\ell_i}$ that minimize

$$C(g_i) := \sum_{k=1}^{\ell_i} (\hat{y}_i(t_{ik}) - g_i(t_{ik}))^2 + \alpha \int g_i''(u)^2 \mathrm{d}u \tag{10}$$

and α is a nonnegative smoothing parameter which can be selected by cross-validation.

Remark 2.1: For greenhouse plant images, since the background of images is homogeneous, the proposed predictive model works well and fast. In more general applied scenarios, such as plant images in the fields, semantic segmentation is commonly used to extract the plant features from images [25]. However, to conduct the semantic segmentation, it needs a large amount of labeled data and the annotation cost is particularly expensive as it requires pixel-wise labeling. Thus we recommend using the proposed method when the background of images is homogeneous, and applying semantic segmentation methods for plants images with more complicated background.

2.2. Adaptive change detection method

After obtaining the predicted plant leaf area $\hat{g}(t)$ from (9, 10), we are ready to present our monitoring and adaptive change detection method for each plant.

Specifically, for every plant trajectory $\hat{g}(t)$, we take N discrete equispaced time $t_k = kT/N, k = 1, 2, \ldots, N$, over the interval $[0, T]$ and denote the corresponding fitted leaf area measurement for the ith plant as $\hat{g}_{i,k} = \hat{g}_i(t_k)$. Note the number of samples N is determined by how often we can take the image data $x_{i,j}$.

To reduce the temporal correlation and increase the detection efficiency, we propose to monitor the standardized relative change to detect the potential change on the leaf area. For the ith plant at time index k, we denote the relative change as

$$r_{i,k} = \frac{\hat{g}_{i,k} - \hat{g}_{i,k-1}}{\hat{g}_{i,k-1}}, \tag{11}$$

where $k > 1$ and let $r_{i,1} = 0$. The relative change $r_{i,k}$ is one version of relative growth rates defined by researchers in plant science, which has reduced temporal correlation and has been widely used in plant growth analysis [30]. Since the mean and variance of the relative change process $r_{i,k}$ vary with the time index k, we further standardize $r_{i,k}$ by

$$z_{i,k} = \frac{r_{i,k} - \bar{r}_k}{s_{r_k}}, \tag{12}$$

where $\bar{r}_k$ and s_{r_k} are the sample mean and sample standard deviation of relative change at time index k for all plants in the control group.

After standardizing those relative changes, we assume those $z_{i,k}$'s are i.i.d with standard normal distribution $N(0, 1)$ for the control samples and have some unknown mean shifts for those drought samples, i.e. with normal distribution $N(\mu, 1)$. Here μ quantifies the magnitude of the change for the drought group compared with the control group. Furthermore, the time when the mean shift occurs is unknown and may vary from plant to plant.

If we had known the exact post-change mean μ, we would essentially face the problem of testing the hypotheses in the change-point model where $z_{i,1}, \ldots, z_{i,\nu-1}$ are i.i.d. $f_0(x) = \mathrm{pdf}$

of $N(0,1)$ and $z_{i,\nu},\ldots,z_{i,n}$ are i.i.d. $f_1(x)=$ pdf of $N(\mu,1)$. At each time k, we repeatedly test the null hypothesis $H_0:\nu=\infty$ (no change) against the alternative hypothesis $H_1:\nu=1,2,\ldots$ (a change occurs at some finite time ν). Thus the log-likelihood ratio statistic at time k becomes

$$W_{i,k}^{*}=\max_{1\le\nu\le k}\log\frac{\prod_{\ell=1}^{\nu}f_0(z_{i,\ell})\prod_{\ell=\nu+1}^{k}f_1(z_{i,\ell})}{\prod_{\ell=1}^{k}f_0(z_{i,\ell})}, \tag{13}$$

which can be recursively computed as

$$W_{i,k}^{*}=\max\left(W_{i,k-1}^{*}+\mu z_{i,k}-\mu^2/2,0\right), \tag{14}$$

for $k=1,\ 2,\ \ldots,$ with the initial value $W_{i,0}^{*}=0$. In the literature, the statistic $W_{i,k}^{*}$ in (14) was first defined by [27] and is called cumulative sum (CUSUM) statistics. The CUSUM procedure is then defined as the first time when the CUSUM statistic exceeds some predefined threshold b: that is, the CUSUM procedure is given by

$$\tau_i^{*}(b)=\inf\{k\ge 1: W_{i,k}^{*}\ge b\}, \tag{15}$$

which enjoys theoretical optimality [22,26].

