Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Note

Utilising variable sorting for normalisation to correct illumination effects in close-range spectral images of potato plants

Puneet Mishra [a,*], *Gerrit Polder* [a], *Aoife Gowen* [b], *Douglas N. Rutledge* [c,d], *Jean-Michel Roger* [e,f]

[a] *Agro-food Robotics, Wageningen University and Research, P.O. Box 17, 6700, AA, Wageningen, the Netherlands*
[b] *UCD School of Biosystems and Food Engineering, University College Dublin, Ireland*
[c] *Université Paris-Saclay, INRAE, AgroParisTech, UMR SayFood, 75005, Paris, France*
[d] *National Wine and Grape Industry Centre, Charles Stuart University, Wagga Wagga, Australia*
[e] *ITAP, INRAE, Institut Agro, University Montpellier, Montpellier, France*
[f] *ChemHouse Research Group, Montpellier, France*

ARTICLE INFO

Article history:
Received 19 May 2020
Received in revised form
13 July 2020
Accepted 14 July 2020
Published online 3 August 2020

Keywords:
Illumination
Phenotyping
Spectroscopy
Non-destructive

Visible and near-infrared spectral imaging is a key non-destructive technique for rapid assessment of biophysical traits of plants. A major challenge with close-range spectral imaging of plants is spectral variation arising from illumination effects, which may mask the signals due to physiochemical differences. In the present work, we describe a new scatter correction technique called variable sorting for normalisation (VSN) and compare its efficiency with that of the commonly used standard normal variate (SNV) technique for the removal of unwanted illumination effects. Spectral images of potato plants were used for testing the correction. The results showed that the VSN outperformed SNV in removing illumination effects from the images of plants. The results show that the VSN approach to illumination correction can support high-throughput plant phenotyping where spectral imaging is used as a continuous monitoring tool.

1. Introduction

Visible and near-infrared (VNIR: 400–1000 nm) spectral imaging is a key non-destructive technique for rapid assessment of biophysical traits of plants (Mishra, Asaari et al., 2017; Mishra, Polder et al., 2020). The visible region (400–670 nm) provides access to different pigments, through their absorption of light, which can be used to characterise photosynthetic activity in plants. In addition, the near-infrared region (670–1000 nm) can be related to leaf chemicals and the internal structure of leaves (Mishra, Asaari et al., 2017). However, spectral imaging of plants suffers from illumination effects due to the interaction of the incoming light with the complex plant geometry (Huang, Luo et al., 2018). Illumination

* *Corresponding author.*
E-mail address: puneet.mishra@wur.nl (P. Mishra).
https://doi.org/10.1016/j.biosystemseng.2020.07.010

effects lead to undesirable spectral variability that is unrelated to plant tissue composition. Much effort has been made to correct these effects, such as extracting local information with a 3D point cloud and fusing with the spectral data (Behmann, Mahlein et al., 2016; Huang, Luo et al., 2018), or including local inclination information into the directional hemispherical model of leaf optical properties (Jay, Bendoula et al., 2016; Morel, Jay et al., 2018). These methods have given improved results in laboratory conditions.

Spectral imaging is typically deployed in high-throughput phenotyping setups in a dedicated cabinet where the plants are taken to be imaged during the experiment. In this framework of high throughput spectral imaging, simple pre-processing methods are preferred due to their relative simplicity and speed. Thus, standard normal variate (SNV) has been widely applied to correct for illumination effects. Vigneau, Ecarnot et al. (2011) showed that the correction of the spectral imaging of maize leaves with SNV improved the performance of the regression models for prediction of nitrogen content. The SNV approach was also taken by Asaari, Mishra et al. (2018) and applied to a high-throughput phenotyping experiment with maize plants. The results also showed that SNV was successfully able to reduce linear illumination effects from the interaction of illumination with the maize leaves. SNV was used in the same high-throughput phenotyping platform for extended experiments related to drought stress induction and recovery of the plants (Asaari, Mertens et al., 2019). SNV has also been extended to commercial platforms and used for the detection of drought in Arabidopsis plants (Mishra, Feller, Schmuck, Nicol, & Nordon, 2019) and for testing the effects of chemicals on plants in a high-throughput scenario (Mishra, Schmuck, Roth, Nicol, & Nordon, 2019). The results were conclusive that SNV was useful for illumination correction. SNV has the advantages of being rapid, model-free and not requiring additional measurements compared to other approaches such as 3D modelling (Behmann, Mahlein et al., 2016) and radiative transfer modelling (Jay, Bendoula et al., 2016; Morel, Jay et al., 2018).

