Potential of high-spectral resolution for field phenotyping in plant breeding: Application to maize under water stress

Maxime Ryckewaert, Nathalie Gorretta, Fabienne Henriot, Alexia Gobrecht, Daphné Heran, Daniel Moura, Ryad Bendoula, Jean-Michel Roger

Potential of high-spectral resolution for field phenotyping in plant breeding: application to maize under water stress

Maxime Ryckewaert[a,b], Nathalie Gorretta[a], Fabienne Henriot[c], Alexia Gobrecht[a], Daphné Héran[a], Daniel Moura[a], Ryad Bendoula[a], Jean-Michel Roger[a,b]

[a] *ITAP, Univ Montpellier, INRAE, Institut Agro, Montpellier, France*
[b] *ChemHouse Research Group, Montpellier, France*
[c] *Limagrain Europe, Chappes, France*

Abstract

Spectroscopy is today and for two decades strongly used in many fields (pharmacy, agriculture, process, medicine...). This use in a very large number of applications is linked to the great spectral richness of the measurement and therefore to the large amount of accessible chemical information. For plant breeding, spectral reflectance in the visible and near-infrared range (VIS-NIR) embeds a lot of information about vegetation (pigments, structure, water, etc.). Discriminatory power between genotypes can be greatly improved by using high spectral resolution. NIR spectroscopy is still limited in the field for phenotyping compared to existing imaging solutions that are easier to implement.

In this study, we will address the potential of high spectral resolution data by using NIR spectroscopy to describe phenotypic responses of maize genotypes to water stress. To that end, data acquired following an experimental design with water-deficient environment are processed using an analysis of

variance method adapted to multivariate data called REP-ASCA. For each factor, this method gives its significance, the loadings describing the impacted spectral regions and the scores to classify observations. For a date with proven water stress, the treatment and genotype factors and the interaction term are significant with a p-value threshold at 0.05. Treatment term loadings highlight the spectral regions impacted by the change in irrigation while those of the genotype factor allows to group genotypes according to the yield potential regardless the irrigation. The interaction term loadings are used as a phenotyping trait related to water stress response. Based on this signature, tolerant genotypes are differentiated from sensitive genotypes according to a ranking based on final yield (R = 0.81). This spectral signature was then applied to another environment with a moderate water deficit. For most genotypes, we were able to recover the ranking previously established by the stressed environment (R = 0.60).

Keywords:

Plant Breeding, Phenotyping, Spectroscopy, REP-ASCA, Drought, Water Stress, Maize, Multivariate Data, Chemometrics

---

1. Introduction

In an agricultural context, drought is defined as any lack of water that prevents crops to reach their yield potential or that affects quality of harvested products. This stress induces changes in yield and quality of many crops around the world. Periods of drought are increasingly frequent and severe and require the development of new genotypes adapted to extreme water stress conditions [1, 2, 3].

Furthermore, stress will have a different impact depending on the development stage at which it occurs [4, 5, 6]. For example, water stress will have less impact on the plant when it occurs during the vegetative stage than the flowering stage. In order to select genotypes according to their ability to respond to this stress, it is necessary to have rapid and efficient phenotyping tools to characterize plants in many situations. Therefore, efforts to improve high-throughput phenotyping tools must be made [7, 8].

Technologies with high spectral resolution are ideal for plant phenotyping [9, 10, 11]. High spectral resolution in the visible and near-infrared range (VIS-NIR) provides rich information on bio-chemical content and plant structure [12]. Indeed, the use of highly resolved spectral data has shown its interest in crop monitoring [13, 14, 9] or in early detection of biotic or abiotic stresses [15, 16, 17]. Additionally, some studies have shown the interest to use high spectral resolution in plant breeding, particularly for yield prediction [18, 19].

In recent years, technologies tend to jointly increase spectral and spatial resolution with the use of hyperspectral cameras for phenotyping crops [20]. These technologies still present operational constraints for agricultural applications (cost, weight, ease of use or acquisition time). On the other hand, less expensive solutions such as micro-spectrometers provide a high spectral resolution at the expense of spatial resolution [21].

However, analyzing spectral data acquired on vegetation is challenging, particularly in plant breeding where the objective is to compare genotypes with similar behaviors. Besides, data collected on several experimental sites with different agronomic and pedoclimatic conditions brings an additional

difficulty. Indeed, the environment has a strong influence on the expression of a large quantity of genes [22, 23] and hence on the phenotypic traits measured. These effects must be taken into account to ensure the representativeness of measured phenotypic traits regardless of the environment and the development state of the plant.

In chemometrics, methods can be used to exploit spectral data [24, 25]. The choice of the method to be used depends on the objective of analysis of the spectral data. We can dissociate four cases of use of spectral data in the order of the most commonly used. First, many applications of NIR spectroscopy are aimed at predicting biochemical variables from spectra. In this case, the best known and most commonly used method is partial least squares regression (PLSR) [26]. A second use case is the use of spectra to predict a class which combines partial least square (PLS) and discriminant analysis (DA) [27]. A third use case is the exploration of spectral data through unsupervised approaches such as Principal Component Analysis (PCA) [28]. The last case, very little used but nevertheless very useful when performing experiments, is the use of methods for the analysis of variance of spectral data.

Experimental design, often called Design of Experiments (DoE) [29] and ANalysis Of VAriance (ANOVA)[30] are commonly used in plant breeding [31]. On the one hand, DoE is a way of organizing experiments so that they allow genotypes to be tested optimally against several variability factors. On the other hand, the analysis of variance is a statistical method to separate different factors of variability.

Usually, methods of analysis of variance are not adapted to spectral data.

Indeed, variables are highly correlated with each other and normality assumptions are not satisfied [32]. The most commonly used method is Analysis of variance - Simultaneous Component Analysis (ASCA) [33] which is widely used in data analysis of laboratory experiments.

In the case of spectral measurements made in the field, errors related to the lack of repeatability of the measurements may affect analysis conclusions. An analysis of variance method called Reduction of repeatability error for Analysis of variance-Simultaneous Component Analysis (REP-ASCA) [34] has recently been developed to reduce this repeatability error and to describe identified factors. REP-ASCA also highlights spectral regions associated with each factor and reduces uncontrolled effects (such as leaf angle, sunlight, temperature, ...).

In this study, we are assessing the potential of high spectral resolution to phenotype different maize genotypes, specifically in the context of adaptation to drought. We will show how REP-ASCA approach applied to spectral data collected in the field can identify spectral signatures as phenotypic traits to describe genotype responses to water stress. For this purpose, two experimental campaigns were set up to study the response of different maize genotypes.

In this article, we will first describe the two experimental campaigns. Then, REP-ASCA will be used to analyze spectra collected in the highest drought environment to obtain spectral signatures related to the different studied factors. Finally, the relevance of these signatures will be discussed in the second environment.

2. Materials and methods

2.1. Field experiments

Two experimental campaigns were conducted on Limagrain experimental sites. The first one took place in 2017 in Aubiat (France) and the second one in 2018 in Nérac (France). The same experimental design was used for both campaigns (Table 1).

Table 1: Summary of the experimental conditions in 2017 and 2018.

Year	Location	Irrigation	Number of genotypes	Number of replicates	Plant density
2017	Aubiat	Irrigated	10 Genotypes	2 Replicates	9.6 Plants/m^2
2017	Aubiat	Rainfed	10 Genotypes	2 Replicates	8.5 Plants/m^2
2018	Nérac	Irrigated	10 Genotypes	2 Replicates	8.5 Plants/m^2
2018	Nérac	Rainfed	10 Genotypes	2 Replicates	8.5 Plants/m^2

This design was organized in a complete and balanced randomized block design to compare ten commercial genotypes under two irrigation conditions. The irrigation conditions or treatments were as follows: one under optimal irrigation conditions with irrigation triggering and another without irrigation, i.e. rainfed.

Ten commercial genotypes, chosen to have contrasting tolerances to water stress, were identified by letters from A to J. Genotypes A and C were known to be highly sensitive while genotypes B, E and F were known to be tolerant. The remaining genotypes (D, G, H, I and J) were selected for their average behavior to water stress. These behaviors differed according to water stress severity. This experimental design had two replicates (two micro-plots) of the same genotype per treatment. For both experimental campaigns, forty micro-plots were sowed in four rows and 6m-long each. In Nérac (2018)

the density was 8.5 plants/m^2 for both treatments. In Aubiat (2017), the densities were 8.5 plants/m^2 and 9.6 plants/m^2 respectively under rainfed and optimal irrigation conditions.

2.1.1. Meteorological data

Temperature, sunshine, rainfall and humidity were locally measured using meteorological stations installed in the fields. In addition, tensiometers (Watermark probes) were installed to describe water reserves for both irrigation conditions. Weather data were collected every hour whereas tensiometric data were collected four times a day.

Besides, vapour pressure deficit (VPD) was used to describe the evaporative demand [35].

2.1.2. Agronomic data

Several agronomic data were collected during plant growing and at harvest: flowering dates, plants counts, grain yield and moisture. All plots were harvested with a twin-plot combine DP4000 BAURAL, collecting grain yield and grain moisture. Grain moisture was measured by NIR (POLYTECH instrument). Grain yield was determined by weighing all the grains in each plot, adjusted to 15% grain moisture and converted to tons per hectare (q/ha), considering the differences in the plot size across trials. Flowering dates were measured as the number of days from June 1st, to the time when silks had emerged on 50% of the plants in a plot. Plant counts were measured as the numbers of plants after emergence. The micro-plot qualitative scorings were measured throughout the period of the experiment in order to eliminate plots where issues other than genotypic related occurred.