However, in our context of plant leaf area monitoring, we do not know the exact value of the post-change mean μ except that $\mu\le-\delta$, where δ is the minimal magnitude of the change that we want to detect. Thus we cannot use the CUSUM $W_{i,n}^{*}$ in (14) directly. One natural idea is to estimate the post-change mean of the ith plant μ_i from observed data, and then plug in the estimated $\hat{\mu}_i$ into the CUSUM statistics in (14). For that purpose, at time k, denote by $\hat{\nu}_k$ the largest $\ell\le k-1$ such that $W_{i,\ell}^{*}=0$. Then the generalized likelihood ratio properties suggest that $\hat{\nu}_k$ is actually the maximum likelihood estimate of the change-point ν at time k, and thus one would expect that the data between time $[\hat{\nu}_k,k]$ would be likely from the post-change distributions, which allows us to provide a reasonable estimate of the post-change mean $\hat{\mu}_i$ at time k. This idea was first rigorously investigated in [23] for detecting positive mean shifts of normal distributions, and here we aim to detect the negative mean shifts. Specifically, at time k, for the ith standardized relative change $z_{i,k}$'s, we define $S_{i,k}$, $T_{i,k}$ by

$$T_{i,k}=k-\hat{\nu}_k,\quad S_{i,k}=\sum_{j=\hat{\nu}_k}^{k-1}z_{i,j}. \tag{16}$$

Then, the estimate of the post-change mean of $z_{i,k}$ can be written by

$$\hat{\mu}_{i,k}=\min\left(\frac{S_{i,k}+s}{T_{i,k}+t},\ -\delta\right)<0, \tag{17}$$

Here $t>0$, $s<0$ are prespecified constants, and s/t can be thought of as a prior estimate of the post-change mean.

By plugging the adaptive estimations $\hat{\mu}_{i,k}$ of the post-change mean μ in the CUSUM statistics in (14), we can derive the adaptive CUSUM statistics by

$$W_{i,k} = \max\left(W_{i,k-1} + \hat{\mu}_{i,k} z_{i,k} - \hat{\mu}_{i,k}^2/2,\ 0\right), \tag{18}$$

where $W_{i,0} = 0$. Based on the adaptive CUSUM statistics, our proposed monitoring procedure can be defined by

$$\tau_i(b) = \inf\{k \geq 1 : W_{i,k} \geq b\}, \tag{19}$$

where b is a constant that can control the false alarm rate.

From the computation viewpoint, the adaptive CUSUM procedure can be implemented in real time due to its recursive form. That is, based on the definition of $(S_{i,k}, T_{i,k})$ in (16), they can be computed recursively over k by

$$\begin{pmatrix} S_{i,k} \\ T_{i,k} \end{pmatrix} = \begin{cases} \begin{pmatrix} S_{i,k-1} + z_{i,k} \\ T_{i,k-1} + 1 \end{pmatrix} & \text{if } W_{i,k-1} > 0, \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \text{if } W_{i,k-1} = 0. \end{cases}$$

Remark 2.2: We should mention that our proposed two-step monitoring framework in HTPP is very flexible. Besides of the adaptive CUSUM procedure, many classical statistical process control (SPC) methods can be used in the second step of our proposed framework. Here, we introduced the adaptive CUSUM as an example and will show it outperforms some classical methods such as Shewhart chart, EWMA chart, standard CUSUM in a case study in the following section. However, as shown in [22,23], the optimality of the adaptive CUSUM requires the i.i.d assumption, which may not hold in practice. Thus some advanced monitoring procedures that can handle the non i.i.d data, e.g. [19,20,31], may also be used in the second step of our proposed framework.

3. Case study

In this section, we will apply our proposed method to the soybean plant data collected at the Nebraska Innovation Campus Greenhouse, High-Throughput Plant Phenotyping Core Facilities (Scanalyzer 3D, LemnaTec Gmbh, Aachen, Germany), University of Nebraska-Lincoln. In Section 3.1, we will introduce the detailed experimental design and the data description. The plant leaf area prediction and detection results will be presented in Section 3.2. We compare our proposed method with exact CUSUM procedures and the classical Shewhart chart.