Although there are many applications of SNV for illumination correction in close range spectral imaging (SI), the method does have the disadvantage that it assumes that all spectral bands are affected in a similar way by the illumination effects, which is not always the case. Even more importantly, SNV suffers from the closure problem, due to the division of the intensity of all the points in the spectrum by its standard deviation. This can lead to ineffective removal of complex illumination effects such as multiplicative effects of higher order (Roger, Boulet et al., 2020) resulting in a deformation of the general shape of the spectra. To overcome such limitations, a new method has been developed in the chemometric domain called variable sorting for normalisation (VSN) (Rabatel, Marini et al., 2020). VSN works by defining a weighting function to identify the variables that are affected by external effects such as illumination rather than the response of interest. In addition, VSN helps to characterise the nature of such external effects, whether they are additive, multiplicative, polynomial or otherwise. Recent applications related to the use of VSN have shown an improvement in model performance compared to traditional pre-processing techniques (Rabatel, Marini et al., 2020; Sun, Subedi et al., 2020).

The aim of the present work is to evaluate the use of the VSN pre-processing approach for illumination correction in spectral imaging of plants. Aa an example, potato plant was chosen for this study. The comparison was performed with the commonly used SNV method and the raw reflectance signal. In order to show the improvement by using VSN, unsupervised cluster analysis was applied to the spectral images to visualise the segmentation of plant parts i.e. leaf blade and the vein.

2. Material and methods

2.1. Spectral imaging of potato plants

The image (a single hyperspectral image of approximately 4 plants) was selected from an experiment for the early detection of diseased potato plants. The image was captured with a V10e spectral line-scan camera from Specim (Oulu, Finland) with top view as illustrated in Fig. 1. Illumination was provided by a 15 W halogen light source. The camera provided a 500 x 656 x 193 data cube with 2 spatial dimensions, and a spectral dimension including 193 bands acquired in the range of 400–1000 nm, at a resolution of 3 nm (FWHM). For illumination a fibre-optic line light was used, using a tungsten halogen light source. The image was corrected with white (>98% reflectivity Spectralon) and dark reference samples. However, it should be noted that this correction is carried out using a flat white calibration reference, therefore despite this correction, spectral images of plants suffer from unavoidable illumination effects due to plant geometry. Image segmentation to separate leaves from the background was done with normalised difference vegetation index. The images were processed using MATLAB 2017b (Natwick, USA).

2.2. Pre-processing with SNV

SNV includes subtraction of the mean signal intensity from the intensity at all wavelengths and then dividing by the standard deviation of the spectrum. In this way, the mean corrects for the offset effect and the standard deviation corrects global differences in intensities (Roger, Boulet et al., 2020). In the case of SI, the spectral cube is first unfolded and SNV is then applied to each spectrum. The SNV transform can be represented as in equation (1):

$$\boldsymbol{X}_{SNV} = \frac{\boldsymbol{x} - \boldsymbol{X}_{mean}}{\boldsymbol{X}_{std}} \quad (1)$$

where $\boldsymbol{X}_{SNV}$ is the transformed spectra, $\boldsymbol{x}$ is the reflectance (or raw signal corresponding to each wavelength), $\boldsymbol{X}_{mean}$ is the mean intensity and $\boldsymbol{X}_{std}$ is the standard deviation of the intensities.

2.3. Pre-processing with VSN

Variable sorting for normalisation (VSN) is a recently developed scatter correction technique that calculates

weights to be applied to wavelengths when applying classical methods such as SNV, multiple scatter correction (MSC) or Detrend (Rabatel, Marini et al., 2020). VSN estimates the weights based on random consensus (RANSAC) algorithm which estimates to what extent a wavelength is affected by size effects (additive and multiplicative offsets by scattering) or by shape effects (chemical-related features). In this way, variables that are strongly related to chemicals have a low weight and negligible role in the calculation of the size effect.

The main benefit of the VSN approach in comparison to the MSC is that it does not require a reference spectrum to perform the weight estimation. In the present work, VSN was implemented as presented by Rabatel, Marini et al. (2020), which involved using the weight while estimating the mean and standard deviation for the SNV. The 3D spectral data cube was unfolded and the VSN was applied in the spectral domain as can be understood from Fig. 2.