2.1.3. Spectral measurements

Spectral measurements were performed with a device designed specifically for this study to collect data over the maize canopy.

This device was designed to address the UAV specifications (Fig. 1). It included a spectrometer (MMS1, Carl Zeiss Spectroscopy GmbH, Germany) with 256 spectral bands defined in a spectral range from 310 nm to 1100 nm and a spectral resolution of 3 nm, an integrating sphere (30-REFL, AvaSphere), a microcomputer (pico-ITX, VIA) and three optical fibers (550 µm-core diameter, numerical aperture of 0.22, Sedi-ATI). Each spectrometer measurement S was normalized by the corresponding integration time t_s to have the signal I as follows:

$$I = S/t_s \tag{1}$$

The incident sunlight I_0 was measured with the integrating sphere connected to the spectrometer. Light reflected by the vegetation I_v was measured at a distance of approximately 1 m above the canopy corresponding to an imprint of a 16 cm-diameter circle. Dark measurements I_d were performed with a shutter in front of the spectrometer to subtract electronic noise from signals. A manual switch was used to toggle between the measurement of incident radiation and canopy radiation with the same spectrometer (Fig. 1).

Finally, reflectance spectrum was obtained according to the following equation:

$$R = \frac{I_v - I_d}{I_0 - I_d} \tag{2}$$

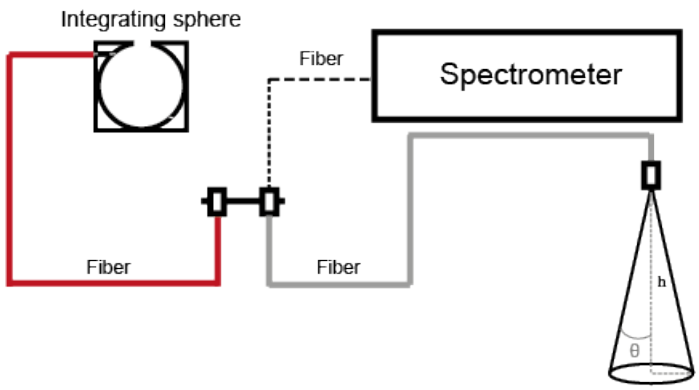


Figure 1: Diagram of the acquisition system; A manual switch was connected to the spectrometer (dot-line), to the integrating sphere (red) and a fiber (grey) for measuring light from vegetation with a numerical aperture ($\theta = 0.22$)

A wheelbarrow as support vector (Fig. 2) was designed to embed the spectral acquisition system. The height being adjustable, it was set at the beginning of each acquisition date at a height of approximately 1 meter from the top of the canopy.

2.1.4. Protocol for spectral field measurements

Spectral measurements were performed on 6 dates per year using the device presented in section 2.1.3. Measurements were made on one of the two central rows of each micro-plot to avoid border effects from the two adjacent genotypes. Acquisition path was defined according to the orientation of the field and the acquisition time in order to minimize shadows caused by the measuring device. A reference measurement with the integrating sphere was acquired systematically for each measurement on vegetation. At the end, 12 spectral reflectances were obtained per micro-plot and 480 spectra over the whole field for each acquisition date.

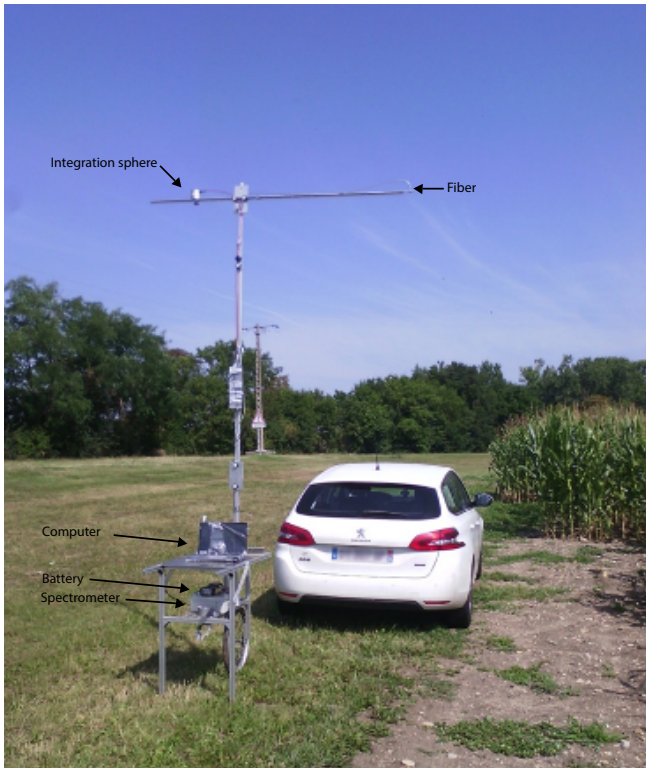


Figure 2: Image of the wheelbarrow embedding the spectral acquisition system, next to a car to visualize the height of the device.

2.2. Data analysis

2.2.1. Theory of REP-ASCA

REP-ASCA method [34] is used to analyze multivariate data associated with a design of experiments. The specificity of REP-ASCA is to reduce effects due to a lack of repeatability of measurements. As detail in [34], REP-ASCA requires a multivariate data set represented by a matrix $\mathbf{X}$ for the analysis of variance and a data set represented by a matrix $\mathbf{W}$ to describe the repeatability error. In this matrix notation, rows correspond to observations and columns to spectral bands. The matrix $\mathbf{W}$ contains centered measures

per packet of repeated measurements and carries only the information related to the repeatability error. The centered measures per packet are obtained by removing, for each spectrum, the average spectrum of the corresponding packet. This operation provides a set of observations containing within-variance.

REP-ASCA approach aims to define first a subspace representative of the repeatability error and then to perform an orthogonal projection of the multivariate data $\mathbf{X}$ to remove this subspace, as detailed in [36]. To define this repeatability error subspace, a Principal Component Analysis (PCA) is performed on $\mathbf{W}$ to give principal components. The choice of the first k components according to criteria specific to the dataset to be realized allows to define this vector subspace named here $\mathbf{D}$.

The orthogonal projection is then performed to obtain a corrected matrix $\mathbf{X}_{\perp}$. This projection can be written as follows:

$$\mathbf{X}_{\perp} = \mathbf{X}(\mathbf{I} - \mathbf{D}\mathbf{D}^{t}) \tag{3}$$

2.2.2. Application of REP-ASCA

The algorithm REP-ASCA , available at github.com/RYCKEWAERT/REP-ASCA, has been implemented with Matlab R2015b (The Mathworks, Natick, MA, USA).

No pretreatment was performed on spectra. Data acquired on the most representative date of a stressed environment were used for REP-ASCA. This dataset of 480 acquired spectra was randomly split into 2 datasets to form the two matrices $\mathbf{X}$ and $\mathbf{W}$ required for REP-ASCA. $\mathbf{X}$ was thus composed of 2/3 of the measurements, i.e. 320 spectra. The remaining measurements

were centered in packets of repeated measures to form the matrix $\mathbf{W}$. $\mathbf{W}$ was thus made up of 160 spectra.

The classical ASCA method [33] was then applied to the corrected spectral data $\mathbf{X}_\perp$, obtained after orthogonalization (eq. 3). As all analysis of variance of multivariate data [33, 34], $\mathbf{X}_\perp$ was decomposed into matrices according to the variances associated with the effects studied and their interactions, namely here genotype, treatment effects and genotype-environment interaction. This decomposition is written as follows:

$$\mathbf{X}_\perp = \mu + \mathbf{X}_G + \mathbf{X}_T + \mathbf{X}_{G\times T} + \mathbf{E} \tag{4}$$

Where μ denotes the overall mean matrix of $\mathbf{X}_\perp$. Here, $\mathbf{X}_G$, $\mathbf{X}_T$ and $\mathbf{X}_{GxT}$ are matrices corresponding to the genotype, treatment and genotype x treatment interaction effects, respectively. The matrix $\mathbf{E}$ represents the residuals. Each spectrum (line) from $\mathbf{X}$ is then decomposed as a sequence of average spectra per modality. For example, the matrix $\mathbf{X}_T$ contains for each row corresponding to an observation, either the average spectrum of the irrigated condition or the average spectrum of the non-irrigated condition.

The second step is to perform a permutation test to compute p-values to determine if the factors are significant. The last step is to reduce the dimensionality of each significant factor by simultaneous component analysis (SCA). For a given factor i, the corresponding matrix X_i is then written as :

$$\mathbf{X}_i = \mathbf{T}_i\mathbf{P}_i^t + \mathbf{R}_i \tag{5}$$

Where $\mathbf{P}_i$ is the matrix containing loadings of the principal components (PC), $\mathbf{T}_i$ the scores and $\mathbf{R}_i$ the residuals of the SCA.