3.1. Experimental design and data description

The experiment was conducted in the Nebraska Innovation Campus Greenhouse, High-Throughput Plant Phenotyping Core Facilities (Scanalyzer 3D, LemnaTec Gmbh, Aachen, Germany), University of Nebraska-Lincoln. One hundred soybean plants were sown on July 3, 2019, among which 50 were randomly assigned to the drought treatment group and

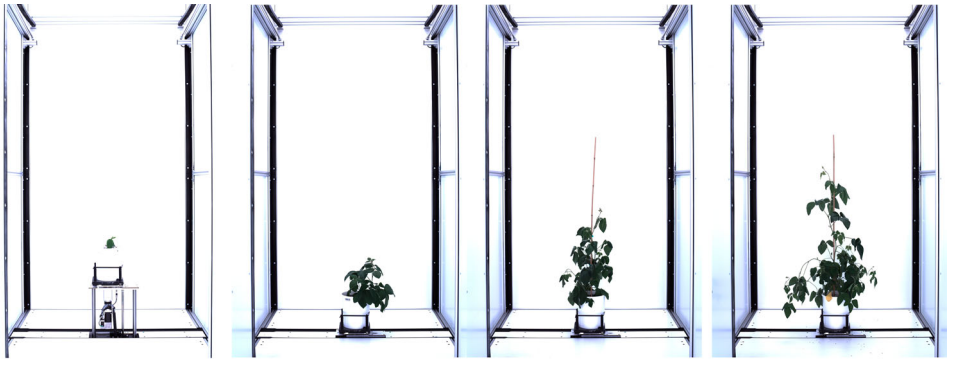


Figure 1. A sequence of side view images of one drought soybean plant taken on July 18, August 7, August 23, and September 8, respectively.

the other 50 were assigned to the control group. Between July 3 and July 17, a uniform watering rate was applied to both groups to ensure all soybean plants were established well before the water treatments. From July 18, the pots were transferred onto automated conveyor belts, with an automated weighing and watering station where the change in pot weight as a result of water evaporation and transpiration can be quantified, and prescribed amounts of water can be precisely applied to the desired level. For the control group, water was added to the pots to a targeted weight of 7 kg (representing 80% field capacity), while water was added to the pots daily to a targeted weight of 5 kg (representing 40% field capacity) for the drought group.

From July 18, 2019, to September 8, 2019, pots were moved into the imaging chamber with RGB camera (maker: Basler) every 2 days. Multiview images were taken at 10 different angles (9 side views at 0, 36, 72, 108, 144, 216, 252, 288, and 324 degrees, and 1 top view at 90 degrees) with side view resolution 6576×4384 and top view resolution 4384×6576 (see Figure 1).

At seven time points during the experiment, a total of 66 plants were destructively sampled (July 29, August 2, August 9, August 14, August 23, August 30, and September 6). Six plants were sampled at the first date, and 10 plants were sampled at subsequent sampling points. For each destructed soybean plant, all plant material above soil was harvested and fractionated into leaves, stems, and pods, and leaf area was measured by a leaf area meter (LI-3100C, LI-COR Biosciences, Lincoln, Nebraska, USA). Thus, after destructive sampling, these plants were destroyed and no further images of those plants can be taken.

At the late stage of the experiment, some branches of soybean plants were tied to the stake in the middle to avoid problematic movement on the conveyor belt because the plants got bushy. Correspondingly, the projected areas of plant images taken after tying decreased. But this did not have an impact on our experiment, since plant leaf growth started to slow down much earlier than the date of tying the branches, which occurred on August 21.

3.2. Results

In this section, the proposed methods in Section 2 are applied to detect when the leaf area growth of the individual soybean plant would slow down due to drought stress.

We trained the predictive model (2) for plant leaf area with the plant RGB images and observed leaf areas. One thing to note is, at the late stage of the experiment, the branches of soybean plants were tied to the stake in the middle to avoid problematic movement on the conveyor belt because the plants got bushy. Correspondingly, the projected areas of plant images taken after tying decreased. To this end, we separated the 66 destructed plants into two groups. The first 36 plants without tying were used to fit model (2) to predict the plant leaf area growth in the early and middle stages of the experiment, while the remaining plants were used to fit the same model to do prediction in the late stage. The detected change in this study occurred about 20 days earlier than the date of tying the branches, which was August 21. Thus, to avoid the tying effect, we only used the first 36 destructed plants to fit model (2). We used 25 of them for training and the other 11 for testing. The tuning parameter λ was selected by cross-validation and the predicted R^2 in the testing set is 0.93. We predict plant leaf areas at every time point when plant images were taken before August 21 using (9) and further obtained the predicted leaf area of each individual soybean plant at the semi-day frequency with the penalized smoothing spline method.