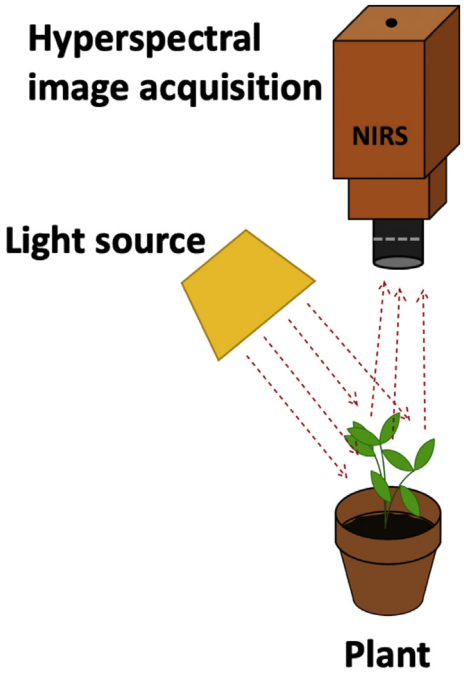

Fig. 1 – A schematic of close range spectral imaging of plants.

2.4. K-means clustering analysis

K-means clustering is used here for data modelling in order to demonstrate the effect of the different pre-processing techniques. K-means is a vector quantisation method which partitions *n* observations into *k* number of clusters in which each observation belongs to the cluster with the closest barycentre. Any number of clusters can be defined; however, it is always necessary to perform an optimisation to find the optimal number of clusters. The cluster number optimisation protocol includes defining a criterion that minimises the within-cluster distances and maximises the between-cluster distances. In the present work, the K-means clustering was implemented using MATLAB's '*Statistics and machine learning toolbox*' and the criterion used to determine the optimal number of clusters was the '*Calinski Harabasz*' index, which is defined as the ratio between the within-cluster dispersion and the between-cluster dispersion.

3. Result and discussion

3.1. Illumination effects in image

A single band from a spectral image of potato plants and corresponding to ~700 nm was taken from the spectral data cube (Fig. 3). The image reveals the presence of illumination effects as the yellowish regions on the leaves. The illumination effects are so dominant in some regions that it masks the differences between the veins and the leaf blade. If any data analysis were performed without removing these effects, the model would therefore not be able to capture the difference between the veins and the leaf blades (Asaari, Mishra et al., 2018; Mishra, Schmuck et al., 2019).

The correction of the illumination effects using VSN pre-processing for the same spectral plane shown in Fig. 3A (~700 nm) is shown in Fig. 3B. VSN identified that the spectral data have multiplicative effects with 3rd order polynomial baseline. Regions of interest from the images in Fig. 3A and B, showing 'zoomed in' regions for specific plant, are shown in Fig. 3C to F. Figure 3B shows that the effects are removed by the VSN pre-processing. After removal of illumination effects, the veins can be distinguished from the leaf blades. Similar effects can be seen when the image is zoomed for the two plants as shown in Fig. 3C to F. In both of the plants, the VSN

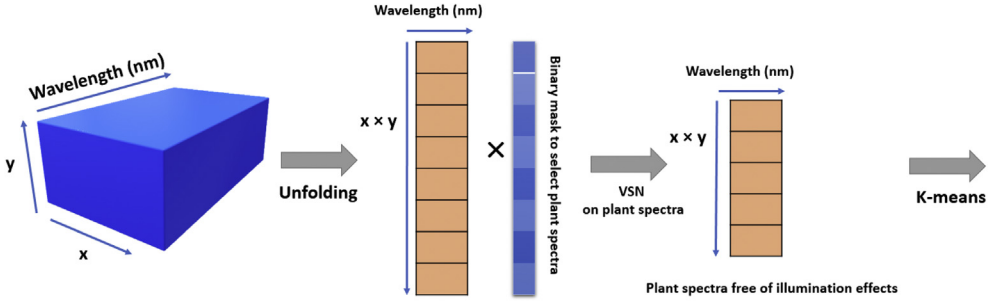

Fig. 2 – Schematic of spectra image unfolding and pre-processing with VSN.