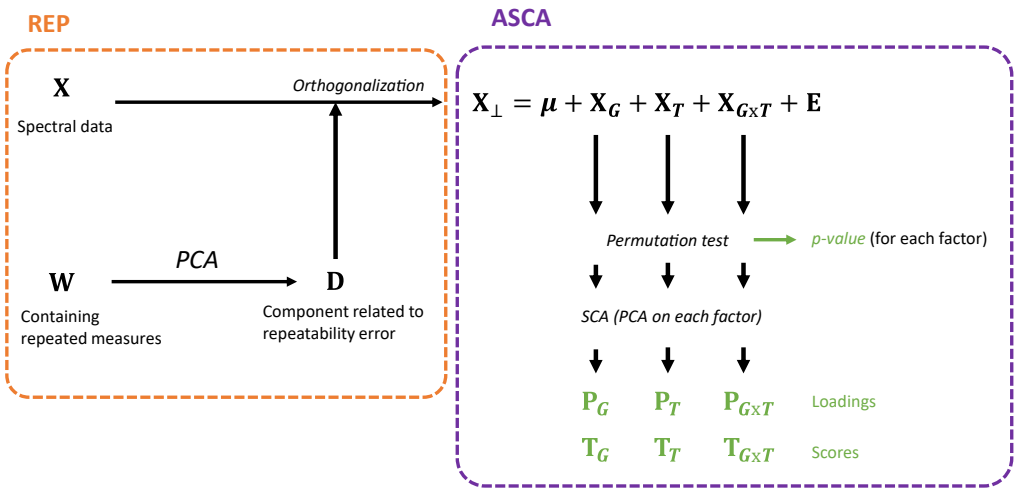


Figure 3: Flow chart to summarize REP-ASCA method.

3. Results and discussion

3.1. Description of experimental campaigns

3.1.1. Climatic conditions

Water stress results from a combination of a lack of water in the soil, high evaporative demand and high temperatures. The assessment of the water stress intensity on the micro-plots and for each date was performed using the available meteorological, i.e. measurements of maximum temperatures, tensiometry and vapour pressure deficit (VPD).

All climatic variables at the acquisition dates are providing in the table 2). In the first column, the acquisition dates are numbered from 1 to 6 in chronological order for each year for descriptive convenience.

For all other dates and with differences that do not exceed 5°C between

$T_{max,d}$ and $T_{max,-3d}$, Over this period, there were therefore no abrupt changes in temperature.

Tensions at the acquisition dates for the two treatments and differences between treatments are given in table 2. A tension value represents the force required to extract water from the soil, so the higher the tension, the drier the soil. In Aubiat (2017), tension values for both treatments are very low, with maximum tensions being reached on the third date with values equal to 73.42 and 50.08 cbar for the rainfed and irrigated treatments respectively. At this date, the difference δP is maximal with a value equal to 23.34 cbar.

In Nérac (2018), the tension value of the rainfed plot is high with a maximum value equal to 231.5 cbar at the third date. For irrigated plot, the tension value increases to 78.5 cbar that is already higher than all value in Aubiat (2017), even in irrigated condition. On this date, a difference of 153 cbar is then observed for tension values between irrigated and rainfed conditions and decreases on the following dates.

In Aubiat (2017), tensiometry and VPD values are very low. In addition, tension differential between irrigated and non-irrigated plots is very low due to the presence of significant rainfall events. These two plots are therefore under optimal water reserve conditions even if a slight differential appears between the two plots.

In Nérac (2018), the tensiometric value is high for the rainfed condition, combined with average evaporative demand and high temperatures. The values obtained in Nérac (2018) are those expected for trials under water stress conditions.

Although the maximum temperatures recorded in 2017 and 2018 are al-

most identical, rainfall and evaporative demand are different between these two environments. Aubiat (2017) has a low value for tension difference between treatments corresponding to low water stress. On the other hand, Nérac (2018) is the environment with the highest water stress on the rainfed condition.

At the measurement dates, plant transpiration is higher in Nérac (2018) than in Aubiat (2017) where rainy episodes occurred. High values correspond to a high evaporative demand. In this situation, the air is dry and increases plant transpiration.

These findings are confirmed by the tensions values. We therefore select the most contrasted dates in terms of drought intensity with the maximum differences in tension values between the two treatments.

On all available acquisition dates, the third date (07/30/18) of Nerac (2018) is the acquisition date when the water stress is the most severe. Spectral data acquired on this date is used for the REP-ASCA study. The third date (02/08/17) of Aubiat (2017) is the acquisition date with a maximal difference between treatments. This date is used as a second environment with low water stress. These two dates correspond to the grain filling stage which is one of the most sensitive stages to water stress. Indeed, at this date drought has an impact on grain filling and therefore on the thousand kernel weight (TKW) [37, 5].

3.1.2. Agronomic variables

The table 3 shows, by treatment, means and standard deviations of yield values obtained at harvest.

In 2017, average yield values for irrigated and rainfed conditions are high,

Table 2: Description of the climatic conditions for each acquisition date: Maximum temperatures $T_{max,d}$(°C) for the acquisition dates and averaged over the 3 previous days of the acquisition dates $T_{max,-3d}$ (°C), rainfed cumulated during the 3 days preceding the acquisition date R_{-3d}(mm), vapour pressure deficit (VPD)(kPA), tensiometry P(cbar) for both treatment and difference between treatments δP(cbar).

N°	Location	Date	$T_{max,d}$ (°C)	$T_{max,-3d}$ (°C)	R_{-3d} (mm)	VPD (kPA)	P (cbar)		δP (cbar)
							rainfed	irrigated	
1	Aubiat	19/06/17	34.5	30.0	6.75	1.53	5.83	5.83	0
2	Aubiat	19/07/17	28.0	33.0	6.75	0.89	13.08	11.41	1.67
3	Aubiat	02/08/17	33.5	31.0	3.5	0.95	73.42	50.08	23.34
4	Aubiat	17/08/17	29.5	30.5	0.25	0.80	43.92	26.01	17.91
5	Aubiat	01/09/17	19.5	34.5	19.8	0.31	2.92	1.42	1.5
6	Aubiat	28/09/17	25.0	24.0	18.3	0.47	1.42	1.42	0
1	Nérac	26/06/18	31.6	29.1	0	1.60	x	x	x
2	Nérac	10/07/18	28.8	31.1	0.6	1.19	145.08	x	<145.08
3	Nérac	30/07/18	30.8	34.2	0.4	1.25	231.5	78.5	153
4	Nérac	20/08/18	29.5	29.9	0.2	1.30	145.17	86.67	58.5
5	Nérac	27/08/18	29.8	30.6	4.0	1.23	229.5	121	108.5
6	Nérac	10/09/18	26.9	29.4	11.8	0.83	239	168	71

with 133.5 q/ha and 116.2 q/ha respectively. In 2018, average yield values are much lower with 104.42 q/ha for irrigated condition and 85.46 q/ha for rainfed condition. Standard deviations are of the same order of magnitude for all the experimental campaigns with values between 9 and 13 q/ha.

Obtaining lower yield values in rainfed condition compared to irrigated condition was expected. In 2017, yield values in both conditions are high. Indeed, as stress is very low (Tab. 2), yield values in rainfed condition remain high (Tab. 3).

Table 3: Average and standard deviation of the yield obtained at harvest for both irrigation conditions and for each environment.

Year	Yield (q/ha)			
	Irrigated		Rainfed	
	Average	Standard deviation	Average	Standard deviation
2017	133.5	9.36	116.2	10.85
2018	104.42	12.48	85.46	9.12

3.1.3. Genotypes classification based on yield

In our study, the percentages of yield loss between two irrigation conditions are used to describe genotypes responses to water stress. Thus, for a given genotype, the lower the percentage of loss between the two irrigation conditions, the less sensitive the genotype will be to water stress. On the contrary, a high percentage of loss between the two irrigation conditions will show a high sensitivity of the genotype to water stress. The yield value of the irrigated condition is also important. Indeed, it shows the yield potential of genotypes under optimal irrigation conditions.

In general, several criteria are needed to characterize resistance to water stress [38]. Climatic variables help to put yield values obtained into context. The 2017 environment does not correspond to a stressed environment. Therefore, it is difficult to assess genotype sensitivity to irrigation conditions in this particular environment. The year 2018 was therefore used to verify that the genotypes were correctly classified according to their sensitivity to water stress.

For each genotype, yield values and the percentage of yield loss between the two irrigation conditions for 2018 are given in table 4.

Table 4: Average yield per treatment and percentage losses for each genotype for 2018.

Genotype	Yield 2018 (q/ha)		
	Irrigated	rainfed	Loss (%)
A	103.22	83.17	19.42
B	105.85	97.01	8.34
C	115.39	88.92	22.94
D	100.82	93.30	7.45
E	78.39	77.13	1.61
F	103.63	92.49	10.74
G	118.55	86.65	26.91
H	104.16	76.29	26.75
I	114.51	81.64	28.70
J	99.68	77.99	21.75

In irrigated conditions, some genotypes have very high yields such as genotypes C and G with average values of 115.39 q/ha and 118.55 q/ha respectively. Others have very low yield values such as genotypes E and J with average values equal to 78.39 q/ha and 99.68 q/ha respectively. In rainfed conditions, genotypes with high yield values are genotypes B and D with respectively 97.01 q/ha and 93.30 q/ha. Genotypes E and J have the lowest yields with values of 77.13 q/ha and 77.99 q/ha respectively.