Figure 2 (a) shows the smoothed mean growth curves of the leaf area in the control group and the drought group, while Figure 2(b) shows the mean difference of the standardized relative change between the control and the drought treatment. Both results show that the mean leaf areas of the soybean plants in the control group are significantly different from those in the drought group, and the difference appears around 13 days after July 18. However, this argument is established at the population level. Later on, we will use our proposed change detection method to provide such information at the individual plant level. We also observe a higher standardized relative change in the drought group around 25 days after July 18. One potential reason is that the well-watered plants might have passed their fast growth period and started to grow toward their maximum sizes, the plant growth would slow down. Correspondingly, the standardized relative changes z would decrease. However, the plants under drought stress need longer time to grow up. So compared to the well-watered plants, they might have higher standardized relative changes in some periods. Yet, since it occurred about 12 days after the mean change was detected, it would not affect the validity of our detection method.

To detect when the leaf area of the individual plant under drought treatment becomes different from control, we monitor the standardized relative change as in (12) obtained from our predictive model. Based on it, we will compare the following four methods:

- Our proposed adaptive CUSUM method in (19), which is denoted by $T_a(b)$.
- The exact CUSUM method in (15), with some specific choices of post-change mean μ. We denote this method by $T_c(b, \mu)$.
- The one-sided EWMA chart ([34]), which is denoted by $T_{ewma}(b) := \inf\{k \geq 1, : -M_k \geq b \times \sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2k}}\}$, where $M_k = \lambda z_k + (1-\lambda)M_{k-1}$, λ is the chosen weight.
- The one-sided Shewhart chart [38], which is denoted by $T_{sc}(b) := \inf\{k \geq 1, : -z_k \geq b\}$.

Note for our proposed method $T_a(b)$, the tuning parameter δ, which represents the minimal magnitude of the change, can be selected by the practitioners based on the domain

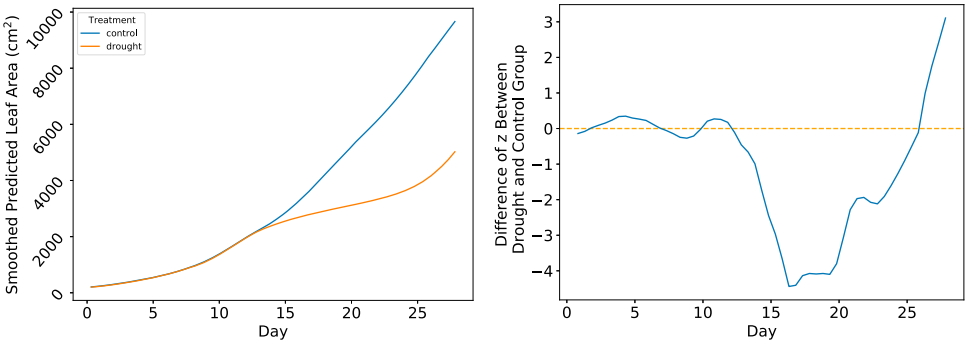


Figure 2. (a) The smoothed mean growth curves of the leaf area in the control and the drought treatment group; (b) the mean difference of the standardized relative change between the control and the drought treatment group. *Here, day 0 corresponds to July 18, the first day when plant images were taken.*

Table 2. A comparison of the average detection delays of the four methods using the standardized relative change with in-control average run length as 60 based on 10, 000 repetitions in Monte Carlo simulation.

Method	Specific	Detection delay
Adaptive CUSUM	$T_a(b = 4.15)$	11.39 (0.58)
CUSUM	$T_c(b = 4.39, \mu = -1)$	11.73 (0.57)
	$T_c(b = 5.37, \mu = -2)$	12.34 (0.56)
One-sided EWMA	$T_{ewma}(b = 2.56, \lambda = 0.1)$	14.88 (1.40)
	$T_{ewma}(b = 2.75, \lambda = 0.2)$	13.27 (1.31)
One-sided Schewhart chart	$T_{sc}(b = 3.42)$	15.26 (0.51)

Note: The standard error of detection delays are reported in the bracket.

knowledge. Here, we simply set $\delta = 1$ and show the results for illustration. Furthermore, we also set $t = 1, s = -1$ for our method. And we set λ =0.1 and 0.2 for the EWMA chart.