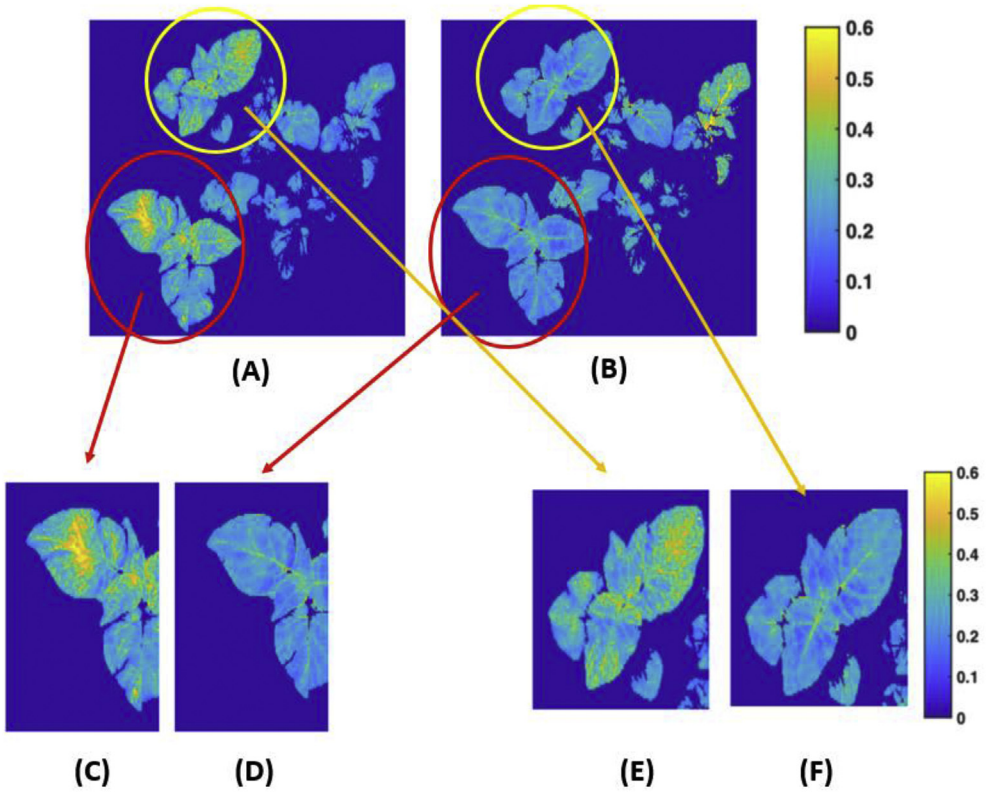

Fig. 3 – Spectral image corresponding to ~700 nm: (A) before any correction, and (B) after VSN correction. The illumination effects have been eliminated. The zoomed plants are (C) plant 1 before correction, (D) plant 1 after correction, (E) plant 2 before correction, and (F) plant 2 after correction.

pre-treatment has led to a reduction of illumination effects and a highlighting of the plant parts.

3.2. *k-means clustering applied to spectral images of plants*

The plant part segmentation results obtained with the k-means clustering are shown in Figs. 4–6. The optimisation for the identification of the optimal number of cluster centroids showed that the raw reflectance gave 4 clusters (Fig. 4A), SNV gave 3 clusters (Fig. 4B) and VSN gave only 2 clusters (Fig. 4C). The clusters are further visualised as cluster maps in Fig. 5. Cluster maps show that the clustering of raw reflectance data (Fig. 5A) only modelled the differences in illumination and does not have any biophysical basis. This is because the regions which were highlighted due to illumination effects in Fig. 3 are identified as belonging to the same cluster. Furthermore, these clusters that were related to illumination effects masked the identification of leaf blade and veins. In the case of SNV, the cluster maps obtained from the three clusters show that there is an improvement in the identification of plant parts compared to the raw reflectance data. This indicates that SNV indeed works for illumination reduction as stated by Vigneau, Ecarnot et al. (2011), Asaari, Mishra et al.

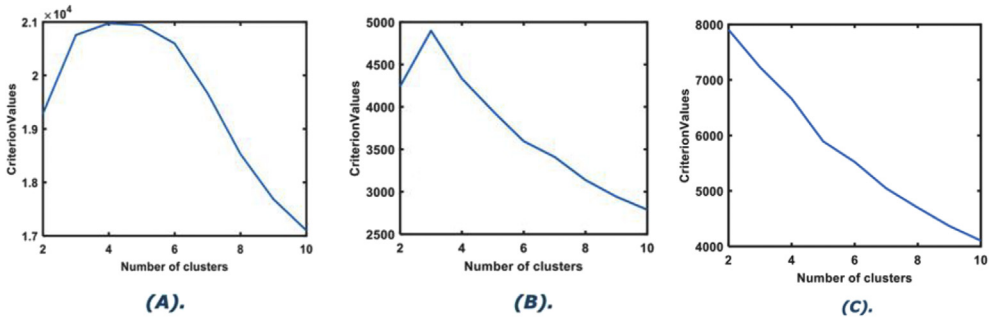

Fig. 4 – The optimisation of K-means clustering and automatic selection of cluster numbers based on 'Calinski Harabasz' index. (A) Reflectance (4 clusters), (B) SNV corrected data (3 clusters) and (C) VSN corrected data (2 clusters).