Genotypes that have a low percentage of losses between irrigated and rainfed conditions are genotypes B, D and E with a loss of 8.34%, 7.45% and 1.61% respectively. Very sensitive genotypes are genotypes A,C G, H, I and J with losses of 19.42%, 22.94%, 26.91%, 26.75%, 28.70% and 21.75% respectively.

Based on agronomic data, this analysis allows us to establish two classes

of genotypes: the first class includes tolerant genotypes (B, D, E and F). For these genotypes, a change in irrigation has little impact on yield. The second class is composed of sensitive genotypes (A, C, G, H, I and J). For these genotypes, yield is strongly impacted by the change in irrigation. Genotype E is atypical as it seems to reach its yield potential even in the absence of water.

3.2. Reduction of repeatability errors with REP-ASCA

3.2.1. Descriptions of the spectral measures

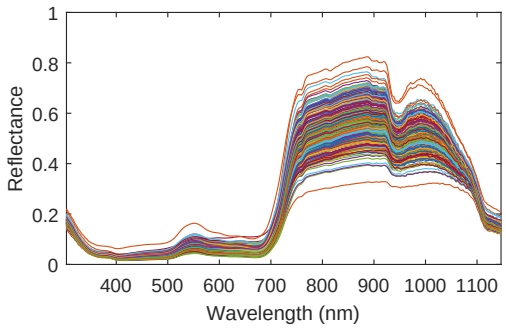


Figure 4: Spectral data for the chosen date (30/07/18 - Nérac).

Figure 4 shows spectra acquired in 2018 corresponding to characteristic spectra of vegetation: presence of specific hollows at 450 nm and 650 nm related to chlorophyll content; the slope between 700 nm and 800 nm corresponding to the red-edge; a large reflected part related to internal structure beyond 750 nm and a hollow at 950-980 nm related to water content. It can be seen that these spectra are affected by a systematic variation effect consisting of a vertical translation visible at 400 nm (additive effect) and a multiplicative effect visible between 800 and 1000 nm. This last effect is particularly visible when reflectance signal intensity is high.

Additive and multiplicative effects are classically observed on vegetation spectra [39]. Additive effects are mainly due to variations in angles formed between normal at the leaf surface and measurement axis [40] during spectral acquisition. Multiplicative effects show variations as a function of wavelength and are essentially related to the increase of optical path [40]. Additive and multiplicative effects are generally addressed by the application of standardization methods such as Standard Normal Variate (SNV) [41] and Variable Sorting for Normalization (VSN) [42]. However, these pretreatments can have deleterious effects on spectra [43, 44, 42]. For this reason, no pretreatment has been applied here.

3.2.2. Selection of the dimension k of the detrimental space

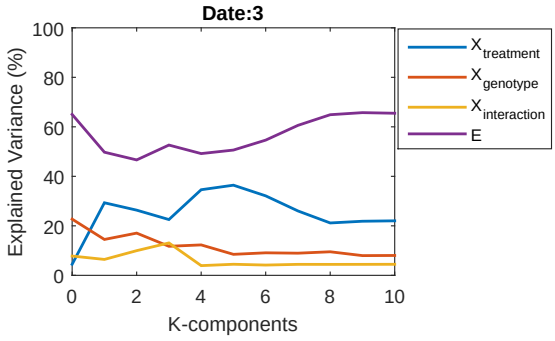


Figure 5: Impact of projections on explained variances for each factor in the analysis of variance on spectral acquired at the chosen date (30/07/18 - Nérac).

The objective is to determine the k dimension of the detrimental subspace corresponding to the repeatability errors. k must be judiciously chosen to define the detrimental subspace corresponding to the repeatability error while avoiding removing any part of the variance related to a factor of interest.

This choice can be driven by various criteria. Here we choose to look at the impact of the components number k removed by orthogonal projection on the variances carried by each factor of the data set $\mathbf{X}_{\perp}$.

Figure 5 shows the evolution of the explained percentage variance of the matrix $\mathbf{X}_{\perp}$ for each of factors studied as a function of the number k of orthogonal projections performed.

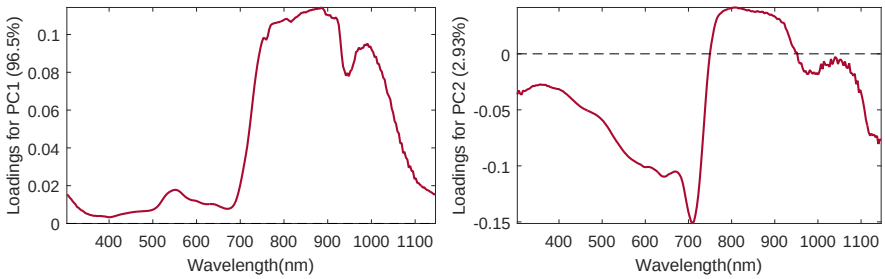
Without correction, the percentage of explained variance of the residual term is very high and reaches 65.01% of the total variance. It is always higher than other factors regardless of k. This percentage reaches a minimum of 46.63% for k=2. For values greater than 2, this percentage increases progressively to reach a maximum value of approximately 65% for k=10.

For the treatment factor, the percentage variance explained is 4.45% of the total variance when no correction is made (k=0). After orthogonal projection according to the first component (k=1), this percentage is then 29.32%. It then varies between 22% and 36% for k ranging between 2 and 10. The treatment term seems to be strongly impacted by the first component related to repeatability error. For the genotype factor, the percentage of variance explained decreases progressively when k increases from 14.5% when k=1 to 8.03% when k=10. For the interaction term, the percentage variance explained increases slightly for k=3 to reach 13.09%. But from 4 projections, it drops sharply to around 3-4% of the total variance. For a value of k greater than or equal to 4, the variance explained for the interaction term no longer changes. The projection then has a negative effect and removes the variance related to this factor. k must therefore be less than 4.

We choose a value of k that minimizes the percentage of residual variance

while maximizing those of the factors of interest. The choice was therefore k=2.

3.2.3. Loadings of the principal components related to the repeatability error



(a) (b)

Figure 6: Description of **W**: Loadings of (a) PC1, (b) PC2.

Loadings of the first component (PC1) of the matrix **W** are visible in figure 6a. They represent 96.5 percent variance in repeatability error. These loadings are all positive and are similar to the spectra obtained on vegetation (Fig. 4). This corresponds to a systematic variation corresponding with the average spectrum. It can therefore be seen that a large part of the repeatability error is due to the systematic effects described above (Fig. 4).

The loadings of the second principal component (PC2) can be seen in figure 6b. These loadings show an opposition between the visible region (300 - 700 nm) and the near-infrared one (700 - 1100 nm). The repeatability error within a micro-plot is expressed here as a difference in the slope between the pigment-sensitive visible part and the structure-sensitive near-infrared part of the cells. In addition, they show a pronounced negative spike at 709 nm. The peak corresponds to a variation in the position of the red-edge and more specifically the start of the red-edge [45].

It is known [46, 47] that the balance between the visible and near-infrared parts, as well as the position of the red-edge can vary from leaf to leaf for the same plant and obviously from one plant to another. It is therefore not surprising to find these deformations in the repeatability error because the acquisitions were done at different points in the plants row and indeed capture the plants under different angles of view.

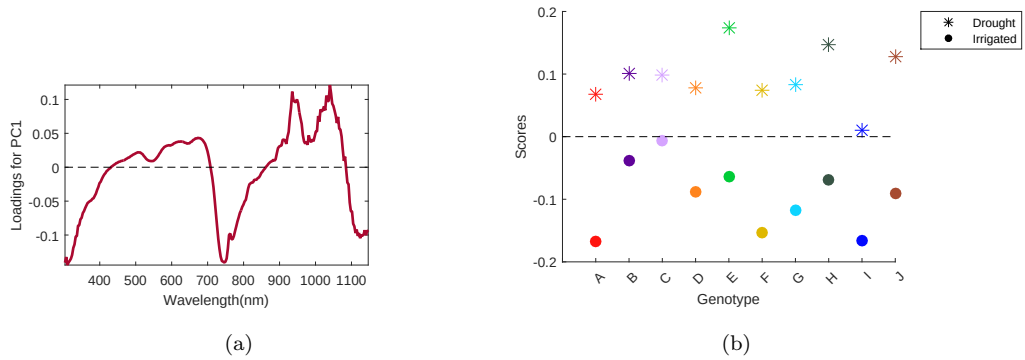
3.3. REP-ASCA results

The repeatability error of the dataset having been reduced by orthogonalization (k=2), ASCA is performed on the corrected dataset $\mathbf{X}_{\perp}$. All factors in equation 4 are significant with p-values < 0.05 according to permutation tests. ASCA thus provides loadings and scores for each term of the equation 4.

The number of principal components obtained for each term is equal to the number of levels minus one. The treatment term is a two-level factor (irrigated/rainfed). So, only one principal component is obtained. For other terms (genotype and interaction), many principal components are obtained. However, for simplicity, results on the first two components will be described.