To evaluate the performance of the four methods, for each value of threshold b, we conduct 10, 000 Monte Carlo simulations to simulate the average run length by random sampling with replacement from the control group. Then, we search for the appropriate value of the threshold b such that the average run length to the false alarm of each procedure equals to 60 in the control group ($N = 60$), so that the false alarm rate keeps the same across the four methods. Then, using the obtained b for each method, we identify the time when the procedure raises an alarm for each of the 50 plants in the drought group. Since we do not know the true change time for each plant, we consider a worst-case scenario when the true change occurs at the beginning of the time series in the drought group. Thus the time when the monitoring procedure raises an alarm can be thought of as the worst-case detection delay. To compare the performance of the four methods, the average and standard error of the worst-case scenario detection delays of the 50 plants in the drought group are summarized in Table 2.

Table 2 demonstrates that our proposed method has the smallest detection delay compared with other methods. While it is not surprising that our adaptive CUSUM procedure, the exact CUSUM procedures, and the EWMA chart are faster than the classical Shewhart chart method due to the cumulative properties, we can observe an interesting phenomenon

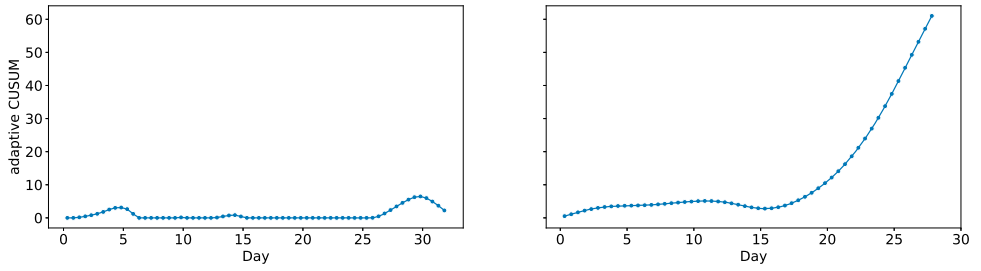


Figure 3. Left: Adaptive CUSUM statistics of one plant from the control group; Right: Adaptive CUSUM statistics of one plant from the drought group.

that our proposed method is also better than the exact CUSUM. One possible explanation is that with our proposed method, the unknown true post-change parameter μ is estimated adaptively from the observed data, while the preset μ in the exact CUSUM method could be away from the true value. Thus our proposed method yields better results. In addition, by comparing the two exact CUSUM results, we found that the average detection delay of the one with the smaller post-change ($\mu = -1$) is much closer to the proposed adaptive CUSUM procedure, indicating that the absolute magnitude of the actual change is likely smaller than 1.

It is important to mention that the detection delays reported in Table 2 indeed represent the worst-case scenario of the delay, where the change happened at the initial time. However, in practice, individual drought plants may start to change at a different unknown time. Thus it would be very challenging to get the actual average detection delay for all plants in the drought group.

To have another illustration of our proposed method, in Figure 3, we show our adaptive CUSUM statistics for one plant from the control group and one from the treatment group. The statistics are very small for the plant in the control group, but become very large after day 15 for the plant under drought stress. This phenomenon implies the adaptive CUSUM statistics are good indicators for the change.

4. Conclusion

In this paper, we developed a framework of monitoring and quickest detection of the individual plant leaf area via real-time imaging data. An efficient ADMM-based algorithm was developed to extract the plant leaf area from the multiview RGB images and the adaptive CUSUM procedure was then proposed to detect the individual plant leaf area change in real time. To the best of our knowledge, efficient detection methods have not been established in the area of high-throughput plant phenotyping. We validated the proposed framework with multiview RGB images of 100 soybean plants, which were collected almost every 2 days from the greenhouse at the University of Nebraska Lincoln in about 2 months. With our proposed online image detection framework, the monitoring and detection can be implemented in real time without keeping all past images in storage. The result showed our method had the smallest detection delay compared with other methods. Further, although the proposed framework was illustrated based on plant leaf area, it can be generalized to monitoring and detection of other morphological, physiological, and