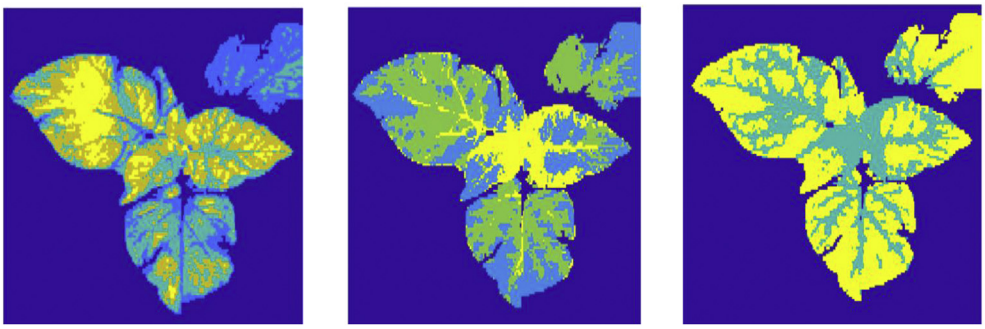

Fig. 5 – The cluster maps obtained from the k-means clustering: (A) Reflectance, (B) SNV pre-processed and (C) VSN pre-processed.

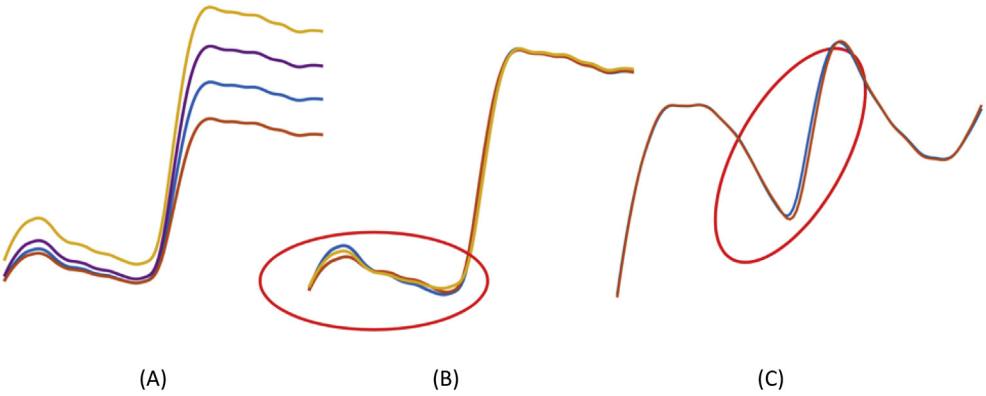

Fig. 6 – The cluster centroids from K-Means: (A) raw reflectance, (B) SNV pre-processed and (C) VSN pre-processed.

(2018), Asaari, Mertens et al. (2019), Mishra, Feller et al. (2019), and Mishra, Schmuck et al. (2019). However, despite the reduction in illumination effects by SNV, the clustering does not result in a clear identification of different plant parts (Fig. 5B); there are still multiple clusters assigned to the same leaves, and the leaf blade and vein parts are found in a single cluster. On the other hand, the cluster maps after VSN showed a clear segmentation of the leaf blade and the vein part (Fig. 5C). In the case of clustering based on the VSN pre-processed data, two clusters were found by the cluster selection criterion, and these two clusters were sufficient to distinguish the leaf blade and veins in the plants. Compared to the cluster maps of raw reflectance, there is a significant improvement in the identification of the veins in the illumination affected parts.

To get an understanding of the cluster maps, the spectra of the cluster centroids are presented in Fig. 6. The cluster centroid shows that the clustering on raw reflectance spectra (Fig. 6A) just captured the differences in global intensities of the pixels. Such differences are the result of illumination differences and do not carry physicochemical information. The clustering on the SNV pre-processed spectra was able to avoid illumination effects and so captured physicochemical information. However, the information was mainly limited to colour differences. Capturing the colour differences by the cluster centroids of SNV shows that the pre-processed data still has its main variability in the colour part of the spectrum, which is masking the chemical differences between the leaf blade and the vein. In the case of the VSN pre-processed data, the clustered centroids have their main differences in the chlorophyll and the red edge parts of the spectrum, both of which are indicators of plant photosynthetic activity. Differences highlighted by VSN in this region can be understood as differences in the photosynthetic activity of the leaf blade and vein, which are at the origin of the separate clusters and the segmentation in the cluster maps.