3.3.1. Treatment term

Loadings. The loadings of the principal component are shown in figure 7. In the visible and near-infrared region, we observe two peaks with the same sign, separated by a hollow centered at 739 nm. There are also two positive peaks located at 945 nm and 1040 nm, surrounding a hollow at 970 nm. A positive slope is also visible between 300 and 420 nm. The hollow located at 739 nm corresponds to the region of the red-edge and more precisely to its inflection



(a) (b)

Figure 7: Decomposition of the treatment term : (a) Loadings of the principal component, (b) average scores obtained by genotype.

point (maximum of the first derivative). The hollow at 970 nm corresponds to the water absorption [48, 49]. The presence of the two peaks at 945 nm and 1040 nm express an enlargement of the hollow at 970 nm. The slope between 300 and 420 nm shows a difference in absorption in the ultraviolet (UV) region between the irrigated and stressed areas. In this region there are both the UV-A absorption and the beginning of the region corresponding to the UV-B absorption [50].

Changes in reflectance in this spectral region can be induced by different levels of flavonoids and phenylpropanoids. Indeed, theses chemical components absorb strongly in UV radiation [51]. UV absorption directly impacts photosynthetic activity, and so, yield [52, 50].

Scores. Average scores obtained on each genotype are proceeded by projecting their mean spectra onto this principal component. The scores obtained for each genotype are presented in figure 7b. The projection corresponds to the vector product between a spectrum measured and this component

producing a score.

The scores obtained in irrigated condition (round symbol) are mainly negative. Conversely, scores obtained in rainfed condition (star symbol) are mainly positive. However, genotype I in rainfed conditions and C in irrigated conditions have scores close to zero. Extreme positive values are also observed with genotype E (in rainfed condition) and negative with genotypes A and I in irrigated condition.

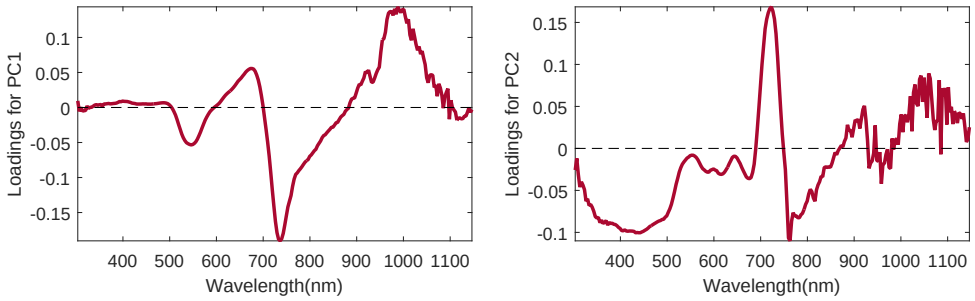
By jointly analyzing scores and loadings of this component (Fig. 7a), we can thus deduce information on traits related to genotype in response to the average behavior per treatment. With negative loadings in 739 nm, a positive score will correspond to a genotype with a low red-edge slope. And conversely, with positive loadings between 970 nm, a positive score will have high reflectance values around the dip at 970 nm. Results are reversed when scores are of different sign. For example, On the other hand, scores produced by this component will be negative when the initial spectra show a marked red-edge and a widening of the absorption peak related to water.

In general, genotypes in rainfed conditions show positive scores (Fig. 7b). Their reflectance spectrum has a less marked inflection point and also a shallower absorption dip than the same genotypes under irrigated conditions. These phenomena can express a reduction in leaf water content and changes in plant metabolism for defense mechanisms regarding drought [53].

Indeed, under water stress, the closure of stomata plays a protective role by reducing water loss and limiting gas exchanges [5, 54]. This stomatal closure could slow down the photosynthetic process of the plant leading to a reduction in yield when sporadic stress occurs [55].

The study of the treatment term by loading interpretation allow to identify the spectral regions impacted by the irrigation change. It is possible to identify atypical behaviors with respect to other genotypes tested for a given treatment. Genotypes with scores close to zero will not have the same mechanisms as other genotypes when faced a given treatment. On the other hand, genotypes with extreme score values will have more pronounced behaviors.

3.3.2. Genotype factor



(a) (b)

Figure 8: Loadings of the genotype term on (a) PC1, (b) PC2.

Loadings. Loadings of the first two principal components (PC1 and PC2) of the genotype term are shown in figure 8. On PC1 (Fig. 8a), loadings in the spectral range 300-500 nm are close to zero. There are two negative peaks located at 545 nm and 736 nm and two positive peaks located at 672 and 980 nm. A slope break at 769 nm occurs in the increasing slope visible from 736 to 980 nm. The dips at 545 nm and 736 nm correspond to the anthocyanin content and the inflection point of the red-edge, respectively. The peaks at 672 nm and 980 nm correspond to the absorption of chlorophyll and water, respectively. PC1 will therefore produce negative scores for plants

having high chlorophyll and water contents and positive scores for those with high anthocyanin contents and a less pronounced slope of the red-edge. This component therefore contrasts genotypes with good photosynthetic activity (negative scores) and genotypes with a high anthocyanin content that may result from the presence of various environmental stresses (positive scores) [56]. The latter therefore will have a lower photosynthetic capacity.

The loadings of the second principal component (PC2) (Fig. 8b) have a large negative part between 350 nm and 500 nm, a peak at 719 nm and a constant slope from 760 nm to 900 nm. The negative part of 350 nm to 500 nm relates to the region of pigment absorption, particularly carotenoids and chlorophyll pigments. The peak at 719 nm is the first part of the red-edge. This component provides additional information to PC1 and also more specific information on the carotenoid content. These pigments can protect the photosynthetic system from excess light [57, 58]. Positive scores on PC2 will reflect lower reflectance at 500 nm and therefore higher carotenoid content. Genotypes with a high carotenoid content may be selected for their ability to be tolerant to excess light [59].

Scores. The plot scores obtained on PC1 and PC2 are shown in figure 9. According to the previous study of loadings (Fig. 8a and 8b), score value per genotype is due to differences in red-edge slope and pigments contents (chlorophylls, anthocyanins and carotenoids). Genotypes B, C, E, H and J have positive scores on PC1. These genotypes and have therefore a higher anthocyanin content and a spectral response with a less pronounced red-edge slope. PC2 mainly separates genotype C and I (negative scores) from genotypes D, E, F and H (positive scores). Genotypes C and I appear to have

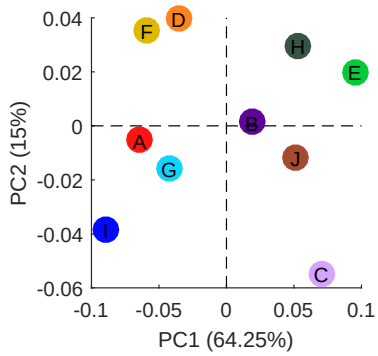
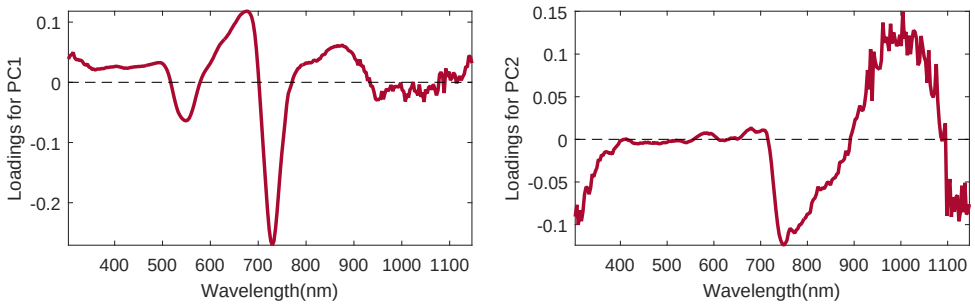


Figure 9: Genotype scores on the first two components (PC1 and PC2) of the genotype term.

higher carotenoid levels than genotypes D, F and H. Genotypes A, G, and I have a higher chlorophyll content and will therefore favor yield potential [59, 60, 61]. Based on these scores, we can describe each genotypes or see the grouping of certain genotypes, i.e. genotypes F and D, genotypes A, G and I and genotypes E and H.

3.3.3. Interaction term



(a) (b)

Figure 10: Loadings of the interaction term on: (a) PC1, (b) PC2.

Loadings. The loadings of the first two components (PC1 and PC2) are shown in figure 10. On PC1 (Fig. 10a), a dip is located at 545 nm and a hump is located at 672 nm. Another major hollow is located at 729 nm and another bump appears from 800 to 922 nm. After 950 nm, loadings values are close to zero. The first part of the loadings of this component (Fig. 10a) has a similar shape to the loadings of PC1 of the genotype term (Fig. 10a) relating to anthocyanin and chlorophyll content.

The loadings of PC2 are shown in figure 10b. Surprisingly, loadings values are equal to zero between 400 and 700 nm. After a sharp decrease between 700 to 750 nm, loading values increase and a positive slope is visible from 750 nm to 1,000 nm. PC2 carries strictly no information on the visible spectral region and thus no information on the pigment content. On the other hand, it expresses changes in the internal structure of the leaves. Indeed, changes such as changes in turgidity and intercellular space can be significant even in the presence of very low water stress [62]. In the short term, this stress can lead to wilting or leaf curling.