biophysical plant traits or indices, such as height, width, size, biomass, water content, normalized difference vegetation index (NDVI), photochemical reflectance index (PRI), and other well-established VIs extracted from hyperspectral images or other types of images [2,12,24,28,37]. Our results contribute to the solutions that aim at addressing the bottleneck in HTPP research and will allow the development of tools for fast and efficient linkages between phenomics traits and genomics expression to speed up the selection of genotypes and improve the decision-making process of plant breeders and plant scientists. Besides, the real-time monitoring procedure is helpful for crop management in precision agriculture, such as irrigation and disease control.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by USDA -NIFA [2020-68013-32371].

References

[1] J. Adams, Y. Qiu, Y. Xu, and J.C. Schnable, *Plant segmentation by supervised machine learning methods*, Plant Phenome J. 3 (2020), pp. e20001.
[2] M.S.M. Asaari, S. Mertens, S. Dhondt, D. Inzé, N. Wuyts, and P. Scheunders, *Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform*, Comput. Electron. Agric. 162 (2019), pp. 749–758.
[3] S. Bashyam, S. Das Choudhury, A. Samal, and T. Awada, *Visual growth tracking for automated leaf stage monitoring based on image sequence analysis*, Remote. Sens. (Basel) 13 (2021), pp. 961.
[4] J. Behmann, J. Steinrücken, and L. Plümer, *Detection of early plant stress responses in hyperspectral images*, ISPRS. J. Photogramm. Remote. Sens. 93 (2014), pp. 98–111.
[5] M. Bosc, F. Heitz, J.P. Armspach, I. Namer, D. Gounot, and L. Rumbach, *Automatic change detection in multimodal serial mri: application to multiple sclerosis lesion evolution*, NeuroImage 20 (2003), pp. 643–656.
[6] S. Boyd, N. Parikh, and E. Chu, *Distributed Optimization and Statistical Learning Via the Alternating Direction Method of Multipliers*, Now Publishers Inc, 2011.
[7] A. Chawade, J. van Ham, H. Blomquist, O. Bagge, E. Alexandersson, and R. Ortiz, *High-throughput field-phenotyping tools for plant breeding and precision agriculture*, Agronomy 9 (2019), pp. 258.
[8] S. Das Choudhury, A. Samal, and T. Awada, *Leveraging image analysis for high-throughput plant phenotyping*, Front. Plant. Sci. 10 (2019), pp. 508.
[9] N. Fahlgren, M.A. Gehan, and I. Baxter, *Lights, camera, action: high-throughput plant phenotyping is ready for a close-up*, Curr. Opin. Plant. Biol. 24 (2015), pp. 93–99.
[10] A.R. Fernie and J. Gutierrez-Marcos, *From genome to phenome: genome-wide association studies and other approaches that bridge the genotype to phenotype gap*, Plant. J. 97 (2019), pp. 5–7.
[11] R.T. Furbank and M. Tester, *Phenomics–technologies to relieve the phenotyping bottleneck*, Trends. Plant. Sci. 16 (2011), pp. 635–644.
[12] J. Gamon, J. Penuelas, and C. Field, *A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency*, Remote. Sens. Environ. 41 (1992), pp. 35–44.
[13] Y. Ge, G. Bai, V. Stoerger, and J.C. Schnable, *Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput rgb and hyperspectral imaging*, Comput. Electron. Agric. 127 (2016), pp. 625–632.