4. *Conclusion*

Close-range spectral imaging of plants suffers from illumination effects due to interaction of the light with the complex plant geometry. Such illumination effects can mask the underlying physicochemical signals from the various plant parts. Present work recommends the use of spectral normalisation techniques before modelling the close-range spectral images of plants. Spectral normalisation techniques such as SNV and

VSN have the key advantage over 3D modelling and radiative transfer modelling that they are fast, do not require extra sensor measurements and do not require any external parameters. SNV can reduce the global differences in the signal intensities, but, in the present work, it did not allow clear identification of the physicochemical differences in plant parts. On the other hand, VSN was successful in removing illumination effects and a clear segmentation of leaf blade and vein was obtained.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Asaari, M. S. M., Mertens, S., Dhondt, S., Inze, D., Wuyts, N., & Scheunders, P. (2019). Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Computers and Electronics in Agriculture, 162*, 749–758.

Asaari, M. S. M., Mishra, P., Mertens, S., Dhondt, S., Inze, D., Wuyts, N., & Scheunders, P. (2018). Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform. *ISPRS Journal of Photogrammetry and Remote Sensing, 138*, 121–138.

Behmann, J., Mahlein, A. K., Paulus, S., Dupuis, J., Kuhlmann, H., Oerke, E. C., & Plumer, L. (2016). Generation and application of hyperspectral 3D plant models: Methods and challenges. *Machine Vision and Applications, 27*(5), 611–624.

Huang, P. K., Luo, X. W., Jin, J., Wang, L. J., Zhang, L. B., Liu, J., & Zhang, Z. G. (2018). Improving high-throughput phenotyping using fusion of close-range hyperspectral camera and low-cost depth sensor. *Sensors, 18*(8).

Jay, S., Bendoula, R., Hadoux, X., Feret, J. B., & Gorretta, N. (2016). A physically-based model for retrieving foliar biochemistry and leaf orientation using close-range imaging spectroscopy. *Remote Sensing of Environment, 177*, 220–236.

Mishra, P., Asaari, M. S. M., Herrero-Langreo, A., Lohumi, S., Diezma, B., & Scheunders, P. (2017). Close range hyperspectral imaging of plants: A review. *Biosystems Engineering, 164*, 49–67.

Mishra, P., Feller, T., Schmuck, M., Nicol, A., & Nordon, A. (2019a). Early detection of drought stress in Arabidopsis thaliana utilsing a portable hyperspectral imaging setup. In *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)* (pp. 1–5). IEEE.

Mishra, P., Polder, G., & Vilfan, N. (2020). Close range spectral imaging for disease detection in plants using autonomous platforms: A review on recent studies. *Curr Robot Rep*.

Mishra, P., Schmuck, M., Roth, S., Nicol, A., & Nordon, A. (2019b). Homogenising and segmenting hyperspectral images of plants and testing chemicals in a high-throughput plant phenotyping setup. In *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)* (pp. 1–5). IEEE.

Morel, J., Jay, S., Feret, J. B., Bakache, A., Bendoula, R., Carreel, F., & Gorretta, N. (2018). Exploring the potential of PROCOSINE and close-range hyperspectral imaging to study the effects of fungal diseases on leaf physiology. *Scientific Reports, 8*.

Rabatel, G., Marini, F., Walczak, B., & Roger, J.-M. (2020). VSN: Variable sorting for normalization. *Journal of Chemometrics, 34*(2). e3164.

Roger, J.-M., Boulet, J.-C., Zeaiter, M., & Rutledge, D. N. (2020). *Pre-processing Methods☆. Reference module in chemistry, molecular sciences and chemical engineering*. Elsevier.

Sun, X., Subedi, P., Walker, R., & Walsh, K. B. (2020). NIRS prediction of dry matter content of single olive fruit with consideration of variable sorting for normalisation pre-treatment. *Postharvest Biology and Technology, 163*, 111140.

Vigneau, N., Ecarnot, M., Rabatel, G., & Roumet, P. (2011). Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in Wheat. *Field Crops Research, 122*(1), 25–31.