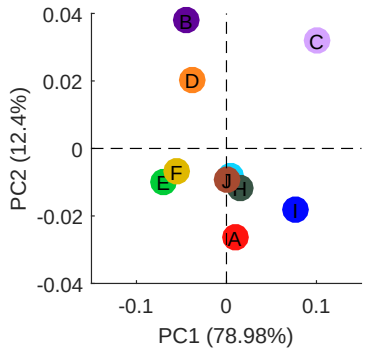


Figure 11: Scores obtained on the first two components PC1, PC2 of the interaction term.

Scores. Average scores obtained on PC1 and PC2 for each genotype are presented in figure 11. Some genotypes are grouped together: genotypes E and F have negative scores on both axes; genotypes G, H and J have scores close to zero on both axes.

According to the loadings of PC1 (Fig. 10a), negative scores for genotypes E and F are due to their good chlorophyll and anthocyanin content in irrigated conditions. On this same component, the positioning of genotype A changes very little between irrigation conditions. For this genotype, the change in irrigation condition has very little impact in the spectral regions related to anthocyanins and the position of the red-edge. For this same genotype, the score observed on PC2 is strongly negative. The difference between the two irrigation conditions then seems to be a difference in the desiccation of internal cells and therefore in the wilting of the leaves. Genotypes B, C and D have positive scores on PC2. Considering that this component reflects internal modification effects of the cells, it can be assumed that these genotypes do not undergo internal structural degradation under water stress.

3.4. Principal components used as proxies

In this part, we study correlation between scores obtained on the different components of the different factors with yield variables measured during experimental campaigns. This study will enable us to classify genotypes according to their tolerance to water stress for the 2018 experimental campaign. A classification of genotypes for the 2017 experimental campaign will then be carried out on the basis of the results of the 2018 campaign.

3.4.1. Correlation between scores and agronomic variables

The table 5 shows correlations between the agronomic variables average yield for the two irrigation conditions (rainfed/irrigated), the percentage of losses and the scores previously calculated for all the terms studied i.e. treatment, genotype and interaction.

Table 5: Correlation between yield-based agronomic variables and treatment term scores $\mathbf{T}_T$; genotype scores on PC1 ($\mathbf{T}_{G,1}$) and on PC2($\mathbf{T}_{G,2}$); and interaction term scores on PC1($\mathbf{T}_{G\text{x}T,1}$) and PC2 ($\mathbf{T}_{G\text{x}T,2}$).

agronomic variables	$\mathbf{T}_T$	$\mathbf{T}_{G,1}$	$\mathbf{T}_{G,2}$	$\mathbf{T}_{G\text{x}T,1}$	$\mathbf{T}_{G\text{x}T,2}$
Yield (irrigated)	-0.42	-0.48	-0.59	0.69	0.17
Yield (rainfed)	-0.12	-0.34	0.03	-0.2	0.71
Average yield	-0.36	-0.51	-0.38	0.41	0.47
Loss percentage (%)	-0.33	-0.24	-0.59	0.81	-0.34

For the treatment term or the genotype term, scores are weakly correlated to the agronomic variables with absolute values of the correlation coefficient below 0.60.

The highest values are obtained for the interaction term: scores on PC1 ($\mathbf{T}_{GxT,1}$) are positively correlated (R=0.69) to the yield in irrigated condition. Those on PC2 ($\mathbf{T}_{GxT,2}$) are correlated (R=0.71) to yield in rainfed condition.

This confirms that pigment content and steep red-edge slope (Fig. 10a) are related to the photosynthetic capacity of the plant and thus its yield under optimal irrigation conditions. Under rainfed conditions, variations in cell structure (degradation) and plant morphology (drying, wilting, coiling) affect yield values.

The correlation between the scores on PC1 of the interaction term ($\mathbf{S}_{GxT,1}$)

and the percentage of yield loss is highest and equal to 0.81. The study of loadings (Fig. 10a) indicate that the scores obtained for the different genotypes express the plant ability to mobilize its photosynthetic system to water stress. It is therefore possible to retrieve the sensitivity ranking of genotypes to stress previously established directly from the scores of this component. Negative scores will therefore correspond to tolerant genotypes and positive scores to stress-sensitive genotypes.

Figure 12, shows the $\mathbf{S}_{GxT,1}$ scores for all genotypes (Fig. 11) as a function of the percentage of losses. Indeed, genotypes B, D, E and F are then classified as tolerant while genotypes J, C, G, H, A and I are classified as rather sensitive. We can find this classification by using the table 4. All genotypes with negative score have a percentage of yield loss below 11%. On the other hand, genotypes with positive scores on this axis have a percentage of yield loss greater than 20%. Genotype E is very atypical, as it is extremely tolerant with a yield loss of less than 2% but this genotype has also a very low yield potential. On the other hand, genotypes B and D are well dissociated from genotypes E and F on the second principal component. These two pairs are therefore dissociated from each other in terms of leaf structure (Fig. 10b).

3.4.2. Application to another environment

Components obtained by REP-ASCA to the data acquired during the 2018 campaign were applied on another environment. The objective is to study the influence of the environment on the scores obtained. As previously mentioned, there was no significant water stress on the 2017 experimental campaign. We will nevertheless apply the principal component of the inter-

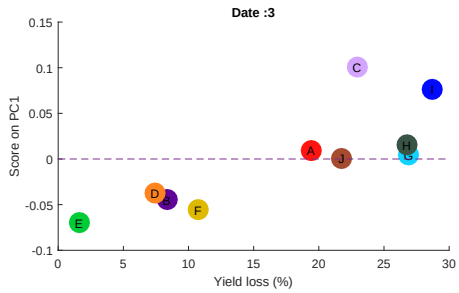


Figure 12: Scores according to the percentages of yield loss.

action term obtained in 2018 to data collected in 2017.

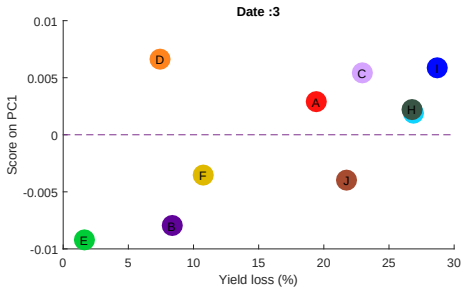


Figure 13: Scores by genotype obtained for 2017 observations based on percentage yield loss in 2018.

In 2017, the water reserve of the rainfed part is slightly reduced without any significant impact on yield. Percentages of yield loss obtained in 2018 are then used to provide a yield-based ranking describing genotypes responses to water stress.

When comparing genotype rankings obtained for the 2018 (Fig. 12) and 2017 (Fig. 13) data, we find strong similarities in behavior. Thus, genotypes B, E and F scored negative in both environments and are classified as tolerant. Genotypes A, C, G, H and I have positive scores in both environments and are classified as sensible to water stress. These similarities are very en-

couraging and show that the loadings obtained from an environment with high water stress can be applied to another environment with lower water stress.

Only genotypes D and J do not meet this classification. It can be hypothesized that these genotypes do not put in place responses identified by the components on the different factors when the intensity of the stress is not sufficient. The spectral signatures obtained on these genotypes in this low water stress environment are not affected in the same way when they are under intense stress.

Although water stress was very limited during the 2017 campaign, we were able to link the scores obtained on this environment with the ranking established on a water-stressed environment. The scores obtained on this component are positively correlated (R = 0.60) with their ability to respond to water stress.

This demonstrates that the method described is appropriate to identify differential responses of genotypes to water stress even in a situation where this stress is ad hoc, without a major impact on yield. This approach offers a promising solution for phenotyping a larger number of genotypes for their drought tolerance without implementing a specific and costly experimental device.

4. Conclusion

In this study, we investigated the added value of using high spectral resolution to describe genotype responses to water stress using the REP-ASCA method. Two experimental campaigns were conducted in 2017 and 2018

where spectra were acquired from a panel of genotypes with different sensitivities to water stress. The 2018 season having a significant difference between the two irrigation conditions was used as a reference study of water stress situation. The spectra acquired during this campaign were analyzed by REP-ASCA. This method reduces errors due to lack of repeatability and provides scores and loadings for each of the identified terms.

We demonstrate that the treatment term loadings highlight the spectral regions impacted by the change in irrigation. When looking at yield potential regardless of treatment, genotypes can be grouped according to the loadings provided by the genotype term. And finally, we found out that the interaction term provides loadings used as new phenotyping traits describing genotype response to stress.

By projecting the observed spectra on these new components, we obtain scores that are useful to describe genotypes. The scores obtained for the interaction term were directly related to the percentage of the yield loss between the two irrigation conditions in a stressed environment. By applying these same components to spectra acquired on the same genotypes but in an environment with moderate water stress, we find the same classification for the majority of genotypes. This component seems to be best expressed when water reserves are limited, but is still relevant in the characterization of genotypes when the stress is light. This shows that REP-ASCA provides robust components applicable to environments with occasional stresses that do not impact yield. The analysis of scores simultaneously with loadings highlights the different strategies used by genotypes to manage water stress. Genotypes without adaptive mechanisms can suffer serious damage in terms

of growth, development and thus yield.

This study shows that the use of high-spectral resolution data, when linked to a method reducing repeatability error, is interesting to classify the behavior of maize genotypes to water stress.