[14] T. Hastie, R. Tibshirani, J.H. Friedman, and J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2, Springer, 2009.
[15] N. Itoh and J. Kurths, *Change-point detection of climate time series by nonparametric method*, in *Proceedings of the World Congress on Engineering and Computer Science*, Vol. 1. Citeseer, 2010, pp. 445–448
[16] T.L. Lai, *Sequential changepoint detection in quality control and dynamical systems*, J. R. Stat. Soc. Ser. B Methodo, 57 (1995), pp. 613–644.
[17] T.L. Lai and H. Xing, *Sequential change-point detection when the pre-and post-change parameters are unknown*, Seq. Anal. 29 (2010), pp. 162–175.
[18] M. Lavielle and G. Teyssiere, *Adaptive detection of multiple change-points in asset price volatility*, in *Long Memory in Economics*, Springer, 2007, pp. 129–156
[19] J. Li, *Nonparametric adaptive cusum chart for detecting arbitrary distributional changes*, J. Qual. Technol. 53 (2021), pp. 154–172.
[20] J. Li and P. Qiu, *Nonparametric dynamic screening system for monitoring correlated longitudinal data*, IIE Trans. 48 (2016), pp. 772–786.
[21] K. Liu, R. Zhang, and Y. Mei, *Scalable sum-shrinkage schemes for distributed monitoring large-scale data streams*, Stat. Sin. 29 (2019), pp. 1–22.
[22] G. Lorden, *Procedures for reacting to a change in distribution*, Ann. Math. Stat. 42 (1971), pp. 1897–1908.
[23] G. Lorden and M. Pollak, *Sequential change-point detection procedures that are nearly optimal and computationally simple*, Seq. Anal. 27 (2008), pp. 476–512.
[24] A. Mazis, S.D. Choudhury, P.B. Morgan, V. Stoerger, J. Hiller, Y. Ge, and T. Awada, *Application of high-throughput plant phenotyping for assessing biophysical traits and drought response in two oak species under controlled environment*, For. Ecol. Manage. 465 (2020), pp. 118101.
[25] A. Milioto, P. Lottes, and C. Stachniss, *Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs*, in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2229–2235
[26] G.V. Moustakides, *Optimal stopping times for detecting changes in distributions*, Ann. Stat. 14 (1986), pp. 1379–1387.
[27] E.S. Page, *Continuous inspection schemes*, Biometrika 41 (1954), pp. 100–115.
[28] P. Pandey, Y. Ge, V. Stoerger, and J.C. Schnable, *High throughput in vivo analysis of plant leaf chemical properties using hyperspectral imaging*, Front. Plant. Sci. 8 (2017), pp. 1348.
[29] M. Pollak, *Optimal detection of a change in distribution*, Ann. Stat. 13 (1985), pp. 206–227.
[30] A. Pommerening and A. Muszta, *Relative plant growth revisited: towards a mathematical standardisation of separate approaches*, Ecol. Modell. 320 (2016), pp. 383–392.
[31] P. Qiu, *Big data? statistical process control can help!*, Am. Stat. 74 (2020), pp. 329–344.
[32] J. Ramsay, N. Heckman, and B. Silverman, *Spline smoothing with model-based penalties*, Behav. Res. Methods. Instrum. Comput. 29 (1997), pp. 99–106.
[33] D.K. Ray, N.D. Mueller, P.C. West, and J.A. Foley, *Yield trends are insufficient to double global crop production by 2050*, PLoS. ONE. 8 (2013), pp. e66428.
[34] S. Roberts, *Control chart tests based on geometric moving averages*, Technometrics 1 (1959), pp. 239–250.
[35] S. Roberts, *A comparison of some control chart procedures*, Technometrics 8 (1966), pp. 411–430.
[36] C. Römer, M. Wahabzada, A. Ballvora, F. Pinto, M. Rossini, C. Panigada, J. Behmann, J. Léon, C. Thurau, C. Bauckhage, and K. Kersting, *Early drought stress detection in cereals: simplex volume maximisation for hyperspectral image analysis*, Funct. Plant. Biol. 39 (2012), pp. 878–890.
[37] J. Rouse, R. Haas, J. Schell, and D. Deering, *Monitoring vegetation systems in the great plains with erts*, NASA Special Publication 351 (1974), pp. 309.
[38] W.A. Shewhart, *Economic Control of Quality of Manufactured Product*, Macmillan and Co Ltd, London, 1931.
[39] A.N. Shiryaev, *On optimum methods in quickest detection problems*, Theory Probab. its Appl. 8 (1963), pp. 22–46.
[40] A. Singh, B. Ganapathysubramanian, A.K. Singh, and S. Sarkar, *Machine learning for high-throughput stress phenotyping in plants*, Trends. Plant Sci. 21 (2016), pp. 110–124.

[41] R. Wang, Y. Qiu, Y. Zhou, Z. Liang, and J.C. Schnable, *A high-throughput phenotyping pipeline for image processing and functional growth curve analysis*, Plant Phenomics 2020 (2020), pp. 8. doi:10.34133/2020/7481687.
[42] Y. Xu, Y. Qiu, and J.C. Schnable, *Functional modeling of plant growth dynamics*, Plant Phenome. J. 1 (2018), pp. 1–10.