Using NIR spectroscopy could be a preliminary study. A variable selection step on the spectral signatures obtained with REP-ASCA could be performed. New spectral indices could be created from a targeted breeding objective. While increasing the number of spectral bands, we need to be cautious about the possible redundancy of the spectral information. The use of high spectral resolution may be less useful in other applications, where spectral indices based on low resolution have proven to be very effective. But, provided that the right precautions are taken when processing the data, as here with REP-ASCA, high spectral resolution will be able to do as well as spectral indices, since the latter can be retrieved from high resolution information.

However, it would be interesting to build up a larger spectral database containing more genotypes and several drought typology environments. In addition, increasing the number of measurement dates for an experimental campaign would make it possible to monitor genotypes by focusing on their resilience to occasional water stress.

This study focused on identifying responses of different maize genotypes to water stress. The approach combining the REP-ASCA method with spectral data is adapted to other breeding objectives (diseases, hot stress or nitrogen deficiency) or even other crops.

References

John Passioura. The drought environment: physical, biological and agricultural perspectives. 2006. URL http://sharif.edu/~ghodsi/PaP/erl212v1.pdf.

David B. Lobell, Wolfram Schlenker, and Justin Costa-Roberts. Climate trends and global crop production since 1980. *Science*, 333(6042):616–620, 2011. URL http://science.sciencemag.org/content/333/6042/616.short.

Ed Hawkins, Thomas E. Fricker, Andrew J. Challinor, Christopher AT Ferro, Chun Kit Ho, and Tom M. Osborne. Increasing influence of heat stress on French maize yields from the 1960s to the 2030s. 2012.

R. Okono. I. 1 Phenotyping drought-stressed crops: key concepts, issues and approaches.

Muhammad Aslam, Muhammad Amir Maqbool, and Rahime Cengiz. *Drought Stress in Maize (Zea mays L.)*. SpringerBriefs in Agriculture. Springer International Publishing, Cham, 2015. ISBN 978-3-319-25440-1 978-3-319-25442-5. doi: 10.1007/978-3-319-25442-5. URL http://link.springer.com/10.1007/978-3-319-25442-5.

Jose L. Araus, C. Sanchez, and Gregory O. Edmeades. Phenotyping maize for adaptation to drought. *Drought phenotyping in crops: from theory to practice*, 1:263–283, 2011. URL https://books.google.fr/books?hl=fr&lr=&id=zRApAwAAQBAJ&oi=fnd&pg=PA138&dq=Phenotyping+for+drought+maize&ots=bBst_LMxMM&sig=hIKPXhHiPSK91cqLaixHR2MAROg.

J. B. Passioura. Phenotyping for drought tolerance in grain crops: when is it useful to breeders? *Functional Plant Biology*, 39(11):851–859, 2012. URL http://www.publish.csiro.au/?paper=FP12079.

Roberto Tuberosa. Phenotyping for drought tolerance of crops in the genomics era. 2012.

Nathalie Vigneau, Martin Ecarnot, Gilles Rabatel, and Pierre Roumet. Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in Wheat. *Field Crops Research*, 122(1):25–31, April 2011. ISSN 03784290. doi: 10.1016/j.fcr.2011.02.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0378429011000451.

Llorenç Cabrera-Bosquet, José Crossa, Jarislav von Zitzewitz, María Dolors Serret, and José Luis Araus. High-throughput phenotyping and genomic selection: The frontiers of crop breeding converge. *Journal of integrative plant biology*, 54(5):312–320, 2012.

José Luis Araus and Jill E. Cairns. Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Science*, 19(1):52–61, January 2014. ISSN 13601385. doi: 10.1016/j.tplants.2013.09.008. URL http://linkinghub.elsevier.com/retrieve/pii/S1360138513001994.

Stéphane Jacquemoud, Wout Verhoef, Frédéric Baret, Cédric Bacour, Pablo J. Zarco-Tejada, Gregory P. Asner, Christophe François, and Susan L. Ustin. PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment*, 113:S56–S66, Septem-

ber 2009. ISSN 00344257. doi: 10.1016/j.rse.2008.01.026. URL http://linkinghub.elsevier.com/retrieve/pii/S0034425709000765.

Martin Ecarnot, Frédéric Compan, and Pierre Roumet. Assessing leaf nitrogen content and leaf mass per unit area of wheat in the field throughout plant cycle with a portable spectrometer. *Field Crops Research*, 140:44–50, January 2013. ISSN 03784290. doi: 10.1016/j.fcr.2012.10.013. URL http://linkinghub.elsevier.com/retrieve/pii/S0378429012003486.

G. R. Mahajan, R. N. Pandey, R. N. Sahoo, V. K. Gupta, S. C. Datta, and Dinesh Kumar. Monitoring nitrogen, phosphorus and sulphur in hybrid rice (Oryza sativa L.) using hyperspectral remote sensing. *Precision Agriculture*, December 2016. ISSN 1385-2256, 1573-1618. doi: 10.1007/s11119-016-9485-2. URL http://link.springer.com/10.1007/s11119-016-9485-2.

Anne-Katrin Mahlein, Ulrike Steiner, Christian Hillnhütter, Heinz-Wilhelm Dehne, and Erich-Christian Oerke. Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant methods*, 8(1):1, 2012. URL http://plantmethods.biomedcentral.com/articles/10.1186/1746-4811-8-3.

Jan Behmann, Jörg Steinrücken, and Lutz Plümer. Detection of early plant stress responses in hyperspectral images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93:98–111, 2014. URL http://www.sciencedirect.com/science/article/pii/S092427161400094X.

Christoph Römer, Mirwaes Wahabzada, Agim Ballvora, Francisco Pinto,

Micol Rossini, Cinzia Panigada, Jan Behmann, Jens Léon, Christian Thurau, Christian Bauckhage, and others. Early drought stress detection in cereals: simplex volume maximisation for hyperspectral image analysis. *Functional Plant Biology*, 39(11):878–890, 2012. URL http://www.publish.csiro.au/?paper=FP12060.

Osval A. Montesinos-López, Abelardo Montesinos-López, José Crossa, Gustavo los Campos, Gregorio Alvarado, Mondal Suchismita, Jessica Rutkoski, Lorena González-Pérez, and Juan Burgueño. Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods*, 13(1):4, 2017. URL https://plantmethods.biomedcentral.com/articles/10.1186/s13007-016-0154-2.

Fernando M. Aguate, Samuel Trachsel, Lorena González Pérez, Juan Burgueño, José Crossa, Mónica Balzarini, David Gouache, Matthieu Bogard, and Gustavo de los Campos. Use of Hyperspectral Image Data Outperforms Vegetation Indices in Prediction of Maize Yield. *Crop Science*, 57(5):2517, 2017. ISSN 0011-183X. doi: 10.2135/cropsci2017.01.0007. URL https://dl.sciencesocieties.org/publications/cs/abstracts/57/5/2517.

Puneet Mishra, Mohd Shahrimie Mohd Asaari, Ana Herrero-Langreo, Santosh Lohumi, Belén Diezma, and Paul Scheunders. Close range hyperspectral imaging of plants: A review. *Biosystems Engineering*, 164:49–67, December 2017. ISSN 1537-5110. doi: 10.1016/j.biosystemseng.2017.09.009. URL https://www.sciencedirect.com/science/article/pii/S1537511017302635.

Krzysztof B. Beć, Justyna Grabska, Heinz W. Siesler, and Christian W. Huck. Handheld near-infrared spectrometers: Where are we heading? *NIR news*, 31(3-4):28–35, 2020. Publisher: SAGE Publications Sage UK: London, England.

John Milton Poehlman and D. A. Sleper. *Breeding field crops*. Iowa State University Press, Ames, 4th ed edition, 1995. ISBN 978-0-8138-2427-7.

K. E. Basford and M. Cooper. Genotype×environment interactions and some considerations of their implications for wheat breeding in Australia This review is one of a series commissioned by the Advisory Committee of the Journal. *Australian Journal of Agricultural Research*, 49(2):153, 1998. ISSN 0004-9409. doi: 10.1071/A97035. URL http://www.publish.csiro.au/?paper=A97035.

Richard G. Brereton, Jeroen Jansen, João Lopes, Federico Marini, Alexey Pomerantsev, Oxana Rodionova, Jean Michel Roger, Beata Walczak, and Romà Tauler. Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools. *Analytical and Bioanalytical Chemistry*, 409(25):5891–5899, 2017.

Federico Marini, Dalene de Beer, Nico A. Walters, André de Villiers, Elizabeth Joubert, and Beata Walczak. Multivariate analysis of variance of designed chromatographic data. A case study involving fermentation of rooibos tea. *Journal of Chromatography A*, 1489:115–125, March 2017. ISSN 00219673. doi: 10.1016/j.chroma.2017.02.007. URL http://linkinghub.elsevier.com/retrieve/pii/S0021967317302133.

Svante Wold, Michael Sjöström, and Lennart Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2):109–130, 2001.

Matthew Barker and William Rayens. Partial least squares for discrimination. *Journal of Chemometrics*, 17(3):166–173, 2003. ISSN 1099-128X. doi: 10.1002/cem.785. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cem.785. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cem.785.

Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

Ronald Aylmer Fisher. *The design of experiments*. Oliver And Boyd; Edinburgh; London, 1937.

Lars Stahle and Svante Wold. Multivariate Analysis of Variance (MANOVA). page 15.

Rodney Alistair Kempton, Paul N. Fox, and Manuela Cerezo. *Statistical methods for plant variety evaluation*. Springer Science & Business Media, 2012.

Marti Anderson and Cajo Ter Braak. Permutation tests for multi-factorial analysis of variance. *Journal of statistical computation and simulation*, 73(2):85–113, 2003. URL http://www.tandfonline.com/doi/abs/10.1080/00949650215733.

A. K. Smilde, J. J. Jansen, H. C. J. Hoefsloot, R.-J. A. N. Lamers, J. van der Greef, and M. E. Timmerman. ANOVA-simultaneous

component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*, 21(13):3043–3048, July 2005. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/bti476. URL https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bti476.

Maxime Ryckewaert, Nathalie Gorretta, Fabienne Henriot, Federico Marini, and Jean-Michel Roger. Reduction of repeatability error for analysis of variance-Simultaneous Component Analysis (REP-ASCA): Application to NIR spectroscopy on coffee sample. *Analytica Chimica Acta*, 1101:23–31, March 2020. ISSN 00032670. doi: 10.1016/j.aca.2019.12.024. URL https://linkinghub.elsevier.com/retrieve/pii/S0003267019314606.

Richard G. Allen, Luis S. Pereira, Dirk Raes, and Martin Smith. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome*, 300(9):D05109, 1998.

Jean-Michel Roger and Jean-Claude Boulet. A review of orthogonal projections for calibration. *Journal of Chemometrics*, page e3045, June 2018. ISSN 08869383. doi: 10.1002/cem.3045. URL http://doi.wiley.com/10.1002/cem.3045.

R. Chapuis, C. Delluc, R. Debeuf, F. Tardieu, and C. Welcker. Resiliences to water deficit in a phenotyping platform and in the field: How related are they in maize? *European Journal of Agronomy*, 42:59–67, October 2012. ISSN 11610301. doi: 10.1016/j.eja.2011.12.006. URL https://linkinghub.elsevier.com/retrieve/pii/S1161030111001444.

E. Farshadfar and J. Sutka. Screening drought tolerance criteria in maize. *Acta Agronomica Hungarica*, 50(4):411–416, 2002.

Sylvain Jay, Ryad Bendoula, Xavier Hadoux, Jean-Baptiste Féret, and Nathalie Gorretta. A physically-based model for retrieving foliar biochemistry and leaf orientation using close-range imaging spectroscopy. *Remote Sensing of Environment*, 177:220–236, May 2016. ISSN 00344257. doi: 10.1016/j.rse.2016.02.029. URL http://linkinghub.elsevier.com/retrieve/pii/S0034425716300566.

Nathalie Vigneau. *Potentiel de l'imagerie hyperspectrale de proximité comme outil de phénotypage: application à la concentration en azote du blé*. PhD thesis, Montpellier, SupAgro, 2010. URL http://www.theses.fr/2010NSAM0026.

R. J. Barnes, M. S. Dhanoa, and Susan J. Lister. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43(5):772–777, July 1989. ISSN 0003-7028, 1943-3530. doi: 10.1366/0003702894202201. URL http://journals.sagepub.com/doi/10.1366/0003702894202201.

Gilles Rabatel, Federico Marini, Beata Walczak, and Jean-Michel Roger. VSN: Variable sorting for normalization. *Journal of Chemometrics*, 34(2), February 2020. ISSN 0886-9383, 1099-128X. doi: 10.1002/cem.3164. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cem.3164.

Yiming Bi, Liang Tang, Peng Shan, Qiong Xie, Yong Hu, Silong Peng, Jie Tan, and Changwen Li. Interference correction by extracting the in-

formation of interference dominant regions: Application to near-infrared spectra. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 129:542–550, August 2014. ISSN 13861425. doi: 10.1016/j.saa.2014.03.080. URL https://linkinghub.elsevier.com/retrieve/pii/S1386142514004910.

Yiming Bi, Kailong Yuan, Weiqiang Xiao, Jizhong Wu, Chunyun Shi, Jun Xia, Guohai Chu, Guangxin Zhang, and Guojun Zhou. A local preprocessing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. *Analytica Chimica Acta*, 909:30–40, February 2016. ISSN 00032670. doi: 10.1016/j.aca.2016.01.010. URL https://linkinghub.elsevier.com/retrieve/pii/S000326701630054X.

Frank Boochs, K. Dockter, Gunther Kupfer, and W. Kuhbauch. Red edge shift as vitality indicator for plants. In *Proceedings of the 16th Congress of the International Society for Photogrammetry and Remote Sensing, held in Kyoto, Japan, on*, pages 1–10. Committee of the 16th International Congress for Photogrammetry and Remote . . . , 1988.

H. W. Gausman and W. A. Allen. Optical parameters of leaves of 30 plant species. *Plant Physiology*, 52(1):57–62, 1973.

H. W. Gausman, W. A. Allen, and D. E. Escobar. Refractive index of plant cell walls. *Applied optics*, 13(1):109–111, 1974.

B. G. Osborne, T. Fearn, and P. T. Hindle. *Practical NIR spectroscopy with applications in food and beverage analysis*. Addison-Wesley Longman Ltd,

Harlow UK, 1993. ISBN 978-0-582-09946-3. URL http://discovery.ucl.ac.uk/267166/.

Susan L. Ustin, David Riaño, and E. Raymond Hunt. Estimating canopy water content from spectroscopy. *Israel Journal of Plant Sciences*, 60 (1):9–23, December 2012. ISSN 0792-9978. doi: 10.1560/IJPS.60.1-2.9. URL http://www.sciencefromisrael.com/openurl.asp?genre=article&id=doi:10.1560/IJPS.60.1-2.9.

J. F. Bornman, S. Reuber, Y. P. Cen, and G. Weissenböck. Ultraviolet radiation as a stress factor and the role of protective pigments. In *SEMINAR SERIES-SOCIETY FOR EXPERIMENTAL BIOLOGY*, volume 64, pages 157–170. Cambridge University Press, 1997.

H. W. Gausman, R. R. Rodriguez, and D. E. Escobar. Ultraviolet Radiation Reflectance, Transmittance, and Absorptance by Plant Leaf Epidermises 1. *Agronomy Journal*, 67(5):720–724, 1975. Publisher: American Society of Agronomy.

Alan H. Teramura. Effects of ultraviolet-B radiation on the growth and yield of crop plants. *Physiologia Plantarum*, 58(3):415–427, July 1983. ISSN 1399-3054. doi: 10.1111/j.1399-3054.1983.tb04203.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1399-3054.1983.tb04203.x.

Tatjana Kavar, Marko Maras, Marjetka Kidrič, Jelka Šuštar Vozlič, and Vladimir Meglič. Identification of genes involved in the response of leaves of Phaseolus vulgaris to drought stress. *Molecular Breeding*, 21

(2):159–172, February 2008. ISSN 1380-3743, 1572-9788. doi: 10.1007/s11032-007-9116-8. URL http://link.springer.com/10.1007/s11032-007-9116-8.

M Chaves, João Maroco, and Joao Pereira. *Understanding plant responses to drought - From genes to the whole plant*, volume 30. January 2003. doi: 10.1071/FP02076.

M. M. Chaves, J. Flexas, and C. Pinheiro. Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of botany*, 103(4):551–560, 2009.

Linda Chalker-Scott. Environmental Significance of Anthocyanins in Plant Stress Responses. *Photochemistry and Photobiology*, 70(1):1–9, July 1999. ISSN 1751-1097. doi: 10.1111/j.1751-1097.1999.tb01944.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1751-1097.1999.tb01944.x/abstract.

Barbara Demmig-Adams. Carotenoids and photoprotection in plants: A role for the xanthophyll zeaxanthin. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1020(1):1–24, October 1990. ISSN 0005-2728. doi: 10.1016/0005-2728(90)90088-L. URL http://www.sciencedirect.com/science/article/pii/000527289090088L.

Anatoly A. Gitelson, Yoav Zur, Olga B. Chivkunova, and Mark N. Merzlyak. Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. *Photochemistry and Photobiology*, 75(3):272–281, 2002.

Mohammad Pessarakli, editor. *Handbook of plant and crop stress*. Books in soils, plants, and the environment. M. Dekker, New York, 2nd ed., rev. and expanded edition, 1999. ISBN 978-0-8247-1948-7.

Houman Homayoun, Morteza Sam Daliri, and Parisa Mehrabi. Effect of drought stress on leaf chlorophyll in corn cultivars (Zea mays). *Middle-East J Sci Res*, 9(3):418–420, 2011. URL http://idosi.org/mejsr/mejsr9(3)11/19.pdf.

Pablo J. Zarco-Tejada, John R. Miller, G. H. Mohammed, Thomas L. Noland, and P. H. Sampson. Vegetation stress detection through chlorophyll a+ b estimation and fluorescence effects on hyperspectral imagery. *Journal of environmental quality*, 31(5):1433–1441, 2002.

J. Levitt and R. Ben Zaken. Effects of Small Water Stresses on Cell Turgor and Intercellular Space. *Physiologia Plantarum*, 34(3):273–279, July 1975. ISSN 0031-9317, 1399-3054. doi: 10.1111/j.1399-3054.1975.tb03835.x. URL http://doi.wiley.com/10.1111/j.1399-3054.1975.tb03835.x.